Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology)
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Genome-wide analysis of OCT4 binding sites in glioblastoma cancer cells*#

Xue-feng FANG[§1,2], Wei-yi ZHANG[§1,2], Na ZHAO[2], Wei YU[2], Dong DING[2], Xu HONG[2], Li-sha LI[2], Hua-rong ZHANG[2], Shu ZHENG[†‡1], Biao-yang LIN[†‡2]

([1]*Key Laboratory of Cancer Prevention and Intervention of Ministry of Education, Key Laboratory of Molecular Biology in Medical Sciences of Zhejiang Province, Cancer Institute, the Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, China*)

([2]*Systems Biology Division, Zhejiang-California International Nanosystems Institute (ZCNI), Zhejiang University, Hangzhou 310029, China*)

[†]E-mail: zhengshu@zju.edu.cn; biaoylin@zju.edu.cn

Received Mar. 1, 2011; Revision accepted June 2, 2011; Crosschecked Sept. 13, 2011

Abstract: OCT4, a member of the POU family of gene products, is an octamer motif-binding transcription factor. As it is known to play a crucial role in cancer processes including proliferation, invasion, and chemoradioresistance, it is important to identify the direct targets of *OCT4* in living cancer cells. Here, chromatin immunoprecipitation-sequencing (ChIP-seq) was used to identify OCT4 binding sites in glioblastoma cancer cells. The results showed that 5438 OCT4 binding sites were localized in the glioblastoma cancer genome and that these sites contained a consensus sequence TTTkswTw (k=T or G, s=C or G, w=A or T), which occurred 3931 times in 2312 OCT4 binding regions. Furthermore, binding motifs of some other transcription factors were identified in OCT4 binding regions. Our results provide a valuable dataset for understanding gene regulation mechanisms underlying the function of *OCT4* in glioblastoma cancer.

Key words: OCT4, Chromatin immunoprecipitation-sequencing (ChIP-seq), DNA binding region, Glioblastoma
doi:10.1631/jzus.B1100059 **Document code:** A **CLC number:** R73

1 Introduction

OCT4, a member of the POU family of gene products (Verrijzer *et al.*, 1992; Dekker *et al.*, 1993; Herr and Cleary, 1995), is an octamer motif-binding transcription factor (TF), which is highly expressed in stem cells (Schoorlemmer *et al*., 1995; Pesce *et al.*, 1998). OCT4 has been considered as a key molecule which could induce somatic cell pluripotency (Takahashi and Yamanaka, 2006; Takahashi *et al*., 2007). OCT4 is also essential for maintaining undifferentiated and pluripotent populations of cells (Yu *et al*., 2007). Kim *et al*. (2009a; 2009b) reported that OCT4 can directly reprogram adult mouse and human neural stem cells to induced pluripotent stem (iPS) cells. These studied showed that OCT4 plays an important role in conferring the stemness of cells.

In general, cancer cells are similar to early embryonic cells with the properties of immortality, undifferentiation, and invasion (Monk and Holding, 2001). So, it is vital to study genes associated with embryogenesis and tumorigenesis. In recent years, there have been some reports on the deregulated expression of *OCT4* in several cancers, such as human embryonal carcinomas (Kraft *et al.*, 1996), testicular germ cell tumors and seminomas (Cheng *et al.*, 2007), lung cancer (Chen *et al*., 2008), and bladder

‡ Corresponding authors
§ The two authors contributed equally to this work
* Project supported by the Ministry of Science and Technology, China (Nos. 2004CB518707, 2006DFA32950, 2006AA02Z4A2, 2006AA02A303, 2007DFC30360, and 2008DFA11320), and the National Natural Science Foundation of China (No. 81101580)
# Electronic supplementary materials: The online version of this article (doi:10.1631/jzus.B1100059) contains supplementary materials, which are available to authorized users

carcinomas (Atlasi *et al.*, 2007). It is possible that aberrant expression of *OCT4* may contribute to the progression of these cancers. Atlasi *et al*. (2007) found that *OCT4* is overexpressed in bladder cancer tissues compared with normal bladder tissues and can enhance bladder cancer cell migration and invasion. Schreiber *et al*. (1994) showed that the expression of *OCT4* was significantly higher in high-grade primary gliomas than in low-grade gliomas at the mRNA and protein levels (Schreiber *et al.*, 1994). Chen *et al*. (2008) demonstrated that the expression of *OCT4* is crucial in maintaining the self-renewing, cancer stem-like, and chemoradioresistance properties of $CD133^+$ lung cancer cells.

