



OPEN

Connectome and regulatory hubs of CAGE highly active enhancers

Mewen Briend¹, Anne Rufiange¹, Louis-Hippolyte Minvielle Moncla¹, Samuel Mathieu¹, Yohan Bossé^{2,3} & Patrick Mathieu^{1,4}✉

Evidence indicates that enhancers are transcriptionally active. Herein, we investigated transcriptionally active enhancers by using cap analysis of gene expression (CAGE) combined with epigenetic marks and chromatin interactions. We identified CAGE-tag highly active (CHA) enhancers as distant regulatory elements with CAGE-tag \geq 90th percentile and overlapping with H3K27ac peaks (4.5% of enhancers). CHA enhancers were conserved between mouse and man and were independent from super-enhancers in predicting cell identity with lower P-values. CHA enhancers had increased open chromatin and a higher recruitment of cell-specific transcription factors as well as molecules involved in 3D genome interactions. HiChIP analysis of enhancer-promoter looping indicated that CHA enhancers had a higher density of anchor loops when compared to regular enhancers. A subset of CHA enhancers and promoters characterized by a high density of chromatin loops and forming hub regulatory units were connected to the promoter of immediate early response genes, genes involved in cancer and encoding for transcription factors. Promoter of genes within hub CHA regulatory units were less likely to be paused. CHA enhancers were enriched in gene variants associated with autoimmune disorders and had looping with causal candidate genes as revealed by Mendelian randomization. Hence, CHA enhancers form a dense hierarchical network of chromatin interactions between regulatory elements and genes involved in cell identity and disorders.

The regulation of gene expression relies on complex interactions between gene promoters and distant *cis*-acting elements often referred to as enhancers¹. Enhancers, which often reside several kilobases away from their target protein coding genes, are enriched in cell-specific transcription factor motifs². As such, enhancers participate in cell identity and developmental program³. The mechanism by which enhancers control the expression of genes involves the looping process, which allows two distant regions on the linear genome to be in close proximity⁴.

Studies have underlined that active enhancers are enriched in the histone mark H3K27ac, which is often used to assess the presence of distant-active regulatory elements⁵. H3K27ac mark, which is deposited by the acetyltransferase EP300⁶, allows the recruitment of transcriptional cofactors such as bromodomain containing protein BRD4⁷. Hence, acetylation of histone is an actively regulated process, which participates to the activity of regulatory elements. Distant-acting elements are associated with the recruitment RNA polymerase II (RNAPII), which produces enhancer-associated RNAs (eRNAs)^{8,9}. Based on cap analysis of gene expression (CAGE)¹⁰, the Functional Annotation of the Mammalian Genome (FANTOM5)¹¹ consortium has underlined that transcription is pervasive in the noncoding genome, often bidirectional and associated with the expression of protein coding genes. FANTOM5 has produced an exhaustive list of actively transcribed enhancer regions in different cell lines and primary cells^{12,13}.

According to the number of connections, enhancers may exert hierarchical activity and establish enhancer-promoter networks^{14,15}. Several processes are likely involved in the control of looping. The loop extrusion model in which CTCF and the cohesin complex interact explains how topologically associated domains (TADs) are established^{16,17}. Also, recruitment of transcription factors and RNA polymerase II (RNAPII) are significant features that participate to the establishment of intra-domain interactions between enhancers and promoters¹⁸. Herein, in H3K27ac defined enhancers we leveraged CAGE analysis to assess: (1) the relationship between enhancer-associated CAGE activity and the chromatin landscape at regulatory elements, (2) the association of enhancer-associated CAGE activity with 3D genome organization, (3) whether looping between enhancer-associated CAGE activity and promoter defines a subset of regulatory units (i.e. loop between a promoter and its enhancers) and (4) whether enhancer-associated CAGE activity is informative for the mapping of

¹Genomic Medicine Laboratory, Quebec Heart and Lung Institute, Laval University, Quebec, Canada. ²Quebec Heart and Lung Institute, Laval University, Quebec, Canada. ³Department of Molecular Medicine, Laval University, Quebec, Canada. ⁴Institut de Cardiologie et de Pneumologie de Québec/Québec Heart and Lung Institute, 2725 Chemin Ste-Foy, Québec, Québec G1V-4G5, Canada. ✉email: patrick.mathieu@fmed.ulaval.ca

disease-associated gene variants to target causal genes. We underlined that a small proportion of enhancers, referred to as CAGE-tag highly active (CHA) enhancers, were associated with the recruitment of cell-specific transcription factors, cell function and a high density of enhancer-promoter contacts. CHA enhancers were also enriched in gene variants associated with autoimmune disorders and had looping with causal candidate genes. A subset of highly 3D connected CHA enhancer-promoter units were associated with transcription elongation at protein coding genes involved in key cellular functions. Our findings establish a distinct role for CHA enhancers in chromatin organization and in enhancer-promoter networks with relevance for cell function and disease.

Results

Transcriptionally active enhancers. We first set out to identify distant regulatory elements in blood B cell line (GM12878) by using H3K27ac (data from ENCODE)^{19–21}. ChIP-seq defined enhancers were identified from H3K27ac peaks and chromatin active regions overlapping gene promoters were excluded (“Methods”). In total, 26,409 regions were identified as enhancers. We next leveraged CAGE data from the FANTOM5 consortium to assess the overlap of transcriptionally active regions with H3K27ac-defined enhancers. H3K27ac is required for the transcriptional activity of enhancer RNA (eRNA)²². Accordingly, in the 26,409 enhancers previously identified, CAGE-tag correlated positively with H3K27ac (spearman correlation = 0.35) (Supplementary Fig. 1). Among the H3K27ac enhancers, those overlapping with active CAGE regions had much higher open chromatin identified by transposase accessible chromatin and high-throughput sequencing (ATAC-seq) assay (Fig. 1a). According to the level of CAGE activity, we found that higher transcription (based on CAGE percentile) at H3K27ac defined enhancers was associated with elevated open chromatin (Supplementary Fig. 2). Considering the previous finding, we defined CAGE-tag highly active (CHA) enhancers as those with CAGE-tag activity \geq 90th percentile, which overlapped H3K27ac peaks (Fig. 1b). The intersection of CAGE-tag \geq 90th percentile with H3K27ac peak regions identified 1197 (1197/26409, 4.5% of enhancers) CHA enhancers (Fig. 1c). CHA enhancers were compared to H3K27ac defined enhancers with CAGE-tag activity $<$ 90th percentile including regions without detectable CAGE-tag activity, which were referred to as regular enhancers. We found that 14% of CHA enhancers vs. 6% of regular enhancers overlapped with CpG islands (CGI) ($P=2.96E-22$, Fisher’s exact test) (Fig. 1d). Also, a higher proportion of CHA enhancers intersected with highly conserved noncoding elements between human and mouse (31% vs. 10% respectively for CHA and regular enhancers, $P=1.21E-80$, Fisher’s exact test) (Fig. 1e). As CHA enhancers were conserved, we hypothesized that these regions may confer cell-specific functions. Annotation of gene ontology (GO) showed an enrichment of lower P-values for cell-specific functions in CHA vs. regular enhancers. As super-enhancers (SEs), which are group of enhancers in close proximity, have been shown to be associated with cell identity³, we compared the overlap between CHA enhancers and SEs. By using H3K27ac we identified 615 SEs in GM12878. We found a poor overlap between the CHA enhancers and SEs (Jaccard index=0.03). Assessment of P-values for cell-specific functions in GO showed lower P-values in CHA enhancers vs. SEs (Fig. 1f). Taken together, the present findings suggest that CHA enhancers are largely independent from SEs and are associated with cell identity.

Chromatin landscape and transcription factors in CHA enhancers. We tested whether the occupancy by the RNA polymerase II (RNAPII) was increased at CHA enhancers. As expected, compared to regular enhancers, CHA enhancers displayed elevated RNAPII ChIP signal ($P < 2.2E-16$, Wilcoxon rank-sum test) (Fig. 2a). Signals for H3K27ac and ATAC-seq were also highly increased in CHA vs. regular enhancers (Fig. 2b,c). Analysis of ChIP-seq signal for EP300, a histone 3 acetyl transferase for lysine 27 (H3K27ac), revealed higher occupancy at CHA compared to regular enhancers ($P < 2.2E-16$, Wilcoxon rank-sum test) (Fig. 2d). Also, the ChIP-seq tag density for BRD4, a reader of histone acetylation and transcriptional cofactor, was much higher in CHA vs. regular enhancers ($P < 2.2E-16$, Wilcoxon rank-sum test) (Fig. 2e). A transcription factor motif analysis showed that several transcription factor motifs such as NF-kappa B, the interferon response factor family, JUN and BATF were enriched in CHA enhancers (Supplementary Table 1). Considering those results, we analyzed ChIP-seq for the transcription factor IRF4 and BATF. This analysis showed significantly elevated ChIP-seq signal for IRF4 and BATF in CHA vs. regular enhancers (Fig. 2f,g). Taken together, these findings highlight a higher recruitment of cell-specific transcription factors at CHA enhancers.

