



Universiteit  
Leiden  
The Netherlands

## Unraveling multifaceted roles of Grainyhead-like transcription factor-2 in breast cancer

Coban, B.

### Citation

Coban, B. (2024, November 5). *Unraveling multifaceted roles of Grainyhead-like transcription factor-2 in breast cancer*. Retrieved from <https://hdl.handle.net/1887/4107667>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4107667>

**Note:** To cite this publication please use the final published version (if applicable).

# Chapter 3

---

## GRHL2-controlled gene expression networks in luminal breast cancer

Published in: Zi Wang<sup>1</sup>, Bircan Coban<sup>1</sup>, Haoyu Wu<sup>2</sup>, Jihed Chouaref<sup>2</sup>, Lucia Daxinger<sup>2</sup>, Michelle T Paulsen<sup>3</sup>, Mats Ljungman<sup>3</sup>, Marcel Smid<sup>4</sup>, John W.M. Martens<sup>4</sup>, Erik HJ Danen<sup>1,5</sup>. GRHL2-controlled gene expression networks in luminal breast cancer. *Cell Commun Signal*. 2023 Jan 23;21(1):15.

<sup>1</sup>Leiden Academic Center for Drug Research, Leiden University, Leiden, The Netherlands; <sup>2</sup>Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands; <sup>3</sup>Departments of Radiation Oncology and Environmental Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA; <sup>4</sup>Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands; <sup>5</sup>correspondence to Erik HJ Danen, [e.danen@lacdr.leidenuniv.nl](mailto:e.danen@lacdr.leidenuniv.nl)

### Abstract

Grainyhead like 2 (GRHL2) is an essential transcription factor for development and function of epithelial tissues. It has dual roles in cancer by supporting tumor growth while suppressing epithelial to mesenchymal transitions (EMT). GRHL2 cooperates with androgen (AR) and estrogen receptors (ER) to regulate gene expression. We explore genome wide GRHL2 binding sites conserved in three ER $\alpha$ /GRHL2 positive luminal breast cancer cell lines by ChIP-Seq. Interaction with the ER $\alpha$ /FOXA1/GATA3 complex is observed, however, only for a minor fraction of conserved GRHL2 peaks. We determine genome wide transcriptional dynamics in response to loss of GRHL2 by nascent RNA Bru-seq using an MCF7 conditional knockout model. Integration of ChIP- and Bru-seq pinpoints candidate direct GRHL2 target genes in luminal breast cancer. Multiple connections between GRHL2 and proliferation are uncovered, including transcriptional activation of ETS and E2F transcription factors. Among EMT-related genes, direct regulation of CLDN4 is corroborated but several targets identified in other cells (including *CDH1* and *ZEB1*) are ruled out by both ChIP- and Bru-seq as being directly controlled by GRHL2 in luminal breast cancer cells. Gene clusters correlating positively (including known GRHL2 targets such as *ErbB3*, *CLDN4/7*) or negatively (including *TGFB1* and *TGFBR2*) with GRHL2 in the MCF7 knockout model, display similar correlation with GRHL2 in ER positive as well as ER negative breast cancer patients. Altogether, this study uncovers gene sets regulated directly or indirectly by GRHL2 in luminal breast cancer, identifies novel GRHL2-regulated genes, and points to distinct GRHL2 regulation of EMT in luminal breast cancer cells.

### Keywords

Breast cancer, luminal, GRHL2, CHIP-seq, BRU-seq, transcription, gene regulation

### Background

The *Grh* gene was discovered in *Drosophila* and its mammalian homologs have three members (*GRHL1*, *GRHL2* and *GRHL3*) [1]. Mice lacking GRHL1, 2, or 3 display neural tube closure defects and a variety of defects in epithelia

of several organs with disruption of epithelial adhesion complexes as a major common event [2-7]. GRHLs support expression of genes encoding key epithelial cell-cell junction proteins in desmosomes, adherens junctions, and tight junctions as well as targets involved in cytoskeletal regulation, membrane trafficking, and guidance cues. Several of these genes have been identified as direct GRHL transcriptional targets [3, 4, 8-15]. ChIP-seq in placenta, kidney, and lung epithelial cells has revealed >5000 GRHL2 binding peaks [11, 14, 15]. GRHL2 depletion in these same tissues identified a few hundred to a thousand genes whose expression was altered. Notably, i) overlap between these different tissues with respect to GRHL2 binding peaks and candidate target genes is limited pointing to common and tissue specific functions of GRHL2 and ii) for many of the GRHL2 target genes regulation appears indirect, which may involve GRHL2-regulated expression of other transcription factors or epigenetic modifiers [16, 17].

*GRHL2* is located on chromosome 8q22 that is frequently amplified in many cancers, including breast cancer, colorectal cancer and oral squamous cell carcinoma [18-20]. GRHL2 acts as an activator or suppressor of target gene transcription by interacting with promotor and enhancer regions in competition or in cooperation with other transcription factors and epigenetic regulators [2]. GRHL2 may enhance proliferation, replicative potential, and evasion of cell death through activation of the *ErbB3* gene, epigenetically promoting expression of *hTERT*, and suppressing death receptor expression [9, 18, 19]. Indeed, GRHL2 expression was negatively correlated with metastasis-free survival in breast cancer patients [21, 22]. By contrast, others have reported that high GRHL2 expression in breast cancer cell lines is associated with sensitivity to anoikis and chemotherapy and reduced tumor initiation capacity [23, 24].

Loss of GRHL2 was reported in gastric cancer and GRHL2 was found down-regulated at the invasive front of breast cancers and loss of GRHL2 expression in primary breast cancers correlated with lymph node metastasis [9, 25]. A key mechanism by which GRHL2 may suppress aspects of tumor progression is through inhibition of epithelial-to-mesenchymal transition

(EMT). GRHL2 acts in a double negative feedback loop with ZEB1 and it activates the expression of miR-200s that, in turn, are in a double negative feedback loop with ZEBs, thereby enforcing the epithelial phenotype [9, 17, 23, 24, 26, 27]. The roles of GRHL2 may be tumor type- and stage-specific through regulating different target genes in different cancers [28].

Breast cancer represents a heterogeneous disease with multiple clinically relevant subtypes appearing to originate from luminal or basal epithelial cells in the duct [29-31]. The luminal subtype accounts for the majority of breast cancer cases and can be treated by therapies targeting estrogen receptor alpha (ER $\alpha$ ) signaling [32]. Recent studies have shown that GRHL2 cooperates with androgen receptor in prostate cancer [33] and with ER $\alpha$  in breast cancer. Like FOXA1, GRHL2 may act as a pioneer factor, promoting chromatin accessibility and GRHL2 has been found to co-occupy enhancer elements with FOXA1, GATA3, and ER $\alpha$  to regulate ER $\alpha$  signaling output in hormone receptor positive breast cancer [34-37].

In this study, we identify genomic binding sites of GRHL2 shared among 3 luminal breast cancer cell lines and find that only a small subset of these GRHL2 peaks is associated with ER binding sites. We integrate this ChIP-seq data with Bru-seq analysis of genes showing transcriptional responses at different time points after conditional GRHL2 knockout in MCF7 cells. For genes showing sustained up- or downregulation in response to GRHL2 deletion, we explore correlations with GRHL2 expression in breast cancer patients. Our findings reveal gene sets regulated directly or indirectly by GRHL2 in luminal breast cancer that partly overlap but also appear markedly distinct from targets identified in other tissues.

## Methods

### Cell lines and plasmids

Human breast cancer cell lines representing the luminal subtype (MCF7, T47D and BT474) were obtained from the American Type Culture Collection. The Hs578T human basal-B breast cancer cell line served as a GRHL2-negative control. Cells were cultured in RPMI1640 medium with 10% fetal bovine

serum, 25 U/mL penicillin and 25 µg/mL streptomycin in the incubator (37°C, 5% CO<sub>2</sub>). For production of lentiviral particles, VSV, GAG, REV and Cas9 or single guide (sg) RNA plasmids were transfected into HEK293 cells using Polyethylenimine (PEI). After 2 days, lentiviral particles were harvested and filtered. Conditional Cas9 cells were generated by infecting parental cells with lentiviral particles expressing the Edit-R Tre3G promoter-driven Cas9 (Dharmacon) and selected by blasticidin. Limited dilution was used to generate Cas9 monoclonal cells. Subsequently, Cas9-monoclonal cells were transduced with U6-gRNA:hPGK-puro-2A-tBFP control non-targeting sgRNAs or GRHL2-specific sgRNAs (Sigma) and selected by puromycin. The EHF plasmid was kindly provided by Dr. Giuseppina Carbone, Institute of Oncology Research, Bellinzona, Switzerland and described previously [38, 39]. The EHF plasmid was transfected into cells using Lipofectamin 2000 according to a protocol provided by the manufacturer.

### Western blot

Cells were lysed by radioimmunoprecipitation (RIPA) buffer (150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate and 0.1% Tris and 1% protease cocktail inhibitor (Sigma-Aldrich, P8340)). Lysates were sonicated and protein concentration was determined by bicinchoninic acid (BCA) assay. Cell lysates were mixed with protein loading buffer, separated by SDS-PAGE, and transferred to a methanol-activated polyvinylidene difluoride (PVDF) membrane (Milipore, The Netherlands). The membrane was blocked with 5% bovine serum albumin (BSA; Sigma-Aldrich) for 1 hour at room temperature (RT). Next, membranes were stained with primary antibody overnight at 4°C and HRP-conjugated secondary antibodies for half hour at room temperature (RT). After staining with Prime ECL Detection Reagent (GE Healthcare Life science), chemoluminescence was detected with an Amersham Imager 600 (GE Healthcare Life science, The Netherlands). The following antibodies were used: GRHL2 (Atlas-Antibodies, hpa004820) Cas9 (Cell Signaling, 14697), and GAPDH (SantaCruz, sc-32233).

### ChIP-seq

Cells were grown in RPMI-1640 complete, serum-containing medium. Cross-linking was performed by 1% formaldehyde for 10 minutes at room temperature (RT). Then 1M glycine (141  $\mu$ l of 1M glycine for 1 ml of medium) was used to quench for 5 minutes at RT. Cells were washed twice with ice-cold PBS containing 5  $\mu$ l/ml phenylmethylsulfonyl fluoride (PMSF). Cells were harvested by centrifugation (2095 g for 5 minutes at 4°C) and lysed with NP40 buffer (150 mM NaCl, 50mM Tris-HCl, 5mM EDTA, 0.5% NP40, 1% Triton X-100) containing 0.1% SDS, 0.5% sodium deoxycholate and protease inhibitor cocktail (EDTA-free Protease Inhibitor Cocktail, Sigma). Chromatin was sonicated to an average size of 300 bp (**Fig. S1**). GRHL2-bound chromatin fragments were immunoprecipitated with anti-GRHL2 antibody (Sigma; HPA004820). Precipitates were washed by NP buffer, low salt (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl (pH 8.1), 150mM NaCl), high salt (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl (pH 8.1), 500mM NaCl) and LiCl buffer (0.25M LiCl, 1%NP40, 1% deoxycholate, 1mM EDTA, 10mM Tris-HCl (pH 8.1)). Chromatin was de-crosslinked by 1% SDS at 65°C. DNA was purified by Phenol:Chloroform:Isoamyl Alcohol (PCI) and then diluted in TE buffer.

In order to examine the quality of our samples before sequencing, ChIP-qPCR (quantitative polymerase chain reaction) was performed to validate interaction of GRHL2 with the promoter region of Claudin-4 (*CLDN4*), a known direct target gene of GRHL2 [4]. The results confirmed the GRHL2 binding site around the *CLDN4* promoter (**Fig. S2**). The following primers were used for ChIP-qPCR: *CLDN4* forward: gtgacctcagcatgggctttga, *CLDN4* reverse: ctctctctgaccagtttctctg, Control (an intergenic region upstream of the *GAPDH* locus) forward: atgggtgccactggggatct, Control reverse: tgccaaagcctaggggaaga, *ZEB1* promoter<sup>#</sup> forward: cggtccttagcaacaagggtt, *ZEB1* promoter<sup>#</sup> reverse: tcgcttggtctaaatgctcg. *ZEB1*<sup>##</sup> forward: gccgccgagcctccaacttt, *ZEB1*<sup>##</sup> reverse: tgctaggagaccgggcggttt, *OVOL2* exon forward: ccttaaactcgcgagtgaagacc, *OVOL2* exon reverse: gtagcgagcttggtgacacc, *CDH1* intron forward: gtatgaacggcaagcctctg, *CDH1* intron reverse: caaggagaccaggaagagaa. ChIP-qPCR data were collected and analyzed using the  $2^{-\Delta\Delta Ct}$  method [40].

For ChIP-seq, library preparation and paired-end (151bp) sequencing were performed by GenomeScan (Leiden, The Netherlands). MCF7, T47D and BT474 had 87393758, 84633440, and 82080866 pair-end reads, respectively.

### ChIP-seq analysis

Less than 5% of adapter sequences were present, and the mean per base sequence quality was >30, indicating high quality reads and no requirement for adapter-trimming (**Fig. S3, S4**). Paired-end reads were mapped to the human reference genome (hg38) using BWA-MEM [41] with default parameters. Over 93% of total reads were mapped to the human genome in T47D and MCF7 and 57.3% in BT474. Phred quality score (Q score) was used to measure base calling accuracy [42]. Q>30 scores (corresponding to a 0.1% error rate [43]) were >86% in T47D and MCF7 and 48.6% in BT474. Reads with low mapping quality ( $\leq Q30$ ) were filtered out. MACS version 2.1.0 [44] was used for peak calling by default settings. The q value was adjusted to 0.1 for BT474 cell line to avoid loss of peaks. The `annotatePeaks` and `MergePeaks` functions from HOMER [45] were used to annotate and overlap peaks, respectively. ChIPseeker was used for the analysis of ChIP-seq peaks coverage plot and the density profile of GRHL2 binding sites [46]. Motif analysis was performed using ChIP-seq peaks with high scores by the MEME-ChIP program with default settings. ChIP-seq data was visualized by the UCSC genome browser. To analyze coverage of GRHL2 peaks at consensus motifs for GRHL2, ER $\alpha$ , FOXA1, and GATA3 binding, the JASPAR 2022 database was used to identify motifs [47]. To analyze colocalization of our GRHL2 binding events with published ER $\alpha$  peaks in luminal breast cancer cells, ChIP-seq data files from a study mapping ER $\alpha$  binding sites in MCF7, BT474, and T47D [48] were intersected using bedtools (v2.3.0) [PMID: 20110278] and ChIP-seq data files from two different studies mapping ER-alpha binding sites in MCF7 were intersected [49, 50].

### Bru-seq

MCF7 cells expressing inducible Cas9 and control non-targeting sgRNAs or GRHL2-specific sgRNAs were exposed to 1  $\mu$ g/ml doxycycline. At different timepoints after doxycycline-induced deletion of GRHL2, cells were

## Chapter 3

---

incubated with a final concentration of 2 mM Bru at 37°C for 30 minutes. Cells were lysed in TRIzol reagent (Sigma) and Bru-labelled nascent RNA was isolated using an anti-BrdU antibody conjugated to magnetic beads [51]. Subsequently, cDNA libraries were generated using the Illumina TruSeq library kit and sequenced using the Illumina NovaSeq 6000 Sequencing System. Sequence reads were strand-specific, paired-ended with read lengths of ~150 nucleotides. Reads were pre-mapped to the ribosomal RNA (rRNA) repeating unit (GenBank U13369.1) and the mitochondrial and EBV genomes (from the hg38 analysis set) using Bowtie2 (2.3.3). Unaligned reads were subsequently mapped to human genome build hg38/GRCh38 using STAR (v 2.5.3a) and a STAR index created from GENCODE annotation version 27 [51, 52].

### **Bru-seq analysis**

To identify GRHL2-regulated genes, an inter-sample comparison analysis was performed comparing RPKM (reads per kilobase per million mapped reads) for each gene in the doxycycline-treated samples compared to the untreated sample, to obtain fold-change (FC) and *p* values. Genes with  $p < 0.05$  and  $FC > 2$  or  $FC < 0.5$  in any of the doxycycline-treated samples relative to untreated cells were filtered. Subsequently, genes responding to Cas9 induction in the context of both GRHL2 sgRNAs were selected and genes responding also in the context of control sgRNA were eliminated from this list. A heatmap was generated by R. The function “fviz\_nbclust()” from the R package “factorextra” was used to determine and visualize the optimal numbers of clusters using the method “within cluster sums of square”. The graph is attached. The STRING database (version 11.5) was used to assign protein interaction networks to Bru-seq data [53].

### **Breast cancer patient mRNA expression data analysis**

A compendium microarray dataset, all Affymetrix U133a, was used, containing RNA expression data of primary tumors of 867 untreated, lymph node negative patients (MA-867 dataset [59]; publicly available at [GSE2034](#), [GSE5327](#), [GSE2990](#), [GSE7390](#) and [GSE11121](#)). Raw .cel files were downloaded, processed with fRMA and batch effects were corrected using ComBat.

