Data in Brief

Profiling of Sox4-dependent transcriptome in skin links tumour suppression and adult stem cell activation

Miguel Foronda [a],[*],[1], Lucia Morgado-Palacin [b],[1], Gonzalo Gómez-López [c], Orlando Domínguez [d], David G. Pisano [c], Maria A. Blasco [a],[*]

[a] Telomeres and Telomerase Group, Molecular Oncology Program, Spanish National Cancer Research Centre (CNIO), C/ Melchor Fernández Almagro, 3, E-28029, Madrid, Spain
[b] Tumour Suppression Group, Molecular Oncology Program, Spanish National Cancer Research Centre (CNIO), C/ Melchor Fernández Almagro, 3, E-28029, Madrid, Spain
[c] Bioinformatics Unit, Structural Biology and Biocomputing Program, Spanish National Cancer Research Centre (CNIO), C/ Melchor Fernández Almagro, 3, E-28029, Madrid, Spain
[d] Genomics Unit, Biotechnology Program, Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro, 3, E-28029 Madrid, Spain

ARTICLE INFO

ABSTRACT

Adult stem cells (ASCs) reside in specific niches in a quiescent state in adult mammals. Upon specific cues they become activated and respond by self-renewing and differentiating into newly generated specialised cells that ensure appropriate tissue fitness. ASC quiescence also serves as a tumour suppression mechanism by hampering cellular transformation and expansion (White AC et al., 2014). Some genes restricted to early embryonic development and adult stem cell niches are often potent modulators of stem cell quiescence, and derailed expression of these is commonly associated to cancer (Vervoort SJ et al., 2013). Among them, it has been shown that recommissioned Sox4 expression facilitates proliferation, survival and migration of malignant cells. By generating a conditional Knockout mouse model in stratified epithelia (Sox4^{cKO} mice), we demonstrated a delayed plucking-induced Anagen in the absence of Sox4. Skin global transcriptome analysis revealed a prominent defect in the induction of transcriptional networks that control hair follicle stem cell (HFSC) activation such as those regulated by Wnt/Ctnnb1, Shh, Myc or Sox9, cell cycle and DNA damage response-associated pathways. Besides, Sox4^{cKO} mice are resistant to skin carcinogenesis, thus linking Sox4 to both normal and pathological HFSC activation (Foronda M et al., 2014). Here we provide additional details on the analysis of Sox4-regulated transcriptome in Telogen and Anagen skin. The raw and processed microarray data is deposited in GEO under GSE58155.

Specifications

Organism/cell line/tissue	*Mus musculus*/keratinocytes/back skin epidermis
Sex	Male
Sequencer or array type	Whole Mouse Genome DNA Microarray (Agilent)
Data format	Raw and analysed
Experimental factors	Telogen vs Anagen. Sox4^{WT} vs Sox4^{cKO} mouse skin.
Experimental features	Telogen (resting) and Anagen (proliferative; 12 days post-plucking) skin from adult Sox4^{WT} vs Sox4^{cKO} mice was collected and total RNA was extracted to study global transcriptome changes, in the absence of Sox4, in active HFSCs.
Consent	Mice were maintained under specific pathogen-free conditions at the CNIO, and mouse experimentation was performed in agreement with the recommendations of FELASA.
Sample source location	N.A.

* Corresponding authors.
 E-mail addresses: miguel.foronda@ed.ac.uk (M. Foronda), mblasco@cnio.es (M.A. Blasco).
[1] Current address: MRC Centre for Regenerative Medicine (The University of Edinburgh), 5 Little France Drive, EH16 4UU Edinburgh, UK.

1. Direct link to deposited data

The deposited data can be found at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58155.

