

RESEARCH ARTICLE

HSA21 Single-Minded 2 (Sim2) Binding Sites Co-Localize with Super-Enhancers and Pioneer Transcription Factors in Pluripotent Mouse ES Cells

Audrey Letourneau¹, Gilda Cobellis², Alexandre Fort³, Federico Santoni¹, Marco Garieri¹, Emilie Falconnet¹, Pascale Ribaux¹, Anne Vannier^{1,4}, Michel Guipponi^{1,4}, Piero Carninci³, Christelle Borel^{1*}, Stylianos E. Antonarakis^{1,4,5*}

1 Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland, **2** Department of Biophysics, Biochemistry and General Pathology, Seconda Università di Napoli, Napoli, Italy, **3** Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan, **4** University Hospitals of Geneva, Geneva, Switzerland, **5** iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland

* stylianos.antonarakis@unige.ch (SEA); christelle.borel@unige.ch (CB)



OPEN ACCESS

Citation: Letourneau A, Cobellis G, Fort A, Santoni F, Garieri M, Falconnet E, et al. (2015) HSA21 Single-Minded 2 (Sim2) Binding Sites Co-Localize with Super-Enhancers and Pioneer Transcription Factors in Pluripotent Mouse ES Cells. PLoS ONE 10(5): e0126475. doi:10.1371/journal.pone.0126475

Academic Editor: Jason Glenn Knott, Michigan State University, UNITED STATES

Received: December 10, 2014

Accepted: April 2, 2015

Published: May 8, 2015

Copyright: © 2015 Letourneau et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequencing data are available at the GEO database under the accession number GSE59379.

Funding: This work was supported by grants from the Swiss National Science Foundation, and the European ERC to SEA, and the Lejeune Foundation to CB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The HSA21 encoded Single-minded 2 (SIM2) transcription factor has key neurological functions and is a good candidate to be involved in the cognitive impairment of Down syndrome. We aimed to explore the functional capacity of SIM2 by mapping its DNA binding sites in mouse embryonic stem cells. ChIP-sequencing revealed 1229 high-confidence SIM2-binding sites. Analysis of the SIM2 target genes confirmed the importance of SIM2 in developmental and neuronal processes and indicated that SIM2 may be a master transcription regulator. Indeed, SIM2 DNA binding sites share sequence specificity and overlapping domains of occupancy with master transcription factors such as SOX2, OCT4 (Pou5f1), NANOG or KLF4. The association between SIM2 and these pioneer factors is supported by co-immunoprecipitation of SIM2 with SOX2, OCT4, NANOG or KLF4. Furthermore, the binding of SIM2 marks a particular sub-category of enhancers known as super-enhancers. These regions are characterized by typical DNA modifications and Mediator co-occupancy (MED1 and MED12). Altogether, we provide evidence that SIM2 binds a specific set of enhancer elements thus explaining how SIM2 can regulate its gene network in neuronal features.

Introduction

Down syndrome (DS) results from trisomy of human chromosome 21 (T21). It is the most frequent live-born aneuploidy, affecting 1 in 750 newborns. DS patients are characterized by a broad range of phenotypes including mental retardation, short stature, muscle hypotonia, congenital heart defects or Alzheimer disease neuropathology [1]. Among the HSA21 genes,

transcription factors are important candidates to explain some DS features. Indeed, transcription factors are known to play a global role in the gene transcription regulation via their direct or indirect binding to promoter and enhancer elements. Consequently, their dysregulation (in trisomic cells for instance) is likely to impact the expression of the target genes leading to the perturbation of a variety of distinct molecular pathways. More than 20 transcription factors or transcription regulators map on HSA21 and may directly or indirectly contribute to the transcriptional regulation [2]. Among them, Single-minded 2 (SIM2) appears to be a relevant candidate to explain some DS features, in particular the cognitive impairment.

SIM2 is a member of the basic helix-loop-helix Per-Arnt-Sim (bHLH/PAS) family of transcription factors. The proteins of this family contain a basic DNA binding domain adjacent to a helix-loop-helix region and a PAS region, essential for the dimerization of the proteins and the proper formation of active transcription factor complexes [3]. They are known to be involved in multiple fundamental biological processes including neurogenesis, hypoxic response, circadian rhythms or toxin metabolism [3–5]. The first single-minded protein was identified in *Drosophila melanogaster* as a key regulator of the midline cell development in the central nervous system (CNS) [6–8]. Interestingly, the *Drosophila Sim* does not only contribute to gene activation in the midline cells [7] but also to indirect gene repression in the lateral CNS, through activation of repressive factors [9, 10]. To form active complexes, the *Drosophila Sim* protein dimerizes with another member of the bHLH-PAS family called Tango [11]. The SIM proteins identified in mammals show a high degree of similarities with their *Drosophila* homolog [12–16]. They contain comparable bHLH and PAS domains and dimerize with the Tango ortholog called ARNT (Ah receptor nuclear translocator). The presence of ARNT is essential for the formation of active complexes since SIM2 does not homodimerize [3]. The murine *Sim2* is expressed early during development in many tissues affected in DS such as developing forebrain, ribs, vertebrae, limb skeletal muscles or kidney [17]. Similarly, the human *SIM2* is expressed during the early fetal life in the central nervous system and in key brain structures involved in learning and memory processes [15, 18]. The expression pattern of *SIM2* and its known function in *Drosophila* suggest that it may be a good candidate to explain some of the DS cognitive features. Interestingly, the transgenic mice harboring three copies of *Sim2* exhibit some of the DS phenotypes namely a moderate impairment of learning and memory as well as a reduced exploratory behavior and sensitivity to pain [19–21]. *Sim2*^{-/-} mutant mice die rapidly after birth due to breathing failure and display rib, vertebral and craniofacial abnormalities [22, 23].

In order to better understand how SIM2 can participate in some DS features, we have further explored its regulatory role in mammalian cells. An accurate list of SIM2 target genes in normal and trisomic cells is required for understanding its role in genetic regulation. We have mapped the SIM2 DNA binding sites using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) in a mouse embryonic stem cell (mES cell) line that stably overexpresses *Sim2* under the control of an inducible system. Using this model, we have identified 1229 regions occupied by SIM2 and showed that the associated target genes fulfill molecular functions related to the DS phenotypes. More importantly, we observed that a significant fraction of SIM2 binding sites overlaps with genomic regions occupied by master transcription factors involved in the genetic control of the pluripotent state, namely SOX2, OCT4, NANOG or KLF4. These regions are characterized by typical enhancer signatures and our data demonstrate that the binding of SIM2 could also predict a super-enhancer activity. Altogether, we provide new evidence of the SIM2 functional capacity.

Materials and Methods

Cell culture

The EBRTcH3 (EB3) parental mouse ES line and the clones overexpressing Sim2-Flag have been previously described in [24, 25].

mES cells were grown on 0.1% gelatin (Sigma #G1890) coated dishes in DMEM high glucose medium (Life technologies #41965) supplemented with 15% Fetal Bovine Serum (FBS HyClone, Thermo Scientific #SH30070), 2mM L-glutamine (Life technologies #25030), 1mM Sodium pyruvate (Life technologies #11360), 100units/ml penicillin/streptomycin (Bioconcept #4-01F00-H), 0.1mM 2-mercaptoethanol (Life technologies #31350), 1000units/ml Leukemia Inhibitory Factor (LIF, Millipore #ESG1107) and 1 μ g/ml tetracycline (Sigma #T7660). Cells were incubated at 37°C in a 5% CO₂ atmosphere. Medium was changed every day and cells were passed every 1 or 2 days using 1X Trypsin-EDTA (Sigma #T4174).

Induction of *Sim2* transgene expression

Culture medium was changed for tetracycline-free medium three hours before passing the cells in order to eliminate the residual tetracycline. Cells were passed using 1X Trypsin-EDTA (Sigma #T4174). Five million cells were plated in each new dish and cultured in the tetracycline-free medium for 26h starting from passage time.

Fluorescence Activated Cell Sorting

Cells were grown in presence or absence of tetracycline for 26 hours and collected using Trypsin-EDTA. Pellets were washed with PBS and 300'000 cells from each line were collected in 300 μ l of PBS supplemented with 2% FBS for the measure of Venus fluorescence by FACS (FACSCalibur platform).

RNA preparation

Total RNA was isolated 26h post induction, concurrently with the crosslinking experiment. RNA samples were prepared using TRIzol reagent (Life technologies #15596) as per the manufacturer's instructions. Quality was assessed using the Agilent 2100 Bioanalyzer (RNA 6000 Nano Kit, #5067) and quantity was measured on a Qubit instrument (Life technologies). RNA was extracted from each of the SIM2 clones (A6, B8 and C4) and from three independent cultures of the EB3 clone.

Reverse-Transcription PCR

Reverse transcription was performed using 1 μ g of total RNA and the SuperScript II Reverse Transcriptase (Life technologies #18064). PCR was performed on 1 μ l of cDNA diluted 10 times using the following primers: TTCGAATGAAGTGCGTCTTG (forward) and ACATGTGCTGTGGAGCTTG (reverse) for mSim2 and TGCCTCATCTGGTACTGCTG (forward) and GAACATGCTGCTCACTGGAA (reverse) for mArnt. The PCR program was the following: 94°C for 5min followed by 10 cycles of 94°C for 30s, 60°C (Δ -1) for 30s, 72°C for 30s, followed by 25 cycles of 94°C for 30s, 50°C for 30s, 72°C for 30s and a final elongation step at 72°C for 7min.

Chromatin immunoprecipitation (ChIP)

ChIP was performed using the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling #9003) according to manufacturer's instructions. Briefly, mES cells SIM2 (A6, B8 and C4) and

mES cells EB3 were cultured in absence of tetracycline for 26h. After the induction, 50 million cells from each clone were crosslinked using 1% formaldehyde. Digestion of the chromatin was performed with 5 μ l of Micrococcal Nuclease for 20 minutes at 37°C, followed by sonication (3 sets of 10-second pulses at 10% amplitude on a Branson Digital Sonifier 450). Independent chromatin immunoprecipitations were performed on SIM2 (A6, B8 and C4) and EB3 chromatin preparations using the equivalent of 20 μ g of chromatin DNA per IP. Each chromatin preparation was incubated at 4°C overnight with 6 μ g of anti-FLAG M2 antibody (Sigma #F3165). The next day, samples were incubated with 30 μ l of Protein G Magnetic Beads for 2h at 4°C, beads were washed and chromatin was eluted from the antibody/Protein G beads complexes. A 2% input sample of each chromatin preparation was saved before the immunoprecipitation as a control. Both input and eluted chromatin samples were reverse crosslinked in presence of Proteinase K for 2 hours at 65°C and DNAs were purified on columns.

