Integrating Model Organism Data around Clinical Genomics

Calum A. MacRae MD, PhD

Brigham and Women's Hospital
Harvard Medical School
Broad Institute of Harvard and MIT
Harvard Stem Cell Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL
BIOMEDICAL RESEARCH INSTITUTE

HARVARD MEDICAL SCHOOL
TEACHING AFFILIATE

Sudden death: the genotype/phenotype problem

- Single gene disorder
- **Laminopathy**
- Perfect segregation (LOD>12)
- Large effect size for SCD: 500 -10,000X risk

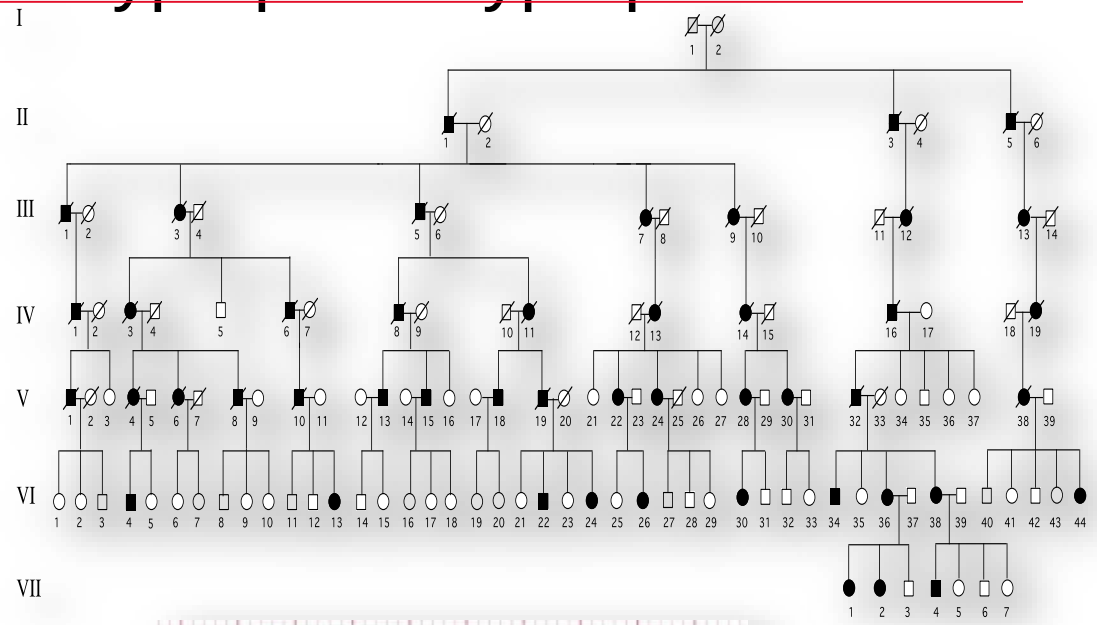
- Multiple phenotypes in a single family
  - Asymptomatic EKG findings
  - CHF
  - Sudden death

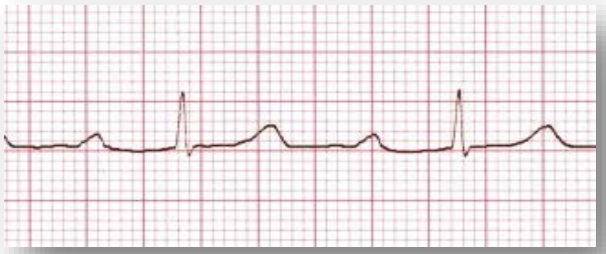
  - 12 different lamin syndromes reported

- “Modifiers”
  - Genetic
  - Epigenetic
  - Environmental

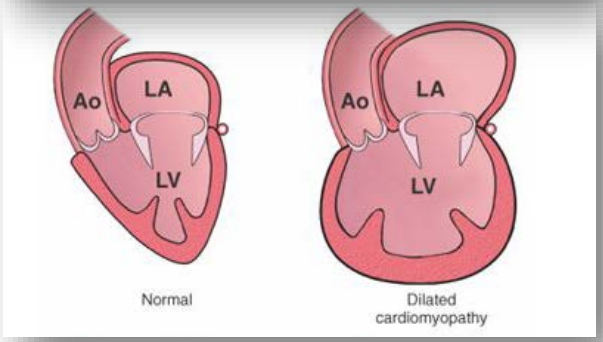
- No empiric support for any modifier model

- Insufficient **additional** information to change clinical care:
  - Improve symptoms
  - Improve outcomes



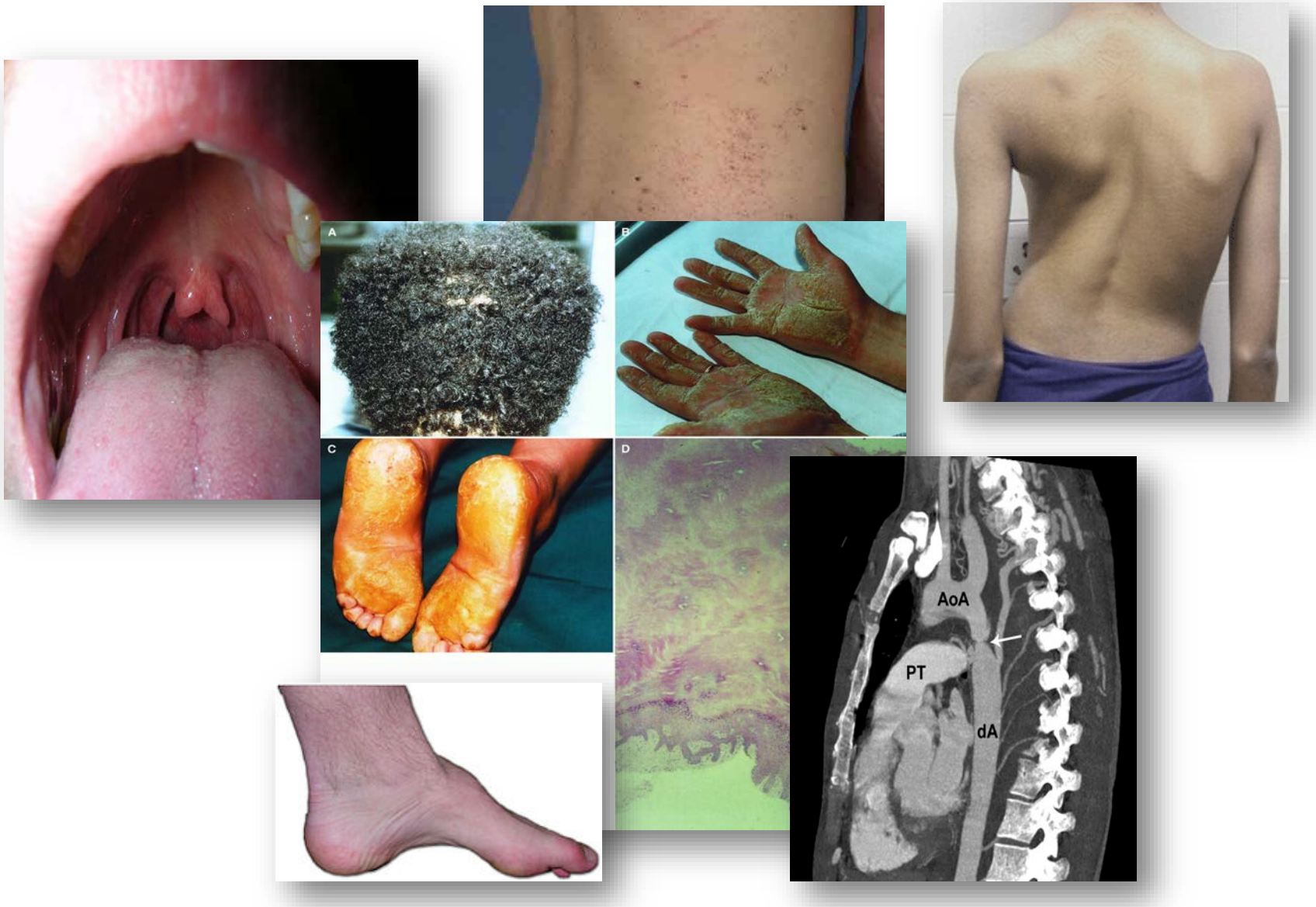


AV Block



DCM

Extracardiac phenotypes in cardiomyopathy



Clinical genomics: the other extreme

- MedSeq Study (CSER)
  - RCT of WGS
  - Healthy primary care cohort
  - Cardiomyopathy cohort
- Likely pathogenic KCNQ1 variant identified in a primary care patient
- PCP/Patient disclosure associated with anxiety attack and immediate concern re sudden death risk
- “Feeling better or living longer”

A. MONOGENIC DISEASE RISK: 1 VARIANT IDENTIFIED

This test identified 1 genetic variant that may be responsible for existing disease or the development of disease in this individual's lifetime.

