

Supplementary Data

The Supplementary Data includes Supplementary Materials and Methods, and Figures (Supplementary Figs. S1–S10).

Supplementary Materials and Methods

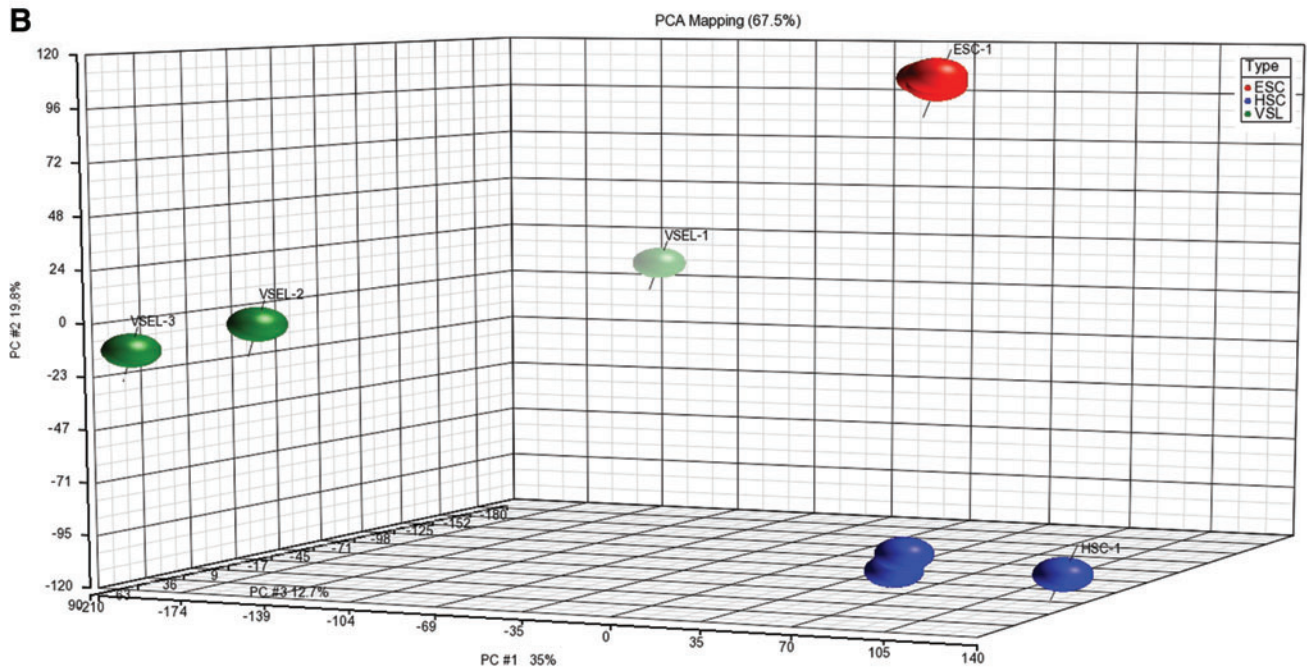
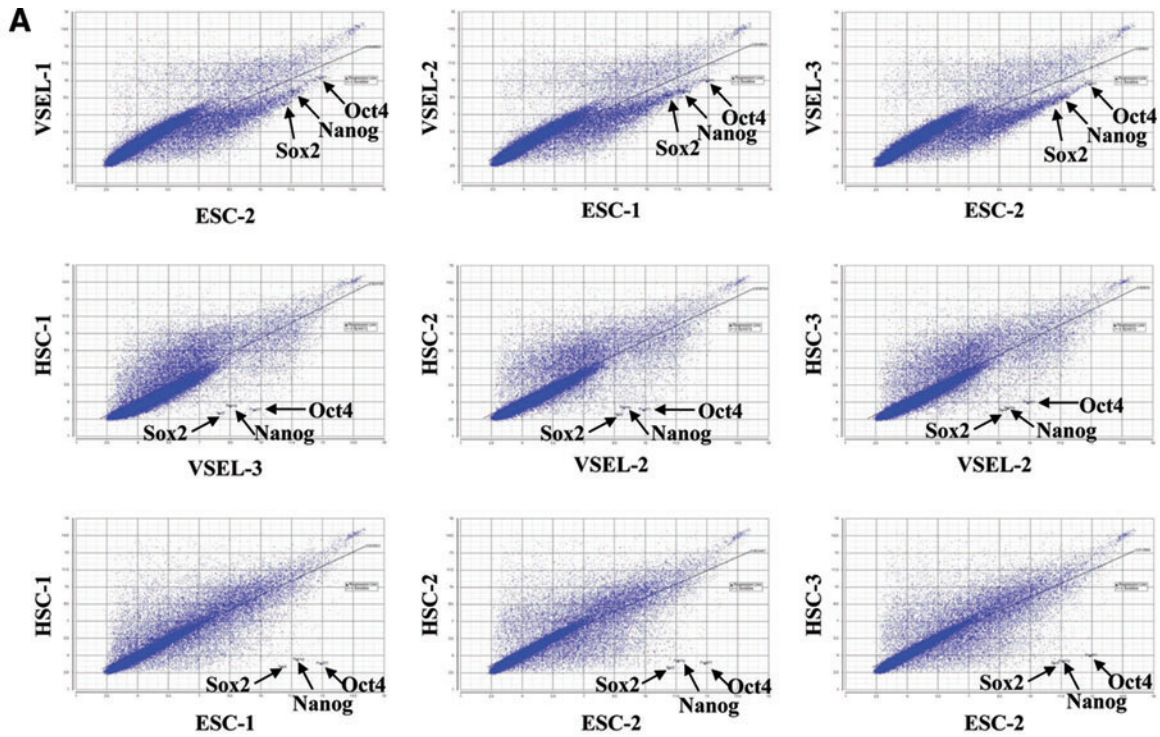
Twenty-cell gene expression profiling

