bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Unique and redundant roles of SOX2 and SOX17 in regulating the germ cell tumour fate

Sina Jostes[1], Martin Fellermeyer[1], Gina Merges[1], Glen Kristiansen[2], Daniel Nettersheim[3*+] and Hubert Schorle[1*+]

[1]Institute of Pathology, Department of Developmental Pathology, University of Bonn Medical School, Bonn, Germany
[2]Institute of Pathology, University of Bonn Medical School, Bonn, Germany
[3]Department of Urology, Urological Research Lab, Translational Urooncology, University Hospital Düsseldorf, Düsseldorf, Germany

* contributed equally

+ correspondence to:

Prof. Dr. Hubert Schorle,
Institute of Pathology, Department of Developmental Pathology, University of Bonn Medical School, Venusberg-Campus 1, Building 62, 53127 Bonn, Germany.
E-mail: Hubert.Schorle@ukbonn.de
Phone: +49 228 287 16342
Fax: +49 228 287 19757

or

Prof. Dr. Daniel Nettersheim
Department of Urology, Urological Research Lab, Translational Urooncology, University Hospital Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany
E-Mail: Daniel.Nettersheim@med.uni-duesseldorf.de
Phone: +49 0211 81 15844

KEY WORDS: SOX2, SOX17, Transcription Factor, Germ Cell Tumour, Pluripotency

LIST OF ABBREVIATIONS:
TGCT = Testicular Germ Cell Tumour
EC = Embryonal Carcinoma
GCNIS = Germ Cell Neoplasia in Situ
PGC = Primordial Germ Cell
ESC = Embryonic Stem Cell
TSS = Transcription Start Site
GC = Germ Cell
GSEA = Gene Set Enrichment Analysis
GEO = Gene Expression Omnibus
AP = Alkaline Phosphatase

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Embryonal carcinomas (ECs) and seminomas are testicular germ cell tumours. ECs display expression of SOX2, while seminomas display expression of SOX17. In somatic differentiation, SOX17 drives endodermal cell fate. However, seminomas lack expression of endoderm markers, but show features of pluripotency. Here, we use ChIP-sequencing to report and compare the binding pattern of SOX17 in seminoma-like TCam-2 cells to SOX2 in EC-like 2102EP cells and SOX17 in somatic cells. In seminoma-like cells, SOX17 was detected at canonical (SOX2/OCT4), compressed (SOX17/OCT4) and other SOX family member motifs. SOX17 directly regulates *TFAP2C*, *PRDM1* and *PRDM14*, thereby maintaining latent pluripotency and supressing somatic differentiation. In contrast, in somatic cells canonical motifs are not bound by SOX17. In sum only 10% of SOX17 binding sites overlap in seminoma-like and somatic cells. This illustrates that binding site choice is highly dynamic and cell type specific. Deletion of SOX17 in seminoma-like cells resulted in loss of pluripotency, marked by a reduction of OCT4 protein level and loss of alkaline phosphatase activity. Further, we found that in EC-like cells SOX2 regulates pluripotency-associated genes, predominantly by partnering with OCT4. In conclusion, SOX17 (in seminomas) functionally replaces SOX2 (in ECs) to maintain expression of the pluripotency cluster.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Testicular cancer is the most common malignancy in young men [1]. The most frequent testicular tumours are the type II germ cell tumours (TGCTs). Based on epigenetic profile, morphology and marker expression TGCTs are classified as seminomas or non-seminomas [1]. Non-seminomas initially present as embryonal carcinomas (ECs), which can further differentiate into mixed non-seminoma [1]. Seminomas and ECs arise from the same precursor lesion (germ cell neoplasia in situ / GCNIS), which is believed to result from primordial germ cells (PGCs) arrested in development [2]. Seminomas retain to the most part the genetic and epigenetic signature of PGCs i.e. they express pluripotency (*LIN28, NANOG, POU5F1* (*OCT4*)) and PGC (*PRDM1, TFAP2C, DMRT1, cKIT*, *SOX17*) markers [3]. The gene expression profile of ECs is highly similar to embryonic stem cells (ESCs), hallmarked by expression of pluripotency markers (*LIN28*, *NANOG*, *POU5F1* (OCT4), *TFAP2C*, *GDF3*, *DNMT3B*, *TDGF1*, *SOX2*) [3]. Interestingly, seminomas are SOX17+ and SOX2-, while ECs are SOX17- and SOX2+ [4].

Both SOX17 and SOX2 are members of the SOX family of transcription factors, which comprises 20 proteins hallmarked by a high mobility group box DNA binding domain [5]. Both SOX2 and SOX17 are able to partner with OCT4 and, as a consequence, recognize and bind specific binding motifs [6,7]. In human and mouse ESCs, SOX2/OCT4 bind to canonical motifs on the DNA (*CTTTGTCATGCAAAT*-like), which are composite SOX (*CATTGTC*-like) and OCT (*ATGCAAAT*-like) motifs. Canonical motifs are located in several enhancers of pluripotency genes [7,8]. By SOX2/OCT4 mediated transactivation of pluripotency markers, ESCs maintain their undifferentiated, pluripotent status [7,8]. An increase of SOX17 levels in ESCs leads to a shift of OCT4 from partnering with SOX2 to SOX17, resulting in binding to compressed SOX17/OCT4 motifs [7,9]. Compressed motifs (*CTTTGTATGCAAAT*-like) are similarly made up from juxtaposition of SOX and OCT binding sites, but lack a central cytosine [7]. Compressed motifs are predominantly found in regulatory regions of somatic genes [7,9]. In mouse and human ESCs, binding of SOX17/OCT4 to compressed motifs initiates endodermal differentiation [7,9]. Although in ESCs the SOX17/OCT4 heterodimer is preferentially bound to the compressed motif, it can also be detected at the canonical motif [7].

In human PGCs SOX17 and OCT4 are also highly expressed [10]. Alike seminomas, human PGCs lack expression of endoderm-associated genes but display a so-called latent pluripotency, i.e. expressing core regulators of pluripotency while being unable to spontaneously differentiate like ECs or ESCs. Three transcription factors, SOX17, PRDM1 and TFAP2C induce the PGC fate in the early embryo and establish and maintain latent pluripotency [10]. We reasoned that in human PGCs and seminomas, SOX17 acts not as an endoderm-initiation factor but supports latent pluripotency. In line with our hypothesis, it was shown by Irie *et al*. that knockdown of SOX17 in seminoma cells results in downregulation of PGC / seminoma and pluripotency markers [10]. We further assumed that SOX2 and SOX17 may fulfil similar roles in maintaining expression of pluripotency genes in EC and seminoma cells, respectively. Since it was shown in mice that the SOX17/OCT4 heterodimer is able to bind to the canonical motif, we hypothesized that in seminomas SOX17 uses at least some of the SOX2 binding sites that were described in ESCs to activate pluripotency.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Here we used high-throughput chromatin immunoprecipitation sequencing (ChIP-seq) analysis to investigate SOX17 and SOX2 genome occupancy in seminoma and EC model cell lines to decipher the functional role of both factors in TGCT biology. Based on our results, we present a model where SOX17 is essential for the seminoma cell fate by activating pluripotency-related genes and suppressing differentiation via transactivation of *TFAP2C*, *PRDM1* and *PRDM14*.

MATERIALS AND METHODS

Cell Culture

TGCT cell lines were cultured as described previously [11]. In brief, TCam-2 cells were cultured in RPMI medium (Thermo Fisher Scientific, Waltham, USA). 2102EP cells were cultured in DMEM medium (Thermo Fisher Scientific). Both media were supplemented with 10% FBS (Thermo Fisher Scientific), 1% penicillin / streptomycin (PAN, Aidenbach, Germany) and 200 mM L-glutamine (PAN).

Tumour Xenotransplantation

For xenotransplantation, 1 x 10^7 cells (TCam-2 or 2102EP) were re-suspended in 500 µl of 4°C cold matrigel (Corning, Amsterdam, Netherlands) and injected subcutaneously into the flank of Crl:NU-*Foxn1*^{nu} mice (Charles River Laboratories, Erkrath, Germany) as described [12] . Tumours were isolated six weeks after injection.

Fixation and Chromatin Preparation

For chromatin preparation from cell lines, 1 x 10^7 cells / immunoprecipitation were fixed for 30 minutes (min) at room temperature (RT) using Diagenode Crosslink Gold (Diagenode, Liège, Belgium). After two washing steps with PBS (Thermo Fisher Scientific), cells were additionally fixed for 10 min in 1% formaldehyde (Applichem, Darmstadt, Germany) and then processed according to the SimpleChIP® Plus Enzymatic Chromatin IP Kit (Cell Signaling Technology, Leiden, Netherlands). For chromatin preparation from tissue, 25 mg tumour tissue / IP were fixed in 1% formaldehyde (Applichem) and processed according to the SimpleChIP® Plus Enzymatic Chromatin IP Kit (Cell Signaling Technology).

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was carried out using 200 µg chromatin lysate and 5 µg antibody. The protocol was performed according to the Simple ChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology). 2% Input (= 2 µg chromatin) and IgG-IP served as controls. For ChIP-qPCR 10 µl of IP samples were amplified using the Genomeplex Single Cell Whole Genome Amplification Kit (WGA4) (Sigma Aldrich, Taufkirchen, Germany) and subjected to qPCR in a 1:40 dilution. For primer information, see **Supp. Table 1**.

