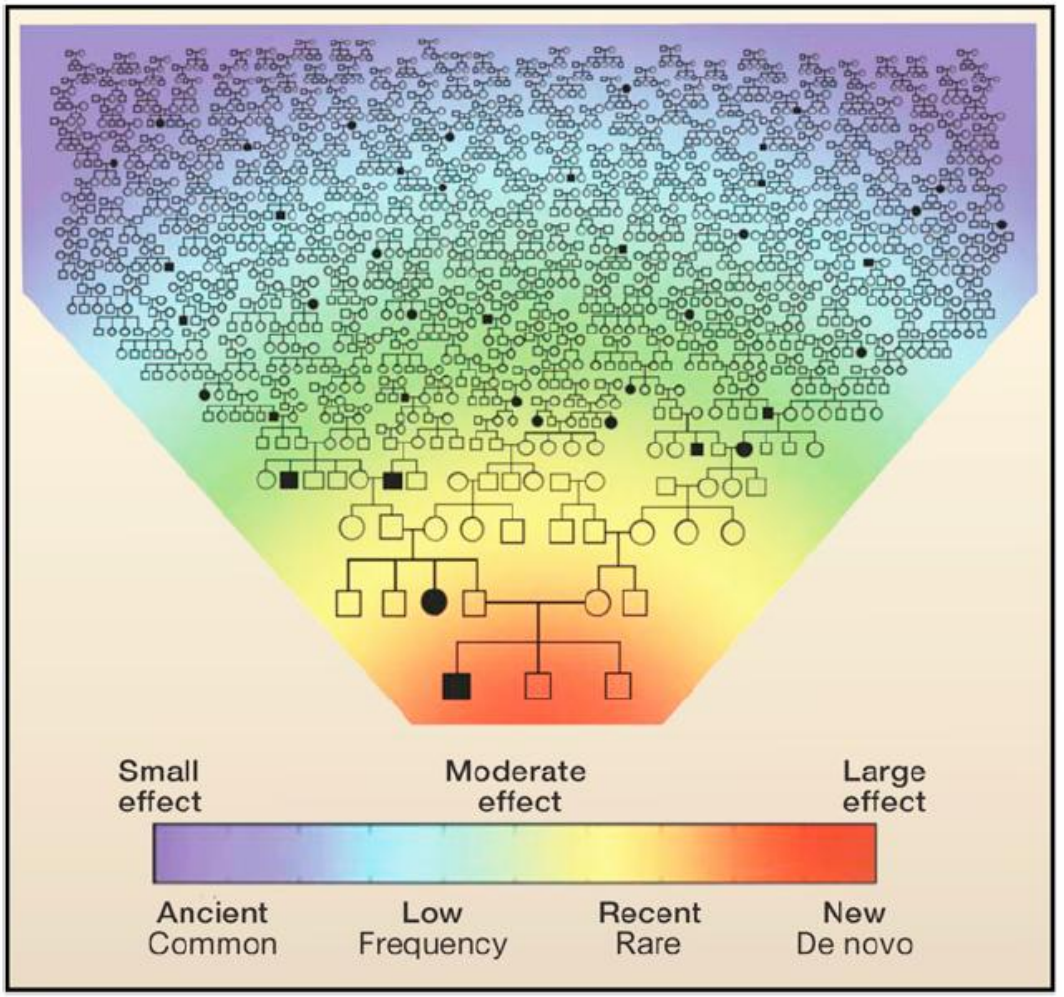
Lessons Learned from Sequencing and Genomics

Richard Gibbs, Baylor College of Medicine

Human Genome Sequencing Center



COI: Co-investment with Life Technology; Co-Founder SeqWright

Lessons (and outline):

1. TECHNOLOGY: Mind-boggling
   (quality not what is used to be)
2. BACKGROUND VARIATION: Enormous
   (hard to find ***important*** variation)
3. GENETICS *RULES*: Sequencing does not solve everything
   (genetics and functional studies most important)
4. ANALYSIS: Building robust pipelines
   (It takes a village….)
4. UBIQUITY: Social trend will be for ubiquitous sequencing
   (cohorts replaced by medical records?)
5. THIS PROJECT: How it might look?
   (a strawman….)