For 20-cell cDNA library synthesis, fluorescence-activated cell sorting-sorted *Sca-1⁺Lin⁻CD45⁻* very small embryonic-like stem cells (VSELs), *Sca-1⁺Lin⁻CD45⁺* hematopoietic stem cells (HSCs), and trypsinized embryonic stem cell (ESC)-D3 were distributed at 20 cells per well into a 384-well plate (Thermo Scientific) containing 4.5 μ L of single-cell lysis buffer [1 \times PCR buffer II (Applied Biosystems), 1.5 mM MgCl₂ (Applied Biosystems), 0.5% NP40, 5 mM DTT, 0.3 U/ μ L RNaseOUT™ Recombinant RNase Inhibitor (Invitrogen), 0.3 U/ μ L RNase Inhibitor (Applied Biosystems), 0.2 ng/ μ L V1(dT)₂₄ primer, and 0.2 mM of each dNTP mix (Invitrogen)] using a MoFlo cell sorter (Dako). The V1(dT)₂₄ primer sequence is “atatggatcggcgcgccgctgacttttttttttttttttttttt.” After transfer into a 0.5-mL-thin microtube, cell lysis was performed at 70°C for 90 s, and followed by immediately cooling on ice for 1 min. After adding 0.3 μ L of reverse transcriptase (RT) mixture [133.3 U/ μ L SuperScript III (Invitrogen), 3.33 U/ μ L RNaguard RNase Inhibitor (GE Healthcare), and 1.3 μ g/ μ L T4 gene 32 protein (Roche)], the RT reaction was incubated at 50°C for 5 min, followed by heat inactivation of the RT enzyme at 70°C for 10 min. After cooling on ice for 1 min and brief centrifugation, to each reaction tube was added 1.0 μ L of Exonuclease I mixture [1 \times Exonuclease I buffer (Takara) and 0.5 U/ μ L Exonuclease I (Takara)] and then incubated at 37°C for 30 min, followed by heat inactivation at 80°C for 25 min. The poly(A) tailing reaction was performed by adding 6 μ L of poly(A) tailing buffer [1 \times PCR buffer II, 1.5 mM MgCl₂, 3 mM dATP (GE Healthcare), 0.1 U/ μ L RNaseH (Invitrogen), and 0.75 U/ μ L TdT (Invitrogen)] and then incubating at 37°C for 15 min with subsequent heat inactivation at 70°C for 10 min. The synthesized poly(A)-tailed RT product was converted to double-stranded cDNA by incubating for 1 polymerase chain reaction (PCR) cycle (3 min at 95°C, 2 min at 50°C, and 3 min at 72°C) using PCR mixture I [1 \times ExTaq buffer, 1 mM dNTP mix, 20 ng/ μ L V3(dT)₂₄ primer, and 0.05 U/ μ L Hot Start ExTaq (Takara)]. The V3(dT)₂₄ primer sequence is “atatctc-gagggcgcgccgcatccttttttttttttttttttttt.” The double-stranded 20-cell cDNA library was further amplified using PCR mixture II [V3(dT)₂₄ primer is replaced by V1(dT)₂₄ primer in PCR mixture I] and 24 PCR cycles (30 s at 95°C, 1 min at 67°C, and 3 min at 72°C with 6-s extension per cycle) and followed by 1 cycle of 10 min at 72°C. The synthesized 24-cycle PCR products were purified using Qiaquick PCR purification kit (Qiagen). For initial screening of the quality of the 20-cell cDNA libraries, purified PCR products diluted 20-fold were used to examine the expression of the indicated genes by real-time quantitative PCR (RQ-PCR). All primers used in RQ-PCR are available upon request. For microarray analysis, T7 promoter sequence was joined to the 5' terminus of 24-cycle PCR products by reaction with PCR mixture III

