**Exploring Grainyhead-like 2 target genes in breast cancer**
Wang, Z.

**Citation**
Wang, Z. (2020, October 6). *Exploring Grainyhead-like 2 target genes in breast cancer*.
Retrieved from https://hdl.handle.net/1887/137309

| | |
|---|---|
| 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/1887/137309](#) |

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

Cover Page

Universiteit Leiden

Leiden University Repository

**Author**: Wang, Z.
**Title**: Exploring Grainyhead-like 2 target genes in breast cancer
**Issue date**: 2020-10-06

# Exploring Grainyhead-like 2 target genes in breast cancer

Zi Wang

**Exploring Grainyhead-like 2 target genes in breast cancer**

# Exploring Grainyhead-like 2 target genes in breast cancer

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus, prof. mr. C.J.J.M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op dinsday, 6 oktober 2020

klokke 10:00 uur

door

Zi Wang   |   王梓

geboren te Tianjin, China in 1988

# Table of Contents

# Chapter 1

## General introduction and outline of this thesis

# 1. Cancer development and progression

## 1.1 Hallmarks

In human cancer development, characteristic hallmarks are sustainability of proliferative signaling, resistance to cellular apoptosis, mutations of tumor suppressor genes, tumor angiogenesis, replicative immortality and tumor metastasis[1].

One of the most fundamental features of cancer is uncontrolled proliferation. Cancer cells acquire the capability to maintain proliferative signaling at activation. Cancer cells may secrete molecules (e.g., TGF-β) to stimulate the surrounding normal cells within the supporting tumor-associated stroma, which reciprocate by providing cancer cells with growth factors[2]. Alternatively, cancer cells may generate growth ligands themselves, to which they may react through expression of related receptors, leading to activation of proliferative signaling pathways. Reciprocal loops are important to maintain the delicate balance between the promotion and inhibition of cell growth[3]. Defects in the negative feedbacks may lead to consistent activation of signaling pathways associated with cell proliferation. For example, PTEN negatively regulates AKT/PKB signaling and intracellular levels of phosphatidylinositol-3,4,5-trisphosphate (PIP3). Loss of PTEN by its promoter methylation results in loss of a brake on PI3K, which promotes cell proliferation and reduces cell apoptosis[1,4].

Tumor suppressor genes negatively regulate cell proliferation. And tumor suppressor gene mutations are widely found in tumor tissues and cancer cells[5-7]. The well-known cancer-related genetic change is *TP53* mutation. Over 75% of *TP53* mutations lead to expression of mutant p53 proteins. Mutant p53 proteins have a dominant-negative effect beyond the remaining wild-type p53 protein[8,9]. Many studies suggest that, in non-small cell lung cancer (NSCLC), *TP53* mutations carry a worse prognosis in patients and are a contributor to cisplatin resistance[10]. Mutations of *TP53* contribute to breast cancer growth through regulation of mevalonate signaling pathway[11].

Apoptosis, also known as programmed cell death, acts as an intrinsic barrier to cancer development. Cancer cells possess the capability to circumvent apoptosis. Cancer

cells may upregulate expression of anti-apoptotic regulators (e.g., BCL2) or growth factors (e.g., IGF-I/II), combined with downregulation of pro-apoptotic factors (e.g., BCL2L4)[1]. In cervical cancer, overexpression of BCL2 contributes to resistance of $As_2O_3$-induced apoptosis[12].

To evacuate metabolic waste and absorb nutrients, tumor angiogenesis is required for cancer development. Genes involved in angiogenesis are promising targets for cancer therapeutic treatment. Vascular endothelial growth factor A (VEGF-A), a typical angiogenesis inducer, facilitates growth and migration of vascular endothelial cells. The disruption of VEGF-A may lead to abnormal blood vessel formation, antibodies against VEGF-A (Bevasizumab) are reported as an effective drug to counteract the progression of NSCLC[13]. PTK787, a VEGFR RTK inhibitor, is reported to significantly increase progression-free survival in colorectal cancer patients[13].

Cancer cells require the capability to unlimitedly replicate DNA for macroscopic tumors. In most normal cells, there are two barriers against unlimited replication, including senescence and crisis[1]. The transition, in which cells originate from a population in crisis and display the potential of unlimited replication, is termed immortalization. Established cell lines possess the feature of immortalization because of unlimited proliferation in culture. Telomeres, the highly conserved DNA sequences at the end of chromosomes, are the substrates for telomerase that is the enzyme responsible for addition of DNA to the ends of chromosomes. Telomerase is composed of the telomerase reverse transciptase (TERT) protein and the noncoding RNA component. Telomerase is positive in most non-immortalized cells[1], whereas telomerase activity is elevated in most cancer cells[14,15].

Most cancer-related mortality is caused by tumor metastasis. The period of metastasis dormancy and pattern of tumor metastasis are determined by cancer types[16]. Autopsy studies show that breast, lung and prostate cancers are more likely to metastasize than other types of cancer, such as bone cancer, which rarely forms metastasis at distant organs outside of primary site[17].

Combination of targeted therapy and cytotoxic agents offers an effective response against oncogenes, while its contribution to anti-metastasis therapy is transient, with an increase in overall survival of several months only[16]. To date, conventional therapeutic treatment of cancers has limited effectiveness in preventing and controlling metastasis in patients with cancer, probably due to the complex nature of tumor metastasis.

## 1.2 EMT and metastasis

The epithelial-mesenchymal transition (EMT) is a multiple-step process that enables the polarized epithelial cell to possess a mesenchymal cell phenotype, which includes increased cell proliferation, enhanced migration capability and invasiveness [18-20]. EMT can be divided into three types with highly different functional consequences[18]. Type 1 EMT is linked to implantation and embryonic development. For type 2 EMT, it is associated with tissue regeneration and organ fibrosis. Type 3 EMT is connected with cancer development and tumor metastasis. Activation of EMT requires molecular reprogramming of epithelium with new biochemical characteristics (**Fig. 1**). Main characteristics of EMT are loss of epithelial markers and gain of mesenchymal markers. The study of classical EMT regulation, to date, concentrates on the prototypic adhesion molecular, E-cadherin [21]. E-cadherin (*CDH1*) emerges as one of the epithelial markers. Loss of E-cadherin is considered to be highly involved in EMT and tumor metastasis[22]. Mutations of *CDH1* are shown to contribute to oral squamous cell carcinoma, breast cancer, NLCSC and thyroid cancer[23-27]. During the process of EMT, loss of cell-cell adhesion and cell junctions trigged by loss of E-cadherin allows cells to be separated from the primary cancer, migrate, invade surrounding tissues and colonize at distant organs. Meanwhile, the re-established functional E-cadherin/catenin complex converts invasive tumor phenotype into a benign and epithelial phenotype[28]. The cells that lose epithelial markers and gain mesenchymal markers are considered to be the cells that eventually go into blood circulation and colonize at distant tissues[29,30]. However, cells at distant colonized sites do not exhibit mesenchymal phenotypes. And equal or even higher E-cadherin expression is observed in tumor metastases relative to the primary tumors, suggesting cells at secondary sites undergo mesenchymal to epithelial transition (MET)[18].

On the other hand, some *in vivo* studies showed that EMT was required for chemotherapy resistance, instead of tumor metastasis[31,32]. Moreover, no convincing evidence of EMT is observed at any stage of tumor tissues so far[33], suggesting the occurrence of EMT in cancer tissues is in doubt.
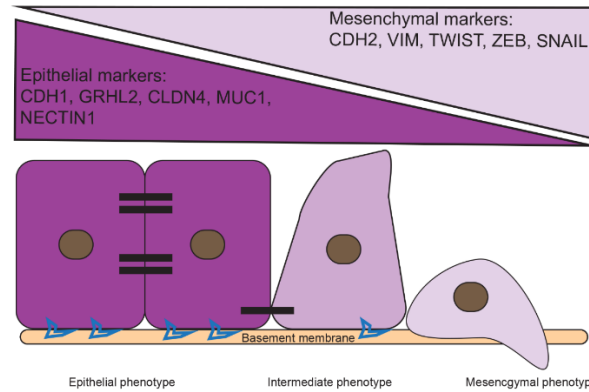


**Fig. 1** EMT is a continuum which can be divided into epithelial, intermediate, and mesenchymal phenotype.

Taken together, the role of EMT in cancer development is still controversial. One of the reasons is the lack of a standardized criteria of EMT definition. Another one is variation on marker expression patterns in distinct tumor samples[34,35].

## 1.3 Therapy resistance

EMT induction allows differentiated cells to gain a multipotent stem cell-like phenotype[36]. Acquisition of cancer stem cell features in the process of EMT contributes to therapeutic resistance. Activated Notch signaling pathway is linked to enhanced cell proliferation in gemcitabine therapy resistance[37]. Both Slug and Snail, important regulators of EMT, are connected with chemotherapy resistance in ovarian cancer[38]. The inducer of EMT, Twist, is associated with hormone therapy resistance in breast cancer through downregulation of estrogen receptor $\alpha$[39].

Tumor microenvironment is also involved in EMT and therapy resistance. Many studies show that the components of the tumor stroma secrete molecules that trigger induction of EMT such as transforming growth factor beta (TGF-β) and interleukin 6 (IL-6)[40]. IL-6 originating from cancer associated fibroblasts (CAFs) is connected to EMT-mediated chemotherapy resistance in NSCLC[41]. As an early tumor suppressor, TGF-β represses

normal epithelial cell growth. However, in the late stage of cancer progression, TGF-β is considered as the most potent EMT-inducing signal in many cancer contexts. In squamous cell carcinoma stem cells, TGF-β-responsive cells are accompanied by the features typically associated with EMT and TGF-β diminishes cisplatin-induced apoptosis via activation of p21[42]. So far, the well-accepted assumption is that paracrine signals from stromal components lead to induction of EMT, resulting in therapy resistance, rather than tumor metastasis.

# 2. Breast cancer

Breast cancer is one of the most common cancers among women globally and becomes the second leading cause of cancer-related mortality[43]. Breast cancer is a highly heterogeneous disease with variable morphologies and clinical implications[44]. Immunohistochemical testing for epidermal growth factor receptor 2 (HER2), estrogen receptor (ER), progesterone receptor (PR) is conventionally used in clinical practice for diagnosis of breast cancer. With the development of high-throughput platforms for gene expression analysis, it has been shown that the response of tumor cell to treatment is in part determined to intrinsic molecular characteristics, suggesting that classification of breast cancer according to molecular characteristics may not only increase the accuracy of disease diagnosis and also therapeutic decision making[45]. The differences in gene expression profile reflect the fundamental differences of breast cancer at the molecular level.

The gene expression profiling divides breast cancer into several subtypes, including basal-like, luminal A, luminal B, HER2-enriched and normal-like subtype. Most basal-like breast cancers are triple negative breast cancer, with absent expression of ER, PR and HER2. The HER2-enriched subtype generally has HER2 gene amplification on chromosome 17q12. Both luminal A and luminal B are positive for ER and PR[46]. Each molecular subtype corresponds to an IHC-defined subtype (**Table 1**), except for normal-like breast cancer that has similarities in IHC status with luminal A subtype.

Identification of molecular subtypes provides valuable information on prediction of treatment response and clinical outcome. Both HER2-enriched and basal-like breast cancer have higher rates of complete response to pre-surgery chemotherapy than

luminal-like breast cancer. Patients with basal-like breast cancer are more likely to have an unfavorable prognosis and a short-term disease-free and overall survival[47].

**Table 1**

| Intrinsic subtype | IHC status | Outcome | Prevalence |
|---|---|---|---|
| **Luminal A** | **ER+, PR+, HER2-, Ki67-** | **Favorable** | **23.7%** |
| **Luminal B** | **ER+, PR+, HER2-, Ki67+** | **Intermediate** | **38.8%** |
| | **ER+, PR+, HER2+, Ki67+** | **Unfavorable** | **14.0%** |
| **Her2-enriched** | **ER-, PR-, HER2+** | **Unfavorable** | **11.2%** |
| **Basal-like** | **ER-, PR-, HER2-, basal markers+** | **Unfavorable** | **12.3%** |
| **Normal-like** | **ER+, PR+, HER2-, Ki67-** | **Intermediate** | **7.8%** |

This table describes the characteristics of each molecular subtype breast cancer cited from Bozhi Shi (45)

## 2.1 Luminal A and luminal B

Luminal-like breast cancer is characterized by high expression of a panel of luminal associated genes/proteins such as ESR, T18/19, GATA3, FOXA1, Cytokeratin-8/18 and Cyclin D1[44,48]. Due to tight cell-cell contacts, luminal-like breast cancer is comparably differentiated and has limited ability to migrate. There are various single subtype predictors (SSPs) used to identify the molecular subtype of an individual breast cancer, none of which could produce substantial agreement in subdividing luminal breast cancers[49]. Despite differences in definition of luminal subtype classification, classification of luminal subtype remains useful and important for clinical practice.

In general, luminal-like breast cancer can be further stratified into luminal A and luminal B according to expression level of proliferation-related genes and HER2 status. Luminal A is the most common subtype and its characteristics include high expression of estrogen receptor (ER), low expression of HER2 and low expression of proliferation-related genes[50,51]. Luminal A breast cancer tends to express genes, such as *GATA3*,

TOB1, *ERBB3* and *SPDEF* that link to a more differentiated and noninvasive phenotype[52].

Luminal A status may be an indicator of lack of chemotherapy benefit, owing to low levels of proliferation-related genes.

As for luminal B, it exhibits low expression of ER, variable expression of HER2 and high expression of proliferation-related proteins (e.g. Ki-67, CCNB1 and MYBL2)[51]. Proliferation is identified as one of the most important features of several prognostic multigene signatures, which distinguishes high-risk luminal B from low-risk luminal A[49,53]. Luminal B subtype may be more invasive and aggressive than luminal A subtype, and it is insensitive to endocrine therapy relative to luminal A subtype, and to chemotherapy compared with HER2-enriched and basal-like subtype. Some clinical studies, in which have differences in subtype definition and chemotherapy received, showed that complete response rate is consistently lower in luminal B subtype relative to HER2-enriched and basal-like subtype[44,54-56].

Luminal B derives from luminal A in term of proliferation-related markers. However, luminal A and luminal B are distinct entities since sequencing data reveal that luminal B has molecular uniqueness, including gene copy number alternations, DNA methylation, and somatic point mutations[57]. High-level DNA amplification and chromosomal aberrations are more frequently examined in luminal B than other subtypes[58,59]. Luminal B is more likely to have higher frequency of *TP53* mutations relative to luminal A[57]. The increased expression of PI3K signaling pathway genes is a feature of luminal B subtype breast cancer[44].

Clinical outcome of luminal B breast cancer is similar to non-luminal subtypes. A multivariate analysis showed that hazard ratio (2.43, *P*<0.001) of luminal B breast cancer patients for relapse-free survival is similar to that (2.53, *P*<0.001) of patients with HER2-enriched breast cancer[51]. Unlike basal-like breast cancer that is more likely to metastasize to brain tissue, luminal B has a preference to relapse to pleura and lung that is similar to luminal A[60].

8

## 2.2 Basal-like

The prevalence of basal-like breast cancer (BLBC) is between 12.3% and 36.7% of all breast cancer cases in different patient cohorts[50,61-64]. The incidence of BLBC is associated with increased parity, early age of menarche, and first full-term pregnancy before age of 26[64-66].

Basal-like breast cancer is a highly aggressive molecular subtype characterized by enrichment of genes expressed by epithelial cells in the basal or outer layer of adult mammary gland[64]. The basal-like subtype is characterized by high expression of Keratin 5 and 17, Laminin, and fatty acid binding protein 7 (FABP7)[62]. Most basal-like breast cancers are negative for ER, PR and HER2, and named triple-negative breast cancer (TNBC). Unlike ER-positive luminal subtype and HER2-enriched subtype breast cancer, basal-like breast cancer typically lacks expression of molecular targets that confers responsiveness to typical target therapies such as tamoxifen or trastuzumab. One of characteristics of BLBC is high proliferation rate[67]. Downregulated expression of Retinoblastoma 1 (*RB*) and Cyclin D1 and elevated expression of E2F3 and Cyclin E contribute to enhanced cell proliferation. Copy number of *CCNE1* is much higher in BLBC than other subtypes, and its expression correlates with unfavorable prognosis for patients with breast cancer[68-70]. *RB* tumor suppressor gene negatively regulates G1 to S cell cycle transition that is required for cell proliferation[71]. Phosphorylated Retinoblastoma 1 promotes G1 to S cell cycle transition by releasing E2F, a transcription factor that activates *CCNE1* expression[64]. Moreover, epidermal growth factor receptor (EGFR) highly expresses in BLBC and facilitates cell proliferation by activation of RAS/MAPK/MAPKK signaling pathway[72]. Basal-like breast cancer has a higher frequency of *TP53* mutations relative to luminal-like and HER2-enriched breast cancer, which is associated with unfavorable prognosis and poor response to systemic therapy[62,73-75].

Basal A breast cancer may exhibit either luminal-like or basal-like morphology. Basal B subtype breast cancer preferentially expresses *CD44*, *VIM*, *AXL* and *SPARC* that are associated with a mesenchymal or cancer stem cell phenotype, which leads to much more invasiveness of basal B than basal A and luminal-like breast cancer.

## 2.3 HER2-enriched

HER2, one member of four membrane receptor tyrosine kinases (RTKs), was firstly identified as a novel gene from rat neuroblastomas NIH 3T3 cells[76]. HER2, located on chromosome 17q12, is amplified or overexpressed in about 15% ~ 20% of all breast cancer cases[77]. The copy number of *HER-2* in breast cancer ranges from 25 to 50, whereas HER2 protein expression may increase 40- to 100-fold[78,79], suggesting copy number of *HER-2* correlates with HER2 protein expression in 90% of breast cancer cases[80]. Amplification of *HER-2* negatively correlates with overall survival and time to relapse in breast cancer patients[80], suggesting *HER-2* is a useful prognostic factor. Many findings suggest that HER2 is a major classifier of breast cancer and target of therapy. Despite a well-accepted finding that the majority of *HER-2* mutations are activating mutations[81], *HER-2* mutations can be found in breast cancers lacking *HER-2* amplification.

To date, a number of HER2-targeted medications have been developed such as small molecule inhibitors (e.g., ZD1839), monoclonal antibody (e.g., Trastuzumab and Pertuzumab), and antibody drug conjugates. HER2 overexpression correlates with the benefit of HER2-directed therapy. For patients with HER2-enriched breast cancer, HER2-targeted therapy is recommended, except for those who have clinical congestive heart failure or compromised left ventricular ejection fraction[82]. HER2-targeted therapy in combination with chemotherapy increases the response rate, progression-free survival and overall survival relative to chemotherapy alone for HER-enriched breast cancer patients[83].

## 2.4 Claudin-low

The Claudin-low subtype is a new molecular subtype of breast cancer [84]. It is characterized by tumor initiating cell genomic signature and lack of tight junction and cell to cell adhesion (e.g., Claudin 3, 4, 7 and E-cadherin)[85]. Claudin-low subtype breast cancer highly expresses mesenchymal markers (e.g., Vimentin) and exhibits enrichment for cancer stem cell features [85]. The majority of Claudin-low breast cancer is negative for ER, PR and HER2, which is similar to basal-like breast cancer. However, Claudin-low breast cancer has inconsistent expression of basal keratins (e.g., Keratin 5, 14 and 17) and does not highly express proliferation-related genes, indicating it is a

slower-cycling breast cancer which distinguishes from basal-like breast cancer. Some studies show that Claudin-low breast cancer has high expression of stromal-specific and lymphocyte- or granulocyte-specific gene signatures relative to other molecular subtypes of breast cancer[86,87].

## 2.5 Representative cell lines of basal-like and luminal-like breast cancer

By virtue of unlimited self-replication, breast cancer cell lines are widely utilized for breast cancer researches[48]. Whether the breast cancer cell lines reflect the molecular characteristics of corresponding tumors is crucial for *in vitro* breast cancer experiments. As described above, breast cancer cell lines are divided into several subtypes according to gene expression profiling. Since there is still no unifying and strict definition for basal-like subtype of breast cancer[88], a few cell lines may be categorized as luminal-like or basal-like simultaneously in different articles (e.g., HCC1500). Although inconsistency of definition of basal-like/luminal-like exist so far, most cell lines have the identical classification in a number of different studies. Here, representative cell lines are listed.

| Basal A | Basal B | Luminal-like | |
|---|---|---|---|
| BT20 | BT549 | CAMA-1 | SK-BR-5 |
| DU4475 | HCC1395 | EVSA-T | SK-BR-3 |
| HCC1143 | HCC1500 | HCC1419 | MPE600 |
| HCC1187 | HCC38 | BT483 | OCUB-F |
| HCC1569 | Hs578T | HCC202 | SUM185PE |
| HCC1599 | MDA-MB-157 | HCC2218 | BT474 |
| HCC1806 | MDA-MB-231 | MCF7 | ZR-75-1 |
| HCC1937 | MDA-MB-435s | MDA-MB-134VI | UACC893 |
| HCC1954 | MDA-MB-436 | MDA-MB-175VII | T47D |
| HCC70 | SUM149PT | MDA-MB-330 | ZR75-1 |
| MDA-MB-468 | SUM159PT | MDA-MB-361 | ZR75-30 |
| SUM102PT | SUM1315MO2 | MDA-MB-415 | SUM44PE |
| SUM190PT | | MDA-MB-453 | SUM52PE |
| SUM225CWN | | | MDA-MB-175 |
| SUM229PE | | | |

# 3. Grainyhead like transcription factors

The Grainyhead (*GRH*) gene family encodes transcription factors with an isoleucine-rich activation domain, a DNA-binding domain, and a dimerization domain[89-91]. After

identification of the first member of the *GRH* gene family in *Drosophila*, GRH homologs have subsequently been identified in other animals such as nematodes and mammals.

Based on whether the family members are associated with the *Drosophila* GRH or *Drosophila* CP2 (*dCP2*), this gene family has been divided into two main categories: GRH like (GRHL) and CP2. A study shows that there is no interaction between GRHL and CP2 so far[90], consistent with the observation that GRHL has no striking identity with *d*CP2 in the protein dimerization domain[92] .

In mammals, there are three mammalian members of the GRHL family, which have been termed GRHL1, GRHL2 and GRHL3 (**Fig. 2**) [93]. These transcription factors adopt a DNA-binding immunoglobulin fold homologous to the core domain of tumor suppressor p53 and they display remarkable amino acid sequence identity with each other, particularly in the functional DNA-binding and dimerization domains[92]. The N-terminal domains of GRHL transcription factors are involved in transcriptional activation and C-terminal regions possess DNA-binding and dimerization domains[94].
The expression patterns of these factors are tissue and developmentally specific, which means they can show differential spatiotemporal expression patterns during development[95].

GRHL proteins are involved in many important biological processes, including cell migration, cell growth and differentiation, through interactions with other transcription factors, gene promoters or partner proteins[96-98].
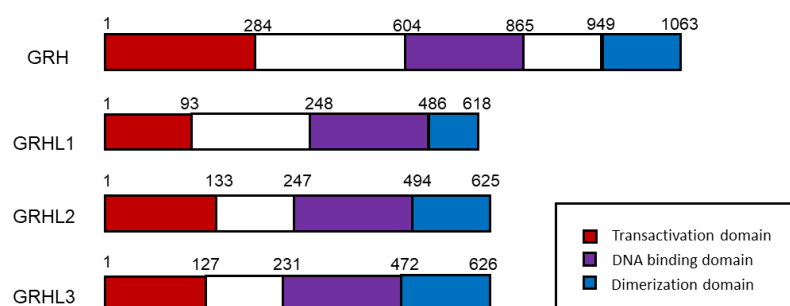


**Fig. 2 Schematic structure of GRH (top lane) and GRHL family members** (Adapted from Alarid (92).

## 3.1 GRHL1

GRHL1, also known as MGR, LBP-32 and TCFCP212, is one of the highly conserved family of β-scaffold transcription factors. GRHL1 can exist as homodimer or can form heterodimers with GRHL3 and GRHL2.

*GRHL1* is transcriptionally and epigenetically inhibited by interaction of HDAC3 with MYCN via their binding to the promoter of *GRHL1*. A *in vivo* experiment shows that mice lacking GRHL1 exhibit hair loss and palmoplantar keratoderma, due to downregulated expression of desmoglein 1[99].

## 3.2 GRHL2

*GRHL2* encodes a 325 amino acid protein and is positive in human brain, placenta, kidney, prostate, thymus, lung, salivary, mammary gland, digest tract and pancreas[88,100,101]. To date, GRHL2 has three identified isoforms. Isoform 1 is the full length of GRHL2. And isoform 2 results from translation at alternative site, therefore, isoform 2 is five amino acids shorter than isoform 1. The isoform 3 has no transcriptional activity due to the loss of 98 amino acid at N-terminal[101].

GRHL2 is involved in non-tumor disease in human and mice. Mutations of *Grhl2* give rise to loss of non-neural ectoderm integrity and abnormal mesenchymal phenotype by downregulation of EMT suppressors *Lamc2*, *Tmprss2*, *Sostdc1* and *Esrp1*[102]. In the process of neural tube closure, GRHL2 is regulated by tet methylcytosine dioxygenase 1 (TET1)[103]. Higher expression of GRHL2 is observed in kidney. And loss of GRHL2 leads to deficient nephric duct lumen expansion and damages epithelial barrier formation and even inhibits lumen expansion in collecting duct epithelia[101]. Mutations in *GRHL2* are observed in progressive autosomal-dominant hearing loss[88,104]. A single nucleotide polymorphism (SNP) found in the human *GRHL2* gene has a correlation with age-related hearing impairment (ARHI) in a American population[88]. Meanwhile, genetic polymorphisms of *GRHL2* may have protective effects on sudden sensorineural hearing loss (SSHL) and may lower the risk of SSHL[105,106]. *GRHL2* mutation is also found in skin diseases[107]. Autosomal recessive missense mutations in *GRHL2* leads to ectodermal dysplasia, which is characterized by pigmentation of oral mucosa and/or tongue abnormal dentition, including hypodontia and enamel hypoplasia[108] In the field of prenatal diagnosis, *GRHL2* deletion may be a marker for diagnosis of bilateral cleft lip through examination of polymorphic DNA marker[109]. In

primary human bronchial epithelial cells, deletion of *GRHL2* leads to failure of establishment of electrical resistance and differentiation of multi-ciliated cells [110].

### 3.3 GRHL3

GRHL3, also known as SOM, TFCP2L4 and VWS2, is only capable of forming multi-protein complexes with the other members of GRHL family, without interaction with CP2.

In line with the finding that GRH family is highly conserved from *Caenorhabditis elegans* to human[92,111], the role of GRHL3 in the maintenance of epidermal integrity in mice is also confirmed. GRHL3 is required for formation and maintenance of the epidermal barrier in mice and lacking of GRHL3 leads to defective skin barrier function, loss of eyelid fusion[112] and deficient wound repair[113], in which other family members fail to compensate for the loss of GRHL3.

*In vivo* experiment shows that knockout of *Grhl3* gives rise to an eye-open at birth phenotype, probably owing to repression of F-actin polymerization, and filopodia formation[114].

A null mutation of *Grhl3* also exhibits spina bifida in mouse[115], suggesting that GRHL3 plays important a role in closure of several structures.

# 4. GRHL1 and GRHL3 in cancer

As a tumor suppressor, *GRHL1* maps to chromosome 2p25.1 that is infrequently involved in aberrations in neuroblastoma[116]. A study shows that GRHL1 expression level positively correlates with favorable patient survival in primary neuroblastoma[117]. GRHL3 has a protective role against several cancers in human and mice. As a tumor suppressor, loss of GRHL3 by a miR-21-dependent network results in squamous cell carcinoma in human, in part due to downregulation of PTEN that is one of GRHL3 targets via activation of PI3K/AKT/mTOR signaling[118]. In the early stage of breast cancer, elevated expression of GRHL3 is observed in both plasma and tumor samples,

whereas reduced expression of GRHL3 in both plasma and tumor samples characterizes advanced stages.

GRHL3, induced by tumor necrosis factor alpha (TNF α), strongly stimulates human umbilical vein endothelial cell migration[119], consistent with the finding that GRHL3 elevates the capacity of cell migration in endothelial cell via activation of AKT and endothelial nitric oxide synthase (eNOS)[120]. Upregulation of GRHL3 is observed in breast cancers related to atypical hyperplasia. But GRHL3 expression in histological grade 1 is higher than that in histological grade 3, similar to the finding that GRHL3 is highly expressed in the early stage of breast cancer[121]. GRHL3 expression correlates with longer breast cancer-specific survival in lymph node-positive group[122], in part due to regulation of E-cadherin mediated by GRHL3[123].

In humans, GRHL3 has three isoforms (SOM1, SOM2 and SOM3) that are derived from differential first usage and alternative splicing and differ in their N terminal domain. These isoforms can dimerize with each other and other members of GRHL family, recognizing the same DNA-binding domain[90].

SOM2 is present in human and mice. But as for SOM1 and SOM3, with different N termini, they are not found in mice and specific to human, which is caused by that the mouse genome lacks the corresponding first exon[90]. Both SOM1 and SOM2 have a highly conserved activation domain  in the N-terminal region, which lacks in SOM3[90] that is less widely expressed related to SOM1 and SOM2. And both of them are transcriptional activators, but have opposing effects on apoptosis and migration in primary human endothelial cell via regulating expression of different target genes[124].

## 5. GRHL2 in Cancer

GRHL2, also known as BOM, ECTDS, TFCP2L3 and DFNA28, is located on human chromosome 8q22.3 that is highly conserved in mammals and amplified or overexpressed in multiple cancers such as oral cancer, breast cancer, prostate cancer and acute myeloid leukemia[125-129]. Amplification of *GRHL2* inhibits death receptor-mediated apoptosis by repressing FAS and DR5 on the cell surface[130]. In oral

squamous cell carcinoma, overexpressed GRHL2 promotes cell proliferation due to upregulation of the human telomerase reverse transcriptase gene expression by GRHL2-mediated DNA methylation[131]. And a gain of *GRHL2* is closely associated with early recurrence of hepatocellular carcinoma[132]. In prostate cancer, *GRHL2* is commonly amplified and overexpressed but expression levels of GRHL2 are not associated with Gleason grade or serum prostate specific antigen levels[128]. A recent study shows that GRHL2 is not only required for cell proliferation, but also for maintenance of androgen receptor (AR) expression[133]. GRHL2 is regulated by AR and co-localized with AR at specific sites on DNA to regulate gene expression[133]. *GRHL2* overexpression is observed in NLCSC cell lines and is associated with poor prognosis [134]. Due to its capability of transforming NIH3T3 fibroblasts, GRHL2 has been identified as the first member of GRH family of transcription factors to induce malignant transformation of cells. In the mouse model, overexpression of GRHL2 promotes breast tumor growth and metastasis in part owing to affecting *microRNA-200*s that directly target *Sec23a* that mediates secretion of metastasis-suppressive proteins[135]. The clinical relevance of GRHL2 in prognosis of patients with breast cancer is demonstrated by the finding of a positively significant association between overexpression of GRHL2 and poor relapse-free survival and increased risk of metastasis[136]. These observations indicate that *GRHL2* is a potential oncogene.

GRHL2 is an epithelial marker. Microarrays of 51 breast cancer cell lines show that *GRHL2* is expressed specifically in epithelial cell lines, with high expression level of *CLDN3, CLDN4, CLDN7, TJP2* and *CD24*[136]. Downregulation of GRHL2 is observed in basal B subtype breast tumors that exhibit mesenchymal gene expression signatures[97]. Expression levels of GRHL2 are associated with the epithelial phenotype, which means cell types with strong epithelial features tend to have higher GRHL2 expression[98,137].

GRHL2 is a suppressor of EMT. GRHL2 activates expression of *CDH1*, *CLDN4* and *OVOL2*[97,138], which are typical epithelial markers. GRHL2 may suppress TGF-β-induced and spontaneous EMT in part by suppression of ZEB1 expression[97,139]. It means that GRHL2 may inhibit tumor metastasis, perhaps through its ability to suppress EMT.

GRHL2 not only upregulates expression of epithelial-specific genes (e.g., *CDH1*, *CLDN4* and *OVOL2*) (**Fig. 3**)[138,140,141], but also downregulates mesenchymal regulators such as *ZEB1*, *ZEB2* and *CDH2* (N-cadherin) [142-144]. *ZEB1* is identified as a direct target gene of GRHL2 due to the finding that GRHL2 negatively regulates expression of *ZEB1* mRNA by binding *ZEB1* promoter in human mammary epithelial cell line (HMLE)[97]. However, in epithelial ovarian cancer (EOC) cells, any binding of GRHL2 is not found around *ZEB1* promoter, suggesting that the interaction of GRHL2 with *ZEB1* is cell context-dependent[98].
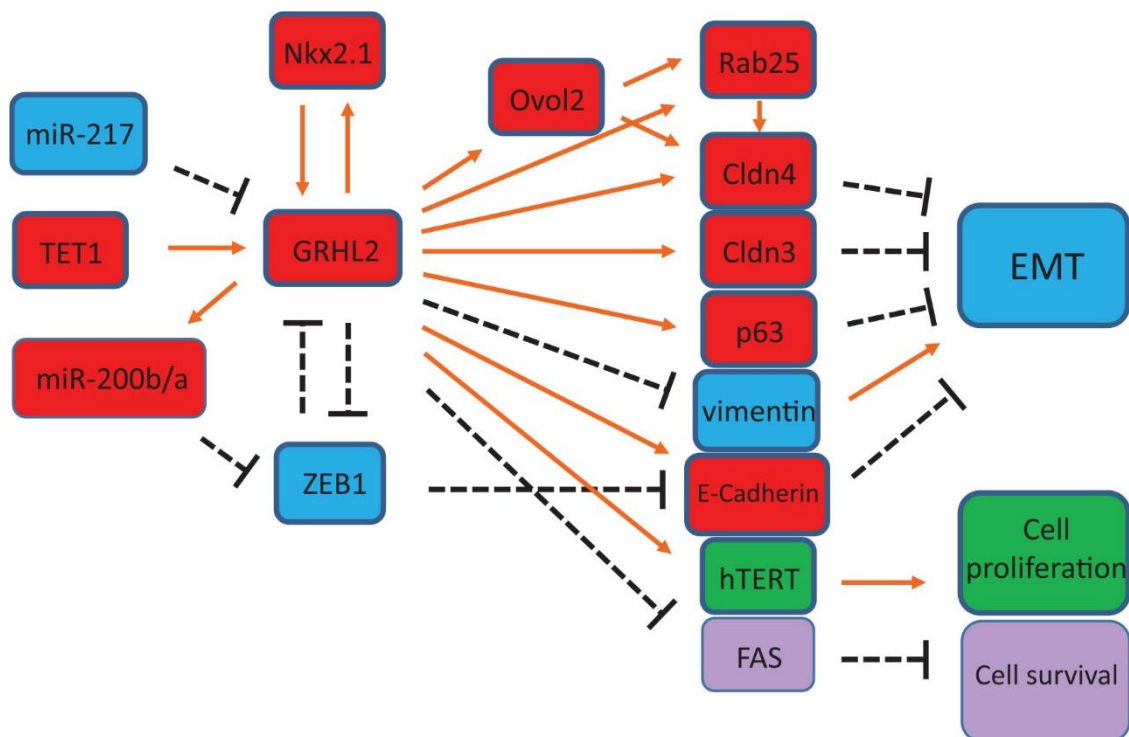


**Fig. 3 The regulatory network of GRHL2**. *Red color*: antagonists of EMT. *Blue color*: agonist of EMT. *Green color*: activator of cell growth. *Violet color*: repressor of cell survival. *Dotted line*: indirect relationship. *Solid line*: indirect relationship. (Adapted from Sun (102).)

GRHL2 also coordinates p63 to maintain integrity of epithelial cell and embryonic morphogenesis[145-147]. GRHL2 is required and necessary for p63 expression in part due to the fact that GRHL2 binding is observed at *TP63* promoter and overexpression of GRHL2 leads to upregulation of p63[147]. Interestingly, p63 knockdown also reduces *GRHL2* promoter activity, which means reciprocal feedback loop exists between GRHL2 and p63[147].

Beside transcriptional regulation, GRHL2 regulates histone modifications to mediate expression of its target genes. Downregulation of GRHL2 leads to reduced active histone marks (H3K4me3 and H3-K9/14ac) at *Cdh1* promoter in mouse kidney cells[94]. In EOC, GRHL2 knockdown significantly increases H3K27me3 levels at the promoters and GRHL2 binding sites of *CDH1* and *miR-200B/200A/429*[98]. Inhibition of the recruitment of histone demethylase JMJD3 induced by GRHL2 results in elevated levels of H3K27me3 at the promoters of GRHL2 target genes in keratinocytes[148].

# 6. Outline of this thesis

The objective of this study was to investigate the expression and function of GRHL2 in different breast cancer subtypes. In **Chapter 2**, we focused on the expression of GRHL2 in human breast cancer and the distinct effects of GRHL2 knockout on aspects of growth versus migration in basal A and luminal-like subtypes. In **Chapter 3**, ChIP-seq was used to explore the genomic landscape of GRHL2 binding sites in basal A and luminal-like subtypes of breast cancer and this data was used to predict shared and distinct GRHL2 target genes. In **Chapter 4**, based on a conditional GRHL2 knockout system, we determined the dynamic changes in genome-wide DNA transcription triggered by loss of GRHL2 in luminal-like breast cancer cells and used the data to predict affected pathways. In **Chapter 5,** ChIP-seq and BrU-seq data were integrated to identify genes whose transcription is controlled by GRHL2 and establish gene expression networks regulated by GRHL2 in luminal-like breast cancer.

# References

1       Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
2       Cheng, N., Chytil, A., Shyr, Y., Joly, A. & Moses, H. L. Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion. *Mol Cancer Res* **6**, 1521-1533, doi:10.1158/1541-7786.MCR-07-2203 (2008).
3       Cooper, J. P. & Youle, R. J. Balancing cell growth and death. *Curr Opin Cell Biol* **24**, 802-803, doi:10.1016/j.ceb.2012.11.003 (2012).
4       Chang, F. *et al.* Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. *Leukemia* **17**, 590-603, doi:10.1038/sj.leu.2402824 (2003).
5       Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49-53, doi:10.1126/science.1905840 (1991).
6       Narla, G. *et al.* KLF6, a candidate tumor suppressor gene mutated in prostate cancer. *Science* **294**, 2563-2566, doi:10.1126/science.1066326 (2001).
7       Leroy, B. *et al.* Analysis of TP53 mutation status in human cancer cell lines: a reassessment. *Hum Mutat* **35**, 756-765, doi:10.1002/humu.22556 (2014).
8       Petitjean, A. *et al.* Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* **28**, 622-629, doi:10.1002/humu.20495 (2007).
9       Willis, A., Jung, E. J., Wakefield, T. & Chen, X. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene* **23**, 2330-2338, doi:10.1038/sj.onc.1207396 (2004).
10      Mogi, A. & Kuwano, H. TP53 mutations in nonsmall cell lung cancer. *J Biomed Biotechnol* **2011**, 583929, doi:10.1155/2011/583929 (2011).
11      Freed-Pastor, W. A. *et al.* Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway. *Cell* **148**, 244-258, doi:10.1016/j.cell.2011.12.017 (2012).
12      Deng, Y. *et al.* Overexpression of Bcl-2 partly inhibits apoptosis of human cervical cancer SiHa cells induced by arsenic trioxide. *Chin Med J (Engl)* **113**, 84-88 (2000).
13      Ferrara, N. & Kerbel, R. S. Angiogenesis as a therapeutic target. *Nature* **438**, 967-974, doi:10.1038/nature04483 (2005).
14      Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015, doi:10.1126/science.7605428 (1994).
15      Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* **11**, 319-330, doi:10.1038/nrg2763 (2010).
16      Wan, L., Pantel, K. & Kang, Y. Tumor metastasis: moving new biological insights into the clinic. *Nature medicine* **19**, 1450-1464, doi:10.1038/nm.3391 (2013).
17      Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57-70, doi:10.1016/S0092-8674(00)81683-9 (2000).
18      Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119**, 1420-1428, doi:10.1172/JCI39104 (2009).
19      Kalluri, R. & Neilson, E. G. Epithelial-mesenchymal transition and its implications for fibrosis. *The Journal of clinical investigation* **112**, 1776-1784, doi:10.1172/JCI20530 (2003).
20      Ma, J. *et al.* Cadherin-12 enhances proliferation in colorectal cancer cells and increases progression by promoting EMT. *Tumour Biol* **37**, 9077-9088, doi:10.1007/s13277-015-4555-z (2016).
21      Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21-45, doi:10.1016/j.cell.2016.06.028 (2016).
22      Petrova, Y. I., Schecterson, L. & Gumbiner, B. M. Roles for E-cadherin cell surface regulation in cancer. *Mol Biol Cell* **27**, 3233-3244, doi:10.1091/mbc.E16-01-0058 (2016).
23      Chen, H. N. *et al.* PDLIM1 Stabilizes the E-Cadherin/beta-Catenin Complex to Prevent Epithelial-Mesenchymal Transition and Metastatic Potential of Colorectal Cancer Cells. *Cancer Res* **76**, 1122-1134, doi:10.1158/0008-5472.CAN-15-1962 (2016).
24      Pannone, G. *et al.* The role of E-cadherin down-regulation in oral cancer: CDH1 gene expression and epigenetic blockage. *Curr Cancer Drug Targets* **14**, 115-127 (2014).

25    Liu, X., Su, L. & Liu, X. Loss of CDH1 up-regulates epidermal growth factor receptor via phosphorylation of YBX1 in non-small cell lung cancer cells. *FEBS letters* **587**, 3995-4000, doi:10.1016/j.febslet.2013.10.036 (2013).

26    Ali, K. M. *et al.* Role of P53, E-cadherin and BRAF as predictors of regional nodal recurrence for papillary thyroid cancer. *Ann Diagn Pathol* **40**, 59-65, doi:10.1016/j.anndiagpath.2019.04.005 (2019).

