Exploring the Wnt enhancer landscape in the mammary gland

Heijmans, N.

Publication date
2021

Citation for published version (APA):
Chapter 4

Functional dissection of the Wnt4 locus to identify regulatory elements that control expression in the mammary gland

Nika Heijmans*, Katrin E. Wiese*, Yorick van de Grift, Renée van Amerongen

Section of Molecular Cytology and Van Leeuwenhoek Centre for Advanced Microscopy, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands.

*contributed equally
Abstract

WNT4 is a signaling molecule that is critical for the development and maintenance of epithelial branches of the mouse mammary gland. Despite its essential function, the exact mechanisms by which Wnt4 expression is regulated at the chromatin level, remain unknown. The aim of this study was to identify regulatory elements that control Wnt4 expression in the mammary gland. By functionally dissecting the topologically associating domain of Wnt4, we identified an active enhancer hub that is linked to the Wnt4 gene. In addition, we focused on Gm13003, a lncRNA in the Wnt4 TAD, as a potential regulator of Wnt4 expression. Together, we provide a detailed overview of regulatory elements that are located in the vicinity of Wnt4, which brings us one step closer to resolving the mechanism that controls expression of this important gene.
Introduction

Wingless-type mouse mammary tumor virus (MMTV) integration site family, member 4 (WNT4) is a highly conserved, locally acting signaling molecule that is important for stem cell regulation in various tissues\(^1\)\(^\text{-}^8\). In the mouse mammary gland, WNT4 is known for its role in branching morphogenesis, at least partially by maintaining the mammary stem cell (MaSC) population and controlling its behaviour\(^9\)\(^\text{-}^12\). Expression of \textit{Wnt4} in the mammary gland is mainly reported to be downstream of progesterone\(^9\)\(^\text{-}^12\), but there is evidence that \textit{Wnt4} can also be expressed independently of progesterone: perinatal expression of \textit{Wnt4} in the mammary gland has been reported to be independent of estrogen receptor a (ER\(a\)) and progesterone receptor (PR) signaling\(^10\). Furthermore, the more significant impairment of outgrowth of \textit{Wnt4}\(^{-/}\) compared to \textit{PR}\(^{-/}\) epithelial tissue after transplantation, also indicates that WNT4 can function as a MaSC regulator independent of progesterone\(^10\). This suggests that \textit{Wnt4} expression in the mammary gland can be regulated dependent or independent of steroid hormones. However, for both mechanisms it is unknown how specific spatiotemporal expression is regulated at the chromatin level.

Tissue-specific gene expression can be controlled by regulatory sequences such as enhancers or repressors, which are mostly located in the non-coding genome. Currently, no regulatory sequences have been described in the literature that control expression of \textit{Wnt4}. The goal of this study is to identify genomic elements that regulate \textit{Wnt4} expression in the mammary gland. Finding regulatory sequences in the genome is challenging, since they don’t have a fixed location or consensus sequence\(^13\). Besides, there are still many questions in the field about how enhancers carry out their function exactly.

Recent studies have taught us that the genome is divided into topologically associating domains (TADs) (Figure 1). The boundaries of TADs are called insulators and contain binding sites for insulator proteins, such as CCCTC-binding factor (CTCF)\(^14\)\(^,^15\). Evidence suggests that in the process of organizing the 3D structure of chromatin, the cohesin protein complex actively forms large DNA loops in a process called loop extrusion\(^16\). When cohesin encounters a boundary protein such as CTCF, CTCF acts as an extrusion barrier and loop extrusion is stalled\(^17\). The orientation of CTCF binding sites at TAD boundaries is relevant, as loops are mostly formed between converging CTCF binding sites\(^15\)\(^,^18\).

Chromatin interactions between enhancers and promoters occur predominantly within a TAD\(^19\)\(^\text{-}^21\). Interestingly, genes located in the same TAD can share expression profiles across cell types and tissues\(^20\)\(^,^22\)\(^,^23\). Besides organizing the co-regulation of gene expression in the same TAD, another important function of TAD formation is to avoid inappropriate DNA interactions between promoters and regulatory sequences. A well-known example in which this function is demonstrated, is the rearrangement of TAD boundaries in the WNT6/IHH/EPHA4/PAX3 locus, resulting in misexpression of developmental genes. Disruption of the TADs in this locus is associated with malformation syndromes in mice and humans\(^24\). Thus, although the effect of
3D genome organization on gene expression differs between loci\textsuperscript{25}, it is clear that it can play an important role in facilitating interactions between promoters and regulatory sequences\textsuperscript{24,26,27}.

![Diagram of topologically associating domains (TADs)](image)

**Figure 1: The genome is divided into topologically associating domains (TADs).** Graphical representation of two different TADs, separated by insulator regions. Arrows within the TADs indicate communication between enhancers and promoters.

Given that DNA regions within a TAD interact more frequently with regions in the same TAD\textsuperscript{19–21}, we decided to start our search for \textit{Wnt4} regulatory sequences in the \textit{Wnt4} TAD by performing a functional dissection of this region. Based on publicly available datasets containing information about active enhancer marks or chromatin accessibility, we selected candidate regulatory sequences (CRSs) that are located in this TAD. We identified an active enhancer hub located in the \textit{Wnt4} TAD that is linked to the \textit{Wnt4} gene. Furthermore, our results indicate a link between \textit{Wnt4} and \textit{Gm13003}, a lncRNA located in the \textit{Wnt4} TAD, >100 kb upstream from the \textit{Wnt4} gene itself. This study does not only provide the first detailed description of regulatory sequences that are linked to the \textit{Wnt4} gene, but is also one of the few to dissect an entire TAD in such detail.
Results

Identification of the Wnt4 TAD

A functional dissection of the Wnt4 TAD requires first of all the identification of the TAD in which the Wnt4 gene is located. We used the 3D Genome Browser to visualize genome interactions and TAD predictions of the Wnt4 locus from publicly available Hi-C datasets (Figure 2). Figure 2A shows an example of one of those Hi-C datasets in which Wnt4 is the only annotated gene in the predicted TAD. The TAD predictions that are visualized in the 3D Genome Browser are calculated according to the so-called directionality index, which is a method that looks at the degree of up- and downstream interaction bias for DNA regions, since it was noted that DNA regions at the periphery of TADs are highly biased in direction of interaction. Upstream regions in a TAD are highly biased towards interacting with downstream regions and vice versa. Using this directional bias, the boundaries of TADs can be predicted. In the depicted triangle, the physical interaction frequency of DNA regions is represented by the color intensity (Figure 2A). A dark blue spot at the top of the triangle connects the TAD boundaries of the predicted Wnt4 TAD, indicating that these genomic regions were found to frequently interact in this particular Hi-C dataset. By aligning a CTCF ChIP-seq track from mammary gland cells, we observed CTCF peaks at the predicted boundaries of the Wnt4 TAD. These double peaks and the orientation of the binding motifs are in line with earlier evidence that suggests that chromatin loops preferentially form between convergently oriented CTCF sites.

We downloaded all 95 TAD coordinate lists from the 3D Genome Browser and aligned the predicted TADs in which the Wnt4 gene is located (Figure 2B). As no specific mouse mammary gland Hi-C dataset was available and TADs have been reported to be conserved across cell types and species, we included mouse and human data from different tissues and cell types. Despite varying lengths, there was an overlapping region of 250 kb that contains boundaries marked by CTCF. This region was selected for a thorough analysis to find genomic elements that regulate Wnt4 expression in the mammary gland.

Selection of candidate regulatory sequences

Before focusing on the regulatory elements of Wnt4, we looked into the expression of Wnt4 in different mammary cell types. RNA in situ hybridization on paraffin sections of the mouse mammary gland showed that Wnt4 is mostly expressed in the luminal layer of the epithelial ducts. Lower Wnt4 levels are detectable in the stroma (Figure 3A). This observation was validated by RNA-sequencing of sorted primary mammary cells (Figure 3B). These results are in accordance with the literature, and confirm that we should focus our search especially on the luminal cell population.
Figure 2: Identification of the topologically associating domain of the Wnt4 locus. A) Hi-C data from CH12 cells, dataset by Rao et al. visualized in the 3D Genome Browser\cite{rao2014}. Coordinates of the total depicted region: chr4:136200000-137250000 (Mm9). Coordinates of the Wnt4 TAD: chr4:136625000-136875000. CTCF ChIP-seq on whole mammary gland tissue is from Shin et al.\cite{shin2018}. TADs are predicted by the 3D Genome Browser, using
Functional dissection of the Wnt4 locus

To select candidate sequences in the Wnt4 TAD that could potentially regulate Wnt4 expression in the mammary gland, we focused on chromatin regions that are accessible for TFs, as open regions have been associated with regulatory elements. We examined publicly available datasets including ATAC-seq and H3K27ac ChIP-seq data from mature luminal, luminal progenitor and basal cells from adult mammary glands. In addition, we generated our own ATAC-seq dataset of luminal and basal cells from pubertal mouse mammary glands. Since Wnt4 is mainly expressed in luminal cells, we selected most candidate sequences (candidates 1-6, 8-13, 15, 17) based on the ATAC-seq dataset and the provided called peaks of mature luminal cells by Dravis et al. Additionally, we selected candidate sequences showing specific ATAC-seq
peaks in basal cells (candidates 7, 14 and 16), to investigate the difference between luminal and basal specific peaks. Three sequences that showed no peaks in the ATAC-seq or H3K27ac ChIP-seq tracks were chosen as negative control regions (n1-n3). Seven of the candidate sequences are located in introns of either Zbtb40 (candidates 1 and 2), which is a neighboring gene, or Wnt4 itself (candidates 12 to 16). Part of candidate 6 overlaps with lncRNA Gm13003. In conclusion, based on datasets for chromatin accessibility or the presence of a known active enhancer mark, we selected 20 regions for further experiments, of which 3 are negative controls and 17 are candidate regulatory sequences (CRSs) (Figure 4B).
Figure 4: Selection of CRSs. A) Visualization of ATAC-seq and ChIP-seq datasets in the IGV Browser\textsuperscript{41}. CRSs = candidate regulatory sequences. The coordinates of the depicted region are chr4:136525000-136925000 (Mm9). Pink arrows indicate position of CRSs, black arrows are negative controls. L = luminal, B = basal, ML = mature luminal, LP = luminal progenitor, MG = mammary gland. We generated the puberty ATAC-seq dataset; adult ATAC-seq and H3K27ac ChIP-seq data are from Dravis et al.\textsuperscript{42}; Shin et al.\textsuperscript{29} published the CTCF ChIP-seq data. B) Zoom into selected peaks in the IGV Browser. Selected candidate sequences are numbered according to genomic location, n1-n3 are negative controls. Seven candidates are located intronic in Zbtb40 or Wnt4 and one candidate (CRS-6) is partially overlapping with lncRNA Gm13003.

Dual luciferase reporter assays reveal transcriptional enhancer activity of CRSs

The first property of the CRSs we investigated was their functional enhancer activity. The gold standard assay for determining enhancer activity is the reporter assay, where the expression of a reporter gene is in direct relation to the strength of the enhancer\textsuperscript{44}. For our reporter assays we selected the BC44 and HC11 cell lines, which are both mammary epithelial cell lines\textsuperscript{45,46}. Wnt4 is expressed in both cell lines (Supplementary Figure 2). However, there is a clear difference in expression levels (BC44: 69.2-77.6 RPKM, hereafter Wnt4\textsuperscript{high}; HC11: 2.8-3.2 RPKM, hereafter Wnt4\textsuperscript{low}). Both BC44 (Wnt4\textsuperscript{high}) and HC11 (Wnt4\textsuperscript{low}) were selected for further experiments, as this would give us the opportunity to find out whether the enhancer activity of CRS is different in Wnt4\textsuperscript{high} or Wnt4\textsuperscript{low} cells.

To test the transcriptional enhancer activity of the CRSs, these sequences needed to be cloned into the vector containing the reporter gene. To reduce size variability, we adjusted the lengths of the sequences (362-2596 base pairs (bp) originally, 664-1217 bp adjusted, Table 1). We added sequences (max. 200 bp, according to reference genome Mm10) to both the 5’ and 3’ ends of the core sequence that was originally called as a peak in the ATAC-seq datasets from Figure 4\textsuperscript{42}. The modification of the core sequence differed per CRS and was based on its original length, sequence complexity (repetitive stretches were avoided), the proximity to a transcriptional start site (TSS) (CRS-12), or proximity to other CRSs (CRS-14).