Disease (Inheritance)	Phenotype	Gene (Variant)	Classification
A1. Romano-Ward syndrome (Autosomal dominant)	QT prolongation with risk for syncope and sudden cardiac arrest	KCNQ1 (c.826delT p.Ser276ProfsX13)	Likely Pathogenic

B. CARRIER RISK: 5 VARIANTS IDENTIFIED

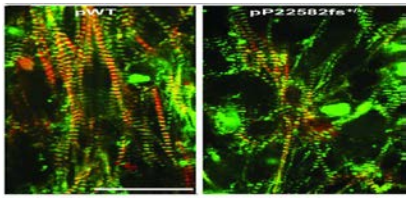
This test identified carrier status for 5 autosomal recessive disorders.

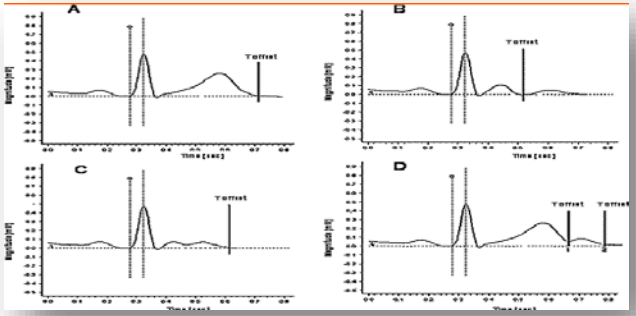
Disease (Inheritance)	Phenotype	Gene (Variant)	Classification	Carrier Phenotype*
B1. Usher syndrome type III (Autosomal recessive)	Hearing loss, retinitis pigmentosa, and vestibular dysfunction	CLRN1 (c.528T>G p.Tyr176X)	Pathogenic	None Reported
B2. Primary congenital glaucoma (Autosomal recessive)	Increased intraocular pressure	CYP1B1 (c.171G>A p.Trp57X)	Pathogenic	Late onset glaucoma (case report only)
B3. Recurrent hydatidiform mole (Autosomal recessive)	Mass or growth that forms inside the womb	NLRP7 (c.337_338insG p.Glu113GlyfsX7)	Pathogenic	None Reported
B4. Jervell and Lange-Nielsen syndrome (Autosomal recessive)	Congenital profound bilateral sensorineural hearing loss and long QT	KCNQ1 (c.826delT p.Ser276ProfsX13)	Likely Pathogenic	Romano-Ward syndrome (see above)
B5. Alpha-N-acetylgalactosaminidase deficiency (Autosomal recessive)	Variable infantile neuroaxonal dystrophy	NAGA (c.479C>G p.Ser160Cys)	Likely pathogenic	None Reported

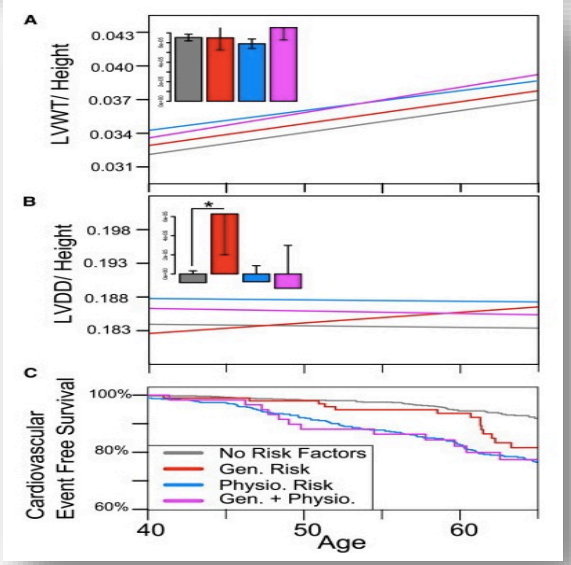
As a carrier for recessive genetic variants, this individual is at higher risk for having a child with one or more of these highly penetrant disorders. To determine the risk for this individual's future children to be affected, the partner of this individual would also need to be tested for these variants. Other biologically related family members may also be carriers of these variants. *Carriers for some recessive disorders may be at risk for certain phenotypes. Please see variant descriptions for more information.

Pathogenicity assessment

- In vitro assays
- In vivo assays
- Patient
  "‥ but my QT was normal"
- Is it always the phenotype?
  - Segregation
  - Penetrance
  - Pleiotropy
- Relationship between all of these metrics and risk obscure
- We need quantitative assays mapped onto people







Bick et al AJHG 2012

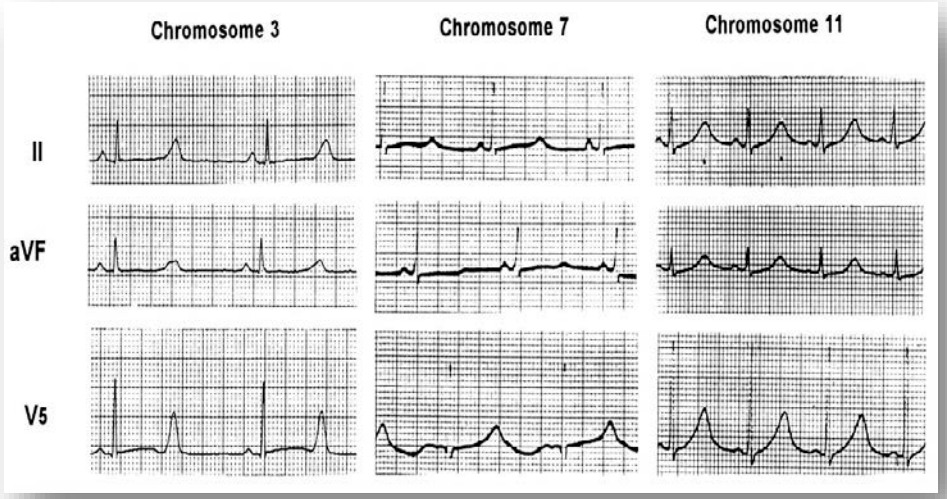
Potential clinical studies

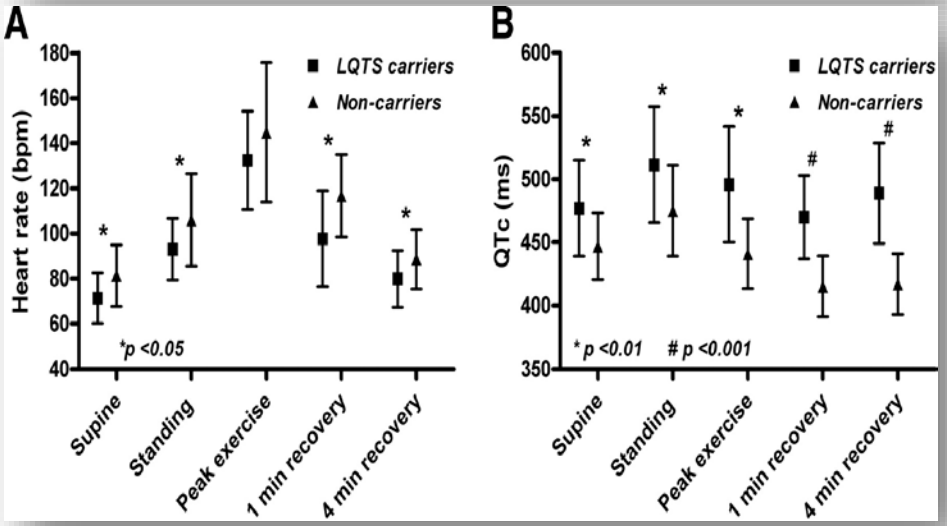
- QT
- QTc
- ECG morphology
- Subclinical /extracardiac phenotypes

- Provoked phenotypes
  - Posture
  - Exercise
  - Recovery

- Signal: noise

- Risk
  - “Am I at risk of sudden death”
  - Is the risk associated with genotype or phenotype?

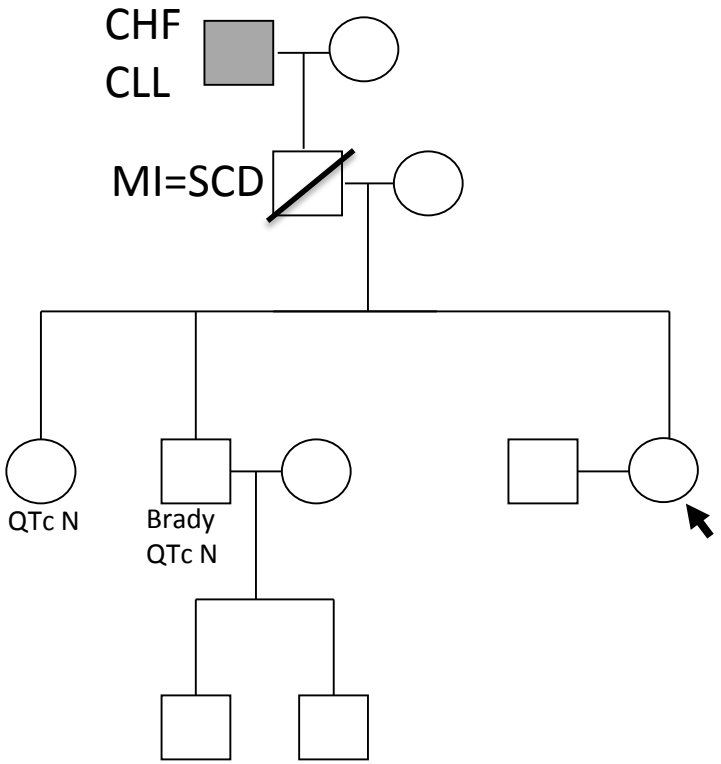




Krahn et el. 2012

Family study reveals ‘overlap syndrome’

