

AI for Health and the future of Medicine: Deciphering and manipulating human disease at single-cell resolution

Manolis Kellis

compbio.mit.edu

 [@manoliskellis](https://twitter.com/manoliskellis)



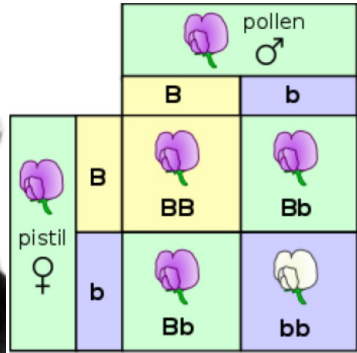
Computer Science and
Artificial Intelligence Lab



Genetics: Ancient Foreshadowings → Mendelian traits → Polygenicity



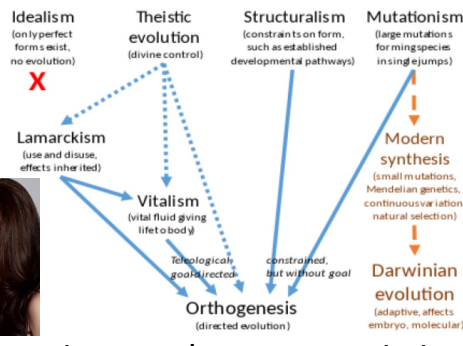
9000BC: Selective breeding of animals/plants
Inheritance: Eye/hair color long understood



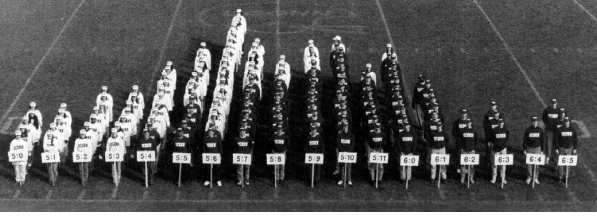
1866: Mendel: Discrete inheritance
 No blending. Dominant/recessive alleles
 Independent assortment



Biometrics: continuous phenotype variation.
Others: Saltationism, orthogenesis, vitalism, neo-Lamarckism, theistic evolution...



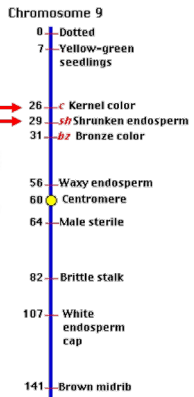
Fisher



1918. Continuous phenotype variation explained by multiple Mendelian loci

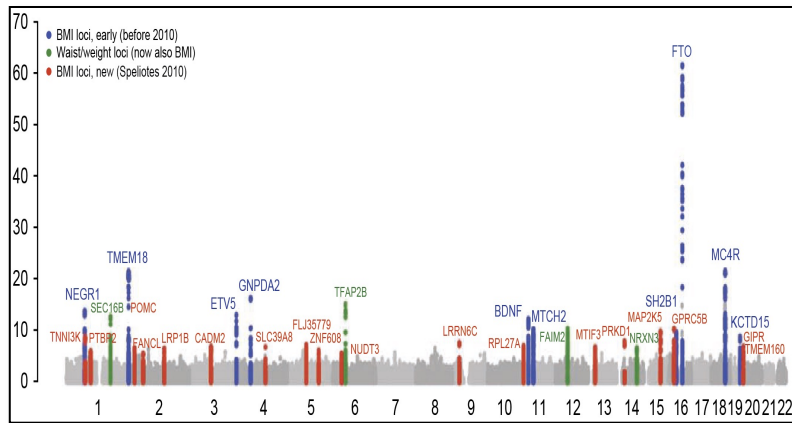


1913: Linkage/mapping, Morgan, Sturtevant
1980s: Mendelian Trait genes mapped



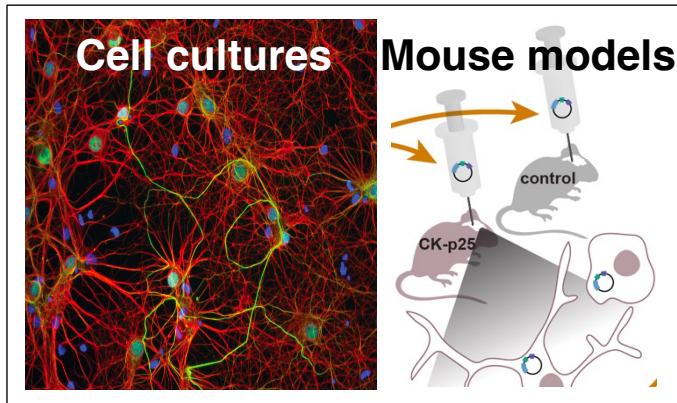
2000s: Human genome. Variation maps. Haplotypes. GWAS. Common/rare variants.

Dissect mechanisms of disease-associated regions

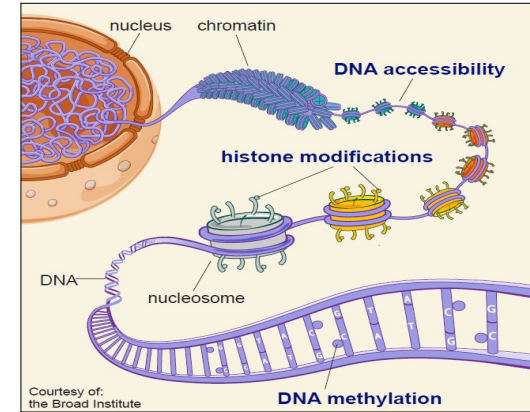


**Start with disease genetics:
Common + rare variants**

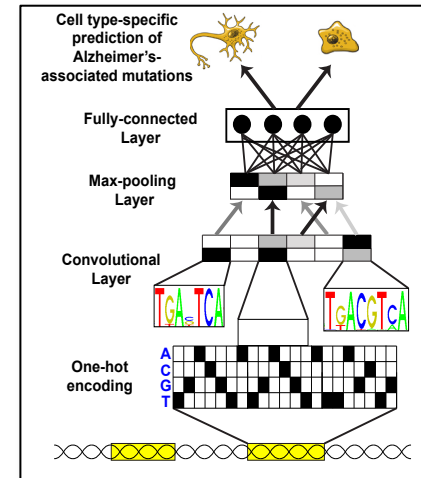
4. Disseminate results



**3. Validate predictions in
human cells + mouse models**



**1. Profile RNA + Epigenome
in healthy + disease samples**



**2. Integrate data to predict driver
genes, regions, cell types³**

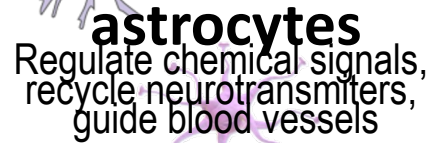
Step 1: Large-scale profiling (e.g. 1500+ brain samples, 20M+ cells)

- Alzheimer's
- FrontoTemporal
- LewyBodyDem
- ALS amyotr.lat.scl
- Huntington's
- AD+Psychosis
- Schizophrenia
- Bipolar Disorder
- Down Syndrome
- Autism, PTSD
- Addiction, Aging
- Cardiac Disease
- Obesity, Exercise



Across traits (N=12, each ~48)

Across individuals (N=1500)

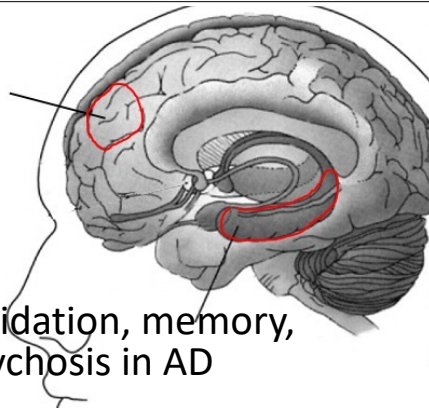


Across cell types (n=75+)

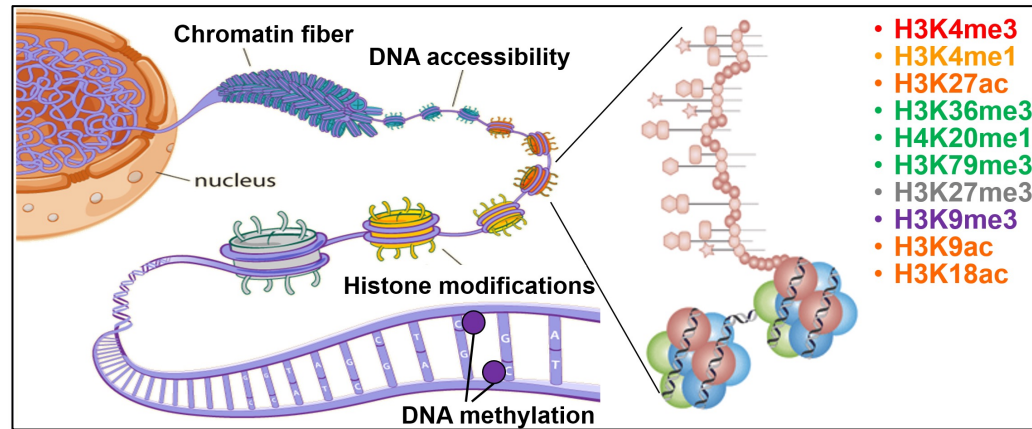
Across individual cells (n=20,000,000+)

Frontal cortex
Cognition, personality, social behavior, psychosis in AD

Hippocampus
Information consolidation, memory, social behavior, psychosis in AD

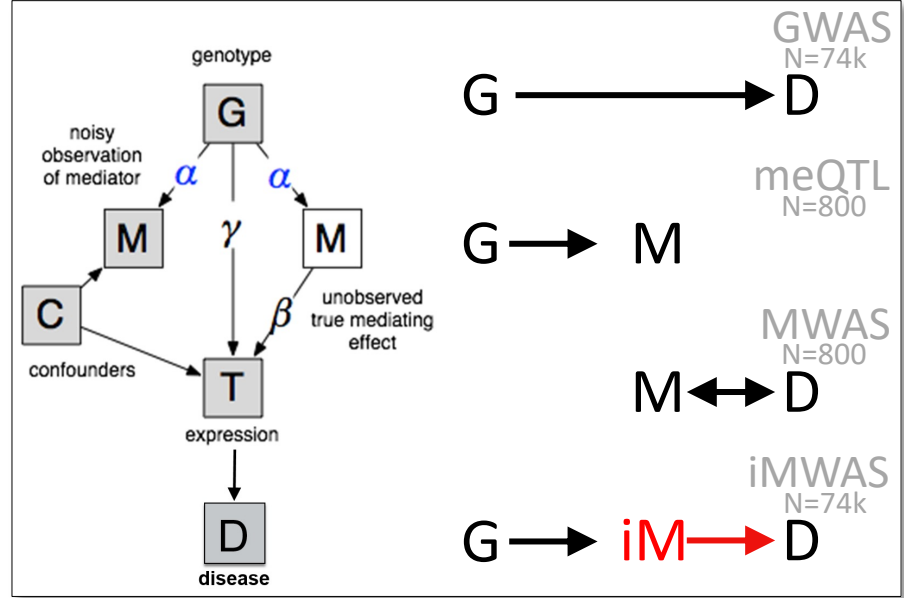
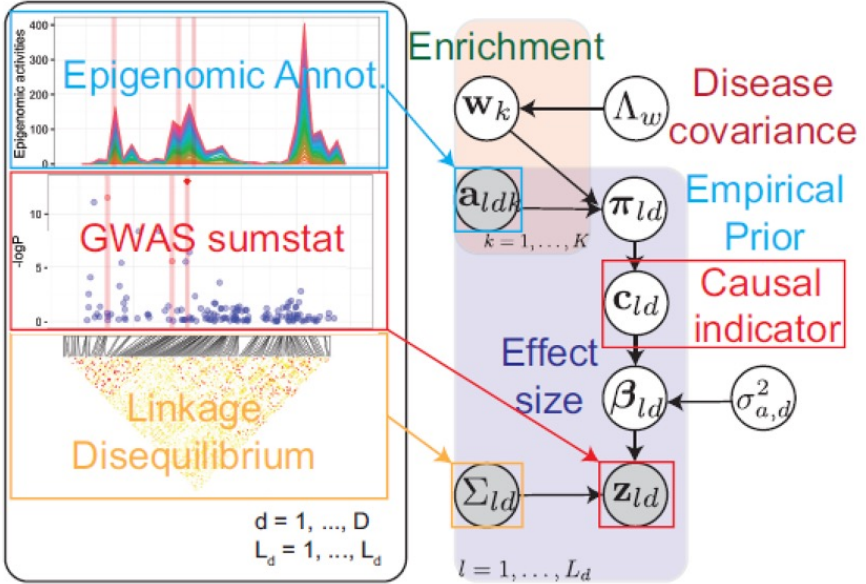


Across brain regions (N=7+)

