Exploring the Wnt enhancer landscape in the mammary gland

Heijmans, N.

Publication date
2021

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

General Discussion
As introduced in chapter 1, the mouse mammary gland is a complex and dynamic tissue. Its development and homeostasis require extensive and controlled communication between different cells and cell types. To regulate cell behavior, signaling molecules have to facilitate communication between cells at exactly the right time and place. WNT ligands are an example of such signaling molecules, and part of a conserved intercellular signaling pathway. Despite our knowledge of the WNT/CTNNB1 pathway downstream of the translated WNT protein, surprisingly little is known about how spatiotemporal expression of Wnt genes themselves is regulated. To understand the basis of spatiotemporal specific cell-cell communication, we need to investigate how cell type specific Wnt expression is regulated.

In this thesis we have focused on this knowledge gap by exploring the enhancer landscape of different Wnt genes in the mammary gland. This has resulted in the generation of an extensive transcriptomic analysis of the postnatal mammary gland and the setup of a workflow for enhancer identification. Subsequently, we used this workflow to get new insights into the regulation of Wnt4 and Wnt2 in the mammary gland. This is the first comprehensive study that explores Wnt gene regulation in detail.

**Transcriptomic analysis of the postnatal mammary gland**

The goal of chapter 2 was to obtain a comprehensive overview of gene expression in the mouse mammary gland at different time points and in different cell types. The selected time points were puberty and different stages of the estrous cycle of adult mice: proestrus, estrus, metestrus, and diestrus. During the estrous cycle, estrogen and progesterone levels fluctuate. Previous studies have shown that this can result in morphological changes during the estrous cycle, although there is no clear consensus in literature about this event. Outgrowth of mammary epithelium has been reported to occur in late proestrus/estrus, but also in diestrus. However, these studies do agree on the timing of regression of the epithelium, which occurs during diestrus. More recently, another paper highlighted the proliferative heterogeneity in the mammary gland during the estrous cycle, and showed that the morphological changes do not have to occur during every cycle. Our results also indicate that the heightened proliferative state that was previously linked to the estrous cycle, does not occur during every cycle, as we did not observe epithelial outgrowth in carmine stainings. Furthermore, our FACS analysis did not show clear differences in cell numbers between the different stages.

The proliferative changes in the mammary gland during the estrous cycle are complex and not well understood. The estrous cycle repeats itself every 4 to 5 days and the question arises why the morphology would change so frequently. This seems like a waste of energy, since the mammary gland has time to change its morphology and prepare for lactation during pregnancy. Furthermore, proliferative estrous cycles can result in more replication-related mutations in the genome than non-proliferative cycles. As the number of menstrual cycles in a woman's lifetime has been associated with breast cancer, it would be interesting to investigate whether proliferative estrous cycles might pose a higher breast cancer risk. As this study was performed...
on mouse mammary glands, it remains to be seen whether this phenomenon of proliferative and non-proliferative cycles is conserved in humans.

In our genome-wide expression analysis, we specifically focused on the expression of \textit{Wnt} genes. These genes encode for WNT ligands, which are important for the development and homeostasis of various tissues, including the mammary gland\cite{13-16}. In previous studies, expression of \textit{Wnt4}, \textit{Wnt5a}, and \textit{Wnt7a} has been linked to steroid hormone activity in different tissues\cite{2,13,14,17-20}. As the luminal cell layer contains estrogen receptor (ER) and progesterone receptor (PR) positive cells\cite{21}, and as \textit{Wnt4} has been explicitly been implicated as a PR-responsive gene\cite{2,13,14}, we expected these \textit{Wnt} genes to be differentially expressed in luminal cells throughout the estrous cycle. Our RNA-seq analysis however, showed that no \textit{Wnt} genes are differentially expressed in luminal or basal cells when comparing different stages of the estrous cycle. Even \textit{Wnt4} was not differentially expressed in luminal cells. As mentioned in \textbf{chapter 2}, one explanation for this discrepancy is that the changes under physiological conditions could be too subtle and might not have been robust enough to pick up in our bulk RNA-seq. It is also possible that \textit{Wnt4} expression in the mammary gland is more heterogenous than previously thought, similar to the heightened proliferative state that has previously been reported\cite{9}. Our results indicate however, that \textit{Wnt4} expression is not only regulated by steroid hormones. The question remains: If not solely regulated by hormones, how is \textit{Wnt4} expression regulated in the mammary gland?

