Navigating progestrone-driven gene regulatory networks in the mammary gland

Aarts, M.T.

Publication date
2025

Chapter 6

General discussion

In this thesis, we explored the chromatin-level mechanisms underlying progesterone-driven gene regulation, with a particular focus on how enhancer elements orchestrate spatiotemporal control of target gene expression. Although progesterone signaling is essential for tissue development and homeostasis, the molecular processes by which the progesterone receptor (PR) engages with chromatin, cooperates with cofactors, and modulates gene expression within the 3D genome architecture remain incompletely understood. Moreover, studies that dissect the transcriptomic regulation of a single target gene across different spatial and temporal contexts are scarce. To address these knowledge gaps, we first developed new tools and optimized conditions to more accurately measure PR signaling (**Chapter 2**). We then explored the tissue- and developmental stage-specific regulation of *Wnt4* by functionally dissecting and comparing the (PR-responsive) *Wnt4* enhancer landscape in the mammary gland, lung, and kidney (**Chapters 3 and 4**). Finally, we identify GRHL2 as a key co-regulator of PR and dissect how PR and GRHL2 co-regulate gene expression within the context of 3D genome organization (**Chapter 5**). Together, these studies provide novel insights into how progesterone signaling operates at the chromatin level to regulate gene expression. They also uncover the complexity of enhancer logic that governs tissue-specific and temporally restricted responses. In this final chapter, I reflect on the broader implications of the research presented in this thesis and highlight remaining questions and future research opportunities.

Advancing the toolkit for progesterone receptor signaling: challenges and opportunities

As outlined in **chapters 1 and 2**, progesterone and the progesterone receptor (PR) remain significantly less studied than estrogen and estrogen receptor (ER) signaling. Consequently, fewer dedicated tools are available to investigate PR signaling, and therefore our molecular understanding of how progesterone regulates cellular processes remains incomplete.

To address this knowledge gap, **chapter 2** presents new strategies for measuring PR signaling. We first optimized cell culture conditions and demonstrated that a certain level of PR expression is needed for proper PR signaling measurements. Additionally, we once more confirm that the use of steroid-depleted (i.e. stripped) serum, enhances sensitivity in PR activity assays. Moreover, we developed and characterized novel PRE-driven luciferase reporters. Among these, the 4xPRE-luciferase reporter proved to be the most sensitive and consistent tool for detecting exogenous PR activity. We also generated PRE-GFP reporter constructs to visualize PR signaling at the single-cell level. Both reporter systems respond to physiological concentrations of the synthetic progestin R5020 and reveal the strong context dependency of PR signaling. Together, these tools and optimized conditions represent valuable resources for the field and will facilitate further investigation into the molecular mechanisms of PR action. In fact, the optimized culture conditions, cell models, and reporter assays developed in **chapter 2** were essential for enabling the new insights described in **chapters 3 and 5** and our constructs that are available on Addgene are frequently requested by the community.

Beyond the availability of molecular tools, a major limitation in PR research is the lack of suitable *in vitro* models[1]. Currently, no PR-responsive mouse mammary epithelial cell lines exist, and only two frequently used human PR-responsive breast cancer cell lines, MCF7 and T47D, are available. These lines, however, do not reflect the characteristics of normal, non-transformed PR responsive breast epithelial cells. Non-transformed PR-responsive human breast or mouse mammary epithelial cells do not exist yet, likely due to the rapid loss of hormone receptor expression during *in vitro* culture of both mouse and human primary mammary epithelial cells. In our own experiments, we even noticed that PR target gene expression was already reduced during epithelial cell extraction, prior to actual culturing. This resulted in a lack of induction of PR targets such as *Wnt4* and *Tnfsf11* (RANKL) in pregnant mouse mammary samples, even though these genes were clearly induced when measured in whole tissue samples (**Chapter 4**). This highlights the urgent need for non-transformed, PR-responsive human breast epithelial cell lines, as well as models representing the full diversity of epithelial subtypes in the breast.

As an alternative to cell lines, recent progress in the mammary and breast organoid field shows potential to study hormone signaling in a more physiological setting. Some culture protocols now allow organoids to maintain expression of PR and, under certain conditions, enable experimental induction of PR target genes[2–7] (**Chapter 4**). In our studies, mouse mammary organoids responded to R5020 stimulation; however, the magnitude and consistency of the response varied considerably. Also, human breast epithelial organoids still exhibit extreme variability, both between donors and between organoids derived from the same donor[8] (unpublished work from our lab). While organoids hold great promise as models for studying PR signaling, current protocols require further optimization to ensure reproducibility and physiological relevance.

Deciphering *Wnt4* regulation: variability and context dependence in tissues and cell lines

In **chapters 3 and 4**, we explored the spatiotemporal regulation of *Wnt4* across the mammary gland, lung, and kidney. To our knowledge, this represents one of the first detailed studies to map the regulatory landscape of a single gene across multiple tissues and developmental contexts. In **chapter 3**, we focused on the postnatal mammary gland and described the mapping and functional dissection of the mouse and human mammary *Wnt4* enhancer repertoire. While multiple enhancers appeared species-specific, the overall/larger 3D regulatory chromatin hub architecture was conserved. Specifically, we identified a conserved distal enhancer, mgCRS4, which contains a CTCF-binding site and stably loops to the *Wnt4* promoter in both mouse and human cells. This enhancer appears to facilitate formation or maintenance of a chromatin hub that bridges the linear distance between distal (progesterone-responsive) regulatory elements and the *Wnt4* promoter. PR and GRHL2 bind to different enhancer elements but both converge on this chromatin hub in 3D space to jointly regulate *Wnt4* expression.

In **chapter 4**, we applied the enhancer identification pipeline developed in **chapter 3** to map the *Wnt4* regulatory landscape in two additional tissues: the kidney and the lung. This analysis included both adult tissues and embryonic developmental stages, offering a unique opportunity to compare enhancer usage across tissue and temporal contexts for a single gene. In the mouse, we identified 12 mammary gland, 10 kidney, and 8 lung enhancers, all distributed throughout the *Wnt4* TAD. The majority of these enhancers were tissue-specific: only five elements were shared between two tissues, and just one enhancer, mgCRS4/kCRS5/luCRS4, was identified in all three (**Chapter 4**). Importantly, the CTCF-mediated looping interaction observed between mgCRS4 and the *Wnt4* promoter in the mammary gland is also predicted to be present in the embryonic lung and adult kidney.