So far, system-wide investigation of the function of *OCT4* has been focused only on embryonic cells. To understand the function of *OCT4* in a genome-wide manner in cancer, we have used chromatin immunoprecipitation-sequencing (ChIP-seq) to determine the genome binding sites of OCT4 in living glioblastoma cells. We report here the genome-wide targets of OCT4 in glioblastoma cells identified by ChIP-seq, which provide us with a comprehensive dataset for improving our understanding of the role of *OCT4* in tumor development.

2 Materials and methods

2.1 Cell culture

U251 cells were obtained from the American Type Culture Collection (Manassas, VA) and maintained in Dulbecco's modified Eagle medium (DMEM) with 10% fetal bovine serum (FBS).

2.2 Immunoblot analysis

Cells were lysed in RIPA cell lysis buffer (25 mmol/L Tris-HCl (pH 7.6), 150 mmol/L NaCl, 0.01 g/ml NP-40, 0.01 g/ml sodium deoxycholate, 1 g/L sodium dodecyl sulfate (SDS)) containing protease inhibitor cocktail kit (Pierce, Rockford, IL, USA). A total of 20 μg of protein was loaded for Western blot analysis. The OCT4 (ab18976) antibody was purchased from Abcam (Cambridge, MA, USA).

2.3 ChIP-seq procedure

The aforementioned anti-OCT4 antibody and normal IgG (SC-2027; Santa Cruz, CA, USA) were used for immunoprecipitation. Cells (1×10^7 U251) were fixed with 1% fresh formaldehyde and lysed in lysis buffer according to the protocol of Fang *et al*. (2010). The nuclei were resuspended and sonicated on ice to chromatin of 200–500 bp DNA. A total of 50 μl of Dynal protein G beads, 5 μg of antibody, and the sonicated chromatin were incubated at 4 °C overnight. The precipitated immunocomplex was treated with proteinase K at 65 °C for 2 h and DNA was purified with QIAquick polymerase chain reaction (PCR) purification kit (Qiagen, USA). ChIP DNA was amplified, ligated with adaptor, and repaired at the DNA end. The 150 bp DNA fragments were isolated from agar gel and sequenced with a Solexa Illumina 2G genetic analyzer.

2.4 ChIP-seq analysis

Solexa pipeline analysis was performed as described by Lin *et al*. (2009) and Fang *et al*. (2010). The SOAP program was used to map the 36 nucleotide sequence tags to the human genome with two mismatches allowed (Wang *et al.*, 2009; Fang *et al*., 2011). We used the MACS program to identify the enriched peaks of the OCT4 binding sites with an IgG ChIP group dataset as control. The program parameters were set as follows: *P*-value cutoff of 1.00×10^{-5}, model fold of 32, tag size of 36 bp of 5′ sequences, and effective size of 2.70×10^9 for the human genome. We annotated the datasets with the HG19 database using the CisGenome program (Ji *et al.*, 2008).

2.5 Validation of ChIP-seq targets by real-time PCR

A list of binding peaks was selected and quantitative real-time PCR was performed to validate the OCT4 binding region dataset. The primers used are listed in Table 1.

2.6 Analysis of distance to transcription start site

We used annotations of the UCSC human genome HG19 database to calculate the distance from transcription start sites (TSSs) to the binding sites. We also considered the strand's direction when we calculated the distance: for the positive strand, the distance was from the TSS start position to the end position of the OCT4 binding region; for the negative strand, the distance was to the OCT4 binding region start position.