ChIP experiments against OCT4, SOX2, NANOG, MED1 and MED12 were performed in the same conditions (same cross-linked pellets) on SIM2 A6 cells and EB3 cells using 6 μ g of the following antibodies: anti-Nanog (D2A3) XP (Cell Signaling #8822), anti-Oct3/4 (N-19) (Santa Cruz #sc-8628), anti-Sox2 (Santa Cruz #sc-17320), anti-Med1 (CRSP1/TRAP220) (Bethyl #A300-793A) and anti-Med12 (Bethyl #A300-774A).

ChIP-Sequencing

Preparation of the libraries for high-throughput sequencing was performed using the ChIP-Seq DNA sample Prep Kit (Illumina #IP-102-1001), following the manufacturer's instructions with some modifications. Libraries were prepared starting from 1.08ng of DNA from SIM2 (A6, B8 and C4) anti-FLAG ChIP, SIM2 (A6, B8 and C4) input, EB3 anti-FLAG ChIP and EB3 input samples. 8ng of starting material were used from ChIP and input DNA from OCT4, SOX2, NANOG, MED1 and MED12 ChIP experiments (Sim2 A6 and EB3 cells). Enrichment of the DNA fragments by PCR was performed using reagents and adapters from the TruSeq RNA Sample Preparation kit (Illumina #RS-122-2001) according to the following program: 98°C for 30s followed by 18 cycles of 98°C for 10s, 60°C for 30s and 72°C for 30s, followed by a final elongation step at 72°C for 5min. PCR clean up was done on Agencourt AMPure XP beads (Beckman Coulter #A63880). Libraries were validated on an Agilent Technologies 2100 Bioanalyzer (DNA1000 chip). Libraries were sequenced on Illumina HiSeq 2000, in single-end sequencing 1x36bp or 1x50bp (4 samples per lane).

mRNA-Sequencing

mRNA-Sequencing libraries were prepared from 500ng of total RNA using the TruSeq RNA Sample Preparation kit (Illumina #RS-122-2001) following Illumina's instructions. Libraries were sequenced on one lane of the Illumina HiSeq 2000 in paired-end sequencing 2x100bp.

Co-immunoprecipitation experiments

SIM2 A6 cells and EB3 cells were grown in tetracycline-free media for 26h and harvested using Trypsin-EDTA. Total protein extract was collected in lysis buffer (50mM Hepes pH 8, 200mM NaCl, 0.1mM EDTA pH 8, 0.5% NP-40, 10% glycerol and protease inhibitors) after 1h at 4°C and centrifugation for 30min at 4°C (13000rpm). 50 μ l of beads (Dynabeads protein G, Life Technologies #10003D) were prepared for the immunoprecipitation by coupling with 2 μ g of antibody for 30min at room temperature. Immunoprecipitation was performed overnight at 4°C using 500 μ g of protein extract and the beads coupled to the following antibodies: anti-Sox2 (Y-17) (Santa Cruz #sc-17320), anti-Oct3/4 (N-19) (Santa Cruz #sc-8628), anti-Klf4 (R&D systems #AF3158), anti-Nanog (N-term and C-term) (Bethyl #A310-110A). Beads were washed

four times for 5min at 4°C in 50mM TrisHCl pH 8, 250mM NaCl, 1% Triton X-100. Elution was performed in 200mM TrisHCl pH 8, 6% SDS, 15% glycerol, 3% β-mercaptoethanol for 10min at 95°C. 5μl of the immunoprecipitated extract were analyzed by western blot using an anti-Flag antibody coupled to HRP (Sigma #A8592, 1:1000 dilution). Each experiment was performed in duplicate.

ChIP-Sequencing analysis

For each sequenced library, reads generated from the sequencing were mapped against the mouse genome (mm9) using the BWA (Burrows-Wheeler Aligner) [26] alignment program with the default parameters (allowing 2 mismatches). Mapped reads were submitted to the HOMER (Hypergeometric Optimization of Motif EnRichment) software (<http://biowhat.ucsd.edu/homer/ngs/index.html>) [27] for the identification of SIM2 DNA binding sites. HOMER was used with the default parameters after removing of the duplicated reads. For each SIM2 clone, peak finding was done first by comparing the SIM2 tags to the input tags (background removal) and second by deleting all the non-specific sites identified in the EB3 control experiment. The genome ontology and motif discovery analyses were performed using HOMER.

Each identified peak was assigned to the closest gene(s) by calculating the distance separating the center of the peak from the TSS. All peaks located in intergenic or intronic regions were assigned to both the closest upstream and downstream genes. Peaks located in exonic or promoter regions were assigned to the unique gene to which they belong.

For OCT4, SOX2, NANOG, MED1 and MED12, peak finding was performed in HOMER by comparing the ChIP tags to the input tags in SIM2 A6 cells and EB3 cells independently. Each peak was then assigned to the closest gene by calculating the distance separating the center of the peak from the TSS.

mRNA-Sequencing analysis

mRNA-Seq reads were mapped against the mouse genome (mm9) using the default parameters of BWA. For each gene, a custom pipeline was used to calculate the exon coverage. This coverage was normalized in reads per kilobase per million (RPKM). Differential expression analysis between *Sim2* expressing cells (A6, B8, C4) and EB3 control cells (3 independent replicates) was performed using the default parameters of EdgeR [28]. A gene was considered differentially expressed if the false discovery rate (FDR) was below 5%.

Gene Ontology and Gene Set Enrichment Analysis (GSEA)

Gene ontology analyses were performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) [29, 30]. Gene Set Enrichment Analysis was performed using the GSEA software [31]. Genes were sorted according to their fold change between *Sim2* expressing cells and EB3 cells (mRNA-Seq data). The GSEA analysis consisted in testing if a particular gene set was randomly distributed in this ranked list or enriched at the beginning or the end of the distribution. A positive enrichment score (ES) reflects enrichment in the upregulated genes whereas a negative ES reflects enrichment in the downregulated genes. This enrichment was considered significant if the FDR corrected p-value was less than 0.05 (after 1000 or 10000 permutations).

ChIA-PET analysis

ChIA-PET data were taken from Zhang *et al.* [32]. Circular representation of the inter-chromosomal and intra-chromosomal interactions was done using the RCircos package (<https://bitbucket.org/henryhzhang/rcircos>).

Overlap between Sim2 binding sites and other features

Enrichment around Sim2 DNA binding sites was performed using the ChIP-Cor Analysis Module of the ChIP-Seq Web Server (<http://ccg.vital-it.ch/chipseq/documents.php>). The relative abundance of each tested feature is reported in a 40kb window around the SIM2 DNA binding sites by comparing the position of the SIM2 peaks with the position of the target features. SOX2, NANOG, OCT4, MED1 and MED12 peak positions were taken from our ChIP-seq data in *Sim2* expressing cells. Other transcription factor peak coordinates were taken from Chen *et al.* [33] (after lift over of the data to mm9). Control MED1 peak coordinates were taken from Kagey *et al.* [34]. Chromatin modification marks were taken from the mouse ENCODE data in the UCSC genome browser mm9 build (<http://genome.ucsc.edu/>): P300 and PolII data are from ES-Bruce4 cells (LICR), H3K4me1 and H3K27ac data are from ES-E14 cells (LICR), H3K4me3 data are from ES-E14 cells (SYDH) and DNaseI HS data are from ES-E14 cells [17].

Overlap between each feature and the SIM2 DNA binding sites was tested using the windowBed command of bedtools, by using a 100bp window interval. The significance of the association was tested using a Fisher's exact test comparing the number of features overlapping with the SIM2 binding sites and the number of features overlapping with a random set of 1229 intervals. The F-score was calculated according to [35].

Accession number

Sequencing data have been submitted to GEO under the accession number: GSE59379.

