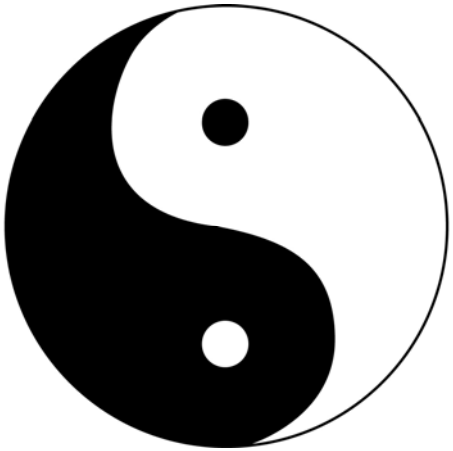
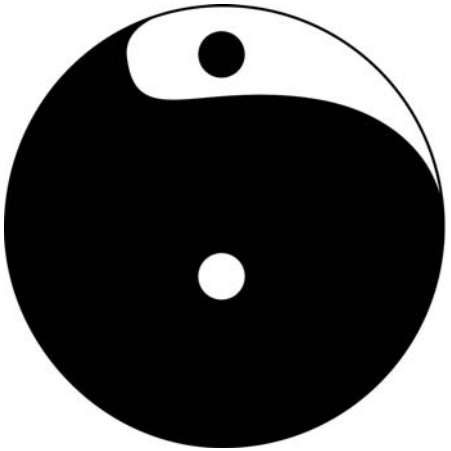
Challenges and Opportunities in Translational Genomics

