ChIP-seq Mapping of Distant-Acting Enhancers and Their *In Vivo* Activities

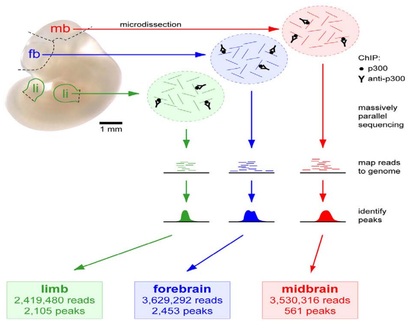
Axel Visel and Len A. Pennacchio

Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

1. SUMMARY

- The genomic location and function of most distant-acting transcriptional enhancers in the human genome remains unknown
- We performed ChIP-seq for various transcriptional coactivator proteins (such as p300) directly from different embryonic mouse tissues, identifying thousands of binding sites
- Transgenic mouse experiments show that p300 and other co-activator peaks are highly predictive of genomic location AND tissue-specific activity patterns of distant-acting enhancers
- Most enhancers are active only in one or very few tissues
- Genomic location of tissue-specific p300 peaks correlates with tissue-specific expression of nearby genes
- Most binding sites are conserved, but the global degree of conservation varies between tissues

2. GENOME-WIDE MAPPING OF P300 IN EMBRYONIC TISSUES

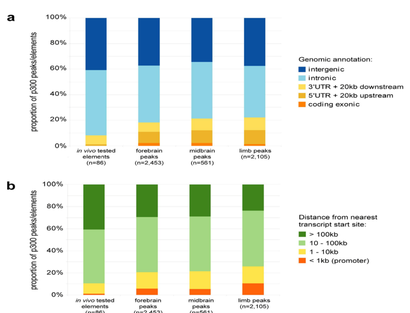


Overview of approach

- different tissues are collected from wild-type mouse embryos at embryonic day e11.5 – shown here: forebrain, midbrain, limb
- chromatin immunoprecipitation (ChIP) is performed directly from tissues, using an antibody directed against the p300 protein
- millions of sequence reads are obtained by massively-parallel sequencing and mapped to the mouse genome
- thousands of significantly p300-enriched genome regions ("peaks") are identified

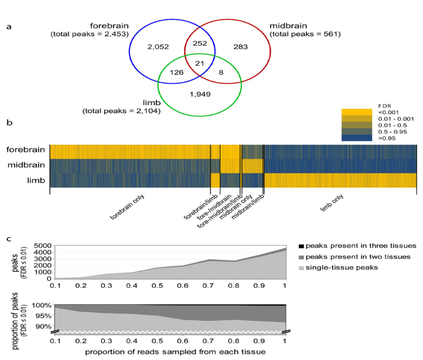
Genomic distribution of p300 peaks

- most peaks are intergenic or intronic, about 20% are close to transcript start or end sites, only a marginal fraction overlaps coding exons
- most peaks are located >10kb away from the nearest transcript start site



Most peaks are tissue-specific

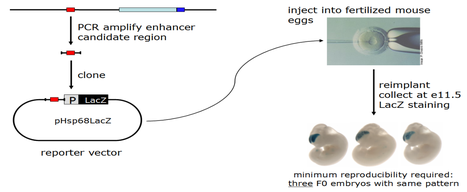
- among three tissues analyzed initially (forebrain, midbrain, limb), only 21 of 4,691 peaks (0.4%) were present in all three tissues, whereas 4,284 (91%) were significantly p300-enriched only in one of the three tissues
- re-sampling of subsets of reads suggests that deeper sequencing will identify additional peaks, most of them again only significantly enriched in a single tissue



3. P300 BINDING PREDICTS *IN VIVO* ENHANCERS

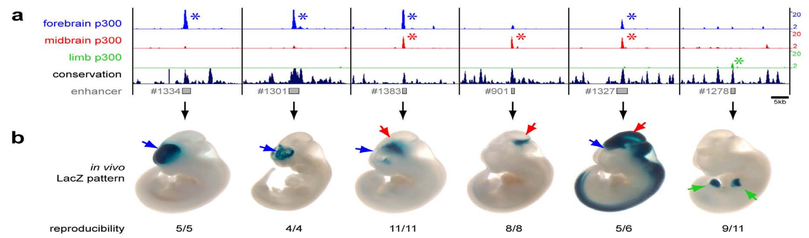
Transgenic enhancer assay

- p300 peaks are considered as enhancer candidate sequences
- human non-coding DNA fragments orthologous to mouse p300 peaks are coupled to an Hsp68 minimal promoter and a LacZ reporter gene
- transgenic mice (F_0 founder embryos) are generated by pronuclear injection and stained for reporter activity at e11.5

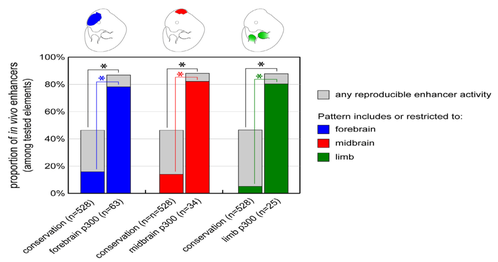


Prediction of *in vivo* enhancer activities

- presence or absence of p300 peaks in one of the three tissues correctly predicts *in vivo* reporter patterns observed in transgenic embryos (selected examples shown)