The CRSs were cloned into a pGL4-minP plasmid, which contains a minimal promoter and a firefly luciferase reporter gene (luc2) (Figure 5A). BC44 and HC11 cells were transiently transfected with a pGL4-minP plasmid containing a CRS and a CMV-Renilla plasmid for normalization. The amount of firefly luciferase in the cell lysate revealed whether a CRS was able to induce transcription of this gene compared to the empty vector (EV), for which a pGL4-minP plasmid without a CRS was used. Figure 5B shows the log2 fold change of firefly luciferase activity compared to the EV. Despite the difference in baseline expression levels of Wnt4, no clear difference was observed in transcriptional enhancer activity of the CRSs between BC44 and HC11 cells. Out of 17 candidates, 7 sequences (CRS-1,-6,-7,-8,-10,-11,-17) showed an average log2 fold change of >1 in firefly luminescence signal compared to the EV in BC44 cells. In HC11 cells, this was the case for 8 sequences (CRS-1,-6,-7,-8,-10,-11,-15,-17). CRS-7 showed the highest activity (average log2 fold changes of 3.2 and 2.6 in BC44 and HC11, respectively). No transcriptional enhancer activity was observed for CRS-2,-3,-4,-5,-9,-12,-13,-14 and -16. Negative controls n1 and n2 also showed no increase in average luciferase luminescence signal compared to the EV. In contrast, negative control n3 did show enhancer activity in both cell lines. This indicates that the
n3 sequences contains binding sites for transcriptional regulators that can induce transcription of the firefly luciferase gene in mouse mammary epithelial cell lines, even though no ATAC-seq or H3K27ac peaks were observed in this region (Figure 4B). Interestingly, as the CRSs in Figure 5B are ordered according to their genomic location, a hub of active enhancers can be observed ranging from CRS-6 to CRS-11. With the exception of CRS-9, these CRSs all showed an average log2 fold change of >1 in firefly luminescence signal compared to the EV in both cell lines. To conclude, dual luciferase assays showed for each individual CRS whether it has transcriptional enhancer activity and revealed a possible enhancer hub, located upstream of the Wnt4 gene.

<table>
<thead>
<tr>
<th>Candidate #</th>
<th>Length peak (bp)</th>
<th>Adjusted length (bp)</th>
<th>Coordinates cloned sequences (Mm10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>785</td>
<td>1185</td>
<td>chr4:136980770-136981954</td>
</tr>
<tr>
<td>2</td>
<td>1013</td>
<td>1217</td>
<td>chr4:136997198-136998414</td>
</tr>
<tr>
<td>3</td>
<td>362</td>
<td>762</td>
<td>chr4:137077015-137077776</td>
</tr>
<tr>
<td>4</td>
<td>1014</td>
<td>1014</td>
<td>chr4:137097326-137098339</td>
</tr>
<tr>
<td>5</td>
<td>686</td>
<td>726</td>
<td>chr4:137121443-137122168</td>
</tr>
<tr>
<td>6</td>
<td>1084</td>
<td>1216</td>
<td>chr4:137164242-137165457</td>
</tr>
<tr>
<td>7</td>
<td>1552</td>
<td>1217</td>
<td>chr4:137200440-137201656</td>
</tr>
<tr>
<td>8</td>
<td>833</td>
<td>1217</td>
<td>chr4:137209519-137210735</td>
</tr>
<tr>
<td>9</td>
<td>743</td>
<td>1143</td>
<td>chr4:137231229-137232371</td>
</tr>
<tr>
<td>10</td>
<td>408</td>
<td>664</td>
<td>chr4:137255464-137256127</td>
</tr>
<tr>
<td>11</td>
<td>305</td>
<td>705</td>
<td>chr4:137263755-137264459</td>
</tr>
<tr>
<td>12</td>
<td>800</td>
<td>740</td>
<td>chr4:137279224-137279963</td>
</tr>
<tr>
<td>13</td>
<td>764</td>
<td>1164</td>
<td>chr4:137281350-137282513</td>
</tr>
<tr>
<td>14</td>
<td>2596</td>
<td>941</td>
<td>chr4:137283024-137283964</td>
</tr>
<tr>
<td>15</td>
<td>686</td>
<td>786</td>
<td>chr4:137287060-137287845</td>
</tr>
<tr>
<td>16</td>
<td>900</td>
<td>1140</td>
<td>chr4:137289790-137290929</td>
</tr>
<tr>
<td>17</td>
<td>564</td>
<td>964</td>
<td>chr4:137307093-137308056</td>
</tr>
<tr>
<td>n1</td>
<td>X</td>
<td>1201</td>
<td>chr4:137070585-137071785</td>
</tr>
<tr>
<td>n2</td>
<td>X</td>
<td>1191</td>
<td>chr4:137138885-137140075</td>
</tr>
<tr>
<td>n3</td>
<td>X</td>
<td>1201</td>
<td>chr4:137243585-137244785</td>
</tr>
</tbody>
</table>

Table 1: Coordinates of candidate regulator sequences. This table contains the coordinates of the selected CRSs (Mm10), based on the information in figure 4 (column: Length peak). Since the sizes of the selected peaks in figure 4 were variable, we adjusted the length of the CRSs (column: Adjusted length). The Mm10 coordinates of these sequences are shown in the last column and these regions were cloned individually into pGL4-minP. Note that the coordinates of 4 CRSs partially overlap with candidate Wnt4 enhancers that were investigated in chapter 3 (4.5/CRS-15; 4.6/CRS-9; 4.7/CRS-14; 4.8/CRS-7).
Figure 5: Enhancer activity of CRSs in BC44 and HC11 cells measured in transient dual luciferase reporter assays. A) Schematic representation of plasmids used in the dual luciferase assay. Cells were transfected with pGL4-minP, which contains the CRS, a minimal promoter and a firefly luciferase gene (luc2), and CMV-Renilla as a transfection control. B) Graph showing the log2 fold change in firefly luminescence signal for each candidate enhancer compared to the EV in BC44 (blue) and HC11 (red) cells. EV is a pGL4-minP plasmid without a CRS. CRSs are on the X-axis, ordered according to their genomic location. Every dot represents one biological replicate (n=4/5, measured in technical triplicates, not depicted). Bars show the mean of replicates and are depicted as visual guidance for the individual data points. Error bars represent the SD of biological replicates.

Functionally linking CRSs to the Wnt4 gene

By performing luciferase reporter assays, we learned which CRSs have the ability to increase transcription. Although this gives us important information regarding the potential enhancer activity of these CRSs, these experiments cannot show us whether these enhancers are in fact Wnt4 enhancers in situ. To functionally link the CRSs to the Wnt4 promoter, we used a CRISPR-based transcriptional activator called dCas9-VPR (Figure 6A). dCas9-VPR is a fusion of a nuclease dead Cas9 with VPR, which consists of a transcriptional activator (Vp64) and two transcription...
factors (TFs) (p65 and Rta)\textsuperscript{47}. The possibility of guiding dCas9-VPR to specific target sequences gave us the opportunity to investigate the link of individual CRSs to the \textit{Wnt4} gene, without removing them from their endogenous genomic context (Figure 6B).

For each individual CRS we designed guideRNAs (gRNAs) that would allow us to direct dCas9-VPR to that particular CRS. The number of gRNAs we cloned per CRS was dependent on their length (approximately 1 gRNA per 150 bp, distributed over the CRS (Supplementary Table 1)). We also cloned gRNAs for negative controls n1-n3 (Figure 6B, black arrows) and for a region 200 bp upstream of the \textit{Wnt4} TSS (Figure 6B, blue arrow).

HC11 and BC44 cells were transfected with a plasmid containing dCas9-VPR and pSpgRNA plasmids containing the gRNAs, which were pooled per CRS. To determine whether guiding a transcriptional activator to a CRS had a transcriptional effect on \textit{Wnt4}, we measured \textit{Wnt4} mRNA levels 48 hours after transfection by qRT-PCR. Empty vector (EV) controls were transfected with dCas9-VPR and an empty pSpgRNA (i.e. not containing a gRNA). Figure 6C shows the fold change in \textit{Wnt4} expression compared to the EV control in HC11 cells. Eleven CRSs (CRS-3, -6, -7, -8, -9, -10, -11, -12, -13, -15, 17) showed a fold change of >2 compared to the control. CRS-6, -7 and -11 show the highest average fold changes (8.1, 9.7, and 8.0, respectively). Interestingly, whereas guiding dCas9-VPR to most CRSs in the \textit{Wnt4} TAD induced some upregulation of \textit{Wnt4} expression, \textit{Wnt4} mRNA levels were not affected when dCas9-VPR was guided to CRS-1 and CRS-2, which are located outside the predicted \textit{Wnt4} TAD. Negative controls n2 and n3 also show an average fold change of >2 compared to the control (2.1 and 3.0, respectively). Contrary to our expectations, \textit{Wnt4} expression was decreased when dCas9-VPR was guided to the promoter region (P).

In BC44 cells, the effect of dCas9-VPR guided to CRSs on \textit{Wnt4} expression was much lower compared to HC11 (Figure 6D). In multiple prior experiments, we had already observed a negative correlation between endogenous gene expression and the effect of dCas9-VPR on gene expression (Chapter 3, Figure 5C; Chapter 5, Supplementary Figure 3), which has also been reported in the literature\textsuperscript{47,48}. As \textit{Wnt4} expression is much higher in BC44 than in HC11 (Supplementary Figure 2), this adds to the evidence that endogenous expression of a gene affects the possible range of inducing gene expression by dCas9-VPR. Similar to what we observed in HC11 cells, \textit{Wnt4} expression was reduced in BC44 cells when dCas9-VPR was guided to the \textit{Wnt4} promoter (P). Although the fold changes in \textit{Wnt4} expression in BC44 cells were lower in general, we observed that a similar group of CRSs showed a higher induction of \textit{Wnt4} expression, namely CRS-6 to CRS-11. Interestingly, this is the same hub of CRSs that showed the highest fold changes in luminescence signal in the luciferase assays (Figure 5B). The fact that CRSs in this hub are not only showing the most enhancer activity, but also the highest increase in \textit{Wnt4} expression in the dCas9-VPR experiments, suggests a role of this specific region in regulating \textit{Wnt4} expression.
Functional dissection of the Wnt4 locus

Figure 6: Upregulation of Wnt4 upon guiding of dCas9-VPR to CRSs. A) Illustration of dCas9-VPR being guided to a CRS using gRNAs specific for that CRS. The dashed grey arrow represents the transcriptional signal that is transferred from the CRS to the Wnt4 promoter. B) Illustration of the Wnt4 locus and the location of the CRSs. Zbtb40 and Cdc42 are the neighboring genes of Wnt4, located outside the selected TAD. Gm13003 is an annotated long non-coding RNA (lncRNA). Pink arrows show the location of the CRSs, black arrows are the negative controls and blue arrow represents the location of the Wnt4 promoter gRNAs. C) Graph depicting Wnt4 expression upon guiding of dCas9-VPR to each individual CRS in HC11 (Wnt4<sup>lo</sup>) cells. CRSs are ordered according to their genomic location. Ctr = control without gRNAs, set to 1 (grey line). N1-3 are negative control regions. P = Wnt4 promoter. Wnt4 RNA levels were measured by qRT-PCR. Ctbp1 was used as housekeeping gene<sup>49</sup>. Biological replicates (n=4) are depicted as yellow dots, bars show the mean of these replicates. Error bars represent the SD. D) Graph depicting Wnt4 expression upon guiding of dCas9-VPR to each individual CRS in BC44 (Wnt4<sup>hi</sup>) cells (n=3 biological replicates each). Experimental details are similar to C).
Neighboring genes of \textit{Wnt4} are not influenced by dCas9-VPR being guided to CRSs

To determine whether the increase in gene expression observed after guiding dCas9-VPR to the CRSs was specific for \textit{Wnt4}, we checked the expression of its two neighboring genes in the same samples as \textbf{figure 6} using qRT-PCR. Located just outside the selected TAD, up- and downstream of \textit{Wnt4} respectively, are genes \textit{Zbtb40} and \textit{Cdc42} (\textbf{Figure 7A}). Importantly, expression of these neighboring genes was not affected when dCas9-VPR was guided to the individual CRSs (\textbf{Figure 7B} and 7C) – even when the CRS in question was closer to the TSS of \textit{Zbtb40} or \textit{Cdc42} than the TSS of \textit{Wnt4}. Accordingly, the links between CRSs and the \textit{Wnt4} gene that we observed using dCas9-VPR seem to be specific and exclusive.

Because we had previously observed a negative correlation between endogenous \textit{Wnt4} expression and the dynamic range of dCas9-VPR (\textbf{Figure 6C} and 6D), we compared gene expression levels of \textit{Zbtb40} and \textit{Cdc42} in HC11 and BC44 cells (\textbf{Figure 7D}). \textit{Cdc42} shows higher expression values than \textit{Wnt4} in both BC44 (143.0-153.1 RPKM) and HC11 (151.0-152.3 RPKM). As discussed before, this may have affected the ability of dCas9-VPR to increase expression of this gene. For \textit{Zbtb40} however, this is not the case, since this gene shows similarly low RPKM values as \textit{Wnt4} in HC11 cells (\textit{Zbtb40}: 1.8-2.2 RPKM in BC44; 1.8-2.6 RPKM in HC11).

To summarize, the functional link we observed between CRSs in this locus and \textit{Wnt4} in dCas9-VPR assays is specific for \textit{Wnt4}. Although potential effects on \textit{Cdc42} may have been obscured as a result of its higher endogenous expression levels, we can confidently link a subset of our CRSs to \textit{Wnt4} and conclude that they do not regulate \textit{Zbtb40}.