Next-Generation Sequencing

ChIP-seq libraries were generated using the MicroPlex Library Preparation Kit v2 (Diagenode). Quality of ChIP-seq libraries was assessed on a 2400 Tapestation (Agilent, Santa Clara, USA) and library pools were quantified using Qubit (Thermo Fisher Scientific). Samples were sequenced on the Illumina High

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Seq 2500 v4 (Illumina, San Diego, USA) using 25 M single-end reads (1 x 75 bp) per sample. Reads were mapped to the human genome (hg38) using CLC Genomics Workbench [12] (Qiagen, Hilden, Germany) and data was analysed using HOMER (Hypergeometric Optimization of Motif EnRichment) Software (http://homer.ucsd.edu/homer/) [13]. For peak calling, regions were considered significantly enriched where signals in the SOX17 / SOX2 IP were at least four times higher compared to control. Binding occupancy profiles were calculated using Easeq https://easeq.net/ [14]. Gene set enrichment analysis (GSEA) on curated gene sets (C2) was performed using the Molecular Signatures Database (MSigDB) http://software.broadinstitute.org/gsea/msigdb/ [15,16]. Venn diagrams were visualized using Venny2.1 http://bioinfogp.cnb.csic.es/tools/venny/ [17]. Visualization of transcription factor binding occupancy was performed using the Integrative Genomics Viewer (IGV) browser https://software.broadinstitute.org/software/igv/ [18,19]. For meta-analysis with previously published SOX17 (GSM1505761, GSM1505762, GSM1505763) and SOX2 ChIP-seq datasets (GSM1505766, GSM1505768), bed-files were downloaded from GEO [20].

CRISPR/Cas9 mediated Deletion of *SOX17* in TCam-2 Cells

To generate *SOX17*-deficient TCam-2 cells (TCam-$2^{\Delta SOX17}$), $1x10^5$ cells were transfected with 500 ng pX330-U6-Chimeric_BB-CBh-hSpCas9 (addgene: #42230) coding for *SOX17*-KO-gRNA1 and *SOX17*-KO-gRNA2 (gRNA1: ACGGGTAGCCGTCGAGCGG, gRNA2: GGCACCTACAGCTACGCGC) in a 1:1 ratio. For transfection efficacy monitoring, cells were transfected with 500 ng pEGFP-N3 (Clontech, # 6080-1). Transfection was performed using FuGENE HD transfection reagent (500 ng DNA : 2.5 µl FuGENE reagent) (Promega, Mannheim, Germany).

RNA and Protein Isolation

Proteins were isolated using RIPA buffer and protein concentrations were determined using the BCA Protein Assay Kit (Thermo Fisher Scientific). Total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany). RNA quality was assessed by photometric measurement of ratios 260/280 nm and 260/230 nm using a NanoDrop photometer (PeqLab, Erlangen, Germany).

Quantitative real-time RT-PCR

cDNA was synthesized using Maxima First Strand cDNA synthesis Kit (Thermo Fisher Scientific). For qRT-PCR, 7.58 ng of cDNA was run in technical triplicates with Maxima SYBR Green qPCR Master Mix (Fermentas, St. Leon-Rot, Germany). Primer sequences are listed in **Supp. Table 2**. qRT-PCR was performed using the ViiA 7 RealTime PCR System (Thermo Fisher Scientific). *GAPDH* was used as housekeeping gene and for data normalization. Relative expression was calculated from Ct (relative expression = $2^{-\Delta\Delta CT}$).

Western blot Analysis

Proteins were separated by SDS gel electrophoresis and transferred onto a PVDF membrane (Carl Roth, Karlsruhe, Germany) using the semi-dry Trans Blot Turbo (Bio-Rad, Düsseldorf, Germany). Membranes were blocked in 5% non-fat milk powder (Applichem) in PBS (Applichem) + 0.05% Tween-20 (Applichem) (PBST). The membrane was then incubated with primary antibody overnight at 4°C, see

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supp. Table 3. After three washing steps in PBST, the membrane was incubated with HRP-conjugated secondary antibody for 1 hour (h) at RT. After three washing steps in PBST, signal was detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific).

Co-Immunoprecipitation

1.5 mg of Dynabeads Protein G (Thermo Fisher Scientific) were coated with 10 µg of primary antibody (see **Supp. Table 3**), according to the manufacturer's instructions. After washing the beads-antibody complex with PBS pH 7.4 (+0.02% Tween-20), 1 mg of whole protein lysate were added and incubated with the beads under constant rotation for 30 min at RT. After three washing steps, the beads-antibody-antigen complexes were eluted in 15 µl low pH glycine buffer + 5 µl Roti-Load (Carl Roth) for 5 min at 95°C. The beads were then separated from the antibody-antigen complex in a magnetic rack and the clear supernatant was loaded on a 12% SDS gel for visualization. 20 µg protein lysate served as 2% input sample.

Immunofluorescence

Cells were washed with PBS (Thermo Fisher Scientific) and fixed in 4% formaldehyde for 10 min at RT. Afterwards, cells were permeabilized using 0.5% Triton X-100 (Applichem) in PBS for 5 min at RT. Cells were washed twice in PBS and blocked in 2% bovine serum albumin in PBS for 1 h at RT. Cells were then incubated with primary antibody (see **Supp. Table 3**) diluted in blocking solution for 2 h at RT or 4°C overnight. After three washing steps in PBS, cells were incubated with Alexa Fluor secondary antibody (Thermo Fisher Scientific) diluted in blocking solution for 1 h at RT or 4°C overnight. Cells were then washed three times in PBS and counterstained with DAPI (Applichem) in PBS for 5 min at RT.

Alkaline Phosphatase Staining

For detection of alkaline phosphatase (AP) activity, cells were seeded in 6-well dishes. At the desired time point of analysis, cells were fixed for 1 min in 4% formaldehyde and stained for AP activity using the Alkaline Phosphatase Detection Kit (Merck Millipore, Darmstadt, Germany).

Statistical Analyses

For statistical analyses of qRT-PCR expression data, significance was calculated using two-tailed students t-test (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$). qRT-PCR measurements were performed in three biological replicates, where error bars indicate standard deviation from the mean. For statistical analyses of motif enrichment within ChIP-seq data, p-values were calculated using Homer Software (http://homer.ucsd.edu/homer/) [13].

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

Binding Sites of SOX17 and SOX2 in TGCT Cells

We hypothesized that in seminoma cells SOX17 and OCT4 partner to regulate pluripotency genes, while SOX17-binding to endodermal genes might be inhibited. By co-immunoprecipitation we verified interaction of SOX17 and OCT4 in seminoma-like TCam-2 cells (**Supp. Fig. 1A-B**), which had already been demonstrated in our previous study [21]. In ESCs, SOX2/OCT4 binds to canonical motifs to regulate pluripotency genes. Because EC cells are reminiscent of ESCs in expressing pluripotency genes we predicted a similar SOX2-OCT4 interaction in EC cells. Using co-immunoprecipitation we demonstrated interaction of SOX2 and OCT4 in the EC line 2102EP (**Supp. Fig. 1C-D**). Next, suitability of SOX17 and SOX2 antibodies for ChIP was tested by immunoblotting of SOX17- and SOX2-precipitated chromatin lysates. We detected a band indicative of the SOX17 protein in the SOX17 ChIP sample and SOX2 protein in the SOX2 ChIP sample (**Supp. Fig. 1E**), confirming that SOX17 and SOX2 were successfully immunoprecipitated.

To determine genome occupancy of SOX17 and SOX2 in TGCT cell lines, we performed ChIP-seq analysis in TCam-2 (SOX17) and 2102EP (SOX2) cells on three independent biological replicates. We detected a total of 19,157 SOX17 peaks in TCam-2 (**Fig. 1A, Supp. Data 1 upon request**) and 17,797 SOX2 peaks in 2102EP (**Fig. 1A, Supp. Data 1 upon request**). Interestingly, 4,630 of SOX17 and SOX2 peaks overlaped, indicating common binding sites of SOX17 and SOX2 in TCam-2 and 2102EP, respectively (**Fig. 1A**). Comparison of SOX17 and SOX2 genome binding occupancy in TGCT cells (+/- 10,000 bp from transcription start site (TSS)), confirms that ChIP-seq signals correlate to each other (r = 0.633) (**Fig. 1B**). SOX17 and SOX2 binding occupancy +/-10,000 bp from TSS further illustrates highest density in a region spanning +/- 1,500 bp from the TSS, an effect that was even more prominent for SOX17 binding in TCam-2 cells (**Fig. 1C**; dashed line). This indicates that both transcription factors mainly bind in close proximity to promoter regions at a similar distance to the TSS.

We next performed HOMER analysis to identify motifs in promoter and genomic regions directly bound by SOX17 in TCam-2 and SOX2 in 2102EP cells. As expected, in TCam-2 cells regions bound by SOX17 showed strongest enrichment for the SOX17 binding motif (-log10 (p-value): 4765) and the compressed motif (-log10 (p-value): 4759) (**Fig. 1D, Supp. Data 2 upon request**). Additionally, SOX17 recognizes and binds related SOX-family binding motifs (SOX3, SOX15, SOX10, SOX6, SOX2, SOX4 and SOX9) (**Supp. Data 2 upon request**). Further, within SOX17-bound regions, we detected TFAP2C (AP2-gamma) binding motifs (-log10 (p-value): 1766) (**Fig. 1D**, **Supp. Data 2 upon request**). TFAP2C is highly expressed in seminomas [3]. This suggests that SOX17 and TFAP2C might cooperate to regulate a set of common target genes in TCam-2. Interestingly, SOX17 peaks also mapped to canonical motifs (-log10 (p-value): 1181) (**Fig. 1D, Supp. Data 2 upon request**). In 2102EP, regions bound by SOX2 showed strongest enrichment for the canonical motif (-log10 (p-value): 5589) (**Fig. 1D, Supp. Data 2 upon request**). Further, we found enrichment of SOX3, SOX2, SOX6, SOX10, SOX15, SOX17, SOX4 and SOX9 binding motifs (**Supp. Data 2 upon request**). Also in 2102EP cells, which express TFAP2C at moderate levels, SOX2-bound regions were enriched for the TFAP2C motif (-log10 (p-value): 363)

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(**Fig. 1D, Supp. Data 2 upon request**). Interestingly, only a minor fraction of SOX2 peaks displayed the compressed motif (-log10 (p-value): 109) (**Fig. 1D**).

