From genomics to medicine: Targeting the regulatory genome in GWAS & Cancer

Manolis Kellis

Broad Institute of MIT and Harvard
MIT Computer Science & Artificial Intelligence Laboratory
Genomic medicine: challenge and promises

1. The promise of genetics
   - Disease mechanism
   - New target genes
   - New therapeutics
   - Personalized medicine

2. The challenge
   - 90+% disease hits non-coding
   - Cell type of action not known
   - Causal variant not known
   - Mechanism not known

GWAS: simple $\chi^2$ statistical test

Hillmer Nature Genetics 2008
Genomic medicine: challenge and promises

3. The remedy
- Annotation of non-coding genome (ENCODE/Roadmap)
- Linking of enhancers to regulators and target genes
- New methods for utilizing them


4. The deliverables
- Relevant cell type
- Target genes
- Causal variant
- Upstream regulator
- Relevant pathways
- Intermediate phenotypes

This talk: From genomics to medicine

Part 1: Decoding the regulatory genome

Enhancers: Enhancers, promoters, transcribed, repressed
Cell types: Predict tissues and cell types of epigenomic activity
Target genes: Link variants to their target genes using eQTLs, activity, Hi-C
Nucleotides: Regulatory consequence of mutation: Conservation, PWMs
Regulators: Upstream regulators whose activity is disrupted by mutation

Part 2: Application to discover new therapeutic leads

GWAS hits:
- CATGCCTG
- CGTGCTTA
  - 93% top hits non-coding → Mechanism? Cell type?
  - Lie in haplotype blocks → Causal variant(s)?

Discover new loci:
- CATGCCTG
- CGTGCTTA
  - Many loci undetected → Pathway-level burden/load
  - Many false positives → Prioritize w/ regulatory annotations

Cancer mutations:
- CATGCCTG
- CATCCTG
  - Loss of function → Protein-coding variants, convergence
  - Gain of function → Regulatory variants, heterogeneity
From genomics to medicine in GWAS & Cancer

1. Characterize the epigenomic landscape
   - Map non-coding elements, uncover cellular circuitry
   - Identify disease-relevant tissues and regulators
     - Immune/inflammatory drivers in Alzheimer’s disease

2. Mechanistic dissection of non-coding disease loci
   - Identify cell type, causal SNP, regulator, targets, pathways
   - Genome editing: adipocyte browning drivers of obesity

3. Genetic + epigenomic discovery of new disease loci
   - Prioritize and experimentally validate variants missed by GWAS
   - New heart repolarization disease genes with cardiac functions

4. Non-coding discovery of new cancer driver genes
   - Gain of function mutations by out-of-context de-repression
   - Prostate cancer drivers: immune evasion, signaling, energy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   - Thousands of weak sites show genetic + epigenetic evidence
Integrative analysis of 111 reference human epigenomes

Roadmap Epigenomics Consortium†, Anshul Kundaje1,2,3, Wouter Meuleman1,2,*, Jason Ernst1,2,4,*, Misha Bilenky5,*, Angela Yen1,2, Alireza Heravi-Moussavi5, Pouya Kheradpour1,2, Zhizhuo Zhang1,2, Jianrong Wang1,2, Michael J. Ziller2,6, Viren Amin7, John W. Whitaker9, Matthew D. Schultz9, Lucas D. Ward1,2, Abhishek Sarkar1,2, Gerald Quon12, Richard S. Sandstrom10, Matthew L. Eaton1,2, Yi-Chieh Wu1,2, Andreas R. Pfenning1,2, Xinchen Wang1,2,11, Melina Claussnitzer1,2, Yaping Liu1,2, Cristian Coarfa7, R. Alan Harris7, Noam Shoresh2, Charles B. Epstein, Elizabeta Gjoneska2,12, Danny Leung8,13, Wei Xie8,13, R. David Hawkins8,13, Ryan Lister9, Chibo Hong14, Philippe Gascard15, Andrew J. Mungall5, Richard Moore8, Eric Chuah8, Angela Tam9, Theresa K. Canfield10, R. Scott Hansen16, Rajinder Kaul16, Peter J. Sabo16, Mukul S. Bansal1,2,17, Annaick Carles18, Jesse R. Dixon8,13, Kai–How Farh2, Soheil Feizi1,2, Rosa Karlic19, Ah-Ram Kim1,2, Ashwinikumar Kulkarni20, Daofeng Li21, Rebecca Lowdon21, GiNell Elliott21, Tim R. Mercer22, Shane J. Neph10, Vitor Onuchic23, Paz Polak2,23, Nisha Rajagopal8,13, Pradipta Ray20, Richard C. Sallari1,2, Kyle T. Siebenthal18, Nicholas A. Sinnott-Armstrong1,2, Michael Stevens21,42, Robert E. Thurman10, Jie Wu24,25, Bo Zhang21, Xin Zhou21, Arthur E. Beaudet26, Laurie A. Boyer11, Philip L. De Jager2,27, Peggy J. Farnham28, Susan J. Fisher29, David Haussler30, Steven J. M. Jones5,31,32, Wei Li33, Marco A. Marra5,32, Michael T. McManus34, Shamil Sunyaev2,23,27, James A. Thompson35,41, Thea D. Tlsty15, Li–Huei Tsai2,12, Wei Wang8, Robert A. Waterland36, Michael Q. Zhang20,37, Lisa H. Chadwick38, Bradley E. Bernstein2,39,40, Joseph F. Costello14,5, Joseph R. Ecker9, Martin Hirst5,18,5, Alexander Meissner2,46,6, Aleksandar Milosavljevic7, Bing Ren8,13, John A. Stamatoyannopoulos10, Ting Wang21,2 & Manolis Kellis1,2,8

