

1 Comprehensive epigenome characterization reveals 2 diverse transcriptional regulation across human 3 vascular endothelial cells

4
5 Ryuichiro Nakato^{1,2,*}, Youichiro Wada^{2,3,*}, Ryo Nakaki⁴, Genta Nagae^{2,4}, Yuki Katou⁵,
6 Shuichi Tsutsumi⁴, Natsu Nakajima¹, Hiroshi Fukuhara⁶, Atsushi Iguchi⁷, Takahide
7 Kohro⁸, Yasuharu Kanki^{2,3}, Yutaka Saito^{2,9,10}, Mika Kobayashi³, Akashi Izumi-Taguchi³,
8 Naoki Osato^{2,11}, Kenji Tatsuno⁴, Asuka Kamio⁴, Yoko Hayashi-Takanaka^{2,12}, Hiromi
9 Wada^{3,13}, Shinzo Ohta¹³, Masanori Aikawa¹⁴, Hiroyuki Nakajima⁷, Masaki Nakamura⁶,
10 Rebecca C. McGee¹⁵, Kyle W. Heppner¹⁵, Tatsuo Kawakatsu¹⁶, Michiru Genno¹⁶,
11 Hiroshi Yanase¹⁶, Haruki Kume⁶, Takaaki Senbonmatsu¹⁷, Yukio Homma⁶, Shigeyuki
12 Nishimura¹⁷, Toutai Mitsuyama^{2,9}, Hiroyuki Aburatani^{2,4}, Hiroshi Kimura^{2,12}, Katsuhiko
13 Shirahige^{2,5}

14
15 ¹Laboratory of Computational Genomics, Institute for Quantitative Biosciences, The
16 University of Tokyo, Tokyo, 113-0032, Japan

17 ²Japan Agency for Medical Research and Development (AMED-CREST), AMED, 1-7-1
18 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

19 ³Isotope Science Center, The University of Tokyo, Tokyo, 113-0032, Japan

20 ⁴Genome Science Division, Research Center for Advanced Science and Technology,
21 The University of Tokyo, Tokyo, 153-8904, Japan

22 ⁵Laboratory of Genome Structure and Function, Institute for Quantitative Biosciences,
23 The University of Tokyo, Tokyo, 113-0032, Japan

24 ⁶Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo
25 113-8655, Japan

26 ⁷Department of Cardiovascular Surgery, Saitama Medical University International
27 Medical Center, Saitama 350-1298, Japan

28 ⁸Department of Clinical Informatics, Jichi Medical University School of Medicine,
29 Shimotsuke, 329-0498, Japan

- 30 ⁹Artificial Intelligence Research Center, National Institute of Advanced Industrial
31 Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
32 ¹⁰Computational Bio Big-Data Open Innovation Laboratory (CBBD-OIL), National
33 Institute of Advanced Industrial Science and Technology (AIST), 3-4-1 Okubo,
34 Shinjuku-ku, Tokyo 169-8555, Japan
35 ¹¹Graduate School of Information Science and Technology, Osaka University, Suita,
36 565-0871, Japan
37 ¹²Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology,
38 Yokohama, 226-8503, Japan
39 ¹³Brain Attack Center, Ohta Memorial Hospital, Fukuyama, 720-0825, Japan
40 ¹⁴The Center for Excellence in Vascular Biology and the Center for Interdisciplinary
41 Cardiovascular Sciences, Cardiovascular Division and Channing Division of Network
42 Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115,
43 USA
44 ¹⁵Lifeline Cell Technology, Frederick, MD 21701, USA
45 ¹⁶Bio-Medical Department, KURABO INDUSTRIES LTD., Neyagawa, Osaka 572-0823,
46 Japan
47 ¹⁷Department of Cardiology, Saitama Medical University International Medical Center,
48 Saitama 350-1298, Japan

49
50 *These authors contributed equally to this work.

51

52 **Contact**

53 Corresponding authors: rnakato@iam.u-tokyo.ac.jp (R.N.); ywada-ky@umin.ac.jp
54 (Y.W.); hkimura@bio.titech.ac.jp (H.K.); kshirahi@iam.u-tokyo.ac.jp (K.S.)

55

56

57 **ABSTRACT**

58 **Background:** Endothelial cells (ECs) make up the innermost layer throughout the entire
59 vasculature. Their phenotypes and physiological functions are initially regulated by
60 developmental signals and extracellular stimuli. The underlying molecular mechanisms
61 responsible for the diverse phenotypes of ECs from different organs are not well
62 understood.

63 **Results:** To characterize the transcriptomic and epigenomic landscape in the vascular
64 system, we cataloged gene expression and active histone marks in nine types of human
65 ECs (generating 148 genome-wide datasets) and carried out a comprehensive analysis
66 with chromatin interaction data. We identified 3,765 EC-specific enhancers, some of
67 which were associated with disease-associated genetic variations. We also identified
68 various candidate marker genes for each EC type. Notably, reflecting the developmental
69 origins of ECs and their roles in angiogenesis, vasculogenesis and wound healing.

70 **Conclusions:** While the importance of several HOX genes for early vascular
71 development and adult angiogenesis in pathological conditions has been reported, a
72 systematic analysis of the regulation and roles of HOX genes in mature tissue cells has
73 been lacking. These datasets provide a valuable resource for understanding the
74 vascular system and associated diseases.

75

76 **Keywords:** Endothelial cells, Histone modifications, Epigenome database, ChIP-seq,
77 Large-scale analysis

78

79

80 INTRODUCTION

81 Endothelial cells (ECs), which make up the innermost blood vessel lining of the
82 body, express diverse phenotypes that affect their morphology, physiological function
83 and gene expression patterns in response to the extracellular environment, including
84 the oxygen concentration, blood pressure and physiological stress. In the kidney, for
85 example, the vascular bed plays a role in the filtration of blood; in the brain, however,
86 the vascular architecture protects the central nervous system from toxins and other
87 components of the blood (1). Endothelial heterogeneity is mainly dependent on both the
88 function of each organ and the developmental lineage of different EC populations, which
89 result in adaptation to the vascular microenvironment. It is widely recognized that
90 certain specific vessels are susceptible to pathological changes, which include those
91 related to atherosclerosis and inflammation (2). Atherosclerosis, which occurs in the
92 muscular and elastic arteries, is a progressive disease characterized by the
93 accumulation of macrophages, and this process is initiated by the expression of cell
94 adhesion molecules, such as P-selectin. In mice, the expression level of P-selectin is
95 higher in the lung and mesentery vesicles compared with the heart, brain, stomach and
96 muscle (3).

97 In clinical practice, the thoracic, radial and gastroepiploic arteries are used for
98 coronary bypass grafts because these arteries have no tendency toward
99 atherosclerosis and hence are therapeutically advantageous in patients with coronary
100 artery plaques (4). In addition, the long-lasting results from coronary bypass graft
101 surgery indicate that vessels transplanted to a new environment differ in their outcome
102 based on their origin as an artery or vein (5, 6). Although the elucidation of the
103 molecular mechanisms underlying EC heterogeneity is critically important for the
104 development of vascular bed-specific remedies, these mechanisms have remained
105 largely unknown because ECs do not display such heterogeneity when cultured *in vitro*.

106 Epigenetic variation is a prime candidate for controlling the heterogeneity among
107 various ECs. Increasing evidence supports the idea that certain site-specific
108 characteristics are epigenetically regulated and easily altered by changes in the human
109 extracellular microenvironment. Previous gene expression studies of many types of
110 human ECs in culture demonstrated that site-specific epigenetic modifications play an

important role in differential gene expression (7). Moreover, our recent reports elucidated that there are different histone modifications present in the same genomic loci, such as *GATA6*, in human umbilical vein endothelial cells (HUVECs) and human dermal microvascular endothelial cells (HMVECs) (8, 9). Despite the discovery of these important insights, we still lack a systematic understanding of how the epigenomic landscape contributes to EC phenotype and heterogeneity. Therefore, there is a great demand for a comprehensive epigenomic catalog of the various EC types.

As a part of the International Human Epigenome Consortium (IHEC) project (10), we collected Chromatin immunoprecipitation followed by sequencing (ChIP-seq) data for the active histone modifications trimethylated H3 at Lys4 (H3K4me3) and acetylated H3 at Lys27 (H3K27ac) in EC DNA from nine different vascular cell types from multiple donors. We implemented large-scale comparative ChIP-seq analysis of these datasets to understand how the diverse phenotypes of ECs are regulated by key genes. All datasets used in this study are publicly available and are summarized on our website (<https://rnakato.github.io/HumanEndothelialEpigenome/>).

RESULTS

Reference epigenome generation across EC types

To establish an epigenetic catalog for different EC types, we generated a total of 491 genome-wide datasets, consisting of 424 histone modification ChIP-seq and 67 paired-end RNA sequencing (RNA-seq) datasets, encompassing a total of 22.3 billion sequenced reads. ECs were maintained as primary cultures with a physiological concentration of vascular endothelial growth factor (VEGF) and a minimal number of passages (fewer than six). We generated genome-wide normalized coverage tracks and peaks for ChIP-seq data and estimated normalized gene expression values for RNA-seq data.

In this study, we selected a subset of 33 EC samples (131 datasets) as a representative set comprising nine types of vessels from the human body (Figure 1A):

- Human aortic endothelial cells (HAoECs),
- Human coronary artery endothelial cells (HCoAECs),
- Human endocardial cells (HENDCs),

- 142 ● Human pulmonary artery endothelial cells (HPAECs),
- 143 ● Human umbilical vein endothelial cells (HUVECs),
- 144 ● Human umbilical artery endothelial cells (HUAECs),
- 145 ● Human common carotid artery endothelial cells (HCCaECs),
- 146 ● Human renal artery endothelial cells (HRAECs),
- 147 ● Human great saphenous vein endothelial cells (HGSVECs).