[1 \times ExTaq buffer, 1 mM dNTP, 20 ng/ μ L T7V1 primer, 20 ng/ μ L V3(dT)₂₄ primer, and 0.05 U/ μ L Hot Start ExTaq] with the following PCR cycles: 1 cycle (5 min 30 s at 95°C, 1 min at 64°C, and 5 min 18 s at 72 °C), 8 cycles (30 s at 95°C, 1 min at 67 °C, and 5 min 18 s at 72°C with 6-s extension per cycle), and 1 cycle of 10 min at 72°C. The T7V1 primer sequence is “ggccagtgaattgtaatacactactataggaggcgatgatgga tccggcgcgccgctgac.” The resultant PCR products were purified using Qiaquick PCR purification kit. Employing 2% agarose gel electrophoresis, DNA byproducts shorter than 300 bp were removed and PCR products longer than 300 bp were prepared using the QIAquick Gel Extraction kit (Qiagen). The gel-eluted T7-primed PCR products were biotin labeled using the GeneChip® 3' in vitro transcription kit (Affymetrix), starting from “In vitro Transcription to Synthesize Labeled aRNA.”

Stem cell transcriptome analysis

The principal component analysis (PCA) plot, scatter plot, and heatmap with hierarchical clustering of the microarray data from VSELs, HSCs, and ESC-D3 cells were prepared using Partek software (Partek, Inc.). The PCA analysis was performed using Partek software with default setting. In heatmap analysis, hierarchical agglomerative clustering with Spearman's rank correlation coefficient and average linkage were applied to both rows (samples) and columns (probe-sets) and heatmaps were produced by arranging the rows and columns according to the clustering outputs. The heatmaps for global transcriptome (45,101 probes) and protein translation and ubiquitination (949 probes) gene ontology-organized gene lists were individually developed. Functional analysis of the transcriptomes of VSELs, HSCs, and ESC-D3 cells was performed using Ingenuity Pathway Analysis (IPA) software version 8.7 (Ingenuity Systems, Inc.) by core and comparison analysis for gene networks, biofunctions, and canonical pathways. For general global transcriptome comparison (Fig. 1), the *P* value cut-off was 4.2172e-005 (false discovery rate [FDR]=0.002 in VSEL vs. HSC comparison) and the expression cut-off value was set as a 2-fold change up or down (1,058 genes were eligible for data analysis). To investigate the unique VSEL characteristics (Fig. 2), the gene list was filtered as 2-fold high or low expressed in the VSEL versus HSC comparison, and at the same time, high or low expressed in the VSEL versus ESC comparison. For VSEL targets specifically high expressed, the *P* value cut-off was 0.00306534 (FDR=0.025 in VSEL vs. HSC) and 362 eligible genes were used for IPA comparison; for VSEL targets specifically low expressed, the *P* value cut-off was 0.00016463 (FDR=0.05 in VSEL vs. HSC) and 516 eligible genes were used for IPA comparisons. The minimum resolution for multiple probes was set at the experimental *P* value. Red and green represented upregulated and downregulated values, respectively. The biofunctions and canonical pathway analysis of the indicated gene lists were performed using the default settings, with a threshold value of 0.05 and Fisher's exact test for scoring method.

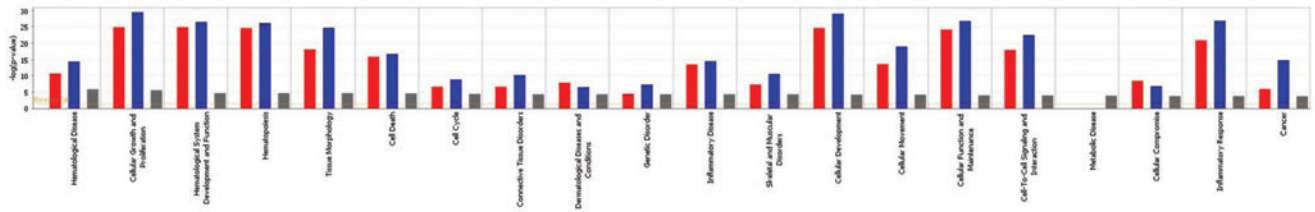


SUPPLEMENTARY FIG. S1. (A) Scatter plot analysis between individual 20-cell transcriptome for the indicated stem cell population. Regression line is depicted as diagonal black line with slope (end of line) and R square (box). The dots corresponding to Oct4, Nanog, and Sox2 are pointed with arrow. (B) PCA mapping of 20-cell microarray results for the indicated stem cell population. Each stem cell transcriptome was projected with first 3 principal components (PCs) as the total variance accounts for 67.5% with 35% for PC1, 19.8% for PC2, and 12.7% for PC3. Note that VSEL-2 and VSEL-3 libraries used in Figs. 1A and 3A were tightly clustered but distinct to VSEL-1, coordinated with heatmap clustering (Fig. 1C). PCA, principal component analysis.

