UNBIASED *PIGGYBAC* MUTAGENESIS SCREENS IDENTIFY TUMORIGENIC AND METASTATIC PATHWAYS IN BREAST CANCER

Inauguraldissertation

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

von

FEDERICA ZILLI

aus Vittorio Veneto, Italien

Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Gerhard M. Christofori

Prof. Dr. Mohamed Bentires-Alj

Prof. Dr. Matthew J. Smalley

Basel, den 13.10.2020

Prof. Dr. Martin Spiess

Dekan

“Whatever decision you have made for your future, you are authorized, and I would say encouraged, to subject it to a continuous examination, ready to change it, if it no longer meets your wishes.” — Rita Levi Montalcini

“Qualunque decisione tu abbia preso per il tuo futuro, sei autorizzato, e direi incoraggiato, a sottoporla ad un continuo esame, pronto a cambiarla, se non risponde più ai tuoi desideri.” — Rita Levi Montalcini

1 | SUMMARY

Breast cancer is the most frequently diagnosed cancer and leading cause of death in female cancer patients worldwide. Metastasis remains the primary cause of solid cancer-evoked mortality. Tumourigenesis and metastatic progression are complex processes, where cancer cells acquire specific mutations or genomic alterations, leave the primary tumour and eventually seed distant metastasis. Despite some progress in understanding breast tumour biology, most of the molecular mechanisms leading to tumour progression and ultimately to metastasis remain undefined.

Activating mutations in *PIK3CA*, which encodes for a catalytic subunit (p110α) of the phosphoinositide 3-kinase (PI3K), are among the most frequent driver alterations in human breast cancer leading to an hyperactivated PI3K pathway signaling. Oncogenic transformation often requires simultaneous and numerous driver oncogenic events. The cancer genome evolves dynamically influenced by the generation of additional mutations and selective forces acting on cancer clones. Not surprisingly, several oncogenic mutations and other genomic alterations often co-occur in cancer cells. This results in tumour heterogeneity, which impacts the clinical outcome of the disease and make their discovery extremely necessary.

In my PhD studies, we focused on the identification of genes involved in breast cancer tumourigenesis and metastasis using a transposon-induced mutagenesis screen. In particular, we performed an unbiased *ex vivo piggyBac* (PB) transposon insertional mutagenesis screen using cancer cells with an activating *PIK3CA*^{H1047R} mutation to identify possible synergistic mechanisms of metastatic colonization. I identified and validated *NFIB* as an inducer of metastatic colonization in breast cancer. Mechanistically and functionally, I demonstrated that, *NFIB* increases the expression of the oxidoreductase *ERO1A* and the growth factor *VEGFA*,

promotes metastasis and shortens overall survival of the animals. Finally, *NFIB-ERO1A-VEGFA* co-expression in basal-like breast cancers correlated with reduced patient survival.

Furthermore, I generated a conditional PB *PIK3CA*^{H1047R} mutant mammary cancer mouse model and performed an *in vivo* PB genome-wide unbiased mutagenesis screen to identify potential synergistic tumorigenic pathways. Serial transplantation of the mutant tumours enriched for tumour initiating cells and targeted sequencing of transposon integration sites revealed almost 7700 common insertion sites (CISs). The most frequently altered genes in the secondary transplanted tumours compared with the primary transplanted tumours may increase tumor initiating ability. My results not only describe potential new targetable pathways of breast cancer tumorigenesis and metastatic colonization, but also highlight the power of genetic screenings for identifying functionally relevant breast cancer driver pathways.

2 | TABLE OF CONTENTS

3 | INTRODUCTION

3.1 Breast cancer

Breast cancer is a heterogeneous disease characterized by cancer cells that proliferate, disseminate, survive and form metastases. It progresses from a premalignant disease to invasive carcinoma and eventually metastasis (Polyak, 2007; Hanahan and Weinberg, 2011). It arises from epithelial cells of the mammary gland as hyperplasia, which is a benign condition. It can progress into atypical ductal hyperplasia, where proliferating cells are arranged in an abnormal way and evolve into ductal carcinoma *in situ* (DCIS), a non-invasive stage, where in the duct appear cancerous cells. During *in situ* carcinomas the myoepithelial cells are epigenetically and phenotypically altered and their number decreases potentially due to the degradation of the basement membrane (Doebar et al., 2019). At the same time, the number of stromal fibroblasts, myofibroblasts, lymphocytes and endothelial cells increases. Loss of myoepithelial cells and basement membrane results in invasive carcinomas, in which tumour cells can invade surrounding tissue and migrate to distant organs and seed, eventually leading to metastasis (**Figure 3-1**) (Massagué and Obenauf, 2016).

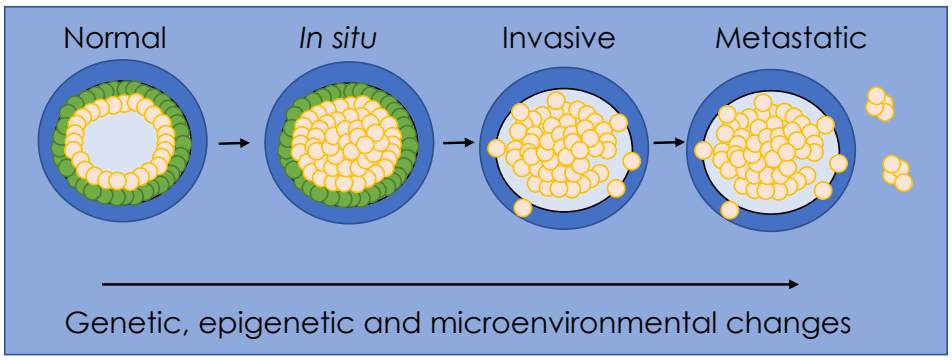


Figure 3-1 | Breast cancer linear progression model

Schematic outline of breast tumour progression. Tumorigenesis goes through defined histologic and clinical stages. It arises from mammary epithelial cells, which progressively accumulate genetic, epigenetic, and microenvironmental alterations.

3.1.1 Breast cancer incidence and classification

Breast cancer is the most frequently diagnosed cancer and leading cause of death in female cancer patients worldwide. It represents the 25% of all cancers diagnosed worldwide (Wyld et al., 2017). The World Health Organization (WHO) measured that breast cancer is impacting 2.1 million women per year, and with 627,000 annual deaths, it represents 15% of all cancer losses among women (Bray et al., 2018). Moreover, lethality is associated with the progression of the disease (metastatic dissemination and colonization) due to resistance to standard treatments (Steeg, 2016).

Breast cancer encompasses a heterogeneous collection of neoplasms with diverse morphologies, molecular phenotypes, responses to therapy, probabilities of relapse and overall survival. Heterogeneity typically exists between the similar type of tumours resulting in subtypes (intertumour heterogeneity). Genome-wide gene expression profiles revealed at least 6 different subclasses and these specific subtypes are characterized by their molecular profiles, morphology, and expression of specific biomarkers (Perou et al., 2000; Sørlie et al., 2001). Breast cancers are then classified as Luminal A (ER and/or PR positive, and HER2 negative); Luminal B (ER and/or PR positive, and HER2 positive); HER2-enriched (ER and PR negative, and HER2 positive) and Basal Like and claudin low, Triple negative breast cancer (TNBC) (ER, PR and HER2 negative) (**Table 3-1**) (Sørlie et al., 2001). TNBCs grow and spread faster than the other types of breast cancer. Proliferation status (Ki67-expression) (Munzone et al., 2012) and clinical parameters such as age, tumour size, lymph node status, histological grade are also taken in consideration (Fragomeni et al., 2018).

Altogether these profiles mostly reflect different clinical prognoses (Perou et al., 2000; Sørlie et al., 2001, 2006; Prat et al., 2010) and responses to therapy (Troester et al. 2004; Prat et al. 2015; Pernas et al. 2019). Patients with luminal A and B, and HER2-enriched subtypes are sensitive to targeted treatments, while patients with triple negative characteristic show poor

prognosis. The status of these markers helps to determine which patients are likely to respond to targeted therapies (i.e., hormonotherapy: tamoxifen or aromatase inhibitors for patients bearing ER+/PR+ and trastuzumab or lapatinib for patients bearing HER2/neu+ tumours) while triple negative patients only have chemotherapy as an alternative.

Molecular subtype	Molecular characteristics	Occurrence	Prognosis
Luminal A	ER positive and/or PR positive. HER-2 negative and low ki67	30-70%	Fairly high survival with low recurrence rate
Luminal B	ER positive and/or PR positive. High ki67 and/or HER-2 positive	10-20%	Fairly high survival but not as Luminal A
HER2-enriched	HER-2 positive. ER and PR negative.	15-20%	Poor prognosis (trastuzumab improved considerably the prognosis)
Basal-like/ claudin-low (TNBC)	ER, PR, HER-2 negative.	10-20%	Poor prognosis

Table 3-1 | Major molecular breast cancer subtypes

Further improvements led to integrative clustering using gene expression and DNA CNAs. This strategy was implemented in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study, where the identification of integrative clusters revealed that there are 10 subtypes (Russnes et al. 2017; Curtis et al. 2012).

3.1.2 Breast cancer driver genes

Comprehensive DNA sequencing have given insights into breast cancer genetics (Yates et al., 2015; Nik-Zainal et al., 2016, Pereira et al., 2016). Moreover these studies also confirmed the importance of driver genes and their mutations, showing association between breast cancer classification and genomic drivers (e.g. *TP53*, *PIK3CA*) (Curtis, C. 2012; Stephens et al. 2012; Pereira et al., 2016; Bailey et al. 2018). Next generation sequencing of DNA from clinical specimens provided insights into breast cancer genetics and contributed to identifying potential

drivers of tumour progression (Yates et al., 2015; Nik-Zainal et al., 2016; Pereira et al., 2016, Robinson et al., 2017, Angus et al., 2019, Bertucci et al. 2019, De Mattos-Arruda et al., 2019).

3.1.3 *PIK3CA* mutations in breast cancer

The PI3K is a family of lipid kinases involved in cell growth, proliferation and is among the most frequently activated pathways in human cancer (Stemke-Hale et al., 2008). Genomic alterations of components of the PI3K pathway are found in over 70% of breast cancers (Saal et al., 2005; Miller et al., 2011). The gene *PIK3CA* encodes the catalytic subunit p110α and its amplification and/or mutation is associated with several kinds of human solid tumours (Bachman et al., 2004; Saal et al., 2005; Karakas et al., 2006;). Of note, activating mutations in the gene *PIK3CA* are found in 40% of ER-positive breast cancers (Banerji et al., 2012; Dey et al., 2017). Major mutations are found on two hot spots, located in the helical domain (exon 9, E542K, E545K) with 33% occurrence, and in the kinase domain (exon 20, H1047R) with 47% occurrence (Bachman et al., 2004; Barbareschi et al., 2007). These mutations lead to a constitutively active enzyme with oncogenic capacity *in vitro* and enhanced tumorigenicity in xenograft models (Zhao et al., 2005; Bader et al., 2006). Alterations in *PIK3CA* are found to occur early in carcinoma development (Miron et al., 2010).

Several mouse model expressing *PIK3CA*^{H1047R} in the mammary gland have been generated by diverse groups (Bader et al., 2006; Adams et al., 2011; Borowsky, 2011; Meyer et al., 2011; Koren and Bentires-Alj, 2013). In our lab we developed a mouse model with whey acid protein (WAP) for conditional mammary specific expression of human *PIK3CA*^{H1047R} to drive Cre recombinase expression. Cre driven by the WAPi-Cre (Wintermantel et al., 2002) results in expression of mutant *PIK3CA H1047R* in alveolar progenitor cells and differentiated secretory luminal cells (Meyer et al., 2011). Finally, we and others have shown that inducible expression of *PIK3CA*^{H1047R} mutation evokes heterogeneous mammary tumours in mice

(Meyer et al., 2011; Koren and Bentires-Alj, 2013; Koren et al., 2015; Van Keymeulen et al., 2015).

3.1.4 Breast cancer heterogeneity

Breast cancers frequently present substantial heterogeneity due to genetic and phenotypic diversity, such as cellular morphology, gene expression, metabolism, motility, and angiogenic, proliferative, immunogenic, and metastatic potential (Marusyk and Polyak, 2010; Hinohara and Polyak, 2019). Heterogeneity typically exists between the similar type of tumours resulting in subtypes (intertumour heterogeneity), or within the tumours of the same type (intratumour heterogeneity) (Visvader, 2011).

Each tumours contains various tumour cell subpopulations and stromal entities, which depending upon their composition can influence survival, therapy responses and global growth of the tumour, posing a major challenge for the clinical management of cancer patients (Koren and Bentires-Alj, 2015). Generation of phenotypic and genetic heterogeneity results from the integration of both genetic and non-genetic factors combined with stochastic events and their delineation is important for diagnostic and therapeutic improvement. Indeed, molecular profiling studies have revealed that heterogeneity is limited not only to cancer cells but also to components of the tumour microenvironment (Hu and Polyak, 2008). Tumour microenvironment can shape tumour cell phenotypes interfering with extraneous elements such as dietary factors, environmental agents, therapy, or diagnosis-induced stress (e.g., biopsy collection) on the tumour cells (Marusyk et al., 2012). Genetic- and microenvironment-mediated epigenetic events can trigger activation and/or prevent the return to quiescence of activated stem cells/progenitor cells, thus trapping the activated cells in a state of continuous renewal. Moreover, heterogeneity in tumour microenvironments has been associated with more invasive phenotypes (Anderson et al., 2006). In conclusion, this intratumour heterogeneity of the microenvironment affects the phenotypic heterogeneity of tumour cells, challenging cancer

therapy and influencing the evolution of cancers by providing different selective and advantageous pressures within the same tumour.

3.1.5 Cancer stem cells and plasticity

The cancer stem cell (CSC) model proposes that only a small subpopulation of tumour cells with stem cell properties drives tumour initiation, progression, recurrence and maintenance because of their indefinite self-renewal capability (Bapat, 2007; Marusyk et al., 2012; Meacham and Morrison, 2013). The term "cancer stem cells" (CSCs) was coined to describe tumorigenic cells that can self-renew (i.e., form tumours when serially passaged at limiting dilutions) and give rise to tumours that display the phenotypic heterogeneity of the parental tumour (Clarke et al., 2006; Nguyen et al., 2012). The CSC model posits that the heterogeneity observed between tumour cells is the result of differences in the "cell of origin" from which they arose. Stem cell variability is often caused by epigenetic changes (Toh et al., 2017), but can also result from clonal evolution of the CSC population where advantageous genetic mutations can accumulate in CSCs and their progeny (Shackleton et al., 2009). Finally, mechanisms occurring during epithelial to mesenchymal transition (EMT) can be associated with cell plasticity (Polyak and Weinberg, 2009; Pastushenko and Blanpain, 2019; Ishay-Ronen and Christofori, 2019; Ishay-Ronen et al., 2019; Nihan Kilinc et al., 2020).

3.1.6 Genetic evolution

The tumorigenesis and progression of breast cancer is an evolutionary process driven by mutations and Darwinian selection (Nowell, 1976), where *de novo* mutations, conferring a fitness and heritable advantage, result in rapid expansion and positive selection of the new clone at the expense of others. Over time, cells can gain additional advantageous mutations in a multistep-manner. Heritable mutations that confer a selective advantage to a clone in its particular microenvironment are referred to as 'drivers', while those that do not immediately

confer a selective advantage are 'passengers' (Yates and Campbell, 2012). The particular combination of driver alterations, together with thousands of passenger mutations and structural rearrangements make each breast cancer unique (Nik-Zainal et al., 2016). In general, high levels of genomic heterogeneity tend to be associated with worse clinical outcomes. In the clonal evolution model, cells acquire mutations that not only give rise to derivatives with different functionalities and behaviour but also serve as a platform for further acquisition of genetic alterations, translating into substantial phenotypic diversity (Kreso and Dick, 2014).

Extensive studies have given comprehensive insights into the diversity of breast cancer genetics (Curtis et al., 2012; Shah et al., 2012, Yates et al., 2015, Nik-Zainal et al., 2016, Pereira et al., 2016, De Mattos-Arruda et al., 2019), showing that the dynamics of cancer evolution is much higher than a natural evolution through multiple linear steps. One of the disadvantages of this model is that it ignores non-genetic variability and does not take into consideration the interactions among clones within the tumour ecosystem.

It is important to remark that the concepts of genetic evolution and CSC are complementary rather than mutually exclusive (**Figure 3-2**), revealing a unified model in which CSCs may evolve and change in frequency under the influence of clonal genetic evolution during tumour progression (Kreso and Dick, 2014).

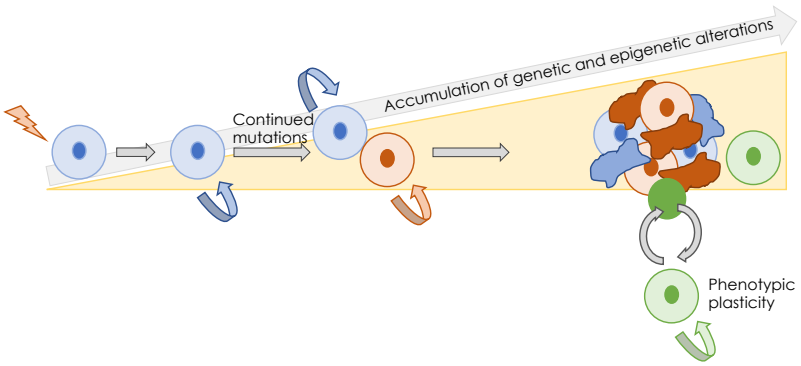


Figure 3-2 | Scheme of the unified model of clonal evolution and cancer stem cells

In this representation, the originating CSC that sustained the first oncogenic mutation gives rise to subclones with self-renewal capabilities that accumulate epigenetic and genetic changes over time. Each different CSC subclone gives rise to intermediate transit-amplifying progenitors that lack self-renewal capabilities. A subset of these progenitors (shown in green) follows a model of tumour cell plasticity and bidirectional conversion between non-CSC to CSC states. This phenotypic change is modulated by microenvironmental stimuli which confer CSC self-renewal capacities to the differentiated cell.

3.2 The metastatic cascade

In most of the cases, breast cancers can progress to metastasis, which remains the primary cause of solid cancer-evoked mortality (Gupta and Massagué, 2006). Breast cancer metastases are found in bone, lung, liver and brain (Chen et al., 2018). The fact that cancer patients may develop metastases years or even decades after the diagnosis of the primary tumour shows that the metastatic process is more complex than our current understanding. Metastasis is responsible of 90% of cancer-associated mortality and it remains the most poorly understood component of cancer pathogenesis (Gupta and Massagué, 2006).

The metastatic process consists of multiple and specific steps that includes local tumour cell invasion, entry into the vasculature followed by the exit of carcinoma cells from the circulation, survival and colonization at the distal sites (Pantel and Brakenhoff, 2004, Massagué and Obenauf, 2016). The process involves a complex interplay between intrinsic tumour cell properties as well as interactions between cancer cells and multiple microenvironments (Ungefroren et al., 2011). To successfully disseminate, metastatic cells acquire properties in addition to those necessary to become neoplastic. Circulating tumour cells (CTCs), both single cells or clusters, enter the blood or lymphatic vessels, survive anoikis while they are detached from the tumour mass and in circulation, exit the blood or lymphatic vessels at a distant organ. However, metastasis is a low-probability, process in which most of these invasive cancer cells perish and only a small proportion manages to infiltrate distant organs and survives as disseminated seeds (Valastyan and Weinberg, 2011). They may not form overt metastases

(Luzzi et al., 1998; Giancotti, 2013), and be blocked as micrometastases in a process termed metastatic dormancy (Sosa et al., 2014). Some cells adapt and reprogram the surrounding stroma, and divide to form macrometastases, a process called colonization (**Figure 3-3**) (D. X. Nguyen, Bos, and Massagué 2009; Faltas 2012).

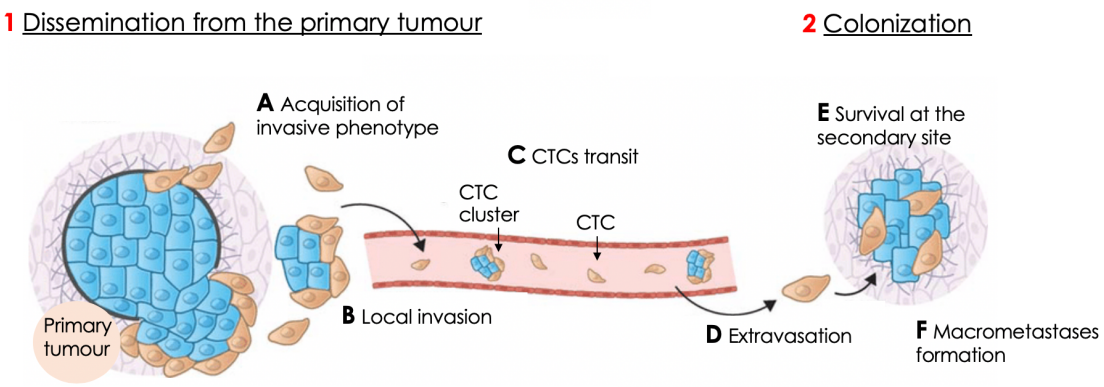


Figure 3-3 | The metastatic cascade

Schematic view of the metastatic cascade. (A) Cancer cells within the primary tumour acquire an invasive phenotype. (B) They can then invade into the surrounding matrix and toward blood vessels, where they intravasate to enter the circulation. (C) In the circulation CTCs survive anoikis. (D) At the distant organ, CTCs exit the circulation and invade into the microenvironment of the foreign tissue. (E) Cancer cells must be able to evade the innate immune response and also survive as a single cell/cluster of cells. (F) To develop into overt macrometastases (colonization step) deposit, the cancer cell must adapt to the microenvironment and initiate proliferation.

3.2.1 Dissemination to distant organs

Cancer cells can migrate away from the primary tumour site and invade neighbouring tissue penetrating the basement membrane (Nguyen et al., 2009). As a primary tumour grows, it needs to develop a blood supply that can support its metabolic needs, a process called angiogenesis (Wyckoff et al., 2000). The onset of neovascularization in primary tumours can be relatively rapid, described as the "angiogenic switch" (Hanahan and Folkman, 1996). The extracellular matrix provides structural and biochemical support and invading tumour cells undergo cytoskeletal rearrangements (Hall, 2009), alter their adhesion properties and degrade the

protein matrix using secretion of extracellular matrix degrading metalloproteinases (MMPs) and cathepsins (Kessenbrock et al., 2010; Clark and Vignjevic, 2015).

Changes in cell phenotype between the epithelial and mesenchymal states defined as the epithelial to mesenchymal transition (EMT; pre-invasion) and mesenchymal to epithelial transition (MET; re-invasion) occur during metastasis (Polyak and Weinberg, 2009; Brabletz et al., 2018; Zhang and Weinberg, 2018). EMT is a physiological process characterized by loss of cell-cell adhesion and cytoskeletal alterations, leading to acquisition of invasive and migratory properties (Nieto et al., 2016). In cancer cells, EMT can promote cell entry into the vasculature, known as intravasation, metastasis formation and support the induction of a stem-cell phenotype (Tiwari et al., 2013; Chaffer et al., 2016; Diepenbruck et al., 2017). Recently, it has become broadly understood that the EMT program is a spectrum of transitional stages between the epithelial and mesenchymal phenotypes, in contrast to a progression that involves a binary choice between full-epithelial and full-mesenchymal phenotypes (Pastushenko et al., 2018; Pastushenko and Blanpain, 2019).

The transition from one state to another is governed by a number of growth factors and signalling pathways and brought evidence that this program is essential for tumour cells to circumvent apoptosis, anoikis, oncogene addiction, and cellular senescence and to escape immune surveillance (Tiwari et al. 2012; Lamouille, Xu, and Derynck 2014; Diepenbruck and Christofori, 2016). However, even if there is abundant evidence that the EMT is associated with motility of multiple cell types, some studies suggested that EMT is not essential to complete the metastatic cascade for in breast and prostate cancer (Fischer et al., 2015; Zheng et al., 2015). This does not diminish the important role for EMT in certain tumours and metastasis (Mani et al., 2008), but because of the complexity of the process, the contribution of EMT to metastasis formation is still under debate (Brabletz et al., 2018).

The new blood vessels can also provide an escape route by which cells can leave the tumour and enter into the body's circulatory blood system. Then cancer cells enter the blood or lymphatic vessels; circulating tumour cells (CTCs) (mostly single cells, less frequently cluster of cells) survive anoikis while they are detached from the tumour mass and in circulation (Aceto et al., 2014; Aceto et al., 2015; Gkountela et al., 2019; Szczerba et al., 2019)

The bloodstream is a very harsh environment and lots of cancer cells die on the way to the secondary site, indicated by the low number of intravasated cells that eventually form metastasis (~0.01%) (Chambers et al., 2002, Quail and Joyce, 2013). Some CTCs reach the distant organ with the possibility to form or not a tumourigenic macro-deposit.

3.2.2 Colonization

CTCs exit the blood or lymphatic vessels at a distant organ, form micrometastatic nodule and adapt and reprogram the surrounding stroma, forming macrometastases (Nguyen et al., 2009). The mesenchymal to epithelial transition (MET) is thought to facilitate colonization (Ocaña et al., 2012; Stankic et al., 2013; Tam and Weinberg, 2013; Li et al., 2015). The growth of a metastatic colony represents the final and most deadly phase in the malignant progression of a tumour and it is triggered by individual properties (*e.g.,* adaptive programs) (Welch and Hurst, 2019) and most importantly by the composition of their microenvironment. The establishment of a supportive metastatic environment occurs prior to the arrival of any carcinoma cell, the so-called premetastatic niche (Hanahan and Weinberg, 2011). Communication with the tumour microenvironment allows invading cancer cells to overcome stromal challenges, settle, and colonize (Oskarsson et al., 2014). Once the cancer cells formed micrometastases, the elevated oxygen and nutrient consumption for continued tumour growth is covered by the attraction of blood vessels leading to detectable metastatic lesions, also called macrometastases. Additionally, metastases can subsequently metastasize, a process described as the shower

metastasis (Weinberg, 2008; Valastyan and Weinberg, 2011). Once a tumour cell colonizes a secondary site, genetic instability inherent in neoplastic cells continues to operate at each cell division, suggesting a continuous and dynamic dissemination and seeding of metastasis (Gundem et al., 2015). Current therapies rarely cure metastatic disease because of the development of drug-resistant metastases, which remain the biggest challenge in oncology.

3.3 Models of metastatic progression and identification of metastatic driver genes

Identification of driver genes of metastatic progression is essential, because metastases are fatal. Several proteins have been reported to contribute to metastasis including its initiation, progression and colonization (Robinson et al., 2017, Krøigård et al. 2018; Angus et al., 2019, Bertucci et al. 2019, De Mattos-Arruda et al., 2019; Koedoot et al., 2019; Priestley et al., 2019). Yet the timing of the occurrence of the metastasis-enabling genomic alterations and the degree of genomic concordance between primary tumours and its metastases remain controversial (Kreike et al. 2007; Koscielny and Tubiana 2010; Stoecklein and Klein 2010).

The linear progression model proposes that the genotype of the metastasis founder cells is highly similar to the primary tumour and the malignant cells pass through multiple successive rounds of genetic changes and selection within the primary tumour microenvironment, before tumour cell dissemination successfully results in a metastatic lesion. From this perspective, metastases are seeded by the most advanced and aggressive clone that should also dominate the primary tumour (Marusyk et al., 2012). In this case, metastasis may have dual effects providing both a local advantage for malignant progression in the primary site and a distal advantage for colonizing a secondary organ (Minn et al., 2005). In recent studies, DNA copy number and somatic mutation patterns, as wells as chromosomal rearrangements were found to be highly similar across primary tumours and matched metastases (Meric-Bernstam et al.,

2014; Tang et al. 2015; Hoadley et al. 2016). These results suggest that most metastatic drivers are established in the primary tumour, despite the substantial heterogeneity seen in the metastases (Siegel et al. 2018). Once metastatic cells reached a secondary site, they might benefit from the niche characteristics that can be in some situation similar to their original niche. As an example, aside from primary tumour ecosystem maintenance, tumour angiogenesis enables tumour cell invasion and dissemination and favours the creation of new secondary tumour ecosystems at metastatic sites (Paran and Paran, 1996; Zuazo-Gaztelu and Casanovas, 2018).

On the contrary, the parallel progression model proposes parallel, independent progression of metastases arising from early disseminated tumour cells and predicts greater disparity between the primary tumour and metastatic lesions. The model emphasizes independent accumulation of genetic and epigenetic alterations as the metastasis is subject to site-specific selection pressures. Indeed, in some cases oncogenic transformation is not sufficient for overt metastatic growth, as shown by the fact that many oncogene-driven mouse models of cancer do not automatically establish distant metastases (*e.g.* *PIK3CA*^{H1047R}) (Koren and Bentires-Alj, 2013) or the observation that some patients have disseminated cancer cells but do not develop clinical metastasis (Kang and Pantel, 2013). Other studies showed that metastatic cell clones clearly do not represent the whole primary tumour population, but only parts of it (Ding et al., 2010; Navin et al., 2011; Schrijver et al. 2018; Iwamoto et al. 2019) suggesting that transformed cells must consequently acquire additional advantageous mutations to gain metastatic ability.

These two models are not mutually exclusive as it has been observed that the genome of late disseminated cells is similar to the genome of the primary tumour at the time of first diagnosis and that simultaneously, in some patients the most distant metastases have acquired additional driver mutations not seen in the primary (Krøigård et al., 2015; Yates et al., 2017;

Krøigård et al., 2017). It is important to stress that different selection processes act on the primary tumor and metastatic cells showing that biopsy and sequencing of metastases may be helpful in providing invaluable mechanistic insight into the biology underlying metastatic progression with the potential to identify novel, potentially druggable, drivers of progression and ultimately improve the treatment of patients with metastatic breast cancer (Hinohara and Polyak, 2019; Yang et al., 2018a).

3.3.1 Nuclear factor I transcriptional factors

The nuclear factor I (NFI) is a family of transcription factors that were found to be essential for adenovirus replication (Gronostajski 2000, Harris et al. 2015). The four members in vertebrates, NFIA, NFIB, NFIC, and NFIX, can bind as hetero and homodimers to the TTGGC(N5) GCCAA symmetric consensus sequence and have a homologous DNA binding domain, resulting in either activation or repression of transcription depending on the context (Gronostajski, 2000). They are expressed in various mammalian tissues in partially overlapping patterns and are important during normal development. NFI TFs promote proliferation and differentiation during the development of multiple organ systems (Becker-Santos et al., 2017), including the central nervous system (Piper et al., 2009, 2010, 2014; Steele-Perkins et al., 2005), mammary gland (Murtagh et al., 2003), and lungs (Gründer et al., 2002). These TFs are also required haematopoiesis (Starnes et al., 2010), osteoblastosis (Pérez-Casellas et al., 2009) and melanocytosis (Chang et al. 2013; Fane et al. 2017). During development NFI TFs promote differentiation at the expense of stem cell self-renewal (Harris et al., 2015), reversely in adult tissues NFI factors promotes the survival and maintenance of slow-cycling adult stem cell populations rather than their differentiation (Chang et al., 2013; Holmfeldt et al., 2013; Martynoga et al., 2013). In support of their role on adult stem cells, recent studies describe the effects of NFI TFs in enforcing tissue homeostasis, functioning as safeguard for the stem cell epigenome preventing irreversible tissue degeneration (Adam et al., 2020).

Emerging evidence has gradually shown NFI expression in various cancers (Chen et al., 2017; Li et al., 2020), highlighting their involvement as potential therapeutic targets or prognostic biomarkers in some cancers.

3.3.2 NFIB transcriptional factor and cancer

Overexpression of NFIB has been implicated in a range of malignancies (**Table 3-2**) (Becker-Santos et al., 2017). The chromosomal region encoding NFIB is amplified in patients with TNBC and overexpression or amplification of NFIB is associated with poor prognosis (Kreike et al. 2007; Moon et al. 2011; Liu et al., 2019a). Moreover, NFIB was shown to enhance TNBC cell survival and progression by suppressing *CDKN2A* (Liu et al., 2019a), and to confer estrogen independency in estrogen receptor-positive breast cancer models (Campbell et al., 2018). Finally, NFIB increases Keratin 14 and 18 positive cell numbers, which possess increased tumorigenic ability (Granit et al., 2018).

NFIB is amplified in ~15% of primary human SCLC (Dooley et al., 2011) and drives tumour initiation and progression in models of SCLC (Denny et al., 2016; Semenova et al., 2016; Wu et al., 2016). Mechanistically, NFIB has been proposed to increase metastasis in SCLC, through the reconfiguration of chromatin accessibility. In particular, chromatin in metastatic lesions displays a widespread increase in accessibility at gene distal regions that were enriched for NFI motifs, and NFIB is specifically associated with the newly open chromatin sites (Denny et al., 2016).

NFIB has also been shown to promote a highly invasive and migratory phenotype in melanoma, where it directly promotes *EZH2* expression, leading to changes in the chromatin state that facilitates this aggressive behaviour (Fane et al., 2017). Furthermore, NFIB operates through the PI3K signalling pathway to induce aggressive gastric cancer (Wu et al., 2018). Additionally, it promotes colorectal cancer cell proliferation and epithelial-mesenchymal

transition (Liu et al., 2019b). NFIB amplifications within squamous cell carcinoma of the esophagus (Yang et al. 2001), large cell neuroendocrine carcinoma of the submandibular gland (Andreasen et al., 2016), and metastatic giant cell tumour of the bone (Quattrini et al., 2015) have also been reported. *NFIB* has been linked to other malignancies through gene fusions, which is frequently the case in adenoid cystic carcinomas (tumours that most commonly arise from salivary and lachrymal glands, although they can also occur in other tissues containing secretory glands such as breast, cervix and vulva) (Persson et al., 2009; Marchiò et al., 2010; Brayer et al., 2016; Xing et al., 2017; Magers et al., 2019).

Type of aberration	Organ/Site	Tumour type	References
Amplification/ overexpression	Lung	Small cell lung cancer (SCLC)	(Denny et al., 2016) (Semenova et al. 2016) (Wu et al., 2016) (Dooley et al., 2011)
Overexpression	Skin	Melanoma	(Fane et al., 2017)
Amplification/ overexpression	Breast	TNBC, ER negative	(Moon et al. 2011) (Liu et al., 2019a) (Campbell et al., 2018)
Amplification	Esophagus	Squamous cell carcinoma	(Yang et al. 2001)
Amplification	Submandibular gland	Large cell neuroendocrine carcinoma	(Andreasen et al., 2016)
Amplification	Bone	Metastatic giant cell tumour	(Quattrini et al., 2015)
Overexpression	Colon/Rectum	Colorectal cancer (CRC)	(Liu et al., 2019b)
Overexpression	Stomach	Gastric cancer	(Wu et al., 2018)
Gene fusions	Salivary, lacrimal and ceruminous glands; breast; vulva	Adenoid cystic carcinomas	(Brayer et al., 2016) (Magers et al., 2019) (Marchiò et al., 2010) (Persson et al., 2009) (Xing et al., 2017)

Table 3-1 | Summary of NFIB alterations reported in cancer.