\textbf{Figure 7: Upregulation of \textit{Wnt4} upon guiding of dCas9-VPR to CRSs is \textit{Wnt4}-specific in this locus.} A) Illustration of the \textit{Wnt4} locus and the location of the CRSs. This focuses on the expression of \textit{Zbtb40} and \textit{Cdc42}, so these genes are highlighted in green and purple, respectively. The locations of CRSs are indicated by pink arrows and negative controls are indicated by black arrows. The blue arrow represents the \textit{Wnt4} promoter. B) Graph depicting \textit{Zbtb40} and \textit{Cdc42} expression upon guiding of dCas9-VPR to each individual CRS in HC11 (\textit{Wnt4} high) cells. Before transfection, gRNAs for each CRS were pooled. Cells were transfected with dCas9-VPR and pSpgRNA, which contain gRNAs. Samples were harvested 48h after transfection. Ctr = control without gRNAs, which is set to 1 (grey line). P = \textit{Wnt4} promoter. The average fold change of \textit{Wnt4} expression (yellow crosses) compared to EV is shown for comparison. Individual data points of \textit{Wnt4} expression are shown in figure 6. C) Same as B), but now for BC44 (\textit{Wnt4} low) cells. D) Using RNA-sequencing, gene expression of \textit{Zbtb40}, \textit{Cdc42} and \textit{Wnt4} was determined in BC44 and HC11 cell lines. This table shows the RPKM (= Read Per Kilobase per Million mapped reads) values from two different replicates from each cell line. \textit{Wnt4} expression values the same as in supplementary figure 2, shown again here for comparison.
Functional dissection of the Wnt4 locus

A

Refseq genes
selected TAD
CRSSs
Mm9, chr:4

B

Zbtb40 and Cdc42 expression upon guiding dCas9-VPR to CRSs in HC11 (Wnt4\textsuperscript{low}) cells

C

Zbtb40 and Cdc42 expression upon guiding dCas9-VPR to CRSs in BC44 (Wnt4\textsuperscript{high}) cells

D

Endogenous expression of Zbtb40, Cdc42 and Wnt4

BC44 (Wnt4\textsuperscript{high})
HC11 (Wnt4\textsuperscript{low})
CRS-6 is in close physical proximity to the Wnt4 promoter

So far, we have established a link between an active enhancer hub, which is located between Gm13003 and the Wnt4 promoter, and the Wnt4 gene. However, we don’t know how these sequences communicate with each other. In fact, how transcriptional information is transferred from regulatory sequences to promoters, remains a big question in the field of gene regulation. One of the mechanisms that has been proposed is called “looping”, which is the formation of chromatin loops that bring the regulatory sequence and promoter in close physical proximity. How the genome is organized in 3D, can be investigated using Chromosome Conformation Capture (3C), which enables quantification of interactions between two different genomic loci. This technique was introduced in 2002 and over the years several new variants have been developed. One of these variants is 4C, which allows for the detection of genome wide physical interactions with any given locus of interest, the so-called viewpoint. Since we are interested in the interactions of the Wnt4 promoter, we selected a sequence close to the Wnt4 TSS (660 bp upstream) as the viewpoint to investigate which DNA regions are in close spatial proximity. Figure 8A and 8B show the quantification of interactions of genomic regions with this viewpoint in HC11 and BC44 cells, respectively. In these figures, the height of the peaks represents the frequency of interactions of that genomic region with the viewpoint (blue triangle). We found that DNA regions within the selected TAD show a higher interaction frequency overall than the DNA outside this TAD, especially in BC44 cells. The red peaks represent statistically significant interactions between the viewpoint and that particular DNA region. In both BC44 and HC11 cells, two regions show statistically significant interactions with the Wnt4 viewpoint. One of those, containing part of the genomic sequence of Gm13003, is detected in both cell lines. Importantly, this region overlaps with CRS-6, which also showed enhancer activity in the luciferase assays and was linked to the Wnt4 gene in dCas9-VPR assays (Figure 5 and 6). In conclusion, our 4C results show that in both HC11 and BC44 cell lines the Wnt4 promoter is in close physical proximity to Gm13003, which overlaps with CRS-6.
Figure 8: Physical interaction between the Wnt4 promoter and CRS-6. Representation of physical interactions between the viewpoint and other regions in the Wnt4 locus in HC11 (A) or BC44 (B) cells using 4C. The blue triangle represents the viewpoint, which is close (660 bp) to the Wnt4 TSS (coordinates transcript Wnt4 in Mm9: chr4:136833550-136852694). The pink arrows show the location of the CRSs and the black arrows are the negative controls. The light pink bar represented the selected TAD of approximately 250 kb. Statistically significant interactions are indicated by red bars. For each cell line, cumulative analysis of biological triplicates are shown.
Expression of Wnt4 and Gm13003 is correlated in cell lines and primary mammary cell types.

Gm13003 is a long non-coding RNA (lncRNA), located between Zbtb40 and Wnt4 in the mouse genome. This lncRNA is not well studied, however it is upregulated >2-fold in organoids derived from mammary carcinomas compared to normal mammary organoids\(^5\). In the past decades, a growing body of evidence has shown that lncRNAs play an important role in gene regulation and help to achieve transcriptional specificity. Not only can they recruit chromatin modifying proteins\(^59-63\), they are also key regulators in 3D genome organization\(^64-67\).

To look further into the link between Wnt4 and lncRNA Gm13003, we analyzed the expression of Gm13003 in BC44 and HC11 cells (Figure 9A). Similar to the expression of Wnt4 in these cell lines (Supplementary Figure 2), Gm13003 is also higher expressed in BC44 than in HC11 cells. We also isolated primary cells from different mammary gland cell types, to confirm that Wnt4 and Gm13003 expression also correlate in vivo. Freshly isolated mouse mammary glands were enzymatically digested and the different populations were separated using FACS (Supplementary Figure 3). Like Wnt4, we find Gm13003 to be mainly expressed in luminal cells (Figure 9B). To conclude, the expression levels of Wnt4 and Gm13003 correlate in different mouse mammary cell lines and primary cell types.

Since Wnt4 and Gm13003 are in close physical proximity in the nucleus (Figure 8) and their expression patterns are correlated across mammary cell lines and cell types (Figure 9), we wondered whether guiding dCas9-VPR to the different CRSs would also increase transcription of Gm13003, similar to Wnt4 expression. To this end, we measured the RNA levels of Gm13003 by qRT-PCR (Figure 10A) in the samples that were previously used in the dCas9-VPR assays (Figure 6 and 7).

CRS-6 overlaps with the Gm13003 TSS and gRNAs for CRS-6 are surrounding this TSS: 4 gRNAs are located up- and 4 gRNAs are located downstream. Guiding dCas9-VPR to CRS-6 resulted in an increase in Gm13003 expression in both HC11 and BC44 cells (Figure 10A and 10B). Thus, unlike Wnt4, Gm13003 could be upregulated when dCas9-VPR was guided to its own promoter. Interestingly, the previously identified hub of active enhancers between Gm13003 and the Wnt4 promoter (CRS-6 to -11) are amongst the CRSs that show the highest fold increase in Gm13003 expression in both HC11 and BC44 cells. In the dCas9-VPR assays measuring Wnt4 or Gm13003 expression, there is a clear overlap of CRSs that show the highest fold change in gene induction, with CRS-6, -7 and -10 in the top 5 for regulating both Wnt4 and Gm13003 expression in both HC11 and BC44 cells. This overlap in regulatory sequences in combination with our earlier observations that the promoters of Wnt4 and Gm13003 are in close physical proximity (Figure 8), and that their expression is correlated in different mammary cell types and cell lines (Figure 9), indicate that Gm13003 might play a regulatory role in Wnt4 expression in the mouse mammary gland.
Functional dissection of the Wnt4 locus

**Figure 9:** Expression of Wnt4 and IncRNA Gm13003 is correlated in BC44 and HC11 cell lines and primary mammary cells. A) Graph depicting qRT-PCR results from BC44 and HC11 cells. Ctbp1 was used as housekeeping gene and the highest expression was set to 1 (n = 3 or 4 biological replicates). B) Graph showing results of qRT-PCR experiments of different mammary gland cell populations from 6 different adult C57BL/6J mice. Luminal (L), basal (B) and stromal (S) cells were isolated using FACS. Ctbp1 was used as housekeeping gene in the qRT-PCR experiments. Values were normalized to the expression values of the cell type that showed highest expression. For visualization purposes, only Wnt4, Gm13003 and Krt18 are shown in this figure. For expression of other markers see supplementary figure 3.

### Legend Figure 11

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>BC44</td>
<td>Luminal</td>
</tr>
<tr>
<td>HC11</td>
<td>Basal</td>
</tr>
<tr>
<td></td>
<td>Adipocytes</td>
</tr>
<tr>
<td></td>
<td>Stromal</td>
</tr>
</tbody>
</table>
Figure 10: Expression of Gm13003 is upregulated when dCas9-VPR is guided to the active enhancer hub between Gm13003 and Wnt4. A) Illustration depicting the location of lncRNA Gm13003 in the Wnt4 locus. Pink arrows represent the location of CRSs and black arrows are negative controls n1, n2 and n3. The blue arrow is the location of the gRNAs for the Wnt4 promoter. B) Expression of Gm13003 in HC11 cells (Wnt4low), upon guiding dCas9-VPR to the CRSs (in the same samples earlier shown in figure 6 and 7). Samples were transfected with pools of gRNAs, specific for each CRS, and dCas9-VPR, which is a transcriptional activator. Samples were harvested 48 hours after transfection and Gm13003 expression was measured using qRT-PCR. Every symbol represents one biological experiment (HC11 n=4; BC44 n=3). The yellow crosses represent the mean of Wnt4 expression, of which the individual data points can be found in Figure 6. For comparison, the mean of Wnt4 expression is shown again in this figure. CRSs are ordered according to their genomic location on the x-axis. Ctr = control without gRNAs, set to 1 (grey line). N1-3 are negative control regions. P = Wnt4 promoter. C) Expression of Gm13003 in BC44 cells (Wnt4high). Except for the cell line, experimental details are the same as in B).
**CRSs are partially conserved between mouse and human**

Similar to the mouse mammary gland, the human breast is also characterized by a branched network of ductal epithelial cells. Multiple genes and pathways are conserved between the mouse and human tissue. In human breast tissue, WNT4 is also mainly expressed in the luminal epithelial cells. To investigate the potential functional relevance of the CRSs for regulating WNT4 expression in human cells, we checked their enhancer activity in luciferase assays in HMLE cells, which is a human mammary gland epithelial cell line (Figure 11A). Similar to the luciferase assays in mouse epithelial cells, we observed that the CRSs located between Gm13003 and Wnt4 (CRS-6 to CRS-11) are amongst the CRSs with the highest enhancer activity. The results of CRS-9 and n3 in HMLE cells also match the results in BC44 and HC11 cells. The main difference we observed is that in mouse epithelial cells CRS-7 shows the highest enhancer activity, while in HMLE cells CRS-8 shows the highest transcriptional enhancer, with an average log2 fold change of 3.9 in luminescence signal compared to the EV control.

The fact that the CRSs also show functional enhancer activity in HMLE cells, suggests that these sequences contain conserved binding sites for TFs that are also expressed in HMLE cells. To check the conservation of the CRSs between mouse and human, we aligned the mouse sequences to the human genome using BLAST. In Figure 11B the hits for CRS-6 to -11, the CRSs located in the active enhancer hub, are visualized. The query sequence is the CRS sequence that was used in our experiments (Table 1) and the first hit is, as expected, a perfect match with the mouse genome. The other hits shown in Figure 11B are from the human genome. Multiple hits are depicted, representing matches with different deposited sequences, including BAC clones and reference sequences. An overview of the alignments that were found between the query sequence (of all CRSs) and hits in the human genome can be found in Supplementary Table 2 with further details in Supplementary Table 3. Out of the sequences checked, CRS-10 is the only CRS in which no region is conserved between mouse and human. All other analyzed CRSs contain larger or smaller blocks that could be aligned to the human genome, with different alignment scores.

**Conserved CTCF binding motif between mouse and human in CRS-6**

Last, we want to highlight the binding of CTCF at Gm13003. In Figure 2A, we only focused on CTCF motif orientation for the CTCF peaks flanking of our region of interest, i.e. those most likely to correspond with the predicted TAD boundaries. However, CTCF ChIP-seq data from whole mammary gland shows two other interesting peaks, one in Gm13003 and one in Wnt4 (Figure 12A). The convergent orientation of these two motifs indicates possible loop formation between those sites. The CTCF ChIP-seq peak located in Gm13003, is contained within CRS-6 (Figure 12B). In addition, and perhaps more interestingly, it also overlaps with conservation hit 2a. The sequence of 2a was aligned to clone RP1-163O16, which contains the sequence of human chromosome 1, location 1p35.1-36.13 (Supplementary Table 2 and 3). This area contains the location of the human WNT4 gene (1p36.12). TRANSFAC analysis of this region revealed a
potential binding site for CTCF (Figure 12C). Thus, although Gm13003 itself is not conserved or annotated in the human genome, the CTCF binding site at CRS-6 is conserved.

