Co-expression patterns define epigenetic regulators associated with neurological dysfunction

Kasper Daniel Hansen
<khansen@jhsph.edu | www.hansenlab.org>
Department of Biostatistics
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University
Genome (DNA) → Phenotype
Epigenetics (transcriptomics)

Genome (DNA) → Epigenetic marks → Expression → Environment → Phenotype → Cell-type → Time
Mendelian disorders of the epigenetic machinery

Genetic defect → Epigenetic defect → Phenotype

(hypothesis)

Recent interest in the epigenetic machinery

**Cancer:** Somatic mutations in EM genes are frequent in many cancers.

**Neurological:** GWAS and rare variant analysis has implicated EM genes in various neurological disorders incl. sz. and autism.
Shared phenotypes in EM disorders

Bjornsson (2015) Genome Research
Kabuki syndrome / intervening on the epigenome

Caused by LOF in KMT2D or KDM6A.

Can the intellectual disability associated with Kabuki syndrome be reversed by changing the epigenome?

The answer is yes  
(caveats: short-term, in mice, Kabuki type I)


Questions

1. Characterize the EM

2. Is the epigenetic function of the EM genes, the most likely cause of disease?

3. Are there expression signatures characteristic of disease candidates?

4. Are there distinct expression signatures between the EM genes involved in neurological dysfunction and cancer?

**Co-expression patterns define epigenetic regulators associated with neurological dysfunction**

Leandros Boukas, James M Havrilla, Aaron R Quinlan, Hans T Bjornsson, Kasper D Hansen

doi: https://doi.org/10.1101/219097
Defining the Epigenetic Machinery using protein domains

Any gene encoding a protein with a domain which can act as
- Reader / Writer / Eraser of DNA methylation.
- Reader / Writer / Eraser of histone methylation / acetylation.
- Chromatin remodeler

295 EM genes
Loss of function (LOF) mutations

We have two copies of each gene. Each copy produces mRNA at the same rate.

A produces mRNA
B produces mRNA

1/2 as much

“dosage sensitive” or “haploinsufficient”
EM genes are highly intolerant to LOF mutations

Using ExAC (Lek et al. 2016)
EM genes are very intolerant to LOF mutations

- EM (Enhancer Modulator) genes
- TF (Transcription Factor) genes
- All other genes

Highly constrained genes

$pLI$ Density

Graph showing density against $pLI$ for different gene categories:
- Blue line: All other genes
- Green line: TF genes

Highly constrained genes highlighted in grey.
EM genes are very intolerant to LOF mutations

OR = 7.7

Highly constrained genes

Density

2 4

0.1 0.5 0.9

pLI

All other genes
TF genes
EM genes
The epigenetic machinery and tissue expression

These epigenetic marks are present in every cell type and at every time point.

Genetic defects act in every cell where the gene is expressed.

The GTEx (genotype-tissue expression) project is profiling ~30 tissues in ~1000 people.
Testis is an outlier tissue

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Testis specificity score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRDM9</td>
<td>1</td>
</tr>
<tr>
<td>PRDM13</td>
<td>1</td>
</tr>
<tr>
<td>PRDM14</td>
<td>1</td>
</tr>
<tr>
<td>CDY2A</td>
<td>1</td>
</tr>
<tr>
<td>BRDT</td>
<td>0.87</td>
</tr>
<tr>
<td>RNF17</td>
<td>0.86</td>
</tr>
<tr>
<td>HDGFL1</td>
<td>0.81</td>
</tr>
<tr>
<td>PRDM7</td>
<td>0.77</td>
</tr>
<tr>
<td>MORC1</td>
<td>0.75</td>
</tr>
<tr>
<td>TDRD15</td>
<td>0.67</td>
</tr>
<tr>
<td>TDRD1</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Tissue specificity scores for various genes in different tissues.
Motivation for co-expression

1 tissue in GTEx

All of GTEx
Co-expression; tissue-specific networks and modules

- **Number of partners**: 20, 70, 120
- **pLI**: 0.1, 0.5, 0.9

For tissues 1, 2, and 28, there are modules labeled as Module 1 and Module 2, with genes A, B, C, and D represented.

- **Density** comparison between random genes, EM genes, and observed genes.
  - **Percentage with pLI > 0.9**: Highly co-expressed, Co-expressed, Not co-expressed.

