Massively parallel sequencing as a clinical research tool

Leslie G. Biesecker, MD
Genetic Disease Research Branch
National Human Genome Research Institute

Outline

- **Genetic architecture**
- **Sequencing**
- **ClinSeq™ study design**
  - **Preliminary results**
  - **Clinically relevant results**
  - **Basic and clinical follow-up studies**

Resource

The ClinSeq Project: Piloting large-scale genome sequencing for research in genomic medicine

Leslie G. Biesecker,[1,2,5] James C. Mullikin,[1,2] Flavia M. Facio,[1] Clesson Turner,[1] Praveen F. Cherukuri,[1] Robert W. Blakesley,[1,2] Gerard G. Bouffard,[1,2] Peter S. Chines,[1] Pedro Cruz,[2] Nancy F. Hansen,[1,2] Jamie K. Teer,[1] Baishali Maskeri,[2] Alice C. Young,[2] NISC Comparative Sequencing Program[1,2] Teri A. Manolio,[1] Alexander F. Wilson,[1] Toren Finkel,[3] Paul Hwang,[3] Andrew Arai,[3] Alan T. Remaley,[3,4] Vandana Sachdev,[3] Robert Shamburek,[3] Richard O. Cannon,[3] and Eric D. Green[1,2]

Pilot Clinical Large-Scale Medical Sequencing (LSMS)

- **Pilot LSMS on a *scale* that is practical yet answers initial questions**
- **The goal is to model the attributes of whole genome LSMS as a clinical research tool**

ClinSeq™

Pilot Clinical Large-Scale Medical Sequencing (LSMS)

- **Pilot LSMS on a *scale* that is practical yet answers initial questions**
- **The goal is to model the attributes of whole genome LSMS as a clinical research tool**
- **Start with subsets and design everything to be scalable**

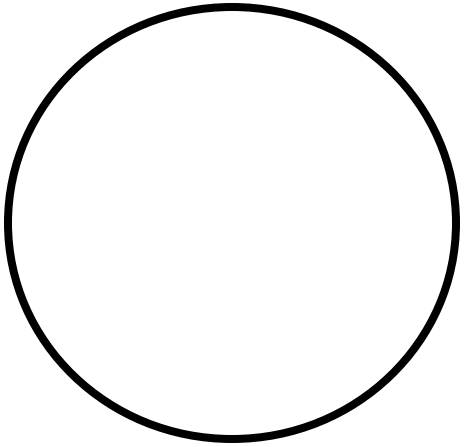
ClinSeq™

Specific aims

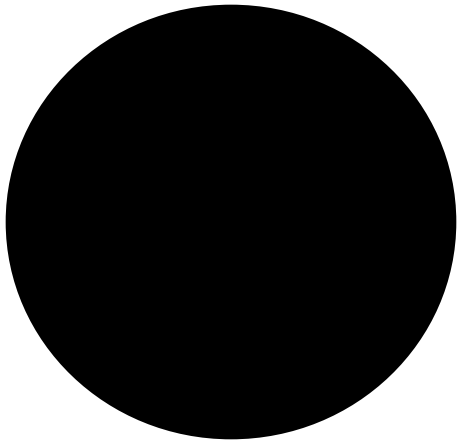
- Develop a robust infrastructure for the generation and use of LSMS in a clinical research setting
- Use LSMS data to address biomedical research questions
- To understand how to interact with subjects re LSMS
- Build a cohort of clinical research subjects consented for whole genome LSMS and f/u studies