CHA enhancers are enriched in 3D genomic interactions. Considering that activity of enhancers could be linked to the looping process, we assessed the occupancy of proteins involved in 3D genome organization at CHA and regular enhancers. We documented the ChIP-signal for RAD21 and SMC3, which are members of the cohesin complex. The tag density signals for RAD21 and SMC3 were markedly elevated in CHA compared to regular enhancers ($P < 2.2E-16$, Wilcoxon rank-sum test) (Fig. 3a and b). These data thus suggested that CHA enhancers were associated with 3D genome organization. Enhancer-promoter interactions in H3K27ac-HiChIP were analyzed in order to establish genome-wide chromatin interactions of CHA and regular enhancers with the promoter of genes (GSE101498). Analysis of H3K27ac-HiChIP showed that 93% of 1D data overlapped with H3K27ac ChIP-seq data, indicating a good agreement to detect and enrich for active chromatin regions in 3D chromatin conformation assay. By using H3K27ac-HiChIP we identified TADs. The overlapping of enhancers with borders of TADs including a buffer of 10kb was similar for CHA (2.3%) vs. regular (1.9%) enhancers ($P=0.40$, Fisher’s exact test) (Fig. 3c). These data suggested that CHA enhancers were occupied by essential factors involved in higher order chromatin organization. However, CHA enhancers were not overrepresented at TAD borders. Hence, it is possible that CHA enhancers are not involved in the regulation of long-range looping at domain boundaries, but instead could be involved in enhancer-promoter contact within these domains.

We next analyzed significant loops and their relationships with CHA and regular enhancers. With a stringent loop calling (q -value $<$ $1E-06$), we detected 52,175 significant loops. We determined the loops involving

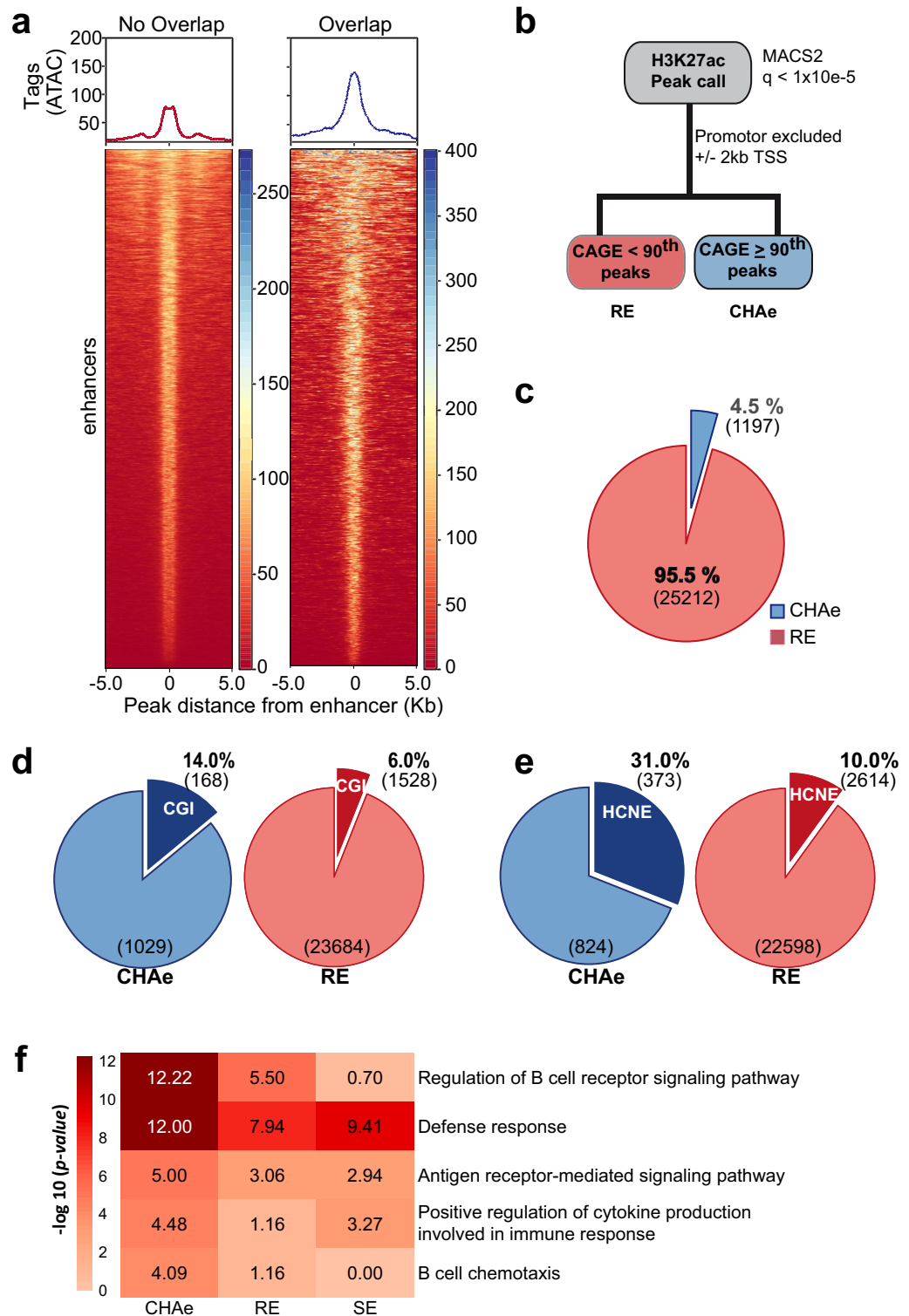


Figure 1. Characterisation of transcriptionally active enhancers. **(a)** Heat map representing ATAC activity (open chromatin) in H3K27ac-defined enhancers with or without overlap with nonzero CAGE-tag regions (deeptools, plotHeatMap). **(b)** Pipeline to identify CHAe and RE. **(c)** Proportions of CHAe and RE. **(d)** Proportions of CHAe and RE overlapping with CGI ($P=2.96E-22$, Fisher’s exact test). **(e)** Proportions of CHAe and RE overlapping with HCNE between human and mouse ($P=1.21E-80$, Fisher’s exact test). **(f)** GO enrichment of CHAe, RE and SE. CHAe CAGE-tag highly active enhancer, RE regular enhancer, SE super-enhancer, CGI CpG island, HCNE human conserved noncoding elements (between man and mouse), GO gene ontology.

