

Research article

Open Access

***In silico* identification of a core regulatory network of OCT4 in human embryonic stem cells using an integrated approach**

Lukas Chavez*¹, Abha S Bais², Martin Vingron², Hans Lehrach¹, James Adjaye¹ and Ralf Herwig¹

Address: ¹Department of Vertebrate Genomics, Max-Planck-Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany and ²Department of Computational Molecular Biology, Max-Planck-Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany

Email: Lukas Chavez* - chavez@molgen.mpg.de; Abha S Bais - bais@pitt.edu; Martin Vingron - vingron@molgen.mpg.de; Hans Lehrach - lehrach@molgen.mpg.de; James Adjaye - adjaye@molgen.mpg.de; Ralf Herwig - herwig@molgen.mpg.de

* Corresponding author

Published: 15 July 2009

Received: 13 January 2009

BMC Genomics 2009, 10:314 doi:10.1186/1471-2164-10-314

Accepted: 15 July 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/314>

© 2009 Chavez et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The transcription factor OCT4 is highly expressed in pluripotent embryonic stem cells which are derived from the inner cell mass of mammalian blastocysts. Pluripotency and self renewal are controlled by a transcription regulatory network governed by the transcription factors OCT4, SOX2 and NANOG. Recent studies on reprogramming somatic cells to induced pluripotent stem cells highlight OCT4 as a key regulator of pluripotency.

Results: We have carried out an integrated analysis of high-throughput data (ChIP-on-chip and RNAi experiments along with promoter sequence analysis of putative target genes) and identified a core OCT4 regulatory network in human embryonic stem cells consisting of 33 target genes. Enrichment analysis with these target genes revealed that this integrative analysis increases the functional information content by factors of 1.3 – 4.7 compared to the individual studies. In order to identify potential regulatory co-factors of OCT4, we performed a *de novo* motif analysis. In addition to known validated OCT4 motifs we obtained binding sites similar to motifs recognized by further regulators of pluripotency and development; e.g. the heterodimer of the transcription factors C-MYC and MAX, a prerequisite for C-MYC transcriptional activity that leads to cell growth and proliferation.

Conclusion: Our analysis shows how heterogeneous functional information can be integrated in order to reconstruct gene regulatory networks. As a test case we identified a core OCT4-regulated network that is important for the analysis of stem cell characteristics and cellular differentiation. Functional information is largely enriched using different experimental results. The *de novo* motif discovery identified well-known regulators closely connected to the OCT4 network as well as potential new regulators of pluripotency and differentiation. These results provide the basis for further targeted functional studies.

Background

Several studies on reprogramming human somatic cells to induced pluripotent stem cells (iPS) have demonstrated that the transduction of only a few transcription factors (TFs) is sufficient for resetting differentiated cells into a molecular state similar to embryonic stem cells (ESCs). While Takahashi et al. [1] and Wernig et al. [2] obtained iPS cells by transduction of the TFs OCT4, SOX2, KLF4, and C-MYC, Yu et al. [3] achieved similar results with a transcription factor set composed of OCT4, SOX2, NANOG, and LIN28. Only the TFs OCT4 and SOX2 are common in both approaches and Huangfu et al. [4] demonstrated that iPS cells can be derived at higher efficiencies by the transduction of these two factors in combination with the supplementation with the histone deacetylase inhibitor -valproic acid.

The TF OCT4 is known as a key regulator for maintaining pluripotency in the mammalian embryo [5-7]. The HMG-box containing TF SOX2 interacts with OCT4 and the SOX2/OCT4 heterodimer complex is able to promote selective gene activation or repression during mammalian embryogenesis [8-11].

Functional data on OCT4 regulatory action is available from heterogeneous sources: to reveal DNA-Protein binding events of OCT4, SOX2 and of the pluripotency associated TF NANOG, chromatin immuno-precipitation followed by microarray experiments (ChIP-on-chip) has been performed using hESCs [12]. Additionally, sequence motifs have been identified, for example the octamer motif *ATTGTCAT* interacting with POU domain factors like the homeodomain containing TF OCT4 and a motif recognized by the SOX2/OCT4 heterodimer complex [13-15]. Mapping of these known transcription factor binding motifs to the promoter sequences of putative OCT4 target genes provides additional evidence for direct binding events.

Although ChIP-on-chip experiments and sequence-based methods have the ability to detect such putative protein-DNA binding sites, these techniques do not allow inference of directional transcriptional dependencies between DNA binding and the effect on regulation of gene expression. In order to test the regulatory influence of OCT4 to the transcription rate of its target genes, Babaie et al. [16] performed RNA interference-mediated suppression of OCT4 function in the H1 hESC line and analyzed the resulting global gene expression changes by microarray experiments. Transcriptional changes induced by OCT4 knockdown are expected to include genes linked with pluripotency, and genes activated upon differentiation along the trophoblast lineage [16].

ChIP-on-chip experiments, promoter sequence analysis and RNA interference provide complementary pieces of information on transcriptional dependencies. In this study, we performed an integrated analysis of these methods in the context of OCT4 dependent regulation of pluripotency and differentiation along the trophoblast lineage in hESCs in order to construct a core network composed of the genes that were detected by all individual experimental approaches. Using this conservative selection, we observed a 1.3-4.7 fold increase of functional information content compared to single experiment analysis. In order to extend the analysis of OCT4 regulation, we performed a comprehensive *in silico* promoter sequence analysis with the OCT4 target genes and identified binding sites related to potential co-factors of OCT4.

Results

Analysis of individual experimental methods

We performed a re-analysis of the OCT4, SOX2 and NANOG ChIP-on-chip data from hESCs (NIH Code: WA09 cells) [12] including the mapping of the 60 mer oligonucleotide probes to an updated NCBI build (v36.1). In total, 230,068 oligonucleotides matched to their original position (+/-100 bp) whereas 141,270 probes were mapped more than 500 bp away from their original position. Processing of the uniquely-mapped probes includes background correction, normalization, fold-enrichment and peak identification and resulted in 308 potential OCT4 target genes (see Figure 1a for a histogram of OCT4 ChIP-on-chip ratios and Materials and Methods for a detailed description of the analysis). Figure 1b shows a histogram of the distances between binding sites and transcription start sites (TSSs) for the 308 direct OCT4 target genes. Most OCT4 binding sites (72.07%) are located less than 3 kb upstream of the TSS. 37.98% of all binding sites are located less than 1 kb upstream of the TSS and a second accumulation of binding sites is observed in the region between -1 kb to -2 kb with nearly a quarter (23.05%) of all binding sites.

Because protein-DNA binding events do not give information on the direction of the regulatory influence of the TF with respect to the transcription rate of its target genes, we complemented the results of the OCT4 ChIP-on-chip experiment with the results of the RNAi mediated OCT4 silencing in hESCs (H1 clone) performed by Babaie et al. [16]. Identifier mapping of the different chip platforms (Agilent oligochips and cDNA microarrays) resulted in 10,065 genes that were represented as cDNA clones on the microarray (see Materials and Methods) and that had promoter regions covered by the Agilent tiling arrays. From the originally published 623 OCT4 target genes [12], 472 were also represented on the cDNA microarray. From the 1,104 genes that show significantly altered expression 72

