Cross-Species Epigenomic Maps of Intergenic Loci for Hypertension and Blood Pressure

Monika Tutaj[1], Chien-Wei Lin[1], Atrayee Ray[2], Chun Yang[1], Cary Stelloh[2], Pengyuan Liu[3], Yong Liu[3], Michael Widlansky[1], Andrew Greene[4], Aron M. Geurts[1], Allen W. Cowley[1], Mingyu Liang[3], Sridhar Rao[2], Anne E. Kwitek[1]

[1]Medical College of Wisconsin, Milwaukee, WI; [2]Versiti Blood Research Institute, Milwaukee, WI; [3]The University of Arizona, Tucson, AZ; [4]The Jackson Laboratory, Bar Harbor, ME, United States.

Background

Many non-coding SNPs identified in Genome-Wide Association Studies (GWAS) may influence blood pressure (BP)-related gene expression through **epigenetic mechanisms**.

Analyzing the **comparative genomic and epigenomic landscapes** in human and rat kidney tissues, particularly focusing on interspecies similarities in **intergenic regions**, can provide insights into these mechanisms.

High-resolution, genome-wide epigenomic maps of key BP-relevant tissues, such as proximal tubule (**PT**) and medullary thick ascending limb (**mTAL**), are crucial for understanding these **regulatory elements**.

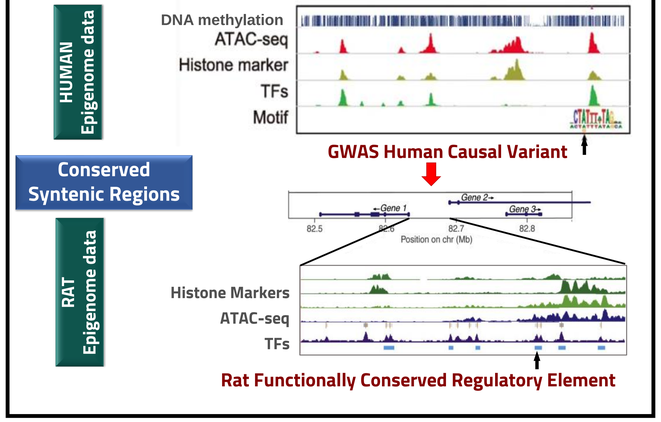
While the UCSC Genome Browser has extensive human data, **rat data is limited**, especially for updated genomes.

Objective

- Identify conserved BP gene regulatory elements in intergenic regions between human and rats for in vivo validation and mechanistic studies.
- Develop a pattern recognition algorithm to detect conserved epigenomic marks in BP GWAS loci and enhance cross-species comparisons using advanced visualization tools.
- Create high-resolution comparative maps by integrating epigenomic data (e.g., ATAC-seq, Hi-C, CUT&Tag, DNA methylation, RNA-seq) from human and rat tissues.
- Build comparative genome and epigenome maps by:
  - Using high-quality reference genomes and annotations
  - Locating conserved sequence and epigenome regions
  - Using relevant kidney tissue data for testing matching algorithms
  - Focusing on LD blocks located at least 10 kb from any protein-coding gene

Selected human BP GWAS LD blocks

	SNP	Chr	Population	LD region hg38 (TOPMED)	LD size
1	rs1173771	5	european	32814922-32832368	17447
2	rs42398	5	trans-ethnic group	96753407-96859577	106171
3	rs12078697	1	european	116471907-116477722	5816
4	rs143112823	3	european	154937749-154995835	58087
5	rs2782980	10	european	114021768-114021788	21
6	rs1475130	14	trans-ethnic group	99758807-99767143	8337
7	rs17477177	7	european	106770331-106776021	5691
8	rs11145807	9	european	136622614-136626445	3832