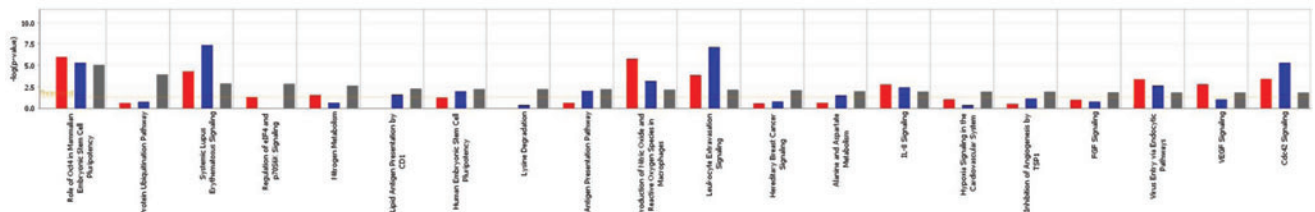
IPA comparison analysis for global transcriptome

VSEL vs HSC
ESC vs HSC
VSEL vs ESC

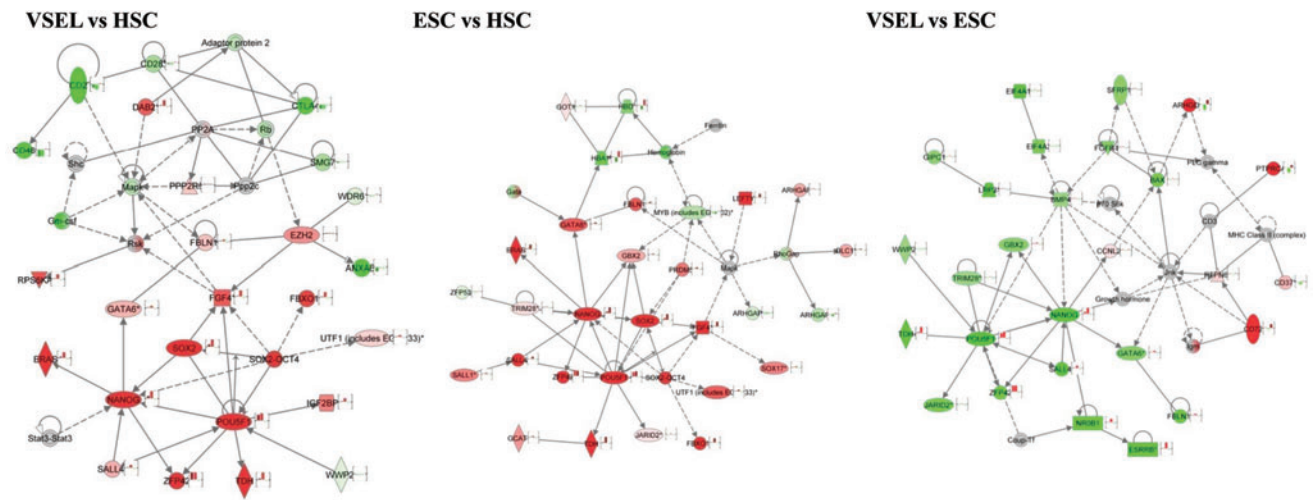
A Bio Functions (Ordered based on VSEL versus ESC comparison)



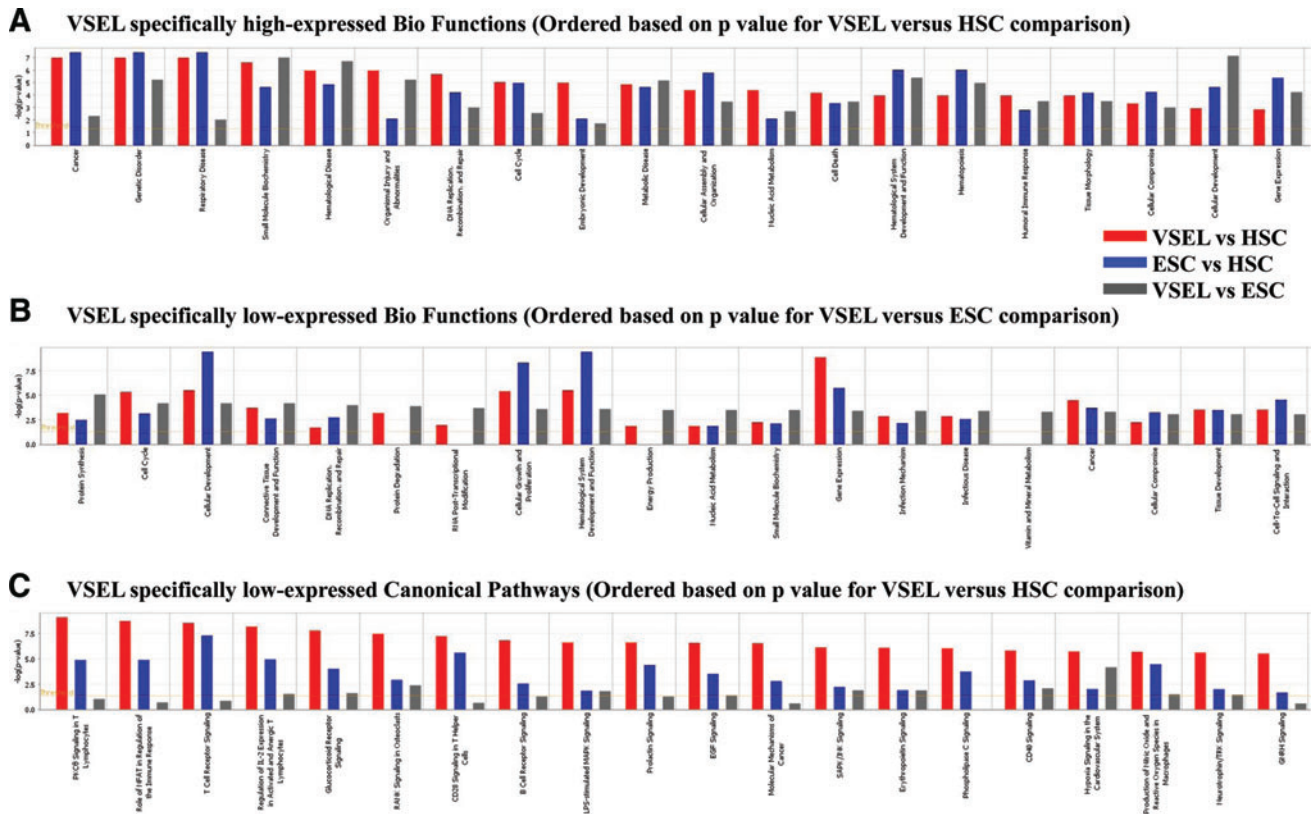
B Canonical Pathways (Ordered based on VSEL versus ESC comparison)