So Far:

SCD a ‘monogenic disorder’ ***but***:

1. Exceptions to the primary mutation
2. Modifiers of severity
3. Differential response to primary therapy
4. Different predisposition to other effectors
   e.g. pain, infection etc
5. Other

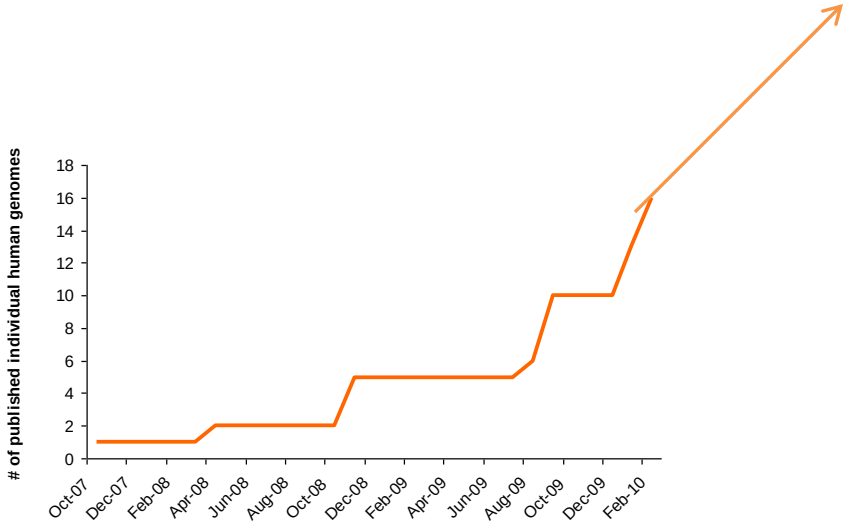
Each of these ‘secondary genetic’ influences represents the same challenge as more complex phenotypes!

Raw sequence production.….<< $50/Gb!!

Wild guess – today 4,000 genomes

Mind boggling!!

5,000 2011 Genomes?
30,000 2012 Genomes?



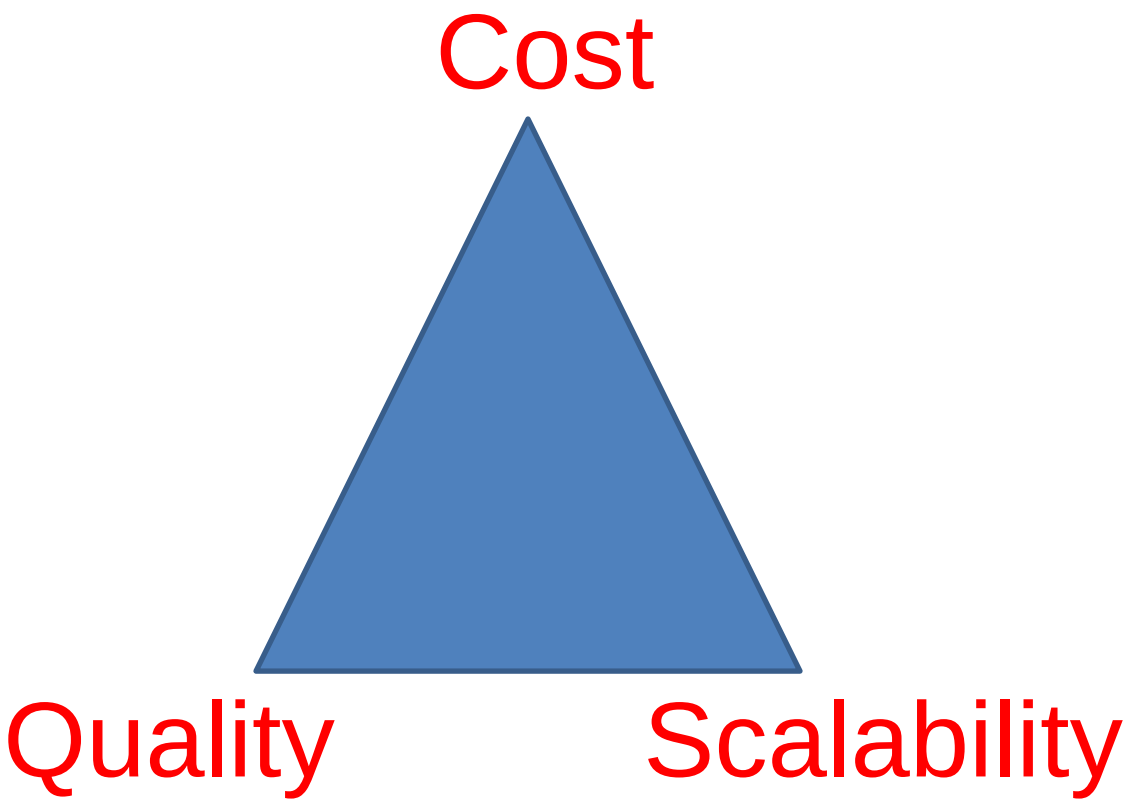
1. TECHNOLOGY

EXOMES INSTEAD OF GENOMES??.