27    Hugo, H. J. *et al.* Epithelial requirement for in vitro proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: oncogenic rather than tumor-suppressive role of E-cadherin. *Breast cancer research : BCR* **19**, 86, doi:10.1186/s13058-017-0880-z (2017).

28    Semb, H. & Christofori, G. The tumor-suppressor function of E-cadherin. *American journal of human genetics* **63**, 1588-1593, doi:10.1086/302173 (1998).

29    Fidler, I. J. & Poste, G. The "seed and soil" hypothesis revisited. *Lancet Oncol* **9**, 808, doi:10.1016/S1470-2045(08)70201-8 (2008).

30    Brabletz, T. *et al.* Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10356-10361, doi:10.1073/pnas.171610498 (2001).

31    Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472-476, doi:10.1038/nature15748 (2015).

32    Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525-530, doi:10.1038/nature16064 (2015).

33    Tarin, D., Thompson, E. W. & Newgreen, D. F. The fallacy of epithelial mesenchymal transition in neoplasia. *Cancer Res* **65**, 5996-6000; discussion 6000-5991, doi:10.1158/0008-5472.CAN-05-0699 (2005).

34    Rodriguez-Gonzalez, F. G., Mustafa, D. A., Mostert, B. & Sieuwerts, A. M. The challenge of gene expression profiling in heterogeneous clinical samples. *Methods* **59**, 47-58, doi:10.1016/j.ymeth.2012.05.005 (2013).

35    Alkatout, I. *et al.* Transcription factors associated with epithelial-mesenchymal transition and cancer stem cells in the tumor centre and margin of invasive breast cancer. *Exp Mol Pathol* **94**, 168-173, doi:10.1016/j.yexmp.2012.09.003 (2013).

36    Singh, A. & Settleman, J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29**, 4741-4751, doi:10.1038/onc.2010.215 (2010).

37    McKeithen, D., Graham, T., Chung, L. W. & Odero-Marah, V. Snail transcription factor regulates neuroendocrine differentiation in LNCaP prostate cancer cells. *Prostate* **70**, 982-992, doi:10.1002/pros.21132 (2010).

38    Haslehurst, A. M. *et al.* EMT transcription factors snail and slug directly contribute to cisplatin resistance in ovarian cancer. *BMC Cancer* **12**, 91, doi:10.1186/1471-2407-12-91 (2012).

39    Vesuna, F. *et al.* Twist contributes to hormone resistance in breast cancer by downregulating estrogen receptor-alpha. *Oncogene* **31**, 3223-3234, doi:10.1038/onc.2011.483 (2012).

40    Santamaria, P. G., Moreno-Bueno, G. & Cano, A. Contribution of Epithelial Plasticity to Therapy Resistance. *Journal of clinical medicine* **8**, doi:10.3390/jcm8050676 (2019).

41    Shintani, Y. *et al.* IL-6 Secreted from Cancer-Associated Fibroblasts Mediates Chemoresistance in NSCLC by Increasing Epithelial-Mesenchymal Transition Signaling. *J Thorac Oncol* **11**, 1482-1492, doi:10.1016/j.jtho.2016.05.025 (2016).

42    Oshimori, N., Oristian, D. & Fuchs, E. TGF-beta promotes heterogeneity and drug resistance in squamous cell carcinoma. *Cell* **160**, 963-976, doi:10.1016/j.cell.2015.01.043 (2015).

43    Li, Z. H., Hu, P. H., Tu, J. H. & Yu, N. S. Luminal B breast cancer: patterns of recurrence and clinical outcome. *Oncotarget* **7**, 65024-65033, doi:10.18632/oncotarget.11344 (2016).

44    Tran, B. & Bedard, P. L. Luminal-B breast cancer and novel therapeutic targets. *Breast cancer research : BCR* **13**, 221, doi:10.1186/bcr2904 (2011).

45    Dai, X. *et al.* Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research* **5**, 2929-2943 (2015).

46    Sotiriou, C. & Pusztai, L. Gene-expression signatures in breast cancer. *The New England journal of medicine* **360**, 790-800, doi:10.1056/NEJMra0801289 (2009).

47    Milioli, H. H., Tishchenko, I., Riveros, C., Berretta, R. & Moscato, P. Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. *BMC medical genomics* **10**, 19, doi:10.1186/s12920-017-0250-9 (2017).

48    Dai, X., Cheng, H., Bai, Z. & Li, J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J Cancer* **8**, 3131-3141, doi:10.7150/jca.18457 (2017).

49     Mackay, A. *et al.* Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. *Journal of the National Cancer Institute* **103**, 662-673, doi:10.1093/jnci/djr071 (2011).

50     Fan, C. *et al.* Concordance among gene-expression-based predictors for breast cancer. *The New England journal of medicine* **355**, 560-569, doi:10.1056/NEJMoa052933 (2006).

51     Hu, Z. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC genomics* **7**, 96, doi:10.1186/1471-2164-7-96 (2006).

52     Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527, doi:10.1016/j.ccr.2006.10.008 (2006).

53     Loi, S. *et al.* Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **25**, 1239-1246, doi:10.1200/JCO.2006.07.1522 (2007).

54     de Ronde, J. J. *et al.* Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response. *Breast Cancer Res Treat* **119**, 119-126, doi:10.1007/s10549-009-0499-6 (2010).

55     Carey, L. A. *et al.* The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13**, 2329-2334, doi:10.1158/1078-0432.CCR-06-1109 (2007).

56     Rouzier, R. *et al.* Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 5678-5685, doi:10.1158/1078-0432.CCR-04-2421 (2005).

57     Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45**, 1113-1120, doi:10.1038/ng.2764 (2013).

58     Bergamaschi, A. *et al.* Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* **45**, 1033-1040, doi:10.1002/gcc.20366 (2006).

59     Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10**, 529-541, doi:10.1016/j.ccr.2006.10.009 (2006).

60     Smid, M. *et al.* Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68**, 3108-3114, doi:10.1158/0008-5472.CAN-07-5644 (2008).

61     Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752, doi:10.1038/35021093 (2000).

62     Sorlie, T. *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* **98**, 10869-10874, doi:10.1073/pnas.191367098 (2001).

63     Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8418-8423, doi:10.1073/pnas.0932692100 (2003).

64     Toft, D. J. & Cryns, V. L. Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies. *Mol Endocrinol* **25**, 199-211, doi:10.1210/me.2010-0164 (2011).

65     Yang, X. R. *et al.* Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev* **16**, 439-443, doi:10.1158/1055-9965.EPI-06-0806 (2007).

66     Millikan, R. C. *et al.* Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat* **109**, 123-139, doi:10.1007/s10549-007-9632-6 (2008).

67     Livasy, C. A. *et al.* Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **19**, 264-271, doi:10.1038/modpathol.3800528 (2006).

68     Agarwal, R. *et al.* Integrative analysis of cyclin protein levels identifies cyclin b1 as a classifier and predictor of outcomes in breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15**, 3654-3662, doi:10.1158/1078-0432.CCR-08-3293 (2009).

69     Keyomarsi, K. *et al.* Cyclin E and survival in patients with breast cancer. *The New England journal of medicine* **347**, 1566-1575, doi:10.1056/NEJMoa021153 (2002).

70     Voduc, D., Nielsen, T. O., Cheang, M. C. & Foulkes, W. D. The combination of high cyclin E and Skp2 expression in breast cancer is associated with a poor prognosis and the basal phenotype. *Human pathology* **39**, 1431-1437, doi:10.1016/j.humpath.2008.03.004 (2008).

71     Sherr, C. J. & McCormick, F. The RB and p53 pathways in cancer. *Cancer Cell* **2**, 103-112 (2002).

72    Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* **5**, 341-354, doi:10.1038/nrc1609 (2005).

73    Aas, T. *et al.* Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nature medicine* **2**, 811-814 (1996).

74    Berns, E. M. *et al.* Complete sequencing of TP53 predicts poor response to systemic therapy of advanced breast cancer. *Cancer Res* **60**, 2155-2162 (2000).

75    Bergh, J., Norberg, T., Sjogren, S., Lindgren, A. & Holmberg, L. Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. *Nature medicine* **1**, 1029-1034 (1995).

76    Shih, C., Padhy, L. C., Murray, M. & Weinberg, R. A. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature* **290**, 261-264 (1981).

77    Krishnamurti, U. & Silverman, J. F. HER2 in breast cancer: a review and update. *Adv Anat Pathol* **21**, 100-107, doi:10.1097/PAP.0000000000000015 (2014).

78    Gutierrez, C. & Schiff, R. HER2: biology, detection, and clinical implications. *Arch Pathol Lab Med* **135**, 55-62, doi:10.1043/2010-0454-RAR.1 (2011).

79    Kallioniemi, O. P. *et al.* ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5321-5325 (1992).

80    Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-182 (1987).

81    Bose, R. *et al.* Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer discovery* **3**, 224-237, doi:10.1158/2159-8290.CD-12-0349 (2013).

82    Giordano, S. H. *et al.* Systemic Therapy for Patients With Advanced Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer: ASCO Clinical Practice Guideline Update. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **36**, 2736-2740, doi:10.1200/JCO.2018.79.2697 (2018).

83    Ahmed, S., Sami, A. & Xiang, J. HER2-directed therapy: current treatment options for HER2-positive breast cancer. *Breast Cancer* **22**, 101-116, doi:10.1007/s12282-015-0587-x (2015).

84    Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**, R76, doi:10.1186/gb-2007-8-5-r76 (2007).

85    Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research : BCR* **12**, R68, doi:10.1186/bcr2635 (2010).

86    West, R. B. *et al.* Determination of stromal signatures in breast carcinoma. *PLoS biology* **3**, e187, doi:10.1371/journal.pbio.0030187 (2005).

87    Palmer, C., Diehn, M., Alizadeh, A. A. & Brown, P. O. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC genomics* **7**, 115, doi:10.1186/1471-2164-7-115 (2006).

88    Peters, L. M. *et al.* Mutation of a transcription factor, TFCP2L3, causes progressive autosomal dominant hearing loss, DFNA28. *Human molecular genetics* **11**, 2877-2885, doi:10.1093/hmg/11.23.2877 (2002).

89    Wang, S. & Samakovlis, C. Grainy head and its target genes in epithelial morphogenesis and wound healing. *Current topics in developmental biology* **98**, 35-63, doi:10.1016/B978-0-12-386499-4.00002-1 (2012).

90    WILANOWSKI, T., CERRUTI, L., Lin-Lin, Z. & CUNNINGHAM, J. M. The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. *Biochemical Journal* **370**, 953-962 (2003).

91    Reese, R. M., Harrison, M. M. & Alarid, E. T. Grainyhead-like Protein 2: The Emerging Role in Hormone-Dependent Cancers and Epigenetics. *Endocrinology* **160**, 1275-1288, doi:10.1210/en.2019-00213 (2019).

92    Wilanowski, T. *et al.* A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. *Mechanisms of development* **114**, 37-50 (2002).

93    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* **240**, 1396-1401, doi:10.1177/1535370215588924 (2015).

94    Werth, M. *et al.* The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835-3845, doi:10.1242/dev.055483 (2010).

95     Auden, A. *et al.* Spatial and temporal expression of the Grainyhead-like transcription factor family during murine development. *Gene Expr Patterns* **6**, 964-970, doi:10.1016/j.modgep.2006.03.011 (2006).

96     Blastyak, A., Mishra, R. K., Karch, F. & Gyurkovics, H. Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Mol Cell Biol* **26**, 1434-1444, doi:10.1128/MCB.26.4.1434-1444.2006 (2006).

97     Cieply, B. *et al.* Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res* **72**, 2440-2453, doi:10.1158/0008-5472.CAN-11-4038 (2012).

98     Chung, V. Y. *et al.* GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943, doi:10.1038/srep19943 (2016).

99     Wilanowski, T. *et al.* Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. *The EMBO journal* **27**, 886-897, doi:10.1038/emboj.2008.24 (2008).

100    Wilanowski, T. *et al.* A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. *Mechanisms of development* **114**, 37-50 (2002).

101    Ma, L., Yan, H., Zhao, H. & Sun, J. Grainyhead-like 2 in development and cancer. *Tumour Biol* **39**, 1010428317698375, doi:10.1177/1010428317698375 (2017).

102    Ray, H. J. & Niswander, L. A. Grainyhead-like 2 downstream targets act to suppress epithelial-to-mesenchymal transition during neural tube closure. *Development* **143**, 1192-1204, doi:10.1242/dev.129825 (2016).

103    Fong, K. S. *et al.* A mutation in the tuft mouse disrupts TET1 activity and alters the expression of genes that are crucial for neural tube closure. *Dis Model Mech* **9**, 585-596, doi:10.1242/dmm.024109 (2016).

104    Vona, B., Nanda, I., Neuner, C., Muller, T. & Haaf, T. Confirmation of GRHL2 as the gene for the DFNA28 locus. *Am J Med Genet A* **161A**, 2060-2065, doi:10.1002/ajmg.a.36017 (2013).

105    Lin, X. *et al.* GRHL2 genetic polymorphisms may confer a protective effect against sudden sensorineural hearing loss. *Mol Med Rep* **13**, 2857-2863, doi:10.3892/mmr.2016.4871 (2016).

106    Han, Y. *et al.* Grhl2 deficiency impairs otic development and hearing ability in a zebrafish model of the progressive dominant hearing loss DFNA28. *Human molecular genetics* **20**, 3213-3226, doi:10.1093/hmg/ddr234 (2011).

107    Petrof, G. *et al.* Mutations in GRHL2 result in an autosomal-recessive ectodermal Dysplasia syndrome. *American journal of human genetics* **95**, 308-314, doi:10.1016/j.ajhg.2014.08.001 (2014).

108    Carpinelli, M. R., de Vries, M. E., Jane, S. M. & Dworkin, S. Grainyhead-like Transcription Factors in Craniofacial Development. *Journal of dental research* **96**, 1200-1209, doi:10.1177/0022034517719264 (2017).

109    Chen, C. P. *et al.* Prenatal diagnosis of an 8q22.2-q23.3 deletion associated with bilateral cleft lip and palate and intrauterine growth restriction on fetal ultrasound. *Taiwan J Obstet Gynecol* **56**, 843-846, doi:10.1016/j.tjog.2017.10.026 (2017).

110    Gao, X., Bali, A. S., Randell, S. H. & Hogan, B. L. GRHL2 coordinates regeneration of a polarized mucociliary epithelium from basal stem cells. *J Cell Biol* **211**, 669-682, doi:10.1083/jcb.201506014 (2015).

111    Venkatesan, K., McManus, H. R., Mello, C. C., Smith, T. F. & Hansen, U. Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. *Nucleic acids research* **31**, 4304-4316 (2003).

112    Boglev, Y. *et al.* The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. *Developmental biology* **349**, 512-522, doi:10.1016/j.ydbio.2010.11.011 (2011).

113    Ting, S. B. *et al.* A homolog of Drosophila grainy head is essential for epidermal integrity in mice. *Science* **308**, 411-413, doi:10.1126/science.1107511 (2005).

114    Yu, Z. *et al.* Grainyhead-like factor Get1/Grhl3 regulates formation of the epidermal leading edge during eyelid closure. *Developmental biology* **319**, 56-67, doi:10.1016/j.ydbio.2008.04.001 (2008).

115    Greene, N. D. & Copp, A. J. Mouse models of neural tube defects: investigating preventive mechanisms. *Am J Med Genet C Semin Med Genet* **135C**, 31-41, doi:10.1002/ajmg.c.30051 (2005).

116    Brodeur, G. M. Neuroblastoma: biological insights into a clinical enigma. *Nat Rev Cancer* **3**, 203-216, doi:10.1038/nrc1014 (2003).

117     Fabian, J. *et al.* GRHL1 acts as tumor suppressor in neuroblastoma and is negatively regulated by MYCN and HDAC3. *Cancer Res* **74**, 2604-2616, doi:10.1158/0008-5472.CAN-13-1904 (2014).

118     Darido, C. *et al.* Targeting of the tumor suppressor GRHL3 by a miR-21-dependent proto-oncogenic network results in PTEN loss and tumorigenesis. *Cancer Cell* **20**, 635-648, doi:10.1016/j.ccr.2011.10.014 (2011).

119     Guardiola-Serrano, F. *et al.* Gene trapping identifies a putative tumor suppressor and a new inducer of cell migration. *Biochemical and biophysical research communications* **376**, 748-752, doi:10.1016/j.bbrc.2008.09.070 (2008).

120     Lukosz, M., Mlynek, A., Czypiorski, P., Altschmied, J. & Haendeler, J. The transcription factor Grainyhead like 3 (GRHL3) affects endothelial cell apoptosis and migration in a NO-dependent manner. *Biochemical and biophysical research communications* **412**, 648-653, doi:10.1016/j.bbrc.2011.08.018 (2011).

121     Panis, C., Pizzatti, L., Herrera, A. C., Cecchini, R. & Abdelhay, E. Putative circulating markers of the early and advanced stages of breast cancer identified by high-resolution label-free proteomics. *Cancer Lett* **330**, 57-66, doi:10.1016/j.canlet.2012.11.020 (2013).

122     Xu, H. *et al.* Clinical implications of GRHL3 protein expression in breast cancer. *Tumour Biol* **35**, 1827-1831, doi:10.1007/s13277-013-1244-7 (2014).

123     Alotaibi, H. *et al.* Enhancer cooperativity as a novel mechanism underlying the transcriptional regulation of E-cadherin during mesenchymal to epithelial transition. *Biochim Biophys Acta* **1849**, 731-742, doi:10.1016/j.bbagrm.2015.01.005 (2015).

124     Haendeler, J. *et al.* Two isoforms of Sister-Of-Mammalian Grainyhead have opposing functions in endothelial cells and in vivo. *Arteriosclerosis, thrombosis, and vascular biology* **33**, 1639-1646, doi:10.1161/ATVBAHA.113.301428 (2013).

125     Garnis, C., Coe, B. P., Zhang, L., Rosin, M. P. & Lam, W. L. Overexpression of LRP12, a gene contained within an 8q22 amplicon identified by high-resolution array CGH analysis of oral squamous cell carcinomas. *Oncogene* **23**, 2582-2586, doi:10.1038/sj.onc.1207367 (2004).

126     Tanner, S. M. *et al.* BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13901-13906, doi:10.1073/pnas.241525498 (2001).

127     Li, Y. *et al.* Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nature medicine* **16**, 214-218, doi:10.1038/nm.2090 (2010).

128     Paltoglou, S. *et al.* Novel androgen receptor co-regulator GRHL2 exerts both oncogenic and anti-metastatic functions in prostate cancer. *Cancer Res*, doi:10.1158/0008-5472.CAN-16-1616 (2017).

129     Cheng, L. *et al.* Identification of genes with a correlation between copy number and expression in gastric cancer. *BMC medical genomics* **5**, 14, doi:10.1186/1755-8794-5-14 (2012).

130     Dompe, N. *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* **108**, E943-951, doi:10.1073/pnas.1100132108 (2011).

131     Kang, X., Chen, W., Kim, R. H., Kang, M. K. & Park, N. H. Regulation of the hTERT promoter activity by MSH2, the hnRNPs K and D, and GRHL2 in human oral squamous cell carcinoma cells. *Oncogene* **28**, 565-574, doi:10.1038/onc.2008.404 (2008).

132     Tanaka, Y. *et al.* Gain of GRHL2 is associated with early recurrence of hepatocellular carcinoma. *J Hepatol* **49**, 746-757, doi:10.1016/j.jhep.2008.06.019 (2008).

133     Paltoglou, S. *et al.* Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer. *Cancer Res* **77**, 3417-3430, doi:10.1158/0008-5472.CAN-16-1616 (2017).

134     Pan, X. *et al.* GRHL2 suppresses tumor metastasis via regulation of transcriptional activity of RhoG in non-small cell lung cancer. *Am J Transl Res* **9**, 4217-4226 (2017).

135     Korpal, M. *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nature medicine* **17**, 1101-1108, doi:10.1038/nm.2401 (2011).

136     Xiang, X. *et al.* Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One* **7**, e50781, doi:10.1371/journal.pone.0050781 (2012).

137     Tan, T. Z. *et al.* Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med* **6**, 1279-1293, doi:10.15252/emmm.201404208 (2014).

138     Senga, K., Mostov, K. E., 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. *Mol Biol Cell* **23**, 2845-2855, doi:10.1091/mbc.E12-02-0097 (2012).

139    Cieply, B., Farris, J., Denvir, J., Ford, H. L. & Frisch, S. M. Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res* **73**, 6299-6309, doi:10.1158/0008-5472.CAN-12-4082 (2013).

140    Aue, A. *et al.* A Grainyhead-Like 2/Ovo-Like 2 Pathway Regulates Renal Epithelial Barrier Function and Lumen Expansion. *J Am Soc Nephrol* **26**, 2704-2715, doi:10.1681/ASN.2014080759 (2015).

141    Pyrgaki, C., Liu, A. & Niswander, L. Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Developmental biology* **353**, 38-49, doi:10.1016/j.ydbio.2011.02.027 (2011).

142    Chen, W. *et al.* Grainyhead-like 2 regulates epithelial plasticity and stemness in oral cancer cells. *Carcinogenesis* **37**, 500-510, doi:10.1093/carcin/bgw027 (2016).

143    Quan, Y. *et al.* Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. *Cancer Biol Ther* **15**, 878-887, doi:10.4161/cbt.28877 (2014).

144    Mooney, S. M. *et al.* The GRHL2/ZEB Feedback Loop-A Key Axis in the Regulation of EMT in Breast Cancer. *Journal of cellular biochemistry* **118**, 2559-2570, doi:10.1002/jcb.25974 (2017).

145    Laurikkala, J. *et al.* p63 regulates multiple signalling pathways required for ectodermal organogenesis and differentiation. *Development* **133**, 1553-1563, doi:10.1242/dev.02325 (2006).

146    Barbareschi, M. *et al.* p63, a p53 homologue, is a selective nuclear marker of myoepithelial cells of the human breast. *Am J Surg Pathol* **25**, 1054-1060 (2001).

147    Mehrazarin, S. *et al.* The p63 Gene Is Regulated by Grainyhead-like 2 (GRHL2) through Reciprocal Feedback and Determines the Epithelial Phenotype in Human Keratinocytes. *The Journal of biological chemistry* **290**, 19999-20008, doi:10.1074/jbc.M115.659144 (2015).

148    Chen, W. *et al.* Grainyhead-like 2 (GRHL2) inhibits keratinocyte differentiation through epigenetic mechanism. *Cell Death Dis* **3**, e450, doi:10.1038/cddis.2012.190 (2012).

# Chapter 2

# The role of GRHL2 in luminal- and basal like breast cancer

Zi Wang[1], Bircan Coban[1], Chen-Yi Liao[1], Yao-Jun Chen[1], Erik HJ Danen[1, *]

[1]Division of Drug Discovery and Safety, Leiden Academic Center for Drug Research, Leiden University, The Netherlands

* Corresponding author: Erik HJ Danen; e.danen@lacdr.leidenuniv.nl; Division of Drug Discovery and Safety, Leiden Academic Center for Drug Research, Leiden University, Einsteinweg 55, 2333CC Leiden, The Netherlands

# Abstract

The transcription factor Grainyhead like 2 (GRHL2) is reported to promote cancer growth in some-, and suppress aspects of cancer progression in other studies. We investigated its role in different breast cancer subtypes. In breast cancer patients, association of GRHL2 expression with prognosis differed for different subtypes and in a large cell line panel, GRHL2 was low or absent in basal B- and expressed in all luminal- and basal A cell lines. In a luminal cell line (MCF7) deletion of GRHL2 triggered cell cycle arrest, loss of colony formation capacity, and downregulation of PCNA and TERT. In parallel, E-cadherin was lost but only a minor increase in EGF-stimulated motility was observed. In a basal A cell line (HCC1806) GRHL2 deletion also suppressed proliferation and colony formation but no changes were seen in PCNA and TERT. Rather, loss of E-cadherin in this case was accompanied by induction of Vimentin and N-cadherin, and conversion to a highly migratory phenotype, further augmented by EGF treatment. These results point to distinct responses to GRHL2 depletion in luminal- and basal-like breast cancers with respect to growth arrest and enhanced motility phenotypes and suggest that GRHL2 may be a candidate target in luminal-like breast cancer.

# Introduction

Breast cancer is the most prevalent malignancy in female globally. Mortality of patients with breast cancer has decreased, resulting from early diagnosis and development of therapies [1-3]. A considerable proportion of knowledge on breast cancer originates from experiments performed with breast cancer cells that cover the various subtypes of this heterogeneous disease [4]. Breast cancer is divided into luminal-like (luminal A and luminal B), epidermal growth factor receptor 2-enriched (HER2-enriched), basal-like (basal A and basal B), claudin-low, and normal-like subtypes based on gene expression profiling [5].

Luminal-like breast cancer is characterized by enrichment of genes/proteins associated with the luminal epithelial phenotype (e.g., ESR1, GATA3 and FOXA1) [4,6]. Basal-like breast cancer is characterized by significant enrichment of basal epithelial cytokeratins, hormone receptor negativity and a high tumor grade and poor prognosis

[7]. Basal-like breast cancer can be further divided into basal A and basal B subtype [8]. The basal A subtype is enriched with basal markers such as cytokeratins, while basal-B exhibits a mesenchymal or a normal-like phenotype with overexpression of several genes related to tumor invasion and tumor stemness [4].

The Grainyhead (*GRH*) gene was originally discovered through a mutation that causes slack and fragile cuticles in *Drosophila*[9]. This gene mutation results in failure of neural tube closure during embryogenesis [10]. The transcription factor GRH family is highly conserved from *Drosophila* to humans. In humans, GRHL1, GRHL2 and GRHL3 are identified as GRH homologs that contain an N-terminal transcriptional activation domain, a central CP2 DNA-binding domain and a C-terminal dimerization domain [11].

GRHL2 has been implicated in cancer development. In some studies, GRHL2 is considered as a tumor suppressor, because it suppresses epithelial-mesenchymal transition (EMT) through upregulation of epithelial markers or downregulation of mesenchymal markers [12-14]. In contrast, *GRHL2* is located on chromosome 8q22 that is frequently amplified or overexpressed in many cancers and hence may rather have an oncogenic function [15]. Indeed, in prostate cancer [16], breast cancer [14], lung cancer [17] and ovarian cancer [18] downregulation of GRHL2 has been associated with inhibition of cell proliferation. Together, this suggests that GRHL2 function may vary depending on the cancer cell context.

In this study, we investigated the role of GRHL2 in different breast cancer subtypes. Our findings show that GRHL2 is absent in basal B-like breast cancer cells, it is expressed in basal A where its deletion triggers a slow growth/high motility phenotype, and it is expressed in luminal-like breast cancer cells where its depletion causes an arrested replication/proliferation.

## Materials and methods

### Cell lines
Human breast cancer cell lines (MCF7, T47D, BT474, HCC1806, BT20, MDA-MB-468, Hs578T) were obtained from the American Type Culture Collection. 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% $CO_2$).

For production of lentiviral particles, VSV, GAG, REV and Cas9 or sgRNA 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 Edit-R Tre3G promotor-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 or GRHL2-specific single guide (sg)RNAs (Sigma) and selected by puromycin.

## 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)). Then cell lysis was sonicated and protein concentration was determined by bicinchoninic acid assay (BCA) assay. Cell lysis was mixed with protein loading buffer. Subsequently, protein was separated by SDS-PAGE gel and then transferred to methanol-activated polyvinylidene difluoride (PVDF) (Milipore, The Netherlands) membrane. The membrane was blocked with 5% bovine serum albumin (BSA, Sigma-Aldrich) for 1 hour at room temperature (RT). Then 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 by Amersham Imager 600 (GE Healthcare Life science, The Netherlands). The following antibodies were used: GRHL2 (Atlas-Antibodies, hpa004820), GAPDH (SantaCruz, sc-32233), PCNA (SantaCruz, sc-56), Vimentin (Abcam, ab8069), N-cadherin (BD, 610920), E-cadherin (Abcam, ab76055), Peroxidase AffiniPure Goat Anti-Rabbit IgG (Jackson ImmunoResearch, 111-035-003), Peroxidase AffiniPure Goat Anti-Mouse IgG (Jackson ImmunoResearch, 115-035-003). Original blots are shown in Supplementary Figures.

## Sulforhodamine B (SRB) assay

Cell proliferation rate was measured by SRB assay. 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.

## Realtime quantitative PCR (RT-qPCR)

Total RNA was isolated using RNEasy Plus Mini Kit (Qiagen). 500 ng RNA was reverse-transcribed into cDNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). The cDNA was mixed with SYBR green master mix (Fisher Scientific) for qPCR. RT-qPCR data were collected and analyzed using $2^{-\Delta\Delta Ct}$ method. The primers are shown in supplementary table 1.

## Immunofluorescence

Cells were fixed with 2% formaldehyde for 15 minutes under slow rotation, permeabilized with 1% Triton in Phosphate buffered saline (PBS) for 10 minutes, and then stained with primary antibodies and secondary antibodies. The following antibodies were used: Vimentin (Abcam, ab8069); E-cadherin (Abcam, ab76055); Hoechst (33258, Abcam); β-catenin (BD, 610153 ); Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody-Alexa Fluor 488 (Thermo Fisher, A-11001); Rhodamine-Phalloidin (Thermo Fisher, R415).

## Three-dimensional (3D) culture

Collagen (2 mg/ml, 70 μl/well) was added into 96-well plates. Plates were kept in the incubator (37°C, 5% $CO_2$) for 1 hour. A subconfluent layer of cells in a T25 flask was detached by 0.25% trypsin and collected. The cell suspension was centrifuged at 230×g for 5 minutes. The cell pellet was resuspended in 50 μl 2% Polyvinylpyrrolidone. The cell suspension was injected into the collagen scaffolds to generate ~200 μm diameter tumor spheroids as described previously [19]. The plates were kept in the incubator and spheroids were observed at the indicated time points under a Nikon

ECLIPSE Ti confocal microscope. At the end of the experiment, plates were incubated with 0.1% Triton, 3.7% formaldehyde, Phalloidin-Rhodamin, and 400 ng/ml Hoechst for 48 hours at 4°C. Plates were washed three times with PBS, and imaged using a Nikon ECLIPSE Ti2 confocal microscope.

### Random migration assay

96 well-plates were coated with collagen (50 µl/well, 20 µg/ml) 1 hour 37°C and washed with PBS. Cells were seeded into the coated 96-well plates at the density of 8000 cells/well overnight and stained with Hoechst (Thermo Fisher 33242) diluted 1:7500 for 45 minutes. Images were taken every 5 minutes on a Nikon TE confocal microscope for 12 hours, at two positions per well. Tracks were analyzed using NIS Elements software. For epidermal growth factor (EGF) stimulation, cells were seeded and cultured in collagen-coated 96-well plate overnight with serum free medium and 50 ng/ml EGF was added 1 hour before tracking cell migration.

### Cell cycle analysis

Cell cycle analysis was performed with a Click-iT EdU Flow Cytometry Kit (Invitrogen). Cells were cultured with 50 µm 5-ethynyl-2-deoxyuridine (EdU) for 4 h and fixed and stained according to the manufacturers protocol for analysis on a BD FACS Canto II.

### Colony formation assays

Cell survival was measured by colony formation assay.  450 cells were seeded into a well of 6-well plate after 4 days of doxycycline treatment. After 7 days, cells were fixed with 4% formaldehyde and stained with Giemsa. Images were analyzed by Image J (ColonyArea package).

### Expression analysis in breast cancer cohorts

The Cancer Genome Atlas [20] (TCGA) and Metabric breast cancer datasets [21] were analyzed for copy number alterations of GRHL2 and correlation with overall survival using cBioPortal [22]. The KM plotter database was analyzed to evaluate the association of GRHL2 expression with overall survival of patients with different subclasses of breast cancer[23].

A

GRHL2

22%

GRHL2 ▐████▌────────────────────▐██▌

Metabric dataset

18%

GRHL2 ▐████▌──────────────────

TCGA dataset

Genetic Alteration ▐ Amplification ▐ Deep deletion ▐ No alternations ▐ Not profiled

B



Metabric dataset

$P < 0.05$

TCGA dataset

$P = 0.06$

■ Cases with amplification(s) in GRHL2
■ Cases without amplification(s) in GRHL2

■ Cases with amplification(s) in GRHL2
■ Cases without amplification(s) in GRHL2

C

Luminal-like

logrank $P = 0.59$

Expression
— Low
— High

D

HER2

logrank $P = 0.31$

Expression
— Low
— High

E

Basal-like

logrank $P = 0.55$

Expression
— Low
— High

**(Last page) Fig. 1. GRHL2 expression in clinical breast cancer datasets. (A)** Oncoprints showing *GRHL2* copy-number alternations from Metabric and TCGA datasets, generated by cBioPortal. **(B)** Kaplan-Meier survival analysis from Metabric and TCGA dataset, generated in cBioPortal. **(C-F)** Kaplan-Meier survival analysis for luminal (ER+) **(C)**, HER2-enriched **(D)**, and basal-like **(E)** subtypes of breast cancer, generated by KM plotter.

**Statistical analyses**

Statistical analyses were performed by GraphPad Prism 8. Details of statistical tests used are shown in the figure legends.

# Results

**GRHL2 is associated with poor prognosis but downregulated in basal B subtype breast cancer**

In order to evaluate the clinical relevance of GRHL2 in breast cancer, GRHL2 alternations were examined in a series of published cohorts. *GRHL2* is located on chromosome 8q22.3, a genomic region that is frequently amplified or overexpressed in many cancers, including breast cancer and prostate cancer [16,24]. *GRHL2* gene amplification was detected in 22% to 18% of breast cancers (Fig. 1A). Kaplan-Meier survival analysis revealed that *GRHL2* gene copy number was associated with a trend toward a lower overall survival rate in patients with breast cancer although this was significant in a Metabric dataset but not in a TCGA dataset (Fig. 1B). We next explored the association of GRHL2 expression levels with overall survival in different breast cancer subtypes[25]. In luminal-like (ER+) breast cancers no association of GRHL2 expression with prognosis was found (Fig. 1C). In HER2-enriched breast cancer there was not significant association of high GRHL2 expression with poor prognosis (Fig. 1D). In basal-like breast cancer on the other hand, there was a slight trend that was not significant towards better prognosis for patients with higher GRHL2 levels (Fig. 1E).

Previous analysis of RNA-seq data organized for a large panel of human breast cancer cell lines showed that *GRHL2* was downregulated in the basal B subtype [12]. Our RNA-seq data further confirmed this finding (Fig. 2A and B). Notably, HCC1500 and SUM149PT may have been misclassified. SUM149PT has been classified as basal A or basal B subtype [8,26] and reported to contain different subpopulations, according to expression level of EpCAM and CD49f surface markers [27]. Likewise, HCC1500 cells

33
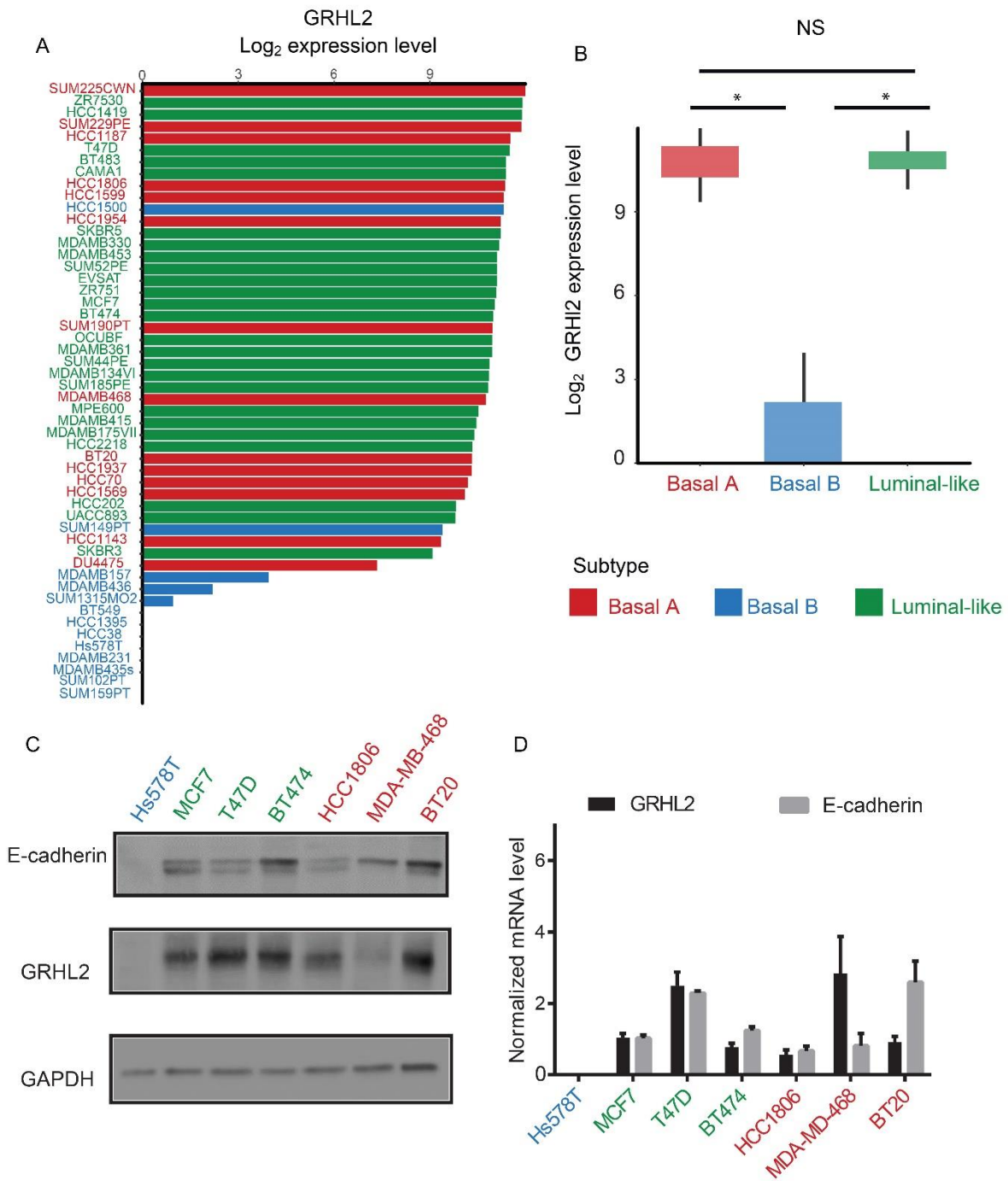
**Fig. 2. GRHL2 expression in a panel of human breast cancer cell lines representing different subtypes. (A and B)** *GRHL2* expression in a panel of >50 human breast cancer cell lines covering luminal-, basal A-, and basal B-like subtypes extracted from RNA-seq data. * indicates *p* < 0.05. Western blot analysis **(C)** and qRT-PCR **(D)** validating downregulation of GRHL2 and its target gene *CDH1* in basal B-like subtype breast cancer. Color codes refer to **B**

have been classified as basal A, due to a predominant population of cells that are positive for EpCAM and CD24 [3] or as basal B, owing to an enrichment for gene clusters associated with cancer stem cell- and invasive phenotypes [8]. To further validate
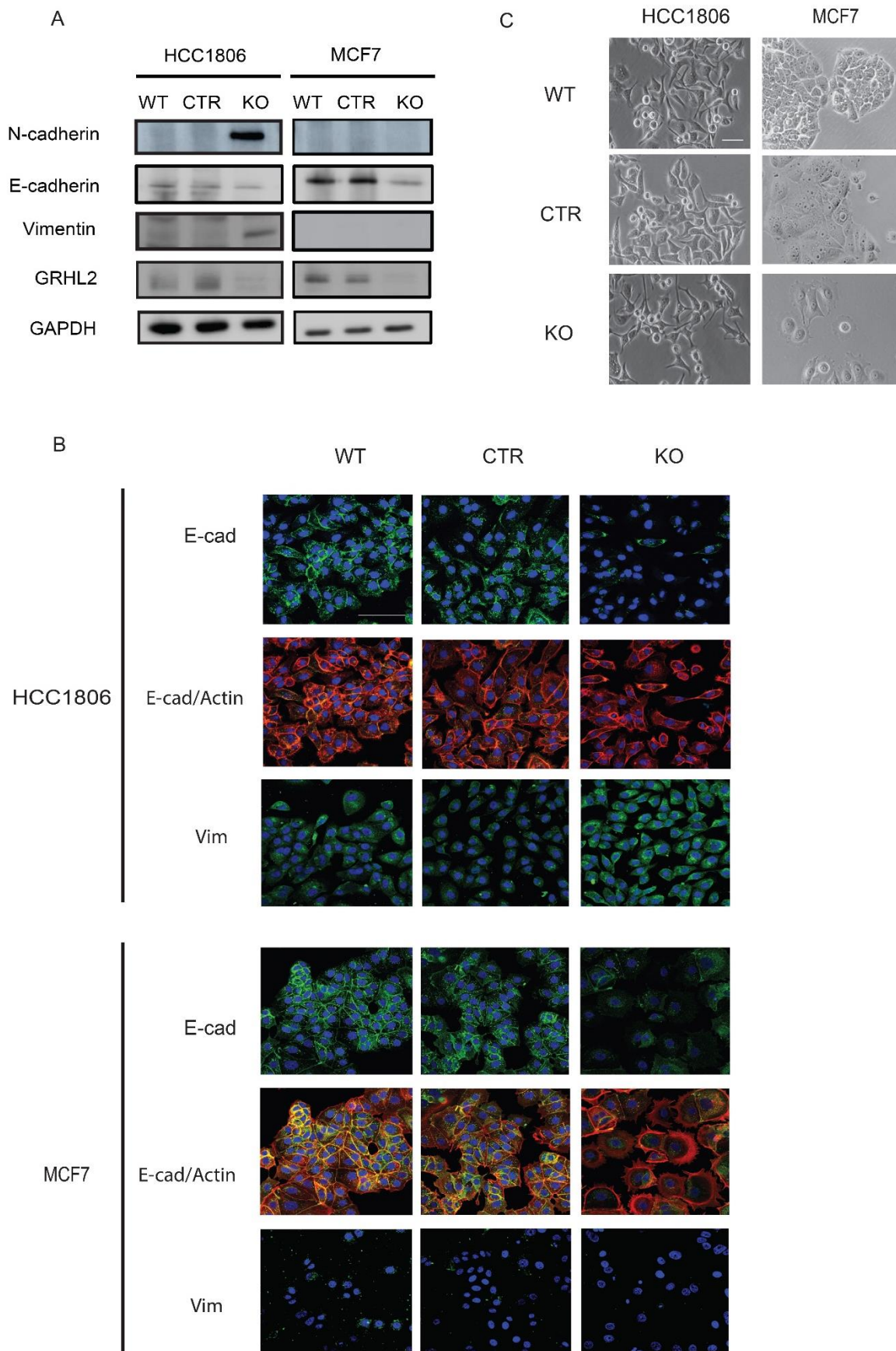
**(Last page) Fig. 3. Response to GRHL2 knockout in luminal- and basal A-like cells – aspects of EMT. (A)** Western blot analysis of the indicated proteins in wild type (WT) and sgCTR and sgGRHL2 transduced MCF7 and HCC1806 cells after 10 days doxycycline-induction. One experiment of two biological replicates is shown. **(B)** Immunofluorescence analysis of the indicated proteins in wild type (WT) and sgCTR and sgGRHL2 transduced MCF7 and HCC1806 cells after 10 days doxycycline-induction. One experiment of two biological replicates is shown. **(C)** Microphotographs showing morphology of wild type (WT) and sgCTR and sgGRHL2 transduced MCF7 and HCC1806 cells after 10 days doxycycline-induction. Scale bars in B and C, 100µm.

specific downregulation of GRHL2 in basal B, Western blot and qPCR were performed to detect GRHL2 protein and mRNA in a smaller panel of breast cancer cell lines. In agreement with the RNA-seq data, GRHL2 protein and mRNA were not detectable in Hs578T cells (basal B subtype), whereas it was expressed in basal A (HCC1806, MDA-MB-468 and BT20) and luminal-like (MCF7, T47D and BT474) subtypes (Fig. 2C and D).