Table 1 Primers for validation of ChIP-target sequence

Gene	Forward primer	Reverse primer
SOX4	CTGCTCTGTCCAGAGGCTTT	AGTGTCGCGAACTCAAATCG
JUNB	TCTCTCAAGCTCGCCTCTTC	ACGTGGTTCATCTTGTGCAG
KLF6	CACAGGCATTGACAGGTACG	ACTTGATTTGCTGGGTTTGG
MAP3K11	GTCCCAGAAGGGGTCTGAGT	ATCCTCACAGGAGGCACTGT
FBXO17	TCCCAACACCAGAGATTTCC	CTTCTCCTGCAAGGATCTCG
FOXA2	AGGTGCTTGAAGAAGCAGGA	GAGACGCAAGGGAGAAGAAA
PIP5K2C	GGAGACTATGGCGTCCTCCT	TTCTCACCGAGTGGGCTAC
PARK7	GTTGCACTCAACCTCAAGCA	AGAGGGGAGTGGTAGGATGC

2.7 Motif scanning and identification

For the identification of the novel motifs, we extended the OCT4 binding sites observed by ChIP-seq to 100 bp 3′ and 5′, then retrieved the sequences in the FASTA format. We used the RepeatMasker program (http://www.repeatmasker.org) to mask all human repeats in these sequences. To find out the up-regulation motifs, a program named Motif-Sampler (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/Motif_Sampler.html) was used (default parameters were applied). We used the WebLogo program (http://weblogo.berkeley.edu/) (Crooks *et al.*, 2004) to calculate the consensus sequences.

2.8 Gene ontology analysis

To analyze the enriched Gene Ontology (GO) terms of all evidence levels and categories statistically, we used the program High-Throughput GoMiner (Zeeberg *et al.*, 2005).

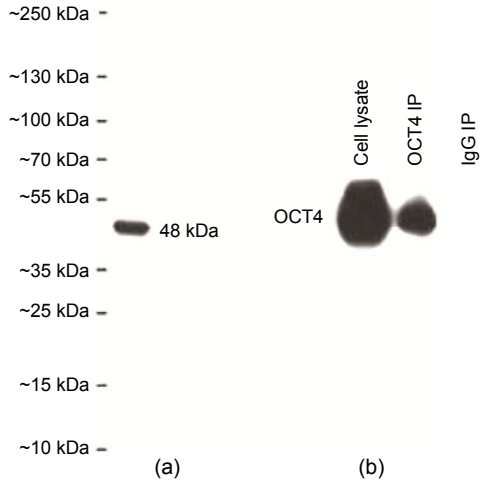

Fig. 1 Analysis of the OCT4 antibody by Western blot
(a) OCT4 protein in U251 cell lysate; (b) Immunoprecipitation (IP)-products using IgG and OCT4 antibodies

3 Results

3.1 Specificity of the antibody

The specificity of the antibody is essential to the success of ChIP-seq experiments. We characterized the anti-OCT4 antibody used in our study using Western blot. The antibody recognized a distinct band of the appropriate size in glioblastoma cells (Fig. 1).

3.2 ChIP-seq analysis of OCT4 binding regions in glioblastoma cells

To observe the genome-wide binding patterns of OCT4, a novel approach for the identification of TF binding sites, ChIP-seq technology, was applied (Barski *et al.*, 2007; Johnson *et al.*, 2007). From sequence analysis, we mapped a total of 217688 and 421892 sequence tags uniquely to the human genome (allowing two mismatches) for IgG and OCT4, respectively. MACS was then applied to analyze the enriched binding regions of OCT4. With a *P*-value of <0.01, 5438 OCT4 binding regions were identified in the human genome with IgG ChIP-seq as a negative control (Table S1). We randomly picked ten genes and confirmed that all ten gene DNA binding regions were enriched in the OCT4 IP DNA compared to the IgG IP DNA, using real-time PCR (Fig. 2). This suggested that the false positive rate of our dataset was negligible. With 5438 binding regions corresponding to 3798 genes, our results suggested that on average there were 1.4 enriched regions per gene.