	Whole-Genome Sequencing (WGS)	Exome Sequencing
Cost	Still costly, but decreasing rapidly	
Technical	No capture step, automatable	
Variation	Uncovers ALL genetic and genomic variation (SNVs + CNVs) Discovery of functional coding and non-coding variation ~3.5 Million variants	Limited to coding and splice-site variants in annotated genes ~20,000 variants
Disease	Suitable for mendelian and complex trait gene identification, as well as sporadic phenotypes caused by de novo SNVs or CNVs	Good for highly penetrant mendelian disease gene identification

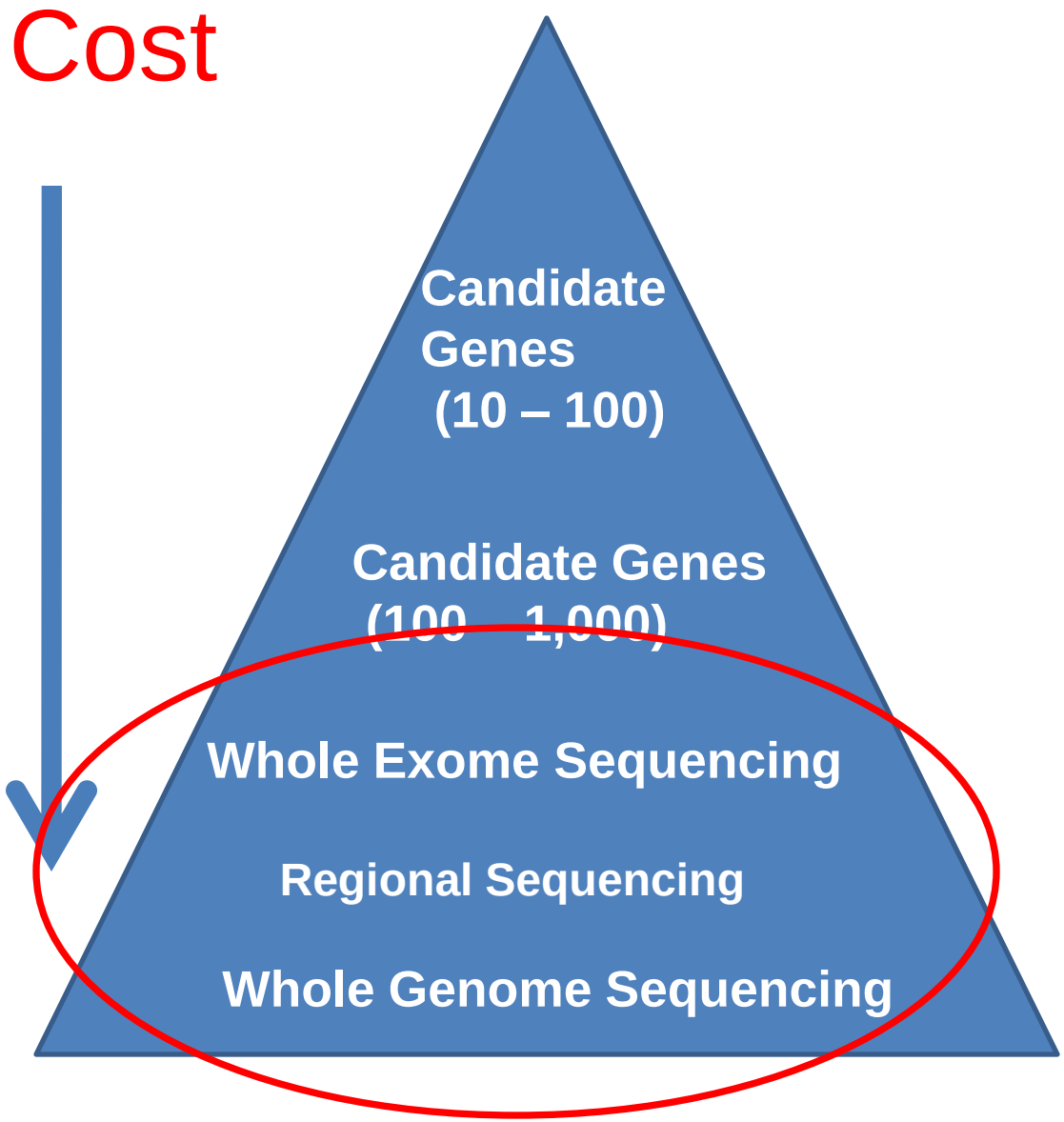
2011, 'The CORE' With Supplements: Aim >99% of exons