148 The detail of 33 EC samples is summarized in Supplementary Table S1.

149 Among the structures lined by these nine EC types, a group of two aortic, six
 150 common carotid and three coronary arteries is known as the “systemic arteries” and
 151 harbors arterial blood with 100 mmHg of oxygen tension and blood pressure in a range
 152 from 140 mmHg to 60 mmHg. Data sets for each cell type comprise samples from
 153 multiple donors, all of which achieved high-quality values as evaluated below. Here we
 154 focused on two histone modifications, H3K4me3 and H3K27ac (Figure 1B), which are
 155 the key markers of active promoters and enhancers (11). Because both H3K4me3 and
 156 H3K27ac exhibit strong, sharp peaks with ChIP-seq analysis, they are more suitable for
 157 identifying shared and/or unique features across EC cell types as compared with other
 158 histone modifications that show broad peaks, such as H3K9me3.

159

160 **Quality validation**

161 To evaluate the quality of obtained ChIP-seq data, we computed a variety of
 162 quality control measures (Supplementary Table S2), including the number of uniquely
 163 mapped reads, library complexity (the fraction of nonredundant reads), GC-content of
 164 mapped reads, genome coverage (the fraction of overlapped genomic areas with at
 165 least one mapped read), the number of peaks, signal-to-noise ratio (S/N) by the
 166 normalized strand coefficient (12), read-distribution bias measured by background
 167 uniformity (12), inter-sample correlation for each EC type and genome-wide correlation
 168 of read density across all-by-all pairs (Supplementary Figure S1). In addition, the peak
 169 distribution around several known positive/negative marker genes was visually
 170 inspected. Low-quality datasets were not used for further analyses.

171 To further validate the reliability of our data, we evaluated the consistency
 172 between the obtained peaks from ChIP-seq and the gene expression values from
 173 corresponding RNA-seq data. We applied a bivariate regression model (13) to estimate

the expression level of all genes based on H3K4me3 and H3K27ac peaks and then calculated the Pearson correlation between the estimated and the observed expression levels from ChIP-seq and RNA-seq, respectively. We used data derived from IMR90 fibroblasts analyzed with the same antibodies as a negative control, and we confirmed that peak distribution of the ChIP-seq data was highly correlated with corresponding RNA-seq data for ECs, but not with IMR90 data (Figure 1C, Supplementary Figure S2 for the full matrix). Therefore, our ChIP-seq data are likely to represent the histone modification states of ECs for annotation.

182

183 **Identification of active promoter and enhancer sites**

We used H3K4me3 and H3K27ac ChIP-seq peaks to define “active promoter (H3K4me3 and H3K27ac)” and “enhancer (H3K27ac only)” sites for each sample (Figure 1B, left). Then we assembled them and defined the common sites among all samples of a given EC type as the reference sites, to avoid differences among individuals. Finally, the reference sites of all nine EC types were merged into a single reference set for ECs (Figure 1B, right). We identified 9,121 active promoter sites (peak width, 2840.8 bp on average) and 23,202 enhancer sites (peak width, 1799.4 bp on average). The averaged peak width became relatively wide due to the merging of multiple contiguous sites.

We compared the distribution of the reference sites with gene annotation information. As expected, active promoter sites were enriched in the transcription start sites (TSSs) of genes, whereas enhancer sites were more frequently dispersed in introns and intergenic regions (Supplementary Figure S3). Among the enhancers, 15,625 (67.3%) were distally located (more than 10 kbp away from the nearest TSSs). The number of enhancer sites was more varied among the nine tissue types, whereas the number of active promoter sites was comparable across the EC types (Figure 2A, upper panel). The large number of HUAEC enhancer sites is possibly due to the small number of samples (two) and a relatively small individual difference (both samples were from newborns). We also evaluated the shared ratio of promoter and enhancer sites across all EC types (Figure 2A, lower panel). We found that nearly 80% of the active promoter sites were shared among multiple EC types. In contrast, 57.7% of the

enhancers were specific to up to two EC types, suggesting that their more diverse distribution across EC types relative to active promoter sites contributes to the EC type-specific regulatory activity. These observations are consistent with previous studies for other cell lines (11, 14).

Evaluation of enhancer sites by PCA

To investigate the diverse distribution of our reference enhancer sites, we used the principal component analysis (PCA) based on the H3K27ac read densities in the integrated EC enhancer sites with the 117 cell lines from the Roadmap Epigenomics Project (14). We found that ECs were well clustered and separated from other cell lines (Figure 2B). Remarkably, HUVECs represented in the Roadmap Epigenomics Project dataset, termed E122, were properly included in the EC cluster (red circle). In contrast, IMR90 cells from our study were included in the non-EC cluster (blue circle). This result supported the reliability of our EC-specific enhancer profiling. It should be noted, however, that the samples for each EC cell type (indicated by different colors) were not well clustered, possibly because the EC type-specific difference is minuscule and is disrupted by differences at the level of the individual.

Identification of enhancer-promoter interactions by ChIA-PET

We sought to identify the corresponding gene for the reference enhancer sites and used chromatin loop data obtained from the Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) data using RNA Polymerase II (Pol II) in HUVECs. We identified 292 significant chromatin loops (false discovery rate (FDR) < 0.05), 49.3% (144 loops) of which connected promoter and enhancer sites. Even when we used all chromatin loops (at least one read pair), 27.4% (8,782 of 31,997) of them linked to enhancer-promoter sites. Remarkably, 48.1% (4,228 of 8,782) of loops connected the distal enhancer sites. In total, we identified 2,686 distal enhancer sites that are connected by chromatin loops. We also detected enhancer-enhancer (3,136, 9.8%) and promoter-promoter (11,618, 36.3%) loops, suggesting physically aggregated chromatin hubs in which multiple promoters and enhancers interact (15). As the ChIA-PET data are derived from RNA Pol II-associated loops in HUVECs, chromatin interactions in active genes could be detected.

237

238 Identification of EC-specific sites

239 Next, we identified EC-specific enhancer sites by excluding any sites from our
240 reference sites that overlapped with those of our IMR90 cells and other cell types from
241 the Roadmap Epigenomics Project, except HUVECs (E122). As a result, we obtained
242 3,765 EC-specific enhancer sites (Supplementary Table S3), some of which were
243 located around known marker genes of ECs with chromatin loops. One example is
244 kinase insert domain receptor (*KDR*; Figure 2C, left), which functions as the VEGF
245 receptor, causing endothelial proliferation, survival, migration, tubular morphogenesis
246 and sprouting (16). The TSS of *KDR* was marked as an active promoter (enriched for
247 both H3K4me3 and H3K27ac) and physically interacted with the EC-specific enhancer
248 sites indicated by H3K27ac, ~50 kbp upstream and downstream of the TSS. Another
249 example is intercellular adhesion molecule 2 (*ICAM2*, Figure 2C, right), which is an
250 endothelial marker and is involved in the binding to white blood cells that occurs during
251 the antigen-specific immune response (17). This gene has two known TSSs, both of
252 which were annotated as active promoters in ECs, and one TSS that was EC specific
253 (black arrow). This EC-specific TSS did not have a ChIA-PET interaction, and, likewise,
254 the enhancer sites within the entire gene body did not directly interact with the adjacent
255 promoter sites, implying the distinctive regulation of the two *ICAM2* promoters.

256

257 Genome-wide association study (GWAS) enrichment analysis

258 To explore the correlation of EC-specific reference enhancer sites with
259 sequence variants associated with disease phenotypes, we obtained reference GWAS
260 single-nucleotide polymorphisms (SNPs) from the GWAS catalog (18) and identified
261 significantly enriched loci by permutation analysis (19). Notably, we identified 67
262 enhancer sites that markedly overlapped with GWAS SNPs associated with “heart”,
263 “coronary” and “cardiac” (Z score > 5.0, Supplementary Table S4). The most notable
264 region was around *CALCRL* and *TFPI* loci (chr2:188146468-188248446, Figure 2D).
265 The EC-specific enhancer region in these loci contained four GWAS risk variants
266 (Figure 2D, red triangles), three of which are associated with coronary artery/heart
267 disease (20, 21). Another example is the *RSPO3* locus (Supplementary Figure S4). The

268 upstream distal enhancer regions of that gene contained four GWAS SNPs that are
269 associated with cardiovascular disease and blood pressure (22, 23).

270

271 **Functional analysis of the reference sites**

272 We next investigated whether any characteristic sequence feature is observed in the
273 EC-specific enhancer and promoter sites (Supplementary Figure S5). We found a
274 putative motif for EC-specific enhancer sites, which is closely similar to the canonical
275 motifs of the homeobox genes *bcd*, *oc*, *Gsc* and *PITX1,2,3* (Figure 3). In fact, most of
276 the EC-specific enhancer sites consisted of enhancers of HGSVECs (47.0%), HRAECs
277 (37.4%) and HUAECs (68.3%) (Supplementary Figure S6), suggesting that these
278 enhancers contain this motif.

279 We also looked into the Gene Ontology (GO) classifications under Biological
280 Process for the enhancer sites using GREAT (24) and found that the enhancer sites
281 (both all sites and EC-specific sites) have GO terms that are more specific to the
282 vascular system (e.g., platelet activation, myeloid leukocyte activation and
283 vasculogenesis), as compared with active promoter sites (e.g., mRNA catabolic process,
284 Supplementary Figure S7). This also suggests that the enhancer sites are more likely to
285 be associated with EC-specific functions, whereas promoter sites are also correlated
286 with the more common biological functions.