ClinSeq™

Conventional View of Genetics

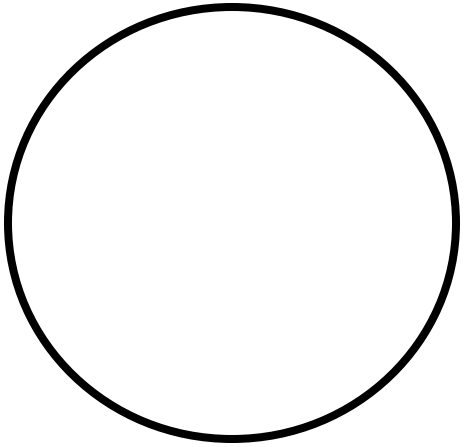


Rare Diseases
Rare Variants

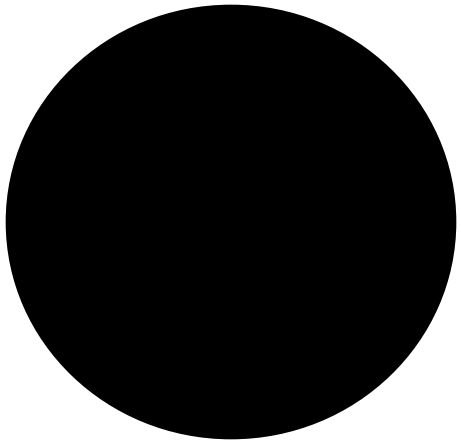


Common Diseases
Common Variants

Conventional View of Genetics



Rare Diseases
Rare Variants

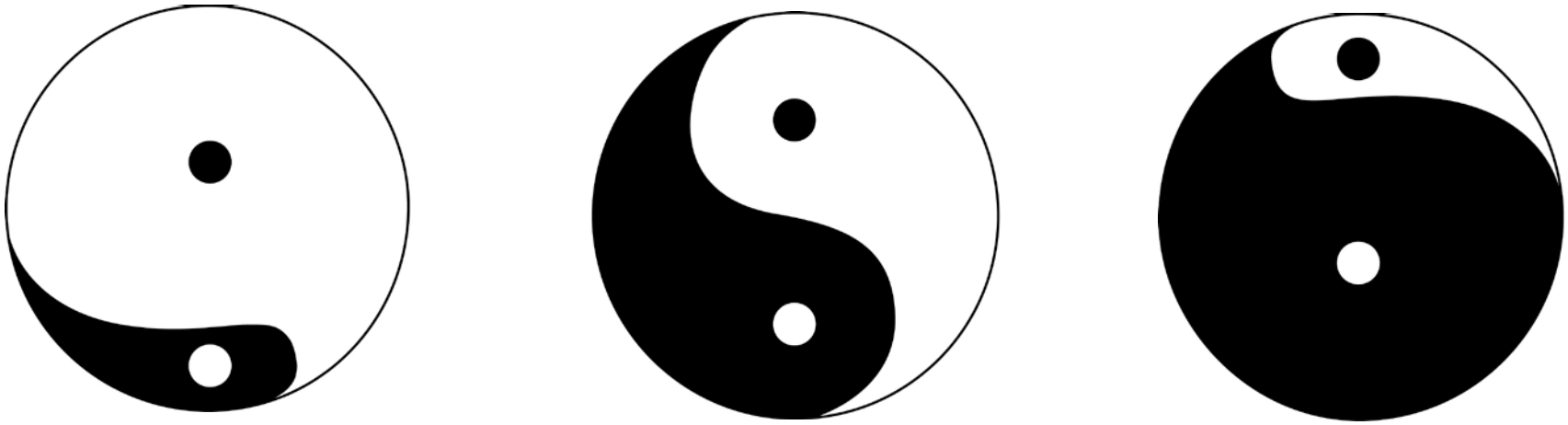


Common Diseases