E-cadherin (CDH1), a previously identified target gene of GRHL2, is a cell surface glycoprotein expressed in epithelial tissues that mediates cell-cell adhesion and is lost during EMT [14,28]. E-cadherin expression indeed correlated with GRHL2 expression and was not detectable in basal B subtype breast cancer, while basal A and luminal-like subtypes were positive for E-cadherin (Fig. 2C and D).

**EMT and migratory responses to GRHL2 loss in luminal versus basal A-like breast cancer cells**

Previous studies showed that loss of GRHL2 can be sufficient to trigger EMT [12]. We studied the response to GRHL2 loss in luminal and basal A-like cells using a conditional knockout approach. In both luminal (MCF7) and basal A-like cells (HCC1806) GRHL2 knockout, but not control sgRNA triggered a reduction in the epithelial marker, E-cadherin (Fig. 3A). However, the induction of mesenchymal markers, Vimentin and N-cadherin was only observed in HCC1806 cells. These Western blot results were confirmed using immunofluorescence. E-cadherin was expressed at cell–cell junctions and in the cytoplasm in HCC1806 and MCF7 cells and GRHL2 knockout led to reduced expression. A concomitant gain of Vimentin expression was only observed in HCC1806 cells (Fig. 3A and B). MCF7 cells showed reduced cell-cell contacts but still formed

islands and cells adopted a more flattened morphology whereas HCC1806 cells were more scattered in response to GRHL2 knockout (Fig. 3C).

**(Last page) Fig. 4. Response to GRHL2 knockout in luminal- and basal A-like cells – migratory behavior. (A)** Analysis of random migration assay showing the average path speed (y-axis) captured at the indicated timepoints during the assay (x-axis) for wild type (WT) and sgCTR and sgGRHL2 transduced MCF7 and HCC1806 cells after 10 days doxycycline-induction. Data are presented as mean ± SEM from 3 biological replicates. Data are statistically analyzed by two-way ANOVA. * indicates $p < 0.05$. **(B)** Analysis of random migration assay showing the average path speed (y-axis) captured at the indicated timepoints during the assay (x-axis) for wild type (WT) and sgCTR and sgGRHL2 transduced MC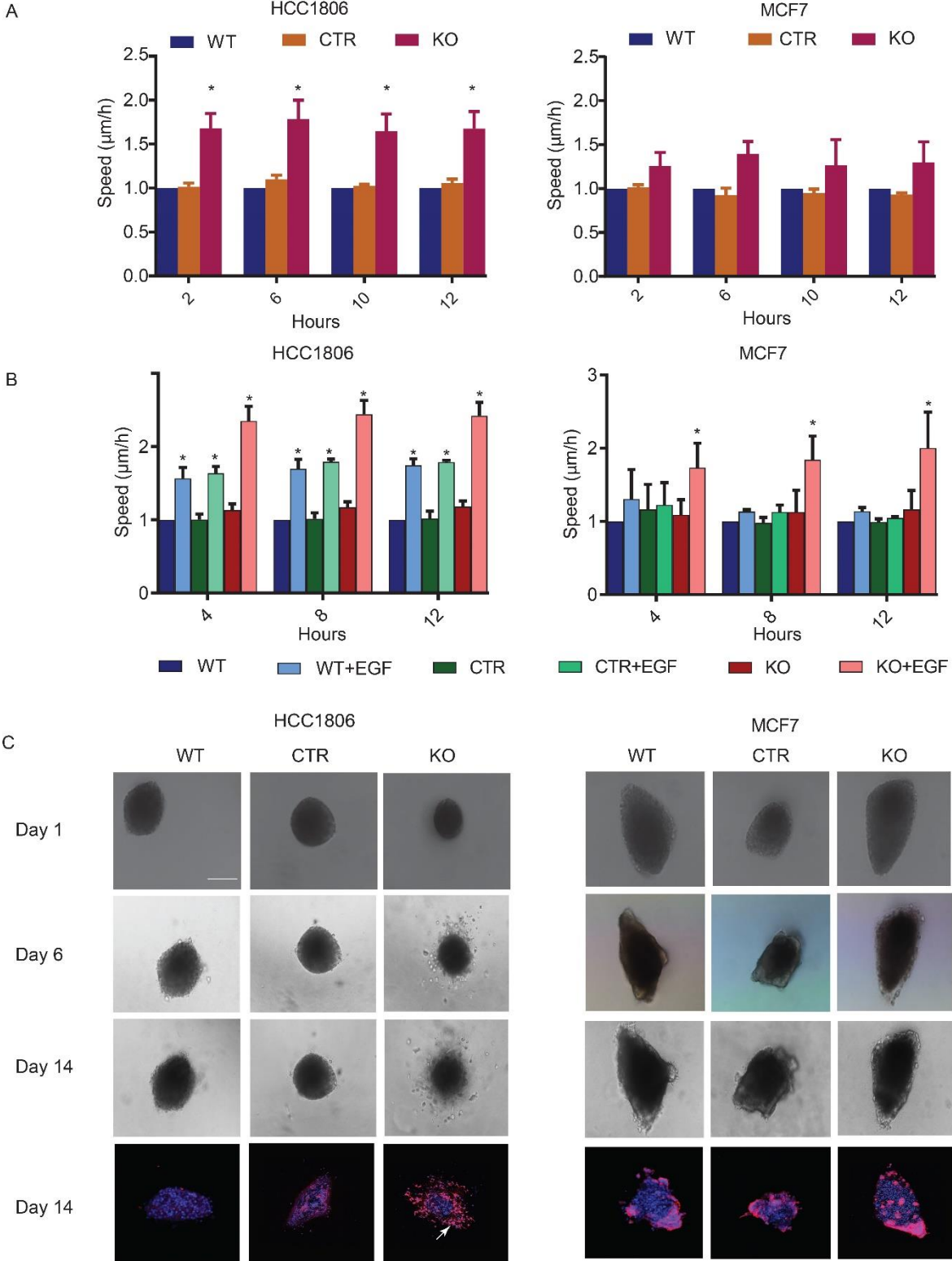F7 and HCC1806 cells after 10 days doxycycline-induction. Cells were left untreated or treated with 50 ng/ml EGF. Data are presented as mean ± SEM from 3 biological replicates. Data are statistically analyzed by two-way ANOVA. * indicates $p < 0.05$. **(C)** DIC (grey) and immunofluorescence images (blue, Hoechst; red, Rhodamin-Phalloidin) captured at the indicated timepoints after spheroid formation of collagen-embedded tumor spheroids derived from wild type (WT) and sgCTR and sgGRHL2 transduced HCC1806 and MCF7 cells after 10 days doxycycline-induction. Arrow shows invaded cells.

A shift from an epithelial to a mesenchymal phenotype is associated with enhanced migratory capability that may contribute to metastasis [31-33]. To investigate the effect of GRHL2 depletion on cell migration in HCC1806 and MCF7 cells, wild type, control sgRNA, and GRHL2 knockout cells were seeded into collagen-coated wells and intrinsic random migration was tracked. By calculating the mean square deviation (MSD) of the path length of each migrating cell, migration speed of the cells was determined. GRHL2 knockout led to an enhanced migration speed in HCC1806 cells whereas it did not significantly affect MCF7 cells (Fig. 4A).

EGF stimulates cell growth and migration in the mammary epithelium, by binding to the EGF receptor (EGFR) [34]. Previous studies showed that EGF promotes cell migration in basal B-like MDA-MB-231 cells [34]. We asked to what extent EGF-induced migration is modulated by GRHL2 loss and EMT status. EGF enhanced cell migration speed in HCC1806 cells and this effect was further increased upon GRHL2 loss (Figure 4B). By contrast, EGF failed to trigger cell migration in MCF7 but loss of GRHL2 resulted in enhanced cell migration in these cells.

Next, we investigated the effect of GRHL2 loss on the ability of luminal and basal A-like tumor spheroids to invade 3D extracellular matrix (ECM) scaffolds. MCF7 and HCC1806 tumor spheroids were generated in collagen matrices and invasion was analyzed over a 2-week period. Wild type and control sgRNA expressing HCC1806 spheroids did not show invasive behavior but GRHL2 knockout spheroids effectively

invaded into the surrounding collagen matrix (Fig. 4C). By contrast, MCF7 spheroids failed to invade the collagen matrix regardless of the presence or absence of GRHL2.

Together, these results indicate that loss of GRHL2 triggers several aspects associated with an EMT and leads to increased cell migration in basal A-like breast cancer cells whereas a partial EMT that enhances the response to EGF but does not lead to enhanced migratory potential per se, is induced in luminal-like breast cancer cells.
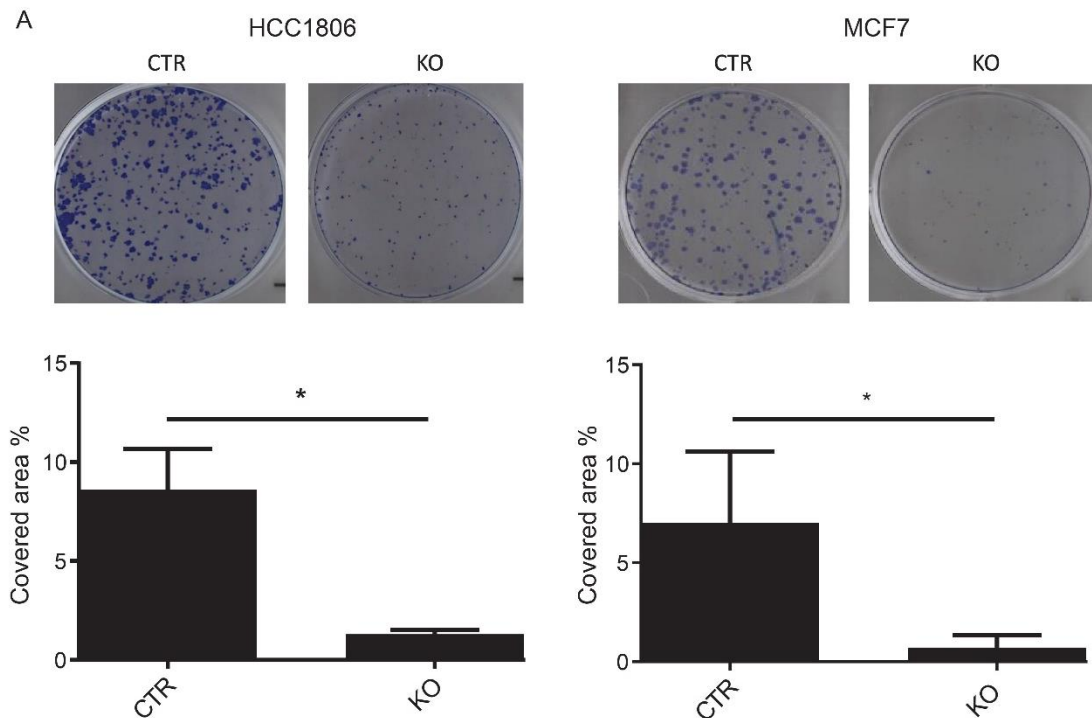


**Fig. 5. Response to GRHL2 knockout in luminal- and basal A-like cells – colony formation capacity.** Representative images and quantification of colony formation potential in control and GRHL2 depleted HCC1806 and MCF7 cells. Data are presented as mean ± SEM from 3 biological replicates. Data are statistically analyzed by t-test. * indicates $p < 0.05$.

## Effects on growth triggered by GRHL2 loss in luminal versus basal A-like breast cancer cells

GRHL2 not only regulates genes involved in epithelial cell adhesion but also supports replication and growth of epithelial cells[2]. To address how GRHL2 loss affected the latter properties in luminal and basal A-like cells, colony formation assays were performed[35]. GRHL2 knockout caused a significant decrease in clonogenic cell survival in both HCC1806 and MCF7 cells (Fig. 5). However, after GRHL2 loss the area
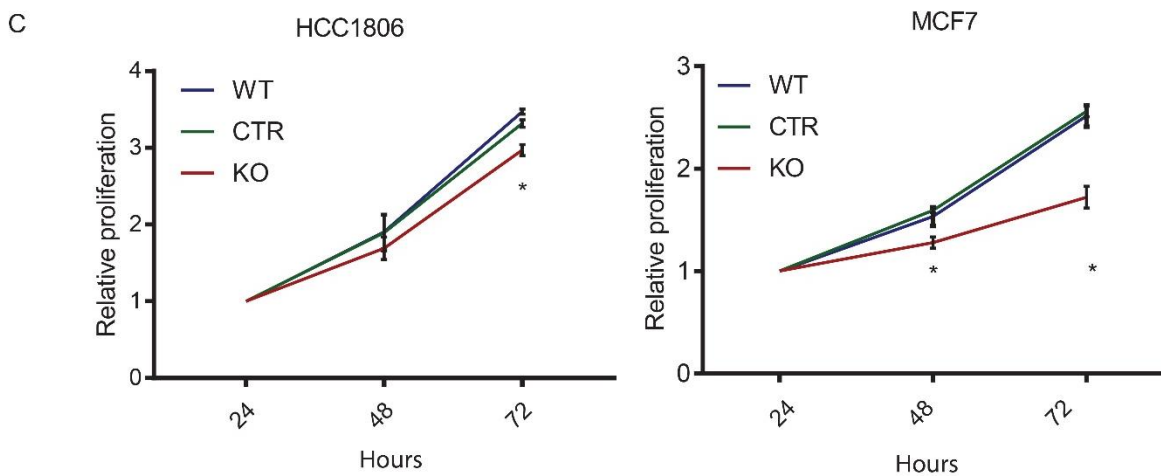
**Fig. 6. Response to GRHL2 knockout in luminal- and basal A-like cells – proliferative capacity.**
**(A and B)** Representative FACS profiles (A) and quantification of cell cycle phase distribution

**(B)** in control and GRHL2 depleted HCC1806 and MCF7 cells. Data are presented as mean ± SEM from 2 and 3 biological replicates for MCF7and HCC1806, respectively. Data are statistically analyzed by t-test comparing KO to CTR cells. * indicates $p < 0.05$. **(C)** Graphs showing results from SRB assay for wild type (WT) and sgCTR and sgGRHL2 transduced MCF7 and HCC1806 cells after 4 days doxycycline-induction and subsequent incubation for the indicated time periods. Data are presented as mean ± SEM from 3 biological replicates. Data are statistically analyzed by t-test comparing CTR and KO to WT.  * indicates $p < 0.05$.

covered by colonies in HCC1806 remained significantly higher than that in MCF7 cells. To address the effect of GRHL2 depletion on cell cycle progression, cell cycle analysis was performed. This demonstrated that a higher percentage of MCF7 cells were in G0/1 compared to HCC1806 and GRHL2 knockout resulted in a G0/1 arrest in MCF7 and a less-pronounced shift to G0/1 in HCC1806 cells (Fig. 6A and B).

To further address the effect of GRHL2 loss on proliferative potential, SRB assays were performed [36,37]. In agreement with the more robust block in colony formation and cell cycle progression observed in MCF7, GRHL2 knockout significantly inhibited cell proliferation of MCF7 cells at 2- and 3-days post seeding whereas a small but significant decrease in proliferation was observed at 3 days in HCC1806 (Fig. 6C).

We next examined candidate GRHL2-controlled genes that may underlie the observed distinct levels of suppression of cell proliferation in MCF7 and HCC1806 cells. GRHL2 has been shown to enhance expression of telomerase reverse transcriptase (TERT) in keratinocytes and oral squamous cell carcinoma cells [38,39], to support expression of proliferating cell nuclear antigen (PCNA) in colorectal cancer cells [40], and to suppress expression of the death receptor FAS in fibrosarcoma cells [24]. Expression of *FAS* was not significantly increased upon GRHL2 knockout in either cell type (Fig. 7). However, expression of *TERT* and *PCNA* mRNA was significantly downregulated in absence of GRHL2 in MCF7 whereas expression was unaltered or even increased in HCC1806 cells.

These results indicate that the impact of GRHL2 depletion on growth characteristics of different breast cancer subtypes may be distinct with luminal-like cells experiencing a robust growth arrest and basal A-like cells maintaining a reduced growth potential.
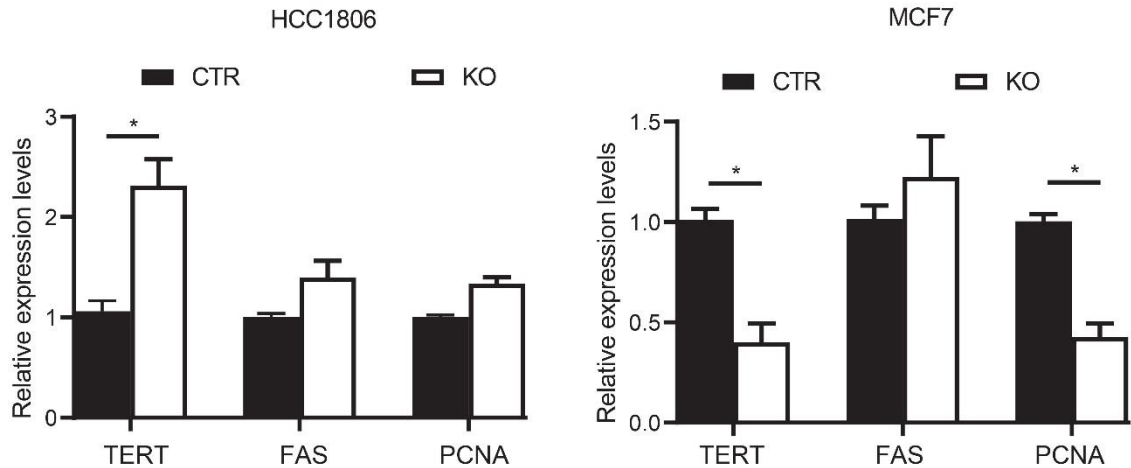
**Fig. 7. Response to GRHL2 knockout in luminal- and basal A-like cells – changes in candidate target genes related to survival and proliferation.** Graphs showing results from qRT-PCR assay for MCF7 and HCC1806 cells transduced with sgCTR and sgGRHL2 after 8 days doxycycline-induction. Data are statistically analyzed by t-test. Data are presented as mean ± SEM from 3 biological replicates. * indicates $p < 0.05$

# Discussion

*GRHL2* is located on chromosome 8q22 and amplified or overexpressed in several cancer types, including breast cancer [11,41]. *In vivo* and clinical studies support an oncogenic role of GRHL2 [11,16-18,42-44]. Our findings corroborate such a role for GRHL2 and demonstrate an association of GRHL2 expression with a trend toward poor prognosis in breast cancer. However, our results indicate that this association is different for different breast cancer subtypes, with a trend, albeit not significant, towards an association with better prognosis in the basal-like subtype. Our study using a panel of >50 human breast cancer cell lines, confirms and extends an earlier report showing that GRHL2 is downregulated in basal B-like breast cancer [12]. This is remarkable given its apparent relation to poor prognosis, since triple negative/basal-like tumors are often aggressive and have a poorer prognosis compared to the ER-positive luminal subtypes [4]. Moreover, basal B-like cells are enriched in EMT markers that are also associated with aggressiveness [4]. Indeed, GRHL2 may play a dual role in breast cancer [11,14,30]and a tumor- or metastasis-suppressive function has been related to its ability to suppress EMT, stemness, and invasion in cell line models and clinical samples [12,14,45]. The function of GRHL2 likely is context-dependent and the consequence of GRHL2 loss depends on the cancer type and the stage of cancer progression.

GRHL2 expression is associated with epithelial markers, including E-cadherin and claudins that are absent in basal B-like breast cancer cells. However, it is unknown whether luminal- and basal A-like breast cancer cells that have an epithelial phenotype with E-cadherin-mediated cell-cell contacts respond similarly to a loss of GRHL2. Our current study indicates that this response is modulated by the balance between a loss of growth stimulation and an induced EMT. This balance appears to be different for luminal and basal-like cells.

We find that expression of E-cadherin is downregulated in response to GRHL2 knockout in luminal-like and basal A-like breast cancer cells, consistent with previous studies [46-49]. However, we find that a further EMT-like shift is not necessarily induced. The acquisition of mesenchymal markers and enhanced cell migration and invasion is seen for HCC1806 (basal A) cells but not in MCF7 cells (luminal-like). These results demonstrate that the enhanced motile properties triggered by a loss of GRHL2 require acquisition of a mesenchymal phenotype. Loss of E-cadherin is not sufficient for enhanced cell migration and invasion of breast cancer cells [50]. The induction of mesenchymal markers, such as N-cadherin and Vimentin that we find to occur only in the HCC1806 cells may contribute to cell migration and invasion. N-cadherin junctions on the cell surface may act as migration tracks [51] and N-cadherin supports the organization of an actin network that drives cell migration [52]. Vimentin, a type III intermediate filament protein, is involved in cell adhesion, migration and signal transduction and emerges in pathologies processes involving epithelial cell migration [53]. Vimentin may facilitate cell migration through upregulation of AXL in breast cancer [54] but overexpression of Vimentin by itself does not enhance cell migration in MCF7 cells [55]. Altogether, our findings and other reports indicate that in order for GRHL2 loss to trigger a shift to a more motile behavior, loss of E-cadherin is not sufficient but a more elaborate transition is required, including loss of epithelial markers such as E-cadherin and gain of mesenchymal markers such as Vimentin and N-cadherin.

Our results show that loss of GRHL2 results in an inhibition of cell proliferation in basal-like and luminal-like breast cancer cells, consistent with earlier findings supporting an oncogenic role of GRHL2 [16,18,40]. However, we find that the impact is different for luminal and basal A-like cells with MCF7 cells experiencing a robust cell cycle and

growth arrest and HCC1806 maintaining a slow growth phenotype with a moderate increase in the fraction of cells in G0/1. This may result from the fact that GRHL2 is an integral part of ER transcriptional complex that induces expression of genes associated with cell proliferation in ER positive MCF7 cells only [11].No induction of cell death is observed in either cell type and expression of *FAS*, which was reported to be controlled by GRHL2, is unaltered [24]. *PCNA* expression is downregulated in MCF7 upon GRHL2, where a growth arrest is observed. Moreover, expression of *TERT*, which has been reported to be epigenetically controlled by GRHL2 [38,39] is attenuated in MCF7 but not HCC1806 following GRHL2 depletion indicating that replicative potential is differentially affected.

Taken together, our findings shed further light on the apparent dual role of GRHL2 in breast cancer. GRHL2 expression supports cell proliferation and suppresses cell motility in breast cancer cells but the outcome of GRHL2 loss differs for different subtypes. In luminal-like cells growth arrest is the main outcome of GRHL2 loss whereas in basal A-like cells reduced growth is accompanied by aspects of EMT and enhanced motility and invasion. This suggests that GRHL2 represents a candidate therapeutic target for luminal-like breast cancer, but interfering with GRHL2 expression or function is senseless in basal B-like breast cancers and may trigger unwanted effects in basal A-like breast cancers.

# References

1      Harbeck, N. & Gnant, M. Breast cancer. *Lancet* **389**, 1134-1150, doi:10.1016/S0140-6736(16)31891-8 (2017).

2      Ma, L., Yan, H., Zhao, H. & Sun, J. Grainyhead-like 2 in development and cancer. *Tumour Biol* **39**, 1010428317698375, doi:10.1177/1010428317698375 (2017).

3      Keller, P. J. *et al.* Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast cancer research : BCR* **12**, R87, doi:10.1186/bcr2755 (2010).

4      Dai, X., Cheng, H., Bai, Z. & Li, J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J Cancer* **8**, 3131-3141, doi:10.7150/jca.18457 (2017).

5      Dai, X. *et al.* Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res* **5**, 2929-2943 (2015).

6      Tran, B. & Bedard, P. L. Luminal-B breast cancer and novel therapeutic targets. *Breast cancer research : BCR* **13**, 221, doi:10.1186/bcr2904 (2011).

7      Prat, A. *et al.* Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* **24 Suppl 2**, S26-35, doi:10.1016/j.breast.2015.07.008 (2015).

8      Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527, doi:10.1016/j.ccr.2006.10.008 (2006).

9      Ming, Q. *et al.* Structural basis of gene regulation by the Grainyhead/CP2 transcription factor family. *Nucleic acids research* **46**, 2082-2095, doi:10.1093/nar/gkx1299 (2018).

10    Nüsslein-Volhard C., W. E., Kluding H. Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : I. Zygotic loci on the second chromosome. *Wilehm Roux Arch Dev Biol.* **193**, doi:10.1007/BF00848156. (1984).

11    Reese, R. M., Harrison, M. M. & Alarid, E. T. Grainyhead-like Protein 2: The Emerging Role in Hormone-Dependent Cancers and Epigenetics. *Endocrinology* **160**, 1275-1288, doi:10.1210/en.2019-00213 (2019).

12    Cieply, B. *et al.* Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res* **72**, 2440-2453, doi:10.1158/0008-5472.CAN-11-4038 (2012).

13    Pifer, P. M. *et al.* Grainyhead-like 2 inhibits the coactivator p300, suppressing tubulogenesis and the epithelial-mesenchymal transition. *Mol Biol Cell* **27**, 2479-2492, doi:10.1091/mbc.E16-04-0249 (2016).

14    Werner, S. *et al.* Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *The Journal of biological chemistry* **288**, 22993-23008, doi:10.1074/jbc.M113.456293 (2013).

15    Garnis, C., Coe, B. P., Zhang, L., Rosin, M. P. & Lam, W. L. Overexpression of LRP12, a gene contained within an 8q22 amplicon identified by high-resolution array CGH analysis of oral squamous cell carcinomas. *Oncogene* **23**, 2582-2586, doi:10.1038/sj.onc.1207367 (2004).

16    Paltoglou, S. *et al.* Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer. *Cancer Res* **77**, 3417-3430, doi:10.1158/0008-5472.CAN-16-1616 (2017).

17    Pan, X. *et al.* GRHL2 suppresses tumor metastasis via regulation of transcriptional activity of RhoG in non-small cell lung cancer. *Am J Transl Res* **9**, 4217-4226 (2017).

18    Faddaoui, A. *et al.* Suppression of the grainyhead transcription factor 2 gene (GRHL2) inhibits the proliferation, migration, invasion and mediates cell cycle arrest of ovarian cancer cells. *Cell Cycle* **16**, 693-706, doi:10.1080/15384101.2017.1295181 (2017).

19    Truong HH, X. J., Ghotra VP, Nirmala E, Haazen L, Le Dévédec SE, Balcioğlu HE, He S, Snaar-Jagalska BE, Vreugdenhil E, Meerman JH, van de Water B, Danen EH. β1 Integrin Inhibition Elicits a Prometastatic Switch Through the TGFb–miR-200–ZEB Network in E-Cadherin–Positive Triple-Negative Breast Cancer. *SCi Signal* (2014).

20    Ciriello, G. *et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163**, 506-519, doi:10.1016/j.cell.2015.09.033 (2015).

21    Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature communications* **7**, 11479, doi:10.1038/ncomms11479 (2016).

22    Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).

23    Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* **123**, 725-731, doi:10.1007/s10549-009-0674-9 (2010).

24    Dompe, N. *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* **108**, E943-951, doi:10.1073/pnas.1100132108 (2011).

25    Holliday, D. L. & Speirs, V. Choosing the right cell line for breast cancer research. *Breast cancer research* **13**, 215 (2011).

26    Su, Y., Pogash, T. J., Nguyen, T. D. & Russo, J. Development and characterization of two human triple-negative breast cancer cell lines with highly tumorigenic and metastatic capabilities. *Cancer Med* **5**, 558-573, doi:10.1002/cam4.616 (2016).

27    Prat, A. *et al.* Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat* **142**, 237-255, doi:10.1007/s10549-013-2743-3 (2013).

28    Teo, K. *et al.* E-cadherin loss induces targetable autocrine activation of growth factor signalling in lobular breast cancer. *Sci Rep* **8**, 15454, doi:10.1038/s41598-018-33525-5 (2018).

29    Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21-45, doi:10.1016/j.cell.2016.06.028 (2016).

30    Frisch, S. M., Farris, J. C. & Pifer, P. M. Roles of Grainyhead-like transcription factors in cancer. *Oncogene*, doi:10.1038/onc.2017.178 (2017).

31    Zhu, H. *et al.* The role of the hyaluronan receptor CD44 in mesenchymal stem cell migration in the extracellular matrix. *Stem cells (Dayton, Ohio)* **24**, 928-935, doi:10.1634/stemcells.2005-0186 (2006).

32    Xu, H. *et al.* The role of CD44 in epithelial-mesenchymal transition and cancer development. *OncoTargets and therapy* **8**, 3783-3792, doi:10.2147/OTT.S95470 (2015).

33    Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359-369, doi:10.1016/s0092-8674(00)81280-5 (1996).

34    Price, J. T., Tiganis, T., Agarwal, A., Djakiew, D. & Thompson, E. W. Epidermal growth factor promotes MDA-MB-231 breast cancer cell migration through a phosphatidylinositol 3'-kinase and phospholipase C-dependent mechanism. *Cancer Res* **59**, 5475-5478 (1999).

35    Franken, N. A., Rodermond, H. M., Stap, J., Haveman, J. & van Bree, C. Clonogenic assay of cells in vitro. *Nature protocols* **1**, 2315-2319, doi:10.1038/nprot.2006.339 (2006).

36    Orellana, E. A. & Kasinski, A. L. Sulforhodamine B (SRB) Assay in Cell Culture to Investigate Cell Proliferation. *Bio-protocol* **6**, doi:10.21769/BioProtoc.1984 (2016).

37    Vichai, V. & Kirtikara, K. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nature protocols* **1**, 1112-1116, doi:10.1038/nprot.2006.179 (2006).

38    Kang, X., Chen, W., Kim, R. H., Kang, M. K. & Park, N. H. Regulation of the hTERT promoter activity by MSH2, the hnRNPs K and D, and GRHL2 in human oral squamous cell carcinoma cells. *Oncogene* **28**, 565-574, doi:10.1038/onc.2008.404 (2009).

39    Chen, W. *et al.* 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* **285**, 40852-40863, doi:10.1074/jbc.M110.103812 (2010).

40    Hu, F., He, Z., Sun, C. & Rong, D. Knockdown of GRHL2 inhibited proliferation and induced apoptosis of colorectal cancer by suppressing the PI3K/Akt pathway. *Gene* **700**, 96-104, doi:10.1016/j.gene.2019.03.051 (2019).

41    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* **240**, 1396-1401, doi:10.1177/1535370215588924 (2015).

42    Yang, X., Vasudevan, P., Parekh, V., Penev, A. & Cunningham, J. M. Bridging cancer biology with the clinic: relative expression of a GRHL2-mediated gene-set pair predicts breast cancer metastasis. *PLoS One* **8**, e56195, doi:10.1371/journal.pone.0056195 (2013).

43    Quan, Y. *et al.* Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. *Cancer Biol Ther* **15**, 878-887, doi:10.4161/cbt.28877 (2014).

44    Butz, H. *et al.* Integrative bioinformatics analysis reveals new prognostic biomarkers of clear cell renal cell carcinoma. *Clinical chemistry* **60**, 1314-1326, doi:10.1373/clinchem.2014.225854 (2014).

45    Cieply, B., Farris, J., Denvir, J., Ford, H. L. & Frisch, S. M. Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res* **73**, 6299-6309, doi:10.1158/0008-5472.CAN-12-4082 (2013).

46      Chen, W. *et al.* Grainyhead-like 2 (GRHL2) knockout abolishes oral cancer development through reciprocal regulation of the MAP kinase and TGF-beta signaling pathways. *Oncogenesis* **7**, 38, doi:10.1038/s41389-018-0047-5 (2018).

47      Chen, W. *et al.* Grainyhead-like 2 regulates epithelial plasticity and stemness in oral cancer cells. *Carcinogenesis* **37**, 500-510, doi:10.1093/carcin/bgw027 (2016).

48      Chung, V. Y. *et al.* GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943, doi:10.1038/srep19943 (2016).

49      Farris, J. C. *et al.* Grainyhead-like 2 Reverses the Metabolic Changes Induced by the Oncogenic Epithelial-Mesenchymal Transition: Effects on Anoikis. *Mol Cancer Res* **14**, 528-538, doi:10.1158/1541-7786.MCR-16-0050 (2016).

50      Sommers, C. L. *et al.* Cell adhesion molecule uvomorulin expression in human breast cancer cell lines: relationship to morphology and invasive capacities. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research* **2**, 365-372 (1991).

51      Shih, W. & Yamada, S. N-cadherin-mediated cell-cell adhesion promotes cell migration in a three-dimensional matrix. *J Cell Sci* **125**, 3661-3670, doi:10.1242/jcs.103861 (2012).

52      Ponti, A., Machacek, M., Gupton, S. L., Waterman-Storer, C. M. & Danuser, G. Two distinct actin networks drive the protrusion of migrating cells. *Science* **305**, 1782-1786, doi:10.1126/science.1100533 (2004).

53      Ivaska, J., Pallari, H. M., Nevo, J. & Eriksson, J. E. Novel functions of vimentin in cell adhesion, migration, and signaling. *Experimental cell research* **313**, 2050-2062, doi:10.1016/j.yexcr.2007.03.040 (2007).

54      Vuoriluoto, K. *et al.* Vimentin regulates EMT induction by Slug and oncogenic H-Ras and migration by governing Axl expression in breast cancer. *Oncogene* **30**, 1436-1448, doi:10.1038/onc.2010.509 (2011).

55      Sommers, C. L. *et al.* Loss of epithelial markers and acquisition of vimentin expression in adriamycin- and vinblastine-resistant human breast cancer cell lines. *Cancer Res* **52**, 5190-5197 (1992).

## Acknowledgements

## Author contributions

EHJD supervised the research. ZW and EHJD conceived, designed the experiments. ZW, BC, CYL and YJC performed the experiments. ZW and EHJD wrote the manuscript. All authors read, reviewed and approved the final manuscript.

## Competing interests

The author(s) declare no competing interests.

# Supplementary

**Supplementary Table S1 RT-qPCR primers.**

| | | |
|---|---|---|
| TERT | Forward | GATATCGTCCAGGCCCAGC |
| TERT | Reverse | CATGGACTACGTCGTGGGAG |
| FAS | Forward | GTGGACCCGCTCAGTACG |
| FAS | Reverse | TCTAGCAACAGACGTAAGAACCA |
| PCNA | Forward | GCCTGACAAATGCTTGCTGAC |
| PCNA | Reverse | TTGAGTGCCTCCAACACCTTC |
| GAPDH | Forward | CCATGGGGAAGGTGAAGGTC |
| GAPDH | Reverse | AGTTAAAAGCAGCCCTGGTGA |

# Chapter 3

## Genome-wide identification of binding sites of GRHL2 in luminal-like and basal A subtypes of breast cancer

Zi Wang[1], Haoyu Wu[2], Lucia Daxinger[2], Erik HJ Danen[1,3]

[1]Leiden Academic Center for Drug Research, Leiden University, Leiden, The Netherlands; [2]Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands; [3]correspondence to Erik HJ Danen, e.danen@lacdr.leidenuniv.nl

# Abstract

Grainyhead like 2 (*GRHL2*) is one of three mammalian homologues of the Grainyhead (*GRH*) gene. It suppresses the oncogenic epithelial-mesenchymal transition (EMT), acting as a tumor suppressor. On the other hand, GHRL2 promotes cell proliferation by increasing human telomerase reverse transcriptase (hTERT) activity, serving as a tumor promoter. According to gene expression profiling, breast cancer can be divided into basal-like (basal A and basal B), luminal-like, HER2 enriched, claudin-low and normal-like subtypes. To identify common and subtype-specific genomic binding sites of GRHL2 in breast cancer, GRHL2 ChIP-seq was performed in three luminal-like and three basal A human breast cancer cell lines. Most binding sites of GRHL2 were found in intergenic and intron regions. 13,351 common binding sites were identified in basal A cells, which included 551 binding sites in gene promoter regions. For luminal-like cells, 6,527 common binding sites were identified, of which 208 binding sites were found in gene promoter regions. Basal A and luminal-like breast cancer cells shared 4711 GRHL2 binding sites, of which 171 binding sites were found in gene promoter regions. The identified GRHL2-binding motifs are all identical to a motif reported for human ovarian cancer, indicating conserved GRHL2 DNA-binding among human cancer cells. Notably, no binding sites of GRHL2 were detected in the promoter regions of several established EMT-related genes, including *CDH1*, *ZEB1*, *ZEB2* and *CDH2* genes. Collectively, this study provides a comprehensive overview of interactions of GRHL2 with DNA and lays the foundation for further understanding of common and subtype-specific signaling pathways regulated by GRHL2 in breast cancer.

# Introduction

Breast cancer is the predominant cause of cancer-related death in women aged 20 to 59 years globally [1]. Based on gene expression profiling, breast cancer can be divided into several subtypes with distinct molecular features, which includes luminal-like (luminal A and luminal B), basal-like (basal A and basal B), human epidermal growth factor receptor 2 (HER2)-enriched, claudin-low and normal-like subtypes[2]. Both luminal-like and basal-like subtypes comprise at least 73% of all breast cancers[2]. Conversions of luminal to basal lineage have been observed in mouse breast cancer models [3,4] but luminal-like and basal-like subtypes differ in prognosis and response to therapy. Therefore, it is important to characterize common features and discordances between them.

The *GRH* gene was discovered in *Drosophila* and its mammalian homologs have three members (*GRHL1*, *GRHL2* and *GRHL3*)[5]. *GRH* deficiency leads to failure of complete neural tube closure, epidermal barrier formation, trachea elongation and epidermal wound response [5-7]. *GRHL2* is one of three mammalian homologues of the *GRH* gene, which has been investigated in cancer development. *GRHL2* is located on chromosome 8q22 that is frequently amplified in many cancers, including breast cancer, colorectal cancer and oral squamous cell carcinoma[8-10]. GRHL2, as an oncogene, positively regulates cell proliferation by enhancing hTERT activity through inhibition of DNA methylation at 5'-CpG island around gene promoter[9]. GRHL2 inhibits cell apoptosis by suppressing death receptor (FAS and DR5) expression in breast cancer cells[8,11]. Knockdown of GRHL2 downregulated HER3 expression, resulting in inhibition of cell proliferation[11]. On the other hand, GRHL2 was previously reported as a suppressor of oncogenic EMT by the loop of GRHL2-miR200-ZEB1 and regulation of the TGF-β pathway[12-14]. These controversial results suggest that the roles of GRHL2 may be tumor-specific through regulating different target genes in different cancers.

Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) is a widely used method to analyze protein-DNA interactions, histone modifications, and nucleosomes on genome-wide scale in living cells by capturing proteins at sites of their binding to DNA[15,16]. Previous findings showed that GRHL2 shares a similar DNA-binding motif with other GRHL family members[13,17,18]. To date, no studies have

investigated the genomic landscape of GRHL2 binding sites across breast cancer subtypes. In this study, we provide a comprehensive overview of binding sites of GRHL2 in the genome of basal A and luminal-like subtypes of breast cancer.

# Methods and materials

### Cell lines

Human breast cancer cell lines representing luminal-like (MCF7, T47D, BT474), basal A (HCC1806, BT20 and MDA-MB-468), and basal B subtypes (Hs578T) were obtained from the American Type Culture Collection. 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% $CO_2$).

### Chromatin immunoprecipitation-sequencing (ChIP-seq)

Cells were grown in RPMI-1640 complete medium. Cross-linking was performed by 1% formaldehyde for 10 minutes at room temperature (RT). Then 1M glycine (141 µl 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 µl/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 eluted 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-PCR was performed to validate interaction of GRHL2 with the promoter region of *CLDN4*, a direct

target gene of GRHL2 [19]. The results confirmed the GRHL2 binding site around the *CLDN4* promoter (Fig. S2). The following primers were used for ChIP-PCR: *CLDN4* forward: gtgacctcagcatgggctttga, *CLDN4* reverse: ctcctcctgaccagtttctctg, Control (an intergenic region upstream of the *GAPDH* locus) forward: atgggtgccactggggatct, Control reverse: tgccaaagcctaggggaaga, *ZEB1* promoter[#] forward: cggtccctagcaacaaggtt, *ZEB1* promoter[#] reverse: tcgcttgtgtctaaatgctcg. *ZEB1*[##] forward: gccgccgagcctccaacttt, *ZEB1*[##] reverse: tgctagggaccgggcggttt, *OVOL2* exon forward: ccttaaatcgcgagtgagacc, *OVOL2* exon reverse: gtagcgagcttgttgacacc, *CDH1* intron forward: gtatgaacggcaagcctctg, *CDH1* intron reverse: caagggagccaggaagagaa. ChIP-PCR data were collected and analyzed using the $2^{-\Delta\Delta Ct}$ method [20].