2. Experimental design, materials and methods

### 2.1. Experimental groups and conditions

To understand the links between cancer and stem cell activation [1,2], we generated Sox4^{cKO} mice in stratified epithelia by crossing Sox4$^{lox/lox}$ with Keratin5-Cre transgenic mice [3,4]. Sox4^{cKO} mice display reduced Sox4 expression in the whole organism and undetectable levels in back and tail skin epidermis [3]. As Sox4 expression is restricted to active HFSCs, its depletion should principally affect this compartment during skin and hair regeneration [5,6]. First, male mice were selected to avoid potential hormonal-dependent hair cycle alterations [7]. We initially performed Telogen skin analysis on Sox4^{WT} vs Sox4^{cKO} mice of 4–5 months of age (n = 3 mice per genotype). At this stage mouse skin enters

into Anagen at a slower pace and asynchronously in small discrete areas, instead of happening in sagittal coordinated waves [7]. During sampling, we therefore avoided collecting Anagen skin patches arising from dominant behaviour among male mice or spontaneous Anagen induction patches [7,8]. In order to study the contribution of *Sox4* to HFSC activation, we performed hair plucking, a bona-fide method for assessing hair regeneration [5,7,8]. In this setting, we used young adult (4–5 months, 3 mice per genotype) and aged mice (17–18 months, 3 mice per genotype). We did not observe any differences in Anagen induction or in the contribution of *Sox4* to modulating HFSC pathways in young versus aged mice [3], therefore all these mice were grouped in this study based solely on their genotype. To induce Anagen, 1 cm^2 hair skin patch was plucked with tweezers and tape-stripped from the lower dorsal part of anaesthetized mice. 12 days later, the plucked patch is in full Anagen and can be identified macroscopically as a darkly pigmented skin area [7,8].

2.2. Skin collection, RNA extraction and microarray hybridization

A critical point to obtain high quality RNA from skin is to preserve the samples immediately after collection by immersing them on RNAlater (QIAGEN); this method works for skin RNA extraction even more efficiently than liquid nitrogen flash-freezing, in our experience.

