NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Exploring long-range genome interaction data using the WashU Epigenome Browser

Xin Zhou[1], **Rebecca F. Lowdon**[1], **Daofeng Li**[1], **Heather A Lawson**[1], **Pamela A.F. Madden**[2], **Joseph F. Costello**[3], and **Ting Wang**[1]

[1]Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108

[2]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO 63110

[3]Brain Tumor Research Center, Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, CA 94143

To the Editor

Eukaryotic chromosomes are a highly organized three-dimensional entity. Chromosome folding is a tightly regulated process[1,2], with important functions including bringing distal regulatory elements into the vicinity of their target gene promoters and arranging the chromosomes into distinct compartments. Mapping and visualizing higher order chromatin interactions are critical to genomics research, especially for interpreting high-throughput epigenomic data[3–6]. Recent technological innovations, including 5C, Hi-C, and ChIA-PET, have facilitated discovery of chromosomal organization principles and folding architectures at unprecedented scales and resolution. Each technology also comes with corresponding computational tools[7,8] to process and visualize its specific data type (Supplementary Notes). However, visualizing and navigating long-range interaction data, as well as integrating these interactions with other epigenomics data, remain a much-desired capability and a daunting challenge[9].

We have extended the WashU Epigenome Browser[10](http://epigenomegateway.wustl.edu) to support multiple types of long-range genome interaction data. The WashU Epigenome Browser was developed for the Roadmap Epigenomics project[11] and incorporates ENCODE data. It currently hosts thousands of epigenome and transcriptome datasets for multiple cell types, tissues, individuals, and species. Long-range interaction data and visualization functions are now integrated with existing data and functions of the Browser, enabling investigators to explore epigenomic data in the context of higher order chromosomal domains, and to generate multiple types of intuitive, publication quality figures of interactions. In Figure 1 we display both long-range interaction data and histone modification data of two human cell lines (IMR90 and K562). Two interacting regions are displayed side-by-side and can be explored simultaneously. Hi-C and ChIA-PET capture similar interaction patterns in the two cell lines over these two regions. Interestingly, these two interacting regions exhibit distinct cell type-specific characteristics in their histone modification profiles (Figure 1). These observations are consistent with the hypothesis that chromatin domains are stable across cell types, but can have different epigenetic profiles in different cells. Genes within each domain are regulated epigenetically in a cell type-specific manner. Integrating higher order chromatin interaction data with other genomic data could potentially reveal novel insights about mechanisms underlying gene and genome regulation.

Correspondence should be addressed to: T.W. (twang@genetics.wustl.edu) or J.F.C (jcostello@cc.ucsf.edu).

The WashU Epigenome Browser combines advanced web technology with intuitive graphical design to visualize long-range interaction data. Pairs of interacting regions can be joined by arcs (Figure 1b, 1c), or indicated by filled rectangles in a heatmap (Figure 1d). The "arc" glyph is suitable to represent sparse interactions (as is common with ChIA-PET, and sometimes found in 5C data sets), and the "heatmap" glyph is best for dense interactions (as is typical for Hi-C and some regions in 5C data sets). The data tracks are interactive: investigators can click on the arcs or heatmap cells to invoke a companion Browser panel (Supplementary Fig. 1), which displays epigenomic data over the distal interacting locus. This companion panel can be navigated independently, enabling comparison of data patterns of interacting loci in the same view. Thus, investigators can observe several loci that are distant in their genomic coordinates but are inferred to be spatially close to each other in the nucleus. Both the "arc" and "heatmap" modes only display interactions that are contained within the current browsing range while omitting interactions beyond the range. To visualize the complete set of interactions, investigators can invoke the "Circlet View" (Supplementary Fig. 2). In the Circlet View, the chromosomal axis curls to form a circle and interactions are displayed as arcs inside the circle. Investigators can choose to display one single chromosome or to include interacting chromosomes to achieve a whole-genome perspective of the interactions. Gene Set View and genomic juxtaposition can be combined with the long-range interaction function to focus on the interaction events within a subregion of the genome (Figure 1). An investigator's own long-range interaction data can be displayed on the Browser via the custom track or Data Hub function.

