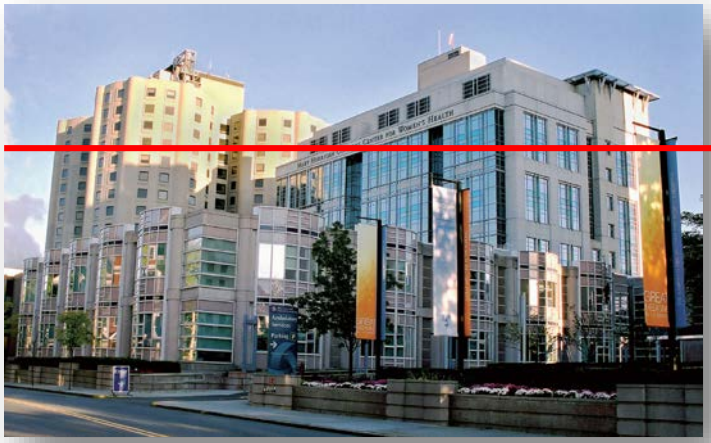


“Vertical Integration” around Clinical Problems

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

A typical clinical problem

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

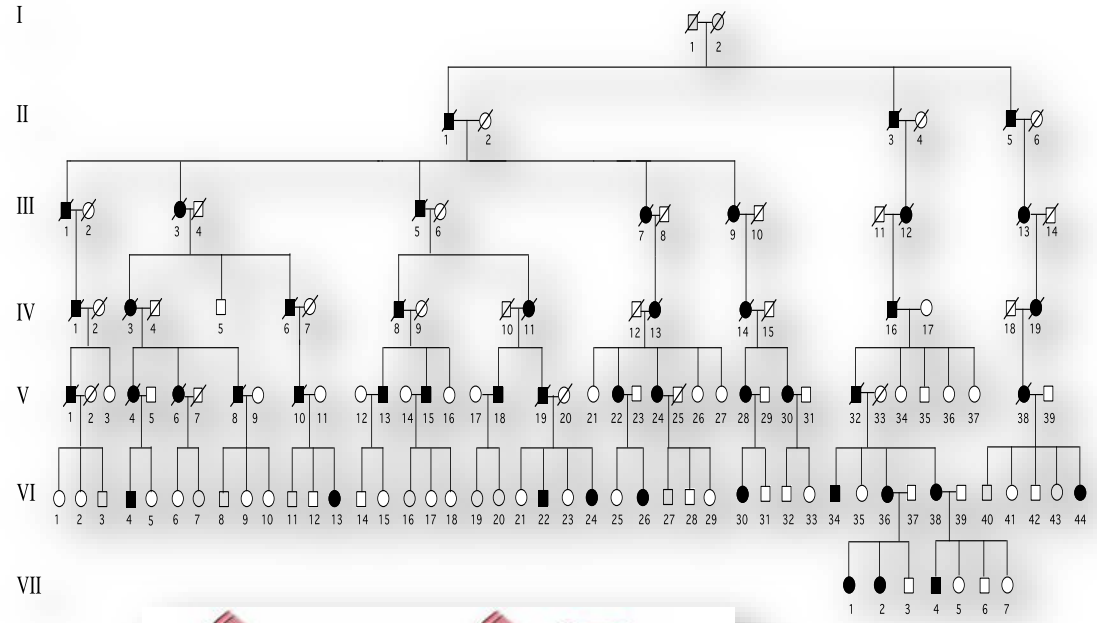
- Multiple phenotypes in a single family
  - Asymptomatic EKG findings
  - CHF
  - Sudden death

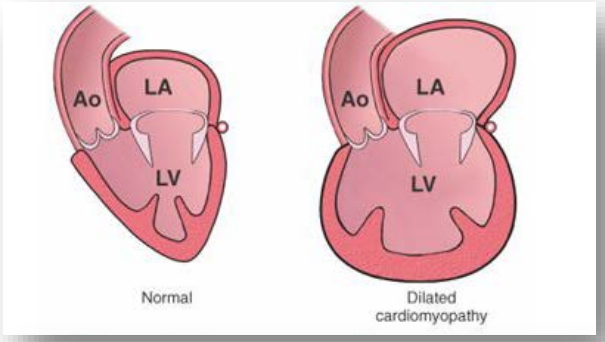
  - 12 different lamin syndromes

- “Modifiers”
  - Genetic
  - Epigenetic
  - Environmental

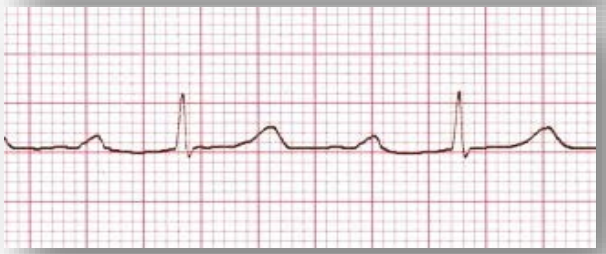
- No empiric support for any model

- **Insufficient information**



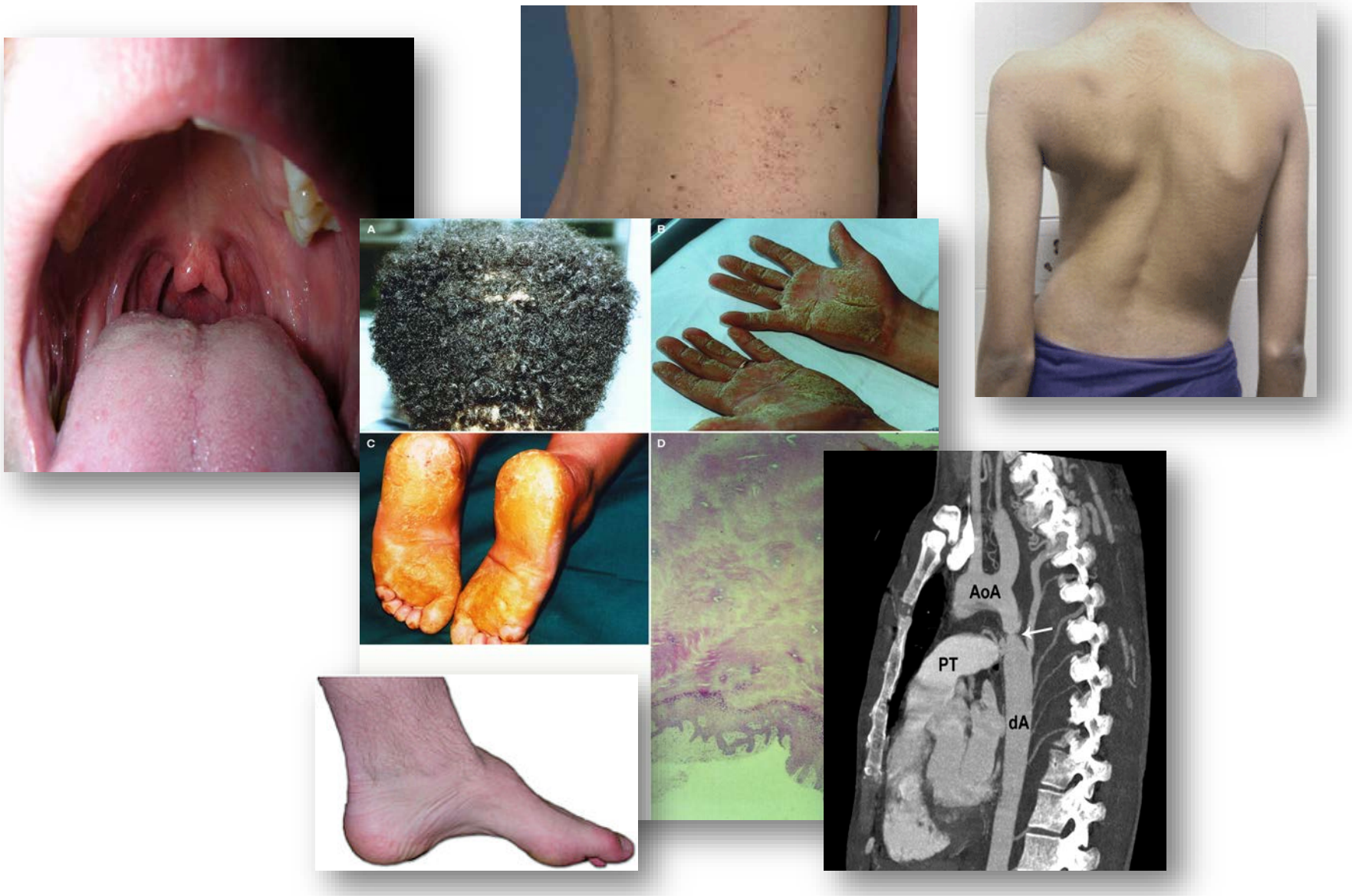


DCM



AV block

Pleiotropic manifestations of DCM genes



Clinical genomics: the other extreme

- Likely pathogenic KCNQ1 variant identified in a primary care patient
- PCP/GC/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.