Leslie Biesecker, M.D.
National Human Genome Research Institute, NIH

Genetic architecture of disease

Rare variants

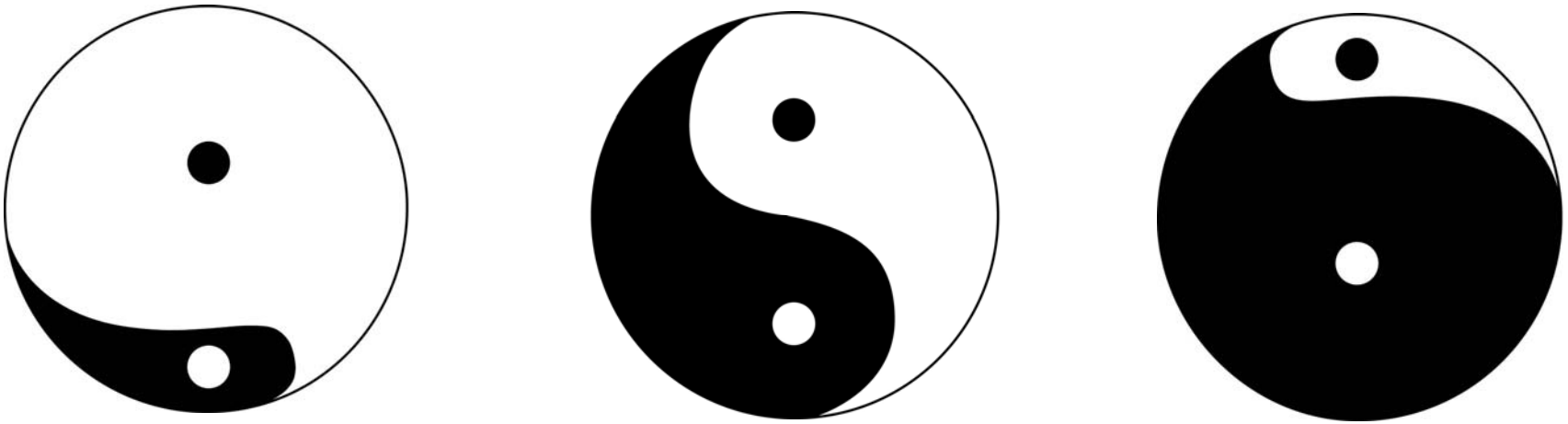




Common variants

Genetic architecture of disease

