RESEARCH ARTICLE

Transcription factors Krüppel-like factor 4 and paired box 5 regulate the expression of the Grainyhead-like genes

Grzegorz Kotarba[1☯], Agnieszka Taracha-Wisniewska[1☯], Michal Miller[2], Michal Dabrowski[2], Tomasz Wilanowski[1] *

1 Faculty of Biology, Institute of Genetics and Biotechnology, University of Warsaw, Warsaw, Poland,
2 Laboratory of Bioinformatics, Nencki Institute of Experimental Biology of Polish Academy of Sciences, Warsaw, Poland

☯ These authors contributed equally to this work.
* t.wilanowski@biol.uw.edu.pl

Abstract

Genes from the Grainyhead-like (*GRHL*) family code for transcription factors necessary for the development and maintenance of various epithelia. These genes are also very important in the development of many types of cancer. However, little is known about the regulation of expression of *GRHL* genes. Previously, there were no systematic analyses of the promoters of *GRHL* genes or transcription factors that bind to these promoters. Here we report that the Krüppel-like factor 4 (KLF4) and the paired box 5 factor (PAX5) bind to the regulatory regions of the *GRHL* genes and regulate their expression. Ectopic expression of KLF4 or PAX5 alters the expression of *GRHL* genes. In KLF4-overexpressing HEK293 cells, the expression of *GRHL1* and *GRHL3* genes was upregulated by 32% and 60%, respectively, whereas the mRNA level of *GRHL2* gene was lowered by 28% when compared to the respective controls. The levels of *GRHL1* and *GRHL3* expression were decreased by 30% or 33% in PAX5-overexpressing HEK293 cells. The presence of minor frequency allele of single nucleotide polymorphism rs115898376 in the promoter of the *GRHL1* gene affected the binding of KLF4 to this site. The evidence presented here suggests an important role of KLF4 and PAX5 in the regulation of expression of *GRHL1-3* genes.

Introduction

The Grainyhead-like (*GRHL*) genes belong to the *TFCP2*/Grainyhead family, and can be found in all animal species and fungi that were studied so far [1]. In mammals this family has two distinct branches, the Grainyhead-like subfamily termed *GRHL1-3* [2,3] and the other subfamily consisting of three genes: *TFCP2*, *TFCP2L1* and *UBP1*, which code for transcription factors with the same names (recently reviewed in [4,5]). The expression of the GRHL factors is tissue- and developmentally-specific, and they are found primarily in epithelial tissues, in organs such as epidermis, oral and olfactory epithelium, kidneys and urogenital tract, stomach and the digestive tract, heart and lung [6]. These factors are involved in a wide variety of bio-logical processes including: skin barrier formation [7,8], neural tube closure [9,10], wound

healing [8,11] and cancer development [12–14]. Furthermore, they are not merely passive markers of tumor growth, but they also directly influence the process of carcinogenesis [12–14]. Although this gene family plays a pivotal role in epithelial maintenance and development, no systematic analyses have ever been carried out to identify and characterize transcription factors binding to promoters or enhancers of *GRHL1-3* genes. The aim of our research was to fill this gap in our knowledge. The results from our previous works as well as literature analyses allowed us to propose two transcription factors that could potentially regulate the expression of *GRHL1-3* genes: the Krüppel-like factor 4 (KLF4) and the paired box 5 factor (PAX5). KLF4 is highly expressed in the epidermis where it is responsible for the formation of skin barrier [15]. The deficiency of this transcription factor is correlated with increased cell proliferation and skin tumorigenesis [16,17]. Previously we observed that in mice lacking the functional *Grhl3* gene the hyperproliferative defect in their epidermis is similar to another barrier defective mouse model, the *Klf4*-nullizygous mice, indicating that loss of these transcription factors leads to analogous phenotypic aberrations [18,19]. In addition, both GRHL3 and KLF4 are involved in wound healing [8,11,20]. The results from our more recent studies indicate that some single nucleotide polymorphisms (SNP), located in the promoter regions of genes from the *GRHL* family, occur with statistically significantly altered frequencies in patients with clear cell renal cell carcinoma (ccRCC) and non-melanoma skin cancer (NMSC) [21,22]. Our bioinformatics analyses revealed that the presence of one of these polymorphisms in patients with ccRCC may alter the binding of KLF4 [22]. PAX genes are critically required during embryogenesis and play important roles in cancer progression. PAX5 was found to interact with the promoter of human telomerase reverse transcriptase (*hTERT*) gene and regulate its transcription, and GRHL2 also regulates the expression of *hTERT* [23,24]. More links between PAX5 and *GRHL* genes were discovered in our studies of the midbrain-hindbrain boundary development in the zebrafish model, showing that the formation of this boundary is governed by the interplay between multiple transcription factors including PAX2/5/8 and GRHL2 orthologue– GRHL2B [25,26]. On the basis of these observations we put forward a hypothesis that KLF4 and PAX5 may regulate the expression of genes from the *GRHL* family. Moreover, the presence of different alleles of SNPs that affect transcription factor binding sites in the promoter regions of *GRHL* genes, may result in altered expression of these genes. For all of the above reasons, we carried out bioinformatics analyses of the regulatory regions for each gene: *GRHL1*, *GRHL2* and *GRHL3*, and identified potential KLF4 and PAX5 binding sites. Subsequently we carried out experiments to investigate whether these transcription factors bind to the regulatory regions of *GRHL* genes and regulate their expression. We also examined whether the presence of minor frequency allele of SNP rs115898376 in the promoter of *GRHL1* gene alters the binding of KLF4 and affects the level of expression of *GRHL1*. Our results confirmed that KLF4 and PAX5 regulate the expression of *GRHL* genes, and that SNP rs115898376 in the promoter of *GRHL1* gene affects the binding of KLF4 to this binding site.

Materials and methods

In silico prediction of KLF4 and PAX5 binding sites in the regulatory elements of *GRHL* genes

Regulatory regions for further analyses were chosen based on Nencki Genomic Database (NGD) [27] which is convergent with Ensembl version 79. For *GRHL1* and *GRHL2* we selected 4000 bp regions flanking the transcription start sites (TSS) (2000 bp upstream and 2000 bp downstream of TSS). For *GRHL3* we selected a wider region of 6000 bp (2000 bp downstream of TSS and 4000 bp upstream), guided by its longer open chromatin and histone modification regions (Fig 1B). We analyzed the occurrence of transcription start sites, DNase I

hypersensitivity sites, CpG islands and patterns of histone H3K9ac acetylation and H3K4me3 methylation (DNase I and histone modification data from any cell type) from Ensembl version 79 in the promoters of *GRHL* genes and concluded that the above-listed regions are the most likely to contain promoters of these genes (Fig 1B). The data were visualized using the WSDL (Web Service Description Language) webservice of the NGD database, accessed with the client PlotGenomic.t2flow provided as S1 File. In this file, all the analysis parameters mentioned above are explicitly provided as example values, viewable and usable upon opening the client in Taverna Workbench [28] http://www.taverna.org.uk/download/workbench/. The genomic coordinates of the selected TSS flanks of the *GRHL1-3* genes are provided in the "Analyzed regions" table in S2 File. For each of these human regions, we downloaded (from Ensembl ver. 79) its pairwise genomic alignments with orthologous sequences from 6 mammalian species: rhesus macaque (rheMac), mouse (mm), dog (canFam), horse (equCab), cow (bosTau), and opossum (monDom). This was achieved via the Ensembl Genome Browser (ver. 79), using the options: Configure this page, Comparative genomics, BLASTz/LASTz alignments; after choosing the orthologous species and the genomic coordinates of the human region; and the output was saved in the Fasta format. Multiple alignments were performed with T-Coffee [29]. Prior to using the pairwise alignments as the input for T-Coffee all the gaps in them were manually removed. We used a command line Version_12.00.7fb08c2 of T-Coffee, run with this command:./t_coffee -in = my_fasta.txt -mode = regular -output = score_html clustalw_aln fasta_aln score_ascii phylip -maxnseq = 150 -maxlen = 60000 -case = upper -seqnos = off -outorder = input -run_name = result -multi_core = 4 -quiet = stdout. For predicting transcription factor binding sites (TFBS) we used MotEvo [30] (ver 1.0) with standard settings, including position weight matrix (PWM) and phylogenetic tree ((((hg:0.048,rheMac:0.048): 0.143,mm: 0.489): 0.030((canFam: 0.224, equCab: 0.149): 0.011,bosTau:0.246):0.047):0.365, monDom:0.481). This phylogenetic tree was used by the authors of MotEvo in their original publication [30]. We used TFBS motifs for KLF4 (KLF4.p3) and PAX5 (PAX5.p2) from the Swiss Regulon [31] motif library for maximum compatibility with MotEvo [30] developed by the same authors. In analyses of regions located downstream and upstream of TSS we considered only those TFBSs which met the requirements of threshold that we set (p aposteriori > 0.8 and weight matrix (WM)-score > 8). Additionally, for KLF4, we chose the previously identified KLF4 motif site within the SNP rs115898376 (C/T) and also weaker motifs (p aposteriori < 0.8 and WM-score < 8) [22]. For the comparison of the MotEvo predictions with the ChIP-seq data from the Cistrome database [32] we used the function "What factors bind in your interval?" of http://dbtoolkit.cistrome.org/.

Cell culture

Human embryonic kidney 293 cells (HEK293) were a kind gift from Ewelina Szymanska. The cell line was authenticated by LGC Standards (Teddington, Middlesex, UK). The cells were maintained in DMEM GlutaMAX medium supplemented with 10% fetal bovine serum (FBS) and 100 IU/mL penicillin-streptomycin at 37˚C in a humidified atmosphere with 5% CO_2. Cell culture components were purchased from Thermo Fisher Scientific.