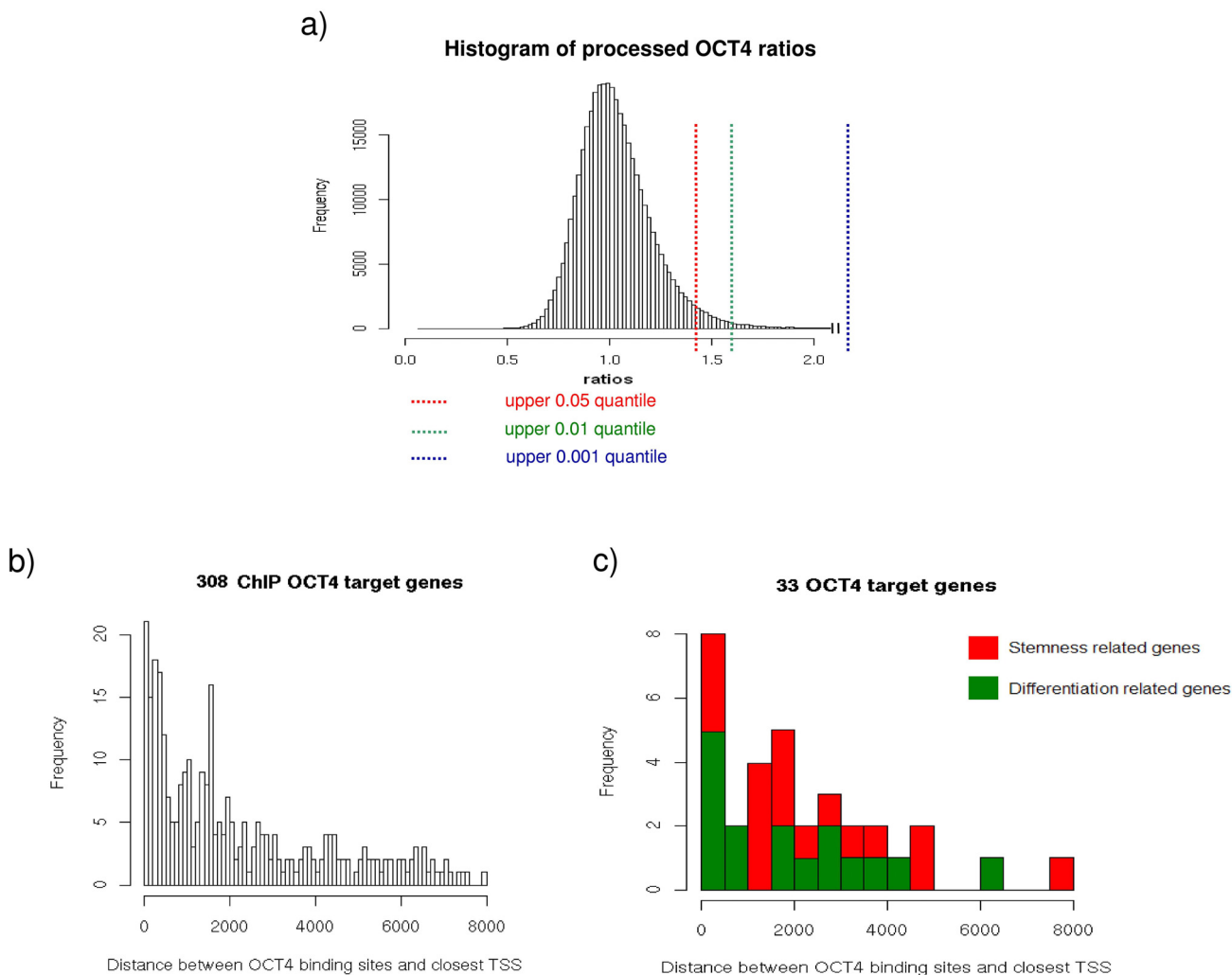


Figure 1

Histogram of processed OCT4 ratios. a) Histogram of quality controlled and normalized OCT4 ChIP-on-chip ratios for the individual probes. The blue line represents the threshold for the upper 0.001 quantile of the distribution (green line: upper 0.01 quantile and red line upper 0.05 quantile, respectively). b) Histogram of distances between OCT4 binding sites and closest TSS for all 308 validated OCT4 target genes, (c) Histogram of distances between OCT4 binding sites and closest TSS for the 33 isolated functional OCT4 target genes. The red fraction corresponds to stemness related genes, the green fraction to differentiation related genes.

hours after the OCT4 knock down, 40 genes (<4%) were also identified as direct OCT4 target genes.

In order to obtain an even more stringent set of OCT4 target genes, we searched the promoter sequences of the targets for the occurrence of the known OCT4-related octamer and SOX-OCT joint motifs within a distance of 8 kb upstream of the respective TSSs (see Material and Methods). Even though we neglect information on binding events caused by OCT4-DNA interactions mediated by unknown cofactors and heterodimer complexes, our results reflect confirmed functional circuitries dependent

on direct OCT4 and SOX-OCT binding. The combination of the three approaches resulted in a set of 33 genes (Figure 2). Nevertheless, it has to be mentioned that a motif could be mapped to the genomic environment of a ChIP-on-chip derived significant peak (distance of less than 1 kb) for only a third of these genes, whereas for the remaining genes the genomic position of the mapped motif is further away from the centre of the peaks.

The complete results of the individual studies (together with the results of the re-analyzed SOX2 and NANOG

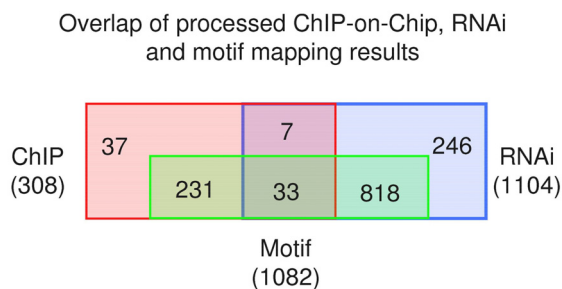


Figure 2
Overlap of the individual studies. Overlap of the re-analyzed OCT4 ChIP-on-chip experiment, the OCT4 RNAi experiment and the motif mapping results with the octamer and sox-oct joint motifs.

ChIP-on-chip experiments) are summarized in Additional file 1.

Positional distribution of OCT4 transcription factor binding sites (TFBSs)

Figure 1c shows a histogram of the distances between binding sites and TSSs for the 33 functional OCT4 target genes. Similar to the histogram of all 308 validated direct OCT4 target genes (see Figure 1b), the majority of OCT4 binding sites (69.68%) are located less than 3 kb upstream of the TSS, 30.30% of all binding sites are located already less than 1 kb upstream of the TSS and a second accumulation of binding sites can be observed in the region between -1 kb to -2 kb with nearly a quarter (24.24%) of all binding sites. Interestingly, slight differences in the distribution of binding site distances to TSSs can be observed when the set of OCT4 target genes is split into functionally distinct subsets (see Figures 1c): 43.75% of the genes that are functionally connected to the process of differentiation (defined by negative regulation by OCT4) have the OCT4 binding site within the 1 kb upstream region of their TSS, whereas only 17.65% of the stemness related genes (defined by positive regulation by OCT4) have the OCT4 binding site within this region. On the other hand, 35.23% of the stemness related genes have an OCT4 binding site within the -1 kb to -2 kb region whereas only 12.5% of the differentiation related genes have the OCT4 binding site within this region. Therefore, it seems that differentiation related genes tend to have an OCT4 binding site closer to their TSS.

OCT4 target genes

Among the 33 genes, several well-known targets of OCT4 can be found as well as genes whose regulatory interaction with OCT4 is less well-described. In general, OCT4 binds to and regulates diverse classes of genes encoding for example transcription factors (TGIF2, EOMES, FOXD3, GSC, TSC22D1, GATA6, OCT4, SOX2, NANOG, PAX6,

CDX2, TCF4), transcriptional regulators (SSBP2), regulators of kinase, transferase, and catalytic activity (GAP43, TDGF1), members of the Wnt receptor signalling pathway (SFRP2, FRAT2, DKK1), and growth factors (FGF2, LEFTY2, TDGF1). A functional classification of the 33 OCT4 target genes is given in Additional file 2.

As an example, Figure 3 illustrates ChIP-on-chip data results in the promoter regions of two target genes. Figure 3a shows two identified peaks located approximately 155 bp and 2027 bp upstream of the transcription start site of CDX2 (Caudal-type homeobox transcription factor 2). A binding event was identified for NANOG at the same genomic positions but not for SOX2. Additionally, the octamer motif was found approximately 233 bp upstream of the TSS. OCT4 negatively regulates the transcription of CDX2, as it is significantly up-regulated upon OCT4 knockdown [16]. This observation is in line with the function of CDX2 which encodes a protein that is important in a broad range of cellular functions such as trophoblast differentiation in human and mouse [17,18] to maintenance of the intestinal epithelial lining of both the small and large intestine [19]. Furthermore, it has been shown that Oct4 directly regulates the expression of Cdx2 in mouse embryonic stem cells [20,21].