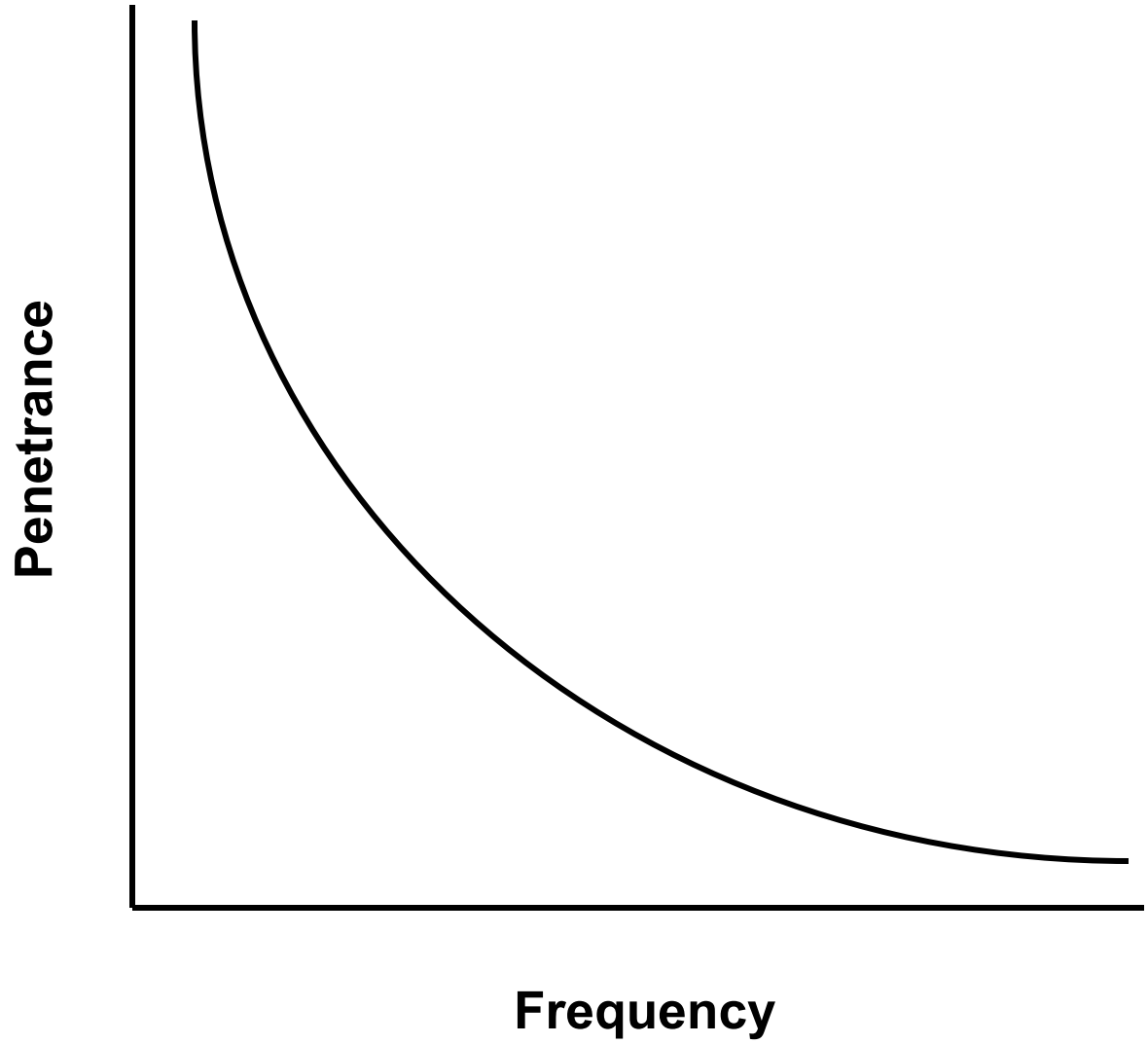
Rare variants



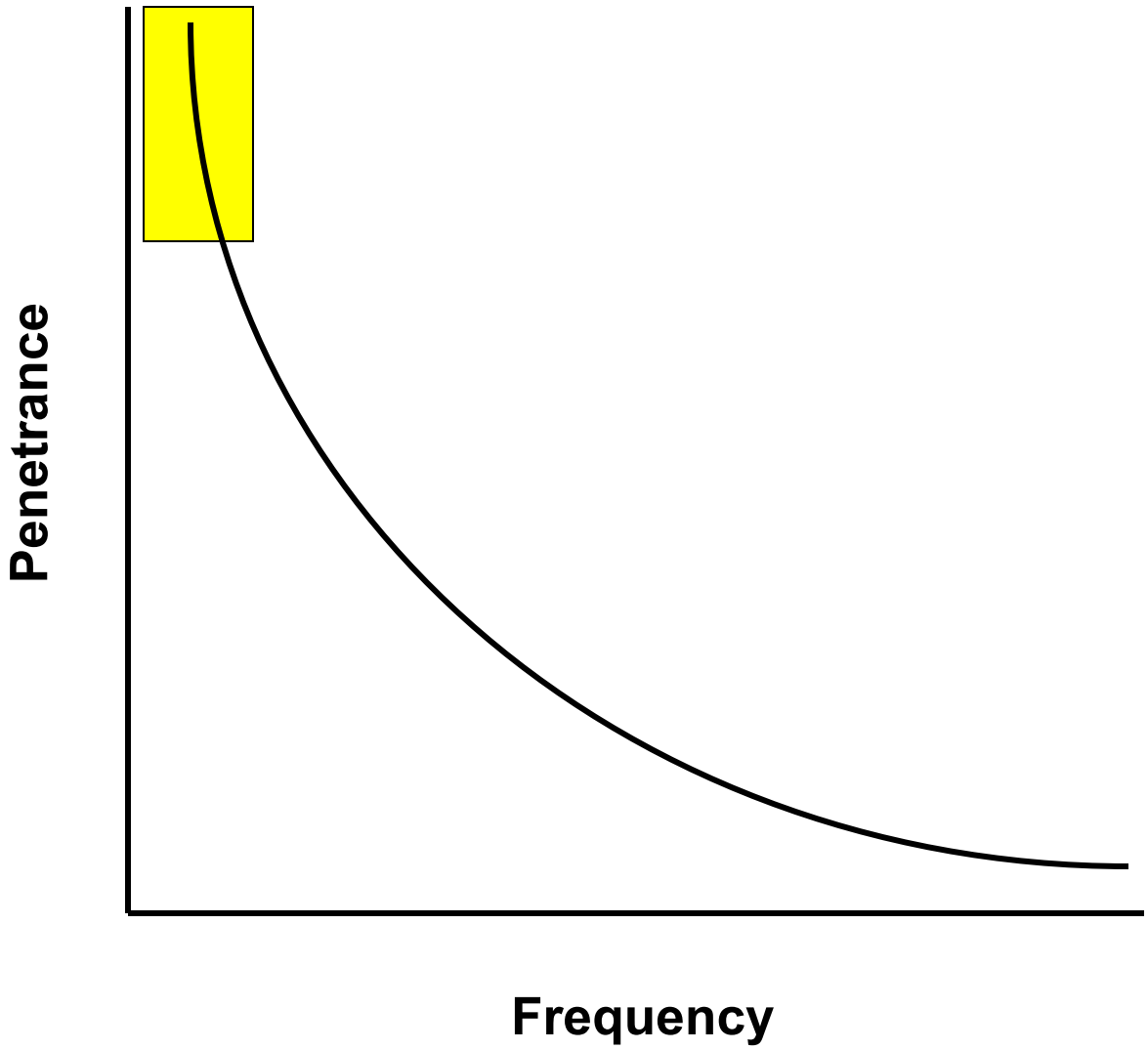
Common variants

We hypothesize that large-scale sequencing will allow a comprehensive assessment of the architecture of the genetic component of disease