19 FEBRUARY 2015 | VOL 518 | NATURE | 317

Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Lisa Chadwick, Brad Bernstein, Joe Costello, Joe Ecker, Martin Hirst, Alex Meissner, Aleks Milosavljevic, Bing Ren, John Stam, Ting Wang
Epigenomic mapping across 100+ tissues/cell types

**Diverse tissues and cells**

- Adult tissues and cells (brain, muscle, heart, digestive, skin, adipose, lung, blood...)
- Fetal tissues (brain, skeletal muscle, heart, digestive, lung, cord blood...)
- ES cells, iPS, differentiated cells (meso/endo/ectoderm, neural, mesench...)

**Diverse epigenomic assays**

- Histone modifications
  - H3K4me3, H3K4me1, H3K36me3
  - H3K27me3, H3K9me3, H3K27/9ac
  - +20 more
- Open chromatin:
  - DNA accessibility
- DNA methylation:
  - WGBS, RRBS, MRE/MeDIP
- Gene expression
  - RNA-seq, Exon Arrays
Summarize multiple marks into chromatin states

ChromHMM: multi-variate hidden Markov model
Enhancer modules, regulators, and target genes

1. Map chromatin states across 127 tissue/cells

2. Group enhancers into modules of common function

3. Predict module regulators using motif enrichment

4. Predict target genes using common activity
Use resulting annotations and networks for GWAS

- Expand each GWAS locus using SNP linkage disequilibrium (LD)
  - Recognize **relevant cell types**: tissue-specific enhancer enrichment
  - Recognize **driver TFs**: enriched motifs in multiple GWAS loci
  - Recognize **target genes**: linked to causal enhancers
GWAS hits in enhancers of relevant cell types