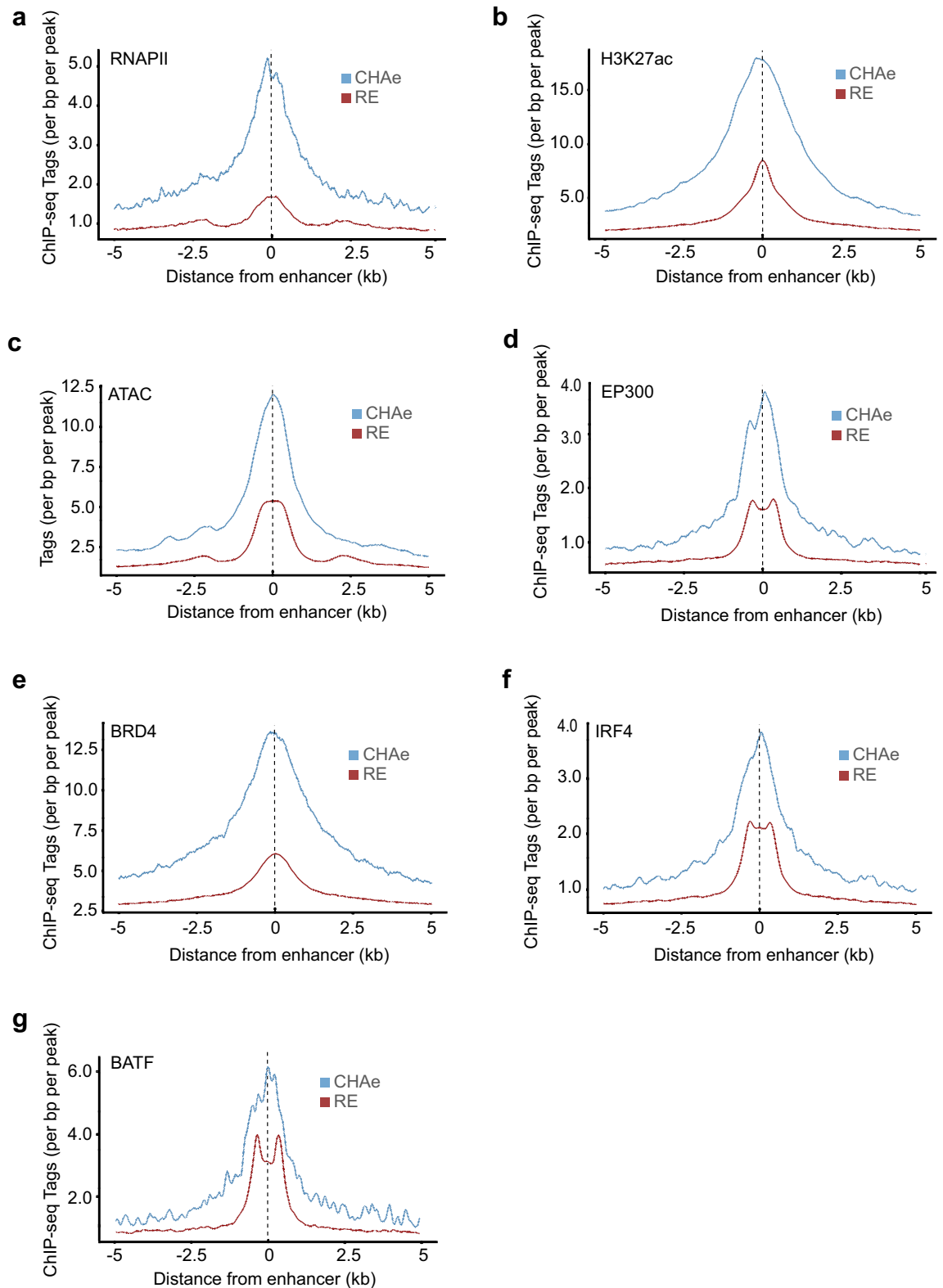


Figure 2. Chromatin and transcription factors in CHA and regular enhancers. Tag density plot centered on CHA and regular enhancers for (a) RNA polymerase II occupancy, (b) ChIP-seq H3K27ac, (c) transposase accessible chromatin, (d) EP300 occupancy, (e) BRD4 occupancy, (f) IRF4 occupancy and (g) BATF occupancy. (for all $P < 2.2E-16$, Wilcoxon rank-sum test). CHAe CAGE-tag highly active enhancer, RE regular enhancer.

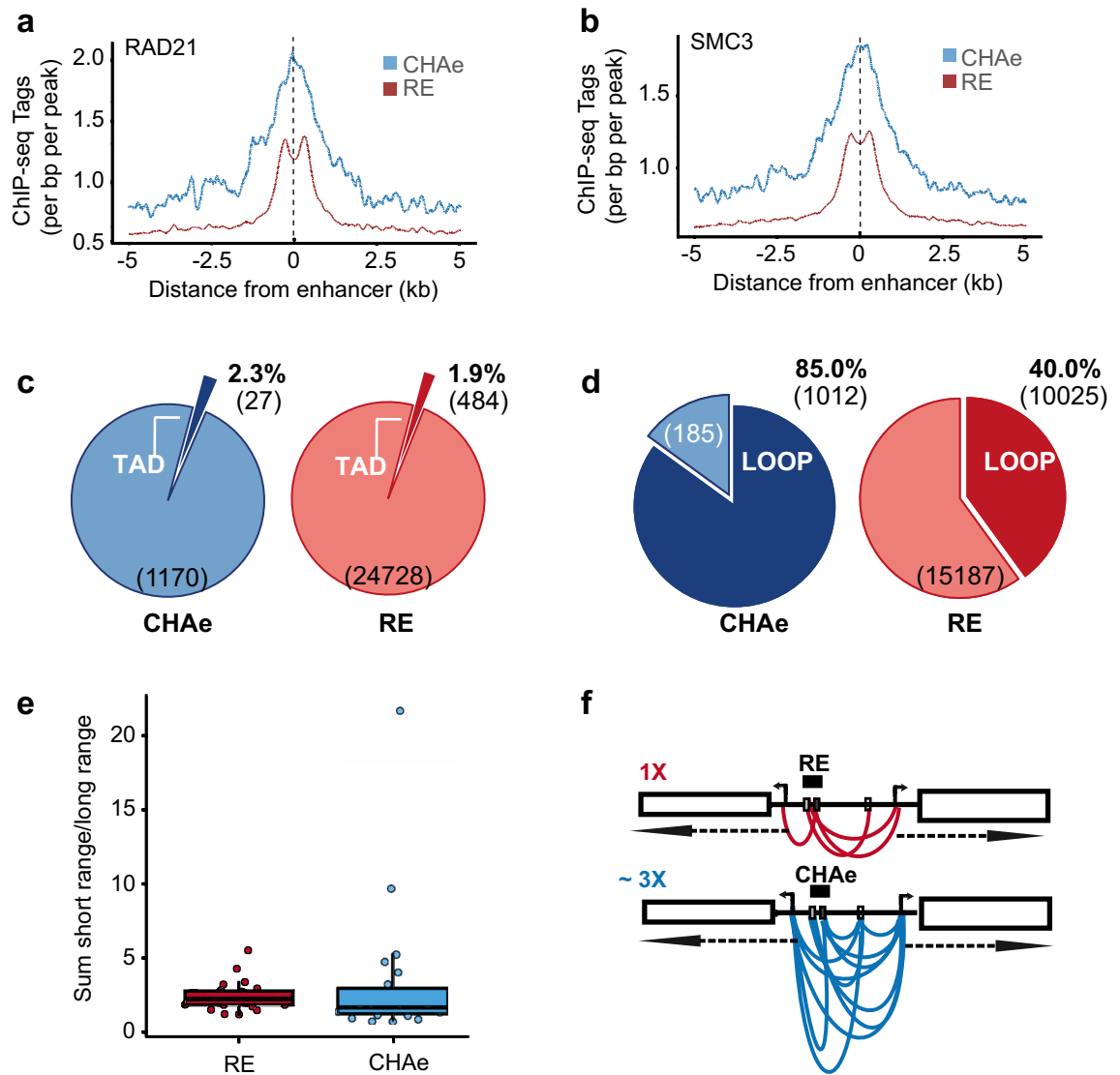


Figure 3. Connectome of CHA and regular enhancers. Tag density plot centered on CHA and regular enhancers for (a) RAD21 occupancy ($P < 2.2E-16$, Wilcoxon rank-sum test) and (b) SMC3 occupancy ($P < 2.2E-16$, Wilcoxon rank-sum test). (c) Proportion of enhancers overlapping with TAD ($P=0.40$, Fisher's exact test). (d) Proportions of enhancers overlapping with anchor loops in H3K27ac-HiChIP ($P=1.87E-214$, Fisher's exact test). (e) Box-plot representing ratio of the sum of short ($<200\text{kb}$) to long distance ($\geq 200\text{kb}$) interactions for each chromosome ($P=0.38$, Wilcoxon rank-sum test). (f) Schematic representation of loop density analysis. CHAe CAGE-tag highly active enhancer, RE regular enhancer, TAD topologically associated domain.

CHA and regular enhancers. The proportion of regulatory elements involved in at least one anchor loop was increased for CHA (85%, 1012/1197) vs. regular enhancers (40%, 10025/25212) ($P=1.87E-214$, Fisher's exact test) (Fig. 3d). The short ($<200\text{kb}$) vs. long distance chromatin interaction ratio calculated for each chromosome ("Methods") involving CHA and regular enhancers were similar (2.4 vs. 3.3, respectively for regular and CHA enhancers, $P=0.38$, Wilcoxon rank-sum test) (Fig. 3e). However, the mean number of loops per distant regulatory element was markedly increased in CHA versus regular enhancers (12.3 ± 12.7 vs. 4.5 ± 5.1 , $P < 2.2E-16$, Wilcoxon rank-sum test). A schematic representation of anchor loops in a typical CHA enhancer vs. a regular enhancer is depicted in Fig. 3f.

CHA enhancers in different cell populations are associated with cell-specific functions and open chromatin.