Figure 1

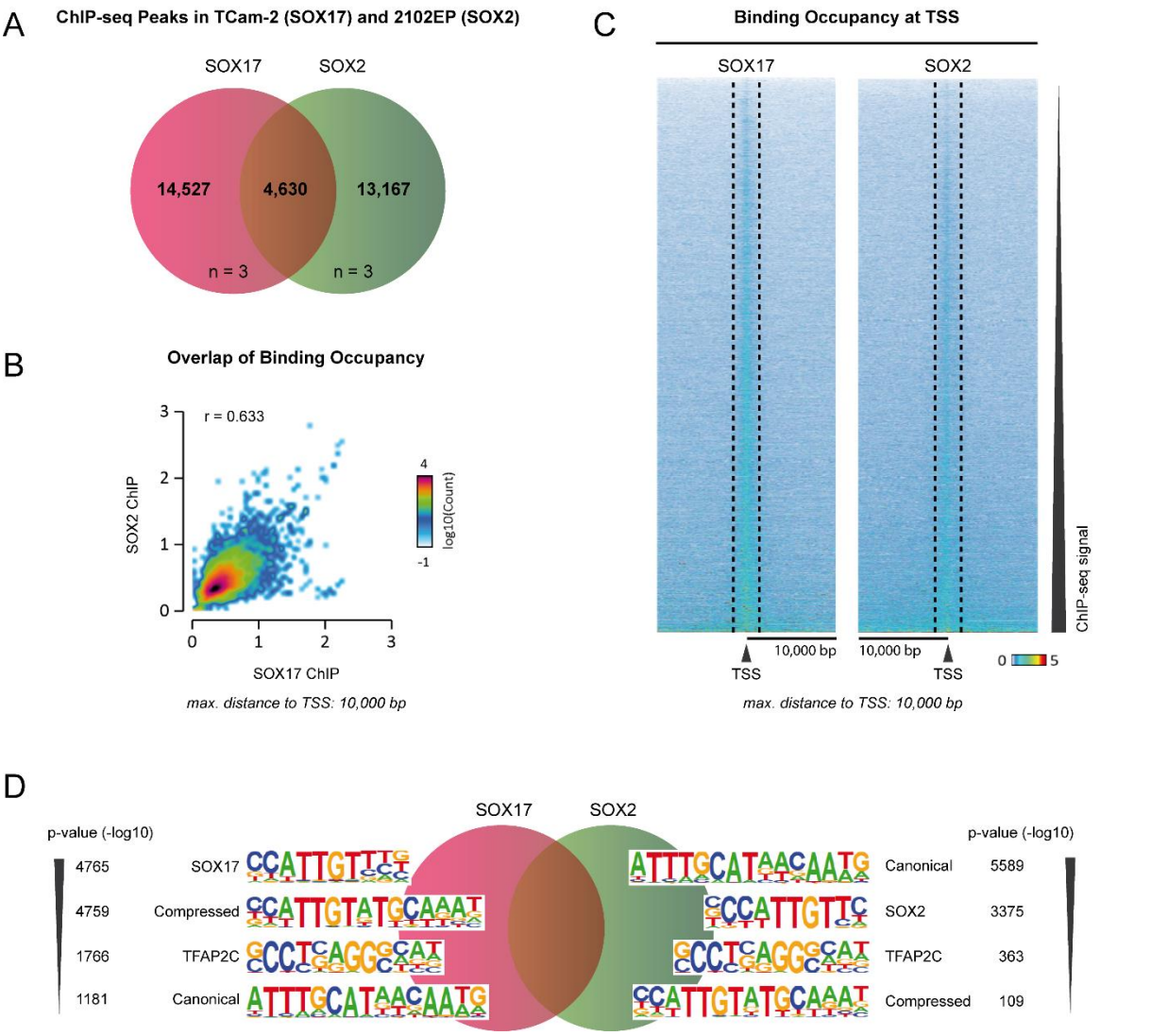


Figure 1: **(A)** Venn diagram depicting the number of annotated peaks identified commonly in all three biological replicates (n = 3) of SOX17 and SOX2 ChIP-seq samples. Red circles = SOX17 ChIP-seq analyses in TCam-2 cells. Green circles = SOX2 ChIP-seq analyses in 2102EP cells. **(B)** 2D histogram of the SOX17 ChIP (on x-axis) and SOX2 ChIP (on y-axis) signals found +/- 10,000 bp from TSS of genes with correlation coefficient (r). **(C)** Heat map of SOX17 and SOX2 ChIP signal 10,000 bp upstream and downstream of the transcription start site (TSS) of genes. Regions were sorted according to decreasing SOX17 or SOX2 ChIP signal. Dashed lines are spanning regions 1,500 bp +/- TSS. **(D)** Transcription factor binding motifs enriched in SOX17 and SOX2 ChIP-seq datasets, determined by HOMER [13]. The relative enrichment to background sequences was calculated as indicated (p-value).

Additionally, we analysed binding motifs identified in those regions that were commonly bound by SOX17 in TCam-2 and SOX2 in 2102EP cells. Expectedly, common DNA binding sites were similarly enriched for SOX motifs and the TFAP2C motif (**Supp. Data 2 upon request**). Strong enrichment was also detected for the canonical motif (-log10 (p-value): 1221) (**Supp. Data 2 upon request**). Lower

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

enrichment was seen for the compressed motif (-log10 (p-value): 173) (**Supp. Data 2 upon request**). Taken together, these findings illustrate that SOX17 peaks in TCam-2 and SOX2 peaks in 2102EP mostly overlap at canonical and SOX-family binding motifs.

The Genes regulated by SOX17 and SOX2 in TGCT cells

We next annotated the SOX17 and SOX2 peaks to genes with the most closely located TSS (only in a region +/- 10,000 bp from TSS). In TCam-2 cells, this included 3,738 genes bound by SOX17 and in 2102EP cells 3,269 genes bound by SOX2. Gene set enrichment analysis of both datasets revealed a significant overlap with SOX2 and NANOG target genes (**Supp. Fig. 2A-B**), confirming their role in regulating pluripotency-associated genes. Since in TGCTs both SOX17 and SOX2 share target genes with NANOG, we analysed whether there is a direct protein-protein interaction. However, co-immunoprecipitation showed no interaction of SOX17 or SOX2 with NANOG (**Supp. Fig. 2C**).

We then meta-analysed SOX17- and SOX2-bound genes with previously published gene expression data of TCam-2 and 2102EP, respectively [22-24]. In TCam-2 cells, we identified 31 genes above an arbitrary expression threshold of ≥ $\log_2 7$ and SOX17 peak score of ≥ 150 (**Fig. 2A**). We discovered that SOX17 is a regulator of the insulin-like growth factor *IGF1,* which is known to be overexpressed in seminoma and TCam-2 cells [25]. Further, we found that 14 of these 31 genes (*NANOG, ZNF511, GLDC, PLEKHG6, XPNPEP1, C8orf76, TFDP2, LRRC59, MIR205, PSAP, LOC728715, SRGAP1, KNTC1, AACS*) were exclusively bound by SOX17 in seminoma cells, but not in endoderm-, mesoderm- or mesendoderm-differentiated ESCs [26] (**Supp. Data 3 upon request**). This suggests that these genes contribute to the germ cell fate and / or latent pluripotency of seminomas. Of these, the pluripotency factor NANOG is a well-established marker of seminomas and ECs [27]. Not surprisingly, we therefore found *NANOG* also among SOX2 target genes in 2102EP cells (**Supp. Data 1 upon request**).

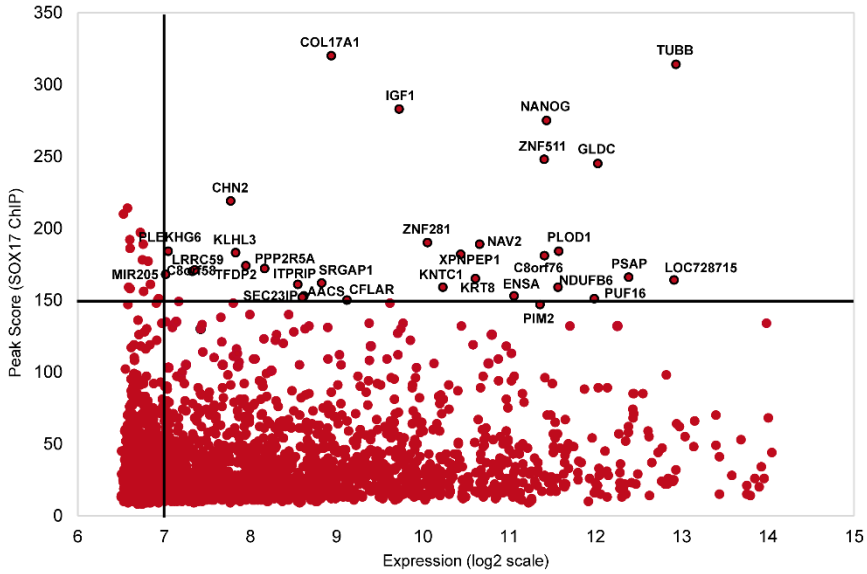
In 2102EP cells we identified 40 genes above an arbitrary gene expression threshold of ≥ $\log_2 7$ and SOX2 peak score of ≥ 200 (**Fig. 2B**). Of these*, GDF3, TDGF1* and *DPPA5* are well-known regulators of pluripotency [28-30]. Further, *OTX2* was described as a gatekeeper of PGC specification in mice [31]. Previously published microarray data reveals a higher expression of *OTX2* in TGCT tissues compared to normal testis tissue [32] (**Supp. Fig. 2D**). *OTX2* was also identified as a direct target of SOX17 in TCam-2 cells (**Supp. Data 1 upon request**). We then compared EC to ESC in order to find genes, which are linked to tumorigenesis. Comparing our SOX2 ChIP-seq data from 2102EP to data from ESC (HUE64) [26] revealed that four of the 40 genes (*MLEC, PIM2, CD99L2,* APOBEC3F) were exclusively bound by SOX2 in 2102EP (**Supp. Data 3 upon request**). *PIM2* is a proto-oncogene regulating survival and proliferation [33]. We found that *PIM2* is highly expressed in 2102EP and TCam-2 cells, but weakly in the Sertoli cell line FS1 and adult fibroblasts (MPAF) (**Supp. Fig. 2E**). This confirms findings from Jiménez-García *et al*., who demonstrated overexpression of *PIM1* and *PIM2* in TGCTs 33. Interestingly, *PIM2* was also bound by SOX17 (peak score: 147) in TCam-2 cells (**Fig. 2A**), suggesting that SOX2 and SOX17 are able to transactivate *PIM2*. Notably, *PIM1* was similarly bound by SOX17 and SOX2 in TGCT cells, albeit with a lower enrichment (SOX17 peak score: 31; SOX2 peak score: 21) (**Supp. Data 1 upon request**). Together, these data demonstrate that known markers of seminomas (*IGF1, NANOG, PIM1/2*) and / or ECs (*GDF3, TDGF1, NANOG, PIM1/2*) are under direct control of SOX17 and SOX2,