RNAseq data retrieved from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) data set [54, 55] was used consisting of targeted sequencing data of 1904 primary breast tumors with matched normal tissues. Data visualization and calculation of co-expression z-scores were performed using cBioPortal (<https://www.cbioportal.org/>).

### SRB assay

For Sulforhodamine B (SRB) assays, cells were seeded into 96-well plates. At indicated time points, cells were fixed with 50% Trichloroacetic acid (TCA, Sigma-Aldrich) for 1 hour at 4 °C and then plates were washed with demineralized water four times and air-dried at RT. Subsequently, 0.4% SRB (60 µl/well) was added and kept for at least 2 hours at RT. The plates were washed five times with 1% acetic acid and air-dried. 10 mM (150 µl/well) Tris was added and kept for half hour at RT with gentle shaking. The absorbance value was measured by a plate-reader Fluostar OPTIMA.

## Results

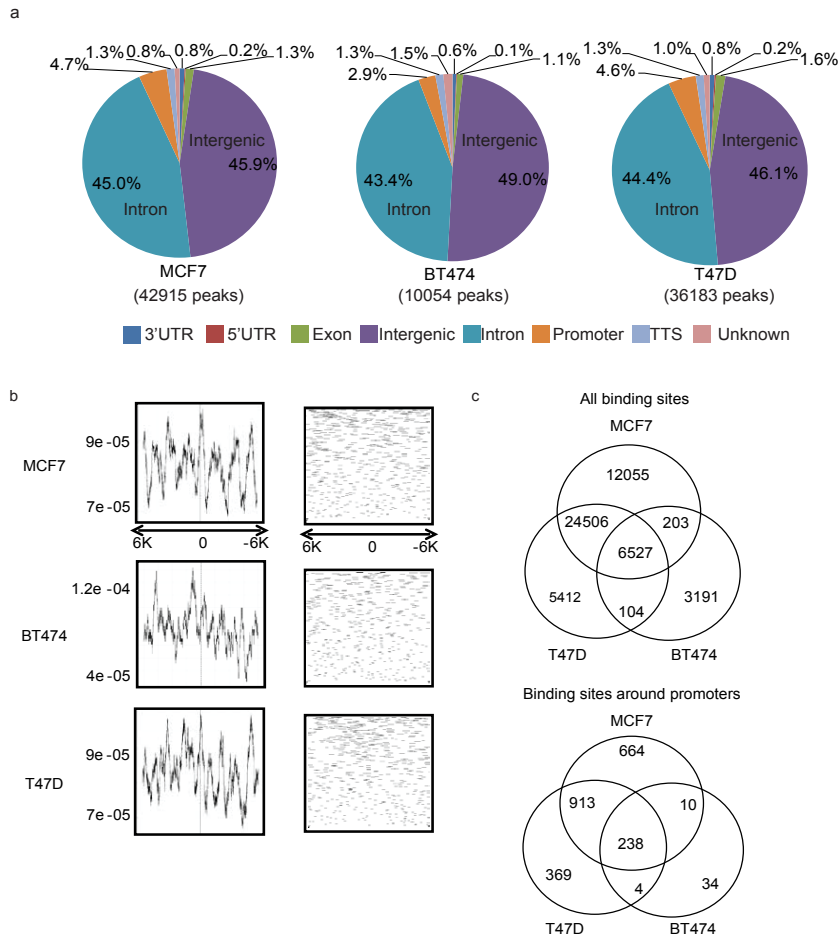
### Genome-wide identification of GRHL2 binding sites in luminal breast cancer cells

To identify GRHL2 binding sites, ChIP-seq was performed in the human luminal breast cancer cell lines, MCF7, T47D and BT474. As a quality control of the ChIP samples, ChIP-qPCR confirmed the interaction of GRHL2 with the promoter region of *CLDN4*, a known direct target gene of GRHL2 [4] in all three luminal, GRHL2-positive cell lines but not in the GRHL2-negative Hs578T human basal-B breast cancer cell line (**Fig. S2**). Subsequently, ChIP-seq was performed and the coverage of peak regions across chromosomes was analyzed [46]. In each sample, GRHL2 was associated with all chromosomes (**Fig. S5**).

GRHL2 binding sites were mainly located in intergenic regions and introns, with ~3-5% of the peaks located in -1000 bp to +100 bp promoter regions (**Fig. 1a**). Analysis of read count frequency and density profiling of GRHL2 binding sites within -6000 bp to +6000 bp of the transcription start site (TSS)

### Chapter 3

showed no enrichment around the TSS (**Fig. 1b**). Intersection of the data in the 3 cell lines identified 6527 conserved GRHL2 binding sites in luminal breast cancer cells. Of these, 238 binding sites located in the -1000 bp to +100 bp regions, representing candidate interactions for direct GRHL2-mediated regulation of gene promoter activity (**Fig. 1c; Table S1**).

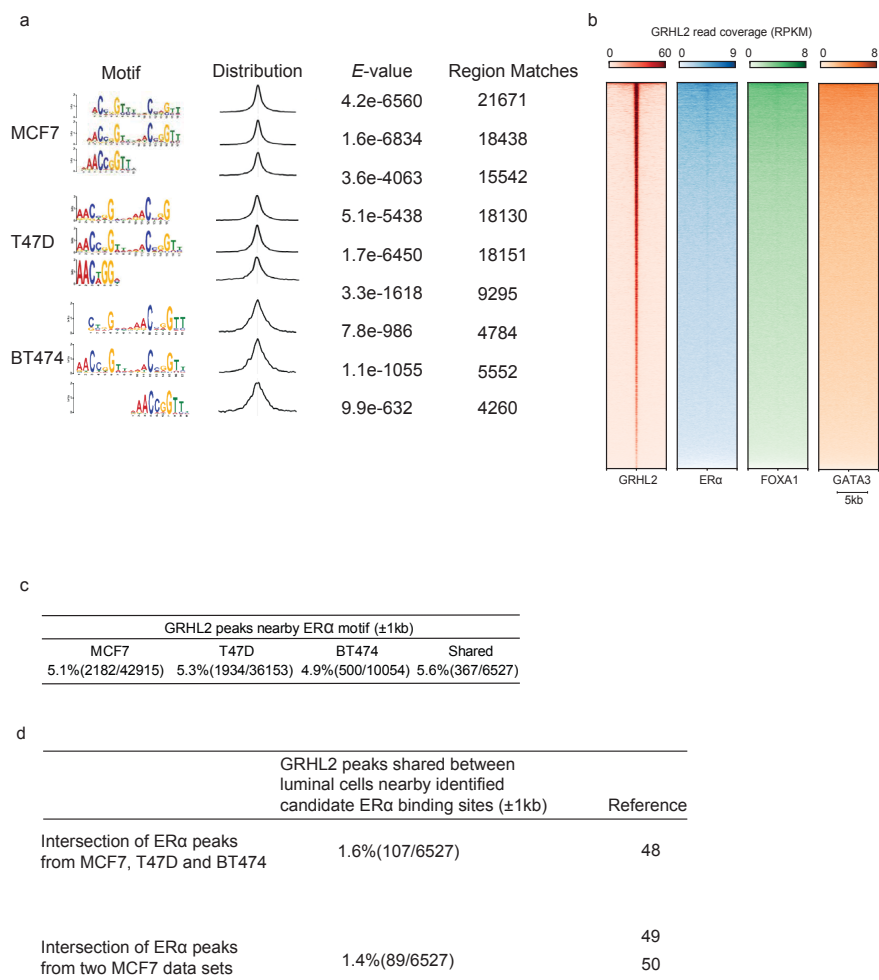


**Fig. 1. GRHL2 ChIP-seq in luminal breast cancer cells. (a)** Percentage of GRHL2 binding sites found at promoter regions, 5' untranslated regions (UTRs), 3' UTRs, exons, introns, intergenic regions, transcription termination sites (TTSs) and

unknown regions in the indicated luminal breast cancer cell lines. Promoter regions are defined as -1000 bp to +100 bp from the transcription start sites (TSS). **(b)** Read count frequency and density profile of GRHL2 binding sites within -6000 bp to +6000 bp of the TSS. Left panels show GRHL2 ChIP-seq read count frequencies in indicated cell lines (Y axis, read count frequency; X axis, genomic region). Right panels show density of ChIP-seq reads for GRHL2 binding sites in the indicated cell lines. **(c)** Venn diagrams showing overlap of GRHL2 binding sites among the three indicated cell lines. Top panel shows overlap for all peaks. Bottom panel shows overlap for peaks within the -1000 to +100 promoter region.

### **A small proportion of GRHL2 peaks is associated with ER $\alpha$ binding**

MEME-ChIP identified 3 GRHL2 binding motifs low *E* values in each cell line (**Fig. 2a**), whose core binding site matched previously published motifs [14, 15, 17, 56]. Based on the published interaction of GRHL2 with ER $\alpha$ , FOXA1, and GATA3 at enhancer elements of target genes [34-36, 57], we addressed to what extent the identified conserved GRHL2 binding sites in luminal breast cancer cells were flanked by putative binding sites for the ER $\alpha$ -mediated transcriptional complex. Heatmap visualization showed concentration of the GRHL2 peaks at the consensus GRHL2 motif [AACCGGTT] as expected (**Fig. 2b**). GRHL2 peaks showed only a weak trace for ER $\alpha$  [AGGTCAnnnTGACCT] and a barely detectable trace for the FOXA1 motif [TGTTT(A/G)C], and no concentration of the GATA3-binding motif [A/T]GATA(A/G) was observed. Indeed, among the shared GRHL2 peaks in luminal breast cancer cells ~5% was flanked by an ER $\alpha$  binding motif within +/- 1000bp (**Fig. 2c**).



**Fig. 2. Association of GRHL2 motif with ER transcriptional complex in luminal breast cancer cells. (a)** DNA-binding motif of GRHL2 in luminal breast cancer. From left to right, the first panel shows the identified motifs in the indicated cell lines. The second panel shows distribution of the best matches to the motif in the sequences. The third panel shows the *E*-value, representing the significance of the motif according to the motif discovery. The last panel shows the number of regions that match the corresponding motif. **(b)** Heatmaps showing the coverage of identified GRHL2 peaks shared between MCF7, BT474 and T47D at GRHL2 motifs

(red) (n=20766), ER $\alpha$  motifs (blue) (n=76564), FOXA1 motifs (Green) (n=88923) and GATA3 motifs (Orange) (n=93403). Note that the read coverage scale differs for the different heatmaps. **(c)** Table indicating the occurrence of ER $\alpha$  consensus motif in a region spanning 1000bp up- and downstream of all GRHL2 peaks either identified in the indicated cell lines (left 3 columns) or shared between the indicated cell lines (right column). **(d)** Table indicating the occurrence of published ER $\alpha$  binding events in a region spanning 1000bp up- and downstream of all GRHL2 peaks shared between MCF7, BT474 and T47D (upper row) or shared between 2 MCF7 datasets (bottom row).

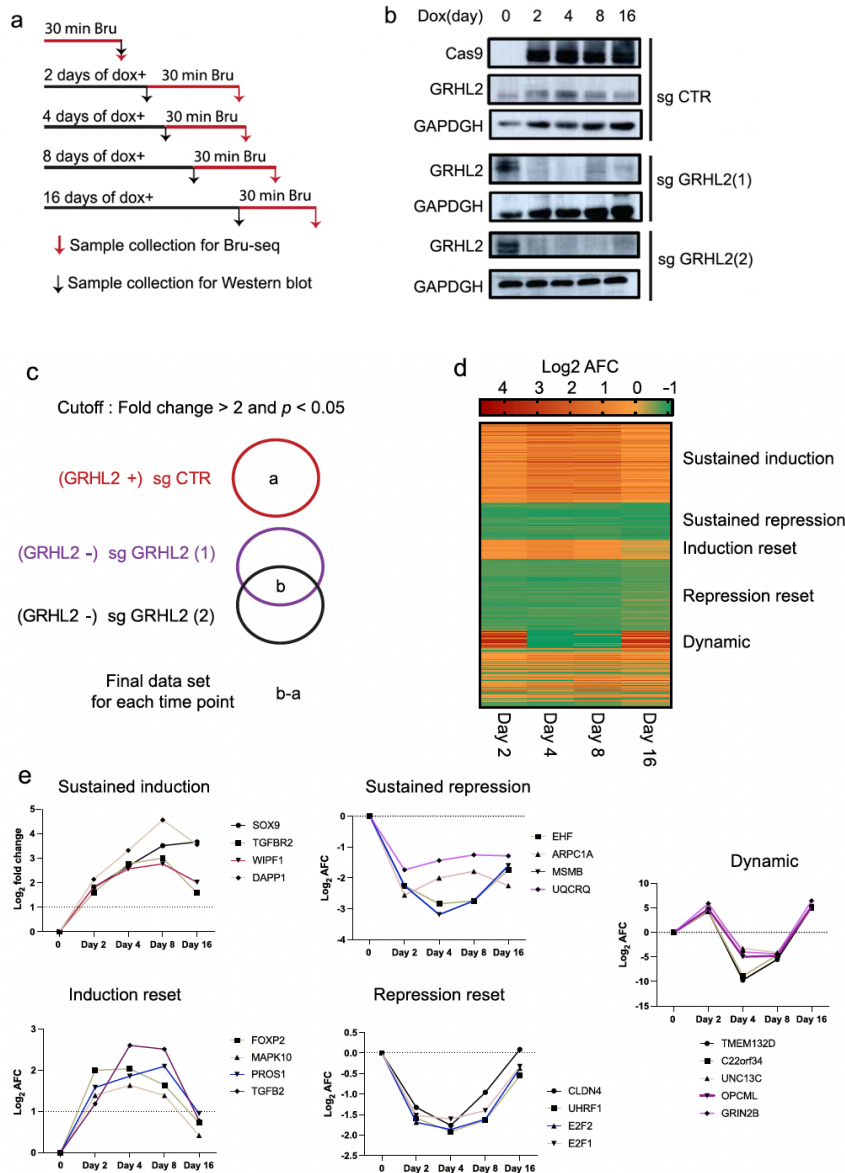
To further address colocalization of GRHL2 and ER $\alpha$  binding in luminal breast cancer, regions flanking +/- 1000bp of the conserved GRHL2 peaks in MCF7, BT474, and T47D were interrogated for the presence of previously reported ER $\alpha$  binding events. For this purpose, ChIP-seq data files from a study mapping ER $\alpha$  binding sites in MCF7, BT474, and T47D [48] and ChIP-seq data files from two studies mapping ER $\alpha$  binding sites in MCF7 were intersected [49, 50]. These studies had used similar culture conditions as ours, using phenol red medium and serum containing estrogen. Only a minor fraction of ~1.5% of conserved GRHL2 peaks identified in our study was flanked by established ER $\alpha$  binding sites in luminal breast cancer cells identified in those studies (**Fig. 2d**). Altogether, this data indicated that the majority of GRHL2 binding sites in luminal breast cancer cells were not associated with the ER $\alpha$ -mediated transcriptional complex.

### Changes in gene transcription in response to GRHL2 loss

Next, we employed nascent RNA Bru-seq to investigate genome-wide dynamic changes in DNA transcription triggered by GRHL2 loss. For this purpose, we made use of a conditional Cas9 MCF7 knockout model expressing a control or 2 different GRHL2 sgRNAs (sgCTR, sgGRHL2(1) and sgGRHL2(2), respectively). At 0, 2, 4, 8, or 16 days after GRHL2 knockout, cells were incubated with bromouridine (BrU) for 30 minutes to label nascent RNA (**Fig. 3a**) and analyzed in parallel by Western blot for the induction of Cas9 and deletion of GRHL2 (**Fig. 3b**). To identify GRHL2-regulated genes, for each time point, the average fold change (AFC) of transcription induced by doxycycline treatment in the two sgGRHL2 and sgCTR samples was determined. 262

# Chapter 3

genes were found to be upregulated and 226 genes were downregulated in at least one time point after GRHL2 loss in both sgGRHL2 samples ( $FC > 2$  or  $FC < 0.5$ ;  $p < 0.05$ ) but not in the sgCTR samples (**Fig. 3c**; **Table S2**).



**Fig. 3.** Bru-seq analysis of transcriptional changes in response to GRHL2 loss in luminal breast cancer MCF7 cells. (a) Bru-seq sample preparation. Bromouridine

(Bru) labeling of nascent RNA was carried out for 30 minutes at the indicated time points after doxycycline (dox)-induced GRHL2 deletion. **(b)** Western blot analysis of GRHL2 expression levels at the indicated time points in sgCTR and sgGRHL2 transduced MCF7 cells. Cas9 induction is monitored and GAPDH serves as loading control. **(c)** Bru-seq data analysis approach. Each circle represents a gene set with differential transcription relative to the condition where no doxycycline was added. **(d)** Heatmap for genes whose transcription was altered in response to GRHL2 depletion. **(e)** Graphs depicting clusters of genes with distinct patterns of transcriptional changes in response to GRHL2 depletion. Graphs represent  $\log_2$  AFC of transcription in sgGRHL2(1) and sgGRHL2(2) cells. “Dynamic”: genes with  $\text{AFC} > 2$ ;  $p < 0.05$  at some and  $\text{AFC} < 0.5$ ;  $p < 0.05$  at other time points. “Sustained induction”: genes with  $\text{AFC} > 2$ ;  $p < 0.05$  at all time points. “Sustained repression”: genes with  $\text{AFC} < 0.5$ ;  $p < 0.05$  at all time points. “Induction reset”: genes with  $\text{AFC} > 2$ ;  $p < 0.05$  at early time points followed by a return to  $1 < \text{AFC} < 2$  at day 16. “Repression reset”: genes with  $\text{AFC} < 0.5$ ;  $p < 0.05$  at early time points followed by a return to  $0.5 < \text{AFC} < 1$  at day 16.