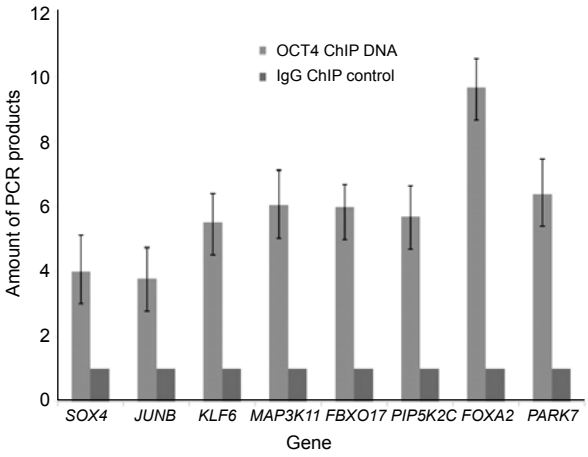

Fig. 2 Quantitative real-time PCR for the confirmation of ChIP-seq peaks

Relative amounts of PCR products from OCT4 ChIP and IgG ChIP are shown as bar graphs with the amount of IgG ChIP normalized to 1. Standard deviations are also shown for OCT4 IP

The OCT4 binding regions were mapped to the human genome (Table S1) using database HG19 with a cutoff-value of 50 kb distance. We calculated the distance from the OCT4 binding region to the known genes' TSSs and tabulated the frequency of OCT4 binding sites across the 5-kb distance interval before and after the TSSs. The OCT4 binding sites were around the TSSs, with about 41.05 % of binding sites mapped within a 5-kb distance (Fig. 3). We analyzed the function of OCT4 binding genes using the GoMiner program with Level 3 of the biology process. The enriched GO terms are shown in Table 2. The top enriched terms included GO:0010467 gene expression, GO:0046907 intracellular transport, GO:0006886 intracellular protein transport, GO:0015031 protein transport, and GO:0006412 translation. Note that OCT4 regulated 338 of the 1085 genes involved in gene expression of biological processes (GO:0010467),

Table 2 GO terms enriched in OCT4 binding genes identified by ChIP-seq

GO category	Number of total genes	Number of changed genes	Enrichment	lg*P*
GO:0010467_gene_expression	1085	338	1.201	−4.89743
GO:0046907_intracellular_transport	282	105	1.435	−4.89674
GO:0006886_intracellular_protein_transport	145	61	1.622	−4.85319
GO:0015031_protein_transport	160	63	1.518	−3.94681
GO:0006412_translation	140	56	1.542	−3.78437
GO:0006397_mRNA_processing	73	33	1.742	−3.56974
GO:0006396_RNA_processing	165	63	1.472	−3.50444
GO:0006986_response_to_unfolded_protein	30	17	2.184	−3.47247
GO:0051789_response_to_protein_stimulus	30	17	2.184	−3.47247
GO:0045184_establishment_of_protein_localization	189	70	1.428	−3.37452
GO:0006376_mRNA_splice_site_selection	12	9	2.891	−3.28214
GO:0006508_proteolysis	187	69	1.422	−3.28129
GO:0008380_RNA_splicing	88	37	1.621	−3.16857
GO:0042060_wound_healing	59	25	1.633	−2.37032
GO:0000375_RNA_splicing__via_transesterification_reactions	35	16	1.762	−2.05301
GO:0006915_apoptosis	430	133	1.192	−2.03365
GO:0002526_acute_inflammatory_response	11	7	2.453	−2.02957
GO:0009408_response_to_heat	11	7	2.453	−2.02957
GO:0051604_protein_maturation	11	7	2.453	−2.02957
GO:0006457_protein_folding	56	23	1.583	−2.02892
GO:0012501_programmed_cell_death	432	133	1.187	−1.96191
GO:0030162_regulation_of_proteolysis	9	6	2.570	−1.91899
GO:0051641_cellular_localization	377	117	1.196	−1.89567
GO:0008219_cell_death	442	135	1.177	−1.86254

