Pleiotropy and dynamics of *cis*-regulatory activities in human evolution

Abstract

Variations in *cis*-regulatory elements (CREs) can lead to differences in gene expression regulation, which likely underlie many phenotypic distinctions between human and non-human primates, as well as between human individuals. The modular nature of CREs has long been thought to confer advantages to their evolvability. However, CREs may be heterogeneous with regard to pleiotropy, with varying breadths of effects on biological processes. To examine the interplay between pleiotropy and dynamics of *cis*-regulatory activities during human evolution, we analyzed two published ChIP-Seq datasets of H3K27ac histone modification, which informs promoter and enhancer activity. One study identified lineage-specific increase of H3K27ac activity in the human embryonic limb by comparing signal in human, rhesus and mouse (Cotney et al., 2013). The other study identified ancestry-specific H3K27ac regions in human lymphoblastoid cell lines by studying individuals of varied ancestry (Kasowski et al., 2013). We examined the tissue-specificity of these lineage- and ancestry-specific regions, or variable sites, using H3K27ac ChIP-Seq data in various human tissues, and observed that the variable sites are active in fewer tissues compared to other sites. We also investigated the putative interactions between enhancers and promoters using DNaseI-based connectivity information, and found that the variable sites are connected to fewer promoters compared to other sites. These combined suggest lower level of pleiotropy of the variable sites. We further explored other features of these variable sites, including conservation, GC content, motif composition, and CTCF/cohesin binding, obtaining a deeper understanding of the properties associated with the evolution of *cis*-regulatory activities.

Introduction

It has been recognized for a long time that a gene, or a mutation in a gene, can affect multiple traits in an organism - a phenomenon called pleiotropy. On the role of pleiotropy during evolution, it is hypothesized that the pleiotropic effects of genes lead to the "cost of complexity", which compromises the capacity of organisms to evolve, because a mutation advantageous to one trait could be disadvantageous to other traits (Wagner and Zhang, 2011). *Cis*-regulatory elements (CREs), especially distal enhancers, control the spatial and temporal expression of genes and are generally thought to have lower level of pleiotropic effects than genes (Carroll et al., 2009). It has been proposed that *cis*-regulatory change is a major driving force of phenotypic adaptation, owing to their modular nature in gene regulation and hence restricted pleiotropy. However, like genes, CREs may be heterogeneous regarding their pleiotropic effects. While some CREs may be active in only one tissue or cell type, others could be functional at multiple developmental stages and/or in multiple tissues. The relationship between pleiotropy and evolution of CREs has been little explored, largely because genome-wide estimates for either aspect have been scarce. Recent development in genomic assays facilitated the global mapping of putative pleiotropic effects and evolutionary dynamics at CREs, allowing us to explore their relationship.

The genome-wide study of CRE evolution has been under active pursuit. Traditionally, people used pure computational methods that rely on models of sequence and