GRHL2-regulated genes were clustered in a heatmap using the AFC at each time point (**Fig. 3d**). Five clusters were identified based on transcriptional dynamics (**Fig. 3e; Table S2**). There was no preference for the subset of genes containing GRHL2 binding sites flanked by ER $\alpha$  binding in either of the clusters. Clusters displaying sustained upregulation of RNA synthesis or a transient induction that subsequently returned to baseline included *TGFB1*, *TGFB2*, and *TGFB2* pointing to enhanced TGF $\beta$  signaling. Other clusters showed sustained downregulation of RNA synthesis following GRHL2 deletion or a transient repression that subsequently returned to baseline. These included genes encoding the epithelial specific ETS transcription factor EHF, the *E2F1* and *E2F2* genes encoding E2F transcription factors involved in cell cycle progression, and the *CLDN4* gene encoding an epithelial tight junction protein. Another cluster showed responses that could be categorized as highly dynamic with alternating increased and decreased transcription.

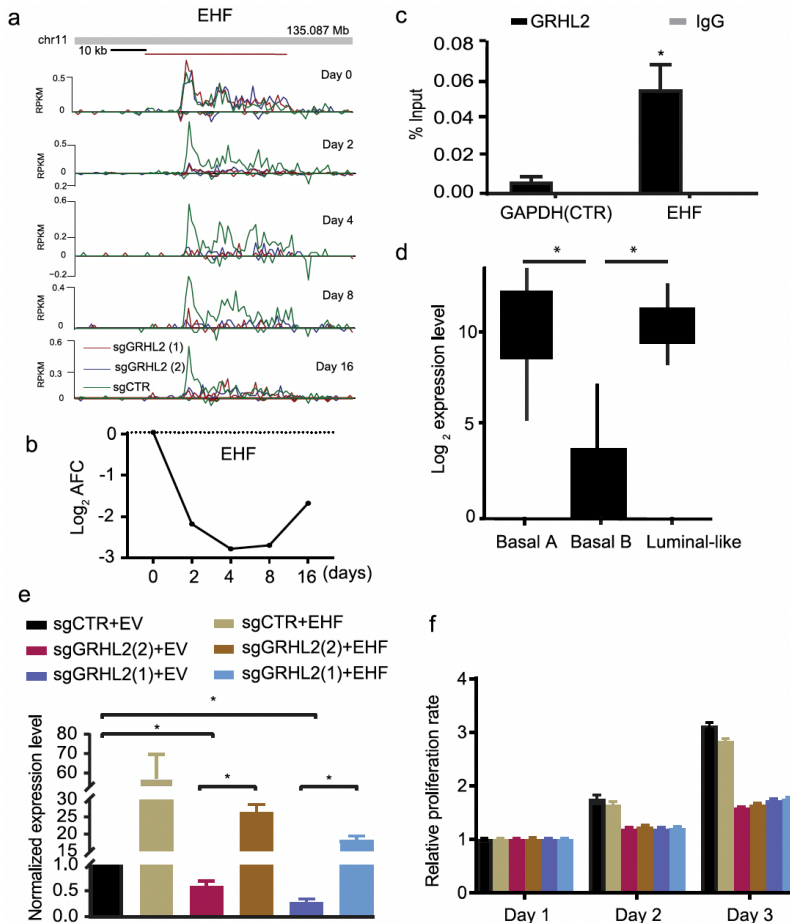
### Identification of candidate genes regulated by GRHL2 promoter binding

GRHL2 can regulate gene transcription through interaction with gene promoter or enhancer elements [2]. We intersected the list of genes whose expression levels were significantly altered after GRHL2 loss in MCF7 at one or

more time points as identified by Bru-seq, with genes harboring GRHL2 binding sites in the -1000 bp to +100 bp promoter regions in MCF7 identified by ChIP-seq. 53 genes were identified where transcriptional regulation could be explained by direct GRHL2 interactions at the promoter region (**Table S2**; genes indicated in bold). Restricting this list to genes harboring GRHL2 binding sites in the promoter regions that were shared in all three luminal breast cancer cell lines, reduced this number to 9 (**Table S1**; genes indicated in bold). The presence or absence of GRHL2 binding sites in the promoter region did not correspond to the dynamic pattern of the transcriptional response of the gene (**Table S2**). Together, this indicated that the majority of the genes showing a transcriptional response to GRHL2 depletion was regulated either by direct interactions at enhancer elements or indirectly, e.g., through GRHL2 regulation of a transcription factor targeting the gene of interest.

### **EHF is a direct GRHL2-target inversely correlated with GRHL2 in breast cancer subtypes**

EHF was identified as a GRHL2 target harboring a GRHL2 binding site in its promoter region that was conserved in all three luminal breast cancer cell lines (**Fig. 3e**; **Table S1,2**). EHF had not been previously reported as a GRHL2 target gene while our Bru-seq tracks showed that EHF transcription was rapidly and continuously attenuated following GRHL2 loss (**Fig. 4a,b**). ChIP-qPCR confirmed the interaction between GRHL2 and the promoter region of the *EHF* gene (**Fig. 4c**). *EHF* is a member of the ETS transcription factor subfamily characterized by epithelial-specific expression [58]. Epithelial markers (e.g., GRHL2, CLDN4 and E-cadherin) are lost in basal B breast cancer cells as compared to the luminal and basal A subtype and we examined whether *EHF* expression followed this pattern. Indeed, RNA-seq data from a panel of 52 human breast cancer cell lines [59] showed a decrease of *EHF* RNA levels in the basal B subtype (**Fig. 4d**).

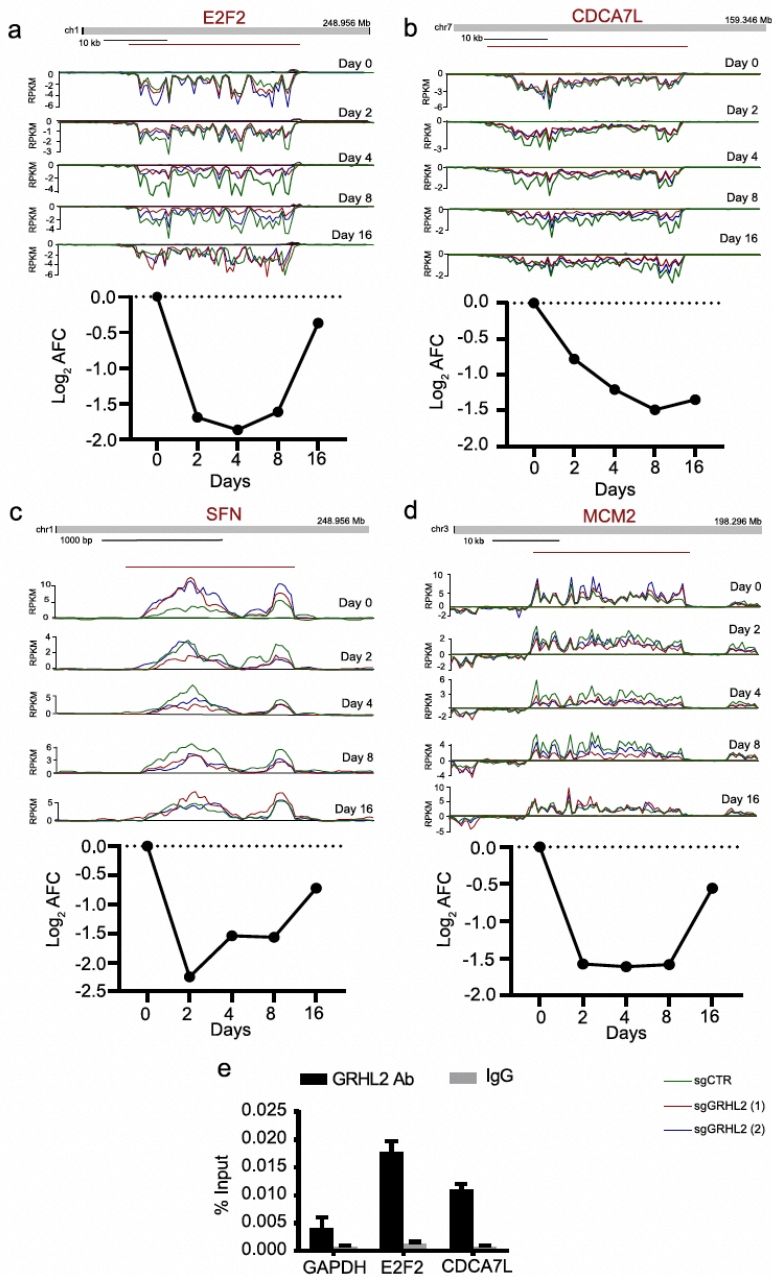


**Fig. 4. EHF represents a direct GRHL2 regulated gene.** **(a)** Bru-seq reads for *EHF* at indicated time points after to GRHL2 deletion. Track colors: green, sgCTR; red, sgGRHL2(1); blue, sgGRHL2(2). Positive y-axis indicates the plus-strand signal of RNA synthesis from left to right and the negative y-axis represents the minus-strand signal of RNA synthesis from right to left. **(b)** Line graph depicting the  $\log_2$  AFC of *EHF* transcription in sgGRHL2(1) and sgGRHL2(2) cells. **(c)** ChIP-qPCR showing enrichment of GRHL2 binding sites in *EHF* promoter region but not in the control *GAPDH* gene. Graph represents the efficiency of indicated genomic DNA co-precipitation with anti-GRHL2 Ab (black bars) or IgG control Ab (grey bars). Signals for IgG control and GRHL2 antibody pulldown samples are normalized to input DNA and are presented as % input with SEM from 3 technical replicates. Data are statistically analyzed by t-test and \* indicates  $p < 0.05$ . **(d)** *EHF* mRNA expression

in a panel of 52 human breast cancer cell lines covering luminal-, basal A-, and basal B subtypes extracted from RNA-seq data. Data is statistically analyzed by t-test and \* indicates  $p < 0.05$ . **(e)** qRT-PCR analysis of expression level of *EHF* mRNA after 4 days of doxycycline treatment of MCF7 cells transduced with dox-inducible Cas9 and sgCTR or sgGRHL2 constructs, in combination with ectopic expression of *EHF* or empty vector (EV) plasmids. Data are presented as mean  $\pm$  SEM from three technical replicates. Data are statistically analyzed by t-test. \* Indicates  $p < 0.05$ . **(f)** Graph showing results from SRB assay after 4 days doxycycline-treatment as in (e) and subsequent culture for the indicated time periods.

Studies in various cancer types have attributed tumor promoting as well as tumor suppressive roles to EHF but its role in breast cancer is largely unknown [60]. GRHL2 loss led to a rapid reduction in MCF7 cell growth and we tested whether ectopically overexpressed *EHF* could enhance proliferation in absence of GRHL2. However, overexpression of *EHF* did not rescue cell proliferation of GRHL2 KO MCF7 cells (**Fig. 4e,f**). The RNA synthesis rates of several other genes supporting cell cycle progression were rapidly suppressed in response to GRHL2 loss, including E2F transcription factors *E2F1* and *E2F2* and other genes such as *CDCA7L* and *MCM2* [61-63] (**Fig. 5a-d; Fig. 7a**). Our ChIP-seq data revealed GRHL2 binding sites in the promoter regions of *E2F2* and *CDCA7L* in MCF7 (**Table S2**) and this finding was corroborated by ChIP-qPCR analysis (**Fig. 5e**). Altogether, these results showed that several genes involved in cell cycle progression are rapidly downregulated following GRHL2 depletion with *EHF*, *E2F2*, and *CDCA7L* representing candidate targets for direct transcriptional regulation by GRHL2 at the gene promoter.

## GRHL2-controlled gene expression networks



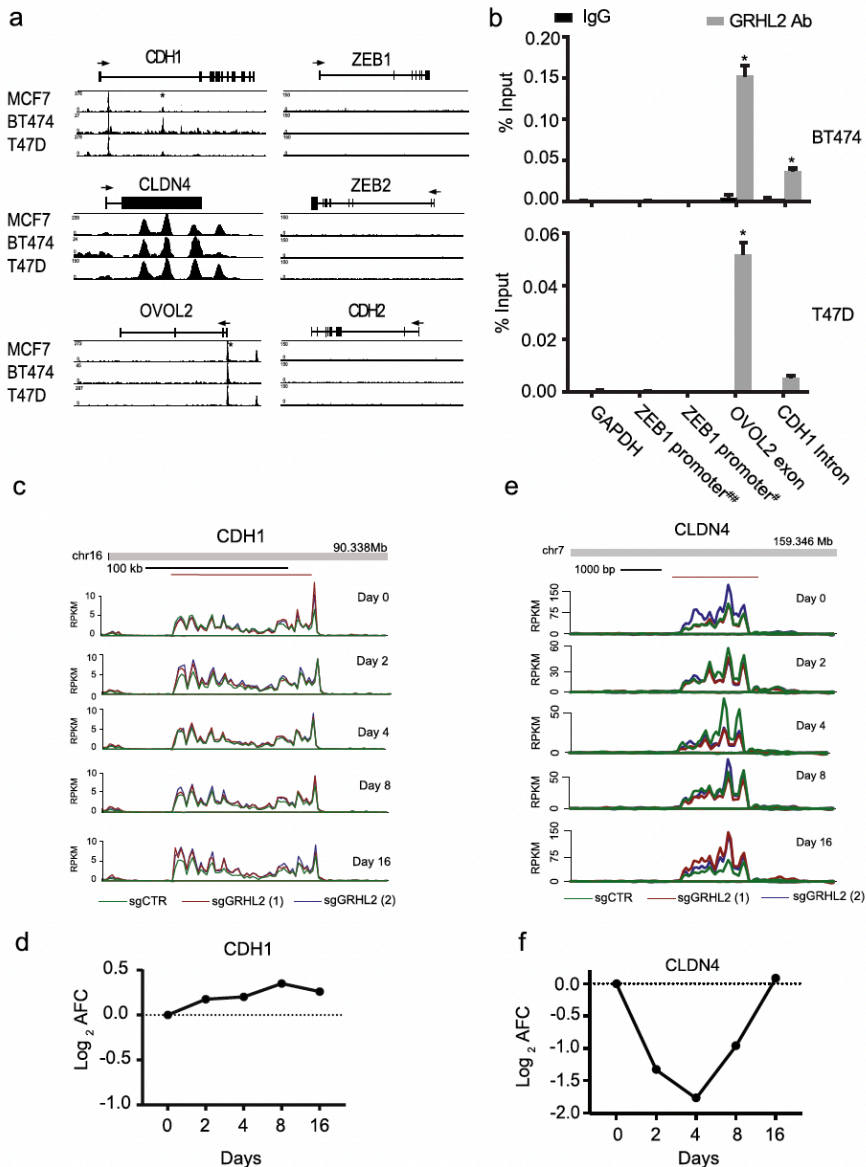
**Fig. 5. Downregulation of RNA synthesis for genes involved in cell cycle progression after GRHL2 loss. (a-d) Top: Bru-seq reads for indicated genes**

at indicated time point after to GRHL2 deletion. Track colors: green, sgCTR; red, sgGRHL2(1); blue, sgGRHL2(2). **Bottom:** Line graphs depicting the log<sub>2</sub> AFC of transcription in sgGRHL2(1) and sgGRHL2(2) cells for the indicated genes. The positive y-axis indicates the plus-strand signal of RNA synthesis from left to right and the negative y-axis represents the minus-strand signal of RNA synthesis from right to left. **(e)** Validation of interaction of GRHL2 binding sites with the promoter regions of indicated genes by ChIP-qPCR. Signals for IgG control and GRHL2 antibody pulldown samples are normalized to input DNA and are presented as % input with SEM from 3 technical replicates. Data are statistically analyzed by t-test and \* indicates  $p < 0.05$ .

### **Regulation of EMT-related genes: CLDN4 but not CDH1, ZEB1, and ZEB2 represent direct GRHL2 targets in luminal breast cancer**

GRHL2 and OVOL2 support an epithelial phenotype and counteract EMT transcription factors such as ZEB1, ZEB2, and SNAIL. Genes encoding epithelial adhesion components such as CLDN4 in tight junctions or E-cadherin (CDH1) in adherens junctions are regulated by this balance [64]. It has been reported that GRHL2 binding sites are present in the intronic region of *CDH1* and in the promoter regions of *CLDN4* and *OVOL2* for activation of transcription, and GRHL2 was reported to bind the *ZEB1* gene as a negative regulator [4, 12, 15, 23, 24, 65].