<table>
<thead>
<tr>
<th>Trait</th>
<th>Abbrev</th>
<th>log2FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>ENSC</td>
<td>4.7</td>
</tr>
<tr>
<td>Height</td>
<td>ENSC</td>
<td>4.0</td>
</tr>
<tr>
<td>Chronic lymphocytic leukaemia</td>
<td>ENSC</td>
<td>6.4</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>Treg</td>
<td>4.8</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>Treg</td>
<td>4.1</td>
</tr>
<tr>
<td>Platelet counts</td>
<td>Th.slm</td>
<td>4.8</td>
</tr>
<tr>
<td>Chronic lymphocytic leukaemia</td>
<td>Th.slm</td>
<td>4.8</td>
</tr>
<tr>
<td>Self-reported allergy</td>
<td>Th.slm</td>
<td>4.9</td>
</tr>
<tr>
<td>Colitis</td>
<td>Th.slm</td>
<td>4.3</td>
</tr>
<tr>
<td>Celiac disease</td>
<td>Th.sln</td>
<td>4.3</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Th.Tst</td>
<td>4.2</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Th.mm</td>
<td>3.6</td>
</tr>
<tr>
<td>Celiac disease + rheumatoid arthritis</td>
<td>Th.mm</td>
<td>4.6</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>Th.mm</td>
<td>0.6</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>Ecor</td>
<td>4.4</td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
<td>Ecor</td>
<td>3.9</td>
</tr>
<tr>
<td>Red blood cell traits</td>
<td>HS.comp</td>
<td>0.9</td>
</tr>
<tr>
<td>Platelet counts</td>
<td>HS.comp</td>
<td>0.0</td>
</tr>
<tr>
<td>Mean platelet volume</td>
<td>HS.comp</td>
<td>0.0</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Ep.r</td>
<td>0.6</td>
</tr>
<tr>
<td>Mitochondrial disorder</td>
<td>Ep.r</td>
<td>4.7</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Nk.r</td>
<td>4.0</td>
</tr>
<tr>
<td>Mean platelet volume</td>
<td>Nk.r</td>
<td>0.2</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>Fat</td>
<td>4.2</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>Fat</td>
<td>4.9</td>
</tr>
<tr>
<td>Height</td>
<td>Fat</td>
<td>4.8</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>Tym</td>
<td>4.2</td>
</tr>
<tr>
<td>Adiponectin levels</td>
<td>Brain</td>
<td>4.3</td>
</tr>
<tr>
<td>Attention deficit hyperactivity disorder</td>
<td>Brain</td>
<td>4.5</td>
</tr>
<tr>
<td>Perinuclear</td>
<td>Heart</td>
<td>4.7</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Heart</td>
<td>4.6</td>
</tr>
<tr>
<td>Aortic root size</td>
<td>Vasc</td>
<td>4.1</td>
</tr>
<tr>
<td>Pulmonary function</td>
<td>Smp</td>
<td>4.2</td>
</tr>
<tr>
<td>Liver enzyme levels (g-glut 1x)</td>
<td>Gal</td>
<td>4.5</td>
</tr>
<tr>
<td>Urate levels</td>
<td>Gmuc</td>
<td>4.0</td>
</tr>
<tr>
<td>Adv. resp. to chemot (neutrophil)</td>
<td>Stom</td>
<td>4.3</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Stom</td>
<td>4.3</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>Stom</td>
<td>4.3</td>
</tr>
<tr>
<td>Insulin-like growth factors</td>
<td>Plac</td>
<td>4.2</td>
</tr>
<tr>
<td>Cytosolic Ca dissociated states</td>
<td>Plac</td>
<td>4.1</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>Liver</td>
<td>4.9</td>
</tr>
<tr>
<td>Cholesterol, total</td>
<td>Liver</td>
<td>0.0</td>
</tr>
<tr>
<td>Cholesterol, total</td>
<td>Liver</td>
<td>0.0</td>
</tr>
<tr>
<td>Lipid metabolism phenotypes</td>
<td>Liver</td>
<td>6.8</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>Liver</td>
<td>0.7</td>
</tr>
<tr>
<td>Cholesterol, total</td>
<td>Liver</td>
<td>4.8</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>Liver</td>
<td>3.9</td>
</tr>
<tr>
<td>Metabolite levels</td>
<td>Liver</td>
<td>3.9</td>
</tr>
<tr>
<td>Platelet counts</td>
<td>Leuk</td>
<td>4.6</td>
</tr>
<tr>
<td>Primary biliary cirrhosis</td>
<td>Leuk</td>
<td>4.7</td>
</tr>
<tr>
<td>Liver parenchymal volume</td>
<td>Leuk</td>
<td>4.7</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
<td>Mncyt</td>
<td>14.6</td>
</tr>
<tr>
<td>Intestinal cell</td>
<td>Mncyt</td>
<td>9.3</td>
</tr>
<tr>
<td>Alzheimer's disease (data missed)</td>
<td>Mncyt</td>
<td>4.9</td>
</tr>
<tr>
<td>Freecampal</td>
<td>Bone</td>
<td>4.5</td>
</tr>
</tbody>
</table>
From genomics to medicine in GWAS & Cancer

1. Characterize the epigenomic landscape
   - Map non-coding elements, uncover cellular circuitry
   - Identify disease-relevant tissues and regulators
   - Immune/inflammatory drivers in Alzheimer’s disease

2. Mechanistic dissection of non-coding disease loci
   - Identify cell type, causal SNP, regulator, targets, pathways
   - Genome editing: adipocyte browning drivers of obesity

3. Genetic + epigenomic discovery of new disease loci
   - Prioritize and experimentally validate variants missed by GWAS
   - New heart repolarization disease genes with cardiac functions