3.3.3 Therapeutic strategies for patients with metastatic breast cancer

Presently, metastatic breast cancer is largely an incurable disease and treatment is focused on therapies designed to prolong a patient's life and to palliate symptoms associated with the disease. Clearly, treatment design cannot be approached with a "one size fits all" protocol, instead, a personalized approach is required (Low et al., 2018; Sachs et al., 2018). This is particularly true when considering that the genomic profile of this malignancy varies from patient to patient, from stage to stage and sometimes from cell to cell (Koren and Bentires-Alj, 2015). While there has been an increase in the understanding of what drives this disease process at the molecular level (Yu et al., 2016; Robinson et al., 2017), with a number of possible therapeutic targets/treatments emerging from that understanding, chemotherapy is still the main treatment option (O'Shaughnessy, 2005; Zeichner et al., 2016).

Unfortunately, chemotherapy was initially designed with the outdated understanding that all cancer cells are created equal and hence equally susceptible to therapies targeting cellular proliferation. In light of research characterizing the heterogeneity of breast and other cancers, particularly the paradigm shift that has occurred with the cancer stem cell hypothesis, cytotoxic therapies need to be delivered in conjunction with additional therapies for a multimodal approach that targets multiple pathways that are critical to the development and progression of these malignancies (Nedeljković and Damjanović, 2019). Many candidate pathways can be exploited in breast cancer patients depending on their unique tumour microenvironment and genetic profile (Joyce, 2005; Place et al., 2011; Nwabo Kamdje et al., 2014; Flister and Bergom, 2018).

While these candidate drugs may all have promising benefits to extend survivorship, realistically, the efficacy of these additional inhibitors depends on the unique profile of the disease and whether the cells express the molecular targets upon which candidate inhibitors can act (McDonald et al., 2016). Furthermore, it has been well established that the tumour

microenvironment or the tumour itself has mechanisms to compensate for the inhibition of one pathway by upregulating others (Valkenburg et al., 2018). Dormancy has also been studied as a potential target of metastatic colonization and therapies that may help sustain the dormant state have been proposed (Marshall et al., 2012).

Overall, metastasis is a complex challenge that requires more than one therapeutic agent for effective inhibition. Therefore, embracing the multimodal/combination therapy model and targeting multiple pathways simultaneously seems to be key to countering the significant genomic and phenotypic alterations presented by metastatic cancer cells (Fares et al., 2020).

3.4 Transposon insertional mutagenesis systems

Eukaryotic genomes contain an abundance of repeated DNA, and some repeated sequences are mobile (Biscotti et al., 2015). DNA transposons are mobile genetic elements that are able of self-directed excision and subsequent reintegration within the host genome. From bacteria to humans, transposable elements have accumulated over time and continue to shape genomes through their mobilization (Kidwell, 2002). In nature these elements (Class II) exist as single units containing the transposase gene flanked by terminal inverted repeats (TIRs) that carry transposase binding sites. Genomic transposition is mediated by an enzyme called transposase, which recognizes the inverted repeats at the ends of the transposon and also recognizes the target sequence, in which it makes a double-strand break and inserts the transposon in a new genomic location (Kaufman and Rio, 1992; Craig, 1997). This is a “cut-and-paste” mechanism that excises the transposon from its original genomic location and inserts it into a new locus **(Figure 3-4)**, (Fischer et al., 2001).

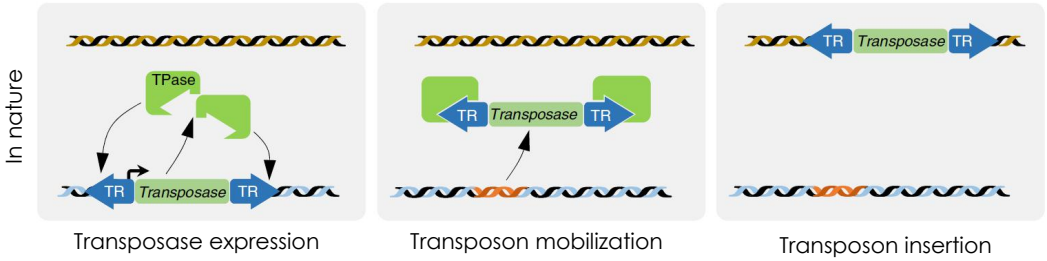


Figure 3-4 | Natural autonomous transposons

In nature, DNA transposons encode and express their own transposase, which then recognizes the flanking terminal repeats (TR), excises the transposon and integrates it into a new locus (Adapted from Mathias J F et al., Nature Genetics 2017).

Transposable elements are involved in a wide variety of biological transactions including genome alteration by element insertion and deletion and by homologous recombination between element copies, viral integration and replication, and the dispersal of a variety of determinants, most notably antibiotic resistance genes (Babakhani and Oloomi, 2018). Thus, transposable elements contributed to genomes evolution (Kidwell, 2002).

3.4.1 Transposon systems as a genetic tool

In invertebrates, transposable elements have long been used for transgenesis and insertional mutagenesis (Johnston, 2002; Thibault et al., 2004). In higher organisms, however, transposons were inactivated millions of years ago and were hence not available as genetic tools until their reconstruction. DNA transposon offers an efficient non-viral method of permanently modifying the genome of mammalian cells. For most applications, it is possible to use transposons as bi-component systems, in which virtually any DNA sequence of interest can be placed between the transposon TIRs and mobilized by *trans*-supplementing the transposase in the form of an expression plasmid **(Figure 3-5)** or mRNA synthesized *in vitro*. In the transposition process, the transposase mediates the excision of the element from its donor

plasmid, followed by reintegration of the transposon into a variety of chromosomal loci (Li et al., 2013).

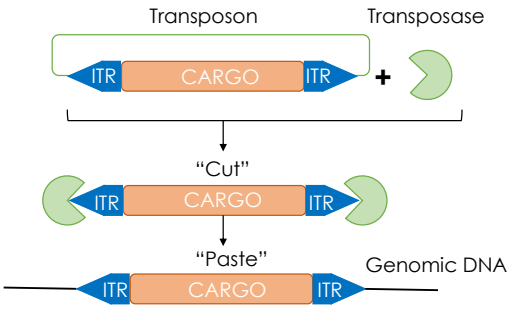


Figure 3-5 | Transposon cut and paste mechanism

The target cargo is inserted into the transposon vector (cargo) between two inverted terminal repeat sequences (ITRs), which are transposon-specific sequences located on both ends of the vector. During transposition, the PB transposase recognizes ITRs and cut the target sequences, thereby moving the contents to the target sites and integrating them into chromosomal sites randomly dispersed in the genome.

Co-transfection of the two vectors permits an efficient "cut and paste" transfer of the transposon from the donor plasmid into the genome (Wilson et al., 2007). Such transposon systems can be effectively used for the generation of recombinant cell lines, transposon-mediated insertional mutagenesis screens in mice and as a therapy for genetic disorders in humans (Ivics and Izsvák, 2010).

3.4.2 *piggyBac* transposon system

The *piggyBac* (PB) transposon was originally isolated from the cabbage looper moth *Trichoplusiani*, in the 1980s (Cary et al., 1989). In general, all members of the PB superfamily recognize a short TTAA motif in the host genome for insertion and excises the transposon without mutation (seamless excision). PB transposition does not require DNA synthesis during the event (Ding et al., 2005; Mitra et al., 2008). PB preferentially integrates in genic regions, especially regions that contain expressed genes. Comparisons of integration profiles of the

transposon systems in human cells showed that PB integrates in intragenic regions and near transcriptional start sites (Meir et al. 2013). PB transposition is unaffected by cargo sizes up to 9.1 kb, showing some reduction in efficiency at 14.3 kb (**Table 3-3**) (Ding et al., 2005). Overproduction inhibition (OPI) is a phenomenon that affects most transposable elements where increasing cellular transposase concentration beyond a certain threshold decreases transposition efficiency, though different studies have described both the presence and lack of OPI for PB (Wilson et al., 2007; Wu et al., 2006).

In mammalian cells, PB exhibits approximately either low (approximately 10%) or no reported local hopping (i.e., the propensity of reintegrating near the original excision site), and had a narrow local hopping distribution of approximately 100 kb from the donor site (Liang et al., 2009; Wang et al., 2008). The rest of the insertion events seem random and evenly distributed across the genome. PB has been shown to transpose efficiently in such different organisms as protozoa, planaria, insects and mammals (Momose et al., 2003; Balu et al., 2005; Wilson et al., 2007; Lukacsovich et al., 2008; Yusa et al., 2009).

3.4.3 *Sleeping beauty* (SB) transposon system

The SB transposon was reconstructed from inactive copies of Tc1/mariner like elements found in a fish genome (Plasterk and Izsva, 1997). SB systems preferentially target their integration into TA dinucleotides leaving a footprint mutation at the donor locus. Transposition occurred randomly across the genome, with low propensity for transcriptionally active regions (Meir et al. 2013) and this makes SB the right choice for human gene therapy protocols which require vectors showing the least preference for target genes. SB undergoes local hopping so it tends to insert in the vicinity of the donor locus.

Currently, the SB system seems to be the most limited with 50 % reduction in transposition efficiency with cargo size of 6 kB compared with 2 kb (**Table 3-3**) (Geurts et al.,

2003). The phenomenon of OPI in SB elements is variable depending on the transposon from this family (Wilson et al., 2007). A hyperactive SB transposase, SB100X, with 100-fold enhancement in efficiency compared with first-generation transposase, was recently identified and shown as effective as *piggyBac* under nonrestrictive conditions (Ivics et al., 2009).

The SB system has been extensively used as a tool for genetic modifications in somatic tissues of a wide range of vertebrate species including humans and the germline of fish, frogs, mice, and rats (Score et al., 2006; Mátés et al., 2007; Huang et al., 2010).

Transposon	Origin	Target	Integration site preference	Capacity for cargo	Local hopping	Footprint
piggyBac (Superfamily *piggyBac*)	*Trichopl* *usiani*	TTAA	Transcription units (introns)	>9Kb	Low	No
Sleeping Beauty (Superfamily *Tc1/mariner*)	Salmon species (reconstr ucted)	TA	Intergenic regions	>10Kb efficiency decrease with size	High	Yes

Table 3-3 │Characteristics of DNA transposon systems used in genomics

3.5 Transposon insertional mutagenesis for cancer gene discovery

Insertional mutagenesis studies can highlight cancer genes or common pathways that are disrupted at low frequency or by processes not immediately obvious from the genome sequence alone. Insertional mutagenesis is a powerful way to elucidate genes involved in a variety of pathways in cultured cells. Moreover, insertional mutagenesis screens provide a functional readout to complement sequencing studies, as genes identified by insertional mutagens are likely to represent both functionally important and evolutionarily conserved cancer genes.

3.5.1 Transposon insertional mutagenesis *in vitro*

PB offers a versatile plasmid-based system for stable cell line generation, especially for generation of stably expressing one or more recombinant gene sequences for biomedical

research, biotechnology or drug development (**Table 3-4**). PB transposase with a second plasmid containing the PB transposon is as an efficient way of transducing mouse embryonic stem (ES) cells, with approximately 28% of cells that survive electroporation transduced (Wang et al., 2008). This allowed the genetic manipulation of Bloom syndrome, RecQ genomic transposition in mESCs (Wang et al., 2009) and inducible genomic transposition in mESCs for gene trapping (Kong et al., 2010). PB system has been further functionally enhanced with a coding sequence (CDS) that provides a level of transposition significantly higher than the native PB system (Bradley, 2007). Multiplexed PB transposon delivery was developed for stable gene transfer in human cells (Kahlig et al., 2009). Furthermore, an inducible version was used for modulating gene expression in myogenic differentiation of human iPSCs (hiPSCs), including a human myopathy (Tanaka et al., 2013) and to reliably incorporate bacterial artificial chromosome (BAC) transgene in human ESCs (Rostovskaya et al., 2012). Precise excision of PB has been used to generate transgene-free iPSCs, whereby transgenes are delivered for reprogramming (Kaji et al., 2009; Yusa et al., 2009). *In vitro*, PB has been used for gene discovery in mouse neural stem cells (Albieri et al., 2010) and human pancreatic cells (You et al., 2011). Moreover, PB has enabled high gene transfer in human T cells and the ability to expand cells to clinical numbers (Wilson et al., 2010) and to identify resistance mechanisms (Zhao et al., 2017). Additionally, it has been used as an alternative vehicle to deliver a guide RNA (gRNA) library for *in vivo* screening, to generate transgenesis in mice and to produce iPSCs (Wang et al., 2016; Xu et al., 2017; Li et al., 2017;).

SB-mediated integration supports highly efficient transgene integration in various cell types (Izsvák et al., 2000). The list includes somatic or germ cells, differentiated or stem cells essentially in all vertebrate species (**Table 3-4**).

Transposon	Species/Tissue/Cells	Purpose	References
piggyBac	Mouse embryonic stem (ES) cells	RecQ genomic transposition/ genetic	(Wang et al., 2008) (Wang et al., 2009)

		manipulation of Bloom Syndrome	(Kong et at., 2010)
piggyBac	Mouse embryonic stem (ES) cells	Gene trapping	(Kong et al., 2010)
piggyBac	human cells	Multiplexed PB transposon delivery /gene transfer	(Kahlig et al., 2009)
piggyBac	human iPSCs (hiPSCs)	Gene modulation expression in myogenic differentiation (myopathy)	(Tanaka et al., 2013)
piggyBac	human ESCs	Incorporation of bacterial artificial chromosome (BAC) transgenesis	(Rostovskaya et al., 2012)
piggyBac	iPSCs	Generation of transgene-free iPSCs	(Kaji et al., 2009) (Yusa et al., 2009)
piggyBac	mouse neural stem cells/ human pancreatic cells	Gene discovery	(Albieri et al., 2010) (You et al., 2011)
piggyBac	human T cells	Gene transfer/ identification of resistance mechanisms	(Wilson et al., 2010) (Zhao et al., 2017)
piggyBac	human iPSCs (hiPSCs)	Vehicle for guide RNA/ library delivery for *in vivo* screen/generation of transgenic and iPSCs	(Wang et al., 2016) (Li et al., 2017) (Xi et al., 2017)
Sleeping Beauty	Somatic/germ cells, differentiated/stem cells (all vertebrate species)	Transgene integration	(Izsvák et al., 2000)
Sleeping Beauty	human cells	Gene delivery	(Geurts et al., 2003)
Sleeping Beauty	human T cells	Gene transfer	(Huang et al. 2006)
Sleeping Beauty	Sheep fibroblast/ mouse iPS	Genetic modifications	(Hu et al., 2011) (Grabundzija et al., 2013)
Sleeping Beauty	various organisms and models	Production of iPSCs	(Davis et al., 2013)
Sleeping Beauty	primary human CD34-positive hematopoietic stem cells/ human primary T cells	Gene transfer via hyperactive SB100X	(Huang et al., 2008) (Singh et al., 2008)
Sleeping Beauty	human cells	Introduction of a chimeric and safety receptor (CAR)	(Monjezi et al., 2017)

Sleeping Beauty	Mouse modes	Generation of transgenic mice	(Huang et al., 2019)

Table 3-4 │Summary of studies using transposon systems as an insertional mutagen

3.5.2 Transposon insertional mutagenesis screen in mice

A typical transposon cargo used for *in vivo* insertional mutagenesis contains splice acceptors (SAs) followed by polyadenylation signals (pA) in both orientations, and a unidirectional promoter upstream of a splice donor (SD). A transposon can either disrupt gene function when it integrates into the body of a gene, thereby intercepting and curtailing transcription through the SA, pA elements, or it can activate expression when inserted upstream of a gene as the promoter, SD module drives expression of downstream sequences **(Figure 3-6)**. The general scheme for *in vivo* screening of cancer genes is to combine either activating/inactivating/bi-functional transposons with a source of transposase, the transposons will then translocate and stochastically activate or inactivate genes in somatic cells (Weber et al., 2020). Where these resultant mutational events are tumourigenic, the tumours that arise would have either oncogene activated or tumour suppressor genes inactivated by the transposon, conferring a selective advantage to the clone. Cancer genes can be identified by exploiting the transposon's molecular fingerprint, which allows amplification and sequencing of transposon genome junction fragments (Copeland and Jenkins, 2010).

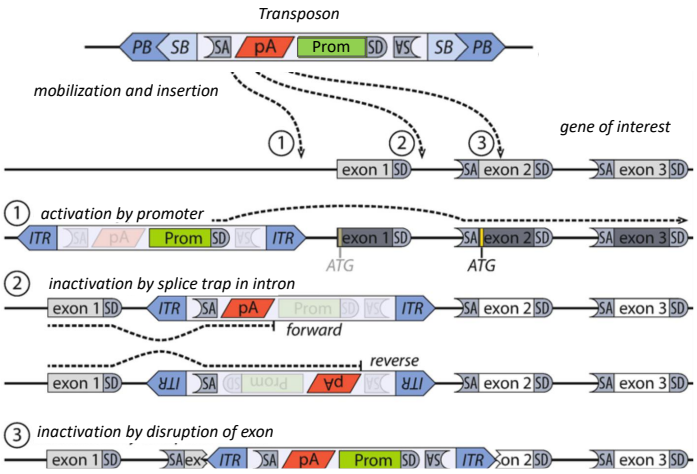


Figure 3-6 | Possible interferences of ATP transposon insertions with gene activity

Possible interferences of transposon insertions with gene activity: (1) insertion upstream of the gene and in forward orientation enables the integrated (Prom) promoter to drive gene expression; to overcome the early termination of translation immediately downstream of the promoter within the transposon repeat, the construct contains a splice donor (SD); (2) insertion into an intron disrupts gene function due to trapping of gene-splicing into the transposon; and (3) direct insertion into a critical exon disrupts gene function. (Adapted from Mathias J F et al., Cancer Discover Genes, Methods in Molecular Biology, 2018)

Mouse lines containing transposons are crossed to a transposase-containing mouse line **(Figure 3-7)**. The progeny that contains both transposon and transposase undergo somatic insertional mutagenesis and cells with insertions in a gene or combination of genes that favour tumorigenesis are positively selected. Therefore, insertional mutagenesis studies in the mouse are a powerful complement to the analysis of human cancer genomes, and facilitate the identification of driver cancer genes (Rad et al., 2015; De La Rosa et al., 2017; Takeda et al., 2017; Chapeau et al. 2017; Weber et al., 2019).

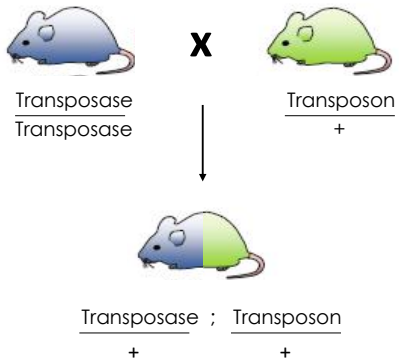


Figure 3-7 | Transgenic mice generation

Scheme of a typical genetic crossing for obtaining F1 experimental cohort mice carrying both the PB transposons and the transposase.

3.5.3 *piggyBac* insertional mutagenesis screen in mice

PB has been successfully used for transgenesis and insertional mutagenesis in mice (S. Wu et al. 2007) . As a transgenic vector, PB has a cargo capacity much bigger than that of retroviruses and lentiviruses. It efficiently produces single-copy integrations that mimic the situation of endogenous genes. PB transposase mice can be used for whole-body or tissue-specific insertional mutagenesis screens (Rad et al., 2010). *Rad et al. 2010* showed that depending on the promoter used for the transposon, mice can generate hematopoietic, solid or mixed tumours **(Figure 3-8)**. As an insertional mutagen, PB effectively inserts into the transcription units and interferes with gene expression. PB integrations can be easily mapped, so that desired animals could be selected before further breeding. For cancer genes discovery, transposons are designed to induce either gain-of-function (GOF) or loss-of-function (LOF) mutations when inserted in or near a gene based on its genetic cargo. This allowed the discovery of new potential driver genes, for example in B-cell lymphoma (BCL) (Weber et al., 2019), in pancreatic cancer (Rad et al., 2015). Furthermore, PB *in vivo* screen is useful for the identification of resistance mechanism in cancer (Chapeau et al., 2017).

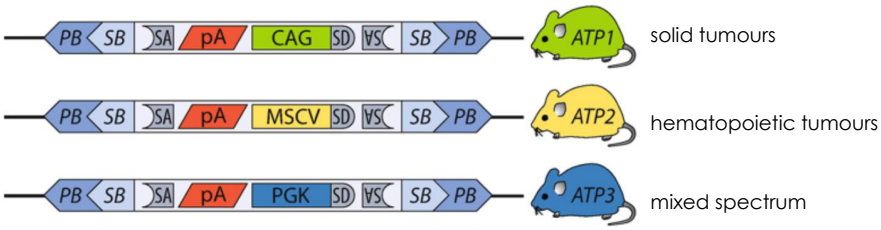


Figure 3-8 | Components of *piggyBac* transposition systems in mice.

Activating transposon constructs ATP1, ATP2, and ATP3: All constructs contain *piggyBac* (*PB*) and *Sleeping Beauty* (*SB*) inverted terminal repeats (ITR) and can be mobilized with both transposases. Bidirectional splice acceptors (SA) and poly-adenylation signals (pA) confer splice trap capabilities irrespective of the transposon's orientation. The constructs contain either the cytomegalovirus enhancer and chicken beta-actin promoter (CAG), the murine stem cell virus long terminal repeat (MSCV), or the phosphoglycerate kinase promoter (PGK). Constitutively activated in a whole-body screen, these promoters preferentially induce solid, hematopoietic, or mixed spectrum tumours, respectively. (Adapted from Mathias J F et al., Cancer Discover Genes, Methods in Molecular Biology, 2018)

3.5.4 Common insertion sites identification

The final step for an insertional mutagenesis screen is the recovery of transposon insertion coordinates in the genome and the identification of genomic locations that are hit by transposons more frequently than expected by chance (common insertion sites, CISs). A semi-quantitative method has been developed to identify *PB* insertion sites (Rad et al. 2015; Bronner et al. 2016; Friedrich et al. 2017), which is based on splinkerette PCR (**Figure 3-9**) (Devon et al., 1995). Splinkerette PCR involves DNA fragmentation by acoustic shearing, followed by ligation of a splinkerette adaptor (a double-stranded oligonucleotide) and PCR amplification of the transposon–genome junction fragments using transposon-specific and adaptor-specific primers. In the first round of PCR, the adapter-ligated DNA fragments that also contain transposon sequence are specifically amplified with primers for the Splinkerette adapter on one end and the transposon inverted terminal repeats (ITR) on the other end. In the second round of PCR, the Illumina NGS-specific adapter overhangs and multiplex barcodes are attached. The design of the splinkerette adaptor ensures that only transposon-

containing adaptor-ligated DNA is amplified. The resulting PCR products are then highly multiplexed next-generation sequenced and mapped to the genome.

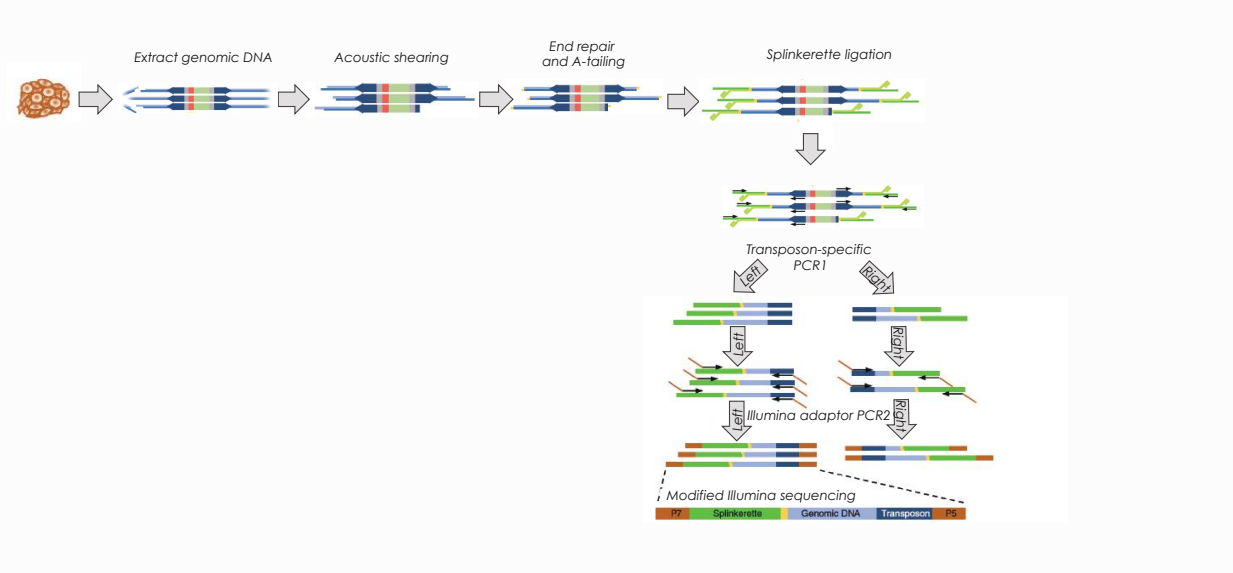


Figure 3-9 | Library preparation steps

First steps include acoustic fragmentation of tumour DNA and enzymatic repair (blunting) of the fragment ends after an A-tail is attached, the splinkerette adaptor is ligated. A first round of PCR with transposon and adaptor-specific primers allows selective amplification of transposon–genome junction fragments. A second nested PCR further amplifies these fragments and extends them with a multiplexing barcode tag and with the Illumina flow cell binding sites P5 and P7.

To correctly identify transposon insertions, the sequencing strategy aims to verify transposon–genomic junctions: the sequencing primer is placed within the transposon ITR so that a part of the transposon sequence, then the insertion point ('TTAA' for *PB* and 'TA' for *SB*) and finally the genomic sequence of the insertion location is sequenced. Therefore, each sequencing read starts with a short stretch of identical sequence representing the transposon and the insertion point. Illumina sequencing instruments detect the exact position of each DNA cluster on the flow cell. However, due to the uniformity of the transposon sequence at the beginning of each fragment on the flow cell, the sequencing run may suffer from insufficient degree of randomness, reduced quality or fail completely. Illumina sequencing requires

modification to circumvent low-complexity template problems. One solution to this problem is to spike the library with PhiX (Mukherjee et al., 2015), a known bacteriophage genome that adds sequence diversity and can later be removed during bioinformatics analysis. The disadvantage of PhiX spiking is the loss of ~20% sequencing yield (Mukherjee et al., 2015).

After reading 75 bp, the template is denatured, the same sequencing primer is re-annealed and the previously skipped 12 bp are sequenced, completing the transposon–genomic junction. Subsequently, the 8-bp indexing tag (‘sample barcode’) is read and finally, after template reversal, a 75-bp reverse read is acquired. QiSeq determines sequence read numbers for each insertion, thus defining the clonal distribution of transposon integrations within a cancer (Bronner et al., 2016). This allows predictions/conclusions about the “importance” and position of cancer genes at the trunks or branches of evolutionary trees (Friedel et al., 2013).

4 | RATIONALE OF THE WORK

Breast cancer is the most frequent and lethal cancer among women worldwide (Fahad Ullah, 2019). In most of the cases, breast cancers can progress to metastasis, which remains the primary cause of solid cancer-evoked mortality. Cancer progression is the result of sequential acquisition of genomic alterations that contributes to subsequent clonal expansion (Nowell, 1976; Hanahan and Weinberg, 2011). Characterization of breast cancer genomic diversity contributed to the identification of numerous putative driver cancer genes (Nik-Zainal et al., 2016, Pereira et al., 2016, Yates et al., 2015; Robinson et al., 2017, Angus et al., 2019, Bertucci et al. 2019, De Mattos-Arruda et al., 2019), yet their functional validation and the nature of the collaborating oncogenic pathways remain elusive. My PhD work aimed to identify and validate these synergistic genes or pathways involved in breast cancer tumorigenesis and metastasis.

One of the most frequent gain of function mutations in breast cancer occurs in the PI3K pathway. Activating mutations of *PIK3CA* are present in ~30% of human breast cancers at all stages. It has been shown that inducible expression of *PIK3CA*^{H1047R} mutation evokes mammary tumours in mice suggesting a causative effect for this mutation in breast tumorigenesis (Meyer et al., 2013; Koren and Bentires-Alj, 2013; Koren et al., 2015).

Transposon-induced mutagenesis screens are an efficient method to identify cancer driver genes. When mobilized by PB transposase, the transposon integrates the genome randomly, and cells with insertions in a gene or combination of genes that favour tumorigenesis are positively selected.

Here, we performed an unbiased *ex vivo* and *in vivo* PB transposon insertional mutagenesis screen to identify possible *PIK3CA* mutant co-occurring mechanisms of tumourigenesis and metastatic progression.

Specifically, I wanted to elucidate:

i. *PIK3CA* synergistic mechanisms of metastatic colonization by PB transposon *ex vivo* insertional mutagenesis screen.

ii. *PIK3CA* synergistic pathways important for breast tumourigenesis by *in vivo* mutagenesis.

iii. Validate the identified candidate breast cancer genes/pathways involved in the progression of the disease.

5 | RESULTS PART I

Research article submitted to *EMBO MOL MED* Journal.

NFIB-ERO1A axis promotes breast cancer metastatic colonization of disseminated tumour cells

Authors: Federica Zilli[1,2*], Pedro Marques Ramos[2,3*], Priska Auf der Maur[1], Charly Jehanno[1], Atul Sethi[1,2,4], Marie-May Coissieux[1,2], Tobias Eichlisberger[2], Laura Bonapace[3], Joana Pinto Couto[1,2,3], Roland Rad[5], Michael Rugaard Jensen[3], Michael B. Stadler[2,4] and Mohamed Bentires-Alj[1,2]

Affiliations:

[1]Department of Biomedicine, Department of Surgery, University Hospital Basel, University of Basel, Switzerland

[2]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

[3]Novartis Institutes for Biomedical Research, Basel, Switzerland

[4]Swiss Institute of Bioinformatics, Basel, Switzerland

[5]Institute of Molecular Oncology and Functional Genomics, the Center for Translational Cancer Research (TranslaTUM), and the Department of Medicine II, TUM School of Medicine, Technische Universität München, Germany; German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

* These authors contributed equally to this work

Running title: NFIB/ERO1A/VEGFA axis induces metastases

Keywords: Breast cancer / ERO1A / Metastasis / NFIB / VEGFA

Correspondence:

M. Bentires-Alj, Pharm., Ph.D.

Department of Biomedicine, University of Basel, University Hospital

Lab 306, Hebelstrasse 20, CH-4031 Basel/Switzerland

Email: m.bentires-alj@unibas.ch

Phone: +41 (0) 61 26 53 313

5.1 Summary

Metastasis is the main cause of deaths related to solid cancers. Active transcriptional programs are known to regulate the metastatic cascade but the molecular determinants of metastatic colonization remain elusive. Using an inducible *piggyBac* (PB) transposon mutagenesis screen, we have shown that overexpression of the transcriptional factor nuclear factor IB (NFIB) alone is sufficient to enhance primary mammary tumour growth and lung metastatic colonization. *NFIB* is thus clinically relevant: it is preferentially expressed in the poor-prognostic group of basal-like breast cancers, and high expression of the *NFIB/ERO1A/VEGFA* pathway correlates with reduced breast cancer patient survival. Mechanistically and functionally, NFIB increases expression of the oxidoreductase *ERO1A*, which enhances VEGFA-mediated angiogenesis and colonization - the last and fatal step of the metastatic cascade.

5.2 Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of death in female cancer patients worldwide (Fahad Ullah, 2019). Metastasis remains the primary cause of solid-cancer-evoked mortality (Gupta and Massagué, 2006). The multistep process of metastasis includes local tumour cell invasion, entry into the vasculature (intravasation), exit from the circulation into the parenchyma of distant organs (extravasation), and colonization. Colonization, which results in the development of clinically manifesting metastasis and fatal disease, is dependent upon resistance of the disseminated tumour cells (DTCs) to immune and host tissue defences, survival in a foreign environment, and tumour initiation capacity (Pantel and Brakenhoff, 2004, Massagué and Obenauf, 2016). However, the molecular determinants of colonization remain elusive.

Numerous oncogenic mutations and other genomic alterations co-occur often in cancer cells and result in tumour heterogeneity, which impinges on the clinical outcome of the disease

(Koren and Bentires-Alj, 2015; Yates and Campbell, 2012; Kennecke et al., 2010). The dynamic evolution of the cancer genome is influenced by the generation of additional mutations and selective forces acting on cancer clones (Kreso and Dick, 2014). Next-generation sequencing of DNA from clinical specimens has provided insights into breast cancer genetics and contributed to identifying potential drivers of tumour progression (Nik-Zainal et al., 2016, Pereira et al., 2016, Yates et al., 2015; Robinson et al., 2017, Angus et al., 2019, Bertucci et al. 2019, De Mattos-Arruda et al., 2019). Mechanistic elucidation of the effects of co-occurring genomic alterations and their functional validation is required to define the causality between these events and their contribution to specific steps of tumour progression to overt metastases.

Activating mutations of *PIK3CA*, which encodes for the p110α catalytic subunit of phosphoinositide 3-kinase (PI3K), are among the most frequent alterations in human breast cancer and lead to an hyperactivated PI3K pathway signalling (Yuan and Cantley, 2008; Zhao and Vogt, 2008; Miller, 2012). We and others have shown that inducible expression of the *PIK3CA*^{H1047R} mutation evokes heterogeneous mammary tumours in mice (Koren et al., 2015; Meyer et al., 2013; Koren and Bentires-Alj, 2013), which indicates a causative effect of *PIK3CA* mutations in mammary tumorigenesis. In contrast, metastases were not found in our *PIK3CA*^{H1047R}-expressing mice (Koren et al., 2015; Koren and Bentires-Alj, 2013), which thus provides a model system to identify collaborating gain- or loss-of-function genomic alterations that contribute to metastasis.

Transposon insertional mutagenesis is a powerful tool in mice for discovering genes related to cancer (Rad et al., 2010; Ding et al., 2005; Dupuy, 2010; Dupuy et al., 2005). Indeed, the fact that transposons are mobile within the genome and alter gene activity in those cells that express the transposase make these systems ideal for whole-genome screens. The *piggyBac* (PB) transposon was engineered to be active in mammalian cells (Ding et al., 2005) and it allows functional identification not only of oncogenes but also of tumour suppressor genes,

depending on the site of insertion and the orientation of the transposon. (De La Rosa et al., 2017; Rad et al., 2015; Takeda et al., 2017; Weber et al., 2019).

