Who watches the watchmen?

*WNT responsive stem cells and the regulation of their niche ligands*

van de Grift, Y.B.C.

Publication date

2023

’Make new mistakes. Make glorious, amazing mistakes. Make mistakes nobody’s ever made before. Don’t freeze, don’t stop, don’t worry that it isn’t good enough, or isn’t perfect, whatever it is: art, or love, or work or family or life.

Whatever it is your scared of doing, Do it.’

Neil Gaiman
Chapter 8

General discussion
Chapter 8

General discussion

In this thesis, we tried to shed light on the spatiotemporal dynamics of WNT-responsive stem cells and the regulation of their niches. This resulted in novel mouse models that 1) allow visualisation of WNT-responsive stem cells and 2) allow simultaneous lineage tracing of multiple stem cell clones and at the same time visualize cell shape and more importantly cell-cell interactions. Stem cell identity is controlled by so-called niche cells that maintain a stem cell population by the secretion of short/long-range ligands (e.g. WNT ligands) \(^1\text{–}^6\). In most mammalian tissues, including the mammary gland, the niche cells and thus the sources of factors that control stem cell identity are not clear. Even if the niche has been identified, the regulation of the specific niche factors is generally still a mystery. It is an outstanding question in the (WNT) field how factors that control stem cell identity are regulated. Hence, if the stem cells are a metaphorical population of citizens and their niche cells the city watchmen that keep them in line, the question remains what ensures tight spatiotemporal control of the signalling ligands coming from the niche: Who watches the Watchmen?
Novel mouse reporter models to visualize and study WNT-responsive stem cell dynamics and their respective niches

Tissue context matters for stem cell behaviour. For example, cross-talk between stem cells and their niche cells, stem cell competition and ligand availability (e.g. due hormonal regulation in the mammary gland) are all factors that determine stem cell dynamics. Many techniques and mouse models can only provide a snapshot in time, in which endogenous context, dynamics over time, or both are lost. In chapter 2 & 3 we characterized two novel mouse models that may help better address these questions.

In chapter 2 we characterized a novel *Axin2* knock-in strain that doubles as a fluorescent reporter and lineage tracing driver for WNT/CTNNB1 responsive cells. Although truly universal WNT/CTNNB1 target genes are rare \(^7\), *Axin2* has been shown to reliably report active WNT/CTNNB1 signalling in cells across tissues \(^8\). Previously published *Axin2* mouse models involve 5’ Knock-in of either *LacZ* \(^8\) or *CreERT2* \(^9\). The advantage of our novel *Axin2* \(^{P2A}-rT3A-T2A-3xNLS-SGFP2\) reporter strain is threefold: 1) Unlike 5’ *Axin2* knock-in mice, our 3’ UTR knock-in leaves the entire *Axin2* gene intact. Meaning that it preserves the *Axin2* copy number whereas other *Axin2* mouse models disrupt the *Axin2* gene and that both 5’ and 3’ UTR transcriptional and translational control are maintained. We did not detect any phenotype associated with homozygosity of the knock-in allele unlike 5’ models (Chapter 2, supplementary Table 1-2) \(^8\)\(^-\)\(^1\)\(^3\). Moreover, SGFP2 is stably expressed over prolonged periods of time (up to 1 year) (This chapter, Fig1A-B) and does not get silenced, which is a known issue with for instance some existing *Lgr5-GFP* reporter strains (personal experience). 2) The 3xNLS-sGFP2 in the reporter strain allows for direct and dynamic visualisation of *Axin2* expression and thus WNT/CTNNB1 activity in individual cells. Our design using P2A/T2A linkers ensures roughly equimolar expression of AXIN2 and sGFP2 \(^1\)\(^4\) (Chapter 2, Fig2C-F) and thus can be used to quantify WNT/CTNNB1 activation. 3) *CreERT2* mouse reporter strains depend on Tamoxifen for the initiation of lineage tracing. Tamoxifen is a selective oestrogen-receptor modulator. In hormone-responsive tissues, such as the mammary gland and ovaries, administration of tamoxifen can alter tissue homeostasis and thereby influence the results \(^1\)\(^5\), \(^1\)\(^6\). It has been reported that in the mouse mammary gland, tamoxifen can skew the distribution of mammary epithelial cell types in a dose dependent manner, and thus caution should be taken when interpreting lineage tracing experiments using high doses of tamoxifen. Particular care should be taken when quantitative short-term analysis are performed \(^1\)\(^5\). Moreover, recent studies suggest that the combination of tamoxifen and CreERT2 is particularly toxic to proliferating cells in some tissues \(^1\)\(^0\), \(^1\)\(^7\).
Alternatively, we designed an Axin2-dependent, doxycycline inducible rtA/TetON-Cre system for lineage tracing. Although breeding is more time consuming, as a triple heterozygous intercross (e.g. $\text{Axin2}^{\text{P2A-rtTA3-T2A-3xNLS-SGFP2}};\text{Tg}^{\text{TetON-Cre}};\text{Rosa26}^{\text{mTmG}}$) is required, Doxycycline is in general less toxic and does not alter tissue homeostasis in hormone-responsive tissues.

