Role of Rpl39l in Translation, and consequences for Pluripotency and Cancer

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Arka Banerjee

Basel, 2024

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Erstbetreuerin: Prof. Dr. Mihaela Zavolan
Zweitbetreuer: Prof. Dr. Erik van Nimwegen
Externer Experte: Prof. Dr. Andre Gerber

Basel, 21. Juni 2022

Prof. Dr. Marcel Mayor
Dekan der Philosophisch-Naturwissenschaftliche Fakultät

Contents

ACKNOWLEDGEMENT

At the time of writing this thesis, I am still a bit anxious about the project, and yet, confident that people will support me through it. An eventual victory over Rpl39l (my mortal enemy at this point) would not be possible without them.

And that is why I want to thank the people associated to me during this project.

After having worked with 3 other PIs, Mihaela surprised me with the idea that PIs too, can be human, and treat a student humanely (sounds like a low bar to start with, but being a PI, this is a huge achievement, at least from my limited sample space of PIs and the examples in Biozentrum alone). She has supported me when I was out of my depth, challenged me when I was flying too high, was my sparring partner on outrageous ideas and above all, gave me the freedom to chart my own course. She gave me the space to make my own mistakes and guided in ways that suited my need. I find her an inspiring leader, thank her for our adventures together and look forward to some new ones.

And then comes Nitish, the guy with all the right ideas. I wish I had listened to him earlier than I did, I would have saved 8 months of the same western blots daily. His guidance, reluctant willingness to jump into questionable (and exciting, of course) project decisions and getting hands dirty, understanding of the field and the banter over daily chai really kept my spirits high. You will be the corresponding author when I finally publish the chai-in-a-microwave paper in Nature methods. I am yet to learn from him the skill of giving a reality check without the slightest hint of discouragement. He is solid PI material, but a bit too free of a soul to enjoy being a PI. Thanks for the fun times. I hope there are more coming.

Lea came into the game quite late, but our lab's personal botanist had quite an impact. Not only is she still dutifully taking care of my plants in the lab, but she also pushed quite a few of the

frustrating experiments forward. However, her biggest support is got to be taking over and scaling-up our chai operations. Without you, the chai sessions would be less fun.

Next up Jeremie, for actually showing me the value of learning maths and almost literally handholding me on the computational aspects on things, along with Mihaela, I have met few people more competent than you. You are made for great things in life.

Maciej and Meric got on-board the bandwagon quite late, but dear lord that energy and stamina you two brought to the project. Maciej, even though I have never eaten the MJS special and I never will, I am sure you are a as good a Chef de Cuisine as you are a scientist. And Meric, you are a high-throughput system on your own. Thank you both for your help.

Niels, what would I do without the German cultural integration that you offer? You make the lunches a seriously interesting affair.

And then there are Ayush, Bikash and Soham, co-victims of doing a PhD, so that sharing the pain gets really smooth. I am grateful for your company, the drinks, and the laughs. You guys are my trampolines, the lower the life dragged me, the higher you shot me up. Kudos to you!

My parents, thanks for trusting me to do my own thing, and resisting the urge, that has ruined the dreams of so many of the Indian youth. You are strong and you are soft. Nothing is more empowering to me, than the thought that you are there for me, no matter what. Mimi, I am doing good, and am happy. Stop worrying about me, you are the younger one. Good things are coming your way, just hold on.

Kirsten, my Samwise Gamgee, you raise me when I am fallen. I cannot possibly state in words, how much I value your presence in my life, but I will try. Life is not easy, and yet, somehow you make it fun. You have the exceptional ability to find light in darkness, or to be the light in the darkness. As much as this PhD has deconstructed my sense of self, you have built it. Thank you.

For the people in my life that I have not named here, your value for me, is more than this page can hold. I hope I do enough to make you feel that value. I have met many people through the course of this PhD, and many became good friends. I learned a great deal from you, and I hope we keep running into each other for eternity, it is a lot of fun.

And finally, the people who scooped me. As much as I dislike the fact that you scooped my work and really made my PhD difficult at a bad time, I thank you for pushing us to get our s**t together and aiming for something higher. Other than that, man that was a shoddy piece of work.

INTRODUCTION

Gene expression

Throughout the known life forms, gene expression is a multi-step process, in which the genetic information encoded in the DNA is converted, via the messenger RNA intermediary (transcription), into proteins (translation). While in bacteria these two processes take place simultaneously at a given gene locus, in eukaryotes, due to the more compartmentalized structure of cells, they are spatially separated, more complex and subject to additional controls (Kozak, 1999). Eukaryotic transcription requires the locus to be accessible to the transcription machinery, which is achieved via chromatin remodelling, and then cell type and context-dependent transcription factors recruit the RNA polymerase (Magaña-Acosta and Valadez-Graham, 2020). As the RNA is transcribed, the addition of 7-methylguanosine cap stabilizes its 5' end, and spliceosomes start removing the introns, while specific signals trigger the 3' end processing complex assembly on the transcript to define the 3' end. This is done by endonucleolytic cleavage followed by the addition of a polyadenosine tail. The completion of splicing can then take place post-transcriptionally, resulting in the mature mRNA, ready to be exported from the nucleus to the cytoplasm (Zhao et al., 2017). All of these transcription and maturation steps are carried out by proteins or nucleoprotein complexes that can be modified in various ways to modulate their activity. Moreover, as these biochemical reactions involve small numbers of molecules, there is substantial variability in the number of transcripts per cell, even when the cells are of the same type (Janes, 2015). Furthermore, work in the past two decades has revealed that gene expression occurs in bursts, not with a constant average rate, and these bursts appear to be largely due to the need for chromatin remodelling at the transcribed locus. Once the locus is "on", a burst of transcripts is produced before the remodelling enzymes regenerate the "off" state of the chromatin (Tunnacliffe and Chubb, 2020).

Accompanied by various RNA-binding proteins (RBPs), the mRNAs are exported to the cytoplasm, sometimes to specific compartments. The RBPs help chaperone the mRNAs to these compartments. The past decade has seen an explosion of work on what is known as "liquid-liquid phase separation" of RNAs and proteins. The term describes the observation that molecules self-organize in the cell so that the cell and its membrane-separated compartments are not homogenous mixtures of various component molecules, but rather inhomogeneous mixtures, generally likened to the droplets of oil and vinegar that intersperse with each other when one mixes these two liquids on a plate (Paci and Lemke, 2017). One of the emerging paradigms is that RNAs serve as scaffolds and organizers of phase-separated droplets. In fact, various granules that have been observed in cells over the years, such as the processing (P)-bodies and stress granules in the cytoplasm, or the nucleolus in the nucleus, are now considered RNA granules (Luo et al., 2018).

In the cytoplasm, mRNAs have the opportunity to encounter ribosomes, the molecular machines that translate the mRNAs into proteins. Ribosome biogenesis is also a highly complex, multi-step process, which involves transcription and processing of the 4 ribosomal RNAs, and assembly of the two ribosome subunits, small and large, with the help of some 80 ribosomal proteins, which also need to be transcribed, translated in the cytoplasm, and imported in the nucleus (Leary and Huang, 2001). The site of ribosome subunit assembly is in fact the nucleolus. The two subunits are then exported as independent components to the cytoplasm. Upon encountering an mRNA in the cytoplasm, the small ribosome subunit, primed with additional factors that include the methionine-initiator transfer RNA, binds to it, and then the scanning process starts in the 5' to 3' direction, until the start codon, typically AUG, is

encountered. This triggers the recruitment of the large ribosome subunit and initiation of translation (Jackson et al., 2010) (Fig 1.).

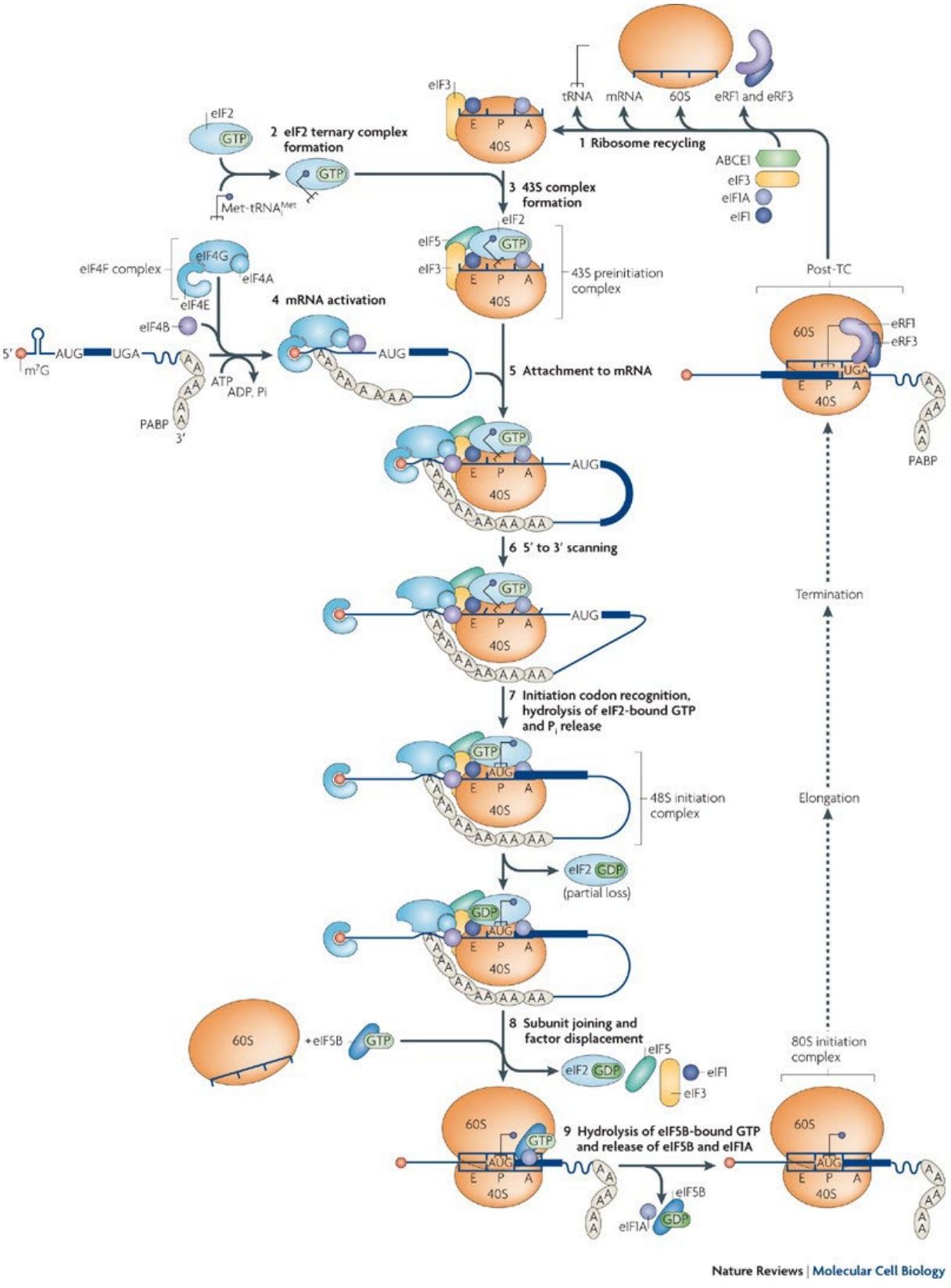


*Figure 1 (from (Jackson et al., 2010)). The canonical pathway of eukaryotic translation initiation is divided into eight stages (**2**–**9**). These stages follow the recycling of post-termination complexes (post-TCs; **1**) to yield separated 40S and 60S ribosomal subunits, and result in the formation of an 80S ribosomal initiation complex, in which Met-tRNA$^{Met}_i$ is base-paired with the initiation codon in the ribosomal P-site and which is competent to start the translation elongation stage. These stages are: eukaryotic initiation factor 2 (eIF2)–GTP–Met-tRNA$^{Met}_i$ ternary complex formation (**2**); formation of a 43S preinitiation complex comprising a 40S subunit, eIF1, eIF1A, eIF3, eIF2–GTP–Met-tRNA$^{Met}_i$ and probably eIF5 (**3**); mRNA activation, during which the mRNA cap-proximal region is unwound in an ATP-dependent manner by eIF4F with eIF4B (**4**); attachment of the 43S complex to this mRNA region (**5**); scanning of the 5′ UTR in a 5′ to 3′ direction by 43S complexes (**6**); recognition of the initiation codon and 48S initiation complex formation, which switches the scanning complex to a 'closed' conformation and leads to displacement of eIF1 to allow eIF5-mediated hydrolysis of eIF2-bound GTP and P_i release (7); joining of 60S subunits to 48S complexes and concomitant displacement of eIF2–GDP and other factors (eIF1, eIF3, eIF4B, eIF4F and eIF5) mediated by eIF5B (**8**); and GTP hydrolysis by eIF5B and release of eIF1A and GDP-bound eIF5B from assembled elongation-competent 80S ribosomes (**9**). Translation is a cyclical process, in which termination follows elongation and leads to recycling (**1**), which generates separated ribosomal subunits. The model omits potential 'closed loop' interactions involving poly(A)-binding protein (PABP), eukaryotic release factor 3 (eRF3) and eIF4F during recycling, and the recycling of eIF2–GDP by eIF2B. Whether eRF3 is still present on ribosomes at the recycling stage is unknown.*

It is especially the initiation step of translation that is heavily regulated, primarily at the level of cap-binding by the eIF4 complex. For example, the number of ribosomes in a cell is dependent of the cell's proliferation rate, as the building blocks for daughter cells need to be synthesized (Rudra and Warner, 2004). For translation to initiate on a given mRNA, the cap structure has to be recognized by the eIF4E translation initiation factor. eIF4E is normally kept in an inactive state by eIF4E-binding proteins (eIF4-BP). Phosphorylation of these proteins by the TORC1 complex (where TOR stands for "target of rapamycin") releases eIF4E, which can interact with the translation-stimulatory eIF4G (Fonseca et al., 2014). Viruses, but also some host transcripts can initiate translation independent of the cap (cap-independent translation). Rather, such transcripts have "internal ribosome entry" sites, which correspond to structural elements in the RNA, where the small ribosome subunit is recruited via interactions with specific RBPs (Komar and Hatzoglou, 2011).

While under proliferative conditions translation is upregulated via TORC1, under stress conditions, translation is inhibited. A main mechanism underlying this inhibition converges on the initiation factor eIF2, which binds GTP and the Met-tRNA$_i$ and transfers the Met-tRNA$_i$ to the small ribosome subunit (Wek, 2018, p. 2). When the protein synthesis at the endoplasmic reticulum (ER) is impaired, the cell activates a mechanism that reduces the further synthesis of

proteins and the accumulation of damage. This mechanism converges on the subunit of eIF2, which is phosphorylated by the protein kinase RNA-like ER kinase (PERK), which is activated upon ER stress (Fig 2.) (Pavitt and Ron, 2012).

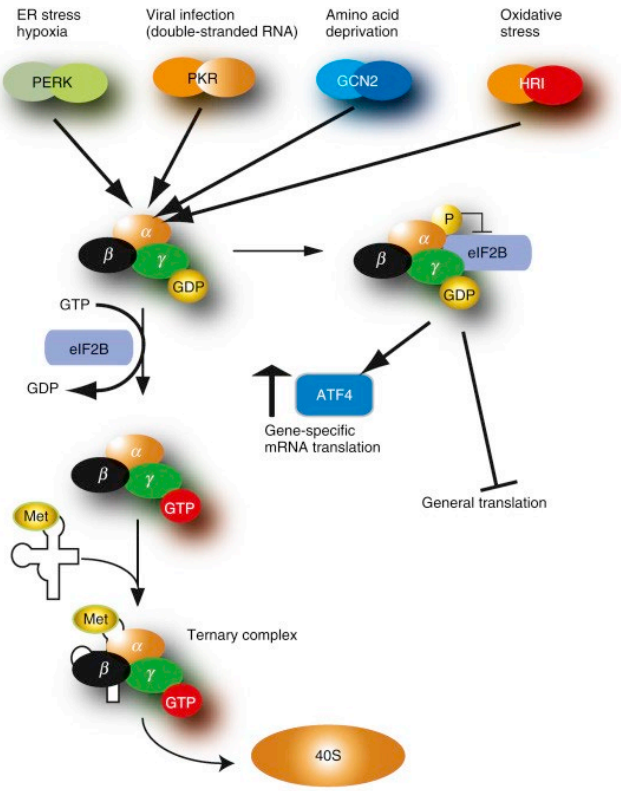


Figure 2 (from (Costa-Mattioli and Klann, 2017)). Translational control by the eIF2α kinases. Different stress conditions activate the eIF2α kinases, PERK [RNA-activated protein kinase (PKR)–endoplasmic reticulum–related kinase], PKR, HRI (hemin-regulated inhibitor kinase), and GCN2 (general control nonderepressible kinase), resulting in phosphorylation of α subunit of eIF2. Phosphorylation of eIF2α represses global rate of translation, by sequestering eIF2B, but stimulates translation of ATF4 mRNA. Studies from Costa-Mattioli et al. (2005) indicate that general control nonderepressible kinase (GCN2) and eIF2α phosphorylation are critical for long-lasting synaptic plasticity and memory.

The ribosome

The bulk of the work described in this thesis aimed to decipher the function of one component of the ribosome. The ribosome is an ancient machine for translating mRNAs into proteins, one of the most consequential steps in life. A ribosome is primarily constituted of multiple rRNAs and several proteins that depend on the kingdom and phylum of the organism that hosts the ribosome. The core functions of the ribosome are: 1) Translation initiation 2) Positioning on the proper codon, 3) Accepting the correct tRNA, 4) Catalysing the peptide bond formation 5) Interaction with release factors leading to the release of the peptide, mRNA and disassembly of ribosome into its subunits. Steps 2-4 are repeated multiple times during the translation of an mRNA, and primarily mediated by the rRNA core (Lafontaine and Tollervey, 2001). This role of rRNA in protein biogenesis is conserved throughout the known living forms. This gives credence to the hypothesis that ribosomes were formed in the RNA world in the very early stages of evolution of life (Bowman et al., 2015). As mentioned already, ribosomes have 2 major components, a small subunit, that detects and scans the mRNA for a start codon, and a large subunit, that joins the small subunit to start the process of translating the mRNA (Yusupova and Yusupov, 2017). The formation of the peptide bond happens at the PTC (peptidyl transferase centre). Studies have shown that the PTC structure is highly conserved between prokaryotes and eukaryotes (Prosdocimi et al., 2020), consisting in 3 sites: A (acceptor), P (peptidyl transfer) and E (exit) sites. Briefly, the mRNA enters the ribosome at the mRNA entry site, then through the 43/48S Pre-initiation complex (PIC) the PTC identifies the start codon, the initiation-codon-specific methionyl-tRNA binds in the A-site of the ribosome, the mRNA-tRNA-amino acid complex moves to the P-site where the peptide bond with the previous amino acid is formed, the finally, the empty tRNA shifts to the E-site and is released from the ribosome. The nascent peptide starts moving into the Nascent Peptide Exit Tunnel (NPET). The process repeats until the ribosome reaches one of the 3 stop codons, where

the tRNA for the respective stop codon binds the A-site, but there is no amino acid to form the peptide complex, hence the nascent peptide is released (Lafontaine and Tollervey, 2001; Yusupova and Yusupov, 2017). While this mechanism of translation in the ribosome is conserved across all living forms, what has changed over the course of evolution is the control, stability, and fine-tuning of the translation process by addition of more protein and rRNA components (Bowman et al., 2020) (Fig 3.a). As the catalytic activity of the ribosome resides in the rRNA, the ribosome is a ribozyme (Steitz and Moore, 2003), in which ribosomal proteins have largely structural functions. It has been observed that as organisms became more complex, the complexity of the ribosomal content and structure went up, along with the size (Lukeš et al., 2011) (Fig 3.b). For example, the fully functional ribosomal complex in Prokaryotes is 70S while in Eukaryotes, it is 80S (Melnikov et al., 2012).

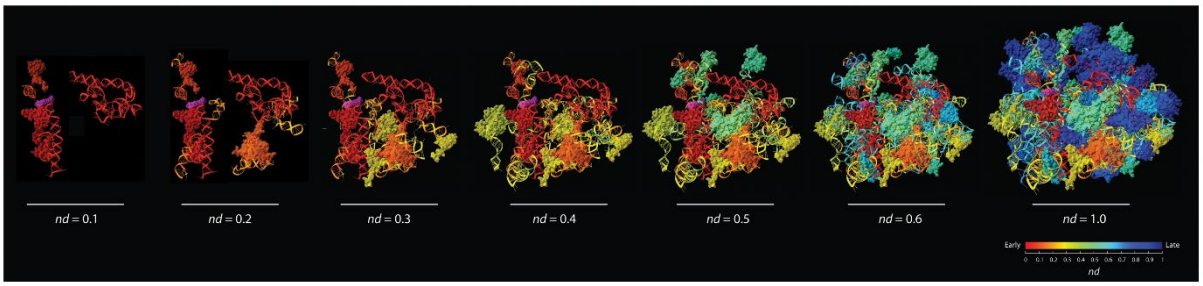


Figure 3.a) Proposed model of ribosome evolution (from (Harish and Caetano-Anollés, 2012)). In early forms of ribosomes rRNA helices interacted with ribosomal proteins to form the peptidyl transferase centre, the core of ribosomal function by mediating nucleotide interactions. In the later stages of ribosomal evolution, this PTC serves to coordinate and balance ribonucleoprotein accretion, eventually building the modern ribosome. Here, rRNA is shown in ribbon representation and the proteins are as space-filling models.

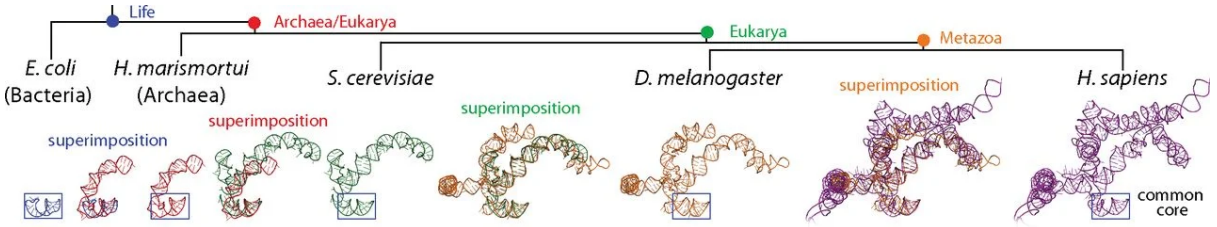


Figure 3.b) The evolution of helix 25/ES 7 (from (Petrov et al., 2014))shows continuous accretion of rRNA onto a constant core. This figure shows on an atomic level how helix 25 of the large ribosomal subunit rRNA started form a small stem loop in the constant core into a major rRNA domain in the metazoans. Every accretion step adds on top of the last rRNA core but the core remains unchanged. Each structure is a direct representation of X-ray diffraction or CryoEM.

In 1957 the central dogma had already hypothesized the presence of a machinery that reads the mRNA to make proteins, while the first ribosomes were detected in 1955 (BARRNETT and PALADE, 1958). However, it wasn't until late 1960s that the evidence of the translating role of ribosomes started piling up. Emil Palade shared his Nobel prize with Albert Claude and Christian de Duve for detecting the ribosome in 1974. Up until the 70s, the proteins involved in formation of ribosomes' core structure (Ribosomal Proteins, RPs), 79 of them in mammals, 68 in archaea and 56 in prokaryotes, were considered to have only one function, to sit in a ribosome, and the only function of the ribosome was to translate mRNA (Melnikov et al., 2012).

Ribosome variants and regulation at the level of the ribosomes

What followed were major publications showing not only the extra ribosomal role of ribosomal proteins (Zhang et al., 2017), but also the control and signalling roles of the ribosomes themselves (Karamyshev and Karamysheva, 2018; Rutkowski et al., 2001). Development of genomic and proteomics analysis techniques gave rise to the understanding that the RPs are encoded in multiple gene copies in the genome (Kenmochi et al., 1998; Kuzumaki et al., 1987). Most of these copies formed by retrotransposition of an original gene, which can been seen from the lack of introns in their coding regions (Wong et al., 2014). In fact, RP pseudogenes have a total of 2000 copies in the genome, largest of any other class of proteins (Zhang et al., 2002) (Fig. 4). While initially it was assumed that pseudogenes can function as backup in case the major RP gene was mutated, it was soon noticed that most of these genes never express,

even when the expression of the "original gene" is abolished (Balasubramanian et al., 2009). In contrast, quite a few that are expressed, do so in parallel with the major gene copy (Guimaraes and Zavolan, 2016). While strongly conserved, some interesting mutations were observed between the main RPs and their parallelly expressed paralog partners. In some cases, the expression of such paralog partners seems to extremely context specific. This gave rise to the interest in the role of the so called 'Ribosomal protein paralogs' (Gerst, 2018; Guimaraes and Zavolan, 2016). It was found that while many of these proteins did have a redundant function with respect to their paralog partner (Jiang et al., 2017), most of the paralog partners were indeed, functionally independent. This thesis focusses on one such RP-paralog pair, RPL39 and RPL39L.

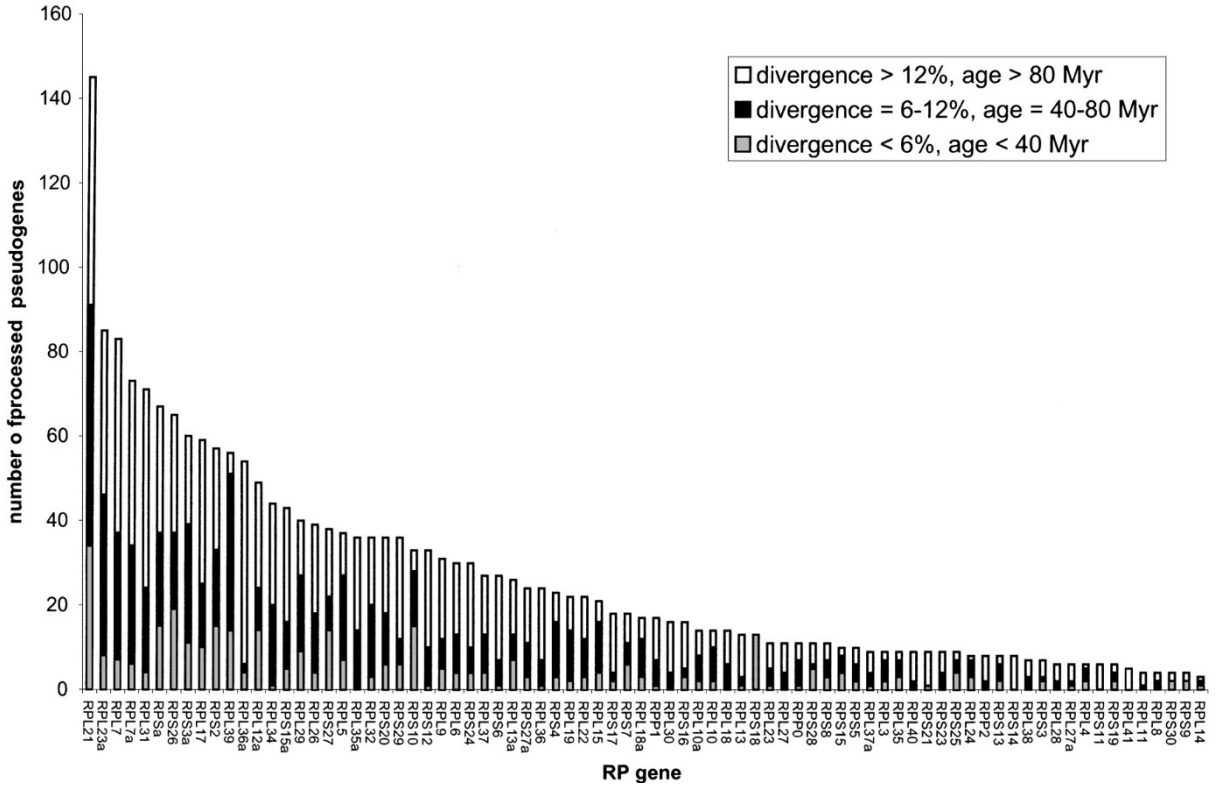


Figure 4 (from (Zhang et al., 2002)). Mammalian genomes have up to 2000 copies of ribosomal proteins as a processed pseudogenes (retrotransposed genes). This figure represents a distribution of the number of gene copies per ribosomal gene along with their estimated age based on the gene sequence divergence from the original ribosomal gene. As can be seen here, there has been consistent retrotransposition of ribosomal genes over the period of last 80 million years. Rpl39 has around 60 copies in various mammalian genomes. Rpl39l is the only copy that expresses both on the mRNA level and protein level.

present in limiting concentrations. In contrast, mRNAs of growth factors and proto-oncogenes

are ‘weak’ competitors (Lodish H. F., 1974). Furthermore, ribosomes purified from bacteria that were grown in different conditions showed (Van Duin J. et al., 1970) variable composition, while still maintaining translation competence. The concept of a ribosomes as having some degree of selectivity for the mRNAs that they translate, selectivity which is modulated through the composition and post-translational modification of ribosomes, has not been part of the mainstream work on translation for some decades, but recently re-emerged. A main contributing finding was that ribosomes containing the Rpl38 RP are specifically needed to translate Hox transcripts during mouse development, while global translation can go on unperturbed when this protein is missing (Kondrashov et al., 2011). While this work has rekindled the interest in the concept of ribosome heterogeneity, other groups argue that most of the mRNA-specific perturbations in translation that are observed upon perturbed expression of individual RPs are due to perturbations in overall ribosome numbers, which disproportionately affect different mRNAs in the cell (Mills et al., 2017).