- ‘Condition-specific’ family history
- Proband
  - Physical exam-**S4 and ESM**
  - QT-**466ms**
  - QTc-**461ms**
  - EKG morphology-**Normal**
  - Echo-**Focal LVH and MV thickening**
  - MRI-**Normal**
  - Provoked phenotypes
    - QTc at 4 mins recovery **400ms**



- Definite abnormalities observed
- ? Phenotype expansion
- ? False positive

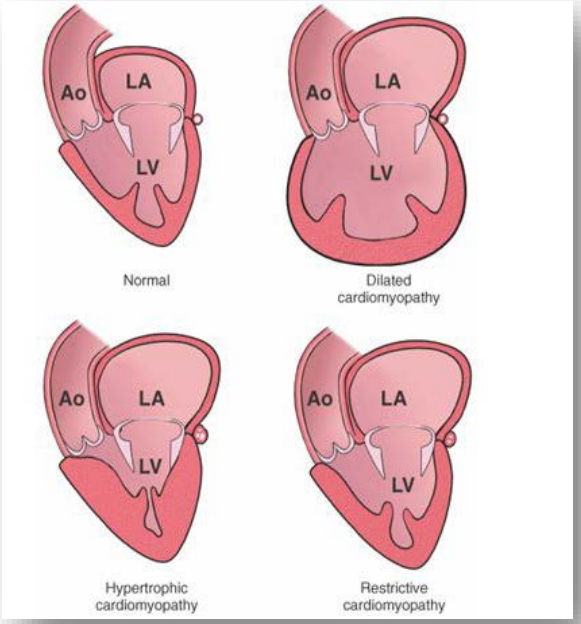
- Genotypic and phenotypic uncertainty
- Actual risk – unmeasured

- Additional clinical and genetic testing>$8000
- Remember this is a ‘**known’ gene and a typical family**

Phenotype is now limiting in multiple arenas

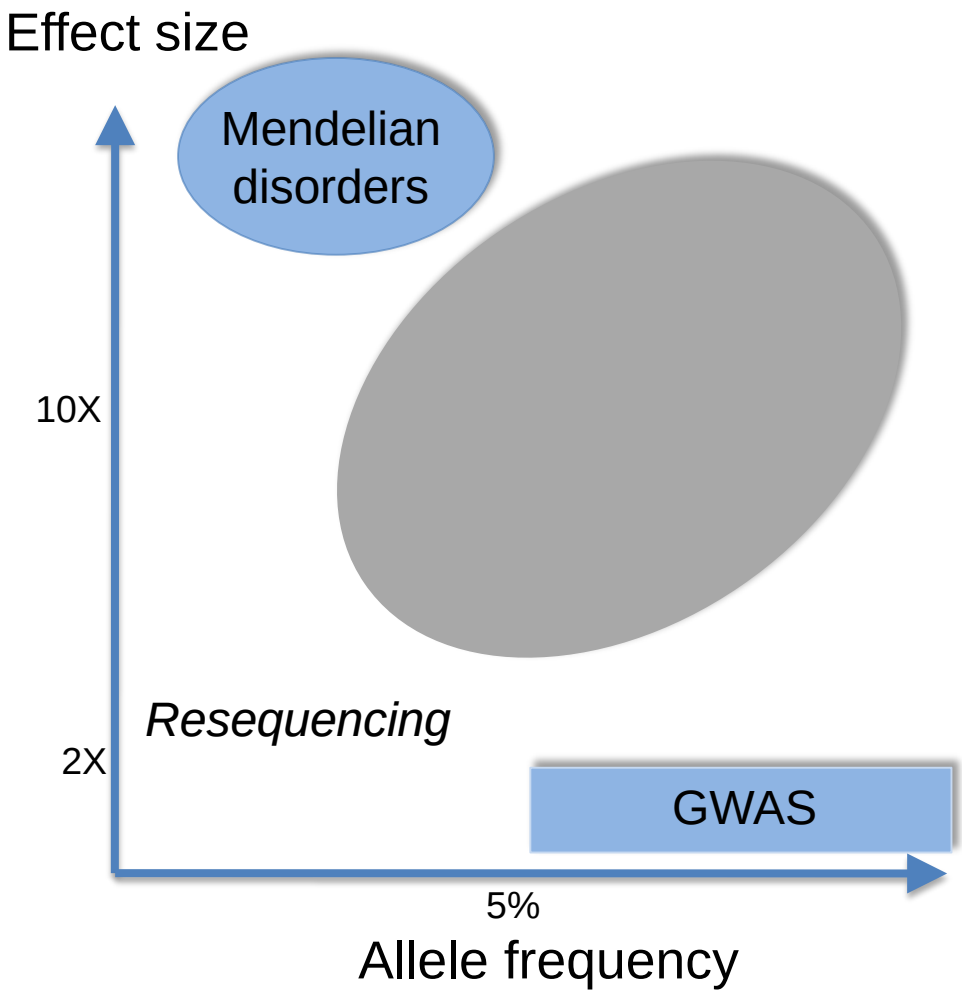
- Clinical care
- Genetics/genomics
- Precision medicine

- Fundamental issues
  - Morphology dominates
  - Semi-subjective at best
  - Late or even end-stage
  - Aggregation for statistical power
  - Legacy – better at measuring same old phenotypes
  - Binary
  - Cross-sectional
  - No systematic perturbations



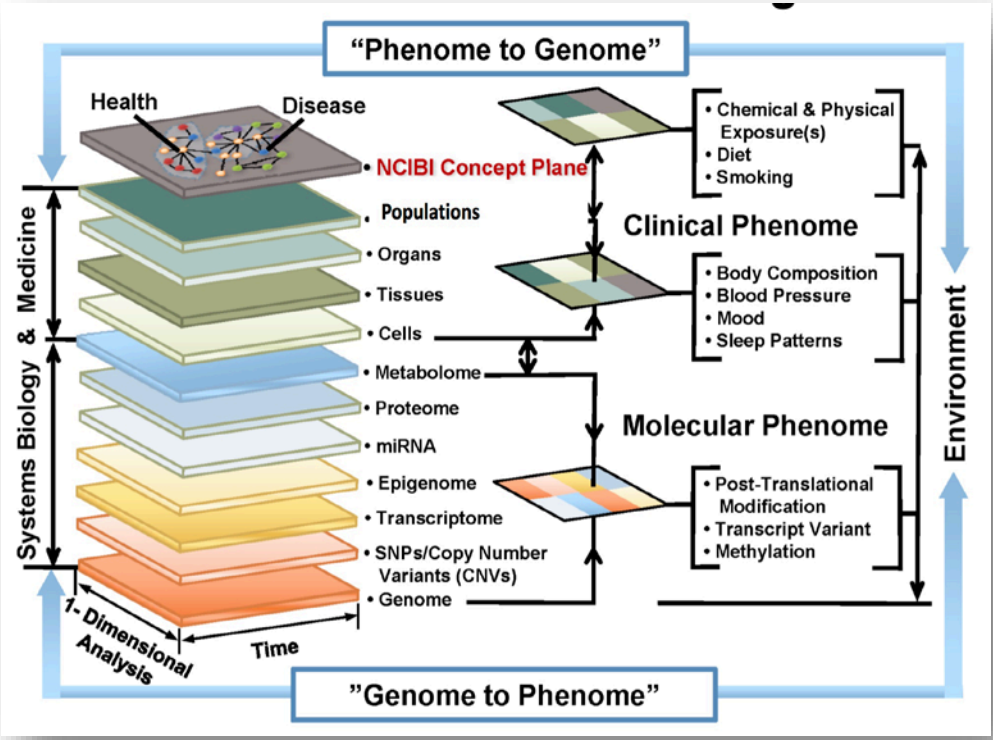
Where is all the information?

- Why are alleles ‘silent’?
  - Inaccessible to current study designs
  - Inaccessible to current assays
  - Unmeasured conditioning variables
- Genetic architecture dependent on phenotypic architecture
  - **Phenotypic resolution**
  - Selection pressures
  - Environmental contributions
  - Not assessed for most disease traits
- Limitations of genetics to date
  - Focused on extreme phenotypes
  - Few prospective cohorts
  - If familiality detectable how many genes involved?
  - Heterogeneity also scales: GWAS
- Disease definitions always evolving
  - Overlapping causes
  - Overlapping therapies
  - **More precise** medicine



How might model organisms help?