We performed an unbiased PB transposon insertional mutagenesis screen in cancer cells with an activating PIK3CA^{H1047R} mutation to identify possible synergistic mechanisms of metastatic colonization. Mechanistically and functionally, we demonstrate that NFIB enhances expression of the oxidoreductase *ERO1A* and of *VEGFA*, promotes metastatic colonization, and shortens overall survival of mice.

The paper explained

Problem

Metastasis is the main cause of solid-cancer-related death and is a major clinical challenge in breast cancer. Although many studies have thoroughly characterized and classified breast tumours, the molecular determinants of metastatic colonization remain elusive. Their delineation and functional validation are urgently needed to improve our understanding of this currently incurable disease.

Results

We engineered a non-metastatic cell line with the doxycycline (dox)-inducible *piggyBac* (PB) transposon mutagenesis system and performed an unbiased *in vivo* PB-mutagenesis genetic screen to identify drivers of metastatic colonization. We have shown that the transcriptional factor nuclear factor IB (NFIB) is necessary and alone sufficient for breast cancer metastasis. NFIB is upregulated in mammary tumourspheres and metastases. Transcriptional profiling of tumours and tumourspheres derived from highly metastatic cell lines showed that this lethal effect of NFIB is mediated by increased expression of the oxidoreductase *ERO1A*. Mechanistically and functionally, ablation of ERO1A in NFIB-overexpressing models decreased VEGFA-mediated angiogenesis, reduced metastases, and prolonged overall survival of the animals. Furthermore, *NFIB/ERO1A/VEGFA* co-expression correlates with a poor prognosis in triple-negative breast cancer (TNBC) patients, emphasizing the clinical importance of this molecular axis in breast cancer metastasis.

Impact

This study provides new molecular insights into the determinants of metastatic colonization. Our work not only highlights the power of genetic screening to identify functionally relevant metastatic networks, but also describes the mechanism by which the transcriptional factor NFIB mediates metastatic colonization. Thus, we have revealed a targetable network that influences colonization, the last and fatal step of the metastatic cascade.

5.3 Results

5.3.1 Transposon mutagenesis confers metastatic potential to *PIK3CA*^{H1047R} mammary cells.

To identify cancer genes relevant to metastatic progression, we engineered a murine non-metastatic mammary cancer cell line derived from *PIK3CA*^{H1047R} mutant tumours (LB-mHR1, here cited as HR1) with a doxycycline (dox)-inducible *piggyBac* (PB) transposon system (HR1.PB) (Ivics et al., 2009, Rad et al., 2010) (Fig 1A). HR1 cells lacking the transposon served as control (HR1.Ctrl). To perform an unbiased *in vivo* PB transposon mutagenesis screen (Fig 1B), we injected HR1.PB and HR1.Ctrl orthotopically into 19 and 14 NOD/SCID mice, respectively, and monitored the animals for tumour growth and metastasis. PB transposon mutagenesis in HR1.PB cells increased tumour incidence (Fig EV1A) and promoted metastasis (Fig EV1B). Metastases were observed in half of the mice injected with HR1.PB cells and nine lung-metastases (LM) -derived cell lines were isolated. To phenotypically characterize the LM cell lines *in vivo*, we injected them into mammary fat pads of NOD/SCID mice. The LM1, LM8, and LM9 lines exhibited a particularly aggressive behaviour characterized by accelerated tumour growth and metastasis (Fig EV1C-E**)**. These results indicate that selected transposon integrations can confer aggressive metastatic behaviour on cells.

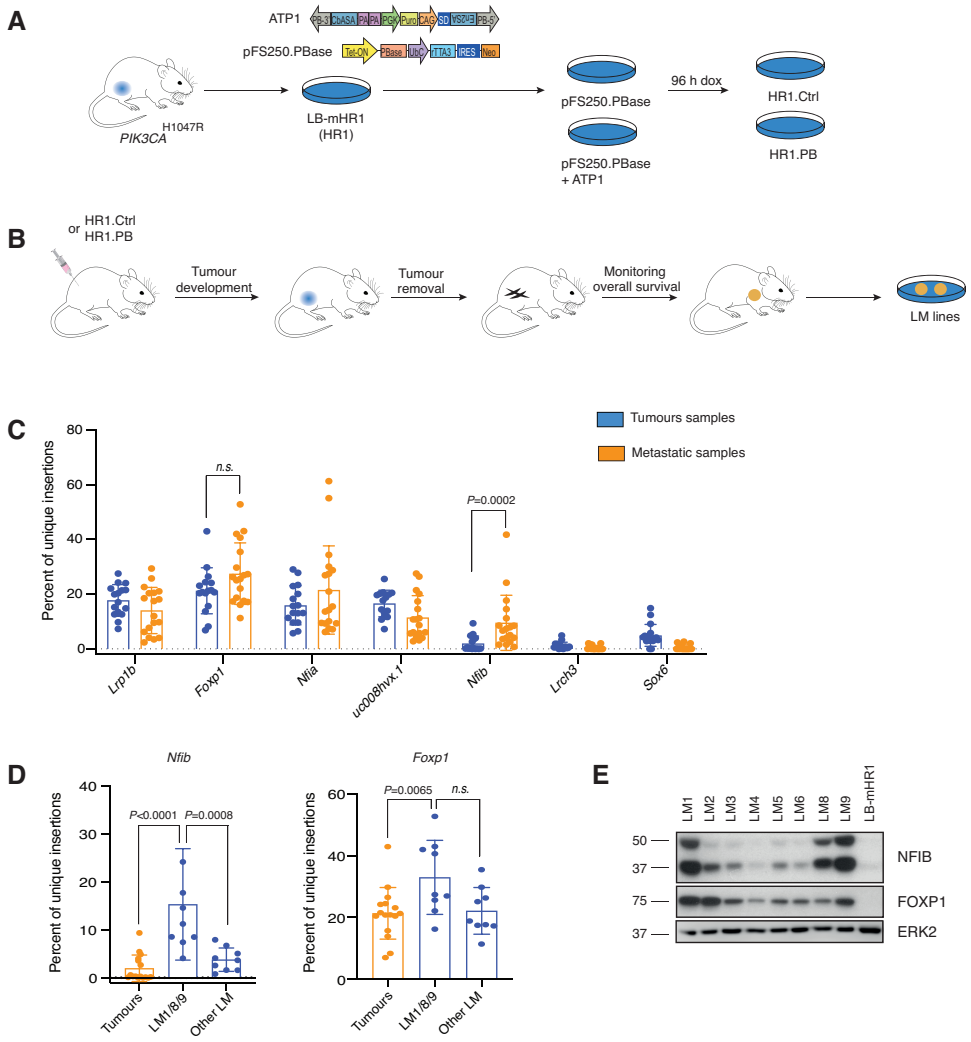


Figure 5-1 | Transposon mutagenesis screen identifies *Nfib* as candidate metastatic inducer in breast cancer.

A Screen design: generation of mammary cancer cell lines LB-HR1 (HR1) from *PIK3CA*^{H1047R} mutant transgenic animals and transfection with the *piggyBac* (PB) system (HR1 Ctrl and HR1.PB cells) (Ivics et al., 2009). PB transposon (ATP1-Puro) and transposase (pFS250.PBase) plasmid design. PB-3'/5', *piggyBac* inverted terminal repeats; CβASA, Carp β-actin splice acceptor; En2SA, Engrailed-2 exon-2 splice acceptor; SD, Foxf2 exon-1 splice donor; pA, bidirectional SV40 polyadenylation signal; PGK, mouse phosphoglycerate kinase 1 promoter; Puro, puromycin resistance; CAG, cytomegalovirus enhancer and chicken beta-actin promoter; Tet-ON, tetracycline-responsive element-tight promoter; PBase, PB transposase; UbC, Ubiquitin C promoter; rTTA3, reverse tetracycline transactivator 3; IRES, internal ribosomal entry site; Neo, Neomycin/G418 selection marker. The cell pools were treated with doxycycline (1 µg/ml) for 96 h *in vitro*.

B *In vivo* screen design: generation of lung metastatic mammary cancer cell lines (LM) from HR1.PB cell lines after orthotopic injections of HR1.PB or HR1.Ctrl.

C Bar graph showing the percentage of unique insertions in the top 25 genes normalized to the total number of insertions in a given sample. Gene are annotated with gene symbol when available otherwise with UCSC ID transcript names. Data are from all the sequenced samples (tumour samples: 16 primary tumours; metastatic samples: 3 lung-metastases, 6 lung macro-

metastases and 9 LM cell lines). Means ± s.d., two-tailed Mann-Whitney U-test, *n.s*. = not significant.

D Bar graphs showing the percentage of unique insertions in *Nfib* and *Foxp1* in tumours, LM1, LM8, LM9 and other lung cell lines normalized to total number of insertions in a given sample. Dots represent individual samples, means ± s.d., two-tailed Mann-Whitney U-test, *n.s*. = not significant.

E Immunoblot analysis of LM and LB-mHR1 cell lines showing NFIB and FOXP1 protein levels. ERK2 served as a loading control.

5.3.2 Insertional PB mutagenesis screening identifies *Nfib* as a metastatic gene in mammary cancer.

To investigate induced alterations that enabled cancer cells to metastasize, we mapped transposon integration sites in 16 primary tumours and 18 metastatic samples, searching for insertions enriched at metastatic sites compared with tumours. We used splinkerette PCR followed by next-generation sequencing from both transposon arms (Friedrich et al., 2017) and identified 1,252 insertions in 1,080 genes (Table EV1). To evaluate the relative abundance of each integration site and minimize PCR-induced amplification effects, we devised a diversity count measure for each integration site within each sample (percentage of unique insertions) that quantifies the number of unique sheared ends rather than all sequencing counts. To select metastatic genes, we searched for genes with more unique insertion sites (based on their diversity counts) in metastatic samples than in tumour samples (Fig 1C). Depending on the transposon integration site, insertions may lead to gene activation or to inactivation. The most frequently altered genes in the primary tumours and/or metastatic samples were *Lrp1b, Nfia*, *Foxp1*, *Nfib, Lrch3,* and *Sox6.* Some of these unique insertions (e.g., *Foxp1*) were found in both metastatic and non-metastatic samples at high frequency, suggesting that they are unlikely to be specifically required for metastasis. Notably, unique insertions in *Nfib* were particularly enriched in a subset of highly metastatic samples (LM1/8/9) compared with the other hits (Fig 1D). The same-strand insertions upstream of both *Nfib* and *Foxp1* transcription starts suggest

them to be candidate oncogenes. Consistent with the pattern of transposon integration, analysis of FOXP1 and NFIB protein abundance in the LM cell lines revealed enhanced levels of FOXP1 in all LM cell lines but selective elevation of NFIB in LM lines with high metastatic potential (Fig 1E). These data suggest the importance of *Nfib* in mammary cancer metastasis.

5.3.3 Depletion of *Nfib* delays mammary cancer growth and abrogates metastases.

To assess the contribution of *Nfib* and *Foxp1* to tumour growth and metastatic progression, we produced knockout (KO) cells for the two genes using CRISPR-Cas 9 technology in the highly metastatic LM1 and LM9 lines and generated oligoclonal pools of cells (Fig EV2A and B). Considering that NFIB has been implicated in skin stem cell maintenance (Chang et al., 2013) and tumour growth in other cancer types (Brayer et al., 2016; Fane et al., 2017; Semenova et al., 2016; Wu et al., 2018, 2016), we addressed its effects on mammary tumours. *Nfib* KO in LM1 and LM9 delayed tumour growth (Fig 2A) and markedly decreased the frequency of tumour-initiating cells (TICs) (Fig 2B and C). Consistently, *Nfib* KO also reduced tumoursphere formation (Fig 2D and E).

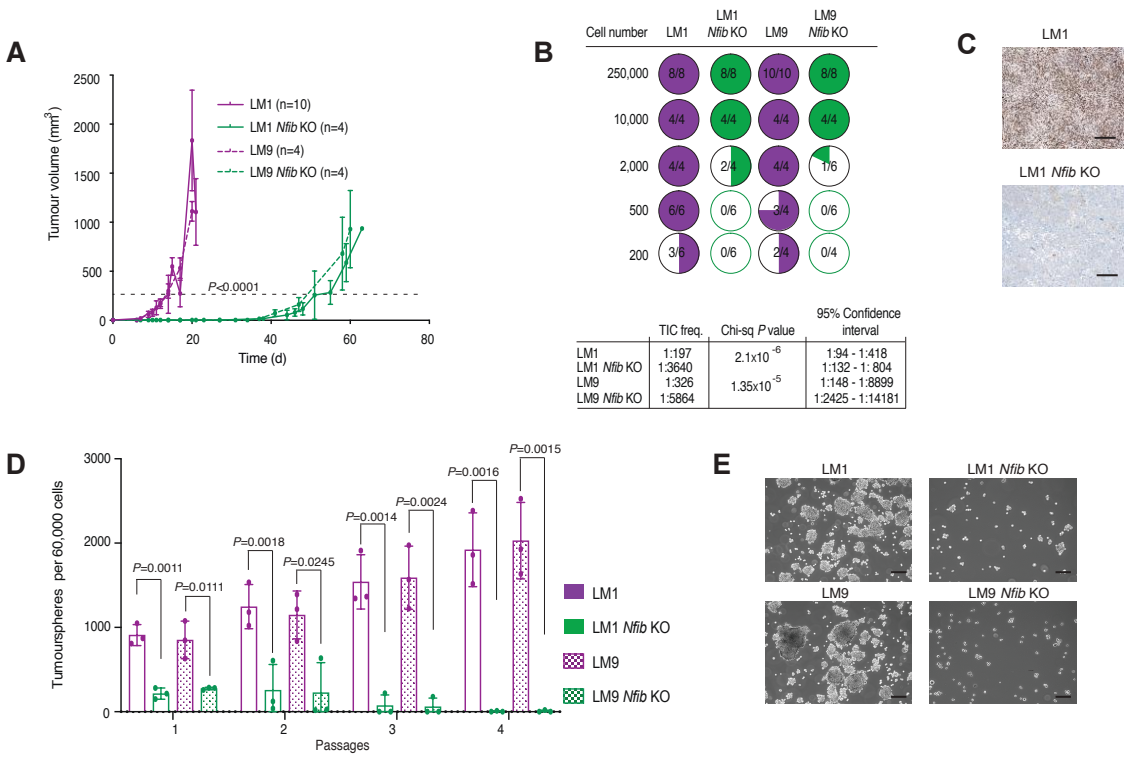


Figure 5-2 | *Nfib* overexpression accelerates tumour growth.

A The kinetics of LM1 (n = 10), LM9 (n = 4), LM1 *Nfib* KO (n = 4), and LM9 *Nfib* KO (n = 4) tumour growth upon orthotopic injection of 250 x 10^3 cells into NOD/SCID mice. Curves show means of tumour volume ± s.d., two-way ANOVA group analysis on the times to reach 250 mm^3 (dashed line).

B Limiting dilution assay: (upper panel) quantification of tumour incidence and respective statistics (Chi-square test and tables) upon orthotopic injection into mice of LM1, LM9, LM1 *Nfib* KO and LM9 *Nfib* KO cells. Lower panel: Tumour initiating cell (TIC) frequencies and 95% confidence intervals were calculated using ELDA (Hu and Smyth, 2009) as described in the experimental procedures.

C Representative images of immunostaining for NFIB of tumours from mice injected with LM and LM *Nfib* KO cancer cells. Scale bar 100 µm.

D Bar graph showing tumoursphere numbers over four passages of LM1, LM9, LM1 *Nfib* KO and LM9 *Nfib* KO, n = 3 biological replicates, means ± s.d., two-tailed Student's t-test.

E Representative images of tumourspheres of LM1 and LM1 *Nfib* KO cancer cell lines at the first passage. Scale bar 100 µm.

Deletion of *Foxp1* increased mammary tumour latency (Fig EV2C) and slightly decreased metastasis compared with the wild type (WT) (Fig EV2D). In contrast, when LM1 and LM9 *Nfib* KO cells were injected orthotopically and the primary tumour removed, we

observed a dramatic abrogation of lung metastasis (orthotopic metastasis assay) compared with controls (Fig 3A and B). The data suggest that the *Nfib* overexpression observed in our PB screen and in metastatic lines may also enhance metastatic colonization. To test this possibility, we injected *Nfib* KO and control cells intravenously (*i.v.*, experimental metastasis assay), and found that *Nfib* KO completely or dramatically impaired metastatic colonization in the LM9 and LM1 lines, respectively (Fig 3C and D). Finally, ablation of *Nfib* using CRISPR-Cas 9 technology also increased tumour latency, decreased the number of circulating tumour cells, and prolonged overall survival of Balb/C mice orthotopically injected with the highly metastatic 4T1 mammary cancer cells (Fig EV3A-D). The results of these functional assays demonstrate the importance of NFIB in mammary cancer metastatic colonization.

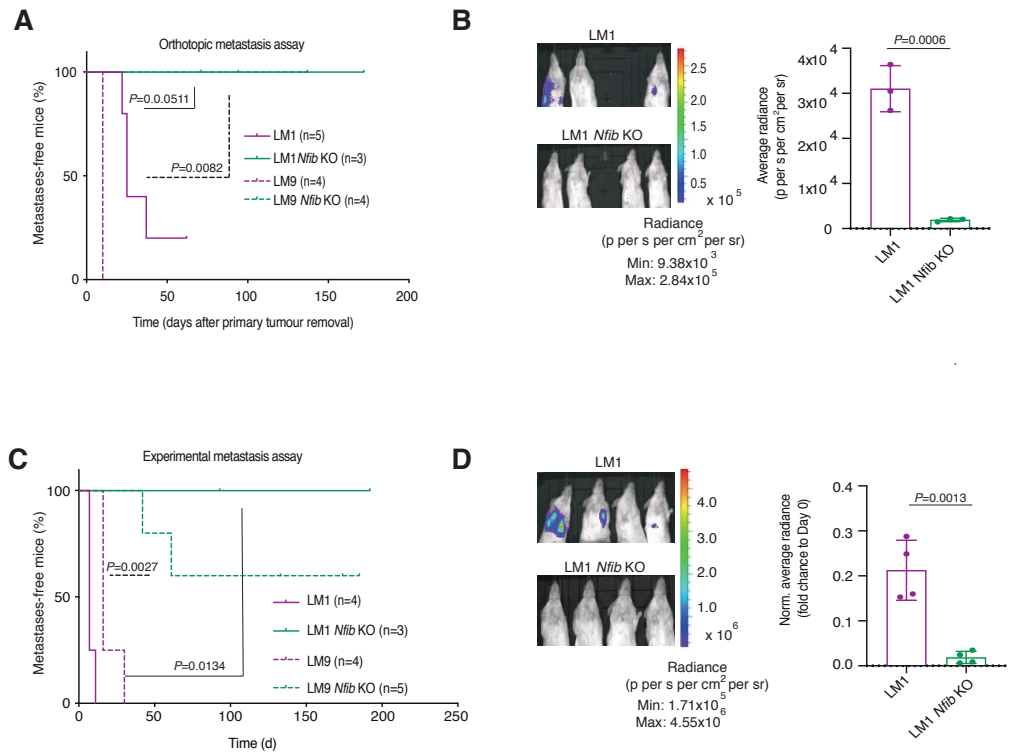


Figure 5-3 | *Nfib* ablation abrogates mammary tumour metastasis.

A Kaplan-Meier plot depicting metastasis onset after tumour removal in mice injected orthotopically with LM1 ($n = 5$), LM9 ($n = 4$), LM1 *Nfib* KO ($n = 3$), or LM9 *Nfib* KO ($n = 4$) cells, two-tailed log-rank test.

B *Nfib* knockout abrogates metastasis in the orthotopic LM1 model. Representative bioluminescence image (left panel) and bar plot quantification (right panel) of mice at 20

(LM1) and 40 (KO) days after primary tumour removal; $n = 3$ mice, means ± s.d., two-tailed Student's *t*-test.

C *Nfib* knockout impairs experimental metastases formation in the LM models. Kaplan-Meier plot showing metastatic incidence of animals inoculated *i.v.* with LM1 ($n = 4$), LM9 ($n = 4$), LM1 *Nfib* KO ($n = 3$), or LM9 *Nfib* KO ($n = 5$) cells, two-tailed log-rank test.

D Representative bioluminescence images (left panel) and bar plot quantification (right panel) of mice injected *i.v.* with LM1 or LM1 *Nfib* KO 16 days after cancer cell injection. $n = 4$ mice, means ± s.d., two-tailed Student's *t*-test.

5.3.4 *Nfib* is sufficient to induce metastasis.

We next asked whether NFIB activity is alone sufficient for inducing metastasis. Using Cas9-Activators with Synergistic Activation Mediators (SAM) (Konermann et al., 2015), we overexpressed *Nfib* from its endogenous promoter in the parental HR1 cell lines (Fig EV4A). This accelerated tumour onset (Fig EV4B), metastasis formation after orthotopic injection (Fig 4A and B), and metastatic colonization after tail-vein injection (Fig 4C and D). Overexpression of *NFIB* in SUM159PT, a human TNBC cell line with low metastatic potential (Fig EV4C), similarly decreased tumour latency (Fig EV4D), and increased metastasis, both in the orthotopic (Fig EV4E and F) and the experimental metastasis assays (Fig 4E and F). Thus, NFIB appears to be sufficient to induce metastasis when overexpressed in non- (HR1) and low- (SUM159PT) metastatic mammary cancer lines.

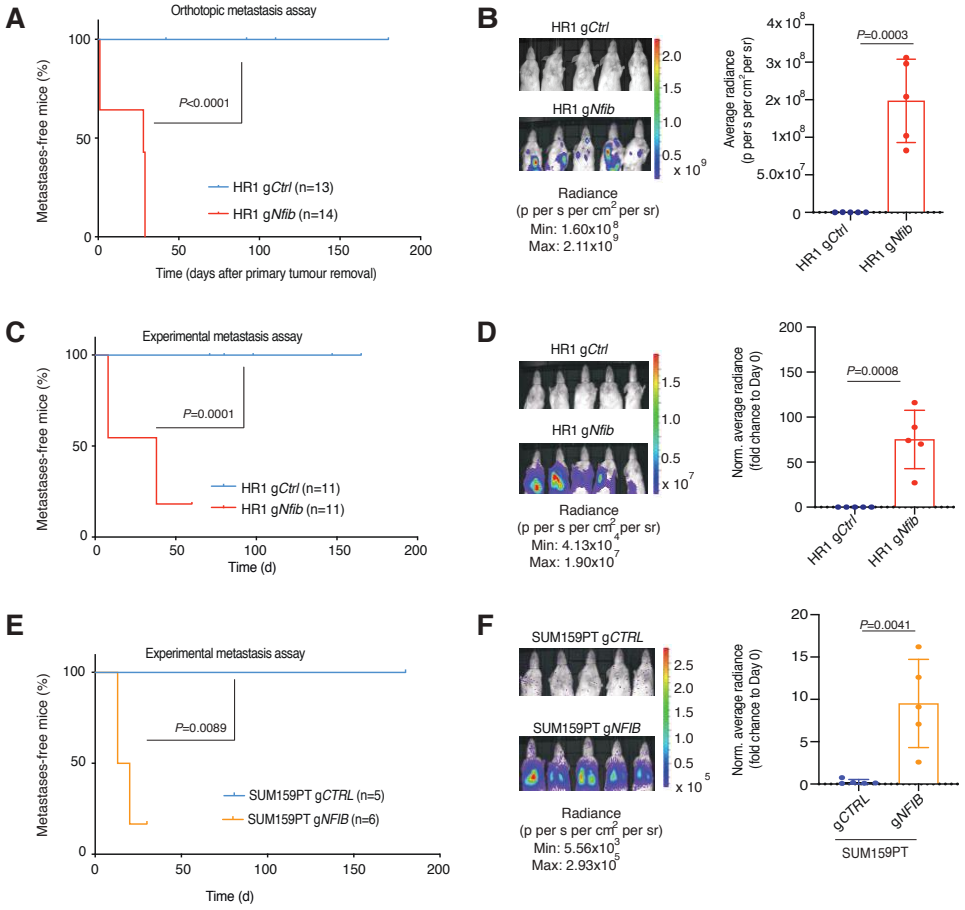


Figure 5-4 | *Nfib* is sufficient to induce metastasis.

A Kaplan-Meier plot depicting metastasis onset after tumour removal in mice injected orthotopically with HR1 g*Ctrl* (n = 13) or HR1 g*Nfib* (n = 14), two-tailed log-rank test.

B *Nfib* overexpression via CRISPR Cas9 SAM in HR1 parental cells enhances metastasis in the orthotopic model. Representative bioluminescence images (left panel) and bar plot quantification (right panel) of metastases at day 20 (HR1 g*Ctrl*) and day 2 (HR1 g*Nfib*) after primary tumour removal. n = 5 mice, means ± s.d., two-tailed Student's t-test.

C *Nfib* overexpression via CRISPR Cas9 SAM in the HR1 parental cells enhances experimental metastases. Kaplan-Meier showing metastatic incidence of animals inoculated *i.v.* with HR1 g*Ctrl* (n = 11) or HR1 g*Nfib* (n = 11) cells, two-tailed log-rank test.

D Representative bioluminescence images (left panel) and bar plot quantification (right panel) of mice *i.v.* injected with HR1 g*Ctrl* or HR1 g*Nfib* at day 16 after cancer cell injection. n = 5, means ± s.d., two tailed Student's t-test.

E *NFIB* overexpression enhances experimental metastases formation in the SUM159PT model. Kaplan-Meier survival analysis of animals inoculated *i.v.* with SUM159PT g*CTRL* (n = 5) or SUM159PT g*NFIB* (n = 6) cells, two-tailed log-rank test.

F Representative bioluminescence images (left panel) and bar plot quantification (right panel) of mice *i.v.* injected with HR1 g*CTRL* or HR1 g*NFIB* at day 13 after cancer cell injection. n = 5 mice, means ± s.d., two-tailed Student's t-test.

5.3.5 NFIB induces *Ero1l/ERO1A* expression.

To determine the molecular mechanism underlying increased metastatic colonization driven by NFIB, we sequenced RNA of tumourspheres and primary tumours derived from *Nfib* high-expression models (LM1 and LM9) and *Nfib* low-expression models (HR1, LM1 *Nfib* KO and LM9 *Nfib* KO). Comparing the 200 most differentially expressed genes with ChIP-seq data of putative transcriptional targets of *Nfib* in the mammary gland (Shin et al., 2016) and epithelial-melanocyte stem cells (Chang et al., 2013) (Table EV2-4), we found endoplasmic reticulum disulphide oxidase 1 like (*Ero1l*) to be the single common gene (Fig EV5A and B). Indeed, abundance of the mouse and human ERO1L/A protein was increased in *Nfib/NFIB*-overexpressing cells (Fig EV5C). Consistently, *Nfib* KO decreased *Ero1l* mRNA levels (Fig EV5D). ERO1L/ERO1A is an oxidoreductase located in the endoplasmic reticulum and involved in the production of hydrogen peroxidase (H_2O_2) (Zito, 2015); it is a poor prognostic factor in cancer (Kim et al., 2018; Yang et al., 2018, Takei et al., 2017; Zhou et al., 2017).

We next examined whether *Ero1l/ERO1A* mediates NFIB effects on metastatic colonization. Notably, knockdown of *ERO1A* in SUM159PT g*NFIB*, using two doxycycline-inducible shRNA constructs (Fig EV5E), reduced metastatic colonization after tail-vein injection and prolonged survival (Fig 5A-C). Altogether, these findings show that *ERO1A* expression is increased in *NFIB*-overexpressing models and suggest that the NFIB-ERO1A axis is critical for metastatic colonization in breast cancer.

5.3.6 *NFIB-ERO1A* induces angiogenesis.

As ERO1A has been shown to enhance *VEGFA* expression (Tanaka et al., 2016), we assessed *VEGFA* mRNA levels in *NFIB* models. *VEGFA* expression increased in cells overexpressing *NFIB* and decreased in cell lines and lung metastases upon *ERO1A* downregulation (Fig 5D), indicating that *VEGFA* expression is *NFIB/ERO1A* dependent. We also assessed angiogenesis

in sections of mice lungs injected intravenously with sh*CTRL* or sh1- or sh2-*ERO1A* and observed a decrease in CD31-positive blood vessels in lung metastases (Fig 5E). Taken together, these data indicate that NFIB is a breast cancer metastasis transcriptional regulator which, through increased expression of *ERO1A* and *VEGFA*, promotes angiogenesis at the metastatic site. This creates a permissive microenvironment for metastatic colonization.

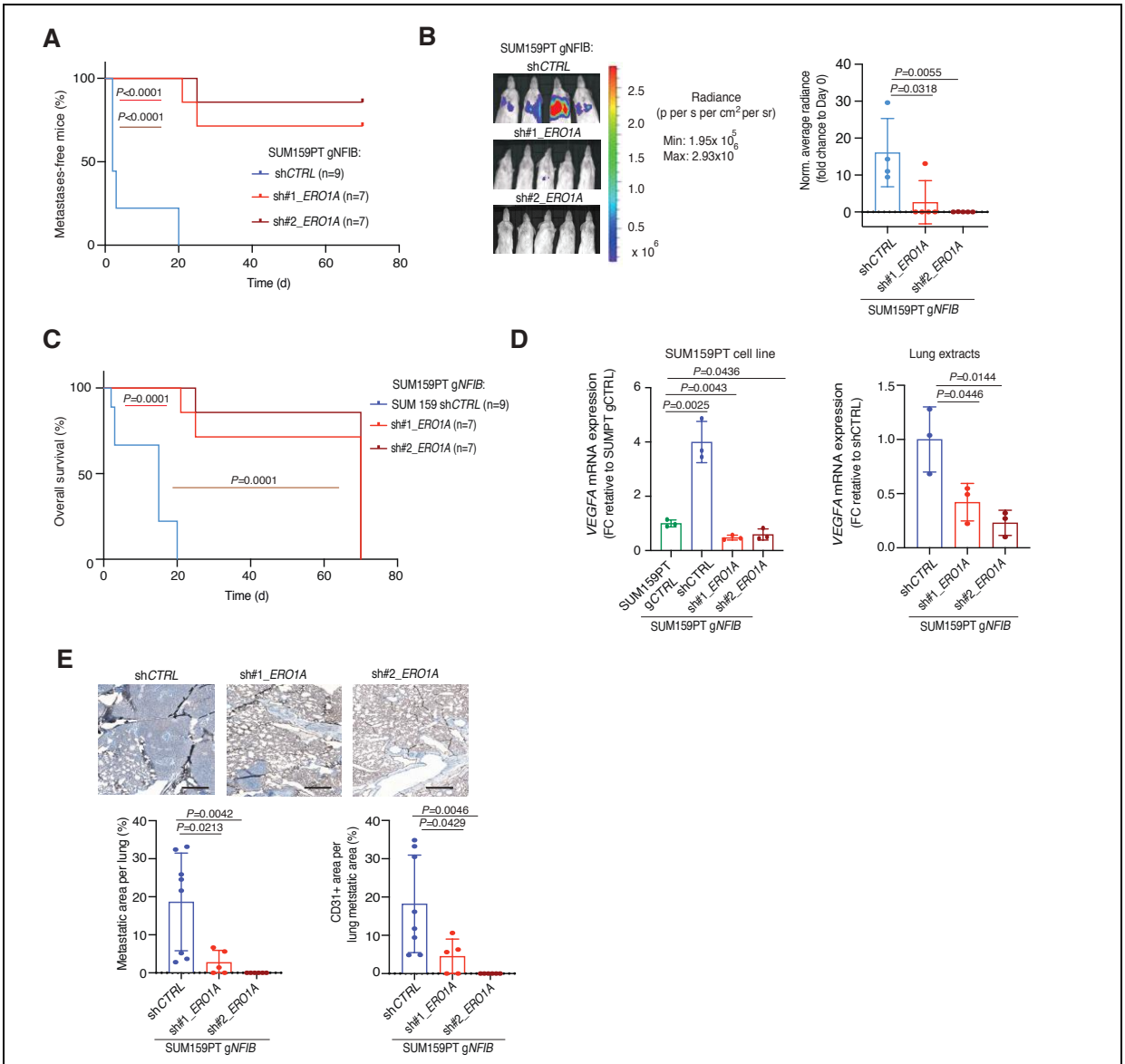


Figure 5-5 | NFIB-ERO1A axis promotes lung metastatic colonization via angiogenesis.

A Kaplan-Meier survival analysis of NOD/SCID mice inoculated *i.v.* with SUM159PT g*NFIB* sh*CTRL* (n = 9), SUM159PT g*NFIB* sh1 *ERO1A* (n = 7), or SUM159PT g*NFIB* sh2 *ERO1A* (n = 7) cells. Two-tailed log-rank test.

B Representative bioluminescence images (left panel) and bar plot quantification (right panel) of mice *i.v.* injected with SUM159PT sh*CTRL* (n = 9), SUM159PT sh1 *ERO1A* (n = 7)

or SUM159PT sh2 *ERO1A* (n = 7) eight days after cancer cell injection, means ± s.d., two-tailed Student's *t*-test.

C Kaplan-Meier survival analysis of overall survival of mice injected *i.v.* with SUM159PT sh*CTRL* (n = 10), SUM159PT sh1 *ERO1A* (n = 12) or SUM159PT sh2 *ERO1A* (n = 13) cells, two-tailed log-rank test.

D Left panel: Bar graph representing mean *VEGFA* mRNA expression in SUM159PT g*CTRL* and SUM159PT g*NFIB* breast cancer cell line upon downregulation of *ERO1A* with two independent shRNAs (sh1 and sh2). Right panel: Bar graph representing mean *VEGFA* mRNA expression in lung extracts from animals inoculated *i.v.* with SUM159PT g*NFIB* sh*CTRL*, SUM159PT sh1 *ERO1A* or SUM159PT sh2 *ERO1A* cells. In both graphs n = 3 biological replicates, means ± s.d., two-tailed Student's *t*-test, FC= fold change.

E Upper panel: Representative images of CD31-stained cells (brown) in lungs. Scale bar, 1 mm. Lower panel: Bar graphs showing lung metastases quantified as percentage of metastatic area per lung area and quantification of CD31 staining in the metastatic area. Means ± s.d., n = 8 sh*CTRL*, n = 5 per sh1 *ERO1A* and n = 6 sh2 *ERO1A*, two-tailed Student's *t*-test.

5.3.7 High *NFIB-ERO1A-VEGFA* expression is associated with poor prognosis in TNBC patients.

NFIB is expressed preferentially in basal-like breast cancers and is a potential prognostic factor in TNBC human breast cancer, as observed in previous reports (Liu et al., 2019; Moon et al., 2011). Analysis of METABRIC (Curtis et al., 2012) confirmed that *NFIB* expression was elevated in basal-like breast cancer, according to PAM50 classification, and in iC10 by integrated cluster classification (Dawson et al., 2013) (Fig 6A). Notably, we found that high expression levels of *NFIB* and the combined signature of *ERO1A* and *NFIB* in patients with breast cancer of the basal-like subtype were predictive of decreased distant metastasis-free survival and overall survival (Fig 6B-D). Furthermore, the combined signature of *ERO1A*, *NFIB, VEGFA* high expression correlated with reduced overall survival in breast cancer patients (Fig 6E).