For ChIP-Seq, library preparation and paired-end sequencing were performed by GenomeScan (Leiden, The Netherlands)

**ChIP-seq data analysis**

Paired-end reads were mapped to the human reference genome (hg38) using BWA-MEM[21] with default parameters. Over 93% of total reads were mapped to the human genome in BT20, HCC1806, MDA-MB-468, T47D and MCF7 cell line. For BT474, ~57.3% reads were mapped. Phred quality score (Q score) was used to measure base calling accuracy, which indicates the probability that a given base is called incorrectly[22]. Q score is logarithmically related to the base calling error probabilities $P$ [22].

$$Q = -10 \log_{10} P$$

Q=30 nominally corresponds to a 0.1% error rate[23]. Reads with scores > Q30 were over 86% in BT20, HCC1806, MDA-MB-468, T47D and MCF7 cell lines. For BT474, reads with scores > Q30 accounted for 48.6%.

To examine whether the paired-end reads were appended with unwanted adapter sequences, an adapter content test was performed. The quality control report (Fig. S3) showed that cumulative presence of adapter sequences was <5% in all cell samples, indicating that all data sets could be further analyzed without adapter-trimming. Per base sequence quality of sequencing was examined, which indicated that all sequencing data were of high quality (Fig. S4) and could be further analyzed.

Reads with low mapping quality (≤ Q30) were filtered out. MACS version 2.1.0[24] was used for peak calling by default settings. q value was adjusted to 0.1 for BT474 cell line to avoid loss of peaks. The annotatePeaks and MergePeaks function from HOMER [25] were used to annotate and overlap peaks, respectively. ChIPseeker was used for analysis of ChIP-seq peaks coverage plot and density profile of GRHL2 binding sites [26]. Motif analysis was performed by ChIP-seq peaks with high scores using the MEME-ChIP program with default settings. ChIP-seq data was visualized by UCSC genome browser.
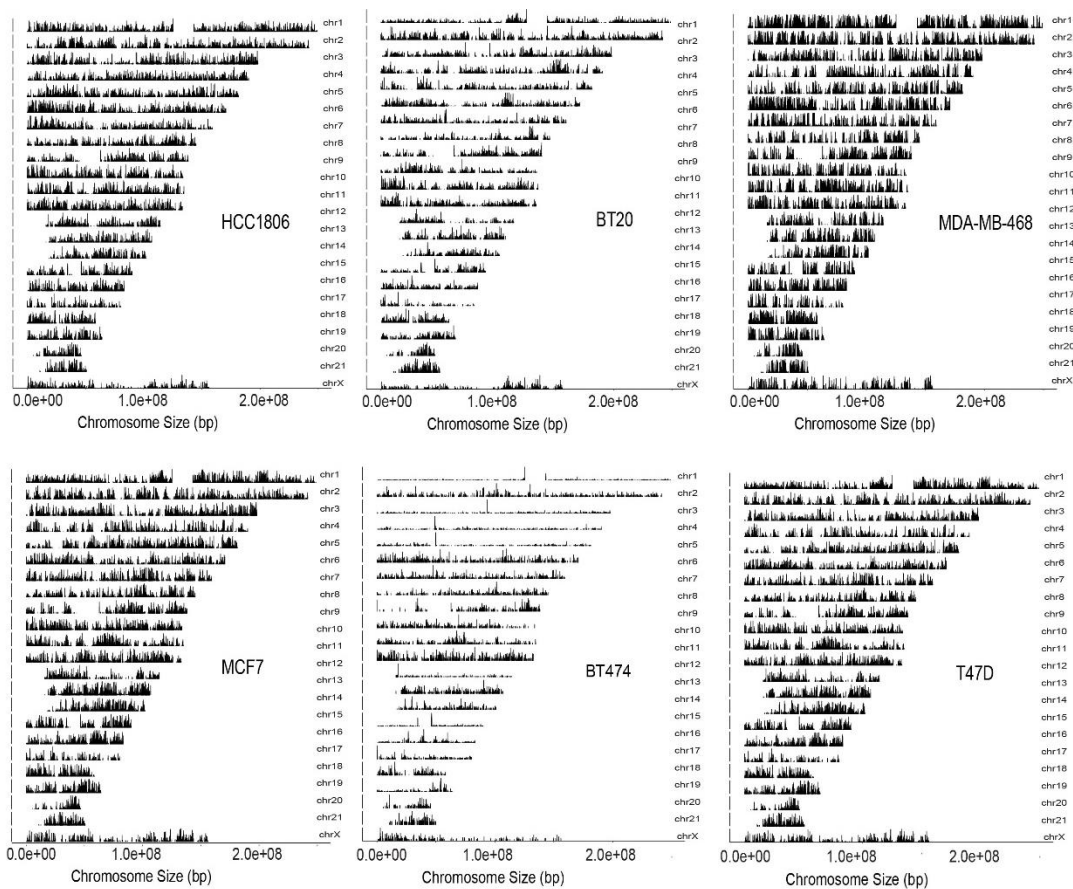


**Fig. 1 Coverage of peak regions across chromosomes.** The graph represents the coverage of GRHL2 binding sites across the chromosomes**.**

# Results

## Genome-wide identification of binding sites of GRHL2 in luminal-like and basal A subtypes of breast cancer

To identify GRHL2 binding sites, ChIP-seq was performed in luminal-like (MCF7, T47D and BT474) and basal A (HCC1806, BT20 and MDA-MB-468) breast cancer cells.

Firstly, the coverage of peak regions across chromosomes was analyzed[26]. In each cell sample, GRHL2 was strongly associated with all chromosomes (Fig. 1).



**3'UTR** ■ **5'UTR** ■ **Exon** ■ **Intergenic** ■ **Intron** ■ **Promoter** ■ **TTS** ■ **Unknown**
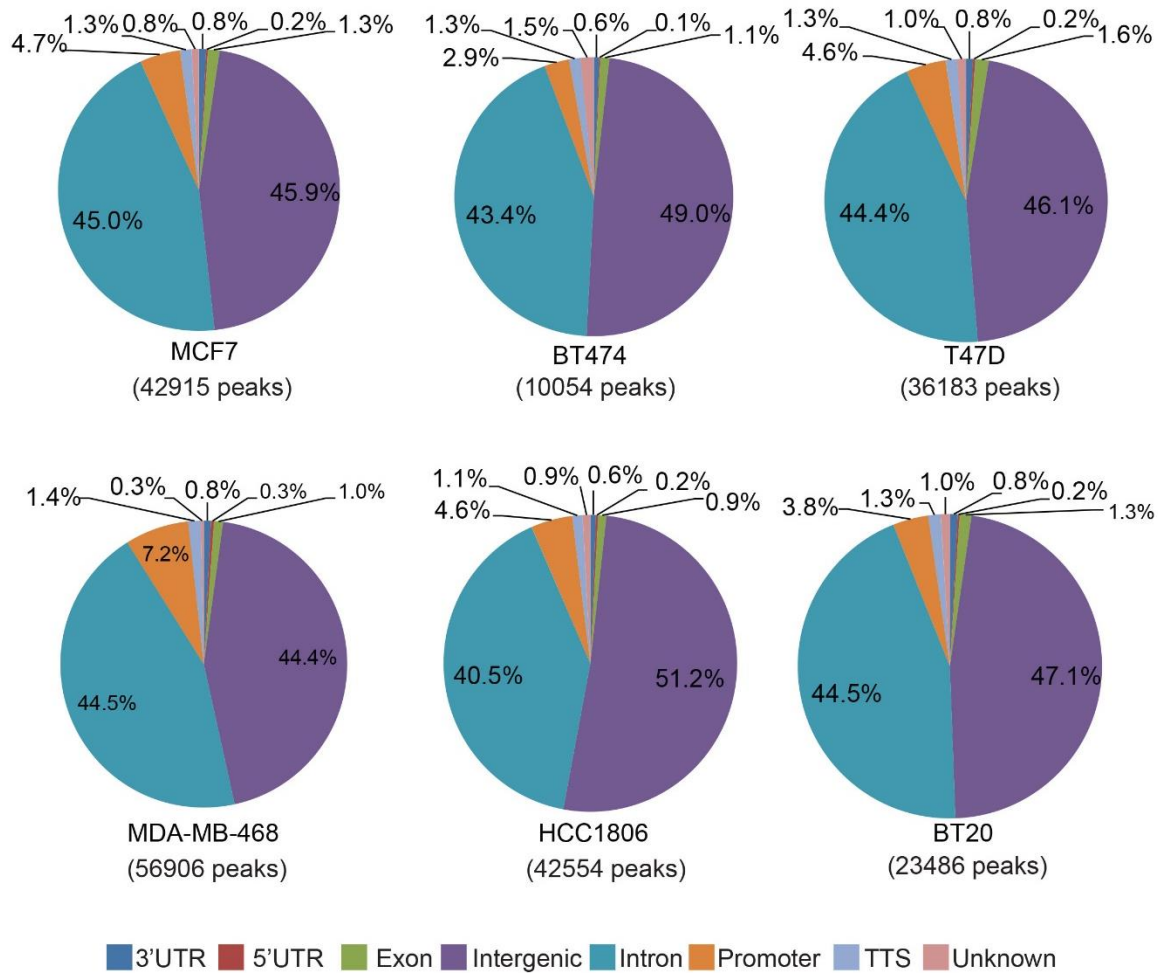
**Fig. 2 Percentage of GRHL2 binding sites found at promoter regions, 5' untranslated regions (UTRs), 3' UTRs, exons, introns, intergenic regions, transcription termination sites (TTSs) and unknown regions.** Promoter regions are defined as -1000 bp to +100 bp from the transcription start sites (TSS).

GRHL2 binding sites were found in intergenic regions, transcription start sites (TSS) promoter regions, introns, exons, transcription termination sites (TTS) and unknown regions (Fig. 2). The majority of peaks was located in intergenic and intron regions in basal A and luminal-like breast cancer cells. Genes where GRHL2 was found to interact with the -1000 bp to +100 bp promoter region in all three luminal (left column), all three basal A (middle column), or all luminal and basal A cell lines tested (right column) were identified and represent likely candidate general and subtype-specific GRHL2 target genes (Table S1).

To further investigate if peaks were enriched in promoter regions, read count frequency and density profiling of GRHL2 binding sites within -6000 bp ~ +6000 bp of the transcription start site (TSS) were analyzed (Fig. 3). Consistent with the annotation of binding sites, which showed most GRHL2 binding sites existed in the intergenic regions, the density of GRHL2 binding sites was not increased in the -1000 bp to +100 bp promoter region of basal A and luminal-like breast cancer cells.

To detect similarities of GRHL2 binding sites between luminal-like and basal A subtype, three luminal-like/basal A data sets were overlapped to identify shared binding sites. 13,351 common binding sites were identified in basal A subtype of breast cancer cells, which included 551 binding sites in gene promoter regions (-1000 bp~ +100 bp from TSS) (Fig. 4a and b). For luminal-like breast cancer cells, 6,527 common binding sites were identified, of which 208 binding sites were found in gene promoter regions (Fig. 4c and d). Basal A and luminal-like subtypes of breast cancer cells shared 4,711 binding sites of GRHL2, of which 171 binding sites were found in gene promoter regions (Fig. 4e and f).

**Identification of a common GRHL2-interaction motif**
The MEME-ChIP program was used to identify motifs, all of which were with statistical significance. In each sample, 3 motifs with high E value were shown (Fig. 5), whose core binding was similar to previously published ones [13,27-29]. Thus, our ChIP-seq data indicated that GRHL2 motif was highly conserved in human and mouse cells.

**GRHL2-binding at EMT-related genes**
GRHL2 and OVOL2 support an epithelial phenotype and counteract EMT transcription factor such as ZEB and SNAIL. Some studies have 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 [12,27,30,31]. In our ChIP-seq data, GRHL2 binding sites were observed at *CDH1* introns and at promoter regions of *CLDN4* and *OVOL2* (Fig. 6) ChIP-PCR was performed to further validate these interactions (Fig. 7). *CLDN4* showed multiple GRHL2 binding sites across the coding and non-coding regions,

suggesting the binding of GRHL2 to multiple regions may be involved in long-distance chromatin interactions as suggested previously [13]. Conversely, no GRHL2 binding was observed at the promoter of *ZEB1* or *ZEB2* (Fig. 6), arguing against mutual regulation through direct interaction as previously suggested [32,33]. To further evaluate this, ChIP-PCR was carried out using primers that have been previously reported to amplify *ZEB1* promoter DNA sequences bound by GRHL2 in human mammary epithelial cells and human ovarian cancer cells. This experiment further confirmed the absence of GRHL2 binding sites around the promoter of *ZEB1* in basal A (HCC1806, BT20, MDA-MB-468) and luminal-like (T47D, BT474) subtype breast cancer cells (Fig. 7). Moreover, interactions of GRHL2 with *CDH1* intron and *OVOL2* promoter regions were validated in these experiments (Fig. 7). Together, these findings suggest that GRHL2 binding sites in EMT-related genes may be cell context-dependent.
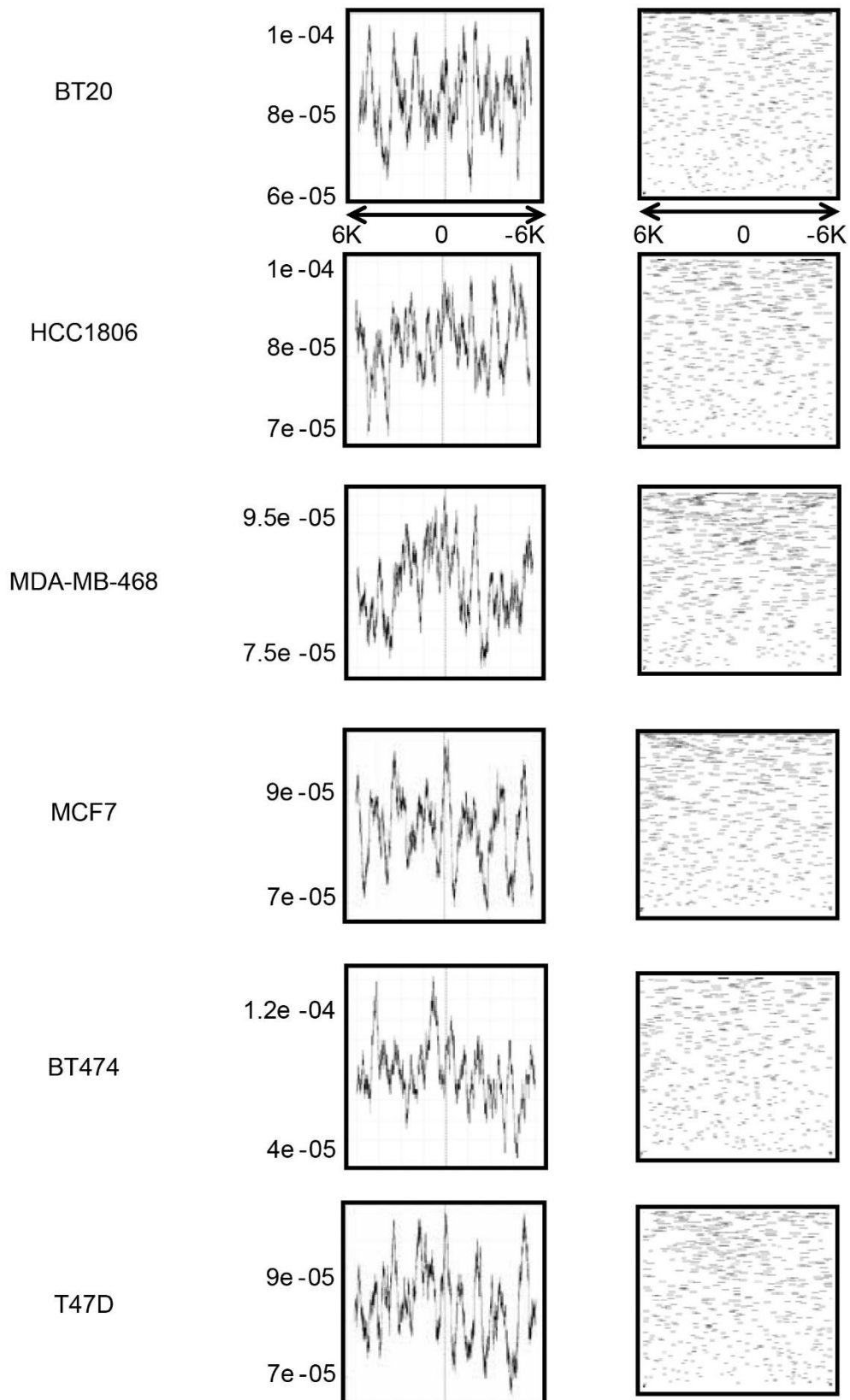
**Fig. 3 The read count frequency and density profile of GRHL2 binding sites within -6000 bp~ +6000 bp of the promoter-TSSs.** On the left side, graphs are for GRHL2 ChIP-seq read count frequency in indicated cell line. X axis represents read count frequency; Y axis is for genomic

region. On the right side, graphs show the density of ChIP-seq reads for GRHL2 binding sites in the indicated cell line.
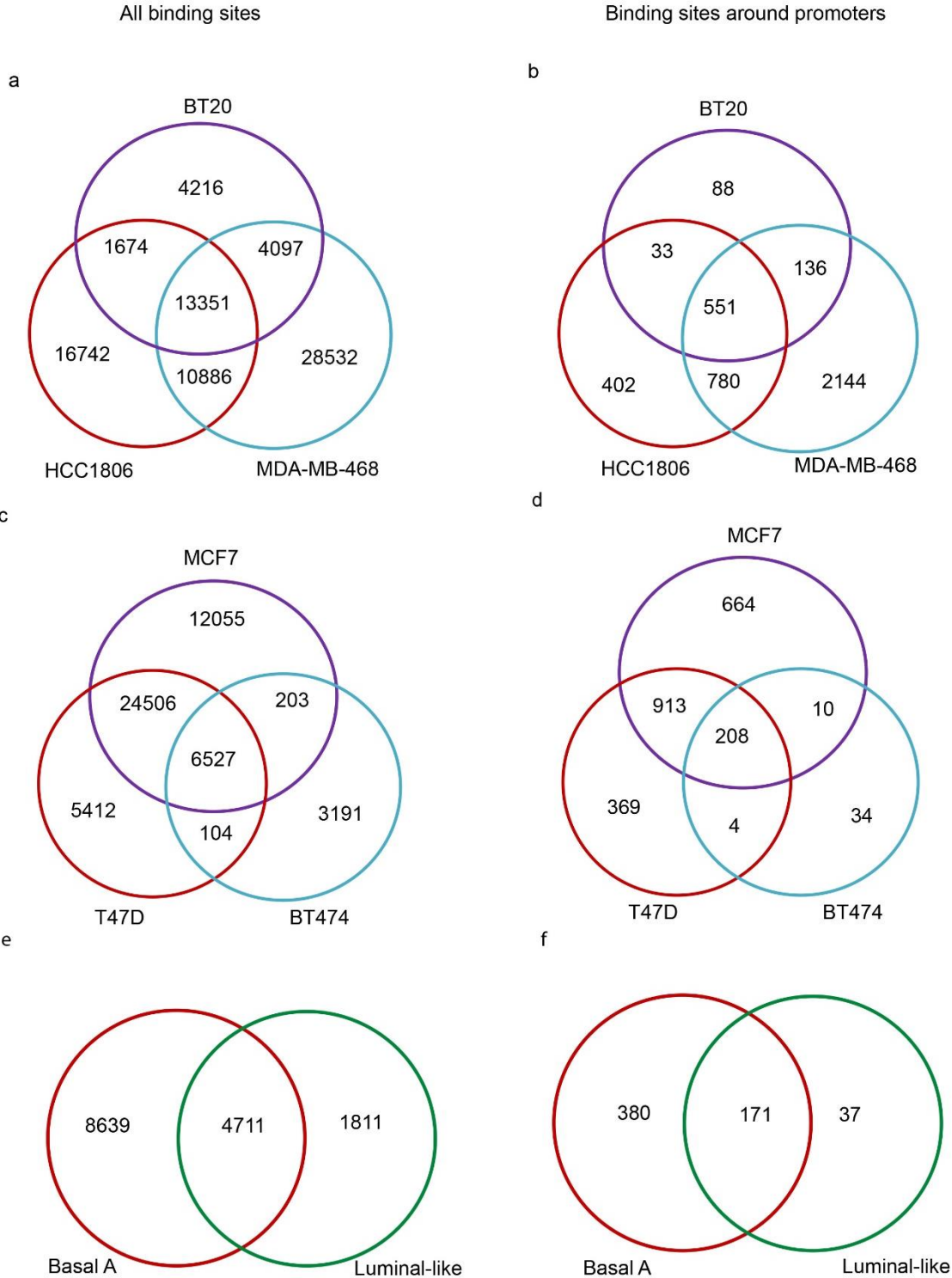


**Fig. 4 Overlap of GRHL2 binding sites.** Overlap of GRHL2 binding sites is identified in the indicated subtypes.
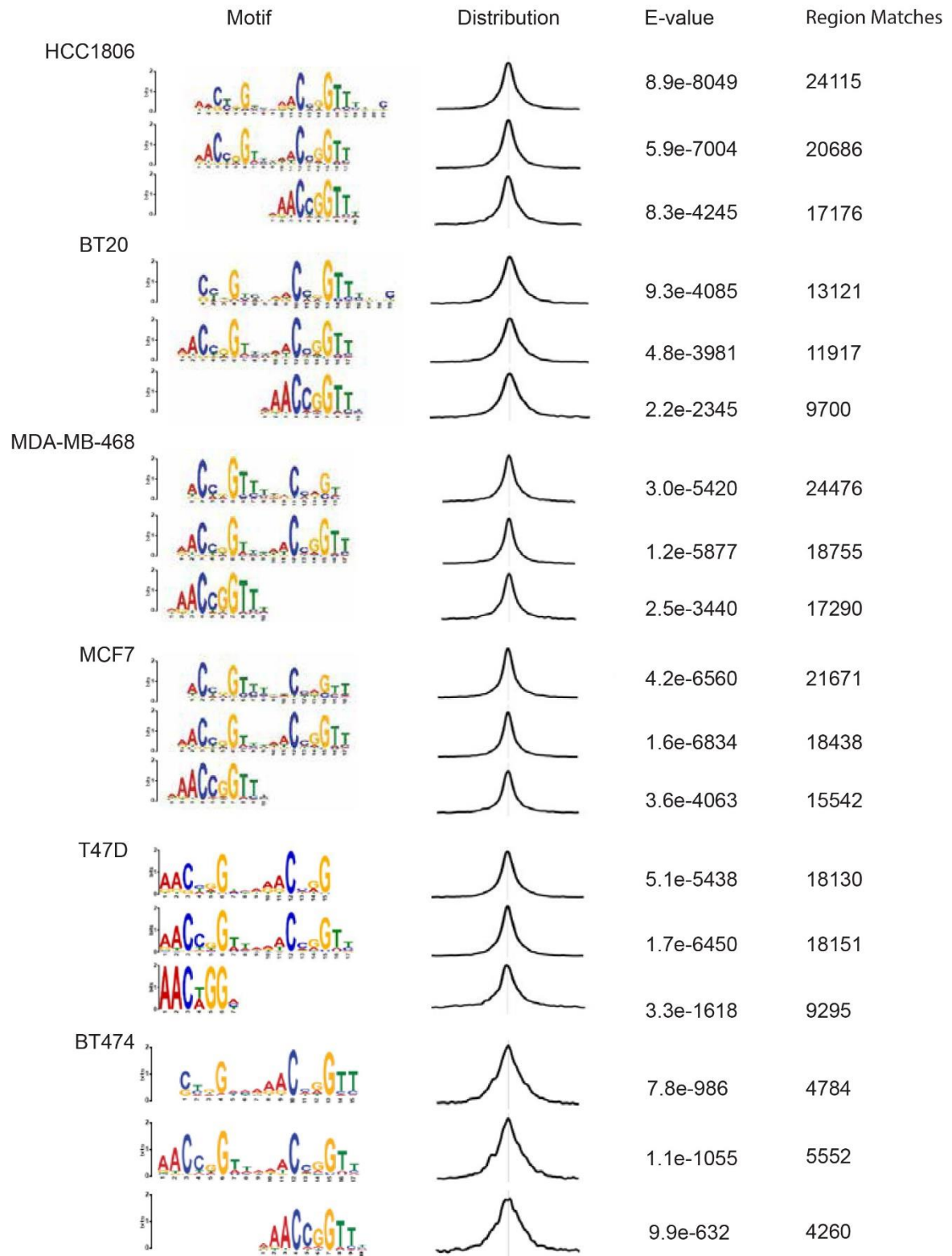
**Fig. 5 DNA-binding motif of GRHL2 in luminal-like and basal A subtypes of breast cancer.** From left to right, the first panel shows the identified motifs in the indicated cells. The second panel shows distribution of the best matches to the motif in the sequences. The third panel shows E-value, the significance of the motif according to the motif discovery. The last panel shows the number of regions that match the corresponding motif.

60
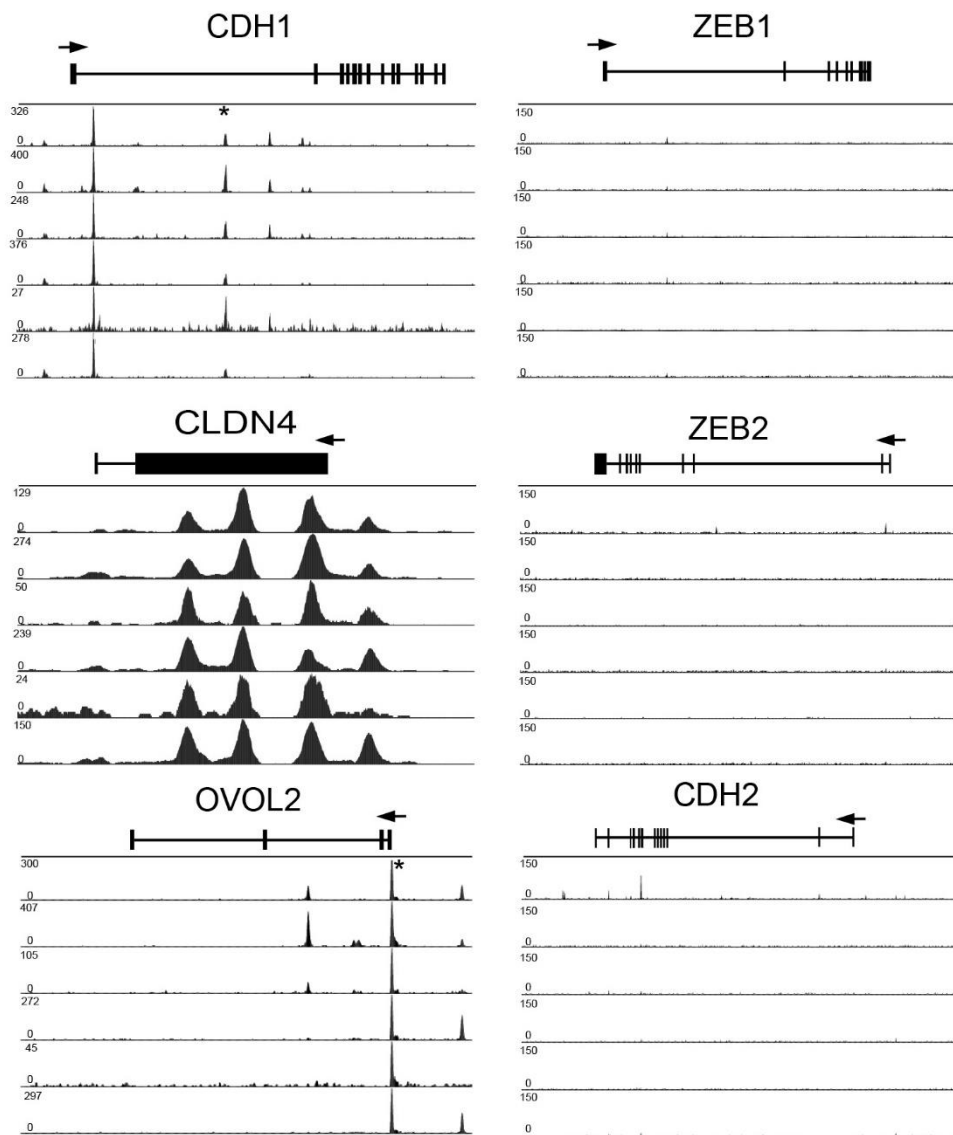
**Fig. 6 GRHL2 ChIP tracks at selected core genes and EMT genes.** ChIP tracks are shown from top to bottom for HCC1806, MDA-MB-468, BT20, MCF7, BT474 and T47D, respectively. The snapshot on the left shows results for 3 identified GRHL2 targets (*CDH1*, *CLDN4* and *OVOL2*) and the snapshot on the right shows results for three not-identified genes encoding proteins associated with EMT (*ZEB1*, *ZEB2* and *CDH2*). The track height is scaled from 0 to the indicated number. Above all tracks, the locus with its exon/intron structure is presented. Binding sites with * are validated by ChIP-PCR.
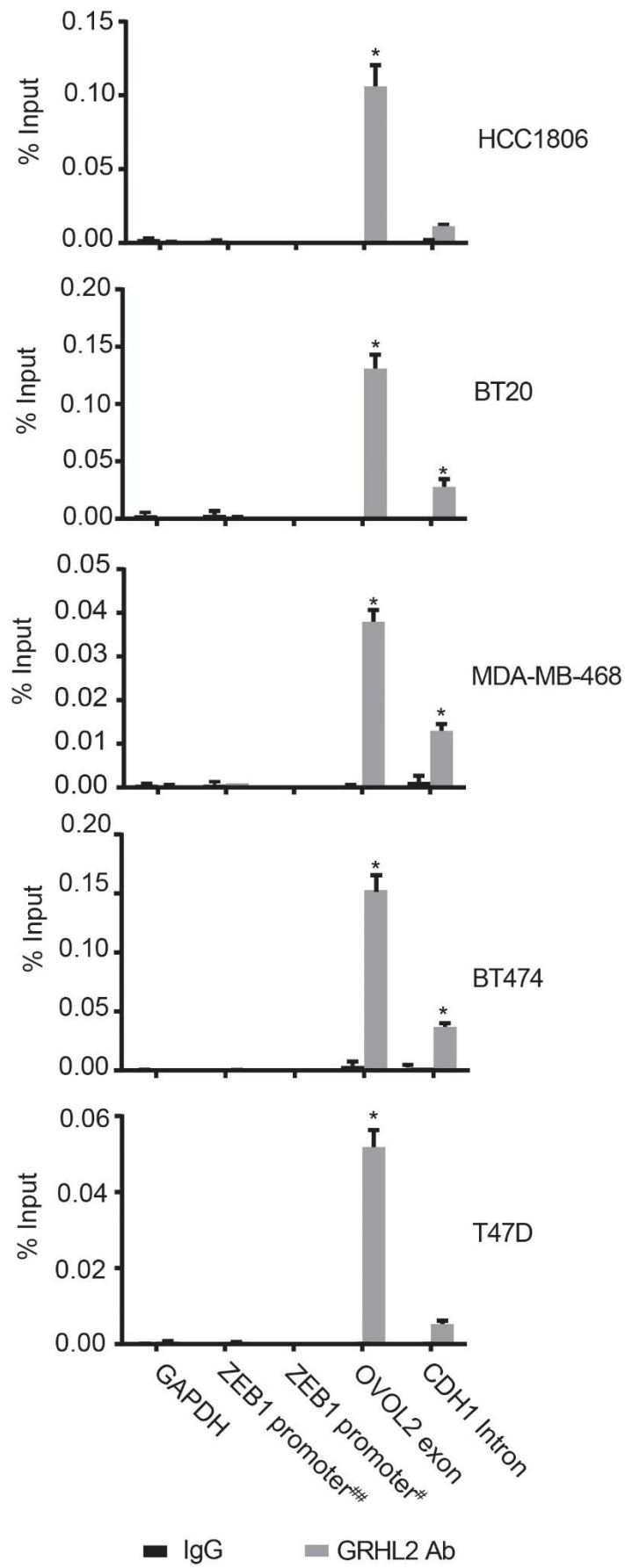
**(Last page) Fig. 7. ChIP-PCR validation of presence and absence of GRHL2 binding sites identified by ChIP-seq.** Graphs represent the efficiency of indicated genomic DNA co-precipitation with anti-GRHL2 Ab (black bars) or IgG control Ab (grey bars). ChIP-PCR showing enrichment of GRHL2 binding sites at *OVOL2* exon and *CDH1* intron, but not *ZEB1* promoter regions. For *ZEB1* detection, primers that were previously reported to successfully amplify GRHL2 binding sites in human mammary epithelial cells (##) and human ovarian cancer cells (#) were used. 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 $p < 0.05$.

# Discussion

Cell type origin is one of the most important factors that determine molecular features of tumors [34]. In general, luminal-like tumor cells are biologically similar to cells derived from inner (luminal) cells lining the mammary ducts, whereas cells of basal-like breast cancer are characterized by features similar to surrounding the mammary ducts[35]. Basal-like breast cancers are associated with a worse prognosis and an increased possibility of cancer metastasis compared with the luminal-like subtype[4,36,37]. Immunohistochemical staining is clinically used to categorize luminal-like breast cancer into luminal A (estrogen receptor (ER) and/or progesterone receptor (PR) positive, HER2 negative) and luminal B (ER and/or PR, and HER2 positive). However, most basal-like breast cancers are negative for ER, PR and HER2, therefore the majority of basal-like breast cancer is triple negative breast cancer (TNBC). Basal-like breast cancer can be further subdivided into basal A and basal B. As for basal A, it is associated with *BRCA1* signatures and resembles basal-like tumors, whereas basal B subtype displays mesenchymal properties and stem/progenitor characteristics[38,39].

In the present study, ChIP-seq was performed to characterize genome-wide binding sites of transcription factor GRHL2 in basal A and luminal-like subtypes of breast cancer. The match with previously a published binding motif shows that GRHL2-interaction with the DNA is highly conserved in human cancer cells. A limited number of binding sites were located in gene promoter regions. Similar to previous reports [13,28], most binding sites were located in introns and intergenic regions of target genes. Such regions may contain enhancers interacting with GRHL2 and GRHL2 has also been reported to regulate histone modifications such as H3K4me3 and H3K4me1[13,40]. Together, this suggests that GRHL2 may regulate gene expression through direct

transcriptional control at promoter regions or through alternative mechanisms including epigenetic mechanisms.

Close to 5000 identified GRHL2 genomic binding sites were shared between all tested basal A and luminal-like cell lines. A similar number of binding sites were found in all basal-like cell lines but were not detected in any of the tested luminal lines. These candidate subtype-specific GRHL2-target sites may serve as a starting point to the unraveling of distinct transcriptional networks in different breast cancer subtypes.

Our analysis of GRHL2 interaction with known EMT-related genes fits previously published findings except for *ZEB1*. It was reported that ZEB1 is regulated by GRHL2 directly and, vice versa, that ZEB1 regulates GRHL2 in a balance between EMT and MET [10-12,32]. However, we did not detect obvious GRHL2 binding sites in the promoter regions of the *ZEB1* or *ZEB2* genes. GRHL2 may regulate *ZEB1* and *ZEB2* indirectly in luminal-like and basal A breast cancers.

Taken together, this study provides a comprehensive genome-wide resource of GRHL2 binding sites and identifies specific and shared binding sites for GRHL2 in luminal-like and basal A subtype breast cancer. Overall, this study lays the foundation for unraveling signaling pathways regulated by GRHL2.

## Acknowledgements

# References

1       Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA: a cancer journal for clinicians* **66**, 7-30 (2016).

2       Yersal, O. & Barutca, S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol* **5**, 412-424, doi:10.5306/wjco.v5.i3.412 (2014).

3       Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155**, 1639-1651, doi:10.1016/j.cell.2013.11.029 (2013).

4       Sonzogni, O. *et al.* Reporters to mark and eliminate basal or luminal epithelial cells in culture and in vivo. *PLoS biology* **16**, e2004049, doi:10.1371/journal.pbio.2004049 (2018).

5       Frisch, S. M., Farris, J. C. & Pifer, P. M. Roles of Grainyhead-like transcription factors in cancer. *Oncogene*, doi:10.1038/onc.2017.178 (2017).

6       Bray, S. J. & Kafatos, F. C. Developmental function of Elf-1: an essential transcription factor during embryogenesis in Drosophila. *Genes & development* **5**, 1672-1683 (1991).

7       Mace, K. A., Pearson, J. C. & McGinnis, W. An epidermal barrier wound repair pathway in Drosophila is mediated by grainy head. *Science* **308**, 381-385, doi:10.1126/science.1107573 (2005).

8       Dompe, N. *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* **108**, E943-951, doi:10.1073/pnas.1100132108 (2011).

9       Chen, W. *et al.* 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* **285**, 40852-40863, doi:10.1074/jbc.M110.103812 (2010).

10      Quan, Y. *et al.* Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. *Cancer Biol Ther* **15**, 878-887, doi:10.4161/cbt.28877 (2014).

11      Werner, S. *et al.* Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *The Journal of biological chemistry* **288**, 22993-23008, doi:10.1074/jbc.M113.456293 (2013).

12      Cieply, B., Farris, J., Denvir, J., Ford, H. L. & Frisch, S. M. Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res* **73**, 6299-6309, doi:10.1158/0008-5472.CAN-12-4082 (2013).

13      Chung, V. Y. *et al.* GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943, doi:10.1038/srep19943 (2016).

14      Gregory, P. A. *et al.* An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol Biol Cell* **22**, 1686-1698, doi:10.1091/mbc.E11-02-0103 (2011).

15      Satoh, J.-i., Kawana, N. & Yamamoto, Y. Pathway analysis of ChIP-Seq-based NRF1 target genes suggests a logical hypothesis of their involvement in the pathogenesis of neurodegenerative diseases. *Gene regulation and systems biology* **7**, 139 (2013).

16      Mundade, R., Ozer, H. G., Wei, H., Prabhu, L. & Lu, T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* **13**, 2847-2852, doi:10.4161/15384101.2014.949201 (2014).

17      Wilanowski, T. *et al.* A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. *Mechanisms of development* **114**, 37-50 (2002).

18      Boglev, Y. *et al.* The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. *Developmental biology* **349**, 512-522, doi:10.1016/j.ydbio.2010.11.011 (2011).

19      Werth, M. *et al.* The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835-3845, doi:10.1242/dev.055483 (2010).

20      Lin, X., Tirichine, L. & Bowler, C. Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. *Plant Methods* **8**, 48, doi:10.1186/1746-4811-8-48 (2012).

21      Liu, C.-M. *et al.* SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. **28**, 878-879 (2012).

22      Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175-185 (1998).

23      Liao, P., Satten, G. A. & Hu, Y. J. PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genet Epidemiol* **41**, 375-387, doi:10.1002/gepi.22048 (2017).

24      Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

25      Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

26      Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).

27      Aue, A. *et al.* A Grainyhead-Like 2/Ovo-Like 2 Pathway Regulates Renal Epithelial Barrier Function and Lumen Expansion. *J Am Soc Nephrol* **26**, 2704-2715, doi:10.1681/ASN.2014080759 (2015).

28      Walentin, K. *et al.* A Grhl2-dependent gene network controls trophoblast branching morphogenesis. *Development* **142**, 1125-1136, doi:10.1242/dev.113829 (2015).

29      Gao, X. *et al.* 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* **110**, 9356-9361 (2013).

30      Senga, K., Mostov, K. E., 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. *Mol Biol Cell* **23**, 2845-2855, doi:10.1091/mbc.E12-02-0097 (2012).

31      Varma, S. *et al.* 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* **287**, 37282-37295, doi:10.1074/jbc.M112.408401 (2012).

32      Cieply, B. *et al.* Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res* **72**, 2440-2453, doi:10.1158/0008-5472.CAN-11-4038 (2012).

33      Xiang, X. *et al.* Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One* **7**, e50781, doi:10.1371/journal.pone.0050781 (2012).

34      Kumar, B. *et al.* Normal breast-derived epithelial cells with luminal and intrinsic subtype-enriched gene expression document inter-individual differences in their differentiation cascade. *Cancer Res*, doi:10.1158/0008-5472.CAN-18-0509 (2018).

35      Wang, X. *et al.* Epigenetic activation of HORMAD1 in basal-like breast cancer: role in Rucaparib sensitivity. *Oncotarget* **9**, 30115-30127, doi:10.18632/oncotarget.25728 (2018).

36      Kennecke, H. *et al.* Metastatic behavior of breast cancer subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 3271-3277, doi:10.1200/JCO.2009.25.9820 (2010).

37      de Silva Rudland, S. *et al.* Statistical association of basal cell keratins with metastasis-inducing proteins in a prognostically unfavorable group of sporadic breast cancers. *The American journal of pathology* **179**, 1061-1072, doi:10.1016/j.ajpath.2011.04.022 (2011).

38      Kao, J. *et al.* Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4**, e6146, doi:10.1371/journal.pone.0006146 (2009).

39      Stingl, J., Eaves, C. J., Zandieh, I. & Emerman, J. T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**, 93-109 (2001).

40      Chung, V. Y. *et al.* The role of GRHL2 and epigenetic remodeling in epithelial-mesenchymal plasticity in ovarian cancer cells. *Commun Biol* **2**, 272, doi:10.1038/s42003-019-0506-3 (2019).