Plasmids

The pcDNA3.1-HA-KLF4 FL plasmid was a kind gift from Michael Ruppert (Addgene plasmid #34593) [33], pcDNA3.1-N-DYK-PAX5 plasmid was purchased from GenScript, the control pcDNA3.1 plasmid was purchased from Invitrogen. In co-transfection experiment no. 1, using In-Fusion® HD Cloning Kit (Takara Bio), according to the protocol provided by the manufacturer, synthetic oligonucleotides of regulatory elements of *GRHL1* (50 bp from -429
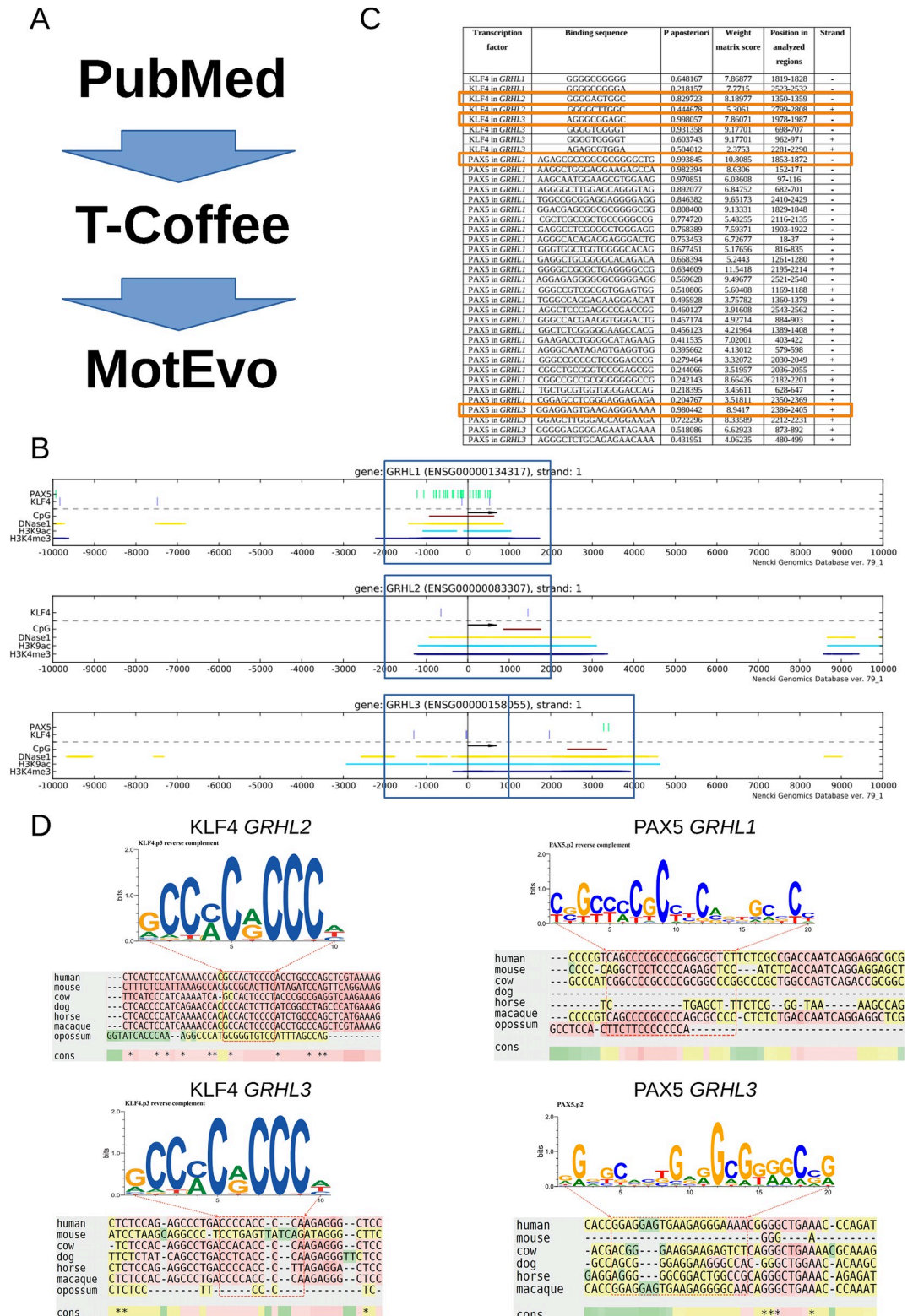
A

PubMed

T-Coffee

MotEvo

C

Transcription factor	Binding sequence	P aposteriori	Weight matrix score	Position in analyzed regions	Strand
KLF4 in *GRHL1*	GGGGCGGGGG	0.648167	7.86877	1819-1828	-
KLF4 in *GRHL1*	GGGGGCGGGGA	0.218157	7.7715	2523-2532	-
KLF4 in *GRHL2*	GGGGAGTGGC	0.829723	8.18977	1350-1359	-
KLF4 in *GRHL2*	GGGGCTTGGC	0.444678	5.3061	2799-2808	+
KLF4 in *GRHL3*	AGGGCGGAGC	0.998057	7.86071	1978-1987	-
KLF4 in *GRHL3*	GGGGTGGGGT	0.931358	9.17701	698-707	-
KLF4 in *GRHL3*	GGGGTGGGGT	0.603743	9.17701	962-971	+
KLF4 in *GRHL3*	AGAGCGTGGA	0.504012	2.3753	2281-2290	+
PAX5 in *GRHL1*	AGAGCGCCGGGGCGGGGCTG	0.993845	10.8085	1853-1872	-
PAX5 in *GRHL1*	AAGGCTGGGAGGAAGAGCCA	0.982394	8.6306	152-171	+
PAX5 in *GRHL1*	AAGCAATGGAAGCGTGGCAAG	0.970851	6.03608	97-116	-
PAX5 in *GRHL1*	AGGGGCTTGGAGCAGGGTAG	0.892077	6.84752	682-701	-
PAX5 in *GRHL1*	TGGCCGCGGAGGAGGGGAGG	0.846382	9.65173	2410-2429	-
PAX5 in *GRHL1*	GGACGAGCGGGCGCGGGGCGG	0.808400	9.13331	1829-1848	+
PAX5 in *GRHL1*	CGCTGGCCGCTGCCGGGCCG	0.774720	5.48255	2116-2135	+
PAX5 in *GRHL1*	GAGGCCTCGGGGCTGGGAGG	0.768389	7.59371	1903-1922	-
PAX5 in *GRHL1*	AGGGCACAGAGGAGGGACTG	0.753453	6.72677	18-37	+
PAX5 in *GRHL1*	GGGTGGCTGGTGGGGCACAG	0.677451	5.17656	816-835	-
PAX5 in *GRHL1*	GAGGCTGCGGGGCACAGACA	0.668394	5.2443	1261-1280	+
PAX5 in *GRHL1*	GGGGCCGCGCTGAGGGGCCG	0.634609	11.5418	2195-2214	+
PAX5 in *GRHL1*	AGGGAGGGGGCGGGGGGAGG	0.569628	9.49677	2521-2540	-
PAX5 in *GRHL1*	GGGCGTCGCGGTGGAGTGG	0.510806	5.60408	1169-1188	+
PAX5 in *GRHL1*	TGGGCCAGGAGAAGGGACAT	0.495928	3.75782	1360-1379	+
PAX5 in *GRHL1*	AGGCTCCCGAGGCCGACCGG	0.460127	3.91608	2543-2562	+
PAX5 in *GRHL1*	GGGCCACGAAGGTGGGACTG	0.457174	4.92714	884-903	-
PAX5 in *GRHL1*	GGCTCTCGGGGGAAGCCACG	0.456123	4.21964	1389-1408	+
PAX5 in *GRHL1*	GAAGACCTGGGGCATAGAAG	0.411535	7.02001	403-422	+
PAX5 in *GRHL1*	AGGGCAATAGAGTGAGGTGG	0.395662	4.13012	579-598	+
PAX5 in *GRHL1*	GGGGAGCCCGCTCCGGACCCG	0.279464	3.32072	2030-2049	+
PAX5 in *GRHL1*	CGGCTGCGGGTCCGGAGCGG	0.244066	3.51957	2036-2055	+
PAX5 in *GRHL1*	CGGCCGCCGCGGGGGGGCCG	0.242143	8.66426	2182-2201	+
PAX5 in *GRHL1*	TGCCTGCGTGGTGGGGACCAG	0.218395	3.45611	628-647	-
PAX5 in *GRHL1*	GGGAGCCTCGGGAGGAGAGA	0.204767	3.51811	2350-2369	+
PAX5 in *GRHL3*	GGAGGAGTGAAGAGGGAAAA	0.980442	8.9417	2386-2405	+
PAX5 in *GRHL3*	GGAGCTTGGGAGACAGGAAGA	0.722296	8.33589	2212-2231	+
PAX5 in *GRHL3*	GGGGGAGGGGGAGAATAGAAA	0.518086	6.62923	873-892	+
PAX5 in *GRHL3*	AGGGCTCTGCAGAGAACAAA	0.431951	4.06235	480-499	+

B

gene: GRHL1 (ENSG00000134317), strand: 1

gene: GRHL2 (ENSG00000083307), strand: 1

gene: GRHL3 (ENSG00000158055), strand: 1

D

KLF4 *GRHL2*

KLF4 *GRHL3*

PAX5 *GRHL1*

PAX5 *GRHL3*

Fig 1. Bioinformatics predictions of TFBS motifs for KLF4 and PAX5 in the promoters of human *GRHL1-3* genes. (**A**) The overall analysis workflow, starting from the selection of the two transcription factors of interest that was based on the literature (PubMed). (**B**) Analysis windows used for multi-alignment, based on promoter features (DNase I-seq, CpG, H3K9ac, H3K4me3) from Ensembl v.79 are marked as blue rectangles within the ± 10 kb flanks of the TSS. Instances of

SwissRegulon TFBS motifs for KLF4 and PAX5 in the promoter taken from the NGD database [27] are also shown. The plots were generated using the webservice of the NGD database [27], accessed with the client PlotGenomic.t2flow, which is provided as S1 File. In this file, all the analysis parameters are provided as example values, viewable and usable upon opening the client in Taverna Workbench [28] http://www.taverna.org.uk/download/workbench/. (**C**) Evolutionarily conserved TFBS motif instances for KLF4 and PAX5 identified using MotEvo [30] with the default parameters. The most conserved motif instances are marked by orange rectangles. (**D**) Multi-alignments around the most conserved instances of motifs KLF4 and PAX5.

https://doi.org/10.1371/journal.pone.0257977.g001

to -379 with respect to the transcription start site, with KLF4 binding site or 40 bp identical fragment without KLF4 binding site), *GRHL2* (60 bp from -672 to -612 or 50 bp identical fragment without KLF4 binding site) and *GRHL3* (60 bp from -1324 to -1264 or 50 bp identical fragment without KLF4 binding site) or synthetic oligonucleotides of regulatory elements of *GRHL1* (60 bp from -180 to -120 or 40 bp identical fragment without PAX5 binding site) and *GRHL3* (60 bp from +3360 to +3420 or 40 bp identical fragment without PAX5 binding site) were cloned into the firefly luciferase vector with SV40 promoter (pGL3-promoter) (Promega). These plasmids were termed KLF4/GRHL1-luc, NoKLF4/GRHL1-luc, KLF4/GRHL2-luc, NoKLF4/GRHL2-luc, KLF4/GRHL3-luc, NoKLF4/GRHL3-luc, PAX5/GRHL1-luc, NoPAX5/GRHL1-luc, PAX5/GRHL3-luc, NoPAX5/GRHL3-luc, respectively. In co-transfection experiment no. 2, using the above mentioned In-Fusion® HD Cloning Kit (Takara Bio) the *GRHL1* promoter fragment (synthetic oligonucleotides 50 bp fragment from -429 to -379 with major or minor frequency allele of SNP rs115898376), were cloned into the firefly luciferase vector with SV40 promoter (pGL3-promoter) (Promega). These plasmids were termed KLF4/GRHL1-luc2 and KLF4/GRHL1-SNP-luc2. All constructs were verified by sequencing. The control vector with *Renilla* luciferase gene (pRL-CMV) was also purchased from Promega. Fragments and primers used for cloning are listed in S1 Table.