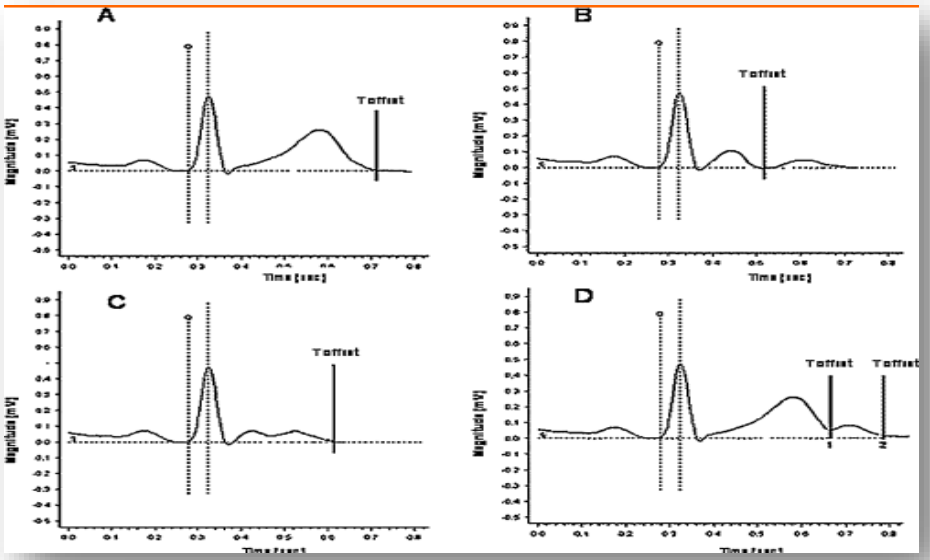
Pathogenicity assessment

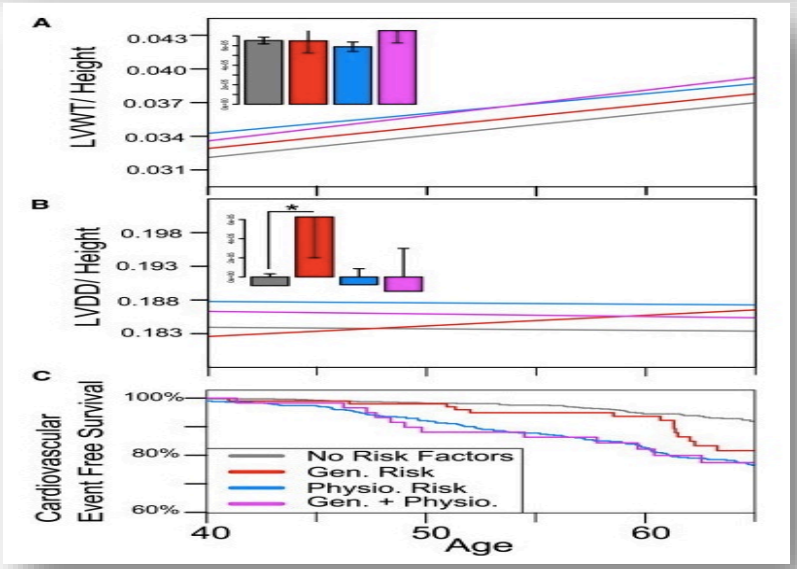
- In vitro assays
- In vivo assays

"... but my QT was normal"

- Segregation
- Penetrance
- Pleiotropy

- Relationship between penetrance and risk obscure





Bick et al AJHG 2012

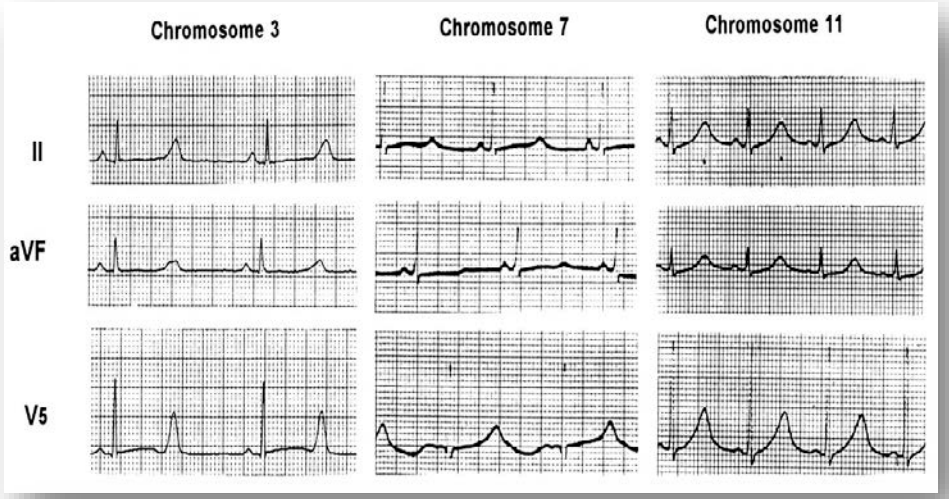
Potential clinical studies

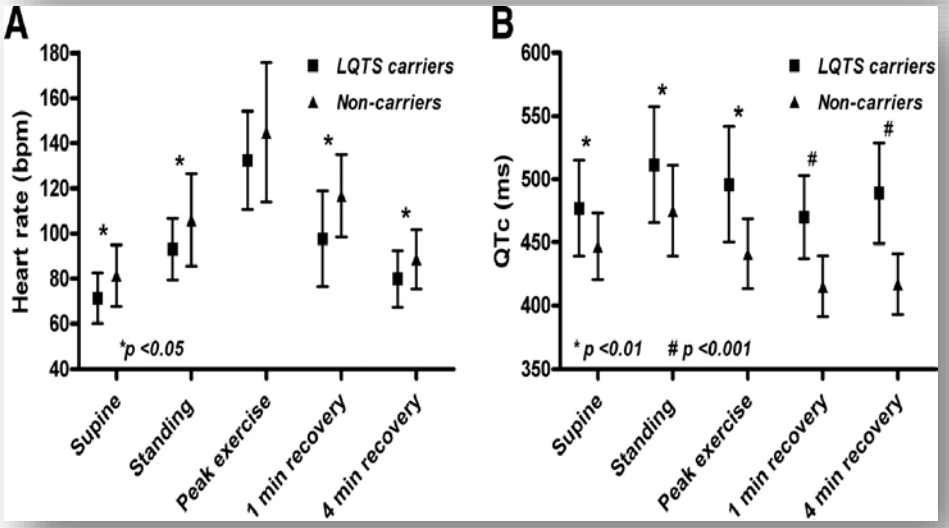
- QT
- QTc
- ECG morphology
- Subclinical /extracardiac phenotypes

- Provoked phenotypes
  - Posture
  - Exercise
  - Recovery

- Signal: noise

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





Family study reveals ‘overlap syndrome’

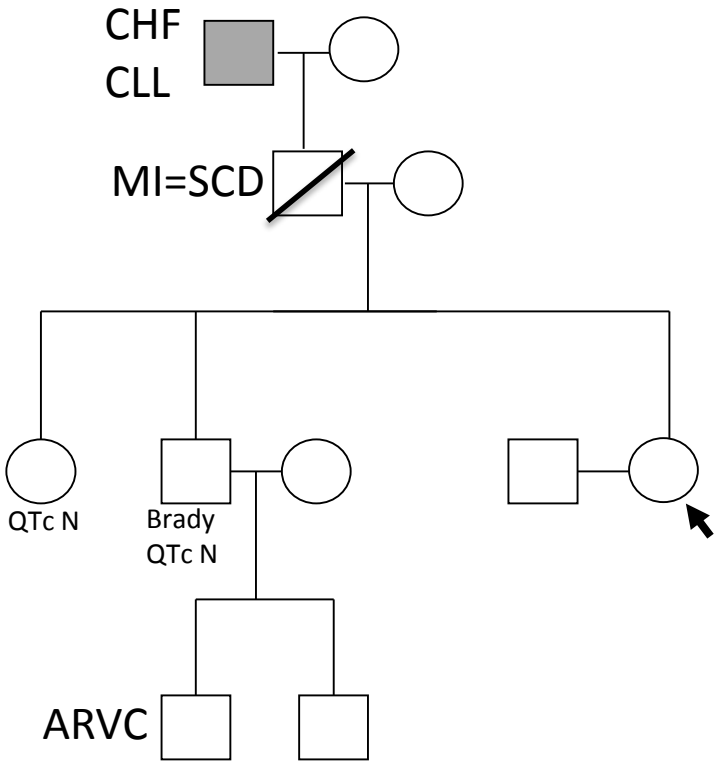
- ‘Condition-specific’ family history
- Physical exam-**S4 and ESM**
- QT-**466ms**
- QTc-**461ms**

- EKG morphology-**Normal**
- Echo-**DUST and MV thickening**
- MRI-**Normal**
- Provoked phenotypes
  - QTc at 4 mins recovery **400ms**