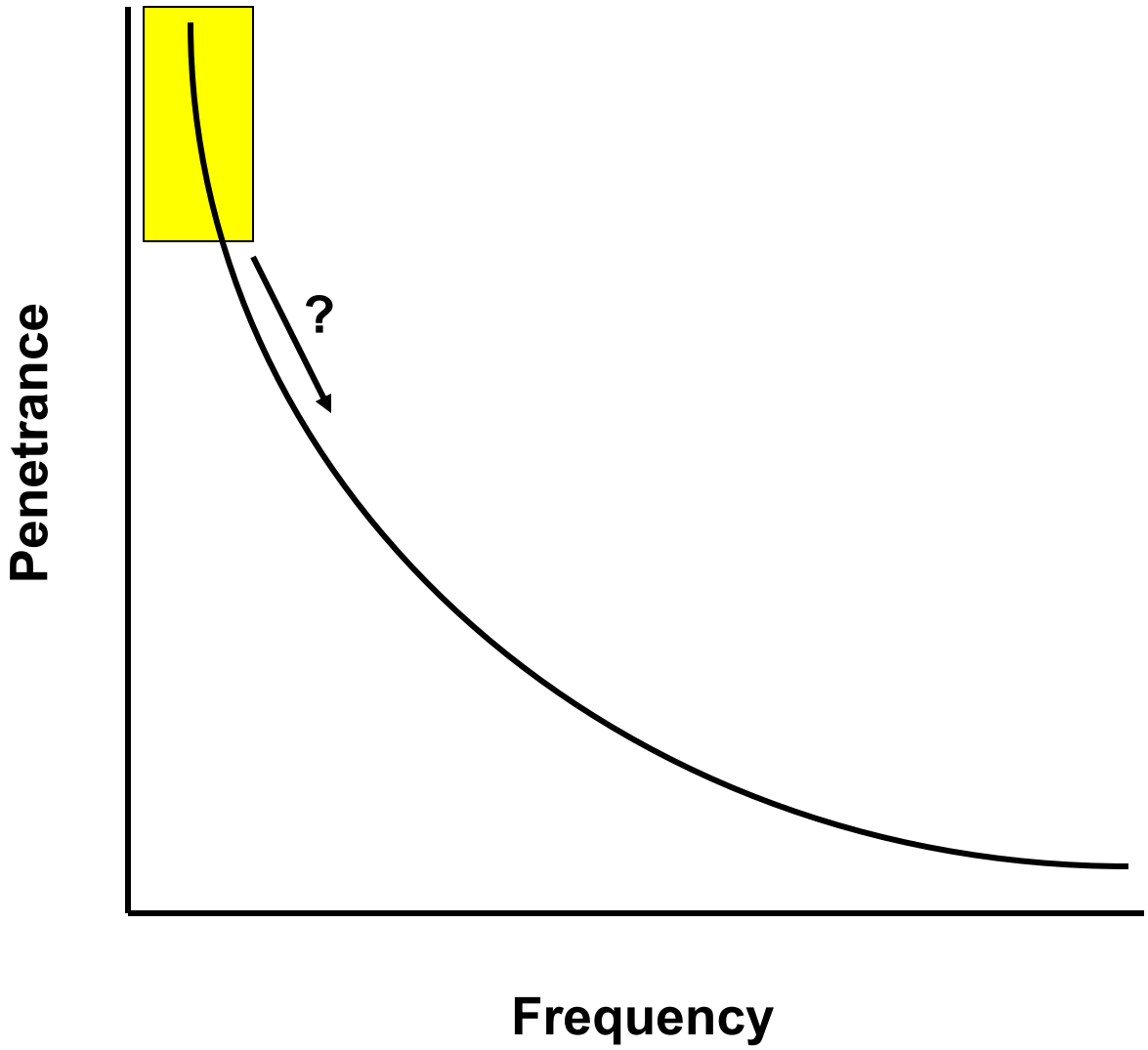
Kinds of mutations



Kinds of mutations



Kinds of mutations



Translational genomics

- Genetic architecture of human disease
- Technologic advances
- New ways to ask and answer biomedical research questions