Figure 1: $\text{Axin2}^{\text{P2A-rtTA3-T2A-3xNLS-SGFP2}};\text{Rosa26}^{\text{mTmG}}$ long term trace experiments. A) Bottom section of the small intestine of a >1 year old mouse showing SGFP2 signal. B) Bottom section of the small intestine of a 1 year trace $\text{Axin2}^{\text{P2A-rtTA3-T2A-3xNLS-SGFP2}};\text{Rosa26}^{\text{mTmG}}$ mouse ($n=1$). Experiments in A-B) were performed after publication of the paper in chapter 2.

Surprisingly, our $\text{Axin2}^{\text{P2A-rtTA3-T2A-3xNLS-SGFP2}}$ does not fully recapitulate results described in 5' Axin2 knock-in models: $\text{Axin2}^{\text{LacZ}}$ 8, $\text{Axin2}^{\text{CreERT2}}$ 9 and $\text{Axin2}^{\text{mGFP}}$ 12. Specifically, LacZ/mGFP reporter expression and CreERT2 mediated recombination was detected in a larger set of adult cell types than SGFP2 expression and rtTA3 induced Cre/lox recombination in our 3' knock-in model 2, 4-6, 8, 9, 12, 19-22. This discrepancy in different Axin2 reporter strains can be due to either one of the following reasons. First, the difference can be technical in nature. For example, the tetO-Cre line 23 that is activated by $\text{Axin2}^{\text{P2A-rtTA3-T2A-3xNLS-SGFP2}}$ in the experiments in chapter 2 is not leaky, but may display a high threshold for activation. In combination with unknown Doxycycline availability in different tissues, the low Axin2 (and thus rtTA3) signal may result in insufficient Cre activation to induce Cre/lox recombination of even the sensitive $\text{Rosa26}^{\text{mTmG}}$ allele. This, however, seems unlikely since repeating the experiments from chapter 2 with a different, less stringent tetO-Cre line (RRID: IMSR_JAX:006234) 24, 25 did not alter our results (data not shown). Second, unlike other published Axin2 mouse strains, our new knock-in model preserves both 5' and 3' transcriptional and translational regulation, and thus fully recapitulates endogenous Axin2 expression (Fig2A-B). Our model may have inadvertently
uncovered that 3'UTR regulation is relevant for maintaining low levels of Axin2 expression in vivo. Several studies have shown that particularly in cancer cells Axin2 can be inhibited by microRNAs through 3'UTR regulation. We speculate that this manner of Axin2 regulation is also present during adult tissue homeostasis and that Axin2 is particularly tightly regulated through its 3'UTR. If true, this underscores the importance of generating reporter mouse models that completely maintain a gene’s endogenous transcriptional and translational regulation, and that care should be taken when interpreting results between different models. For example, our Axin2P2A-rtTA3-T2A-3xNLS-SGFP2 does not detect Axin2 expression in the pubertal mammary gland (data not shown), whereas the Axin2mGFP model detect mGFP expression in a subset of basal cells and more strikingly over half of the Pdgfra+ mammary fibroblasts. This observation is also supported by Axin2LacZ and Axin2CreERT2. Further experiments confirm that WNT/CTNNB1 signalling is indeed active in these fibroblasts and that it actively contributes to mammary gland homeostasis. Thus the signal detected in these Axin2 models is biologically relevant, however, they possibly misrepresent the actual Axin2 expression and overexaggerate the Axin2+ population. The differences between the 5' prime and 3' prime Axin2 mouse models also illustrates the limits of detection in fluorescent reporter mouse models and how that may affect the interpretation of results. Thus the current Axin2 reporter models offer a case study for how knock-in design can affect reporter read-out.