For skin collection we first hair-clipped the mice and then culled them by cervical dislocation to minimize the time elapsed before tissue harvesting. Dorsal back skin was then cut and pulled with forceps and placed dermis side up in a Petri dish; then, the dermal fat was thoroughly removed by scrapping with a scalpel, immediately immersed in RNAlater in a new Petri dish and cut into 3 × 3 small pieces. RNAlater-soaked samples were preserved in 1.5 mL of fresh RNAlater solution at 4 °C overnight and − 80 °C thereafter until RNA extraction. After removing RNAlater completely, we then performed tissue disruption and RNA extraction. For this, tissue explants were placed on a 2 mL homogenisation tube containing 1 mL of TRIzol reagent (Life Technologies) and 5 zirconium oxide beads per 3 × 3mm^2 tissue bits. After 2 runs (4 cycles) on programme 2 in a Precellys-24 Tissue disruptor (Precellys), we proceeded to TRIzol-based RNA extraction; we adjusted the hydrophobicity of TRIzol's chloroform-extracted aqueous phase by adding ethanol up to a final concentration of 36%, and applied it directly to a silica spin column for RNA purification (RNeasy, QIAGEN). If RNA integrity number (RIN) is not greater than 7.5, additional RNA extraction from the original sample might be required to fulfil this Quality Control (QC). Nevertheless, a sample with RIN 6.7 was successfully analysed in this study. As an alternative approach we recommend using the Fibrous Tissue RNeasy minikit (QIAGEN) that includes a Proteinase K step, and gives high RNA yields and good RIN values. RNA integrity numbers
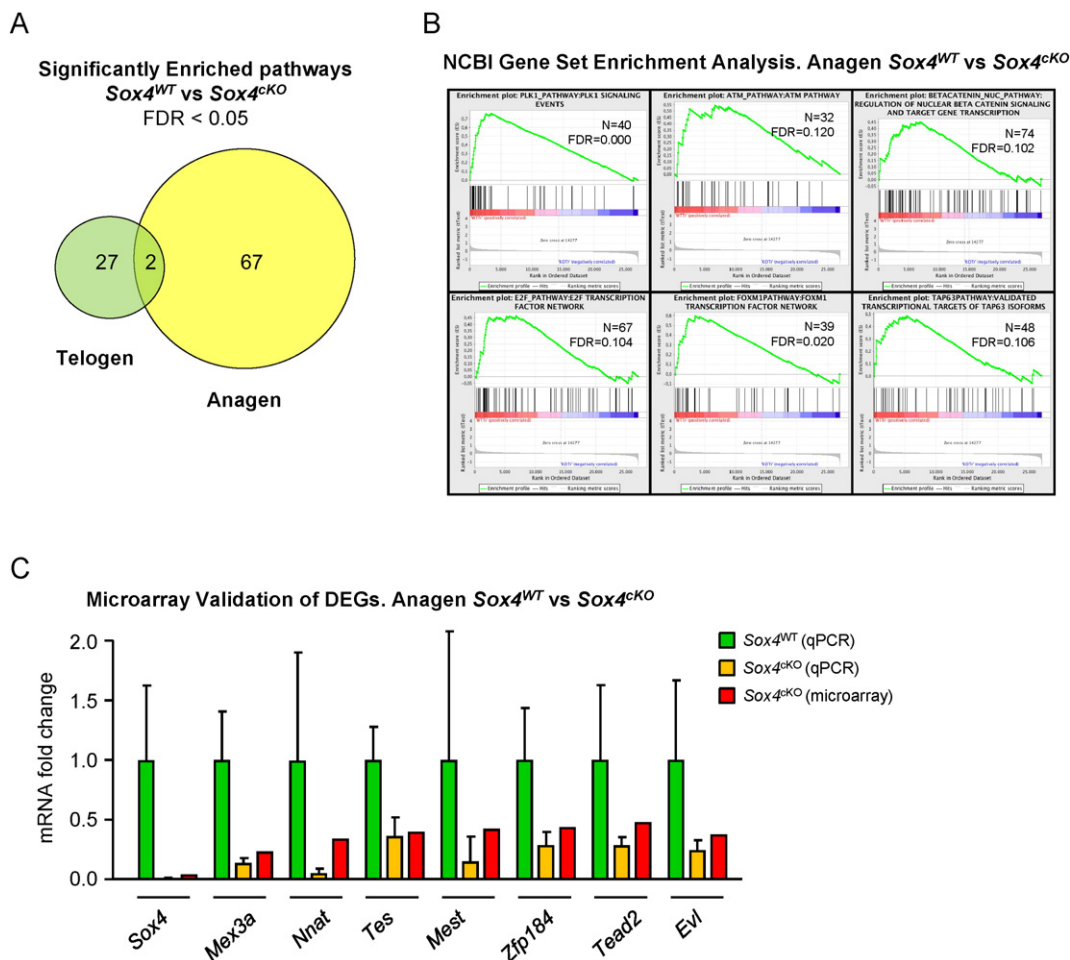
Fig. 1. A) Venn's diagram showing the overlap between the significantly enriched pathways in *Sox4^{WT}* vs *Sox4^{cKO}* mouse skin in Telogen (green) and Anagen (yellow). Only pathways showing FDR < 0.05 were included for this comparison. Note that most deregulated pathways belong to the Anagen group, reinforcing a role for *Sox4* in this condition, and that the overlap is minimal between both categories. B) Enrichment plots for selected pathways, obtained from the NCI repository at FDR < 0.150, where N indicates the number of genes included in each list, and FDR means false discovery rate (q value). The red to blue bar represents the ranked list of genes (red = *Sox4^{WT}*, blue = *Sox4^{cKO}*). Genes showing differential expression between genotypes are located at the edges of the bar, and similarly-expressed genes are in the centre. The Y axis represents the enrichment score (ES). *Plk1, ATM, nuclear β-catenin, E2F, FoxM1* and *TAP63* pathways are shown from left to right and top to bottom. C) Comparison of the values obtained by qPCR and microarray in *Sox4^{WT}* vs *Sox4^{cKO}* mouse skin, for the selected DEGs. Fold change relative to *Sox4^{WT}* (set to 1) is shown. N = 6 mice per genotype.

were obtained on an Agilent 2100 Bioanalyzer. A final amount of 100 ng of total RNA was labelled using a Low Input Quick Amp Labelling kit ("One-color Microarray-based Gene Expression analysis kit" v6.5, Agilent) following the manufacturer's instructions. Labelled samples were further purified on silica spin columns (RNeasy, QIAGEN). Finally, samples were hybridized to a 60K Whole Mouse Genome DNA Microarray (Agilent design ID 028005 SurePrint G3 Mouse GE 8x60K Microarray) in a SureHyb chamber (Agilent) in a final volume of 50 μL during 17 h at 65 °C. Subsequently, arrays were washed and images obtained with a G2505C DNA microarray scanner (Agilent).