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

respectively. Additionally, we present *OTX2*, which is overexpressed in TGCTs, as a *bona fide* target of SOX17 and SOX2.

Figure 2

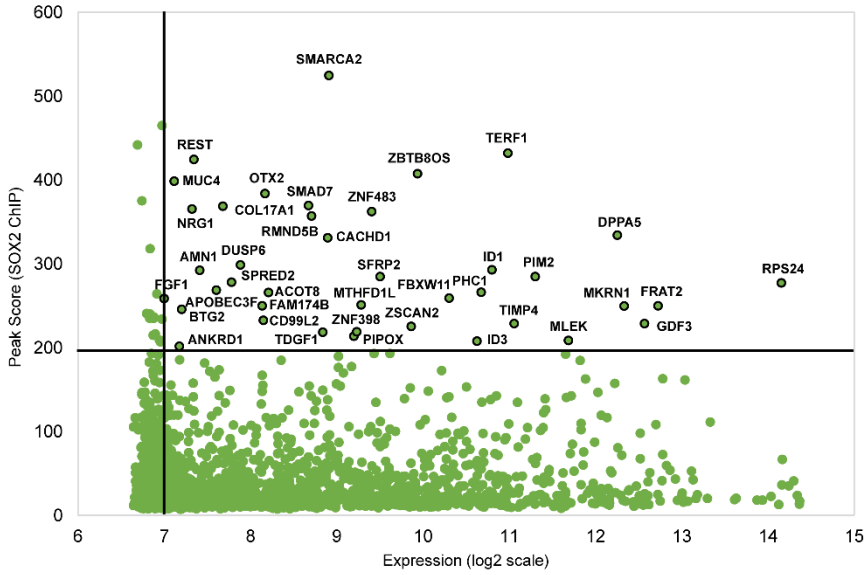
A



Gene	Gene Expression	Peak Score
COL17A1	8.94	320
TUBB	12.94	314
IGF1	9.73	283
NANOG	11.44	275
ZNF511	11.41	248
GLDC	12.03	245
CHN2	7.77	219
ZNF281	10.06	190
NAV2	10.66	189
PLOD1	11.58	184
PLEKHG6	7.05	184
KLHL3	7.83	183
XPNPEP1	10.44	182
C8orf76	11.41	181
TFDP2	7.95	174
PPP2R5A	8.17	172
LRRC59	7.36	171
C8orf58	7.33	170
MIR205	7.02	168
PSAP	12.39	166
KRT8	10.61	165
LOC728715	12.92	164
SRGAP1	8.83	162
ITPRIP	8.55	161
NDUFB6	11.57	159
KNTC1	10.23	159
ENSA	11.06	153
AACS	8.62	153
SEC23IP	8.61	152
PUF60	11.99	151

Top 30 genes (Gene expression > 7; Peak Score > 150; Peak Score-sorted)

B



Gene	Gene Expression	Peak Score
SMARCA2	8.91	525
TERF1	10.99	432
REST	7.35	425
ZBTB8OS	9.94	408
MUC4	7.12	399
OTX2	8.17	384
SMAD7	8.67	370
COL17A1	7.68	369
NRG1	7.32	365
ZNF483	9.41	362
RMND5B	8.71	357
DPPA5	12.26	334
CACHD1	8.90	331
DUSP6	7.89	299
ID1	10.80	293
AMN1	7.41	292
PIM2	11.30	285
SFRP2	9.50	285
SPRED2	7.78	278
RPS24	14.16	277
APOBEC3F	7.60	269
PHC1	10.68	266
ACOT8	8.21	266
FBXW11	10.31	259
FGF1	7.00	259
MTHFD1L	9.28	251
FRAT2	12.73	250
FAM174B	8.14	250
MKRN1	12.34	250
BTG2	7.20	246

Top 30 genes (Gene expression > 7; Peak Score > 200; Peak Score-sorted)

Figure 2: **(A)** Left: Scatterplot of genes, which are bound by SOX17 within a distance +/- 10,000 bp from TSS. The relative gene expression of each gene in TCam-2 cells is given on the x-axis, SOX17 ChIP-seq peak score is given on the y-axis. Right: Table of top 30 genes (peak-score-sorted) with a gene expression > 7. **(B)** Left: Scatterplot of genes, which are bound by SOX2 within a distance +/- 10,000 bp from TSS. The relative gene expression of each gene in 2102EP cells is given on the x-axis, SOX2 ChIP-seq peak score is given on the y-axis. Right: Table of top 30 genes (peak-score-sorted) with a gene expression > 7.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SOX17 and SOX2 regulate Germ Cell- and Pluripotency-associated Genes in TGCTs

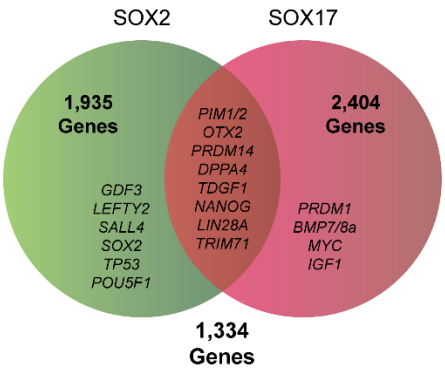
Next, we asked whether we could find additional genes under control of SOX17 in TCam-2 and SOX2 in 2102EP that are related to germ cell (GC) development, cancer and / or pluripotency. We found that SOX17 in TCam-2 binds and possibly regulates the cancer-related gene *MYC* [34], the GC specifier *PRDM1* [35] and members of the BMP pathway (*BMP7, BMP8a*), known to contribute to GC differentiation in mice [36] (**Fig. 3A**). SOX2-bound genes in EC cells included regulators of proliferation, cancer and pluripotency (*GDF3, LEFTY2, SALL4, SOX2, TP53, POU5F1 (OCT4)*) (**Fig. 3A**). Comparison of SOX17 and SOX2-bound genes in TCam-2 and 2102EP indicates that 1,334 genes overlap, representing approximately one third of all regulated genes of each factor (**Fig. 3A**). Among these, we found key-regulators of GC development and / or pluripotency, such as *PRDM14*, *DPPA4*, *TDGF1*, *LIN28A*, and *TRIM71* (in addition to *PIM1/2*, *OTX2* and *NANOG)* (**Fig. 3A**). All of these genes are highly expressed in TCam-2 and 2102EP, except for *TDGF1,* which is expressed in 2102EP only (**Supp. Fig. 3A**). Next, we asked which DNA motifs SOX17 and SOX2 use to regulate these genes. Canonical motifs were found within regulatory regions of *PIM2*, *TRIM71*, *PRDM14, GDF3* and *SOX2,* while compressed motifs were found in *PRDM1, MYC* and *IGF1* (**Fig. 3B - E).** The other SOX17 and SOX2 target genes (*PIM1, OTX2, DPPA4, TDGF1, NANOG* and *LIN28A*) seem to be regulated independent of OCT4, since neither compressed nor canonical motifs were detected (**Fig. 3B - E**).

In order to confirm that the SOX2 regulatory network is also established and maintained *in vivo*, we xenografted 2102EP and TCam-2 cells into the flank of nude mice and isolated tumours after six weeks. 2102EP xenografts formed compact EC tumours, while TCam-2 cells reprogrammed to an EC-like state, as published previously [22]. Both tumours highly expressed SOX2 and were subjected to SOX2 ChIP-qPCR. Similar to 2102EP cells *in vitro*, we were able to detect strong enrichment of SOX2 in the regulatory regions of *DPPA4*, *LEFTY2*, *LIN28A*, *TDGF1, NANOG, GDF3, TP53, PRDM14, SALL4* and *SOX2* in 2102EP tumours (n = 4) (**Supp. Fig. 4A**). Although less pronounced, TCam-2-derived reprogrammed EC-like tumours similarly demonstrate enrichment of SOX2 in the described regulatory regions in at least two of three samples (**Supp. Fig. 4B**).

