The identification of novel long non-coding RNAs and immune-related long non-coding RNAs in triple negative breast cancers

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

von

Venkatesh Kancherla

2022

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Luigi M. Terracciano

Prof. Michael N. Hall

Dr. Charlotte K. Y. Ng

Dr. Anamaria Necsulea

Basel, 14.12.2021

Prof. Dr. Marcel Mayor

Dekan

IF YOU FAIL, NEVER GIVE UP

BECAUSE **FAIL** MEANS

"FIRST ATTEMPT IN LEARNING"

DONT FEAR OF FACING FAILURE

IN THE FIRST ATTEMPT,

BECAUSE EVEN THE SUCCESSFUL MATH

STARTS WITH **"ZERO"** ONLY

-   **DR. A. P. J. ABDUL KALAM**

Table of contents

Summary

Triple-negative breast cancers (TNBCs) are aggressive tumors which lack estrogen receptor (ER), progesterone receptor (PR) and HER2 overexpression and account for 15%-20% of breast cancers. To understand the heterogeneity of TNBCs, they have been classified into different molecular subtypes. Studies reported that lncRNAs were dysregulated in many types of cancers and also play an important role in biological functions of TNBCs including cell cycle, migration, cell proliferation, cell invasion, double strand DNA break repair and metastasis.

But still attempts for the identification and characterizing the function of lncRNAs in TNBCs, especially in the subtypes of TNBCs are very limited. With the difficulties in discovering lncRNAs and predicting their functions through experimental approaches, the increase in genomics and transcriptomics data availability computational approaches might serve as screening to identify lncRNA targets for functional characterization[1]. Many studies have also reported that the tumor microenvironment plays an important role in tumor progression. Immune dysregulation is also one of the hallmarks of cancer[2]. TNBCs are more aggressive tumors and at present the standard of care treatment option is cytotoxic chemotherapy[3]. Currently, immunotherapy is under study in treating TNBC with immune checkpoint inhibitors and many other targets. From the very recent study, it is seen that several immunology related lncRNAs involved in different cancer types were identified with computational pipeline ImmLnc[4]. Very few immune lncRNAs have been studied in TNBCs so far[5,6].

The present doctoral work explores the role of long non-coding RNAs in Triple negative breast cancers (TNBCs). In the first part of the project we applied *in-silico* tools to identify novel lncRNAs in the TNBCs and to infer their functions using co-expression network analysis. The second part of the project focussed on the identification of immune lncRNAs in TNBCs. Altogether, the present thesis reports valuable information about lncRNA function in the subtypes of TNBCs and how they may contribute to the shaping of the immune environment in TNBCs.

List of Abbreviations

BL1: Basal-like 1

BL2: Basal-like 2

BLIA: Basal-like immune activated

BLIS: Basal-like immunosuppressed

BRCA: Breast Invasive Carcinoma

CPC: Coding Potential Calculator

FUSCC: Fudan University Shanghai Cancer Center

GLOBOCAN: Global cancer observatory statistics

GSVA: Gene set variation analysis

IHC: Immunohistochemistry

IM: Immunomodulatory

IPS: Immunophenoscore

LAR: Luminal Androgen Receptor

lincRNA: Long intergenic non-coding RNA

lncRES: Long non-coding RNA Enrichment Score

lncRNA: Long non-coding RNA

lncRNAKB: Long non-coding RNA Knowledge Base

M: Mesenchymal

MHC: Major histocompatibility complex

miRNA: MicroRNA

MSL: Mesenchymal Stem-like

ncRNA: non-coding RNA

NODE: National Omics Data Encyclopedia

ORF: Open reading frame

PC : Protein-coding

PLS-DA: Partial least squares discriminant analysis

RBP: RNA Binding Protein

RF: Random Forest

rRNA: Ribosomal RNA

RS: Rank score

siRNA: Small interfering RNA

ssGSEA: Single sample gene set enrichment analysis

SVM: Support vector machines

TCGA: The Cancer Genome Atlas

TILs: Tumor infiltrating lymphocytes

TIMER: Tumor IMmune Estimation Resource

TIS: Tumor inflammation signature

TME: Tumor Microenvironment

TNBC: Triple negative breast cancer

tRNA: Transfer RNA

VST: Variance Stabilizing Transformation

WGCNA: Weighted gene co-expression network analysis

List of lncRNAs, mRNAs and proteins

AFAP-AS1: Actin Filament Associated Protein 1 Antisense RNA 1

ANRIL: Antisense non-coding RNA in the INK4 locus

ANRILA: Androgen Receptor Negatively Regulated LncRNA

BRCA1 - Breast Cancer 1

CCND1: Cyclin D1

CTLA4: Cytotoxic T-Lymphocyte Associated Protein 4

DANCR: Differentiation Antagonizing Non-Protein Coding RNA

DDX11-AS1: DEAD/H-Box Helicase 11 Antisense RNA 1

FOXP4-AS1: Forkhead Box P4 Antisense RNA 1

FZD7: Frizzled Class Receptor 7

GAS5: Growth Arrest Specific 5

HCP5: HLA class I histocompatibility antigen protein P5

HIF1A: Hypoxia Inducible Factor 1 Subunit Alpha

HOTAIR: HOX Transcript Antisense RNA

HOTTIP: HOXA transcript at the distal tip

HOXD: Homeobox D Cluster

HP1: Heterochromatin protein 1

ITGB2: Integrin Subunit Beta 2

LAG3: Lymphocyte Activating 3

LINC-ROR: Long Intergenic Non-Protein Coding RNA, Regulator Of Reprogramming

LINK-A: Long intergenic non−coding RNA for kinase activation

LINP1: LncRNA In Non-Homologous End Joining Pathway 1

LSD1: lysine-specific demethylase 1

LUCAT1: Lung Cancer Associated Transcript 1

MALAT1: Metastasis Associated Lung Adenocarcinoma Transcript 1

NAMPT: Nicotinamide Phosphoribosyltransferase Antisense RNA 1

NEAT1: Nuclear Paraspeckle Assembly Transcript 1

NFYA: Nuclear Transcription Factor Y Subunit Alpha

NRON: Non-Coding Repressor Of NFAT

OCT4: Octamer-binding transcription factor 4

PANDAR: Promoter Of CDKN1A Antisense DNA Damage Activated RNA

PARP: Poly (ADP-ribose) polymerase

PD1: Programmed cell death protein 1

PDCD4-AS1: Programmed Cell Death 4 Antisense RNA 1

PDL1: Programmed death-ligand 1

PRC2: Polycomb repressive complex 2

PTCSC3: Papillary Thyroid Carcinoma Susceptibility Candidate 3

PVT1: Plasmacytoma variant translocation 1

RMST: Rhabdomyosarcoma 2 Associated Transcript

RXRA: Retinoid X Receptor Alpha

SNHG12: Small Nucleolar RNA Host Gene 12

SOX9: SRY-Box Transcription Factor 9

TERC: Telomerase RNA Component

TIGIT: T Cell Immunoreceptor With Ig And ITIM Domains

TUG1: Taurine Up-Regulated 1

TYMSOS: TYMS Opposite Strand RNA

VTCN1: V-Set Domain Containing T Cell Activation Inhibitor 1

XIST: X Inactive Specific Transcript

ZEB2: Zinc Finger E-Box Binding Homeobox 2

ZNF703: Zinc Finger Protein 703

1 Introduction

1.1 Breast Cancer and its classification

Breast cancer is one of the major cancer types that is the most common in women and fifth leading cause of cancer death worldwide. According to Global cancer observatory statistics (GLOBOCAN), breast carcinoma is diagnosed in women with approx. 2.3 million new cases and an estimated 685,000 deaths in 2020 worldwide[7].

In the majority of the countries, breast carcinoma cases were almost one in four cases and one in six cancer deaths and ranked first in incidence and mortality rates, followed by colorectal, lung, and cervical cancers *(Figure 1.1)*. The incidence and mortality rates were different between the countries. Higher incidence rates and lower mortality rates were seen in Australia, New Zealand, North America and the European countries[7] *(Figure 1.1)*. Higher incidence rates could be due to awareness about screening tools and the early detection may lead to lower mortality rates.

Breast cancer is a heterogeneous disease which has been classified into different subtypes to better understand their biological and clinical behaviour. Based only on immunohistochemistry (IHC) of cellular markers, breast cancer has been classified into three subtypes (a) hormone-receptor-positive defined as ER-positive and/or PR-positive, HER2 negative (b) HER2 positive (c) triple negative breast cancer (TNBC), defined by the ER negative, PR negative and HER2 negative *(Figure 1.2)*.
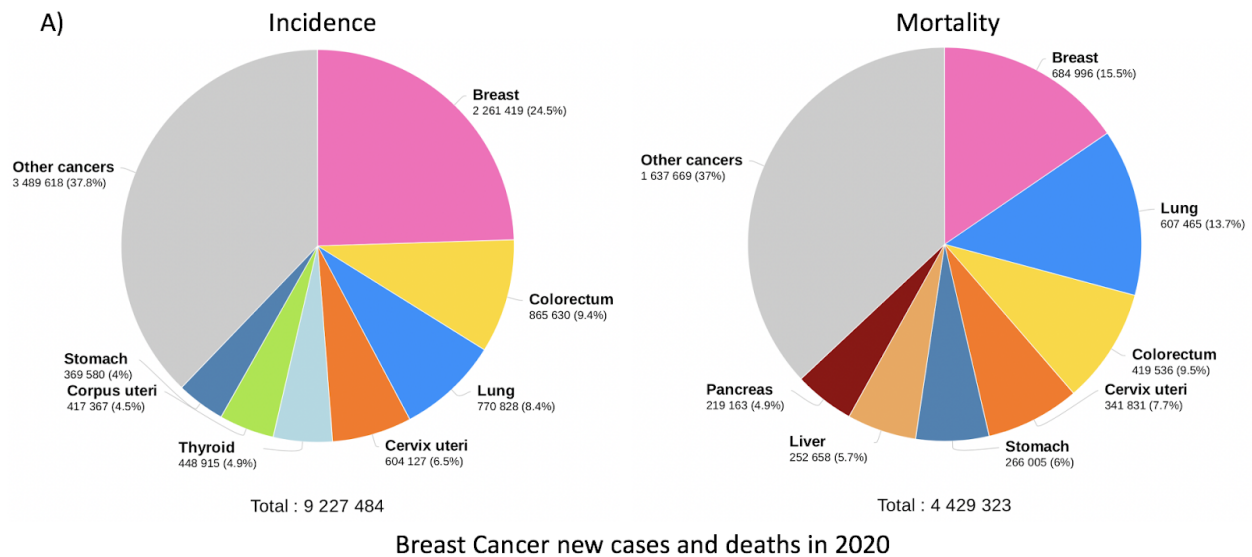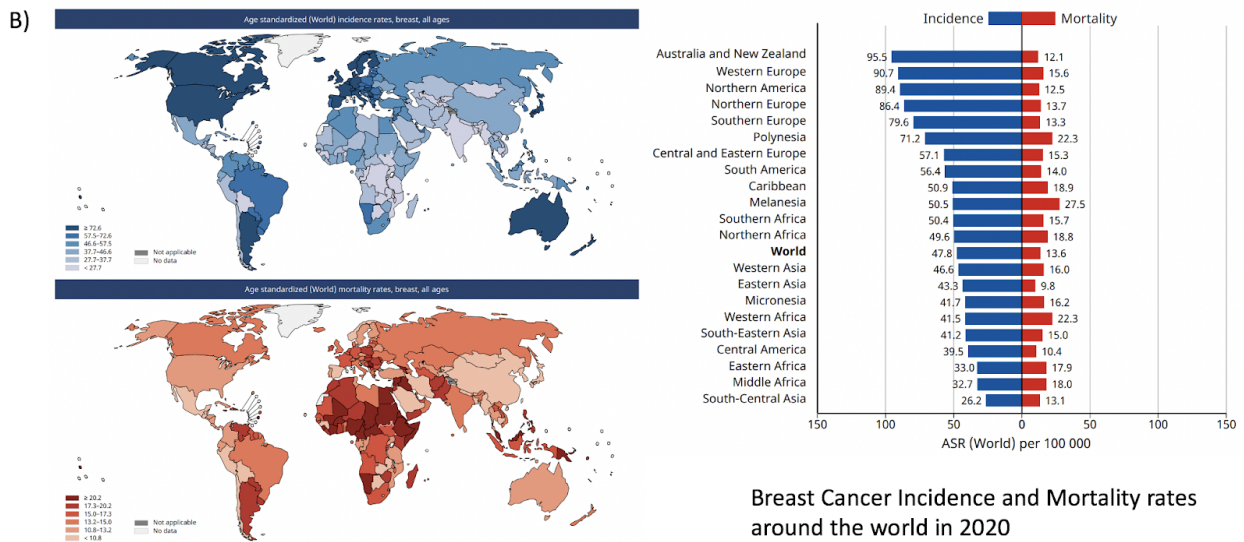
Figure 1.1: *(A) Pie charts present the distribution of incidence and deaths for the different cancer types in 2020 for incidence and mortality. (B) Bar chart of region-specific incidence and mortality age- standardized rates for breast cancer in 2020. Rates are shown in descending order of the world age-standardized rate. Source: GLOBOCAN 2020[7].*

Taking advantage of available transcriptomic studies, breast cancers have also been categorized into different molecular subtypes like luminal A, luminal B, basal-like and

HER2-like[8] (**Figure 1.2**). Luminal A breast cancers are characterized by the ER-positive and/or PR-positive, HER2 negative and have low levels of *Ki-67* protein. They account for 50-60% of all breast cancers, low histological grade tumors and have good prognosis. Luminal B breast cancers represent 15-20% of breast cancers and are ER-positive and/or PR-positive positive, HER2 positive/negative which have higher levels of *Ki-67* and worse prognosis than Luminal A tumors. Basal-like tumors are frequently negative for ER, PR and HER2 and have an overlap with triple negative phenotype cancers. HER2-like tumors represent 15-20% of breast cancer subtypes which have a poor prognosis and are frequently negative for ER, PR and positive for HER2[9].
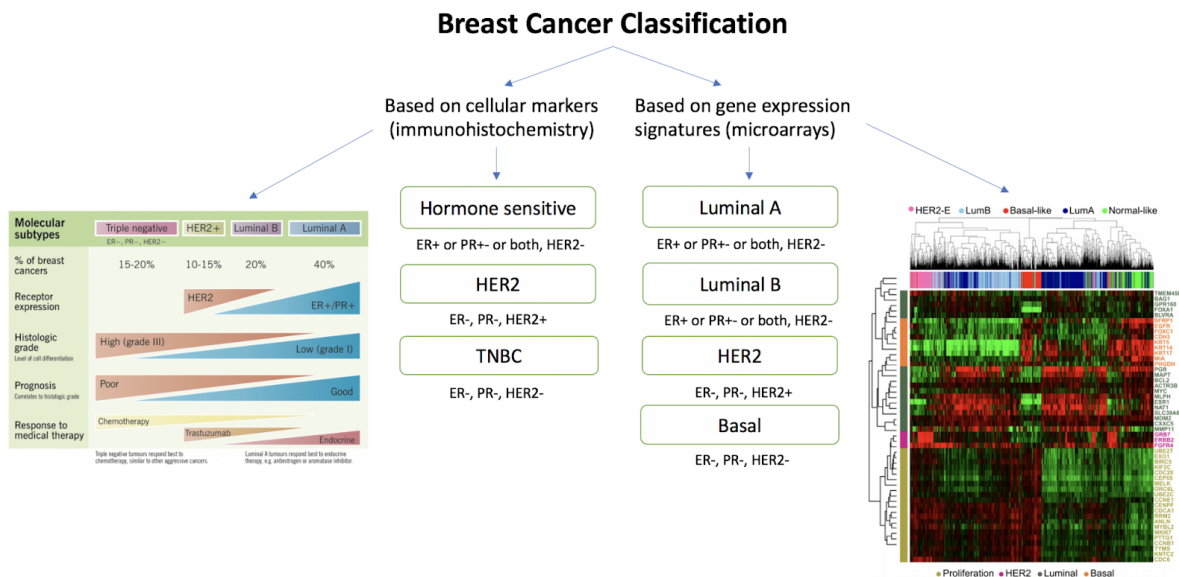
Figure 1.2: *Classification of breast cancer based on immunohistochemistry and gene expression signatures[10].*

1.2 Triple Negative Breast Cancer

Among the subtypes of breast cancer, triple negative breast cancer (TNBC) which is defined by the estrogen receptor negative, progesterone receptor negative and HER2 negative accounts for 15-20% of breast cancers, an aggressive subtype with poorer prognosis than other subtypes of breast cancer and patients have a high incidence of *BRCA1/BRCA2* gene mutations[11]. Pregnancy is an important risk factor for women in developing triple negative breast cancer compared to other subtypes of breast cancer[10]. From the previous studies it is seen that women with multiple pregnancies have a higher chance of developing triple negative breast cancer[12,7]. Till date cytotoxic chemotherapy is the main standard treatment and several targeted therapies such as *PARP* inhibitors are being investigated for treating TNBC patients[10]. TNBCs have shown to be immunogenic, and from the recent study pembrolizumab (PD1 receptor specific monoclonal antibody) combined with chemotherapy can be suggested to treat TNBC[13,3].

1.3 Classification of TNBCs

Triple negative breast cancer is a heterogeneous breast cancer group and aggressive phenotype, lacking molecular targets. To understand clinical and biological behavior of triple negative breast cancers Lehmann *et al.* proposed a computational approach, applied on published datasets to classify TNBCs into six subtypes, namely basal-like1 (BL1), basal-like2 (BL2), mesenchymal (M), immunomodulatory (IM), mesenchymal stem-like (MSL) and luminal androgen receptor (LAR) based on differential gene expression profiles and gene ontologies[14] *(Figure 1.3)*. Each subtype of TNBC has different characteristics and is enriched in different pathways. BL1 is a proliferative

subtype enriched in cell cycle, DNA replication and cell proliferation pathways. BL2 subtype tumors are involved in glycolysis, growth factor signaling pathways. M subtype is more related to cell differentiation pathways and extracellular matrix interaction. The IM subtype is enriched in immune processes like T-cell receptor pathway, JAK-STAT signaling pathway and Th1/Th2 pathways. MSL subtype shares several transcriptomic similarities and are enriched in genes related to mesenchymal stem cells and cell motility. LAR subtype displays a luminal-like gene expression enriched in metabolic processes and hormone signaling pathways[15].

Apart from *Lehmann et al.* molecular subtyping, *Burstein et al.* proposed a non-negative matrix factorization model to classify TNBCs into four subtypes such as Basal-like immune activated (BLIA) defined by cytokines and STAT signal transduction gene expression, Basal-like immunosuppressed (BLIS) defined by immunosuppressive activation inhibitor VTCN1, Mesenchymal (M) which is expressed by the growth factor receptors and Luminal Androgen receptor (LAR) defined by androgen receptor activation (AR) and luminal like expression pattern[16] (*Figure 1.4*).

Similarly, *Lie et al. (FUSCC classification)* applied *k*-means clustering on the gene expression data including both mRNAs, lncRNAs and classified 165 TNBC tumors into four subtypes like Basal-like immunosuppressive (BLIS), Immunomodulatory (IM), Mesenchymal-like (MES) and Luminal Androgen receptor (LAR)[17] (*Figure 1.4*). FUSCC subtype classification was significantly associated with *Lehmann et al.* subtype classification[17] and also consistent with *Burstein et al.* subtypes[18].

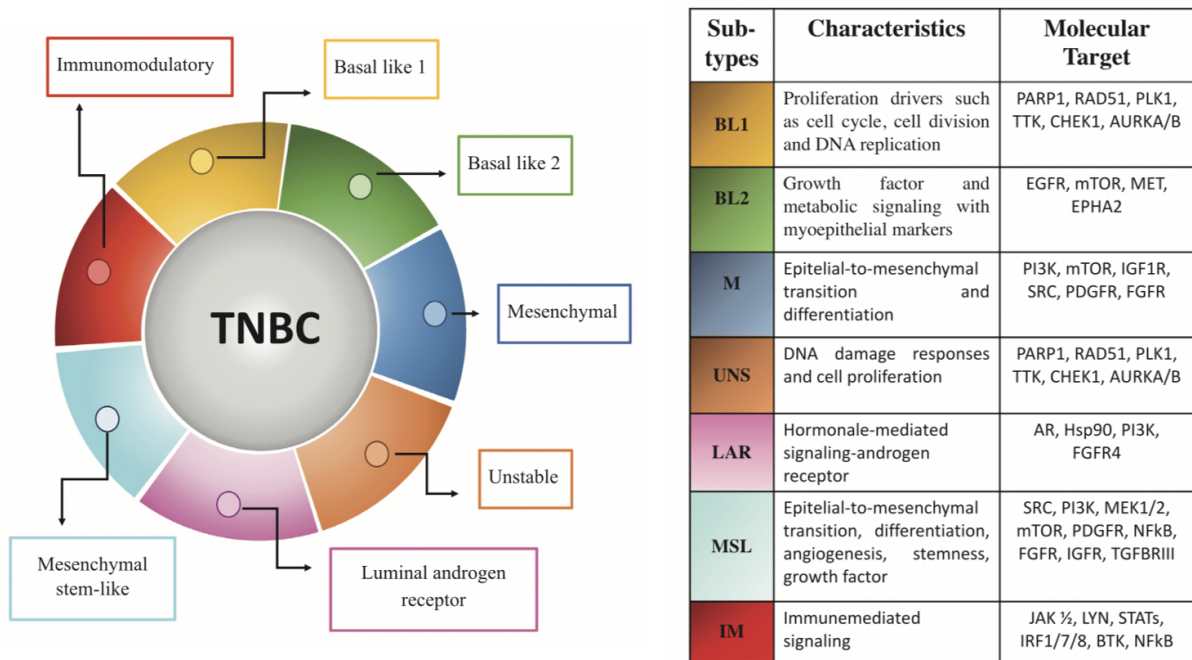| Sub-types	Characteristics	Molecular Target
BL1	Proliferation drivers such as cell cycle, cell division and DNA replication	PARP1, RAD51, PLK1, TTK, CHEK1, AURKA/B
BL2	Growth factor and metabolic signaling with myoepithelial markers	EGFR, mTOR, MET, EPHA2
M	Epitelial-to-mesenchymal transition and differentiation	PI3K, mTOR, IGF1R, SRC, PDGFR, FGFR
UNS	DNA damage responses and cell proliferation	PARP1, RAD51, PLK1, TTK, CHEK1, AURKA/B
LAR	Hormonale-mediated signaling-androgen receptor	AR, Hsp90, PI3K, FGFR4
MSL	Epitelial-to-mesenchymal transition, differentiation, angiogenesis, stemness, growth factor	SRC, PI3K, MEK1/2, mTOR, PDGFR, NFkB, FGFR, IGFR, TGFBRIII
IM	Immunemediated signaling	JAK ½, LYN, STATs, IRF1/7/8, BTK, NFkB

Figure 1.3: Triple Negative Breast Cancers classification according to Lehmann et al[19]
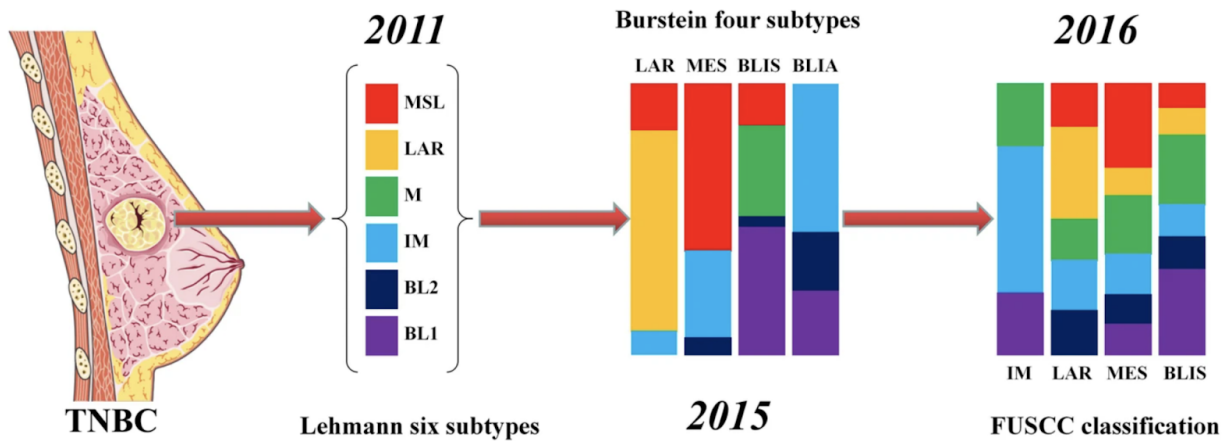

Figure 1.4: Molecular subtyping of Triple negative breast cancers by Lehmann classification into six subtypes, Burstein classification into four subtypes and recent FUSCC classification into four subtypes[18]

1.4 Non coding RNAs

Approximately 20,000-25,000 protein coding genes were estimated in the human genome which account for around 2% of the human genome and the remaining 98% of the genome were regarded as "dark matter" before, due to their inability to code into proteins[20]. Implementation of next generation sequencing technologies have revolutionized genome research and found that 90% of the human genome is transcribed into RNA transcripts[21]. Among these RNA transcripts, the majority are noncoding RNAs (ncRNAs) that do not encode into proteins and are identified with several mechanisms like DNA or RNA based duplications of genomic sequences[22]. Several types of ncRNAs have been widely studied in the past decade. These include small housekeeping RNAs like transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and microRNAs (miRNAs), circular RNAs, small interfering RNAs (siRNAs) and long noncoding RNAs (lncRNAs)[20]. Each of these ncRNAs is transcribed with any of the three RNA polymerases[23] (RNA polymerase I, RNA polymerase II and RNA polymerase III). Among these ncRNAs, tRNA and rRNA play an important role in mRNA translation. miRNAs are small single stranded non-coding RNAs with a length of 22 nucleotides that are transcribed into primary miRNAs from DNA sequences and processed into precursor miRNAs and mature miRNAs[24]. They play an important role in regulating mRNA expression[25]. Circular RNAs are a class of ncRNAs, where 3` and 5` ends are linked together by back splicing of exons of a pre-mRNA[26]. siRNAs are a length of 20-30 nucleotides that regulate the gene expression and can be used to study the function of genes either by *in-vivo* or *in-vitro* experiments to treat cancer and other diseases[27].

Long non-coding RNAs are also a class of ncRNAs which are longer than 200 nucleotides and have different properties compared to other ncRNA types[28].

1.5 Emergence of lncRNAs

LncRNAs have different types of origins[29] 1) Protein coding genes might undergo mutations which interfere with ORFs that lead to structural damage and transform into a lncRNA 2) Two different sequences together generate an expressed noncoding sequence after chromosomal rearrangement 3) non-coding RNA duplication events might cause adjacent repeats and increase the length of the transcripts 4) Transposable element insertion creates a functional lncRNA sequence *(Figure 1.5)*.
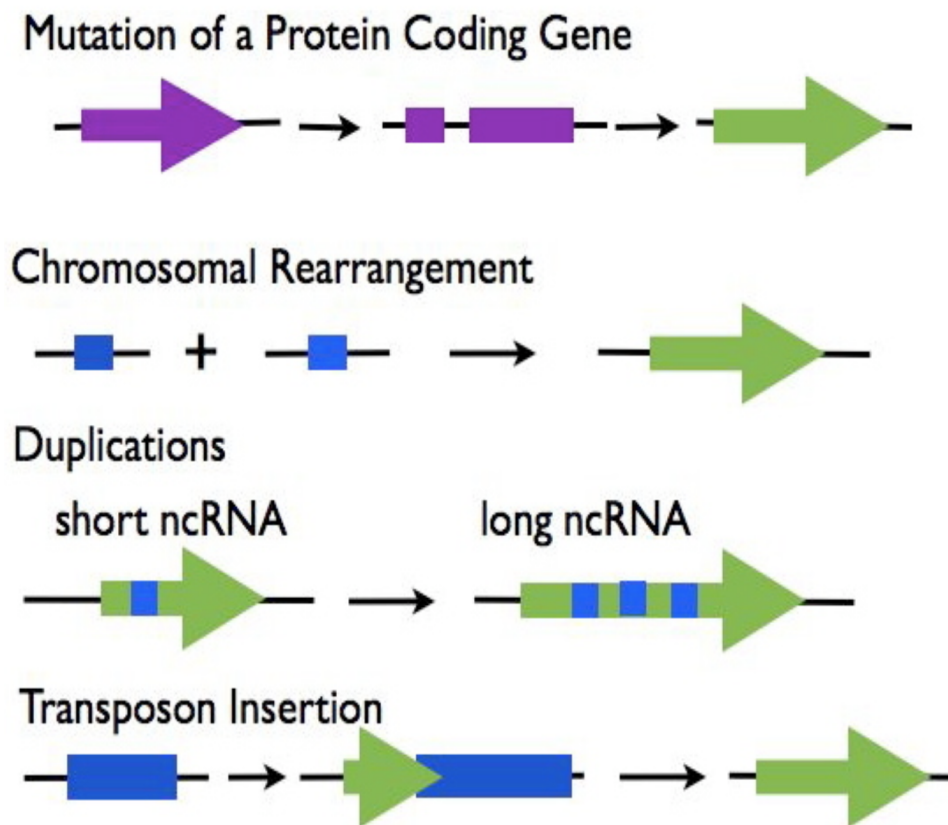
Figure 1.5: Emergence of functional lncRNAs[29]

1.6 Long non coding RNAs

Long non-coding RNAs are the ncRNA transcripts with length greater than 200 base pairs, lack open reading frames and do not code into proteins[30,31,32]. They are typically very low expressed, have fewer exons compared to protein coding genes and have been involved in several cellular functions. lncRNAs are transcribed by RNA polymerase II and are mostly spliced at 5` capped and some are polyadenylated at 3` end[33]. Approximately, 30,000 lncRNA transcripts have been described in GENCODE reference annotation[34]. lncRNAKB database, which is an integration of six lncRNA annotation databases CHESS[35], FANTOM5[36,37], LNCipedia[38], NONCODE[39], MiTranscriptome[40] and BIGTranscriptome[41], contains 224,286 lncRNA transcripts from 77,199 lncRNA genes in human[42]. Many lncRNAs have been described as biomarkers in different cancer types[43], diseases[44] and are involved in several cellular functions such as cell proliferation[45], cell cycle arrest[46], cell invasion[47] and cellular differentiation[48].

1.7 Classification of lncRNAs

Based on the lncRNA location in the genome

lncRNAs have been classified into five categories based on their genomic locations and their relation to protein coding genes (a) Sense overlapping lncRNAs (b) Antisense lncRNAs (c) Bidirectional lncRNAs (d) Intronic lncRNAs (e) Long intergenic non coding RNAs (*Figure 1.6*).

Sense overlapping are lncRNA sequences that overlap with the exons of protein coding genes on the same strand. The overlap might be with the complete sequence of the protein coding gene or a part of the protein coding gene[49]. In total, there are 189 sense overlapping lncRNAs present in long non-coding RNA gencodev27 (*GRCh38.p10*) annotation[34]. Several sense overlapping lncRNAs have been reported to be associated with different cancer types. For example: *LINC00189* is a sense overlapping lncRNA that has been studied in various cancer types like bladder squamous cell carcinoma[50], renal cell carcinoma[51], cervical cancer[52]. Another sense overlapping lncRNA *HCP5* (HLA class I histocompatibility antigen protein P5), is involved in many autoimmune diseases and acts as an oncogene in many cancers[53]. *HCP5* also contributes to cisplatin resistance in human TNBCs by regulating *PTEN* expression[54]. It is highly expressed in spleen and thymus[55] and plays an important role in immune regulation[4].

Antisense lncRNAs are lncRNA sequences that overlap with the exons of protein coding genes on the antisense strand. There are around 5500 antisense lncRNAs in long non-coding RNA gencodev27 (*GRCh38.p10*) annotation[34]. Most of the antisense lncRNAs are represented as "*Genename-AS*". For example: *ITGB2-AS1,* which is located on an antisense strand of protein coding gene *ITGB2* on chromosome 21, could promote migration and invasion of breast cancer[56]. Using strand specific sequencing, antisense lncRNAs are identified to be true transcripts[49].

Bidirectional lncRNAs or divergent lncRNAs were identified based on the discovery of bidirectional transcription, which are located on the opposite strand and within 1kb of

the promoter of the protein coding genes[57]. It is difficult to identify bidirectional lncRNAs, because of unreliable annotation of promoters and transcriptional start sites in the genome[58]. Most of the bidirectional promoters were associated with neuronal genes and were also found close to protein coding genes which were related to cellular functions like cell cycle[59]. There are only 19 bidirectional lncRNAs present in genocode. For example: *TYMSOS* a bidirectional lncRNA involved in cell proliferation, cell migration and cell invasion in gastric cancer regulating *ZNF703* by interacting with miRNA *miR-4739*[60].
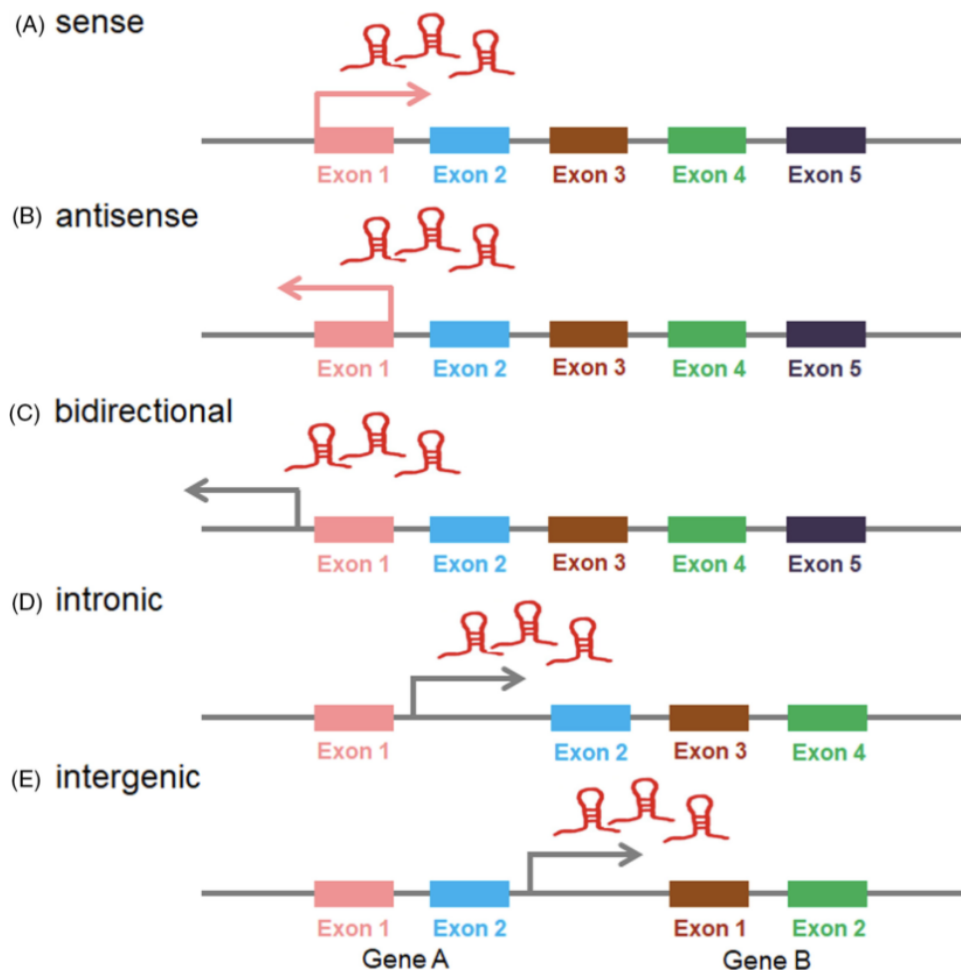
Figure 1.6: Classification of lncRNAs based on their proximity to the neighboring protein coding genes in the genome[49,61]

Intronic lncRNAs are lncRNAs which derive entirely from an intron of a second transcript. Intronic lncRNAs may appear in two schemes such as sense intronic lncRNAs which are transcribed in the introns of a protein coding gene and do not overlap any exons and antisense intronic lncRNAs, transcribed in the opposite direction of a gene[62]. There are approximately 900 sense intronic lncRNAs in the gencode annotation[34]. Some (~13% in Gencode) of the sense intronic lncRNAs are represented as "*Genename-IT*".

Long intergenic non-coding RNAs (lincRNAs) are a class of long non-coding RNAs that have no overlap with any of the annotated protein coding genes[63]. lincRNAs are associated with chromatin signatures and affect expression of genes[64]. Gencode v27 annotation annotates around 7,500 lincRNAs. lincRNAs are evolutionarily conserved and involved in several functional roles in human diseases and development[63]. Many studies reported lincRNAs such as *MALAT1*[65], *LINK-A*[66] and *PVT1*[67] as potential biomarkers in several cancer types.

Based on the lncRNA effects on DNA sequences

Long non-coding RNAs have been localized in many places within the cell such as chromatin, nucleus, cytoplasm and exosomes[68], which may suggest that it may show an effect on DNA sequences[49]. Based on their effects they have been classified into *cis*-acting lncRNAs and *trans*-acting lncRNAs *(Figure 1.7)*. lncRNAs that are in close proximity to the protein coding gene (neighbouring genes) on which they regulate are

cis-lncRNAs and lncRNAs have an effect on genes that are on other chromosomes (distal genes)[69]. *cis*-lncRNAs mainly function through chromatin modifications.
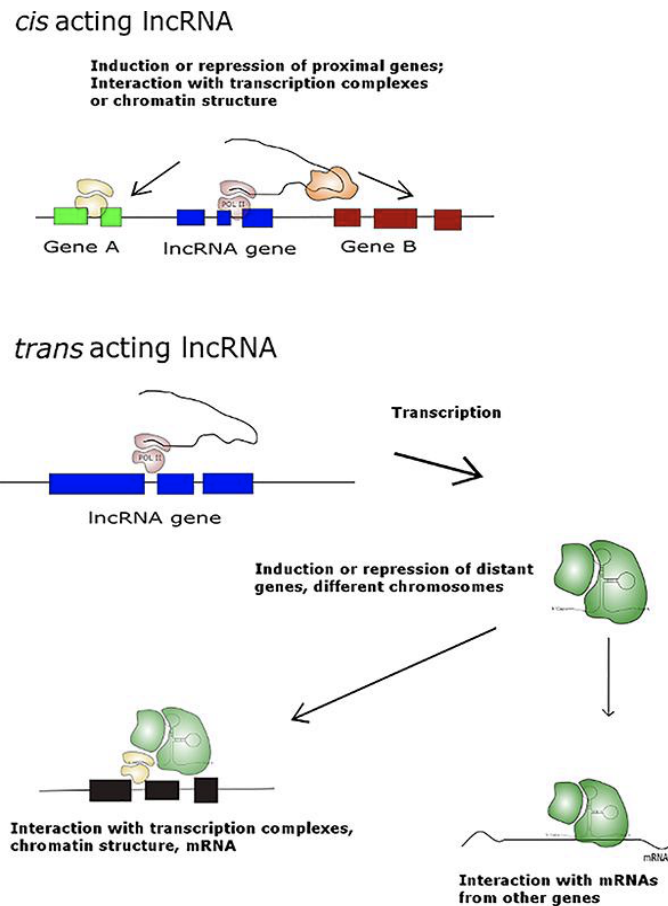
Figure 1.7: Based on the effects of lncRNAs on DNA sequences cis-acting and trans-acting lncRNAs were identified[70]

lncRNAs such as *H19*[71], *HOTTIP*[72], *XIST*[73] are the most well characterized *cis*-lncRNAs which have a transcriptional regulation as functional mechanism[49]. For example: *HOTTIP* maintains active chromatin across 5` end of all *HOXA* genes and regulates the expression of the genes in close proximity[74]. Another *cis*-lncRNA *XIST* induces silencing of genes on the *X* chromososme, due to its interaction with polycomb repressive

complex2 (PRC2)[75]. On the other hand, *HOTAIR* is a well studied *trans*-lncRNA located within the *HOXC* (Homeobox C) gene cluster on chromosome 12 and is co-expressed with *HOXC* genes[76]. *HOTAIR* interacts with the PRC2 complex and promotes the repression of the *HOXD*[77] gene cluster on chromosome 2.

Based on the lncRNA mechanism of action

Long non-coding RNAs might regulate the expression of genes with several mechanisms such as interacting with mRNAs, miRNAs, DNAs and proteins at transcriptional and post-transcriptional levels[62,78]. Based on these mechanisms of action, lncRNAs are classified into four categories such as **decoys, guides, scaffolds and molecular signals**[79] *(Figure 1.8)*.

lncRNAs act as molecular ***decoys*** which interact and inhibit the functions of miRNAs, transcription factors, and proteins. They are regarded as negative regulators which compete with sequences and structures for interacting[80] *(Figure 1.8)*. lncRNA *Gas5* (Growth arrest specific 5) located on chromosome 1 acts as a molecular decoy for glucocorticoid response element (GRE)[81]. It interacts with the DNA binding domain of glucocorticoid receptor, which blocks it from binding with glucocorticoid response element and supresess glucocorticoid-responsive genes. *Gas5* is considered as a riborrepressor which influences several functions like cell survival and metabolism[81]. Another lncRNA *PANDA* (Promoter of CDKN1A Antisense DNA damage activated RNA) located on chromosome 6, functions as a decoy interacts with the *NF-YA* which is a transcription factor to control the pro-apoptotic target genes expression[59].
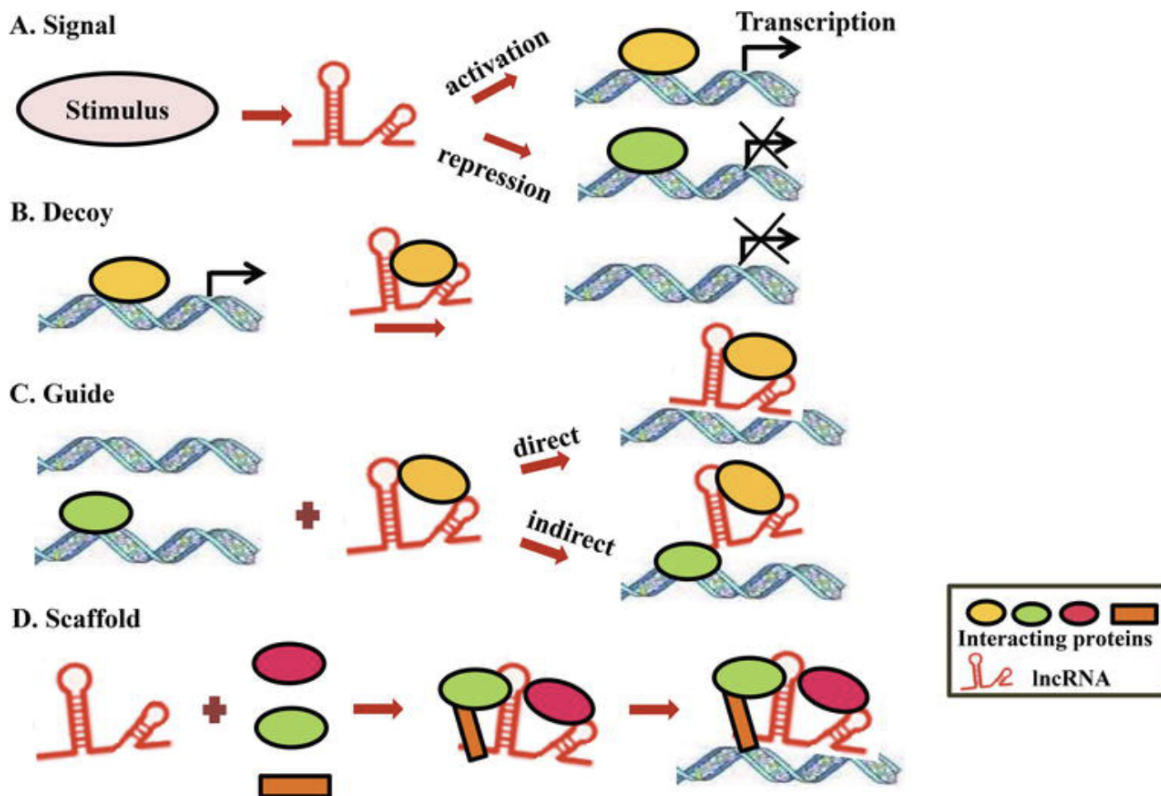
Figure 1.8: lncRNA categorization based on the mechanism of functions[79]

Guides are the lncRNAs which bind with the RNA binding proteins (RBPs) and direct them to specific targets[80] *(Figure 1.8)*. These lncRNAs might interact with either cis-acting (neighboring genes) or trans-acting genes (distant genes) by acting as activators or suppressors[82]. Some of the lncRNAs such as *HOTTIP, CCND1, COLDAIR, XIST* act as guides in cis and lncRNAs such as *HOTAIR, lincRNA-p21 and PRC2* complex RNAs acts as guides in trans[82].

Scaffolds bring the bound proteins into a chromatin remodeling complex and other complexes or in spatial proximity[79] *(Figure 1.8)*. lncRNAs like *TERC* (telomerase RNA)

acts as scaffold for its proteins, *HOTAIR* functions as scaffold interact with *LSD1* and also *PRC2* complexes[83], *ANRIL*[84] as an antisense lncRNA and also *XIST*[85] interacts with both *PRC1* and *PRC2* chromatin remodeling complexes. lncRNAs also act as scaffolds for targeting and further heterochromatin protein 1 (HP1) localization[82].

Molecular signals are the lncRNAs which might silence or activate the genes based on the stimulus[79] ***(Figure 1.8)***. They act as markers for several biological events and are involved in DNA damage response, regulation of gene expression and developmental stage. *lincRNA ROR* (Regulator of reprogramming) acts as signal and is targeted by pluripotency factors *OCT4* and *SOX2*[86]. Other lncRNAs *COLDAIR* and *COOLAIR* serve as signals involved in stress responses[79], *PANDA* and *lincRNA-p21* are both involved in DNA damage by interacting with transcription factors, and both *HOTAIR* and *HOTTIP* lncRNAs which are found in the distal end of the *HOXA* gene cluster appear as molecular signals of anatomic specific expression and are involved in embryonic development[82].

1.8 LncRNAs in TNBCs

Several studies reported lncRNAs expression is dysregulated in TNBCs and plays a crucial role in the regulation of biological processes like apoptosis, cell proliferation, promoting epithelial-mesenchymal transition (EMT), cell migration and invasion[61]. *RMST* (Rhabdomyosarcoma 2 Associated Transcript) lncRNA located on chromosome 12, might induce the block of G0/G1 phase thus inhibiting cell proliferation in TNBCs. It has

a tumor suppressor role which inhibits cell proliferation, cell migration and promotes apoptosis[87] (*Figure 1.9*).
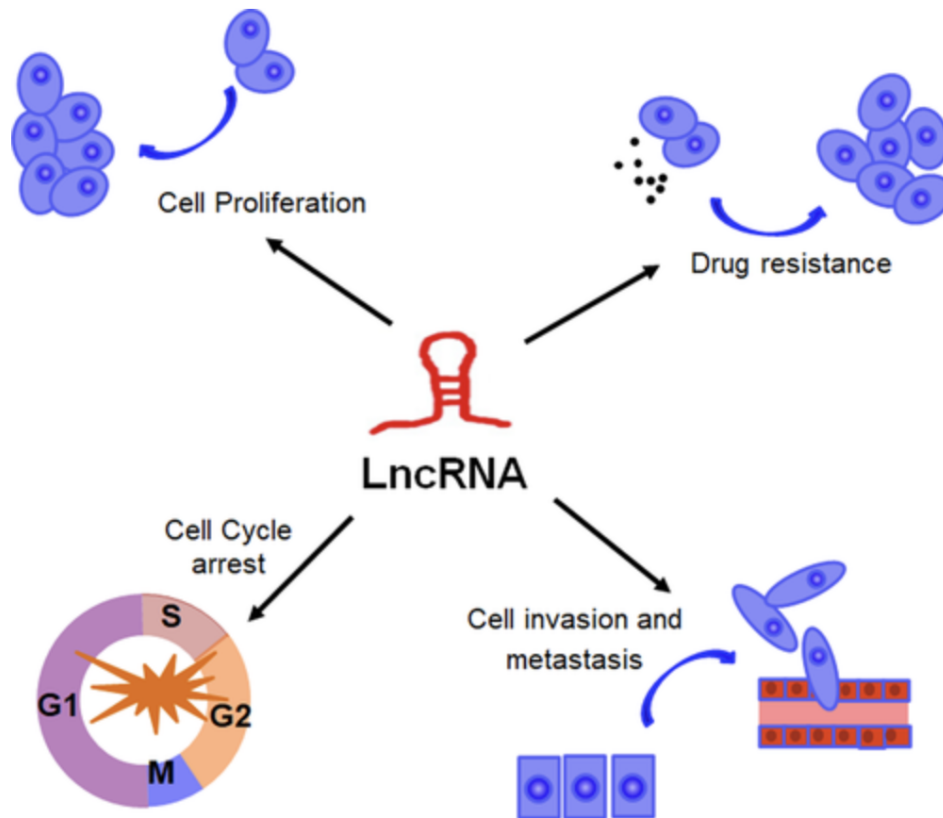
Figure 1.9: *Functions of lncRNAs in Triple Negative Breast Cancers*[61]

LUCAT1 might interact and regulate miRNA *miR-5702* and contribute to cell cycle progression and metastasis in TNBCs[88] (*Figure 1.10*). *MIR100HG* binds with p27 to inhibit cell cycle in G1 phase[89] and another lncRNA *LINC00339* promote tumor progression by sponging to *miR-377-3p* and activating *miR-377-3p/HOXC6* signalling pathway in TNBCs[90] (*Figure 1.10*).
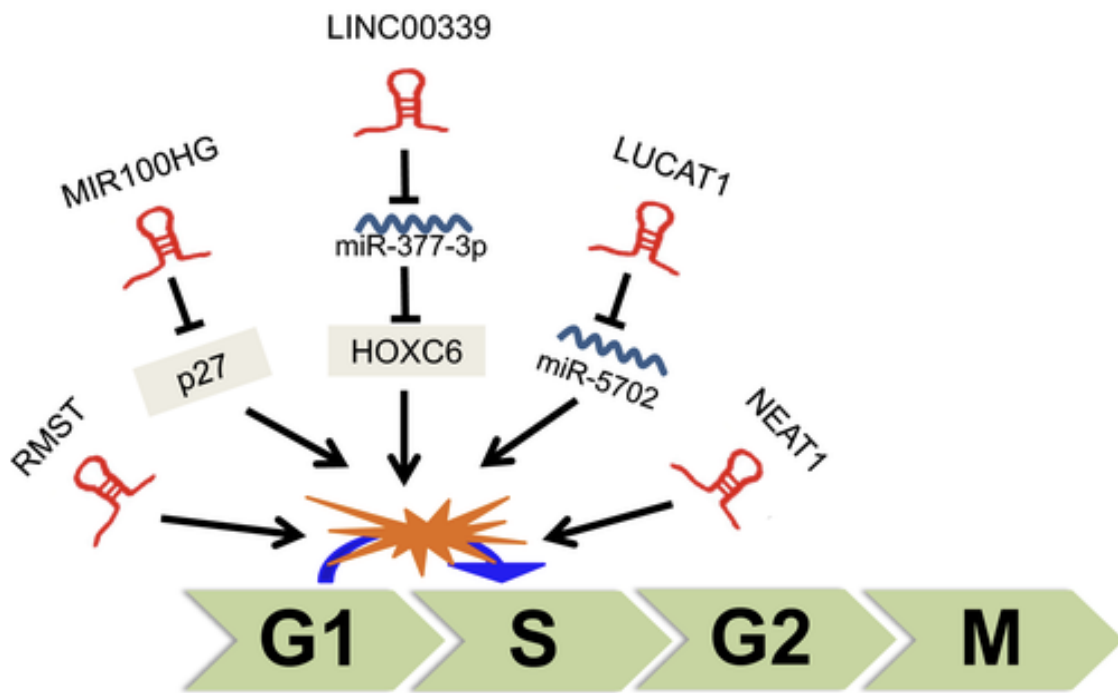
Figure 1.10: Some of the lncRNAs involved in regulation of cell cycle[61]

LncRNA *NEAT1* plays an oncogenic role and regulates the cell cycle progression in TNBCs[91] *(Figure 1.10)*. *GAS5* is downregulated in TNBCs and inhibits cell proliferation and cell invasion through interacting with miRNA *miR-196a-5p*[92] and also regulates cell apoptosis[93].

Overexpression of *LINC00096* interacts with *miR-383-5p* and regulates expression of *RBM3* thereby enhancing the cell proliferation and invasion[94]. *PAPAS* lncRNA acts as a ceRNA based on its interaction with *miR-34a* and promotes cell migration and invasion[95]. Several lncRNAs were reported as molecular signals, based on their interaction with transcription factors to regulate the target genes expression. Some other lncRNAs interact with proteins, other RNAs and are regarded as scaffolds. LncRNA *LINC02095* interacts with transcription factor *SOX9* and promotes cell

proliferation in breast cancer[96]. It was reported that lncRNA *DANCR* acts as an oncogene which enhances *PI3K/AKT* signaling by activating serine phosphorylation of *RXRA* and leading to TNBC carcinogenesis[97]. *PDCD4-AS1* is an antisense lncRNA that regulates the expression and promotes the stability of mRNA *PDCD4*, forms RNA duplex and controls cancer progression in TNBCs[98].

From the previous studies it was reported that lncRNAs might cause cell proliferation in TNBCs by regulating other lncRNAs. For instance, lncRNA *DRHC* downregulates the *HOTAIR* expression and inhibits cell proliferation in TNBCs[99]. *PTCSC3* overexpression inhibits TNBC cell proliferation by downregulating lncRNA *H19*[100]. LncRNA *NRON* overexpression led to downregulated lncRNA *snaR*, and inhibited cell proliferation in TNBC cells[101]. LncRNA *AWPPH* overexpression led to *FZD7* upregulation and promoted the growth of triple negative breast cancer[102]. *LUCAT1* modulates the expression of miRNA *miR-5702* and promotes carcinogenesis and metastasis in triple negative breast cancer[88]. LncRNA *ANRIL* promotes tumorigenesis via sponging miRNA *miR-199a* in TNBC cells[103]. Few other lncRNAs *MALAT1*[104], *LINC01096*[105], *TROJAN*[106], *LINC00152*[107], *SNHG12*[108] were also shown playing an important role in TNBC tumorigenesis and apoptosis. Antisense lncRNA *AFAP1-AS1*, regulates *MTH1* expression by interacting with miRNA *miR-145* and promotes cell proliferation and invasion in TNBCs[109].

It has been reported that TNBC patients have a high risk of developing metastasis, which is one of the major reasons for the deaths worldwide[110]. Many lncRNAs have

been identified for their role in metastasis and cell invasion. LncRNA *LINC00096* regulates *miR-383-5p/RBM3* pathway which in turn promotes invasion and cell proliferation in TNBCs[94].

Overexpression of antisense lncRNA *NAMPT-AS* regulates the *NAMPT* gene epigenetically and promotes metastasis and tumor development in TNBC[111]. Another lncRNA *CCAT1* downregulates miRNA *miR-218* expression which in turn promotes TNBC invasion and migration[112]. In TNBC *MIR503HG* works as a tumor suppressive lncRNA which interacts with *miR-103/OLFM4* axis and inhibits cell invasion and migration[113]. Antisense lncRNA *ZEB2-AS1* identified as an oncogene which positively regulates *ZEB2* gene expression and promotes metastasis and cell proliferation in TNBC[114]. *ARNILA* which is a androgen receptor negatively induced lncRNA functions as ceRNA by interacting with miRNA *miR-204* and promotes metastasis and cell invasion in TNBC[115] *(Figure 1.11)*.
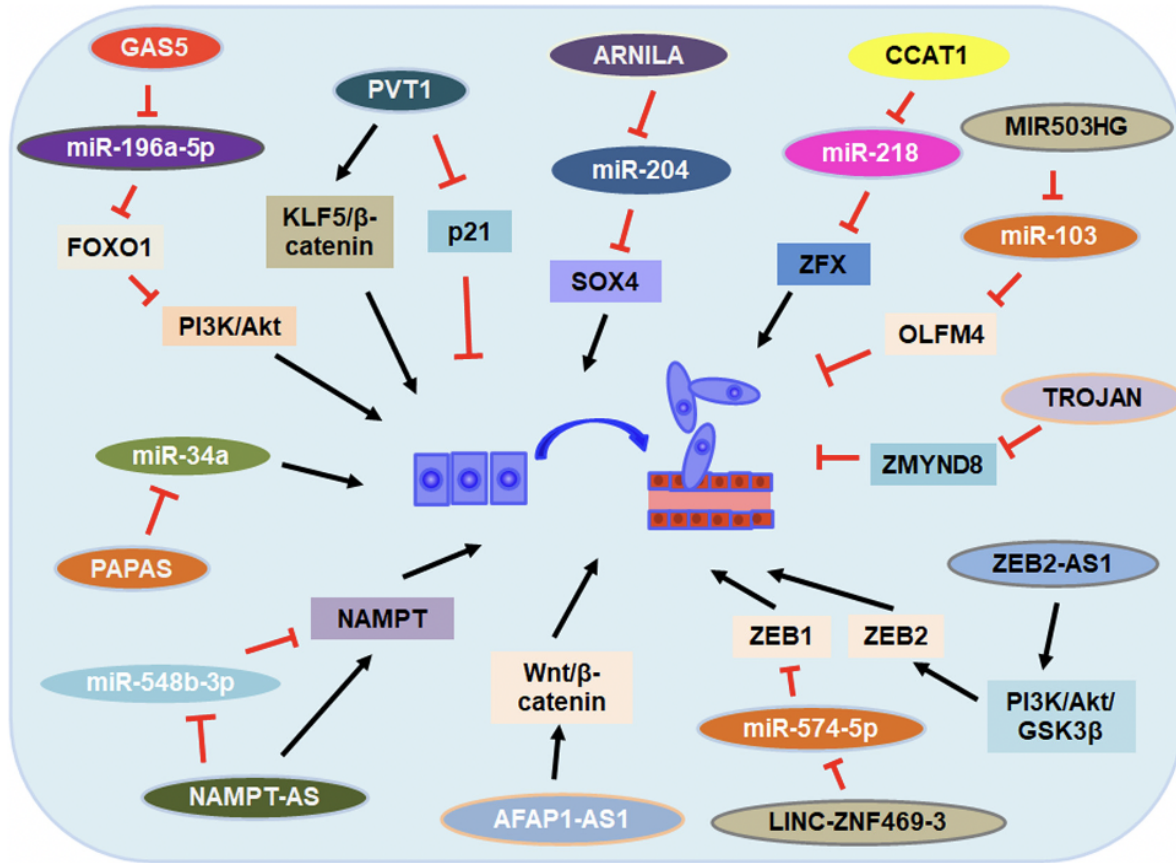
Figure 1.11: *lncRNAs regulating cell metastasis and invasion in Triple negative breast cancer[61]*

Previous studies show that the lncRNAs target various genes and are involved in regulating drug resistance[61] like chemotherapy drugs, radiotherapy and immunotherapy drugs in TNBCs ***(Figure 1.12)***. For example: *LINP1* enhances repair of DNA double strand breaks by serving as a scaffold linking *Ku80, DNA-PKcs* and also plays a key role in radiotherapy resistance in breast cancer[116]. *LINK-A* is overexpressed in TNBC and predicts immunosuppression and immunotherapy resistance[6]. The same lncRNA also interacts with Phosphatidylinositol-3,4,5-trisphosphate and confers resistance to

AKT inhibitors[66]. LncRNA *AWPPH* and miRNA *miR-21* interaction might regulate TNBC cell proliferation and chemosensitivity[117]. There are other lncRNAs *BORG, H19, NEAT1, HIF1A-AS2, TUG1, GAS5* that mediate chemotherapy resistance in TNBCs *(Figure 1.12)*. All these studies give us an idea about lncRNAs regulation for chemotherapy and radiotherapy drugs resistance in TNBCs. More research needs to be done in future about their mechanisms in TNBCs.
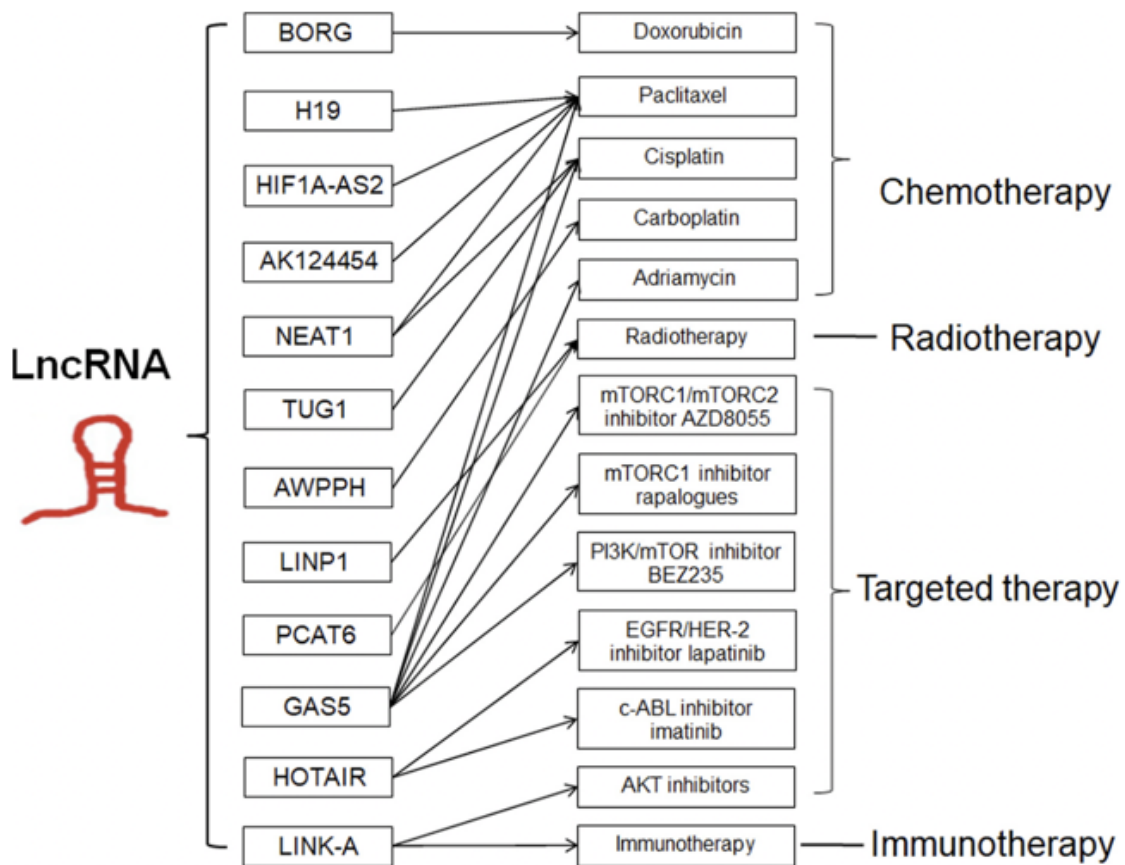
Figure 1.12: *lncRNAs involved in the regulation of immunotherapy, radiotherapy and chemotherapy drugs resistance[61]*

1.9 LncRNAs from immune point of view

Immune system is an important component for every being's survival. It protects from foreign invaders such as parasites, viruses and bacterial attacks. For disease treatment and drug development, it is very much important to know about the immune system. It involves many types of organs, cells, tissues and proteins which are needed for the immune system to work. Many studies showed that the immune system has the potential to suppress the development of tumors[118–120]. Many lncRNAs are expressed in immune cells (Macrophages, CD8+ T cells, CD4+ T cells, B cells, Neutrophils, Dendritic cells) and studies showed that they play an important role in innate and adaptive immune system, linked to regulation of immune response and disease development[121,122].

CD8+ T cells (Cytotoxic T lymphocytes / Killer T cells) are the most prominent immune cells that protect from infected cells or cancer cells. They originate in the thymus of the human body. CD8+ T cells secrete cytokines and play an important role in antitumor effects. They produce cytotoxic granules perforin and granzymes to induce programmed apoptosis. lncRNAs were associated with CD8+ T cells for the promotion of apoptosis and regarded as functional regulators, which can act as a biomarker for immunotherapy[123]. It has been reported that CD8+ T cells also defend viruses and fight against tumors. Study showed that lncRNA *Morbid* regulates CD8+ T cell survival in response to viral infection[124]. Inhibition of *NEAT1* lncRNA repressed tumor development in hepatocellular carcinoma through regulation of miR-155/Tim-3 pathway[125]. Expression of *Lnc-Tim3* influences the CD8+ T cell exhaustion which is associated with anti-tumor

immunity and can be used as a potential target for better immune response[126] *(Figure 1.13)*.
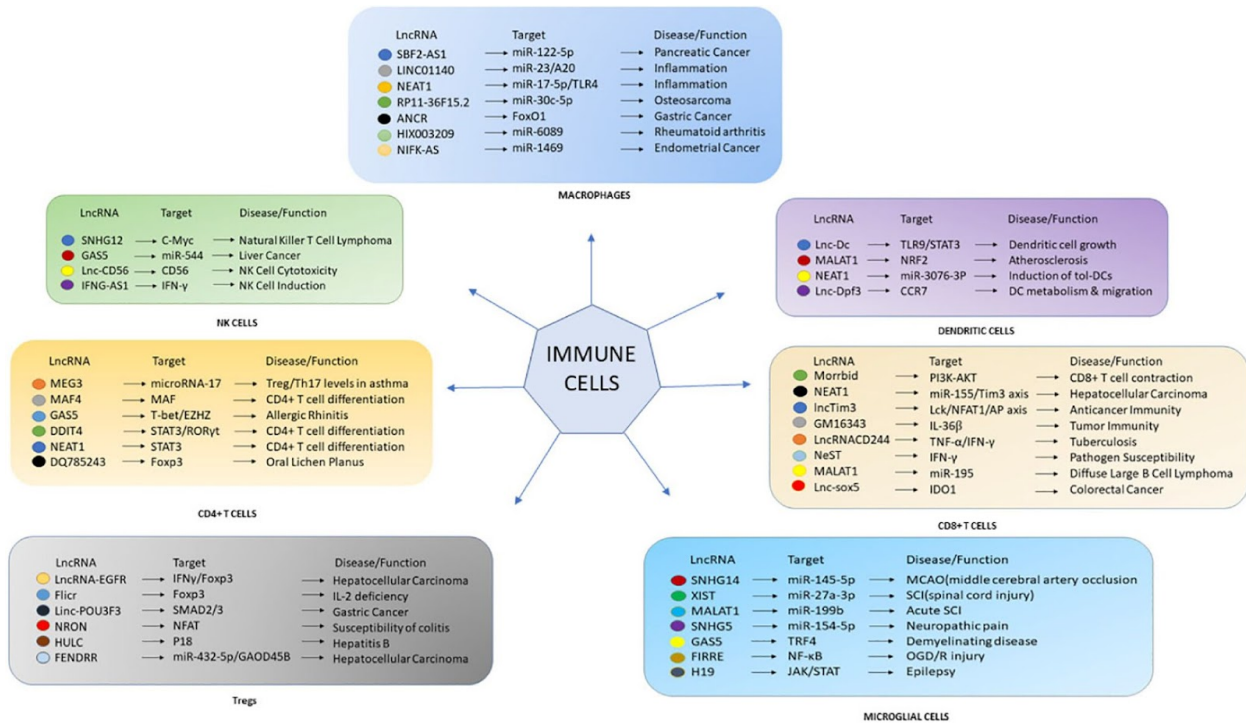
Figure 1.13: lncRNAs association with immune cells by interacting with different molecular targets[123].

Macrophage cells originate from blood monocytes and play an important role in regulating immune response to detect and kill bacteria or pathogens. They also produce antigens to T-cells and activate other cells by releasing cytokines. Studies showed lncRNAs were involved in macrophage growth and can activate or inactivate cell signalling pathways, preventing apoptosis. With the help of molecular targets several lncRNAs were associated with macrophage differentiation and other diseases. For instance: *HOTAIR* lncRNA plays an important role in NF-kB activation and is proposed

as a biomarker for immune response in macrophages[127]. LncRNA *ANCR* overexpression promotes invasion and metastasis via regulation of FoxO1 to slow down macrophage M1 polarization[128]. It has been shown that lncRNA *AK085865* plays an important role in regulating macrophage polarization both in-vitro and in-vivo[129].

Dendritic cells are a type of antigen presenting cells (APCs) which mainly process antigens on cell surface to the T-cells in the immune system. They basically link the innate and adaptive immune systems. Experiments showed that dendritic cells process antigens and play an important role in converting proteins to peptides on the MHC molecules which are recognized by T lymphocytes. lncRNAs are involved in dendritic cell growth, apoptosis and disease development. For example: Knockdown of lncRNA *NEAT1* induces tolerogenic phenotype in dendritic cells with the help of NLRP3 inflammasome[130]. lncRNA *Lnc-DC* was regarded as an important regulator for dendritic cell growth and immune responses via modulation of TLR9/STAT3 pathway[131].

Neutrophils are the type of granulocytes, make up 40 - 70% of all white blood cells in the human body[132]. They help in healing damaged tissues and infections. lncRNAs were seen playing a key role in regulating spontaneous neutrophil apoptosis[133] and also accelerate immune escape of ovarian cancer cells and might act as target for therapy[123]. *HOTTIP* promotes interleukin 6 expression and upregulates PDL1 in neutrophils and accelerates immune escape in ovarian cancer cells[134].

CD4+ T cells are the type of T cells that express CD4 protein on its surface, regulate immune response and activate CD8+ T cells and B cells by secreting cytokines[135]. They are considered as the helper T cells and are important to activate macrophages and dendritic cells for immune defenses from harmful infections. They have different subtypes T-helper 1, T-helper 2, T-helper 9, T-helper 17, Tregs, and follicular helper T-cell to counteract virus infections[136]. It has been reported that in CD4+ T cells, lncRNAs are identified as epigenetic and transcriptomic regulators by interacting with miRNA targets, transcription factors[137]. They were regarded as important for T-cell activation and differentiation[138]. LncRNA *MEG3* acts as a ceRNA by interacting with microRNA-17 and affects Th17/Treg balance in asthma patients[139]. *Linc-MAF-4* targets *MAF* gene and regulates differentiation of Th1/Th2 cells in multiple sclerosis patients[140].

B-cells originate in the bone marrow and produce antibodies called immunoglobulins to fight against infections. Several immunoglobulins like IgG, IgM, IgA, IgE and IgD play distinct roles in both innate and adaptive immune responses[141]. lncRNA expression is associated with transcription factors and regulates the development of B-cells and cell cycle regulation[123]. lncRNAs *CRNDE* and *OIP5-AS* were highly expressed in proliferating B-cells and associated with cell cycle gene regulation[142,143].

2 Aim of the research project

Long non-coding RNAs have been found dysregulated in many cancer types and also play an important role in cancer progression. Several studies previously reported on the role of lncRNAs in triple negative breast cancer progression. However, attempts for the identification and characterizing the function of lncRNAs in triple negative breast cancers, especially in the subtypes of TNBCs are very limited. Several lncRNAs were regarded as immune regulators in different diseases and cancer types which might act as immunotherapy targets. In TNBCs, the function of lncRNAs in immune response is not well explored.

My PhD project is focused on detecting novel lncRNAs, identifying overexpressed lncRNAs specific to the subtypes and their functional characterization by applying in-silico approaches in a publicly available dataset (*FUSCC TNBC cohort*). Additionally, we also aimed at identifying immune lncRNAs and their association with TNBC subtypes and selecting signature lncRNAs related to the immunogenicity of the samples.

3 Materials and Methods

3.1 Data collection, Read alignment and *Ab-initio* assembly

We downloaded the raw fastqs of 317 TNBCs and 44 non-tumor samples of FUSCC cohort (Accession: *OEP000155)* that were prepared with a strand specific library and sequenced using total RNA-seq with the rRNA depletion protocol from the National Omics Data Encyclopedia (NODE)[144] (http://www.biosino.org/node). RNA sequencing quality was evaluated for all the samples using fastqc[145] and multiqc[146]. *HISAT2[147]* index was built using GRCh38 genome and gencode v27 annotation files with --ss and --exon options (**Figure 1**). RNA-seq reads were aligned to the GRCh38 reference genome using *HISAT2[147],* followed by reference based assembly with *StringTie[147]* using gencode v27 annotation. Next, a transcriptome was finally created by merging all assemblies using --merge option from *StringTie*. Raw read counts were extracted using *featureCounts[148]*. TPM was extracted using *TPMcalculator[149]*. Raw fastqs of 1,040 breast cancer samples were downloaded from The Cancer Genome Atlas (TCGA) data portal[150]. The data was aligned to the GRCh38 genome using *HISAT2[147]* and raw counts extracted with *featureCounts[148].*

3.2 Identification of novel lncRNAs

In addition to gencode-annotated lncRNAs, for identifying novel lncRNAs we used *gffcompare[151]* output to compare the assembled transcriptome to the annotation file and look for transcripts assigned with classification codes that shows a relationship between a transcribed fragment and the near by reference transcript *(Figure 1)*. From the output

of *gffcompare,* annotated transcripts found in gencode and *lncRNAKB*[42] were set side and focussed on unannotated transcripts. The unannotated transcripts with class code "i", "u", "x" that represent intronic, intergenic and antisense to protein-coding genes were then further filtered. First, all the unannotated transcripts with length less than 200 bp and without exons were removed. Next, based on the expression, single exon transcripts with FPKM > 1 in at least one sample and multi exon transcripts with FPKM > 0.5 in at least one sample were kept for further steps. These transcripts were then used as input into *lncRNApipe*[152] which uses *CPC2, RNAFold* and *Infernal* tools to detect novel lncRNA transcripts. FASTA sequences of the transcripts were used to calculate coding potential by *CPC2*[153] with *UniRef90*[154] protein database and transcripts with coding probability > 0.5 were flagged as "coding" and removed from the analysis. The resulting ncRNAs were then used to predict secondary structure minimum free energy using *RNAFold*[155] from the *ViennaRNA*[155] package. The lower the entropy, the more stable it is and thus more likely to be a ncRNA transcript. The remaining ncRNA transcripts were further annotated using *cmscan* from *Infernal*[156] which searches *Rfam*[157] database and discards transcripts that have any match with other RNA families. Finally, the lncRNA transcripts found after using all the tools and filtering options, were termed as novel lncRNAs.

3.3 TNBC subtype classification

For TNBC subtype classification, Lehmann et. al subtyping approach was used[15]. Initially, variance stabilizing transformations (VST) function from *DESeq2*[158] was applied on raw counts data to produce transformed data on log2 scale which is normalized with

respect to library sizes. The normalized data was then used as an input into the TNBCtype[14] subtyping tool (https://cbc.app.vumc.org/tnbc/) to classify TNBCs. It uses gene signature (n=2,188) for TNBC subtype prediction. The output from the tool shows each sample is assigned to one of the TNBC subtypes (BL1, BL2, M, LAR, MSL and IM) based on the highest correlation coefficient and p-value < 0.05. The transformed data of 317 TNBCs and 44 normal samples of the FUSCC cohort was used for principal component analysis to check the distribution between the samples.

3.4 Differential Expression Analysis

Annotated lncRNAs and novel lncRNAs detected, were used for differential expression analysis between the subtypes of TNBCs in the FUSCC cohort. LncRNAs with low expression (< 1 log-counts per million in ≥ 30 samples) were filtered out. Using negative binomial distribution, raw counts were modeled and a Wald test was applied from DESeq2[158] to perform differential expression analysis. Differentially expressed lncRNAs were identified based on $| log2FC | > 1$ and FDR < 0.05. Followed by, differential analysis was also performed between TNBCs (n=317) and normals (n=44) in the FUSCC cohort for both annotated and novel lncRNAs. Similarly, differential analysis for annotated lncRNAs was also performed between TNBCs (n=178) and non-TNBCs (n=862) in the TCGA cohort.

3.5 Co-expression network analysis

Co-expression network was constructed using *Weighted gene co-expression network analysis (WGCNA)*[159] R package with 14,020 protein-coding genes and 12,390 lncRNAs (both annotated and novel lncRNAs) after filtering out genes which are not expressed in at least 10 samples. Modules were identified using a dynamic tree cut approach, followed by calculating the moduleEigengenes (1st principal component of each module). To identify the relationship between modules and phenotypes, spearman correlation was applied between the eigengenes (1st principal component of each module) and TNBC subtypes. As we were primarily interested in upregulated lncRNAs, modules which show significant positive correlation with the subtype and having higher number of lncRNAs were considered as interesting modules for further analysis.

3.6 Functional enrichment analysis

Protein-coding genes correlated with any of the lncRNAs in the modules were used as targets for identifying the lncRNA function. For this, modules with higher numbers of lncRNAs co-expressed with protein-coding genes and their association with subtypes of TNBCs, were selected. All the protein-coding genes in the selected modules were used for Pathway enrichment analysis. We used *ClusterprofileR*[160] and applied Gene Ontology, KEGG and Reactome analyses for estimating enrichment in biological processes and pathways for interesting modules. Pathways with a *p-adjust < 0.05* were selected.

3.7 Immune lncRNA identification

Immune lncRNAs were identified with a three step process. 1) lncRNA - Immune pathways association 2) lncRNA - Immune cells correlation 3) lncRNA - Tumor purity correlation.

lncRNAs and immune pathways association:

To identify lncRNAs that associate with immune pathways, the computational approach *ImmLnc[4]* was used. In brief, for each lncRNA-protein coding (PC) gene pair, a partial correlation coefficient was calculated incorporating the pairwise correlation coefficients between the lncRNA expression, the PC gene expression and tumor purity, where tumor purity was estimated using *ESTIMATE[161]*. The rank score for the lncRNA-PC gene pair was then defined as the -log10(P-value) of the partial correlation, signed by the direction of the partial correlation coefficient. For each lncRNA, the PC genes were then ranked based on rank scores with the given lncRNA and used for gene set enrichment analysis (GSEA) of immune pathways derived from *Immport[162]* an open access data portal. Combining the enrichment scores and p-values from the GSEA analysis, a lncRNA enrichment score (lncRES), ranging from -1 to 1, was calculated for each immune pathway. lncRNAs positively and negatively associated with immune pathways were selected based on |lncRES score| > 0.995 and p-adjusted value (FDR) < 0.05.

Association between lncRNAs and immune cells:

To check the association between lncRNAs and Immune cells, we first estimated the immune infiltration levels from Tumor Immune Estimation Resource (TIMER 2.0)[163] database using the gene expression data of FUSCC TNBCs. Spearman's rank correlation was applied between the lncRNA expression data and immune cells infiltration levels. lncRNAs correlated with immune cells were identified based on $|r| > 0.4$ and p-value < 0.05.

Correlation between lncRNAs and tumor purity:

Similarly, correlation between lncRNAs expression and tumor purity (computed using *ESTIMATE*[161]) of each sample, was calculated using spearman correlation and selected lncRNAs based on correlation coefficient $r < -0.4$ and significant p-adjusted value.

The final set of immune lncRNAs were defined as the overlap between all the above three steps (lncRNAs - immune pathways, lncRNAs - immune cells and lncRNAs - tumor purity).

3.8 Immune lncRNA score calculation with ssGSEA

Using the expression of immune lncRNAs, *ssGSEA*[164] was applied to calculate immune lncRNA score using the *GSVA*[165] R package. Based on the median of immune lncRNA score, immune clusters Immune-High (IH) and Immune-Low (IL) were identified.

3.9 Immunogenicity of the samples

Immunophenoscore (IPS) is an aggregated score calculated using the *getIPS* function supplied by The cancer Immunome Atlas[166], which is based on the expression of representative genes of four immunophenotypes in tumor samples: MHC molecules (antigen processing), immunomodulators, suppressor cells (Tregs and MDSCs) and effector cells (CD8+ T cells and CD4+ T cells)[166]. It also calculates mean expression (z-score) of all the four immune genesets. The higher IPS aggregated score, the more immunogenic the samples are, which is correlated with response to immunotherapy[166]. Tumor inflammation signature (TIS) is a score which is associated with response to anti-PD-1 treatment[167]. TIS score is estimated by applying gene set variation analysis (GSVA) on the expression of 18 gene signatures that enriches for clinical benefit of immunotherapy. Tumors with higher TIS score are associated with better response to anti-PD-1 blockade[168]. Using the gene expression data other immune related scores of TNBCs were calculated. Mean of the expression levels of two cytolytic effectors granzyme A (*GZMA*) and perforin (*PRF1*), a quantitative score of cytolytic activity (*CYT*) has been calculated[169]. Similarly, based on the expression levels of MHC-I geneset[170], a mean expression was calculated which is regarded as MHC score for TNBC patients. Immune score which represents immune cell infiltration, was calculated from the *ESTIMATE*[161] R package. Stromal TILs (sTILs) and Intratumoral TILs (iTILs) information from H&E staining of TNBC patients was used from the clinical data of the FUSCC cohort[144]. *Consensus TME*[171] is an integrative tool which uses multiple immune genesets to calculate the abundance of multiple cell types of TNBC samples.

3.10 Identification of most important lncRNAs related to Immune High

To identify the significant important lncRNAs related to the immune-high cluster, we divided 317 TNBCs randomly into training (75% of samples) and testing (remaining 25% of samples) datasets. Using the training dataset, differential expression analysis was performed between identified Immune-High and Immune-Low clusters. We applied an algorithm *biosigner*[172] to select lncRNAs that are most important for classifier performance. It iterates through feature selection steps which include resampling, ranking of lncRNAs according to their importance in the model and the identification of significant lncRNAs until the model is restricted to significant lncRNAs or when there is no feature left for evaluation. This algorithm is wrapped around three binary classifiers: partial least squares discriminant analysis (PLS-DA), random forest (RF) and support vector machines (SVM). As a final result, the algorithm assigns each significant lncRNA to a tier for each of the three machine learning approaches. S tier is for lncRNAs in the final signature, where lncRNAs are regarded significant in all iterations of feature selection; A tier - lncRNAs found significant in all but the last iteration followed by B, C, D and E tiers. The performance of the final models was evaluated on the testing dataset.

4 Results

4.1 Identification of novel lncRNAs in TNBCs

To identify novel lncRNAs in TNBCs , we downloaded data for 317 TNBCs and 44 normals of the FUSCC cohort which was sequenced using paired end strand-specific total RNA-seq (rRNA depletion protocol). All the samples were aligned to the GRCh38 reference genome using *HISAT2* and *ab-initio* assembly with *StingTie.* A total of 15,754 lncRNAs and 19,802 protein coding genes were annotated by gencode v27[34] gene annotation in FUSCC samples *(Figure 4.1A)*. Annotated transcripts and protein-coding genes were excluded from the detection of novel lncRNAs and the remaining transcripts that represent intronic, intergenic and antisense to protein-coding genes were then used to detect novel lncRNAs.

We applied bioinformatic tools and filters to detect novel lncRNAs *(Figure 4.1A)*. For this, the unannotated transcripts with class code "i", "u", "x" that represent intronic, intergenic and antisense to protein-coding genes were used. Initially, some of the transcripts were filtered out that have length < 200 bp and without any exons. Followed by, single exon transcripts with FPKM > 1 in at least one sample and multi exon transcripts with FPKM > 0.5 in at least one sample were used for further analysis. *CPC2* (Coding Potential Calculator)[153] was run on transcripts to estimate coding potential and removed transcripts that have coding probability > 0.5 which were flagged as "coding". The ncRNA transcripts were then filtered based on the stability of the secondary structure using *RNAFold* followed by structure and sequence similarities with any other

RNA families from the *Rfam* database using *Infernal.* From the above filtering steps we finally identified a total of 5,716 novel lncRNAs (12,372 lncRNA transcripts) **(Figure 4.1A)**. Among them 86.7% were located in intergenic regions (Intergenic), 2.4% were antisense exonic overlap with protein-coding genes (Antisense Exonic), 10.7% were antisense intronic that overlap with protein coding genes **(Figure 4.1B)**.
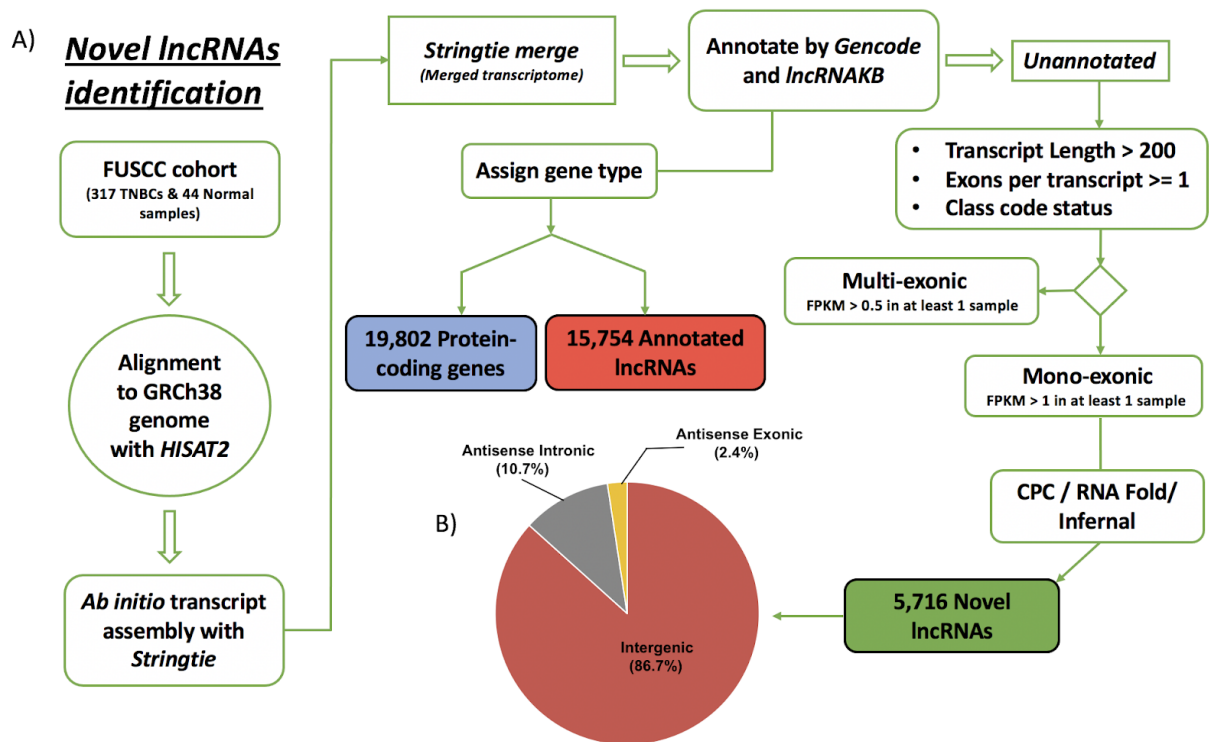
Figure 4.1*: Identification and characterization of novel lncRNAs in TNBCs A) Computational workflow to identify novel lncRNAs in 361 TNBCs. B) Pie chart showing identified novel lncRNAs gene types percentage.*

Looking at the characteristics of the detected novel lncRNAs, they were shorter than annotated lncRNAs and protein-coding transcripts in length **(Figure 4.2A)**. Novel lncRNA transcripts have fewer exons when compared with annotated lncRNAs and

protein-coding transcripts *([NL: median=2.0, mean=2.4]; [AL: median=3.0, mean=3.7]; [PC: median=6.0, mean=8.8])* *(Figure 4.2B)*. We observed that the novel lncRNAs were low expressed *([NL: median=1.9, sd=1.1]; [AL: median=2.1, sd=1.3]; [PC: median=3.7, sd=1.9])* than protein-coding and annotated lncRNA transcripts (*p-value* < 0.001) *(Figure 4.2C)*. Also we found that, similar to annotated lncRNAs, novel lncRNA transcripts have extremely low coding probability when compared to protein-coding transcripts. There is a good separation between the coding potential score distributions *(Figure 4.2D)*. The 317 TNBC samples were categorized into BL1 (n=65), BL2 (n=24), M (n=55), LAR (n=62), MSL (n=39) and IM (n=72) using TNBCtype[14] subtyping tool.
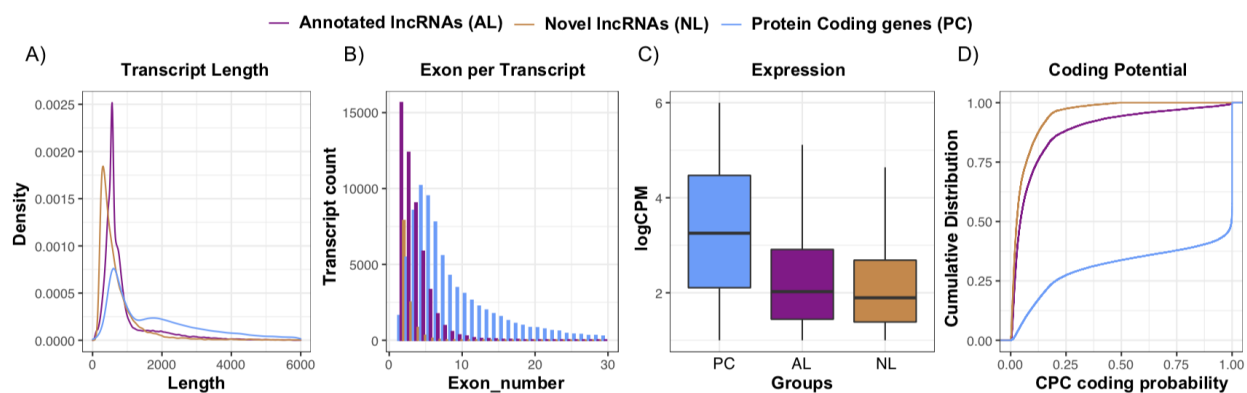
Figure 4.2: Characterization of identified novel lncRNAs comparing with protein-coding genes, annotated lncRNAs A) Transcript length B) Exon number C) Expression D) Coding Potential.

4.2 Differentially expressed lncRNAs in TNBCs

To identify lncRNAs that defined each TNBC subtype, initially raw read counts of lncRNAs (including annotated lncRNAs and novel lncRNAs) were obtained using *featureCounts[148]*. Using negative binomial distribution, raw counts were modeled and a

Wald test was applied from *DESeq2*[158] to perform differential expression analysis. We identified 2,039 lncRNAs (648 up-regulated and 1391 down-regulated in BL1) between BL1 and other subtypes combined, 518 lncRNAs (202 up-regulated and 316 down-regulated in BL2) between BL2 and other subtypes combined, 1684 lncRNAs
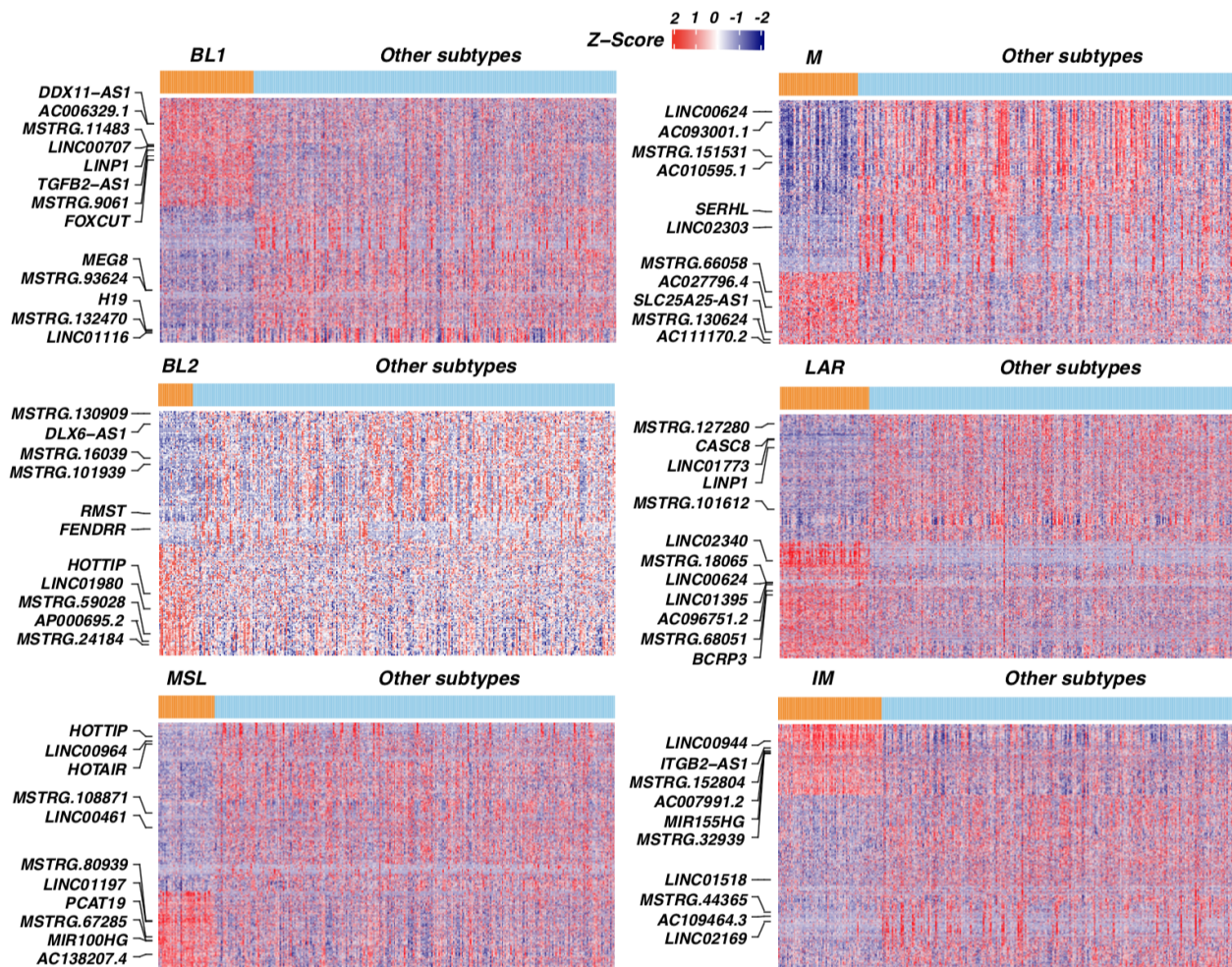
Figure 4.3: *Differential expression analysis between each subtype of TNBCs compared to remaining subtypes combined. Heatmap depicts differentially expressed lncRNAs which were selected based on |log2FC| > 1 and FDR < 0.05. Highlighted annotated lncRNAs were already reported in different cancer types.*

(490 up-regulated and 1194 down-regulated in IM) between IM and other subtypes combined, 1593 lncRNAs (564 up-regulated and 1029 down-regulated in M) between M and other subtypes combined, 2055 lncRNAs (497 up-regulated and 1558 down-regulated in MSL) between MSL and other subtypes combined, 2503 lncRNAs (1309 up-regulated and 1194 down-regulated in LAR) between LAR and other subtypes combined were considered differentially expressed based on | log2FC | > 1 and false discovery rate (FDR) < 0.05 *(Figure 4.3)*.