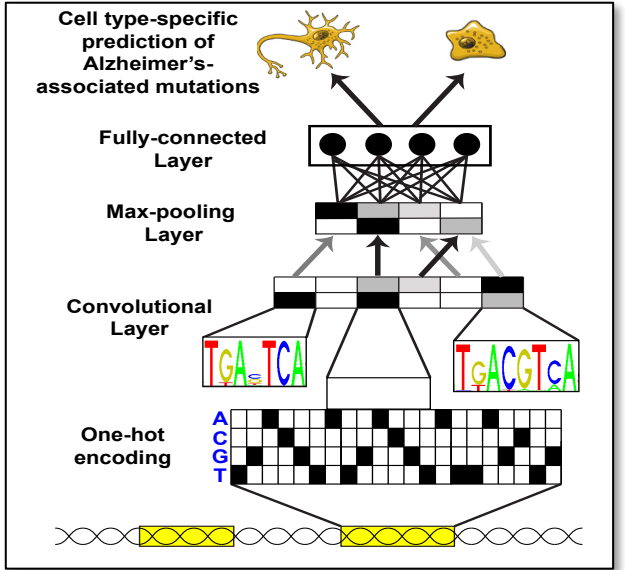
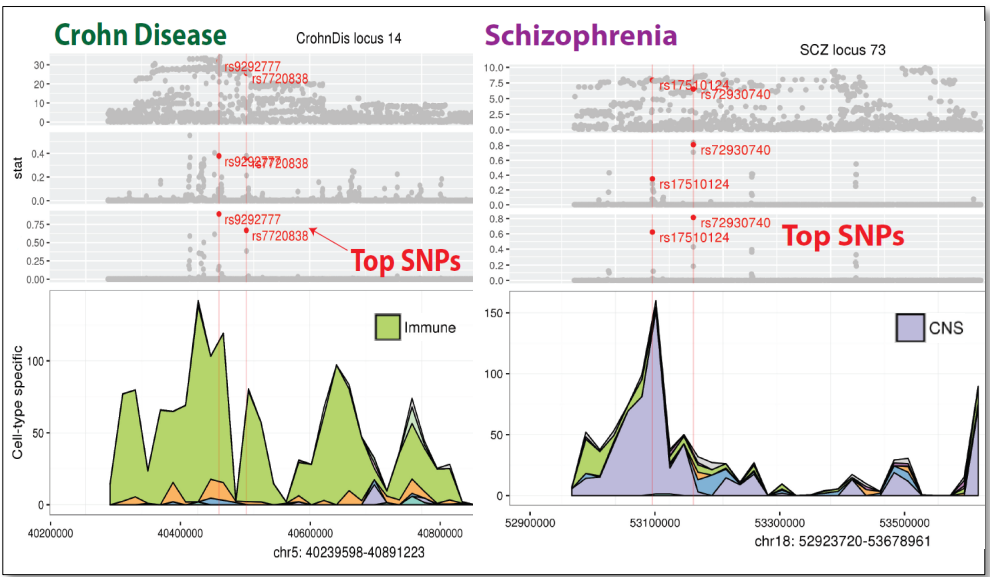


Across scRNA (N=1000), scATAC (N=500)

Step 2: Data integration → driver genes/regions/cells/pathways

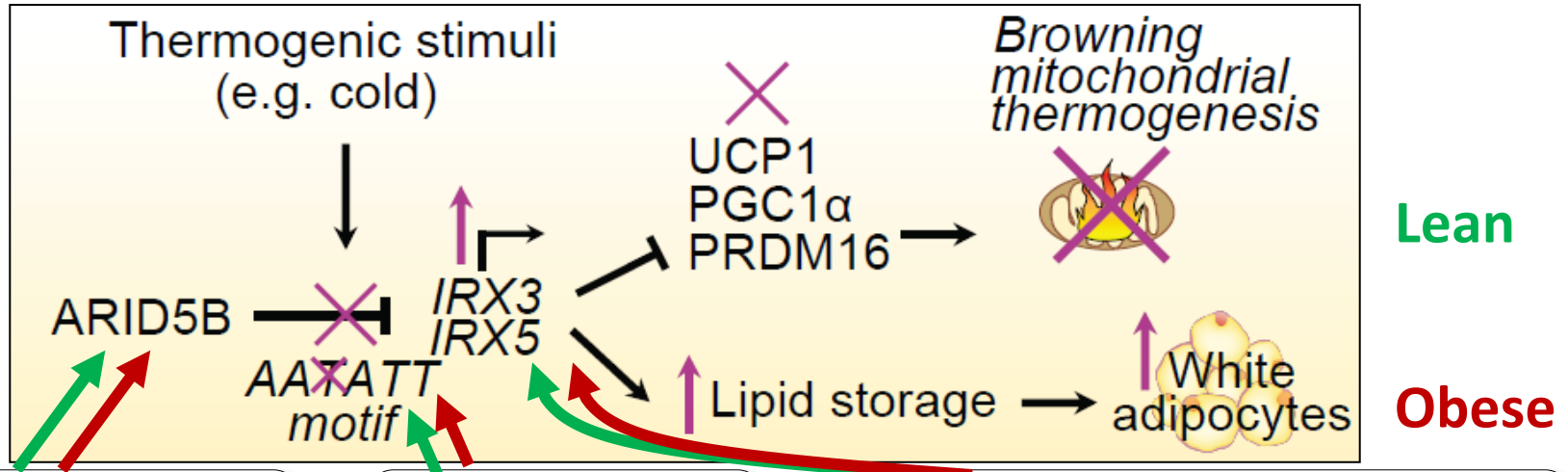


Bayesian multi-trait GWAS integration Causality inference, Mediation



Predict causal variants and cell types CNNs: Predict function from seq

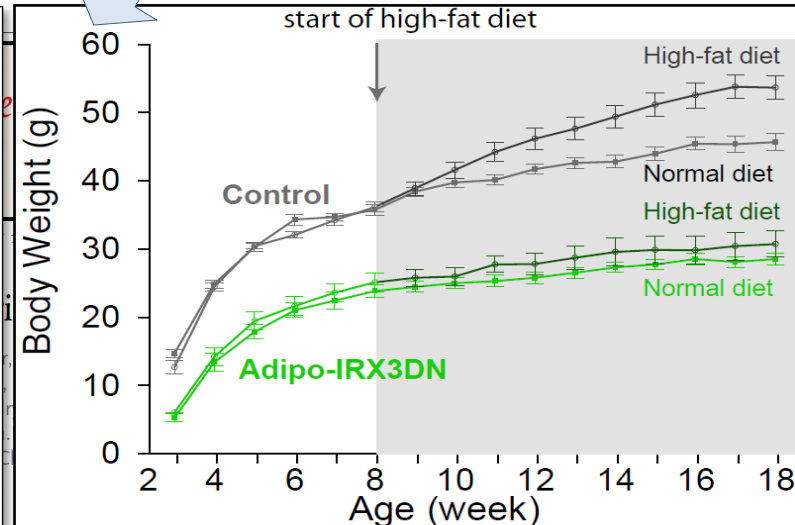
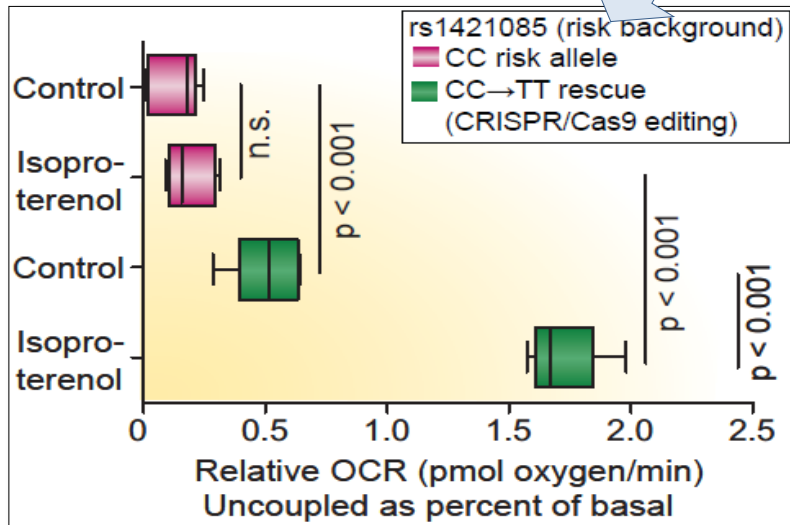
Step 3. Manipulate circuitry → reverse disease phenotypes



Incr. ARID5B → Lean
 Decr ARID5B → Obese

C-to-T → Lean
 T-to-C → Obese

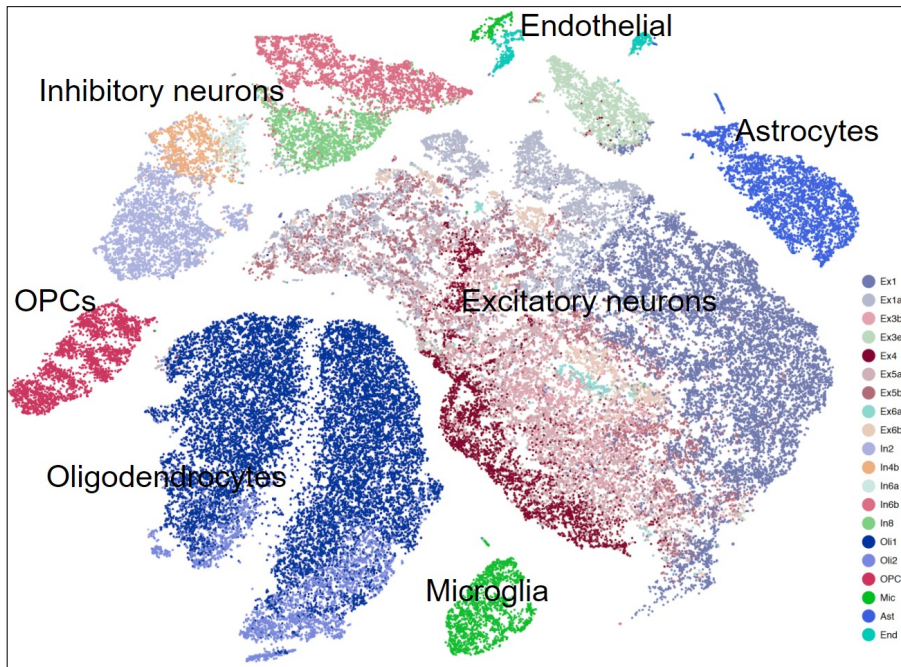
Decrease IRX3, IRX5 → Lean
 Increase IRX3, IRX5 → Obese



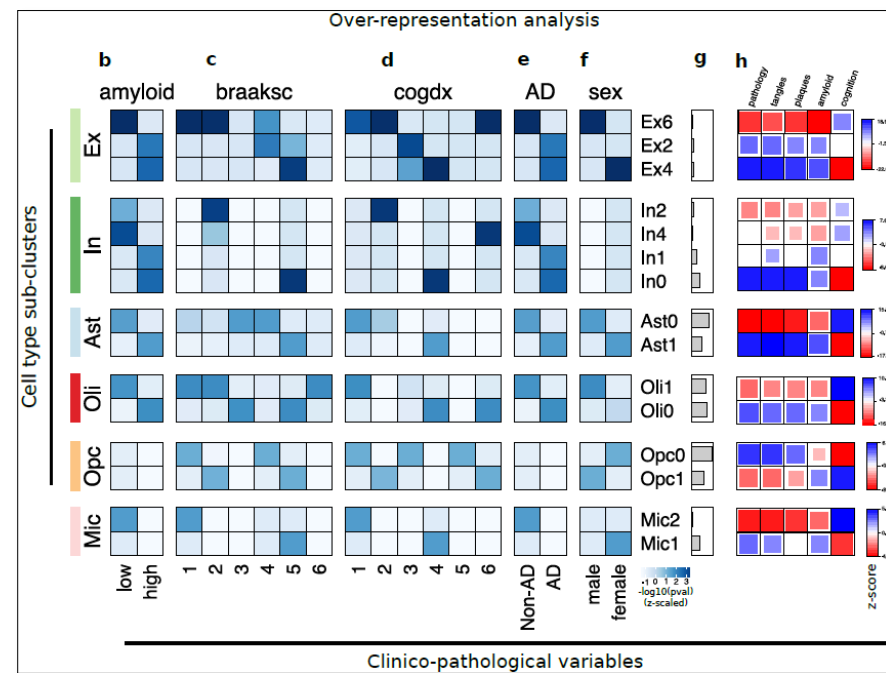
CRISPR-edit human fat cells
 → able to burn calories again

IRX3 KD → Burn calories in their sleep
 → 54% weight loss. Can't gain weight

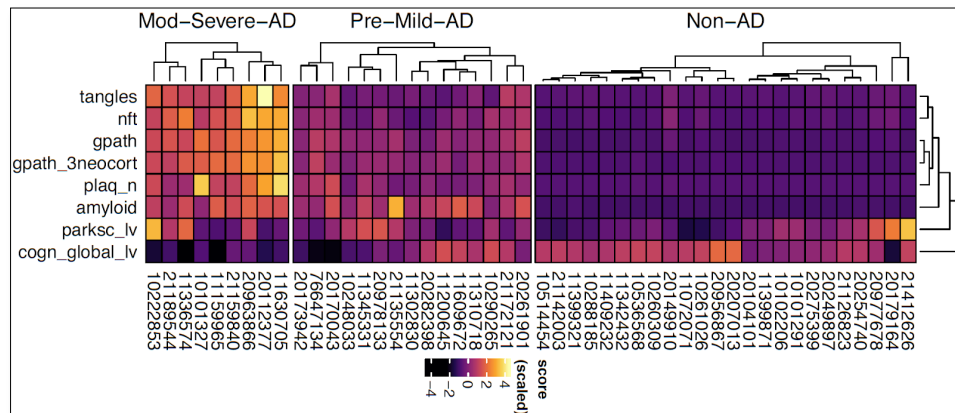
Single-cell profiling across 1400+ brain samples, 5M+ cells