![Figure 11: Conservation of CRSs.](image)

**A** Enhancer activity of CRSs in HMLE cells

**B** Graphical representation of alignments between mouse (taxid:10088) and human (taxid:9606) in BLAST of conserved regions in the CRSs. For every CRS, hit 1 is a validation from the mouse genome. Other hits are from the human genome. Alignment scores represent the overall quality of an alignment and are provided by BLAST. Higher scores represent a higher degree of similarity. In this figure, only hits with an E-value of <0.05 are shown. For more details on the hits, see supplementary table 2 and 3.
**Discussion**

The goal of this study was to identify regulatory sequences that control Wnt4 expression in the mammary gland. Previously published HiC\(^\text{15}\) and CTCF ChIP-seq data\(^\text{29}\), combined with the location and orientation of putative CTCF binding sites in the Wnt4 locus, indicate that a TAD is formed, in which Wnt4 is the only coding gene (Figure 2). Based on the peak in the CTCF ChIP-seq data\(^\text{29}\) in the Wnt4 gene and the orientation of the putative CTCF binding site, combined with the results from our 4C experiments (Figure 8), we hypothesize that a smaller regulatory loop is formed within this TAD, that brings CRS-6 and the Wnt4 promoter in close physical proximity. Regulatory loops have been described before as structures that are formed within TADs to facilitate communication between enhancers and promoters\(^\text{74}\). The formation of the regulatory loop in the Wnt4 TAD might contribute to tissue-specific Wnt4 expression, as it has been suggested that these regulatory loops depend on the binding of tissue-specific TFs\(^\text{73}\). The tissue-specific transcripts of Gm13003 could also play a role in stabilizing the regulatory loop, as recent studies have shown that CTCF-RNA interactions are required for genome organization\(^\text{88,89}\).

Within the Wnt4 TAD, we selected CRSs, based on datasets about active enhancers marks (Figure 4). In our dCas9-VPR assays, we discovered that transcriptional information can be
transferred from CRSs to the \textit{Wnt4} promoter (Figure 6). Furthermore, the sequences of CRS-6, -7, -8, -10 and -11 show functional enhancer activity on their own (Figure 5 and 11).

Based on the results presented in this chapter, we propose a model in which tissue-specific \textit{Wnt4} expression is regulated by functionally active enhancers that are located in a regulatory loop that is formed between \textit{Gm13003} and \textit{Wnt4} (Figure 13).

Yet, there are remaining questions related to this model. For example, does the binding of CTCF in the \textit{Wnt4} gene block transcription? If the formation of the regulatory loop relies on tissue-specific TFs, which TFs are involved in the formation of this specific loop? Why is there a clear difference in the dCas9-VPR assay between CRS-5 and CRS-7, while their distance to the \textit{Wnt4} TSS is not substantially different when \textit{Gm13003} is “pulled” to the \textit{Wnt4} promoter? Do the CTCF proteins that form the regulatory loop form a roadblock for chromatin-scanning TFs? Answers to these questions are needed to further develop the proposed model of regulation of \textit{Wnt4} expression.

Aligning the mouse CRSs to the human genome, revealed that small stretches of CRSs are conserved (Figure 11). One CRS that is partially conserved, but did not show transcriptional enhancer activity in the luciferase reporter assay, is CRS-9. This CRS has a high peak in the ATAC-seq datasets (which is associated with open chromatin), as well as in H3K27ac datasets (which is associated with active enhancers), especially in mature luminal cells (Figure 3). CRS-9 is also linked to \textit{Wnt4} in the dCas9-VPR assays (Figure 6). These results indicate that CRS-9 does have a regulatory role, and could possibly act as a repressor, similar to CRS-16.

Future experiments have to reveal how essential these regulatory sequences are for \textit{Wnt4} expression. By using other methods from the CRISPR toolbox, including regular CRISPR-Cas9 gene editing, the CRSs can be cut from the genome individually and subsequently \textit{Wnt4} expression can be measured. Different combinations of enhancer sequences should be removed, as enhancer redundancy could act as a safeguard to prevent aberrant gene expression after deletion of individual enhancers. It also has to be taken into account that removing large sections of DNA using CRISPR-Cas9 might alter the 3D genome structure, which can influence gene expression in and by itself.

Our results suggest that the annotated lncRNA \textit{Gm13003} might play a regulatory role in \textit{Wnt4} expression. However, the question remains which part of \textit{Gm13003} is involved in \textit{Wnt4} expression: the genomic sequence of \textit{Gm13003}, the actual lncRNA (i.e, the \textit{Gm13003} transcript) or both. Our luciferase (Figure 5 and 11) and dCas9 assays (Figure 6) show that CRS-6, which extends upstream of the \textit{Gm13003} TSS and comprises only part of the \textit{Gm13003} transcript region, has functional enhancer activity and can be linked to \textit{Wnt4}. Furthermore, our 4C experiments show that CRS-6 is in close physical proximity to the \textit{Wnt4} promoter (Figure 8). The CTCF binding motifs in CRS-6 and the \textit{Wnt4} gene are conserved between mouse and human (Figure 12), which lends further support to the formation of a CTCF-dependent chromatin loop in human cells.
Functional dissection of the Wnt4 locus

Figure 13: Proposed model of regulation of Wnt4 expression in mammary gland cells. The insulator sequences of the Wnt4 TAD are located between Zbtb40 and Cdc42. Both TAD boundaries contain two CTCF binding sites. Within the Wnt4 TAD, a regulatory loop is formed by CTCF that can bind to Gm13003 and Wnt4. Formation of this loop causes physical interaction between the upstream region of Gm13003, which contains CRS-6, and the Wnt4 promoter. The regulatory loop also facilitates the transfer of transcriptional information from CRS-6 to -11, which are all located between Gm13003 and Wnt4. Note that we do not have experimental evidence that cohesin plays a role in the formation of these specific loops, but based on its role described in literature\textsuperscript{76–79}, the cohesin complex is included in our model.
The lncRNA itself could be regulating Wnt4 expression as well. Various studies have pointed out the role of lncRNAs in gene regulation and organization of the 3D genome\textsuperscript{81–87}. One indication that Gm13003 could play a role in Wnt4 expression is that the expression of Gm13003 and Wnt4 correlates in mammary cell lines and primary mammary cell types (Figure 9). Furthermore, CRSs in the identified enhancer hub were not only linked to Wnt4 expression, but also to Gm13003 (Figure 10). Recently, studies have shown that CTCF requires RNA binding to function\textsuperscript{88,89}. As mentioned above, this might be part of the mechanism behind cell type specific Wnt4 expression in the mammary gland, as in the mammary gland Gm13003 is only expressed in luminal cells (Figure 9). One way to determine whether the transcript of Gm13003 is involved in Wnt4 expression, is to knock-down Gm13003. This can be achieved by using methods such as siRNA or CRISPR-Cas13, which specifically targets RNA\textsuperscript{90,91}, including lncRNAs\textsuperscript{92}. Furthermore, to unravel the role of the 3D genome in Wnt4 expression in more detail, it will be interesting to manipulate the CTCF binding sites at this locus, especially the conserved CTCF site at CRS-6/Gm13003 (Figure 12). Using CRISPR-Cas9, CTCF sites can be removed or inverted. These experiments will tell us more about the importance of the genomic structure on Wnt4 expression.

In conclusion, we performed a dissection of the Wnt4 TAD, spanning ~250 kb, searching for regulatory elements that control Wnt4 expression. This resulted in the identification of several active enhancers that are located upstream of Wnt4. Based on our combined experimental findings, we propose a model in which Wnt4 expression in the mammary gland is controlled by the formation of a smaller regulatory loop within the larger Wnt4 TAD. Although the Wnt4 TAD is exceptional in this respect because it harbors only a single coding gene, our workflow can be used for other loci to investigate cell type specific gene regulation.

**Materials and methods**

**Cell culture**

Mammary epithelial BC44 cells (a kind gift of Marie-Anne Deugnier, Institute Curie, Paris, France) were cultured in RPMI1640 + L-Glut (Gibco, cat. #), supplemented with 10% FBS (Gibco, cat# 10270-106) and 5 mg/ml insulin (Sigma-Aldrich, cat. #I9278). HC11 mammary epithelial cells were also cultured in RPMI1640 + L-Glut with 10% FBS and 5 mg/ml insulin, with the addition of 10 ng/ml EGF (Peprotech, cat. #AF-100-15-B). The human mammary epithelial cell line HMLE (a kind gift of Christina Scheel, Helmholtz Center Munich, Germany) was cultured in DMEM/F12 (Gibco), supplemented with 20 ng/ml EGF, 0.5 mg/ml Hydrocortisone (VWR, cat. #SAFSH4001) and 10 mg/ml insulin.

**RNA-seq cell lines**

Total RNA was isolated from BC44 and HC11 cells using a combination of TRIzol reagent (Fisher Scientific, #15608948) and RNeasy mini kit (Qiagen, #74104), including on-column DNase
digestion (Qiagen, RNase-free DNase set #79254). Briefly, cells were lysed by directly adding 1 ml TRizol to 10 cm culture dishes. The homogenate was transferred to Phasemaker tubes and mixed with chloroform. After separation, the aqueous RNA layer was mixed with 1 volume of 70% ethanol and transferred to RNeasy spin columns. RNA purification and DNase digestion were performed according to manufacturer’s instructions.

All library preparation and sequencing steps were performed by MAD: Dutch Genomics Service & Support Provider, Swammerdam Institute for Life Sciences, University of Amsterdam. RNA integrity was analyzed with an Agilent RNA ScreenTape system (RIN >= 9.6). ERCC RNA Spike-In Mix 1 (ThermoFisher Scientific #4456740) was added to samples prior to polyA enrichment (NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs) and stranded- library preparation (NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (New England BioLabs). Assessment of the size distribution of the libraries with indexed adapters was performed using a 2200 TapeStation System with Agilent D1000 ScreenTapes (Agilent Technologies). The NEBNext Library Quant Kit for Illumina (New England BioLabs) was used according to manufacturer’s instructions to quantify the libraries on a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific). Libraries were clustered and sequenced on a Illumina NextSeq 550 Sequencing System (NextSeq 500/550 Mid Output Kit (300 Cycles), 2 x 150 bp). The snakemake workflow VIPER (Visualization Pipeline for RNA-seq analysis)\(^{93}\) was used to map and align raw sequencing data to the mouse genome (mm9, STAR\(^{94}\) version 2.7.1a). Raw transcript counts were further processed using edgeR\(^{95,96}\) (version 3.28.0) and limma\(^{97}\) (version 3.42.1) packages with R version 3.6.2\(^{98}\). A cutoff of CPM > 0.5 in at least 2 libraries was applied to filter out genes with low counts prior to trimmed mean of M-values (TMM) normalization.

**RNA in situ**

The RNAscope 2.5 HD Duplex Assay was used for RNA in situ. Freshly isolated mammary glands were dissected and fixed for 24 hours in 4% PFA. Samples were dehydrated through ascending grades of ethanol, cleared in Histo-Clear II (National Diagnostics cat. #HS-200) and embedded in paraffin. Sections of 5 μm were cut using a microtome, after which the Formalin-Fixed Paraffin-Embedded (FFPE) Sample Preparation and Pretreatment protocol was followed (Document Number 322452). One adjustment of the pretreatment was the 30x dilution of Protease Plus with PBS. We optimized this protocol for mammary gland sections and did the Target Retrieval Treatment for 15 minutes and Protease Plus Treatment for 30 minutes. Second part of the RNAscope experiments was done according to RNAscope. 2.5 HD Duplex Detection Kit (Chromogenic) User Manual (Document Number 322500-USM). Probes that were used are: Channel 1: *Krt14* (Cat No. 422521) and Channel 2: *Wnt4* (Cat No. 401101-C2). Bright field images were taken on a Zeiss LSM 510 META microscope using a 20x objective in combination with the Axiocam HRc.
The 3rd and 4th mammary gland fat pads from 6 puberty (P35) FVB/N mice were dissected, chopped to small pieces and incubated for 2 hours at 37°C in an orbital shaker in a digestion mix composed of DMEM/F12, 5% FBS, 1% Penicillin / Streptomycin, 25 mM HEPES and 300 U/ml Collagenase IV. Red blood cells were lysed with ACK Lysing Buffer and single cell suspensions were generated by consecutive digestion steps with Trypsin-EDTA and DNAseI. Cells were stained in HBSS containing 10% FBS with the following antibodies (eBioscience): anti-Mouse CD45-Biotin (clone 30-F11), anti-Mouse CD31-Biotin (clone 390), anti-Mouse TER-119-Biotin (clone Ter-119), anti-Mouse CD326-PE (clone G8.8), anti-Mouse CD49f-FITC (clone GoH3) and Streptavidin-APC. DAPI was used for live/dead cell discrimination. 55,000 luminal, basal and stromal cells were sorted into ice-cold 10 mM HEPES containing 10% FBS using a BD FACSria III equipped with a 100 μm nozzle at 20 psi. FITC was excited with a 488 nm laser and emission was filtered using a 530/30 nm bandpass filter. PE was measured using a 561 nm laser and 582/15 nm bandpass filter. DAPI was excited with a 407 nm laser and emission was filtered using a 450/50 nm bandpass filter. APC was measured using a 633 nm laser and 660/20 nm bandpass filter. Cells were sorted with a plate voltage of 2500 V using the 4-Way Purity precision mode. ATAC-seq samples were prepared according to the protocol from Buenrostro et al.99,100. Briefly, freshly sorted cells were washed once in 50 ml ice-cold PBS and gently resuspended in 50 ml cold lysis buffer (10mM Tris-HCl, pH 7,4; 10mM NaCl; 3mM MgCl2 and 0.1% NP-40) to isolate nuclei. The transposase reaction (Illumina Nextera DNA Library Preparation Kit, FC-121-1030) was performed for 30 minutes at 30°C and DNA was purified using MinElute PCR purification columns (Qiagen, #28004). DNA fragments were amplified using NEBNext Ultra II Q5 Master Mix (#M05445) and custom primers: Luminal forward: AATGATACGCGACACCGATCTACACTACGTCCAGCAGTCGATGATGT*G Luminal reverse: CAAGCAGAAGACGGCATACGAGATGTAGAGGGTCTCGTGGGCTCGGAGATG*T Basal forward: AATGATACGCGACACCGATCTACACTACGTCCAGCAGTCGATGATGT*G Basal reverse: CAAGCAGAAGACGGCATACGAGATGTAGAGGGTCTCGTGGGCTCGGAGATG*T.