Some of the overexpressed lncRNAs specific to the subtypes in this study have been reported previously about their role in different cancers. BL1 subtype lncRNAs such as, *AC006329.1* is one of the recurrence associated six-lncRNAs identified that distinguish high risk and low risk patients in colon cancer[173]. *LINC00707* was upregulated and led to cancer progression in bladder cancer[174] and also promotes cell proliferation and invasion in colorectal cancer through *miR-206/FMNL2* axis[175]. *LINP1* promotes acute myeloid leukemia progression through HNF4α/AMPK/WNT5A signaling pathway and might be used as a therapeutic target for treatment[176]. It also regulates dna double strand break repair in TNBCs[116]. *BANCR* (BRAF activated non-coding RNA) acts as an oncogene in melanoma[177], lung cancer[178], gastric[179] and colorectal cancers[180] associated with clinical progression and poor prognosis. *BLACAT1* expression is upregulated in colorectal cancer which leads to poor prognosis and affects cell proliferation[181]. *DDX11-AS1* promotes cell-cycle progression, cancer development in hepatocellular carcinoma[182], non-small cell lung cancer[183], gastric cancer[184] and colorectal cancer[185]. *FOXP4-AS1* is overexpressed and regulates proliferation in colorectal cancer[186],

promoting the growth of prostate cancer[187]. *MEG8* lncRNA, which is downregulated in the BL1 subtype, is also downregulated in osteoarthritis, regulates apoptosis and might be a target for treatment[188]. Other lncRNAs *LINC01116*, *H19* were also downregulated in the BL1 subtype.

BL2 subtype upregulated lncRNAs, for example *HOTTIP* has been reported as an oncogene which is associated with cell proliferation and migration in lung cancer[189], hepatocellular carcinoma[190], colorectal cancer[191]. Upregulation of *LINC01980* in esophageal squamous cell carcinoma enhances tumor progression[192]. *LINC00520* promotes metastasis and tumor growth in melanoma[193]. *FENDRR* is downregulated in the BL2 subtype, also showed downregulation in gastric cancer, associated with poor prognosis and regulates metastasis[194]. *DLX6-AS1* downregulation decreased cell proliferation, migration and invasion in epithelial ovarian cancer[195].

M subtype overexpressed lncRNAs such as *SLC25A25-AS1* play an oncogenic role in non-small cell lung cancer through *miRNA-195-5p/ITGA2* axis[196]. Other lncRNAs *AC027796.4, AC111170.3, AC111170.2* were upregulated and *FENDRR* is downregulated in M subtype which has been reported as a chromatin modifier which affects gene expression[83].

MSL subtype upregulated lncRNAs include *MIR100HG* which is overexpressed and linked to cetuximab resistance through wnt signaling along with miRNAs *miR100* and *miR125b*[197]. *MEG3 which* promotes tumor development of osteoblasts through

wnt/beta-catenin signaling[198]. In addition, several other lncRNAs (for example: *AC138207.4, LINC01013, LINC01197, PCAT19*) were also upregulated, whereas others (for example: *HOTTIP, HOTAIR*) were downregulated in MSL subtype.

LAR subtype overexpressed lncRNAs, for instance *LINC00624* acts as a decoy and promotes hepatocellular carcinoma progression[199]. *BCRP3* has been part of the ceRNA network and found as a potential biomarker in lung adenocarcinoma. CASC8 is downregulated in the LAR subtype, and is involved in regulating glycolysis in bladder cancer[200]. Along with these, several other known lncRNAs have been upregulated and downregulated in the LAR subtype.

IM subtype upregulated lncRNAs, *ITGB2-AS1* associated with immune signatures and chemotherapy outcome in acute myeloid leukemia[201], *LINC00944* immune related lncRNA associated with breast cancer prognosis, *LINC01871* is among the immune lncRNA signatures in treating melanoma[202] and *MIR155HG* a biomarker associated with immune checkpoint expression levels in multiple cancers[203]. In particular, we also identified novel lncRNAs upregulated and downregulated in respective subtypes. Therefore, primarily these deregulated lncRNAs can be looked at to predict their functional and regulatory mechanisms related to the subtypes of TNBCs with *in-silico* approaches.

With the principal component analyses of the 361 samples (317 TNBCs and 44 normals) we observed that the samples were clustered together with respective to their

subtypes and displayed a similar distribution based on the expression of protein-coding genes, annotated lncRNAs and novel lncRNAs *(Figure 4.4)*. The LAR subtype is far away from the other subtypes in PCA and this observation is consistent with the identification of most differentially expressed lncRNAs specific to the LAR subtype.
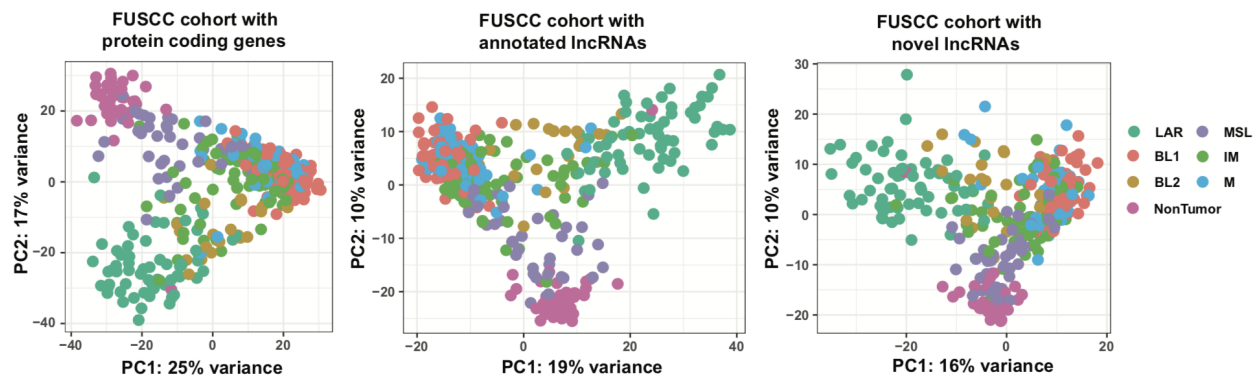
Figure 4.4: PCA showing the samples segregation with respect to their subtypes based on coding genes, annotated lncRNAs and novel lncRNAs.

4.3 Functional evaluation of lncRNAs with co-expression network

To identify the biological pathways in which lncRNAs are involved in each subtype of TNBCs, we constructed a co-expression network analysis with 14,020 protein-coding genes (PC) and 12,390 lncRNAs (both annotated (AL) and novel lncRNAs (NL)). Using Weighted gene co-expression network analysis (WGCNA) we identified 20 modules *(Figure 4.5A)*.

As lncRNAs have low expression compared to mRNAs, we focussed on upregulated lncRNAs. We calculated module eigengenes (1st principal component of modules) and applied spearman correlation to define the relationship between the modules and the

subtypes. Among the identified modules, M8 (r = 0.33, p-value < 0.001; PC = 66, AL = 7, NL = 4), M5 (r = 0.34, p-value < 0.001; PC = 56, AL = 1, NL = 0), M1 (r = 0.44, p-value < 0.001; PC = 410, AL = 21, NL = 6) and M17 (r = 0.33, p-value < 0.001; PC = 65, AL = 1, NL = 0) were having a significant positive corrrelation with BL1 subtype. We observed that M3 (r = 0.2, p-value < 0.001; PC = 221, AL = 12, NL = 3) is the only module that has a positive correlation with BL2 subtype. Module M2 (r = 0.61, p-value < 0.001; PC = 1063, AL = 213, NL = 53) had a strong positive significant corrrelation with IM subtype. M18 (r = 0.17, p-value = 0.003; PC = 52, AL = 2, NL = 0), M20 (r = 0.16, p-value = 0.004; PC = 58, AL = 2, NL = 0), M8 (r = 0.12, p-value = 0.04; PC = 66, AL = 7, NL = 4), M16 (r = 0.12, p-value = 0.04; PC = 312, AL = 21, NL = 1), M1 (r = 0.17, p-value = 0.003; PC = 410, AL = 21, NL = 6) and M5 (r = 0.17, p-value = 0.002; PC = 56, AL = 1, NL = 0) modules showed a positive correlation with M subtype. M9 (r = 0.5, p-value < 0.001; PC = 330, AL = 20, NL = 6) is the only module that showed strong positive significant correlation with the MSL subtype. Modules M15 (r = 0.66, p-value < 0.001; PC = 55, AL = 28, NL = 19) and M19 (r = 0.32, p-value < 0.001; PC = 60, AL = 5, NL = 0) were positively correlated with the LAR subtype *(Figure 4.5B)*.

For identifying lncRNA function, we selected 6 interesting modules one for each subtype (M1 ~ BL1, M3 ~ BL2, M2 ~ IM, M15 ~ LAR, M16 ~ M, M9 ~ MSL) based on the highest number of lncRNAs (upregulated annotated and novel lncRNAs) co-expressed with protein coding genes and significant positive correlation between modules and subtypes *(Figure 4.6A, Table 4.1)*. Pathway Analysis was performed on the

protein-coding genes in these modules to identify pathways related to each subtype.

(Figure 4.6B, Appendix Table 1)

Table 4.1: Modules identified from WGCNA with number of co-expressed protein-coding and lncRNAs			
Modules	**Protein-coding**	**Annotated LncRNAs**	**Novel LncRNAs**
M19	60	5	0
M13	71	1	0
M12	50	3	0
M6	1435	54	1
M18	52	2	0
M8	66	7	4
M20	58	2	0
M4	253	1	0
M10	45	3	1
M16	312	21	1
M1	410	21	6
M5	56	1	0
M14	48	4	0
M17	65	1	0
M7	84	0	0
M11	49	1	0
M15	55	28	19
M2	1063	213	53
M3	221	12	3
M9	330	20	6
Colored modules are selected for further analysis based on the number of co-expressed annotated and novel lncRNAs along with Module - Phenotype correlation			

A)

Identified modules

B)

Legend:
- Protein coding genes
- Upreg Annotated lncRNAs
- Upreg Novel lncRNAs

Module – Phenotype relationships

Module	BL1	BL2	M	IM	MSL	LAR
M19 (60, 5, 0)	0.21 (2e-04)	0.052 (0.4)	0.084 (0.1)	-0.39 (9e-13)	-0.29 (2e-07)	0.32 (7e-09)
M13 (71, 1, 0)	0.15 (0.007)	-0.097 (0.08)	0.083 (0.1)	-0.031 (0.6)	-0.037 (0.5)	-0.11 (0.06)
M12 (50, 3, 0)	0.11 (0.04)	-0.089 (0.1)	0.08 (0.2)	0.0011 (1)	-0.047 (0.4)	-0.096 (0.09)
M6 (1435, 54, 1)	0.054 (0.3)	-0.04 (0.5)	-0.0096 (0.9)	0.033 (0.6)	-0.029 (0.6)	-0.03 (0.6)
M18 (52, 2, 0)	0.28 (3e-07)	-0.078 (0.2)	0.17 (0.003)	0.017 (0.8)	-0.13 (0.02)	-0.31 (3e-08)
M8 (66, 7, 4)	0.33 (2e-09)	-0.057 (0.3)	0.12 (0.04)	-0.083 (0.1)	-0.32 (4e-09)	-0.053 (0.3)
M20 (58, 2, 0)	0.16 (0.004)	-0.048 (0.4)	0.16 (0.004)	-0.016 (0.8)	-0.23 (5e-05)	-0.084 (0.1)
M4 (253, 1, 0)	0.17 (0.002)	0.017 (0.8)	-0.016 (0.8)	-0.08 (0.2)	-0.3 (7e-08)	0.16 (0.005)
M10 (45, 3, 1)	0.14 (0.02)	-0.044 (0.4)	0.12 (0.04)	-0.098 (0.08)	-0.14 (0.01)	0.001 (1)
M16 (312, 21, 1)	0.092 (0.1)	-0.048 (0.4)	0.12 (0.04)	-0.057 (0.3)	-0.12 (0.04)	-0.016 (0.8)
M1 (410, 21, 6)	0.44 (9e-17)	-0.024 (0.7)	0.16 (0.003)	0.06 (0.3)	-0.36 (3e-11)	-0.36 (6e-11)
M5 (56, 1, 0)	0.33 (2e-09)	-0.069 (0.2)	0.17 (0.002)	-0.034 (0.5)	-0.23 (5e-05)	-0.23 (4e-05)
M14 (48, 4, 0)	0.27 (1e-06)	0.069 (0.2)	0.16 (0.005)	-0.012 (0.8)	-0.2 (2e-04)	-0.29 (2e-07)
M17 (65, 1, 0)	0.33 (1e-09)	-0.0027 (1)	0.08 (0.2)	0.01 (0.9)	-0.21 (1e-04)	-0.25 (8e-06)
M7 (84, 0, 0)	0.17 (0.003)	-0.026 (0.6)	0.074 (0.2)	-0.029 (0.6)	-0.15 (0.006)	-0.065 (0.3)
M11 (49, 1, 0)	0.15 (0.009)	-0.056 (0.3)	0.097 (0.08)	0.035 (0.5)	-0.14 (0.01)	-0.12 (0.03)
M15 (55, 28, 19)	-0.4 (1e-13)	0.066 (0.2)	-0.25 (9e-06)	-0.1 (0.07)	0.054 (0.3)	0.66 (1e-41)
M2 (1063, 213, 53)	-0.22 (5e-05)	-0.013 (0.8)	-0.42 (7e-15)	0.61 (3e-34)	0.19 (5e-04)	-0.17 (0.002)
M3 (221, 12, 3)	-0.44 (2e-16)	0.19 (6e-04)	0.11 (0.05)	-0.11 (0.04)	0.19 (6e-04)	0.18 (0.002)
M9 (330, 20, 6)	-0.48 (3e-19)	0.021 (0.7)	-0.14 (0.01)	0.1 (0.08)	0.49 (6e-21)	0.088 (0.1)

***Figure 4.5:** A) Gene dendrogram derived by the hierarchical clustering. The colors represent the identified modules by the Dynamic Tree Cut. B) Module – Phenotype relationship plot shows subtypes on x-axis, correlation coefficient and in the parentheses p-value were shown in the box. Identified modules were on the y-axis with the bar plot showing the number of protein-coding genes, annotated lncRNAs, novel lncRNAs which were co-expressed in the parentheses.*

Module M1 showed a significant positive correlation with BL1 subtype (r = 0.44, p-value < 0.001; PC = 410, AL = 21, NL= 6) in which protein-coding genes were enriched with pathways related to cell cycle, DNA replication and double strand break repair which play an important role in cell differentiation **(Figure 4.6B, Appendix Table 1)**. Among the 21 annotated lncRNAs, 6 lncRNAs *AC006329.1* (Antisense RNA located on chromosome 7), *AC013391.2* (sense overlapping RNA located on chromosome 5), *AC016205.1* (Intergenic lncRNA located on chromosome 18), *DDX11-AS1* (Antisense

RNA located on chromosome 12), *FOXP4-AS1* (Antisense RNA located on chromosome 6) and *AP001505.2* (Intergenic lncRNA located on chromosome 21) were upregulated in TNBCs when compared with normals in the FUSCC cohort and also upregulated in TNBCs when compared with non-TNBCs in the TCGA cohort *(Figure 4.7, Table 4.2)*.
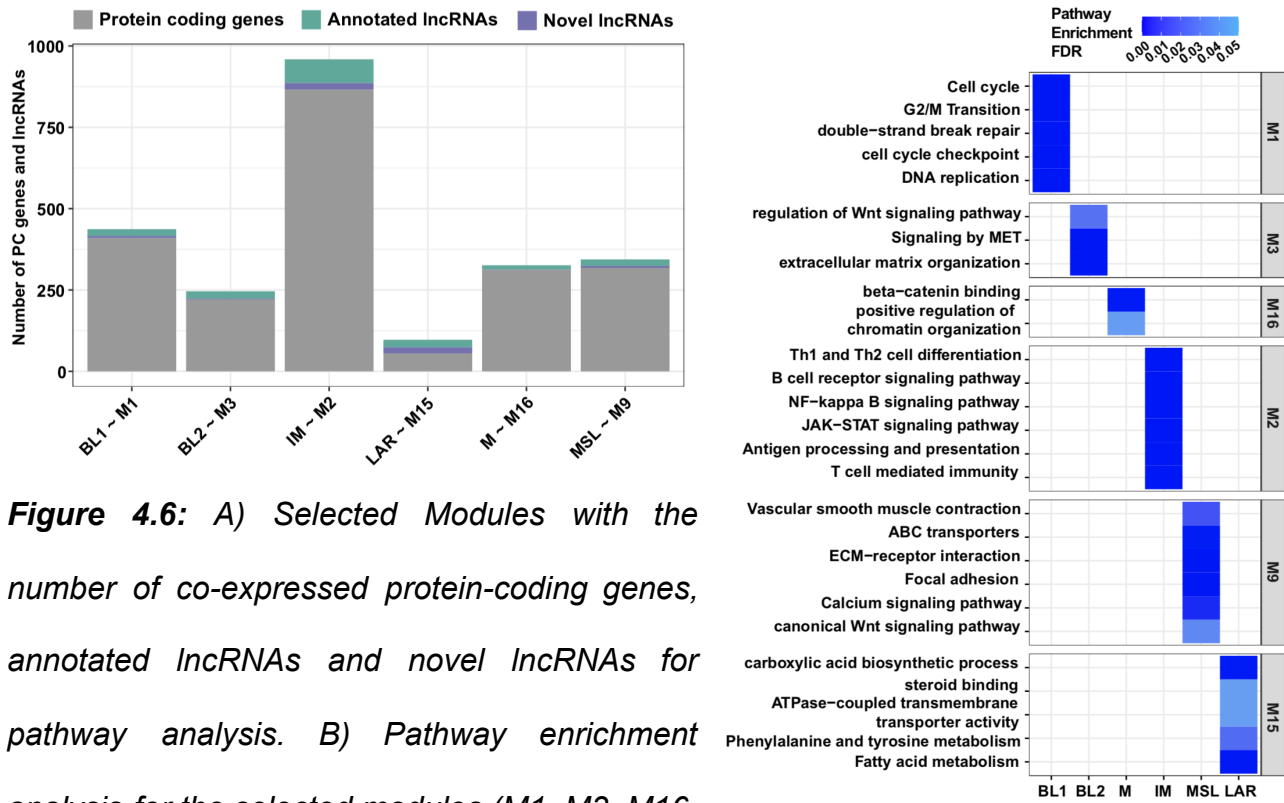
Figure 4.6: A) Selected Modules with the number of co-expressed protein-coding genes, annotated lncRNAs and novel lncRNAs for pathway analysis. B) Pathway enrichment analysis for the selected modules (M1, M3, M16, M2, M9 and M15). Most significantly enriched pathways which have FDR < 0.05 were displayed in the heatmap.

Among the novel lncRNAs in this **Module M1**, two of them *MSTRG.11483* (Antisense Exonic to *LRRC41*) and *MSTRG.9061* (Antisense Intronic to *MTFR1L*) located on chromosome 1 were also upregulated in TNBCs when compared with normals *(Figure 4.7, Table 4.2)*

IncRNA Type	GeneSymbols	WGCNA	FUSCC_BL1vs Other_log2FC	FUSCC_BL1vs Other_padj	FUSCC_TNBCvs Nor_log2FC	FUSCC_TNBCvs Nor_padj	TCGA_TNBCvs nonT_log2FC	TCGA_TNBCvs nonT_padj
Annotated	AC006329.1	M1 (BL1)	1.564306621	1.97E-09	3.114122711	5.88E-23	1.4309	7.56E-41
Annotated	AC013391.2	M1 (BL1)	1.023075524	4.48E-06	3.45499961	5.32E-18	0.59186	0.0313
Annotated	AC016205.1	M1 (BL1)	1.441486435	3.93E-13	2.487463076	3.04E-24	1.084	5.54E-21
Annotated	AP001505.2	M1 (BL1)	1.136935167	3.41E-12	1.111219846	1.78E-07	1.497	4.61E-52
Annotated	DDX11-AS1	M1 (BL1)	1.056538245	1.95E-10	1.70895777	1.26E-17	1.2021	3.94E-60
Annotated	FOXP4-AS1	M1 (BL1)	1.417952699	4.18E-12	2.177561536	1.26E-17	2.0093	1.51E-67
Novel	MSTRG.11483	M1 (BL1)	1.134830075	5.62E-13	3.309345386	1.50E-63	NA	NA
Novel	MSTRG.9061	M1 (BL1)	1.071385576	1.01E-06	2.776862036	5.34E-13	NA	NA

Table 4.2: IncRNAs upregulated in BL1 vs other subtypes and also upregulated in TNBCs vs Normals and TNBCs vs non-TNBCs

These IncRNAs are also associated with the Module M1 in WGCNA
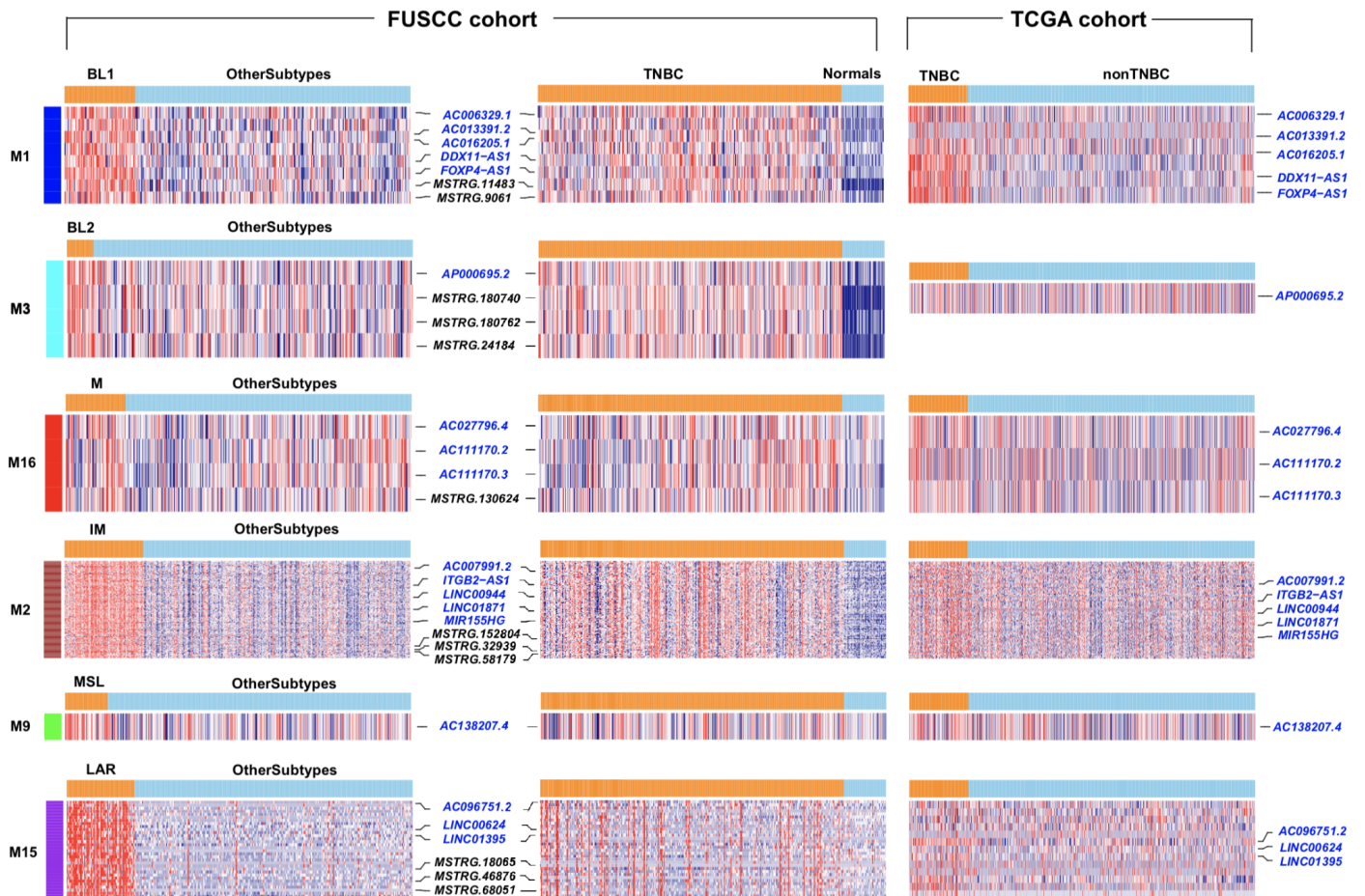
Figure 4.7: *Heatmap showing the expression of interesting lncRNAs that were co-expressed with protein-coding genes in the identified modules. They were overexpressed in their respective subtypes (vs Other subtypes) and TNBCs (vs Normals) in the FUSCC cohort. Annotated lncRNAs were also seen overexpressed in TNBCs (vs non-TNBCs) in the TCGA cohort.*

Module M3 showed association with BL2 subtype (r = 0.2, p-value < 0.001; PC = 221, AL = 12, NL = 3). The coding genes in M3 were enriched in pathways including extracellular matrix organization and signalling by MET *(Figure 4.6B, Appendix Table 1)*. Annotated lncRNA *AP000695.2* (Antisense RNA located on chromosome 21) and novel lncRNAs *MSTRG.180740*, *MSTRG.180762* (Intergenic lncRNAs located on chromosome 5) and *MSTRG.24184* (Intergenic lncRNA located on chromosome 1) were upregulated in BL2 (vs Other subtypes) and TNBCs (vs Normals) *(Figure 4.7, Table 4.3)*.

Module M16 was associated with M subtype (r = 0.12, p-value = 0.04; PC = 312, AL = 21, NL = 1) Protein coding genes co-expressed with lncRNAs are associated with pathways like beta-catenin binding and chromatin organization *(Figure 4.6B, Appendix Table 1)*. lncRNAs such as *AC027796.4* (Antisense RNA) and *AC111170.2* (sense intronic) were located on chromosome 17 and novel lncRNA *MSTRG.130624* (Antisense intronic to *MAVS*) was located on chromosome 20 were upregulated in the M subtype compared with other subtypes and also upregulated in TNBCs in the other comparisons *(Figure 4.7, Table 4.4)*

Table 4.3: lncRNAs upregulated in BL2 vs other subtypes and also upregulated in TNBCs vs Normals and TNBCs vs non-TNBCs								
These lncRNAs are also associated with the Module M3 in WGCNA								
lncRNA Type	GeneSymbols	WGCNA	FUSCC_BL2vs Other_log2FC	FUSCC_BL2v sOther_padj	FUSCC_TNBC vsNor_log2FC	FUSCC_TNB CvsNor_padj	TCGA_TNBCvsnon T_log2FC	TCGA_TNBCvsno nT_padj
Annotated	AP000695.2	M3 (BL2)	1.021902	0.001039	1.840553	7.61E-17	0.21515	0.037515
Novel	MSTRG.180740	M3 (BL2)	1.1399	2.40E-05	3.069444	1.95E-71	NA	NA
Novel	MSTRG.180762	M3 (BL2)	1.042025	0.0019399	3.330093	1.83E-27	NA	NA
Novel	MSTRG.24184	M3 (BL2)	1.068983	0.0190337	4.514372	2.02E-66	NA	NA

Table 4.4: lncRNAs upregulated in M (Mesenchymal) vs other subtypes and also upregulated in TNBCs vs Normals and TNBCs vs non-TNBCs								
These lncRNAs are also associated with the Module M16 in WGCNA								
lncRNA Type	GeneSymbols	WGCNA	FUSCC_Mvs Other_log2FC	FUSCC_Mvs Other_padj	FUSCC_TNBC vsNor_log2FC	FUSCC_TNBCvs Nor_padj	TCGA_TNBCvs nonT_log2FC	TCGA_TNBCvs nonT_padj
Annotated	AC027796.4	M16 (M)	0.512343	0.0009368	0.4179891	0.00957841	0.3679657	2.04E-13
Annotated	AC111170.2	M16 (M)	0.369826	0.0462604	0.569518	0.001212	0.610838	6.97E-05
Annotated	AC111170.3	M16 (M)	0.273536	0.0464764	0.4118934	0.001710	0.240441	0.029869
Novel	MSTRG.130624	M16 (M)	0.458688	0.0006191	0.540190	6.85E-05	NA	NA

Module M9 was significantly associated with MSL subtype (r = 0.49, p-value < 0.001; PC = 330, AL = 20, NL = 6), where protein coding genes were enriched in pathways related to focal adhesion, calcium signaling pathway, Wnt signaling pathway *(Figure 4.6B, Appendix Table 1)*. Among the lncRNAs in this module, only a single annotated lncRNA *AC138207.4* (sense intronic located on chromosome 17) was found interesting based on its upregulation in all the comparisons *(Figure 4.7, Table 4.5)*

Module M15 was significantly associated with the LAR subtype (r = 0.66, p-value < 0.001; PC = 55, AL = 28, NL = 19). Coding genes in M15 were related to metabolic processes including fatty acid metabolism, tyrosine metabolism *(Figure 4.6B, Appendix Table 1)*. Annotated lncRNAs in this module included *LINC00624* (Antisense RNA located on chromosome 1), *LINC01395* (intergenic lncRNA on chromosome 11) and 11 other annotated lncRNAs that were upregulated in LAR subtype (vs Other subtypes) and in TNBCs (vs Normals) in FUSCC. These lncRNAs were also upregulated in TNBCs when compared with non-TNBCs in the TCGA cohort. 19 novel lncRNAs were upregulated in LAR subtype and also upregulated in TNBCs when compared to normals in the FUSCC cohort *(Figure 4.7, Table 4.6)*.

Table 4.5: lncRNAs upregulated in MSL vs other subtypes and also upregulated in TNBCs vs Normals and TNBCs vs non-TNBCs								
These lncRNAs are also associated with the Module M9 in WGCNA								
lncRNA Type	GeneSymbols	WGCNA	FUSCC_MSLvsOther_log2FC	FUSCC_MSLvsOther_padj	FUSCC_TNBCvsNor_log2FC	FUSCC_TNBCvsNor_padj	TCGA_TNBCvsnonT_log2FC	TCGA_TNBCvsnonT_padj
Annotated	AC138207.4	M9 (MSL)	0.60511	8.69E-09	0.22872	0.02837	0.189326	0.0002640

Table 4.6: lncRNAs upregulated in LAR vs other subtypes and also upregulated in TNBCs vs Normals and TNBCs vs non-TNBCs								
These lncRNAs are also associated with the Module M15 in WGCNA								
lncRNA Type	GeneSymbols	WGCNA	FUSCC_LARvsOther_log2FC	FUSCC_LARvsOther_padj	FUSCC_TNBCvsNor_log2FC	FUSCC_TNBCvsNor_padj	TCGA_TNBCvsnonT_log2FC	TCGA_TNBCvsnonT_padj
Annotated	AC096751.2	M15 (LAR)	4.11354	3.01E-23	1.86771	0.006711	1.20978	0.000135
Annotated	AC112243.1	M15 (LAR)	5.1722	3.89E-28	3.43207	8.12E-05	2.9216	3.99E-08
Annotated	AL138881.1	M15 (LAR)	4.8289	4.33E-34	2.96894	1.22E-06	0.87540	0.000197
Annotated	AL353803.2	M15 (LAR)	4.05401	7.19E-39	1.57671	0.011026	1.13226	4.03E-06
Annotated	AL354719.2	M15 (LAR)	4.35503	3.36E-50	2.50768	5.97E-07	0.84371	0.000153
Annotated	AL355916.2	M15 (LAR)	4.85988	3.15E-25	2.74466	0.000544	0.60841	2.61E-05
Annotated	AL356804.1	M15 (LAR)	5.08723	7.08E-40	2.35281	0.000408	0.760616	0.008963
Annotated	AP000688.1	M15 (LAR)	3.90924	2.01E-53	1.93102	2.24E-06	0.83853	2.32E-22
Annotated	BCRP3	M15 (LAR)	3.48897	2.57E-61	1.92468	7.71E-07	0.297582	0.019586
Annotated	LINC00624	M15 (LAR)	5.57183	1.21E-29	4.80985	2.13E-12	1.34036	4.80E-44

Annotated	LINC01213	M15 (LAR)	3.24421	1.11E-32	2.26868	1.55E-06	0.502478	0.001892
Annotated	LINC01395	M15 (LAR)	4.627	3.89E-54	2.55204	5.23E-08	1.54568	1.14E-11
Annotated	SERHL	M15 (LAR)	5.0369	2.35E-60	2.53677	8.54E-08	0.26308	0.0001319
Novel	MSTRG.134576	M15 (LAR)	6.5190	9.18E-36	2.1012	0.003763	NA	NA
Novel	MSTRG.147129	M15 (LAR)	3.00198	2.22E-36	1.81308	3.07E-07	NA	NA
Novel	MSTRG.151531	M15 (LAR)	6.26632	3.15E-31	2.1783	0.011647	NA	NA
Novel	MSTRG.151876	M15 (LAR)	4.61020	5.69E-17	3.0178	0.00318	NA	NA
Novel	MSTRG.164552	M15 (LAR)	4.96758	5.51E-35	3.05017	3.54E-08	NA	NA
Novel	MSTRG.164590	M15 (LAR)	6.1196	1.73E-47	4.65654	3.79E-14	NA	NA
Novel	MSTRG.168187	M15 (LAR)	4.06378	1.37E-19	3.10725	1.21E-07	NA	NA
Novel	MSTRG.168188	M15 (LAR)	3.54981	3.63E-11	2.25059	0.011247	NA	NA
Novel	MSTRG.173491	M15 (LAR)	3.78888	1.81E-18	1.6941	0.01719	NA	NA
Novel	MSTRG.18065	M15 (LAR)	4.497092	1.54E-16	4.18623	1.48E-06	NA	NA
Novel	MSTRG.199816	M15 (LAR)	4.05619	1.96E-43	2.70232	6.34E-11	NA	NA
Novel	MSTRG.220493	M15 (LAR)	4.72934	2.06E-21	2.14181	0.002780	NA	NA
Novel	MSTRG.37951	M15 (LAR)	4.02430	6.76E-25	2.69451	3.42E-06	NA	NA
Novel	MSTRG.46876	M15 (LAR)	5.59951	1.23E-44	4.95237	2.56E-19	NA	NA
Novel	MSTRG.51267	M15 (LAR)	4.99786	2.21E-20	3.32044	1.77E-06	NA	NA
Novel	MSTRG.59089	M15 (LAR)	3.7384	5.61E-33	2.05059	0.000133	NA	NA
Novel	MSTRG.68051	M15 (LAR)	4.98243	1.53E-16	2.64519	0.014355	NA	NA
Novel	MSTRG.91115	M15 (LAR)	5.10277	1.00E-63	2.15576	3.78E-06	NA	NA
Novel	MSTRG.9506	M15 (LAR)	2.96637	4.83E-13	2.68869	2.89E-08	NA	NA

Module M2, associated with the IM subtype (r = 0.61, p-value < 0.001; PC = 1063, AL = 213, NL = 53), was enriched with immune-response genes involved in B-cell receptor, T-cell mediated pathway, JAK-STAT signaling pathway, antigen processing and presentation *(Figure 4.6B, Appendix Table 1)* . A total of 130 annotated and 43 novel lncRNAs in Module M2 were also upregulated in all respective differential analysis comparisons *(Figure 4.7, Appendix Table 2)*.

These co-expression relationships provide functional evidence demonstrating that the identified modules were enriched for biological processes and pathways from Gene Ontology (GO), KEGG and REACTOME. Among the identified modules, 6 modules include a higher number of lncRNAs, both annotated and novel lncRNAs, suggesting that these lncRNAs might play an important role in regulating the processes that support the TNBC subtypes.

4.4 The Identification of immune lncRNAs in TNBCs

A three step process was applied to identify immune lncRNAs. 1) lncRNAs association with immune pathways 2) lncRNAs correlation with immune cells infiltration 3) lncRNAs correlation with tumor purity *(Figure 4.8)*.
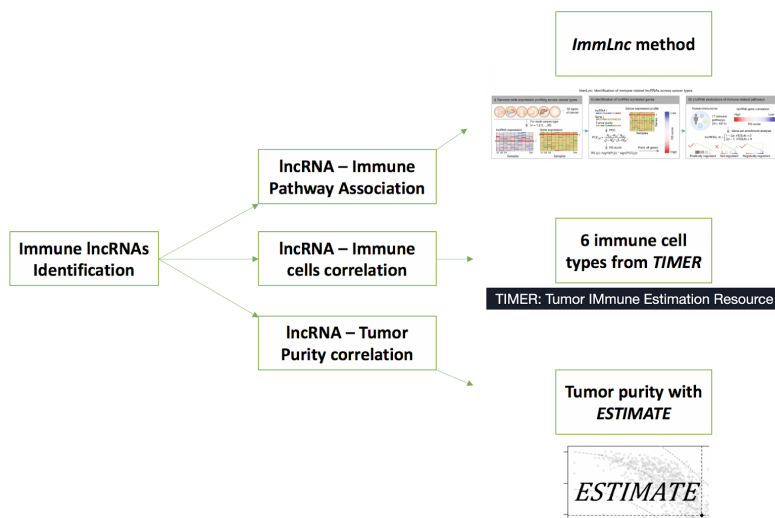
Figure 4.8: Identification of *Immune lncRNAs in TNBCs.* Flowchart showing 3 steps to identify Immune lncRNAs.

As a first step to identify lncRNAs that associate immune pathways, we applied a computational approach *ImmLnc[4]*, where we used protein coding (PC), annotated lncRNAs and novel lncRNAs expression data of 317 FUSCC TNBCs with tumor purity as a covariable, focussing on widely used 16 Immune pathways from *ImmPort[204]*. We identified 2,201 lncRNAs (1,782 annotated lncRNAs and 419 novel lncRNAs) that were positively associated with 16 immune pathways based on lncRES score > 0.995 and p-adjusted value (FDR) < 0.05. Similarly, 1,577 lncRNAs (1,276 annotated lncRNAs and 301 novel lncRNAs) were negatively associated with 16 immune pathways based on lncRES score < -0.995 and p-adjusted value (FDR) < 0.05 *(Figure 4.9)*. Majority of lncRNAs were associated with immune pathway antigen Processing and Presentation,

followed by TCR signaling pathway, Cytokine Receptors, Cytokines and BCR signaling pathway. Immune pathways such as Antigen Processing and Presentation[205], Cytokines[206] and TCR signaling[207] were used as potential targets for Immunotherapy in treating cancer *(Figure 4.9)*.
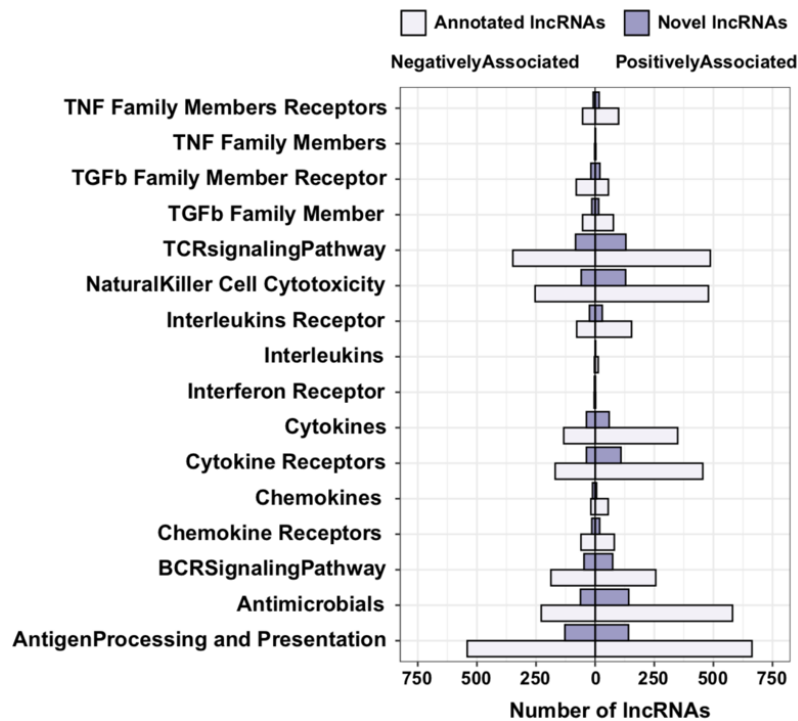
Figure 4.9: Bar plot showing the number of lncRNAs identified with immune pathways regulation.

It has been reported that the immune reaction by immune cell infiltration is potentially important in the tumor microenvironment (TME)[208]. As we observed that lncRNAs are associated with immune pathways, one could hypothesize that these lncRNAs may also correlate with immune cell infiltration in tumors. To evaluate the correlation between lncRNAs and Immune cells, we first estimated the immune infiltration levels from the Tumor Immune Estimation Resource (TIMER 2.0)[163] database using the gene expression data of FUSCC TNBCs. Spearman's rank correlation was then computed between the lncRNA expression data and immune cells infiltration levels. We identified

1,006 lncRNAs (804 annotated and 202 novel lncRNAs) that showed significant positive

correlation with any of the immune cell infiltrates (285 lncRNAs for B-cells, 777 lncRNAs

for CD4+ T-cells, 198 lncRNAs for CD8+ T-cells, 296 lncRNAs for Neutrophils, 418

lncRNAs for Myeloid dendritic cells and 97 lncRNAs for Macrophages) based on r > 0.4

and p-adjusted value < 0.05, but didn't find any lncRNAs negative correlation with

immune cells *(Figure 4.10)*.
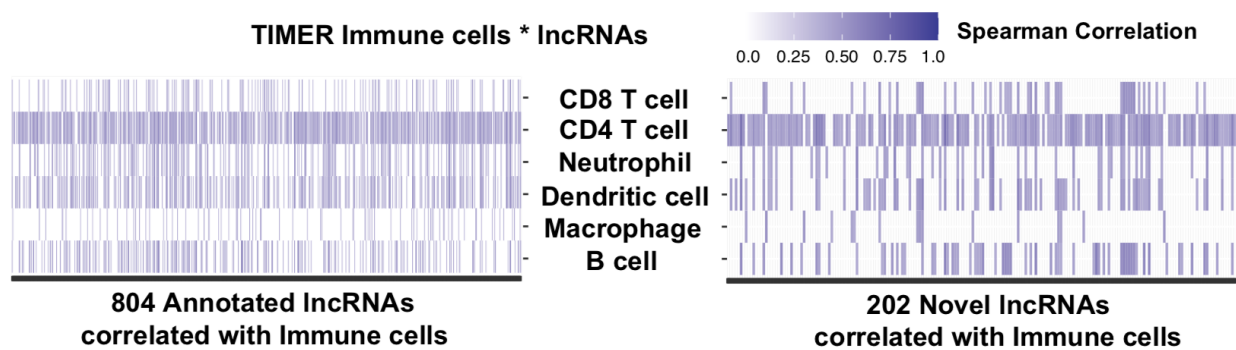
Figure 4.10: Heatmap showing the correlation between immune cells and lncRNAs.

From the previous studies it was seen that protein coding genes that were negatively

correlated with tumor purity, were also positively correlated with immune cells[209]. We

observed a negatively skewed distribution between lncRNAs and tumor purity

suggesting expression of lncRNAs were inversely correlated with tumor purity *(Figure*

4.11). A total of 872 lncRNAs (699 annotated lncRNAs and 173 novel lncRNAs) which

were negatively correlated with tumor purity, were also positively correlated with
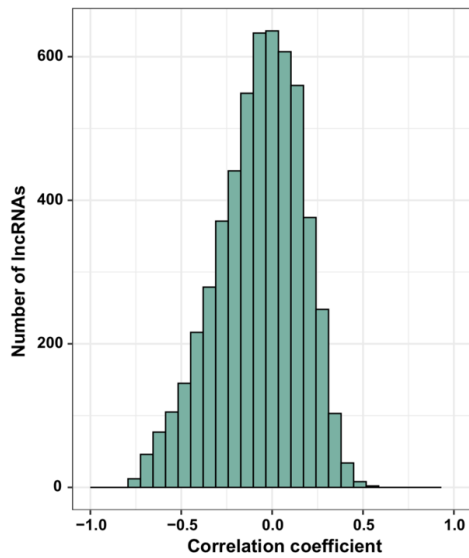
immune cells.

Figure 4.11: Distribution of correlation coefficients between tumor purity and lncRNAs expression.

It has been reported that protein-coding genes associating immune pathways, are likely to be positively correlated with immune cell infiltration and are inversely correlated with tumor purity are immunology related genes and can be used for precision medicine development[209-210]. Similarly, lncRNAs have been found as regulators of immune pathways and might be used as a valuable resource to advance identification of immunotherapy targets[4]. To identify immunology related lncRNAs in TNBCs we overlapped results from all the three steps (lncRNAs - immune pathways, lncRNAs - immune cells and lncRNAs - tumor purity) and found 602 immune lncRNAs which needed a further study *(Figure 4.12, Appendix Table 3)*.

Figure 4.12: Upset plot showing the total number of immune lncRNAs identified from all the three steps.
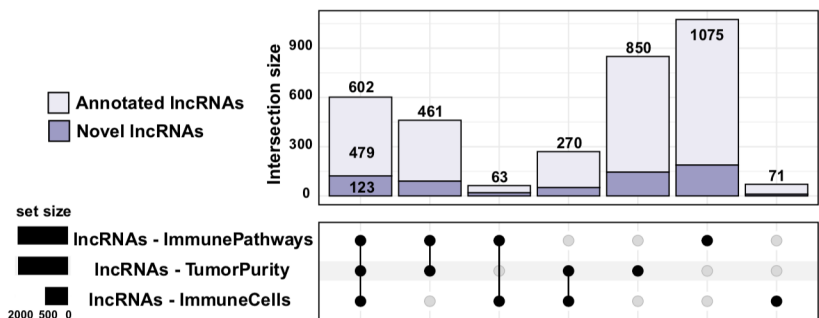
Same procedure was also performed on the TCGA cohort TNBCs (*n=178*) and identified 382 immune lncRNAs *(Figure 4.13)*.
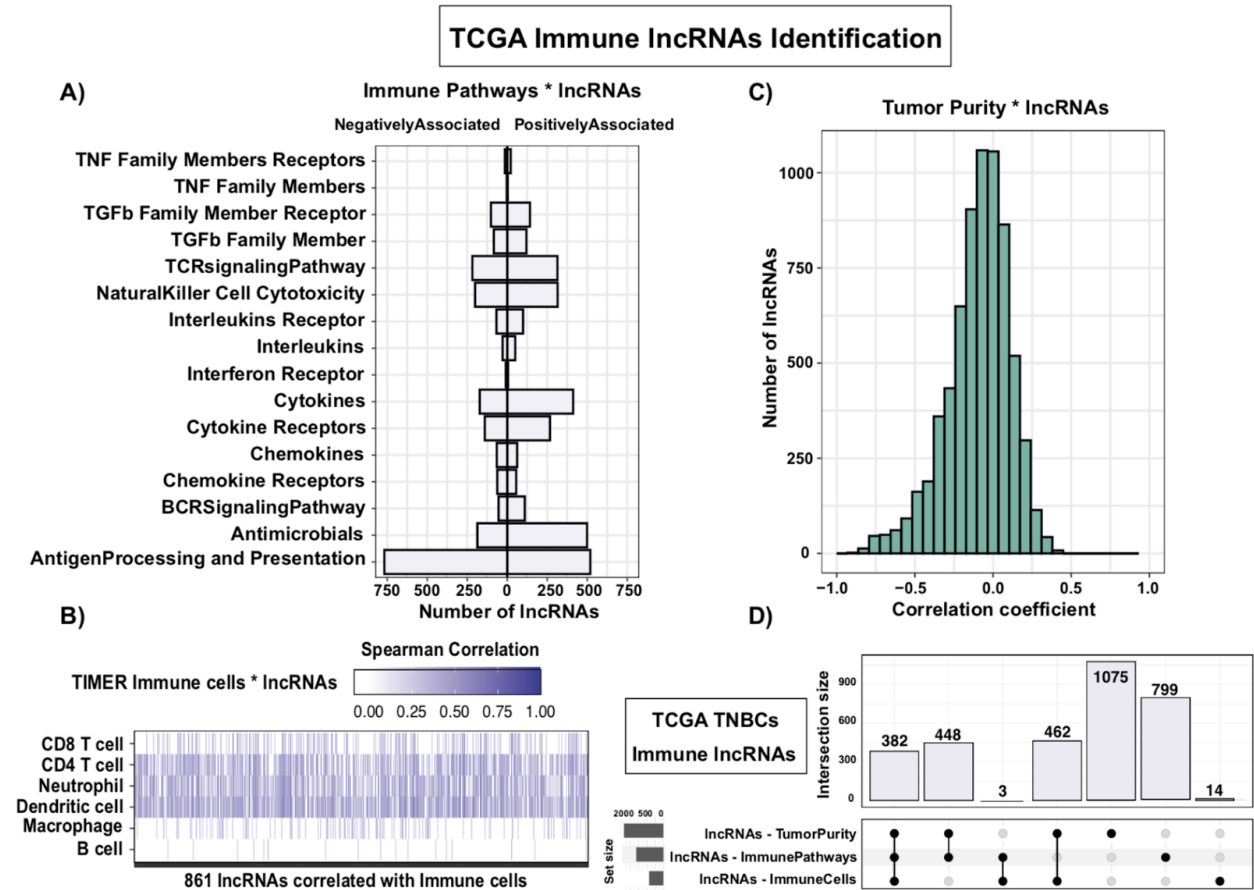
Figure 4.13: Immune lncRNAs identification in TCGA TNBCs A) Bar plot showing the number of lncRNAs identified with immune pathways association. B) Heatmap showing the correlation between immune cells and lncRNAs. C) Distribution of correlation coefficients between tumor purity and lncRNAs expression. D) Upset plot showing the total number of immune lncRNAs identified from all the three steps in TCGA TNBCs.

A significant overlap between the identified immune lncRNAs (*n=479 [FUSCC] and n=382 [TCGA]*) from both the cohorts, showed that TCGA analysis supports the validity of FUSCC analysis (*p-value < 0.001, odds ratio = 5.6*) *(Figure 4.14)*

FUSCC cohort

348   132   250

TCGA cohort

p-value < 0.0001
odds ratio = 5.6
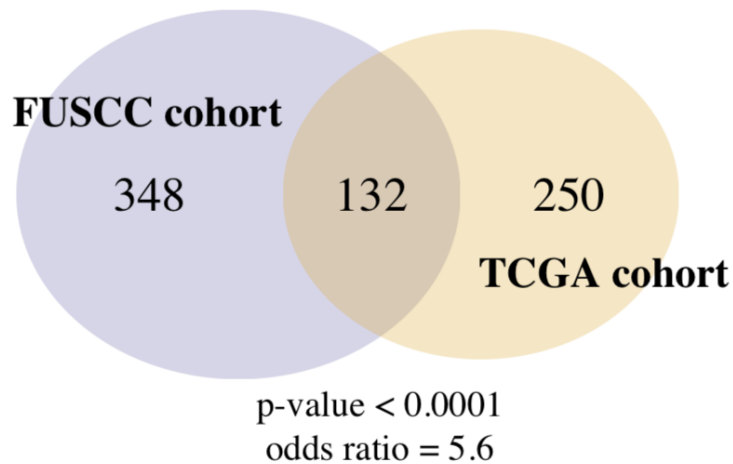
Figure 4.14: Venn diagram with only annotated immune lncRNAs overlap between FUSCC and TCGA cohorts.

To evaluate how the identified immune lncRNAs (*n=602*) enrich in the subtypes of TNBCs of FUSCC cohort, single sample gene set enrichment analysis (ssGSEA) was applied using their expression to calculate immune lncRNA score. We observed that the IM subtype had higher immune lncRNA score compared to other subtypes of TNBCs, followed by MSL, BL2, LAR, BL1 and M (*p-value < 0.001 for all comparisons*) *(Figure 4.15A)*. Among 602 immune lncRNAs, 479 (80%) were annotated lncRNAs and 123 (20%) were novel lncRNAs. Among annotated immune lncRNAs, the most frequent were antisense (35.4%) followed by intergenic (33.5%). The remainder was split between sense overlapping (3.9%), sense intronic (13.7%), processed transcript (3.9%), 3prime overlapping ncRNA (0.2%). Among novel lncRNAs the majority of them were intergenic (55.7%) followed by antisense intronic (29.5%) and antisense exonic (14.8%) *(Figure 4.15B)*.
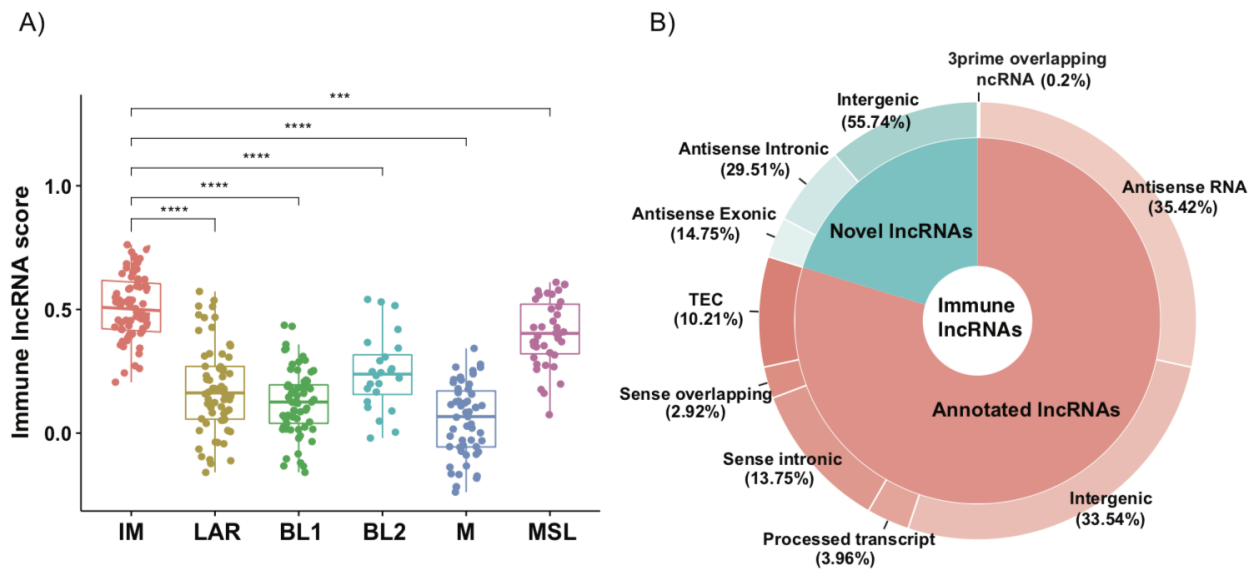
Figure 4.15: *A) Immune lncRNA ssGSEA score between the TNBC subtypes. B) Pie chart showing the gene types of novel and annotated Immune lncRNAs.*

To understand the role of immune lncRNAs, we looked into antisense and sense lncRNAs that overlap corresponding protein coding genes and might regulate their exprression. Spearman correlation was applied between expression of antisense and sense lncRNAs and the corresponding protein coding genes it overlaps. We found the lncRNAs (both annotated and novel) are positively correlated with their corresponding coding genes with an r > 0.5 and p-value < 0.05 *(Figure 4.16A-B)*, but didn't find any negative correlation. lncRNAs highlighted in red color were also found strongly correlated with their corresponding protein coding genes in the TCGA cohort *(Figure 4.16A)*.

Protein-coding genes which were strongly correlated with the above mentioned antisense and sense lncRNAs were enriched with the immune pathways like T-cell activation, B-cell receptor, chemokine signaling pathways *(Figure 4.16C)*.
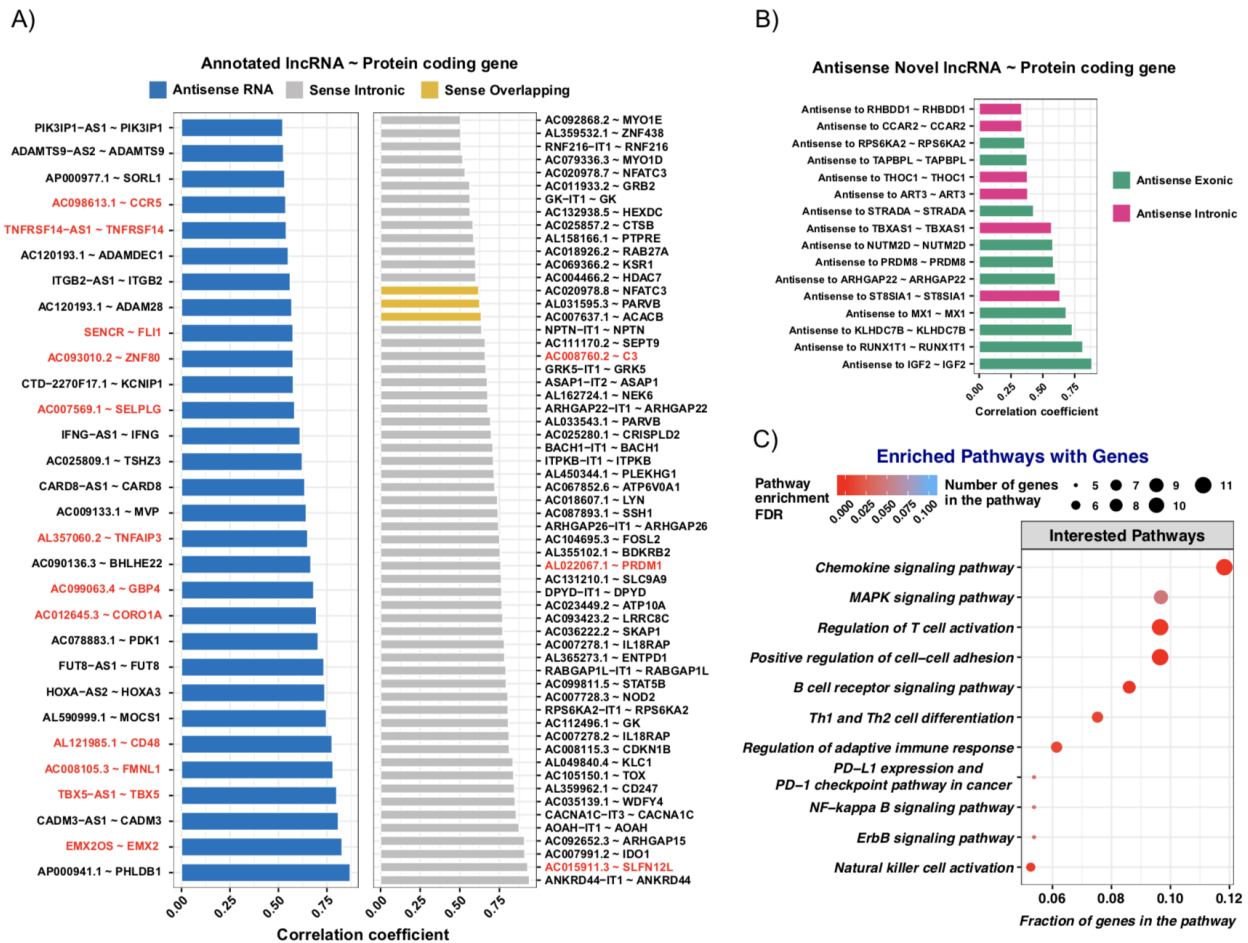
Figure 4.16: A) Bar plot with antisense and sense annotated immune lncRNAs showing positive correlation with their corresponding protein-coding gene it overlaps (Immune lncRNA ~ Protein-coding gene). x-axis represent correlation coefficients and those highlighted in red color on y-axis also showed strong positive correlation with their corresponding protein coding genes in the TCGA cohort. B) Bar plot with antisense novel immune lncRNAs showing positive correlation with their corresponding protein-coding gene it overlaps. C) Pathway enrichment analysis performed on the

protein-coding genes that were strongly correlated with antisense/sense immune lncRNAs. Size and color of dots represent the number of genes involved in the specific pathway and the statistical significance (adjusted p-value), respectively.

For example: *CD48* is an immune cell activation marker involved in a wide variety of immune responses[211] and showed strong correlation with *AL121985.1* (Antisense RNA located on chromosome 1) expression. *CORO1A* disruption in T-cells promotes allograft tolerance[212] and is strongly correlated with *AC012645.3* (Antisense RNA located on chromosome 16) expression. *FLI-1* is reported to play an important role in immunity and autoimmune diseases[213] and is correlated with lncRNA *SENCR* (*FLI-1 AS1*) (Antisense RNA located on chromosome 1) that regulates smooth muscle cell and endothelial cell (EC) phenotypes[214].

4.5 Identification of Immune clusters

It has been suggested that cancer subtyping based on immune profiles may inform which patients might respond to immunotherapy and which may not[210,215]. High immune evasion, low lymphocyte infiltration and low CD8+ might be related to Immune-low cluster, whereas higher CD8+ and higher expression of immune checkpoint inhibitors *PD-1, CTLA-4* might be related to Immune-high cluster and might respond to immunotherapy[210]. Based on this hypothesis, we classified TNBCs into two clusters Immune-High (IH) and Immune-Low (IL) clusters based on the median of immune lncRNA score (ssGSEA score) *(Figure 4.17A)*. Among the TNBC subtypes IM and MSL were significantly associated with the IH group (both p-value < 0.001, odds ratio > 1,

(Figure 4.17A). On the other hand, subtypes M (p-value < 0.001, odds ratio < 1), BL1 (p-value < 0.001, odds ratio < 1) and LAR (p-value < 0.004, odds ratio < 1) were significantly associated with the IL group *(Figure 4.17A)*. We observed that IH patients had a higher distribution of cytolytic activity (CYT), major histocompatibility (MHC) and immune scores compared to IL patients *(Figure 4.17B)*. Tumor inflammation signature (TIS) measure has been shown to be enriched for patients who might respond to anti-PD1 agent pembrolizumab[168].
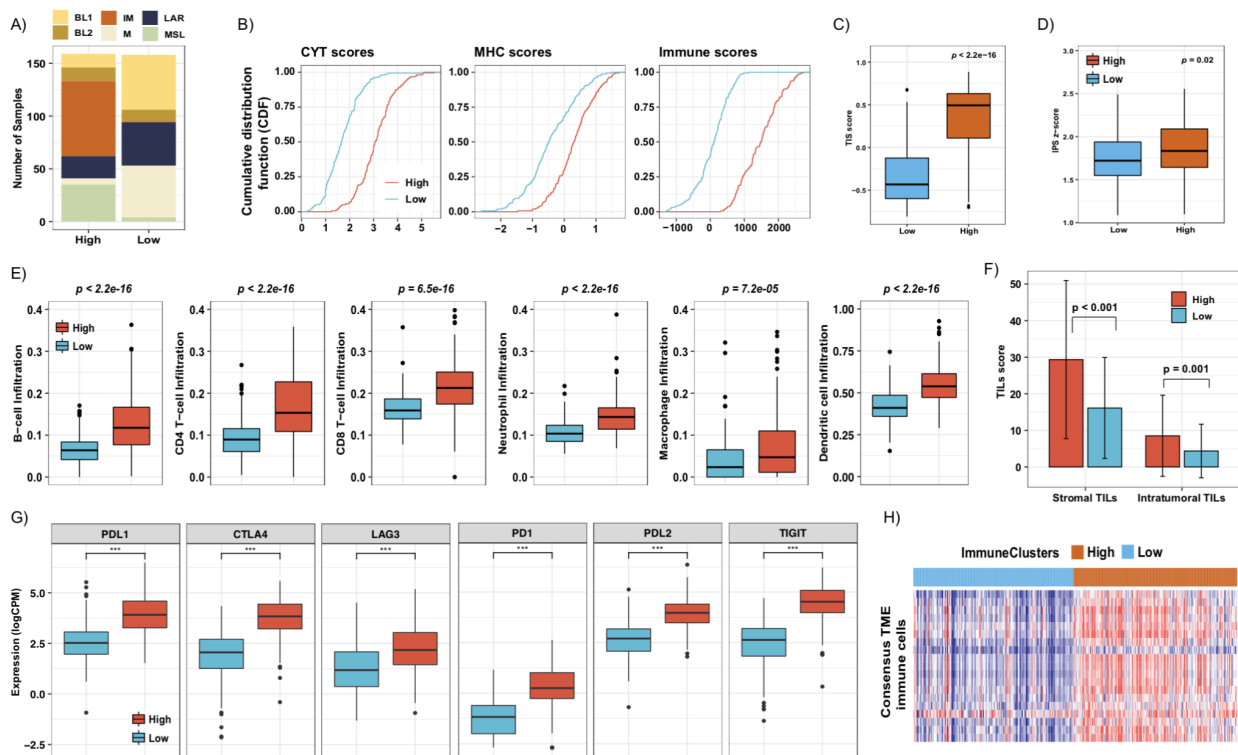
Figure 4.17: Classification of TNBCs into immune clusters A) Association of TNBC subtypes with the identified Immune clusters. B, C and D) Distribution of immunogenic scores between immune clusters. E) TIMER immune cells abundance between the identified immune clusters. F) Average stromal and intratumoral tumor-infiltrating lymphocyte (TIL) score between Immune-High and Immune-Low clusters. G) Immune

checkpoint inhibitors PDL1, CTLA4, LAG3, PD1, PDL2 and TIGIT expression between the immune clusters. H) Consensus TME (Tumor Micro Environment) immune cells abundance between the identified immune clusters.

We observed the TIS score is higher in the IH cluster when compared to the IL cluster *(Figure 4.17C)*. Similarly, immunophenoscore (IPS) was regarded as a predictor of response to anti-CTLA-4 and anti-PD1 antibodies[166]. IPS is higher in IH tumors compared to IL tumors *(Figure 4.17D)*. We also found higher levels of immune cell infiltration (B-cells, CD4 T-cells, CD8 T-cells, Neutrophils, Macrophages, Dendritic cells) in IH than IL clusters ($p < 0.001$) *(Figure 4.17E)*. Higher CD8 T-cell infiltration might suggest that IH cluster patients respond to therapy[216]. Tumor infiltrating lymphocytes (TILs) have been demonstrated as a biomarker for response to immune checkpoint blockade[217]. H&E staining of FUSCC TNBCs predicted higher stromal and intra-tumoral TILs score in IH than IL clusters *(Figure 4.17F)*. Higher expression of several immune checkpoints such as *PD1, CTLA4, PDL1, LAG3 and TIGIT* were observed in IH, compared to IL tumors *(Figure 4.17G)*. Immune cell abundance from Consensus TME was observed higher in the IH cluster compared to the IL cluster *(Figure 4.17H)*. All the above results suggest that Immune-High patients might be immunotherapy responders.

4.6 Signature lncRNAs related to Immune-High

Dysregulated immune lncRNAs among the immune clusters in training dataset were used to identify biomarkers related to Immune-High cluster with the wrapper algorithm *biosigner* around three classifiers (*PLS-DA, Random Forest and SVM*) which iterates

through selection steps that include resampling, feature ranking according to their importance in the model and the identification of significant lncRNAs, building the final model *(Figure 4.18A)*.
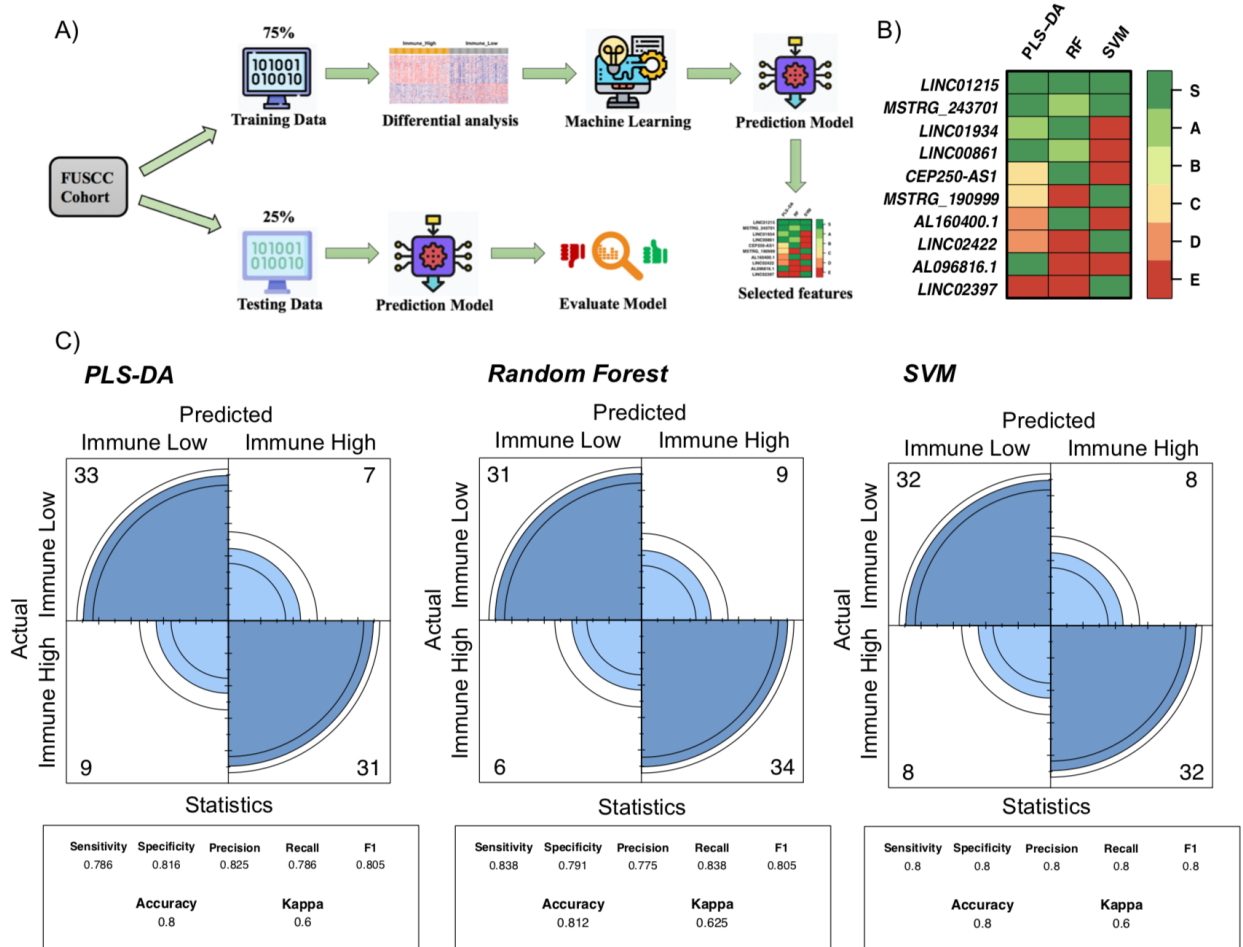
Figure 4.18: A) Computational workflow to identify signature immune lncRNAs. B) Identified signature immune lncRNAs related to Immune-High cluster from the machine learning approaches. C) Confusion matrix plot showing the sensitivity, specificity and accuracy for the three machine learning approaches.

It returned 10 significant signature lncRNAs and the trained model for each of the classifiers. Among them annotated lncRNA *LINC01215* (*"S"* tier - final signature with all three machine learning approaches) and novel lncRNA *MSTRG_243701* (*"S"* tier - final signature with PLS-DA and SVM approaches, whereas *"A"* tier with RF but still can be used for future prediction) which passed through all selection steps were regarded as the most important signature lncRNAs related to the Immune-High cluster. The other 8 lncRNAs (*LINC01934, LINC00861, CEP250-AS1, MSTRG_190999, AL160400.1, LINC02422, AL096816.1, LINC02397*) were detected as *"S"* signature lncRNAs only in a specific single machine learning approach **(Figure 4.18B)**. Finally, computed the performances on testing dataset and observed good prediction values with all three machine learning approaches (*PLS-DA*: sensitivity=0.79, specificity=0.82 and AUC=0.80; *RF*: sensitivity=0.84, specificity=0.80 and AUC=0.81; *SVM*: sensitivity=0.80, specificity=0.80 and AUC=0.80) **(Figure 4.18C)**. These results show a good performance model for categorizing the TNBCs into Immune-High and Immune-Low clusters and also identified the most important significant signature lncRNAs belonging to Immune-High cluster. All identified 10 signature lncRNAs were upregulated in Immune-High compared to Immune-Low **(Table 4.7)**.

gene name	log2FC	padj	chr	start	end	gene type	lncRNA type

Table 4.7: Identified 10 signature lncRNAs related to Immune-High with ML approaches were upregulated in Immune-High when compared with Immune-low

gene name	log2FC	padj	chr	start	end	gene type	lncRNA type
LINC01215	1.809304295	5.48E-15	chr3	108125821	108138610	lincRNA	Annotated lncRNA
MSTRG.243701	1.494521438	9.50E-16	chrX	79368888	79372643	antisense_exonic	Novel lncRNA
LINC01934	1.156149482	4.22E-10	chr2	181101932	181399559	lincRNA	Annotated lncRNA
LINC00861	1.497092708	4.69E-12	chr8	125922308	125951249	lincRNA	Annotated lncRNA
FO393401.1	1.031703841	2.13E-07	chr20	35476203	35490982	antisense_RNA	Annotated lncRNA
MSTRG.190999	0.8868440975	0.0006	chr6	25350576	25450627	antisense_exonic	Novel lncRNA
AL160400.1	1.219057445	1.69E-10	chr6	25014952	25042170	lincRNA	Annotated lncRNA
LINC02422	1.965432687	5.13E-16	chr12	31877079	31887203	lincRNA	Annotated lncRNA
AL096816.1	1.553965182	3.41E-13	chr6	107697299	107700218	TEC	Annotated lncRNA
LINC02397	1.362967126	6.71E-05	chr12	92466451	92492091	lincRNA	Annotated lncRNA

5 Discussion

Many long non-coding RNAs have been suggested to play a role in Triple Negative Breast Cancers (TNBCs), but their association with the subtypes of TNBCs has been very limited. Similarly, research on immune lncRNAs in TNBCs has not been well studied previously. So, here in this study we applied *in-silico* approaches to understand the lncRNAs role in the subtypes of TNBCs and their immune regulation in TNBCs.

To detect novel lncRNAs in TNBCs, we analyzed the transcriptomic data of 317 TNBC tumors and 44 normal samples. With some strict filtering steps and by applying bioinformatic tools *CPC2, RNAFold and Infernal* we identified 5,716 novel lncRNAs. Among these 86.7% were intergenic lncRNAs that do not overlap any of the protein-coding genes, followed by antisense intronic and antisense exonic lncRNAs. Compared with annotated lncRNAs and protein-coding genes, novel lncRNA transcripts had fewer exons, shorter in length and expressed at low levels. These observations are not unexpected given that intergenic regions are less well described and novel lncRNAs likely have features that make them more difficult to detect than previously annotated lncRNAs.

We identified subtype-specific lncRNAs, among which some of them have been studied about their role in different cancer types. So, to identify their functional role in the TNBC subtypes with *in-silico* approaches, initially co-expression network analysis was performed with protein-coding genes, annotated lncRNAs and novel lncRNAs with *WGCNA*[159]. This revealed 20 modules, among which we identified 6 modules with

higher number of lncRNAs co-expressed with protein-coding genes and having a significant positive correlation between modules and subtypes (M1 ~ BL1, M3 ~ BL2, M2 ~ IM, M15 ~ LAR, M16 ~ M, M9 ~ MSL). Apart from these 6 modules, Module M6 also has many lncRNAs (annotated *n=54*, novel *n=1*) co-expressed with coding genes, but there is no correlation between this module and any of the subtypes. We found lncRNAs in the subtype-specific modules that regulate cell cycle, DNA replication, immune pathways, focal adhesion and metabolic processes. For example: *FOXP4-AS1* is a lncRNA upregulated in all the three comparisons (BL1 vs other subtypes, TNBCs vs normals and TNBCs vs non-TNBCs) and based on the co-expression network it might be related to DNA replication and G2/M transition. It has been reported that *FOXP4-AS1* promotes cell proliferation, migration and plays an important role in colorectal cancer progression[186], is involved in cervical cancer progression via regulating miR-136-5p/CBX4 Axis[218]. Another lncRNA *DDX11-AS1,* which is also upregulated in the BL1 subtype and in TNBCs when compared to normals and non-TNBCs, has been studied as a novel oncogene in various cancer types[219]. *LINC00944,* upregulated in IM subtype, has been reported as a prognostic biomarker in breast cancer[220] and is involved in the process of liver metastasis in colorectal cancer[221]. *MIR155HG*, which is also upregulated in the IM subtype, has been regarded as a prognostic biomarker and associated with immune infiltration in different cancer types[203]. Similarly, other lncRNAs which play an important role in the respective subtypes, have been studied as biomarkers in multiple cancers. LncRNAs have been identified as regulators of gene expression in the immune system and have been shown to play a prominent role in cancer progression. Immune lncRNAs have been studied in different

cancers. Some immune lncRNAs have been studied in TNBCs so far[5,6], but there are more to be investigated. To identify immune lncRNAs in TNBCs, we applied *ImmLnc[4]* and detected 602 immune lncRNAs (479 were annotated and 123 were novel lncRNAs) associated with immune pathways, positively correlated with immune cells and negatively correlated with tumor purity. We further looked at the immune lncRNAs which were antisense and sense, as they overlap protein-coding genes and might regulate their expression. lncRNAs that are positively correlated with their corresponding coding genes were seen regulating T-cell activation, B-cell receptor immune pathways. We observed 175 immune lncRNAs (138 annotated and 37 novel lncRNAs) were highly expressed in TNBCs when compared with normals based on log2FC > 1 and FDR < 0.05. Some of the identified immune lncRNAs have been reported in previous studies. For instance: *H19* lncRNA has been reported about its role in promoting macrophage activation and hepatic inflammation in cholestatic liver disease[222]. *RRN3P2* is one of the lncRNAs in an immune lncRNA signature identified to predict the prognosis of bladder cancer[223]. *MIR155HG* lncRNA overexpression in chronic obstructive pulmonary disease patients led to regulation of M1/M2 polarization[224]. Immune lncRNA *MEG3* acts as a competing endogenous RNA sponges *miRNA-17* to regulate *Treg/Th17* levels in asthma patients[139]. There has been a study on immune lncRNAs *HCP5* and *MIAT* which acts as a ceRNA and sponges miRNA *miR-150-5p* which activates the expression of *PDL-1[225]*.

We further classified TNBCs into two immune clusters Immune-High and Immune-Low, based on the median of immune lncRNA score calculated by applying *ssGSEA*. We

observed higher immune cell infiltration, immune scores, TILs score and expression of immune checkpoint inhibitors in Immune-High cluster compared to Immune-Low cluster, which inform about patients in Immune-High cluster might respond to immunotherapy. As a final aim of our study, machine learning approaches were applied and identified 10 signature lncRNAs that are related to Immune-High clusters. Among these, 2 lncRNAs (*LINC01215 and MSTRG_243701*) were regarded as the most important signature lncRNAs related to immune-high cluster, from all the three machine learning approaches. *LINC01215* has been reported among the seven-lncRNA prognostic signatures which might predict prognosis in breast cancer and is associated with immune cells and tumor mutation burden[226]. The other 8 lncRNAs were regarded as signatures in any of the machine learning approaches. In these *LINC01934* has been identified as cancer related lncRNA, regulating immune pathways in multiple cancer types[4]. *LINC00861* was significantly associated with immune checkpoints *PD-1*, *CTLA4* and found as a potential target for immunotherapy to treat prostate cancer[227]. lncRNAs *LINC02422* and *LINC02397* were specific to the immune subtype (samples which are identified for the pathway targetable by combinatorial immunotherapy) when compared to Basal and CIN (chromosomal instability) subtypes identified in HPV neg HNSCCs (head and neck squamous cell carcinomas)[228].

There are few limitations in this study. First, due to the lack of external total RNA-seq datasets of TNBCs, subtype-specific lncRNAs were validated only in the TCGA dataset which was generated by mRNA-seq, which would not have detected many intergenic lncRNAs. Second, our study used *in-silico* methods to identify lncRNAs associated with specific subtypes but lacks experimental validation of lncRNA function in the subtypes

of TNBCs. Third, we have used only RNA-sequencing to understand the lncRNAs in TNBCs. The study could be extended by looking at subtype-specific lncRNA association with alterations in copy number alterations and DNA methylation by integrating multi-omics analysis.

In summary, we detected novel lncRNAs and provided lncRNA association with the subtypes of TNBCs through co-expression network analysis. Our results provide the foundation for the functions of these lncRNAs to be confirmed by *in-vitro* approaches. In this study we also identified lncRNA associated with immune pathways, immune cells and tumor purity which suggest that these lncRNAs may play a role in regulating the immune environment of TNBCs. The subtype-specific lncRNAs and signature immune lncRNAs identified in this study provide valuable insights into lncRNA functions in TNBCs.

6 Conclusions

Although lncRNAs have been investigated in multiple cancers, including triple negative breast cancers, their discovery and characterization remain challenging and many have likely not been discovered due to their molecular features, such as their low expression. Indeed, the identification and characterization of novel lncRNAs, especially the role of lncRNAs in subtypes of TNBCs remain in the early stage. The molecular heterogeneity between TNBC subtypes on the mRNA level is likely associated with a similar degree of heterogeneity in terms of lncRNAs. Our *in-silico* analysis provides a glimpse into the subtype-specific pathways in which lncRNAs may play a role.

Given the recent development of immunotherapy in TNBCs, it is important that we gain a better understanding of the molecular mechanisms that regulate the immune environment of TNBCs. Currently, little is known about how lncRNAs regulate immune response. Our study identified a few lncRNAs that show particularly strong association with a variety of immune features. Future functional experiments would be required to determine if these lncRNAs would indeed regulate the tumor immune environment in TNBCs.

In conclusion, our findings in this study revealed the presence of many novel lncRNAs not previously known, and identified lncRNAs associated with TNBC subtypes and the tumor immune environment.

References

1. Sun, Z. & Zhu, Y. Systematic comparison of RNA-Seq normalization methods using measurement error models. *Bioinformatics* **28**, 2584–2591 (2012).

2. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* vol. 144 646–674 (2011).

3. Katz, H. & Alsharedi, M. Immunotherapy in triple-negative breast cancer. *Med. Oncol.* **35**, 1–9 (2017).

4. Li, Y. *et al.* Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat. Commun.* **11**, 1000 (2020).

5. Li, Y.-X., Wang, S.-M. & Li, C.-Q. Four-lncRNA immune prognostic signature for triple-negative breast cancer Running title: Immune lncRNAs predict prognosis of TNBC. *Math. Biosci. Eng.* **18**, 3939–3956 (2021).

6. Hu, Q. *et al.* Oncogenic lncRNA downregulates cancer cell antigen presentation and intrinsic tumor suppression. *Nat. Immunol.* **20**, 835–851 (2019).

7. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* (2021) doi:10.3322/caac.21660.

8. Dai, X., Li, Y., Bai, Z. & Tang, X.-Q. Molecular portraits revealing the heterogeneity of breast tumor subtypes defined using immunohistochemistry markers. *Sci. Rep.* **5**, 14499 (2015).

9. Cho, N. Molecular subtypes and imaging phenotypes of breast cancer. *Ultrasonography* vol. 35 281–288 (2016).

10. Uscanga-Perales, G. I., Santuario-Facio, S. K. & Ortiz-López, R. Triple negative

breast cancer: Deciphering the biology and heterogeneity. *Medicina Universitaria* **18**, 105–114 (2016).

11. Bareche, Y. *et al.* Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann. Oncol.* **29**, 895–902 (2018).

12. Phipps, A. I. *et al.* Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. *J. Natl. Cancer Inst.* **103**, 470–477 (2011).

13. Yu, L.-Y. *et al.* New Immunotherapy Strategies in Breast Cancer. *Int. J. Environ. Res. Public Health* **14**, (2017).

14. Chen, X. *et al.* TNBCtype: A Subtyping Tool for Triple-Negative Breast Cancer. *Cancer Inform.* **11**, 147 (2012).

15. Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).

16. Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin. Cancer Res.* **21**, 1688–1698 (2015).

17. Liu, Y.-R. *et al.* Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast Cancer Res.* **18**, (2016).

18. Yin, L., Duan, J.-J., Bian, X.-W. & Yu, S.-C. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* **22**, 1–13 (2020).

19. Gerratana, L. *et al.* Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. *Cancer Treat. Rev.* **68**, 102–110 (2018).

20. Li, Z., Zhao, W., Wang, M. & Zhou, X. Gene Expression Profiling in Cancer. in *Gene Expression Profiling in Cancer* (IntechOpen, 2019).

21. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* **9**, e1003569 (2013).

22. Qu, Z. & Adelson, D. L. Evolutionary conservation and functional roles of ncRNA. *Frontiers in Genetics* vol. 3 (2012).

23. Li, Z., Zhao, W., Wang, M. & Zhou, X. The Role of Long Noncoding RNAs in Gene Expression Regulation. in *Gene Expression Profiling in Cancer* (IntechOpen, 2019).

24. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol.* **9**, 402 (2018).

25. MicroRNA. https://www.sciencedirect.com/topics/neuroscience/microrna.

26. Yu, C.-Y. & Kuo, H.-C. The emerging roles and functions of circular RNAs and their generation. *J. Biomed. Sci.* **26**, 1–12 (2019).

27. Dana, H. *et al.* Molecular Mechanisms and Biological Functions of siRNA. *Int. J. Biomed. Sci.* **13**, 48 (2017).

28. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: past, present, and future. *Genetics* **193**, 651–669 (2013).

29. Evolution and Functions of Long Noncoding RNAs. *Cell* **136**, 629–641 (2009).

30. Guo, C.-J. *et al.* Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* **181**, 621–636.e22 (2020).

31. Quinn, J. J. *et al.* Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes Dev.* **30**, 191–207 (2016).

32. Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110–1122 (2015).

33. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).

34. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

35. Pertea, M. *et al.* CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **19**, 1–14 (2018).

36. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).

37. Abugessaisa, I. *et al.* FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Res.* **49**, D892–D898 (2021).

38. Volders, P.-J. *et al.* LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* **47**, D135–D139 (2019).

39. Fang, S. *et al.* NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* **46**, D308–D314 (2018).

40. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics* vol. 47 199–208 (2015).

41. You, B.-H., Yoon, S.-H. & Nam, J.-W. High-confidence coding and noncoding transcriptome maps. *Genome Res.* **27**, 1050 (2017).

42. Seifuddin, F. *et al.* lncRNAKB, a knowledgebase of tissue-specific functional annotation and trait association of long noncoding RNA. *Sci Data* **7**, 326 (2020).

43. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253–1261 (2015).

44. Di Mauro, S. *et al.* Circulating Coding and Long Non-Coding RNAs as Potential Biomarkers of Idiopathic Pulmonary Fibrosis. *Int. J. Mol. Sci.* **21**, (2020).

45. Chen, J., Liu, S. & Hu, X. Long non-coding RNAs: crucial regulators of gastrointestinal cancer cell proliferation. *Cell Death Discov* **4**, 50 (2018).

46. Ghafouri-Fard, S., Shoorei, H., Anamag, F. T. & Taheri, M. The Role of Non-Coding RNAs in Controlling Cell Cycle Related Proteins in Cancer Cells. *Front. Oncol.* **10**, (2020).

47. Zhang, L. *et al.* LncRNA WTAPP1 promotes cancer cell invasion and migration in NSCLC by downregulating lncRNA HAND2-AS1. *BMC Pulm. Med.* **20**, (2020).

48. Chen, J., Wang, Y., Wang, C., Hu, J.-F. & Li, W. LncRNA Functions as a New Emerging Epigenetic Factor in Determining the Fate of Stem Cells. *Front. Genet.* **11**, (2020).

49. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **10**, 924 (2013).

50. Zhang, X. *et al.* Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder. *Oncotarget* **7**, 66069–66076 (2016).

51. Xu, D. *et al.* An optimal prognostic model based on gene expression for clear cell renal cell carcinoma. *Oncol. Lett.* **20**, 2420–2434 (2020).

52. Zhang, Y. *et al.* Identification of Potential Prognostic Long Non-Coding RNA

Biomarkers for Predicting Recurrence in Patients with Cervical Cancer. *Cancer Manag. Res.* **12**, 719–730 (2020).

53. Long non-coding RNA HCP5 in cancer. *Clin. Chim. Acta* **512**, 33–39 (2021).

54. Downregulation of long noncoding RNA HCP5 contributes to cisplatin resistance in human triple-negative breast cancer via regulation of PTEN expression. *Biomed. Pharmacother.* **115**, 108869 (2019).

55. Liu, Y. *et al.* A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.* **4**, e1000041 (2008).

56. Liu, M. *et al.* LncRNA ITGB2-AS1 Could Promote the Migration and Invasion of Breast Cancer Cells through Up-Regulating ITGB2. *Int. J. Mol. Sci.* **19**, (2018).

57. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).

58. Jarroux, J., Morillon, A. & Pinskaya, M. History, Discovery, and Classification of lncRNAs. in *Long Non Coding RNA Biology* 1–46 (Springer, Singapore, 2017).

59. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629 (2011).

60. Gu, Y. *et al.* TYMSOS drives the proliferation, migration, and invasion of gastric cancer cells by regulating ZNF703 via sponging miR-4739. *Cell Biol. Int.* (2021) doi:10.1002/cbin.11610.

61. Zhang, W., Guan, X. & Tang, J. The long non–coding RNA landscape in triple–negative breast cancer. *Cell Proliferation* vol. 54 (2021).

62. Balas, M. M. & Johnson, A. M. Exploring the mechanisms behind long noncoding RNAs and cancer. *Noncoding RNA Res* **3**, 108–117 (2018).

63. Ransohoff, J. D., Wei, Y. & Khavari, P. A. The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol.* **19**, 143 (2018).

64. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* vol. 458 223–227 (2009).

65. Jin, C., Yan, B., Lu, Q., Lin, Y. & Ma, L. Reciprocal regulation of Hsa-miR-1 and long noncoding RNA MALAT1 promotes triple-negative breast cancer development. *Tumour Biol.* **37**, 7383–7394 (2016).

66. Lin, A. *et al.* The LINK-A lncRNA interacts with PtdIns(3,4,5)P3 to hyperactivate AKT and confer resistance to AKT inhibitors. *Nature Cell Biology* vol. 19 238–251 (2017).

67. Tseng, Y.-Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512**, 82–86 (2014).

68. Gudenas, B. L. & Wang, L. Prediction of LncRNA Subcellular Localization with Deep Learning from Sequence Features. *Sci. Rep.* **8**, 16385 (2018).

69. Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Pauler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC Biology* vol. 11 59 (2013).

70. Vieira, A. S., Dogini, D. B. & Lopes-Cendes, I. Role of non-coding RNAs in non-aging-related neurological disorders. *Braz. J. Med. Biol. Res.* **51**, (2018).

71. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse H19 gene. *Nature* vol. 351 153–155 (1991).

72. Burgess, D. J. Non-coding RNA: HOTTIP goes the distance. *Nature reviews. Genetics* vol. 12 300 (2011).

73. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is

expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).

74. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).

75. Maenner, S. *et al.* 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* **8**, e1000276 (2010).

76. Sun, M. & Kraus, W. L. From Discovery to Function: The Expanding Roles of Long NonCoding RNAs in Physiology and Disease. *Endocr. Rev.* **36**, 25–64 (2015).

77. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).

78. Karlsson, O. & Baccarelli, A. A. Environmental Health and Long Non-coding RNAs. *Current Environmental Health Reports* vol. 3 178–187 (2016).

79. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim. Biophys. Acta* **1842**, 1910–1922 (2014).

80. Srijyothi, L., Ponne, S., Prathama, T., Ashok, C. & Baluchamy, S. Roles of Non-Coding RNAs in Transcriptional Regulation. in *Transcriptional and Post-transcriptional Regulation* (IntechOpen, 2018).

81. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8 (2010).

82. Kevin C. Wang, H. Y. C. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904 (2011).

83. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with

chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11667–11672 (2009).

84. Kotake, Y. *et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956–1962 (2011).

85. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).

86. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117 (2010).

87. Wang, L. *et al.* Long non-coding RNA (LncRNA) RMST in triple-negative breast cancer (TNBC): Expression analysis and biological roles research. *J. Cell. Physiol.* **233**, 6603–6612 (2018).

88. Mou, E. & Wang, H. LncRNA LUCAT1 facilitates tumorigenesis and metastasis of triple-negative breast cancer through modulating miR-5702. *Biosci. Rep.* **39**, (2019).

89. Wang, S. *et al.* LncRNA MIR100HG promotes cell proliferation in triple-negative breast cancer through triplex formation with p27 loci. *Cell Death Dis.* **9**, 805 (2018).

90. Wang, X. *et al.* Long noncoding RNA Linc00339 promotes triple-negative breast cancer progression through miR-377-3p/HOXC6 signaling pathway. *J. Cell. Physiol.* **234**, 13303–13317 (2019).

91. Shin, V. Y. *et al.* Long non-coding RNA NEAT1 confers oncogenic role in triple-negative breast cancer through modulating chemoresistance and cancer

stemness. *Cell Death Dis.* **10**, 270 (2019).

92. Li, S. *et al.* Long noncoding RNA GAS5 suppresses triple negative breast cancer progression through inhibition of proliferation and invasion by competitively binding miR-196a-5p. *Biomed. Pharmacother.* **104**, 451–457 (2018).

93. Pickard, M. R. & Williams, G. T. Regulation of apoptosis by long non-coding RNA GAS5 in breast cancer cells: implications for chemotherapy. *Breast Cancer Res. Treat.* **145**, 359–370 (2014).

94. Tian, Y., Xia, S., Ma, M. & Zuo, Y. LINC00096 Promotes the Proliferation and Invasion by Sponging miR-383-5p and Regulating RBM3 Expression in Triple-Negative Breast Cancer. *Onco. Targets. Ther.* **12**, 10569–10578 (2019).

95. Kong, Y., Geng, C. & Dong, Q. LncRNA PAPAS may promote triple-negative breast cancer by downregulating miR-34a. *J. Int. Med. Res.* **47**, 3709–3718 (2019).

96. Tariq, A. *et al.* LncRNA-mediated regulation of SOX9 expression in basal subtype breast cancer cells. *RNA* vol. 26 175–185 (2020).

97. Tang, J. *et al.* LncRNA DANCR upregulates PI3K/AKT signaling through activating serine phosphorylation of RXRA. *Cell Death Dis.* **9**, 1167 (2018).

98. Jadaliha, M. *et al.* A natural antisense lncRNA controls breast cancer progression by promoting tumor suppressor gene mRNA stability. *PLoS Genet.* **14**, e1007802 (2018).

99. Yu, F., Wang, L. & Zhang, B. Long non-coding RNA DRHC inhibits the proliferation of cancer cells in triple negative breast cancer by downregulating long non-coding RNA HOTAIR. *Oncol. Lett.* **18**, 3817–3822 (2019).