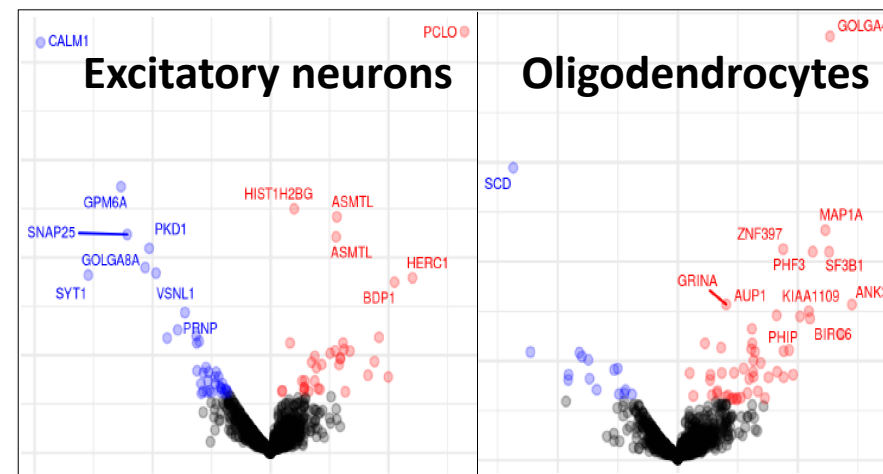
scRNA-seq in 48 individuals: cell type diversity



AD vs. non-AD subgroups in each cell type

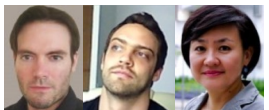


Phenotypic diversity across individuals

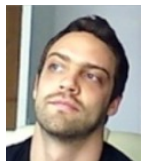
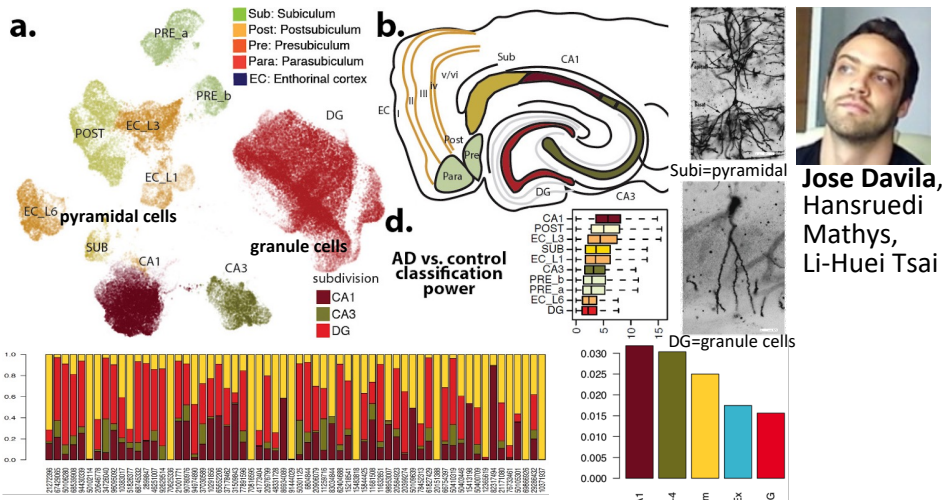


• Somatic mutation burden clusters in Golgi, vesicle trafficking, intermediate filaments

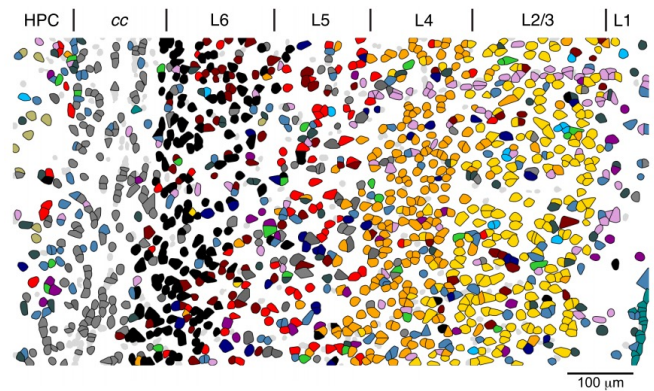
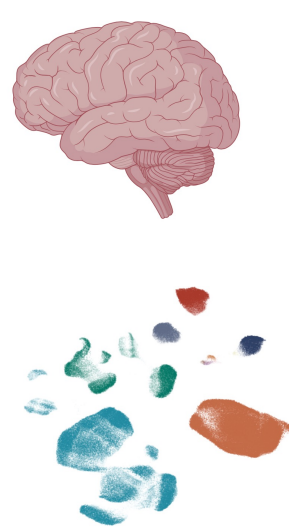
With Li-Huei Tsai
Mathys, Davila *et al*
Nature, 2019



Deep Learning for Spatial Transcriptomics + Single-cell Integration



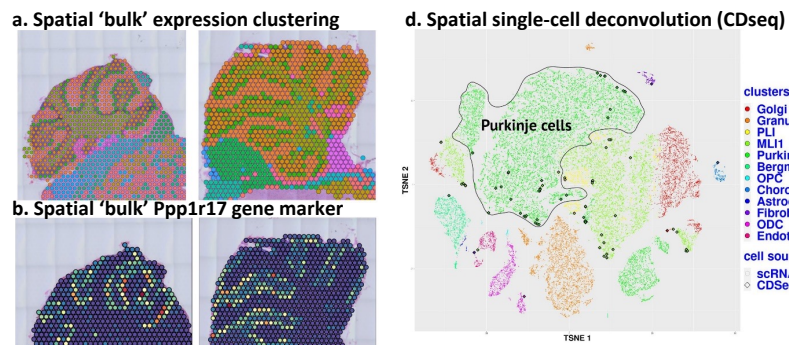
Jose Davila,
Hansruedi
Mathys,
Li-Huei Tsai



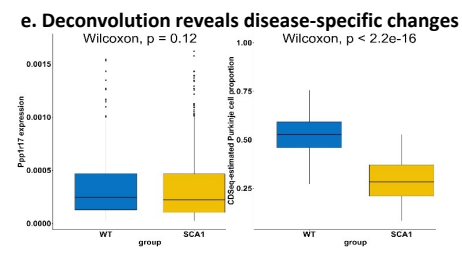
Na Sun, Djuna Von Maydell,
Guillaume Leclerc, Fatima
Gunter-Rahman

Captured sub-regions across individuals
snRNA profiles capture hippocampal sub-regions

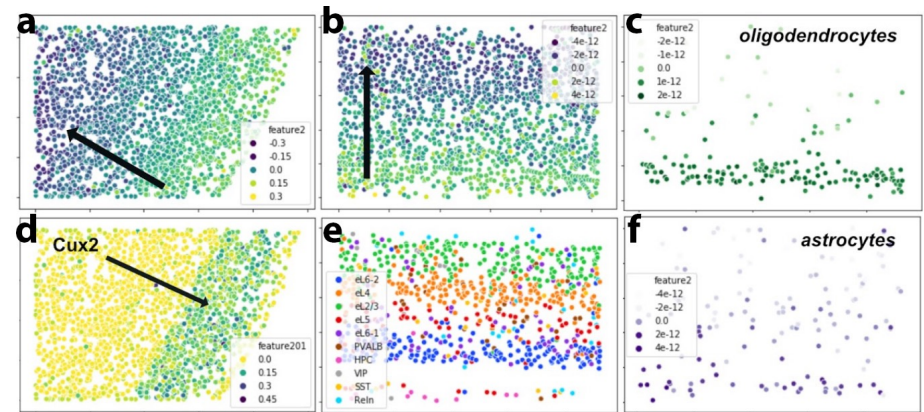
Davila *et al*, In submission



Kai Kang,
Kiki Galani,
Myriam
Heiman



Spatial + single-cell possible, but slow, scarce, expensive



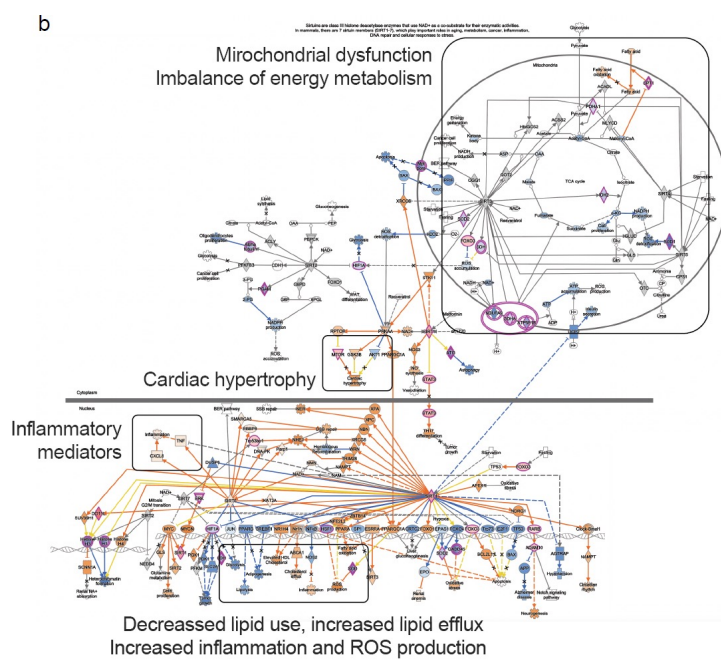
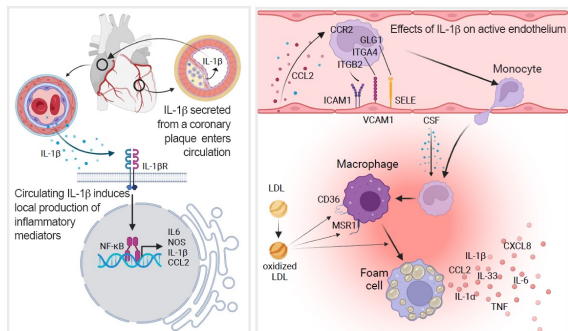
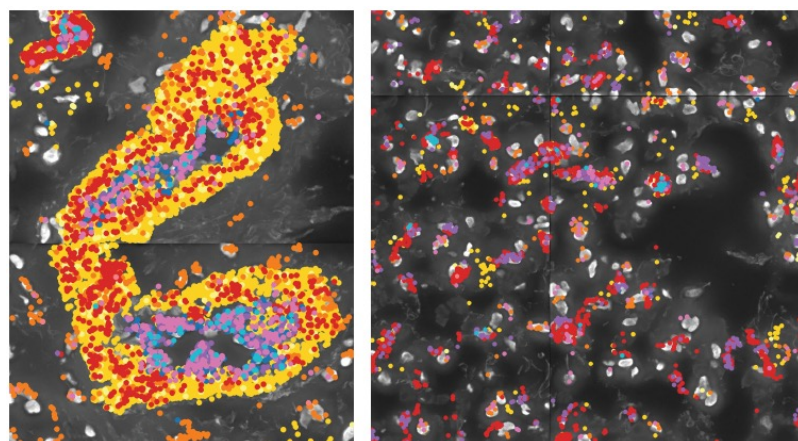
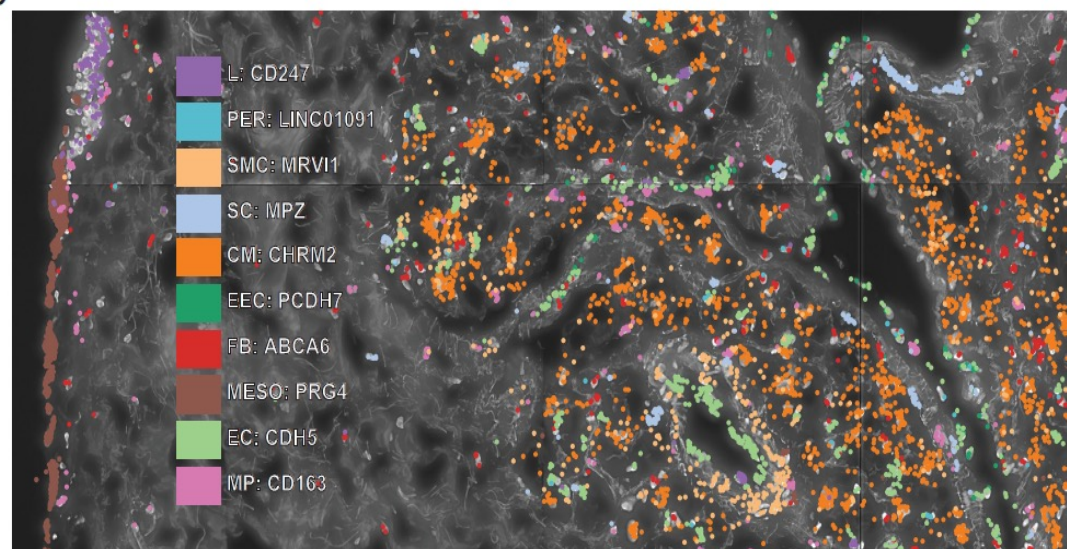
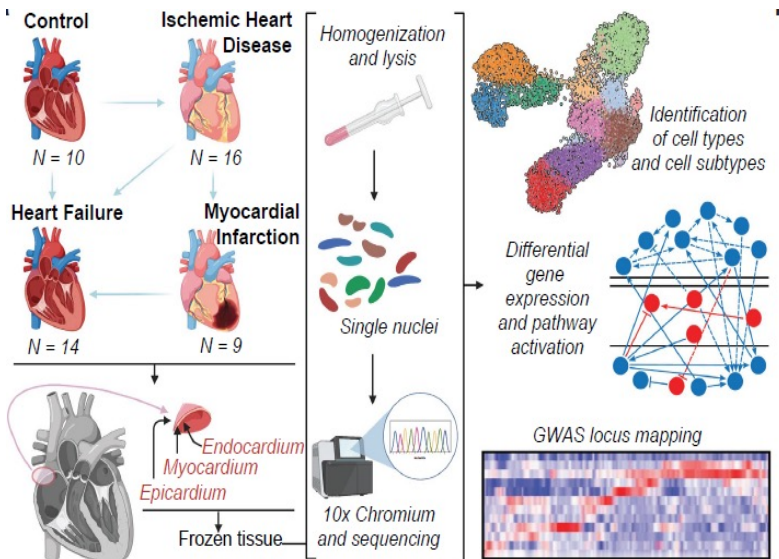
- Train deep-learning model to predict spatial positioning information for single-cell data from expression patterns
- Recover neuronal layer information, well-recognized information
- Astrocytes and oligodendrocytes also show spatial positioning