THE SEQUENCING TRIAD



Current methods have lower quality……

Sequencing for discovery:



Known genes

Gene interactions

Coding mutation burden

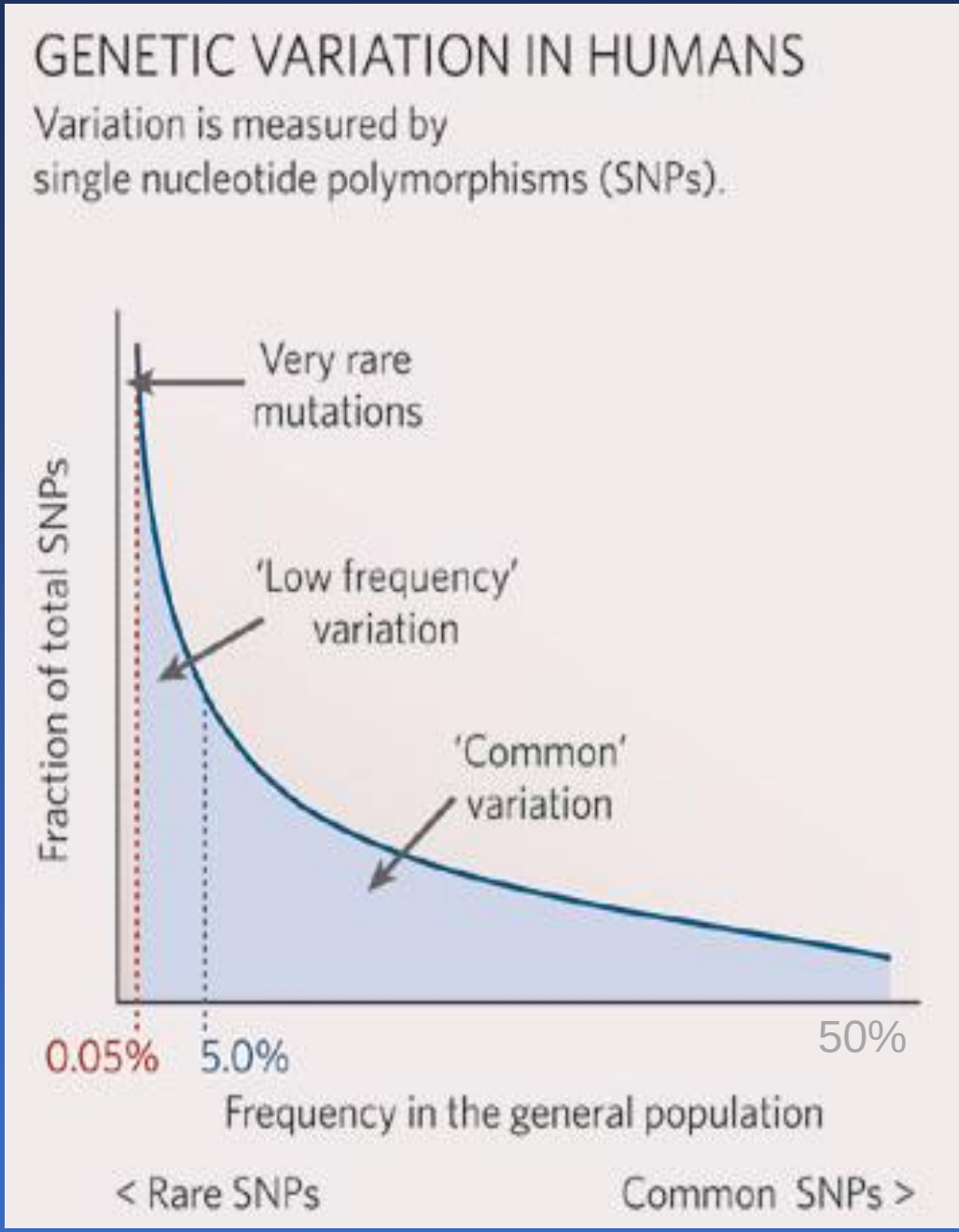
Exons + flanks

Whole genome variation

The Site Frequency Spectrum (SFS)