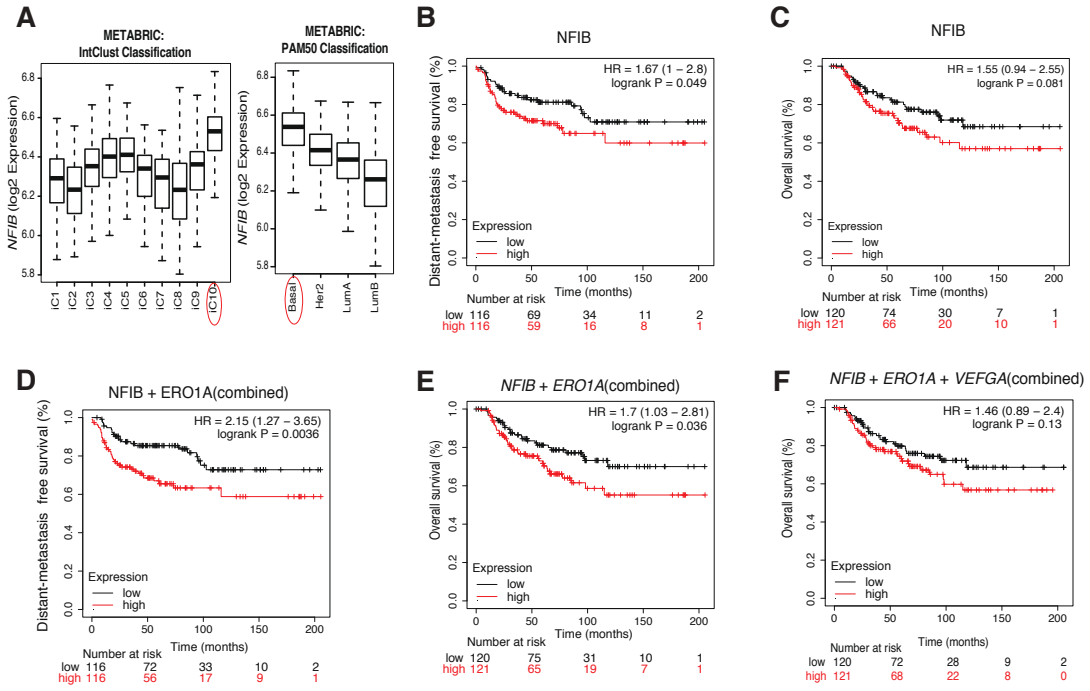


Figure 5-6 | *NFIB-ERO1A-VEGFA* overexpression is associated with reduced survival in TNBC patients.

A Higher *NFIB* expression observed in iC10 and basal breast cancer subtype compare to the other subtypes using intClust (Dawson et al., 2013) and PAM50 classifications in the METABRIC cohort.

B, C Distant-metastasis-free survival (B) and overall survival (C) plots generated using the Kaplan-Meier Plotter (Győrffy et al., 2010) based on the signal intensity of the *NFIB* probe (213029_at).

D, E Distant-metastasis-free survival (D) and overall survival (E) plots generated using the Kaplan-Meier Plotter based on the mean expression using the signal intensity of the *NFIB* (213029_at) combined with the *ERO1A* (218498_s_at) probes.

F Overall survival plot generated using the Kaplan-Meier Plotter based on the mean expression using the signal intensity of the *NFIB* (213029_at) combined with the *ERO1A* (218498_s_at) and *VEGFA* (210512_ s_at) probes.

In all the graphs the cut-off was automatically set to split patients in two groups (median), high and low. All the plots were generated using signal intensity of the different probes in Affymetrix microarray gene expression data from TNBC patients of The Cancer Genome Atlas. Number of patients (*n*) and *p* values (two tailed log-rank test) are presented in the panels.

5.4 Expanded view figures

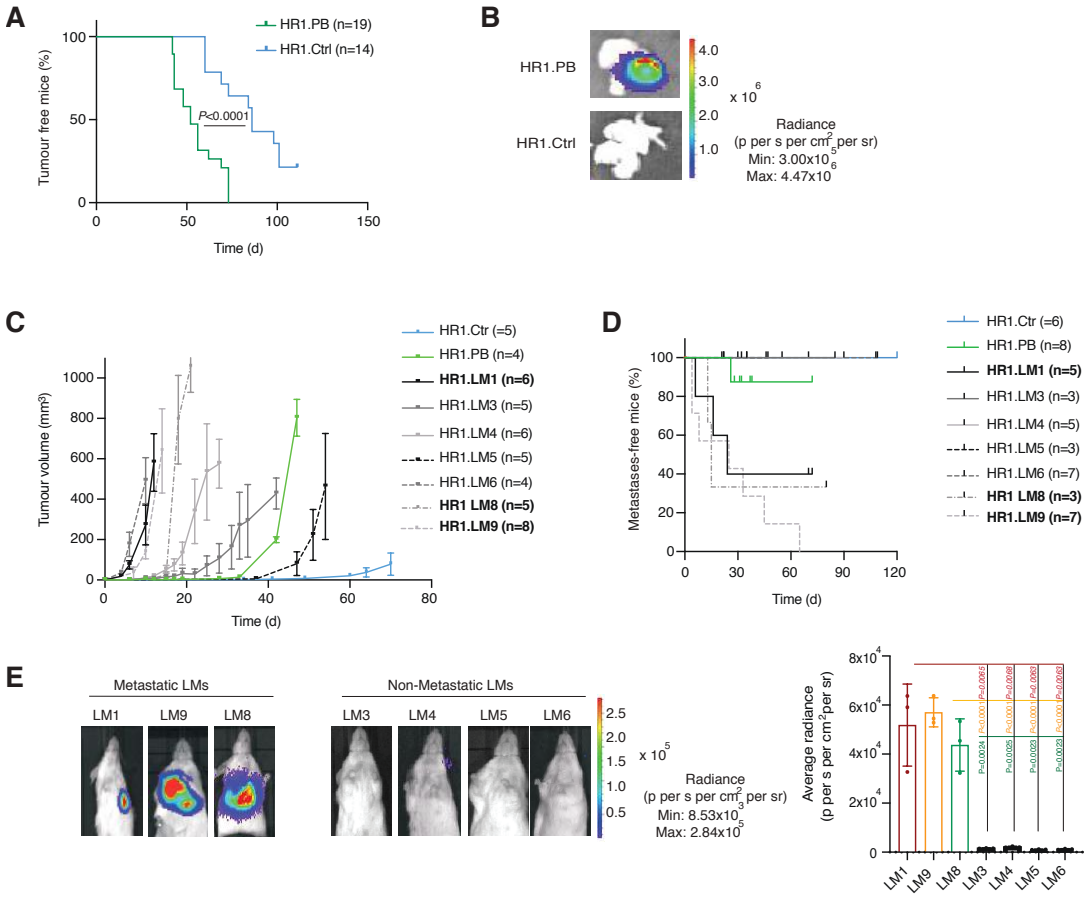


Figure 5-EV1 | *In vivo* unbiased transposon screen in the non-metastatic *PIK3CA*^{H1047R} mammary tumour-derived cells allows the isolation of aggressive lung metastatic lines.

A Kaplan-Meier plot depicting tumour incidence in NOD/SCID animals transplanted in the mammary fat pad wit 10^6 cells of HR1.Ctrl (n = 14) or HR1.PB cells (n = 19), two-tailed log-rank test.

B Representative bioluminescence acquisition of lungs from mice injected with HR1.PB and HR1.Ctrl cells as in A.

C Graph representing the kinetics of HR1.Ctrl (n = 5), HR1.PB (n = 4), HR1.LM1 (n = 6), HR1.LM3 (n = 5), HR1.LM4 (n = 6), HR1.LM5 (n = 5), HR1.LM6 (n = 4), and HR.LM9 (n = 8) tumour growth upon orthotopic injection of 250 x 10^3 cells in NOD/SCID mice.

D Kaplan-Meier plot showing metastasis onset after tumour removal in mice injected with HR1.Ctrl (n = 6), HR1.PB (n = 8), HR1.LM1 (n = 5), HR1.LM3 (n = 3), HR1.LM4 (n = 5), HR1.LM5 (n = 3), HR1.LM6 (n = 7), or HR1.LM9 (n = 7).

E Representative bioluminescence images (left panel) and bar plot quantification (right panel) of lung metastatic (LMs) and non-metastatic LM cell lines after primary tumour removal. n = 3 mice, means ± s.d., two-tailed Student's t-test.

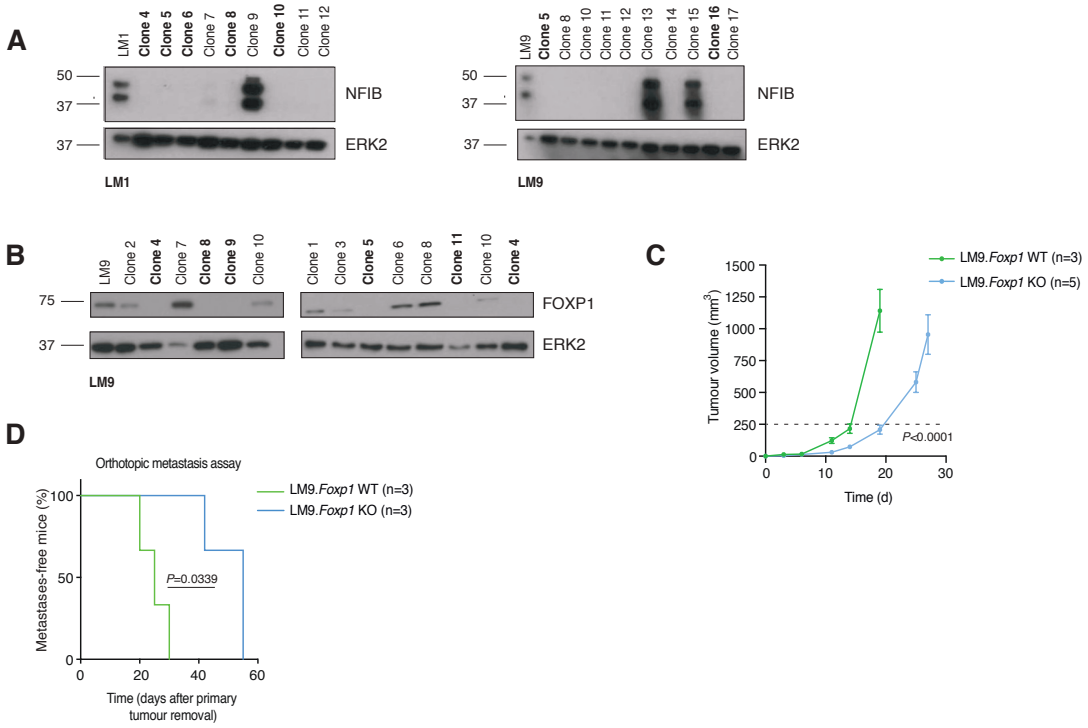


Figure 5-EV2 | *Foxp1* depletion decreases mammary tumour onset and metastasis.

A Immunoblots showing protein levels of NFIB in single-cell derived clones of LM1 and LM9 after *Nfib* knockout (KO). Single cell clones were pooled for the KOs (LM1 KOs clones: 4, 5, 6, 8, 10; LM9 KOs clones: 5,16; in bold). ERK2 served as a loading control.

B Immunoblots showing protein levels of FOXP1 in single-cell derived clones of LM9 after *Foxp1* knockout. Single cell clones were pooled (LM9 WT clones: 1,2,3,6,7,10; KOs: 4,5,8,9,11; in bold). ERK2 served as a loading control.

C Graph representing the kinetics of LM9 *Foxp1* WT (n = 3) or KO (n = 5) tumour growth. Curves show means of tumour volume ± s.d., two-tailed Student's t-test on the times to reach 250 mm^3 (dashed line).

D Kaplan-Meier plot depicting metastasis onset after tumour removal in mice injected orthotopically with LM9 *Foxp1* WT (n = 3) or KO (n = 3), two-tailed log-rank test.

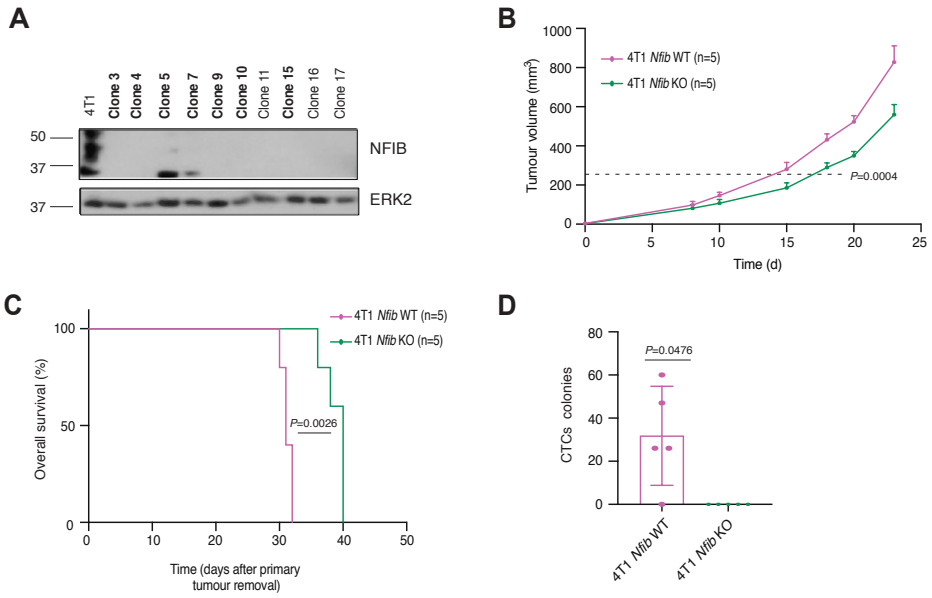


Figure 5-EV3 | *Nfib* ablation in a highly metastatic mammary cancer line increases overall survival in mice.

A Immunoblots showing protein levels of NFIB in single-cell derived clones of 4T1 after *Nfib* knockout. Single cell clones were pooled: WT clones: 5,7; KOs clones: 3, 4, 9, 10, 15; in bold. ERK2 served as a loading control.

B Graph representing the kinetics of 4T1 *Nfib* WT ($n = 5$) and 4T1 *Nfib* KO ($n = 5$) tumour growth upon orthotopic injection of 250 x 10^3 cells into BalbC animals. Curves show means of tumour volume ± s.d., two-tailed Student's *t*-test on the times to reach 250 mm^3 (dashed line).

C Kaplan-Meier plot depicting survival analysis after tumour removal from mice injected orthotopically with 4T1 *Nfib* WT ($n = 5$) or 4T1 *Nfib* KO ($n = 5$), two-tailed log-rank test.

D Quantification of CTCs colonies from mice orthotopically injected with 4T1 *Nfib* WT ($n = 5$) and 4T1 *Nfib* KO ($n = 5$). Means ± s.d., two-tailed Mann-Whitney U-test.

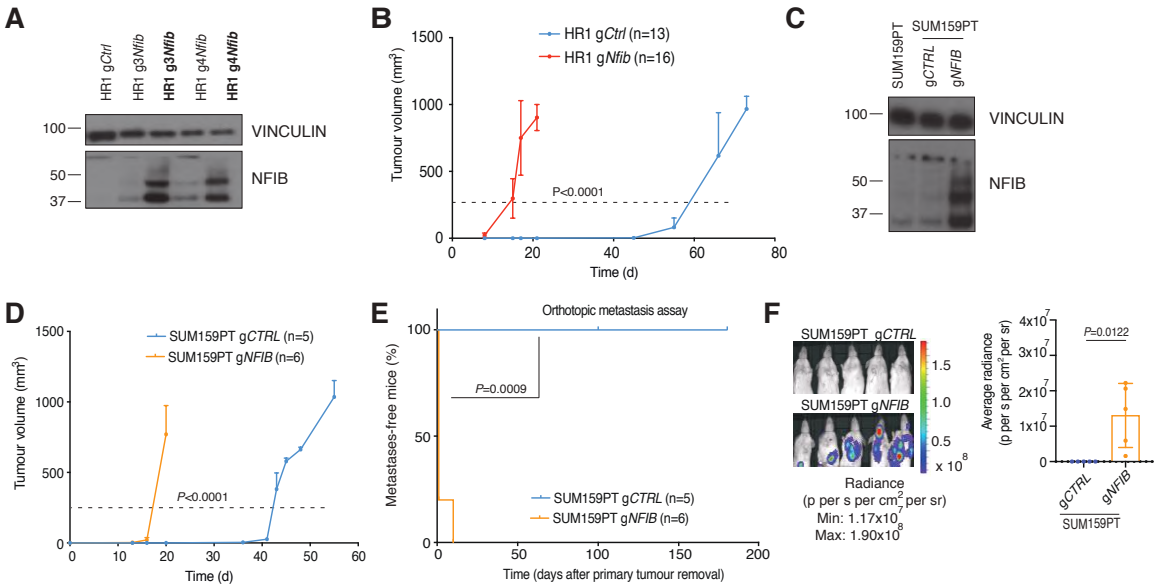


Figure 5-EV4 | Endogenous *Nfib/NFIB* overexpression in HR1 and SUM159PT cell lines decreases tumour latency and enhances metastasis.

A Immunoblot showing proteins levels of NFIB in HR1 cells overexpressing *Nfib* via the CRISPR Cas 9 SAM system (in bold: cell lines after two months of culture *in vitro*). VINCULIN served as a loading control.

B Graph representing the kinetics of HR1 g*Ctrl* (n = 13) and HR1 g*Nfib* (n = 16) tumour growth upon orthotopic injection of 250 x 10^3 cells in NOD/SCID mice. Curves show means of tumour volume ± s.d., two-tailed Student's t-test on the times to reach 250 mm^3 (dashed line).

C Immunoblots showing protein levels of NFIB in SUM159PT cells overexpressing *NFIB* via the CRISPR Cas 9 SAM system. VINCULIN served as a loading control.

D Graph representing the kinetics of SUM159PT g*CTRL* (n = 5) and g*NFIB* (n = 6) tumour growth upon orthotopic injection of 250 x 10^3 cells into NSG mice. Curves show means of tumour volume ± s.d., two-tailed Student's t-test on the times to reach 250 mm^3 (dashed line).

E Kaplan-Meier plot depicting metastasis onset after tumour removal from mice injected orthotopically with SUM159PT g*CTRL* (n = 5) or g*NFIB* (n = 6), two-tailed log-rank test.

F Representative bioluminescence images (left panel) and quantification (right panel) of mice at day 30 (SUM159PT g*CTRL*) and day 3 (SUM159PT g*NFIB*) after primary tumour removal. n = 5 mice, means ± s.d., two-tailed Student's t-test.

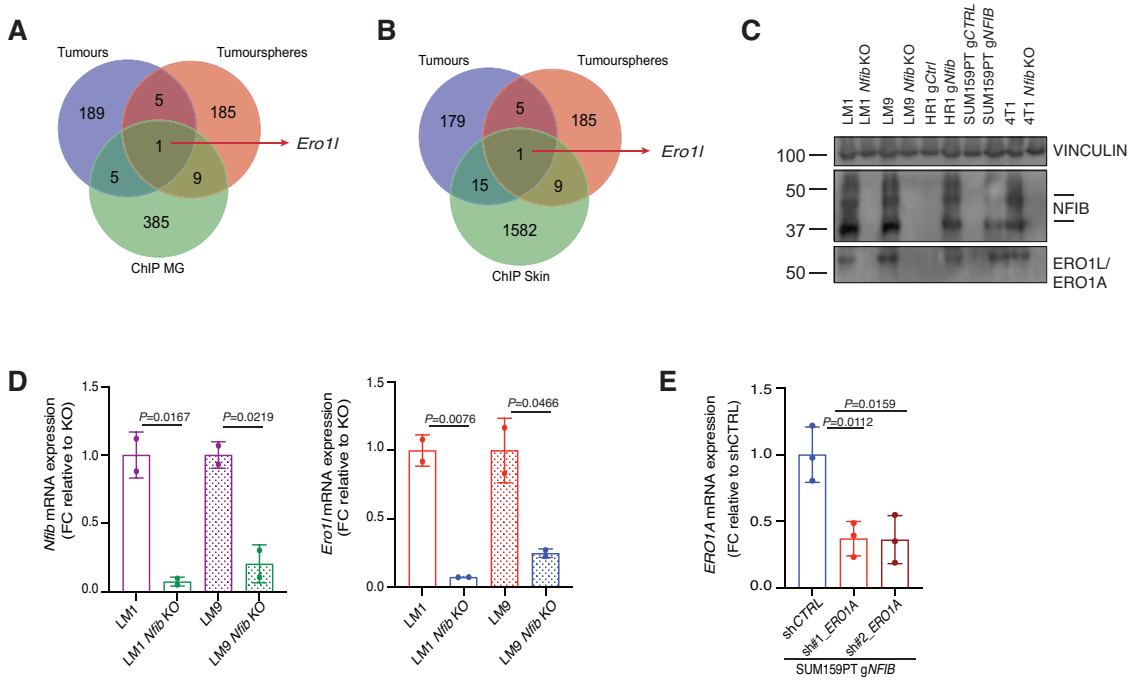


Figure 5-EV5 | *Nfib* increases *Ero1l/ERO1A* expression.

A, B Intersection of the top 200 differentially expressed genes (sorted by statistical significance) in tumourspheres and tumours (*Nfib* high-expression vs. the respective controls) and putative transcriptional targets of *Nfib* determined by ChIP-seq in mammary gland (Shin et al., 2016) (A) and epithelial-melanocyte stem cells (Chang et al., 2013) (B) (association rule: Basal+extension: 5000 bp upstream, 1000 bp downstream, 10000 bp max extension, curated regulatory domains included).

C Immunoblots showing proteins levels of NFIB and ERO1L/ERO1A in tumours derived from mammary fat pad injection of LM and LM *Nfib* KO, HR1g*Ctrl* and HR1 g*Nfib*, SUM159PT g*CTRL* and SUM159PT g*NFIB,* and 4T1 and 4T1 *Nfib* KO cell lines. VINCULIN served as a loading control.

D Bar graph representing mean *Nfib* and *Ero1l* mRNA expression in tumourspheres derived from LM and LM KO cancer cells at the first passage. *Ero1l* is downregulated upon *Nfib* KO in this model. In both graphs $n = 2$ biological replicates, means ± s.d., two-tailed Student's *t*-test, FC= fold change.

E Bar graph representing mean *ERO1A* mRNA expression in the SUM159PT g*NFIB* cell line upon *ERO1A* knockdown. Means ± s.d., $n = 3$ biological replicates, two-tailed Student's *t*-test, FC= fold change.

5.5 Discussion

Metastasis, the final, lethal hallmark of cancer, remains a major burden for breast cancer patients but there are numerous mechanisms that may prevent circulating cancer cells from colonizing distant organs. Only when these are circumvented and metastases develop is the prognosis of patients dismal (Massagué and Obenauf, 2016). Several interacting oncogenic pathways contribute to metastasis, but the exact molecular mechanisms of colonization are not fully understood. A detailed understanding of these molecular mechanisms is urgently needed.

Previous studies showed that *NFIB* governs epithelial melanocyte stem cell behaviour and facilitates melanoma cell migration and invasion (Chang et al., 2013; Fane et al., 2017). NFIB has also been implicated in models of small-cell lung carcinoma (SCLC) and breast cancer (Denny et al., 2016; Semenova et al., 2016; Campbell et al., 2018; Liu et al., 2019; Moon et al., 2011), and *NFIB* or *ERO1A* has been associated with different tumour types (Becker-Santos et al., 2017; Yang et al., 2018; Zhou et al., 2017). NFIB was shown to enhance TNBC cell survival and progression by suppressing *CDKN2A* (Liu et al., 2019), and to confer estrogen independency in estrogen receptor-positive breast cancer models (Campbell et al., 2018).

Here, we provide new insights into the effects of the NFIB-ERO1A axis on breast cancer progression to metastasis. Via an unbiased *in vivo* PB functional genetic screen, we have shown that NFIB induces mammary cancer metastatic progression and colonization. We have collected functional and mechanistic evidence that NFIB is a mammary cancer metastatic transcriptional regulator. First, depletion of NFIB delayed mammary tumour growth and abrogated metastases in three different cell lines when using both the orthotopic and experimental metastasis assays. Second, *Nfib* KO reduced tumoursphere formation and the frequency of tumour-initiating cells. Third, endogenous Nfib/NFIB overexpression was

sufficient to evoke metastasis in non-metastatic murine and human mammary cancer lines in both the orthotopic and experimental metastasis assays. It also decreased overall survival of the animals.

Mechanistically we have discovered, and provide functional validation, that ERO1A is critical for NFIB-evoked metastatic colonization in mammary cancer. The NFIB-ERO1A axis promotes angiogenesis at secondary sites via increased *VEGFA* expression, resulting in a permissive microenvironment for colonization. Clinically, we have shown that high expression of the *NFIB/ERO1A/VEGFA* axis correlates with reduced breast cancer patient survival. These recorded effects of the NFIB-ERO1A axis on breast cancer progression to metastasis identify a targetable network for cancer therapy.

5.6 Materials and methods

Cell lines

LB-mHR1 is a mouse mammary cancer cell line derived from a transgenic mouse expressing *PIK3CA*^{H1047R} under the CAG-promoter (Meyer et al. 2011). In brief, a mammary tumour was dissociated into single cells through mechanical and enzymatic digestion. The cells were sorted using GFP and established in DMEM supplemented with 10% FBS. To generate HR1.PB cells, LB-mHR1 cells were engineered with the PB system (Ivics et al., 2009, Rad et al., 2010) by transduction with the pFS250-PBase vector (HR1.Ctrl) followed by transfection with an ATP1 construct. To induce transposon mobilization and generate pools of mutagenized cells, HR1.PB cells were treated with dox for up to 96 h *in vitro*. LM (1 to 9) cell lines are lung-derived tumour cell lines isolated from NOD/SCID mice bearing HR1.PB derived mammary tumours. In brief, lungs were dissociated into single cells and cultured until tumour cell colonies became apparent. For increased purity, tumour cells were sorted for GFP. HEK293T and 4T1 cell lines were purchased from the ATCC and cultured according to the ATCC protocol. LB-mHR1, 4T1, LM1-9, and HEK293T cells lines were cultured in DMEM supplemented with 10% FCS. SUM159PT cells were kindly provided by Dr. Charlotte Kupperwasser (Boston, USA). SUM159PT cells were cultured in Ham's F12 with 5% fetal calf serum (Gibco, Invitrogen), 5 µg/ml bovine insulin (Sigma), 1 µg/ml hydrocortisone (Sigma), 1x penicillin/streptomycin (Gibco, Invitrogen). SUM159PT cell line identity was confirmed and routinely tested using short tandem repeat (STR) sequencing; all cell lines were tested routinely for mycoplasma contamination.

Animal experiments

All *in vivo* experiments were performed in accordance with the Swiss animal welfare ordinance and approved by the cantonal veterinary office Basel-Stadt. Female NOD/SCID, NSG, and BALB/c mice were maintained in the Friedrich Miescher Institute for Biomedical Research

and the Department of Biomedicine animal facilities in accordance with Swiss guidelines on animal experimentation. Mice were maintained in a sterile controlled environment (a gradual light–dark cycle with light from 7:00 to 17:00, 21-25°C, 45-65 % humidity). For orthotopic engraftment of cancer cell lines in the limiting dilution assay, 250 x 10^3,10 x 10^3, 2 x 10^3, 500 and 200 LM cells were suspended in 50 µl Matrigel:PBS (1:1) and injected into the fourth mouse mammary gland of six- to eight-week-old female NOD/SCID mice. The frequencies of tumour-initiating cells in the different conditions were calculated and compared statistically using the Extreme Limiting Dilution Analysis (ELDA) online tool (Hu and Smyth, 2009). For the PB *in vivo* screen, HR1.Ctrl and HR1.PB cells (1 x 10^6) were resuspended in 50 µl Matrigel:PBS (1:1) and injected into the mammary fat pads of four- to eight-week-old female NOD/SCID mice. For orthotopic engraftment of cancer cell lines, LB-mHR1 or LM cells (250 x 10^3 cells) were resuspended in 50 µl Matrigel:PBS (1:1) and injected into the mammary fat pads of four- to eight-week-old female NOD/SCID mice. SUM159PT or 4T1 (250 x 10^3 cells) cells were injected into the mammary fat pads of four- to eight-week-old female NSG or BALB/c mice, respectively. Tumours were measured with Vernier calipers and volume calculated as 0.5 x [(large diameter) x (smallest diameter)^2]. Tumours were resected before they reached 1500 mm^3 and mice were monitored regularly for signs of metastatic outgrowth and distress. For survival studies, animals were sacrificed when tumours reached 1500 mm^3 or when they showed any signs of distress (e.g., breathing disorders, weight loss, or immobility). All orthotropic experimental procedures (tumour resection and tumour cell implantation) were undertaken on anaesthetized mice by a single investigator according to protocols approved by the cantonal veterinary office Basel-Stadt. Experimental metastasis assays were performed by injecting 100 x 10^3 cells suspended in 100 µl of PBS into tail veins (i.v.). After intravenous injection of LM, LB-mHR1, or SUM159PT cells, *in vivo* bioluminescence imaging was performed to confirm injection and to monitor metastatic outgrowth. For bioluminescent

imaging, mice were injected *i.p*. with 100 µl of D-luciferin (15 mg/mL, Biosynth L8220). Mice were anesthetized with isoflurane (2% in 1 L/min oxygen) and bioluminescence imaging performed using an IVIS Lumina XR instrument (Caliper LifeSciences) upon injection of luciferin. Acquisition times ranged from 3 to 10 min.

Splinkerette PCR for the amplification of transposon integration sites

For PB sequencing, we adapted the splinkerette PCR protocol described previously (Friedrich et al., 2017). Sequenced samples: tumour samples (16 primary tumours) and 18 metastatic samples (3 lung-metastases, 6 lung macro-metastases and 9 LM cell lines). Genomic DNA was isolated (using DNeasy Blood and Tissue Kit; Qiagen), sheared to a fragment length of 250 bp with a Covaris sonicator. After end repair and A-tailing, purified DNA fragments were ligated to a splinkerette adaptor (obtained after annealing of top 5′-gttcccatggtactactcatataatacgactcactataggtgacagcgagcgct-3′ and bottom 5′-gcgctcgctgtcacctatagtgagtcgtattataatttttttttcaaaaaaa-3′). Transposon-containing fragments were enriched by 18 cycles of transposon-specific PCR1 for the 5′ transposon ends in a unique library (5′-gacggattcgcgctatttagaaagagag-3′ for the 5′ arm of PB, and common splinkerette primer 5′-gttcccatggtactactcata-3′). Bar coding of individual samples and completion of Illumina adaptor sequences were achieved by an additional 12 cycles (primary tumours) and 18 cycles (metastatic samples) of transposon-specific PCR and a custom array of 35 unique bar-coding primers. For the 5′ arm, we used 5′-aatgatacggcgaccaccgagatctacacatgcgtcaattttacgcagactatc- 3′ and for the splinkerette side, we used 5′-caagcagaagacggcatacgagatcggtXXXXXXXXtaatacgactcactatagg-3′ primers. The Xs represent the bar code of 8 nucleotides. After magnetic bead purification (Beckman Ampure XP), libraries were assembled in two pools and sequenced on an HiSeq Desktop Sequencer (Illumina): HiSeq 2500 and MiSeq, Rapid Run, Paired-End, 2 x 100 bp, with 15% PhiX; 7 pM was loaded on to the instrument.

Mapping of insertion sequences to the mouse genome and identification of integration sites

Pre-processing and alignment: Paired-End reads were first pre-processed by removing the expected transposon-derived sequence (5'-TAGGGTTAA-3') from the beginning of the first read (read-pairs with non-matching first reads were discarded; typically around 1%) using the preprocessReads function from QuasR (version 1.12.0). Read pairs were then aligned to the mouse genome (BSgenome.Mmusculus.UCSC.mm10) using the QuasR qAlign function and parameters "-m 1 --best --strata --maxins 1000"; only uniquely mapping pairs with up to 1000 bp between-pair distance were reported. Mapping rates were recorded and non-mapped read pairs were further aligned against the non-mobilized transposon sequence to estimate the probable fraction of read pairs that originate from non-mobilized copies of the transposon. For each aligned read pair, the *piggyBac* insertion coordinate was identified as the coordinate of the first (most 5') base of the first read.

Integration site identification and quantification: For each unique integration site, the number of distinct supporting alignments (distinct read pairs), and the genomic sequence from the four base pairs on the same strand as the first read directly upstream of the insertion site were recorded. Only integration sites with the expected upstream TTAA sequence were used for the downstream analysis.

Association of integration sites with genes: Coordinates of known genes (exons/introns, 5'-untranslated region [UTR], coding sequence [CDS], 3'UTR) were obtained from the TxDb.Mmusculus.UCSC.mm10.knownGene Bioconductor package (version 3.2.2). Promoter regions were defined as regions 2,000 bp upstream of known transcript start sites. Integration sites were matched against these genomic regions to identify overlaps on any strand and orientation, selecting the first overlap in the case of multiple overlaps. Integration sites were classified hierarchically as follows: sites without overlaps to any transcript were labeled as

promoter sites (in the case of an overlap with a promoter region) or intergenic sites. All other sites were labeled with the first region type that they overlapped, in the following order: 5'UTR, CDS, 3'UTR, intron, or ncRNA (defined as an overlap with a transcript without annotated CDS). Sites were further labeled according to their orientation with respect to the associated gene (same or opposite). Finally, sites were grouped according to the gene they overlapped (including promoter sites) or, for intergenic sites, according to the pair of flanking genes.

Selection of the enriched unique integration sites: For each insertion the diversity has been calculated (the number of distinct alignment-pairs starting at the insertion coordinate) (Chapeau et al., 2017). For each gene the diversity was summed up resulting in the number of unique insertions in a gene. Finally, this value was divided by the total number of unique insertions for a given sample. In order to find putative metastatic genes, we looked at the genes with a higher number of unique insertions in the metastatic than in the tumour samples.