- Clinical overlap syndromes observed
- ? Phenotype expansion
- ? False positive

- Genotypic and phenotypic uncertainty
- Actual risk – unmeasured

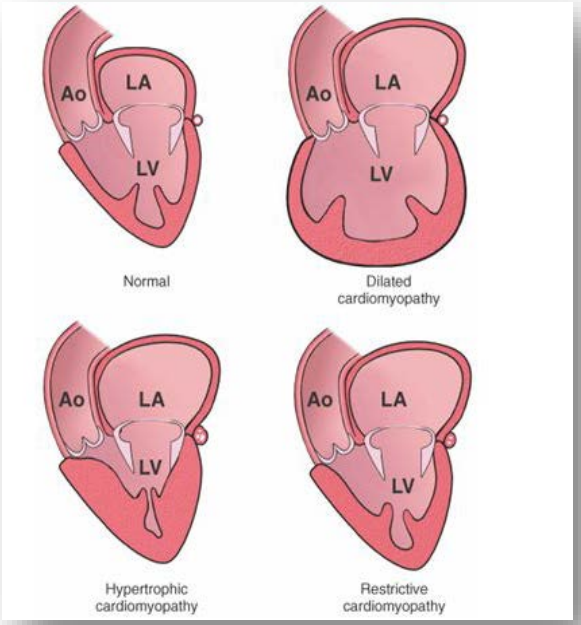
- Cost->$8000

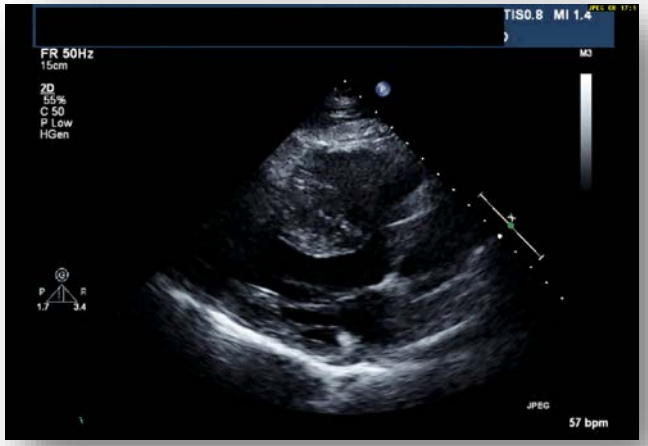


Phenotype is now limiting in multiple arenas

- Clinical care
- Genetics
- Personalized medicine

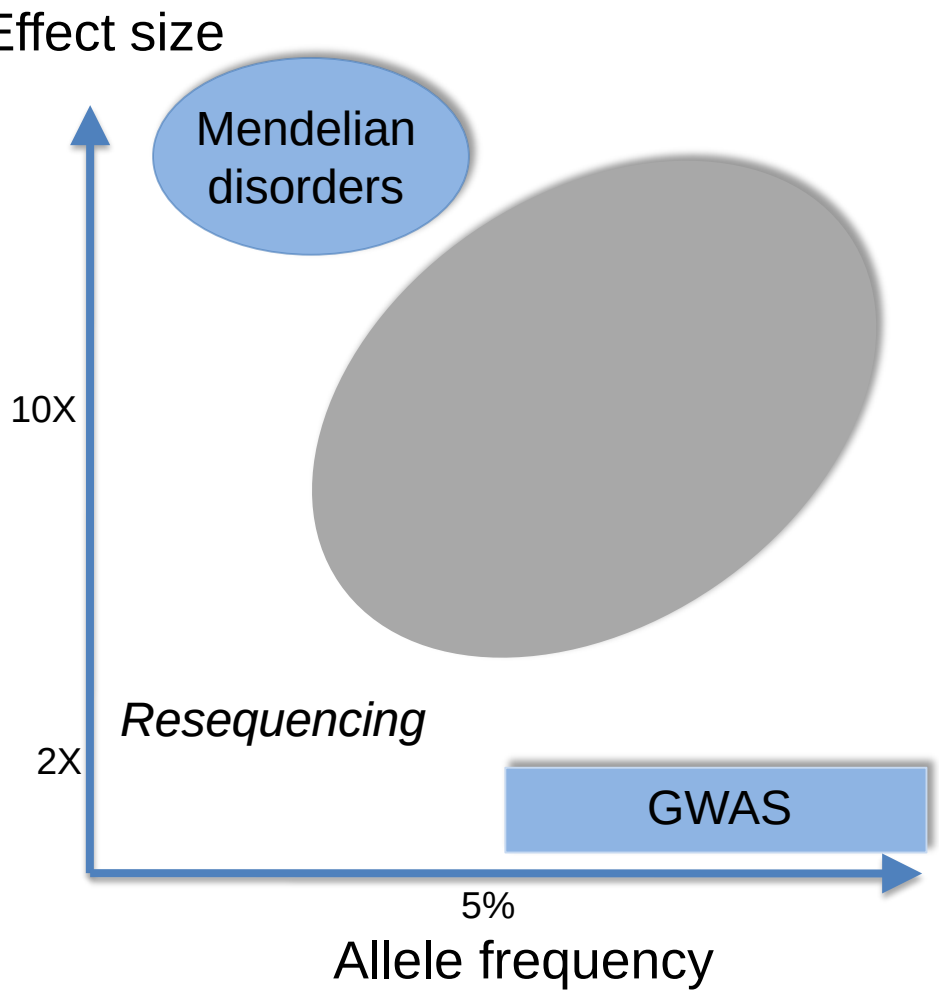
- Fundamental issues
  - Morphology dominates
  - Aggregation
  - Legacy – better at measuring same old phenotypes
  - Semi-subjective at best
  - Cross-sectional
  - Binary
  - Late stage





Where is all the information?

- **Silent alleles**
  - Inaccessible to current study designs
  - Inaccessible to current assays
  - Unmeasured conditioning variables
- Genetic architecture dependent on phenotypic architecture
  - **Phenotypic resolution**
  - Selection pressures
  - Environmental contribution
  - Not assessed for most disease traits
- Limitations of genetic studies to date
  - Focused on extreme phenotypes
  - Few prospective cohorts
  - If familiality detectable how many genes involved?
  - Heterogeneity scales-GWAS inflation



How might model organisms help?

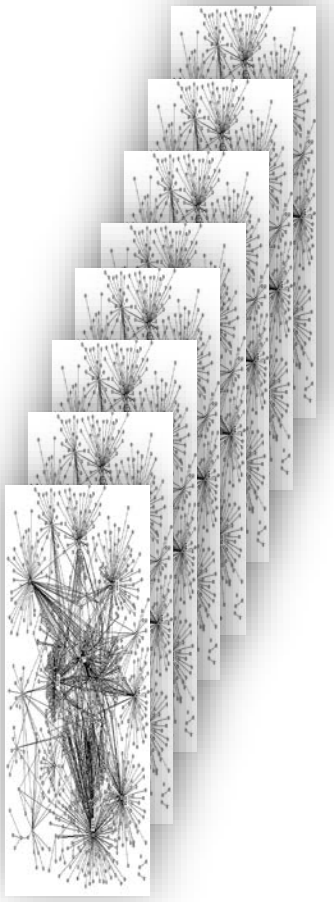
- Saturation screens: to identify all of the genes for a given trait
  - Phenotype anchoring for validation
  - Extreme perturbation
  - Not just F3 recovery but all of the alleles (phenotype)

- Reverse genetics: Manipulate each gene and explore phenotypic ‘universe’
  - KOMP, Zebrafish mutant project, other organisms
  - **Phenotype expansion feasible including functional genomics**

- Test empiric predictions of genotype-phenotype correlation at scale
  - Iterative validation and refinement

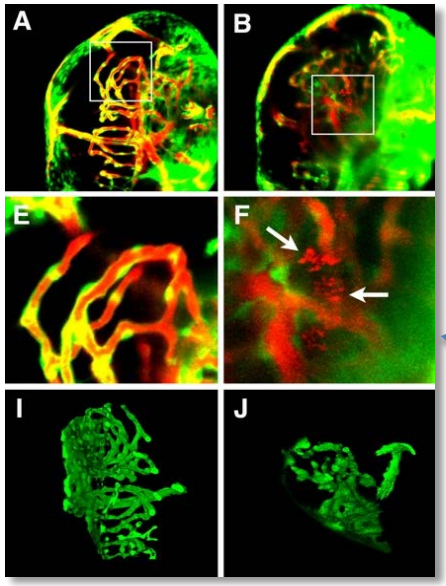
- Environmental modeling: generate **provoked phenotypes**
  - Dynamic responses
  - Few attempts at in vivo disease screens across environmental space
  - Drug discovery as a special case

- **Identify gaps in genetic or phenotypic architecture**

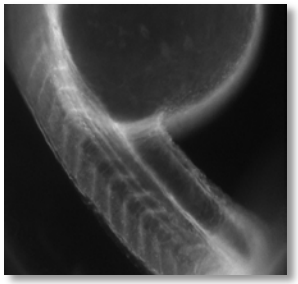


‘Massively’ parallel phenotyping

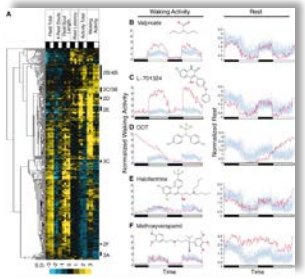
Vascular permeability
(Statin/COX2)



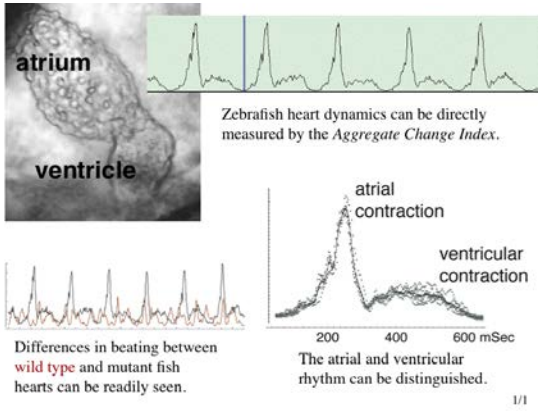
Molecular imaging

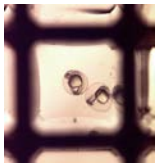


Behavior (Caffeine)