As a second example, Figure 3b shows two identified peaks located approximately 1304 bp and 4212 bp upstream of the TSS of FOXD3 (Forkhead box protein D3). Binding events were identified for NANOG at the same genomic position but not for SOX2. The octamer motif was found approximately 4679 bp upstream of the TSS. Moreover, OCT4 has a positive regulatory influence on the transcription of FOXD3 as it is significantly down-regulated 72 hours after RNAi mediated OCT4 depletion [16]. This observation is consistent with the function of Foxd3 in mouse embryonic stem cells, as it is required for maintenance of progenitor cells in the inner cell mass and in the trophoblast [22,23]. Additionally, it has been shown that Foxd3 has an important role in repressing differentiation, promoting self-renewal, and maintaining survival of mouse ESCs [24].

Further evidence for the regulatory influence of OCT4 on the 33 target genes has been agglomerated from published experimental studies. Additional file 3 contains a glossary for the 33 core OCT4 target genes that summarizes further independent published experimental validations on the regulatory influence of OCT4 to its presented target genes.

Integration of data enriches functional content of OCT4 target gene set

Enrichment analysis [25] revealed that the functional information content of the gene set is accelerated (factors of 1.3 – 4.7) by integrating the results of the individual

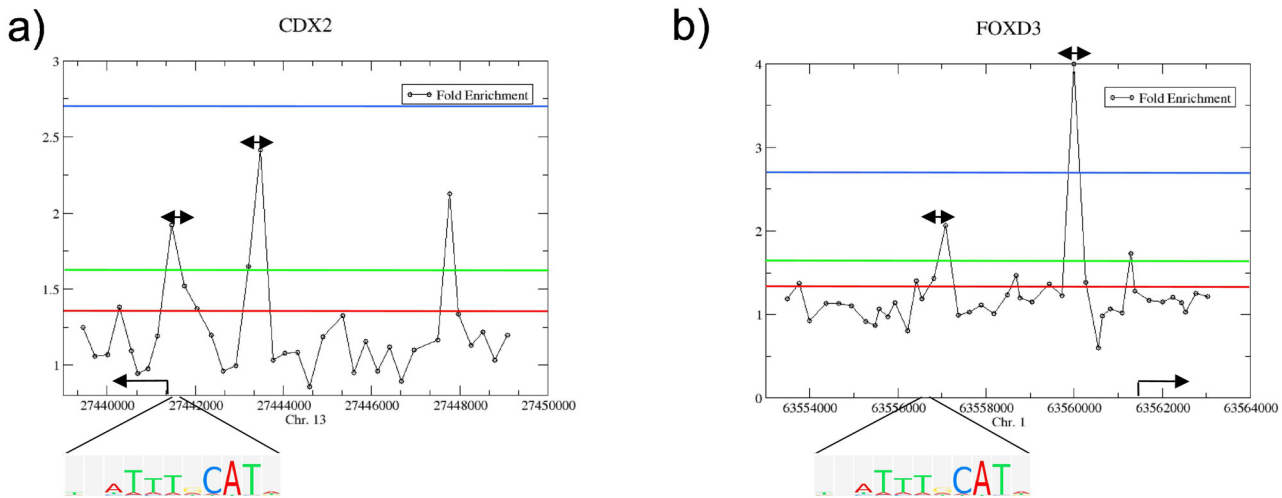


Figure 3
Example peaks taken from the isolated set of OCT4 target genes. Example peaks within the promoter regions of a) CDX2 and b) FOXD3, two of the isolated OCT4 target genes. The coloured lines refer to the thresholds corresponding to the quantiles given in Figure 1a. The motifs represent the octamer motif mapped to the respective promoter regions. The two-sided arrows illustrate the sub sequences taken for the *de novo* motif discovery.

studies. Figure 4 shows the increase in the percentage of genes connected to gene ontology terms, for example "GO:0003700: Transcription factor activity", with respect to the original set of ChIP-on-chip targets (origChIP, 22.72%), the re-analyzed set of ChIP-on-chip target genes (ChIP, 26.62%), the additional filtering according to the RNAi experiment (+RNAi, 35%) and additional motif mapping (+Motif, 36.36%). The same trend can be observed with other GO terms such as "GO:0030154: Cell differentiation", "GO:0045165: Cell fate commitment", "GO:0009790: Embryonic development", "GO:0008283: Cell proliferation", and others (see Figure 4). Table 1 shows the top enriched gene ontology terms (see Additional file 2 for the complete results of the enrichment analysis).

OCT4 core regulatory network

The resulting OCT4 core regulatory network, also incorporating the information on direct target genes from the re-analyzed SOX2 (red lines) and NANOG (blue lines) ChIP-on-chip experiments, is shown in Figure 5. The network distinguishes genes that are suppressed (left side) from those that are activated (right side) by OCT4. Among the 33 genes a high fraction is annotated with transcription factor activity (GO:0003700, indicated as rhombuses). Furthermore, a classification in hESCs specific genes (red boxes) and genes that are associated with the process of differentiation (green boxes) was performed by accessing several further public sources [16-18,26]. White boxed genes could not be annotated using these sources, but the information about up or down regulation after the OCT4 knock-down indicates, whether the respective gene

is functional connected to the process of differentiation or to the maintenance of pluripotency.

An additional level of gene regulation has been added to this core OCT4 target network by further literature and database mining (Additional file 4). This additional figure shows the core network extended by known up- and downstream target genes of the respective TFs as given by TRANSFAC [19] and by another published work [27].

Further interactions of the OCT4 target genes were revealed using the ConsensusPathDB [28], a database that integrates the content of 12 different interaction databases with heterogeneous foci. As an example, Additional file 5 shows known interaction partners of CDX2. Among these interactions, interestingly, a physical interaction is observed between CDX2 and PAX6, that is another important differentiation associated TF included in the presented set of OCT4 target genes. The core OCT4 network is represented in SBML format (Additional file 6) which can be used for further studies, e.g. mathematical modelling.

De novo motif discovery

Transcription factor co-localizations targeted by multiple transcription factors are sites that integrate the external signalling pathways to the transcriptional regulatory circuitry governed by OCT4, SOX2, and NANOG and these sites may serve as focal points for the assembly of further regulatory nucleoprotein complexes [29]. In order to test for further regulatory co-factors of OCT4, we performed a