The following PCR conditions were used: 72°C for 5 minutes; 98°C for 30 seconds, 8 cycles of 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 1 minute. Libraries were purified with the MinElute PCR purification kit. An additional, single left-sided purification with AMPure XP beads (Beckman Coulter A63880) was performed to remove primer dimers. Libraries were quantified with qRT-PCR before paired-end sequencing on an Illumina NextSeq 550 system (2 x 75 bp, performed by MAD: Dutch Genomics Service & Support Provider, Swammerdam Institute for Life Sciences, University of Amsterdam). Sequencing data were processed using the ATAC-Seq pipeline of the Kundaje lab (https://github.com/kundajelab/atac_dnase_pipelines) with default parameters and mouse genome version Mm9.
FACS primary mammary cells

Freshly isolated mammary glands from FVB/N mice were minced and enzymatically digested for 2 hours in 9.2 ml DMEM/F12, 5% FCS, 1%P/S, 25 mM HEPES (Gibco, cat. #15630056) and 300 U/ml Collagenase IV (Gibco, cat. #17104019) (10 ml per mouse and 4 glands). Adipocytes were transferred to TRIzol LS (Invitrogen, cat. #10296028) and stored at -80°C. After resuspension in a mix (1:3) of HBSS (Gibco) with 2% FBS (Gibco, cat# 10270-106) and ACK solution (Gibco, cat. #A1049201), cells were incubated at RT for 5 min. HBSS (13 ml) was added and cells were spun down for 5 min, 1000 rpm, 4°C. Pellets were resuspended in 2 ml pre-warmed 0.05% Trypsin-EDTA (Gibco) and incubated for 5 minutes at 37°C. Serum-free DMEM (3 ml, pre-warmed) and 1mg/ml DNaseI was added and after mixing well, 8 ml of DMEM with 10% FCS was added to stop trypsinization. Before antibody staining, cells were filtered through a 40 mm mesh. The following antibodies were used: EpCAM-PE (1:100, eBioscience, 12-5791-82, clone G8.8), CD49f-FITC (1:100, eBioscience, 11-0495-82, clone GoH3), CD45-Bio (1:100, eBioscience, 13-0451-82, clone 30-F11), CD31-Bio (1:100, eBioscience, 13-0311-81, clone 390), Ter119-Bio (1:100, eBioscience, 13-5921-81, clone TER-119) and Streptavidin-APC (1:200, eBioscience, 17-4317-82). During antibody staining, cells were in 200 ml HBSS supplemented with 10% HF and kept in the dark on ice for 20 minutes. Before sorting, cells were stained with DAPI and with DAPI (1:5000) and filtered through a 50 mm mesh. Cell sorting was performed using a BD FACS Aria III. FITC was excited with a 488 nm laser and emission was filtered using a 530/30 nm bandpass filter. PE was measured using a 561 nm laser and 582/15 nm bandpass filter. DAPI was excited with a 407 nm laser and emission was filtered using a 450/50 nm bandpass filter. APC was measured using a 633 nm laser and 660/20 nm bandpass filter. Cells were sorted with a plate voltage of 2500 V using the 4-Way Purity precision mode. Cells were collected in TRIzol LS and stored at -80°C.

Cloning pGL4-minP constructs

We cloned a minimal promoter (identical to the minimal promoter of pGL4.23 (Promega): TAGAGGGTATATAATGGAAGCTCGACTTCCAG) at the HindIII site of pGL4.20, which contains a firefly luciferase gene (luc2) and a puromycin resistance gene. This pGL4.20 plasmid with the minimal promoter is referred to as pGL4-minP. The pGL4-minP vector was digested with XhoI and BglII and 1X Tango buffer (Thermo Scientific). Sequences of CRSs (that belong to the coordinates shown in Table 1) were ordered as double stranded gBlocks from IDT. Restriction sites were added to the CRSs (5’: XhoI or Sall, 3’: BglII or BamHI). The default combination was XhoI/BglII, but when the CRS contained one of those restriction sites, Sall and/or BamHI were chosen. Tubes from IDT were spun down to ensure that the DNA was at the bottom of the tube. DNA was diluted with 0.2 mm filtered water to get a concentration of 10 ng/ml, after which the tubes were incubated at 50°C for 20 minutes. gBlocks were digested with the XhoI/BglII, Xho/ BamHI, Sall/BglII, or Sall/BamHI with the appropriate buffer (Thermo Scientific). After digestion, gBlocks were purified using the GeneJET PCR Purification kit (Thermo Scientific, cat. #K0701). pGL4-minP vector and CRS inserts were ligated using T4 ligase (Thermo Fisher Scientific, cat.
Chapter 4

#EL0016) and T4 DNA ligase buffer in a ratio of 1:3 (vector:insert) O/N at 16°C. Transformation was done using DH5α bacteria and a 1 minute heat shock in a 42°C water bath. Mixtures were incubated in a shaker at 37°C for 30 minutes at 500 rpm after adding 250 ml LB medium, and plated on LB agar plates with ampicillin. Minipreps were performed using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, #K0502) and the correct inserts were checked with digestion checks. In addition, plasmids were sequence verified using the following primers: forward (CTAGCAAAATAGGCTGTCCC), reverse (CTTCTTAATGGCTGTCCC).

### Cloning pSpgRNA constructs

For the dCas9 assays, gRNAs were cloned into pSpgRNA (Addgene, #47108) at the BbsI site. For designed gRNAs, CRISPOR (http://crispor.tefor.net) was used. Cloning was done according to the Zhang lab protocol and minipreps were done using the the GeneJET Plasmid Miniprep Kit. Correct inserts sequence verified using the following primer: GAGGGCCTATTTCCCATGATT. All gRNAs that were used in dCas9 assays can be found in Supplementary Table 1.

### Dual luciferase assays

BC44, HC11 (both 20.000 cells per well) or HMLE cells (80.000 cells per well) were plated in a 24 well plate 24 hours prior to transfection. X-tremeGENE HP DNA Transfection Reagent (Sigma-Aldrich, cat. #636626001) was used for transfection according to manufacturer's protocol (1:1 ratio for BC44 and HC11, 1:3 ratio for HMLE). Transfections were done in triplo, using 300 ng pGL4_minP (EV or containing the CRS sequence), 100 ng CMV-Renilla (transfection control), and 100 ng eGFP (to check transfection efficiency) in each well. Cells were harvested 48 hours after transfection and plates were stored in -80°C. For the luciferase assay, cells were lysed in 1X Passive Lysis Buffer (50 ml per well (Promega, cat# E1941)) according to manufacturer’s instructions. For the reactions, non-commercial firefly and Renilla luciferase reagents (LAR) were used. Firefly LAR is composed of 200 mM Tris-HCl (pH 8.0), 15 mM MgSO₄, 0.1 mM EDTA (pH 8.0), 25 mM DTT (Fisher cat. 10792782), 1 mM ATP pH 7.0 (Sigma cat. A2383), 0.2 mM Coenzyme A (Sigma cat. C3144), 200 mM D-Luciferin pH 6.0-7.0 (Biosynth cat. L-8200) with a final pH of 8.0. The buffer of Renilla LAR contains 25 mM Na₂P₂O₇₅₆(Carl Roth cat. T883.1), 10 mM NaAc, 15 mM EDTA, 500 mM Na₂SO₄, 500 mM NaCl with a final pH of 5.0. The following components were freshly added to the Renilla LAR buffer before every assay: 50 mM phenylbenzothiazole (Santa Cruz Biotechnology cat. sc-391075) and 4 mM benzyl-coelenterazine (Nanolight cat. 301-500). Firefly and Renilla luciferase activity was measured in a GloMax Navigator (Promega, cat# GM2000). The following protocol was used in the GloMax: Injection of 50 ml non-commercial firefly LAR, 2 seconds pre-measurement delay, 10 seconds measurement firefly reporter, injection of 50 ml non-commercial Renilla LAR, 2 seconds pre-measurement delay, 10 seconds measurement Renilla reporter. Firefly luciferase values were normalized to their own Renilla luciferase values.
dCas9 assays

BC44 and HC11 cells were plated with a density of 100,000 cells per well in a 6 well plate. After 24 hours, cells were transfected with prepared pools of pSpgRNA (500 ng per well, containing the gRNAs, pools were made per promoter/CRS) and dCas9-VPR (1500 ng, Addgene #63798) using X-tremeGENE HP DNA Transfection Reagent in a 1:1 ratio. Cells were harvested after 48 hours and stored in -80°C until further analysis.

qRT-PCR

For BC44 and HC11 WT cell lines, dCas9 samples, and FACS sorted primary mammary cells, RNA was isolated from Trizol according to the manufacturer’s protocol. cDNA synthesis was performed from 200 ng - 4 mg of RNA using SuperScript IV Reverse Transcriptase (Invitrogen, cat. #18090200) and Random Hexamers (Invitrogen, cat. #N8080127) according to manufacturer’s guidelines with the addition of Ribolock RNase Inhibitor (Thermo Scientific, cat. #EO0328). After cDNA synthesis was completed, samples were diluted 10x. qRT-PCR reactions were performed using a QuantStudio 3 Real-Time PCR System. For the reactions, 5 ml of diluted cDNA was added to a mix of 4 ml 5X HOT FIREPol EvaGreen qPCR Mix Plus (ROX) (Solis Biodyne, cat. #08-24-00008), 2 ml primers (from a 10 mM stock) and 10 ml nuclease-free water. Reactions were performed in triplicates in a 96x0.2 ml plate (BIOplastics, cat. #AB17500). The following stages were included in the reaction: 2 minutes at 50.0°C and 15 minutes at 95.0°C, then 40 cycles of 15 seconds at 95.0°C and 1 minute at 60.0°C, followed by the melting curve stage. See Table 3 for used primers.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward / Reverse</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cebp1</td>
<td>Forward</td>
<td>GTGCCCTGATGTACCATACTACCA</td>
</tr>
<tr>
<td>Cebp1</td>
<td>Reverse</td>
<td>GCCAATTCGCCAGATGTATCTCTCA</td>
</tr>
<tr>
<td>Wnt4</td>
<td>Forward</td>
<td>ACTGCACTCCCTCTGTCT</td>
</tr>
<tr>
<td>Wnt4</td>
<td>Reverse</td>
<td>TGGCCCTTGCACCTGAAAA</td>
</tr>
<tr>
<td>Zbtb40</td>
<td>Forward</td>
<td>GTATCTACCGGATGCAAAG</td>
</tr>
<tr>
<td>Zbtb40</td>
<td>Reverse</td>
<td>CCTGTTGAGCGGGGTCT</td>
</tr>
<tr>
<td>Cdc42</td>
<td>Forward</td>
<td>ACAACAAACAAATCCCATCG</td>
</tr>
<tr>
<td>Cdc42</td>
<td>Reverse</td>
<td>TGGCCTGAGATCTACAAAA</td>
</tr>
<tr>
<td>Gmi13003</td>
<td>Forward</td>
<td>CAGGACACAGCTGCAA</td>
</tr>
<tr>
<td>Gmi13003</td>
<td>Reverse</td>
<td>CGGCCTGATCTTCTGTCT</td>
</tr>
<tr>
<td>Krt18</td>
<td>Forward</td>
<td>AGATGACACCAATCACTACAAA</td>
</tr>
<tr>
<td>Krt18</td>
<td>Reverse</td>
<td>CTGACGACGCTTGTGACTCTG</td>
</tr>
<tr>
<td>Krt14</td>
<td>Forward</td>
<td>ATCGAGGACCTGAGAAGC</td>
</tr>
<tr>
<td>Krt14</td>
<td>Reverse</td>
<td>TCGATCTACAGGAGGACATT</td>
</tr>
<tr>
<td>Vimentin</td>
<td>Forward</td>
<td>CAGCCTACGAGAAATTCG</td>
</tr>
<tr>
<td>Vimentin</td>
<td>Reverse</td>
<td>CCATTCCTGAGGAGGTCAAG</td>
</tr>
<tr>
<td>Cfd</td>
<td>Forward</td>
<td>CATGCTGCGCTACTGEE</td>
</tr>
<tr>
<td>Cfd</td>
<td>Reverse</td>
<td>CACAGAGTGTGCATCAGTC</td>
</tr>
</tbody>
</table>