Common Variants

But, biology is continuous, not dichotomous…

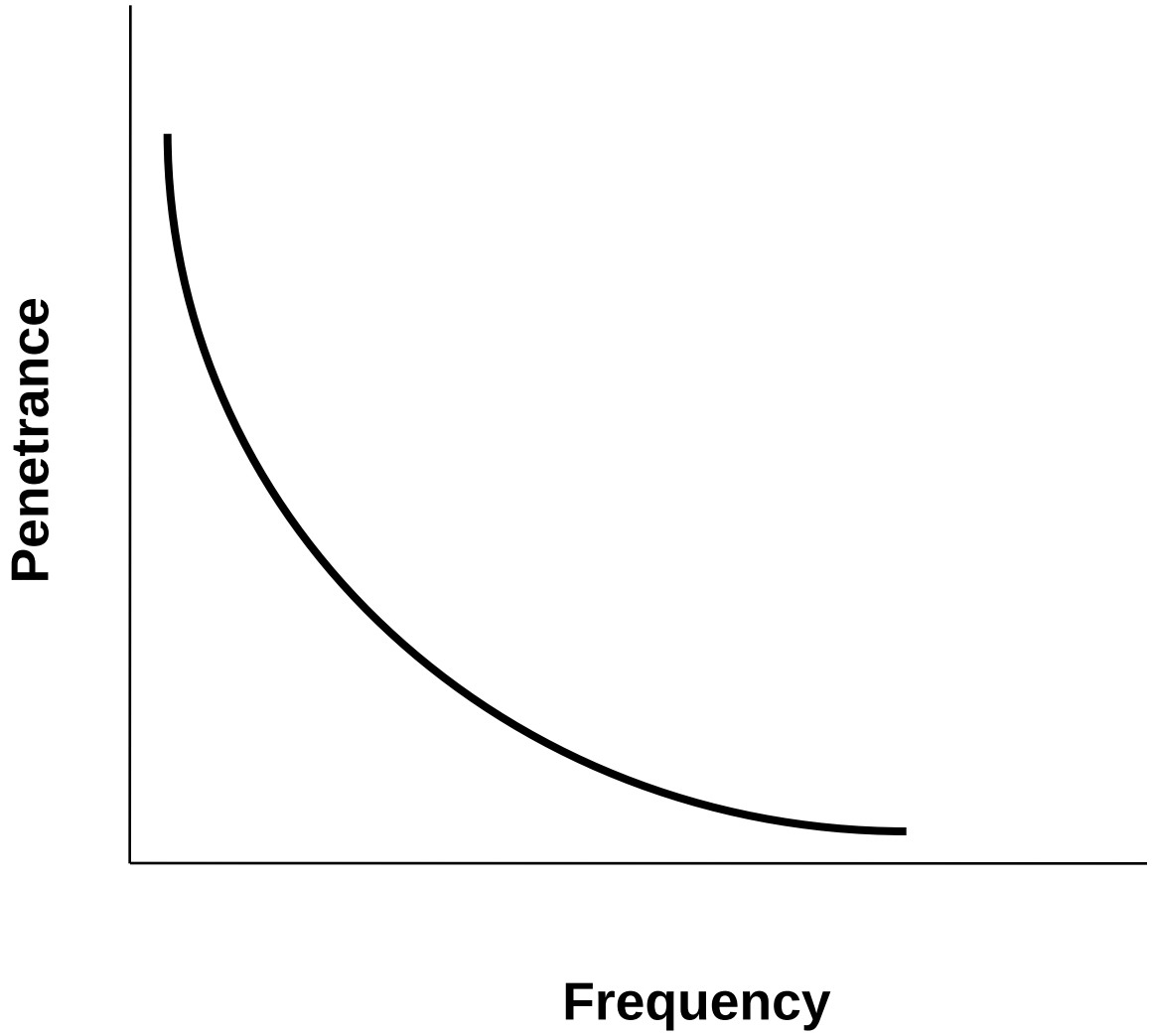
Genetic Architecture

Common variants

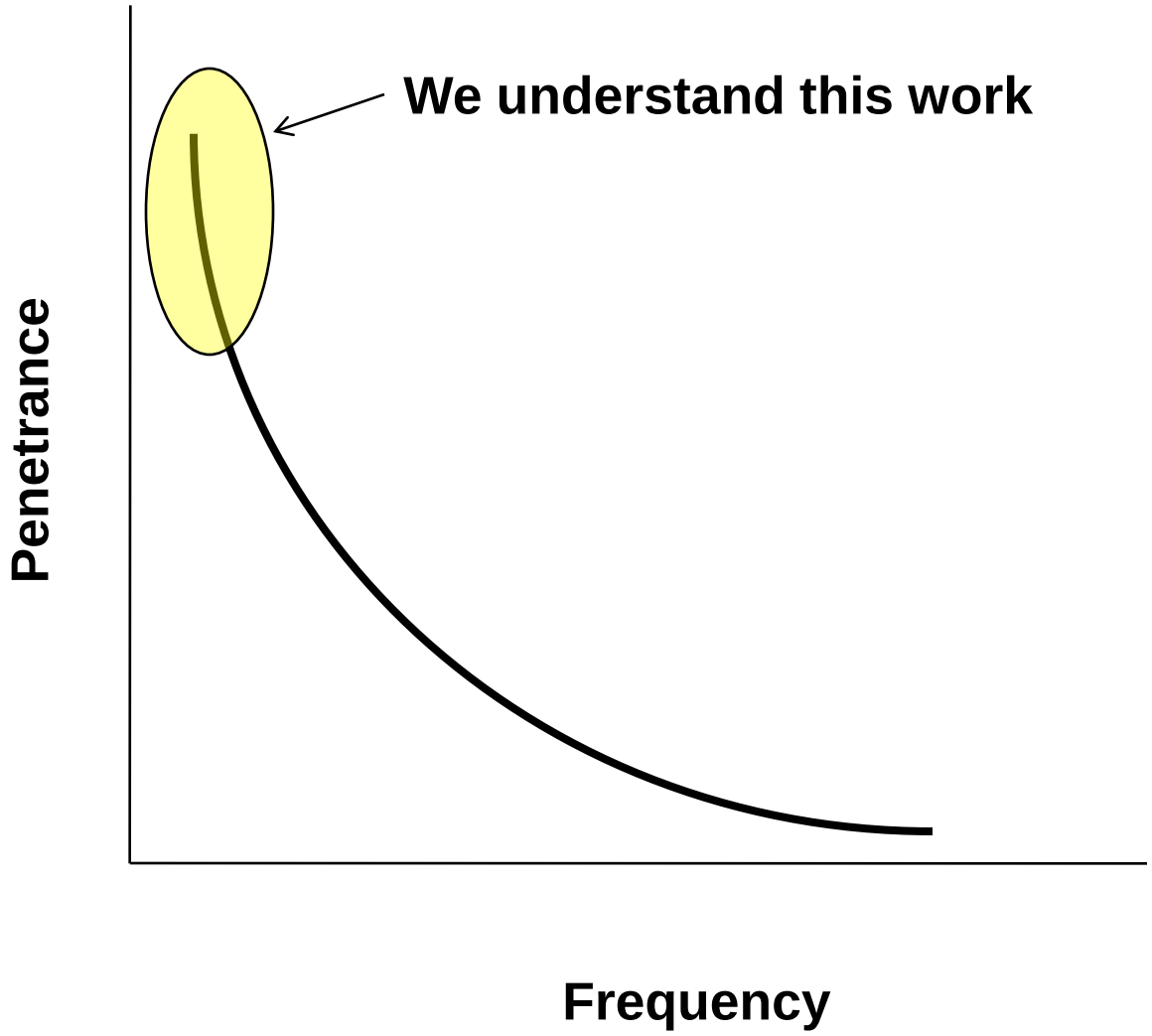


Rare variants

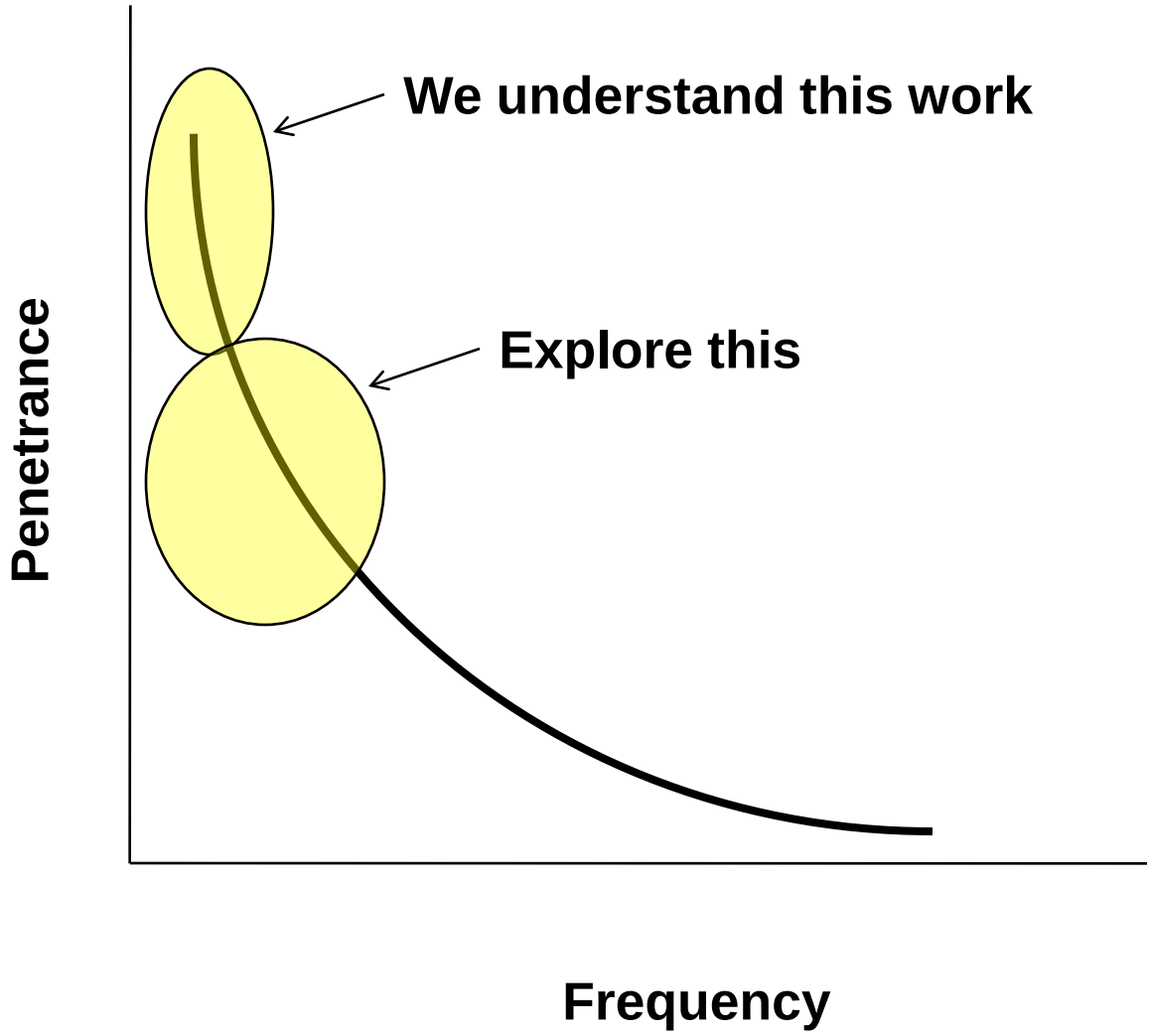
Penetrance & frequency



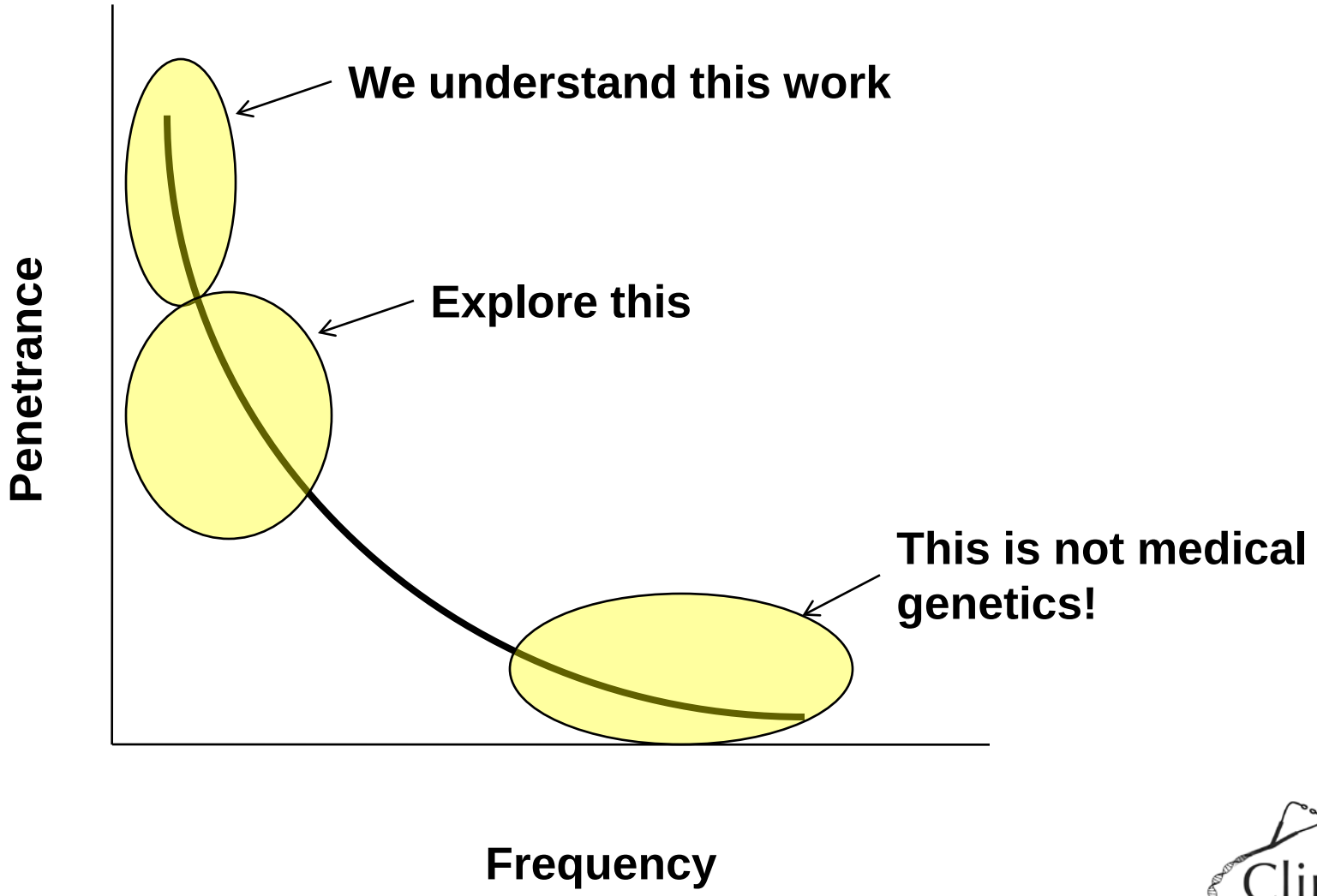