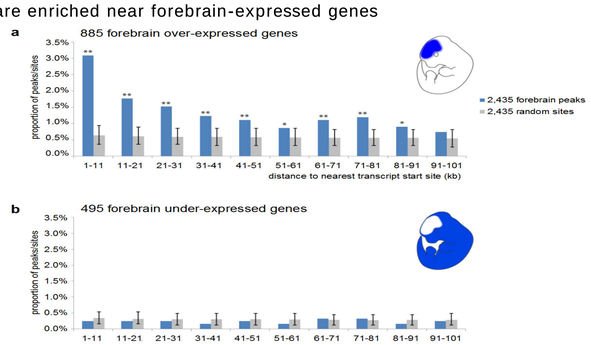
- in an initial large-scale assessment (n=86), we found that p300 peaks correctly predicted *in vivo* activity of enhancers in ~80% of cases, representing a dramatic improvement compared to conservation-based enhancer prediction methods



4. P300 PEAKS CORRELATE WITH TISSUE-SPECIFIC GENE EXPRESSION

Peaks are enriched near genes expressed in the same tissue

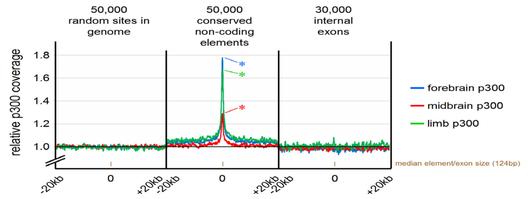
- genome-wide distribution of forebrain-p300 peaks was compared to microarray gene expression data from the same tissue
- forebrain p300 peaks are enriched near forebrain-expressed genes
- forebrain p300 peaks are NOT enriched near forebrain-underexpressed genes (compared to rest of body)
- a similar correlation was observed for limb peaks near limb-expressed genes (not shown)

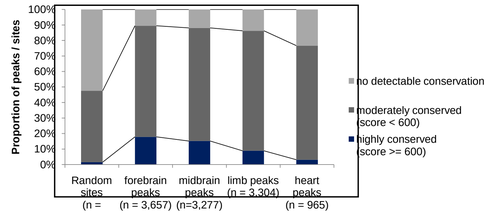


5. MOST P300 PEAKS ARE CONSERVED

p300 is enriched at conserved noncoding sequences

- extremely conserved noncoding sequences are enriched in p300 binding, consistent with their known enrichment in developmental enhancers (see Pennacchio et al. 2006; Visel et al. 2008)

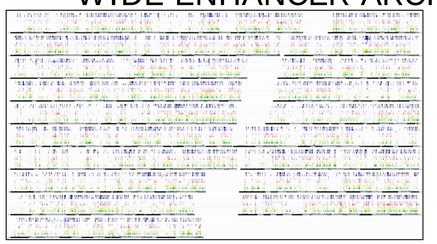




Most p300 peaks are constrained

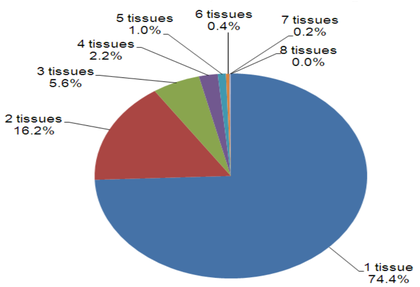
- the majority of p300-bound regions identified in forebrain, midbrain and limb overlap sequences that are under detectable evolutionary constraint in vertebrates
- proportion of extremely conserved peaks differs between tissues

6. A COMPREHENSIVE VIEW OF GENOME-WIDE ENHANCER ARCHITECTURE



Multiple Tissues

- based on the initial proof of principle, data for a total of 8 embryonic tissues was obtained (forebrain, midbrain, hindbrain, neural tube, heart, liver, limb, face; all at e11.5)
- 35 million non-redundant mapped reads identify 20,000 enhancer candidate sequences with predicted tissue-specific activities genome-
- most of the identified enhancer candidate sequences are bound by p300 only in one or two tissues
- very few regions are bound by p300 in most or all tissues, consistent with the hypothesis that these regions represent tissue-specific transcriptional enhancers
- large-scale transgenic validation of p300 peaks from all 8 tissues is in progress



LITERATURE AND RESOURCES

ChIP-seq identification of enhancers and activity patterns:

Visel et al. (2009), *Nature* 457:854-858

Visel et al. (2009) *Nature* 461:199-205.

Pennacchio and Visel (2010) *Nature Genetics* 42:557-8.

Blow et al. (2010) *Nature Genetics* 42:806-10.

Conservation-guided identification of developmental enhancers:

Visel et al. (2008) *Nature Genetics* 40:158-160

Pennacchio et al. (2006) *Nature* 444:499-502

Access to *In Vivo* Data:

http://enhancer.lbl.gov (Vista Enhancer Browser)

also see: Visel et al. (2007) *Nucleic Acids Research* 35:D88-92

This work was supported by the U.S. Department of Energy under Contract no. AC02-05CH11231

DISCLAIMER