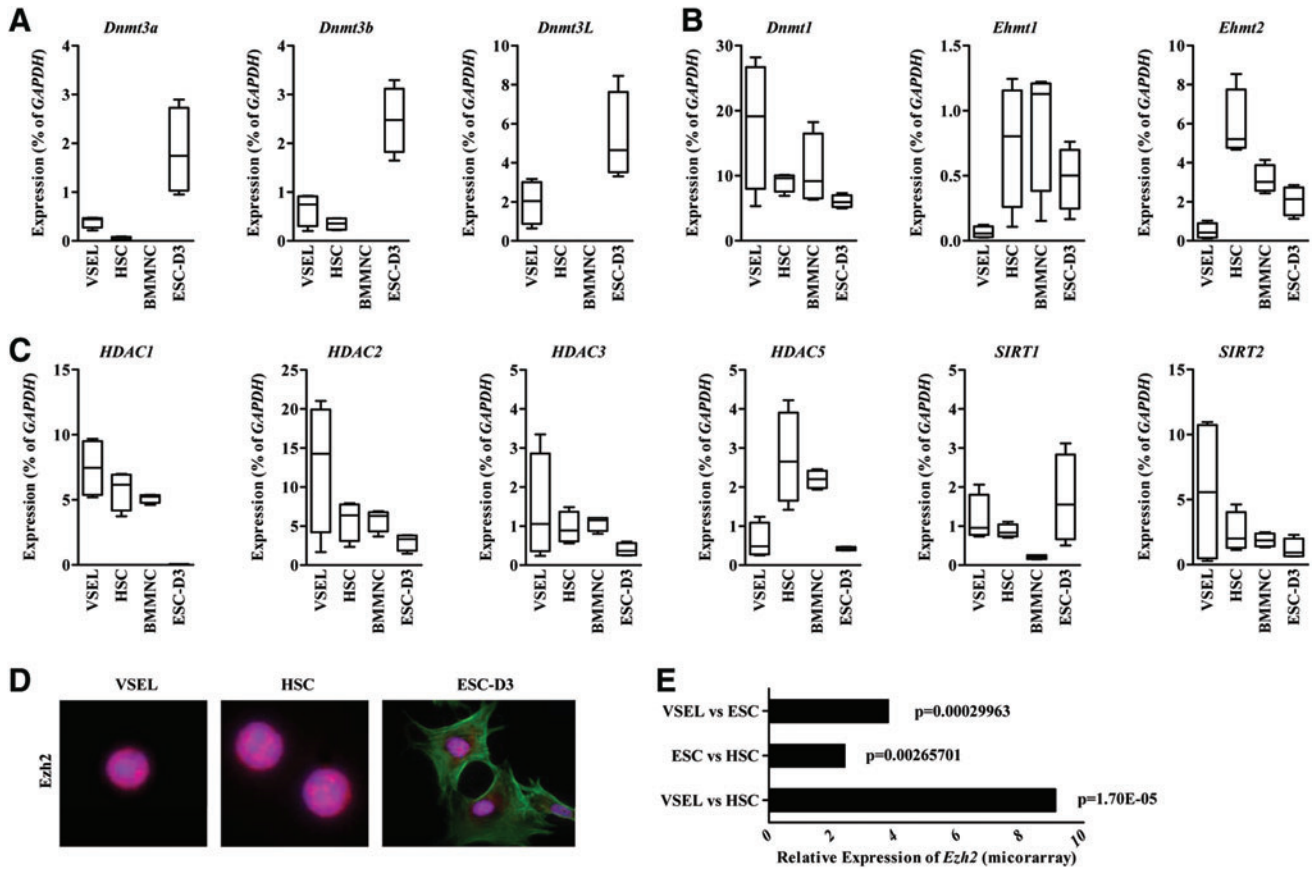
SUPPLEMENTARY FIG. S2. Top 20 for biofunction (A) and canonical pathways (B), which are highly represented in IPA comparison analysis for the indicated stem cell global transcriptome. Note that the graphs are ordered by experimental *P* value in VSEL versus ESC comparison. Yellow line indicates the threshold (0.05). ESC, embryonic stem cell; IPA, ingenuity pathway analysis.



