Enhancer jungles establish robust tissue-specific regulatory control in the human genome

Shan Li1, Ivan Ovcharenko1,*

1Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, 20892, USA

Abstract

An increasing number of studies suggests that functionally redundant enhancers safeguard development via buffering gene expression against environmental and genetic perturbations. Here, we identified over-represented clusters of enhancers (enhancer jungles or EJs) in human B lymphoblastoid cells. We found that EJs tend to associate with genes involved in the activation of the immune system response. Although spanning multiple genes, the enhancers within an EJ tend to collaborate with each other in regulating a single gene. The employment of homotypic transcription factor binding sites (TFBSs) in EJ enhancers and heterotypic TFBSs between constituent enhancers within an EJ may safeguard a robust transcriptional output of the target gene. EJ enhancers evolve under a weaker selective pressure compared to regular enhancers (REs), and approximately 35% of EJs do not have orthologues in the mouse genome. In GM12878, these unconserved EJs appear to regulate genes associated with the adaptive immune system response, while the conserved EJs are associated with innate immunity. The unconserved EJs elevate the expression of genes in human relative to mouse, thus facilitating the environmental adaptation of the organism during evolution. In short, the existence of EJs is a common regulatory architecture conferring a robust regulatory control for key lineage genes.

Keywords

gene regulation; enhancer redundancy; activation of immune system response; human adaptation during evolution

*Correspondence: ovcharen@nih.gov.
Author contribution statement
IO designed the study. SL performed the computational analysis. SL and IO wrote the manuscript. Both authors read and approved the final manuscript.

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest
The authors declare that they have no competing interests.
Introduction

The precise and robust regulation of gene expression is a cornerstone for normal development and cellular function of complex organisms. A class of cis-regulatory elements, called enhancers, plays a pivotal role in orchestrating spatiotemporally precise regulatory programs of gene expression. Enhancers are normally identified based on genome-wide chromatin profiling methods. They range in length from 100–1000 bp (1) and outnumber genes by approximately an order of magnitude (2). Their ability to activate transcription over long genomic distance is the most striking feature that sets them apart from promoters. In addition to the relative distance to the transcription start site, enhancers act to drive transcription independent of their orientation to their cognate promoters. Together, these features allow a gene to be regulated by multiple enhancers, conferring redundant regulatory repertoires to buffer the risk of lethality caused by environmental or genetic variability.

With our greater appreciation of the complexity of the genomic organization, it is becoming apparent that enhancers do not function in isolation. Many developmental genes are regulated by multiple enhancers with overlapping or distinct spatiotemporal activities (3–5). In *Drosophila melanogaster*, functionally redundant enhancers (referred to as “shadow enhancers”) have been identified for several developmental genes (6). Each enhancer separately is dispensable for proper gene function in standard laboratory conditions, but the presence of primary and shadow enhancers is essential for buffering individual enhancer activity to facilitate a robust transcriptional output or, in other words, canalization of developmental processes, as coined by Levine and colleagues (7). Recently, another study performed by the Pennacchio, Dickel and Visel groups showed that redundancy is widespread among developmental enhancers in mammals (8). Applying genome editing to the pairwise redundant limb enhancers located in the loci of two critical limb developmental genes, *Gli3* and *Shox2*, they observed that the deletion of pairs of redundant enhancers, but not single enhancers, caused noticeable changes in limb skeletal morphology. More importantly, they found that more than 1000 genes associated with five or more enhancers exhibited redundant activity patterns (with highly overlapping activity patterns in the same tissue), especially those genes encoding transcription factors (TFs) regulating developmental expression in limb, heart, and brain tissues, indicating that developmentally expressed genes are commonly associated with enhancer redundancy to provide protection against loss of individual regulatory functions. These studies have raised several questions in regard to the role of redundant enhancers in gene regulation. Is enhancer redundancy exceptional for developmental processes? Is it common for a gene to be associated with cluster(s) of enhancers? Do enhancers located in genomic proximity tend to code a redundant regulatory program to fine-tune the transcription of a single gene? What are the regulatory relationships among these closely located and simultaneously active enhancers? How were these enhancer clusters established during evolution?

To systematically explore these questions, in this study, we applied an HMM-based approach to identify a set of overrepresented clusters of enhancers (dubbed enhancer jungles or EJs) that locate close to each other in a B-lymphoblastoid cell line, GM12878, and observed that the constituent enhancers within EJs are enriched in the loci of genes regulating activation of immune system response. In spite of spanning two to three genes,
their constituent enhancers tend to collaborate with each other to fine-tune only one gene. EJs enhancers are relatively lowly constrained on the sequence level compared to regular enhancers (REs). The unconserved EJs elevate the expression of their associated genes involved in adaptive immune response to better protect the organisms against exposure to new pathogenic insults that arise from physiological or environmental changes during evolution.