Penetrance & frequency



Penetrance & frequency



Penetrance & frequency

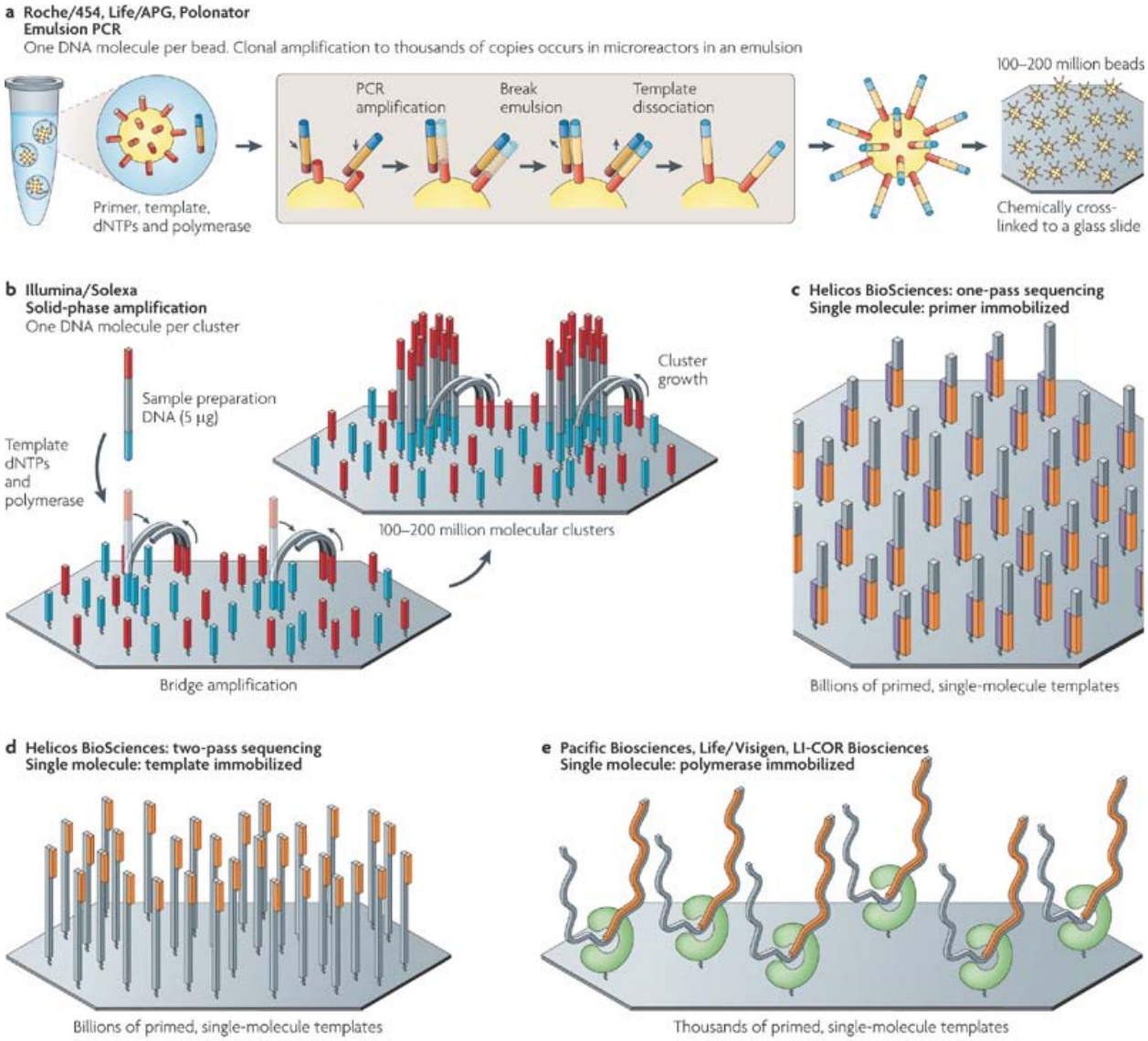


Sequencing Technologies (2007)

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

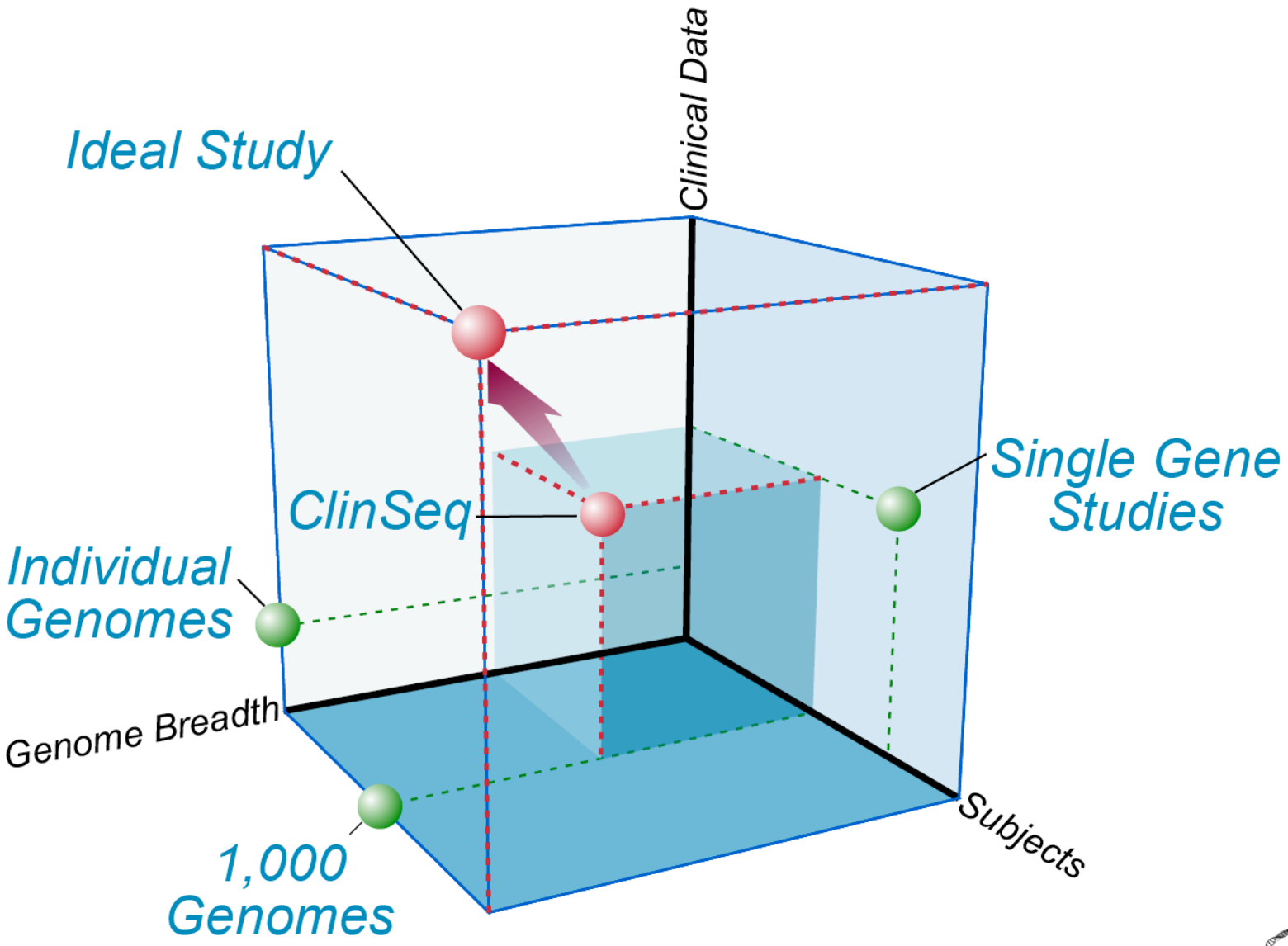
Equivalent to two 30X mammalian genomes/year