Because mgCRS4 overlaps with kCRS5 and luCRS4, we leveraged our mgCRS4 *in vivo* enhancer knockout mouse model (**Chapter 3**) to assess the functional importance of this element in all three tissues. In the mammary gland, mgCRS4 deletion had no detectable effect on *Wnt4* expression, even when tested across multiple developmental timepoints and under diverse exogenous hormone treatments (**Chapter 3 and 4**). Similarly, in the kidney, *Wnt4* expression was unchanged at both E14.5 and in the adult. In the lung, we did not observe significant expression differences in adult tissue, though a trend toward decreased expression was observed. However, in the E14.5 lung,

6

homozygous luCRS4 knockout mice exhibited significantly reduced *Wnt4* expression compared to wild-type controls, confirming a role for luCRS4 in the embryonic lung. Despite this change in gene expression, no overt phenotypic abnormalities were observed in gross embryonic or adult lung morphology and function.

The absence of detectable effects in the mammary gland and kidney could be attributed to two, non-mutually exclusive mechanisms. First, although mgCRS4/kCRS5/luCRS4 forms a stable, CTCF-mediated loop with the *Wnt4* promoter, likely bringing distal regulatory elements into proximity, this interaction may not be strictly required for enhancer-promoter communication. CTCF is known to stabilize such interactions and promotes robust gene expression in other contexts[9], but its loss does not always result in substantial expression changes[10–12]. Indeed, previous studies have shown that CTCF and Cohesin knockdown often causes only modest transcriptional effects[10–12]. Moreover, essentially even so-called 'stable' CTCF-CTCF loops are generally rare, occurring only a few percent of the time[13], suggesting that deletion of the CTCF-mediated loop within the *Wnt4* TAD may increase expression variability rather than abolish expression altogether. Although we did observe high variability in *Wnt4* expression, it remains challenging to distinguish whether this reflects true biological consequences of the deletion or is driven by technical or biological noise.

Second, enhancer redundancy may compensate for the loss of mgCRS4/kCRS5/luCRS4. This phenomenon, wherein multiple enhancers function cooperatively or redundantly to maintain robust transcriptional output, is well-documented[14]. This limited gene dosage effect, together with compensation, and ultimately a rescue in gene expression at later developmental stages therefore likely explain the lack of anatomical or physiological abnormalities in homozygous animals. Our *in vivo* data also suggest that *Wnt4* expression is tightly regulated and only modestly inducible, up to ~2.5-fold during early pregnancy or following extended progesterone exposure, whereas other PR target genes such as *Tnsf11* (RANKL) and *Rspo1* were induced ~12-fold and >40-fold, respectively (**Chapter 4**). These findings support the hypothesis that even small changes in *Wnt4* expression could have significant physiological consequences in the mammary gland. However, whether increased *Wnt4* levels lead to enhanced side-branching remains controversial, as studies have reported conflicting outcomes upon ectopic expression[15,16].

Given the marginal expression changes and high variability observed in our *in vivo* models, it is possible that subtle but biologically relevant effects of enhancer deletion are being masked – or that the critical experiment is yet to be performed. Additionally, *Wnt4* expression is confined to relatively low abundant subpopulations of cells within adult mammary, kidney, and lung tissues, resulting in difficulties in measuring differences in gene expression levels when whole tissues are used for bulk expression analyses. Of course, this would also further complicate the interpretation of enhancer function *in vivo*. Although including additional timepoints and replicates could help reduce noise and increase the statistical power to convincingly measure small effect sizes, this would require substantially more animals, which raises ethical and practical concerns.

In contrast to the *in vivo* situation, *Wnt4* induction was robust in cell lines with high PR expression. In human MCF7/PR and T47DS cells, *Wnt4* expression can be increased up to ~5-fold (although on average it is also most often in the 2-fold range), and baseline expression is also high in the non-PR-responsive mouse BC44 cell line (**Chapters 2 and 3**). In **chapter 3**, we generated mgCRS4 enhancer knockouts in both the mouse BC44 mammary cells and human MCF7 breast cancer cells. While the deleted sequences differed slightly, all knockouts disrupted the GRHL2-binding site, and in the case of MCF7, the CTCF site as well. Notably, both models exhibited functional, if slightly different consequences: baseline *Wnt4* expression was reduced ~2-fold in BC44 cells, and progesterone-induced *Wnt4* expression was abolished in MCF7 knockouts. We cannot definitively explain the discrepancies in transcriptional responses between cell lines and the *in vivo* models based on our current data. However, cultured cells differ markedly from their *in vivo* counterparts in cell identity, signaling environment, and chromatin organization. Differences in growth factor availability, transcription factor activity, chromatin-modifying enzymes, and even chromosomal composition likely underlie the observed divergence in *Wnt4* regulation. Putting mouse/human differences aside, it seems that on the one hand full progesterone responsiveness may require multicellular interactions[17], while on the other hand an intact developmental context and cellular plasticity offer buffering strategies that are lacking in cultured cell lines but are critical to ensure developmental robustness *in vivo*[18].

Lessons from public omics data: reproducibility, reuse, and reality

In recent years, genome-wide technologies—genomics, epigenomics, transcriptomics, metabolomics, and proteomics, collectively referred to as omics approaches, have evolved rapidly. While still relatively expensive for those with limited financial resources, these techniques are becoming increasingly affordable, less time-consuming, and thus more accessible to the broad scientific community. Most omics datasets are deposited in public repositories, such as the NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), either in raw or processed formats. In addition, larger initiatives such as the Encyclopedia of DNA Elements[19] (ENCODE, https://www.encodeproject.org) and ChIP-Atlas[20] (https://chip-atlas.org) curate and re-analyze collections of datasets and make them available in multiple formats. Numerous research groups also dedicate time to developing user-friendly online tools that allow other researchers to easily explore these datasets with a user-friendly interface and to generate new hypotheses[21].

Although data sharing has clear theoretical and practical advantages, in reality, researchers still often prefer to generate their own datasets. As a result, public datasets remain underused, and their integration into new projects is limited. The majority of contemporary research papers include newly generated omics data, yet too frequently only a few top candidates are followed up, or relatively narrow questions are addressed. Consequently, a wealth of existing data remains largely unexplored, even though they represent a valuable and cost-efficient resource for the broader scientific community.