Translational genomics

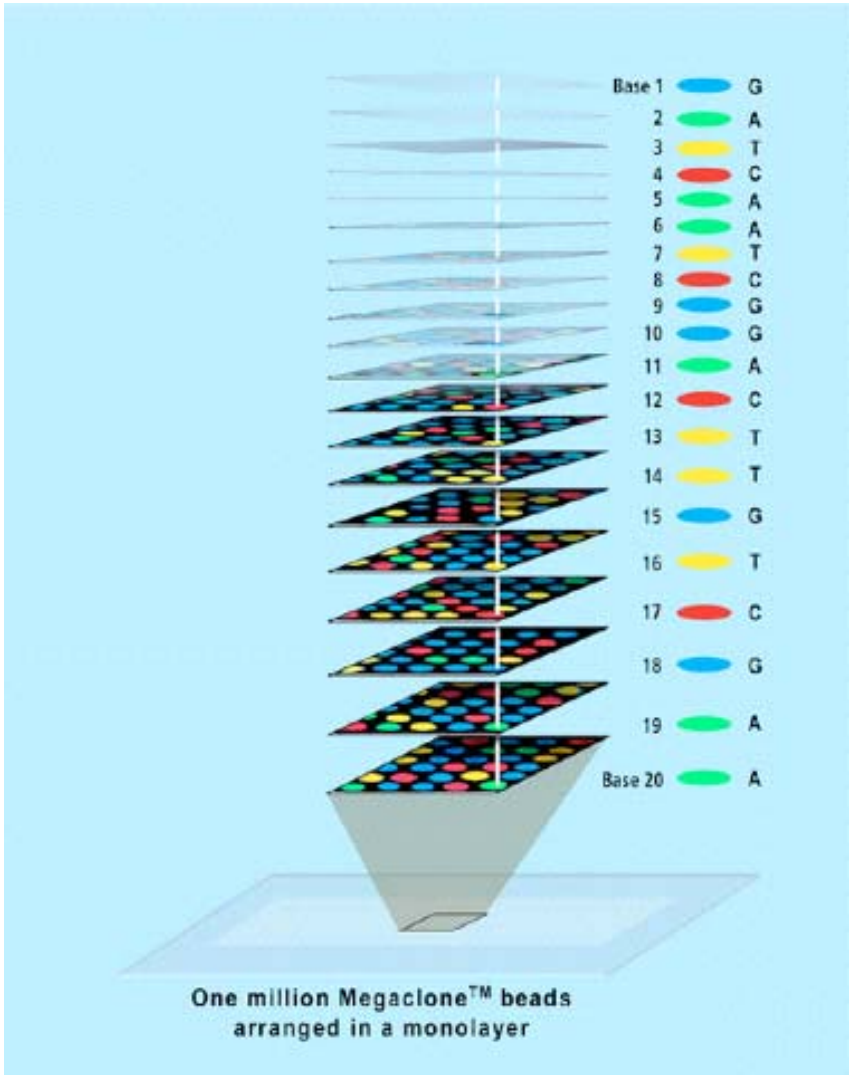
- **Genetic architecture of human disease**
- **Technologic advances**
- **New ways to ask and answer biomedical research questions**
  - ***Pilot whole genome sequence acquisition of individual subjects as a clinical research tool from consent to the return of results***