100. Wang, N., Hou, M., Zhan, Y. & Sheng, X. LncRNA PTCSC3 inhibits triple-negative

breast cancer cell proliferation by downregulating lncRNA H19. *J. Cell. Biochem.* **120**, (2019).

101. Niu, L., Fan, Q., Yan, M. & Wang, L. LncRNA NRON down-regulates lncRNA snaR and inhibits cancer cell proliferation in TNBC. *Biosci. Rep.* **39**, (2019).

102. Wang, K., Li, X., Song, C. & Li, M. LncRNA AWPPH promotes the growth of triple-negative breast cancer by up-regulating frizzled homolog 7 (FZD7). *Biosci. Rep.* **38**, (2018).

103. Xu, S.-T. *et al.* Long non-coding RNA ANRIL promotes carcinogenesis via sponging miR-199a in triple-negative breast cancer. *Biomed. Pharmacother.* **96**, 14–21 (2017).

104. Zuo, Y., Li, Y., Zhou, Z., Ma, M. & Fu, K. Long non-coding RNA MALAT1 promotes proliferation and invasion via targeting miR-129-5p in triple-negative breast cancer. *Biomed. Pharmacother.* **95**, 922–928 (2017).

105. Wang, G.-P. *et al.* LINC01096 knockdown inhibits progression of triple-negative breast cancer by increasing miR-3130-3p. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 7445–7456 (2019).

106. Jin, X. *et al.* The endogenous retrovirus-derived long noncoding RNA TROJAN promotes triple-negative breast cancer progression via ZMYND8 degradation. *Sci Adv* **5**, eaat9820 (2019).

107. Shen, X., Zhong, J., Yu, P., Zhao, Q. & Huang, T. YY1-regulated LINC00152 promotes triple negative breast cancer progression by affecting on stability of PTEN protein. *Biochem. Biophys. Res. Commun.* **509**, 448–454 (2019).

108. Wang, O. *et al.* C-MYC-induced upregulation of lncRNA SNHG12 regulates cell

proliferation, apoptosis and migration in triple-negative breast cancer. *Am. J. Transl. Res.* **9**, 533–545 (2017).

109. Zhang, X. *et al.* lncRNA AFAP1-AS1 promotes triple negative breast cancer cell proliferation and invasion via targeting miR-145 to regulate MTH1 expression. *Sci. Rep.* **10**, 7662 (2020).

110. Yao, Y., Chu, Y., Xu, B., Hu, Q. & Song, Q. Risk factors for distant metastasis of patients with primary triple-negative breast cancer. *Biosci. Rep.* **39**, (2019).

111. Zhang, H. *et al.* Correction: Epigenetic Regulation of NAMPT by NAMPT-AS Drives Metastatic Progression in Triple-Negative Breast Cancer. *Cancer Res.* **81**, 3145 (2021).

112. Han, C., Li, X., Fan, Q., Liu, G. & Yin, J. CCAT1 promotes triple-negative breast cancer progression by suppressing miR-218/ZFX signaling. *Aging* **11**, 4858–4875 (2019).

113. Fu, J. *et al.* LncRNA MIR503HG inhibits cell migration and invasion via miR–103/OLFM4 axis in triple negative breast cancer. *Journal of Cellular and Molecular Medicine* vol. 23 4738–4745 (2019).

114. Zhang, G. *et al.* Long non-coding RNA ZEB2-AS1 promotes the proliferation, metastasis and epithelial mesenchymal transition in triple-negative breast cancer by epigenetically activating ZEB2. *J. Cell. Mol. Med.* **23**, 3271–3279 (2019).

115. Yang, F. *et al.* An androgen receptor negatively induced long non-coding RNA ARNILA binding to miR-204 promotes the invasion and metastasis of triple-negative breast cancer. *Cell Death Differ.* **25**, 2209–2220 (2018).

116. Zhang, Y. *et al.* Long noncoding RNA LINP1 regulates repair of DNA double-strand

breaks in triple-negative breast cancer. *Nat. Struct. Mol. Biol.* **23**, 522–530 (2016).

117. Liu, A.-N. *et al.* LncRNA AWPPH and miRNA-21 regulates cancer cell proliferation and chemosensitivity in triple-negative breast cancer by interacting with each other. *J. Cell. Biochem.* **120**, 14860–14866 (2019).

118. Candeias, S. M. & Gaipl, U. S. The Immune System in Cancer Prevention, Development and Therapy. *Anti-Cancer Agents in Medicinal Chemistry* vol. 16 101–107 (2015).

119. Terremovic, D. & Vile, R. G. The Immune System in Cancer: If It Isn't Broken, Can We Fix It? *Cancer Immune Therapy* 204–229 (2002) doi:10.1002/3527600795.ch10.

120. Doss, M. Cancer Prevention and Treatment with Immune System Boosting Interventions. doi:10.20944/preprints201912.0269.v1.

121. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).

122. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).

123. Long non-coding RNA: An immune cells perspective. *Life Sci.* **271**, 119152 (2021).

124. Kotzin, J. J. *et al.* The long noncoding RNA Morrbid regulates CD8 T cells in response to viral infection. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11916–11925 (2019).

125. Yan, K. *et al.* Repression of lncRNA NEAT1 enhances the antitumor activity of CD8 T cells against hepatocellular carcinoma via regulating miR-155/Tim-3. *The International Journal of Biochemistry & Cell Biology* vol. 110 1–8 (2019).

126. Ji, J. *et al.* Long non-coding RNA Lnc-Tim3 exacerbates CD8 T cell exhaustion via binding to Tim-3 and inducing nuclear translocation of Bat3 in HCC. *Cell Death Dis.* **9**, 478 (2018).

127. Obaid, M. *et al.* LncRNA HOTAIR regulates lipopolysaccharide-induced cytokine expression and inflammatory response in macrophages. *Sci. Rep.* **8**, 15670 (2018).

128. Xie, C., Guo, Y. & Lou, S. LncRNA ANCR Promotes Invasion and Migration of Gastric Cancer by Regulating FoxO1 Expression to Inhibit Macrophage M1 Polarization. *Dig. Dis. Sci.* **65**, 2863–2872 (2020).

129. Zhang, Y. *et al.* Long non-coding RNA AK085865 ablation confers susceptibility to viral myocarditis by regulating macrophage polarization. *J. Cell. Mol. Med.* **24**, 5542–5554 (2020).

130. Zhang, M. *et al.* Knockdown of NEAT1 induces tolerogenic phenotype in dendritic cells by inhibiting activation of NLRP3 inflammasome. *Theranostics* **9**, 3425–3442 (2019).

131. Zhuang, L., Tian, J., Zhang, X., Wang, H. & Huang, C. Lnc-DC regulates cellular turnover and the HBV-induced immune response by TLR9/STAT3 signaling in dendritic cells. *Cell. Mol. Biol. Lett.* **23**, 43 (2018).

132. Cells and Organs of the Immune System. *Immunology for Pharmacy* 1–14 (2012) doi:10.1016/b978-0-323-06947-2.10001-x.

133. Jiang, N. *et al.* Identification of key protein-coding genes and lncRNAs in spontaneous neutrophil apoptosis. *Sci. Rep.* **9**, 15106 (2019).

134. Shang, A. *et al.* Long non-coding RNA HOTTIP enhances IL-6 expression to potentiate immune escape of ovarian cancer cells by upregulating the expression of

PD-L1 in neutrophils. *J. Exp. Clin. Cancer Res.* **38**, 411 (2019).

135. Luckheeram, R. V., Zhou, R., Verma, A. D. & Xia, B. CD4+T Cells: Differentiation and Functions. *Clin. Dev. Immunol.* **2012**, (2012).

136. Swain, S. L., McKinstry, K. K. & Strutt, T. M. Expanding roles for CD4⁺ T cells in immunity to viruses. *Nat. Rev. Immunol.* **12**, 136–148 (2012).

137. Gao, X., Liu, L., Min, X., Jia, S. & Zhao, M. Non-Coding RNAs in CD4 T Cells: New Insights Into the Pathogenesis of Systemic Lupus Erythematosus. *Front. Immunol.* **11**, 568 (2020).

138. West, K. A. & Lagos, D. Long Non-Coding RNA Function in CD4 T Cells: What We Know and What Next? *Noncoding RNA* **5**, (2019).

139. Qiu, Y.-Y. *et al.* LncRNA-MEG3 functions as a competing endogenous RNA to regulate Treg/Th17 balance in patients with asthma by targeting microRNA-17/ RORγt. *Biomed. Pharmacother.* **111**, 386–394 (2019).

140. Zhang, F., Liu, G., Wei, C., Gao, C. & Hao, J. Linc-MAF-4 regulates Th1/Th2 differentiation and is associated with the pathogenesis of multiple sclerosis by targeting MAF. *FASEB J.* **31**, 519–525 (2017).

141. LeBien, T. W. & Tedder, T. F. B lymphocytes: how they develop and function. *Blood* **112**, 1570–1580 (2008).

142. Hu, W. Faculty Opinions recommendation of Long Noncoding RNA Expression during Human B-Cell Development. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature* (2016) doi:10.3410/f.725803718.793520919.

143. Winkle, M., Kluiver, J. L., Diepstra, A. & van den Berg, A. Emerging roles for long noncoding RNAs in B-cell development and malignancy. *Crit. Rev. Oncol. Hematol.*

120, 77–85 (2017).

144. Jiang, Y.-Z. *et al.* Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell* **35**, 428–440.e5 (2019).

145. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

146. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

147. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).

148. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

149. Vera Alvarez, R., Pongor, L. S., Mariño-Ramírez, L. & Landsman, D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**, 1960–1962 (2019).

150. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

151. Geo Pertea, M. P. GFF Utilities: GffRead and GffCompare. *F1000Res.* **9**, (2020).

152. Triff, K. *et al.* Assessment of histone tail modifications and transcriptional profiling during colon cancer progression reveals a global decrease in H3K4me3 activity. *Biochim. Biophys. Acta Mol. Basis Dis.* **1863**, 1392–1402 (2017).

153. Kang, Y.-J. *et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).

154. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).

155. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).

156. Eric P. Nawrocki, S. R. E. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933 (2013).

157. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* **31**, 439 (2003).

158. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

159. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* vol. 9 (2008).

160. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* vol. 16 284–287 (2012).

161. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).

162. Bhattacharya, S. *et al.* ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data* **5**, 180015 (2018).

163. Li, T. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020).

164. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).

165. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

166. Charoentong, P. *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* **18**, 248–262 (2017).

167. Damotte, D. *et al.* The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *J. Transl. Med.* **17**, 1–10 (2019).

168. Danaher, P. *et al.* Pan-cancer adaptive immune resistance as defined by the Tumor Inflammation Signature (TIS): results from The Cancer Genome Atlas (TCGA). *J Immunother Cancer* **6**, 63 (2018).

169. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

170. Lauss, M. *et al.* Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat. Commun.* **8**, (2017).

171. Jiménez-Sánchez, A., Cast, O. & Miller, M. L. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res.* **79**, 6238–6246 (2019).

172. Rinaudo, P., Boudah, S., Junot, C. & Thévenot, E. A. biosigner: A New Method for the Discovery of Significant Molecular Signatures from Omics Data. *Front. Mol.*

Biosci. **3**, (2016).

173. Zhou, M. *et al.* Recurrence-Associated Long Non-coding RNA Signature for Determining the Risk of Recurrence in Patients with Colon Cancer. *Mol. Ther. Nucleic Acids* **12**, 518–529 (2018).

174. Gao, T. & Ji, Y. Long Noncoding RNA LINC00707 Accelerates Tumorigenesis and Progression of Bladder Cancer via Targeting miR-145/CDCA3 Regulatory Loop. *Urol. Int.* **105**, 891–905 (2021).

175. Shao, H.-J., Li, Q., Shi, T., Zhang, G.-Z. & Shao, F. LINC00707 promotes cell proliferation and invasion of colorectal cancer via miR-206/FMNL2 axis. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 3749–3759 (2019).

176. Shi, J. *et al.* LncRNA LINP1 regulates acute myeloid leukemia progression via HNF4α/AMPK/WNT5A signaling pathway. *Hematol. Oncol.* **37**, 474–482 (2019).

177. Flockhart, R. J. *et al.* BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome Res.* **22**, 1006–1014 (2012).

178. Chen, J.-X., Chen, M., Zheng, Y., Wang, S.-Y. & Shen, Z.-P. Up-regulation of BRAF activated non-coding RNA is associated with radiation therapy for lung cancer. *Biomed. Pharmacother.* **71**, 79–83 (2015).

179. Li, L., Zhang, L., Zhang, Y. & Zhou, F. Increased expression of LncRNA BANCR is associated with clinical progression and poor prognosis in gastric cancer. *Biomed. Pharmacother.* **72**, 109–112 (2015).

180. Guo, Q. *et al.* BRAF-activated long non-coding RNA contributes to colorectal cancer migration by inducing epithelial-mesenchymal transition. *Oncol. Lett.* **8**,

869–875 (2014).

181. Su, J. *et al.* Long noncoding RNA BLACAT1 indicates a poor prognosis of colorectal cancer and affects cell proliferation by epigenetically silencing of p15. *Cell Death Dis.* **8**, e2665 (2017).

182. Xu, M. *et al.* Landscape analysis of lncRNAs shows that DDX11-AS1 promotes cell-cycle progression in liver cancer through the PARP1/p53 axis. *Cancer Lett.* **520**, 282–294 (2021).

183. Feng, X., Yang, S., Zhou, S., Deng, S. & Xie, Y. Long non-coding RNA DDX11-AS1 promotes non-small cell lung cancer development via regulating PI3K/AKT signalling. *Clin. Exp. Pharmacol. Physiol.* **47**, 1622–1631 (2020).

184. Ren, Z., Liu, X., Si, Y. & Yang, D. Long non-coding RNA DDX11-AS1 facilitates gastric cancer progression by regulating miR-873-5p/SPC18 axis. *Artificial Cells, Nanomedicine, and Biotechnology* vol. 48 572–583 (2020).

185. Tian, J.-B., Cao, L. & Dong, G.-L. Long noncoding RNA DDX11-AS1 induced by YY1 accelerates colorectal cancer progression through targeting miR-873/CLDN7 axis. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 5714–5729 (2019).

186. Li, J. *et al.* Long non-coding RNA FOXP4-AS1 is an unfavourable prognostic factor and regulates proliferation and apoptosis in colorectal cancer. *Cell Prolif.* **50**, (2017).

187. Wu, X., Xiao, Y., Zhou, Y., Zhou, Z. & Yan, W. LncRNA FOXP4-AS1 is activated by PAX5 and promotes the growth of prostate cancer by sequestering miR-3184-5p to upregulate FOXP4. *Cell Death Dis.* **10**, 472 (2019).

188. Xie, W. *et al.* lncRNA MEG8 is downregulated in osteoarthritis and regulates

chondrocyte cell proliferation, apoptosis and inflammation. *Exp. Ther. Med.* **22**, 1153 (2021).

189. Sang, Y. *et al.* Up-regulation of long non-coding HOTTIP functions as an oncogene by regulating HOXA13 in non-small cell lung cancer. *Am. J. Transl. Res.* **8**, 2022–2032 (2016).

190. Quagliata, L. *et al.* Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology* **59**, 911–923 (2014).

191. Liu, T. *et al.* The Long Non-coding RNA HOTTIP Is Highly Expressed in Colorectal Cancer and Enhances Cell Proliferation and Invasion. *Molecular Therapy - Nucleic Acids* vol. 19 612–618 (2020).

192. Zhang, S. *et al.* Upregulation of a novel lncRNA LINC01980 promotes tumor growth of esophageal squamous cell carcinoma. *Biochem. Biophys. Res. Commun.* **513**, 73–80 (2019).

193. Luan, W. *et al.* Long non-coding RNA LINC00520 promotes the proliferation and metastasis of malignant melanoma by inducing the miR-125b-5p/EIF5A2 axis. *J. Exp. Clin. Cancer Res.* **39**, 96 (2020).

194. Xu, T.-P. *et al.* Decreased expression of the long non-coding RNA FENDRR is associated with poor prognosis in gastric cancer and FENDRR regulates gastric cancer cell metastasis by affecting fibronectin1 expression. *Journal of Hematology & Oncology* vol. 7 (2014).

195. Zhao, J. & Liu, H.-R. Down-regulation of long noncoding RNA DLX6-AS1 defines good prognosis and inhibits proliferation and metastasis in human epithelial ovarian

cancer cells via Notch signaling pathway. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 3243–3252 (2019).

196. Chen, J., Gao, C. & Zhu, W. Long non-coding RNA SLC25A25-AS1 exhibits oncogenic roles in non-small cell lung cancer by regulating the microRNA-195-5p/ITGA2 axis. *Oncology Letters* vol. 22 (2021).

197. Lu, Y. *et al.* lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/β-catenin signaling. *Nat. Med.* **23**, 1331–1341 (2017).

198. Li, X.-G. *et al.* LncRNA MEG3 promotes proliferation and differentiation of osteoblasts through Wnt/β-catenin signaling pathway. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 4521–4529 (2019).

199. Li, Z. *et al.* Gain of LINC00624 Enhances Liver Cancer Progression by Disrupting the Histone Deacetylase 6/Tripartite Motif Containing 28/Zinc Finger Protein 354C Corepressor Complex. *Hepatology* **73**, 1764–1782 (2021).

200. Hu, R., Zhong, P., Xiong, L. & Duan, L. Long Noncoding RNA Cancer Susceptibility Candidate 8 Suppresses the Proliferation of Bladder Cancer Cells via Regulating Glycolysis. *DNA Cell Biol.* **36**, 767–774 (2017).

201. Wang, J. *et al.* Identification and validation of inferior prognostic genes associated with immune signatures and chemotherapy outcome in acute myeloid leukemia. *Aging* **13**, (2021).

202. Xiao, B. *et al.* Identification and validation of immune-related lncRNA prognostic signatures for melanoma. *Immun Inflamm Dis* **9**, 1044–1054 (2021).

203. Peng, L., Chen, Z., Chen, Y., Wang, X. & Tang, N. MIR155HG is a prognostic biomarker and associated with immune infiltration and immune checkpoint

molecules expression in multiple cancers. *Cancer Med.* **8**, 7161–7173 (2019).

204. Dunn, P. J. *et al.* ImmPort: Shared research data for bioinformatics and immunology. *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2015) doi:10.1109/bibm.2015.7359752.

205. Mpakali, A. & Stratikos, E. The Role of Antigen Processing and Presentation in Cancer and the Efficacy of Immune Checkpoint Inhibitor Immunotherapy. *Cancers* **13**, (2021).

206. Qiao, J. & Fu, Y.-X. Cytokines that target immune killer cells against tumors. *Cell. Mol. Immunol.* **17**, 722–727 (2020).

207. Chandran, S. S. & Klebanoff, C. A. T cell receptor–based cancer immunotherapy: Emerging efficacy and pathways of resistance. *Immunological Reviews* vol. 290 127–147 (2019).

208. Ino, Y. *et al.* Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. *Br. J. Cancer* **108**, 914–923 (2013).

209. Rhee, J.-K. *et al.* Impact of Tumor Purity on Immune Gene Expression and Clustering Analyses across Multiple Cancer Types. *Cancer Immunol Res* **6**, 87–97 (2018).

210. García-Mulero, S. *et al.* Lung metastases share common immune features regardless of primary tumor origin. *J Immunother Cancer* **8**, e000491 (2020).

211. McArdel, S. L., Terhorst, C. & Sharpe, A. H. Roles of CD48 in regulating immunity and tolerance. *Clin. Immunol.* **164**, 10 (2016).

212. Jayachandran, R. *et al.* Disruption of Coronin 1 Signaling in T Cells Promotes Allograft Tolerance while Maintaining Anti-Pathogen Immunity. *Immunity* vol. 50

152–165.e8 (2019).

213. Li, Y., Luo, H., Liu, T., Zacksenhaus, E. & Ben-David, Y. The ets transcription factor Fli-1 in development, cancer and disease. *Oncogene* **34**, 2022–2031 (2014).

214. Lyu, Q. *et al.* SENCR stabilizes vascular endothelial cell adherens junctions through interaction with CKAP4. *Proceedings of the National Academy of Sciences* vol. 116 546–555 (2019).

215. Berger, A. C. *et al.* A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell* **33**, 690–705.e9 (2018).

216. Obeid, J. M., Hu, Y., Erdag, G., Leick, K. M. & Slingluff, C. L., Jr. The heterogeneity of tumor-infiltrating CD8+ T cells in metastatic melanoma distorts their quantification: how to manage heterogeneity? *Melanoma Res.* **27**, 211–217 (2017).

217. Paijens, S. T., Vledder, A., de Bruyn, M. & Nijman, H. W. Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell. Mol. Immunol.* **18**, 842–859 (2020).

218. Zhao, J., Yang, T. & Li, L. LncRNA FOXP4-AS1 Is Involved in Cervical Cancer Progression via Regulating miR-136-5p/CBX4 Axis. *Onco. Targets. Ther.* **13**, 2347 (2020).

219. Feng, Y., Wu, M., Hu, S., Peng, X. & Chen, F. LncRNA DDX11-AS1: a novel oncogene in human cancer. *Hum. Cell* **33**, 946–953 (2020).

220. de Santiago, P. R. *et al.* Immune-related lncRNA LINC00944 responds to variations in ADAR1 levels and it is associated with breast cancer prognosis. *Life Sci.* **268**, 118956 (2021).

221. Chen, D. *et al.* Genome-wide analysis of long noncoding RNA (lncRNA) expression in colorectal cancer tissues from patients with liver metastasis. *Cancer Med.* **5**,

1629–1639 (2016).

222. Li, X. *et al.* Cholangiocyte-Derived Exosomal lncRNA H19 Promotes Macrophage
Activation and Hepatic Inflammation under Cholestatic Conditions. *Cells* **9**, (2020).

223. Lai, C. *et al.* A robust signature of immune-related long non-coding RNA to predict
the prognosis of bladder cancer. *Cancer Med.* **10**, 6534–6545 (2021).

224. Li, N., Liu, Y. & Cai, J. LncRNA MIR155HG regulates M1/M2 macrophage
polarization in chronic obstructive pulmonary disease. *Biomed. Pharmacother.* **117**,
109015 (2019).

225. Xu, S. *et al.* Long Noncoding RNAs Control the Modulation of Immune Checkpoint
Molecules in Cancer. *Cancer Immunol Res* **8**, 937–951 (2020).

226. Liu, Z. *et al.* A lncRNA prognostic signature associated with immune infiltration and
tumour mutation burden in breast cancer. *J. Cell. Mol. Med.* **24**, 12444–12456
(2020).

227. Hu, W., Wang, Y., Fang, Z., He, W. & Li, S. Integrated Characterization of
lncRNA-Immune Interactions in Prostate Cancer. *Front Cell Dev Biol* **9**, 641891
(2021).

228. Huang, C. *et al.* Proteogenomic insights into the biology and treatment of
HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* **39**,
361–379.e16 (2021).

Appendix

Genomic Analysis Revealed New Oncogenic Signatures in *TP53*-Mutant Hepatocellular Carcinoma

Venkatesh Kancherla[1], Samir Abdullazade[1], Matthias S. Matter[1], Manuela Lanzafame[1], Luca Quagliata[1], Guglielmo Roma[2], Yujin Hoshida[3], Luigi M. Terracciano[1], Charlotte K. Y. Ng[1,4]* and Salvatore Piscuoglio[1]*

[1] Institute of Pathology, University Hospital Basel, Basel, Switzerland, [2] Department of Biology, University of Naples Federico II, Naples, Italy, [3] Division of Liver Diseases, Department of Medicine, Liver Cancer Program, Tisch Cancer Institute, Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [4] Department of Biomedicine, University of Basel, Basel, Switzerland

The *TP53* gene is the most commonly mutated gene in human cancers and mutations in *TP53* have been shown to have either gain-of-function or loss-of-function effects. Using the data generated by The Cancer Genome Atlas, we sought to define the spectrum of *TP53* mutations in hepatocellular carcinomas (HCCs) and their association with clinicopathologic features, and to determine the oncogenic and mutational signatures in *TP53*-mutant HCCs. Compared to other cancer types, HCCs harbored distinctive mutation hotspots at V157 and R249, whereas common mutation hotspots in other cancer types, R175 and R273, were extremely rare in HCCs. In terms of clinicopathologic features, in addition to the associations with chronic viral infection and high Edmondson grade, we found that *TP53* somatic mutations were less frequent in HCCs with cholestasis or tumor infiltrating lymphocytes, but were more frequent in HCCs displaying necrotic areas. An analysis of the oncogenic signatures based on the genetic alterations found in genes recurrently altered in HCCs identified four distinct *TP53*-mutant subsets, three of which were defined by *CTNNB1* mutations, 1q amplifications or 8q24 amplifications, respectively, that co-occurred with *TP53* mutations. We also found that mutational signature 12, a liver cancer-specific signature characterized by T>C substitutions, was prevalent in HCCs with wild-type *TP53* or with missense *TP53* mutations, but not in HCCs with deleterious *TP53* mutations. Finally, whereas patients with HCCs harboring deleterious *TP53* mutations had worse overall and disease-free survival than patients with *TP53*-wild-type HCCs, patients with HCCs harboring missense *TP53* mutations did not have worse prognosis. In conclusion, our results highlight the importance to consider the genetic heterogeneity among *TP53*-mutant HCCs in studies of biomarkers and molecular characterization of HCCs.

Keywords: *TP53* mutations, somatic mutations, copy number alterations, mutational signature, oncogenic signature

114

INTRODUCTION

Hepatocellular carcinomas (HCCs) display extensive histologic, transcriptomic and genetic diversity (Lee et al., 2004; Boyault et al., 2007; Chiang et al., 2008; Hoshida et al., 2009; Fujimoto et al., 2012; Guichard et al., 2012; Ahn et al., 2014; Schulze et al., 2015; The Cancer Genome Atlas Research Network, 2017). On the genetic level, genes involved in liver metabolism, Wnt and p53 signaling have been shown to be recurrently altered (Fujimoto et al., 2012; Guichard et al., 2012; Ahn et al., 2014; Schulze et al., 2015; The Cancer Genome Atlas Research Network, 2017). The most frequently mutated protein-coding genes are *CTNNB1* (encoding β-catenin) and *TP53* (encoding p53), both mutated in 20–40% of HCCs (Fujimoto et al., 2012; Guichard et al., 2012; Ahn et al., 2014; Schulze et al., 2015; The Cancer Genome Atlas Research Network, 2017).

TP53 is the most frequently mutated gene in human cancers (Kandoth et al., 2013). The p53 protein modulates multiple cellular functions, including transcription, DNA synthesis and repair, cell cycle arrest, senescence and apoptosis (Vogelstein et al., 2000). Mutations in *TP53* can abrogate these functions, leading to genetic instability and progression to cancer (Vogelstein et al., 2000). Across 12 major cancer types (excluding HCC), 42% of cancers harbored *TP53* somatic mutations, with at least 20% mutational rate in 10/12 cancer types and *TP53* mutations are associated with inferior prognosis and unfavorable clinicopathologic parameters, such as tumor stage (Kandoth et al., 2013). Furthermore, *TP53*-mutant tumors are highly enriched among tumors driven by copy number alterations (CNAs), with most remaining *TP53*-mutant tumors associated with the presence of somatic mutations in the Wnt and/or the RAS-RAF-ERK signaling pathways (Ciriello et al., 2013).

The pattern of *TP53* mutations is reminiscent of both an oncogene and a tumor suppressor gene (Vogelstein et al., 2013). The majority (86%) of *TP53* mutations are in the DNA-binding domain (Olivier et al., 2010; Kandoth et al., 2013). Most mutations in the DNA-binding domain are missense (88%) and approximately 1/3 of missense mutations affect the hotspot residues R175, G245, R248, R249, R273, and R282 (Olivier et al., 2010). Outside the DNA-binding domain, most mutations (∼60%) are nonsense or frameshift (Olivier et al., 2010). Mutant p53 proteins may lose the tumor suppressive functions and exert dominant-negative activities, but may also gain new oncogenic properties (Olivier et al., 2010; Muller and Vousden, 2014). Indeed, on the immunohistochemical level, p53 is generally detectable to various extents in samples with missense mutations but is undetectable in samples with truncating or frameshift mutations (Hall and McCluggage, 2006; Soussi et al., 2014).

In HCC, *TP53* mutational frequency has been reported to range between 22 and 33% (Fujimoto et al., 2012; Guichard et al., 2012; Cleary et al., 2013; Kan et al., 2013; Ahn et al., 2014; Jhunjhunwala et al., 2014; Shiraishi et al., 2014; Totoki et al., 2014; Weinhold et al., 2014; Schulze et al., 2015; Fujimoto et al., 2016; The Cancer Genome Atlas Research Network, 2017). However, the frequency varies between geographic regions, etiological

factors and carcinogen exposure, with more frequent *TP53* mutations in regions where hepatitis B virus (HBV) infection is endemic (Fujimoto et al., 2012; Guichard et al., 2012; The Cancer Genome Atlas Research Network, 2017). Similar to other cancer types, *TP53*-mutant HCCs have been associated with features linked to poor prognosis, including high levels of alpha-fetoprotein, high Edmondson grade, expression of stem-like markers, and activation of pro-oncogenic signaling pathways (Kiani et al., 2002; Breuhahn et al., 2004; Lee et al., 2004; Peng et al., 2004; Boyault et al., 2007; Chiang et al., 2008; Hoshida et al., 2009; Goossens et al., 2015). Furthermore, patients with *TP53*-mutant HCCs tend to have shorter overall (OS) and disease-free survival (DFS) (Yano et al., 2007; Woo et al., 2011; Cleary et al., 2013). However, it appears that not all *TP53* mutations in HCCs are equal. For instance, one of the most common mutation hotspots affecting residues R248/249 has an overall frequency of ∼10% among *TP53*-mutant HCCs (Fujimoto et al., 2012, 2016; Ahn et al., 2014; Schulze et al., 2015; The Cancer Genome Atlas Research Network, 2017). In particular, the R249S mutation resulting from G>T transversion has specifically been linked to the combined effect of aflatoxin B1 exposure and HBV infection (Bressac et al., 1991; Hsu et al., 1991) and this mutation is detected in >75% of HCC from areas with high aflatoxin B1 exposure (Gouas et al., 2009; Kew, 2010). Further hotspot mutations affecting preferentially HCC are located at the residues V157 and H193 (both at ∼2%) (Fujimoto et al., 2012, 2016; Ahn et al., 2014; Schulze et al., 2015; The Cancer Genome Atlas Research Network, 2017). Both R249S and V157F have been associated with stem cell-like traits and poor prognosis in HCC patients (Villanueva and Hoshida, 2011; Woo et al., 2011).

Finally, molecular classification studies have invariably grouped *TP53*-mutant HCCs under the umbrella of the aggressive subclass, but it is also clear that this subclass is molecularly, biologically and clinically heterogeneous (Boyault et al., 2007; Hoshida et al., 2009; Goossens et al., 2015).

Given the diverse pattern of *TP53* mutations, taking advantage of The Cancer Genome Atlas (TCGA) dataset, in this study we sought to determine the pattern of *TP53* somatic mutations in HCCs and its association with clinicopathologic features. Additionally, as *TP53* mutations are associated with HCC molecular subclasses with poor prognosis, we sought to define the oncogenic and mutational signatures among *TP53*-mutant HCCs.

MATERIALS AND METHODS

Sample Selection and Histologic Assessment

From TCGA liver hepatocellular carcinoma (LIHC) project (The Cancer Genome Atlas Research Network, 2017), 373 tumors with available somatic mutational data[1] (accessed April 2017) (Gao et al., 2013) were included in the study. Images of diagnostic hematoxylin & eosin (H&E) slides were retrieved from the cbioportal and reviewed by three expert hepatopathologists (SA,

---

[1]http://www.cbioportal.org

115

MSM and LMT) according to the guidelines by the World Health Organization (Bosman et al., 2010) to define the presence or absence of cholestasis, Mallory bodies, tumor infiltrating lymphocytes (TILs), vessel infiltration and necrotic areas. 4-point scale Edmondson and Steiner system was adopted for tumor grading as previously described (Edmondson and Steiner, 1954; Alexandrov et al., 2013). Clinical information were obtained from the cbioportal (Gao et al., 2013).

Classification of *TP53* Somatic Mutations

TP53 somatic non-synonymous and splice region mutations for the 373 HCCs were retrieved from the cbioportal (accessed April 2017) (Gao et al., 2013). *TP53* mutations were stratified according to (i) the mutation type as single-nucleotide missense mutations (also encompassing synonymous mutations affecting splice region, Supplementary Methods and Supplementary Table S1) or deleterious mutations (encompassing splice site, nonsense, in-frame, and frameshift mutations); (ii) whether the mutations were within or outside of the DNA-binding domain. For correlative analyses with clinicopathologic parameters, the sample (TCGA-DD-A1EE) with three *TP53* mutations (A161S, H193R and C277*) was classified as harboring deleterious mutation.

The spectrum of *TP53* mutations in non-LIHC TCGA datasets were retrieved from the cbioportal (accessed June 2017, Supplementary Table S2) (Gao et al., 2013). Mutation (lolliplot) diagrams and Oncoprints were generated using cbioportal (Gao et al., 2013).

Genomic and Transcriptomic Data Analysis

Gene-level copy number ("gistic2_thresholded," 370/373 samples) and expression ("IlluminaHiSeq," 367/373 samples) data were retrieved from the UCSC Xena Functional Genomics Browser[2] accessed April 2017). Gene-level copy number data were used to define genomic regions with differential frequencies of copy number alterations between HCCs with missense *TP53* mutations, with deleterious *TP53* mutations, or with wild-type *TP53*. Copy number states −2, −1, 0, 1, and 2 were considered homozygous deletion, heterozygous loss, copy number neutral, gain and high-level gain/amplification, respectively.

Transcriptomic data were in the form of gene-level, log-transformed, upper-quartile-normalized RSEM values. Molecular classification was performed according to Hoshida et al. (2009), using the Nearest Template Prediction: http://software.broadinstitute.org/cancer/software/genepattern. The R package limma was used to perform quantile normalization and for differential expression analysis. Multiple correction was performed using the Benjamini–Hochberg method. Genes with adjusted *P*-value < 0.05 were considered as differentially expressed.

The number of somatic mutations per sample was obtained from the cbioportal (Gao et al., 2013).

[2]https://xenabrowser.net/datapages/?cohort=TCGA%20Liver%20Cancer%20(LIHC)

Oncogenic Signatures

Oncogenic signature ("oncosign") classification and the selection of genomic features as 'selected functional elements' (SFEs) input data were performed as described by Ciriello et al. (2013). Specifically, we selected 29 significantly mutated genes that have previously been reported as cancer genes (Futreal et al., 2004; Fujimoto et al., 2012; Kandoth et al., 2013; Lawrence et al., 2014), 27 recurrent amplifications and 34 recurrent deletions as SFEs (Supplementary Methods). Robustness of the subclasses was assessed by removing 5, 10, or 20% of samples, reclassifying the reduced dataset, and calculating the Jaccard coefficients over 20 runs (Ciriello et al., 2013). Enrichment of genomic alterations was assessed using Chi-squared and Fisher's exact tests, as described (Ciriello et al., 2013).

Mutational Signatures

Decomposition of mutational signatures was performed using deconstructSigs (Rosenthal et al., 2016), based on the set of 30 mutational signatures ("signature.cosmic") (Alexandrov et al., 2013; Nik-Zainal et al., 2016), for the 358 samples with at least 30 somatic mutations. Mutational signatures with >20% weight were considered to have substantial contribution to the overall mutational landscape. For each sample, the mutation signature with the highest weight was considered the dominant mutational signature.

Pathway Analysis

Pathways analysis was performed using Ingenuity Pathway Analysis as previously described (Piscuoglio et al., 2014; Martelotto et al., 2015). *P* < 0.001 was considered significant (Supplementary Methods).

Statistical Analysis

Associations between *TP53* mutations and clinical/histologic features were assessed using Mann–Whitney *U*, Chi-squared or Fisher's exact tests as appropriate. Survival analyses were performed using the Kaplan-Meier method and the log-rank test. Univariate and multivariate analyses for OS and DFS were performed using the Cox proportional-hazards model. Mutual exclusivity and co-occurrence of somatic mutations were defined using the cbioportal (Gao et al., 2013). Statistical analyses comparing copy number profiles and defining genes up-regulated when gained or amplified and genes down-regulated when lost were performed as previously described (Supplementary Methods) (Piscuoglio et al., 2014). All tests were two-sided. *P* < 0.05 were considered statistically significant. Statistical analyses were performed with R v3.1.2 or SPSS v24 (IBM, Münchenstein, CH).

RESULTS

Clinicopathologic Characterization and Molecular Classification of HCCs

TP53 mutation status was available for 373 HCCs subjected to whole-exome sequencing by TCGA (The Cancer Genome

116

Atlas Research Network, 2017). Analysis of the clinical details of the patients revealed that the median age at diagnosis was 61 (range 16–90) and that 67.5% were male (Supplementary Table S3). Half of the patients were Caucasian (50.8%), with most remaining patients being Asian (43.9%). The most frequent primary risk factor was alcohol consumption (33.1%), followed by HBV (30.0%) and hepatitis C virus (HCV) infection (15.9%). Overall, history of at least one primary risk factor was noted in 74.2% patients (Supplementary Table S3).

We performed a comprehensive histopathologic review of the diagnostic H&E slides for all 373 included cases to assess Edmondson grade, the presence of cholestasis, Mallory bodies, vessel infiltration, necrotic areas, and TILs (**Figure 1** and Supplementary Table S3). Most samples were of intermediate grade, with 33.2, 60.6, and 5.4% graded as of Edmondson grades 2, 3, and 4, respectively. No sample was classified as of Edmondson grade 1. Cholestasis, Mallory bodies, vessel infiltration, necrotic areas, and TILs were present in 21.6, 22.0, 34.1, 24.8, and 47.3% of cases, respectively.

Molecular classification was performed for the 367 HCCs for which expression data were available according to Hoshida et al. (2009). 31.3, 21.5, and 47.2% of HCCs were classified as S1, S2 and S3, respectively (Supplementary Table S3).

Spectrum of *TP53* Somatic Mutations in HCCs

Given that *TP53* is one of the most frequently mutated genes in HCCs and its diverse spectrum of mutations in human cancers, we sought to define the spectrum and type of *TP53* mutations found in HCCs. A total of 116 somatic non-synonymous *TP53* mutations and 2 synonymous *TP53* mutations affecting splice regions were identified in 115 (30.8%) cases, including one case with three distinct mutations and one case with two. Missense (including missense and synonymous mutations affecting splice

region, Supplementary Methods and Supplementary Table S1) and deleterious (including nonsense, frame-shift, in-frame, splice site) mutations accounted for 73 (62%) and 45 (38%), respectively (**Figure 2**). Compared to other cancer types characterized by the TCGA, there was no difference between HCC and non-HCC tumor types in terms of the ratio of missense vs deleterious mutations ($P = 0.197$, Fisher's exact test).

Of the 73 missense and synonymous mutations affecting splice region, 51 (70%) affected known hotspot residues (Chang et al., 2016; Gao et al., 2017) and all but one (99%) affected the DNA-binding domain (**Figure 2A**). All missense mutations were predicted to be pathogenic by at least 2/5 *in silico* mutation effect predictors, with the two synonymous mutations affecting splice region also predicted to be disease causing (Supplementary Methods and Supplementary Table S1). The most frequent hotspot mutations were R249S (11/73, 15%), H193R (4/73, 5%), and R248Q/W (4/73, 5%). V157F, a mutation not considered to be a hotspot residue (Chang et al., 2016; Gao et al., 2017) but was reported as a mutation hotspot in HCCs (Woo et al., 2011), accounted for 4/73 (5%) of the missense mutations (**Figure 2A**). Compared to other cancer types, mutations affecting V157 and R249 accounted for greater proportions of the missense mutations in HCCs than in other cancer types (4/73, 5% vs. 22/1787, 1.2%, $P = 0.017$ and 11/73, 15% vs. 21/1787, 1.2%, $P < 0.001$, respectively, Fisher's exact tests, **Figures 2A,B**). In particular, R249S accounted for <0.5% of *TP53* missense mutations in non-HCC TCGA samples, but accounted for 15% of the missense mutations in HCCs ($P < 0.001$, Fisher's exact test). In contrast, the most frequent hotspots in non-HCC tumors R273 (178/1787, 10.0% of missense mutations) and R175 (112/1787, 6.3%) were only observed once and not at all, respectively, in HCCs ($P = 0.008$ and $P = 0.020$, respectively, Fisher's exact tests).

FIGURE 1 | Histologic features of hepatocellular carcinoma. Low-power view of hepatocellular carcinomas with tumor infiltrating lymphocytes **(A)**, necrotic areas **(B)**, vessel infiltration **(C)**, Mallory bodies **(D)**, cholestasis **(E)** and of high Edmondson grade **(F)**. Red arrows indicate the relevant histologic features.

117

FIGURE 2 | The distribution and the spectrum of *TP53* mutations. The distribution and spectrum of *TP53* missense **(A,B)** and deleterious **(C,D)** mutations in hepatocellular carcinoma **(A,C)** and in non-liver TCGA datasets **(B,D)**. Diagrams represent the protein domains of p53 encoded by the *TP53* gene. The presence of a mutation is shown on the *x* axis (lollipop), and the frequency of mutations is shown on the *y* axis. Missense mutations are presented as green circles, deleterious mutations (i.e., nonsense, frameshift, splice-site and in-frame) are depicted in black and brown circles. Plots were generated using cBioPortal tools (http://www.cBioPortal.org) and curated manually.

The 45 deleterious mutations comprised 13 (29%) nonsense point mutations, 20 (44%) frameshift small insertions or deletions (indels), 3 (7%) in-frame indels and 9 (20%) mutations affecting splice sites. Unlike missense mutations, the 45 deleterious mutations were spread across the *TP53* gene, with 32 (71%) in the DNA-binding domain, 3 (7%) in the tetramerization motif and 10 (22%) outside of these two domains (**Figure 2C**). In other cancer types, recurrent truncating mutations were observed at R196 (44/926, 4.8% of deleterious *TP53* mutations) and R213 (56/926, 6.0%), both of which were not observed in HCC (**Figures 2C,D**).

Our results demonstrate that the spectrum of *TP53* mutations in HCCs is distinct from that in non-HCC tumors, with HCC-specific recurrent hotspot mutations and a near absence of highly recurrent *TP53* mutations found in other cancer types.

TP53 Status Correlates with Specific Histopathologic and Clinical Features of HCCs

Next, we sought to define whether *TP53* mutation status correlated with clinicopathologic parameters. *TP53* mutations were more frequently found in male patients (35.9% vs. 20.7% in female; $P = 0.003$, Fisher's exact test) and in patients with at least one primary risk factor (35.1% vs. 20.9%; $P = 0.013$, Fisher's exact test), especially in HCCs associated with HBV/HCV infection (53.1% vs. 39.7%; $P = 0.021$, Fisher's exact test, **Table 1**). Patients from different racial backgrounds were associated with different *TP53* mutational frequencies ($P = 0.001$, Chi-squared test, **Table 1**). Black or African Americans had the highest frequency of *TP53* mutations (70.6% vs. Asians, 36.5%, $P = 0.009$, and vs. Caucasians, 22.8%, $P < 0.001$, Fisher's exact tests), while Asians displayed more frequent *TP53* mutations than Caucasians ($P = 0.006$, Fisher's exact test). No

association with age of patients or Child-Pugh classification was observed.

Correlation with histologic features revealed that *TP53*-mutant HCCs were associated with high Edmondson grade, accounting for 12.1, 38.5, and 65.0% of cases classified as Edmondson grades 2, 3, and 4, respectively ($P < 0.001$, Chi-squared test, **Table 1**). *TP53* mutations were less frequent in HCCs associated with cholestasis (17.5% vs. 38.4%; $P = 0.003$, Fisher's exact test) and were more frequent in HCCs with necrotic areas (43.5% vs. 26.9%; $P = 0.004$, Fisher's exact test, **Table 1**). The presence of TILs was associated with less frequent *TP53* mutations (37.4% vs. 62.6%; $P = 0.013$, Fisher's exact test; **Table 1**). No association was found between *TP53* mutation status and the presence of Mallory Bodies or vessel infiltration.

Further analyses comparing HCCs with missense or deleterious mutations showed that patients with HCCs with deleterious *TP53* mutations were slightly older than those with missense mutations (median 64 vs. 58, $P = 0.049$, Mann–Whitney U test, Supplementary Table S4). After excluding one patient (TCGA-DD-A1EE) with both deleterious mutation (C277*) and hotspot missense (H193R) mutations, the ages between the two groups were not different ($P = 0.058$, Mann–Whitney U test). Of note, *TP53* recurrent hotspots V157F, R158H, H193R, Y205, and R249S were exclusively found in tumors of high Edmondson grade (grades 3/4, $P = 0.038$, Fisher's exact test, compared to HCCs with other *TP53* mutations).

Correlating *TP53* status with molecular classification, (Hoshida et al., 2009) *TP53*-mutant HCCs were preferentially enriched in the S1 and S2 subclasses (36.5% and 42.5% vs. 21.8% in S3, $P = 0.001$, Chi-squared test, **Table 1**). Stratifying *TP53*-mutant HCCs into those with missense or deleterious mutations did not reveal association between *TP53* mutation types and molecular classification ($P = 0.459$, Chi-squared test, Supplementary Table S4).

118

TABLE 1 | Analyses of *TP53* status and clinicopathologic parameters in the 373 HCCs from The Cancer Genome Atlas cohort (The Cancer Genome Atlas Research Network, 2017).

		TP53 status		P-value
		Mutant [N (%)]	Wild-type [N (%)]	
Age (n = 372)	Median years	59	61	0.200
Gender (n = 372)	Female	25 (20.7)	96 (79.3)	**0.003**
	Male	90 (35.9)	161 (64.1)	
Child-Pugh classification grade (n = 243)	A	65 (29.4)	156 (70.6)	0.754
	B	7 (33.3)	14 (66.7)	
	C	0 (0)	1 (100)	
Race (n = 362)	America Indian or Alaskan native	1 (50)	1 (50)	**<0.001**
	Asian	58 (36.5)	101 (63.5)	
	Black or African American	12 (70.6)	5 (29.4)	
	Caucasian	42 (22.8)	142 (77.2)	
History of Primary Risk Factors (n = 353)	At least one risk factor	92 (35.1)	170 (64.9)	**0.013**
	No risk factor	19 (20.9)	72 (79.1)	
Edmondson Grade (n = 373)	2	15 (12.1)	109 (87.9)	**<0.001**
	3	87 (38.5)	139 (61.5)	
	4	13 (65.0)	7 (35.0)	
Cholestasis (n = 370)	Absent	101 (38.4)	189 (65.2)	**0.003**
	Present	14 (17.5)	66 (82.5)	
Mallory Bodies (n = 373)	Absent	94 (32.3)	197 (67.7)	0.280
	Present	21 (25.6)	61 (74.4)	
Vessel infiltration (n = 370)	Absent	72 (29.5)	172 (70.5)	0.407
	Present	43 (34.1)	83 (65.9)	
Necrotic areas (n = 371)	Absent	75 (26.9)	204 (73.1)	**0.004**
	Present	40 (43.5)	52 (56.5)	
Infiltrating lymphocytes (n = 372)	Absent	72 (62.6)	124 (48.2)	**0.013**
	Present	43 (37.4)	133 (51.8)	
Molecular classification by Hoshida et al. (2009, n = 367)	S1	42 (36.5)	73 (63.5)	**0.001**
	S2	31 (42.5)	42 (57.5)	
	S3	39 (21.8)	140 (78.2)	

Statistical comparisons were performed using Mann–Whitney U test, Fisher's exact test or Chi-Squared test. P < 0.05 was considered to be statistically significant.

These results demonstrate that, additional to the well-established associations with the male gender, HBV/HCV infection and high Edmondson grade, *TP53* mutations were less frequent in HCCs with cholestasis or TILs, but were more frequent in HCCs with necrotic areas.

Genomic Instability Is Not Associated with *TP53* Mutation Type

Next, we compared the number of somatic genetic alterations between *TP53*-wild-type and mutant cases. Mutational burden was higher in *TP53*-mutant HCCs, HCCs with missense *TP53* mutations and HCCs with deleterious *TP53* mutations than *TP53*-wild-type cases ($P < 0.001$, $P < 0.001$ and $P = 0.004$, respectively, Mann–Whitney U tests), but no difference was observed between cases with missense or deleterious mutations ($P = 0.799$, Mann–Whitney U test, Supplementary Figure S1A). Similarly, *TP53*-mutant HCCs, HCCs with missense *TP53* mutations and HCCs with deleterious *TP53* mutations all harbored higher number

of genes affected by CNAs compared with *TP53*-wild-type cases ($P < 0.001$, $P < 0.001$ and $P = 0.001$, respectively, Mann–Whitney U tests, Supplementary Figure S1B), with no difference between cases with missense or deleterious *TP53* mutations ($P = 0.352$, Mann–Whitney U test, Supplementary Figure S1B).

Consistent with their increased chromosomal instability, *TP53*-mutant HCCs displayed more frequent gains of chromosomes 1p, 3, 10p and 19p and losses of half the genome, notably of chromosomes 4, 5, 10q, 14, 17p, 18 and 19 (Supplementary Figures S2A–C). The CNA landscapes between HCCs with *TP53* missense or deleterious mutations were remarkably similar (Supplementary Figure S2D).

To identify potential CNA drivers associated with *TP53* mutations, we interrogated the genes overexpressed when gained and genes downregulated when lost in the regions that showed differential CNA frequencies between *TP53*-mutant and *TP53*-wild-type cases (Supplementary Figure S2A). Pathway analysis of the copy number-regulated genes revealed that *TP53*-mutant

119

FIGURE 3 | Oncogenic signature classes in *TP53*-mutant hepatocellular carcinoma. The pattern of mutations in *TP53, CTNNB1, BAP1, RB1, JAK1* and *KEAP1* in hepatocellular carcinoma **(A)**. Number of *TP53*-mutant samples classified as OSC1, OSC2, OSC3, and OSC4, according to the color key in A **(B)**. Number of mutational **(C)** and copy number **(D)** 'selected functional elements' (SFEs) in the different subclasses. The distribution of mutational vs copy number SFEs in *TP53*-mutant cases **(E)**. The shade of red is proportional to the number of samples for a given (*x,y*) position. Heatmap shows the mutational and copy number SFEs altered in at least 5% of the samples in at least one oncogenic signature class **(F)**. Shades of red and blue are proportional to the number of samples with a given genetic alteration, according to the color key. Plot in **(A)** was generated using cBioPortal (http://www.cBioPortal.org) and curated manually.

cases displayed deregulation in pathways associated with EIF2 signaling, protein ubiquitination pathway, RNA polymerase-II complex and DNA repair pathways, and in molecular and cellular functions related to cell death and survival, cell cycle, DNA replication, recombination and repair (Supplementary Figure S3).

TP53-Mutant HCCs Displayed Heterogeneous Oncogenic Signatures

In HCCs, *TP53* and *CTNNB1* mutations were largely mutually exclusive (*P* = 0.028, **Figure 3A**) (Fujimoto et al., 2012; Guichard et al., 2012; Schulze et al., 2015; The Cancer Genome

Atlas Research Network, 2017). Additionally, *TP53* and *BAP1* mutations were also mutually exclusive (*P* = 0.004; **Figure 3A**). In contrast, *TP53* mutations co-occurred with *RB1, JAK1* and *KEAP1* mutations (*P* = 0.028, *P* = 0.034 and *P* = 0.044, respectively, **Figure 3A**). These observations suggest that *TP53*-mutant HCCs likely constitute a genetically heterogeneous subclass and may be subclassified into categories with distinct oncogenic signatures.

To define the oncogenic signatures in *TP53*-mutant HCCs, we performed unsupervised partitioning of the samples into classes with distinct patterns of likely 'driver' genetic alterations (or 'selected functional elements,' SFEs), (Ciriello

120

et al., 2013) including mutations in 29 significantly mutated genes, amplifications in 27 recurrently amplified regions, and homozygous deletions in 34 recurrently deleted regions (see Materials and Methods). Among the 144 TP53-mutant HCCs with mutational and CNA data, we found median of 2 mutational (range 0–11) and 2.5 CNA (range 0–13) SFEs in each case and identified four robust oncogenic signature classes (OSCs, **Figures 3B–E** and Supplementary Figure S4A). HCCs with missense or deleterious TP53 mutations did not cluster separately ($P = 0.305$, Chi-squared test, **Figure 3B**), nor HCCs of distinct transcriptomic subclasses (Supplementary Figure S4B).

Inspection of the SFEs that characterized each OSC revealed that OSC1 was defined by the presence of CTNNB1 mutations (100%, $P < 0.001$, Fisher's exact test, **Figure 3F**). The most frequent alteration in OSC2 was 8q24.21 amplification (encompassing MYC, 67%, $P < 0.001$, Fisher's exact test), while the most frequent alterations in OSC4 were 1q21.3 (encompassing CHD1L and HORMAD1, 60%) and 1q42.2 (encompassing TARBP1 and EXO1, 63%) amplifications (both $P < 0.001$, Fisher's exact tests, **Figure 3F**). OSC3 was notable for lacking highly recurrent genetic alterations, with the most frequent alteration being 11q13.3 amplification (CCND1, 23%, $P = 0.011$, Fisher's exact test). Additionally, ARID1A mutations were enriched in OSC1 (35%, $P < 0.001$, Fisher's exact test), while 10q23.21 deletion (PTEN, 20%) and 6p25.2 amplification (VEGFA, 23%) were enriched in OSC4 ($P = 0.020$ and $P = 0.001$, respectively, Fisher's exact tests). We also found that OSC1 harbored higher number of mutational SFEs and lower number of CNA SFEs ($P < 0.001$ and $P = 0.002$, respectively, Mann–Whitney U tests, **Figures 3C,D**) compared to other classes. By contrast, OSC4 harbored higher number of CNA SFEs than the other classes ($P < 0.001$, respectively, Mann–Whitney U test, **Figure 3D**). The TP53 R249S hotspot mutation was not associated with specific OSC classes ($P = 0.591$, Chi-squared test). Finally, OSC1/2 were more frequently associated with the presence of TILs than OSC3/4 ($P = 0.028$, Chi-squared test). No other associations between histologic or clinicopathologic parameters and OSCs were found.

These observations are concordant with the observation that tumors are primarily driven by either somatic mutations or CNAs but rarely both (Ciriello et al., 2013) (**Figure 3E** and Supplementary Figures S4C,D). Furthermore, we identified subclasses of TP53-mutant HCCs likely driven by co-occurring CTNNB1 mutations, 8q24.21 (MYC) amplification or 1q amplification in a mutually exclusive manner.

Mutational Signatures in *TP53*-Mutant HCCs

The somatic mutational landscapes are shaped by endogenous and/or environmental biological and chemical processes (Alexandrov et al., 2013). More than 10 mutational signatures have been identified in liver cancers, including two liver cancer-specific signatures 12 and 16 of unknown etiology, both of which are characterized by frequent T>C substitutions but with different sequence contexts (Alexandrov et al., 2013).

To determine whether TP53-mutant HCCs harbored distinct mutational signatures compared to TP53-wild-type HCCs, we inferred the underlying mutational processes for the 358 HCCs with at least 30 somatic mutations (Alexandrov et al., 2013; Nik-Zainal et al., 2016). The age-associated signature 5, (Alexandrov et al., 2015) and the liver cancer-specific signatures 12 and 16 contributed substantially ($\geq$20% weight) to the mutational landscapes in 17.0, 12.8, and 53.4% of the samples, respectively (**Figure 4**). Together, 72.9% of HCCs harbored signatures 5, 12 or 16 as the dominant signatures (14.0, 10.6, and 48.3%, respectively).

A comparison of the mutational signatures with substantial contribution ($\geq$20%) to the mutational landscapes of TP53-mutant or TP53-wild-type HCCs revealed that only the aflatoxin-associated signature 24 was enriched among TP53-mutant HCCs (7/114, 6.1% vs. 4/244, 1.6%, $P = 0.042$, Fisher's exact test).

We further compared the mutational signatures between HCCs with missense or deleterious TP53 mutations. Interestingly, while 18.6% (13/70) of samples with missense TP53 mutations displayed substantial contribution from signature 12, only 4.5% (2/44) of samples harboring deleterious TP53 mutations did ($P = 0.044$, Fisher's exact test), with signature 12 being the dominant signature in 15.7% (11/70) and 2.3% (1/44) of samples with missense or deleterious TP53 mutations, respectively ($P = 0.027$, Fisher's exact test, **Figure 4**). No difference in other signatures was observed. The aflatoxin-associated signature 24 was enriched among R249S-mutant HCCs compared other TP53-mutant HCCs (4/11, 36% vs. 3/103, 3%, $P = 0.001$ for substantial contribution and 3/11, 27% vs. 2/103, 2%, $P = 0.006$ for dominant signature, Fisher's exact tests).

Taken together, our results suggest that the different types of TP53 mutations were associated with distinct mutational processes. Specifically, signature 12 was rarely found in HCCs with deleterious TP53 mutations.

Distinct Types of *TP53* Mutations Are Associated with Different Prognoses

Previous studies found that associations between the types of TP53 mutations and prognoses in breast, and head and neck cancers (Olivier et al., 2006; Ozcelik et al., 2007; Vegran et al., 2013; Lapke et al., 2016). Here we hypothesized that patients with HCCs harboring TP53 missense or deleterious mutations may display different prognoses. Considering the patients with available data on OS ($n = 372$) or DFS ($n = 321$), we found that patients with TP53-mutant HCCs displayed a more aggressive behavior including shorter OS and DFS than TP53-wild-type patients ($P = 0.018$ and $P = 0.005$, respectively, log-rank tests, **Figure 5**). Patients with missense or deleterious TP53 mutations did not differ in OS or DFS ($P = 0.129$ and $P = 0.148$, respectively, log-rank tests, **Figure 5**). Importantly, while patients with deleterious TP53 mutations had worse OS and DFS than TP53-wild-type patients ($P = 0.004$ and $P = 0.001$, respectively, log-rank tests, **Figure 5**), there was no difference in OS or DFS between patients with missense TP53 mutations and those wild-type for TP53 ($P = 0.192$ and $P = 0.084$, respectively, log-rank tests, **Figure 5**).

121
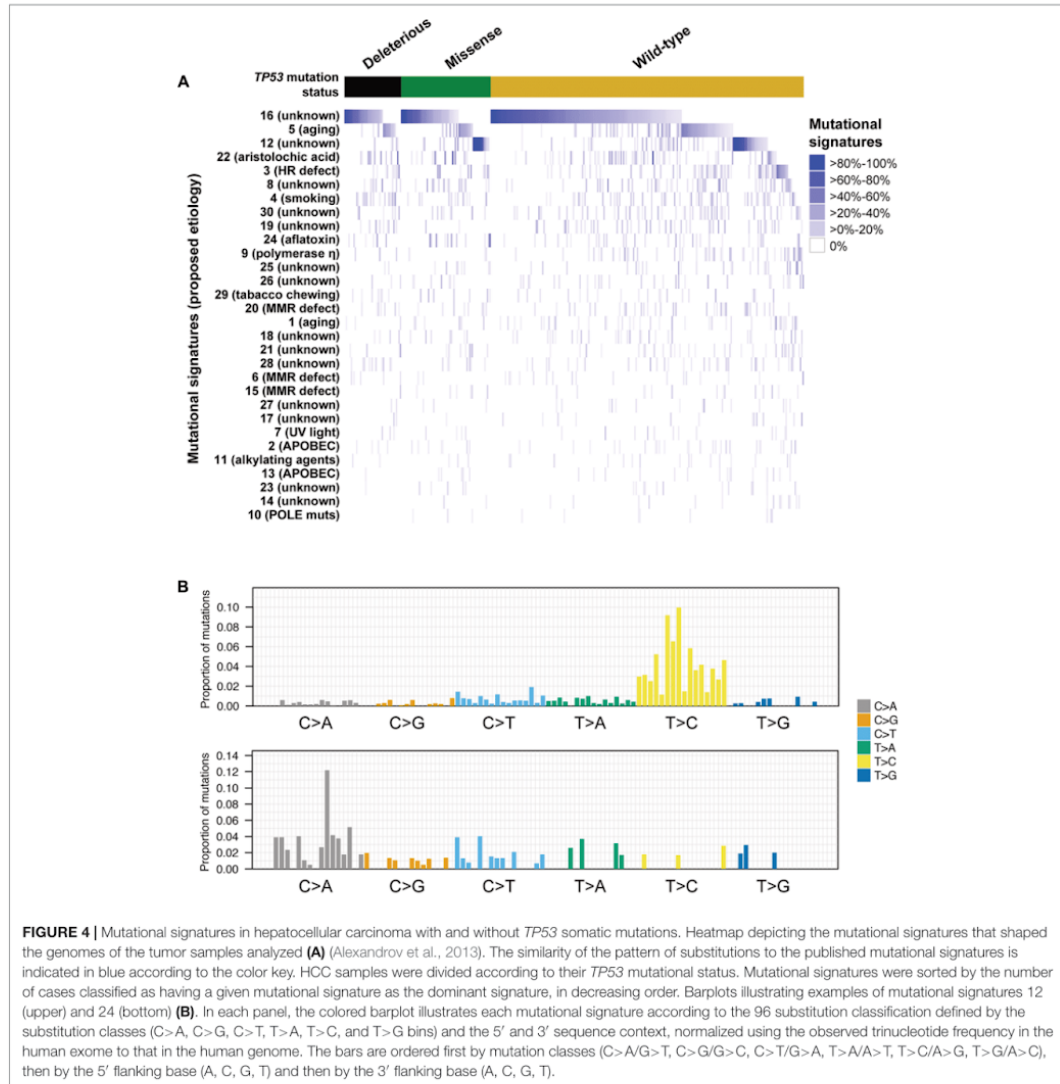
FIGURE 4 | Mutational signatures in hepatocellular carcinoma with and without *TP53* somatic mutations. Heatmap depicting the mutational signatures that shaped the genomes of the tumor samples analyzed **(A)** (Alexandrov et al., 2013). The similarity of the pattern of substitutions to the published mutational signatures is indicated in blue according to the color key. HCC samples were divided according to their *TP53* mutational status. Mutational signatures were sorted by the number of cases classified as having a given mutational signature as the dominant signature, in decreasing order. Barplots illustrating examples of mutational signatures 12 (upper) and 24 (bottom) **(B)**. In each panel, the colored barplot illustrates each mutational signature according to the 96 substitution classification defined by the substitution classes (C>A, C>G, C>T, T>A, T>C, and T>G bins) and the 5′ and 3′ sequence context, normalized using the observed trinucleotide frequency in the human exome to that in the human genome. The bars are ordered first by mutation classes (C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, T>G/A>C), then by the 5′ flanking base (A, C, G, T) and then by the 3′ flanking base (A, C, G, T).

As an exploratory analysis, we asked whether OSCs or mutational signatures of *TP53*-mutant HCCs were prognostic. Compared to OSC1 (28 months), OSC2 (26 months) and OSC3 (median not reached), OSC4 was associated with the shortest median OS of 14 months, although the difference was not statistically significant ($P = 0.366$, log-rank test; Supplementary Figure S4E). Univariate Cox regression analyses revealed that the aflatoxin-associated signature 24 (HR 3.275, CI 1.279–8.384, $P = 0.013$), HBV infection status and the presence of necrotic areas were associated with poor prognosis (Supplementary Table S5). However, in a multivariate analysis, mutational signature 24 was not an independent prognostic indicator ($P = 0.242$; Supplementary Table S5).

Taken together, our results showed only patients with deleterious *TP53* mutations but not missense *TP53* mutations were associated with significantly worse OS and DFS in this cohort.

122

FIGURE 5 | *TP53* mutation status is associated with worse overall and disease-free survival. Overall **(A)** and disease-free survival **(B)** of HCC patients with and without *TP53* somatic mutations using the Kaplan–Meier method. Median survival for each group is indicated in parentheses. Statistical comparisons were performed using log-rank tests. *P* < 0.05 was considered statistically significant.

DISCUSSION

In this study, we performed a detailed analysis of *TP53* somatic mutational spectrum in HCCs, with nearly all missense mutations (98%) and most deleterious mutations (73%) affecting the DNA-binding domain. Notably, we found that the residues mutated in HCCs differed from those in other cancer types. Hotspot mutations R249S and V157F were common in HCCs but extremely rare in other cancers, while mutations affecting R175 and R273, two of the most frequently mutated residues in other cancers, were nearly absent in HCCs. This latter observation also applies to other HCC datasets (Ahn et al., 2014; Schulze et al., 2015), suggesting that *TP53* mutational spectrum in HCC is distinct from that in other cancers.

To determine the genotype–phenotype correlation between *TP53* mutation status and clinicopathologic parameters, we performed a detailed assessment of histologic features using H&E slides. We confirmed the established associations with the male gender, HBV/HCV infection and high Edmondson grade. Additionally, *TP53* mutations were associated with the presence of necrotic areas, and accordingly, with the absence of cholestasis, a feature more frequently observed in well-differentiated HCCs. Finally, we observed that the presence of TILs was associated with less frequent *TP53* mutations, in line with the favorable prognosis associated with tumors with high TILs in other tumor types (Mahmoud et al., 2011).

Analysis of the mutational signatures revealed that signatures 16 of unknown etiology and the age-associated signature 5 (Alexandrov et al., 2015) were the most prevalent in HCCs. We also found that signature 12 of unknown etiology, characterized by frequent T>C substitutions, was prevalent in *TP53*-wild-type and HCCs with missense *TP53* mutations but were largely absent in those with deleterious *TP53* mutation. A previous study reported that the W3 signature, which was highly similar

to signature 12 (Fujimoto et al., 2012), was associated with the age of patients. Here we found no difference in the age of patients when we considered tumors with strictly missense or deleterious *TP53* mutations (i.e., excluding one patient with both types). The basis of signature 12 is thus unclear and further studies are required to elucidate its biological significance.

Adopting the algorithm of "oncosign" (Ciriello et al., 2013), we identified four robust subclasses of *TP53*-mutant HCCs with distinct oncogenic signatures. Of these classes, one subclass was likely driven by co-occurring *CTNNB1* mutations, while two subclasses were likely driven by amplicon drivers on 1q and 8q. 1q21 amplification has been linked to hepatocarcinogenesis, with *ALC1* (*CHD1L*) overexpression in HCC cells shown to promote G1/S phase transition and to inhibit apoptosis (Ma et al., 2008). The authors further suggested that the oncogenic function of ALC1 might be associated with its role in promoting cell proliferation by down-regulating p53 expression (Ma et al., 2008). The 1q21 amplicon also contains *HORMAD1*, a gene that has been shown to drive chromosomal instability in breast cancer (Watkins et al., 2015). As for 8q24, in addition to the well-known oncogenic role of *MYC*, previous studies have also shown that *MYC* amplification is an indicator of malignant potential and poor prognosis in HCC (Lin et al., 2010), and that the co-occurrence of *MYC* amplification and p53 alteration may contribute to HCC progression (Kawate et al., 1999). The remaining subclass did not have highly recurrent genetic alterations. Interestingly, this subclass was numerically, though not statistically, associated with the most favorable OS among the four classes. One may speculate that *TP53*-mutant HCCs lacking additional drivers may constitute a less aggressive subclass. Of note, the features that characterized the four OSCs were largely mutually exclusive, suggesting that distinct oncogenic processes are operative in non-overlapping subsets of *TP53*-mutant HCCs.

123

TP53 mutation status predicts worse OS and DFS in HCC patients (Yano et al., 2007; Woo et al., 2011; Cleary et al., 2013). However, we found that patients with deleterious mutations, but not those with missense mutations, were associated with worse OS and DFS compared to patients wild-type for *TP53*. This is in line with other tumor types, in which different types of *TP53* mutations have been associated with different prognoses (Olivier et al., 2006; Ozcelik et al., 2007; Vegran et al., 2013; Lapke et al., 2016). In fact, the risk of death or relapse for patients harboring deleterious mutation is 2.3 times (HR = 2.36 and 2.063, respectively) higher than *TP53*-wild-type patients. The prognosis for patients with missense mutations appears to sit between those with wild-type *TP53* or deleterious *TP53* mutations, albeit not statistically different from either group. It is conceivable that the prognostic significance of the type of *TP53* mutations may be confirmed in a larger cohort with extensive follow-up.

It has been suggested that *TP53* missense mutations have varying capacity to transactivate p53 target genes and to alter the responsiveness to chemotherapeutic agents in breast cancer (Jordan et al., 2010). A differential expression analysis using the HCC TCGA dataset comparing HCCs with *TP53* missense mutations and those with *TP53* deleterious mutations identified *TP53* itself as up-regulated but did not identify significantly altered genes (data not shown). Furthermore, HCCs harboring the missense mutations functionally shown to lack the ability to transactivate genes with p53 response elements (Jordan et al., 2010) did not differ from HCCs with other missense mutations on the transcriptomic level (data not shown). It is thus unclear precisely how the various *TP53* mutations may differentially alter the transcriptomic landscape of HCCs. Further functional studies may be required to elucidate how the types of *TP53* mutations may affect its biological functions.

In HCC molecular characterization studies to date, HCCs are typically classified as *TP53*-wild-type or *TP53*-mutant, where all *TP53* mutations were treated as equal (Fujimoto et al., 2012; The Cancer Genome Atlas Research Network, 2017). However, many studies have demonstrated that *TP53* can be affected by either (or both) gain-of-function or loss-of-function mutations, with missense mutations preferentially displaying gain-of-function or neomorphic properties (Muller and Vousden, 2014). Our study has demonstrated that HCCs with missense or deleterious *TP53* mutations display similar clinicopathologic features, mutational/CNA burden and oncogenic signatures, but are associated with distinct mutational signatures. Clinically, while patients with tumors harboring deleterious *TP53* mutations had worse prognosis compared to those wild-type for *TP53*, there was no statistically significant difference between those with missense mutations and those wild-type for *TP53*. Our study highlights the importance to consider the type of *TP53* mutations in studies of biomarkers and molecular characterization of HCCs.

Our study has limitations. Despite TCGA being the largest genomic study of HCC, it is by no means the only large-scale study. However, as one of our aims was to define clinicopathologic correlates, we chose TCGA as it is the only study with publicly available H&E slides for pathology review. Secondly, the power of the OS and DFS analyses was limited due to the cohort size. Further studies may reveal whether prognosis is related to the type of *TP53* mutations, as has been shown in other cancers. Thirdly, our analyses did not consider the non-coding genome due to the nature of the sequencing performed by the TCGA. Given the frequent mutations in non-coding regions such as *TERT* promoter, *MALAT1* and *NEAT1* (Fujimoto et al., 2012; Schulze et al., 2015), it is conceivable that additional oncogenic signatures within *TP53*-mutant HCCs may emerge.

CONCLUSION

Our study highlights the genetic heterogeneity among *TP53*-mutant HCCs and that patients with HCCs harboring different types of *TP53* mutations may be associated with distinct prognoses. Future work will be required to elucidate whether the co-occurring genetic alterations act synergistically with *TP53* mutations to promote carcinogenesis in HCCs.

AUTHOR CONTRIBUTIONS

CKYN and SP conceived and supervised the study. SA, MSM, and LMT performed the histologic review. VK performed the bioinformatics analyses under supervision of CKYN. VK, CKYN, and SP performed the statistical analyses. VK performed the transcriptomic classification under supervision of YH. VK, ML, LQ, GR, LMT, CKYN, and SP analyzed the data and critically discussed the results. CKYN and SP wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00002/full#supplementary-material

124

REFERENCES

Ahn, S. M., Jang, S. J., Shim, J. H., Kim, D., Hong, S. M., Sung, C. O., et al. (2014). Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* 60, 1972–1982. doi: 10.1002/hep.27198

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., et al. (2015). Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47, 1402–1407. doi: 10.1038/ng.3441

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421. doi: 10.1038/nature12477

Bosman, F. T., Carneiro, F., Hruban, R. H., and Theise, N. D. (2010). *WHO Classification of Tumours of the Digestive System World Health Organization.* Lyon: IARC.

Boyault, S., Rickman, D. S., De Reynies, A., Balabaud, C., Rebouissou, S., Jeannot, E., et al. (2007). Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 45, 42–52. doi: 10.1002/hep.21467

Bressac, B., Kew, M., Wands, J., and Ozturk, M. (1991). Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* 350, 429–431. doi: 10.1038/350429a0

Breuhahn, K., Vreden, S., Haddad, R., Beckebaum, S., Stippel, D., Flemming, P., et al. (2004). Molecular profiling of human hepatocellular carcinoma defines mutually exclusive interferon regulation and insulin-like growth factor II overexpression. *Cancer Res.* 64, 6058–6064. doi: 10.1158/0008-5472.CAN-04-0292

Chang, M. T., Asthana, S., Gao, S. P., Lee, B. H., Chapman, J. S., Kandoth, C., et al. (2016). Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* 34, 155–163. doi: 10.1038/nbt.3391

Chiang, D. Y., Villanueva, A., Hoshida, Y., Peix, J., Newell, P., Minguez, B., et al. (2008). Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Res.* 68, 6779–6788. doi: 10.1158/0008-5472.CAN-08-0742

Ciriello, G., Miller, M. L., Aksoy, B. A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* 45, 1127–1133. doi: 10.1038/ng.2762

Cleary, S. P., Jeck, W. R., Zhao, X., Chen, K., Selitsky, S. R., Savich, G. L., et al. (2013). Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology* 58, 1693–1702. doi: 10.1002/hep.26540

Edmondson, H. A., and Steiner, P. E. (1954). Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies. *Cancer* 7, 462–503. doi: 10.1002/1097-0142(195405)7:3<462::AID-CNCR2820070308>3.0.CO;2-E

Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., et al. (2016). Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* 48, 500–509. doi: 10.1038/ng.3547

Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K. A., Hosoda, F., Nguyen, H. H., et al. (2012). Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* 44, 760–764. doi: 10.1038/ng.2291

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., et al. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183. doi: 10.1038/nrc1299

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6:l1. doi: 10.1126/scisignal.2004088

Gao, J., Chang, M. T., Johnsen, H. C., Gao, S. P., Sylvester, B. E., Sumer, S. O., et al. (2017). 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* 9:4. doi: 10.1186/s13073-016-0393-x

Goossens, N., Sun, X., and Hoshida, Y. (2015). Molecular classification of hepatocellular carcinoma: potential therapeutic implications. *Hepat. Oncol.* 2, 371–379. doi: 10.2217/hep.15.26

Gouas, D., Shi, H., and Hainaut, P. (2009). The aflatoxin-induced TP53 mutation at codon 249 (R249S): biomarker of exposure, early detection and target for therapy. *Cancer Lett.* 286, 29–37. doi: 10.1016/j.canlet.2009.02.057

Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I. B., et al. (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* 44, 694–698. doi: 10.1038/ng.2256

Hall, P. A., and McCluggage, W. G. (2006). Assessing p53 in clinical contexts: unlearned lessons and new perspectives. *J. Pathol.* 208, 1–6. doi: 10.1002/path.1913

Hoshida, Y., Nijman, S. M., Kobayashi, M., Chan, J. A., Brunet, J. P., Chiang, D. Y., et al. (2009). Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* 69, 7385–7392. doi: 10.1158/0008-5472.CAN-09-1089

Hsu, I. C., Metcalf, R. A., Sun, T., Welsh, J. A., Wang, N. J., and Harris, C. C. (1991). Mutational hot spot in the p53 gene in human hepatocellular carcinomas. *Nature* 350, 427–428. doi: 10.1038/350427a0

Jhunjhunwala, S., Jiang, Z., Stawiski, E. W., Gnad, F., Liu, J., Mayba, O., et al. (2014). Diverse modes of genomic alteration in hepatocellular carcinoma. *Genome Biol.* 15, 436. doi: 10.1186/s13059-014-0436-9

Jordan, J. J., Inga, A., Conway, K., Edmiston, S., Carey, L. A., Wu, L., et al. (2010). Altered-function p53 missense mutations identified in breast cancers can have subtle effects on transactivation. *Mol. Cancer Res.* 8, 701–716. doi: 10.1158/1541-7786.MCR-09-0442

Kan, Z., Zheng, H., Liu, X., Li, S., Barber, T. D., Gong, Z., et al. (2013). Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* 23, 1422–1433. doi: 10.1101/gr.154492.113

Kandoth, C., Mclellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339. doi: 10.1038/nature12634

Kawate, S., Fukusato, T., Ohwada, S., Watanuki, A., and Morishita, Y. (1999). Amplification of *c-myc* in hepatocellular carcinoma: correlation with clinicopathologic features, proliferative activity and p53 overexpression. *Oncology* 57, 157–163. doi: 10.1159/000012024

Kew, M. C. (2010). Epidemiology of chronic hepatitis B virus infection, hepatocellular carcinoma, and hepatitis B virus-induced hepatocellular carcinoma. *Pathol. Biol.* 58, 273–277. doi: 10.1016/j.patbio.2010.01.005

Kiani, C., Chen, L., Wu, Y. J., Yee, A. J., and Yang, B. B. (2002). Structure and function of aggrecan. *Cell Res.* 12, 19–32. doi: 10.1038/sj.cr.7290106

Lapke, N., Lu, Y. J., Liao, C. T., Lee, L. Y., Lin, C. Y., Wang, H. M., et al. (2016). Missense mutations in the TP53 DNA-binding domain predict outcomes in patients with advanced oral cavity squamous cell carcinoma. *Oncotarget* 7, 44194–44210. doi: 10.18632/oncotarget.9925

Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., et al. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501. doi: 10.1038/nature12912

Lee, J. S., Chu, I. S., Heo, J., Calvisi, D. F., Sun, Z., Roskams, T., et al. (2004). Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* 40, 667–676. doi: 10.1002/hep.20375

Lin, C. P., Liu, C. R., Lee, C. N., Chan, T. S., and Liu, H. E. (2010). Targeting c-Myc as a novel approach for hepatocellular carcinoma. *World J. Hepatol.* 2, 16–20. doi: 10.4254/wjh.v2.i1.16

Ma, N. F., Hu, L., Fung, J. M., Xie, D., Zheng, B. J., Chen, L., et al. (2008). Isolation and characterization of a novel oncogene, amplified in liver cancer 1, within a commonly amplified region at 1q21 in hepatocellular carcinoma. *Hepatology* 47, 503–510. doi: 10.1002/hep.22072

Mahmoud, S. M., Paish, E. C., Powe, D. G., Macmillan, R. D., Grainge, M. J., Lee, A. H., et al. (2011). Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.* 29, 1949–1955. doi: 10.1200/JCO.2010.30.5037

Martelotto, L. G., De Filippo, M. R., Ng, C. K., Natrajan, R., Fuhrmann, L., Cyrta, J., et al. (2015). Genomic landscape of adenoid cystic carcinoma of the breast. *J. Pathol.* 237, 179–189. doi: 10.1002/path.4573

Muller, P. A., and Vousden, K. H. (2014). Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* 25, 304–317. doi: 10.1016/j.ccr.2014.01.021

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54. doi: 10.1038/nature17676

125

High Expression of FAP in Colorectal Cancer Is Associated With Angiogenesis and Immunoregulation Processes

Mairene Coto-Llerena[1†], Caner Ercan[1†], Venkatesh Kancherla[1], Stephanie Taha-Mehlitz[2,3], Serenella Eppenberger-Castori[1], Savas D. Soysal[3], Charlotte K. Y. Ng[4], Martin Bolli[3], Markus von Flüe[3], Guillaume P. Nicolas[5], Luigi M. Terracciano[1], Melpomeni Fani[6*] and Salvatore Piscuoglio[1,2,3*]

[1] Institute of Pathology and Medical Genetics, University Hospital Basel, Basel, Switzerland, [2] Visceral Surgery Research Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland, [3] Department of Visceral Surgery, Clarunis University Centre for Gastrointestinal and Liver Diseases, St. Clara Hospital and University Hospital Basel, Basel, Switzerland, [4] Department for BioMedical Research (DBMR), University of Bern, Bern, Switzerland, [5] Division of Nuclear Medicine, University Hospital Basel, Basel, Switzerland, [6] Division of Radiopharmaceutical Chemistry, University Hospital Basel, University of Basel, Basel, Switzerland

Fibroblast activation protein α (FAP) plays an important role in tissue remodeling and helps tumor cells invade surrounding tissue. We sought to investigate FAP as a prognostic molecular marker in colorectal cancer (CRC) using immunohistochemical and transcriptomic data. FAP expression and clinicopathological information were obtained from The Cancer Genome Atlas data set. The association of FAP expression and tissue cellular heterogeneity landscape was explored using the xCell method. We evaluated FAP protein expression in a cohort of 92 CRCs and 19 non-tumoral tissues. We observed that FAP was upregulated in tumors both at the mRNA and protein levels, and its expression was associated with advanced stages, poor survival, and consensus molecular subtype 4. FAP expression was also associated with angiogenesis and collagen degradation. We observed an enrichment in immune-cell process–related genes associated with FAP overexpression. Colorectal cancers with high FAP expression display an inflamed phenotype enriched for macrophages and monocytes. Those tumors showed enrichment for regulatory T cell populations and depletion of T_H1 and natural killer T cells, pointing to an immunosuppressive environment. Colorectal cancers with high levels of stromal FAP are associated with aggressive disease progression and survival. Our results suggest that FAP plays additional roles in tumor progression such as modulation of angiogenesis and immunoregulation in the tumor microenvironment.

Keywords: colorectal cancer, cancer-associated fibroblast, immunohistochemistry, FAP, gene expression

INTRODUCTION

Tumor-infiltrating immune cells as well as cancer-associated fibroblasts (CAFs) are important components of the tumor microenvironment. In human cancer, the tumor microenvironment has been suggested as a new component for the classification of malignant tumors including colorectal cancers (CRCs) (1–3). In particular, CAFs play important roles in modulating tumor development

126

and prognosis via releasing proteolytic enzymes, growth factors, and immunomodulatory cytokines (4, 5).

Fibroblast activation protein α (FAP, also called seprase) is a prolyl-specific serine proteinase, highly upregulated in fibroblasts especially at sites of active tissue remodeling, including wound healing and fibrosis (6, 7). In CRC, previous studies reported the detection for FAP in more than 93% of the tumor. Among those, 30% showed high intensity for FAP staining (8). High FAP expression has been proposed as a biomarker for disease progression in metastatic CRCs (9). Similarly in rectal cancer, high FAP expression after preoperative chemoradiotherapy has been associated with poor prognosis (9). Given the scientific evidence, FAP has been considered as a candidate for targeted therapy in CRC. So far, diverse approaches, including FAP-targeting vaccines and immunotherapies, have been used in preclinical studies to deplete FAP-positive cells (10–12). Although these approaches have shown encouraging results in preclinical studies, those tested in clinical trials have shown limited efficacy, even in combination with chemotherapy (13–15). Furthermore, FAP-targeting radioligands have been used for *in vivo* imaging and targeted radionuclide therapy for a variety of cancers including CRC (16, 17).

Most of the functions described for FAP are associated with its enzymatic activity involved in tissue remodeling, which helps tumor cells invade the surrounding tissue, penetrate the blood vessel wall, and travel to form distant metastasis (18–21). Recent evidence suggested that FAP in CAFs could also play a critical role in regulating antitumor immune response by inducing tumor-promoting inflammation (22–24). This is particularly interesting because the majority of CRC patients are resistant to immunotherapies, especially to immune checkpoint blockades (25).

In our study, we sought to investigate FAP as a molecular marker in CRC using immunohistochemical and transcriptomic data. To investigate other potential roles of FAP in CRC, we explored its association with the clinicopathological characteristics of our in-house cohort. We further investigated its association at the mRNA level with molecular features, pathways and cell type populations in the tumor microenvironment using The Cancer Genome Atlas (TCGA) data set.

MATERIALS AND METHODS

Patients and Specimen Characteristics

One hundred primary unselected, non-consecutive CRCs treated at the University Hospital Basel between the years 2006 and 2012 were included in this study. A tissue microarray (TMA) of these 100 tumors was constructed. Briefly, tissue cylinders with a diameter of 1 mm were punched from morphologically representative areas of each donor block and brought into one recipient paraffin block (30 × 25 mm) using the TMA GrandMaster® (TMA-GM; 3D-Histech Ltd.; Sysmex AG, Horgen, Switzerland) technology. Each punch was derived from the center of the tumor in an area with no necrosis so that each TMA spot consisted of more than 50% tumor cells. For 30 cases, non-malignant adjacent mucosa was selected from

the same donor block. The study was performed in accordance with the Helsinki Declaration and approved by the ethics committee (Ethics Committee of Basel, EKBB, no. EKBB 361/12). Data were collected retrospectively in a non-stratified and non-matched manner including patient age, tumor diameter, location, pT/pN stage, grade, histologic subtype, vascular invasion, and clinical outcome. Intratumoral and peritumoral lymphocytic inflammation was evaluated using the original hematoxylin-eosin (H&E) slides of the resection specimens used as donor block. The tumor grade was categorized as low and high ($\geq$50, <50% gland formation, respectively). The clinical outcome measure of interest was overall survival time.

Immunohistochemistry

Immunohistochemistry (IHC) was performed using an anti-FAP antibody (Vitatex, Stony Brook, NY, USA; seprase/FAPα; dilution 1:100). Staining was performed on a Leica Bond III IHC staining system (Muttenz, Switzerland) using DAB as chromogen. Immunoreactivity was evaluated semiquantitatively as the proportion of positive staining in stromal cells in 10% increments, as well as the maximal staining intensity (0 = none, 1 = weak, 2 = intermediate, 3 = strong) by two experienced pathologists with expertise in gastrointestinal pathology (C.E. and L.M.T.). In terms of the percentage of FAP-positive cells, samples containing <10% of positive cells were classified as low, whereas samples containing at least 10% of positive cells were classified as high as suggested by Henry et al. (8). In terms of FAP staining intensity, samples with intensities 0 or 1 were considered low, whereas samples with intensities 2 or 3 were considered high (8). In addition, 20 cases positive for FAP on TMA were reevaluated using whole sections from formalin-fixed paraffin-embedded tissue to study FAP expression heterogeneity. FAP immunostaining was evaluated both in stroma adjacent to the invasive tumor front and within the tumor center.

Tumor–Stromal Ratio

For all tumors, the tumor–stromal ratio on 4 μm H&E-stained tissue sections was calculated as described previously (26), and the stromal percentage was estimated per 10% intervals. Tumors were divided into stroma-high (>50%) and stroma-low ($\leq$50%) groups according to their highest score.

Assessment of Tumor Budding

Tumor budding was evaluated according to the International Tumor Budding Consensus Conference (ITBCC) method (27) and was defined according to ITBCC as single tumor cells or tumor cell clusters of up to four cells. Whole H&E-stained tissue sections of the tumors were used. One pathologist (C.E.) searched all tumor slides throughout at low magnification. Densest budding area at the invasive front (hot spot) was selected by visual estimation. Tumor buds in this area were counted at 20× magnification (field area, 0.785 mm^2). Density of tumor buds was assigned into three grades: grade 1 (BD-1): 0–4 buds; grade 2 (BD-2): 5–9 buds; and grade 3 (BD-3): $\geq$10 buds.

127

Microsatellite Instability

Immunohistochemical analyses of mismatch repair proteins were performed for expression of the four mismatch repair proteins MLH1, MSH2, MSH6, and PMS2 as previously described (28). Tissue samples with tumor cells lacking nuclear staining for at least one of these proteins were considered to have a positive microsatellite instability (MSI) screening status, hereafter referred to as MSI. Negative MSI screening status based on immunohistochemical staining is hereafter referred to as microsatellite-stable.

Analysis of TCGA Data Set

FPKM gene-level expression data for TCGA colorectal carcinoma cohort (29) with 622 tumors and 51 non-tumoral tissues, defined as "solid tissue normal," were obtained from TCGA Genomics Data Commons harmonized data portal using *TCGAbiolinks* R package (30). The expression of *FAP* was compared between tumors and normal tissues using the Student *t*-test. Tumor samples were classified into FAP-high and FAP-low groups based on the threshold of mean + 3 standard deviations of normal tissues. Clinical information was obtained from the Human Protein Atlas (Pathology Atlas) (31) CRC project for 596 TCGA CRCs.

Raw read counts of the TCGA CRC Project downloaded using *TCGAbiolinks* package (30) were used for differential expression analysis using the *edgeR* package (32). Genes with low expression (<1 log-counts per million in ≥50 samples) were filtered out. Normalization was performed using the "TMM" (weighted trimmed mean) method (33), and differential expression was assessed using the quasi-likelihood *F* test. Genes with log-fold change >2 and false discovery rate (FDR) <0.05 were considered differentially expressed. Pathway enrichment analysis of the upregulated genes from the differential analysis between the *FAP*-high and *FAP*-low groups was performed using clusterProfiler package (34), which supports Gene Ontology, KEGG, and Reactome Pathways. Significantly enriched pathways were selected based on FDR <0.05. Gene set enrichment analysis (GSEA) of all analyzed genes ranked based on signed *p* value according to the direction of the log-fold change was performed using the *fgsea* package (35). Gene Ontology gene sets from MSigDB (36) were used to identify significantly upregulated/downregulated pathways. Molecular subtyping was performed using *CMScaller* package (37), and the 622 TCGA CRCs were classified into 97 CMS1, 170 CMS2, 95 CMS3, 195 CMS4 subtypes, and 65 unclassified. Cell type enrichment analysis was performed with FPKM gene expression data using *xCell* gene signatures-based method for cell types (38).

Statistical Analysis

Statistical comparisons between categorical variables were performed using χ^2 test or Fisher exact test where appropriate. Statistical comparisons between numeric variables were performed using *t*-test, Mann–Whitney *U*-test, or paired Wilcoxon test. Survival analysis was performed using the Kaplan–Meier method and log-rank test. For the TCGA cohort, stratification of *FAP* expression for overall survival analysis was performed using the *maxstat* R package (39). Univariate Cox

regression analyses were performed to investigate the association between overall survival and clinical variables. Variables significant in univariate Cox regression analyses were included in multivariate Cox regression analysis. All tests were two-sided, and $p < 0.05$ was considered statistically significant. Statistical power of statistical tests was estimated by 100 iterations of bootstrapping. Statistical analysis was carried out with Prism (v7.0; San Diego, CA, USA) and R (v.3.6.1; R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/).

RESULTS

Expression of FAP in Colorectal Cancers

We analyzed the expression of FAP at the protein level in stromal cells by staining and scoring a TMA containing 100 CRC samples and 30 non-tumoral adjacent tissues using IHC (**Figure 1A**). After excluding samples for which the tissue core was absent or had poor staining quality, 92 CRCs and 19 non-tumoral colon samples were available for analysis. We observed a significantly higher percentage of FAP-positive stromal cells in tumors compared to non-tumoral tissues ($p < 0.0001$; **Figure 1B**). FAP expression was detected in 91% (84/92) of the tumors. High frequency (at least 10%) of FAP-positive cells was found in 78% (72/92) of CRCs, whereas high FAP intensity was observed in 66% (61/92, **Figure 1** and **Supplementary Figure 1A**). Sixty-three percent (58/92) had high frequency of FAP-positive cells with elevated intensity (2 or 3). Although FAP staining was detectable in 79% (15/19) of the non-tumoral tissue samples, both the number of positive cells and the intensity were significantly lower compared to tumoral tissue (both $p < 0.0001$; **Figure 1B**). Similar results were observed when we considered the 19 matched pairs of CRCs and non-tumoral colon samples ($p = 0.005$ for frequency and $p = 0.01$ for intensity; **Supplementary Figure 1B**). Of the 15 non-tumoral tissue samples with detectable FAP expression, only a single case showed a high percentage (≥10%) of FAP-positive (intensity 2) cells.

We then investigated the association between FAP expression (both frequency and intensity) and the clinicopathological characteristics of our cohort (**Table 1**). We found that the frequency of FAP-positive cells and that of cells with high FAP intensity were significantly more frequently found in CRCs with advanced stages ($p = 0.01$ and $p < 0.05$, respectively, **Figure 1D**). Similarly, high FAP intensity was also associated with tumors showing high tumor grade and lymphovascular invasion ($p < 0.05$ and $p = 0.03$, respectively, **Table 1**). The lymphatic invasion was also associated with a high frequency of FAP-positive cells ($p = 0.09$, **Figure 1E** and **Table 1**) and a high FAP intensity ($p < 0.01$, **Figure 1E** and **Table 1**). We further observed a stepwise increase in tumor budding according to the percentage of FAP-positive cells ($p = 0.009$, **Supplementary Figure 1C**). No significant association was found with age, sex, tumor location, presence of MSI, stroma-to-tumor ratio, or venous invasion (**Table 1**).

Furthermore, we determined whether there was an association between the frequency of FAP-positive cells and overall survival. High frequency of FAP-positive cells, but not FAP

128

FIGURE 1 | FAP protein is overexpressed in CRC and is associated with worse overall survival. **(A)** Representative examples of FAP expression by immunohistochemistry. FAP expression in (I) high percentage of tumor-associated stromal cells (100%); (II) tumor-associated stroma with low number of positive cells (5%); (III) normal tissue showing negative staining for FAP; (IV) FAP high-intensity staining in tumor-associated stromal cells (high); (V) FAP low-intensity staining in tumor-associated stromal cells (low); and (VI) normal tissue showing FAP high-intensity staining in stroma cells. Scale bar: 50 μm. **(B)** Percentage of cells showing positive FAP staining in normal stroma and tumor-associated stroma (above). Scoring of FAP staining based on intensity in normal and tumor-associated stroma (below). **(C)** Overall survival analysis (Kaplan–Meier) of CRC patients from the TMA cohort after stratification for high and low percentage of FAP-positive cells. Number of CRCs with high and low percentages of FAP-positive cells (above), as well as high and low FAP intensity (below) in **(D)** early and advanced tumor stages, and **(E)** positive and negative lymphatic invasion. Statistical analyses were performed using Fisher exact test for categorical variables, Mann–Whitney U-test for numeric variables and log-rank (Mantel–Cox) test for survival analysis. All tests were two-sided, and $p < 0.05$ was considered statistically significant. Data in **(B)** are represented as mean ± SD. *Statistical power >70%.

intensity, was associated with worse overall survival ($p = 0.02$; **Figure 1C** and **Supplementary Figure 1D**). Similarly, univariate and multivariate Cox regression analyses showed that FAP expression is an independent predictor of overall survival (**Supplementary Table 1**).

To cross-validate our results, we retrieved the gene expression data of 622 CRC cases from TCGA (29). In agreement with the data obtained by IHC on the TMA (**Figure 1B**), tumor samples expressed significantly higher levels of FAP compared to normal tissues ($p < 2.2e\text{-}16$; **Figure 2A**). Association of FAP expression with clinicopathological parameters demonstrated that FAP overexpression was associated with more advanced tumor stage ($p = 0.02$, **Table 2** and **Figure 2C**). In terms of outcome, we observed a trend toward worse overall survival in patients with tumors with high FAP expression ($p = 0.06$; **Figure 2B** and **Supplementary Figure 2**). Univariate analysis found tumor stages and tumor location as predictors of overall survival, whereas FAP expression showed a trend to it ($p = 0.06$; **Supplementary Table 2**).