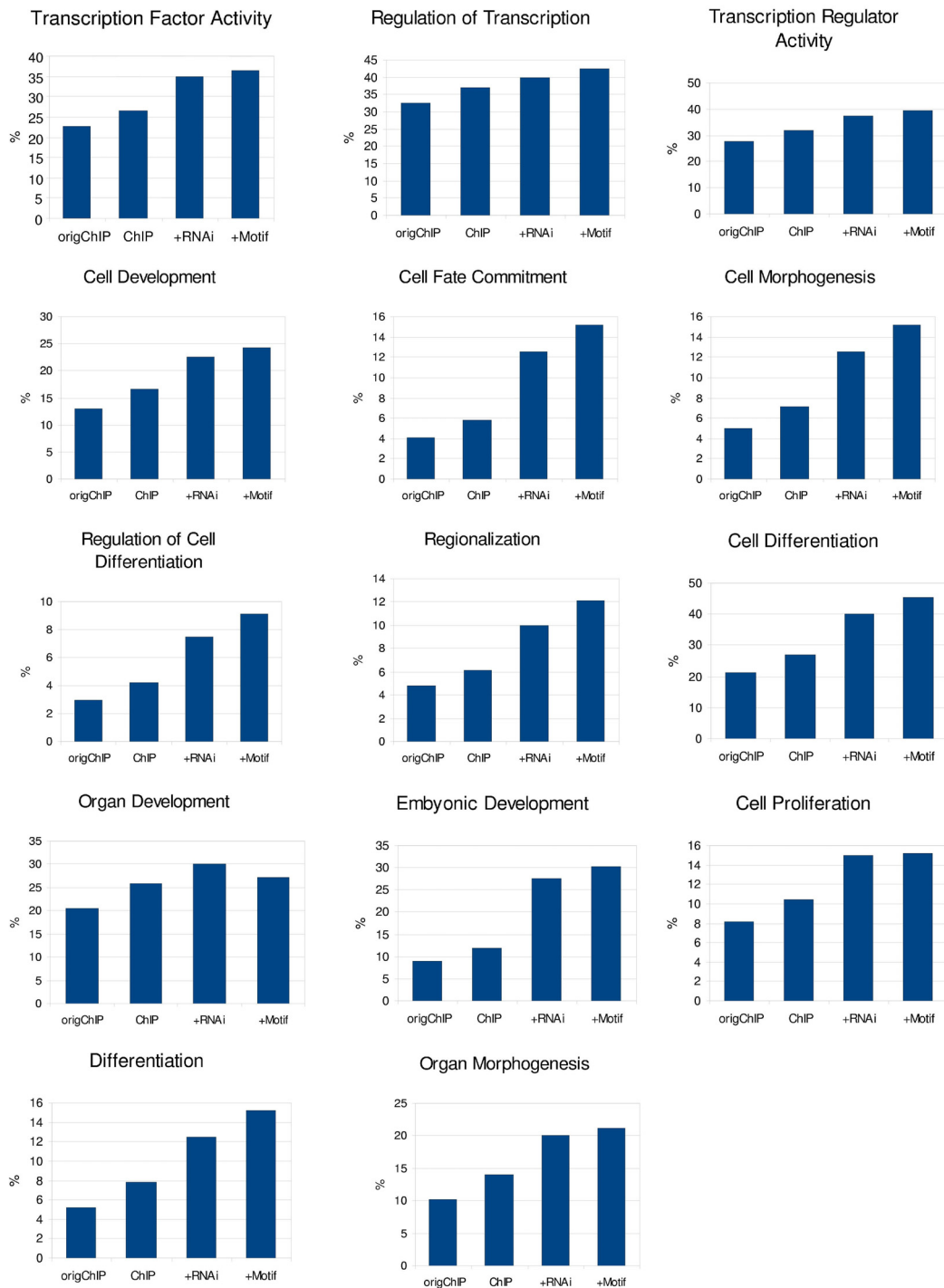


Figure 4
Enrichment of functional information through integrated analysis. Increase of functional information content of selected gene lists. As an example, the gene ontology term "GO:0003700 Transcription factor activity" is described in detail (from left to right): OCT4 target genes using the original ChIP-on-chip data (origChIP), using the re-analyzed OCT4 ChIP-on-chip data (ChIP), after integration of the results of the OCT4 RNAi experiment (+RNAi) and after integration of the motif mapping (+Motif). The corresponding absolute numbers of selected target genes are 122, 82, 14 and 12 (not indicated within the illustration).

Table 1: Enrichment analysis

Term	Count	%	PValue
GO:0007275~multicellular organismal development	18	54.55%	0.000000
GO:0003700~transcription factor activity	12	36.36%	0.000000
GO:0009653~anatomical structure morphogenesis	16	48.48%	0.000000
GO:0032502~developmental process	20	60.61%	0.000000
GO:0009790~embryonic development	10	30.30%	0.000000
GO:0048856~anatomical structure development	17	51.52%	0.000000
GO:0030528~transcription regulator activity	13	39.39%	0.000002
GO:0030154~cell differentiation	15	45.45%	0.000002
GO:0048869~cellular developmental process	15	45.45%	0.000002
GO:0007389~pattern specification process	6	18.18%	0.000010
GO:0032501~multicellular organismal process	19	57.58%	0.000019
GO:0045165~cell fate commitment	5	15.15%	0.000022
GO:0043565~sequence-specific DNA binding	8	24.24%	0.000029
GO:0003677~DNA binding	14	42.42%	0.000051
GO:0031323~regulation of cellular metabolic process	16	48.48%	0.000063
GO:0019222~regulation of metabolic process	16	48.48%	0.000095
GO:0009887~organ morphogenesis	7	21.21%	0.000105
GO:0031325~positive regulation of cellular metabolic process	7	21.21%	0.000107
GO:0009893~positive regulation of metabolic process	7	21.21%	0.000154
GO:0001824~blastocyst development	3	9.09%	0.000356
GO:0045449~regulation of transcription	14	42.42%	0.000413
GO:0019219~regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	14	42.42%	0.000519
GO:0050789~regulation of biological process	19	57.58%	0.000533
GO:0065007~biological regulation	20	60.61%	0.000539
GO:0006350~transcription	14	42.42%	0.000597
GO:0009798~axis specification	3	9.09%	0.000640
GO:0050794~regulation of cellular process	18	54.55%	0.000754
GO:0010468~regulation of gene expression	14	42.42%	0.000756
GO:0006355~regulation of transcription, DNA-dependent	13	39.39%	0.000890

Top enriched gene ontology categories for the 33 OCT4 target genes as received by DAVID [25] (see Additional file 2 for the complete results).

de novo motif discovery analysis based on specific promoter regions of the 308 direct OCT4 target genes derived from the re-analysis of the ChIP-on-chip data (sub-sequences of length 200 bp around the identified peaks; such selected regions are highlighted as arrows in Figure 3). The selected sub-sequences were repeat-masked [30] and used as input for several *de novo* motif discovery algorithms (see Material and Methods). We identified 12 unique sequence motifs of higher quality (Figure 6 and Additional file 7). These motifs were compared against two existing databases of known motifs (TRANSFAC [19] and JASPAR [31]) using the STAMP tool [32] (for the complete results see Additional files 8 and 9). Motifs similar to the octamer and sox-oct joint motifs were discovered. Additionally, sequence motifs were identified that are potentially recognized by factors involved in maintaining pluripotency and development. For the 33 core OCT4 target genes, Figure 6 lists the individual genes that contain the discovered motifs within their promoter regions.

Interestingly, we identified a motif similar to the binding site of MYC-MAX (see Figure 6), a heterodimer of the transcription factors C-MYC and MAX, a pre-requisite for C-MYC transcriptional activity that leads to cell growth and

proliferation [33,34]. C-MYC has been utilized in a set of four transcription factors for deriving iPS cells from somatic cells [1,2]. The C-MYC related motif was identified in the promoter region of PIP5K1C which encodes PIPKI gamma of the phosphatidylinositol pathway [35]. C-MYC is even further connected to the presented OCT4 regulatory network as it is a direct target of TCF4 (Transcription factor 7-like 2) whose transcription is enhanced by OCT4 function and which is a CTNBN1 binding protein that regulates Wnt signalling, cell cycle, and cell proliferation [19]. Moreover, C-MYC is a direct target of E2F/DP, a complex essential for cell proliferation [36]. Additionally, E2F/DP directly binds to ORC1L (Origin recognition complex subunit 1-like), another member of the network which encodes for a chromatin binding protein that plays a role in DNA replication initiation and mitotic cell cycle, induces caspase activation, cell surface exposure of phosphatidylserine, and DNA fragmentation during apoptosis [19]. A motif similar to the binding site of E2F-1 was obtained by the *de novo* motif discovery. Based on an analogous *de novo* motif discovery performed on the sub-sequences of validated peak regions taken from the promoter regions of the 33 core OCT4 target genes only, a motif similar to the binding site of E2F was obtained

