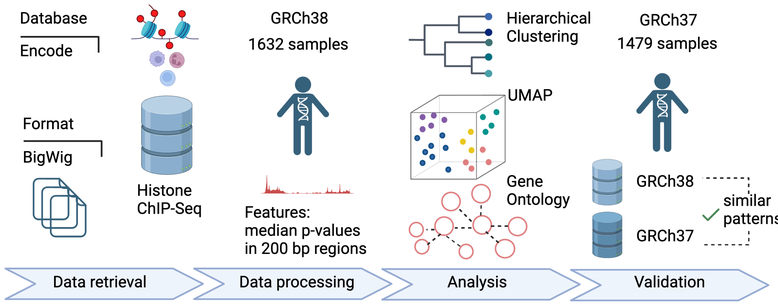
A bird's eye view of relationships between epigenetic modifiers across immune cells

Anastasiia Kim, Nicholas Lubbers, Christina R. Steadman, Karissa Y. Sanbonmatsu
Los Alamos National Laboratory

Abstract

Exposure to environmental toxins and infections leaves epigenetic signatures on chromosomes. The avalanche of epigenomic data is difficult to parse for biological interpretation given non-linear complex patterns and relationships. This attractive challenge in epigenomic data lends itself to machine learning for discerning infectivity and susceptibility. Here, we establish a baseline for understanding the typical clustering patterns of epigenetic features in a non-diseased state. We provide a comprehensive overview of the relationships between various epigenetic modifiers and immune cell types across all chromosomes. By establishing what constitutes a normal or typical epigenomic configuration, we set a foundation against which changes in the disease state can be measured.

Approach



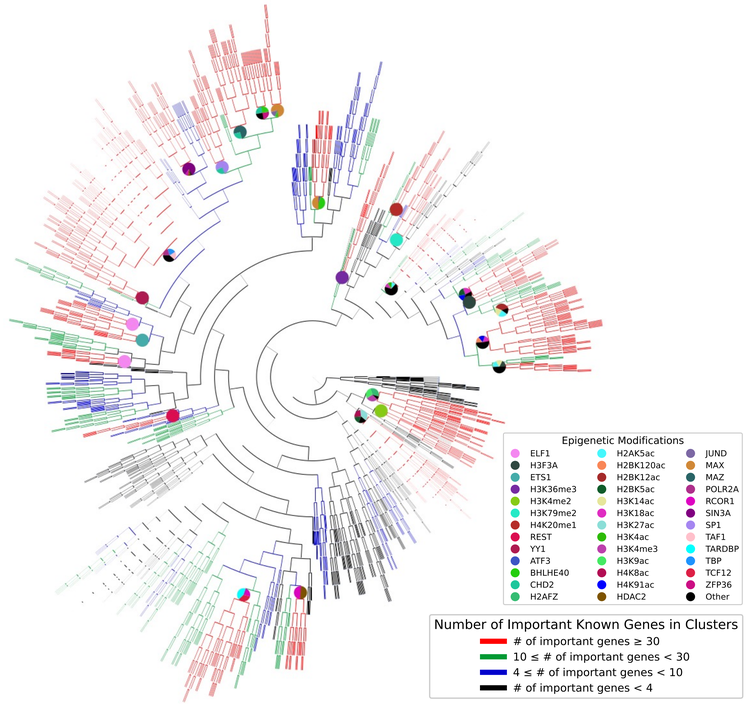
A flowchart of our approach.

- Hierarchical clustering using Pearson correlation.
- UMAP visualization.
- Entropy-based analysis across chromosomes.
- Identification of genes among important features that drive clustering.
- Co-occurrence analysis of epigenetic modifiers.
- Gene linkage with Gene Ontology for functional insights.
- Validation on another dataset.

Numerous known genes drove the clustering

A considerable number of epigenetic modifiers tended to group together.

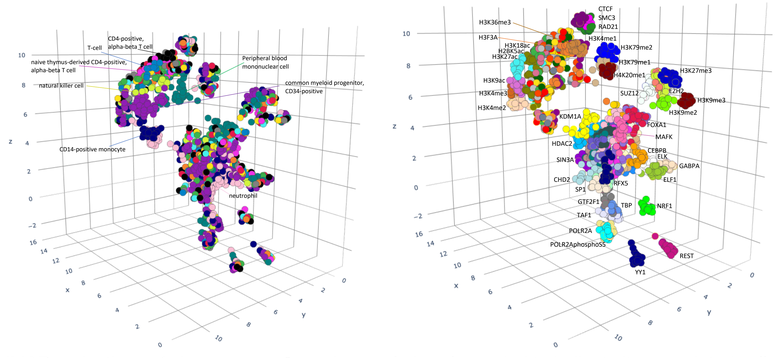
- Many important known genes, defined as genome regions with p-value < 0.05 in all samples within the cluster, were abundant in most clusters, as indicated by the color-coded branches.
- Several pie charts, showing the count of distinct epigenetic modifiers within each cluster, consisted of only one or two colors, signifying that clusters were characterized by one or two specific epigenetic modifiers, respectively.