**Results**

**B cell EJs span large genomic domains associated with genes involved in activation of immune response**

This study aims to identify over-represented clusters of enhancers (dubbed enhancer jungles or EJs) in the human genome and characterize their genomic features, evolutionary constraints and biological functions. To identify EJs, we initially screened the EUs (enhancers within 1kb of each other are combined into stitched enhancer regions, dubbed enhancer unit, see Methods) in B-lymphoblastoid cell lines (GM12878) by using a HMM trained for GM12878 (see Methods). In total, we identified 85 clusters of contiguous enhancer units spanning large genomic regions, with their median size two orders of magnitude larger than that of regular enhancers (in GM12878, 120 kb versus 1kb) (Figure 1A, Table 1, Methods). The large enhancer domains of EJs generally span multiple genes (median = 2) (Figure 1B). The 85 EJs correspond to 524 enhancers (1.5% of all enhancers, Table 1), which have similar genomic locations as REs (Figure S1). EJ enhancers also have similar levels of GC content and non-repetitive region content as compared with REs (Figure S2). However, further characterization of the genomic features of EJ enhancers revealed that their proximal genes have higher expression level compared to those associated with REs (Figure 1C), enclosing transcription factors essential for B cell identity and survival, such as IKZF2/3 (9–12), STAT3 (13), NFATC2 (14,15), and IRF8 (16) (Table S1). The overall GO enrichment shows that the EJ enhancers are largely enriched in the regulatory domains of genes associated with the toll-like receptor signaling pathway, regulation of leukocyte/lymphocyte/B cell activation, activation of innate immune response, activation of immune response, and regulation of leukocyte/lymphocyte/B cell proliferation, as compared with REs (Figure 2; Methods). Notably, toll-like receptors play crucial roles in the innate immune system and signal through the recruitment of specific adaptor molecules, leading to activation of the transcription factors NF-κB and IRFs, which dictate the outcome of innate immune responses (17). In short, the genes related to critical B cell-specific biological pathways, such as activation of immune response, may necessitate precise regulatory control established by multiple enhancers. We, therefore, speculate that their coding mutations are likely to be linked to a deleterious phenotype change. By contrast, the deleterious effect of the mutations occurring at EJ enhancers might be buffered due to the redundancy of the enhancers. Indeed, the genes proximal to EJs are enriched with coding GWAS traits, whereas their EJ enhancers are depleted with non-coding GWAS traits, although the gene loci associated with EJ enhancers have slightly greater enrichment of GWAS traits compared with those associated with REs (Figure 1D, Figure S3).
Zooming in on the EJ associated with the gene GFI1, which is required to support B cell differentiation (18), we observed that the EJ (chr1:92,949,208–93,050,811) comprises five enhancer units encompassing two genes: GFI1 and EVI5 (Figure 1E). Notably, despite the fact that GFI1 is differentially highly expressed (TPM = 13.14) in lymphocytes, EVI5 has a very low expression (TPM = 1.14) in lymphocytes but is highly expressed across many other tissues (Figure S4). Although sharing the same EJ, not all associated genes but GFI1 seem to obtain elevated expression, which motivated us to raise the questions: Do EJ enhancers regulate multiple genes separately or do they collaborate with one another to orchestrate the expression of a single gene? Do EJ enhancers harbor homologous or heterogenous TFBS profiles? To answer these questions, we first needed to gain insight into the sequence features and epigenetic features of EJs underlying their regulatory mechanism.

**Heterogeneous EJ enhancers collaborate to orchestrate expression of a single gene**

To investigate the sequence features of EJ enhancers, we examined the occupancy in the two classes of enhancers of 54 different TFs using the available ChIP-seq data for the GM12878 cell line from ENCODE (Materials and Methods). Overall, the majority of essential B cell-specific TFs exert a slightly stronger enrichment in EJ enhancers compared with REs (Figure 3A), including the pioneer factors PU.1 (19) and EBF1 (20,21), the master regulator PAX5 (22), and critical TFs governing B cell identity and functions such as IRF4 (16), IKZF1 (9,11), NFATC1 (23), NF-κB (24), and MEF2C (25). The stronger TF binding in EJ enhancers compared to REs motivated us to conjecture that homotypic TF binding (defined as chromosomal segments that contain at least two TF ChIP-seq peaks within a 1kb window) might be employed by EJ enhancers to a larger extent. Indeed, the homotypic binding of key B cell-specific transcription factors including PAX5, NFATC1, IKZF1, IRF4, EBF1 and PU.1 is enriched in EJ enhancers relative to REs (Figure 3B). Notably, the constituent enhancers within an EJ tend to harbor heterogeneous cohorts of TFs as compared with expectation (Figure 3C). Taken together, the usage of homotypic TF binding in an EJ enhancer and the employment of heterotypic TF binding across the constituent enhancers within an EJ may provide regulatory robustness for a stable transcriptional output, whereby the shutting down of one TF or one biological pathway will not affect the expression of the target gene to a large extent.