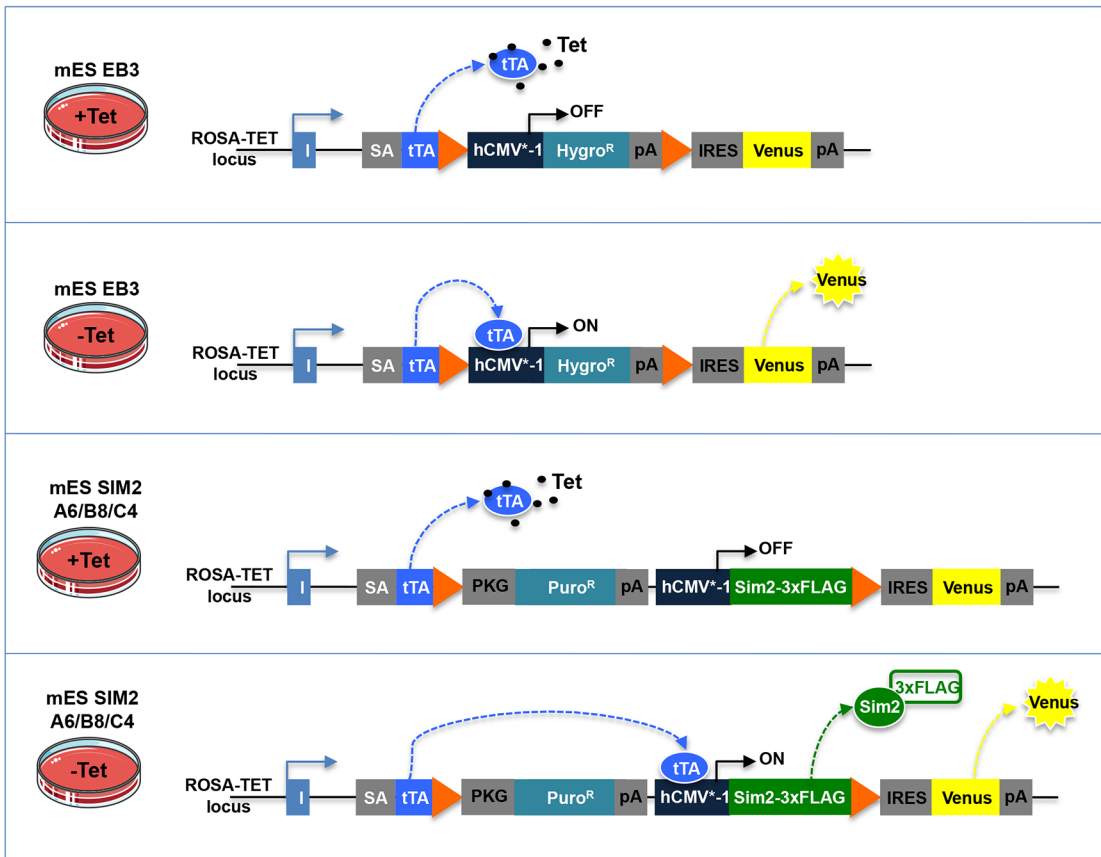
Results

Identification of SIM2 DNA binding sites

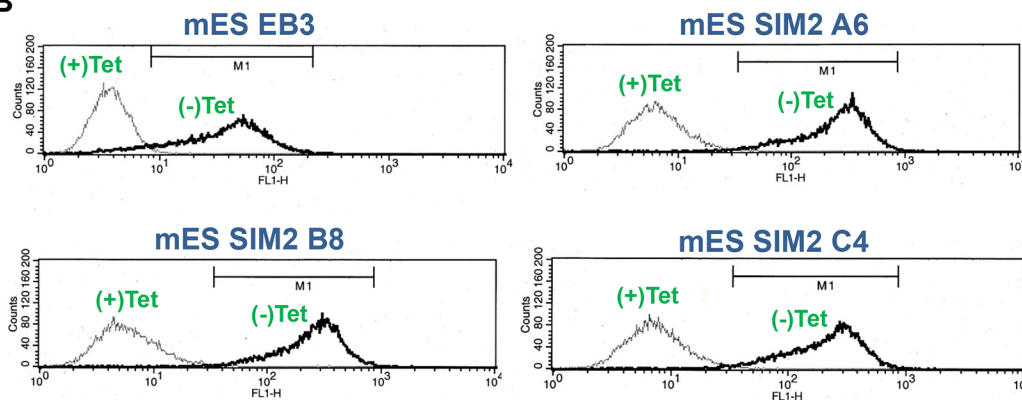
We used of a mES cell line that stably overexpresses a Flag-tagged version of the mouse *Sim2* gene [24]. This model is based on a ROSA-TET system allowing the inducible overexpression of the *Sim2*-FLAG transgene upon removal of tetracycline from the culture media. We analyzed three different mES clones harboring the *Sim2* construct (named A6, B8 and C4) as well as the EBRTcH3 (EB3) parental line as a negative control (Fig 1A). The *Venus* transgene, inserted downstream of the construct, was used in the four lines as an internal control to verify the inducible system. A FACS (Fluorescence Activated Cell Sorting) analysis revealed that 26 hours of induction were sufficient to promote the expression of *Venus* in the four lines (Fig 1B). We confirmed the presence of *Sim2* transcripts in the induced A6, B8 and C4 expressing clones as opposed to the EB3 parental line by Reverse-Transcription PCR (RT-PCR, Fig 1C). Finally, we showed that the *Arnt* partner, essential for the formation of active transcription factor complexes, was expressed in all the lines (Fig 1D).

We then performed ChIP-Sequencing in the A6, B8, C4 and EB3 lines after 26 hours of induction in tetracycline-free medium by using an anti-Flag antibody. ChIP and input DNAs were sequenced on the Illumina HiSeq 2000. 36 to 68 million reads were generated and mapped against the mouse genome (mm9) using the BWA aligner (Burrows-Wheeler Aligner) [26]. We then used HOMER [27] to analyze the reads and identify the SIM2 binding sites. We used the A6, B8 and C4 clones as biological replicates and identified 2387, 2137 and 631 peaks

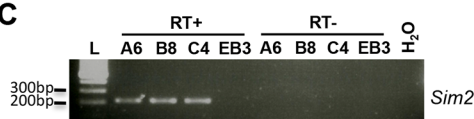
A



B



C



D

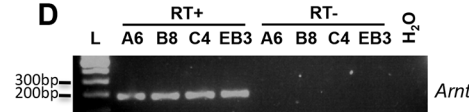


Fig 1. Cellular model. a. Schematic representation of the inducible ROSA-TET system. The mES SIM2 clones A6, B8 and C4 contain a Flag-tagged version of the mouse *Sim2* gene under the control of a modified human CMV promoter (hCMV*-1). In presence of tetracycline in the culture media (+Tet), the tetracycline-regulatable transactivator (tTA) is trapped and cannot bind the hCMV*-1 promoter. Upon removal of tetracycline (-Tet), the tTA binds the hCMV*-1 promoter inducing the expression of the *Sim2*-Flag-IRES-Venus construct. The mES EB3 parental line does not contain the *Sim2*-Flag transgene. The puromycin-resistant (Puro^R) and hygromycin-resistant (Hygro^R) cassettes are used for the clone selection process. SA: Splice Acceptor; pA: poly-

adenylation site; IRES: Internal Ribosome Entry Site; orange triangles represent loxP sites. Modified from [25] **b**. Fluorescence-activated cell sorting (FACS) analysis of *Sim2* expressing and non-expressing clones. Cells were grown in the presence (+Tet) or absence (-Tet) of Tetracycline during 26 hours. The y-axis represents the number of cells and the x-axis the fluorescence intensity. **c and d**. Agarose gel electrophoresis results of reverse-transcription PCR assay (RT). Total RNA from +Tet cells was reverse transcribed and amplified using primers specific to *Sim2* (c) or *Arnt* (d) in presence (RT+) or absence (RT-) of reverse transcriptase. L: loading marker; H₂O: PCR negative control.

doi:10.1371/journal.pone.0126475.g001

in each line, respectively (Fig 2A). After exclusion of the non-specific peaks detected in the EB3 parental line, we selected the binding sites that were identified in at least 2 replicate experiments. We described a total of 1229 SIM2 specific binding sites, including 346 sites common to the 3 *Sim2* expressing lines (S1 Table). The majority of these 1229 peaks were located in intergenic (57%) and intronic (37%) regions of the genome (Fig 2B). We found that 80% of the SIM2 peaks were located in a 100kb window around a known transcription start site (TSS) (Fig 2C). A total of 32 SIM2 binding sites were found in promoter regions defined by a window of -1kb/+300bp around the TSS (Fig 2B).

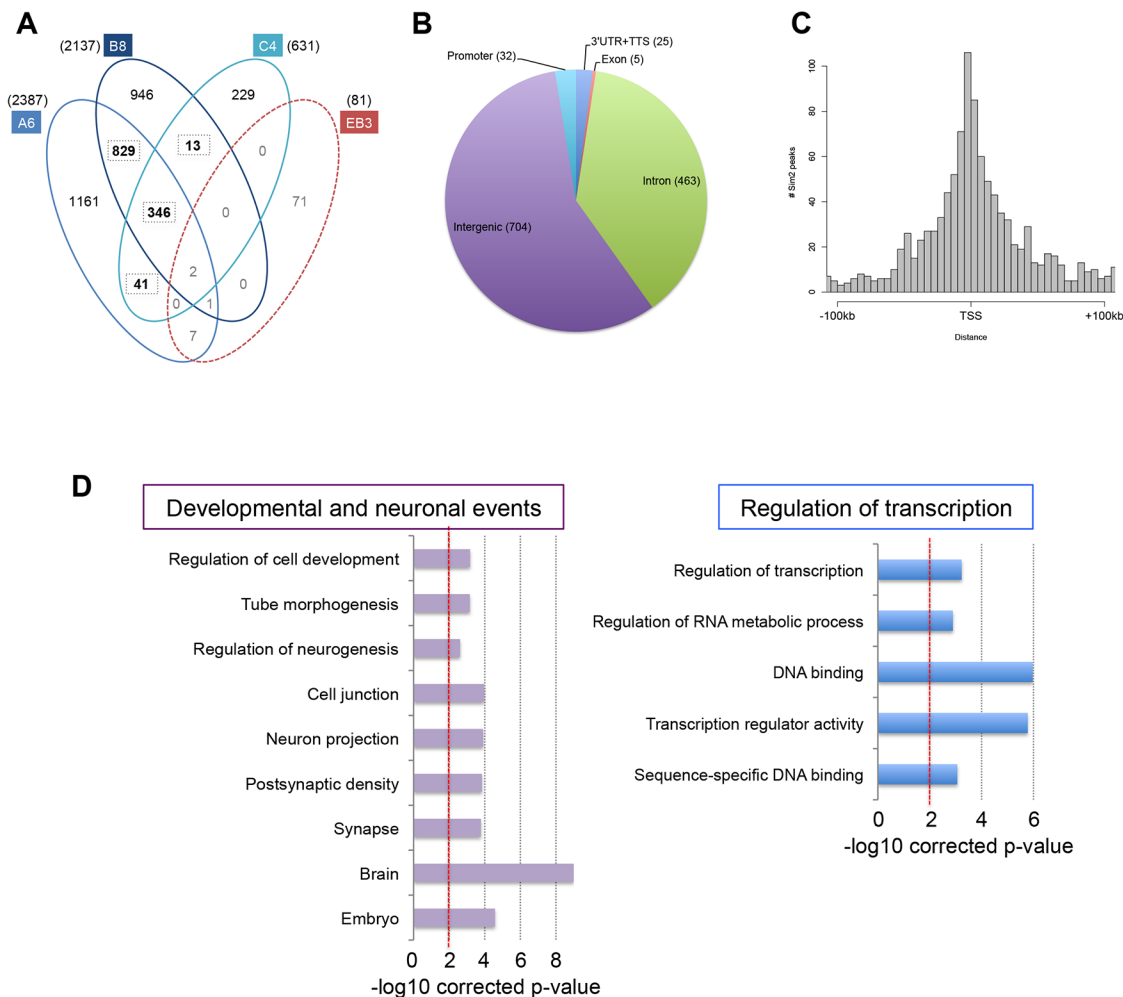


Fig 2. Identification and characterization of the SIM2 DNA binding sites by ChIP-seq. **a**. Venn diagram of the number of SIM2 binding sites identified by ChIP-seq in each SIM2 clones (A6, B8, C4) and EB3 line. The sum of the bold numbers is equal to the 1229 SIM2 DNA binding sites found in at least 2 SIM2 clones. **b**. The pie chart shows the genomic distribution of these 1229 sites. **c**. Distribution of the distances between the SIM2 DNA binding sites and the closest transcription start site (TSS). **d**. Selection of gene ontology terms significantly over-represented in the list of genes associated to a SIM2 DNA binding site.