In **chapters 3, 4, and 5**, we actively leveraged public omics resources, downloading a wide range of datasets from NCBI GEO, ENCODE, ChIP-Atlas and other repositories in both raw and processed formats. However, using these datasets came with multiple challenges. First, no universally accepted or 'gold standard' data analysis workflows currently exist. Analysis pipelines vary greatly between studies, and as computational tools evolve rapidly, results produced by different or even updated versions of similar pipelines can differ considerably. Moreover, some pipelines become outdated or unsupported, making it difficult if not impossible to precisely reproduce the results presented in the papers for which the omics datasets were first generated[22]. Furthermore, in many cases, the documentation and metadata accompanying deposited datasets is incomplete, unclear, or missing altogether, which hampers reproducibility. Processed data formats are often poorly annotated, making it difficult to use them as a starting point for further analyses. Additionally, not all researchers have the necessary bioinformatics expertise or access to computational infrastructure required to process raw data independently. We encountered all of these issues firsthand during our research and we strongly advocate that researchers depositing data in public repositories indeed strive to work according to FAIR data principles, such that their data are indeed suitable for re-analysis and re-use.

Even when properly documented and stored, technical or biological hurdles can emerge. As an illustrative example, we encountered difficulties when analyzing PR ChIP-seq datasets in our attempt to generate a high-confidence PR peak set. We downloaded 31 PR ChIP-seq datasets from ChIP-Atlas[20], all generated in T47D cells, and analyzed these datasets using the exact same pipeline. To select the high confidence peaks, we looked for common peaks across all datasets by counting the number of datasets in which a certain peak was called.

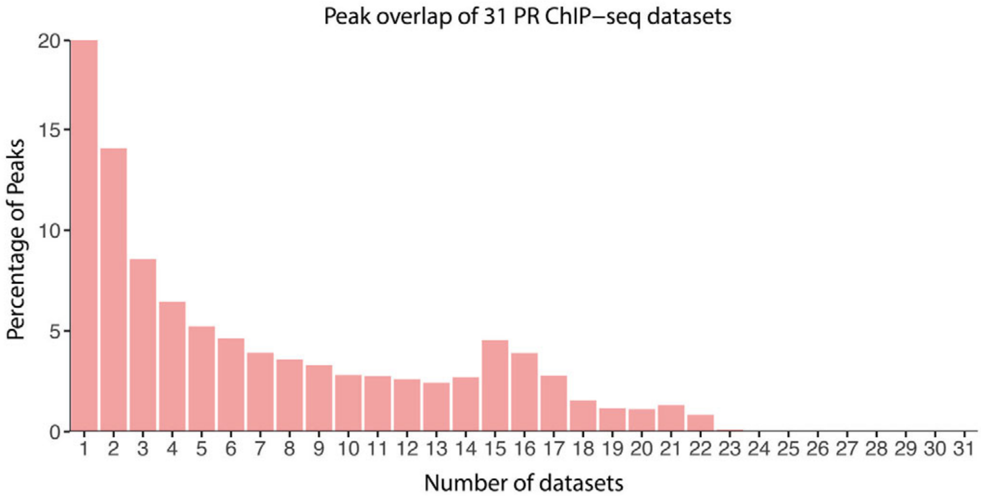


Fig1. Peaks of 31 publicly available PR ChIP-seq datasets poorly overlap
Bar graph depicting the ChIP-seq peak overlap of 31 publicly available PR datasets downloaded from ChIP-Atlas[20]. Peaks were defined as overlapping if they shared at least one base pair. Percentage of peaks was calculated using the accumulative number of peaks of the 31 ChIP-seq datasets.

We were initially surprised by the low overlap between datasets, even among replicates from the same experiment. Notably, approximately 20% of peaks were found in only one of the 31 datasets, and ~71% were present in fewer than one-third of them (<10 of the 31 datasets) (**Fig1**). This low reproducibility has also been noted for CUT&RUN analysis and inspired the development of the ICEBERG method, which uses a large number of replicates to discover true binding events[23]. Of course, when using publicly available data, despite the common target (here: PR), cell line (here: T47D) and analysis pipeline (here: pipeline by ChIP-Atlas), substantial differences will exist between datasets from different labs in terms of ChIP-seq protocols, antibodies used for pull down, ligands and concentrations used for stimulation, cell culture media, treatment durations, and other experimental parameters. These factors invariably contributed to the variability we observed but still, this variability is concerning, complicates downstream analyses and raises critical questions about data reliability: which peaks can be considered "true"? Which dataset should one prioritize for analysis? And what biological conclusions can be drawn from a peak observed in a single dataset? Because of the inherent variability, it is tempting to generate your own dataset to be 'certain' of the PR genomic interactions under your specific conditions. However, if even experimental replicates show minimal overlap[23], does this not simply add to the growing collection of poorly reproducible data? Also, what drives this lack of reproducibility? Are these limitations intrinsic to the techniques, or is the biology even more dynamic than we currently appreciate? Each of these is a fair question, yet none can be definitively answered at present. Ultimately, for **chapter 5** we selected a ChIP-seq dataset that came from the same cell type, was stimulated with the same physiological R5020 concentration as we used for our RNA-seq and came from the same lab as the PR HiChIP dataset we used.

While publicly available datasets are an invaluable asset, their effective reuse is hindered by the challenges just mentioned. The development and adoption of standardized analysis pipelines, consistent file formats, and better annotation practices would significantly improve the utility and reproducibility of public omics data. I also encourage researchers to deal with the above-mentioned challenges and not cast aside, but make greater use of existing datasets, at least for hypothesis generation, instead of defaulting to the generation of new datasets for every project or publication. Additionally, during my PhD, I invested time into developing a basic level of bioinformatic proficiency, which opened many doors in my research. This not only enabled me to re-analyze public datasets and strengthen my own findings, but also fundamentally changed how I interpret the results presented in other studies. In my opinion, a basic level of bioinformatic literacy is becoming essential for modern biological research and should be more prominently integrated into academic training.

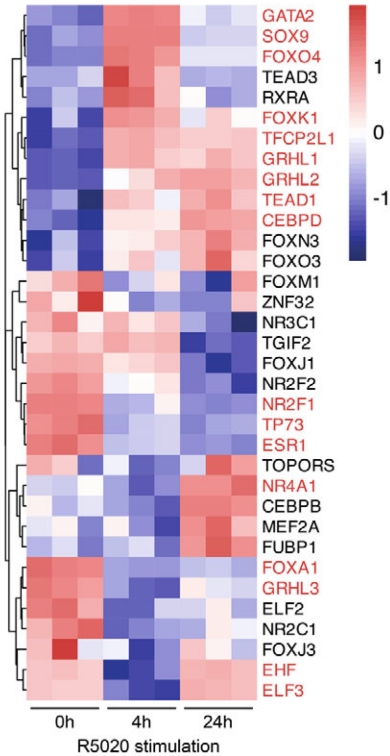


Fig2. Progesterone regulates a wide array of breast related transcription factors
Heatmap of bulk RNA-seq data from T47DS human breast cancer cells, showing unsupervised clustering and expression changes of 44 candidate *WNT4* regulatory transcription factors in response to treatment with R5020 for either 4 or 24 hours. Expression values depicted as z-scores for n=3 replicates.