# Supplemental data
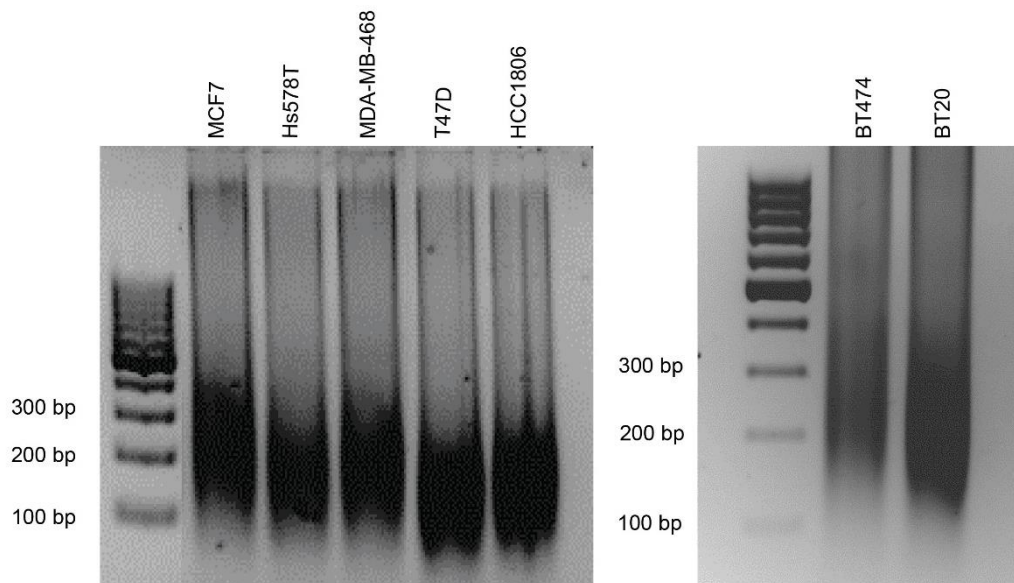


**Fig. S1. DNA fragmentation analysis by agarose gel electrophoresis.** After sonication, indicated samples were purified and loaded to 2% agarose gel.

**(Next page)Fig. S2. ChIP-PCR 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 PCR 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, T47D), 3 GRHL2-positive basal-A cell lines (BT20, HCC1806, MDA-MB-468), and 1 GRHL2-negative basal-B cell line (Hs578T).

Fig. S2

MCF7

BT20

T47D

HCC1806

BT474

MDA-MB-468

Hs578T

68
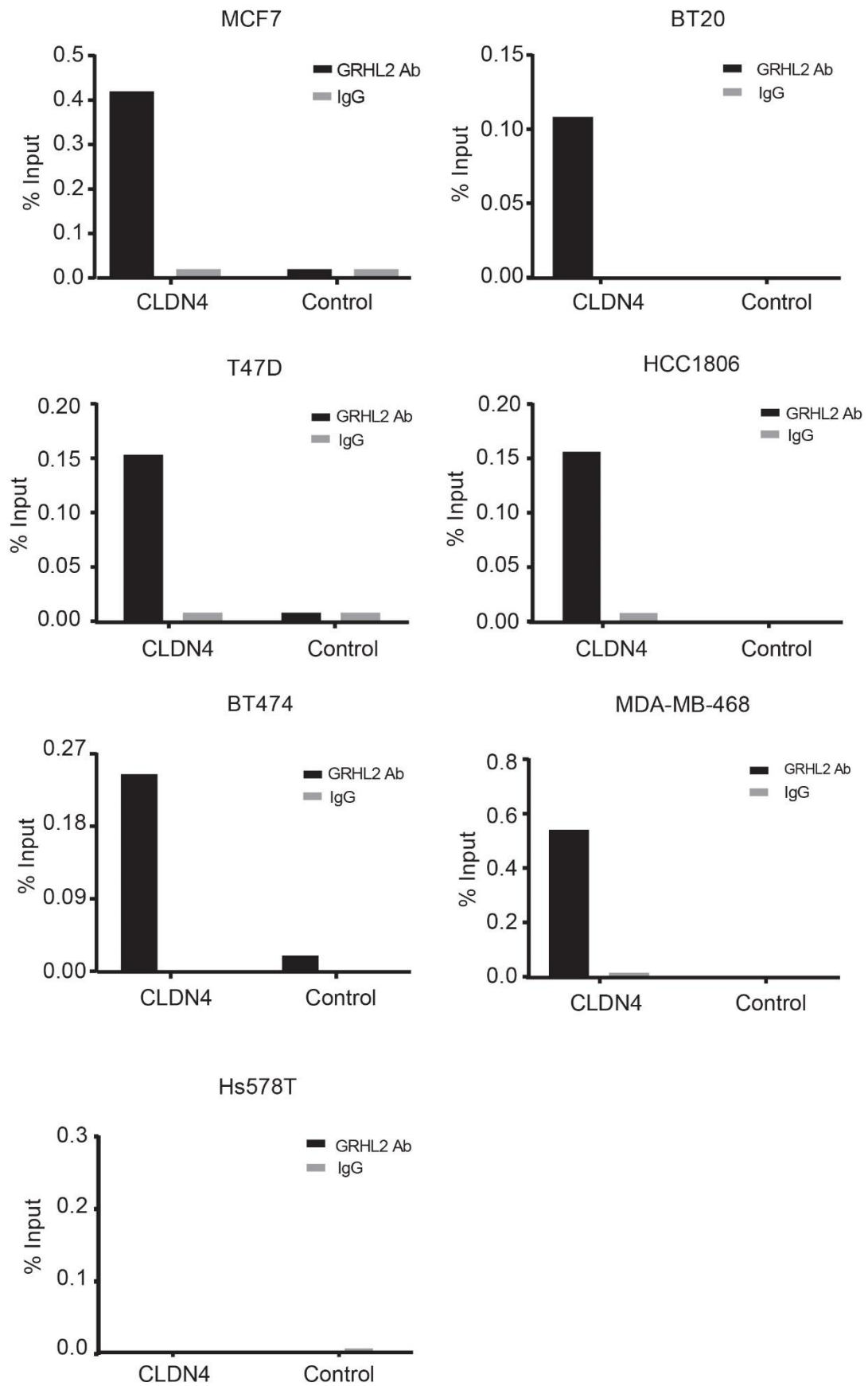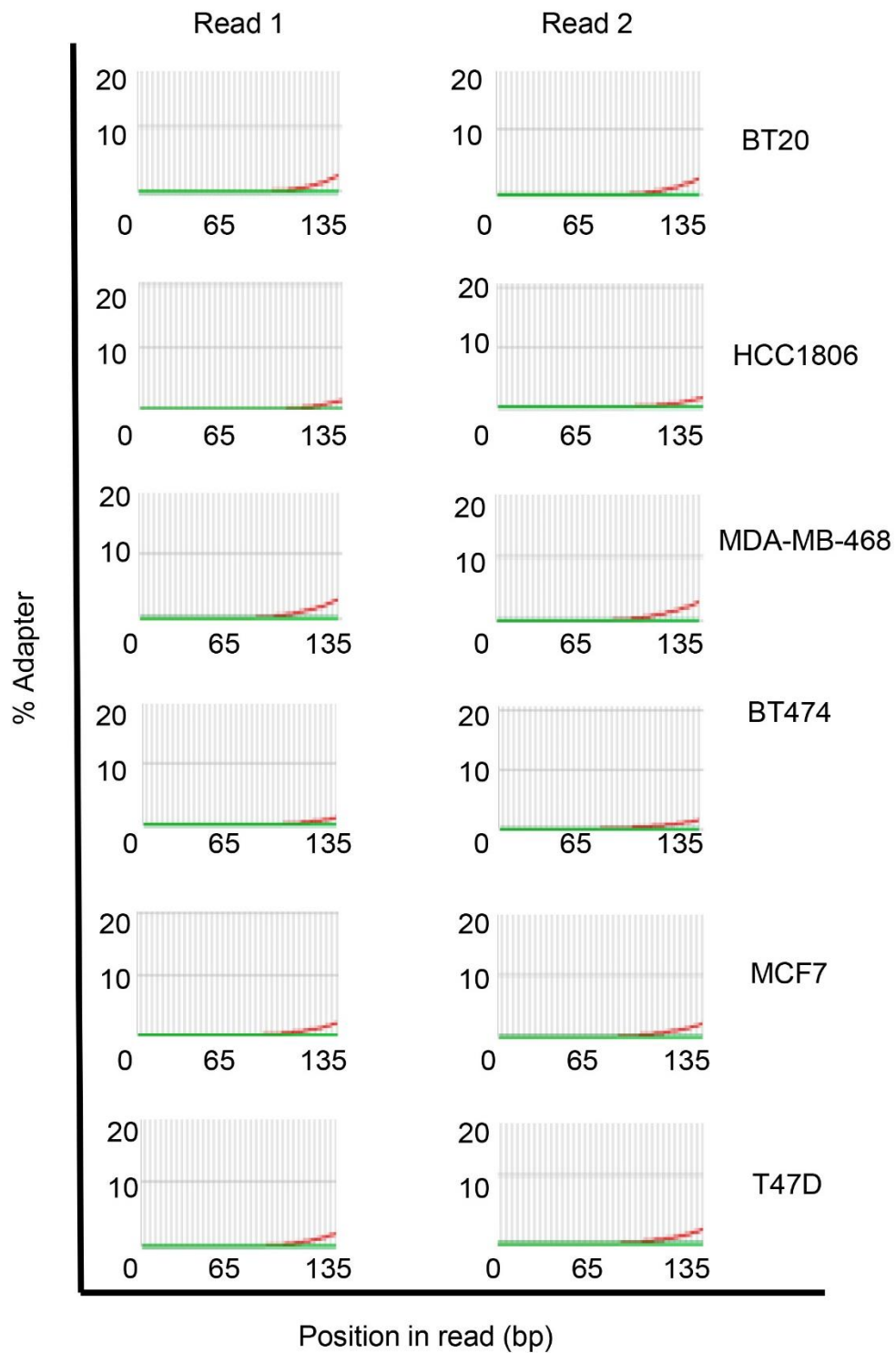
**Fig. S3. Cumulative presence of adapter sequences.** Results show that cumulative presence of adapter sequences is less than 5% in each cell sample, indicating that the data sets could be further analysed without adapter-trimming.
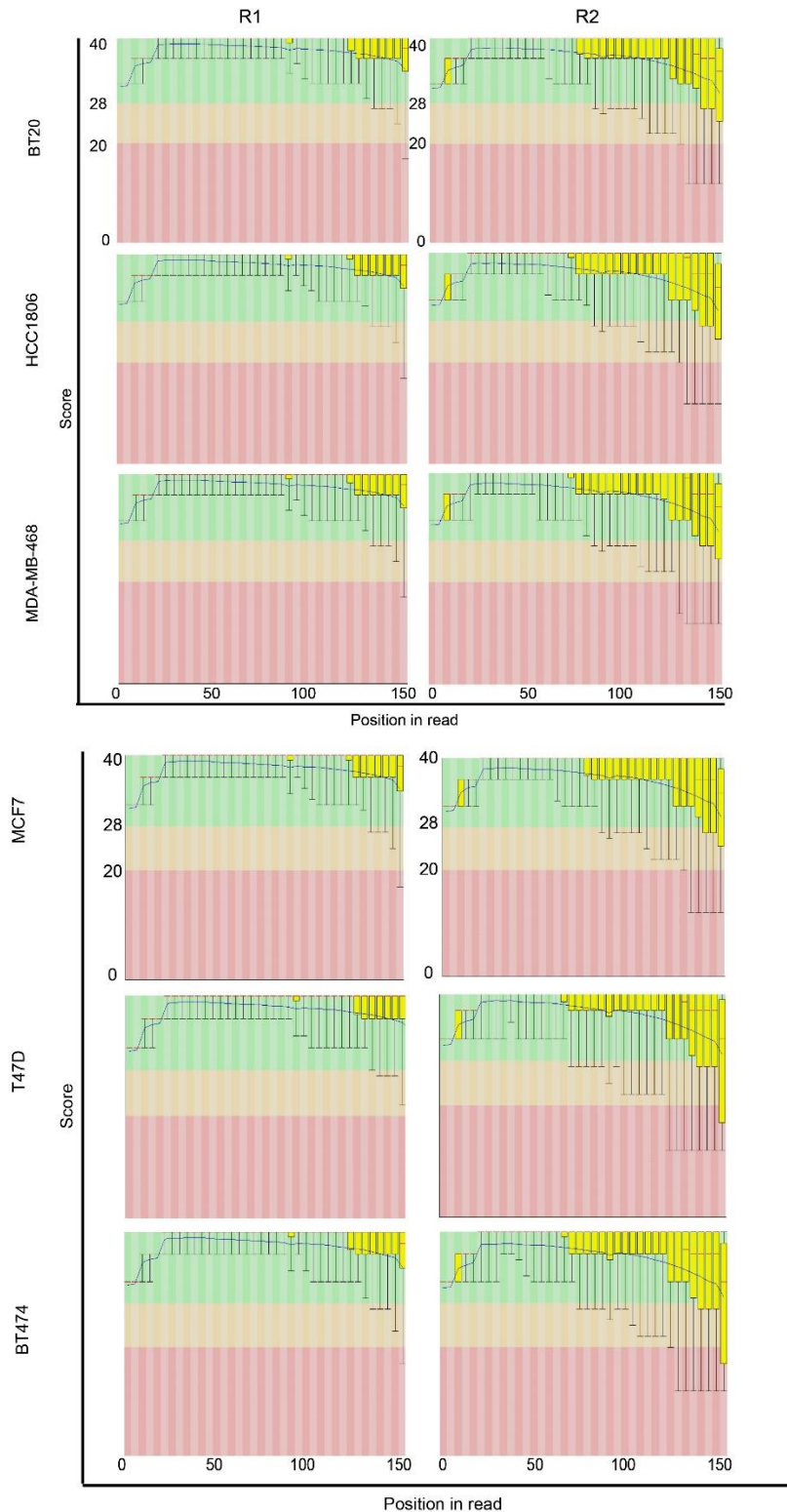
**Fig. S4. Per base sequence quality for all sequencing data sets.** Y axis is divided into good quality calls (coloured green), reasonable quality calls (coloured orange) and poor-quality calls (coloured red). In general, it is normal to observe base calls dropping into the red area towards the end of reads. The blue line representing the mean quality of base calls consistently stayed in the green area, indicating that sequencing data sets were of high quality.

70

**Table S1. Candidate GRHL2 target genes.** List of genes where GRHL2 was found to interact with the -1000 bp to +100 bp promoter region in all three basal A, all three luminal-like, or all basal A and luminal-like cell lines tested.

*Specific target genes in basal A subtype:*

| | | | | |
|---|---|---|---|---|
| TTK | RNF224 | DHDDS | SUPT7L | TFDP2 |
| CCDC171 | RAB23 | ILF2 | NMRAL1 | DNAH3 |
| PATJ | ZNF436 | SH2D4A | ZNF75D | CEP44 |
| HRCT1 | IVNS1ABP | ETV6 | NFKBIZ | AP2S1 |
| LOC100419583 | SRFBP1 | PLCD1 | STXBP5-AS1 | TFAP2A |
| PRG3 | PKD1L2 | NEK1 | A2ML1 | INHBA |
| CSTF2T | NCEH1 | PYM1 | TEX261 | VAT1 |
| NAPA | GGH | TMEM9 | MUC20 | HNRNPA1L2 |
| TGM5 | GOLGA5 | ADGRL2 | S100A9 | MDGA1 |
| HMGN4 | PPP1R13L | WWP1 | ATF3 | MIR548XHG |
| LOC153684 | ADGRF1 | LOC283299 | MTAP | LINC00989 |
| KLHDC9 | CDC25C | PDE12 | GDF5 | WDR5B |
| OTX1 | FGD3 | EYA2 | LOC100128770 | TSC22D1 |
| LINC01637 | SYTL2 | SLPI | RAD23A | ZNF563 |
| NUP155 | LIMCH1 | GTF3C2 | DUSP10 | PI3 |
| ZNF700 | PSMB6 | CASC4 | ADK | MINK1 |
| MORN3 | TMEM154 | ADGRG7 | PLS3 | FER1L4 |
| CEP162 | ZNF468 | TMEM173 | RNF224 | LOC101927822 |
| CARD6 | CLCN3 | TENT4A | MCFD2 | IGF2BP3 |
| NXT1 | DPY19L4 | SPATA32 | LINC01094 | UBP1 |
| C11orf52 | LINC00474 | KIAA0319 | LINC01344 | CBR4 |
| ZNF280D | RNF19A | IMPDH1 | ZNF682 | SOX2 |
| SEMA4B | DST | ALG10B | LEMD2 | MIR5684 |
| KIF16B | KLHDC7A | MORC4 | THAP11 | LOC105376114 |
| ACSL1 | TMEM256 | HMGB2 | PPP2R2C | SPINT2 |
| MIR4432HG | TAB3 | IFT22 | GSDMC | SLC29A2 |
| RPN2 | OXR1 | SYNE2 | THADA | RGL1 |
| RPSAP58 | KMT5C | MICB-DT | SLC35F2 | BCLAF1 |
| COPS5 | TMEM102 | KHDC4 | MIR7706 | TOMM70 |
| PUM2 | SAMD12-AS1 | ZNF221 | PIGV | DHRS4-AS1 |
| CFDP1 | MUC15 | BCAR4 | ANAPC5 | KLK8 |
| SRP68 | KIF5C | SCNM1 | ETFBKMT | EPS8 |
| PIGA | GLP2R | STK3 | STK38L | POU2F3 |
| RNVU1-15 | GPBP1L1 | EBPL | LOC93429 | ZNF552 |
| UFSP1 | YWHAZ | ANKRD2 | HRH1 | ZNF562 |
| HSPA4L | ACIN1 | PTGR2 | CRYM-AS1 | ADGRE2 |
| MIR3165 | CITED4 | FZR1 | CAST | LOC103021296 |
| ATP2C1 | ADGRF2 | MIR378J | TLCD2 | VPS50 |
| PGAM1P5 | FLJ42969 | SNX27 | RXRA | SNX3 |
| LYSMD3 | Septin8 | MUC1 | ABT1 | C11orf74 |
| ZC3H15 | FNTA | DTL | PHF23 | LRG1 |
| TRIT1 | GS1-124K5.11 | RPAIN | MDH1B | CEBPB-AS1 |
| NFIA | HS3ST1 | ABCA12 | KRTDAP | KYAT3 |

| | | | | |
|---|---|---|---|---|
| ME1 | IL36RN | LINC00869 | B3GALT4 | PHETA1 |
| SLTM | ZNF695 | MSANTD4 | COL4A5 | COTL1 |
| RNF225 | ZFAND6 | LOC101928008 | BTN3A3 | PITPNM1 |
| NFKBIA | BCAS2 | MIR135B | PGRMC2 | TMPRSS11E |
| PSORS1C1 | GSE1 | GLB1 | ACKR2 | PIP4P2 |
| STAP2 | RNU6-2 | TGIF1 | CD63 | RASGEF1B |
| ZNF844 | AFG3L2 | ZNF284 | PA2G4 | CHRNE |
| DDX12P | UPK1A | GPR156 | HMGCR | HS6ST2 |
| BLNK | GTF2H4 | WDR75 | ZNF234 | LDHA |
| MIR4422 | ENO1-AS1 | RCBTB1 | TMPRSS4 | RALY |
| RASL11A | ATG14 | OSGIN1 | ATAT1 | PDE4DIP |
| MIR4799 | IKBKE | HACE1 | TRIM16L | LOC100506113 |
| WBP1 | BCL9 | LINC01393 | MTMR11 | KIAA0513 |
| UPK2 | RBM47 | PDSS1 | LRRC23 | PGLS |
| S100A12 | ALKBH8 | GCNT2 | ANKMY1 | INTS1 |
| VOPP1 | DTWD2 | XDH | ZNF140 | TM4SF4 |
| ACE2 | LARS | ATP1A1 | RGS3 | LINC01634 |
| LOC101928977 | EIF4G1 | CRTAM | MON2 | VSIG10L |
| UCA1 | RNF222 | CHD8 | SLMO2-ATP5E | SLC2A11 |
| IL1RN | EXOC6 | TIGD2 | LINC02447 | TSHZ1 |
| SCGB1B2P | MCCC1 | ARHGAP27 | ADAMTS6 | GGTLC2 |
| GJB3 | SFTA2 | IKBKG | EIF2AK2 | KDM5C |
| SMIM13 | DLGAP1-AS2 | PDCD10 | HIPK1-AS1 | BNIPL |
| UNC13D | MGC32805 | C7orf77 | SEC11A | PEX26 |
| MRPL1 | RPIA | ELF3 | DENND6A-DT | IMPDH1 |
| FEZF1 | CCDC26 | GORAB | EEF2 | KAT14 |
| ALDH4A1 | ACAD9 | GTPBP3 | PLS1 | GBA |
| BEND5 | ANXA11 | PLEKHF2 | LOC100294145 | EIF1AD |
| RNF32 | ANKRD54 | S100P | CD164 | LOC101927151 |
| CSNK1G3 | KDM6B | REXO2 | LINC01588 | MRPL24 |
| CORO2B | RETREG1 | SCAMP3 | C1orf226 | CAB39L |
| LINC01559 | DEPDC1-AS1 | CENPA | TCHHL1 | IL17RE |
| LINC01354 | SLC37A4 | AATF | PPARD | TSKU |

*Specific target genes in luminal-like subtype*

| | | | | |
|---|---|---|---|---|
| EPHA1 | SSR4P1 | OR7E91P | MIR6070 | ARRDC3 |
| FLJ31356 | TMEM40 | ADGRF4 | TGM1 | CLDN4 |
| IQCK | ZNF440 | MESP1 | NIPSNAP1 | MGP |
| LOC101927391 | TGIF1 | TUFT1 | PTPN14 | CARD14 |
| NAALADL2 | ZNF433 | ARSD | FRRS1 | CCDC12 |
| RIMS1 | SMG8 | TIGAR | GAR1 | ZNF823 |
| EDEM2 | PRR15L | TMEM79 | PPOX | MIR4676 |
| LINC01213 | RBBP8NL | LOC101927318 | TMPRSS11F | LINC00359 |
| GRAMD2B | ANXA9 | AMD1 | MIR4513 | PDGFB |
| NIPAL2 | AFG1L | FAF1 | ZNF799 | LINC00456 |
| SCAMP4 | GGTLC1 | AIFM1 | IFRD1 | ZNF274 |
| VEPH1 | DAZAP1 | RPL41 | ZNF20 | TRPC4AP |
| KCNJ13 | ZNF44 | STX19 | SYTL5 | SLC4A7 |
| ARHGAP32 | NEU1 | BATF | VGLL1 | CMTR2 |

| | | | | |
|---|---|---|---|---|
| MIR6773 | RPL32P3 | ANKRD22 | BCAS1 | ZMYND8 |
| TYSND1 | OVOL2 | LOC100506098 | GPNMB | LRP10 |
| CNP | LOC101927911 | ELF5 | CFAP45 | EEA1 |
| ADIPOQ | ST3GAL4 | TMPRSS13 | GPR108 | LOC344967 |
| LOC101927272 | SLC9A1 | RNVU1-14 | NXT1 | DLG4 |
| BBOX1 | SEMA4A | UBE2A | RAP2B | ERP27 |
| CLDN8 | ITFG2 | IGSF9 | EHF | EPN3 |
| FMN1 | CD46 | ARHGEF19 | LINC02408 | PDCD2 |
| PKP2 | SLITRK6 | ERBB3 | PSCA | MAPK10 |
| FKBP2 | GRAMD1C | VIPAS39 | EIF2B5 | HRH1 |
| LINC00346 | DNAJC5B | DNAAF5 | JUP | LIMA1 |
| RBL2 | ALDH3B2 | ARHGAP24 | EEF1E1 | WSB2 |
| EPB41L1 | UBALD2 | SBNO1 | C1orf116 | PGLYRP2 |
| ZBTB20 | LINC01405 | GMPR2 | TBL1X | YAP1 |
| RASAL2 | BMF | SNORA38 | EHF | ROCK1P1 |
| SLC25A45 | P2RY6 | SORT1 | PLA2G4B | SLC41A3 |
| ZER1 | IKZF2 | LOC100132781 | RNU5B-1 | ZNHIT6 |
| ATAD3B | LINC00885 | CHD3 | LOC100129917 | HIST2H2AB |
| GMEB1 | C4orf3 | IVL | RAB25 | TRIL |
| CBLB | CDS1 | MACROD1 | KRT80 | ZNF443 |
| TJP2 | | | | |

*Common target genes between basal-like and luminal-like subtypes*

| | | | | |
|---|---|---|---|---|
| ARHGEF38 | ERLNC1 | PIM2 | LOC101927296 | LOC102724064 |
| GINS2 | MIR6784 | MTERF2 | SLC40A1 | PIK3C2G |
| PGR | NME7 | CRISP3 | DSCAM-AS1 | LOC148709 |
| MIR4328 | KLK12 | SLFN12 | SFTPA2 | LOC102724163 |
| PROM2 | ATP6V0A4 | SLC10A5 | TP53INP2 | ZP1 |
| MUCL1 | PPEF1 | FMO9P | PURG | ASCL2 |
| DLX5 | LINC00938 | HIST2H2BF | PRIM2 | JADE1 |
| PDE4D | KRTAP3-1 | | | |

# Chapter 4

# Dynamic changes in nascent RNA after GRHL2 loss in luminal-like breast cancer

Zi Wang[1], Yao-Jun Chen[1], Mats Ljungman[2], Erik HJ Danen[1,3]

[1]Leiden Academic Center for Drug Research, Leiden University, Leiden, The Netherlands; [2]University of Michigan, Ann Arbor, MI 48109; [3]correspondence to Erik HJ Danen, e.danen@lacdr.leidenuniv.nl

# Abstract

GRHL2 drives expression of key epithelial genes and supports proliferation, survival, and epithelial differentiation. It plays a dual role in cancer by stimulating proliferation and suppressing EMT. GRHL2 has been reported to act as a transcription factor as well as a modulator of gene expression through epigenetic mechanisms. The relevant genetic programs controlled by GRHL2 in cancer are not resolved. In the present study, the response to GRHL2 loss in luminal breast cancer cells was studied by combining an MCF7 conditional knockout model with Bru-seq analysis. The rate of RNA synthesis of 264 and 244 genes was upregulated or downregulated, respectively, for at least one out of four time points following GRHL2 loss ranging from 1-16 days. Five dynamic response patterns were characterized and GRHL2-controlled canonical pathways and signaling networks were identified. Collectively, this study characterizes patterns of RNA synthesis regulated by GRHL2 and identifies signaling pathways regulated by GRHL2.

# Introduction

GRHL2 is a mammalian homolog of the *Drosophila* Grainyhead gene. GRHL2 has a crucial role in neural tube closure, epithelial cell morphology, cancer cell proliferation and migration [1-3]. It is widely accepted that GRHL2 has dual roles in cancer development [4,5]. GRHL2 can inhibit epithelial to mesenchymal transition (EMT) by upregulating E-cadherin and Claudin4 [6] and downregulating ZEB1 [7,8]. On the other hand, GRHL2 is frequently overexpressed or amplified in breast cancer [9], lung cancer [10], and ovarian cancer [11] and high expression of GRHL2 was associated with histological differentiation and lymphatic metastasis in pancreatic carcinoma [12].

The relevant genetic programs controlled by GRHL2 in cancer are not resolved. GRHL2 has been reported to act as a transcription factor as well as a modulator of gene expression through epigenetic mechanisms [13,14]. Gene regulation includes transcriptional initiation, RNA processing, post-transcriptional modification, translation and post-translational modification. Conventional RNA-seq is used for analysis of steady-state RNA levels whereas bromouridine sequencing (Bru-seq) measures nascent RNA, allowing for direct assessment of changes in DNA transcription [15,16]. Bru is relatively non-toxic as compared to other ribonucleotide analogs and is widely used to label nascent RNA in vitro and in cells [17,18] [19].

In this study, we used Bru-seq to investigate genome-wide dynamic changes of nascent RNA induced by GRHL2 loss in an MCF7 conditional knockout model. Following identification of differentially expressed genes in response to GRHL2 loss, bioinformatics analysis was performed to predict signaling networks regulated by GRHL2. Thus, GRHL2-controlled gene networks were unraveled.


## Materials and methods

### Cell culture and lentiviral transduction

MCF7 human breast cancer cells were obtained from the American Type Culture Collection. Cells were cultured in RPMI1640 medium with 10% fetal bovine serum, 25 U/mL penicillin and 25 µg/mL streptomycin at 37°C and 5% $CO_2$. 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 promotor-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.

**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 assay (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).

**Bru-seq**

At different timepoints after doxycycline-induced deletion of GRHL2, cells were 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 [15]. Subsequently, cDNA libraries were generated using the Illumina TruSeq library

kit and sequenced using the Illumina NovaSeq 6000 Sequencing System. Sequencing and read mapping were carried out as previously described [15,20]

## Bioinformatics 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 sgGRHL2 (1) as well as sgGRHL2 (2) were selected and genes responding also in the context of sgCTR were eliminated from this list. Canonical pathways and networks analysis was performed with the Ingenuity Pathways Analysis (IPA) software (Ingenuity Systems, USA). A heat map was generated by R. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) [21,22] was utilized to identify signaling pathways associated with GRHL2 loss. Gene Ontology (GO) terms (biological process, cellular component and protein class) analysis was performed by Protein Analysis Through Evolutionary Relationships (PANTHER) database [23].

## ChIP-PCR

Chromatin preparation was described previously 24. For ChIP-PCR, chromatin fragments were immunoprecipitated with control IgG or anti-GRHL2 antibodies (Sigma; HPA004820). Precipitates were eluted 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. The following primers were used for ChIP-PCR: E2F2 forward, tcctgggaagaggaatgatg; E2F2 reverse, caggcagcttgggagagtag; CDCA7L forward, tttggggcttgttttgtttt; CDCA7L reverse: ggtgtggaggcctactgtgt; control (an intergenic region upstream of the GAPDH locus) forward, atgggtgccactggggatct; control reverse, tgccaaagcctaggggaaga. ChIP-PCR data were analyzed using the $2^{-\Delta\Delta Ct}$ method [25].
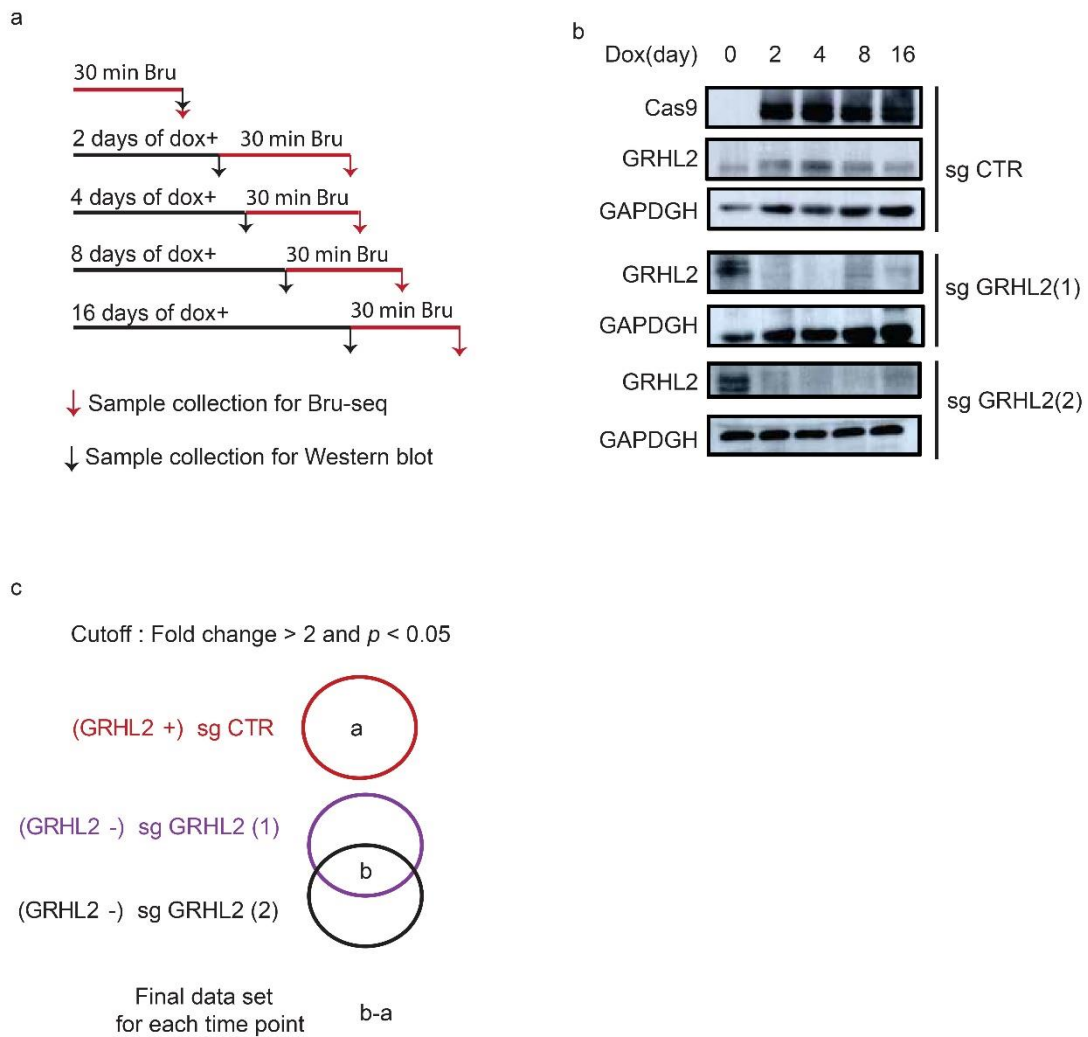


**Fig. 1 Bru-seq sample preparation and Bru-seq data analysis strategy. (a)** Bromouridine (Bru) labeling of nascent RNA was carried out for 30 minutes at the indicated time points after doxycycline (dox)-induced GRHL2 deletion. **(b)** Western blot analysis of GRHL2 expression levels at the indicated time points in sgCTR and sgGRHL2 transduced MCF7 cells. GAPDH serves as loading control. **(c)** Strategy for Bru-

seq data analysis. Each circle represents a gene set with differential transcription relative to the condition where no doxycycline was added.

## Results

### Dynamic regulation of RNA synthesis in response to GRHL2 loss

Using conditional CRISPR-Cas9 MCF7 cells, Bru-seq was carried out to investigate the dynamic changes in DNA transcription triggered by GRHL2 loss. At 0, 2, 4, 8, or 16 days after doxycycline-induced GRHL2 knockout, cells were incubated with Bru for 30 minutes to label nascent RNA (Fig. 1a) or they were analyzed by Western blot to examine the expression of GRHL2 protein (Fig. 1b).

To identify GRHL2-regulated genes, for each time point, the $\log_2$ average fold change (AFC) of transcription induced by doxycycline treatment in the two sgGRHL2 and the control sgRNA sample was determined. A list of genes was generated whose transcription was altered in both sgGRHL2 samples (FC>2; $p$<0.05 or FC<0.5; $p$<0.05) but not in the sgCTR sample (Fig. 1c). Using these criteria, 264 genes were upregulated and 244 genes were downregulated in at least one time point after GRHL2 loss (Table S1).

### Distinct dynamic patterns of response to GRHL2 depletion

GRHL2-regulated genes were clustered in a heat map using the AFC at each time point in sgGRHL2 (1) and sgGRHL2 (2) cells (Fig. 2a). In response to GRHL2 loss, one cluster of genes exhibited rapid and continuing upregulation in RNA synthesis. For instance, *LAMB3* encoding the β2 unit of the trimeric basement membrane protein laminin-332 [26] was rapidly induced after GRHL2 loss (Fig. 2b). Another cluster showed rapid and sustained downregulation of RNA synthesis following GRHL2 deletion. This cluster included *UBB*, encoding
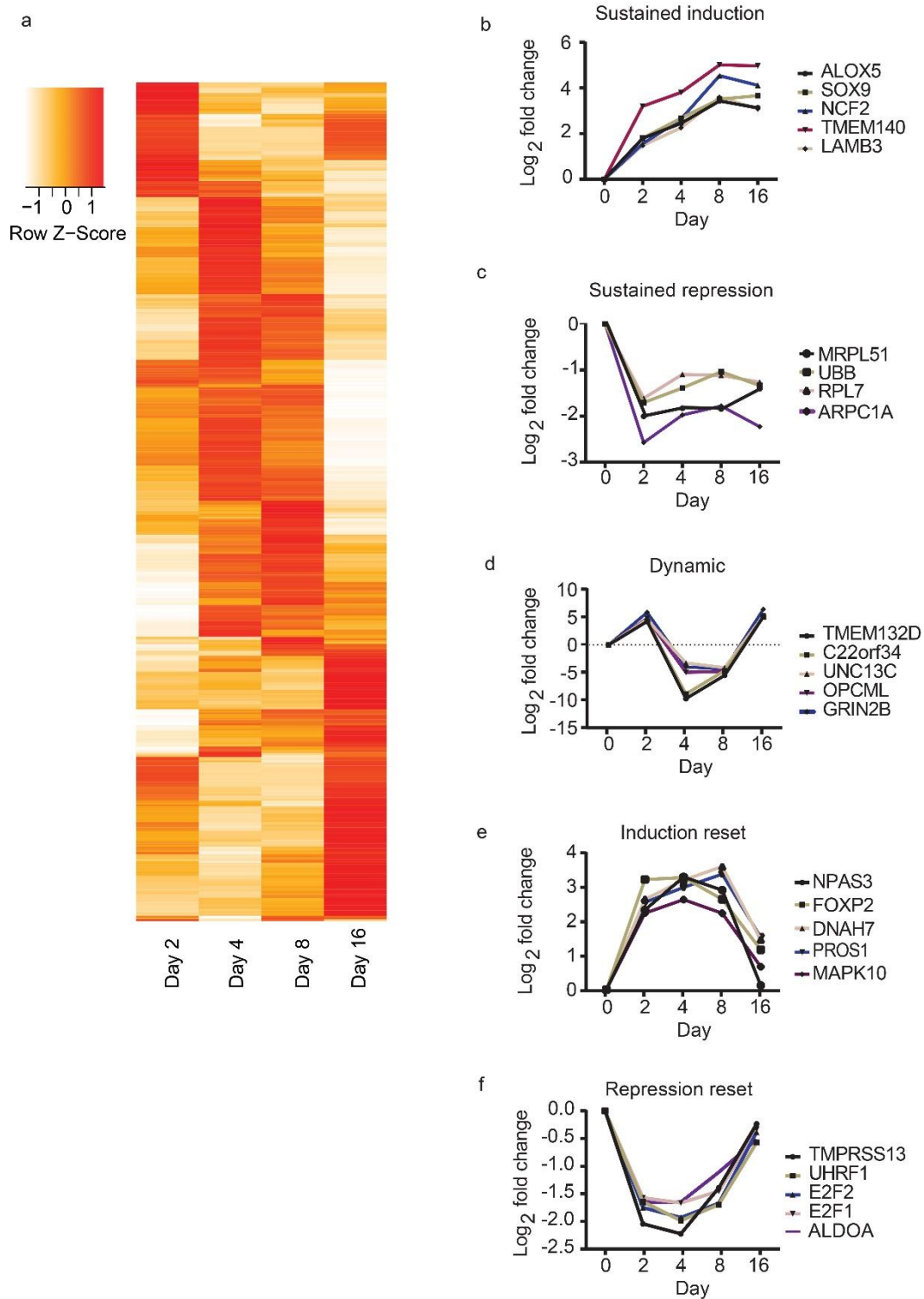
**Fig. 2 Dynamic changes in RNA synthesis following GRHL2 loss. (a)** Heat map for GRHL2 loss response genes. **(b-f)** After GRHL2 abrogation, genes are categorized according to RNA synthesis patterns. The line graph depicts the $\log_2$ average fold change (AFC) of transcription in sgGRHL2 (1) and sgGRHL2 (2) cells. **Dynamic**: genes

with AFC>2; p<0.05 at some and AFC<0.5; p<0.05 at other time points. **Sustained induction**: genes with AFC>2; p<0.05 at all time points. **Sustained repression:** genes with AFC<0.5; p<0.05 at all time points. **Induction reset:** genes with AFC>2 at early time points followed by a return to 1<AFC<2 at day16. **Repression reset:** genes with AFC<0.5 at early time points followed by a return to 0.5<AFC<1 at day16.

the highly conserved ubiquitin protein that is involved in the regulation of protein degradation, signaling, and gene expression [27]. The downregulation in RNA synthesis of *UBB* was observed at each labeling period (Fig. 2c). Another cluster of genes displayed a dynamic transcriptional response following GRHL2 loss. For example, GRHL2 loss enhanced transcription of the *GRIN2B* gene (encoding GluN2B, a subunit of NMDA-type glutamate-gated ion channels [28]) within 2 days, followed by a repression at day 4 and 8, and followed by another peak of enhanced transcription at day 16 after GRHL2 deletion (Fig. 2d). The "induction reset" cluster included genes whose transcription was transiently induced initially followed a repression phase where transcription returned to baseline. This cluster included *FOXP2* encoding a forkhead transcription factor (Fig. 2e). The "repression reset" cluster showed an opposite pattern with an initial repression that returned to baseline at later timepoints, and included the *E2F1* gene encoding the E2F1 transcription factor involved in cell survival and proliferation [29] (Fig. 2f).

**Predicted signaling networks regulated by GRHL2**

IPA software was utilized to elucidate GRHL2-regulated signaling pathways and networks from the differently expressed genes at the different time points following GRHL2 depletion. Canonical pathway results predicted changes in Granzyme A signaling, remodeling of epithelial adherens junctions, mTOR signaling, and DNA methylation and transcriptional repression signaling at each time point after GRHL2 loss (Table 1). At early time points (day 2, 4 and 8), significantly enriched canonical pathways included EIF2 signaling, germ cell-

sertoli cell junction signaling, pancreatic adenocarcinoma signaling, and a transcriptional regulatory network in embryonic stem cells (Table 1).