287

288 **Differential analysis and clustering across EC types**

289 One important issue of this study is to clarify the epigenomic/transcriptomic
290 diversity across EC cell types. To circumvent variances at the level of the individual in
291 each cell type observed (Figure 2C) and different S/N ratios, we fitted the value of peak
292 intensity on the reference enhancer sites among samples using generalized linear
293 models with the quantile normalization. By implementing a PCA, we confirmed that
294 different cell samples in the same EC type were properly clustered (Figure 4A). The
295 PCA also showed that different EC types can be divided into two subgroups based on
296 the epigenomic landscape, corresponding to upper body (HAoEC, HCoAEC, HPAEC,
297 HCCaEC and HENDC, purple circle) and lower body (HUVEC, HUAEC, HGSVEC and
298 HRAEC) origins. A PCA based on gene expression data showed similar results to that
299 based on the H3K27ac profile, although in the gene expression analysis HUAECs were

300 more similar to heart ECs (Supplementary Figure S8).

301 To further investigate this tendency, we implemented a multiple-group
302 differential analysis with respect to H3K4me3, H3K27ac and gene expression data to
303 obtain sites and genes whose values were significantly varied between any of the nine
304 cell types. With the threshold $FDR < 1e-5$, we identified 753 differential H3K4me3 sites
305 (differential promoters, DPs; 8.3% from 9,121 active promoter sites), 2,979 differential
306 H3K27ac sites (differential enhancers, DEs; 9.2% from 32,323 active promoter and
307 enhancer sites) and 879 differentially expressed genes (DEGs; 2.1% from 41,880
308 genes). As expected, DPs and DEs were more enriched around DEGs, as compared
309 with all genes. DPs were enriched within ~10 kbp from TSSs, whereas DEs were more
310 broadly distributed (~100 kbp) (Supplementary Figure S9), indicative of the
311 longer-range interactions between enhancers and their corresponding genes.

312 We then implemented k-means clustering ($k = 6$) to characterize the overall
313 variability of DEGs, DEs (Figure 4B) and DPs (Supplementary Figure S10). Although k
314 $= 6$ was empirically defined and might not be biologically optimal to classify the nine EC
315 types, the results could capture the differential patterns. The upregulated genes were
316 roughly categorized into upper and lower body-specific EC types (Figure 4B), even
317 though diverse expression patterns were observed overall. In particular, the expression
318 patterns of the EC types around the heart (HCoAEC, HAoEC and HPAEC) were similar
319 (cluster 3 of DP and DEG), consistent with the anatomical proximity of these ECs.
320 HENDCs had uniquely expressed genes (cluster 6 of DEGs). Considering that most of
321 the DEGs and Des are cooperatively enriched in more than one EC type, these nine EC
322 types may use distinct combinations of multiple genes, rather than exclusively
323 expressed individual genes, for their specific phenotype.

324

325 **DEGs that contribute to EC functions**

326 Our clustering analysis also identified important genes for EC functions as
327 DEGs (Figure 4B). For example, heart and neural crest derivatives expressed 2
328 (*HAND2*) and GATA binding protein 4 (*GATA4*) were expressed in HAoECs, HENDCs
329 and HPAECs (cluster 3). *HAND2* physically interacts with *GATA4* and the histone
330 acetyltransferase p300 to form the enhanceosome complex, which regulates
331 tissue-specific gene expression in the heart (25). Another example is the related family

bHLH transcription factor with YRPW motif 2 (*HEY2*, also called *Hrt2*), a positive marker for arterial EC specification (26), which was grouped to cluster 5 and was expressed specifically in aorta-derived ECs but not in vein-derived ECs (HUVECs and HGSVECs). HRAECs showed uniquely upregulated genes, including cadherin 4 (*CDH4*); the protein product of this gene mediates cell-cell adhesion, and mutation of this gene is significantly associated with chronic kidney disease in the Japanese population (27). Interestingly, at TSSs of *CDH4* and *HEY2* loci, H3K4me3 was also enriched in some EC types in which the genes were not expressed, whereas the H3K27ac enrichment pattern at TSSs was correlated with the expression level (Figure 4C). This variation in H3K4me3 with/without H3K27ac enrichment may reflect the competence of expression, which cannot be fully captured by gene expression analysis.

DEGs also contained several notable gene families. One example is the claudin family, a group of transmembrane proteins involved in barrier and pore formation (28). Whereas *CLDN5* has been reported as a major constituent of the brain EC tight junctions that make up the blood-brain barrier (29), we found that seven other genes belonging to the claudin family (*CLDN1*, 7, 10, 11, 12, 14 and 15) were expressed in ECs, and their expression pattern varied across EC types (Supplementary Figure S11). For example, in HUVECs, *CLDN11* was highly expressed but *CLDN14* was not, although the two claudins share a similar function for cation permeability (30). These observations suggest that distinct usages of specific claudin proteins may result in different phenotypes with respect to vascular barrier function. Consequently, these DEGs are thus usable as a reference marker set for each EC type.

354

355 **Homeobox genes are highly differentially expressed across EC types**

We also found that DEGs identified in our analysis contained genes that were not previously acknowledged as relevant to the different EC types. Most strikingly, quite a few homeobox (*HOX*) genes were differentially expressed (cluster 1 in Figure 4B and Figure 5A). The human genome has four *HOX* clusters (*HOXA*, *B*, *C* and *D*), each of which contains 9–11 genes essential for determining the body axes during embryonic development, as well as regulating cell proliferation and migration in diverse organisms (31). These genes are transcribed sequentially in both time and space, following their

positions within each cluster (32). Figure 5A shows that genes in *HOX* clusters *A*, *B* and *D* were highly expressed in all EC types, except HENDCs, possibly because HENDCs are derived from cardiac neural crest, whereas the other EC types are derived from mesoderm (33). *HOXC* genes were moderately expressed in HRAECs, HUVECs and HUAECs, but not in the upper-body ECs. More interestingly, perhaps, *HOXD* genes were not expressed in HPAECs, despite their similar expression pattern relative to other EC types around the heart (Figure 4B). This result implies the distinct use of *HOX* paralogs, especially *HOXD* genes, in ECs. We also found that the 5' *HOX* genes (blue bars in Figure 5A) tended to be selectively expressed in EC types derived from the lower body (HGSVECs, HRAECs, HUAECs and HUVECs). Considering the collinearity of their activation during axial morphogenesis, it is conceivable that the type-specific expression of *HOX* clusters, especially in 5' *HOX* genes, reflects the developmental origin of EC types and that distinct activation of *HOX* genes collectively maintains the diversity of the circulatory system.

It has been suggested that the more 3' *HOX* genes tend to promote the angiogenic phenotype in ECs, whereas the more 5' *HOX* genes tend to be inhibitory with respect to that phenotype (31). For example, *HOXD3* may promote wound healing and invasive or migratory behavior during angiogenesis in ECs (34). In contrast, *HOXD10* may function to inhibit EC migration by muting the downstream effects of other pro-angiogenic *HOX* genes (e.g., *HOX3* paralogs), and thus human ECs that overexpress *HOXD10* fail to form new blood vessels (35). Figure 5A shows that *HOXD10* was highly expressed in HGSVECs, which evokes the inhibition of the angiogenic phenotype as regulated by *HOXD10* in this cell type.

In addition to *HOX* genes, multiple non-*HOX* homeobox genes were also differentially regulated across ECs. For example, cluster 3 in the DEGs contained NK2 homeobox 5 (*Nkx-2.5*), which is essential for maintenance of ventricular identity (36); paired like homeodomain 2 (*PITX2*) and paired related homeobox 1 (*PRRX1*), which are both associated with the atrial fibrillation and cardioembolic ischemic stroke variants loci (37-39); and Meis homeobox 1 (*MEIS1*), which is required for heart development in mice (40). These were all associated with the GO term "blood vessel morphogenesis (GO:0048514)". Interestingly, *PITX3* was mainly expressed in HENDCs (cluster 6),

394 unlike *PITX2*. Another example is Mesenchyme Homeobox 2 (*MEOX2*, also known as
395 Gax; cluster 4), which regulates senescence and proliferation in HUVECs (41) and was
396 also expressed in HUAECs and HGSVECs but not in other cell types. Taken together
397 with the finding that some binding motifs of homeobox genes including *PITX* were
398 identified among the EC-specific enhancer sites (Figure 3), these data suggest that
399 distinct combinations of proteins coded by *HOX* and non-*HOX* homeobox genes play a
400 key role in mature human ECs for angiogenesis, vasculogenesis and wound healing, in
401 addition to their function during the development and proliferation of ECs.

402

403 **Enhancers in the telomeric domain were upregulated within the *HOXD* cluster**

404 Lastly, we investigated the epigenomic landscape around the *HOXD* cluster
405 (Figure 5B). It has been reported that the mammalian *HOXD* cluster is located between
406 two enhancer-rich topologically associating domains (TADs), the centromeric domain
407 (C-DOM) and the telomeric domain (T-DOM), which are activated during limb and digit
408 development, respectively (42). By using public Hi-C (genome-wide Chromosome
409 Conformation Capture (3C)) data for HUVECs (43) to detect the T-DOM and C-DOM
410 (bottom black bars), we observed the presence of EC enhancers in the T-DOM (Figure
411 5B), as in early stages of limb development (42). Of note, two long non-coding RNAs,
412 *LINC01117* (*Hotdog*) and *LINC01116* (*Twin of Hotdog*), which physically contact the
413 expressed *HOXD* genes and are activated during cecum budding (44), had ChIA-PET
414 loops and showed similar expression patterns with *HOXD* genes in ECs (Figure 5A and
415 5B). In the T-DOM, some enhancers are likely to be activated in most EC types (Figure
416 5B, black arrow), whereas others are active only in ECs from the lower body (blue
417 arrow), suggesting a physical interaction between these enhancers and each *HOXD*
418 gene in a constitutive and a cell type-specific manner, respectively. A detailed view of
419 the genomic region from *HOXD1* to *HOXD11* (Figure 5C) shows that H3K4me3 and
420 H3K27ac are specifically enriched within the *HOXD10* locus in HGSVECs, which is
421 consistent with their gene expression pattern. Because the *HOXD10* locus did not have
422 ChIA-PET loops in HUVECs, there might be HGSVEC-specific chromatin loops.