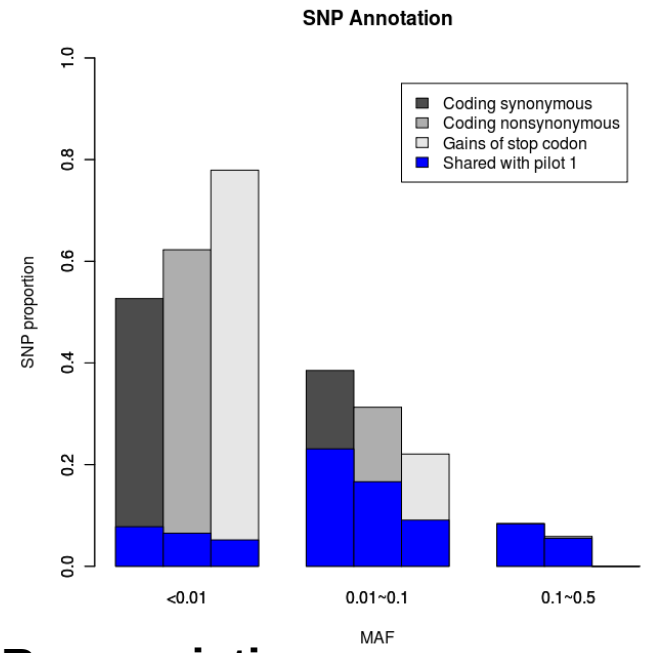
What does our knowledge of population variation tell us about the challenges for generally solving the challenges of common diseases?

What is the nature of BACKGROUND rare variation?

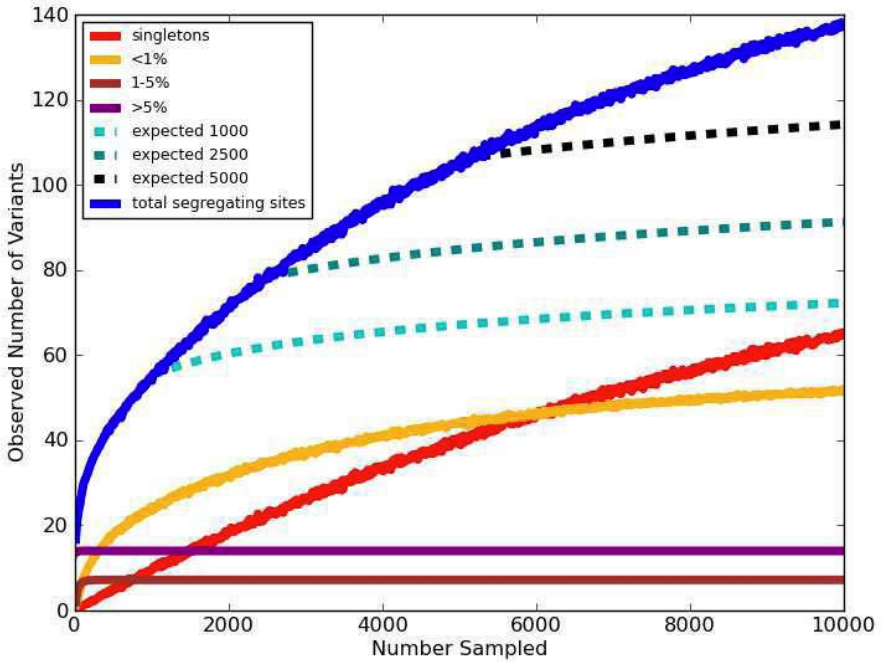


Rare Variants in the Site-Frequency –Spectrum

- ~10,000 ns vs reference
- 1 - 300 novel ns variants per person
- More ***novel*** functional variants than *previously* expected
- ***Huge*** impact on assessing functional significance



Rare variation more likely to be deleterious!!