The overview of \textit{Wnt} expression in different mammary gland cell types revealed a complex expression pattern of individual \textit{Wnt} genes across different cell types, with far less variation in a temporal sense (comparing puberty to adulthood). At the moment, for none of these \textit{Wnt} genes it is known how these defined spatiotemporal expression patterns are regulated in the mammary gland. In \textbf{chapter 3, 4, and 5} we focused on this topic, and aimed to unravel mechanisms of \textit{Wnt} gene regulation at the chromatin level.

\textbf{Strategies for enhancer identification}

Specific spatiotemporal expression patterns can be established by genomic regulatory elements, such as enhancers\cite{22-24}. In \textbf{chapter 3}, we set up a workflow to identify enhancers that can induce \textit{Wnt} expression. In recent years, new technologies have emerged that can be used for discovering and validating enhancers (Table 1). As these methods are clearly useful for investigating these regulatory sequences, all of them have distinct advantages and disadvantages.
Currently, there is no single method that can unequivocally identify enhancers on its own. Therefore, it is important to combine complimentary techniques for the discovery, characterization and validation of enhancers. This ensures that enhancer identification is supported at multiple levels. Strong support of enhancer identification should include the following criteria:

1. The presence of enhancer-associated chromatin marks at the endogenous genomic location of the candidate enhancer, such as histone modifications or chromatin accessibility.
2. Evidence of functional enhancer activity in plasmid-based reporter assays, such as the dual luciferase assay.
3. Demonstration of a link between the candidate enhancer and target gene, which can be achieved by CRISPRinterference (CRISPRi) or CRISPRactivation (CRISPRa).

We started setting up our workflow for enhancer identification in chapter 3 by selecting candidate enhancers based on sequence conservation or specific chromatin signatures. Luciferase reporter assays in which we tested the functional enhancer activity of these candidate enhancers, revealed that it is more effective to select candidates based on enhancer-associated chromatin marks than based on sequence conservation. We think, however, that analyzing sequence conservation of specific candidate enhancers (e.g. using BLAST) can be useful to further dissect the molecular sequence architecture. So rather than using conservation at the start of a project to select candidate enhancers, this tool might be more useful after the validation of enhancers (as done in chapter 4).

The datasets for chromatin accessibility or histone modifications that were publicly available when starting the project of chapter 3, were not optimal. First, the epithelial cell types were usually not separated, which is inconvenient when searching for enhancers that regulate cell type specific expression. Second, not all datasets were from virgin mice; in some cases pregnant or lactating mice were used. None of them were derived from pubertal animals. We still decided to use and visualize these datasets, because of the lack of cell-type specific mammary gland datasets at the time.

Functional enhancer activity is another important characteristic that should be tested when identifying new enhancers. We chose to clone the candidate enhancers individually in the pGL4-minP plasmid, rather than performing massively parallel reporter assays (MPRAs). The main reason is that there currently is a limit on the length of the oligos that are synthesized for MPRAs. These experiments allow for checking enhancer activity of <200 bp DNA sequences. Since our candidate enhancers were >200 bp (and enhancers are described to be 50-1500 bp), we decided to clone these sequences individually, although this method is low-throughput. We must be aware of the main disadvantage of testing enhancer activity using reporter constructs, namely that enhancers are tested outside of their native genomic context. Epigenetic modifications or specific 3D structures of the chromatin at the endogenous genomic location of the enhancer will have no effect on the enhancer activity of the enhancer when it is tested in a plasmid based reporter assay. On the other hand, it allows for testing sequences that would otherwise reside in closed (i.e. non-accessible) chromatin regions. Taken together, reporter assays remain an essential tool for enhancer identification, as the enhancer activity of the candidate enhancers can be tested directly.