Genome editing by CRISPR-Cas9

A single sgRNA that produced a frameshift mutation in first and second exons was designed using the Zhang's lab online CRISPR (http://crispr.mit.edu) and cloned into a modified PX330 (Addgene plasmid 42230) in which the puromycin cassette was replaced by red fluorescent protein (RFP; provided by the group of M. Bühler at the FMI). The sgRNA sequences selected for *Nfib* (based on the lowest number of predicted off-targets and highest predicted efficiency) were, guide 1: 5'-CACCGCTCCGGGAAAGTGCGTTTTA-3' and 5'-AAACTAAAACGCACTTTCCCGGAGC-3'; guide 2: 5'-CACCGTAGGCAATTGCACGGACGTG-3' and 5'-AAACCACGTCCGTGCAATTGCCTAC -3. The sgRNA sequences selected for *Foxp1* were, guide1: 5'-CACCGCTTCGTGACACTCGGTCCAA-3' and 5'-AAACTTGGACCGAGTGTCACGAAGC-3'; guide 2: 5'-CACCGTAGTAAGTGGTTGCCACCGC-3' and 5'-

AAACGCGGTGGCAACCACTTACTAC-3'. The sgRNA vectors were transduced into LM and 4T1 cells and the cells sorted for RFP positivity into 96-well plates. Single cell clones were expanded and screened by immunoblotting. Different numbers of clones were pooled in equal proportions to minimize undesired off-target and clonal effects (LM1 KO = 5, LM9 KO =2, 4T1 WT = 2, 4T1 KO = 5). For human *NFIB* overexpression, we used SAM-engineered Cas9 activation complexes (Konermann et al., 2015). LB-mHR1 and SUM159PT cells were lentiviral infected in order to express the full murine and human *NFIB* from an endogenous promoter and maintained in culture for two months. sgRNA sequences were designed using the Zhang's lab online CRISPR (http://sam.genome-engineering.org) and cloned into the lentiviral vector Addgene; plasmid 61427. Lenti-sgRNA-MS2 was digested with BsmbI and purified in a gel. The two annealed primers were ligated using the golden gate reaction. The sgRNA sequences selected for mouse *Nfib* were, guide 2: 5'-CACCGCAGGAGGAGGAGGAGTAAAG-3' and 5'-AAACCTTTACTCCTCCTCCTCCTGC-3'; guide 3: 5'-CACCGTGGGGGAGGCGCGCGGGAGG-3' and 5'-AAACCCTCCCGCGCGCCTCCCCCAC -3'; guide 4: 5'-CACCGTGTGGAGAGGCTGGTGCAAA-3' and 5'-AAACTTTGCACCAGCCTCTCCACAC-3'. The sgRNA sequences selected for human *NFIB* were, guide 1: 5'-CACCGAGCTGAGCCATCCATTCCTC-3' and 5'-AAACGAGGAATGGATGGCTCAGCTC; guide 2: 5'-CACCGACTAGGCTTGCAGTAAACGC-3' and 5'-AAACGCGTTTACTGCAAGCCTAGTC; guide 3: 5'-CACCGGAAGAGACTTGTCAGTATA-3' and 5'-AAACTATACTGACAAGTCTCTCCC-5'.

Lentiviral infections

The ATP1 vector was described previously (Rad et al., 2010). The transposon has PB inverted terminal repeats and can therefore be mobilized with the transposon system. ITRs were cloned into pBlueScript and the following genetic elements introduced between the ITRs of the transposon: Carp β-actin splice acceptor (CβASA), En2SA splice acceptor from exon 2 of the mouse Engrailed-2 gene, Lun-SD from exon 1 of the mouse Foxf2 gene, and two bidirectional SV40 polyAs. Promoter elements carried by the transposons were unique to individual transposons: CAG (CMV immediate early enhancer and chicken beta-actin gene promoter) for ATP1. The piggyBac transposase was cloned out of the Super PiggyBac Transposase expression (System Biosciences) by PCR (primers 5' AGCTAGCACCGGTCGGAATTGTACCCAATTCGTTAAG 3' and 5' AGAATTCTTAATTAATTCTGGCGGCCGTTACG 3') and cloned into a doxycycline-inducible vector pFS250 (a kind gift from Novartis). For human *ERO1A* downregulation, we used two shRNA constructs (V2THS_85712 and V2THS_85710: Dharmacon pTRIPZ). Non-targeting shRNAs (pTRIPZ) were used as controls. A dual green fluorescent protein-luciferase 2 reporter (eGFP-Luc2 (Liu et al., 2010)) vector was used for *in vivo* bioluminescence imaging experiments. Lentiviral batches were produced using PEI transfection on 293T cells as previously described (Britschgi et al., 2017). The titer of each lentiviral batch was determined in 4T1, LM, FMI-LB-mHR1, and SUM159PT cells. Cells were infected overnight in the presence of polybrene (8 µg/ml). p250.PBase selection was performed with 500 µg/ml Neomicin G418 (InvivoGen) and applied 48 h after infection. ATP1 selection was performed with 2.5 µg/ml puromycin (Sigma) and applied 48 h after transfection. The SAM engineered Cas9 activation complex (Konermann et al., 2015) consists of three lentiviral vectors: Lenti MS2-P65-HSF1_Hygro (Addgene plasmid 61426), Lenti-sgRNA-MS2_Zeo (Addgene; plasmid 61427), and Lenti dCAS-VP64_Blast (Addgene 61425). Cells were infected first with

a 1:1 mass ratio of Lenti MS2-P65-HSF1 and Lenti dCAS-VP64 at a MOI of 10 viral particles per cell; selection was carried out in 1 µg/ml Blasticidin (Gibco) and 500 µg/ml Hygromicin (InvivoGen) after a 24-h transduction. Selection in 500 µg/ml Zeomicin (Invitrogene) was also applied after a 24-h infection with Lenti-sgRNA-MS2_Zeo. Cells were cultured for two months i*n vitro* before performing experiments.

Fluorescence-activated cell sorting

For FACS, cell lines were detached using trypsin-EDTA, resuspended in growth medium and counted. Cells were passed through a 40-µm strainer (Falcon) and resuspended in PBS with 1% FCS. DAPI (0.2%, Invitrogen) was added (1:250) 2 min before cell sorting. Single cells were gated on the basis of their forward and side-scatter profiles and pulse-width was used to exclude doublets. Dead cells (DAPI bright) were gated out and RFP and GFP bright cells were selected. FACS was carried out with a BD FACSAria III (Becton Dickinson) using a 70-µm nozzle for SUM159PT, 4T1, HR1 and LM cell lines.

Immunoblotting

Cells were lysed in RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 1x protease inhibitor cocktail (Complete Mini, Roche), 0.2 mM sodium orthovanadate, 20 mM sodium fluoride, and 1 mM phenylmethylsulfonyl fluoride. Extracted proteins were measured using a DC protein assay kit (BioRad #5000112) and their concentrations adjusted. Whole-cell lysates, immunoprecipitates or nuclear cell lysates (30-50 µg) were subjected to 8% SDS– PAGE, transferred to PVDF membranes (Immobilon-P, Millipore), and blocked for 1 h at room temperature with 5% milk in PBS–0.1% Tween 20. Membranes were then incubated overnight with antibodies as indicated and exposed to secondary HRP-coupled anti-mouse or -rabbit antibody at 1:5,000–10,000 for 1 h at room temperature. For each of the blots presented, the results represent at least three independent experiments. The following antibodies were used: anti-NFIB (Sigma,

HPA003956-100UL), anti-Vinculin (Abcam, ab18058), anti-ERK2 (sc-1647, Santa Cruz), anti-FOXP1 (Cell signaling, 4402), and anti-ERO1L (Abnova, H0003001-M01, Clone 4G3).

RNA isolation and qPCR

Total RNA was extracted using a Qiagene, RNeasy Plus Mini kit (cat. number 74136) according to the manufacture's protocol. Total RNA (1 µg) was transcribed using the iScript cDNA synthesis kit (BioRad, cat. number 170-8891). PCR and fluorescence detection were performed using the StepOnePlus Sequence Detection System or the 7500 ABI Fast Cycler (Applied Biosystems) according to the manufactures' protocols in a reaction of 20 µl containing 1x TaqMan Universal PCR Master Mix (Applied Biosystems) and 25-50 ng cDNA. Primetime qPCT IDT assays (Integrated DNA technologies) were used for quantification of *Nfib* (Mm.PT.58.16634012), *Gapdh* (Mm.PT.39a.1), *Foxp1* (Mm.PT.58.2734907), *Ero1l* (Mm.PT.58.6905093), *Vegfa* (Mm.PT.58.1400306), *NFIB* (Hs.PT.58.40046929), *ERO1A* (Hs.PT.58.2554238), *VEGFA* (Hs.PT.58.1149801), and *HPRT1* (Hs.PT.58.45621572). All measurements were performed in technical duplicate and the arithmetic mean of the Ct values used for calculations: target gene mean Ct values were normalized to the respective housekeeping genes (Hprt1 or Gapdh), mean Ct values (internal reference gene, Ct), and then to the experimental control. The values obtained were $2^{-\Delta\Delta Ct}$ expressed as fold changes in regulation compared with the experimental control using the $2^{-\Delta\Delta Ct}$ method of relative quantification.

Tumoursphere cultures

LM and HR1 cells were plated in ultra-low attachment plates (Corning) for six days at 10,000 cells per mL in DMEM:F12 supplemented with 1x B27 (Gibco, Invitrogen), 20 ng/mL human or mouse EGF (PeproTech), 20 ng/mL basic FGF (PeproTech), and 1x penicillin/ streptomycin (Gibco, Invitrogen)(Hilsenbeck et al., 2008). Primary tumourspheres were dissociated with 0.05% trypsin and replated at the same density for six further days for secondary tumoursphere

assessment; additional rounds of culture were performed. All measurements were performed in technical triplicates.

Circulating tumour cell quantification

Circulating tumour cells (CTCs) were isolated from the peripheral blood of animals bearing tumours just before tumour resection. Peripheral blood was plated in DMEM medium supplemented with 10% FCS and colonies counted on day 10 of culture. The number of CTCs was expressed as the total number of colonies in the dish divided by the volume of blood taken.

RNA sequencing and analysis

Tumours RNA preparation and sequencing: Sorted GFP positive cells from tumours were collected and total RNA was extracted using a Qiagene, RNeasy Plus Mini kit (cat. number 74136) according to the manufacture's protocol. RNA integrity was measured on an Agilent 2100 Bioanalyzer using RNA Pico reagents (Agilent Technologies). The library was prepared using Illumina TruSeq stranded mRNA-seq preparation kit (according to recommendations from the manufacturer). Library quality was measured on an Agilent 2100 Bioanalyzer for product size and concentration. Single-end libraries were sequenced by an Illumina HiSeq 2500 (50-nt read length).

Tumourspheres RNA preparation and sequencing: cells were dissociated with 0.05% trypsin, and total RNA was extracted using a Qiagene, RNeasy Plus Mini kit (cat. number 74136) according to the manufacture's protocol. RNA integrity was measured on an Agilent 2100 Bioanalyzer using RNA Pico reagents (Agilent Technologies). The library was prepared using Illumina TruSeq stranded mRNA-seq preparation kit (according to recommendations from the manufacturer). Library quality was measured on an Agilent 2100 Bioanalyzer for product size and concentration. Single-end libraries were sequenced by a NextSeq 500 (76-nt read length).

Analysis: reads were aligned to the mouse genome (UCSC version mm10) with STAR(Dobin et al., 2013). The output was sorted and indexed with samtools. Stand-specific coverage tracks per sample were generated by tiling the genome in 20-bp windows and counting 5' end of reads per window using the function bamCount from the bioconductor package bamsignals (https://bioconductor.org/packages/release/bioc/html/bamsignals.html). These window counts were exported in bigWig format using the bioconductor package rtracklayer (https://bioconductor.org/packages/release/bioc/html/rtracklayer.html). The qCount function of QuasR (https://bioconductor.org/packages/release/bioc/html/QuasR.html) was used to count the number of reads (5' ends) overlapping the exons of each gene, assuming an exon union model. Differential gene expression analysis was performed using the limma-voom framework (Carriço et al., 2011).

Immunohistochemistry

All xenograft tissues were fixed in 4% paraformaldehyde for 24 h at 4°C. Samples were then dehydrated, embedded in paraffin, sectioned (3-4 µm), and processed for haematoxylin/eosin staining and for immunohistochemistry. All immunohistochemistry experiments were performed using a Ventana DiscoveryXT instrument (Roche Diagnostics) following the Research IHC DAB Map XT procedure. For Nfib staining, slides were pretreated with CC1 for 90 min. NFIB primary antibody (1:250) was incubated for 44 min, followed by secondary antibody incubation and detection. NFIB scores were expressed as positive=1 and negative=0. The RUO Discovery Universal procedure was used for CD31 staining, with a CC1 pre-treatment (40 min) and incubation with a rat anti-CD31 antibody (SZ31 dianova, 1:50) for 1 h at 37°C. Next, a polymer Immpress goat anti-rat (Vector Lab, MP-7444) was applied for 28 min at 37°C and the peroxidase reaction revealed with the Discovery ChromoMap DAB kit (Ventana, Roche diagnostics). Counter staining was performed with hematoxylin II and bluing

reagent (Ventana, Roche diagnostics). Two images of 2-8 lung metastases of each condition were captured and quantified manually or with Image j (Fiji).

CD31 quantification

The images were analysed using Fiji open source software(Gao et al., 2013). First, the lung area for each tissue section was measured by manual selection. Second, for each metastasis in each tissue section, a region of interest (ROI) was selected manually. Each ROI was then subjected to background subtraction (rolling ball with a 50-px radius) followed by color deconvolution using [H DAB] vectors. The resulting color channel (2) corresponding to the DAB stain was then thresholded (using the value for all images [222]) to measure the CD31-positive areas.

Microscopy image acquisition

For immunohistochemistry sections, images were captured using a Nikon Ni-e upright microscope coupled to a PRIOR slide loader. Acquisition was performed with a 4x AIR objective with a DS-Fi3 camera using NIS software. Quantification of the images was performed using Fiji. Phase-contrast imaging used an inverted Zeiss Axioscope (10x, NA =0.25 and 5x, NA= 0.15) equipped with an Axiocam 503 mono 60N-C camera (pixel size 4.54 µm) and images were acquired using the Zen lite software.

Data analysis from publicly available datasets

We used cBioPortal (Cerami et al., 2012, Gao et al., 2013) for the *NFIB* and *ERO1A* expression correlation study with publicly available data (Curtis et al., 2012, Pereira et al., 2016). The results here are based in part on data generated by the TCGA Research Network (https://www.cancer.gov/tcga). Relapse-free survival, distant-metastasis-free survival and overall survival of *NFIB* and *ERO1A* were generated using the 2017 version of KMplotter (Győrffy et al. 2010) (http://kmplot.com/analysis/index.php?p=service&cancer=breast). Venn

diagrams were produced using Bioinformatics & Evolutionary Genomics http://bioinformatics.psb.ugent.be/webtools/Venn/.

Statistical data analysis

The standard laboratory practice randomization procedure was used for cell-line groups and animals of the same age and sex. The investigators were not blinded to allocation during experiments and outcome assessment. The number of mice was calculated by power analysis using data from small pilot experiments. Values represent the means ±s.d., unless stated otherwise. *P* values were determined using unpaired two-tailed *t*-tests and statistical significance set at P=0.05. The variance was similar between the groups compared. Biological replicates correspond to different cell lines and tumour material. Technical replicates are tests or assays run on the same sample multiple times. For statistical analysis and for reporting of the number of experimental entities, biological replicates were used. The means of technical replicates, if available, were used for analysis and visualization. Data were tested for normal distribution, Student's t-tests and two-way ANOVA (if normally distributed) or nonparametric Mann-Whitney U-test or Wilcoxon tests were applied unless stated otherwise. Kaplan–Meier plots were generated using the survival calculation tool from Graphpad Prism and significance was calculated using the two-tailed log-rank test at $P < 0.05$. For the analysis of tumour growth, we extracted for each individual sample the time to reach a tumour volume of 250 mm^3 as follows: a linear model was fit to the log of the tumour volume as a function of time using all volumes greater than zero and least squares fitting in R (lm function). The resulting fit typically had an R-squared of 0.7 or greater. The estimated times were then used to compare conditions, either in a two-way ANOVA or Student's t-test, as indicated.

Data availability

Tumours mRNA-seq data (GEO accession GSE144392): Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144392

Enter token gxifiyyytbunxgf into the box

Tumourspheres mRNA-seq data (GEO accession GSE144393):

Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144393

Enter token efcfoeiynfmfzox into the box

piggyBac screen genomic GEO accession (GSE144898):

Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144898

Enter token ivebimgmpncrtsz into the box

Acknowledgements

We thank members of the Bentires-Alj laboratory for advice and discussions. The authors are grateful for the support of Genomics Core Facility at FMI and the BSSE-ETHZ, to the Animal facility of the University of Basel and FMI. We thank S. Bichet (FMI) for help with immunohistochemistry, L. Sauteur for developing algorithm for immunohistochemistry quantification, G. Galli for discussion and advice, S. Smallwood and S. Dessus-Babus for optimizing and running the PB-seq experiments. We thank C. Kuperwasser and S. Ethier for the SUM159 cells. Research in the Bentires-Alj laboratory is supported by the Swiss Initiative for Systems Biology- SystemsX, the European Research Council (ERC advanced grant 694033 STEM-BCPC), the Swiss National Science Foundation, Novartis, the Krebsliga Beider Basel, the Swiss Cancer League, the Swiss Personalized Health Network (Swiss Personalized Oncology driver project) and the Department of Surgery of the University Hospital Basel.

Author contributions

F.Z., P.R.M., M.B.-A. conceived the study and wrote the manuscript. F.Z., P.R.M., designed and performed the experiments, analysed the data and interpreted the results. P.-A.D.M. performed computational analysis of RNA-seq, analysed the data and interpreted the results. C.J. performed intersection between RNA-seq data and Nfib putative targets, analysed the data and interpreted the results. A. S. and M.B.S performed computational analysis of RNA-seq and PB-seq experiments, analysed the data and interpreted the results. M.-M.C. performed and optimized immunohistochemical staining, analysed the data and interpreted the results. L.B. generated LB-mHR1 cell line. J.P.C. performed experiments for the generation of HR1-PB cell lines. T.E. performed PB library preparation and quality control for PB-seq. M.R.J. contributed to the conceptual development of the project. R.R provided PBase and ATP1 vectors and contributed to the conceptual development of the project. All authors read and approved the final manuscript.

Disclosure of potential conflicts of interest

F.Z., P.A.D.M., C.J., A. S., M.-M.C., R.R., and M.B.-A. declare no competing interests. P.R.M, L.B., J.P.C., T.E., M.R.J. and M.B.S. are employees of Novartis Pharma AG.

6 | RESULTS PART II

6.1 Results

6.1.1 Generation of a conditional PB/*PIK3CA*^{H1047R} mammary cancer mouse model.

To identify *PIK3CA*^{H1047R} mutant collaborative mechanisms involved in breast cancer progression we used an unbiased *piggyBac* (PB) genome-wide mutagenesis system suitable for *in vivo* cancer screens (Rad et al., 2010). Particularly, we developed a conditional PB/*PIK3CA* mutant pregnancy induced mouse model by expressing the PB system and *PIK3CA*^{H1047R} mutation in the mammary epithelial cells (**Fig 6-1**). First, PB-transposase animals were crossed to WAPiCre mice in which the expression of recombinase Cre is controlled by the whey acidic protein (WAP) promoter. Second, ATP1 transposon mice were crossed to *PIK3CA*^{H1047R} animals (**Fig 6-1A**). We used a bi-functional transposon in which the cargo was modified with a promoter element, a splice donor, two splice acceptors facing in opposite orientation and bidirectional SV40 polyAs (**Fig 6-1B**) (Rad et al., 2010). Chromosome 10 harbors the transposon donor locus in 20 copies (Rad et al. 2015; Friedrich et al. 2017). For activation of transposition and mutation expression we finally crossed PB-transposase/WAPiCre and *PIK3CA*^{H1047R}/ATP1 transposon mutant mice generating W/P/A/H transgenic mice. In quadruple-transgenic mice, once the mouse is pregnant *PIK3CA*^{H1047R} and PB transposase are expressed. Consequently, the transposase can mobilize the transposon specifically in the secretory mammary epithelial cells, where the WAP is manly active. Mice lacking the ATP1 transposon served as control (W/P/H).

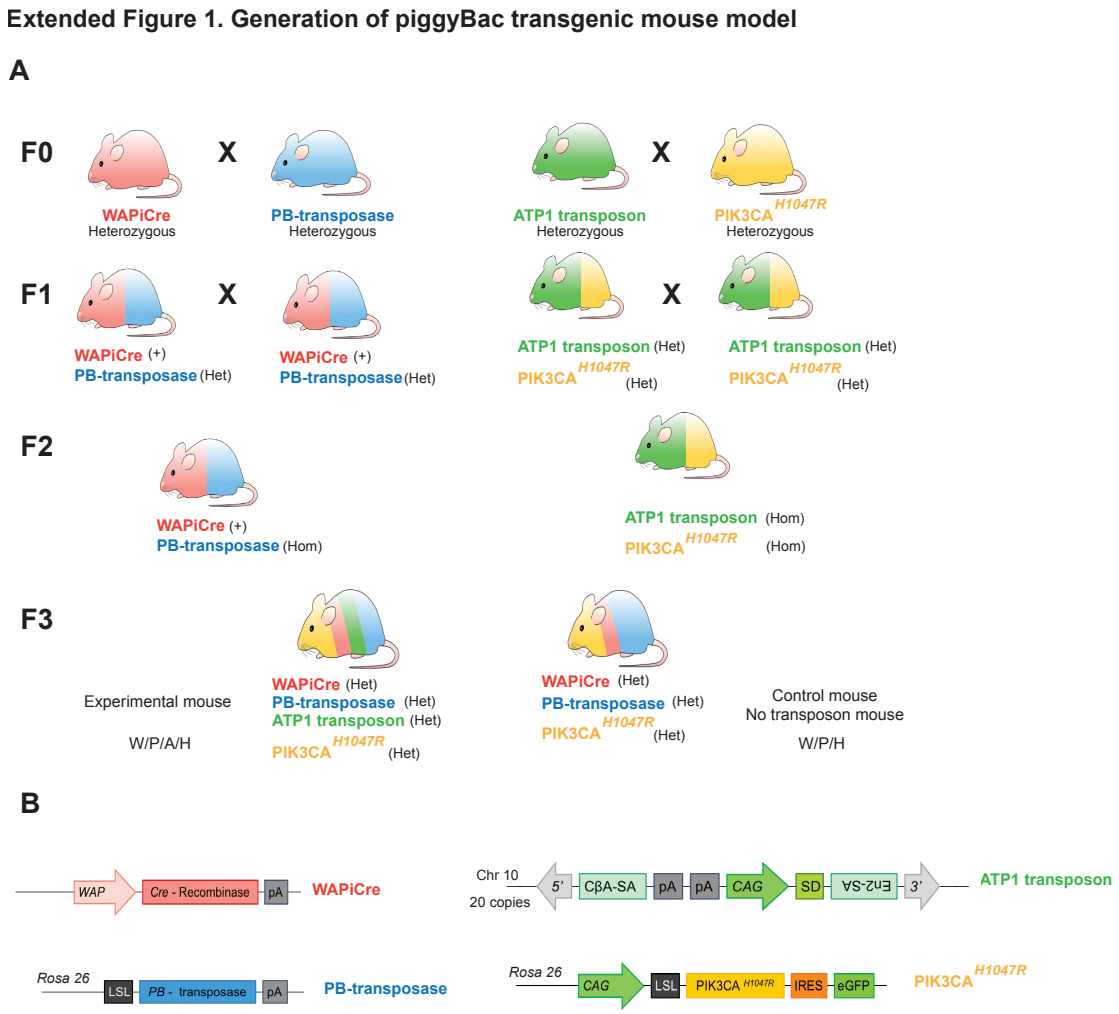


Figure 6-1 | Generation of a *piggyBac/PIK3CA*^{H1047R} mammary cancer mouse model. A, Schematic overview of the breeding strategy and mouse lines: PB-transposase (Rad et al., 2010) transgenic mice were crossed to WapiCre mice and ATP1(Rad et al., 2010) transposon transgenic mice to *PIK3CA*^{H1047R} transgenic mice (Meyer et al., 2011, 2013) generating PB-transposase/WapiCre and ATP1 transposon/ *PIK3CA*^{H1047R} mice. Experimental mice were obtained by crossing PB-transposase/WapiCre to ATP1 transposon/*PIK3CA*^{H1047R} mice (W/P/A/H). Mice without ATP1 transposon were used as control (W/P/H). F0, F1, F2 and F3 are the filial generations. **B,** Graphical representation of the constructs used in this study: WAPi-CreERT2 (Koren and Bentires-Alj, 2013); *RosaPB* lox-STOP-lox (LSL) mice expressing *piggyBac* transposase (Rad et al., 2010); ATP1 transposon (Rad et al., 2010) and *RosaPB* LSL *PIK3CA*^{H1047R} (Koren and Bentires-Alj, 2013).

6.1.2 Transposon mutagenesis enhanced tumourigenic ability in mice.

Mice with an active PB system developed more mammary tumours (**Fig 6-2A and B**) and with a shorter latency (**Fig 6-2C**) compared with control animals (W/H/P). This suggests that the PB system generated advantageous mutations that enhance tumour progression.

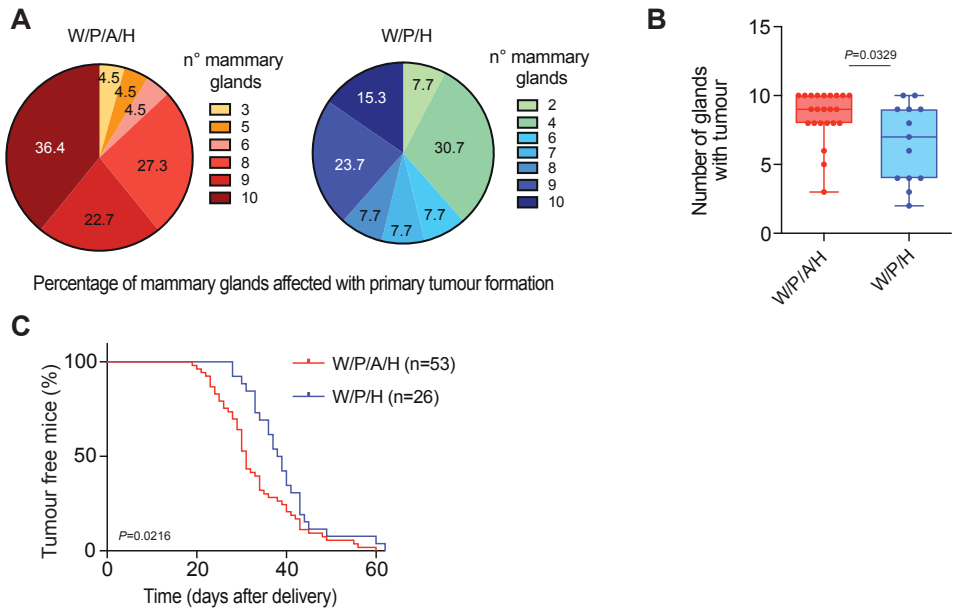


Figure 6-2 | Decreased tumour latency in PB mutagenized mice. A, Pie chart showing the percentage of mammary glands per mouse affected with primary tumour in W/P/A/H and W/P/H mice. **B,** Box plot showing the number of mammary glands per mouse affected with primary tumour. Means ± s.d., two tailed Mann-Whitney U-test. **C,** Kaplan-Meier plot depicting tumour onset in W/P/A/H (n = 53) and W/P/H (n = 19) transgenic mice, two-tailed log-rank test.

6.1.3 Increased frequency and accelerated tumour formation after secondary transplantation.

To evaluate the tumourigenic potential and select for cancer cells with a tumour initiation ability, we performed serial transplantation of tumour pieces into the cleared fat pad of recipient NOD/SCID mice (**Fig 6-3A**). Six control and eleven W/P/A/H donor tumours were used for primary transplantation (**Fig 6-3B**) and four control and nine W/P/A/H mutagenized donor tumours for secondary transplantation respectively (**Fig 6-3C**). Tumourigenic ability and growth on primary transplantations were not affected (**Fig 6-3D, Extended Fig 6-1A and B**). In contrast, W/P/A/H tumours had increased frequency in tumour formation and faster growth compared with control mice, after the secondary transplantation (**Fig 6-3E, Extended Fig 6-1C and D**). Notably, almost the totality of the W/P/A/H donor pieces engrafted after the

secondary transplantation, showing positive selection of cancer cells with tumourigenic ability (**Extended Fig 6-1E and F**). Conversely, almost half of the control donor tumour pieces did not grow a tumour after the primary transplantation (**Extended Fig 6-1E)** and only 15% grew tumours after the secondary transplantation (**Extended Fig 6-1F).**

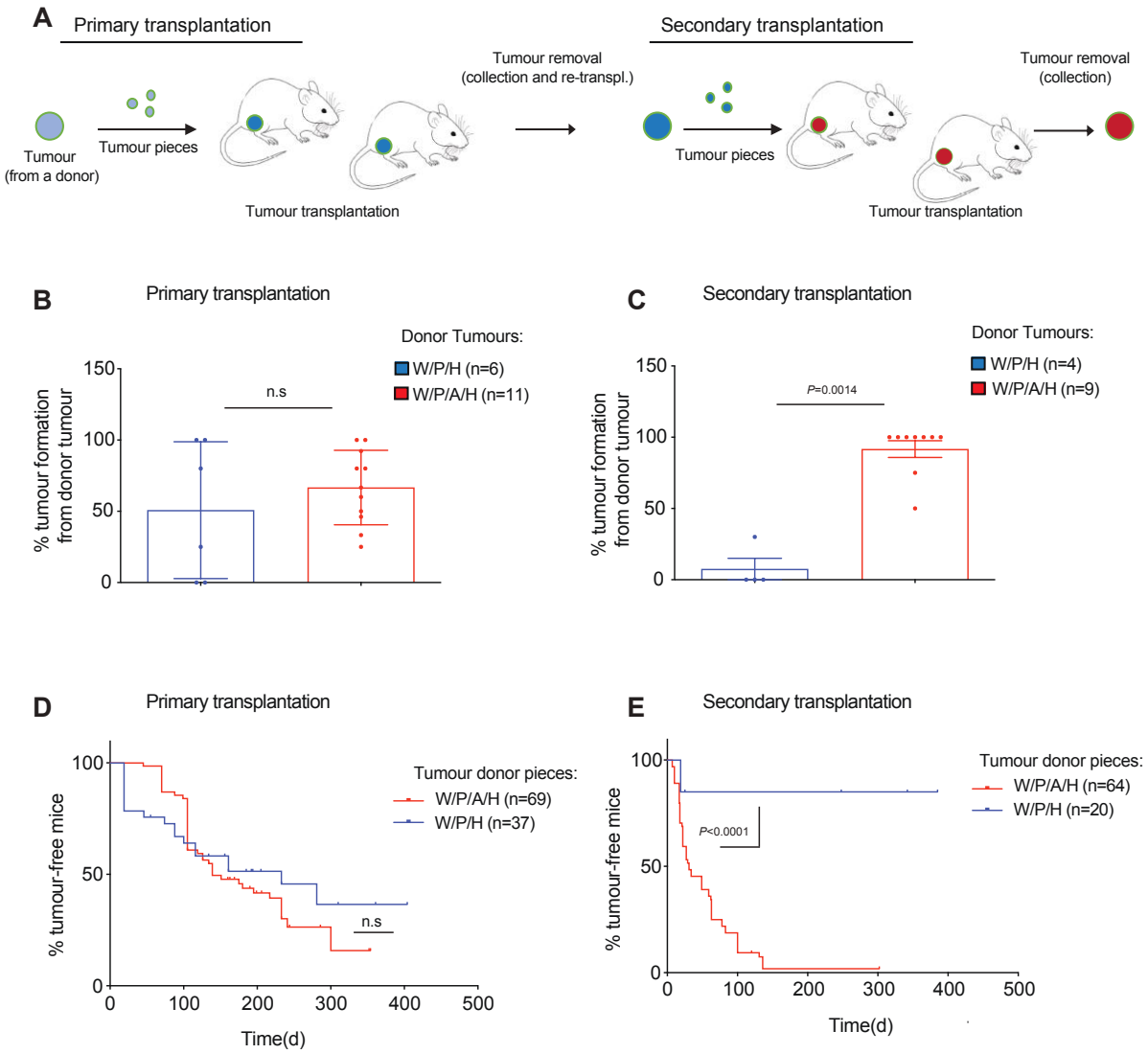


Figure 6-3 | Increased tumour frequency and accelerated tumour formation after W/P/A/H tumours secondary transplantation. A, *In vivo* serial transplantation experimental design: donor tumours are cut in pieces of similar size and transplanted into the cleared fat pad of recipient NOD/SCID mice. **B,** Box plot showing the percentage of tumour formation from each donor tumour W/P/H (n = 6) or W/P/A/H (n = 11) in primary transplantation. Means ± s.d., two tailed Mann-Whitney U-test, *n.s.* = not significant. **C,** Box plot showing the percentage of tumour formation from each donor tumour W/P/H (n = 4) or W/P/A/H (n = 9) in secondary transplantation. Means ± s.d., two tailed Mann-Whitney U-test. **D,** Kaplan-Meier plot depicting tumour incidence of W/P/A/H (n = 69) and W/P/H (n = 37) tumour donor pieces (n = 37) after primary transplantation in NOD/SCID mice; two tailed log-rank test, *n.s.* = not

significant. **E**, Kaplan-Meier plot depicting tumour incidence of W/P/A/H (n = 64) and W/P/H (n = 20) tumour donor pieces (n = 20) after secondary transplantation in NOD/SCID mice; two tailed log-rank test.

6.1.4 Identification of candidate tumour driver genes.

Next, we performed targeted sequencing of the transposon integration sites in the 22 primary and 31 secondary tumours, searching for insertions enriched in secondary compared with primary transplanted tumours (**Fig 6-4**). The sequencing data show that the transposon integration was efficient and unbiased in the genome (**Extended Fig 6-2A and B**). Similar to previous report (Liang et al., 2009), our data from the PB transposon showed slight local hopping (**Extended Fig 6-2c**). We observed a mild peak of integration at the donor chromosome 10, thus we excluded putative hits from this chromosome.

Splinkerette PCR and sequencing from both transposon arms (Friedrich et al. 2017) allowed us to consider the diversity counts (div counts) within each sample; each uniquely sheared end was counted rather than all sequencing counts, thereby minimizing PCR-induced amplification effects. Targeted sequencing of the transposon integration sites in primary and secondary transplanted tumours revealed almost 12000 genes (**Supplementary Table 1**). To assess the most significant common insertion sites (CIS) we used a gene-centric common insertion site (gCIS) calling method (Brett et al., 2011), which identifies 7,764 CISs found in at least two tumours (**Fig 6-4A and Supplementary Table 2**). Two parameters were considered for each CIS to determine the confidence for selection: the number of tumour samples in which the CIS gene was targeted and the average of normalized div counts that estimate the frequency of insertion of the CIS within samples. CIS with high confidence were chose as potential driver cancer genes.