4. Non-coding discovery of new cancer driver genes
   - Gain of function mutations by out-of-context de-repression
   - Prostate cancer drivers: immune evasion, signaling, energy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   - Thousands of weak sites show genetic + epigenetic evidence
Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer’s disease

Elizabeta Gjoneska¹,²*, Andreas R. Pfenning²,³*, Hansruedi Mathys¹, Gerald Quon²,³, Anshul Kundaje²,³,⁴, Li-Huei Tsai¹,²§ & Manolis Kellis²,³§
Immune activation + neural repression in human + mouse

Two contrasting signatures of immune activation vs. neural repression

Sample mouse brain epigenomics during neurodegeneration

Is inflammation simply a consequence of neuronal loss?
Genetic evidence for immune vs. neuronal components

Indicates immune cell dysregulation is causal component

Microglial cells: resident immune cells of adult brain
Macrophages: infiltrate brain in neurodegeneration

Only increasing (immune) enhancers enriched in AD-associated SNPs

Neuronal cell types are depleted for AD-associated SNPs
HaploReg: systematic mining of GWAS variants

- Start with any list of SNPs or select a GWA study
  - Mine ENCODE and Roadmap epigenomics data for hits
  - Hundreds of assays, dozens of cells, conservation, motifs
  - Report significant overlaps and link to info/browser

- Try it out: [http://compbio.mit.edu/HaploReg](http://compbio.mit.edu/HaploReg)

Ward, Kellis NAR 2011
From genomics to medicine in GWAS & Cancer

1. Characterize the epigenomic landscape
   - Map non-coding elements, uncover cellular circuitry
   - Identify disease-relevant tissues and regulators
   - Immune/inflammatory drivers in Alzheimer’s disease

2. Mechanistic dissection of non-coding disease loci
   - Identify cell type, causal SNP, regulator, targets, pathways
   - Genome editing: adipocyte browning drivers of obesity

3. Genetic + epigenomic discovery of new disease loci
   - Prioritize and experimentally validate variants missed by GWAS
   - New heart repolarization disease genes with cardiac functions

4. Non-coding discovery of new cancer driver genes
   - Gain of function mutations by out-of-context de-repression
   - Prostate cancer drivers: immune evasion, signaling, energy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   - Thousands of weak sites show genetic + epigenetic evidence
2. Mechanistic dissection of a non-coding disease locus

- Identify cell type, causal SNP, regulator, targets, process
- Genome editing demonstrates variant causality
- Adipocyte browning drivers of obesity

Melina Claussnitzer
1. Establish relevant **tissue/cell type**
2. Establish downstream **target** gene(s)
3. Establishing **causal** nucleotide variant
4. Establish upstream **regulator** causality
5. Establish **cellular** phenotypic consequences
6. Establish **organismal** phenotypic consequences

This talk: Apply these to the FTO locus
FTO region: strongest association with obesity

- Associated with **obesity**, Type 2 Diabetes, Cardiovascular traits
- 89 variants in LD, spanning 47kb, intron 1 of FTO gene
- No protein-altering variants: regulatory role? Target gene, tissue?
1. Tissue: Chromatin states predict adipocyte function

Progenitors of white/beige adipocytes
2. Targets: 3D folding and expr. genetics indicate IRX3+IRX5

Cohort of 20 homozygous risk and 18 homozygous non-risk individuals: Genotype-dependent expression?

Topological domains span 2.5Mb
Implicate 8 candidate genes

eQTL targets: IRX3 and IRX5

Risk allele: increased expression (gain-of-function)
3. Causal SNP: motif enrichment + conservation: rs1421085

Regulatory motifs enriched in BMI GWAS hits

Causal nucleotide rs1421085: risk alters T to C, abolishes AT-rich motif

rs1421085 T -> C obesity risk allele chr16: 53800954
4. Regulator: Causality and epistasis of ARID5B repressor

Regulatory model: risk allele disrupts a repressor

- Risk (C) rs1421085
- Non-risk (T) rs1421085

Cis/trans conditional analysis

- Repression of enhancer, IRX3 and IRX5 all require both TF and motif
- Disrupting motif (CC), or repressing ARID5B (siRNA) \(\rightarrow\) de-repression
Steps 5-6. Does this circuitry actually lead to obesity?