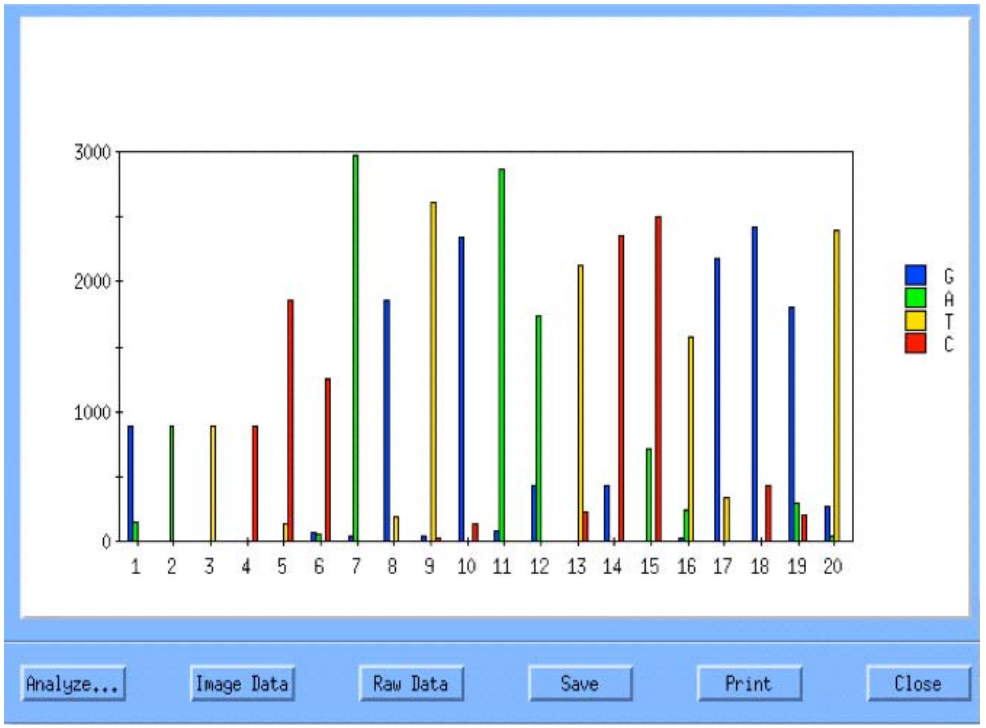
Sequencing technologies

NHGRI supported sequencing centers ~ 2 x 10^{11} bp/yr

Equivalent to 60 - 70 mammalian genomes/year

Solid phase sequencing





Up to 10^9 bp per run
$100,000 genome

ClinSeq: A translational research project in clinical genomics

Medical & Statistical Genetics

NHLBI

ClinSeq

NIH Clinical Center

NIH Intramural Sequencing Center

Specific aims

1. Develop a robust infrastructure for the generation and use of LSMS in a clinical research setting
2. Use LSMS data to develop novel approaches to clinical biomedical research
3. To understand how to interact with subjects re LSMS
4. Genetically dissect a phenotype with complex genetic architecture

Initial approach

- **Phenotype 1,000 subjects**
  - **Framingham risk score**
  - **250 patients into each of four bins**
    - **Low, medium, high risk, known disease**
- ***Initially* sequence 400 candidate genes**
  - **8,000,000 capillary reads**
- **Follow-up studies**
- **Interpret variants and validate *some***
- **Return results**

ClinSeq GeneList V6

.

ABCA1	*CD40*	*HDAC2*	*MTP*	*ROS1*
ABCG1	*CD40LG*	*HMGCR*	*MVK*	*RXRA*
ABCG5	*CDKN1A*	*HMOX1*	*MYLK*	*SAR1B*
ABCG8	*CDKN2A*	*HMOX2*	*NCCT*	*SCARB1*
ACE	*CDKN2B*	*ICAM1*	*NFKB1*	*SCD*
ACTA2	*CELSR2*	*IL18*	*NKCC2*	*SELL*
ADH1C	*CETP*	*IL8*	*NOS1*	*SELP*
AGT	*CFH*	*INSIG2*	*NOS3*	*SELS*
AGTR1	*CIITA*	*ITGA2*	*NOX1*	*SERPINE1*
ADIPOR1	*CRP*	*ITGB1*	*NPC1*	*SIRT1*
ADIPOR2	*CXCR4*	*ITGB2*	*NPC1L1*	*SIRT3*
ADIPOQ	*CYBA*	*ITGB3*	*NPR1*	*SOAT1*
ALOX	*DGAT1*	*ITGB5*	*NRIH2*	*SOAT2*
ALOX5AP	*DGAT2*	*ITGB7*	*NRIH3*	*SOCS3*
ANGPTL3	*DUSP1*	*JAK3*	*NRF1*	*SOD2*
ANGPTL4	*ENPEP*	*KALRN*	*OLR1*	*SOD3*
ANRIL	*ESR1*	*KCNJ8*	*OR13G1*	*SORT1*
APOA1	*F13B*	*KCNMB1*	*P2RY12*	*SREBF1*
APOA2	*F2*	*KIF6*	*PAPPA*	*SREBF2*
APOA5	*F5*	*KL*	*PCSK9*	*STAT1*
APOB	*F7*	*LCAT*	*PDGFB*	*TAS2R50*
APOBEC1	*FABP2*	*LDLR*	*PDGFRB*	*TBXA2R*
APOC1	*FABP4*	*LDLRAP1*	*PER1*	*TCF1*
APOC2	*FAM5C*	*LGALS2*	*PITX2*	*TCF7L2*
APOC3	*FAS*	*LIPC*	*PLA2G4A*	*THBD*
APOC4	*FGB*	*LIPE*	*PLA2G7*	*THBS4*
APOE	*FGG*	*LIPG*	*PLAT*	*TIMP1*
APOM	*FLI1*	*LPA*	*PLTP*	*TLR4*
AR	*FOS*	*LPL*	*PON1*	*TLR8*
ARF	*FTO*	*LRP6*	*PPARA*	*TNFRSF1A*
ARG1	*GAL*	*LRP8*	*PPARD*	*TNFSF4*
ATF4	*GALNT2*	*LTA*	*PPARG*	*TRIB1*
BDKRB2	*GAS6*	*LTA4H*	*PRDX2*	*TRIB3*
BMI1	*GATA2*	*LTC4S*	*PRDX3*	*UCP2*
BSDL	*GCKR*	*MBL2*	*PRKG1*	*UCP3*
C11ORF2	*GCLC*	*MEF2A*	*PSMA6*	*USF1*
CALM1	*GCLM*	*MLXIPL*	*PSRC1*	*VAMP8*
CAPG	*GDF5*	*MMAB*	*PTGIS*	*VEGF*
CAV1	*GJA4*	*MMP3*	*PTGS1*	*VLDLR*
CCL2	*GP1BA*	*MMP9*	*PTGS2*	*VNN1*
CCR2	*GP1BB GP5*	*MOGAT1*	*RAP1B*	
CCR5	*GP6*	*MOGAT2*	*RBKS*	
CD36	*GPX1*	*MPO*	*ROMK*	