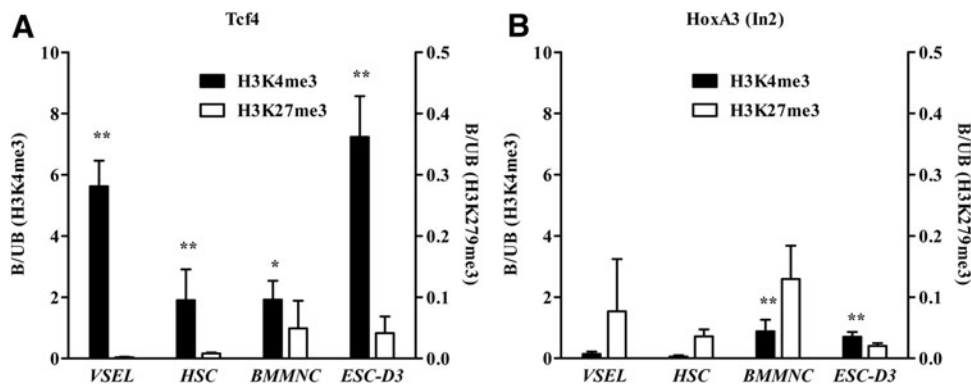
SUPPLEMENTARY FIG. S3. Highly represented gene network interacting with Oct4-Nanog-Sox2 pluripotency network in VSEL versus HSC (left), ESC versus HSC (middle), and VSEL versus ESC (right) comparison. The gene networks were illustrated with overlaying all experimental values for the indicated stem cell comparison dataset. Bar graphs next to genes indicate the fold change of the corresponding comparisons. High and low expressions of genes are represented as red and green colors, respectively. HSC, hematopoietic stem cells.



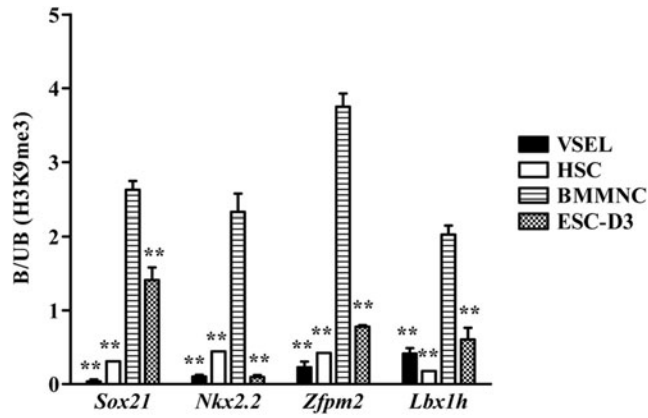
SUPPLEMENTARY FIG. S4. Top 20 for biofunction, which are highly represented in IPA comparison analysis using VSELS specifically high-expressed (A) and low-expressed (B) gene lists. The VSELS specifically high- and low-expressed genes are defined as increased and decreased expression in both VSEL versus HSC and VSEL versus ESC comparisons, respectively. (C) Top 20 for canonical pathway in IPA comparison analysis for VSELS specifically low-expressed gene lists. The graphs are ordered by experimental P value in the indicated stem cell comparison. Yellow line indicates the threshold (0.05). VSELS, very small embryonic-like stem cells.