In Chapter 3, we characterized a novel multicolour lineage tracing reporter that offers several advantages over existing models: from a technical perspective, we integrated state of the art fluorescent proteins (mTq2, mVenus and mScarlet-I) with optimal brightness, spectral compatibility and photostability that should allow us to readily detect signal from each subpopulation and minimize technical artifacts. We purposely did not use GFP or its derivates so that this channel is available to combine the PRIME reporter with for example WNT reporters such as Axin2P2A-rtTA3-T2A-3xNLS-SGFP2 or stem cell reporters such as Lgr5EGFP-IRES-CreERT2.

From a biological perspective, we purposely chose to tether the fluorescent proteins to the membrane for a couple of reasons. Similar to existing models, the PRIME reporter can detect and visualize (stem) cell dynamics and competition. However, by tethering the fluorescent proteins to the membrane this reporter simultaneously also allows straightforward measurements of cell shape, migration and cell-cell interactions (Chapter 3, Fig2-3), as such we expect it to be useful for investigations into cell dynamics during embryonic development and in adult tissues.
Figure 2: Axin2 endogenous regulation and knock-in mouse models. A) Illustration of post transcriptional regulation of Axin2 mRNA transcripts by miRNAs that target the 3'UTR and a schematic illustration of 5'UTR and 3'UTR reporter inserts in the Axin2 locus for several knock-in mouse models. B) Cartoon illustrating the resulting mRNA transcripts from several Axin2 reporter models and their possible post-transcriptional regulation.

Moreover, the membrane signal generated by the PRIME reporter is sufficient to readily segment individual cells in both 2D and 3D within clones in vivo and in vitro for further downstream quantitative analysis.

When combined with a universal CreERT2 driver such as Rosa26\textsuperscript{CreERT2}, multicolour lineage tracing reporters, like PRIME, can be used for unbiased lineage tracing. This has previously been shown to be a viable strategy with the Confetti reporter to identify novel tissue dynamics in for example the mouse mammary gland. PRIME offers the additional advantage of better visualizing tissue morphology, cell migration, cell shape and cell-cell contact thereby offering a more complete view than (partial) nuclear reporters. Moreover, because the
membrane signal of individual cells can be readily segmented, more quantitative studies can be performed by either combining Rosa26<sup>PRIME</sup>, Rosa26<sup>CreERT2</sup> with e.g. a third signalling reporter or post mortem staining and quantification. Contrary to expectations, the PRIME reporter does not recombine in near-equal ratios and did not appear to recombine in all tissues even in combination with a universal driver such as Rosa26<sup>CreERT2</sup> (Chapter 3, Fig 4, supplementary Figure 1&3). It is unclear whether this discrepancy is due to expression of the PRIME reporter, suboptimal universal expression of Rosa26<sup>CreERT2</sup>, availability of active Tamoxifen per tissue, toxicity of CreERT2 and tamoxifen or a combination thereof. It does underscore that care should be taken when interpreting results from genetic reporters.

With the recent onset of single cell technologies, and in particular single cell barcode lineage tracing, one can wonder whether these developments render mouse models such as Axin2<sup>P2A-rtTA3-T2A-3xNLS-SGFP2</sup> and Rosa26<sup>PRIME</sup> obsolete. However, lineage tracing models, either marker-based or unbiased, are ideally suited to complement (or even be combined with) current barcoding technologies. These models can experimentally support pseudotime trajectories. Marker-based lineage tracing model can offer experimental context in scRNA-seq datasets where the stem cell driver is too lowly expressed or present in too few cells to detect with current single cell technologies (about ~10-30% of mRNA transcripts per cell can be detected depending on the technique performed). Unbiased or marker-based lineage tracing provides relational information for a population of cells that can be followed over time. Novel developments in the field of spatial transcriptomics offer a complementary approach where protein detection (immunofluorescence or immunohistochemistry based) and RNA expression (via e.g. a ~300 RNA probe set at subcellular level) can be combined on FFPE sections thereby offering spatial context to -omics based inventories (such as the 10X Xenium or Visium platforms). Combining these novel techniques with more traditional lineage tracing models will provide a complete picture of lineage and progeny, gene expression, cell shape and cell-cell interactions within either healthy tissues or tumour biology. In the context of the overarching question of this thesis, of particular interest will be additional bioinformatics analysis using packages that can infer cell-cell communication from scRNA-seq data (e.g. CellphoneDB 2.0) or spatial transcriptomics data (e.g. BulkSignalR) to better understand stem cell – niche cell interactions and provide receptor-ligand candidates that can be further tested experimentally.
Regulatory mechanisms of WNT ligand expression in potential niche cells in the mammary gland

The spatiotemporal regulation of stem cells by niche cells through secreted factors is essential for maintaining tissue homeostasis in adult tissues. WNT signalling has been implicated in almost all stages of mammary gland development. Numerous components of the WNT/CTNNB1 signalling pathway have been described at various stages of mammary gland development, including a wide range of WNT ligands, receptors, downstream effectors and transcription factors. Notably, WNT/CTNNB1 signalling is frequently dysregulated in breast cancer. However, unlike for example colorectal cancer where mutations accumulate at downstream components of the WNT pathway, in breast cancer most alteration occur at the level of the receptor-ligand interface. In chapter 4-7 we examined the (epi)genetic regulation of several sources of WNT ligands in the mammary gland.