To link the candidate enhancers to the target genes, several methods are available from the CRISPR toolbox. CRISPR-interference (CRISPRi) and CRISPR-activation (CRISPRa) are used to study promoters and enhancers in their native genomic context by enforcing enhancer repression or activation, respectively. The first CRISPRi experiments showed that a nuclease dead Cas9
(dCas9) can be used for repressing transcription initiation or elongation, as it can block binding of transcription factors (TFs) or form an obstruction for RNA Pol II\textsuperscript{53–56}. The first time CRISPRi was used to target enhancers, dCas9 was fused to the Krüppel-associated box (KRAB) repressor. The dCas9-KRAB fusion was guided to the HS2 enhancer in the globin locus control region, which resulted in the silencing of multiple globin genes\textsuperscript{57}. In contrast, CRISPRa uses fusions of dCas9 with transcriptional activators, such as Vp64, p65AD, or a combination of multiple activators (Vp64, p65, Rta (VPR))\textsuperscript{39,40,58,59}. Another way of achieving transcriptional activation using CRISPRa, is by targeting the epigenome. The fusion of dCas9 to the histone acetyltransferase p300 can acetylate H3K27 at specific genomic locations, which can result in gene activation\textsuperscript{38}. Both dCas9-VPR and dCas9-p300 have been used to link genomic regulatory elements to target genes\textsuperscript{38,60,61}. In our workflow for enhancer identification, we used dCas9-VPR to link our candidate enhancer to the target gene. We selected this approach after testing multiple dCas9 fusions (dCas9-KRAB, dCas9-p300, dCas9-VPR) (data not shown), which showed the most robust results for dCas9-VPR. We hypothesize that the acetylation of our candidate enhancers by dCas9-p300 was not sufficient to increase gene transcription. This can be linked to the fact that some of our candidate enhancers were selected based on histone marks such as H3K27ac, which would imply that they were already open and activated and cannot be further activated by dCas9-p300. In fact, guiding dCas9-p300 to an already acetylated region might actually form an obstruction for the TFs that would bind to the DNA. dCas9-VPR on the other hand, includes a transcriptional activator (Vp64) and two TFs (p65 and Rta), which artificially “switch on” the enhancer. Most likely, different dCas9 modifiers are suitable for different genomic loci, depending on their epigenetic state. We have to keep in mind that the exact mechanism of CRISPRi and CRISPRa are not known. These dCas9 tools however, are still useful tools to investigate possible links between regulatory sequences and target genes.

**Regulation of Wnt2 and Wnt4 expression in the mammary gland**

Although the workflow presented in chapter 3 can provide a framework for enhancer identification, there are additional methods that can further help understand how expression of specific genes is regulated. We implemented some of these methods in chapter 4 and 5, in which we focused on the expression of \textit{Wnt4} and \textit{Wnt2}, respectively. In these chapters, we refer to our candidate enhancer sequences as CRSs (candidate regulatory sequences). The reasoning behind this, is that the only method in our workflow that provides information about functional enhancer activity, is the luciferase reporter assay. The other methods that we used for enhancer identification, can also be used to identify regulatory sequences that have other functions, such as repressors. In chapter 4, CRS-9 and CRS-16 are examples of such candidate repressors. Both CRS-9 and CRS-16 were linked to the \textit{Wnt4} gene in dCas9-VPR assays, but did not show any enhancer activity in the luciferase reporter assays. CRS-9 showed clear peaks in multiple ATAC-seq datasets, especially in luminal cells. Interestingly, CRS-16 had a specific peak in the basal ATAC-seq dataset. Since \textit{Wnt4} is not expressed at basal cells, CRS-16 might play a role in repressing \textit{Wnt4} expression in this cell type, although additional experiments are needed to prove this.
Another development in chapter 4 and 5 compared to chapter 3, was the use of cell type specific datasets for enhancer associated marks. In 2018, ATAC-seq and H3K27ac datasets were published from mature luminal, luminal progenitor, and basal mammary cells from WT mice. In addition, we generated our own ATAC-seq datasets of luminal, basal, and stromal cell types from pubertal animals. This was especially valuable for selecting Wnt2 CRSs, since Wnt2 is not expressed in epithelial cells (chapter 2) and to our knowledge, no stromal specific ATAC-seq (or H3K27ac ChIP-seq) dataset has been published before.