Progress

- Enrollment began January, 2007
- Nearly 350 patients enrolled mid May
- >900,000 sequence reads to date
- Results already returned

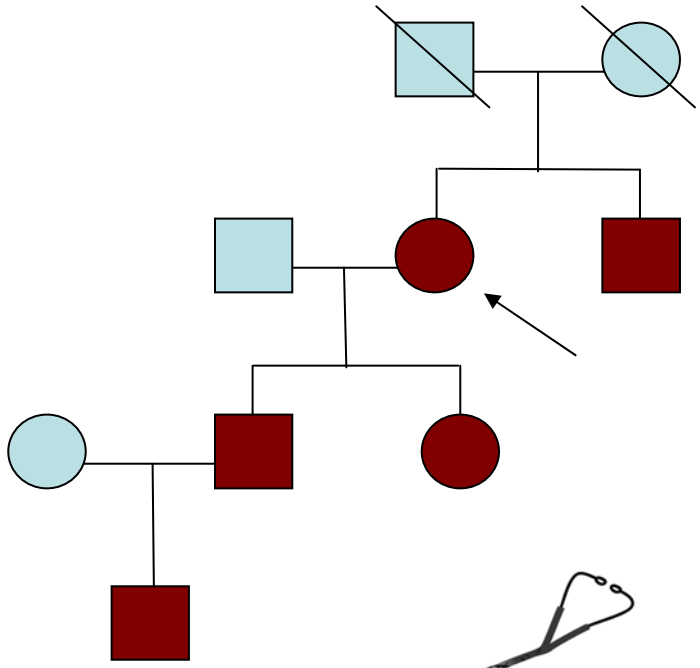
- **N=75 ClinSeq**
  - A50S
  - A391T
  - T726I
  - A391T
  - A391T
  - A391T
  - P685L
  - Y188X
  - A391T
  - V827I
  - A50S
  - T726I
  - R744Q:A391T
  - A391T
- **N=5 HapMap**
  - A391T

Initial ClinSeq LDLR variants

Medical history

- 65 yo female
- High cholesterol diagnosed at 25 years
- RX: atorvastatin, ezetimibe, hctz, lisinopril, niacin
- Coro Ca^{++} 1,726
- Chol 172, Trig 50, HDL 75, LDLd 88

- Family members diagnosed & treatment started



ClinSeq

What we will accomplish

- Develop molecular, bioinformatic & medical approaches for clinical LSMS
- Comprehensively dissect the genetic architecture of a phenotype
- Pilot approaches to personalized health care
- Ascertain the subject's views of LSMS

What we will accomplish

- Develop molecular, bioinformatic & medical approaches for clinical LSMS
- Comprehensively dissect the genetic architecture of a phenotype
- Pilot approaches to personalized health care
- Ascertain the subject's views of LSMS
- Proceed to exome or genome sequencing - fully consented!

Public resource

- Clinical data deposited in dbGAP
- Sequences in trace archive
- Lymphoblast cell lines available
- Plasma/serum available (limited)