Micortex:
~ 11,000 samples, ~ 25 kb

3: GENETICS Vignette I: NIMH/NHGRI Autism Sequencing

1,000 Autism cases from NIMH Collection

1,000 Matched controls

-Collaboration BI and BCM (Sequencing)
- Multiple analysis centers
- Aimed for all exonic variation
-Statistical differences between mutation burden?

New 'Autism Genes'?

Two years later……..a lot of data!!:

	Case			Control			Total
	Male	Female	Sum	Male	Female	Sum	
BCM	440	65	505	240	251	491	996
BI	344	86	430	177	202	379	809
			935			870	

> 500,000 variant sites

Large QA/QC effort

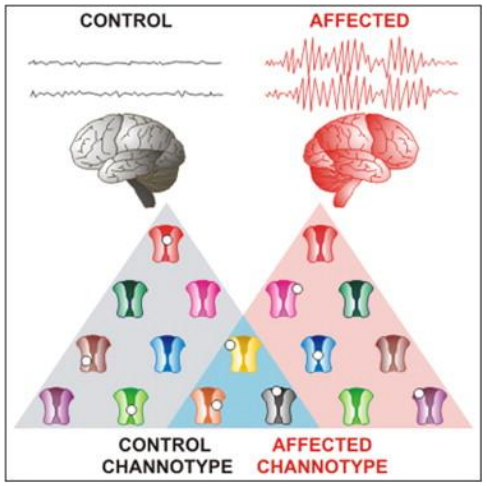
Battery of tests

No new loci (yet)

Testing for *de novo* mutations..

3: GENETICS Vignette II: Epilepsy Candidate Gene Sequencing

- 237 ion channel genes
- Case vs control study'
- Few familys
- No definitive 'hits'
- Lots of hypotheses!



Jeff Noebels

Cell,145:1036-1048 (2011),

3: GENETICS Vignette III: Autism Candidate Gene Sequencing

Human Molecular Genetics, 2011 **1–10**
doi:10.1093/hmg/ddr243

Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders

Christian P. Schaaf[1,†], Aniko Sabo[2,†], Yasunari Sakai[1,4], Jacy Crosby[5], Donna Muzny[2], Alicia Hawes[2], Lora Lewis[2], Humeira Akbar[2], Robin Varghese[2], Eric Boerwinkle[5], Richard A. Gibbs[1,2,*] and Huda Y. Zoghbi[1,3,4,6,*]

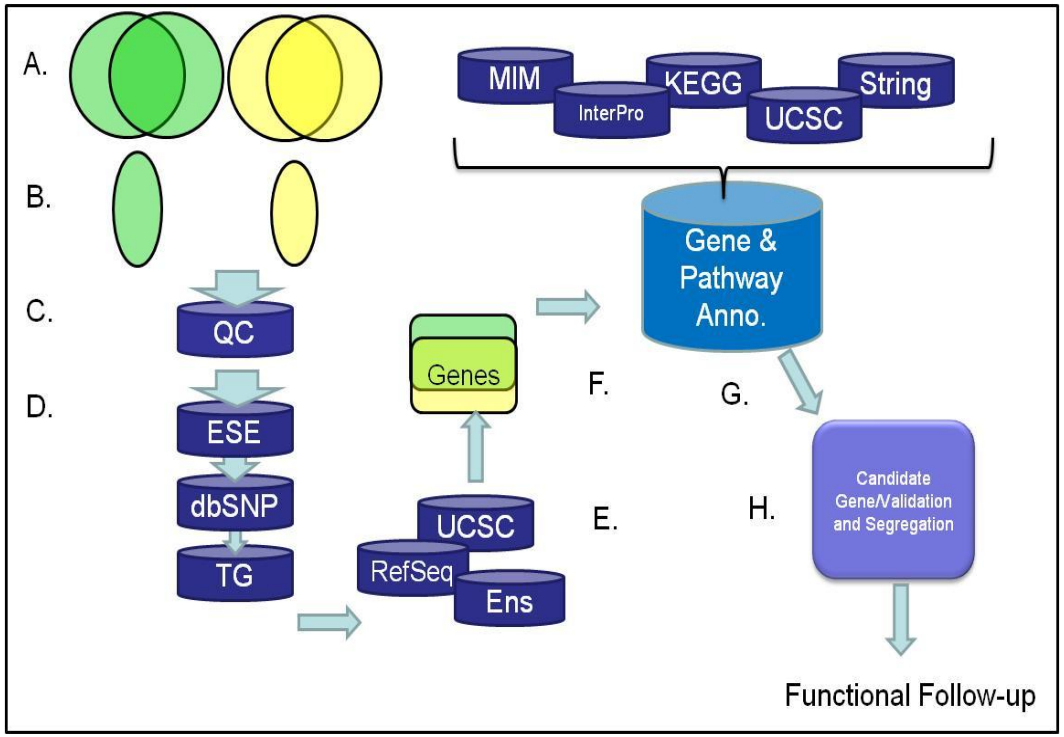
Observed	Case	Control	Total
Oligogenic event	18	6	24
No oligogenic event	321	370	691
Total	339	376	715