Next, using IPA, networks associated with multiple biological functions and diseases were identified and ranked according to the score. The top 10 networks scored >22, indicating that the likelihood that genes in these networks were not connected was $<10^{-22}$ (Table 2). Overall, networks associated with

**Table 1. Top 10 canonical pathways responding to GRHL2 loss at the indicated time points generated by IPA.**

| Canonical pathway | -log(p-value) | Downregulated | Upregulated |
|---|---|---|---|
| **Day 2** | | | |
| EIF2 Signaling | 6.33 | 16/224 (7%) | 0/224 (0%) |
| Granzyme A Signaling | 4.96 | 5/20 (25%) | 0/20 (0%) |
| Remodeling of Epithelial Adherens Junctions | 4.95 | 7/69 (10%) | 1/69 (1%) |
| DNA Methylation and Transcriptional Repression Signaling | 4.91 | 6/34 (18%) | 0/34 (0%) |
| Germ Cell-Sertoli Cell Junction Signaling | 4.10 | 7/172 (4%) | 4/172 (2%) |
| Transcriptional Regulatory Network in Embryonic Stem Cells | 3.74 | 6/54 (11%) | 0/54 (0%) |
| Pancreatic Adenocarcinoma Signaling | 3.53 | 3/109 (3%) | 5/109 (5%) |
| Epithelial Adherens Junction Signaling | 3.19 | 7/153 (5%) | 2/153 (1%) |
| Signaling by Rho Family GTPases | 2.82 | 5/243 (2%) | 6/243 (2%) |
| mTOR Signaling | 2.77 | 6/210 (3%) | 4/210 (2%) |
| **Day 4** | | | |
| EIF2 Signaling | 5.77 | 16/224 (7%) | 0/224 (0%) |
| Remodeling of Epithelial Adherens Junctions | 5.59 | 6/69 (9%) | 3/69 (4%) |
| Granzyme A Signaling | 4.75 | 5/20 (25%) | 0/20 (0%) |
| DNA Methylation and Transcriptional Repression Signaling | 4.67 | 6/34 (18%) | 0/34 (0%) |
| Transcriptional Regulatory Network in Embryonic Stem Cells | 3.51 | 6/54 (11%) | 0/54 (0%) |
| Pancreatic Adenocarcinoma Signaling | 3.24 | 3/109 (3%) | 5/109 (5%) |
| Germ Cell-Sertoli Cell Junction Signaling | 3.11 | 6/172 (3%) | 4/172 (2%) |
| Breast Cancer Regulation by Stathmin1 | 2.62 | 7/200 (4%) | 3/200 (2%) |
| Inhibition of Angiogenesis by TSP1 | 2.59 | 1/34 (3%) | 3/34 (9%) |
| mTOR Signaling | 2.47 | 6/210 (3%) | 4/210 (2%) |
| **Day 8** | | | |
| Remodeling of Epithelial Adherens Junctions | 5.60 | 6/69 (9%) | 3/69 (4%) |
| EIF2 Signaling | 5.12 | 15/224 (7%) | 0/224 (0%) |
| Granzyme A Signaling | 4.75 | 5/20 (25%) | 0/20 (0%) |
| DNA Methylation and Transcriptional Repression Signaling | 4.67 | 6/34 (18%) | 0/34 (0%) |
| Transcriptional Regulatory Network in Embryonic Stem Cells | 3.52 | 6/54 (11%) | 0/54 (0%) |
| Pancreatic Adenocarcinoma Signaling | 3.25 | 3/109 (3%) | 5/109 (5%) |
| Germ Cell-Sertoli Cell Junction Signaling | 3.12 | 6/172 (3%) | 4/172 (2%) |
| Inhibition of Angiogenesis by TSP1 | 2.59 | 1/34 (3%) | 3/34 (9%) |
| Synaptogenesis Signaling Pathway | 2.54 | 6/312 (2%) | 7/312 (2%) |
| mTOR Signaling | 2.47 | 6/210 (3%) | 4/210 (2%) |
| **Day 16** | | | |
| Granzyme A Signaling | 5.68 | 5/20 (25%) | 0/20 (0%) |
| mTOR Signaling | 3.26 | 5/210 (2%) | 4/210 (2%) |
| Synaptogenesis Signaling Pathway | 3.14 | 2/312 (1%) | 9/312 (3%) |
| Remodeling of Epithelial Adherens Junctions | 3.01 | 3/69 (4%) | 2/69 (3%) |
| GP6 Signaling Pathway | 2.69 | 0/119 (0%) | 6/119 (5%) |
| DNA Methylation and Transcriptional Repression Signaling | 2.21 | 3/34 (9%) | 0/34 (0%) |
| Opioid Signaling Pathway | 2.18 | 0/250 (0%) | 8/250 (3%) |
| Axonal Guidance Signaling | 2.10 | 2/486 (0%) | 10/486 (2%) |
| T Helper Cell Differentiation | 2.06 | 0/73 (0%) | 4/73 (5%) |
| Glioma Invasiveness Signaling | 2.04 | 1/74 (1%) | 3/74 (4%) |

diseases including cancer, networks associated with cell cycle, DNA, and RNA regulation, and networks associated with cell-to-cell signaling and interaction

were predicted at most timepoints tested. At day 2, the most enriched networks were associated with cancer, protein synthesis and RNA damage and repair. At day 4, the top enriched networks were linked to developmental disorder, embryonic development and organismal development, in which *AURKB* and *E2F1* represented core genes that were most interconnected with other genes (Fig. 3b). The top enriched networks at day 8 were associated with cell cycle, cellular assembly, DNA replication, recombination and repair. At day 16, the top molecular networks predicted by IPA were closely related to cell to cell signaling and interaction, nervous system development, RNA damage and repair (Fig. 3d).

**Table 2. Top 10 networks responding to GRHL2 loss at the indicated time points generated by IPA.**

| Top Diseases and Functions | Score | Focus Molecules |
|---|---|---|
| **Day 2** | | |
| Cancer, Protein Synthesis, RNA Damage and Repair | 62 | 33 |
| Cellular Assembly and Organization, Developmental Disorder, DNA Replication, Recombination, and Repair | 59 | 32 |
| Cancer, Organismal Injury and Abnormalities, Respiratory Disease | 59 | 32 |
| Connective Tissue Disorders, Developmental Disorder, Hereditary Disorder | 48 | 28 |
| Cellular Development, Connective Tissue Development and Function, Tissue Development | 34 | 22 |
| Cell Cycle, Hereditary Disorder, Organismal Injury and Abnormalities | 30 | 20 |
| Cell Death and Survival, Neurological Disease, Organismal Injury and Abnormalities | 26 | 18 |
| Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair | 24 | 17 |
| Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Cellular Function and Maintenance | 24 | 17 |
| Endocrine System Development and Function, Protein Synthesis, Small Molecule Biochemistry | 22 | 16 |
| **Day 4** | | |
| Developmental Disorder, Embryonic Development, Organismal Development | 67 | 35 |
| Cellular Assembly and Organization, DNA Replication, Recombination, and Repair, Post-Translational Modification | 58 | 32 |
| Cancer, Protein Synthesis, RNA Damage and Repair | 55 | 31 |
| Cellular Assembly and Organization, Cellular Function and Maintenance, Hematological Disease | 55 | 31 |
| Developmental Disorder, Hereditary Disorder, Metabolic Disease | 49 | 29 |
| Cell Death and Survival, Cellular Movement, Post-Translational Modification | 37 | 24 |
| Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair | 37 | 24 |
| Cellular Compromise, Energy Production, Nucleic Acid Metabolism | 31 | 21 |
| Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder | 31 | 21 |
| Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Nervous System Development and Function | 25 | 18 |

| Day 8 | | |
|---|---|---|
| Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair | 64 | 34 |
| Cancer, Protein Synthesis, RNA Damage and Repair | 58 | 32 |
| Nervous System Development and Function, Neurological Disease, Organ Morphology | 58 | 32 |
| Cancer, Hematological Disease, Immunological Disease | 52 | 30 |
| Energy Production, Nucleic Acid Metabolism, Small Molecule Biochemistry | 42 | 26 |
| Cellular Assembly and Organization, DNA Replication, Recombination, and Repair, Post-Translational Modification | 33 | 22 |
| Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder | 31 | 21 |
| Cancer, Hematological Disease, Organismal Injury and Abnormalities | 31 | 21 |
| Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair | 29 | 20 |
| Cardiovascular System Development and Function, Cell Morphology, Organ Development | 25 | 18 |
| **Day 16** | | |
| Cell-To-Cell Signaling and Interaction, Nervous System Development and Function, RNA Damage and Repair | 58 | 30 |
| Cellular Development, Cellular Growth and Proliferation, Embryonic Development | 45 | 25 |
| Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder | 37 | 22 |
| Gastrointestinal Disease, Inflammatory Disease, Inflammatory Response | 33 | 20 |
| Carbohydrate Metabolism, Lipid Metabolism, Small Molecule Biochemistry | 31 | 19 |
| Cardiovascular Disease, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair | 26 | 17 |
| Cancer, Organismal Injury and Abnormalities, Reproductive System Disease | 26 | 17 |
| Cell Cycle, Protein Synthesis, RNA Damage and Repair | 24 | 16 |
| Cancer, Organismal Injury and Abnormalities, Reproductive System Disease | 24 | 16 |
| Hereditary Disorder, Neurological Disease, Organismal Injury and Abnormalities | 22 | 15 |

## Predicted biological processes and functional pathways regulated by GRHL2

The PANTHER classification system was utilized to identify biological process, cellular component and protein classification predicted to be associated with the genes regulated in response to GRHL2 depletion. Biological processes at all 4 timepoints included cellular process, metabolic process, and biological regulation (Fig. 4a). For cellular component, differentially transcribed genes induced by GRHL2 loss were predominantly involved in organelle, extracellular region, protein-containing complex, membrane and cell junction for each time point (Fig. 4b). In terms of protein classification, genes transcriptionally affected by GRHL2 depletion were enriched in nucleic acid binding, hydrolase, transcription factor, signaling molecule and enzyme modulator for all time points (Fig. 4c). Subsequently, DAVID was utilized to investigate whether GRHL2-regulated genes identified by Bru-seq were enriched for known functional pathways. At each time point, GRHL2-regulated genes were significantly enriched in pathways associated with cancer, focal adhesion and ECM receptor interaction, and several signaling pathways (Fig. 5a). DAVID also identified enrichment of signaling pathways including those involved in viral carcinogenesis, ribosome, alcoholism and systemic lupus erythematosus

signaling at each time point whereas pathways involved in biosynthesis of amino acids, carbon metabolism, transcriptional misregulation in cancer, DNA replication and cell cycle signaling were identified at early timepoints (Fig. 5b).

We and others have reported that GRHL2 loss is associated with growth arrest [11,30]. Consistent with this notion, the RNA synthesis rate of several genes involved in cell cycle progression and DNA replication were rapidly suppressed in response to GRHL2 loss (i.e., *E2F2, CDCA7L, SFN and MCM2*) [31-34] (Fig. 6a-d). Our previous ChIP-seq data revealed that GRHL2 binding sites were observed at the promoter regions of *E2F2* and *CDCA7L* [24] and this finding was corroborated by ChIP-PCR analysis (Fig. 6e). These results suggested that *E2F2* and *CDCA7L* are directly regulated by GRHL2 and inhibition of cell proliferation mediated by GRHL2 loss may be associated with repression of *E2F2* and/or *CDCA7L*.
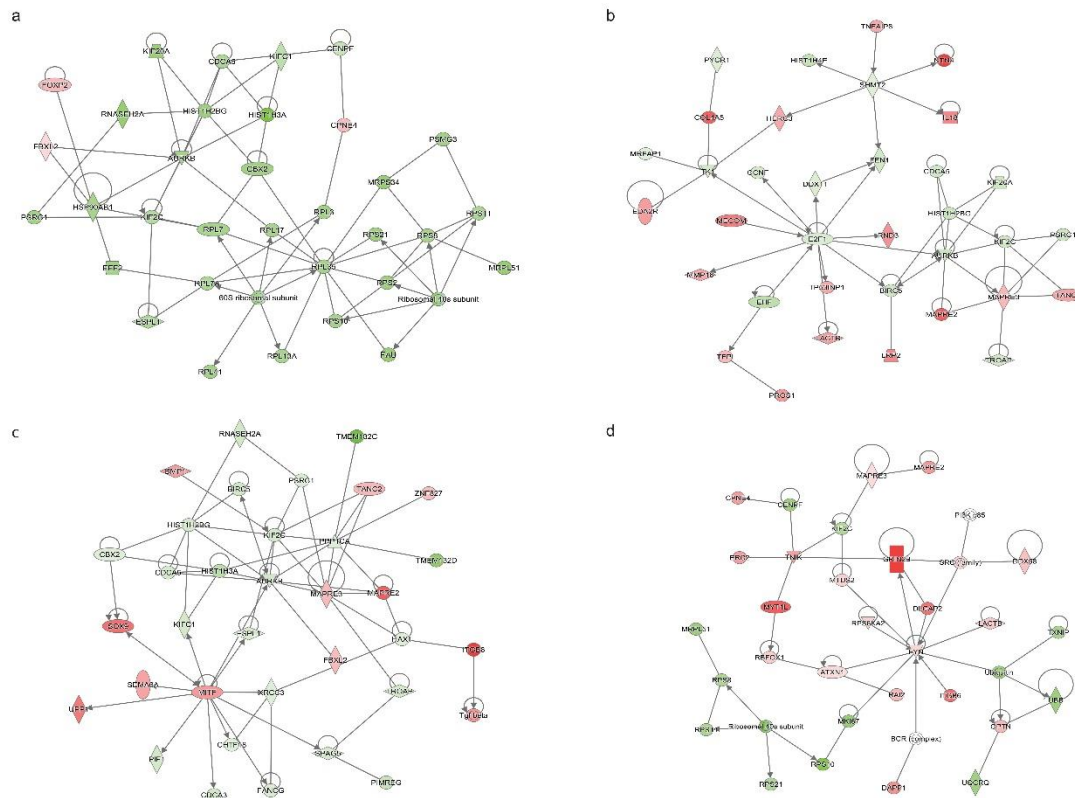


**Fig. 3 Networks with the highest scores according to differentially transcribed genes after GRHL2 loss by IPA. (a-d)** Networks for day 2, 4, 8 and 16 respectively. The

intensity of the node color indicates up- (red) and down regulation (green). Single-way arrows indicate one gene regulating another, two-sided arrows indicate co-regulation, looped arrows indicate self-regulation.
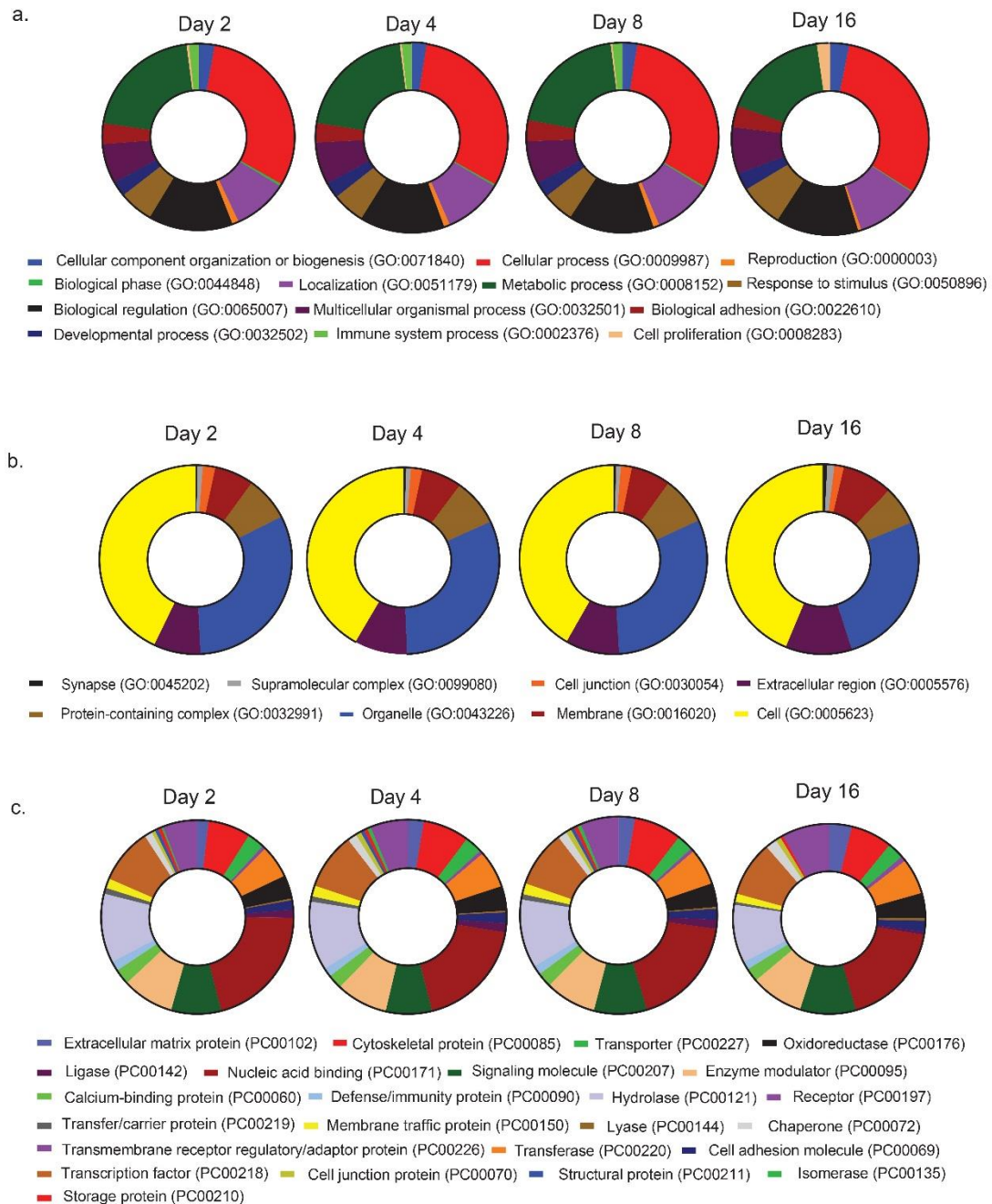


**Fig. 4 PANTHER gene ontology enrichment analysis of differentially transcribed genes after GRHL2 loss.** Enrichment analyses were carried out for biological process **(a)**, cellular component **(b)** and protein classification **(c).**

## RNA synthesis of CDH1 is not altered after GRHL2 loss

*CDH1* encodes E-cadherin, a cell-cell adhesion receptor involved in maintenance of the epithelial phenotype [35]. *CDH1* has been proposed to represent a direct target gene of GRHL2 [8,36,37]. Other studies [2,8,38], and our unpublished results (Wang et al, manuscript under revision) showed that GRHL2 loss gives rise to reduced expression of E-cadherin protein in MCF7 cells [8]. However, our previous ChIP-seq data [24] revealed that GRHL2 binding sites were not observed at the *CDH1* promoter region, consistent with other findings [8,13,39]. Moreover, we did not observe any downregulation of *CDH1* nascent RNA synthesis in the first 16 days after GRHL2 loss (Fig. 8a and b), Together, these findings indicate that the *CDH1* gene is not a direct target for transcriptional regulation by GRHL2. Rather, *CDH1* may be regulated indirectly through other transcriptional regulators [40] or by GRHL2-mediated post-transcriptional modification (e.g., miR200) [7,13,41] at later timepoints.
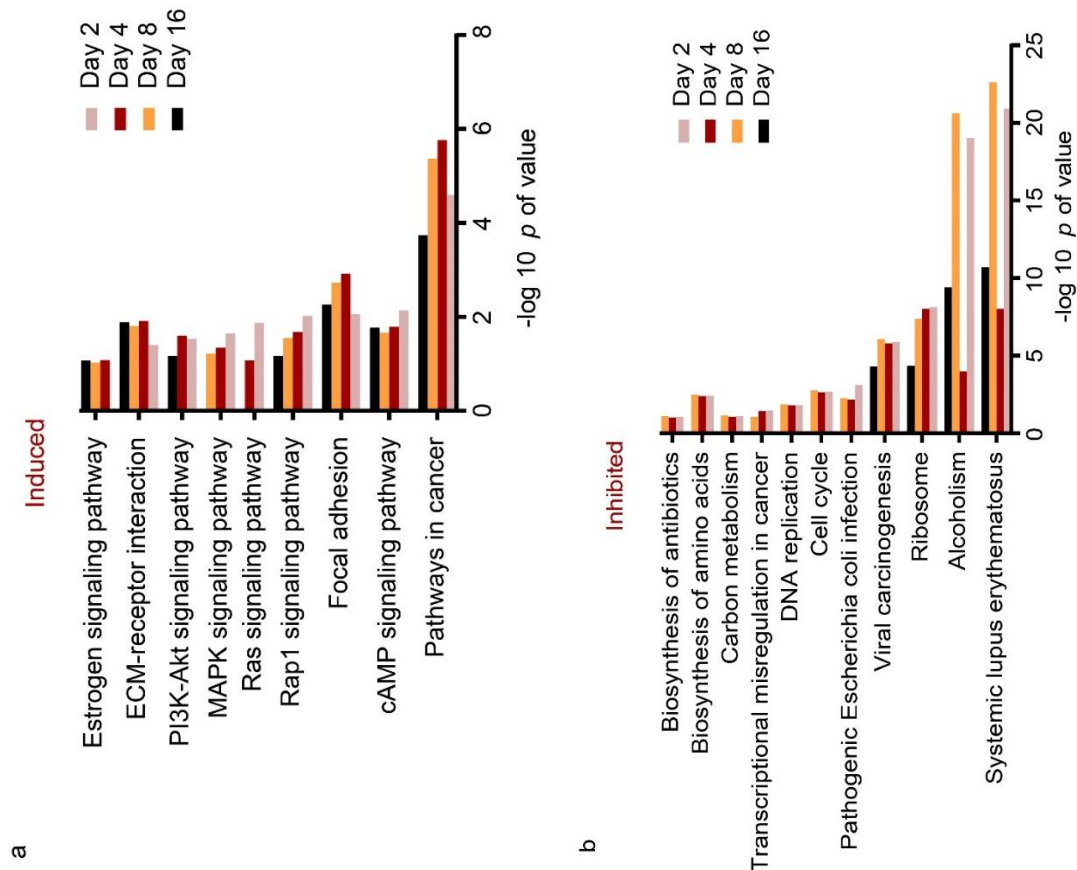
**(Last page) Fig. 5 DAVID analysis of differentially transcribed genes after GRHL2 loss.** Enriched pathways identified by DAVID for induced **(a)** and suppressed **(b)** gene sets at indicated time points.
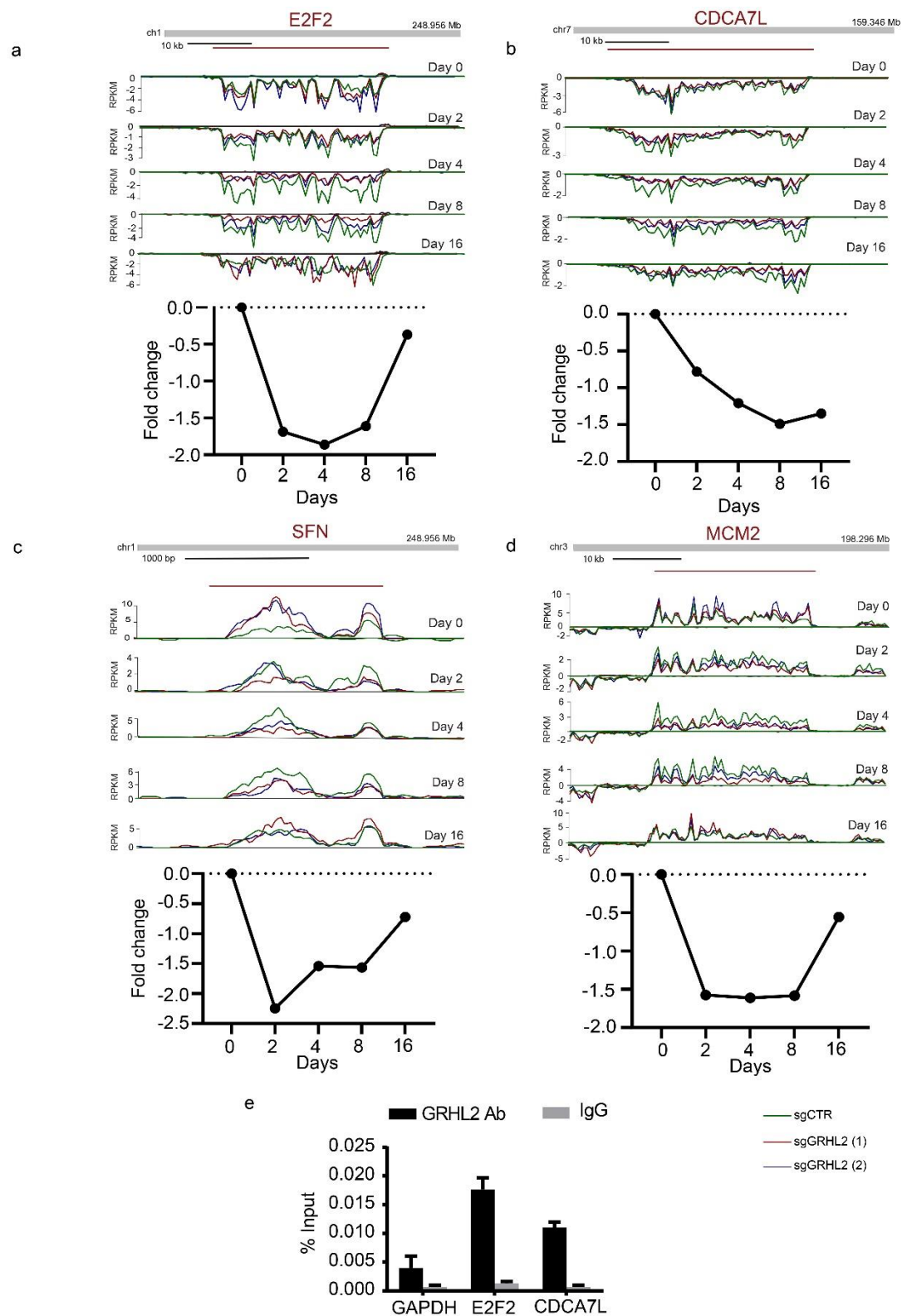
**(Last page) Fig. 6 Downregulation of RNA synthesis for genes involved in cell cycle progression after GRHL2 loss. (a-d) Top:** Bru-seq reads for indicated genes at indicated time point in response to GRHL2 deletion. **Bottom:** Line graphs depicting the log$_2$ AFC of transcription in sgGRHL2 (1) and sgGRHL2 (2) cells. 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-PCR. 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 $p < 0.05$.

# Discussion

The expression level of individual mRNAs is determined by the RNA synthesis and degradation rates. Characterization of global RNA dynamics provides insight into mechanisms of cell signaling [42]. In this study, we examined genome-wide time-resolved responses of RNA synthesis after GRHL2 loss in luminal-like breast cancer cells. We used Bru-seq to capture changes in RNA synthesis [16] in a conditional GRHL2 knockout model. We identified 264 induced and 244 repressed genes in at least one time point following GRHL2 loss. These genes exhibit diverse patterns of RNA synthesis that are divided into sustained induction, sustained repression, induction reset, dynamic and repression reset. Genes with similar patterns of RNA synthesis may be regulated by similar means and the fact that patterns of transcription induction are similar to the patterns of transcription repression, suggests that transcriptional induction and repression may involve similar mechanisms [16].

Bioinformatics analysis identifies several signaling pathways that are enriched at each time point analyzed after GRHL2 deletion (i.e., Granzyme A signaling, remodeling of epithelial adherens junctions, mTOR signaling and DNA methylation, and transcriptional repression signaling). Granzyme A induces caspase-independent apoptosis by dysregulation of mitochondrial metabolism

and generation of reactive oxygen species (ROS) in the mitochondrion [43]. Some repressed genes caused by GRHL2 loss (i.e., *HIST1H1C*, *HIST1H1D*, *HIST1H1E*, *HMGB2*, and *NME1*) are linked to Granzyme A signaling, of which HMGB2 is a positive regulator of proliferation and negative mediator of apoptosis [44,45]. The adherens junctions are specialized structures that encircle epithelial cells and maintain the architectural integrity of epithelial tissues [46,47]. A total of 11 genes, including *HIST1H1C*, *HIST1H1D*, *HIST1H1E*, *HMGB2ACTB*, *ACTG1, ARPC1A, MAPRE2, NME1, TUBA1B, TUBB*, and *TUBB4B*, are identified to be involved in remodeling of epithelial adherens junctions caused by GRHL2 loss. mTOR is a protein kinase that is involved in cell metabolism, proliferation and survival [48]. A cluster of GRHL2 loss responsive genes (*FAU, PLD1, PRKD1, RND3, RPS10, RPS11, RPS2, RPS21, RPS6KA2*, and *RPS8*) are associated with mTOR signaling. The activation of the AKT/mTOR pathway can trigger EMT through upregulation of ZEB1 [49], which has a negative feedback with GRHL2 [7].

Notably, we demonstrate that *CDH1* RNA synthesis is not altered following GRHL2 loss. This is in agreement with our previous report that CDH1 is not identified as a GRHL2 target by ChIP-seq in breast cancer cells [24], demonstrating that E-cadherin downregulation must occur in an indirect manner in our luminal breast cancer model. Others have identified *CDH1* as a direct GRHL2 target in normal epithelia [8] suggesting that the mechanism of E-cadherin regulation significantly differs between non-transformed epithelial cells and cancer cells that retain epithelial characteristics.

IPA network analysis shows that signaling networks exhibit numerous similarities among different time points but the most enriched networks vary over time. PANTHER analysis reveals that biological processes, cellular component, and protein classifications associated with networks regulated by

GRHL2 loss are conserved over time. DAVID analysis shows that the genes whose transcription is attenuated after GRHL2 loss are associated with important functions, including DNA replication, which is consistent with our previous finding that GRHL2 loss leads to a G0/1 arrest (Wang et al, manuscript under revision). A group of repressed genes are enriched for cell cycle and DNA replication including *E2F1, E2F2, MCM7, CDC20, ESPL1, MCM2, PTTG1, SFN, RNASEH2A* and *FEN1*. *E2F2* is a member of E2F transcription factor family that has a crucial role in the control of cell cycle and DNA replication [50].Our ChIP-PCR validates the presence of GRHL2 binding sites in the *E2F2* promoter region. Additionally, previous studies show that cell division cycle associated 7 like (*CDCA7L*) is a positive regulator of cell proliferation in prostate cancer and glioma [51,52]. The existence of GRHL2 binding sites in the *CDCA7L* promoter region is also verified by ChIP-PCR. These findings suggest that GRHL2 may regulate DNA replication and cell cycle by multiple mechanisms, including direct transcriptional modulation of *E2F2* and *CDCA7L*. Moreover, we establish *EHF* as a direct GRHL2 target gene.

Taken together, in this study we identify GRHL2-regulated genes, we find five main patterns by which RNA synthesis is altered in response to depletion of GRHL2, and we provide new insights into the dynamics of GRHL2-mediated signaling networks. Additionally, our findings reveal how regulation of epithelial genes such as *CDH1* can be strikingly different in normal and cancer cells involving direct GRHL2-binding or indirect mechanisms.

## Acknowledgements

# Reference

1       Ma, L., Yan, H., Zhao, H. & Sun, J. Grainyhead-like 2 in development and cancer. *Tumour Biol* **39**, 1010428317698375, doi:10.1177/1010428317698375 (2017).

2       Xiang, X. *et al.* Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One* **7**, e50781, doi:10.1371/journal.pone.0050781 (2012).

3       Pyrgaki, C., Liu, A. & Niswander, L. Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. *Developmental biology* **353**, 38-49, doi:10.1016/j.ydbio.2011.02.027 (2011).

4       Werner, S. *et al.* Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *The Journal of biological chemistry* **288**, 22993-23008, doi:10.1074/jbc.M113.456293 (2013).

5       Reese, R. M., Harrison, M. M. & Alarid, E. T. Grainyhead-like Protein 2: The Emerging Role in Hormone-Dependent Cancers and Epigenetics. *Endocrinology* **160**, 1275-1288, doi:10.1210/en.2019-00213 (2019).

6       Cieply, B. *et al.* Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res* **72**, 2440-2453, doi:10.1158/0008-5472.CAN-11-4038 (2012).

7       Cieply, B., Farris, J., Denvir, J., Ford, H. L. & Frisch, S. M. Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res* **73**, 6299-6309, doi:10.1158/0008-5472.CAN-12-4082 (2013).

8       Werth, M. *et al.* The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835-3845, doi:10.1242/dev.055483 (2010).

9       Dompe, N. *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* **108**, E943-951, doi:10.1073/pnas.1100132108 (2011).

10      Pan, X. *et al.* GRHL2 suppresses tumor metastasis via regulation of transcriptional activity of RhoG in non-small cell lung cancer. *Am J Transl Res* **9**, 4217-4226 (2017).

11      Faddaoui, A. *et al.* Suppression of the grainyhead transcription factor 2 gene (GRHL2) inhibits the proliferation, migration, invasion and mediates cell cycle arrest of ovarian cancer cells. *Cell Cycle* **16**, 693-706, doi:10.1080/15384101.2017.1295181 (2017).

12      Wang, G., Pan, J., Zhang, L. & Wang, C. Overexpression of grainyhead-like transcription factor 2 is associated with poor prognosis in human pancreatic carcinoma. *Oncology letters* **17**, 1491-1496, doi:10.3892/ol.2018.9741 (2019).

13      Chung, V. Y. *et al.* GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943, doi:10.1038/srep19943 (2016).

14      Chung, V. Y. *et al.* The role of GRHL2 and epigenetic remodeling in epithelial-mesenchymal plasticity in ovarian cancer cells. *Commun Biol* **2**, 272, doi:10.1038/s42003-019-0506-3 (2019).

15      Paulsen, M. T. *et al.* Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67**, 45-54, doi:10.1016/j.ymeth.2013.08.015 (2014).

16      Kirkconnell, K. S., Paulsen, M. T., Magnuson, B., Bedi, K. & Ljungman, M. Capturing the dynamic nascent transcriptome during acute cellular responses: The serum response. *Biol Open* **5**, 837-847, doi:10.1242/bio.019323 (2016).

17      Ohtsu, M. *et al.* Novel DNA microarray system for analysis of nascent mRNAs. *DNA Res* **15**, 241-251, doi:10.1093/dnares/dsn015 (2008).

18      Haider, S. R., Juan, G., Traganos, F. & Darzynkiewicz, Z. Immunoseparation and immunodetection of nucleic acids labeled with halogenated nucleotides. *Experimental cell research* **234**, 498-506, doi:10.1006/excr.1997.3644 (1997).

19      Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848, doi:10.1126/science.1162228 (2008).

20      Paulsen, M. T. *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* **110**, 2240-2245, doi:10.1073/pnas.1219192110 (2013).

21      Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

22      Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

23      Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research* **47**, D419-D426, doi:10.1093/nar/gky1038 (2019).

24      Wang, Z., Wu, H., Daxinger, L. & Danen, E. H. J. Genome-wide identification of binding sites of GRHL2 in luminal-like and basal A subtypes of breast cancer. *bioRxiv*, 2020.2002.2013.946947, doi:10.1101/2020.02.13.946947 (2020).

25      Haring, M. *et al.* Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**, 11, doi:10.1186/1746-4811-3-11 (2007).

26      Marinkovich, M. P. Tumour microenvironment: laminin 332 in squamous-cell carcinoma. *Nat Rev Cancer* **7**, 370-380, doi:10.1038/nrc2089 (2007).

27      Popovic, D., Vucic, D. & Dikic, I. Ubiquitination in disease pathogenesis and treatment. *Nature medicine* **20**, 1242-1253, doi:10.1038/nm.3739 (2014).

28      Stroebel, D., Casado, M. & Paoletti, P. Triheteromeric NMDA receptors: from structure to synaptic physiology. *Current opinion in physiology* **2**, 1-12, doi:10.1016/j.cophys.2017.12.004 (2018).

29      Trimarchi, J. M. & Lees, J. A. Sibling rivalry in the E2F family. *Nature reviews. Molecular cell biology* **3**, 11-20, doi:10.1038/nrm714 (2002).

30      Quan, Y. *et al.* Downregulation of GRHL2 inhibits the proliferation of colorectal cancer cells by targeting ZEB1. *Cancer Biol Ther* **15**, 878-887, doi:10.4161/cbt.28877 (2014).

31  Tian, Y. *et al.* CDCA7L promotes hepatocellular carcinoma progression by regulating the cell cycle. *Int J Oncol* **43**, 2082-2090, doi:10.3892/ijo.2013.2142 (2013).

32  Miao, B. *et al.* The transcription factor FLI1 promotes cancer progression by affecting cell cycle regulation. *International journal of cancer*, doi:10.1002/ijc.32831 (2019).

33  Cheng, A. C., Shen, C. J., Hung, C. M. & Hsu, Y. C. Sulforaphane Decrease of SERTAD1 Expression Triggers G1/S Arrest in Breast Cancer Cells. *J Med Food* **22**, 444-450, doi:10.1089/jmf.2018.4195 (2019).

34  Cheung, C. H. Y. *et al.* MCM2-regulated functional networks in lung cancer by multi-dimensional proteomic approach. *Sci Rep* **7**, 13302, doi:10.1038/s41598-017-13440-x (2017).

35  Mendonsa, A. M., Na, T. Y. & Gumbiner, B. M. E-cadherin in contact inhibition and cancer. *Oncogene* **37**, 4769-4780, doi:10.1038/s41388-018-0304-2 (2018).

36  Aue, A. *et al.* A Grainyhead-Like 2/Ovo-Like 2 Pathway Regulates Renal Epithelial Barrier Function and Lumen Expansion. *J Am Soc Nephrol* **26**, 2704-2715, doi:10.1681/ASN.2014080759 (2015).

37  Jolly, M. K. *et al.* E-Cadherin Represses Anchorage-Independent Growth in Sarcomas through Both Signaling and Mechanical Mechanisms. *Mol Cancer Res* **17**, 1391-1402, doi:10.1158/1541-7786.MCR-18-0763 (2019).

38  Paltoglou, S. *et al.* Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer. *Cancer Res* **77**, 3417-3430, doi:10.1158/0008-5472.CAN-16-1616 (2017).

39  Varma, S. *et al.* 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* **287**, 37282-37295, doi:10.1074/jbc.M112.408401 (2012).

40  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* **1868**, 584-591, doi:10.1016/j.bbcan.2017.06.006 (2017).

41  Park, S. M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & development* **22**, 894-907, doi:10.1101/gad.1640608 (2008).

42  Yamada, T. *et al.* 5'-Bromouridine IP Chase (BRIC)-Seq to Determine RNA Half-Lives. *Methods Mol Biol* **1720**, 1-13, doi:10.1007/978-1-4939-7540-2_1 (2018).

43  Lieberman, J. Granzyme A activates another way to die. *Immunol Rev* **235**, 93-104, doi:10.1111/j.0105-2896.2010.00902.x (2010).

44  Fu, D. *et al.* HMGB2 is associated with malignancy and regulates Warburg effect by targeting LDHB and FBP1 in breast cancer. *Cell Commun Signal* **16**, 8, doi:10.1186/s12964-018-0219-0 (2018).

45  Kwon, J. H. *et al.* Overexpression of high-mobility group box 2 is associated with tumor aggressiveness and prognosis of hepatocellular carcinoma. *Clin Cancer Res* **16**, 5511-5521, doi:10.1158/1078-0432.CCR-10-0825 (2010).

46  D'Souza-Schorey, C. Disassembling adherens junctions: breaking up is hard to do. *Trends Cell Biol* **15**, 19-26, doi:10.1016/j.tcb.2004.11.002 (2005).

47    Rudini, N. & Dejana, E. Adherens junctions. *Curr Biol* **18**, R1080-1082, doi:10.1016/j.cub.2008.09.018 (2008).

48    Laplante, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274-293, doi:10.1016/j.cell.2012.03.017 (2012).

49    Lau, M. T., So, W. K. & Leung, P. C. Fibroblast growth factor 2 induces E-cadherin down-regulation via PI3K/Akt/mTOR and MAPK/ERK signaling in ovarian cancer cells. *PLoS One* **8**, e59083, doi:10.1371/journal.pone.0059083 (2013).

50    Lafta, I. J. E2F6 is essential for cell viability in breast cancer cells during replication stress. *Turk J Biol* **43**, 293-304, doi:10.3906/biy-1905-6 (2019).

51    Lin, T. P. *et al.* R1 Regulates Prostate Tumor Growth and Progression By Transcriptional Suppression of the E3 Ligase HUWE1 to Stabilize c-Myc. *Mol Cancer Res* **16**, 1940-1951, doi:10.1158/1541-7786.MCR-16-0346 (2018).

52    Ji, Q. K. *et al.* CDCA7L promotes glioma proliferation by targeting CCND1 and predicts an unfavorable prognosis. *Mol Med Rep* **20**, 1149-1156, doi:10.3892/mmr.2019.10349 (2019).