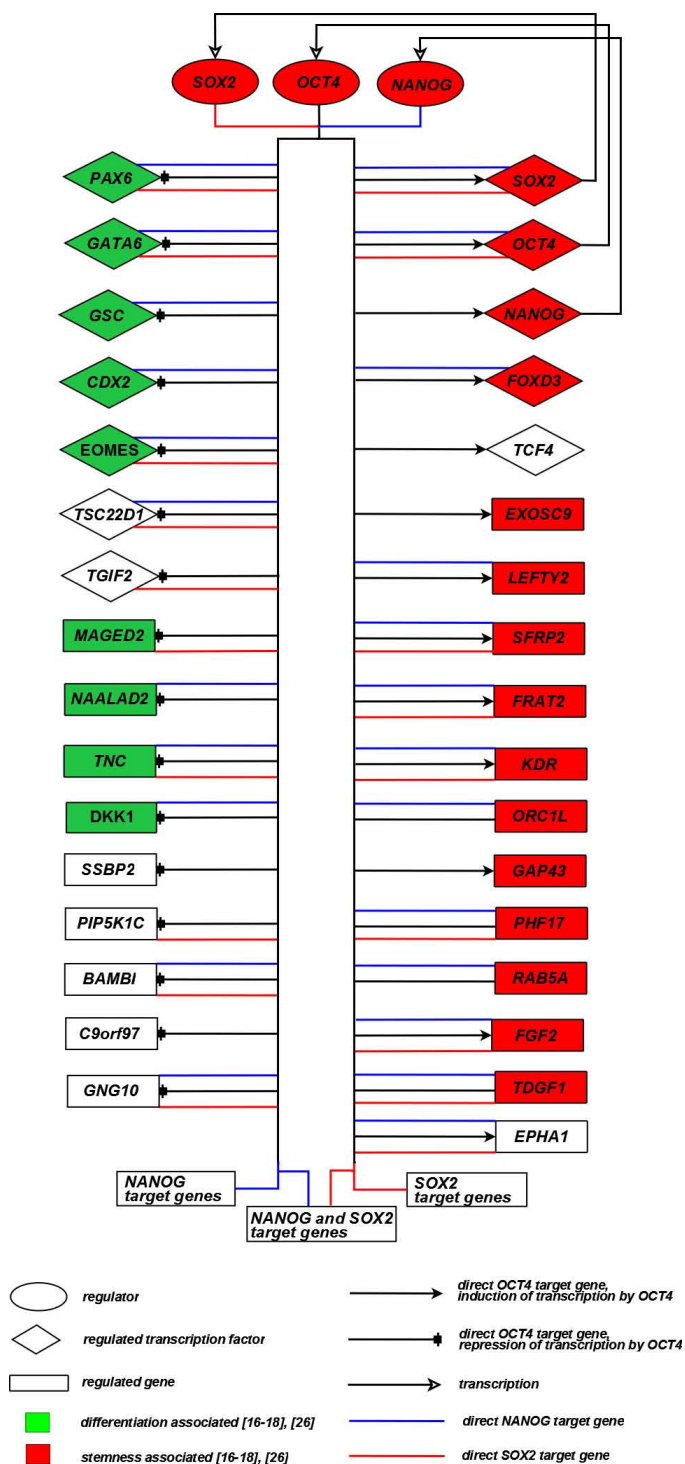


Figure 5
OCT4 core regulatory network. Core OCT4 transcriptional regulatory network identified by the integrative analysis of the re-analysed ChIP-on-Chip data, the OCT4 RNAi knock-down and the sequence-based octamer and sox-oct motif mapping. Green boxes represent genes associated with differentiation and red boxes indicate genes being specific for hESCs as annotated by several further public sources [16-18,26]. For white boxed genes no detailed annotation about differentiation or stemness characteristics was found by literature research. The network also incorporates the information on direct target genes from the re-analysed SOX2 (red lines) and NANOG (blue lines) ChIP-on-chip experiments.

ID	Motif	Complexity	Bits	Genes containing motif in promoter sequence (-8kb to TSS)	Similar to known motifs in TRANSFAC and JaspAr (selected from Additional File 7 and 8)
1		1.6479	25.03	RAB5A, <i>TGIF2</i>	<i>HNF4alpha</i> , <i>Pax-5</i> , <i>STAT5A</i> , <i>HNF6</i> , <i>POU6F1</i> , <u><i>NR2F1</i></u> , <u><i>Pax6</i></u>
2		1.5123	17.72	<i>C9orf97</i> , <i>CDX2</i> , <i>EOMES</i> , EPHA1, EXOSC9, FGF2, FOXD3, FRAT2, <i>GSC</i> , KDR, LEFTY2, <i>MAGED2</i> , NAALAD2, NANOG, ORC1L, <i>PAX6</i> , PHF17, <i>PIP5K1C</i> , SFRP2, SOX2, <i>SSBP2</i> , TCF4, TDGF1, <i>TGIF2</i> , <i>TNC</i> , <i>TSC22D1</i>	<i>Oct-4</i> , <i>Oct-1</i> , <i>Octamer</i> , <u><i>HMG-IY</i></u> , <u><i>STAT1</i></u>
3		1.4507	47.47	-	<i>HOXA5</i> , <i>HOXA4</i> , <u><i>Pax-4</i></u> , <i>E2F-1</i>
4		1.4497	26.79	SOX2	<i>Pax-5</i> , <i>Pax-4</i> , <i>Sp3</i> , <u><i>GATA2</i></u> , <u><i>NFKB1</i></u>
5		1.4391	17.16	<i>CDX2</i> , <i>DKK1</i> , FGF2, KDR, LEFTY2, NAALAD2, <i>PIP5K1C</i> , RAB5A, SFRP2, TDGF1	<i>Pitx2</i> , <i>EBF</i> , <i>HOXA7</i> , <i>OCT-x</i> , <u><i>REST</i></u> , <u><i>BRCA1</i></u> , <u><i>HMG-1</i></u> , <u><i>ZNF354C</i></u>
6		1.3337	24.98	<i>BAMBI</i> , FOXD3, FRAT2, SOX2, <i>SSBP2</i> , <i>TSC22D1</i>	<i>PEA3</i> , <i>Tra-1</i> , <u><i>Eip74EF</i></u>
7		1.2740	36.48	-	<i>STAT1</i> , <i>SOX</i> , <u><i>ELK4</i></u> , <u><i>GABPA</i></u>
8		1.1468	39.68	<i>PIP5K1C</i>	<i>POU3F2</i> , <u><i>Mycn</i></u> , <u><i>MYC-MAX</i></u> , <u><i>REST</i></u>
9		1.0990	33.37	-	<i>Mat1-Mc</i> , <i>SRY</i> , <i>SOX</i> , <u><i>REST</i></u> , <u><i>SP1</i></u> , <i>Sox5</i> , <i>FOXI1</i> , <i>Foxd3</i>
10		0.8941	16.92	<i>CDX2</i> , <i>EOMES</i> , EPHA1, EXOSC9, FOXD3, FRAT2, GAP43, <i>GNG10</i> , <i>GSC</i> , KDR, LEFTY2, <i>MAGED2</i> , NANOG, <i>PAX6</i> , PHF17, <i>PIP5K1C</i> , RAB5A, SFRP2, SOX2, <i>SSBP2</i> , TCF4, TDGF1, <i>TGIF2</i> , <i>TNC</i> , <i>TSC22D1</i>	<i>Oct-1</i> , <i>Oct-4</i> , <i>Octamer</i> , <i>OCT-x</i> , <u><i>HMG-IY</i></u> , <u><i>STAT1</i></u> , <i>Pax4</i> , <u><i>Foxa2</i></u>
11		0.6719	20.60	<i>BAMBI</i> , <i>DKK1</i> , EPHA1, FRAT2, <i>GATA6</i> , <i>GSC</i> , KDR, NANOG, <i>PAX6</i> , OCT4, RAB5A, SOX2, TDGF1, <i>TNC</i>	<i>ZNF219</i> , <i>VDR</i> , <i>UF1H3BETA</i> , <i>KROX</i> , <u><i>Pax4</i></u> , <u><i>RREB1</i></u> , <u><i>Klf4</i></u> , <u><i>SP1</i></u>
12		0.6323	16.31	<i>BAMBI</i> , <i>C9orf97</i> , <i>CDX2</i> , <i>DKK1</i> , EPHA1, EXOSC9, FOXD3, FRAT2, GAP43, <i>GATA6</i> , <i>GNG10</i> , <i>GSC</i> , KDR, LEFTY2, <i>MAGED2</i> , NANOG, ORC1L, <i>PAX6</i> , <i>PIP5K1C</i> , OCT4, RAB5A, SFRP2, SOX2, TDGF1, <i>TGIF2</i> , <i>TNC</i> , <i>TSC22D1</i>	<i>RAP1</i> , <i>FOXPI</i> , <i>Egr-3</i> , <i>Egr-2</i> , <i>Egr-1</i> , <u><i>CF2-II</i></u> , <i>Foxd3</i> , <u><i>Pax6</i></u>

Figure 6
De novo motif discovery. 12 motifs (second column) identified by a *de novo* motif discovery approach, sorted by their complexity value derived from the di-mer distribution (third column). The second and tenth motifs have a high similarity with known OCT4 related motifs. The fifth column lists the genes from the network that contain the motif within their promoter regions. Genes being suppressed by OCT4 are stated italic, genes being activated by OCT4 are stated bold. The last column shows selected similarity matched motifs from TRANSFAC (stated italic, [19]) and JaspAr (stated underlined, [31]) as received by STAMP [32] (see Additional files 8 and 9 for the full results).

(consensus sequence *GSmkGs*) and could be mapped to the promoter regions of *ORC1L* and further *OCT4* target genes (data not shown).