1. Establish relevant tissue/cell type: pre-adipocytes
2. Establish downstream target gene(s): IRX3 and IRX5
3. Establishing causal nucleotide variant: rs1421085
4. Establish upstream regulator causality: ARID5B
5. Establish cellular phenotypic consequences
6. Establish organismal phenotypic consequences
Expression analysis to recognize target processes

Search for genes co-expressed with IRX3 and IRX5 (n=20 indiv.)

Risk allele: shift from dissipation to storage

Negative correlation: mitochondria
Positive correlation: lipid storage

Reflected in cellular phenotypes

Risk carriers: increased mito
Non-risk: increased adipocytes

Risk allele: shift from dissipation to storage
Test model by systematic perturbations

- C-to-T motif rescue (anti-obesity phenotypes)
- T-to-C motif disruption (pro-obesity phenotypes)

- ARID5B KD (obesity)
- ARID5B OE (anti-obesity)
- IRX3, IRX5 knock-down★ (anti-obesity phenotypes)
- IRX3, IRX5 overexpression (pro-obesity phenotypes)

**Lean**

**Obese**

**Thermogenic stimuli (e.g. cold)**

- UCP1
- PGC1α
- PRDM16

**Browning mitochondrial thermogenesis**

**AATATT motif**

**White adipocytes**

**Lipid storage**
Irx3 adipose repression: anti-obesity phenotypes in mice

54% reduced body weight

Resistance to high-fat diet

Increased energy dissipation
- No reduction in appetite
- No increase in exercise
- In thermoneutral conditions
- Day and night (not exercise)
Single-nucleotide editing reverses thermogenesis in humans

rs1421085 editing alters IRX3+IRX5 expression (500,000 and 1 million nucleotides away!)

rs1421085 causality: C-to-T editing rescues IRX3/IRX5 expression, ARID5B repression, thermogenesis, developmental expression
Model: beige ↔ white adipocyte development

Shift therapeutic focus from brain to adipocytes
1. Establish relevant tissue/cell type: pre-adipocytes
2. Establish downstream target gene(s): IRX3 and IRX5
3. Establishing causal nucleotide variant: rs1421085
4. Establish upstream regulator causality: ARID5B
5. Establish cellular phenotypic consequences: thermogenesis
6. Establish organismal phenotypic consequences: body weight
The FTO locus is not unusual, 100s to go!

Across 11 well-powered association studies:
- BMI
- Bone Mineral Density
- Bipolar
- BreastCancer
- Cholesterol
- CrohnIBDUC
- Height
- Menarche
- QT
- RheumatoidArthritis
- Schizophrenia

895 associated loci

572 (64%) have no protein-coding variants
From genomics to medicine in GWAS & Cancer

1. Characterize the epigenomic landscape
   – Map non-coding elements, uncover cellular circuitry
   – Identify disease-relevant tissues and regulators
   – Immune/inflammatory drivers in Alzheimer’s disease

2. Mechanistic dissection of non-coding disease loci
   – Identify cell type, causal SNP, regulator, targets, pathways
   – Genome editing: adipocyte browning drivers of obesity

3. Genetic + epigenomic discovery of new disease loci
   – Prioritize and experimentally validate variants missed by GWAS
   – New heart repolarization disease genes with cardiac functions

4. Non-coding discovery of new cancer driver genes
   – Gain of function mutations by out-of-context de-repression
   – Prostate cancer drivers: immune evasion, signaling, energy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   – Thousands of weak sites show genetic + epigenetic evidence
3. Genetic + epigenomic discovery of new disease loci

• Prioritize and validate variants missed by GWAS
• New heart repolarization genes with cardiac functions

Xinchen Wang
Characterizing sub-threshold variants in heart arrhythmia

Focus on sub-threshold variants (e.g. rs1743292 P=10^{-4.2})

Trait: QRS/QT interval

(1) Large cohorts, (2) many known hits
(3) well-characterized tissue drivers
Enhancers overlapping GWAS loci share functional properties