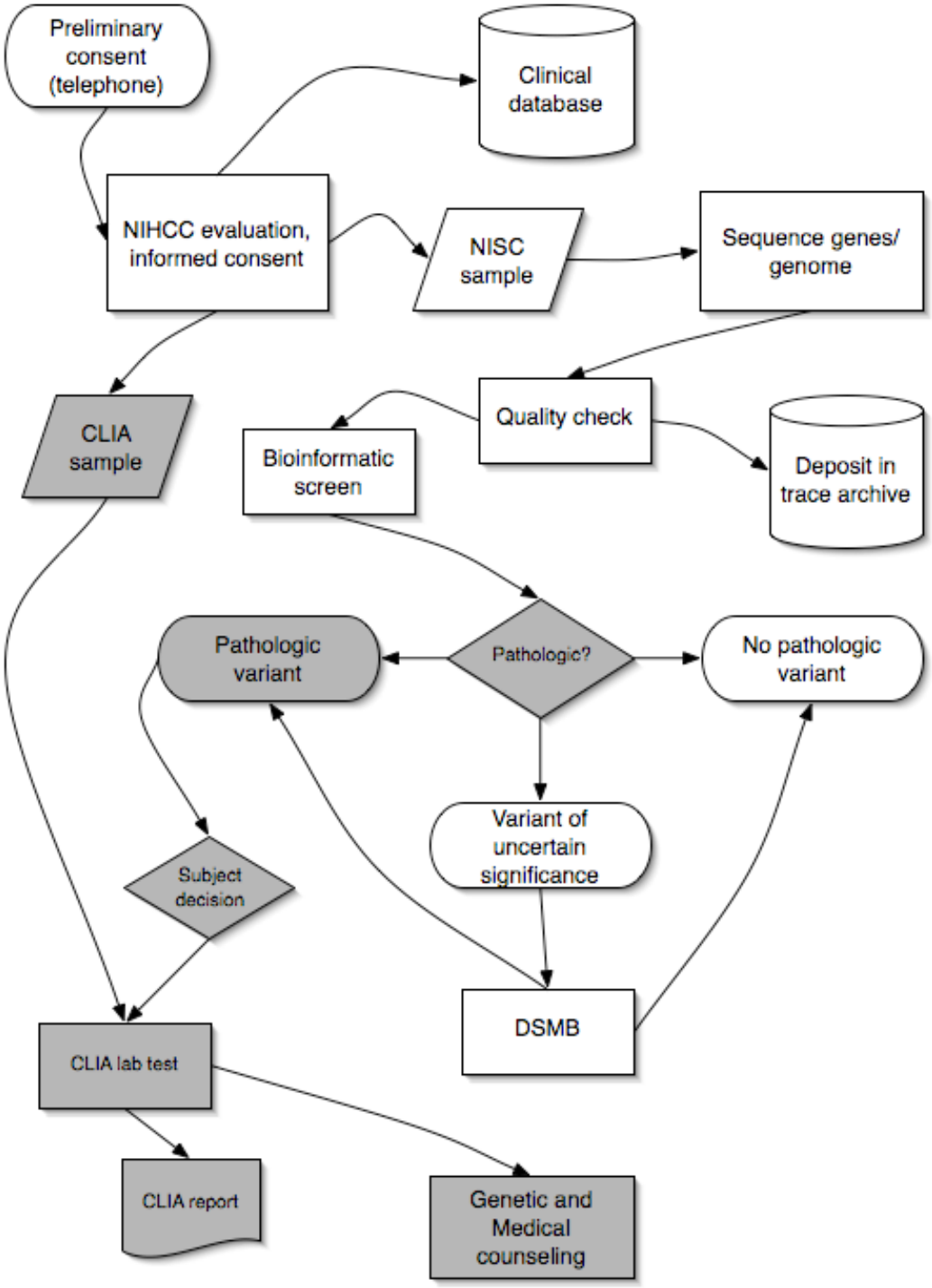
Hypothesis generating clinical research

- Collect patients with broadly defined clinical phenotype
- Sort on genomic attributes
- Refine phenotype of selected patients

Collaborators

- **NISC**
  - **Jim Mullikin, Bob Blakesly, Gerry Bouffard, Pedro Cruz, Nancy Hanson, Morgan Park, Alice Young**
- **NHGRI**
  - **Eric Green, Flavia Facio, Paul Gobourne, Jennifer Johnston, Teri Manolio, Jamie Teer, Clesson Turner, Alec Wilson**
- **NHLBI**
  - **Richard Cannon, Andrew Arai, Paul Hwang, Toren Finkel, Vandana Sachdev, Bob Shamburek**
- **NIHCC**
  - **Alan Remaley**

Results flow to subjects: Variants of clear clinical significance



Results flow to subjects: Variants of uncertain clinical significance