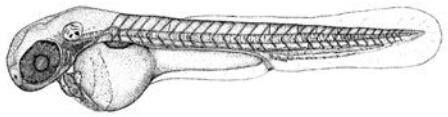


Contractility

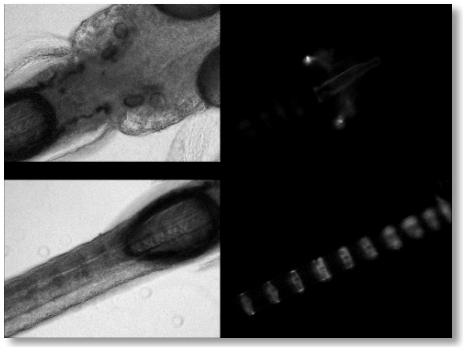


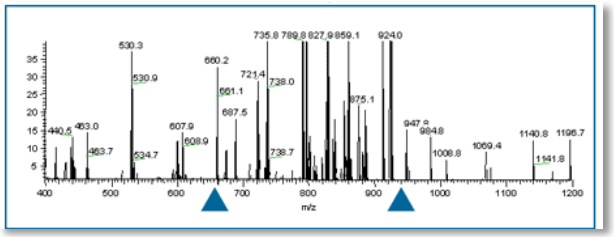


Phenoclustering



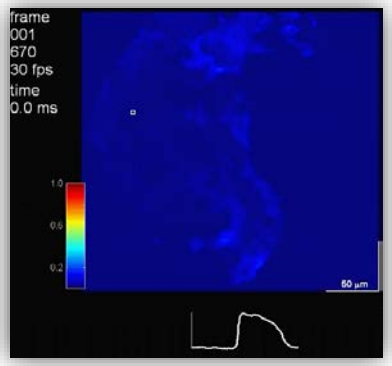
Bone calcification



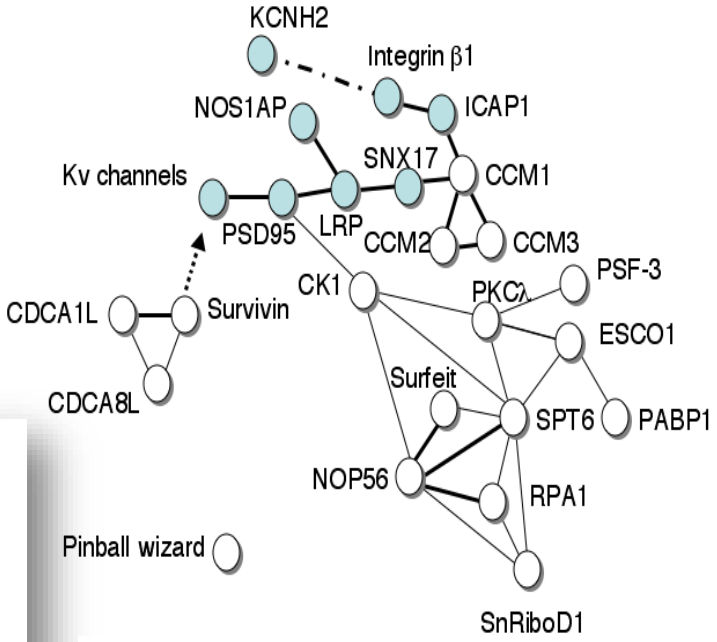


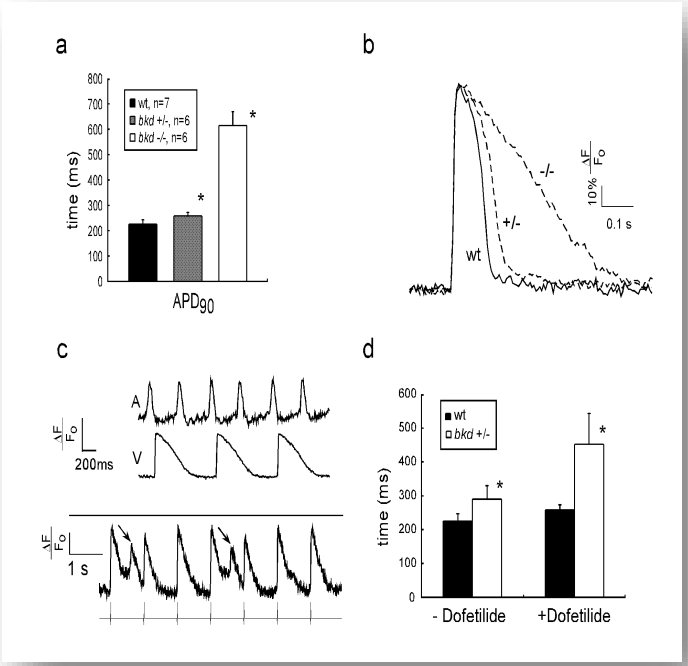
Metabolomics – biomarker discovery

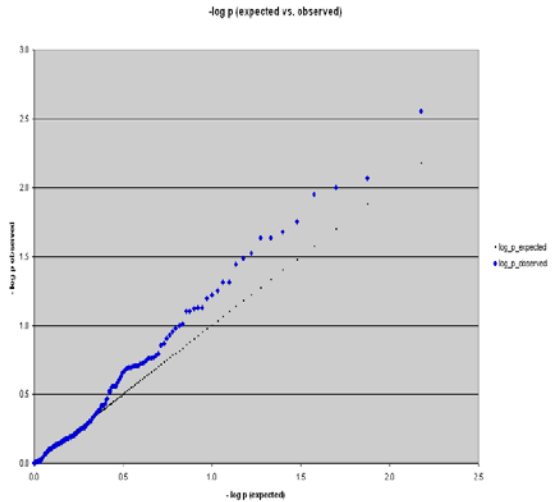
Complex phenotype shelf screen: QT



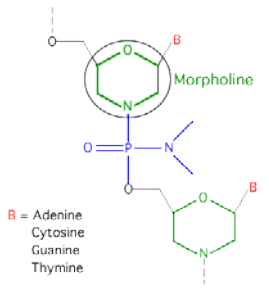
Optical mapping

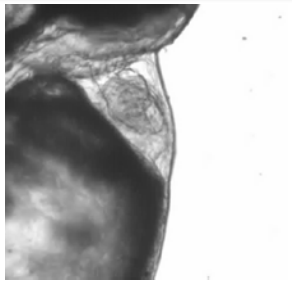


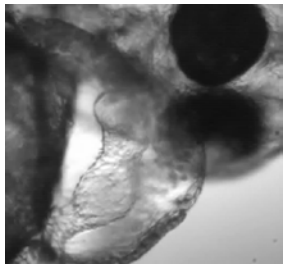


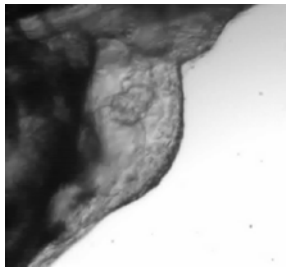


Empiric network validation



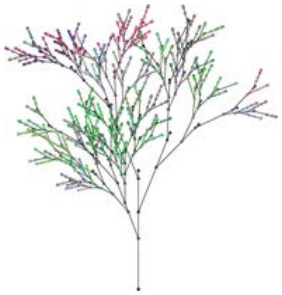


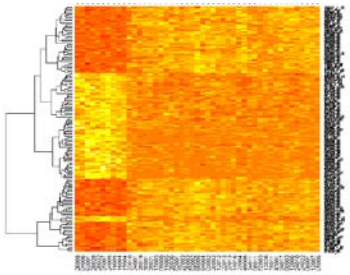




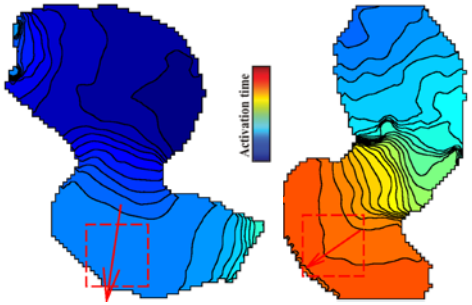
Prediction process begins by collecting published morpholino-phenotype association data

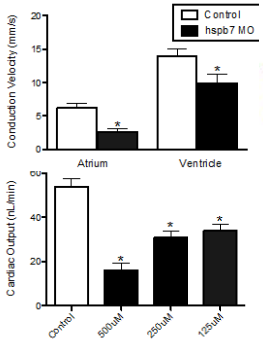
- Gene expression
- Phylogenetic footprint
- Functional annotation
- Protein domain data
- Mapped PPI/GI

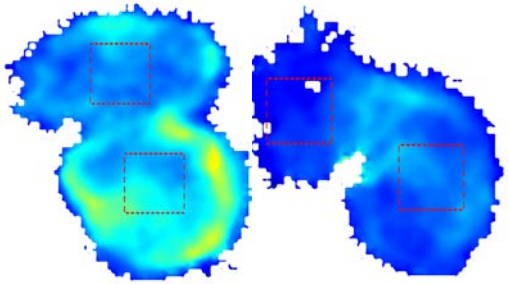




Gene feature information is used in a machine-learning framework to identify patterns and predict novel associations for each phenotype







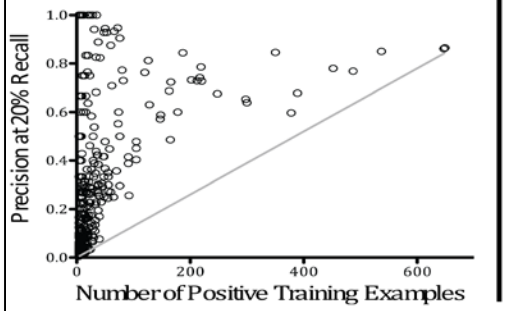
Predictions are tested experimentally, new associations are discovered. The process can be repeated to make more specific predictions.

Iterative prediction and testing

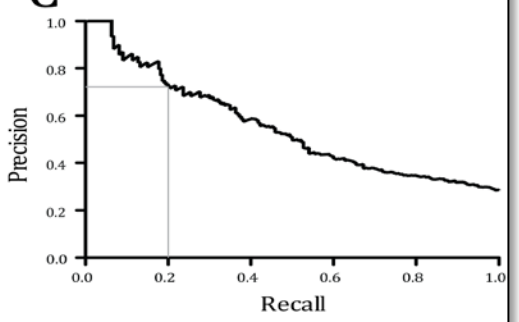
A

	Guilt-by-profiling	Guilt-by-association
Training Data	- 338 anatomical terms - 767 training genes - 34,506 training gene anatomical term associations - 15,106 test genes	- 338 anatomical terms grouped into 9 major anatomical categories - 295,296 total training gene pairs - 114 million test pairs
Features	- Expression (both in specific tissues and in expression sub-units derived based on correlation over 60 experimental datasets) - GO and KEGG annotations - Protein domain data - Orthology in specific species	- Co-expression - Similarity in GO and KEGG annotations - Similarity in domain data - Phylogenetic profile - Mapped protein-protein and genetic interactions from other species
Learning Process	- Separate Random Forest classifier trained for every phenotypic term, used to generate gene term associations	- Random Forest classifier used to construct 9 functional linkage networks, one for each major anatomical term. Edge weights correspond to strength of gene-gene associations
Score Combination	- Gene phenotype associations transferred from functional linkage network probabilistically - Logistic regression used to generate combined scores	

B



C



767 training genes

Training set

338 phenotypes in parallel

Predict across genome

Validate empirically

Retrain

Musso et al. 2014

Initial predictions for CHF: Ca^{2+} handling

85% genes validated for single phenotype

- No false negatives in samples around 50th or 99th centile
- Top 2 GWAS hits for CHF in top 20

