Systematic Localization of Common Disease-Associated Variation in Regulatory DNA
Matthew T. Maurano et al.
Science 337, 1190 (2012);
DOI: 10.1126/science.1222794

Systematic Localization of Common Disease-Associated Variation in Regulatory DNA
Ma Xiaopeng
2012-12-6
Stam lab

John Stamatoyannopoulos

**Science**
September 7, 2012
Systematic Localization of Common Disease-Associated Variation in Regulatory DNA

**Cell**
September 14, 2012
Circuitry and Dynamics of Human Transcription Factor Regulatory Networks

**Nature**
September 6, 2012
An extensive human regulatory lexicon encoded in transcription factor footprints

**Nature**
September 6, 2012
An integrated encyclopedia of DNA elements in the human genome

**Nature**
September 6, 2012
The accessible chromatin landscape of the human genome

**Genome Res**
September 4, 2012
Personal and Population Genomics of Human Regulatory Variation

**Genome Res**
September 4, 2012
What does our genome encode?

**Genome Res**
September 4, 2012
Widespread plasticity in CHG occupancy linked to DNA methylation
Background: GWAS

A Catalog of Published Genome-Wide Association Studies
http://www.genome.gov/gwastudies/
Background: DHS

- **DNase I**
  - Hypersensitivity
- **DHS**
  - DNase I hypersensitive sites
- **DNase-seq**
  - A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome

From Guo Weilong

The accessible chromatin landscape of the human genome

Lingyun Song and Gregory E. Crawford, (2010)
Data I: GWAS

• GWAS
  – 920 GWAS studies
  – 207 diseases and 447 quantitative traits
  – 6,011 trait- SNP associations (nocoding 5,654)
  – 5,386 distinct SNPs (nocoding 5,134)
GWAS

A

920 Genome Wide Association Studies

- Neurological/behavioral: 13%
- Parasitic/bacterial Disease: 1%
- Quantitative traits: 7%
- Radiographic parameters: 1%
- Serum Metabolites: 11%
- Viral Disease: 2%
- Aging: 9%
- Autoimmune disease: 12%
- Cancer: 12%
- Miscellaneous: 7%
- Lipid: 3%
- Kidney, Lung, Liver: 3%
- Hematologic parameters: 3%
- Drug metabolism: 4%
- Diabetes: 4%
- Cardiovascular: 9%

B

Distribution of GWAS SNPs vs. RefSeq

- >1Mb: 1.2%
- 1-100Mb: 4.9%
- >100kb-1Mb: 20.2%
- >50-100kb: 7.8%
- >1-50kb: 23.4%
- Coding: 4.9%
- Introns: 41.2%
- Promoter: 1.4%
Data II: DHS

• DHSs
  – 349 cell and tissue samples
    • 85 cell types in ENCODE
    • 264 samples from Roadmap Epigenomics Program
  – 233 diverse fetal tissue samples (~60 to 160d)
<table>
<thead>
<tr>
<th>Cell_line</th>
<th>Description</th>
<th>Fetal?</th>
<th>#DHSs</th>
<th>#SNPs</th>
<th>Pub?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A549</td>
<td>Epithelial cell line derived from a lung carcinoma tissue</td>
<td>N</td>
<td>117,992</td>
<td>180</td>
<td>Y</td>
</tr>
<tr>
<td>AG04449</td>
<td>Fetal buttock/thigh fibroblast</td>
<td>Y</td>
<td>174,802</td>
<td>202</td>
<td>Y</td>
</tr>
<tr>
<td>AG04450</td>
<td>Fetal lung fibroblast</td>
<td>Y</td>
<td>150,114</td>
<td>187</td>
<td>Y</td>
</tr>
<tr>
<td>AG09309</td>
<td>Adult human toe fibroblast</td>
<td>N</td>
<td>197,301</td>
<td>266</td>
<td>Y</td>
</tr>
<tr>
<td>AG09319</td>
<td>Adult human gum tissue fibroblasts</td>
<td>N</td>
<td>137,192</td>
<td>190</td>
<td>Y</td>
</tr>
<tr>
<td>AG10803</td>
<td>Adult human abdominal skin fibroblasts</td>
<td>N</td>
<td>171,903</td>
<td>224</td>
<td>Y</td>
</tr>
<tr>
<td>AoAF</td>
<td>Normal Human Aortic Adventitial Fibroblast Cells</td>
<td>N</td>
<td>169,477</td>
<td>261</td>
<td>Y</td>
</tr>
<tr>
<td>BE2_C</td>
<td>Human Neuroblastoma cell line</td>
<td>N</td>
<td>168,003</td>
<td>259</td>
<td>Y</td>
</tr>
<tr>
<td>BJ</td>
<td>Skin fibroblasts</td>
<td>N</td>
<td>162,671</td>
<td>246</td>
<td>Y</td>
</tr>
</tbody>
</table>
## DHS