Spatial transcriptomics single-cell deconvolution

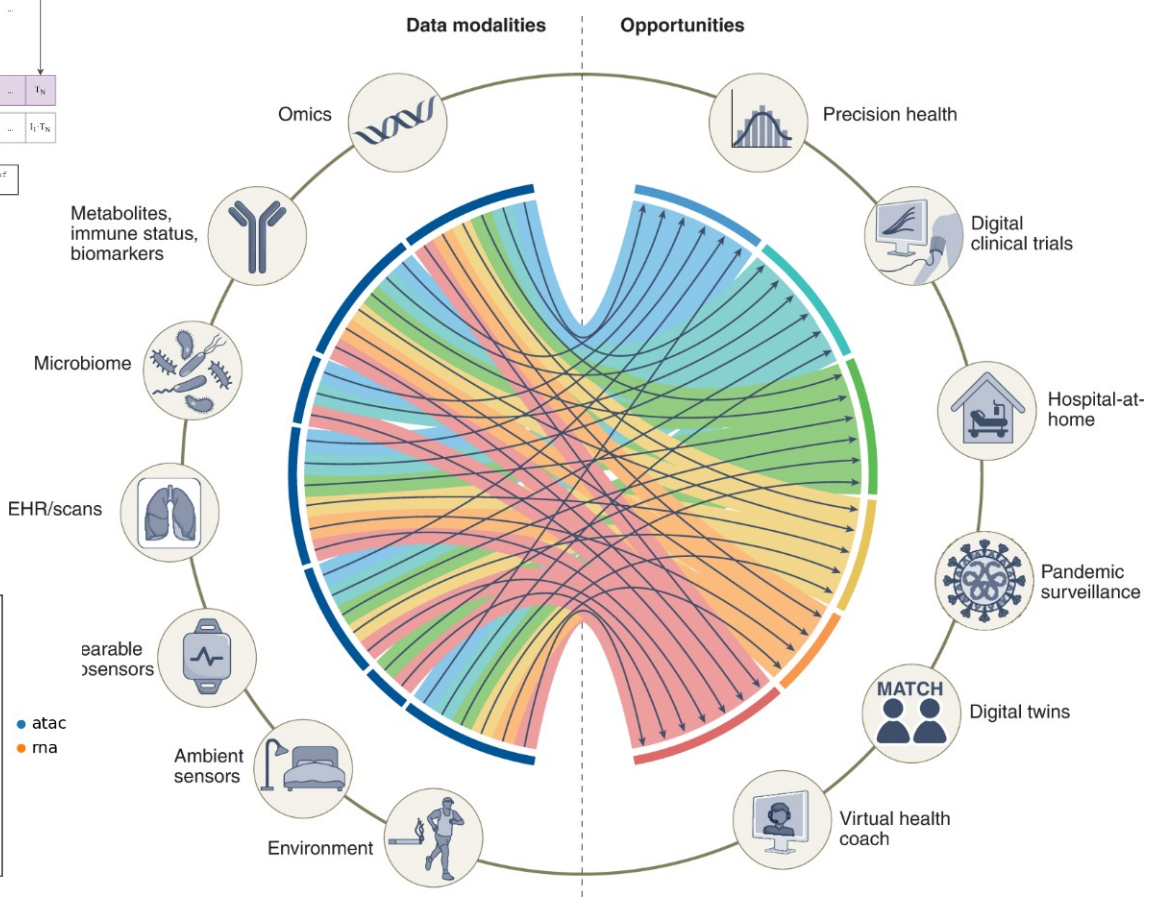
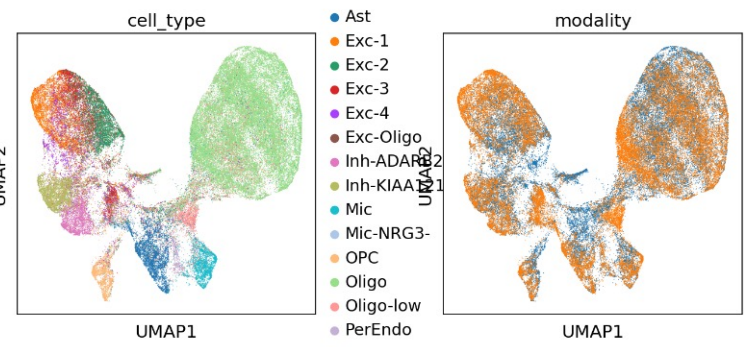
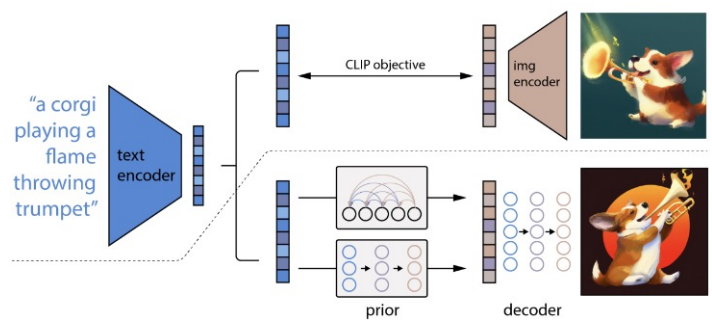
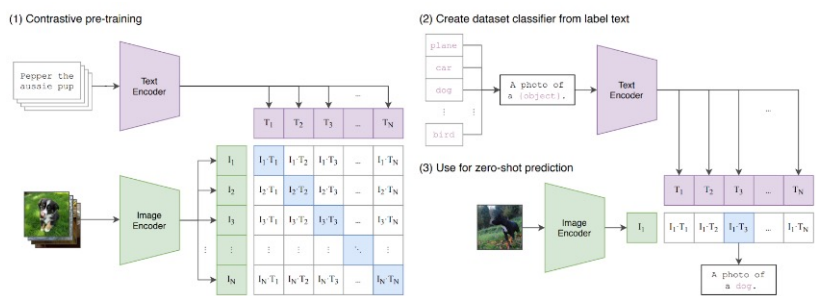
Kang *et al*, In preparation

Van Maydell, Sun, *et al*, In preparation

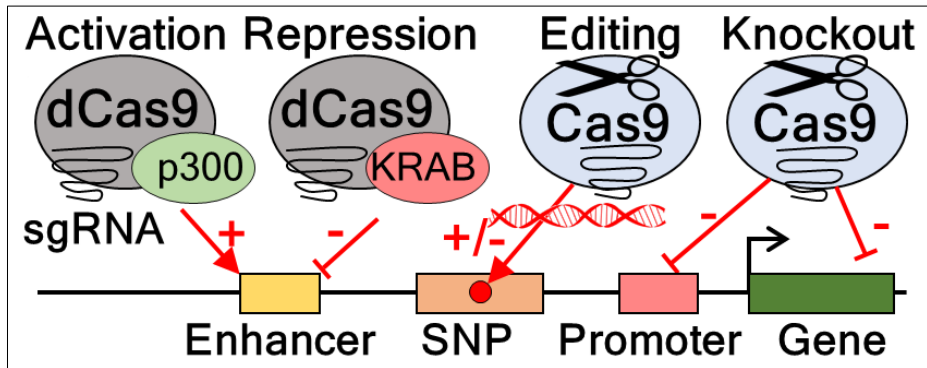
Sub-cellular transcriptomics, single-cell RNA, and imaging in heart: Coronary Artery Disease



Deep Learning transformers across modalities: scRNA, scEpigenomics, Imaging, Phenotypes



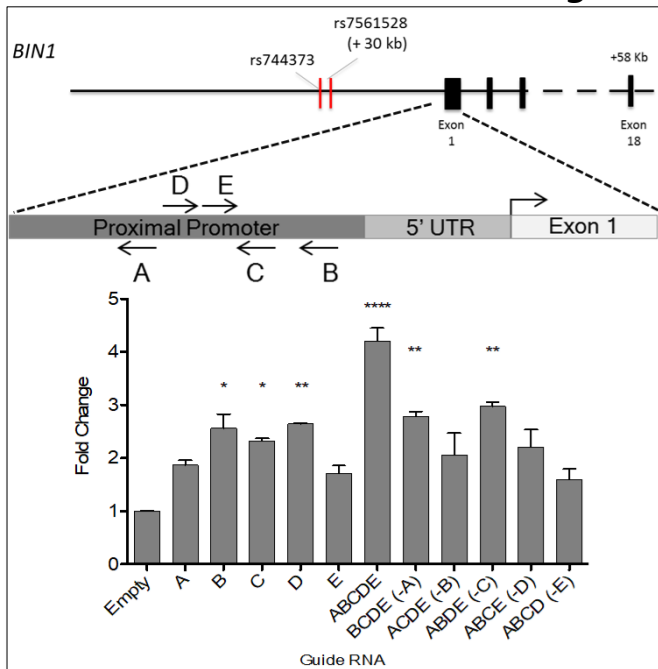
Modular and programmable CRISPR-Cas9/dCas9 system



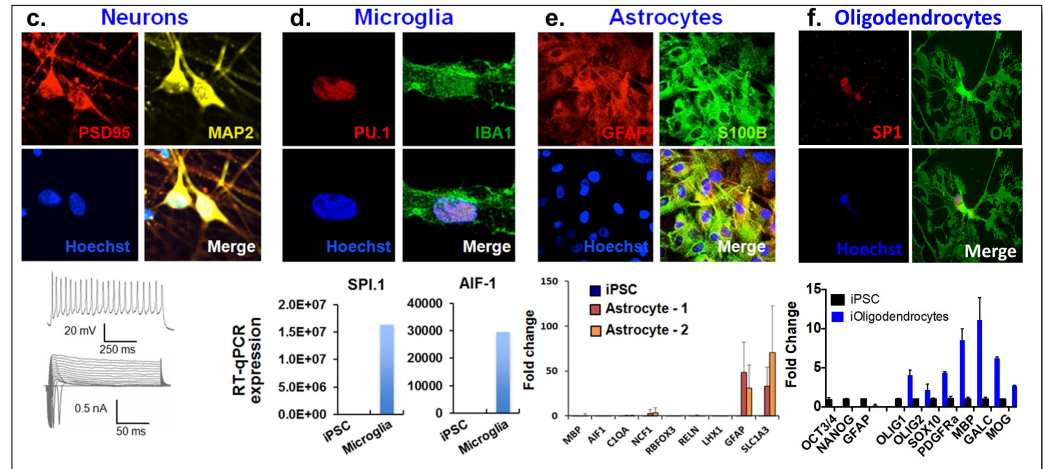
- **Activation:** CRISPR-dCas9+p300
- **Repression:** CRISPR-dCas9+KRAB
- **Editing:** CRISPR-Cas9 + repair template
- **Knockout:** CRISPR-Cas9 cutting

Modularity:

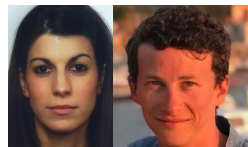
- Pick perturbation type (3 lines)
- Pick cell type (differentiation)
- Pick target (sgRNA + repair template)
- Induce (Dox/Tet control)
- Environmental modulations (+Aβ)
- Cross-cell-type effects (2D/3D co-culture)