2.3. Gene expression analysis

Images were analysed with Agilent Feature Extraction (FE) Software (ver. 10.1.1). FE performs spot quantitation and systematic gradient correction by spatial de-trending algorithms. QC reports provided by FE (which inform on signal qualities, background level and overall sensitivity within well-established acceptance thresholds) were examined for the absence of anomalies. We then performed background subtraction using the *normexp* method [9]. To normalize the dataset, we performed loess within-array normalization and quantile approach for between-array normalization. Differentially expressed genes were obtained by applying linear models with R limma package [10] (Bioconductor project, http://www.bioconductor.org). To account for multiple hypothesis testing, the estimated significance level (p value) was adjusted using Benjamini & Hochberg false discovery rate (FDR) correction [11]. Those genes with FDR < 0.05 were selected as differentially expressed genes (DEGs). Gene set enrichment analysis (GSEA) was studied using annotations from NCI [12], KEGG [13] and Reactome [14] databases. We additionally studied some custom-made gene lists obtained from genomic datasets publicly available [3,5]. Genes were ranked according to their limma moderated t statistic. After Kolmogorov–Smirnov testing, those gene sets showing FDR < 0.25, a well-established cut-off for the identification of biologically relevant gene sets [15], were considered enriched between $Sox4^{WT}$ vs $Sox4^{cKO}$ mice (Fig. 1A and B).

2.4. Microarray validation

To validate the DEGs and the significantly changed pathways, we extracted RNA from similar groups of mice following RNA later and RNeasy protocols (described above). We retrotranscribed 500 ng of total RNA using iSCRIPT Advanced (Bio-Rad) following the manufacturer's instructions. We used 1/10th dilutions of the resulting cDNA for subsequent quantitative-real time PCR (qPCR) in a 7900HT 384-well real-time fast thermocycler (Applied Biosystems) using SYBR $2\times$ qPCR master mix (Life Technologies). Fold change (FC) of genes of interest was further validated by comparing the results obtained in the microarray. Most DEGs displayed consistent FC in microarray versus qPCR values (Fig. 1C).

3. Discussion

Adult tissues are composed by specialised cells that need to be constantly fuelled up with fresh cells upon tissue damage or during normal homeostatic turnover. In one of the most paradigmatic examples, skin replenishment is carried out by its tissue stem cells, which have a high proliferative potential when compared to their progeny, but remain in a relatively quiescent state until the appropriate signals pull them out of their niches to migrate, differentiate and specialise [16]. On one hand, it is acknowledged that stem cell function decline is one of the hallmarks of ageing, and molecular ageing is linked to decreased stem cell activation potential [17,18]. On the other hand, adult stem cell activation constitutes a prerequisite to license carcinogenesis initiation and progression, at least in a two-step chemically-induced mouse skin carcinogenesis model [1]. Therefore, this leads to the speculation that there might be common signals that link stem cell activation with ageing

and cancer. We proved that *Sox4* is a key factor that warrants normal hair follicle stem cell activation during hair regeneration and wound healing. Interestingly, abrogation of *Sox4* expression in skin prevented chemically-induced tumorigenesis and globally reduced mRNA levels correlated with diminished incidence of spontaneous cancer in mice [3]. By analysing the transcriptome of *Sox4*-depleted skin during hair regeneration we observed deficiencies in the induction of pathways that modulate HFSC activation, but we also detected strong deregulation of groups of genes that are typically associated to cancer. In summary, our results demonstrate that understanding the *Sox4*-regulated transcriptome might be an interesting subject of study for effective cancer therapies and controlled regenerative medicine approaches, and thus we hope that this data resource constitutes a valuable tool for the research community.