Table 3: Primers used for qRT-PCR reactions.
4C-seq was performed as previously described\(^{106}\), with some modifications. Briefly, 1x10^6 BC44 or HC11 cells were crosslinked with 2% (v/v) formaldehyde in PBS/10% FBS for 10 minutes at RT in 15 cm cell culture dishes while shaking. Cells were isolated by using a cell scraper, lysed for 10 minutes in lysis buffer and homogenized with a dounce-homogenizer. Nuclei were digested with 200 U DpnII (NEB) in DpnII buffer (NEB) complemented with 0.3% SDS and 2% Triton X-100 for 4 hours, an additional 200 U DpnII O/N and another 200 U DpnII for 4 hours the following day at 37°C while shaking. The first ligation was performed with 100 U T4 DNA ligase O/N at 16°C in a total volume of 15 ml. Following decrosslinking and phenol/chloroform extraction, the DNA was digested NlaIII in Cutsmart buffer (NEB) O/N at 37°C while shaking. After the digestion a second ligation was carried out with T4 DNA ligase (Thermo Scientific) in a total volume of 15 ml O/N at 16°C. The DNA was extracted with a phenol/chloroform extraction and purified with the ChIP DNA Clean & Concentrator Kit (Zymo Research). 50-100 ng template per reaction and 1mg in total was used to amplify PCR-specific libraries for the viewpoint using primers listed in Table 4 and the Expand Long Template PCR system (Sigma). Successful reactions were pooled and purified with the High-pure PCR Product Purification Kit (Sigma). Equal amounts of libraries were combined and all library preparation and sequencing steps were performed by MAD: Dutch Genomics Service & Support Provider, Swammerdam Institute for Life Sciences, University of Amsterdam. For processing raw 4C-seq data, the publicly available 4C mapping pipeline for mapping (Mm9) and filtering 4C data was used (https://github.com/deWitLab/4C_mapping) using default settings. The R package PeakC\(^{107}\) was used to analyze processed triplicates of 4C-seq data and identify significantly interacting regions within a window size of 400kb.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewpoint reading primer</td>
<td>CCAGAACTCAGGGCGATC</td>
</tr>
<tr>
<td>Viewpoint non-reading primer</td>
<td>TGGGTGTCTGTTCGATTCTT</td>
</tr>
<tr>
<td>Sequencing primer BC44 Wnt4 1</td>
<td>AATGATACGGCGACCCAGGATCTACACTCTTTCCCCTACACGACGC TCTTCCGATCTGTAGCCCCAGAACTCAGGGCGATC</td>
</tr>
<tr>
<td>Sequencing primer BC44 Wnt4 2</td>
<td>AATGATACGGCGACCCAGGATCTACACTCTTTCCCCTACACGACGC TCTTCCGATCTACATAGCCAGAACTCAGGGCGATC</td>
</tr>
<tr>
<td>Sequencing primer HC11 Wnt4 1</td>
<td>AATGATACGGCGACCCAGGATCTACACTCTTTCCCCTACACGACGC TCTTCCGATCTACATAGCCAGAACTCAGGGCGATC</td>
</tr>
<tr>
<td>Sequencing primer HC11 Wnt4 2</td>
<td>AATGATACGGCGACCCAGGATCTACACTCTTTCCCCTACACGACGC TCTTCCGATCTACATAGCCAGAACTCAGGGCGATC</td>
</tr>
</tbody>
</table>

Table 4: Primers used in 4C experiments.

Acknowledgements

We would like to thank Selina van Leeuwen and other members of the MAD: Dutch Genomics Service & Support Provider, Swammerdam Institute for Life Sciences, University of Amsterdam for their sequencing service. We also would like to thank our colleagues at the section of Molecular Cytology at the Swammerdam Institute for Life Sciences for discussions and feedback during
the project. This research was funded by the Netherlands Organization for Scientific Research (NWO ALW VIDI 864.13.002. to R.v.A.) and the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement #706443 (K.E.W.).

Author contributions
K.E.W., R.v.A., and N.H. conceived the study. K.E.W. and N.H. designed and performed most of the experiments. N.H. performed the TAD analysis, RNAscope, luciferase assays, and conservation analysis. K.E.W. performed the ATAC-sequencing, dCas9 assays, FACS and qRT-PCR. Y.v.d.G. performed the 4C experiments. R.v.A. supervised the study. All authors analyzed and interpreted the data. N.H. wrote the manuscript with input from all authors.
Supplementary Figure 1: RNA in situ negative control images. Numbers 1 and 2 refer to the different mice and correspond to the same numbers in Figure 3A. For the negative controls, RNAscope 2-plex Negative Control Probe Mix was used, which contains premix probes for DapB in both channels. Nuclei were counterstained with hematoxylin. Scale bars are 20 μm.
Supplementary Figure 2: Wnt4 expression in BC44 and HC11 cell lines. Graph showing RNA-sequencing results of BC44 and HC11 WT samples. RPKM = reads per kilobase per million mapped reads. Two replicates per cell line were sequenced.

Supplementary Figure 3: Gene expression in sorted primary mammary cell types. Graph showing expression of cell type specific markers of different mammary gland cell types. Mammary gland cells from 6 adult C57BL/6J mice were sorted using FACS. In qRT-PCR experiments, Ctbp1 was used as housekeeping gene and values were normalized to the sample with the highest expression per gene. Krt18 is a marker for luminal cells, Krt14 for basal cells, Vimentin for fibroblasts and Cfd (Adipsin) for adipocytes\textsuperscript{108,109}. 

108,109
<table>
<thead>
<tr>
<th>CRS/P/NC</th>
<th>Size (bp)</th>
<th>#</th>
<th>Strand</th>
<th>Forward oligo</th>
<th>Reverse oligo</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>600</td>
<td></td>
<td>-</td>
<td>caccgGCCGAGGATGGGTGTTACT</td>
<td>aaacAGGTAACCCCCATACTCTGCCGcc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgACCGCGGATCCCCGCTGTTG</td>
<td>aaacCAGAGCCGGAAGTCGGGGCAGCt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGCCGTCGAGCATGCCGCTCCGCC</td>
<td>aaacGCCGACCCAGCCTGAGCAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGCGCTACCCCGAGGCGAGC</td>
<td>aaacGCTGCTCGGAGGCTGAGGcc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgAGCGGACCACTCGGGCGTAAT</td>
<td>aaacATTGTGCTGGGCGGCGGACAg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgACCGGGGAGGAGCGGACGAC</td>
<td>aaacAGGAGTACCGCCCGGGGAGCc</td>
</tr>
<tr>
<td>1</td>
<td>1185</td>
<td></td>
<td>+</td>
<td>caccgCTAGGGAGGTGAGTTGAC</td>
<td>aaacCATGAATCTACCTGCTACTGcc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCACCGAGAGACAGCTGATG</td>
<td>aaacCATGACTACTGAGAGGTTAc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCACCAGCACAGAGCTCATTG</td>
<td>aaacCAATGAGCTCTGTGCTGGTc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgTTCTGCAGCCAACTACGGTC</td>
<td>aaacGACCGTAGTTGGCTGCAGAAc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGGTTCCTCAGTCCATCGTCC</td>
<td>aaacGGACGATGGACTGAGGACCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCGAGAGCTCCCCTGGTCGGC</td>
<td>aaacGCCGACCAGGGGAGCTCTCgc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGATTTGCATATGCAATGAGG</td>
<td>aaacCCTCATTGCATATGGCAAATc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGCATCGCCACCATGCCTTGG</td>
<td>aaacCCAAGGCATGGTGATTGCAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgTGTAGGTAACCAGCCACACC</td>
<td>aaacGGTGTGGCTGATCTACCTAc</td>
</tr>
<tr>
<td>2</td>
<td>1217</td>
<td></td>
<td>+</td>
<td>caccgCACCAGGAGGCTGAGTAGAC</td>
<td>aaacGTCACTTGGTGACAGGATTAc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCCCCAACCTGGGGAGTACCTT</td>
<td>aaacAAAGGATCCCCAGGTTGGGc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCATTCTCAAGATGCTGCCCT</td>
<td>aaacGGAGGACCATTTGATTAAGTtc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGAAGTTGCTGAGATTGCTGCA</td>
<td>aaacGTCGCAAAGTTCTACAGACTCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGCTCACACCCCGCTGTCGGC</td>
<td>aaacGCCAGACCGGGGTGTTAGGc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCTGGAGACACTCAGTCAAAC</td>
<td>aaacTGGTTTGACTGAGTGTCCAGc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGGCACAGCACCCACGTGCGT</td>
<td>aaacACGCACGTGGGTGCTGTGCCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCTACGTCGCTGTCGTCCAC</td>
<td>aaacTCACAGTGGGTTGACCGAGc</td>
</tr>
<tr>
<td>3</td>
<td>762</td>
<td></td>
<td>+</td>
<td>caccgCTTCTGTCGACCCCGCCTTG</td>
<td>aaacCAGAGGCAAGGGGAGGAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgTTGTCACCCCTGCCGTGTG</td>
<td>aaacCACACGGCAGGGGTGACACc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCATCATGTGACAGCGCCCTG</td>
<td>aaacCAGTTCTACAGCTACTTCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgATCAGACGTTGGTTGAGCTTA</td>
<td>aaacGGGAGCAGAGGGGAGTTTc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgCCGAGGACGCCCTGACGAT</td>
<td>aaacCTTCTACAGCTACTTCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgCTGAACTCCGGAGTACGGAG</td>
<td>aaacCTCCGTACTCCGGAGTTCAGc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgCTCGCCGACAGCCTACAGAG</td>
<td>aaacCTCTGTAGGCTGTCGGCGAGc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAAAGGACTTACCTGAGTAC</td>
<td>aaacGTCGAGCAGAGGGAGGAGCc</td>
</tr>
<tr>
<td>4</td>
<td>1014</td>
<td></td>
<td>+</td>
<td>caccgCTATGTTGACAAAGGCTGTTG</td>
<td>aaacCAAGCTGATGCTGAGTTTc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgATCAGACTGCTGCTGAGCCTT</td>
<td>aaacAAGGACAGGCTGTTGAGATc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgCATGAGAGCTGCTGCTGTA</td>
<td>aaacATTGCTGCTGCTGCTGATc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAGAGGAGGCTGAGGCTG</td>
<td>aaacTGAGACGCTAGCTACTTCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGACGAGACCGCTGACGAT</td>
<td>aaacATGAGATGAGATTGCTGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAGCTTCAAGGCTGTGAGAT</td>
<td>aaacGACGAGACTGAGTTTCTc</td>
</tr>
<tr>
<td>5</td>
<td>726</td>
<td></td>
<td>+</td>
<td>caccgCTATGTTGACAAAGGCTGTTG</td>
<td>aaacCAAGCTGATGCTGAGTTTc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgATCAGACTGCTGCTGAGCCTT</td>
<td>aaacAAGGACAGGCTGTTGAGATc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgCATGAGAGCTGCTGCTGTA</td>
<td>aaacATTGCTGCTGCTGCTGATc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAGAGGAGGCTGAGGCTG</td>
<td>aaacTGAGACGCTAGCTACTTCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGACGAGACCGCTGACGAT</td>
<td>aaacATGAGATGAGATTGCTGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAGCTTCAAGGCTGTGAGAT</td>
<td>aaacGACGAGACTGAGTTTCTc</td>
</tr>
<tr>
<td>6</td>
<td>1216</td>
<td></td>
<td>-</td>
<td>caccgAAAGGACTTACCTGAGTAC</td>
<td>aaacGTCGAGCAGAGGGAGGAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgCATGAGAGCTGCTGCTGTA</td>
<td>aaacATTGCTGCTGCTGCTGATc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAGAGGAGGCTGAGGCTG</td>
<td>aaacTGAGACGCTAGCTACTTCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGACGAGACCGCTGACGAT</td>
<td>aaacATGAGATGAGATTGCTGTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGAGCTTCAAGGCTGTGAGAT</td>
<td>aaacGACGAGACTGAGTTTCTc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgTTTACTACATCGGATGTAAC</td>
<td>aaacGTAAACCGAGTATGAAATAc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>caccgGCCATCCCCCTAGGTTAAAC</td>
<td>aaacGTATACCTAGGGAGGATCCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>caccgGGGTGCGCTAGTGATGTC</td>
<td>aaacGACGATTACCTCTAGCAGCCc</td>
</tr>
</tbody>
</table>
### Functional dissection of the Wnt4 locus