doi:10.1371/journal.pone.0126475.g002

Characterization of the putative SIM2 target genes

In order to identify putative SIM2 target genes, we assigned each of the 1229 peaks to the closest gene(s) by calculating the distance separating the center of the peak from the TSS. Peaks located in intergenic or intronic regions were assigned to both the closest upstream and downstream TSSs whereas peaks located in exons or promoter regions were assigned to a unique gene. In total, 1992 different genes were associated to one or more SIM2 binding site(s). We used DAVID (Database for Annotation, Visualization and Integrated Discovery) [29, 30] to perform a gene ontology analysis and check if this list of genes was enriched for specific biological functions (Fig 2D and S2 Table). Interestingly, the analysis revealed a significant enrichment for genes involved in developmental processes and more specifically in neurogenesis, including regulation of cell development (Benjamini corrected p-value $p = 6.69e-04$), tube morphogenesis ($p = 6.89e-04$), regulation of neurogenesis ($p = 0.002$) or regulation of nervous system development ($p = 0.002$). The same analysis revealed the over-representation of genes expressed in brain and embryonic tissues ($p = 9.36e-10$ and $p = 2.66e-05$, respectively) as well as cellular components such as synapse ($p = 1.65e-04$) or neuron projection ($p = 1.24e-04$) (S2 Table). These results confirmed the role of SIM2 in developmental events, specifically in the nervous system. Additionally, we found that those SIM2-associated genes were also significantly involved in mechanisms of transcription regulation, as revealed by gene ontology terms such as regulation of transcription ($p = 5.78e-04$), DNA binding ($p = 1.06e-06$) or transcription factor activity ($p = 5.81e-06$). These results show that SIM2 can control the expression of other transcription factors in the genome, suggesting that it may be an important master regulator. Interestingly, the list of SIM2 targets was also enriched for genes involved in cancer pathways, as revealed by the KEGG pathway analysis ($p = 1.46e-04$) (S2 Table). This finding is consistent with the reported involvement of SIM2 in several cancers [36–40].

Validation by mRNA-sequencing and ChIA-PET data analyses

In order to further validate the SIM2 target genes, we have investigated the changes of mRNA levels induced by the overexpression of *Sim2*. We used mRNA-sequencing to study the transcriptome of the A6, B8, C4 and EB3 lines. Total RNA was collected concurrently with the ChIP-Seq experiment and sequenced on the Illumina instrument. The reads generated were mapped against the mouse genome using BWA and normalized in RPKM (Reads per Kilobase per Million) in order to compare the expression level of each gene between *Sim2* expressing and non-expressing cells. We first verified the expression level of *Sim2* in both conditions and confirmed its overexpression in A6, B8 and C4 (196.04, 212.42 and 168.40 RPKM, respectively) as opposed to the EB3 cells which showed very low levels of endogenous *Sim2* transcripts (0.13, 0.09 and 0.10 RPKM in each of the 3 replicates, respectively) (Fig 3A). We also confirmed that *Arnt*, the SIM2 co-factor, is expressed at similar levels in all cell lines, with an average RPKM level of 19.7 (Fig 3B). This shows that the overexpression of SIM2-FLAG does not influence the level of endogenous *Arnt*. The SIM2-FLAG activity is therefore limited by the endogenous levels of *Arnt*, restricting the formation of active complexes to physiological ranges.

We then used EdgeR [28] to perform a differential expression analysis between *Sim2* expressing and non-expressing cells. This analysis revealed that 300 RefSeq genes were significantly upregulated and 230 genes significantly downregulated when *Sim2* is overexpressed in mES cells ($FDR < 0.05$) (Fig 3C and S3 Table). However, the gene ontology analysis did not show enrichment for a specific type of biological functions (data not shown). A Gene Set Enrichment Analysis (GSEA) revealed that the SIM2 targets were significantly enriched among the most upregulated genes in *Sim2* overexpressing cells (Enrichment Score $ES = 0.37$, $FDR < 1e-04$, Fig 3D), with 18.7% of the differentially expressed genes associated to a SIM2

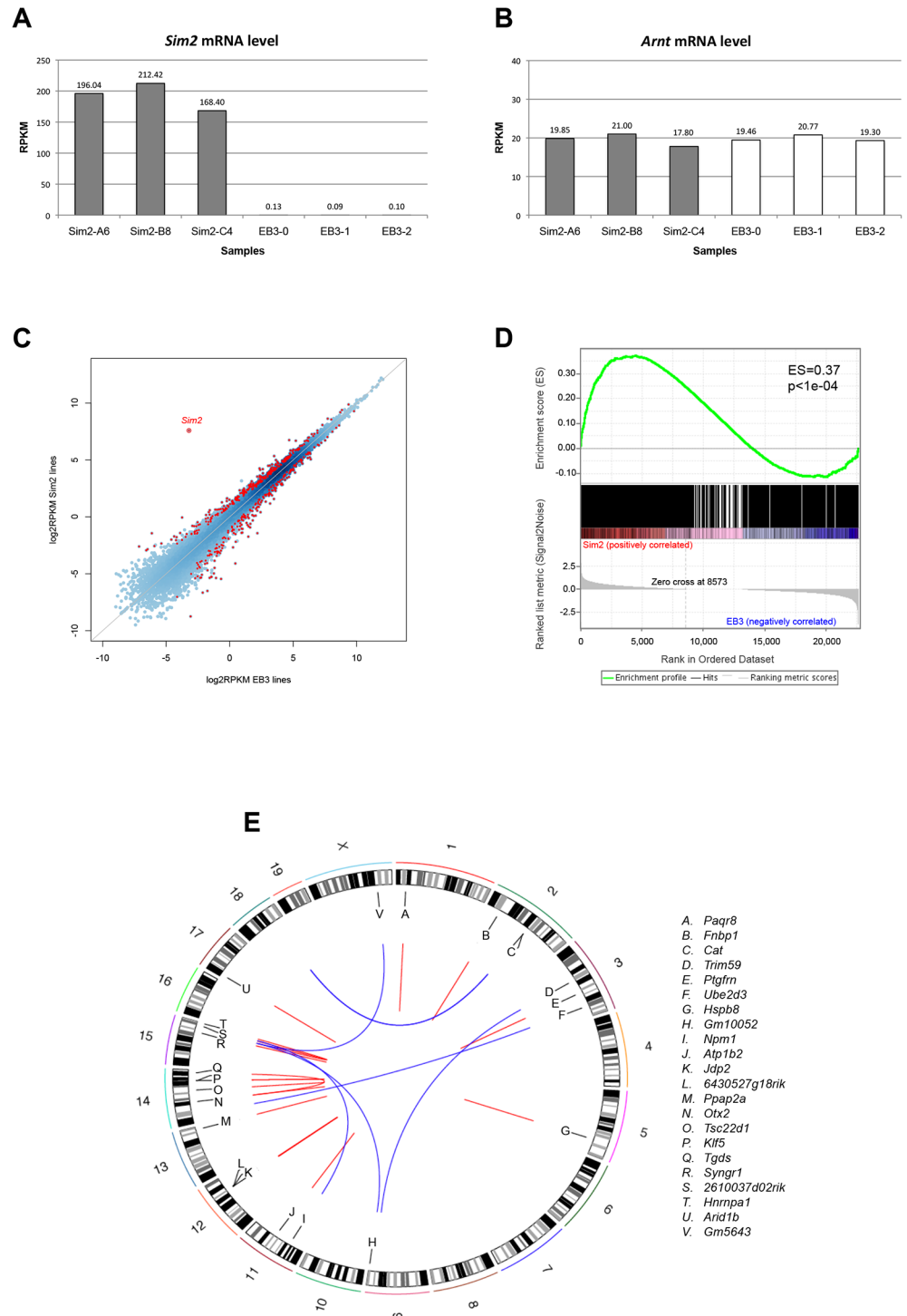


Fig 3. mRNA-sequencing and ChIA-PET analyses. *Sim2* (a) and *Arnt* (b) mRNA levels (RPKM) in A6, B8, C4 SIM2 clones and three EB3 replicates. c. Comparison of the gene expression level (mean log₂ RPKM between the 3 replicates) between *Sim2* expressing cells (y-axis) and EB3 control cells (x-axis). Each blue dot represents a gene; differentially expressed genes (EdgeR FDR<0.05) are shown in red. The diagonal line represents the expected distribution of genes equally expressed between *Sim2* expressing and non-expressing cells. d. Enrichment of *Sim2* targets among the genes upregulated in *Sim2* expressing clones as revealed by the GSEA analysis. Genes were sorted according to their expression fold change between *Sim2* expressing and non-expressing cells (x-axis, 0 showing the most upregulated gene). Black vertical bars show the position of the SIM2 targets in the ranked list. The enrichment score (ES in green) significantly deviates

from zero at the beginning of the distribution showing that the SIM2 targets are not randomly distributed in the ranked list but enriched among the most upregulated genes. **p**: FDR corrected p-value **e**. ChIA-PET interactions occurring between SIM2 DNA binding sites and promoters of genes differentially expressed in *Sim2* expressing cells compared to EB3 cells (FDR<0.05). Blue lines show inter-chromosomal interactions and red lines intra-chromosomal interactions.

doi:10.1371/journal.pone.0126475.g003

DNA binding site. Interestingly, 99 genes previously associated to a SIM2 peak were found significantly dysregulated (24 down- and 75 up-regulated) in *Sim2* expressing cells (FDR<0.05) and can be considered as direct targets of SIM2 ([S4 Table](#)). A gene ontology analysis revealed that this list of 99 genes was not enriched for any particular biological functions (data not shown).

Importantly, SIM2 may also bind genomic regions that are not located in a direct proximity of a target gene. Indeed, the formation of chromatin loops within the nucleus is known to promote the interaction between promoters and regulatory regions located distantly to regulate the gene transcription. The ChIA-PET method (Chromatin Interaction Analysis by Paired-End Tag sequencing) has been developed to identify such long-range interactions [[41](#), [42](#)]. We used datasets of ChIA-PET chromatin interactions associated with RNA polymerase II available in mouse ES cells [[32](#)] to investigate the existing interactions between the SIM2 binding loci and distant gene promoters. Among the 1229 genomic regions bound by SIM2, 206 were found to physically interact with one or several gene promoter(s) occupied by a RNAPII transcriptionally active complex in mES cells ([S5 Table](#)). We observed 102 inter-chromosomal interactions, suggesting that SIM2 could act in *trans* to regulate the expression of distant targets. In contrast, 265 interactions occur between loci located on the same chromosome. Most of these intra-chromosomal interactions (63%) connected SIM2 binding loci and gene promoters distant from less than 100kb. Overall, the RNA polymerase II ChIA-PET datasets identified 310 unique transcribed gene promoters that physically interact with at least one SIM2 binding site. Interestingly, 22 of those genes were significantly dysregulated by the overexpression of *Sim2* (EdgeR FDR<0.05) ([Table 1](#)). We considered those as putative SIM2 targets since their expression level is associated with the binding of SIM2 in their promoter region ([Fig 3E](#)).