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3

A



max. distance to TSS: 10,000 bp

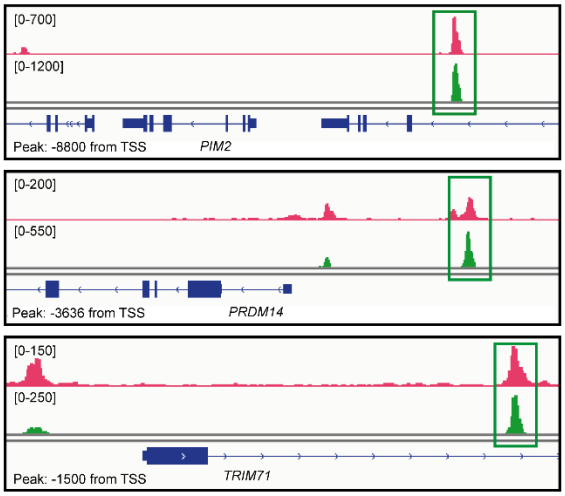
B

Target Gene	Distance to TSS [bp]	Transcription Factor	Description
PIM1	-468	SOX17	-
	+2087	SOX2	-
PIM2	-8763/-8826	SOX17/SOX2	Canonical
PRDM14	-3635/-3638	SOX17/SOX2	Canonical
OTX2	-299/-312	SOX17/SOX2	-
DPPA4	-91/-89	SOX17/SOX2	-
TDGF1	+3/+61	SOX17/SOX2	-
NANOG	-3712	SOX17	-
	-1656	SOX2	-
LIN28A	-38/-43	SOX17/SOX2	-
	+8355/+8338	SOX17/ SOX2	-
TRIM71	-1494/-1503	SOX17/SOX2	-
	+5198/+5207	SOX17/ SOX2	Canonical

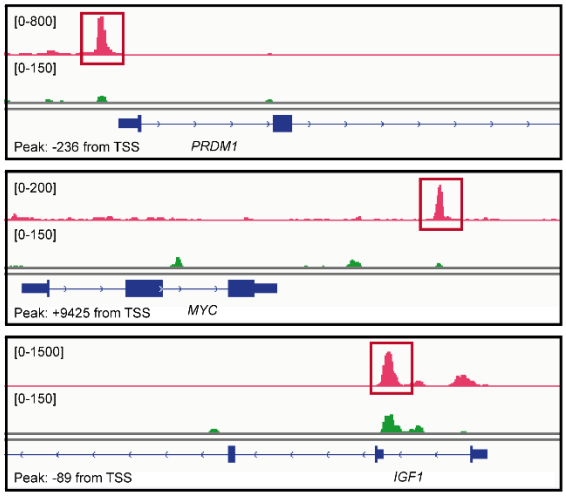
Target Gene	Distance to TSS [bp]	Transcription Factor	Description
GDF3	+1045	SOX2	Canonical
LEFTY2	-4363	SOX2	-
SALL4	-9206	SOX2	-
	+2767	SOX2	-
SOX2	-3749	SOX2	Canonical
	+4222	SOX2	Canonical
TP53	-2133	SOX2	-
POU5F1	-1361	SOX2	-

Target Gene	Distance to TSS [bp]	Transcription Factor	Description
PRDM1	-236	SOX17	Compressed
BMP7	+8355	SOX17	-
BMP8a	+9313	SOX17	-
MYC	+9425	SOX17	Compressed
IGF1	-89	SOX17	Compressed

C



D



E

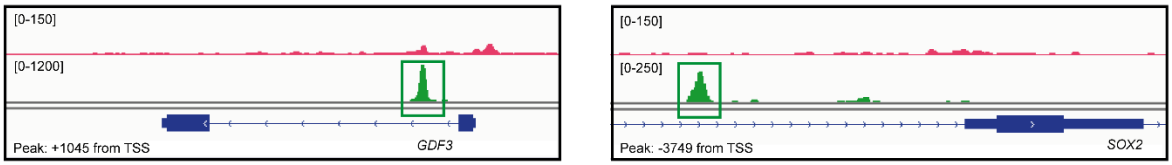


Figure 3: (A) Venn diagram depicting the number of genes bound by SOX17, SOX2 or SOX17 and SOX2 (binding site within a distance +/- 10,000 from TSS) in TCam-2 and 2102EP, respectively. Genes representative for each group are listed. **(B)** Description of binding sites (relative to TSS) of SOX17 and SOX2 in the regulatory regions of the genes listed in (A). **(C-E)** DNA binding profiles of SOX17 and SOX2 at canonical motifs (green boxes) of *PIM2*, *PRDM14*, *TRIM71* **(C)**, at compressed motifs (red boxes) of *PRDM1*, *MYC*, *IGF1* **(D)** and at canonical motifs (green boxes) of *GDF3*, *SOX2* **(E)**. Distance of binding sites to TSS are indicated in lower left corner. Scales (upper left corner) were adjusted to allow for best comparison.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Germ cell-related Function of SOX17 in Seminoma Cells

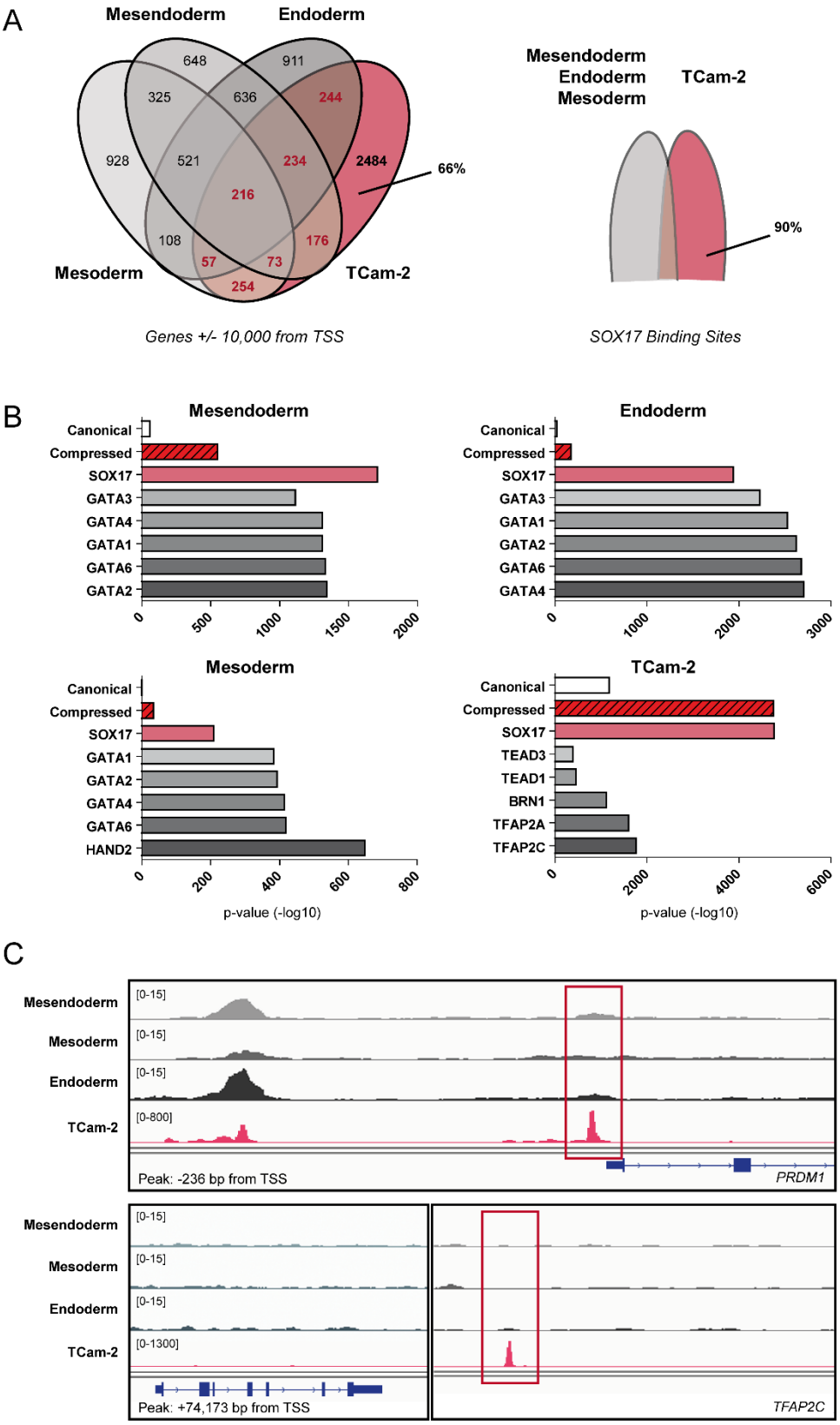
Next, we asked how the role of SOX17 in seminomas is different to the role of SOX17 as endodermal factor. Detailed analysis revealed that from 3,738 genes bound by SOX17 in TCam-2 (10,000 bp +/- from TSS), 66% are exclusively found in TCam-2, but not in mesoderm-, endoderm- or mesendoderm-differentiated ESCs (10,000 bp +/- from TSS) [26] (**Fig. 4A**). Even more strikingly, 90% of the *bona fide* SOX17 binding sites identified in TCam-2 were not bound by SOX17 in differentiated ESCs [26] (**Fig. 4A**). This shows that the DNA binding sites of SOX17 differ dramatically between TCam-2 and somatic lineages. The most prominent difference is that in TCam-2 we found strong enrichment of SOX17 at the canonical motif, which was absent in somatic cells (**Fig. 4B**). Surprisingly, also enrichment for the compressed motif was much stronger in TCam-2 cells compared to differentiated ESCs (**Fig. 4B**). This shows that SOX17 fulfills its function in seminoma cells not only via the canonical, but also the compressed motif.