Chromatin Immunoprecipitation assays (ChIP)

HEK293 cells cultured in 6-well plates (well diameter 34.8 mm) were transfected with 1.0 μg of pcDNA3.1-HA-KLF4 FL or pcDNA3.1-N-DYK-PAX5 plasmid, using Lipofectamine 3000 (Thermo Fischer Scientific) according to the protocol provided by the manufacturer. After 24 h cells were crosslinked with 1% formaldehyde for 10 min, and subjected to sonication (25 min, 30 s on/30 s off) (Bioruptor, Diagenode) to generate ~500 bp DNA fragments. Chromatin was isolated with the Imprint Chromatin Immunoprecipitation Kit (Sigma-Aldrich), according to the manufacturer's instructions. DNA fragments were immunoprecipitated with 3 μg of anti-KLF4 antibody (ab106629, Abcam) or 12.5 μg of anti-DYK (FLAG) antibody (ab1162, Abcam) and 1 μg of normal rabbit IgG as negative control (10500C, Thermo Fisher Scientific). The immunoprecipitated DNA was analyzed using quantitative real-time PCR (qPCR). To calculate real-time PCR results delta-delta cycle threshold (Ct) method was used. The fold changes related to 10% input delta Ct were calculated as $2^{-\Delta\Delta Ct}$ [34]. For negative control for KLF4 we used the *ZNF333* coding region (primers provided with the Imprint Chromatin Immunoprecipitation Kit). Negative control for PAX5 binding was *KRAS* gene which does not contain PAX5 binding sites [35]. Primers used for ChIP are listed in S2 Table. All ChIP assays were performed in triplicate. Statistical differences were determined by Student's t-test. $P \leq 0.05$ was considered statistically significant.

Nuclear extract preparation and Electrophoretic Mobility Shift Assays (EMSA)

HEK293 cells grown on 100 mm plates were transfected (Lipofectamine 3000, Thermo Fischer Scientific) with 14.1 μg of pcDNA3.1-HA-KLF4 FL or pcDNA3.1-N-DYK-PAX5 plasmids.

After 24 h, nuclear extracts were prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fischer Scientific).

Binding reactions for EMSA were performed using 10 μg of nuclear extract proteins with 100 nM of 20–30 bp double-stranded 5'-fluorescein-labeled oligonucleotides (synthesized by Genomed) containing KLF4 or PAX5 binding site sequences with or without 100-fold molar excess of non-labeled double-stranded cold competitors in EMSA binding buffer: 10 mM Tris-HCl, pH 7.5; 50 mM KCl, 3.5 mM DTT, 0.25% Tween 20, 5 mM $MgCl_2$, 50 ng/μl of poly(dI-dC) and 2.5% glycerol in a final volume of 20 μl. Binding reactions were incubated on ice for 30 minutes, after which protein-DNA complexes were separated by electrophoresis on 5% nondenaturing polyacrylamide 1 x Tris-borate-EDTA (TBE), 10% glycerol gel (120 V, 1.5 h) in 1 x TBE buffer and visualized on a Typhoon FLA 9000 laser scanner. To identify proteins bound to DNA probes, nuclear extracts were incubated with 3 μg anti-KLF4 antibody (ab106629, Abcam) or 1 μg of anti-PAX5 antibody (ab15164, Abcam) on ice, for 30 minutes prior to the addition of fluorescein-labeled DNA probes. The sequences of the oligonucleotides used in EMSA are listed in S3 Table.

Total RNA extraction, reverse transcription and qRT-PCR assays

HEK293 cells cultured in 6-well plates were transfected with 1.0 μg of pcDNA3.1-HA-KLF4 FL or pcDNA3.1-N-DYK-PAX5, or pcDNA3.1-empty plasmids (negative control), using Lipofectamine 3000 (Thermo Fischer Scientific) in accordance with the manufacturer's protocol. After 24 h, total RNA was extracted and purified using Total RNA Mini Plus Kit (A&A Biotechnology) following manufacturer's instructions. The yield of RNA was estimated spectrophotometrically from absorbance at 260 nm, and RNA purity was evaluated according to the A_{260}/A_{280} and A_{260}/A_{230} ratio (NanoDrop ND-1000, Thermo Fisher Scientific). The integrity of total RNA was confirmed by the presence of sharp bands in UV light corresponding to 18S and 28S rRNA when separated by electrophoresis on 1.5% agarose. Next, 1 μg of total RNA was reverse-transcribed into first-strand cDNA with ReadyScript cDNA Synthesis Mix (Sigma-Aldrich). Quantitative real-time PCR was performer in triplicate using TaqMan Fast Universal PCR Master Mix No AmpErase UNG (Thermo Fischer Scientific) and TaqMan Probes (Thermo Fischer Scientific) on a StepOne Plus Real Time PCR system (Applied Biosystems). Gene expression levels were normalized to hypoxanthine phosphoribosyltransferase 1 housekeeping gene (*HPRT1*) (the highest stability based on literature [36]) and calculated using the $2^{-\Delta\Delta Ct}$ method [34]. The following TaqMan Probes were used in experiments: Hs01119372_m1 for *GRHL1*, Hs00227745_m1 for *GRHL2*, Hs00297962_m1 for *GRHL3*, Hs02800695_m1 for *HPRT1*. Statistical differences for relative expression levels were determined using Student's t-test. $P \leq 0.05$ was considered statistically significant.

Reporter gene assays

HEK293 cells cultured in 6-well plates (well diameter 34.8 mm) were transfected (Lipofectamine 3000, Thermo Fischer Scientific) with 500 ng of pcDNA3.1-HA-KLF4 FL or pcDNA3.1-N-DYK-PAX5, or pcDNA3.1-empty plasmids, 25 ng pRL-CMV and 500 ng of the firefly luciferase vector with KLF4 or PAX5 binding sites (described in section *Plasmids*). Using the Dual-Luciferase Reporter Assay System (Promega), after 24 h, cells were lysed and luciferase activity was measured using Tecan Infinite M1000 PRO luminometer. Relative reporter activity was calculated and normalized based on *Renilla* luciferase activity. All experiments were carried out in triplicate. Statistical evaluations were performed using Student's t-test. $P \leq 0.05$ was considered statistically significant.

Results

Bioinformatics analyses of potential binding sites for KLF4 and PAX5 in the regulatory regions of the *GRHL1-3* genes

As detailed in the Introduction, the hypothesis that KLF4 and PAX5 may be biologically interesting transcription factors–regulators of expression of genes from the *GRHL* family was based on our previous work and analyses of scientific literature. Working with this hypothesis, we set ourselves a goal of bioinformatically identifying the most likely TFBSs for these two transcription factors in the vicinity of *GRHL1-3* gene promoters. The flow of our analysis is illustrated in Fig 1A. To identify the promoter regions of *GRHL* genes, we utilized transcription start sites established by the FANTOM consortium [37] and chromatin data (openness, histone methylation and acetylation patterns) from the ENCODE project [38], available via the NGD database [27]. From the NGD we also obtained the genomic coordinates of the instances of TFBS motifs for KLF4 and PAX5 (Fig 1B). In order to identify the most likely binding sites for either of these transcription factors, among the sometimes large numbers of instances of a particular TFBS motif in a given promoter, we employed phylogenetic footprinting with the MotEvo program [30], following multi-alignments of the regions orthologous to the analyzed promoter windows in 7 mammalian species with T-coffee [29]. MotEvo employs Bayesian reasoning and returns a posterior probability that a given TFBS motif's instance is functional in the chosen species (here–human), given the motif strength in that species (Weight Matrix score) and its evolutionary conservation.

These bioinformatic analyses allowed us to identify 8 evolutionarily conserved putative binding sites for KLF4 in the regulatory regions of *GRHL1-3* genes and 29 evolutionarily conserved putative binding sites for PAX5 in the regulatory regions of *GRHL1* and *GRHL3* genes (Fig 1C and S4 Table). We did not find any evolutionarily conserved putative binding sites for PAX5 in the promoter region of *GRHL2*, and for this reason we excluded the PAX5/GRHL2 pair from further investigations. For experimental analyses for each *GRHL* gene we selected one predicted binding site for KLF4 or PAX5 (with the highest posterior probability (P aposteriori)), marked by orange rectangles in Fig 1C. The corresponding multi-alignments, matched to the logos of the respective TFBS motifs are shown in Fig 1D. We note that the selected predicted binding site for PAX5 in *GRHL3* is conserved in human and macaque only. All the potential binding sites for KLF4 and PAX5, together with their genomic coordinates and the human genome coordinates of the multi-aligned windows, are listed in S4 Table and S2 File. Moreover, for KLF4, we chose the previously identified KLF4 motif site with the SNP (C/T) (rs115898376) [22].

KLF4 and PAX5 bind to the regulatory regions of the Grainyhead-like genes

To determine whether the bioinformatically predicted binding sites in the regulatory regions of *GRHL* genes are direct targets for KLF4 and PAX5, ChIP analyses were carried out. Chromatin from HEK293 cells transfected with either KLF4 or DYK-tagged PAX5 expressing vectors was immunoprecipitated with anti-KLF4 or anti-FLAG antibody, or normal IgG, and then examined by qPCR using primers listed in S2 Table. As a result, we found that the investigated DNA fragments were significantly enriched by anti-KLF4 antibody in comparison with nonspecific IgG, indicating KLF4 binding to the sequences of *GRHL1*, *GRHL2* and *GRHL3* promoters (Fig 2A). The data also showed a strong enrichment of the qPCR signal with anti-FLAG antibody in comparison with normal IgG, demonstrating that PAX5 binds to the *GRHL1* promoter sequence and also to the enhancer sequence located in the first intron of
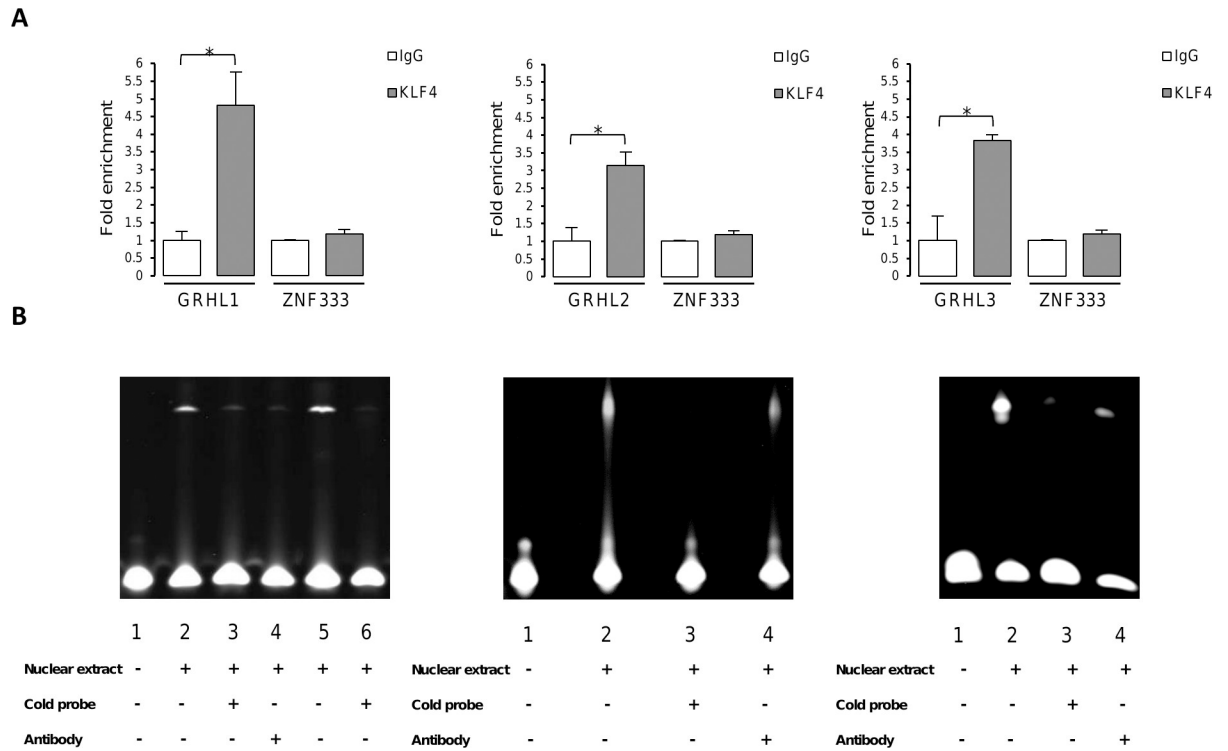
A