Public resources to generate novel hypotheses

In Chapter 4 we illustrate how the development of user-friendly tools designed to probe public repositories of -omics datasets can be harnessed for in silico analyses of gene regulatory mechanisms and novel hypothesis generation. Chapter 5 follows up on these predictions and tests if they can be validated experimentally. Publicly available -omics datasets harbour a wealth of data that often remains largely unexplored in the corresponding publication and can be repurposed to generate novel hypotheses and address specific questions. As a case study to illustrate what can be inferred from public datasets without any prior expertise in analysing -omics datasets, we chose to investigate the gene regulation of a specific Wnt gene, Wnt7b, in the healthy mouse mammary gland. The publication of novel -omics atlases and tools to analyse them is ever increasing and thus likely Chapter 4 is already outdated upon print of this thesis. Nevertheless, it provides a clear example of what is possible in silico without generating your own datasets and prior bioinformatics expertise. However, if remotely possible I would encourage researchers to develop the necessary bioinformatic skills even when resources to perform these -omics experiments are not available in the lab. Dedicated reanalysis of -omics data published in public repositories such as NCBI GEO, ENCODE, ChIP-Atlas or databases maintained by labs and institutions themselves can provide novel insights beyond the scope of the existing user-friendly online tools used in chapter 4, as illustrated and exploited in chapters 5-7.
Hormonal regulation of WNT ligands in the mammary gland

In Chapter 5 & 6 we investigated the epigenetic regulation of two luminal specific WNT ligands, WNT7B and WNT4, in both mouse and human breast epithelial cells. These are of particular interest because the epithelial network of the mammary gland is known to expand in response to ovarian hormones estradiol and progesterone. In the ductal tree, only a subpopulation of luminal cells is hormone responsive (ER+/PGR+) \(^{53, 54}\). Moreover, WNT4 has previously been implicated as a direct target gene of PGR in hormone responsive luminal cells \(^{55-57}\). Side-branching of epithelial ducts during early pregnancy is completely ablated in \(Wnt4^{-/-}\) mice \(^{55}\). Thus, it is tempting to speculate that in response to estradiol and progesterone, hormone responsive luminal cells secrete WNT ligands that promote proliferation and thereby control the expansion of the epithelial network of the mammary gland \(^{53}\).

The genomic organisation of WNT loci in the mammary gland

In Chapter 5 we confirm and expand upon our initial hypothesis on \(Wnt7b\) gene regulation in the mammary gland from Chapter 4. By chromatin conformation capture techniques we identify the 3D organisation of the \(Wnt7b\) TAD and confirm that this is conserved in human breast cells. By our own 4C assay and reanalysing publicly available datasets we detect an intraTAD loop in both human and mouse cells that shows high interaction frequency and harbours species specific cis-regulatory elements (mCRS7-11 an hCRS5-11 respectively) that can be directly linked to the \(Wnt7b/WNT7B\) promoter. We termed this loop the \(Wnt7b/WNT7B\) locus. These cis-regulatory elements are accessible in cells that express \(Wnt7b/WNT7B\) in vivo (luminal progenitor and mature luminal cells) and are marked by H3K27 acetylation in both human and mouse breast cells.

In Chapter 6 we explored the epigenetic regulation of \(Wnt4/WNT4\) gene expression in mouse and human breast epithelial cells. By Hi-C and CTCF ChiaPET we identified a consensus \(Wnt4\) TAD in both mouse and human cells in which \(Wnt4/WNT4\) is surprisingly the only protein coding gene. Within this TAD we identified two sets of candidate cis-regulatory elements (mCRSs) that act either PGR-dependent or independent. Surprisingly, only one mCRS directly interacts with the \(Wnt4\) promoter, mCRS4. This enhancer-promoter loop is present regardless of \(Wnt4\) gene expression or cell type. We show that this loop is also established in mammary gland cells in vivo and that this loop is mediated by CTCF. Interestingly, other mCRSs that can induce \(Wnt4\) gene expression do so by intermedial looping to mCRS4 rather than the \(Wnt4\) promoter. A knockout of mCRS4 is sufficient to significantly reduce \(Wnt4\) gene expression. mCRS4 acts as a regulatory chromatin hub that
controls Wnt4 gene expression in the mouse mammary gland. We confirm that this regulatory chromatin hub is conserved in human breast epithelial cells by in-depth reanalysis of publicly available CTCF ChiaPET and scATAC-seq datasets. Deletion of hCRS7 (human mCRS4 equivalent) abrogates PGR-mediated WNT4 induction in MCF7 cells. Together both chapters provide a framework for the 3D organisation of WNT gene loci in the mammary gland and in-depth analysis of how the genomic organisation may contribute to spatiotemporal regulation of gene expression of WNT genes in both human and mouse mammary epithelial cells.