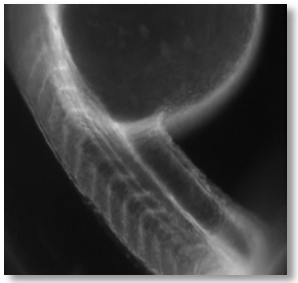
- Saturation screens: to identify all of the genes for a given trait
  - Phenotype anchoring for validation
  - Extreme perturbation
  - Not just F3 recovery but all of the alleles (phenotype)
- Reverse genetics: Manipulate each gene and explore phenotypic ‘universe’
  - KOMP, Zebrafish mutant project, other organisms
  - **Phenotype expansion feasible including functional genomics**
- Environmental modeling: generate **provoked phenotypes**
  - Dynamic responses
  - Few attempts at in vivo disease screens across environmental space
  - Drug discovery as a special case
- **Identify gaps in genetic or phenotypic architecture**
- **Iterative systems level modeling**
- **Mapped to human genotype and phenotype**



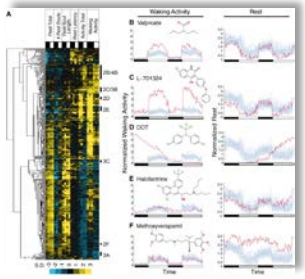
Zak Kohane

Model organisms: scalable parallel phenotyping

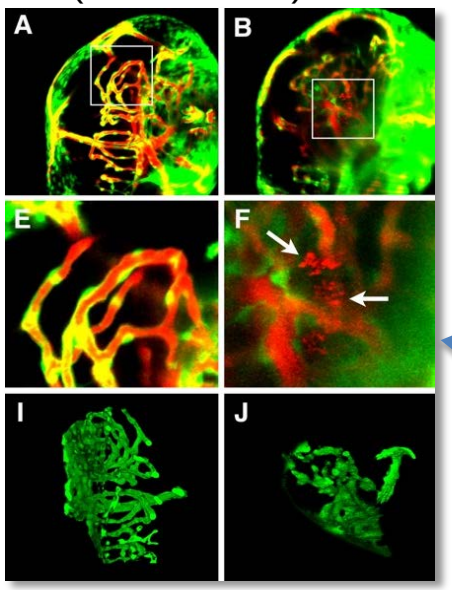
Molecular imaging



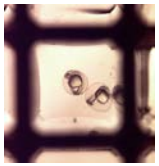
Behavior (Caffeine)



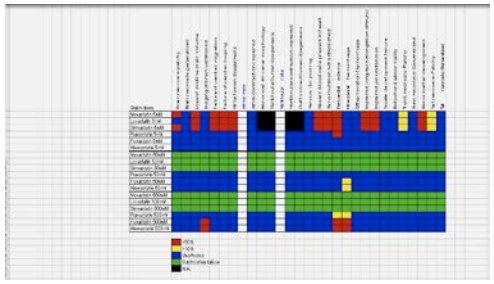
Vascular permeability
(Statin/COX2)

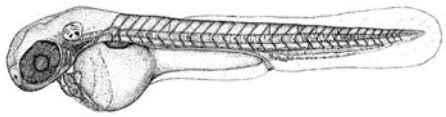


Contractility

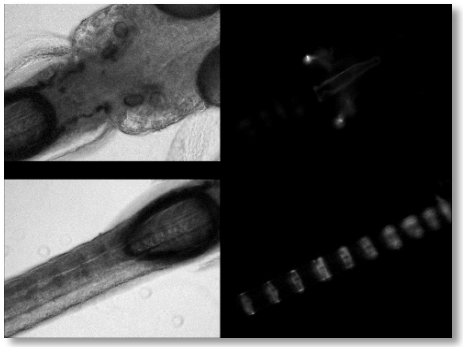


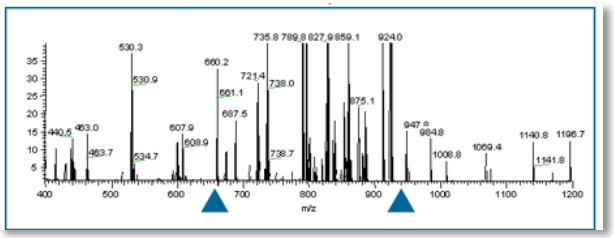
Phenoclustering





Bone calcification

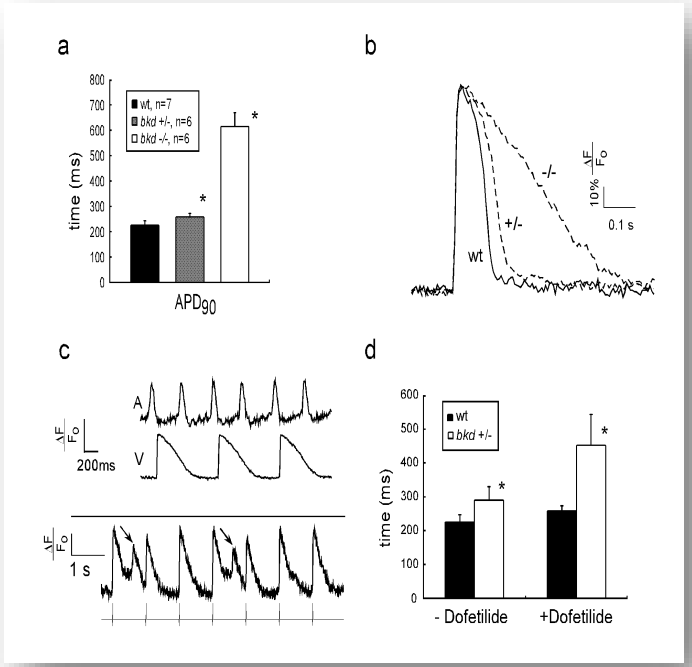


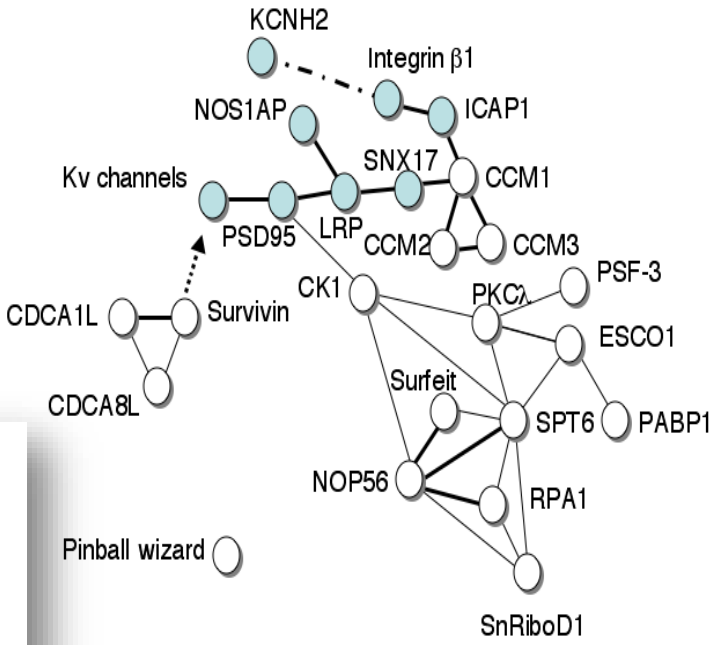


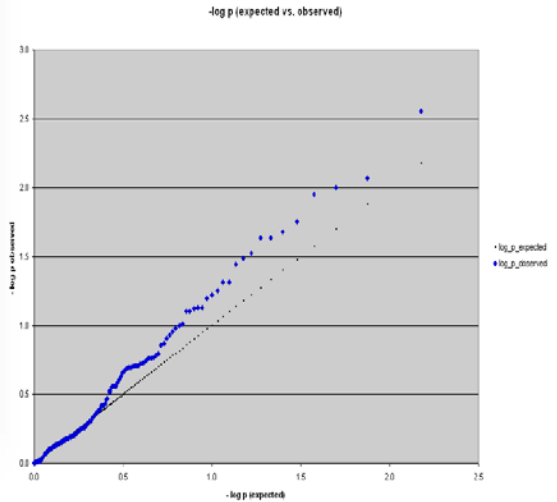
Metabolomics – biomarker discovery

A shelf screen for QT

Optical mapping







Blocks in translation: AFib Genetics

- Formal kin-cohort study-220 families
- High narrow sense heritability high
- Environmental triggers

- Large Mendelian loci identified

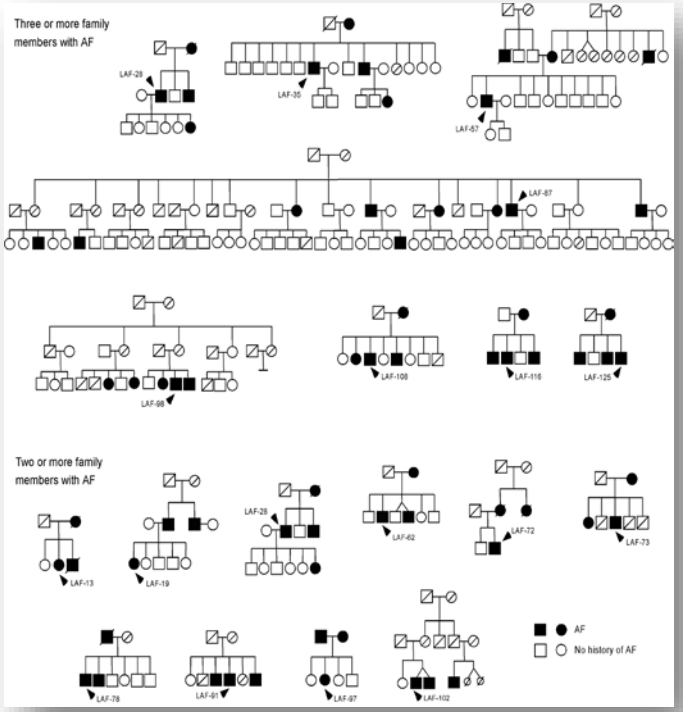
- ~ 10% of heritability explained by GWAS loci