But how does the role of SOX17 in seminomas compare to the role of SOX17 in PGCs? Previous studies demonstrated that human PGC specification is governed by a transcription factor network of SOX17, TFAP2C and PRDM1, where SOX17 transactivates *TFAP2C* and *PRDM1* [10,37,38]. In TCam-2, we found that SOX17 binds to a compressed motif within the *PRDM1* promoter, which is not bound by SOX17 in somatic lineages (**Fig. 4C**). Surprisingly, in TCam-2 *TFAP2C* was not bound by SOX17 +/- 10,000 bp from TSS. However, we found in total six binding sites of SOX17 in distal regulatory regions (-19,773; +35,685; +74,173; +124,640; +175,507; +268,206) upstream and downstream of the *TFAP2C* TSS (**Supp. Data 1 upon request**). The highest ChIP-seq peak score was calculated for a compressed motif, which lies closest to the *TFAP2C* TSS (+74,173). Again, this binding site was not bound by SOX17 in somatic lineages (**Fig. 4C**). Thus, in seminomas the transcription factor network (SOX17-PRDM1-TFAP2C) described in human PGC specification seems to be in place, but absent in differentiated ESCs.

Figure 4: (A) Left: Venn diagram depicting the number of genes putatively regulated by SOX17 (SOX17 binding site within a distance +/- 10,000 from TSS) in mesendoderm-, endoderm- and mesoderm-differentiated ESCs, as well as TCam-2. Right: Venn diagram depicting the overlap of SOX17 binding sites in TCam-2 cells with SOX17 in mesendoderm-, endoderm- and/or mesoderm-differentiated ESCs. **(B)** Transcription factor binding motifs enriched within SOX17 ChIP-seq peaks of mesendoderm-, endoderm- and mesoderm-differentiated ESCs, as well as TCam-2, as determined by HOMER [13]. **(C)** Upper panel: DNA binding profile of SOX17 upstream of *PRDM1* in mesendoderm-, endoderm- and mesoderm-differentiated ESCs, as well as TCam-2. Red box indicates the compressed motif bound by SOX17 in TCam-2 cells. Lower panel: DNA binding profile of SOX17 downstream of *TFAP2C* in mesendoderm-, endoderm- and mesoderm-differentiated ESCs, as well as TCam-2. Red box indicates the compressed motif bound by SOX17 in TCam-2 cells. Distance of binding sites to TSS are indicated in lower left corner. Scales (upper left corner) were adjusted to allow for best comparison.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Loss of SOX17 in Seminoma Cells results in Differentiation

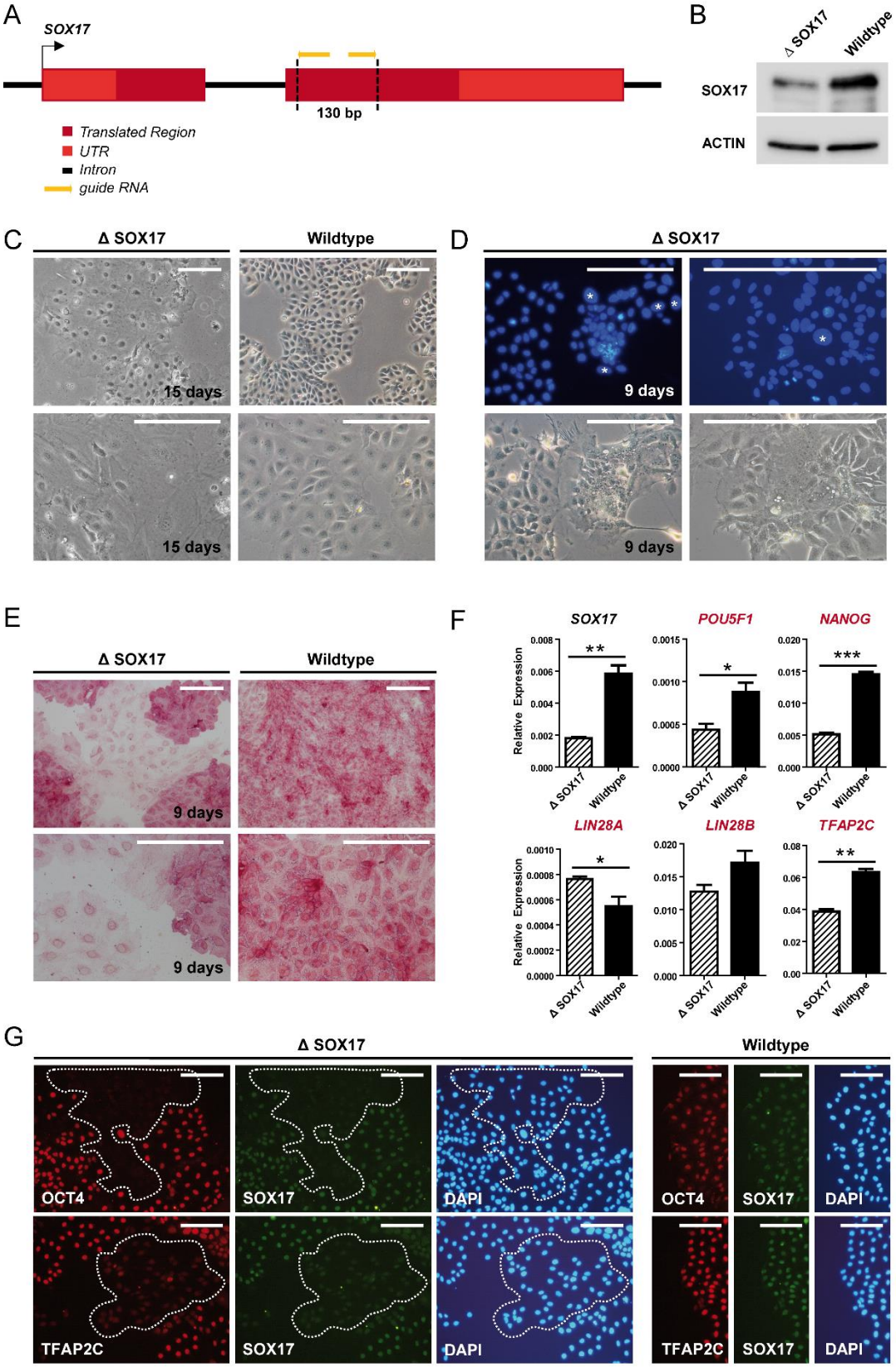
To test whether SOX17 regulates latent pluripotency of seminomas, we used CRISPR/Cas9-mediated gene-editing to delete a region of approximately 130 bp within the *SOX17* coding sequence (**Fig. 5A**). *SOX17* is encoded on chromosome [8], which has six copies in TCam-2 cells [39]. Therefore, gene editing of the *SOX17* locus in TCam-2 cells results in a mixed population of cells having one to six alleles of *SOX17* deleted. The loss of SOX17 protein in the gene-edited TCam-2 bulk population (referred to as TCam-2$^{\Delta SOX17}$) was shown by Western blot analysis (**Fig. 5B**). TCam-2$^{\Delta SOX17}$ cells show morphologically distinct areas suggestive of differentiation (**Fig. 5C**). They display polynucleated trophoblast giant cell morphology (**Fig. 5D**), which is reminiscent of TCam-2 cells differentiated by treatment with TGFβ, EGF and FGF4 [40]. In these morphologically distinct areas, TCam-2$^{\Delta SOX17}$ are negative for alkaline phosphatase (AP) activity, indicating loss of pluripotency (**Fig. 5E**). qRT-PCR analysis confirmed a downregulation of pluripotency and / or seminoma markers (*POU5F1, TFAP2C, NANOG* and *LIN28B* (but not *LIN28A*)) in TCam-2$^{\Delta SOX17}$ cells (**Fig. 5F**). Immunofluorescence staining confirmed reduction of OCT4 and TFAP2C protein levels in those cells that display reduced levels or complete loss of SOX17 protein (**Fig. 5G**). Further, we detected downregulation of additional GC- and / or pluripotency-associated markers (*PRDM14, PRDM1, KIT, NANOS3, SPRY4, ALPL*), independent of whether they were identified as direct SOX17 targets or not (**Supp. Fig. 5A**).

Next, we asked into which lineage or cell type TCam-2$^{\Delta SOX17}$ differentiate. Since SOX17 is known as a master regulator of the endodermal lineage, differentiation into the endoderm lineage seems highly unlikely. The fact that endodermal-differentiation factors *FOXA1* or *FOXA2* were not upregulated further supported this notion (**Supp**. **Fig. 5B**) 41. In addition, mesoderm or ectoderm differentiation markers were not upregulated (except for mild induction of *PAX6*), nor was SOX2 (**Supp**. **Fig. 5B**). Therefore, TCam-2 cells do not differentiate into somatic lineages upon loss of SOX17. Due to the resemblance of TCam-2$^{\Delta SOX17}$ cells to trophoblast giant cells we screened for expression of extra-embryonic lineage markers. Loss of SOX17 resulted in significant upregulation of the trophoblast giant cell marker *HAND1,* as well as *GATA3, GATA6* and *αHCG* (**Supp**. **Fig. 5C**). *CDX2* and *EOMES* levels remained unchanged (**Supp**. **Fig. 5C**). Upregulation of GATA3 was confirmed by immunofluorescence staining (**Supp**. **Fig. 5D**). Together this suggests that downregulation of SOX17 results in extra-embryonic differentiation.