<table>
<thead>
<tr>
<th>CRS/P/NC</th>
<th>Size (bp)</th>
<th>#</th>
<th>Strand</th>
<th>Forward oligo</th>
<th>Reverse oligo</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1217</td>
<td>1</td>
<td>+</td>
<td>caccgTGGCTAGAGCTTGACGCCATCTG</td>
<td>aaacAGATGCCTCAGCTCTGGAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgCCTCTGGATGCTGCGTCAAAC</td>
<td>aaacGTTATACGCCGCTACGAGCAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>+</td>
<td>caccgCCTAGGGAAAATCTCAACGGCTGCTC</td>
<td>aaacGGTGGCTACTGCTTCCCTGACCAc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgTCCACCCTTACATCTGGG</td>
<td>aaacCCGAGTATGAGGGGTTGGAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgCTCTTTGACATACCTAACGGCTCGGCC</td>
<td>aaacGGCGCTATGAGGGGTTGGAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgCTCTTTGACATACCTAACGGCTCGGCC</td>
<td>aaacGGCGCTATGAGGGGTTGGAGCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgTTTACAGGGCTGCAACAGC</td>
<td>aaacGTTGTCGAGGCGCTTGGAGGc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-</td>
<td>caccgCTTACAGGGCTGCAACAGC</td>
<td>aaacGTTGTCGAGGCGCTTGGAGGc</td>
</tr>
<tr>
<td>8</td>
<td>1217</td>
<td>1</td>
<td>+</td>
<td>caccgGCACTGAGGCGTTACTTATCTG</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgGTAGTGGCTAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>+</td>
<td>caccgGGACTGAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>+</td>
<td>caccgGACTGAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>+</td>
<td>caccgGACTGAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>+</td>
<td>caccgGACTGAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgGACTGAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-</td>
<td>caccgGACTGAGGCGCTACGAGC</td>
<td>aaacACTAAATACGGCCCTTACGTGCC</td>
</tr>
<tr>
<td>9</td>
<td>1143</td>
<td>1</td>
<td>+</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>+</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>+</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>+</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>+</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-</td>
<td>caccgATCCCGTCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td>10</td>
<td>664</td>
<td>1</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td>11</td>
<td>705</td>
<td>1</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td>12</td>
<td>740</td>
<td>1</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td>13</td>
<td>1164</td>
<td>1</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>+</td>
<td>caccgGATCCGCTGATTTGCAAGC</td>
<td>aaacGTTGGAACATCCACGGGTATAC</td>
</tr>
<tr>
<td>CRS/P/NC</td>
<td>Size (bp)</td>
<td>#</td>
<td>Strand</td>
<td>Forward oligo</td>
<td>Reverse oligo</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>---</td>
<td>--------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>14</td>
<td>941</td>
<td>1</td>
<td>-</td>
<td>caccgTGGAAGCACACTAGTAGGTCG</td>
<td>aacacGACCTACAGTTGCTCCACac</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgTCAGTAGTACAGGAACAGCCCGGCGC</td>
<td>aacacGCGGGCTGGCGTAACTGACaG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGCCACCTAGGCTGGCGGCGGAGCAAC</td>
<td>aacacAGCCCAACGCTAGCTAGGACCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgAATCTCTAGGCTGACAGTAC</td>
<td>aacacGAGTCCGAATTCCACAGGAGTTe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgAACGGACCTAGTCCGAGGTCG</td>
<td>aacacTCAGGCCACATGAGGCCAGGTCc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgCCTTCCGTCCTACAGGACCGC</td>
<td>aacacCTTGTCGGTAGAAGGGGAAGAc</td>
</tr>
<tr>
<td>15</td>
<td>786</td>
<td>1</td>
<td>+</td>
<td>caccgGAAAGGCTACAGGCGACAGCCCG</td>
<td>aacacTCAGTGGCGCTAGCAGGTCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgGATCTCAGGCGGACACAGCCTG</td>
<td>aacacCCAGTCGCGCTCAGGTCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGGATGACGGGTCTGCTCAGG</td>
<td>aacacATCCGACTAGGTCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>+</td>
<td>caccgAGAATTCTCCTGCTGAGTGGTCG</td>
<td>aacacACTACAGGATTTACAGGTCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgGAAGGCTGGGCCTGTCACAGG</td>
<td>aacacCTAGTCGACAGGATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgCAATGCGACACGCCACAGGTGCTC</td>
<td>aacacGCTTAGCTTTACCAGGAGGAc</td>
</tr>
<tr>
<td>16</td>
<td>1140</td>
<td>1</td>
<td>+</td>
<td>caccgAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgGAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>+</td>
<td>caccgGAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>+</td>
<td>caccgGAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgGAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgGAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>+</td>
<td>caccgGAATGCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td>17</td>
<td>964</td>
<td>1</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>+</td>
<td>caccgGATCTCTGATGGCATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td>n1</td>
<td>1201</td>
<td>1</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td>n2</td>
<td>1191</td>
<td>1</td>
<td>+</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>+</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td>n3</td>
<td>1201</td>
<td>1</td>
<td>+</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>+</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>-</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>+</td>
<td>caccgGTCTTTAAAAGTTGGGATGAC</td>
<td>aacacGTCAGGAAAGCGGGACTTC</td>
</tr>
</tbody>
</table>
**Supplementary Table 1: Oligos used for cloning gRNAs.** The CRS/P/NC column indicates the CRS, promoter, or negative control. In the next column, the length of the CRS/P/NC is shown. For every 150 bp, approximately 1 gRNA was designed. The gRNAs are dispersed over the sequence of interest and can be located on the plus or minus strand.

<table>
<thead>
<tr>
<th>CRS</th>
<th>Hit</th>
<th>Cover</th>
<th>E-value</th>
<th>Ident.</th>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>NM_198248.1</td>
<td>Mus musculus zinc finger and BTB domain containing 40 (Zbtb40), mRNA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11%</td>
<td>2,00E-12</td>
<td>76.76%</td>
<td>NM_014870.4</td>
<td>Homo sapiens zinc finger and BTB domain containing 40 (ZBTB40), transcript variant 2, mRNA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11%</td>
<td>2,00E-12</td>
<td>76.76%</td>
<td>NM_001330398.1</td>
<td>Homo sapiens zinc finger and BTB domain containing 40 (ZBTB40), transcript variant 3, mRNA</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11%</td>
<td>2,00E-12</td>
<td>76.76%</td>
<td>NM_001083621.1</td>
<td>Homo sapiens zinc finger and BTB domain containing 40 (ZBTB40), transcript variant 1, mRNA</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>11%</td>
<td>2,00E-12</td>
<td>76.76%</td>
<td>AL035703.21</td>
<td>Human DNA sequence from clone RP1-61A9 on chromosome 1p35.2-36.13, complete sequence</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>11%</td>
<td>2,00E-12</td>
<td>76.76%</td>
<td>AB007947.2</td>
<td>Homo sapiens KIAA0478 mRNA for KIAA0478 protein</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL671011.9</td>
<td>Mouse DNA sequence from clone RP23-95O23 on chromosome 4, complete sequence</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>40%</td>
<td>3,00E-36</td>
<td>73.77%</td>
<td>AL035703.21</td>
<td>Human DNA sequence from clone RP1-61A9 on chromosome 1p35.2-36.13, complete sequence</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>11%</td>
<td>4,00E-09</td>
<td>71.92%</td>
<td>BC114607.1</td>
<td>Homo sapiens zinc finger and BTB domain containing 40, mRNA (cDNA clone MGC:133098 IMAGE:40027)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11%</td>
<td>4,00E-09</td>
<td>71.92%</td>
<td>AK095273.1</td>
<td>Homo sapiens cDNA FLJ37954 fis, clone CTONG2009270, weakly similar to Mus musculus transcriptional repressor RPS8 (rp58) mRNA</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11%</td>
<td>4,00E-09</td>
<td>71.92%</td>
<td>AB007947.2</td>
<td>Homo sapiens KIAA0478 mRNA for KIAA0478 protein</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11%</td>
<td>5,00E-08</td>
<td>71.23%</td>
<td>NM_014870.4</td>
<td>Homo sapiens zinc finger and BTB domain containing 40 (ZBTB40), transcript variant 2, mRNA</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11%</td>
<td>5,00E-08</td>
<td>71.23%</td>
<td>NM_001330398.1</td>
<td>Homo sapiens zinc finger and BTB domain containing 40 (ZBTB40), transcript variant 3, mRNA</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>11%</td>
<td>5,00E-08</td>
<td>71.23%</td>
<td>NM_001083621.1</td>
<td>Homo sapiens zinc finger and BTB domain containing 40 (ZBTB40), transcript variant 1, mRNA</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL671011.9</td>
<td>Mouse DNA sequence from clone RP23-95O23 on chromosome 4, complete sequence</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>34%</td>
<td>2,00E-49</td>
<td>78.65%</td>
<td>AL358788.17</td>
<td>Human DNA sequence from clone RP11-466J21 on chromosome 1, complete sequence</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL671011.9</td>
<td>Mouse DNA sequence from clone RP23-95O23 on chromosome 4, complete sequence</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>19%</td>
<td>2,00E-07</td>
<td>79.52%</td>
<td>AL591122.20</td>
<td>Human DNA sequence from clone RP11-415K20 on chromosome 1, complete sequence</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3%</td>
<td>0.041</td>
<td>88.89%</td>
<td>NG_042164.1</td>
<td>Homo sapiens oxysterol binding protein like 2 (OSBPL2), RefSeqGene on chromosome 20</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3%</td>
<td>0.041</td>
<td>88.89%</td>
<td>AL354836.13</td>
<td>Human DNA sequence from clone RP11-157P1 on chromosome 20, complete sequence</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL671011.9</td>
<td>Mouse DNA sequence from clone RP23-95O23 on chromosome 4, complete sequence</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>34%</td>
<td>7,00E-48</td>
<td>77.34%</td>
<td>AL591122.20</td>
<td>Human DNA sequence from clone RP11-415K20 on chromosome 1, complete sequence</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>6%</td>
<td>0.008</td>
<td>84.78%</td>
<td>AC277883.1</td>
<td>Homo sapiens chromosome 15 clone CH17-21B9, complete sequence</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>6%</td>
<td>0.008</td>
<td>84.78%</td>
<td>AC024375.13</td>
<td>Homo sapiens chromosome 15, clone RP11-572A8, complete sequence</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>6%</td>
<td>0.008</td>
<td>84.78%</td>
<td>AC135628.8</td>
<td>Homo sapiens chromosome 15, clone RP11-883G10, complete sequence</td>
<td></td>
</tr>
<tr>
<td>CRS</td>
<td>Hit</td>
<td>Cover</td>
<td>E-value</td>
<td>Ident.</td>
<td>Accession</td>
<td>Description</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL671011.9</td>
<td>Mouse DNA sequence from clone RP23-95O23 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>2a-b</td>
<td>17%</td>
<td>3,00E-24</td>
<td>76.88%</td>
<td>AL031279.1</td>
<td>Human DNA sequence from clone RP1-163O16 on chromosome 1p35.1-36.13, complete sequence</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3%</td>
<td>0.049</td>
<td>87.18%</td>
<td>NG_021394.1</td>
<td>Homo sapiens LDL receptor related protein 4 (LRP4), RefSeqGene on chromosome 11</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3%</td>
<td>0.049</td>
<td>87.18%</td>
<td>AC021573.10</td>
<td>Homo sapiens chromosome 1, clone RP11-411D10, complete sequence</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL671011.9</td>
<td>Mouse DNA sequence from clone RP23-95O23 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5%</td>
<td>0.014</td>
<td>78.12%</td>
<td>AC092500.2</td>
<td>Homo sapiens chromosome 3 clone RP11-111P21, complete sequence</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5%</td>
<td>0.014</td>
<td>78.12%</td>
<td>AC006515.7</td>
<td>Homo sapiens 3p22-7 PAC RPC15-1053D16 (Roswell Park Cancer Institute Human PAC Library) complete</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2%</td>
<td>0.049</td>
<td>88.89%</td>
<td>AC097649.3</td>
<td>Homo sapiens BAC clone RP11-92D1 from 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2%</td>
<td>0.049</td>
<td>88.89%</td>
<td>DQ304649.1</td>
<td>Homo sapiens anaphase promoting complex subunit 10 (ANAPC10) gene, complete cds</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL808114.11</td>
<td>Mouse DNA sequence from clone RP23-146A17 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>2a-b</td>
<td>25%</td>
<td>2,00E-51</td>
<td>82.63%</td>
<td>AL031279.1</td>
<td>Human DNA sequence from clone RP1-163O16 on chromosome 1p35.1-36.13, complete sequence</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3%</td>
<td>0.046</td>
<td>85.71%</td>
<td>AL445624.3</td>
<td>Homo sapiens chromosome 9 BAC RP11-512L9, complete sequence</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33%</td>
<td>8,00E-28</td>
<td>73.75%</td>
<td>AL031279.1</td>
<td>Human DNA sequence from clone RP1-163O16 on chromosome 1p35.1-36.13, complete sequence</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7%</td>
<td>0.008</td>
<td>82.00%</td>
<td>AC092265.3</td>
<td>Homo sapiens chromosome 1 clone RP4-516G21, complete sequence</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6%</td>
<td>0.008</td>
<td>82.98%</td>
<td>AL137076.6</td>
<td>Human DNA sequence from clone RP5-893G23 on chromosome 1p34.2-36.11, complete sequence</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8%</td>
<td>0.003</td>
<td>82.00%</td>
<td>AC092265.3</td>
<td>Homo sapiens chromosome 1 clone RP4-516G21, complete sequence</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>6%</td>
<td>0.008</td>
<td>82.98%</td>
<td>AL137076.6</td>
<td>Human DNA sequence from clone RP5-893G23 on chromosome 1p34.2-36.11, complete sequence</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>64%</td>
<td>1,00E-45</td>
<td>69.76%</td>
<td>AH010731.2</td>
<td>Homo sapiens secreted glycoprotein Wnt4 (WNT4) gene, partial cds</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>64%</td>
<td>1,00E-45</td>
<td>69.76%</td>
<td>NG_008974.1</td>
<td>Homo sapiens Wnt family member 4 (WNT4), RefSeqGene on chromosome 1</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>64%</td>
<td>1,00E-45</td>
<td>69.76%</td>
<td>AL445253.13</td>
<td>Human DNA sequence from clone RP4-660K3 on chromosome 1, complete sequence</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>64%</td>
<td>1,00E-45</td>
<td>69.76%</td>
<td>AL031281.6</td>
<td>Human DNA sequence from clone RP1-224A6 on chromosome 1p35.1-36.23, complete sequence</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5%</td>
<td>0.008</td>
<td>88.10%</td>
<td>AC084871.4</td>
<td>Homo sapiens BAC clone RP11-701P16 from 4, complete sequence</td>
</tr>
</tbody>
</table>
## Functional dissection of the Wnt4 locus

### Supplementary Table 2: BLAST hits from the human genome.

This table contains information about the alignments that were found between the mouse CRSs and the human genome using BLAST. The first column shows the number of the CRS and the second column shows the alignment number (these numbers correspond to the numbers in Supplementary Table 3). The query sequence in the BLAST alignment was the mouse CRS, and the first hit of every CRS is a validation of that sequence. Multiple blocks per CRS could be aligned to the human genome, therefore multiple hits are depicted per CRS. The “Cover” column shows the percentage of the sequence that is identical to the query sequence and the E-value is a statistical value for this alignment. The column “Ident.” shows the percentage that is identical between the mouse and human genome for every hit. The last two columns show the accession number of the data and a description of the project/clone.