The WashU Epigenome Browser continues to evolve. We report here a new development that innovates how long-range genome interaction data can be visualized and explored. As long-range interaction data becomes increasingly available, the Epigenome Browser will have a significant impact on our understanding of how eukaryotic genomes function as nonlinear systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Dekker J. Gene regulation in the third dimension. Science. 2008; 319:1793–1794.10.1126/science. 1152850 [PubMed: 18369139]

2. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012; 489:109–113.10.1038/nature11279 [PubMed: 22955621]

3. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293.10.1126/science.1181369 [PubMed: 19815776]

4. Handoko L, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nature genetics. 2011; 43:630–638.10.1038/ng.857 [PubMed: 21685913]

5. Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380.10.1038/nature11082 [PubMed: 22495300]

6. Li G, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. [PubMed: 22265404]

7. Lajoie BR, van Berkum NL, Sanyal A, Dekker J. My5C: web tools for chromosome conformation capture studies. Nature methods. 2009; 6:690–691.10.1038/nmeth1009-690 [PubMed: 19789528]

8. Maunakea AK, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010; 466:253–257.10.1038/nature09165 [PubMed: 20613842]

9. Stamatoyannopoulos JA. What does our genome encode? Genome research. 2012; 22:1602–1611.10.1101/gr.146506.112 [PubMed: 22955972]

10. Zhou X, et al. The human epigenome browser at washington university. Nat Methods. 2011; 8:989–990. [PubMed: 22127213]

11. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology. 2010; 28:1045–1048.10.1038/nbt1010-1045
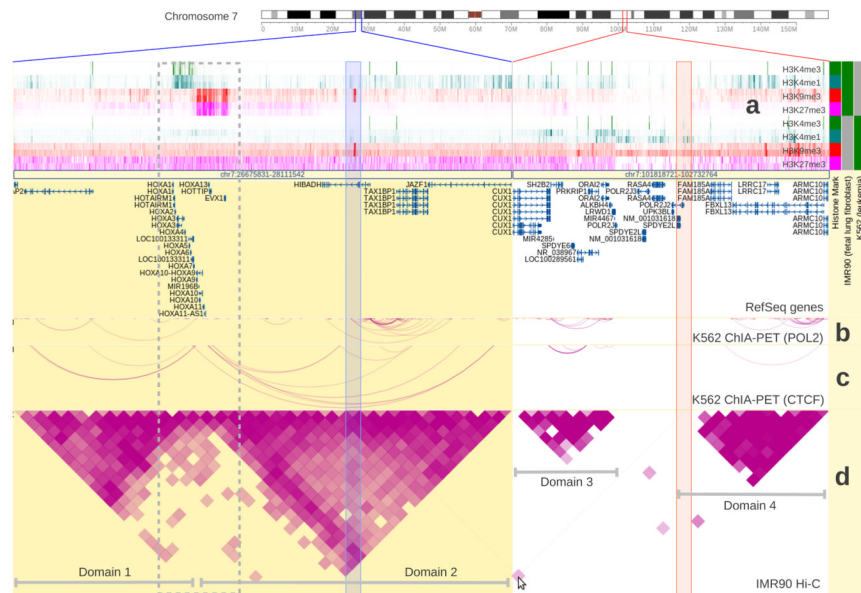
Figure 1.
Two genomic regions (marked by blue or red lines on the chromosome ideogram) with stable chromatin domains are characterized by cell type-specific epigenetic profiles. The chromatin domains are marked by three long-range interaction tracks on two cell types. (**a**) Heatmap view of histone modification profiles of both IMR90 and K562 cells. (**b**) and (**c**) are two ChIA-PET tracks from K562 cells with different DNA binding proteins (*POL2* and *CTCF* respectively, ENCODE data). (**d**) is a Hi-C track from IMR90 cells[5]. The triangle shapes in the Hi-C track depict chromatin domains in IMR90 cells (labeled Domains 1, 2, 3, 4), and the arcs in the ChIA-PET tracks indicate similar domain structure in K562 cells, suggesting that chromatins in these regions are organized similarly in both IMR90 and K562 cells. Histone modification tracks are shown in distinct colors to indicate the modification (red and fuchsia for repressive histone marks, green and teal for active histone marks). Histone modification marks are also identified by the color blocks in the metadata color map on the right. In the first region, the HOXA gene cluster at the boundary of the two domains shows a bipartite histone modification pattern (enclosed by a grey dotted box) in IMR90 cells. The left part of the cluster is covered by active marks (H3K4me3 and H3K4me1) and the right part by repressive marks (H3K9me3 and H3K27me3). This is consistent with the normal, co-linear expression pattern of this locus. In K562 cells this pattern is replaced by a uniform repressive pattern (depleted of active histone marks and enriched in repressive H3K27me3 mark). In the second region, IMR90 and K562 cells do not show significant differences in their histone modification patterns. A cell in the Hi-C track suggests an interaction event between the two regions (the spot is indicated by the cursor, the underlying interacting regions are indicated by the two semi-transparent columns). This cell represents a highly probable interaction event between two remote genomic regions. Additional information is provided in Supplementary Notes.