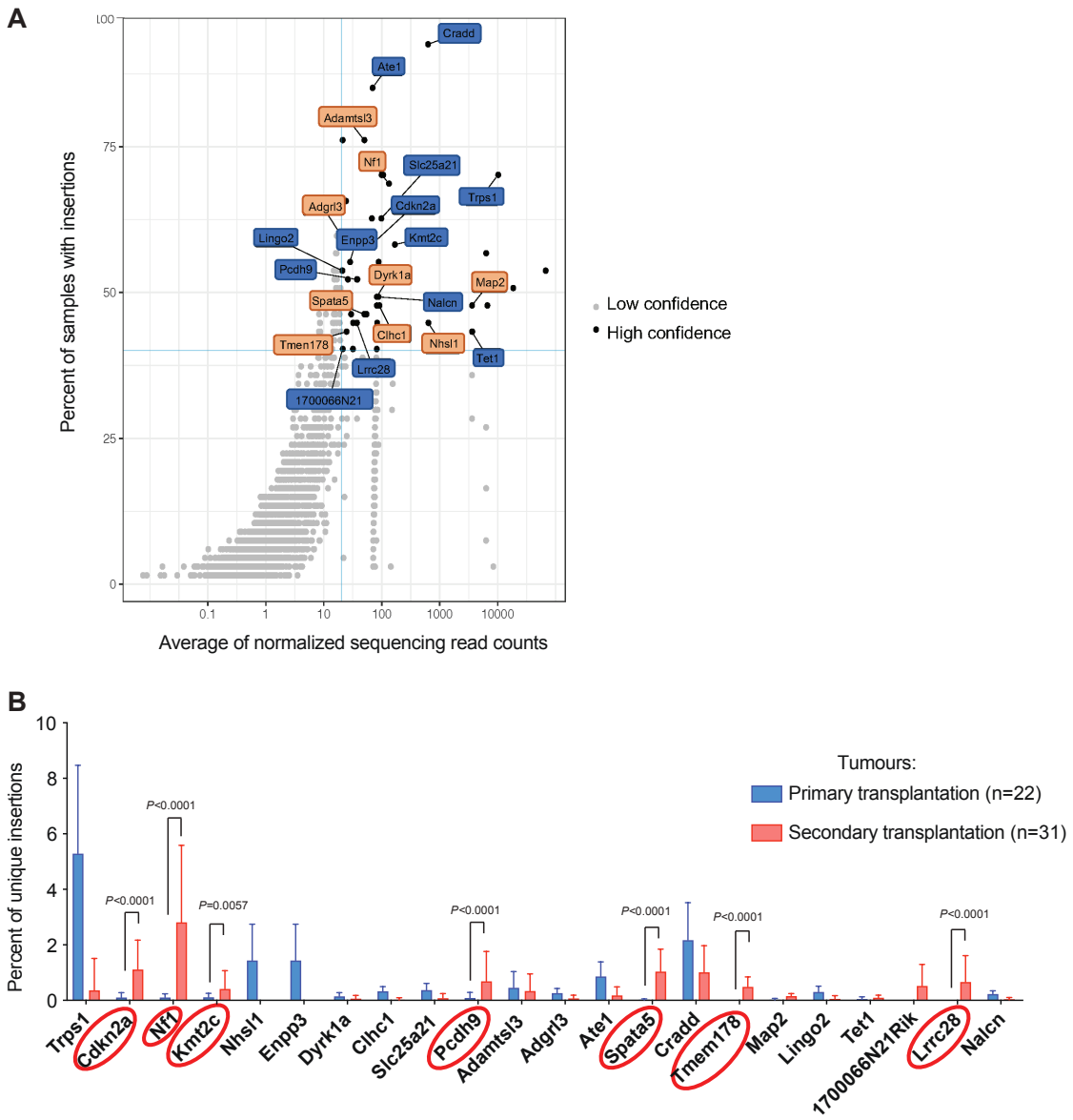
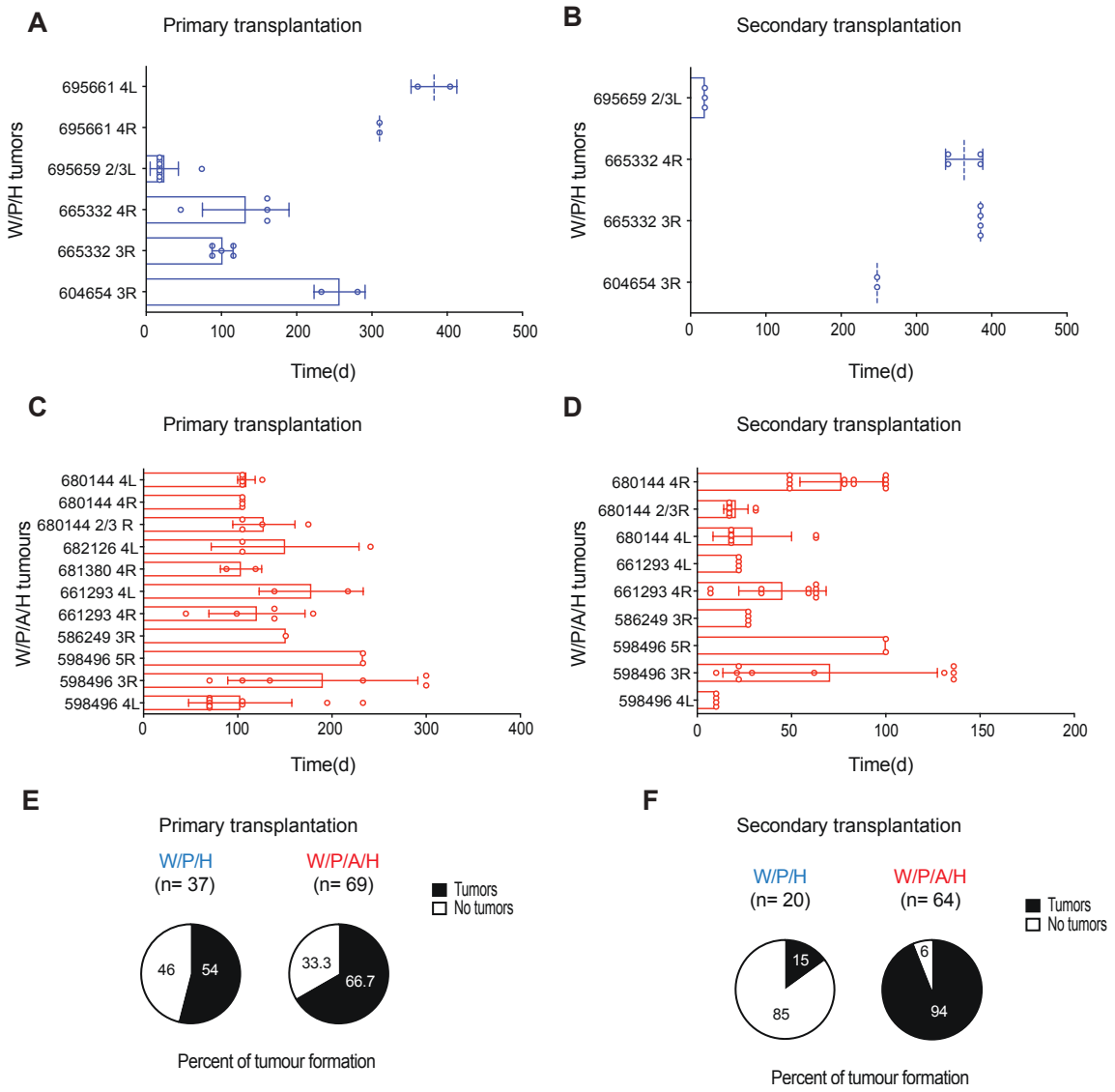


Figure 6-4 | Transposon mutagenesis screen identifies candidate mammary tumour driver genes. A, Genetic landscape of PB insertions in the total 53 tumours sequenced. **B,** Bar graph showing the percentage of unique insertions in the top 25 genes normalized to the total number of insertions in a given sample. Gene are annotated with gene symbol when available otherwise with UCSC ID transcript names. Data are from all the sequenced samples (22 primary tumours and 31 secondary tumours).

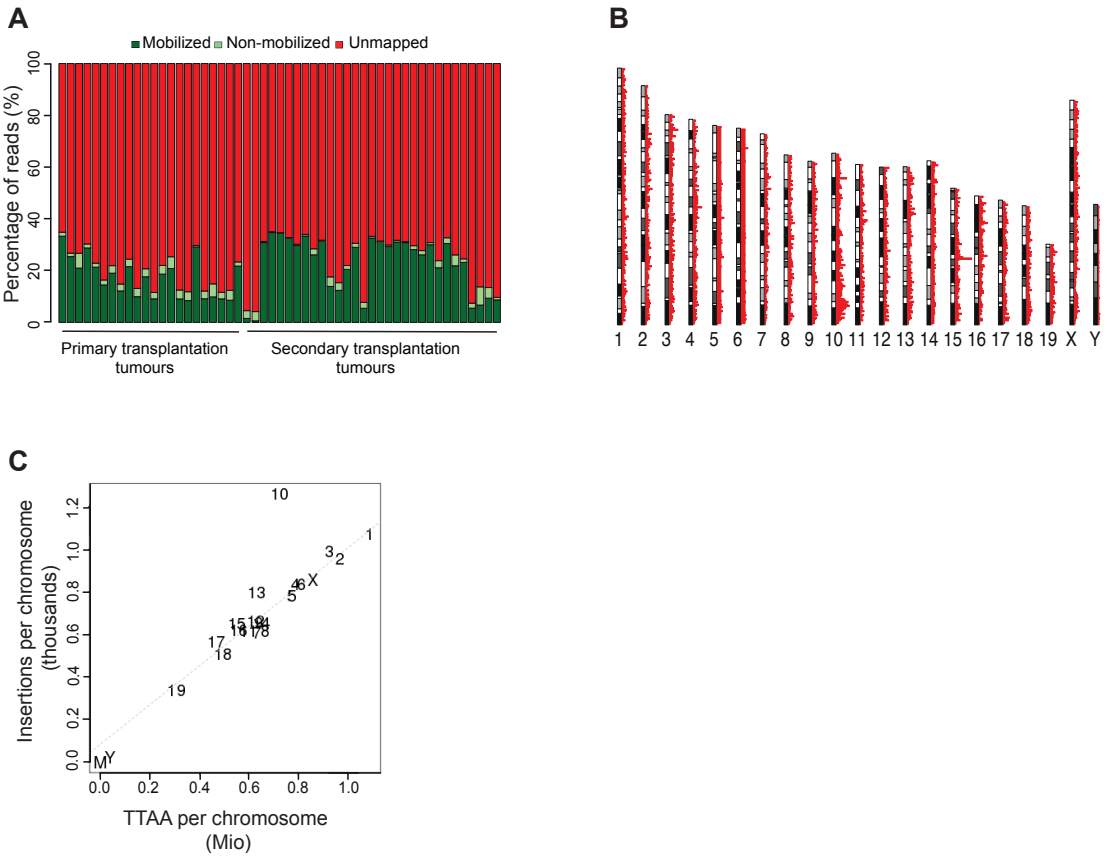
Finally, we selected for unique insertions enriched in the secondary transplanted tumour compared with the primary (see Methods). Among the most frequently altered genes in the secondary transplanted tumours compared with the primary transplanted tumours were *Nf1, Cdkn2a, Kmt2c, Pcdh9, Lrrc28, Tmen178,* suggesting an effect of these genes on tumour formation ability. The same-strand insertions upstream of *Nf1* and *Tmen178* transcription start

sites also suggest they are candidate oncogenes, on the contrary the opposite-strand insertions in *Cdkn2a, Kmt2c, Pcdh9* and *Lrrc28* genes suggest they are candidate tumour suppressor genes.

6.2 Extended Figures



Extended Figure 6-1 | W/PA/H tumours show enhanced tumourigenic ability after secondary tumour transplantation. W/P/A/H tumours had increased frequency and acceleration in tumour formation after secondary transplantation. **A, B** Box plot showing the time (days) of tumour growth of W/P/H, control, from different tumours donors during primary (**A**) and secondary transplantation (**b**). **C, D** Box plot showing the time (days) of tumour growth of W/P/A/H, control, from different tumours donors during primary (**c**) and secondary transplantation (**D**). Each dot corresponds to a single transplanted tumour piece. **E, F** Pie chart showing the percent of tumour formation after primary (**E**) and secondary transplantation (**F**) in W/P/A/H and W/P/H pieces donor tumour.



Extended Figure 6-2 | Transposon mobilization in transplanted tumours. A Mapping statistic for individual samples, with fractions of read pairs that mapped uniquely to the mouse genome ("Mobilized"), that mapped to the transposon ("Non-mobilized"), or that could not be uniquely mapped ("Unmapped). Insertion per chromosome per TTAA present in the chromosome. **B,** Graphic representation of the genomic insertion distribution in each chromosome. **C,** Graph representing the insertion distribution per chromosome per TTAA per chromosome.

6.3 Discussion

Here we used PB transposon system in an *PIK3CA* mutant background mouse model to identify genes that propagate and accelerate tumorigenesis. To enrich and select for tumour initiating cells, we serially transplanted PB/*PIK3CA*^{H1047R} mutagenized and control tumours. We used DNA shearing followed by splinkerette PCR and deep sequencing to allow quantification of the amplicon diversity based on the sheared end and estimation of the clonality of each tumour sample. 7,764 CISs have been identified from the sequencing of the W/H/A/P transplanted tumours. Slightly local transposon hopping effect was observed and deep sequencing allowed us to define a method to predict GOF or LOF transposon insertions. The most frequently targeted tumour suppressor was the gene *Trps1*, which is already known to be involved in mammary tumourigenesis (Rangel et al., 2016). Other known tumour suppressor genes identified included *Kmt2c Cdkn2a* and *Rasa1* (Suárez-Cabrera et al., 2017).

The nature of CIS genes targeted by PB suggests that diverse genetic and epigenetic posttranslational alterations occur in these tumours. For instance, a number of alterations of known chromatin remodelling factors, including *Arid1b*, four different *Hdac* genes, *Brd4*, and *Kmt2c*. Genes enriched in secondary transplanted tumour, suggesting their involvement in tumor initiation. We used two cohort datasets, METABRIC (Curtis et al., 2012) and TCGA (Koboldt et al., 2012), to examine whether the expression of the most significant candidate genes could be prognostic for human breast cancer and if they co-occurred with *PIK3CA* mutation. *KMT2C* and *PIK3CA* co-occurred significantly ($p < 0{,}001$) and *PCDH9* did not co-occurred significantly with *PIK3CA. KMT2C* (also referred to as *MLL3*) encodes an histone 3 Lysine 4 (H3K4) methyltransferase and is one of the most commonly mutated genes in breast cancer (Liu et al., 2015; Gala et al., 2018). Notably *KMT2C* is a key regulator of ERα activity (Sato and Akimoto, 2017), that sustain the co-occurrence with *PIK3CA* mutation, which mainly occurs in luminal (ER positive) breast cancers. *KMT2C* is a predicted tumour suppressor gene,

indeed higher expression levels of *KMT2C* in luminal breast cancers were predictive of increased relapse-free survival (Györffy et al. 2010).

PCDH9 encodes a member of the protocadherin family, and cadherin superfamily, of transmembrane proteins containing cadherin domains and has been correlated with ovarian and hepatocellular carcinoma (Chen et al., 2015; Lv et al., 2017; Shi et al., 2019).

Overall, our work highlights how *in vivo* transposon-mediated mutagenesis can be used to elucidate driver genetic mechanisms of cancer propagation.

6.4 Materials and methods

Mice

WAPiCre mouse line, the generation of *PIK3CA*^{H1047R} mice (pure FVB) (Meyer et al., 2011, 2013), PB-transposase and ATP1 transposon mice (Rad et al., 2010, 2015) were previously described. Mouse colonies and female NOD/SCID were maintained in the Friedrich Miescher Institute for Biomedical Research and the Department of Biomedicine animal facilities in accordance with Swiss guidelines on animal experimentation. Mice were maintained in a sterile controlled environment (a gradual light–dark cycle with light from 7:00 to 17:00, 21-25°C, 45-65 % humidity).

Genotyping: WAPiCre allele was carried out using WAP11 5'-GAAAAGCACCAGGAGAAGTCAC-3' and WAP12 5'-GACACAGCATTGGAGTCAGAAG -3' primers, which amplified a 900-bp product; *PIK3CA*^{H1047R} allele was carried out using H1047FW 5'- TGGCCAGTACCTCATGGATT -3' and H1047RV 5'- GCAATACATCTGGGCTACTTCAT-3', which amplified a 600-bp product; PB-transposase knock-in allele was carried out using using BpA5F 5'-GCTGGGGATGCGGTGGGCTC-3' and Rosa3R 5'-GGCGGATCACAAGCAATAATAACCTGTAGTTT-5' primers, which amplified a 250-bp product and PB-transposase wild-type *Rosa26* allele was carried out using Rosa5F 5'-CCAAAGTCGCTCTGAGTTGTTATCAG-3' and Rosa3R 5'-GGCGGATCACAAGCAATAATAACCTGTAGTTT-3', which amplify a 450-bp product; ATP1- transposon allele was carried out using ATP FW 5'-CTCGTTAATCGCCGAGCTAC3' and ATP RV 5'-GGCGGATCACAAGCAATAATAACCTGTAGTTT-3', which amplified a 808-bp product.

Animal experiments

All *in vivo* experiments were performed in accordance with the Swiss animal welfare ordinance and approved by the cantonal veterinary office Basel-Stadt. Mice were maintained in a sterile controlled environment (a gradual light–dark cycle with light from 7:00 to 17:00, 21-25°C, 45-65 % humidity). For serial transplantation assay, donor tumours were orthotopically transplanted (primary and secondary transplantations) in the cleared fat pad of recipient NOD/SCID mice. More specifically donor tumour pieces were cut with similar size and engrafted in the cleared fat pad of the fourth mammary gland of NOD/SCID mice. One mammary gland per mouse was used for primary transplantation and two mammary glands were used for secondary transplantation. Tumours were measured with Vernier calipers and volume calculated as 0.5 x [(large diameter) x (smallest diameter)^2]. Tumours were measured with Vernier calipers and volume calculated as 0.5 x [(large diameter) x (smallest diameter)^2]. Tumours were resected before they reached 1500 mm^3 and mice were monitored regularly for signs of metastatic outgrowth and distress. For survival studies, animals were sacrificed when tumours reached 1500 mm^3 or when they showed any signs of distress (e.g., breathing disorders, weight loss, or immobility). All orthotropic experimental procedures (tumour resection and tumour cell implantation) were undertaken on anaesthetized mice by a single investigator according to protocols approved by the cantonal veterinary office Basel-Stadt.

Splinkerette PCR for the amplification of transposon integration sites

For PB sequencing, we adapted the splinkerette PCR protocol described previously (Friedrich et al., 2017). Sequenced samples: 22 primary transplanted tumours and 31 secondary transplanted tumours. Genomic DNA was isolated (using DNeasy Blood and Tissue Kit; Qiagen), sheared to a fragment length of 250 bp with a Covaris sonicator. After end repair and A-tailing, purified DNA fragments were ligated to a splinkerette adaptor (obtained after annealing of top 5′- gttcccatggtactactcatataatacgactcactataggtgacagcgagcgct-3′ and bottom 5′-

gcgctcgctgtcacctatagtgagtcgtattataattttttttcaaaaaaa-3′). Transposon-containing fragments were enriched by 18 cycles of transposon-specific PCR1 for the 5′ transposon ends in a unique library (5′-gacggattcgcgctatttagaaagagag-3′ for the 5′ arm of PB, and common splinkerette primer 5′-gttcccatggtactactcata-3′). Bar coding of individual samples and completion of Illumina adaptor sequences were achieved by an additional 29 cycles of transposon-specific PCR and a custom array of 37 unique bar-coding primers. For the 5′ arm, we used 5′-aatgatacggcgaccaccgagatctacacatgcgtcaattttacgcagactatc- 3′ and for the splinkerette side, we used 5′-caagcagaagacggcatacgagatcggtXXXXXXXXtaatacgactcactatagg-3′ primers. The Xs represent the bar code of 8 nucleotides. After magnetic bead purification (Beckman Ampure XP), libraries were assembled in two pools and sequenced on an HiSeq Desktop Sequencer (Illumina): HiSeq 2500 and MiSeq, Rapid Run, Paired-End, 2 x 100 bp, with 15% PhiX; 7 pM was loaded on to the instrument.

Mapping of insertion sequences to the mouse genome and identification of integration sites

Pre-processing and alignment: Paired-End reads were first pre-processed by removing the expected transposon-derived sequence (5'-TAGGGTTAA-3') from the beginning of the first read (read-pairs with non-matching first reads were discarded; typically around 1%) using the preprocessReads function from QuasR (version 1.12.0,[33]). Read pairs were then aligned to the mouse genome (BSgenome.Mmusculus.UCSC.mm10) using the QuasR qAlign function and parameters "-m 1 --best --strata --maxins 1000"; only uniquely mapping pairs with up to 1000 bp between-pair distance were reported. Mapping rates were recorded and non-mapped read pairs were further aligned against the non-mobilized transposon sequence to estimate the probable fraction of read pairs that originate from non-mobilized copies of the transposon. For each aligned read pair, the *piggyBac* insertion coordinate was identified as the coordinate of the first (most 5') base of the first read.

Integration site identification and quantification: For each unique integration site, the total number of supporting alignments (any read pair), the number of supporting alignments from the most frequent fragment (most frequent read pair), the number of distinct supporting alignments (distinct read pairs) and the genomic sequence from the four base pairs on the same strand as the first read directly upstream of the insertion site were recorded. Normalized counts were calculated by dividing raw counts through the total number of alignments in a given sample and multiplied with 1e5 (counts per 100,000 insertions). For the downstream analysis, only integration sites with the expected upstream TTAA sequence were used.

Association of integration sites with genes: Coordinates of known genes (exons/introns, 5'-untranslated region [UTR], coding sequence [CDS], 3'UTR) were obtained from the TxDb.Mmusculus.UCSC.mm10.knownGene Bioconductor package (version 3.2.2). Promoter regions were defined as regions 2,000 bp upstream of known transcript start sites. Integration sites were matched against these genomic regions to identify overlaps on any strand and orientation, selecting the first overlap in the case of multiple overlaps. Integration sites were classified hierarchically as follows: sites without overlaps to any transcript were labeled as promoter sites (in the case of an overlap with a promoter region) or intergenic sites. All other sites were labeled with the first region type that they overlapped, in the following order: 5'UTR, CDS, 3'UTR, intron, or ncRNA (defined as an overlap with a transcript without annotated CDS). Sites were further labeled according to their orientation with respect to the associated gene (same or opposite). Finally, sites were grouped according to the gene they overlapped (including promoter sites) or, for intergenic sites, according to the pair of flanking genes.

Gene-centric common integration site calling: In order to identify genes with over-represented integration sites, we adapted the gCIS method (Brett et al., 2011). Briefly, for each region of interest (e.g. gene or intergenic region), the observed number of integration sites is compared with the expected number. For this, the numbers of integration sites within the region

(sites.inside) and on the same chromosome outside of the region (sites.outside) are compared with the numbers of TTAA sites in the sequence of the region (ttaa.inside) and on the same chromosome outside of the region (ttaa.outside). A region with enriched integration sites has a ratio sites.inside/sites.outside greater than the expected ratio for that size of region and chromosome (ttaa.inside/ttaa.outside), and p-value for the region was calculated using Fisher's exact test.

Selection of the enriched unique integration sites: For each insertion the diversity has been calculated (the number of distinct alignment-pairs starting at the insertion coordinate) (Chapeau et al., 2017). For each gene the diversity has been summed up resulting in the number of unique insertions in a gene. Finally, this value was divided by the total number of unique insertions for a given sample. In order to find putative oncogenic genes, we looked at the genes with the higher number of unique insertions in the secondary transplanted tumours compared with the primary transplanted tumour samples.

Data analysis from publicly available datasets. We used cBioPortal (Cerami et al. 2012, Gao et al. 2013) for the *PIK3CA*, *NF1, CDKN2A, KMT2C, PCDH9, LRRC28* and *TMEN178* expression correlation study with publicly available data (Curtis, et al. 2012, Pereira, et al. 2016). The results here are based in part on data generated by the TCGA Research Network (https://www.cancer.gov/tcga). Relapse-free survival and distant-metastasis-free survival of *PIK3CA*, *NF1, CDKN2A, KMT2C, PCDH9, LRRC28* and *TMEN178* were generated using the 2017 version of KMplotter (Györffy et al. 2010) (http://kmplot.com/analysis/index.php?p=service&cancer=breast).

Statistical data analysis

Values represent the means ±s.d., unless stated otherwise. *P* values were determined using unpaired two-tailed *t*- tests and statistical significance set at *P*=0.05. The variance was similar between the groups compared. Biological replicates correspond to different tumour

material. Technical replicates are tests or assays run on the same sample multiple times. For statistical analysis and for reporting of the number of experimental entities, biological replicates were used. The means of technical replicates, if available, were used for analysis and visualization. Data were tested for normal distribution, Student's t-tests and two-way ANOVA (if normally distributed) or nonparametric Mann-Whitney U-test or Wilcoxon tests were applied unless stated otherwise. Kaplan–Meier plots were generated using the survival calculation tool from Graphpad Prism and significance was calculated using the two-tailed log-rank test at $P < 0.05$.

7 | DISCUSSION AND OUTLOOK

Current models of transposon-mediated forward genetic screen are valuable and functional tools to discover cancer genes and to better understand cancer pathogenesis (Rad et al., 2014; Chapeau et al., 2017; Friedrich et al., 2017; De La Rosa et al., 2017; Weber et al., 2019). These models are based on random mutagenesis and subsequent clonal expansion and neoplastic transformation of a cell carrying cancer-causing mutations. Because of its unbiased nature it is possible to discover cancer drivers that are difficult to find with other approaches. By using transposon unbiased genetic *ex vivo* and *in vivo* models we revealed new mechanism of breast tumorigenesis and metastasis, which could provide valuable information for developing targeted therapeutics for breast cancer.

7.1 *NFIB* induces metastasis

Metastasis is the primary cause of solid cancer-related mortality and delineation of its molecular mechanisms is a high priority for development of successful treatments. Although, several studies showed the causative effect for *PIK3CA* mutations in breast tumorigenesis (Meyer et al., 2013; Koren et al., 2015), metastases were not found in the *PIK3CA*^{H1047R} expressing mice (Koren and Bentires-Alj, 2013; Koren et al., 2015) suggesting that additional gain- or loss-of-function genomic alterations may be required for metastatic progression.

Using a transposon insertional mutagenesis screen in a *PIK3CA* mutant background, we found that NFIB evokes breast cancer metastatic progression and colonization. NFIB has been reported to drive tumour initiation and progression in different cancer types (Denny et al., 2016; Semenova et al., 2016; Fane et al., 2017; Wu et al., 2018) and to be implicated in the regulation of breast cancer growth and progression (Campbell et al., 2018; Liu et al., 2019a). However, our functional validation *in vivo* shows its implication in breast cancer formation and progression. Furthermore, we show that *in vivo NFIB* overexpression is sufficient to escalate

the metastatic colonization step and ultimately reduce animal survival. I conclude that NFIB has two effects, providing growth advantages both in the primary tumour and in the lung microenvironment. My studies provide new insights into the effects of the NFIB-ERO1A axis in breast cancer progression to metastasis. Mechanistically, I discovered that the NFIB-ERO1A axis promotes angiogenesis at the secondary site, resulting in a permissive microenvironment for metastatic colonization.

7.2 The NFIB-ERO1A axis promotes angiogenesis at the secondary site

These findings lead to further questions. How is the NFIB-ERO1A axis inducing angiogenesis at the metastatic site? Other studies showed that ERO1A increased angiogenesis via oxidative protein folding of VEGF and enhancement of *VEGF* mRNA expression (Tanaka et al., 2016). It is known that the formation of a disulphide bond by ERO1A is accompanied by the production of ROS such as hydrogen peroxide (Zito, 2015). Moreover, it has been shown that ROS induce stabilisation of HIF-1A protein through inhibition of the prolyl hydroxylase enzyme (Bell et al., 2007; Yan et al., 2010). Further investigation is needed to explore if ROS, which are products of ERO1A can upregulate the mRNA expression of *VEGF* eventually via stabilisation of HIF-1A protein.

7.3 *NFIB* has a pleiotropic effect during development and cancer progression

During oncogenesis, microenvironmental stress such as chronic inflammation, accumulation of ROS, or hypoxia may promote clonal expansion of cells with genetic or epigenetic abnormalities. These cells then acquire further mutations or epigenetic alterations and become founder tumour cells that initiate cancer. Thus epigenetic mechanisms contribute to the intratumoural heterogeneity (Easwaran et al., 2014). A recent study examining *NFIB* in SCLC suggest that in this context, *NFIB* modifies chromatin architecture in a manner that results in

increased tumour heterogeneity (Denny et al., 2016). Another report shows that *NFIB* promotes dynamic changes in the chromatin state of melanoma cells to facilitate migration, invasion and metastasis (Fane et al., 2017). Whether NFIB-induced breast cancer progression and metastasis is triggered by epigenetic events remains unknown, investigating this possibility is warranted.

Our transcriptomic analysis show increased expression of stem-cell like and mesenchymal/epithelial related genes in tumourspheres and primary tumours derived from *Nfib* high-expression models. *NFIB* has been shown to induce EMT in colorectal and gastric cancer by activation of AKT and STAT3 (Wu et al. 2018; Z. Y. Liu et al. 2019). Additionally, a recent study demonstrated that cancer cells display multiple tumour subpopulations associated with different EMT stages: from epithelial to completely mesenchymal states. The intermediate hybrid states presented increased cellular plasticity, invasiveness and metastatic potential (Pastushenko et al., 2018). Interestingly, they found that one of the core family TFs to induce these transient states are the NFIs. Functional validation is needed to investigate whether *NFIB* is a core TFs in promoting intermediate states that allow colonization at the secondary site.

NFIB is a pleiotropic TF that creates a permissive tumour microenvironment and alters the global chromatin architecture within a subset of cancers to drive metastasis. The correlation between NFIB effects in development and tumorigenesis highlight its importance as prognostic marker for tumour progression. Additional studies, similar to mine, that investigate a tumorigenic effect of NFIB within different types of cancer may highlight potential downstream therapeutic candidates that could be targeted more effectively.

7.4 *PIK3CA* synergistic pathways in breast cancer progression

Each breast cancer can carry several mutations (Alexandrov et al., 2013), and a normal human cell can acquire 7–15 somatic mutations before malignant transformation (Beerenwinkel et al., 2007). Thus, most mutations in breast cancer are likely to be passenger mutations that do not

contribute to tumorigenesis or tumour progression. Functional screens that can identify the driver mutations in breast cancer are very useful in this regard.

Serial transplantation into the mammary fat pads of recipient mice allowed us to select for mutations responsible for enhanced tumour initiation ability. Synergistic mutational events not only induced tumour formation but also allowed their maintenance through serial transplantation rounds. I identified several candidate genes that could contribute to tumour progression with *PIK3CA* and one of the most promising is *KMT2C*, which is an epigenetic modulator that encodes an H3K4 histone methyltransferase, and is one of the most frequently mutated genes in ERα-positive breast cancer (Sato and Akimoto, 2017). Mechanistically, it has been demonstrated that *FOXA1* interacts with the chromatin *KMT2C* to facilitate monomethylation of H3K4 at enhancer elements, resulting in the potential for transcription from these enhancer regions (Jozwik et al., 2016). Moreover it is a tumour suppressor gene and is an essential ERα coactivator, and its depletion correlates with significantly shorter progression-free survival on anti-estrogen therapy (Liu et al., 2015; Gala et al., 2018). Crosstalk between the PI3K and ERα contributes to resistance, and activation of the PI3K pathway has been associated with the failure of endocrine therapy (Mayer et al., 2017). PI3K pathway inhibitors have emerged as promising therapeutic agents for estrogen receptor ERα-positive breast cancers. However they display limited efficacy in clinical settings often due to resistance mechanisms and toxicity (Vasan et al., 2019). Hence, combinatorial strategies to circumvent resistance should be considered. Recent studies identified that knockdown of the *KMT2C* gene leads to the suppression of estrogen-dependent genes (Gala et al., 2018). Of note, *KMT2D*, another epigenetic member of the KMT2 family, was reported to enhance ERα activity in *PIK3CA* mutant breast cancer treated with a PI3K-selective inhibitor (Toska et al., 2017). Conclusively, from a therapeutic point of view, these results indicate that targeting of

epigenetic modulators such as KMT2C in combination with PI3K pathway inhibition should be investigated.

8 | REFERENCES

Aceto, N., Bardia, A., Miyamoto, D.T., Donaldson, M.C., Wittner, B.S., Spencer, J.A., Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell *158*, 1110–1122.

Aceto, N., Toner, M., Maheswaran, S., and Haber, D.A. (2015). En Route to Metastasis: Circulating Tumor Cell Clusters and Epithelial-to-Mesenchymal Transition. Trends in Cancer *1*, 44–52.

Adam, R.C., Yang, H., Ge, Y., Infarinato, N.R., Gur-Cohen, S., Miao, Y., Wang, P., Zhao, Y., Lu, C.P., Kim, J.E., et al. (2020). NFI transcription factors provide chromatin access to maintain stem cell identity while preventing unintended lineage fate choices (Springer US).

Adams, J.R., Xu, K., Liu, J.C., Ruiz Agamez, N.M., Loch, A.J., Wong, R.G., Wang, W., Wright, K.L., Lane, T.F., Zacksenhaus, E., et al. (2011). Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation. Cancer Res. *71*, 2706–2717.

Albieri, I., Onorati, M., Calabrese, G., Moiana, A., Biasci, D., Badaloni, A., Camnasio, S., Spiliotopoulos, D., Ivics, Z., Cattaneo, E., et al. (2010). A DNA transposon-based approach to functional screening in neural stem cells. J. Biotechnol. *150*, 11–21.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Anderson, A.R.A., Weaver, A.M., Cummings, P.T., and Quaranta, V. (2006). Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment. Cell *127*, 905–915.

Andreasen, S., Persson, M., Kiss, K., Homøe, P., Heegaard, S., and Stenman, G. (2016). Genomic profiling of a combined large cell neuroendocrine carcinoma of the submandibular gland. Oncol. Rep. *35*, 2177–2182.

Angus, L., Smid, M., Wilting, S.M., van Riet, J., Van Hoeck, A., Nguyen, L., Nik-Zainal, S., Steenbruggen, T.G., Tjan-Heijnen, V.C.G., Labots, M., et al. (2019). The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nat. Genet. *51*, 1450–1458.

Babakhani, S., and Oloomi, M. (2018). Transposons: the agents of antibiotic resistance in bacteria. J. Basic Microbiol. *58*, 905–917.