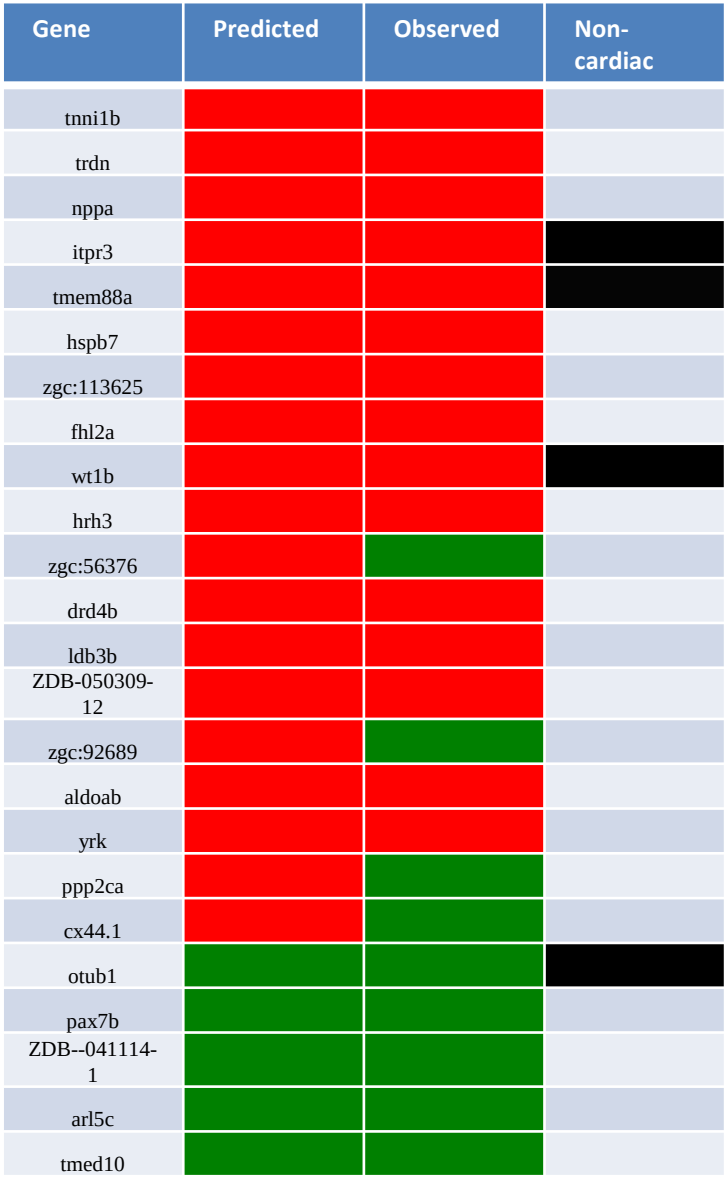
All but 1 remaining genes in top 40 with cardiac phenotypes but only with high resolution assays

Note

- Morpholinos
- Low resolution training phenotypes
- Dominated by anatomy and GO

Iterative refinement underway

- **Addition of emerging genomic data**
- **More specific training sets**



Blocks in translation: AF Genetics

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

- Large Mendelian loci identified

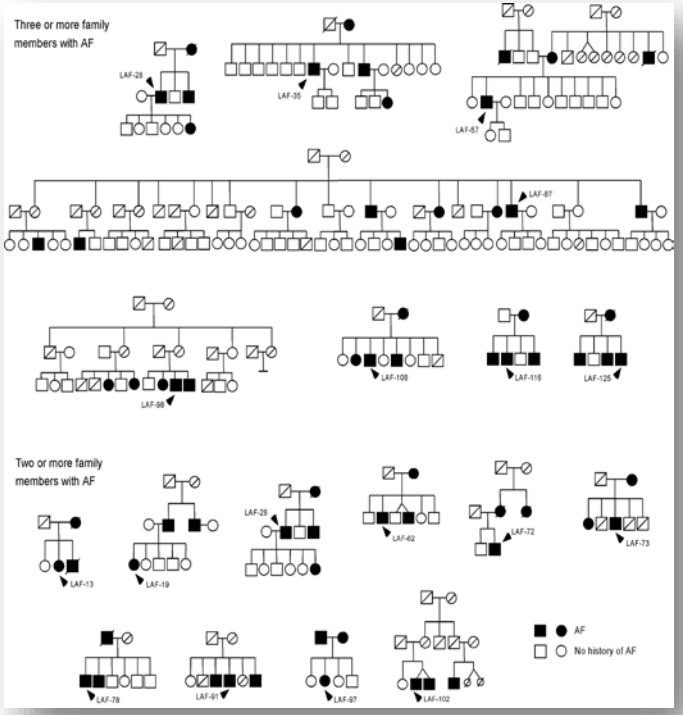
- ~ 10% of heritability explained by GWAS loci

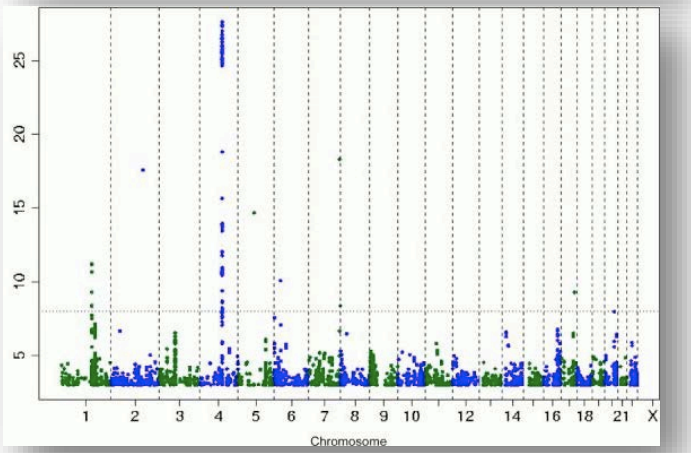
- Missing intermediate effect sizes

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

Need to:

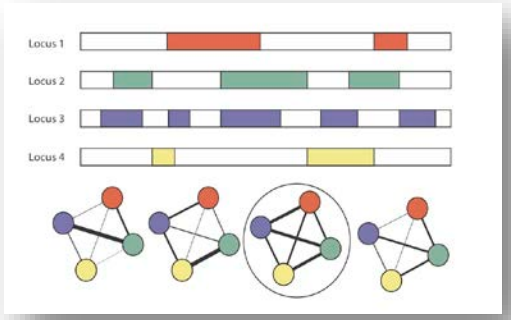
- Explore existing pathways identified in man
- Define better phenotypes
  - Biomarkers
  - New structural or functional assays
  - “AF threshold”

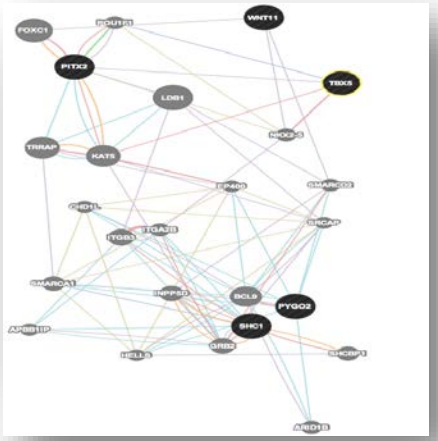


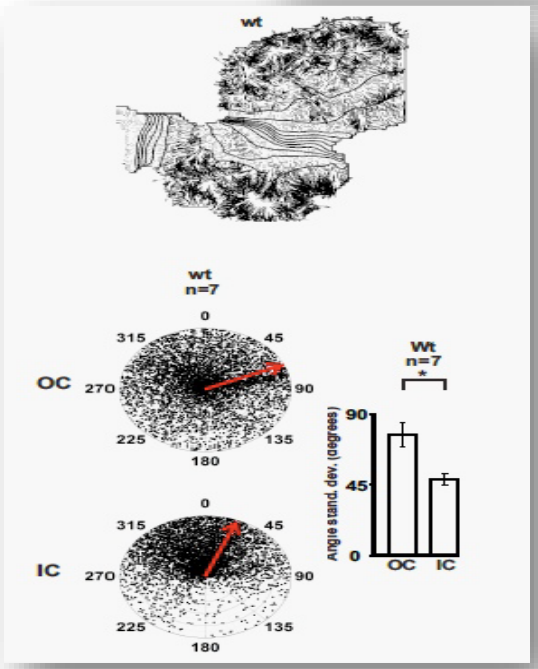


Identify shared networks across loci

- 8 GWAS loci for AFib
- All genes/miRNAs/linc RNAs within 500kb
- Used prior functional networks
- Permutation to maximize functional linkage information
- Network of Wnt pathway genes identified
  - Perturb primary cell circuitry in heart
- Human phenotype rate-limiting