Dendrogram of chromosome 6 reflecting a clustering of samples influenced predominantly by the specific traits of epigenetic modifiers.

Samples clustered first by the epigenetic modifiers and then by the cell types

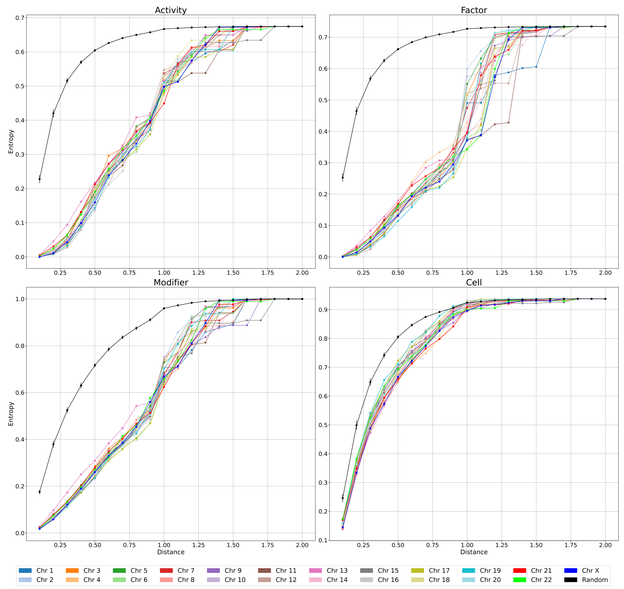
- The main distinction in clustering was notably between histone and non-histone modifiers. After that, the clustering was predominantly shaped by the modifiers.
  - Histone 2 and 3 acetylation modifiers commonly co-occurred in the same clusters.
  - The trio of RAD21, SMC3, and CTCF was also frequently found to co-occur, consistent with their participation with the cohesin complex.
  - TAF1 and TBP were frequently found in the same cluster with GTF2F1, consistent with their involvement in RNA Polymerase II transcription.
  - EZH2 and EZH2phosphoT487 were clustered together with SUZ12 and H3K27me3 which aligns with the function of the PRC2 complex in laying down the H3K27me3 mark.



UMAP projections of chromosome 6, each sample is color-coded to indicate its associated cell type (left) and epigenetic modifier (right).

Clustering behavior was consistent across all chromosomes

- Entropy made a dramatic downward jump around 10 clusters (where the distance was around 1) for the activity, factor, and epigenetic modifier plots.
- Comparing the entropy of randomly shuffled labels to original curves showed less meaningful cell type clustering than clustering by other identifiers.



The weighted entropy values with respect to activity type, factor, modifier, and cell type, as a function of the distance (1 - Pearson) used to cut the hierarchical clustering dendrogram.

Many of the known genes were associated with microRNAs and snoRNAs

- Several significant Gene Ontology terms were enriched in our important genes including those involved with chromatin remodeling and nucleosome occupancy, mRNA splicing and regulation, and immune responses.
  - The GO term "Innate immune response to mucosa" was aligned with the h2BC10, H2BC11, H2BC6, H2BC7, H2BC8, and RNASE3 genes, which are directly related to the immune system processes.
- KEGG pathway analysis revealed important genes that were connected to microRNAs, which were best annotated and studied in the context of cancer.

Conclusions

- Evaluation on another dataset revealed that the same epigenetic modifiers tended to cluster together regardless of the dataset.
- Many of the known genes that drove clustering were associated with microRNAs and snoRNAs, which are key to regulating gene expression and are thought to be dysregulated during aberrant immune function and in various cancers.
- The genes constantly regulated by epigenetic modifiers, regardless of the type modifier or the extent of permissiveness, are those involved in the formation of chromatin, gene expression processes, and small noncoding RNAs. Identifying the epigenetic regulation of noncoding RNAs may be key to understanding disease states, which requires support via empirical data.

Los Alamos NATIONAL LABORATORY
National Nuclear Security Administration
Managed by Triad National Security, LLC, for the U.S. Department of Energy's NNSA.
This study was supported by the Defense Threat Reduction Agency, Grant DTRA1308139949.