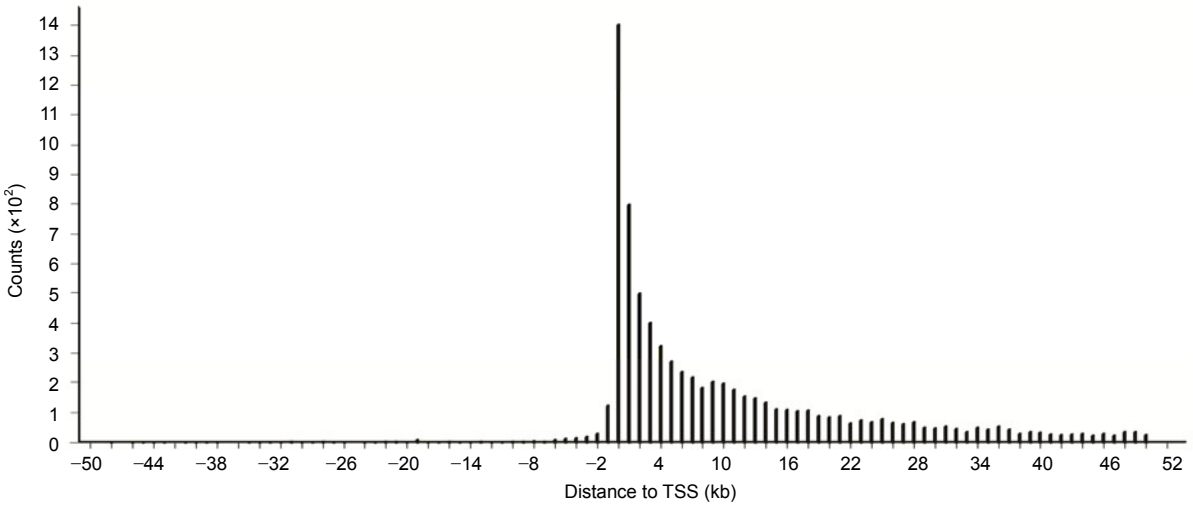

Fig. 3 Histogram of OCT4 binding sites around annotated transcription start sites (TSSs)
Frequencies of OCT4 island binding sites were calculated every 2 kb (*Y*-axis). Relative distance to TSSs is shown on the *X*-axis, negative and positive values indicate localizations 5′ and 3′, respectively, to TSSs

56 of the 140 genes involved in gene translation (GO:0006412), and 133 of the 430 genes involved in apoptosis (GO:0006915).

3.3 Identification of DNA binding consensus sequences and OCT4-associated TFs in the OCT4 binding regions

We applied the program MotifSampler to identify the enriched and unique consensus sequences in the binding regions of human OCT4 in glioblastoma cells. A consensus sequence of TTTkswTw (w=A or T, k=T or G, s=C or G) with a very high log likelihood score (17971.1) was found. The Table S2 shows this consensus sequence's output matrix and Table S3 shows 3931 instances of this motif in 2312 OCT4 binding regions. The logo of the consensus sequence is shown in Fig. 4. The MotifSampler program was used to systematically search for other known potential TFs which bind to the OCT4 binding regions and cooperate to regulate gene expression. All matrices of the TF's motifs were scanned and the matched matrices were tabulated and calculated. Among the top known TF's matrices identified as *OCT* cooperators, matching in more than 30% of cases, the *hepatocyte nuclear factor* (*HNF*) family, the *fork head transcription factor* (*FOX*) family, the *OCT* family, the *sex-determining region of Y-chromosome* (*SRY*) family, and several other TFs including *CEBP*, *PIT1_Q6*, *POU1F1_Q6*, and *PBX1_01* (Table 3) were found.