Rpl39/Rpl39l

Rpl39 is a 51amino acid RP that is specific to Eukaryotes and sits on the X-chromosome. It is an Arginine and Lysine-rich RP that lines the NPET of the ribosome. The amino acid sequence of Rpl39 is 100% conserved within Mammalia, and the NPET-lining motifs of Rpl39 are highly conserved throughout the Eukaryotes (Fig 5.) (Tsui et al., 1996). Structural studies have found Rpl39 to be one of the first RPs to sit into the newly forming NPET during ribosome biogenesis in the nucleus (Wilson et al., 2020). While the exact role of Rpl39 in the ribosome is yet to be determined, it has been shown that the knock-out of Rpl39 is non-lethal in yeast but makes the yeast cells sensitive to translation error-inducing drugs like Paromomycin. This indicates that Rpl39 has a role in translation speed and fidelity (Peisker et al., 2008). Rpl39 has been shown interact with Hsp70 (Gumiero et al., 2016) and with the protruding arm of n-beta Nascent

polypeptide associated complex (n-βNAC) through crosslinking (Gamerdinger et al., 2019), but the exact significance of these interactions is yet to be elucidated. With crosslinking experiments Rpl39 has also been implicated in the folding of transmembrane alpha helices of the transmembrane proteins in the NPET (Zhang et al., 2013). This folding of the transmembrane alpha helix in the NPET is necessary for proper translocation of the transmembrane protein to the endoplasmic reticulum (ER). Molecular dynamics simulations have shown that the Arginines of Rpl39 that line the NPET are necessary for the formation of a dynamically charged surface of the NPET (Petrone et al., 2008). In fact, the R28 of Rpl39 falls on the second constriction site of the NPET and can be hypothesized to slow down the movement of nascent peptide in the tunnel, helping with the folding of the nascent peptide. The RAC-dependent interaction of Rpl39 with the Hsp70 chaperone is also hypothesized to underlie the role of Rpl39 in translation fidelity.

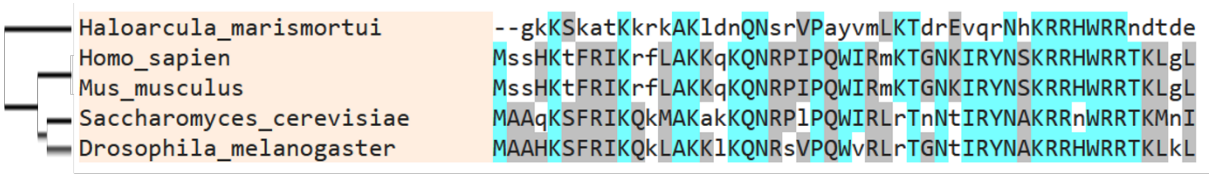


Figure 5. Alignment of Rpl39 proteins from Mammalia, Insecta, Fungi and Archaea species, showing the incredible conservation of the Rpl39 sequence over different phylums and kingdoms of life. Haloarcula is one of the earliest organisms to have evolved Rpl39 out of Rpl23. The conservation of the sequence focuses primarily on the Arginines and the Lysines. Most conserved positions are highlighted in cyan, partially conserved amino acids are coloured in gray.

Rpl39l, on the other hand, is also a 51 amino acid protein that emerged by the retrotransposition of Rpl39 transcript some 80 million years ago, making it one of the newest RP to evolve (Guimaraes and Zavolan, 2016; Wong et al., 2014). Rpl39l was found in apes and rodents but not in ungulates, though a thorough evolutionary analysis has not been done so far. While Rpl39 is located on X-chromosome, Rpl39l is on the 14th chromosome in mouse and the 3rd chromosome in humans. Mouse Rpl39l differs by 3 amino acid from Rpl39 and the human

Rpl39l by 4 amino acids. The high amino acid sequence similarity, combined with the interspersed distribution of the mutated amino acid throughout the sequence, renders it extremely challenging, if not impossible, to make an antibody that can distinguish Rpl39 and Rpl39l (data not shown). Of all the amino acid differences between Rpl39 and Rpl39l only R28Q is conserved between human and mouse. This makes the Arginine (R) of Rpl39 and Glutamine (Q) of special interest to this study (Fig 6.). In fact, the interaction between n-βNAC and Rpl39, as well as the constriction of the NPET has been shown to be through R28 (Gamerdinger et al., 2019). Assuming that the conservation of an amino acid is indicative of its vital function, we hypothesized that Rpl39l impacts a ribosome's interaction with NAC and the transmembrane alpha helix of transmembrane proteins. Hence, the study of Rpl39l could shed new light on the functioning of NPET and its role in control of translation rate and fidelity.

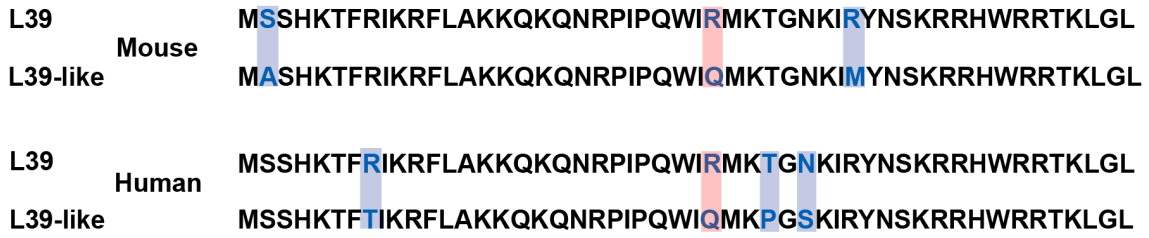


Figure 6. Sequence conservation between Rpl39 and Rpl39l in mouse and human. Despite multiple difference between Rpl39 and Rpl39l in both mouse and human (highlighted in blue), the only mutation conserved between the two is R28Q (highlighted in red).

As mentioned previously, RP paralogs tend to have context specific expression and interactions. Rpl39l is not an exception. mRNA studies have shown the Rpl39l-transcript to be expressed in pluripotent cells, cancer cells (Uechi et al., 2002), and, most strongly, in the germ cells lineage. While the mRNA expression of Rpl39l was known in different cell types, no study, hitherto has been able to detect the protein in all these systems. The challenge came from the previously described trouble with generating specific antibodies and the fact that the trypsinization step in

a normal proteomic MS-MS study generates extremely small peptides that don't support Rpl39l detection (Fig 7.). These troubles in detection of this protein are compounded by its low expression level with respect to its paralog partner Rpl39 in most cell types. Although we tried both antibodies available in the market and developed in-house, we failed to detect the protein by these means. One publication has detected the protein using a 2D gel followed by MALDI-TOF analysis, but this is a method difficult to execute and replicate. The publication also used serum of a mouse immunised against Rpl39l to detect the protein, but the authors turned down our request to share the serum (Sugihara et al., 2010). This led us to developing our own proteomic method for detecting Rpl39l with very high sensitivity and precision.

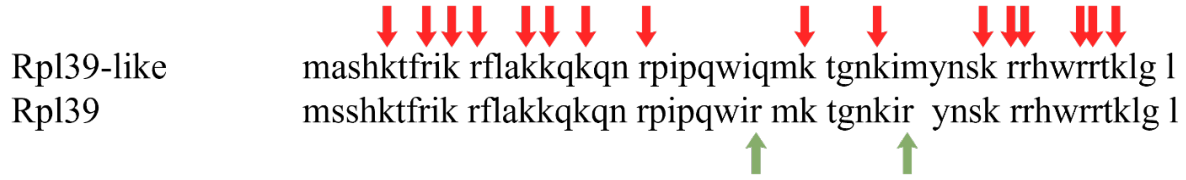


Figure 7. Here we show the number of Trypsin digestion target-sites on Rpl39l in mouse (Depicted in red arrows). Trypsin cuts at Lysine (K) and Arginine (R). Rpl39 has 2 more targets for Trypsin digestion than Rpl39l (Depicted in green arrows). This is the primary reason why Rpl39l protein could not be dected by proteomics over the years.

Main Project: Characterization of Rpl39l function in embryonic stem cells, cell differentiation and cancer

I will describe here my main project, which aimed to determine the function of Rpl39l in all the cell types in which its mRNA has been observed, namely embryonic stem cells, spermatocytes, and cancer cells. I will present the material as I have obtained it, following specific hypotheses, validating, or rejecting them. As the material is quite extensive, we are preparing separate manuscripts, in particular on the ESC and cancer cell data.

Once the protein detection method was developed, we started surveying various cells and tissue samples, primarily the ones reported to express the Rpl39l mRNA. Not too surprisingly, we found that, these cells do express the RPl39l protein, though the main question that we were

able to answer is what the relative proportion of Rpl39l and Rpl39. We were able to detect Rpl39l in embryonic stem cells from both mouse and human, mesenchymal stem cells from human bone marrow, human and mouse sperm cells, multiple human cancer cell lines and human breast-cancer tissue samples. To check if Rpl39l is an actual RP, and not just a protein that shares sequence similarity with another RP and doesn't actually sit in the ribosome, we purified the ribosomes from E14 mouse embryonic stem cell lines under high salt concentration (300mM KCl), such that all the peripheral proteins associated with ribosomes fall off. We still detected a strong Rpl39l signal proving that Rpl39l is in fact an RP.

The strong expression of Rpl39l in sperm cells makes these cells strong candidates for studying the role of this protein. The location of Rpl39 and Rpl39l in the genome further hinted at the role that Rpl39l potentially plays in spermatogenesis. The two hypotheses that we started from then were: 1) since Rpl39 is located on the X-chromosome, contexts like spermatogenesis, where one of the haploid cells does not have access to the X-chromosome, could activate Rpl39l expression to compensate the lack of Rpl39 or 2) Rpl39 production is halted during meiosis by Meiotic sex chromosome inactivation (MSCI) (Cloutier and Turner, 2010), and a autosomal paralog of Rpl39 is expressed to compensate the lack of Rpl39, a process very similar to Rpl10 and Rpl10L regulation in spermatogenesis (Jiang et al., 2017, p. 10) as shown by Jiang et al. 2017. In fact, the work of Zou et al. 2021 showed that KO of Rpl39l in the mouse model led to spermatogenesis failure in the male mice (Zou et al., 2021). While these two hypotheses made sense in terms of the role of Rpl39l in sperm cells, they were neither sufficient to explain the role of Rpl39l in stem cells and in cancer, nor the anticorrelation of Rpl39l transcript expression and relapse free survival (RFS) of breast cancer patient, as shown by Guimaraes and Zavolan, 2016. One trivial explanation for the expression of Rpl39l transcript in cancer was that the genomic region where Rpl39l sits in humans, ~480bp downstream of Bcl6, is prone to breakage and reintegration in cancer (Lossos et al., 2003, p. 6), and potentially, Rpl39l has an elevated

copy number in the genome, leading to spurious expression. This would make Rpl39l expression a downstream effect of genomic instability associated to lower Relapse Free Survival (RFS) cancer. A recent publication showed that Rpl39l over-expression was in fact associated to spurious demethylation of the specific genomic region in certain cancers (Yan et al., 2019). These were also major considerations during the early days of this study, prompting us to focus primarily on embryonic stem cells (ESCs), in parallel to sperm cells. A simple explanation for Rpl39l expression in ESCs would have been that the Rpl39l protein is waiting for the spermatogenesis to start, and then it would show its activity. To get additional insights into this idea, we looked at the protein level expression of Rpl10l, an Rpl10 paralog that has been shown to compensate Rpl10 expression inhibition induced by MSCI during spermatogenesis, in ESCs (Rpl10 is located on X-chromosome, but Rpl10l is not) (Jiang et al., 2017). We found no Rpl10l expression in ES cells, indicating that the "waiting for spermatogenesis" might not be the case after all, and that the early stages of spermatogenesis can sufficiently produce the ribosomal paralog proteins required for this process. This indicated that Rpl39l has a very specific role in translation in pluripotent cells. This idea was further strengthened by the fact that Rpl39l expression was also detected in hMSCs derived from human bone marrow, that go on to become bone and cartilage cells (Hanna et al., 2018; Somoza et al., 2014), which have no known expression of Rpl39l transcript.

As hinted above, embryonic development is the prime scene where RPs and their especially their paralogs have been shown to be necessary, in a ribosome-dependent manner. Rpl38 is necessary for the IRES-dependent translation of HOX transcripts, and the KO of Rpl38, while keeping translation intact for most transcripts, blocks Hox transcripts' translation, leading to KO mice having major skeletal patterning defect, leading to unviable mice (Kondrashov et al., 2011). Rpl22 and Rpl22l have been shown to not only regulate the splicing of pre-mRNA to regulate morphogenesis, but also to be regulating each other's transcripts by directly interacting

with the transcript and targeting the transcript for degradation, as a simple mechanism to maintain the right protein balance in the cells (Zhang et al., 2017, p. 22). Rpl3l is integral to the development of striate muscle, while Rpl3 is necessary for myotube formation and growth (Chaillou et al., 2016). Rps27 and Rps27l have both been shown to target p53 through their interaction with E3 ubiquitin ligase Mdm2 in a much wider context than just ES cells (Xiong et al., 2011). Similarly, Rpl13 has been implicated in IRES dependent recruitment of Foot-and-Mouth Disease Virus (FMDV) transcripts and hence FMDV infection (Han et al., 2020).

The major challenge to teasing apart the molecular role of Rpl39l is that it sits deep inside the ribosome (Fig 8.). Even if one of the antibodies that we have generated would have worked, we would have not gotten to the Rpl39l in the ribosome, but only the Rpl39l waiting to be loaded in the ribosome. While it is tempting to tag the N or C terminus of this protein by a small tag like V5 or His, as has been done before for Rpl39, the tag would have to pass through the NPET to become available for an antibody to bind. This would have to led to disruption of translation in the Rpl39l specific ribosomes, probably skewing our observations. Similar issues also kept us from trying approaches with proximity-tagging systems like BioID, given how huge these tags are compared to our 51 AA protein.

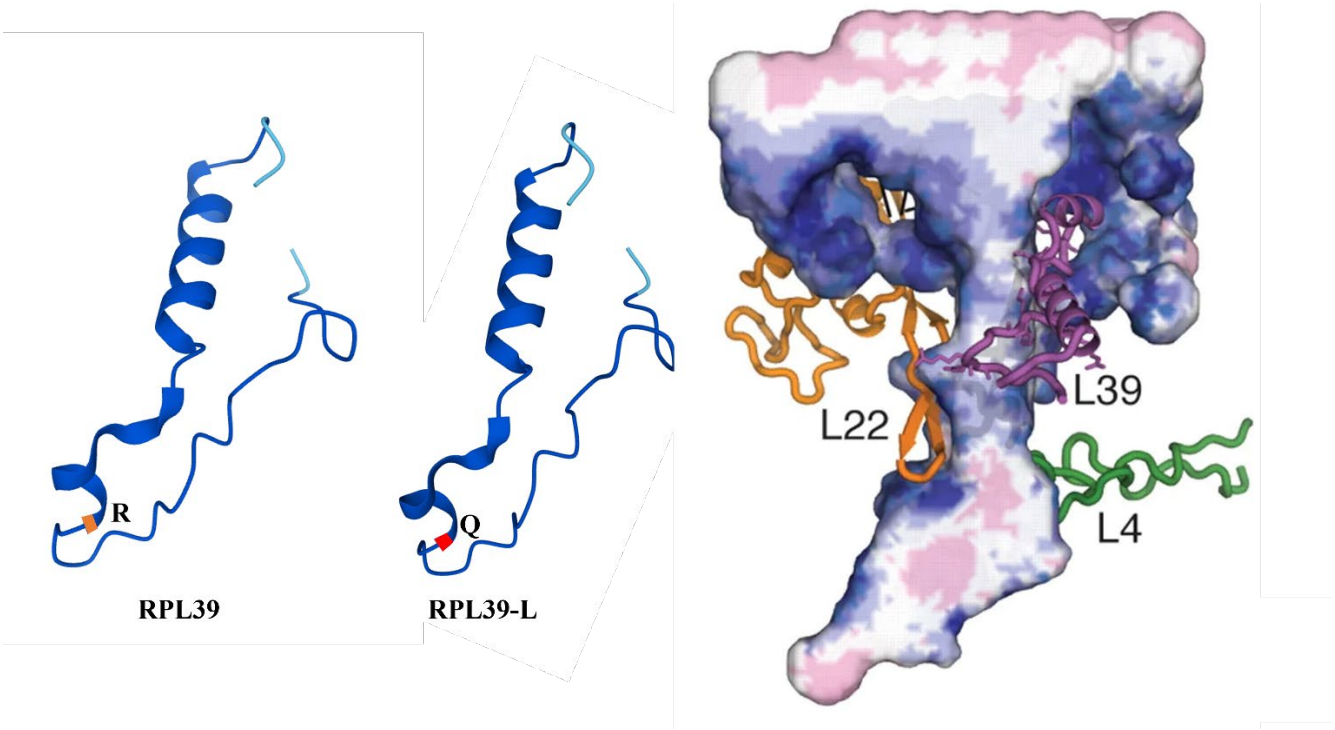


Figure 8. This figure shows (from left to right) the structure of Rpl39 as determined by X-ray crystallography and Rpl39l structure as predicted by Alphafold. The image on the right is a depiction of NPET with the PTC at the bottom and the vestibule at the top from Petrone et al.2008. Rpl39 has been shown to be sitting in the vestibular region, surrounded by other protiens and rRNA, projecting structure and amino acid side chains deep into the NPET.

We decided that a KO system of this protein in an ESC line would be the best line of approach, given that it can be used not only to recapitulate early embryonic states of the organism (Guan et al., 1999), but can also be differentiated in sperm cells *in vitro* (Kerkis et al., 2007). The idea was to use CRISPR-Cas9 system to disrupt the gene expression in E14 mouse embryonic stem cell (mESC) line.

E14 is a fast-growing mESC with a doubling time of 13-15 hours. It is a classical model for studying differentiation and pluripotency in embryonic stem cells. E14, like many other embryonic stem cell lines, is derived from the inner cell mass (ICM) of a pre-implantation blastocyst (Thompson et al., 1989). The cells show remarkable stability of pluripotency in feeder-free culture conditions and with media supplemented with just Leukemia Inhibitory Factor (LIF). The pluripotency in these cells is measured by looking at the expression of Sox2,

Nanog and Oct4 (Xu et al., 2011). A disruption of the expression levels of these factors leads to differentiation and acts as a marker of differentiation. Loss of pluripotency, when too strong, can also be detected by the change in the morphology of the classical dome shaped colony to signature flattened cells with extensions (Morgani et al., 2017). Loss of pluripotency is also associated to a reduced rate of cellular replication and increased global translation (Tahmasebi et al., 2019). While these parameters are rather straightforward to measure, and have simple outcomes, the impact of disrupted pluripotency on differentiation is more challenging to analyse. One of the primary ways of analysing the impact of a certain manipulation in mESCs is to look at the spontaneous differentiation of the cells in the absence of any pluripotency maintaining signal, in our case, LIF. Such studies allow us to look at the prevalence, strength, and fidelity of differentiation along all the three germ lineages: Ectoderm, Mesoderm and Endoderm (Ngondo et al., 2018). A change in the above-mentioned marker expression is a disruption of differentiation capacity. Differentiation on the spermatogenic lineage is more targeted in nature, and requires retinoic acid signalling, along with the absence of pluripotency maintenance factors in the medium, in a defined serum-free medium (Kerkis et al., 2007).

Pluripotency and translation have a complex interplay. Studies have shown that stem cells have a generally lower global translation compared to differentiated cells (Tahmasebi et al., 2019). In fact, the increase in global translation associated to differentiation is also reflected in the increased rate in overall translation efficiency (TE), i.e., number of 80S ribosomes loaded on an mRNA (Sampath et al., 2008). While this phenomenon was first observed in mESCs, later studies have also found it in haematopoietic stem cells, quiescent epidermal cells, and hair follicle stem cells and in *Drosophila melanogaster* germline stem cells (Saba et al., 2021). So, a low level of global protein production is a feature of pluripotency. Despite the low global protein production, ribosome biogenesis is generally upregulated in pluripotent cells, including mESCs (Zaidi et al., 2016). In eukaryotes, ribosome biogenesis can be summarised in 4 steps

(Saba et al., 2021): 1) pre-rRNA production in the nucleolus, 2) import into the nucleus of ribosomal proteins and ribosome assembly associated proteins, 3) pre-40S and pre-60S ribosomal subunits getting exported to the cytosol, 4) processing of the ribosomal proteins in the cytoplasm and eventual assembly into functional 80S translating ribosomes. Several proteins involved in ribosome biogenesis are upregulated in pluripotent compared to differentiated cells, like fibrillarin, an rRNA 2′-O-methyltransferase involved in rRNA processing in the nucleus, which is highly expressed in mESCs and has been found to be necessary for pluripotency maintenance (Watanabe-Susaki et al., 2014). Another interesting example would be Myc, one of the 4 Yamanaka Factors (OSKM, M for Myc), which is involved in the transcription of rRNA and many ribosomal proteins, and is protected against proteasomal degradation in mESCs (Chappell and Dalton, 2013). In fact, mouse haematopoietic stem cells transcribe rRNA at a rate 2-4 times higher than their lineage restricted progeny. It has been shown that when ESCs differentiate to embryonic bodies (EBs) or to neural progenitors, the rRNA synthesis is downregulated (Blair et al., 2017).

It is an obvious dichotomy that the stem cells do not produce a lot of protein but produce a lot of protein producing machinery. One idea in this direction is that low protein synthesis reduces the damage to the cells happening from huge proteogenic activity and maintains self-renewal and longevity (Signer et al., 2014). While this can be one of the explanations, the consensus is that it is not sufficient since translation inhibitors like cycloheximide do not increase the longevity of the mESCs a lot (Bulut-Karslioglu et al., 2016). Another idea in this regard is that the pluripotent cells do not really need a huge proteome given their lack of specialized roles. Another possibility is that the low protein content of the cells allows for a faster switch of the proteome from pluripotent state to the specialised cell state that the cells is differentiating to. This idea is supported by the short half-life of many pluripotency factors and the requirement of high proteasome activity to maintain the ESC function and viability (Buckley et al., 2012).

Nevertheless, the apparent mismatch between ribosome production and protein output in pluripotent cells still needs an explanation.

Similar to the low proteogenic levels, high ribosome biogenesis can also be a pluripotency signalling biproduct. For example, given the numerous targets of the Myc transcription factor, activation of Myc to promote pluripotency via some of its targets could lead to overexpression of ribosome-associated genes (van Riggelen et al., 2010). But given that the ribosome content is directly linked to a cell's growth rate, it has been hypothesized that high ribosome numbers in ESCs are necessary for triggering cell division signals dependent on ribosome number in the cells (Cheng et al., 2019; Dez and Tollervey, 2004). Multiple studies have supported the idea that a pluripotent cell must have a large number of latent ribosomes to quickly generate large numbers to make the differentiation step efficient (Saba et al., 2021). But major studies are yet to be conducted in this context to understand the details of this global proteome redesign.

The targeted proteomics analysis showed that Rpl39l is present in a few, perhaps up to 10% of the ribosomes in mESCs. While in the Rpl39l KO lines of E14 mESCs showed strong differentiation phenotypes and statistically significant phenotypes in terms of global translation, proliferation and cell death, the mechanism of such a wide-ranging effect by a rather rare subset of ribosomes is indeed an interesting subject that merits further research. Given the similarity of Rpl39l to Rpl39, we expected that Rpl39l sits on the exit tunnel too. The NPET, for a long time was believed to be only a passive conduit for the nascent peptide, but in the last decades, studies have shown that the tunnel plays an important role in the regulation of translation, protein folding and translocation targeting (Deutsch, 2003). X-ray structure studies have found the NPET to be primarily composed of rRNA, which gives it an overall negative charge (Nissen et al., 2000). Along with the rRNA, the tunnel is also composed of Rpl4, Rpl17 (L22 in prokaryotes) and Rpl39 (L23 in prokaryotes) (Nissen et al., 2000). The tunnel has the first constriction site made by L4 and L17 (the only constriction site in prokaryotes), and a second

constriction site in Eukarya, due to an extra loop of L4 and L39 (Nissen et al., 2000). The evolution of Rpl39 to replace prokaryotic Rpl23 has not only led to the elongation of the NPET but also the reduction in the tunnel radius near the tunnel exit (Armache et al., 2010). The work of Dao Duc, 2019 has shown that while the negative charge in the tunnel comes from the rRNA, the tunnel surface is interspersed with positively charged amino acids Arginine and Lysine. In fact, the consensus motif of the tunnel lining amino acids has a ~30% enrichment of these two amino acids and is conserved across multiple kingdoms of life (Dao Duc et al., 2019). Interestingly, the positively charged amino acids are frequently co-occurring at the same sites upon alignments of the tunnel lining protein sequences across Bacteria, Archaea and Eukarya (Dao Duc et al., 2019). This conservation of sequence and charge property of the tunnel is probably not without its reason.

Thanks to the recent improvement in CryoEM capabilities, folding of an NC in the NPET has been studied on a structural level. It has been shown that both hydrophilic and hydrophobic NCs can fold into alpha helices in the lower tunnel near the exit (Wilson and Beckmann, 2011). While a very small tunnel radius can affect the elongation rate during protein synthesis by slowing the movement of NC and not leaving enough space in the PTC for the new amino acylated tRNA to bind, too wide a tunnel radius could increase conformation sampling, increasing the probability of misfolding events early on during translation (Chen et al., 2013). Studies have shown that the increase in radius followed by the constriction site creates an entropic barrier for a diffusing particle, that is neutralised if the particle is positively charged (Duc and Song, 2018). There are multiple studies showing the role of the tunnel electrostatics in modulation of the NC elongation rate and stalling (Duc and Song, 2018; Lu and Deutsch, 2008). Proteins like SecM in *E. coli* have been studied in context of their translation stalling capability where the stalling signal originates in the NPET due to the folding of the transmembrane alpha helix (Zhang et al., 2015).

This brings us to the next function of NPET, to modulate translocation. A stalled SecM nascent chain(NC)-ribosome complex is given a translation boost by the pull generated by the translocon (Zhang et al., 2015). While mammalian systems lack the ribosome-Hsp70 interaction of this sort, multiple other interactors of ribosome like the RAC, SRP and the translocon complex are still dependent on the signalling through NPET for recruitment on the ribosome and proper folding of the protein getting translated at the time (Lee et al., 2021). The proper functioning and interaction of the NPET in eukaryotes appears to be valuable enough for the system to evolve specific adapter proteins to keep an empty NPET occupied, lest it gets spurious interactions with SRP and the translation complex. This protein complex is called NAC, composed of n-betaNAC and n-alphaNAC, where the protrusion created by the C-terminus of the n-betaNAC goes deep into the NPET and actively samples the Arginines in the tunnel (Gamerdinger et al., 2019). Translation of a NC pushes the N-terminus of the n-betaNAC out of the tunnel, making the conditions for recruitment of different factors like SRP.

In context of NPET evolution, the evolution of tunnel radius-restricting proteins like Rpl39 have reduced the solvent accessible volume at the vestibular region of the NPET (Dao Duc et al., 2019). Studies have suggested that the reduction in the vestibular space leads to less favourable environment for folding of proteins inside the ribosome, which in itself is an indicator of change in the division of labour between eukaryotic and prokaryotic chaperone networks (Pechmann et al., 2013). In prokaryotes, the bacterial trigger factors determine chaperon recruitment for both the co-translational protein folding and stress-induced protein denaturing. In eukaryotes, on the other hand, specialised and distinct chaperone networks evolved to deal with these functions. While in prokaryotes, ribosomal folding of NC does some of the folding heavy lifting to support the chaperone network, in eukaryotes, specific families of chaperones like Prefoldin do the job (Kramer et al., 2009). The evolution of Rpl39 also changes the way in which SRP is recruited. In bacteria, uL23 acts to recognise the NC before it

emerges from the tunnel and help recruit the SRP (Bornemann et al., 2008), but eL39 allows the ribosome to get rid of the necessity to signal from within the ribosome for SRP binding and instead is directly involved in SRP binding when the NC protrudes out of the NPET (Berndt et al., 2009). It can also be hypothesised that the evolution of eL39 and the associated tunnel constriction as a barrier against macrolide antibiotics invading the NPET on the lines of L4 and L22 mutations that give rise to Puromycin resistance in Bacteria (Tu et al., 2005; Zaman et al., 2007).

This study, to elucidate the role of Rpl39l, has been designed based on our current understanding for the role of Rpl39 and its significance, with specific focus on its NPET location, role in folding of transmembrane protein, translation fidelity and speed. Even though the exact role of Rpl39 is not completely clear beyond its immediate interactors, studying Rpl39l holds the promise to shed some more light on the functioning of Rpl39 as well. Specifically, translation fidelity, while a major driver of evolution, has not been studied much in the context of specific ribosomal proteins not directly sitting on the PTC. Overall, we are looking at the nature of a very tiny and rare protein having a major effect on translation through their location on NPET in the behemoth we call Ribosome, we are looking at an object of tiny dimensions, casting major shadows.

The manuscript describing the function of Rpl39l in mESCs is shown in Supplement 1.

Supplement 1: Ribosomal protein Rpl39l maintains the fidelity of mRNA translation in mouse embryonic stem cells

Arka Banerjee[1], Maciej Jerzy Smialek[1], Meric Ataman[1], Lea Mues[1], Alexander Schmidt[1], Nitish Mittal[1,*], Mihaela Zavolan[1,*]

[1]Biozentrum, University of Basel, Basel, Switzerland

[2]Department of Pathology, University of Basel, Basel, Switzerland

[*]correspondence: nitish.mittal@unibas.ch; mihaela.zavolan@unibas.ch

Introduction

Intimately coupled to the nutrient intake and proliferation rate of cells (Haneke et al. 2020; Edinger and Thompson 2002), protein synthesis is carried out by the highly conserved molecular machine that is the ribosome. In all free-living organisms, ribosomes have the same basic architecture consisting of a small and a large subunit. In mammals, the small 40S subunit contains the 18S ribosomal RNA (rRNA) and 33 ribosomal proteins (RPs), while the large 60S subunit contains 3 rRNAs - 5S, 5.8S and 28S - and 46 RPs. Gene duplications, in particular by retrotransposition, have resulted in ribosomal protein paralogs (Wong et al. 2014), many with strikingly conserved patterns of tissue-specific expression (Guimaraes and Zavolan 2016). The functions of the latter have been only partially characterized. For example, Rpl10l, an autosomal-chromosome-encoded paralog of the X chromosome-linked Rpl10, works redundantly with Rpl10, maintaining the ribosome function upon X chromosome inactivation in the meiotic phase of spermatogenesis (Jiang et al. 2017). Rpl22 controls pre-mRNA splicing, antagonistically with its Rpl22l1 paralog, to modulate morphogenesis in both zebrafish and mouse (Zhang et al. 2017). The muscle-specific Rpl3L is necessary for striated muscle

formation but actively inhibits myotube growth, which is exclusively Rpl3-dependent (Chaillou et al. 2016). Rps27 and Rps27l control the activity of p53 through interaction with the E3 ubiquitin-protein ligase MDM2, the short-lived Rps27l being both a p53 target and a p53 regulator (Xiong et al. 2011).