# Supplemental data

**Table S1. List of genes whose transcription is altered in response to GRHL2 deletion and their classification into subgroups according to their dynamic pattern of regulation.**

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| 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 | Repression reset |
| AC013652.1 | 2.43 | 3.59 | 4.10 | 1.57 | Induction reset |
| AC019209.3 | 1.43 | 3.61 | 4.91 | 4.29 | Sustained induction |
| 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 | |
| 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 |
| 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 |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|------|--------|--------|--------|---------|---------|
| 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 | |
| ANKRD1 | 1.39 | 4.68 | 5.06 | 1.90 | Induction reset |
| 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 |
| ASB9 | 10.25 | 13.53 | 11.41 | 4.43 | Sustained induction |
| ATP10D | 3.42 | 10.07 | 11.67 | 5.55 | Sustained induction |
| ATP5O | 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 | Induction reset |
| 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 |
| CADM1 | 1.89 | 2.93 | 3.51 | 1.49 | Induction reset |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| CADPS | 53.78 | 0.06 | 0.03 | 51.71 | Dynamic |
| CAMK1D | 1.92 | 3.14 | 3.93 | 1.98 | Induction reset |
| 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 |
| CDCA7L | 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 |
| 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 | |
| CLIP4 | 1.91 | 3.35 | 4.85 | 3.69 | |
| CNTN4 | 22.17 | 0.05 | 5.10 | 26.44 | |
| COL4A5 | 3.79 | 7.76 | 10.60 | 5.61 | Sustained induction |
| COLQ | 2.50 | 2.83 | 3.71 | 1.81 | Induction reset |
| CORO2A | 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 |
| DAPP1 | 4.40 | 9.94 | 23.48 | 11.76 | Sustained induction |
| DDX11 | 0.37 | 0.37 | 0.41 | 0.67 | Repression reset |
| DDX12P | 0.33 | 0.27 | 0.30 | 0.58 | Repression reset |
| DDX41 | 0.33 | 0.48 | 0.52 | 0.77 | |
| DDX58 | 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 | Induction reset |
| DOCK4 | 2.36 | 3.98 | 5.25 | 1.92 | Induction reset |
| DOCK8 | 2.24 | 3.15 | 3.16 | 2.68 | Sustained induction |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| DOK5 | 24.10 | 0.01 | 0.03 | 18.16 | Dynamic |
| DUSP10 | 2.30 | 2.61 | 4.07 | 2.72 | Sustained induction |
| E2F1 | 0.35 | 0.33 | 0.38 | 0.80 | Repression reset |
| E2F2 | 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 |
| EEF1B2 | 0.44 | 0.47 | 0.43 | 0.39 | Sustained repression |
| 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 |
| GRIN2B | 59.19 | 0.06 | 0.05 | 86.59 | Dynamic |
| GRK5 | 2.15 | 3.92 | 3.74 | 2.42 | Sustained induction |
| GULP1 | 2.50 | 4.50 | 6.29 | 3.19 | Sustained induction |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| H2AFZ | 0.17 | 0.20 | 0.24 | 0.36 | Sustained repression |
| HAX1 | 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 | Induction reset |
| 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 |
| 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 |
| 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 |
| HIST1H4B | 0.20 | 0.11 | 0.17 | 0.52 | Repression reset |
| HIST1H4D | 0.19 | 0.11 | 0.17 | 0.40 | Sustained repression |
| HIST1H4E | 0.19 | 0.17 | 0.25 | 0.54 | Repression reset |
| HIST1H4F | 0.21 | 0.11 | 0.13 | 0.41 | Sustained repression |
| HIST1H4H | 0.36 | 0.26 | 0.38 | 0.58 | Repression reset |
| HIST2H3D | 0.19 | 0.17 | 0.16 | 2.15 | |
| 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 |
| HMGB2 | 0.25 | 0.23 | 0.29 | 0.46 | Sustained repression |
| HMMR | 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 |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|------|--------|--------|--------|---------|---------|
| 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 | Induction reset |
| 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 | |
| KIAA0513 | 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 |
| KIF5C | 1.26 | 2.82 | 3.91 | 3.14 | |
| KIFC1 | 0.46 | 0.29 | 0.29 | 0.55 | Repression reset |
| LACTB | 1.58 | 3.33 | 3.83 | 3.25 | |
| LAD1 | 0.32 | 0.41 | 0.56 | 1.13 | |
| LAMA3 | 2.00 | 3.34 | 3.66 | 2.74 | |
| LAMB3 | 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 | |
| LIMCH1 | 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 | |
| LINC00885 | 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 |
| MAPK10 | 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 | |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| 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 |
| MIR5087 | 0.19 | 0.24 | 0.33 | 0.59 | Repression reset |
| 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 |
| MTUS2 | 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 | Induction reset |
| 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 |
| NPY1R | 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 | |
| P2RY2 | 0.22 | 0.35 | 0.54 | 1.53 | |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| PACSIN3 | 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 |
| PGLYRP2 | 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 |
| PPP1CA | 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 |
| PRICKLE2-AS3 | 4.11 | 9.01 | 9.17 | 4.29 | Sustained induction |
| PRKD1 | 3.24 | 5.01 | 7.05 | 2.39 | Sustained induction |
| PROS1 | 2.99 | 3.62 | 4.27 | 1.93 | Induction reset |
| PRSS23 | 2.39 | 4.01 | 5.50 | 5.31 | Sustained induction |
| PSG5 | 2.07 | 7.09 | 9.24 | 7.10 | Sustained induction |
| PSMB6 | 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 |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| 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 |
| RND3 | 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 |
| RNU5E-4P | 0.19 | 0.16 | 0.31 | 0.59 | Repression reset |
| 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 |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| 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 |
| SH3TC2 | 0.16 | 0.22 | 0.26 | 0.26 | Sustained repression |
| 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 | |
| SLC16A3 | 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 | |
| SLIT3 | 4.89 | 0.03 | 0.63 | 3.96 | |
| SNORD3A | 0.10 | 0.13 | 0.19 | 0.27 | Sustained repression |
| SNORD3B-1 | 0.19 | 0.23 | 0.32 | 0.67 | 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 |
| SPATA18 | 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 | |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|------|--------|--------|--------|---------|---------|
| SUN2 | 0.35 | 0.33 | 0.43 | 0.67 | Repression reset |
| SYNPO | 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 |
| TACSTD2 | 0.32 | 0.30 | 0.43 | 0.64 | Repression reset |
| TANC2 | 2.06 | 3.16 | 3.40 | 1.56 | Induction reset |
| TENM2 | 25.34 | 0.03 | 6.08 | 23.79 | |
| TFPI | 1.65 | 2.39 | 3.30 | 2.04 | |
| TGFB2 | 2.27 | 6.07 | 5.70 | 1.71 | Induction reset |
| TGFBI | 2.92 | 6.30 | 9.58 | 4.64 | Sustained induction |
| TGFBR2 | 3.05 | 6.81 | 8.02 | 2.99 | Sustained induction |
| THAP11 | 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 |
| TMEM140 | 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 |
| TMPRSS13 | 0.25 | 0.23 | 0.39 | 0.85 | Repression reset |
| 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 | Sustained induction |
| TPI1 | 0.23 | 0.34 | 0.41 | 0.52 | 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 |
| TYRO3 | 0.20 | 0.28 | 0.33 | 0.50 | Repression reset |
| 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 |
| UHRF1 | 0.33 | 0.27 | 0.32 | 0.68 | Repression reset |

| Gene | D2 AFC | D4 AFC | D8 AFC | D16 AFC | Cluster |
|---|---|---|---|---|---|
| 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 | |
| VMP1 | 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 |
| WIPI1 | 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 |
| ZBTB20 | 3.17 | 2.79 | 5.25 | 2.21 | Sustained induction |
| ZMAT4 | 1.72 | 5.72 | 14.92 | 5.98 | Sustained induction |
| 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 | Induction reset |
| ZNF827 | 2.41 | 3.25 | 3.03 | 1.78 | Induction reset |
| ZWINT | 0.29 | 0.26 | 0.33 | 0.64 | Repression reset |

# Chapter 5

## Identification of direct GRHL2 target genes in luminal-like breast cancer through integration of ChIP-seq and Bru-seq

Zi Wang[1], Yao-Jun Chen[1], Mats Ljungman[2], Erik HJ Danen[1,3]

[1]Leiden Academic Center for Drug Research, Leiden University, Leiden, The Netherlands; [2]University of Michigan, Ann Arbor, MI 48109; [3]correspondence to Erik HJ Danen, e.danen@lacdr.leidenuniv.nl

# Abstract

The transcription factor GRHL2 has been reported to induce and repress gene expression, which is cell context-dependent. While several studies have addressed expression and function of GRHL2 in breast cancer with different conclusions, the profile of GRHL2 target genes in breast cancer has not been characterized. In the present study, ChIP-seq analysis of GRHL2-binding genes in MCF7 cells was integrated with Bru-seq analysis of genes showing transcriptional responses to conditional CRIPR-Cas9 knockout of GRHL2 in MCF7 cells. This identified 48 direct target genes of GRHL2 in MCF7 cells. Signaling pathways and networks associated with these direct GRHL2 target genes were explored using IPA. Notably, in line with our previous report that the *CDH1* promotor lacks GRHL2-binding sites, RNA synthesis of *CDH1*, encoding the epithelial adhesion receptor E-cadherin, was not altered following GRHL2 deletion, demonstrating *CDH1* is not a direct target gene of GRHL2. Instead, the epithelial-specific transcription factor EHF/ESE3, a transcription factor that, like GRHL2, suppresses EMT, was identified as a direct target gene of GRHL2. *EHF* was downregulated at all time points after GRHL2 deletion and, like *GRHL2*, *EHF* was specifically absent in basal B-like breast cancer in a pan-subtype human breast cancer cell line panel. EHF has been implicated in tumor initiating properties. However, its overexpression failed to rescue proliferation in GRHL2-depleted breast cancer cells. Collectively, this study identifies direct target genes of GRHL2 and their related signaling pathways and sheds light on the epithelial factors GRHL2 and EHF in luminal-like breast cancer MCF7 cells.

# Introduction

The transcription factor GRHL2 has been reported to activate and repress gene expression [1,2]. Identification of GRHL2 direct target genes is significant for

understanding GRHL2 biological functions and signaling pathways regulated by GRHL2. A negative feedback loop exists between ZEB1 and GRHL2, in which GRHL2 acts as an inhibitory regulator of EMT, inhibiting ZEB1 expression directly [2]. Additionally, GRHL2 is involved in differentiation of epithelial cells, morphogenesis of epithelial tubes and maintenance of epithelial phenotype through activating expression of epithelial-specific genes such as *CDH1* and *CLDN4* [3]. EHF derives from the ETS transcription factor family characterized by epithelial-specific expression [4]. Multiple lines of evidence demonstrate a significant role of EHF in the regulation of cell proliferation and differentiation [4,5].

Integrating RNA-seq and ChIP-seq data is a typical method to identify target genes. However, target genes identified by RNA-seq include direct and indirect target genes. By contrast, Bru-seq analyzes changes in nascent RNA, is not affected by post-transcriptional regulation, and hence monitors direct changes in transcription [6,7]. Here, we integrated ChIP-seq analysis of GRHL2-binding genes in MCF7 cells with Bru-seq analysis of genes showing transcriptional responses to conditional CRISPR-Cas9 knockout of GRHL2 in MCF7 cells. Direct target genes of GRHL2 and their related signaling pathways were identified and the interaction between *GRHL2, EHF*, and *CDH1* in luminal-like breast cancer (MCF7 cells) was explored.

## Materials and methods

### Cell culture

MCF7 human breast cancer cells were obtained from the American Type Culture Collection. Cells were cultured in RPMI1640 medium with 10% fetal bovine serum, 25 U/mL penicillin and 25 µg/mL streptomycin at 37°C and 5%

CO$_2$. Generation of MCF7 conditional GRHL2 knockout (KO) cells was previously described (see Chapter 4).
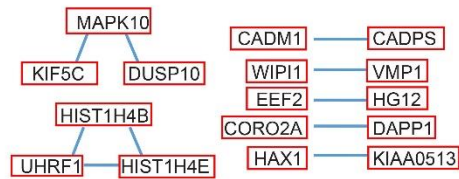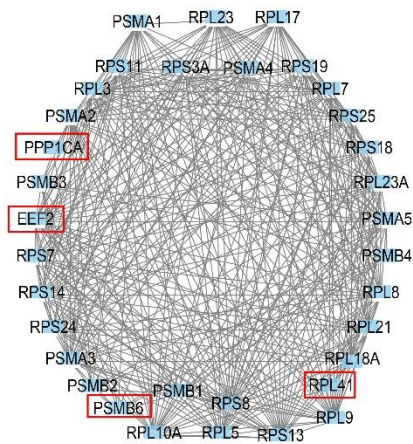


**Fig. 1 Protein-protein interaction networks among GRHL2 direct target genes.** Based on identification of GRHL2 direct target genes through integration of ChIP-seq and Bru-seq data, **(a)** direct and **(b)** indirect protein-protein interaction networks among GRHL2 direct target genes are predicted. Red rectangles indicate direct target genes of GRHL2.

### ChIP-PCR

Chromatin preparation for chromatin immunoprecipitation-sequencing (ChIP-seq) has been previously described [8]. For ChIP-PCR, chromatin fragments were immunoprecipitated with control IgG or anti-GRHL2 antibodies (Sigma; HPA004820). Precipitates were eluted 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. The following primers were used for ChIP-PCR: *EHF* forward: ctgaaaagaacagtcaccacca, *EHF* reverse: tggccaactcacacgttagt, control (an intergenic region upstream of the *GAPDH* locus) forward: atgggtgccactggggatct, control reverse: tgccaaagcctaggggaaga. ChIP-PCR data were analyzed using the $2^{-\Delta\Delta Ct}$ method [9].

**Bru-seq**

Bru-seq and associated bioinformatics analysis was described previously (see Chapter 4). In short, at different timepoints after doxycycline-induced deletion of GRHL2, MCF7 conditional KO cells were incubated with Bru, cells were lysed, and Bru-labelled nascent RNA was isolated using an anti-BrdU antibodies. Subsequently, cDNA libraries were generated, sequenced, and reads were mapped to Genome Reference Consortium human genome (build 37). To identify GRHL2-regulated genes, reads per kilobase per million mapped reads for each gene in the doxycycline-treated samples were compared to the untreated samples, genes with $p<0.05$ and FC>2 or FC<0.5 were filtered, and genes responding to Cas9 induction in the context of sgGRHL2 but not in the context of sgCTR were selected.

**Bioinformatics analysis**

Canonical pathways and networks analysis were performed with the Ingenuity Pathways Analysis (IPA) software (Ingenuity Systems, USA). The STRING database [10] was used to predict direct protein-protein interactions (PPI), which were visualized by Cytoscape v3.7.2 [11].

**Table 1.** Top 20 canonical pathways identified by GRHL2 direct target genes using IPA.

| Canonical Pathways | -log(p-value) | Ratio |
|---|---|---|
| Germ Cell-Sertoli Cell Junction Signaling | 2.26 | 0.0181 |
| Diphthamide Biosynthesis | 2.19 | 0.333 |
| Production of Nitric Oxide and Reactive Oxygen Species in Macrophages | 2.13 | 0.0162 |
| Activation of IRF by Cytosolic Pattern Recognition Receptors | 2.11 | 0.0328 |
| SAPK/JNK Signaling | 1.7 | 0.0198 |
| Sumoylation Pathway | 1.7 | 0.0198 |
| CDK5 Signaling | 1.65 | 0.0187 |
| Pancreatic Adenocarcinoma Signaling | 1.64 | 0.0185 |
| Cholecystokinin/Gastrin-mediated Signaling | 1.58 | 0.0171 |
| D-myo-inositol (1,4,5,6)-Tetrakisphosphate Biosynthesis | 1.46 | 0.0148 |
| D-myo-inositol (3,4,5,6)-tetrakisphosphate Biosynthesis | 1.46 | 0.0148 |
| Granzyme A Signaling | 1.44 | 0.0588 |
| Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses | 1.41 | 0.0139 |
| 3-phosphoinositide Degradation | 1.39 | 0.0134 |
| D-myo-inositol-5-phosphate Metabolism | 1.38 | 0.0133 |
| HMGB1 Signaling | 1.34 | 0.0127 |
| 3-phosphoinositide Biosynthesis | 1.33 | 0.0126 |
| Tec Kinase Signaling | 1.33 | 0.0125 |

## Sulforhodamine B (SRB) assay

Cell proliferation rate was measured by SRB assay. Cells were seeded into 96-well plates. At the indicated time points, cells were fixed with 50% Trichloroacetic acid (TCA, Sigma-Aldrich) for 1 hour at 4 °C and plates were washed with demineralized water 4 times and air-dried at room temperature (RT). Subsequently, plates were incubated with 0.4% SRB (60 µl/well) for at least 2 hours at RT. The plates were washed five times with 1% acetic acid and air-dried. Plates were incubated with 10 mM Tris (150 µl/well) for 30 minutes at RT with gentle shaking. The absorbance value was measured by a Fluostar OPTIMA plate-reader.

## Realtime quantitative PCR (RT-qPCR)

Total RNA was isolated using RNEasy Plus Mini Kit (Qiagen). 500 ng RNA was reverse-transcribed into cDNA using the RevertAid H Minus First Strand cDNA

Synthesis Kit (Thermo Fisher Scientific). The cDNA was mixed with SYBR green master mix (Fisher Scientific) for qPCR. RT-qPCR data were collected and analyzed using $2^{-\Delta\Delta Ct}$ method. The following primers were used: *GAPDH* forward: ccatggggaaggtgaaggtc, *GAPDH* reverse: agttaaaagcagccctggtga. *EHF* forward: ctgccctgagtggagattgg, *EHF* reverse: tgcccttgccttcacagaaa.

**Colony formation assays**

Cell survival was measured by colony formation assay. 450 cells were seeded into a well of 6-well plate after 4 days of doxycycline treatment. After 7 days, cells were fixed with 4% formaldehyde and stained with Giemsa. Images were analyzed by Image J (ColonyArea package).

**Table 2.** Networks based on direct GRHL2 genes predicted by IPA.

| Top Diseases and Functions | Score | Focus Molecules |
|---|---|---|
| Cancer, Organismal Injury and Abnormalities, Reproductive System Disease | 24 | 13 |
| Cell Cycle, Cellular Development, Cellular Growth and Proliferation | 22 | 12 |
| Cell Death and Survival, Cellular Development, Cellular Movement | 9 | 6 |
| Cellular Assembly and Organization, Embryonic Development, Organismal Development | 2 | 1 |
| Cell Cycle, Cellular Development, Embryonic Development | 2 | 1 |
| Drug Metabolism, Molecular Transport, Organismal Functions | 2 | 1 |
| Cell Morphology, Cellular Movement, Immune Cell Trafficking | 2 | 1 |
| Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry | 2 | 1 |

# Results

**Candidate direct GRHL2 target genes whose transcription is altered in response to GRHL2 depletion**

In order to identify direct target genes of GRHL2 in luminal breast cancer (MCF7 cells), integration of ChIP-seq and Bru-seq data was performed. 48 differentially expressed genes identified by Bru-seq were confirmed as direct GRHL2 target genes by ChIP-seq **(Table S1)**, including *EHF/ESE3, E2F2, CDCA7L* and *FOXP2*. Protein-protein interaction (PPI) networks among the direct GRHL2 target genes were constructed using STRING database [10] **(Fig. 1).**

**Predicted signaling networks regulated by GRHL2**

IPA analysis was performed to predict signaling pathways associated with the identified GRHL2 direct target genes. The signaling pathways were ranked according to *p* value. **Table 1** demonstrates the top 20 signaling pathways, including SAPK/JNK signaling and CDK5 signaling that are involved in cell proliferation [12,13]. Germ cell-Sertoli cell junction signaling was identified as the top signaling pathway predicted by genes regulated by GRHL2 loss. Notably, this same pathway was also predicted by the Bru-seq data alone at 2-8 days post deletion of GRHL2 **(see Chapter 4 Table 1)**.

Subsequently, networks related to multiple biological functions and diseases associated with the identified GRHL2 direct target genes were predicted by IPA and then ranked based on the score. The top networks were enriched in cancer, organismal injury and abnormalities, reproductive system disease **(Table 2)**, of which top 3 networks are shown in **Fig. 2.** The main functions of GRHL2 direct target genes were associated with cell cycle, growth, and proliferation **(Table 2)**.

**RNA synthesis of CDH1 is not altered after GRHL2 loss**

*CDH1* encodes E-cadherin, a cell-cell adhesion receptor involved in maintenance of the epithelial phenotype [14]. *CDH1* has been proposed to represent a direct target gene of GRHL2 [3,15,16]. Other studies [3,17,18], and our unpublished results (Wang et al, manuscript under revision) showed that GRHL2 loss gives rise to reduced expression of E-cadherin protein in MCF7 cells [3]. However, our ChIP-seq data revealed that GRHL2 binding sites were not observed in the *CDH1* promoter region, consistent with other findings [3,19,20]. Moreover, we did not observe any downregulation of *CDH1* nascent RNA synthesis in the first 16 days after GRHL2 loss **(Fig. 3a and c)**, Together, these findings indicate that the *CDH1* gene is not a direct target for transcriptional
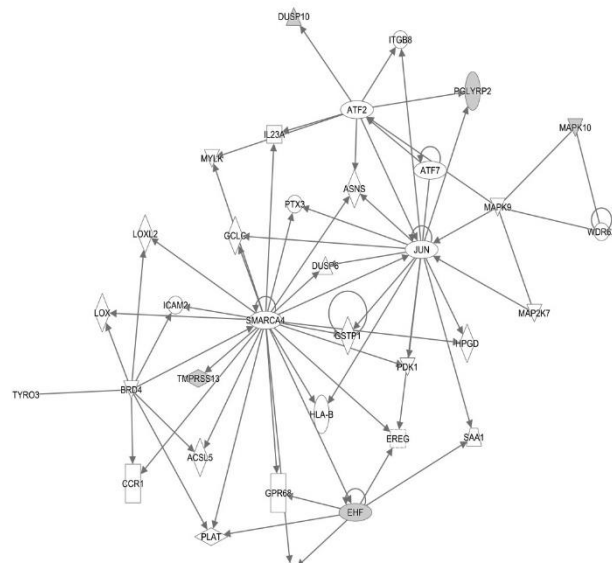
a

b

c

117
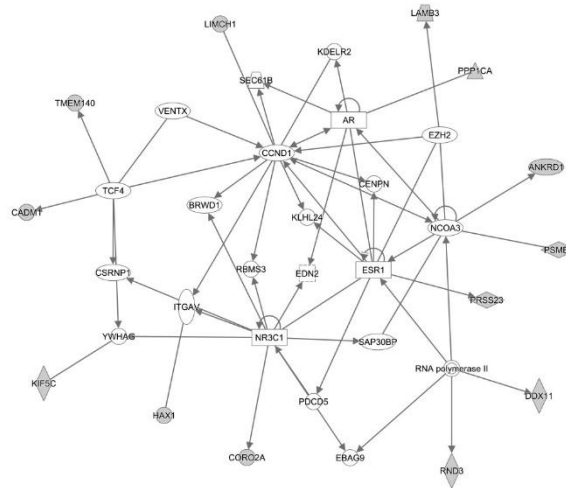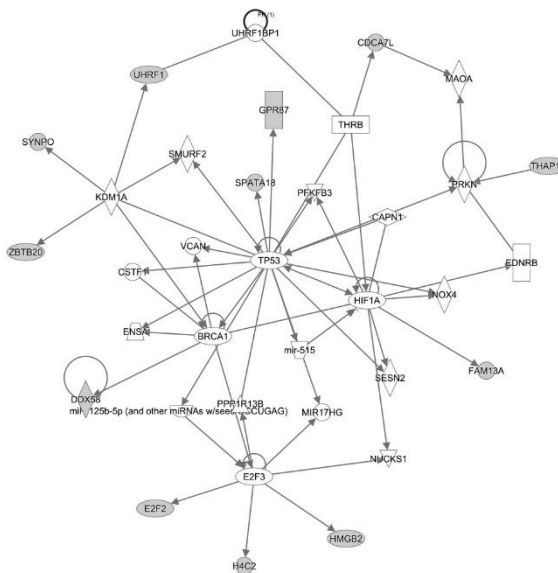
**(Last page) Fig. 2 Predicted networks based on direct GRHL2 target genes**. **(a-c)** Top 3 networks with the highest scores predicted by IPA based on the 48 direct GRHL2 target genes. Single-way arrows indicate one gene regulating another, two-sided arrows indicate co-regulation, looped arrows indicate self-regulation.

regulation by GRHL2. Rather, *CDH1* RNA levels may be regulated indirectly through other transcriptional regulators [21] or by GRHL2-mediated post-transcriptional modification (e.g., miR200) [2,19,22] at later timepoints.

## EHF is co-regulated with GRHL2

*EHF* was identified as a direct target of GRHL2 that was rapidly and continuously attenuated following GRHL2 loss **(Table S1; Fig. 3b and d)**. ChIP-PCR confirmed the interaction between GRHL2 and the promoter region of the *EHF* gene **(Fig. 3e)**. EHF is a member of the ETS transcription factor subfamily characterized by epithelial-specific expression [4]. Epithelial markers (e.g., GRHL2 and E-cadherin) are specifically reduced in basal B subtype breast cancer [1]. We examined whether *EHF* expression was also low in basal B versus other breast cancer subtypes. Indeed, RNA-seq data for a large panel of breast cancer cell lines showed loss of *EHF* in the basal B subtype **(Fig. 3f)** and qRT-PCR confirmed this finding in a smaller panel of breast cancer cell lines **(Fig. 3g)**.
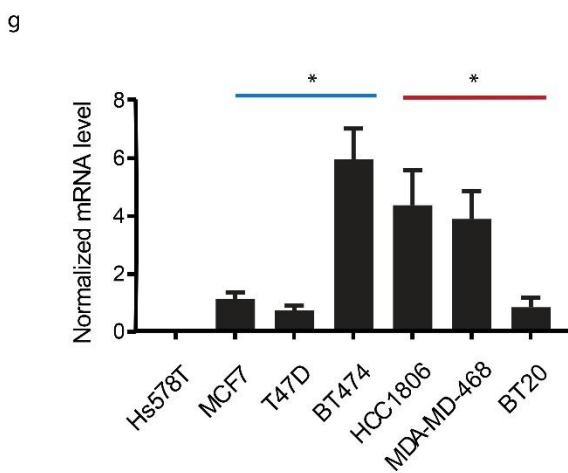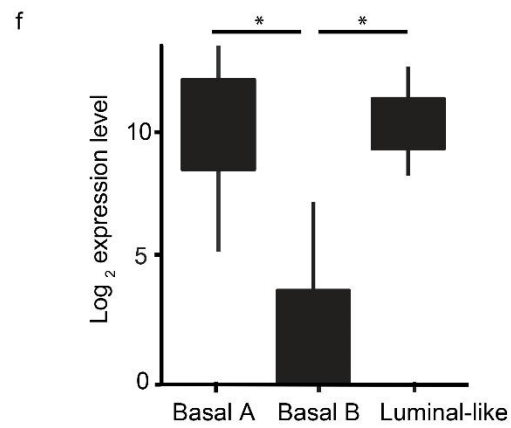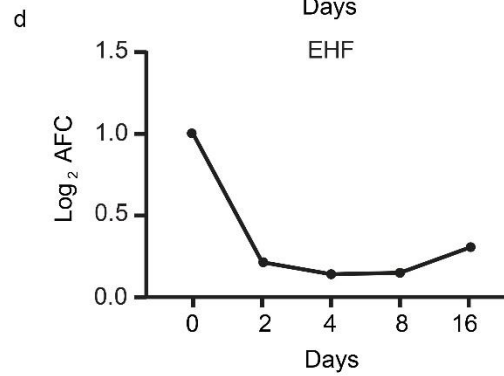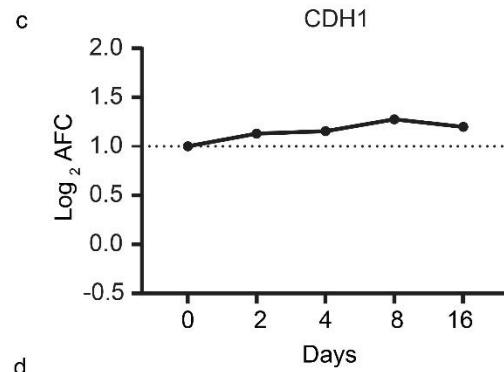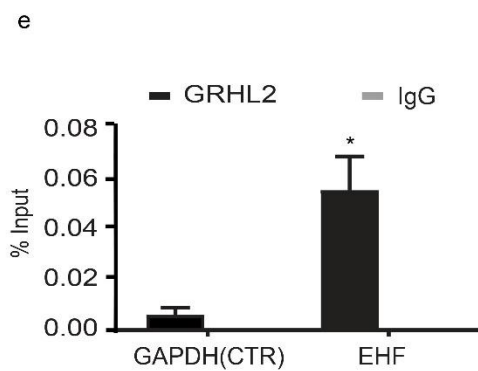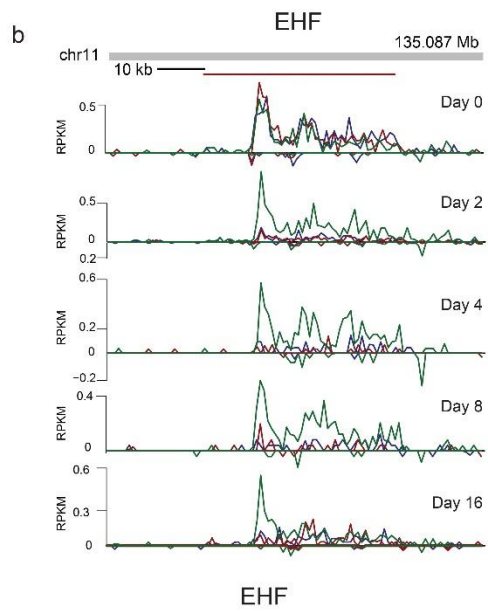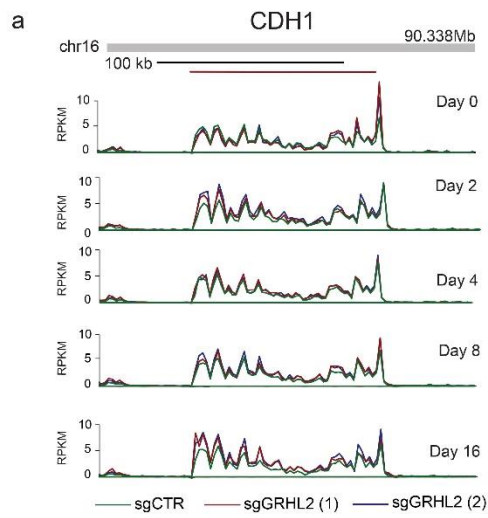
**Fig. 3 EHF but not CDH1 is a direct GRHL2 target gene in luminal breast cancer. (a and b)** Bru-seq reads for the *CDH1* and *EHF* gene at the indicated time points after GRHL2 loss. **(c and d)** Line graphs depicting the $\log_2$ AFC of *CDH1* and *EHF* transcription in MCF7 sgGRHL2 (1) and sgGRHL2 (2) cells at the indicated timepoints after doxycycline-induced GRHL2 deletion. 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)** ChIP-PCR 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 $p < 0.05$. **(f)** *EHF* mRNA expression in a panel of >50 human breast cancer cell lines covering luminal-, basal A-, and basal B-like subtypes extracted from RNA-seq data. Data is statistically analyzed by t-test and * indicates $p < 0.05$. **(g)** qRT-PCR validating downregulation of *EHF* mRNA in basal B-like subtype breast cancer. Blue and red lines represent luminal- and basal A-like subtypes of breast cancer, respectively. Data are presented as mean ± SEM from 2 biological replicates performed in triplicate. Normalized mRNA expression in each cell line is compared to the Hs578T basal B subtype cell line using t-test, * indicates $p < 0.05$.

## EHF overexpression does not rescue proliferation in GRHL2 KO cells

We previously found that GRHL2 loss induced inhibition of cell proliferation accompanied by a downregulation of *EHF* in MCF7 cells (Wang et al, manuscript under revision). Our previous work, exploring only Bru-seq analysis, also showed that EHF was connected to cell cycle regulators such as E2F1 **(see Chapter 4 Fig. 3)** and EHF has been implicated in tumor initiation and tumorigenesis [5,23]. To determine whether GRHL2 loss inhibits cell proliferation through reduced expression of its direct target, *EHF*, we investigated the effect of ectopically overexpressed *EHF* on cell proliferation. As shown in **Fig. 4a~d**, overexpression of *EHF* did not rescue GRHL2 loss-induced inhibition of cell proliferation and cell survival. Taken together, these findings indicate that *EHF* is a direct GRHL2-target whose expression is reduced in absence of GRHL2 but is not solely responsible for the proliferation arrest caused by GRHL2 depletion in luminal breast cancer cells.
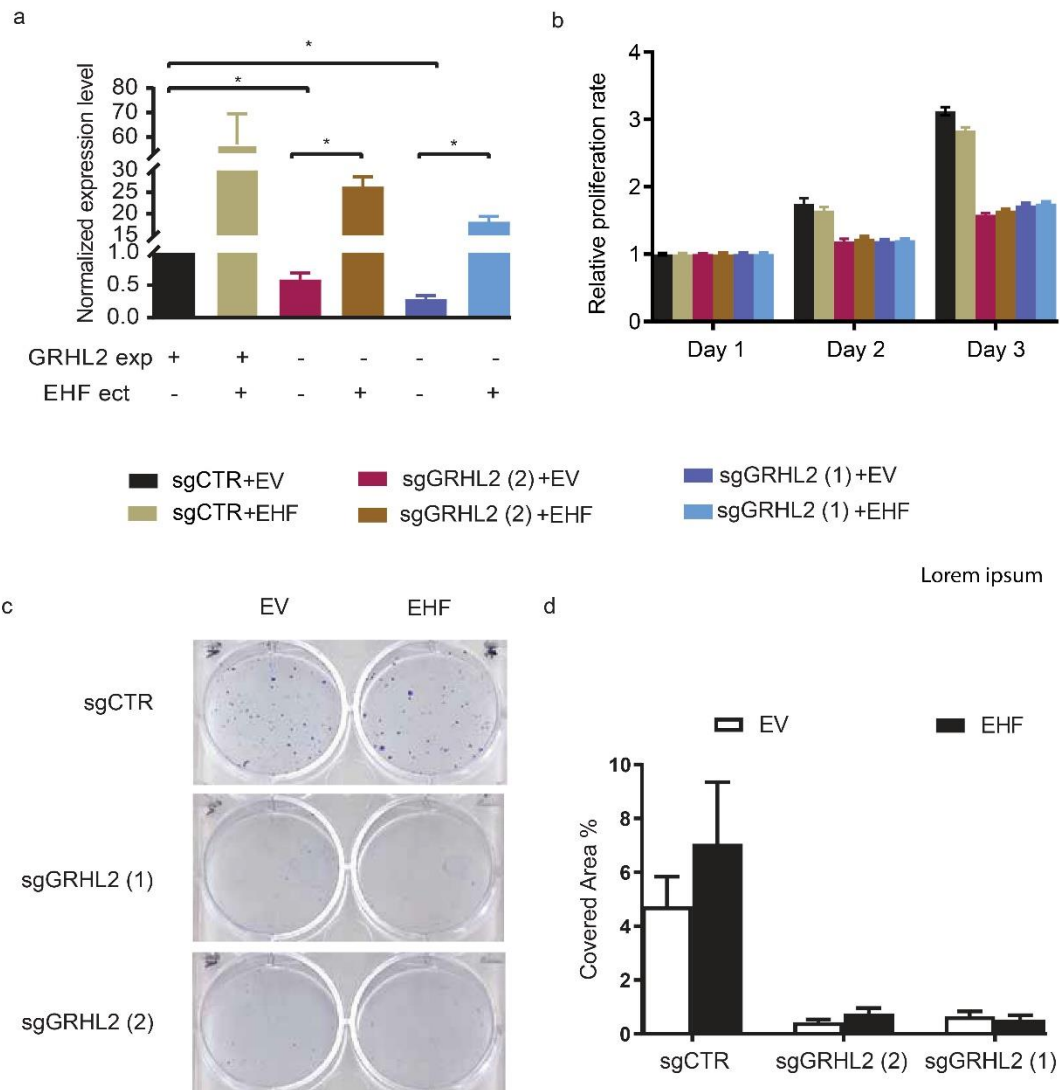
**Fig. 4 Ectopic expression of EHF does not rescue proliferation upon GRHL2-loss. (a)** Examination of expression level of *EHF* mRNA after 4 days of doxycycline treatment and transfection with *EHF* or EV plasmid by qRT-PCR in MCF7 cells transduced with sgCTR and sgGRHL2. Data are presented as mean ± SEM from three technical replicates. EV, empty vector; EHF, *EHF* cDNA plasmid. Data are statistically analyzed by t-test. * indicates *p* < 0.05. "GRHL2 exp" represents GRHL2 expression; "EHF ect" represents *EHF* ectopic overexpression. **(b)** Graph showing results from SRB assay after 4 days doxycycline-induction and transfection with *EHF*/EV plasmid and subsequent incubation for the indicated time periods. **(c,d)** Representative images of colony formation assay **(c)** and quantification of colony formation potential **(d)** for sgCTR and sgGRHL2 transduced MCF7 cells after 4 days doxycycline-induction and transfection with *EHF* or EV plasmid and subsequent incubation for the indicated time periods. Data are presented as mean ± SEM from 2 biological replicates performed in triplicate. Data are statistically analyzed by t-test.

# Discussion

By integrating ChIP-seq analysis with Bru-seq analysis of a conditional KO model, we identify 48 high confidence direct target genes of GRHL2 in luminal breast cancer MCF7 cells. Of 48 direct target genes, 44 genes can encode proteins, whose interactions are investigated. The other 4 encode non-coding RNAs. Interestingly, in the signaling networks identified by IPA using differentially expressed genes identified from Bru-seq, the hub genes are not the direct target genes of GRHL2 **(see Chapter 4 Fig. 3)**. Rather, GRHL2-regulated genes interact with candidate hub genes providing entry points for GRHL2 regulation of these signaling networks. SAPK/JNK and CDK5 signaling pathways associated with cell proliferation are included in the top 10 canonical pathways identified by genes regulated by GRHL2 directly. In addition, the functions of direct GRHL2 target genes are mainly enriched in cell cycle, growth and proliferation. This indicates that in luminal breast cancer MCF7 cells GRHL2 acts predominantly as a regulator of cell proliferation, which is consistent with our previous findings (Wang et al, manuscript under revision).

Notably, we demonstrate that *CDH1* RNA synthesis is not altered following GRHL2 loss, which confirms and extends an earlier report that *CDH1* mRNA has no obvious changes in response to GRHL2 depletion [24]. Our Bru-seq result is in line with the fact that the *CDH1* gene is not identified as a GRHL2 target by ChIP-seq. Together, this demonstrates that E-cadherin downregulation must occur in an indirect manner in our luminal breast cancer model. *CDH1* was identified as a direct target gene of GRHL2 in normal epithelial [3] due to remodeling of the *CDH1* promoter caused by an interaction of GRHL2 at an intron of *CDH1*. However, there was no existence of GRHL2 binding sites in the promoter region of *CDH1*, which means GRHL2 has no direct contact with a

DNA consensus motif from -1000 bp to 100 bp from the transcription start site (TSS) of *CDH1* [3]. To date, ChIP-seq data reveals no GRHL2 binding in the *CDH1* promoter region in human prostate cancer cells [17], ovarian cancer cells [19] and murine kidney cells [15]. Taken together, the *CDH1* gene is not a direct target of GRHL2.

We establish *EHF* as a direct GRHL2 target gene whose expression, like GRHL2, is lost in the basal B subtype of breast cancer. EHF was previously implicated in cell proliferation [5,25,26]. On the other hand, EHF may act as a tumor suppressor, resulting from its role in controlling differentiation, the cancer stem-like phenotype [27] and transcriptional repression of genes positively regulated by MAP kinase signaling cascades [28]. The biological function of EHF may be cell context-dependent. In our study, inhibition of cell proliferation caused by GRHL2 loss in MCF7 cells is accompanied by reduced expression of *EHF* but ectopic expression of *EHF* is not sufficient to rescue proliferation triggered by GRHL2 loss. Altogether, these findings unravel multiple connections between GRHL2 and regulation of proliferation and point to cooperative roles of GRHL2 target genes (e.g., *EHF, E2F2* and *CDCA7L*) in sustaining proliferation.

Taken together, this study identifies direct target genes of GRHL2 and their related signaling pathways. It also explores the co-regulation and function of the epithelial factors GRHL2 and EHF in luminal-like breast cancer.

# Reference

1 Cieply, B. *et al.* Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res* **72**, 2440-2453, doi:10.1158/0008-5472.CAN-11-4038 (2012).

2 Cieply, B., Farris, J., Denvir, J., Ford, H. L. & Frisch, S. M. Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res* **73**, 6299-6309, doi:10.1158/0008-5472.CAN-12-4082 (2013).

3 Werth, M. *et al.* The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835-3845, doi:10.1242/dev.055483 (2010).

4 Kas, K. *et al.* 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* **275**, 2986-2998, doi:10.1074/jbc.275.4.2986 (2000).

5 Lv, Y. *et al.* Increased expression of EHF contributes to thyroid tumorigenesis through transcriptionally regulating HER2 and HER3. *Oncotarget* **7**, 57978-57990, doi:10.18632/oncotarget.11154 (2016).

6 Imamachi, N. *et al.* BRIC-seq: a genome-wide approach for determining RNA stability in mammalian cells. *Methods* **67**, 55-63, doi:10.1016/j.ymeth.2013.07.014 (2014).

7 Paulsen, M. T. *et al.* Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67**, 45-54, doi:10.1016/j.ymeth.2013.08.015 (2014).

8 Wang, Z., Wu, H., Daxinger, L. & Danen, E. H. J. Genome-wide identification of binding sites of GRHL2 in luminal-like and basal A subtypes of breast cancer. *bioRxiv*, 2020.2002.2013.946947, doi:10.1101/2020.02.13.946947 (2020).

9 Haring, M. *et al.* Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**, 11, doi:10.1186/1746-4811-3-11 (2007).

10 Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**, D607-D613, doi:10.1093/nar/gky1131 (2019).

11 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

12 Pozo, K. & Bibb, J. A. The Emerging Role of Cdk5 in Cancer. *Trends in cancer* **2**, 606-618, doi:10.1016/j.trecan.2016.09.001 (2016).