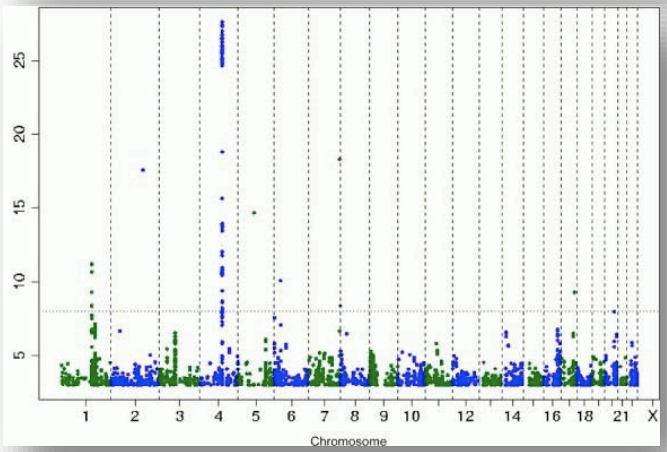
- Missing intermediate effect sizes

- Difficult to clone genes where large effects because we cannot reliably identify unaffected individuals
- Different major effects in each family

Need to:

- Explore existing pathways identified in man
- Define better phenotypes
  - Biomarkers
  - New structural or functional assays
  - “AF threshold”





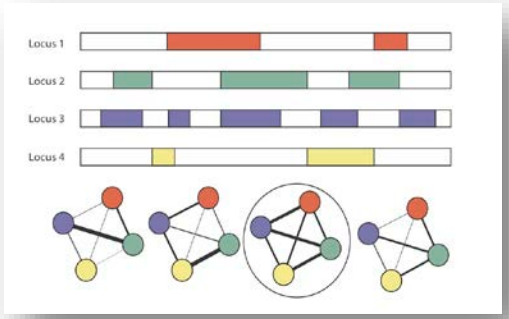
Genotype but no phenotype

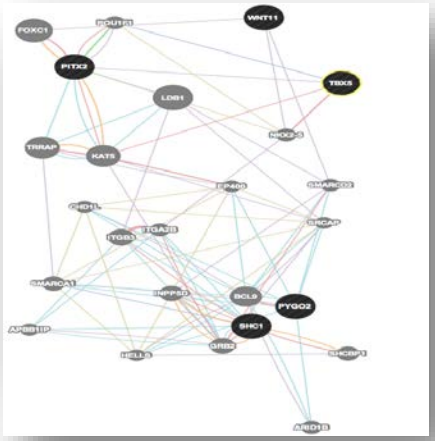
- 12 GWAS loci for AFib
- All genes/miRNAs/linc RNAs within 3Mb

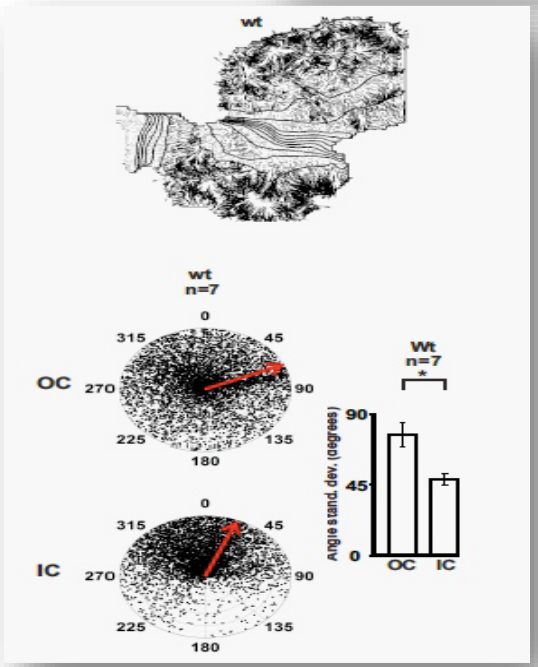
- Identify shared network
- Permutation to maximize functional linkage information

- Network of cell coupling pathway genes identified
  - Perturb primary cell circuitry in heart

- Human phenotype rate-limiting
- No pre-event biology

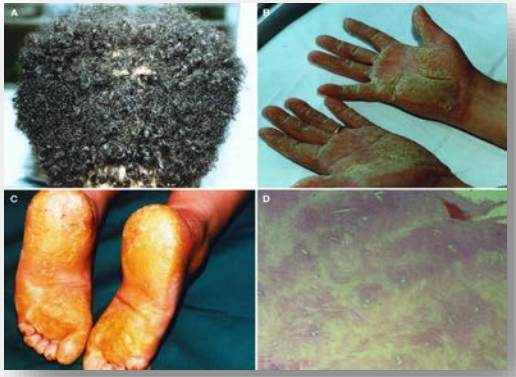




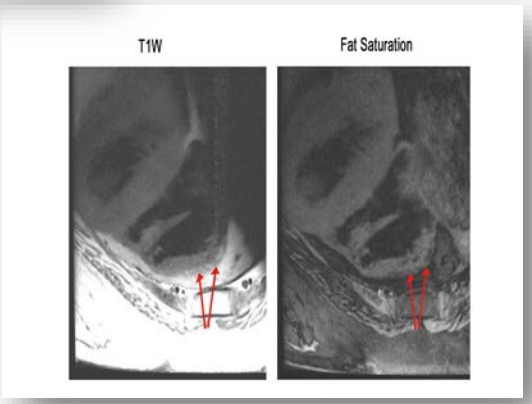


Modeling chronic disease in 5 days in a fish

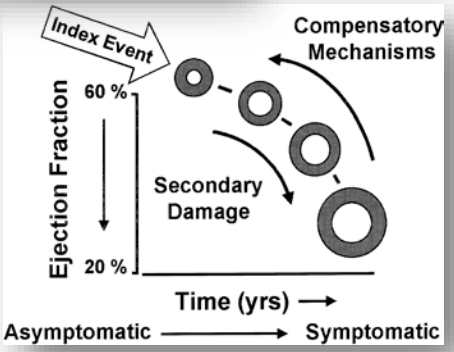
- Poorly penetrant



- Arrhythmia
- Sudden death
- Cutaneous abnormalities

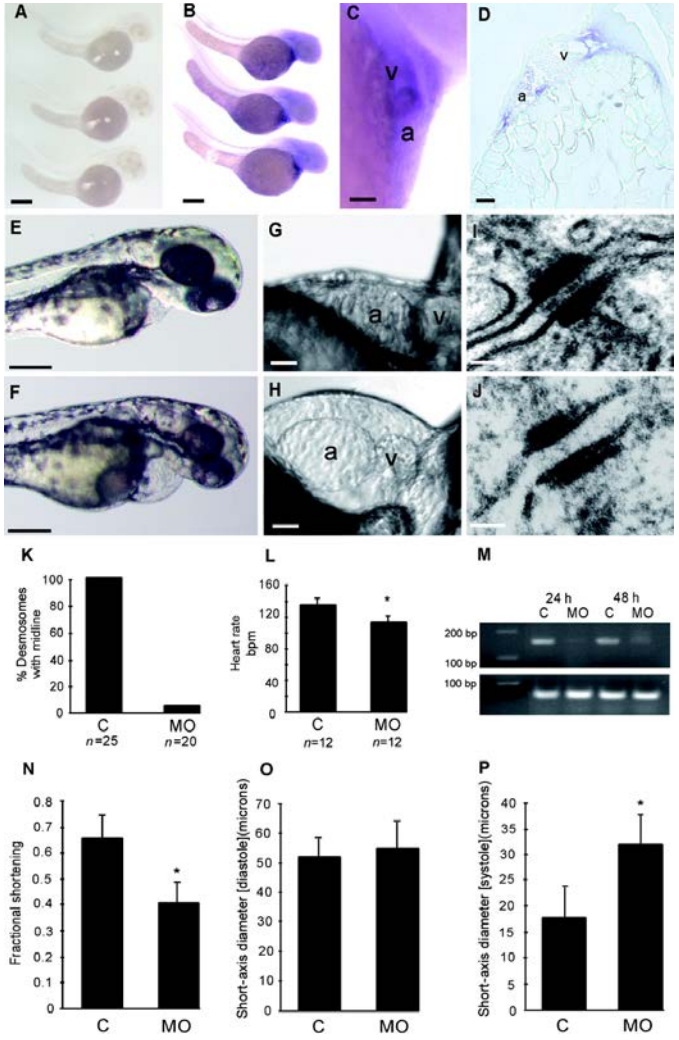


- Contractile abnormalities
  - Congestive heart failure
  - Biomarker abnormalities(nt-BNP)