A so-far poorly studied RP paralog with tissue-specific pattern of expression is RPL39L. The expression of its mRNA is highly specific to male reproductive tissue (Rohozinski et al. 2009), but it has also been observed in pathological contexts such as ovarian (Rohozinski et al. 2009) breast (Guimaraes and Zavolan 2016), and lung cancers, probably due to a cancer-specific hypomethylation of a CpG island (Yan et al. 2019). The *RPL39L* gene is encoded on an autosomal chromosome, 16 in mice and 3 in humans, in contrast to *RPL39*, which is located on the X chromosome in both species. The lack of introns in the coding region indicates that *RPL39L* has been generated by retrotransposition (Uechi et al. 2002), relatively recently in evolution (Wong et al. 2014). Protein-level expression data for RPL39L is almost entirely lacking (Sugihara et al. 2010). The likely reason is the high arginine and lysine content of the protein, leading to extensive tryptic digestion in standard tandem mass spectrometry experiments, which hinders its detection (Garcia et al. 2007).

Consistent with the predominant expression of *Rpl39l* in male reproductive tissues, a recent study reported that mice with a frameshifting deletion in the *Rpl39l* locus exhibit spermatogenic defects (Zou et al. 2021). How this defect comes about has not been clarified, although it is thought to involve the regulation of translation of particular mRNAs. The functions of Rpl39l in other cell types in which its mRNA has been detected remains entirely unexplained.

To elucidate the function of Rpl39l across the cell types in which it is present, we have constructed Rpl39l-deficient mouse embryonic stem cell (mESC) lines, then characterized their molecular profiles and the impact of Rpl39l depletion on the spontaneous and sperm cell differentiation. With ribosome profiling we demonstrated that Rpl39l impacts the translation of membrane-localized, secreted proteins, and sperm cell-specific proteins, explaining the unusual defects in differentiation of Rpl39l-deficient mESCs. Analyses of translation accuracy further showed that the depletion of Rpl39l leads to translation errors, which can explain the endoplasmic reticulum stress and the defects in protein synthesis. Our results elucidate the function of an elusive RP paralog, highlighting unique aspects of protein synthesis in pluripotent cells.

Results

Detection of Rpl39l protein in mammalian pluripotent cells

The large ribosomal component Rpl39 is encoded by a 3 exon-gene in both human and mouse, its coding region (CDS) spanning all of the 3 exons (Fig. 1a). In contrast, the CDS of the paralogous *Rpl39l* is contained in the long, terminal exon (Uechi et al. 2002), indicative of its emergence in evolution by retrotransposition. The *Rpl39l* CDS is highly conserved, both relative to *Rpl39* and between species (Fig. 1a). The conservation is even higher at the protein level: while the 51 amino acids (AA)-long Rpl39 is identical between mouse and human, Rpl39l differs from it by 3 substitutions in mice and by 4 substitutions in humans. Only one of these substitutions, R28Q, is shared between species (Fig. 1a).

The expression of *Rpl39l* has almost exclusively been studied at mRNA level. The Human Protein Atlas (https://www.proteinatlas.org/, (Uhlén et al. 2015)), which contains data up to the

resolution of single cells, shows that the mRNA level of *RPL39L* is highest in the male germ cell lineage, followed by the extravillous trophoblast and then by other of cell types (Fig. 1b). Protein-level expression data for RPL39L is almost entirely lacking (Sugihara et al. 2010). The high similarity of RPL39 and RPL39L makes the development of a specific antibody very challenging and such antibodies do not seem to be available (data not shown). To circumvent these issues and obtain evidence of RPL39L protein expression we have modified the standard tandem mass spectrometry workflow (see Methods), acetylating the lysines in the proteome to direct trypsin digestion to arginines only. We then applied this approach to samples from a variety of human and mouse cell types: mouse sperm cells (as positive control), the mESC line E14, human pluripotent and cancer cell lines. As expected from the mRNA level measurements, we found that the RPL39L/Rpl39l protein is also expressed beyond germ cells, in pluripotent and cancer cells (Fig. 1c). The ratio of Rpl39l:Rpl39 expression, estimated by targeted proteomics, is in the range of ~1:1000 to ~1:1, highest in germ cells (Fig. 2b). To determine whether Rpl39l is incorporated into ribosomes, we carried out mass spectrometric analysis of purified ribosomes from the E14 mESC line. We found that Rpl39l is also present in these samples, in a similar ratio to Rpl39, ~1:20, as in the total lysate (Fig. 2b). Finally, to determine whether this breadth of expression is typical for an RP gene known to be specific for testis, we searched again the Human Protein Atlas (https://www.proteinatlas.org/, (Uhlén et al. 2015)) for *RPL10L* and found the expression of this gene to be more narrowly focused on testis cell types compared to *RPL39L* (Fig. 1b, compare with 1c). Our data thus provide direct and conclusive evidence of RPL39L protein expression in germ cells, pluripotent cells and cancer cell lines. Along with the expression patterns of other RP paralogs, the data also suggest that the function of RPL39L goes beyond spermatogenesis and is manifested as a component of the ribosome, in mRNA translation.

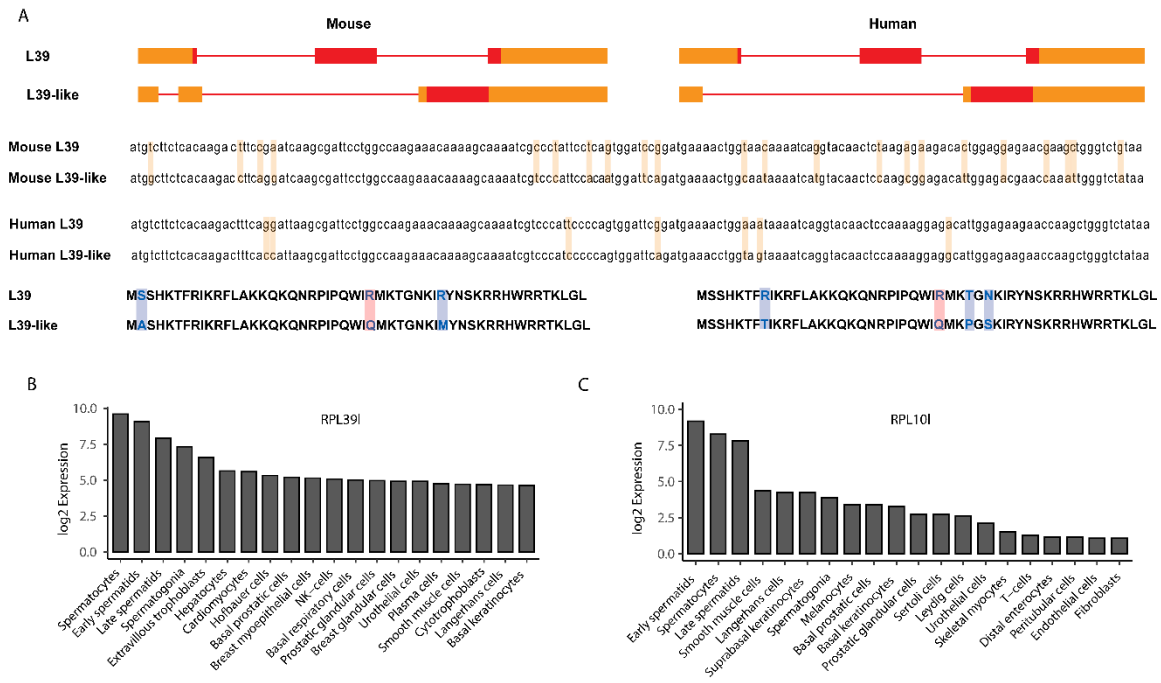


Figure 3. Sequence and expression pattern of Rpl39l. A. Rpl39l is a retrotransposed copy of Rpl39, as indicated by the complete coding region being contained in a single exon, unlike Rpl39. B. Expression of Rpl39l in different cell types according to human protein atlas (https://www.proteinatlas.org/). C. Expression of Rpl10l in different cell types according to the human protein atlas.

KO of *Rpl39l* impairs the pluripotency of E14 mESCs

To investigate the role of Rpl39l in pluripotent cells we used CRISPR-based genome editing to delete the coding region of the gene in E14 mESCs. To minimize the chance of interfering with *Rpl39*, we designed two sets of guide RNAs matching non-coding regions around the *Rpl39l* CDS (Fig. 2a). For each set of sgRNAs we tested multiple clones, originating from independent editing events. To mitigate sgRNA-specific off-target effects, we decided to analyze four distinct clones, two for each sgRNA pair, for which we validated both the editing of the *Rpl39l*

locus by qPCR amplification using probes that flank the edited region and the resulting depletion of Rpl39l protein by targeted proteomics (Fig. 2c). All of these four clones had *Rpl39l* locus deletions, but they presented a spectrum of editing events. Specifically, all clones (labeled as 1.17, 1.20, 2.9 and 2.11) had homozygous deletions of the *Rpl39l* locus, with 2.11 harboring different deletion products on the two chromosomes (Fig 2d). The expression of Rpl39l protein was completely lost in all clones but 1.17, which had a residual Rpl39l expression of ~25% of that observed in wildtype E14 cells. This could be due to an ectopic integration of *Rpl39l* (Boroviak et al. 2017), thus providing us with a "partial rescue" clone, to serve as an additional control that allows us to study dose-dependent effects of Rpl39l.

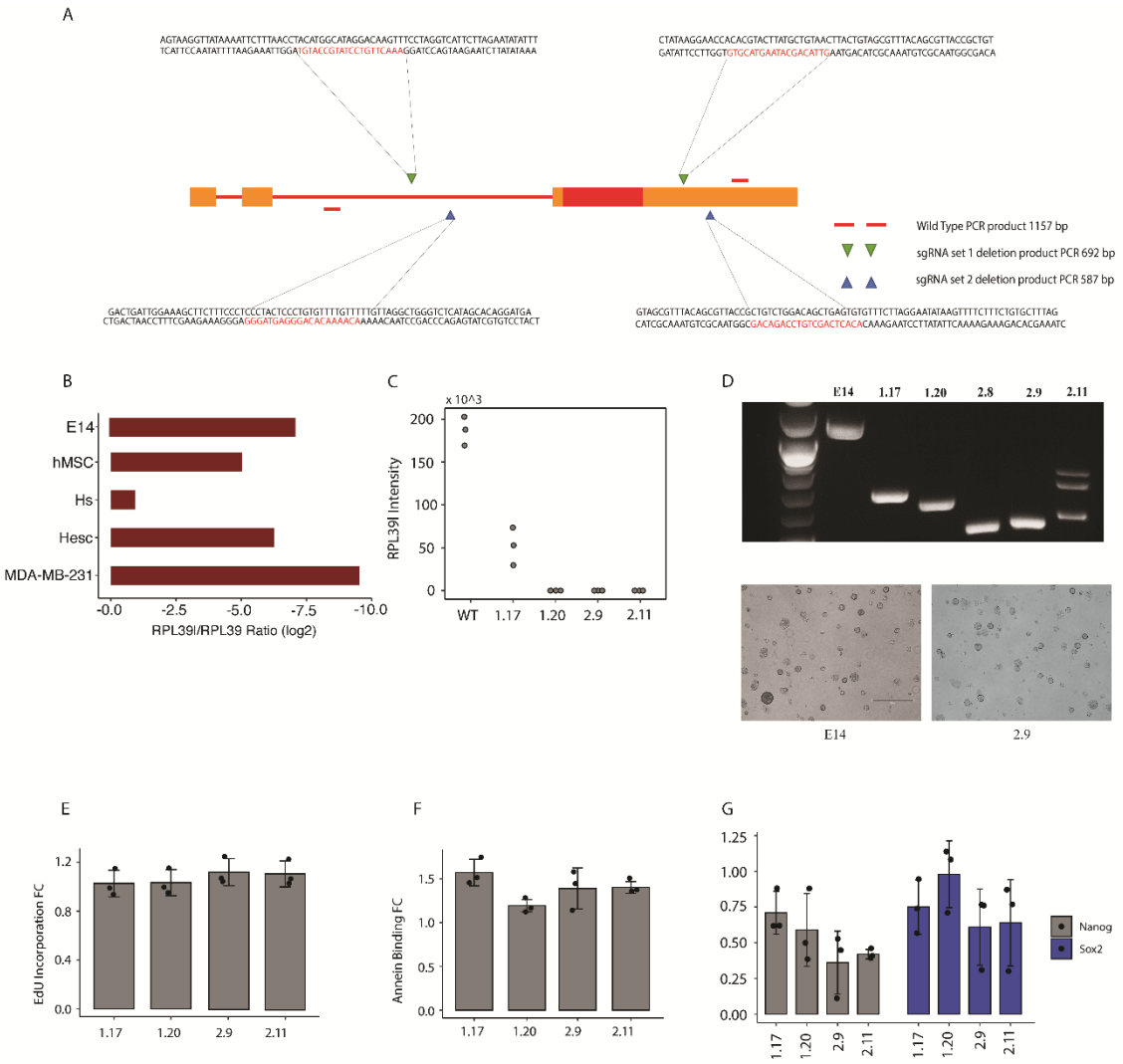


Figure 4. Construction of Rpl39l KO clones. A. Strategy for CRISPR excision of Rpl39l coding region using a dual sgRNA system with Cas9 to minimise the risk of off-target effect on Rpl39. B. Protein expression level of Rpl39l normalised by Rpl39 expression level in different cells. c) Protein expression levels in different Rpl39l KO lines. d) Loci size after excision of the Rpl39l coding region and images of colonies formed by the KO lines. E. Relative EdU incorporation with respect to the WT E14 cells indicates similar proliferation rate. F) Increased annexin binding in the KO lines compared to WT indicates increased apoptosis. G. Reduced level of Nanog and Sox2 expresison compared to WT in the KO lines indicates reduced pluripotency.

The *Rpl39l* KO cells were viable and formed colonies that were similar in structure to those formed by the WT ESCs (Fig. 2d). Treatment with 5-ethynyl-2-deoxiuridine (EdU) for 2 hours did not reveal significant differences in dye incorporation, indicating that the *RPL39l* KO does not obviously alter mESC proliferation (Fig. 2e). The proportion of apoptotic cells in the culture, indicated by anexin staining, was ~1.5-fold higher in the KO lines, though it remained

below 20 percent of the entire cell population (Fig. 2f). Finally, we measured the expression levels of the Nanog and Sox2 pluripotency markers by western blot and found them to be significantly reduced in the KO compared to WT E14 lines (Fig. 2g). These results demonstrate that the *Rpl39l* KO does not affect the viability of mESCs, but it does reduce their pluripotency.

Spontaneous differentiation is disrupted in *Rpl39l* KO mESCs

The reduced expression of pluripotency markers in *Rpl39l* KO E14 cells prompted us to investigate whether these cells also exhibit differentiation defects. As sperm cells have the highest expression of Rpl39l, we first asked whether the KO affects the differentiation of mESCs along the spermatogenic lineage. We subjected both wildtype (WT) and *Rpl39l* KO E14 lines to *in vitro* differentiation along the spermatogenic lineage using a previously described protocol (Kerkis et al. 2007). The emergence of spermatocyte-like cells in the culture of WT E14 mESCs demonstrates that the protocol works as expected (Fig. 3a). In contrast, we could not identify any spermatocyte-like cells in the cultures of *Rpl39l* KO lines (Fig. 3a), and the expression of germ cell lineage markers Stella, Dazl and Haprin was significantly reduced relative to differentiating WT cells (Fig. 3b). These results demonstrate that Rpl39l is indeed necessary for spermatogenesis *in vitro*. However, given that the expression of Rpl39l is not restricted to the spermatogenic lineage (Fig. 1), we wondered whether other developmental lineages may also be affected.

To address this, we carried out the spontaneous differentiation of mESCs *in vitro*, by culturing *Rpl39l* KO and WT E14 cells in leukemia inhibiting factor (LIF)-free medium under non-adherent conditions (Ngondo et al. 2018). We then analyzed the expression of lineage markers

both by qRT-PCR and by immunofluorescence-based quantification of protein levels in embryoid bodies (Fig. 3c). The qRT-PCR revealed that the expression of ectodermal lineage markers *Nestin, Fgf5 and Pax6* was higher (Fig. 3d), while that of extraembryonic endoderm markers *Gata6, Gata4 and Dab2* was reduced in the organoids generated from the KO cells compared to those generated from WT E14 cells (Fig. 3e). The mesoderm markers *Brachyury, Fgf8 and Actc1* showed only small and inconsistent changes across cell lines (Fig. S2). Protein level quantification of Nestin (ectoderm) and Gata4 (endoderm) confirmed these results (Fig. 3f), thus demonstrating that the KO of *Rpl39l* impairs the spontaneous differentiation of mESCs, with the largest impact on the ectoderm and endoderm differentiation.

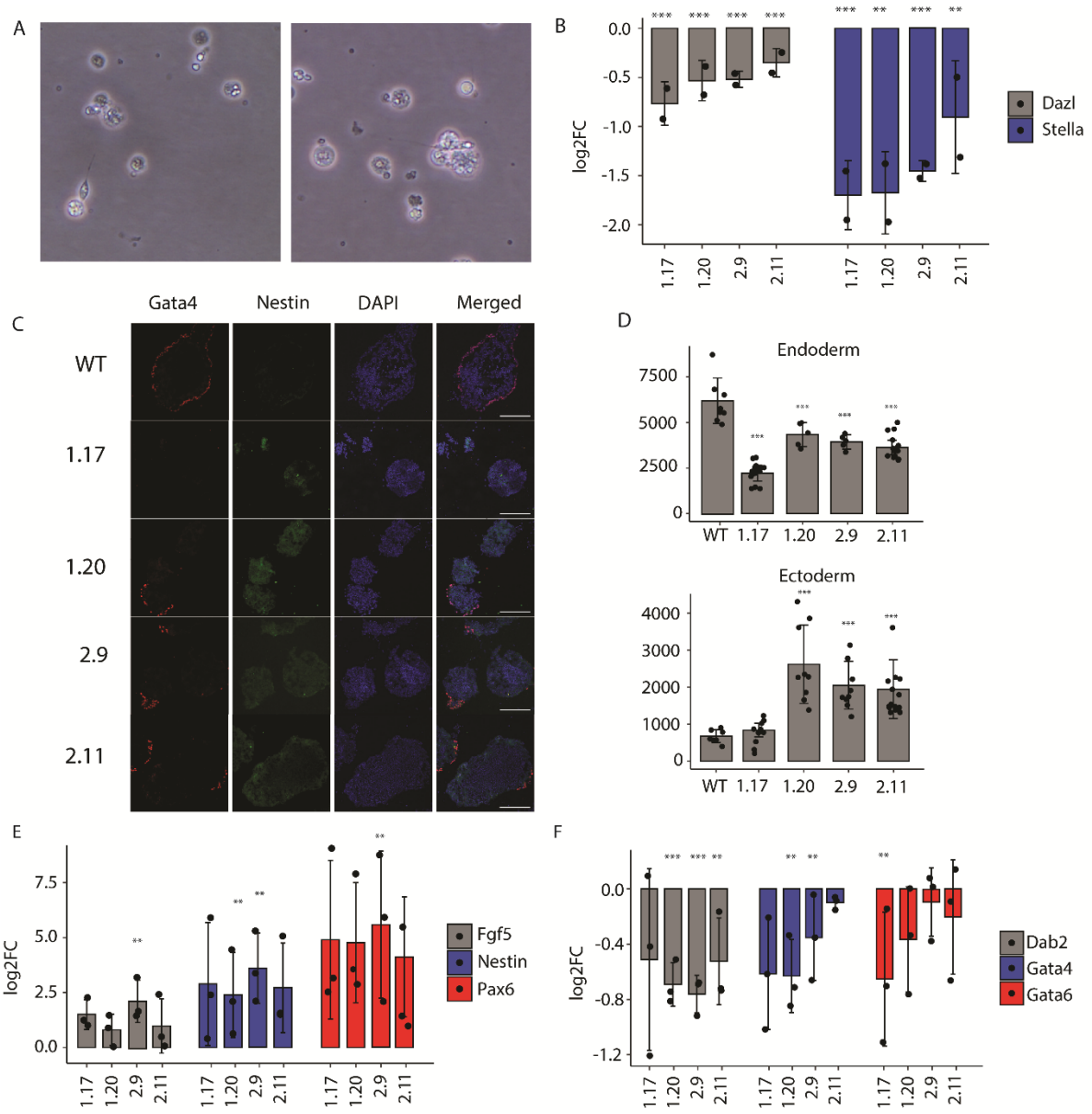


Figure 5. Differentiation defects of Rpl39l KO E14 cell lines. .) Sperm-cell like structures obtained in the in vitro spermatogenesis culture of WT cells, B. Reduced expression of germ cell lineage markers in KO lines undergoing spermatogenesis when compared to WT line. C. and D. Strong reduction in the expression of the endoderm marker Gata4 and increased expression of the ectoderm marker Nestin in spontaneously differentiation KO lines. E. and F. Other markers for endodermal and ectodermal differentiation behave similar to Gata4 and Nestin respectively, in spontaneously differentiating KO lines.

Translation of membrane-associated proteins is reduced *Rpl39l* KO lines

The fact that Rpl39l is incorporated in the ribosomes of mESCs (Fig. 1b) suggests that it may be necessary for the synthesis of a class of proteins with specific relevance to this cell type. To

first determine whether Rpl39l impacts translation globally, we measured the incorporation of O-propargyl-puromycin (OPP) analog of puromycin. We found a slight (10-15%) though significant upregulation in translation rate in Rpl39l KO cells (Fig. S3), further validated by the increased ratio of the area under the curve corresponding to polysome fractions vs. the 80S fraction in polysome profiles (Fig. 4a, Fig. S3). Thus, the KO of *Rpl39l* induces a detectable increase in global translation. To identify the classes of mRNAs that are specifically impacted by the loss of Rpl39l, we carried out ribosome footprinting (Ingolia et al. 2012), sequencing ribosome-protected mRNA fragments (RPFs) from both WT and mutant E14 cell lines.

The RPF data fulfilled expected quality criteria such as a high proportion of reads mapping to the coding regions of mRNAs (Fig. S3) and the 3-nucleotide periodicity of ribosome footprints (Fig. 4b). We did not observe any specific clustering of the samples, indicating that the differences in gene expression between the different clones were relatively small (not shown). We further sequenced the mRNAs from these cells and calculated the translation efficiency (TE) per mRNA as the ratio of the RPF and mRNA read density along the CDS (see Methods). Focusing on mRNAs whose TE was significantly perturbed in at least one of the KO lines relative to the WT, we found that while the differences in RPF and TE of KO clones relative to WT were not large, they were highly consistent (Fig. 4c -wheel). The mRNA levels of the corresponding transcripts remained relatively unchanged. Gene ontology analysis showed that transcripts encoding components of the endomembrane system, including Golgi apparatus and endoplasmic reticulum had the highest and consistent reduction in TE across KO lines (Fig. 4d). Many of these proteins are predicted to have signal peptides and transmembrane domains (Fig. S3), which could explain the enrichment in ion-binding proteins in this set as well (Fig. 4d). In contrast, we found a significantly increased TE for transcripts associated with the cellular response to chemical and oxidative stress (Fig. 4e). These results demonstrate that the

KO of *Rpl39l* induces cellular response to stress and impairs the translation of proteins that can be broadly classified as membrane-associated.

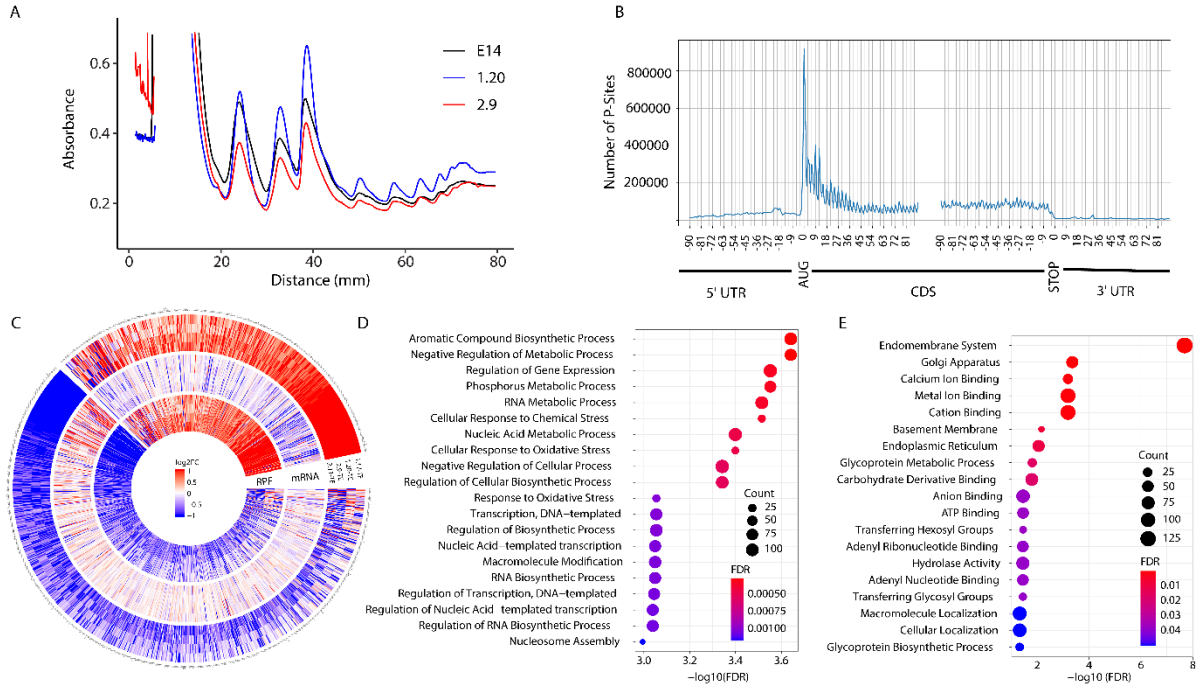


Figure 6. Ribosome footprint analysis of E14 cell lines. A. Example polysome profiles including the WT E14 cell cline as well two of the Rpl39l KO clones. B. Distribution of inferred P-sites around the start and stop codons of transcripts revealing the expected 3-nucleotide periodicity. C. Differences in mRNA-seq, ribosome footprint and translation efficiency (TE) between KO and WT clones. Only genes with a significant change in TE in at least one clone are included. The range of log-fold-changes is capped at +/-1. D. and E. Gene Ontology analysis of upregulated and downregulated genes.

Rpl39l KO disrupts ER-associated functions

The impaired translation of membrane-associated proteins on the background of increased cellular response to stress led us to investigate whether the ER-related functions are affected by the *Rpl39l* KO. First, we measured the levels phosphorylated PKR-like ER kinase (Perk), a main sensor of ER stress (Jäger et al. 2012) and of its phosphorylation target, Eif2a (Harding et al. 1999), which is responsible for regulation of many stress response genes. Western blotting

showed that both of these markers had elevated expression in *Rpl39l* KO compared to WT E14 cells, confirming the results of the ribosome footprint analysis (Fig. 5a-b).

The ER being the site of glycosylation-dependent quality control of proteins and given that some glycosylases had reduced TE in Rpl39l KO lines, we sought to measure the overall abundance of glycosylated proteins in our different mESC lines. We thus separated proteins on a PAGE gel and used the Pro-Q™ Emerald 488 Glycoprotein staining kit to label glycosylated proteins. We then compared the intensity of the fluorescence integrated over entire lanes, each corresponding to a specific cell line. As shown in Fig. 5c, the intensity of the signal was reduced in *Rpl39l* KO cell lines compared to the WT E14 line, indicating a defect in protein glycosylation. Thus, the KO of *Rpl39l* leads to ER stress and disrupts ER-associated protein glycosylation and quality control.