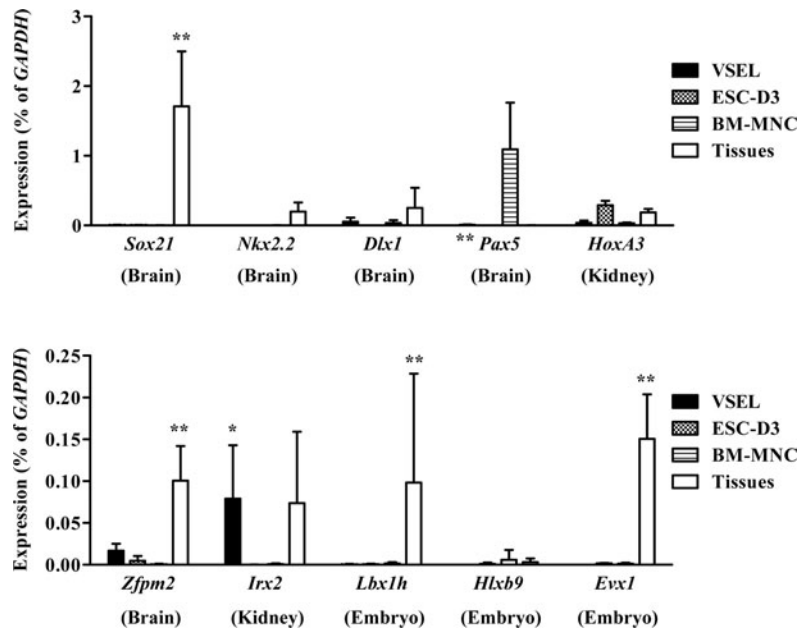
SUPPLEMENTARY FIG. S5. Expression of DNA methyltransferases and histone and modifying enzyme transcripts in murine VSELs. RQ-PCR results for DNA methyltransferases (*Dnmts*, **A**), euchromatic histone methyltransferase (*Ehmts*, **B**), and several histone deacetylases (*HDACs*, **C**) using the indicated single-cell level cDNA library (24-cycle PCR product). Expression level (% of GAPDH) was shown as the boxed region with the black bar for the median and the “whiskers” for the extreme values; $n=4$. **(D)** Immunostaining of Ezh2 protein in the indicated cells. Nucleus was stained with DAPI blue and cytoplasm for ESC-D3 cells was stained with F-actin (FITC conjugate pallodin). **(E)** The relative expression of *Ezh2* from microarray data comparison between the indicated stem cell populations. P value was depicted beside each bar. RQ-PCR, real-time quantitative PCR; Ezh2, enhancer of zeste drosophila homolog 2.



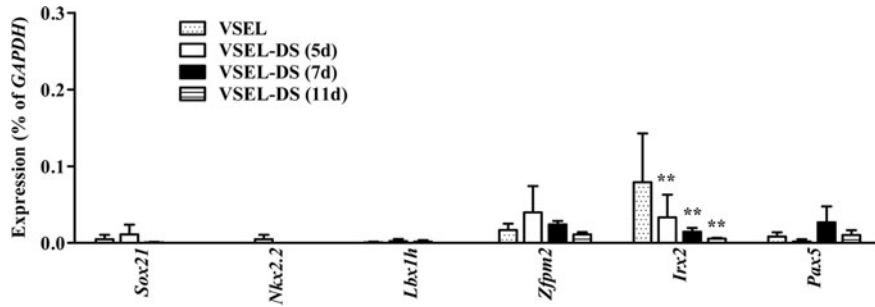
SUPPLEMENTARY FIG. S6. Quantitative ChIP (qChIP) analysis about H3K4me3 and H3K27me3 modification in the non-BD-enriched loci in the indicated cells. Note that the promoter region of *Tcf4* (**A**) and intron 2 (Ins2) of the *HoxA3* locus (**B**) in VSELs were enriched with only H3K4me3 and H3K27me3 modification, respectively. The enrichment of H3K4me3 (filled box) and H3K27me3 (open box) modification shown as the mean \pm S.D. ($n=4$) was plotted to left and right y axis, respectively. * $P < 0.05$, ** $P < 0.01$ compared with H3K27me3. Two-way ANOVA with Bonferroni posttests was used for statistical analysis. BD, bivalent domain; ANOVA, analysis of variance.