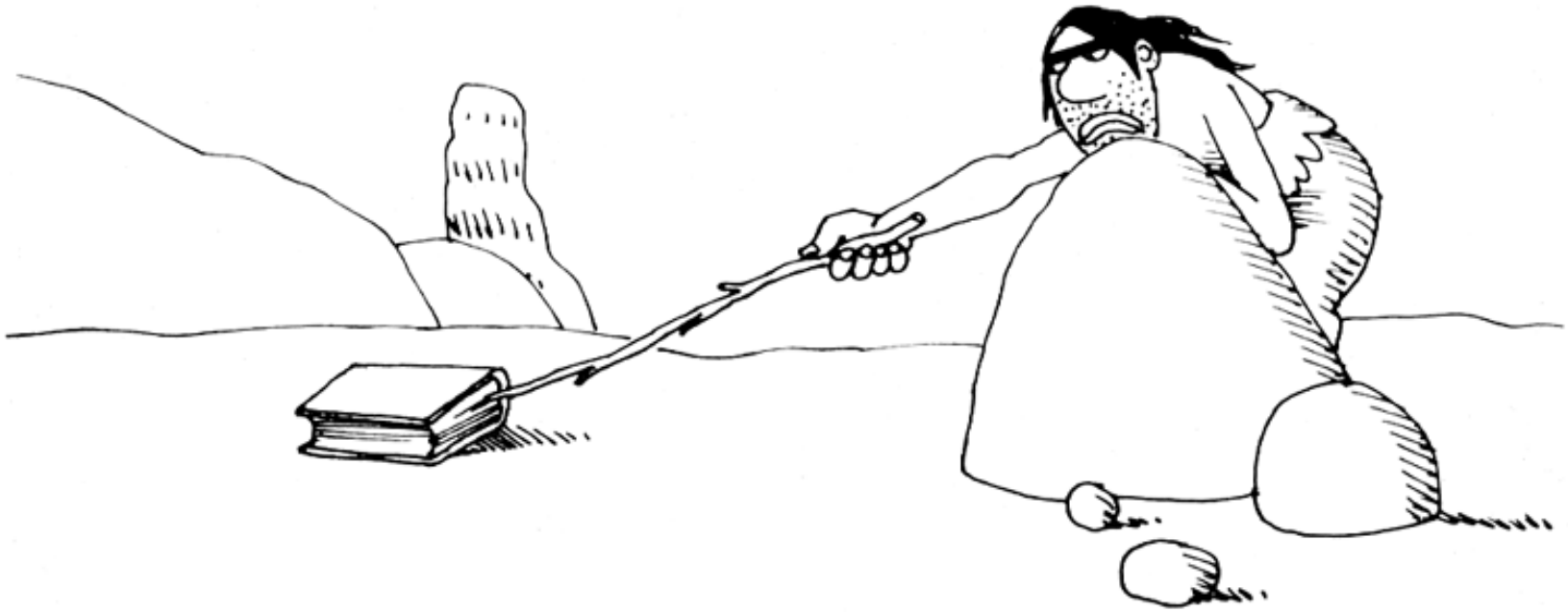
“Next-gen” sequencing



>2 x 10^{11} bp per run

Clinical Genomics Research Space





Approach

- **Initial target phenotype: Atherosclerosis**
  - **Phenotype 1,000 subjects**
    - **Coronary calcium, lipidemia, & related indices**
  - **Sequence 200-400 candidate genes**
  - **Follow-up studies**
    - **Bioinformatic & bench**
  - **Interpret variants and validate *some***
  - **Return results**

ClinSeq™

ABCA1
ABCG5
ABCG8
ACE
ACTA2
AGTR1
ALOX
ANGPTL3
ANGPTL4
ANRIL
APOA2
APOA5
APOB
APOBEC1
APOC1
APOC2
APOC3
APOC4
APOE
APOM
ARF
ARG1
BDKRB2
CALM1
CAPG
CAV1
CCL2
CCR2
CDKN1A
CDKN2A
CDKN2B
CELSR2
CETP
CRP
CYBA
DUSP1
ENPEP
FABP4
FAM5C
FAS
FLI1
FOS
FTO
GAL
GALNT2
GCKR
GDF5
HMGCR
HMOX1
ICAM1
INSIG2
ITGA2
ITGB1
ITGB2
ITGB3
ITGB5
ITGB7
KCNMB1
LCAT
LDLR
LDLRAP1
LIPC
LIPE
LIPG
LPL
LRP6
MLXIPL
MMAB
MMP9
MPO
MTP
MVK
MYLK
NCCT
NFKB1
NKCC2
NOS3
NPC1
NPC1L1
NPR1
PCSK9
PDGFB
PDGFRB
PER1
PITX2
PLA2G7
PLAT
PLTP
PON1
PPARA
PPARD
PPARG
PRDX2
PSRC1
PTGS2
RBKS
ROMK
SCARB1
SELL
SELS
SOAT1
SOAT2
SOCS3
SORT1
SREBF1
SREBF2
TCF1
TCF7L2
THBS4
TLR4
TRIB1
USF1
VAMP8
VEGF
VLDLR
VNN1