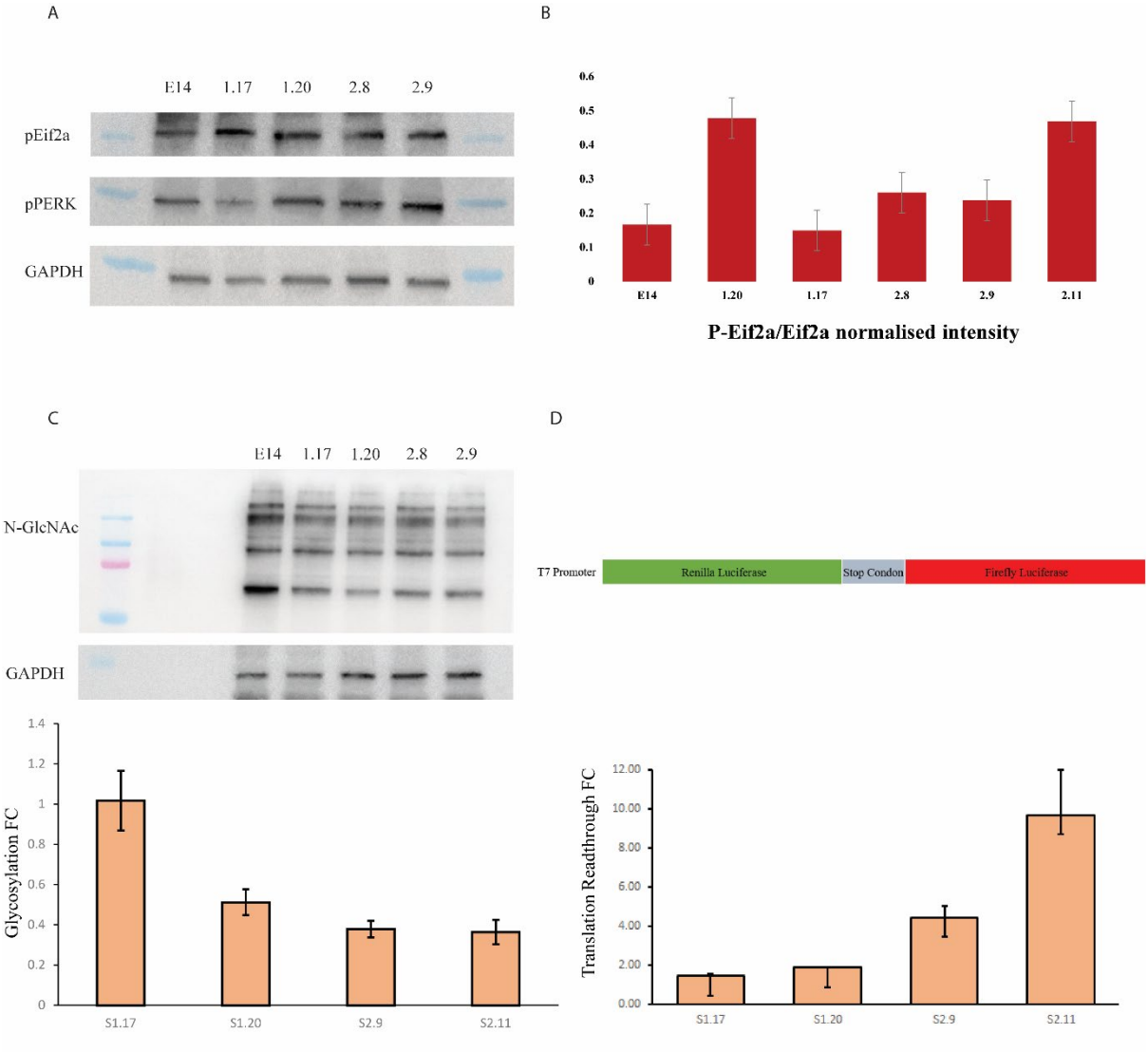


Figure 7. Rpl39l KO disrupts ER function. A. and B. Western blot and quantification indicating increased ER stress (phosphorylation of Eif2a and PERK) in KO vs. WT E14 lines. C. Glycosylation staining strength goes down in KO lines compared to the WT. D. rate of translation readthrough, normalised by the readthrough rate of ribosomes in WT E14 cells.

KO of Rpl39l impairs translation fidelity

The apparent ER stress along with the increased ubiquitination in the Rpl39l KO lines, made us wonder whether the cells experience an increased production of erroneous proteins, as a result reduced translational fidelity. This defect can manifest itself, for example, as an increase in the rate of stop codon readthrough, which can be measured with dual reporter constructs (Halvey et al. 2012). We used a construct composed of the SV40 promoter, the coding region

of Renilla luciferase, a short peptide-coding sequence containing an in-frame stop codon and finally the coding region of the firefly luciferase. While the expression of Renilla is constitutive, the firefly luciferase is expressed only upon stop codon readthrough, and the ratio of the two signals is a proxy for the rate of stop codon readthrough that can be compared between cell lines. Strikingly, *Rpl39l* KO lines had much higher firefly to Renilla expression ratio than the WT line, indicating a reduced termination of translation at the appropriate stop codon (Fig. 5d). Altogether, these results lead us to the following model: Rpl39l, which differs in a few amino acids from the canonical Rpl39 protein, is incorporated in a fraction of mESC ribosomes, and participates in the control of translation fidelity, similar to Rpl39 (Peisker et al. 2008). Lack of Rpl39l results in increased translation errors, ER stress and perturbed production of transmembrane and secreted proteins. While these perturbations are not very large, they impact the ability of mESC to differentiate and lead to a spermatogenic defect.

Discussion

Although cell type-specific gene expression is primarily controlled at epigenetic and transcriptional level, numerous studies in the past decade have revealed the importance of translation-related mechanisms in the establishment and maintenance of cell identity (reviewed in (Saba et al. 2021)). MRNA translation is carried out by ribosomes, which are highly conserved molecular machines, dating back to the last universal common ancestor of living cellular organisms (Mushegian 2008). Although a ribosome can in principle translate any of the mRNAs expressed in a cell, the protein output, dependent on the efficiency of translation initiation and the speed of peptide chain elongation, varies widely among mRNAs (Schwanhäusser et al. 2011; Riba et al.). Variations in ribosome concentration across cells also induce mRNA-dependent changes in protein output (Lodish 1974; Mills and Green 2017; Guimaraes et al. 2020). In recent years, many new examples of cell type-specific translational programs controlled by cell type-specific ribosomal protein paralogs (Kondrashov et al. 2011)

or by post-translational modifications of ribosome-associated proteins (Genuth and Barna 2018) have been described. The human and mouse genomes encode around 20 RP paralogs, their range of expression across tissues being generally more restricted compared to the canonical RPs (Guimaraes and Zavolan 2016). Among the RP paralogs, RPL39L is seemingly the most elusive so far. Its functional relevance has only started to emerge last year, when a study showed that the Rpl39l deficiency leads to spermatogenesis defects in mice (Zou et al. 2021). Interestingly, while enriched in testis, the expression of the *Rpl39l* mRNA is not restricted to this organ. It has been observed in various other cell types including ESCs and, most puzzling, cancer cells (Rohozinski et al. 2009; Guimaraes and Zavolan 2016; Yan et al. 2019). These observations prompted us to study the role of Rpl39l in mESCs and to evaluate its potential to modulate translation. Our main findings are as follows.

First, with a targeted proteomics approach we were able to detect and quantify for the first time the Rpl39l protein across a range of human and mouse cell types. The relative Rpl39l:Rpl39 ratio varied from ~1:1000 in triple negative breast cancer cell line MDA-MB-231 to ~1:1 in mouse sperm cells. In mESCs the relative Rpl39l:Rpl39 abundance was in the ~1:20 range, similar in total lysate and purified ribosomes. This supports the notion that Rpl39l functions as a component of the ribosome, a few percent of mESC ribosomes containing Rpl39l.

Second, we demonstrated that Rpl39l contributes to mESC pluripotency and differentiation along multiple lineages. Similar to Rpl10l, Rpl39l is a recently evolved autosomal chromosome-encoded paralog of an X chromosome-encoded RP. These paralogs are thought to take part in the formation of testis-specific ribosomes, necessary for spermatogenesis (Sugihara et al. 2010). Indeed, Rpl10l has been shown to compensate for the reduced level of Rpl10 during meiotic sex chromosome inactivation (Jiang et al. 2017), while mice deficient in Rpl39l exhibit spermatogenesis defects (Zou et al. 2021). In contrast to Rpl10l however, Rpl39l

expression is not restricted to testis, but extends to other cell types. This suggests that Rpl39l is required in cell types other than male germ cells. Indeed, by depleting Rpl39l in E14.tg2a mESCs, we found that the protein contributes to the maintenance of pluripotency and is necessary for differentiation along multiple lineages. Rpl39l KO cells not only lose the ability to differentiate into sperm cells, but also show defects in spontaneous differentiation: the expression of ectoderm markers is increased and of endoderm markers decreased in Rpl39l KO cell-derived embryoid bodies compared to embryoid bodies derived from WT mESCs. CRISPR/Cas9-based editing in mESCs is broadly used to explore gene function.

Third, we demonstrated that the loss of Rpl39l in mESCs leads to ER stress and disproportionately affects the synthesis of glycosylated, membrane-associated and secretory proteins, due to increase in translation error. Thus, Rpl39l-containing ribosomes could be considered specialized ribosomes that support the synthesis of higher fidelity protein product. Pluripotent stem cells and their health are crucial for fitness of a species, not only from the perspective of developmental fidelity, but also in terms of wound healing. Lower translation error leads to reduced speed of ageing due to proteostatic stress. Structural studies have shown that the Rpl39 homolog is located in the peptide exit tunnel (PET) of the ribosome (Nissen et al. 2000; Zhang et al. 2013), where it interacts with both hydrophilic and hydrophobic nascent polypeptide chains, and specifically with the signal anchor (SA) sequence of ER-targeted transmembrane proteins (Zhang et al. 2013). Yeast Rpl39 also interacts with the protein folding catalyst Hsp70 (Gumiero et al. 2016), which may explain why the deletion of Rpl39 leads to impaired translation fidelity (Dresios et al. 2000). While Hsp70 class chaperone and ribosome interaction in mammals is missing, Rpl39l-containing ribosomes may have similar interactions with other chaperone classes (Hsp1a1 and Hsp2a1 in mammals) to modulate the translation fidelity in ESCs.

mRNA-specific changes in translation can be detected with ribosome profiling (Wang et al. 2021). Here we applied this powerful technique to study the impact of a specialized ribosome in translation. Changes in translation efficiency occurred in specific classes of mRNAs, with very high consistency among Rpl39l KO E14 lines. Most downregulated were membrane-associated proteins like proteins associated to ER and Golgi apparatus, proteins associated to cell adherence like basement membrane proteins and proteins involved in protein glycosylation. Upregulated in TE were transcripts involved in stress response, oxidative damage, protein phosphorylation and vesicle docking. The oxidation stress in Rpl39l KO cells is suggested by multiple lines of evidence. First is the perturbation of mitochondrial respiration, as the Rpl39l KO has been associated with disrupted mitochondrial function (Zou and Qi 2021) and with perturbed translation of proteins involved in mitochondrial respiration, shown in the present study. A second contributor to cellular stress is the ER-associated unfolded protein response (UPR) (Read and Schröder 2021), which we read out in the KO lines via the increased phoshphorylation of PERK and Eif2a. While pluripotent cells have higher resistance to ER stress than differentiated cells, a certain level of cellular UPR is needed for normal differentiation along the endodermal and ectodermal lineages (Kratochvílová et al. 2016). However, the general increase of UPR in ESCs increases their susceptibility to apoptosis (Kratochvílová et al. 2016), as we have observed, in fact, in our study.

The molecular mechanisms underlying the spermatogenic defect in Rpl39l KO lines remain to be identified in detail. Given the importance of glycosylation for spermatogenesis (Akintayo and Stanley 2019), the reduced translation of spermatogenesis-specific glycosylation genes in Rpl39l KO cells may already provide a sufficient explanation (Song et al. 2013). However, our data also showed that the translation factor Eif2s3y, associated with the maintenance of mESC pluripotency (Li et al. 2016) and required for spermatogenesis (Liu et al. 2021), is down-

regulated in Rpl39l KO cells, due to a decrease in its mRNA level. This indicates that multiple layers of regulation lead to the spermatogenesis defect in Rpl39l KO animals (Zou et al. 2021).

While our work elucidates the function of a specialized ribosome in mESCs, it also raises additional questions that could be addressed in the future. First, the implications of Rpl39l KO on the *in vivo* development are insufficiently understood. As already mentioned, an Rpl39l-deficient mouse was obtained via a CRISPR/Cas9-induced frameshifting deletion in *Rpl39l* (Zou et al. 2021), but the viability of various Rpl39l mutants is not known and the obtained mouse model was not investigated beyond the spermatogenesis defect. At the molecular level, it will be interesting to investigate further how ES cells fine-tune their translation output and the quality of synthesized proteins using modular ribosome components. The role of Rpl39l in cancers also remains to be studied, though we expect that some of the observations we made in ESCs will also apply to cancer cells. Given that the abundance of Rpl39l in cancer cell lines is still lower than in mESCs, it will be interesting to determine whether specialized ribosomes that represent less than a percent of cellular ribosomes can influence cellular behaviors in a meaningful way.

In conclusion, Rpl39l is a modular component of ribosomes that can be exchanged against Rpl39 for tightening the translation fidelity. Through this function, Rpl39l exerts a disproportionate impact on physiology of the cells it expresses in and is involved in critical functions like secretion, cellular adhesion and pluripotency.

Methods

Cell culture

In the absence of feeder cells, the WT E14 mESC line was cultured in Dulbecco's Modified Eagle Media (DMEM) (Gibco 31966), which contained 15% fetal bovine serum (FBS; Gibco

16141079) tested for optimal mESC growth, NEAA (Gibco 11140050), Sodium Pyruvate (Gibco 11360039), 100 U/mL LIF (Millipore ESG1107) and 0.1 mM 2-ß-mercaptoethanol (Merck ES-007-E), on 0.2% gelatin coated plates. The culture medium was changed daily, and the cells were passaged every second or third day. Cells were cultured at 37°C in 5% CO2.

Rpl39l CRISPR in mouse embryonic stem cells

All transfections for the CRISPR knockout were performed using the lipofectamine 2000 reagent (Life technologies), according to manufacturer's instructions. SgRNAa for the *Rpl39l* gene were designed in pairs, from the sequences upstream and downstream of the 5' and 3'UTRs, respectively. Upstream sgRNAs were cloned into the px330 plasmid backbone with an mCherry marker, and downstream sgRNAs were cloned in the px330 backbone with a GFP marker. The cells positive for both mCherry and GFP expression were FACS sorted.

For single cell colony selection, FACS sorted cells were diluted to a concentration of 0.75 cells per 100 µl in the ESC culture media. 100µl of this solution was added to each well of a 96 well plate. The cells were allowed to grow for 2 days before wells with single clones were marked.

Forward primers for CRISPR KO validation were designed upstream of the upstream sgRNAs and the reverse primers were designed downstream of the downstream sgRNAs. In clones with homozygous KO the PCR product should consist of a single band, migrating lower (692bp and 587bp in KO, while 1157bp in WT) than the wild type band when run on a 1.3% agarose gel. The bands were excised and gel purified using the Qiagen Gel purification Kit, according to manufacturer's instructions. The purified DNA was cloned into a plasmid

backbone using Zero Blunt Topo kit (ThermoFischer) according to manufacturer's instructions. This plasmid was sequenced to validate the complete excision of *Rpl39l*.

LC-MS analysis

Sample preparation

Cells were collected and lysed in 50 µl lysis buffer (2 M Guanidinium-HCl, 0.1 M HEPES, 5 mM TCEP, pH = 8.3) using strong ultra-sonication (10 cycles, Bioruptor, Diagnode). Protein concentration was determined by BCA assay (Thermo Fisher Scientific 23227) using a small sample aliquot. Sample aliquots containing 50 µg of total proteins were supplemented with lysis buffer to 50 µl, reduced for 10 min at 95 °C and alkylated at 10 mM iodoacetamide for 30 min at 25 °C followed by adding N-acetyl-cysteine to a final concentration of 12.5 mM to quench any iodoacetamide excess. For global proteomics analyses, proteins were directly digested by incubation with sequencing-grade modified trypsin (1/50, w/w; Promega, Madison, Wisconsin) overnight at 37°C. For targeted analysis of RPL39 and RPL39L, a mixture containing 100 fmol of heavy reference peptides were added to the samples. Then, protein samples were propionylated by adding N-(Propionyloxy)-succinimid (0.15 M in DMSO) to a final concentration of 22.5 mM and incubating for 2h at 25°C with shaking at 500 rpm. This step prevents trypsin from cleaving after lysine residues and lead to the generation of larger peptides suited for LC-MS analysis for our target proteins. To quench the labelling reaction, 1.5 µL aqueous 1.5 M hydroxylamine solution were added and samples were incubated for another 10 min at 25°C shaking at 500 rpm. Subsequently, the pH of the samples was increased to 11.9 by adding 1 M potassium phosphate buffer (pH 12) and incubated for 20 min at 25°C shaking at 500 rpm to remove propionic acid linked to peptide hydroxyl groups. The reaction was stopped by adding 2 M hydrochloric acid until a pH < 2 was reached. After adding 70 ul of a 1 M TEAB (pH=8.5), proteins were digested by incubation with sequencing-grade modified trypsin (1/50,

w/w; Promega, Madison, Wisconsin) overnight at 37°C. For all samples, the generated peptide were cleaned up using iST cartridges (PreOmics, Munich, Germany) according to the manufacturer's instructions. Samples were dried under vacuum and stored at -80 °C until further use.

For global proteomics analysis, sample aliquots comprising 25 µg of peptides were labeled with isobaric tandem mass tags (TMT 10-plex, Thermo Fisher Scientific) as described previously (Ahrné et al. 2016). Shortly, peptides were resuspended in 20 µl labeling buffer (2 M urea, 0.2 M HEPES, pH 8.3) and 5 µL of each TMT reagent were added to the individual peptide samples followed by a 1 h incubation at 25°C, shaking at 500 rpm. To control for ratio distortion during quantification, a peptide calibration mixture consisting of six digested standard proteins mixed in different amounts was added to each sample before TMT labeling (for details see (Ahrné et al. 2016)). To quench the labelling reaction, 1.5 µL aqueous 1.5 M hydroxylamine solution were added and samples were incubated for another 10 min at 25°C shaking at 500 rpm followed by pooling of all samples. The pH of the sample pool was increased to 11.9 by adding 1 M potassium phosphate buffer (pH 12) and incubated for 20 min at 25°C shaking at 500 rpm to remove TMT labels linked to peptide hydroxyl groups. Subsequently, the reaction was stopped by adding 2 M hydrochloric acid until a pH < 2 was reached. Finally, peptide samples were further acidified using 5 % TFA, desalted using Sep-Pak Vac 1cc (50 mg) C18 cartridges (Waters) according to the manufacturer's instructions and dried under vacuum.

TMT-labeled peptides were fractionated by high-pH reversed phase separation using a XBridge Peptide BEH C18 column (3,5 µm, 130 Å, 1 mm x 150 mm, Waters) on an Agilent 1260 Infinity HPLC system. Peptides were loaded on column in buffer A (20 mM ammonium formate in water, pH 10) and eluted using a two-step linear gradient from 2% to 10% in 5 minutes and

then to 50% buffer B (20 mM ammonium formate in 90% acetonitrile, pH 10) over 55 minutes at a flow rate of 42 µl/min. Elution of peptides was monitored with a UV detector (215 nm, 254 nm) and a total of 36 fractions were collected, pooled into 12 fractions using a post-concatenation strategy as previously described (Wang et al. 2011) and dried under vacuum.

Global proteomics LC-MS analysis

Dried peptides were resuspended in 0.1% aqueous formic acid and subjected to LC–MS/MS analysis using a Q Exactive HF Mass Spectrometer fitted with an EASY-nLC 1000 (both Thermo Fisher Scientific) and a custom-made column heater set to 60°C. Peptides were resolved using a RP-HPLC column (75µm × 30cm) packed in-house with C18 resin (ReproSil-Pur C18–AQ, 1.9 µm resin; Dr. Maisch GmbH) at a flow rate of 0.2 µLmin-1. The following gradient was used for peptide separation: from 5% B to 15% B over 10 min to 30% B over 60 min to 45 % B over 20 min to 95% B over 2 min followed by 18 min at 95% B. Buffer A was 0.1% formic acid in water and buffer B was 80% acetonitrile, 0.1% formic acid in water.

The mass spectrometer was operated in DDA mode with a total cycle time of approximately 1 s. Each MS1 scan was followed by high-collision-dissociation (HCD) of the 10 most abundant precursor ions with dynamic exclusion set to 30 seconds. For MS1, 3e6 ions were accumulated in the Orbitrap over a maximum time of 100 ms and scanned at a resolution of 120,000 FWHM (at 200 m/z). MS2 scans were acquired at a target setting of 1e5 ions, maximum accumulation time of 100 ms and a resolution of 30,000 FWHM (at 200 m/z). Singly charged ions and ions with unassigned charge state were excluded from triggering MS2 events. The normalized collision energy was set to 35%, the mass isolation window was set to 1.1 m/z and one microscan was acquired for each spectrum.

The acquired raw-files were searched against a protein database containing sequences of the predicted SwissProt entries of *homo sapiens* (www.ebi.ac.uk, release date 2019/03/07, in total 20,796 sequences) or *mus musculus* (www.ebi.ac.uk, release date 2019/03/07, in total 17,414 sequences), the six calibration mix proteins (Ahrné et al. 2016) and commonly observed contaminants using the SpectroMine software (Biognosys, version 1.0.20235.13.16424). Standard Pulsar search settings for TMT ("TMT_Quantification") were used and resulting identifications and corresponding quantitative values were exported on the PSM level using the "Export Report" function. Acquired reporter ion intensities in the experiments were employed for automated quantification and statistical analysis using our in-house developed SafeQuant R script (v2.3, (Ahrné et al. 2016)). This analysis included adjustment of reporter ion intensities, global data normalization by equalizing the total reporter ion intensity across all channels, summation of reporter ion intensities per protein and channel, calculation of protein abundance ratios and testing for differential abundance using empirical Bayes moderated t-statistics. Finally, the calculated p-values were corrected for multiple testing using the Benjamini–Hochberg method.

Targeted LC-MS analysis of Rpl39l

Parallel reaction-monitoring (PRM) assays (Peterson et al. 2012; Gallien et al. 2012) were generated from a mixture containing 25 fmol/µL of each proteotypic N-(Propionyloxy)-succinimid labeled heavy reference peptide (SSHKTFR (RPL39 human and mouse), SSHKTFTIKR (RPL39L human), ASHKTFR (rpl39l mouse), JPT Peptide Technologies GmbH). 2 µL of this standard peptide mix were subjected to LC–MS/MS analysis using a Q Exactive plus Mass Spectrometer fitted with an EASY-nLC 1000 (both Thermo Fisher Scientific) and a custom-made column heater set to 60°C. Peptides were resolved using a EasySpray RP-HPLC column (75µm × 25cm, Thermo Fisher Scientific) and a pre-column setup

at a flow rate of 0.2 µL/min. The mass spectrometer was operated in DDA mode. Each MS1 scan was followed by high-collision-dissociation (HCD) of the precursor masses of the imported isolation list and the 20 most abundant precursor ions with dynamic exclusion for 20 seconds. Total cycle time was approximately 1 s. For MS1, 3e6 ions were accumulated in the Orbitrap cell over a maximum time of 50 ms and scanned at a resolution of 70,000 FWHM (at 200 m/z). MS1 triggered MS2 scans were acquired at a target setting of 1e5 ions, a resolution of 17,500 FWHM (at 200 m/z) and a mass isolation window of 1.4 Th. Singly charged ions and ions with unassigned charge state were excluded from triggering MS2 events. The normalized collision energy was set to 27% and one microscan was acquired for each spectrum.

The acquired raw-files were searched using the MaxQuant software (Version 1.6.2.3) against the same human and mouse database mentioned above using default parameters except protein, peptide and site FDR were set to 1 and Lys8, Arg10 and propyl (K) were added as variable modifications. The search results were imported into Skyline (v21.1.0.278) (MacLean et al. 2010) to build a spectral library and assign the most intense transitions to each peptide. An unscheduled mass isolation list containing all peptide ion masses was exported and imported into the Q Exactive Plus operating software for PRM analysis. For PRM-MS analysis, peptide samples were resuspended in 0.1% aqueous formic acid. Due to the required protein propionylation for rpl39 and rpl39l LC-MS analysis, the heavy reference peptides were already spiked in at a concentration of 2 fmol of heavy reference peptides per 1 µg of total endogenous peptide mass during sample preparation (see above). The samples were subjected to LC–MS/MS analysis on the same LC-MS system described above using the following settings: The MS2 resolution of the orbitrap was set to 17,500/140,000 FWHM (at 200 m/z) and the fill time to 50/500ms for heavy/light peptides. AGC target was set to 3e6, the normalized collision energy was set to 27%, ion isolation window was set to 0.4 m/z and the first mass was fixed to 100 m/z. A MS1 scan at 35,000 resolution (FWHM at 200 m/z), AGC target 3e6 and fill time

of 50 ms was included in each MS cycle. All raw-files were imported into Skyline for protein / peptide quantification. To control for sample amount variations during sample preparation, the total ion chromatogram (only comprising precursor ions with two to five charges) of each sample was determined using Progenesis QI software (Nonlinear Dynamics (Waters), Version 2.0) and used for normalization of light (endogenous) peptide abundances.

Spontaneous differentiation

Spontaneous differentiation of mESC lines was carried out in conventional mESC medium (as stated above) with 10% fetal bovine serum and no LIF (Kurosawa, 2007). To avoid attachment, Embryoid Bodies (EBs) were generated by growing 750'000 cells in suspension for 6 days in non-adherent dishes (Greiner Bio-One 633181). Spheroids were allowed to grow without disturbing the plate for the first 3 days. The medium was changed every second day afterwards. The EBs were harvested after 6 days with 25ml Pasteur Pipettes and washed with 1X PBS thrice before further analysis.

Spermatogenic differentiation

Differentiation of mESC toward sperm cells was done according to the protocol by Kerkis et al. 2007. In short, E14 cells were harvested and plated as a hanging media drop on Petri dishes filled with PBS on the bottom of the plate, with 1250 ESC per 25 µl in each drop. The resultant EBs were transferred onto Petri dishes (10–15 EB per dish) after 3 days in hanging drop culture. Neurobasal medium (Gibco 21103049), supplemented with B27 (Invitrogen 17504044) was used as the differentiation medium, according to the manufacturer's protocol 0.1 uM retinoic acid final concentration (Sigma R2625-50MG) was added to the culture medium one day after

EB transfer, and the medium was replaced every two days to minimize deterioration. The contents of each plate were harvested after 4 days.

Polysome profiling and Ribo-Seq

For Ribo-Seq analysis, wild-type and Rpl39l KO E14 mESCs were propagated in 5x15cm Petri Dishes (Falcon, 353025) per sample as described in the cell culture section. The media was replenished three hours prior cell collection at the confluency of 50-70%. Before harvesting, the cells were treated with 100µg/ml cycloheximide (CHX) (Sigma, G7698) for 15 minutes at 5% CO2, 37ºC, to freeze the elongating ribosomes. The cells were harvested on ice in a cold room. Cells were washed twice with ice cold DPBS (Lonza, BE17-512Q) containing 100µg/ml CHX. Cells were scraped, collected, spun down, flash frozen and stored at -80 °C.

The cell pellet was resuspended in 900µl of ice cold polysome lysis buffer (20 mM Tris-HCl (Sigma, T294), pH=7.5; 100 mM NaCl (Sigma, 71386); 10 mM MgCl2 (Sigma, 63069); 1% Triton X100 (Sigma, T8787); freshly added 2 mM DTT (Sigma, 646563); 100 µg/ml cycloheximide (Sigma, G7698); 400U of RNAsin plus RNase inhibitor (Promega, N261B); 20U of Turbo DNase (Ambion, AM2238) and Complet mini, EDTA-free protease inhibitors (Roche, 11836170001)) by pipetting up and down. After resuspending the pellet, the sample was incubated for 5 minutes at 4°C with continuous rotation (50rpm), passed through a 23G needle (Braun, 4657640) for minimum 10 times, followed by additional 5 min incubation at 4°C with continuous rotation (50rpm). The cell lysate was clarified by centrifugation at 3000g/3min/4°C followed by centrifugation of supernatant from the first step at 10’000g/5min/4°C. 50 µl of clarified lysate from each sample was kept aside for mRNAseq library preparation and was snap frozen and stored at -80 °C. The optical density (OD) of

remaining lysate was measured at A_{260} with a NanoDrop2000. For Ribo-Seq analysis, the lysate equivalent to OD (A_{260}) = 7 was subjected to RNase I digestion (5U per OD, 35U in total ; Invitrogen, AM2294) at 22 °C/ 20 minutes with continuous rotation at 1000 rpm in thermoblock (eppendorf). RNase I was inactivated by adding 10µl of RNase inhibitor SuperaseIN (Invitrogen, AM2696) in each reaction. Equal amount of undigested lysate was also used to obtain a polysome profile.

10-50% linear sucrose gradient was prepared using Gradient Master instrument (Biocomp) according to the manufacturer's instructions. In brief, 10 and 50% sucrose (Sigma, 84100) solution was prepared in buffer containing 50 mM Tris-HCl , pH=7.5 (Sigma, T294); 50 mM NH4Cl (Sigma, 09718); 12 mM MgCl2 (Sigma, 63069); 100 µg/ml CHX (Sigma, G7698); 0.5mM DTT (Sigma, 646563) and 10µl SuperaseIN (Invitrogen, AM2696). A 14x89mm tube (Beckman Coulter, 331372) (used in rotor SW-41/TH-641) (Beckmann Coulter) with a long cap that can hold 800 µl of sample was used to prepare the gradient. First 10% solution was laid in the tube followed by 50% solution that was under-laid with the help of a long syringe. The gradient was prepared by using a pre-program for 10% to 50% sucrose gradient in Gradient master 108 instrument (Biocomp). The gradient was cooled down at 4°C by keeping it in the fridge for a minimum of one hour. Undigested (for polysome profile) and digested samples (for Ribo-Seq) were loaded onto pre-cooled 10-50% sucrose gradient and centrifuged at 35'000 rpm for 3 hrs at 4°C in a SW-41Ti rotor (Beckmann Coulter). Finally, all gradients were monitored at wavelength A_{254} and fractionated using Piston Gradient Fractionator (Biocomp). 30 separate fractions of 0.37ml were collected in 1.5ml Eppendorf tubes for each digested and undigested ribosome profiles using a Gilson collector attached with the fractionator.