AD: Bin1 enhancer activation with multiple sgRNAs

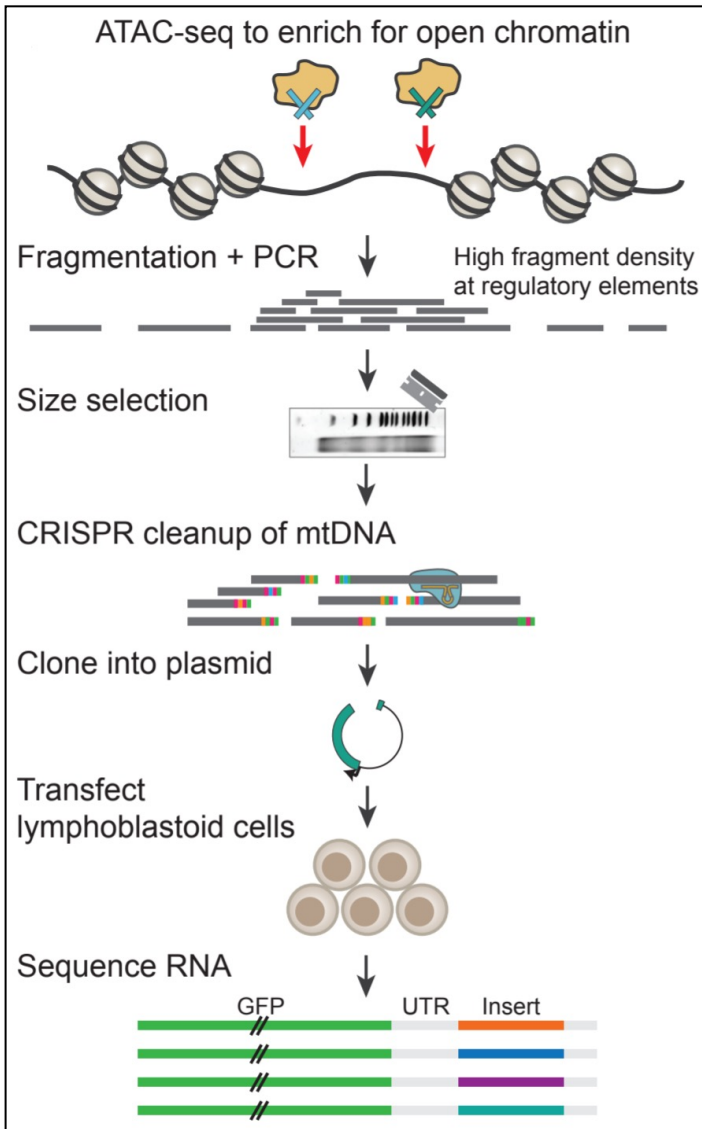


Apply in iPSCs, differentiate into NPCs, neurons, astrocytes, oligodendrocytes, microglia

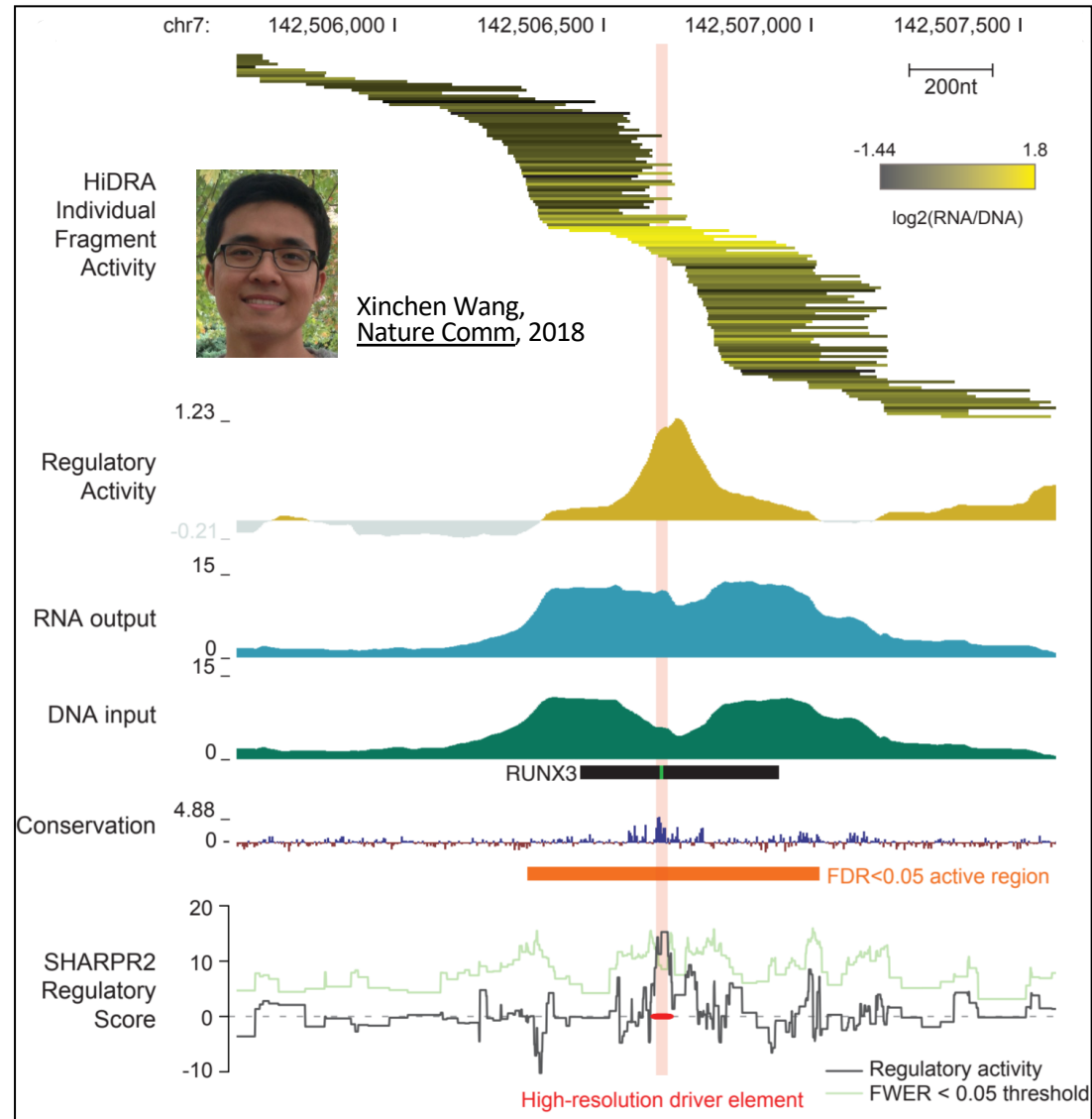


Maria Kousi, Michael Gutbrod, In preparation
Collaboration with Li-Huei Tsai, Kevin Eggan, Nikos Daskalakis et al

Ultra-high-throughput assays: 7M tests + high-res

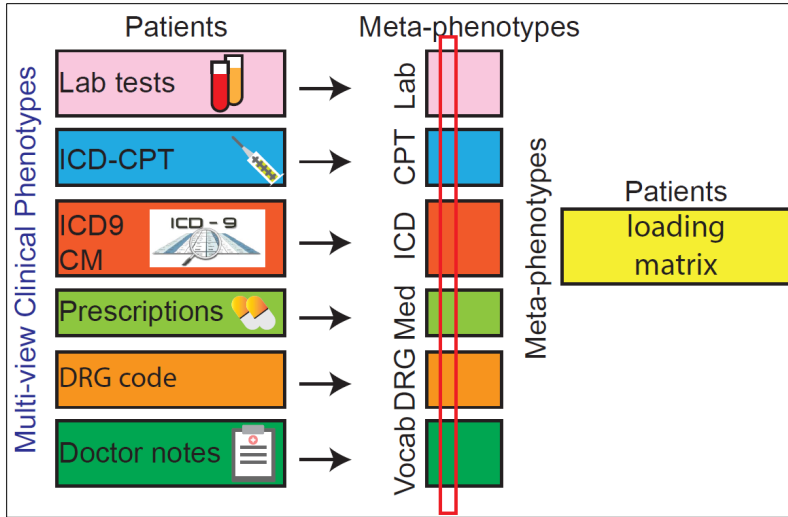


ATAC selection → No synthesis → 7M tests
 3'UTR incorp. → Self-transcribe → No barcode
 Dense, random start/end → Region tiling

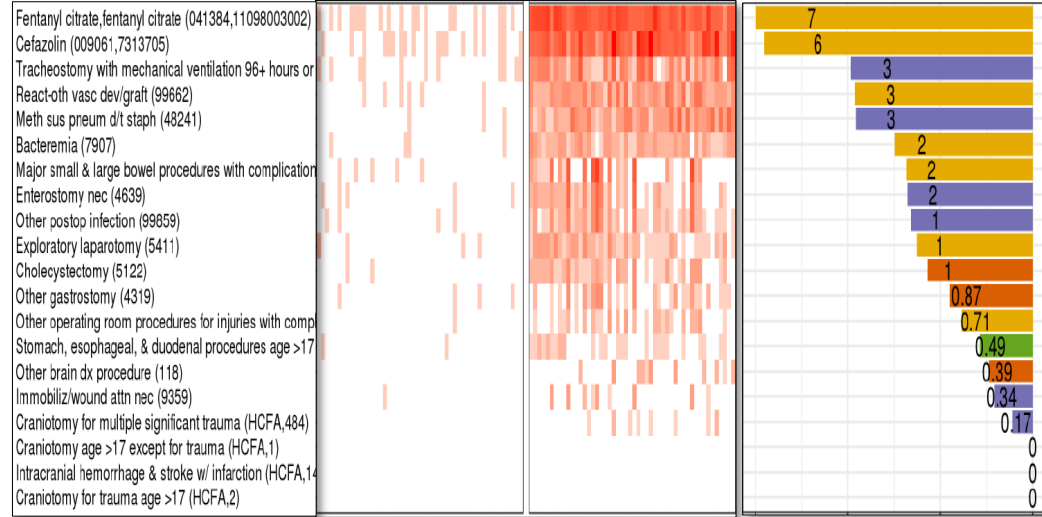


High-resolution inference of driver nucleotides
 → Exploit differences between neighboring fragments
 → Driver nucleotides match motifs, evolut. conservation

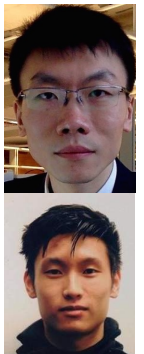
Genomics + EHR integration across 1M+ patients



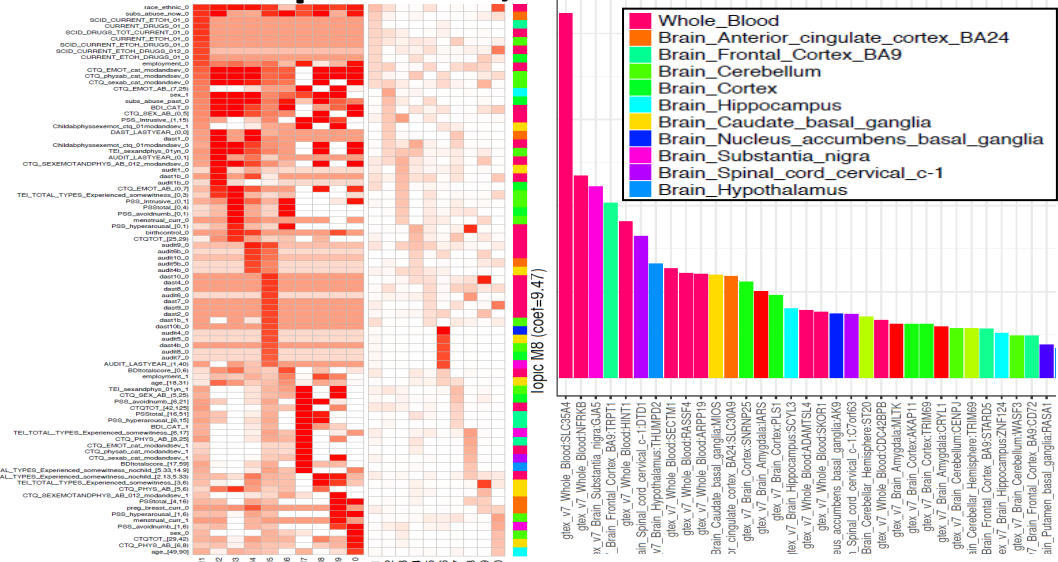
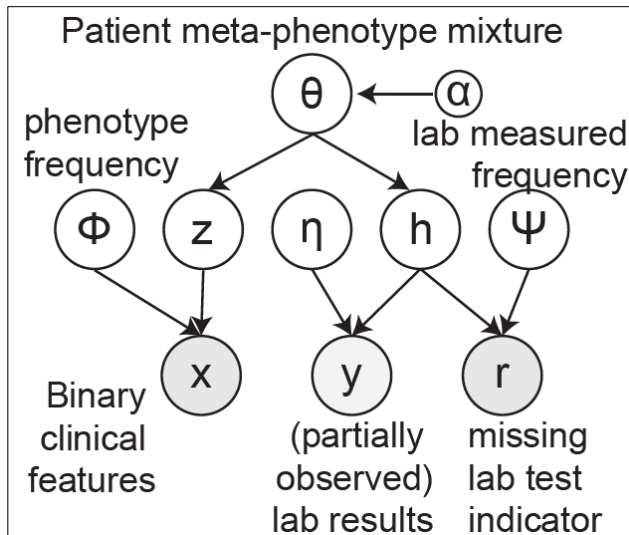
Integrate ICD9, DRG, lab tests, prescriptions, doctor notes



Learned topics cut across EHR data types, enable imputation, completion. correction



Yue Li, Andy Shea, Nature Com.