Jing Leng 7/8/14 6:22 PM
Deleted: feature

Jing Leng 7/8/14 6:22 PM
Deleted: as a class

Jing Leng 7/8/14 6:22 PM
Deleted: their

Jing Leng 7/8/14 6:22 PM
Deleted: queried

Jing Leng 7/8/14 6:22 PM
Deleted: regions

Jing Leng 7/8/14 6:22 PM
Deleted: in regard to

Jing Leng 7/8/14 6:22 PM
Deleted: were

Jing Leng 7/8/14 6:22 PM
Deleted: functional genomics

Jing Leng 7/8/14 6:22 PM
Deleted: change

corresponding function evolution, and examined the sequence alignments across species to identify conserved as well as lineage-accelerated CREs (Bejerano et al., 2004; Prabhakar et al., 2008). More recently, comparative studies using chromatin immuno-precipitation coupled with sequencing (ChIP-Seq) have provided maps of putative CRE dynamics, either across species or across individuals within the same species. Such maps provide direct measurements of CRE dynamics in specific biological contexts based on biochemical properties, having the advantage of making no assumptions on the sequence-function correlation in CREs, which is still poorly understood. Among such studies, the ones that targeted the histone modification H3K27ac inform the dynamic landscapes of putative active promoters and enhancers. To investigate properties associated with variable *cis*-regulatory activities in human evolution, we used data from two studies that conducted H3K27ac ChIP-Seq experiments. One study explored the *cis*-regulatory evolution of human embryonic limb development by comparing H3K27ac signal in human, rhesus and mouse, and identified lineage-specific increase of H3K27ac activity in human (Cotney et al., 2013). The other study investigated chromatin state variation in lymphoblastoid cell lines (LCLs) across human individuals of varied ancestry, and identified ancestry-specific H3K27ac regions (Kasowski et al., 2013). To examine the relationship between pleiotropy and evolutionary dynamics at CREs, we studied various properties of the lineage- and ancestry-specific H3K27ac regions, referred to as variable sites collectively in the remainder of this manuscript. The pleiotropic effects of CREs may be manifested in two major aspects - that the elements are active in multiple conditions themselves, and/or that the elements regulate multiple target genes. To investigate both of these two aspects, we utilized data from relevant genomic assays, focusing on the contrast between the variable sites and the rest of the sites identified in respective studies. In addition, we investigated other properties that are closely related to pleiotropy – conservation, GC content, motif content, CTCF/cohesion binding. Finally, we examined the potential functional impacts of the ancestry-specific regions by cross-referencing eQTLs studies in human LCLs.

Jing Leng 7/8/14 6:22 PM
Deleted: functional divergence

Jing Leng 7/8/14 6:22 PM
Deleted: provide

Jing Leng 7/8/14 6:22 PM
Deleted: sites

Jing Leng 7/8/14 6:22 PM
Deleted: elements

Materials and Methods

Lineage-specific sites: H3K27ac peaks exhibiting lineage-specific increase of signal in the human embryonic limb compared to rhesus and mouse (Cotney et al., 2013). In this dataset, four sets of lineage-specific sites were identified, representing four time points sampled in human. Four sets of all H3K27ac peaks called at individual time points in human from this study were also obtained. Note that "lineage-specific" in this context does not mean that the H3K27ac enriched regions in human do not have corresponding DNA sequence in the other two species. Rather, it means that the orthologous sequences are present in all three species, but in human they have higher level of H3K27ac ChIP-Seq signal.

Jing Leng 7/8/14 6:22 PM
Deleted: regions

Ancestry-specific sites: H3K27ac peaks displaying ancestry-specific signal in the lymphoblastoid cell lines across nineteen individuals in four human populations (Kasowski et al., 2013). All H3K27ac peaks called in human LCLs from this study were also obtained. Note that "ancestry-specific" sites in this context only indicate the differential H3K27ac signal, not the presence or absence of the DNA sequences.

Jing Leng 7/8/14 6:22 PM
Deleted: regions

Roadmap Epigenomics project data: H3K27ac ChIP-Seq data in adult human primary tissue samples, including adipose tissue, adrenal gland, aorta, bladder, brain (anterior

Jing Leng 7/8/14 6:22 PM
Deleted: Project

Jing Leng 7/8/14 6:22 PM
Deleted: : H3K27ac peaks

caudate, cingulate gyrus, hippocampus middle, inferior temporal lobe, mid frontal lobe), duodenum smooth muscle, esophagus, gastric, left ventricle, lung, ovary, pancreas, psoas muscle, right atrium, right ventricle, sigmoid colon, small intestine, spleen, thymus. The peaks were called using the same methods as that in Cotney et al. In brief, ChIP-seq reads were aligned to hg19 using Bowtie, retaining uniquely mapped reads while also discarding duplicated reads (same strand and same start site). A sliding window of 500 bp with a 25 bp step-size was used to scan the genome. In each window, the numbers of reads from H3K27ac as well as input control experiments were counted. Raw read number in the input control experiment was scaled to match the sequencing depth of each H3K27ac experiment. The number of expected reads in a window based on uniform distribution of total mapped reads in H3K27ac ChIP-Seq along a specific chromosome was calculated. Significance of enrichment in each window was calculated using a Poisson model, in which the null model is a larger number of input control read counts or expected read counts. Significantly enriched windows (p value $\leq 10^{-5}$) within 1kb distance were merged into a single region.