423

424

425 DISCUSSION

426 In this study, we analyzed the epigenomic status of the active histone
427 modifications H3K4me3 and H3K27ac in 33 samples from nine different EC types by
428 ChIP-seq, RNA-seq, ChIA-PET and Hi-C analyses. The integrative ChIP-seq analysis
429 based on the samples from the human tissues of multiple donors is hampered by both
430 individual variations and technical noise derived from tissue sample acquisition under
431 various conditions (e.g., race, sex, age and the sample acquirement process),
432 compared with the smaller difference among EC types. To overcome these issues, at
433 least in part, we developed a robust procedure for comparative epigenome analysis,
434 combined with chromatin interaction data. We successfully identified 3,765 EC-specific
435 enhancer sites, 67 of which were highly significantly overlapping with GWAS SNPs. We
436 aim to expand this analysis to other core histone marks including suppressive markers
437 (e.g., H3K27me3 and H3K9me3) and apply semi-automated genome annotation
438 methods (45). Because this type of genome annotation strategy with its associated
439 assembling of broad marks is more sensitive to noise, more stringent quality control of
440 tissue data will be required.

441 The PCA showed that EC types can be divided into those from the upper and
442 from the lower body (Figure 4A). The nine EC types tend to use distinct combinations of
443 multiple genes, rather than exclusively expressed genes, for their specific phenotype
444 (Figure 4B and Figure 5A). Our results identified key marker genes that were
445 differentially expressed across EC types, such as homeobox genes. The importance of
446 several *HOX* genes for early vascular development and adult angiogenesis in
447 pathological conditions has been reported (46). However, a systematic analysis of the
448 regulation and roles of homeobox genes in mature tissue cells has been lacking. We
449 identified a regulatory motif enriched in EC-specific enhancers that is very similar to
450 those of *homeobox* genes and distinct epigenome states and chromatin conformations
451 of *HOX* gene clusters and flanking regions in different EC types. Taken together, our
452 data suggest the distinct roles and combinatorial usage of *HOX* genes during
453 development and in regulating EC phenotypes throughout the body.

454

455

456 **CONCLUSION**

457 The primary goal of the IHEC project is to generate high-quality reference
458 epigenomes and make them available to the scientific community (10). To this end, we
459 established an epigenetic catalog of various human ECs and implemented
460 comprehensive analysis to elucidate the diversity of the epigenomic and transcriptomic
461 landscape across EC types. The dataset presented in this study will be an important
462 resource for future work on understanding the human cardiovascular system and its
463 associated diseases.

465 **METHODS**

466 **Tissue preparation**

467 ECs were isolated from the vasculature and maintained as primary cultures, as
468 reported (47, 48). Briefly, HAOECs, HCoAECs, HENDCs, HPAECs and HUVECs were
469 isolated from the various vessels by incubating the vessels with collagenase at 37°C for
470 30 minutes. The aortic root was used for HAOEC isolation. Cells were plated in tissue
471 culture-treated flasks (Iwaki Glass Co. Ltd., cat. No.) and cultured for one or more
472 passages in modified VasculLife VEGF Endothelial Medium (Lifeline Cell Technology). A
473 reduced concentration of VEGF that was lower than 5 ng/mL was tested in preliminary
474 cell culture experiments and then optimized to be as low as possible considering cell
475 growth and viability (data not shown). The VEGF concentration was lowered to 250
476 pg/mL, which is lower than the standard culture conditions, to more closely replicate in
477 vivo concentrations (49). ECs were separated from non-ECs using immunomagnetic
478 beads. Fibroblasts were first removed using anti-fibroblast beads and the appropriate
479 magnetic column (Miltenyi Biotec). The remaining cells were then purified using
480 Dynabeads and anti-CD31 (BAM3567, R&D Systems). When positive selection was
481 used, the bead-bound cells were removed from the cell suspension prior to
482 cryopreservation.

483 HCCaECs and HRECs were prepared by an explant culture method (48).
484 HGSVECs were isolated from discarded veins taken from patients at Saitama Medical
485 University International Medical Center.

486 Quality control was performed using a sterility test (for bacteria, yeast and

fungi), a PCR-based sterility test (for hepatitis B and C, HIV-I and -II and mycoplasma) and immunostaining-based characterization for von Willebrand factor (vWF) (>95% cells are positively stained (50)) and alpha-actin, and viability was determined by both counting and trypan blue staining.

Cell culture

Purified ECs were cultured in VasculLife VEGF Endothelial Medium (Lifeline Cell Technology) with 250 pg/mL VEGF. Cells were maintained at 37°C in a humidified 5% CO₂ incubator, and the medium was changed every 3 days. The cells used in the experiments were from passage 6 or less. The cryopreservation solution used consisted of VasculLife VEGF Endothelial medium, containing 250 pg/mL VEGF, 12% Fetal Bovine Serum (FBS) and 10% Dimethylsulfoxide (DMSO).

RNA-seq analysis

Poly(A)-containing mRNA molecules were isolated from total RNA and then converted to cDNA with oligo(dT) primers using a TruSeq RNA Sample Preparation kit v2 (Illumina) and were sequenced with a HiSeq 2500 system (Illumina). We applied sequenced paired-end reads to kallisto version 0.43.1 (51) with the "--rf-stranded -b 100" option, which estimates the transcript-level expression values as Transcripts Per Kilobase Million (TPM, Ensembl gene annotation GRCh37). These transcript-level expression values were then assembled to the gene-level by tximport (52). We also obtained RNA-seq data from IMR90 cells from the Sequence Read Archive (SRA) (www.ncbi.nlm.nih.gov/sra) under accession number SRR2952390. The full list of gene expression data is available at the NCBI Gene Expression Omnibus (GEO) under the accession number GSE131953.

ChIP

For each EC sample, two million ECs were plated on a 15-cm culture plate and cultured until confluency. The cells were crosslinked for 10 minutes using 1% paraformaldehyde. After quenching using 0.2 M glycine, cells were collected using a scraper, resuspended in SDS lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 1% SDS, 1

mM EDTA; pH 8.0) and fragmented by sonication (Branson, Danbury, CT, USA; 10 minutes). Samples were stored at -80°C before use. To perform ChIP, antibodies against histone modifications (CMA304 and CMA309 for H3K4me3 and H3K27ac, respectively) (53) were used in combination with protein G Sepharose beads (GE Healthcare Bio-Sciences AB, Sweden). The prepared DNA was quantified using Qubit (Life Technologies/Thermo Fisher Scientific), and >10 ng of DNA was processed, as described below. The primer sequences for ChIP-qPCR were as follows: for H3K4me3, KDR (Fw: CCACAGACTCGCTGGGTAAT, Rv: GAGCTGGAGAGTTGGACAGG) and GAPDH (Fw: CGCTCACTGTTCTCTCCCTC, Rv: GACTCCGACCTTCACCTT CC); for H3K27ac, ANGPTL4 (Fw: TAGGGGAATGGGTAGGGAAG, Rv: AGTTCTCAGGCAGGTGGAGA) and GATA2 (Fw: AGACGACCCCAACTGACATC, Rv: CCTTCAAATGCAGACGCTTT) and, as a negative control, HBB (Fw: GGGCTGAGGGTTTGAAGTCC, Rv: CATGGTGTCTGTTTGAGGTTGC).

ChIP-seq analysis

Sequencing libraries were made using the NEBNext ChIP-Seq Library Prep Master Mix Set of Illumina (New England Biolabs). Sequenced reads were mapped to the human genome using Bowtie version 1.2 (54) allowing two mismatches in the first 28 bases per read and outputting only uniquely mapped reads (-n2 -m1 option). Peaks were called by DROMPA version 3.5.1 (55) using the stringent parameter set (-sm 200 -pthre_internal 0.00001 -pthre_enrich 0.00001) to mitigate the effect of technical noise. The mapping and peak statistics are summarized in Supplementary Table S2.

Quality validation of ChIP-seq samples

We checked the quality of each sample based on the peak number, library complexity and GC content bias by DROMPA; the normalized strand coefficient and background uniformity by SSP (12); inter-sample correlation (jaccard index of peak overlap) by bedtools (<https://github.com/arq5x/bedtools2>); and the pairwise correlations of read coverage by deepTools version 2.5.0 (56) (Supplementary Figure S1).

549 **Regression analysis of ChIP-seq data**

550 To estimate the expression level of a gene from the level of its histone
551 modifications, we implemented the linear regression analysis proposed by Karlic *et al.*
552 (13) with minor modifications. We built a two-variable model to predict the expression
553 level for each mRNA as follows:

$$f(x_1, x_2) = a + b_1x_1 + b_2x_2$$

554 where x_1 and x_2 are the log-scale basepair coverage in a region of 4-kbp surrounding
555 the TSSs covered by obtained peaks of H3K4me3 and H3K27ac, respectively. We used
556 the level of protein-coding mRNA in autosomes as an estimation of the level of histone
557 modifications. We then learned the parameters a , b_1 and b_2 using all of the EC samples
558 and the IMR90 sample to minimize the differences between observed and expected
559 values. Using the learned parameter set, we predicted the expression value for each
560 mRNA and calculated the Pearson correlation between observed and expected values.