Author contributions

MF performed most of the experiments. LM-P helped with microarray validation. GG-L and DG-P performed Bioinformatics analysis. OD performed microarray hybridization and scanning. MF and MAB conceived and designed the experiments, interpreted results and wrote the paper. MAB supervised the project and secured funding. All the authors commented and approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

References

[1] A.C. White, J.K. Khuu, C.Y. Dang, J. Hu, K.V. Tran, A. Liu, S. Gomez, Z. Zhang, R. Yi, P. Scumpia, M. Grigorian, W.E. Lowry, Stem cell quiescence acts as a tumour suppressor in squamous tumours. Nat. Cell Biol. 16 (1) (2014) 99–107, http://dx.doi.org/10.1038/ncb2889.

[2] S.J. Vervoort, R. van Boxtel, P.J. Coffer, The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? Oncogene 32 (2013) 3397–3409.

[3] M. Foronda, P. Martinez, S. Schoeftner, G. Gomez-López, R.P. Schneider, J.M. Flores, D.G. Pisano, M.A. Blasco, Sox4 links tumor suppression to accelerated aging in mice by modulating stem cell activation. Cell Rep. 8 (2) (2014) 487–500, http://dx.doi.org/10.1016/j.celrep.2014.06.031 (24).

[4] A. Ramirez, A. Page, A. Gandarillas, J. Zanet, S. Pibre, M. Vidal, L. Tusell, A. Genesca, D.A. Whitaker, D.W. Melton, J.L. Jorcano, A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated recombination. Genesis 39 (1) (2004) 52–57.

[5] W.H. Lien, L. Polak, M. Lin, K. Lay, D. Zheng, E. Fuchs, In vivo transcriptional governance of hair follicle stem cells by canonical Wnt regulators. Nat. Cell Biol. 16 (2) (2014) 179–190, http://dx.doi.org/10.1038/ncb2903.

[6] V. Greco, T. Chen, M. Rendl, M. Schober, H.A. Pasolli, N. Stokes, J. Dela Cruz-Racelis, E. Fuchs, A two-step mechanism for stem cell activation during hair regeneration. Cell Stem Cell 4 (2) (2009) 155–169, http://dx.doi.org/10.1016/j.stem.2008.12.009 (6).

[7] S. Muller-Rover, B. Handjiski, C. van der Veen, S. Eichmuller, K. Foitzik, I.A. McKay, K.S. Stenn, R. Paus, A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. J. Investig. Dermatol. 117 (1) (2001) 3–15.

[8] K.B. Jensen, R.R. Driskell, F.M. Watt, Assaying proliferation and differentiation capacity of stem cells using disaggregated adult mouse epidermis. Nat. Protoc. 5 (5) (2010) 898–911, http://dx.doi.org/10.1038/nprot.2010.39.

[9] M.E. Ritchie, J. Silver, A. Oshlack, M. Holmes, D. Diyagama, A. Holloway, G.K. Smyth, A comparison of background correction methods for two-colour microarrays. Bioinformatics 23 (20) (2007) 2700–2707.

[10] G.K. Smyth, J. Michaud, H.S. Scott, Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21 (9) (2005) 2067–2075 (1).

[11] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57 (1) (1995) 289–300.

[12] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41 (Database issue) (2013) D991–D995, http://dx.doi.org/10.1093/nar/gks1193.

[13] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42 (Database issue) (2014) D199–D205, http://dx.doi.org/10.1093/nar/gkt1076.

[14] D. Croft, A.F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M.R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, P. D'Eustachio, The Reactome pathway knowledgebase. Nucleic Acids Res. 42 (Database issue) (2014) D472–D477, http://dx.doi.org/10.1093/nar/gkt1102.

[15] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.

[16] G. Solanas, S.A. Benitah, Regenerating the skin: a task for the heterogeneous stem cell pool and surrounding niche. Nat. Rev. Mol. Cell Biol. 14 (11) (2013) 737–748.

[17] N.E. Sharpless, R.A. DePinho, How stem cells age and why this makes us grow old. Nat. Rev. Mol. Cell Biol. 8 (2007) 703–713.

[18] C. López-Otín, M.A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. Cell 153 (6) (2013) 1194–11217, http://dx.doi.org/10.1016/j.cell.2013.05.039 (6).