**Identifying novel transcriptional regulators of WNT genes in the mammary gland**

We initially developed a pipeline to identify novel transcriptional regulator of WNT genes in the mammary gland to better understand Wnt4/WNT4 regulation beyond PGR. In this manner we identified pioneer factor GRHL2 as potential novel regulators of the luminal lineage and of WNT4 in the human breast in chapter 6. We show that GRHL2 is enriched at hCRS7 (mCRS4-like) in MCF7 cells and co-localizes with histone methylase KMT2C. Moreover, known mature luminal lineage factors FOXA1 and PGR both associate with other hCRSs within the regulatory hub. PGR binds to hCRS8 and 9 (mCRS3-like) upon stimulation with R5020 and co-localizes with histone acetylase EP300 in MCF7 cells. We propose a working model where in both human and mouse luminal breast cells WNT4/Wnt4 expression is dependent on a regulatory chromatin hub that interacts with the Wnt4/WNT4 promoter by CTCF-mediated looping. This regulatory hub brings together a transcription factor complex that consists of GRHL2, FOXA1, PGR and histone modifiers. GRHL2 acts as a pioneering factor that binds to hCRS7 (mCRS4-like) in luminal progenitor cells. Together with KMT2C, GRHL2 primes the WNT4 regulatory hub by the addition of H3K4me1 marks. In mature luminal cells FOXA1 binds to hCRS9 (mCRS3-like). Upon exposure to progesterone under the influence of the hormone cycle, PGR and EP300 bind to hCRS8 and 9 in the regulatory hub. EP300 adds H3K27ac marks to cis-regulatory elements in the regulatory hub and thereby promotes WNT4 activation.

We applied the pipeline that was developed for WNT4 in chapter 6 to Wnt7b/WNT7B in chapter 5. In this manner we show that ESR1 becomes highly enriched at mCRSs within the Wnt7b locus in immortalised mouse mammary epithelial cells upon stimulation with Estradiol. We confirm that ESR1 is also enriched at hCRSs in the WNT7B locus in human breast epithelial cells. Moreover, GRHL1 and GRHL2, two pioneer factors identified in our screen, colocalize with ESR1 together with histone methylase KMT2C.
Figure 3: Hormone dependent regulation of WNT genes. A) Schematic showing the relative serum hormone levels of oestrogen and progesterone during the human menstrual cycle. B) Cartoon illustrating our proposed model of hormone dependent WNT regulation in the mammary gland. A transcription factor complex consisting of luminal specific transcription factors GRHL2, FOXA1, PGR and ESR1 and epigenetic modifiers activates specific WNT genes dependent on hormone availability during the menstrual cycle. An ESR1 and PGR switch within this complex drives different luminal transcriptional programs during different stages of the menstrual cycle. C) Illustration of PGR/ESR1 transcription factor complexes proposed by Mohammed et al. 2014.
GRHL2 has previously been shown to interact with ESR1 and KMT2C with GRHL2. Binding of GRHL1 and 2 to hCRSs in the WNT7B locus has functional consequences in human breast epithelial cells. A Knockdown of GRHL transcription factors in MCF7 cells reduces WNT7B expression on average by ~30% and abolishes DNA accessibility at hCRSs at which ESR1, GRHL1, GRHL2 and KMT2C associate. Gene expression of both ESR1 and GRHL1 and 2 is restricted to the luminal lineage. We propose a working model for WNT7B regulation in luminal breast epithelial cells where GRHL1 and 2 act as pioneering transcription factors that bind to hCRSs within the WNT7B locus in luminal progenitor cells. Together with KMT2C, GRHL1 and 2 prime the WNT7B locus for WNT7B gene activation by the addition of H3K4me1 marks. Upon exposure to Estradiol, ESR1 co-localizes at hCRSs within the WNT7B locus that are bound by GRHLs to induce H3K27ac and WNT7B transcription.