In our ChIP-seq data, a conserved intronic GRHL2 binding site was observed in *CDH1* that was validated by ChIP-qPCR (**Fig. 6a,b**). However, while GRHL2 was found to transcriptionally activate *CDH1* in earlier reports [4, 21, 33] we did not observe downregulation of *CDH1* nascent RNA synthesis in the first 16 days after GRHL2 loss (**Fig. 6c,d**). No GRHL2 peaks were associated with *CDH2* (encoding N-cadherin, a mesenchymal marker) while GRHL2 binding was conserved in the promoter regions of *CLDN4* and *OVOL2* (**Fig. 6a,b; Fig S2**). *CLDN4* also showed multiple GRHL2 binding sites across the coding and non-coding regions. *CLDN4* transcription was suppressed at 2, 4, and 8 days after GRHL2 depletion but recovered at 16 days (**Table S2; Fig. 6c,d**) whereas *OVOL2* was not affected (data not shown).



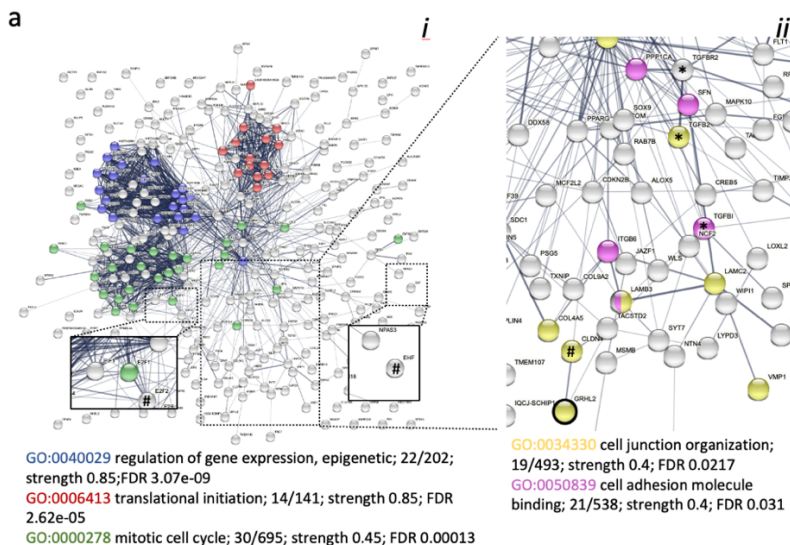
**Fig. 6. Regulation of EMT related genes by GRHL2. (a)** ChIP tracks for the indicated genes in three luminal breast cancer cell lines. The track height is scaled from 0 to the indicated number. The locus with its exon/intron structure is presented above the tracks. \*Indicates binding sites validated by ChIP-qPCR in (b). **(b)** ChIP-qPCR validation of presence and absence of GRHL2 binding sites identified by ChIP-seq. Graphs represent the efficiency

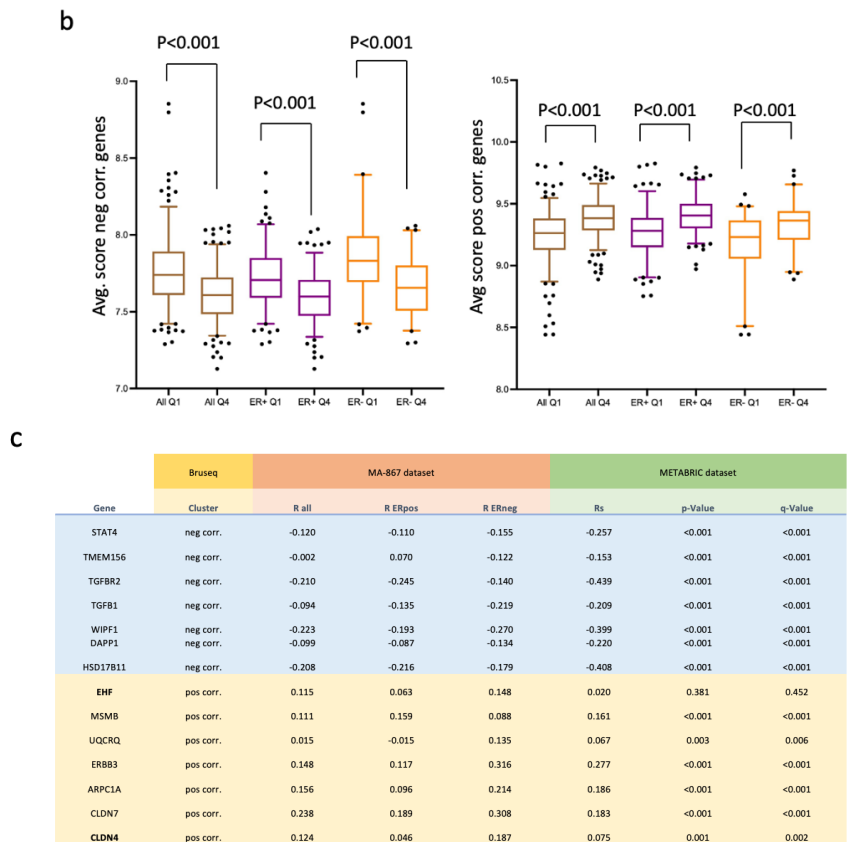
of indicated genomic DNA co-precipitation with anti-GRHL2 Ab (grey bars) or IgG control Ab (black bars). Note enrichment of GRHL2 binding at *OVOL2* exon and *CDH1* intron, but not at *ZEB1* promoter regions. For *ZEB1* detection, ChIP-qPCR was performed using primers that have been previously reported to amplify *ZEB1* promoter DNA sequences bound by GRHL2 in human mammary epithelial cells and in PEO1 but not OVCA429 human ovarian cancer cells (indicated by ##) [17, 23] and another primer set that did not confirm GRHL2 promoter interaction in ovarian cancer cells (indicated by #) [17]. Signals for IgG control and GRHL2 antibody pulldown samples were normalized to input DNA and presented as % input with SEM from 3 technical replicates. Data were statistically analyzed by t-test and \* indicates  $p < 0.05$ . **(c,e)** Bru-seq reads for indicated genes at indicated time point after to GRHL2 deletion. Track colors: green, sgCTR; red, sgGRHL2(1); blue, sgGRHL2(2). **(d,f)** Line graphs depicting the  $\log_2$  AFC of transcription in sgGRHL2(1) and sgGRHL2(2) cells for the indicated genes. The positive y-axis indicates the plus-strand signal of RNA synthesis from left to right and the negative y-axis represents the minus-strand signal of RNA synthesis from right to left.

No GRHL2 binding was observed at the promoter or other regions of *ZEB1* or *ZEB2* as opposed to findings in mammary epithelial cells [24] (**Fig. 6a**). ChIP-qPCR was performed using primers that have been previously reported to amplify *ZEB1* promoter DNA sequences bound by GRHL2 in human mammary epithelial cells and in PEO1 but not OVCA429 human ovarian cancer cells [17, 23] and another primer set that did not detect GRHL2 promoter interaction in ovarian cancer cells [17] (**Fig. 6b**). This confirmed the absence of GRHL2 binding in the promoter of *ZEB1* in luminal breast cancer cells. In agreement, no significant changes in transcription of *ZEB1* and *ZEB2* genes were observed after GRHL2 loss in MCF7 (data not shown). Together, these results indicated that *CLDN4* is a direct GRHL2 target while *CDH1*, *ZEB1*, or *ZEB2* are unlikely to represent direct GRHL2 target genes in luminal breast cancer cells. The latter may be regulated at later timepoints indirectly through other transcriptional regulators [66] or by GRHL2-mediated post-transcriptional modification [17, 23, 67].

## Validation of GRHL2 associations in breast cancer patients

All genes identified by Bru-seq in the MCF7 conditional GRHL2 KO model falling in the categories “sustained induction/repression” or “induction/repression reset”, were imported in the STRING database to visualize clusters representing enriched functionalities regulated by GRHL2. Three clusters of proteins associated with i) epigenetic regulation of gene expression (including proteins also connected to GO:0098532, histone H3-K27 trimethylation; not shown), ii) translation initiation, and iii) mitosis were clearly visible (**Fig. 7a**). This was in agreement with the growth suppression observed in response to GRHL2 depletion (**Fig. 4**) and earlier reports involving GRHL2 in histone methylation [17]. E2F1 and E2F2 were connected to the mitosis cluster but EHF showed no connections. No connections of these clusters with GRHL2 were visible but the interaction of GRHL2 with CLDN4 was shown as well as co-expression of GRHL2 with TACSTD2, a transmembrane receptor regulating cell proliferation and migration in development and cancer [68]. The TGFB1, TGFB2, and TGFB2 axis was not closely connected to GRHL2 but both were surrounded by genes encoding extracellular matrix (ECM) components (e.g., laminin subunits and collagen chains), the ITGB6 integrin subunit, and LOXL2 encoding an ECM crosslinking enzyme [69] pointing to modification of ECM production and adhesion.





**Fig. 7. Gene clusters responding to GRHL2 depletion and their correlation with GRHL2 in breast cancer tissues. (a)** STRING derived protein interaction analysis of genes displaying sustained up- or down regulation in response to GRHL2 depletion in MCF7. GO terms are color marked as indicated. *i*, entire network with boxes showing zoom-in on indicated regions; *ii*, zoom in on indicated region showing different GO terms. # Indicates GRHL2 targets identified by promoter binding. \* Indicates TGFβ signaling axis. **(b)** Average expression (log2 scale) in the MA-867 patient dataset of the cluster of genes negatively (left panel) or positively associated with GRHL2 (right panel) in MCF7 KO model. Patients were divided in 4 quartiles according to the level of GRHL2 expression. Q1, lowest GRHL2 expression; Q4, GRHL2 highest expression; All, all patients grouped together; ER+, ER

positive patients grouped; ER-, ER negative patients grouped. Boxplots display the median with 25th–75th percentile and dots represent lower 5% and upper 95% samples. P values determined by t test (two-sided). (c) Correlation with GRHL2 in MA-867 and METABRIC datasets for indicated genes negatively or positively correlated with GRHL2 in MCF KO model analyzed by Bru-seq. For MA-867 dataset, R-values for all patients grouped together, ER positive patients, or ER negative patients are shown. For METABRIC dataset, correlation, p-value, and q-values are shown as determined in BioPortal.

We addressed to what extent GRHL2 regulated gene clusters identified in our conditional MCF7 KO model predicted associations with GRHL2 gene expression in breast cancer patients. We focused on all genes where the control sgRNA gave  $0.75 < FC < 1.5$  at each time point after GRHL2 KO while both GRHL2 sgRNAs triggered either  $FC < 0.75$  in at least 3 time points (positive correlation with GRHL2) or  $FC > 1.5$  in at least 3 time points (negative correlation with GRHL2). We made use of a cohort of 867 untreated breast cancer patients (MA-867 dataset [59]) and ranked patients in 4 quartiles according to the level of GRHL2 expression. The average expression of predicted negatively correlated and positively correlated gene clusters based on the MCF7 conditional KO model, displayed a significant correlation with GRHL2 expression in the same direction when all patients were treated as one group (**Fig. 7b**). Moreover, behavior in the MCF7 conditional KO model correctly predicted the correlation of gene clusters with GRHL2 expression when ER positive and ER negative patients were separately tested.

At the individual gene level, Pearson correlation coefficients for association with GRHL2 when all patients of the MA-867 dataset were treated as one group, were in the range  $-0.31 < R < 0.29$  indicating that associations while in the same orientation were weak. This included *EHF* and *CLDN4* that were subject to promoter binding by GRHL2 and belonged to the positively correlated gene cluster (**Fig. 4c, 6a, 7c; Table S2**). We also analyzed the METABRIC data set consisting of RNAseq data of 1904 primary breast tumors. Here, co-expression analysis using cBioPortal showed a significant correlation in the

same direction as predicted by the MCF7 conditional KO model for *CLDN4* but not *EHF* (**Fig. 7c**). For the negatively correlated gene cluster, *TGFBR2* as well as *TGFB1* showed a significant correlation in the same direction in the METABRIC data set, further establishing suppression of TGF $\beta$  signaling by GRHL2 in breast cancer cells. Notably, the large majority of genes in both clusters did not harbor promoter binding sites, further indicating that regulation at enhancer sites or indirect mechanisms prevailed.

### Discussion

We report genome-wide binding sites of the transcription factor GRHL2 that are conserved across 3 human luminal breast cancer cell lines. The match with previously published binding motifs in other cell types shows conservation of GRHL2-DNA interaction but we find that the spectrum of GRHL2 targets differs considerably from those identified in other cells. A limited number of binding sites were located at gene promoter regions. Similar to previous reports [14, 17], most binding sites were located in introns and intergenic regions. Such regions may contain enhancers interacting with GRHL2 and GRHL2 has been reported to regulate histone modifications such as H3K4me3 and H3K4me1 [17, 70]. Notably, GRHL2 can regulate ER $\alpha$  signaling output in hormone receptor positive breast cancer by co-occupying enhancer elements with FOXA1, GATA3, and ER $\alpha$  [34-36]. Co-occupation of enhancers by ER $\alpha$  and GRHL2 has been shown to be regulated by ER $\alpha$  phosphorylation at Ser118 [57]. Indeed, we detect an ER $\alpha$  binding motif in the vicinity of GRHL2 peaks, but this represents only a minor fraction of the identified GRHL2 binding sites. Moreover, intersection of our identified GRHL2 peaks with published ER $\alpha$  binding events in the same series of luminal breast cancer cells cultured under the same conditions further indicates that GRHL2 binds most of the targets found by us in absence of ER $\alpha$ , FOXA1, and GATA3. A study intersecting binding sites for GRHL2, FOXA1, and ER $\alpha$  in MCF7 cells also found that most GRHL2 binding sites did not overlap with FOXA1 or ER $\alpha$  binding but ~30% did show overlap [71]. Our exclusive focus on GRHL2 binding sites that are conserved across three luminal breast cancer cell lines may have selected for those sites binding only GRHL2. Together, these studies indicate that enhancers occupied by ER $\alpha$ , FOXA1, and GATA3 frequently also

bind GRHL2, but a majority of conserved GRHL2 binding sites in luminal breast cancer cells do not overlap with binding of the ER $\alpha$  signaling complex.

Using a conditional KO model, we identify genes whose transcription is regulated by GRHL2 in luminal breast cancer cells. Notably, the gene clusters showing up- or downregulation in response to GRHL2 loss show a significant, albeit low level of correlation with GRHL2 expression in breast cancer patients. By using Bru-seq we focus on changes in the rate of nascent RNA synthesis caused by GRHL2 depletion [72]. Differences with studies using steady state RNA-seq may be due to post transcriptional mechanisms of regulation not addressed in our analysis, including RNA stability. We observed diverse responses to GRHL2 depletion, including enhanced or repressed transcription that can be sustained, transient or dynamic type of response. The fact that patterns of transcription induction are similar to the patterns of transcription repression is in line with the fact that GRHL2 has been reported to act as a positive as well as a negative regulator of gene transcription. However, indirect mechanisms involving other transcriptional activators or repressors may also be triggered by GRHL2 depletion.

GRHL2 expression appears to support cancer growth and even disease progression in most tumor types investigated [18-22, 37]. Indeed, GRHL2 drives expression of several genes promoting cell survival and proliferation [9, 18, 19]. Our study agrees with this as GRHL2 loss rapidly affects a cluster of genes involved in cell cycle progression and causes a gradual decrease in proliferation in MCF7 cells. A group of genes whose transcription is reduced following loss of GRHL2 is involved in cell cycle progression and DNA replication including the epithelial specific ETS family transcription factor *EHF*, E2F transcription factors *E2F1* and *E2F2* and other genes such as *CDCA7L* and *MCM2* [60-63]. We show that *EHF*, *E2F2* and *CDCA7L* represent previously unidentified GRHL2 target genes that can be subject to direct regulation at promotor regions. *EHF* has been previously implicated in ovarian, gastric and prostate cancer [73-75] but our findings point to cooperative roles of GRHL2 target genes including *EHF* and E2Fs in sustaining proliferation.