Enhancer-promoter connectivity information: Non-promoter and promoter connection of DNaseI hypersensitivity sites (DHSs), based on the correlation (no less than Pearson correlation coefficient of 0.7) of DNaseI signal across 79 cell types (Thurman et al., 2012). [[Intend to not include Yip et al., will explain why]]

Conservation: PhastCons scores across placental mammals for the hg19 assembly were downloaded from the UCSC Genome Browser and used to assess cross-species conservation. Human population variation data from 1000 Genomes Project phase 1 release were used to evaluate within-human conservation. Only SNP variation data in low coverage samples were used and heterozygosity was calculated as 2*AF*(1-AF) (AF is allele frequency).

Motif content: 532 transcription factor (TF) motifs were obtained from the JASPAR database. Motif occurrences in the hg19 genome were identified using FIMO (p value $\leq 10^{-5}$).

CTCF and cohesin ChIP-Seq: CTCF and cohesion ChIP-Seq peaks in the mouse embryonic limb were obtained from (DeMare et al., 2013), those in the human LCL were obtained from ENCODE project GM12878 experiments (Consortium, 2012).

eQTL data in human LCLs: *cis*-eQTL sites were obtained from GEUVADIS project (Lappalainen et al., 2013).

Jing Leng 7/8/14 6:22 PM
Deleted: **Enhancer-promoter connections**: Non-promoter and promoter connection of DNAse hypersensitivity sites (DHSs), based on the correlation of signal across XX tissues [[cite paper from Stam lab]]. ... [1]

Results

We focus on the contrast between the lineage- and ancestry-specific sites, or variable sites, versus the rest of the elements in Cotney et al. or Kasowski et al respectively.

Jing Leng 7/8/14 6:22 PM
Deleted: respective datasets

Jing Leng 7/8/14 6:22 PM
Deleted: of variable sites

Tissue-specificity

Using H3K27ac peaks called in the selected adults tissues in Roadmap Epigenomics Project, we examined in how many tissues a peak in the Cotney or the Kasowski dataset was also called a peak in the Epigenomics data. The fewer the number of tissues in which a peak was called in the Epigenomics data, the more tissue-specific it is, and therefore the less pleiotropic it potentially is. We observed that the variable sites (both distal and proximal) show a spectrum of tissue-specificity, indicating their heterogeneity in terms of potential pleiotropy. We also observed that the variable sites are more tissue-specific than the other sites (Figure 1), implying that the variable sites possess lower level of

Jing Leng 7/8/14 6:22 PM
Deleted: are more tissue-specific than the other sites (Figure 1), implying that the variable sites have lower level of pleiotropy

pleiotropic effects. We note that in this analysis, each tissue was counted as one, regardless of their similarity between each other. We show that the result remain robust if we merge closely related tissues into one [[sup figure]]. We note that the raw numbers may vary depending on the definition of tissue or tissue types as well as how finely dissected the raw materials are. Nevertheless, the contrast of distribution for variable sites versus other sites should not be affected.

Promoter targeting

Using the connectivity information between distal and proximal DHSs (Thurman et al., 2012), we studied the number of candidate target promoters of distal regions. The fewer the number of promoters that are connected to a putative enhancer, the less pleiotropic the enhancer may be. We observed that the distal variable sites on average are connected to fewer promoters than the other sites (Figure 2), suggesting that the variable sites possess lower level of pleiotropic effects.

Conservation and other properties

Variable sites versus other sites: lower level of conservation, lower GC content, lower motif content, lower (but not all significant) CTCF/cohesin binding. For ancestry-specific sites, they overlap fewer eQTLs, but is not significant. [[need to expand, XJM]]

Discussion

The evo-devo field has traditionally pictured *cis*-regulatory elements as being very modular (Carroll et al., 2009), with several distinct elements controlling the tissue- or stage-specific expression of a developmental gene. This is an attractive feature for evolution, since the changes in CREs for a gene could introduce morphological changes without influencing the other functions of the same gene. As such, *cis*-regulatory changes have been proposed as the primary driver during organismal evolution (Carroll et al., 2009). However, some CREs could be functional in multiple contexts, and confer pleiotropic effects when mutated. It is thus natural to ask how pleiotropic are CREs, and how their level of pleiotropy is related to their evolutionary dynamics.