Taken together, our results suggest that FAP expression may be prognostic in CRC.

Transcriptomic Analysis of *FAP* Expression in Colorectal Tumors

To further understand the possible role of FAP in CRC, we analyzed the transcriptomic data from the TCGA data set ($n = 622$). We investigated the association of FAP expression with the CRC molecular subtypes, and we observed a statistically significant association with tumors classified as CMS1 and CMS4 ($p = 0.02$ and $p < 0.001$; **Figure 3A** and **Table 2**). CMS1 and CMS4 have been reported to be associated with an upregulation of immune response genes and epithelial-to-mesenchymal transition, respectively (40). In particular, the association with CMS4 suggests a more aggressive origin of these tumors characterized by FAP overexpression.

Then, we performed a differential expression analysis between CRCs with high vs. low FAP expression. The differential expression analysis revealed 655 up- and 9 downregulated genes. Consistent with previous reports (41, 42), we found that the upregulated genes were enriched in functions related to collagen degradation, extracellular organization, regulation of cell-to-cell adhesion,

129

TABLE 1 | Association between FAP protein expression and clinicopathological features in the TMA cohort.

Clinical features	Frequency			Intensity		
	High FAP expression	Low FAP expression	p	High FAP expression	Low FAP expression	p
Age (years)						
<59	9 (82%)	2 (18%)	0.64	7 (64%)	4 (36%)	0.11
60–69	16 (73%)	6 (27%)		11 (50%)	11 (50%)	
70–79	24 (86%)	4 (14%)		23 (82%)	5 (18%)	
>80	23 (74%)	8 (26%)		19 (61%)	12 (39%)	
Sex						
Male	42 (78%)	12 (22%)	1	34 (63%)	20 (37%)	0.66
Female	30 (79%)	8 (21%)		26 (68%)	12 (32%)	
Tumor location						
Cecum	14 (82%)	3 (18%)	0.96	13 (76%)	4 (24%)	0.73
Ascending colon	12 (80%)	3 (20%)		8 (53%)	7 (47%)	
Transverse colon	4 (80%)	1 (20%)		4 (80%)	1 (20%)	
Descending colon	6 (67%)	3 (33%)		5 (56%)	4 (44%)	
Sigmoid colon	20 (77%)	6 (23%)		16 (62%)	10 (38%)	
Rectum	16 (80%)	4 (20%)		13 (65%)	7 (35%)	
Stage						
I	10 (59%)	7 (41%)	0.03	11 (65%)	6 (35%)	0.14
II	19 (70%)	8 (30%)		13 (48%)	14 (52%)	
III	25 (93%)	2 (7%)		20 (74%)	7 (26%)	
IV	18 (86%)	3 (14%)		16 (76%)	5 (24%)	
Grade						
Low	51 (75%)	17 (25%)	0.26	40 (59%)	28 (41%)	<0.05
High	21 (88%)	3 (12%)		20 (83%)	4 (17%)	
Tumor: stroma*						
Stroma low	55 (77%)	16 (23%)	1	46 (65%)	25 (35%)	1
Stroma high	17 (68%)	4 (32%)		14 (66%)	7 (34%)	
Microsatellite instability*						
MSI	9 (90%)	1 (10%)	0.45	9 (90%)	1 (10%)	0.09
MSS	62 (77%)	19 (23%)		50 (62%)	31 (38%)	
Lymphatic invasion						
Positive	26 (90%)	3 (10%)	0.09	24 (83%)	5 (17%)	<0.01**
Negative	34 (71%)	14 (29%)		24 (50%)	24 (50%)	
Venous invasion*						
Positive	16 (76%)	5 (24%)	1	13 (61%)	8 (38%)	1
Negative	44 (76%)	14 (24%)		35 (60%)	23 (40%)	
Lymphovascular invasion*						
Positive	32 (84%)	6 (16%)	0.2	29 (76%)	9 (24%)	0.03
Negative	33 (72%)	13 (28%)		24 (52%)	22 (48%)	

All 2 × 2 contingency tables were analyzed with Fisher exact tests. All others by χ^2 test.
*Patients with data not available, unknown, and discrepancies.
**Statistical power estimated by bootstrapping.

and cell junction organization (**Figure 3B**). Additionally, pathways involved in epithelial cell proliferation, invasion, and immune surveillance such as regulation of Wnt signaling, ERK1, and ERK2 cascade and angiogenesis were also associated with *FAP* upregulation (**Figure 3B**). We further performed a GSEA by ranking all expressed genes based on the signed *p*-value from the differential expression analysis according to the direction of the log-fold

130

FIGURE 2 | *FAP* is upregulated in colorectal tumors. **(A)** *FAP* expression level in tumors ($n = 622$) compared to non-tumoral tissue ($n = 51$) in the TCGA data set. **(B)** Overall survival analysis (Kaplan–Meier) of CRC patients from the TCGA data set stratified by *FAP* expression (see also **Supplementary Figure 2**). **(C)** Colorectal cancer patient characteristics grouped by *FAP* expression. Statistical comparisons of clinical and molecular parameters between tumors with high vs. low *FAP* expression are shown in **Table 2**. Statistical comparisons were performed Mann–Whitney *U*-test in **(A)** and by log-rank test in **(B)**.

change (**Supplementary Figure 3**). We observed that *FAP* overexpression was associated with processes related to epithelial-to-mesenchymal transition, angiogenesis, tissue remodeling, epithelial cell proliferation, and immune response (**Supplementary Figures 3, 4**).

Because we observed that *FAP* overexpression was associated with immune-related processes (**Figure 3C** and **Supplementary Figure 3**), we sought to explore which immune components may be involved. We observed that CRCs with high *FAP* expression showed high expression of genes such as *FOXP3*, *CTL4*, *ICOS*, and *KLRG1* (**Figure 3C**), which are usually expressed in immune cell populations such as regulatory T cells (Tregs) or in populations showing an immunosuppressive phenotype. To evaluate whether *FAP* overexpression was associated with the enrichment of specific immune cell populations, we used xCell, a method to perform cell type enrichment analysis from gene expression data for 64 immune and stromal cell types, on the TCGA CRC cohort. Overall, we found that tumors with high *FAP* expression were enriched for both immune and stromal cell types compared to tumors with low *FAP* expression (**Figure 3D** and **Supplementary Figure 5**). Populations such as endothelial cells and fibroblasts were found to be more abundant in *FAP*-overexpressing CRC ($p = 3.1e\text{-}13$ and $p = 6.9e\text{-}11$, respectively; **Figures 3D,E**). Similarly, macrophages, monocytes, and Tregs were also enriched in *FAP*-overexpressing samples. Interestingly, *FAP*-overexpressing CRCs were depleted for populations associated with antitumoral responses such as

T_H1 cells and natural killer T (NKT) cells. No significant differences were found regarding CD8 T cells, T_H2 cells, or B cells.

Taken together, our results suggest that FAP may contribute to the poor prognosis of CRC by modulating the tumor microenvironment not only by driving angiogenesis but also by promoting a more protumorigenic environment.

FAP Distribution in the Tumor Center and Invasive Tumor Front

Given that the transcriptomic analysis suggests that FAP may promote a protumorigenic environment, we investigated whether the localization of FAP in the tumor could be associated with its role in CRC invasion. We took advantage of 20 CRCs that showed positive FAP staining in the TMA and stained whole tissue sections to define FAP heterogeneity. Using whole sections, we were able to visualize the tumor center and the invasive front in 95% (19/20) of the samples. FAP staining was visible in both tumor center and tumor front in all 19 cases (**Figure 4A**). Compared to the tumor center, FAP-positive cells were more frequently found at the invasive front ($p = 0.03$; **Figure 4B**). The high frequency of FAP-positive cells at the invasive front was associated with advanced tumor stage ($p = 0.03$; **Figure 4C**; right), but no significant association was found between the frequency of FAP-positive cells in the tumor center and tumor stage (**Figure 4C**; left). The presence of lymphovascular invasion was significantly associated with high

131

TABLE 2 | Association between *FAP* mRNA expression and clinicopathological features in the TCGA cohort.

Clinical features		Low *FAP* expression	High *FAP* expression	*p*
		n (%)	n (%)	
Gender (*n* = 591)	Female (*n* = 271)	133 (49.1%)	138 (50.9%)	0.81
	Male (*n* = 320)	154 (48.1%)	166 (51.9%)	
AJCC stages (*n* = 571)*	Stage I + II (*n* = 316)	167 (52.8%)	149 (47.2%)	0.02
	Stage III + IV (*n* = 255)	111 (43.5%)	144 (56.5%)	
Microsatellite instability (*n* = 115)*	MSI (*n* = 11)	7 (63.6%)	4 (36.4%)	0.42
	MSS (*n* = 104)	53 (51.0%)	51 (49.0%)	
CRC subtyping (*n* = 556)*	CMS1 (*n* = 97)	36 (37.1%)	61 (62.9%)	<0.001**
	CMS2 (*n* = 170)	129 (75.9%)	41 (24.1%)	
	CMS3 (*n* = 94)	66 (70.2%)	28 (29.8%)	
	CMS4 (*n* = 195)	17 (8.7%)	178 (91.3%)	
CRC location (*n* = 597)*	Cecum (*n* = *n* = 106)	56 (52.8%)	50 (47.2%)	0.4
	Ascending colon (*n* = 86)	31 (36.0%)	55 (64.0%)	
	Descending colon (*n* = 20)	9 (45.0%)	11 (55.0%)	
	Transverse colon (*n* = 38)	20 (52.6%)	18 (47.4%)	
	Sigmoid colon (*n* = 155)	75 (48.4%)	80 (51.6%)	
	Hepatic flexure (*n* = 26)	12 (46.2%)	14 (53.8%)	
	Splenic flexure (*n* = 7)	4 (57.0%)	3 (43.0%)	
	Rectosigmoid junction (*n* = 49)	24 (49.0%)	25 (51.0%)	
	Rectum (*n* = 110)	60 (55.0%)	50 (45.0%)	

Patients with data not available, unknown, and discrepancies.
**Statistical power >70% (estimated by bootstrapping).*
All 2 × 2 contingency tables were analyzed with Fisher exact tests. All others by χ^2 test.

FIGURE 3 | *FAP* is associated with pathways involved in tumor growth, invasion, and immunosuppressive tumor microenvironment. **(A)** Colorectal cancers grouped by *FAP* expression. Statistical comparison of molecular subtypes between tumors with high vs. low *FAP* expression are shown in **Table 2**. **(B)** Figure shows selected significantly enriched pathways from pathway enrichment analysis of the upregulated genes from a differential expression analysis between *FAP*-high vs. *FAP*-low. The size of the dots indicates the number of upregulated genes in each pathway. The color of the dot indicates FDR, and *x* axis represents the fraction of upregulated genes in the pathway. **(C)** Gene set enrichment analysis plots of GO immune response, where *x* axis shows ranked list of genes (ranked by the *p*-values signed according to the direction of the differential expression analysis between *FAP*-high and *FAP*-low CRCs), and the vertical bars on the *x* axis show the genes that belong to gene set. The *y* axis shows the enrichment score of the gene set. Heatmaps below show selected genes in the GO process. **(D)** Representative pictures showing the enrichment of fibroblast on high *FAP* expression (left) and low number of fibroblasts on low *FAP* expression (right). **(E)** Heatmap shows the enrichment of immune and stromal cell types in the TCGA CRC cohort, as defined by the xCell method. Samples were stratified into *FAP*-high and *FAP*-low groups and then sorted based on their xCell fibroblasts scores. Cell types that showed statistically significant difference between *FAP*-high and *FAP*-low groups are shown. Statistical comparisons were performed using Mann–Whitney *U*-tests.

132

FIGURE 4 | FAP expression is enriched in the tumor front area in colorectal tumor. **(A)** Representative micrographs of FAP distribution in the tumor center and at the tumor front area. (I) FAP expression is higher at the tumor front compared to the tumor center; (II) FAP expression does not show significant difference between tumor center and tumor front, whereas neighbor non-neoplastic colonic mucosal stroma is completely negative. Scale bar: 200 μm. **(B)** Percentage of cells showing FAP-positive staining in the tumor front and tumor center areas ($n = 19$). **(C)** Percentage of cells showing FAP-positive staining in tumor center (left) and front (right) in early and advanced tumor stage. **(D)** Percentage of cells showing FAP-positive staining in tumor center (left) and front (right) with lymphovascular invasion. Statistical analyses were performed using Fisher exact tests for categorical variables and Mann–Whitney U-tests for numeric variables. All tests were two-sided, and $p < 0.05$ was considered statistically significant. Data in **(B–D)** are represented as mean ± SD.

frequency of FAP-positive cells in tumor center (**Figure 4D**, $p = 0.04$) but not in tumor front.

These results are in line with the results we obtained by TMA, underlying the important role of FAP in tumor invasion.

DISCUSSION

FAP has been shown to be overexpressed in tumor-associated stromal cells in epithelial tumors (43) and its presence has been associated with worse prognosis. Moreover, tumors showing upregulation of FAP present a high level of microvessel density (44), which is also a marker for poor prognosis in several epithelial cancers (45–47). *In vivo* studies have shown that FAP overexpression in breast and colonic xenograft models leads to more rapid development of subcutaneous tumors and enhanced tumor growth (44). By contrast, inhibition of FAP activity in colorectal xenograft models results in tumor growth attenuation (48). Together, there is substantial evidence supporting the role of FAP in tumor proliferation and metastasis (49). Moreover, little is known about the molecular role of FAP in CRC and its potential in modulating the tumor microenvironment. This is particularly important because immune checkpoint inhibitors have demonstrated little or no clinical activity in the majority of patients with metastatic CRC (50).

In the present study, we investigated the expression of FAP both at the RNA and protein levels in two independent cohorts of CRC and its association with clinicopathological features. Similarly to previous reports evaluating FAP expression using IHC (8, 9, 51, 52), FAP was found upregulated in tumors compared to non-tumoral tissues and was associated with poor survival in both cohorts. In line with the poor prognosis, we found that high frequency of FAP-positive cells and high FAP intensity at the protein and RNA levels were associated with advanced stages. Although our results are discrepant from previous works showing FAP expression was also elevated in early-stage CRC (8, 51), we also found that FAP expression was associated with lymphatic invasion and tumor budding. Lymphatic invasion has been used to estimate the aggressiveness of colorectal tumors (53), whereas tumor budding is a surrogate for epithelial-to-mesenchymal transition and is associated with poor prognosis (54). The findings related to pathological features are thus in agreement with the association with advanced stages. Taken together, the association of FAP expression on the mRNA and protein levels with multiple clinicopathological features known to be associated with poor prognosis supports our finding that FAP is also associated with tumor aggressiveness.

Most studies on FAP have focused on its potential value as a prognostic marker in epithelial cancers, but little is known on

133

how and why FAP may be prognostic. We, therefore, analyzed the transcriptomic data from 622 CRCs in the TCGA. We found that the majority of tumors with high *FAP* expression were classified as CMS1 and CMS4 of the consensus molecular subtypes. CMS4 tumors have been reported to overexpress proteins implicated in stromal invasion, mesenchymal activation, and complement pathways. The enrichment for CMS4 tumors among tumors with high *FAP* expression suggests a more invasive tumor phenotype, which agrees with our findings from the TMA. In fact, our analysis of the TCGA transcriptomic data also showed that CRCs with *FAP* overexpression were enriched for stromal cell types, in particular for fibroblasts and mesenchymal stem cells. Furthermore, most of the pathways associated with *FAP* overexpression were related to extracellular matrix, junction remodeling, and collagen degradation, all of which have previously been associated with FAP expression. We further found that epithelial-to-mesenchymal transition, angiogenesis, and epithelial cell proliferation were also enriched in *FAP*-overexpressing CRCs. Our results are in line with previous studies that have demonstrated the involvement of FAP in epithelial-to-mesenchymal transition, angiogenesis, and in tumorigenesis (44, 49, 55). Additionally, our IHC analysis showed that the frequency of FAP-positive cells was associated with tumor budding score, a very well-known pathological marker associated with epithelial-to-mesenchymal transition and tumor invasion. Overall, the findings from our analysis of the TCGA are further supported by the observation that FAP was preferentially found at the tumor invasive front, in agreement with Sandberg et al. (56), suggesting a role in tissue invasion and metastasis.

In addition to CMS4, we also found that *FAP*-overexpressing CRCs were enriched in the subtype CMS1. While the most widely described function of FAP on CAFs in CRC was extracellular matrix remodeling (57), new evidence suggests that FAP on CAFs also has critical roles in regulating antitumor immune response by inducing tumor-promoting inflammation (58). Indeed, the CMS1 subtype displays upregulation of proteins involved in immune response pathways (40). We also found that *FAP* overexpression is associated with an upregulation of genes involved in immune cell response, suggesting that FAP may promote an inflamed environment. Recent reports have proposed that FAP expression may be associated with resistance to immune therapies (22, 59). Using the xCell algorithm, we found that populations such as T_H1 and NKT cells were suppressed, whereas Tregs were enriched in *FAP*-overexpressing CRCs, suggesting an immunosuppressive environment in these tumors. Chen et al. (23) reported that FAP expression promotes immunosuppression in a CRC tumor model via the upregulation of CCL2. CCL2, a member of the C-C chemokine family, regulates the recruitment of myeloid cells, mostly macrophages and monocytes (60), into inflamed sites to promote tumor growth (61, 62). In breast cancer, Costa et al. (63) showed that CAF with high expression of FAP was associated with increased $CD25^+FOXP3^+$ T lymphocytes via the modulation of B7H3, DPP4, and CD73. Accordingly, we also observed that macrophage and monocyte populations are enriched in *FAP*-overexpressing CRCs.

Although the statistical power of our study is limited by the relatively small sample size of the TMA cohort, our study is still one of the largest cohorts evaluating FAP expression in CRCs. Our results support the well-known role of FAP in promoting tumor growth and invasion. Based on our results, we speculate that one of the mechanisms by which FAP promotes tumorigenesis is linked to its ability to recruit endothelial cells and to induce angiogenesis, together with its enzymatic activity. Moreover, FAP will orchestrate a broad panel of other cells to push microenvironment toward an immunosuppressive environment, thus providing a niche for a more aggressive CRC phenotype.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA data portal (https://portal.gdc.cancer.gov/).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Basel, EKBB, number EKBB 361/12. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MC-L, LT, MF, and SP: conceived the study. MC-L and SP: supervised the study. VK: performed bioinformatic analysis. CE and LT: performed the histologic review and immunohistochemical evaluation. MC-L, CE, ST-M, and SP: analyzed the results. SS, MB, MvF, GN, and MF: carefully discussed the results. MC-L, CE, and VK: wrote the manuscript that was revised by CN and SP. All authors edited and approved the final draft of the manuscript and are accountable for all aspects of the work.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00979/full#supplementary-material

Supplementary Figure 1 | (A) Table displays the distribution of frequency of FAP-positive cells and FAP intensity in the TMA cohort. **(B)** Frequency (left) and intensity (right) of FAP-positive cells in 19 non-malignant tissues and paired malignant tissues. **(C)** Percentage of cells showing FAP-positive staining (left) or intensity (right) associated with tumor budding score. **(D)** Overall survival analysis (Kaplan–Meier) of CRC patients from the TMA cohort after stratification (below) for tumors with high FAP intensity vs tumors with low FAP intensity and (above) for tumors with high frequency of FAP-positive cells and/or high FAP intensity vs tumors with low frequency of FAP-positive cells and low FAP intensity. Data in **(B,C)** are represented as mean $\pm$ SD Statistical analyses were performed using Fisher exact tests for categorical variables in **(B)**, ANOVA in **(C)** and log-rank (Mantel–Cox) tests for survival analysis in **(D)**. All tests were two-sided, and $p <$ 0.05 was considered statistically significant. *Statistical power >70%.

Supplementary Figure 2 | The plot shows the optimal cutpoint for the stratification of TCGA CRC cohort into two groups, *FAP*-high and *FAP*-low, for overall survival analysis.

Supplementary Figure 3 | (A) Gene Set Enrichment Analysis plot shows selected significantly enriched Gene Ontology gene sets from MSigDB database

in *FAP*-high tumors in the TCGA CRC cohort. **(B)** Gene Set Enrichment Analysis plots of Gene Ontology molecular functions, where x-axis shows ranked list of genes (ranked by the *p*-values signed according to the direction of the differential expression analysis between *FAP*-high and *FAP*-low CRCs) and the vertical bars on the x-axis show the genes that belong to gene set. The y-axis shows the enrichment score of the gene set. Heatmaps below show selected genes in each GO process. Pearson correlation coefficients of the expression of the genes with *FAP* expression are shown to the right. NES: normalized enrichment score. P adj, *p*-value adjusted for multiple testing (i.e., FDR).

Supplementary Figure 4 | Correlation of *FAP* expression with **(A)** immune markers; **(B)** epithelial to mesenchymal transition (EMT) markers and **(C)** angiogenesis markers. Statistical analyses were performed using Pearson correlation tests.

Supplementary Figure 5 | Boxplots show the enrichment scores of each cell type between *FAP*-high and *FAP*-low groups in the TCGA CRC cohort, as defined by xCell. p-values were calculated using t-tests.

Supplementary Table 1 | Univariate and multivariate Cox regression analyses for the effect of FAP protein expression and clinicopathological parameters on overall survival in the TMA cohort.

Supplementary Table 2 | Univariate and multivariate Cox regression analyses for the effect of *FAP* mRNA expression and clinicopathological parameters on overall survival in the TCGA cohort.

REFERENCES

1. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the "Immunoscore" in the classification of malignant tumours. *J Pathol.* (2014) 232:199–209. doi: 10.1002/path.4287
2. Park JH, Richards CH, McMillan DC, Horgan PG, Roxburgh CSD. The relationship between tumour stroma percentage, the tumour microenvironment and survival in patients with primary operable colorectal cancer. *Ann Oncol.* (2014) 25:644–51. doi: 10.1093/annonc/mdt593
3. Eriksen AC, Sørensen FB, Lindebjerg J, Hager H, Christensen RD, Kjær-Frifeldt S, et al. The prognostic value of tumour stroma ratio and tumour budding in stage II colon cancer. A nationwide population-based study *Int J Colorectal Dis.* (2018) 33:1115–24. doi: 10.1007/s00384-018-3076-9
4. Kuzet S-E, Gaggioli C. Fibroblast activation in cancer: when seed fertilizes soil. *Cell Tissue Res.* (2016) 365:607–19. doi: 10.1007/s00441-016-2467-x
5. Koliaraki V, Pallangyo CK, Greten FR, Kollias G. Mesenchymal cells in colon cancer. *Gastroenterology.* (2017) 152:964–79. doi: 10.1053/j.gastro.2016.11.049
6. Niedermeyer J, Garin-Chesa P, Kriz M, Hilberg F, Mueller E, Bamberger U, et al. Expression of the fibroblast activation protein during mouse embryo development. *Int J Dev Biol.* (2001) 45:445–7.
7. Jacob M, Chang L, Pure E. Fibroblast activation protein in remodeling tissues. *Curr Mol Med.* (2012) 12:1220–43. doi: 10.2174/156652412803833607
8. Henry LR Lee H-O, Lee JS, Klein-Szanto A, Watts P, Ross EA, et al. Clinical implications of fibroblast activation protein in patients with colon cancer. *Clin Cancer Res.* (2007) 13:1736–41. doi: 10.1158/1078-0432.CCR-06-1746
9. Saigusa S, Toiyama Y, Tanaka K, Yokoe T, Okugawa Y, Fujikawa H, et al. Cancer-associated fibroblasts correlate with poor prognosis in rectal cancer after chemoradiotherapy. *Int J Oncol.* (2011) 38:655–63. doi: 10.3892/ijo.2011.906
10. Lee J, Fassnacht M, Nair S, Boczkowski D, Gilboa E. Tumor immunotherapy targeting fibroblast activation protein, a product expressed in tumor-associated fibroblasts. *Cancer Res.* (2005) 65:11156–63. doi: 10.1158/0008-5472.CAN-05-2805
11. Ostermann E, Garin-Chesa P, Heider KH, Kalat M, Lamche H, Puri C, et al. Effective immunoconjugate therapy in cancer models targeting a serine protease of tumor fibroblasts. *Clin Cancer Res.* (2008) 14:4584–92. doi: 10.1158/1078-0432.CCR-07-5211

12. Loeffler M, Krüger JA, Niethammer AG, Reisfeld RA. Targeting tumor-associated fibroblasts improves cancer chemotherapy by increasing intratumoral drug uptake. *J Clin Investig.* (2009) 119:421. doi: 10.1172/JCI26532C1
13. Narra K, Mullins SR, Lee H-O, Strzemkowski-Brun B, Magalong K, Christiansen VJ, et al. Phase II trial of single agent Val-boroPro (talabostat) inhibiting fibroblast activation protein in patients with metastatic colorectal cancer. *Cancer Biol Therap.* (2007) 6:1691–9. doi: 10.4161/cbt.6.11.4874
14. Eager RM, Casey Cunningham C, Senzer NN, Stephenson J, Anthony SP, O'Day SJ, et al. Phase II assessment of talabostat and cisplatin in second-line stage IV melanoma. *BMC Cancer.* (2009) 9:263. doi: 10.1186/1471-2407-9-263
15. Eager RM, Cunningham CC, Senzer N, Richards DA, Raju RN, Jones B, et al. Phase II trial of talabostat and docetaxel in advanced non-small cell lung cancer. *Clin Oncol.* (2009) 21:464–72. doi: 10.1016/j.clon.2009.04.007
16. Giesel FL, Kratochwil C, Lindner T, Marschalek MM, Loktev A, Lehnert W, et al. 68Ga-FAPI PET/CT: biodistribution and preliminary dosimetry estimate of 2 DOTA-containing FAP-targeting agents in patients with various cancers. *J Nucl Med.* (2019) 60:386–92. doi: 10.2967/jnumed.118.215913
17. Kratochwil C, Flechsig P, Lindner T, Abderrahim L, Altmann A, Mier W, et al. 68Ga-FAPI PET/CT: tracer uptake in 28 different kinds of cancer. *J Nuclear Med.* (2019) 60:801–5. doi: 10.2967/jnumed.119.227967
18. Kelly T. Fibroblast activation protein-alpha and dipeptidyl peptidase IV (CD26): cell-surface proteases that activate cell signaling and are potential targets for cancer therapy. *Drug Resist Updat.* (2005) 8:51–8. doi: 10.1016/j.drup.2005.03.002
19. Christiansen VJ, Jackson KW, Lee KN, McKee PA. Effect of fibroblast activation protein and α2-antiplasmin cleaving enzyme on collagen Types I, III, and IV. *Arch Biochem Biophys.* (2007) 457:177–86. doi: 10.1016/j.abb.2006.11.006
20. Henriksson ML, Edin S, Dahlin AM, Oldenborg P-A, Öberg Å, Van Guelpen B, et al. Colorectal cancer cells activate adjacent fibroblasts resulting in FGF1/FGFR3 signaling and increased invasion. *Am J Pathol.* (2011) 178:1387–94. doi: 10.1016/j.ajpath.2010.12.008
21. Higashino N, Koma Y-I, Hosono M, Takase N, Okamoto M, Kodaira H, et al. Fibroblast activation protein-positive fibroblasts promote tumor progression through secretion of CCL2 and interleukin-6 in esophageal squamous cell carcinoma. *Lab Invest.* (2019) 99:777–92. doi: 10.1038/s41374-018-0185-6

135

22. Yang X, Lin Y, Shi Y, Li B, Liu W, Yin W, et al. FAP promotes immunosuppression by cancer-associated fibroblasts in the tumor microenvironment via STAT3–CCL2 signaling. *Cancer Res.* (2016) 76:4124–35. doi: 10.1158/0008-5472.CAN-15-2973

23. Chen L, Qiu X, Wang X, He J. FAP positive fibroblasts induce immune checkpoint blockade resistance in colorectal cancer via promoting immunosuppression. *Biochem Biophys Res Commun.* (2017) 487:8–14. doi: 10.1016/j.bbrc.2017.03.039

24. Xia Q, Geng F, Zhang F-F, Liu C-L, Xu P, Lu Z-Z, et al. Cyclophosphamide enhances anti-tumor effects of a fibroblast activation protein α-based DNA vaccine in tumor-bearing mice with murine breast carcinoma. *Immunopharmacol Immunotoxicol.* (2017) 39:37–44. doi: 10.1080/08923973.2016.1269337

25. Singh PP, Sharma PK, Krishnan G, Craig Lockhart A. Immune checkpoints and immunotherapy for colorectal cancer. *Gastroenterol Rep.* (2015) 3:289–97. doi: 10.1093/gastro/gov053

26. Mesker WE, Junggeburt JMC, Szuhai K, de Heer P, Morreau H, Tanke HJ, et al. The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. *Cell Oncol.* (2007) 29:387–98. doi: 10.1155/2007/175276

27. Lugli A, Kirsch R, Ajioka Y, Bosman F, Cathomas G, Dawson H, et al. Recommendations for reporting tumor budding in colorectal cancer based on the International Tumor Budding Consensus Conference (ITBCC) 2016. *Mod Pathol.* (2017) 30:1299–311. doi: 10.1038/modpathol. 2017.46

28. Kishore S, Piscuoglio S, Kovac MB, Gylling A, Wenzel F. 3′-UTR poly (T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair–deficient cancers. *Cancer Res.* (2014) 74:224–34. doi: 10.1158/0008-5472.CAN-13-2100

29. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* (2012) 487:330–7. doi: 10.1038/nature11252

30. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* (2016) 44:e71. doi: 10.1093/nar/gkv1507

31. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. *Science.* (2017) 357:eaan2507. doi: 10.1126/science.aan2507

32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

33. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* (2010) 11:R25. doi: 10.1186/gb-2010-11-3-r25

34. Yu G. clusterProfiler: An universal enrichment tool for functional and comparative study. *BioRxiv.* (2018) BioRxiv:256784. doi: 10.1101/256784

35. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv.* (2016) BioRxiv:060012. doi: 10.1101/060012

36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

37. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci Rep.* (2017) 7:16618. doi: 10.1038/s41598-017-16747-x

38. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* (2017) 18:220. doi: 10.1186/s13059-017-1349-1

39. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. *Comput Statist Data Anal.* (2003) 43:121–37. doi: 10.1016/S0167-9473(02)00225-6

40. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med.* (2015) 21:1350–6. doi: 10.1038/nm.3967

41. Santos AM, Jung J, Aziz N, Kissil JL, Puré E. Targeting fibroblast activation protein inhibits tumor stromagenesis growth in mice. *J Clin Invest.* (2009) 119:3613–25. doi: 10.1172/JCI38988

42. Lee H-O, Mullins SR, Franco-Barraza J, Valianou M, Cukierman E, Cheng JD. FAP-overexpressing fibroblasts produce an extracellular matrix that enhances invasive velocity and directionality of pancreatic cancer cells. *BMC Cancer.* (2011) 11:245. doi: 10.1186/1471-2407-11-245

43. Garin-Chesa P, Old LJ, Rettig WJ. Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody target in human epithelial cancers. *Proc Natl Acad Sci USA.* (1990) 87:7235–9. doi: 10.1073/pnas.87. 18.7235

44. Huang Y. Seprase promotes rapid tumor growth and increased microvessel density in a mouse model of human breast cancer. *Cancer Res.* (2004) 64:2712–6. doi: 10.1158/0008-5472.CAN-03-3184

45. Hartenbach EM, Olson TA, Goswitz JJ, Mohanraj D, Twiggs LB, Carson LF, et al. Vascular endothelial growth factor (VEGF) expression and survival in human epithelial ovarian carcinomas. *Cancer Lett.* (1997) 121:169–75. doi: 10.1016/S0304-3835(97)00350-9

46. Joo HJ, Oh DK, Kim YS, Lee KB, Kim SJ. Increased expression of caveolin-1 and microvessel density correlates with metastasis and poor prognosis in clear cell renal cell carcinoma. *BJU Int.* (2004) 93:291–6. doi: 10.1111/j.1464-410X.2004.04604.x

47. Guetz GD, Des Guetz G, Uzzan B, Nicolas P, Cucherat M, Morere J-F, et al. Microvessel density and VEGF expression are prognostic factors in colorectal cancer. Meta-analysis of the literature. *Br J Cancer.* (2006) 94:1823–32. doi: 10.1038/sj.bjc.6603176

48. Cheng JD, Dunbrack RL, Valianou M, Rogatko A, Alpaugh RK, Weiner LM. Promotion of tumor growth by murine fibroblast activation protein, a serine protease, in an animal model. *Cancer Res.* (2002) 62:4767–72.

49. Koczorowska MM, Tholen S, Bucher F, Lutz L, Kizhakkedathu JN, De Wever O, et al. Fibroblast activation protein-α, a stromal cell surface protease, shapes key features of cancer associated fibroblasts through proteome and degradome alterations. *Mol Oncol.* (2016) 10:40–58. doi: 10.1016/j.molonc.2015. 08.001

50. Le DT, Uram JN, Wang H, Bartlett B, Kemberling H, Eyring A, et al. PD-1 blockade in tumors with mismatch repair deficiency. *J Clin Oncol.* (2015) 33:LBA100. doi: 10.1200/jco.2015.33.18_suppl.lba100

51. Wikberg ML, Edin S, Lundberg IV, Van Guelpen B, Dahlin AM, Rutegård J, et al. High intratumoral expression of fibroblast activation protein (FAP) in colon cancer is associated with poorer patient prognosis. *Tumour Biol.* (2013) 34:1013–20. doi: 10.1007/s13277-012-0638-2

52. Son GM, Kwon M-S, Shin D-H, Shin N, Ryu D, Kang C-D. Comparisons of cancer-associated fibroblasts in the intratumoral stroma and invasive front in colorectal cancer. *Medicine.* (2019) 98:e15164. doi: 10.1097/MD.0000000000015164

53. Akagi Y, Adachi Y, Ohchi T, Kinugasa T, Shirouzu K. Prognostic impact of lymphatic invasion of colorectal cancer: a single-center analysis of 1,616 patients over 24 years. *Anticancer Res.* (2013) 33:2965–70.

54. Koelzer VH, Zlobec I, Lugli A. Tumor budding in colorectal cancer—ready for diagnostic practice? *Human Pathol.* (2016) 47:4–19. doi: 10.1016/j.humpath.2015.08.007

55. Wu Q-Q, Zhao M, Huang G-Z, Zheng Z-N, Chen Y, Zeng W-S, et al. Fibroblast activation protein (FAP) overexpression induces epithelial–mesenchymal transition (EMT) in oral squamous cell carcinoma by down-regulating dipeptidyl peptidase 9 (DPP9). *OncoTargets Therap.* (2020) 13:2599–611. doi: 10.2147/OTT.S243417

56. Sandberg TP, Maaike PM, Oosting J, Rob AE, Sier CFM, Mesker WE. Increased expression of cancer-associated fibroblast markers at the invasive front and its association with tumor-stroma ratio in colorectal cancer. *BMC Cancer.* (2019) 19:284. doi: 10.1186/s12885-019-5462-2

57. Park JE, Lenter MC, Zimmermann RN, Garin-Chesa P, Old LJ, Rettig WJ. Fibroblast activation protein, a dual specificity serine protease expressed in reactive human tumor stromal fibroblasts. *J Biol Chem.* (1999) 274:36505–12. doi: 10.1074/jbc.274.51.36505

58. Fearon DT. The carcinoma-associated fibroblast expressing fibroblast activation protein and escape from immune surveillance. *Cancer Immunol Res.* (2014) 2:187–93. doi: 10.1158/2326-6066.CIR-14-0002

59. Wen X, He X, Jiao F, Wang C, Sun Y, Ren X, et al. Fibroblast activation protein-α-positive fibroblasts promote gastric cancer progression and resistance to immune checkpoint blockade. *Oncol Res Featur Preclini Clin Cancer Therap.* (2017) 25:629–40. doi: 10.3727/096504016X14768383625385

136

60. Charo IF, Taubman MB. Chemokines in the pathogenesis of vascular disease. *Circul Res.* (2004) 95:858–66. doi: 10.1161/01.RES.0000146672.10582.17

61. Steiner JL, Angela Murphy E. Importance of chemokine (CC-motif) ligand 2 in breast cancer. *Int J Biol Mark.* (2012) 27:179–85. doi: 10.5301/JBM.2012.9345

62. Chun E, Lavoie S, Michaud M, Gallini CA, Kim J, Soucy G, et al. CCL2 promotes colorectal carcinogenesis by enhancing polymorphonuclear myeloid-derived suppressor cell population and function. *Cell Rep.* (2015) 12:244–57. doi: 10.1016/j.celrep.2015.06.024

63. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. *Cancer Cell.* (2018) 33:463–79.e10. doi: 10.1016/j.ccell.2018.01.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

137

Research Paper

Adenylosuccinate lyase is oncogenic in colorectal cancer by causing mitochondrial dysfunction and independent activation of NRF2 and mTOR-MYC-axis

Stephanie Taha-Mehlitz[1,2*], Gaia Bianco[1*], Mairene Coto-Llerena[1,3*], Venkatesh Kancherla[3], Glenn R. Bantug[4], John Gallon[1], Caner Ercan[3], Federica Panebianco[1], Serenella Eppenberger-Castori[3], Marco von Strauss[2], Sebastian Staubli[2], Martin Bolli[2], Ralph Peterli[2], Matthias S. Matter[3], Luigi M. Terracciano[5,6], Markus von Flüe[2], Charlotte K.Y. Ng[7], Savas D Soysal[2], Otto Kollmar[2], Salvatore Piscuoglio[1,3✉]

1. Visceral Surgery and Precision Medicine Research Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland.
2. Clarunis, Department of Visceral Surgery, University Center for Gastrointestinal and Liver Diseases, St. Clara Hospital and University Hospital Basel, Switzerland.
3. Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland.
4. Department of Biomedicine, Immunobiology, University of Basel, Basel, Switzerland.
5. Department of Pathology, Humanitas Clinical and Research Center, IRCCS, Milan, Italy.
6. Humanitas University, Department of Biomedical Sciences, Milan, Italy.
7. Department for BioMedical Research, University of Bern, Bern, Switzerland.

*These authors equally contributed to this work.

✉ Corresponding author: Dr. Salvatore Piscuoglio, Visceral Surgery and Precision Medicine Research Laboratory, Department of Biomedicine, Hebelstrasse 20, 4031 Basel, Switzerland. Tel: +41613286874; Fax: +41612653194. E-mail: s.piscuoglio@unibas.ch.

Abstract

Rationale: Adenylosuccinate lyase (ADSL) is an essential enzyme for *de novo* purine biosynthesis. Here we sought to investigate the putative role of ADSL in colorectal carcinoma (CRC) carcinogenesis and response to antimetabolites.

Methods: ADSL expression levels were assessed by immunohistochemistry or retrieved from The Cancer Genome Atlas (TCGA) dataset. The effects of ADSL silencing or overexpression were evaluated on CRC cell proliferation, cell migration and cell-cycle. *In vivo* tumor growth was assessed by the chicken chorioallantoic membrane (CAM). Transfected cell lines or patient-derived organoids (PDO) were treated with 5-fluorouracil (5-FU) and 6-mercaptopurine (6-MP) and drug response was correlated with ADSL expression levels. Metabolomic and transcriptomic profiling were performed to identify dysregulated pathways and ADSL downstream effectors. Mitochondrial respiration and glycolytic capacity were measured using Seahorse; mitochondrial membrane potential and the accumulation of ROS were measured by FACS using MitoTracker Red and MitoSOX staining, respectively. Activation of canonical pathways was assessed by immunohistochemistry and immunoblotting.

Results: ADSL expression is significantly increased in CRC tumors compared to non-tumor tissue. ADSL-high CRCs show upregulation of genes involved in DNA synthesis, DNA repair and cell cycle. Accordingly, ADSL overexpression accelerated progression through the cell cycle and significantly increased proliferation and migration in CRC cell lines. Additionally, ADSL expression increased tumor growth *in vivo* and sensitized CRCs to 6-MP *in vitro*, *ex vivo* (PDOs) and *in vivo* (CAM model). ADSL exerts its oncogenic function by affecting mitochondrial function via alteration of the TCA cycle and impairment of mitochondrial respiration. The KEAP1-NRF2 and mTORC1-cMyc axis are independently activated upon ADSL overexpression and may favor the survival and proliferation of ROS-accumulating cells, favoring DNA damage and tumorigenesis.

Conclusions: Our results suggest that ADSL is a novel oncogene in CRC, modulating mitochondrial function, metabolism and oxidative stress, thus promoting cell cycle progression, proliferation and migration. Our results also suggest that ADSL is a predictive biomarker of response to 6-mercaptopurine in the pre-clinical setting.

Key words: colorectal cancer, ADSL, mitochondria, fumarate, mTOR-MYC-axis

138

Introduction

Glucose and nucleotide metabolism are altered in cancer to promote cell reprogramming and survival under stress conditions [1]. *De novo* purine nucleotide synthesis, in particular, has been shown to play a critical role in cancer cell growth under nutrient-limited conditions, a common feature of the tumor microenvironment [2,3]. Adenylosuccinate lyase (ADSL) is an essential enzyme for *de novo* purine biosynthesis and the purine nucleotide cycle [4,5], where it plays a key role in the regulation of adenosine monophosphate levels in the cells. Given its crucial role in both cellular replication and metabolism, it is not surprising that ADSL has been found dysregulated in several malignancies, including breast [6] and endometrial [6,7] cancers as well as glioma [8]. In particular, *ADSL* expression has been shown to be significantly upregulated and to be oncogenic in triple-negative breast cancer by enhancing tumor growth and invasiveness, at least partially through the regulation of the long non-coding RNA *MIR22HG* and, indirectly, *cMYC* and *cMYC* target genes expression [6]. Similarly, *ADSL* has been shown to enhance the aggressiveness of endometrial cancer cells by increasing cell proliferation and migration, as well as invasive potential via the regulation of the natural killer cells lectin-like receptor C3 (KLRC3) expression by fumarate [7].

The use of antimetabolites is a common strategy in the treatment of cancer [9,10]. Antimetabolites are small molecules that resemble nucleotide metabolites and act by inhibiting the activity of enzymes involved in DNA synthesis [10]. Fluorouracil (5-FU), a pyrimidine analogue [9,11], is a notable example and is the most commonly used constituent of chemotherapy combination regimens to treat colorectal cancers (CRC) [9,12]. However, almost half of CRC patients develop resistance to 5-FU-based chemotherapy [13,14]. A previous study found that *ADSL* activity was increased in pre-neoplastic colonic lesions [15], but a comprehensive assessment of ADSL in CRC has not been performed. In this study, we sought to define the putative role of ADSL in CRC carcinogenesis and response to 5-FU using a combination of both *in vitro, ex vivo,* and *in vivo* models. Furthermore, given the central role of ADSL in *de novo* purine synthesis, we evaluated ADSL expression as a predictive biomarker of response to the purine analogue 6-mercaptopurine (6-MP).

Methods

Patients and specimen characteristics

One hundred primary non-consecutive CRCs treated at the University Hospital Basel between the years 2006 and 2012 were included in this study. A tissue microarray (TMA) of these 100 tumors was constructed. For thirty cases non-malignant adjacent mucosa was selected from the same donor block (**Supplementary Methods**). Forty-three primary CRCs and forty-five metastases, including matched primary tumors and metastases of thirty-seven patients, treated at the University Hospital Basel between the years 1995 and 2015 were also used to construct a second TMA. Both studies were performed in accordance with the Declaration of Helsinki and approved by the ethics committee (Ethics Committee of Basel, EKBB 361/12). Data were collected retrospectively in a non-stratified and non-matched manner including patient age, gender, location, stage, grade, vascular invasion, and clinical outcome.

Analysis of The Cancer Genome Atlas (TCGA) dataset

Gene-level expression data for the TCGA CRC cohort [16] with 622 tumors and 51 normal tissues were obtained from Genomics Data Commons (**Supplementary Methods**). Tumor samples were classified as *ADSL*-overexpressing (n = 218) and non-*ADSL*-overexpressing (n = 404) based on the threshold of mean + 2 standard deviations of the normal tissues. Clinical information was obtained for 596 CRCs. Differential expression analysis was performed using the *edgeR* package [17]. Gene set enrichment analysis was performed using the *fgsea* [18] package with genes ranked based on signed p-value according to the direction of the log-fold change (**Supplementary Methods**).

Cell culture

CRC-derived cell lines (HT-29, SW480, Caco-2, and DLD-1) were authenticated by short tandem repeats and cultures were confirmed to be free of mycoplasma infection using the PCR-based Universal Mycoplasma Detection kit (**Supplementary Methods**). NCM460 normal colonic cell line was purchased from INCELL Corporation and cultured according to manufacturer's instructions.

Transient gene knockdown and overexpression

Transient gene knockdown was conducted using ON-TARGET plus siRNA transfection. Gene overexpression was conducted using jetPRIME™ (#9154226, Polyplus-transfection, **Supplementary Methods**).

139

Immunoblot

Total protein lysates containing protease and phosphatase inhibitors were treated with reducing agent, loading buffer, boiled and loaded into SDS-PAGE gels (4-12%). Proteins were transferred to nitrocellulose membranes, blocked for 1 h and incubated overnight at 4 °C with the following primary antibodies: anti-ADSL (1 : 750; A000525, Sigma), anti-B actin (1 : 5000; A5441; Sigma), anti-c-Myc (1 : 1000, MA1-980, ThermoFisher), anti-S6 (1 : 1000, 2217S, Cell Signaling), anti-phospho-S6 (Ser235/236) (1 : 2000, 4858S, Cell Signaling), anti 2-SC antibody (crb2005017d, Discovery Antibodies), anti-NRF2 (1 : 1000, D1C9, Cell Signaling), anti-KEAP1 (1 : 1000, D6B12, Cell Signaling) and anti-TOM20 (1 : 1000, D8T4N, Cell Signaling). Membranes were subsequently incubated with secondary goat anti-mouse or anti-rabbit antibodies for 1 h at room temperature. Band intensity was quantified using ImageJ software and the ratio of proteins of interest/loading control was normalized to their control counterparts (**Supplementary Methods**).

Cell proliferation assay

Cell proliferation was assessed using CellTiter-Glo™ Luminescent Cell Viability Assay (#G7570, Promega, **Supplementary Methods**). For drug treatment, transfected cells were re-plated in 96-well plates, and after 4 h, 5-Fluorouracil (5-FU) (#S1209, Selleckchem), 6-MP (#S1305, Selleckchem) or equal concentration of vehicle (DMSO) was added. After 48 h of treatment, cell proliferation was assayed using CellTiter-Glo™. Results were normalized to the vehicle.

Migration assay

Cell migration was assayed using the xCELLigence Real-Time Cell Analyzer (RTCA) DP (ACEA Biosciences, **Supplementary Methods**). All experiments were performed in triplicate. Results are shown as mean±SD. Statistical significance was assessed by multiple t-test.

Cell cycle analysis by flow cytometry

Cells were transfected with control or ADSL encoding plasmid. 24 h post transfection transfection medium was replaced by either complete medium or glucose-depleted medium and the cells were left to grow for an additional 48 h. Cells were stained with DAPI (10236276001, Sigma-Aldrich) and analyzed by flow cytometry using the BD FACS Canto II cytometer (BD Biosciences). Data were analyzed using the FlowJo software version 10.5.3 (https://www.flowjo.com, **Supplementary Methods**).

Generation of CRC patient-derived organoids (PDOs)

CRC-patient liver metastases were obtained from the University Hospital Basel following patient consent and ethical approval (Ethics Committee of Basel, EKBB, number EKBB 2019-02118). Surgically resected tissues were transported to the laboratory in ice-cold MACS buffer (cat #130-100-008, Miltenyi) and were processed as previously described [19,20] (**Supplementary Methods**). The medium was supplemented with growth factors as previously described [21].

Drug Treatment in PDOs

Tumor organoids were plated as single cells in a 96-well plate at a density of 1×10^4 cells in 10 μL Matrigel droplets. Before treatment, cells were allowed to recover and form organoids for 2 to 3 days. At day 3, 6-MP at a final concentration of 2.5 μM or 5-FU at concentrations of 1.25, 2.5 and 5 μM or in combination was added to the medium, and cell viability was assessed after 5 day using CellTiter-Glo 3D reagent (#G9682, Promega). Luminescence was measured on Varioskan Microplate Reader (ThermoFisher Scientific). Results were normalized to DMSO control. All experiments were performed twice in quadruplicate.

Chorioallantoic membrane (CAM) assay

Fertilized chicken eggs were obtained at day 1 of gestation and incubated at 37 °C with 60% humidity for 9 days. At this time, an artificial air sac was formed, and the CAM was accessed as previously described [22] (**Supplementary Methods**). Caco-2 cells growing in tissue culture were inoculated on CAMs at 2×10^6 cells per CAM, on three to four CAMs each. Embryos were maintained at 37 °C for 4 days after which tumors at the site of inoculation were excised using surgical forceps. Images of each tumor were acquired using a Canon EOS 1100D digital camera. Tumor size measurements were performed by averaging the volume (height*width*width) of each tumor using ImageJ, as previously described [23].

Immunohistochemistry (IHC)

Colonic cell lines were harvested by scraping with cold PBS and fixed in 10% Paraformaldehyde (PFA) for 30 min at room temperature. Similarly, immediately after excision, CAM assay-explanted tumors were fixed in 10% PFA for 2 days at room temperature. PFA-fixed and paraffin-embedded cell pellets and tumors were cut as 4 μm-thick sections. Sections were stained with hematoxylin and eosin (H&E) according to standard protocols.

140

Immunohistochemical staining was performed on a Benchmark immunohistochemistry staining system (Bond, Leica) with BOND polymer refine detection solution for DAB, using anti-ADSL (1 : 100, A000525, Sigma HP), anti-cleaved caspase 3 (1 : 200, 9661S, Cell signaling), anti-c-Myc (1 : 200, MA1-980, ThermoFisher) and anti-phospho-S6 (Ser235/236) (1 : 400, 4858S, Cell signaling) primary antibodies as substrate. Images were acquired using an Olympus BX46 microscope.

For the analysis of the CRC TMA, ADSL immunoreactivity was scored semi-quantitatively by multiplying the proportion of ADSL positive cells (%) and the staining intensity (0 = none; 1 = weak; 2 = intermediate; 3 = strong). The median value of the score in the tumor tissues (score median = 40) was used as cut-off and samples containing a score lower than the median were defined as ADSL-low (ADSL score < 40) while samples containing a score higher than the median were defined as ADSL-high (ADSL score > 40). For all tumors, the tumor-stroma content was calculated as described previously [24] (**Supplementary Methods**).

For the quantification of cleaved caspase 3 (Cl.Casp3) staining performed on the tumors extracted from the CAMs, Cl.Casp3 positive cells were counted in a single high power field (HPF) with the highest density of positive cells, at 400x magnification.

RNA extraction and quantitative PCR

RNA extraction from snap-frozen tissues was performed using TRIzol Reagent (Invitrogen) according to manufacturer's guidelines (Supplementary Methods). cDNA was synthesized from 400 ng of total RNA. Quantitative RT-PCR analysis was performed using SYBR Green. Expression of GAPDH was used to normalize. mRNA fold expression change was calculated by the 2-ΔΔCT method as previously described [25]. All samples were analyzed in triplicate.

Metabolite extraction and multiple pathway targeted analysis (LC-MS Analysis)

Cell lysates were extracted and homogenized as previously described [26,27] (**Supplementary Methods**). Dried sample extracts were re-suspended in MeOH:H_2O (4 : 1, v/v) prior to LC-MS/MS analysis according to the total protein content. Extracted samples were analyzed by hydrophilic interaction liquid chromatography coupled to tandem mass spectrometry (HILIC - MS/MS) in both positive and negative ionization modes as previously described [28] (**Supplementary Methods**).

Analysis of metabolomic data

Comparative analysis of metabolic profiles between ADSL overexpressing and control cells was carried out using MetaboDiff [29]. Data were normalized using variance stabilizing normalization before comparing mean metabolite levels using the student's t test. P values were adjusted using the Benjamini Hochberg procedure. Metabolic correlation modules were defined using the weighted gene coexpression network analysis methodology [30], and identified modules were named according to the most abundant pathway annotation in the module. The metabolic correlation modules were then correlated with ADSL overexpression to identify modules significantly associated with the ADSL overexpressing phenotype.

RNA sequencing and pathway analysis

Biological replicates were generated for all the samples analyzed. Total RNA was extracted from cells at 75% confluence using TRIZOL according to the manufacturer's guidelines. RNA was treated and library generated as described in **Supplementary material and methods**. Differential expression analysis was performed using the DESeq2 Wald test. Gene set enrichment analysis was performed using the *fgsea* R package [18] and the Hallmark gene set from the Molecular Signatures Database [31], using the ranked t statistics from the DESeq2 Wald test. Pathways with false discovery rate (FDR) < 0.05 were considered to be significant. Results were visualized using ggplot2 [32].

MitoTracker and MitoSOX staining

For the analysis of changes in mitochondrial membrane potential, 48 h post-transfection *ADSL* overexpressing and control Caco-2 cells were stained with Mitotracker Red (MTR, M22425, Thermo Fisher). Briefly, 5×10^5 cells were resuspended in 100 µL of pre-warmed (37 °C) FACS buffer (0.1% BSA, dilute in PBS) and placed in a 96-well U plate. Then, 100 µL of MTR (100 nM) were added to the wells and cells were incubated 20 min at 37 °C. The reaction was stopped by addition of cold DMEM-10% FBS and incubating for 5 min at 4 °C. Cells were washed twice with DMEM-10% FBS and resuspend in 100 µL of FACS buffer.

For analysis of mitochondrial production of ROS, 48 h post-transfection control and *ADSL* overexpressing Caco-2 cells were stained with Mitosox red (M36008,Thermo Fisher). Briefly, 1×10^5 cells were resuspended in 100 µL of pre-warmed HBSS (Ca^{2+} and Mg^{2+}) and placed in a 96-well U plate. 100 µL of Mitosox red (20 µM) diluted in prewarmed HBSS was added to each well. Cells were incubated 20

min at 37 °C. After incubation, cells were washed twice with FACS buffer and resuspended in 100 µL of FACS buffer.

All samples were analyzed using the CytoFLEX cytometer (Beckman). Information about the percentage of positive cells as well as Mean Fluorescence Intensity were recorded for all the samples and analyzed by FlowJo software version 10.5.3.

OCR and ECAR measurements

For analysis of the Oxygen consumption rate (OCR) (in pmol/min) and extracellular acidification rate (ECAR) (in mpH/min), the Seahorse XF-24 metabolic extracellular flux analyzer was used (Seahorse Bioscience). Caco-2 and *NCM40D* were plated onto Seahorse cell plates (8×10^4 cells per well) in serum-free unbuffered DMEM medium (Sigma-Aldrich). Perturbation profiling of the use of metabolic pathways was achieved by the addition of oligomycin (1 µM), Carbonyl cyanide-4-(trifluoro-methoxy)phenylhydrazone (**FCCP**) (0.5 µM) and rotenone (1.3 µM) plus antimycin A (20 µM) (all from Sigma-Aldrich). Experiments with the Seahorse system were done with the following assay conditions: 2 min mixture; 2 min wait; and 4–5 min measurement. Metabolic parameters were then calculated as previously described [33].

Fumarate measurement

Fumarate concentration in Caco-2 and NCM460 cells was measured using the Fumarate Assay Kit (Abcam, 102516) following the manufacturer's instructions. Briefly, cells were plated in 96-well plates 8 h after transfection and supernatant was collected 24, 48, 72 and 96 h post-transfection. The amount of fumarate was quantified relative to fumarate standard curve.

Fumarate rescue experiment

Caco-2 cells were re-plated at a density of 5×10^3 cells/well in a 96-well plate 8 h post transfection. Dimethyl fumarate (Selleckchem, #S2586) or vehicle (DMSO) were added to cells 18 h before measuring the cell index for the respective time points to a final concentration of 50 µM. Proliferation was measured using the CellTiter-Glo™ Luminescent Cell Viability Assay (Promega, #G7570) at 24, 48, 72 and 96 h after re-plating. Results were normalized to 4 h.

Statistical analysis

Statistical analyses were performed using Prism software v6.0 (GraphPad Software) and R. For *in vitro*, *in vivo* and *ex vivo* studies, statistical significance was determined by the two-tailed unpaired Student's t-test. The differences were considered statistically significant at $p < 0.05$. All experiments were performed at least twice. The statistical parameters (e.g., the exact value of n, p values) have been noted in the figures and figure legends. Bar plots with error bars show mean ± SD. Statistical comparisons for categorical variables were performed using χ^2 or Fisher's exact tests where appropriate. Statistical comparisons for numeric and ordinal variables were performed using t-test, Mann-Whitney U test, paired Wilcoxon test or Cochran-Armitage test for trend where appropriate. Ordinal regression analysis was performed using cumulative link models using the 'ordinal' R package. Survival analysis was performed using the Kaplan-Meier method and log-rank test. For the TCGA cohort, stratification of ADSL expression for overall survival analysis was performed using the 'maxstat' [34] R package.

Results

ADSL is overexpressed in CRCs but its expression does not increase with disease stage

We recently analyzed the large-scale perturbation screen of the project DRIVE (Deep RNAi Interrogation of Viability Effects in cancer), in which 7,837 genes were targeted with a median of 20 shRNAs per gene in 398 cancer cell lines across a variety of malignancies [35]. In this analysis, we identified the *ADSL* gene as one of the top putative oncogenes in CRC. Indeed, 33 out of 35 CRC cell lines profiled displayed significantly decreased cell viability upon *ADSL* knockdown ($p < 0.001$; **Figure 1A**). To confirm the putative oncogenic role of *ADSL* and its potential role as biomarker in CRC, we performed ADSL IHC on two TMAs, one containing 100 CRC samples and 30 non-tumoral adjacent tissues and the other containing 43 CRC primary samples and 45 metastatic tissues. In the first TMA, after excluding the samples with missing tissue core or poor staining, 73 tumor samples and 13 non-tumoral tissue cores were analyzed for ADSL protein expression. Immunoscore results showed that ADSL expression was significantly increased in tumor tissues compared to non-tumor tissue areas ($p = 0.001$; **Figure 1B-C**). Additionally, a regression analysis found that ADSL immune scores showed a downward trend with increasing disease stage ($p = 0.035$; **Figure 1D**). In the second TMA containing CRC primary and metastatic tissues, we observed no difference in ADSL expression levels between primary and all types of metastasis (liver, lung, small intestine, lymph node, abdominal wall, bone, bladder, cutaneous, ovarian; **Figure 1E-F** and **Figure S1A**) or between primary and paired liver metastasis (**Figure**

S1B). Our results suggest that ADSL is overexpressed in CRC but its expression is not increased upon tumor progression.

We then asked whether ADSL expression was associated with any clinicopathological feature in our primary CRC patient cohort (**Table S1**). We found that ADSL-low tumors were associated with lymphatic and lymphovascular invasion, as well as with low tumor-stroma ratio (all p = 0.001; **Table S1**), all indicators of poor prognosis. No association with microsatellite instability, tumor grade (**Table S1**), or patient survival (**Figure S1C**) was observed.

To corroborate the results obtained with our in-house cohorts we evaluated the mRNA expression of *ADSL* in the 622 CRCs and 51 normal tissues in the TCGA cohort [36]. In accordance with the IHC results,

we found that *ADSL* expression was significantly higher in CRC samples compared to normal tissue (p < 0.001; **Figure 1G**). Additionally, similar to our TMA cohort of primary CRCs, ADSL transcript levels showed a downward trend with increasing disease stages (p < 0.001; **Figure 1H** and **Table S2**). However, the difference in *ADSL* expression levels between early- and advanced-stage patients was very modest (**Figure 1H**). We also found that ADSL levels were slightly higher in CRC tumors classified as CMS2 [37] (p < 0.05 compared to other CMS subtypes, **Figure S2A**), a molecular subtype characterized by epithelial differentiation and strong upregulation of WNT and MYC downstream targets [37]. No association between *ADSL* expression and patient overall survival was found (**Figure S2B**).
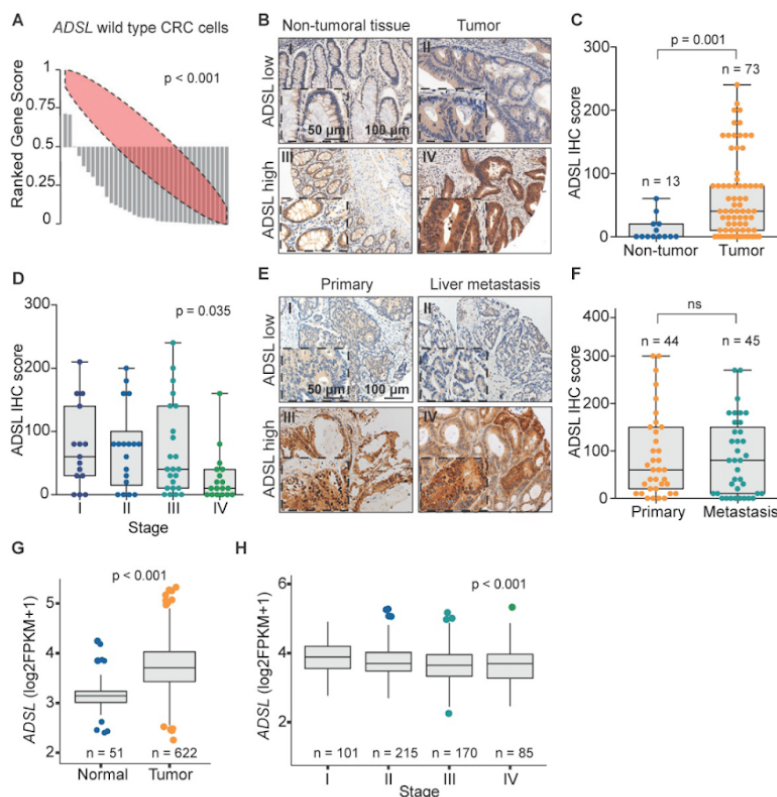
Figure 1. *ADSL* is overexpressed in CRC. (A) Rank profile of cell viability upon knockdown of the *ADSL* gene in CRC cell lines with no genetic alteration in the *ADSL* gene. Each bar in the waterfall plot represents one cell line. The value of each bar represents the rank of viability for the *ADSL* gene among all genes knocked down for a given cell line in the project DRIVE. Thus ranks close to zero represent reduced viability while ranks close to one indicate cell growth upon knockdown of *ADSL*. The red ellipse in the rank profile represents a no-change (random) viability band. **(B)** Representative micrographs showing ADSL staining in CRC and non-tumoral adjacent tissue: (I) non-tumoral tissue with low ADSL staining score; (II) tumor tissue with low ADSL staining score; (III) non-tumoral tissue with high ADSL staining score and (IV) tumor tissue with high ADSL staining score. Semi-quantitative scoring of ADSL staining in **(C)** CRCs and non-tumoral adjacent tissues and **(D)** CRCs stratified by disease stage. **(E)** Representative micrographs showing ADSL staining in CRC primary and metastasis tissue: (I) primary tumor with low ADSL staining score; (II) metastasis tissue with low ADSL staining score; (III) primary tumor with high ADSL staining score and (IV) metastasis tissue with high ADSL staining score. **(F)** Semi-quantitative scoring of ADSL staining in CRC primary and metastasis tissue. **(G)** *ADSL* transcript expression in CRCs and normal tissues in the TCGA cohort [36]. **(H)** *ADSL* transcript expression in CRCs in the TCGA cohort, stratified by disease stage. Statistical significance was assessed for **(C, F-G)** by unpaired student's t-tests and for **(D-H)** by ordinal regression. Scale bars 50-100 μm.

143

Taken together our data support the hypothesis that ADSL acts as a potential oncogene in CRC and suggest its possible implication in the early stage of colorectal carcinogenesis.

ADSL has oncogenic-like properties in *in vitro* and *in vivo* models of CRC

To validate the potential oncogenic role of *ADSL* we screened ADSL protein expression in CRC cell lines and selected the SW480 and DLD-1 cell lines, with high endogenous levels, for knock-down experiments, and the Caco-2 and HT-29 cells, with low endogenous levels, for overexpression (**Figure S3A-B**). We first sought to validate the results from the project DRIVE and assessed if *ADSL* knock-down would affect cell viability in CRC. By silencing *ADSL* using a siRNA approach, we achieved a reduction of ADSL protein expression by 70% and 80% 48 h post-transfection in DLD-1 and SW480 cells, respectively (**Figure 2A** and **Figure S3C**). In both cell lines, transient *ADSL* knock-down significantly decreased the proliferation rate compared to control cells (**Figure 2B** and **Figure S3D**). Similarly, both DLD-1 and SW480 *ADSL*-silenced cells showed a significant reduction in the migration potential compared to control cells (**Figure 2C** and **Figure S3E**). These data support our hypothesis that ADSL is a novel oncogene in CRC. However, the phenotype observed upon *ADSL* depletion may also imply that *ADSL* is an essential gene in colonic cells. To test this hypothesis, we silenced *ADSL* expression in the normal colonic cell line NCM460 (**Figure S3F**). Of note, ADSL levels are already lower in normal cells compared to colorectal tumor cells (**Figure S3A**). *ADSL* gene knock-down did not significantly impact the viability of NCM460 cells (**Figure S3G**), suggesting that while it confers a proliferative advantage to tumor cells, it does not have the same effect/impact in normal colonic cells.

To mimic the conditions observed in human CRC tissues and to validate the oncogenic role of ADSL, we assessed cell proliferation and migration potential upon forced overexpression of *ADSL* in Caco-2 and HT-29 cells. We achieved 35% and 25% increased ADSL protein expression 48 h post-transfection in Caco-2 and HT-29 cells, respectively (**Figure 2D** and **Figure S3H**). In support of our hypothesis, overexpression of *ADSL* in Caco-2 and HT-29 led to a significant increase in the proliferation rate compared to control cells (**Figure 2E** and **Figure S3I**). Similarly, the migration rate was significantly increased in *ADSL*-overexpressing cells compared to control cells (**Figure 2F** and **Figure S3J**).

To demonstrate the specificity of our results, we performed rescue experiments. Specifically, we restored ADSL expression levels in DLD-1 and Caco-2 cells where *ADSL* was silenced or overexpressed, respectively (**Figure 2G**). Restoring ADSL to levels similar to the endogenous expression rescued the phenotype, thus suggesting that the effects induced by *ADSL* modulation on cell proliferation were on-target (**Figure 2H**).

To further demonstrate the putative oncogenic role of ADSL in CRC, we modulated *ADSL* expression and xeno-transplanted *ADSL*-overexpressing or control CRC cells into the chicken chorioallantoic membrane (CAM) to assess tumor growth *in vivo*. Engraftment of tumor cells in the CAM has been successfully used as a fast and reproducible model of tumorigenesis [38,39]. Briefly, twenty-four hours post-transfection, CRC cells were harvested, re-suspended in Matrigel and seeded into the CAMs (**Methods**). Four days later, the eggs were screened for tumor formation and tumors were harvested (**Figure 2I**). *ADSL*-overexpressing Caco-2 cells formed significantly larger tumors compared to control cells (**Figure 2J**). IHC analysis confirmed that the resected tumors were indeed of human origin and that ADSL overexpression could still be detected five days post-transfection (**Figure 2K** and **Figure S3K**).

Our *in vitro* and *in vivo* results provide evidence of the role of ADSL as an oncogene in colorectal carcinogenesis.

Dysregulation of ADSL affects cell cycle progression and increases the DNA damage marker γH2AX

To understand which signaling pathways might be affected as a result of *ADSL* overexpression we classified CRC patient samples from the TCGA cohort as *ADSL*-overexpressing and non-*ADSL*-overexpressing (**Methods**) and then performed differential gene expression and gene set enrichment analysis (GSEA). In accordance with the role of *ADSL* in *de novo* DNA synthesis, we found an upregulation of genes involved in the purine nucleoside biosynthetic process pathways as well as DNA replication and repair in the *ADSL*-overexpressing samples (p adj = 0.025; **Figure 3A** and **Figure S4A-B**). We found that expression of the 'replicative' DNA polymerase subunits *POLA2, POLD2,* and *POLE2* genes, as well as the proliferative marker *PCNA* correlated with *ADSL* expression (**Figure S4B**). Genes involved in the firing of DNA replication (*CDC45, MCM2, CDT1*) also showed significant positive correlation with *ADSL* in CRCs (**Figure S4B**). Of note, we found enrichment of genes involved in DNA repair, and specifically DNA repair-associated DNA synthesis (**Figure 3A** and **Figure S4C**). In particular, S-phase checkpoint genes involved in the

downstream protection of stalled DNA replication forks (e.g. *RAD51*) as well as S-phase signaling mediators (e.g. *CHEK2*), previously reported to be upregulated in CRCs, showed significant positive correlation with *ADSL* expression (**Figure S4C**) [40]. We therefore tested whether forced expression of ADSL might drive DNA replication stress in CRC by immunostaining H2AX phosphorylation at serine 139 (γH2AX), a sensitive indicator of both DNA damage

and DNA replication stress [41]. Indeed, *ADSL*-silenced DLD-1 cells showed weaker staining for γH2AX compared to control cells, while forced expression of *ADSL* in Caco-2 strongly upregulated γH2AX protein staining (**Figure 3B**). *ADSL*-overexpressing CAM tumors also showed higher expression of the DNA damage marker γH2AX (**Figure 3C** and **Figure S5A**), further suggesting a role of ADSL in DNA replication stress.
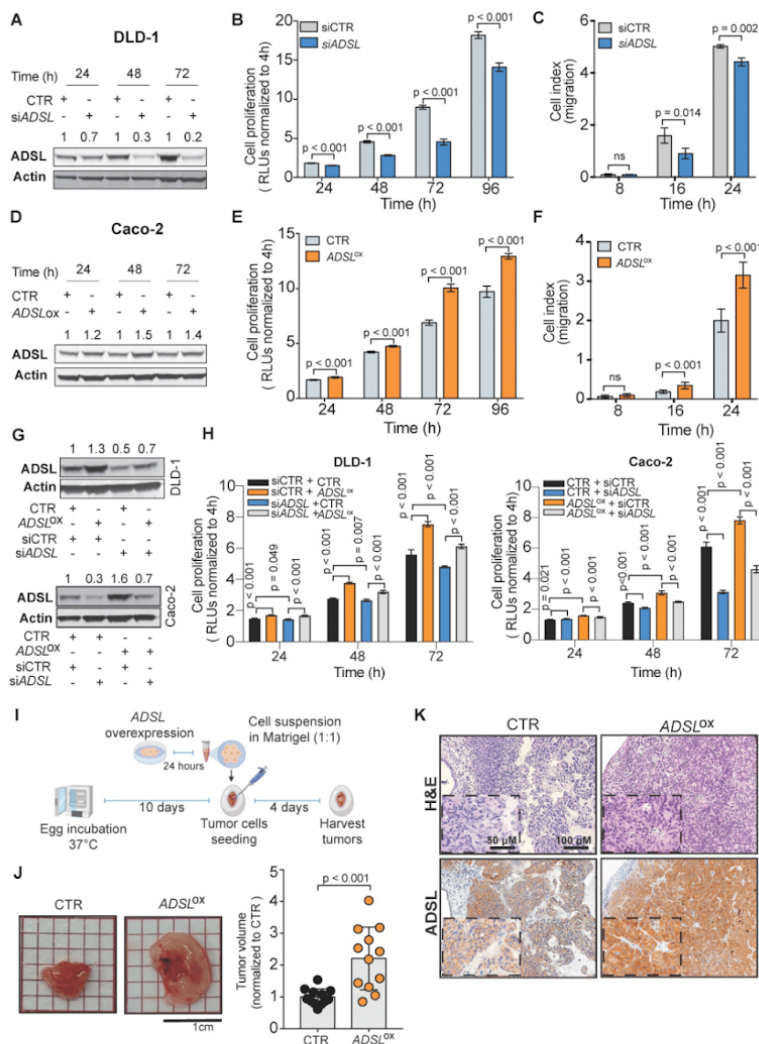
Figure 2. *ADSL* has oncogenic-like properties in *in-vitro* and *in-vivo* models of CRC (**A**) Immunoblot showing ADSL expression in the DLD-1 cell line at 24, 48 and 72 h post siRNA transfection. (**B**) Proliferation and (**C**) migration capacity of DLD-1 *ADSL*-silenced cells compared to control cells. (**D**) Immunoblot showing ADSL expression in Caco-2 cells at 24, 48, and 72 h post vector transfection. (**E**) Proliferation and (**F**) migration capacity of Caco-2 *ADSL*-overexpressing cells compared to control cells. (**G**) Immunoblot showing ADSL expression in DLD-1 cells (upper panel) and Caco-2 cells (lower panel) 72 h post siRNA and/or vector transfection. (**H**) Rescue experiment showing proliferation capacity of DLD-1 (left) and Caco-2 (right) cells with siRNA and/or vector transfection compared to control cells. (**I**) Schematic illustration of the CAM assay. (**J**) Representative pictures and quantification of volume of Caco-2 control (CTR) and *ADSL*-overexpressed (*ADSLox*) resected tumors 4 days post-implantation of cells in CAMs

145

(n = 12 tumors from 2 independent experiments); mean ± SD. Values are normalized to the mean of CTR. (**K**) Representative pictures of Caco-2 tumors extracted 4 days post-implantation. Tissue sections were immunostained with ADSL. Scale bars 20-50 µm. For (**C**) and (**F**) 8, 16 and 24 correspond to 32, 40 and 48 h post transfection, respectively.
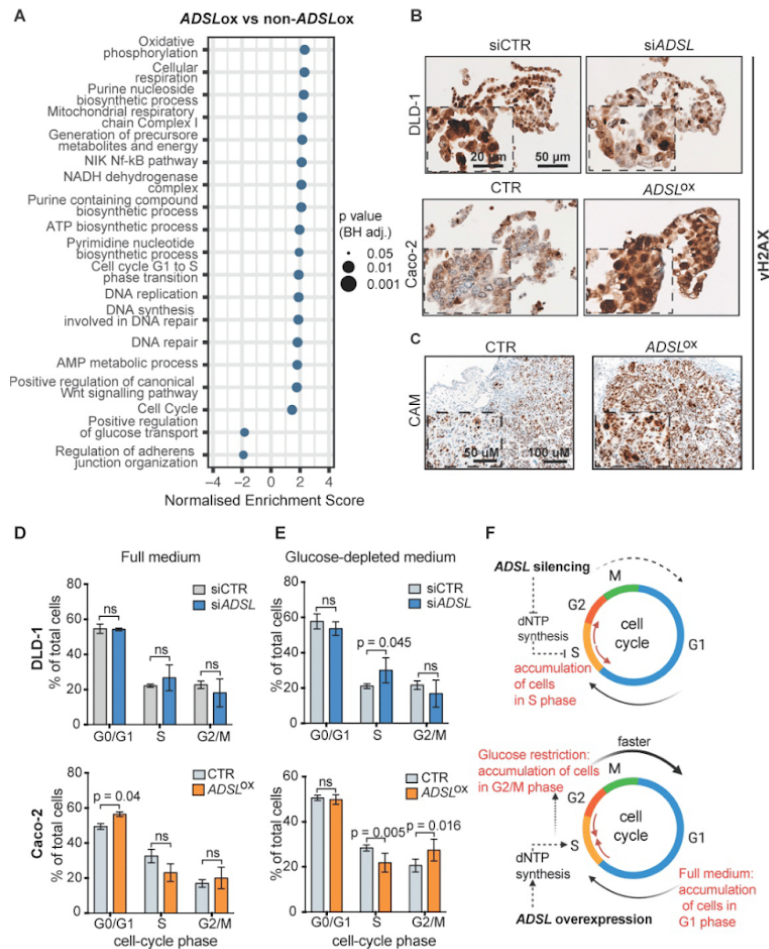
Figure 3. Dysregulation of ADSL affects cell cycle and induces replication stress (**A**) Normalized enrichment scores of significantly up-regulated pathways identified by gene set enrichment analysis (GSEA) in *ADSL*-overexpressing (*ADSL*ox) versus non-*ADSL*-overexpressing (non-*ADSL*ox) CRCs. NES = normalized enrichment score. (**B**) Representative pictures of control and *ADSL*-silenced DLD-1 cells and control *ADSL*-overexpressed Caco-2 cells immunostained with the DNA damage marker γH2AX. Scale bars 20-50 µm. (**C**) Representative pictures of Caco-2 tumors extracted 4 days post-implantation. Tissue sections were immunostained with γH2AX. Scale bars 50-100 µm. (**D-E**) Flow cytometry analysis of DAPI-stained DLD-1 (blue) and Caco-2 (orange) upon *ADSL* transient downregulation or upregulation in (**D**) complete medium and (**E**) glucose-deprived medium. (**F**) Schematic representation of *ADSL* silencing or *ADSL* overexpression impact on cell cycle in CRC cell lines. Data are mean ± SD (**D-E**) n ≥ 3 replicates. For all experiments, statistical significance was assessed by multiple t-tests.

Pathways related to cell cycle, specifically G1 to S phase transition, were also significantly upregulated in *ADSL*-overexpressing versus non-*ADSL*-overexpressing CRCs (p adj = 0.025; **Figure 3A** and **Figure S4D**). We found that *ADSL* expression correlated with many cyclins and cell-cycle checkpoints related genes. We also observed a correlation between the expression levels of *ADSL* and the *MYC* oncogene, whose expression has been shown to be indirectly regulated by ADSL in triple-negative breast cancer [6] and whose upregulation of target genes is a signature of the CMS2 subtype [37]. In accordance with the enrichment in the CMS2 molecular subtype, *ADSL*-overexpressing tumors also showed significant enrichment of the canonical WNT signaling pathway (p = 0.037; **Figure 3A**).

Given the role of *ADSL* in *de novo* DNA synthesis and that the cell cycle pathway, specifically G1 to S phase transition, was also significantly upregulated in *ADSL*-overexpressing CRCs, we hypothesized that modulation of *ADSL* expression might also affect cell cycle in CRC cells. We therefore assessed the distribution of cells in the different cell-cycle phases upon modulation of *ADSL* expression by flow cytometry (FACS) (**Figure S5B**). Under normal growth conditions (complete cell culture medium), transient *ADSL* knock-down showed a tendency to increase the percentage of DLD-1 and SW480 cells in the S-phase, while *ADSL* overexpression significantly increased the fraction of Caco-2 and HT-29 cells in the G1 phase (**Figure 3D** and **Figure S5C**). To determine if the accumulation of *ADSL*-overexpressing cells in the G1 phase was a result of a block in G1 phase or an indicator of faster cell cycle and increased proliferation rate, we synchronized cells at the G1-S transition checkpoint by glucose deprivation [42] and again analyzed cell cycle upon modulation of *ADSL* expression. Given the key role of *ADSL* in *de novo* purine synthesis and glucose being the primary substrate of this metabolic pathway, we expected to have a stronger phenotype upon glucose restriction. Indeed, under glucose-depleted growth conditions, *ADSL*-silenced DLD-1 and SW480 cells could not proceed from S to G2/M phase, which resulted in significant accumulation of cells in S phase compared to control cells, while no significant difference in the proportion of cells in the S phase was observed between *ADSL*-silenced and control cells in full medium (**Figure 3E** and **Figure S5D**). By contrast, *ADSL* overexpression facilitated cell transition from S-phase to G2/M phase in both Caco-2 and HT-29 CRC cells (**Figure 3E** and **Figure S5D**).

Together with the exploratory analysis performed on the TCGA cohort, our data suggest the involvement of ADSL in the deregulation of cell cycle and DNA repair/replication mechanisms during the process of colorectal oncogenesis. In particular, our results show that ADSL overexpression accelerates progression through the cell cycle (**Figure 3F**).

ADSL expression levels predict response to 6-mercaptopurine in *in vitro*, *in vivo* and in patient-derived organoids

The antimetabolite 5-FU is the most commonly used chemotherapeutic agent for CRC treatment. To evaluate whether ADSL expression could modulate response to 5-FU, we compared the viability of Caco-2 *ADSL*-overexpressing and control cells upon treatment with 5-FU. *ADSL* overexpression did not significantly affect response to 5-FU *in vitro* (**Figure S6A-B**). 5-FU is, however, a pyrimidine analogue,

while ADSL is involved in the conversion of the intermediate molecule SAICAR into AICAR and fumarate in the purine nucleotide cycle (**Figure 4A**). We therefore asked whether *ADSL* expression levels may instead better predict response to a purine analogue antimetabolite, such as 6-mercaptopurine (6-MP). 6-MP competes with inosine monophosphate (IMP) thus inhibiting the same biosynthetic pathway in which ADSL plays a role (**Figure 4A**). As hypothesized, *ADSL* overexpression significantly sensitized Caco-2 cells to 6-MP (p < 0.001; **Figure S6C-D**). Caco-2 cells overexpressing *ADSL* treated with 6-MP also showed positivity for the apoptotic marker cleaved caspase 3 (**Figure S6E**). To determine whether modulating *ADSL* expression levels would also affect response to 6-MP *in vivo*, we pre-treated *ADSL*-overexpressing Caco-2 cells with 6-MP for 24 h before implantation in the CAM and then screened for tumor formation. In accordance with our results obtained *in vitro*, treatment with 6-MP significantly reduced tumor volume only upon *ADSL* overexpression (**Figure 4B**). Indeed, no significant difference could be appreciated when comparing the tumors derived from control cells treated with DMSO or 6-MP, while the volume of tumors derived from *ADSL*-overexpressing cells pre-treated with 6-MP was significantly smaller compared to *ADSL*-overexpressing cells pre-treated with DMSO (**Figure 4C**), suggesting that *ADSL* expression also modulates response to 6-MP *in vivo*. *ADSL*-overexpressing tumors pre-treated with 6-MP also showed a significantly stronger positive signal for cleaved caspase 3, as well as morphological features of apoptosis in H&E staining (e.g. hypereosinophilic cytoplasm, nuclear fragmentation, "apoptotic bodies", **Figure 4D-E** and **Figure S7**), demonstrating that 6-MP induces apoptosis more efficiently in *ADSL*-overexpressing CRCs and supporting the importance of an efficient purine nucleotide cycle for the oncogenic activity mediated by ADSL.

Accumulating evidence indicates that CRC-patient derived organoids (PDO) retain molecular features of the original tumor, and they are amenable to drug screening and response prediction in a preclinical setting [43–45]. We therefore decided to explore whether differences in ADSL level may predict response to 5-FU or 6-MP in four PDOs generated from liver metastasis of CRC patients (PDO1 to PDO4; **Table S3**). The PDOs showed different levels of ADSL protein expression by immunoblot (**Figure 4F**). Immunohistochemical analysis of ADSL showed cytoplasmic immunoreactivity in all original tissues and their derived organoids, with ADSL expression levels that were consistent with the immunoblot results (**Figure**

147

4G). All the PDOs retained the histopathological features of the original tumor tissue from which they were derived (**Figure 4G** and **Figure S8**). Indeed, hematoxylin and eosin (H&E) staining confirmed that tumor-derived organoids resembled the patient-specific morphology heterogeneity, ranging from cystic to solid/compact phenotype (**Figure 4G**), as well as a similar staining pattern for colorectal-specific markers (**Figure S8**).
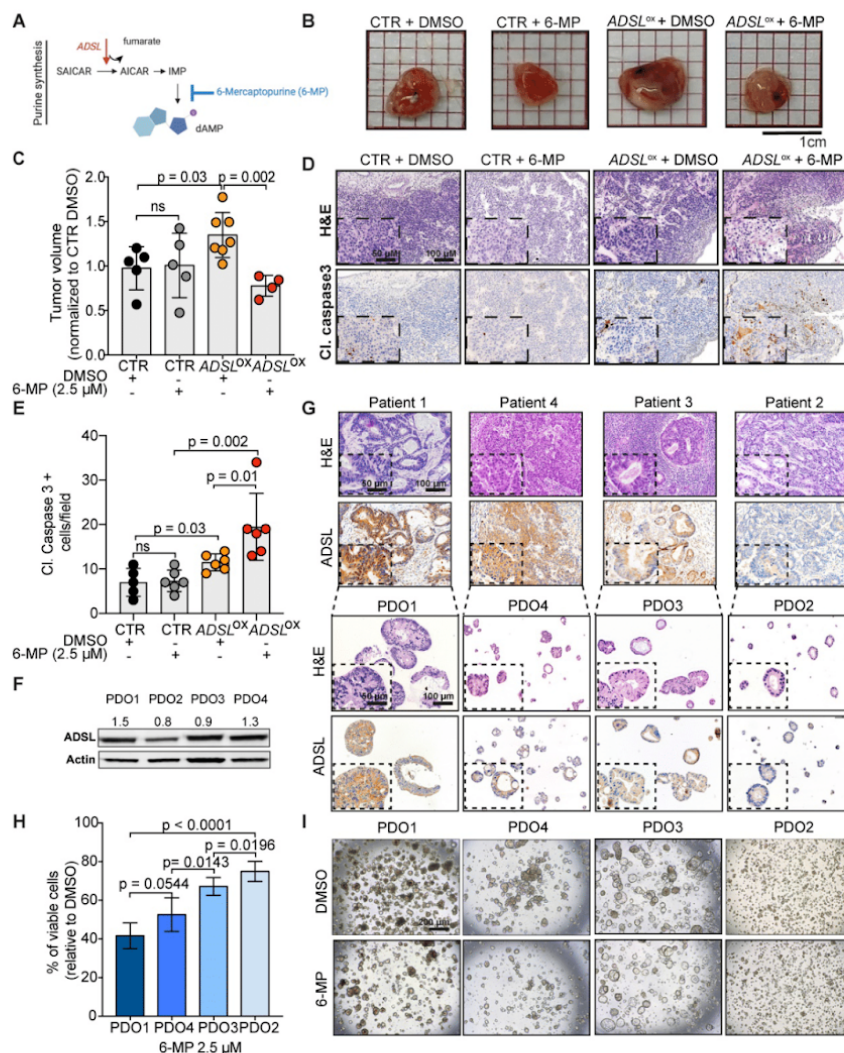
Figure 4. ADSL expression levels predict response to 6-mercaptopurine in *in vitro*, *in vivo* and PDOs. (**A**) Schematic diagram of the role of *ADSL* as well as 6-MP inhibitory effect in the *de novo* purine biosynthetic process. (**B**) From the left: pictures of Caco-2 control cells pre-treated with DMSO, Caco-2 control cells pre-treated with 6-MP, *ADSL*-overexpressed Caco-2 cells pre-treated with DMSO and *ADSL*-overexpressed Caco-2 cells pre-treated with 6-MP implanted in CAMs and grown for 4 days post-implantation. (**C**) Quantification of tumor growth derived from the CAM experiment (n = at least 4 tumors from 2 independent experiments); mean ± SD. Values are normalized to the mean of CTR DMSO. (**D**) Representative micrographs of Caco-2 tumors extracted 4 days post-implantation. Tissue sections were stained with Hematoxylin-eosin (H&E) and immunostained with the apoptotic marker cleaved Caspase 3 in the different treatment conditions. (**E**) Quantification of cleaved caspase 3 positive cells in the tumor from (**C**). (**F**) Immunoblot showing ADSL protein expression in the different CRC-PDOs. Quantification is relative to the loading control (Actin). (**G**) Representative pictures of matched tissue-organoids pairs. Both tissues and organoids sections were stained with Hematoxylin-eosin (H&E) (upper panel) and ADSL antibody (lower panel). (**H**) Percentage of viable cells relative to DMSO in CRC-PDOs treated with 2.5 μM of 6-MP. (**I**) Representative pictures of the organoid growth after 5 days treatment with 2.5 μM of 6-MP. Data are mean ± SD (**B**, **E**, **H**) n ≥ 3 replicates. For all experiments, statistical significance was assessed by unpaired t-test. Scale bars are 50 and 100 μm for (**D**, **G**) and 200 μm for (**I**).

Using these CRC-PDOs, we explored whether differences in ADSL level may affect response to 5-FU or 6-MP in a pre-clinical setting. Contrary to the results obtained *in vitro*, 5-FU treatment had a moderate but significant effect in reducing cell viability in ADSL-high PDOs (PDO1 and PDO4) compared to ADSL-low PDOs (PDO2 and PDO3; **Figure S9A**), suggesting that ADSL may also be partially involved in the cell response to 5-FU. On the other hand, consistent with our *in vitro* results, 6-MP treatment on PDOs demonstrated a strong ADSL-level-dependent response. In fact, PDOs with high ADSL expression (PDO1 and PDO4) showed higher sensitivity to 6-MP compared to PDOs with lower ADSL expression (PDO2 and PDO3) in terms of cell viability (**Figure 4H**) and cell proliferation (**Figure 4I**). To test whether 6-MP and 5-FU may act synergistically in CRC, we treated PDO1 and PDO3, respectively with high and low ADSL expression, with different dosages of 5-FU alone or in combination with a fixed dose of 6-MP. While the addition of 6-MP to 5-FU significantly reduced the percentage of cell viability in PDO1, the difference between treatment with 6-MP alone or in combination with 5-FU was minor (approximately 60% viable cells with 6-MP alone versus 50% in combination with all three concentrations of 5-FU; **Figure S9B**, left). On the contrary, PDO3 was resistant to both treatments alone or in combination (**Figure S9B**, right). Overall our data do not support the synergistic effect of 6-MP and 5-FU and suggest that ADSL overexpression may specifically sensitize CRCs to treatment with 6-MP.

ADSL overexpression causes mitochondrial dysfunction and oxidative stress

To understand the molecular basis of oncogenic properties of ADSL we performed targeted metabolomic and global transcriptomic analysis on *ADSL*-overexpressing and control Caco-2 cells (**Figure 5A**). In line with the well-defined role of ADSL, *ADSL*-overexpressing cells showed a significant fold increase in metabolites involved in purine as well as pyrimidine biosynthetic processes (**Figure S10A**). Likewise, metabolites from mitochondrial tricarboxylic acid (TCA) cycle and beta-oxidation were also significantly enriched (**Figure S10A**). Mirroring the metabolomic data, gene set enrichment analysis performed on the transcriptomic data showed that oxidative phosphorylation and lipid metabolism were indeed dysregulated in *ADSL*-overexpressing cells, together with the up-regulation of Myc targets (**Figure 5B**). Given that mitochondrial processes such as oxidative phosphorylation and TCA cycle were significantly dysregulated in both omics' analyses, we hypothesized a role for ADSL in

regulating mitochondrial function in CRC. Indeed, *ADSL*-overexpressing CRCs from the TCGA cohort showed an enrichment in genes involved in oxidative phosphorylation and mitochondrial respiratory chain (**Figure 3A**). To further explore this hypothesis, we measured the mitochondrial membrane potential of *ADSL*-overexpressing and control cells using MitoTracker Red. In line with the up-regulation of oxidative phosphorylation seen at the gene expression level, *ADSL*-overexpressing cells showed significantly higher membrane potential (**Figure 5C**). However, this was not reflected in a higher mitochondrial respiration rate (**Figure 5D**). By contrast, *ADSL*-overexpressing cells showed significantly reduced oxygen consumption rate (OCR) compared to control cells (**Figure 5D**). Specifically, *ADSL* overexpression significantly impaired both basal and ATP-coupled respiration as well as maximal respiratory capacity in Caco-2 cells (**Figure 5E**). Leak respiration and non-mitochondrial respiration were also significantly reduced (**Figure S10B**).

Previous studies have demonstrated that generation of reactive oxygen species (O2(*-), ROS) is partially regulated by the membrane potential ($\Delta\psi m$), such that an increase in mitochondrial membrane potential due to dysfunctional electron transport favors ROS production [46]. Indeed, *ADSL* overexpression significantly increased mitochondrial ROS production in Caco-2 cells, as measured using MitoSOX staining (**Figure 5F**).

Reduced aerobic respiration is usually accompanied by a higher glycolytic rate in cancer cells [47]. However, *ADSL*-overexpressing cells additionally showed a significantly lower basal and maximal glycolytic capacity (**Figure 5G-H**). Accordingly, we also demonstrated that the impact of *ADSL* overexpression on cell cycle progression was significantly greater upon glucose depletion (**Figure 3D**).

ADSL catalyzes the conversion of SAICAR to AICAR as well as S-AMP to AMP, and in both reactions fumarate is generated. Fumarate is a TCA cycle intermediate and a well-known oncometabolite which has been connected to ADSL oncogenicity in endometrial cancer [7]. Additionally, fumarate accumulation has been shown to cause respiratory chain defects [48] as well as to impair glycolysis [49]. Accordingly, we found that *ADSL* overexpression led to a significant fumarate accumulation in Caco-2 cells in a dose-dependent manner (**Figure 5I** and **Figure S10C**). Treating parental cells with dimethyl fumarate (DMF), a cell-permeable form of fumarate, significantly impared cellular respiration, albeit to a lesser extent compared to *ADSL* overexpression (**Figure 5J**). Specifically, DMF significantly impacted

149

basal and ATP-coupled respiration of Caco-2 cells, whilst there was no observable effect on maximal respiratory capacity and leak respiration (**Figure S10D**). Similarly, DMF significantly impaired the glycolytic activity of parental Caco-2 cells (**Figure 5K**), both at their basal and maximal capacity (**Figure S10E**).