We then turned to the question of whether EJ enhancers constitute a functional unit based on a novel regulatory paradigm or simply represent an assembly of enhancers that independently regulate different sets of genes. First of all, we would like to know whether the genes overlapping an EJ are their regulatory targets, considering that enhancers tend to loop to, and are associated with, adjacent genes in order to activate their transcription (26). To explore this question, we first examined the co-activity between EJ enhancers and their overlapping genes, taking advantage of chromatin signatures of enhancers and the transcription of EJ-spanned genes measured across multiple tissues (Methods). Surprisingly, the correlations of the co-activity between EJ enhancers and their overlapping genes is smaller than that of random enhancer-gene pairs with the same distances (Figure 4B), leading to the hypothesis that these constituent enhancer elements within an EJ tend to regulate only a subset of the overlapping genes by skipping nearby genes. In other words, only part of the EJ-spanned genes are highly expressed due to a targeted regulation by
redundant and interconnected EJ enhancers. Taking advantage of the 5 kb resolution Hi-C map (27), we examined the expression level of the genes overlapping an EJ and compared the subset of genes linked to the constituent EJ enhancers with those not linked to EJs. We observed that the EJ-spanned genes linked to EJ enhancers show a significantly elevated expression (median RPKM = 6.822) compared with those that are not linked (median RPKM = 0.026) (Figure 4C), suggesting that only a subset of EJ-spanned genes is regulated by EJ. In turn, the most highly expressed genes (the top 10% of genes, dubbed top, that have the largest fold enrichment compared with their median level expression across tissues) are likely to have more 3-D contacts with EJ constituent enhancers, as compared with the remaining EJ-spanned genes, dubbed nonTop genes (Figure 4D). Interestingly, we observed that an EJ tends to regulate only one gene (Figure 4E), although, on average, EJs encompass two genes (Figure 1B). Going back to the example of the EJ overlapping the locus of GFI1 (Figure 1E), the five enhancer units, indeed, together regulate GFI1 while skipping EVI5, which is highly expressed in other tissues (Figure S5). Together, these results indicate that the constituent enhancers within an EJ incline to skip nearby genes and collaborate with each other to regulate single gene expression.

**Recently acquired enhancers from EJs elevate gene expression levels in humans and are associated with B cell activation**

We next asked how these enhancers collectively established the overrepresented clusters during evolution. Did they arise from ancestral DNA or have they recently emerged? To answer this question, we first needed to obtain an empirical understanding of the evolutionary stability of EJ enhancers. An enhancer is considered to be conserved if it has an evolutionarily conserved region (ECR) (≥100 bp and with ≥70% identity) between human and mouse (28,29). Overall, approximately 55% (291/524) of EJ enhancers are conserved, as compared with 62% (21784/35242) of REs are conserved (Fisher’s exact test p-value = 2.06e-3) (Figure 5A). This observation reveals that EJ enhancers evolved under weaker selective pressure compared with REs, which to some extent may be due to the redundancy among EJ enhancers as discussed above (Figure 3BC). We next defined the conservation of EJs based on their fraction of conserved constituent enhancers. Specifically, the EJs with at least 60% of conserved enhancers were considered as conserved EJ, whereas the EJs with no more than 20% of conserved enhancers were considered as human-specific (recently acquired) EJs (Figure S6). Overall, 47 out of 85 EJs were conserved, while 30 out of 85 EJs were human-specific (Table S2). We speculated that the human-specific EJs are likely to elevate the expression of their target genes compared to that of their mouse orthologues. To explore this hypothesis, we compared the expression of genes overlapping the two groups of EJs (only the genes with Hi-C contacts to EJs were considered) in human and mouse B cells, separately. The genes associated with human-specific EJs, indeed, exhibited a strongly elevated expression relative to mouse, as compared with those associated with conserved EJs (Figure 5B). In addition, the unconserved EJs were chiefly associated with activation of multiple key components of immune response, such as platelet activation, leukocyte activation, and positive regulation of lymphocyte activation. By contrast, the conserved EJs were associated with the toll-like receptor signaling pathways which play important roles in innate immune responses to given pathogens (Figure 5C) (30).
Next, we asked whether the EJ conserved and human-specific enhancers harbor distinct cohorts of TFs that underlie the distinct biological pathways associated with these two classes of EJs. By examining the TF binding profiles of the two sets of EJ enhancers, we observed that most B cell-specific TFs had a greater enrichment in EJ conserved enhancers compared to EJ human-specific enhancers (Figure 5D), which to some extent is due to the enrichment of homotypic TF binding in the conserved EJ enhancers (Figure S7). Notably, the enhancers of the same category (conserved or human-specific) within an EJ did not exhibit greater similarity of TF binding profiles with each other, as compared with the enhancers of different kinds (conserved versus human-specific) within an EJ (Figure S8), further corroborating the heterogeneous TF binding patterns of the constituent EJ enhancers (Figure 3C). Different from the majority of B cell-specific TFs, Ikzf1 was exceptionally enriched in EJ human-specific enhancers compared with conserved enhancers (Figure 5D). Ikzf1 (IKAROS) is one of the earliest regulators of lymphoid lineage identity and a guardian of lymphocyte homeostasis (10). The role of the IKAROS gene family in mature B cells is highlighted by the studies of Ikzf3 (AIOLOS), in which the Ikzf3 deficiency in mature B cells causes an increase in B cell receptor (BCR)-mediated proliferative response (31). More importantly, Ikzf3 functions with Ikzf1 to raise the threshold for antigen-stimulated B-cell activation and, thereby, controls the magnitude of an immune response (32). Notably, in GM12878, the human-specific EJ enhancers associated with Ikzf3 were bound by several Ikzf1s, further advocating not only the role of human-specific EJs in the activation of key components of immune response including B cell activation (Figure 5C), but also the role of the interactions among IKAROS family in mature B cells to control their ability to respond to antigen stimulation by raising the bar for B cell activation. Moreover, as the B cell receptor (BCR)/T cell receptor (TCR)-based system is one of the hallmarks of adaptive immunity in jawed vertebrates, allowing them to recognize almost any antigen, the unconserved EJs are likely to be associated with adaptive immunity that evolved along with the emergence of vertebrates(33). By contrast, the conserved EJs are likely to associated with innate immunity that is present in all pants and metazoans, and even in some unicellular organisms (33).