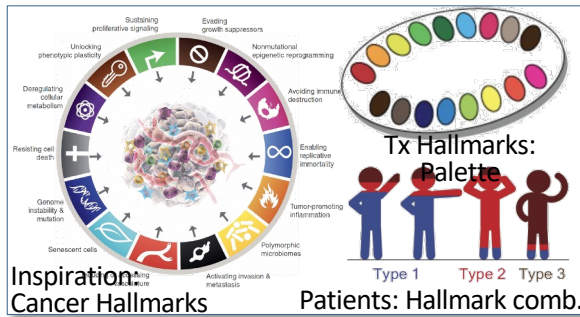


Multi-modal hierarchical Bayesian NMAR model (MixEHR)

Integrate w/ eQTL-based expression genes/cells of action for complex traits

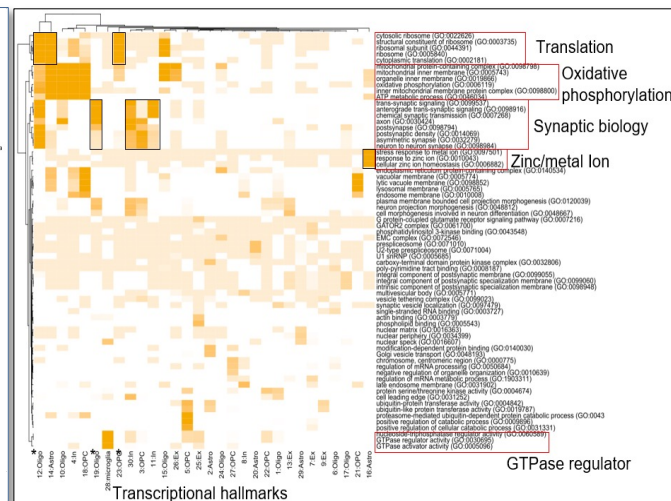
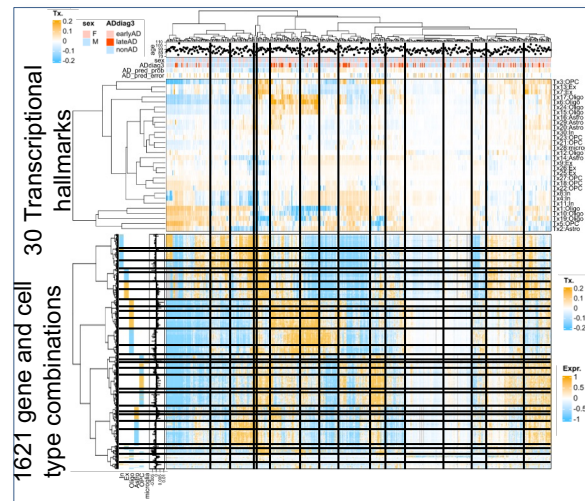
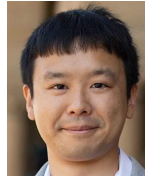
Maria Kousi, Michael Guttorf, in preparation
Collaboration with Liuhui Tang, Kevin Egan, Nikos Bakas, et al.

Disease hallmarks: patient subtyping & personalized medicine

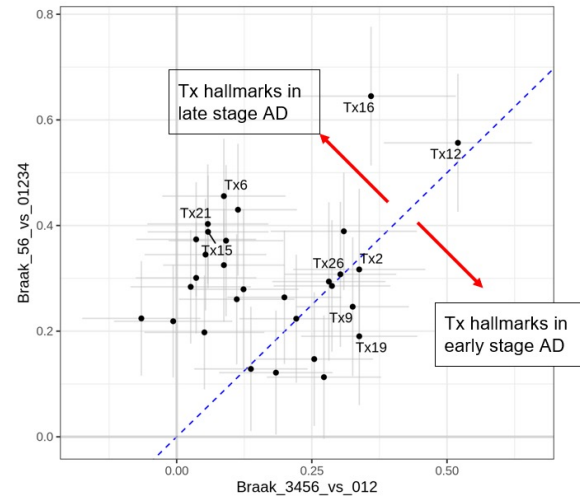


- Pathway-centric disease heterogeneity
- Each patient: pathway-specific burdens

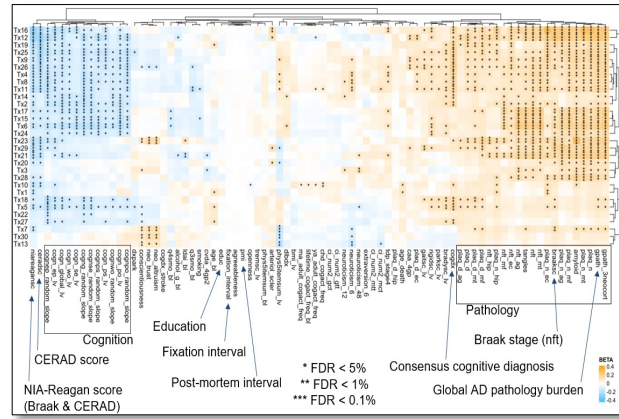
Yosuke Tanigawa



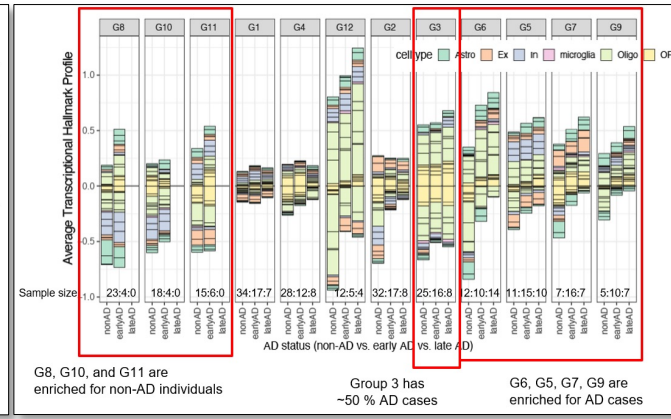
- 30 Transcriptional hallmarks: distinct pathways → Distinct painting of each patient



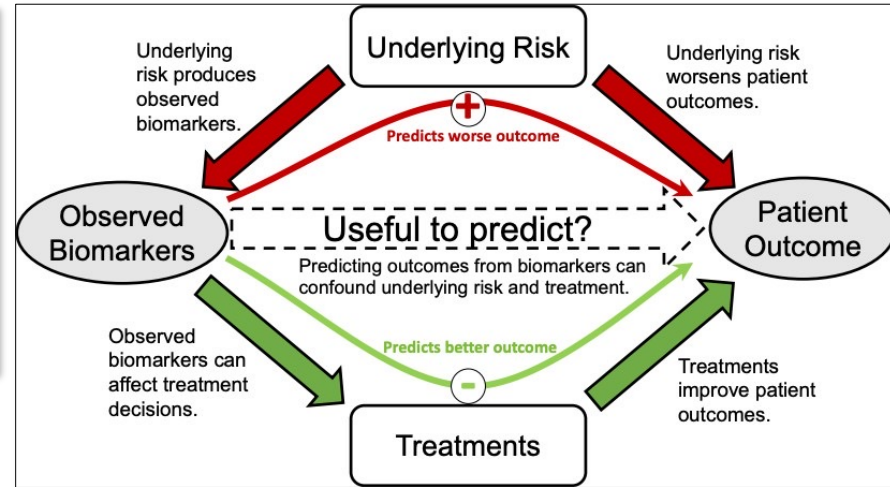
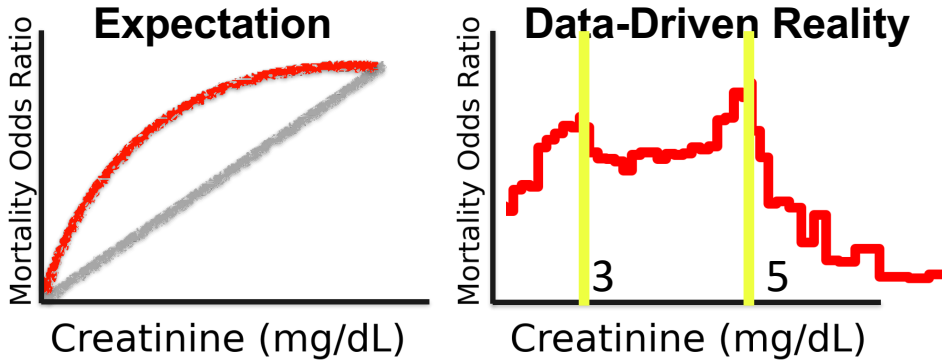
- Distinct early- vs. late-AD Hallmarks



- Tx Hallmarks → Distinct patient subgroups → Diagnosis & Prognosis → Treatment

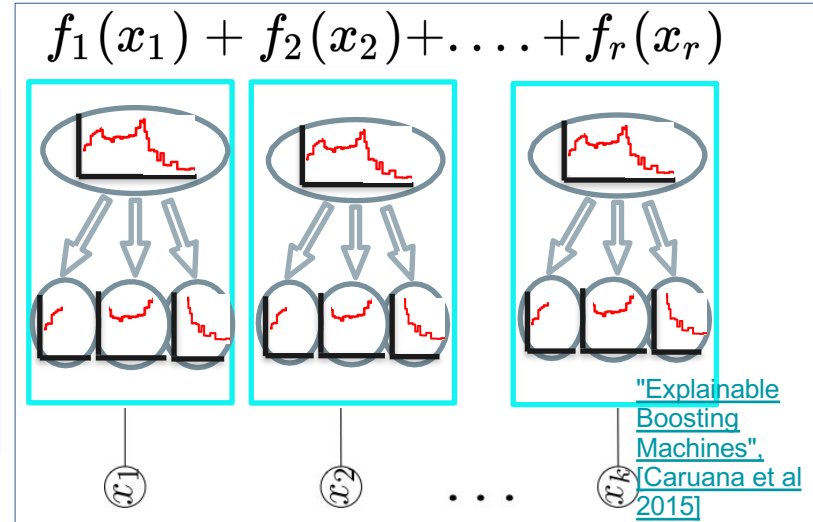
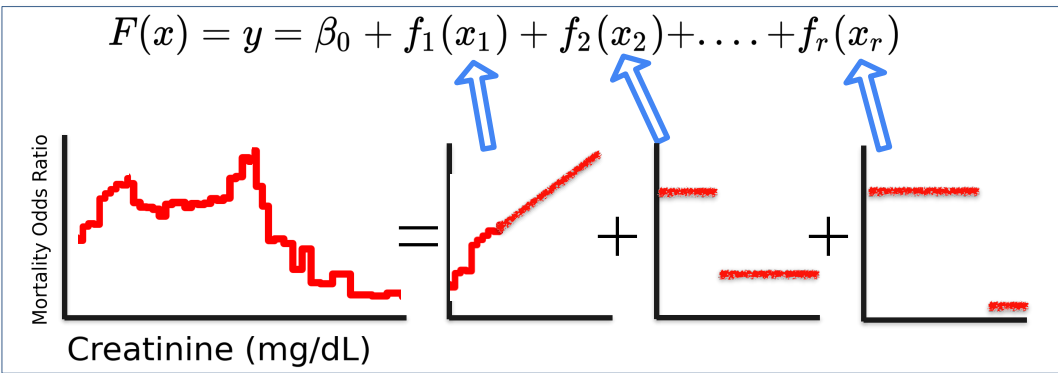


Death by Round Numbers and Sharp Thresholds: EHR and AI mining



- Elevated creatinine → indicator of renal failure, so expect mortality risk to increase w/creatinine
- In practice, much more complex curve: many discontinuities, sharp thresholds, round numbers
- Reason: round thresholds trigger treatment interventions that lower risk
- Naïve AI model: would recommend increasing creatinine levels to reduce mortality risk: Death!

- Goodhart Law: “When a measure becomes a target, it ceases to be a good measure”
- EHR: “When a biomarker leads to treatment, it ceases to be a good biomarker”