Discussion

We have performed an integrative study of complementary experimental techniques in order to identify the core gene regulatory network of *OCT4* in the context of maintaining pluripotency and preventing differentiation along the trophoblast lineage. This network represents a rather conservative selection since we have chosen only genes with high evidence that showed significant results in ChIP-on-chip analysis, RNAi analysis and promoter sequence analysis.

Figure 2 illustrates the limited overlap between the different technologies. For example, only 40 genes intersect between the ChIP-on-chip (12%) and the RNAi experiments (4%). This observation is in line with the results of a similar approach comparing the overlap of altered gene expression after *Oct4* silencing and TF binding in mouse ESCs (<9% overlap) [20]. Genes that show altered gene expression but no binding site may be regulated by an inter-dependent network, where loss of expression of one factor ultimately leads to the suppression of the others [16]. Additionally, the RNAi targets relate also to downstream effects independent of direct protein-DNA binding of *OCT4* which explains the higher number of RNAi targets (1,104) compared to ChIP-on-chip targets (308). Alternatively, TF binding may not be limited to the promoter region interrogated by the tiling arrays. On the other hand, genes having *OCT4* binding sites but do not show altered expression may be regulated by a more complex system of *OCT4* co-factors, epigenetic modifications like de-/methylation of CpG di-nucleotides within promoter regions or simply at later time points during differentiation into one of the three germ layers or into the trophectoderm lineage. Hence, independent validations such as the accessed knockdown experiment are critical in distinguishing functional from non-functional circuitries [20].

Although the presented network is rather conservative and potentially neglects genes regulated by *OCT4* together with unknown interaction partners, it represents the functional regulatory circuitry of direct *OCT4* target genes in hESCs as deduced by the available data. It is a well-known problem of both ChIP-on-chip experiments and motif prediction analyses to generate a large number of false positives. Additionally, RNAi experiments do not only reveal direct but – to a much higher extent – indirect targets. Thus, having a rather conservative process for identifying *OCT4* target genes has the benefit of narrowing down this large number of false positives. An indicator of this is the fact that the integration of the different experi-

ments purifies and enriches functional content of the resulting targets in all investigated functional classes by factors of 1.5 – 4 as is shown in Figure 4.

Recent studies report *OCT4/Oct4* expression in the adult, most frequently in the bone marrow of both humans and mice, particularly in hematopoietic and mesenchymal stem cells as well as in various sub-populations of multipotent progenitors [37]. It has been suggested that *Oct4* may not only be crucial for the maintenance of pluripotency in embryonic cells but may also play an important role for the self-renewal of somatic stem cells [37]. However, Lengner et al. [37] have shown that *Oct4*, even if expressed at low levels in somatic cells, is dispensable for the self-renewal of somatic stem cells, and for the regeneration of tissue in the adult, and is only rarely activated in somatic tumors. Based on these observations, we do not consider *OCT4/Oct4* to be a key player for transcriptional regulation of pluripotency in either mesenchymal stem cells and other adult stem cells. The identified core regulatory network of *OCT4* was created in the context of human embryonic stem cells for maintaining pluripotency and preventing differentiation along the trophoblast lineage.

Our *de novo* motif discovery approach did not only reveal known *OCT4* binding sites but also motifs similar to binding sites recognized by regulators that are known to interact with components of the *OCT4* regulatory network as well as genes that may have important functions as downstream effectors of *OCT4* but not yet described. Besides the co-factors presented above, a predominantly occurring motif is similar to a binding site recognized by Sp1 (Specificity protein 1), a transcription regulator that plays a role in TGF beta induced cell migration and mesenchymal transition, regulates angiogenesis, heart contraction, and aberrant expression is associated with several types of cancer. Yang et al. proposed that Sp1 or Sp3 play a critical role in controlling the transcriptional activity of *OCT4* by direct binding and an overexpression study showed that Sp1 positively regulates *OCT4* promoter activity [38]. The Sp1 motif was identified within the promoter regions of *OCT4* and of other *OCT4* target genes (see Figure 6). Sp1 is closely connected to the network as it binds to *FGF2*, *C-MYC*, *HOXB7*, *Spp1* (the latter two genes are upregulated by Sp1), and interacts with *Egr-1*. Moreover, Sp1 interacts with *CP2A*, a TF which in turn regulates *PAX6* [19] (not indicated in the extended network), a transcription factor which is a member of the differentiation related *OCT4* target genes. From the mouse model it is known that Sp1 binds to *Foxa1* and *Cdx2*. *Egr-1* (Early growth response 1) is a transcription factor that acts in apoptosis, angiogenesis, cell differentiation, regulates TNF production, cell proliferation and adhesion and aberrant expression of the gene is associated with several

types of cancer. HOXB7 (Homeo box B7) is a transcriptional activator and functions in DNA double strand break repair by nonhomologous end joining. Both, Egr-1 and HOXB7 bind to the promoter region of FGF2 and a motif similar to the binding site of Egr-1 was obtained (see Figure 6).

As another example, the *de novo* motif discovery identified a binding site similar to a motif recognized by STAT1 (Signal transducer and activator of transcription 1), a gene that mediates DNA replication, cell proliferation, apoptosis, and cell cycle regulation. It is known that STAT1 binds to C-MYC and is upregulated by Sp1. Several of the OCT4 target genes show a putative STAT1 binding site within their promoter region (see Figure 6).

HNF4A (Hepatocyte nuclear factor 4 alpha) has a known binding site similar to one of the obtained motifs. It is a transcription factor that inhibits GH1 induced STAT5 and JAK2 phosphorylation and functions in hepatocyte differentiation and blood coagulation. HNF4A expression is upregulated by Sp1 and is a target of GATA6, a transcription factor which is a member of the differentiation related OCT4 target genes. RAB5A and TGIF2 show a putative HNF4A binding site within their promoter region (see Figure 6).

PAX4 (Paired box gene 4) is a putative RNA polymerase II transcription factor that acts in positive regulation of cell proliferation and motifs similar to the known binding site for PAX4 were obtained. PAX4 itself has a binding site for HNF4A which is a downstream target of GATA6 [19]. There are putative PAX4 binding sites within the promoter regions of several OCT4 target genes (see Figure 6).

The computed OCT4 core regulatory network can be utilized in multiple ways. Well-characterized OCT4 target genes will help in extending the OCT4 network by suggesting further experimental work. The relatively high proportion of TFs in the OCT4 target set can be used for further inhibition studies or protein-DNA binding experiments. This leads to an extended radius of the network. For example, Additional file 4 shows that OCT4 has a positive regulatory effect on FGF2. FGF2 re-stimulation experiments performed by Greber et al. in hESCs revealed BMP4 as a downstream target of FGF2 signaling [27]. BMP4 expression was activated upon OCT4 knockdown in the original experiment as well, so both experiments consistently confirm that BMP4 is a negatively regulated downstream target of OCT4. Such an extended network and even the constructed core regulatory network will ultimately help in the study of stemness and early embryogenesis. Figure 4 shows functional enrichment for "embryonic development" that is increasing from 8%-30% with the integrative approach.

Finally, the identification of targets and co-factors of OCT4 might help in the design of iPS reprogramming protocols that use different TFs in order to generate and monitor cell status. C-MYC has already been successfully applied within a set of TFs for generating iPS cells through reprogramming. Figure 6 gives a guided hint for testing a variety of these co-factors.

Conclusion

The OCT4 dependent functional transcriptional regulatory network important in the analysis of human stem cell characteristics and cellular differentiation was reconstructed using an integrative approach. Functional information is largely enriched using an overlay of different experimental results. The *de novo* motif discovery points out several well known regulators closely connected to the network as well as less described potential downstream regulators of pluripotency and differentiation.

