Compound cis-regulatory elements with both boundary and enhancer sequences in the human genome

Daudi Jjingo¹, Jianrong Wang¹, Andrew B. Conley¹, Victoria V. Lunyak²,³ and I. King Jordan¹,3,*

¹School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA, ²Buck Institute for Research on Aging, Novato, CA 94945, USA and ³PanAmerican Bioinformatics Institute, Santa Marta, Magdalena, Colombia

ABSTRACT

Motivation: It has been suggested that presumably distinct classes of genomic regulatory elements may actually share common sets of features and mechanisms. However, there has been no genome-wide assessment of the prevalence of this phenomenon.

Results: To evaluate this possibility, we performed a bioinformatic screen for the existence of compound regulatory elements in the human genome. We identified numerous such colocated boundary and enhancer elements from human CD4⁺ T cells. We report evidence that such compound regulatory elements possess unique chromatin features and facilitate cell type-specific functions related to inflammation and immune response in CD4⁺ T cells.

Contact: king.jordan@biology.gatech.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on May 10, 2013; revised on August 26, 2013; accepted on September 15, 2013

1 INTRODUCTION

Several types of cis-regulatory elements have been identified and classified. They include well-defined elements such as transcriptional promoters (Goldberg, 1979), enhancers (Banerji et al., 1981), boundary elements (Udvardy et al., 1985) and enhancer-blocking insulators (Kellum and Schedl, 1991). They also include the less discernible elements like silencers (Laimins et al., 1986), promoter-tethering elements (Calhoun et al., 2002) and locus control regions (Grosveld et al., 1987).

Among all cis-regulatory elements, enhancers exhibit the highest flexibility and modularity (Lomvardas et al., 2006; Shen et al., 2012). Mechanistically, enhancers recruit transcription factors, which can loop over long genomic distances to reach promoters, thereby giving enhancers the ability to influence the expression of distal genes. This long-range capacity of enhancers can, however, be inhibited by boundary elements, particularly enhancer-blocking insulators (Kellum and Schedl, 1991). This boundary element insulating activity protects genes in domains located on the active sides of boundaries against activating or repressive regulatory effects of both flanking and distant domains. In this way, enhancer-blocking insulators play a critical role in facilitating the specificity of interactions between enhancers and genes located in the same chromosomal domains (West and Fraser, 2005). As such, boundaries and enhancers have hitherto been considered to be functionally antagonistic, and thus to occupy distinct and separate loci in the genome. To date no genomic loci have been reported to simultaneously encode the functional capacities of both enhancers and boundaries on a genome-wide scale.

Nevertheless, it has previously been suggested that boundaries and enhancers might actually use a common set of regulatory features and strategies, and more generally, that many of the accepted distinctions between classes of regulatory elements may be overstated (Gaszner and Felsenfeld, 2006). Considering this possibility, together with the coordinated regulatory activities of boundaries and enhancers, we sought to evaluate whether there actually exist colocated compound boundary-enhancer loci in the human genome. We found that numerous compound boundary-enhancer loci do in fact exist in the human genome, and have epigenetic and regulatory features that are distinct from those seen for individual regulatory elements of either class.

2 METHODS

2.1 Boundaries, enhancers and compound elements

Sets of boundary elements (n = 2542) (Wang et al., 2012) and enhancers (n = 23 574) (Fernandez and Miranda-Saavedra, 2012) previously predicted to function in CD4⁺ T cells were mapped to the human genome reference sequence (hg18). Compound regulatory elements (designated B+E) were defined as boundary regions (~8 kb) that contain or overlap with enhancer elements (~1 kb), and the locations of canonical non-compound boundaries (designated B−E) and solitary enhancers (designated E−B) were retained for comparison. A binomial test of enrichment was performed to check for statistical overrepresentation of enhancers within boundary elements compared with their background genomic frequency.

2.2 Chromatin analysis

Four genome-wide functional genomic datasets for CD4⁺ T cells were analyzed to characterize the chromatin environment in-and-around the regulatory elements studied here. These include ChIP-seq generated genomic distributions for eight different histone modifications (Barski et al., 2007), genomic sites for 95 710 DNase I hypersensitive (DHS) sites (Boyle et al., 2008), ChIP-seq generated genomic locations of ~2 million Pol II binding sites (Barski et al., 2007) and ~8.3 million RNA-seq tags (Barski et al., 2010). For each of these datasets, fold-enrichment plots in-and-around regulatory elements were computed and normalized using genomic background averages. For histone marks, the combined histone
fold-enrichment index was computed as the sum of fold enrichments for individual marks (Supplementary Fig. S1 and Supplementary Methods). Differences in fold enrichment for compound and non-compound regulatory elements were evaluated using paired Mann–Whitney U-tests.