Taken together, Chapter 5 and 6 imply that this transcription factor complex might have broader implications beyond WNT4 and WNT7B regulation and controls luminal hormone dependent gene expression in human breast cells. Chapter 5 and 6 suggest that GRHL2, KMT2C (and FOXA1) can both associate with ESR1 and PGR, and that co-localisation of ESR1 and PGR with this complex is dependent on binding of their respective ligands. In the case of WNT4 and WNT7B, it seems that PGR and ESR1 binding to this complex is mutual exclusive (Fig3B), however, it has previously been reported that upon progesterone stimulation in MCF7 cells ESR1 and PGR can co-localize in a similar transcriptional complex with one of the two acting as the dominant factor (Fig3C). Interestingly, the same paper shows that PGR also associates with GATA3, a known luminal lineage factor, and TRPS1, a transcriptional repressor that inhibits GATA regulated genes. Together, most well-known luminal lineage factors seem to associate with PGR in this transcription factor complex, thus this suggests that this transcription factor complex may act as the main downstream regulator of luminal identity in the mammary gland.

The ESR1/PGR/GRHL2 transcription factor complex finetunes hormone dependent gene expression in luminal cells in response to varying systemic levels of oestrogen and progesterone during the menstrual cycle. We propose that hormone dependent WNT gene expression, and possibly hormone dependent signalling in general, is mediated by this transcription factor complex and that depending on the stage of the hormone cycle ESR1 and PGR are interchangeable (Fig3A-C). As such, this transcription factor complex acts to convert a systemic endocrine signal into a localized, paracrine proliferative signal.
In this transcription factor complex GRHL1 and 2 act as a novel class of luminal lineage specific pioneering factors that prime luminal progenitors and mature luminal cells for hormone-dependent transcriptional activation cooperatively with KMT2C. Interestingly, recent studies have shown that recruitment of the COHESIN complex at distal enhancers is dependent on GRHL2 during differentiation of embryonic stem cells and in ovarian cancer cell lines. Moreover, KMT2C is a histone methylase that adds H3K4me1 marks that primes enhancers for activation and has previously been shown to promote enhancer-promoter looping. Loss of GRHL expression in MCF7 cells abrogates chromatin accessibility in hCRS7 (Chapter 6, Fig6G). Thus, GRHLs pioneering activity functions two-fold: it opens the chromatin and primes enhancers via KMT2C and simultaneously (or consecutively) promotes enhancer-promoter loop formation by recruiting the COHESIN complex.

**Protein structure predictions of PGR/ESR1-GRHL2 interaction**

In Chapters 4-7 we focus our search on Wnt gene expression regulation using (epi)genetic -omics datasets and largely ignore proteomics or protein structures. Recent developments in protein structure prediction such as AlphaFold2, RoseTTAFold and ESMFold are worth mentioning and have already revolutionized Cryo-EM and protein crystal structure experiments by supporting experimental data and fuelling in situ novel hypotheses. For example, ColabFold, an open-source protein structure prediction platform, is able to predict protein-protein complexes up to a certain size limit. Although likely part of a larger complex of transcription factors and therefore providing an incomplete picture, we can run the ESR1-GRHL2 complex proposed in chapter 5 and PGR-GRHL2 complex proposed in chapter 6 in ColabFold to predict the structure of these complexes (Fig4A-B). In the future, this can help identify regions to mutate and disrupt these interactions experimentally. In accordance to what is expected based on literature, PGR and ESR1 are predicted to interact with the transactivating domain of GRHL2 and specific amino acids that have previously been shown to disrupt the activity of this transactivating domain are predicted to interact with PGR and ESR1 (Fig4E-F). Interestingly, ColabFold predicts similar binding of GRHL2 to both ESR1 and PGR suggesting that this binding might be mutually exclusive and thus competitive between both steroid receptors (Fig4A-B). Notably, these predictive methods still suffer several drawbacks. There is a limit to the size of complexes ColabFold can predict (the reason why we try to predict monomer PGR/ESR1-GRHL2 complexes although both steroid receptors and GRHL2 are known to act as dimers) and none of the aforementioned techniques can as of yet predict the effects of ligand...
or DNA binding on the protein (complex) structure. First steps are being taken to mitigate this problem: AlphaFill, an AlphaFold2 variant that can integrate small molecule cofactors and ions into AlphaFold2 predicted structures, was recently published (Fig4C-D). Based on these predictions and the experimental evidence from chapter 5 and 6, a follow-up of PGR/ESR1-GRHL2 is warranted, both in terms of protein complex composition and dynamics upon changes in hormone status.