Another addition in chapter 4 and 5, was the topologically associating domain (TAD) analysis at the Wnt4 and Wnt2 loci. As enhancers are more likely to communicate with promoters in the same TAD, we decided to focus specifically on the TAD of our genes of interest when selecting CRSs. We analyzed the TADs of Wnt4 and Wnt2 by downloading the TAD coordinates of various datasets from the 3D Genome Browser. At the Wnt4 locus, there was an overlapping region in all datasets that coincided with the CTCF binding motifs and their orientation. We selected this as our region of interest to select Wnt4 CRSs. For the Wnt2 locus, we performed a similar TAD analysis, but those results were less conclusive. Interestingly, our TAD analyses of the Wnt4 and Wnt2 loci were supported by 4C experiments, which showed a higher interaction frequency of the Wnt4 or Wnt2 promoter with regions within the selected TAD than with regions outside the selected TAD. 4C is another experiment that we added to our protocol in chapter 4 and 5 compared to chapter 3, mainly to identify regions that showed a physical interaction with the Wnt4 or Wnt2 promoter in the native chromatin context. Indeed, by integrating our findings from the luciferase reporter assays, dCas9-VPR experiments and 4C analysis, we were able to identify elements that were recognized in each of these assays as being of interest, resulting in the identification of a regulatory hub between Gm13003 and Wnt4 in the mammary epithelium (chapter 4) and CRS-12 as a stromal specific Wnt2 mammary gland enhancer (chapter 5).

Currently, there is a debate in the field of gene regulation regarding the role of TADs. Although multiple studies have indicated that TAD formation is important for gene regulation by bringing enhancers and promoters in closer proximity, other reports suggest that this role is more nuanced. Loss of TAD boundaries caused by depletion of critical factors such as CTCF or subunits of the cohesin complex, quite surprisingly resulted only in modest gene expression changes in some cases. Also more specific experimental rearrangements of TADs at the Sox9-Kcnj2 locus did not result in major gene expression changes. On the other hand, disruption of TADs in the WNT6/IHH/EPHA4/PAX3 locus leads to de novo enhancer-promoter interactions, which causes misexpression of developmental genes. The basis of this inconsistency has been unclear. Recently, a study that focused directly on the impact of chromosome topology on enhancer-promoter interactions in Drosophila embryos, showed that interactions of elements that are in the same TAD, occur independently of TAD formation (Figure 1). Inter-TAD interactions however, were lost after disruption of TADs. This suggests that the role of TADs is different for intra- and inter-TAD interactions. Future research is needed to unravel the mechanisms and specificity of gene regulation by TAD formation.
To summarize, we have supported our findings about regulation of Wnt4 and Wnt2 expression in the mammary gland on multiple levels:

1. TAD analysis of the locus of interest, by downloading publicly available HiC datasets from the 3D Genome Browser.
2. Analysis of different enhancer-associated chromatin features, such as chromatin accessibility or histone modifications, using generated or published ATAC-seq or H3K27ac ChIP-seq datasets, respectively.
4. Showing evidence of a link between candidate sequences and Wnt4 or Wnt2 by CRISPRa.
5. Investigating physical interactions between the Wnt4 or Wnt2 promoter and candidate sequences by performing 4C experiments.