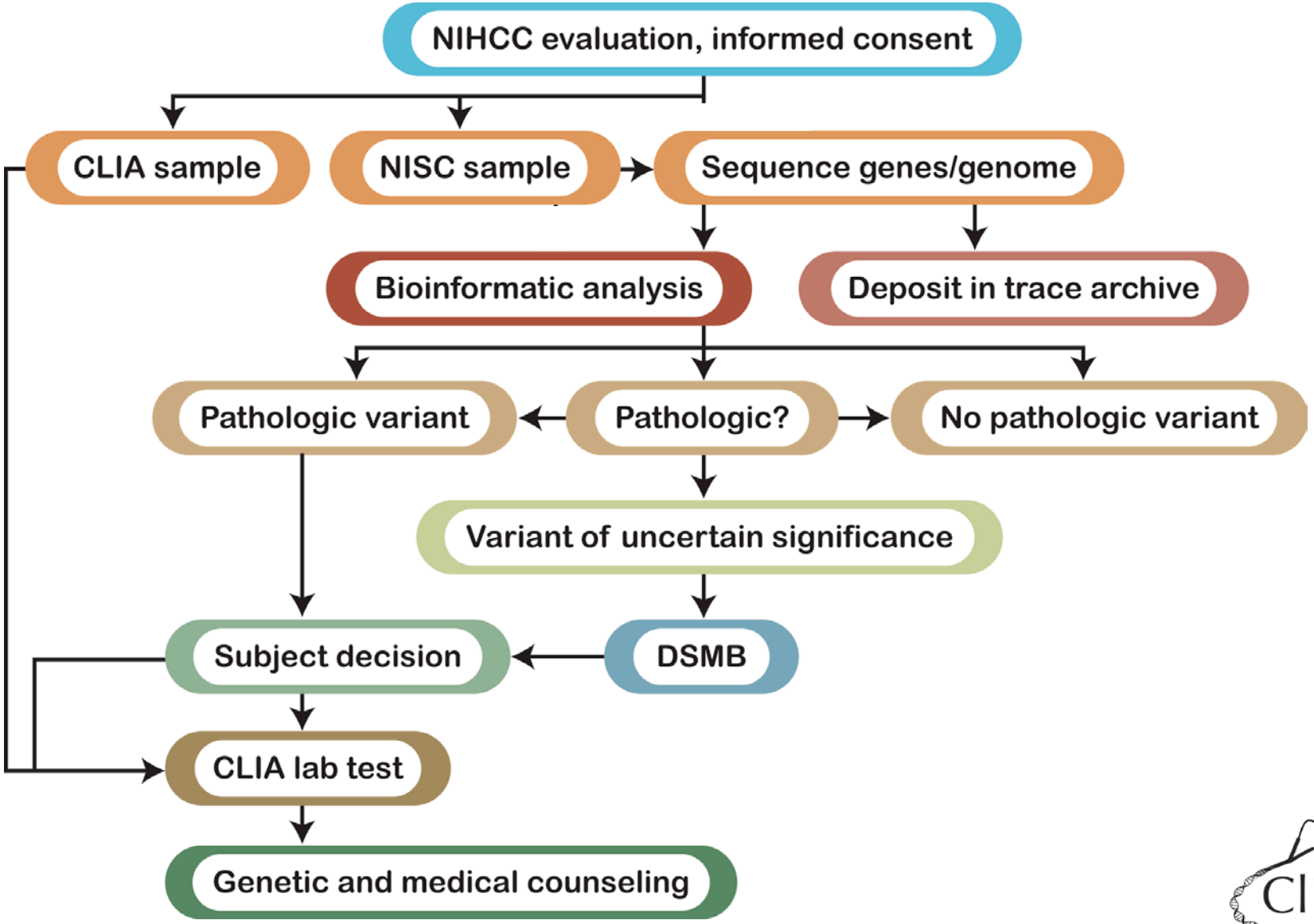
ClinSeq™

Clinical evaluation

- **Family history (semiautomated)**
- **Medical history (form-driven)**
- **Coronary calcium score (MDCT)**
- **Echo/electro-cardiography**
- **Clinical & research bloods**

ClinSeq™

Data flow


NIHCC evaluation, informed consent
CLIA sample
NISC sample
Sequence genes/genome
Bioinformatic analysis
Deposit in trace archive
Pathologic variant
Pathologic?
No pathologic variant
Variant of uncertain significance
Subject decision
DSMB
CLIA lab test
Genetic and medical counseling

Progress: Enrollment and Sequence

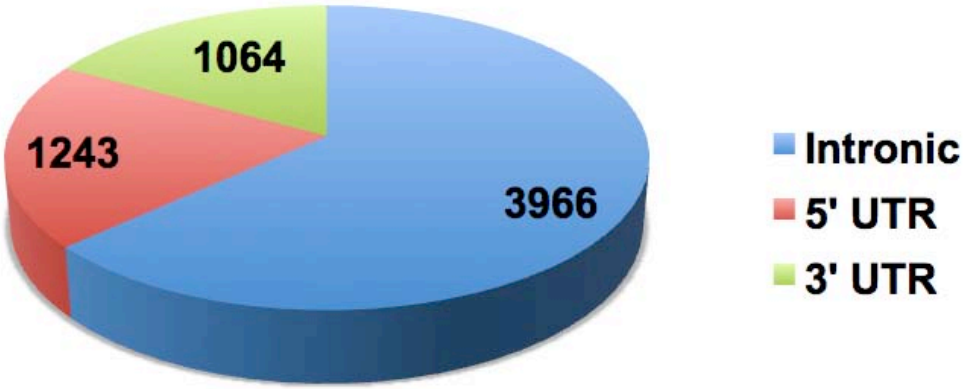
- Enrollment began January, 2007
- >775 patients enrolled Nov 30, 2009
- 454 DNAs sequenced – PCR/3730
  - 425 ClinSeq subject samples
  - 29 HapMap samples
- 251 genes
  - 2,725 genomic target sequences
- > 2.3 M sequence reads to date

Progress: Enrollment and Sequence

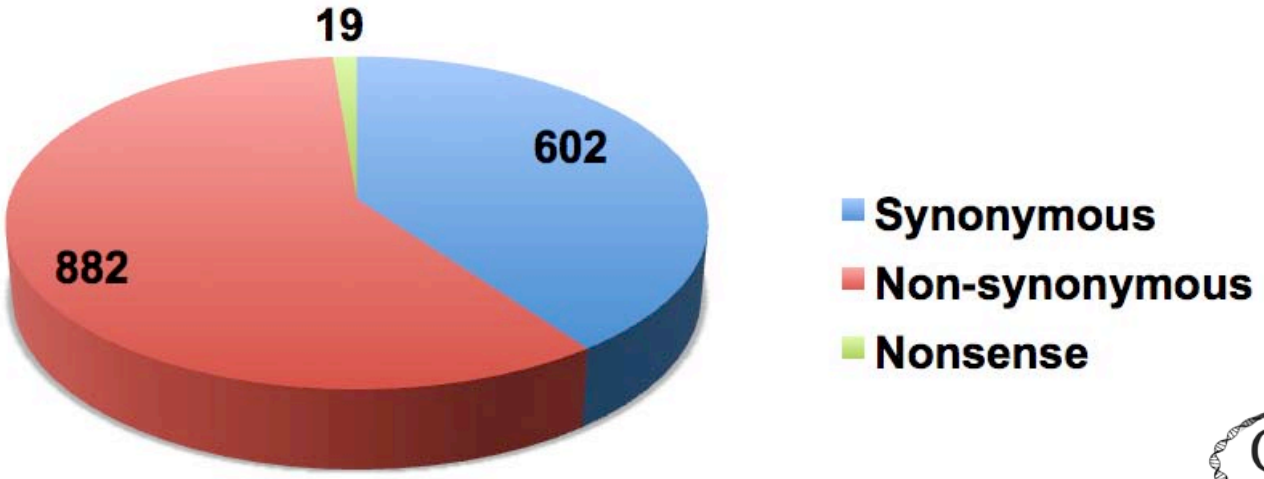
- Enrollment began January, 2007
- >775 patients enrolled Nov 30, 2009
- 454 DNAs sequenced – PCR/3730
  - 425 ClinSeq subject samples
  - 29 HapMap samples
- 251 genes
  - 2,725 genomic target sequences
- > 2.3 M sequence reads to date

- One 50X WGS sequence completed Nov 2009
- Eight exomes done; scaling to 100-200 per year