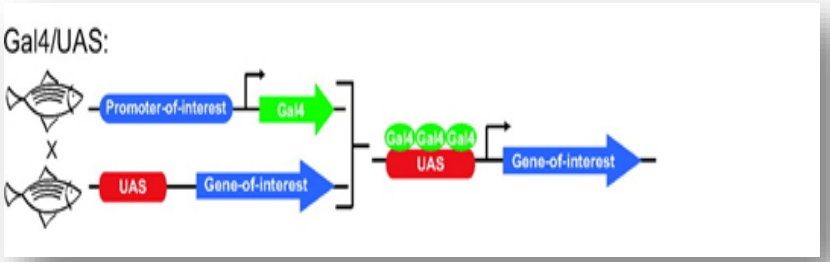


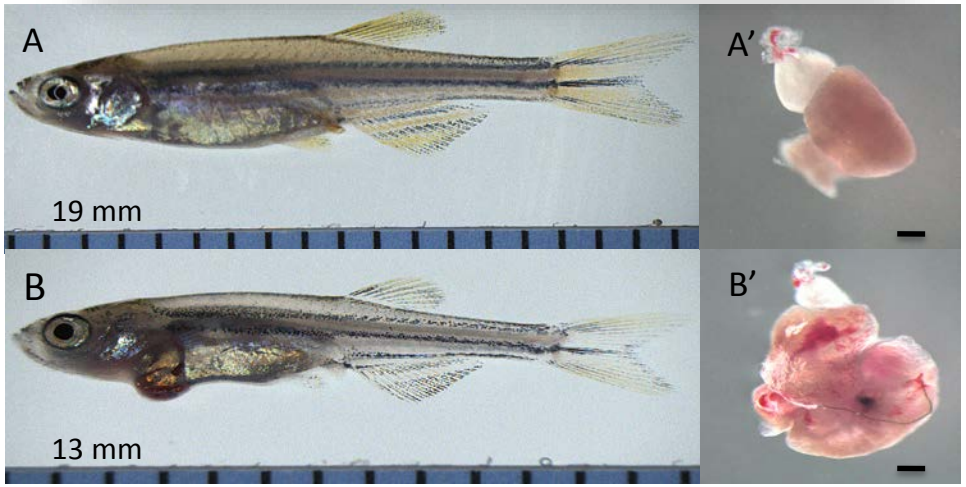
- Desmosomal gene mutations
- Mechanism unclear
  - Wnt signaling perturbed

Genotype anchoring



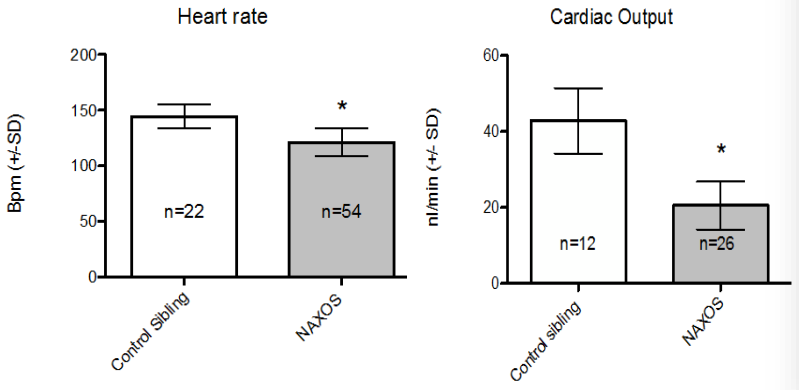
- Multiple disease alleles modeled
  - Morpholino, CRISPR, rescue, transgenesis
  - Recapitulate structure and function
- Modeling human allelic series
- Conditional germline mutant (GAL4::UAS)
  - Allows screening