The next step in investigating the regulation of Wnt4 and Wnt2 expression in the mammary gland, is to validate the functional relevance of the identified enhancers by mutating or deleting their genomic sequences using CRISPR-Cas9. Also candidate repressors (e.g. CRS-9 and -16 for Wnt4) can be deleted to investigate their possible repressive role in Wnt4 expression. When measuring Wnt4 or Wnt2 expression after mutating identified enhancers, it has to be taken into account that the transcriptional effect of these mutations can be buffered by so-called shadow enhancers. Some regulatory networks are known to contain enhancers that have similar tasks, to guarantee robustness of gene expression. These shadow or redundant enhancers ensure...
that gene expression is maintained in case of mutations, which can lead to false negatives when deleting functional enhancers using CRISPR-Cas9. In addition to mutating genomic enhancer sequences, it would be interesting to investigate the effect of mutating the putative CTCF binding sites at the TAD boundaries, since the importance of these boundaries is highly locus specific (as mentioned above). The identified putative CTCF binding sites can be removed or their orientation can be reversed. Subsequent 4C or Hi-C experiments can show the effect of TAD disruption on the physical interactions of the Wnt promoters.

Although Wnt4 expression in the mammary gland has been linked to steroid hormones in previous studies, our research has so far focused on hormone-independent regulation of this gene. There are multiple options to expand our workflow to also investigate genomic elements that regulate Wnt4 via binding of nuclear receptors, such as PR or ER. First, PR/ER ChIP-seq datasets can be included in the phase of CRS selection. Regions that have a peak in PR/ER binding and are located in the region of interest (e.g. the TAD), can be added to the list of CRSs. At this stage it would be useful to search for putative PR/ER binding sites in these sequences using TRANSFAC. Second, plasmid based reporter assays that are used to check the enhancer activity of the candidate enhancers, can be performed with and without progesterone or estrogen treatment. It is important that in this case cell lines are used that stably express PR or ER, which is not the case for BC44 or HC11. Reporter assays can also be performed using mutated versions of the candidate enhancer, in which for example one or multiple PR/ER binding sites have been removed. Third, the genomic sequences of these putative binding sites in the candidate enhancer can be mutated using CRISPR-Cas9, to check their biological relevance in Wnt4 expression. Adding these methods to our workflow will allow us to definitively discriminate between enhancers that act downstream of steroid hormones and those that do not.

**Future perspectives in enhancer biology**

To further unravel the mechanisms of enhancer function, identified and validated enhancers can be dissected at the nucleotide level. We know that TFs bind to enhancers, which contain clusters of sequence-specific binding motifs. A study using Ciona intestinalis embryos showed that the order, orientation, and spacing (referred to as “syntax”) of TF binding sites in enhancers play an important role in their function. They showed that enhancers with low affinity TF binding sites can still establish robust tissue specific gene expression patterns, because of compensation by an optimal syntax. The observation that enhancer syntax is important for gene regulation, suggests a more complex mechanism for enhancer function than just TF binding to high affinity consensus binding motifs. Therefore, it is important that we not only focus on high affinity consensus motifs, but also on low sequence specificity and syntax, to better understand the rules of enhancer function.

To understand the molecular mechanism of gene regulation, it is important to find out which TFs are involved in activating gene expression. There are several methods available to identify these proteins. First, one could analyze single-cell RNA sequencing (scRNA-seq) datasets to
uncover co-expression of TFs and the gene of interest. This can yield a list of candidate regulators, which can be further validated by knock-out or knock-down experiments, to analyze a possible link with the gene of interest. Another method that can identify relevant TFs, is unbiased genomic locus proteomics (GLoPro), which uses a dCas9 fused to the peroxidase enzyme APEX2 (CASPEX) (Figure 2A)\(^85\). Using gRNAs, specific genomic locations such as promoters and enhancers can be targeted. At these DNA regions, APEX2 can label proteins with biotin derivates that are in close proximity, which can be identified using quantitative mass spectrometry analysis. Proximity labeling can also be used to identify proteins that interact with long non-coding RNAs (lncRNAs). We made a start with the preparations for such an analysis by generating BC44 and HC11 cell lines that stably express and doxycycline inducible CASPEX (data not shown). Recently, CRISPR-assisted RNA-protein interaction detection (CARPID) was introduced, which uses dCasRx-based RNA targeting (Figure 2B)\(^86\). dCasRx can be guided to RNA transcripts, but, similar to dCas9, does not cleave the target. The biotin ligase BASU is fused to dCasRx and will biotinylate proteins that are close proximity, such as RNA binding proteins (RBPs). In chapter 4 we discussed a possible role of lncRNA Gm13003 in Wnt4 regulation. Using a technique such as CARPID, possible regulators can be identified that are involved in Wnt4 expression via Gm13003 binding.