WGCNA
Removing unwanted variation in co-expression networks

Addressing confounding artifacts in reconstruction of gene co-expression networks

Princy Parsana, Claire Ruberman, Andrew E. Jaffe, Michael C. Schatz, Alexis Battle, Jeffery T. Leek

doi: https://doi.org/10.1101/202903

Simple solution: remove the top singular values; they will represent artifacts

Systematic noise degrades gene co-expression signals but can be corrected

Saskia Freytag\textsuperscript{1,2*}, Johann Gagnon-Bartsch\textsuperscript{3}, Terence P. Speed\textsuperscript{1,2,3} and Melanie Bahlo\textsuperscript{1,2,4}

How do we measure if it works?
Random groups of genes

Supplementary Figure S5. Removing noise in co-expression analysis by removing principal components.

We remove unwanted variation in our co-expression analysis by removing 4 principal components from the expression matrix in all tissues. (a) The distribution of pairwise correlations between randomly sampled genes, serving as a negative control, for 9 out of the 28 tissues. (b) The distribution of pairwise correlations between 80 genes coding for ribosomal proteins, which serve as a positive control, for the same tissues as in (a).
Supplementary Figure S5. Removing noise in co-expression analysis by removing principal components.

We remove unwanted variation in our co-expression analysis by removing 4 principal components from the expression matrix in all tissues.

(a) The distribution of pairwise correlations between randomly sampled genes, serving as a negative control, for 9 out of the 28 tissues.

(b) The distribution of pairwise correlations between 80 genes coding for ribosomal proteins, which serve as a positive control, for the same tissues as in (a).
Co-expression; tissue-specific networks and modules

Module partners across tissues

- not partners
- partners
- stable partners

Number of partners

- 20
- 70
- 120

pLI

- 0.1
- 0.5
- 0.9

Density

- random genes
- EM genes

Number of genes with >= 75 module partners

Percentage with pLI > 0.9

Highly co-expressed
Co-expressed
Not co-expressed

Observed
Co-expression is associated with LOF intolerance.

![Graph showing co-expression and LOF intolerance](image)

**Number of partners**
- 20
- 70
- 120

**pLI**
- 0.1
- 0.5
- 0.9

**Module partners across tissues**
- (a)
- (b)
- (c)
- (d)
- (e)
- (f)

**Number of genes with >= 75 module partners**

**Density**

**Percentage with pLI > 0.9**

**Highly co-expressed**

**Co-expressed**

**Not co-expressed**

**EM Genes**

**Observed**
Co-expression is associated with LOF intolerance

Figure 4. A large subset of the components of the epigenetic machinery are highly co-expressed. (a) Schematic illustrating our definition and identification of module partners. WGCNA was used to construct tissue-specific co-expression networks and modules for 28 tissues profiled in GTEx. We determined if two EM genes were module partners (part of the same module in 10-14 tissues) or stable module partners (part of the same module in >14 tissues). (b) The number of module partners for each EM gene, in decreasing order. (c) The module partner matrix for EM genes, ordered by its number of module partners (b). We define 3 groups of EM genes, “highly co-expressed”, “co-expressed” and “not co-expressed” based on their number of module partners. (d) The pLI for each EM gene, ordered by the its number of module partners as in (b). (e) The size of the (highly) co-expressed group of EM genes compared to 300 draws of 270 random genes, where the random genes are selected to have a similar expression level across tissues compared to EM genes (Supplementary Figure S7).
Permutations

number of partners

genes

EM
Transcription factors
Kinases/Phosphorylases
Co-expression is associated with LOF intolerance
Co-expression is associated with neurological dysfunction
Acknowledgements

Leandros Boukas

James Havrilla

Hans Bjornsson

Aaron Quinlan

Co-expression patterns define epigenetic regulators associated with neurological dysfunction

Leandros Boukas, James M Havrilla, Aaron R Quinlan, Hans T Bjornsson, Kasper D Hansen

doi: https://doi.org/10.1101/219097

(looking for postdocs)
Domains

(a) Distribution of CCR local constraint for high and low pLI genes.

(b) Box plot showing mean difference in CCR local constraint between high and low pLI genes.

(c) Range of CCR local constraint for various domains:
- Chromodomain
- Zinc finger, PHD type
- Tudor domain
- PWWP domain
- Mbt repeat
- WD40 repeat
- Ankyrin repeat
- Bromodomain
- Zinc finger, CXXC type