562 **Definition of reference promoter and enhancer sites**

563 As shown in Figure 1B, we defined active promoters and enhancers as
564 “H3K4me3 sites overlapping with H3K27ac sites by ≥ 1 bp” and “H3K27ac sites not
565 overlapping with H3K4me3 sites”, respectively, based on the annotation of the
566 Roadmap Epigenomics consortium (14). Peaks from sex chromosomes were excluded
567 to ignore sex-specific difference. To avoid the effect of individual differences, the
568 common sites among all samples were used as the reference sites for each cell type.
569 Then the reference sites of all cell types were merged into the reference sites of ECs.
570 Multiple sites that were within 100 bp of each other were merged to avoid multiple
571 counts of large individual sites. The generated reference promoter and enhancer sites
572 are available at the GEO under the accession number GSE131953.

574 **Identification of EC-specific sites**

575 We called peaks for H3K4me3 and H3K27ac for all 117 cell lines in the
576 Roadmap Epigenomics Project by DROMPA with the same parameter set and excluded
577 the sites in the reference promoter and enhancer sites of ECs that overlapped the
578 H3K4me3 peaks (promoter sites) or H3K27ac peaks (enhancer sites) of all cells except

for E122 (HUVECs) from the Roadmap Epigenomics Project. Similarly, we further excluded the sites that overlapped H3K27ac peaks of IMR90 cells generated by this study, to avoid the protocol-dependent false-positive peaks. The resulting sites were used as EC-specific sites. We also defined “distal enhancer sites” as those that are >10 kbp from the nearest TSS. These sites are summarized in Table S2.

GWAS enrichment analysis

We implemented GWAS enrichment analysis using a strategy similar to that of Lake *et al.* (19). We obtained reference SNPs from the GWAS Catalog [18]. We then calculated the occurrence probability of GWAS SNPs associated with the terms “heart”, “coronary” and “cardiac” in 100-kb regions centered on all EC-specific enhancer sites and investigated their statistical significance by random permutations. We extended each enhancer site to a 100-kb region to consider linkage disequilibrium with GWAS SNPs. We identified the enhancer sites with a Z-score > 5.0. We shuffled the enhancer sites randomly within each chromosome, ignoring the centromeric region, using bedtools shuffle command.

Differential analysis of multiple groups for histone modification and gene expression

We applied the ANOVA-like test in edgeR (57) based on the normalized read counts of H3K4me3 ChIP-seq data in active promoters and H3K27ac ChIP-seq data in active promoters and enhancers, as well as gene expression data, while fitting the values among samples to estimate dispersion using generalized linear models. For RNA-seq data, the count data were fitted using a generalized linear model, and the Z-score was calculated based on logged values. For ChIP-seq data, we also applied the quantile normalization to peak intensity in advance of the fitting because this model does not consider the different S/N ratios among samples (58). This normalization assumes that the S/N ratio for most of the common peaks should be the same among all samples in which the same antibody was used. Supplementary Figure S12 shows the distribution patterns of the H3K27ac read density normalized for quantile normalization for all ECs.

610 **Chromatin interaction analysis**

611 We used ChIA-PET data mediated by RNA Pol II for HUVECs (59). We
612 acquired fastq files from the GEO under accession number GSE41553, applied Mango
613 (60) with default parameter settings and identified the 943 significant interactions (1,886
614 sites, FDR < 0.05). For Hi-C analysis, we acquired .hic files for HUVECs from the GEO
615 under accession number GSE63525 and applied Juicer (43) to obtained the TAD
616 structure (Figure 5B).

617 **Motif analysis**

618 We used MEME-ChIP version 5.0.1 (61) with the parameter set “-meme-mod
619 zoops -meme-minw 6 -meme-maxw 14” with the motif data
620 “JASPAR2018_CORE_non-redundant.meme”.
621

622 **List of abbreviations**

623 EC: Endothelial cells; S/N: signal-to-noise ratio; PCA: principal component
624 analysis; Pol II: RNA Polymerase II; FDR: false discovery rate; GWAS: Genome-wide
625 association study; SNP: single-nucleotide polymorphism; GO: Gene Ontology; DP:
626 differential promoter; DE: differential enhancer; DEG: differentially expressed gene;
627 TAD: topologically associating domain; C-DOM: the centromeric domain; T-DOM: the
628 telomeric domain; vWF: von Willebrand factor; TPM: Transcripts Per Kilobase Million
629

630 **Declarations**

631 ***Ethics approval and consent to participate***

632 Human clinical specimens were prepared from discarded tissue during surgery
633 under consent of donors at Saitama Medical University International Medical Center
634 according to the Institutional Review Board (IRB) protocol 15-209, at Ohta Memorial
635 Hospital according to the IRB protocols 068 and 069 and at The University of Tokyo
636 according to the IRB protocol G3577. Primary ECs were isolated at The University of
637 Tokyo according to the IRB protocol 17-311. Other ECs were prepared from a
638 commercial biobank (Lifeline Cell Technology, Frederick, MD). DNA and RNA samples
639 were prepared at The University of Tokyo according to the IRB protocol 12-81.
640

641 ***Consent for publication***

642 As to purchased cells, not applicable. As to primary cultivated cells from human
643 tissue, written informed consent was obtained from the patients for publication of their
644 individual details and accompanying images in this manuscript. The consent form is
645 held by the authors and is available for review by the Editor-in-Chief.

647 ***Availability of data and materials***

648 The raw sequencing data and processed files are available at the GEO under
649 the accession numbers GSE131953 (ChIP-seq) and GSE131681 (RNA-seq) with links
650 to BioProject accession number PRJNA532996. The data summary, quality control
651 results, the full list of ChIA-PET loops and visualization figures are available on the EC
652 analysis website (<https://rnakato.github.io/HumanEndothelialEpigenome/>).

654 ***Competing interests***

655 The authors declare no competing interests.

657 ***Funding***

658 This work was supported by a grant from the Japan Agency for Medical
659 Research and Development (AMED-CREST, Grant ID JP16gm0510005h0006), the
660 Basis for Supporting Innovative Drug Discovery and Life Science Research from AMED
661 (to H.K. and K.S.) and a grant-in-aid for Scientific Research (17H06331 to R.N. and
662 JP18H05527 to H.K.).

664 ***Authors' contributions***

665 R. Nakato designed the studies, performed bioinformatics analysis and drafted
666 the manuscript. Y.W. designed the studies, carried out the molecular genetic studies and
667 drafted the manuscript. R. Nakaki, N.N., S.T., T. Kohro, N.O., Y.S. and T.M. performed
668 bioinformatics analyses. G.N. and K.T. performed mRNA-seq. Y. Katou performed library
669 preparation and sequencing. Y. Kanki performed cell culture, prepared samples and
670 drafted the manuscript. M.K. and A.K. performed cell culture and prepared samples.
671 Y.H-T. and A.I-T. prepared reagents and performed chromatin immunoprecipitation. H.F.,

672 A.I., H.N., M.N., T.S., S.N. H.W., S.O., M.A., R.C.M., K.W.H., T. Kawakatsu, M.G., H.Y., H.
673 Kume, and Y.H. prepared primary cell cultures. H.A. and K.S. designed the studies and
674 drafted the manuscript. H.Kimura prepared antibodies, designed the studies and drafted
675 the manuscript.

676

677 **Acknowledgements**

678 Advanced Medical Graphics (MA, USA) prepared the graphics of the organs
679 and tissues. Niinami Hiroshi continuously supported domestic sample collection and
680 preparation. We thank Ryoza Omoto for providing the pipeline from the surgical
681 operating room to the research laboratory.

682

683 **Competing Interests**

684 The authors declare no competing interests.

685

686 **References**

- 687 1. Aird WC. Phenotypic heterogeneity of the endothelium: II. Representative
688 vascular beds. Circulation research. 2007;100(2):174-90.
- 689 2. Thompson RC, Allam AH, Lombardi GP, Wann LS, Sutherland ML, Sutherland
690 JD, et al. Atherosclerosis across 4000 years of human history: the Horus study of four
691 ancient populations. Lancet. 2013;381(9873):1211-22.
- 692 3. Eppihimer MJ, Wolitzky B, Anderson DC, Labow MA, Granger DN.
693 Heterogeneity of expression of E- and P-selectins in vivo. Circulation research.
694 1996;79(3):560-9.
- 695 4. Kaufer E, Factor SM, Frame R, Brodman RF. Pathology of the radial and
696 internal thoracic arteries used as coronary artery bypass grafts. The Annals of thoracic
697 surgery. 1997;63(4):1118-22.
- 698 5. Aird WC. Phenotypic heterogeneity of the endothelium: I. Structure, function,
699 and mechanisms. Circulation research. 2007;100(2):158-73.
- 700 6. Sabik JF, 3rd, Raza S, Blackstone EH, Houghtaling PL, Lytle BW. Value of
701 internal thoracic artery grafting to the left anterior descending coronary artery at
702 coronary reoperation. Journal of the American College of Cardiology.
703 2013;61(3):302-10.
- 704 7. Chi JT, Chang HY, Haraldsen G, Jahnsen FL, Troyanskaya OG, Chang DS, et
705 al. Endothelial cell diversity revealed by global expression profiling. Proceedings of the