The appropriate fractions that contain 80S monosomes were processed for Ribo-Seq library preparation by combining the protocol from ref. (PMID: 22836135; PMID: 27380875). In brief, RNA was isolated from the appropriate monosomes fraction by using the hot phenol method. RNA fragments of appropriate size (28-32 nt) were obtained by running samples on 15% polyacrylamide denaturing TBE-Urea gel and visualized by SYBR Gold dye (Life Technologies). Size selected RNA was dephosphorylated by T4 polynucleotide kinase (PNK, New England Biolabs, B0201S) treatment for 1 hour at 37ºC. PNK was heat inactivated and RNA was purified using phenol chloroform method and overnight precipitation of RNA in ethanol. RNA amount equivalent to 12ng was used to prepare the sequencing library using SMARTer® smRNA-Seq Kit for Illumina® (Takara 635031) according to the kit manual till cDNA synthesis. After cDNA synthesis, rRNA was depleted by using protocol and probe defined in ref.(PMID: 22836135) followed by final PCR according to Smarter kit with multiplexing barcodes. The PCR product was purified on 8% polyacrylamide native TBE gel and submitted for sequencing at genomic facility Basel.

qRT-PCR

RNA from Spontaneous and Spermatogenic differentiation was isolated using AccuPure Cell/Blood RNA Mini Kit (96) (AccuBioMed, R10096) using iColumn24 Robot (AccuBioMed) with DNase1 treatment and 50ul volume elution. Reverse transcription was performed from 500-1000ng of RNA using SuperScript IV First-Strand Synthesis System (Invitrogen, 18091200) using Random hexamers (Promega, 300453). qPCR reaction was performed in 20ul volume using PowerSYBR Green PCR Master Mix (AppliedBiosystems, 4367659) and QuantStudio 3 System (Applied Biosystems, A28567) using Comparative CT

(ΔΔCT) analysis and Standard Run mode. Rrm2 was treated as endogenous control for all analyses. The primer sequences used are in table 1.

Table 1.

Primer ID	Primer Sequence
Oct4_Fw	CAACTCCCGAGGAGTCCCA
Oct4_Rev	CTGGGTGTACCCCAAGGTGA
Nanog_Fw	CAGAAAAACCAGTGGTTGAAGA
Nanog_Rev	GCAATGGATGCTGGGATACTC
Sox2_Fw	CACAGATGCAACCGATGCA
Sox2_Rev	GGTGCCCTGCTGCGAGTA
Nestin_Fw	CTCTTCCCCCTTGCCTAATACC
Nestin_Rev	TTTAGGATAGGGAGCCTCAGACAT
Fgf5_Fw	GCCTGTCCTTGCTCTTCCTCAT
Fgf5_Rev	GGAGAAGCTGCGACTGGTGA
Pax6_Fw	TAACGGAGAAGACTCGGATGAAGC
Pax6_Rev	CGGGCAAACACATCTGGATAATGG
Gata6_Fw	GAAGCGCGTGCCTTCATC
Gata6_Rev	GTAGTGGTTGTGGTGTGACAGTTG
Gata4_Fw	CTGGAAGACACCCCAATCTC
Gata4_Rev	GTAGTGTCCCGTCCCATCTC
Dab2_Fw	TTGATGATGTGCCTGATGCT
Dab2_Rev	TTTGCTTGTGTTGTCCCTGA
Rrm2_Fw	CCGAGCTGGAAAGTAAAGCG
Rrm2_Rev	ATGGGAAAGACAACGAAGCG
Dazl_Fw	GCCAGCACTCAGTCTTCATC
Dazl_Rev	GTTGGAGGCTGCATGTAAGT
Stella_Fw	CAGCCGTACCTGTGGAGAACAAGAG
Stella_Rev	AGCCCTGGGCCTCACAGCTT

Spontaneous differentiation organoids immunofluorescence

Organoids were fixed using 4% PFA in PBS at 4 °C for 2hrs. After two washes in PBS, the organoids were transferred to 10% sucrose solution in PBS and stored at 4 °C for 1 day. This was followed by a 1-day incubation in 20% sucrose in PBS at 4 °C, and finally by a 1-day incubation in 30% sucrose in PBS at 4 °C. Next, the organoids were embedded in PolyFreeze Tissue Freezing Medium (Sigma, SHH0025) in ibidi slides (ibiTreat, 80826), snap frozen on dry ice and stored at −80 °C until cryosectioning. Cryosections of 12-µm thickness were made on Superfrost Ultra Plus Gold Adhesion slides (Thermo Fisher, 11976299) using a Leica Microsystems cryostat. Slides were stored at −80 °C or processed immediately. For IF, slides were air-dried at RT for 1 h, washed in PBS three times 5 min each and permeabilised in 0.2% Triton X-100 in PBS for 30 min at RT. The cryosections on each slide were then circumscribed using ImmEdge Hydrophobic Barrier Pen (Vector Labs) and blocking solution (1% BSA, 5% mouse serum in 0.2% Triton X-100) was added for 30 min at RT. Primary antibodies against Gata4-Alexa594 (SantaCruz, sc-25310) and Nestin-Alexa488 (SantaCruz, sc-23927) were diluted 1:50 in blocking solution (with 0.1% Tween) and incubated in dark at 4 °C overnight. In the end slides were washed 3 times with 0.1%Tween in PBS and mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200-10). The images were acquired using Zeiss LSM800 confocal microscope and analyses using Fiji software (PMID: 22743772) and visualized and deposited using OMERO.

mESCs collonies preparation and immunofluorescence

16hrs prior to fixation 30.000 of wild-type or Rpl39l KO mESCs were seeded on 10 Well 6.7mm Diagnostic Microscope Slides (Thermo Scientific, ER-208B-CE24) coated with 0.2% gelatin (Sigma, G1393). Cells were seeded in 40ul volume in 1:500 dilution of TetraSpeck™ Microspheres, 0.2 µm, fluorescent blue/green/orange/dark red (Invitrogen, T7280) for light

source shift corrections. mESCs collonies were fixed using 4% PFA in PBS RT for 1h. Then the slides were washed in PBS three times 5 min each and permeabilised in 0.1% Triton X-100 in PBS for 30 min at RT. Followed by 3 times washes with PBS, 50ug/ml of Concavalin-A-Alexa594 (Invitrogen, C11253) was incubated for 30min RT. Then slides were incubated for 1h RT in blocking solution (3% mouse serum in 0.1% Triton X-100 in PBS). Later the primary antibody against rRNA-Alexa488 (SantaCruz, sc-33678) was incubated on slides for 1h RT (1:50 dilution in blocking solution). In the end slides were washed 3 times with PBS and mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories, H-1200-10). The images were acquired using Zeiss LSM800 confocal microscope and analyses using Fiji software (PMID: 22743772) and visualized and deposited using OMERO.

Western Blot

Samples were collected from cell culture by scraping, centrifuged (5min @210rcf), washed and resuspended in RIPA buffer containing Phosphatase inhibitor (Roche 4906837001) and protease inhibitor cocktail (Roche 118361530001). After sonication (2x 4 times Amp60%, 0,5) on Hilsher UP50H probe sonicator and centrifugation (5min @ 5000rpm) of the samples, a BCA measurement was executed to determine the protein concentration of the samples. For the SDS-Gel electrophoresis, 30ug of the samples were mixed with 4x Lämmli Buffer containing b-mercaptoethanol, heated up for 10minutes at 95°C and centrifuged before loading onto a 10well 5-20% gradient gel (BioRad 456-1093). The finished gels were blotted via semi dry method (20V, 400mA, 45min) onto a nitrocellulose membrane (GE, Amersham Protran Premium 0,2um NC), dyed with ponceau, washed with 1x TBST, blocked with 5% BSA in 1x TBST and kept overnight at 4°C rotating in the respective primary Antibody (refer to table 2). The membranes were washed afterwards 3x 10min with 1xTBST then put into secondary Antibody (concentrations in table 2) at RT for 1h and washed again. They were imaged in the

Fusion FX from VILBER (Software FUSIONFX7 Edge 18.11) with Amersham ECL Western Blotting Detection Reagents RPN2106 from GE Life science.
For loading control the membranes were put in gapdh (Histone H3 for eif2alpha and P-eif2alpha)

Table 2:

Antibody	Use concentration	Supplier
Eif2alpha Rabbit mAb	1:1000	Cell signaling technology 99722S
P-eif2alpha (119A11)	1:1000	Cell signaling technology 3597S
Anti-Rabbit mAb	1:1000	Cell signalling technology 7074P2
Perk Rabbit mAb	1:1000	Cell signalling technology 3192
P-perk Rabbit mAb	1:1000	Cell signalling technology 3179
Nanog	1:1000	Cell signalling technology 8822S
Sox2	1:1000	Biolegend 23064
Oct4	1:1000	Abcam 18976 Lot: GR3243479-1
Gata4	1:1000	Santa Cruz 25310
Histone H3	1:1000	Cell signaling technology 9715
Gapdh	1:1000	Sigma G9295

Glycosylation gel staining

After the PAGE gel was electrophoresed as mentioned in the previous section, the gel was fixed using Thermo Fischer Pro-Q Emerald-488 Glycosylation gel staining kit (P21857) according to the manufacturer's protocol. Briefly, the gel was fixed in a mixture of Acetic acid: Methanol: Distilled Water at a ratio of 10:40:50 overnight. The gel was then acidified with 3 washes of 10 min each 3% Acetic acid, followed by oxidation of the gel using the provided Oxidation Reagent for 30 min. This was followed by three rounds of washing and staining in the Pro-Q

Emerald 488 dye, diluted 1:50 with Gel staining buffer. The final stained gel was again washed thrice in 3% Acetic acid for 10 min each and imaged in Fusion FX from VILBER (Software FUSIONFX7 Edge 18.11).

Readthrough dual luciferase assay

For the assay the dual luciferase reporter assay system from promega was used (ref nr. E1980). 24h after transfection the cells were lysed with 20uL 1x passive lysis buffer per well and 30min shaking at room temperature. 10uL of each sample was transferred into a white microplate (greiner bio-one, 96well, PS, F-bottom, Lumitrac#600, ref655074). For the Luciferase measurement, 50uL of LAR II was added to the samples and read out with the plate reader Mithras2 LB 943 MultimodeReader, program: mikroWin 2010.

Afterwards 50uL Stop & Glo substrate, a mixture of 1:50 (Stop & Glo Reagent : Stop & Glo Buffer), was pipette in each well for the Renilla measurement.

Outlook

In this work, we have successfully shown that not only is Rpl39l a protein that sits in the ribosome, but we also showed that a protein with relatively low abundance with respect to the ortholog seems to wide ranging effect on ES cells. While the ES cells in their undifferentiated state showed some minor changes in terms of the translation rate, proliferation and apoptosis, we observed strong changes in terms of glycosylation defect and ER stress, along with major increase in translation errors. We also observed that there was a change in the translation efficiency of genes from specific functional categories. Although not included in the current manuscript, our analysis of ribosome footprints indicated that in Rpl39l KO lines, some amino acids were incorporated faster than others compared to the WT cells. Thus, rather unintuitive changes seem to culminate into a major defect in the core function of an ES cells, its differentiation capacity. We observed a significant reduction in the expression of endodermal lineage markers and an overexpression of ectodermal markers during spontaneous differentiation, in the KO lines when compared to the WT line. While marker expression change is not enough to conclude the nature of disruption of ES cell differentiation, compared to other results reported in the literature, where high variability between ES lines has also been reported, it is clear that the disruption is significant. The significant increase in the ER stress, the reduction in global glycosylation and the reduced global TE of transmembrane proteins is in line with what we expect from a defect from a protein with similar roles as that of Rpl39. To study this phenomenon further, we did try to generate a Rpl39 KO line of ES cells, but we could not get viable lines (data not shown), probably due to Rpl39 KO being either lethal or completely disrupting pluripotency. Hence, to study the mechanism of Rpl39l more directly, we decided that we will take a three-pronged approach:1) Generating Rpl39 KO yeast lines with mouse Rpl39l expression, to understand the nature of defect in translation. Previous studies in yeast have shown that Rpl39 KO is not lethal but increases translation speed and reduces translation fidelity. This makes yeast a good model to study the role of Rpl39l in the absence

of Rpl39. Previous studies have successfully used yeast a model for understanding the roles of mammalian ribosomal protein, and the existence of yeast KO library makes it a very feasible model for this purpose. 2) Using the mouse Rpl39/Rpl39l-expressing yeast as a source of purified Rpl39l-containing ribosomes we will study the structure of NPET using CryoEM. We previously tried to overexpress Rpl39l in mammalian cells through different promoters to achieve high Rpl39l levels, but we observed that pushing the proportion of Rpl39l-containing ribosomes in E14 cells or cancer cell lines beyond 10-15% is lethal to the cells. We have already reached the 6Å resolution for ribosome structure using purified Rpl39-ribosomes from the Rpl39l KO lines. But we could not get the structure of the Rpl39l-ribosome due to this enrichment challenge. Having the yeast lines will solve this problem of getting purified Rpl39l-ribosomes. The CryoEM study will help us understand how Rpl39l and the associated change in the amino acids exposed in the NPET changes the tunnel structure and behaviour. 3) Finally, we plan BS3 crosslinking of Rpl39l-expressing mouse ES line lysate to apply crosslinking peptide mass spectrometry for identifying the direct interactors of Rpl39l in the mammalian system. BS3 is a crosslinker that forms an NHS bond with the side chain of lysine at the range of ~1.6 Å (Schmidt and Robinson, 2014). We are expecting not only to crosslink the interactors to Rpl39l but also the nascent peptides. This will give us an insight into the exact proteins, if any, that Rpl39l ribosome helps translate specifically, and also a better understanding of the chaperone network interaction with Rpl39 and Rpl39l, something that is not yet studied for the mammalian system. The experiments are currently in progress and the results of these studies will be updated in the thesis as they are generated.

Additional work: Exploring the origin of the difference in innate immune responses between mouse strains

As already mentioned in the main introduction to this thesis, gene expression is a stochastic process. Single cell sequencing technologies have revolutionized the way gene expression is being analysed, and much effort is dedicated to characterizing the breadth and explore the consequences of variability in gene expression between cells of the same type. Having joined the Zavolan lab with a background in innate immune response (Guha et al., 2019; Pradhan et al., 2019; Sur et al., 2015) and given the lab's interest in single cell gene expression, I decided to propose addressing questions that piqued my interest as a M.Sc. student, namely 1) Why are mice of different strains prone to give generally stronger or weaker immune responses? 2) Why do the individuals of a macrophage population respond to a stimuli so heterogeneously?

Macrophages, one of the first line of defence in mammalian immune system, are one of the core members of the innate immune response (Espinosa and Rivera, 2016). The monocyte-macrophage lineage of cells is conserved across all vertebrates and is involved not only in immune response, but also wound healing, homeostasis and development (Lavin et al., 2015). Extremely plastic in terms of effector functions, various functionally distinct groups of macrophage populations are found throughout the body. A rough classification of macrophages has been made by analogy with that of the T lymphocytes, which also activate macrophages, into M1 (pro-inflammatory, or classically activated) and M2 (anti-inflammatory or alternatively activated). Macrophages stimulated with Interferon-γ, Lipopolysachharide (LPS) or Tumor Necrosis Factor (TNF) become M1 macrophage, while treatment with IL-1R, IL-4, IL-10, IL-13 etc leads to M2 polarization (Martinez and Gordon, 2014). Given the variability in function and activation states, it is still unclear, how far the current definition of macrophage functions holds true beyond Mammalia, though there is some evidence of functional diversification in ectothermic mammals (Edholm et al., 2017). While the function of macrophage adaptability

and variation has been studied, little is understood about the source of such variation (Ravasi et al., 2002). While studies comparing macrophage variability between twins have shown that there are non-genetic sources of variability in these cells, like environment and exposure history, little has been studied in the direction of intrinsic variability of these cells (Ingelfinger et al., 2022).

The paradigm of M1 and M2 polarization in macrophage, on the lines of Th1 and Th2 in T-cells, is a model to understand the functional and molecular states of a stimulated macrophage cells. M1 polarised macrophage, also called classically activated macrophage, is associated to pro-inflammatory stimulations and behaviour, normally seen in the context of bacterial and viral infections, marked by production of iNOS. The M2 polarization stands for anti-inflammatory type, associated to tissue residence, wound healing and developmental function. While M1 polarization is a rather simple model with a single mode of function, M2 has M2a, b, c and d for different activity modes marked by context-specific cytokine production (Chávez-Galán et al., 2015). While M1-M2 model started as a model assuming discrete states, studies have shown that M1 and M2 are rather two ends of a spectrum. Nevertheless, it is still thought that overall, the M1-M2 model provides a good framework to study the status and behaviour of macrophage cells (Orecchioni et al., 2019).

Balb/C and C57bl6 are two commonly used mouse strains for immunological studies, and they are Th2 and Th1 biased, respectively (Mills et al., 2000). On an innate immune system level, this boils down to the fact that the macrophage cells derived from the mice mentioned, have a corresponding bias in terms of M1 and M2 (Watanabe et al., 2004). When subjected to the exact same stimulus, the peritoneal macrophages from Balb/C have been shown to produce lower levels of pro-inflammatory cytokines like Tnfα and Il12, while the anti-inflammatory cytokine Il-12 is released at a higher level than the peritoneal macrophages of C57Bl6 (Watanabe et al., 2004). Similar phenotypic pattern has also been shown for resident macrophages from various

sites of the two mice (Taylor et al., 2008). In fact, this difference in macrophage behaviour from the two mice is wired on the level of arginine metabolism, where Balb/C macrophages are unable to efficiently channel the Arginine to produce nitric oxide species like the C57bl6 macrophages, instead producing higher levels of Citrulline, an Arginine metabolic pathway intermediate, necessary for wound healing (Mills et al., 2000, p. 1).

In this study, we sought to explore the seemingly ubiquitous observation of a macrophage cell population not responding evenly to a specific stimulus. We looked at peritoneal macrophage cells from mice with different immune background and used the power of single cell RNA transcriptomics to understand the intrinsic variability of macrophage gene expression in the absence of specific stimulation with antigens. My role in the project was to prepare the macrophages for single cell sequencing and help with the biological background of the analysis, while the data analysis itself was carried out by Dr.Jeremie Breda.

The manuscript describing the results is shown in Supplement 2.

Supplement 2: A novel approach to single cell analysis to reveal intrinsic differences in immune marker expression in unstimulated macrophages from BALB/c and C57Bl/6 mouse strains

Jeremie Breda, Arka Banerjee, Rajesh Jayachandran, Jean Pieters, Mihaela Zavolan

Introduction

The large heterogeneity in the response of macrophages to stimuli is well established [1,2]. Multiple techniques have been used to decipher the nature and cause of such heterogeneity and several hypotheses have emerged, though all invoke stochastic expression of individual genes across macrophages [3]. Some studies focused on the role of general factors such as the scaling of gene expression parameters with cell size [4], while others emphasized the role of cell surface receptors and interactions with the microenvironment in generating macrophage diversity [5]. The latter aspect is receiving renewed attention in the context of tumorigenesis [6]. As in other populations of cells or organisms, the diversity of macrophages is viewed as a ‘‘bet hedging’’ strategy [7,8] for the immune system, whereby a range of cellular phenotypes is induced upon immune activation, in anticipation of phenotypic variation of the pathogen. This strategy enables the immune system to rapidly respond to diverse types of rapidly evolving pathogens [9]. An interesting system for investigating the origin and potential implications of macrophage heterogeneity consists of the laboratory mouse strains C57BL/6 and BALB/c, which naturally differ in their innate immune response. C57BL/6 mice give prototypical Th1-biased immune responses, while BALB/c mice have Th2-biased immune responses [10,11]. It has further been shown that macrophages isolated from these mice respond by producing different levels of the microbicidal nitric oxide molecule [11] and with different dynamics to stimulation [12].

Stimulated macrophages are grouped into two major classes, classically activated (M1) and alternatively activated (M2) macrophages, along the lines of T-cell Th1 and Th2 respectively. While the classically activated macrophage is associated with a pro-inflammatory immune response, the alternatively activated immune response is associated with an anti-inflammatory-like immune response [13]. Genes like Nos2 and Tnfa are considered markers of M1 polarization, while Arg1 and Tgfb are markers of M2 polarization [14]. Increasingly many studies are shedding light on the nature of M1 and M2 polarization, and whether they represent two ends of a spectrum of cell states [15]. For example, previous investigations of the variability of unstimulated macrophages reported differences in the expression of M1 and M2 marker genes, even though the studies were done on the RAW264.7 cells line [2], which have been shown to have the least variation of all sources of macrophage [16]. While definitely not exhaustive, the M1-M2 paradigm provides a good framework to analyze behavior, response and activity in macrophage cells from unstimulated mice.

As methods like immunocytochemistry or FACS are restricted to relatively few cells or cell types and few genes at a time [17], more comprehensive approaches, based on single cell transcriptome sequencing, could provide new insights into the intrinsic heterogeneity of macrophage populations in different mouse strains. Single cell sequencing (scRNA-seq) enables the estimation of expression levels for several thousands of genes in thousands of individual cells at a time. This, in turn, enables the identification of novel cell types, reconstruction of differentiation trajectories, and elucidation of the dynamics of responses to various perturbations [18]. Interestingly, although a number of single cell studies have been carried out in mouse model systems (e.g. [19]), a comparison of the unstimulated macrophage populations of C57BL/6 and BALB/c strains has not been done. Here we applied scRNA-seq to unstimulated peritoneal macrophages obtained from specific-pathogen-free mice, with the aim of determining whether differences are already detectable in pre-stimulation macrophage

populations from these mouse strains and whether they can explain some of the response variability that is observed upon stimulation.

Results

Gene expression profiles identify immune cell types in the mouse peritoneum

To gain insight into the cell types that populate the peritoneum of unstimulated mice, we prepared and sequenced cells from the peritoneal lavage of 3 BALB/c mice, aiming to obtain two technical replicates from each mouse by splitting the isolated cells into two distinct pools before single cell sample preparation. One of the technical replicates for mouse #3 yielded very few reads and was discarded. For all other replicates, we estimated the relative abundances of transcripts from all genes in every cell with a method that we have recently developed [20] and then projected these data on the first two principal components. As shown in Fig 1A, for all mice and all replicates, cells separated along the two first principal components, with one large subpopulation distributed along PC1, another one along PC2 and a smaller subpopulation of cells at intermediate PC1 coordinates. We assigned a cell type to each cell hierarchically, starting from B cells, then T cells and RBCs and finally macrophages (see Methods). Determining the average expression of markers typically associated with macrophages (M), B/T cells and erythrocytes (RBC) (see Table S1), as well as for other hematopoietic cell types (not shown) in each sample (individual marker distributions shown in Fig. S1), we found that indeed, B/T cell and RBC markers clearly identified specific subpopulations with high expression (Fig S2A-B). Macrophages and B cells were the most abundant cell types in the peritoneal lavage of unstimulated mice (Fig. S2C).

We then carried out the same experiment in two distinct C57Bl/6 mice. Applying the same analysis, we identified similar proportions of cells, macrophages and B cells being the two main cell types in the peritoneal lavage. The relative distribution of the cell types in the PC1-PC2 coordinate system appeared similar between the two mouse strains (Fig. 1A vs. 1C), and the distinction between the two main cell subpopulations was even clearer than for the BALB/c mice, probably due to the higher sequencing depth achieved in this experiment (Table S2). We then assessed quantitatively whether the relative proportion of cell types was similar between individual mice. Indeed, we found that within each strain, macrophages and B cells originated proportionally from each mouse, whereas the very rare RBCs and T cells came primarily from individual animals (Fig. S3). Altogether, these data indicate that scRNA-seq identifies similar cell populations in the peritoneal lavage of both BALB/c and C57Bl/6 mice.

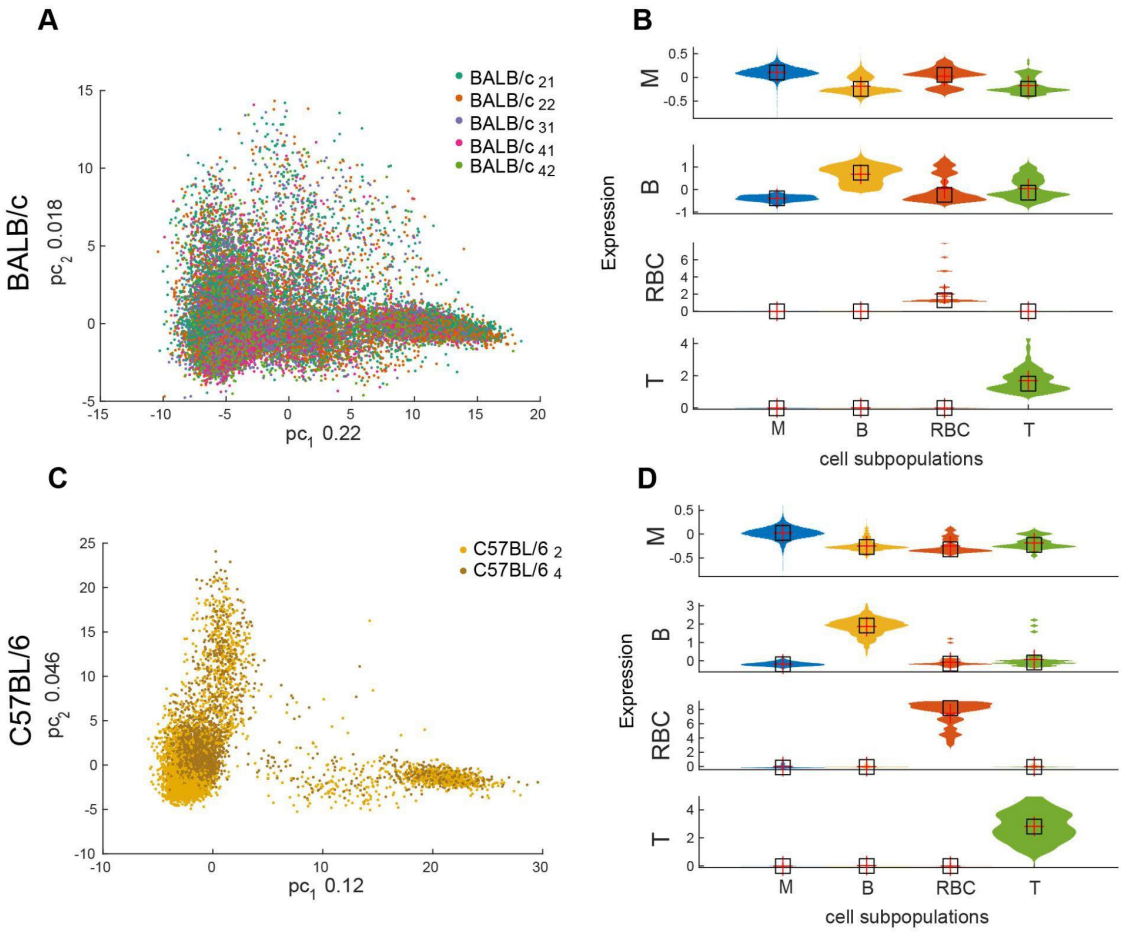


Figure 1. The main cell populations identified in the peritoneal lavage of mice. A. Projection of single cells from 3 BALB/c mice (1-2 technical replicates each) on the first two principal components of gene expression. The fraction of variance explained by each component is indicated on the axes. Individual samples are indicated by color. **B.** Violin plots showing the distribution of the average expression levels of markers of different cell types in cells from all samples from panel A. **C-D.** Similar to A-B, but for the 2 C57bl/6 mice.

The genes that vary most across macrophages are immune response-related.

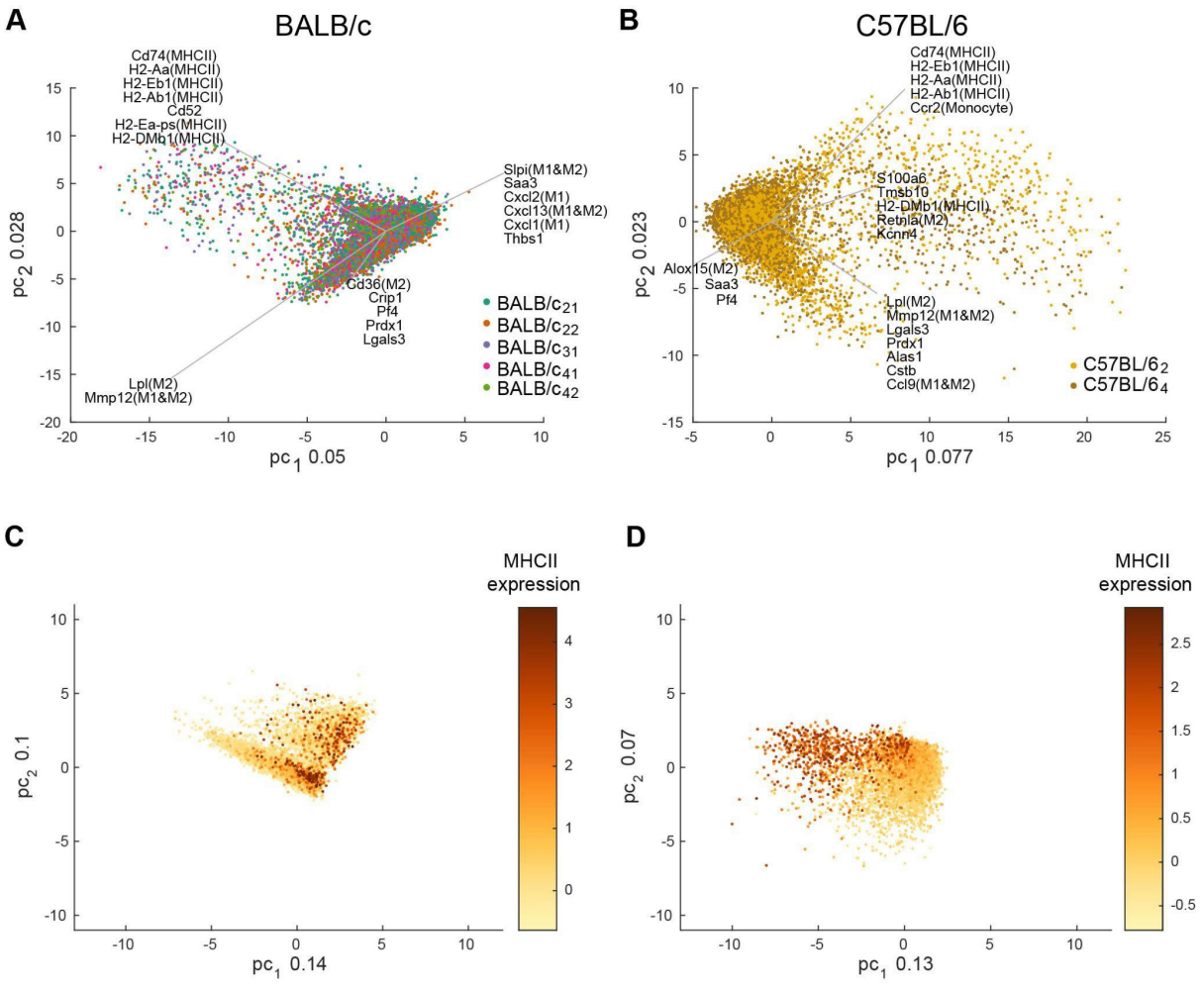


Figure 2. Summary of gene expression in mouse peritoneal macrophages. A-B. Projection of single macrophages identified in the BALB/c (A) and C57bl/6 (B) mice on the first two principal components of gene expression. The gene expression data was centered within each

sample before carrying out the analysis. Individual samples are indicated by color. The genes with highest expression variability in the data set are also indicated, along with their direction of variation (see Methods). **C-D.** Similar as in A-B, but in the gene expression space of known M1/M2 markers. Cells are colored by the average expression of MHC class II genes.