Transcription factor dynamics downstream of progesterone signaling

In **chapter 3**, we identified PR as a transcriptional regulator of *GRHL2* expression, suggesting that PR may regulate the expression of its own co-regulators. In the same chapter, we presented a set of 44 transcription factors with expression levels that correlated positively with chromatin accessibility of their respective binding motifs in human breast cells. To narrow down this list, we focused on transcription factors restricted to the luminal lineage, resulting in seven candidates, including *GRHL1* and *GRHL2*, which we explored further (**Chapter 3**). Later, using RNA-seq in T47DS cells, we tested whether these seven luminal cell-restricted transcription factors were directly regulated by progesterone signaling. We found that all seven candidates responded directly to PR activation: *SOX9*, *GRHL1*, and *GRHL2* were upregulated, whereas *FOXA1*, *ESR1*, *EHF*, and *ELF3* were downregulated. This begs the question whether additional transcription factors from the original group of 44, those not necessarily luminal-enriched, might also be regulated by PR.

Indeed, a heatmap including all 44 transcription factors identified as both expressed and motif-accessible in human breast cells, shows that, in addition to *GRHL1*, *GRHL2*, and *SOX9*, several other transcription factors are also directly upregulated following four hours of progesterone treatment, including *GATA2*, *FOXO4*, *FOXK1*, *TFCP2L1*, *TEAD1*, and CEBPD (**Fig2**). Moreover, *FOXO4*, *FOXK1*, *TFCPL1*, *TEAD1*, and *CEBPD* remain elevated after 24 hours of treatment (**Fig2**). Conversely, *NR2F1*, *TP73*, *NR4A1*, and *GRHL3* are directly downregulated (**Fig2**). These factors may thus play previously unrecognized roles in the transcriptional response to progesterone in the mammary gland and are worthy of follow up. Of particular interest, *TEAD1* is not only regulated by PR but TEAD motifs were also enriched under GRHL2 ChIP-seq peaks as well as under PR/GRHL2 co-bound regions (**Chapter 5**), suggesting potential crosstalk between PR, GRHL2, and TEAD factors (**Fig2**). It is important to note that while the RNA-seq data were obtained from the T47DS breast cancer cell line, the initial expression and motif accessibility data were derived from tissue-derived human breast cells. As such, baseline expression levels of these transcription factors may differ between these different conditions.

Interplay between hormone receptors and co-regulators: competition or cooperation?

Following our observations in **chapter 3** that both PR and GRHL2 regulate *Wnt4*, we further explored the genome-wide co-regulatory mechanisms between PR and GRHL2 in **chapter 5**. We demonstrated that PR and GRHL2 physically interact independent of progesterone and co-regulate target genes by binding distal enhancer regions that engage in long-range chromatin looping. A similar mechanism has been described for ER and GRHL2[24–26], supporting a broader role for GRHL2 as a pioneering factor involved in both estrogen and progesterone signaling. In physiological settings such as the estrous/menstrual cycle and pregnancy, periods of high estrogen and progesterone levels overlap. This raises key questions: how is GRHL2 distributed between ER and PR during dual hormone exposure (**Fig3A**)? Do ER and PR share additional co-factors and are these distributed in a similar manner? How do PR and ER modulate each other's activity under these conditions? How does protein complex composition and enhancer occupancy change with rising and falling levels of estrogen and progesterone? More broadly, do steroid hormone receptors generally share pioneering factors or other co-factors? While these questions remain to be fully resolved, they point to a fascinating area of hormone regulated gene expression with potential implications for breast cancer biology.

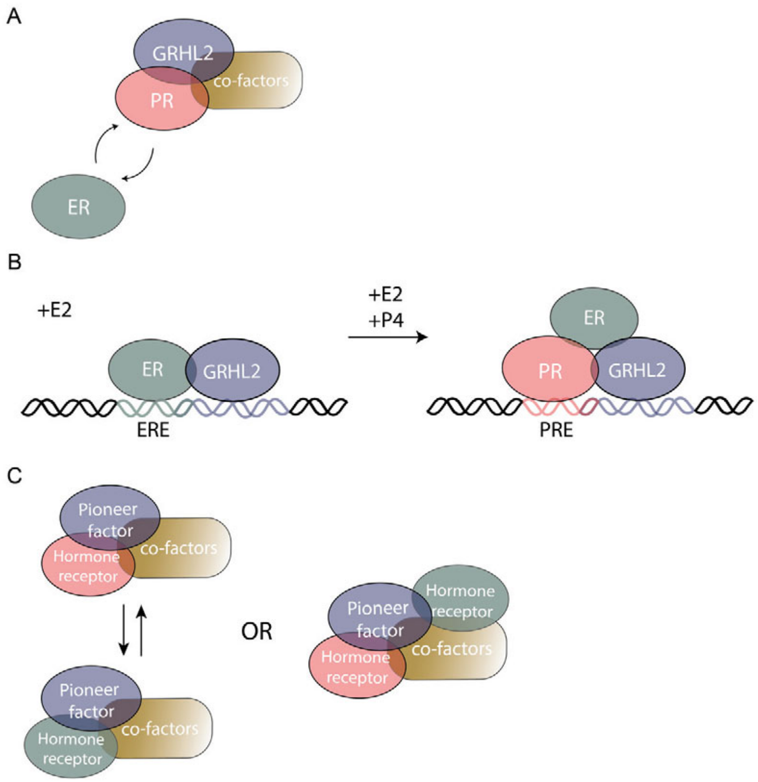
One intriguing question is if ER and PR physically interact. Prior studies have shown that PR and ER can interact in the presence of both hormones[27,28]. Upon estrogen presence, additional progesterone treatment can redirect ER to PR binding sites, thereby reshaping the gene regulatory landscape in favor of progesterone responsive transcription[28,29]. This observation has led to the idea that PR activation, in the presence of estrogen, antagonizes estrogen signaling[28,30,31], possibly by also sequestering shared co-factors such as GRHL2 in addition to ER itself (**Fig3B**). However, other reports suggest that progesterone can promote estrogen signaling[32], whereas no clear changes in estrogen responsive gene expression have been reported, and studies looking into the antagonistic effects are all performed in the context of breast cancer, rather than normal breast physiology[28,29,33,34]. Nonetheless, hormone receptor interplay is not unique to PR and ER. Other nuclear hormone receptors, such as the androgen receptor (AR)[35] and glucocorticoid receptor (GR)[36], are also reported to modulate ER function[37]. For instance, in dual estrogen-androgen stimulatory conditions, ER can be redirected by AR to androgen-responsive loci, resulting in anti-proliferative effects in breast tissue[35]. Additionally, PR/GR interaction and interplay has been described in dual progesterone and glucocorticoid conditions[38]. This highlights the need for a better understanding of the interaction between the different hormone receptors in breast tissue in general, with PR still being particularly understudied.