Several studies have shown that GRHL2 suppresses EMT [9, 17, 23, 24, 26, 27]. This may explain its reported role as a suppressor of local tissue invasion and metastasis [9, 25]. In fact, a similar function may also be involved in the many examples where GRHL2 is positively associated with tumor progression and metastasis. GRHL2 may prevent a complete EMT and maintain cancer cells in a hybrid EMT state that is believed to be crucial for cancer cell plasticity, which supports invasion and metastasis [76, 77]. Our results concerning GRHL2 interactions with known EMT-related genes are partly in disagreement with previously published findings. First, we demonstrate that *CDH1* RNA synthesis is not altered following GRHL2 loss, despite an intronic binding site that is conserved in the three luminal cell lines. No binding site is observed in the -1000/+100 promotor region but we detect GRHL2 binding in the region from -6000 bp to -1000 bp relative to the TSS of the *CDH1* gene, consistent with an earlier study reporting a contact of GRHL2 upstream of the *CDH1* promoter [4]. Although this may facilitate long-distance interactions with the promoter region through chromatin looping [4], loss of this interaction, nor that at the intronic GRHL2 binding site, causes a reduction in *CDH1* transcription in the first 16 days after GRHL2 deletion in our study. Our findings do not rule out *CDH1* regulation through indirect, post transcriptional mechanisms including RNA stability that are not measured in Bru-seq and may underlie findings in studies using RNA-seq or PCR analyses, or at the level of translation. Second, it has been reported that *ZEB1* is regulated by GRHL2 directly and, vice versa, that *ZEB1* regulates *GRHL2* in a balance between EMT and MET [9, 20, 23, 24]. We do not detect GRHL2 binding sites in the promoter, or other regions of the *ZEB1* or *ZEB2* genes. This potential discrepancy cannot be explained by technical differences as we have confirmed the lack of GRHL2 binding in the ChIP-seq analysis by ChIP-qPCR using primers that amplified *ZEB1* and *ZEB2* regions bound by GRHL2 in human mammary epithelial cells and human ovarian cancer cells in other studies [17, 23]. Rather, this may point to differences in GRHL2 interactions in different cell types. Nevertheless, the fact that we do not detect GRHL2-binding sites in *ZEB1* or *ZEB2* is in line with our Bru-seq analysis indicating that transcription of the *ZEB1* and *ZEB2* genes is not affected by GRHL2 depletion in the first 16 days. Together, this data indicates that *CDH1*, *ZEB1*, and *ZEB2* genes do not

represent direct transcriptional targets of GRHL2 in luminal breast cancer and their regulation may occur through post-transcriptional regulation in this cellular context. Our data do confirm *CLDN4* as a direct target gene with GRHL2-binding in the promoter region and transcriptional suppression in response to GRHL2 depletion in luminal breast cancer cells.

The fact that in our study GRHL2 supports gene networks involved in cell proliferation and that a tumor/metastasis suppressing function related to its suppression of EMT is less evident, agrees with earlier studies and with the location of GRHL2 on chromosome 8q22, a region that is amplified in various cancers, including breast cancer. One explanation for the discrepancy between different studies including our own is the possibility that GRHL2 interacts with- and regulates genes in a context-dependent manner. A meta-analysis combining all RNA-seq, micro-array, and ChIP-seq experiments, identified common candidate genes for regulation by GRH or GRHL1-3. The authors noticed a striking lack of correlation between findings in normal epithelia as compared to cancerous cells with *CDH1* being identified as a target in normal epithelia but not cancer [78]. Likewise, the findings reported in our study represent candidate GRHL2-regulated genes and pathways in luminal breast cancer that partly overlap but are also distinct from GRHL2 regulation in normal epithelia and other cancer types.

### Conclusions

Taken together, this study provides a comprehensive genome-wide resource of GRHL2 binding sites conserved across luminal breast cancer cells. In a conditional KO model, we identify groups of genes whose transcription is positively or negatively controlled by GRHL2 and find 5 main patterns of dynamic regulation. The association with GRHL2 of gene clusters in the KO model predicts the correlation with GRHL2 expression in breast cancer patients. The dominant response to GRHL2 depletion in luminal breast cancer cells is suppression of proliferation and we identify clusters of genes reflecting this response including direct regulation of ETS and E2F transcription factors by GRHL2. An EMT response to GRHL2 loss is limited and our findings indicate that regulation of epithelial genes can be strikingly different in normal and

## Chapter 3

---

cancer cells involving direct GRHL2-mediated transcriptional control or indirect mechanisms.

### Abbreviations

GRHL2, Grainyhead like 2

EMT, epithelial to mesenchymal transition

HER2, human epidermal growth factor receptor 2

ChIP-seq, chromatin immunoprecipitation followed by high-throughput DNA sequencing

qPCR, quantitative polymerase chain reaction

Bru-seq, bromouridine sequencing

RIPA buffer, radioimmunoprecipitation buffer

BCA, bicinchoninic acid

PVDF, polyvinylidene difluoride

BSA, bovine serum albumin

RT, room temperature

IPA, Ingenuity Pathways Analysis

DAVID, Database for Annotation, Visualization, and Integrated Discovery

GO, Gene Ontology

PANTHER, Protein Analysis Through Evolutionary Relationships

AFC, average fold change

PPI, protein-protein interaction

### Availability of data and materials:

Chip-seq data supporting the results of this article is available at the UCSC Genome Browser [<https://genome.ucsc.edu/s/hwuRadboudumc/ZWang>].

Bru-seq data supporting the results of this article is available at Gene Expression Omnibus (GEO) database, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (Accession No. GSE222353).

### Competing interests:

The authors declare that they have no competing interests. Zi Wang was supported by a grant from the China Scholarship Council. Bircan Coban was supported by the Dutch Cancer Society (KWF Research Grant #10967).

### Acknowledgements:

We thank Dr. Giuseppina Carbone, Institute of Oncology Research, Bellinzona, Switzerland, for kindly providing the EHF plasmid.

### Author contributions:

ZW designed, executed and analyzed the experiments, prepared figures, and wrote the manuscript; BC analyzed experimental and patient data, prepared figures, and wrote the manuscript, HW assisted with design and analysis of the ChIP-seq experiment and critically read the manuscript; JC analyzed the ChIP-seq experiment and critically read the manuscript LD assisted with design of the ChIP-seq experiment and critically read the manuscript; MTP executed processing and sequencing of Bru-seq samples and critically read the manuscript; ML assisted with design and analysis of the Bru-seq experiment and critically read the manuscript; MS analyzed patient data and critically read the manuscript; JM analyzed patient data and critically read the manuscript; EHJD initiated the study, designed the experiments, and wrote the manuscript.

### References

1. Frisch SM, Farris JC, Pifer PM: Roles of Grainyhead-like transcription factors in cancer. *Oncogene* 2017.
2. Wang S, Samakovlis C: Grainy head and its target genes in epithelial morphogenesis and wound healing. *Curr Top Dev Biol* 2012, 98:35-63.
3. Wilanowski T, Caddy J, Ting SB, Hislop NR, Cerruti L, Auden A, Zhao LL, Asquith S, Ellis S, Sinclair R *et al*: Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. *The EMBO journal* 2008, 27(6):886-897.
4. Werth M, Walentin K, Aue A, Schonheit J, Wuebken A, Pode-Shakked N, Vilianovitch L, Erdmann B, Dekel B, Bader M *et al*: The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* 2010, 137(22):3835-3845.
5. Pyrgaki C, Liu A, Niswander L: Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Developmental biology* 2011, 353(1):38-49.
6. Ting SB, Caddy J, Hislop N, Wilanowski T, Auden A, Zhao LL, Ellis S, Kaur P, Uchida Y, Holleran WM *et al*: A homolog of Drosophila grainy head is essential for epidermal integrity in mice. *Science* 2005, 308(5720):411-413.

7. Rifat Y, Parekh V, Wilanowski T, Hislop NR, Auden A, Ting SB, Cunningham JM, Jane SM: Regional neural tube closure defined by the Grainy head-like transcription factors. *Developmental biology* 2010, 345(2):237-245.
8. Boglev Y, Wilanowski T, Caddy J, Parekh V, Auden A, Darido C, Hislop NR, Cangkrama M, Ting SB, Jane SM: The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. *Developmental biology* 2011, 349(2):512-522.
9. Werner S, Frey S, Riethdorf S, Schulze C, Alawi M, Kling L, Vafaizadeh V, Sauter G, Terracciano L, Schumacher U *et al*: Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *The Journal of biological chemistry* 2013, 288(32):22993-23008.
10. Caddy J, Wilanowski T, Darido C, Dworkin S, Ting SB, Zhao Q, Rank G, Auden A, Srivastava S, Papenfuss TA *et al*: Epidermal wound repair is regulated by the planar cell polarity signaling pathway. *Developmental cell* 2010, 19(1):138-147.
11. Gao X, Vockley CM, Pauli F, Newberry KM, Xue Y, Randell SH, Reddy TE, Hogan BL: Evidence for multiple roles for grainyhead-like 2 in the establishment and maintenance of human mucociliary airway epithelium.[corrected]. *Proceedings of the National Academy of Sciences of the United States of America* 2013, 110(23):9356-9361.
12. Senga K, Mostov KE, Mitaka T, Miyajima A, Tanimizu N: Grainyhead-like 2 regulates epithelial morphogenesis by establishing functional tight junctions through the organization of a molecular network among claudin3, claudin4, and Rab25. *Molecular biology of the cell* 2012, 23(15):2845-2855.
13. Kohn KW, Zeeberg BM, Reinhold WC, Pommier Y: Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. *PloS one* 2014, 9(6):e99269.
14. Walentin K, Hinze C, Werth M, Haase N, Varma S, Morell R, Aue A, Potschke E, Warburton D, Qiu A *et al*: A Grhl2-dependent gene network controls trophoblast branching morphogenesis. *Development* 2015, 142(6):1125-1136.
15. Aue A, Hinze C, Walentin K, Ruffert J, Yurtdas Y, Werth M, Chen W, Rabien A, Kilic E, Schulzke JD *et al*: A Grainyhead-Like 2/Ovo-Like 2 Pathway Regulates Renal Epithelial Barrier Function and Lumen Expansion. *Journal of the American Society of Nephrology : JASN* 2015, 26(11):2704-2715.
16. Pifer PM, Farris JC, Thomas AL, Stoilov P, Denvir J, Smith DM, Frisch SM: Grainyhead-like 2 inhibits the coactivator p300, suppressing tubulogenesis and the epithelial-mesenchymal transition. *Molecular biology of the cell* 2016, 27(15):2479-2492.

17. Chung VY, Tan TZ, Tan M, Wong MK, Kuay KT, Yang Z, Ye J, Muller J, Koh CM, Guccione E *et al*: GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Scientific reports* 2016, 6:19943.
18. Dompe N, Rivers CS, Li L, Cordes S, Schwickart M, Punnoose EA, Amler L, Seshagiri S, Tang J, Modrusan Z *et al*: A whole-genome RNAi screen identifies an 8q22 gene cluster that inhibits death receptor-mediated apoptosis. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108(43):E943-951.
19. Chen W, Dong Q, Shin KH, Kim RH, Oh JE, Park NH, Kang MK: Grainyhead-like 2 enhances the human telomerase reverse transcriptase gene expression by inhibiting DNA methylation at the 5'-CpG island in normal human keratinocytes. *The Journal of biological chemistry* 2010, 285(52):40852-40863.
20. Quan Y, Jin R, Huang A, Zhao H, Feng B, Zang L, Zheng M: Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. *Cancer Biol Ther* 2014, 15(7):878-887.
21. Xiang X, Deng Z, Zhuang X, Ju S, Mu J, Jiang H, Zhang L, Yan J, Miller D, Zhang HG: Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PloS one* 2012, 7(12):e50781.
22. Yang X, Vasudevan P, Parekh V, Penev A, Cunningham JM: Bridging cancer biology with the clinic: relative expression of a GRHL2-mediated gene-set pair predicts breast cancer metastasis. *PloS one* 2013, 8(2):e56195.
23. Cieply B, Farris J, Denvir J, Ford HL, Frisch SM: Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer research* 2013, 73(20):6299-6309.
24. Cieply B, Riley Pt, Pifer PM, Widmeyer J, Addison JB, Ivanov AV, Denvir J, Frisch SM: Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer research* 2012, 72(9):2440-2453.
25. Xiang J, Fu X, Ran W, Chen X, Hang Z, Mao H, Wang Z: Expression and role of grainyhead-like 2 in gastric cancer. *Medical oncology* 2013, 30(4):714.
26. Brabletz S, Brabletz T: The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer? *EMBO reports* 2010, 11(9):670-677.
27. Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S, Morris M, Wyatt L, Farshid G, Lim YY *et al*: An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Molecular biology of the cell* 2011, 22(10):1686-1698.

28. Mlacki M, Kikulska A, Krzywinska E, Pawlak M, Wilanowski T: Recent discoveries concerning the involvement of transcription factors from the Grainyhead-like family in cancer. *Experimental biology and medicine* 2015, 240(11):1396-1401.
29. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98(19):10869-10874.
30. Cancer Genome Atlas N: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490(7418):61-70.
31. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: Molecular portraits of human breast tumours. *Nature* 2000, 406(6797):747-752.
32. Forouzanfar MH, Foreman KJ, Delossantos AM, Lozano R, Lopez AD, Murray CJ, Naghavi M: Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* 2011, 378(9801):1461-1484.
33. Paltoglou S, Das R, Townley SL, Hickey TE, Tarulli GA, Coutinho I, Fernandes R, Hanson AR, Denis I, Carroll JS *et al*: Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer. *Cancer research* 2017, 77(13):3417-3430.
34. Cocce KJ, Jasper JS, Desautels TK, Everett L, Wardell S, Westerling T, Baldi R, Wright TM, Tavares K, Yllanes A *et al*: The Lineage Determining Factor GRHL2 Collaborates with FOXA1 to Establish a Targetable Pathway in Endocrine Therapy-Resistant Breast Cancer. *Cell Rep* 2019, 29(4):889-903 e810.
35. Holding AN, Giorgi FM, Donnelly A, Cullen AE, Nagarajan S, Selth LA, Markowitz F: VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERα at enhancers in breast cancer. *Genome biology* 2019, 20(1):91.
36. Chi D, Singhal H, Li L, Xiao T, Liu W, Pun M, Jeselsohn R, He H, Lim E, Vadhi R *et al*: Estrogen receptor signaling is reprogrammed during breast tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 2019, 116(23):11437-11443.
37. Reese RM, Harrison MM, Alarid ET: Grainyhead-like Protein 2: The Emerging Role in Hormone-Dependent Cancers and Epigenetics. *Endocrinology* 2019, 160(5):1275-1288.
38. Tugores A, Le J, Sorokina I, Snijders AJ, Duyao M, Reddy PS, Carlee L, Ronshaugen M, Mushegian A, Watanaskul T *et al*: The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor

downstream of MAPK signaling cascades. *The Journal of biological chemistry* 2001, 276(23):20397-20406.

39. Cangemi R, Mensah A, Albertini V, Jain A, Mello-Grand M, Chiorino G, Catapano CV, Carbone GM: Reduced expression and tumor suppressor function of the ETS transcription factor ESE-3 in prostate cancer. *Oncogene* 2008, 27(20):2877-2885.

40. Lin X, Tirichine L, Bowler C: Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. *Plant Methods* 2012, 8(1):48.

41. Liu CM, Wong T, Wu E, Luo R, Yiu SM, Li Y, Wang B, Yu C, Chu X, Zhao K *et al*: SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics* 2012, 28(6):878-879.

42. Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998, 8(3):175-185.

43. Liao P, Satten GA, Hu YJ: PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genet Epidemiol* 2017, 41(5):375-387.

44. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al*: Model-based analysis of ChIP-Seq (MACS). *Genome biology* 2008, 9(9):R137.

45. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010, 38(4):576-589.

46. Yu G, Wang LG, He QY: ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015, 31(14):2382-2383.

47. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Perez N *et al*: JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2022, 50(D1):D165-D173.

48. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O, Ellis IO, Green AR *et al*: Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012, 481(7381):389-393.

49. Michaloglou C, Crafter C, Siersbaek R, Delpuech O, Curwen JO, Carnevalli LS, Staniszevska AD, Polanska UM, Cheraghchi-Bashi A, Lawson M *et al*: Combined Inhibition of mTOR and CDK4/6 Is Required for Optimal Blockade

of E2F Function and Long-term Growth Inhibition in Estrogen Receptor-positive Breast Cancer. *Mol Cancer Ther* 2018, 17(5):908-920.

50. Lai CF, Flach KD, Alexi X, Fox SP, Ottaviani S, Thiruchelvam PT, Kyle FJ, Thomas RS, Launchbury R, Hua H *et al*: Co-regulated gene expression by oestrogen receptor alpha and liver receptor homolog-1 is a feature of the oestrogen response in breast cancer cells. *Nucleic Acids Res* 2013, 41(22):10228-10240.

51. Paulsen MT, Veloso A, Prasad J, Bedi K, Ljungman EA, Magnuson B, Wilson TE, Ljungman M: Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* 2014, 67(1):45-54.

52. Paulsen MT, Veloso A, Prasad J, Bedi K, Ljungman EA, Tsan YC, Chang CW, Tarrier B, Washburn JG, Lyons R *et al*: Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proceedings of the National Academy of Sciences of the United States of America* 2013, 110(6):2240-2245.

53. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P *et al*: The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021, 49(D1):D605-D612.

54. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486(7403):346-352.

55. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ *et al*: The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* 2016, 7:11479.

56. Gao X, Vockley CM, Pauli F, Newberry KM, Xue Y, Randell SH, Reddy TE, Hogan BL: Evidence for multiple roles for grainyhead-like 2 in the establishment and maintenance of human mucociliary airway epithelium. *Proceedings of the National Academy of Sciences* 2013, 110(23):9356-9361.