We next focused on the macrophage subpopulation in the two mouse strains. To avoid 'batch effects' due to technical factors we centered the data from all mice of a given strain before analyzing the variability in gene expression in the dataset (Fig. 2A-B). Strikingly, in both strains, the genes contributing most to the two main directions of variation in expression across macrophages (Table S3) came from the major histocompatibility II complex or were previously associated with the M1/M2 polarization [15]. To investigate further the significance of this observation, we carried out principal component analysis using only the expression levels of these genes (Table S4). As shown in Fig. 2C-D, in both mouse strains, the expression of M1 and M2 marker defines two orthogonal axes of variation, emanating from a common center of 'amorphous' gene expression. Further marking the expression level of MHCII molecules in individual cells reveals a distinct subpopulation of MHCII-expressing cells, but these are not specifically associated with one of the main axes of M1-M2 marker expression in Fig. 2C-D. As it has been reported that the MHCII expression level in macrophages is cell cycle-dependent [21], we determined the correlation between the average expression of cell cycle genes and the average MHCII gene expression across cells finding it to be higher (Pearson r = 0.26, p-value = 10^{-300} in BALB/c and Pearson r = 0.4, p-value = 10^{-300} in C57Bl/6 mice) than the correlation with the average expression of M1/M2 markers (correlation between cell cycle genes and M1 marker genes: Pearson r = -0.0051, p-value = 0.46 in BALB/c and Pearson r = 0.19, p-value = 9.7 10^{-81} in C57Bl/6 mice; correlation between cell cycle genes and M2 marker genes: Pearson r = 0.03, p-value = 2 10^{-5} in BALB/c and Pearson r = 0.16, p-value = 1.1 10^{-59} in C57Bl/6 mice).

Thus, the expression of immune markers is highly variable in unstimulated macrophages from both C57BL/6 and BALB/c strains.

A 'priming' signature can be identified in macrophages from unstimulated mice

To determine whether these axes of gene expression variation are related to the M1/M2 polarization observed upon stimulation with immunological stimuli, we analyzed our data along with a previously published bulk sequencing data set of macrophages that were stimulated *in vitro* with lipopolysaccharide (LPS) and interferon γ (IFNγ), stimuli that induce M1 polarization, or IL-4, which induces M2a polarization [22]. We centered the bulk gene expression data sets on the unstimulated samples and carried out principal component analysis to identify the first two principal axes of variation in the bulk data. We then determined the distribution of each bulk sample as well as each single cell in our data in the space defined by PC1 and PC2 of the bulk data (see Methods). Bulk samples were distributed along a curved trajectory that traced well the time after stimulation. Whereas by construction most of the single cells were located at the center, along with the unstimulated bulk samples, a distinct subpopulation of cells was distributed along the axis of variation corresponding to 4-8 hours after *in vitro* stimulation with LPS+IFNγ (Fig. 3A-B). These results indicate that unstimulated macrophage populations contain cells that with an expression profile that is similar to that observed observed after 4-8 hours of LPS+IFNγ stimulation *in vitro*.

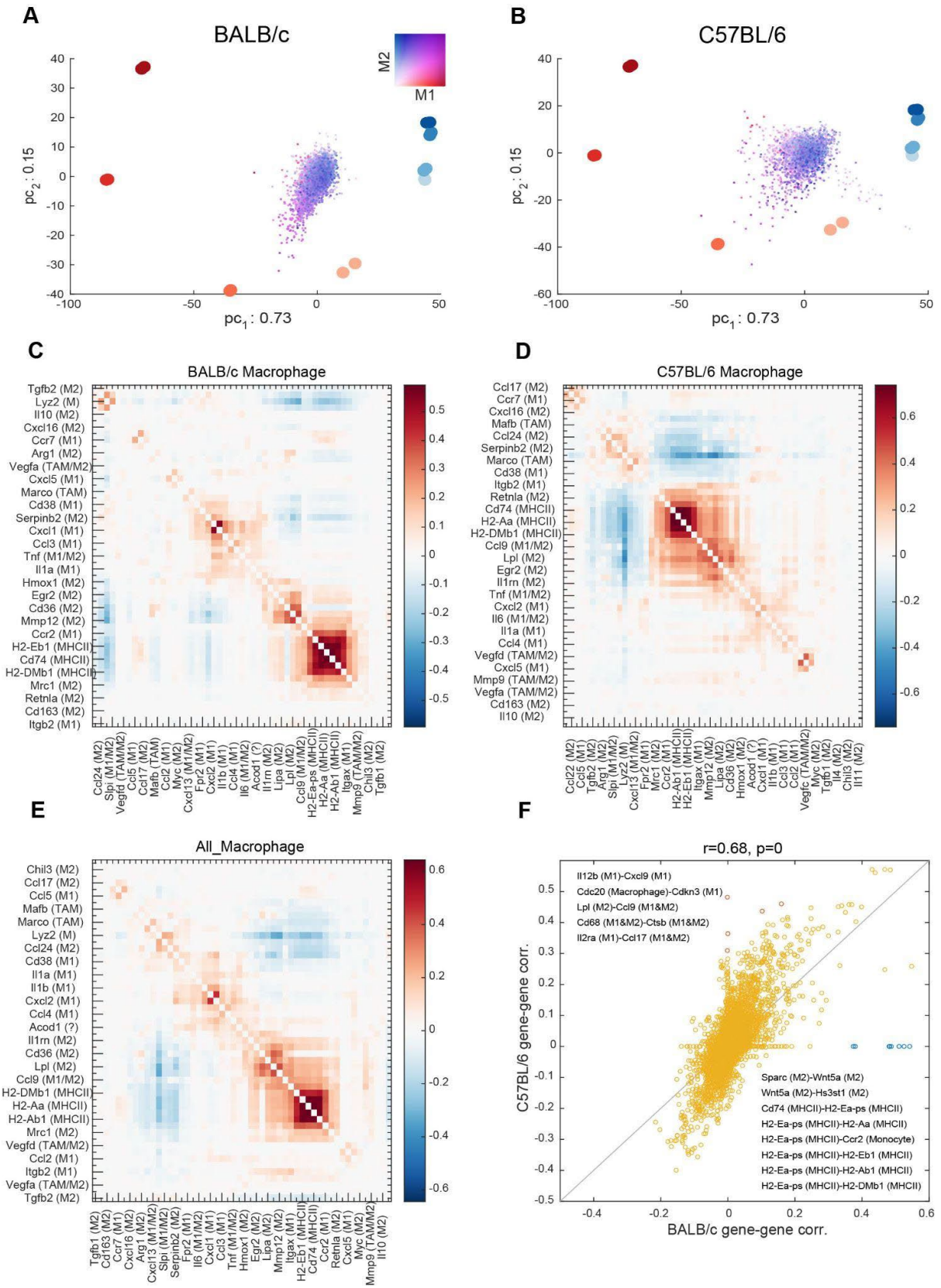
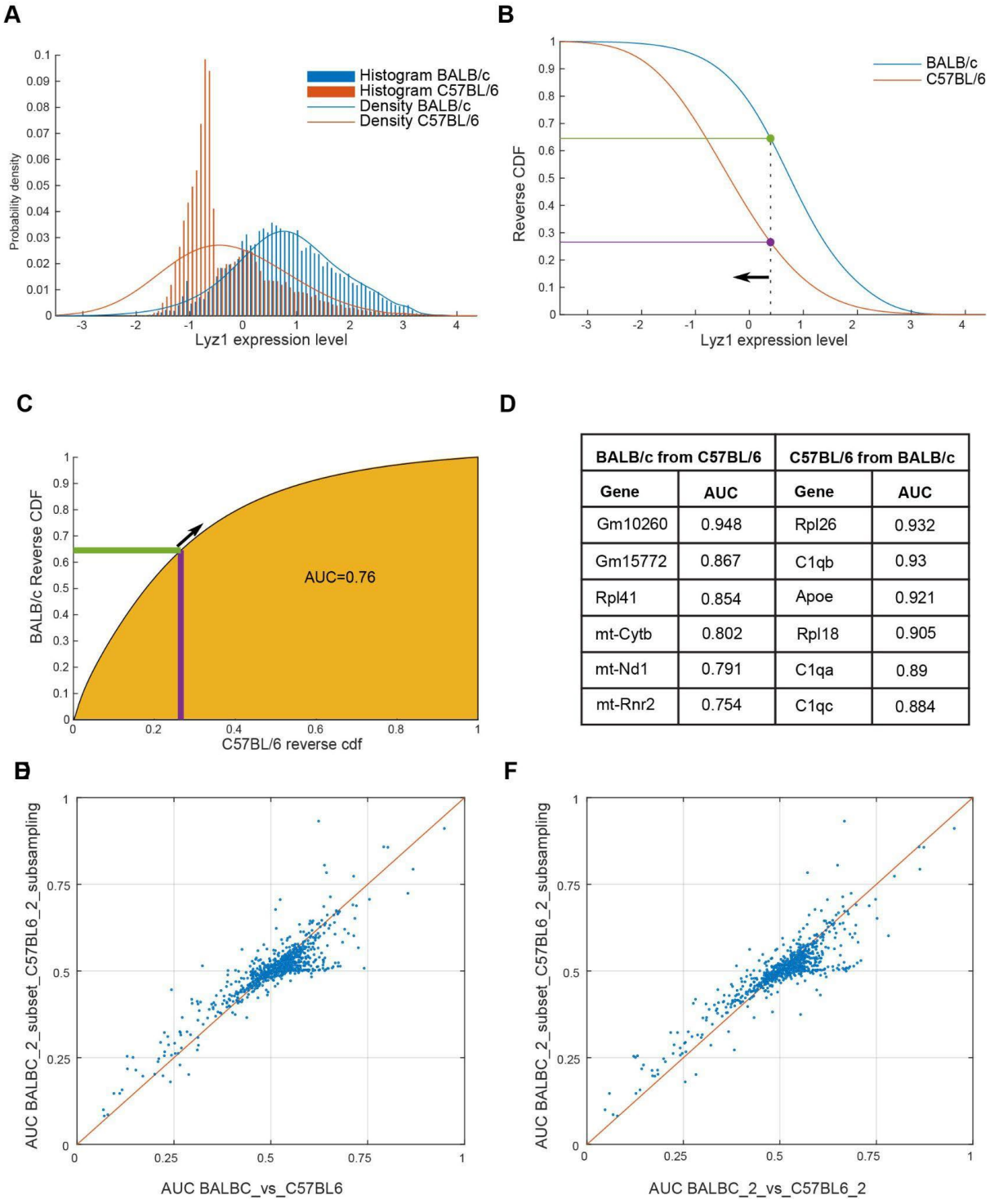

A
BALB/c
M2
M1
pc2 : 0.15
pc1 : 0.73
B
C57BL/6
pc2 : 0.15
pc1 : 0.73
C
BALB/c Macrophage
Tgfb2 (M2)
Lyz2 (M)
Il10 (M2)
Cxcl16 (M2)
Ccr7 (M1)
Arg1 (M2)
Vegfa (TAM/M2)
Cxcl5 (M1)
Marco (TAM)
Cd38 (M1)
Serpinb2 (M2)
Cxcl1 (M1)
Ccl3 (M1)
Tnf (M1/M2)
Il1a (M1)
Hmox1 (M2)
Egr2 (M2)
Cd36 (M2)
Mmp12 (M2)
Ccr2 (M1)
H2-Eb1 (MHCII)
Cd74 (MHCII)
H2-DMb1 (MHCII)
Mrc1 (M2)
Retnla (M2)
Cd163 (M2)
Itgb2 (M1)
D
C57BL/6 Macrophage
Ccl17 (M2)
Ccr7 (M1)
Cxcl16 (M2)
Mafb (TAM)
Ccl24 (M2)
Serpinb2 (M2)
Marco (TAM)
Cd38 (M1)
Itgb2 (M1)
Retnla (M2)
Cd74 (MHCII)
H2-Aa (MHCII)
H2-DMb1 (MHCII)
Ccl9 (M1/M2)
Lpl (M2)
Egr2 (M2)
Il1rn (M2)
Tnf (M1/M2)
Cxcl2 (M1)
Il6 (M1/M2)
Il1a (M1)
Ccl4 (M1)
Vegfd (TAM/M2)
Cxcl5 (M1)
Mmp9 (TAM/M2)
Vegfa (TAM/M2)
Cd163 (M2)
Il10 (M2)
E
All_Macrophage
Chil3 (M2)
Ccl17 (M2)
Ccl5 (M1)
Mafb (TAM)
Marco (TAM)
Lyz2 (M)
Ccl24 (M2)
Cd38 (M1)
Il1a (M1)
Il1b (M1)
Cxcl2 (M1)
Ccl4 (M1)
Acod1 (?)
Il1rn (M2)
Cd36 (M2)
Lpl (M2)
Ccl9 (M1/M2)
H2-DMb1 (MHCII)
H2-Aa (MHCII)
H2-Ab1 (MHCII)
Mrc1 (M2)
Vegfd (TAM/M2)
Ccl2 (M1)
Itgb2 (M1)
Vegfa (TAM/M2)
Tgfb2 (M2)
F
r=0.68, p=0
Il12b (M1)-Cxcl9 (M1)
Cdc20 (Macrophage)-Cdkn3 (M1)
Lpl (M2)-Ccl9 (M1&M2)
Cd68 (M1&M2)-Ctsb (M1&M2)
Il2ra (M1)-Ccl17 (M1&M2)
Sparc (M2)-Wnt5a (M2)
Wnt5a (M2)-Hs3st1 (M2)
Cd74 (MHCII)-H2-Ea-ps (MHCII)
H2-Ea-ps (MHCII)-H2-Aa (MHCII)
H2-Ea-ps (MHCII)-Ccr2 (Monocyte)
H2-Ea-ps (MHCII)-H2-Eb1 (MHCII)
H2-Ea-ps (MHCII)-H2-Ab1 (MHCII)
H2-Ea-ps (MHCII)-H2-DMb1 (MHCII)
C57BL/6 gene-gene corr.
BALB/c gene-gene corr.

Figure 3. Analysis of M1/M2 bias in unstimulated mouse peritoneal macrophages. A-B. Projection of single macrophages from BALB/c (A) and C57bl/6 (B) mice in the gene expression space of bulk RNA seq samples from macrophages stimulated in vitro with (red), known to induce M1 bias or , known to induce M2 bias (blue). For the bulk data, the shade indicates the stimulation time, 2, 4, hours. 2 replicates of fold-changes with respect to time 0 (unstimulated cells) were available for each time point. The centered single cell data were anchored to the center of the bulk data, which corresponds to the gene expression state of unstimulated macrophages. The color of individual single cells indicates the relative expression of M1-M2 markers, as indicated by the 2-dimensional color legend. **C-D.** Pairwise correlation of M1/M2/MHC class II gene expression levels across individual macrophages from BALB/c (C) and C57bl/6 (D) mice. **E.** Similar, but the analysis was carried out across all cells, from both mouse strains. **F.** Scatter plot of pairwise correlation coefficients of M1/M2/MHC class II gene expression levels between the two mouse strains.

An immediate question is whether the M1 markers are expressed in individual unstimulated cells in a coordinated manner, under the control of a common regulator, or rather independently of each other, and thus as a result of stochastic transcriptional activation of individual marker genes. To answer this question we constructed the matrix of pairwise correlations of marker expression (Fig. 3C-E), finding that blocks of stably correlated expression emerge. The highest correlations are within the MHC class II molecules, and also between some of the M1 markers (Cxcl1-Cxcl2, Il2ra-Ccl17), as would be expected from macrophage cells with some degree of polarization [15]. Interestingly, some M2 (e.g. Tgfbi, Lpl, Lipa, and Trem2) and M1 (e.g. Atf3, Bcl2a1d, Bcl2a1a and Psme2)-associated genes are also correlated in expression with MHCII molecules, while a set of M2 markers (Itgam, Fn1, Alox15, Ccl24 and Cd14) are anti-correlated in expression with both MHC class II molecules and the M1 markers (Fig. 3E). MHCII

molecules have been previously found on M2 macrophage cells, although below the threshold required for effective antigen presentation [23]. Interestingly, the pairwise correlations in marker expression are similar between mice (Pearson r = 0.68, p-value < 10^-300, Fig. 3F). Only a few relationships are strain-specific, indicated with distinct colors in Fig. 3F. Correlations specific for the BALB/c strain involve primarily MHCII molecules, but also Wnt5a, a ligand that promotes the containment of bacterial infections by macrophages [24]. In contrast, molecules whose expression correlates more in the C57Bl/6 compared to the BALB/c mice, correspond to chemokines surface receptors and enzymes (Fig. 3F) that modulate T cell immunity (e.g. [25–27]). These data indicate that M1/M2 markers are expressed in a coordinated manner in the macrophages from both mouse strains, indicating that these correlations are induced by upstream regulators.

Evidence of stronger polarization in unstimulated macrophages from C57BL/6 compared to BALB/c mice



BALB/c from C57BL/6		C57BL/6 from BALB/c	
Gene	**AUC**	**Gene**	**AUC**
Gm10260	0.948	Rpl26	0.932
Gm15772	0.867	C1qb	0.93
Rpl41	0.854	Apoe	0.921
mt-Cytb	0.802	Rpl18	0.905
mt-Nd1	0.791	C1qa	0.89
mt-Rnr2	0.754	C1qc	0.884

Figure 4. Strain-specificity of macrophage gene expression. A. Example inference of the gene expression level distribution across cells from individual mouse strains, taking into account the uncertainty associated with the observed read counts in single cells. The histograms show relative transcription activities [20] corresponding to inferred expression levels in individual cells, while the curves show the probability densities of relative transcription activities, calculated by taking into account the uncertainty in the point estimates of gene expression in individual cells. **B.** Reverse cumulative distributions of the same gene's expression across cells from BALB/c (blue) and C57bl/6 (red) mice. The proportion of cells from BALB/c mice that have an expression level at least as high as the value indicated by the dashed vertical line is shown with the green and purple lines for the two strains, respectively. **C.** Inference of the AUC measure of strain-specificity of expression. The cumulative distributions shown in panel B are traversed from left to right, recording the proportion of cells from either strain whose expression level of a given gene is at least as high as indicated on the x-axis. This results in the indicated curve, the area under which is then calculated. **D.** Genes with the highest specificity of expression in BALB/c and C57bl/6 are indicated.

We next asked whether there are genes that distinguish unstimulated macrophages from the two mouse strains, that is, whether there are genes with systematically higher expression in macrophages from one or the other mouse strain. To answer this question, we first constructed for each gene, the histogram of expression levels in macrophages from each mouse strain and the probability distribution function (Fig. 4A), taking into account the Sanity-provided error bars (from the normalization method Sanity [20] (see Methods)). This is crucial as the uncertainty of gene expression level estimates strongly depends on the expression levels themselves. Next, we constructed the reverse cumulative distributions of the gene's expression

level across cells from each mouse strain (Fig. 4B). For each expression value *x*, these distributions give the probability that a macrophage from a given strain expresses the gene of interest at a level at least as high as *x*. Traversing these two distributions from high to low expression levels *z*, we recorded the probability that macrophages from the two mouse strains express at a level at least as high as *z*. These points define the curve shown in Fig. 4C, with the values from C57Bl/6 mice on the x-axis and those from the BALB/c mice on the y-axis. A high area under this curve (AUC) indicates that a higher proportion of cells from the BALB/c mouse has high expression level of the gene of interest and thus the gene is more 'specific' for BALB/c macrophages. Conversely, low AUC indicates that higher proportions of cells from the C57Bl/6 mouse have high expression levels of the gene of interest and thus the gene is more 'specific' for C57Bl/6 macrophages. This analysis showed that most of the genes that are 'specific' to the C57Bl/6 mouse are immune response genes, while the few genes specific to the BALB/c mouse encode ribosomal and mitochondrial proteins (Table S5).

To ensure that these results are not due to differences in the number of cells or the number of reads per cell obtained from different strains, we carried out a randomization test. Specifically, we took the data from BALB/c mouse #2 with 7919 cells and an average of 11967 reads per cell and the data from C57Bl/6 mouse #2 with 6134 cells and 37653 reads per cell and subsampled them, so that we obtain the same number of cells (6134) and a similar average number of reads per cell (11967 and 13029, for BALB/c and C57bl6, respectively) from both mice. We then repeated the AUC analysis and the results remained unchanged (Fig. 4D-E), indicating that the specificity of gene expression that we inferred above is not due to differences in cell or mRNA capture between experiments.

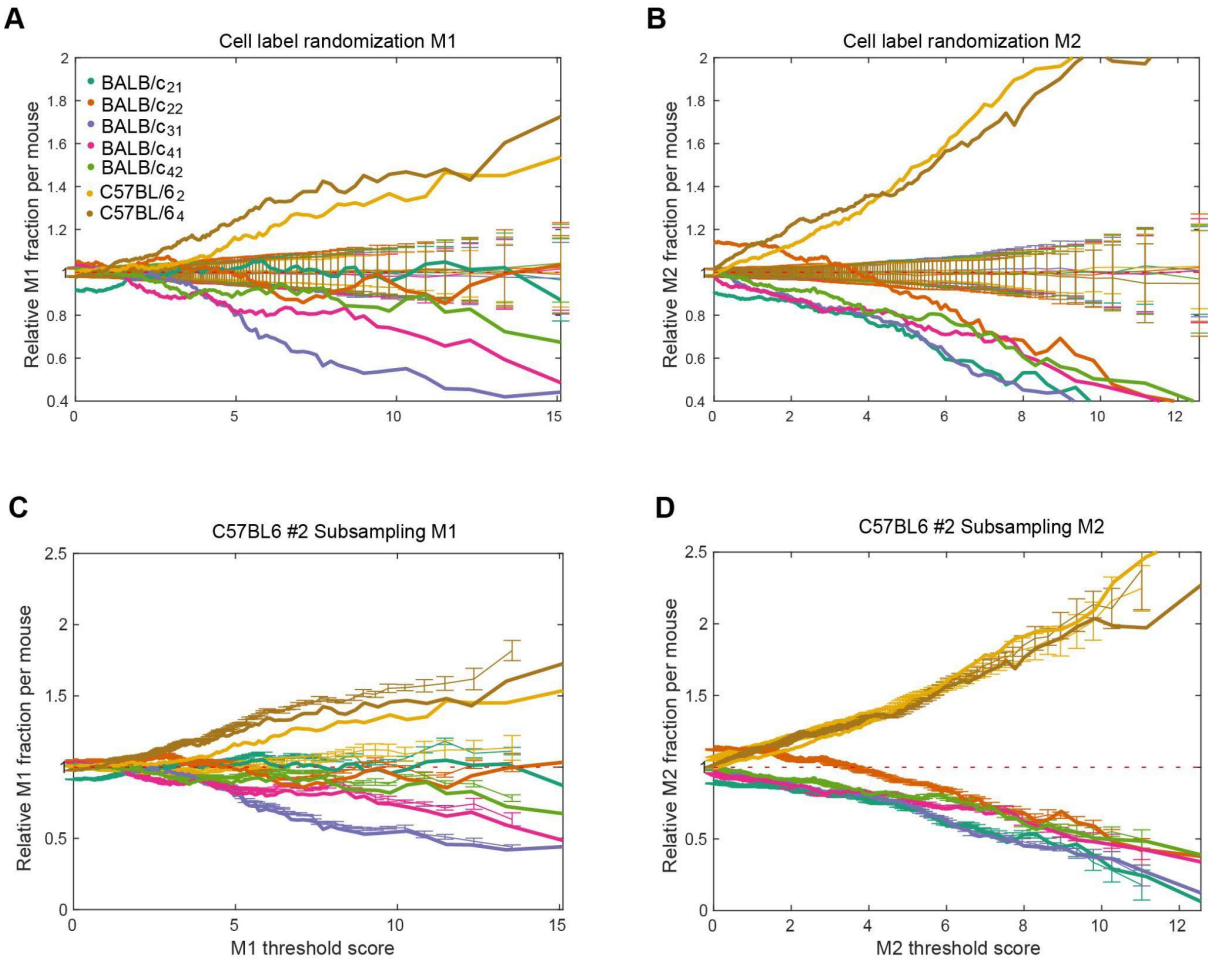


Figure 5. M1/M2 bias in unstimulated macrophages from BALB/c and C57bl/6 mice. A. M1 marker gene expression in samples from the two mouse strains. Cells from all samples were sorted by the average expression of M1 markers and then the expression range was traversed from lowest to highest, to calculate the ratio between the proportion of cells from a given sample whose expression was at least as high as the value indicated on the x-axis and the proportion of cells from that sample in the entire data set. Individual samples (thick lines) are indicated by color. For comparison, cells labels were permuted 10 times, the same analysis was repeated, and the mean and standard deviation of these results are shown in thin lines. **B.** Similar analysis for M2 marker gene expression. **C-D.** Effect of sequencing depth on the inferred degree of M1 (C) / M2 (D) bias in unstimulated macrophages. The analysis described for panel A was carried out for 10 subsamplings of the reads from the C57bl/6 mouse #2. These results are shown in thin lines with error bars, along the results on the original data, which are shown in thick lines.

The striking specificity of expression of immune genes in the C57bl/6 mice prompted us to investigate further if the degree of macrophage polarization may be higher in this strain. To answer this question, we calculated the total expression of M1 and M2 markers in each cell from each of the strain. We then traversed the expression range of M1 and M2 markers from lowest to highest, and for each expression level x, we calculated the ratio between the proportion of cells from each mouse with an expression level at least as high as x and the proportion of cells from that mouse in the entire data set. Surprisingly, we found that both C57bl/6 mice showed an enrichment of cells with high marker expression, and this was the case for both M1 and M2 markers (Fig. 5A-B). To evaluate the significance of these enrichments we randomly permuted the labels of the cells and repeated the analysis, finding that the enrichment signal disappears. Thus cells from the C57bl/6 mice do indeed show higher expression of both M1/M2 markers. Then, to ensure that the enrichments are not due to differences in sequencing depth and cell numbers in the two experiments, we subsampled reads from the C57bl/6 mouse #2, to enforce an average number of reads per cell similar to those in BALB/c mice. We did not subsample the C57bl/6 mouse #4, for comparison. Repeating the subsampling 10 times, we obtained the results shown in Fig. 5C-D. Namely, the subsampling strongly reduced the enrichment of cells with high M1 marker expression, but did not affect the enrichment of cells with high M2 marker expression in the C57bl/6 mouse. These results demonstrate that the peritoneal macrophages of C57bl/6 mice exhibit evidence of polarization especially towards the M2 state, in the absence of specific stimulation.

Discussion

For reasons that remain poorly understood, individuals in a population, as well as cells within an organism are not equally susceptible to infection. Macrophages are a first line of defense against various pathogens and for this reason, many studies have attempted to identify macrophage-dependent factors that can explain the variability in the response to infections. Distributed throughout the body, macrophages have different origins and functions [5]. Although simplistic, a widely-used classification has been into M1 and M2 types, also known as classically and alternatively activated macrophages [28]. Classically activated macrophages are induced by interferon gamma produced by Th1 T cells, and are involved in the response to intracellular infection such as, for example, with the Mycobacterium tuberculosis. In contrast, alternative macrophage activation by the interleukins IL4 and IL13 produced by Th2 T cells leads to the M2 type, associated with allergies, parasitic infection and wound healing phenotype [5]. Interestingly, the laboratory mouse strains C57bl/6 and BALB/c naturally differ in their innate immune response, with C57bl/6 mice giving prototypical Th1-biased immune responses and BALB/c mice giving Th2-biased immune responses [10,11]. Consequently, these mice have different susceptibility to pathogens such as Leishmania [29].

To gain a better understanding of the intrinsic differences in the innate immune cells of these mouse strains, we have carried out single cell sequencing of peritoneal macrophages from specific pathogen-free mice. Our main results are as follows. First, by comparing single cell gene expression with the gene expression of populations of macrophages stimulated with either canonical or alternatively-activating stimuli, we found that the macrophages from specific pathogen-free animals show some degree of polarization, roughly corresponding to 4-8 hours of *in vitro* stimulation. This provides new insights into the degree of macrophage heterogeneity in mice that are not exposed to specific pathogens. However, identifying raw datasets from *in vitro* stimulation experiments has been surprisingly difficult, and we hope that more will be

available in the future. This will be important for understanding the behavior of macrophages in different contexts, including the much-studied tissue-resident macrophages. If fine-grained timelines of macrophage response to various stimuli were established, they would allow much finer-grained analyses of the heterogeneity and 'distribution of labor' among macrophages from different sites, thus contributing to an improved understanding of immune responses and susceptibility to infections.