Table 3 Frequencies of known TF binding sites in the OCT4 binding regions

Known TF	Number of occurrences	Percentage of total sites (%)
SRY_01	6039	61.14
CEBP_01	5733	58.04
HNF3_Q6	5508	55.76
POU1F1_Q6	4359	44.13
FOX_Q2	4222	42.74
GATA_Q6	3857	39.05
SRY_02	3841	38.88
NKX62_Q2	3837	38.84
HFH3_01	3780	38.27
PIT1_Q6	3778	38.25
CEBP_Q2_01	3762	38.08
STAT5A_04	3595	36.39
TBP_01	3593	36.37
CRX_Q4	3573	36.17
DBP_Q6	3556	36.00
STAT4_01	3429	34.71
FREAC7_01	3395	34.37
HNF1_Q6	3378	34.20
TEF1_Q6	3314	33.55
STAT5A_03	3282	33.23
AML1_Q6	3045	30.83

Fig. 4 Consensus sequence of TTTkswTw (k=T or G, s=C or G, w=A or T) with log-likelihood score of 17971.1 identified by the MotifSampler program
The over-represented sequences were used as input for the WebLogo program (http://weblogo.berkeley.edu/) to display the consensus sequence graphically

4 Discussion

Identification of the binding sites of TFs on the genome is important for understanding the mechanism of function of genes. ChIP-combined next-generation sequencing can identify physical interactions between proteins and DNA in living cells and provide a whole-genome view of protein-DNA interactions (Barski and Zhao, 2009; Gilchrist *et al.*, 2009; Visel *et al.*, 2009). OCT4 has been considered as a major factor which could induce pluripotency in both mouse and human somatic cells. It plays a key role in the process of self-renewal and pluripotency. Recently, *OCT4* was consistently detected in several cancers and was shown to be able to contribute to cancer processes including proliferation and invasion, and to the property of chemoradioresistance. The diverse functions of *OCT4* in cancer suggested that we should study its target genes genome-wide.

In the present study, we used ChIP-seq to identify OCT4 binding regions in human glioblastoma cancer cells and found that OCT4 binding regions were enriched for AT with a consensus sequence TTTkswTw. The consensus sequence of OCT4 binding regions is consistent with previous in vitro studies in stem cells (Williams *et al*., 2004; Okumura-Nakanishi *et al*., 2005; Tay *et al*., 2008), which showed that the OCT4 protein binds DNA through AT-rich sequences with the motifs TTTGCwT (w=T or A). Marson *et al*. (2008) used ChIP-seq to identify binding regions of OCT4, SOX2, NANOG, and TCF3, and discovered that these four factors co-occupy 14230 sites in the genome. They showed that about 25% of binding sites were within 8 kb of the TSSs of 3289 annotated genes, 25% were within genes with a distance of more than 8 kb to the TSSs, and almost 50% occurred in intergenic regions far away from annotated start sites. In our study, we found that about 66.79% of OCT4 binding regions (3632 of 5438) mapped within 8 kb of the TSSs and about 42.72% (2323 binding regions) mapped within 2 kb of the TSSs in glioblastoma cancer cells.

ChIP-seq technology has been applied to the identification of target genes of *OCT4*, *SOX2*, and *NANOG* in human embryonic stem cells by Boyer *et al*. (2005). They identified 605 of the OCT4 binding promoter regions for known protein-coding genes in human embryonic stem cells. In our study, we found that 279 regions corresponding to 93 genes overlapped with our data. This suggests that the genes that are regulated by *OCT4* are conserved in embryonic stem cells and brain tumor cells.

Previous studies suggested that the OCT4 protein could either activate or repress specific target genes through interaction with different partner proteins such as *SOX2* and *NANOG* (Tay *et al*., 2008). *OCT4* target specificity may depend on its interacting proteins, which may include other TFs, co-activators, repressors and other proteins. These interacting proteins may interact with OCT4 to modulate OCT4's role as a TF. In the present study, we searched for the potential known TF binding motif in OCT4 binding regions. These TFs may act as *OCT4* cooperators and regulate expression of many genes together with *OCT4* in glioblastoma cancer.

