Finding epigenetic blood biomarkers for Alzheimer's Disease

Amulya Garimella[1], Easwaran Ramamurthy[2], Andreas Pfenning, Ph.D.[2]

[1] University of Pittsburgh Cancer Institute Academy (Hillman Academy)

[2] Computational Biology Department, Carnegie Mellon University

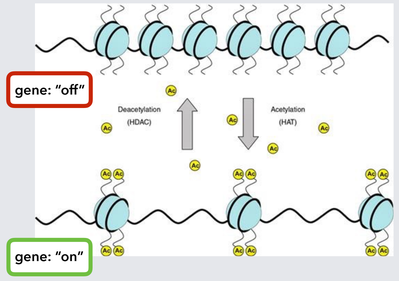
Background

Alzheimer's Disease (AD) = **immune basis in blood + brain**[1]

How can we find epigenetic AD blood biomarkers?

H3K27ac: histone acetylation, marker of nearby open chromatin[2]

- related to areas on genome that are linked to AD

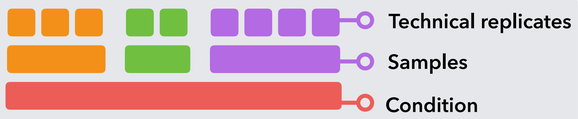


This figure[3] (adapted) shows the process of histone acetylation.

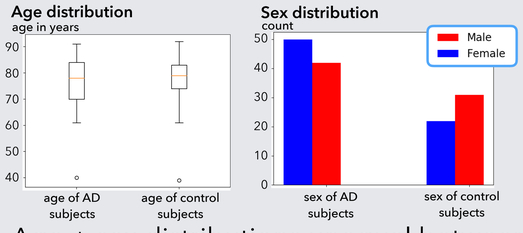
Hypothesis

Differential H3K27ac ChIP-seq peaks between AD and control samples in blood data can be used as an **epigenetic blood biomarker for AD.**

Dataset

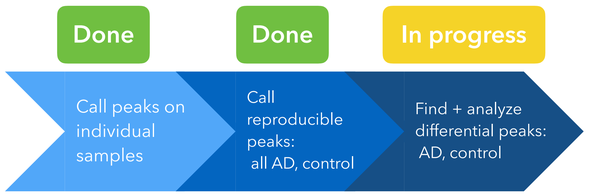


- **Buffy-coat H3K27ac ChIP-seq data - Kellis Lab @ MIT**
  - Postmortem human brain
  - Project previously designed by Pfenning Lab
- 145 subjects - 92 AD, 53 control



- Age + sex distributions unequal between AD and control

Method



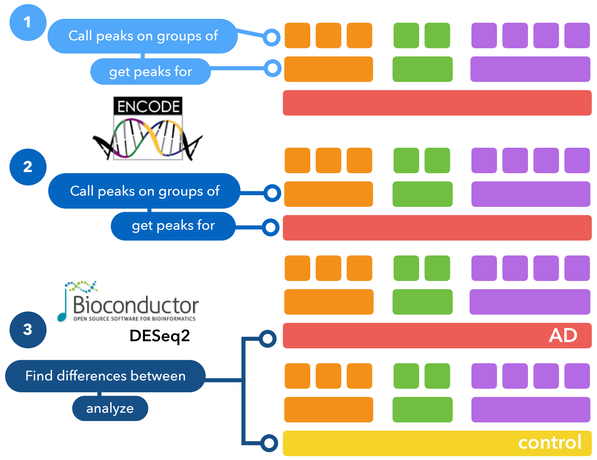
Peak calling - ENCODE ChIP-seq analysis pipeline[4]

- Automated input creation and pipeline runs

Differential peaks - DESeq2[5]

Analysis - GREAT[6], MEME-ChIP[7]

Model



Individual peak calling: results

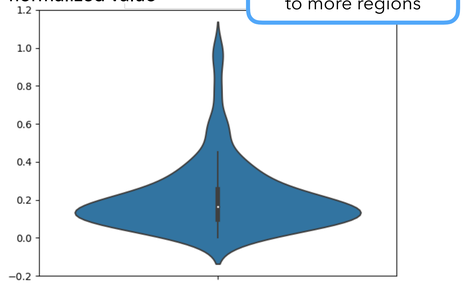
For a given *sample*, where does the histone acetylation occur?