The KEAP1-NRF2 and the mTOR-cMYC axis are independently activated upon *ADSL* overexpression in CRC

To test the causality between fumarate increase and the phenotype driven by ADSL dysregulation, we performed a rescue experiment by adding DMF to DLD-1 cells after transfection with siRNAs against
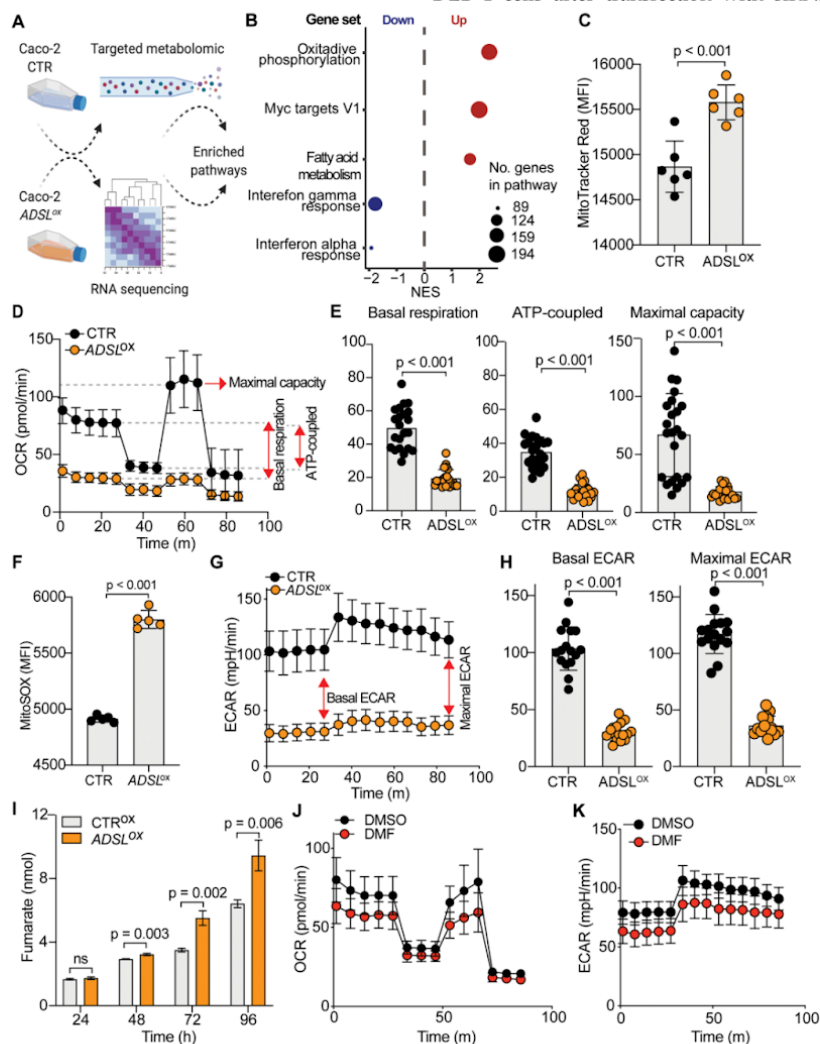
Figure 5. ADSL overexpression causes mitochondrial dysfunction and oxidative stress. (**A**) Schematic representation of the multi-omics approaches performed on Caco-2 cells control and overexpressing ADSL. (**B**) Normalized enrichment scores of significantly up- and down-regulated pathways identified by gene set enrichment analysis in control and *ADSL*-overexpressing Caco-2 cells. (**C**) Mitochondrial membrane potential measured by flow cytometry in control and ADSL overexpressing Caco-2 cells stained with MitoTracker Red. Y-axis represents mean fluorescence intensity (MFI). (**D**) Oxygen-consumption rate (OCR) of control and *ADSL*-overexpressing Caco-2 cells in real time under basal conditions and after drug-induced mitochondrial stress [73] with oligomycin, FCCP and rotenone plus antimycin A. (**E**) Mean basal, ATP-coupled and maximal respiratory capacity of the cells in (**D**). (**F**) Mitochondrial reactive oxygen species measured by flow cytometry in control and *ADSL*-overexpressing Caco-2 cells stained with MitoSOX. Y-axis represents mean fluorescence intensity (MFI). (**G**) ECAR of control and *ADSL* overexpressing Caco-2 cells in real time after treatment as in **D**. (**H**) Mean basal and maximal ECAR of the cells in **D**. (**I**) Extracellular fumarate level (nmol) in control and *ADSL*-overexpressing Caco-2 cells 24, 48, 72 and 96 h post transfection. (**J**, **K**) OCR (**J**) and ECAR of parental Caco-2 cells treated with DMSO or fumarate (50 μM) in real time after treatment as in **D**.

ADSL or control siRNAs. The addition of DMF significantly increased the proliferation rate of *ADSL*-silenced cells (**Figure 6A**). However, fumarate only partially rescued the phenotype induced by *ADSL* gene knock-down. We therefore asked which other molecular mechanism could account for the pro-proliferative role of *ADSL* in CRC. Given that *ADSL*-overexpressing Caco-2 cells displayed a significant enrichment for *MYC*-targets (**Figure 5B** and **Figure S11A**) and that *ADSL* was previously shown to activate the c-MYC pathway in triple negative breast cancer [6], we therefore asked whether a similar mechanism occurs in CRC.

While *MYC* mRNA levels only mildly increased

upon *ADSL* overexpression (**Figure S11B**), we detected a 1.5- and 2-fold increase in the c-MYC protein levels in *ADSL*-overexpressing Caco-2 cells at 48 and 72 h post-transfection, respectively (**Figure 6B**). The increase in c-MYC protein was further detected by immunohistochemistry (**Figure 6C**) and similar results were also obtained in the HT-29 cell line (**Figure S11C-D**). Accordingly, c-MYC and ADSL protein levels significantly and positively correlated in CRC cell lines (Pearson r = 0.99, p = 0.004; **Figure 6D**). Supporting our *in vitro* data, *ADSL*-high tumors are significantly enriched in the CRC molecular subtype CMS2 (**Figure S2A**), characterized by a strong upregulation of *MYC* targets.
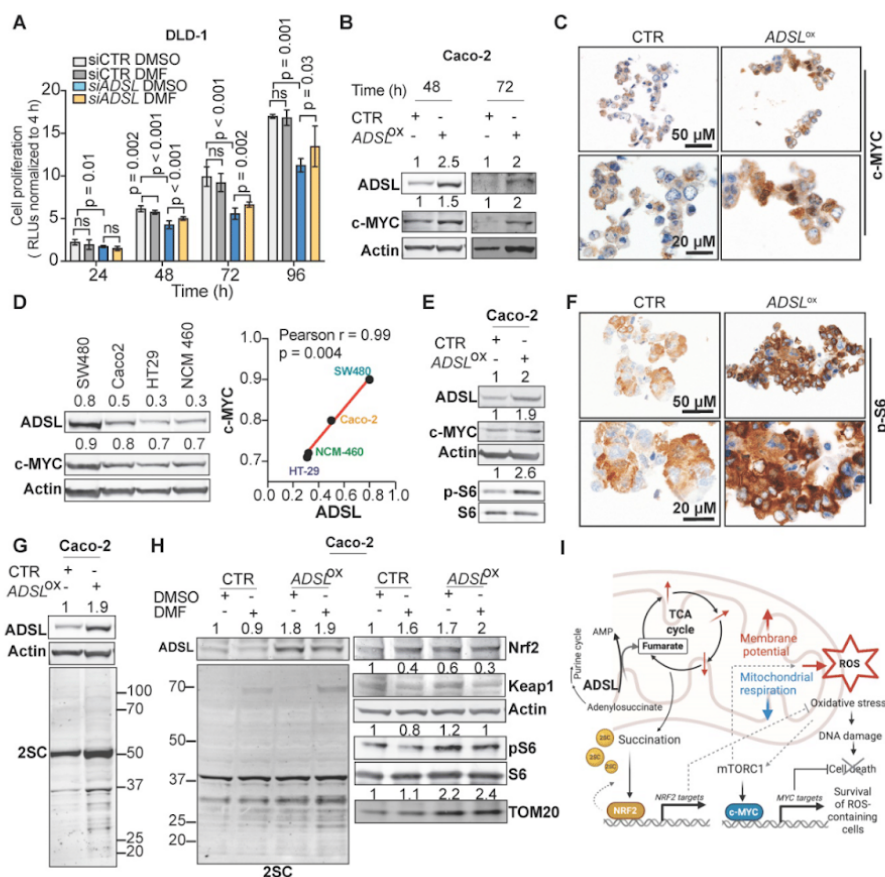
Figure 6. ADSL overexpression induces succination and mTOR activation and increases c-MYC protein expression. (**A**) Proliferation capacity of DLD-1 control and *ADSL*-silenced cells treated with DMSO or fumarate (50 µM). (**B**) Immunoblot showing ADSL and c-MYC expression in the Caco-2 cells at 48 and 72 h post transfection. (**C**) Representative pictures of control and *ADSL*-overexpressing Caco-2 cells immunostained with c-MYC. Scale bars 20-50 µm. (**D**) Immunoblot showing ADSL and c-MYC expression in CRC cell lines (left). Correlation (linear regression) between c-MYC (Y-axis) and ADSL (X-axis) levels of expression (relative to actin). (**E**) Immunoblot showing ADSL, c-MYC and phospho and total S6 expression in Caco-2 cells 48 h post-transfection. (**F**) Representative pictures of control and *ADSL*-overexpressed Caco-2 cells immunostained with phospho-S6 (p-S6). Scale bars 20-50 µm. (**G**) Immunoblot showing ADSL and succination (2-SC) levels in Caco-2 cells 48 h post-transfection. (**H**) Immunoblot showing ADSL and succination (2-SC) levels (left) and KEAP1, NRF2, TOM20, total and phospho-S6 in control and *ADSL*-overexpressing Caco-2 cells treated with DMSO or fumarate (50 µM). (**I**) Schematic representation of ADSL-driven pro-oncogenic effects in CRC cells.

Next, we asked through which mechanism *ADSL* overexpression was able to induce c-MYC. It was previously shown that ADSL levels modulate Akt phosphorylation in endometrial cancer and that TCA cycle defects, specifically the accumulation of α-ketoglutarate (αKG), as well as ROS increase can activate the mTOR signaling pathway [7,50–53]. mTORC1, a major regulator of cellular metabolism and energetic state, has been shown to regulate *MYC* mRNA translation [54,55]. Here we observed that S6 phosphorylation, a marker of mTORC1 activation, increased upon *ADSL* overexpression in both Caco-2 and HT-29 cells (**Figure 6E-F** and **Figure S11E**). Similarly, *ADSL* knock-down reduced phospho-S6 and c-MYC levels in DLD-1 cells (**Figure S11F**).

Fumarate has been shown to modify thiol groups in several proteins by forming *S*-(2-succinyl) Cys (2SC) adducts, a process termed as succination [56]. Succination of KEAP1, a repressor of the transcription factor nuclear factor, erythroid 2 like 2 (NFE2L2/NRF2), is known to promote NRF2 stabilization and nuclear translocation [57], with consequent induction of stress response genes [57–59]. Using an anti-2SC antibody we tested whether *ADSL* overexpression, via increasing Krebs cycle intermediates such as fumarate, would increase succination in CRC cells. Indeed, we detected higher 2SC levels in *ADSL*-overexpressing Caco-2 cells compared to control (**Figure 6G**), while silencing *ADSL* reduced succination to the endogenous level (**Figure S11G**). Modulation of *ADSL* had similar effects on 2SC levels in DLD-1 cells (**Figure S11H**). The addition of fumarate also increased 2SC levels in Caco-2, however the succination pattern slightly differed from that induced by *ADSL* overexpression (**Figure 6H**, left panel). As expected, 2SC levels correlated with reduced KEAP1 levels and NRF2 upregulation upon both the addition of fumarate and *ADSL* overexpression (**Figure 6H**, right panel). However, fumarate alone did not induce S6 phosphorylation nor c-MYC up-regulation, indicating that ADSL overexpression activates the mTOR-MYC axis through mechanisms other than fumarate accumulation (**Figure 6H** and **Figure S11I**). Additionally, only *ADSL* overexpression and not fumarate increased the expression of the mitochondrial outer membrane protein TOM20, a common marker of mitochondrial mass and/or biogenesis (**Figure 6H**), which is also a typical downstream effect mediated by c-MYC activation [60].

Taken together our data indicate that ADSL overexpression independently activates the NRF2 stress response pathway and the mTORC1-cMYC axis. Activation of both NRF2 and mTORC1-cMYC may help protect CRC cells from oxidative stress while allowing cell survival and proliferation of DNA damage-prone tumor clones (**Figure 6I**).

Discussion

In the present study, we identified ADSL as a novel putative oncogene in CRC. ADSL is a key enzyme of *de novo* purine biosynthesis [4,5]. Increased levels of ADSL have been observed in several cancer types [6–8] and have been shown to increase proliferation, migration, and invasive capability of endometrial and triple-negative breast cancer cell lines [6,7]. Here, we have demonstrated that ADSL is upregulated in CRC at both the mRNA and protein levels. However, while our TMA cohort of CRC primary tumors suggested that ADSL expression was reduced in advanced-stage tumors, we did not observe any difference in ADSL expression in our second TMA cohort consisting of primary CRC tumors and CRC metastases. On the mRNA level, *ADSL* expression levels also showed a downward trend with increasing stages, albeit with very small differences in expression levels between tumors of different stages. Although it is unclear whether ADSL levels are genuinely reduced in advanced-stage CRCs, our results suggest that ADSL plays an oncogenic role in CRC initiation rather than tumor progression, which would be consistent with the upregulated *ADSL* activity in colon pre-neoplastic lesions [15].

By modulating the expression levels of *ADSL in vitro,* we showed that *ADSL* promotes proliferation and migration in CRC cells, as well as tumor growth *in vivo* in the CAM model. Additionally, we showed that ADSL regulates cell cycle progression. Indeed, silencing of *ADSL* blocked cells in the S-phase while upon forced expression of *ADSL*, cells progressed faster through the cell cycle. Accordingly, *ADSL*-overexpressing CRCs showed overexpression of DNA synthesis and cell-cycle related pathways. In particular, we found that *ADSL* overexpression correlates with the expression of many genes coding for DNA polymerases and S-phase signaling checkpoints, many of which are part of a "DNA replication signature" found dysregulated in colorectal cancer [40]. These findings demonstrate the oncogenic effect of *ADSL* and emphasize that *ADSL* may be a potential therapeutic target in CRC.

The pyrimidine analogue 5-FU is one of the most commonly used chemotherapeutic agents for CRC [9,13,61]. Given the essential role of *ADSL* in DNA synthesis and cell cycle, we investigated its potential as a predictive biomarker of response to 5-FU *in vitro* and *ex vivo. ADSL* overexpression did not significantly affect response to 5-FU in CRC cells, but it partially affected response in CRC-PDOs. The discordant

results may reflect the differential cellular response between traditional two-dimensional cell culture models derived from single clones compared to PDOs that better represent tumor heterogeneity [62], underlining the importance of using multiple preclinical models for accurate drug-response prediction. The pivotal role played by *ADSL* in *de novo* purine biosynthesis led us to hypothesize that a purine analogue rather than a pyrimidine analogue may be more effective on *ADSL*-overexpressing CRCs. Indeed, our data strongly indicate that *ADSL* overexpression significantly sensitizes CRCs to 6-MP both *in vitro* and *in vivo*. Additionally, we showed that CRC-PDOs with high ADSL expression respond better to 6-MP compared to ADSL-low PDOs. As a chemotherapeutic agent, 6-MP is used in the treatment of acute lymphoblastic leukemia [63]. In the context of gastrointestinal diseases, 6-MP as an immunosuppressant is one of the cornerstones of treatment in inflammatory bowel diseases [64,65], but there is a substantial lack of studies on the use of 6-MP in chemotherapy regimens in CRC. To our knowledge, only one study has previously investigated the potential use of 6-MP as a chemotherapeutic agent in CRC [66], with no detection of any substantial clinical benefit. We speculate that patient stratification based on *ADSL* expression may help better dissect the potential benefit of 6-MP in the treatment of CRC. Unfortunately, due to the lack of clinical studies evaluating the effect of 6-MP in CRC patients, we cannot draw any conclusion regarding the predictive value of ADSL expression in response to 6-MP treatment in the overall survival or disease-free survival in patients.

Using targeted metabolomics and transcriptomics analysis we showed that ADSL overexpression mainly affects mitochondrial function, leading to the accumulation of Krebs cycle intermediates and altered oxidative phosphorylation. In particular, we demonstrated that ADSL overexpression lowered the oxygen consumption rate (OCR) in CRC cells. Reduced cellular respiration is a common strategy used by cancer cells to survive in hypoxic conditions [67]. Although impairment in oxidative metabolism is usually associated with increased glycolysis in cancer [67], we found that ADSL overexpression impairs glycolysis as well, thus suggesting that other metabolic pathways might fuel the energetic and anabolic demands of *ADSL*-overexpressing CRC cells, such as fatty acid metabolism that was also found significantly up-regulated upon *ADSL* overexpression.

Fumarate is a TCA cycle intermediate and a product of ADSL catalyzed reactions. Fumarate has been shown to inhibit both mitochondrial respiration [48] and glycolysis [49], as well as to mediate ADSL oncogenic properties in other cancer types [7]. Although the forced expression of *ADSL* increased fumarate abundance which we postulated contributed to the significant impairment of mitochondrial respiration and glycolysis in the parental CRC cells, treatment with exogenous fumarate only partially mimicked the metabolic phenotype induced by *ADSL* overexpression. Our data indicate that dysregulation of additional metabolites or target genes is required for the full oncogenic potential of *ADSL* in CRC. Indeed, we also found that *ADSL* overexpression, but not fumarate, induces mTORC1 activation and c-MYC protein increase in CRC cell lines. Our results are in line with the current literature reporting that ADSL indirectly modulates Akt phosphorylation and c-MYC activation in other cancer types [6,7]. Both mTOR and c-MYC are well-known master regulators of cellular metabolism and have been shown to regulate the TCA cycle and mitochondrial function, thus driving metabolic rewiring in cancer [68–70]. In particular, both mTOR and c-MYC can stimulate the synthesis of mitochondria-targeted proteins and promote mitochondrial biogenesis [60,70], which could partially explain the upregulation of nuclear-encoded mitochondrial genes. Decreased mitochondrial respiration coupled with increased membrane potential can result in ROS accumulation. Indeed, *ADSL* overexpression increased mitochondrial ROS levels in Caco-2 cells and triggered the accumulation of DNA-damage, as proved by the overexpression of γH2AX. Given that defects in both the TCA cycle and ROS accumulation can activate the mTOR signaling pathway [7,50-53], we speculate that, upon *ADSL* overexpression, the accumulation of ROS leads to the activation of the mTORC1-cMYC axis. mTOR and c-MYC in turn induce mitochondrial biogenesis through a positive feedback loop which exacerbates the phenotype. Our current data cannot discriminate whether the c-MYC increase is induced by mTORC1 activation or vice versa, as well as to what extent mitochondrial dysfunction with subsequent ROS generation causes, or is caused by, the activation of the mTORC1-cMYC pathway. Additional experiments are required to elucidate in more detail how *ADSL* overexpression drives these mechanisms.

Fumarate accumulation is also known to cause protein succination. Succination of regulatory proteins such as KEAP1 and p62, causes the activation of the NRF2-mediated response, which promotes tumorigenesis by enhancing ROS detoxification. Indeed, both fumarate and *ADSL* overexpression induced higher levels of protein succination as well as

153

the activation of the KEAP1-NRF2 pathway. While the KEAP1-Nrf2 pathway might appear as an independent pathway, the mTORC1-cMYC and the KEAP1-Nrf2 pathway are tightly interconnected and act synergistically in cancer initiation driven by ROS accumulation. Indeed, it has been shown that activation of both NRF2 stress-response and mTORC1-cMYC axis are required for the survival and expansion of ROS-abundant cells in the early stages of hepatocarcinogenesis [71]. Additionally, while NRF2 is a regulator of redox homeostasis in quiescent cells, in the presence of an active PI3K-Akt pathway, it can also drive the expression of genes involved in glutamine metabolism and the pentose phosphate pathway [72]. We therefore hypothesized that by inducing mitochondrial dysfunction and ROS generation, ADSL acts as an oncogene in CRC by up-regulating both NRF2 and mTORC1-cMYC axis.

In conclusion, our investigation highlights the multifaceted role of *ADSL* as a new oncogene in colorectal cancers and strongly supports a role for *ADSL* overexpression in sensitizing tumor cells to 6-MP. Specifically, we demonstrated that *ADSL* overexpression in CRC: 1) induces dysregulation of the Krebs cycle and mitochondrial dysfunction, with consequent 2) activation of the mTOR-cMYC pathway, and 3) NRF2 stress response. Our results show that ADSL overexpression is pleiotropic in CRC, in the sense that it induces metabolic and mitochondrial dysfunction with consequent oxidative stress and ROS accumulation via a series of interconnected pathways, which favors the survival of stressed CRC cells and enable the accumulation of DNA damage.

Abbreviations

2SC: *S*-(2-succinyl) Cys; 5-FU: 5-fluorouracil; 6-MP: 6-mercaptopurine; ADSL: adenylosuccinate lyase; AICAR: 5-Aminoimidazole-4-carboxamide-1-β-d-ribofuranoside; AJCC: American Joint Committee on Cancer; CAM: chicken chorioallantoic membrane; CIN: chromosomal instability; Cl. Caspase: cleaved caspase; CMS: consensus molecular subtype; CRC: colorectal carcinoma; DAB: 3,3'-Diaminobenzidine; DAPI: 4',6-diamidino-2-phenylindole; DAPI-A: 4',6-diamidino-2-phenylindole-area; dAMP: deoxyadenosine monophosphate; DMEM: Dulbecco's Modified FEagle Medium; DMSO: dimethyl sulfoxide; DRIVE: deep RNAi interrogation of viability effects in cancer; FACS: fluorescence-activated cell sorting; FBS: fetal bovine serum; FDR: false discovery rate; FSC-A: forward scatter area; FSC-W: forward scatter width; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; GSEA: gene set enrichment analysis; H&E: hematoxylin and eosin; IHC: immunohistochemistry; IMP: inosine monophosphate; KRT20: keratin 20; KLRC3: natural killer cells lectin-like receptor C3; NES: normalized enrichment score; PAS: periodic acid-Schiff; PCR: polymerase chain reaction; PBS: phosphate-buffered saline; PDO: patient-derived organoids; PFA: Paraformaldehyde; PI: propidium iodide; RTCA: Real-Time Cell Analyzer; RT-PCR: reverse transcription polymerase chain reaction; SAICAR: phosphoribosylaminoimidazolesuccinocarboxamide; SD: standard deviation; SSC-A: side scatter-area; TCGA: The Cancer Genome Atlas; TMA: tissue microarray; TRG: tumor regression grade.

Supplementary Material

Supplementary figures and tables.
http://www.thno.org/v11p4011s1.pdf

154

Competing Interests

The authors have declared that no competing interest exists.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144: 646–74.
2. Keller KE, Tan IS, Lee Y-S. SAICAR stimulates pyruvate kinase isoform M2 and promotes cancer cell survival in glucose-limited conditions. Science. 2012; 338: 1069–72.
3. Wang X, Yang K, Xie Q, et al. Purine synthesis promotes maintenance of brain tumor initiating cells in glioma. Nat Neurosci. 2017; 20: 661–73.
4. Jurecka A, Zikanova M, Kmoch S, Tylki-Szymańska A. Adenylosuccinate lyase deficiency. J Inherit Metab Dis. 2015; 38: 231–42.
5. Gooding JR, Jensen MV, Dai X, et al. Adenylosuccinate Is an Insulin Secretagogue Derived from Glucose-Induced Purine Metabolism. Cell Rep. 2015; 13: 157–67.
6. Zurlo G, Liu X, Takada M, et al. Prolyl hydroxylase substrate adenylosuccinate lyase is an oncogenic driver in triple negative breast cancer. Nat Commun. 2019; 10: 5177.
7. Park H, Ohshima K, Nojima S, et al. Adenylosuccinate lyase enhances aggressiveness of endometrial cancer by increasing killer cell lectin-like receptor C3 expression by fumarate. Lab Invest. 2018; 98: 449–61.
8. Bardot V, Dutrillaux AM, Delattre JY, et al. Purine and pyrimidine metabolism in human gliomas: relation to chromosomal aberrations. Br J Cancer. 1994; 70: 212–8.
9. Vodenkova S, Buchler T, Cervena K, Veskrnova V, Vodicka P, Vymetalkova V. 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. Pharmacol Ther. 2020; 206: 107447.
10. Luengo A, Gui DY, Vander Heiden MG. Targeting Metabolism for Cancer Therapy. Cell Chem Biol. 2017; 24: 1161–80.
11. Longley DB, Paul Harkin D, Johnston PG. 5-Fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer. 2003; 3: 330-8.
12. Sargent D, Sobrero A, Grothey A, et al. Evidence for cure by adjuvant therapy in colon cancer: observations based on individual patient data from 20,898 patients on 18 randomized trials. J Clin Oncol. 2009; 27: 872–7.
13. Christensen S, Van der Roest B, Besselink N, et al. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. Nat Commun. 2019; 10: 4571.
14. Touil Y, Igoudjil W, Corvaisier M, et al. Colon cancer cells escape 5FU chemotherapy-induced cell death by entering stemness and quiescence associated with the c-Yes/YAP axis. Clin Cancer Res. 2014; 20: 837–46.
15. Terzuoli L, Carlucci F, Martino AD, et al. Determination of p185 and adenylosuccinate lyase (ASL) activity in preneoplastic colon lesions and intestinal mucosa of human subjects. Clin Biochem. 1998; 31: 523–8.
16. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487: 330–7.
17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139–40.
18. [Internet] Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. 2016. https://github.com/ctlab/fgsea
19. Sato T, Stange DE, Ferrante M, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology. 2011; 141: 1762–72.
20. Fujii M, Shimokawa M, Date S, et al. A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. Cell Stem Cell. 2016 Jun 2; 18:827-838.
21. van de Wetering M, Francies HE, Francis JM, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. Cell. 2015; 161: 933–45.
22. Zijlstra A, Mellor R, Panzarella G, et al. A quantitative analysis of rate-limiting steps in the metastatic cascade using human-specific real-time polymerase chain reaction. Cancer Res. 2002; 62: 7083–92.
23. Lauzier A, Normandeau-Guimond J, Vaillancourt-Lavigueur V, et al. Colorectal cancer cells respond differentially to autophagy inhibition *in vivo*. Sci Rep. 2019; 9: 11316.
24. Mesker WE, Junggeburt JMC, Szuhai K, et al. The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. Cell Oncol. 2007; 29: 387–98.
25. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001; 25: 402–8.
26. Roci I, Gallart-Ayala H, Schmidt A, et al. Metabolite Profiling and Stable Isotope Tracing in Sorted Subpopulations of Mammalian Cells. Anal Chem. 2016; 88: 2707-13.
27. Ivanisevic J, Zhu Z-J, Plate L, et al. Toward 'Omic Scale Metabolite Profiling: A Dual Separation–Mass Spectrometry Approach for Coverage of Lipid and Central Carbon Metabolism. Anal Chem. 2013; 85: 6876-84.
28. Gallart-Ayala H, Konz I, Mehl F, et al. A global HILIC-MS approach to measure polar human cerebrospinal fluid metabolome: Exploring
29. gender-associated variation in a cohort of elderly cognitively healthy subjects. Anal Chim Acta. 2018; 1037: 327-337.
29. Mock A, Warta R, Dettling S, Brors B, Jäger D, Herold-Mende C. MetaboDiff: An R package for differential metabolomic analysis. Bioinformatics. 2018; 34: 3417–8.
30. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559.
31. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems. 2015; 6: 417-425.
32. Wickham, H. Ggplot2: Elegant Graphics for Data Analysis. New York, USA; Springer; 2009.
33. Gubser PM, Bantug GR, Razik L, et al. Rapid effector function of memory CD8+ T cells requires an immediate-early glycolytic switch. Nat Immunol. 2013; 14: 1064–72.
34. [Internet] Torsten H. maxstat: Maximally Selected Rank Statistics. 2017. https://CRAN.R-project.org/package=maxstat
35. McDonald ER 3rd, de Weck A, Schlabach MR, et al. Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell. 2017; 170: 577–92.e10.
36. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018; 173: 291–304.e6.
37. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015; 21: 1350–6.
38. Stupack DG, Teitz T, Potter MD, et al. Potentiation of neuroblastoma metastasis by loss of caspase-8. Nature. 2006; 439: 95–9.
39. Hagedorn M, Javerzat S, Gilges D, et al. Accessing key steps of human tumor progression *in vivo* by using an avian embryo model. Proc Natl Acad Sci U S A. 2005; 102: 1643–8.
40. Pillaire M-J, Selves J, Gordien K, et al. A "DNA replication" signature of progression and negative outcome in colorectal cancer. Oncogene. 2010; 29: 876-87.
41. Furuta T, Takemura H, Liao Z-Y, et al. Phosphorylation of histone H2AX and activation of Mre11, Rad50, and Nbs1 in response to replication-dependent DNA double-strand breaks induced by mammalian DNA topoisomerase I cleavage complexes. J Biol Chem. 2003; 278: 20303–12.
42. Colombo SL, Palacios-Callender M, Frakich N, et al. Molecular basis for the differential use of glucose and glutamine in cell proliferation as revealed by synchronized HeLa cells. Proc Natl Acad Sci U S A. 2011; 108: 21069–74.
43. Ganesh K, Wu C, O'Rourke KP, et al. A rectal cancer organoid platform to study individual responses to chemoradiation. Nat Med. 2019; 25: 1607–14.
44. Buzzelli JN, Ouaret D, Brown G, Allen PD, Muschel RJ. Colorectal cancer liver metastases organoids retain characteristics of original tumor and acquire chemotherapy resistance. Stem Cell Res. 2018; 27: 109–20.
45. Ooft SN, Weeber F, Dijkstra KK, et al. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. Sci Transl Med. 2019; 11: eaay2574
46. Murphy MP. How mitochondria produce reactive oxygen species. Biochemical Journal. 2009; 417; 1-13.
47. Kitazawa M, Hatta T, Sasaki Y, et al. Promotion of the Warburg effect is associated with poor benefit from adjuvant chemotherapy in colorectal cancer. Cancer Sci. 2020; 111: 658–66.
48. Tyrakis PA, Yurkovich ME, Sciacovelli M, et al. Fumarate Hydratase Loss Causes Combined Respiratory Chain Defects. Cell Rep. 2017; 21: 1036–47.
49. Kornberg MD, Bhargava P, Kim PM, et al. Dimethyl fumarate targets GAPDH and aerobic glycolysis to modulate immunity. Science. 2018; 360: 449–53.
50. Carbonneau M, M Gagné L, Lalonde M-E, et al. The oncometabolite 2-hydroxyglutarate activates the mTOR signalling pathway. Nat Commun. 2016; 7: 12700.
51. Yalcin S, Marinkovic D, Mungamuri SK, et al. ROS-mediated amplification of AKT/mTOR signalling pathway leads to myeloproliferative syndrome in Foxo3(-/-) mice. EMBO J. 2010; 29: 4118–31.
52. Kim J-H, Choi TG, Park S, et al. Mitochondrial ROS-derived PTEN oxidation activates PI3K pathway for mTOR-induced myogenic autophagy. Cell Death Differ. 2018; 25: 1921–37.
53. Li M, Zhao L, Liu J, et al. Multi-mechanisms are involved in reactive oxygen species regulation of mTORC1 signaling. Cell Signal. 2010; 22: 1469–76.
54. Wall M, Poortinga G, Hannan KM, Pearson RB, Hannan RD, McArthur GA. Translational control of c-MYC by rapamycin promotes terminal myeloid differentiation. Blood. 2008; 112: 2305–17.
55. Sodi VL, Khaku S, Krutilina R, et al. mTOR/MYC Axis Regulates O-GlcNAc Transferase Expression and O-GlcNAcylation in Breast Cancer. Mol Cancer Res. 2015; 13: 923–33.
56. Bardella C, El-Bahrawy M, Frizzell N, et al. Aberrant succination of proteins in fumarate hydratase-deficient mice and HLRCC patients is a robust biomarker of mutation status. J Pathol. 2011; 225: 4–11.
57. Kinch L, Grishin NV, Brugarolas J. Succination of Keap1 and activation of Nrf2-dependent antioxidant pathways in FH-deficient papillary renal cell carcinoma type 2. Cancer Cell. 2011; 20: 418–20.
58. Adam J, Hatipoglu E, O'Flaherty L, et al. Renal Cyst Formation in Fh1-Deficient Mice Is Independent of the Hif/Phd Pathway: Roles for Fumarate in KEAP1 Succination and Nrf2 Signaling. Cancer cell. 2011; 20: 524-37.

59. Ooi A, Wong J-C, Petillo D, et al. An Antioxidant Response Phenotype Shared between Hereditary and Sporadic Type 2 Papillary Renal Cell Carcinoma. Cancer Cell. 2011; 20; 511-23.

60. Morrish F, Hockenbery D. MYC and Mitochondrial Biogenesis. Cold Spring Harbor Perspectives in Medicine. 2014; 4; a014225.

61. Bracht K, Nicholls AM, Liu Y, Bodmer WF. 5-Fluorouracil response in a large panel of colorectal cancer cell lines is associated with mismatch repair deficiency. British Journal of Cancer. 2010; 103: 340-6.

62. Nagle PW, Plukker JTM, Muijs CT, van Luijk P, Coppes RP. Patient-derived tumor organoids for prediction of cancer treatment response. Semin Cancer Biol. 2018; 53: 258–64.

63. Bostrom B, Erdmann G. Cellular pharmacology of 6-mercaptopurine in acute lymphoblastic leukemia. Am J Pediatr Hematol Oncol. 1993; 15: 80–6.

64. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE. ACG Clinical Guideline: Management of Crohn's Disease in Adults. Am J Gastroenterol. 2018; 113: 481–517.

65. Munnig-Schmidt E, Zhang M, Mulder CJ, Barclay ML. Late-onset Rise of 6-MMP Metabolites in IBD Patients on Azathioprine or Mercaptopurine. Inflamm Bowel Dis. 2018; 24: 892–6.

66. Lokich J, Moore C, Anderson N, et al. Phase I–II trial of 14 day infusional 6-Mercaptopurine in advanced colorectal cancer. Eur J Cancer. 1998; 34: 584-5.

67. DeBerardinis RJ, Chandel NS. We need to talk about the Warburg effect. Nat. Metabolism. 2020; 2: 127-129.

68. Anderson NM, Li D, Peng HL., et al. The TCA cycle transferase DLST is important for MYC-mediated leukemogenesis. Leukemia. 2016; 30: 1365–74.

69. Wise DR, DeBerardinis RJ, Mancuso A, et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proc Natl Acad Sci U S A. 2008; 105: 18782–7.

70. Morita M, Gravel S-P, Hulea L, et al. mTOR coordinates protein synthesis, mitochondrial activity and proliferation. Cell Cycle. 2015; 14: 473–80.

71. Umemura A, He F, Taniguchi K, et al. p62, Upregulated during Preneoplasia, Induces Hepatocellular Carcinogenesis by Maintaining Survival of Stressed HCC-Initiating Cells. Cancer Cell. 2016; 29: 935–48.

72. Mitsuishi Y, Taguchi K, Kawatani Y, et al. Nrf2 redirects glucose and glutamine into anabolic pathways in metabolic reprogramming. Cancer Cell. 2012; 22: 66–79.

73. Wu M, Neilson A, Swift AL, et al. Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. Am J Physiol Cell Physiol. 2007; 292: C125–36.

Transcriptional Enhancer Factor Domain Family member 4 Exerts an Oncogenic Role in Hepatocellular Carcinoma by Hippo-Independent Regulation of Heat Shock Protein 70 Family Members

Mairene Coto-Llerena [ID],[1,2*] Nadia Tosti,[1*] Stephanie Taha-Mehlitz,[2,3] Venkatesh Kancherla,[1] Viola Paradiso,[1] John Gallon [ID],[2] Gaia Bianco,[2] Andrea Garofoli,[1] Souvik Ghosh [ID],[4] Fengyuan Tang,[5] Caner Ercan,[1] Gerhard M. Christofori,[5] Matthias S. Matter,[1] Raoul A. Droeser,[3] Mihaela Zavolan,[4] Savas D. Soysal,[3] Markus von Flüe,[3] Otto Kollmar,[3] Luigi M. Terracciano,[1] Charlotte K. Y. Ng [ID],[6] and Salvatore Piscuoglio [ID][1,2]

Transcriptional enhancer factor domain family member 4 (TEAD4) is a downstream effector of the conserved Hippo signaling pathway, regulating the expression of genes involved in cell proliferation and differentiation. It is up-regulated in several cancer types and is associated with metastasis and poor prognosis. However, its role in hepatocellular carcinoma (HCC) remains largely unexplored. Using data from The Cancer Genome Atlas, we found that TEAD4 was overexpressed in HCC and was associated with aggressive HCC features and worse outcome. Overexpression of TEAD4 significantly increased proliferation and migration rates in HCC cells *in vitro* as well as tumor growth *in vivo*. Additionally, RNA sequencing analysis of *TEAD4*-overexpressing HCC cells demonstrated that *TEAD4* overexpression was associated with the up-regulation of genes involved in epithelial-to-mesenchymal transition, proliferation, and protein-folding pathways. Among the most up-regulated genes following *TEAD4* overexpression were the 70-kDa heat shock protein (HSP70) family members *HSPA6* and *HSPA1*A. Chromatin immunoprecipitation–quantitative real-time polymerase chain reaction experiments demonstrated that TEAD4 regulates *HSPA6* and *HSPA1A* expression by directly binding to their promoter and enhancer regions. The pharmacologic inhibition of HSP70 expression in *TEAD4*-overexpressing cells reduced the effect of TEAD4 on cell proliferation. Finally, by overexpressing *TEAD4* in yes-associated protein (*YAP*)/transcriptional coactivator with PDZ binding motif (*TAZ*)-knockdown HCC cells, we showed that the effect of TEAD4 on cell proliferation and its regulation of HSP70 expression does not require YAP and TAZ, the main effectors of the Hippo signaling pathway. *Conclusion:* A novel Hippo-independent mechanism for TEAD4 promotes cell proliferation and tumor growth in HCC by directly regulating HSP70 family members. (*Hepatology Communications* 2021;5:661-674).

Transcriptional enhancer factor domain (TEAD) proteins are a family of transcription factors that bind the consensus 5′-CATTCCA/T-3′ sequence through their shared TEA DNA binding domain.[1] TEAD-regulated transcription strictly depends on the binding of TEAD transcription factors with various coactivators,[2,3] among which are the yes-associated protein (YAP) and transcriptional

coactivator with PDZ binding motif (TAZ), two major effectors of the Hippo signaling pathway,[4,5] and the vestigial-like protein 1 (VGLL1).[6] In mammals, TEADs are highly conserved and widely expressed[7,8] and play a pivotal role in development by mediating cell proliferation and organ size control through the Hippo signaling pathway.[9] In cancer cells, TEADs regulate proliferation, migration, differentiation, epithelial-to-mesenchymal transition (EMT), apoptosis, and invasion.[10] In particular, TEAD factors regulate the expression of progrowth factors, such as connective tissue growth factor (CTGF),[10,11] cysteine-rich angiogenic inducer 61 (Cyr61),[11] AXL receptor tyrosine kinase (AXL),[12] MYC proto-oncogene bHLH transcription factor (Myc),[13] baculoviral inhibitor of apoptosis repeat-containing 5 (survivin),[13] and insulin-like growth factor binding protein 5 (IGFBP5).[6] Furthermore, overexpression of TEADs and their association with a poor clinical outcome have been reported in many cancer types, including breast, lung, prostate, colon, and gastric cancers as well as melanoma and glioblastoma.[14-16]

TEAD family member 4 (TEAD4), the gene encoding for the transcriptional enhancer factor (TEF)-3, is overexpressed in several tumor types, such as breast and gastric cancers,[16] as well as hepatoblastoma, the most common type of pediatric liver cancer.[17] Notably, TEAD4 acts as an oncogene through Hippo signaling-dependent[18] and signaling-independent mechanisms.[6,15] In hepatocellular carcinoma (HCC), it is known that TEAD4 acts through the Hippo signaling pathway, wherein the TEAD4/YAP complex cooperates with forkhead box M1 (FOXM1) in inducing chromosome instability.[19] It has also been shown that up-regulation of the YAP2/TEAD4 axis by sirtuin 1 (SIRT1) deacetylation of YAP2 promotes HCC cell proliferation.[20] Furthermore, TEAD4/YAP and hepatocyte nuclear factor 4α (HNF4α) regulate hepatocarcinogenesis by reciprocal repression in mice and rats.[21] Thus, until now, TEAD4 was thought to play a YAP-dependent role in HCC oncogenesis.

In this study, we show that TEAD4 is overexpressed in a subset of HCC and that it promotes cell proliferation in HCC cells both *in vitro* and *in vivo*. We further show that TEAD4 directly regulates HSP70 in a Hippo-independent mechanism.

ARTICLE INFORMATION:

From the [1]Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland; [2]Visceral Surgery and Precision Medicine Research Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland; [3]Clarunis, Department of Visceral Surgery, University Center for Gastrointestinal and Liver Diseases, St. Clara Hospital and University Hospital Basel, Basel, Switzerland; [4]Computational and Systems Biology, Biozentrum; [5]Department of Biomedicine, University of Basel, Basel, Switzerland; [6]Department for BioMedical Research, University of Bern, Bern, Switzerland.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Salvatore Piscuoglio, Ph.D.
Visceral Surgery and Precision Medicine Research Laboratory
Department of Biomedicine, University of Basel
Hebelstrasse, 20

Basel 4031, Switzerland
E-mail: s.piscuoglio@unibas.ch
Tel.: +41 61 328 68 74

Materials and Methods

CELL LINES

HCC-derived cell lines (HepG2, SNU449, HLE, and Huh7) were maintained in a 5% CO_2-humidified atmosphere at 37°C and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin (Bio-Concept, Allschwil, Switzerland), and 1% minimal essential medium–nonessential amino acids (ThermoFisher Scientific, Basel, Switzerland). Stable *YAP*- and *TAZ*-knockdown Huh7 and SNU449 cell lines (described in Tang et al.[22]) and their respective controls were maintained in complete medium supplemented with puromycin (1 μg/mL). All cell lines were confirmed negative for mycoplasma infection using the polymerase chain reaction (PCR)-based Universal Mycoplasma Detection kit (American Type Culture Collection, Manassas, VA). For HSP70 inhibitor 1 (KNK437; CAS 218924-25-5; Calbiochem, Sigma-Aldrich, St. Louis, MO), cells were incubated with 100 μM of the inhibitor and the corresponding dimethyl sulfoxide (DMSO) control.

HUMAN SAMPLES

Ten human, unselected, nonconsecutive HCCs retrieved from the archives of the Institute of Medical Genetics and Pathology at the University Hospital Basel (Basel, Switzerland) were included in this study. These 10 samples had already been prescreened for HSP70 protein expression. The study was performed in accordance with the Helsinki Declaration and approved by the Ethics Committee of Basel. Data were collected retrospectively in a nonstratified and nonmatched manner, including patient age, tumor diameter, location, prognostic tumor/prognostic node stage, grade, histologic subtype, and vascular invasion.

IMMUNOHISTOCHEMISTRY

Immunohistochemistry (IHC) was performed as described.[23] Details are presented in the Supporting Information.

PLASMIDS AND TRANSFECTION

For *TEAD4* overexpression, pLV[Exp]-enhanced green fluorescent protein (EGFP)/Neo-EF1A> hTEAD4 was designed and ordered on the vector builder website (https://en.vectorbuilder.com/), and the empty control vector was plasmid cytomegalovirus (pCMV)-mir EGFP. For *TEAD4* silencing, pSuper-retro-puro empty vector and pSuper-retro-puro sh*TEAD*1/3/4 were adapted from Zhao et al.[11] The expression vectors were transiently transfected using the jetPRIME transfection reagent (Polyplus, Illkirch, France) following the manufacturer's instructions. The expression of the plasmids was evaluated by western blot and quantitative reverse-transcription (qRT)-PCR analysis. Cells were harvested 48 hours after transfection for further experiments.

PROTEIN EXTRACTION AND WESTERN BLOT RNA EXTRACTION AND qRT-PCR AND RNA SEQUENCING

Protein extraction and western blot RNA extraction along with qRT-PCR and RNA sequencing (RNA-seq) are detailed in the Supporting Information.

PROLIFERATION AND MIGRATION ASSAYS

Proliferation and cell migration assays were performed using the xCELLigence Real-Time Cell Analysis-dual purpose (ACEA Biosciences, San Diego, CA) system. All experiments were performed in triplicate. Results are shown as mean ± SD. Statistical significance was assessed by the *t* test.

WOUND-HEALING ASSAY

Twenty-four hours after transfection, a 100-μL micropipette tip was used to create a scratch in the cell monolayer in each well of the six-well plate. To monitor the wound closure, representative phase-contrast images of each well were taken at 0, 8, 12, and 24 hours after the scratch was made. Each experimental condition was evaluated in triplicate. Statistical significance was assessed by the *t* test.

CHORIOALLANTOIC MEMBRANE ASSAY

Fertilized chicken eggs were obtained from a local hatchery (Gepro Geflügelzucht AG, Flawil,

663

159

Switzerland) at day 1 of gestation and were incubated at 37°C with 60% humidity for 10 days. The cells were harvested 24 hours after transfection, suspended (2 × 10⁶ cells per chorioallantoic membrane [CAM] assay) in 10 μL of medium (DMEM), mixed 1:1 with matrigel (Matrigel Matrix; Ref. 354234; Corning, Tewksbury, MA), and grafted onto the CAM of 9-day-old chicken embryos. The chicken embryos were maintained *in ovo* at 37°C for another 4 days, followed by removal of the tumors. Pictures of each tumor were taken with a Canon EOS 1100D digital camera. Tumor size measurements were performed by averaging the volume (height × width × width) of each tumor, using ImageJ as described.[24]

CELLTITER-GLO CELL VIABILITY ASSAY

CellTiter-Glo (G7573; Promega, Dübendorf, Switzerland) was used to determine the number of viable cells based on adenosine triphosphate content. Twenty-four hours after transfection, cells were seeded in a 96-well plate. After 8 hours, drug treatment was added to the cells and cell viability was measured by adding 100 μL of CellTiter-Glo/well at 24, 48, and 72 hours posttreatment. Statistical significance was assessed by multiple *t* test. In sorafenib experiments, results were normalized to DMSO. Curve fitting was performed using Prism software (GraphPad Software, San Diego, CA) and the nonlinear regression equation.

ANALYSIS OF CHROMATIN IMMUNOPRECIPITATION SEQUENCING DATA

Chromatin immunoprecipitation (ChIP)-seq data for TEAD4 in the HepG2 human liver cell line were produced using mouse monoclonal immunoglobulin G (IgG) raised against recombinant protein TEAD4 and were obtained from the Gene Expression Omnibus (accession number GSM1010875).[25] Processed data as an hg19 bigWig file were loaded into Integrative Genomics Viewer,[26] and the "Find Motif" tool was used to search for the most conserved bases of *TEAD4* binding motif within regions of *TEAD4* ChIP-seq peaks. TEAD4 binding motif was obtained from the JASPAR 2018 database.[27] TEAD4 enhancers were called from the *TEAD4* broadPeak BED file using the Cognitive Reliability and Error Analysis Method (CREAM),[28] with the options WScutoff = 1.5, Minlength = 1000, and peakNumMin = 2. CREAM calls enhancers based on clusters of peaks, taking into account the distribution of distances between peaks in ChIP-seq data from a given sample.

ChIP

Cells from four 10-cm Petri dishes at 70%-80% confluence were crosslinked in 1% formaldehyde for 10 minutes with continuous shaking. The crosslinking was stopped by adding 0.15 M glycine while continuing shaking. After collecting the cells by scraping, the pellet was washed 3 times with cold phosphate-buffered saline. The nuclei of the cells were isolated and lysed. Chromatin shearing was performed using Bioruptor Pico (Diagenode, Liège, Belgium). The number of cycles and the settings were as described.[29] At the same time, the antibody was coupled with magnetic protein G beads (100-03D; Invitrogen, Carlsbad, CA) by incubating 75 μL of protein G beads with 10 μg of TEAD4 antibody (TEF-3, sc-101184; Santa Cruz Biotechnology, Dallas, TX) or 10 μg of mouse IgG (sc-2025; Santa Cruz Biotechnology) as a negative control for 1 hour with constant rotation. At the end of the sonication process, an aliquot of the chromatin was kept as input control for every sample and an equal amount of sonicated chromatin was incubated with the previously produced antibody-coupled magnetic beads at 4°C overnight while rotating. The samples were washed and then eluted (all buffers are as described in the original protocol from Blecher-Gonen et al.[30]). Ribonuclease treatment and then proteinase K treatment were performed on all samples, followed by overnight reverse crosslinking at 65°C with continuous shaking. DNA purification was performed using Agencourt AMPure XP (A63880; Beckman Coulter, Brea, CA). TEAD4 abundance on specific target gene promoters was quantified by qRT-PCR compared to IgG negative control. Primer sequences are listed in Supporting Table S1.

STATISTICAL ANALYSIS

Statistical analyses were performed using GraphPad Prism version 6.0 and R. For *in vitro* and *in vivo* studies, statistical significance was determined by the

two-tailed unpaired Student t test. Differences were considered statistically significant at $P < 0.05$. All experiments were performed at least twice. The statistical parameters (i.e., the exact value of n, P values) are noted in the figures and figure legends. Results are shown as mean ± SD.

Analysis of The Cancer Genome Atlas Liver Data Set

Analysis of The Cancer Genome Atlas (TCGA) liver data set is described in the Supporting Information.

Results

TEAD4 OVEREXPRESSION PROMOTES HCC TUMOR GROWTH *IN VITRO* AND *IN VIVO*

As *TEAD4* overexpression has been associated with poor clinical outcome in a number of cancer entities, we asked whether *TEAD4* overexpression plays a similar role in HCC. We first confirmed that *TEAD4* was overexpressed in HCC using the TCGA data set. Compared to the nontumoral livers, HCCs showed *TEAD4* overexpression ($P < 0.001$, Mann-Whitney U test) in 21% (n = 73/371) of cases (Supporting Fig. S1A,B). Additionally, we observed that *TEAD4* expression was associated with aggressive HCC features, such as high Edmondson grade, high stage, and histologic changes, more frequently observed in severe diseases, such as the presence of necrosis, pleomorphic, and multinucleated cells ($P < 0.05$; Supporting Fig. S1C). Additionally, *TEAD4*-overexpressing tumors were more enriched in Hoshida molecular subtype S1, which is associated with higher proliferation and worse outcome (Supporting Fig. S1C). Furthermore, we observed that *TEAD4* overexpression was associated with worse overall survival and was independently prognostic in patients with HCC in univariate and multivariate models (hazard ratio, 1.52; 95% confidence interval, 1.00-2.29) (Supporting Fig. S1D; Supporting Table S2).

Having determined that *TEAD4* is indeed overexpressed in HCC, we modulated the expression of *TEAD4* in HCC cell lines, examined its effect on cell growth, and further exploited these models to dissect the possible oncogenic role of *TEAD4*. Using a complementary DNA construct, we overexpressed *TEAD4* in two HCC cell lines with low endogenous levels of TEAD4 expression (Huh7 and SNU449; Supporting Fig. S2A,B). We further used shorthairpin RNAs to silence *TEAD4* in two HCC cell lines with high endogenous levels of *TEAD4* expression (HLE and HepG2; Supporting Fig. S2A,B). We observed that *TEAD4* overexpression significantly increased proliferation as well as cell migration in both Huh7 and SNU449 cells (Fig. 1A,B; Supporting Fig. S2C,D). On the contrary, *TEAD4*-silenced HLE and HepG2 cells significantly proliferated and migrated less compared to control cells (Fig. 1A,B; Supporting Fig. S2C,D). The wound-healing assay further confirmed the role of TEAD4 in regulating the migration potential of HCC cells. Indeed, forced expression of *TEAD4* in Huh7 led to an increase in the gap closure rate, while *TEAD4* silencing slowed the speed of gap closure in HLE cells ($P < 0.01$, Mann-Whitney U test) (Fig. 1C). Finally, because sorafenib is still the primary choice for first-line treatment of patients with HCC, we tested if the modulation of TEAD4 expression might affect the response to sorafenib. Our results showed that overexpression or silencing of *TEAD4* did not alter the sensitivity of the cells to sorafenib (Supporting Fig. S2E).

To further demonstrate the oncogenic role of TEAD4 in HCC, we xenotransplanted *TEAD4*-overexpressing and control Huh7 cells into the chicken CAM and assessed tumor growth *in vivo* (Fig. 2A). Engraftment of tumor cells in the CAM has been successfully used as a model of tumorigenesis.[31] Indeed, it has been shown that this densely vascularized extraembryonic tissue enables a fast, reproducible, and precise analysis of the principal steps of tumor progression and is a powerful tool for preclinical *in vivo* studies.[32] Briefly, 24 hours posttransfection, cells were harvested and resuspended in matrigel before being seeded into the CAMs. After 4 days, eggs were screened for tumor formation and tumors were harvested for quantification and further analysis. In accordance with the results obtained *in vitro*, *TEAD4*-overexpressing engrafted cells formed significantly larger tumors compared to control cells (Fig. 2B,C; $P < 0.05$, Mann-Whitney U test). IHC and western blot analysis confirmed that the resected tumors were indeed of human origin and that *TEAD4* overexpression could still be detected 5 days posttransfection (Fig. 2D).

665
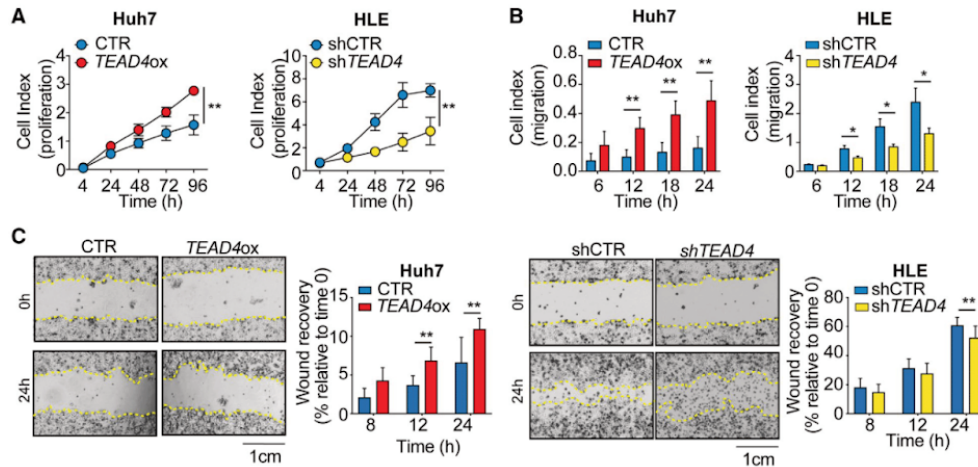
FIG. 1. Modulation of *TEAD4* expression level impacts cell proliferation and migration. (A) Proliferation and (B) migration kinetics of *TEAD4*-overexpressing (*TEAD4ox*) Huh7 and *TEAD4*-silenced (sh*TEAD4*) HLE cells compared to their respective controls. (C) Wound-healing assay in *TEAD4*-overexpressing Huh7 and *TEAD4*-silenced HLE cells compared to their respective controls at specific time points. Representative pictures at time points 0 and 24 hours after scratch are shown for each condition. Scale bar in (C), 1 cm. Statistical significance was determined by the *t* test; *$P < 0.05$, **$P < 0.01$. Data are mean ± SD. Abbreviations: h, hours; sh, short hairpin.

Taken together, our *in vitro* and *in vivo* data provide compelling evidence of the oncogenic role of TEAD4 in HCC. In particular, we demonstrate the role of TEAD4 in promoting proliferation and migration of liver cancer cells.

TEAD4 REGULATES CELL DIFFERENTIATION AND PROLIFERATION AND THE EXPRESSION OF HSP70 FAMILY GENES

To define the transcriptional changes induced by *TEAD4* overexpression, we performed RNA-seq on Huh7 cells overexpressing *TEAD4* (n = 4) or control vector (n = 4; Fig. 3A). Differential expression analysis identified 569 up-regulated and 316 down-regulated genes following *TEAD4* overexpression (quasi-likelihood F test, adjusted $P < 0.05$) (Fig. 3B; Supporting Table S3). Gene set enrichment analysis (GSEA)[33] revealed the up-regulated genes were involved in pathways, such as myogenesis, EMT, and the P53 pathway, all associated with cell differentiation and proliferation (Fig. 3C). Conversely, we observed a down-regulation of gene sets

involved in metabolism, such as protein secretion and bile acid metabolism (Fig. 3C). These results are consistent with the aggressive phenotype we observed in our *in vitro* and *in vivo* models following *TEAD4* overexpression as well as with the well-known role of TEAD proteins in regulating EMT and proliferation.[10]

In-depth analysis of the RNA-seq data revealed altered expression of the HSP family members, including some small HSPs (*HSPB9* and *HSPB1*), HSP40 (*DNAJA4*), and HSP70 (*HSPA1A, HSPA1B,* and *HSPA6*). Furthermore, the set of HSP70 genes defined by Klimczak et al.[34] was found to be significantly enriched in the *TEAD4*-overexpressing Huh7 cells ($P < 0.05$, GSEA; Fig. 3D), while other HSP gene sets (HSP40 and HSP90) were not significantly altered. In line with this, GSEA using Gene Ontology (GO) pathways showed enrichment of gene sets downstream of the HSP genes, such as those involved in HSP-binding protein-folding chaperones and unfolded protein binding (Fig. 3E). Surprisingly, the canonical Hippo targets seem not to be up-regulated in *TEAD4*-overexpressing cells, suggesting a potential Hippo-independent role of *TEAD4* in regulating HSP70 family members.

666
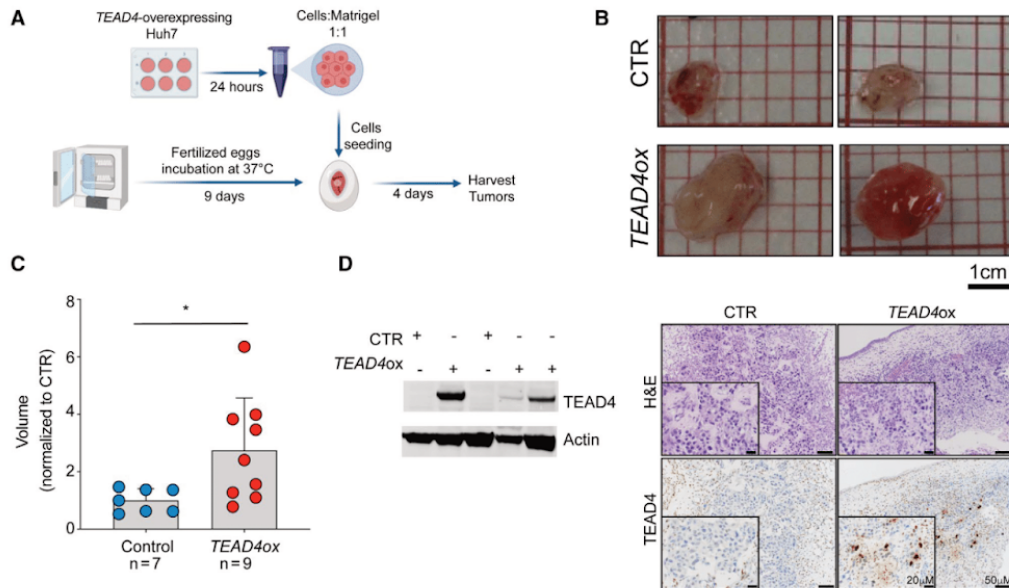
FIG. 2. *TEAD4* overexpression increases tumor growth *in vivo*. (A) Schematic representation of the CAM assay. (B) Photographs of *TEAD4*-overexpressing or control Huh7 cells implanted in CAMs and grown for 4 days postimplantation. (C) Volume of tumors derived from the CAM experiment (n ≥ 7 tumors over two independent experiments). Values are normalized to the mean volume of control. (D) TEAD4 expression in Huh7 tumors extracted 4 days postimplantation analyzed by western blot (left) and IHC (right). Tumoral cells were immunostained with the TEAD4 antibody. Scale bars, (B) 1 cm and (D) 50 μm and 20 μm. Statistical significance was determined by the Mann-Whitney U test; *P < 0.05. Data are mean ± SD. Abbreviations: H&E, hematoxylin and eosin; ox, overexpression.

Among the HSP70 genes, *HSPA1A* and *HSPA6* were the most significantly up-regulated genes following *TEAD4* overexpression (Fig. 3B). Using qPCR, we confirmed that overexpression of *TEAD4* led to a significant increase in *HSPA6 and HSPA1A* expression in our *in vitro* models (Huh7 and SNU449; Fig. 3F). Our results suggested that HSP70 is a novel family of TEAD4 target genes in HCC.

TEAD4 DIRECTLY REGULATES THE TRANSCRIPTION OF HSP70 FAMILY GENES

The HSP70 family comprises 13 gene products that differ from each other by expression level, subcellular location, and amino acid composition. Following exposure to different stressors, HSP70s bind to misfolded proteins and prevent their aggregation. As HSP70 family members and their overexpression have been shown to play an oncogenic role in different cancer types,[35,36] including HCC,[37,38] we asked if HSP70 family genes were direct transcriptional targets of TEAD4. Because *HSPA6* and *HSPA1A* were the most up-regulated genes following *TEAD4* overexpression (Fig. 3B), we focused on these two members of the family. We analyzed a set of peak calls from TEAD4 ChIP-seq performed on the HepG2 cell line.[25] Interestingly, we observed a TEAD4 peak within the *HSPA6* promoter region, ~100 base pairs (bp) upstream of the transcription starting site (TSS), that contained two regions with the TEAD4 binding motif (chr1:161493541-161493545 and chr1:161493648-161493652; Fig. 4A). Additional analysis of the peak calls using CREAM[28] identified a cluster of TEAD4 peaks ~50 kilobases (kb) upstream of the *HSPA6* promoter (i.e., a TEAD4 enhancer, chr1:31,776,368-31,832,367; Fig. 4A) and another enhancer ~47 kb downstream of the *HSPA1A* promoter (chr6:31,830,291-38,133,978; Fig. 4A), both well

667

163

within the 100-kb range of most promoter-enhancer interactions.[39] Peaks in both enhancers and at the gene promoters were also confirmed to contain the TEAD4 binding motif (Fig. 4A).

To confirm that TEAD4 directly binds to the promoter region of the *HSPA6* gene and the enhancer region of *HSPA1A*, we performed ChIP in *TEAD4*-overexpressing Huh7 cells, using anti-TEAD4
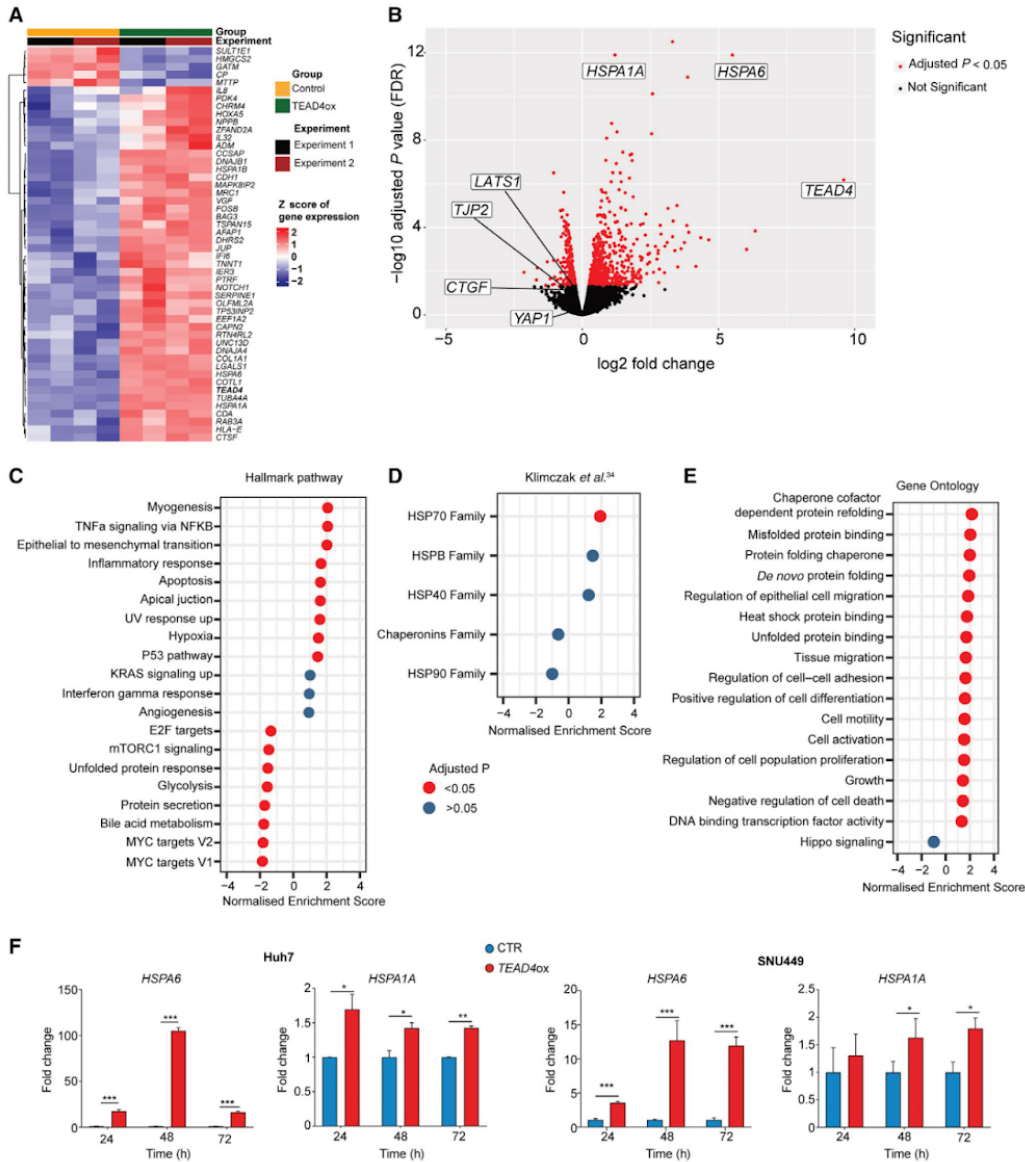
FIG. 3. TEAD4 regulates cell differentiation and proliferation and the expression of HSP70 family genes. (A) Heatmap showing the top 50 differentially expressed genes in *TEAD4*-overexpressing Huh7 cells. (B) Volcano plot of −log10 adjusted *P* value against log2 fold change to illustrate the differentially expressed genes between *TEAD4*-overexpressing Huh7 cells and control cells. Red dots indicate differentially expressed genes (adjusted *P* < 0.05). (C-E) GSEA plots show selected significantly enriched gene sets in *TEAD4*-overexpressing cells: Hallmark gene sets from (C) MSigDB database, (D) HSP family gene set, and (E) GO gene sets from the MSigDB database. (F) The Huh7 cell line was transiently transfected with a vector overexpressing *TEAD4* and the corresponding control. RNA was extracted at different time points. mRNA expression levels of *HSPA1A* (left) and *HSPA6* (right) were measured by qRT-PCR using *GAPDH* as the internal control. Data are shown as fold change relative to control. Results represent three independent experiments. Statistical significance was determined by the Mann-Whitney U test; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. Data are mean ± SD. Abbreviations: E2F, E2 factor; FDR, false discovery rate; *GAPDH*, glyceraldehyde 3-phosphate dehydrogenase; LATS1, large tumor suppressor kinase 1; mRNA, messenger RNA; mTORC1, mammalian target of rapamycin complex 1; NFKB, nuclear factor kappa B; ox, overexpression; TJP2, tight junction protein 2; TNFa, tumor necrosis factor alpha; UV, ultraviolet.

....

antibody or IgG followed by qPCR. Using primers spanning the predicted TEAD4 binding region on the *HSPA6* promoter, the *HSPA1A* enhancer region, and the known TEAD4 target *CTGF* promoter,[7] we confirmed that TEAD4 indeed binds to the *HSPA6* promoter and *HSPA1A* enhancer region (Fig. 4B).

To corroborate the results obtained in *vitro* we performed IHC staining on 10 HCC samples using TEAD4 and HSP70 antibodies. Analysis of this cohort revealed that samples positive for TEAD4 expression were also positive for HSP70 and vice versa ($P = 0.06$, Spearman correlation test; Fig. 4C). Of these 10 samples, eight were positive for HSP70 and two were negative. Six out of eight HCCs positive for HSP70 were also positive for TEAD4 (Fig. 4C), while the two HCCs negative for HSP70 were also negative for TEAD4.

Taken together, our results demonstrate that *HSPA6* and *HSPA1A* are transcriptional targets of TEAD4 in HCC through both TEAD4-promoter and TEAD4-enhancer interactions.

ONCOGENIC EFFECTS OF TEAD4 ARE MEDIATED THROUGH ITS UP-REGULATION OF HSP70

Given that the HSP70 family genes are direct transcriptional targets of TEAD4 and that HSP70s are known to play an oncogenic role in various cancer types, including HCC,[37] we hypothesized that the oncogenic effects of TEAD4 may be at least partially mediated by its up-regulation of HSP70. We therefore inhibited HSP70 pharmacologically in *TEAD4*-overexpressing cells with the benzylidene lactam compound KNK473 (Fig. 4C). KNK437 has been shown to inhibit the induction of HSPs, including HSP70, HSP40, and HSP105, *in vitro*.[40]

Additionally, KNK437 was shown to revert E2 factor (E2F) transcription factor 1-mediated *HSP40* induction, which plays a role in promoting colorectal cancer tumor growth and metastasis *in vitro* and *in vivo*.[41] Supporting our hypothesis, treatment of *TEAD4*-overexpressing Huh7 cells with 100 μM of KNK437 abolished the effect of *TEAD4* overexpression on cell proliferation at 48 and 72 hours posttreatment (Fig. 4D). No significant reduction on cell proliferation was observed in control cells following treatment with KNK437.

Taken together, our results demonstrate that TEAD4 promotes tumor growth in HCC at least partially through direct regulation of the HSP70 family members. We further showed that the pharmacologic inhibition of HSP70 induction reverted the increased cell proliferation phenotype induced by TEAD4 overexpression.

TEAD4 REGULATES HSP70 AND CELL PROLIFERATION INDEPENDENT OF HIPPO SIGNALING

The transcriptional activity of TEAD proteins is highly dependent on the binding of their C-terminal protein interaction domain to several coactivators.[42,43] Among those cofactors, YAP and TAZ,[44,45] two major Hippo signaling pathway transcriptional coactivators, are the most well studied. When the Hippo signaling pathway is active, YAP and TAZ proteins are located in the nucleus, thus binding to TEADs and inducing transcription of TEAD downstream targets (Fig. 5A). Conversely, when the Hippo pathway is inactive, phosphorylation of YAP and TAZ allows TEADs to bind to other cofactors and activate alternative transcriptional targets (Fig. 5A).[46] The role of TEAD4 in HCC

669

165
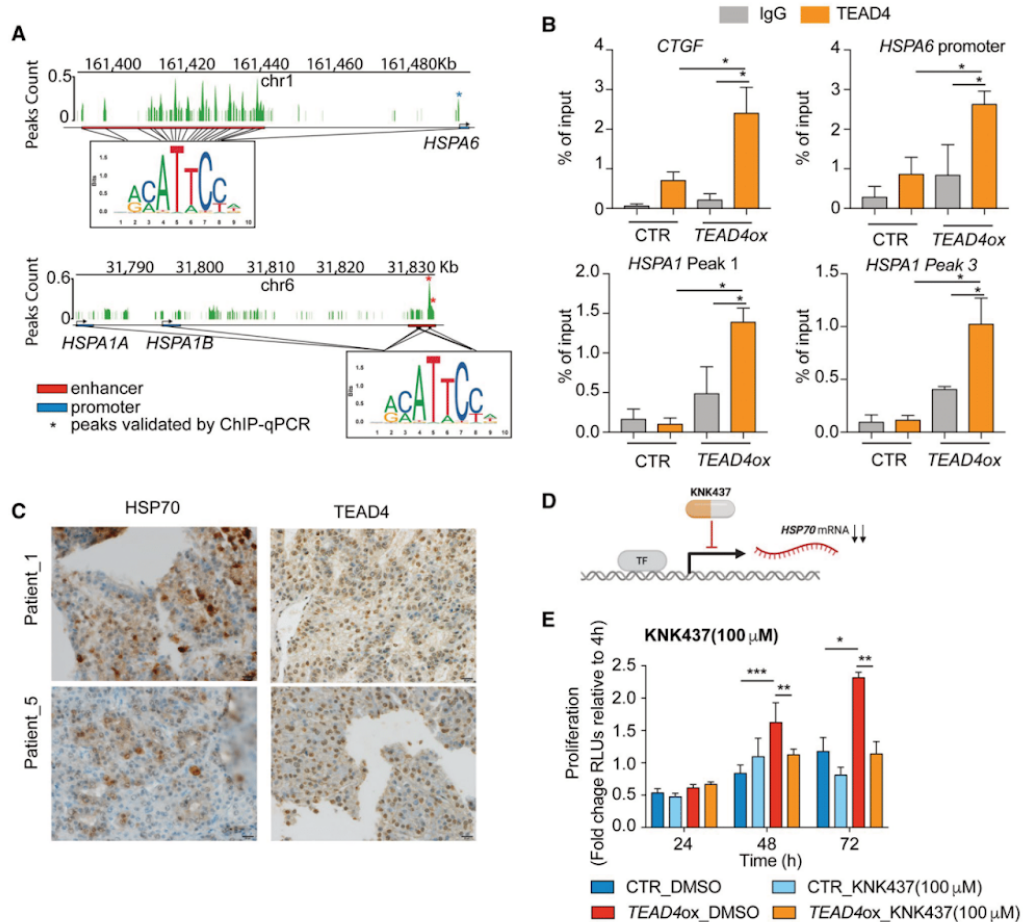
FIG. 4. TEAD4 impacts cell proliferation by directly binding HSP70 promoters and associated enhancers. (A) TEAD4 ChIP-seq peaks at the promoters of *HSPA6* and *HSPA1A,* and regions called as TEAD4 enhancers within 100 kb of *HSPA6* and *HSPA1A,* with TEAD4 binding motifs labeled. Enhancers and promoters are labeled as red and blue blocks, respectively. The TEAD4 binding motif logos are shown below the track, and positions in the *HSPA6* and in *HSPA1A* enhancers harboring this motif are indicated (this motif was also found at the promoter of all three genes). Asterisks indicate peaks validated by ChIP-qPCR. (B) ChIP-qPCR showing enrichment for TEAD4 binding at the *HSPA6* promoter and the *HSPA1A* enhancer. *CTGF* was used as a positive control. (C) Representative immunostain of TEAD4 and HSP70 on human HCCs. Magnification ×40; scale bar 20 μm. (D) Schematic representation of the mechanisms of action of KNK437 on HSP70. (E) Proliferation of *TEAD4*-overexpressing Huh7 cells treated with 100 μM of KNK437 at various time points, relative to proliferation at 4 hours. DMSO was used as control. Results represent three independent experiments. Statistical significance was determined by the *t* test; *P < 0.05, **P < 0.01, ***P < 0.001. Data are mean ± SD. Abbreviations: chr, chromosome; h, hours; mRNA, messenger RNA; ox, overexpression; RLU, relative light unit; TF, transcription factor.

carcinogenesis has so far been linked to the Hippo signaling pathway.[19,21] Interestingly, analysis of RNA-seq data from *TEAD4*-overexpressing Huh7 cells revealed no significant change in well-known targets and/or effectors of the Hippo signaling pathway (e.g., *YAP1*, large tumor suppressor kinase 1 [*LATS1*], tight junction protein 2 [*TJP2*], and *CTGF*; Fig. 3B) or in the Hippo signaling pathway as a whole (*q* > 0.01 GO; Fig. 3E),
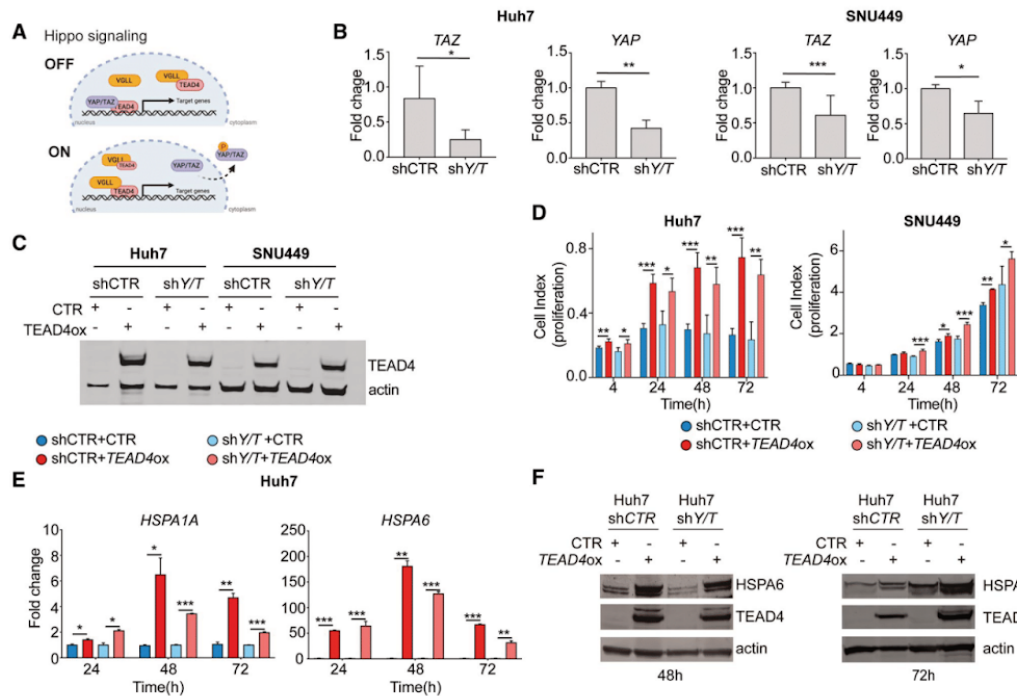
FIG. 5. The effect of TEAD4 on cell proliferation is independent of expression of cofactors YAP and TAZ. (A) Schematic representation of *TEAD4* transcriptional activity associated with YAP/TAZ cofactor expression. (B) Expression levels of *YAP* and *TAZ* were measured in stable *YAP/TAZ*-knockdown (sh*Y/T*) and control (shCTR) Huh7 and SNU449 cell lines by qRT-PCR normalized to *GAPDH* as internal control. (C) TEAD4 expression in stably *TEAD4*-overexpressing and control Huh7 and SNU449 cell lines with knockdown of *YAP/TAZ* or control (shCTR). (D) Proliferation kinetics of stable *TEAD4*-overexpressing and control Huh7 cell lines with *YAP/TAZ* knockdown or control (shCTR) (left) and in *TEAD4*-overexpressing and control SNU449 stable cell lines with *YAP/TAZ* knockdown and control (shCTR) (right). (E) *HSPA6* and *HSPA1A* expression levels in TEAD4-overexpressing and control Huh7 cell lines with knockdown of *YAP/TAZ* or control (shCTR) were measured by qRT-PCR using *GAPDH* as reference. (F) Expression levels of *HSPA6* in *TEAD4*-overexpressing and control Huh7 cell lines with knockdown of *YAP/TAZ* or control (shCTR) were measured by western blot at 48 and 72 hours posttransfection. Data are shown as fold change to control. Results represent three independent experiments. Statistical significance was determined by *t* test; *P < 0.05, **P < 0.01, ***P < 0.001. Data are mean ± SD. Abbreviations: *GAPDH*, glyceraldehyde 3-phosphate dehydrogenase; h, hours; ox, overexpression; sh, short hairpin; VGLL, vestigial-like protein 1.

suggesting that the oncogenic properties of *TEAD4* in HCC are independent of the Hippo pathway.

To test the hypothesis that TEAD4 does not require Hippo signaling to promote cell proliferation in liver cancer cells, we overexpressed *TEAD4* in stable *YAP/TAZ*-knockdown and control liver cancer cell lines (Supporting Fig. S3A,B; Fig. 5B,C). We first confirmed that the stable *YAP/TAZ*-knockdown cells had impaired Hippo signaling, as indicated by reduced *Cyr61, CTGF,* and vimentin expression (Supporting Fig. S3A,B). When *TEAD4* was overexpressed

(Supporting Fig. S3C), we observed that the proliferation rate was significantly increased in both stable *YAP/TAZ*-knockdown and control Huh7 and SNU449 cells (*P* < 0.001 and *P* < 0.05, respectively; Fig. 5D), suggesting that the presence or absence of the cofactors YAP and TAZ did not alter how *TEAD4* modulates cell proliferation *in vitro*.

Finally, to confirm that TEAD4 regulates HSP70 expression independent of Hippo signaling, we measured *HSPA1A* and *HSPA6* expression in the stable *YAP/TAZ*-knockdown and control cells. In line with

671

167

our RNA-seq analysis, *HSPA1A* and *HSPA6* were up-regulated following *TEAD4* overexpression in both the stable *YAP/TAZ*-knockdown and the control cells (Fig. 5E; Supporting Fig. S3D), demonstrating that TEAD4 regulates HSP70 expression independent of its YAP and TAZ coactivators. Additionally, we also performed western blot analysis to evaluate if the changes observed at the RNA level were also observed at the protein level. Indeed, we observed that the overexpression of *TEAD4* in cells with knockdown of YAP/TAZ led to an increased level of HSP70 at the protein level (Fig. 5F).

Our results suggest that the oncogenic effects of *TEAD4* overexpression in HCC are, at least in part, independent of the expression of the Hippo signaling YAP and TAZ coactivators.

Discussion

TEAD4 is a member of the TEAD transcriptional enhancer factor family composed of four members (TEAD1-4). Deregulation of *TEAD4* expression, as with other members of the family, has been extensively reported in several tumor types.[16,47,48] In the present study, we showed that TEAD4 promotes cell proliferation and tumor growth in HCC by directly regulating HSP70 family members through, at least in part, a Hippo-independent mechanism.

After confirming the overexpression of *TEAD4* in a subset of HCC in the TCGA data set, we evaluated the functional relevance of *TEAD4* up-regulation in HCCs by performing a series of *in vitro* and *in vivo* experiments. We demonstrated that *TEAD4* overexpression promoted tumor growth, cell proliferation, and migration in liver cancer cells while *TEAD4* silencing had the opposite effect. Our observations are consistent with the oncogenic role played by TEAD4 in other tumor types. Indeed, it has been shown that down-regulation of *TEAD4* expression hampers cancer cell proliferation and invasiveness in *in vitro* and *in vivo* models of colorectal and gastric cancers.[48,49] Our findings demonstrate that TEAD4 also acts as an oncogene in HCC.

As a transcription factor, TEAD4 regulates the transcription of many genes, activating or repressing several downstream pathways, such as cell growth, differentiation, and apoptosis.[11] Consistent with the role of TEAD4 in cell differentiation, the gene expression profile of *TEAD4*-overexpressing Huh7 cells showed up-regulation of pathways, such as spermatogenesis, EMT, and myogenesis. Surprisingly, chaperone folding and HSP pathways were also found to be significantly enriched in *TEAD4*-overexpressing cells. Specifically, we identified two members of the HSP70 family (*HSPA6 and HSPA1A*) as among the most up-regulated genes. HSP70B and HSP70-1, the protein products of *HSPA6* and *HSPA1A*, respectively, are members of the ubiquitous and highly conserved HSP70 family of molecular chaperones. In normal cell conditions, chaperones are expressed at very low levels, ensuring the correct folding and transport of newly synthesized proteins.[50] However, chaperone levels increase during cell cycle and development[51] as well as in response to cellular stress as occurs, for example, in tumors. Overexpression of HSP70 family members has been shown to play an oncogenic role in different cancer types,[35,36] including HCC.[37,38] In HCC, HSP70 has been reported to drive cell migration,[52] and genetic ablation of HSP70 was able to markedly impair chemically induced liver tumorigenesis and tumor progression.[38] Notably, HSP70 is a clinically important marker in HCC diagnosis; HSP70 expression, together with that of glutamine synthetase and glypican 3, is commonly used to differentiate early and low-grade tumors from dysplastic nodules.[53]

Our reanalysis of the ChIP-seq data from the HepG2 cell line[25] revealed two mechanisms by which TEAD4 can regulate the expression of HSP70 family genes; first by binding sites mapped less than 1 kb from the TSS of *HSPA6* and second through TEAD4 enhancer regions ~50 kb upstream of the *HSPA6* promoter and ~47 kb downstream of *HSPA1A*. Both putative TEAD4 enhancers and the HSP70 promoter regions were found to harbor TEAD4 binding motif, and we confirmed the direct binding of TEAD4 at these loci by ChIP-qPCR in *TEAD4*-overexpressing Huh7 cells. In fact, we demonstrated that the increased cell proliferation resulting from TEAD4 overexpression was reversed by the pharmacologic inhibition of HSP70. We note that although KNK437 is a nonspecific HSP70 inhibitor, it has been shown to inhibit only inducible HSPs but has no effect on the expression of constitutively expressed HSP family members, including HSC70 and HSP90.[40] Therefore, although KNK437 is not specific to HSP70, it would only inhibit HSPs induced following *TEAD4*

672

168

overexpression. Our results thus indicate that the oncogenic effect of TEAD4 in HCC acts through its regulation of HSP70.

The transcription factor TEAD4 and its coactivators YAP and TAZ are considered major downstream effectors of the conserved Hippo signaling pathway.[18] However, in colorectal and prostate cancers,[6,15] TEAD4 has been found to play a role in EMT and cell proliferation independent of YAP/TAZ expression. Here, we demonstrated that *TEAD4* overexpression increased liver cancer cell proliferation independent of the expression of YAP/TAZ. Supporting our YAP/TAZ-knockdown experiments, our RNA-seq data from *TEAD4*-overexpressing cells revealed no significant changes on Hippo signaling following *TEAD4* overexpression. Furthermore, we showed that *TEAD4* overexpression induced *HSPA1A* and *HSPA6* expression in the context of YAP/TAZ knockdown. Together, our results indicate that TEAD4 regulates HSP70 expression and liver cancer cell proliferation independent of Hippo signaling.

In conclusion, we showed that TEAD4 plays an oncogenic role in HCC and that its oncogenic effect is mediated in part by its regulation of HSP70 genes. More importantly, our results unveil a novel role of TEAD4 outside its canonical Hippo-dependent mechanism.

REFERENCES

1) Bürglin TR. The TEA domain: a novel, highly conserved DNA-binding motif. Cell 1991;66:11-12.
2) Pobbati AV, Hong W. Emerging roles of TEAD transcription factors and its coactivators in cancers. Cancer Biol Ther 2013;14:390-398.
3) Xiao JH, Davidson I, Matthes H, Garnier JM, Chambon P. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell 1991;65:551-568.
4) Chen L, Chan SW, Zhang X, Walsh M, Lim CJ, Hong W, et al. Structural basis of YAP recognition by TEAD4 in the hippo pathway. Genes Dev 2010;24:290-300.
5) **Li Z**, **Zhao B**, Wang P, Chen F, Dong Z, Yang H, et al. Structural insights into the YAP and TEAD complex. Genes Dev 2010;24:235-240.
6) Pobbati AV, Chan SW, Lee I, Song H, Hong W. Structural and functional similarity between the Vgll1-TEAD and the YAP-TEAD complexes. Structure 2012;20:1135-1140.
7) Benhaddou A, Keime C, Ye T, Morlon A, Michel I, Jost B, et al. Transcription factor TEAD4 regulates expression of myogenin and the unfolded protein response genes during C2C12 cell differentiation. Cell Death Differ 2012;19:220-231.
8) Joshi S, Davidson G, Le Gras S, Watanabe S, Braun T, Mengus G, et al. TEAD transcription factors are required for normal primary
9) Zhao B, Li L, Lei Q, Guan K-L.. The Hippo-YAP pathway in organ size control and tumorigenesis: an updated version. Genes Dev 2010;24:862-874.
10) **Zhang H**, **Liu C-Y**, Zha Z-Y, Zhao B, Yao J, Zhao S, et al. TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition. J Biol Chem 2009;284:13355-13362. Erratum in: J Biol Chem 2019;294:5808.
11) Zhao B, Ye X, Yu J, Li L, Li W, Li S, et al. TEAD mediates YAP-dependent gene induction and growth control. Genes Dev 2008;22:1962-1971.
12) **Xu MZ**, **Chan SW**, Liu AM, Wong KF, Fan ST, Chen J, et al. AXL receptor kinase is a mediator of YAP-dependent oncogenic functions in hepatocellular carcinoma. Oncogene 2011;30:1229-1240.
13) Lu L, Li Y, Kim SM, Bossuyt W, Liu P, Qiu Q, et al. Hippo signaling is a potent in vivo growth and tumor suppressor pathway in the mammalian liver. Proc Natl Acad Sci U S A 2010;107:1437-1442.
14) Knight JF, Shepherd CJ, Rizzo S, Brewer D, Jhavar S, Dodson AR, et al. TEAD1 and c-Cbl are novel prostate basal cell markers that correlate with poor clinical outcome in prostate cancer. Br J Cancer 2008;99:1849-1858.
15) Liu Y, Wang G, Yang Y, Mei Z, Liang Z, Cui A, et al. Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial–mesenchymal transition and metastasis in a YAP-independent manner. Oncogene 2016;35:2789-2800.
16) **Zhou Y**, **Huang T**, Cheng ASL, Yu J, Kang W, To KF. The TEAD family and its oncogenic role in promoting tumorigenesis. Int J Mol Sci 2016;17:138.
17) Zhang J, Liu P, Tao J, Wang P, Zhang Y, Song X, et al. TEA domain transcription factor 4 is the major mediator of yes-associated protein oncogenic activity in mouse and human hepatoblastoma. Am J Pathol 2019;189:1077-1090.
18) Yu F-X, Zhao B, Guan K-L. Hippo pathway in organ size control, tissue homeostasis, and cancer. Cell 2015;163:811-828.
19) Weiler SME, Pinna F, Wolf T, Lutz T, Geldiyev A, Sticht C, et al. Induction of chromosome instability by activation of yes-associated protein and forkhead box M1 in liver cancer. Gastroenterology 2017;152:2037-2051.e22.
20) Mao B, Hu F, Cheng J, Wang P, Xu M, Yuan F, et al. SIRT1 regulates YAP2-mediated cell proliferation and chemoresistance in hepatocellular carcinoma. Oncogene 2014;33:1468-1474.
21) Cai W-Y, Lin L-Y, Hao H, Zhang S-M, Ma F, Hong X-X, et al. Yes-associated protein/TEA domain family member and hepatocyte nuclear factor 4-alpha (HNF4α) repress reciprocally to regulate hepatocarcinogenesis in rats and mice. Hepatology 2017;65:1206-1221.
22) Tang F, Gao R, Jeevan-Raj B, Wyss CB, Kalathur RKR, Piscuoglio S, et al. LATS1 but not LATS2 represses autophagy by a kinase-independent scaffold function. Nat Commun 2019;10:5755.
23) **Coto-Llerena M**, **Ercan C**, Kancherla V, Taha-Mehlitz S, Eppenberger-Castori S, Soysal SD, et al. High expression of FAP in colorectal cancer is associated with angiogenesis and immunoregulation processes. Front Oncol 2020;10:979.
24) Lauzier A, Normandeau-Guimond J, Vaillancourt-Lavigueur V, Boivin V, Charbonneau M, Rivard N, et al. Colorectal cancer cells respond differentially to autophagy inhibition in vivo. Sci Rep 2019;9:11316.
25) ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.
26) Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178-192.

myoblast differentiation in vitro and muscle regeneration in vivo. PLoS Genet 2017;13:e1006600.

673

169

27) Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 2018;46:D260-D266. Erratum in: Nucleic Acids Res 2018;46:D1284.

28) Tonekaboni SAM, Mazrooei P, Kofia V, Haibe-Kains B, Lupien M. CREAM: Clustering of genomic REgions analysis method. bioRxiv 2017; https://doi.org/10.1101/222562.

29) Dimitrova Y, Gruber AJ, Mittal N, Ghosh S, Dimitriades B, Mathow D, et al. TFAP2A is a component of the ZEB1/2 network that regulates TGFB1-induced epithelial to mesenchymal transition. Biol Direct 2017;12:8.

30) Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, Amann-Zalcenstein D, Lara-Astiaso D, Amit I. High-throughput chromatin immuno-precipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat Protoc 2013;8:539-554.

31) Stupack DG, Teitz T, Potter MD, Mikolon D, Houghton PJ, Kidd VJ, et al. Potentiation of neuroblastoma metastasis by loss of caspase-8. Nature 2006;439:95-99.

32) Azoitei N, Hoffmann CM, Ellegast JM, Ball CR, Obermayer K, Gößele U, et al. Targeting of KRAS mutant tumors by HSP90 inhibitors involves degradation of STK33. J Exp Med 2012;209:697-711.

33) Sergushichev A. Algorithm for cumulative calculation of gene set enrichment statistic. Sci Tech J Inf Technol Mech Opt 2016;16:956-959.

34) Klimczak M, Biecek P, Zylicz A, Zylicz M. Heat shock proteins create a signature to predict the clinical outcome in breast cancer. Sci Rep 2019;9:7507.

35) Rohde M, Daugaard M, Jensen MH, Helin K, Nylandsted J, Jäättelä M. Members of the heat-shock protein 70 family promote cancer cell growth by distinct mechanisms. Genes Dev 2005;19:570-582.

36) Fuller KJ, Issels RD, Slosman DO, Guillet JG, Soussi T, Polla BS. Cancer and the heat shock response. Eur J Cancer 1994;30A:1884-1891.

37) **Xiong J, Jiang X-M**, Mao S-S, Yu X-N, Huang X-X. Heat shock protein 70 downregulation inhibits proliferation, migration and tumorigenicity in hepatocellular carcinoma cells. Oncol Lett 2017;14:2703-2708.

38) **Cho W, Jin X, Pang J, Wang Y**, Mivechi NF, Moskophidis D. The molecular chaperone heat shock protein 70 controls liver cancer initiation and progression by regulating adaptive DNA damage and mitogen-activated protein kinase/extracellular signal-regulated kinase signaling pathways. Mol Cell Biol 2019;39:e00391-18.

39) Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell Res 2012;22:490-503.

40) Yokota S, Kitahara M, Nagata K. Benzylidene lactam compound, KNK437, a novel inhibitor of acquisition of thermotolerance and heat shock protein induction in human colon carcinoma cells. Cancer Res 2000;60:2942-2948.

41) Yang S, Ren X, Liang Y, Yan Y, Zhou Y, Hu J, et al. KNK437 restricts the growth and metastasis of colorectal cancer via targeting DNAJA1/CDC45 axis. Oncogene 2020;39:249-261.

42) Anbanandam A, Albarado DC, Nguyen CT, Halder G, Gao X, Veeraraghavan S. Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. Proc Natl Acad Sci U S A 2006;103:17225-17230.

43) Vassilev A, Kaneko KJ, Shu H, Zhao Y, DePamphilis ML. TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. Genes Dev 2001;15:1229-1241.

44) Jiao S, Li C, Hao Q, Miao H, Zhang L, Li L, et al. VGLL4 targets a TCF4-TEAD4 complex to coregulate Wnt and Hippo signalling in colorectal cancer. Nat Commun 2017;8:14058.

45) Belandia B, Parker MG. Functional interaction between the p160 coactivator proteins and the transcriptional enhancer factor family of transcription factors. J Biol Chem 2000;275:30801-30805.

46) Meng Z, Moroishi T, Guan K-L. Mechanisms of Hippo pathway regulation. Genes Dev 2016;30:1-17.

47) **Wang C, Nie Z**, Zhou Z, Zhang H, Liu R, Wu J, et al. The interplay between TEAD4 and KLF5 promotes breast cancer partially through inhibiting the transcription of p27Kip1. Oncotarget 2015;6:17658-17697.