Bachman, K.E., Argani, P., Samuels, Y., Silliman, N., Ptak, J., Szabo, S., Konishi, H., Karakas, B., Blair, B.G., Lin, C., et al. (2004). The PIK3CA Gene is Mutated with High Frequency Bio sci MATERIALS AND METHODS No t D ist r ibu te. Therapy *3*, 772–775.

Bader, A.G., Kang, S., and Vogt, P.K. (2006). Cancer-specific mutations in PIK3CA are oncogenic in vivo. Proc. Natl. Acad. Sci. U. S. A. *103*, 1475–1479.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385.e18.

Balu, B., Shoue, D.A., Fraser, M.J., and Adams, J.H. (2005). High-efficiency transformation of Plasmodium falciparum by the lepidopteran transposable element piggyBac. *102*.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. Nature *486*, 405–409.

Bapat, S.A. (2007). Evolution of cancer stem cells. Semin. Cancer Biol. *17*, 204–213.

Barbareschi, M., Buttitta, F., Felicioni, L., Cotrupi, S., Barassi, F., Del Grammastro, M., Ferro, A., Palma, P.D., Galligioni, E., and Marchetti, A. (2007). Different prognostic roles of mutations in the helical and kinase domains of the PIK3CA gene in breast carcinomas. Clin. Cancer Res. *13*, 6064–6069.

Becker-Santos, D.D., Lonergan, K.M., Gronostajski, R.M., and Lam, W.L. (2017). Nuclear Factor I/B: A Master Regulator of Cell Differentiation with Paradoxical Roles in Cancer. EBioMedicine *22*, 2–9.

Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., and Nowak, M.A. (2007). Genetic progression and the waiting time to cancer. PLoS Comput. Biol. *3*, 2239–2246.

Bell, E.L., Klimova, T.A., Eisenbart, J., Moraes, C.T., Murphy, M.P., Budinger, G.R.S., and Chandel, N.S. (2007). The Qo site of the mitochondrial complex III is required for the transduction of hypoxic signaling via reactive oxygen species production. J. Cell Biol. *177*, 1029–1036.

Bertucci, F., Ng, C.K.Y., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N., Soria, J.C., Dien, A.T., Adnani, Y., Kamal, M., et al. (2019). Genomic characterization of metastatic breast cancers. Nature *569*, 560–564.

Biscotti, M.A., Olmo, E., and Heslop-Harrison, J.S. (Pat. (2015). Repetitive DNA in eukaryotic genomes. Chromosom. Res. *23*, 415–420.

Borowsky, A.D. (2011). Choosing a mouse model: Experimental biology in context-the utility and limitations of mouse models of breast cancer. Cold Spring Harb. Perspect. Biol. *3*, 1–16.

Brabletz, T., Kalluri, R., Nieto, M.A., and Weinberg, R.A. (2018). EMT in cancer. Nat. Rev. Cancer *18*, 128–134.

Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system y. *35*.

Bray, F., Ferlay, J., and Soerjomataram, I. (2018). Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. 394–424.

Brayer, K.J., Frerich, C.A., Kang, H., and Ness, S.A. (2016). Recurrent fusions in MYB and MYBL1 define a common, transcription factor-driven oncogenic pathway in salivary gland adenoid cystic carcinoma. Cancer Discov. *6*, 176–187.

Brett, B.T., Berquam-Vrieze, K.E., Nannapaneni, K., Huang, J., Scheetz, T.E., and Dupuy, A.J. (2011). Novel molecular and computational methods improve the accuracy of insertion site analysis in sleeping Beauty-Induced tumors. PLoS One *6*.

Britschgi, A., Duss, S., Kim, S., Couto, J.P., Brinkhaus, H., Koren, S., De Silva, D., Mertz, K.D., Kaup, D., Varga, Z., et al. (2017). The Hippo kinases LATS1 and 2 control human breast cell fate via crosstalk with ERα. Nat. Publ. Gr. *541*, 4–1.

Bronner, I.F., Otto, T.D., Zhang, M., Udenze, K., Wang, C., Quail, M.A., Jiang, R.H.Y., Adams, J.H., and Rayner, J.C. (2016). Quantitative insertion-site sequencing (QIseq) for high throughput phenotyping of transposon mutants. Genome Res. *26*, 980–989.

Campbell, T.M., Castro, M.A.A., de Oliveira, K.G., Ponder, B.A.J., and Meyer, K.B. (2018). Era binding by transcription factors NFIB and YBX1 enables FGFR2 signaling to modulate estrogen responsiveness in breast cancer. Cancer Res. *78*, 410–421.

Carriço, C., Costa, J.M., and Santos, T.C. (2011). Carriço, C., Costa, J.M. & Santos, T.C. descrição da larva de neocordulia machadoi santos, costa & carriço, 2010 (Odonata: Corduliidae) para o Brasil. Biota Neotrop. *11*, 71–73.

Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E., and Fraser, M.J. (1989). Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. Virology *172*, 156–169.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Chaffer, C.L., Juan, B.P.S., Lim, E., and Weinberg, R.A. (2016). EMT , cell plasticity and metastasis. Cancer Metastasis Rev. 645–654.

Chambers, A.F., Groom, A.C., and MacDonald, I.C. (2002). Dissemination and growth of cancer cells in metastatic sites. Nat. Rev. Cancer *2*, 563–572.

Chang, C.Y., Pasolli, H.A., Giannopoulou, E.G., Guasch, G., Gronostajski, R.M., Elemento, O., and Fuchs, E. (2013). NFIB is a governor of epithelial–melanocyte stem cell behaviour in a shared niche. Nature *495*, 98–102.

Chapeau, E.A., Gembarska, A., Durand, E.Y., Mandon, E., Estadieu, C., Romanet, V., Wiesmann, M., Tiedt, R., Lehar, J., de Weck, A., et al. (2017). Resistance mechanisms to TP53-MDM2 inhibition identified by in vivo piggyBac transposon mutagenesis screen in an Arf $^{-/-}$ mouse model. Proc. Natl. Acad. Sci. *114*, 3151–3156.

Chen, W., Hoffmann, A.D., Liu, H., and Liu, X. (2018). Organotropism: new insights into molecular mechanisms of breast cancer metastasis. Npj Precis. Oncol. *2*.

Chen, Y., Xiang, H., Zhang, Y., Wang, J., and Yu, G. (2015). Loss of PCDH9 is associated with the differentiation of tumor cells and metastasis and predicts poor survival in gastric cancer. Clin. Exp. Metastasis *32*, 417–428.

Clark, A.G., and Vignjevic, D.M. (2015). Modes of cancer cell invasion and the role of the microenvironment. Curr. Opin. Cell Biol. *36*, 13–22.

Clarke, M.F., Dick, J.E., Dirks, P.B., Eaves, C.J., Jamieson, C.H.M., Jones, D.L., Visvader, J., Weissman, I.L., and Wahl, G.M. (2006). Cancer stem cells - Perspectives on current status and future directions: AACR workshop on cancer stem cells. Cancer Res. *66*, 9339–9344.

Copeland, N.G., and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. Nat. Rev. Cancer *10*, 696–706.

Craig, N.L. (1997). TARGET SITE SELECTION IN TRANSPOSITION.

Curtis, C., et al (2012). The Cancer Genome Atlas Network. Nature *490*, 61–70.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D.,

Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Davis, R.P., Nemes, C., Varga, E., Freund, C., Kosmidis, G., Gkatzis, K., de Jong, D., Szuhai, K., Dinnyés, A., and Mummery, C.L. (2013). Generation of induced pluripotent stem cells from human foetal fibroblasts using the Sleeping Beauty transposon gene delivery system. Differentiation *86*, 30–37.

Dawson, S.J., Rueda, O.M., Aparicio, S., and Caldas, C. (2013). A new genome-driven integrated classification of breast cancer and its implications. EMBO J. *32*, 617–628.

Denny, S.K., Yang, D., Chuang, C.H., Brady, J.J., Lim, J.S.S., Grüner, B.M., Chiou, S.H., Schep, A.N., Baral, J., Hamard, C., et al. (2016a). Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. Cell *166*, 328–342.

Denny, S.K., Yang, D., Sage, J., William, J., Winslow, M.M., Denny, S.K., Yang, D., Chuang, C., Brady, J.J., Lim, J.S., et al. (2016b). Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility Article Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. 1–15.

Devon, R.S., Porteous, D.J., and Brookes, A.J. (1995). Splinkerettes - improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res. *23*, 1644–1645.

Dey, N., De, P., and Leyland-Jones, B. (2017). PI3K-AKT-mTOR inhibitors in breast cancers: From tumor cell signaling to clinical trials. Pharmacol. Ther. *175*, 91–106.

Diepenbruck, M., and Christofori, G. (2016). Epithelial-mesenchymal transition (EMT) and metastasis: Yes, no, maybe? Curr. Opin. Cell Biol. *43*, 7–13.

Diepenbruck, M., Tiede, S., Saxena, M., Ivanek, R., Kalathur, R.K.R., Lüönd, F., Meyer-Schaller, N., and Christofori, G. (2017). MIR-1199-5p and Zeb1 function in a double-negative feedback loop potentially coordinating EMT and tumour metastasis. Nat. Commun. *8*, 1–14.

Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., et al. (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature *464*, 999–1005.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell *122*, 473–483.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doebar, S.C., Krol, N.M., van Marion, R., Brouwer, R.W.W., van Ijcken, W.F.J., Martens, J.M., Dinjens, W.N.M., and van Deurzen, C.H.M. (2019). Progression of ductal carcinoma in situ to invasive breast cancer: comparative genomic sequencing. Virchows Arch. *474*, 247–251.

Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., et al. (2011). Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. *25*, 1470–1475.

Dupuy, A.J. (2010). Transposon-based screens for cancer gene discovery in mouse models. Semin. Cancer Biol. *20*, 261–268.

Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *436*.

Easwaran, H., Tsai, H.C., and Baylin, S.B. (2014). Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. Mol. Cell *54*, 716–727.

Fahad Ullah, M. (2019). Breast Cancer: Current Perspectives on the Disease Status.

Faltas, B. (2012). Cornering metastases: Therapeutic targeting of circulating tumor cells and stem cells. Front. Oncol. *2 JUL*, 1–7.

Fane, M.E., Chhabra, Y., Hollingsworth, D.E.J., Simmons, J.L., Spoerri, L., Oh, T.G., Chauhan, J., Chin, T., Harris, L., Harvey, T.J., et al. (2017). NFIB Mediates BRN2 Driven Melanoma Cell Migration and Invasion Through Regulation of EZH2 and MITF. EBioMedicine *16*, 63–75.

Fares, J., Kanojia, D., Rashidi, A., Ulasov, I., and Lesniak, M.S. (2020). Landscape of combination therapy trials in breast cancer brain metastasis. Int. J. Cancer.

Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T.C., Choi, H., El Rayes, T., Ryu, S., Troeger, J., et al. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature *527*.

Fischer, S.E.J., Wienholds, E., and Plasterk, R.H.A. (2001). Regulated transposition of a fish transposon in the mouse germ line. *98*, 6759–6764.

Flister, M.J., and Bergom, C. (2018). Genetic Modifiers of the Breast Tumor Microenvironment. Trends in Cancer *4*, 429–444.

Fragomeni, S.M., Sciallis, A., and Jeruss, J.S. (2018). Molecular Subtypes and Local-Regional Control of Breast Cancer. Surg. Oncol. Clin. N. Am. *27*, 95–120.

Friedel, R.H., Friedel, C.C., Bonfert, T., Shi, R., Rad, R., and Soriano, P. (2013). Clonal Expansion Analysis of Transposon Insertions by High-Throughput Sequencing Identifies Candidate Cancer Genes in a PiggyBac Mutagenesis Screen. PLoS One *8*, 1–12.

Friedrich, M.J., Rad, L., Bronner, I.F., Strong, A., Wang, W., Weber, J., Mayho, M., Ponstingl, H., Engleitner, T., Grove, C., et al. (2017). Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *12*.

Gala, K., Li, Q., Sinha, A., Razavi, P., Dorso, M., Sanchez-Vega, F., Chung, Y.R., Hendrickson, R., Hsieh, J.J., Berger, M., et al. (2018). KMT2C mediates the estrogen dependence of breast cancer through regulation of ERα enhancer function. Oncogene *37*, 4692–4710.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. *6*, pl1–pl1.

Geurts, A.M., Yang, Y., Clark, K.J., Liu, G., Cui, Z., Dupuy, A.J., Bell, J.B., Largaespada, D.A., and Hackett, P.B. (2003). Gene Transfer into Genomes of Human Cells by the Sleeping Beauty Transposon System. Mol. Ther. *8*, 108–117.

Giancotti, F.G. (2013). XMechanisms governing metastatic dormancy and reactivation. Cell *155*, 750.

Gkountela, S., Castro-Giner, F., Szczerba, B.M., Vetter, M., Landin, J., Scherrer, R., Krol, I., Scheidmann, M.C., Beisel, C., Stirnimann, C.U., et al. (2019). Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. Cell *176*, 98-112.e14.

Grabundzija, I., Wang, J., Sebe, A., Erdei, Z., Kajdi, R., Devaraj, A., Steinemann, D., Szuhai, K., Stein, U., Cantz, T., et al. (2013). Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. Nucleic Acids Res. *41*, 1829–1847.

Granit, R.Z., Masury, H., Condiotti, R., Fixler, Y., Gabai, Y., Glikman, T., Dalin, S., Winter, E., Nevo, Y., Carmon, E., et al. (2018). Regulation of Cellular Heterogeneity and Rates of Symmetric and Asymmetric Divisions in Triple-Negative Breast Cancer. Cell Rep. *24*, 3237–3250.

Gronostajski, R.M. (2000). Roles of the NFI/CTF gene family in transcription and development. Gene *249*, 31–45.

Gründer, A., Ebel, T.T., Mallo, M., Schwarzkopf, G., Shimizu, T., Sippel, A.E., and Schrewe, H. (2002). Nuclear factor I-B (Nfib) deficient mice have severe lung hypoplasia. Mech. Dev. *112*, 69–77.

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil, E., Brewer, D.S., Kallio, H.M.L., Högnäs, G., Annala, M., et al. (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353–357.

Gupta, G.P., and Massagué, J. (2006). Cancer Metastasis: Building a Framework. Cell *127*, 679–695.

Győrffy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res. Treat. *123*, 725–731.

Hall, A. (2009). The cytoskeleton and cancer. Cancer Metastasis Rev. *28*, 5–14.

Hanahan, D., and Folkman, J. (1996). Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell *86*, 353–364.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell.

Harris, L., Genovesi, L.A., Gronostajski, R.M., Wainwright, B.J., and Piper, M. (2015). Nuclear factor one transcription factors: Divergent functions in developmental versus adult stem cell populations. Dev. Dyn. *244*, 227–238.

Hilsenbeck, S., Michalowska, A.M., Medina, D., Green, J.E., Tsimelzon, A., Behbod, F., Zhang, M., Kittrell, F., Landis, M.D., Rosen, J.M., et al. (2008). Identification of Tumor-Initiating Cells in a p53-Null Mouse Model of Breast Cancer. Cancer Res. *68*, 4674–4682.

Hinohara, K., and Polyak, K. (2019). Intratumoral Heterogeneity: More Than Just Mutations. Trends Cell Biol. *29*, 569–579.

Hoadley, K.A., Siegel, M.B., Kanchi, K.L., Miller, C.A., Ding, L., Zhao, W., He, X., Parker, J.S., Wendl, M.C., Fulton, R.S., et al. (2016). Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases. PLoS Med. *13*, 1–20.

Holmfeldt, P., Pardieck, J., Saulsberry, A.C., Nandakumar, S.K., Finkelstein, D., Gray, J.T., Persons, D.A., and McKinney-Freeman, S. (2013). Nfix is a novel regulator of murine hematopoietic stem and progenitor cell survival. Blood *122*, 2987–2996.

Hu, M., and Polyak, K. (2008). Molecular characterisation of the tumour microenvironment in breast cancer. Eur. J. Cancer *44*, 2760–2765.

Hu, Y., and Smyth, G.K. (2009). ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods *347*, 70–78.

Hu, S., Ni, W., Sai, W., Zhang, H., Cao, X., Qiao, J., Sheng, J., Guo, F., and Chen, C. (2011). Sleeping Beauty-mediated knockdown of sheep myostatin by RNA interference. Biotechnol. Lett. *33*, 1949–1953.

Huang, X., Guo, H., Kang, J., Choi, S., Zhou, T.C., Tammana, S., Lees, C.J., Li, Z., Milone, M., Levine, B.L., et al. (2008). Sleeping Beauty Transposon-mediated Engineering of Human Primary T Cells for Therapy of CD19 + Lymphoid Malignancies. *16*, 580–589.

Huang, X., Guo, H., Tammana, S., Jung, Y., Mellgren, E., Bassi, P., and Zhou, X. (2010). Gene Transfer Efficiency and Genome-Wide Integration Profiling of Sleeping Beauty , Tol2 , and PiggyBac Transposons in Human Primary T Cells. YMTHE *18*, 1803–1813.

Huang, X., Wilber, A.C., Bao, L., Tuong, D., Tolar, J., Orchard, P.J., Levine, B.L., June, C.H., Mcivor, R.S., Blazar, B.R., et al. (2019). Stable gene transfer and expression in human primary T cells by the Sleeping Beauty transposon system. *107*, 483–492.

Ishay-Ronen, D., and Christofori, G. (2019). Targeting cancer cell metastasis by converting cancer cells into fat. Cancer Res. *79*, 5471–5475.

Ishay-Ronen, D., Diepenbruck, M., Kalathur, R.K.R., Sugiyama, N., Tiede, S., Ivanek, R., Bantug, G., Morini, M.F., Wang, J., Hess, C., et al. (2019). Gain Fat—Lose Metastasis: Converting Invasive Breast Cancer Cells into Adipocytes Inhibits Cancer Metastasis. Cancer Cell *35*, 17-32.e6.

Ivics, Z., and Izsvák, Z. (2010). The expanding universe of transposon technologies for gene and cell engineering. Mob. DNA *1*, 1–15.

Ivics, Z., Li, M.A., Mátés, L., Boeke, J.D., Nagy, A., Bradley, A., and Izsvák, Z. (2009). Transposon-mediated genome manipulation in vertebrates. Nat. Methods *6*, 415–422.

Iwamoto, T., Niikura, N., Ogiya, R., Yasojima, H., Watanabe, K. ichi, Kanbayashi, C., Tsuneizumi, M., Matsui, A., Fujisawa, T., Iwasa, T., et al. (2019). Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. Sci. Rep. *9*, 1–8.

Izsvák, Z., Ivics, Z., and Plasterk, R.H. (2000). Sleeping beauty, a wide host-range transposon vector for genetic trensformation in vertebrates. J. Mol. Biol. *302*, 93–102.

Johnston, D.S. (2002). THE ART AND DESIGN OF GENETIC SCREENS : DROSOPHILA MELANOGASTER. *3*, 176–188.

Joyce, J.A. (2005). Therapeutic targeting of the tumor microenvironment. Cancer Cell *7*, 513–520.

Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S., and Carroll, J.S. (2016). FOXA1

Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. Cell Rep. *17*, 2715–2723.

Kahlig, K.M., Saridey, S.K., Kaja, A., Daniels, M.A., George, A.L., and Matthew, H. (2009). Multiplexed transposon-mediated stable gene transfer in human cells.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771–776.

Kang, Y., and Pantel, K. (2013). Tumor Cell Dissemination: Emerging Biological Insights from Animal Models and Cancer Patients. Cancer Cell *23*, 573–581.

Karakas, B., Bachman, K.E., and Park, B.H. (2006). Mutation of the PIK3CA oncogene in human cancers. Br. J. Cancer *94*, 455–459.

Kaufman, P.D., and Rio, D.C. (1992). P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. Cell *69*, 27–39.

Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M.C.U., Voduc, D., Speers, C.H., Nielsen, T.O., and Gelmon, K. (2010). Metastatic behavior of breast cancer subtypes. J. Clin. Oncol. *28*, 3271–3277.

Kessenbrock, K., Plaks, V., and Werb, Z. (2010). Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. Cell *141*, 52–67.

Kidwell, M.G. (2002). Transposable elements and the evolution of genome size in eukaryotes. Genetica *115*, 49–63.

Kim, K.M., An, A.R., Park, H.S., Jang, K.Y., Moon, W.S., Kang, M.J., Lee, Y.C., Ku, J.H., and Chung, M.J. (2018). Combined expression of protein disulfide isomerase and endoplasmic reticulum oxidoreductin 1-α is a poor prognostic marker for non-small cell lung cancer. Oncol. Lett. *16*, 5753–5760.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., et al. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Koedoot, E., Fokkelman, M., Rogkoti, V.M., Smid, M., van de Sandt, I., de Bont, H., Pont, C., Klip, J.E., Wink, S., Timmermans, M.A., et al. (2019). Uncovering the signaling landscape controlling breast cancer cell migration identifies novel metastasis driver genes. Nat. Commun. *10*, 1–16.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature.

Kong, J., Wang, F., Brenton, J.D., and Adams, D.J. (2010). Slingshot : a PiggyBac based transposon system insertional mutagenesis. *38*, 1–9.

Koren, S., and Bentires-Alj, M. (2013). Mouse models of PIK3CA mutations: One mutation initiates heterogeneous mammary tumors. FEBS J. *280*, 2758–2765.

Koren, S., and Bentires-Alj, M. (2015). Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. Mol. Cell *60*, 537–546.

Koren, S., Reavie, L., Couto, J.P., De Silva, D., Stadler, M.B., Roloff, T., Britschgi, A.,

Eichlisberger, T., Kohler, H., Aina, O., et al. (2015a). PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. Nature *525*, 114–118.

Koren, S., Reavie, L., Couto, J.P., De Silva, D., Stadler, M.B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., et al. (2015b). PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. Nature *525*, 114–118.

Koscielny, S., and Tubiana, M. (2010). Parallel progression of tumour and metastases. Nat. Rev. Cancer *10*, 156.

Kreike, B., van Kouwenhove, M., Horlings, H., Weigelt, B., Peterse, H., Bartelink, H., and van de Vijver, M.J. (2007). Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. Breast Cancer Res. *9*, 1–14.

Kreso, A., and Dick, J.E. (2014). Evolution of the cancer stem cell model. Cell Stem Cell *14*, 275–291.

Krøigård, A.B., Larsen, M.J., Lænkholm, A.V., Knoop, A.S., Jensen, J.D., Bak, M., Mollenhauer, J., Kruse, T.A., and Thomassen, M. (2015). Clonal expansion and linear genome evolution through breast cancer progression from pre-invasive stages to asynchronous metastasis. Oncotarget *6*, 5634–5649.

Krøigård, A.B., Larsen, M.J., Brasch-Andersen, C., Lænkholm, A.V., Knoop, A.S., Jensen, J.D., Bak, M., Mollenhauer, J., Thomassen, M., and Kruse, T.A. (2017). Genomic analyses of breast cancer progression reveal distinct routes of metastasis emergence. Sci. Rep. *7*, 1–9.

Krøigård, A.B., Larsen, M.J., Lænkholm, A.V., Knoop, A.S., Jensen, J.D., Bak, M., Mollenhauer, J., Thomassen, M., and Kruse, T.A. (2018). Identification of metastasis driver genes by massive parallel sequencing of successive steps of breast cancer progression. PLoS One *13*, 1–18.

De La Rosa, J., Weber, J., Friedrich, M.J., Li, Y., Rad, L., Ponstingl, H., Liang, Q., De Quirós, S.B., Noorani, I., Metzakopian, E., et al. (2017). A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating tumor suppressor genes. Nat. Genet. *49*, 730–741.

Lamouille, S., Xu, J., and Derynck, R. (2014). Fakultas Psikologi Dan Sosial Budaya Universitas Islam Indonesia Yogyakarta. Nat. Rev. Mol. Cell Biol. *15*, 178–196.

Li, M.A., Pettitt, S.J., Eckert, S., Ning, Z., Rice, S., Cadinanos, J., Yusa, K., Conte, N., and Bradley, A. (2013). The piggyBac Transposon Displays Local and Distant Reintegration Preferences and Can Cause Mutations at Noncanonical Integration Sites. Mol. Cell. Biol. *33*, 1317–1330.

Li, S., Zhang, A., Xue, H., Li, D., and Liu, Y. (2017). One-Step piggyBac Transposon-Based CRISPR / Cas9 Activation of Multiple Genes. Mol. Ther. Nucleic Acid *8*, 64–76.

Li, Y., Gong, W., Ma, X., Sun, X., Jiang, H., and Chen, T. (2015). Smad7 maintains epithelial phenotype of ovarian cancer stem.like cells and supports tumor colonization by mesenchymal.epithelial transition. Mol. Med. Rep. *11*, 309–316.

Li, Y., Sun, C., Tan, Y., Li, L., Zhang, H., Liang, Y., Zeng, J., and Zou, H. (2020). Transcription levels and prognostic significance of the NFI family members in human cancers. PeerJ *2020*, 1–34.

Liang, Q., Kong, J., Stalker, J., and Bradley, A. (2009). Chromosomal Mobilization and

Reintegration of Sleeping Beauty and PiggyBac Transposons. *408*, 404–408.

Liu, H., Patel, M.R., Prescher, J.A., Patsialou, A., Qian, D., Lin, J., Wen, S., Chang, Y.F., Bachmann, M.H., Shimono, Y., et al. (2010). Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models. Proc. Natl. Acad. Sci. U. S. A. *107*, 18115–18120.

Liu, L., Kimball, S., Liu, H., Holowatyj, A., and Yang, Z.-Q. (2015). Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. Oncotarget *6*.

Liu, R.Z., Vo, T.M., Jain, S., Choi, W.S., Garcia, E., Monckton, E.A., Mackey, J.R., and Godbout, R. (2019a). NFIB promotes cell survival by directly suppressing p21 transcription in TP53-mutated triple-negative breast cancer. J. Pathol. *247*, 186–198.

Liu, Z.Y., Chen, J.H., Yuan, W.Z., Ruan, H.L., Shu, Y., Ji, J.T., Wu, L., Tang, Q., Zhou, Z.L., Zhang, X.D., et al. (2019b). Nuclear factor I/B promotes colorectal cancer cell proliferation, epithelial-mesenchymal transition and 5-fluorouracil resistance. Cancer Sci. *110*, 86–98.

Low, S.K., Zembutsu, H., and Nakamura, Y. (2018). Breast cancer: The translation of big genomic data to cancer precision medicine. Cancer Sci. *109*, 497–506.

Lukacsovich, T., Hamada, N., Miyazaki, S., and Kimpara, A. (2008). A New Versatile Gene-Trap Vector for Insect Transgenics. *175*, 168–175.

Luzzi, K.J., MacDonald, I.C., Schmidt, E.E., Kerkvliet, N., Morris, V.L., Chambers, A.F., and Groom, A.C. (1998). Multistep nature of metastatic inefficiency: Dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am. J. Pathol. *153*, 865–873.

Lv, J., Zhu, P., Zhang, X., Zhang, L., Chen, X., Lu, F., Yu, Z., and Liu, S. (2017). PCDH9 acts as a tumor suppressor inducing tumor cell arrest at G0/G1 phase and is frequently methylated in hepatocellular carcinoma. Mol. Med. Rep. *16*, 4475–4482.

Magers, M.J., Iczkowski, K.A., Montironi, R., Grignon, D.J., Zhang, S., Williamson, S.R., Yang, X., Wang, M., Osunkoya, A.O., Lopez-Beltran, A., et al. (2019). MYB-NFIB gene fusion in prostatic basal cell carcinoma: clinicopathologic correlates and comparison with basal cell adenoma and florid basal cell hyperplasia. Mod. Pathol. *32*, 1666–1674.

Mani, S.A., Guo, W., Liao, M., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. 704–715.

Marchiò, C., Weigelt, B., and Reis-Filho, J.S. (2010). Adenoid cystic carcinomas of the breast and salivary glands (or "The strange case of Dr Jekyll and Mr Hyde" of exocrine gland carcinomas). J. Clin. Pathol. *63*, 220–228.

Marshall, J.C.A., Collins, J.W., Nakayama, J., Horak, C.E., Liewehr, D.J., Steinberg, S.M., Albaugh, M., Vidal-Vanaclocha, F., Palmieri, D., Barbier, M., et al. (2012). Effect of inhibition of the lysophosphatidic acid receptor 1 on metastasis and metastatic dormancy in breast cancer. J. Natl. Cancer Inst. *104*, 1306–1319.

Martynoga, B., Mateo, J.L., Zhou, B., Andersen, J., Achimastou, A., Urbán, N., van den Berg, D., Georgopoulou, D., Hadjur, S., Wittbrodt, J., et al. (2013). Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. Genes Dev. *27*, 1769–1786.

Marusyk, A., and Polyak, K. (2010). Tumor heterogeneity: Causes and consequences.

Biochim. Biophys. Acta - Rev. Cancer *1805*, 105–117.

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? Nat. Rev. Cancer *12*, 323–334.

Massagué, J., and Obenauf, A.C. (2016). Metastatic colonization by circulating tumour cells. Nature *529*.

Mátés, L., Izsvák, Z., Ivics, Z., Delbrück, M., and Medicine, M. (2007). Technology transfer from worms and flies to vertebrates : transposition-based genome manipulations and their future perspectives. *8*, 1–19.

De Mattos-Arruda, L., Sammut, S.J., Ross, E.M., Bashford-Rogers, R., Greenstein, E., Markus, H., Morganella, S., Teng, Y., Maruvka, Y., Pereira, B., et al. (2019). The Genomic and Immune Landscapes of Lethal Metastatic Breast Cancer. Cell Rep. *27*, 2690-2708.e10.

Mayer, I.A., Abramson, V.G., Formisano, L., Balko, J.M., Estrada, M. V., Sanders, M.E., Juric, D., Solit, D., Berger, M.F., Won, H.H., et al. (2017). A Phase Ib Study of Alpelisib (BYL719), a PI3Ka-Specific Inhibitor, with Letrozole in ERþ/HER2 Metastatic Breast Cancer. Clin. Cancer Res. *23*, 26–34.

McDonald, E.S., Clark, A.S., Tchou, J., Zhang, P., and Freedman, G.M. (2016). Clinical diagnosis and management of breast cancer. J. Nucl. Med. *57*, 9S-16S.

Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328–337.

Meir, Y.J., Lin, A., Huang, M., Lin, J., Weirauch, M.T., Chou, H., Lin, S.A., and Wu, S.C. (2019). A versatile , highly efficient , and potentially safer piggyBac transposon system for mammalian genome manipulations.

Meyer, D.S., Brinkhaus, H., Müller, U., Müller, M., Cardiff, R.D., and Bentires-Alj, M. (2011). Luminal expression of PIK3CA mutant H1047R in the mammary gland induces heterogeneous tumors. Cancer Res. *71*, 4344–4351.

Meyer, D.S., Koren, S., Leroy, C., Brinkhaus, H., Müller, U., Klebba, I., Müller, M., Cardiff, R.D., and Bentires-Alj, M. (2013a). Expression of PIK3CA mutant E545K in the mammary gland induces heterogeneous tumors but is less potent than mutant H1047R. Oncogenesis *2*.

Meyer, D.S., Koren, S., Leroy, C., Brinkhaus, H., Mu, U., Klebba, I., and Mu, M. (2013b). Expression of PIK3CA mutant E545K in the mammary gland induces heterogeneous tumors but is less potent than mutant H1047R.

Miller, T.W. (2012). Initiating breast cancer by PIK3CA mutation. Breast Cancer Res. *14*, 2010–2011.

Miller, T.W., Rexer, B.N., Garrett, J.T., and Arteaga, C.L. (2011). Mutations in the phosphatidylinositol 3-kinase pathway: Role in tumor progression and therapeutic implications in breast cancer. Breast Cancer Res. *13*.

Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. Nature *436*, 518–524.

Miron, A., Varadi, M., Carrasco, D., Li, H., Luongo, L., Kim, H.J., Park, S.Y., Cho, E.Y., Lewis, G., Kehoe, S., et al. (2010). PIK3CA mutations in in situ and invasive breast

carcinomas. Cancer Res. *70*, 5674–5678.

Mitra, R., Fain-Thornton, J., and Craig, N.L. (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. EMBO J. *27*, 1097–1109.

Momose, T., Gehring, W.J., Salo, E., and Gonza, C. (2003). Transgenic planarian lines obtained by electroporation using transposon-derived vectors and an eye-specific GFP marker.

Monjezi, R., Miskey, C., Gogishvili, T., Schleef, M., Schmeer, M., Einsele, H., Ivics, Z., and Hudecek, M. (2017). Enhanced CAR T-cell engineering using non-viral Sleeping Beauty transposition from minicircle vectors. 186–194.

Moon, H., Hwang, K., Kim, J., Sung, H., Lee, M., Jung, E., Ko, E., Han, W., and Noh, D. (2011a). NFIB is a potential target for estrogen receptor-negative breast cancers. Mol. Oncol. *5*, 538–544.

Moon, H.G., Hwang, K.T., Kim, J.A., Kim, H.S., Lee, M.J., Jung, E.M., Ko, E., Han, W., and Noh, D.Y. (2011b). NFIB is a potential target for estrogen receptor-negative breast cancers. Mol. Oncol. *5*, 538–544.

Mukherjee, S., Huntemann, M., Ivanova, N., Kyrpides, N.C., and Pati, A. (2015). Large-scale contamination of microbial isolate genomes by illumina Phix control. Stand. Genomic Sci. *10*, 1–4.

Munzone, E., Botteri, E., Sciandivasci, A., Curigliano, G., Nolè, F., Mastropasqua, M., Rotmensz, N., Colleoni, M., Esposito, A., Adamoli, L., et al. (2012). Prognostic value of Ki-67 labeling index in patients with node-negative, triple-negative breast cancer. Breast Cancer Res. Treat. *134*, 277–282.

Murtagh, J., Martin, F., and Gronostajski, R.M. (2003). The Nuclear Factor I (NFI) Gene Family in Mammary Gland Development and Function. J. Mammary Gland Biol. Neoplasia *8*, 241–254.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. Nature *472*, 90–95.

Nedeljković, M., and Damjanović, A. (2019). Mechanisms of Chemotherapy Resistance in Triple-Negative Breast Cancer—How We Can Rise to the Challenge. Cells *8*, 957.

Nguyen, D.X., Bos, P.D., and Massagué, J. (2009). Metastasis: From dissemination to organ-specific colonization. Nat. Rev. Cancer *9*, 274–284.

Nguyen, L. V., Vanner, R., Dirks, P., and Eaves, C.J. (2012). Cancer stem cells: An evolving concept. Nat. Rev. Cancer *12*, 133–143.

Nieto, M.A., Huang, R.Y.Y.J., Jackson, R.A.A., and Thiery, J.P.P. (2016). EMT: 2016. Cell.