In order to assess whether the findings were generalizable, we also mapped CHA enhancers in human coronary artery smooth muscle cells (HCASMC) and normal human epidermal keratinocytes (NHEK). In HCASMC, we mapped 358 CHA enhancers which corresponds to 1.8% of H3K27ac-defined enhancers, whereas in NHEK we identified 852 CHA enhancers (2.5% of enhancers). We found a poor overlap between CHA enhancers and SEs in HCASMC (Jaccard = 0.003) and NHEK (Jaccard = 0.01). Compared to regular enhancers and SEs ($n = 1168$), the annotation of regulatory regions showed lower P values in CHA enhancers for HCASMC-specific functions such as muscle contraction, regulation of vasoconstriction and regulation of

vascular smooth muscle cell proliferation (Suppl. Fig. 3A). Compared to regular enhancers, CHA enhancers in HCASMC were enriched in open chromatin (Suppl. Fig. 3B). Also, occupancy by transcription factors such as JUN and TCF21, which are involved in the regulation of HCASMC function, was markedly increased in CHA versus regular enhancers (Suppl. Fig. 3C,D). Analysis of H3K27ac-HiChIP in HCASMC, showed a higher number of high-confidence anchor loops at CHA compared to regular enhancers (7.7 ± 8.1 loops per CHA enhancer vs. 4.5 ± 5.6 loops per regular enhancer, $P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test). Among the genes connected to CHA enhancers in HCASMC there are some key transcription factors such as *KLF3*, *SOX15* and *FOXP1*, which are involved myogenic cell fate^{23,24}. Consistently, several myogenic genes such as *MYH9*, *MYL12A* and *MYO1C* are also connected to CHA enhancers. The 852 CHA enhancers in NHEK were annotated and enriched for cell-specific function such as skin development, cornification and regulation of hair follicle development in comparison to regular enhancers and SEs ($n = 809$) (Suppl. Fig. 4A). Compared to regular enhancers, CHA enhancers from NHEK had increased open chromatin and occupancy for MYC, a transcription factor highly expressed in basal cell layer of epidermis and hair follicles (Suppl. Fig. 4B,C)²⁵. Also, the occupancy of GRHL3, a transcription factor involved in the differentiation of keratinocyte²⁶, was increased in CHA compared to regular enhancers (Suppl. Fig. 4D). Analysis of H3K27ac-HiChIP in HaCaT, a keratinocyte cell line, showed after loop calling a higher number of anchor loops involving CHA compared to regular enhancers (10.1 ± 9.1 loops per CHA enhancer vs. 6.1 ± 6.0 loops per regular enhancer, $P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test). In HaCaT, gene promoters connected to CHA enhancers include several genes involved in the cornification process such as genes located in the human keratin cluster locus at 12q13 (*KRT80*, *KRT7*, *KRT83*, *KRT84*, *KRT5*, *KRT72* and *KRT8*) as well as keratin-associated proteins genes (*KRTAP*) (Suppl. Fig. 5)²⁷. The Jaccard index for the CHA enhancers between GM12878 versus HCASMC (0.016), GM12878 versus NHEK (0.018) and HCASMC versus NHEK (0.069) showed that CHA enhancers between these three cell types were largely unique and cell type specific (Suppl. Fig. 6). Taken together, these data highlight that CHA enhancers are related to cell identity and have a high occupancy of transcription factors and are highly connected to gene promoters with cell-specific functions.

Hierarchy among the connectome of CHA enhancer-promoter. Enhancers interacting with gene promoters, which are part of the cell chromatin connectome, form dense networks of regulatory units. According to the level of connection, distant regulatory elements and their target gene promoters form hierarchical networks with functional relevance¹⁵. We hypothesized that among the gene promoters connected to CHA enhancers those with the highest level of contact with CHA elements formed regulatory hubs, which may include genes of significant relevance for cellular functions. We defined hub CHA regulatory units as enhancers and promoters with the highest number of CHA loops converging on promoters (≥ 90 th percentile), which were compared to non-hub CHA regulatory units (CHA enhancer-promoter loops < 90 th percentile). In total, hub CHA regulatory units mapped 89 unique highly connected gene promoters. Genes within hub CHA regulatory units were enriched in immediate early response genes (fold-enrichment: 8.7, $P = 9.07 \times 10^{-7}$, hypergeometric test), in genes involved in cancer (COSMIC database) (fold-enrichment: 3.6, $P = 0.0001$, hypergeometric test) and encoding for transcription factors as reported in the TF checkpoint database (fold-enrichment: 2.1, $P = 0.01$, hypergeometric test) (Fig. 4a). Among the different genes within hub CHA regulatory units, *MYC* and *IRF4* are unique as they are transcription factors, immediate early response genes and are involved in cancer (Fig. 4b). Figure 4c shows the locus at 8q24.21 where the promoter of *MYC* has many loops with highly interconnected CHA enhancers (CHA enhancers at ~ 550 – 160 kb from the promoter). One of the most connected protein coding gene in hub CHA regulatory units is *ETV6*, which encodes for a transcriptional repressor involved in hematopoiesis and leukemia²⁸. Figure 4d represents the locus at 12p13.2 where *ETV6* is densely connected to distant regulatory elements. Dynamic modelling of the enhancer-promoter 3D structure showed a multi loop aggregate forming a rosette-like structure at hub CHA regulatory unit for *ETV6*, which spans more than 500kb (Fig. 4e).

Hub CHA regulatory units and transcription elongation. We next hypothesized that hub CHA regulatory units include elements with increased active chromatin features. Among CHA enhancers, hub regulatory elements had increased open chromatin and occupancy by BRD4 (Fig. 5a,b). Considering the previous finding, we wondered whether gene promoters within hub CHA regulatory units may exhibit a higher recruitment of molecules involved in the transcription process. We thus assessed the recruitment of RNAPII phosphorylated on serine 2 (RNAPII-S2p), which is a marker of the elongation phase during the transcription process, at genes within hub CHA regulatory units²⁹. A metagene analysis revealed that the recruitment of RNAPII-S2p was increased at the transcription start site (TSS) and gene body of genes within hub CHA regulatory units vs. non-hub CHA regulatory units (Fig. 5c). To further assess if genes within hub CHA regulatory units were associated with increased processivity of the RNAPII, we analyzed global run-on sequencing (GRO-seq) data (GSE60454). Genes with nonzero tag density at the TSS (-50 bp and $+300$ bp) were considered for the analysis (“Methods”). We calculated the pause index, which is the ratio of RNAPII tag signal at the TSS to the gene body. The pause index was significantly increased at the promoter of genes connected to non-hub CHA vs. hub CHA regulatory units (median [95% CI]: 5.15 [4.61–5.78] vs. 2.96 [2.22–4.43], $P = 0.02$, Wilcoxon rank-sum test). When using a cut-off of 2 to assess paused genes (pause index > 2)³⁰, there was a trend for a higher proportion of paused genes in non-hub CHA vs. hub CHA regulatory units (73% vs. 64%, $P = 0.096$, Fisher’s exact test). These data suggested that genes within hub CHA regulatory units are less likely to be paused.

CHA enhancers and autoimmune disorders. Gene variants associated with complex trait disorders are enriched in the noncoding genome. We wondered whether autoimmune gene variants were enriched in CHA enhancers. Autoimmune variants prioritized from the Probabilistic Identification of Causal SNPs (PICS) were downloaded for the analysis (“Methods”) ^{31,32}. We found that PICS autoimmune variants resided at proximity