Discussion

We identified large enhancer domains consisting of overrepresented clusters of enhancers, or EJs, in the human genome. The constituent enhancers within an EJ locate in close genomic proximity but not too close (> 12.5 kb apart from each other). In GM12878, the genes associated with EJs exhibit higher expression level compared with regular enhancers, and are likely to be involved in activation of immune response. The majority of B cell-specific TFs have slightly greater enrichment and are more likely to function in the fashion of homotypic binding in EJ enhancers relative to regular enhancers. Nevertheless, the constituent EJ enhancers harbored heterogeneous cohorts of TFs with each other to provide the target genes with robust regulatory control. Notably, although spanning multiple genes, the EJ enhancers appear to form a functional unit regulating one gene while skipping nearby genes, rather than an assembly of enhancers regulating different sets of genes.

Overall, EJ enhancers evolved under a weaker evolutionary constraint compared to regular enhancers (Figure 5). Remarkably, around 35% (30/85) of EJs are human-specific, with
more than 80% of their constituent enhancers having no orthologous counterpart in mouse. The emergence of these human-specific EJs, indeed, elevates the expression level of the target gene in human as compared to mouse. The human-specific EJs are associated with activation of multiple components of the immune system response that are parts of adaptive immunity. By contrast, the conserved EJs are associated with toll-like receptor signaling pathways that play essential roles in innate immunity. Overlapping the EJ conserved and EJ unconserved enhancers with GWAS studies, we observed that the immune response-related traits are largely enriched in the EJ unconserved enhancers as compared with the EJ conserved enhancers (Figure 5E), which might be associated with the enrichment of homotypic TFBS clusters in the EJ conserved enhancers, whereby the existence of multiple copies of binding sites for the same TF may buffer the risk caused by mutations occurring at one of the binding sites.

Our study has identified a panel of redundant enhancers that act as regulatory buffers for the transcription of key genes associated with cellular identity and diseases, suggesting that a large degree of redundancy of regulatory control is not the only privilege for developmental processes when facing environmental and genetic variability (8,34,35). In fact, the usage of EJs is widely spread across the human genome. We applied the HMM to 27 tissues (Materials and methods) and observed that 5,631 genes are associated with at least one EJ in at least one tissue. This number could be overestimated since EJs appears to regulate a single gene from afar (Figure 3, Figure 4), although spanning two to three genes on average (Figure S9). There are still several unanswered questions that require follow up analyses. Specifically, do the constituent enhancers confer functional redundancy in an additive or synergistic fashion? Is an EJ more than the sum of its enhancer elements? Do EJs tend to function following a hierarchical or competitive logic? Since the constituent enhancers within an EJ do not seem to be strongly interlinked based on the high-resolution Hi-C data from GM12878 (Figure S10), EJ enhancers may collaborate with each other either in an additive or a competitive model to regulate their target genes. In the latter potential mode, multiple enhancers compete for association with a shared promoter, buffering individual enhancer activities to facilitate a constant transcription output. However, approaches of genome editing, such as the CRISPER-Cas9 system, offer the opportunity to manipulate specific genomic loci to systematically address these questions.