57. Helzer KT, Szatkowski Ozers M, Meyer MB, Benkusky NA, Solodin N, Reese RM, Warren CL, Pike JW, Alarid ET: The Phosphorylated Estrogen Receptor alpha (ER) Cistrome Identifies a Subset of Active Enhancers Enriched for Direct ER-DNA Binding and the Transcription Factor GRHL2. *Mol Cell Biol* 2019, 39(3).

58. Kas K, Finger E, Grall F, Gu X, Akbarali Y, Boltax J, Weiss A, Oettgen P, Kapeller R, Libermann TA: ESE-3, a novel member of an epithelium-specific

ets transcription factor subfamily, demonstrates different target gene specificity from ESE-1. *The Journal of biological chemistry* 2000, 275(4):2986-2998.

59. Koedoot E, Wolters L, Smid M, Stoilov P, Burger GA, Herpers B, Yan K, Price LS, Martens JWM, Le Devedec SE *et al*: Differential reprogramming of breast cancer subtypes in 3D cultures and implications for sensitivity to targeted therapy. *Scientific reports* 2021, 11(1):7259.

60. Luk IY, Reehorst CM, Mariadason JM: ELF3, ELF5, EHF and SPDEF Transcription Factors in Tissue Homeostasis and Cancer. *Molecules* 2018, 23(9).

61. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC: The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer* 2017, 17(6):337-351.

62. Labib K, Tercero JA, Diffley JF: Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* 2000, 288(5471):1643-1647.

63. Ji QK, Ma JW, Liu RH, Li XS, Shen FZ, Huang LY, Hui L, Ma YJ, Jin BZ: CDCA7L promotes glioma proliferation by targeting CCND1 and predicts an unfavorable prognosis. *Mol Med Rep* 2019, 20(2):1149-1156.

64. De Craene B, Berx G: Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* 2013, 13(2):97-110.

65. Varma S, Cao Y, Tagne JB, Lakshminarayanan M, Li J, Friedman TB, Morell RJ, Warburton D, Kotton DN, Ramirez MI: The transcription factors Grainyhead-like 2 and NK2-homeobox 1 form a regulatory loop that coordinates lung epithelial cell morphogenesis and differentiation. *The Journal of biological chemistry* 2012, 287(44):37282-37295.

66. Goossens S, Vandamme N, Van Vlierberghe P, Berx G: EMT transcription factors in cancer development re-evaluated: Beyond EMT and MET. *Biochim Biophys Acta Rev Cancer* 2017, 1868(2):584-591.

67. Park SM, Gaur AB, Lengyel E, Peter ME: The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 2008, 22(7):894-907.

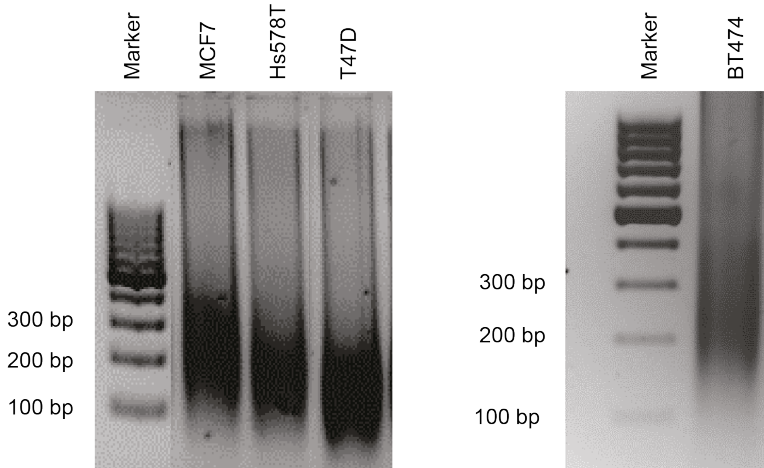
68. McDougall AR, Tolcos M, Hooper SB, Cole TJ, Wallace MJ: Trop2: from development to disease. *Dev Dyn* 2015, 244(2):99-109.

69. Liburkin-Dan T, Toledano S, Neufeld G: Lysyl Oxidase Family Enzymes and Their Role in Tumor Progression. *Int J Mol Sci* 2022, 23(11).

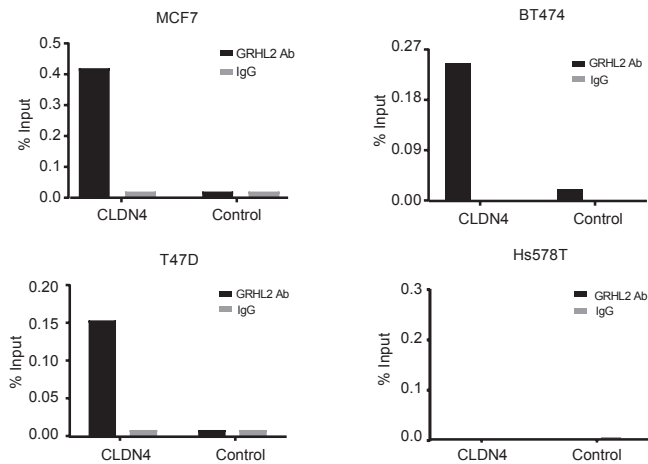
70. Chung VY, Tan TZ, Ye J, Huang RL, Lai HC, Kappei D, Wollmann H, Guccione E, Huang RY: The role of GRHL2 and epigenetic remodeling in epithelial-mesenchymal plasticity in ovarian cancer cells. *Communications biology* 2019, 2:272.

71. Jozwik KM, Chernukhin I, Serandour AA, Nagarajan S, Carroll JS: FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell Rep* 2016, 17(10):2715-2723.
72. Kirkconnell KS, Paulsen MT, Magnuson B, Bedi K, Ljungman M: Capturing the dynamic nascent transcriptome during acute cellular responses: The serum response. *Biology open* 2016, 5(6):837-847.
73. Shi J, Qu Y, Li X, Sui F, Yao D, Yang Q, Shi B, Ji M, Hou P: Increased expression of EHF via gene amplification contributes to the activation of HER family signaling and associates with poor survival in gastric cancer. *Cell Death Dis* 2016, 7(10):e2442.
74. Cheng Z, Guo J, Chen L, Luo N, Yang W, Qu X: Knockdown of EHF inhibited the proliferation, invasion and tumorigenesis of ovarian cancer cells. *Mol Carcinog* 2016, 55(6):1048-1059.
75. Albino D, Civenni G, Rossi S, Mitra A, Catapano CV, Carbone GM: The ETS factor ESE3/EHF represses IL-6 preventing STAT3 activation and expansion of the prostate cancer stem-like compartment. *Oncotarget* 2016, 7(47):76756-76768.
76. Pastushenko I, Blanpain C: EMT Transition States during Tumor Progression and Metastasis. *Trends Cell Biol* 2019, 29(3):212-226.
77. Coban B, Bergonzini C, Zweemer AJM, Danen EHJ: Metastasis: crosstalk between tissue mechanics and tumour cell plasticity. *Br J Cancer* 2021, 124(1):49-57.
78. Mathiyalagan N, Miles LB, Anderson PJ, Wilanowski T, Grills BL, McDonald SJ, Keightley MC, Charzynska A, Dabrowski M, Dworkin S: Meta-Analysis of Grainyhead-Like Dependent Transcriptional Networks: A Roadmap for Identifying Novel Conserved Genetic Pathways. *Genes* 2019, 10(11).

## Supplementary Figures and Tables

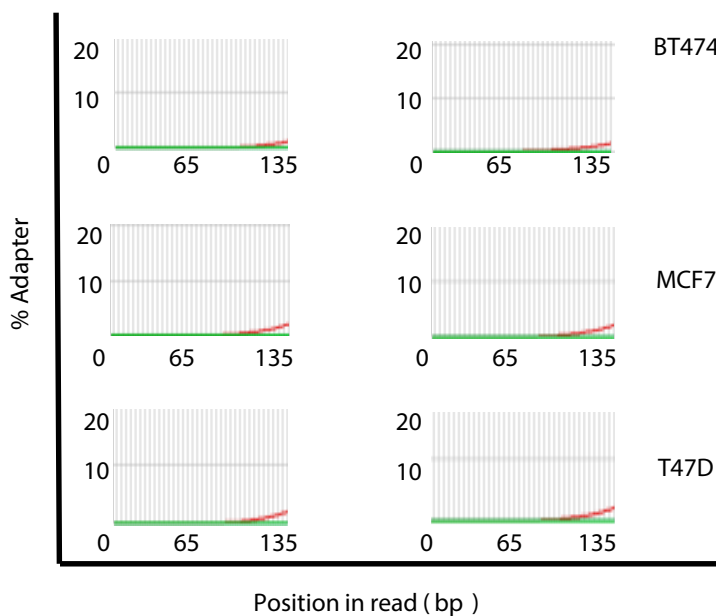


**Fig. S1. DNA fragmentation analysis by agarose gel electrophoresis.** After sonication, indicated samples were purified and loaded on 2% agarose gel.

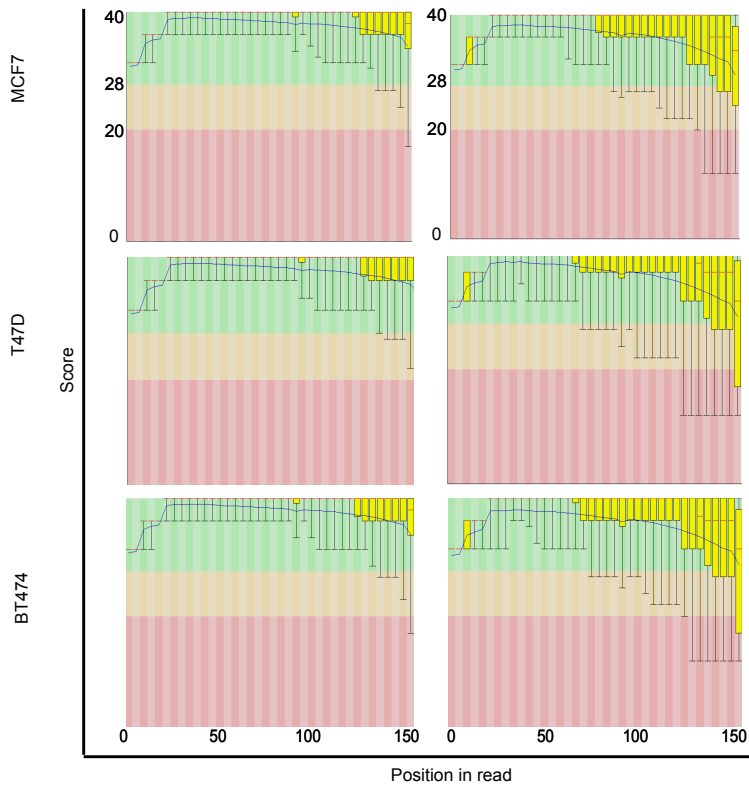


**Fig. S2. ChIP-qPCR validation of the isolated genomic DNA fragments.** Graphs represent the efficiency of *CLDN4* genomic DNA co-precipitation with anti-GRHL2

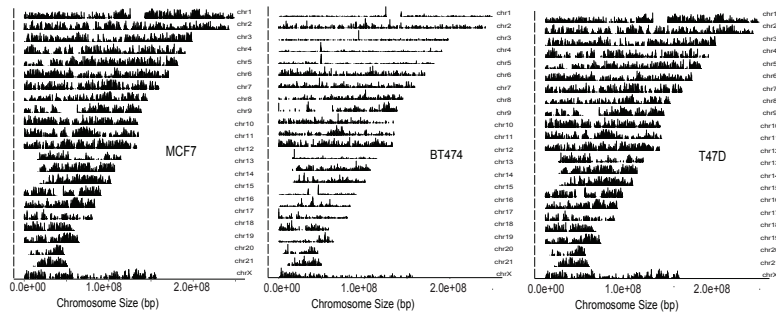
Ab (black bars) or IgG control Ab (grey bars). Detection was performed by qPCR using primers targeting the promoter region of *CLDN4* or targeting the intergenic region upstream of the *GAPDH* locus (Control). Results are shown for 3 GRHL2-positive luminal cell lines (MCF7, BT474 and T47D) and 1 GRHL2-negative basal-B cell line (Hs578T).



**Fig. S3. Cumulative presence of adapter sequences.** Results show that cumulative presence of adapter sequences is less than 5% in each cell sample, indicating that the data sets could be further analyzed without adapter-trimming.



**Fig. S4. Per base sequence quality for all sequencing data sets.** Y axis is divided into high quality calls (green), reasonable quality calls (orange) and poor-quality calls (red). Analysis shows that the mean quality of base calls, indicated by the blue line, consistently remained in the green area, indicating that sequencing data sets were of high quality.



**Fig. S5. Coverage of peak regions across chromosomes.** Graphs represent the coverage of GRHL2 binding sites across all chromosomes in the indicated cell lines.

## GRHL2-controlled gene expression networks

---

ADAT3	ERBB3	LOC731157	SCOC
ADGRF4	ERP27	LRP10	SCOC-AS1
ADIPOQ	FAF1	MACC1	SEMA4A
ADK	FKBP2	MACROD1	SFTPA2
AIFM1	FLJ31356	<b>MAPK10</b>	SHPK
AIMP1	FMN1	MCEE	SLC10A5
ALDH3B2	FMO9P	MESP1	SLC25A45
AMD1	FOXA1	MGP	SLC40A1
ANKRD22	FRRS1	MIR4328	SLC41A3
ANXA9	GAR1	MIR4513	SLC4A7
ARHGAP24	GGTLC1	MIR4676	SLC9A1
ARHGAP32	GINS2	MIR6070	SLFN12
ARHGEF19	GMDS-AS1	MIR6773	SLITRK6
ARHGEF38	GMEB1	MIR6784	SMG8
ARRDC3	GMPR2	MIR8072	SNORA38
ARSD	GPNMB	MTERF2	SNORD13
ASCL2	GPR108	MUCL1	SORT1
ATAD3B	GRAMD1C	NAALADL2	SSR4P1
ATP5S	GRAMD3	NEBL-AS1	ST3GAL4
ATP6V0A4	HIST2H2AB	NEU1	STX17-AS1
B4GAT1	HIST2H2BF	NFATC4	STX19
BATF	HRH1	NIPAL2	SYTL5
BBOX1	IFRD1	NIPSNAP1	TBL1X
BCAS1	IGSF9	NME7	TGIF1
BMF	IKZF2	NXT1	TGM1
C1orf116	IQCK	OR7E91P	TIGAR
C4orf3	ITFG2	OVOL2	TJP2
CARD14	IVL	P2RY6	TMEM40
CBLB	JADE1	PAN2	TMEM79
CCDC12	JUP	PDCD2	TMPRSS11F
CD46	KCNJ13	PDE4D	<b>TMPRSS13</b>
CDC42SE1	KLK12	PDGFB	TP53INP2
CDS1	KRT80	PGAP3	TRIL
CFAP45	KRTAP3-1	<b>PGLYRP2</b>	TRPC4AP
CHD3	LACE1	PGR	TUFT1

## Chapter 3

CLCN3	LIMA1	PIK3C2G	TYSND1
<b>CLDN4</b>	LINC00346	PIM2	UBALD2
CLDN8	LINC00359	PKP2	UBE2A
CMTR2	LINC00437	PLA2G4B	URB1-AS1
CNP	LINC00456	PPEF1	VEPH1
COMT	<b>LINC00885</b>	PPOX	VGLL1
CREB3L4	LINC00938	PRIM2	VIPAS39
CRISP3	LINC01213	PROM2	WSB2
CSE1L-AS1	LINC01405	PRR15L	YAP1
CSTF1	LOC100129917	PSCA	<b>ZBTB20</b>
DAZAP1	LOC100132781	PTPN14	ZER1
DLG4	LOC100506207	PURG	ZMYND8
DLX5	LOC100506804	RAB25	ZNF20
DNAAF5	LOC100507175	RAP2B	ZNF274
DNAJC5B	LOC101927272	RASAL2	ZNF433
DSCAM-AS1	LOC101927296	RBBP8NL	ZNF44
EDEM2	LOC101927318	RBL2	ZNF440
EEA1	LOC101927391	RBM47	ZNF443
EEF1E1	LOC101927755	RIMS1	ZNF567
<b>EHF</b>	LOC101927911	RNF32	ZNF799
EIF2B5	LOC101929441	RNU5B-1	ZNF823
ELF5	LOC101929718	RNVU1-14	ZNHIT6
EPB41L1	LOC102724163	ROCK1P1	ZP1
EPHA1	LOC148709	RPL32P3	
<b>EPN3</b>	LOC344967	<b>RPL41</b>	

**Table. S1. Candidate GRHL2 target genes in luminal breast cancer cells displaying promoter interaction.** GRHL2 promotor interactions identified by ChIP-seq in 3 luminal breast cancer cell lines are listed. Genes also identified by Bru-seq in MCF7 conditional KO model showing up- or downregulation at one or more timepoints in response to GRHL2 loss are indicated in bold.