![Figure 2: Proximity labeling techniques to identify TFs and RBPs. A) Genomic locus proteomics (GLoPro)\(^85\): Illustration of a genomic locus, containing one gene and two enhancers. dCas9-APEX2 (CASPEX) is guided to the promoter region using specific gRNAs, where APEX2 biotinylates proteins that are in close proximity. B) CRISPR-assisted RNA-protein interaction detection (CARPID)\(^86\): Illustration of a lncRNA with RNA binding proteins (RBP). Using specific gRNAs, dCasRx-BASU is guided to the lncRNA of interest. dCasRx can target RNA transcripts, but does have nuclease activity. BASU biotinylates proteins that are in close proximity, such as RBPs. Proteins that are biotinylated in the GLoPro or CARPID workflow can be identified using Western blot or mass spectrometry analysis.]
Another element of enhancer function, which we did not study in this thesis, is the role of enhancer RNAs (eRNAs). eRNAs are transcripts from genomic enhancer sequences and have been implicated in gene regulation\textsuperscript{87–92}. Multiple studies have shown evidence that eRNAs can interact with epigenetic modifying enzymes or can stabilize chromatin loops that promote enhancer-promoter interactions\textsuperscript{88,93–95}. Their exact mechanism of action however, remains unknown. It has been suggested that enhancer transcription itself also plays a role in methylation of H3K4, independent of eRNA transcripts\textsuperscript{96}. The idea that enhancer transcription, in addition to the resulting eRNAs, can play a role in gene regulation, was also indicated by a study that showed interference of gene transcription by transcription of intragenic enhancers\textsuperscript{97}. In chapter 2, we showed a transcriptomic analysis from different mammary gland cell populations, but this experiment was not sensitive enough to detect eRNAs\textsuperscript{98}. For eRNA detection, global run-on sequencing (GRO-seq)\textsuperscript{99,100}, precision run-on nuclear sequencing (PRO-seq)\textsuperscript{101}, or cap analysis gene expression (CAGE)\textsuperscript{102–104} can be used. To understand the role of eRNAs in gene regulation, it is imperative that we invest in cataloging eRNA transcripts and their spatio-temporal expression patterns.

Concluding remarks
In this thesis we have shown for the first time a detailed functional dissection of $Wnt$ loci in mammary gland cells. This has given us new insights into how cell type specific expression patterns of $Wnt$ genes are established in the mammary gland. Ultimately, this will help us better understand how Wnt ligands can regulate cell behavior in an organized spatio-temporal manner, thereby controlling tissue development and homeostasis of the mammary gland.

These are exciting times in the field of enhancer biology, in which these new technologies enable us to further advance our understanding of gene regulation. Despite this progress, many questions remain unanswered\textsuperscript{45,105–108}: What is the role of TADs in gene regulation? Does the 3D structure of the chromatin regulate gene expression or is it, at least in some cases, merely the result of enhancer-promoter communication? What is the exact mechanism of action of enhancers? What is the minimum distance in 3D between enhancers and promoters to trigger transcription? How does an enhancer know which promoter to act on? How is enhancer redundancy coordinated? What is the exact role of enhancer syntax and how specific is this? What is the role of eRNAs, or of the total interaction between DNA, RNA and proteins in enhancer hubs for that matter? Performing functional dissections of individual genomic loci, enhancers, their syntax, their transcripts, and chromatin context using advanced technologies will help us unravel the fascinating mechanisms of gene regulation in the coming years.
References