Figure 5: (A) CRISPR/Cas9-gene editing strategy of SOX17 locus in TCam-2 cells. **(B)** Immunoblot of SOX17 in gene-edited TCam-2$^{\Delta SOX17}$ cells. β-ACTIN (ACTIN) served as loading control. **(C)** Morphology of TCam-2$^{\Delta SOX17}$ cells (left) and TCam-2 wildtype cells (right) 15 days following gene-editing. Scale bar = 250 µm. **(D)** Hoechst staining of TCam-2$^{\Delta SOX17}$ cells (upper panel) and corresponding brightfield images (lower panel) 9 days following gene-editing. Multi-nucleated cells are indicated with an asterisk. Scale bar = 250 µm. **(E)** Alkaline phosphatase staining of TCam-2 wildtype cells and TCam-2$^{\Delta SOX17}$ cells 9 days following gene-editing. Scale bar = 250 µm. **(F)** qRT-PCR of selected germ cell markers and SOX17 target genes (red labels) in TCam-2$^{\Delta SOX17}$ cells and TCam-2 cells transiently transfected with a GFP-coding plasmid as control (wildtype), 72 h following transfection. Expression was normalized to *GAPDH* as housekeeper. Significance was calculated using two-tailed students t-test (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$). Error bars indicate standard deviation from the mean. **(G)** Immunofluorescence staining of OCT4, TFAP2C and SOX17 protein in TCam-2$^{\Delta SOX17}$ cells and TCam-2 wildtype cells. Cell nuclei were stained with DAPI. Scale bar = 125 µm.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5



bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DISCUSSION

Here, we have investigated the binding-patterns and target genes of SOX17 and SOX2 in seminoma and EC cell lines, respectively. We demonstrated that both transcription factors bind to canonical motifs by partnering with OCT4, as well as to SOX-family DNA binding motifs. This way SOX17 and SOX2 regulate a common set of pluripotency and germ cell-related genes (*PRDM14, DPPA4, TDGF1, NANOG, LIN28A, TRIM71, OTX2, PIM2*) (**Fig. 6**). Additionally, in TCam-2 cells SOX17 binds to compressed motifs or SOX motifs (not bound by SOX2 in ECs), thereby regulating the PGC specifiers *PRDM1* and *TFAP2C,* BMP members *BMP7* and *BMP8a* and the cancer-related genes *MYC* and *IGF1* (**Fig. 6**). In 2102EP EC cells, SOX2 further binds canonical elements or SOX motifs (not bound by SOX17 in TCam-2), thereby regulating additional pluripotency genes (*GDF3, LEFTY2, SALL4, SOX2* and *POU5F1*) (**Fig. 6**). In contrast to SOX17, SOX2 did not bind to compressed motifs. This is in line with the previous observation of Jauch *et al.* that due to sterical hindrance the SOX2/OCT4 complex is occluded from the compressed motif [42]. Our findings suggest that in seminomas and probably PGCs, SOX17 is able to functionally replace SOX2 in its role in maintaining pluripotency.

The role of SOX2 in transcriptional regulation was already well-described for ESCs. Here, we have shown that SOX2 in EC cells regulates similar target genes as SOX2 in ESCs, hence it is a key regulator of pluripotency genes. In addition to already well-established markers of pluripotency, we found *PIM2* and *OTX2* to be SOX2 target genes in ECs. Overexpression of *PIM1* and *PIM2* in mice resulted in an increase in inflammatory signatures as well as an overall higher cell proliferation rate in testicular tissues [33]. Both *PIM2* and *OTX2* are overexpressed in TGCT, but not much is known about their role in human TGCTs. So, future studies will address role of *PIM1/2* and *OTX2* for human TGCT development in more detail.

SOX17 is known as a regulator of endodermal cell fate decisions in ESCs and as a key specifier of human PGC fate. Here we demonstrate that only 10% of the SOX17 binding sites in TCam-2 are shared by SOX17 binding sites in differentiated (endoderm-, mesoderm-, mesendoderm-) lineages. The canonical motif was not found among SOX17 binding elements in differentiated lineages. This suggests that SOX17 binding to canonical motifs is specific to seminoma cells (and possibly PGCs). This is a highly important finding, as it demonstrates that the DNA-occupancy of SOX17 is strictly dependent on the cell type analysed. We speculate that either cofactors and / or locus accessibility regulated by epigenetic changes such as DNA methylation or histone modifications might guide SOX17 to its respective target gene. Since seminoma cells highly express pluripotency-associated genes, we predicted that SOX17 transactivates these genes via the canonical motif [21]. Indeed, we found that in seminoma cells SOX17 binds to canonical motifs allocated within regulatory regions of pluripotency genes (such as *PRDM14, TRIM71*). Additionally, SOX17 binds compressed motifs within the regulatory regions of GC- and cancer-associated genes (such as *IGF1, MYC, TFAP2C, PRDM1, PRDM14*). Therefore, partnering with OCT4 is required in order to fulfil the GC- and cancer-specific function of SOX17 in seminoma cells and to maintain latent pluripotency (**Fig. 6**).

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6

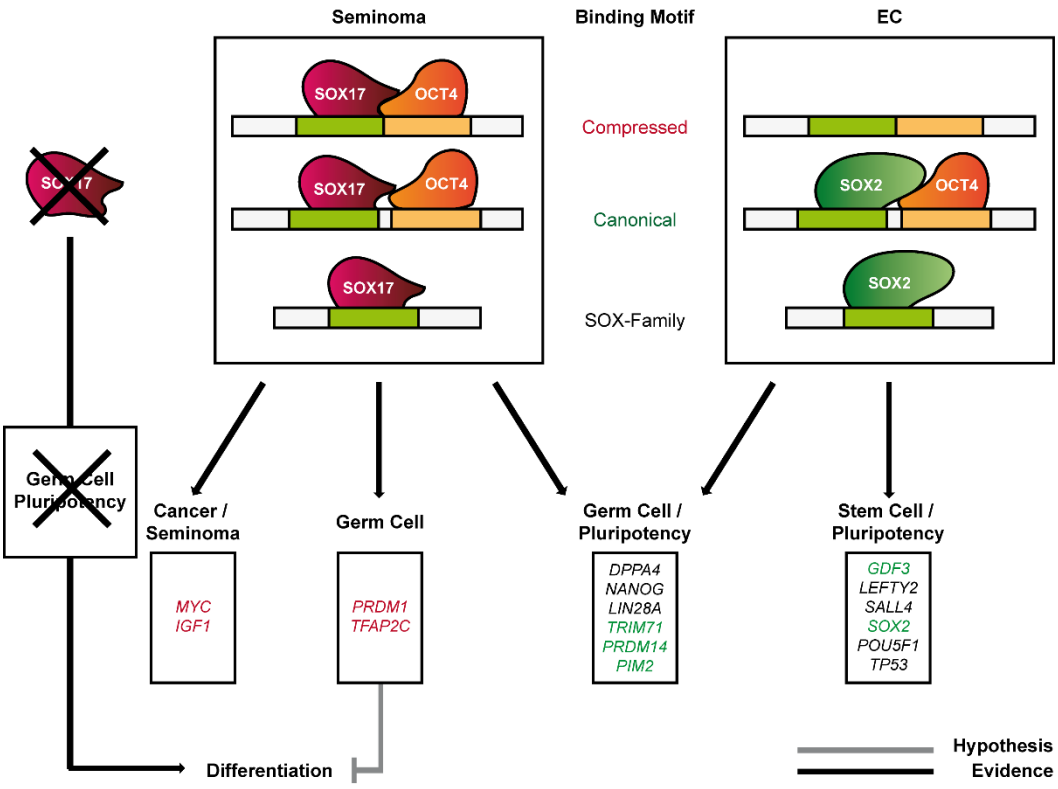


Figure 6: Schematic representation of the role of SOX17 in seminomas (left) and SOX2 in ECs (right). Genes regulated via compressed motifs are written in red labels, genes regulated via canonical motifs are written in green labels. Genes regulated via solitaire SOX motifs are written in black labels.

Thus, the role of SOX17 in seminomas seems to be reminiscent of the role of SOX17 in human PGCs, which is to maintain latent pluripotency and to suppress somatic differentiation via transactivation of *TFAP2C* and *PRDM1*. In human PGCs, SOX17, TFAP2C and PRDM1 form a tripartite transcription factor network, which prohibits mesoderm and neural differentiation and maintains expression of pluripotency- and germ cell-associated genes. Irie *et al.* proposed SOX17 as the most upstream transcription factor within this network [10]. Later, Kobayashi *et al*. showed that SOX17 and TFAP2C have a pivotal role in specification of human PGCs and that SOX17 stimulates *PRDM1* expression [35]. In the end, in PGC-like cells all three transcription factors form a self-regulatory network including feedback loops regulating their own expression [35,38]. Interestingly, PRDM14 was recently introduced as a fourth factor (next to SOX17, TFAP2C and PRDM1) maintaining human PGC pluripotency [43]. A different study further suggested that an aberrant overexpression of *PRDM14* during PGC development could favour a malignant transformation of the cells [44]. We here demonstrate that in seminoma cells SOX17 binds multiple elements within *PRDM14* regulatory regions. Thus, SOX17 not only regulates the GC network in seminoma cells by induction of *PRDM1* and *TFAP2C*, but additionally via *PRDM14* expression. Additionally, SOX17 contributes to the latent pluripotency state of seminoma cells by directly transactivating a set of pluripotency-related genes (such as *DPPA4, NANOG, LIN28A, TRIM71*) (**Fig.**

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6). Collectively, these findings highlight the role of SOX17 in the germline and demonstrate the dependency of seminoma cells for SOX17-driven transactivation of pluripotency- and GC-associated genes.

ACKNOWLEDGEMENTS

We kindly thank Blanca Randel, Gaby Beine, Angela Egert and Andrea Jäger for technical assistance. Further, we thank the NGS Core Facility of the University Hospital Bonn for the generation of the sequencing data and the Core Unit for Bioinformatic Analysis of the University Hospital Bonn for their assistance with the data analysis.

FUNDING

This work was supported by the Wilhelm-Sander-Stiftung [grant number 2016.088.1 to H.S. / D.N.] and the DFG (Scho 503/16-1 to H. S.)