B

Fig 2. KLF4 binds to the regulatory regions of *GRHL* genes. (**A**) Quantitative ChIP-PCR analysis of KLF4 occupancy of *GRHL1*, *GRHL2* and *GRHL3* regulatory regions was performed in HEK293 cells transfected with pcDNA3.1-HA-KLF4 FL. *ZNF333* was used to identify non-specific interactions. Chromatin was immunoprecipitated with anti-KLF4 antibody or nonspecific antibody. The amount of DNA amplified from immunoprecipitated DNA was normalized to that amplified from input DNA. Data are shown as means ± SEM from experiments independently performed in triplicate, *significantly different at p≤ 0.05. (**B**) EMSA analysis performed with probes including KLF4 binding sequences: left panel–in the region -409/-400 (lane 2) or -409/-400 with the minor frequency allele of SNP rs115898376 (C/T) (lane 5) of the *GRHL1* promoter; central panel–in the region -650/-641 (lane 2) of the *GRHL2* promoter; right panel–in the region -1302/-1293 (lane 2) of the *GRHL3* promoter. Lane 1: probe only. Cold probe: unlabeled double-stranded oligonucleotides including -409/-400 (left panel, lane 3) or -409/-400 with SNP rs115898376 (C/T) (left panel, lane 6) or -650/-641 (central panel, lane 3) or -1302/-1293 (right panel, lane 3) regions of *GRHL* genes (100-fold molar excess of competitors). Where indicated, 3 μg anti-KLF4 antibody was added per lane (ab106629, Abcam).

GRHL3 gene (Fig 3A). We found no enrichment of *ZNF333* or *KRAS* (negative controls for KLF4 or PAX5, respectively) qPCR signals, indicating that KLF4 and PAX5 binding was specific (Figs 2A and 3A).

Our ChIP results have shown that both KLF4 and PAX5 transcription factors bind to the predicted regulatory regions of *GRHL* genes. In order to confirm KLF4 and PAX5 binding site sequences (selected on the basis of bioinformatics analyses) in the *GRHL* regulatory elements, we designed a number of fluorescein-labeled oligonucleotide probes for EMSA experiments. We performed EMSA with nuclear extracts from HEK293 cells following their transfection with either KLF4 or DYK-tagged PAX5 expressing vectors. We discovered that KLF4 binds to oligonucleotide probes containing binding sites: -409/-400 of *GRHL1*, -650/-641 of *GRHL2* and -1302/-1293 of *GRHL3* promoters, as shown by the representative results in Fig 2B. Cold competitor probes containing KLF4 binding sites effectively competed with respective fluorescein-labeled probes. Interestingly, the -409/-400 fragment with the minor frequency allele of SNP rs115898376 (C/T) was also bound by KLF4. Anti-KLF4 antibody completely abolished the binding, indicating that the protein-DNA interactions were specific. Similarly, PAX5 interacted with labeled probes containing fragments -147/-128 of *GRHL1* promoter and +3386/+3405 of *GRHL3* enhancer. Results are shown in Fig 3B. A 100-fold molar excess of identical
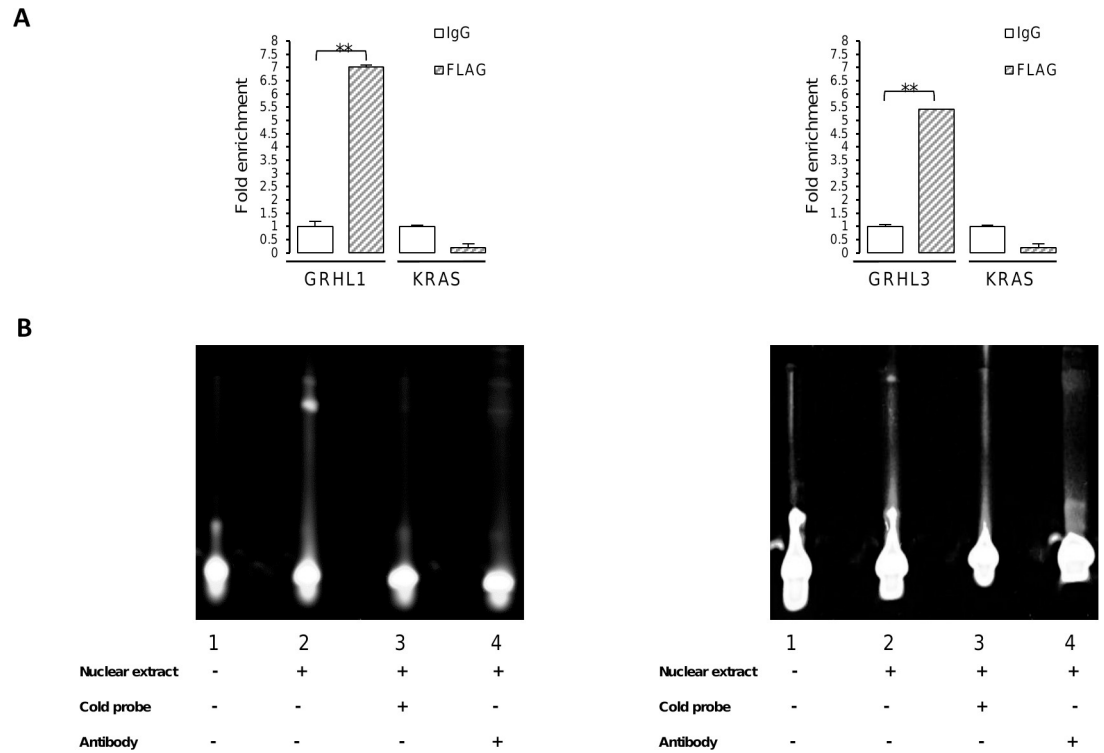
Fig 3. PAX5 binds to the regulatory regions of *GRHL* genes. (A) Quantitative ChIP-PCR analysis of PAX5 occupancy of the *GRHL1* and *GRHL3* regulatory regions was performed in HEK293 cells transfected with pcDNA3.1-N-DYK-PAX5. *KRAS* was used to identify non-specific interactions. Chromatin was immunoprecipitated with anti-DYK (FLAG) antibody or nonspecific antibody. The amount of DNA amplified from immunoprecipitated DNA was normalized to that amplified from input DNA. Data are shown as means ± SEM of experiments independently performed in triplicate, **significantly different at $p \leq 0.01$. **(B)** EMSA analysis performed with probes including PAX5 binding sequences: left panel–in the region -147/-128 (lane 2) of the *GRHL1* promoter; right panel–in the region +3386/+3405 (lane 2) of the *GRHL3* enhancer. Lane 1: probe only. Cold probe: unlabeled double-stranded oligonucleotides including -147/-128 (left panel, lane 3) or +3386/-3405 (right panel, lane 3) regions of *GRHL* genes (100-fold molar excess of competitors). Where indicated, 1 µg anti-PAX5 antibody was added per lane (ab15164, Abcam).

but unlabeled probes resulted in a decrease of fluorescence signal, and 1 µg of anti-PAX5 antibody eliminated protein-DNA complexes, showing that PAX5 binding was specific. Quantitative results of EMSA experiments are presented in S1 Fig.

Ectopic expression of KLF4 or PAX5 alters the expression of *GRHL* genes

By regulating the expression of its target genes, KLF4 may function as a suppressor or an oncogene in a context-dependent manner [39]. Contrary to its oncogenic effects in B cell lymphomas, PAX5 seems to play a role of an anti-proliferative protein in many malignancies [40–42]. However, it was still unknown whether these transcription factors regulate the expression of *GRHL* genes. For this purpose, HEK293 cells were transiently transfected with either KLF4 or PAX5 expressing plasmids, and the control empty vector. Next, using qRT-PCR we investigated the levels of expression of *GRHL1-3* genes. In KLF4-overexpressing HEK293 cells, the expression of *GRHL1* and *GRHL3* genes was upregulated by 32% or 60%, respectively, whereas the mRNA level of *GRHL2* gene was reduced by 28% when compared to the respective control cells (Fig 4A). We also observed a statistically significant decrease by 30% or 33% in the *GRHL1* and *GRHL3* mRNA levels in PAX5-overexpressing HEK293 cells (Fig 4B).
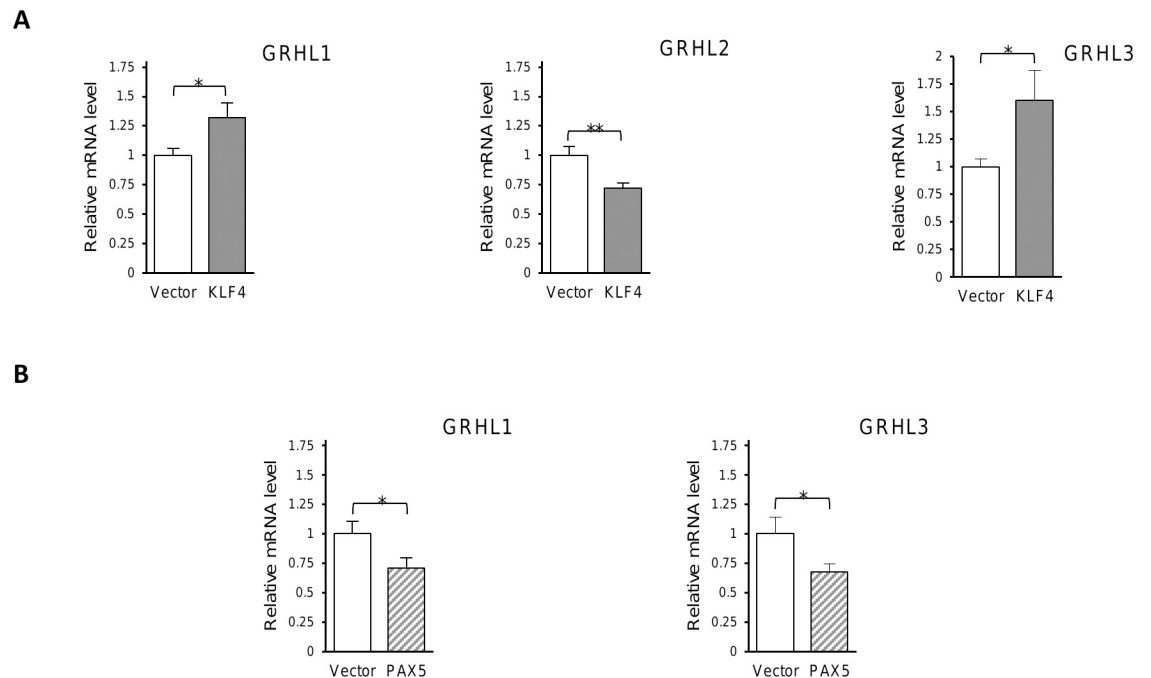
A