The enriched analysis of GO terms suggested that *OCT4* is involved in some important biological processes, including gene expression (GO:0010467), translation (GO:0006412), mRNA processing (GO:0006397), wound healing (GO:0042060), and apoptosis (GO:0006915).

In summary, our genome-wide analysis of protein-DNA interaction in glioblastoma cells may help to explain the role of *OCT4* in glioblastoma carcinogenesis.

References

Atlasi, Y., Mowla, S.J., Ziaee, S.A., Bahrami, A.R., 2007. *OCT-4*, an embryonic stem cell marker, is highly expressed in bladder cancer. *Int. J. Cancer*, **120**(7): 1598-1602. [doi:10.1002/ijc.22508]

Barski, A., Zhao, K., 2009. Genomic location analysis by ChIP-seq. *J. Cell. Biochem*., **107**(1):11-18. [doi:10.1002/jcb.22077]

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., Zhao, K., 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, **129**(4):823-837. [doi:10.1016/j.cell.2007.05.009]

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al*., 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**(6):947-956. [doi:10.1016/j.cell.2005.08.020]

Chen, Y.C., Hsu, H.S., Chen, Y.W., Tsai, T.H., How, C.K., Wang, C.Y., Hung, S.C., Chang, Y.L., Tsai, M.L., Lee, Y.Y., *et al*., 2008. Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells. *PLoS One*, **3**(7):e2637. [doi:10.1371/journal.pone.0002637]

Cheng, C.J., Wu, Y.C., Shu, J.A., Ling, T.Y., Kuo, H.C., Wu, J.Y., Chang, E.E., Chang, S.C., Huang, Y.H., 2007. Aberrant expression and distribution of the OCT-4 transcription factor in seminomas. *J. Biomed. Sci*., **14**(6):797-807. [doi:10.1007/s11373-007-9198-7]

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res*., **14**(6):1188-1190. [doi:10.1101/gr.849004]

Dekker, N., Cox, M., Boelens, R., Verrijzer, C.P., van der Vliet, P.C., Kaptein, R., 1993. Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature*, **362**(6423):852-855. [doi:10.1038/362852a0]

Fang, X., Yu, W., Li, L., Shao, J., Zhao, N., Chen, Q., Ye, Z., Lin, S.C., Zheng, S., Lin, B., 2010. ChIP-seq and functional analysis of the *SOX2* gene in colorectal cancers. *OMICS*, **14**(4):369-384. [doi:10.1089/omi.2010.0053]

Fang, X., Yoon, J., Li, L., Yu, W., Shao, J., Hua, D., Zheng, S., Hood, L., Goodlett, D.R., Foltz, G., Lin, B., 2011. The *SOX2* response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. *BMC Genomics*, **12**(1):11. [doi:10.1186/1471-2164-12-11]

Gilchrist, D.A., Fargo, D.C., Adelman, K., 2009. Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. *Methods*, **48**(4):398-408. [doi:10.1016/j.ymeth.2009.02.024]

Herr, W., Cleary, M.A., 1995. The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev*., **9**(14):1679-1693. [doi:10.1101/gad.9.14.1679]

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., Wong, W.H., 2008. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, **26**(11):1293-1300. [doi:10.1038/nbt.1505]

Johnson, D.S., Mortazavi, A., Myers, R.M., Wold, B., 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**(5830):1497-1502. [doi:10.1126/science.1141319]

Kim, J.B., Greber, B., Arauzo-Bravo, M.J., Meyer, J., Park, K.I., Zaehres, H., Scholer, H.R., 2009a. Direct reprogramming of human neural stem cells by *OCT4*. *Nature*, **461**(7264):643-649. [doi:10.1038/nature08436]

Kim, J.B., Sebastiano, V., Wu, G., Arauzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., *et al*., 2009b. Oct4-induced pluripotency in adult neural stem cells. *Cell*, **136**(3):411-419. [doi:10.1016/j.cell.2009.01.023]