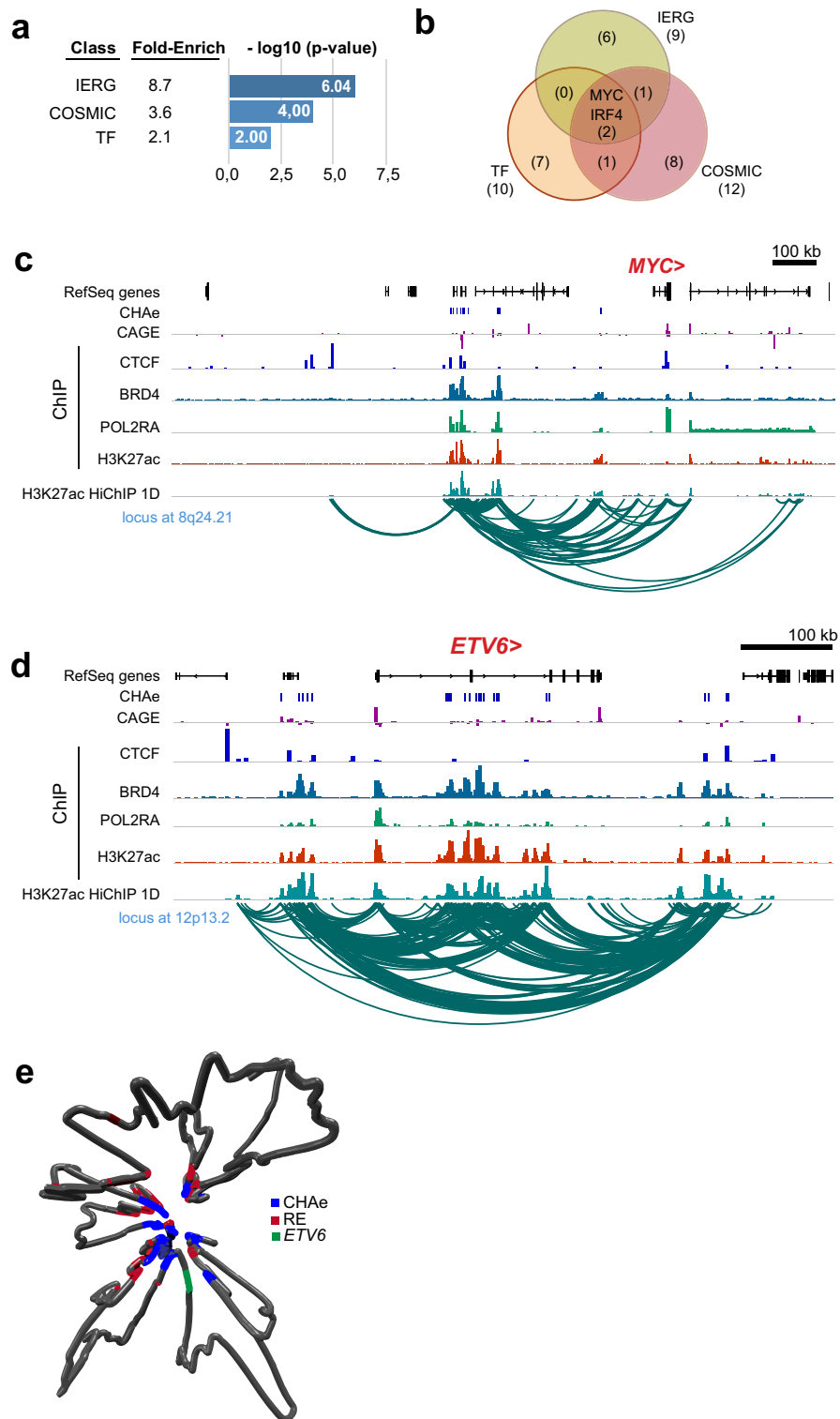


Figure 4. Characterization of hub and non-hub CHA regulatory units. **(a)** Enrichment of genes mapped by hub CHA regulatory unit. **(b)** Representation in Venn diagram of transcription factors (TF), genes involved in cancer (COSMIC) and IERG mapped by hub CHA regulatory unit. **(c,d)** Representative loci of hub CHA regulatory units; tracks represent genes, CHAe, CAGE, ChIP-seq CTCF, ChIP-seq BRD4, ChIP-seq POL2RA, ChIP-seq H3K27ac, H3K27ac-HiChIP 1D and arcs of significant loops; mapped genes are in red (Integrative Genomics Viewer). **(e)** 3D DNA modelling structure of hub CHA regulatory unit for *ETV6*; green: promoter region of *ETV6*; blue: CHAe; red: RE. CHAe CAGE-tag highly active enhancer, RE regular enhancer, IERG immediate early response gene, TF transcription factor.

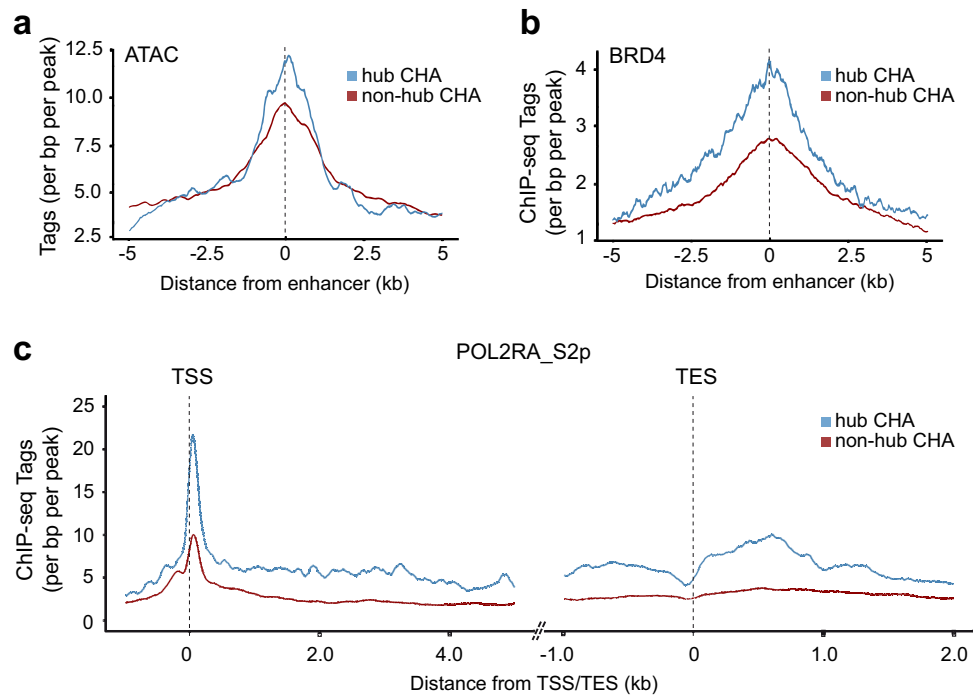


Figure 5. Chromatin landscape and recruitment of RNA polymerase at hub and non-hub CHA regulatory units. (a) Tag density plot of ATAC-seq in hub CHA versus non-hub CHA ($P=0.001$, Wilcoxon rank-sum test). (b) Tag density plot of ChIP-seq BRD4 in hub and non-hub CHA regulatory unit ($P < 2.2E-16$, Wilcoxon rank-sum test). (c) Metagenome analysis of RNAPII phosphorylated on serine 2 (POL2RA_S2p) in hub versus non-hub CHA regulatory unit. TSS transcription start site, TES transcription end sites ($P < 2.2E-16$, Wilcoxon rank-sum test).

($P < 2.2E-16$, relative distance, Kolmogorov-Smirnov test) and within CHA enhancers (fold-enrichment:10.4, $P < 2.2E-16$, binomial test) (Fig. 6a) (“Methods”). By using CHA enhancer-promoter looping, we mapped 138 autoimmune variants to gene promoters (Supplementary Table 2). For several variants, genes mapped by chromatin looping were upstream or downstream from the closest annotated gene to the SNP (Supplementary Table 2). We next ask whether genes mapped by CHA enhancer-promoter contacts may be causal in Mendelian Randomization. For each gene-disease pair mapped by CHA enhancer-promoter looping, we performed Mendelian randomization analysis for 4 major autoimmune disorders (rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease, type 1 diabetes) (“Methods”). Blood expression quantitative trait loci (eQTL) from eQTLGen, the largest blood eQTL resource including 31,684 samples, was leveraged to assess the exposition. In total, Mendelian randomization could be performed (independent instrumental variables ≥ 3 , P_{eQTL} for instrumental variables $< 5E-08$) for 40 CHA enhancer 3D mapped gene-disease pairs. We found that 33% (13/40) of CHA enhancer 3D mapped genes were significant causal candidates for 4 major autoimmune disorders ($P_{causal} < 0.05$) (Supplementary Table 3). For instance, PICS variants rs77013147 and rs12946510 at 17q12 are associated with several autoimmune disorders. Gene variants rs77013147 and rs12946510 are within CHA enhancers, which have chromatin loops with the promoter of *IKZF3* (Fig. 6b). The alleles T-rs12946510 ($P_{eQTL}=1.42E-243$) and C-rs77013147 ($P_{eQTL}=4.71E-190$) are associated with a higher blood expression of *IKZF3*. Motif analysis showed that the gene variant rs12946510 is within a transcription factor motif in which the risk allele T (freq EUR=0.47) (Supplementary Fig. 7) increases the binding for MEF2C (Fig. 6c), a transcription factor involved in B cell proliferation³³. This region is localized in a highly conserved noncoding region between human and mouse (Fig. 6b). *IKZF3* encodes for an IKAROS family zinc finger 3 protein, which is involved in lymphocyte proliferation³⁴. In Mendelian randomization, the blood expression of *IKZF3* was positively associated with the risk of systemic lupus erythematosus (beta=0.41, SE=0.05, $P_{IVW}=1.01E-16$) (Fig. 6d). At 21q21.2, PICS variant A-rs28735854 (freq EUR=0.15) is associated with rheumatoid arthritis. At this risk locus, intergenic variant rs28735854 is within a CHA enhancer having loops with the promoters of *IFNGR2* and *IFNAR2* (Fig. 6e). In Mendelian randomization, the expression of *IFNGR2* (beta=-0.12, SE=0.02, $P_{IVW}=7.78E-12$) and *IFNAR2* (beta=-0.08, SE=0.03, $P_{IVW}=0.004$) in blood cells were negatively associated with the risk of rheumatoid arthritis.