B

Fig 4. Overexpression of KLF4 or PAX5 alters mRNA levels of *GRHL* genes. (**A**-**B**) The mRNA expression levels of *GRHL1*, *GRHL2* and *GRHL3* genes in HEK293 cells transiently overexpressing KLF4 (**A**) or PAX5 (**B**). The results represent relative expression of the respective target gene vs *HPRT* gene. Data are shown as means ± SEM of experiments independently performed in triplicate, *significantly different at p≤ 0.05, **significantly different at p≤ 0.01.

https://doi.org/10.1371/journal.pone.0257977.g004

KLF4 and PAX5 change the activity of *GRHL* regulatory regions

We performed co-transfection studies using fragments of regulatory regions of *GRHL* genes fused to the reporter gene–luciferase. In experiment no. 1 we transiently transfected HEK293 cells with reporter constructs containing *Luc* gene under the control of DNA fragments in which the KLF4 and PAX5 binding site sequences were either present or deleted. The results obtained indicate that in cells co-transfected with the vector including the KLF4 binding site, the expression of *Luc* gene fused to *GRHL1* promoter fragment was efficiently repressed (the *Luc* enzyme activity was downregulated by 39%). The *Luc* enzyme activity was also downregulated by 33% under the control of *GRHL2* promoter fragment and upregulated by 125% under the control of *GRHL3* promoter fragment (Fig 5A). Analogous experiment showed a decrease in luciferase activity from plasmid with *GRHL1* promoter (by 43%) and an increase in the case of a plasmid containing a *GRHL3* enhancer element (by 157%) in HEK293 cells co-transfected with the vector coding for PAX5 (Fig 6A).

In experiment no. 2 we transiently transfected HEK293 cells with reporter constructs with *Luc* gene under the control of DNA fragments containing KLF4 binding site sequences in the *GRHL1* promoter, with either the major frequency allele of SNP rs115898376 or the minor frequency allele of this SNP. Our results showed that, in cells co-transfected with the KLF4 expressing vector, the expression of *Luc* gene fused to *GRHL1* promoter fragment was efficiently repressed (the *Luc* enzyme activity was downregulated by 44%), moreover the presence of minor frequency allele of SNP rs115898376 in the cloned fragment of *GRHL1* gene resulted in further reduced expression (the *Luc* enzyme activity downregulated by further 20%) in comparison with the same region with the major frequency allele of this SNP (Fig 7).
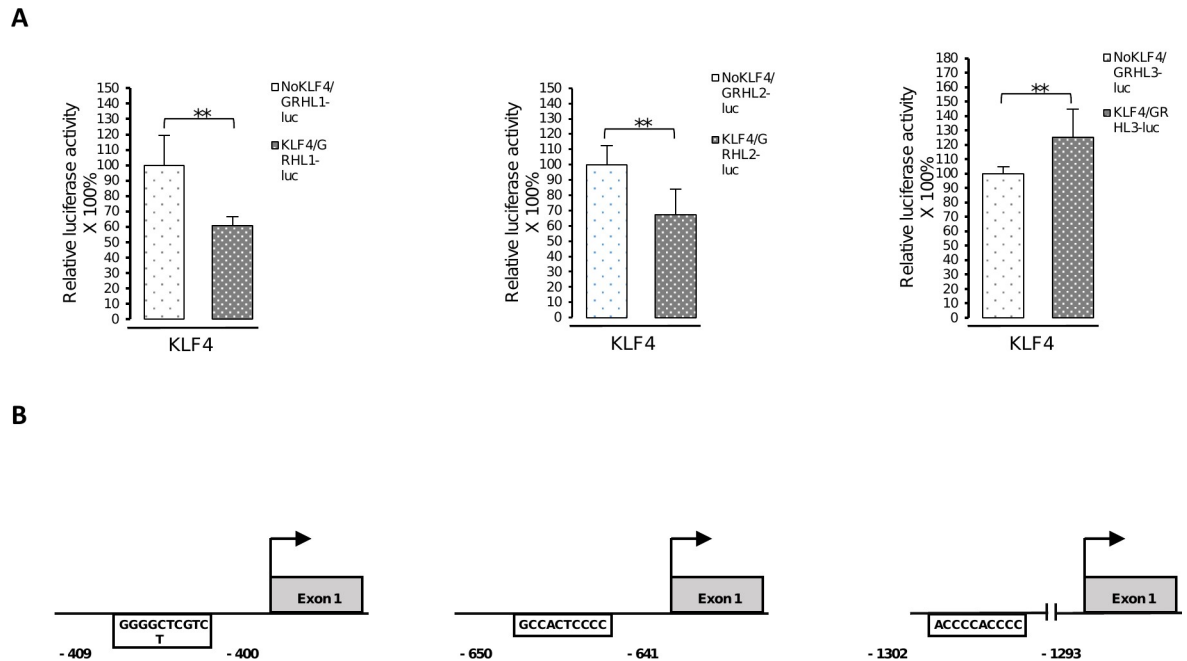
A

B

Fig 5. KLF4 changes the activity of *GRHL* regulatory regions. (**A**) HEK293 cells were transfected with pcDNA3.1-KLF4 plasmid, pGL3-promoter vector containing the luciferase gene under the control of the fragment of the *GRHL1* or *GRHL2* or *GRHL3* regulatory sequence with or without binding sites for KLF4 (named: KLF4/GRHL1-3-luc or NoKLF4/GRHL1-2-luc) and pRL-CMV *Renilla* luciferase control reporter vector. Data are shown as means ± SEM of experiments independently performed in triplicate, *significantly different at p≤ 0.05, **significantly different at p≤ 0.01. (**B**) Schematic representation of locations of the KLF4 binding site sequences in the promoters of *GRHL1*, *GRHL2* and *GRHL3* genes cloned into luciferase vectors.

Discussion

The aim of our research was to identify additional transcription factors regulating the expression of genes from the *GRHL* family, because the precise regulation of the expression of *GRHL* genes is essential for their function and affects the functioning of encoded proteins. The role of GRHL transcription factors is well defined and is associated with the regulation of expression of genes involved in epithelial proliferation and differentiation [7,43]. Imbalance between these processes is frequently observed in a variety of diseases including atopic dermatitis, psoriasis, idiopathic pulmonary fibrosis and cancer [14,44,45]. To date, only fragmentary studies investigated the regulation of expression of *GRHL* genes by transcription factors. For this reason in the present work we employed comprehensive literature and bioinformatics analyses aimed at predicting which transcription factors may regulate *GRHL* expression. The completion of above analyses allowed us to identify KLF4 and PAX5 as putative regulators of *GRHL* expression. Our results show that KLF4 binds to the promoters of *GRHL1-3*, while PAX5 interacts with *GRHL1* and *GRHL3* regulatory regions. While most of these TF-binding results are novel, by comparison with the ChIP-seq data from the Cistrome database [32] we found out the PAX5 binding site that we focused on in *GRHL1* intersects a peak of PAX5-binding identified in the ChIP-seq experiment performed in lymphoblastoid B-cells (S2 File) [46]. We discovered that overexpression of KLF4 and PAX5 transcription factors in HEK293 cells have opposite effects on the mRNA levels of the *GRHL* genes. KLF4 positively regulates *GRHL1* and *GRHL3* expression and negatively regulates the *GRHL2* gene, while PAX5 has the ability to decrease *GRHL1* and *GRHL3* mRNA levels. Earlier studies revealed that the development of many types of cancer is often accompanied by changes in the levels of expression of the genes
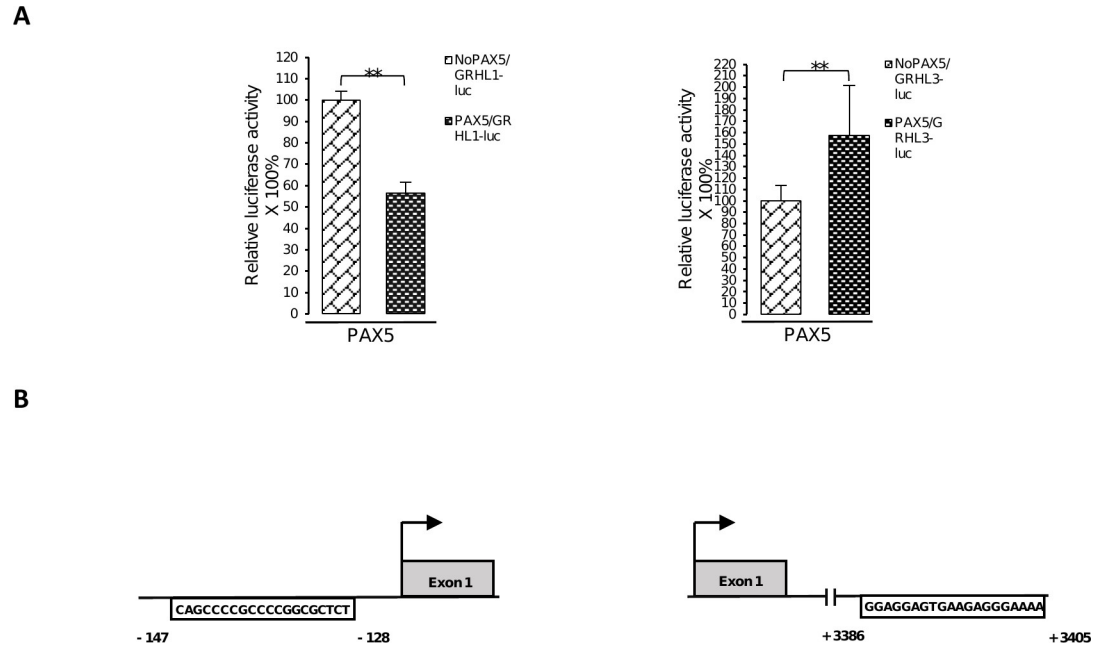
A

B

Fig 6. PAX5 changes the activity of *GRHL* regulatory regions. (**A**) HEK293 cells were transfected with pcDNA3.1-N-DYK-PAX5 plasmid, pGL3-promoter vector containing the luciferase gene under the control of the fragment of the *GRHL1* or *GRHL3* regulatory sequence with or without binding sites for PAX5 (named: PAX5/GRHL1-luc or PAX5/ GRHL3-luc or NoPAX5/GRHL1-luc or NoPAX5/GRHL3-luc and pRL-CMV *Renilla* luciferase control reporter vector. Data are shown as means ± SEM of experiments independently performed in triplicate, **significantly different at p≤ 0.01. (**B**) Schematic representation of locations of the PAX5 binding site sequences in the promoter of *GRHL1* gene and in the enhancer of *GRHL3* gene cloned into luciferase vectors.