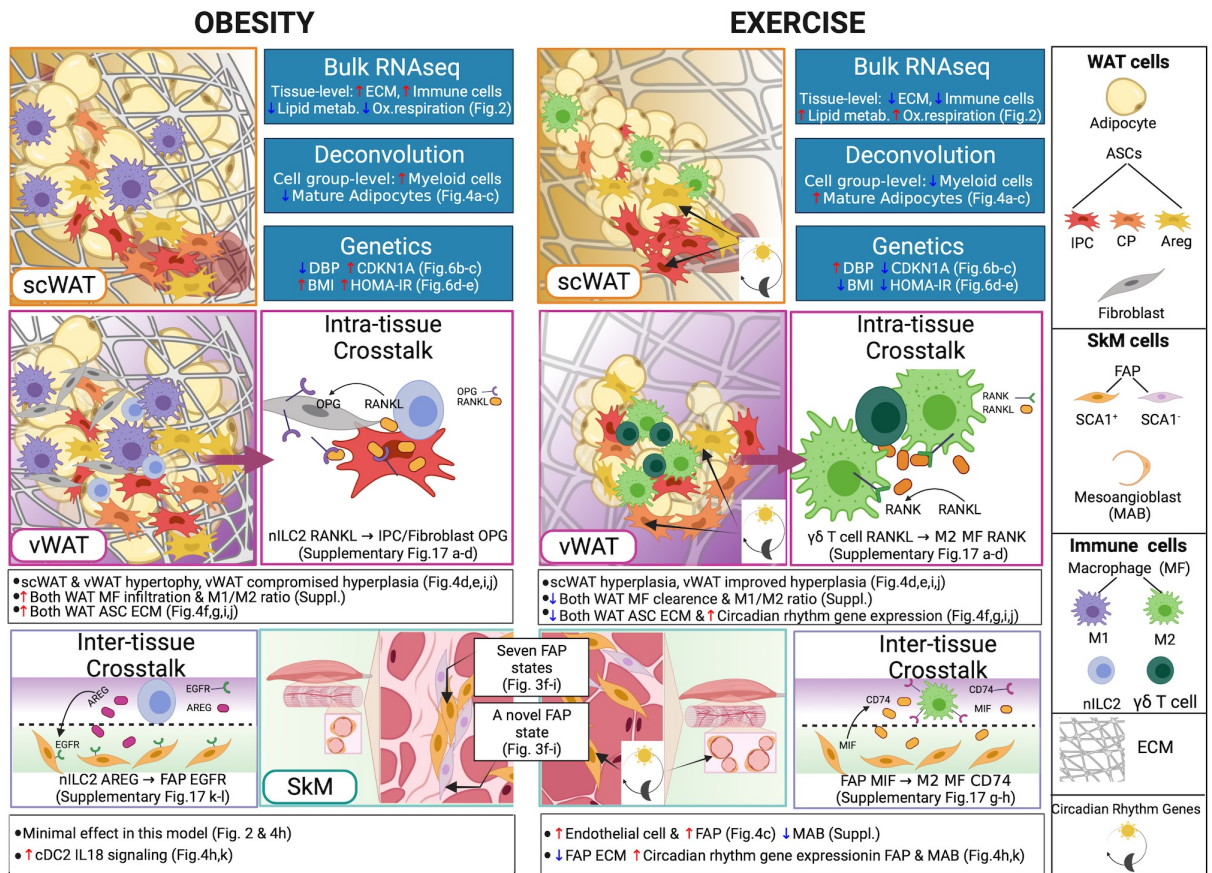
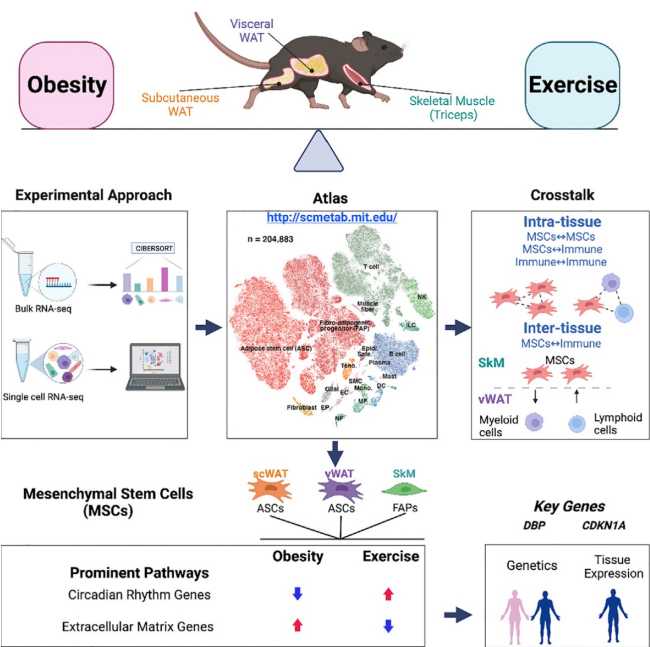


- Generalized Additive Models (GAM): Flexible Glass-Box Models
- Decompose each variable into its drivers



- GAM with Boosted Trees
- Deal with sharp thresholds

Multi-tissue effects: Impact of exercise + obesity



Yang CellMetabolism 22

- Single-cell Multi-Tissue effects
- Exercise rewires your metabolic tissues to burn more calories
- Stem cell reprogramming, cell-cell communication, immune processes

MIT Massachusetts Institute of Technology

Education | Research | Innovation | Admissions + Aid | Campus Life | News | Alumni | About MIT

Explore websites, people, and locations

Too resources for:

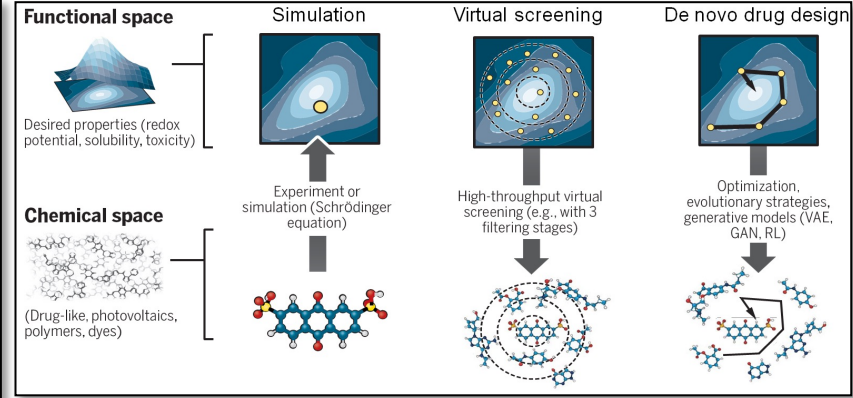
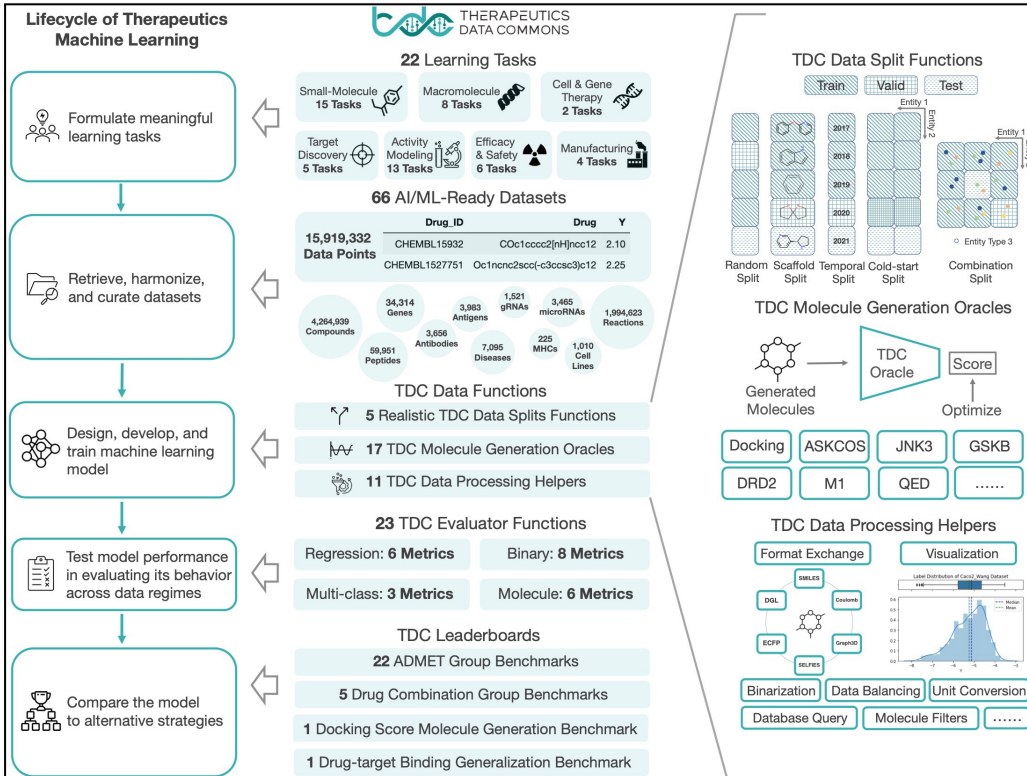
- _prospirative students
- _current students
- _faculty & staff
- _alumni
- _parents
- _David '99 and MIT
- _all resources

A new study maps out the cells, genes, and cellular pathways involved in exercise-induced weight loss. "High-fat diets push all of these cells and systems in one way, and exercise seems to be pushing them nearly all in the opposite way," says Manolis Kellis.

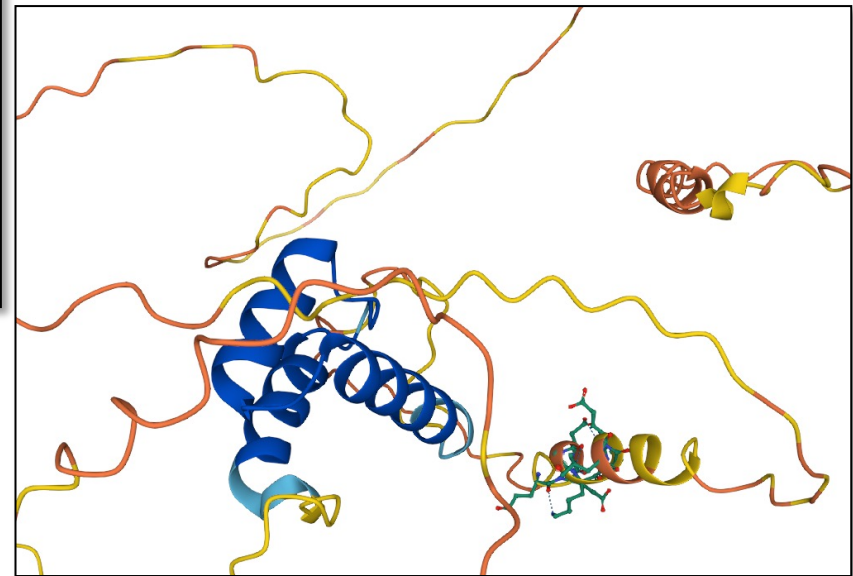
Oct 6, 2022 | Fall story | Thank | Explore more spotlights

Network-level therapeutic development

Connect TF ↔ motif ↔ SNP ↔ enhancer ↔ gene ↔ CellType ↔ protein ↔ drug ↔ pathway ↔ disease



Deep Learning Models Molecule Synthesis



Multiple therapeutic modalities. E.g. IRX3
Target: Protein, target sites, enhancer, RNA



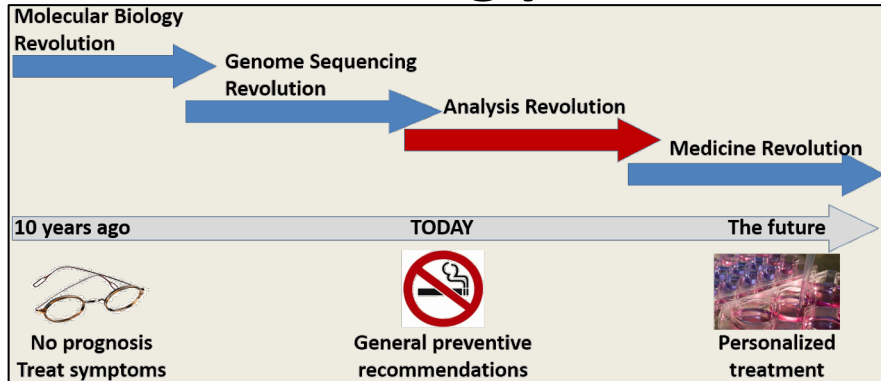
Henry Herce
Brad Pentelute
Marinka Zitnik

Disease still reigns



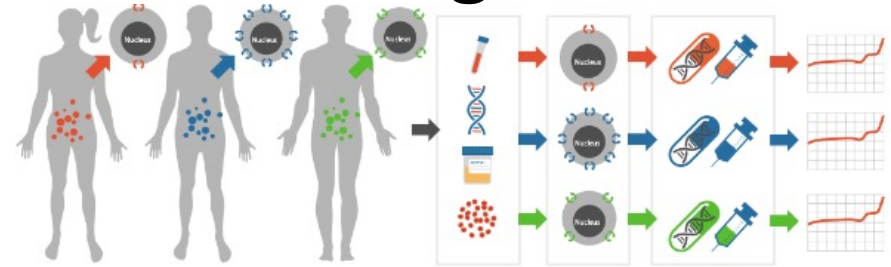
- **My own family:** Obesity, cancer, stroke, diabetes
- **My own predispositions:** obesity, blindness, cancer.
- **Genetics:** Each of us in this room carries mutations
- **Environment:** pollution, nutrition, sedentary lifestyle
- **Systemic disorders:** obesity, diabetes, cancer, heart
- **Pathogens:** infections, immune dysregulation, cancer
- **Lifespan:** Alzheimer's, new diseases

Transforming pharma



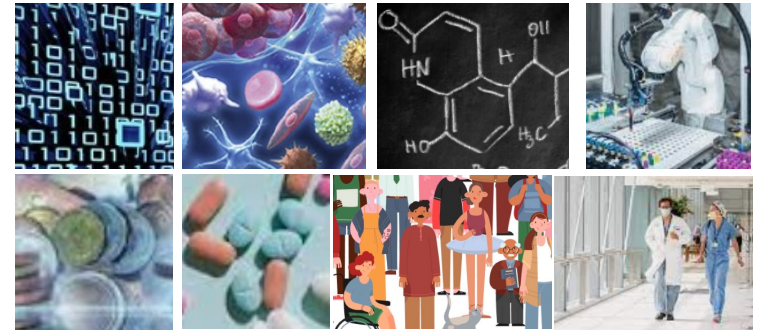
- Always surprised → Prognosis: Mendelian, Polygenic Scores
- Misdiagnosis → Better biomarkers, Multi-modal diagnosis
- Treat manifestations → Address root causes, causal hallmarks
- Monolithic: AD, T2D, Cancer → Heterogeneity: symptoms+causes
- Monolithic: AD → Understand components: Ab, tau, infl, lipids
- Silos: tissues, departments → Interplay, commonalities, sharing
- Treatment too late → Preventive personalized interventions