Discussion

In the present work, we underlined that enhancers with a high level of transcriptional activity as determined by CAGE, referred to as CHA enhancers, were associated with active chromatin features and the recruitment of cell-specific transcription factors. We underlined novel findings where CHA enhancers were cell-specific and

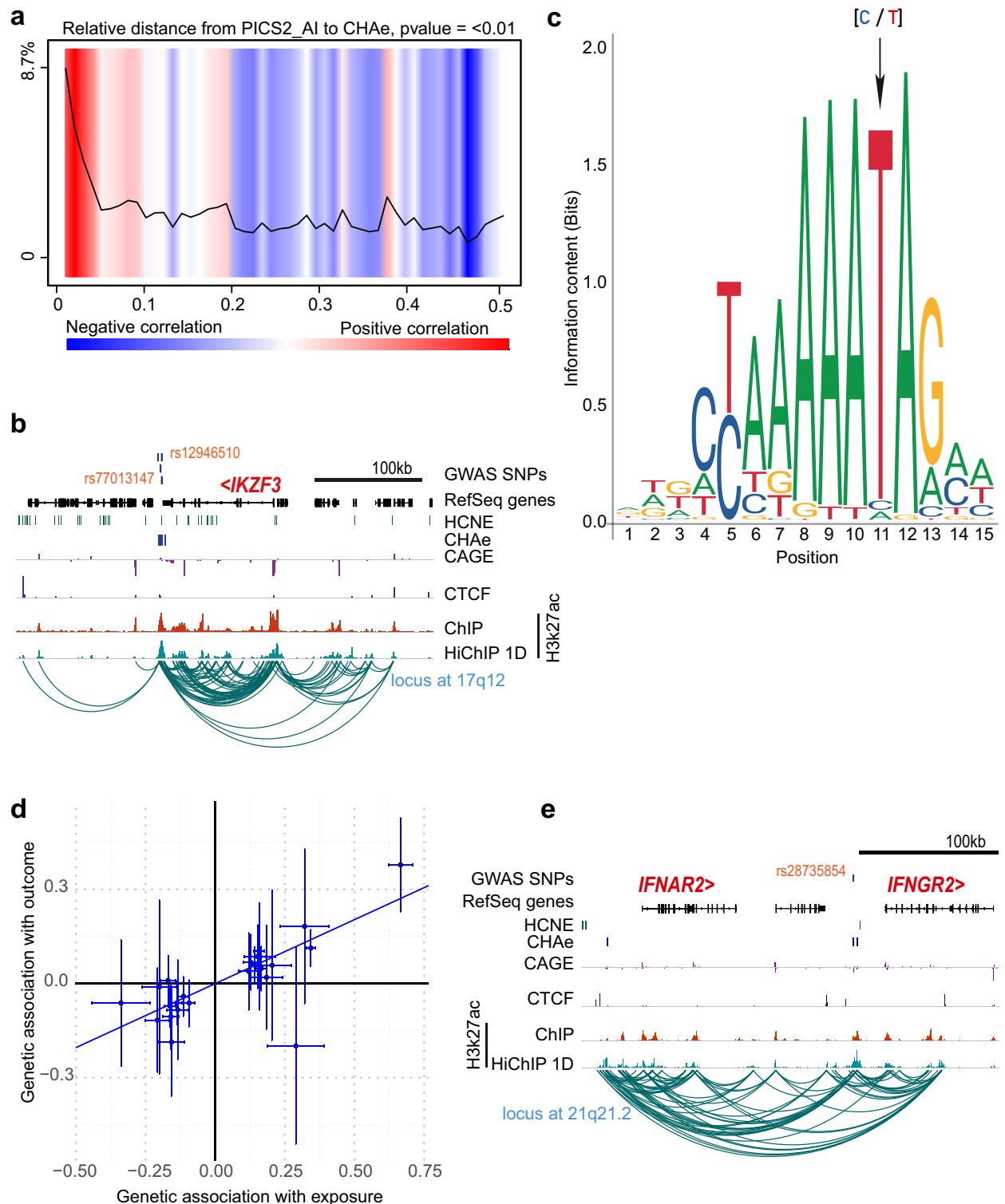


Figure 6. Relationships between CHA enhancers and gene variants for autoimmune diseases. **(a)** Heat map representing the relative distance between Probabilistic Identification of Causal SNPs (PICS) autoimmune variants and CHA enhancers ($P < 2.2E-16$, Kolmogorov-Smirnov test) (GenometriCorr). **(b)** *IKZF3* risk locus for lupus erythematosus; tracks represent SNPs, genes, HCNE, CHA enhancers, CAGE, ChIP-seq CTCF, ChIP-seq H3K27ac, H3K27ac-HiChIP 1D and arcs of significant loops; mapped genes are in red. **(c)** Sequence logo for MEF2C; arrow indicates rs12946510 base position where the reference allele is C and the risk allele is T. **(d)** Inverse variance weighted MR for *IKZF3* in systemic lupus erythematosus. **(e)** *IFNAR2/IFNGR2* risk locus for rheumatoid arthritis; tracks represent SNPs, genes, HCNE, CHA enhancers, CAGE, ChIP-seq CTCF, ChIP-seq H3K27ac, H3K27ac-HiChIP 1D and arcs of significant loops; mapped genes are in red. CHAe CAGE-tag highly active enhancer, HCNE human conserved noncoding elements.

did not significantly overlap with SEs. Compared to regular enhancers and SEs, CHA enhancers had lower GO P-values for cell identity functions. Hub CHA regulatory units established a dense network of enhancer-promoter interactions and were connected to immediate early response genes and transcription factors. Genes within hub CHA regulatory units had higher transcription elongation. CHA enhancers were enriched in autoimmune-associated gene variants prioritized by PICS and enhancer-promoter mapping identified causal candidate genes. Overall, these data underlined that highly transcribed enhancers establish a dense connectome with regulatory activity and having implications on cellular functions and disorders.

We observed that CHA enhancers were conserved and enriched in CGI. These findings suggested that CHA enhancers may exert a control over key regulatory genes. To this effect, we found lower P-values in cell-specific GO for CHA vs. regular enhancers and SEs. CGI are enriched at gene promoters and have been shown to support transcriptional activity³⁵. Though regular enhancers were largely poor in CGI, CHA enhancers were significantly enriched in CGI. A recent analysis has underlined that a small subset of enhancers characterized by the presence of CGI were stronger, enriched in transcription factor binding and highly conserved³⁶. These data are in line with the present findings and suggest that CHA enhancers define a subset of active regulatory elements likely involved in cell-specific functions and development.

Previous work conducted in GM12878 underlined that transcriptionally active regulatory elements are enriched in the recruitment of transcription factors³⁷. Herein, we provide the confirmation in GM12878 that CHA enhancers are associated with higher recruitment of cell-specific transcription factors and we expand these findings to HCASMC and epidermal keratinocytes as a proof of generalization. These data support that a high level of transcription in H3K27ac defined enhancers is associated with chromatin remodelling in genomic regions enriched for cell-specific functions. A growing body of evidence suggest that enhancer-promoter interactions, which control gene expression, is actively regulated by the recruitment of several molecules including 3D architectural proteins^{4,38}. Data also suggests that transcriptional activity and noncoding RNAs participate in the establishment of genomic interactions³⁹. The present work supports that a high level of transcription in enhancers is associated with a higher density of genomic interactions. CHA enhancers were associated with an elevated recruitment of 3D architectural molecules (RAD21 and SMC3) and with higher density of genomic loops.

Among the CHA enhancers, we found that densely connected elements and forming hub CHA regulatory units were connected to gene promoters of immediate early response genes and transcription factors. Some genes within hub CHA regulatory units, such as *ETV6*, were connected with several regulatory elements and formed a rosette-like structure spanning several hundred of kilobases. These data suggested that some genes with important functions have connections with multiple highly transcribed enhancers. Multiple enhancers converging on gene promoters may afford robustness with lesser impact of stochastic events on the control of gene expression⁴⁰. Among the gene promoters within hub CHA regulatory units, several were early immediate response genes with roles in the inflammatory response. We found that hub CHA enhancers had increased open chromatin and recruitment of BRD4, a transcriptional cofactor⁴¹. Of significance, genes within hub CHA regulatory units had higher recruitment of elongating RNAPII. Moreover, these genes had lower pause index. Hence, hub regulatory units including highly transcribed enhancers is associated with an increased transcription elongation at target genes necessitating robust gene control.

The majority of gene variants associated with complex trait disorders are located in the noncoding genome⁴². Studies have consistently shown that noncoding gene variants are enriched in regulatory elements⁴³. Herein, we found a strong enrichment of PICS autoimmune variants in CHA enhancers. Enhancer-promoter mapping of autoimmune SNPs showed that CHA enhancers were often connected to genes located several kilobases upstream or downstream from the closest annotated gene, which is an accordance with previous studies. Hence, 3D mapping of CHA enhancers may help identify causal candidate genes in genome-wide association studies and reposition drugs for novel treatment opportunities. Consistently, a significant proportion of disease-associated variants mapped by CHA enhancer-promoter contacts identified causal candidate genes in Mendelian randomization. For instance, *IKZF3*, which was mapped by CHA enhancers, is a strong causal candidate for systemic lupus erythematosus as shown by Mendelian randomization. Lenalidomide, which is approved in the treatment of multiple myeloma, targets *IKZF3* for degradation⁴⁴. Drugs targeting this pathway could thus be repositioned for the treatment of lupus⁴⁵.