Relatively little is known about how pioneering factors and co-regulators behave under dual hormone conditions. GRHL2 is a known shared pioneering factor of PR (**Chapter 5**) and ER[24–26], and the luminal pioneering factor FOXA1 is a shared between ER and AR[39–41]. While it is clear that hormone receptors share transcriptional co-regulators, it is not yet established whether the presence of multiple hormones leads to exclusion of one receptor from key interactions, or whether co-factors are bound simultaneously (**Fig3C**). The latter scenario seems more plausible, as it would allow for rapid switching between transcriptional programs in response to changing hormonal cues. In such a model, ER, PR and other hormone receptors may dynamically divide access to a limited pool of shared co-factors, modulating each other's activity through competitive binding. Our own hormone-depleted RIME and co-immunoprecipitation experiments revealed a robust interaction between PR and GRHL2, while no interaction between ER and GRHL2 was detected under either these unstimulated conditions or upon progesterone treatment (**Chapter 5**). This suggests that PR–GRHL2 may represent a default interaction, which is possibly maintained in the presence of both hormones. In contrast, under estrogen stimulated conditions, ER may potentially displace PR from GRHL2 to form ER–GRHL2 complexes. Definitively testing this model will require systematic immunoprecipitation assays across a range of hormonal conditions to dissect the dynamics of receptor–co-factor associations, as well as a better understanding of how stable these different protein (sub)complexes are or whether temporary local high concentrations in 3D chromatin space allow them to be detected as such, as depicted in our working model for *Wnt4* gene regulation in **chapter 4**.

In summary, nuclear hormone receptors frequently rely on shared pioneering factors and transcriptional co-regulators, such as GRHL2, to orchestrate gene regulatory programs. However, exposure to multiple hormones may reallocate these co-factors, directing them to distinct target genes based on the hormonal context. This redistribution may serve as a key mechanism through which hormone receptors co-operate and fine-tune gene expression. Future research is needed to determine whether co-factors are simultaneously engaged by multiple receptors or selectively reallocated. Exploring these scenarios is critical to understand how hormone signaling networks shape gene expression in health and disease.

Fig3 (right). Models for hormone receptor interactions and sharing of co-factors

A) Model where PR and ER interchangeably interact with GRHL2 and other cofactors depending on the hormones present. Chapter 5 lends support for the existence of a GRHL2/PR/co-factor complex in both the absence and presence of progesterone. B) Model for PR antagonization of ER on the DNA level. In the presence of estrogen (E2), ER and GRHL2 are bound to estrogen responsive elements (ERE) activating ER target genes. Upon progesterone (P4) addition, activated PR can sequester ER and GRHL2, redirecting it to progesterone responsive elements (PRE) and activating PR regulated genes. C) Generalized model where hormone receptors interact with pioneering factors and co-factors. The remaining question is if whether the presence of multiple hormones leads to exchange of the hormone receptors binding to the pioneering factor (left) or whether pioneering and co-factors are bound simultaneously (right).



Linking enhancers and transcription factors to their target genes: challenges and future directions

In **chapters 3, 4, and 5**, we aimed to identify enhancers and link them to their target genes. When focusing on a single gene, such as *Wnt4* in **chapters 3 and 4**, we confined our search to the topologically associating domain (TAD) in which the gene is located. Within this region, candidate enhancers were selected based on chromatin accessibility (ATAC-seq/DNase-seq) and enrichment for the active enhancer mark H3K27ac, a commonly used approach[19], also in genome-wide studies. However, once we identified candidate enhancers, functionally linking them to their target gene remained a major challenge. Many studies still assign enhancers to the nearest gene or define arbitrary windows (e.g., 5–100 kb) around the transcription start site[42]. This approach is far from accurate, as it is now well-established that enhancers can act over large genomic distances, where 33–73% of enhancers are estimated not to regulate their nearest gene43–45. In this context, accurate prediction of enhancer–gene interactions is inherently difficult.

6

High-resolution chromatin conformation capture techniques, such as Hi-C, 5C, ChIA-PET, and HiChIP, have emerged as some of the most effective tools for identifying enhancer–promoter interactions. Yet still, these methods come with limitations. High resolution is required to identify DNA-DNA interactions beyond general chromatin architecture and pinpoint specific interactions. Techniques like ChIA-PET and HiChIP provide valuable insights into transcription factor involvement at enhancer–promoter loops, but they are limited to a single factor per experiment. This restricts our ability to assess co-involvement of DNA-DNA loops, which is critical since transcription factors rarely act alone. Furthermore, physical DNA interactions do not necessarily imply transcriptional activity. For example, stable CTCF-mediated enhancer–promoter loops are often present even in cell types where the associated gene is inactive (as shown in **chapter 3**). Therefore, to be able to determine if an enhancer promotor loop is functional, chromatin interaction data should always be interpreted in conjunction with gene expression data.

In **chapter 3**, we used a multi-faceted approach, combining reporter assays, dCas9-based perturbations, 4C, bioinformatic loop predictions, and enhancer knockouts, to functionally link candidate enhancers to *Wnt4*. While this strategy yielded valuable insights, (combinations of) enhancer knockouts remain the most conclusive to confirm enhancer–gene relationships. Yet, generating these knockouts is far from straightforward. It is time-consuming and labor-intensive and as discussed above while effective in cell lines, can prove more difficult and less informative *in vivo* (**Chapters 3 and 4**). Redundancy and developmental robustness complicate the interpretation of single enhancer knockouts and necessitate multiple mouse models with combinatorial enhancer deletions—an approach that is not only costly and time-consuming but also will continue to require animal experiments. At the same time, the limited conservation of individual enhancer sequences between mice and humans underscores the need for additional, physiologically relevant human model systems as well.