<table>
<thead>
<tr>
<th></th>
<th># DHS position</th>
<th>% genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3,899,693</td>
<td>42.2%</td>
</tr>
<tr>
<td>Every cell</td>
<td>198,180</td>
<td>~2.1%</td>
</tr>
</tbody>
</table>
Representative of GWAS SNP in DHSs
GWAS SNPs is concentrated in DHS

Data showing that GWAS SNPs are concentrated in DHS regions. The pie chart and bar graph illustrate the distribution and replication status of these SNPs.
Proportions of GWAS SNPs in fetal-origin DHSs

- SNPs in DHSs found only in adult tissues (n=348) - 11.9%
- SNPs in fetal-origin DHSs (n=1,696) - 57.8%
- SNPs in fetal-specific DHSs (n=887) - 30.3%
GWAS SNP in DHSs shows phenotype-specific enrichment for fetal regulatory elements.

* P < 0.05 (binomial)

Baseline = 88.1% GWAS SNPs in fetal–origin DHSs
GWAS SNP-regulated genes

• An example
## GWAS SNP-regulated genes

<table>
<thead>
<tr>
<th>Disease or trait</th>
<th>$r$</th>
<th>Target gene</th>
<th>Function</th>
<th>Distance (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>1</td>
<td>SYNGAP1*</td>
<td>Axon formation; component of NMDA complex</td>
<td>411</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>1</td>
<td>TRIB1*</td>
<td>NF-κB regulation</td>
<td>95</td>
</tr>
<tr>
<td>Time to first primary tooth</td>
<td>0.99</td>
<td>PRDM1*</td>
<td>Craniofacial development</td>
<td>452</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>0.99</td>
<td>NLRP3</td>
<td>Response to bacterial pathogens</td>
<td>20</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>0.98</td>
<td>AHI1*</td>
<td>White matter abnormalities</td>
<td>149</td>
</tr>
<tr>
<td>QRS duration</td>
<td>0.96</td>
<td>SCN10A*</td>
<td>Sodium channel involved in cardiac conduction</td>
<td>181</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>0.96</td>
<td>TACC2*</td>
<td>Tumor suppressor</td>
<td>411</td>
</tr>
<tr>
<td>Schizophrenia/brain imaging</td>
<td>0.95</td>
<td>KIF1A*</td>
<td>Neuron-specific kinesin involved in axonal transport</td>
<td>428</td>
</tr>
<tr>
<td>Brain structure</td>
<td>0.94</td>
<td>CXCR6*</td>
<td>Chemokine receptor involved in glial migration</td>
<td>357</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>0.94</td>
<td>CTSB*</td>
<td>Cysteine proteinase linked to articular erosion</td>
<td>359</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>0.93</td>
<td>HSPG2*</td>
<td>Ovarian tumor suppressor</td>
<td>268</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>0.93</td>
<td>ZP1*</td>
<td>Known autoantigen</td>
<td>153</td>
</tr>
<tr>
<td>ADHD</td>
<td>0.93</td>
<td>PDLIM5*</td>
<td>Neuronal calcium signaling</td>
<td>328</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>0.88</td>
<td>MAP3K1*</td>
<td>Response to growth factors</td>
<td>158</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>0.88</td>
<td>CNTN4</td>
<td>Neuronal cell adhesion</td>
<td>306</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>0.81</td>
<td>FXR1*</td>
<td>Cognitive function</td>
<td>120</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>0.75</td>
<td>ACAD10*</td>
<td>Mitochondrial oxidation of fatty acids</td>
<td>343</td>
</tr>
<tr>
<td>Lupus</td>
<td>0.74</td>
<td>STAT4</td>
<td>Mediates IL-12 immune response and Th1 differentiation</td>
<td>113</td>
</tr>
</tbody>
</table>
Heterogeneous allele: allele-specific DNase I sensitivity

HCPEpiC  T186: G16

HPF  T138: G9

SAEC  T276: G144

HAEpiC  T29: C13

SAEC  T46: C26

HCT116  T28: C19

CEBP A  T

Smad3  T
Common disease-associated variants cluster in regulatory network
De novo identification of pathogenic cell types

Crohn’s disease

- immune cells (n=15)
- CD34+ (n=1)
- thymus (n=10)
- ES/primitive (n=9)
- intestine (n=28)
- other (n=268)

Fold enrichment of SNPs in DHSs

GWAS P-value threshold
Take-home messages

• Most noncoding GWAS SNP in DHS
• 88% such DHS are active during fetal development
• ChIA-PET/ 5C: a powerful tool to find SNP’s target gene
• De novo identification of pathogenic cell
• Thank you!