Kraft, H.J., Mosselman, S., Smits, H.A., Hohenstein, P., Piek, E., Chen, Q., Artzt, K., van Zoelen, E.J., 1996. Oct-4 regulates alternative platelet-derived growth factor α receptor gene promoter in human embryonal carcinoma cells. *J. Biol. Chem.*, **271**(22):12873-12878. [doi:10.1074/jbc.271.22.12873]

Lin, B., Wang, J., Hong, X., Yan, X., Hwang, D., Cho, J.H., Yi, D., Utleg, A.G., Fang, X., Schones, D.E., *et al*., 2009. Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor. *PLoS One*, **4**(8):e6589. [doi:10.1371/journal.pone.0006589]

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al*., 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, **134**(3):521-533. [doi:10.1016/j.cell.2008.07.020]

Monk, M., Holding, C., 2001. Human embryonic genes re-expressed in cancer cells. *Oncogene*, **20**(56):8085-8091. [doi:10.1038/sj.onc.1205088]

Okumura-Nakanishi, S., Saito, M., Niwa, H., Ishikawa, F., 2005. Oct-3/4 and Sox2 regulate *Oct-3/4* gene in embryonic stem cells. *J. Biol. Chem*., **280**(7):5307-5317. [doi:10.1074/jbc.M410015200]

Pesce, M., Wang, X., Wolgemuth, D.J., Scholer, H., 1998. Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. *Mech. Dev*., **71**(1-2):89-98. [doi:10.1016/S0925-4773(98)00002-1]

Schoorlemmer, J., Jonk, L., Sanbing, S., van Puijenbroek, A., Feijen, A., Kruijer, W., 1995. Regulation of *Oct-4* gene expression during differentiation of EC cells. *Mol. Biol. Rep*., **21**(3):129-140. [doi:10.1007/BF00997235]

Schreiber, E., Merchant, R.E., Wiestler, O.D., Fontana, A., 1994. Primary brain tumors differ in their expression of octamer deoxyribonucleic acid-binding transcription factors from long-term cultured glioma cell lines. *Neurosurgery*, **34**(1):129-135. [doi:10.1227/00006123-199401000-00019]

Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**(4):663-676. [doi:10.1016/j.cell.2006.07.024]

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**(5):861-872. [doi:10.1016/j.cell.2007.11.019]

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., Rigoutsos, I., 2008. MicroRNAs to *Nanog*, *Oct4* and *Sox2* coding

regions modulate embryonic stem cell differentiation. *Nature*, **455**(7216):1124-1128. [doi:10.1038/nature07299]

Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., van Leeuwen, H.C., Strating, M.J., van der Vliet, P.C., 1992. The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.*, **11**(13):4993-5003.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., *et al.*, 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**(7231):854-858. [doi:10.1038/nature07730]

Wang, W., Zhang, P., Liu, X., 2009. Short read DNA fragment anchoring algorithm. *BMC Bioinformatics*, **10**(s1):S17. [doi:10.1186/1471-2105-10-S1-S17]

Williams, D.C.Jr., Cai, M., Clore, G.M., 2004. Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.*Hoxb1*-DNA ternary transcription factor complex. *J. Biol. Chem.*, **279**(2):1449-1457. [doi:10.1074/jbc.M309790200]

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al.*, 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**(5858):1917-1920. [doi:10.1126/science.1151526]

Zeeberg, B.R., Qin, H., Narasimhan, S., Sunshine, M., Cao, H., Kane, D.W., Reimers, M., Stephens, R.M., Bryant, D., Burt, S.K., *et al.*, 2005. High-throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics*, **6**(1):168. [doi:10.1186/1471-2105-6-168]

List of electronic supplementary materials

Table S1 Regions enriched for OCT4 in human cancer cells

Table S2 MotifSampler's report: the consensus sequences in the binding regions of human OCT4 in glioblastoma cells

Table S3 MotifSampler's report: instances of the motif in OCT4 binding regions