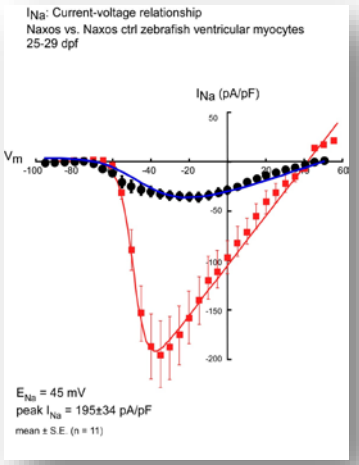


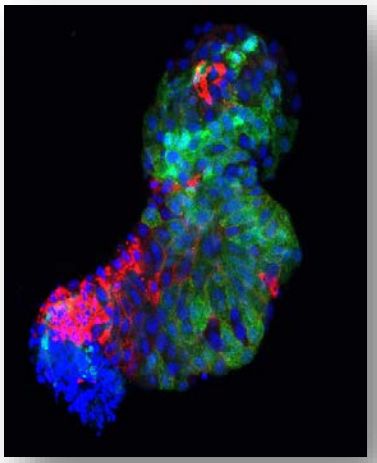


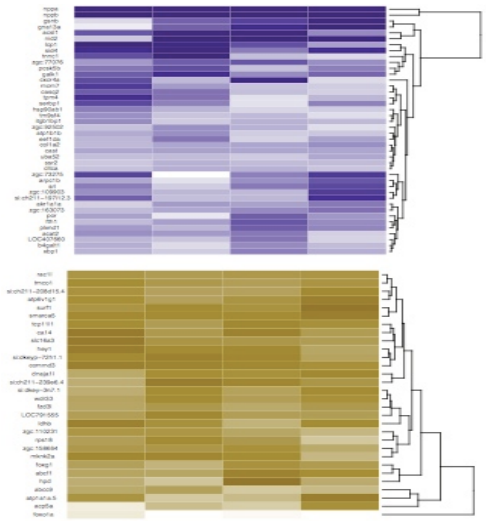
— =1µm

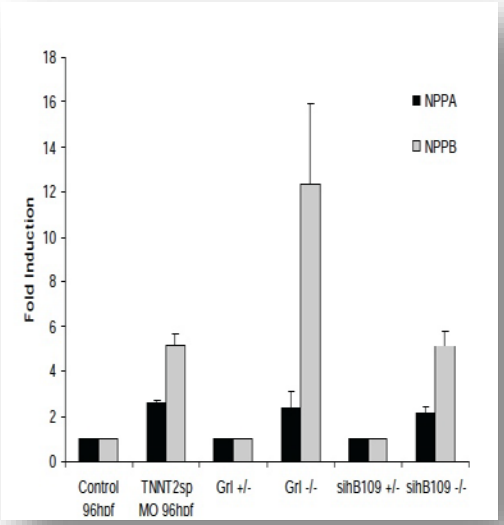
Phenotype anchoring

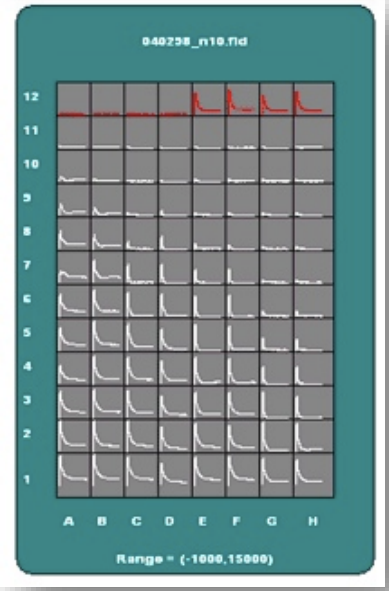










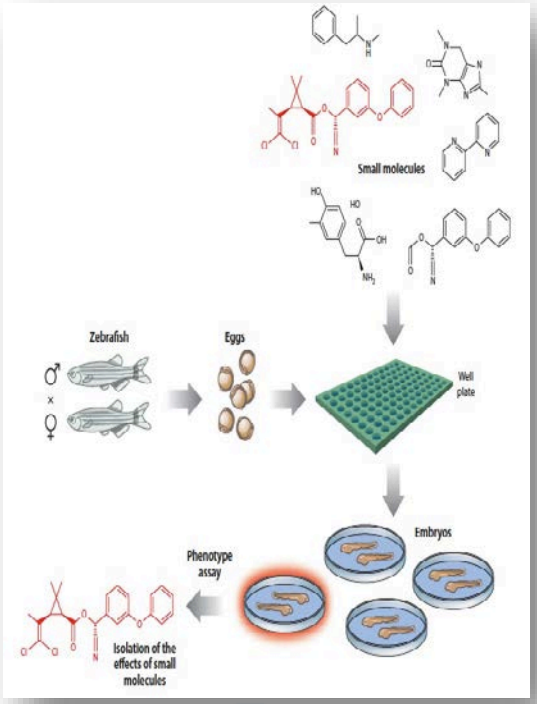


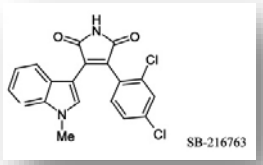
Transcriptomics

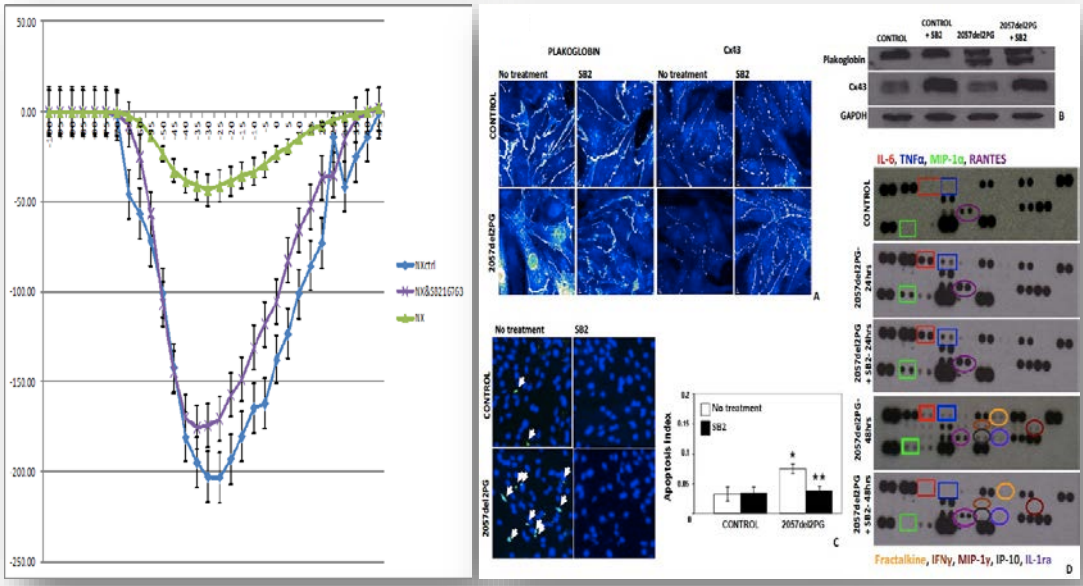
Natriuretic peptide reporter

Luciferase for screen

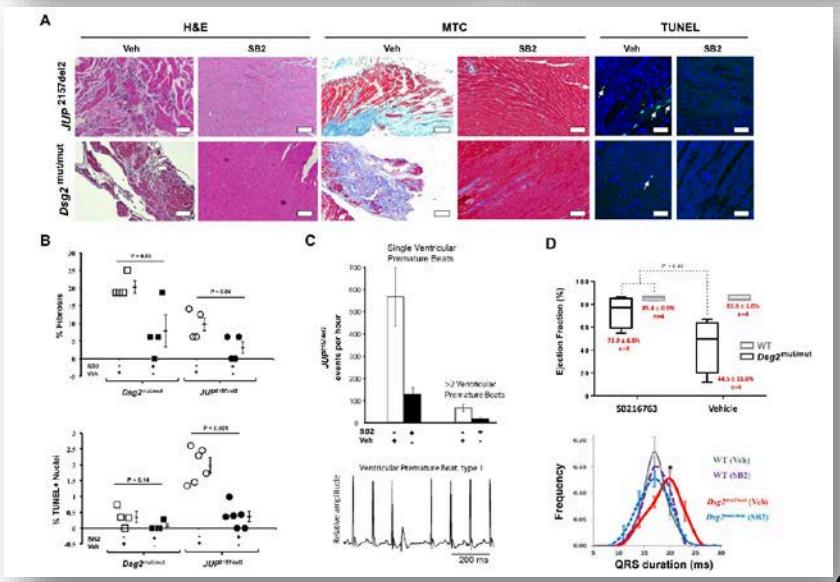
High throughput screen





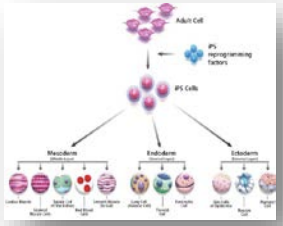


Asimaki et al.



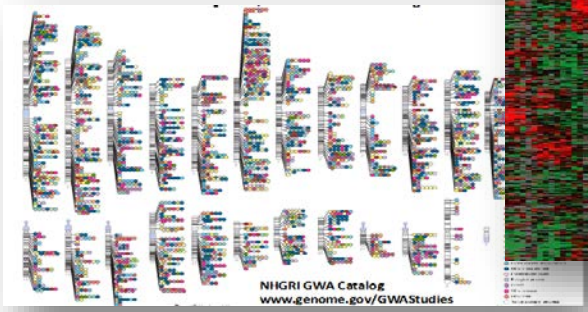
Chelko et al

The ‘phenotype gap’



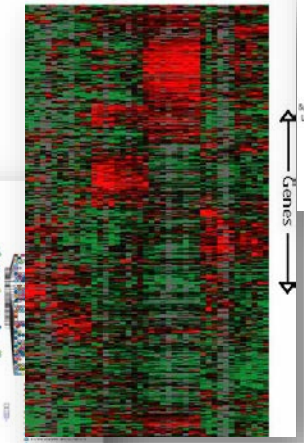
Single cell multipliers

Genome



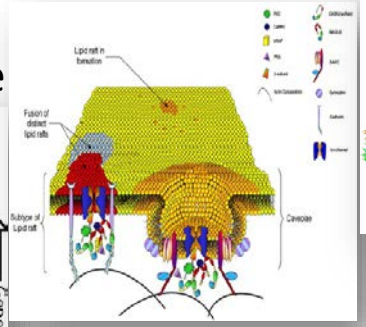
10^9

Transcriptome



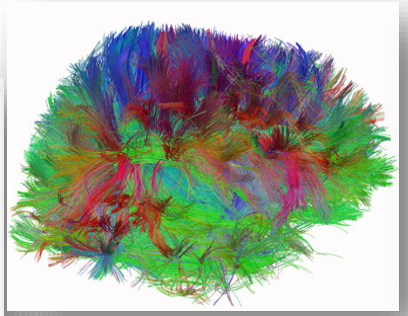
10^{16}

Proteome



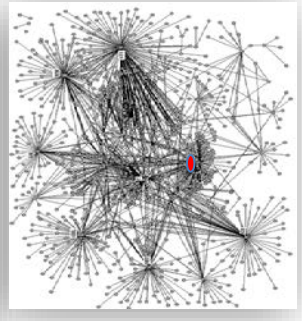
10^{20}

Connectome



10^{69}

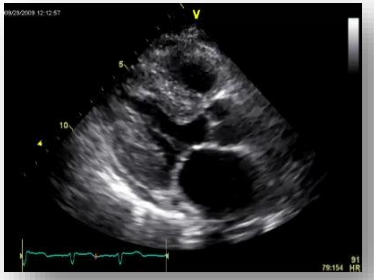
Physiome



Timescales
Perturbations

Exposome

All clinical phenotypes



10^4

We need new translatable human phenotypes

- Current syndromes are really aggregates of many different disorders dating from ~1800s
  - Diabetes
  - High blood pressure
  - Cardiovascular diseases

- Different clinical outcomes
- Different therapeutic responses

Glucose
Taste

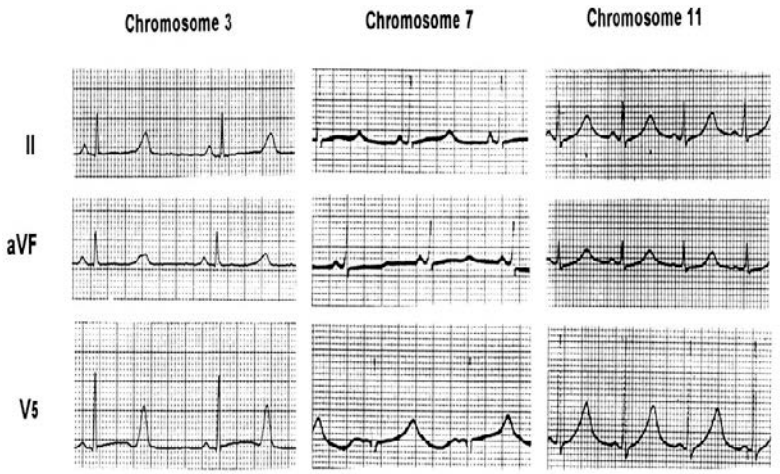
Cholesterol
Visible

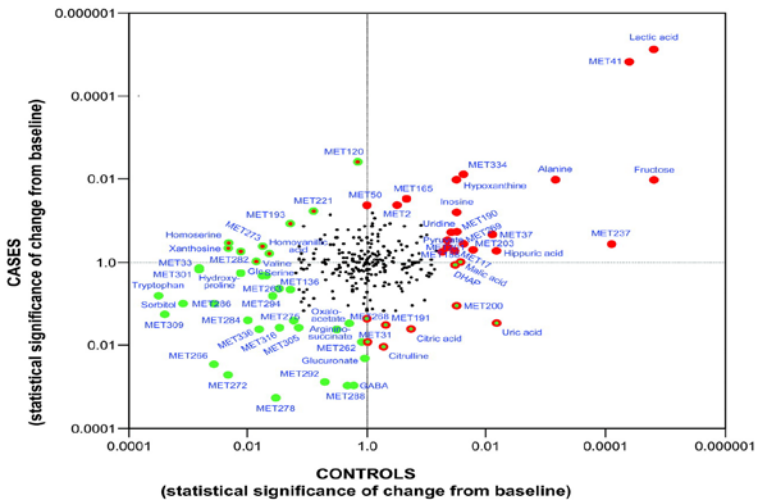
- We have focused on measuring serendipitous endpoints more precisely

- Deliberate reduction in complexity
- Limited dimensionality
- No clear organizing stimulus

Reappraisal of existing data types

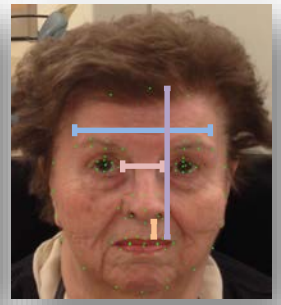
- **Focus on resolution and computability**
- Collect structured data in eHR
- Reanalysis of existing datasets
  - **Standardized acquisition**
  - New analytic approaches
    - Machine learning defines new EKG subsets
  - Infrastructure
    - Storage
    - Computation
    - Data display
- Functional genomics
  - New comprehensive datasets
  - e.g. Metabolomics