from the *GRHL* family [12–14]. Our previous work showed that the *GRHL1* and *GRHL3* expression is downregulated in the samples of human NMSC [21]. Also, our results indicated that the *Grhl1*-null mice display increased susceptibility to chemically-induced skin tumorigenesis [47] and deletion of *Grhl3* in the epidermis increases the occurrence of squamous cell carcinoma (SCC) of the skin [48]. Moreover, silencing of *GRHL2* expression in non-tumorigenic kidney cell line resulted in increased cell proliferation and increased resistance to apoptosis [22]. The above findings indicate that *GRHL1*, *GRHL2* and *GRHL3* serve protective role against NMSC and ccRCC. KLF4 plays a pivotal role in regulating various cellular processes including proliferation, differentiation, development, maintenance of homeostasis and apoptosis [49,50]. This transcription factor can function either as a tumor suppressor or as an oncogene depending on different cellular context [39]. Some studies showed that induction of KLF4 in basal keratinocytes initiates squamous epithelial dysplasia [51]. On the other hand, KLF4 is a tumor suppressor in a UVB-induced mouse skin tumor model [52] and its expression is decreased in samples from patients with SCC and basal cell carcinoma of the skin [53]. Additionally, knockdown of KLF4 in human epidermal squamous cell carcinoma SCC13 cell line was associated with increased cell growth, migration and adhesion [53]. Other studies revealed that KLF4 expression was lower in ccRCC tumors than in patient normal control samples both at the protein and mRNA levels, while the KLF4 overexpression arrested the cell cycle progress in ccRCC cell lines [54]. Given the above, KLF4 has an important function in tumorigenesis and it is possible that its role as a tumor suppressor in SCC and ccRCC is related to its ability to regulate the expression of all three *GRHL* genes. Furthermore, *Klf4*-null mice die immediately after birth due to the lack of an impermeable skin barrier [19], and the same
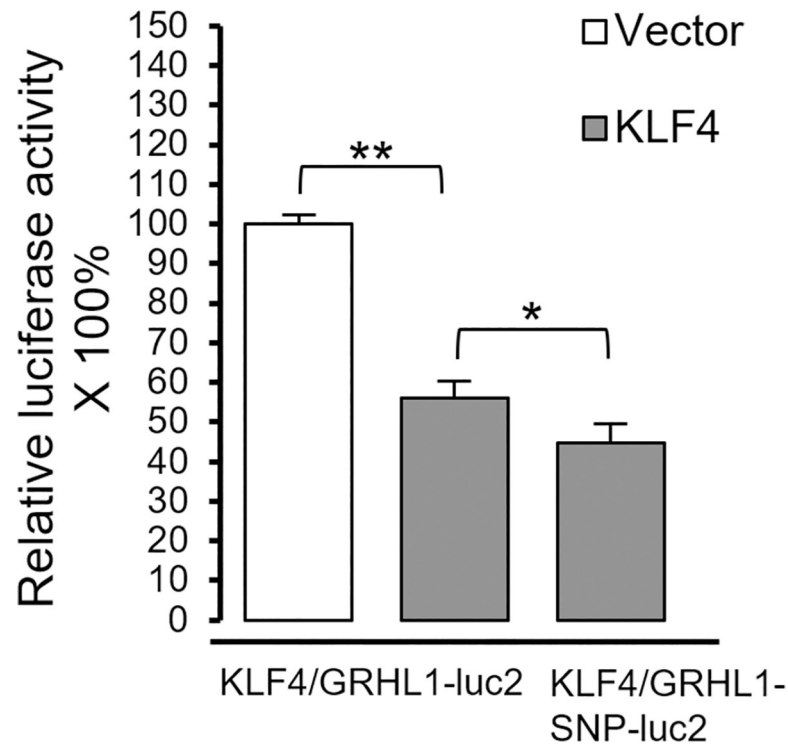
Fig 7. Impact of the presence of the minor frequency allele of SNP rs115898376 in the promoter of *GRHL1* gene on KLF4 binding. HEK293 cells were transfected with pcDNA3.1-KLF4 or pcDNA3.1-empty plasmid, pGL3-promoter vector containing the luciferase gene under the control of the fragment of the *GRHL1* regulatory sequence with either the major or the minor frequency allele of SNP rs115898376 (named KLF4/GRHL-luc2 and KLF4/GRHL1-SNP-luc2), and pRL-CMV *Renilla* luciferase control reporter vector. The presence of the minor frequency allele of SNP rs115898376 in the promoter of *GRHL1* gene alters the functioning of the KLF4 binding site.

https://doi.org/10.1371/journal.pone.0257977.g007

phenotype is observed in the *Grhl3*-null mice [8]. Although all KLF factors recognize related sequences, KLF4 is the only KLF with an expression pattern specific to skin and gut [55]. The expression of KLF4 is unchanged in the *Grhl3*-null mice [8] which suggests that KLF4 is likely to regulate the expression of *Grhl3* in this context as well.

In this study we discovered that PAX5 negatively regulates the expression of *GRHL1* and *GRHL3*. PAX5 is expressed in aggressive N-type neuroblastoma cell lines, and downregulation of this transcription factor significantly reduced their proliferation rate and tumorigenic phenotype [56]. Moreover, PAX5 plays an important role during oral carcinogenesis where increase in PAX5 expression (at both mRNA and protein levels) was observed in oral SCC cell lines, compared to human normal oral keratinocytes [57]. We can assume that the oncogenic effects of PAX5 may be associated with its negative regulation of expression of *GRHL1* and *GRHL3*, since *GRHL1* has been shown to be a tumor suppressor in neuroblastoma [58], whereas *GRHL3* is a tumor suppressor in head and neck SCC [59].

In the present project we also attempted to investigate some of the functional binding sites for KLF4 and PAX5 in the regulatory regions of *GRHL* genes (Figs 5–7). Our results prove that all of the investigated DNA fragments respond to the regulation by KLF4 or PAX5, as predicted by us. All the changes in luciferase activity are statistically significant. However, the direction of these changes is not always consistent with the direction of changes of expression of endogenous *GRHL* genes upon overexpression of KLF4 or PAX5. The expression of *GRHL1* is increased upon ectopic overexpression of KLF4 (Fig 4A) but the reporter gene assays of one KLF4 binding site suggest that the opposite should be true (Figs 5A and 7). Similar discrepancy

can be observed regarding the regulation of *GRHL3* by PAX5 (Figs 4B and 6A). These apparent contradictions can be explained by the findings that, at least in some cases, variation in the number of binding sites for a single transcription factor is sufficient to encode activation versus repression of the regulated gene [60]. In the promoter of the *GRHL1* gene we identified two putative binding sites for KLF4, while in the promoter of the *GRHL3* gene we identified four putative binding sites for PAX5 (Fig 1C and S4 Table). If simple rearrangements of TFBSs can encode qualitatively different responses to a single transcription factor [60] then it is not surprising that the effect of only one TFBS investigated in reporter assays can be opposite to the effect of an arrangement of a number of TFBSs in the endogenous locus. Another possible explanation for the above-mentioned discrepancies can be provided by "squelching" [61,62]. Overexpression of a transcriptional transactivator can sometimes repress the transcription of its target genes. This phenomenon is likely to be caused by the titration of limiting transcriptional coactivators [62]. "Squelching" in mammalian cells is limited to episomal target genes (such as our reporter gene constructs), but it does not affect the target genes present on cellular chromosomes [61], which is why the direction of expression changes shown in Fig 4 is likely to be correct. One more possible explanation is that the outcome of the transcription factor binding in the promoter of an endogenous gene is modified by interaction with enhancer(s), which are not reflected by the reporter construct. Nevertheless, our results prove that KLF4 and PAX5 bind to these sites, and that this binding has a regulatory effect.

We are cognizant of other shortcomings of overexpression experiments. Unbalancing gene dosage can lead to various artifacts and systemic errors, hence the results of overexpression experiments should be verified at native expression and/or *in vitro* [62]. To address this concern, we confirmed our findings by EMSA which is an *in vitro* assay based on a different principle (Figs 2 and 3). Our experiments were repeated only in triplicate, which is why our work is of preliminary nature. Nevertheless, our experimental results are consistent with the outcome of bioinformatic studies and literature analyses discussed above, which adds weight to the validity of our research. The observed changes in gene expression are relatively small (Figs 4–7). This can be explained by the fact that overexpression experiments are inherently quantitative and not qualitative [63]. Substantial fractions of overexpressed transcription factors can be present in inactive forms as they often require posttranslational modifications to fulfil their functions [63,64]. Future experiments will require animal models and samples obtained from human patients to determine whether the regulatory events studied by us are relevant in biology and medicine.

The results from our previous studies indicated that the minor frequency allele of SNP rs115898376, located in the promoter region of *GRHL1* gene, occurs at significantly altered frequencies in patients with ccRCC and the presence of this polymorphism may alter KLF4 binding [22]. In this work we confirmed this hypothesis to be true. We cloned 50 bp fragment of *GRHL1* promoter, with two putative KLF4 binding sites containing an either major or minor frequency allele of this SNP, into the plasmid with reporter luciferase gene. We found that cotransfection with KLF4 expressing plasmid decreased luciferase activity, moreover luciferase level was lower when using reporter vector containing the minor frequency allele of SNP indicating that the presence of this allele affected the binding of KLF4 to cloned fragment (Fig 7). However, the presence of the minor frequency allele of this SNP does not completely abolish KLF4 binding as in EMSA assays with probes containing major or minor frequency allele of SNP rs115898376 we found no difference in KLF4 binding ability (Fig 2B).

In conclusion, our results reveal that both KLF4 and PAX5, by binding to the regulatory regions of *GRHL* genes. are able to regulate the expression of these genes. Moreover, the presence of different alleles of SNP rs115898376 in the promoter region of the *GRHL1* gene alters KLF4 binding and influences its ability to regulate *GRHL1* transcription.

Supporting information

S1 Fig. Quantitative results of EMSA experiments performed with probes including KLF4 (**A**) or PAX5 (**B**) binding sequences. The relative decrease in signals for cold probe or probe with KLF4 or PAX5 antibody compared with probes with nuclear extract (100%) was measured by densitometry (ImageJ). Data are shown as means ± SEM of two or three experiments. * one experiment.
(PDF)

S1 Table. List of oligonucleotides used for cloning.
(DOC)

S2 Table. List of ChIP qRT-PCR primers.
(DOC)

S3 Table. List of EMSA oligonucleotide probes.
(DOC)

S4 Table. List of identified putative binding sites for KLF4 and PAX5 in the regulatory regions of *GRHL1-3* genes.
(DOC)

S1 File. Taverna Workbench workflow (a NGD webservice client) used to prepare Fig 1B.
(T2FLOW)

S2 File. Genomic coordinates of binding sites for KLF4 and PAX5 in the promoter regions of *GRHL* genes.
(XLS)

S1 Raw images.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Michal Dabrowski, Tomasz Wilanowski.