<table>
<thead>
<tr>
<th>Enhancer characteristic</th>
<th>GWAS (red) vs. all LV enhancers (blue)</th>
<th>Fold difference</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K27ac density</td>
<td></td>
<td>3.10</td>
<td>1.54x10^-4</td>
</tr>
<tr>
<td>Activity in cardiac tissues (Proportion overlap)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. heart</td>
<td>R. atrium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. ventricle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity in non-cardiac tissues (# non-cardiac tissues with activity)</td>
<td></td>
<td>0.59</td>
<td>9.05x10^-3</td>
</tr>
<tr>
<td>LV-specific hypomethylation (Proportion of LV-specific hypomethylated CpGs)</td>
<td></td>
<td>2.34</td>
<td>1.07x10^-5</td>
</tr>
<tr>
<td>LV-specific hypermethylation (Proportion of LV-specific hypermethylated CpGs)</td>
<td></td>
<td>0.39</td>
<td>0.60</td>
</tr>
<tr>
<td>Primate conservation (Average conservation best 100nt window)</td>
<td></td>
<td>1.14</td>
<td>6.82x10^-5</td>
</tr>
<tr>
<td>Fetal heart DNase I hypersensitivity (DNase reads / kb)</td>
<td></td>
<td>1.45</td>
<td>5.25x10^-4</td>
</tr>
<tr>
<td>CAGE-seq fetal heart (CAGE-seq reads / kb)</td>
<td></td>
<td>0.83</td>
<td>1.44x10^-3</td>
</tr>
<tr>
<td>CAGE-seq adult heart (CAGE-seq reads / kb)</td>
<td></td>
<td>0.89</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Train machine learning model to prioritize sub-threshold loci
9 of 11 tested loci show allelic activity, chromatin interactions
Functional evidence for sub-threshold target genes

GWAS-only discovery requires 60k → 140k cohorts
From genomics to medicine in GWAS & Cancer

1. Characterize the epigenomic landscape
   – Map non-coding elements, uncover cellular circuitry
   – Identify disease-relevant tissues and regulators
   – Immune/inflammatory drivers in Alzheimer’s disease

2. Mechanistic dissection of non-coding disease loci
   – Identify cell type, causal SNP, regulator, targets, pathways
   – Genome editing: adipocyte browning drivers of obesity

3. Genetic + epigenomic discovery of new disease loci
   – Prioritize and experimentally validate variants missed by GWAS
   – New heart repolarization disease genes with cardiac functions

4. Non-coding discovery of new cancer driver genes
   – Gain of function mutations by out-of-context de-repression
   – Prostate cancer drivers: immune evasion, signaling, energy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   – Thousands of weak sites show genetic + epigenetic evidence
4. Non-coding discovery of cancer driver genes

Gain of function mutations by out-of-context de-repression
Prostate cancer drivers: immune evasion, signaling, energy

Richard Sallari
Nicholas Sinnott-Armstrong
Mathieu Lupien
This talk: From genomics to medicine

Part 1: Decoding the regulatory genome

- **Regions:** Enhancers, promoters, transcribed, repressed
- **Cell types:** Predict tissues and cell types of epigenomic activity
- **Target genes:** Link variants to their target genes using eQTLs, activity, Hi-C
- **Nucleotides:** Regulatory consequence of mutation: Conservation, PWMs
- **Regulators:** Upstream regulators whose activity is disrupted by mutation

![Diagram of gene regions: Enhancers, Promoters, Transcribed, Repressed]

- **Enhancers Promoters Transcribed Repressed**

Part 2: Application to discover new therapeutic leads

- **Dissect GWAS hits:**
  - CATGCCTG
  - CGTGCTGA
  - 93% top hits non-coding → Mechanism? Cell type?
  - Lie in haplotype blocks → Causal variant(s)?

- **Discover new loci:**
  - CATGCCTG
  - CGTGCTGA
  - Many loci undetected → Pathway-level burden/load
  - Many false positives → Prioritize w/ regulatory annotations

- **Cancer mutations:**
  - CATGCCTG
  - CATCCCTG
  - Loss of function → Protein-coding variants, convergence
  - Gain of function → Regulatory variants, heterogeneity

- **GWAS hits:**
  - Regions: Enhancers, promoters, transcribed, repressed
  - Cell types: Predict tissues and cell types of epigenomic activity
  - Target genes: Link variants to their target genes using eQTLs, activity, Hi-C
  - Nucleotides: Regulatory consequence of mutation: Conservation, PWMs
  - Regulators: Upstream regulators whose activity is disrupted by mutation
Regulatory convergence of dispersed driver mutations

Common mutations in regulatory plexus of each gene

Richard Sallari
Cancer genes are more likely to be up-regulated
Over-expressed genes harbor non-coding driver mutations

Gain of function drivers by activation of distal enhancers

New model for discovery of new cancer driver genes
Discover new cancer genes using non-coding drivers