The present work provided evidence that a subset of enhancers typified by a very high transcriptional activity are associated with active chromatin features and the recruitment of cell-specific transcription factors. We show that CHA enhancers do not appreciably overlap with SEs and provide better assessment of cell-specific functions. Hierarchy among the CHA enhancers was observed as hub CHA regulatory units were associated with higher transcription elongation at connected genes. CHA enhancers may help establish 3D genome regulatory networks involved in the control of gene expression during development and under pathologic conditions.

Methods

CAGE. CAGE normalized expression matrix for the human enhancers was downloaded from the Functional Annotation of the Mammalian Genome (FANTOM5) and data for GM12878 were extracted. Replicates (n=3) were combined by calculating the mean for CAGE-tag value at each annotation with at least one nonzero value. CAGE bam hg19 files were downloaded from FANTOM5. Bam replicates were sorted and merged before being converted to sam with samtools (v1.10). These files were processed with HOMER (v4.11) using makeUCSCfile for each strand and combined into a bedgraph for data visualization in Integrative Genomics Viewer (IGV)⁴⁶.

Processing of ChIP-seq and ATAC-seq. Replicate FASTQ files from GM12878 for ATAC-seq (ENC-SR095QNB) and ChIP-seq for H3K27ac (ENC-SR000AKC), POLR2A/RNAPII (ENC-SR000DKT), POLR2A_phosphoS2 (ENC-SR000DZK), EP300 (ENC-SR000DZD), IRF4 (ENC-SR000BGY), BATF (ENC-SR000BGT),

SMC3 (ENCSR000DZP0), RAD21 (ENCSR000EAC) and CTCF (ENCSR000DZN) were downloaded from the Encyclopedia of DNA Elements (ENCODE). SRA files from ChIP-seq for BRD4 were downloaded from GEO DataSets (GSE62912). For HCASMC, SRA files from GSE1369 (ChIP-seq JUN), GSE124011 (ChIP-seq TCF21 and H3K27ac, ATAC-seq) were downloaded. For NHEK, FASTQ files from ENCODE (ENCSR000ALK) (ChIP-seq H3K27ac), (ENCSR000DCY) (FAIRE-seq) and SRA files from GSE32883 (ChIP-seq MYC) and GSE94471 (ChIP-seq GRHL3) were downloaded. Parallel-fastq-dump (v0.6.5) was used to extract FASTQ from SRA files. FASTQ files were aligned with Bowtie2 (v2.3.5.1) on hg19 and converted to bam files with samtools. Duplicate reads were removed with picard-tools and replicates were merged. The bigWig tracks were generated with deepTools (v3.3.2) bamCoverage and the sorted bam files⁴⁷; bigWig files were visualized in IGV along with h19 gene annotations. Peak call for H3K27ac ChIP was performed with MACS2 (v2.2.7.1) with a threshold q-value at 1E-05 and with a normalization for the input. H3K27ac peaks overlapping the promoter of protein coding genes (TSS±2kb) (GENCODE version 35 in build 37) and the ENCODE blacklist regions were removed by using bedtools (v2.27.1) to obtain the H3K27ac defined enhancers.

Super-enhancers, gene ontology and motifs analyses. Super-enhancers (SEs) annotation was performed by using HOMER findPeaks.pl with the option -style super, data was normalized by input (-i). Jaccard index between annotations were performed by using bedtools with the jaccard option. Gene ontology for annotations was performed by using HOMER annotatePeaks.pl function with the option -go. Transcription factor motif enrichment analysis for annotations was performed by using HOMER findMotifsGenome.pl⁴⁸.

CGI and human conserved noncoding regions. Bed files for CGI and human conserved noncoding elements between man and mouse (HCNE) were downloaded from UCSC and the Atlas of Noncoding Conserved Regions in Animals (ANCORA) respectively⁴⁹. The percentage of referent annotations (CHA and regular enhancers) overlapping with CGI and HCNE (queries) was determined by using the midpoint of query. If a referent annotation had more than one overlap with a query it was reported only once (bedtools intersect with -u option). Proportions of overlap between each group of referents (CHA vs. regular enhancers) were compared with the Fisher's exact test performed with the scipy library in Python (3.8.10).

Tag density plot and heat map. Aligned sam files were normalized for a total of 10 million tag counts and processed with HOMER⁴⁸ makeTagDirectory function with unique reads and MAPQ score > 10. Tag density values centered on annotations were generated by using HOMER with the annotatePeaks.pl function using bins of 10 (-hist 10) and covering 10 kb (-size 10000). Tag count normalized matrix of ChIP-seq processed by HOMER were subtracted by the tag count matrix of input. Plots were generated with ggplot2 in R (4.0.3). Heat map centered on annotations was performed by using deepTools; deepTools computeMatrix function was performed with the ATAC-seq bigWig file and the heat map was generated with the plotHeatMap function. Tag values obtained from HOMER between different annotations were compared by a Wilcoxon rank-sum test performed in R.

Analysis of HiChIP. FASTQ files from H3K27ac-HiChIP for GM12878, HCASMC (GSE101498) and HaCaT (GSE151193) were downloaded and aligned by using HiC-Pro (2.11.4) with default settings⁵⁰. The 1D track for H3K27ac was generated by using deepTools bamCoverage from sorted bam files obtained from HiC-Pro. Loop call was performed with FitHiChIP (v10.0) with default settings and high confidence loops (q-value < 1E - 06) were identified for downstream analyses. Identification of loops overlapping with regulatory elements and gene promoters was performed by using bedtools with the intersect function. Gene promoters were annotated by using a region ± 2kb from the transcription start site and GENCODE version 35 in build 37. TADs were determined by using HiCExplorer (v2.2.1.1) hicFindTADs function and adding a buffer of 10kb. For each chromosome, the ratio of the sum of short (< 200 kb) to long distance (≥ 200 kb) interactions were calculated for loops involving CHA and regular enhancers.

Modelling of 3D interactions. The modelling of 3D interactions was performed by using CSynth⁵¹. CSynth provides an interactive interface to perform in real-time the modelling of chromatin fiber interactions derived from HiC technology. Annotation file in bed format of the region of interest (ROI) was created to which RGB color code was associated for each functional element (enhancers and gene promoters). A contact file of the ROI was obtained from the normalized interaction file from FitHiChIP. CSynth was used with default value parameters except for pushapartforce (set at 75).

Enrichment tests. Immediate early response genes curated by Arner E. et al. were retrieved for enrichment analysis¹². Genes associated with cancer from the COSMIC database⁵² and transcription factors curated by TFcheckpoint were downloaded⁵³. Genes to be tested were intersected with annotated gene lists and hypergeometric tests were performed by using the Python library scipy.stats. Whole genome protein coding genes were used as the background.

Metagene analysis. Transcription start (TSS) and end sites (TES) of mapped genes on each strand were processed by HOMER annotatePeaks.pl (-hist 10) using the POL2RA_phosphoS2 normalized data obtained with HOMER makeTagDirectory with unique reads and MAPQ score >10. Data were combined into a metagene tag file around the TSS (- 1000 and +5000) and TES (- 1000 and +2000). Plots were generated with ggplot2 in R. Tag values between different annotations were compared by a Wilcoxon rank-sum test performed in R.

GRO-seq. SRA files from GRO-seq data in GM12878 (GSE60454) were downloaded and converted to FASTQ files with parallel-fastq-dump. Files were trimmed with cutadapt and aligned with Bowtie2 on hg19. Pause index was calculated as the ratio of reads centered on the TSS including a window of (– 50bp to +300bp) to the gene body normalized for its length⁵⁴. Files in sam format were normalized and processed with HOMER with makeTagDirectory and analyzeRNA.pl -pausing 300. Genes having nonzero tag values at the promoter were retrieved to obtain the pause index from the HOMER output file and compared between groups with the Wilcoxon rank-sum test performed in R. Genes in pause were identified by using a pause index > 2 and compared between groups with a Fisher's exact test performed with the Python library scipy.stats.

Analysis with autoimmune gene variants. GWAS autoimmune variants processed by the Probabilistic Identification of Causal SNPs (PICS)³² were downloaded. PICS is a Bayesian approach for the fine-mapping of causal candidate variants based on the linkage disequilibrium (LD) and the association data. PICS autoimmune data were processed from the GWAS catalog as of January 11, 2019. Autoimmune diseases with at least 5 significant genome-wide significant variants ($P_{\text{GWAS}} < 5E-08$) were processed. Variants were clumped with PLINK at $r^2 > 0.5$ within a 250 kb window and were processed by running PICS2³¹. Variants with PICS probability ≥ 0.025 were kept as candidate causal SNPs. Enrichment of autoimmune PICS variants (query) in annotation (referent) was performed by using the R package GenometriCorr⁵⁵. From the output of GenometriCorr, we reported the relative distance Kolmogorov–Smirnov test and the projection test (binomial test). Variants overlapping with CHA enhancers and connected to gene promoters were assessed for a transcription factor motif analysis with the Predicting Regulatory Functional Effect by Approximate P-value Estimation (PERFECTOS-APE) by using the HOCOMOCO-11 human collection. PERFECTOS-APE is a prediction algorithm to assess regulatory role of SNPs on transcription factor binding sites.