Data curation: Michal Miller, Michal Dabrowski.

Formal analysis: Grzegorz Kotarba, Agnieszka Taracha-Wisniewska, Michal Miller, Michal Dabrowski.

Funding acquisition: Tomasz Wilanowski.

Investigation: Grzegorz Kotarba, Agnieszka Taracha-Wisniewska, Michal Miller, Michal Dabrowski.

Methodology: Michal Dabrowski, Tomasz Wilanowski.

Project administration: Tomasz Wilanowski.

Resources: Michal Dabrowski, Tomasz Wilanowski.

Software: Michal Dabrowski.

Supervision: Michal Dabrowski, Tomasz Wilanowski.

Validation: Grzegorz Kotarba, Agnieszka Taracha-Wisniewska, Michal Miller.

Visualization: Grzegorz Kotarba, Agnieszka Taracha-Wisniewska, Michal Miller, Michal Dabrowski.

Writing – original draft: Grzegorz Kotarba, Michal Dabrowski.

Writing – review & editing: Agnieszka Taracha-Wisniewska, Tomasz Wilanowski.

References

1. Traylor-Knowles N, Hansen U, Dubuc TQ, Martindale MQ, Kaufman L, Finnerty JR. The evolutionary diversification of LSF and Grainyhead transcription factors preceded the radiation of basal animal lineages. BMC Evol Biol. 2010; 10:101. Epub 2010/04/20. https://doi.org/10.1186/1471-2148-10-101 1471-2148-10-101 [pii]. PMID: 20398424; PubMed Central PMCID: PMC2873413.

2. Wilanowski T, Tuckfield A, Cerruti L, O'Connell S, Saint R, Parekh V, et al. A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead. Mech Dev. 2002; 114(1–2):37–50. Epub 2002/08/15. https://doi.org/10.1016/s0925-4773(02)00046-1 PMID: 12175488.

3. Ting SB, Wilanowski T, Cerruti L, Zhao LL, Cunningham JM, Jane SM. The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. The Biochemical journal. 2003; 370(Pt 3):953–62. Epub 2003/01/29. https://doi.org/10.1042/BJ20021476 PMID: 12549979; PubMed Central PMCID: PMC1223248.

4. Kotarba G, Krzywinska E, Grabowska AI, Taracha A, Wilanowski T. TFCP2/TFCP2L1/UBP1 transcription factors in cancer. Cancer Lett. 2018; 420:72–9. Epub 2018/02/08. https://doi.org/10.1016/j.canlet.2018.01.078 PMID: 29410248.

5. Taracha A, Kotarba G, Wilanowski T. Neglected Functions of TFCP2/TFCP2L1/UBP1 Transcription Factors May Offer Valuable Insights into Their Mechanisms of Action. Int J Mol Sci. 2018; 19(10):2852. Epub 2018/09/23. https://doi.org/10.3390/ijms19102852 PMID: 30241344; PubMed Central PMCID: PMC6213935.

6. Auden A, Caddy J, Wilanowski T, Ting SB, Cunningham JM, Jane SM. Spatial and temporal expression of the Grainyhead-like transcription factor family during murine development. Gene Expr Patterns. 2006; 6(8):964–70. Epub 2006/07/13. https://doi.org/10.1016/j.modgep.2006.03.011 PMID: 16831572.

7. Wilanowski T, Caddy J, Ting SB, Hislop NR, Cerruti L, Auden A, et al. Perturbed desmosomal cadherin expression in grainy head-like 1-null mice. The EMBO journal. 2008; 27(6):886–97. Epub 2008/02/22. https://doi.org/10.1038/emboj.2008.24 PMID: 18288204; PubMed Central PMCID: PMC2274933.

8. Ting SB, Caddy J, Hislop N, Wilanowski T, Auden A, Zhao LL, et al. A homolog of Drosophila grainy head is essential for epidermal integrity in mice. Science. 2005; 308(5720):411–3. Epub 2005/04/16. https://doi.org/10.1126/science.1107511 PMID: 15831758.

9. Rifat Y, Parekh V, Wilanowski T, Hislop NR, Auden A, Ting SB, et al. Regional neural tube closure defined by the Grainy head-like transcription factors. Developmental biology. 2010; 345(2):237–45. https://doi.org/10.1016/j.ydbio.2010.07.017 PMID: 20654612.

10. Ting SB, Wilanowski T, Auden A, Hall M, Voss AK, Thomas T, et al. Inositol- and folate-resistant neural tube defects in mice lacking the epithelial-specific factor Grhl-3. Nature medicine. 2003; 9(12):1513–9. Epub 2003/11/11. https://doi.org/10.1038/nm961 PMID: 14608380.

11. Caddy J, Wilanowski T, Darido C, Dworkin S, Ting SB, Zhao Q, et al. Epidermal wound repair is regulated by the planar cell polarity signaling pathway. Dev Cell. 2010; 19(1):138–47. Epub 2010/07/21. https://doi.org/10.1016/j.devcel.2010.06.008 PMID: 20643356; PubMed Central PMCID: PMC2965174.

12. Frisch SM, Farris JC, Pifer PM. Roles of Grainyhead-like transcription factors in cancer. Oncogene. 2017; 36(44):6067–73. Epub 2017/07/18. https://doi.org/10.1038/onc.2017.178 PMID: 28714958.

13. Kotarba G, Taracha-Wisniewska A, Wilanowski T. Grainyhead-like transcription factors in cancer—Focus on recent developments. Experimental biology and medicine (Maywood, NJ). 2020; 245(5):402–10. Epub 2020/02/06. https://doi.org/10.1177/1535370220903009 PMID: 32008358; PubMed Central PMCID: PMC7082887.

14. Mlacki M, Kikulska A, Krzywinska E, Pawlak M, Wilanowski T. Recent discoveries concerning the involvement of transcription factors from the Grainyhead-like family in cancer. Experimental biology and medicine. 2015; 240(11):1396–401. Epub 2015/06/13. https://doi.org/10.1177/1535370215588924 PMID: 26069269; PubMed Central PMCID: PMC4935299.

15. Jaubert J, Cheng J, Segre JA. Ectopic expression of kruppel like factor 4 (Klf4) accelerates formation of the epidermal permeability barrier. Development. 2003; 130(12):2767–77. Epub 2003/05/09. https://doi.org/10.1242/dev.00477 PMID: 12736219.

16. Li J, Zheng H, Yu F, Yu T, Liu C, Huang S, et al. Deficiency of the Kruppel-like factor KLF4 correlates with increased cell proliferation and enhanced skin tumorigenesis. Carcinogenesis. 2012; 33(6):1239–46. Epub 2012/04/12. https://doi.org/10.1093/carcin/bgs143 PMID: 22491752; PubMed Central PMCID: PMC3388492.

17. Sen GL, Boxer LD, Webster DE, Bussat RT, Qu K, Zarnegar BJ, et al. ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. Dev Cell. 2012; 22(3):669–77. Epub 2012/03/01. https://doi.org/10.1016/j.devcel.2011.12.001 PMID: 22364861; PubMed Central PMCID: PMC3306457.

18. Ting SB, Caddy J, Wilanowski T, Auden A, Cunningham JM, Elias PM, et al. The epidermis of grhl3-null mice displays altered lipid processing and cellular hyperproliferation. Organogenesis. 2005; 2(2):33–5. Epub 2005/04/01. https://doi.org/10.4161/org.2.2.2167 PMID: 19521564; PubMed Central PMCID: PMC2634083.

19. Segre JA, Bauer C, Fuchs E. Klf4 is a transcription factor required for establishing the barrier function of the skin. Nat Genet. 1999; 22(4):356–60. Epub 1999/08/04. https://doi.org/10.1038/11926 PMID: 10431239.

20. Ou L, Shi Y, Dong W, Liu C, Schmidt TJ, Nagarkatti P, et al. Kruppel-like factor KLF4 facilitates cutaneous wound healing by promoting fibrocyte generation from myeloid-derived suppressor cells. The Journal of investigative dermatology. 2015; 135(5):1425–34. Epub 2015/01/13. https://doi.org/10.1038/jid.2015.3 PMID: 25581502; PubMed Central PMCID: PMC4402119.

21. Kikulska A, Rausch T, Krzywinska E, Pawlak M, Wilczynski B, Benes V, et al. Coordinated expression and genetic polymorphisms in Grainyhead-like genes in human non-melanoma skin cancers. BMC Cancer. 2018; 18(1):23. Epub 2018/01/06. https://doi.org/10.1186/s12885-017-3943-8 PMID: 29301499; PubMed Central PMCID: PMC5755140.

22. Pawlak M, Kikulska A, Wrzesinski T, Rausch T, Kwias Z, Wilczynski B, et al. Potential protective role of Grainyhead-like genes in the development of clear cell renal cell carcinoma. Molecular carcinogenesis. 2017; 56(11):2414–23. Epub 2017/05/26. https://doi.org/10.1002/mc.22682 PMID: 28543713.

23. Ramlee MK, Wang J, Toh WX, Li S. Transcription Regulation of the Human Telomerase Reverse Transcriptase (hTERT) Gene. Genes (Basel). 2016; 7(8):50. Epub 2016/08/23. https://doi.org/10.3390/genes7080050 PMID: 27548225; PubMed Central PMCID: PMC4999838.

24. Kang X, Chen W, Kim RH, Kang MK, Park NH. Regulation of the hTERT promoter activity by MSH2, the hnRNPs K and D, and GRHL2 in human oral squamous cell carcinoma cells. Oncogene. 2009; 28 (4):565–74. Epub 2008/11/19. https://doi.org/10.1038/onc.2008.404 [pii] PMID: 19015635; PubMed Central PMCID: PMC2919678.

25. Dworkin S, Darido C, Georgy SR, Wilanowski T, Srivastava S, Ellett F, et al. Midbrain-hindbrain boundary patterning and morphogenesis are regulated by diverse grainy head-like 2-dependent pathways. Development. 2012; 139(3):525–36. Epub 2012/01/10. https://doi.org/10.1242/dev.066522 PMID: 22223680.

26. Dworkin S, Jane SM. Novel mechanisms that pattern and shape the midbrain-hindbrain boundary. Cellular and molecular life sciences: CMLS. 2013; 70(18):3365–74. Epub 2013/01/12. https://doi.org/10.1007/s00018-012-1240-x PMID: 23307071.

27. Krystkowiak I, Lenart J, Debski K, Kuterba P, Petas M, Kaminska B, et al. Nencki Genomics Database —Ensembl funcgen enhanced with intersections, user data and genome-wide TFBS motifs. Database (Oxford). 2013; 2013:bat069. Epub 2013/10/04. https://doi.org/10.1093/database/bat069 PMID: 24089456; PubMed Central PMCID: PMC3788330.

28. Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, et al. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics. 2004; 20(17):3045–54. Epub 2004/06/18. https://doi.org/10.1093/bioinformatics/bth361 PMID: 15201187.

29. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. J Mol Biol. 2000; 302(1):205–17. https://doi.org/10.1006/jmbi.2000.4042 PMID: 10964570

30. Arnold P, Erb I, Pachkov M, Molina N, van Nimwegen E. MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences. Bioinformatics. 2011; 28(4):487–94. https://doi.org/10.1093/bioinformatics/btr695 PMID: 22334039

31. Pachkov M, Balwierz PJ, Arnold P, Ozonov E, van Nimwegen E. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic acids research. 2013; 41(Database issue):D214–20. Epub 2012/11/28. https://doi.org/10.1093/nar/gks1145 PMID: 23180783; PubMed Central PMCID: PMC3531101.

32. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic acids research. 2019; 47(D1):D729–35. Epub 2018/11/22. https://doi.org/10.1093/nar/gky1094 PMID: 30462313; PubMed Central PMCID: PMC6324081.

33. Lin CC, Liu LZ, Addison JB, Wonderlin WF, Ivanov AV, Ruppert JM. A KLF4-miRNA-206 autoregulatory feedback loop can promote or inhibit protein translation depending upon cell context. Mol Cell Biol. 2011; 31(12):2513–27. Epub 2011/04/27. https://doi.org/10.1128/MCB.01189-10 PMID: 21518959; PubMed Central PMCID: PMC3133414.

34. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001; 25(4):402–8. https://doi.org/10.1006/meth.2001.1262 PMID: 11846609

35. Bougel S, Renaud S, Braunschweig R, Loukinov D, Morse HC 3rd, Bosman FT, et al. PAX5 activates the transcription of the human telomerase reverse transcriptase gene in B cells. J Pathol. 2010; 220 (1):87–96. Epub 2009/10/07. https://doi.org/10.1002/path.2620 PMID: 19806612; PubMed Central PMCID: PMC3422366.

36. Radonić A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun. 2004; 313(4):856–62. Epub 2004/01/07. https://doi.org/10.1016/j.bbrc.2003.11.177 PMID: 14706621.

37. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. 2015; 16:22. Epub 2015/02/28. https://doi.org/10.1186/s13059-014-0560-6 PMID: 25723102; PubMed Central PMCID: PMC4310165.

38. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489 (7414):57–74. Epub 2012/09/08. https://doi.org/10.1038/nature11247 PMID: 22955616; PubMed Central PMCID: PMC3439153.

39. Ghaleb AM, Yang VW. Kruppel-like factor 4 (KLF4): What we currently know. Gene. 2017; 611:27–37. Epub 2017/02/27. https://doi.org/10.1016/j.gene.2017.02.025 PMID: 28237823; PubMed Central PMCID: PMC5391259.

40. Krenacs L, Himmelmann AW, Quintanilla-Martinez L, Fest T, Riva A, Wellmann A, et al. Transcription factor B-cell-specific activator protein (BSAP) is differentially expressed in B cells and in subsets of B-cell lymphomas. Blood. 1998; 92(4):1308–16. Epub 1998/08/08. https://doi.org/10.1182/blood.V92.4.1308 PMID: 9694719.

41. Liu W, Li X, Chu ES, Go MY, Xu L, Zhao G, et al. Paired box gene 5 is a novel tumor suppressor in hepatocellular carcinoma through interaction with p53 signaling pathway. Hepatology (Baltimore, Md). 2011; 53(3):843–53. Epub 2011/02/15. https://doi.org/10.1002/hep.24124 PMID: 21319196.

42. Vidal LJ-P, Perry JK, Vouyovitch CM, Pandey V, Brunet-Dunand SE, Mertani HC, et al. PAX5α Enhances the Epithelial Behavior of Human Mammary Carcinoma Cells. Molecular Cancer Research. 2010; 8(3):444–56. https://doi.org/10.1158/1541-7786.MCR-09-0368 PMID: 20197384

43. Boglev Y, Wilanowski T, Caddy J, Parekh V, Auden A, Darido C, et al. The unique and cooperative roles of the Grainy head-like transcription factors in epidermal development reflect unexpected target gene specificity. Developmental biology. 2011; 349(2):512–22. https://doi.org/10.1016/j.ydbio.2010.11.011 PMID: 21081122

44. Chen W, Xiao Liu Z, Oh JE, Shin KH, Kim RH, Jiang M, et al. Grainyhead-like 2 (GRHL2) inhibits keratinocyte differentiation through epigenetic mechanism. Cell death & disease. 2012; 3:e450. https://doi.org/10.1038/cddis.2012.190 PMID: 23254293; PubMed Central PMCID: PMC3542624.

45. Varma S, Mahavadi P, Sasikumar S, Cushing L, Hyland T, Rosser AE, et al. Grainyhead-like 2 (GRHL2) distribution reveals novel pathophysiological differences between human idiopathic pulmonary fibrosis and mouse models of pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol. 2014; 306 (5):L405–19. https://doi.org/10.1152/ajplung.00143.2013 PMID: 24375798; PubMed Central PMCID: PMC3949053.

46. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol Cell. 2013; 52(1):25–36. Epub 2013/10/01. https://doi.org/10.1016/j.molcel.2013.08.037 PMID: 24076218; PubMed Central PMCID: PMC3811135.

47. Mlacki M, Darido C, Jane SM, Wilanowski T. Loss of Grainy head-like 1 is associated with disruption of the epidermal barrier and squamous cell carcinoma of the skin. PloS one. 2014; 9(2):e89247–e. https://doi.org/10.1371/journal.pone.0089247 PMID: 24586629.

48. Darido C, Georgy SR, Wilanowski T, Dworkin S, Auden A, Zhao Q, et al. Targeting of the tumor suppressor GRHL3 by a miR-21-dependent proto-oncogenic network results in PTEN loss and tumorigenesis. Cancer Cell. 2011; 20(5):635–48. Epub 2011/11/19. https://doi.org/10.1016/j.ccr.2011.10.014 PMID: 22094257.

49. Ghaleb AM, Nandan MO, Chanchevalap S, Dalton WB, Hisamuddin IM, Yang VW. Krüppel-like factors 4 and 5: the yin and yang regulators of cellular proliferation. Cell Research. 2005; 15(2):92–6. https://doi.org/10.1038/sj.cr.7290271 PMID: 15740636

50. Evans PM, Liu C. Roles of Krupel-like factor 4 in normal homeostasis, cancer and stem cells. Acta biochimica et biophysica Sinica. 2008; 40(7):554–64. Epub 2008/07/08. https://doi.org/10.1111/j.1745-7270.2008.00439.x PMID: 18604447; PubMed Central PMCID: PMC2668950.

51. Foster KW, Liu Z, Nail CD, Li X, Fitzgerald TJ, Bailey SK, et al. Induction of KLF4 in basal keratinocytes blocks the proliferation-differentiation switch and initiates squamous epithelial dysplasia. Oncogene. 2005; 24(9):1491–500. https://doi.org/10.1038/sj.onc.1208307 PMID: 15674344.

52. Choi WJ, Youn SH, Back JH, Park S, Park EJ, Kim KJ, et al. The role of KLF4 in UVB-induced murine skin tumor development and its correlation with cyclin D1, p53, and p21(Waf1/Cip1) in epithelial tumors of the human skin. Archives of dermatological research. 2011; 303(3):191–200. Epub 2010/12/07. https://doi.org/10.1007/s00403-010-1101-0 PMID: 21132436.

53. Li J, Zheng H, Yu F, Yu T, Liu C, Huang S, et al. Deficiency of the Kruppel-like factor KLF4 correlates with increased cell proliferation and enhanced skin tumorigenesis. Carcinogenesis. 2012; 33(6):1239–46. https://doi.org/10.1093/carcin/bgs143 PMID: 22491752

54. Song E, Ma X, Li H, Zhang P, Ni D, Chen W, et al. Attenuation of kruppel-like factor 4 facilitates carcinogenesis by inducing g1/s phase arrest in clear cell renal cell carcinoma. PloS one. 2013; 8(7):e67758. Epub 2013/07/19. https://doi.org/10.1371/journal.pone.0067758 PMID: 23861801; PubMed Central PMCID: PMC3702498.

55. Pearson R, Fleetwood J, Eaton S, Crossley M, Bao S. Kruppel-like transcription factors: a functional family. Int J Biochem Cell Biol. 2008; 40(10):1996–2001. Epub 2007/10/02. https://doi.org/10.1016/j.biocel.2007.07.018 PMID: 17904406.

56. Baumann Kubetzko FB, di Paolo C, Maag C, Meier R, Schäfer BW, Betts DR, et al. The PAX5 oncogene is expressed in N-type neuroblastoma cells and increases tumorigenicity of a S-type cell line. Carcinogenesis. 2004; 25(10):1839–46. https://doi.org/10.1093/carcin/bgh190 PMID: 15155532

57. Norhany S, Kouzu Y, Uzawa K, Hayama M, Higo M, Koike H, et al. Overexpression of PAX5 in oral carcinogenesis. Oncology reports. 2006; 16(5):1003–8. Epub 2006/10/04. https://doi.org/10.3892/or.16.5.1003 PMID: 17016584.

58. Fabian J, Lodrini M, Oehme I, Schier MC, Thole TM, Hielscher T, et al. GRHL1 Acts as Tumor Suppressor in Neuroblastoma and Is Negatively Regulated by MYCN and HDAC3. Cancer research. 2014; 74 (9):2604–16. https://doi.org/10.1158/0008-5472.CAN-13-1904 PMID: 24419085.

59. Georgy SR, Cangkrama M, Srivastava S, Partridge D, Auden A, Dworkin S, et al. Identification of a Novel Proto-oncogenic Network in Head and Neck Squamous Cell Carcinoma. J Natl Cancer Inst. 2015; 107(9):djv152. Epub 2015/06/13. https://doi.org/10.1093/jnci/djv152 PMID: 26063791; PubMed Central PMCID: PMC4836819.

60. White MA, Kwasnieski JC, Myers CA, Shen SQ, Corbo JC, Cohen BA. A Simple Grammar Defines Activating and Repressing cis-Regulatory Elements in Photoreceptors. Cell Rep. 2016; 17(5):1247–54. Epub 2016/10/27. https://doi.org/10.1016/j.celrep.2016.09.066 PMID: 27783940; PubMed Central PMCID: PMC5123866.

61. Natesan S, Rivera VM, Molinari E, Gilman M. Transcriptional squelching re-examined. Nature. 1997; 390(6658):349–50. Epub 1997/12/06. https://doi.org/10.1038/37019 PMID: 9389473.

62. Gibson TJ, Seiler M, Veitia RA. The transience of transient overexpression. Nature methods. 2013; 10 (8):715–21. Epub 2013/08/01. https://doi.org/10.1038/nmeth.2534 PMID: 23900254.

63. Moriya H. Quantitative nature of overexpression experiments. Mol Biol Cell. 2015; 26(22):3932–9. Epub 2015/11/07. https://doi.org/10.1091/mbc.E15-07-0512 PMID: 26543202; PubMed Central PMCID: PMC4710226.

64. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol Sci. 2014; 35(2):76–85. Epub 2014/01/07. https://doi.org/10.1016/j.tips.2013.11.005 PMID: 24388790; PubMed Central PMCID: PMC3954851.