48) **Zhou Y, Huang T**, Zhang J, Wong CC, Zhang B, Dong Y, et al. TEAD1/4 exerts oncogenic role and is negatively regulated by miR-4269 in gastric tumorigenesis. Oncogene 2017;36:6518-6530.

49) **Tang J-Y, Yu C-Y**, Bao Y-J, Chen L, Chen J, Yang S-L, et al. TEAD4 promotes colorectal tumorigenesis via transcriptionally targeting YAP1. Cell Cycle 2018;17:102-109.

50) Martin J, Hartl FU. Molecular chaperones in cellular protein folding. BioEssays 1994;16:689-692.

51) Kiang JG, Tsokos GC. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology. Pharmacol Ther 1998;80:183-201.

52) Liu C-C, Jan Y-J, Ko B-S, Wu Y-M, Liang S-M, Chen S-C, et al. 14-3-3σ induces heat shock protein 70 expression in hepatocellular carcinoma. BMC Cancer 2014;14:425.

53) Tommaso LD, Di Tommaso L, Franchi G, Park YN, Fiamengo B, Destro A, et al. Diagnostic value of HSP70, glypican 3, and glutamine synthetase in hepatocellular nodules in cirrhosis. Hepatology 2007;45:725-734.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1656/suppinfo.

Infiltration by IL22-Producing T Cells Promotes Neutrophil Recruitment and Predicts Favorable Clinical Outcome in Human Colorectal Cancer

Nadia Tosti[1], Eleonora Cremonesi[2], Valeria Governa[2], Camilla Basso[3], Venkatesh Kancherla[1], Mairene Coto-Llerena[1], Francesca Amicarella[2], Benjamin Weixler[4], Silvio Däster[5], Giuseppe Sconocchia[6], Pietro Edoardo Majno[3], Dimitri Christoforidis[3], Luigi Tornillo[1], Luigi Terracciano[1], Charlotte K.Y. Ng[1,2], Salvatore Piscuoglio[1,7], Markus von Flüe[7,8], Giulio Spagnoli[2,6], Serenella Eppenberger-Castori[1], Giandomenica Iezzi[2,3], and Raoul Andre Droeser[8]

ABSTRACT

Immune cell infiltration in colorectal cancer effectively predicts clinical outcome. IL22, produced by immune cells, plays an important role in inflammatory bowel disease, but its relevance in colorectal cancer remains unclear. Here, we addressed the prognostic significance of IL22[+] cell infiltration in colorectal cancer and its effects on the composition of tumor microenvironment. Tissue microarrays (TMA) were stained with an IL22-specific mAb, and positive immune cells were counted by expert pathologists. Results were correlated with clinicopathologic data and overall survival (OS). Phenotypes of IL22-producing cells were assessed by flow cytometry on cell suspensions from digested specimens. Chemokine production was evaluated *in vitro* upon colorectal cancer cell exposure to IL22, and culture supernatants were used to assess neutrophil migration *in vitro*. Evaluation of a testing ($n = 425$) and a validation TMA ($n = 89$) revealed that high numbers of IL22 tumor-infiltrating immune cells were associated with improved OS in colorectal cancer. *Ex vivo* analysis indicated that IL22 was produced by CD4[+] and CD8[+] polyfunctional T cells, which also produced IL17 and IFNγ. Exposure of colorectal cancer cells to IL22 promoted the release of the neutrophil-recruiting chemokines CXCL1, CXCL2, and CXCL3 and enhanced neutrophil migration *in vitro*. Combined survival analysis revealed that the favorable prognostic significance of IL22 in colorectal cancer relied on the presence of neutrophils and was enhanced by T-cell infiltration. Altogether, colorectal cancer–infiltrating IL22-producing T cells promoted a favorable clinical outcome by recruiting beneficial neutrophils capable of enhancing T-cell responses.

Introduction

Colorectal cancer is the third most common cause of cancer-related death worldwide (1). Infiltration of immune cells into colorectal cancer

[1]Institute of Pathology, University Hospital of Basel, University of Basel, Basel, Switzerland. [2]Department of Biomedicine, University of Basel, Basel, Switzerland. [3]Department of Surgery, Ente Ospedaliero Cantonale and Università Svizzera Italiana, Lugano, Switzerland. [4]Klinik für Allgemein-, Viszeral- und Gefäßchirurgie, Charité Universitätsmedizin Berlin, Campus Benjamin Franklin, Berlin, Germany. [5]Department of Colorectal Surgery, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia. [6]Institute of Translational Pharmacology, National Research Council, Rome, Italy. [7]Department of Biomedicine, Visceral Surgery Research Laboratory, Clarunis, Basel, Switzerland. [8]University Center for Gastrointestinal and Liver Diseases, Clarunis, University of Basel, Basel, Switzerland.

tumors predicts clinical outcome more effectively than tumor–node–metastasis staging (2). Infiltration by cytotoxic T cells, Th type 1 cells, T regulatory cells, and neutrophils associates with favorable outcome in human colorectal cancer (3–6), whereas the role of IL17-producing T cells (Th17) is still debated (2, 7).

IL22 is a cytokine of the IL10 family produced by different cell types of the innate immune system, including group 3 innate lymphoid cells (ILC3), and of the adaptive immune system, including Th17, naïve CD4[+] T cells polarized upon exposure to TNFα and IL6, also known as Th22 cells (8), and CD8[+] T cells (9, 10). IL22 receptors, including IL22R αchain and IL10Rβ chain, are uniquely expressed by keratinocytes and a variety of epithelial cells, including intestinal cells (11). IL22 plays key roles in wound healing and tissue repair, and in the maintenance of the "barrier" functions of skin and of intestinal and bronchial epithelial layers (12–14). In these anatomic regions, IL22 synergizes with IL17 and TNFα to promote the expression of proteins involved in host defense (15–17) and innate immunity against bacterial infections. IL22 also induces epithelial cell proliferation and upregulation of genes encoding prosurvival molecules (18–21), and may protect the liver, intestine, and lungs from tissue destruction (19–24). IL22 also plays a role in the maintenance of host–microbiota symbiosis (25).

The fact that IL22 promotes inflammation and concomitantly prevents tissue destruction is intriguing. Notably, exogenous IL22 delivery is sufficient to promote inflammation in mice (26, 27). The functional relevance of IL22 depends on specific tissue microenvironment. For example, in the absence of IL17, IL22 promotes tissue repair in the lung (18). In the colon, IL22 protects against experimental chronic colitis and promotes intestinal wound healing following acute

171

intestinal injury (28). However, IL22 has also been suggested to play a role in the pathogenesis of a variety of human diseases, including psoriasis, arthritis, and inflammatory bowel disease (IBD; refs. 13, 29–31).

There is a paucity of data about the prognostic significance of IL22 in cancer (12, 32–34), in particular its role in human colorectal cancer remains unclear. In murine models, "uncontrolled" IL22 production promotes colorectal cancer development (33), possibly by direct effects on stem cells (35) or by enhancing cancer cell proliferation (36). Accordingly, in a murine model of colon cancer induced by administration of pathogenic *Helicobacter hepaticus* and carcinogenic azoxymethane in immunodeficient animals, IL22 produced by innate lymphoid cells mediates a protumorigenic effect via Stat3 activation in epithelial cells (37, 38). IL22 produced by human tumor-infiltrating lymphocytes also promotes colorectal cancer cell proliferation *in vivo* in a xenograft model (37). On the other hand, IL22 plays a key role in the control of genotoxic damage induced by carcinogens in colon epithelial stem cells, thereby limiting mutagenesis and cancer outgrowth (39).

Most of these studies have addressed direct effects of IL22 on tumor cells, whereas its potential ability to condition the colorectal cancer tumor microenvironment has not been explored in comparable detail (40). Prognostic significance of IL22 expression in human colorectal cancer, the overwhelming majority (>90%) of which is not associated to IBD or to clinically relevant chronic colitis, has not been investigated in large cohorts of patients. To fill this knowledge gap, in this study we used two tissue microarrays (TMA), collectively including >500 clinically annotated colorectal cancer specimens, to investigate the impact of IL22 on clinical outcome and its possible influence on tumor microenvironment composition.

Materials and Methods

TMA construction

The TMAs were constructed at the Pathology Biobank at the University Hospital of Basel (Basel, Switzerland). Unselected, nonconsecutive, formalin-fixed, paraffin-embedded primary colorectal cancer tissue blocks were used as donor blocks. Tissue cylinders with a diameter of 1 mm were punched from morphologically representative areas of each donor block and brought into one recipient paraffin block (30 × 25 mm), using the TMA GrandMaster (TMA-GM; 3D-Histech Ltd, Sysmex AG) technology. Each punch was derived from the center of the tumor in an area with no necrosis so that each TMA spot consisted of more than 50% tumor cells. Approval by the local ethics committee (Ethik Kommission beider Basel, EKBB) for the use of this clinically annotated TMA was obtained in advance, as stated in previous publications (6, 7, 41).

Clinicopathologic features

Clinicopathologic data for patients included in the TMAs were collected retrospectively in a nonstratified and nonmatched manner. The larger TMA set was a subset of a previously published TMA (42) including patients undergoing surgery from 1987 to 1996. The validation TMA was built with surgical specimens of patients who underwent surgery at the University Hospital of Basel (Basel, Switzerland) in the years from 2007 to 2012. Clinicopathologic characteristics are listed in Supplementary Tables S1 and S2. Overall survival (OS) was defined as primary endpoint. Available follow-up data for the testing and validation cohort had a mean event-free follow-up time of 115 and 36 months, respectively.

IHC

IHC IL22 staining was assessed on two TMAs consisting of 538 and 100 colorectal cancer specimens, respectively. After excluding samples for which the tissue punch was absent or had poor staining quality, 425 and 89 colorectal cancers samples, respectively, were available. Staining was performed using the BenchMark ULTRA IHC System (Ventana Medical Systems, Inc.), following the manufacturer's instructions, with an IL22-specific mAb (Creative Diagnostic #DCABH-2900, clone JNH9G22F3, dilution 1:100) and iVIEW-DAB as chromogen. Immunoreactivity was scored as number of tumor-infiltrating IL22$^+$ immune cells by experienced pathologists (L. Tornillo and L. Terracciano). IL17 and CD66b staining protocols have been reported in previous studies (6, 7).

Clinical specimen collection and processing

Freshly excised clinical specimens were collected from patients undergoing surgical treatment at University Hospital of Basel (Basel, Switzerland) and St. Claraspital (Basel, Switzerland). Informed written consent was obtained from all patients whose specimens were analyzed for this study.

Tumor or healthy tissue fragments were snap frozen for RNA extraction or enzymatically digested (2 mg/mL collagenase IV, Worthington Biochemical Corporation and 0.2 mg/mL DNAse I, Sigma-Aldrich, for 1 hour at 37°C) to obtain single-cell suspensions. In addition, peripheral blood mononuclear cells from healthy donors were isolated by density gradient centrifugation (Histopaque-1077, Sigma-Aldrich; 400 × g for 30 minutes at room temperature, without break). Use of human samples in this study was approved by local ethical authorities (Ethikkommission Nordwest und Zentralschweiz, EKNZ 2014–388).

The Cancer Genome Atlas analysis

Gene expression data [as fragments per kilobase of transcript per million mapped reads values (FPKM)] from 597 colorectal cancer samples and 51 normal colorectal mucosa specimens, were obtained from The Cancer Genome Atlas (TCGA) Genomics Data Commons harmonized data portal using *TCGA biolinks* R package (43). Clinical information regarding 597 colorectal cancers (see Supplementary Table S3) was retrieved from the Human Protein Atlas (44). After normalization, expression (FPKM values) of genes encoding IL22 and a panel of cytokines, chemokines, and immune markers was retrieved and correlations between immune markers and patients' OS were evaluated.

Flow cytometry and cell sorting

Cell suspensions obtained from colorectal cancer and tumor-free colonic mucosa were incubated with 50 ng/mL phorbol 12-myristate 13-acetate, 1 µg/mL ionomycin, and 5 µg/mL Brefeldin A (Sigma-Aldrich) for 5 hours. Cells were fixed with the Intracellular Fixation and Permeabilization Buffer Set (eBioscience), following the manufacturer's instructions, and surface stained with fluorochrome-conjugated antibodies specific for human CD3 (clone SK7), CD4 (clone SK3), CD8 (clone SK1), and CD56 (clone B159), all from BD Biosciences. Following permeabilization, intracellular staining with antibodies specific for human IFNγ (BD Bioscience, clone 25723.11), IL22 (R&D Systems, clone 142928), and IL17 (eBioscience, clone 64DEC17) was performed. Cells were analyzed by FACSCalibur or Fortessa Flow Cytometers (BD Biosciences). Primary colorectal cancer cells were sorted from tumor cell suspensions by magnetic microbeads conjugated to EpCAM-specific Antibodies (MACS MicroBeads, Miltenyi Biotec, catalog no. 130–061–101), following the manufacturer's

172

instructions. Cell purity was >97%, as evaluated by flow cytometry. Data were analyzed using FlowJo Software (Tree Star).

Real-time reverse transcription PCR assays

Total RNA was extracted from stored colorectal cancer tissues or sorted cell populations using NucleoSpin RNA Kit (Macherey-Nagel, catalog no. 740955.50), following the manufacturer's instructions, and quantified by Spectrophotometry (NanoDrop, Thermo Fisher Scientific). RNA was reverse transcribed using the Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT, Invitrogen). Quantitative PCR was performed in the ABI prism 7700 sequence detection system, using SYBR Green (Roche) or TaqMan Universal Master Mix, No AmpErase UNG (Applied Biosystems), and commercially available primer sequences. All primers are listed in Supplementary Table S4. Each gene was assayed in duplicate wells, using 20 ng template each. Expression of individual genes was analyzed by using the $2^{-\Delta\Delta C_t}$ method (45), as relative to the expression of GAPDH housekeeping gene.

Cell lines

LS180, HT29, Colo205, HCT15, SW480, SW620, and DLD-1 human colorectal cancer cell lines were purchased from the European Collection of Cell Cultures (period 2013–2015), and immediately stored in liquid nitrogen. Cells used for individual experiments were thawed from original cryopreserved aliquots and maintained in culture, for a maximum of 10 passages, in RPMI1640 (Gibco) or, for HT29 in McCoy's 5A Medium (Sigma-Aldrich) or, for SW480 and SW620 in L15 Medium (Leibovitz, Sigma-Aldrich), supplemented with 10% FBS, GlutaMAX-I, and Kanamycin (Gibco). Absence of *Mycoplasma* contamination in cultured cells was verified by PCR testing prior to investigations.

Chemokine induction in colorectal cancer cell lines

Colorectal cancer cells from LS180 and HT29 established cell lines were plated in 24-well plates (Sigma-Aldrich, 3.5×10^5 cells/well in 0.5 mL) in culture media and then incubated with 10 and 100 ng/mL concentrations of IL22 (R&D Systems) at 37°C. After 4 hours, chemokine expression was assessed by quantitative PCR.

ELISA

Chemokine content in culture supernatants was assessed by ELISA using CXCL1 and CXCL3 DuoSet ELISA (R&D Systems, catalog nos. DY275 and DY801, respectively), following the manufacturer's instructions. Results were collected by Spectrophotometer (BioTek Instruments) using the SoftMax Pro 6 software.

Cell proliferation assay

Colorectal cancer cells from LS180 and HT29 established cell lines were plated in 24-well plates (Sigma-Aldrich, 10^5 cells/well in 0.5 mL) in RPMI1640 and McCoy's 5A Medium (Sigma-Aldrich), respectively, and then incubated with 10 or 100 ng/mL recombinant human IL22 (R&D Systems, catalog no. 782-IL/CF) for 4 days at 37°C. Cell proliferation was quantified by CyQUANT Cell Proliferation Assay (Thermo Fisher Scientific, catalog no. C7026), following the manufacturer's instructions.

Migration assay

CD8[+] T cells and neutrophils were sorted from peripheral blood of healthy donors by Magnetic Microbeads (MACS MicroBeads, Miltenyi Biotec, catalog no. 130-045-201, and EasySep Human Neutrophil Isolation Kit, Stemcell Technologies, catalog no. 17957, respec-

tively), according to the manufacturer's instructions, to a purity of >98%, as confirmed by flow cytometry. Chemotaxis assays were performed using 96-well transwell plates with 5-μm pore size membranes (Corning Costar). Supernatants from LS180 or HT29 cells, left untreated or treated overnight with 10 or 100 ng/mL of IL22, were added to the bottom chambers (250 μL/well). In specific experiments, colorectal cancer cell line supernatants were depleted of CXCL1 and/or CXCL3, prior to use, by using specific capture antibodies (R&D Systems, clone 20326, catalog no. MAB275 and clone 49801, catalog no. MAB160, respectively). CD8[+] T cells and neutrophils (1.5×10^4 cells/well in 80 μL) were placed in the top chamber and allowed to migrate for 60 minutes at 37°C. The number of cells that migrated into the bottom chamber was quantified by flow cytometry. The extent of cell migration was expressed as a migration index, calculated as number of cells migrated toward supernatants/number of cells migrated toward control medium.

Statistical analysis

IL22[+] tumor-infiltrating immune cells were counted on each of the 425 and 89 colorectal cancer plus 16 nonmalignant cores, respectively. After having proven an association between the number of IL22[+] infiltrating cells and OS by univariate Cox regression, an optimal threshold was estimated by regression tree analysis (rpart Statistical Package Software R package, version 3.4.1, 2017-06-30). The obtained threshold was found to be almost equal to the median value. Subsequently, continuous values were dichotomized subdividing half of the collective as colorectal cancer with low or high IL22 immune infiltration. χ^2 or Fisher exact tests were used to determine the association between IL22 infiltration and clinicopathologic discrete features, as well as the Wilcoxon signed rank-sum test for comparison with continuous values. Survival curves were depicted according to the Kaplan–Meier method and compared with the log-rank test. Moreover, individual survival curves were compared one by one and the P values were adjusted according to the Benjamini–Hochberg method, which controls the FDR and the expected proportion of false discoveries among the rejected hypotheses (Survminer R package).

IL17[+] and CD66b[+] cells were evaluated as reported previously (6, 7). Upon staining with antibodies specific for IL17 (goat polyclonal anti-human IL17, R&D Systems) or CD66b (BioLegend, clone G10F5), numbers of positive cells per punch were scored by experienced pathologists. By regression tree analysis (rpart package), cut-off values for both markers, IL17 and CD66b, were set at 10 cells per punch. After dichotomization, Kaplan–Meier curves were plotted, and compared by log-rank test.

The assumption of proportional hazards was verified for all markers by analyzing the correlation of Schoenfeld residuals and the ranks of individual failure times. Any missing clinicopathologic information was assumed to be missing at random.

All P values were two-sided and considered significant at $P < 0.05$. Analyses were performed by using the Statistical Package Software R (version 3.4.1, 2017-06-30, http://www.r-project.org or higher) and GraphPad Prism 7 Software (GraphPad Software).

Results

High density of IL22[+] cells is associated with favorable prognosis in human colorectal cancer

IL22 protein expression was first evaluated by IHC on a testing TMA set including 425 primary colorectal cancer specimens (Supplementary Table S1). As expected, positive staining was clearly detected on infiltrating immune cells. However, a more diffuse staining was also

173
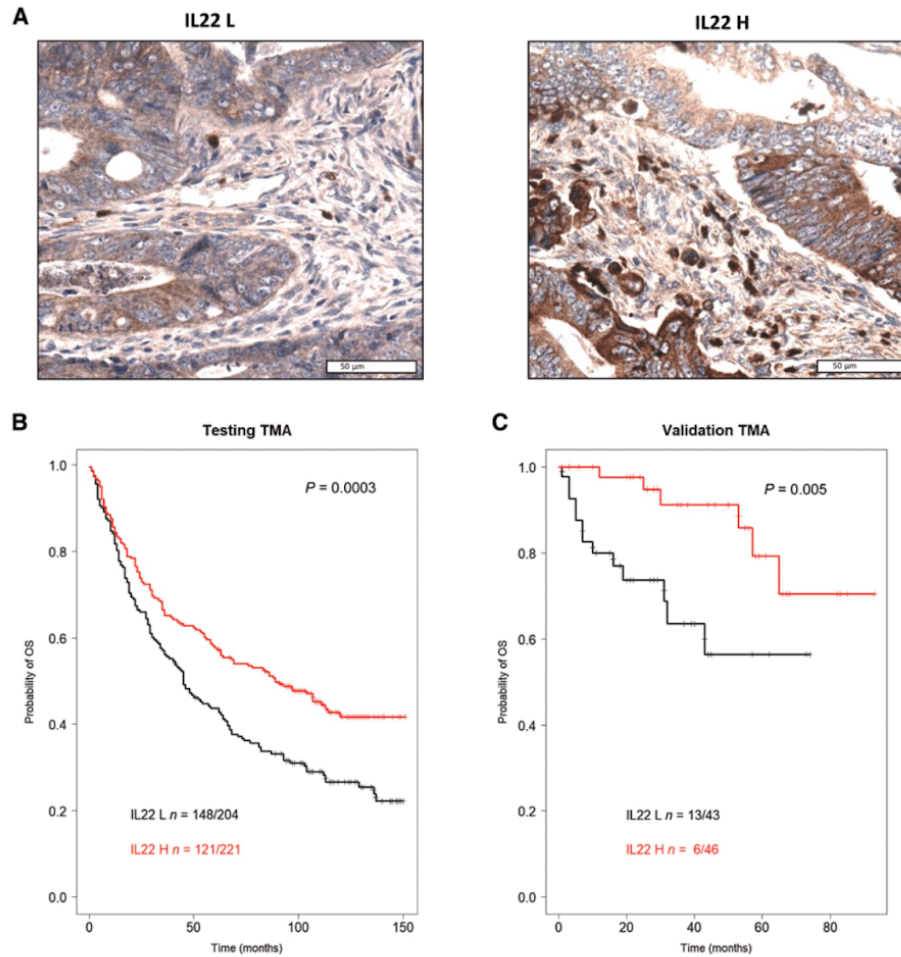
Figure 1.
High density of IL22$^+$ cells is associated with favorable prognosis in human colorectal cancer. **A,** Representative pictures of colorectal cancers with low (left) or high (right) infiltration by IL22$^+$ immune cells (scale bar, 50 μm). **B** and **C,** Kaplan–Meier curves depicting the probability of OS of patients dichotomized according to low or high number of IL22$^+$ colorectal cancer–infiltrating immune cells in the test TMA ($n = 425$) and in the validation TMA ($n = 89$). Statistical significance was assessed by log-rank test. H, high; L, low.

detectable on tumor cells (**Fig. 1A**). To clarify this issue, IL22 gene expression was investigated in colorectal cancer cells isolated from established cell lines with consistently negative results (Supplementary Fig. S1), suggesting that colorectal cancer cell positivity might be due to the IL22 fraction bound to its specific receptor on epithelial cells.

Therefore, we focused our analysis on tumor-infiltrating immune cells. IL22$^+$ cells were counted in each TMA core. IL22$^+$ cells were

detectable within both normal colonic tissues and colorectal cancer, although in the latter case at a significantly higher density ($P = 0.039$). Observed continuous values ranged between 0 and 300 cells per core with median and mean values of 20 and 39 cells per core, respectively. Median value was used to dichotomize tumor specimens into IL22-low and -high groups. Representative examples of colorectal cancer displaying low and high infiltration are displayed in **Fig. 1A**. As

174

Tosti et al.

Table 1. Correlation of IL22$^+$ cell density with clinicopathologic features in $n = 425$ colorectal cancer specimens.

Characteristics	IL22 low		IL22 high		P^a
	$N = 221$	(100%)	$N = 204$	(100%)	
Age					
Years (median, mean)	71, 69.5	(40–90)	71, 70.5	(40–96)	$P = 0.318$
Sex					
Female	117	(52.9)	104	(51.0)	$P = 0.687$
Male	104	(47.1)	100	(49.0)	
Diameter					
mm (median, mean)	50, 52.4	(10–170)	50, 51.0	(7–160)	$P = 0.527$
Tumor location					
Left sided	136	(61.5)	138	(67.6)	$P = 0.167$
Right sided	85	(38.5)	65	(31.9)	
Histologic subtype					
Mucinous	9	(4.1)	19	(9.3)	$P = 0.494$
Nonmucinous	212	(95.9)	185	(90.7)	
pT stage					
pT1	7	(3.2)	8	(3.9)	$P = 0.043^b$
pT2	26	(11.8)	32	(15.7)	
pT3	140	(63.3)	138	(67.6)	
pT4	39	(17.6)	21	(10.3)	
pN stage					
pN0	99	(44.8)	112	(54.9)	$P = 0.059^c$
pN1	64	(29.0)	51	(25.0)	
pN2	48	(21.7)	35	(17.2)	
Tumor grade					
G1	7	(3.2)	2	(1.0)	$P = 0.255$
G2	193	(87.3)	185	(90.7)	
G3	11	(5.0)	12	(5.9)	
Vascular invasion					
Absent	150	(67.9)	155	(76.0)	$P = 0.099$
Present	62	(28.1)	44	(21.6)	
Tumor border					
Pushing	54	(24.4)	64	(31.4)	$P = 0.151$
Infiltrating	156	(70.6)	135	(66.2)	
PTL inflammation					
Absent	167	(75.6)	145	(71.1)	$P = 0.162$
Present	45	(20.4)	54	(26.5)	
Microsatellite stability					
Deficient	30	(13.6)	33	(16.2)	$P = 0.452$
Proficient	191	(86.4)	171	(83.8)	
5-year survival rate					
(95% CI)	0.43	0.36–0.50	0.58	0.51–0.65	$P = 0.004$

Abbreviations: CI, confidence interval; PTL, peritumoral lymphocytic.
^{a}Correlation between the IL22-low and IL22-high subgroup. Percentage may not add to 100% due to missing values of some variables.
^{b}Overall difference in subgroup analysis due to T4.
^{c}Overall difference in subgroup analysis due to N0.

summarized in **Table 1**, the presence of high numbers of IL22$^+$ cells was slightly associated with lower T stage ($P = 0.043$). Instead, no significant association was found with histologic subtype ($P = 0.494$), lymph node metastases ($P = 0.059$), tumor grade ($P = 0.255$), vascular invasion ($P = 0.099$), tumor border configuration ($P = 0.151$), and microsatellite instability ($P = 0.452$).

Survival analysis at 5 years showed that patients with tumors characterized by high numbers of IL22$^+$ infiltrating cells had significantly higher survival probability than those with tumors displaying low IL22$^+$ cell numbers [OS, 58%; confidence interval (CI), 51%–65.0% vs. 43%; CI, 36%–50%; $P = 0.004$; **Table 1**].

Kaplan–Meier survival curve analysis revealed that the favorable prognostic effect of IL22$^+$ cell infiltration remained constant over time ($P = 0.0003$; **Fig. 1B**). The positive prognostic impact of high IL22$^+$ cell infiltration was confirmed upon staining of a TMA with an independent validation cohort (Supplementary Table S2) of 89 patients with colorectal cancer ($P = 0.005$; **Fig. 1C**). In this second cohort, for which data regarding adjuvant treatment were also available, multivariate analysis showed that IL22 retained an impact on OS irrespective of adjuvant chemotherapy ($P = 0.02$; Supplementary Table S5).

IL22 was expressed by polyfunctional T cells predictive of prolonged survival

We next investigated phenotypes and functions of colorectal cancer–infiltrating IL22$^+$ cells. Flow cytometric analysis of single-cell suspensions obtained from colorectal cancer clinical specimens revealed that IL22$^+$ cells were mostly found within conventional CD3$^+$

CANCER IMMUNOLOGY RESEARCH

Figure 2.
IL22 is expressed by conventional polyfunctional T cells. Single-cell suspensions obtained from freshly excised clinical specimens of colorectal cancer and tumor-free colonic tissues were surface stained with antibodies specific for CD3, CD4, CD8, and CD56 and intracellularly stained for IL17 and IFNγ, and analyzed by flow cytometry. **A,** Representative phenotypic analysis. **B,** Percentages of cells positive for the indicated marker or cytokine within the IL22$^+$ gate ($n \le 15$). **C,** Percentages of IL17$^+$ or IFNγ$^+$ cells within the IL22$^+$ cell gate in colorectal cancer (CRC TISSUE) and corresponding tumor-free colonic tissues (CTRL TISSUE). Means and SDs are indicated by bars. Statistical significance was assessed by the Wilcoxon signed rank test.
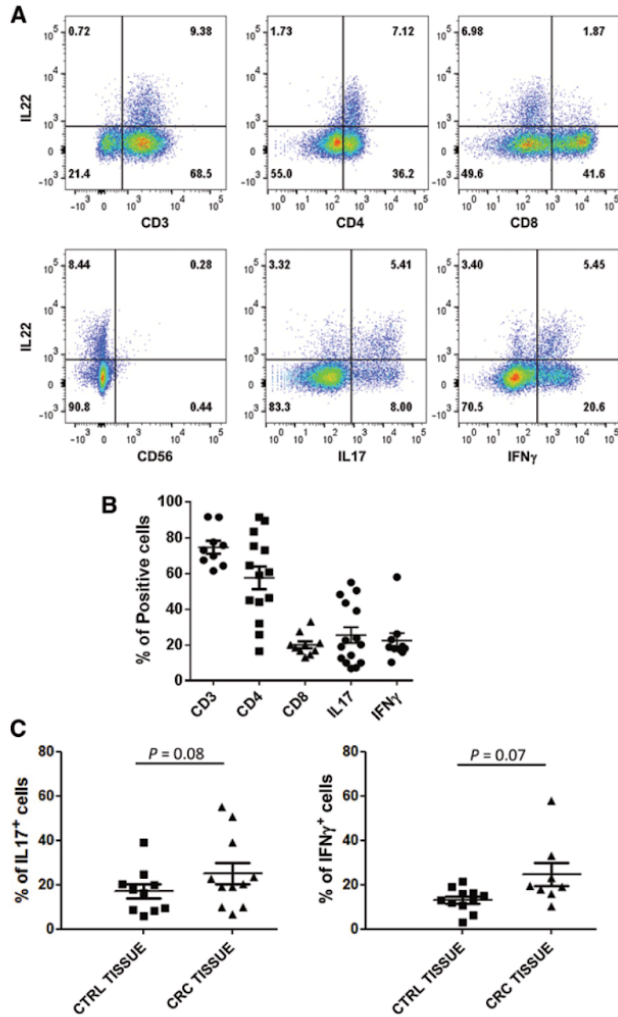
T cells and expressed CD4 and, to lower percentages, CD8 molecules (**Fig. 2A** and **B**). Expression of IL22 at the single-cell level, based on intracellular cytokine staining, was not significantly increased in cells infiltrating tumors as compared with those present in normal tissues (median fluorescence intensity within gated CD3$^+$ IL22$^+$ cells in tumors vs. nontumor tissues: 54 ± 32 vs. 53 ± 47; $P = 0.88$ and mean fluorescence intensity within CD3$^+$ IL22$^+$ cells in tumors vs. nontumor tissues: 125 ± 97 vs. 85 ± 63; $P = 0.15$). However, a large fraction of IL22$^+$ cells also showed production of additional cytokines, such as IL17 and IFNγ, thus indicating that they were polyfunctional T cells (**Fig. 2C**), as reported previously (7, 46). Notably, this subset of IL22-producing cells showed a trend toward an increase in colorectal cancer as compared with control tissues (**Fig. 2C**).

Consistent with their "Th17-like" phenotype, numbers of infiltrating IL22$^+$ cells were significantly higher within colorectal cancer characterized by high infiltration by IL17$^+$ cells (Supplementary Fig. S2A). In TCGA cohort, *IL22* gene expression significantly correlated with that of *IL17* gene ($r = 0.578$; $P < 0.0001$, see **Table 2**). In the TMA cohort, tumors displaying high densities of both IL22$^+$ and

176

Table 2. Correlations between expression of *IL22* gene and selected cytokine/chemokine genes in TCGA cohort ($n = 597$).

	Spearman correlation coefficient	*P*
IL17A	0.578	<0.00001
CXCL2	0.447	<0.00001
CXCL1	0.401	<0.00001
CXCL3	0.393	<0.00001
CXCL8	0.152	0.0002

IL17[+] cells were characterized by enhanced survival probability (Supplementary Fig. S2B).

IL22 stimulated colorectal cancer cells to release neutrophil-recruiting chemokines

TMA staining data indicated that colorectal cancer infiltration by IL22-producing T cells was associated with improved survival. This was an unexpected finding, because in murine colorectal cancer models, IL22 has mostly been shown to play a direct protumorigenic role, by enhancing tumor cell proliferation (35).

Indeed, we found that human colorectal cancer cells do express IL22 receptors, that is, IL22 Rα and IL10Rβ chains (Supplementary Fig. S3A and S3B), thus potentially responding to direct IL22-mediated effects. However, when we evaluated the direct effects of IL22 on LS180 and HT29 colorectal cancer cell lines, we did not observe any significant impact of IL22 on colorectal cancer cell proliferation *in vitro* (Supplementary Fig. S3C and S3D).

In contrast, IL22 treatment consistently increased gene expression of CXCL1, CXCL2, CXCL3 neutrophil-recruiting chemokines in two different colorectal cancer cell lines (**Fig. 3A**; Supplementary Fig. S4A). CXCL1 and CXCL3 protein release was also detectable in culture supernatants (**Fig. 3B**; Supplementary Fig. S4B), and enhanced neutrophil migration *in vitro* was accordingly observed (**Fig. 3C**).

Increased expression of genes encoding T-cell–recruiting chemokines, such as CCL22, CXCL9, and CXCL11, was also boosted by IL22 in LS180, but not in HT29 cells (**Fig. 3A**; Supplementary Fig. S4A).

IL22[+] T cells displayed cross-talk with neutrophils in colorectal cancer samples

On the basis of these findings, we hypothesized that a cross-talk between tumor and beneficial immune cell populations could underlie the favorable prognostic significance of IL22[+] cell infiltration in colorectal cancer.

Consistent with our *in vitro* data, we found a significant positive correlation between expression of *IL22* gene and that of *CXCL1* ($r = 0.401$; $P < 0.00001$), *CXCL2* ($r = 0.447$; $P < 0.00001$), and *CXCL3* ($r = 0.393$; $P < 0.00001$) genes in colorectal cancer specimens from TCGA database (**Table 2**). Upon *ex vivo* analysis on primary colorectal cancer cells sorted from clinical specimens, we found that expression of genes encoding neutrophil-recruiting chemokines was significantly enhanced in tumor cells derived from samples displaying high IL22 expression as compared with those purified from tumors displaying no or low IL22 expression (**Fig. 3D**). Accordingly, increased CD66b expression, consistent with higher neutrophil densities, was detected in IL22-high versus IL22-low tumors (**Fig. 3E**).

In our TMA cohort, we observed that high neutrophil densities, as defined by CD66b marker–specific staining, were significantly ($P < 0.00001$) associated with higher IL22[+] colorectal cancer–infiltrating cell numbers (**Fig. 4A**) and high *IL22* gene expression (Supplementary Fig. S5A).

The positive prognostic significance of high IL22[+] cell infiltration in colorectal cancer was lost in the absence of CD66b[+] cells, thus indicating that the beneficial effect of IL22-producing cells required neutrophil recruitment (**Fig. 4B**). A similar trend ($P = 0.09$) was also detectable in TCGA data (Supplementary Fig. S5B).

Neutrophils costimulate antigen-driven activation of colorectal cancer–infiltrating CD8[+] T cells (6). Remarkably, the presence of high CD8[+] or CD3[+] cell infiltration, in addition to IL22[+] and CD66b[+] cells, further enhanced their prognostic significance (**Fig. 4C**; Supplementary Fig. S6).

Discussion

Tumor infiltration by immune cell populations characterized by different cytokine production profiles heavily impacts clinical outcome of human colorectal cancer (3, 4, 47, 48); however, the role of immune cells producing IL22, a cytokine involved in tissue repair processes at mucosal surfaces, has remained elusive.

In experimental murine models, direct effects of IL22 on epithelial stem cells and tumor cells associate with tumor progression. IL22 also plays a role in IBD pathogenesis, potentially resulting in colorectal cancer outgrowth (37, 38). However, only ≤2% of sporadic human colorectal cancers are associated with clinically relevant IBD (49, 50). The prognostic significance of tumor infiltration by IL22[+] cells in the majority of colorectal cancer cases, diagnosed in the absence of a clinically significant chronic colitis/inflammation, has not been thoroughly investigated. High expression of DOTL1, an IL22-induced methyltransferase, associates with poor patient survival (35). However, no data regarding the prognostic impact of *IL22* gene or protein expression were provided.

This is the first study evaluating the impact of colorectal cancer–infiltrating IL 22[+] cells on patient survival. Upon analysis of two independent cohorts including 425 and 89 primary colorectal cancer cases, respectively, we unexpectedly observed that high infiltration by IL22[+] cells was associated with favorable prognosis. Evaluation at transcriptional level in TCGA cohort also provided data in-line with our findings of protein expression from TMA. Thus, these independent analyses concur in indicating IL22 expression as favorable predictive factor in human colorectal cancer.

Phenotypic analysis revealed that IL22-producing cells mostly comprised of polyfunctional Th17 and CD8[+] T cells, producing IL17 and IFNγ, in addition to IL22. These findings are in-line with previous studies reporting IL22 production by Th17 cells in human sporadic colorectal cancers (35, 46). The role of IL17 and Th17 cells in colorectal cancer is still debated. IL17 is suggested to be protumorigenic in mice and negatively influences colorectal cancer prognosis in humans (51). However, colorectal cancer–infiltrating Th17 cells might play a dual role depending on their tissue localization (7). Here, we showed that IL17 was positively associated with favorable prognosis only in the presence of high density of IL22[+] tumor-infiltrating cells, thus indicating a distinct role of polyfunctional Th17 cells as compared with cells producing IL17 only. On the other hand, in accordance with previous studies (35, 40), CD3[−] IL22[−] cells, possibly including ILCs, appeared to represent a minor fraction of human colorectal cancer–infiltrating IL22-producing cells. In a previously reported IBD-associated colorectal cancer mouse model (35), IL22-producing ILCs sustained tumor progression in immunodeficient animals. It is conceivable that in colorectal cancer subtypes driven by distinct pathogenesis, IL22 is produced by different cell populations, possibly endowed with distinct functional roles and prognostic significance.

The observed association between infiltration by IL22[+] cells and improved prognosis contradicts previous studies showing a
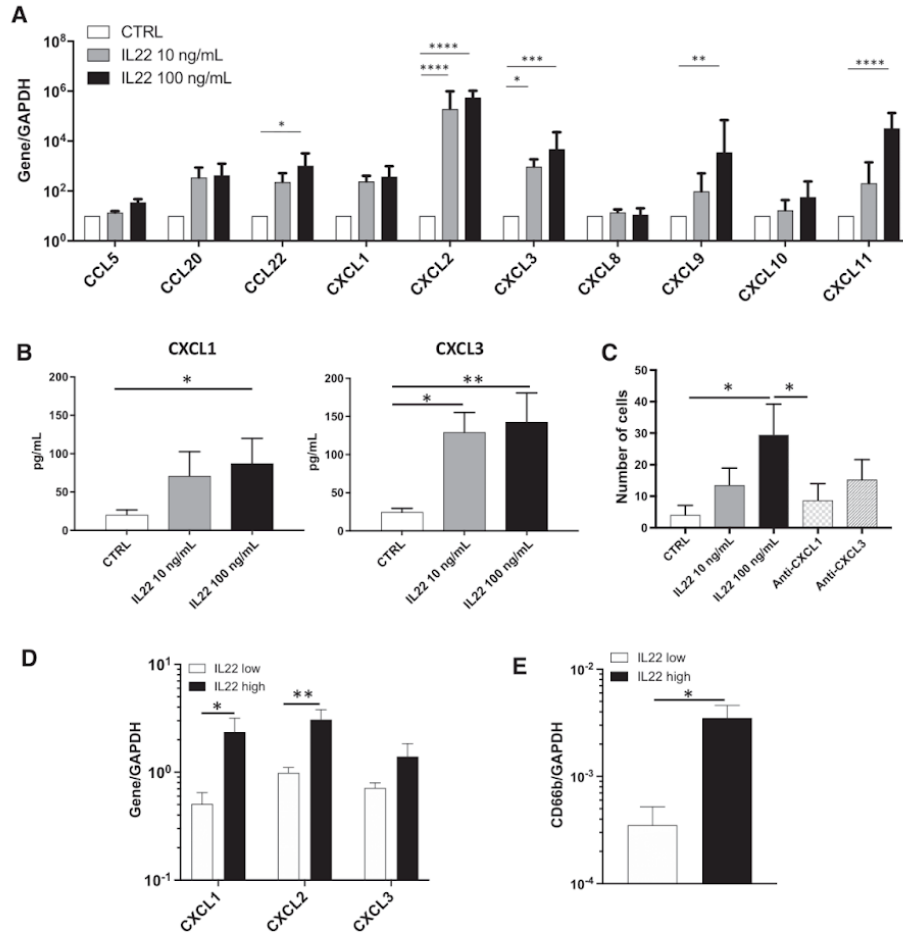
Figure 3.
IL22 enhances the release of neutrophil-recruiting chemokines by colorectal cancer cells. Colorectal cancer cells from the LS180 cell line were treated with IL22 (10 or 100 ng/mL), as indicated. **A,** After 4 hours, expression levels of genes encoding the indicated chemokines were analyzed by qRT-PCR, using *GAPDH* as reference gene. Means and SD from five independent experiments are shown. Statistical significance was assessed by two-way ANOVA test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$). **B,** After an overnight culture, CXCL1 and CXCL3 chemokine contents in culture supernatants were measured by ELISA. Means and SD from three independent experiments are shown. Statistical significance was assessed by Mann–Whitney test (*, $P < 0.05$; **, $P < 0.01$). **C,** Migration of neutrophils, isolated from healthy donors, toward culture supernatants from LS180 cells, untreated (CTRL) or exposed to the indicated doses of IL22 for an overnight period, was assessed in the presence or absence of CXCL3- and CXCL1-blocking reagents. Means and SD from three independent experiments are shown. Statistical significance was assessed by one-way ANOVA test (*, $P < 0.05$). **D,** Expression of CXCL1, CXCL2, and CXCL3 genes was evaluated in primary colorectal cancer cells sorted, based on EpCAM expression, from cell suspensions obtained upon enzymatic digestion of fresh colorectal samples. IL22 gene expression was assessed in corresponding whole colorectal cancer tissues, and, using the median of detected values as cutoff, tumors were classified as IL22 high or IL22 low. Expression of the indicated chemokines genes in colorectal cancer cells from IL22-high versus IL22-low tumors is depicted. **E,** Expression of the neutrophil marker CD66b in whole colorectal cancer tissues of IL22-high versus IL22-low tumors. All indicated P values were assessed by Mann–Whitney test (*, $P < 0.05$; **, $P < 0.01$).

178

Tosti et al.
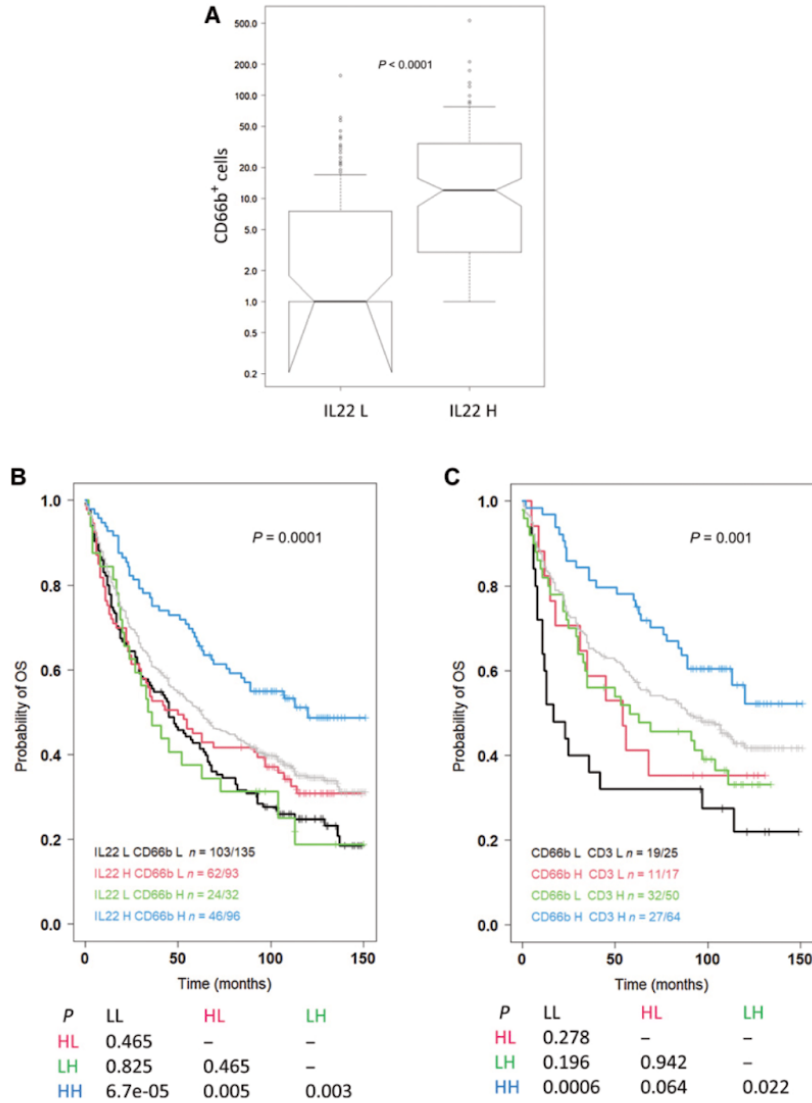
Figure 4.
The positive prognostic effect of IL22$^+$ cells depends on tumor-infiltrating neutrophils and on their interplay with T cells. **A,** Distribution of CD66b$^+$ cells according to the low or high IL22$^+$ colorectal cancer–infiltrating cells. **B,** Kaplan–Meier curves depicting the probability of OS in the entire collective (gray line) and in patients stratified according to combinations of IL22$^+$ and CD66b$^+$ cell density in the overall testing TMA ($n = 356$ instead of 425 because of CD66b missing values). **C,** Kaplan–Meier curves depicting the probability of OS in patients with tumors characterized by IL22$^+$-high density in the entire collective (gray line) and stratified according to combinations of CD66b$^+$ and CD3$^+$ cell infiltration ($n = 156$ instead of 221 because of missing CD66$^+$ and/or CD3$^+$ values). H, high; L, low.

CANCER IMMUNOLOGY RESEARCH

179

protumorigenic direct effect of IL22 on tumor cells (52, 53). Although we did not observe any significant effect when we evaluated the impact of IL22 on colorectal cancer cell proliferation *in vitro* according to published protocols (52, 53), it is possible that IL22 plays a dual role during different stages of the disease. While IL22 might contribute to tumor development during early phases of oncogenesis, in clinically detectable tumors with an established microenvironment, it might help to recruit beneficial immune cell populations.

Here, exposure of colorectal cancer cells to IL22 resulted in increased secretion of neutrophil-recruiting chemokines, including CXCL1, CXCL2, and CXCL3, and enhanced neutrophil migration *in vitro*. Consistently, in human colorectal cancer samples, expression of IL22 positively correlated with expression of neutrophil-recruiting chemokines in whole-tumor tissues and in purified tumor cells. This was associated with a higher expression of the CD66b neutrophil marker. The positive prognostic effect of IL22$^+$ cells was dependent on the presence of neutrophils.

Neutrophils are key players in the immune response against infectious challenges. However, their role in tumor immunobiology has long been neglected, possibly due to important differences in the granulocyte compartment between humans and experimental mice (54). A number of studies indicate that neutrophils may exert antitumor effects, potentially of high clinical relevance (55). In previous works, we and others (6, 55–57) showed that neutrophil infiltration associates with favorable prognosis in colorectal cancer. In addition, colorectal cancer–associated neutrophils costimulate antigen-triggered tumor-infiltrating T cells (6). Our data further underline the critical prognostic relevance of neutrophil infiltration in colorectal cancer microenvironment and unravel a previously unsuspected ability of IL22 to shape its composition. By stimulating production of neutrophil-recruiting chemokines in tumor cells, IL22-producing T cells might favor colorectal cancer infiltration by neutrophils, ultimately enhancing T-cell activation and expansion, and favoring a more positive clinical outcome.

In this scenario, the nature of stimuli favoring differentiation and recruitment of IL22-producing T cells and their antigenic specificities remains to be clarified. Also, it must be noted that the magnitude of IL22-induced effects ultimately depends on responsiveness of cells from individual tumors. Indeed, expression and functionality of IL22 receptors on tumor cells or their chemokine production capacity are determined by their genetic and epigenetic make-up (3, 47). This likely accounts for the variability of IL22-mediated effect across individual cases and the possible lack of neutrophilic infiltration in spite of high IL22 expression.

In conclusion, upon analysis of a large cohort of colorectal cancer cases, we found an unexpected positive prognostic effect of tumor-infiltrating IL22-producing T cells and showed the capacity of IL22 to recruit beneficial neutrophils into the tumor microenvironment by triggering the production of specific chemokines by colorectal cancer cells. This knowledge may be exploited for more precise tumor prognostication and may suggest the development of innovative treatments aimed at enhancing colorectal cancer infiltration by beneficial immune cell populations.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

N. Tosti: Resources, data curation. **E. Cremonesi:** Resources, data curation. **V. Governa:** Data curation. **C. Basso:** Data curation, formal analysis, data acquisition for the revision. **V. Kancherla:** Formal analysis. **M. Coto-Llerena:** Data curation, formal analysis. **F. Amicarella:** Supervision, writing–review and editing. **B. Weixler:** Writing–review and editing. **S. Däster:** Writing–review and editing. **G. Sconocchia:** Funding acquisition, writing–review and editing. **P.E. Majno:** Writing–review and editing. **D. Christoforidis:** Writing–review and editing. **L. Tornillo:** Supervision, validation, writing–review and editing. **L. Terracciano:** Supervision, writing–review and editing. **C.K.Y. Ng:** Formal analysis, methodology, writing–review and editing. **S. Piscuoglio:** Formal analysis, supervision, validation, writing–review and editing. **M. von Flüe:** Writing–review and editing. **G. Spagnoli:** Conceptualization, supervision, writing–review and editing. **S. Eppenberger-Castori:** Data curation, formal analysis, writing–review and editing. **G. Iezzi:** Conceptualization, supervision, funding acquisition, methodology, writing–original draft. **R.A. Droeser:** Conceptualization, supervision, funding acquisition, methodology, writing–original draft.

Acknowledgments

References

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin 2014;64: 9–29.
2. Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol 2017;14:717–34.
3. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313:1960–4.
4. Tosolini M, Kirilovsky A, Mlecnik B, Fredriksen T, Mauger S, Bindea G, et al. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, Th2, Treg, Th17) in patients with colorectal cancer. Cancer Res 2011;71: 1263–71.
5. Frey DM, Droeser RA, Viehl CT, Zlobec I, Lugli A, Zingg U, et al. High frequency of tumor-infiltrating FOXP3$^+$regulatory T cells predicts improved survival in mismatch repair-proficient colorectal cancer patients. Int J Cancer 2010;126;2635–43.
6. Governa V, Trella E, Mele V, Tornillo L, Amicarella F, Cremonesi E, et al. The interplay between neutrophils and CD8$^+$ T cells improves survival in human colorectal cancer. Clin Cancer Res 2017;23:3847–58.
7. Amicarella F, Muraro MG, Hirt C, Cremonesi E, Padovan E, Mele V, et al. Dual role of tumour-infiltrating T helper 17 cells in human colorectal cancer. Gut 2017;66:692–704.
8. Duhen T, Geiger R, Jarrossay D, Lanzavecchia A, Sallusto F. Production of interleukin 22 but not interleukin 17 by a subset of human skin-homing memory T cells. Nat Immunol 2009;10:857–63.
9. Qin WZ, Chen LL, Pan HF, Leng RX, Zhai ZM, Wang C, et al. Expressions of IL-22 in circulating CD4+/CD8+ T cells and their correlation with disease activity in SLE patients. Clin Exp Med 2011;11:245–50.
10. Nograles KE, Zaba LC, Shemer A, Fuentes-Duculan J, Cardinale I, Kikuchi T, et al. IL-22-producing "T22" T cells account for upregulated IL-22 in atopic dermatitis despite reduced IL-17-producing TH17 T cells. J Allergy Clin Immunol 2009;123:1244–52.
11. Wolk K, Kunz S, Witte E, Friedrich M, Asadullah K, Sabat R. IL-22 increases the innate immunity of tissues. Immunity 2004;21:241–54.
12. Hernandez P, Gronke K, Diefenbach A. A catch-22: interleukin-22 and cancer. Eur J Immunol 2018;48:15–31.

180

13. Ouyang W, O'Garra A. IL-10 family cytokines IL-10 and IL-22: from basic science to clinical translation. Immunity 2019;50:871–91.

14. Lindemans CA, Calafiore M, Mertelsmann AM, O'Connor MH, Dudakov JA, Jenq RR, et al. Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration. Nature 2015;528:560–4.

15. Liang SC, Tan XY, Luxenberg DP, Karim R, Dunussi-Joannopoulos K, Collins M, et al. Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. J Exp Med 2006;203:2271–9.

16. Aujla SJ, Chan YR, Zheng M, Fei M, Askew DJ, Pociask DA, et al. IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. Nat Med 2008;14:275–81.

17. Eyerich S, Eyerich K, Pennino D, Carbone T, Nasorri F, Pallotta S, et al. Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. J Clin Invest 2009;119:3573–85.

18. Sonnenberg GF, Nair MG, Kirn TJ, Zaph C, Fouser LA, Artis D. Pathological versus protective functions of IL-22 in airway inflammation are regulated by IL-17A. J Exp Med 2010;207:1293–305.

19. Zenewicz LA, Yancopoulos GD, Valenzuela DM, Murphy AJ, Karow M, Flavell RA. Interleukin-22 but not interleukin-17 provides protection to hepatocytes during acute liver inflammation. Immunity 2007;27:647–59.

20. Radaeva S, Sun R, Pan HN, Hong F, Gao B. Interleukin 22 (IL-22) plays a protective role in T cell-mediated murine hepatitis: IL-22 is a survival factor for hepatocytes via STAT3 activation. Hepatology 2004;39:1332–42.

21. Sugimoto K, Ogawa A, Mizoguchi E, Shimomura Y, Andoh A, Bhan AK, et al. IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. J Clin Invest 2008;118:534–44.

22. Zenewicz LA, Yancopoulos GD, Valenzuela DM, Murphy AJ, Stevens S, Flavell RA. Innate and adaptive interleukin-22 protects mice from inflammatory bowel disease. Immunity 2008;29:947–57.

23. Pickert G, Neufert C, Leppkes M, Zheng Y, Wittkopf N, Warntjen M, et al. STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. J Exp Med 2009;206:1465–72.

24. Simonian PL, Wehrmann F, Roark CL, Born WK, O'Brien RL, Fontenot AP. γδ T cells protect against lung fibrosis via IL-22. J Exp Med 2010;207:2239–53.

25. Goto Y, Obata T, Kunisawa J, Sato S, Ivanov II, Lamichhane A, et al. Innate lymphoid cells regulate intestinal epithelial cell glycosylation. Science 2014;345:1254009.

26. Liang SC, Nickerson-Nutter C, Pittman DD, Carrier Y, Goodwin DG, Shields KM, et al. IL-22 induces an acute-phase response. J Immunol 2010;185:5531–8.

27. Zheng Y, Danilenko DM, Valdez P, Kasman I, Eastham-Anderson J, Wu J, et al. Interleukin-22, a T(H)17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. Nature 2007;445:648–51.

28. Mizoguchi A. Healing of intestinal inflammation by IL-22. Inflamm Bowel Dis 2012;18:1777–84.

29. Sabat R, Ouyang W, Wolk K. Therapeutic opportunities of the IL-22-IL-22R1 system. Nat Rev Drug Discov 2014;13:21–38.

30. Mizoguchi A, Yano A, Himuro H, Ezaki Y, Sadanaga T, Mizoguchi E. Clinical importance of IL-22 cascade in IBD. J Gastroenterol 2018;53:465–74.

31. Eyerich K, Dimartino V, Cavani A. IL-17 and IL-22 in immunity: driving protection and pathology. Eur J Immunol 2017;47:607–14.

32. Nagakawa H, Shimozato O, Yu L, Takiguchi Y, Tatsumi K, Kuriyama T, et al. Expression of interleukin-22 in murine carcinoma cells did not influence tumour growth *in vivo* but did improve survival of the inoculated hosts. Scand J Immunol 2004;60:449–54.

33. Huber S, Gagliani N, Zenewicz LA, Huber FJ, Bosurgi L, Hu B, et al. IL-22BP is regulated by the inflammasome and modulates tumorigenesis in the intestine. Nature 2012;491:259–63.

34. Zhuang Y, Peng LS, Zhao YL, Shi Y, Mao XH, Guo G, et al. Increased intratumoral IL-22-producing CD4(+) T cells and Th22 cells correlate with gastric cancer progression and predict poor patient survival. Cancer Immunol Immunother 2012;61:1965–75.

35. Kryczek I, Lin Y, Nagarsheth N, Peng D, Zhao L, Zhao E, et al. IL-22+CD4+ T cells promote colorectal cancer stemness via STAT3 transcription factor activation and induction of the methyltransferase DOT1L. Immunity 2014;40:772–84.

36. Sun D, Lin Y, Hong J, Chen H, Nagarsheth N, Peng D, et al. Th22 cells control colon tumorigenesis through STAT3 and polycomb repression complex 2 signaling. Oncoimmunology 2015;5:e1082704.

37. Jiang R, Wang H, Deng L, Hou J, Shi R, Yao M, et al. IL-22 is related to development of human colon cancer by activation of STAT3. BMC Cancer 2013;13:59.

38. Kirchberger S, Royston DJ, Boulard O, Thornton E, Franchini F, Szabady RL, et al. Innate lymphoid cells sustain colon cancer through production of interleukin-22 in a mouse model. J Exp Med 2013;210:917–31.

39. Gronke K, Hernández PP, Zimmermann J, Klose CSN, Kofoed-Branzk M, Guendel F, et al. Interleukin-22 protects intestinal stem cells against genotoxic stress. Nature 2019;566:249–53.

40. Crome SQ, Nguyen LT, Lopez-Verges S, Yang SYC, Martin B, Yam JY, et al. A distinct innate lymphoid cell population regulates tumor-associated T cells. Nat Med 2017;23:368–75.

41. Weixler B, Cremonesi E, Sorge R, Muraro MG, Delko T, Nebiker CA, et al. OX40 expression enhances the prognostic significance of CD8 positive lymphocyte infiltration in colorectal cancer. Oncotarget 2015;6:37588–99.

42. Zlobec I, Karamitopoulou E, Terracciano L, Piscuoglio S, Iezzi G, Muraro MG, et al. TIA-1 cytotoxic granule-associated RNA binding protein improves the prognostic performance of CD8 in mismatch repair-proficient colorectal cancer. PLoS One 2010;5:e14282.

43. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGA biolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.

44. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science 2017;357:eaan2507.

45. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 2001;25:402–8.

46. Doulabi H, Rastin M, Shabahangh H, Maddah G, Abdollahi A, Nosratabadi R, et al. Analysis of Th22, Th17 and CD4+cells co-producing IL-17/IL-22 at different stages of human colon cancer. Biomed Pharmacother 2018;103:1101–16.

47. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 2013;39:782–95.

48. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer 2012;12:298–306.

49. Munkholm P. Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. Aliment Pharmacol Ther 2003;18:1–5.

50. Vagefi PA, Longo WE. Colorectal cancer in patients with inflammatory bowel disease. Clin Colorectal Cancer 2005;4:313–9.

51. De Simone V, Pallone F, Monteleone G, Stolfi C. Role of TH17 cytokines in the control of colorectal cancer. Oncoimmunology 2013;2:e26617.

52. Song B, Ma Y, Liu X, Li W, Zhang J, Liu J, et al. IL-22 promotes the proliferation of cancer cells in smoking colorectal cancer patients. Tumor Biol 2016;37:1349–56.

53. De Simone V, Franzè E, Ronchetti G, Colantoni A, Fantini MC, Di Fusco D, et al. Th17-type cytokines, IL-6 and TNF-α synergistically activate STAT3 and NF-kB to promote colorectal cancer cell growth. Oncogene 2015;34:3493–503.

54. Hagerling C, Werb Z. Neutrophils: critical components in experimental animal models of cancer. Semin Immunol 2016;28:197–204.

55. Ponzetta A, Mantovani A, Jaillon S. Dissecting neutrophil complexity in cancer. Emerg Top Life Sci 2017;1:457–70.

56. Galdiero MR, Bianchi P, Grizzi F, Di Caro G, Basso G, Ponzetta A, et al. Occurrence and significance of tumor-associated neutrophils in patients with colorectal cancer. Int J Cancer 2016;139:446–56.

57. Ponzetta A, Carriero R, Carnevale S, Barbagallo M, Molgora M, Perucchini C, et al. Neutrophils driving unconventional T cells mediate resistance against murine sarcomas and selected human tumors. Cell 2019;178:346–60.

181

ZNRF3 and RNF43 cooperate to safeguard metabolic liver zonation and hepatocyte proliferation

Graphical abstract
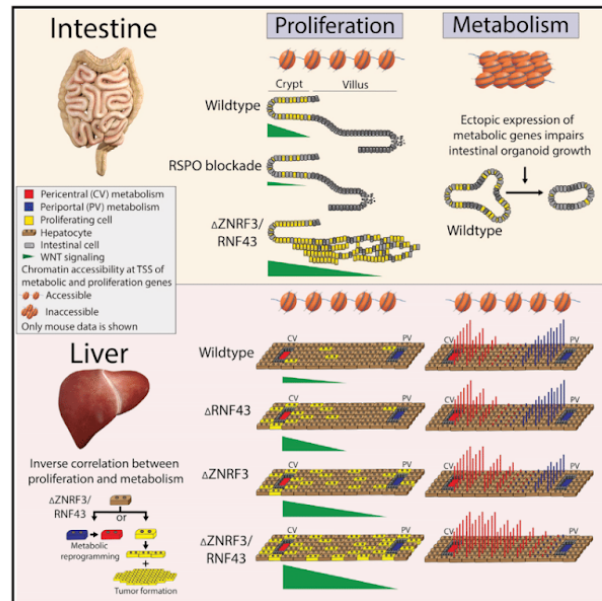
Authors

Tianliang Sun, Stefano Annunziato, Sebastian Bergling, ..., Prisca Liberali, Feng Cong, Jan S. Tchorz

Correspondence
tlsun1983@gmail.com (T.S.),
jan.tchorz@novartis.com (J.S.T.)

In brief

How ISCs and hepatocytes differentially use WNT signaling is unclear. Sun et al. now reveal that ZNRF3/RNF43 balance WNT signaling in AXIN2+ hepatocytes to prevent proliferation while preserving metabolic zonation, and ZNRF3/RNF43 deletion promotes liver tumors. Restricted chromatin accessibility in ISCs prevents the expression of β-Catenin-regulated metabolic enzymes.

Highlights

- Intestinal stem cells (ISCs) and hepatocytes differentially use WNT signaling

- WNT/β-Catenin activity correlates with proliferation of ISCs and hepatocytes

- ZNRF3/RNF43 balance WNT signaling and their deletion promotes liver tumor formation

- Inverse correlation between metabolism and proliferation in hepatocytes and ISCs

CellPress

Cell Stem Cell

CellPress

Article

ZNRF3 and RNF43 cooperate to safeguard metabolic liver zonation and hepatocyte proliferation

Tianliang Sun,[1,*] Stefano Annunziato,[1] Sebastian Bergling,[1] Caibin Sheng,[1] Vanessa Orsini,[1] Pascal Forcella,[1] Monika Pikiolek,[1] Venkatesh Kancherla,[2,3] Sjoerd Holwerda,[1] Dilek Imanci,[1] Fabian Wu,[1] Ludivine Challet Meylan,[4,5] Lea F. Puehringer,[1] Annick Waldt,[1] Mevion Oertli,[1] Sven Schuierer,[1] Luigi M. Terracciano,[6,7] Stefan Reinker,[1] Heinz Ruffner,[1] Tewis Bouwmeester,[1] Andreas W. Sailer,[1] Elizabeth George,[8] Guglielmo Roma,[1] Antoine de Weck,[1] Salvatore Piscuoglio,[2,3] Felix Lohmann,[1] Ulrike Naumann,[1] Prisca Liberali,[4,5] Feng Cong,[8] and Jan S. Tchorz[1,9,*]

[1]Novartis Institutes for BioMedical Research, Novartis Pharma AG, Basel, Switzerland
[2]Institute for Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland
[3]Visceral Surgery and Precision Medicine Research Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland
[4]Friedrich Miescher Institute for Biomedical Research (FMI), Basel, Switzerland
[5]University of Basel, Basel, Switzerland
[6]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (Milan), Italy
[7]IRCCS Humanitas Research Hospital, Rozzano (Milan), Italy
[8]Novartis Institutes for BioMedical Research, Novartis Pharma AG, Cambridge, MA, USA
[9]Lead contact
*Correspondence: tlsun1983@gmail.com (T.S.), jan.tchorz@novartis.com (J.S.T.)
https://doi.org/10.1016/j.stem.2021.05.013

SUMMARY

AXIN2 and LGR5 mark intestinal stem cells (ISCs) that require WNT/β-Catenin signaling for constant homeostatic proliferation. In contrast, AXIN2/LGR5+ pericentral hepatocytes show low proliferation rates despite a WNT/β-Catenin activity gradient required for metabolic liver zonation. The mechanisms restricting proliferation in AXIN2+ hepatocytes and metabolic gene expression in AXIN2+ ISCs remained elusive. We now show that restricted chromatin accessibility in ISCs prevents the expression of β-Catenin-regulated metabolic enzymes, whereas fine-tuning of WNT/β-Catenin activity by ZNRF3 and RNF43 restricts proliferation in chromatin-permissive AXIN2+ hepatocytes, while preserving metabolic function. ZNRF3 deletion promotes hepatocyte proliferation, which in turn becomes limited by RNF43 upregulation. Concomitant deletion of RNF43 in ZNRF3 mutant mice results in metabolic reprogramming of periportal hepatocytes and induces clonal expansion in a subset of hepatocytes, ultimately promoting liver tumors. Together, ZNRF3 and RNF43 cooperate to safeguard liver homeostasis by spatially and temporally restricting WNT/β-Catenin activity, balancing metabolic function and hepatocyte proliferation.

INTRODUCTION

Organs with high turnover of cells are maintained by defined epithelial tissue stem cell populations, which have increased proliferative and regenerative potential compared with their more differentiated descendants. In the small intestine, shedding of cells at the tip of the villus requires constant replenishment of cells, provided by intestinal stem cells (ISCs). ISCs are characterized by high WNT/β-Catenin activity indicated by expression of the β-Catenin target genes *Axin2* and *Lgr5*, similar to tissue stem cells in other organs. AXIN2 and LGR5 are therefore widely considered as tissue stem cell markers (Barker et al., 2010; Clevers et al., 2014). How ISCs prevent the expression of β-Catenin-controlled metabolic genes, while using WNT/β-Catenin signaling to proliferate, remains unclear. In the homeostatic liver, AXIN2 and LGR5 are expressed in pericentral hepatocytes (Ang

et al., 2019; Planas-Paz et al., 2016; Sun et al., 2020), and a centro-portal WNT/β-Catenin activity gradient is essential to control metabolic liver zonation (Benhamouche et al., 2006; Burke et al., 2009; Ma et al., 2020; Planas-Paz et al., 2016; Preziosi et al., 2018; Rocha et al., 2015; Torre et al., 2011; Yang et al., 2014). Zonated WNT/β-Catenin signaling compartmentalizes complementary functions in hepatocytes along the hepatic blood flow, enabling the liver to metabolize nutrients and xenobiotics, as well as to produce and recycle proteins (Russell and Monga, 2018). Although AXIN2+ hepatocytes expressing glutamine synthetase (GS) were initially introduced as liver stem cells, on the basis of AXIN2 lineage tracing studies (Wang et al., 2015), several studies using a variety of approaches contested this finding (Ang et al., 2019; Chen et al., 2020a; He et al., 2021; Planas-Paz et al., 2016; Sun et al., 2020; Wei et al., 2021). The mechanisms preventing proliferation of AXIN2+ hepatocytes in the

presence of constant WNT/β-Catenin activity that is necessary to maintain metabolic zonation remained elusive.

Although the overall low turnover of hepatocytes during homeostasis does not require overt proliferation, hepatocytes throughout the liver can re-enter the cell cycle and replace damaged hepatocytes on demand (Michalopoulos and DeFrances, 1997; Miyajima et al., 2014; Stanger, 2015). Increased WNT/β-Catenin signaling and upregulation of AXIN2 and LGR5 are important for this regenerative response (Planas-Paz et al., 2019; Sun et al., 2020; Zhao et al., 2019). Hyperactive WNT/β-Catenin signaling is a hallmark of many liver tumors and pericentral hepatocytes are highly susceptible to malignant transformation into hepatocellular carcinoma (HCC) in experimental models (Ang et al., 2019; Perugorria et al., 2019; Russell and Monga, 2018). Many HCCs express pericentral metabolic enzymes (Adebayo Michael et al., 2019), and deregulated WNT/β-Catenin signaling disrupts metabolic zonation (Colnot, 2011; Gebhardt and Matz-Soja, 2014). This highlights the importance of safeguarding mechanisms restricting uncontrolled WNT/β-Catenin signaling and hepatocyte proliferation while fine-tuning WNT/β-Catenin activity, which is necessary to ensure liver homeostasis.

Using a combined approach of chromatin accessibility profiling, spatial transcriptomics profiling, lineage tracing, characterization of injury models, and WNT/β-Catenin pathway modulation in mice, we now reveal mechanisms balancing metabolic function and proliferation in hepatocytes and ISCs. Importantly, we identify a tumor suppressor mechanism in hepatocytes, limiting hepatic WNT/β-Catenin activity and balancing hepatocyte proliferation and metabolic function.

RESULTS

Chromatin accessibility and WNT/β-Catenin activity levels control proliferation and metabolic gene expression in ISCs and AXIN2+ hepatocytes

Comparative gene set enrichment analysis (GSEA) suggests that AXIN2+ hepatocytes use WNT/β-Catenin signaling primarily for the expression of metabolic genes, whereas ISCs express mostly cell cycle genes (Sun et al., 2020). Consistently, ISCs show much higher proliferation rates compared with AXIN2+ hepatocytes (80% versus 2%) under homeostatic conditions (Sun et al., 2020). Although AXIN2+ hepatocytes express β-Catenin-regulated metabolic proteins GS and CYP2E1, the same genes are not expressed in ISCs despite their high WNT/β-Catenin activity (Figure 1A). We hypothesized that differences in chromatin accessibility might contribute to the lack of metabolic gene expression observed in ISCs and absence of cell cycle gene expression in AXIN2+ hepatocytes, respectively. AXIN2+ and AXIN2− hepatocytes, as well as AXIN2+ ISCs, were isolated from BAC-transgenic AXIN2-CreERT2^{EGFP} mice (Sun et al., 2020) following tamoxifen (TAM)-induced EGFP expression in AXIN2+ hepatocytes, and subjected to assay for transposase-accessible chromatin sequencing (ATAC-seq) (Figure 1B). Principal-component analysis indicates similarities between AXIN2+ and AXIN2− hepatocytes and marked differences in the overall chromatin accessibility compared with AXIN2+ ISCs (Figure S1A). Importantly, ATAC signals for housekeeping genes between hepatocytes and ISCs show only minor differences

(Figure S1B), which were accounted for by normalizing data in further analyses. Transcriptional start site (TSS)-proximal regulatory elements driving pericentral and periportal metabolic gene expression are more accessible in AXIN2+ and AXIN2− hepatocytes compared with AXIN2+ ISCs (Figure 1C). Moreover, treatment of intestinal organoids derived from wild-type (WT) mice with the histone deacetylase inhibitor Trichostatin A (TSA), which opens chromatin (Yoshida et al., 1995), resulted in the expression of WNT-regulated metabolic genes and impaired growth (Figures S1C–S1F). Lentiviral transduction of intestinal organoids with either GS-EGFP plus mCherry-control or CYP2E1-mCherry plus EGFP-control, showed that cells expressing only a fluorescent marker increased significantly more than cells also overexpressing a metabolic gene (Figures S1G–S1J). This suggests an inverse correlation between metabolic gene expression and proliferation.

Interestingly, chromatin accessibility for metabolic genes in hepatocytes seems to be independent of the status of WNT/β-Catenin signaling, as periportal hepatocytes (AXIN2−) and pericentral hepatocytes (AXIN2+) showed similar chromatin accessibility for both periportal and pericentral metabolic genes (Figure 1C), which was not altered upon abrogation of hepatic WNT/β-Catenin signaling with AlbCre-mediated deletion of LGR4/5 in the liver (Lgr4/5dLKO mice) (Figures S1K and S1L). Surprisingly, TSS-proximal regulatory elements of cell cycle genes were accessible in both ISCs and hepatocytes (Figure 1D), and their accessibility was not altered in hepatocytes of LGR4/5dLKO mice (Figure S1M). Such positive ATAC signal at the TSS of non-expressed genes may prime hepatocytes to quickly react to instructive upstream signals. However, this suggests that other, possibly non-chromatin-based mechanisms, prevent proliferation of AXIN2+ hepatocytes.

We next assessed how fine-tuning of WNT/β-Catenin activity levels affects hepatocyte function, probing the hypothesis that hepatocyte proliferation correlates with WNT/β-Catenin signaling dose. First, we analyzed WNT/β-Catenin pathway activation in AXIN2-LacZ reporter mice during early postnatal liver development or adult homeostasis or following different liver injuries (Figure 2A), all of which are conditions involving WNT/β-Catenin signaling (Planas-Paz et al., 2016; Sun et al., 2020). In naive AXIN2-LacZ mice, WNT/β-Catenin activity was restricted to pericentral and adjacent parenchymal hepatocytes, as indicated by βGal staining in Axin2-LacZ mice (Figure S2A) or Axin2 mRNA in situ hybridization (ISH) (Figures 2B and 2F). At postnatal day (P) 2 or following allyl alcohol (AA), 3,5-dicarbethoxy-1,4-dihydrocollidine (DDC) or partial hepatectomy (PHx) liver injury, Axin2 expression expanded into other liver zones and increased in pericentral hepatocytes (Figures 2B, 2F, and S2A). Hepatocyte proliferation, indicated by costaining for the proliferation marker Ki67 and the hepatocyte marker albumin (ALB), increased in all models that showed elevated WNT/β-Catenin signaling (Figures 2C and 2G). Likewise, high Axin2 ISH signals in the intestinal crypt correspond to high proliferation (Figures 2D–2G). Importantly, we show that the increase in WNT/β-Catenin activity highly correlates with the increase in proliferation, across the different injury models, postnatal liver development, and in the small intestine (Figure 2H).

Injections of R-Spondin1 (RSPO) protein, which potentiates WNT/β-Catenin signaling via LGR4–6 receptors (de Lau et al.,

Cell Stem Cell
Article

CellPress
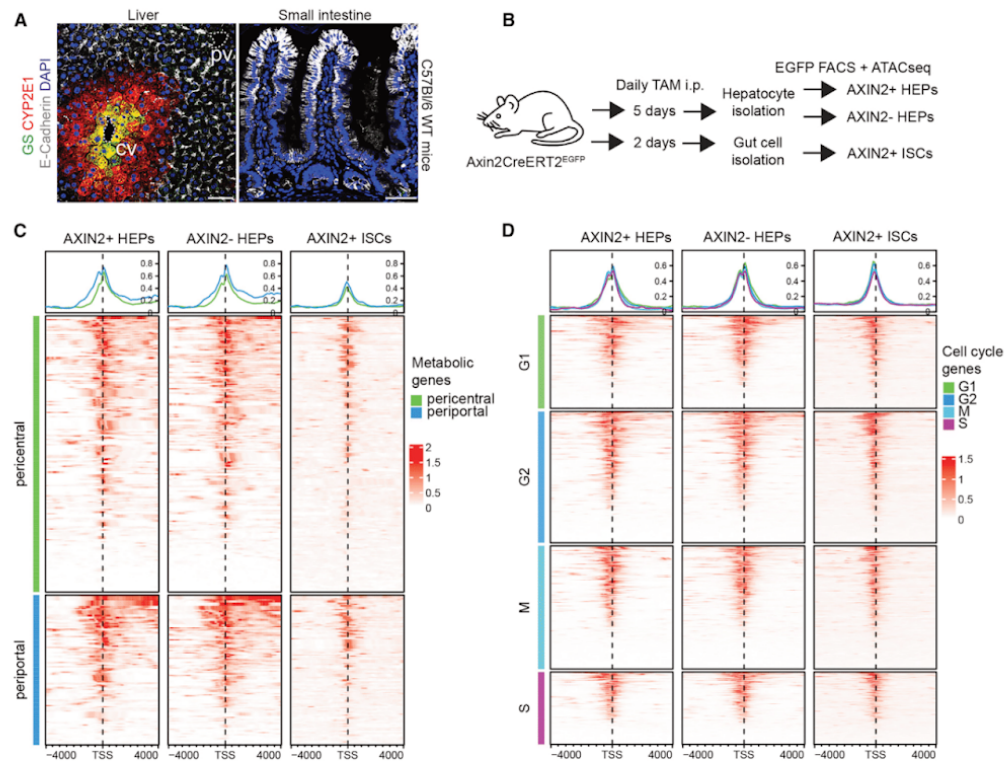
Figure 1. Differential use of WNT/β-Catenin signaling in AXIN2+ ISCs and hepatocytes

(A) GS, CYP2E1, and E-cadherin costaining in C57BL/6 mouse liver and small intestine. CV, central vein; PV, portal vein. Scale bars, 50 μm.

(B) Isolation of AXIN2+ and AXIN2− hepatocytes (HEPs), and AXIN2+ ISCs, for subsequent ATAC-seq.

(C and D) Chromatin accessibility (ATAC) signal centered at transcription start site (TSS; ±4 kb) of metabolic genes (C) or cell cycle genes (D) as observed in AXIN2− and AXIN2+ HEPs or AXIN2+ ISCs. ATAC signal (averaged replicates: n = 6 AXIN2+ HEPs, n = 5 AXIN2- HEPs and AXIN2+ ISCs) sorted by strength per group).

2011; Ruffner et al., 2012), promoted a dose-dependent increase in *Axin2* mRNA ISH signals (Figures 3A–3C) and proliferation (Figures 3D and 3E) in naive C57BL/6 hepatocytes, indicating a direct correlation between RSPO-induced WNT/β-Catenin activity and hepatocyte proliferation (Figure 3F). Likewise, RSPO caused a dose-dependent increase in proliferation in the intestinal crypt (Figures S2B and S2C). In contrast, scavenging of endogenous RSPO protein by doxycycline (DOX)-inducible expression of the extracellular domain of zinc and ring finger 3 tethered to the plasma membrane (ZNRF3ECD-TM) (Planas-Paz et al., 2016) dramatically reduced WNT/β-Catenin signaling and impaired proliferation, resulting in defective homeostatic renewal of the small intestine (Figures S2D–S2I). These data extend previous findings showing that LGR4/5 deletion impairs renewal in the crypt-villus unit (de Lau et al., 2011), suggesting that fine-tuning of WNT/β-Catenin signaling also controls proliferation outside of the liver. Significantly lower expression of *Rspo3* and *Axin2* in the hepatic pericentral zone compared

with the intestinal crypt and transit-amplifying cell (TA) compartment (Figures S2J–S2L) may explain why AXIN2+ hepatocytes do not readily proliferate during homeostasis.

Although LGR4/5 deletion during liver development abrogated metabolic zonation (Planas-Paz et al., 2016), it remained unclear whether RSPO-LGR4/5-mediated WNT/β-Catenin signaling is also critical to maintain WNT/β-Catenin signaling and metabolic zonation in adult mice. TAM-induced deletion of LGR4/5, specifically in adult AXIN2+ cells, disrupted GS and CYP2E1 expression, indicating that LGR4/5 are essential to maintain metabolic zonation in adult AXIN2+ hepatocytes (Figures S2M–S2O). In contrast, RSPO injections promoted a dose-dependent increase in GS and CYP2E1 expression (Figures 3D and S2P–S2R), whereas the opposite occurred following ZNRF3ECD-TM expression (Planas-Paz et al., 2016). To further study the role of RSPO-induced WNT/β-Catenin signaling in balancing metabolic zonation and proliferation in hepatocytes, we performed spatial transcriptomics analyses in livers following either PBS

185
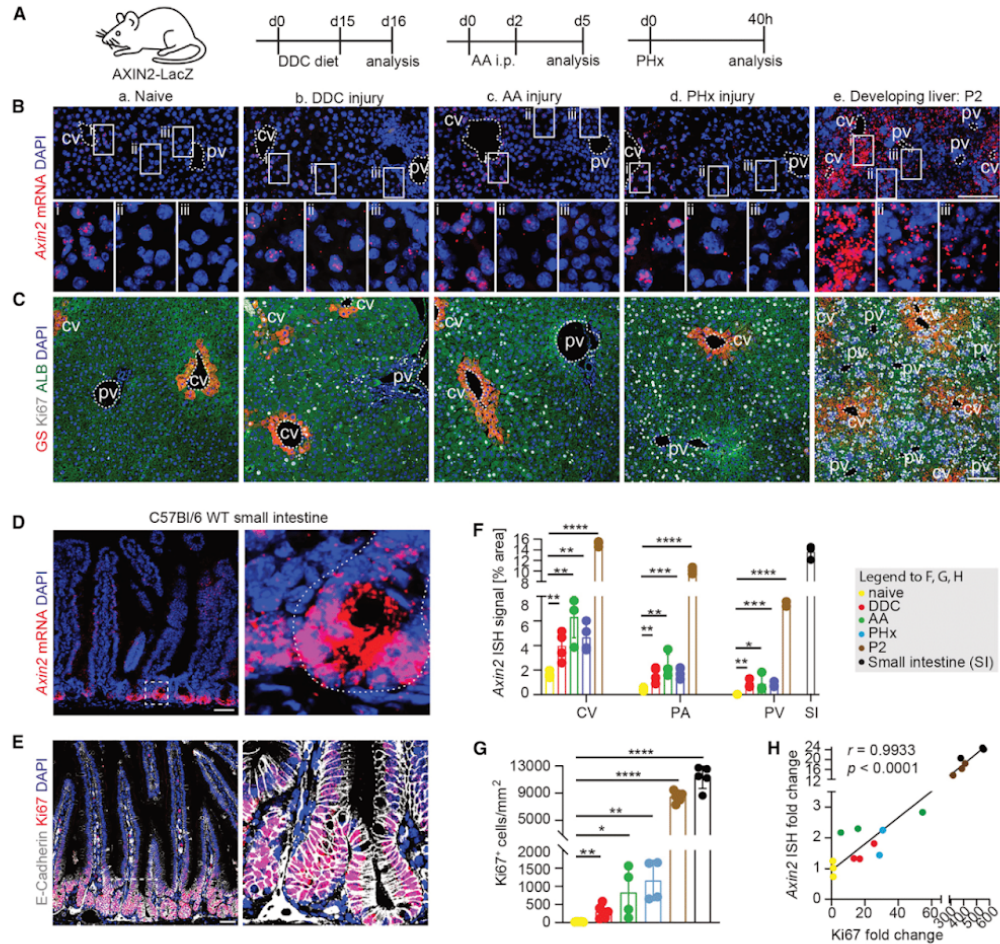
Figure 2. Fine-tuning of WNT/β-Catenin signaling levels controls proliferation in liver and intestine
(A) Liver injury models.
(B) *Axin2* ISH staining in livers from indicated mice. Magnified insets show central vein (cv; i), parenchymal (ii), and portal vein (pv; iii) hepatocytes.
(C) GS, Ki67, and ALB costaining in the indicated mice.
(D and E) *Axin2* ISH staining (D) and costaining for Ki67 and E-cadherin (E) in small intestine (SI) of WT mice.
(F and G) Quantification of *Axin2* mRNA ISH signals (F) and Ki67 staining (G) in hepatocytes of indicated mice and WT intestinal crypts.
(H) Correlation of WNT/β-Catenin activity (*Axin2* mRNA ISH increase) with the increase in proliferation (Ki67) in hepatocytes of indicated mice and WT intestinal crypts. Each dot represents one mouse. Color codes indicate experimental models. In SI, crypt and transit amplifying zone was quantified (F and G).
Data represent mean ± SD. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. Two-tailed unpaired Student's t tests (F and G) or Pearson correlation (H) was used. Scale bars, 50 μm (D and E) and 100 μm (B and C).

or RSPO injection. Data quality assessment for two replicate liver sections suggests similar amounts of unique molecular identifiers (UMIs) and expressed genes across samples (Figure S2S). Spatial gene expression analyses overlaid onto the hematoxylin and eosin (H&E)-stained livers indicated widespread upregulation of the WNT/β-Catenin targets *Lgr5* and *Lect2* upon RSPO in-

jection, whereas increased expression of *Mki67* confirms increased proliferation (Figure 3G). Averaged expression values for the top 12 periportal or pericentral metabolic genes (Sun et al., 2020) enabled reconstructing the metabolic zones in PBS-treated mice on a spatial grid (Figure 3G) or when projected onto a uniform manifold approximation and projection (UMAP)
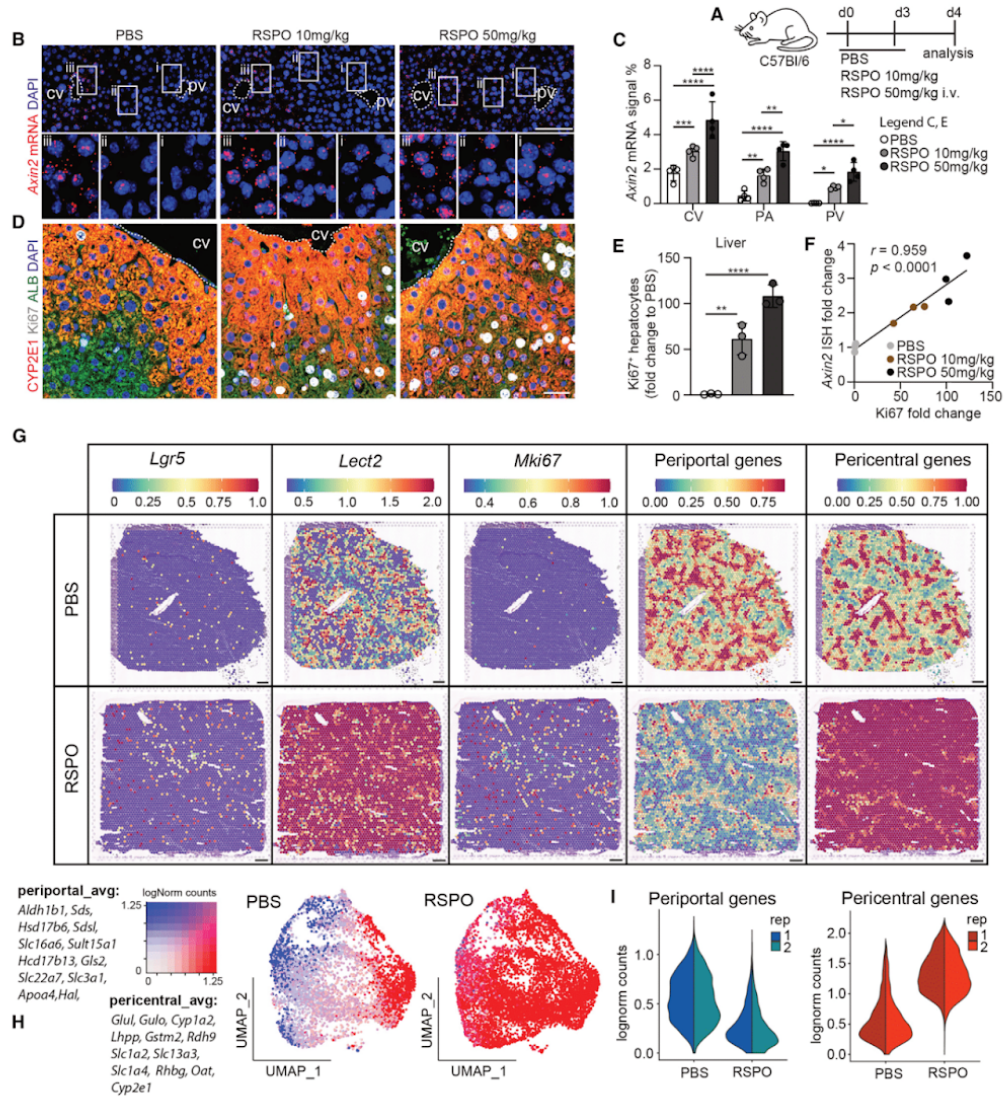
Figure 3. RSPO-induced WNT/β-Catenin signaling in the liver
(A) RSPO and PBS dosing and analysis.
(B and C) *Axin2* ISH (B) and quantification (C) in the indicated livers.
(D and E) CYP2E1, ALB, and Ki67 costaining (D) and quantification for Ki67 (E) in the indicated livers.
(F) Correlation of WNT/β-Catenin activity (*Axin2* mRNA ISH increase) with the increase in proliferation (Ki67) in hepatocytes of indicated mice.
(G) Spatial transcriptomics expression in log-normalized (logNorm) counts for selected genes in mice treated with PBS or RSPO.
(H) UMAP projection of gene expression at each spot colored by periportal (blue) and pericentral (red) metabolic gene expression signature.
(I) Violin plots showing periportal or pericentral metabolic gene expression signatures in replicate livers.
CV, central vein; PV, portail vein. Data represent mean ± SD. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. Two-way ANOVA with Sidak's test (C), one-way ANOVA with Sidak's test (E), or Pearson correlation (F) was used. Scale bars, 30 μm (B), 100 μm (D), and 500 μm (G).

representation of the local gene expression (Figure 3H). RSPO-treated livers almost entirely lost metabolic zonation as pericentral metabolic genes were expressed throughout the liver and periportal gene expression was diminished (Figures 3G–3I). Collectively, our data indicate that WNT/β-Catenin levels correlate with cell proliferation in hepatocytes and ISCs and that fine-tuning of WNT/β-Catenin activity is essential to balance hepatic metabolic zonation and proliferation.

ZNRF3 and RNF43 enable metabolic liver zonation while preventing liver tumor formation

RSPO binding to LGR4–6 receptors increases WNT/β-Catenin signaling by clearing the cell-surface transmembrane E3 ubiquitin ligases ZNRF3 and its homolog ring finger 43 (RNF43), which promote WNT receptor turnover, from the plasma membrane (Hao et al., 2012; Koo et al., 2012). Co-ISH showed coexpression of *Znrf3* and *Rnf43* in pericentral and adjacent parenchymal hepatocytes, whereas *Znrf3* was expressed in hepatocytes throughout the liver (Figure 4A). Although we have previously shown that their combined systemic deletion induced hepatocyte proliferation and extended metabolic zonation (Planas-Paz et al., 2016), confounding factors due to the global deletion did not allow final assessment of their role as potential gatekeepers of WNT/β-Catenin signaling in hepatocytes (Planas-Paz et al., 2016). Therefore, we now deleted ZNRF3 and RNF43 specifically in hepatocytes using Ad5Cre virus and studied the long-term consequence on metabolic zonation and hepatocyte proliferation (Figures 4B and S3A). A single intravenous (i.v.) injection of Ad5Cre into ZNRF3/RNF43^{FLOX} mice caused robust deletion of ZNRF3 and RNF43 specifically in hepatocytes (ZNRF3/RNF43$^{\Delta Hep}$), as indicated by absence of *Znrf3* and *Rnf43* ISH signals in most hepatocytes, compared with Ad5EGFP-injected control mice (Figure S3B). Importantly, Ad5 virus transduced nearly 80% of all hepatocytes, whereas only few other liver cells and negligible amounts of cells in other organs were infected, suggesting successful and specific Ad5-mediated hepatocyte transduction (Figures S3C–S3G). ZNRF3/RNF43$^{\Delta Hep}$ mice showed a dramatic increase in liver weight 1 year after combined deletion of ZNRF3 and RNF43 (Figure 4C). Strikingly, all ZNRF3/RNF43$^{\Delta Hep}$ mice developed multiple liver tumors (Figure 4D). ZNRF3/RNF43$^{\Delta Hep}$ livers showed increased hepatocyte proliferation throughout the liver (Figures 4E and 4F) as well as a substantial expansion of the pericentral metabolic zone, indicated by a 3-fold increase in GS+ hepatocytes (Figures 4G and 4H) and CYP2E1 expression throughout the entire liver (Figures 4I and 4J). Most of the tumors from ZNRF3/RNF43$^{\Delta Hep}$ mice were histologically classified as HCC composed of well to moderately differentiated hepatocytes forming trabeculae of multiple cell layers and focally displaying moderate nuclear pleomorphism with prominent nucleoli (Figure 4K). However, we also observed adenomas, few mixed HCC-cholangiocarcinomas (HCC-CCC), and nodular regenerative hyperplasia (NRH) in ZNRF3/RNF43$^{\Delta Hep}$ mice (Figures S3H–S3K). ZNRF3/RNF43$^{\Delta Hep}$ HCC were CCND1 and CYP2E1 positive, suggesting WNT/β-Catenin activity, and Ki67 staining indicated proliferation (Figure S3L). They expressed the periportal hepatocyte marker ARG1, which differentiates HCC from non-hepatic tumors and is expressed in the majority of well-differentiated HCC (Yan et al., 2010). Absence of the biliary marker SOX9 further confirms

good to moderate differentiation of ZNRF3/RNF43$^{\Delta Hep}$ HCC (Figure S3L). Next, we injected Axin2CreERT2^{EGFP} mice (Sun et al., 2020) with TAM to label AXIN2+ hepatocytes and induced HCC formation using the carcinogen diethylnitrosamine (DEN), enabling tumor development tracing from AXIN2+ hepatocytes for 1 year (Figure 4L). The percentage of EGFP+ DEN^{AXIN2+} HCC correlated with the percentage of initially labeled pericentral hepatocytes, confirming that AXIN2+ hepatocytes are prone to tumor formation (Figures 4M–4O), similar to LGR5+ pericentral hepatocytes (Ang et al., 2019), likely because DEN is metabolized by CYP2E1 (Kang et al., 2007). Although histological characterization of DEN^{AXIN2+} HCC revealed a similar profile as seen in ZNRF3/RNF43$^{\Delta Hep}$ HCC, including CCND1 expression, only 20% expressed CYP2E1 and GS (Figure S3M). Interestingly, 80% of the DEN^{AXIN2+} HCC expressed the periportal hepatocyte marker ARG1 (Figure S3M), although these tumors originated from AXIN2+ pericentral hepatocytes. This suggests loss of metabolic zonation and aberrant expression of metabolic enzymes in these HCC.