2.3 Gene expression analysis
Gene expression in 79 human tissues was computed for RefSeq gene annotations as previously described (Jjingo et al., 2011) using the Novartis Expression Atlas (Su et al., 2004). Gene expression levels were compared between B+E, B−E and E−B elements using Mann–Whitney U-tests.

2.4 Gene set enrichment analysis
For gene set enrichment analysis, we evaluated the distribution of functionally coherent sets of genes, as defined by shared Gene Ontology (GO) categories or presence in the same Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, around compound (B+E) versus non-compound (B−E and E−B) boundary elements. The hypergeometric test was used to evaluate the significance of the enrichment of genes within a defined functional group around sets of regulatory elements.

3 RESULTS AND DISCUSSION

3.1 Compound regulatory element discovery approach
We evaluated the existence of compound cis-regulatory elements (designated as B+E) in the human genome by searching for genomic loci that are predicted to function simultaneously as both boundary elements and enhancers (Fig. 1A). Analyses of the genomic distributions of histone modifications have led to the discovery of characteristic patterns at several genomic regulatory features like boundary elements (Wang et al., 2012) and enhancers (Birney et al., 2007; Visel et al., 2009). These regulatory element-specific histone modification profiles have been used to develop algorithms that can accurately predict regulatory elements from genome-wide ChIP-seq datasets. For example, ChIP-seq data for histone modifications and RNA Pol II-binding have been used to perform a genome-wide prediction of human chromatin boundaries (Wang et al., 2012). Likewise, enhancers have been predicted in several human cell lines (Heintzman et al., 2009). We analyzed the locations of boundaries and enhancers predicted in this way for human CD4+ T cells. There are 2542 predicted boundary elements (Wang et al., 2012) and 23574 predicted enhancers (Fernandez and Miranda-Saavedra, 2012).

3.2 Enrichment of compound boundary-enhancer elements in the human genome
We intersected the human genome coordinates of predicted boundary elements with those of predicted enhancers and found 690 genomic locations with colocated boundary and enhancer annotations (Fig. 1B). These compound regulatory elements represent ~27% of all boundary elements in our dataset. The boundary element predictions used here cover broader genomic regions (8 kb) than the enhancer predictions (1 kb); thus, compound boundary elements may be colocated with multiple enhancers. We compared the observed occurrence of B+E elements against their expected level of occurrence, based on the background genomic frequencies of the individual element classes, to ensure that their presence could not be attributed to chance alone. A binomial test of enrichment revealed predicted enhancers to be significantly enriched within predicted boundary elements relative to their genomic background frequency (Z = 5.65, P = 1.6 e-08); there are 6.6-fold more predicted enhancers occurring in boundaries than can be expected by chance alone. B+E elements are much more enriched around clusters of enhancers than around enhancers outside boundaries (Supplementary Fig. 1E).

This overrepresentation of enhancers within predicted boundary regions is surprising, as boundaries have only been known to have a presumably antagonistic enhancer-blocking activity (Kellum and Schedl, 1991). Conversely, this finding supports the proposition that classes of regulatory elements typically considered to be distinct actually share sets of features and mechanisms (Gaszner and Felsenfeld, 2006). In any case, the enrichment of enhancers within predicted boundary element regions suggests an important functional role for these compound regulatory elements.

Fig. 1. Compound regulatory elements and their features in the human genome. (A) A compound regulatory element possessing both boundary (blue) and enhancer (red) sequences. (B) Overlap between predicted enhancers and boundaries. (C, D) Enrichment profiles and average fold enrichments for histone modifications and Pol2 binding in-and-around boundary elements (blue lines). (E) Average gene expression for genes proximal (within 10 kb) of boundary elements in CD4+ T cells (gray) and 78 other tissues (white)
3.3 Compound boundary-enhancer elements possess unique regulatory features

The enrichment of enhancers within boundary element regions suggests that compound B+E regulatory elements represent a functionally distinct combination of their individual constituent elements. In that case, we expect to observe distinct regulatory features, e.g., chromatin and expression profiles, for compound B+E elements when compared with B−E and E−B elements. To test this prediction, we compared chromatin and expression profiles from CD4+ T cells for B+E elements versus B−E and E−B elements. This was done using ChIP-seq data for eight histone modifications (Wang et al., 2008) to evaluate the chromatin modification state, DHS site data to evaluate the openness of local chromatin and PolII and RNA-seq data to evaluate transcriptional states.