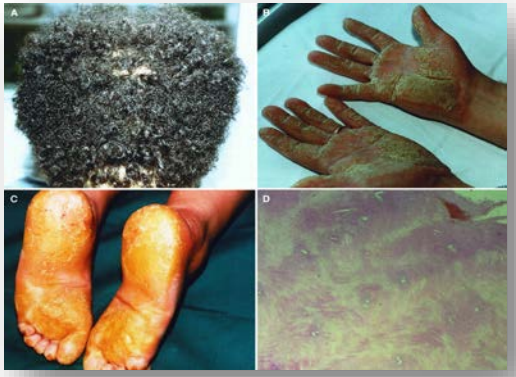




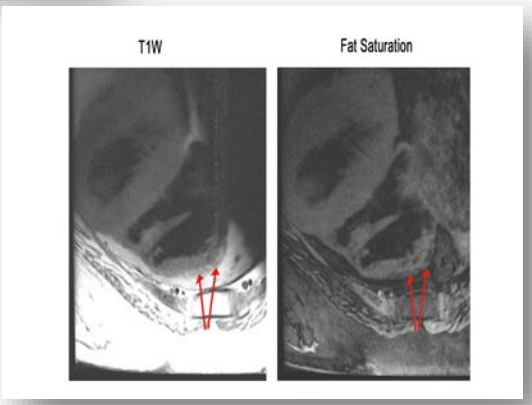


Modeling chronic disease in 5 days: ARVC

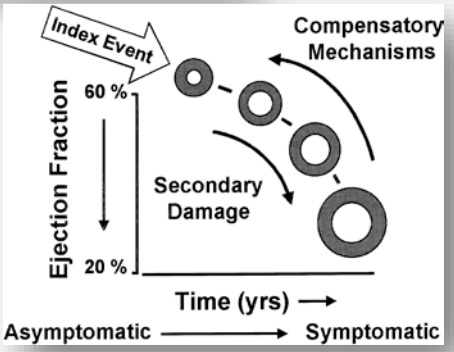
- Poorly penetrant



- Arrhythmia
- Sudden death
- Cutaneous abnormalities

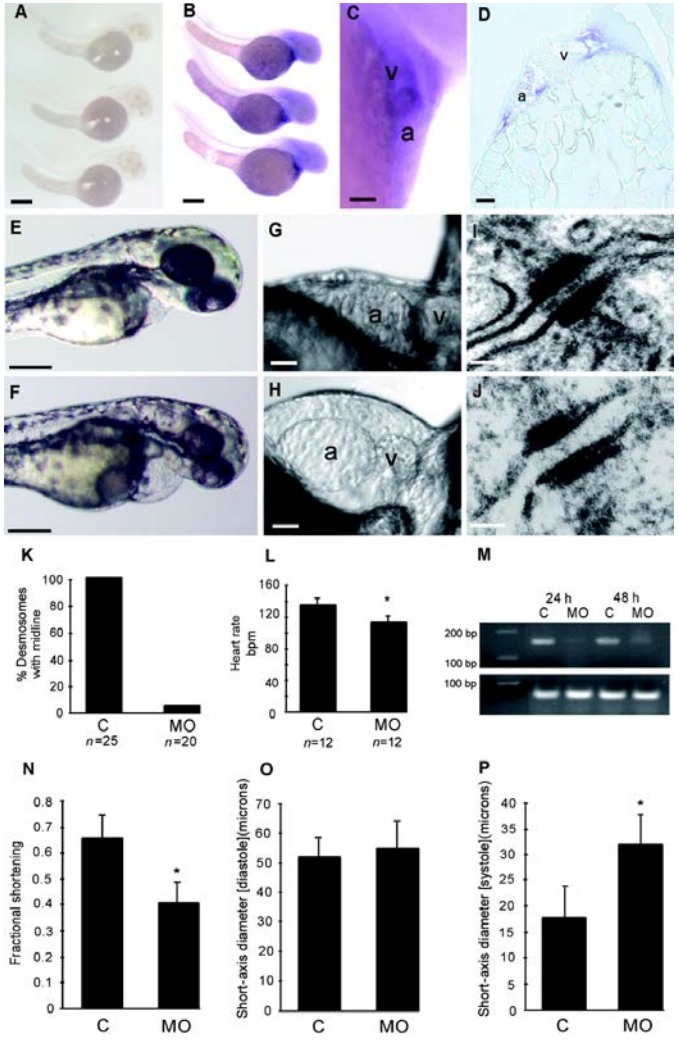


- Contractile abnormalities
  - Congestive heart failure
  - Biomarker abnormalities(nt-BNP)

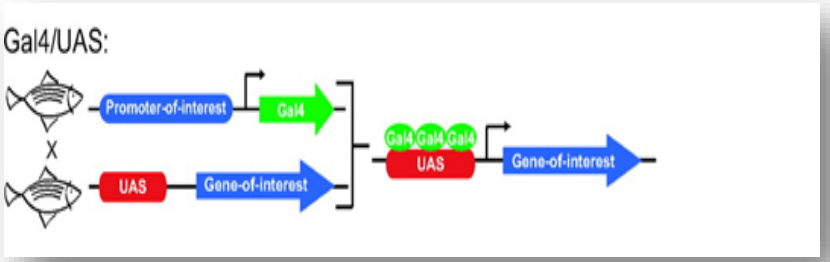


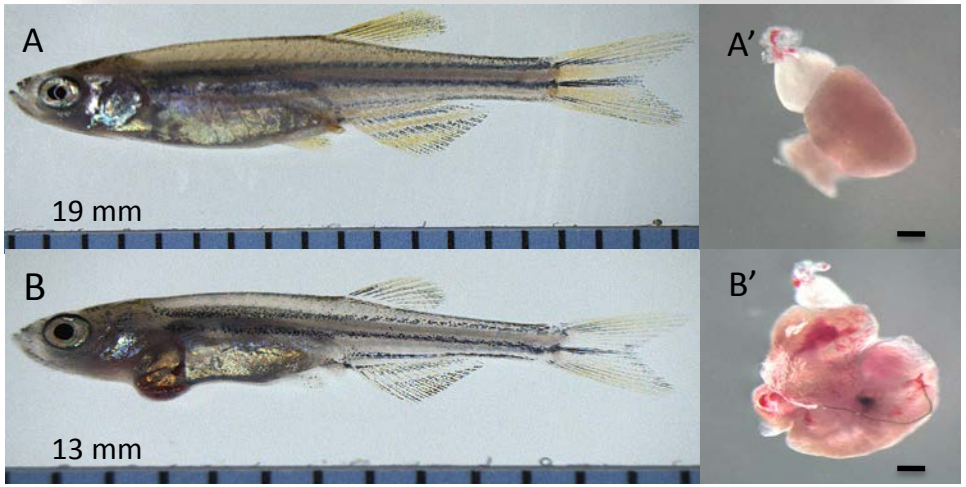
- Desmosomal gene mutations
- Mechanism unclear
  - Wnt signaling perturbed

Genotype anchoring



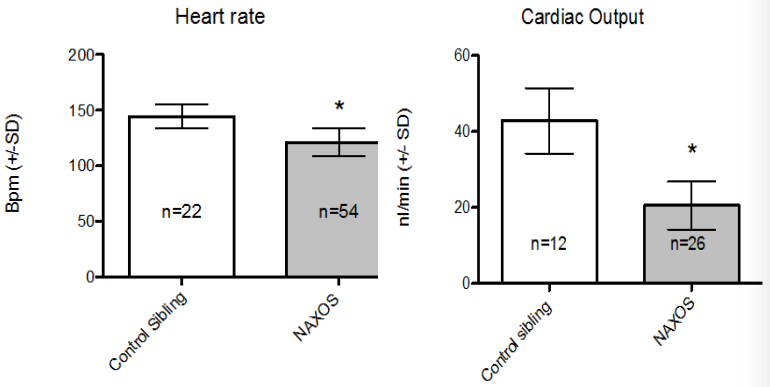
- Multiple disease alleles modeled
  - Morpholino, CRISPR, rescue, transgenesis
  - Recapitulate structure and function
- Modeling human allelic series
- Conditional germline mutant (GAL4::UAS)
  - Allows screening