706 National Academy of Sciences of the United States of America. 2003;100(19):10623-8.
707 8. Kanki Y, Kohro T, Jiang S, Tsutsumi S, Mimura I, Suehiro J, et al. Epigenetically
708 coordinated GATA2 binding is necessary for endothelium-specific endomucin
709 expression. *The EMBO journal*. 2011;30(13):2582-95.
710 9. Tozawa H, Kanki Y, Suehiro J, Tsutsumi S, Kohro T, Wada Y, et al.
711 Genome-wide approaches reveal functional interleukin-4-inducible STAT6 binding to the
712 vascular cell adhesion molecule 1 promoter. *Mol Cell Biol*. 2011;31(11):2196-209.
713 10. Stunnenberg HG, International Human Epigenome C, Hirst M. The
714 International Human Epigenome Consortium: A Blueprint for Scientific Collaboration
715 and Discovery. *Cell*. 2016;167(5):1145-9.
716 11. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al.
717 Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*.
718 2011;473(7345):43-9.
719 12. Nakato R, Shirahige K. Sensitive and robust assessment of ChIP-seq read
720 distribution using a strand-shift profile. *Bioinformatics*. 2018;34(14):2356-63.
721 13. Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. Histone modification
722 levels are predictive for gene expression. *Proceedings of the National Academy of*
723 *Sciences of the United States of America*. 2010;107(7):2926-31.
724 14. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky
725 M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*.
726 2015;518(7539):317-30.
727 15. Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Verstegen M, Geeven
728 G, et al. Enhancer hubs and loop collisions identified from single-allele topologies.
729 *Nature genetics*. 2018;50(8):1151-60.
730 16. Waltenberger J, Claesson-Welsh L, Siegbahn A, Shibuya M, Heldin CH.
731 Different signal transduction properties of KDR and Flt1, two receptors for vascular
732 endothelial growth factor. *The Journal of biological chemistry*. 1994;269(43):26988-95.
733 17. Lyck R, Enzmann G. The physiological roles of ICAM-1 and ICAM-2 in
734 neutrophil migration into tissues. *Current opinion in hematology*. 2015;22(1):53-9.
735 18. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et
736 al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies,
737 targeted arrays and summary statistics 2019. *Nucleic Acids Res*.
738 2019;47(D1):D1005-D12.
739 19. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, et al. Integrative
740 single-cell analysis of transcriptional and epigenetic states in the human adult brain.
741 *Nature biotechnology*. 2018;36(1):70-80.

- 742 20. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an
743 Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circulation*
744 *research*. 2018;122(3):433-43.
- 745 21. Coronary Artery Disease Genetics C. A genome-wide association study in
746 Europeans and South Asians identifies five new loci for coronary artery disease. *Nature*
747 *genetics*. 2011;43(4):339-44.
- 748 22. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging
749 Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet*.
750 2019;104(1):65-75.
- 751 23. Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, et al.
752 The genetics of blood pressure regulation and its target organs from association studies
753 in 342,415 individuals. *Nat Genet*. 2016;48(10):1171-84.
- 754 24. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT
755 improves functional interpretation of cis-regulatory regions. *Nature biotechnology*.
756 2010;28(5):495-501.
- 757 25. Dai YS, Cserjesi P, Markham BE, Molkentin JD. The transcription factors
758 GATA4 and dHAND physically interact to synergistically activate cardiac gene
759 expression through a p300-dependent mechanism. *The Journal of biological chemistry*.
760 2002;277(27):24390-8.
- 761 26. Aranguren XL, Agirre X, Beerens M, Coppiello G, Uriz M, Vandersmissen I, et
762 al. Unraveling a novel transcription factor code determining the human arterial-specific
763 endothelial cell signature. *Blood*. 2013;122(24):3982-92.
- 764 27. Yoshida T, Kato K, Yokoi K, Oguri M, Watanabe S, Metoki N, et al. Association
765 of genetic variants with chronic kidney disease in Japanese individuals with or without
766 hypertension or diabetes mellitus. *Exp Ther Med*. 2010;1(1):137-45.
- 767 28. Morita K, Furuse M, Fujimoto K, Tsukita S. Claudin multigene family encoding
768 four-transmembrane domain protein components of tight junction strands. *Proc Natl*
769 *Acad Sci U S A*. 1999;96(2):511-6.
- 770 29. Morita K, Sasaki H, Furuse M, Tsukita S. Endothelial claudin: claudin-5/TMVCF
771 constitutes tight junction strands in endothelial cells. *The Journal of cell biology*.
772 1999;147(1):185-94.
- 773 30. Amasheh S, Fromm M, Gunzel D. Claudins of intestine and nephron - a
774 correlation of molecular tight junction structure and barrier function. *Acta physiologica*.
775 2011;201(1):133-40.
- 776 31. Gorski DH, Walsh K. Control of vascular cell differentiation by homeobox
777 transcription factors. *Trends in cardiovascular medicine*. 2003;13(6):213-20.

778 32. Kmita M, Duboule D. Organizing axes in time and space; 25 years of colinear
779 tinkering. *Science*. 2003;301(5631):331-3.

780 33. Srivastava D. Making or breaking the heart: from lineage determination to
781 morphogenesis. *Cell*. 2006;126(6):1037-48.

782 34. Uyeno LA, Newman-Keagle JA, Cheung I, Hunt TK, Young DM, Boudreau N.
783 Hox D3 expression in normal and impaired wound healing. *The Journal of surgical*
784 *research*. 2001;100(1):46-56.

785 35. Myers C, Charboneau A, Cheung I, Hanks D, Boudreau N. Sustained
786 expression of homeobox D10 inhibits angiogenesis. *The American journal of pathology*.
787 2002;161(6):2099-109.

788 36. Targoff KL, Colombo S, George V, Schell T, Kim SH, Solnica-Krezel L, et al.
789 Nkx genes are essential for maintenance of ventricular identity. *Development*.
790 2013;140(20):4203-13.

791 37. Christophersen IE, Rienstra M, Roselli C, Yin X, Geelhoed B, Barnard J, et al.
792 Large-scale analyses of common and rare variants identify 12 new loci associated with
793 atrial fibrillation. *Nature genetics*. 2017;49(6):946-52.

794 38. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson
795 A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature*.
796 2007;448(7151):353-7.

797 39. Neurology Working Group of the Cohorts for H, Aging Research in Genomic
798 Epidemiology Consortium tSGN, the International Stroke Genetics C. Identification of
799 additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide
800 association studies. *The Lancet Neurology*. 2016;15(7):695-707.

801 40. Wang H, Liu C, Liu X, Wang M, Wu D, Gao J, et al. MEIS1 Regulates
802 Hemogenic Endothelial Generation, Megakaryopoiesis, and Thrombopoiesis in Human
803 Pluripotent Stem Cells by Targeting TAL1 and FLI1. *Stem Cell Reports*.
804 2018;10(2):447-60.

805 41. Gohn CR, Blue EK, Sheehan BM, Varberg KM, Haneline LS. Mesenchyme
806 Homeobox 2 Enhances Migration of Endothelial Colony Forming Cells Exposed to
807 Intrauterine Diabetes Mellitus. *Journal of cellular physiology*. 2017;232(7):1885-92.

808 42. Andrey G, Montavon T, Mascres B, Gonzalez F, Noordermeer D, Leleu M, et al.
809 A switch between topological domains underlies HoxD genes collinearity in mouse limbs.
810 *Science*. 2013;340(6137):1234167.

811 43. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et
812 al. A 3D map of the human genome at kilobase resolution reveals principles of
813 chromatin looping. *Cell*. 2014;159(7):1665-80.

814 44. Delpretti S, Montavon T, Leleu M, Joye E, Tzika A, Milinkovitch M, et al.
815 Multiple enhancers regulate Hoxd genes and the Hotdog LncRNA during cecum
816 budding. *Cell reports*. 2013;5(1):137-50.

817 45. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS.
818 Unsupervised pattern discovery in human chromatin structure through genomic
819 segmentation. *Nature methods*. 2012;9(5):473-6.

820 46. Kachgal S, Mace KA, Boudreau NJ. The dual roles of homeobox genes in
821 vascularization and wound healing. *Cell adhesion & migration*. 2012;6(6):457-70.

822 47. Wada Y, Sugiyama A, Yamamoto T, Naito M, Noguchi N, Yokoyama S, et al.
823 Lipid accumulation in smooth muscle cells under LDL loading is independent of LDL
824 receptor pathway and enhanced by hypoxic conditions. *Arteriosclerosis, thrombosis,
825 and vascular biology*. 2002;22(10):1712-9.

826 48. Kobayashi M, Inoue K, Warabi E, Minami T, Kodama T. A simple method of
827 isolating mouse aortic endothelial cells. *Journal of atherosclerosis and thrombosis*.
828 2005;12(3):138-42.

829 49. Vermeulen PB, Salven P, Benoy I, Gasparini G, Dirix LY. Blood platelets and
830 serum VEGF in cancer patients. *Br J Cancer*. 1999;79(2):370-3.

831 50. Au-Yeung KK, Woo CW, Sung FL, Yip JC, Siow YL, O K.
832 Hyperhomocysteinemia activates nuclear factor-kappaB in endothelial cells via
833 oxidative stress. *Circ Res*. 2004;94(1):28-36.

834 51. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
835 quantification. *Nature biotechnology*. 2016;34(5):525-7.

836 52. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq:
837 transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4:1521.

838 53. Kimura H, Hayashi-Takanaka Y, Goto Y, Takizawa N, Nozaki N. The
839 organization of histone H3 modifications as revealed by a panel of specific monoclonal
840 antibodies. *Cell structure and function*. 2008;33(1):61-73.

841 54. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient
842 alignment of short DNA sequences to the human genome. *Genome biology*.
843 2009;10(3):R25.

844 55. Nakato R, Itoh T, Shirahige K. DROMPA: easy-to-handle peak calling and
845 visualization software for the computational analysis and validation of ChIP-seq data.
846 *Genes to cells : devoted to molecular & cellular mechanisms*. 2013;18(7):589-601.

847 56. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al.
848 deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic
849 acids research*. 2016;44(W1):W160-5.

850 57. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in
851 RNA sequencing data using observation weights. *Nucleic acids research*.
852 2014;42(11):e91.

853 58. Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality
854 management to whole-genome annotation. *Briefings in bioinformatics*.
855 2017;18(2):279-90.