Materials and Methods

Data availability

The GRCh37 (hg19) assembly of the human genome used in this study was downloaded from the UCSC Genome Browser ((36,37); http://genome.ucsc.edu/). The ChIP-seq of epigenetic marks, DNase-seq and RNA-seq data of the human genome were downloaded from the Roadmap Epigenomics Project (38). The ChIP-seq peaks (narrowPeak format) of TFs and histone marks of GM12878 cell line were downloaded from the USCS Genome Browser (39). Gene Ontology (GO) terms were downloaded from the Gene Ontology Consortium (40,41). The pairwise alignments of hg19 to mm9 were downloaded from the UCSC genome browser (ftp://hgdownload.soe.ucsc.edu/goldenPath/hg19/vsMm9/axtNet/) (39). The H3K27ac ChIP-seq peaks and gene expression data of 13 tissues (Liver, Bone

Genomics. Author manuscript; available in PMC 2021 May 01.
Marrow, Cerebellum, Cortex, Heart, Intestine, Kidney, Lung, Olfactory bulb, Placenta, Spleen, Testes, Thymus) in adult mouse were downloaded from the NCBI Gene Expression Omnibus (GEO) (accession number: GSE29184). The Hi-C contact map with 5 kb resolution in GM12878 was downloaded from GEO (accession number: GSE63525). The gene expression in resting B cells of adult mouse was obtained from the study of Kieffer-Kwon et al. (42). The NHGRI GWAS Catalog was downloaded in September 2016 (43).

### Identification of enhancers and enhancer jungles

Putative enhancers were defined as 1 kb regions extending from the central coordinate of the overlapping H3K27ac and H3K4me1 peaks. Promoters were defined as regions within the upstream 1,500 bp and downstream 500 bp of the transcriptional start sites of genes annotated in the UCSC Genome Browser (37,44), which were excluded from the putative enhancer regions. The enhancers located within 1 kb from each other were further merged to a unit (dubbed enhancer unit, or EU), considering that they might stem from a large enhancer region. We applied a two-state HMM to predict EJs in a tissue. Specifically, any pair of two consecutive EUs corresponded to the hidden state of either “cluster pair” or “non-cluster pair”. The Baum-Welch algorithm was applied to estimate the parameters of the two-state HMM. The HMM was trained independently for each tissue. The inter-EU distances were log<sub>2</sub>-transformed, and the top 40% quantile was used as the cutoff to define the enhancer clusters and initialize the transition and emission probabilities of the model. Once the HMM had converged according to the EM algorithm, the Viterbi algorithm was applied to decode the hidden states of all the consecutive EUs. Finally, only the EU pairs that were decoded as a “cluster pairs” and located more than 12.5kb apart from each other were defined as “cluster pairs”. Otherwise the EU pairs were defined as “non-cluster pairs”. Stitching the decoded cluster pairs, we defined the clusters consisting of at least five EUs as EJs. Notably, we used 12.5 kb as the lower boundary to explicitly exclude super enhancers from our detected enhancer jungles, considering that super enhancers were Med1/H3K27ac signal-enriched regions composed of concatenated neighboring enhancers within 12.5 kb (34). We identified EJs in 27 human tissues/cell lines with available gene expression data (Table S3) that were selected as representatives of the tissue/cell line groups based on the type of biological materials according to Roadmap Epigenomics Project (38).

### GO enrichment analysis

Functional enrichment of enhancers was performed using the online Genomic Regions Enrichment of Annotations Tool (GREAT) version 3.0.0. The default distance parameter was applied for the regulatory domain assignment of genes, and the single nearest gene rule was applied to associate enhancers with genes. GO biological process terms were included only if they satisfied the strict criteria in at least one category of enhancers: 1) binomial p-value ≤ 1e-10 and hypergeometric p-value ≤ 1e-3; 2) a minimum binomial observed region hits of 10 and hypergeometric observed gene hits of five; 3) a minimum binomial region and hypergeometric gene set fold enrichment of two. The binomial fold enrichment was used to evaluate the enrichment of GOs.
TFBSs enrichment in different classes of enhancers

We used the available ChIP-seq TFBS data to calculate the TFBS enrichment in EJ enhancers and regular enhancers in GM12878. We randomly sampled 10,000 DNase I Hypersensitive site (DHS) regions as the control set, all of which were enlarged to 1 kb regions from the central coordinate of the DHS peaks. The fold enrichment of a TFBS in an enhancer class (positive set) was calculated as a ratio of its density in the positive set to that in the control set, where the density was computed as the number of TF ChIP-seq peaks overlapping the enhancer set normalized by the total length of the enhancer set. The fold enrichment of a TFBS in EJ enhancers relative to regular enhancers was then computed as the ratio of its fold enrichment in EJ enhancers to that in regular enhancers. When comparing the TFBS profile similarity of EJ enhancers, five-fold random regular pairs with the same inter-enhancer distance were extracted as the control. Notably, since the neighboring constituent enhancers of an EJ located closer to each other compared to the majority of remaining neighboring regular enhancers based on our algorithm, we only focused on the non-adjacent EJ enhancer pairs when generating control-enhancer pairs and comparing the similarity of TFBS profile between enhancers.