Methods

ChIP-on-chip data re-analysis

For identifying DNA regions occupied by OCT4, SOX2 and NANOG, Boyer et al. [12] performed two ChIP-on-chip experiments for each of the three transcription factors. They utilized a set of ten promoter arrays containing in total 399,309 60 mer oligonucleotides, and the design of the oligonucleotides was based on the NCBI build 35 of the human genome. As an update, the oligonucleotides were mapped [39] to the NCBI build 36.1 (hg18, Mar. 2006) and the updated allocation of 373,181 uniquely matched oligonucleotides to their genomic positions served as reference for the subsequent peak-finding. For all uniquely mapped oligonucleotides, the available raw expression data was background corrected, array-wise quantile normalized and replicates were normalized between arrays by applying Bioconductor's limma package [40]. For each oligonucleotide, a fold-enrichment was calculated by dividing the averaged signal intensities of the immunoprecipitated replicates by the averaged signal intensities of the whole-genome replicates. A histogram of the ratios from the OCT4 replicates is given in Figure 1. Potential binding events were defined as A) one oligonucleotide having a ratio within the upper 0.001 quantile of the total ratio distribution or B) two neighbouring oligonucleotides within a window of 1 kb where one oligonucleotide has a ratio within the upper 0.01 quantile and the other one has a ratio within the upper 0.05 quantile of the total ratio distribution. These two filters were defined by considering possible binding events between sonicated DNA fragments of an estimated averaged length of 550 bp and the oligonucleotides on the arrays with respect to the density of the newly assigned oligonucleotides relative to the genomic promoter regions. All identified peaks were connected to the closest TSS, if one exists within a distance of 10 kb. Genomic positions of TSSs were based on

Ensembl [41] and were downloaded via Biomart [42]. By this procedure, 308 genes of the original 623 identified OCT4 targets could be identified.

Motif mapping of promoter sequences

The promoter regions of the combined targets taken from the validated CHIP-on-chip and from the RNAi experiments were tested for the occurrence of the octamer and the SOX-OCT joint motifs. Position-specific count matrices were retrieved from TRANSFAC database v12.1 [19] for the octamer (TRANSFAC id V\$OCT_Q6) and the SOX-OCT joined motifs (TRANSFAC ids V\$OCT4_02 and V\$OCT4_01). These were converted to regularized and scaled position-specific scoring matrices (PSSMs) using an in-house implementation of the method of Rahmann et al. [43]. For each of the genes, -8 Kb to +2 Kb of the TSS were retrieved from the ENSEMBL (version 47) database and scanned with the PSSMs for the maximum scoring hit on each sequence. To focus on the upstream promoter region and have motifs of reasonable quality, only the subset of maximum scoring hits which lay in the upstream region of -8 Kb and scored above 70% of the maximum attainable score for a given PSSM were recorded.

Enrichment analysis

Enrichment analysis was conducted with the DAVID platform [25]. Official gene symbols were used as input, the *Homo Sapiens* species was selected as background and DAVID was executed with default parameter settings.

De novo motif discovery

In order to test for further regulatory co-factors of OCT4, we performed a *de novo* motif discovery analysis based on specific promoter regions of the 308 direct OCT4 target genes derived from the re-analysis of the CHIP-on-Chip data. By taking the genomic positions of the identified peaks as a reference (that is the position of an oligonucleotide or the centre of oligonucleotides detecting a peak, respectively), we assembled the sub-sequences of length 200 bp around the peaks (bandwidth of length 100 bp, as an example see arrows in Figure 3). The selected sub-sequences were repeat masked [30] and used as input for the TAMO package, a *de novo* motif discovery framework [44] that incorporates AlignACE [45], MDScan [46] and MEME [47]. The motif discovery was performed following the given sample code except the clustering module. Additionally, we used the Gibbs Motif Sampler [48] implementation of the CisGenome [49] framework with default parameter settings. All obtained motifs were compared to each other by applying the *minaligndiff* function of the TAMO distribution and when motifs occur with an alignment difference < 0.2, only the motif with the highest Bit score is further considered. Motifs with Bit score < 15 were discarded. Secondly, we computed entropy of the dimer distribution of the motif sequence as a measure for the

motif complexity. Motifs with complexity score < 0.6 were discarded. The remaining 12 unique motifs were sorted by their complexity value and are shown in Figure 6.

Database matching of discovered motifs

The discovered motifs were compared against two databases of known motifs using the STAMP tool [32]. Motifs were compared against the TRANSFAC (v11.3) [19] and JASPAR (v3) [31] databases using the recommended default parameter settings.

Abbreviations

iPS: induced pluripotent stem cells; TF: transcription factor; ESCs: embryonic stem cells; TSSs: transcription start sites; TFBS: transcription factor binding sites; PSSMs: position-specific scoring matrices.

Authors' contributions

LC carried out the CHIP-on-chip data re-analysis, the analysis of the positional distribution of OCT4 transcription factor binding sites, the integration of the individual methods, created the OCT4 core regulatory network, performed the functional enrichment analyses of the OCT4 target genes, the *de novo* motif discovery and the database matching of discovered motifs and drafted the manuscript. ASB performed the motif mapping to the promoter sequences. MV and HL participated in the design of the study and performed manuscript editing. JA and RH generated and analysed the RNAi data, conceived the study, participated in its design and co-ordination and helped to draft the manuscript. All authors have read and approved the final manuscript.

Additional material

Additional file 1

Summary of the results of the individual studies. The table includes all genes identified by the OCT4, SOX2, and NANOG CHIP-on-chip experiments [12] and by the RNAi mediated silencing of OCT4 function with subsequent microarray analysis [16]. The yellow column shows the influence of OCT4 to the transcription rate of its target genes (0 = no influence, 1 = OCT4 enhances transcription, -1 = OCT4 suppresses transcription). The green columns indicate OCT4 binding to its target genes as published by Boyer et al. and as identified via the presented data validation; the numbers indicate the distance between the TSS of the appropriate gene and the closest peak. The red and blue columns are constructed analogue but show the results for the SOX2 and NANOG experiments. The light brown columns represent the results of the octamer and SOX-OCT joint motif mapping (0 = promoter region contains the motif, 1 = promoter region does not contain the motif, n.a. = promoter region was not tested; we tested only the promoter regions of OCT4 target genes as identified by the re-analyzed CHIP-on-chip and by the RNAi experiments).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S1.xls>]

Additional file 2

Functional Annotation and Enrichment Analysis. The table includes a list of the 33 OCT4 target genes together with their full gene names, a list of enriched functional groups and the results of the functional annotation clustering as received by Dennis et al. [25].

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S2.xls]

Additional file 3

Glossary of target genes. A glossary for the 33 core OCT4 target genes that summarizes further independent published experimental validations on the regulatory influence of OCT4 to its presented target genes.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S3.doc]

Additional file 4

Extended network. An additional level of gene regulation has been added to the core OCT4 target network (Figure 5) by further literature and database mining. This additional figure shows the core network extended by known up- and downstream target genes of the respective TFs as given by TRANSFAC [19] and by another published work [27].

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S4.png]

Additional file 5

CDX2 subnetwork from ConsensusPathDB. The image illustrates the CDX2 centred sub-network as received from the ConsensusPathDB [28] and points out several known downstream target genes as well as a physical interaction between CDX2 and PAX6, another important differentiation associated TF included in the presented set of OCT4 target genes.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S5.png]

Additional file 6

OCT4 network as SBML. The text file contains the core OCT4 network in SBML format.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S6.txt]

Additional file 7

Discovered motifs as probability matrices. The text file includes the 12 identified motifs as probability matrices as received from the TAMO package [44].

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S7.txt]

Additional file 8

Motif database matching results- TRANSFAC. The pdf file contains the results from comparing the 12 discovered motifs to the TRANSFAC database (v11.3, [19]) using the STAMP tool [32].

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S8.pdf]

Additional file 9

Motif database matching results- JASPAR. The pdf file contains the results from comparing the 12 discovered motifs to the JASPAR database (v3, [31]) using the STAMP tool [32].

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-314-S9.pdf]

Acknowledgements

We thank Sean O'Keefe, Stefan Haas and Szymon Kielbasa for help in promoter sequence analysis. The work was funded by the EU within its 7th Framework Programme with the grant APO-SYS (HEALTH-F4-2007-200767), the German Ministry of Science within its National Genome Research Network (NGFN-Plus, FKZ01GS08111), the BMBF programme (Cell-based Regenerative Medicine-01GN0530) and the Max Planck Society.