856 59. Papantonis A, Kohro T, Baboo S, Larkin JD, Deng B, Short P, et al. TNFalpha
857 signals through specialized factories where responsive coding and miRNA genes are
858 transcribed. *The EMBO journal*. 2012;31(23):4404-14.

859 60. Phanstiel DH, Boyle AP, Heidari N, Snyder MP. Mango: a bias-correcting
860 ChIA-PET analysis pipeline. *Bioinformatics*. 2015;31(19):3092-8.

861 61. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets.
862 *Bioinformatics*. 2011;27(12):1696-7.

863 62. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al.
864 Metascope provides a biologist-oriented resource for the analysis of systems-level
865 datasets. *Nat Commun*. 2019;10(1):1523.

866

867

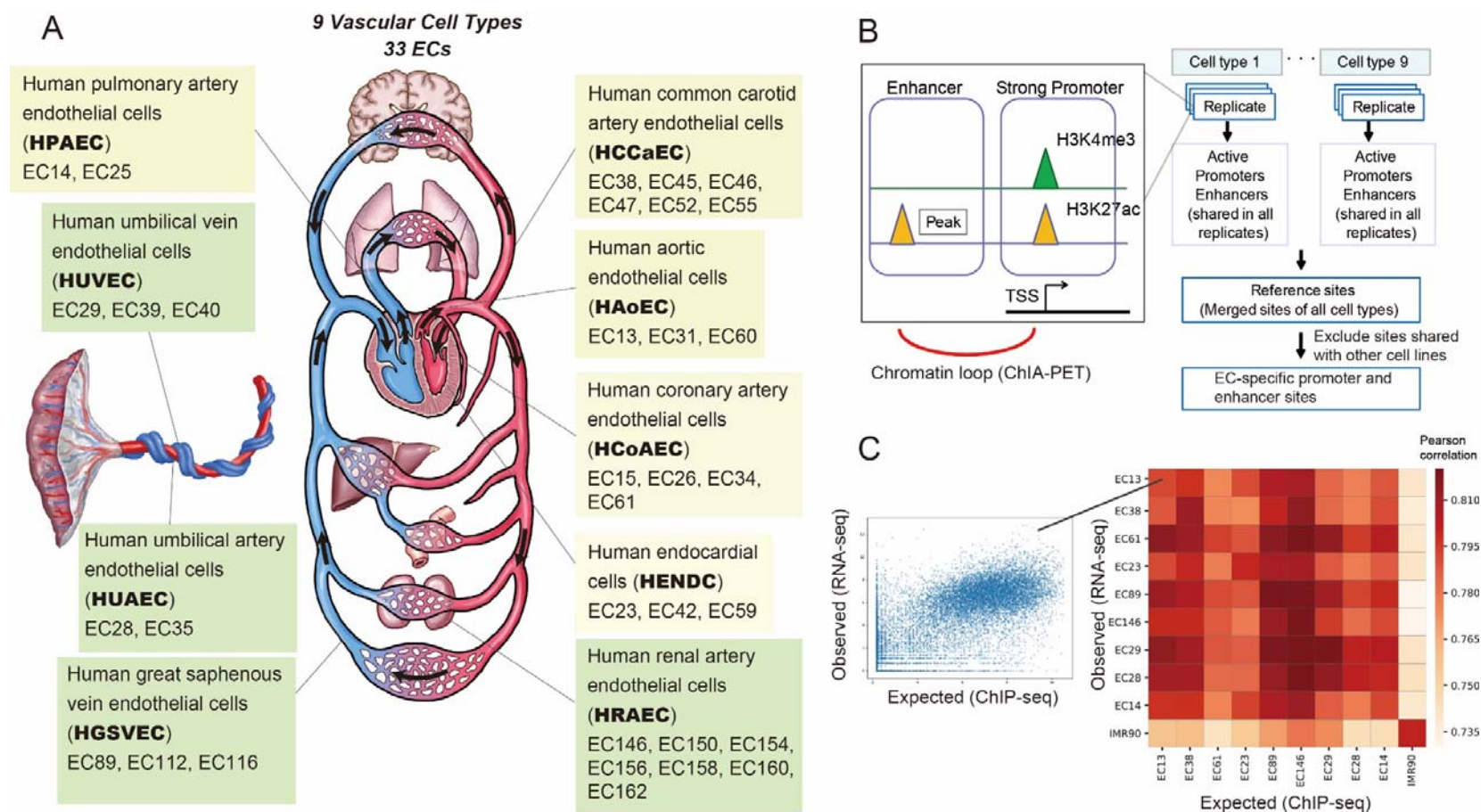


Figure 1. Summary of the cell types and histone modifications analyzed in this project. (A) Schematic illustration of the cardiovascular system, nine EC types and 33 individual samples (prefixed EC*) used in this paper. The yellow and green boxes indicate EC types

872 from upper body and lower body, respectively. (B) Workflow to identify the reference sites for ECs. The active promoter and enhancer
 873 sites of each sample were identified. For each cell type, the shared sites across all samples were extracted as the reference sites.
 874 These were integrated into a single set of reference sites for ECs, which was used for the downstream analyses. ChIA-PET data were
 875 utilized to identify the corresponding gene for the reference enhancer sites. (C) Correlation between observed and expected (from
 876 ChIP-seq analysis using linear regression model) gene expression data. Left: example scatterplot of observed and expected gene
 877 expression level for genes (data from EC13). Right: Pearson correlation heatmap for representative samples of nine cell types and
 878 IMR90 cells (as a negative control).
 879

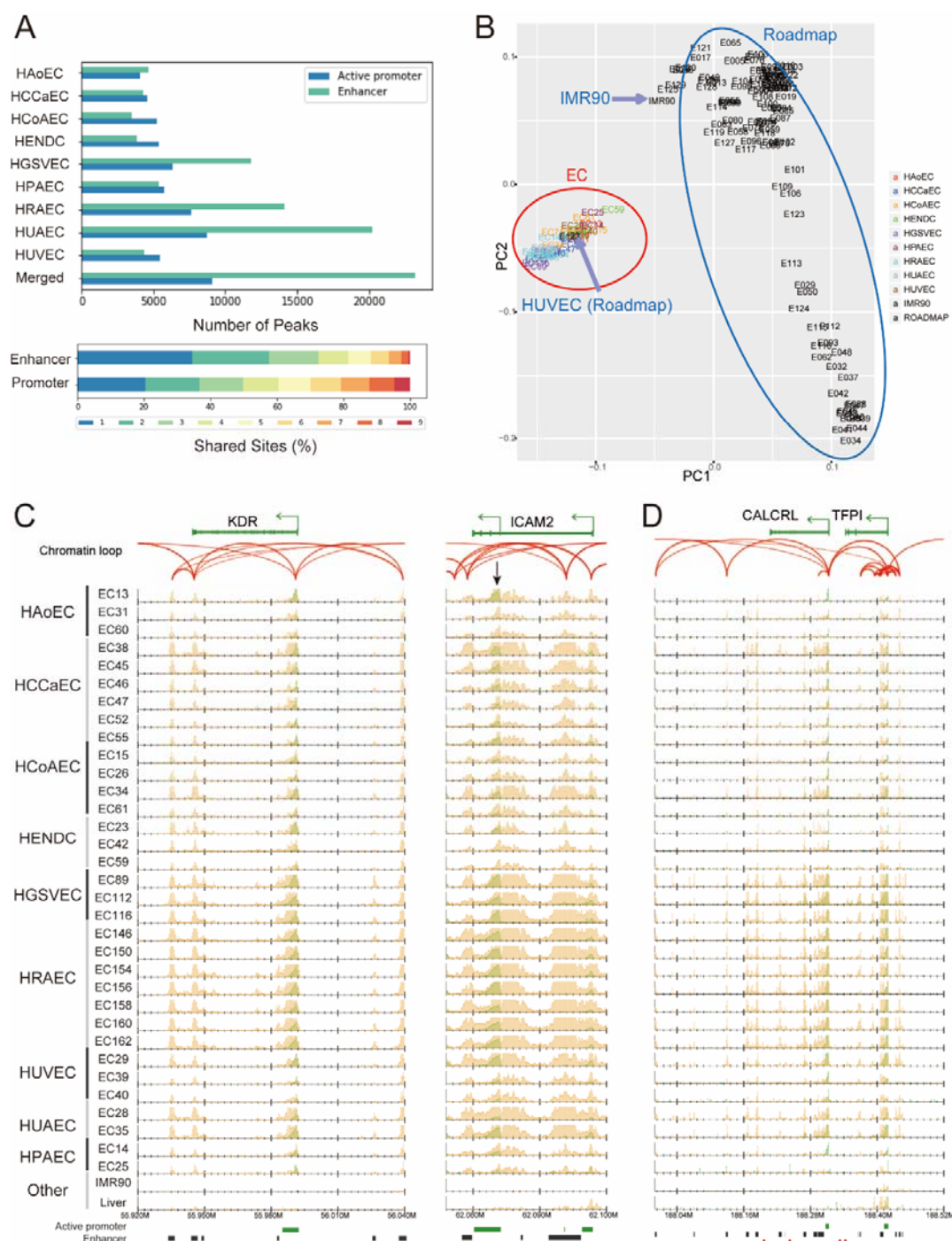


Figure 2. ChIP-seq data indicate variation in the chromatin status of ECs. (A) Top: The number of active promoter and enhancer sites for the nine cell types along with the merged reference sites. Bottom: The percentage of the reference active promoter and enhancer sites shared by one to nine of the EC types. (B) PCA plot using H3K27ac read