Furthermore, we found that the apparent polarization was due to the coherent expression of a subset, but not all, of the M1 and M2 markers. Most unexpected however was the finding that a higher proportion of cells from the C57Bl/6 mouse showed increased marker expression, especially of the M2 type, which could not be explained by technical factors such as differences in the number of sampled cells or reads per cell. This suggests that the pro-inflammatory bias of C57bl/6 mice is associated with an intrinsically higher activation of immune response genes in baseline conditions. Indeed, asking which genes can distinguish the macrophages from the two mouse strains we found that higher expression of immune response identifies macrophages from the C57Bl/6, while higher expression of ribosomal and mitochondrial genes identifies macrophages from the BALB/c mouse. Analysis of single cell sequencing data poses numerous challenges [30]. Although many computational tools have been developed, methods that are anchored in the biological and biophysical nature of the data are slow to develop. However, understanding the and appropriately modeling the sources of variability in gene expression levels provides increased precision in identifying genes with variable expression between cells and reconstruction differentiation trajectories [20]. Our analysis took advantage of the estimates of uncertainty in gene expression levels to identify genes whose expression distinguishes macrophages from the different mouse strains. This method can be used to define markers in other contexts. This will be important for future studies because markers have generally been defined from among cell surface proteins, and the protein level gene expression correlates

relatively poorly with the mRNA level gene expression, especially at the level of single cells. In conclusion, our results provide a baseline for understanding differences in the immune response of different mouse strains, as well as a computational framework for uncovering such differences.

Methods

Sample preparation

For extraction of mouse peritoneal macrophage, we followed the protocol detailed by (PMID: **20110936**). A total of 3 mice from each mouse type (Balb/c and C57Bl6) were used for the study. In short, the mice were euthanized using CO2 as per the protocol validated by the ethics committee. The abdomen of the mouse was cleaned with 70% ethanol. A small incision was made at the outer skin of the peritoneum and the inner skin lining of the peritoneum was exposed. 5ml of ice-cold 3% fetal calf serum (FCS) in 1X PBS was injected into the cavity using a 27G needle. The peritoneum was gently massaged to dislodge the immune cells in the cavity. A 25G needle was used to collect the fluid and was placed in a sterile ice-cold tube on ice. The peritoneal lavage was centrifuged at 200g for 5 min, and resuspended in RPMI-1640 medium supplemented with 10% heat inactivated FCS. The cells were plated in a 10 cm plate in 37 degree celsius and 5% CO2. The plate washed 3 times after 6 hours with 1X PBS supplemented with 3% FCS to remove the floating cells. The adherent cells were scraped and used for 10X Genomics Single cell transcriptomics library preparation according to the manufacturer's protocol.

Reads mapping and demultiplexing

The sequencing reads have been mapped and demultiplexed using *cellranger-2.1.0* (https://github.com/10XGenomics/cellranger) on the *Ensembl* genome assembly and gene annotation GRCm38, release 91 (http://ftp.ensembl.org/pub/release-91/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.primary_assembly.fa.gz and http://ftp.ensembl.org/pub/release-91/gtf/mus_musculus/Mus_musculus.GRCm38.91.gtf.gz)

Data normalization and analysis

The count table obtained from cellranger was normalized using the Sanity 1.0 (https://github.com/jmbreda/Sanity)[20]. The downstream analysis of the data was done using Matlab R2019b (9.7.0.1190202). The vectors of transcription activities for all genes across the cells of the different mice were initially analyzed together for all samples from a given strain. This analysis showed systematic differences in the mean activity between mice. To correct these batch effects, we centered the data within a strain, first obtaining the log2-fold differences in the transcription activity of each gene in each cell of a particular sample, and then concatenating these vectors over samples from the same strain.

Marker-based cell type identification

The identity of each cell was assigned as follows. We explored the expression of markers for the common hematopoietic cells (erythrocytes, macrophages, granulocytes, B and T cells, NK cells, dendritic cells, endothelial and epithelial cells, platelets, hematopoietic stem cells) compiled from the literature and found expression for only B and T cells, macrophage and erythrocyte markers. As many of the macrophage markers seemed to be less specific than the markers for these other cell types (Fig. S1A-B) used a hierarchical approach to assign a type to each of the cells. Specifically, we started with the B cells, for which the markers had

multimodal distribution (Fig. S1B, Fig. S2A). We decomposed the average B cell marker distribution in three gaussian components, and assigned the B cell type to the cells in the gaussian component with the highest mean, using as expression threshold the value where a cell had equal probability of being part of the highest and second highest components. We then took the remaining cells and constructed a 2-dimensional scatter of T cell and erythrocyte marker expression as shown in Fig. S2B. Only very few cells had some expression of markers for these two cell types, and we chose thresholds of expression that uniquely identified cells with highest marker expression. As these cell types are extremely rare in our samples, different choices for these thresholds do not affect our results. The remaining cells were assigned the macrophage cell type.

Identification of most variable genes

To identify the genes most responsible for the variation among macrophages we used the first two components of the gene expression data in single macrophages and multiplied the squared projection of each gene on each component by the fraction of variance captured by the corresponding component. We then summed the contributions of each gene to the two components.

Pairwise correlations of marker genes

The correlations shown in Fig. 3C-E are Pearson correlation coefficients of the log transcription activity (see Data normalization and analysis) of each pair of macrophage subtypes marker genes (see table S4).

Randomization of cell labels

We use the summed transcription activity of the M1/M2 markers as proxy for the respective polarization states. To assess the significance of the apparent enrichment of M1/M2-polarized cells among the cells from C57BL6 mice, we performed 100 random permutations of the cell labels. For each permutation, we computed the proportion of polarized cells coming from each mouse relative to the proportion of the different mice in the whole dataset. A cell was called M1/M2-polarized if the summed transcription activity of the respective markers was above a given threshold. We varied the threshold from 0 to the maximum summed transcription activities of markers across all cells, binning the interval of expression into 200 bins. Thus, at a threshold of 0 marker expression, the proportion of cells from a given mouse is that in the entire set of cells, and, as the threshold of maker expression increases, we observe an increase or decrease in this relative proportion depending on which strain the cells with the highest marker expression come from (see Fig 5A-B).

Subsampling of reads and cells

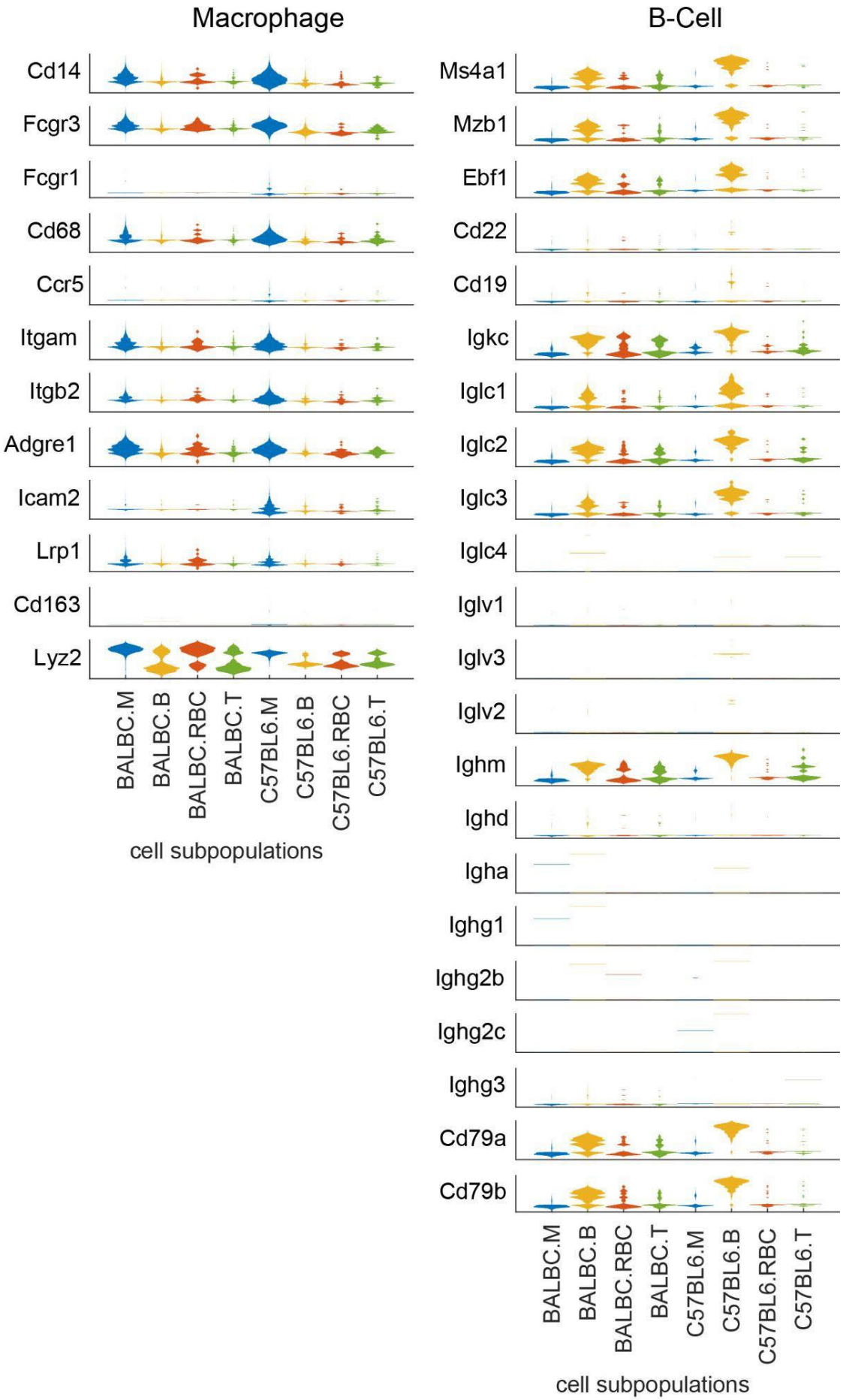
We performed a subsampling of the reads from sample C57BL6_2 in order to match the lower reads per cell in the BALB/c samples. The average number of reads per cell being 37653 in C57BL6_2 and 12375.2 in the BALB/c samples, we randomly selected a fraction 12375.2/37568 = 0.329 of the C57BL6_2 reads directly on the fastq file. We performed 9 independent sub-samplings. To establish that the strain-specific genes that we identified are not due to differences in the experimental procedure for sample preparation, we subsequently performed a subsampling of the BALBC_2 cells so as to match the lower total number of cells captured from the C57BL6_2 sample.

Analysis of bulk RNA-seq data from polarized macrophages

Processed data from bulk RNA-seq experiments of bone marrow macrophages stimulated *in vitro* with either LPS and IFNγ(to induce M1 polarization) or IL-4 (to induce M2a polarization) for 1, 2, 4, and 8 hours was obtained from [22], (https://www.ebi.ac.uk/biostudies/files/S-EPMC5159803/srep37446-s2.xls sheet S6.DEgene_expression). The data consisted in expression values of ~16'532 genes whose expression differed significantly in at least one of the time points after either of the stimuli relative to that at time 0. More precisely, the table contained fold-changes of these genes at all time points relative to time 0. We centered the data for each gene and projected the samples on the first 2 principal components. We then took the single cell gene expression matrix from our experiments, centered these values across cells, and projected them on the PC1-PC2 space from the bulk sequencing, applying a scaling factor of 10, to improve visualization. We colored individual points (samples in the case of bulk sequencing and single cells in the case of our data) by the average expression of M1 and M2 marker genes (Table S4), normalized to the min-max expression interval.

Supplementary Figures

A


Macrophage
B-Cell
Cd14
Fcgr3
Fcgr1
Cd68
Ccr5
Itgam
Itgb2
Adgre1
Icam2
Lrp1
Cd163
Lyz2
Ms4a1
Mzb1
Ebf1
Cd22
Cd19
Igkc
Iglc1
Iglc2
Iglc3
Iglc4
Iglv1
Iglv3
Iglv2
Ighm
Ighd
Igha
Ighg1
Ighg2b
Ighg2c
Ighg3
Cd79a
Cd79b
BALBC.M
BALBC.B
BALBC.RBC
BALBC.T
C57BL6.M
C57BL6.B
C57BL6.RBC
C57BL6.T
cell subpopulations
cell subpopulations

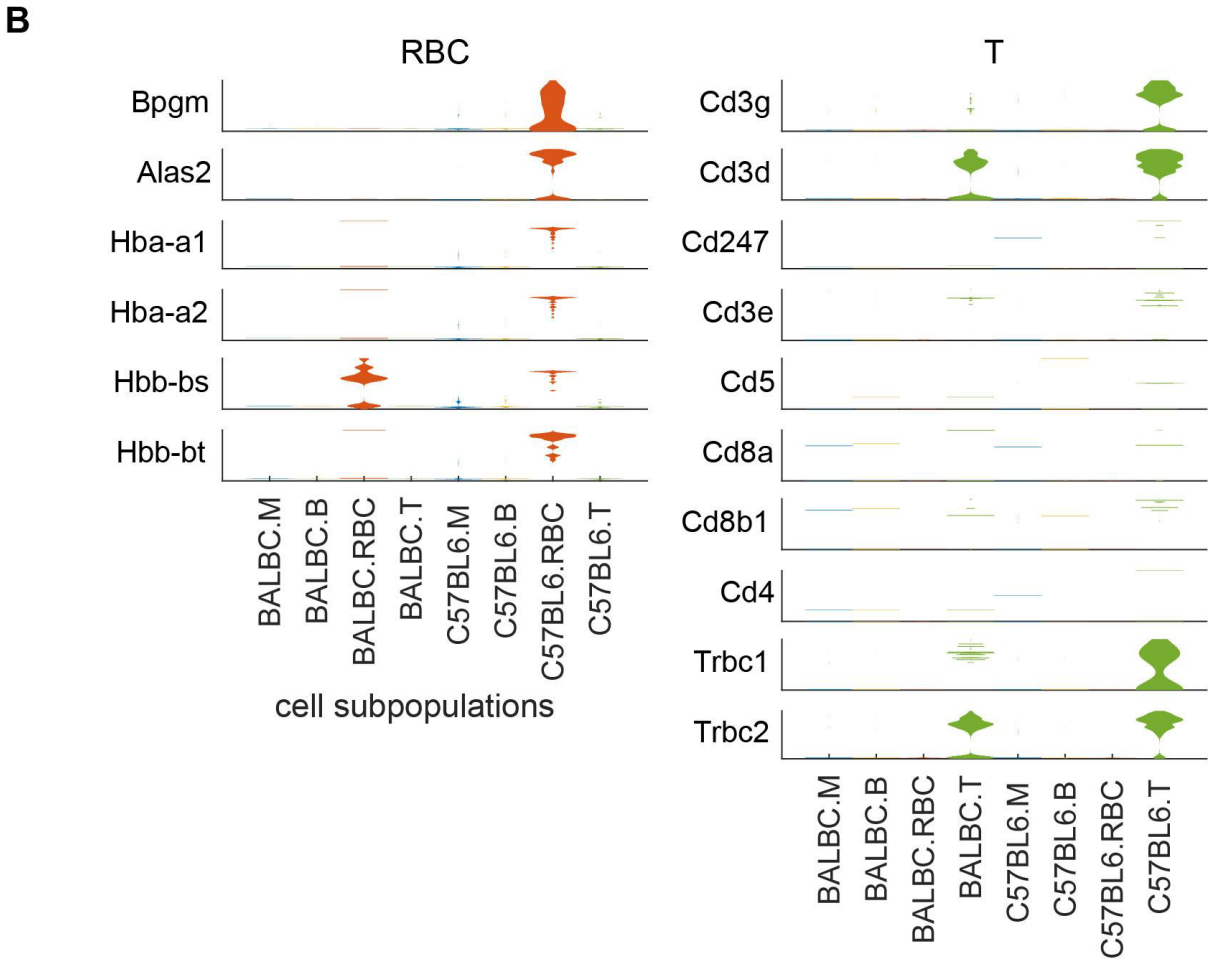


Figure S1. Distribution of cell type-specific makers across cells. A-B. Violin plots showing the expression levels of individual macrophage and B cell makers (**A**) and erythrocytes and T cell markers (**B**) in samples from all the analyzed animals. The y-axis indicates the natural log fold-change (relative to the mean across all cells) in transcriptional activity [20] for each marker (range of variation varies between markers).

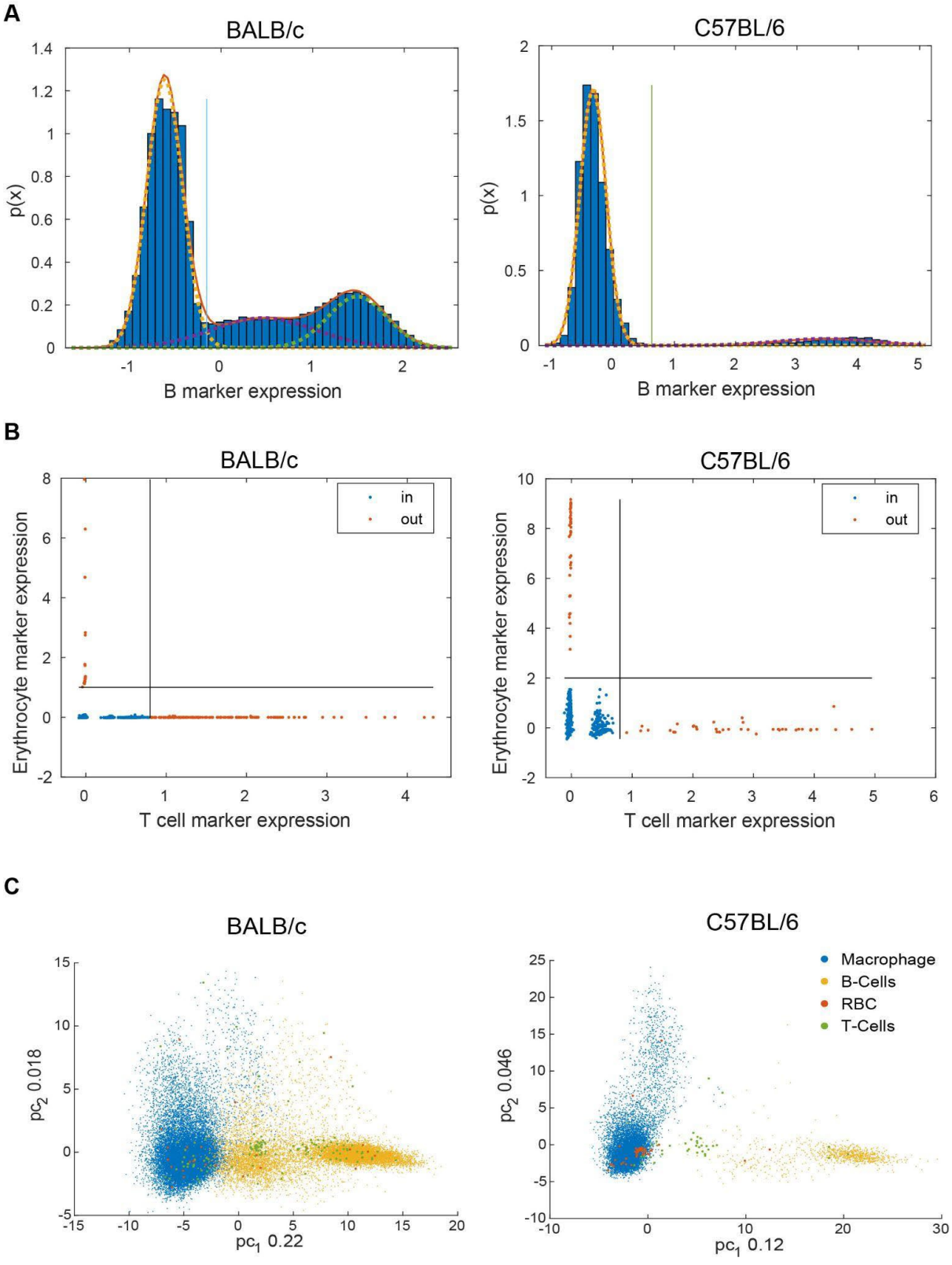


Figure S2. Distribution of average cell type-specific marker expression in cells from the two mouse strains. A. Histogram of B marker expression. **B.** Scatter plots of T cell vs. erythrocyte marker expression. **C.** Projection of all cells from each mouse strain on the first two

principal components of gene expression. The cell type is indicated by the color. In each panel the left figure corresponds to BALB/c and right figure to C57bl/6 mice.

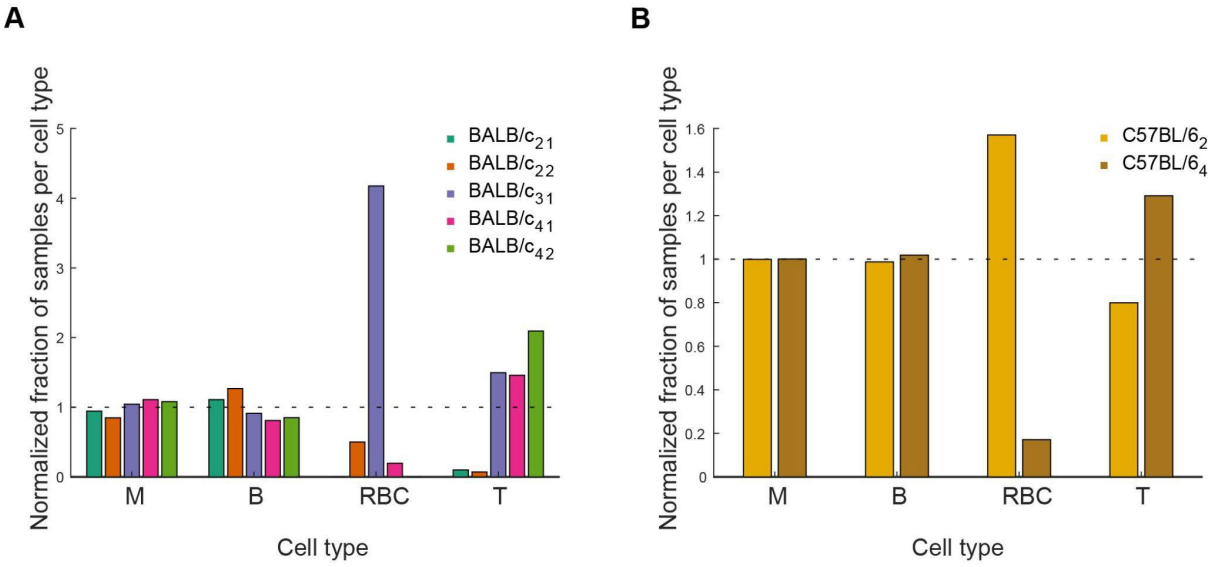
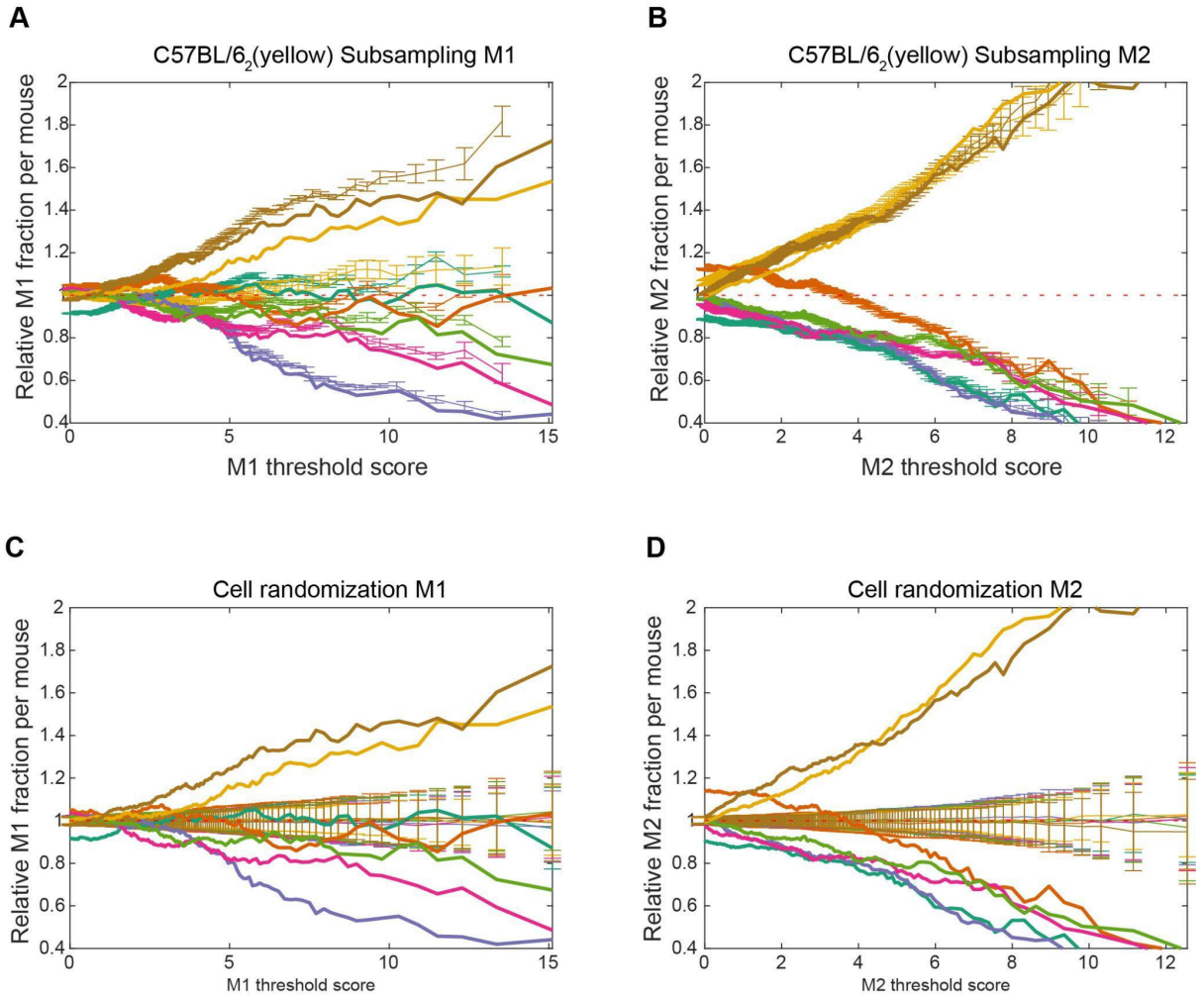


Figure S3. Relative proportions of individual cell types in samples from BALB/c (A) and C57bl/6 (B) mouse samples. The proportions of cells of a given type from individual samples were divided by the proportion of cells of any type in that sample in the data set.



Bibliography

1. Sorg, C. Heterogeneity of macrophages in response to lymphokines and other signals. *Mol. Immunol.* **19**, 1275–1278 (1982).

2. Dijkstra, C. D. & Damoiseaux, J. G. Macrophage heterogeneity established by immunocytochemistry. *Prog. Histochem. Cytochem.* **27**, 1–65 (1993).

3. Ravasi, T. *et al.* Generation of diversity in the innate immune system: macrophage heterogeneity arises from gene-autonomous transcriptional probability of individual inducible genes. *J. Immunol.* **168**, 44–50 (2002).

4. Ramsey, S. *et al.* Transcriptional noise and cellular heterogeneity in mammalian

macrophages. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 495–506 (2006).

5. Gordon, S. & Plüddemann, A. Tissue macrophages: heterogeneity and functions. *BMC Biol.* **15**, 53 (2017).

6. Komohara, Y., Jinushi, M. & Takeya, M. Clinical significance of macrophage heterogeneity in human malignant tumors. *Cancer Sci.* **105**, 1–8 (2014).

7. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* vol. 135 216–226 (2008).

8. Martins, A. J. *et al.* Environment Tunes Propagation of Cell-to-Cell Variation in the Human Macrophage Gene Network. *Cell Syst* **4**, 379–392.e12 (2017).

9. Qian, C. *et al.* Heterogeneous macrophages: Supersensors of exogenous inducing factors. *Scand. J. Immunol.* **90**, e12768 (2019).

10. Watanabe, H., Numata, K., Ito, T., Takagi, K. & Matsukawa, A. Innate immune response in Th1- and Th2-dominant mouse strains. *Shock* **22**, 460–466 (2004).

11. Santos, J. L. *et al.* Differential sensitivity of C57BL/6 (M-1) and BALB/c (M-2) macrophages to the stimuli of IFN-gamma/LPS for the production of NO: correlation with iNOS mRNA and protein expression. *J. Interferon Cytokine Res.* **26**, 682–688 (2006).

12. Soudi, S. *et al.* Comparative study of the effect of LPS on the function of BALB/c and C57BL/6 peritoneal macrophages. *Cell J.* **15**, 45–54 (2013).

13. Yunna, C., Mengru, H., Lei, W. & Weidong, C. Macrophage M1/M2 polarization. *Eur. J. Pharmacol.* **877**, 173090 (2020).

14. Jablonski, K. A. *et al.* Novel Markers to Delineate Murine M1 and M2 Macrophages. *PLoS One* **10**, e0145342 (2015).

15. Chávez-Galán, L., Olleros, M. L., Vesin, D. & Garcia, I. Much More than M1 and M2 Macrophages, There are also CD169(+) and TCR(+) Macrophages. *Front. Immunol.* **6**, 263 (2015).

16. Pavillon, N. & Smith, N. I. Immune cell type, cell activation, and single cell heterogeneity revealed by label-free optical methods. *Sci. Rep.* **9**, 17054 (2019).

17. Yu, Y.-R. A. *et al.* A Protocol for the Comprehensive Flow Cytometric Analysis of Immune Cells in Normal and Inflamed Murine Non-Lymphoid Tissues. *PLoS One* **11**, e0150606 (2016).

18. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).

19. Lantz, C., Radmanesh, B., Liu, E., Thorp, E. B. & Lin, J. Single-cell RNA sequencing uncovers heterogenous transcriptional signatures in macrophages during efferocytosis. *Sci. Rep.* **10**, 14333 (2020).

20. Breda, J., Zavolan, M. & van Nimwegen, E. Bayesian inference of gene expression states from single-cell RNA-seq data. *Nat. Biotechnol.* (2021) doi:10.1038/s41587-021-00875-x.

21. Xaus, J. *et al.* The expression of MHC class II genes in macrophages is cell cycle dependent. *J. Immunol.* **165**, 6364–6371 (2000).

22. Lu, L. *et al.* Time Series miRNA-mRNA integrated analysis reveals critical miRNAs and targets in macrophage polarization. *Sci. Rep.* **6**, 37446 (2016).