References

1. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131(5)**:861-72.
2. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R: **In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state.** *Nature* 2007, **448(7151)**:318-24.
3. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA: **Induced pluripotent stem cell lines derived from human somatic cells.** *Science* 2007, **318(5858)**:1917-20.
4. Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA: **Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.** *Nat Biotechnol* 2008, **26(11)**:1269-75.
5. Nichols J, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A: **Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.** *Cell* 1998, **95(3)**:379-91.
6. Pesce M, Gross MK, Scholer HR: **In line with our ancestors: Oct-4 and the mammalian germ.** *Bioessays* 1998, **20(9)**:722-32.
7. Adjaye J, Bolton V, Monk M: **Developmental expression of specific genes detected in high-quality cDNA libraries from single human preimplantation embryos.** *Gene* 1999, **237(2)**:373-83.
8. Pesce M, Scholer HR: **Oct-4: gatekeeper in the beginnings of mammalian development.** *Stem Cells* 2001, **19(4)**:271-8.
9. Botquin V, Hess H, Fuhrmann G, Anastasiadis C, Gross MK, Vriend G, Scholer HR: **New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2.** *Genes Dev* 1998, **12(13)**:2073-90.
10. Nishimoto M, Fukushima A, Okuda A, Muramatsu M: **The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2.** *Mol Cell Biol* 1999, **19(8)**:5453-65.
11. Yuan H, Corbi N, Basilico C, Dailey L: **Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3.** *Genes Dev* 1995, **9(21)**:2635-45.
12. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122(6)**:947-56.
13. Scholer HR, Balling R, Hatzopoulos AK, Suzuki N, Gruss P: **Octamer binding proteins confer transcriptional activity in early mouse embryogenesis.** *EMBO J* 1989, **8(9)**:2551-7.
14. Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P, Ng HH: **Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells.** *Mol Cell Biol* 2005, **25(14)**:6031-46.

15. Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P: **Transcriptional regulation of nanog by OCT4 and SOX2.** *J Biol Chem* 2005, **280(26)**:24731-7.
16. Babaie Y, Herwig R, Greber B, Brink TC, Wruck W, Groth D, Lehrach H, Burdon T, Adjaye J: **Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells.** *Stem Cells* 2007, **25(2)**:500-10.
17. Adjaye J, Huntriss J, Herwig R, BenKahla A, Brink TC, Wierling C, Hultschig C, Groth D, Yaspo ML, Picton HM, Gosden RG, Lehrach H: **Primary differentiation in the human blastocyst: comparative molecular portraits of inner cell mass and trophectoderm cells.** *Stem Cells* 2005, **23(10)**:1514-25.
18. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J: **Interaction between Oct3/4 and Cdx2 determines trophoctoderm differentiation.** *Cell* 2005, **123(5)**:917-29.
19. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E: **TRANSFAC: transcriptional regulation, from patterns to profiles.** *Nucleic Acids Res* 2003, **31(1)**:374-8.
20. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nat Genet* 2006, **38(4)**:431-40.
21. Carpenter L, Zernicka-Goetz M: **Directing pluripotent cell differentiation using "diced RNA" in transient transfection.** *Genesis* 2004, **40(3)**:157-63.
22. Hanna LA, Foreman RK, Tarasenko IA, Kessler DS, Labosky PA: **Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo.** *Genes Dev* 2002, **16(20)**:2650-61.
23. Tompers DM, Foreman RK, Wang Q, Kumanova M, Labosky PA: **Foxd3 is required in the trophoblast progenitor cell lineage of the mouse embryo.** *Dev Biol* 2005, **285(1)**:126-37.
24. Liu Y, Labosky PA: **Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3.** *Stem Cells* 2008, **26(10)**:2475-84.
25. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, Integrated Discovery.** *Genome Biol* 2003, **4(5)**:P3.
26. Assou S, Le Carrouer T, Tondeur S, Strom S, Gabelle A, Marty S, Nadal L, Pantescio V, Reme T, Hugnot JP, Gasca S, Hovatta O, Hamamah S, Klein B, De Vos J: **A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas.** *Stem Cells* 2007, **25(4)**:961-73.
27. Greber B, Lehrach H, Adjaye J: **Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal.** *Stem Cells* 2007, **25(2)**:455-64.
28. Kamburov A, Wierling C, Lehrach H, Herwig R: **Consensus-PathDB—a database for integrating human functional interaction networks.** *Nucleic Acids Res* 2009:D623-8.
29. Chen X, Vega VB, Ng HH: **Transcriptional Regulatory Networks in Embryonic Stem Cells.** *Cold Spring Harb Symp Quant Biol* 2008.
30. Smit AFA, RHPG: **RepeatMasker 2008** [<http://repeatmasker.org>].
31. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A: **JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.** *Nucleic Acids Res* 2008:D102-6.
32. Mahony S, Benos PV: **STAMP: a web tool for exploring DNA-binding motif similarities.** *Nucleic Acids Res* 2007:W253-8.
33. Lebel R, McDuff FO, Lavigne P, Grandbois M: **Direct visualization of the binding of c-Myc/Max heterodimeric b-HLH-LZ to E-box sequences on the hTERT promoter.** *Biochemistry* 2007, **46(36)**:10279-86.
34. Lu X, Vogt PK, Boger DL, Lunec J: **Disruption of the MYC transcriptional function by a small-molecule antagonist of MYC/MAX dimerization.** *Oncol Rep* 2008, **19(3)**:825-30.
35. Narkis G, Ofir R, Landau D, Manor E, Volokita M, Hershkowitz R, Elbedour K, Birk OS: **Lethal contractural syndrome type 3 (LCCS3) is caused by a mutation in PIP5K1C, which encodes PIPKI gamma of the phosphatidylinositol pathway.** *Am J Hum Genet* 2007, **81(3)**:530-9.
36. Maehara K, Yamakoshi K, Ohtani N, Kubo Y, Takahashi A, Arase S, Jones N, Hara E: **Reduction of total E2F/DP activity induces senescence-like cell cycle arrest in cancer cells lacking functional pRB and p53.** *J Cell Biol* 2005, **168(4)**:553-60.
37. Lengner CJ, Camargo FD, Hochedlinger K, Welstead GG, Zaidi S, Gokhale S, Scholer HR, Tomilin A, Jaenisch R: **Oct4 expression is not required for mouse somatic stem cell self-renewal.** *Cell Stem Cell* 2007, **1(4)**:403-15.
38. Yang HM, Do HJ, Oh JH, Kim JH, Choi SY, Cha KY, Chung HM: **Characterization of putative cis-regulatory elements that control the transcriptional activity of the human Oct4 promoter.** *J Cell Biochem* 2005, **96(4)**:821-30.
39. Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12(4)**:656-64.
40. Smyth GK: **Limma: linear models for microarray data.** In *Bioinformatics and Computational Biology Solutions using R and Bioconductor* Edited by: Gentleman VCR, Dudoit S, Irizarry R, Huber W. New York: Springer; 2005.
41. Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V, Cutts T, Down T, Eyras E, Fernandez-Suarez XM, Gane P, Gibbins B, Gilbert J, Hammond M, Hotz HR, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Lehvaslaiho H, McVicker G, Melsopp C, Meidl P, Mongin E, Pettett R, Potter S, Proctor G, Rae M, Searle S, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Ureta-Vidal A, Woodward KC, Cameron G, Durbin R, Cox A, Hubbard T, Clamp M: **An overview of Ensembl.** *Genome Res* 2004, **14(5)**:925-8.
42. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W: **BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis.** *Bioinformatics* 2005, **21(16)**:3439-40.
43. Rahmann S, Muller T, Vingron M: **On the power of profiles for transcription factor binding site detection.** *Stat Appl Genet Mol Biol* 2003, **2**:Article7.
44. Gordon DB, Nekudova L, McCallum S, Fraenkel E: **TAMO: a flexible, object-oriented framework for analyzing transcriptional regulation using DNA-sequence motifs.** *Bioinformatics* 2005, **21(14)**:3164-5.
45. Hughes JD, Estep PW, Tavazoie S, Church GM: **Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae.** *J Mol Biol* 2000, **296(5)**:1205-14.
46. Liu XS, Brutlag DL, Liu JS: **An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments.** *Nat Biotechnol* 2002, **20(8)**:835-9.
47. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *Proc Int Conf Intell Syst Mol Biol* 1994, **2**:28-36.
48. Thompson W, Rouchka EC, Lawrence CE: **Gibbs Recursive Sampler: finding transcription factor binding sites.** *Nucleic Acids Res* 2003, **31(13)**:3580-5.
49. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH: **An integrated software system for analyzing ChIP-chip and ChIP-seq data.** *Nat Biotechnol* 2008, **26(11)**:1293-300.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