886 densities. All EC samples in this paper (red circle) as well as 117 cell lines from
 887 Roadmap Epigenomics Project (blue circle) are shown. The label colors indicate the EC
 888 types. (C, D) Normalized read distribution of H3K4me3 (green) and H3K27ac (orange)
 889 in representative gene loci (C) *KDR* and *ICAM2* and (D) *CALCRL* and *TFPI* for all ECs
 890 and two other tissues (liver from the Roadmap and the IMR90 cells in this study).
 891 Chromatin loops based on ChIA-PET (read-pairs) are represented by red arches. Green
 892 bars, black bars and red triangles below each graph indicate active promoter sites,
 893 enhancer sites and GWAS SNPs, respectively.
 894
 895

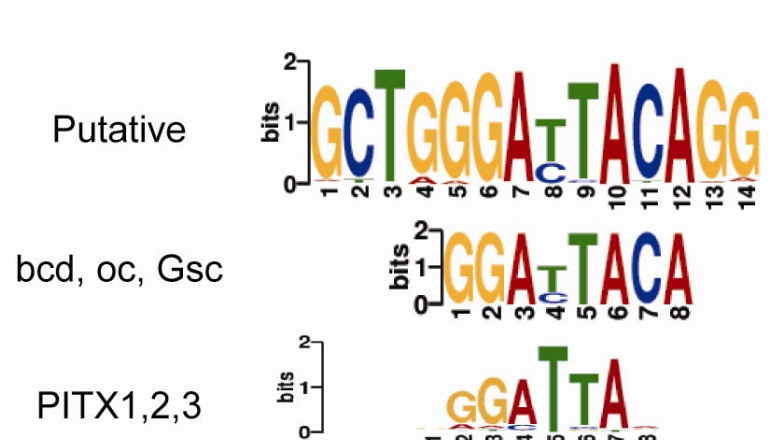


Figure 3. The identified *de novo* motif from EC-specific enhancer sites. The two related canonical motifs derived from JASPAR database are also shown.

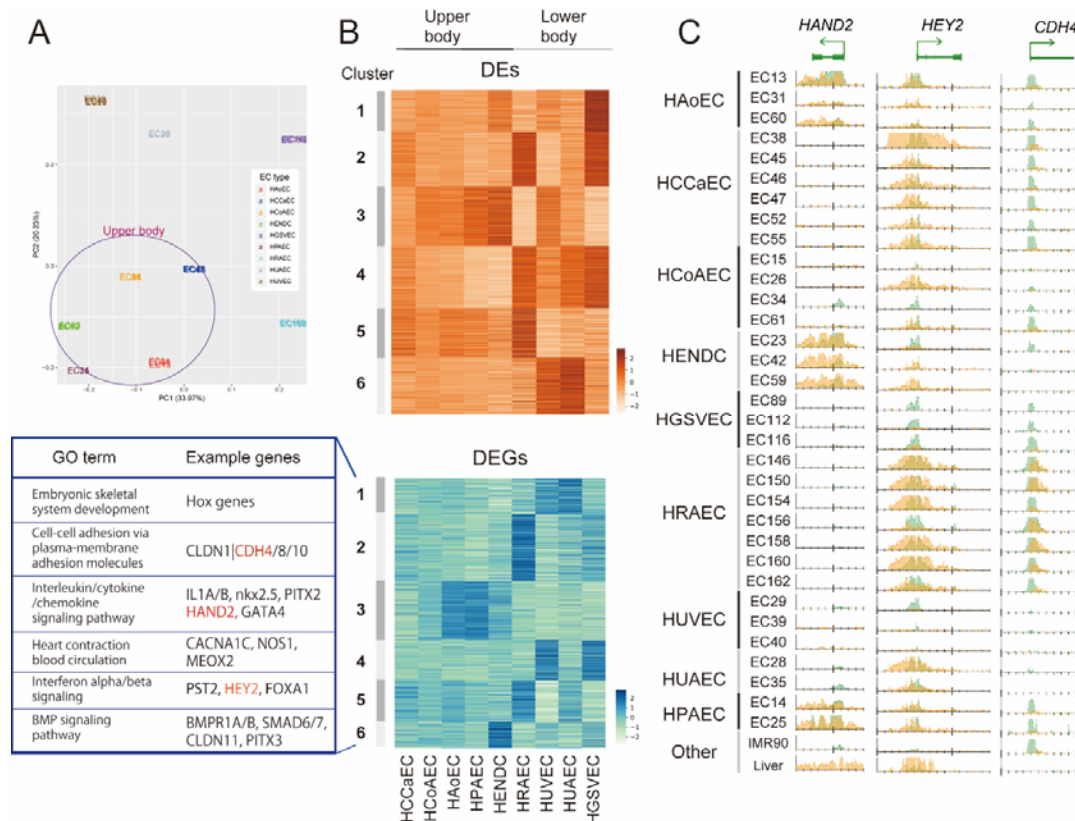


Figure 4: Comparative analysis of enhancer sites and gene expression across EC types. (A) PCA plot of EC samples based on H3K27ac read density fitted by generalized linear models. The color of samples indicates EC types. Samples from upper body are circled. (B) A k-means clustering (k = 6) analysis of DEs (upper) and DEGs (lower) across EC types (a representative for each type) based on Z-scores. The example genes and related GO terms obtained by Metascape (62) for DEG clusters are also shown. (C) Read distribution of H3K4me3 (green) and H3K27ac (orange) for the genes highlighted in red in panel (B).

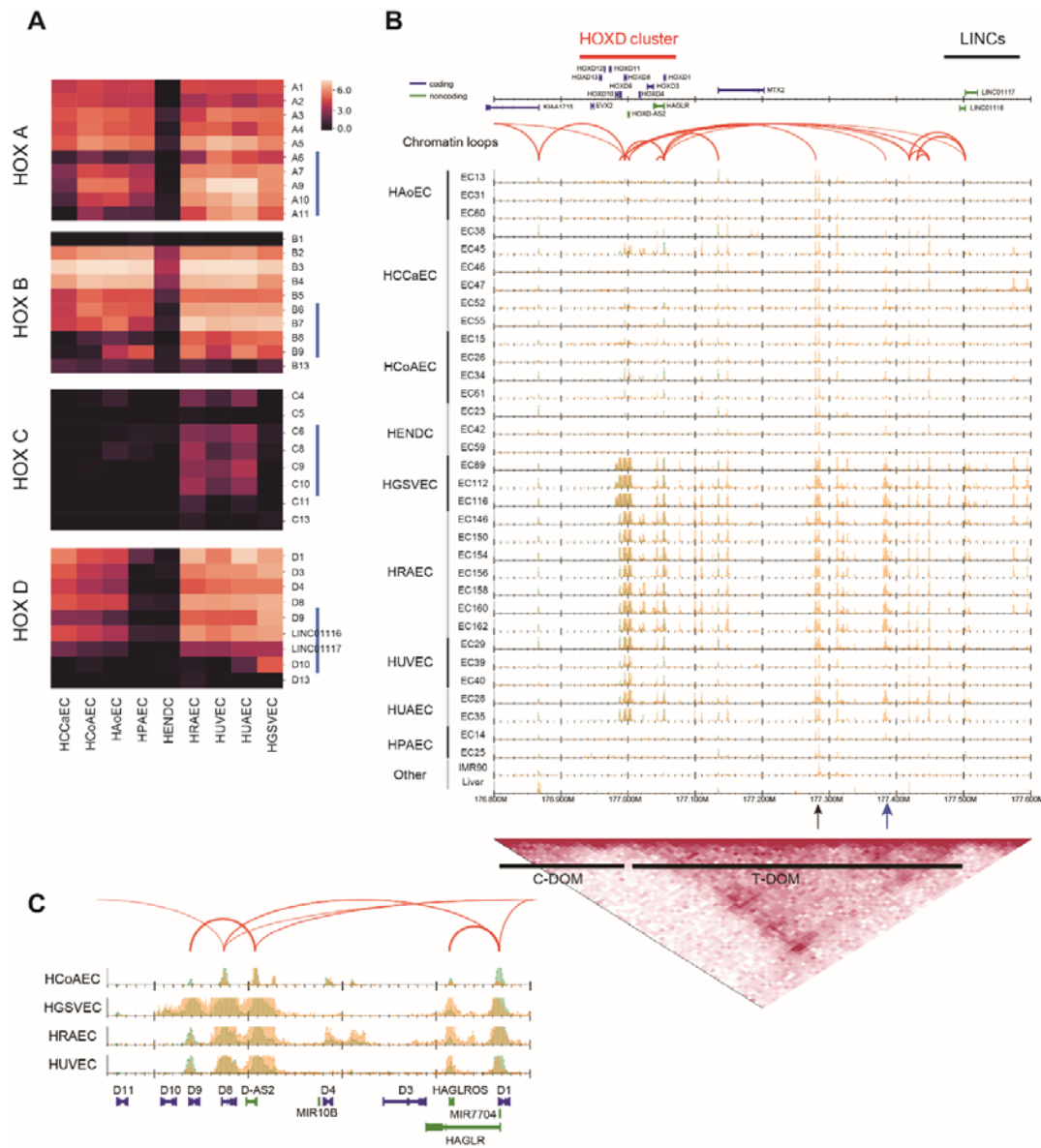


Figure 5: Differential expression of *HOX* genes. (A) Heatmaps visualizing the gene expression level (logged Transcripts Per Million (TPM)) of four *HOX* clusters and two long non-coding RNAs, LINC01117 (Hotdog) and LINC01116 (Twin of Hotdog). Blue vertical bars indicate the 5' *HOX* genes. (B) Read distribution around the *HOXD* cluster (chr2: 176.8–177.6 Mbp). Bottom: topological interaction frequency, telomeric domain (T-DOM) and centromeric domain (C-DOM) identified by Hi-C data for HUVECs. (C) Comparison of read profiles around the *HOXD* region for four EC types.