Personalizing Medicine



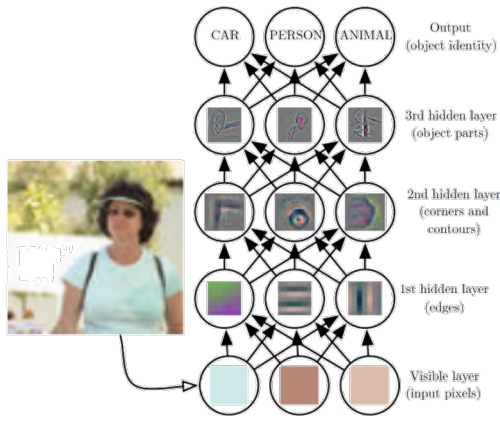
- **Polygenicity:** Thousands of variants
- **Convergence:** Small number of common pathways
- **Hallmarks of disease:** causal pathways
- **Manipulation:** reverse disease circuitry
- **Individualized treatment:** combine pathways
- **Each Patient:** different combination
- **Burden:** Accumulation of pathway perturbations
- **Omics:** Genetic, epigenomic, transcript, proteomic

Call to action: Coalition

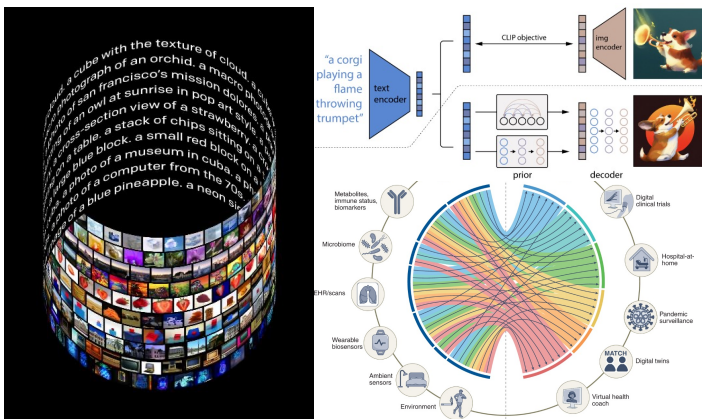
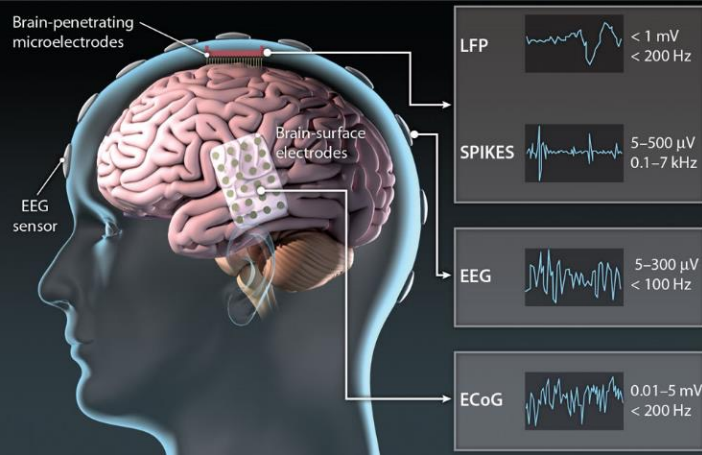


- **CS:** ML, DeepNN, DNA code, circuitry, big data
- **Bio:** High-throughput profiling + manipulation
- **Chemistry:** Libraries, synthesis, modularity
- **Biotech:** New technol. for rewiring, delivery
- **Finance:** long-term 10-year 20-year 'biobonds'
- **Pharma:** partnership, pre-competitive sharing
- **Patients:** empowrmnt, personalization, sharing
- **Hospitals:** combine cohorts, increase power

The role of CS+AI in Transforming Medicine

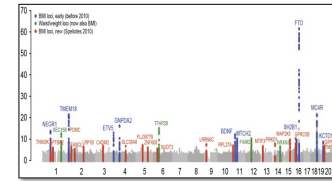


- Computational programming language of the genome
- Systems-level interventions for rewiring cellular circuitry
- Information-theoretic prioritization of experiments
- Genomic transformation of food production/resilience
- Design organisms that use CO₂ for energy production
- Understand/reverse pseudotime progression of aging
- Brain information storage, HCI direct information transfer
- Brain-inspired new deep learning cognitive architectures
- Million-fold multiplex perturbations and measurements
- Hardware acceleration, compressed computing
- Robustness-first programming/robotics paradigms
- Evolution/Evolvability-first system design paradigms
- Complete multi-modal understand. of EHR/medical state
- Design of new proteins/drugs/therapeutic structures
- Cracking & reversing circuitry of Alzheimer's, Obesity, Psychiatric, Cardiac, Immune, Cancer, all of disease
- Representation learning as a window to artificial intelligence
- Systems-level understanding of biological functions and processes



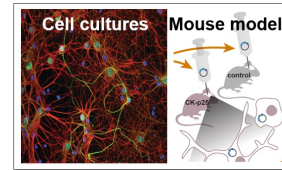
So come to MIT/CSAIL/CompBio – and join us! :-)

Big Data
+ Machine Learning
+ Experimental Validation
→ Impact + Fun!

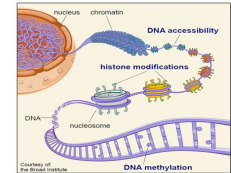


Start with disease genetics:
Common + rare variants

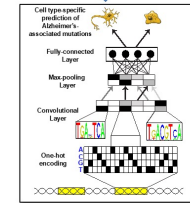
4. Disseminate results



3. Validate predictions in
human cells + mouse models



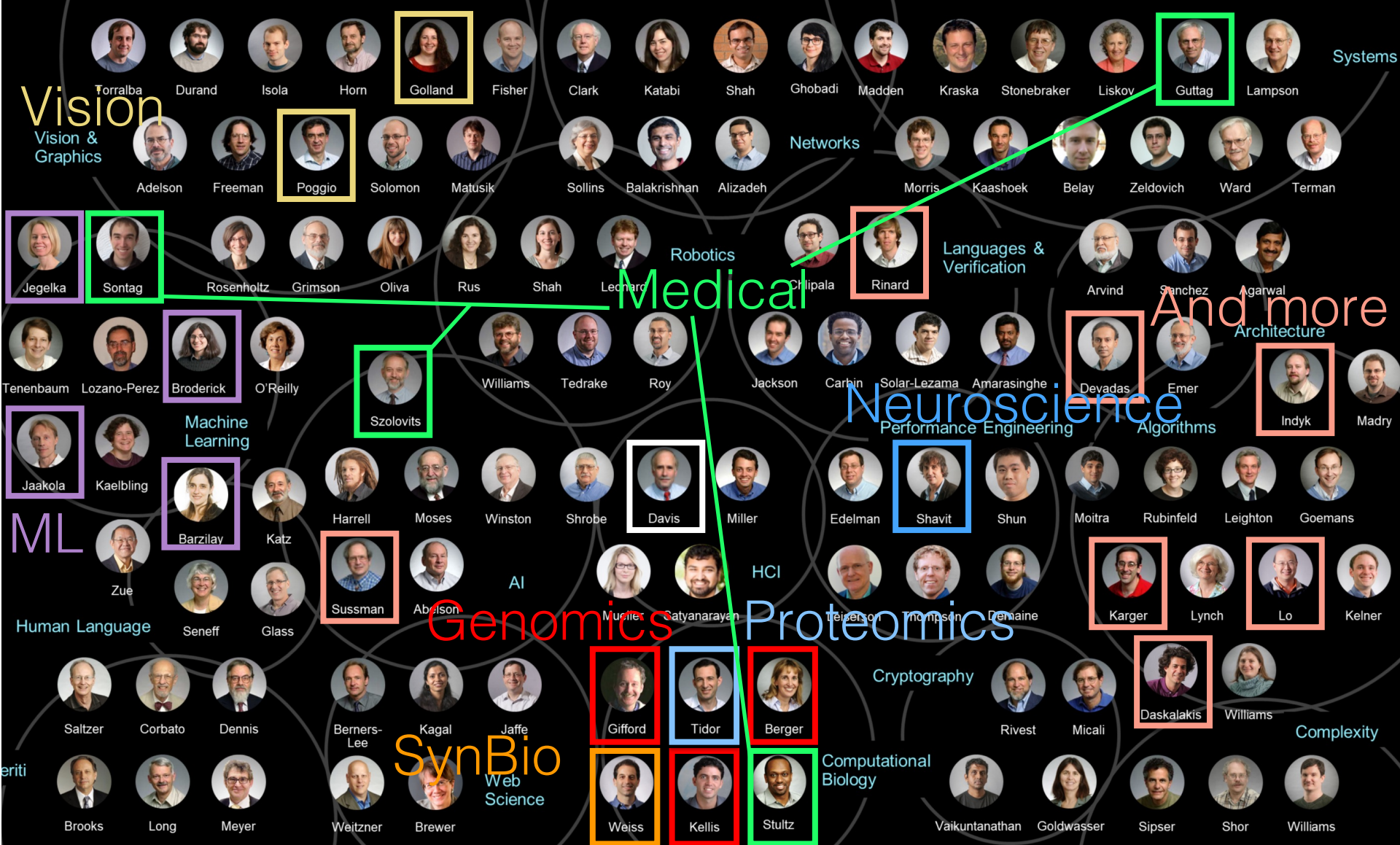
1. Profile RNA + Epigenome
in healthy + disease samples



2. Integrate data to predict driver
genes, regions, cell types²

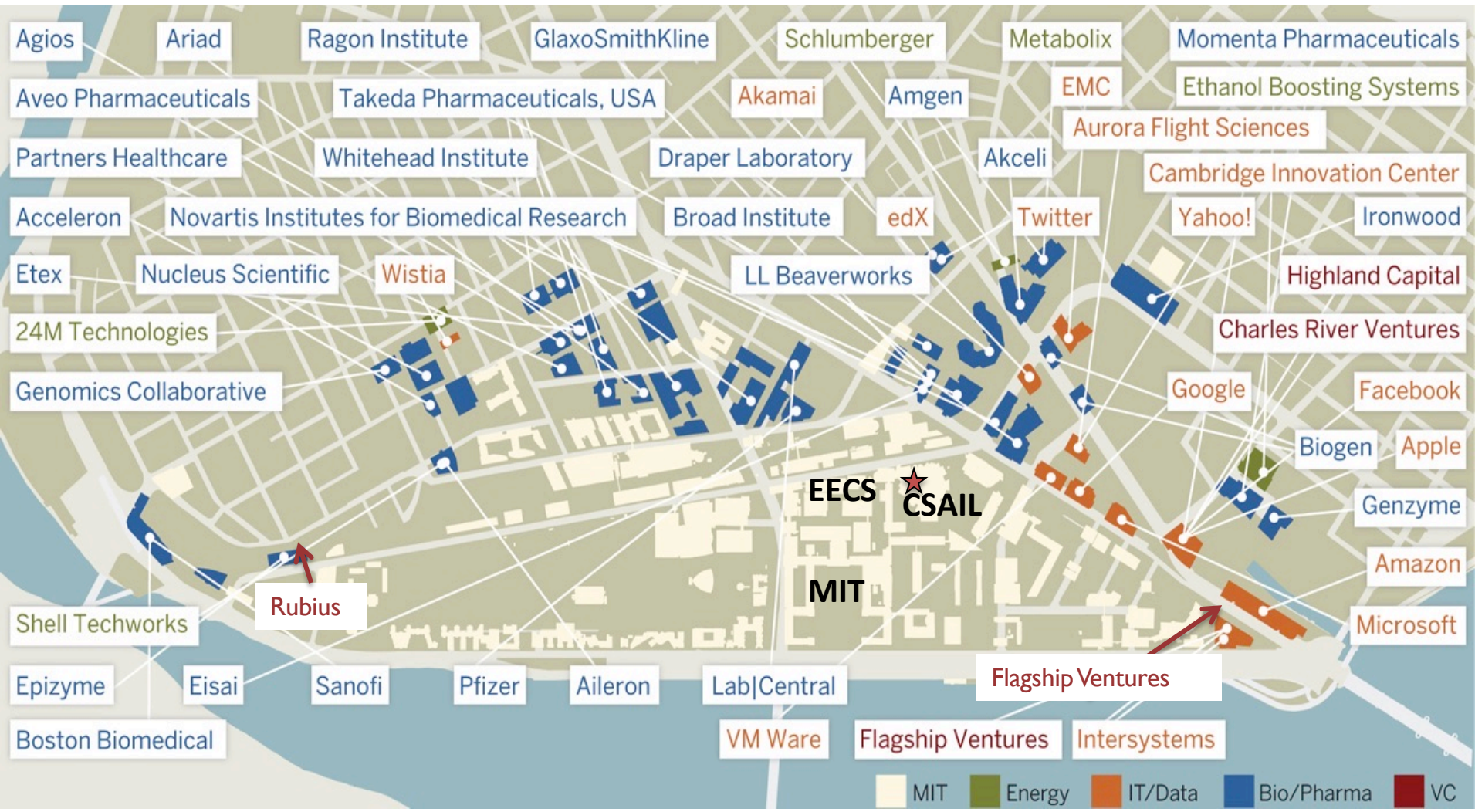
compbio.mit.edu – Prof. Manolis Kellis





CSAIL faculty with research programs in CB span many areas: ML, AI, Vision, Theory, Systems, Languages, Architecture, Computational Biology.

Pharma/Biotech surrounds MIT - Lo (many started by MIT faculty + trainees)



Prof. Philip Sharp, 2017, modified by H. Sive