9,696 Non-exonic, Non-dbSNP Variants

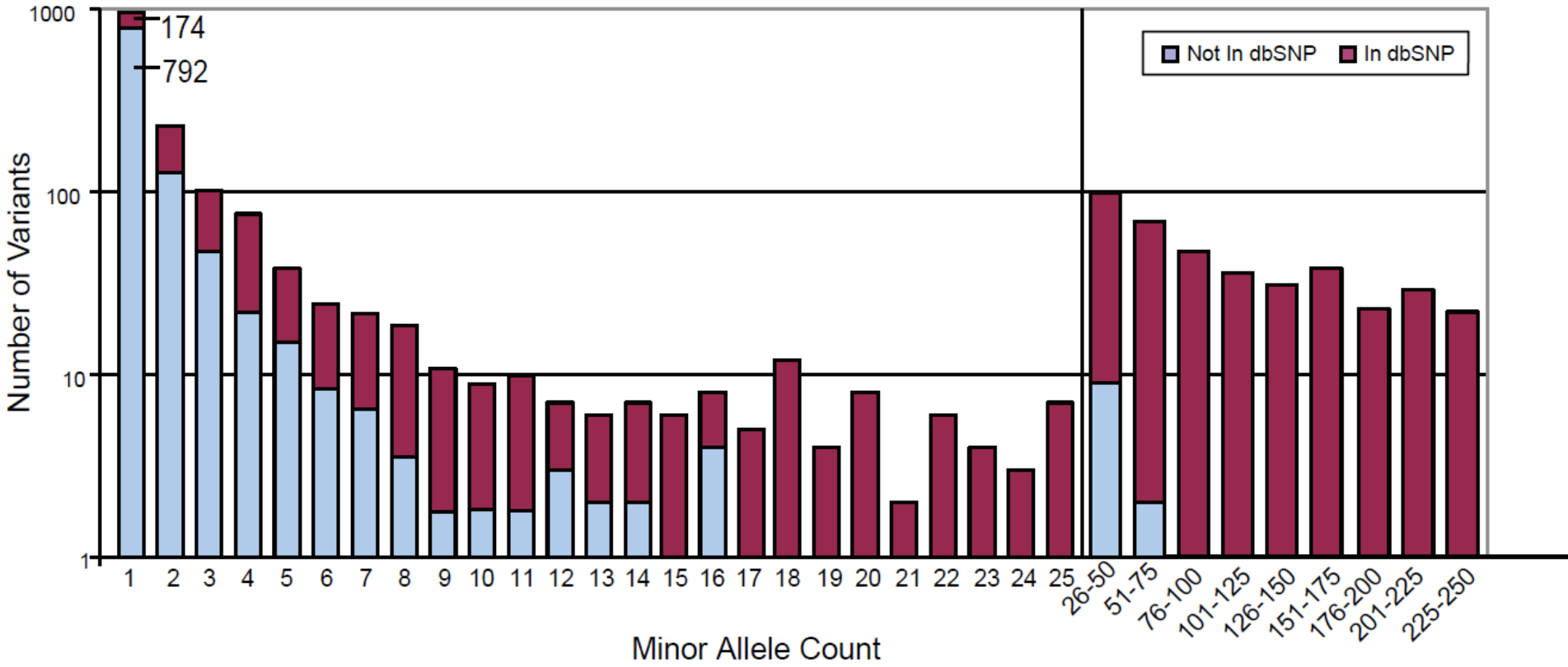


2,395 Exonic, Non-dbSNP Variants



ClinSeq™

Uncommon Variants Common



Clinically Relevant Variants

- **LDLR**
  - **p.W87G**
  - **c.313+1G>A**
  - **p.Y188X**
  - **p.G218VfsX47**
  - **p.P685L**
  - **p.N688QfsX29**
  - **p.W87X**
- **APOB**
  - **p.R3527Q**

Clinically Relevant Variants

- **LDLR**
  - **p.W87G**
  - **c.313+1G>A**
  - **p.Y188X**
  - **p.G218VfsX47**
  - **p.P685L**
  - **p.N688QfsX29**
  - **p.W87X**
- **APOB**
  - **p.R3527Q**

8/400 = 1/50
About 10X the population frequency

Implications

- Subjects are aware of and seek out genetic assessment
- Scanning a genome for pathologic variants and returning results practical
- Eight families with 75-100 people probably affected or at risk for FH
- Studying approach to case finding

ClinSeq™

Sub-projects underway

- Positive controls
- Validating recent assoc. of rare variants & phenotypes
- Sequencing genes under GWAS peaks for rare, high penetrance variants
- Testing associations of candidate genes with phenotype
- Control cohort for other sample sets
- Search for miRNA variants
- cDNA sequencing pilot to measure expression & isoforms
- Capture method refinement
- Patient motivations and preferences for results of medical sequencing
- Testing automated vs. manual pedigree acquisition

ClinSeq™

Classical Clinical Research

- Precisely delineate disorders
- Exhaustively phenotype patients
- Generate genetic data
- Determine causation
- Design follow-on studies

ClinSeq™

Hypothesis-Generating Clinical Genomics

- Precisely delineate disorders
- Exhaustively phenotype patients
- Generate genetic data
- Determine causation
- Design follow-on studies

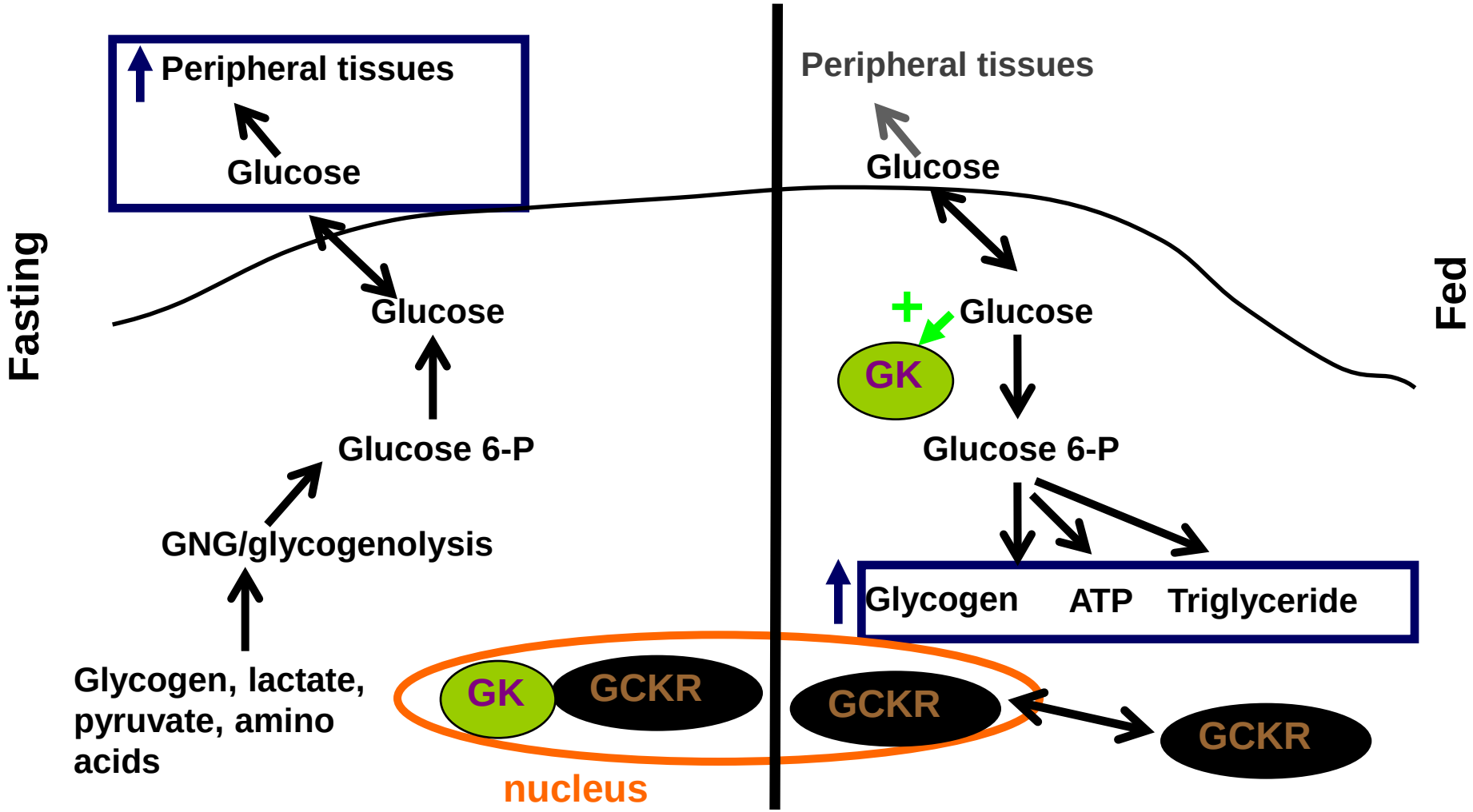
- Generate gen*omic* data
- Design follow-on studies
- Exhaustively phenotype patients
- Precisely delineate disorders
- Determine causation

ClinSeq™

Rare GCKR Variants

Also: 1135insA

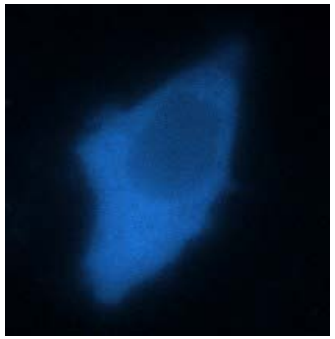
GCKR Function (Liver)