For each of these datasets, enrichment plots showing fold enrichment compared with genomic background levels were computed for 20 kb regions centered on B+E elements versus B−E and E−B elements (Fig. 1C and D and Supplementary Fig. S1). In addition, the overall average fold enrichment levels across these regions were determined (bar graphs in Fig. 1 and Supplementary Fig. S1). When considered jointly, the eight histone modifications show significantly higher enrichment for compound B+E regions than seen for B−E and E−B regions. These particular histone modifications (see Supplementary Fig. S1A and B) were chosen owing to their previously characterized associations with boundary elements and/or enhancers (Fernandez and Miranda-Saavedra, 2012; Heintzman et al., 2007; Heintzman et al., 2009). With respect to the individual histone modifications, 7 of 8 histone modifications, all of which are active modifications, show increased enrichment around the compound B+E elements (Supplementary Fig. S1B). The sole exception to this pattern is seen for the repressive modification H3K27me3. Furthermore, it can be seen that the overall levels of histone modifications are higher for the active side of the boundaries (boundary start position till +10 kb) than for the repressive side (−10 kb until boundary start position), and this effect is also more pronounced for compound B+E elements than B−E and E−B elements (Fig. 1C, bar graph).

When averaged across the open side of elements, similar patterns of greater B+E enrichment compared with B−E and E−B regions can be seen for Pol II binding data, DHS sites and RNA-seq data (bar graphs in Fig. 1D and Supplementary Fig. 1C and D). The RNA-seq data show a qualitatively distinct pattern compared with the other datasets with an extremely marked peak close to the start and center positions. This pattern could indicate that B+E elements most actively protect gene expression in their most proximal regions and could also point to a specific role for expression of non-coding RNAs in establishing boundary element and enhancer activity. Support for both of these possibilities has previously been reported (Lunyak et al., 2007; Wang et al., 2012).

It is, therefore, apparent that compound B+E elements possibly modulate chromatin structure and facilitate transcriptional changes in a more profound manner than do B−E and E−B elements.

3.4 Compound boundary-enhancer elements enhance cell type-specific gene expression

The distinct chromatin changes and higher transcriptional activity across B+E elements suggest that these regulatory elements may help to facilitate higher expression levels of proximal genes (within 10 kb) than B−E elements. Because enhancers boost gene expression levels, we expect their inclusion into boundary element regions to result in higher expression of nearby genes. To test this prediction, we compared the relative expression levels of genes on the active sides of B+E versus B−E and E−B elements. As expected, B+E elements yield average expression levels qualitatively similar to those of E−B and greater than B−E elements.

Furthermore, this effect can be seen to be cell type-specific (Fig. 1E), as the expression is much more pronounced in the CD4+ T cells where the regulatory elements were predicted compared with a panel of 78 additional cell types and tissues (Fig. 1E). Additionally, B+E associated genes are significantly enriched for CD4+ T cell-specific expression compared with genes not associated with B+E elements (Supplementary Table S1).

3.5 Potential functional significance for compound boundary-enhancer elements

Gene set enrichment, based on GO and KEGG pathway annotations, was used to evaluate the potential functional significance of B+E elements for CD4+ T cells. To do this, the set of genes that lie within 10 kb of B+E elements were evaluated for evidence of coherent functional signatures that could be related to T cell-specific or immune-related function. This analysis revealed genes that are significantly enriched around B+E elements and encode proteins with functions that are directly relevant to CD4+ T cell activity; these genes involved in the chemokine signaling pathway (GO:007098).

Chemotaxis, growth, differentiation and apoptosis of inflammatory cells, like T-lymphocytes and eosinophils, are achieved via the chemokine signaling pathway, which is largely dependent on the activation of PIK3 kinases (Britten 2013; Curnock et al., 2002; Klarenbeek et al., 2013). Chemokine signaling pathway genes are enriched around compound B+E elements (hypergeometric test; \( P = 9.0 \times 10^{-13} \)), compared with B−E boundaries (\( P = 0.0055 \)) and E−B (\( P = 5.4 \times 10^{-9} \)). Additionally, pathway genes proximal to B+E elements are expressed at levels comparable with those of solitary enhancers and higher than B−E elements. (Fig. 2A and B and Supplementary Fig. S2). This is specifically exemplified by the PIK3 gene, which is functionally central to the chemokine signaling pathway (Fig. 2D). PIK3 is expressed at higher levels in CD4+ T cells (\( SI = 3463 \)) relative to other human tissues (average SI = 755), and there are two B+E compound elements that can be seen to flank the gene thus helping to maintain its relatively open chromatin environment (Fig. 2C).

4 CONCLUSIONS

This study supports the existence of compound regulatory elements that encode both boundary and enhancer activities with relevance to T cell-specific functions. These findings are consistent with previous postulations of overlap between regulatory
elements, suggesting that regulatory elements from different classes can share mechanistic features, modes of action and even location.

Funding: Fulbright foundation [to D.J.]; The School of Biology, Georgia Institute of Technology [to I.K.J., D.J., J.W. and A.B.C.]; Alfred P. Sloan Research Fellowship [to I.K.J.]. The authors are also grateful to the Encyclopedia of DNA Elements (ENCODE) consortium and acknowledge the use of its datasets.

Conflict of Interest: none declared.

REFERENCES