Emerging tools such as CRISPR-based dCas9 systems designed to block transcription factor–DNA interactions on a specific enhancer thereby preventing transcription factor action on genes, have potential to help studying local regulatory effects[46] but these methods have still not been widely adopted. In addition, epigenetic editing, an area of growing interest[47], may offer new ways to perturb enhancer function by modifying histone marks, potentially offering one way to tackle current challenges. In general, the development of novel techniques to more efficiently and precisely disrupt enhancer–

promoter interactions would significantly advance our ability to identify new layers of gene regulation. To understand the mechanisms of enhancer regulation, ChIP-seq remains an invaluable tool for mapping transcription factor–DNA interactions. However, it has important caveats: it requires crosslinking and reflects an average binding signal across many cells. Moreover, if we are interested in co-regulation or collaborative binding between multiple factors, two separate ChIP-seq datasets do not tell us whether the factors co-bind the same site, bind simultaneously, or even bind in the same cell. Addressing these questions would require single-cell ChIP-seq technologies that can measure multiple factors concurrently, something that does not yet exist but may become feasible given the rapid advances in single-cell methods.

A further challenge arises when we slightly shift the question: instead of asking whether a transcription factor can bind a specific enhancer, what if we want to identify which transcription factors are jointly bound at a specific enhancer or locus at any one moment? In **chapter 3**, we used a bioinformatic pipeline to predict potential regulators of *Wnt4*, resulting in a list of candidate factors. From this list, we focused on GRHL factors, inspecting ChIP-seq data and performing overexpression and knockdown experiments to confirm their involvement. While this approach was successful and informative, ideally, we would have a direct, unbiased method to identify the full complement of proteins, at a specific genomic locus[48.] One existing technique for this is genomic locus proteomics using a dCas9-APEX system, which allows proximity labeling and identification of proteins associated with a specific DNA region[49,50]. In theory, this method is very promising and would allow the study of locus-specific protein complexes. However, many of our colleagues have experienced that this technique lacks robustness and reproducibility, limiting its current utility. A more reliable, and ideally higher throughput version of this approach could open a new frontier in studying enhancer regulation. Finally, live-cell microscopy-based techniques capable of visualizing specific genomic loci in real time[51] are likely to become invaluable tools. Such approaches could help capture the dynamic nature of transcription factor binding and enhancer–promoter interactions—phenomena that are far more dynamic than static, population-averaged methods like ChIP-seq and chromatin capture tend to suggest. As these technologies continue to evolve, they will allow us to better understand the temporal and spatial dynamics of enhancer function and gene regulation.

Concluding remarks

In this thesis, we shed light on the chromatin-level molecular mechanisms of progesterone-driven gene regulation. We developed tools and optimized culture conditions to accurately measure and visualize progesterone receptor signaling activity. These tools serve as an important first step in enabling new experimental approaches in our own work and will provide a valuable resource for others investigating outstanding questions in PR signaling. We furthermore provide the first in-depth functional dissection of the *Wnt4* enhancer landscape across the mammary gland, kidney, and lung. This has not only given us new insights into how tissue specific expression patterns of *Wnt4* are established, but we also had the unique opportunity to compare enhancer sets of a single gene in different tissues at multiple developmental stages. We found that while *Wnt4* enhancer sets are highly tissue-specific, chromatin conformation remains largely conserved. Moreover, we demonstrated that *Wnt4* is co-regulated by both PR and GRHL factors and further explored the genome wide functional interaction of PR and GRHL2. We provided the first evidence of functional interaction between PR and GRHL2 and showed that these transcription factors physically interact and co-regulate gene expression through binding to distal regulatory elements that loop to their target gene promoters.

With PR gaining more attention, and in light of the rapid advancements in chromatin conformation capture, single-cell, and locus-specific techniques, we are now entering an exciting era in enhancer biology. These technological innovations, combined with renewed interest in PR signaling, offer new exiting opportunities to uncover the chromatin-level mechanisms that drive progesterone-responsive gene regulatory networks. In the years ahead, we can expect to gain critical insights into the complexity, specificity, and dynamic regulation of gene expression by progesterone and, for those interested in tissue-specific gene expression, in the context-dependent regulatory mechanisms of individual genes.

Statements and declarations

Declaration of the use of generative AI

In this scientific work, no generative AI tools were used to generate content. During the writing process, ChatGPT was used to improve writing style of the first draft. The output was critically reviewed and partly re-written by the author. The author takes full responsibility for the final text, content, claims and references in this work.

References

1. van Amerongen, R., Bentires-Alj, M., van Boxtel, A.L., Clarke, R.B., Fre, S., Suarez, E.G., Iggo, R., Jechlinger, M., Jonkers, J., Mikkola, M.L., *et al.* (2023). Imagine beyond: recent breakthroughs and next challenges in mammary gland biology and breast cancer research. J. Mammary Gland Biol. Neoplasia *28*, 17. https://doi.org/10.1007/s10911-023-09544-y.
2. Jardé, T., Lloyd-Lewis, B., Thomas, M., Kendrick, H., Melchor, L., Bougaret, L., Watson, P.D., Ewan, K., Smalley, M.J., and Dale, T.C. (2016). Wnt and Neuregulin1/ErbB signalling extends 3D culture of hormone responsive mammary organoids. Nat. Commun. *7*, 13207. https://doi.org/10.1038/ncomms13207.
3. Jamieson, P.R., Dekkers, J.F., Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2017). Derivation of a robust mouse mammary organoid system for studying tissue dynamics. Dev. Camb. Engl. *144*, 1065–1071. https://doi.org/10.1242/dev.145045.
4. Caruso, M., Huang, S., Mourao, L., and Scheele, C.L.G.J. (2022). A Mammary Organoid Model to Study Branching Morphogenesis. Front. Physiol. *13*, 826107. https://doi.org/10.3389/fphys.2022.826107.
5. Sahu, S., Albaugh, M.E., Martin, B.K., Patel, N.L., Riffle, L., Mackem, S., Kalen, J.D., and Sharan, S.K. (2022). Growth factor dependency in mammary organoids regulates ductal morphogenesis during organ regeneration. Sci. Rep. *12*, 7200. https://doi.org/10.1038/s41598-022-11224-6.
6. Sachs, N., de Ligt, J., Kopper, O., Gogola, E., Bounova, G., Weeber, F., Balgobind, A.V., Wind, K., Gracanin, A., Begthel, H., *et al.* (2018). A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. Cell *172*, 373-386.e10. https://doi.org/10.1016/j.cell.2017.11.010.
7. Dekkers, J.F., van Vliet, E.J., Sachs, N., Rosenbluth, J.M., Kopper, O., Rebel, H.G., Wehrens, E.J., Piani, C., Visvader, J.E., Verissimo, C.S., *et al.* (2021). Long-term culture, genetic manipulation and xenotransplantation of human normal and breast cancer organoids. Nat. Protoc. *16*, 1936–1965. https://doi.org/10.1038/s41596-020-00474-1.
8. Rosenbluth, J.M., Schackmann, R.C.J., Gray, G.K., Selfors, L.M., Li, C.M.-C., Boedicker, M., Kuiken, H.J., Richardson, A., Brock, J., Garber, J., *et al.* (2020). Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. Nat. Commun. *11*, 1711. https://doi.org/10.1038/s41467-020-15548-7.
9. Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D.R., and Zhao, K. (2017). CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. Mol. Cell *67*, 1049-1058.e6. https://doi.org/10.1016/j.molcel.2017.08.026.
10. Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., *et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111*, 996–1001. https://doi.org/10.1073/pnas.1317788111.
11. Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22. https://doi.org/10.1016/j.cell.2017.05.004.
12. Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., *et al.* (2017). Cohesin

