**Genome Analysis**

**HUGIn: Hi-C Unifying Genomic Interrogator**

Joshua S. Martin\(^1\), Zheng Xu\(^4\), Alex P. Reiner\(^5\), Karen L. Mohlke\(^1\), Patrick Sullivan\(^1,6\), Bing Ren\(^7\), Ming Hu\(^8,*\) and Yun Li\(^1,2,3,*\)

\(^1\)Department of Genetics, \(^2\)Department of Biostatistics, \(^3\)Department of Computer Science, University of North Carolina, Chapel Hill NC 27599, USA, \(^4\)Department of Statistics, University of Nebraska, Lincoln, NE 68583, USA, \(^5\)Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, \(^6\)Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, \(^7\)University of California, San Diego School of Medicine, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, UCSD Moores Cancer Center, 9500 Gilman Drive, La Jolla, California 92039, USA, \(^8\)Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA.

*To whom correspondence should be addressed.

**Abstract**

**Motivation:** High throughput chromatin conformation capture (3C) technologies, such as Hi-C and ChIA-PET, have the potential to elucidate the functional roles of non-coding variants. However, most of published genome-wide unbiased chromatin organization studies have used cultured cell lines, limiting their generalizability.

**Results:** We developed a web browser, HUGIn, to visualize Hi-C data generated from 21 human primary tissues and cell lines. HUGIn enables assessment of chromatin contacts both constitutive across and specific to tissue(s) and/or cell line(s) at any genomic loci, including GWAS SNPs, eQTLs and cis-regulatory elements, facilitating the understanding of both GWAS and eQTLs results and functional genomics data.

**Availability:** HUGIn is available at [http://yunliweb.its.unc.edu/HUGIn](http://yunliweb.its.unc.edu/HUGIn).

**Contact:** yunli@med.unc.edu and hum@ccf.org

**Supplementary information:**

---

1 **Introduction**

Elucidating the functional roles of the vast majority (>80-90%) of non-coding variants identified from GWAS, and subsequently finding their target gene(s), are pressing and daunting tasks. Investigators have only recently realized the power and value of high throughput chromatin conformation capture (3C) technologies, such as Hi-C (Lieberman-Aiden et al., 2009), TCC (Kalhor et al., 2011) and ChIA-PET (Fullwood et al., 2009), for potential target gene identification (Pombo and Dillon, 2015; Dekker et al., 2017).

Several 3D genome browsers, including WashU Epigenome Browser (Zhou et al., 2013) and Juicebox (Durand et al., 2016), have been developed to visualize 3C-based data. However, the vast majority of genome-wide unbiased chromatin organization studies published to date have used cultured cell lines, limiting their generalizability. Understanding chromatin structure and understanding complex trait genetic architecture are each daunting challenges. Understanding both simultaneously is even more daunting, but of paramount importance to unveil genetic mechanisms underlying complex diseases (Gamazon et al., 2015; Jakobsdottir et al., 2009).

2 **Methods**

We developed HUGIn (Hi-C Unifying Genomic Interrogator) to host our recently published compendium of Hi-C data across 14 human primary tissues and 7 cell lines (Schmitt et al., 2016) ([http://yunliweb.its.unc.edu/HUGIn](http://yunliweb.its.unc.edu/HUGIn)). For each tissue or cell line, HUGIn visualizes the observed and expected read counts between two genomic loci, and the corresponding statistical significance, for long range chromatin interactions, in both heatmap and virtual 4C plots (detailed in Supplementary Material Section 1 and Fig. S1). Compared to existing 3D genome browsers, HUGIn has five major unique features. First, HUGIn simultaneously visualizes a compendium of Hi-C data, which is continually updated as new data become available (Supplementary Material Section 2). Second, it directly enables detection of chromosomal organizations both specific to, and constitutive across, tissue(s) and/or cell line(s). Third, HUGIn can display Hi-C data anchored at genomic locus of any size, suggesting potential target gene(s) for GWAS variants in a tissue or cell line informative manner. This feature is particularly important for evaluating a GWAS locus, which can range from a single nucleotide to a region of linkage disequilibrium extending over tens of thousands of base pairs (Smith et al., 2005). Fourth, HUGIn also hosts gene expression and a rich collection of epigenomic data, including typical and super enhancers, CTCF binding sites, frequently interacting regions (FIREs) (Schmitt et al., 2016), and several core histone modifications (Supplementary Material Section 1). Lastly, HUGIn generates a list of most likely target gene(s) for any...
Results

Figure 1 is a virtual 4C plot in which we display a long range chromatin interaction between the type 2 diabetes (T2D)- and adiponectin-associated SNP rs6450176 with the gene FST (Civelek et al., 2017). Although the SNP is located within the ARL15 gene, it was shown to be an eQTL for FST but not ARL15 in subcutaneous adipose tissue. In addition, expression level of FST but not ARL15 is correlated with adiponectin levels (Civelek et al., 2017). More details and the virtual 4C plot for the rs6450176-FST interaction are shown in Fig. S2. Two more examples, a long range interaction between BMI-associated SNP rs9930506 and the gene IRX3 (Smemo et al., 2014), and between schizophrenia-associated SNP rs1191551 and the gene FOXG1 (Won et al., 2016), are shown in Fig. S3-S4. We further investigated all intergenic SNPs in the NHGRI catalog of published GWAS for seven diseases including schizophrenia, leukemia, Alzheimer’s disease, autism, depression, type 1 diabetes and type 2 diabetes (Supplementary Material Section 3). We found that, across all diseases examined, >85% of HUGIn-annotated genes are not the gene closest to the GWAS SNP (Table S1); and that HUGIn-annotated gene lists tend to provide stronger evidence for enrichment with biologically relevant terms or gene sets (Tables S2-S3).

Conclusions

With the continuing accumulation of high resolution 3C-based data in disease-relevant human tissues, we expect that HUGIn will become an increasingly valuable tool for many investigators, including biologists interested in fundamental chromosome organization across cell lines, tissue types, or developmental stages, and geneticists seeking to understand the genetic architecture and mechanism underlying complex diseases and traits.

Acknowledgements

We would like to thank Dr. Michael L. Boehnke for critical read of an earlier version of this manuscript. We also would like to thank TOPMed investigators for providing helpful comments to improve the usefulness of HUGIn to their research.

Funding

This work is supported by NIH grants R01HG006292 (YL), R01HL129132 (YL and APR) and U54DK107977 (MH and BR). Conflict of Interest: none declared.

References

Article short title