Chromatin Contacts

Hi-C data from GM12878 B-lymphoblastoid cell lines at 5 kilobases (kb) resolution were retrieved from Rao’s work (GSE6352) (27). The Iterative Correction and Eigenvector decomposition (ICE) algorithm (45) was implemented in the Hi-Corrector package (46) to remove biases. We then used Fit-Hi-C (47) to detect statistically significant interactions with the parameters `-U=2000000, -L=50000` along with the threshold of FDR = 0.01. The Hi-C bin overlapping the center of the middle enhancer of an EU was assigned as the Hi-C bin associated with the EU.

Identification of enhancer orthologue

We mapped human enhancers to the mouse genome using the UCSC chained and netted pairwise alignment between hg19 and mm9. More specifically, the alignments that overlap the human enhancer and include mm9 were extracted. An enhancer was considered to be conserved if it had an evolutionarily conserved region (ECR) ( ≥100 bp and with ≥70% identity) between human and mouse (28,29).

Correlation between enhancer activity and associated gene expression

The GM12878 enhancers overlapped with H3K27ac peaks across 27 tissues. For the enhancers with multiple overlapping peaks in a tissue, the maximum signal value of all the overlapping peaks was used to estimate the activity of the enhancer in the tissue. Given a pair of EJ enhancers and its associated gene, five-fold random RE and gene pairs with the same distance were extracted as the control. The Pearson correlation coefficient between the enhancer H3K27ac signal quantifications and the gene expression measurement (RPKM) across 27 tissues was used to evaluate the correlation between enhancer activity and its associated gene expression.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine. The authors are grateful to Dorothy L. Buchhagen for critical reading of the manuscript.