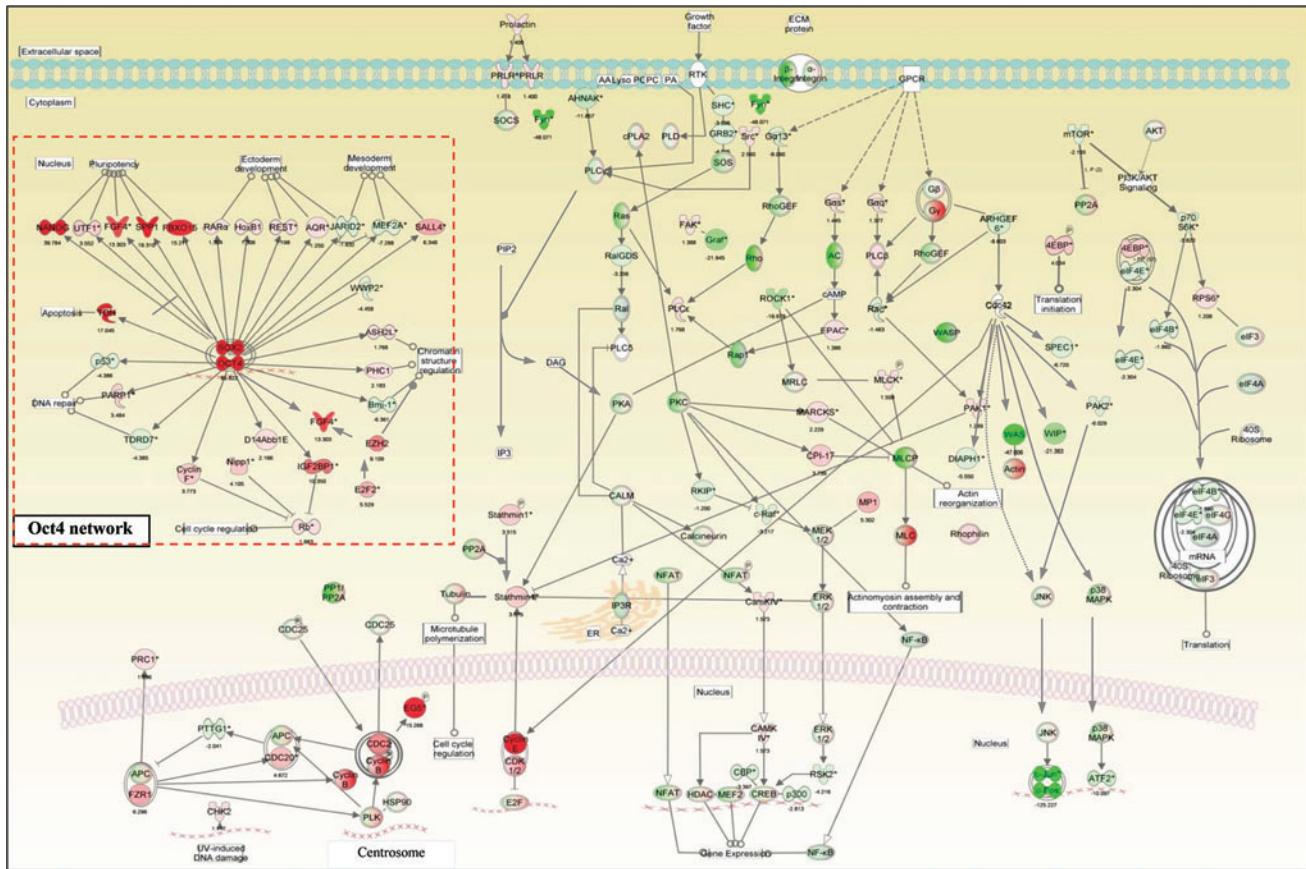
SUPPLEMENTARY FIG. S7. qChIP analysis for H3K9me3 for BD target gene promoters in the indicated cells. The enrichment of H3K9me3 was represented as the ratio for bound (B) to unbound (UB) fractions and shown as the mean \pm S.D. ($n=4$). $**P < 0.01$ as compared with BMMNCs. Two-way ANOVA with Bonferroni posttests was used for statistical analysis. BMMNCs, bone marrow mononuclear cells.



SUPPLEMENTARY FIG. S8. Repression of BD target gene expression in freshly isolated VSELs. RQ-PCR analysis for BD target gene expression in the indicated cells. Regular RT products from fluorescence-activated cell sorting-sorted ($\sim 20,000$ cells) or cultured ESC-D3 cells were used for RQ-PCR. Tissues used for positive control are shown below gene list. The embryo and brain or kidney tissues are prepared from embryonic day 11.5 to 13.5 embryos and 4-week-old C57BL/6 mice, respectively. Expression level (% of GAPDH) was shown as the mean \pm S.D. ($n=4$). $*P < 0.05$, $**P < 0.01$ as compared with BMMNCs. $**Pax5$ indicates that P values for all the experimental groups are less than 0.01. In this case, one-way ANOVA with Bonferroni posttests was used for statistical analysis; RT, reverse transcriptase.



SUPPLEMENTARY FIG. S9. Gene expression of BD target genes during VSEL-DS formation. RQ-PCR analysis for BD target gene expression in cells isolated from the indicated days of VSEL-DSs. Expression level (% of GAPDH) was shown as the mean \pm S.D. ($n=4$). $**P < 0.01$ as compared with freshly isolated VSELS. Two-way ANOVA with Bonferroni posttests was used for statistical analysis.



SUPPLEMENTARY FIG. S10. Summary for VSEL signaling network. Compared with HSCs, VSELS significantly express at a high level Oct4 pluripotent network (red dotted box) and several cell cycle checkpoint-related genes. At the same time, they particularly show the low expression of the protein turnover-related, growth factor/mitogen-activated (e.g., Ras, PI3K, PKA, PKC, and PLC), small G protein-related signaling molecules. Relative expression is shown as red (high expressed) and blue (low expressed) colors with overlaying all experimental values from VSEL versus HSC comparison. Fold change for gene expression between VSEL and HSC is shown below indicated genes.