**Dysregulation of the GRHL2/ESR1/PGR/FOXA1 complex in breast cancer**

Most components of this transcription factor complex are frequently dysregulated in luminal breast cancer. Dysregulation of ESR1, PGR and FOXA1 are known characteristic of luminal A breast cancer and KMT2C is mutated in 10.2-12.8% of luminal breast cancer subtypes. GRHL2 is amplified in 12.2% of all luminal breast cancer cases. GRHL2 promotes proliferation and inhibits epithelial to mesenchymal transition in breast cancer. GRHL2 deletion in luminal breast cancer cell lines causes cell cycle arrest and in a mouse transplantation breast cancer model GRHL2 silencing reduced primary tumour growth and a reduction of metastasis, in agreement with its negative association with patient survival. Taken together, this highlights that GRHL2 amplification has functional consequences for luminal breast cancer and implies that dysregulation of this transcription factor complex might also be a defining feature of luminal breast cancer (Fig5).

**Stromal sources of WNT in the mammary gland**

In Chapter 7 we investigated the epigenetic regulation of Wnt2/WNT2, a WNT ligand that is secreted by the mammary gland stroma. Stromal sources of WNT, and particularly WNT2, have previously been implicated to contribute to stem cell maintenance in the mouse mammary gland. Wnt2 expression is enriched in mammary gland fibroblasts, and in situ hybridisation experiments have demonstrated that stromal Wnt2 is mostly expressed at the stromal-epithelial border. Thus, WNT2 is ideally localised to act as a local stem cell niche factor.

In both mouse and human (mammary) fibroblast cells we detect an intraTAD loop, dubbed the Wnt2/WNT2 locus, that harbours Wnt2/WNT2 by Hi-C and CTCF ChiaPET. Within this loop we identify mCRS12, a candidate cis-regulatory element that displays enhancer activity in luciferase assays, can specifically activate Wnt2 in mammary associated fibroblasts (MAF) once activated by CRISPRa and closely associates with the Wnt2 promoter within the Wnt2 locus in MAF cells but not mammary epithelial cells. We used our pipeline from Chapter 6 to identify potential transcriptional regulators of Wnt2/WNT2 in breast fibroblasts.
General discussion

A

B

C

D

E

F

GRHL2

GRHL2

GRHL2

GRHL2

GRHL2

GRHL2

GRHL2

Progestrone

Progesterone

PGR

PGR

PGR

PGR

PGR

PGR

PGR

8
Figure 4: Protein folding algorithms predict PGR/ESR1:GRHL2 interacting domains. A-B) ColabFold predicts monomeric A) PGR-B-GRHL2 complex and B) ESR1-GRHL2 complex and identifies common interacting domain in GRHL2 based on protein sequence. Protein structure was visualised in the PDB. Protein sequence was downloaded from UniProtKB (as of 16/03/2023). PGR-B: P06401, ESR1: P03372, GRHL2: Q6ISB3. C-D) AlphaFill predicts C) progesterone binding in PGR based on protein sequence and the chemical structure of progesterone, and D) estradiol binding in ESR1 based on protein sequence and the chemical structure of estradiol. (E) Protein sequence GRHL2: PGR-B complex. Transactivating domain from GRHL2 is indicated in blue and DNA binding domain in yellow. Previously published residues important for transactivating activity by GRHL2 are indicated by the asterisk and both these residues and their interacting partner(s) in PGR-B are highlighted in red. F) Protein sequence GRHL2:ESR1 complex. Transactivating domain from GRHL2 is indicated in blue and DNA binding domain in yellow. Previously published residues important for transactivating activity by GRHL2 are indicated by the asterisk and both these residues and their interacting partner(s) in PGR-B are highlighted in red.

Figure 5: Genetic alteration of the GRHL / Hormone receptor complex in breast cancer. Frequency of genetic alterations in breast cancer of the proposed components of the GRHL / Hormone receptor complex. Alteration frequency is derived CBioportal, PanCancer Atlas database.