## GRHL2-controlled gene expression networks

Gene	Fold change				Cluster
	D2 AFC	D4 AFC	D8 AFC	D16 AFC	
ABCA4	0,93	2,23	5,48	3,87	
AC005821.1	2,79	6,46	12,24	6,54	Sustained induction
AC005972.4	3,53	4,29	6,38	2,59	Sustained induction
AC007952.4	0,17	0,19	0,32	0,61	Repression reset
AC008703.1	5,56	6,60	7,51	3,23	Sustained induction
AC009262.1	3,58	4,70	5,98	3,37	Sustained induction
AC010653.3	0,21	0,48	0,55	0,67	
AC013652.1	2,43	3,59	4,10	1,57	Induction reset
AC019209.3	1,43	3,61	4,91	4,29	
AC022166.1	0,00	0,00	0,00	35,23	
AC027277.2	0,26	0,23	0,38	0,68	Repression reset
AC027288.3	2,30	3,95	3,23	0,94	Induction reset
AC051619.5	5,12	4,51	3,69	2,64	Sustained induction
AC055854.1	0,50	0,32	0,27	0,45	Sustained repression
AC068633.1	7,14	0,00	0,00	12,59	Dynamic
AC083967.1	0,52	0,32	0,37	0,40	
AC084880.1	0,30	0,25	0,46	0,62	Repression reset
AC087762.1	4,99	13,04	20,78	6,49	Sustained induction
AC092167.1	8,40	7,04	5,45	2,93	Sustained induction
AC092422.1	85,82	0,02	0,04	63,37	Dynamic
AC098934.1	0,22	0,17	0,25	0,59	Repression reset
AC099520.1	2,93	4,54	5,11	2,30	Sustained induction
AC099753.1	91,52	0,00	0,00	211,15	Dynamic
AC103770.1	2,34	3,65	3,32	3,27	Sustained induction
AC109326.1	0,29	0,27	0,37	0,70	Repression reset
AC245014.3	0,26	0,23	0,32	0,56	Repression reset
ACKR3	0,70	0,61	1,36	2,68	
ACOXL	4,98	9,44	26,98	10,20	Sustained induction
ACTB	0,34	0,44	0,42	0,51	Repression reset
ACTG1	0,34	0,55	0,62	0,67	
ADCY5	1,23	2,33	4,74	6,93	
ADGRE3	3,38	0,05	0,08	11,85	Dynamic

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
AFF3	1,89	3,18	3,47	2,41	
AGPAT4	4,16	8,04	15,78	7,77	Sustained induction
AL049839.2	2,22	4,43	9,16	6,73	Sustained induction
AL132708.1	2,08	1,97	3,83	2,12	
AL137003.2	3,61	5,82	6,54	2,45	Sustained induction
AL137145.2	3,01	3,87	5,79	3,08	Sustained induction
AL139383.1	2,80	3,34	2,39	1,15	Induction reset
AL158066.1	0,53	0,14	0,15	0,29	
AL158847.1	1,50	2,01	3,17	3,33	
AL354740.1	2,61	4,43	4,57	2,31	Sustained induction
AL359976.1	12,40	13,35	31,68	3,11	Sustained induction
AL390726.6	8,20	8,75	9,71	4,52	Sustained induction
AL590004.4	3,58	6,52	15,14	6,33	Sustained induction
ALDH1A3	1,89	6,70	12,97	5,21	
ALDOA	0,33	0,33	0,48	0,71	Repression reset
ALOX5	3,48	5,47	10,74	8,74	Sustained induction
AMPH	1,75	2,73	6,38	4,16	
<b>ANKRD1</b>	1,39	4,68	5,06	1,90	
ANKRD29	3,32	6,49	8,05	6,63	Sustained induction
ANOS1	1,93	3,05	3,15	2,13	
ANXA3	5,19	8,86	6,80	2,41	Sustained induction
AP000880.1	0,16	0,12	0,12	0,41	Sustained repression
AP000924.1	1,55	4,57	11,84	8,49	
AP002761.4	0,23	0,26	0,40	1,11	
APRT	0,26	0,38	0,41	0,84	Repression reset
ARHGAP18	2,78	4,04	3,91	1,80	Induction reset
ARHGAP22	1,88	4,43	7,13	3,45	
ARHGAP42	2,55	4,45	4,23	1,84	Induction reset
ARHGEF28	0,68	0,57	0,44	0,42	
ARHGEF39	0,36	0,30	0,29	0,64	Repression reset
ARPC1A	0,17	0,25	0,29	0,21	Sustained repression
ARSJ	4,10	11,15	10,85	2,69	Sustained induction

## GRHL2-controlled gene expression networks

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
ATP10D	3,42	10,07	11,67	5,55	Sustained induction
ATP50	0,26	0,37	0,33	0,29	Sustained repression
ATP8A2	0,23	0,24	0,61	0,64	
ATXN1	2,17	3,38	3,79	2,18	Sustained induction
AURKB	0,32	0,20	0,26	0,50	Repression reset
BBC3	1,47	1,98	2,07	3,27	
BIRC5	0,38	0,25	0,25	0,49	Sustained repression
BMP1	1,99	3,51	5,20	3,49	
BOC	1,73	2,50	3,81	1,98	
C14orf80	0,28	0,34	0,36	1,28	
C1orf105	2,83	3,30	7,20	2,74	Sustained induction
C21orf58	0,42	0,27	0,32	0,73	Repression reset
C22orf34	19,17	0,00	0,04	33,04	Dynamic
<b>CADM1</b>	1,89	2,93	3,51	1,49	
<b>CADPS</b>	53,78	0,06	0,03	51,71	Dynamic
CAMK1D	1,92	3,14	3,93	1,98	
CAPN8	4,32	5,39	11,47	5,66	Sustained induction
CBX2	0,30	0,33	0,47	0,89	Repression reset
CCNF	0,38	0,28	0,31	0,72	Repression reset
CD109	2,27	2,56	4,55	2,54	Sustained induction
CDC20	0,27	0,30	0,30	0,67	Repression reset
CDCA3	0,33	0,26	0,27	0,51	Repression reset
CDCA5	0,36	0,25	0,29	0,64	Repression reset
<b>CDCA7L</b>	0,58	0,43	0,36	0,39	
CDH18	3,27	6,40	7,10	5,15	Sustained induction
CDKN2B	2,93	9,08	13,19	6,17	Sustained induction
CELF3	0,27	0,14	0,37	1,14	
CEMIP	2,07	1,74	4,07	3,24	
CENPF	0,50	0,29	0,28	0,40	Sustained repression
CFL1	0,32	0,42	0,45	0,63	Repression reset
CHTF18	0,33	0,34	0,43	1,12	
<b>CLDN4</b>	0,40	0,29	0,52	1,06	

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
CNTN4	22,17	0,05	5,10	26,44	Dynamic
COL4A5	3,79	7,76	10,60	5,61	Sustained induction
COL9A2	0,26	0,12	0,25	0,61	Repression reset
COLQ	2,50	2,83	3,71	1,81	Induction reset
<b>CORO2A</b>	1,50	2,23	2,49	2,76	
CPNE4	3,77	3,01	6,43	8,86	Sustained induction
CPQ	2,74	4,86	4,35	2,33	Sustained induction
CPXM2	0,40	0,27	0,29	0,62	Repression reset
CREB5	2,58	8,80	12,08	5,22	Sustained induction
CTNNA3	9,27	16,04	14,13	4,17	Sustained induction
CTNND2	2,53	3,46	3,60	1,57	Induction reset
CYB561	0,35	0,46	0,56	0,92	
CYC1	0,27	0,34	0,43	0,95	Repression reset
<b>DAPP1</b>	4,40	9,94	23,48	11,76	Sustained induction
<b>DDX11</b>	0,37	0,37	0,41	0,67	Repression reset
<b>DDX12P</b>	0,33	0,27	0,30	0,58	Repression reset
DDX41	0,33	0,48	0,52	0,77	
<b>DDX58</b>	2,25	2,17	3,26	4,00	Sustained induction
DDX60L	2,17	3,23	3,85	3,82	Sustained induction
DISC1	2,25	3,20	3,52	1,43	Induction reset
DLGAP2	20,26	0,01	0,02	24,32	Dynamic
DNAH5	3,07	4,84	6,54	3,08	Sustained induction
DNAH7	3,11	3,96	4,70	1,87	Induction reset
DNM3	1,78	2,03	3,69	1,55	
DOCK4	2,36	3,98	5,25	1,92	Induction reset
DOCK8	2,24	3,15	3,16	2,68	Sustained induction
DOK5	24,10	0,01	0,03	18,16	Dynamic
<b>DUSP10</b>	2,30	2,61	4,07	2,72	Sustained induction
E2F1	0,35	0,33	0,38	0,80	Repression reset
<b>E2F2</b>	0,31	0,28	0,33	0,77	Repression reset
EDA2R	3,81	3,40	2,15	1,75	Induction reset
EEF1A1	0,36	0,45	0,50	0,56	Repression reset

## GRHL2-controlled gene expression networks

Gene	Fold change				Cluster
	D2 AFC	D4 AFC	D8 AFC	D16 AFC	
EEF2	0,31	0,47	0,51	0,75	
EFNB2	1,18	1,44	2,17	3,08	
EHF	0,21	0,14	0,15	0,30	Sustained repression
ELL2	2,04	2,30	3,50	1,47	Induction reset
EPAS1	2,35	3,41	6,43	3,06	Sustained induction
EPB41L4A	2,23	3,14	3,26	2,28	Sustained induction
EPN3	0,22	0,30	0,46	1,10	
ERC2	2,15	4,80	8,43	8,33	Sustained induction
ESPL1	0,36	0,30	0,31	0,70	Repression reset
F2R	2,29	6,41	8,42	5,27	Sustained induction
FAM13A	2,62	2,74	5,96	3,59	Sustained induction
FAM83D	0,40	0,30	0,29	0,55	Repression reset
FANCG	0,29	0,30	0,37	0,72	Repression reset
FAU	0,26	0,29	0,33	0,38	Sustained repression
FBN2	1,94	2,24	5,55	2,60	
FBXL2	2,16	3,08	3,13	1,79	Induction reset
FEN1	0,30	0,26	0,30	0,50	Repression reset
FGF12	2,39	4,68	3,75	1,65	Induction reset
FHL2	2,47	3,00	4,81	3,10	Sustained induction
FLT1	5,25	7,58	9,04	3,34	Sustained induction
FLT3	3,73	5,93	5,81	3,09	Sustained induction
FOXP2	3,99	4,09	3,12	1,65	Induction reset
FRY	4,41	7,10	10,36	4,92	Sustained induction
FSTL4	2,31	4,03	7,89	4,83	Sustained induction
FTL	0,20	0,27	0,35	0,42	Sustained repression
FYN	1,58	3,13	4,03	2,52	
GALNT17	7,35	0,03	0,03	6,59	Dynamic
GAPDH	0,19	0,30	0,38	0,53	Repression reset
GBP2	9,04	5,34	5,68	3,34	Sustained induction
GLDN	2,12	2,36	3,32	1,39	Induction reset
GPR155	2,74	4,26	3,92	2,03	Sustained induction
GPR87	3,32	3,32	6,20	3,14	Sustained induction

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
GRK5	2,15	3,92	3,74	2,42	Sustained induction
GULP1	2,50	4,50	6,29	3,19	Sustained induction
H2AFZ	0,17	0,20	0,24	0,36	Sustained repression
<b>HAX1</b>	0,28	0,38	0,44	0,61	Repression reset
HDX	3,74	6,57	15,96	7,88	Sustained induction
HERC3	1,59	2,96	4,04	1,55	
HIST1H1C	0,15	0,10	0,14	0,44	Sustained repression
HIST1H1D	0,14	0,09	0,14	0,50	Sustained repression
HIST1H1E	0,12	0,11	0,14	0,37	Sustained repression
HIST1H2AB	0,16	0,08	0,12	0,47	Sustained repression
HIST1H2AE	0,20	0,10	0,15	0,40	Sustained repression
HIST1H2AI	0,12	0,08	0,11	0,38	Sustained repression
HIST1H2AJ	0,09	0,08	0,09	0,33	Sustained repression
HIST1H2AL	0,25	0,10	0,11	1,18	
HIST1H2AM	0,14	0,08	0,11	0,51	Repression reset
HIST1H2APS4	0,40	0,14	0,23	0,54	Repression reset
HIST1H2BF	0,24	0,13	0,19	0,46	Sustained repression
HIST1H2BG	0,27	0,21	0,31	0,72	Repression reset
HIST1H2BH	0,26	0,15	0,21	0,61	Repression reset
HIST1H2BI	0,13	0,10	0,11	0,31	Sustained repression
HIST1H2BK	0,14	0,10	0,13	0,30	Sustained repression
HIST1H2BM	0,16	0,10	0,11	0,29	Sustained repression
HIST1H2BO	0,14	0,09	0,14	0,41	Sustained repression
HIST1H3A	0,14	0,06	0,13	0,57	Repression reset
HIST1H3G	0,17	0,10	0,15	0,55	Repression reset
HIST1H3H	0,24	0,16	0,25	0,66	Repression reset
HIST1H3I	0,35	0,23	0,12	1,52	
HIST1H3J	0,16	0,09	0,11	0,60	Repression reset
HIST1H4A	0,12	0,12	0,11	0,49	Sustained repression
<b>HIST1H4B</b>	0,20	0,11	0,17	0,52	Repression reset
HIST1H4D	0,19	0,11	0,17	0,40	Sustained repression
<b>HIST1H4E</b>	0,19	0,17	0,25	0,54	Repression reset

## GRHL2-controlled gene expression networks

Gene	Fold change				Cluster
	D2 AFC	D4 AFC	D8 AFC	D16 AFC	
HIST1H4H	0,36	0,26	0,38	0,58	Repression reset
HIST1H4J	0,21	0,08	0,19	0,53	Repression reset
HIST2H2BE	0,39	0,27	0,34	0,66	Repression reset
HIST2H3D	0,19	0,17	0,16	2,15	Dynamic
HIST4H4	0,28	0,23	0,34	0,77	Repression reset
HJURP	0,36	0,28	0,28	0,53	Repression reset
HLA-DQB1	3,02	3,67	8,50	3,86	Sustained induction
<b>HMGB2</b>	0,25	0,23	0,29	0,46	Sustained repression
<b>HMMR</b>	0,54	0,32	0,25	0,39	
HR	0,22	0,25	0,29	0,82	Repression reset
HSD17B11	4,28	5,50	13,27	3,41	Sustained induction
HSP90AA1	0,19	0,41	0,33	0,24	Sustained repression
HSP90AB1	0,33	0,57	0,57	0,59	
HSPA8	0,26	0,35	0,37	0,41	Sustained repression
HSPE1	0,31	0,36	0,28	0,22	Sustained repression
IGSF21	3,81	0,01	0,01	7,50	Dynamic
IL18	1,84	4,63	8,54	3,08	
INCENP	0,30	0,25	0,28	0,58	Repression reset
IQCJ-SCHIP1	2,79	3,41	5,34	1,83	Induction reset
ISM1	1,75	3,28	4,45	1,57	
ITGB6	3,58	7,80	41,07	26,91	Sustained induction
JAZF1	2,15	5,12	5,89	2,88	Sustained induction
KC6	6,12	8,94	12,80	6,63	Sustained induction
KCNJ3	2,03	3,56	6,40	4,08	Sustained induction
KCNK5	0,32	0,17	0,19	0,63	Repression reset
KCNMA1	1,33	1,96	4,48	4,79	
<b>KIAA0513</b>	2,04	3,14	3,92	2,80	Sustained induction
KIAA2012	3,22	9,67	16,82	6,56	Sustained induction
KIF20A	0,26	0,23	0,20	0,37	Sustained repression
KIF2C	0,41	0,25	0,28	0,48	Sustained repression
<b>KIF5C</b>	1,26	2,82	3,91	3,14	
KIFC1	0,46	0,29	0,29	0,55	Repression reset

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
LAD1	0,32	0,41	0,56	1,13	
LAMA3	2,00	3,34	3,66	2,74	
<b>LAMB3</b>	2,80	4,79	12,02	8,49	Sustained induction
LAMC2	2,16	5,02	9,71	4,78	Sustained induction
LHFPL2	1,33	2,47	3,55	2,16	
<b>LIMCH1</b>	1,86	2,56	4,45	2,02	
LINC00473	7,90	6,05	6,28	2,90	Sustained induction
LINC00871	6,51	4,62	1,99	2,96	
<b>LINC00885</b>	0,30	0,17	0,19	0,42	Sustained repression
LINC01191	2,10	5,99	7,73	3,26	Sustained induction
LINC01214	9,20	17,88	43,41	27,34	Sustained induction
LINC01239	4,14	12,86	31,60	7,78	Sustained induction
LINC01619	0,83	0,51	0,51	0,42	
LIPH	2,58	3,13	3,99	1,76	Induction reset
LOXL2	2,08	4,69	7,79	4,69	Sustained induction
LRP2	3,27	4,71	4,54	2,99	Sustained induction
LUCAT1	2,68	6,05	8,36	3,28	Sustained induction
LYPD1	3,00	5,93	9,53	3,41	Sustained induction
LYPD3	0,17	0,17	0,31	0,73	Repression reset
MAF	0,00	0,00	19,71	31,48	
MAP1B	2,55	9,39	16,96	6,44	Sustained induction
<b>MAPK10</b>	2,62	3,10	2,61	1,33	Induction reset
MAPRE2	3,71	8,07	16,19	10,23	Sustained induction
MAPRE3	1,78	2,44	3,22	2,17	
MCF2L2	2,37	3,05	3,15	1,64	Induction reset
MCM2	0,34	0,33	0,33	0,68	Repression reset
MCM7	0,36	0,30	0,37	0,58	Repression reset
MCTP1	2,92	5,77	5,33	4,05	Sustained induction
MDGA2	3,42	6,23	7,58	3,42	Sustained induction
MECOM	3,20	5,13	5,05	2,54	Sustained induction
MIR222HG	1,63	3,09	5,04	3,01	
MIR3681HG	4,80	0,07	0,06	6,43	Dynamic