SIM2 co-localizes and interacts with master transcription factors

We then investigated whether SIM2 preferentially binds to specific DNA motifs. Using the HOMER algorithm, we identified five motifs significantly enriched in the SIM2 DNA binding sites (p-value<1E-50, [Fig 4](#)). Similar motifs were found when we independently analyzed peaks located in promoters, gene bodies or intergenic regions (data not shown). Interestingly, four of those enriched motifs were highly similar to motifs previously described in mouse ES cells for the binding of master transcription factors involved in the control of pluripotency: OCT4, SOX2, NANOG and KLF4 ([Fig 4](#)). Three of them (OCT4, SOX2 and NANOG commonly called OSN) are known to constitute the core of all mechanisms regulating the transcription program of ES cells and participating in the maintenance of their pluripotent state [[43](#)].

The binding motifs similarities as well as the key role of these master transcription factors led us to further investigate the possible overlap between the regions occupied by SIM2 and the binding sites of the OSN factors in the ES cells. To do so, we first generated additional ChIP-sequencing data for OCT4, SOX2 and NANOG in the *Sim2* overexpressing cells (A6 clone) and the EB3 parental line. The numbers of binding sites identified for each factor are summarized in [S6 Table](#). We investigated the distribution of those binding sites in a 40kb window around the SIM2 peaks. Interestingly, for the three factors, we found an increased frequency of binding

Table 1. List of Sim2 target genes identified using the ChIA-PET data.

Gene name	Sim2 target		Associated Sim2 peak	Distance from peak (bp)
	logFC	FDR		
<i>2610037D02RIK</i>	-1.38	4.26E-02	Merged-chr15-96084454-1	29,585
<i>6430527G18RIK</i>	0.68	1.33E-02	Merged-chr12-88165302-1	60,462
<i>Arid1b</i>	0.65	7.12E-03	Merged-chr17-4948580-1	46,493
<i>Atp1b2</i>	0.90	2.51E-03	Merged-chr11-69395541-1	23,903
<i>Cat</i>	0.55	2.68E-02	Merged-chr18-65738596-1	Inter (chr2-chr18)
			Merged-chr18-65741928-1	Inter (chr2-chr18)
<i>Fnbp1</i>	0.70	1.70E-02	Merged-chr2-30916463-1	81,065
<i>Gm10052</i>	-0.64	7.38E-03	Merged-chr15-88698126-1	Inter (chr9-chr15)
<i>Gm5643</i>	-0.57	2.00E-02	Merged-chr15-88698126-1	Inter (chrX-chr15)
<i>Hnrmpa1</i>	-0.52	4.26E-02	Merged-chr15-88698126-1	14,372,818
<i>Hspb8</i>	0.76	2.59E-02	Merged-chr5-116858161-1	14,712
<i>Jdp2</i>	1.21	1.45E-02	Merged-chr12-86962262-1	21,897
			Merged-chr12-86961177-1	20,812
<i>Klf5</i>	0.81	2.23E-04	Merged-chr14-99739553-1	41,644
			Merged-chr14-99739095-1	41,186
<i>Npm1</i>	-0.48	4.07E-02	Merged-chr15-97424436-1	Inter (chr11-chr15)
<i>Otx2</i>	-1.90	5.44E-12	Merged-chr14-49274868-1	12,320
<i>Paqr8</i>	1.10	1.31E-02	Merged-chr1-20857728-1	22,969
<i>Ppap2a</i>	0.63	2.62E-02	Merged-chr13-113561150-1	29,951
<i>Ptgfrn</i>	0.62	1.67E-02	Merged-chr3-100967534-1	53,333
<i>Syng1</i>	0.86	8.95E-04	Merged-chr15-80049889-1	128,116
<i>Tgds</i>	-0.82	3.31E-02	Merged-chr14-49274868-1	69,257,119
<i>Trim59</i>	-0.51	4.26E-02	Merged-chr9-110849737-1	Inter (chr3-chr9)
<i>Tsc22d1</i>	0.77	9.81E-05	Merged-chr14-76915540-1	100,773
<i>Ube2d3</i>	-0.78	9.81E-05	Merged-chr14-20569154-1	Inter (chr3-chr14)

doi:10.1371/journal.pone.0126475.t001

sites at the localization of the SIM2 peaks, suggesting that SIM2 occupies preferentially the same genomic loci as these master transcription factors in the genome (Fig 5A). In the *Sim2*

Rank	SIM2 binding motifs	p-val	% of DNA binding sites	Best match in mES cells	Score
1		1e-256	33.52%	Pou5f1	0.93
2		1e-134	61.84%	---	---
3		1e-101	27.91%	SOX2	0.93
4		1e-70	14.89%	NANOG	0.92
5		1e-63	26.44%	KLF4	0.93

Fig 4. Motifs enriched in Sim2 DNA binding sites.

doi:10.1371/journal.pone.0126475.g004

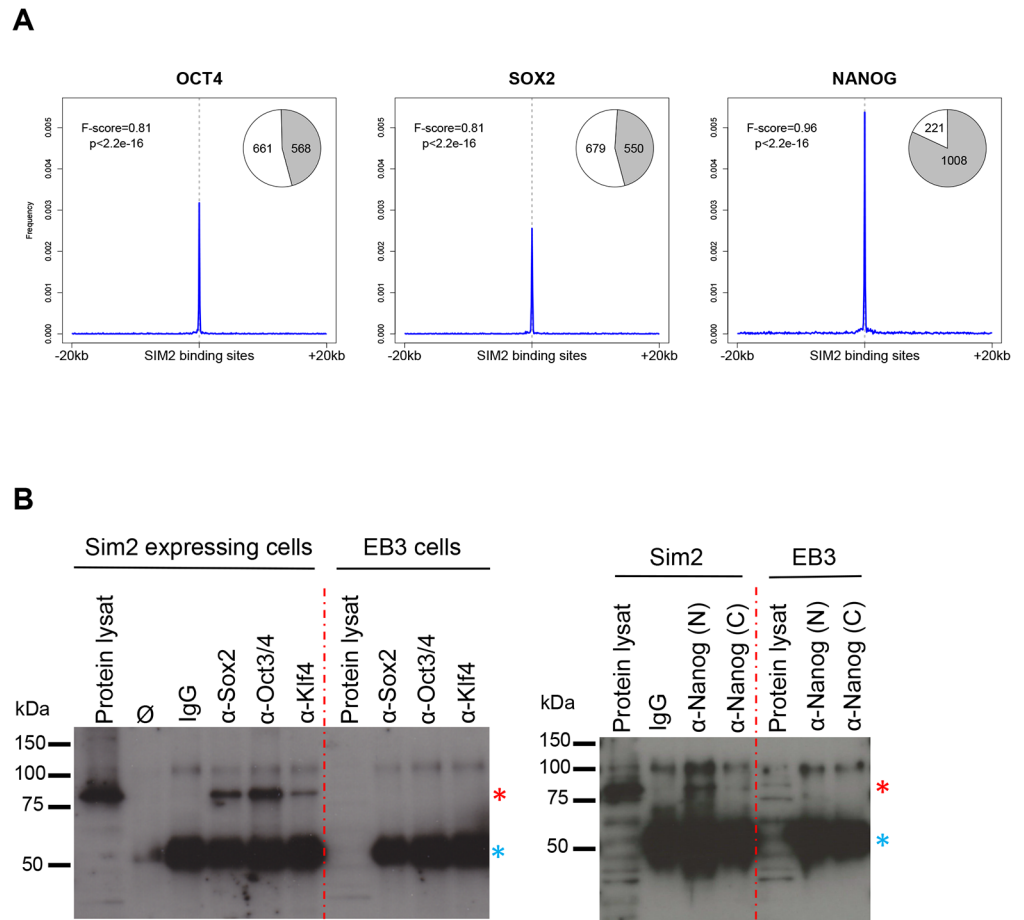


Fig 5. Overlapping SIM2 occupancy with master transcription factor binding sites. **a.** Frequency distribution of OCT4, SOX2 and NANOG DNA binding sites in a 40kb window centered to the newly identified SIM2 DNA binding sites. Plots show a significant enrichment for the OSN binding sites at the SIM2 peak localization in SIM2 A6 expressing cells. Pie charts show the proportion of SIM2 DNA binding sites overlapping with the OCT4, SOX2 or NANOG binding sites (in grey) (100bp window). p = Fisher's exact test p -value; F score: measure of the significance of the association (1 = perfect match). **b.** Protein co-immunoprecipitation experiments of SIM2-FLAG with endogenous OCT4, SOX2, KLF4 (left panel) and NANOG (right panel). Cellular protein extracts from *Sim2* expressing cells (A6) or EB3 cells were immunoprecipitated by using antibodies directed against each of the pluripotency factors (N-terminal and C-terminal part of NANOG) or IgG as a negative control for co-immunoprecipitation. Associated proteins were immunoblotted using an anti-FLAG antibody. Red star shows the SIM2-FLAG protein, blue star the signal given by the recognition of the IgG heavy chains. ∅: Beads only; kDa: kilodaltons; protein lysat: protein lysate.

doi:10.1371/journal.pone.0126475.g005

expressing cells, 82% of SIM2 peaks overlap with a DNA binding site for NANOG, 46.2% with a binding site for OCT4 and 44.75% with a binding site for SOX2 (100bp window) (Fig 5A). Comparison with a random set of peaks revealed that this overlap is significantly higher than expected by chance ($p < 2.2 \times 10^{-16}$). We validated those results by examining ChIP-Seq data previously published for OCT4, SOX2 and NANOG as well as other pluripotency factors including KLF4 and ESRRB [33]. This comparison revealed the same enrichment for the binding of these factors at the SIM2 peaks (S1 Fig). Altogether, these data suggest that SIM2 could co-occupy a number of loci with master transcription factors involved in the control of the pluripotent state. We performed protein co-immunoprecipitation experiments to test if SIM2 interacts with partners of the OSN protein complex. Importantly, detectable amounts of SIM2-FLAG