To further characterize EGFP+ DEN^{AXIN2} HCC (n = 8) and ZNRF3/RNF43$^{\Delta Hep}$ HCC (n = 12), we performed RNA sequencing (RNA-seq) of isolated tumors and compared their expression profiles with human HCC from The Cancer Genome Atlas (TCGA) liver dataset (Cancer Genome Atlas Research Network, 2017) harboring hotspot mutations in *CTNNB1* (n = 89) and loss-of-function mutations in the *AXIN1* gene (n = 23). After batch effect correction for species, multidimensional scaling plot (MDS) (Figure 4P) and clustering heatmap (Figure S3N) revealed that both DEN^{AXIN2} HCC and ZNRF3/RNF43$^{\Delta Hep}$ HCC groups are more similar to human *CTNNB1* mutant HCC rather than to *AXIN1* mutant HCC. However, the majority of DEN^{AXIN2} HCC showed low *Axin2* and *Glul* (encoding GS) expression similar to *AXIN1* mutant HCC (Figure 4Q). ZNRF3/RNF43$^{\Delta Hep}$ HCC showed moderate expression of both genes and weak GS staining, whereas *CTNNB1* mutant HCC showed comparably higher *Axin2* and *Glul* expression (Figures 4Q and S3L), as well as more GS staining (Adebayo Michael et al., 2019), suggesting differences in the metabolic profile of these tumors, possible because of more pronounced β-Catenin activity in *CTNNB1* mutant HCC.

Next, we subjected livers from ZNRF3/RNF43$^{\Delta Hep}$ and control mice 4 weeks after Ad5 injection, or those injected with RSPO or PBS, to bulk RNA-seq to further characterize the consequence of activating the RSPO-LGR4/5-ZNRF3/RNF43 module. Both RSPO injections and ZNRF3/RNF43 deletion promoted the expression of hallmark gene sets associated with proliferation and MYC targets activation, whereas KRAS signaling, Notch signaling, coagulation, epithelial-to-mesenchymal transition (EMT), and angiogenesis were reduced (Figure S4A). Interestingly, several gene sets showed an opposite regulation in livers with RSPO treatment compared with those with ZNRF3/RNF43 deletion, such as adipogenesis, cholesterol homeostasis, oxidative phosphorylation, P53 pathway, and TGF-β signaling. Possibly, RSPO injections promoted a slightly different gene expression program because of the uneven distribution of ZNRF3 (periportal and pericentral) and RNF43 (only pericentral). It is also possible that secreted factors, induced by RSPO in extra-hepatic tissues, influenced gene expression in the liver.

Cell Stem Cell
Article

CellPress
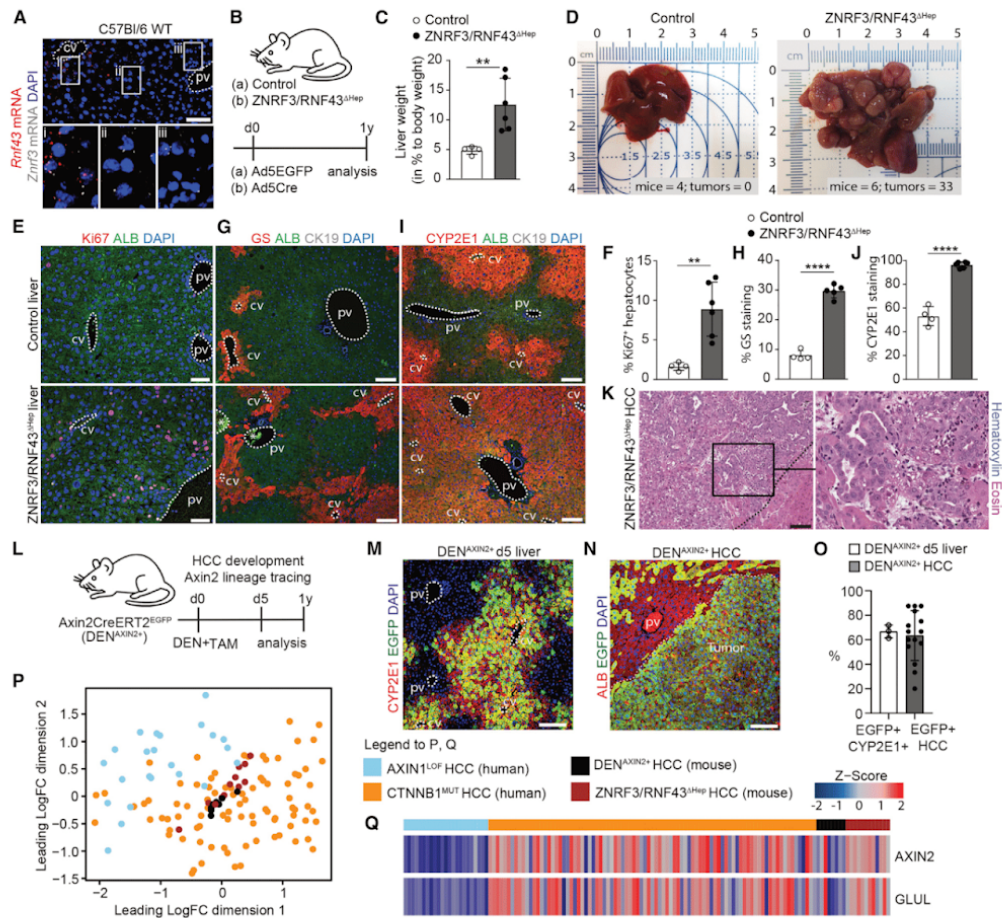
Figure 4. ZNRF3 and RNF43 are guardians of the pericentral niche and control hepatic WNT/β-Catenin signaling levels
(A) *Znrf3* and *Rnf43* co-ISH staining in naive liver.
(B) Aging of ZNRF3/RNF43$^{\Delta Hep}$ and control mice (generation scheme in Figure S3A).
(C) Liver weight increase in ZNRF3/RNF43$^{\Delta Hep}$ mice.
(D) Liver tumor photographs and incidence in ZNRF3/RNF43$^{\Delta Hep}$ mice.
(E and F) Ki67 and ALB costaining (E) and Ki67 quantification (F) in indicated mice.
(G and H) GS, ALB, and CK19 costaining (G) and GS quantification (H) in indicated mice.
(I and J) CYP2E1, ALB, and CK19 costaining (I) and CYP2E1 quantification (J) in indicated mice.
(K) H&E staining of ZNRF3/RNF43$^{\Delta Hep}$ HCC. Image acquisition involved stitching.
(L) AXIN2 lineage tracing in DEN HCC model.
(M and N) Costaining for EGFP and CYP2E1 (M) or EGFP and ALB (N) in the indicated mice.
(O) EGFP labeling efficiency of CYP2E1+ hepatocytes during initiation of lineage tracing and similar percentage of EGFP+ HCCs after 1 year.
(P and Q) Multidimensional scaling plot (MDS) (P) and expression of *AXIN2* and *GLUL* in indicated HCCs. CV, central vein; PV, portal vein.
CV, central vein; PV, portal vein. Data represent mean ± SD. **p < 0.01 and ****p < 0.0001. Two-tailed unpaired Student's t test was used (C, F, H, and J). Scale bars, 50 μm (E) and 100 μm (A, G, I, K, M, and N).

As the fatty acid metabolism gene set was upregulated in both groups (Figure S4A), we assessed whether ZNRF3/RNF43$^{\Delta Hep}$ mice developed a nonalcoholic steatohepatitis (NASH)-like phenotype. Although a small subset of ZNRF3/RNF43$^{\Delta Hep}$ tumors showed steatosis, the majority of tumors and neighboring liver tissue showed even reduced steatosis compared with 1-year-old age-matched control mice (Figures S4B and S4C). We further did not observe significant fibrosis in these mice (Figures S4D and S4E), suggesting that the metabolic changes in ZNRF3/RNF43$^{\Delta Hep}$ mice did not promote a NASH-like phenotype, consistent with previous reports suggesting that abrogated rather than activated WNT/β-Catenin activity was associated with steatosis (Go et al., 2014). The increased gene program driving proliferation, the resulting high penetrance of liver tumor formation, and the dramatic expansion of pericentral metabolic enzymes collectively suggest that ZNRF3 and RNF43 are essential gatekeepers of hepatic WNT/β-Catenin signaling, balancing metabolic zonation and hepatocyte proliferation.

ZNRF3/RNF43 deletion promotes clonal hepatocyte expansion and *de novo* induction of pericentral metabolic genes

To assess whether ZNRF3/RNF43 deletion promotes expansion of AXIN2+ hepatocytes toward the periportal zone or whether pericentral metabolic genes are induced *de novo* in this region, we performed lineage tracing of AXIN2+ hepatocytes with ZNRF3/RNF43 deletion. Therefore, we generated ZNRF3/RNF43$^{EGFP\Delta AXIN2+}$ mice that allow TAM-induced deletion of ZNRF3/RNF43 in AXIN2+ hepatocytes labeled with EGFP. Four weeks after low-dose TAM injection, ZNRF3/RNF43$^{EGFP\Delta AXIN2+}$ mice showed EGFP-labeled hepatocyte clones in the CYP2E1+ zone consisting of two or more cells, whereas Axin2CreERT2^{EGFP} controls showed mostly individual labeled hepatocytes (Figures 5A–5C). To enable lineage tracing of hepatocytes with ZNRF3/RNF43 deletion across the different hepatic zones, we introduced an R26-EGFP reporter into ZNRF3/RNF43$^{\Delta Hep}$ mice to generate ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice (Figure 5D). Although R26^{HepEGFP} control mice showed mostly individual EGFP-labeled hepatocytes in all liver zones, ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice showed clonal growth of labeled hepatocytes 4 weeks after low-dose Ad5Cre injection and further increase in clone size after 8 weeks (Figures 5E and 5F). Interestingly, only a subset of hepatocytes with ZNRF3/RNF43 deletion showed clonal expansion (Figures 5C and 5F). Although individual EGFP+ periportal hepatocytes in ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice showed *de novo* expression of the pericentral metabolic enzymes CYP2E1 or GS, this was not observed in EGFP+ periportal hepatocytes of R26^{HepEGFP} controls. Only the larger hepatocyte clones in ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice were Ki67+ but showed only minimal induction of CYP2E1 and GS expression (Figures 5G–5I). Also during regeneration following PHx, a subset of EGFP-labeled hepatocytes in ZNRF3/RNF43$^{EGFP\Delta AXIN2+}$ mice (Figures S5A–S5C) or ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice (Figures S5D–S5F) showed enhanced clonal expansion compared with their respective controls. The *de novo* induction of WNT/β-Catenin-controlled metabolic genes and hepatocyte proliferation in ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice outside of the zone of homeostatic hepatic WNT/β-Catenin activity suggests that hepatocytes in this zone have access to WNT ligands that are required for RSPO-LGR4/5-ZNRF3/RNF43-

controlled WNT/β-Catenin signaling. Indeed, ISH for several WNT ligands shows that although *Wnt9b* is highly enriched in the pericentral zone as previously reported (Ma et al., 2020; Preziosi et al., 2018; Wang et al., 2015), low levels of *Wnt7a*, *Wnt7b*, and *Wnt9b* are available to periportal hepatocytes, likely enabling WNT/β-Catenin signaling in this zone following ZNRF3/RNF43 deletion (Figures S5G and S5H).

To further characterize the effects of ZNRF3/RNF43 deletion on metabolic zonation and hepatocyte proliferation, we performed spatial transcriptomics on ZNRF3/RNF43$^{\Delta Hep}$ and control livers 4 weeks after Ad5Cre or Ad5EGFP injection, respectively. Comparable numbers of sequenced genes and UMIs for two replicate samples from two individual mice per group suggest no bias for downstream analyses (Figure S6A). Fourteen different clusters were identified and enabled reconstruction of the metabolic zones by identifying hepatocyte populations on the basis of differentially expressed genes (periportal: clusters 2, 3, and 9; pericentral: clusters 1, 6, and 12; midzonal: cluster 0) and blood vessels (cluster 7). Although ZNRF3/RNF43$^{\Delta Hep}$ mice showed changes in the expression of the genes defining periportal (e.g., *Hsd17b13*, *Pck1*, *Arg1*) or pericentral (e.g., *Glul*, *Cyp2e1*, *Oat*) hepatocytes, compared with control mice, unsupervised clustering correctly located spots to the corresponding liver zone in both groups (Figures S6B and 6C). As the diameter of the spots is about twice the diameter of a hepatocyte, we also observed smaller clusters with a mixed hepatocyte-immune cell signature (Figure S6B) and excluded those to not bias the results of our comparative gene expression analyses. Using *Mki67* expression values (Figures 6A, 6B, and S6B), we found proliferating hepatocytes to be enriched in cluster 13. Cluster identity of the spatial transcriptomics spots (Figure 6C) localized the proliferating hepatocytes (cluster 13) mostly between the periportal and pericentral liver zones in control mice, consistent with recent reports suggesting that midzonal hepatocytes show increased proliferative potential during liver homeostasis (Chen et al., 2020a; He et al., 2021; Wei et al., 2021). Importantly, spots with proliferating hepatocytes (cluster 13) were more abundant in ZNRF3/RNF43$^{\Delta Hep}$ mice, confirming increased proliferation upon ZNRF3/RNF43 deletion (Figure 6C). To compare these highly proliferative hepatocytes with the complement hepatocyte pool, we performed differential gene expression (Figure 6D) and gene set enrichment (Figure S6C) analyses of cluster 13 versus the combined periportal or pericentral clusters in ZNRF3/RNF43$^{\Delta Hep}$ mice. Gene sets associated with proliferation (Figure S6C) and leading-edge cell cycle genes (Figure 6D) were highly enriched in cluster 13 compared with periportal and pericentral hepatocyte clusters. Moreover, metabolic gene sets (Figure S6C) and pericentral and periportal metabolic genes (Figure 6D) were reduced in cluster 13. Interestingly, MYC target genes were upregulated in cluster 13 compared with the metabolically active complement hepatocyte clusters. Together, our spatial transcriptomics profiling revealed a cluster of highly proliferative hepatocytes with reduced metabolic gene expression, similar to what we observed during clonal tracing of hepatocytes with ZNRF3/RNF43 deletion.

ZNRF3 and RNF43 cooperate to safeguard hepatic function and proliferation

Interestingly, somatic mutations in *ZNRF3* or *RNF43* are rare, and we did not observe simultaneous mutations among 1,431 human

Cell Stem Cell
Article

CellPress

Figure 5. ZNRF3/RNF43 deletion promotes hepatocyte proliferation and *de novo* induction of pericentral metabolic gene expression

(A) TAM-induced ZNRF3/RNF43 deletion in EGFP-labeled AXIN2+ hepatocytes.

(B and C) EGFP and CYP2E1 costaining (B) and clone size of EGFP-labeled hepatocytes (C) in the indicated mice.

(D) Ad5Cre-induced deletion of ZNRF3/RNF43 in EGFP-labeled hepatocytes.

(E and F) EGFP, ALB, and CK19 costaining (E) and clone size of EGFP-labeled hepatocytes (F) in the indicated mice.

(G–I) EGFP, Ki67, and CYP2E1 costaining (G), EGFP and GS costaining (H), and quantification of CYP2E1 and GS staining intensity in EGFP+ hepatocytes (I) in the indicated mice.

CV, central vein; PV, portal vein. Yellow arrowhead, single EGFP-labeled hepatocytes; white arrowhead, hepatocyte clones (B and E). Data represent mean ± SD. *p < 0.05, **p < 0.01, and ****p < 0.0001. Two-tailed unpaired Student's t test (I) or two-way ANOVA (C and F) was used. Scale bars, 50 μm (B, E, G, and H).

HCC (Figure S6D). To assess whether potential functional redundancy may constitute a safeguarding mechanism preventing uncontrolled hepatocyte proliferation and securing integrity of metabolic zonation, we generated mice with individual Ad5Cre-mediated deletion of RNF43 (RNF43$^{\Delta Hep}$) or ZNRF3 (ZNRF3$^{\Delta Hep}$) and compared them with ZNRF3/RNF43$^{\Delta Hep}$ and control mice 4 weeks after Ad5Cre or Ad5EGFP injection, respectively (Figure 7A). RNF43$^{\Delta Hep}$ mice showed increased *Axin2* mRNA ISH signals restricted to the hepatic zones where RNF43 is normally expressed (Figures 7B, 7C, 4A, and S3B). Although GS and CYP2E1 expression was not affected by RNF43 deletion (Figures 7D, 7E, S7A, and S7B), pericentral and adjacent parenchymal hepatocytes showed increased proliferation (Figures 7D and 7F). This suggests that RNF43 restricts WNT/β-Catenin signaling and hepatocyte proliferation in AXIN2+ hepatocytes where it is normally expressed. Likewise, *Axin2* mRNA ISH signals (Figures 7B and 7C) and hepatocyte proliferation (Figures 7D and 7F) were increased in all three liver zones in ZNRF3$^{\Delta Hep}$ mice, consistent

191

Figure 6. Spatial transcriptomics profiling reveals cluster of proliferating hepatocytes
(A) UMAP plots of spatial transcriptomics data showing clusters in ZNRF3/RNF43$^{\Delta Hep}$ and control mice (clusters with hepatocyte populations and blood cells were marked).
(B) Violin plot showing *Mki67* expression in identified hepatocyte clusters (replicates from two mice).
(C) Spatial projection of selected clusters onto H&E stainings of the indicated mice. Scale bars, 500 μm.
(D) Heatmap showing expression (*Z* score) of cell cycle genes, pericentral metabolic genes, and periportal metabolic genes in cluster 13 versus periportal and pericentral clusters (randomly downsampled to the number of spots in cluster 13 [n = 134]) in ZNRF3/RNF43$^{\Delta Hep}$ mice. CV, central vein; PV, portal vein.

with the non-zonated expression profile of ZNRF3 (Figures 4A and S3B). Although GS expression was not significantly increased in ZNRF3$^{\Delta Hep}$ mice, CYP2E1 expression was moderately expanded upon ZNRF3 deletion. In contrast, combined deletion of ZNRF3 and RNF43 dramatically expanded the pericentral metabolic program as indicated by CYP2E1 expression throughout the entire liver and increased numbers of GS+ hepatocytes (Figures 7D, 7E, S7A, and S7B). ZNRF3/RNF43$^{\Delta Hep}$ mice showed significantly more induction of WNT/β-Catenin signaling, associated with increased hepatocyte proliferation throughout the liver compared with RNF43$^{\Delta Hep}$ or ZNRF3$^{\Delta Hep}$ mice (Figures 7B–7D and 7F). These findings were confirmed following TAM-induced deletion of RNF43 (RNF43$^{\Delta AXIN2+}$ mice), ZNRF3 (ZNRF3$^{\Delta AXIN2+}$ mice), and ZNRF3/RNF43 (ZNRF3/RNF43$^{\Delta AXIN2+}$ mice) in AXIN2+ cells using Axin2-CreERT2 mice (Figures S7C–S7G). Interestingly, proliferating hepatocytes in RNF43$^{\Delta Hep}$, ZNRF3$^{\Delta Hep}$, and ZNRF3/RNF43$^{\Delta Hep}$ mice showed reduced expression of CYP2E1 (Figure 7D, magnified insets). Moreover, lineage tracing

of AXIN2+ hepatocytes with either RNF43 or ZNRF3 deletion showed clonal hepatocyte expansion (Figures S7H–S7J), although not to the extent of combined ZNRF3/RNF43 deletion (Figures 5A–5C). This suggests that ZNRF3 and RNF43 cooperate to restrict hepatic WNT/β-Catenin signaling during homeostasis. Consistent with previous findings (Koo et al., 2012), deletion of ZNRF3/RNF43 in AXIN2+ cells resulted in a dramatic increase of *Axin2* expression and cell proliferation in the small intestine, associated with aberrant growth (Figures S7K and S7L).

Interestingly, ZNRF3 deletion induced *Rnf43* expression, which is itself a β-Catenin target gene (Hao et al., 2012), throughout the liver of ZNRF3$^{\Delta Hep}$ mice (Figures 7G and 7H), likely explaining why RNF43 deletion concomitant to loss of ZNRF3 increased hepatocyte proliferation in all three liver zones and not just where RNF43 is expressed during homeostasis. Likewise, co-ISH for *Axin2* and *Rnf43* showed *Rnf43* upregulation in several liver injury models with activated WNT/β-Catenin signaling (Figure 7I). This suggests that *Rnf43* upregulation

CellPress
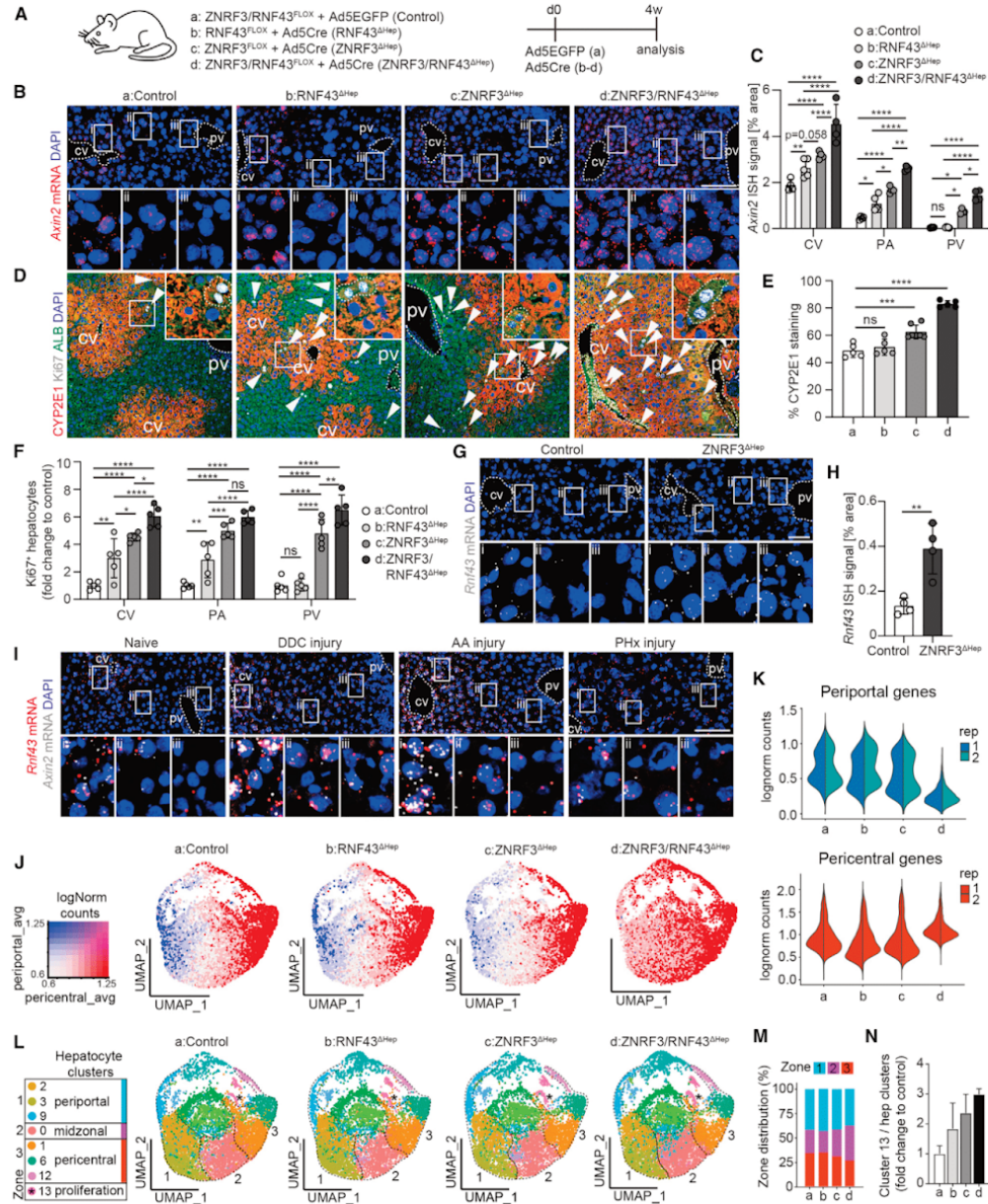
Figure 7. ZNRF3 and RNF43 cooperate to safeguard hepatic metabolic zonation and hepatocyte proliferation

(A) Generation and analysis of ZNRF3$^{\Delta Hep}$, RNF43$^{\Delta Hep}$, ZNRF3/RNF43$^{\Delta Hep}$, and control mice.

(B and C) Axin2 ISH (B) and quantification (C) in the indicated mice.

193

may limit WNT/β-Catenin activity to physiological levels, preventing uncontrolled hepatocyte proliferation and disruption of metabolic zonation. Collectively, our data indicate that ZNRF3 and RNF43 cooperate to safeguard hepatic function by spatially and temporally restricting hepatocyte proliferation while maintaining liver zonation and metabolic function.

To further substantiate this conclusion, we reconstructed metabolic zones in control, RNF43$^{\Delta Hep}$, ZNRF3$^{\Delta Hep}$, and ZNRF3/RNF43$^{\Delta Hep}$ livers using spatial transcriptomics (Figures 7J–7N and S7M–S7Q). In control, RNF43$^{\Delta Hep}$ and ZNRF3$^{\Delta Hep}$ livers the expression of pericentral and periportal metabolic genes followed a similar spatial pattern with comparable expression values, with CYP2E1 expression being moderately increased and periportal genes slightly reduced in ZNRF3$^{\Delta Hep}$ livers. In contrast, in ZNRF3/RNF43$^{\Delta Hep}$ livers the pericentral metabolic gene program expanded across the entire liver, displacing periportal metabolic gene expression (Figures 7J, K, and S7N–S7Q). As mentioned above, unsupervised clustering correctly identified hepatocytes in the three hepatic zones, irrespective of the changes in the expression of β-Catenin-controlled metabolic genes (Figures 6A, 6C, and S6B). All mice showed a similar distribution of clusters associated with the different zones (Figures 7L and 7M), further suggesting *de novo* expression of pericentral metabolic genes in periportal hepatocytes of ZNRF3/RNF43$^{\Delta Hep}$ livers, rather than expansion of pericentral hepatocytes into the periportal zone. The increase in cluster 13 abundance (Figures 7L–7N) correlates with the increased hepatocyte proliferation seen in RNF43$^{\Delta Hep}$, ZNRF3$^{\Delta Hep}$, and ZNRF3/RNF43$^{\Delta Hep}$ livers (Figures 7D and 7F). Together, our data establish that ZNRF3 and RNF43 cooperate to balance metabolic zonation and proliferation in the liver.

DISCUSSION

A large number of liver tumors are characterized by hyperactive WNT/β-Catenin signaling, and hepatocytes with WNT/β-Catenin activity have high susceptibility to give rise to liver tumors in experimental models (Ang et al., 2019; Perugorria et al., 2019; Russell and Monga, 2018). It is therefore conceivable that constant hepatic WNT/β-Catenin signaling, which is necessary to ensure metabolic homeostasis, also increases the risk for unwanted proliferation. The liver therefore requires tunable mechanisms restricting hepatocyte proliferation in homeostasis and enabling it during regeneration, while preserving WNT/β-Catenin-controlled metabolic liver zonation. Because in naive mice, the majority of AXIN2+ ISCs proliferate, in contrast to AXIN2+ hepatocytes (Sun et al., 2020), we hypothesized that differences between these two cell types may highlight mechanisms preventing proliferation of AXIN2+ hepatocytes during liver homeo-

stasis. We speculated that restricted chromatin accessibility of cell cycle gene-regulatory elements might be responsible for the absence of proliferation of AXIN2+ hepatocytes. Despite their very low cycling rates, both AXIN2+ and AXIN2− hepatocytes displayed accessible chromatin at the TSS of cell cycle genes similar to the signal observed in highly proliferative ISCs, suggesting a state permissive for cell proliferation. Likewise, chromatin accessibility at the TSS was detected at pericentral and periportal metabolic genes in both AXIN2+ and AXIN2− hepatocytes. Such positive ATAC signal at the TSS of non-expressed genes could imply a promoter enabled to quickly respond to an instructing upstream signal, but additional experimental analysis would be required to validate this proposed model. Nevertheless, intriguingly, this finding may provide one mechanistic element for how hepatocytes in all liver zones can readily proliferate on demand or adopt a different metabolic profile following WNT/β-Catenin pathway modulation (Colnot, 2011; Russell and Monga, 2018; Sun et al., 2020). In contrast, ISCs differ in their epigenetic profile from their more differentiated descendants (Jadhav et al., 2017). We now show that the chromatin status in ISCs is not permissive for the expression of WNT/β-Catenin-regulated metabolic genes. TSA promoted the expression of hepatic metabolic genes in intestinal organoids, presumably because of increased chromatin accessibility at their TSS, while impairing organoid growth. It is conceivable that silencing of these metabolic genes may support the proliferative capacity of ISCs, allowing them to use WNT/β-Catenin signaling primarily to drive stem cell proliferation and tissue homeostasis. However, inducing chromatin accessibility specific for metabolic genes in ISCs will be needed to validate this hypothesis. In addition, impaired intestinal organoid growth following overexpression of metabolic genes suggests that metabolic gene expression can also influence ISC proliferation independently of the aforementioned possible competition for β-Catenin. Although our data clearly demonstrate an inverse correlation between metabolic gene expression and proliferation in ISCs, the responsible mechanisms remain to be identified.

Although this excludes that restricted chromatin accessibility prevents proliferation of AXIN2+ hepatocytes while securing metabolic liver zonation during liver homeostasis, we now show that spatiotemporal fine-tuning of WNT/β-Catenin activity is directing these processes. Liver regeneration is characterized by the concerted action of many cell types and signaling pathways (Michalopoulos and DeFrances, 1997; Miyajima et al., 2014; Stanger, 2015). Several reports showed that loss of WNT/β-Catenin pathway components impair or delay liver regeneration (Perugorria et al., 2019; Planas-Paz et al., 2016, 2019; Russell and Monga, 2018; Sekine et al., 2006; Tan et al., 2006; Zhao et al., 2019), whereas WNT/β-Catenin activation

(D–F) CYP2E1, ALB, and Ki67 costaining (D) and quantification for CYP2E1 (E) and Ki67 (F) in the indicated mice (arrows mark proliferating hepatocytes; magnified insets show reduced CYP2E1 expression in Ki67+ hepatocytes).
(G and H) *Rnf43* ISH (G) and quantification (H) in control and ZNRF3$^{\Delta Hep}$ mice.
(I) *Axin2* and *Rnf43* co-ISH, showing *Rnf43* upregulation following AA, DDC, or PHx injury.
(J and K) Co-expression UMAP plot of periportal (blue) and pericentral (red) metabolic gene signature (J) and expression distribution (K) in the indicated mice.
(L–N) UMAP plots with spots color-coded by cluster membership (L) reveal similar percentage of clusters related to the three liver zones (M) and differences in cluster 13 abundance (N).
CV, central vein; PV, portal vein. Data represent mean ± SD. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001; ns, not significant. Two-way ANOVA with Sidak's test (C and F) or one-way ANOVA with Sidak's test (E) or unpaired Student's t test (H) was used. Scale bars, 50 μm (G and I) and 100 μm (B and D).

enhances this process (Alvarado et al., 2016; Planas-Paz et al., 2016). We now show a RSPO-mediated dose-dependent increase in proliferation of homeostatic hepatocytes, as well as a remarkable correlation between WNT/β-Catenin activity and hepatocyte proliferation across several liver injury models and postnatal liver development, suggesting that spatial and temporal fine-tuning of WNT/β-Catenin activity directs hepatocyte proliferation. Moreover, we show that a dose-dependent RSPO-mediated increase in WNT/β-Catenin signaling expanded the pericentral metabolic program throughout the liver, replacing the periportal metabolic gene program. Our data extend previous findings, suggesting that deregulated hepatic WNT/β-Catenin signaling impairs metabolic zonation (Colnot, 2011; Russell and Monga, 2018), by using spatial resolution in conjunction with genome-wide gene expression profiling. RSPO injections or ZNRF3/RNF43 deletion also increased ISC proliferation, whereas scavenging of endogenous RSPO caused the opposite effect. However, significantly higher RSPO3 levels in the crypt niche and resulting higher WNT/β-Catenin pathway activity, compared with the liver, direct constant cell proliferation required for intestinal homeostasis. Collectively, our data suggest that fine-tuning of WNT/β-Catenin levels enables metabolic liver zonation while keeping basal hepatocyte proliferation rates low.

Importantly, we show that ZNRF3 and RNF43 safeguard liver homeostasis by restricting hepatocyte proliferation and preserving metabolic zonation. Unlike LGR4 and LGR5, which largely differ in their capacity to direct hepatic WNT/β-Catenin signaling (Planas-Paz et al., 2016), ZNRF3 and RNF43 cooperate to keep WNT/β-Catenin activity under control. LGR4 is dominant over LGR5, suggested by its ability to fully compensate for the loss of LGR5 and the almost complete lack of hepatic WNT/β-Catenin activity upon LGR4 deletion (Planas-Paz et al., 2016). Individual deletion of ZNRF3 or RNF43 increased WNT/β-Catenin activity and induced hepatocyte proliferation in zones where these genes are normally expressed. Periportal hepatocytes express only ZNRF3, whereas pericentral hepatocytes co-express both ZNRF3 and RNF43. Lack of RSPO availability in zone 1 may only require ZNRF3 to block WNT/β-Catenin signaling and may enable induction of WNT/β-Catenin signaling and proliferation in periportal hepatocytes following moderate expression of RSPO3 in this zone following injury (Planas-Paz et al., 2019). In contrast, co-expression of ZNRF3 and RNF43 may be required to limit WNT/β-Catenin signaling in pericentral hepatocytes because of permanent availability of RSPO3 in zone 3. Induced WNT/β-Catenin signaling and proliferation in periportal hepatocytes upon RSPO injection or deletion of ZNRF3 (or ZNRF3/RNF43) suggests that availability of low levels of WNT ligands in this zone is sufficient to activate the pathway and highlights the importance of this gatekeeping mechanism. RNF43 is upregulated throughout the liver in response to ZNRF3 deletion, attenuating WNT/β-Catenin activity to balance hepatocyte proliferation while mostly retaining metabolic zonation. Likewise, RNF43 upregulation in response to different types of liver injury suggests that the same mechanism may restrict uncontrolled hepatocyte proliferation and retain metabolic zonation during liver regeneration. In contrast, combined deletion of ZNRF3 and RNF43 increased WNT/β-Catenin signaling to such an extent that metabolic zonation was impaired and uncontrolled

proliferation resulted into HCC formation. Interestingly, a preprint manuscript suggests that ZNRF3/RNF43 deletion in hepatocytes induces NASH and promotes HCC formation in response to chronic liver injury (Mastrogiovanni et al., 2020). Although we did not observe NASH in our mice, their study supports the tumor-suppressive role of ZNRF3/RNF43.

Recently, it was reported that regenerating hepatocytes undergo division of labor, where a subset maintains metabolic function whereas others proliferate (Chen et al., 2020b; Walesky et al., 2020). Our clonal tracing data now suggest that this functional segregation also occurs in hepatocytes with active WNT/β-Catenin signaling. Hepatocytes with marked metabolic gene expression remained mostly as single clones, suggesting that this segregation may not be transient with hepatocytes switching between a metabolic and proliferative state. However, we cannot exclude that larger hepatocyte clones, which mostly showed reduced metabolic gene expression at the point of analysis, had a different metabolic status earlier in the experiment. Importantly, also following ZNRF3/RNF43 deletion in the majority of hepatocytes, only a subset was proliferating and showed reduced metabolic gene expression compared with neighboring hepatocytes. More recently, midzonal hepatocytes were shown to have increased proliferative capacity during liver homeostasis (Chen et al., 2020a; He et al., 2021; Wei et al., 2021). Although this was not readily detected by Ki67 staining, our spatial transcriptomics analysis identified a cluster of hepatocytes with proliferative signature and reduced metabolic gene expression in zone 2. Together, this suggests an inverse correlation between metabolic gene expression and proliferation, similar to what we observed in ISCs. Given the importance of WNT/β-Catenin signaling in promoting both metabolic function and hepatocyte proliferation (Russell and Monga, 2018), it will be essential to elucidate the additional mechanisms directing the functional assignment in hepatocytes with WNT/β-Catenin activity. The proliferative hepatocyte cluster we identified by spatial transcriptomics showed increased MYC signaling, and MYC is a WNT/β-Catenin target gene driving proliferation (He et al., 1998). Similarly to ISCs in the intestine, the influence of metabolic gene expression on hepatocyte proliferation and the mechanisms behind the inverse correlation remain to be clarified.

Human CTNNB1 mutant HCC show robust expression of pericentral metabolic enzymes and are generally more differentiated and less proliferative. In contrast, human HCC with loss of function in upstream regulators of WNT/β-Catenin signaling show modest WNT/β-Catenin activity, less metabolic enzyme expression, and more proliferation (Boyault et al., 2007; Hsu et al., 2000). Although comparative RNA-seq analyses suggest that the overall gene expression profile in mouse ZNRF3/RNF43$^{\Delta Hep}$ HCC is more similar to that in CTNNB1 mutant rather than AXIN1 mutant HCC, they show lower WNT/β-Catenin activity and GS expression than CTNNB1 mutant HCC. It is possible that the intact β-Catenin in tumors with mutations in upstream negative regulators accounts for the differential gene regulation of metabolic and proliferation genes. Better understanding the transcriptional regulation of metabolic and proliferative genes by β-Catenin will be necessary to identify the mechanisms inducing diverging functions in hepatocytes, maintaining metabolic function while allowing the proliferative capacity necessary to maintain liver mass. In addition, more work will be needed to identify

potential partners supporting ZNRF3 and RNF43 to safeguard the liver during homeostasis and regeneration while promoting proliferation on demand. For example, the WNT/β-Catenin pathway cooperates with several other cell cycle regulators, such as YAP/TAZ (Azzolin et al., 2014). Hepatic YAP activity seems to follow a porto-central activity gradient (Fitamant et al., 2015) and is involved in regulating hepatocyte plasticity and proliferation (Bhushan et al., 2020; Li et al., 2019; Yimlamai et al., 2014).

As diverse β-Catenin-controlled metabolic genes have a different expression pattern (e.g., GS is restricted to one or two pericentral hepatocyte layers, whereas CYP2E1 is expressed in all AXIN2+ hepatocytes), it is likely that different thresholds exist for β-Catenin to induce the expression of these genes. Likewise, moderate WNT/β-Catenin signaling induction in ZNRF3$^{\Delta Hep}$ mice was sufficient to expand CYP2E1 expression, whereas only in ZNRF3/RNF43$^{\Delta Hep}$ or RSPO-injected mice, the pericentral metabolic gene program largely replaced the periportal program. Reconstructing hepatic zones using unsupervised clustering of spatial transcriptomics data and using lineage tracing of hepatocytes with induced ZNRF3/RNF43 deletion, we show that *de novo* upregulation of pericentral metabolic genes in periportal hepatocytes rather than the expansion of pericentral hepatocytes toward this zone causes metabolic reprogramming of the liver. Understanding how the expression of different metabolic enzymes is distributed alongside the centro-portal WNT/β-Catenin activity gradient will be important to shed additional light on the division of labor in functionally diverse hepatocyte subsets.

In summary, our data show that fine-tuning of WNT/β-Catenin activity in hepatocytes, rather than their intrinsic responsiveness controlled by chromatin accessibility, maintains the balance between hepatocyte proliferation and metabolic zonation. Functional redundancy in ZNRF3 and RNF43 enables spatial and temporal control of WNT/β-Catenin activity, preventing uncontrolled proliferation and tumor formation.

Limitations of the study

First, in addition to the upregulation of metabolic genes that we observed, TSA-mediated effects on chromatin would be expected to occur in a locus-unspecific manner, and thus additional gene expression changes may have influenced ISC proliferation. Second, adenoviral immunogenicity may affect hepatic gene expression and proliferation. However, both our knockout (KO) and control mice received adenovirus injections, and we used Axin2CreERT2-mediated gene deletion for phenotype validation. Third, although 10X Visium offers whole-genome spatial transcriptomics, the diameter of the spots is about twice the diameter of a hepatocyte. Although we excluded spots with a mixed hepatocyte-immune cell signature from the analyses and correctly identified hepatocytes assigned to the different hepatic zones, spots might also detect gene expression from endothelial cells lining the captured hepatocytes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse models
  - *In vivo* mouse studies
- METHOD DETAILS
  - Immunohistochemistry (IHC) and *in situ* hybridization (ISH)
  - Intestinal organoid culture and staining
  - Lentiviral packaging
  - RNA extraction, reverse transcription and quantitative RT-PCR (qPCR)
  - Bulk RNA-seq of mouse livers and HCCs, and count generation
  - Human HCC data analysis
  - Bulk RNA-seq data analysis of mice with ZNRF3/RNF43 deletion or RSPO injection, and respective control mice
  - Spatial transcriptomics and bioinformatics analysis
  - ATAC sequencing of hepatocytes and AXIN2+ intestinal cells
  - Human HCC somatic mutation analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

Cell Stem Cell
Article

REFERENCES

Adebayo Michael, A.O., Ko, S., Tao, J., Moghe, A., Yang, H., Xu, M., Russell, J.O., Pradhan-Sundd, T., Liu, S., Singh, S., et al. (2019). Inhibiting glutamine-dependent mTORC1 activation ameliorates liver cancers driven by β-catenin mutations. Cell Metab. *29*, 1135–1150.e6.

Alvarado, T.F., Puliga, E., Preziosi, M., Poddar, M., Singh, S., Columbano, A., Nejak-Bowen, K., and Monga, S.P. (2016). Thyroid hormone receptor β agonist induces β-catenin-dependent hepatocyte proliferation in mice: implications in hepatic regeneration. Gene Expr. *17*, 19–34.

Ang, C.H., Hsu, S.H., Guo, F., Tan, C.T., Yu, V.C., Visvader, J.E., Chow, P.K.H., and Fu, N.Y. (2019). Lgr5+ pericentral hepatocytes are self-maintained in normal liver regeneration and susceptible to hepatocarcinogenesis. Proc. Natl. Acad. Sci. U S A *116*, 19530–19540.

Azzolin, L., Panciera, T., Soligo, S., Enzo, E., Bicciato, S., Dupont, S., Bresolin, S., Frasson, C., Basso, G., Guzzardo, V., et al. (2014). YAP/TAZ incorporation in the β-catenin destruction complex orchestrates the Wnt response. Cell *158*, 157–170.

Barker, N., Bartfeld, S., and Clevers, H. (2010). Tissue-resident adult stem cell populations of rapidly self-renewing organs. Cell Stem Cell *7*, 656–670.

Benhamouche, S., Decaens, T., Godard, C., Chambrey, R., Rickman, D.S., Moinard, C., Vasseur-Cognet, M., Kuo, C.J., Kahn, A., Perret, C., and Colnot, S. (2006). Apc tumor suppressor gene is the "zonation-keeper" of mouse liver. Dev. Cell *10*, 759–770.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B *57*, 289–300.

Bhardwaj, V., Heyne, S., Sikora, K., Rabbani, L., Rauer, M., Kilpert, F., Richter, A.S., Ryan, D.P., and Manke, T. (2019). snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. Bioinformatics *35*, 4757–4759.

Bhushan, B., Molina, L., Koral, K., Stoops, J.W., Mars, W.M., Banerjee, S., Orr, A., Paranjpe, S., Monga, S.P., Locker, J., et al. (2020). Yap-associated protein is crucial for constitutive androstane receptor-driven hepatocyte proliferation but not for induction of drug metabolism genes in mice. Hepatology *73*, 2005–2022.

Boyault, S., Rickman, D.S., de Reyniès, A., Balabaud, C., Rebouissou, S., Jeannot, E., Hérault, A., Saric, J., Belghiti, J., Franco, D., et al. (2007). Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology *45*, 42–52.

Burke, Z.D., Reed, K.R., Phesse, T.J., Sansom, O.J., Clarke, A.R., and Tosh, D. (2009). Liver zonation occurs through a beta-catenin-dependent, c-Myc-independent mechanism. Gastroenterology *136*, 2316–2324.e1–e3.

Cancer Genome Atlas Research Network (2017). Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell *169*, 1327–1341.e23.

Chen, F., Jimenez, R.J., Sharma, K., Luu, H.Y., Hsu, B.Y., Ravindranathan, A., Stohr, B.A., and Willenbring, H. (2020a). Broad distribution of hepatocyte proliferation in liver homeostasis and regeneration. Cell Stem Cell *26*, 27–33.e4.

Chen, T., Oh, S., Gregory, S., Shen, X., and Diehl, A.M. (2020b). Single-cell omics analysis reveals functional diversification of hepatocytes during liver regeneration. JCI Insight *5*, e141024.

Clevers, H., Loh, K.M., and Nusse, R. (2014). Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. Science *346*, 1248012.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71.

Colnot, S.P.C. (2011). Liver zonation. In Molecular Pathology of Liver Diseases S, P. Monga, ed. (New York: Springer Science and Business Media).

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. *43*, D662–D669.

de Lau, W., Barker, N., Low, T.Y., Koo, B.K., Li, V.S., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M., et al. (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature *476*, 293–297.

Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al.; MC3 Working Group; Cancer Genome Atlas Research Network (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, 271–281.e7.

Fitamant, J., Kottakis, F., Benhamouche, S., Tian, H.S., Chuvin, N., Parachoniak, C.A., Nagle, J.M., Perera, R.M., Lapouge, M., Deshpande, V., et al. (2015). YAP inhibition restores hepatocyte differentiation in advanced HCC, leading to tumor regression. Cell Rep. *10*, 1692–1707.

Gebhardt, R., and Matz-Soja, M. (2014). Liver zonation: novel aspects of its regulation and its impact on homeostasis. World J. Gastroenterol. *20*, 8491–8504.

Go, G.W., Srivastava, R., Hernandez-Ono, A., Gang, G., Smith, S.B., Booth, C.J., Ginsberg, H.N., and Mani, A. (2014). The combined hyperlipidemia caused by impaired Wnt-LRP6 signaling is reversed by Wnt3a rescue. Cell Metab. *19*, 209–220.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

Gu, Z., Eils, R., Schlesner, M., and Ishaque, N. (2018). EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics *19*, 234.

Hao, H.X., Xie, Y., Zhang, Y., Charlat, O., Oster, E., Avello, M., Lei, H., Mickanin, C., Liu, D., Ruffner, H., et al. (2012). ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. Nature *485*, 195–200.

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Identification of c-MYC as a target of the APC pathway. Science *281*, 1509–1512.

He, L., Pu, W., Liu, X., Zhang, Z., Han, M., Li, Y., Huang, X., Han, X., Li, Y., Liu, K., et al. (2021). Proliferation tracing reveals regional hepatocyte generation in liver homeostasis and repair. Science *371*, eabc4346.

Hounkpe, B.W., Chenou, F., de Lima, F., and De Paula, E.V. (2021). HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Res. *49* (D1), D947–D955.

Hsu, H.C., Jeng, Y.M., Mao, T.L., Chu, J.S., Lai, P.L., and Peng, S.Y. (2000). Beta-catenin mutations are associated with a subset of low-stage hepatocellular carcinoma negative for hepatitis B virus and with favorable prognosis. Am. J. Pathol. *157*, 763–770.

Jadhav, U., Saxena, M., O'Neill, N.K., Saadatpour, A., Yuan, G.C., Herbert, Z., Murata, K., and Shivdasani, R.A. (2017). Dynamic reorganization of chromatin accessibility signatures during dedifferentiation of secretory precursors into Lgr5+ intestinal stem cells. Cell Stem Cell *21*, 65–77.e5.

Kang, J.S., Wanibuchi, H., Morimura, K., Gonzalez, F.J., and Fukushima, S. (2007). Role of CYP2E1 in diethylnitrosamine-induced hepatocarcinogenesis in vivo. Cancer Res. *67*, 11141–11146.

Koo, B.K., Spit, M., Jordens, I., Low, T.Y., Stange, D.E., van de Wetering, M., van Es, J.H., Mohammed, S., Heck, A.J., Maurice, M.M., and Clevers, H. (2012). Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. Nature *488*, 665–669.

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. bioRxiv, 060012.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

Li, W., Yang, L., He, Q., Hu, C., Zhu, L., Ma, X., Ma, X., Bao, S., Li, L., Chen, Y., et al. (2019). A homeostatic Arid1a-dependent permissive chromatin state

197

licenses hepatocyte responsiveness to liver-injury-associated YAP signaling. Cell Stem Cell 25, 54–68.e5.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 1, 417–425.

Lugert, S., Vogt, M., Tchorz, J.S., Müller, M., Giachino, C., and Taylor, V. (2012). Homeostatic neurogenesis in the adult hippocampus does not involve amplification of Ascl1(high) intermediate progenitors. Nat. Commun. 3, 670.

Ma, R., Martínez-Ramírez, A.S., Borders, T.L., Gao, F., and Sosa-Pineda, B. (2020). Metabolic and non-metabolic liver zonation is established non-synchronously and requires sinusoidal Wnts. eLife 9, e46206.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell 151, 671–683.

Mastrogiovanni, G., Pacini, C., Kakava, S., Arnes-Benito, R., Bradshaw, C.R., Davies, S., Saeb-Parsy, K., Koo, B.-K., and Huch, M. (2020). Loss of RNF43/ZNRF3 predisposes to Hepatocellular carcinoma by impairing liver regeneration and altering liver fat metabolism. bioRxiv. https://doi.org/10.1101/2020.09.25.313205.

Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 28, 1747–1756.

Michalopoulos, G.K., and DeFrances, M.C. (1997). Liver regeneration. Science 276, 60–66.

Mitchell, C., and Willenbring, H. (2008). A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice. Nat. Protoc. 3, 1167–1170.

Miyajima, A., Tanaka, M., and Itoh, T. (2014). Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. Cell Stem Cell 14, 561–574.

Perugorria, M.J., Olaizola, P., Labiano, I., Esparza-Baquer, A., Marzioni, M., Marin, J.J.G., Bujanda, L., and Banales, J.M. (2019). Wnt-β-catenin signalling in liver development, health and disease. Nat. Rev. Gastroenterol. Hepatol. 16, 121–136.

Planas-Paz, L., Orsini, V., Boulter, L., Calabrese, D., Pikiolek, M., Nigsch, F., Xie, Y., Roma, G., Donovan, A., Marti, P., et al. (2016). The RSPO-LGR4/5-ZNRF3/RNF43 module controls liver zonation and size. Nat. Cell Biol. 18, 467–479.

Planas-Paz, L., Sun, T., Pikiolek, M., Cochran, N.R., Bergling, S., Orsini, V., Yang, Z., Sigoillot, F., Jetzer, J., Syed, M., et al. (2019). YAP, but not RSPO-LGR4/5, signaling in biliary epithelial cells promotes a ductular reaction in response to liver injury. Cell Stem Cell 25, 39–53.e10.

Preziosi, M., Okabe, H., Poddar, M., Singh, S., and Monga, S.P. (2018). Endothelial Wnts regulate β-catenin signaling in murine liver zonation and regeneration: A sequel to the Wnt-Wnt situation. Hepatol. Commun. 2, 845–860.

Ritchie, G.R.S., and Flicek, P. (2015). Functional annotation of rare genetic variants. In Assessing Rare Variation in Complex Traits: Design and Analysis of Genetic Studies, E. Zeggini and A. Morris, eds. (New York: Springer), pp. 57–70.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rocha, A.S., Vidal, V., Mertz, M., Kendall, T.J., Charlet, A., Okamoto, H., and Schedl, A. (2015). The angiocrine factor Rspondin3 is a key determinant of liver zonation. Cell Rep. 13, 1757–1764.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393.

Ruffner, H., Sprunger, J., Charlat, O., Leighton-Davies, J., Grosshans, B., Salathe, A., Zietzling, S., Beck, V., Therier, M., Isken, A., et al. (2012). R-Spondin potentiates Wnt/β-catenin signaling through orphan receptors LGR4 and LGR5. PLoS ONE 7, e40976.

Russell, J.O., and Monga, S.P. (2018). Wnt/β-Catenin signaling in liver development, homeostasis, and pathobiology. Annu. Rev. Pathol. 13, 351–378.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 459, 262–265.

Schuierer, S., and Roma, G. (2016). The exon quantification pipeline (EQP): a comprehensive approach to the quantification of gene, exon and junction expression from RNA-seq data. Nucleic Acids Res. 44, e132.

Sekine, S., Lan, B.Y., Bedolli, M., Feng, S., and Hebrok, M. (2006). Liver-specific loss of beta-catenin blocks glutamine synthesis pathway activity and cytochrome p450 expression in mice. Hepatology 43, 817–825.

Serra, D., Mayr, U., Boni, A., Lukonin, I., Rempfler, M., Challet Meylan, L., Stadler, M.B., Strnad, P., Papasaikas, P., Vischi, D., et al. (2019). Self-organization and symmetry breaking in intestinal organoid development. Nature 569, 66–72.

Stanger, B.Z. (2015). Cellular homeostasis and repair in the mammalian liver. Annu. Rev. Physiol. 77, 179–200.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A 102, 15545–15550.

Sun, T., Pikiolek, M., Orsini, V., Bergling, S., Holwerda, S., Morelli, L., Hoppe, P.S., Planas-Paz, L., Yang, Y., Ruffner, H., et al. (2020). AXIN2+ pericentral hepatocytes have limited contributions to liver homeostasis and regeneration. Cell Stem Cell 26, 97–107.e6.

Tan, X., Behari, J., Cieply, B., Michalopoulos, G.K., and Monga, S.P. (2006). Conditional deletion of beta-catenin reveals its role in liver growth and regeneration. Gastroenterology 131, 1561–1572.

Torre, C., Perret, C., and Colnot, S. (2011). Transcription dynamics in a physiological process: β-catenin signaling directs liver metabolic zonation. Int. J. Biochem. Cell Biol. 43, 271–278.

Walesky, C.M., Kolb, K.E., Winston, C.L., Henderson, J., Kruft, B., Fleming, I., Ko, S., Monga, S.P., Mueller, F., Apte, U., et al. (2020). Functional compensation precedes recovery of tissue mass following acute liver injury. Nat. Commun. 11, 5785.

Wang, B., Zhao, L., Fish, M., Logan, C.Y., and Nusse, R. (2015). Self-renewing diploid Axin2(+) cells fuel homeostatic renewal of the liver. Nature 524, 180–185.

Wei, Y., Wang, Y.G., Jia, Y., Li, L., Yoon, J., Zhang, S., Wang, Z., Zhang, Y., Zhu, M., Sharma, T., et al. (2021). Liver homeostasis is maintained by midlobular zone 2 hepatocytes. Science 371, eabb1625.

Yan, B.C., Gong, C., Song, J., Krausz, T., Tretiakova, M., Hyjek, E., Al-Ahmadie, H., Alves, V., Xiao, S.Y., Anders, R.A., and Hart, J.A. (2010). Arginase-1: a new immunohistochemical marker of hepatocytes and hepatocellular neoplasms. Am. J. Surg. Pathol. 34, 1147–1154.

Yang, J., Mowry, L.E., Nejak-Bowen, K.N., Okabe, H., Diegel, C.R., Lang, R.A., Williams, B.O., and Monga, S.P. (2014). β-catenin signaling in murine liver zonation and regeneration: a Wnt-Wnt situation! Hepatology 60, 964–976.

Yimlamai, D., Christodoulou, C., Galli, G.G., Yanger, K., Pepe-Mooney, B., Gurung, B., Shrestha, K., Cahan, P., Stanger, B.Z., and Camargo, F.D. (2014). Hippo pathway activity influences liver cell fate. Cell 157, 1324–1338.

Yoshida, M., Horinouchi, S., and Beppu, T. (1995). Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. BioEssays 17, 423–430.

Zhao, L., Jin, Y., Donahue, K., Tsui, M., Fish, M., Logan, C.Y., Wang, B., and Nusse, R. (2019). Tissue repair in the mouse liver following acute carbon tetrachloride depends on injury-induced Wnt/β-catenin signaling. Hepatology 69, 2623–2635.

198

Cell Stem Cell
Article

🧬 CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-Ki67	Invitrogen	Cat#14-5698-82; clone SolA15
Rabbit monoclonal anti-Ki67	Thermo Fisher Scientific	Cat#RM9106
Rabbit polyclonal anti-SOX9	Milipore	Cat#AB5535
Rabbit polyclonal anti-ARG1	Sigma-Aldrich	Cat#HPA003595
Rabbit polyclonal anti-LYVE-1	Angiobio	Cat#151202
Rabbit monoclonal anti-Iba1	Abcam	Cat# ab178846; clone EPR16588 RRID:AB_2636859
Mouse monoclonal anti-E-Cadherin	BD Biosciences	Cat#610181; clone 36 RRID:AB_397580
Rabbit polyclonal anti-CYP2E1	Sigma-Aldrich	Cat#HPA009128 RRID:AB_1078613
Rat monoclonal anti CK19	Developmental Studies Hybridoma Bank	Cat#TROMA-III; RRID:AB_2133570
Rabbit polyclonal anti-Beta Galactosidase	Fitzgerald	Cat#20C-CR7001RP; RRID:AB_1282752
Goat polyclonal anti-GFP	Abcam	Cat#AB6673; RRID:AB_305643
Rabbit polyclonal anti-GFP	Invitrogen	Cat#A11122 RRID:AB_221569
Rabbit polyclonal anti-Glutamine Synthetase	Abcam	Cat#AB49873; RRID:AB_880241
Rabbit monoclonal anti CCND1	Abcam	Cat#AB16663
Chicken polyclonal anti-Albumin	Sigma-Aldrich	Cat#SAB3500217; RRID:AB_10640041
OmniMap anti-Rb HRP	Roche Diagnostics	Cat#760-4311
OmniMap anti-Rt HRP	Roche Diagnostics	Cat#760-4457
Bacterial and virus strains		
Ad5-CMV-GFP virus	Vector Development Laboratory	N/A
Ad5-CMV-Cre virus	Vector Development Laboratory	N/A
Ecoli stock: Plv[exp]-Puro-CMV > mCherry	VectorBuilder	VB201117-1148vrb
Ecoli stock: Plv[exp]-Puro-CMV > EGFP	VectorBuilder	VB900088-2219pdm
Ecoli stock: Plv[exp]-Puro-CMV > mCyp2e1 [ORF032106](ns):T2A:mCherry	VectorBuilder	VB201117-1125nwh
Ecoli stock: Plv[exp]-Puro-CMV > mGlul [ORF031394](ns):T2A:EGFP	VectorBuilder	VB201117-112tew
Chemicals, peptides, and recombinant proteins		
Tamoxifen	Sigma-Aldrich	Cat#T5648; CAS: 10540-29-1
Trichostatin A	Sigma-Aldrich	Cat#T1952
Hematoxylin II	Roche Diagnostics	Cat# 790-2208
Bluing Reagent	Roche Diagnostics	Cat#760-2037
RSPO1	In house	N/A
Doxycycline hyclate	Sigma-Aldrich	Cat#D-9891-10G
Allyl alcohol	Sigma-Aldrich	Cat#240532
Metacam	Boehringer Ingelheim	Cat#10001293
3,5-Dicarbethoxy-1,4-dihydrocollidine (DDC)	Sigma-Aldrich	Cat#137030; CAS: 632-93-9
Temgesic	Indivior	Cat#7680419310018
PharmLyse Red Blood Cell lysis buffer	BD Biosciences	Cat#555899
Titanium Taq DNA Polymerase	Clontech	Cat#639242
Mayer's Haematoxylin	Dako	Cat#S3309
DAPI	Sigma-Aldrich	Cat#D9542

(Continued on next page)

199

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
N2	Thermo Fisher Scientific	Cat#17502048
B27	Thermo Fisher Scientific	Cat#17504044
mEGF	R&D Systems	Cat#2028-EG-200
GlutaMAX Supplement	Thermo Fisher Scientific	Cat#A1286001
DMSO	Sigma-Aldrich	Cat#D8418
Murine NOGGIN	PeproTech	Cat#250-38-100
ADVANCED DMEM/F12 HEPES	STEMCELL Technologies	Cat#36254
IntestiCult Organoid Growth Medium	STEMCELL Technologies	Cat#06005
ROCK inhibitor Y-27632	STEMCELL Technologies	Cat#72302
Wnt3a-CM	In house	N/A
TrypLE	Thermo Fisher Scientific	Cat#12604013
N-acetylcysteine	Sigma-Aldrich	Cat#A0737
CHIR99021	STEMCELL Technologies	Cat#72054
Matrigel	Coring	Cat#356231
PFA	VWR INTERNATIONAL AG	Cat#ELMI15714-S
EGTA	Sigma-Aldrich	Cat#E0396; CAS: 67-42-5
HEPES	Sigma-Aldrich	Cat#15630056; CAS: 7365-45-9
Penicillin/Streptomycin	Thermo Fisher Scientific	Cat#15140122
$CaCl_2$	Sigma-Aldrich	Cat#21115; CAS: 10043-52-4
Phosphate-Buffered Saline	Thermo Fisher Scientific	Cat#10010
Collagenase A	Roche	Cat#11088793001
William's E Medium	Thermo Fisher Scientific	Cat#22551089
Normal Donkey Serum	Sigma-Aldrich	Cat#D9663
Mini Protease Inhibitor Cocktail	Roche	Cat#11836153001
Fetal Bovine Serum	Thermo Fisher Scientific	Cat#16140071
Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12	Thermo Fisher Scientific	Cat#11039047
Hank's Balanced Salt Solution	Thermo Fisher Scientific	Cat#14175129
Opal fluorophore 570	Akoya Bioscience	Cat#FP1488001KT
Opal fluorophore 690	Akoya Bioscience	Cat#FP1489001KT
Critical commercial assays		
RNeasy plus mini kit	QIAGEN	Cat#74134
RNeasy micro kit	QIAGEN	Cat#74004
MinElute Reaction Cleanup Kit	QIAGEN	Cat#28206
Total RNA Purification Kit	NORGEN	Cat#37500
ChromoMap DAB Kit	Roche Diagnostics	Cat#760-159
Purple Kit	Roche Diagnostics	Cat#760-229
RNase-Free DNase I Kit	NORGEN	Cat#25710
High Capacity cDNA Reverse Transcription kit	Thermo Fisher Scientific	Cat#4368814
TaqMan Universal PCR Master Mix	Thermo Fisher Scientific	Cat#4364338
Oil Red O staining kit	Abcam	Cat#ab150678
Picro-Sirius Red Stain Kit	Abcam	Cat#ab150681
Lenti-X concentrator	Takara	Cat#631232
ViewRNA ISH Tissue Assay Core Kit	Thermo Fisher Scientific	Cat#19931
ViewRNA ISH Tissue Assay Blue Kit	Thermo Fisher Scientific	Cat#19932
Zenon Alexa Fluor 488 Rabbit IgG labeling kit	Thermo Fisher Scientific	Cat#Z25302; RRID:AB_2572214
Visium Spatial Gene Expression Reagent Kits	10x Genomics	Cat#PN-1000186
Visum Spatial Gene ExpressionSlide Kit	10x Genomics	Cat#PN-1000185

Cell Stem Cell
Article

CellPress

Continued		
REAGENT or RESOURCE	**SOURCE**	**IDENTIFIER**
Visum Spatial Gene ExpressionSlide	10x Genomics	Cat#PN-2000233
Dual Index Kit TT Set A, 96 rxns	10x Genomics	Cat#PN-1000215
Agilent RNA 6000 Nano Kit	Agilent Technologies	Cat#5067-1511
Agencourt AMPure XP PCR purification Kit	Beckman Coulter	Cat#A63880
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat#Q32854
Illumina Nextera DNA Unique Dual Indexes	Illumina	Cat#20027213
Illumina Tagment DNA TDE1 Enzyme and Buffer Kit	Illumina	Cat#20034197
Zymo DNA Clean and Concentrator Kit	Zymo Research	Cat#D4014
NEBNex High Fidelity 2x PCR Mix	New England Biolabs	Cat#M0541S
RNAscope Multiplex Fluorescent Reagent v2 Assay Kit	ACD	Cat#323100
Deposited data		
ATAC-seq data	This paper	PRJNA716814
Spatial transcriptomics data	This paper	PRJNA705085
RNA-seq data	This paper	PRJNA716805
Experimental models: cell lines		
HEK293T	ATCC	CRL-11268
Experimental models: organisms/strains		
Mouse: Axin2CreERT2: B6-Tg(mAxin2-mCherry-creRT2)Npa	Planas-Paz et al., 2019	N/A
Mouse: Axin2-LacZ: A16C57BL/6J-Tg(Axin2-LacZ)NPa	Planas-Paz et al., 2016	N/A
Mouse: Axin2-LacZ, Znrf3ECD-TM: C57BL/6J-Tg(Axin2-LacZ)Npa x B4B6-Tg(CAG-ZNRF3-ECDTmiTG)475Npa	Planas-Paz et al., 2016	N/A
Mouse: Axin2CreERT2-R26-EGFP: B6.C-Rosa26tm(CAG-StopEGFPattNeo)Npa x B6-Tg(mAxin2-mCherry-creRT2)Npa	This paper	N/A
Mouse: Znrf3/Rnf43 flox: B6.Cg-Znrf3 < tm1.2 < 5A7 > Npa > ; Rnf43 < tm1.2 < 6-7g > Npa >	Planas-Paz et al., 2016	N/A
Mouse: Znrf3/Rnf43dKO, Axin2CreERT2: B6.Cg-Znrf3 < tm1.2 < 5A7 > Npa > ; Rnf43 < tm1.2 < 6-7g > Npa x B6-Tg(mAxin2-mCherry-creRT2)Npa	This paper	N/A
Mouse: Znrf3/Rnf43dKO, Axin2CreERT2, R26-EGFP: B6.Cg-Znrf3 < tm1.2 < 5A7 > Npa > ; Rnf43 < tm1.2 < 6-7g > Npa x B6-Tg(mAxin2-mCherry-creRT2)Npa; B6.C-Rosa26tm(CAG-StopEGFPattNeo)Npa	This paper	N/A
Mouse: Znrf3: B6.Cg-Znrf3 < tm1.2(5A7)Npa >	This paper	N/A
Mouse: Rnf43: B6-Rnf43 < tm1.2(6-7g)Npa	This paper	N/A
Mouse: LSL-EGFP: B6.C-Rosa26tm(CAG-StopEGFPattNeo)Npa	Lugert et al., 2012	N/A
Mouse: Znrf3KO, Axin2CreERT2: B6.Cg-Znrf3 < tm1.2(5A7)Npa;B6-Tg(mAxin2-mCherry-CreERT2)Npa	This paper	N/A
Mouse: Znrf3KO, Axin2CreERT2, R26-EGFP: B6.Cg-Znrf3 < tm1.2(5A7)Npa;B6-Tg(mAxin2-mCherry-CreERT2)Npa; B6.C-Rosa26tm(CAG-StopEGFPattNeo)Npa	This paper	N/A

201

CellPress

Cell Stem Cell
Article

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: Rnf43KO, Axin2CreERT2: B6-Rnf43 < tm1.2(6-7g)Npa;B6-Tg(mAxin2-mCherry-CreERT2)Npa	This paper	N/A
Mouse: Rnf43KO, Axin2CreERT2, R26-EGFP: B6-Rnf43 < tm1.2(6-7g)Npa;B6-Tg(mAxin2-mCherry-CreERT2)Npa; B6.C-Rosa26tm(CAG-StopEGFPattNeo)Npa	This paper	N/A
Mouse: Lgr4/5d flox: B6-Lgr4tm2Npa;B6-Lgr5tm1Npa;	Planas-Paz et al., 2016	N/A
Mouse: Lgr4/5dKO, AlbCre: B6-Lgr4tm2Npa;B6-Lgr5tm1Npa; B6.Cg-Tg(Alb-cre)21Mgn/J	This paper	N/A
Oligonucleotides		
In situ probes for mouse *Axin2* mRNA	Affymetrix	Cat#VX-01
In situ probes for mouse *Znrf3* mRNA	Affymetrix	Cat#VX-01
In situ probes for mouse *Rnf43* mRNA	Affymetrix	Cat#VX-01
Taqman primers for mouse *Cyp2e1* (Mm00491127_m1)	Thermo Fisher Scientific	Cat#4331182
Taqman primers for mouse *Glul* (Mm00725701_s1)	Thermo Fisher Scientific	Cat#4331182
In situ probes for mouse *Axin2* mRNA	ACD	Cat#400331-C3
In situ probes for mouse *Rspo3* mRNA	ACD	Cat#402011-C2
In situ probes for mouse *Wnt7a* mRNA	ACD	Cat#401121
In situ probes for mouse *Wnt7b* mRNA	ACD	Cat#401131-C2
In situ probes for mouse *Wnt9b* mRNA	ACD	Cat#405091
In situ probes for mouse *Wnt10a* mRNA	ACD	Cat#401061
Recombinant DNA		
Plasmid: pLP/VSVG	Thermo Fisher Scientific	A43237
Plasmid: pLP2	Thermo Fisher Scientific	A43237
Plasmid: pLP1	Thermo Fisher Scientific	A43237
Software and algorithms		
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml; RRID:SCR_016368
Aperio ImageScope software v12.1.0.5029	Leica Biosystems	https://www.leicabiosystems.com/digital-pathology/manage/aperio-imagescope/; RRID:SCR_014311
Color Deconvolution algorithm, Aperio ImageScope	Leica Biosystems	https://www.leicabiosystems.com/digital-pathology/analyze/ihc/aperio-color-deconvolution-algorithm/
Nuclear algorithm, Aperio ImageScope	Leica Biosystems	https://www.leicabiosystems.com/digital-pathology/analyze/ihc/aperio-nuclear-algorithm/
Zeiss AxioVision software Vs40 v4.8.2.0	Zeiss	https://www.micro-shop.zeiss.com/en/us/softwarefinder/software-categories/zen-blue/; RRID:SCR_002677
GraphPad Prism software v7.03	GraphPad	https://www.graphpad.com/scientific-software/prism/; RRID:SCR_002798
ImageJ software v1.51w	ImageJ	https://imagej.net/Welcome; RRID:SCR_003070
Olympus FluoView software FV31S-SW v2.1.2.2	Olympus	https://www.olympus-lifescience.com/en/laser-scanning/fv3000/

202

Cell Stem Cell
Article

CellPress

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R	The R Foundation R-version: 4.0.2	https://www.r-project.org/
R-studio server pro	R-studio version 1.0.153 RStudio Server Pro Version 1.3.1093-1	https://www.rstudio.com https://www.rstudio.com
Edge R Bioconductor package	Robinson et al., 2010 edgeR v3.30.3	http://bioconductor.org/packages/release/bioc/html/edgeR.html; RRID_012802
Limma-voom	Ritchie et al.l, 2015 Limma v3.44.3	https://bioconductor.org/packages/release/bioc/html/limma.html
Gene mapping	Schuierer and Roma, 2016	N/A
Spotfire	Tibco Spotfire Analyst 7.11.	https://spotfire.tibco.com
snakePipes	Bhardwaj et al., 2019	https://github.com/maxplanck-ie/snakepipes
DiffBind	Ross-Innes et al., 2012 Version: 3.0.3	https://bioconductor.org/packages/release/bioc/html/DiffBind.html
EnrichedHeatmap	Gu et al., 2018 Version: 1.20.0	https://bioconductor.org/packages/release/bioc/html/EnrichedHeatmap.html
cutadapt	Martin, 2011 Version 1.16	https://cutadapt.readthedocs.io/en/stable/
deeptools	Ramírez et al., 2016 Version 3.1.2	https://deeptools.readthedocs.io/en/develop/
multiqc	Ewels et al., 2016 Version 3.1.2	https://multiqc.info/
SAMtools	Version 1.9	http://www.htslib.org/
MACS2	Zhang et al., 2008 Version 2.2.7.1	https://pypi.org/project/MACS2/
Seurat R-package	Seurat v3.2.3	http://satijalab.org/seurat/
STTKIT	STTKIT v0.0.9	https://github.com/lima1/sttkit
fgsea Bioconductor R-package	fgsea v1.14.0	https://bioconductor.org/packages/release/bioc/vignettes/fgsea/inst/doc/fgsea-tutorial.html
ComplexHeatmap Bioconductor R-package	ComplexHeatmap v2.4.3	https://jokergoo.github.io/ComplexHeatmap-reference/book/
biomaRt Bioconductor R-package	biomaRt v2.44.4	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
org.Mm.eg.db gene annotation Bioconductor R-package	org.Mm.eg.db_3.11.4	https://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html
msigdbr	msigdbr v7.2.1	https://cran.r-project.org/web/packages/msigdbr/index.html
ggplot2 Bioconductor R-package	ggplot2 v3.3.2	https://cran.r-project.org/web/packages/ggplot2/index.html
Spaceranger 10X Genomics software	Spaceranger v1.0.0	https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/installation

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jan Tchorz (jan.tchorz@novartis.com)

Materials availability
Transgenic mouse lines used in this study (or the basic lines enabling reproduction of these lines) are available from the lead contact, Jan Tchorz (jan.tchorz@novartis.com), with a completed Materials Transfer Agreement.

Data and code availability
ATACseq raw data for hepatocytes from LGR4/5^FLOX and LGR4/5dLKO livers, AXIN2+ and AXIN2- hepatocytes, and AXIN2+ ISCs are archived at the NIH Sequence Read Archive under accession number SRA: SRP312196 [PRJNA716814]. Bulk RNA-seq and

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse models

LGR4/5dLKO mice were generated by crossing AlbCre with LGR4/5^{FLOX} mice (Planas-Paz et al., 2016). LGR4/5$^{\Delta AXIN2+}$ mice were generated by crossing Axin2CreERT2 mice (Sun et al., 2020) with LGR4/5^{FLOX} mice followed by TAM injection. RNF43$^{\Delta AXIN2+}$, ZNRF3$^{\Delta AXIN2+}$ and ZNRF3/RNF43$^{\Delta AXIN2+}$ mice were generated by crossing Axin2CreERT2 mice with RNF43^{FLOX}, ZNRF3^{FLOX} and ZNRF3/RNF43^{FLOX} mice (Planas-Paz et al., 2016), respectively, followed by TAM injection. RNF43$^{EGFP\Delta AXIN2+}$, ZNRF3$^{EGFP\Delta AXIN2+}$ and ZNRF3/RNF43$^{EGFP\Delta AXIN2+}$ mice were generated by crossing R26-STOP-EGFP (Lugert et al., 2012) with RNF43$^{\Delta AXIN2+}$, ZNRF3$^{\Delta AXIN2+}$ and ZNRF3/RNF43$^{\Delta AXIN2+}$ mice, respectively. ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice were generated by crossing R26-STOP-EGFP with ZNRF3/RNF43^{FLOX} mice (yielding ZNRF3/RNF43$^{FLOX;EGFP}$ mice) followed by injection of Ad5Cre. Generation of Axin2CreERT2^{EGFP} (Sun et al., 2020), AXIN2-LacZ, AXIN2-LacZ/ZNRF3ECD-TM, RNF43^{FLOX} and ZNRF3^{FLOX} (Planas-Paz et al., 2016) mice has been described.

In vivo mouse studies

For ATACseq, hepatocytes from LGR4/5^{FLOX} (control) and LGR4/5dLKO livers, as well as AXIN2+/AXIN2-hepatocytes and AXIN2+ ISCs from AXIN2CreERT2^{EGFP} mice were isolated. To isolate AXIN2+ ISCs, AXIN2CreERT2^{EGFP} mice were injected i.p. with 2 mg of Tamoxifen (Sigma-Aldrich, T-5648) on two consecutive days and euthanized 50 hours after the first injection. Single cells were prepared from the intestine followed by EGFP sorting. To isolate AXIN2+/− hepatocytes during homeostasis, AXIN2CreERT2^{EGFP} mice were injected i.p. with 2 mg of Tamoxifen for 5 consecutive days and euthanized after 6-7 weeks and purity was checked as described (Sun et al., 2020). To delete LGR4/5 in AXIN2+ cells, LGR4/5$^{\Delta AXIN2+}$ mice were injected i.p. with 2 mg of Tamoxifen for 5 consecutive days and euthanized 14 days after the first injection. To delete ZNRF3/RNF43, ZNRF3 or RNF43 in AXIN2+ cells, ZNRF3/RNF43$^{\Delta AXIN2+}$, ZNRF3$^{\Delta AXIN2+}$ and RNF43$^{\Delta AXIN2+}$ mice were injected i.p. with one dose of 2mg (high dose) Tamoxifen and euthanized 4 weeks later. To perform lineage tracing of hepatocytes with a deletion of ZNRF3/RNF43, ZNRF3 or RNF43 in AXIN2+ cells, ZNRF3/RNF43$^{EGFP\Delta AXIN2+}$, ZNRF3$^{EGFP\Delta AXIN2+}$ and RNF43$^{EGFP\Delta AXIN2+}$ mice were injected i.p. with one dose of 0.7mg (low dose for males) or 0.25 mg (low dose for females) Tamoxifen and euthanized 4 weeks or 8 weeks later. These different doses were chosen to allow for comparable labeling efficiency that was different in males and females. For Ad5-CMV-Cre (Ad5Cre)-mediated ZNRF3 and/or RNF43 deletion, ZNRF3/RNF43^{FLOX}, ZNRF3^{FLOX} or RNF43^{FLOX} mice were injected i.v. with a single dose of 10E9 plaque-forming units (pfu) (high dose) Ad5Cre and euthanized after 4 weeks or 1 year. ZNRF3/RNF43^{FLOX} mice that bear the same dose and time course of Ad5-CMV-EGFP (Ad5EGFP) were used as relative control groups. To check transduction and Cre recombination, R26-STOP-EGFP mice were injected i.v. with a single dose of 10E9 pfu Ad5Cre and euthanized 4 weeks later. For tracing the hepatocytes with Ad5Cre-mediated ZNRF3 and RNF43 deletion, ZNRF3/RNF43$^{FLOX;EGFP}$ mice were injected i.v. with a single dose of 10E7 pfu (low dose) Ad5Cre (yielding ZNRF3/RNF43$^{EGFP\Delta Hep}$ mice) and euthanized after 4 weeks or 8 weeks. R26-STOP-EGFP mice that bear the same dose and time course of Ad5Cre were used as relative control groups. To trace the AXIN2+ hepatocytes in the DEN HCC model, AXIN2CreERT2^{EGFP} mice were injected with TAM (0.7mg) and DEN (10mg/kg) at p15 and euthanized after 5 days (to assess labeling) or 1 year (to assess tumors). To study RSPO1-induced WNT/β-Catenin activation, C57BL/6 male mice were injected i.v. with either 10 mg/kg or 50 mg/kg RSPO1 or PBS (control) according to the regimen shown in Figure 3A. For doxycycline (DOX)-induced ZNRF3ECD-TM expression (Planas-Paz et al., 2016), Axin2-LacZ or Axin2-LacZ/ZNRF3ECD-TM mice were administered with 2 mg/ml DOX in drinking water for 10 days. Two-thirds partial hepatectomy (PHx) was essentially performed as described (Mitchell and Willenbring, 2008). Briefly, isoflurane/O$_2$ was used for anesthesia and 0.1 mg/kg s.c. Temgesic (Buprenorphine; Indivior) for analgesia. The left and median liver lobes were ligated with a silk filament at the base before resection. The peritoneum was closed with continuous sutures (Silkam 6/0), whereas the skin was closed using simple interrupted suture (Dafilon, 6/0). Normothermia was guaranteed using a heating plate (32°C) until recovery from anesthesia. For postoperative analgesia Temgesic and Meloxicam 5 mg/kg (Metacam) were used for the next day. To trace the hepatocytes with a deletion of ZNRF3 and RNF43 in AXIN2+ cells after PHx, ZNRF3/RNF43$^{EGFP\Delta AXIN2+}$ mice were injected i.p. with one dose of 0.7mg (low dose for males) or 0.25mg (low dose for females) Tamoxifen, subjected to PHx 1 week later and euthanized 2 weeks after PHx. For tracing the hepatocytes with Ad5Cre-mediated deletion of ZNRF3 and RNF43 after PHx, ZNRF3/RNF43$^{FLOX;EGFP}$ mice were injected i.v. with one dose of 10E7 pfu Ad5-Cre (yielding ZNRF3/RNF43$^{EGFP\Delta Hep/}$ mice), subjected to PHx 2 weeks later and euthanized 2 weeks after PHx. AXIN2-LacZ mice were dissected 40 hours after the PHx. Periportal hepatocyte injury in AXIN2-LacZ mice was induced by Allyl alcohol (AA) (Sigma-Aldrich 240532) injection i.p. at a concentration of 0.03 μl/g followed by a second dose of 0.06 μl/g (high dose) 3 days later. Three days after the second AA injection, mice were analyzed. Liver damage and resulting ductular reaction was induced using diet supplemented with 0.1% 3,5-Dicarbethoxy-1,4-dihydrocollidine (DDC; Sigma-Aldrich, 137030) for 16 days and analyzed immediately after end of the diet. Interruption of the diet for 48 h allowed recovering body weight. All mice within an experiment had similar DDC diet regimen. Age-matched adult males and females were assigned in similar numbers to each experimental group. Animals were monitored daily for wellbeing. All mice had unrestricted access to water and food and animal experimentation was conducted in accordance with animal law of Basel-Stadt, Switzerland. The Novartis Campus animal facilities

204

Cell Stem Cell
Article

CellPress

comprise an SPF animal breeding facility and a clean facility for experimental surgery and physiology. Biosecurity and pathogen exclusions follow the Federation of European Laboratory Animal Science Associations (FELASA) health monitoring guidelines, and animals are screened quarterly. Mice were housed in Allentown XJ individually ventilated cages in a 12:12 light:dark cycle. Environmental enrichments included nestlets, wood sticks and mouse houses. The maximum housing density in the breeding and experimental areas was 4 and 5 mice, respectively. Most experiments were performed with grouped mice, single housing was used for aggressive or wounded animals.

METHOD DETAILS

Immunohistochemistry (IHC) and in situ hybridization (ISH)

Tissue samples were freshly frozen in O.C.T. embedding medium (Tissue TEK) or fixed for 48h in formalin and embedded in paraffin. IHC was performed using either a Roche Discovery XT or Roche Discovery ULTRA. Primary antibodies used in this study were rabbit anti-Ki67 (Thermo Scientific, RM-9106), rat anti-Ki67 (Invitrogen, 14-5698-82), rabbit anti-CCND1 (Abcam, 16663), rabbit anti-SOX9 (Milipore, Cat#AB5535), rabbit anti-ARG1 (Sigma-Aldrich, Cat#HPA003595), rat anti-CK19 (DSHB, TROMA-III), rabbit anti-Beta Galactosidase (Fitzgerald, 20C-CR7001RP), goat anti-GFP (Abcam, AB6673), rabbit anti-GFP (Invitrogen, A11122), rabbit anti-Glutamine Synthetase (Abcam, AB49873), rabbit anti-CYP2E1 (Sigma-Aldrich, HPA009128), chicken anti-Albumin (Sigma-Aldrich, SAB3500217), rabbit anti-Iba1 (Abcam, ab178846), mouse anti-E-Cadherin (BD Biosciences, 610181), rabbit anti-LYVE1 (Angiobio, Cat#151202). The TROMA-III, developed by R. Kemler, was obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. For immunofluorescence using two antibodies derived from the same species in the same protocol, Zenon Alexa Fluor 488 Rabbit IgG labeling kit (Thermo Fisher Scientific, Z25302) was used according to the manufacturer's instructions. Alexa DyLight- or Cy-coupled (Jackson ImmunoResearch) secondary antibodies were used for primary antibody detection, and either Mayer's Haematoxylin (Dako, S3309) or 4'-6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich, D9542) were used for counterstaining. For non-fluorescent stainings, the following reagents from Roche Diagnostics, Ventana Discovery product series, were used with either a Roche Discovery XT or Roche Discovery ULTRA: OmniMap anti-Rb HRP (760-4311), OmniMap anti-Rt HRP (760-4457), UltraMap anti-Rb Alk Phos (760-4314), ChromoMap DAB Kit (760-159), Purple Kit (760-229), Hematoxylin II (790-2208) and Bluing Reagent (760-2037). For Oil Red O staining, PFA-fixed slides from fresh frozen samples were processed with the Oil Red O Stain Kit (Abcam, ab150678) according to the manufacturer's instructions. For Sirius Red staining, formalin-fixed paraffin-embedded (FFPE) sections were stained with the Leica Autostainer using a standard protocol. Whole slide images were scanned and stitched by using an Aperio XT Scanscope (Leica) and imaged using ImageScope software.

For representative images, whole-liver lobes were examined histologically in multiple replicates. Automated image analysis and quantification of CYP2E1, Ki67, βGal and GS staining on cross sections of liver lobes was performed using the Color Deconvolution algorithm from the Aperio ImageScope software. Data was presented as percent positive stained area per total analysis area or percent positive cells per total cell number, unless stated otherwise. The length of crypt-villus units and Ki67-positive cells in the mouse intestine were measured and quantified manually using ImageScope. Immunofluorescence costaining for CYP2E1, Ki67, βGal or GS with other markers were used for illustration in the Figures. Analyses and distribution assessment of EGFP, GS or CYP2E1 positive hepatocytes in the clonal analyses, or when assessing different lobular zones, were performed by manual quantification percentage of positive cells. ISH for Axin2 (probes VB1-10588-VT), Znrf3 (probes VB6-VPPRJ3G) and Rnf43 (probes VB1-VPYMJHK) were performed as described (Planas-Paz et al., 2016). Briefly, 10 μm fresh-frozen liver cryosections were fixed with 10% neutral-buffered formalin overnight at 4°C, washed, dehydrated and air-dried followed by protease digestion. Sections were hybridized for 2.5 h with probes at 40°C followed by pre-amplification and amplification according to the manufacturer's instructions. Fast red and fast blue substrates were used to detect the alkaline phosphatase type 1 (VB1) and type 6 (VB6) conjugated probes. ISH for Axin2 (400331-C3), Rspo3 (402011-C2), Wnt7a (401121), Wnt7b (401131-C2), Wnt9b (405091), Wnt10a (401061) using RNAscope Multiplex Fluorescent Reagent v2 Assay Kit (ACD 323100) was performed according to the manufacturer's instructions. Briefly, 5μm sections were cut from FFPE samples and dried at 37°C overnight. The next day, standard protease digestion was performed to the tissue followed by probe hybridization and signal amplification. Probes were then stained for detection with Opal™ fluorophores 570 (Akoya Bioscience, FP1488001KT) and 690 (Akoya Bioscience, FP1497001KT). ISH images were acquired using an Olympus laser scanning confocal microscope FV3000. The percent ISH signal per analyses area was quantified using ImageJ software. Periportal and pericentral hepatocytes were defined as the first 3 layers of hepatocytes around the respective vein, whereas the parenchyma/midzonal area was defined as the hepatocytes in between. For intestine organoid stainings, 20x z stack images were acquired via ImageXpress (Molecular Devices). All maximum intensity projections (MIP) of a specific well and channel were stitched together according to the x/y coordinates provided by the microscope to obtain MIP well overviews. Representative organoid images from Figures S1E, 1G, and 1I were cropped from the MIP well overviews. Organoid areas were measured by ImageJ as shown in Figure S1F. EGFP-positive or mCherry-positive cells at day 4 and day 6 were counted manually.

Intestinal organoid culture and staining

Organoids were generated from murine small intestine as previously described (Sato et al., 2009). In brief, the small intestine was opened longitudinally, cleaned with ice cold PBS and sliced into small fragments. The tissue was then incubated in 2.5 mM EDTA/PBS at 4°C for 30 min on a shaker. Supernatant was removed and intestine pieces were then shaken vigorously in DMEM/F12

205

with 0.1% BSA. The suspension was collected and passed through a 70 μm strainer. The remaining tissue pieces were suspended with DMEM/F12 with 0.1% BSA. The above step was repeated four times and all supernatant fractions were combined and centrifuged at 300 g for 5 min at 4°C. Supernatant was removed and the pellet was re-suspended into Matrigel and plated into 24 well plates. Organoids were kept in IntestiCult Organoid Growth Medium (STEMCELL Technologies, 06005) with 100 μg/ml Penicillin-Streptomycin for amplification and maintenance. Organoids were collected 5 to 7 days after passaging and digested with TrypLE (Thermo Fisher Scientific, 12604013) for 20 min at 37°C. Single alive cells were sorted by FACS as described (Serra et al., 2019) and collected in ENR medium composed of advanced DMEM/F12 with 15 mM HEPES (STEMCELL Technologies, 36254) supplemented with 100 μg/ml Penicillin-Streptomycin, 1x Glutamax (Thermo Fisher Scientific, A1286001), 1x N2 (Thermo Fisher Scientific, 17502048), 1x B27 (Thermo Fisher Scientific, 17504044), N-acetylcysteine 1 mM (Sigma-Aldrich, A0737), mEGF 50 ng/ml (R&D Systems, 2028-EG-200), Noggin 100 ng/ml (Pepro Tech, 250-38-100) and RSPO1 500 ng/ml (in house). 5 μl droplets containing 3000 single cells mixed 1:1 with Matrigel (Corning, 356231) were seeded in 96-well plates. After 20 min of solidification at 37°C, 100 μl of medium was added (day 0). From day 0 to day 1, ENR was supplemented with 20% Wnt3a-conditioned medium (Wnt3a-CM), 10 μM Y-27632 (ROCK inhibitor, STEMCELL Technologies, 72302) and 3 μM of CHIR99021 (GSK3B inhibitor, STEMCELL Technologies, 72054). From day 1 to day 3 ENR was supplemented with 20% Wnt3a-CM and 10 μM Y-27632. From day 3 to day 5, only ENR was added to the cells. Wnt3a-CM was produced in-house with Wnt3a L-cells. On day 2, organoids were treated with 1 μM of TSA (Sigma-Aldrich, T1952) or DMSO. On day 3.5, organoids were collected for RNA extraction. For the organoid growth experiment, organoids were treated with 1 μM of TSA on day 2. The medium was changed with fresh TSA (1 μM) on day 3 and 5. Bright field images were taken with a 20x objective daily from day 3 to day 6 and organoids size was measured using ImageJ. On day 6, organoids were spun down at 3000 rpm for 10 min at 10°C and fixed with 4% PFA for 45 min at RT. For lentiviral infection, organoids were transduced on day 3 with EGFP virus together with Cyp2e1-mCherry virus, or GS-EGFP virus together with mCherry virus, and images were acquired on day 4 (early time point) and day 6 (late time point) for analysis. Number of EGFP+ (mCherry-) cells and mCherry+ (EGFP-) cells were counted from each condition on day 4 and day 6 for ratio analysis. Immunostaining of TSA-treated intestine organoids in 96-well plates were performed as previously described (Serra et al., 2019). Briefly, organoids were permeabilized with 0.5% Triton X-100 for 1 hour and blocked with 3% Donkey Serum (Sigma-Aldrich) in PBS with 0.1% Triton X-100 for 1 hour. Primary and secondary antibodies were diluted in blocking buffer and applied as indicated under the Immunohistochemistry (IHC) and *in situ* hybridization (ISH) section.

Lentiviral packaging

The constructs CMV-mCherry, CMV-EGFP, CMV-mCyp2e1-T2A-mCherry and CMV-mGlul-T2A-EGFP were packaged into lentiviral particles, respectively, using HEK293T cells (ATCC, CRL-11268). Lentiviral packaging was performed in 15 cm cell culture dishes. 24 h after cell plating, transfection mixes were prepared by mixing 20 μg of each plasmid with helper plasmids (pLP/VSVG, pLP1 and pLP2), 2xHBS buffer and 1M CaCl2 at RT for 15 minutes before adding dropwise to HEK293T culture. Medium was changed after 24 h and virus was collected 48 h after the transfection. Lentivirus was concentrated with Lenti-X concentrator according to the manufacturer's instructions (Takara, 631232).

RNA extraction, reverse transcription and quantitative RT-PCR (qPCR)

Total RNA was isolated from intestinal organoids using Total RNA Purification Kit (NORGEN, 37500) according to the manufacturer's instructions. RNA purification was performed in triplicate and 2 wells of a 24 well (50 μl droplet containing 30,000 single cells on day 0) were pooled per sample. DNAase treatment was performed using the RNase-Free DNase I Kit (NORGEN, 25710). RNA was reverse transcribed using the high-capacity cDNA reverse transcription kit (Thermo Fisher Scientific, 4368814). PCR reactions were run on a CFX96 Touch Deep Well Real-Time PCR Detection System (Bio-Rad) and data processing was performed using comparative Ct method (ΔCt). The following Taqman assays were used: *Glul* (Mm00725701_s1), *Cyp2e1* (Mm00491127_m1).

Bulk RNA-seq of mouse livers and HCCs, and count generation

Gene expression for liver samples was measured using RNA sequencing technology. Total RNA from samples was isolated using RNeasy mini plus kit (QIAGEN, 74134) according to the manufacturer's instructions. Amount and quality of the RNA were measured with the Agilent RNA 6000 Nano Kit (Agilent Technologies, 5067-1511). RNA sequencing libraries were prepared using the Illumina TruSeq RNA Sample Prep Kit v2 (Illumina, RS-122-2001) and sequenced using the Illumina HiSeq2500 platform. Samples were sequenced to a length of 2x76 base-pairs. Read pairs were mapped to the mouse genome (GRCh37) and the mouse gene transcripts from Ensembl (Cunningham et al., 2015) by using an in-house gene quantification pipeline (Schuierer and Roma, 2016). Genome and transcript alignments were used to calculate gene counts based on Ensembl gene IDs and for the 28 to 44 million raw reads we obtained alignment rates of an average of 81% to the mm10 reference.

Human HCC data analysis

Mutational and expression data (raw counts) from The Cancer Genome Atlas (liver dataset; LIHC) were downloaded using *TCGAbiolinks* (Colaprico et al., 2016). From this cohort we identified 89 samples harboring one of the common activating hotspot mutations in the *CTNNB1* gene (*S45P, S45F, S37C, T41A, G34V, S37Y, S37F, S37A, S45Y, N387Y, N387I, W383C, N387K, K335T, I35S, T41I, S33P, H36P, D32N, D32Y, K335I, S33C, G34R, D32G, S33Y, G34E, D32V*) and 23 samples harboring loss of function mutations in the *AXIN1* gene. From these samples we extracted the expression profile (as raw counts) using RNA-seq data.

Cell Stem Cell
Article

Expression data from these samples were analyzed together with the RNA-seq data obtained from the 8 DEN^{AXIN2} and 12 ZNRF3/RNF43$^{\Delta Hep}$ HCC samples. Data intersection resulted in 11,982 common genes between the mouse cohorts and TCGA LIHC cohort (human) for further analysis (only genes with the corresponding mouse gene were selected). Genes with low expression (< 1 log-counts per million in ≥ 5 samples) were filtered out. Normalization was performed using the "TMM" (weighted trimmed mean) method from the edgeR (Robinson et al., 2010) package. We removed batch effects from expression data using the species information (mouse and human) and used top 10% highly variable genes for MDS (Multidimensional scaling) plot and heatmap. Heatmap was generated using ComplexHeatmap package (Gu et al., 2016).