## GRHL2-controlled gene expression networks

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
MIR9-3HG	0,35	0,31	0,43	0,78	Repression reset
MITF	1,61	3,58	5,49	2,22	
MKI67	0,42	0,28	0,27	0,38	Sustained repression
MMP16	2,17	2,97	3,27	2,91	Sustained induction
MPPED2	0,75	0,59	0,41	0,35	
MRFAP1	0,33	0,47	0,50	0,59	
MRPL17	0,33	0,40	0,48	0,74	Repression reset
MRPL51	0,25	0,28	0,28	0,38	Sustained repression
MRPS34	0,25	0,28	0,29	0,65	Repression reset
MSMB	0,21	0,11	0,15	0,33	Sustained repression
<b>MTUS2</b>	2,69	2,31	5,20	2,93	Sustained induction
MYT1L	32,55	0,05	0,00	37,70	Dynamic
NBEA	2,41	3,23	3,26	1,63	Induction reset
NCF2	2,94	6,23	23,18	17,27	Sustained induction
NECTIN4	0,26	0,30	0,48	0,96	Repression reset
NEK10	1,50	2,78	6,78	2,59	
NHS	1,47	2,47	3,20	1,40	
NHSL2	3,09	4,61	6,21	3,48	Sustained induction
NLGN1	7,50	0,02	0,00	17,12	Dynamic
NME1	0,28	0,38	0,37	0,39	Sustained repression
NPAS3	2,72	4,13	3,49	1,06	Induction reset
NPM1P27	0,27	0,40	0,50	0,33	Sustained repression
<b>NPY1R</b>	0,81	0,33	0,20	0,19	
NR2C2AP	0,23	0,24	0,34	0,58	Repression reset
NRG2	3,76	8,84	10,44	3,37	Sustained induction
NRP1	2,18	3,57	3,13	1,58	Induction reset
NRXN3	10,82	0,94	1,45	11,75	
NT5DC2	0,28	0,42	0,46	0,85	Repression reset
NTN4	5,48	13,04	18,44	9,78	Sustained induction
NUDT1	0,26	0,23	0,28	0,67	Repression reset
OPCML	27,66	0,03	0,04	38,49	Dynamic
OPTN	1,69	3,76	6,50	5,12	

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
PACSN3	0,31	0,37	0,36	0,91	Repression reset
PALM2	2,42	4,65	4,73	1,72	Induction reset
PALM2- AKAP2	3,43	4,12	5,24	2,77	Sustained induction
PAPSS2	2,57	6,14	7,88	3,34	Sustained induction
PAQR5	1,21	1,66	1,91	3,05	
PCAT29	4,40	5,18	7,17	4,08	Sustained induction
PCSK2	63,00	0,05	0,00	99,27	Dynamic
<b>PGLYRP2</b>	0,18	0,09	0,03	0,27	Sustained repression
PGM2L1	2,30	2,73	4,43	2,62	Sustained induction
PHACTR3	40,57	0,04	0,06	47,82	Dynamic
PHGDH	0,22	0,29	0,31	0,65	Repression reset
PHLDB2	3,97	8,37	11,66	5,28	Sustained induction
PID1	1,84	3,60	10,27	7,22	
PIF1	0,45	0,28	0,28	0,55	Repression reset
PIK3IP1-AS1	5,12	6,17	7,66	3,53	Sustained induction
PIMREG	0,24	0,24	0,21	0,62	Repression reset
PKP1	0,27	0,18	0,19	0,73	Repression reset
PLCE1	2,78	9,19	9,46	2,97	Sustained induction
PLCXD2	4,35	9,62	12,79	5,07	Sustained induction
PLD1	3,98	7,92	7,69	2,78	Sustained induction
PLEKHH2	3,60	4,43	4,97	2,15	Sustained induction
PLIN4	0,11	0,03	0,09	0,34	Sustained repression
PLIN5	0,19	0,04	0,12	0,41	Sustained repression
PMP22	3,24	4,41	4,84	3,59	Sustained induction
POP7	0,30	0,30	0,40	0,71	Repression reset
PPARG	2,63	4,37	10,14	3,87	Sustained induction
PPIAP22	0,11	0,18	0,20	0,18	Sustained repression
<b>PPP1CA</b>	0,27	0,32	0,42	0,63	Repression reset
PPP1R14B	0,28	0,38	0,54	0,85	
PRELID1	0,33	0,36	0,41	0,60	Repression reset
PRICKLE2-AS1	2,03	3,23	3,48	1,32	Induction reset

## GRHL2-controlled gene expression networks

Gene	Fold change				Cluster
	D2 AFC	D4 AFC	D8 AFC	D16 AFC	
PROS1	2,99	3,62	4,27	1,93	Induction reset
<b>PRSS23</b>	2,39	4,01	5,50	5,31	Sustained induction
PSG5	2,07	7,09	9,24	7,10	Sustained induction
<b>PSMB6</b>	0,31	0,37	0,45	0,47	Sustained repression
PSMC3	0,34	0,45	0,48	0,67	Repression reset
PSMD2	0,36	0,52	0,61	0,63	
PSMG3	0,31	0,33	0,39	0,79	Repression reset
PSRC1	0,30	0,23	0,33	0,63	Repression reset
PTTG1	0,33	0,26	0,23	0,45	Sustained repression
PYCR1	0,24	0,37	0,37	0,90	Repression reset
QARS	0,30	0,44	0,43	0,66	Repression reset
RAB7B	3,72	4,73	14,45	7,71	Sustained induction
RAI2	2,11	3,70	8,48	4,74	Sustained induction
RBFOX1	2,27	0,74	0,52	3,17	
RBFOX3	34,53	0,01	0,02	39,26	Dynamic
RCAN2	9,68	1,28	20,29	16,55	
RECQL4	0,28	0,33	0,30	0,89	Repression reset
REEP4	0,24	0,27	0,34	0,99	Repression reset
RETREG1	2,37	3,64	3,37	2,27	Sustained induction
RFTN1	3,06	4,72	6,29	4,49	Sustained induction
RN7SL2	0,22	0,34	0,42	0,71	Repression reset
RN7SL3	0,29	0,45	0,54	0,86	
RN7SL4P	0,18	0,29	0,36	0,74	Repression reset
RNASEH2A	0,25	0,23	0,27	0,61	Repression reset
<b>RND3</b>	2,35	3,96	5,16	3,64	Sustained induction
RNF150	3,46	6,36	7,42	2,48	Sustained induction
RNF219-AS1	54,69	0,00	0,00	71,81	Dynamic
RNU1-120P	0,17	0,16	0,27	0,62	Repression reset
RNU1-122P	0,15	0,15	0,27	0,62	Repression reset
RNU2-63P	0,18	0,20	0,34	0,79	Repression reset
RNU4-1	0,17	0,18	0,30	0,91	Repression reset
RNU5D-1	0,14	0,29	0,43	0,53	Repression reset

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
RNVU1-6	0,17	0,15	0,21	0,67	Repression reset
RNVU1-7	0,20	0,22	0,24	0,59	Repression reset
RPL13A	0,32	0,41	0,49	0,64	Repression reset
RPL17	0,33	0,40	0,48	0,57	Repression reset
RPL3	0,30	0,48	0,55	0,72	
RPL35	0,32	0,38	0,41	0,58	Repression reset
RPL41	0,30	0,38	0,49	0,69	Repression reset
RPL7	0,33	0,47	0,46	0,42	Sustained repression
RPL7A	0,36	0,44	0,47	0,48	Sustained repression
RPL8	0,30	0,37	0,37	0,50	Repression reset
RPL9P9	0,15	0,27	0,18	0,25	Sustained repression
RPS10	0,32	0,37	0,34	0,28	Sustained repression
RPS11	0,35	0,37	0,40	0,46	Sustained repression
RPS2	0,28	0,34	0,35	0,61	Repression reset
RPS21	0,28	0,32	0,36	0,48	Sustained repression
RPS6KA2	2,48	2,79	3,13	2,78	Sustained induction
RPS8	0,33	0,41	0,38	0,47	Sustained repression
RTN1	2,33	5,02	4,89	2,28	Sustained induction
S100A14	0,34	0,52	0,50	0,57	
SAMD12	2,42	3,61	4,11	2,57	Sustained induction
SAMD12-AS1	3,99	6,35	7,19	4,12	Sustained induction
SAMD9	0,00	0,00	19,92	37,58	
SAPCD2	0,24	0,27	0,28	0,76	Repression reset
SCARNA12	0,14	0,20	0,26	0,26	Sustained repression
SCARNA13	0,22	0,32	0,48	0,47	Sustained repression
SCARNA21	0,08	0,10	0,10	0,22	Sustained repression
SCARNA7	0,17	0,31	0,55	0,31	
SDC1	0,28	0,35	0,48	0,97	Repression reset
SEMA6A	2,34	3,48	4,69	3,74	Sustained induction
SEPT8	0,35	0,66	0,82	1,11	
SESN3	6,90	10,94	12,91	7,02	Sustained induction
SFN	0,21	0,34	0,34	0,61	Repression reset

## GRHL2-controlled gene expression networks

Gene	Fold change				Cluster
	D2 AFC	D4 AFC	D8 AFC	D16 AFC	
SHC4	2,49	2,64	5,43	2,78	Sustained induction
SHMT2	0,22	0,39	0,42	0,82	Repression reset
SLC12A4	1,98	3,93	4,88	5,01	
<b>SLC16A3</b>	0,16	0,17	0,28	0,91	Repression reset
SLC1A1	6,72	11,29	17,89	10,80	Sustained induction
SLC22A1	1,68	4,62	6,00	3,78	
SLC22A15	2,05	2,94	3,12	1,68	Induction reset
SLC25A5	0,24	0,27	0,33	0,40	Sustained repression
SLC9A3	0,31	0,00	0,00	103,49	Dynamic
SLIT3	4,89	0,03	0,63	3,96	Dynamic
SMAGP	0,37	0,26	0,35	0,56	Repression reset
SNORD3A	0,10	0,13	0,19	0,27	Sustained repression
SNORD3B-1	0,19	0,23	0,32	0,67	Repression reset
SNORD3B-2	0,14	0,18	0,26	0,62	Repression reset
SOCS2-AS1	6,58	8,39	9,17	3,76	Sustained induction
SORCS2	1,61	2,38	3,41	2,64	
SOX9	3,49	6,38	11,38	12,66	Sustained induction
SOX9-AS1	3,93	3,43	2,82	2,14	Sustained induction
SPAG5	0,34	0,30	0,31	0,45	Sustained repression
<b>SPATA18</b>	4,04	4,13	2,47	1,69	Induction reset
SPEG	1,99	2,92	4,04	4,71	
SPOCK1	2,23	3,00	16,72	4,48	Sustained induction
SSNA1	0,22	0,27	0,39	0,74	Repression reset
SSRP1	0,33	0,44	0,45	0,58	Repression reset
ST3GAL5	1,49	2,64	4,09	3,44	
STAT4	2,50	6,67	5,66	2,00	Sustained induction
STUM	1,75	2,22	10,66	13,14	
SULF1	0,60	0,31	0,26	0,17	
SUN2	0,35	0,33	0,43	0,67	Repression reset
<b>SYNPO</b>	3,53	6,07	6,99	4,11	Sustained induction
SYNPR	8,34	0,01	0,01	10,33	Dynamic
SYT7	0,34	0,30	0,40	0,97	Repression reset

## Chapter 3

	Fold change				
Gene	D2 AFC	D4 AFC	D8 AFC	D16 AFC	Cluster
TANC2	2,06	3,16	3,40	1,56	Induction reset
TENM2	25,34	0,03	6,08	23,79	Dynamic
TFPI	1,65	2,39	3,30	2,04	
TGFB2	2,27	6,07	5,70	1,71	Induction reset
TGFB1	2,92	6,30	9,58	4,64	Sustained induction
TGFB2	3,05	6,81	8,02	2,99	Sustained induction
<b>THAP11</b>	0,20	0,23	0,29	0,81	Repression reset
THEG	0,23	0,07	0,15	0,19	Sustained repression
TIMP3	2,61	5,50	11,16	10,20	Sustained induction
TK1	0,32	0,26	0,31	0,70	Repression reset
TMC7	2,24	3,32	4,00	2,10	Sustained induction
TMEM107	0,22	0,20	0,31	0,47	Sustained repression
TMEM132C	38,52	0,00	0,00	50,13	Dynamic
TMEM132D	19,00	0,00	0,02	34,41	Dynamic
<b>TMEM140</b>	9,19	13,96	32,13	31,21	Sustained induction
TMEM156	4,33	4,22	20,11	4,15	Sustained induction
TMEM54	0,12	0,20	0,29	0,95	Repression reset
<b>TMPRSS13</b>	0,25	0,23	0,39	0,85	Repression reset
<b>TMPRSS4</b>	0,33	0,23	0,45	1,02	
TNFAIP8	2,41	2,92	3,91	1,62	Induction reset
TNIK	3,56	8,40	17,63	7,90	Sustained induction
TONSL	0,33	0,36	0,39	1,10	
TP53INP1	3,11	3,01	3,30	2,22	Sustained induction
TP63	1,61	4,51	34,52	16,43	
TPI1	0,23	0,34	0,41	0,52	Repression reset
TRAIP	0,38	0,25	0,28	0,67	Repression reset
TROAP	0,36	0,28	0,27	0,56	Repression reset
TSPAN5	2,38	3,86	6,41	2,92	Sustained induction
TUBA1B	0,19	0,22	0,25	0,45	Sustained repression
TUBB	0,30	0,33	0,36	0,56	Repression reset
TUBB4B	0,23	0,29	0,36	0,58	Repression reset
TXNIP	0,31	0,22	0,36	0,44	Sustained repression

## GRHL2-controlled gene expression networks

Gene	Fold change				Cluster
	D2 AFC	D4 AFC	D8 AFC	D16 AFC	
U1	0,24	0,18	0,33	0,79	Repression reset
U3	1,24	2,17	1,95	1,80	
UBB	0,31	0,38	0,49	0,40	Sustained repression
UBE2C	0,28	0,21	0,22	0,58	Repression reset
UBE2QL1	2,07	4,53	5,32	4,16	Sustained induction
UBL4A	0,30	0,22	0,30	0,58	Repression reset
<b>UHRF1</b>	0,33	0,27	0,32	0,68	Repression reset
UNC13C	24,62	0,10	0,06	38,12	Dynamic
UPP1	2,23	5,94	9,75	5,93	Sustained induction
UQCRQ	0,30	0,37	0,42	0,41	Sustained repression
USH2A	13,62	0,02	0,03	16,11	Dynamic
USP35	0,70	1,09	1,54	2,77	
<b>VMP1</b>	2,60	2,37	3,02	1,93	Induction reset
VSTM2B	7,96	0,00	0,00	36,91	Dynamic
WIPF1	3,54	5,91	6,80	4,04	Sustained induction
<b>WIPI1</b>	2,38	3,25	4,19	2,56	Sustained induction
WLS	2,59	3,48	3,91	2,16	Sustained induction
XRCC3	0,35	0,40	0,44	0,93	Repression reset
YPEL2	3,21	2,79	4,50	3,36	Sustained induction
Z93241.1	0,24	0,21	0,31	0,63	Repression reset
<b>ZBTB20</b>	3,17	2,79	5,25	2,21	Sustained induction
ZMAT4	1,72	5,72	14,92	5,98	
ZNF365	3,77	6,06	11,52	6,65	Sustained induction
ZNF385B	2,87	4,23	3,30	1,93	Induction reset
ZNF462	1,97	3,33	3,52	1,65	
ZNF827	2,41	3,25	3,03	1,78	Induction reset
ZWINT	0,29	0,26	0,33	0,64	Repression reset

**Table S2. GRHL2-regulated genes identified by Bru-seq in MCF7 conditional KO model.** AFC for indicated genes at the indicated timepoints (days) post induction of GRHL2 KO identified by Bru-seq in MCF7 conditional KO model and assignment

## Chapter 3

---

to clusters is shown. Genes also displaying promoter interaction identified by ChIP-seq in MCF7 cells are indicated in bold.