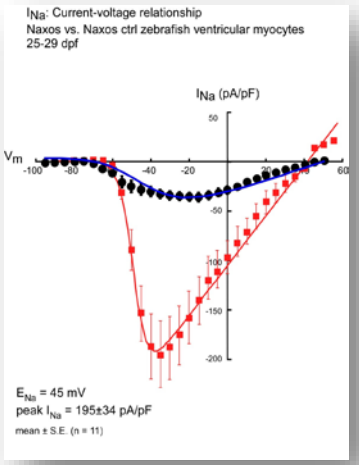


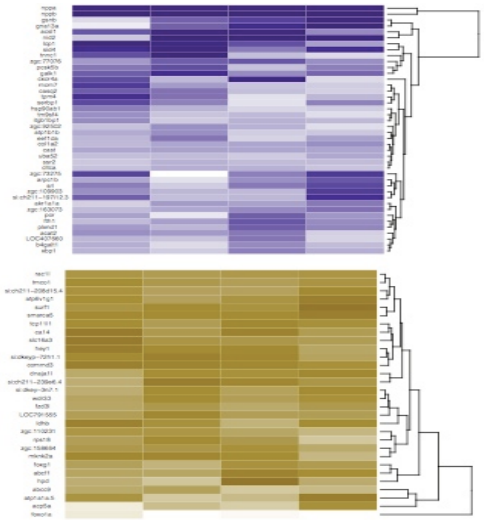


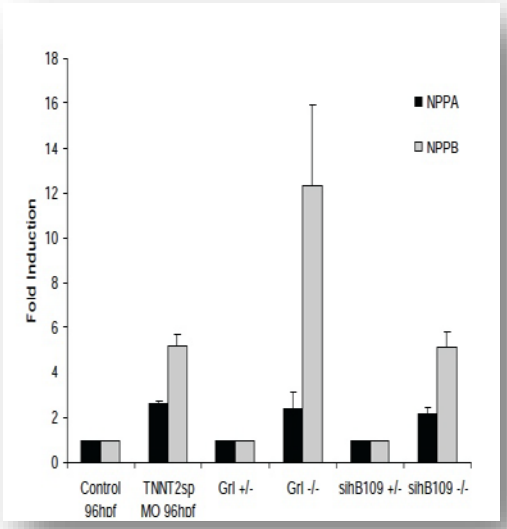
— =1µm

Phenotype anchoring

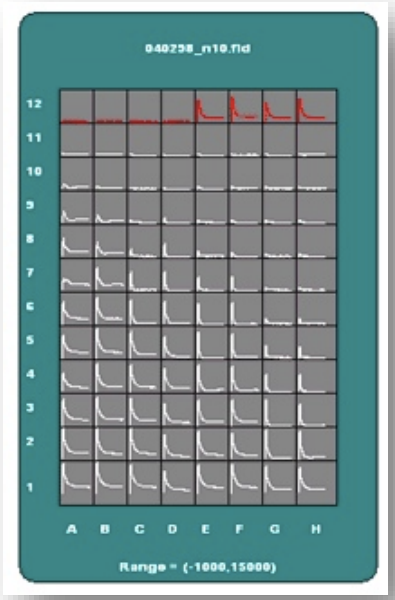






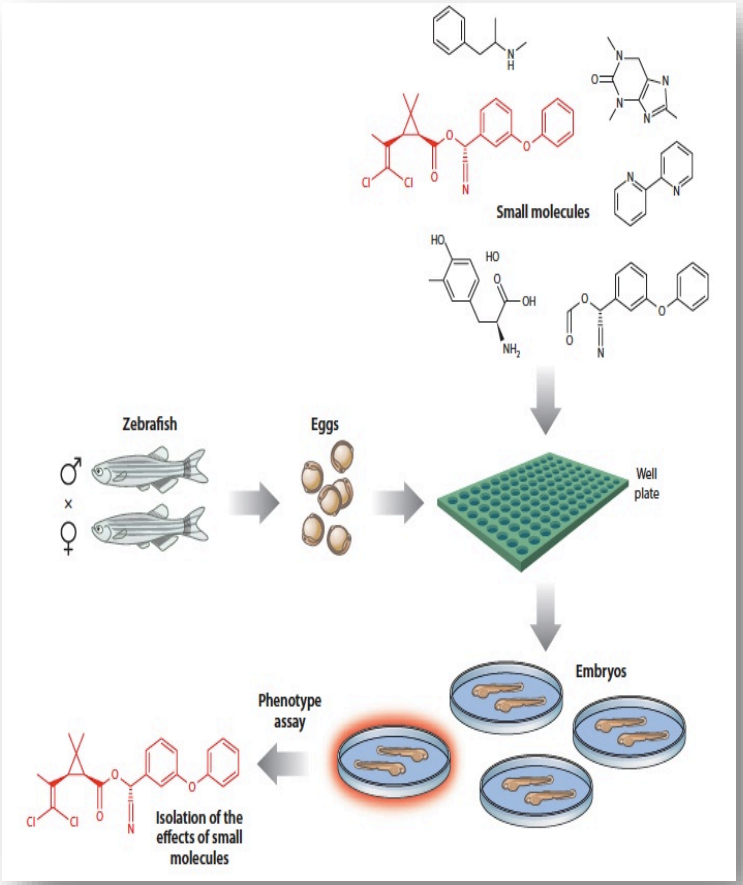


Natriuretic peptide reporter



Luciferase for screen

Screen logic and results



4800 - 70K

Viability
Heart rate and contractility

Nppb::luciferase- 96 well

382

Stringency set empirically

Nppb::luciferase- 96 well

54

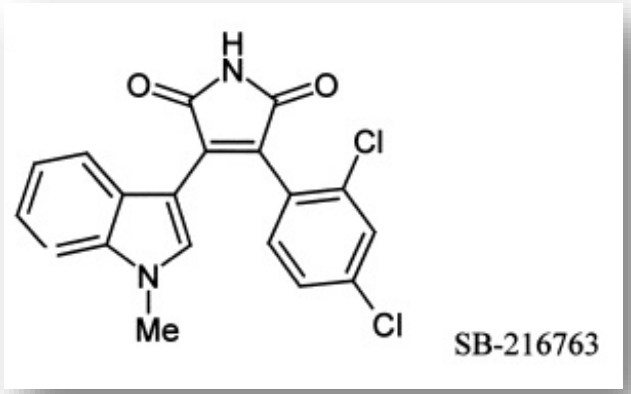
Replicate testing in 50+ larvae

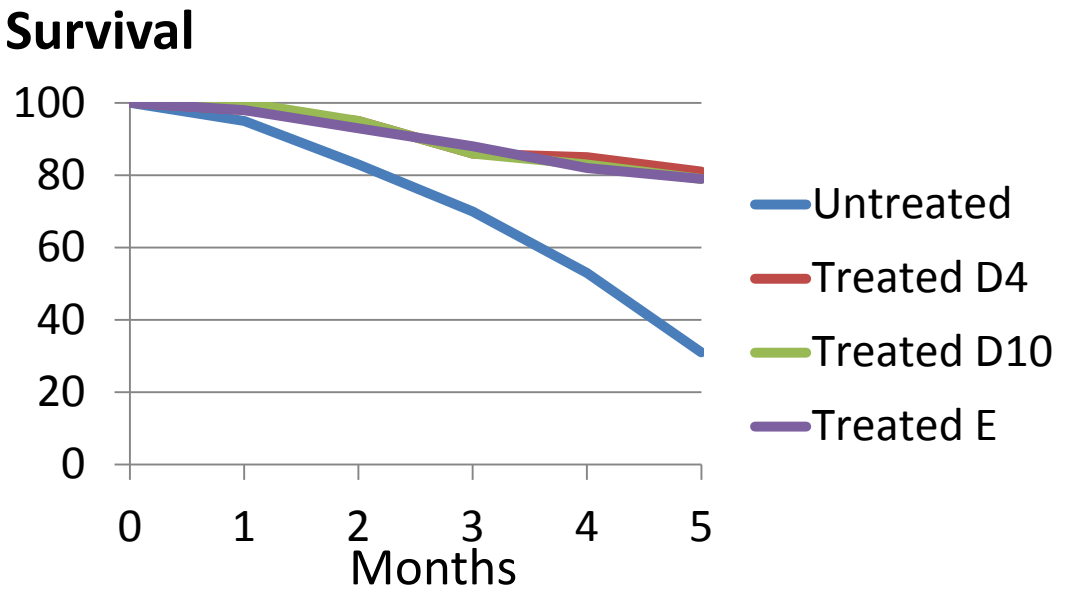
5

Extensive testing in fish, mouse and iPS

1

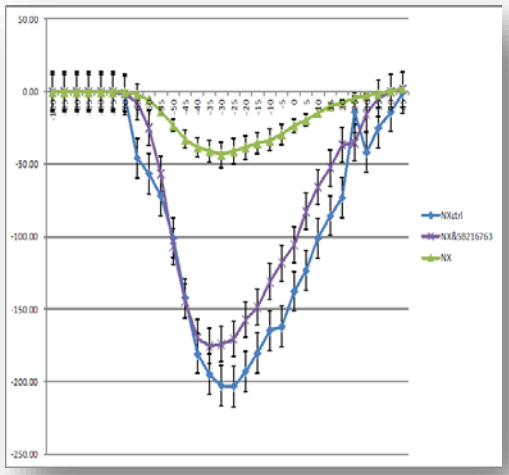
Primary ‘hit’ SB2 rescues ARVC phenotype





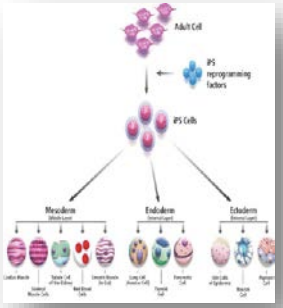
- **Normalizes NPPB expression**
- **Rescues conduction, sodium current and calcium defects**
- **Modest effect on contractility**
- **Improves survival in 3 fish ARVC models, 2 mouse models and rescues iPS defect**
- **Forward trafficking**

Asimaki et al. STM 2014



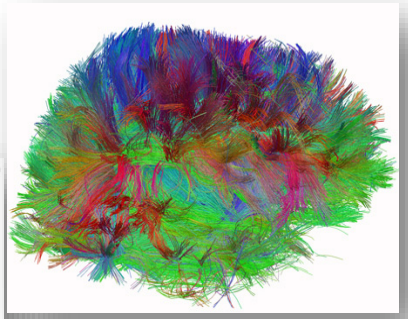
Human iPS (PKG mutant)

The scale of the phenotype ‘gap’



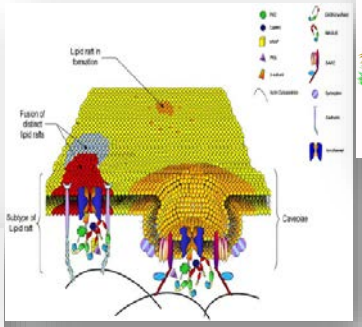
Single cell multipliers

Connectome



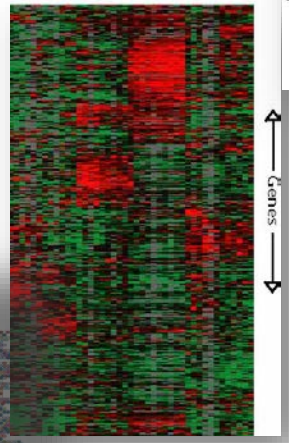
10^{69}

Proteome



10^{20}

Transcriptome

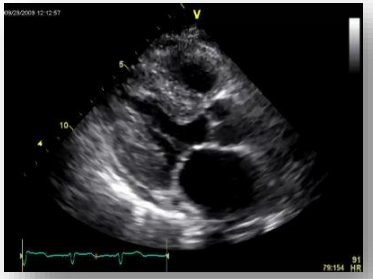


10^{16}

Genome

10^{9}

Clinical Medicine



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

- We have focused on measuring serendipitous endpoints more precisely

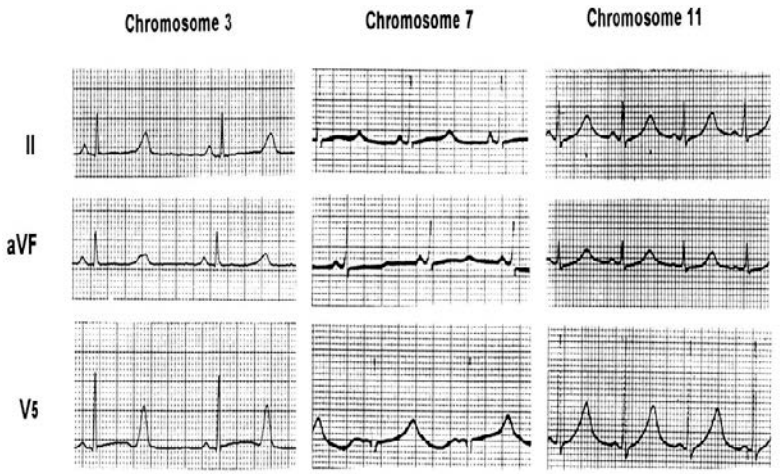
- Deliberate reduction in complexity
- Limited dimensionality
- No clear organizing stimulus