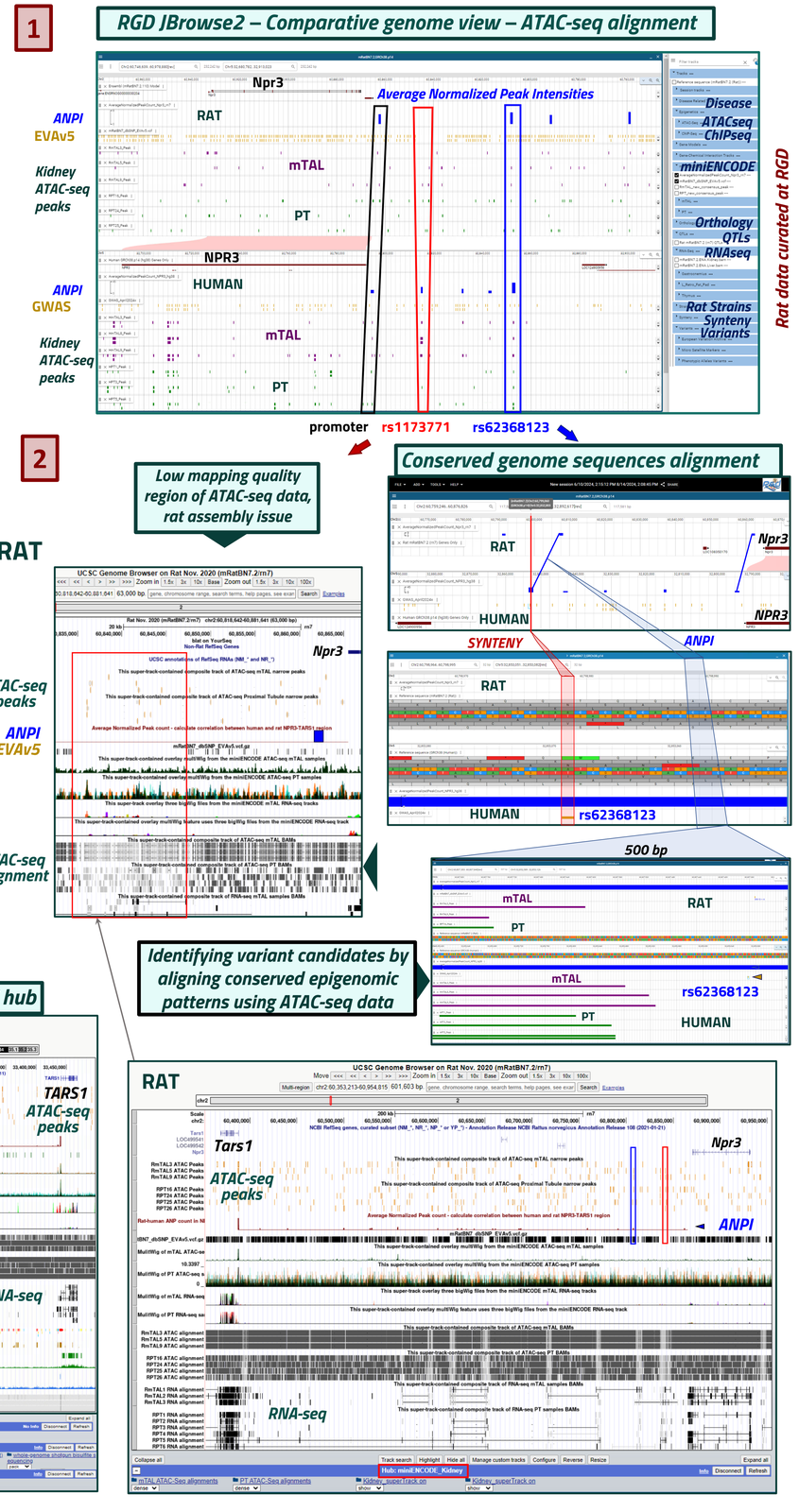


Results

1. We analyzed **ATAC-seq** peak data from mTAL and PT kidney samples (3-4 replicates) in NPR3-TARS1 intergenic region. Conserved epigenomic patterns were identified via **canonical correlation analysis** of average normalized peak intensities (**ANPI**) between human and rat (CC = 0.94 with p-value = 0.0001 by permutation test). This data is accessible through a **cross-species JBrowse2** browser integrated with RGD annotations, including variants and gene interaction tracks.

2. We present **NPR3 region** - an 18 kb linkage disequilibrium block on human chromosome 5 (GRCh38) and a corresponding region on rat chromosome 2 (mRatBN7.2), featuring **GWAS variants rs1173771 and rs62368123** associated with hypertension and body mass, respectively.

3. We created a **miniENCODE Kidney hub** and connected it to the UCSC Genome Browser resource.



Methods

1. Proximal tubule **(PT)** and medullary thick ascending limb **(mTAL)** samples were isolated from human (cadaveric donors or nephrectomies) and rat kidney segments.

2. Human and rat ATAC-seq data were mapped to GRChg38 and mRatBN7.2 genome references. Open chromatin regions were identified using the MACS2 tool.

3. Random intergenic regions (>50 kb) and those near kidney-expressed genes were selected to evaluate pattern recognition algorithms for developing the comparative epigenome tool.

4. Human and rat intergenic regions were divided into equal number of segments respectively. For each segment in each sample, the maximum normalized count value was calculated, resulting in separate matrices for human and rat data. Canonical correlation analysis (CCA), performed using R package CCA, was then applied to identify the maximum correlation between the intergenic region matrices of human and rat. P-value of the correlation was assessed by permutation test using Wilk's statistics.

Conclusion

- **Genetic and Epigenetic Contributions to Blood Pressure Regulation:** Linking noncoding SNPs to blood pressure regulation and hypertension pathogenesis is challenging due to their distance from protein-coding genes.
- **Identification of Conserved Regulatory Regions:** Aligning human and rat data using high-quality reference genomes and synteny information helps identify sequence conservation. However, discovering conserved regulatory regions without sequence conservation remains challenging. We focus on shared epigenomic features for this purpose.
- **Developing Comparative Epigenome Tools:** Our comparative epigenome maps and pattern recognition algorithm identify conserved epigenomic marks as demonstrated by correlations in ATAC-seq data at the NPR3-TARS1 intergenic region (CC = 0.94 with p-value = 0.0001 by permutation test).
- **Visualization for Insight and Intervention:** Advanced visualization tools like JBrowse2, integrated into the Rat Genome Database (RGD), improve the organization of genomic and epigenomic data to prioritize epigenomic conservation and support the development of targeted interventions.

Grant support: P01 HL149620 , R01 HL064541