Loss Eliminates All Loop Domains. Cell *171*, 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026.

13. Jusuf, J.M., Grosse-Holz, S., Gabriele, M., Mach, P., Flyamer, I.M., Zechner, C., Giorgetti, L., Mirny, L.A., and Hansen, A.S. (2025). Genome-wide absolute quantification of chromatin looping. BioRxiv Prepr. Serv. Biol., 2025.01.13.632736. https://doi.org/10.1101/2025.01.13.632736.
14. Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., *et al.* (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243. https://doi.org/10.1038/nature25461.
15. Bradbury, J.M., Edwards, P.A.W., Niemeyer, C.C., and Dale, T.C. (1995). Wnt-4 Expression Induces a Pregnancy-like Growth Pattern in Reconstituted Mammary Glands in Virgin Mice. Dev. Biol. *170*, 553–563. https://doi.org/10.1006/dbio.1995.1236.
16. Kim, Y.C., Clark, R.J., Pelegri, F., and Alexander, C.M. (2009). Wnt4 is not sufficient to induce lobuloalveolar mammary development. BMC Dev. Biol. *9*, 55. https://doi.org/10.1186/1471-213X-9-55.
17. Tanos, T., Sflomos, G., Echeverria, P.C., Ayyanan, A., Gutierrez, M., Delaloye, J.-F., Raffoul, W., Fiche, M., Dougall, W., Schneider, P., *et al.* (2013). Progesterone/RANKL is a major regulatory axis in the human breast. Sci. Transl. Med. *5*, 182ra55. https://doi.org/10.1126/scitranslmed.3005654.
18. Xiao, L., Fan, D., Qi, H., Cong, Y., and Du, Z. (2022). Defect-buffering cellular plasticity increases robustness of metazoan embryogenesis. Cell Syst. *13*, 615-630.e9. https://doi.org/10.1016/j.cels.2022.07.001.
19. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.
20. Zou, Z., Ohta, T., and Oki, S. (2024). ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. Nucleic Acids Res. *52*, W45–W53. https://doi.org/10.1093/nar/gkae358.
21. van de Grift, Y.B.C., Heijmans, N., and van Amerongen, R. (2020). How to Use Online Tools to Generate New Hypotheses for Mammary Gland Biology Research: A Case Study for Wnt7b. J. Mammary Gland Biol. Neoplasia *25*, 319–335. https://doi.org/10.1007/s10911-020-09474-z.
22. Misra, B.B., Langefeld, C., Olivier, M., and Cox, L.A. (2019). Integrated omics: tools, advances and future approaches. J. Mol. Endocrinol. *62*, R21–R45. https://doi.org/10.1530/JME-18-0055.
23. Nordin, A., Pagella, P., Zambanini, G., and Cantù, C. (2024). Exhaustive identification of genome-wide binding events of transcriptional regulators. Nucleic Acids Res. *52*, e40. https://doi.org/10.1093/nar/gkae180.
24. Reese, R.M., Helzer, K.T., Allen, K.O., Zheng, C., Solodin, N., and Alarid, E.T. (2022). GRHL2 Enhances Phosphorylated Estrogen Receptor (ER) Chromatin Binding and Regulates ER-Mediated Transcriptional Activation and Repression. Mol. Cell. Biol. *42*, e00191-22. https://doi.org/10.1128/mcb.00191-22.
25. Jozwik, K.M., Chernukhin, I., Serandour, A.A., Nagarajan, S., and Carroll, J.S. (2016). FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. Cell Rep. *17*, 2715–2723. https://doi.org/10.1016/j.celrep.2016.11.028.