Bulk RNA-seq data analysis of mice with ZNRF3/RNF43 deletion or RSPO injection, and respective control mice

Read pairs were mapped to the mouse genome (GRCh37) and the mouse gene transcripts from Ensembl (Cunningham et al., 2015) by using an in-house gene quantification pipeline (Schuierer and Roma, 2016). Genome and transcript alignments were used to calculate gene counts based on Ensembl gene IDs and for the 28 to 35 million raw reads we obtained alignment rates around 80%. We removed two samples (CTRL_5, PBS_2) from further analysis based on outlier behavior in a PCA analysis and discarded genes with counts < 1 cpm in more than 60% of samples per group for all groups. We applied the differential gene expression data analysis Bioconductor package edgeR (Robinson et al., 2010). In brief, weights for linear modeling were estimated using voom() from the limma package and linear models were fitted with lmFit() (Ritchie and Flicek, 2015). An empirical Bayes eBayes() moderation was used to determine statistics for differentially expressed genes in two comparisons (ZNRF3/RNF43$^{\Delta Hep}$ (ZRdKO) versus control (CTRL), and 50mg/kg RSPO- versus PBS-treated mice). We applied the Benjamini–Hochberg correction for multiple hypothesis testing (Benjamini and Hochberg, 1995). GSEA (Figure S4A) was performed as described below in the 'Spatial transcriptomics and bioinformatics analysis' section, except that we used logFC as the input gene scores.

Spatial transcriptomics and bioinformatics analysis

Samples from RNF43$^{\Delta HEP}$, ZNRF3$^{\Delta HEP}$, ZNRF3/RNF43$^{\Delta HEP}$ and control mice, as well as 50mg/kg RSPO-or PBS-treated mice were used for spatial transcriptomics. Sections of a thickness of 10 μm were placed on a Visium Spatial Gene Expression Slide (10x Genomics, PN-2000233), stored at −80°C for 1 - 7 days, and processed according to the manufacturer instructions with the following parameters: we used 12 minutes for tissue permeabilization, 12 - 16 cycles of PCR for cDNA amplification, and 12 - 13 cycles for the sample index PCR (Dual Index Kit TT Set A, 10X Genomics, PN-1000215). Libraries were pooled, loaded at a concentration of 300 pM on an NovaSeq 6000 SP or S2 flow cell, and sequenced at 28 and 120 bases for read 1 and 2, respectively, using v1 reagents (Illumina).

Fastq reads from the Illumina sequencer NovaSeq 6000 were preprocessed by 10X Genomics spaceranger-1.0.0 software including spatial de-barcoding, read-alignment to mm10 and UMI-generation. QC-metrics, such as total aligned reads, median number of UMIs or genes per spot for the different samples, are all in the acceptable range (Figures S2S, S6A, and S7M). For the follow-up data analysis, we used the R-package Seurat v3.2.3 (http://satijalab.org/seurat/) (Stuart et al., 2019), R 4.0.2 and the Seurat-wrapper STTKIT 0.0.9 (https://github.com/lima1/sttkit) for initial fast processing of the spaceranger processed data on command-line in batch-mode. During this step a list of Seurat objects were generated, consisting of one object for each sample with SCTransformed expression values as well as log-normalized counts. We used Seurat's anchor-based integration approach to combine all 4x2 samples as described in (https://satijalab.org/seurat/articles/integration_introduction.html) using 3000 variable genes generating one large collection of spots for further processing. A 2-dimensional UMAP embedding was generated based on the first 30 PCs. For cluster identification, we used the graph-based clustering approach as implemented in the 'FindNeighbors()' function also choosing the first 30 PCs as input and the 'FindClusters()' function with the cluster resolution parameter set to 0.5. We found 15 clusters and checked for any biases by extensive exploration of the spot-cluster distribution in the UMAP representation split by sample or replicate ID but we could not identify any replicate or sample bias in the clusters.

To identify periportal (zone 1) and pericentral (zone 3) regions, we used the 12 periportal and 12 pericentral marker genes with highest absolute log2fc in the waterfall plot Figure 1K in (Sun et al., 2020), respectively. For visualization of the spatial expression patterns of interest, we used the min.cutoff and max.cutoff clamping option in Seurat's SpatialPlot() function whose values correspond to the lower and upper limits in the respective color scale legend.

For differential gene expression analysis of the clusters, we used the Wilcoxon test as implemented in Seurat's FindMarker() function. In some cases, we performed gene set enrichment analysis (GSEA) (Subramanian et al., 2005) with the Bioconductor R package fgsea (Korotkevich et al., 2021) using the negative log10 of the adjusted p value multiplied by sign(logFC) from the respective differential expression analysis result as input score and the MSigDB hallmark gene sets (Liberzon et al., 2015) as input gene set collection. We used the appropriate leading edge genes as determined by fgsea in the heatmap of Figure 6D. The leading edge subset consists of those genes that contribute most to a given enrichment signal. The respective heatmap was created with the Bioconductor R package complexHeatmap (Gu et al., 2016).

We found a dominating expression of hepatocyte markers (Alb) in all spots. Hepatocytes are considerably larger compared to many other cell types found in the liver (Marguerat et al., 2012) diluting their expression signals considerably. Also, given the limited resolution of the current Visium chip, it was not surprising that we were not able to detect regions with substantial expression of marker genes for some non-parenchymal cells such as biliary cells. In order to identify traces of other (small) cell types, we checked

							Table 1: Selected modules and their associated pathways from the co-expression network analysis		
ID	Description	Count	p.value	p.adj	qvalue	Protein coding genes	Subtypes	Modules
GO	DNA replication	75	1.34E-68	3.04E-65	2.44E-65	ALYREF/ATAD5/BARD1/BLM/BRCA2/BRIP1/CCNA2/CCNE1/CCNE2/CDC45/CDC6/CDC7/CDK1/CDK2/CDK2AP1/CDT1/CHAF1A/CHAF1B/CHEK1/CHEK2/CLSPN/DBF4/DDX11/DNA2/DONSON/DTL/E2F7/E2F8/EME1/ESCO2/EXO1/FBXO5/FEN1/GINS1/GINS2/GINS3/GINS4/GMNN/HMGA1/MCM10/MCM2/MCM3/MCM4/MCM5/MCM6/MCM7/MCM8/MSH6/NASP/ORC1/ORC6/PCLAF/PCNA/PIF1/POLA1/POLA2/POLE2/POLQ/PRIM2/RAD51/RFC2/RFC4/RFC5/RMI1/RMI2/RNASEH2A/RRM1/RRM2/SLBP/SSRP1/TICRR/TIMELESS/TIPIN/TOPBP1/WDHD1	BL1	M1
GO	cell cycle checkpoint	45	4.10E-35	7.16E-33	5.76E-33	AURKA/AURKB/BLM/BRIP1/BUB1/BUB1B/CCNB1/CDC20/CDC25C/CDC45/CDC6/CDK1/CDK2/CDT1/CENPF/CHEK1/CHEK2/CLSPN/DCLRE1B/DNA2/DONSON/DTL/E2F1/E2F7/E2F8/EME1/GTSE1/KNTC1/MAD2L1/MSH2/NDC80/ORC1/PCNA/PLK1/SPDL1/TICRR/TIMELESS/TIPIN/TOP2A/TOPBP1/TRIP13/TTK/WDR76/ZWILCH/ZWINT	BL1	M1
GO	double-strand break repair	41	7.74E-28	6.52E-26	5.24E-26	AUNIP/BARD1/BLM/BRCA2/BRIP1/CDC45/CDC7/CDCA5/CHEK1/CHEK2/DCLRE1B/DDX11/DEK/DNA2/EME1/ESCO2/FANCB/FEN1/FOXM1/GINS2/GINS4/MCM2/MCM3/MCM4/MCM5/MCM6/MCM7/MCM8/MSH2/NSD2/PARPBP/POLA1/POLQ/RAD51/RAD51AP1/RAD54L/RMI1/RMI2/TIMELESS/TRIP13/XRCC2	BL1	M1
Reactome	G2/M Transition	27	3.67E-14	4.87E-13	3.35E-13	AURKA/BORA/CCNA2/CCNB1/CCNB2/CDC25A/CDC25C/CDK1/CDK2/CENPF/CENPJ/CEP152/CKAP5/FOXM1/GTSE1/HAUS8/HMMR/LIN9/MYBL2/NEK2/PKMYT1/PLK1/PLK4/TPX2/TUBA1B/TUBA1C/TUBB	BL1	M1
Kegg	Cell cycle	53	9.85E-51	1.12E-48	8.50E-49	BUB1/BUB1B/CCNA2/CCNB1/CCNB2/CCNE1/CCNE2/CCNE2/CDC20/CDC25A/CDC25C/CDC45/CDC6/CDC6/CDC7/CDK1/CDK2/CDK2/CDKN2D/CDKN2D/CHEK1/CHEK1/CHEK2/DBF4/E2F1/E2F1/E2F5/ESPL1/MAD2L1/MCM2/MCM3/MCM4/MCM4/MCM5/MCM5/MCM6/MCM7/MCM7/MCM7/ORC1/ORC1/ORC6/ORC6/P	BL1	M1

						CNA/PKMYT1/PKMYT1/PLK1/PTTG1/PTTG1/RBL1/SKP2/SKP2/TTK		
GO	extracellular matrix organization	58	9.43E-51	1.43E-47	1.14E-47	ADAM12/ADAMTS14/ADAMTS2/AEBP1/ANTXR1/BGN/BMP1/COL10A1/COL11A1/COL12A1/COL18A1/COL1A1/COL1A2/COL3A1/COL5A1/COL5A2/COL5A3/COL6A1/COL6A2/COL6A3/COL8A1/COL8A2/COMP/CREB3L1/EFEMP2/EMILIN1/FAP/FBLN1/FBLN2/FBN1/FN1/GREM1/HSPG2/HTRA1/ITGA11/ITGA5/ITGB5/LOX/LOXL1/LRP1/LUM/MATN3/MFAP5/MMP11/MMP14/MMP2/MYH11/NID1/NID2/PDPN/PHLDB1/POSTN/SERPINE1/SPARC/SULF1/THBS1/TIMP2/VCAN	BL2	M2
Reactome	Signaling by MET	8	7.78E-06	0.0001	9.63E-05	COL11A1/COL1A1/COL1A2/COL3A1/COL5A1/COL5A2/COL5A3/FN1	BL2	M2
GO	regulation of Wnt signaling pathway	11	0.002	0.0272	0.0217	ADGRA2/COL1A1/CTHRC1/DACT1/DACT3/DKK3/GREM1/HIC1/LRP1/ROR2/SULF1	BL2	M2
GO	positive regulation of chromatin organization	7	0.0008	0.042	0.0385	AKAP8/AKAP8L/BRD4/PRKD2/RNF40/SPHK2/TRIM28	MES	M3
GO	beta-catenin binding	9	6.74E-06	0.0007	0.0006	AXIN1/BCL9/BCL9L/CARM1/DVL3/KDM6B/MED12/SETD1A/SUFU	MES	M3
GO	T cell mediated immunity	34	1.55E-18	7.11E-17	5.41E-17	B2M/BTN3A2/BTN3A3/CCR2/CD1D/CD8A/CTSC/CTSH/FCGR2B/FOXP3/GZMM/HLA-A/HLA-E/HLA-F/IL12RB1/IL18/IL18R1/IL18RAP/IL7R/LILRB1/MICB/MYO1G/NLRP3/P2RX7/PRF1/PTPRC/RAB27A/RFTN1/RIPK3/RSAD2/SASH3/TBX21/TNFRSF1B/WAS	IM	M4

Kegg	Antigen processing and presentation	70	9.92E-34	1.97E-32	1.50E-32	B2M/CD4/CD4/CD74/CD8A/CD8A/CD8A/CD8B/CIITA/CIITA/CIITA/CTSS/HLA-A/HLA-A/HLA-DMA/HLA-DMA/HLA-DMB/HLA-DMB/HLA-DOA/HLA-DPA1/HLA-DPA1/HLA-DPB1/HLA-DPB1/HLA-DRA/HLA-DRA/HLA-E/HLA-E/HLA-E/HLA-F/HLA-F/IFI30/IFI30/KLRD1/KLRD1/PSME1/PSME1/PSME2/PSME2/TAP1/TAP1	IM	M4
Kegg	JAK-STAT signaling pathway	38	9.07E-07	5.04E-06	3.84E-06	CCND2/CCND3/CSF2RB/CSF2RB/CSF3R/IL10RA/IL10RB/IL12RB1/IL15/IL15RA/IL21R/IL2RA/IL2RB/IL2RG/IL4R/IL6R/IL6R/IL7/IL7/IL7R/IRF9/JAK1/JAK1/JAK2/JAK3/JAK3/OSM/PIK3CD/PTPN6/SOCS1/SOCS1/STAT1/STAT2/STAT2/STAT4/STAT5A/STAT5A/STAT5A	IM	M4
Kegg	NF-kappa B signaling pathway	48	7.56E-20	7.78E-19	5.92E-19	ATM/ATM/BIRC3/BLNK/BTK/BTK/CARD11/CARD11/CARD11/CCL19/CCL19/CCL4/CD14/CD40/CD40/CD40/CFLAR/CYLD/DDX58/LAT/LAT/LCK/LCK/LTB/LTB/LY96/LYN/LYN/LYN/MAP3K14/MAP3K14/NFKBIA/PLCG2/PRKCB/PRKCQ/SYK/SYK/TLR4/TNFAIP3/TNFRSF13C/TNFRSF13C/TNFSF13B/TNFSF13B/TNFSF14/TRAF1/TRAF3/VCAM1/ZAP70	IM	M4
Kegg	B cell receptor signaling pathway	42	1.43E-18	1.42E-17	1.08E-17	BLNK/BTK/BTK/CARD11/CARD11/CARD11/CD22/CD22/CD72/CD72/CD79A/CD79B/DAPP1/FCGR2B/IFITM1/IFITM1/INPP5D/LILRB1/LILRB1/LILRB2/LILRB4/LYN/LYN/LYN/NFATC2/NFKBIA/PIK3AP1/PIK3AP1/PIK3CD/PLCG2/PPP3CC/PRKCB/PTPN6/RAC2/RAC2/RAC2/RASGRP3/SYK/SYK/VAV1/VAV1/VAV1	IM	M4

Kegg	Th1 and Th2 cell differentiation	75	7.40E-40	2.94E-38	2.24E-38	CD247/CD3D/CD3D/CD3E/CD3G/CD3G/CD4/CD4/HLA-DMA/HLA-DMA/HLA-DMB/HLA-DMB/HLA-DOA/HLA-DPA1/HLA-DPA1/HLA-DPB1/HLA-DPB1/HLA-DRA/HLA-DRA/IL12RB1/IL2RA/IL2RB/IL2RG/IL4R/JAK1/JAK1/JAK2/JAK3/JAK3/LAT/LAT/LCK/LCK/NFATC2/NFKBIA/PPP3CC/PRKCQ/RUNX3/STAT1/STAT4/STAT5A/STAT5A/STAT5A/TBX21/ZAP70	IM	M4
GO	canonical Wnt signaling pathway	13	0.0033	0.0352	0.02835	BICC1/DAB2/FOXO1/IGFBP4/IGFBP6/MITF/RBMS3/SEMA5A/SFRP2/SFRP4/SOX17/TMEM88/TNN	MSL	M5
Kegg	Calcium signaling pathway	18	0.0001	0.0040	0.00345	ADCY4/AVPR1A/AVPR1A/EDNRB/F2R/F2R/FGF7/FLT1/FLT1/FLT4/HGF/HTR2B/KDR/PDE1A/PDE1B/PDE1B/PDGFD/PDGFRA	MSL	M5
Kegg	Focal adhesion	23	8.31E-08	4.24E-06	3.61E-06	FLT1/FLT1/FLT4/HGF/IGF1/ITGA1/KDR/LAMA2/LAMA2/LAMA4/LAMA4/LAMB1/LAMB1/PDGFD/PDGFRA/RELN/TNN/TNN/TNXB/TNXB/TNXB/TNXB/VWF	MSL	M5
Kegg	ECM-receptor interaction	16	5.18E-09	6.90E-07	5.87E-07	FREM1/ITGA1/LAMA2/LAMA2/LAMA4/LAMA4/LAMB1/LAMB1/RELN/TNN/TNN/TNXB/TNXB/TNXB/TNXB/VWF	MSL	M5
Kegg	ABC transporters	8	3.92E-05	0.0011	0.0009	ABCA10/ABCA6/ABCA8/ABCA8/ABCA9/ABCC9/ABCC9/ABCG2	MSL	M5
Kegg	Vascular smooth muscle contraction	11	0.0014	0.0157	0.0134	ADCY4/ADCY5/ADCY5/AVPR1A/AVPR1A/CALCRL/NPR1/NPR1/PTGIR/RAMP2/RAMP2	MSL	M5
Reactome	Fatty acid metabolism	7	3.64E-06	0.0004	0.0004	ALOX15B/CRAT/CROT/ELOVL5/GGT1/HPGD/SCP2	LAR	M6
Reactome	Phenylalanine and tyrosine metabolism	2	0.0007	0.0244	0.0217	HGD/KYAT1	LAR	M6

GO	ATPase-coupled transmembrane transporter activity	3	0.0029	0.0436	0.0325	ABCA12/ABCC11/ABCC2	LAR	M6
GO	steroid binding	3	0.0024	0.0436	0.03256	APOD/AR/SCP2	LAR	M6
GO	carboxylic acid biosynthetic process	8	5.73E-06	0.0007	0.00061	ACSM1/ALOX15B/ELOVL5/GGT1/HAAO/HPGD/KYAT1/SCP2	LAR	M6

lncRNA Type	GeneSymbols	WGCNA	FUSCC_IMvs Other_log2FC	FUSCC_IMv sOther_padj	FUSCC_TNBCv sNor_log2FC	FUSCC_TNBCv sNor_padj	TCGA_TNBCvsn onT_log2FC	TCGA_TNBCvs nonT_padj
Annotated	AC004585.1	M2 (IM)	1.4688	2.54E-20	2.709210272	7.40E-31	0.5767294755	1.14E-05
Annotated	AC004687.1	M2 (IM)	1.5481	1.19E-29	1.898927141	1.12E-22	0.7436812866	2.32E-07
Annotated	AC004847.1	M2 (IM)	1.176715654	9.93E-16	1.043748944	4.66E-07	0.5351487703	2.95E-07
Annotated	AC005224.3	M2 (IM)	1.15698825	2.01E-16	1.243610488	1.85E-10	0.7281389877	1.78E-05
Annotated	AC005224.4	M2 (IM)	1.101729039	7.92E-16	1.514910506	1.08E-16	0.7172863227	8.35E-09
Annotated	AC005842.1	M2 (IM)	1.316338784	3.63E-06	1.572239779	0.000521	0.4864541269	0.00254
Annotated	AC006369.1	M2 (IM)	1.37837552	2.35E-18	1.650755504	3.24E-13	0.709799496	4.73E-15
Annotated	AC007336.2	M2 (IM)	1.064900082	1.08E-07	2.58586588	1.95E-13	0.4546312885	0.0037390
Annotated	AC007381.1	M2 (IM)	1.297745154	3.31E-06	2.936613707	1.09E-10	1.757688023	3.16E-13
Annotated	AC007386.2	M2 (IM)	1.821011236	4.50E-16	2.176182349	2.82E-12	1.154104731	8.02E-15
Annotated	AC007569.1	M2 (IM)	2.000552671	1.83E-26	1.722139357	4.91E-10	0.5918573939	6.55E-05
Annotated	AC007728.2	M2 (IM)	1.286662585	1.05E-32	1.536769111	4.71E-17	0.3530267975	3.77E-05
Annotated	AC007991.2	M2 (IM)	1.963217205	3.15E-17	3.231505877	5.22E-20	2.98270188	1.19E-34

Table 2: lncRNAs upregulated in IM vs other subtypes and also upregulated in TNBCs vs Normals and TNBCs vs non-TNBCs

These lncRNAs are also associated with the Module M2 in WGCNA

Annotated	AC008033.3	M2 (IM)	1.711017782	2.53E-19	1.072901137	3.73E-05	0.6193148377	0.0004226942507
Annotated	AC008957.1	M2 (IM)	1.136619422	7.80E-17	1.172043084	9.98E-09	0.3452527237	0.004877423662
Annotated	AC009133.3	M2 (IM)	1.808502178	2.72E-19	1.43884906	3.67E-06	0.5116760887	0.002040631197
Annotated	AC010247.1	M2 (IM)	1.081679304	1.57E-10	1.008141397	2.97E-05	0.3906405316	0.00288620663
Annotated	AC011944.1	M2 (IM)	1.184410889	4.57E-11	1.657835596	1.20E-08	1.62797519	8.89E-36
Annotated	AC012236.1	M2 (IM)	1.801250269	7.81E-13	2.574865752	9.09E-11	1.456095417	8.17E-25
Annotated	AC012645.3	M2 (IM)	1.125017344	1.46E-13	1.098392599	1.94E-06	0.6875676868	6.10E-12
Annotated	AC013264.1	M2 (IM)	1.277259271	6.46E-07	1.89749689	9.97E-06	0.5521483503	0.004234008924
Annotated	AC017002.3	M2 (IM)	1.274839669	1.58E-14	2.940781824	8.56E-17	0.6999475336	2.49E-13
Annotated	AC021678.2	M2 (IM)	1.7822223	1.26E-14	1.219002412	0.001236687098	0.914385061	5.36E-08
Annotated	AC022182.2	M2 (IM)	1.373085176	1.97E-12	1.064703381	0.0002173142951	0.4031205817	0.000333593972
Annotated	AC023202.1	M2 (IM)	1.918740171	2.28E-12	2.642307586	3.74E-08	1.547947154	2.22E-07
Annotated	AC023590.1	M2 (IM)	1.500639502	6.68E-11	2.319875427	1.58E-11	0.6178795789	6.44E-08
Annotated	AC023796.2	M2 (IM)	1.822461341	1.26E-10	3.050528523	2.11E-10	1.348140301	7.28E-26
Annotated	AC074366.1	M2 (IM)	1.45805045	2.90E-14	1.975800357	9.12E-08	0.5881427245	0.02670740816
Annotated	AC078883.1	M2 (IM)	1.481482583	5.95E-24	1.49322919	1.77E-13	1.208801764	3.10E-43
Annotated	AC078883.3	M2 (IM)	1.309133639	1.43E-17	1.299215748	4.48E-08	1.232347574	2.80E-31
Annotated	AC079089.1	M2 (IM)	1.400449592	1.02E-08	2.869756157	4.72E-15	2.518112715	2.68E-43
Annotated	AC079209.1	M2 (IM)	1.47594666	1.77E-14	2.119199668	4.67E-10	0.7188629529	2.95E-09

Annotated	AC079316.2	M2 (IM)	1.009253652	3.40E-23	1.781687546	4.67E-36	0.8940236945	0.0009881640585
Annotated	AC083837.1	M2 (IM)	1.465919319	1.22E-22	1.440483438	3.61E-12	1.354530017	1.90E-18
Annotated	AC083949.1	M2 (IM)	1.303021573	8.66E-18	2.151451684	1.24E-14	0.8548347232	1.68E-21
Annotated	AC092145.1	M2 (IM)	1.627777015	2.30E-13	2.349297867	9.11E-09	0.9887803865	8.97E-07
Annotated	AC092652.3	M2 (IM)	1.359037742	3.02E-24	1.02748162	5.73E-08	0.9266090343	1.43E-09
Annotated	AC093010.2	M2 (IM)	1.479774113	3.97E-16	2.252815792	1.66E-18	0.8406849817	1.30E-07
Annotated	AC093583.1	M2 (IM)	1.247194543	5.76E-08	2.260213884	7.87E-09	1.201027207	3.74E-14
Annotated	AC096667.1	M2 (IM)	1.120087098	6.48E-11	1.841043154	4.47E-13	0.279446138	0.007174947379
Annotated	AC096734.1	M2 (IM)	1.662205236	3.43E-13	1.942181271	1.45E-07	1.536232537	5.49E-09
Annotated	AC097375.1	M2 (IM)	1.830720642	9.00E-08	2.846183909	1.11E-07	0.7642330094	0.007026361061
Annotated	AC098613.1	M2 (IM)	1.032586513	1.38E-11	1.661202588	7.04E-15	0.2971343268	0.003412554281
Annotated	AC099063.4	M2 (IM)	1.514884845	2.28E-19	1.307921926	1.40E-06	0.48344232	0.000364796335
Annotated	AC099524.1	M2 (IM)	1.734899436	2.52E-21	2.229195621	9.85E-14	0.7640230957	1.43E-08
Annotated	AC104261.1	M2 (IM)	1.799520054	3.82E-13	2.852137552	2.83E-10	1.233126142	2.01E-07
Annotated	AC104699.1	M2 (IM)	2.037349823	6.90E-20	2.188239092	1.51E-11	1.070898252	2.28E-53
Annotated	AC104971.4	M2 (IM)	1.215322219	4.23E-06	1.676979258	1.97E-05	0.7278519004	0.001034607734
Annotated	AC106882.1	M2 (IM)	1.532105409	2.11E-10	1.896970555	1.95E-07	1.541346026	1.57E-14
Annotated	AC108206.1	M2 (IM)	1.444138087	5.75E-05	3.10223107	2.85E-09	0.7246903199	0.01059235466
Annotated	AC109446.3	M2 (IM)	1.493775044	1.05E-16	2.011131818	1.85E-10	0.3380109293	0.001761418221

Annotated	AC120193.1	M2 (IM)	1.54649702	5.92E-13	2.957778753	9.45E-21	1.358850214	2.76E-10
Annotated	AC139720.1	M2 (IM)	1.69910284	1.56E-14	1.811955352	3.34E-07	1.022270688	4.05E-07
Annotated	AC243960.1	M2 (IM)	1.663635449	2.00E-21	1.113596537	5.27E-05	0.6667712296	4.00E-07
Annotated	AC246787.2	M2 (IM)	1.780595716	7.48E-09	1.872428719	0.000164289818	0.9153793769	0.0002435159735
Annotated	AF127936.1	M2 (IM)	1.322023021	1.97E-24	1.040346765	1.94E-08	0.2557935159	0.02533175981
Annotated	AF127936.2	M2 (IM)	1.40926962	2.10E-19	1.062664373	1.98E-06	0.2902589415	0.02638381571
Annotated	AL022067.1	M2 (IM)	1.381050958	7.39E-18	1.307745397	2.77E-09	0.8937822292	1.81E-08
Annotated	AL023653.1	M2 (IM)	1.560475264	3.15E-19	1.49164874	4.57E-08	0.5370797009	2.68E-08
Annotated	AL096816.1	M2 (IM)	1.782953262	7.39E-21	1.241416039	3.17E-06	0.4953141167	0.002435945182
Annotated	AL121748.1	M2 (IM)	1.514443375	1.29E-12	3.023090109	4.67E-13	0.5011827938	0.02435688602
Annotated	AL121820.1	M2 (IM)	1.180877753	1.85E-10	2.698673594	5.74E-13	1.760779832	2.14E-53
Annotated	AL121985.1	M2 (IM)	1.68825363	5.66E-23	1.720114077	3.64E-12	0.5973885492	4.92E-08
Annotated	AL139020.1	M2 (IM)	1.563138422	0.000795	3.380209863	1.06E-07	1.200734635	4.00E-07
Annotated	AL139125.1	M2 (IM)	1.450064835	1.11E-17	1.835407805	5.27E-11	0.495053299	0.004478746337
Annotated	AL158210.1	M2 (IM)	1.420495723	1.65E-10	1.468809499	2.36E-05	1.105299195	2.33E-06
Annotated	AL160400.1	M2 (IM)	1.557117649	5.07E-25	1.651966568	1.74E-13	0.5832301925	6.75E-05
Annotated	AL161781.2	M2 (IM)	1.681317114	1.93E-05	2.451184725	2.08E-05	1.162341901	2.05E-05
Annotated	AL365361.1	M2 (IM)	1.823923286	1.56E-28	1.434327911	3.94E-10	0.5546399877	3.05E-05
Annotated	AL512631.1	M2 (IM)	1.717922473	2.34E-07	1.98754403	0.0001221740717	0.8691818957	9.77E-05

Annotated	AL590764.1	M2 (IM)	1.385921576	8.99E-24	1.679896752	6.71E-15	0.4934138389	2.85E-07
Annotated	AL591468.1	M2 (IM)	2.009775931	2.79E-24	2.912350466	4.00E-15	1.038923034	5.32E-08
Annotated	AL592164.1	M2 (IM)	1.88315965	1.42E-21	1.927381205	3.87E-09	0.8676015179	0.0001827736408
Annotated	AOAH-IT1	M2 (IM)	1.248891439	7.87E-20	1.04157968	8.28E-08	0.6542575436	0.006279294788
Annotated	AP003419.1	M2 (IM)	1.248488489	6.11E-12	1.321465728	4.89E-06	0.3439841753	1.11E-06
Annotated	AP003557.1	M2 (IM)	1.203851921	2.53E-14	1.256930235	2.44E-07	0.7079879023	0.0001928027473
Annotated	AP003774.4	M2 (IM)	1.295037748	4.24E-09	1.89292921	1.72E-06	0.7247223694	1.48E-07
Annotated	C9orf139	M2 (IM)	1.149762936	3.41E-18	1.38014716	3.68E-14	0.2967668543	0.0005114496552
Annotated	CHRM3-AS2	M2 (IM)	1.357118311	2.33E-11	1.554165279	1.12E-08	1.507134384	6.31E-17
Annotated	FAM30A	M2 (IM)	1.971865777	6.99E-17	1.856913792	1.92E-09	1.003907159	8.80E-08
Annotated	IFNG-AS1	M2 (IM)	1.903226697	2.67E-22	2.082524059	2.42E-15	0.6297203237	0.001694576876
Annotated	IL21-AS1	M2 (IM)	1.554491709	2.65E-13	3.506938021	1.89E-17	1.215316056	3.79E-09
Annotated	ITGB2-AS1	M2 (IM)	1.117457069	2.31E-18	1.331150954	3.35E-15	0.8050269058	1.13E-12
Annotated	LINC00158	M2 (IM)	1.320943766	9.31E-14	1.002516848	6.50E-05	0.9614213249	1.32E-07
Annotated	LINC00239	M2 (IM)	1.094599886	2.98E-09	1.956088495	1.90E-11	1.132411892	1.74E-17
Annotated	LINC00402	M2 (IM)	1.592425538	8.43E-11	2.768367353	1.12E-17	0.889779032	6.13E-06
Annotated	LINC00426	M2 (IM)	1.484636746	5.07E-25	1.91599036	5.48E-22	0.5613594889	1.14E-05
Annotated	LINC00427	M2 (IM)	1.554944854	1.41E-13	2.162876042	1.93E-08	0.4683952879	0.006283735219
Annotated	LINC00487	M2 (IM)	1.111661064	1.88E-08	1.633779998	6.93E-10	1.177174717	5.43E-18

Annotated	LINC00528	M2 (IM)	1.494687709	2.46E-29	1.26047022	6.78E-10	0.6568451294	2.18E-09
Annotated	LINC00544	M2 (IM)	1.131238272	1.06E-07	1.890109615	3.35E-09	0.4832949959	0.01944464685
Annotated	LINC00582	M2 (IM)	1.769782463	9.67E-12	1.259845495	0.0005860728292	1.10132265	7.00E-09
Annotated	LINC00861	M2 (IM)	1.401876411	4.87E-13	1.67700047	2.57E-12	0.5966136477	0.0001320333803
Annotated	LINC00892	M2 (IM)	1.10349204	5.58E-09	1.292325548	1.09E-07	0.4763333938	0.001117068
Annotated	LINC00944	M2 (IM)	1.403205443	3.61E-17	1.406987677	3.00E-10	1.881665312	2.67E-40
Annotated	LINC00996	M2 (IM)	1.620223012	4.21E-22	1.285436533	2.72E-08	0.8100666388	5.12E-09
Annotated	LINC01215	M2 (IM)	1.703241591	5.16E-15	2.689125662	5.82E-20	0.7873717653	4.98E-05
Annotated	LINC01281	M2 (IM)	1.667150521	3.05E-17	4.609775427	4.92E-29	0.9009883311	2.11E-06
Annotated	LINC01727	M2 (IM)	1.206864941	1.51E-09	2.927809873	1.42E-15	0.572641952	0.008374895226
Annotated	LINC01781	M2 (IM)	1.663116467	6.24E-11	2.656864421	3.50E-11	0.96541081	1.44E-05
Annotated	LINC01800	M2 (IM)	1.262885561	5.95E-11	1.784936734	1.08E-09	1.039520897	3.64E-08
Annotated	LINC01857	M2 (IM)	1.479914188	6.56E-17	2.937830223	3.04E-24	0.5888037041	1.65E-05
Annotated	LINC01871	M2 (IM)	1.767943786	1.04E-27	1.066434452	1.13E-05	1.499732813	7.81E-27
Annotated	LINC01934	M2 (IM)	1.597100322	2.38E-27	1.639056309	9.81E-15	0.3874035698	0.008650097223
Annotated	LINC01943	M2 (IM)	1.279890626	5.23E-26	2.611395638	6.79E-36	0.3573813624	1.59E-06
Annotated	LINC02084	M2 (IM)	1.291061958	3.10E-16	1.513504456	1.82E-09	0.816944019	3.63E-10
Annotated	LINC02100	M2 (IM)	1.413361391	4.96E-22	1.659983428	2.59E-13	1.12072533	2.38E-11
Annotated	LINC02195	M2 (IM)	1.611576883	4.27E-12	3.250725534	3.88E-14	1.259291851	9.63E-13

Annotated	LINC02273	M2 (IM)	1.463192539	4.27E-14	2.198485402	3.18E-15	0.4516429574	0.007306668297
Annotated	LINC02320	M2 (IM)	1.325198776	1.14E-08	1.924738484	1.08E-06	1.694113055	1.04E-11
Annotated	LINC02325	M2 (IM)	1.546052484	7.26E-23	2.064338839	2.51E-16	0.6190352926	0.0006041876266
Annotated	LINC02341	M2 (IM)	1.089983926	1.31E-08	2.759546688	4.45E-15	1.432881373	1.62E-16
Annotated	LINC02384	M2 (IM)	1.797160742	1.44E-21	1.046470932	4.42E-05	0.716120104	5.91E-08
Annotated	LINC02397	M2 (IM)	1.569150124	4.52E-11	2.957388776	1.40E-22	0.9558303984	5.35E-07
Annotated	LINC02416	M2 (IM)	2.031938661	4.81E-34	3.11228156	1.02E-17	1.239114679	2.90E-11
Annotated	LINC02422	M2 (IM)	1.890941524	5.11E-17	2.380530919	2.71E-13	1.19657107	3.55E-08
Annotated	LINC02446	M2 (IM)	2.192092612	1.54E-28	2.818918585	6.59E-21	1.221506142	8.98E-12
Annotated	MIR155HG	M2 (IM)	1.458086052	1.68E-29	2.119972671	1.33E-29	0.9361582195	1.39E-17
Annotated	MIR4432HG	M2 (IM)	1.475010893	1.29E-08	3.290752516	2.18E-13	1.541530066	8.35E-08
Annotated	NRIR	M2 (IM)	1.525545765	8.23E-16	1.329231918	1.70E-07	0.7984934922	6.15E-08
Annotated	OSTN-AS1	M2 (IM)	1.783369102	2.48E-12	2.482139163	5.90E-08	1.37891384	6.50E-10
Annotated	PCED1B-AS1	M2 (IM)	1.507254397	4.43E-34	1.698536903	3.13E-21	0.3699447452	0.000574689808
Annotated	PIK3CD-AS1	M2 (IM)	1.383296835	9.12E-08	1.497410225	0.0005496893539	0.9806578661	4.70E-13
Annotated	PRKCQ-AS1	M2 (IM)	1.148737671	5.34E-13	1.226503515	6.99E-10	2.366790087	1.29E-77
Annotated	TCL6	M2 (IM)	1.141089101	3.37E-06	2.072397196	9.19E-12	0.6842288174	0.002364140288
Annotated	TRBV11-2	M2 (IM)	1.67578761	6.73E-16	1.916424507	7.78E-07	0.6910260895	3.18E-05
Annotated	TRG-AS1	M2 (IM)	1.580048467	5.32E-36	1.280389889	7.33E-12	0.6354767645	1.52E-08

Annotated	USP30-AS1	M2 (IM)	1.753644371	5.79E-35	1.805726829	1.85E-14	0.4736020837	1.62E-07
Annotated	XXYLT1-AS2	M2 (IM)	1.734333143	4.24E-19	1.928061763	3.60E-08	0.3692250917	0.02932404894
Annotated	ZBTB20-AS1	M2 (IM)	1.456371871	3.58E-21	2.520791876	4.30E-21	0.7410409369	1.79E-09
Novel	MSTRG.110466	M2 (IM)	1.729477868	3.99E-09	2.53134608	1.47E-07	NA	NA
Novel	MSTRG.117302	M2 (IM)	1.483352135	2.78E-08	1.155766998	0.002111224307	NA	NA
Novel	MSTRG.117305	M2 (IM)	1.40505107	8.60E-07	1.036996174	0.009786008865	NA	NA
Novel	MSTRG.117309	M2 (IM)	1.514144216	4.49E-07	1.050276407	0.01923602896	NA	NA
Novel	MSTRG.129038	M2 (IM)	1.831885325	1.53E-08	1.651213778	0.001571355985	NA	NA
Novel	MSTRG.129474	M2 (IM)	1.495134904	2.62E-12	1.544614958	8.91E-08	NA	NA
Novel	MSTRG.145126	M2 (IM)	1.102895531	1.61E-06	2.481039919	9.04E-12	NA	NA
Novel	MSTRG.152804	M2 (IM)	1.124635485	3.30E-27	2.349053487	4.30E-45	NA	NA
Novel	MSTRG.184441	M2 (IM)	1.351829777	3.03E-12	1.042871492	0.0003516748957	NA	NA
Novel	MSTRG.19363	M2 (IM)	1.038655934	2.04E-06	1.653167337	6.78E-06	NA	NA
Novel	MSTRG.197046	M2 (IM)	1.470711081	3.25E-23	1.430175392	3.59E-12	NA	NA
Novel	MSTRG.197093	M2 (IM)	1.578378489	6.28E-13	1.669684698	7.21E-06	NA	NA
Novel	MSTRG.197101	M2 (IM)	1.417753616	7.56E-09	1.67238905	3.49E-05	NA	NA
Novel	MSTRG.197112	M2 (IM)	1.870720008	4.97E-12	1.649323632	0.0005034994435	NA	NA
Novel	MSTRG.199161	M2 (IM)	1.574631333	4.59E-13	2.284715111	1.62E-08	NA	NA
Novel	MSTRG.202102	M2 (IM)	1.098370841	4.23E-09	1.198912025	2.60E-07	NA	NA

Novel	MSTRG.20391	M2 (IM)	1.70376874	2.25E-15	1.976626618	9.19E-13	NA	NA
Novel	MSTRG.211518	M2 (IM)	1.788362074	1.30E-07	2.27826235	1.60E-05	NA	NA
Novel	MSTRG.22231	M2 (IM)	1.135098044	1.07E-11	1.843360076	6.39E-12	NA	NA
Novel	MSTRG.22240	M2 (IM)	1.527883893	4.07E-11	1.549174542	0.000185	NA	NA
Novel	MSTRG.22253	M2 (IM)	1.073719472	2.64E-09	2.134559332	8.33E-12	NA	NA
Novel	MSTRG.224171	M2 (IM)	1.932655919	1.13E-24	2.4349658	3.98E-21	NA	NA
Novel	MSTRG.231604	M2 (IM)	1.103534271	1.82E-05	2.610279617	1.53E-15	NA	NA
Novel	MSTRG.245261	M2 (IM)	1.183050199	2.70E-07	1.713000121	1.82E-05	NA	NA
Novel	MSTRG.24814	M2 (IM)	2.000715934	1.43E-17	3.017131169	2.93E-21	NA	NA
Novel	MSTRG.32939	M2 (IM)	3.205791405	9.84E-34	3.926354698	4.13E-25	NA	NA
Novel	MSTRG.42942	M2 (IM)	1.004446874	7.64E-07	1.241447924	4.02E-05	NA	NA
Novel	MSTRG.47861	M2 (IM)	1.562928077	3.05E-19	1.640201835	3.10E-07	NA	NA
Novel	MSTRG.49951	M2 (IM)	1.936301507	5.71E-18	1.622856711	4.93E-07	NA	NA
Novel	MSTRG.58179	M2 (IM)	1.386811909	8.55E-08	2.631780046	3.20E-16	NA	NA
Novel	MSTRG.61661	M2 (IM)	1.202869941	9.62E-12	1.326139987	2.71E-05	NA	NA
Novel	MSTRG.62392	M2 (IM)	1.490811219	8.17E-27	1.670576887	4.44E-17	NA	NA
Novel	MSTRG.72763	M2 (IM)	1.826326528	1.01E-17	1.995002638	1.09E-07	NA	NA
Novel	MSTRG.72802	M2 (IM)	1.600301882	6.79E-15	1.255671714	0.0001187	NA	NA
Novel	MSTRG.78559	M2 (IM)	1.335802343	9.77E-12	2.189770913	2.89E-10	NA	NA

Novel	MSTRG.78576	M2 (IM)	1.23030302	1.02E-13	1.342764889	2.02E-06	NA	NA
Novel	MSTRG.78590	M2 (IM)	1.322096337	6.10E-16	1.532650968	2.40E-08	NA	NA
Novel	MSTRG.78591	M2 (IM)	1.208377325	1.05E-12	1.459086443	1.06E-08	NA	NA
Novel	MSTRG.78592	M2 (IM)	1.205207378	1.56E-11	1.451400394	5.10E-08	NA	NA
Novel	MSTRG.81610	M2 (IM)	1.659327178	2.65E-17	2.054004241	3.29E-14	NA	NA
Novel	MSTRG.84843	M2 (IM)	1.308356084	1.29E-12	1.682334946	1.41E-08	NA	NA
Novel	MSTRG.84858	M2 (IM)	1.524231223	2.99E-11	1.760778831	2.21E-05	NA	NA
Novel	MSTRG.87407	M2 (IM)	1.506495999	7.86E-21	1.621467288	9.37E-11	NA	NA

Table 3: 602 Immune lncRNAs (479 Annotated and 123 Novel lncRNAs)				
lncRNAs	**Chromosome**	**Ensembl_ID**	**lncRNA Type**	**Groups**
A2M-AS1	chr12	ENSG00000245105	antisense_RNA	Annotated lncRNAs
AC002091.2	chr17	ENSG00000266389	Intergenic	Annotated lncRNAs
AC002480.1	chr7	ENSG00000225541	sense_overlapping	Annotated lncRNAs
AC002480.2	chr7	ENSG00000232759	antisense_RNA	Annotated lncRNAs
AC002480.4	chr7	ENSG00000238033	Intergenic	Annotated lncRNAs
AC004067.1	chr4	ENSG00000273447	antisense_RNA	Annotated lncRNAs
AC004241.1	chr12	ENSG00000257433	antisense_RNA	Annotated lncRNAs
AC004466.2	chr12	ENSG00000274737	sense_intronic	Annotated lncRNAs
AC004491.1	chr7	ENSG00000279996	TEC	Annotated lncRNAs
AC004520.1	chr7	ENSG00000273237	antisense_RNA	Annotated lncRNAs

AC004584.3	chr17	ENSG00000263120	sense_intronic	Annotated lncRNAs
AC004585.1	chr17	ENSG00000266088	Intergenic	Annotated lncRNAs
AC004687.1	chr17	ENSG00000265206	antisense_RNA	Annotated lncRNAs
AC004706.1	chr17	ENSG00000261996	Intergenic	Annotated lncRNAs
AC004847.1	chr7	ENSG00000260997	sense_overlapping	Annotated lncRNAs
AC004865.2	chr1	ENSG00000239636	antisense_RNA	Annotated lncRNAs
AC004988.1	chr7	ENSG00000203446	antisense_RNA	Annotated lncRNAs
AC005224.3	chr17	ENSG00000266378	Intergenic	Annotated lncRNAs
AC005224.4	chr17	ENSG00000266709	Intergenic	Annotated lncRNAs
AC005291.1	chr17	ENSG00000264067	antisense_RNA	Annotated lncRNAs
AC005332.7	chr17	ENSG00000274712	Intergenic	Annotated lncRNAs
AC005479.2	chr14	ENSG00000259005	Intergenic	Annotated lncRNAs
AC005840.5	chr12	ENSG00000276718	antisense_RNA	Annotated lncRNAs

AC005899.6	chr17	ENSG00000274341	Intergenic	Annotated lncRNAs
AC006033.2	chr7	ENSG00000272908	Intergenic	Annotated lncRNAs
AC006058.1	chr3	ENSG00000261786	Intergenic	Annotated lncRNAs
AC006369.1	chr2	ENSG00000236213	antisense_RNA	Annotated lncRNAs
AC007278.1	chr2	ENSG00000234389	sense_intronic	Annotated lncRNAs
AC007278.2	chr2	ENSG00000236525	sense_intronic	Annotated lncRNAs
AC007336.3	chr16	ENSG00000279030	TEC	Annotated lncRNAs
AC007384.1	chr7	ENSG00000237513	Intergenic	Annotated lncRNAs
AC007386.2	chr2	ENSG00000232613	antisense_RNA	Annotated lncRNAs
AC007569.1	chr12	ENSG00000257221	antisense_RNA	Annotated lncRNAs
AC007637.1	chr12	ENSG00000256139	sense_overlapping	Annotated lncRNAs
AC007686.3	chr14	ENSG00000273729	antisense_RNA	Annotated lncRNAs
AC007728.2	chr16	ENSG00000261644	antisense_RNA	Annotated lncRNAs

AC007728.3	chr16	ENSG00000270120	sense_intronic	Annotated lncRNAs
AC007991.2	chr8	ENSG00000253838	sense_intronic	Annotated lncRNAs
AC008033.3	chr12	ENSG00000273824	Intergenic	Annotated lncRNAs
AC008074.3	chr2	ENSG00000260101	Intergenic	Annotated lncRNAs
AC008083.3	chr12	ENSG00000258181	Intergenic	Annotated lncRNAs
AC008105.3	chr17	ENSG00000267121	antisense_RNA	Annotated lncRNAs
AC008115.3	chr12	ENSG00000275560	sense_intronic	Annotated lncRNAs
AC008467.1	chr5	ENSG00000249476	Intergenic	Annotated lncRNAs
AC008555.5	chr19	ENSG00000271109	Intergenic	Annotated lncRNAs
AC008667.1	chr5	ENSG00000249526	Intergenic	Annotated lncRNAs
AC008691.1	chr5	ENSG00000249738	antisense_RNA	Annotated lncRNAs
AC008760.2	chr19	ENSG00000276980	sense_intronic	Annotated lncRNAs
AC008957.1	chr5	ENSG00000250155	antisense_RNA	Annotated lncRNAs

AC009090.5	chr16	ENSG00000279803	TEC	Annotated lncRNAs
AC009093.10	chr16	ENSG00000284685	processed_transcript	Annotated lncRNAs
AC009093.2	chr16	ENSG00000260517	Intergenic	Annotated lncRNAs
AC009093.7	chr16	ENSG00000279106	TEC	Annotated lncRNAs
AC009126.1	chr5	ENSG00000247121	antisense_RNA	Annotated lncRNAs
AC009133.1	chr16	ENSG00000238045	antisense_RNA	Annotated lncRNAs
AC009948.1	chr2	ENSG00000223960	antisense_RNA	Annotated lncRNAs
AC009951.1	chr2	ENSG00000279166	TEC	Annotated lncRNAs
AC010175.1	chr12	ENSG00000256427	Intergenic	Annotated lncRNAs
AC010247.1	chr19	ENSG00000254887	processed_transcript	Annotated lncRNAs
AC010335.3	chr19	ENSG00000280121	TEC	Annotated lncRNAs
AC010883.1	chr2	ENSG00000234936	antisense_RNA	Annotated lncRNAs
AC011816.2	chr3	ENSG00000272334	Intergenic	Annotated lncRNAs

AC011899.2	chr7	ENSG00000233038	antisense_RNA	Annotated lncRNAs
AC011933.2	chr17	ENSG00000264853	sense_intronic	Annotated lncRNAs
AC011939.3	chr15	ENSG00000275638	antisense_RNA	Annotated lncRNAs
AC012020.1	chr3	ENSG00000279277	TEC	Annotated lncRNAs
AC012363.1	chr2	ENSG00000224789	antisense_RNA	Annotated lncRNAs
AC012368.1	chr2	ENSG00000225889	antisense_RNA	Annotated lncRNAs
AC012645.3	chr16	ENSG00000261416	antisense_RNA	Annotated lncRNAs
AC015726.1	chr17	ENSG00000270091	Intergenic	Annotated lncRNAs
AC015819.1	chr18	ENSG00000273669	Intergenic	Annotated lncRNAs
AC015911.3	chr17	ENSG00000267074	sense_intronic	Annotated lncRNAs
AC016831.1	chr7	ENSG00000226380	Intergenic	Annotated lncRNAs
AC016831.5	chr7	ENSG00000271204	Intergenic	Annotated lncRNAs
AC016957.2	chr12	ENSG00000276136	Intergenic	Annotated lncRNAs

AC017002.1	chr2	ENSG00000224959	Intergenic	Annotated lncRNAs
AC018445.3	chr18	ENSG00000279637	TEC	Annotated lncRNAs
AC018445.4	chr18	ENSG00000279981	TEC	Annotated lncRNAs
AC018607.1	chr8	ENSG00000254325	sense_intronic	Annotated lncRNAs
AC018926.2	chr15	ENSG00000276533	sense_intronic	Annotated lncRNAs
AC020656.2	chr12	ENSG00000274979	Intergenic	Annotated lncRNAs
AC020978.10	chr16	ENSG00000279649	TEC	Annotated lncRNAs
AC020978.7	chr16	ENSG00000262514	sense_intronic	Annotated lncRNAs
AC020978.8	chr16	ENSG00000263276	sense_overlapping	Annotated lncRNAs
AC021016.2	chr2	ENSG00000261338	sense_overlapping	Annotated lncRNAs
AC021188.1	chr2	ENSG00000230747	antisense_RNA	Annotated lncRNAs
AC022706.1	chr17	ENSG00000267364	Intergenic	Annotated lncRNAs
AC022916.1	chr17	ENSG00000266947	antisense_RNA	Annotated lncRNAs

AC023449.2	chr15	ENSG00000274307	sense_intronic	Annotated lncRNAs
AC024084.1	chr7	ENSG00000230649	Intergenic	Annotated lncRNAs
AC025031.4	chr12	ENSG00000275481	Intergenic	Annotated lncRNAs
AC025031.5	chr12	ENSG00000278896	TEC	Annotated lncRNAs
AC025280.1	chr16	ENSG00000260859	sense_intronic	Annotated lncRNAs
AC025524.2	chr8	ENSG00000254119	Intergenic	Annotated lncRNAs
AC025569.1	chr12	ENSG00000258168	antisense_RNA	Annotated lncRNAs
AC025580.2	chr15	ENSG00000259354	antisense_RNA	Annotated lncRNAs
AC025809.1	chr19	ENSG00000267662	antisense_RNA	Annotated lncRNAs
AC025809.2	chr19	ENSG00000278875	TEC	Annotated lncRNAs
AC025857.2	chr8	ENSG00000269899	sense_intronic	Annotated lncRNAs
AC026202.2	chr3	ENSG00000233912	antisense_RNA	Annotated lncRNAs
AC027097.1	chr18	ENSG00000267040	antisense_RNA	Annotated lncRNAs

AC027097.2	chr18	ENSG00000267787	antisense_RNA	Annotated lncRNAs
AC027682.7	chr16	ENSG00000280214	TEC	Annotated lncRNAs
AC027801.1	chr17	ENSG00000204584	antisense_RNA	Annotated lncRNAs
AC027808.2	chr15	ENSG00000259649	antisense_RNA	Annotated lncRNAs
AC034102.6	chr12	ENSG00000258317	antisense_RNA	Annotated lncRNAs
AC034199.1	chr5	ENSG00000250274	Intergenic	Annotated lncRNAs
AC035139.1	chr10	ENSG00000228403	sense_intronic	Annotated lncRNAs
AC036222.2	chr17	ENSG00000266222	sense_intronic	Annotated lncRNAs
AC037459.3	chr8	ENSG00000253200	antisense_RNA	Annotated lncRNAs
AC039056.2	chr15	ENSG00000278493	sense_intronic	Annotated lncRNAs
AC040160.2	chr16	ENSG00000280163	TEC	Annotated lncRNAs
AC046185.3	chr17	ENSG00000279369	TEC	Annotated lncRNAs
AC048341.2	chr12	ENSG00000257354	Intergenic	Annotated lncRNAs

AC048382.5	chr15	ENSG00000275120	antisense_RNA	Annotated lncRNAs
AC048382.6	chr15	ENSG00000276278	antisense_RNA	Annotated lncRNAs
AC051619.6	chr15	ENSG00000259539	3prime_overlapping_ncRNA	Annotated lncRNAs
AC067852.6	chr17	ENSG00000267632	sense_intronic	Annotated lncRNAs
AC068631.1	chr3	ENSG00000197099	antisense_RNA	Annotated lncRNAs
AC068898.2	chr10	ENSG00000251413	sense_intronic	Annotated lncRNAs
AC069208.1	chr12	ENSG00000255864	Intergenic	Annotated lncRNAs
AC069366.2	chr17	ENSG00000265801	sense_intronic	Annotated lncRNAs
AC069528.2	chr3	ENSG00000279320	TEC	Annotated lncRNAs
AC073283.1	chr2	ENSG00000225187	antisense_RNA	Annotated lncRNAs
AC073912.2	chr12	ENSG00000256299	Intergenic	Annotated lncRNAs
AC073957.3	chr7	ENSG00000273151	antisense_RNA	Annotated lncRNAs

AC074032.1	chr12	ENSG00000272368	antisense_RNA	Annotated lncRNAs
AC078883.1	chr2	ENSG00000225205	antisense_RNA	Annotated lncRNAs
AC078883.3	chr2	ENSG00000232788	antisense_RNA	Annotated lncRNAs
AC079015.1	chr8	ENSG00000253988	Intergenic	Annotated lncRNAs
AC079298.3	chr4	ENSG00000280241	antisense_RNA	Annotated lncRNAs
AC079316.2	chr12	ENSG00000279176	TEC	Annotated lncRNAs
AC079336.3	chr17	ENSG00000265337	sense_intronic	Annotated lncRNAs
AC079336.7	chr17	ENSG00000280245	TEC	Annotated lncRNAs
AC079921.2	chr4	ENSG00000249685	Intergenic	Annotated lncRNAs
AC083837.1	chr8	ENSG00000261618	Intergenic	Annotated lncRNAs
AC083862.2	chr7	ENSG00000272941	antisense_RNA	Annotated lncRNAs
AC083949.1	chr2	ENSG00000224875	antisense_RNA	Annotated lncRNAs
AC087672.3	chr8	ENSG00000254288	Intergenic	Annotated lncRNAs

AC087752.3	chr8	ENSG00000253878	sense_intronic	Annotated lncRNAs
AC087893.1	chr12	ENSG00000274598	sense_intronic	Annotated lncRNAs
AC090061.1	chr12	ENSG00000247363	Intergenic	Annotated lncRNAs
AC090136.3	chr8	ENSG00000254102	antisense_RNA	Annotated lncRNAs
AC090204.1	chr8	ENSG00000247134	Intergenic	Annotated lncRNAs
AC090510.2	chr15	ENSG00000274403	antisense_RNA	Annotated lncRNAs
AC090510.3	chr15	ENSG00000278769	antisense_RNA	Annotated lncRNAs
AC090559.1	chr11	ENSG00000255197	antisense_RNA	Annotated lncRNAs
AC090948.1	chr3	ENSG00000271964	antisense_RNA	Annotated lncRNAs
AC090948.2	chr3	ENSG00000272498	antisense_RNA	Annotated lncRNAs
AC090948.3	chr3	ENSG00000272529	antisense_RNA	Annotated lncRNAs
AC091133.1	chr17	ENSG00000230532	antisense_RNA	Annotated lncRNAs
AC092135.1	chr16	ENSG00000279294	TEC	Annotated lncRNAs

AC092135.3	chr16	ENSG00000279841	TEC	Annotated lncRNAs
AC092139.4	chr16	ENSG00000279476	TEC	Annotated lncRNAs
AC092376.2	chr16	ENSG00000277954	antisense_RNA	Annotated lncRNAs
AC092652.3	chr2	ENSG00000283994	sense_intronic	Annotated lncRNAs
AC092868.2	chr15	ENSG00000259735	sense_intronic	Annotated lncRNAs
AC093010.2	chr3	ENSG00000241490	antisense_RNA	Annotated lncRNAs
AC093278.2	chr5	ENSG00000261269	sense_overlapping	Annotated lncRNAs
AC093330.2	chr18	ENSG00000279811	TEC	Annotated lncRNAs
AC093423.2	chr1	ENSG00000230735	sense_intronic	Annotated lncRNAs
AC093525.8	chr16	ENSG00000279520	TEC	Annotated lncRNAs
AC093609.1	chr2	ENSG00000230587	Intergenic	Annotated lncRNAs
AC093726.2	chr7	ENSG00000273183	antisense_RNA	Annotated lncRNAs
AC096667.1	chr2	ENSG00000283839	antisense_RNA	Annotated lncRNAs

AC096921.2	chr3	ENSG00000261468	sense_overlapping	Annotated lncRNAs
AC098613.1	chr3	ENSG00000223552	antisense_RNA	Annotated lncRNAs
AC099063.4	chr1	ENSG00000284734	antisense_RNA	Annotated lncRNAs
AC099343.3	chr4	ENSG00000271646	Intergenic	Annotated lncRNAs
AC099494.2	chr16	ENSG00000279415	TEC	Annotated lncRNAs
AC099521.3	chr16	ENSG00000279693	TEC	Annotated lncRNAs
AC099524.1	chr16	ENSG00000261218	Intergenic	Annotated lncRNAs
AC099804.1	chr17	ENSG00000279570	TEC	Annotated lncRNAs
AC099811.5	chr17	ENSG00000278829	sense_intronic	Annotated lncRNAs
AC100830.1	chr15	ENSG00000259635	antisense_RNA	Annotated lncRNAs
AC100830.3	chr15	ENSG00000265967	antisense_RNA	Annotated lncRNAs
AC104170.1	chr1	ENSG00000227070	antisense_RNA	Annotated lncRNAs
AC104260.2	chr15	ENSG00000280304	TEC	Annotated lncRNAs

AC104564.4	chr17	ENSG00000264290	sense_intronic	Annotated lncRNAs
AC104695.3	chr2	ENSG00000270640	sense_intronic	Annotated lncRNAs
AC104699.1	chr2	ENSG00000224220	antisense_RNA	Annotated lncRNAs
AC105046.1	chr8	ENSG00000261026	sense_overlapping	Annotated lncRNAs
AC105150.1	chr8	ENSG00000254048	sense_intronic	Annotated lncRNAs
AC105749.1	chr3	ENSG00000281100	TEC	Annotated lncRNAs
AC107959.1	chr8	ENSG00000245025	antisense_RNA	Annotated lncRNAs
AC108134.3	chr16	ENSG00000262370	Intergenic	Annotated lncRNAs
AC108463.3	chr2	ENSG00000271590	Intergenic	Annotated lncRNAs
AC110048.2	chr15	ENSG00000277152	Intergenic	Annotated lncRNAs
AC110769.3	chr2	ENSG00000279957	TEC	Annotated lncRNAs
AC111170.2	chr17	ENSG00000267263	sense_intronic	Annotated lncRNAs
AC112496.1	chrX	ENSG00000241886	sense_intronic	Annotated lncRNAs

AC113615.1	chr4	ENSG00000251408	Intergenic	Annotated lncRNAs
AC114490.1	chr1	ENSG00000241014	processed_transcript	Annotated lncRNAs
AC116366.1	chr5	ENSG00000234290	antisense_RNA	Annotated lncRNAs
AC120193.1	chr8	ENSG00000253535	antisense_RNA	Annotated lncRNAs
AC121247.1	chr3	ENSG00000225399	Intergenic	Annotated lncRNAs
AC121761.1	chr12	ENSG00000257497	antisense_RNA	Annotated lncRNAs
AC124319.4	chr17	ENSG00000280248	TEC	Annotated lncRNAs
AC130448.1	chr16	ENSG00000279202	TEC	Annotated lncRNAs
AC130469.2	chr19	ENSG00000279396	TEC	Annotated lncRNAs
AC131210.1	chr3	ENSG00000240888	sense_intronic	Annotated lncRNAs
AC132872.3	chr17	ENSG00000275888	antisense_RNA	Annotated lncRNAs
AC132938.5	chr17	ENSG00000279066	sense_intronic	Annotated lncRNAs
AC133065.2	chr16	ENSG00000262222	antisense_RNA	Annotated lncRNAs

AC133065.4	chr16	ENSG00000263013	sense_intronic	Annotated lncRNAs
AC133065.6	chr16	ENSG00000280153	TEC	Annotated lncRNAs
AC133644.2	chr2	ENSG00000273445	antisense_RNA	Annotated lncRNAs
AC134043.2	chr5	ENSG00000279204	TEC	Annotated lncRNAs
AC136475.1	chr11	ENSG00000251661	antisense_RNA	Annotated lncRNAs
AC138331.1	chr12	ENSG00000258066	Intergenic	Annotated lncRNAs
AC139887.1	chr4	ENSG00000233799	antisense_RNA	Annotated lncRNAs
AC139887.4	chr4	ENSG00000272588	antisense_RNA	Annotated lncRNAs
AC144831.1	chr17	ENSG00000261888	Intergenic	Annotated lncRNAs
AC147651.3	chr7	ENSG00000237181	antisense_RNA	Annotated lncRNAs
AC244034.2	chr1	ENSG00000261000	Intergenic	Annotated lncRNAs
AC244034.3	chr1	ENSG00000279946	TEC	Annotated lncRNAs
ACTA2-AS1	chr10	ENSG00000180139	antisense_RNA	Annotated lncRNAs

ADAMTS9-AS2	chr3	ENSG00000241684	antisense_RNA	Annotated lncRNAs
ADAMTSL4-AS1	chr1	ENSG00000203804	processed_transcript	Annotated lncRNAs
ADPGK-AS1	chr15	ENSG00000260898	antisense_RNA	Annotated lncRNAs
AF111169.4	chr14	ENSG00000259081	antisense_RNA	Annotated lncRNAs
AF127936.1	chr21	ENSG00000226751	Intergenic	Annotated lncRNAs
AF127936.2	chr21	ENSG00000232884	Intergenic	Annotated lncRNAs
AF165147.1	chr21	ENSG00000232855	Intergenic	Annotated lncRNAs
AJ003147.3	chr16	ENSG00000279330	TEC	Annotated lncRNAs
AL008726.1	chr20	ENSG00000271984	antisense_RNA	Annotated lncRNAs
AL021707.1	chr22	ENSG00000225450	antisense_RNA	Annotated lncRNAs
AL021707.3	chr22	ENSG00000230149	antisense_RNA	Annotated lncRNAs
AL021707.5	chr22	ENSG00000244491	antisense_RNA	Annotated lncRNAs
AL021707.6	chr22	ENSG00000272669	antisense_RNA	Annotated lncRNAs

AL021707.8	chr22	ENSG00000273096	sense_intronic	Annotated lncRNAs
AL022067.1	chr6	ENSG00000269919	sense_intronic	Annotated lncRNAs
AL024509.3	chr6	ENSG00000235733	antisense_RNA	Annotated lncRNAs
AL031595.3	chr22	ENSG00000280434	sense_overlapping	Annotated lncRNAs
AL031728.1	chr1	ENSG00000231105	antisense_RNA	Annotated lncRNAs
AL033543.1	chr22	ENSG00000279175	sense_intronic	Annotated lncRNAs
AL035587.1	chr6	ENSG00000231113	antisense_RNA	Annotated lncRNAs
AL035681.1	chr22	ENSG00000235513	antisense_RNA	Annotated lncRNAs
AL049840.4	chr14	ENSG00000269958	sense_intronic	Annotated lncRNAs
AL078604.2	chr6	ENSG00000237927	Intergenic	Annotated lncRNAs
AL096816.1	chr6	ENSG00000280135	TEC	Annotated lncRNAs
AL096855.1	chr1	ENSG00000232912	antisense_RNA	Annotated lncRNAs
AL109741.1	chr1	ENSG00000225938	antisense_RNA	Annotated lncRNAs

AL109806.1	chr20	ENSG00000280387	TEC	Annotated lncRNAs
AL109811.1	chr1	ENSG00000226849	antisense_RNA	Annotated lncRNAs
AL109955.1	chr20	ENSG00000218018	antisense_RNA	Annotated lncRNAs
AL110504.1	chr14	ENSG00000258749	antisense_RNA	Annotated lncRNAs
AL118508.4	chr20	ENSG00000279447	TEC	Annotated lncRNAs
AL121985.1	chr1	ENSG00000228863	antisense_RNA	Annotated lncRNAs
AL133245.1	chr2	ENSG00000272754	Intergenic	Annotated lncRNAs
AL133342.1	chr20	ENSG00000278231	Intergenic	Annotated lncRNAs
AL133355.1	chr10	ENSG00000260461	sense_overlapping	Annotated lncRNAs
AL133384.2	chr10	ENSG00000279796	TEC	Annotated lncRNAs
AL133467.1	chr14	ENSG00000258572	Intergenic	Annotated lncRNAs
AL135818.1	chr14	ENSG00000258875	processed_transcript	Annotated lncRNAs
AL136295.2	chr14	ENSG00000259321	Intergenic	Annotated lncRNAs

AL136295.7	chr14	ENSG00000278784	sense_intronic	Annotated lncRNAs
AL137145.2	chr10	ENSG00000215244	Intergenic	Annotated lncRNAs
AL138895.1	chr9	ENSG00000235119	Intergenic	Annotated lncRNAs
AL139241.1	chr10	ENSG00000230928	antisense_RNA	Annotated lncRNAs
AL157394.1	chr10	ENSG00000261438	sense_overlapping	Annotated lncRNAs
AL157786.1	chr10	ENSG00000232934	antisense_RNA	Annotated lncRNAs
AL158166.1	chr10	ENSG00000227076	sense_intronic	Annotated lncRNAs
AL160272.1	chr9	ENSG00000228512	Intergenic	Annotated lncRNAs
AL160400.1	chr6	ENSG00000272053	Intergenic	Annotated lncRNAs
AL161773.1	chr13	ENSG00000279237	TEC	Annotated lncRNAs
AL162274.2	chr10	ENSG00000277959	antisense_RNA	Annotated lncRNAs
AL162274.3	chr10	ENSG00000279982	TEC	Annotated lncRNAs
AL162414.1	chr9	ENSG00000227531	Intergenic	Annotated lncRNAs

AL162586.1	chr9	ENSG00000225032	antisense_RNA	Annotated lncRNAs
AL162724.1	chr9	ENSG00000227200	sense_intronic	Annotated lncRNAs
AL353593.1	chr1	ENSG00000269890	antisense_RNA	Annotated lncRNAs
AL353593.2	chr1	ENSG00000269934	antisense_RNA	Annotated lncRNAs
AL353597.1	chr6	ENSG00000227920	Intergenic	Annotated lncRNAs
AL353801.2	chr10	ENSG00000234504	antisense_RNA	Annotated lncRNAs
AL355076.2	chr14	ENSG00000258760	Intergenic	Annotated lncRNAs
AL355102.1	chr14	ENSG00000258412	sense_intronic	Annotated lncRNAs
AL355304.1	chr6	ENSG00000232618	Intergenic	Annotated lncRNAs
AL355581.1	chr6	ENSG00000227678	antisense_RNA	Annotated lncRNAs
AL355816.1	chr1	ENSG00000272982	antisense_RNA	Annotated lncRNAs
AL355816.2	chr1	ENSG00000273221	antisense_RNA	Annotated lncRNAs
AL356020.1	chr14	ENSG00000258731	antisense_RNA	Annotated lncRNAs

AL356124.2	chr6	ENSG00000233351	antisense_RNA	Annotated lncRNAs
AL356356.1	chr1	ENSG00000237781	antisense_RNA	Annotated lncRNAs
AL357033.3	chr20	ENSG00000276317	antisense_RNA	Annotated lncRNAs
AL357033.4	chr20	ENSG00000277496	antisense_RNA	Annotated lncRNAs
AL357054.4	chr6	ENSG00000272463	Intergenic	Annotated lncRNAs
AL357060.2	chr6	ENSG00000237499	antisense_RNA	Annotated lncRNAs
AL357078.3	chr1	ENSG00000272506	Intergenic	Annotated lncRNAs
AL357793.1	chr1	ENSG00000226640	Intergenic	Annotated lncRNAs
AL359183.1	chr10	ENSG00000279406	TEC	Annotated lncRNAs
AL359532.1	chr10	ENSG00000272914	sense_intronic	Annotated lncRNAs
AL359962.1	chr1	ENSG00000233411	sense_intronic	Annotated lncRNAs
AL359962.2	chr1	ENSG00000273160	Intergenic	Annotated lncRNAs
AL365273.1	chr10	ENSG00000240527	sense_intronic	Annotated lncRNAs

AL365361.1	chr1	ENSG00000259834	Intergenic	Annotated lncRNAs
AL390066.1	chr1	ENSG00000224950	Intergenic	Annotated lncRNAs
AL390318.1	chr10	ENSG00000279819	TEC	Annotated lncRNAs
AL391425.1	chr10	ENSG00000274461	processed_transcript	Annotated lncRNAs
AL391832.3	chr1	ENSG00000258082	Intergenic	Annotated lncRNAs
AL450344.1	chr6	ENSG00000224029	sense_intronic	Annotated lncRNAs
AL512306.2	chr1	ENSG00000240219	Intergenic	Annotated lncRNAs
AL512306.3	chr1	ENSG00000240710	antisense_RNA	Annotated lncRNAs
AL513365.2	chr1	ENSG00000225891	antisense_RNA	Annotated lncRNAs
AL590764.1	chrX	ENSG00000228427	antisense_RNA	Annotated lncRNAs
AL590999.1	chr6	ENSG00000235033	antisense_RNA	Annotated lncRNAs
AL591623.1	chr1	ENSG00000225300	sense_intronic	Annotated lncRNAs
AL591848.3	chr1	ENSG00000260698	Intergenic	Annotated lncRNAs

AL592078.1	chr1	ENSG00000233875	antisense_RNA	Annotated lncRNAs
AL596244.1	chr9	ENSG00000261534	sense_overlapping	Annotated lncRNAs
AL645568.1	chr1	ENSG00000203739	antisense_RNA	Annotated lncRNAs
AL844908.1	chr21	ENSG00000272825	antisense_RNA	Annotated lncRNAs
ANKRD44-IT1	chr2	ENSG00000236977	sense_intronic	Annotated lncRNAs
AOAH-IT1	chr7	ENSG00000230539	sense_intronic	Annotated lncRNAs
AP000662.2	chr11	ENSG00000254602	sense_overlapping	Annotated lncRNAs
AP000695.3	chr21	ENSG00000279365	TEC	Annotated lncRNAs
AP000763.3	chr11	ENSG00000256448	antisense_RNA	Annotated lncRNAs
AP000892.3	chr11	ENSG00000280143	TEC	Annotated lncRNAs
AP000919.4	chr18	ENSG00000272625	antisense_RNA	Annotated lncRNAs
AP000941.1	chr11	ENSG00000255176	antisense_RNA	Annotated lncRNAs
AP000977.1	chr11	ENSG00000246790	antisense_RNA	Annotated lncRNAs

AP001528.2	chr11	ENSG00000255471	antisense_RNA	Annotated lncRNAs
AP001528.3	chr11	ENSG00000280339	TEC	Annotated lncRNAs
AP001962.1	chr4	ENSG00000279913	TEC	Annotated lncRNAs
AP002807.1	chr11	ENSG00000255031	antisense_RNA	Annotated lncRNAs
AP003354.2	chr8	ENSG00000254281	Intergenic	Annotated lncRNAs
AP003557.1	chr11	ENSG00000255299	Intergenic	Annotated lncRNAs
AP003716.1	chr11	ENSG00000251637	Intergenic	Annotated lncRNAs
AP003717.1	chr11	ENSG00000255847	antisense_RNA	Annotated lncRNAs
ARHGAP22-IT1	chr10	ENSG00000248682	sense_intronic	Annotated lncRNAs
ARHGAP26-IT1	chr5	ENSG00000230789	sense_intronic	Annotated lncRNAs
ASAP1-IT2	chr8	ENSG00000280543	sense_intronic	Annotated lncRNAs
BACH1-IT1	chr21	ENSG00000248476	sense_intronic	Annotated lncRNAs
BACH1-IT2	chr21	ENSG00000228817	Intergenic	Annotated lncRNAs

BFSP2-AS1	chr3	ENSG00000249993	antisense_RNA	Annotated lncRNAs
BISPR	chr19	ENSG00000282851	Intergenic	Annotated lncRNAs
BX324167.2	chr22	ENSG00000279954	antisense_RNA	Annotated lncRNAs
C1orf220	chr1	ENSG00000213057	Intergenic	Annotated lncRNAs
C22orf34	chr22	ENSG00000188511	Intergenic	Annotated lncRNAs
C8orf31	chr8	ENSG00000177335	processed_transcript	Annotated lncRNAs
C9orf139	chr9	ENSG00000180539	antisense_RNA	Annotated lncRNAs
CACNA1C-AS1	chr12	ENSG00000246627	antisense_RNA	Annotated lncRNAs
CACNA1C-IT3	chr12	ENSG00000256721	sense_intronic	Annotated lncRNAs
CADM3-AS1	chr1	ENSG00000225670	antisense_RNA	Annotated lncRNAs
CARD8-AS1	chr19	ENSG00000268001	antisense_RNA	Annotated lncRNAs
CARMN	chr5	ENSG00000249669	Intergenic	Annotated lncRNAs
CHRM3-AS2	chr1	ENSG00000233355	antisense_RNA	Annotated lncRNAs

CTD-2270F17.1	chr5	ENSG00000253647	antisense_RNA	Annotated lncRNAs
CYTOR	chr2	ENSG00000222041	Intergenic	Annotated lncRNAs
DBH-AS1	chr9	ENSG00000225756	antisense_RNA	Annotated lncRNAs
DLGAP1-AS1	chr18	ENSG00000177337	antisense_RNA	Annotated lncRNAs
DLGAP1-AS2	chr18	ENSG00000262001	antisense_RNA	Annotated lncRNAs
DLGAP4-AS1	chr20	ENSG00000232907	antisense_RNA	Annotated lncRNAs
DPYD-IT1	chr1	ENSG00000232542	sense_intronic	Annotated lncRNAs
EMX2OS	chr10	ENSG00000229847	antisense_RNA	Annotated lncRNAs
EPHA1-AS1	chr7	ENSG00000229153	antisense_RNA	Annotated lncRNAs
FAM30A	chr14	ENSG00000226777	Intergenic	Annotated lncRNAs
FLG-AS1	chr1	ENSG00000237975	antisense_RNA	Annotated lncRNAs
FLJ27354	chr1	ENSG00000231999	antisense_RNA	Annotated lncRNAs
FLJ31104	chr5	ENSG00000227908	antisense_RNA	Annotated lncRNAs

FO393401.1	chr20	ENSG00000230155	antisense_RNA	Annotated lncRNAs
FUT8-AS1	chr14	ENSG00000276116	antisense_RNA	Annotated lncRNAs
GHRLOS	chr3	ENSG00000240288	antisense_RNA	Annotated lncRNAs
GK-IT1	chrX	ENSG00000229331	sense_intronic	Annotated lncRNAs
GRK5-IT1	chr10	ENSG00000228485	sense_intronic	Annotated lncRNAs
H19	chr11	ENSG00000130600	processed_transcript	Annotated lncRNAs
HCP5	chr6	ENSG00000206337	sense_overlapping	Annotated lncRNAs
HOXA-AS2	chr7	ENSG00000253552	antisense_RNA	Annotated lncRNAs
IFNG-AS1	chr12	ENSG00000255733	antisense_RNA	Annotated lncRNAs
ITGB2-AS1	chr21	ENSG00000227039	antisense_RNA	Annotated lncRNAs
ITPKB-IT1	chr1	ENSG00000228382	sense_intronic	Annotated lncRNAs
LINC-PINT	chr7	ENSG00000231721	antisense_RNA	Annotated lncRNAs
LINC002481	chr4	ENSG00000246526	Intergenic	Annotated lncRNAs

LINC00299	chr2	ENSG00000236790	Intergenic	Annotated lncRNAs
LINC00324	chr17	ENSG00000178977	Intergenic	Annotated lncRNAs
LINC00402	chr13	ENSG00000235532	Intergenic	Annotated lncRNAs
LINC00426	chr13	ENSG00000238121	Intergenic	Annotated lncRNAs
LINC00487	chr2	ENSG00000205837	Intergenic	Annotated lncRNAs
LINC00494	chr20	ENSG00000235621	Intergenic	Annotated lncRNAs
LINC00528	chr22	ENSG00000269220	Intergenic	Annotated lncRNAs
LINC00539	chr13	ENSG00000224429	Intergenic	Annotated lncRNAs
LINC00598	chr13	ENSG00000215483	Intergenic	Annotated lncRNAs
LINC00649	chr21	ENSG00000237945	antisense_RNA	Annotated lncRNAs
LINC00702	chr10	ENSG00000233117	Intergenic	Annotated lncRNAs
LINC00861	chr8	ENSG00000245164	Intergenic	Annotated lncRNAs
LINC00877	chr3	ENSG00000241163	Intergenic	Annotated lncRNAs

LINC00892	chrX	ENSG00000233093	processed_transcript	Annotated lncRNAs
LINC00896	chr22	ENSG00000236499	Intergenic	Annotated lncRNAs
LINC00921	chr16	ENSG00000281005	Intergenic	Annotated lncRNAs
LINC00926	chr15	ENSG00000247982	Intergenic	Annotated lncRNAs
LINC00937	chr12	ENSG00000226091	Intergenic	Annotated lncRNAs
LINC00944	chr12	ENSG00000256128	Intergenic	Annotated lncRNAs
LINC00954	chr2	ENSG00000228784	Intergenic	Annotated lncRNAs
LINC00987	chr12	ENSG00000237248	Intergenic	Annotated lncRNAs
LINC00996	chr7	ENSG00000242258	Intergenic	Annotated lncRNAs
LINC01050	chr13	ENSG00000271216	Intergenic	Annotated lncRNAs
LINC01094	chr4	ENSG00000251442	Intergenic	Annotated lncRNAs
LINC01126	chr2	ENSG00000279873	Intergenic	Annotated lncRNAs
LINC01146	chr14	ENSG00000258867	Intergenic	Annotated lncRNAs

LINC01160	chr1	ENSG00000231346	Intergenic	Annotated lncRNAs
LINC01197	chr15	ENSG00000248441	Intergenic	Annotated lncRNAs
LINC01215	chr3	ENSG00000271856	Intergenic	Annotated lncRNAs
LINC01281	chrX	ENSG00000235304	Intergenic	Annotated lncRNAs
LINC01336	chr5	ENSG00000250889	Intergenic	Annotated lncRNAs
LINC01358	chr1	ENSG00000237352	Intergenic	Annotated lncRNAs
LINC01422	chr22	ENSG00000235271	Intergenic	Annotated lncRNAs
LINC01422	chr22	ENSG00000223704	Intergenic	Annotated lncRNAs
LINC01484	chr5	ENSG00000253686	Intergenic	Annotated lncRNAs
LINC01550	chr14	ENSG00000246223	Intergenic	Annotated lncRNAs
LINC01588	chr14	ENSG00000214900	Intergenic	Annotated lncRNAs
LINC01619	chr12	ENSG00000257242	processed_transcript	Annotated lncRNAs
LINC01679	chr21	ENSG00000237989	Intergenic	Annotated lncRNAs

LINC01684	chr21	ENSG00000237484	Intergenic	Annotated lncRNAs
LINC01772	chr1	ENSG00000226029	Intergenic	Annotated lncRNAs
LINC01800	chr2	ENSG00000234572	Intergenic	Annotated lncRNAs
LINC01801	chr19	ENSG00000267767	Intergenic	Annotated lncRNAs
LINC01857	chr2	ENSG00000224137	Intergenic	Annotated lncRNAs
LINC01871	chr2	ENSG00000235576	Intergenic	Annotated lncRNAs
LINC01914	chr2	ENSG00000234362	Intergenic	Annotated lncRNAs
LINC01934	chr2	ENSG00000234663	Intergenic	Annotated lncRNAs
LINC01943	chr2	ENSG00000280721	Intergenic	Annotated lncRNAs
LINC02033	chr3	ENSG00000279727	Intergenic	Annotated lncRNAs
LINC02084	chr3	ENSG00000272282	Intergenic	Annotated lncRNAs
LINC02099	chr8	ENSG00000253490	Intergenic	Annotated lncRNAs
LINC02100	chr5	ENSG00000248693	Intergenic	Annotated lncRNAs

LINC02202	chr5	ENSG00000245812	Intergenic	Annotated lncRNAs
LINC02245	chr2	ENSG00000237638	Intergenic	Annotated lncRNAs
LINC02273	chr4	ENSG00000245954	Intergenic	Annotated lncRNAs
LINC02325	chr14	ENSG00000246084	Intergenic	Annotated lncRNAs
LINC02328	chr14	ENSG00000258733	Intergenic	Annotated lncRNAs
LINC02362	chr4	ENSG00000249096	Intergenic	Annotated lncRNAs
LINC02384	chr12	ENSG00000251301	Intergenic	Annotated lncRNAs
LINC02397	chr12	ENSG00000205056	Intergenic	Annotated lncRNAs
LINC02422	chr12	ENSG00000255760	Intergenic	Annotated lncRNAs
LINC02446	chr12	ENSG00000256039	Intergenic	Annotated lncRNAs
LINC02551	chr11	ENSG00000254842	Intergenic	Annotated lncRNAs
LUCAT1	chr5	ENSG00000248323	Intergenic	Annotated lncRNAs
MEG3	chr14	ENSG00000214548	Intergenic	Annotated lncRNAs

MEG9	chr14	ENSG00000223403	Intergenic	Annotated lncRNAs
MIAT	chr22	ENSG00000225783	Intergenic	Annotated lncRNAs
MIR100HG	chr11	ENSG00000255248	processed_transcript	Annotated lncRNAs
MIR155HG	chr21	ENSG00000234883	Intergenic	Annotated lncRNAs
MMP2-AS1	chr16	ENSG00000260135	Intergenic	Annotated lncRNAs
MMP25-AS1	chr16	ENSG00000261971	antisense_RNA	Annotated lncRNAs
MSTRG.10288	chr1	NA	Antisense Intronic	Novel lncRNAs
MSTRG.106451	chr18	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.107081	chr19	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.107232	chr19	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.108144	chr19	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.115529	chr2	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.120715	chr2	NA	Novel_Intergenic	Novel lncRNAs

MSTRG.125572	chr2	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.125734	chr2	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.129182	chr2	NA	Antisense Intronic	Novel lncRNAs
MSTRG.129474	chr2	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.133983	chr20	NA	Antisense Exonic	Novel lncRNAs
MSTRG.13489	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.137226	chr21	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.137994	chr21	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.139361	chr21	NA	Antisense Exonic	Novel lncRNAs
MSTRG.142844	chr22	NA	Antisense Exonic	Novel lncRNAs
MSTRG.142852	chr22	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.143424	chr22	NA	Antisense Intronic	Novel lncRNAs
MSTRG.143603	chr22	NA	Novel_Intergenic	Novel lncRNAs

MSTRG.144049	chr22	NA	Antisense Exonic	Novel lncRNAs
MSTRG.144329	chr3	NA	Antisense Intronic	Novel lncRNAs
MSTRG.145377	chr3	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.145392	chr3	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.147468	chr3	NA	Antisense Intronic	Novel lncRNAs
MSTRG.147886	chr3	NA	Antisense Intronic	Novel lncRNAs
MSTRG.150102	chr3	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.150110	chr3	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.152804	chr3	NA	Antisense Intronic	Novel lncRNAs
MSTRG.154000	chr3	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.159261	chr3	NA	Antisense Intronic	Novel lncRNAs
MSTRG.159809	chr4	NA	Antisense Intronic	Novel lncRNAs
MSTRG.165346	chr4	NA	Novel_Intergenic	Novel lncRNAs

MSTRG.165368	chr4	NA	Antisense Exonic	Novel lncRNAs
MSTRG.165446	chr4	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.165706	chr4	NA	Antisense Intronic	Novel lncRNAs
MSTRG.17316	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.176506	chr4	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.179715	chr5	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.184441	chr5	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.187178	chr5	NA	Antisense Intronic	Novel lncRNAs
MSTRG.18773	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.18794	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.18811	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.188411	chr5	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.188412	chr5	NA	Novel_Intergenic	Novel lncRNAs

MSTRG.190999	chr6	NA	Antisense Exonic	Novel lncRNAs
MSTRG.191655	chr6	NA	Antisense Intronic	Novel lncRNAs
MSTRG.197046	chr6	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.197962	chr6	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.201343	chr6	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.201592	chr6	NA	Antisense Intronic	Novel lncRNAs
MSTRG.202102	chr6	NA	Antisense Exonic	Novel lncRNAs
MSTRG.202570	chr6	NA	Antisense Exonic	Novel lncRNAs
MSTRG.20391	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.210283	chr7	NA	Antisense Intronic	Novel lncRNAs
MSTRG.214349	chr7	NA	Antisense Intronic	Novel lncRNAs
MSTRG.215815	chr8	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.217576	chr8	NA	Antisense Intronic	Novel lncRNAs

MSTRG.219081	chr8	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.222175	chr8	NA	Antisense Exonic	Novel lncRNAs
MSTRG.22230	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.22231	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.22253	chr1	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.224171	chr8	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.226619	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.228278	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.231549	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.231602	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.231604	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.234620	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.235204	chr9	NA	Novel_Intergenic	Novel lncRNAs

MSTRG.237566	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.237683	chr9	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.243701	chrX	NA	Antisense Exonic	Novel lncRNAs
MSTRG.246513	chrX	NA	Antisense Intronic	Novel lncRNAs
MSTRG.24814	chr1	NA	Antisense Exonic	Novel lncRNAs
MSTRG.28873	chr10	NA	Antisense Intronic	Novel lncRNAs
MSTRG.30501	chr10	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.32077	chr10	NA	Antisense Exonic	Novel lncRNAs
MSTRG.32233	chr10	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.32600	chr10	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.32939	chr10	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.34232	chr10	NA	Antisense Exonic	Novel lncRNAs
MSTRG.37798	chr11	NA	Antisense Exonic	Novel lncRNAs

MSTRG.37975	chr10	NA	Antisense Intronic	Novel lncRNAs
MSTRG.41457	chr11	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.45689	chr11	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.47108	chr11	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.47861	chr11	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.47928	chr11	NA	Antisense Intronic	Novel lncRNAs
MSTRG.49951	chr11	NA	Antisense Intronic	Novel lncRNAs
MSTRG.50720	chr12	NA	Antisense Exonic	Novel lncRNAs
MSTRG.51465	chr12	NA	Antisense Intronic	Novel lncRNAs
MSTRG.51467	chr12	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.53744	chr12	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.53753	chr12	NA	Antisense Intronic	Novel lncRNAs
MSTRG.53755	chr12	NA	Antisense Intronic	Novel lncRNAs

MSTRG.53756	chr12	NA	Antisense Intronic	Novel lncRNAs
MSTRG.53757	chr12	NA	Antisense Intronic	Novel lncRNAs
MSTRG.53758	chr12	NA	Antisense Intronic	Novel lncRNAs
MSTRG.55870	chr12	NA	Antisense Exonic	Novel lncRNAs
MSTRG.56715	chr12	NA	Antisense Intronic	Novel lncRNAs
MSTRG.57561	chr12	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.58179	chr12	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.62392	chr13	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.66214	chr13	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.71459	chr14	NA	Antisense Exonic	Novel lncRNAs
MSTRG.73030	chr12	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.75221	chr14	NA	Antisense Intronic	Novel lncRNAs
MSTRG.78432	chr15	NA	Antisense Intronic	Novel lncRNAs

MSTRG.78591	chr15	NA	Antisense Intronic	Novel lncRNAs
MSTRG.81534	chr15	NA	Antisense Intronic	Novel lncRNAs
MSTRG.81610	chr15	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.84843	chr16	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.87407	chr16	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.91976	chr17	NA	Antisense Intronic	Novel lncRNAs
MSTRG.9439	chr1	NA	Antisense Intronic	Novel lncRNAs
MSTRG.95350	chr17	NA	Novel_Intergenic	Novel lncRNAs
MSTRG.95384	chr17	NA	Antisense Intronic	Novel lncRNAs
MSTRG.96856	chr17	NA	Antisense Exonic	Novel lncRNAs
MSTRG.98426	chr18	NA	Antisense Intronic	Novel lncRNAs
MYHAS	chr17	ENSG00000272975	antisense_RNA	Annotated lncRNAs
NPTN-IT1	chr15	ENSG00000281183	sense_intronic	Annotated lncRNAs

NRIR	chr2	ENSG00000225964	antisense_RNA	Annotated lncRNAs
PCED1B-AS1	chr12	ENSG00000247774	processed_transcript	Annotated lncRNAs
PIK3IP1-AS1	chr22	ENSG00000228839	antisense_RNA	Annotated lncRNAs
PPP3CB-AS1	chr10	ENSG00000221817	processed_transcript	Annotated lncRNAs
PRKCQ-AS1	chr10	ENSG00000237943	processed_transcript	Annotated lncRNAs
PSMB8-AS1	chr6	ENSG00000204261	Intergenic	Annotated lncRNAs
PSMD6-AS2	chr3	ENSG00000239653	antisense_RNA	Annotated lncRNAs
RABGAP1L-IT1	chr1	ENSG00000223525	sense_intronic	Annotated lncRNAs
RNF216-IT1	chr7	ENSG00000237738	sense_intronic	Annotated lncRNAs
RORA-AS1	chr15	ENSG00000245534	antisense_RNA	Annotated lncRNAs
RPS6KA2-IT1	chr6	ENSG00000232082	sense_intronic	Annotated lncRNAs
RRN3P2	chr16	ENSG00000103472	processed_transcript	Annotated lncRNAs
SDCBP2-AS1	chr20	ENSG00000234684	antisense_RNA	Annotated lncRNAs

SENCR	chr11	ENSG00000254703	antisense_RNA	Annotated lncRNAs
SERPINB9P1	chr6	ENSG00000230438	Intergenic	Annotated lncRNAs
SMIM25	chr20	ENSG00000224397	Intergenic	Annotated lncRNAs
SPAG5-AS1	chr17	ENSG00000227543	processed_transcript	Annotated lncRNAs
TBX5-AS1	chr12	ENSG00000255399	antisense_RNA	Annotated lncRNAs
TCL6	chr14	ENSG00000187621	processed_transcript	Annotated lncRNAs
TMC3-AS1	chr15	ENSG00000259343	antisense_RNA	Annotated lncRNAs
TNFRSF14-AS1	chr1	ENSG00000238164	antisense_RNA	Annotated lncRNAs
TNK2-AS1	chr3	ENSG00000224614	antisense_RNA	Annotated lncRNAs
TNRC6C-AS1	chr17	ENSG00000204282	processed_transcript	Annotated lncRNAs
TRG-AS1	chr7	ENSG00000281103	antisense_RNA	Annotated lncRNAs
TSPOAP1-AS1	chr17	ENSG00000265148	antisense_RNA	Annotated lncRNAs
U62317.5	chr22	ENSG00000273272	Intergenic	Annotated lncRNAs

U73169.1	chr3	ENSG00000213600	sense_intronic	Annotated lncRNAs
USP30-AS1	chr12	ENSG00000256262	antisense_RNA	Annotated lncRNAs
Z84484.1	chr6	ENSG00000224666	antisense_RNA	Annotated lncRNAs
Z84723.1	chr16	ENSG00000260803	processed_transcript	Annotated lncRNAs
Z95114.3	chr22	ENSG00000279805	TEC	Annotated lncRNAs
Z98044.1	chr1	ENSG00000281937	Intergenic	Annotated lncRNAs
Z99289.2	chr6	ENSG00000270661	Intergenic	Annotated lncRNAs
ZBTB20-AS1	chr3	ENSG00000241560	antisense_RNA	Annotated lncRNAs

Acknowledgements

Firstly, from the bottom of my heart I want to thank Dr. Salvatore Piscuglio for giving me the opportunity to perform my PhD in his group on a very interesting project and for always believing in me. And I'm grateful to Dr. Charlotte Ng for her continuous guidance in bioinformatics and motivating me all the time during the course of my PhD degree. I could not have imagined having better supervisors for my PhD research.

Besides my advisors, I am deeply grateful to the rest of my thesis committee Prof. Luigi M. Terracciano, Prof. Michael N. Hall and Dr. Anamaria Necsulea for their support and encouragement during the committee meetings, also for the questions which made me think outside the box and widen my research from multiple perspectives.

I would like to offer my special thanks to Dr. Mairene Coto for her great support and suggestions during my work. A great mentor like her is always needed. I especially benefited from her guidance before my presentations. Also I would like to extend my sincere thanks to Dr. John Gallon and Dr. Luca Roma for their valuable inputs and helping me out technically during the last year of my research. I also greatly appreciate the support given by Dr. Gaia Bianco during my PhD. They all gave me great scientific advice, shared knowledge and suggestions during the discussions. Thanks a lot for being the people I can always rely on.

I would also like to thank my former colleagues Dr. Viola Paradiso, Dr. Andrea Garoffoli and Dr. Nadia Tosti for the great time we have had in the last four years and for their help when I started working here. I also wish to thank Dr. Manuella Lanzafame, Dr.

Maria Rosaria De Filippo, Dr. Hesam Montazeri, Stephanie Taha, Lana Fourie for their insightful comments and support during the lab meetings.

I also had great pleasure working with Dr. Federica Panebianco, Caner Ercan, Mattia Marinucci, Gina Faye Boot in the group. Many thanks to the members from the other department Dr. Andrej Benjak and Dr. Desiree Schnidrig for their valuable suggestions during the lab meetings.

Thanks also to all the colleagues in the Institute of Pathology for their time and helpful advice for my work during the meetings in the last four years. I really appreciate their great support.

Finally, I would like to express my sincere gratitude to my father Prasada Rao Kancherla and mother Naga Lakshmi Kancherla for their encouragement and supporting me spiritually all through my studies. They made me do what I wanted to do. And I cannot end without appreciating my wife Madhuri Kusumanchi. She has sacrificed many things for me in the last few years and helped me to overcome the depression during the hard times.

vielen dank noch einmal _/_