Mendelian randomization. CHA enhancer-promoter looping mapped genes for 4 major autoimmune disorders (rheumatoid arthritis⁵⁶, systemic lupus erythematosus⁵⁷, inflammatory bowel disease⁵⁸, type 1 diabetes⁵⁹) were considered for causal inference in Mendelian Randomization (data were downloaded from the GWAS Catalog; see URLs). In cases where genes mapped by variants were associated to 'Autoimmunity (multiple disease)' in PICS processed data we considered these genes potentially associated with the 4 major autoimmune disorders. Two-sample Mendelian randomization by using multiple *cis* instruments was performed by using blood eQTLs from eQTLGen as the exposition⁶⁰. SNPs associated with the blood expression ($P_{\text{eQTL}} < 5E-08$) within a window of 500kb from the TSS of the gene of interest and independent ($r^2 < 0.1$ based on the 1000G EUR reference panel) were selected as instrumental variables. SNPs were clumped with PLINK1.9 and using the genotypes of European population in the 1000 Genomes Project. Mendelian randomization was performed for genes with at least 3 instrumental variables with the inverse variance weighted (IVW) method obtained from the R package MendelianRandomisation. Gene-disease pairs nominally significant ($P_{\text{IVW}} < 0.05$) in Mendelian randomization were considered as potential causal candidates.

Ethics approval and consent to participate. Datasets from ENCODE, GEO DataSets, FANTOM5, eQTLGen and PICS were publicly available and did not require ethical approval.

Data availability

All the data are available in the manuscript including in the supplementary tables and figures. Sequencing files from ENCODE, FANTOM5 and GEO DataSets were publicly available. Accession number and URLs are provided in the manuscript "Methods" section.

Received: 3 August 2022; Accepted: 30 March 2023

Published online: 05 April 2023

References

- Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
- Zabidi, M. A. & Stark, A. Regulatory enhancer–core-promoter communication via transcription factors and cofactors. *Trends Genet. TIG* **32**, 801–814 (2016).
- Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
- Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
- Jin, Q. *et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J.* **30**, 249–262 (2011).
- Lee, J.-E. *et al.* Brd4 binds to active enhancers to control cell identity gene induction in adipogenesis and myogenesis. *Nat. Commun.* **8**, 2217 (2017).
- Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
- Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15776–15781 (2003).
- Noguchi, S. *et al.* FANTOM5 CAGE profiles of human and mouse samples. *Sci. Data* **4**, 170112 (2017).
- Arner, E. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
- Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
- Agelopoulos, M., Foutadakis, S. & Thanos, D. The causes and consequences of spatial organization of the genome in regulation of gene expression. *Front. Immunol.* **12**, 682397 (2021).

15. Huang, J. *et al.* Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat. Commun.* **9**, 943 (2018).
16. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
17. Nichols, M. H. & Corces, V. G. A CTCF code for 3D genome architecture. *Cell* **162**, 703–705 (2015).
18. Di Giandomartino, D. C., Polyzos, A. & Apostolou, E. Transcription factors: building hubs in the 3D space. *Cell Cycle Georget. Tex* **19**, 2395–2410 (2020).
19. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
20. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
21. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
22. Kang, Y., Kim, Y. W., Kang, J. & Kim, A. Histone H3K4me1 and H3K27ac play roles in nucleosome eviction and eRNA transcription, respectively, at enhancers. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **35**, e21781 (2021).
23. Himeda, C. L., Ranish, J. A., Pearson, R. C. M., Crossley, M. & Hauschka, S. D. KLF3 regulates muscle-specific gene expression and synergizes with serum response factor on KLF binding sites. *Mol. Cell. Biol.* **30**, 3430–3443 (2010).
24. Meeson, A. P. *et al.* Sox15 and Fhl3 transcriptionally coactivate Foxk1 and regulate myogenic progenitor cells. *EMBO J.* **26**, 1902–1912 (2007).
25. Watt, F. M., Frye, M. & Benitah, S. A. MYC in mammalian epidermis: How can an oncogene stimulate differentiation?. *Nat. Rev. Cancer* **8**, 234–242 (2008).
26. Klein, R. H. *et al.* GRHL3 binding and enhancers rearrange as epidermal keratinocytes transition between functional states. *PLoS Genet.* **13**, e1006745 (2017).
27. Eckhart, L., Lippens, S., Tschachler, E. & Declercq, W. Cell death by cornification. *Biochim. Biophys. Acta* **1833**, 3471–3480 (2013).
28. Hock, H. & Shimamura, A. ETV6 in hematopoiesis and leukemia predisposition. *Semin. Hematol.* **54**, 98–104 (2017).
29. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
30. Lladser, M. E., Azofeifa, J. G., Allen, M. A. & Dowell, R. D. RNA Pol II transcription model and interpretation of GRO-seq data. *J. Math. Biol.* **74**, 77–97 (2017).
31. Taylor, K. E., Ansel, K. M., Marson, A., Criswell, L. A. & Farh, K.K.-H. PICS2: Next-generation fine mapping via probabilistic identification of causal SNPs. *Bioinforma. Oxf. Engl.* <https://doi.org/10.1093/bioinformatics/btab122> (2021).
32. Farh, K.K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
33. Wilker, P. R. *et al.* Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation. *Nat. Immunol.* **9**, 603–612 (2008).
34. Lazarian, G. *et al.* A hotspot mutation in transcription factor IKZF3 drives B cell neoplasia via transcriptional dysregulation. *Cancer Cell* **39**, 380–393.e8 (2021).
35. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
36. Bell, J. S. K. & Vertino, P. M. Orphan CpG islands define a novel class of highly active enhancers. *Epigenetics* **12**, 449–464 (2017).
37. Joly Beuparlant, C. *et al.* Metagenome profiles analyses reveal regulatory element's factor-specific recruitment patterns. *PLoS Comput. Biol.* **12**, e1004751 (2016).
38. Agelopoulos, M., Foutadakis, S. & Thanos, D. The causes and consequences of spatial organization of the genome in regulation of gene expression. *Front. Immunol.* **12**, 682397 (2021).
39. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
40. Tsai, A., Alves, M. R. & Crocker, J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *eLife* **8**, e45325 (2019).
41. Rahnamoun, H. *et al.* RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nat. Struct. Mol. Biol.* **25**, 687–697 (2018).
42. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–110 (2015).
43. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
44. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
45. Lipsky, P. E. *et al.* Biological impact of iberdomide in patients with active systemic lupus erythematosus. *Ann. Rheum. Dis.* <https://doi.org/10.1136/annrheumdis-2022-222212> (2022).
46. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
47. Ramirez, F. *et al.* deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–165 (2016).
48. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
49. Engström, P. G., Fredman, D. & Lenhard, B. Ancora: A web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. *Genome Biol.* **9**, R34 (2008).
50. Servant, N. *et al.* HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
51. Todd, S. *et al.* CSynth: An interactive modelling and visualization tool for 3D chromatin structure. *Bioinforma. Oxf. Engl.* **37**, 951–955 (2021).
52. Tate, J. G. *et al.* COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
53. Chawla, K., Tripathi, S., Thommesen, L., Lægveid, A. & Kuiper, M. TFcheckpoint: A curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinforma. Oxf. Engl.* **29**, 2519–2520 (2013).
54. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
55. Favorov, A. *et al.* Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput. Biol.* **8**, e1002529 (2012).
56. Ha, E., Bae, S.-C. & Kim, K. Large-scale meta-analysis across East Asian and European populations updated genetic architecture and variant-driven biology of rheumatoid arthritis, identifying 11 novel susceptibility loci. *Ann. Rheum. Dis.* **80**, 558–565 (2021).
57. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).
58. de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
59. Forgetta, V. *et al.* Rare genetic variants of large effect influence risk of type 1 diabetes. *Diabetes* **69**, 784–795 (2020).
60. Vösa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).

Author contributions

P.M., M.B. and A.R. conceived and designed analyses. A.R., M.B. and L.H.M.M. completed the figures. M.B. completed the tables. S.M. and P.M. completed MR analyses. P.M. and M.B. drafted the manuscript. All the authors critically reviewed the manuscript and provided intellectual inputs.

Funding

This work was supported by the Quebec Heart and Lung Institute Fund. Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases. P.M. is the recipient of the Joseph C. Edwards Foundation granted to Université Laval.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-32669-3>.

Correspondence and requests for materials should be addressed to P.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023