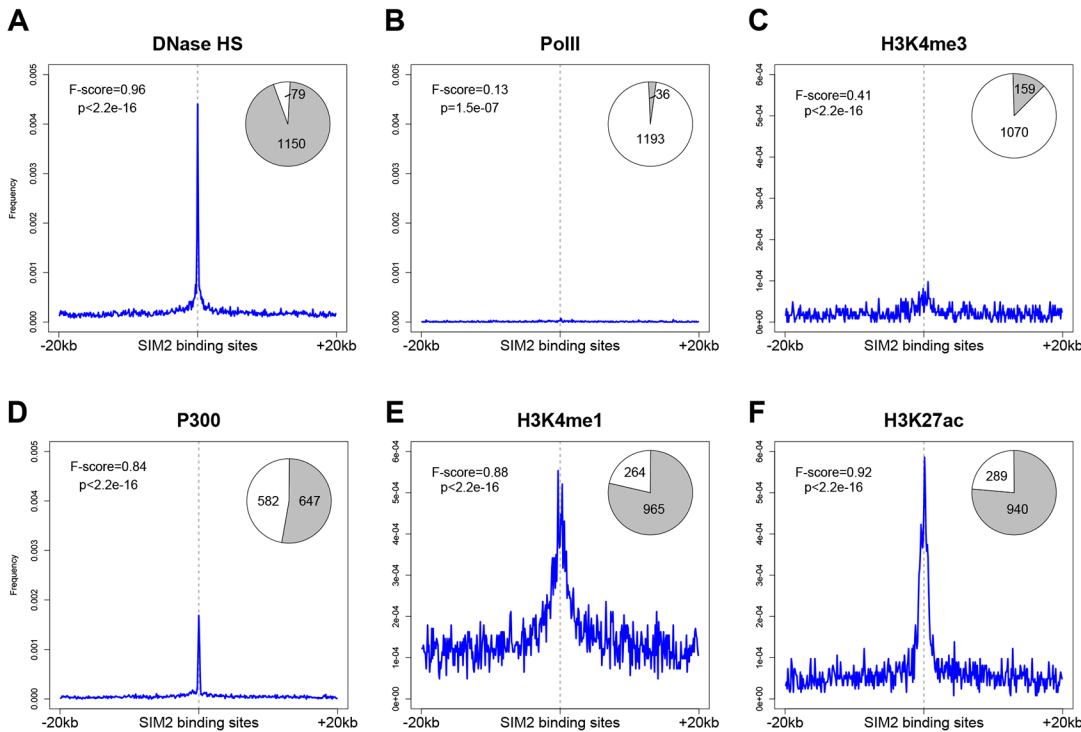


Fig 6. SIM2 DNA binding sites colocalize with known enhancer marks. Distribution of chromatin modification marks in a 40kb window centered to the SIM2 DNA binding sites: DNaseI hypersensitivity signal (a), RNA polymerase II (b), H3K4me3 (c), P300 (d), H3K4me1 (e) and H3K27ac (f). Pie charts show the proportion of SIM2 peaks overlapping each of these marks (in grey) (100bp window). p = Fisher's exact test p -value; F score: measure of the significance of the association (1 = perfect match). Data were taken from the mouse ENCODE project in the UCSC genome browser mm9 build (<http://genome.ucsc.edu/>).

doi:10.1371/journal.pone.0126475.g006

were co-immunoprecipitated with antibodies against endogenous SOX2, OCT3/4, KLF4 and NANOG in a total cellular protein extract from *Sim2* expressing mES cells (Fig 5B and S2 Fig). These results support our hypothesis and suggest that SIM2 interacts independently with these four transcription factors. SIM2 might therefore be a partner of the OSN complex carrying a key regulatory role in ES cells.

SIM2 marks enhancer and super-enhancer regions

A previous study has described the co-binding of the OSN master transcription factors in ES cells as predictive for enhancer activity [33]. In order to test if the binding of SIM2 could also predict such cis-regulatory activity, we analyzed the distribution of chromatin modification marks (available from the mouse ENCODE project) in the vicinity of the SIM2 binding sites. We first observed a significant increase of chromatin accessibility at the loci occupied by SIM2, as revealed by the enrichment for DNaseI hypersensitivity (HS) (Fig 6A). We then examined the distribution of P300, H3K4me1 (monomethylation of lysine 4 of histone 3) and H3K27ac (acetylation of lysine 27) to investigate the enhancer profile in the genomic regions bound by SIM2. We found that the SIM2 binding regions significantly overlap with these enhancer marks, suggesting that the presence of SIM2 may coincide with an enhancer activity, as previously described for the pluripotency factors (Fig 6D–6F). Consistently, SIM2 binding sites were found to significantly overlap with the typical enhancers described by Whyte *et al.* [44] (S1 Fig). In contrast, marks for promoter signals such as RNA polymerase II occupancy or

H3K4me3 (trimethylation of lysine 4) were poorly enriched, suggesting that SIM2 cannot extensively predict a promoter activity (Fig 6B and 6C).

A recent study reported the existence of a sub-category of enhancers known as super-enhancers [44]. We found that a significant fraction of Sim2 binding sites overlap with those (S1 Fig) and thus further investigated this correlation. These super-enhancer regions are characterized by the co-occupancy of OCT4, SOX2 and NANOG. They mainly differ from the typical enhancers by the length of the DNA regions they span and by the increased presence of the Mediator coactivator complex. Additionally, they possess a specific transcription factor signature enriched for KLF4 and ESRRB but excluding other ES cell factors such as CTCF or c-Myc. Interestingly, by examining the binding profile of all these factors (data taken from Chen *et al.* [33]) in the genomic regions occupied by SIM2, we found indeed that CTCF and c-Myc were not enriched as opposed to KLF4 and ESRRB (S1 Fig). We performed ChIP-sequencing to investigate the binding genomic regions of MED1 and MED12, the main constituents of the Mediator complex, in the *Sim2* expressing cells. We revealed a significant overlap between SIM2 binding sites and each of these factors (p -value $< 2.2E-16$, Fig 7). ChIP-seq data available for the MED1 protein [34] confirmed the significant enrichment at the SIM2 binding sites (S1 Fig). Altogether, these results suggest that SIM2 is implicated in conventional enhancers as well as in regulatory functions of super-enhancers.

Discussion

In this study, we have shown how the identification and characterization of the SIM2 DNA binding sites improve the understanding of its molecular function and potential role in the manifestations of DS.

SIM2 target genes confirmed the contribution of SIM2 to the DS cognitive impairment

We have identified 1229 binding loci for SIM2 and shown that a significant fraction of target genes located in the vicinity were involved in neuronal development processes. These results suggest that SIM2 may be a candidate gene for some DS phenotypes, in particular the cognitive impairment. These findings validate the hypotheses established so far with studies performed in *Drosophila* and in mouse models [6–8, 19–23]. In addition, our study enabled the discovery

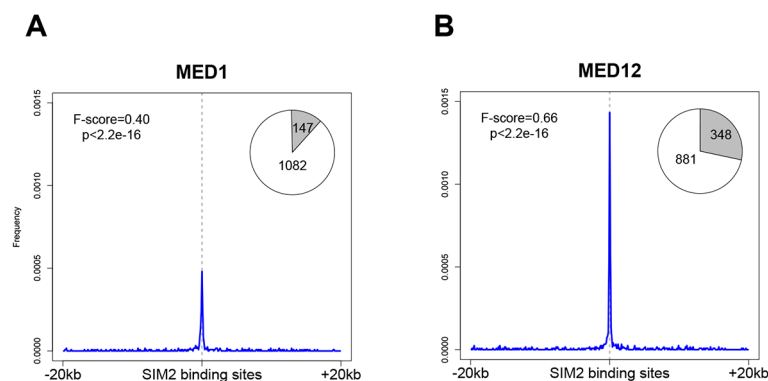


Fig 7. The Mediator complex colocalizes with the SIM2 DNA binding sites. Frequency distribution of MED1 (a) and MED12 (b) DNA binding sites in a 40kb window centered to the SIM2 peaks. Pie charts show the proportion of SIM2 DNA binding sites overlapping MED1 or MED12 DNA binding sites (in grey) (100bp window). p = Fisher's exact test p -value; F score: measure of the significance of the association (1 = perfect match).

doi:10.1371/journal.pone.0126475.g007

of target genes involved in mechanisms of transcription regulation revealing that SIM2 may have a role of master transcription factor and act upstream of important mechanisms controlling the gene expression in mES cells. By combining RNAPII ChIA-PET data, SIM2 binding sites and the transcriptome analysis in *Sim2* expressing cells, we have established a set of 22 genes that could be considered as direct targets of SIM2. Among them, several genes are known to be involved in molecular functions that could possibly be related to DS manifestations. For instance, the *OTX2* gene (Orthodenticle Homeobox 2) is known as an important transcription factor for the control of brain and craniofacial development [45]. Mutations of this gene have been linked to craniofacial malformations in both mouse and human [46, 47]. Patients harboring *OTX2* mutations present a microphthalmia syndrome associated to multiple features resembling those of DS such as developmental delay, hypotonia or short stature [48]. Similarly, the *ARID1B* gene (AT Rich Interactive Domain 1B), a member of the SWI/SNF chromatin remodeling complex, has been recently associated to cognitive impairment [49] and more specifically to the Coffin-Siris syndrome characterized by intellectual disability, severe speech delay and typical facial features [50, 51]. Finally, the *SYNGR1* gene (Synaptogyrin 1) also constitutes an interesting target gene given its role in synaptic plasticity, as revealed by the *Syng1* knockout mice [52].