<table>
<thead>
<tr>
<th>CRS</th>
<th>Hit</th>
<th>Cover</th>
<th>E-value</th>
<th>Ident.</th>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td></td>
<td>4%</td>
<td>0.029</td>
<td>91.67%</td>
<td>NG_007366.2</td>
<td>Homo sapiens interleukin 12 receptor subunit beta 1 (IL12RB1), RefSeqGene (LRG_72) on chromosome 1</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>4%</td>
<td>0.029</td>
<td>91.67%</td>
<td>AY771996.1</td>
<td>Homo sapiens interleukin 12 receptor, beta 1 (IL12RB1) gene, complete cds</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>4%</td>
<td>0.029</td>
<td>91.67%</td>
<td>AC020904.7</td>
<td>Homo sapiens chromosome 19 clone CTB-52I2, complete sequence</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>4%</td>
<td>0.029</td>
<td>91.67%</td>
<td>AC008569.7</td>
<td>Homo sapiens chromosome 19 clone CTC-548K16, complete sequence</td>
</tr>
<tr>
<td>11</td>
<td>5%</td>
<td></td>
<td>0.029</td>
<td>86.05%</td>
<td>AC008732.9</td>
<td>Homo sapiens chromosome 16 clone CTD-2519M14, complete sequence</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td></td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td>2</td>
<td>14%</td>
<td>0.001</td>
<td>66.85%</td>
<td></td>
<td>NG_008974.1</td>
<td>Homo sapiens Wnt family member 4 (WNT4), RefSeqGene on chromosome 1</td>
</tr>
<tr>
<td>3</td>
<td>14%</td>
<td>0.001</td>
<td>66.85%</td>
<td></td>
<td>AL031281.6</td>
<td>Human DNA sequence from clone RP1-224A6 on chromosome 1p35.1-36.23, complete sequence</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td></td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td>2a-c</td>
<td>49%</td>
<td>4,00E-27</td>
<td>68.45%</td>
<td></td>
<td>NG_008974.1</td>
<td>Homo sapiens Wnt family member 4 (WNT4), RefSeqGene on chromosome 1</td>
</tr>
<tr>
<td>3a-c</td>
<td>49%</td>
<td>4,00E-27</td>
<td>68.45%</td>
<td></td>
<td>AL031281.6</td>
<td>Human DNA sequence from clone RP1-224A6 on chromosome 1p35.1-36.23, complete sequence</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td></td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td>2</td>
<td>25%</td>
<td>6,00E-05</td>
<td>67.76%</td>
<td></td>
<td>NG_008974.1</td>
<td>Homo sapiens Wnt family member 4 (WNT4), RefSeqGene on chromosome 1</td>
</tr>
<tr>
<td>3</td>
<td>25%</td>
<td>6,00E-05</td>
<td>67.76%</td>
<td></td>
<td>AL031281.6</td>
<td>Human DNA sequence from clone RP1-224A6 on chromosome 1p35.1-36.23, complete sequence</td>
</tr>
<tr>
<td>1a-g</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td></td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td>2</td>
<td>27%</td>
<td>6,00E-58</td>
<td>77.00%</td>
<td></td>
<td>NG_008974.1</td>
<td>Homo sapiens Wnt family member 4 (WNT4), RefSeqGene on chromosome 1</td>
</tr>
<tr>
<td>3</td>
<td>27%</td>
<td>6,00E-58</td>
<td>77.00%</td>
<td></td>
<td>AL031281.6</td>
<td>Human DNA sequence from clone RP1-224A6 on chromosome 1p35.1-36.23, complete sequence</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
<td>0.0</td>
<td>100.00%</td>
<td></td>
<td>AL645468.11</td>
<td>Mouse DNA sequence from clone RP23-246F18 on chromosome 4, complete sequence</td>
</tr>
<tr>
<td>2a-b</td>
<td>38%</td>
<td>9,00E-23</td>
<td>71.15%</td>
<td></td>
<td>AL031281.6</td>
<td>Human DNA sequence from clone RP1-224A6 on chromosome 1p35.1-36.23, complete sequence</td>
</tr>
<tr>
<td>3</td>
<td>5%</td>
<td>0.47</td>
<td>78.33%</td>
<td></td>
<td>AC067735.36</td>
<td>Homo sapiens 12 BAC RP11-396F22 (Roswell Park Cancer Institute Human BAC Library) complete sequence</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>CRS</th>
<th>Hit</th>
<th>Alignments</th>
<th>Length</th>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Gaps</th>
<th>Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-1185</td>
<td>1185</td>
<td>2138</td>
<td>0.0</td>
<td>1185/1185(100%)</td>
<td>0/1185(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>690-825</td>
<td>136</td>
<td>82.4</td>
<td>2,00E-12</td>
<td>109/142(77%)</td>
<td>11/142(7%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>690-825</td>
<td>136</td>
<td>82.4</td>
<td>2,00E-12</td>
<td>109/142(77%)</td>
<td>11/142(7%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>690-825</td>
<td>136</td>
<td>82.4</td>
<td>2,00E-12</td>
<td>109/142(77%)</td>
<td>11/142(7%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>690-825</td>
<td>136</td>
<td>82.4</td>
<td>2,00E-12</td>
<td>109/142(77%)</td>
<td>11/142(7%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>690-825</td>
<td>136</td>
<td>82.4</td>
<td>2,00E-12</td>
<td>109/142(77%)</td>
<td>11/142(7%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1-1217</td>
<td>1217</td>
<td>2195</td>
<td>0.0</td>
<td>1217/1217(100%)</td>
<td>0/1217(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>33-319</td>
<td>287</td>
<td>161</td>
<td>3,00E-36</td>
<td>225/305(74%)</td>
<td>23/305(7%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>618-770</td>
<td>153</td>
<td>82.4</td>
<td>2,00E-12</td>
<td>112/154(73%)</td>
<td>2/154(1%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2c</td>
<td>450-503</td>
<td>54</td>
<td>45.5</td>
<td>0.60</td>
<td>44/56(79%)</td>
<td>2/56(3%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>624-768</td>
<td>145</td>
<td>72.5</td>
<td>4,00E-09</td>
<td>105/146(72%)</td>
<td>2/146(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>624-768</td>
<td>145</td>
<td>72.5</td>
<td>4,00E-09</td>
<td>105/146(72%)</td>
<td>2/146(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>624-768</td>
<td>145</td>
<td>72.5</td>
<td>4,00E-09</td>
<td>105/146(72%)</td>
<td>2/146(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>624-768</td>
<td>145</td>
<td>68.0</td>
<td>5,00E-08</td>
<td>104/146(71%)</td>
<td>2/146(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>624-768</td>
<td>145</td>
<td>68.0</td>
<td>5,00E-08</td>
<td>104/146(71%)</td>
<td>2/146(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>624-768</td>
<td>145</td>
<td>68.0</td>
<td>5,00E-08</td>
<td>104/146(71%)</td>
<td>2/146(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1-762</td>
<td>762</td>
<td>1375</td>
<td>0.0</td>
<td>762/762(100%)</td>
<td>0/762(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>319-581</td>
<td>263</td>
<td>205</td>
<td>2,00E-49</td>
<td>210/267(79%)</td>
<td>7/267(2%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>1-1014</td>
<td>1014</td>
<td>1829</td>
<td>0.0</td>
<td>1014/1014(100%)</td>
<td>0/1014(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>1b</td>
<td>8-103</td>
<td>96</td>
<td>47.3</td>
<td>0.14</td>
<td>74/107(69%)</td>
<td>12/107(11%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>368-449</td>
<td>82</td>
<td>67.1</td>
<td>2,00E-07</td>
<td>66/83(80%)</td>
<td>2/83(2%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>497-609</td>
<td>113</td>
<td>52.7</td>
<td>0.003</td>
<td>82/114(72%)</td>
<td>4/114(3%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>770-805</td>
<td>36</td>
<td>48.2</td>
<td>0.041</td>
<td>32/36(89%)</td>
<td>0/36(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>770-805</td>
<td>36</td>
<td>48.2</td>
<td>0.041</td>
<td>32/36(89%)</td>
<td>0/36(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1-726</td>
<td>726</td>
<td>1310</td>
<td>0.0</td>
<td>726/726(100%)</td>
<td>0/726(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>272-524</td>
<td>253</td>
<td>199</td>
<td>7,00E-48</td>
<td>198/256(77%)</td>
<td>3/256(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>400-445</td>
<td>46</td>
<td>50.0</td>
<td>0.008</td>
<td>39/46(85%)</td>
<td>2/46(4%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>400-445</td>
<td>46</td>
<td>50.0</td>
<td>0.008</td>
<td>39/46(85%)</td>
<td>2/46(4%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>400-445</td>
<td>46</td>
<td>50.0</td>
<td>0.008</td>
<td>39/46(85%)</td>
<td>2/46(4%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1-1216</td>
<td>1216</td>
<td>2194</td>
<td>0.0</td>
<td>1216/1216(100%)</td>
<td>0/1216(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>398-436</td>
<td>39</td>
<td>58.1</td>
<td>9,00E-05</td>
<td>36/39(92%)</td>
<td>0/39(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>887-1056</td>
<td>170</td>
<td>122</td>
<td>3,00E-24</td>
<td>133/173(77%)</td>
<td>3/173(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>516-554</td>
<td>39</td>
<td>49.1</td>
<td>0.049</td>
<td>34/39(87%)</td>
<td>0/39(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>516-554</td>
<td>39</td>
<td>49.1</td>
<td>0.049</td>
<td>34/39(87%)</td>
<td>0/39(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1-1217</td>
<td>1217</td>
<td>2195</td>
<td>0.0</td>
<td>1217/1217(100%)</td>
<td>0/1217(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>494-761</td>
<td>268</td>
<td>123</td>
<td>3,00E-24</td>
<td>194/272(71%)</td>
<td>11/272(4%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>989-1019</td>
<td>31</td>
<td>48.2</td>
<td>0.049</td>
<td>29/31(94%)</td>
<td>0/31(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>989-1019</td>
<td>31</td>
<td>48.2</td>
<td>0.049</td>
<td>29/31(94%)</td>
<td>0/31(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1-1217</td>
<td>1217</td>
<td>2195</td>
<td>0.0</td>
<td>1217/1217(100%)</td>
<td>0/1217(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>408-470</td>
<td>63</td>
<td>50.0</td>
<td>0.014</td>
<td>50/64(78%)</td>
<td>1/64(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>408-470</td>
<td>63</td>
<td>50.0</td>
<td>0.014</td>
<td>50/64(78%)</td>
<td>1/64(1%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>162-197</td>
<td>36</td>
<td>48.2</td>
<td>0.049</td>
<td>32/36(89%)</td>
<td>0/36(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>162-197</td>
<td>36</td>
<td>48.2</td>
<td>0.049</td>
<td>32/36(89%)</td>
<td>0/36(0%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1-1143</td>
<td>1143</td>
<td>2062</td>
<td>0.0</td>
<td>1143/1143(100%)</td>
<td>0/1143(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>456-665</td>
<td>210</td>
<td>213</td>
<td>2,00E-51</td>
<td>176/213(83%)</td>
<td>4/213(1%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>884-959</td>
<td>76</td>
<td>47.3</td>
<td>0.16</td>
<td>59/78(76%)</td>
<td>5/78(6%)</td>
<td>Plus/Minus</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1020-1058</td>
<td>39</td>
<td>48.2</td>
<td>0.046</td>
<td>36/42(86%)</td>
<td>3/42(7%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1-664</td>
<td>664</td>
<td>1198</td>
<td>0.0</td>
<td>664/664(100%)</td>
<td>0/664(0%)</td>
<td>Plus/Plus</td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Table 3: Details alignments mouse CRS sequences with hits from the human genome. This table provides detailed information about the different alignments of CRSs that were found between the mouse and human genome using BLAST\textsuperscript{z}. The first column represents the number of CRS and the second column the alignment that was found (these numbers are linked to the numbers found in Supplementary Table 2). The third column shows the area of the original CRS that was aligned to the human genome and the fourth column shows the length of that area. The columns "Score" and "Expect" show the alignment score and statistical value, respectively. The column "Identities" shows the percentage of the sequence that was aligned and the "Gap" column shows the percentage of the hit that was not aligned. The last column indicates the strand on which the sequence is located.