Nihan Kilinc, A., Sugiyama, N., Reddy Kalathur, R.K., Antoniadis, H., Birogul, H., Ishay-Ronen, D., George, J.T., Levine, H., Kumar Jolly, M., and Christofori, G. (2020). Histone deacetylases, Mbd3/NuRD, and Tet2 hydroxylase are crucial regulators of epithelial–mesenchymal plasticity and tumor metastasis. Oncogene *39*, 1498–1513.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science (80-.). *194*, 23–28.

O'Shaughnessy, J. (2005). Extending Survival with Chemotherapy in Metastatic Breast Cancer. Oncologist *10*, 20–29.

Ocaña, O.H., Córcoles, R., Fabra, Á., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., and Nieto, M.A. (2012). Metastatic Colonization Requires the Repression of the Epithelial-Mesenchymal Transition Inducer Prrx1. Cancer Cell *22*, 709–724.

Oskarsson, T., Batlle, E., and Massagué, J. (2014). Metastatic stem cells: Sources, niches, and vital pathways. Cell Stem Cell *14*, 306–321.

Paran, H., and Paran, D. (1996). Angiogenesis and tumor growth. N. Engl. J. Med. *334*, 921.

Pastushenko, I., and Blanpain, C. (2019). EMT Transition States during Tumor Progression and Metastasis. Trends Cell Biol. *29*.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468.

Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K.M., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., et al. (2016). The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat. Commun. *7*.

Pérez-Casellas, L.A., Wang, X., Howard, K.D., Rehage, M.W., Strong, D.D., and Linkhart, T.A. (2009). Nuclear Factor I transcription factors regulate IGF binding protein 5 gene transcription in human osteoblasts. Biochim. Biophys. Acta - Gene Regul. Mech. *1789*, 78–87.

Pernas, S., Petit, A., Climent, F., Paré, L., Perez-Martin, J., Ventura, L., Bergamino, M., Galván, P., Falo, C., Morilla, I., et al. (2019). PAM50 Subtypes in Baseline and Residual Tumors Following Neoadjuvant Trastuzumab-Based Chemotherapy in HER2-Positive Breast Cancer: A Consecutive-Series From a Single Institution. Front. Oncol. *9*.

Perou, C.M., Sørile, T., Eisen, M.B., Van De Rijn, M., Jeffrey, S.S., Ress, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Persson, M., Andrén, Y., Mark, J., Horlings, H.M., Persson, F., and Stenman, G. (2009). Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc. Natl. Acad. Sci. U. S. A. *106*, 18740–18744.

Piper, M., Moldrich, R.X., Lindwall, C., Little, E., Barry, G., Mason, S., Sunn, N., Kurniawan, N.D., Gronostajski, R.M., and Richards, L.J. (2009). Multiple non-cell-autonomous defects underlie neocortical callosal dysgenesis in Nfib-deficient mice. Neural Dev. *4*, 1–16.

Piper, M., Barry, G., Hawkins, J., Mason, S., Lindwall, C., Little, E., Sarkar, A., Smith, A.G., Moldrich, R.X., Boyle, G.M., et al. (2010). NFIA controls telencephalic progenitor cell differentiation through repression of the Notch effector Hes1. J. Neurosci. *30*, 9127–9139.

Piper, M., Barry, G., Harvey, T.J., McLeay, R., Smith, A.G., Harris, L., Mason, S., Stringer, B.W., Day, B.W., Wray, N.R., et al. (2014). NFIB-Mediated Repression of the Epigenetic Factor Ezh2 Regulates Cortical Development. J. Neurosci. *34*, 2921–2930.

Place, A.E., Jin Huh, S., and Polyak, K. (2011). The microenvironment in breast cancer

progression: Biology and implications for treatment. Breast Cancer Res. *13*.

Plasterk, R.H., and Izsva, Z. (1997). Molecular Reconstruction of Sleeping Beauty , a Tc1 -like Transposon from Fish , and Its Transposition in Human Cells. *91*, 501–510.

Polyak, K. (2007). Science in medicine Breast cancer : origins and evolution. J. Clin. Invest. *117*, 3155–3163.

Polyak, K., and Weinberg, R.A. (2009). Transitions between epithelial and mesenchymal states: Acquisition of malignant and stem cell traits. Nat. Rev. Cancer *9*, 265–273.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*.

Prat, A., Fan, C., Fernández, A., Hoadley, K.A., Martinello, R., Vidal, M., Viladot, M., Pineda, E., Arance, A., Muñoz, M., et al. (2015). Response and survival of breast cancer intrinsic subtypes following multi-agent neoadjuvant chemotherapy. BMC Med. *13*, 1–11.

Priestley, P., Baber, J., Lolkema, M.P., Steeghs, N., de Bruijn, E., Shale, C., Duyvesteyn, K., Haidari, S., van Hoeck, A., Onstenk, W., et al. (2019). Pan-cancer whole-genome analyses of metastatic solid tumours. Nature *575*, 210–216.

Quattrini, I., Pollino, S., Pazzaglia, L., Conti, A., Novello, C., Ferrari, C., Pignotti, E., Picci, P., and Benassi, M.S. (2015). Prognostic role of nuclear factor/IB and bone remodeling proteins in metastatic giant cell tumor of bone: A retrospective study. J. Orthop. Res. *33*, 1205–1211.

Rad, R., Rad, L., Wang, W., Cadinanos, J., Vassiliou, G., Rice, S., Campos, L.S., Yusa, K., Banerjee, R., Li, M.A., et al. (2010). PiggyBac transposon mutagenesis: A tool for cancer gene discovery in mice. Science (80-.). *330*, 1104–1107.

Rad, R., Rad, L., Wang, W., Strong, A., Ponstingl, H., Bronner, I.F., Mayho, M., Steiger, K., Weber, J., Hieber, M., et al. (2014). A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer.

Rad, R., Rad, L., Wang, W., Strong, A., Ponstingl, H., Bronner, I.F., Mayho, M., Steiger, K., Weber, J., Hieber, M., et al. (2015). A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. Nat. Genet. *47*, 47–56.

Rangel, R., Lee, S., Ban, K.H., Guzman-rojas, L., and Mann, M.B. (2016). Transposon mutagenesis identifies genes that cooperate with mutant Pten in breast cancer progression.

Robinson, D.R., Wu, Y.-M., Lonigro, R.J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V., et al. (2017). Integrative clinical genomics of metastatic cancer. Nature.

Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A.J.H., Anastassiadis, K., and Stewart, A.F. (2012). Transposon-mediated BAC transgenesis in human ES cells. *40*.

Russnes, H.G., Lingjærde, O.C., Børresen-Dale, A.L., and Caldas, C. (2017). Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. Am. J. Pathol. *187*, 2152–2162.

Saal, L.H., Holm, K., Maurer, M., Memeo, L., Su, T., Wang, X., Yu, J.S., Malmström, P.O., Mansukhani, M., Enoksson, J., et al. (2005). PIK3CA mutations correlate with hormone

receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. Cancer Res. *65*, 2554–2559.

Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F., Balgobind, A.V., Wind, K., Gracanin, A., Begthel, H., et al. (2018). A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. Cell *172*, 373-386.e10.

Sato, K., and Akimoto, K. (2017). Expression Levels of KMT2C and SLC20A1 Identified by Information-theoretical Analysis Are Powerful Prognostic Biomarkers in Estrogen Receptor-positive Breast Cancer. Clin. Breast Cancer *17*, e135–e142.

Schrijver, W.A.M.E., Selenica, P., Lee, J.Y., Ng, C.K.Y., Burke, K.A., Piscuoglio, S., Berman, S.H., Reis-Filho, J.S., Weigelt, B., Van Diest, P.J., et al. (2018). Mutation profiling of key cancer genes in primary breast cancers and their distant metastases. Cancer Res. *78*, 3112–3121.

Score, P.R., Belur, L.R., Frandsen, J.L., Guerts, J.L., Yamaguchi, T., Somia, N. V, Hackett, P.B., Largaespada, D.A., and Mcivor, R.S. (2006). Sleeping Beauty -Mediated Transposition and Long-term Expression in Vivo : Use of the LoxP / Cre Recombinase System to Distinguish Transposition-Specific Expression. Mol. Ther. *13*, 617–624.

Semenova, E.A., Kwon, M. chul, Monkhorst, K., Song, J.Y., Bhaskaran, R., Krijgsman, O., Kuilman, T., Peters, D., Buikhuisen, W.A., Smit, E.F., et al. (2016). Transcription Factor NFIB Is a Driver of Small Cell Lung Cancer Progression in Mice and Marks Metastatic Disease in Patients. Cell Rep. *16*, 631–643.

Shackleton, M., Quintana, E., Fearon, E.R., and Morrison, S.J. (2009). Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. Cell *138*, 822–829.

Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature *486*, 395–399.

Shi, C., Yang, Y., Zhang, L., Yu, J., Qin, S., Xu, H., and Gao, Y. (2019). Mir-200a-3p promoted the malignant behaviors of ovarian cancer cells through regulating PCDH9. Onco. Targets. Ther. *12*, 8329–8338.

Shin, H.Y., Willi, M., Yoo, K.H., Zeng, X., Wang, C., Metser, G., and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat. Genet. *48*, 904–911.

Siegel, M.B., He, X., Hoadley, K.A., Hoyle, A., Pearce, J.B., Garrett, A.L., Kumar, S., Moylan, V.J., Brady, C.M., Van Swearingen, A.E.D., et al. (2018). Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer. J. Clin. Invest. *128*, 1371–1383.

Singh, H., Manuri, P.R., Olivares, S., Dara, N., Dawson, M.J., Huls, H., Hackett, P.B., Kohn, D.B., Shpall, E.J., Champlin, R.E., et al. (2008). Redirecting Specificity of T-Cell Populations For CD19 Using the Sleeping Beauty System. 2961–2972.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., Van De Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A. *98*, 10869–10874.

Sørlie, T., Wang, Y., Xiao, C., Johnsen, H., Naume, B., Samaha, R.R., and Børresen-Dale, A.L. (2006). Distinct molecular mechanisms underlying clinically relevant subtypes of breast

cancer: Gene expression analyses across three different platforms. BMC Genomics *7*, 1–15.

Sosa, M.S., Bragado, P., and Aguirre-Ghiso, J.A. (2014). Mechanisms of disseminated cancer cell dormancy: An awakening field. Nat. Rev. Cancer *14*, 611–622.

Stankic, M., Pavlovic, S., Chin, Y., Brogi, E., Padua, D., Norton, L., Massagué, J., and Benezra, R. (2013). TGF-β-Id1 signaling opposes twist1 and promotes metastatic colonization via a mesenchymal-to-epithelial transition. Cell Rep. *5*, 1228–1242.

Starnes, L.M., Sorrentino, A., Ferracin, M., Negrini, M., Pelosi, E., Nervi, C., and Peschle, C. (2010). A transcriptome-wide approach reveals the key contribution of NFI-A in promoting erythroid differentiation of human CD34+ progenitors and CML cells. Leukemia *24*, 1220–1223.

Steeg, P.S. (2016). Targeting metastasis. Nat. Rev. Cancer *16*, 201–218.

Steele-Perkins, G., Plachez, C., Butz, K.G., Yang, G., Bachurski, C.J., Kinsman, S.L., Litwack, E.D., Richards, L.J., and Gronostajski, R.M. (2005). The Transcription Factor Gene Nfib Is Essential for both Lung Maturation and Brain Development. Mol. Cell. Biol. *25*, 685–698.

Stemke-Hale, K., Gonzalez-Angulo, A.M., Lluch, A., Neve, R.M., Kuo, W.L., Davies, M., Carey, M., Hu, Z., Guan, Y., Sahin, A., et al. (2008). An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. Cancer Res. *68*, 6084–6091.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. Nature *486*, 400–404.

Suárez-Cabrera, C., Quintana, R.M., Bravo, A., Casanova, M.L., Page, A., Alameda, J.P., Paramio, J.M., Maroto, A., Salamanca, J., Dupuy, A.J., et al. (2017). A transposon-based analysis reveals RASA1 is involved in triple-negative breast cancer. Cancer Res. *77*, 1357–1368.

Szczerba, B.M., Castro-Giner, F., Vetter, M., Krol, I., Gkountela, S., Landin, J., Scheidmann, M.C., Donato, C., Scherrer, R., Singer, J., et al. (2019). Neutrophils escort circulating tumour cells to enable cell cycle progression. Nature *566*, 553–557.

Takeda, H., Wei, Z., Koso, H., Rust, A.G., Yew, C.C.K., Mann, M.B., Ward, J.M., Adams, D.J., Copeland, N.G., Jenkins, N.A., et al. (2017). A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. Nat. Genet. *47*, 142–150.

Takei, N., Yoneda, A., Sakai-sawada, K., Kosaka, M., and Minomi, K. (2017). Hypoxia-inducible ERO1 α promotes cancer progression through modulation of integrin- β 1 modification and signalling in HCT116 colorectal cancer cells. Sci. Rep. 1–11.

Tam, W.L., and Weinberg, R.A. (2013). The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat. Med. *19*, 1438–1449.

Tanaka, A., Woltjen, K., Miyake, K., Hotta, A., Ikeya, M., and Yamamoto, T. (2013). Efficient and Reproducible Myogenic Differentiation from Human iPS Cells : Prospects for Modeling Miyoshi Myopathy In Vitro. *8*.

Tanaka, T., Kutomi, G., Kajiwara, T., Kukita, K., and Kochin, V. (2016). Cancer-associated oxidoreductase ERO1- a drives the production of VEGF via oxidative protein folding and regulating the mRNA level. *114*, 1227–1234.

Tang, M.H.E., Dahlgren, M., Brueffer, C., Tjitrowirjo, T., Winter, C., Chen, Y., Olsson, E., Wang, K., Törngren, T., Sjöström, M., et al. (2015). Remarkable similarities of chromosomal rearrangements between primary human breast cancers and matched distant metastases as revealed by whole-genome sequencing. Oncotarget *6*, 37169–37184.

Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Demsky, M., Fawcett, R., Francis-lang, H.L., et al. (2004). A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *36*, 283–287.

Tiwari, N., Gheldof, A., Tatari, M., and Christofori, G. (2012). EMT as the ultimate survival mechanism of cancer cells. Semin. Cancer Biol. *22*, 194–207.

Toh, T.B., Lim, J.J., and Chow, E.K.H. (2017). Epigenetics in cancer stem cells. Mol. Cancer *16*, 1–20.

Toska, E., Osmanbeyoglu, H.U., Castel, P., Chan, C., Hendrickson, R.C., Elkabets, M., Dickler, M.N., Scaltriti, M., Leslie, C.S., Armstrong, S.A., et al. (2017). PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. Science (80-.). *355*, 1324–1330.

Troester, M.A., Hoadley, K.A., Sørlie, T., Herbert, B.S., Børresen-Dale, A.L., Lønning, P.E., Shay, J.W., Kaufmann, W.K., and Perou, C.M. (2004). Cell-type-specific responses to chemotherapeutics in breast cancer. Cancer Res. *64*, 4218–4226.

Ungefroren, H., Sebens, S., Seidl, D., Lehnert, H., and Hass, R. (2011). Interaction of tumor cells with the microenvironment. Cell Commun. Signal. *9*, 18.

Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: Molecular insights and evolving paradigms. Cell *147*, 275–292.

Valkenburg, K.C., De Groot, A.E., and Pienta, K.J. (2018). Targeting the tumour stroma to improve cancer therapy. Nat. Rev. Clin. Oncol. *15*, 366–381.

Vasan, N., Toska, E., and Scaltriti, M. (2019). Overview of the relevance of PI3K pathway in HR-positive breast cancer. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. *30*, x3–x11.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.

Wang, G., Yang, L., Grishin, D., Rios, X., Ye, L.Y., Hu, Y., Li, K., Zhang, D., Church, G.M., and Pu, W.T. (2016). Efficient , footprint-free human iPSC genome editing by consolidation of Cas9 / CRISPR and piggyBac technologies. Nat. Protoc. *12*, 88–103.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *105*.

Wang, W., Bradley, A., and Huang, Y. (2009). A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells. 667–673.

Weber, J., de la Rosa, J., Grove, C.S., Schick, M., Rad, L., Baranov, O., Strong, A., Pfaus, A., Friedrich, M.J., Engleitner, T., et al. (2019). PiggyBac transposon tools for recessive screening identify B-cell lymphoma drivers in mice. Nat. Commun. *10*.

Weber, J., Braun, C.J., Saur, D., and Rad, R. (2020). In vivo functional screening for systems-level integrative cancer genomics. Nat. Rev. Cancer.

Weinberg, R.A. (2008). Mechanisms of malignant progression. Carcinogenesis *29*, 1092–1095.

Welch, D.R., and Hurst, D.R. (2019). Defining the Hallmarks of Metastasis. Cancer Res. *79*, 3011–3027.

Wilson, M.H., Coates, C.J., and George, A.L. (2007). PiggyBac transposon-mediated gene transfer in human cells. Mol. Ther. *15*, 139–145.

Wilson, M.H., Maiti, S.N., Singh, H., Olivares, S., Dawson, M.J., Huls, H., Lee, D.A., Rao, P.H., Kaminski, J.M., Nakazawa, Y., et al. (2010). piggyBac Transposon = Transposase System to Generate CD19-Specific T Cells for the Treatment of B-Lineage Malignancies. *437*, 427–437.

Wintermantel, T.M., Mayer, A.K., Schütz, G., and Greiner, E.F. (2002). Targeting mammary epithelial cells using a bacterial artificial chromosome. Genesis *33*, 125–130.

Woalder (2017). 乳鼠心肌提取 HHS Public Access. Physiol. Behav. *176*, 139–148.

Wu, C., Zhu, X., Liu, W., Ruan, T., Wan, W., and Tao, K. (2018). NFIB promotes cell growth, aggressiveness, metastasis and EMT of gastric cancer through the Akt/Stat3 signaling pathway. Oncol. Rep. *40*, 1565–1573.

Wu, N., Jia, D., Ibrahim, A.H., Bachurski, C.J., Gronostajski, R.M., and MacPherson, D. (2016). NFIB overexpression cooperates with <i>Rb/p53</i> deletion to promote small cell lung cancer. Oncotarget *7*.

Wu, S., Ying, G., Wu, Q., and Capecchi, M.R. (2007). Toward simpler and faster genome-wide mutagenesis in mice. Nat. Genet. *39*, 922–930.

Wu, S.C., Meir, Y.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S., and Kaminski, J.M. (2006). piggyBac is a flexible and highly active transposon as compared to Sleeping Beauty , Tol2 , and Mos1 in mammalian cells. *103*, 1–6.

Wyckoff, J.B., Jones, J.G., Condeelis, J.S., and Segall, J.E. (2000). A critical step in metastasis: In vivo analysis of intravasation at the primary tumor. Cancer Res. *60*, 2504–2511.

Wyld, L., Markopoulos, C., Leidenius, M., and Senkus-Konefka, E. (2017). Breast Cancer Management for Surgeons: A European Multidisciplinary Textbook.

Xing, D., Bakhsh, S., Melnyk, N., Isacson, C., Ho, J., Huntsman, D.G., Blake Gilks, C., Ronnett, B.M., and Horlings, H.M. (2017). Frequent NFIB -associated Gene Rearrangement in Adenoid Cystic Carcinoma of the Vulva. Int. J. Gynecol. Pathol. *36*, 289–293.

Xu, C., Qi, X., Du, X., Zou, H., Gao, F., Feng, T., Lu, H., and Li, S. (2017). piggyBac mediates efficient in vivo CRISPR library screening for tumorigenesis in mice. *114*.

Yan, L., Colandrea, V.J., and Hale, J.J. (2010). Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: Small molecule-based therapeutics for anemia. Expert Opin. Ther. Pat. *20*, 1219–1245.

Yang, D., Denny, S.K., Greenside, P.G., Chaikovsky, A.C., Brady, J.J., Ouadah, Y., Granja, J.M., Jahchan, N.S., Lim, J.S., Kwok, S., et al. (2018a). Intertumoral Heterogeneity in SCLC Is Influenced by the Cell Type of Origin. Cancer Discov. 1–17.

Yang, S., Yang, C., Yu, F., Ding, W., Hu, Y., Cheng, F., Zhang, F., Guan, B., Wang, X., Lu, L., et al. (2018b). Endoplasmic reticulum resident oxidase ERO1-Lalpha promotes hepatocellular carcinoma metastasis and angiogenesis through the S1PR1/STAT3/VEGF-A pathway. Cell Death Dis. *9*.

Yang, Z.Q., Imoto, I., Pimkhaokham, A., Shimada, Y., Sasaki, K., Oka, M., and Inazawa, J. (2001). A novel amplicon at 9p23-24 in squamous cell carcinoma of the esophagus that lies proximal to GASC1 and harbors NFIB. Japanese J. Cancer Res. *92*, 423–428.

Yates, L.R., and Campbell, P.J. (2012). Evolution of the cancer genome. Nat. Rev. Genet. *13*, 795–806.

Yates, L.R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T., Alexandrov, L.B., Larsimont, D., Davies, H., et al. (2015). Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat. Med. *21*, 751–759.

Yates, L.R., Knappskog, S., Wedge, D., Farmery, J.H.R., Gonzalez, S., Martincorena, I., Alexandrov, L.B., Van Loo, P., Haugland, H.K., Lilleng, P.K., et al. (2017). Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell *32*, 169-184.e7.

You, L., Chang, D., Du, H., and Zhao, Y. (2011). Biochemical and Biophysical Research Communications Genome-wide screen identifies PVT1 as a regulator of Gemcitabine sensitivity in human pancreatic cancer cells. Biochem. Biophys. Res. Commun. *407*, 1–6.

Yu, Z., Perera, I.R., Daeneke, T., Makuta, S., Tachibana, Y., Jasieniak, J.J., Mishra, A., Bäuerle, P., Spiccia, L., and Bach, U. (2016). Indium tin oxide as a semiconductor material in efficient p-type dye-sensitized solar cells. NPG Asia Mater. *8*, e305.

Yuan, T.L., and Cantley, L.C. (2008). PI3K pathway alterations in cancer : variations on a theme. 5497–5510.

Yusa, K., Rad, R., Takeda, J., and Bradley, A. (2009). Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *6*.

Zeichner, S.B., Terawaki, H., and Gogineni, K. (2016). A review of systemic treatment in metastatic triple-negative breast cancer. Breast Cancer Basic Clin. Res. *10*, 25–36.

Zhang, Y., and Weinberg, R.A. (2018). Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. Front. Med. *12*, 361–373.

Zhao, L., and Vogt, P.K. (2008). Class I PI3K in oncogenic cellular transformation. Oncogene *27*, 5486–5496.

Zhao, J.J., Liu, Z., Wang, L., Shin, E., Loda, M.F., and Roberts, T.M. (2005). The oncogenic properties of mutant p110α and p110β phosphatidylinositol 3-kinases in human mammary epithelial cells. Proc. Natl. Acad. Sci. U. S. A. *102*, 18443–18448.

Zhao, X., Pak, E., Ornell, K.J., Murphy, M.F.P., Mackenzie, E.L., Chadwick, E.J., Ponomaryov, T., Kelleher, J.F., and Segal, R.A. (2017). A transposon screen identifies loss of primary cilia as a mechanism of resistance to SMO inhibitors. Cancer Discov. *7*, 1436–1439.

Zheng, X., Carstens, J.L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., Wu, C.C., Lebleu, V.S., and Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature *527*, 525–530.

Zhou, B.O., Wang, G., Gao, S., Chen, Y.E., Jin, C., and Wang, Z. (2017). Expression of ERO1L in gastric cancer and its association with patient prognosis. 2298–2302.

Zito, E. (2015). Free Radical Biology and Medicine ERO1 : A protein disul fi de oxidase and H 2 O 2 producer. *83*, 299–304.

Zuazo-Gaztelu, I., and Casanovas, O. (2018). Unraveling the role of angiogenesis in cancer

ecosystems. Front. Oncol. *8*, 1–13.

9 | ACKNOWLEDGEMENTS

First, I would like to thank Momo for giving me the opportunity to work on such interesting projects and for his continuous support throughout the years. The high standards of science performed in the lab and the freedom to experiment made these years really fruitful. I also want to thank Gerhard, Matt and Uli for being in my thesis committee and supporting and advising me during the entire PhD.

I would like to acknowledge the former and current lab members of the Bentires lab, who helped me with numerous advices and critical discussions. Sharing their knowledges and help contributed to the improvement of my scientific skills, finally making me a better scientist. I thank them also for all the moments of sincere support during these years.

Special thanks go to the FMI and DBM core facilities, all internal and external collaborators from Novartis Basel for their valuable contribution.

Last but not least, I would like to thank all my friends at FMI, DBM, Biozentrum and outside. Particular thanks go to my family, who has always supported me in my choices and believed in them with my same passion.

10 |APPENDICES

10.1 Abbreviations

CAG	Modified chicken β-actin promoter
CDS	Coding sequence
CIS	Common insertion sites
CNAs	Copy number alterations
CTC	Circulating tumour cell
CRC	colorectal cancer
Cre-ERT2	Tamoxifen inducible activation of Cre-recombinase
CSC	Cancer stem cell
Ctrl	Control
CβASA	Carp β-actin splice acceptor
d	Days
DAPI	4′,6-Diamidin-2-phenylindol
DCIS	Ductal carcinoma *in situ*
DNA	Deoxyribonucleic acid
Dox	Doxycycline
EGFP	Enhanced Green fluorescent protein
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
ERO1A	Endoplasmic Reticulum Oxidoreductase 1 Alpha
Ero1l	ERO1-like protein alpha
ERK	Extracellular-signal-regulated kinase
ES cell	Embryonic stem cell
EZH2	Enhancer of zeste homolog 2

FACS	Fluorescence-activated cell sorting
GOF	Gain of function
H3K4	Histone 3 Lysine 4
HER-2	Receptor tyrosine-protein kinase erbB-2
HIF	hypoxia-inducible factors
iPSCs	Induced pluripotent stem cells
IRES	Internal ribosome entry site
IRS	Insulin receptor substrate
ITRs	Inverted terminal repeat sequences
KD	Knock down
KMT2C	Lysine methyltransferase 2C
KO	Knock out
LOF	Loss of function
MET	Mesenchymal to epithelial transition
METABRIC	Molecular Taxonomy of Breast Cancer
MMP	Metalloproteinases
MMTV	Mouse mammary tumour virus
MSCV	Murine stem cell virus long terminal repeat
NFIB	Nuclear factor IB (NFIB)
NS	Not significant
NGS	Next generation sequencing
OPI	Overproduction inhibition
pA	poly-adenylation
PB	*piggyBac*
PCDH9	Protocadherin 9

PCR	Polymerase chain reaction
PGK	Phosphoglycerate kinase
PI3K	Phosphoinositol 3-kinase
PR	Progesterone receptor
rtTA	Tetracycline transactivator protein
SA	Splice acceptors
SCLC	Small cell lung carcinoma
SB	*Sleeping beauty*
s.d.	Standard deviation
TCGA	The cancer genome atlas
TNBC	Triple negative breast cancer
TSS	Transcription starting site
VEGFA	Vascular endothelial growth factor A
WAP	Whey acidic protein
WHO	World Health Organization
WT	Wild type

10.2 List of figures and tables

Introduction

Results Part I

10.3 Curriculum Vitae

PERSONAL INFORMATION

Name	ZILLI
First name	Federica
Address	4 rue du General de Gaulle, 68770 Bartenheim (France)
Phone (lab)	+41 61 26 53319
Mobile	+33 77 20 47516
Email	federica.zilli@unibas.ch
Date of birth	March 9th, 1990
Nationality	Italian

EXPERIENCE AND EDUCATION

May 2015-Present **PhD student at Friedrich Miescher Institute (FMI), then Department of Biomedicine (DBM),** Tumour heterogeneity metastasis and resistance. Prof. Dr. Bentires-Alj laboratory, *Basel (Switzerland).*

- Research topic: Breast cancer translational research

2020-2021 **Continuing professional development (ECPM) in Project Management in The Life Science Industry** at the University of Basel, *Basel (Switzerland).*

Feb 2017-Nov 2017 **Antelope@Novartis Career Development Program,** career program at the University of Basel, Novartis Campus, *Basel (Switzerland).*

Oct 2014-Mar 2015 **Research Fellowship at IFOM, FIRC Institute of Molecular Oncology Foundation**, Ubiquitin and signal transduction. Group of Dr Simona Polo *Milan (Italy)*

- Research topic: HECT E3 ligases substrates in lung adenocarcinoma

Jan 2011-Oct 2011 **Master's degree intern at IFOM, FIRC Institute of Molecular Oncology Foundation**, *Milan (Italy)*

- Thesis: "Identification of substrates for the uncharacterized HECW1 E3 ligase through a novel 'molecular trap' technique".

2012-2014 **Master of Science in applied research in biomedicine, *University of Milano**, Faculty of Science and Technology, Milan (Italy)***

- Subjects studied: Cellular and Molecular Biology, Oncology, Physiology and Immunology

Mar 2012-Sep 2012 **Bachelor's degree intern at National Laboratory CIB at the Area Science Park**, Experimental oncology. Group of Prof Giannino del Sal, *Trieste (Italy)*

- Thesis: "Study of the cooperation between the overexpression of Twist and the silencing of BRD7 in the transformation of epithelial cells".

2009-2012 **Bachelor of Science in Medical Biotechnology, *University of Trieste,*** *Faculty of Medicine, Trieste (Italy)*

- Subjects studied: Cellular and Molecular Biology, Inorganic and Organic Chemistry, Biochemistry and Physics

AWARDS & SCHOLARSHIPS

2020/21 Selected for **Continuing professional development (ECPM) in Project Management in The Life Science Industry**

The program provides a profound background on understanding the tasks and responsibilities of a project manager with a focus on project documents required for conducting projects successfully, covers major aspects of managing effective project teams, teaches the interactions needed with different project stakeholders and finally thoroughly explains how to practically apply different project management standards to all kinds of pharmaceutical development projects. *University of Basel, Switzerland*

2017 Selected for the **Antelope@Novartis Career Development Program**

Antelope is aimed at highly qualified female doctoral students from all faculties. Over a 10-month period, the program supports early career researchers to systematically plan and promote their careers and prepare them for future leadership and management assignments. Antelope is a program at the University of Basel. *Basel (Switzerland).*

2016/14 **Scholarship for academic achievement.** FIAT & Chrysler, *Turin (Italy)*

2011-12 **Scholarship for academic achievement.** University of Trieste, *Trieste (Italy)*

2009 **Scholarship for High School Diploma achievement.** Cison di Valmarino,

Treviso, (Italy)

CERTIFICATES & COURSES

- **Certificate of Design and Interpretation of Clinical Trials 2020** authorized by *John Hopkins University* and offered for *Coursera*
- **Certificate of Good Clinical Practice 2020** NIDA Clinical Trials Network
- **International PhD Student Course Frontiers in Metastasis Biology 2018** *12-14th September 2018, University of Basel, Switzerland*
- **Animal experimenter degree (LTK1)**

  Course in accordance with Ordinance Governing Training and Ongoing Education of Specialists for Animal Experiments FELASA-B (LTK1), *University of Zürich, Switzerland*

SELECTED POSTERS

PiggyBac genetic screening reveals that the NFIB/ERO1A axis promotes metastatic colonization in breast cancer

Federica Zilli[#], Pedro Marques Ramos[#], Atul Sethi, Marie-May Coissieux, Priska Auf der Maur, Charly Jehanno, Laura Bonapace, Joana Pinto Couto, Tobias Eichlisberger, Kirsten D. Mertz, Zsuzsanna Varga, Roland Rad, Michael Rugaard Jensen, Michael B. Stadler and Mohamed Bentires-Alj.

FARO Oncoday Symposium 2020, Basel, Switzerland

NFIB promotes metastatic colonization in breast cancer via angiogenesis

Federica Zilli[#], Pedro Ramos[#], Priska Auf der Maur, Atul Sethi, Marie-May Coissieux, Charly Jehanno[1], Laura Bonapace, Joana Pinto Couto, Tobias Eichlisberger, Kirsten D. Mertz, Zsuzsanna Varga, Sophie Dessus-Babus, Hubertus Kohler, Francesco Hofmann, Allan Bradley, Roland Rad, Michael Rugaard Jensen, Michael B. Stadler and Mohamed Bentires-Alj

BBC Annual Meeting on personalized breast cancer treatments 2019, November 14-15th 2019, Basel, Switzerland

Transposon-mediated genome-wide mutagenesis screen identifies potential drivers of breast cancer

Federica Zilli, Pedro Ramos, Priska Auf der Maur, Tobias Eichlisberger, Michael Stadler, Roland Rad, Ryoko Okamoto, Michael Rugaard Jensen, Francesco Hofmann, Emilie Chapeau, Mohamed Bentires-Alj

PhD Retreat, June 7-9th 2018, Quarten-Walensee, Switzerland

Identification of mutant *PIK3CA* collaborative mutations in breast cancer tumorigenesis by transposon-mediated genome-wide mutagenesis

Federica Zilli, Pedro Ramos, Tobias Eichlisberger, Ryoko Okamoto, Mohamed Bentires-Alj

ENBDC, May 12-14th 2016, Weggis, Switzerland

PiggyBac transposon mutagenesis screen identifies potential drivers of breast tumorigenesis

Pedro Ramos, **Federica Zilli**, Tobias Eichlisberger, Joana Pinto Couto, Michael B. Stadler, Sophie Dessus-Babus, Tim Roloff, Hubertus Kohler, Francesco Hoffman, Michael Rugaard Jensen and Mohamed Bentires-Alj.

FMI Annual meeting, September 16-18th 2015, Grindewald, Switzerland

RELEVANT CONFERENCES ATTENDED

- FARO Oncoday Symposium 2020, Basel, Switzerland (**Selected poster presentation**, title: PiggyBac genetic screening reveals that the NFIB/ERO1A axis promotes metastatic colonization in breast cancer)
- ENBDC, 2020, Weggis, Switzerland (Poster presenter)
- BBC Annual Meeting 2019, November 14-15th 2019, Basel, Switzerland (Poster presenter)

- PhD Retreat, June 7-9th 2018, Quarten-Walensee, Switzerland (**Selected poster presentation**, title: Transposon-mediated genome-wide mutagenesis screen identifies potential drivers of breast cancer)
- Gordon Research Conference on Mammary Gland Biology, May 26-31 th 2018, Lucca, Italy (Poster presenter)
- ENBDC, March 15-17 th 2018, Weggis, Switzerland (Poster presenter)
- FMI Annual Meeting 2016, September 14-16t th 2016, Basel, Switzerland (Poster presenter)
- ISREC Symposium 2016, September 7-10 th 2016, Lausanne, Switzerland (Poster presenter
- BBC first annual meeting on: Personalized breast cancer treatments, November 17-18th 2016, Basel, Switzerland (Poster presenter)
- ENBDC, May 12-14 th 2016, Weggis, Switzerland (Poster presenter)
- FMI Annual Meeting 2015, September 16-18th 2015, Grindelwald, Switzerland (Poster presenter)

LANGUAGES

Excellent knowledge of both spoken and written English. I have intermediate knowledge in German and French. My mother tongue is Italian.