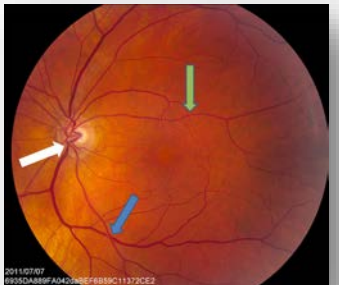


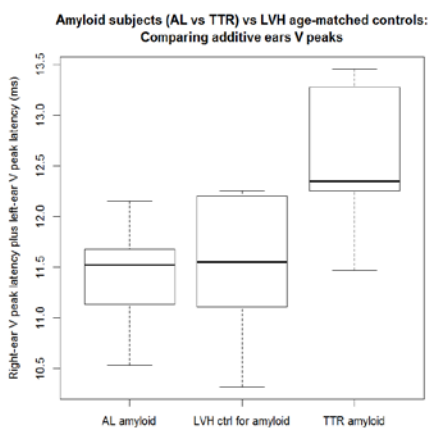


‘Next generation’ phenotyping

- Ambient technologies
  - Patient entered data- integrated with EHR
  - Symptom ontologies
  - Integrated autonomic testing
  - Retinal scans
  - Thermography
- Novel devices
- Exposome
  - GPS and geospatial maps
- Drug responses
  - Microdosing
  - Caffeine example







Auditory evoked responses

Stride length

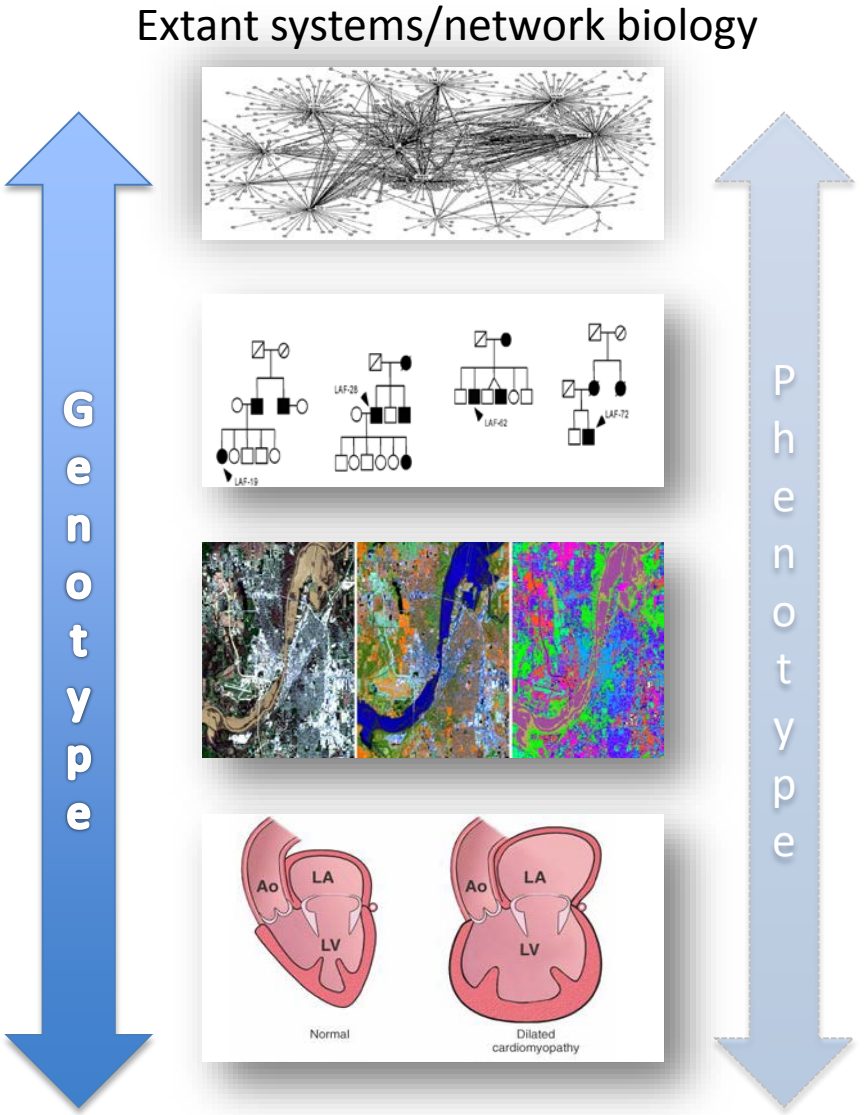
Rigorous probability estimates

- Family history-quantitative
- Population lifetime risk studies
- Network structures and responses
- Measured exposures

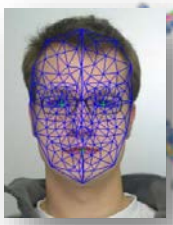
- Shared phenotypic lexicon
- ‘Mechanistic’ phenotypes
- Bidirectional ‘learning’ information systems
- Co-clinical modeling

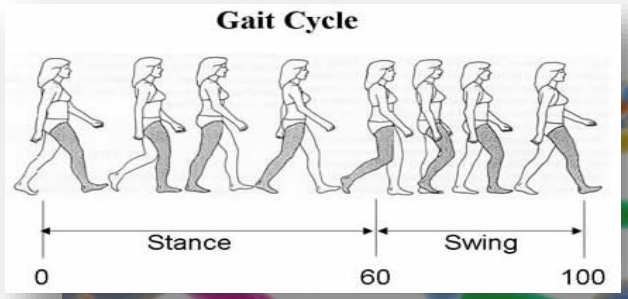
- **All at population scale**

- **Comprehensive multi-scale dynamic phenotyping**



Integrating clinical care and translation





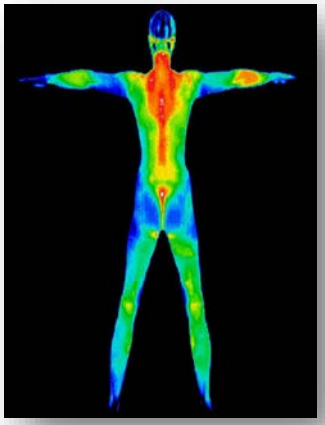
Proxsense

‘Next generation computable physical exam’

- **Technology benchmarking and validation**
- **Controlled phenotyping environment**
- **Mapping onto existing paradigms**
- **Massive increase in information content**

Summary

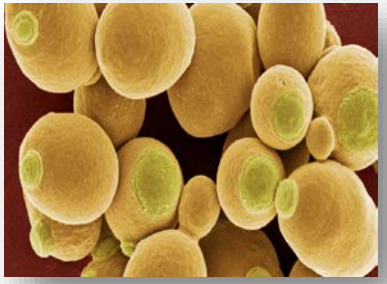
- Genome interpretation requires knowledge of conditioning variable
  - Pretest probabilities
  - Family history
  - Exposures
  - Baseline population data



- Scalable animal modeling is emerging as a partner for clinical genomics
  - Genotype and phenotype anchoring
  - Allow iterative validation of *in silico* models
  - Systems level understanding of disease
  - **Embedding drug discovery in the clinic**

- Phenotypic innovation and therapy align genomic discovery, clinical care redesign and cost
  - **Shared lexicon for translation**
  - **Exploit and extend existing model organism data**
  - **Genomes/phenomes/perturbations and networks**
  - Avoid unaffordable duplication

- Establish a new **minimal clinical dataset** to maximize information content
  - Symptoms
  - **Cellular - universality**
  - **Quantitative/linear/stimulus-response pairs**
  - Complement current clinical care, genomics, eHealth
  - Embedded in clinical platforms with education



Acknowledgements

- **Sunny Kapur, Samantha Beik, Evan Wilson**
- **Andreas Werdich, Gabe Musso**, David Milan, Patrick Ellinor, Albert Kim, Daniela Panakova, Jordan Shin**,** Stephanie Eby, Andrea Giokas, Ian Jones, Khaled Sabeh, Matt Killeen, Jeff Winterfield, Shannon Coy, Eva Plovie, Adrian Low, **Amy Kelly**, Amy Doherty, Ward Capoen, Chris D'Amato, MingSum Lee, Patrick Sips, David White,
- **Randy Peterson**, John Mably, Iain Drummond, **Fritz Roth**

- **Angeliki Asimaki, Jeff Saffitz, Steve Chelko, Dan Judge**

- **Robert Green, Heidi Rehm, Jason Vassy, and The MedSeq Study**
- Scott Weiss, Ed Silverman, Elliot Antman, Raju Kucherlapati, Dick Maas
- Neal Lakdawala, Travis Hinson, Carolyn Ho

- **Krishna Yeshwant, Anthony Phillipakis**
- **Dan Roden, Garret Fitzgerald**
- **Kricket and Jon Seidman, Mark Fishman**

- **NHLBI, NIGMS, NINDS, Leducq, MDA, BHF, Harvard Stem Cell Institute, Burroughs Wellcome Foundation**
- **BRI Seed grant and Watkins Leadership Award**