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Cheng L, Albers P, Berney DM, Feldman DR, Daugaard G, Gilligan T, Looijenga LHJ. Testicular cancer. *Nat Rev Dis Primers* 2018;4:29.

2. Berney DM, Looijenga LHJ, Idrees M, Oosterhuis JW, Meyts ER-D, Ulbright TM, Skakkebaek NE. Germ cell neoplasia in situ (GCNIS): evolution of the current nomenclature for testicular pre-invasive germ cell malignancy. *Histopathology* 2016;69:7–10.

3. Nettersheim D, Jostes S, Schneider S, Schorle H. Elucidating human male germ cell development by studying germ cell cancer. *Reproduction* 2016;152:R101–13.

4. de Jong J, Stoop H, Gillis AJM, van Gurp RJHLM, van de Geijn G-JM, de Boer M, Hersmus R, Saunders PTK, Anderson RA, Oosterhuis JW, Looijenga LHJ. Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications. *The Journal of pathology* 2008;215:21–30.

5. Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* 2013;12:15–30.

6. Nettersheim D, Heimsoeth A, Jostes S, Schneider S, Fellermeyer M, Hofmann A, Schorle H. SOX2 is essential for *in vivo* reprogramming of seminoma-like TCam-2 cells to an embryonal carcinoma-like fate. *Oncotarget* 2016;7:47095.

7. Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Ng CKL, Herath W, Lili S, Hutchins AP, Robson P, et al. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *The EMBO journal* 2013;32.

8. Tapia N, Maccarthy C, Esch D, Marthaler AG, Tiemann U, Araúzo-bravo MJ, Jauch R, Cojocaru V, Schöler HR. Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Nature Publishing Group* 2015;

9. Stefanovic S, Abboud N, Désilets S, Nury D, Cowan C, Pucéat M. Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *J Cell Biol* 2009;186:665–73.

10. Irie N, Weinberger L, Tang WWC, Kobayashi T, Viukov S, Manor YS, Dietmann S, Hanna JH, Surani MA. SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell* 2015;160.

11. Nettersheim D, Fronhoffs F, Haas N, Waha A, Honecker F, Waha A, Kristiansen G, Schorle H. Analysis of TET Expression/Activity and 5mC oxidation during normal and malignant germ cell development. *PloS one* 2013;8.

12. Nettersheim D, Haas N, Leinhaas A, Schlatt S, Schorle H. Establishment of a versatile seminoma model indicates cellular plasticity of germ cell tumor cells. *Genes Chromosomes and Cancer* 2012;51.

13. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* 2010;38:576–89.

14. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nature structural & molecular biology* 2016;23:349–57.

15. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40.

16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:15545–50.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

17. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007;

18. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.

19. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics* 2012;14:178–92.

20. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* 2001;30:207–10.

21. Nettersheim D, Arndt I, Sharma R, Riesenberg S, Jostes S, Schneider S, Hölzel M, Kristiansen G, Schorle H. The cancer/testis-antigen PRAME supports the pluripotency network and represses somatic and germ cell differentiation programs in seminomas. *British journal of cancer* 2016;115:454–64.

22. Nettersheim D, Jostes S, Sharma R, Schneider S, Hofmann A, Ferreira HJ, Hoffmann P, Kristiansen G, Esteller MB, Schorle H. BMP Inhibition in Seminomas Initiates Acquisition of Pluripotency via NODAL Signaling Resulting in Reprogramming to an Embryonal Carcinoma. *PLoS genetics* 2015;11:e1005415.

23. Nettersheim D, Jostes S, Fabry M, Honecker F, Schumacher V, Kirfel J, Kristiansen G, Schorle H. A signaling cascade including ARID1A, GADD45B and DUSP1 induces apoptosis and affects the cell cycle of germ cell cancers after romidepsin treatment. *Oncotarget* 2016;5.

24. Jostes S, Nettersheim D, Fellermeyer M, Schneider S, Hafezi F, Honecker F, Schumacher V, Geyer M, Kristiansen G, Schorle H. The bromodomain inhibitor JQ1 triggers growth arrest and apoptosis in testicular germ cell tumours in vitro and in vivo. *J Cell Mol Med* 2016;21:1300–14.

25. van der Zwan YG, Rijlaarsdam MA, Rossello FJ, Notini AJ, de Boer S, Watkins DN, Gillis AJM, Dorssers LCJ, White SJ, Looijenga LHJ. Seminoma and embryonal carcinoma footprints identified by analysis of integrated genome-wide epigenetic and expression profiles of germ cell cancer cell lines. *PloS one* 2014;9:e98330.

26. Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, Gnirke A, Meissner A. Transcription factor binding dynamics during human ES cell differentiation. *Nature* 2015;518:344–9.

27. Hart AH, Hartley L, Parker K, Ibrahim M, Looijenga LHJ, Pauchnik M, Chow CW, Robb L. The pluripotency homeobox gene NANOG is expressed in human germ cell tumors. *Cancer* 2005;104:2092–8.

28. Qian X, Kim JK, Tong W, Villa-Diaz LG, Krebsbach PH. DPPA5 Supports Pluripotency and Reprogramming by Regulating NANOG Turnover. *Stem cells (Dayton, Ohio* 2015;34:588–600.

29. Levine AJ, Brivanlou AH. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. *Development* 2005;133:209–16.

30. Fiorenzano A, Pascale E, D'Aniello C, Acampora D, Bassalert C, Russo F, Andolfi G, Biffoni M, Francescangeli F, Zeuner A, Angelini C, Chazaud C, et al. Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency. *Nat Commun* 2016;7:12589.

31. Zhang J, Zhang M, Acampora D, Vojtek M, Yuan D, Simeone A, Chambers I. OTX2 restricts entry to the mouse germline. *Nature* 2018;562:595–9.

32. Eckert D, Biermann K, Nettersheim D, Gillis AJM, Steger K, Jäck H-M, Müller AM, Looijenga LHJ, Schorle H. Expression of BLIMP1/PRMT5 and concurrent histone H2A/H4 arginine 3 dimethylation in fetal germ cells, CIS/IGCNU and germ cell tumors. *BMC developmental biology* 2008;8:106.

33. Jiménez-García MP, Lucena-Cacace A, Robles-Frías MJ, Narlik-Grassow M, Blanco-Aparicio C, Carnero A. The role of PIM1/PIM2 kinases in tumors of the male reproductive system. *Sci Rep* 2016;6:38079.

bioRxiv preprint doi: https://doi.org/10.1101/643114; this version posted May 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

34. Chen H, Liu H, Qing G. Targeting oncogenic Myc as a strategy for cancer treatment. *Cancer* 2018;3:5.

35. Kobayashi T, Zhang H, Tang WWC, Irie N, Withey S, Klisch D, Sybirna A, Dietmann S, Contreras DA, Webb R, Allegrucci C, Alberio R, et al. Principles of early human development and germ cell program from conserved model systems. *Nature* 2017;546:416–20.

36. Puglisi R, Montanari M, Chiarella P, Stefanini M, Boitani C. Regulatory role of BMP2 and BMP7 in spermatogonia and Sertoli cell proliferation in the immature mouse. *Eur J Endocrinol* 2004;151:511–20.

37. Sasaki K, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, Ohta H, Moritoki Y, Iwatani C, Tsuchiya H, Nakamura S, Sekiguchi K, et al. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* 2015;17:178–94.

38. Kojima Y, Sasaki K, Yokobayashi S, Sakai Y, Nakamura T, Yabuta Y, Nakaki F, Nagaoka S, Woltjen K, Hotta A, Yamamoto T, Saitou M. Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. *Cell Stem Cell* 2017;21:517–532.e5.

39. De Jong J, Stoop H, Gillis AJM, Hersmus R, Van Gurp RJHLM, Van De Geijn GJM, Van Drunen E, Baeten J, Kitazawa S, Van Roozendaal K, Looijenga LHJ. Further characterization of the first seminoma cell line TCam-2. *Genes Chromosomes and Cancer* 2008;47.

40. Nettersheim D, Gillis AJ, Looijenga LH, Schorle H. TGF-β1, EGF and FGF4 synergistically induce differentiation of the seminoma cell line TCam-2 into a cell type resembling mixed non-seminoma. *International journal of andrology* 2011;34.

41. Wang P, McKnight KD, Wong DJ, Rodriguez RT, Sugiyama T, Gu X, Ghodasara A, Qu K, Chang HY, Kim SK. A molecular signature for purified definitive endoderm guides differentiation and isolation of endoderm from mouse and human embryonic stem cells. *Stem cells and development* 2012;21:2273–87.

42. Jauch R, Aksoy I, Hutchins AP, Ng CKL, Tian XF, Chen J, Palasingam P, Robson P, Stanton LW, Kolatkar PR, Merino F, Merino F. Conversion of Sox17 into a Pluripotency Reprogramming Factor by Reengineering Its Association with Oct4 on DNA. *Stem Cells* 2011;

43. Sybirna A, Tang W, Dietmann S, Gruhn WH. A critical but divergent role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. *biorxivorg* 2019;

44. Gell JJ, Zhao J, Hunt TJ, Clark AT. PRDM14 is expressed in germ cell tumors with constitutive overexpression altering human germline differentiation and proliferation. *Stem Cell Res* 2018;27:46–56.

45. Chakravarthy H, Boer B, Desler M, Mallanna SK, McKeithan TW, Rizzino A. Identification of DPPA4 and other genes as putative Sox2:Oct-3/4 target genes using a combination of in silico analysis and transcription-based assays. *J Cell Physiol* 2008;216:651–62.

46. Cimadamore F, Amador-Arjona A, Chen C, Huang C-T, Terskikh AV. SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. *Proceedings of the National Academy of Sciences* 2013;110:E3017–26.