The genome-wide assessments of pleiotropy and evolution of CREs are difficult, and we have just begun to gain more systematic understanding through high-throughput experiments such as ChIP-Seq. Though ChIP-Seq does not yield readouts of "function" in the very strict sense, it has been shown to correlate well with other function-based assays, especially when the H3K27ac histone modification was used as a ChIP target (Cotney et al., 2012; Cotney et al., 2013; Rada-Iglesias et al., 2011). In this study, we used various H3K27ac ChIP-Seq dataset as a proxy for CRE usage. H3K27ac sites were considered candidate enhancers and promoters active in the specific biological system of choice, and variable sites identified through signal comparison were considered elements likely changed in their activity or function (Cotney et al., 2013; Kasowski et al., 2013). To measure the tissue-specificity of putative enhancers and promoters identified in Cotney et al. and Kasowski et al., we used H3K27ac ChIP-Seq data from the Roadmap Epigenomics project, and considered putative CREs overlapping peaks in Roadmap Epigenomics tissues as being active in those tissues, thus likely possessing functions in them as well. Likewise, when estimating the interaction of enhancers and promoters, DNaseI-based connectivity information has its caveats, though it shows good correlation

Jing Leng 7/8/14 6:22 PM
Deleted: . ... [2]

Jing Leng 7/8/14 6:22 PM
Deleted: cohesion

Jing Leng 7/8/14 6:22 PM
Deleted: of CREs

Jing Leng 7/8/14 6:22 PM
Deleted: assessment

Jing Leng 7/8/14 6:22 PM
Deleted: both

Jing Leng 7/8/14 6:22 PM
Deleted: is

Jing Leng 7/8/14 6:22 PM
Deleted: terms of

Jing Leng 7/8/14 6:22 PM
Deleted: , we used H3K27ac ChIP-Seq data from the Epigenomics projects. Likewise, when estimating the interaction of enhancers and promoters, DHS-based connectivity information has its caveats.

with assays such as 5C and ChIA-PET (Thurman et al., 2012). We recognize that there is still a long way to go from these biochemical activity-based assays to demonstrating the functional impacts of those putative elements on phenotypic traits in an organism. Nonetheless, given that comparative, genome-wide "function-based" data are not available, our study presents a valuable attempt to explore the relationship between pleiotropy and evolution of *cis*-regulatory activities.

We note that the small sample sizes of both Cotney et al. and Kasowski et al., as well as the fact that each of the studies included only one biological context (embryonic limb or LCL) limit our ability to make more generalized conclusions about the results presented in this paper…

An earlier study reported a positive correlation between histone modification conservation between human and mouse and its stability across human cell lines (Woo and Li, 2012) …

Jing Leng 7/8/14 6:22 PM
Deleted: assays

Jing Leng 7/8/14 6:22 PM
Deleted: nor will be available in the near future,

Jing Leng 7/8/14 6:22 PM
Deleted: provides

References

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321-1325.

Carroll, S.B., Grenier, J., and Weatherbee, S. (2009). From DNA to diversity: molecular genetics and the evolution of animal design (John Wiley & Sons).

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Cotney, J., Leng, J., Oh, S., DeMare, L.E., Reilly, S.K., Gerstein, M.B., and Noonan, J.P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. Genome research *22*, 1069-1080.

Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell *154*, 185-196.

DeMare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. Genome research *23*, 1224-1234.

Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J.B., Kundaje, A., Liu, Y., Boyle, A.P., Zhang, Q.C., Zakharia, F., and Spacek, D.V. (2013). Extensive variation in chromatin states across humans. Science *342*, 750-752.

Lappalainen, T., Sammeth, M., Friedländer, M.R., Ac't Hoen, P., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., and Ferreira, P.G. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature.

Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., FitzPatrick, D.R., and Afzal, V. (2008). Human-specific gain of function in a developmental enhancer. Science *321*, 1346-1350.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., and Vernot, B. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.
Wagner, G.P., and Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. Nature Reviews Genetics *12*, 204-213.
Woo, Y.H., and Li, W.-H. (2012). Evolutionary conservation of histone modifications in mammals. Molecular biology and evolution *29*, 1757-1767.