Breast Cancer
- Mutation
- Structural Variant
- Amplification
- Deep Deletion
- Multiple Alterations
In this manner we identified FOSL1 as a potential candidate for \textit{Wnt2/WNT2} transcriptional regulation. FOSL1 expression is enriched in mammary stromal cells of the human breast (Chapter 7, Fig 4). Note that unlike in the human breast, Fosl1 is more broadly expressed in mouse breast cells, particularly in basal cells (Chapter 7, Fig 4). mCRS12 contains a FOSL1 binding motif and although no mammary fibroblasts ChIP-seq datasets are available, ChIP-seq data from mouse embryonic fibroblasts (MEF) show enrichment of FOSL1 at mCRS12. The mCRSs do not display strong sequence conservation and do not show epigenetic features associated with active cis-regulatory elements in human mammary fibroblasts. However, the \textit{WNT2} locus does display strong enrichment in both mammary fibroblast specific chromatin accessibility and H3K27 acetylation. At these human cis-regulatory sites FOSL1 binding is enriched (as shown other human tissues since no human mammary fibroblast data is available).

Interestingly, Fosl1/FOSL1 is a direct target gene of WNT/CTNNB1 signalling \textsuperscript{90-92}. As already described earlier in this chapter, a subset of mammary fibroblasts expresses \textit{Axin2} and responds to WNT/CTNNB1 signalling. Moreover, these \textit{Axin2}+ fibroblasts express \textit{Wnt2}, and this is dependent on the activation of canonical WNT signalling \textsuperscript{12}. It is tempting to speculate that \textit{Wnt2} activation in \textit{Axin2}+ mammary fibroblasts is dependent on FOSL1 binding to mCRS12 and that \textit{Fosl1} expression in these cells is mediated by activation of WNT/CTNNB1 signalling. Basal cells in the mouse mammary gland are also known to express \textit{Axin2} and respond to active canonical WNT signalling, thus possibly explaining \textit{Fosl1} expression in this cell population, although here it doesn’t result in \textit{Wnt2} expression (Chapter 6, Fig1B, Chapter 7, Fig4). It has previously been reported that \textit{Wnt2} expression is enriched in \textit{Gli2}+ mouse mammary gland stromal cells and that \textit{Gli2} is required for \textit{Wnt2} expression in these cells \textsuperscript{93}. In our transcription factor identification pipeline we confirm that GLI2 expression and activity are also enriched in human breast fibroblasts (Chapter 7, Fig4), but we did not detect enrichment of GLI2 transcription factor binding motifs in any of our candidate regulatory sequences of \textit{Wnt2/WNT2}. This suggests that GLI2 is important for \textit{Wnt2} expression, but may not directly control \textit{Wnt2} transcription. Rather, we suggest that GLI2 acts upstream in the transcriptional network that controls \textit{Wnt2} expression and controls gene expression of the transcriptional regulators that directly control \textit{Wnt2} expression in mouse mammary fibroblasts.
FOSL1 has previously been reported to be strongly enriched at enhancers in triple negative breast cancer cells, where it is a key factor and promotes proliferation and cancer cell viability. FOSL1 induces epithelial to mesenchymal transition in breast cancer cells and is associated with a poor prognosis in triple negative breast cancer patients. Whether FOSL1 dysregulation affects WNT2 expression and if altered expression of WNT2 has functional consequences in breast cancer is unclear and warrants further investigation.

Concluding remarks

In this thesis we have developed and characterized novel mouse models that can help better understand WNT signalling in stem cells. To our knowledge we generated the first Axin2 reporter that fully recapitulates endogenous Axin2 expression. Additionally, we characterized a novel multicolour lineage tracing reporter, Rosa26PRIME, which like previously published models can trace multiple clones simultaneously and visualize stem cell dynamics but also offers the additional advantage that it can detect cell shape and cell-cell interaction. Individual cells within clones can be easily segmented for further quantitative analysis.

We have shown for the first time a detailed structural and functional dissection of Wnt/WNT loci in mouse and human breast cells. This has provided us new insights in how cell type specific Wnt/WNT expression patterns are established in mammary gland cells. We have identified novel (candidate) transcriptional (co-)regulators that mediate spatiotemporal Wnt/WNT in mouse and human breast cells. We show that a complex consisting of GRHL transcription factors, KMT2C and cofactors controls Wnt7b/WNT7B and Wnt4/WNT4 interchangeably with ESR1 and PGR in a hormone dependent manner in luminal breast cells. Moreover, we have identified FOSL1 as a potential candidate transcriptional regulator for Wnt2/WNT2 in mammary gland fibroblasts.

Taken together, we initiated the first steps to better understand the spatiotemporal gene regulation of WNT ligands upstream of WNT/CTNNB1 signalling in the mammary gland. This has provided novel insights in how proliferative WNT signals are regulated in the mammary gland and hormone signalling in luminal epithelial breast cells and results in a first framework for how WNT gene expression in the stem cell niche may be controlled.
General discussion

References


Chapter 8