13 Nishina, H., Wada, T. & Katada, T. Physiological roles of SAPK/JNK signaling pathway. *J Biochem* **136**, 123-126, doi:10.1093/jb/mvh117 (2004).

14 Mendonsa, A. M., Na, T. Y. & Gumbiner, B. M. E-cadherin in contact inhibition and cancer. *Oncogene* **37**, 4769-4780, doi:10.1038/s41388-018-0304-2 (2018).

15 Aue, A. *et al.* A Grainyhead-Like 2/Ovo-Like 2 Pathway Regulates Renal Epithelial Barrier Function and Lumen Expansion. *J Am Soc Nephrol* **26**, 2704-2715, doi:10.1681/ASN.2014080759 (2015).

16      Jolly, M. K. *et al.* E-Cadherin Represses Anchorage-Independent Growth in Sarcomas through Both Signaling and Mechanical Mechanisms. *Mol Cancer Res* **17**, 1391-1402, doi:10.1158/1541-7786.MCR-18-0763 (2019).

17      Paltoglou, S. *et al.* Novel Androgen Receptor Coregulator GRHL2 Exerts Both Oncogenic and Antimetastatic Functions in Prostate Cancer. *Cancer Res* **77**, 3417-3430, doi:10.1158/0008-5472.CAN-16-1616 (2017).

18      Xiang, X. *et al.* Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One* **7**, e50781, doi:10.1371/journal.pone.0050781 (2012).

19      Chung, V. Y. *et al.* GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943, doi:10.1038/srep19943 (2016).

20      Varma, S. *et al.* 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* **287**, 37282-37295, doi:10.1074/jbc.M112.408401 (2012).

21      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* **1868**, 584-591, doi:10.1016/j.bbcan.2017.06.006 (2017).

22      Park, S. M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & development* **22**, 894-907, doi:10.1101/gad.1640608 (2008).

23      Gong, Y. C. *et al.* miR206 inhibits cancer initiating cells by targeting EHF in gastric cancer. *Oncol Rep* **38**, 1688-1694, doi:10.3892/or.2017.5794 (2017).

24      Walentin, K. *et al.* A Grhl2-dependent gene network controls trophoblast branching morphogenesis. *Development* **142**, 1125-1136, doi:10.1242/dev.113829 (2015).

25      Cheng, Z. *et al.* Knockdown of EHF inhibited the proliferation, invasion and tumorigenesis of ovarian cancer cells. *Mol Carcinog* **55**, 1048-1059, doi:10.1002/mc.22349 (2016).

26      Shi, J. *et al.* 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* **7**, e2442, doi:10.1038/cddis.2016.346 (2016).

27      Albino, D. *et al.* The ETS factor ESE3/EHF represses IL-6 preventing STAT3 activation and expansion of the prostate cancer stem-like compartment. *Oncotarget* **7**, 76756-76768, doi:10.18632/oncotarget.12525 (2016).

28      Tugores, A. *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* **276**, 20397-20406, doi:10.1074/jbc.M010930200 (2001).

**Table S1.** GRHL2 direct target genes.

| | |
|---|---|
| SPATA18 | CADM1 |
| FOXP2 | WIPI1 |
| ZBTB20 | DUSP10 |
| MIR5087 | GPR87 |
| THAP11 | RND3 |
| DDX11 | TMEM140 |
| PSMB6 | SYNPO |
| UHRF1 | MTUS2 |
| TMPRSS13 | KIAA0513 |
| E2F2 | KIF5C |
| HAX1 | DDX58 |
| EEF2 | VMP1 |
| HMGB2 | LAMB3 |
| TYRO3 | FAM13A |
| EPN3 | DAPP1 |
| PPP1CA | PGLYRP2 |
| RPL41 | LINC00885 |
| HIST1H4B | EHF |
| HIST1H4E | DDX12P |
| ANKRD1 | NPY1R |
| GPR155 | PRSS23 |
| MAPK10 | CORO2A |
| PRICKLE2-AS3 | CADPS |
| LIMCH1 | CDCA7L |

# Chapter 6

**General summary and discussion**

**Breast cancer**

Among women, breast cancer is the most prevalent cancer type, accounting for 25% of all cancers [1]. According to gene expression profiling, breast cancer is mainly categorized into 5 major subtypes, including basal-like, luminal-like, human epidermal growth factor 2 (HER2)-enriched, claudin-low, and normal-like subtypes. Luminal- and basal-like subtypes have the highest prevalence and this dissertation focuses on these subtypes.

**Existing therapies for basal-like breast cancers**

At least 60% of basal-like breast cancers are negative for estrogen receptor (ER), progesterone receptor (PR) and HER2 [2], and are named triple negative breast cancer (TNBC). Basal-like breast cancer cell lines are further divided into basal A, being more luminal-like and basal B exhibiting a cancer stem cell like profile [3,4]. The cell lines generally reflect the genomic heterogeneity and gene copy number observed in the primary tumors [3] Due to the triple negative phenotype, TNBC responds poorly to existing targeted drugs. Multiple new potential therapeutic targets for TNBC are emerging such as poly(ADP-ribose) polymerase (PARP) 1, immune-checkpoints, androgen receptor, and epigenetic proteins [5]. PARP inhibitors (PARPi's) target PARP1, which is involved in cell proliferation, tumor transformation and DNA damage response. At present, several PARPi's (e.g., Olaparib, Veloparib, Niraparib, Talazoparib, and CEP7722) are under evaluation for clinical management. Among these candidate inhibitors, a monotherapy of olaparib induces partial responses in a cohort of mostly breast cancer patients with *BRCA1/2* mutations [6], suggesting it is a promising targeted drug. Cytotoxic chemotherapy still plays an important role for TNBC such as anthracyclines and taxanes. The effectiveness of chemotherapy strategies with dose dense (i.e. increasing dose intensity by short interval admission of standard-dose chemotherapy) or metronomic polychemotherapy has been validated [7]. Platinum agents (e.g. Cisplatin) have gained focus in the TNBC treatment, based on promising preclinical and clinical findings. Single cisplatin treatment allows patients with TNBC to achieve complete or near-complete responses in preoperative therapy [7]. On the other hand, it is well documented that TNBC responds to chemotherapy initially, but complete response at the early stage is not correlated with overall survival [8]. There is

no standard chemotherapy strategy for TNBC patients with specific tumor size, grade and lymph node [8]. The contribution of chemotherapy is still debated. A small proportion of TNBC are positive for androgen receptor (AR) providing an opportunity for targeted therapy. It has been shown that single AR antagonist bicalutamide treatment leads to a 19% benefit rate with a median progression-free survival of 12 weeks, with well-tolerated adverse side effects [9].

**Existing therapies for luminal-like breast cancers**

Luminal-like breast cancer is divided into two subgroups: luminal A and luminal B subtypes. The major difference between them is that luminal B has relatively lower expression of ER-related genes and higher expression of genes associated with cell growth (e.g., *CCND1*, *MKI67* and *MYBL2*) [10]. Luminal breast cancer cell lines are typically not further subdivided. Luminal A breast cancer is positive for ER and PR, with lower expression of proliferation-related genes (e.g., *MKI67*). Estrogen functions as a stimulus for growth and development of human mammary tissue by binding to ER. Selective ER modulators (SERMs) have high affinity for ER and inhibit activation of estrogen-mediated signaling in ER positive breast cancer. Furthermore, SERMs possess the capability to reduce expression of ER, further reducing the response to endogenous estrogen [11]. Tamoxifen is widely used in first-line clinical practice to repress growth of ER positive breast cancer by binding to the ER alpha [12,13]. In recent years, aromatase inhibitors (AIs) such as arimidex and aromasin, that inhibit estrogen biosynthesis, are replacing tamoxifen as hormone therapy. AIs are characterized by their capability to reduce estrogen biosynthesis as much as 98%, together with superiority over tamoxifen in terms of adverse side effects [14].

Luminal B breast cancer is less sensitive to endocrine therapy relative to the luminal A subtype, and to chemotherapy in comparison with basal-like and HER2 enriched breast cancer. Based on the aberrant activation of insulin-like growth factor 1 (IGF-1) signaling in luminal B tumors, IGF-1 inhibitors (e.g., OSI-906) and antibodies against IGF-1 receptor (MIK-0646) are developed and may lead to new candidate drugs [15]. Luminal B tumors are characterized by overexpression of fibroblast growth factor receptor 1 (FGFR1), indicating that antibodies and inhibitors against FGFR1 may also

represent potential drugs. Indeed, small molecular inhibitors such as TKI-258 and AZD-4547 are in phase 2 clinical testing [15]. Furthermore, the addition of PI3K inhibitors to endocrine therapy results in increased inhibition of growth in luminal B breast cancer cell lines, suggesting PI3K inhibitors may contribute to the treatment of luminal B breast cancers [16]. A negative feedback loop between mTOR and IGF-1 signaling has been described by which inhibition of mTOR signaling induces increased expression of IRS1, which in turn activates AKT signaling. A preclinical study shows that the combination of ridaforolimus (mTOR inhibitor) and dalotuzumab (antibody against IGF-1R) is a potential effective treatment against luminal B breast cancer [15].

**Distinct modes of E-cadherin regulation in normal epithelia and cancer cells**

GRHL2 is a transcription factor that activates or represses gene expression through indirect and direct binding to the promoter of the genes and histone modification [17,18]. GRHL2 is specifically expressed in cells with epithelial features. Several studies show that reduction of GRHL2 expression triggers epithelial to mesenchymal transition (EMT) with upregulation of mesenchymal markers (e.g., Vimentin and ZEB1) and downregulation of epithelial genes (e.g., *CDH1*) in breast cancer [17,19,20]. In **Chapter 2**, our results show that GRHL2 loss results in downregulation of E-cadherin (gene name: *CDH1*) in basal A (HCC1806 cells) and luminal-like (MCF7 cells) subtype breast cancer. However, our ChIP-seq data in **Chapter 3** reveal that GRHL2 binding is not observed at the promoter region of *CDH1* in any of the basal A (HCC1806, BT20 and MDA-MB-468 cells) and luminal-like (MCF7, T47D and BT474 cells) subtype breast cancer cell lines tested. In addition, our Bru-seq data in **Chapter 4** demonstrate that RNA synthesis of *CDH1* is not altered in response to GRHL2 deletion in MCF7 luminal-like breast cancer cells. This indicates that GRHL2 loss leads to E-cadherin downregulation while *CDH1* RNA synthesis is maintained, suggesting that GRHL2 regulates *CDH1* post-transcriptionally. This is in contrast to an earlier report showing that GRHL2 directly regulates transcription of *CDH1* in mouse inner medullary collecting duct cells by contacting the *CDH1* promoter through a chromatin loop [18]. This may suggest that GRHL2-mediated regulation of E-cadherin (and perhaps other epithelial genes) is markedly different in non-transformed epithelia versus epithelial cancer cells, such as the breast cancer cell lines tested by us.

**Biological functions of GRHL2 in cancer – from proliferation to invasion**

In **Chapter 2**, the effect of GRHL2 loss on EMT, cell migration (2D random migration), and cell invasion capacity (3D collagen gel invasion) are investigated. Interestingly, E-cadherin downregulation is seen in MCF7 as well as HCC1806 cells, while other effects appear to be subtype specific. GRHL2 loss triggers enhanced cell migration and invasion in HCC1806 cells, concomitantly with enhanced expression of mesenchymal markers i.e., N-cadherin and Vimentin [21,22], which are associated with cell migration. By contrast, in MCF7 cells GRHL2 loss does not lead to expression of mesenchymal markers and does not enhance the ability to invade. Others have shown *in vivo* and *in vitro* that exogenous expression N-cadherin allows MCF7 cells to invade more efficiently [23]. MCF7 cells expressing both N-cadherin and E-cadherin were also found to have the ability to invade, indicating that the mesenchymal marker, N-cadherin, may be a determinant of cell invasion, rather than loss of E-cadherin [23]. Taken together, these findings suggest that cell invasion triggered by GRHL2 deletion is associated with the emergence of mesenchymal markers and may be subtype specific.

In luminal cells GRHL2 may support proliferation and invasion through steroid hormone signaling. Activation of ERβ signaling contributes to inhibition of IMP3, which is an mRNA-binding protein that influences expression or localization of invasion-related mRNAs [24]. MCF7 cells are reported to express ERβ modestly [24,25] and ERβ signaling is regulated by GRHL2 that acts as a part of ER transcriptional complex to stimulate transcription of ER target genes [26]. Our Bru-seq data show that following GRHL2 deletion, *IMP3* RNA synthesis is upregulated at day 16, indicating that GRHL2 is involved in crosstalk between ERβ and IMP3. Hence, GRHL2 deletion may be concomitant with inactivation of ERβ signaling, which results in cell invasion through upregulation of IMP3 in HCC1806 cells.

Our study indicates that GRHL2 may have differential biological functions for basal A and luminal-like breast cancer. In **Chapter 3**, GRHL2 binding sites across breast cancer subtypes are profiled by Chip-seq to further investigate the differences in

molecular mechanisms. GRHL2 binding motifs are found, the landscape of GRHL2 binding sites is mapped, and, based on overlap between three luminal and three basal breast cancer cells lines, subtype-specific and common binding sites and candidate GRHL2-regulated genes are identified. Our findings to some extent confirm and extend previous findings in other cell types, such as the distribution of GRHL2 binding sites and GRHL2 binding motifs.

We also find novel GRHL2-interacting genes. However, ChIP-seq alone is insufficient to identify GRHL2 target genes, due to the fact that binding sites identified by ChIP-seq do not necessarily imply GRHL2-mediated gene expression. Therefore, to identify direct target genes of GRHL2 in breast cancer, Bru-seq is carried out in a conditional CRISPR/Cas9 MCF7 model in **Chapter 4**. Bru-seq is an innovative method to capture changes in initial transcription based on labeling and isolation of nascent RNA using bromouridine (Bru) [27]. Bru-seq measures transcription near transcription start sites (TSSs) and can capture initial transcription rapidly, by which the effects of post-transcriptional regulation on gene expression are eliminated such as RNA binding proteins and microRNAs. This chapter reveals dynamic changes in nascent RNA in response to GRHL2 loss. Through subsequent bioinformatics analysis such as Ingenuity Pathway analysis (IPA) and Gene Ontology terms analysis we provide new insights into molecular mechanisms that may underlie GRHL2 biological functions. As the generation of a conditional basal-A model was unsuccessful, we could so far only perform this analysis for the luminal MCF7 model system.

In **Chapter 2,** cell cycle analysis shows a rapid G0/1 arrest triggered by GRHL2 loss in HCC1806 and MCF7 cells, which is correlated with the findings in **Chapter 4** that demonstrates inhibition of genes involved in cell cycle/DNA replication signaling (*E2F1, E2F2, MCM7, CDC20, ESPL1, MCM2, PTTG1, SFN, RNASEH2A, FEN1*) by Bru-seq. In order to identify direct target genes of GRHL2, ChIP-seq and Bru-seq data are integrated in **Chapter 4**. In this way, 48 novel direct target genes of GRHL2 are identified, some of which are involved in cell cycle/DNA replication signaling. Integrating the data from **Chapters 2-4** provide novel insights into the molecular mechanism of GRHL2-mediated cell cycle/DNA replication regulation. It is well

documented that downregulation of GRHL2 inhibits cell proliferation in multiple human cell lines [20,28,29]. However, little is known about the molecular mechanism of GRHL2 loss-mediated inhibition of cell growth. Several direct target genes of GRHL2 have been previously implicated in cell growth such as *EHF* and *E2F2* [30,31]. Downregulation of EHF is associated with inhibition of cell growth in ovarian cancer cells [31]. However, our data shows that ectopic overexpression of EHF did not rescue GRHL2 loss-triggered inhibition of cell growth in MCF7 cells. One possible explanation is that EHF is not associated with cell proliferation in MCF7 cells, indicating that the GRHL2-EHF biological function is cell context dependent. Another possibility is that GRHL2 loss inhibits cell growth through multiple signaling pathways, which include several target genes. Rescue of one arm of the signaling network in this case is not sufficient to rescue GRHL2 loss-mediated cell proliferation. Indeed, there is currently no single target gene described that can rescue proliferation upon GRHL2 loss on its own, further pointing to cooperative roles of GRHL2 direct target genes (e.g., *E2F2* and *CDCA7L*) in cell proliferation.

**GRHL2-regulated functions as therapeutic targets in breast cancer**

Findings from us and others demonstrate that GRHL2 is involved in cell proliferation and cell cycle progression. Our data suggest that GRHL2 loss-induced cell invasion and multiple aspects of EMT occur in basal A but not luminal-like breast cancers. This implies that GRHL2 may represent a promising therapeutic target especially for luminal-like breast cancer. Although there is no molecular inhibitor targeting GRHL2 to date, inhibitors targeting downstream effectors of GRHL2-regulated proliferation signaling could be identified.

Miniature chromosome maintenance 7 (*MCM7*), a direct target gene of GRHL2 based on our ChIP-seq and Bru-seq data **(Chapter 3 and 4)**, is involved in the initiation of DNA replication. Reduced expression of MCM7 results in apoptosis in RB deficient cells and overexpression of MCM7 is associated with chemotherapy resistance [32]. Our findings show that GRHL2 loss-induced inhibition of cell proliferation is concomitant

with a downregulation of *MCM7* RNA synthesis, suggesting that targeting MCM7 may be a potential strategy for luminal breast cancer treatment.

As a widely used molecular inhibitor against 3-hydroxy-3-methylglutaryl coenzyme A (HMGCoA), simvastatin (SVA) that belongs to the statins may have anti-tumor effects. This was demonstrated by a*n in vivo* mouse model study showing SVA treatment inhibits breast cancer cell growth, through inhibiting MCM7 expression [32]. Additionally, SVA was reported to re-sensitize tamoxifen resistant breast cancer to chemotherapy [32,33]. Thus, the small molecular inhibitor SVA against HMGCoA represents a candidate drug for anti-breast cancer therapy that inhibits a GRHL2 target MCM7. The hexameric protein MCM2-7 complex is composed of six distinct subunits (named MCM2 to MCM7) with an AAA$^+$ ATPase that is targeted by the drug ciprofloxacin. MCM2 is identified as a direct GRHL2 target in our studies **(Chapter 3 and 4)**. Ciprofloxacin, a common and approved human fluoroquinolone inhibitor, was reported to inhibit cell growth by repressing the DNA helicase activity of MCM2-7 [34]. Other quinolone inhibitors such as 271327 and 314850 can selectively target MCM2-7 and may also be further developed for anti-breast cancer therapy. Targeting the enzymes involved in cancer metabolism, which facilitates tumorigenesis and metastasis, may be an alternative strategy for anti-cancer treatment. We identify an interesting candidate in this respect, the glycolytic enzyme aldolase A (ALDOA). ALDOA is a critical enzyme associated with cancer metabolic reprogramming and metastasis [35] that we find to be regulated by GRHL2 **(Chapter 3 and 4)**. The interaction of ALDOA with γ-actin is linked to enhanced metastatic potential of cancer cells and disruption of this interaction by the small molecular inhibitor raltegravir may represent a potential therapeutic strategy [35].

In the set of genes whose expression is attenuated in response to GRHL2 loss, involvement in cell cycle signaling is significantly enriched. A typical hallmark of cancer progression is dysregulation of the cell cycle, in which Cyclin D-dependent kinase 4/6 (CDK4/6) activity is enhanced by activated mitogenic pathways such as PI3K-AKT-mTOR and RAS-RAF-MEK-ERK signaling. CDK4/6 inhibitors, including abemaciclib, palbociclib and ribociclib interfere directly in the cell cycle and synergize with endocrine therapy [36]. As the first drug of CDK4/6 inhibitors, palbociclib benefits luminal-like breast

cancer patients when combined with letrozole or fulvestrant, leading to over 10-months improvement in median progression-free survival (mPFS) [37]. This approach has manageable adverse side effects such as neutropenia [37]. As the latest CDK4/6 inhibitor, abemaciclib can act as a monotherapy for luminal-like advanced/metastatic breast cancer, through inhibition of phosphorylation of RB and induction of G1-phase cell cycle arrest in RB-proficient cancer [36]. Meanwhile, abemaciclib induces breast cancer cell senescence with accumulated β- galactosidase, instead of cell apoptosis [38]. An *in vivo* study shows that abemaciclib can penetrate the blood-brain barrier and functions at lower does relative to palbociclib [39,40]. Compared with palbociclib and ribociclin, abemaciclib has higher selectivity for the CDK4/Cyclin D1 complex but there is no evidence for differences in antitumor effects between these drugs for patients with HR positive and HER2 negative breast cancer [36,38]. Unexpectedly, abemaciclib is also reported to facilitate antitumor immunity through enhancing tumor antigen presentation and inhibiting regulatory T cell proliferation. Moreover, type III interferons and interferon-related genes (e.g., *STAT1, IRF9* and *NLRC5*) are upregulated in the presence of abemaciclib in *in vivo* breast cancer models [38,41]. These observations suggest that abemaciclib triggers breast cancer cytostasis, rather than cell apoptosis, and consolidates antigen presentation to stimulate cytotoxic T cells. Collectively, GRHL2 and the components of signaling pathways regulated by GRHL2 may be potential targets for anti-breast cancer treatment.

**Strengths and limitations of the integrated approaches in this study**

Our study is the first to map the genome-wide landscape of GRHL2 binding sites across breast cancer subtypes (i.e., basal A and luminal-like). This is essential to understand potential subtype specific biological roles of GRHL2. In comparison to ChIP-chip (ChIP-microarray), ChIP–seq generates higher resolution, greater coverage and less noise data [42]. More replications of the ChIP-seq experiments would enhance confidence. On the other hand, by focusing only on candidate targets that are shared within 3 cell lines for each subtype or across all 6 cell lines, robustness of the data is ensured. Moreover, ChIP-PCR is used to validate key targets. Another aspect is the processing of the data to identify bona fide binding sites as well as deciding on the

most promising target genes with binding sites spanning across the genes rather than clustering at promoter regions.

The integration of ChIP-seq with our Bru-seq experiment significantly enhances the ability to find direct GRHL2 target genes. Bromouridine sequencing (Bru-seq) is an innovative method to measure RNA synthesis, based on metabolic labeling using bromouridine [43]. The expression level of an individual mRNA is determined by RNA production and degradation [44]. In contrast to standard RNA-seq, Bru-seq monitors changes in transcription and eliminates effects of post-transcriptional regulation/RNA stability on mRNA expression level [45]. Here, we make use of a conditional CRISPR/Cas9 system to delete GRHL2 gene and study the impact on genome-wide RNA synthesis. The CRISPR/Cas9 system has some intrinsic limitations including off-target effects. The fact that we could so far only design the model for MCF7 precludes a strategy such as used for ChIP-seq where overlay of data in different cell lines enhances robustness. Nevertheless, the overlap between ChIP-seq and Bru-seq represents a powerful approach to the identification of GRHL2 direct target genes and is a starting point for unraveling of GRHL2-regulated signaling networks as we do in **Chapter 4**.

A limitation of our study is the use of two-dimensional (2D) cultures where the cell environment differs significantly from the microenvironment of intact living tissues such as the interaction with extracellular matrix (ECM), the concentrations of essential nutrients, and tissue architecture [46]. While three-dimensional (3D) cultures are not used to collect samples for ChIP-seq or Bru-seq, we make use of a 3D collagen model to study the effect of GRHL2 deletion on invasion. Here, a collagen concentration is used that has been reported to mimic physiological and pathological tissue stiffness [47-49]. We show that effects of GRHL2 deletion on migration in 2D culture are recapitulated in 3D **(Chapter 2)**. Importantly, our results point to differential effects of GRHL2 loss on cell migration in 2D and 3D in basal A versus luminal-like breast cancer. Moreover, we identify novel GRHL2-regulated genes and signaling networks, we find GRHL2 targets that are conserved across epithelial tissues (other reports) and cancer cells (our work), and we discover an alternative relation between GRHL2 and the *CDH1*

gene, by which *CDH1* is not a direct target gene suggesting that mechanisms identified in non-transformed epithelial tissues (i.e. direct GRHL2 binding and regulation of *CDH1*) appear to be altered in cancer cells.

# Reference

1    Abrahams, H. J. *et al.* Risk factors, prevalence, and course of severe fatigue after breast cancer treatment: a meta-analysis involving 12 327 breast cancer survivors. *Ann Oncol* **27**, 965-974, doi:10.1093/annonc/mdw099 (2016).

2    Mehanna, J., Haddad, F. G., Eid, R., Lambertini, M. & Kourie, H. R. Triple-negative breast cancer: current perspective on the evolving therapeutic landscape. *Int J Womens Health* **11**, 431-437, doi:10.2147/IJWH.S178349 (2019).

3    Neve, R. M. *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527, doi:10.1016/j.ccr.2006.10.008 (2006).

4    Dai, X., Cheng, H., Bai, Z. & Li, J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J Cancer* **8**, 3131-3141, doi:10.7150/jca.18457 (2017).

5    Lee, A. & Djamgoz, M. B. A. Triple negative breast cancer: Emerging therapeutic modalities and novel combination therapies. *Cancer Treat Rev* **62**, 110-122, doi:10.1016/j.ctrv.2017.11.003 (2018).

6    Tutt, A. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet* **376**, 235-244, doi:10.1016/S0140-6736(10)60892-6 (2010).

7    Isakoff, S. J. Triple-negative breast cancer: role of specific chemotherapy agents. *Cancer J* **16**, 53-61, doi:10.1097/PPO.0b013e3181d24ff7 (2010).

8    O'Reilly, E. A. *et al.* The fate of chemoresistance in triple negative breast cancer (TNBC). *BBA Clin* **3**, 257-275, doi:10.1016/j.bbacli.2015.03.003 (2015).

9    Gucalp, A. *et al.* Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic Breast Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 5505-5512, doi:10.1158/1078-0432.CCR-12-3327 (2013).

10   Cheang, M. C. *et al.* Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute* **101**, 736-750, doi:10.1093/jnci/djp082 (2009).

11   Farzaneh, S. & Zarghi, A. Estrogen Receptor Ligands: A Review (2013-2015). *Sci Pharm* **84**, 409-427, doi:10.3390/scipharm84030409 (2016).

12   Jameera Begam, A., Jubie, S. & Nanjan, M. J. Estrogen receptor agonists/antagonists in breast cancer therapy: A critical review. *Bioorg Chem* **71**, 257-274, doi:10.1016/j.bioorg.2017.02.011 (2017).

13   Cronin-Fenton, D. P. & Damkier, P. Tamoxifen and CYP2D6: A Controversy in Pharmacogenetics. *Advances in pharmacology* **83**, 65-91, doi:10.1016/bs.apha.2018.03.001 (2018).

14   Potter, B. V. L. SULFATION PATHWAYS: Steroid sulphatase inhibition via aryl sulphamates: clinical progress, mechanism and future prospects. *Journal of molecular endocrinology* **61**, T233-T252, doi:10.1530/JME-18-0045 (2018).

15   Tran, B. & Bedard, P. L. Luminal-B breast cancer and novel therapeutic targets. *Breast cancer research : BCR* **13**, 221, doi:10.1186/bcr2904 (2011).

16   Creighton, C. J. *et al.* Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast cancer research : BCR* **12**, R40, doi:10.1186/bcr2594 (2010).

17   Chung, V. Y. *et al.* GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* **6**, 19943, doi:10.1038/srep19943 (2016).

18   Werth, M. *et al.* The transcription factor grainyhead-like 2 regulates the molecular composition of the epithelial apical junctional complex. *Development* **137**, 3835-3845, doi:10.1242/dev.055483 (2010).

19   Xiang, X. *et al.* Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One* **7**, e50781, doi:10.1371/journal.pone.0050781 (2012).

20   Werner, S. *et al.* Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *The Journal of biological chemistry* **288**, 22993-23008, doi:10.1074/jbc.M113.456293 (2013).

21   Mrozik, K. M., Blaschuk, O. W., Cheong, C. M., Zannettino, A. C. W. & Vandyke, K. N-cadherin in cancer metastasis, its emerging role in haematological malignancies and potential as a therapeutic target in cancer. *BMC Cancer* **18**, 939, doi:10.1186/s12885-018-4845-0 (2018).

22    Cheng, C. W. *et al.* MicroRNA-30a inhibits cell migration and invasion by downregulating vimentin expression and is a potential prognostic marker in breast cancer. *Breast Cancer Res Treat* **134**, 1081-1093, doi:10.1007/s10549-012-2034-4 (2012).

23    Hazan, R. B., Phillips, G. R., Qiao, R. F., Norton, L. & Aaronson, S. A. Exogenous expression of N-cadherin in breast cancer cells induces cell migration, invasion, and metastasis. *J Cell Biol* **148**, 779-790, doi:10.1083/jcb.148.4.779 (2000).

24    Hinsche, O., Girgert, R., Emons, G. & Grundker, C. Estrogen receptor beta selective agonists reduce invasiveness of triple-negative breast cancer cells. *Int J Oncol* **46**, 878-884, doi:10.3892/ijo.2014.2778 (2015).

25    Vladusic, E. A., Hornby, A. E., Guerra-Vladusic, F. K., Lakins, J. & Lupu, R. Expression and regulation of estrogen receptor beta in human breast tumors and cell lines. *Oncol Rep* **7**, 157-167, doi:10.3892/or.7.1.157 (2000).

26    Reese, R. M., Harrison, M. M. & Alarid, E. T. Grainyhead-like Protein 2: The Emerging Role in Hormone-Dependent Cancers and Epigenetics. *Endocrinology* **160**, 1275-1288, doi:10.1210/en.2019-00213 (2019).

27    Venkata Narayanan, I. *et al.* Transcriptional and post-transcriptional regulation of the ionizing radiation response by ATM and p53. *Sci Rep* **7**, 43598, doi:10.1038/srep43598 (2017).

28    Chen, W. *et al.* 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* **285**, 40852-40863, doi:10.1074/jbc.M110.103812 (2010).

29    Hu, F., He, Z., Sun, C. & Rong, D. Knockdown of GRHL2 inhibited proliferation and induced apoptosis of colorectal cancer by suppressing the PI3K/Akt pathway. *Gene* **700**, 96-104, doi:10.1016/j.gene.2019.03.051 (2019).

30    Chen, L., Yu, J. H., Lu, Z. H. & Zhang, W. E2F2 induction in related to cell proliferation and poor prognosis in non-small cell lung carcinoma. *International journal of clinical and experimental pathology* **8**, 10545-10554 (2015).

31    Cheng, Z. *et al.* Knockdown of EHF inhibited the proliferation, invasion and tumorigenesis of ovarian cancer cells. *Mol Carcinog* **55**, 1048-1059, doi:10.1002/mc.22349 (2016).

32    Li, J. *et al.* Simvastatin and Atorvastatin inhibit DNA replication licensing factor MCM7 and effectively suppress RB-deficient tumors growth. *Cell Death Dis* **8**, e2673, doi:10.1038/cddis.2017.46 (2017).

33    Gopalan, A., Yu, W., Sanders, B. G. & Kline, K. Eliminating drug resistant breast cancer stem-like cells with combination of simvastatin and gamma-tocotrienol. *Cancer Lett* **328**, 285-296, doi:10.1016/j.canlet.2012.10.003 (2013).

34    Simon, N. *et al.* Ciprofloxacin is an inhibitor of the Mcm2-7 replicative helicase. *Biosci Rep* **33**, doi:10.1042/BSR20130083 (2013).

35    Chang, Y. C. *et al.* Therapeutic Targeting of Aldolase A Interactions Inhibits Lung Cancer Metastasis and Prolongs Survival. *Cancer Res* **79**, 4754-4766, doi:10.1158/0008-5472.CAN-18-4080 (2019).

36    Kotake, T. & Toi, M. Abemaciclib for the treatment of breast cancer. *Expert Opin Pharmacother* **19**, 517-524, doi:10.1080/14656566.2018.1448787 (2018).

37    Serra, F. *et al.* Palbociclib in metastatic breast cancer: current evidence and real-life data. *Drugs Context* **8**, 212579, doi:10.7573/dic.212579 (2019).

38    Corona, S. P. & Generali, D. Abemaciclib: a CDK4/6 inhibitor for the treatment of HR+/HER2-advanced breast cancer. *Drug Des Devel Ther* **12**, 321-330, doi:10.2147/DDDT.S137783 (2018).

39    Raub, T. J. *et al.* Brain Exposure of Two Selective Dual CDK4 and CDK6 Inhibitors and the Antitumor Activity of CDK4 and CDK6 Inhibition in Combination with Temozolomide in an Intracranial Glioblastoma Xenograft. *Drug Metab Dispos* **43**, 1360-1371, doi:10.1124/dmd.114.062745 (2015).

40    McCartney, A. *et al.* The role of abemaciclib in treatment of advanced breast cancer. *Ther Adv Med Oncol* **10**, 1758835918776925, doi:10.1177/1758835918776925 (2018).

41    Goel, S. *et al.* CDK4/6 inhibition triggers anti-tumour immunity. *Nature* **548**, 471-475, doi:10.1038/nature23465 (2017).

42    Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680, doi:10.1038/nrg2641 (2009).

43    Paulsen, M. T. *et al.* Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67**, 45-54, doi:10.1016/j.ymeth.2013.08.015 (2014).

44    Yamada, T. *et al.* 5'-Bromouridine IP Chase (BRIC)-Seq to Determine RNA Half-Lives. *Methods Mol Biol* **1720**, 1-13, doi:10.1007/978-1-4939-7540-2_1 (2018).

45      Kirkconnell, K. S., Paulsen, M. T., Magnuson, B., Bedi, K. & Ljungman, M. Capturing the dynamic nascent transcriptome during acute cellular responses: The serum response. *Biol Open* **5**, 837-847, doi:10.1242/bio.019323 (2016).

46      Yuan, Z. *et al.* Development of a 3D Collagen Model for the In Vitro Evaluation of Magnetic-assisted Osteogenesis. *Sci Rep* **8**, 16270, doi:10.1038/s41598-018-33455-2 (2018).

47      Kim, J. E., Reynolds, D. S., Zaman, M. H. & Mak, M. Characterization of the mechanical properties of cancer cells in 3D matrices in response to collagen concentration and cytoskeletal inhibitors. *Integr Biol (Camb)* **10**, 232-241, doi:10.1039/c8ib00044a (2018).

48      Hasebe, T., Tsuda, H., Tsubono, Y., Imoto, S. & Mukai, K. Fibrotic focus in invasive ductal carcinoma of the breast: a histopathological prognostic parameter for tumor recurrence and tumor death within three years after the initial operation. *Jpn J Cancer Res* **88**, 590-599, doi:10.1111/j.1349-7006.1997.tb00423.x (1997).

49      Provenzano, P. P. *et al.* Collagen density promotes mammary tumor initiation and progression. *BMC Med* **6**, 11, doi:10.1186/1741-7015-6-11 (2008).

## Samenvatting

Van de transcriptiefactor Grainyhead like 2 (GRHL2) wordt gerapporteerd dat het in sommige gevallen de groei van kanker bevordert en in andere studies aspecten van de progressie van kanker onderdrukt. In **Hoofdstuk 2** hebben we de rol ervan onderzocht in luminale en basale A subtypes van borstkanker. Bij borstkankerpatiënten verschilde de associatie van GRHL2-expressie met prognose voor verschillende subtypen en in een groot cellijnpaneel was GRHL2 laag of afwezig in basaal B- en aanwezig in alle luminale en basale A-cellijnen. In een luminale cellijn (MCF7) veroorzaakte het verlies van GRHL2 een blokkering van de celcyclus, verlies van kolonievormingscapaciteit en downregulatie van PCNA en hTERT. Tegelijkertijd ging E-cadherine verloren, maar er werd slechts een kleine toename in door EGF gestimuleerde motiliteit waargenomen. In een basale A-cellijn (HCC1806) onderdrukte de verwijdering van GRHL2 ook proliferatie en kolonievorming, maar er werden geen veranderingen gezien in PCNA en hTERT. Verlies van E-cadherine ging in dit geval gepaard met inductie van vimentine en N-cadherine en conversie naar een sterk migrerend fenotype, verder versterkt door EGF-behandeling. Deze resultaten wijzen op verschillende reacties op verlies van GRHL2 in luminale en basaalachtige borstkankers met betrekking tot groeiachterstand motiliteit, en suggereren dat GRHL2 een kandidaat-doelwit kan zijn bij luminale borstkanker.

GRHL2 ChIP-seq werd uitgevoerd in drie luminaalachtige en drie basale A-borstkankercellijnen van de mens om gemeenschappelijke en subtype-specifieke genomische bindingsplaatsen van GRHL2 bij borstkanker te identificeren (**Hoofdstuk 3**). De meeste bindingsplaatsen van GRHL2 werden gevonden in intergene en intronregio's. 13.351 gemeenschappelijke bindingsplaatsen werden geïdentificeerd in basale A-cellen, waaronder 551 bindingsplaatsen in gen promotor gebieden. Voor luminaalachtige cellen werden 6.527 gemeenschappelijke bindingsplaatsen geïdentificeerd, waarvan 208 bindingsplaatsen werden gevonden in genpromotorgebieden. Basale A- en luminaalachtige borstkankercellen deelden 4711 GRHL2-bindingsplaatsen, waarvan 171 bindingsplaatsen werden gevonden in genpromotorgebieden. De geïdentificeerde GRHL2-bindende motieven zijn allemaal identiek aan een motief dat is gerapporteerd voor menselijke eierstokkanker, wat wijst

op geconserveerde GRHL2-DNA-binding tussen menselijke kankercellen. Er zijn met opmerkelijk genoeg geen bindingsplaatsen van GRHL2 gedetecteerd in de promotorregio's van verschillende bekende EMT-gerelateerde genen, waaronder CDH1-, ZEB1-, ZEB2- en CDH2-genen. Gezamenlijk biedt deze studie een uitgebreid overzicht van interacties van GRHL2 met DNA en legt het de basis voor verder begrip van gemeenschappelijke en subtype specifieke signaalroutes gereguleerd door GRHL2 bij borstkanker.

**Hoofdstuk 4** toonde aan dat dynamische veranderingen optreden in de aanmaak van nieuw RNA na verlies van GRHL2 in luminale borst door Bru-seq. De respons op GRHL2-verlies in luminale borstkankercellen werd bestudeerd door een MCF7-conditioneel knock-outmodel te combineren met Bru-seq-analyse. De snelheid van RNA-synthese van 264- en 244-genen werd respectievelijk voor ten minste één van de vier tijdspunten opwaarts of neerwaarts gereguleerd na verlies van GRHL2, variërend van 1-16 dagen. Er werden vijf dynamische responspatronen gekarakteriseerd en GRHL2-gecontroleerde canonieke signaleringsroutes en netwerken werden geïdentificeerd. Gezamenlijk karakteriseert dit hoofdstuk patronen van RNA-synthese gereguleerd door GRHL2 en identificeert het signaleringsroutes gereguleerd door GRHL2.

In **Hoofdstuk 5** werd ChIP-seq-analyse van GRHL2-bindende genen in MCF7-cellen geïntegreerd met Bru-seq-analyse van genen die transcriptionele reacties op conditionele CRIPR-Cas9-knock-out van GRHL2 in MCF7-cellen laten zien. Dit identificeerde 48 directe doelwitgenen van GRHL2 in MCF7-cellen. Signaleringsroutes en netwerken geassocieerd met deze directe GRHL2-doelgenen werden onderzocht met behulp van IPA sofware. Met name, in overeenstemming met ons vorige rapport dat de CDH1-promotor GRHL2-bindingsplaatsen mist, werd de RNA-synthese van CDH1, die codeert voor de epitheliale adhesie-receptor E-cadherine, niet veranderd na GRHL2-deletie, wat aantoont dat CDH1 geen direct doelgen van GRHL2 is. In plaats daarvan werd de epitheliale transcriptiefactor, EHF/ ESE3, een transcriptiefactor die, zoals GRHL2, EMT onderdrukt, geïdentificeerd als een direct doelgen van GRHL2. EHF was op alle tijdstippen neerwaarts gereguleerd na verwijdering van GRHL2 en,

net als GRHL2, was EHF specifiek afwezig bij basale B-achtige borstkanker in een pan-subtype menselijk borstkankercellijnpaneel. EHF is betrokken bij tumor-initiërende eigenschappen. De overexpressie ervan slaagde er echter niet in de proliferatie van borstkankercellen waarin GRHL2 was uitgeschakeld te redden. Gezamenlijk identificeert deze studie directe doelgenen van GRHL2 en hun gerelateerde signaalroutes en werpt licht op de epitheliale factoren GRHL2 en EHF in luminale-achtige borstkanker MCF7-cellen.

# Curriculum Vitae

Zi Wang was born on 1$^{st}$ November 1988 in Tianjin, China. After he obtained his Master of Medicine Degree at Nankai University, he conducted PhD research in Erik Danen's group at Leiden University, with the support of China Scholarship Council in 2015. In his PhD project, he focused on identification of Grainyhead-like 2 target genes in breast cancer.

# List of publications

**Phenotypic screening of cannabinoid receptor 2 ligands shows different sensitivity to genotype**

Hillger, J. M., le Roy, B., **Wang, Z.**, Mulder-Krieger, T., Boomsma, D. I., Slagboom, P. E., Danen, E. H. J., I. Jzerman AP, Heitman, L. H.

Biochem Pharmacol 2017, 130: 60-70

**Genome-wide identification of binding sites of GRHL2 in luminal-like and basal A subtypes of breast cancer.**

**Zi Wang**, Haoyu Wu, Lucia Daxinger, Erik HJ Danen

bioRxiv, 2020.02.13.946947

**The role of GRHL2 in luminal- and basal like breast cancer**

**Zi Wang**, Bircan Coban, Chen-Yi Liao, Yao-Jun Chen, Erik HJ Danen

Scientific Reports, under revision

**GRHL2 regulated signaling networks in breast cancer**

**Zi Wang**, Yao-Jun Chen, Mats Ljungman, Erik HJ Danen

Manuscript in preparation