Overexpression of SIM2 may influence the pluripotency signature of mES cells

The enrichment for enhancer marks at the SIM2 binding loci as well as the relatively small number of peaks located in promoter regions show that SIM2 is mainly recruited to distant regulatory elements for the regulation of its target genes. Here, we reported that SIM2 could bind genomic loci occupied by master transcription factors involved in the control of ES cell pluripotency. Co-immunoprecipitation experiments even suggest that SIM2 can physically interact with these factors and raise the possibility that they are part of the same protein complex. Interestingly, a study using the *Drosophila* model validated this hypothesis by showing functional interactions between SIM, SOX and POU transcription factors for the control of midline gene expression [53]. An interesting hypothesis has recently been developed regarding the role of pioneer transcription factors in the cells [54]. Indeed, it is well assumed that the recruitment of transcription factors is highly dependent on the chromatin state and that epigenetic modifications will likely influence their binding on the target sequences. Pioneer factors are known to act upstream of classical transcription factors to promote their binding on enhancer regions by modifying the chromatin landscape in order to improve its accessibility. The Forkhead box protein A1 (FOXA1) constitutes a typical example of pioneer factor acting during neuronal differentiation by changing enhancer chromatin signatures to promote the binding of subsequent factors [55, 56]. Interestingly, it has been proposed that OCT4, SOX2 and KLF4 could play a role of pioneer factors at distal enhancers during pluripotency reprogramming [57–59]. Thus, we can hypothesize that the enhancer sequences bound by SIM2 may initially be occupied by pioneer factors such as OCT4, SOX2, NANOG or KLF4 to modify the chromatin structure and facilitate the recruitment of SIM2 in response to specific differentiation signals.

This colocalization also raises the possibility that SIM2 interferes with the binding of the master transcription factors and thus could modify the pluripotent state of mES cells. This function has been previously reported for other factors including CDX2 (Caudal type homeobox 2) [60]. Indeed, it was shown that CDX2 has the ability to interfere with the binding of OCT4, SOX2 and NANOG, inducing the downregulation of their target genes. Since these pluripotency factors are known to control their own expression through auto-regulatory loops, it is likely that the binding interference of CDX2 contributes to the OSN downregulation and

thus to the initiation of differentiation processes. The same type of mechanisms could be proposed to explain the function of SIM2, especially since the *Sox2*, *Nanog*, *Klf4* and *Esrrb* genes belong to the list of targets associated to SIM2 binding sites in our results.

Interestingly, we have observed that SIM2 can also mark a particular subtype of enhancers called super-enhancers. Those are known to be associated with genes, mostly transcription factors, essential for the maintenance of the ES cell identity [44]. This observation raises the hypothesis that the binding of SIM2 in super-enhancer regions could modify the sensitive balance controlling the transcription program of ES cells and then promote the transition towards specific pathways, most probably neuronal differentiation. Consistently, several transcriptome studies have shown an increased expression of SIM2 in the early stages of the neuronal differentiation [61–63]. The mechanisms responsible for this transition are probably tightly controlled and we hypothesize that the dysregulation of SIM2 could disturb this fragile equilibrium. Further experiments will certainly help to understand the role of SIM2 in the differentiation processes.

Our data open interesting perspectives for the understanding of the mechanisms underlying the DS phenotypes and emphasize the benefit of using an ES cell model to study the function of HSA21 transcription factors.

Supporting Information

S1 Fig. Frequency distribution of published transcription factor binding sites, typical enhancers and super-enhancers in a 40kb window around the SIM2 peaks. Pie charts give the number of SIM2 peaks overlapping with the binding sites of each of the transcription factors, typical enhancers or super-enhancers (in grey) (100bp window). Typical enhancers and super enhancers data were taken from Whyte *et al.* [44]. MED1 ChIP-seq data were taken from Kagey *et al.* [34] and other data from Chen *et al.* [33].
(TIF)

S2 Fig. Protein co-immunoprecipitation experiments of SIM2-FLAG with endogenous OCT4, SOX2, KLF4 (left panel) and NANOG (right panel) (replication of the experiment shown Fig 4B). Cellular protein extracts from *Sim2* expressing cells (A6) or EB3 cells were immunoprecipitated by using antibodies directed against each of the pluripotency factors (N-terminal and C-terminal part of NANOG) or IgG as a negative control for co-immunoprecipitation. Associated proteins were immunoblotted using an anti-FLAG antibody. Red star shows the SIM2-FLAG protein, blue star the signal given by the recognition of the IgG heavy chains. Ø: Beads only; kDa: kilodaltons; protein lysat: protein lysat was loaded as an input control for the immunoblot.
(TIF)

S1 Table. List of Sim2 DNA binding sites.
(XLSX)

S2 Table. Gene Ontology analysis on the putative Sim2 targets.
(XLSX)

S3 Table. List of differentially expressed genes.
(XLSX)

S4 Table. List of Sim2 target genes dysregulated in *Sim2* expressing cells.
(XLSX)

S5 Table. ChIA-PET interactions.

(XLSX)

S6 Table. Number of Oct4, Sox2 and Nanog binding sites in the Sim2 expressing cells and the EB3 parental line.

(PDF)

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation and the European ERC to S.E.A, as well as the Lejeune foundation to C.B. We thank Francine Chopard for corrections on the manuscript.

Author Contributions

Conceived and designed the experiments: AL GC CB SEA. Performed the experiments: AL GC EF PR AV MG CB. Analyzed the data: AL AF FS MG. Contributed reagents/materials/analysis tools: GC AF PC. Wrote the paper: AL CB SEA.

References

1. Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet.* 2004; 5(10):725–38. PMID: [15510164](#)
2. Gardiner K. Transcriptional dysregulation in Down syndrome: predictions for altered protein complex stoichiometries and post-translational modifications, and consequences for learning/behavior genes ELK, CREB, and the estrogen and glucocorticoid receptors. *Behav Genet.* 2006; 36(3):439–53. PMID: [16502135](#)
3. Kewley RJ, Whitelaw ML, Chapman-Smith A. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol.* 2004; 36(2):189–204. PMID: [14643885](#)
4. Crews ST. Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes Dev.* 1998; 12(5):607–20. PMID: [9499397](#)
5. Panda S, Hogenesch JB, Kay SA. Circadian rhythms from flies to human. *Nature.* 2002; 417(6886):329–35. PMID: [12015613](#)
6. Thomas JB, Crews ST, Goodman CS. Molecular genetics of the single-minded locus: a gene involved in the development of the *Drosophila* nervous system. *Cell.* 1988; 52(1):133–41. PMID: [3345559](#)
7. Nambu JR, Franks RG, Hu S, Crews ST. The single-minded gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. *Cell.* 1990; 63(1):63–75. PMID: [2242162](#)
8. Nambu JR, Lewis JO, Wharton KA Jr, Crews ST. The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell.* 1991; 67(6):1157–67. PMID: [1760843](#)
9. Xiao H, Hrdlicka LA, Nambu JR. Alternate functions of the single-minded and rhomboid genes in development of the *Drosophila* ventral neuroectoderm. *Mech Dev.* 1996; 58(1–2):65–74. PMID: [8887324](#)
10. Estes P, Mosher J, Crews ST. *Drosophila* single-minded represses gene transcription by activating the expression of repressive factors. *Dev Biol.* 2001; 232(1):157–75. PMID: [11254355](#)
11. Sonnenfeld M, Ward M, Nystrom G, Mosher J, Stahl S, Crews S. The *Drosophila* tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Development.* 1997; 124(22):4571–82. PMID: [9409674](#)
12. Ema M, Suzuki M, Morita M, Hirose K, Sogawa K, Matsuda Y, et al. cDNA cloning of a murine homologue of *Drosophila* single-minded, its mRNA expression in mouse development, and chromosome localization. *Biochem Biophys Res Commun.* 1996; 218(2):588–94. PMID: [8561800](#)
13. Yamaki A, Noda S, Kudoh J, Shindoh N, Maeda H, Minoshima S, et al. The mammalian single-minded (SIM) gene: mouse cDNA structure and diencephalic expression indicate a candidate gene for Down syndrome. *Genomics.* 1996; 35(1):136–43. PMID: [8661114](#)
14. Moffett P, Dayo M, Reece M, McCormick MK, Pelletier J. Characterization of msim, a murine homologue of the *Drosophila* sim transcription factor. *Genomics.* 1996; 35(1):144–55. PMID: [8661115](#)

15. Dahmane N, Charron G, Lopes C, Yaspo ML, Maunoury C, Decorte L, et al. Down syndrome-critical region contains a gene homologous to *Drosophila sim* expressed during rat and human central nervous system development. *Proc Natl Acad Sci U S A*. 1995; 92(20):9191–5. PMID: [7568099](#)
16. Chrast R, Scott HS, Chen H, Kudoh J, Rossier C, Minoshima S, et al. Cloning of two human homologs of the *Drosophila* single-minded gene SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome chromosomal region. *Genome Res*. 1997; 7(6):615–24. PMID: [9199934](#)
17. Fan CM, Kuwana E, Bulfone A, Fletcher CF, Copeland NG, Jenkins NA, et al. "Expression patterns of two murine homologs of *Drosophila* single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome." *Mol Cell Neurosci*. 1996; 7(6):519. PMID: [8875433](#)
18. Rachidi M, Lopes C, Charron G, Delezoide AL, Paly E, Bloch B, et al. Spatial and temporal localization during embryonic and fetal human development of the transcription factor SIM2 in brain regions altered in Down syndrome. *Int J Dev Neurosci*. 2005; 23(5):475–84. PMID: [15946822](#)
19. Ema M, Ikegami S, Hosoya T, Mimura J, Ohtani H, Nakao K, et al. Mild impairment of learning and memory in mice overexpressing the mSim2 gene located on chromosome 16: an animal model of Down's syndrome. *Hum Mol Genet*. 1999; 8(8):1409–15. PMID: [10400987](#)
20. Chrast R, Scott HS, Madani R, Huber L, Wolfer DP, Prinz M, et al. Mice trisomic for a bacterial artificial chromosome with the single-minded 2 gene (*Sim2*) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of Down syndrome. *Hum Mol Genet*. 2000; 9(12):1853–64. PMID: [10915774](#)
21. Meng X, Peng B, Shi J, Zheng Y, Chen H, Zhang J, et al. Effects of overexpression of *Sim2* on spatial memory and expression of synapsin I in rat hippocampus. *Cell Biol Int*. 2006; 30(10):841–7. PMID: [16963290](#)
22. Goshu E, Jin H, Fasnacht R, Sepenski M, Michaud JL, Fan CM. *Sim2* mutants have developmental defects not overlapping with those of *Sim1* mutants. *Mol Cell Biol*. 2002; 22(12):4147–57. PMID: [12024028](#)
23. Shambloott MJ, Bugg EM, Lawler AM, Gearhart JD. Craniofacial abnormalities resulting from targeted disruption of the murine *Sim2* gene. *Dev Dyn*. 2002; 224(4):373–80. PMID: [12203729](#)
24. De Cegli R, Romito A, Iacobacci S, Mao L, Lauria M, Fedele AO, et al. A mouse embryonic stem cell bank for inducible overexpression of human chromosome 21 genes. *Genome Biol*. 2010; 11(6):R64. doi: [10.1186/gb-2010-11-6-r64](#) PMID: [20569505](#)
25. Masui S, Shimosato D, Toyooka Y, Yagi R, Takahashi K, Niwa H. An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic acids research*. 2005; 33(4):e43. PMID: [15741176](#)
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. doi: [10.1093/bioinformatics/btp324](#) PMID: [19451168](#)
27. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38(4):576–89. doi: [10.1016/j.molcel.2010.05.004](#) PMID: [20513432](#)
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. doi: [10.1093/bioinformatics/btp616](#) PMID: [19910308](#)
29. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44–57. doi: [10.1038/nprot.2008.211](#) PMID: [19131956](#)
30. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. 2009; 37(1):1–13. doi: [10.1093/nar/gkn923](#) PMID: [19033363](#)
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102(43):15545–50. PMID: [16199517](#)
32. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature*. 2013; 504(7479):306–10. doi: [10.1038/nature12716](#) PMID: [24213634](#)
33. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008; 133(6):1106–17. doi: [10.1016/j.cell.2008.04.043](#) PMID: [18555785](#)
34. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010; 467(7314):430–5. doi: [10.1038/nature09380](#) PMID: [20720539](#)