26. Holding, A.N., Giorgi, F.M., Donnelly, A., Cullen, A.E., Nagarajan, S., Selth, L.A., and Markowetz, F. (2019). VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERa at enhancers in breast cancer. Genome Biol. *20*, 91. https://doi.org/10.1186/s13059-019-1698-z.
27. Ballaré, C., Uhrig, M., Bechtold, T., Sancho, E., Di Domenico, M., Migliaccio, A., Auricchio, F., and Beato, M. (2003). Two domains of the progesterone receptor interact with the estrogen receptor and are required for progesterone activation of the c-Src/Erk pathway in mammalian cells. Mol. Cell. Biol. *23*, 1994–2008. https://doi.org/10.1128/MCB.23.6.1994-2008.2003.
28. Mohammed, H., Russell, I.A., Stark, R., Rueda, O.M., Hickey, T.E., Tarulli, G.A., Serandour, A.A., Birrell, S.N., Bruna, A., Saadi, A., *et al.* (2015). Progesterone receptor modulates ERα action in breast cancer. Nature *523*, 313–317. https://doi.org/10.1038/nature14583.
29. Singhal, H., Greene, M.E., Tarulli, G., Zarnke, A.L., Bourgo, R.J., Laine, M., Chang, Y.-F., Ma, S., Dembo, A.G., Raj, G.V., *et al.* (2016). Genomic agonism and phenotypic antagonism between estrogen and progesterone receptors in breast cancer. Sci. Adv. *2*, e1501924. https://doi.org/10.1126/sciadv.1501924.
30. Vignon, F., Bardon, S., Chalbos, D., and Rochefort, H. (1983). Antiestrogenic effect of R5020, a synthetic progestin in human breast cancer cells in culture. J. Clin. Endocrinol. Metab. *56*, 1124–1130. https://doi.org/10.1210/jcem-56-6-1124.
31. Gill, P.G., Vignon, F., Bardon, S., Derocq, D., and Rochefort, H. (1987). Difference between R5020 and the antiprogestin RU486 in antiproliferative effects on human breast cancer cells. Breast Cancer Res. Treat. *10*, 37–45. https://doi.org/10.1007/BF01806133.
32. Daniel, A.R., Gaviglio, A.L., Knutson, T.P., Ostrander, J.H., D'Assoro, A.B., Ravindranathan, P., Peng, Y., Raj, G.V., Yee, D., and Lange, C.A. (2015). Progesterone receptor-B enhances estrogen responsiveness of breast cancer cells via scaffolding PELP1- and estrogen receptor-containing transcription complexes. Oncogene *34*, 506–515. https://doi.org/10.1038/onc.2013.579.
33. Brisken, C., and Scabia, V. (2020). 90 YEARS OF PROGESTERONE: Progesterone receptor signaling in the normal breast and its implications for cancer. J. Mol. Endocrinol. *65*, T81–T94. https://doi.org/10.1530/JME-20-0091.
34. Agnoletto, A., and Brisken, C. (2025). Hormone Signaling in Breast Development and Cancer. In A Guide to Breast Cancer Research: From Cellular Heterogeneity and Molecular Mechanisms to Therapy, T. Sørlie and R. B. Clarke, eds. (Springer Nature Switzerland), pp. 279–307. https://doi.org/10.1007/978-3-031-70875-6_15.
35. Hickey, T.E., Selth, L.A., Chia, K.M., Laven-Law, G., Milioli, H.H., Roden, D., Jindal, S., Hui, M., Finlay-Schultz, J., Ebrahimie, E., *et al.* (2021). The androgen receptor is a tumor suppressor in estrogen receptor–positive breast cancer. Nat. Med. *27*, 310–320. https://doi.org/10.1038/s41591-020-01168-7.
36. Karmakar, S., Jin, Y., and Nagaich, A.K. (2013). Interaction of glucocorticoid receptor (GR) with estrogen receptor (ER) α and activator protein 1 (AP1) in dexamethasone-mediated interference of ERα activity. J. Biol. Chem. *288*, 24020–24034. https://doi.org/10.1074/jbc.M113.473819.
37. Sun, G., Zhao, C., Han, J., Wu, S., Chen, Y., Yao, J., and Li, L. (2024). Regulatory mechanisms of steroid hormone receptors on gene transcription through chromatin interaction and enhancer reprogramming. Cell. Oncol. Dordr. Neth. *47*, 2073–2090. https://doi.org/10.1007/s13402-024-01011-y.

38. Ogara, M.F., Rodríguez-Seguí, S.A., Marini, M., Nacht, A.S., Stortz, M., Levi, V., Presman, D.M., Vicent, G.P., and Pecci, A. (2019). The glucocorticoid receptor interferes with progesterone receptor-dependent genomic regulation in breast cancer cells. Nucleic Acids Res. *47*, 10645–10661. https://doi.org/10.1093/nar/gkz857.
39. Rangel, N., Fortunati, N., Osella-Abate, S., Annaratone, L., Isella, C., Catalano, M.G., Rinella, L., Metovic, J., Boldorini, R., Balmativola, D., *et al.* (2018). FOXA1 and AR in invasive breast cancer: new findings on their co-expression and impact on prognosis in ER-positive patients. BMC Cancer *18*, 703. https://doi.org/10.1186/s12885-018-4624-y.
40. Metovic, J., Borella, F., D'Alonzo, M., Biglia, N., Mangherini, L., Tampieri, C., Bertero, L., Cassoni, P., and Castellano, I. (2022). FOXA1 in Breast Cancer: A Luminal Marker with Promising Prognostic and Predictive Impact. Cancers *14*, 4699. https://doi.org/10.3390/cancers14194699.
41. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241. https://doi.org/10.1101/gad.176826.111.
42. Clarke, C.L., and Graham, J.D. (2012). Non-overlapping progesterone receptor cistromes contribute to cell-specific transcriptional outcomes. PloS One *7*, e35859. https://doi.org/10.1371/journal.pone.0035859.
43. Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113. https://doi.org/10.1038/nature11279.
44. Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., *et al.* (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98. https://doi.org/10.1016/j.cell.2011.12.014.
45. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., *et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606. https://doi.org/10.1038/ng.3286.
46. Shariati, S.A., Dominguez, A., Xie, S., Wernig, M., Qi, L.S., and Skotheim, J.M. (2019). Reversible Disruption of Specific Transcription Factor-DNA Interactions Using CRISPR/Cas9. Mol. Cell *74*, 622-633.e4. https://doi.org/10.1016/j.molcel.2019.04.011.
47. Jacob, J., van Loosen, Q.C., van den Berg van Saparoea, A.C.H., Sarno, F., and Verschure, P.J. (2024). Bridging bioengineering and epigenetics: from technical innovations to clinical applications. Epigenetics Commun. *4*, 8. https://doi.org/10.1186/s43682-024-00031-x.
48. Merika, M., and Thanos, D. (2001). Enhanceosomes. Curr. Opin. Genet. Dev. *11*, 205–208. https://doi.org/10.1016/s0959-437x(00)00180-5.
49. Myers, S.A., Wright, J., Peckner, R., Kalish, B.T., Zhang, F., and Carr, S.A. (2018). Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nat. Methods *15*, 437–439. https://doi.org/10.1038/s41592-018-0007-1.
50. Gao, X.D., Tu, L.-C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S.A., Zhu, L.J., Wolfe, S.A., *et al.* (2018). C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat. Methods *15*, 433–436. https://doi.org/10.1038/s41592-018-0006-2.
51. Platania, A., Erb, C., Barbieri, M., Molcrette, B., Grandgirard, E., de Kort, M.A.C., Pomp, W., Meaburn, K., Taylor, T., Shchuka, V.M., *et al.* (2024). Transcription processes compete with loop extrusion to homogenize promoter and enhancer dynamics. Sci. Adv. *10*, eadq0987. https://doi.org/10.1126/sciadv.adq0987.