References

1. Schaffner W. (2015) Enhancers, enhancers - from their discovery to today's universe of transcription enhancers. Biol Chem, 396, 311–327. [PubMed: 25719310]
2. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C and Snyder M. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57–74. [PubMed: 22955616]
3. Goto T, Macdonald P and Maniatis T. (1989) Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell, 57, 413–422. [PubMed: 2720776]
4. Yuh CH and Davidson EH. (1996) Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development, 122, 1069–1082. [PubMed: 8620834]
5. Werner T, Hammer A, Wahluhl M, Bosl MR and Wegner M. (2007) Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. Nucleic Acids Res, 35, 6526–6538. [PubMed: 17897962]
6. Barolo S. (2012) Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays, 34, 135–141. [PubMed: 22083793]
7. Hong JW, Hendrix DA and Levine MS. (2008) Shadow enhancers as a source of evolutionary novelty. Science, 321, 1314. [PubMed: 18772429]
8. Osterwalder M, Barozzi I, Tissieres V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Pflaiz-Frick I, Pickle CS. et al. (2018) Enhancer redundancy provides phenotypic robustness in mammalian development. Nature, 554, 239–243. [PubMed: 29420474]
9. Schwickert TA, Tagoh H, Gultekin S, Dakic A, Axelsson E, Minnich M, Ebert A, Werner B, Roth M, Cimmino L. et al. (2014) Stage-specific control of early B cell development by the transcription factor Ikaros. Nat Immunol, 15, 283–293. [PubMed: 24509509]
10. Georgopoulos K. (2017) The making of a lymphocyte: the choice among disparate cell fates and the Ikaros enigma. Genes Dev, 31, 439–450. [PubMed: 28385788]
11. Sellars M, Kastner P and Chan S. (2011) Ikaros in B cell development and function. World J Biol Chem, 2, 132–139. [PubMed: 21765979]
12. Klug CA, Morrison SJ, Masek M, Hahn K, Smale ST and Weissman IL. (1998) Hematopoietic stem cells and lymphoid progenitors express different Ikaros isoforms, and Ikaros is localized to heterochromatin in immature lymphocytes. Proc Natl Acad Sci U S A, 95, 657–662. [PubMed: 9435248]
13. Chou WC, Levy DE and Lee CK. (2006) STAT3 positively regulates an early step in B-cell development. Blood, 108, 3005–3011. [PubMed: 16825489]
14. Ranger AM, Hodge MR, Gravalles E, Ouakka M, Davidson L, Alt FW, de la Brousse FC, Hoey T, Grusby M and Glimcher LH. (1998) Delayed lymphoid repopulation with defects in IL-4-driven responses produced by inactivation of NF-ATc. Immunity, 8, 125–134. [PubMed: 9462518]
15. Yoshida H, Nishina H, Takimoto H, Marengeere LE, Wakeham AC, Bouchar D, Kong YY, Ohteki T, Shahinian A, Bachmann M. et al. (1998) The transcription factor NF-ATc1 regulates lymphocyte proliferation and Th2 cytokine production. Immunity, 8, 115–124. [PubMed: 9462517]
16. Shukla V and Lu R. (2014) IRF4 and IRF8: Governing the virtues of B Lymphocytes. Front Biol (Beijing), 9, 269–282. [PubMed: 25506356]
17. Kawasaki T and Kawai T. (2014) Toll-like receptor signaling pathways. Front Immunol, 5, 461. [PubMed: 25309543]
18. Fraszczak J, Helness A, Chen R, Vadas C, Robert F, Khandanpour C and Moroy T. (2016) Threshold Levels of Gfi1 Maintain E2A Activity for B Cell Commitment via Repression of Id1. PLoS One, 11, e0160344. [PubMed: 27467586]
19. DeKoter RP and Singh H. (2000) Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science, 288, 1439–1441. [PubMed: 10827957]
20. Nechanitzky R, Akbas D, Scherer S, Gyory I, Hoyler T, Ramamoorthy S, Diefenbach A and Grosschedl R. (2013) Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. Nat Immunol, 14, 867–875. [PubMed: 23812095]
21. Boller S, Li R and Grosschedl R. (2018) Defining B Cell Chromatin: Lessons from EBF1. Trends Genet, 34, 257–269. [PubMed: 29336845]
22. Cobaleda C, Schebesta A, Delogu A and Busslinger M. (2007) Pax5: the guardian of B cell identity and function. Nat Immunol, 8, 463–470. [PubMed: 17440452]
23. Giampaolo S, Schebesta A, Delogu A and Busslinger M. (2007) Pax5: the guardian of B cell identity and function. Nat Immunol, 8, 463–470. [PubMed: 17440452]
24. Gerondakis S and Siebenlist U. (2010) Roles of the NF-kappaB pathway in lymphocyte development and function. Cold Spring Harb Perspect Biol, 2, a000182. [PubMed: 20452952]
25. Hergetz J, Unrau L, Hauschidt F, Fischer M, Kriebitzsch N, Alawi M, Indenbirken D, Spohn M, Muller U, Zieglar M. et al. (2016) Essential control of early B-cell development by Mef2 transcription factors. Blood, 127, 572–581. [PubMed: 26600426]
26. Ong CT and Corces VG. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet, 12, 283–293. [PubMed: 21358745]
27. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, 159, 1665–1680. [PubMed: 25497547]
28. Nobrega MA, Ovcharenko I, Afzal V and Rubin EM. (2003) Scanning human gene deserts for long-range enhancers. Science, 302, 413. [PubMed: 14563999]
29. Ovcharenko I, Nobrega MA, Loots GG and Stubbs L. (2004) ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res, 32, W280–286. [PubMed: 15215395]
30. Ali H, Haribabu B, Richardson RM and Snyderman R. (1997) Mechanisms of inflammation and leukocyte activation. Med Clin North Am, 81, 1–28. [PubMed: 9012753]
31. Wang JH, Avitahl N, Cariappa A, Friedrich C, Ikeda T, Renold A, Andrikopoulos K, Liang L, Pillai S, Morgan BA, et al. (1998) Aiolos regulates B cell activation and maturation to effector state. Immunity, 9, 543–553. [PubMed: 9806640]
32. Cortes M, Wong E, Koipally J and Georgopoulos K. (1999) Control of lymphocyte development by the Ikaros gene family. Curr Opin Immunol, 11, 167–171. [PubMed: 10322160]
33. Parra D, Takizawa F and Sunyer JO. (2013) Evolution of B cell immunity. Annu Rev Anim Biosci, 1, 65–97. [PubMed: 25340015]
34. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI and Young RA. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell, 153, 307–319. [PubMed: 23582322]
35. Frankel N, Davis GK, Vargas D, Wang S, Payne F and Stern DL. (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature, 466, 490–493. [PubMed: 20512118]
36. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D. (2002) The human genome browser at UCSC. Genome Res, 12, 996–1006. [PubMed: 12045153]
37. Casper J, Zweig AS, Villarreal C, Tynier C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Karolchik D. et al. (2017) The UCSC Genome Browser database: 2018 update. Nucleic Acids Res.
38. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature, 518, 317–330. [PubMed: 25693563]

39. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG. et al. (2013) ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res, 41, D56–63. [PubMed: 23193274]

40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 25, 25–29. [PubMed: 10802651]

41. (2019) The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res, 47, D330–D338. [PubMed: 30393331]

42. Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, Resch W, Baek S, Pruett N, Grontved L. et al. (2013) Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell, 155, 1507–1520. [PubMed: 24360274]

43. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L. et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res, 42, D1001–1006. [PubMed: 24316577]

44. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F. et al. (2006) The UCSC Genome Browser Database: update 2006. Nucleic Acids Res, 34, D590–598. [PubMed: 16381938]

45. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, Dekker J and Mirny LA. (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods, 9, 999–1003. [PubMed: 22941365]

46. Li W, Gong K, Li Q, Alber F and Zhou XJ. (2015) Hi-Corrector: a fast, scalable and memory-efficient package for normalizing large-scale Hi-C data. Bioinformatics, 31, 960–962. [PubMed: 25391400]

47. Ay F, Bailey TL and Noble WS. (2014) Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res, 24, 999–1011. [PubMed: 24501021]

48. (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet, 45, 580–585. [PubMed: 23715323]

49. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science, 348, 648–660. [PubMed: 25954001]
• There are large clusters of redundant enhancers in the human genome spanning hundreds of kilobases each; these have been dubbed enhancer jungles.
• Enhancer jungles span multiple gene loci but regulate a single gene.
• Enhancer jungles underly the immune response.
• Acquisition of enhancer jungles is associated with elevated and evolutionary stable level of gene expression.
Figure 1.
EJs are large enhancer domains associated with immune system response in GM12878. A) Distribution of EJ size (kb). B) Number of genes spanned by EJs, the dashed line indicates the median value. C) Genes associated with EJs have higher expression compared with those associated with regular enhancers. D) Fold enrichment of GO terms of the two categories of enhancers. E) Density (number per kb) of GWAS coding/non-coding traits in EJ enhancers and regular enhancers. F) An example of an EJ containing five EUs spanning two genes, with only one gene, GFI1, highly expressed in GM12878. The graph of the genomic coordinates with the track of H3K27ac and H3K4me1 enrichment data was obtained from the UCSC genome browser (36) (http://genome.ucsc.edu; UCSC hg19 assembly). The RNA-seq data were obtained from the GTEx midpoint milestone data release (V6, October 2015) (48,49).
Fig. 2.
Fold enrichment of GO terms of the two categories of enhancers.
Figure 3.
EJ enhancers have slightly stronger enrichment of TF binding. A) Enrichment of TF binding in EJ enhancers and REs. The legend shows the range of fold enrichment of TF binding in a set of enhancers relative to the negative control (see Methods). Only TFs enriched (Hypergeometric p-value $\leq 1e^{-3}$) with fold enrichment $\geq 1.2$ in at least one class of enhancers were included. B) Enrichment of homotypic binding of TFs in EJ enhancers relative to REs, where $Enrichment(TFA) = \frac{\text{total occurrence of homotypic binding of } TFA \text{ in EJs}}{\text{total EJ enhancer length}}$ relative to REs

$C) Jaccard index of TFBS profile of pairwise enhancers. P-value was calculated based on Wilcoxon test.
Figure 4. EJ enhancers tend to regulate a single gene that is highly expressed.
A) Pearson correlation coefficient between enhancer activity and EJ-spanned gene expression across 27 tissues. B) EJ-spanned genes with Hi-C links to EJ enhancers have higher expression than those without links. C) The most highly expressed genes (dubbed top) have larger number of 3D Hi-C contacts to EJ enhancers compared with the nonTop genes. D) Overall EJ enhancers tend to regulate only one gene. P-values were calculated based on the Wilcoxon test.
Figure 5. EJ enhancers elevate gene expression. The unconserved EJs are associated with B cell activation.

A) Fraction of enhancers that have ≥1 evolutionarily conserved regions (ECRs). P-values were calculated based on Fisher’s exact test. B) Comparison of gene expression associated with the two categories of EJs in human and mouse, respectively. C) Fold enrichment of GO terms of the two categories of EJ enhancers. D) Fold enrichment of TF ChIP-seq peaks in different sets of enhancers. The first column corresponds to enrichment of TFs in EJ unconserved enhancers compared to all DHS regions. The second column corresponds to enrichment of TFs in EJ conserved enhancers compared to DHS regions. The third column corresponds to enrichment of TFs in EJ unconserved enhancers relative to conserved ones. E) Density of GWAS traits in EJ unconserved and conserved enhancers, respectively.
Table 1.

EJ summary for the GM12878 cell line (EJ – enhancer jungle; EU – enhancer unit).

| EJs | EJ EUs | EJ enhancers | Total EUs | Total enhancers | Fraction of EUs within EJs | Fraction of enhancers within EJs | Average number of EUs per EJ |
|-----|--------|--------------|-----------|-----------------|-----------------------------|---------------------------------|-----------------------------|
| 85  | 460    | 524          | 30,369    | 35,766          | 0.015                       | 0.015                           | 5.4                         |