35. Santoni FA, Hartley O, Luban J. Deciphering the code for retroviral integration target site selection. *PLoS Comput Biol.* 2010; 6(11):e1001008. doi: [10.1371/journal.pcbi.1001008](https://doi.org/10.1371/journal.pcbi.1001008) PMID: [21124862](https://pubmed.ncbi.nlm.nih.gov/21124862/)
36. Kwak H-I, Gustafson T, Metz RP, Laffin B, Schedin P, Porter WW. Inhibition of breast cancer growth and invasion by single-minded 2s. *Carcinogenesis.* 2007; 28(2):259–66. PMID: [16840439](https://pubmed.ncbi.nlm.nih.gov/16840439/)
37. Aleman MJ, DeYoung MP, Tress M, Keating P, Perry GW, Narayanan R. Inhibition of Single Minded 2 gene expression mediates tumor-selective apoptosis and differentiation in human colon cancer cells. *Proc Natl Acad Sci U S A.* 2005; 102(36):12765–70. PMID: [16129820](https://pubmed.ncbi.nlm.nih.gov/16129820/)
38. Deyoung MP, Scheurle D, Damania H, Zylberberg C, Narayanan R. Down's syndrome-associated single minded gene as a novel tumor marker. *Anticancer Res.* 2002; 22(6A):3149–57. PMID: [12530058](https://pubmed.ncbi.nlm.nih.gov/12530058/)
39. He Q, Li G, Su Y, Shen J, Liu Q, Ma X, et al. Single minded 2-s (SIM2-s) gene is expressed in human GBM cells and involved in GBM invasion. *Cancer Biol Ther.* 2010; 9(6):430–6. PMID: [20448453](https://pubmed.ncbi.nlm.nih.gov/20448453/)
40. Lu B, Asara JM, Sanda MG, Arredouani MS. The role of the transcription factor SIM2 in prostate cancer. *PLoS One.* 2011; 6(12):e28837. doi: [10.1371/journal.pone.0028837](https://doi.org/10.1371/journal.pone.0028837) PMID: [22174909](https://pubmed.ncbi.nlm.nih.gov/22174909/)
41. Fullwood MJ, Han Y, Wei CL, Ruan X, Ruan Y. Chromatin interaction analysis using paired-end tag sequencing. *Curr Protoc Mol Biol.* 2010;Chapter 21:Unit 21 15 1–25.
42. Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, Vega V, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* 2010; 11(2):R22. doi: [10.1186/gb-2010-11-2-r22](https://doi.org/10.1186/gb-2010-11-2-r22) PMID: [20181287](https://pubmed.ncbi.nlm.nih.gov/20181287/)
43. Young RA. Control of the embryonic stem cell state. *Cell.* 2011; 144(6):940–54. doi: [10.1016/j.cell.2011.01.032](https://doi.org/10.1016/j.cell.2011.01.032) PMID: [21414485](https://pubmed.ncbi.nlm.nih.gov/21414485/)
44. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 2013; 153(2):307–19. doi: [10.1016/j.cell.2013.03.035](https://doi.org/10.1016/j.cell.2013.03.035) PMID: [23582322](https://pubmed.ncbi.nlm.nih.gov/23582322/)
45. Beby F, Lamonerie T. The homeobox gene Otx2 in development and disease. *Exp Eye Res.* 2013; 111:9–16. doi: [10.1016/j.exer.2013.03.007](https://doi.org/10.1016/j.exer.2013.03.007) PMID: [23523800](https://pubmed.ncbi.nlm.nih.gov/23523800/)
46. Matsuo I, Kuratani S, Kimura C, Takeda N, Aizawa S. Mouse Otx2 functions in the formation and patterning of rostral head. *Genes Dev.* 1995; 9(21):2646–58. PMID: [7590242](https://pubmed.ncbi.nlm.nih.gov/7590242/)
47. Chassaing N, Sorrentino S, Davis EE, Martin-Coignard D, Iacovelli A, Paznekas W, et al. OTX2 mutations contribute to the otocephaly-dysgnathia complex. *J Med Genet.* 2012; 49(6):373–9. doi: [10.1136/jmedgenet-2012-100892](https://doi.org/10.1136/jmedgenet-2012-100892) PMID: [22577225](https://pubmed.ncbi.nlm.nih.gov/22577225/)
48. Schilter KF, Schneider A, Bardakjian T, Soucy JF, Tyler RC, Reis LM, et al. OTX2 microphthalmia syndrome: four novel mutations and delineation of a phenotype. *Clin Genet.* 2010; 79(2):158–68.
49. Hoyer J, Ekici AB, Endeles S, Popp B, Zweier C, Wiesener A, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am J Hum Genet.* 2012; 90(3):565–72. doi: [10.1016/j.ajhg.2012.02.007](https://doi.org/10.1016/j.ajhg.2012.02.007) PMID: [22405089](https://pubmed.ncbi.nlm.nih.gov/22405089/)
50. Santen GW, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, et al. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet.* 2012; 44(4):379–80. doi: [10.1038/ng.2217](https://doi.org/10.1038/ng.2217) PMID: [22426309](https://pubmed.ncbi.nlm.nih.gov/22426309/)
51. Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, Hibi-Ko Y, et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet.* 2012; 44(4):376–8. doi: [10.1038/ng.2219](https://doi.org/10.1038/ng.2219) PMID: [22426308](https://pubmed.ncbi.nlm.nih.gov/22426308/)
52. Janz R, Sudhof TC, Hammer RE, Unni V, Siegelbaum SA, Bolshakov VY. Essential roles in synaptic plasticity for synaptogyrin I and synaptophysin I. *Neuron.* 1999; 24(3):687–700. PMID: [10595519](https://pubmed.ncbi.nlm.nih.gov/10595519/)
53. Ma Y, Certel K, Gao Y, Niemitz E, Mosher J, Mukherjee A, et al. Functional interactions between Drosophila bHLH/PAS, Sox, and POU transcription factors regulate CNS midline expression of the slit gene. *J Neurosci.* 2000; 20(12):4596–605. PMID: [10844029](https://pubmed.ncbi.nlm.nih.gov/10844029/)
54. Magnani L, Eeckhoutte J, Lupien M. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet.* 2011; 27(11):465–74. doi: [10.1016/j.tig.2011.07.002](https://doi.org/10.1016/j.tig.2011.07.002) PMID: [21885149](https://pubmed.ncbi.nlm.nih.gov/21885149/)
55. Serandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* 2011; 21(4):555–65. doi: [10.1101/gr.111534.110](https://doi.org/10.1101/gr.111534.110) PMID: [21233399](https://pubmed.ncbi.nlm.nih.gov/21233399/)
56. Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell.* 2008; 132(6):958–70. doi: [10.1016/j.cell.2008.01.018](https://doi.org/10.1016/j.cell.2008.01.018) PMID: [18358809](https://pubmed.ncbi.nlm.nih.gov/18358809/)
57. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell.* 2012; 151(5):994–1004. doi: [10.1016/j.cell.2012.09.045](https://doi.org/10.1016/j.cell.2012.09.045) PMID: [23159369](https://pubmed.ncbi.nlm.nih.gov/23159369/)

58. Bergsland M, Ramskold D, Zaouter C, Klum S, Sandberg R, Muhr J. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* 2011; 25(23):2453–64. doi: [10.1101/gad.176008.111](https://doi.org/10.1101/gad.176008.111) PMID: [22085726](https://pubmed.ncbi.nlm.nih.gov/22085726/)
59. Liber D, Domaschenz R, Holmqvist PH, Mazzarella L, Georgiou A, Leleu M, et al. Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. *Cell Stem Cell.* 2010; 7(1):114–26. doi: [10.1016/j.stem.2010.05.020](https://doi.org/10.1016/j.stem.2010.05.020) PMID: [20621055](https://pubmed.ncbi.nlm.nih.gov/20621055/)
60. Nishiyama A, Xin L, Sharov AA, Thomas M, Mowrer G, Meyers E, et al. Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. *Cell Stem Cell.* 2009; 5(4):420–33. doi: [10.1016/j.stem.2009.07.012](https://doi.org/10.1016/j.stem.2009.07.012) PMID: [19796622](https://pubmed.ncbi.nlm.nih.gov/19796622/)
61. Wu JQ, Habegger L, Noisa P, Szekely A, Qiu C, Hutchison S, et al. Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proc Natl Acad Sci U S A.* 2010; 107(11):5254–9. doi: [10.1073/pnas.0914114107](https://doi.org/10.1073/pnas.0914114107) PMID: [20194744](https://pubmed.ncbi.nlm.nih.gov/20194744/)
62. Lin M, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, Zheng D, et al. RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One.* 2011; 6(9):e23356. doi: [10.1371/journal.pone.0023356](https://doi.org/10.1371/journal.pone.0023356) PMID: [21915259](https://pubmed.ncbi.nlm.nih.gov/21915259/)
63. Hubbard KS, Gut IM, Lyman ME, McNutt PM. Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. *F1000Research.* 2013; 2:35. doi: [10.12688/f1000research.2-35.v1](https://doi.org/10.12688/f1000research.2-35.v1) PMID: [24358889](https://pubmed.ncbi.nlm.nih.gov/24358889/)