Statistical model for excess of rare/somatic variants (correct for background rate in region, state, tumor)
New genes in common ‘cancer hallmarks’ pathways

Convergence in immune, signaling, mitochondrial functions
Regulatory genomics & epigenomics of disease

1. Characterizing the epigenomic landscape
   - Exploit dynamics to learn modules and networks
   - Identifying disease-relevant tissues and regulators

2. Mechanistic dissection of disease-associated loci
   - Relevant cell type, causal nucleotide, upstream regulator, target genes, cellular signatures, organism phenotypes
   - CRISPR/Cas9 editing of causal SNP reverses phenotype

3. Epigenomics reveals weak but functional variants
   - Prioritize and experimentally validate variants at $10^{-5}$
   - Cohort of 69k individuals when 147k would be needed

4. Dispersed mutations underlie cancer dysregulation
   - Regulatory convergence analysis reveals new cancer genes
   - Out-of-context de-repression as cancer gain-of-function strategy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   - Thousands of weak sites show genetic + epigenetic evidence
5. Genetic + Epigenetic variation in Alzheimer’s brains

Thousands of weak sites show genetic + epigenetic evidence
‘Mendelian randomization’ at the genome-wide level

Matt Eaton
Phil De Jager
Methylation in 750 Alzheimer patients/controls

- Patients followed for 10+ years with cognitive evaluations
- Brain samples donated post-mortem methylation/genotype
- Seek predictive features: SNPs, QTLs, mQTLs, regulation

Methylation variation in 723 individuals

Relate to genotype and AD variation
Discover 50,000 methylation QTLs (meQTLs)

Genome-wide effect of genotype on methylome
Alzheimer’s-associated methylation differences

- Methylation more informative than SNPs
- Enhancers more informative than promoters
Methylation differences a causal component of AD

Methylation probes altered in AD are enriched in AD-associated SNPs

Set-wise causality testing

AD predictive power reduced after removing meQTL effect
Regulatory genomics & epigenomics of disease

1. Characterizing the epigenomic landscape
   - Characterizing the epigenomic landscape
   - Exploit dynamics to learn modules and networks
   - Identifying disease-relevant tissues and regulators

2. Mechanistic dissection of disease-associated loci
   - Relevant cell type, causal nucleotide, upstream regulator, target genes, cellular signatures, organism phenotypes
   - CRISPR/Cas9 editing of causal SNP reverses phenotype

3. Epigenomics reveals weak but functional variants
   - Prioritize and experimentally validate variants at $10^{-5}$
   - Cohort of 69k individuals when 147k would be needed

4. Dispersed mutations underlie cancer dysregulation
   - Regulatory convergence analysis reveals new cancer genes
   - Out-of-context de-repression as cancer gain-of-function strategy

5. Genetic + Epigenetic variation in Alzheimer’s brains
   - Thousands of weak sites show genetic + epigenetic evidence
1. Epigenomics Roadmap Project

Anshul Kundaje
Wouter Meuleman
Jason Ernst

Roadmap Epigenomics
ENCODE
NIH, NHGRI, Common Fund

2. Mechanistic dissection of FTO

Melina Claussnitzer
Simon Dankel
Kyoung-Han Kim

Chi-Chung Hui
Hans Hauner
Gunnar Melgren

3. Cardiac sub-threshold variants

Xinchen Wang
Laurie Boyer
David Milan
Chris Newton-Cheh

4. Cancer convergence

Richard Sallari
Nicholas Sinnott-Armstrong
Mathieu Lupien

5. Genetics and Epigenomics of Alzheimer’s Disease

Matt Eaton
Gerald Quon
Lori Chibnik

Phil De Jager
David Bennett
From genomics to medicine in GWAS & Cancer

1. Characterize the epigenomic landscape
   – Map non-coding elements, uncover cellular circuitry
   – Identify disease-relevant tissues and regulators
   – Immune/inflammatory drivers in Alzheimer’s disease

2. Mechanistic dissection of non-coding disease loci
   – Identify cell type, causal SNP, regulator, targets, pathways
   – Genome editing: adipocyte browning drivers of obesity

3. Genetic + epigenomic discovery of new disease loci
   – Prioritize and experimentally validate variants missed by GWAS
   – New heart repolarization disease genes with cardiac functions

4. Non-coding discovery of new cancer driver genes
   – Gain of function mutations by out-of-context de-repression
   – Prostate cancer drivers: immune evasion, signaling, energy