23. Hornell, T. M. C., Beresford, G. W., Bushey, A., Boss, J. M. & Mellins, E. D. Regulation of the class II MHC pathway in primary human monocytes by granulocyte-macrophage colony-stimulating factor. *J. Immunol.* **171**, 2374–2383 (2003).

24. Jati, S. *et al.* Wnt5A Signaling Promotes Defense Against Bacterial Pathogens by Activating a Host Autophagy Circuit. *Front. Immunol.* **9**, 679 (2018).

25. Chinen, T. *et al.* An essential role for the IL-2 receptor in Treg cell function. *Nat. Immunol.* **17**, 1322–1333 (2016).

26. Gonzalez-Leal, I. J. *et al.* Cathepsin B in antigen-presenting cells controls mediators of the Th1 immune response during Leishmania major infection. *PLoS Negl. Trop. Dis.* **8**, e3194 (2014).

27. Metzemaekers, M., Vanheule, V., Janssens, R., Struyf, S. & Proost, P. Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands. *Front. Immunol.* **8**, 1970 (2017).

28. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* **6**, 13 (2014).

29. Shankar, A. H., Morin, P. & Titus, R. G. Leishmania major:Differential Resistance to Infection in C57BL/6 (High Interferon-α/β) and Congenic B6.C-H-28c(Low Interferon-α/β) Mice. *Exp. Parasitol.* **84**, 136–143 (1996).

30. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).

Bibliography

Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., Marquez, V., Mielke, T., Thomm, M., Berninghausen, O., Beatrix, B., Söding, J., Westhof, E., Wilson, D.N., Beckmann, R., 2010. Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. Proc. Natl. Acad. Sci. 107, 19754–19759. https://doi.org/10.1073/pnas.1010005107

Balasubramanian, S., Zheng, D., Liu, Y.-J., Fang, G., Frankish, A., Carriero, N., Robilotto, R., Cayting, P., Gerstein, M., 2009. Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. Genome Biol. 10, R2. https://doi.org/10.1186/gb-2009-10-1-r2

BARRNETT, R.J., PALADE, G.E., 1958. APPLICATIONS OF HISTOCHEMISTRY TO ELECTRON MICROSCOPY. J. Histochem. Cytochem. 6, 1–12. https://doi.org/10.1177/6.1.1

Bergmann, T.J., Pisoni, G.B., Molinari, M., Bergmann, T.J., Pisoni, G.B., Molinari, M., 2016. Quality control mechanisms of protein biogenesis: proteostasis dies hard. AIMS Biophys. 3, 456–478. https://doi.org/10.3934/biophy.2016.4.456

Berndt, U., Oellerer, S., Zhang, Y., Johnson, A.E., Rospert, S., 2009. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. Proc. Natl. Acad. Sci. 106, 1398–1403. https://doi.org/10.1073/pnas.0808584106

Blair, J.D., Hockemeyer, D., Doudna, J.A., Bateup, H.S., Floor, S.N., 2017. Widespread Translational Remodeling during Human Neuronal Differentiation. Cell Rep. 21, 2005–2016. https://doi.org/10.1016/j.celrep.2017.10.095

Bornemann, T., Jöckel, J., Rodnina, M.V., Wintermeyer, W., 2008. Signal sequence–independent membrane targeting of ribosomes containing short nascent peptides

within the exit tunnel. Nat. Struct. Mol. Biol. 15, 494–499. https://doi.org/10.1038/nsmb.1402

Bowman, J.C., Hud, N.V., Williams, L.D., 2015. The Ribosome Challenge to the RNA World. J. Mol. Evol. 80, 143–161. https://doi.org/10.1007/s00239-015-9669-9

Bowman, J.C., Petrov, A.S., Frenkel-Pinter, M., Penev, P.I., Williams, L.D., 2020. Root of the Tree: The Significance, Evolution, and Origins of the Ribosome. Chem. Rev. 120, 4848–4878. https://doi.org/10.1021/acs.chemrev.9b00742

Buckley, S.M., Aranda-Orgilles, B., Strikoudis, A., Apostolou, E., Loizou, E., Moran-Crusio, K., Farnsworth, C.L., Koller, A.A., Dasgupta, R., Silva, J.C., Stadtfeld, M., Hochedlinger, K., Chen, E.I., Aifantis, I., 2012. Regulation of Pluripotency and Cellular Reprogramming by the Ubiquitin-Proteasome System. Cell Stem Cell 11, 783–798. https://doi.org/10.1016/j.stem.2012.09.011

Bulut-Karslioglu, A., Biechele, S., Jin, H., Macrae, T.A., Hejna, M., Gertsenstein, M., Song, J.S., Ramalho-Santos, M., 2016. Inhibition of mTOR induces a paused pluripotent state. Nature 540, 119–123. https://doi.org/10.1038/nature20578

Chaillou, T., Zhang, X., McCarthy, J.J., 2016. Expression of muscle-specific ribosomal protein L3-like impairs myotube growth. J. Cell. Physiol. 231, 1894–1902. https://doi.org/10.1002/jcp.25294

Chappell, J., Dalton, S., 2013. Roles for MYC in the Establishment and Maintenance of Pluripotency. Cold Spring Harb. Perspect. Med. 3, a014381. https://doi.org/10.1101/cshperspect.a014381

Chávez-Galán, L., Olleros, M.L., Vesin, D., Garcia, I., 2015. Much More than M1 and M2 Macrophages, There are also CD169+ and TCR+ Macrophages. Front. Immunol. 6.

Chen, C., Wang, E., Liu, P., Xiao, Y., 2013. Simulation study of the role of the ribosomal exit tunnel on protein folding. Phys. Rev. E 87, 022701. https://doi.org/10.1103/PhysRevE.87.022701

Cheng, Z., Mugler, C.F., Keskin, A., Hodapp, S., Chan, L.Y.-L., Weis, K., Mertins, P., Regev, A., Jovanovic, M., Brar, G.A., 2019. Small and Large Ribosomal Subunit Deficiencies Lead to Distinct Gene Expression Signatures that Reflect Cellular Growth Rate. Mol. Cell 73, 36-47.e10. https://doi.org/10.1016/j.molcel.2018.10.032

Cloutier, J.M., Turner, J.M.A., 2010. Meiotic sex chromosome inactivation. Curr. Biol. 20, R962–R963. https://doi.org/10.1016/j.cub.2010.09.041

Costa-Mattioli, M., Klann, E., 2017. 4.15 - Translational Control Mechanisms in Synaptic Plasticity and Memory☆, in: Byrne, J.H. (Ed.), Learning and Memory: A Comprehensive Reference (Second Edition). Academic Press, Oxford, pp. 311–328. https://doi.org/10.1016/B978-0-12-809324-5.21110-3

Dao Duc, K., Batra, S.S., Bhattacharya, N., Cate, J.H.D., Song, Y.S., 2019. Differences in the path to exit the ribosome across the three domains of life. Nucleic Acids Res. 47, 4198–4210. https://doi.org/10.1093/nar/gkz106

Deutsch, C., 2003. The Birth of a Channel. Neuron 40, 265–276. https://doi.org/10.1016/S0896-6273(03)00506-3

Dez, C., Tollervey, D., 2004. Ribosome synthesis meets the cell cycle. Curr. Opin. Microbiol., Growth and development 7, 631–637. https://doi.org/10.1016/j.mib.2004.10.007

Duc, K.D., Song, Y.S., 2018. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. PLOS Genet. 14, e1007166. https://doi.org/10.1371/journal.pgen.1007166

Edholm, E.-S., Rhoo, K.H., Robert, J., 2017. Evolutionary Aspects of Macrophages Polarization, in: Kloc, M. (Ed.), Macrophages: Origin, Functions and Biointervention, Results and Problems in Cell Differentiation. Springer International Publishing, Cham, pp. 3–22. https://doi.org/10.1007/978-3-319-54090-0_1

Espinosa, V., Rivera, A., 2016. First Line of Defense: Innate Cell-Mediated Control of Pulmonary Aspergillosis. Front. Microbiol. 7.

Fonseca, B.D., Smith, E.M., Yelle, N., Alain, T., Bushell, M., Pause, A., 2014. The ever-evolving role of mTOR in translation. Semin. Cell Dev. Biol., Development of the urogenital system & mTOR Signalling & Tight Junctions in Health and Disease 36, 102–112. https://doi.org/10.1016/j.semcdb.2014.09.014

Gamerdinger, M., Kobayashi, K., Wallisch, A., Kreft, S.G., Sailer, C., Schlömer, R., Sachs, N., Jomaa, A., Stengel, F., Ban, N., Deuerling, E., 2019. Early Scanning of Nascent Polypeptides inside the Ribosomal Tunnel by NAC. Mol. Cell 75, 996-1006.e8. https://doi.org/10.1016/j.molcel.2019.06.030

Gerst, J.E., 2018. Pimp My Ribosome: Ribosomal Protein Paralogs Specify Translational Control. Trends Genet. 34, 832–845. https://doi.org/10.1016/j.tig.2018.08.004

Guan, K., Rohwedel, J., Wobus, A.M., 1999. Embryonic stem cell differentiation models: cardiogenesis, myogenesis, neurogenesis, epithelial and vascular smooth muscle cell differentiation in vitro. Cytotechnology 30, 211–226. https://doi.org/10.1023/A:1008041420166

Guha, D., Banerjee, A., Mukherjee, R., Pradhan, B., Peneva, M., Aleksandrov, G., Suklabaidya, S., Senapati, S., Aich, P., 2019. A probiotic formulation containing Lactobacillus bulgaricus DWT1 inhibits tumor growth by activating pro-inflammatory responses in macrophages. J. Funct. Foods 56, 232–245. https://doi.org/10.1016/j.jff.2019.03.030

Guimaraes, J.C., Zavolan, M., 2016. Patterns of ribosomal protein expression specify normal and malignant human cells. Genome Biol. 17, 236. https://doi.org/10.1186/s13059-016-1104-z

Gumiero, A., Conz, C., Gesé, G.V., Zhang, Y., Weyer, F.A., Lapouge, K., Kappes, J., von Plehwe, U., Schermann, G., Fitzke, E., Wölfle, T., Fischer, T., Rospert, S., Sinning, I., 2016. Interaction of the cotranslational Hsp70 Ssb with ribosomal proteins and rRNA

depends on its lid domain. Nat. Commun. 7, 13563. https://doi.org/10.1038/ncomms13563

Han, S., Sun, S., Li, P., Liu, Q., Zhang, Z., Dong, H., Sun, M., Wu, W., Wang, X., Guo, H., 2020. Ribosomal Protein L13 Promotes IRES-Driven Translation of Foot-and-Mouth Disease Virus in a Helicase DDX3-Dependent Manner. J. Virol. 94, e01679-19. https://doi.org/10.1128/JVI.01679-19

Hanna, H., Mir, L.M., Andre, F.M., 2018. In vitro osteoblastic differentiation of mesenchymal stem cells generates cell layers with distinct properties. Stem Cell Res. Ther. 9, 203. https://doi.org/10.1186/s13287-018-0942-x

Harish, A., Caetano-Anollés, G., 2012. Ribosomal History Reveals Origins of Modern Protein Synthesis. PLOS ONE 7, e32776. https://doi.org/10.1371/journal.pone.0032776

Ingelfinger, F., Gerdes, L.A., Kavaka, V., Krishnarajah, S., Friebel, E., Galli, E., Zwicky, P., Furrer, R., Peukert, C., Dutertre, C.-A., Eglseer, K.M., Ginhoux, F., Flierl-Hecht, A., Kümpfel, T., De Feo, D., Schreiner, B., Mundt, S., Kerschensteiner, M., Hohlfeld, R., Beltrán, E., Becher, B., 2022. Twin study reveals non-heritable immune perturbations in multiple sclerosis. Nature 603, 152–158. https://doi.org/10.1038/s41586-022-04419-4

Jackson, R.J., Hellen, C.U.T., Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. 11, 113–127. https://doi.org/10.1038/nrm2838

Janes, K.A., 2015. Cell-to-Cell Transcript Variability: Seeing Signal in the Noise. Cell 163, 1566–1568. https://doi.org/10.1016/j.cell.2015.12.010

Jiang, L., Li, T., Zhang, X., Zhang, B., Yu, C., Li, Y., Fan, S., Jiang, X., Khan, T., Hao, Q., Xu, P., Nadano, D., Huleihel, M., Lunenfeld, E., Wang, P.J., Zhang, Y., Shi, Q., 2017. RPL10L Is Required for Male Meiotic Division by Compensating for RPL10 during

Meiotic Sex Chromosome Inactivation in Mice. Curr. Biol. 27, 1498-1505.e6. https://doi.org/10.1016/j.cub.2017.04.017

Karamyshev, A.L., Karamysheva, Z.N., 2018. Lost in Translation: Ribosome-Associated mRNA and Protein Quality Controls. Front. Genet. 9.

Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T.J., Tanaka, T., Page, D.C., 1998. A map of 75 human ribosomal protein genes. Genome Res. 8, 509–523. https://doi.org/10.1101/gr.8.5.509

Kerkis, A., Fonseca, S.A.S., Serafim, R.C., Lavagnolli, T.M.C., Abdelmassih, S., Abdelmassih, R., Kerkis, I., 2007. In Vitro Differentiation of Male Mouse Embryonic Stem Cells into Both Presumptive Sperm Cells and Oocytes. Cloning Stem Cells 9, 535–548. https://doi.org/10.1089/clo.2007.0031

Komar, A.A., Hatzoglou, M., 2011. Cellular IRES-mediated translation. Cell Cycle 10, 229–240. https://doi.org/10.4161/cc.10.2.14472

Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Xue, S., Ishijima, J., Shiroishi, T., Barna, M., 2011. Ribosome-Mediated Specificity in Hox mRNA Translation and Vertebrate Tissue Patterning. Cell 145, 383–397. https://doi.org/10.1016/j.cell.2011.03.028

Kozak, M., 1999. Initiation of translation in prokaryotes and eukaryotes. Gene 234, 187–208. https://doi.org/10.1016/S0378-1119(99)00210-3

Kramer, G., Boehringer, D., Ban, N., Bukau, B., 2009. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. 16, 589–597. https://doi.org/10.1038/nsmb.1614

Kuzumaki, T., Tanaka, T., Ishikawa, K., Ogata, K., 1987. Rat ribosomal protein L35a multigene family: molecular structure and characterization of three L35a-related pseudogenes. Biochim. Biophys. Acta 909, 99–106. https://doi.org/10.1016/0167-4781(87)90031-5

Lafontaine, D.L.J., Tollervey, D., 2001. The function and synthesis of ribosomes. Nat. Rev. Mol. Cell Biol. 2, 514–520. https://doi.org/10.1038/35080045

Lavin, Y., Mortha, A., Rahman, A., Merad, M., 2015. Regulation of macrophage development and function in peripheral tissues. Nat. Rev. Immunol. 15, 731–744. https://doi.org/10.1038/nri3920

Leary, D.J., Huang, S., 2001. Regulation of ribosome biogenesis within the nucleolus. FEBS Lett. 509, 145–150. https://doi.org/10.1016/S0014-5793(01)03143-X

Lee, K., Ziegelhoffer, T., Delewski, W., Berger, S.E., Sabat, G., Craig, E.A., 2021. Pathway of Hsp70 interactions at the ribosome. Nat. Commun. 12, 5666. https://doi.org/10.1038/s41467-021-25930-8

Lossos, I.S., Akasaka, T., Martinez-Climent, J.A., Siebert, R., Levy, R., 2003. The BCL6 gene in B-cell lymphomas with 3q27 translocations is expressed mainly from the rearranged allele irrespective of the partner gene. Leukemia 17, 1390–1397. https://doi.org/10.1038/sj.leu.2402997

Lu, J., Deutsch, C., 2008. Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. J. Mol. Biol. 384, 73–86. https://doi.org/10.1016/j.jmb.2008.08.089

Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F., Gray, M.W., 2011. How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life 63, 528–537. https://doi.org/10.1002/iub.489

Luo, Y., Na, Z., Slavoff, S.A., 2018. P-Bodies: Composition, Properties, and Functions. Biochemistry 57, 2424–2431. https://doi.org/10.1021/acs.biochem.7b01162

Magaña-Acosta, M., Valadez-Graham, V., 2020. Chromatin Remodelers in the 3D Nuclear Compartment. Front. Genet. 11.

Martinez, F.O., Gordon, S., 2014. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Rep 6.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., Yusupov, M., 2012. One core, two shells: bacterial and eukaryotic ribosomes. Nat. Struct. Mol. Biol. 19, 560–567. https://doi.org/10.1038/nsmb.2313

Mills, C.D., Kincaid, K., Alt, J.M., Heilman, M.J., Hill, A.M., 2000. M-1/M-2 Macrophages and the Th1/Th2 Paradigm. J. Immunol. 164, 6166–6173. https://doi.org/10.4049/jimmunol.164.12.6166

Morgani, S., Nichols, J., Hadjantonakis, A.-K., 2017. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. BMC Dev. Biol. 17, 7. https://doi.org/10.1186/s12861-017-0150-4

Ngondo, R.P., Cirera-Salinas, D., Yu, J., Wischnewski, H., Bodak, M., Vandormael-Pournin, S., Geiselmann, A., Wettstein, R., Luitz, J., Cohen-Tannoudji, M., Ciaudo, C., 2018. Argonaute 2 Is Required for Extra-embryonic Endoderm Differentiation of Mouse Embryonic Stem Cells. Stem Cell Rep. 10, 461–476. https://doi.org/10.1016/j.stemcr.2017.12.023

Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A., 2000. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. Science 289, 920–930. https://doi.org/10.1126/science.289.5481.920

Orecchioni, M., Ghosheh, Y., Pramod, A.B., Ley, K., 2019. Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS–) vs. Alternatively Activated Macrophages. Front. Immunol. 10.

Paci, G., Lemke, E.A., 2017. Shining a Light on Phase Separation in the Cell. Cell 168, 11–13. https://doi.org/10.1016/j.cell.2016.12.018

Pavitt, G.D., Ron, D., 2012. New Insights into Translational Regulation in the Endoplasmic Reticulum Unfolded Protein Response. Cold Spring Harb. Perspect. Biol. 4, a012278. https://doi.org/10.1101/cshperspect.a012278

Pechmann, S., Willmund, F., Frydman, J., 2013. The Ribosome as a Hub for Protein Quality Control. Mol. Cell 49, 411–421. https://doi.org/10.1016/j.molcel.2013.01.020

Peisker, K., Braun, D., Wölfle, T., Hentschel, J., Fünfschilling, U., Fischer, G., Sickmann, A., Rospert, S., 2008. Ribosome-associated Complex Binds to Ribosomes in Close Proximity of Rpl31 at the Exit of the Polypeptide Tunnel in Yeast. Mol. Biol. Cell 19, 5279–5288. https://doi.org/10.1091/mbc.E08-06-0661

Petrone, P.M., Snow, C.D., Lucent, D., Pande, V.S., 2008. Side-chain recognition and gating in the ribosome exit tunnel. Proc. Natl. Acad. Sci. U. S. A. 105, 16549–16554. https://doi.org/10.1073/pnas.0801795105

Petrov, A.S., Bernier, C.R., Hsiao, C., Norris, A.M., Kovacs, N.A., Waterbury, C.C., Stepanov, V.G., Harvey, S.C., Fox, G.E., Wartell, R.M., Hud, N.V., Williams, L.D., 2014. Evolution of the ribosome at atomic resolution. Proc. Natl. Acad. Sci. 111, 10251–10256. https://doi.org/10.1073/pnas.1407205111

Pradhan, B., Guha, D., Naik, A.K., Banerjee, A., Tambat, S., Chawla, S., Senapati, S., Aich, P., 2019. Probiotics L. acidophilus and B. clausii Modulate Gut Microbiota in Th1- and Th2-Biased Mice to Ameliorate Salmonella Typhimurium-Induced Diarrhea. Probiotics Antimicrob. Proteins 11, 887–904. https://doi.org/10.1007/s12602-018-9436-5

Prosdocimi, F., Zamudio, G.S., Palacios-Pérez, M., Torres de Farias, S., V. José, M., 2020. The Ancient History of Peptidyl Transferase Center Formation as Told by Conservation and Information Analyses. Life 10, 134. https://doi.org/10.3390/life10080134

Ravasi, T., Wells, C., Forest, A., Underhill, D.M., Wainwright, B.J., Aderem, A., Grimmond, S., Hume, D.A., 2002. Generation of Diversity in the Innate Immune System: Macrophage Heterogeneity Arises from Gene-Autonomous Transcriptional

Probability of Individual Inducible Genes. J. Immunol. 168, 44–50. https://doi.org/10.4049/jimmunol.168.1.44

Rudra, D., Warner, J.R., 2004. What better measure than ribosome synthesis? Genes Dev. 18, 2431–2436. https://doi.org/10.1101/gad.1256704

Rutkowski, D.T., Lingappa, V.R., Hegde, R.S., 2001. Substrate-specific regulation of the ribosome– translocon junction by N-terminal signal sequences. Proc. Natl. Acad. Sci. 98, 7823–7828. https://doi.org/10.1073/pnas.141125098

Saba, J.A., Liakath-Ali, K., Green, R., Watt, F.M., 2021. Translational control of stem cell function. Nat. Rev. Mol. Cell Biol. 22, 671–690. https://doi.org/10.1038/s41580-021-00386-2

Sampath, P., Pritchard, D.K., Pabon, L., Reinecke, H., Schwartz, S.M., Morris, D.R., Murry, C.E., 2008. A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation. Cell Stem Cell 2, 448–460. https://doi.org/10.1016/j.stem.2008.03.013

Schmidt, C., Robinson, C.V., 2014. A comparative cross-linking strategy to probe conformational changes in protein complexes. Nat. Protoc. 9, 2224–2236. https://doi.org/10.1038/nprot.2014.144

Signer, R.A.J., Magee, J.A., Salic, A., Morrison, S.J., 2014. Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature 509, 49–54. https://doi.org/10.1038/nature13035

Somoza, R.A., Welter, J.F., Correa, D., Caplan, A.I., 2014. Chondrogenic Differentiation of Mesenchymal Stem Cells: Challenges and Unfulfilled Expectations. Tissue Eng. Part B Rev. 20, 596–608. https://doi.org/10.1089/ten.teb.2013.0771

Steitz, T.A., Moore, P.B., 2003. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. Trends Biochem. Sci. 28, 411–418. https://doi.org/10.1016/S0968-0004(03)00169-5

Sugihara, Y., Honda, H., Iida, T., Morinaga, T., Hino, S., Okajima, T., Matsuda, T., Nadano, D., 2010. Proteomic Analysis of Rodent Ribosomes Revealed Heterogeneity Including Ribosomal Proteins L10-like, L22-like 1, and L39-like. J. Proteome Res. 9, 1351–1366. https://doi.org/10.1021/pr9008964

Sur, A., Pradhan, B., Banerjee, A., Aich, P., 2015. Immune Activation Efficacy of Indolicidin Is Enhanced upon Conjugation with Carbon Nanotubes and Gold Nanoparticles. PLOS ONE 10, e0123905. https://doi.org/10.1371/journal.pone.0123905

Tahmasebi, S., Amiri, M., Sonenberg, N., 2019. Translational Control in Stem Cells. Front. Genet. 9.

Taylor, J.M., Ziman, M.E., Canfield, D.R., Vajdy, M., Solnick, J.V., 2008. Effects of a Th1- versus a Th2-biased immune response in protection against Helicobacter pylori challenge in mice. Microb. Pathog. 44, 20–27. https://doi.org/10.1016/j.micpath.2007.06.006

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., Melton, D.W., 1989. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell 56, 313–321. https://doi.org/10.1016/0092-8674(89)90905-7

Tsui, S.K., Lee, S.M., Fung, K.P., Waye, M.M., Lee, C.Y., 1996. Primary structures and sequence analysis of human ribosomal proteins L39 and S27. Biochem. Mol. Biol. Int. 40, 611–616. https://doi.org/10.1080/15216549600201203

Tu, D., Blaha, G., Moore, P.B., Steitz, T.A., 2005. Structures of MLSBK Antibiotics Bound to Mutated Large Ribosomal Subunits Provide a Structural Explanation for Resistance. Cell 121, 257–270. https://doi.org/10.1016/j.cell.2005.02.005

Tunnacliffe, E., Chubb, J.R., 2020. What Is a Transcriptional Burst? Trends Genet. 36, 288–297. https://doi.org/10.1016/j.tig.2020.01.003

Uechi, T., Maeda, N., Tanaka, T., Kenmochi, N., 2002. Functional second genes generated by retrotransposition of the X-linked ribosomal protein genes. Nucleic Acids Res. 30, 5369–5375. https://doi.org/10.1093/nar/gkf696

van Riggelen, J., Yetil, A., Felsher, D.W., 2010. MYC as a regulator of ribosome biogenesis and protein synthesis. Nat. Rev. Cancer 10, 301–309. https://doi.org/10.1038/nrc2819

Watanabe, H., Numata, K., Ito, T., Takagi, K., Matsukawa, A., 2004. INNATE IMMUNE RESPONSE IN TH1- AND TH2-DOMINANT MOUSE STRAINS. Shock 22, 460–466. https://doi.org/10.1097/01.shk.0000142249.08135.e9

Watanabe-Susaki, K., Takada, H., Enomoto, K., Miwata, K., Ishimine, H., Intoh, A., Ohtaka, M., Nakanishi, M., Sugino, H., Asashima, M., Kurisaki, A., 2014. Biosynthesis of Ribosomal RNA in Nucleoli Regulates Pluripotency and Differentiation Ability of Pluripotent Stem Cells. Stem Cells 32, 3099–3111. https://doi.org/10.1002/stem.1825

Wek, R.C., 2018. Role of eIF2α Kinases in Translational Control and Adaptation to Cellular Stress. Cold Spring Harb. Perspect. Biol. 10, a032870. https://doi.org/10.1101/cshperspect.a032870

Wilson, D.M., Li, Y., LaPeruta, A., Gamalinda, M., Gao, N., Woolford, J.L., 2020. Structural insights into assembly of the ribosomal nascent polypeptide exit tunnel. Nat. Commun. 11, 5111. https://doi.org/10.1038/s41467-020-18878-8

Wilson, D.N., Beckmann, R., 2011. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. Curr. Opin. Struct. Biol. 21, 274–282. https://doi.org/10.1016/j.sbi.2011.01.007

Wong, Q.W.-L., Li, J., Ng, S.R., Lim, S.G., Yang, H., Vardy, L.A., 2014. RPL39L is an example of a recently evolved ribosomal protein paralog that shows highly specific tissue expression patterns and is upregulated in ESCs and HCC tumors. RNA Biol. 11, 33–41. https://doi.org/10.4161/rna.27427

Xiong, X., Zhao, Y., He, H., Sun, Y., 2011. Ribosomal protein S27-like and S27 interplay with p53-MDM2 axis as a target, a substrate and a regulator. Oncogene 30, 1798–1811. https://doi.org/10.1038/onc.2010.569

Xu, X., Pantakani, D.V.K., Lührig, S., Tan, X., Khromov, T., Nolte, J., Dressel, R., Zechner, U., Engel, W., 2011. Stage-Specific Germ-Cell Marker Genes Are Expressed in All Mouse Pluripotent Cell Types and Emerge Early during Induced Pluripotency. PLOS ONE 6, e22413. https://doi.org/10.1371/journal.pone.0022413

Yan, P., Yang, X., Wang, J., Wang, S., Ren, H., 2019. A novel CpG island methylation panel predicts survival in lung adenocarcinomas. Oncol. Lett. 18, 1011–1022. https://doi.org/10.3892/ol.2019.10431

Yusupova, G., Yusupov, M., 2017. Crystal structure of eukaryotic ribosome and its complexes with inhibitors. Philos. Trans. R. Soc. B Biol. Sci. 372, 20160184. https://doi.org/10.1098/rstb.2016.0184

Zaidi, S.K., Boyd, J.R., Grandy, R.A., Medina, R., Lian, J.B., Stein, G.S., Stein, J.L., 2016. Expression of Ribosomal RNA and Protein Genes in Human Embryonic Stem Cells Is Associated With the Activating H3K4me3 Histone Mark. J. Cell. Physiol. 231, 2007–2013. https://doi.org/10.1002/jcp.25309

Zaman, S., Fitzpatrick, M., Lindahl, L., Zengel, J., 2007. Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in Escherichia coli. Mol. Microbiol. 66, 1039–1050. https://doi.org/10.1111/j.1365-2958.2007.05975.x

Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N., Sui, S.-F., 2015. Mechanisms of ribosome stalling by SecM at multiple elongation steps. eLife 4, e09684. https://doi.org/10.7554/eLife.09684

Zhang, Y., O'Leary, M.N., Peri, S., Wang, M., Zha, J., Melov, S., Kappes, D.J., Feng, Q., Rhodes, J., Amieux, P.S., Morris, D.R., Kennedy, B.K., Wiest, D.L., 2017. Ribosomal

Proteins Rpl22 and Rpl22l1 Control Morphogenesis by Regulating Pre-mRNA Splicing. Cell Rep. 18, 545–556. https://doi.org/10.1016/j.celrep.2016.12.034

Zhang, Y., Wölfle, T., Rospert, S., 2013. Interaction of Nascent Chains with the Ribosomal Tunnel Proteins Rpl4, Rpl17, and Rpl39 of Saccharomyces cerevisiae. J. Biol. Chem. 288, 33697–33707. https://doi.org/10.1074/jbc.M113.508283

Zhang, Z., Harrison, P., Gerstein, M., 2002. Identification and Analysis of Over 2000 Ribosomal Protein Pseudogenes in the Human Genome. Genome Res. 12, 1466–1482. https://doi.org/10.1101/gr.331902

Zhao, B.S., Roundtree, I.A., He, C., 2017. Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. 18, 31–42. https://doi.org/10.1038/nrm.2016.132

Zou, Q., Yang, L., Shi, R., Qi, Y., Zhang, X., Qi, H., 2021. Proteostasis regulated by testis-specific ribosomal protein RPL39L maintains mouse spermatogenesis. iScience 24, 103396. https://doi.org/10.1016/j.isci.2021.103396