In Vitro

- **GCKR (yellow) is required to sequester GKR (blue/cyan) in the nucleus**

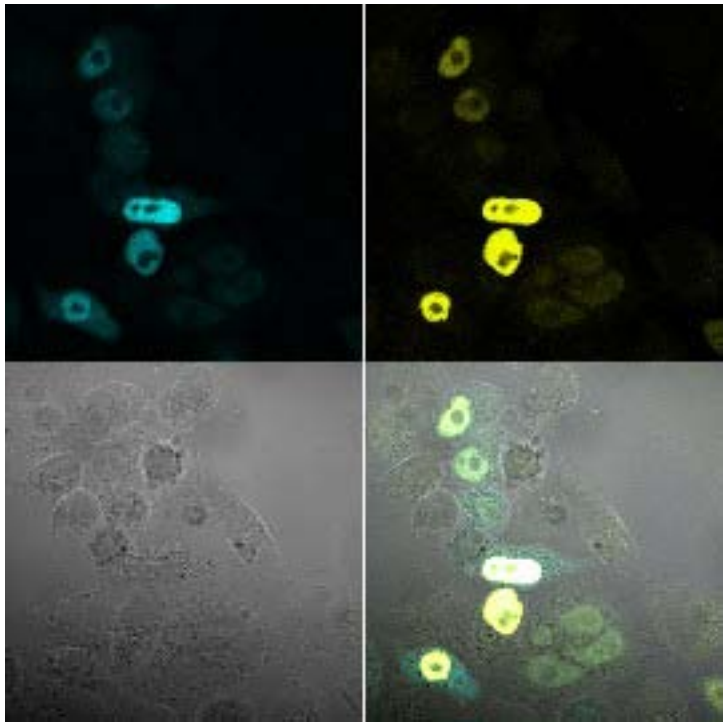
CFP-GCK in HeLa



CFP-GCK and YFP-GCKR in HeLa

GCK

GCKR



phase

overlay

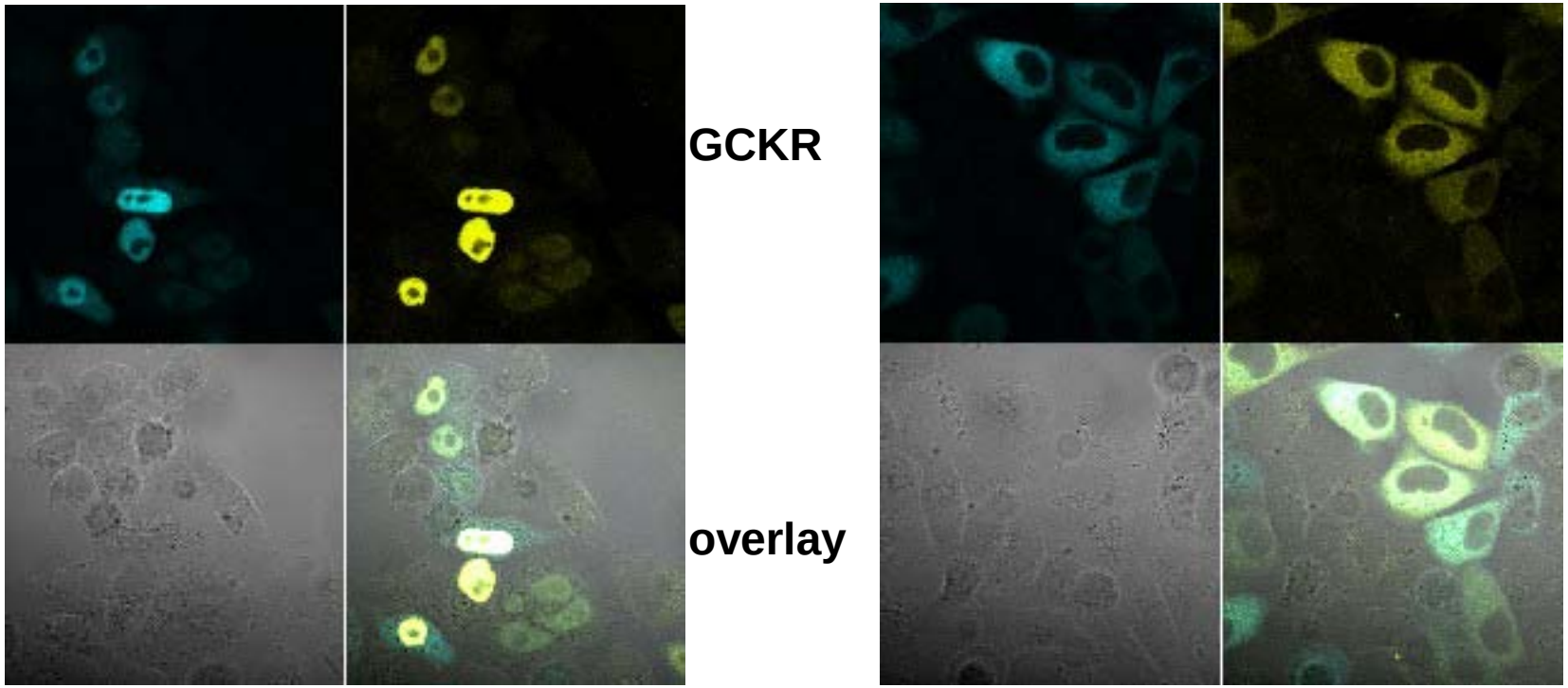
p.Q234P

- **Protein expressed & stable**
- **GCK primarily cytoplasmic**

wt **p.Q234P**

GCK **GCKR**

phase **overlay**



Subject's Views

- **Who wants it & why?**
- **What subset of results they want?**
- **How do they view & interpret results?**
- **How to consent for LSMS?**

Initial Qualitative Survey

- **Participants highly educated**
  - **~85% with post-baccalaureate education**
- **Open ended questions, *post hoc* coding**
  - **Primarily interested in contributing to new genetics research**
  - **Primarily interested in study to find personal genetic health information**

ClinSeq™

Implications

- **Large numbers of patients are interested**
- **Can consent subjects to whole genome interrogation**
- **Clinically relevant results can be sifted**
- **Subjects can receive and act upon results**
- **Scaling:**
  - **Next-Gen: Exome -> Whole genome sequencing**
  - **More diverse population**
  - **Discovering associations of variants and phenotype**
  - **Learning subject's view in real setting**

ClinSeq™

The Future of Clinical Research

- **MUCH more data**

ClinSeq™

The Future of Clinical Research

- **MUCH more data**
- **More bioinformatics than pipetting**

ClinSeq™

The Future of Clinical Research

- **MUCH more data**
- **More bioinformatics than pipetting**
- **Transdisciplinary work essential**

ClinSeq™

Transdisciplinary Approach

Medical & Statistical Genetics

NIH Intramural Sequencing Center

NHLBI

NIH Clinical Research Center

NIDDK NATIONAL INSTITUTE OF DIABETES AND DIGESTIVE AND KIDNEY DISEASES

ClinSeq™

Collaborators

- **NISC**
  - **Jim Mullikin, Bob Blakesley, Gerry Bouffard, Pedro Cruz, Nancy Hanson, Morgan Park, Alice Young, Praveen Cherukuri**
- **NHGRI**
  - **Eric Green, Flavia Facio, Jennifer Johnston, Teri Manolio, Jamie Teer, Clesson Turner, Alec Wilson, Matt Rees, Francis Collins, Shurjo Sen**
- **NHLBI**
  - **Richard Cannon, Andrew Arai, Paul Hwang, Toren Finkel, Vandana Sachdev, Bob Shamburek, Alan Remaley**
- **NIDDK**
  - **Monica Skarulis, Krisina Röther**
- **U Wash**
  - **Evan Eichler**