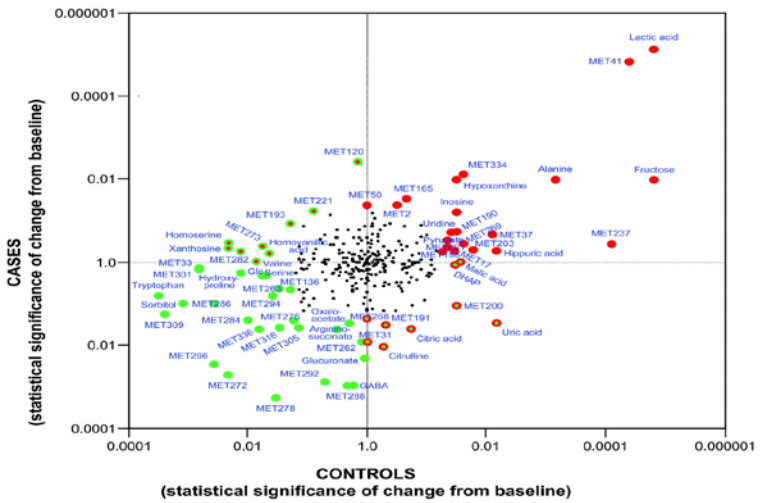
Glucose
Taste

Cholesterol
Visible

Reappraisal of existing data types

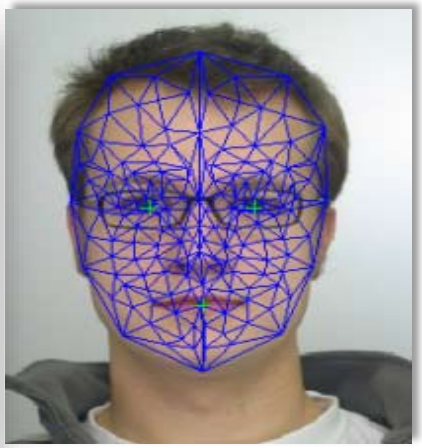
- **Computable phenotypes**
- Collect structured data in eHR
- Reanalysis of existing datasets
  - **Standardized acquisition**
  - New analytic approaches
    - Machine learning defines new EKG subsets
  - Infrastructure
    - Storage
    - Computation
    - Data display
- Functional genomics
  - New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

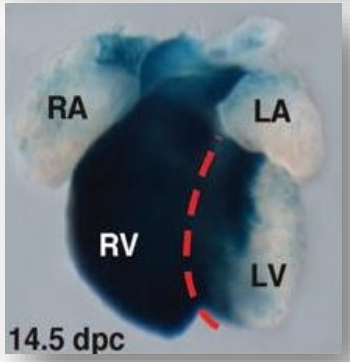


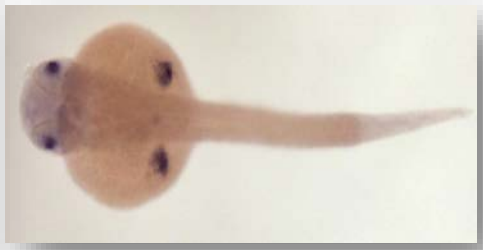


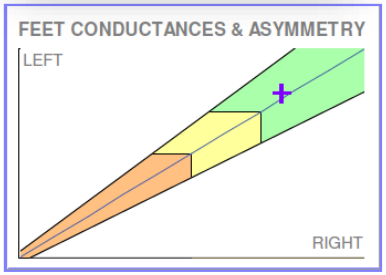
Quantitative and orthogonal phenotyping

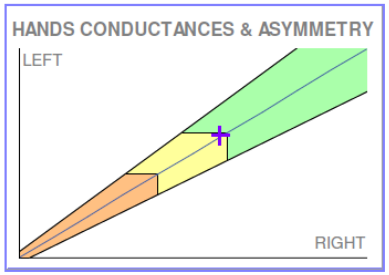
- Facial recognition in the clinic
  - Specific features correlate with disease
  - No systematic studies in adults
  - iPhone based imaging
  - Mapping of fiduciary points
  - **Dominant vocal frequencies**
  - **Hands**
  - **HCM vs Noonans**
  - **Subclinical VCF**
  - **Aortic valve disease subsets**
- Digital physical exam
- Anatomy v physiology





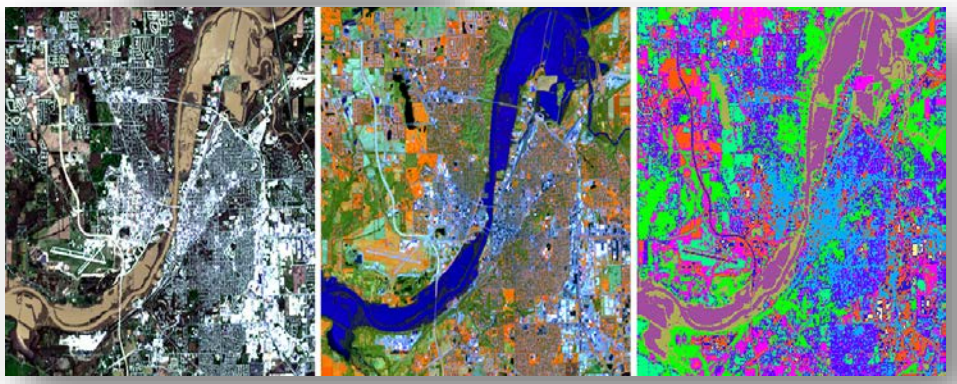




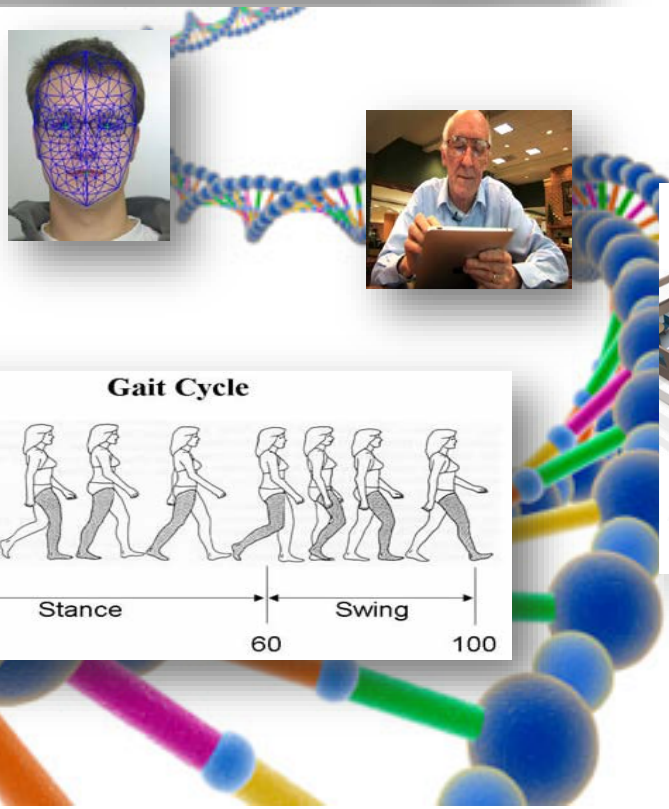


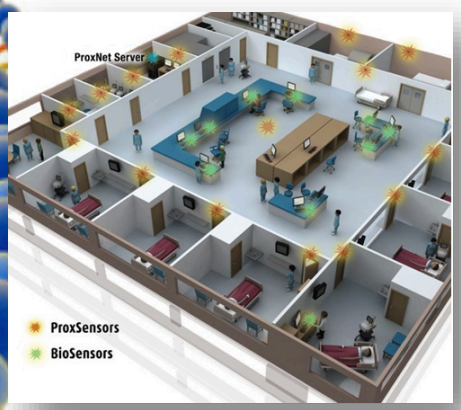
Implementing new technologies-examples

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



Collecting phenotypic narratives in clinic





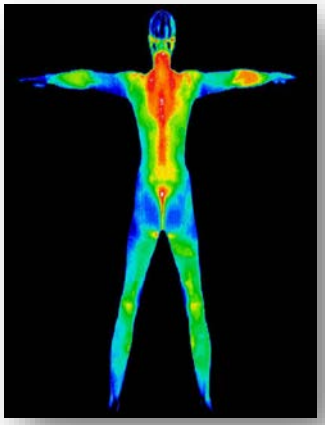
Proxsense

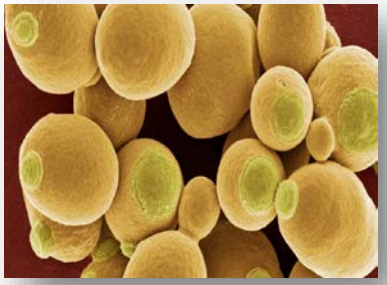
‘Next generation physical exam’

Technology validation
Mapping new onto old
Controlled phenotyping environment
Massive increase in information content

Summary

- Current external pressures demand revolution in multiple elements of the translational cycle
  - Comprehensive approaches to phenotyping to maximize yield from genomics
  - Clinical investment in R&D infrastructure
  - New translational teams-curation, biology, clinical care
  - **Prove clinical utility**
- Model organisms offer scalable *in vivo* genetics/biology/chemical genomics
  - Build fundamental biology around clinical problems
- Phenotypic innovation aligns discovery, clinical care and cost
  - Shared lexicon
  - **Genomes/phenomes/perturbations and networks**
  - Avoid unaffordable duplication
- Establish a new **minimal clinical dataset** for 21st century
  - **Rooted in fundamental biology rather than technology**
  - Complement current clinical care, genomics, eHealth
  - Accelerate translation
  - Portable, affordable…..





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**

- Neal Lakdawala, Travis Hinson, Carolyn Ho
- Bernadette Donnellan, Sioux Gilli, Anne Fuhlbrigge
- Lesley Solomon, Jackie Slavik, Barbara Bierer, Paul Anderson
- Ed Silverman, Scott Weiss, Elliot Antman, Robert Green, Heidi Rehm, Raju Kucherlapati, Dick Maas
- Joel Katz, Bruce Levy, David Cohen, Thomas Michel
- Peter Libby, Pat O'Gara, Marc Pfeffer

- **Krishna Yeshwant, Anthony Phillipakis**
- **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