ENCODE Pipeline process:

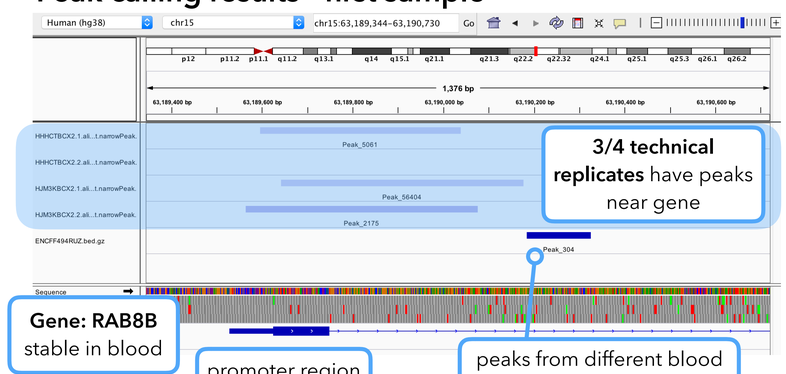
1. FASTQ (sequencing data) files aligned to genome w/ BWA[8]
2. Peaks called w/ MACS2[9]
3. QC metrics generated
   - Ex. FRiP, Reproducibility test

QC example - FRiP

(Fraction of Reads in Peaks)

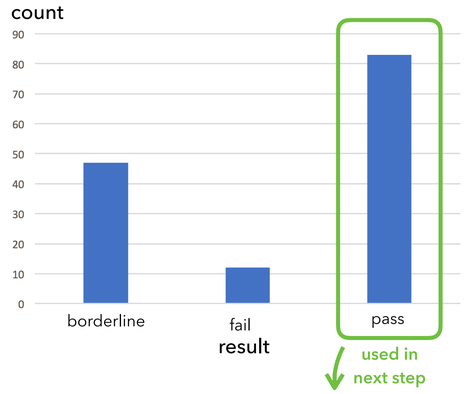


Peak calling results - first sample



Comparison data: We downloaded the call sets from the ENCODE portal (Sloan et al. 2016) (https://www.encodeproject.org/) with the following identifier: ENCBS770TLM

Reproducibility test results



Call reproducible peaks: AD, control

For a given *condition*, where does the histone acetylation occur?

- Used QC metrics from individual peak calling
- 55 high quality samples for AD (6 batches), 29 high quality samples for control (3 batches)
- Pooled sample peaksets into batches - 10 samples each
- Ran pipeline on these batches
- **Output generated:** two peaksets – one for AD, one for control

Summary & Next Steps

Summary

- **Peaks were called** on individual samples
- **QC was performed** to filter out samples w/ reproducible peaks
- **Noise was reduced** in peaksets and overall peaksets were generated

Next Steps

- **Differential peak identification**
  - Where does H3K27ac bind in AD as opposed to control?
- **Differential peak analysis** w/ DESeq2
- Find associated **genes, motifs, pathways**

Future Directions

Which cell types are involved in differential acetylation between AD and control samples?

- **Deconvolve** whole blood data
- **Monocyte profiling** between AD and control samples

Acknowledgements

Neurogenomics Laboratory @ CMU

Andreas Pfenning, Ph.D.
Easwaran Ramamurthy
Irene Kaplow, Ph.D.
Cathy Su
Michael Kleyman
Ashley Brown

Kellis Laboratory @ MIT

Lei Hou, Ph.D
Manolis Kellis, Ph.D

Kundaje Laboratory @ Stanford

Jin Lee
Anshul Kundaje, Ph.D

Hillman Academy

University of Pittsburgh Department of Biomedical Informatics

David Boone, Ph.D
Solomon Livshits
Patrick Flaherty
The students at CoSBBI

References

1. Gjoneska, E., Pfenning, A. R., Mathys, H., Quon, G., Kundaje, A., Tsai, L. H., & Kellis, M. (2015). Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. Nature, 518(7539), 365.
2. Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., ... & Boyer, L. A. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences, 107(50), 21931-21936.
3. Pons, D., de Vries, F. R., van den Elsen, P. J., Heijmans, B. T., Quax, P. H., & Jukema, J. W. (2009). Epigenetic histone acetylation modifiers in vascular remodelling: new targets for therapy in cardiovascular disease. European heart journal, 30(3), 266-277.
4. Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., ... & Chen, Y. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome research, 22(9), 1813-1831.
5. Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology, 15, 550. doi: 10.1186/s13059-014-0550-8.
6. Cory Y McLean, Dave Bristor, Michael Hiller, Shoa L Clarke, Bruce T Schaar, Craig B Lowe, Aaron M Wenger, and Gill Bejerano. "GREAT improves functional interpretation of cis-regulatory regions". Nat. Biotechnol. 28(5):495-501, 2010. PMID 20436461
7. Philip Machanick and Timothy L. Bailey, "MEME-ChIP: motif analysis of large DNA datasets", Bioinformatics 27(12): 1696-1697, 2011. [full text]
8. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168]
9. Zhang et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol (2008) vol. 9 (9) pp. R137
10. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57.
11. Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., ... & Onate, K. C. (2017). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic acids research, 46(D1), D794-D801.