Aniko Sabo

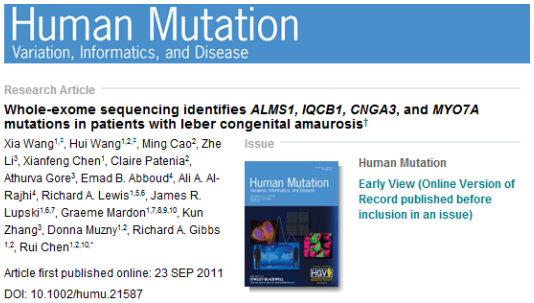
Christian Schaaf

Huda Zoghbi

Contrast to: Solving Mendelian diseases:



-Dozens are being reported each month
-> 50 underway at BCM
- 10% new genes
- 50% 'oops'
- remainder in process


Human Mutation
Variation, Informatics, and Disease

Research Article

Whole-exome sequencing identifies *ALMS1*, *IQCB1*, *CNGA3*, and *MYO7A* mutations in patients with leber congenital amaurosis†

Xia Wang[1,2], Hui Wang[1,2,*], Ming Cao[2], Zhe Li[3], Xianfeng Chen[1], Claire Patenia[2], Athurva Gore[3], Emad B. Abboud[4], Ali A. Al-Rajhi[4], Richard A. Lewis[1,5,6], James R. Lupski[1,6,7], Graeme Mardon[1,7,8,9,10], Kun Zhang[3], Donna Muzny[1,2], Richard A. Gibbs[1,2], Rui Chen[1,2,10,*]

Article first published online: 23 SEP 2011

DOI: 10.1002/humu.21587

Issue

Human Mutation

Early View (Online Version of Record published before inclusion in an issue)


JBMR
Journal of Bone and Mineral Research

JOURNAL INFORMATION

Home
Aims and Scope
Editorial Board
News
Accepted Articles
Current Issue
All Issues
Virtual Issues

INFORMATION FOR CONTRIBUTORS

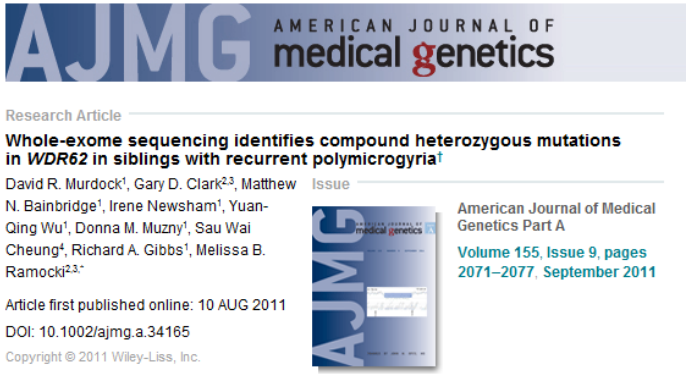
Home > Current Issue > Article

Mutations in *SERPINF1* cause Osteogenesis imperfecta Type VI

Erica P. Homan, Frank Rauch, Ingo Grafe, Caressa Lietman, Jennifer A. Doll, Brian Dawson, Terry Bertin, Dobrawa Napierala, Roy Morello, Richard Gibbs, Lisa White, Rika Miki, Daniel H. Cohn, Susan Crawford, Rose Travers, Francis H. Glorieux, Brendan Lee

Abstract

Osteogenesis imperfecta (OI) is a spectrum of genetic disorders characterized by bone fragility. It is caused by dominant mutations affecting the synthesis and/or structure of type I


AJMG
AMERICAN JOURNAL OF medical genetics

Research Article

Whole-exome sequencing identifies compound heterozygous mutations in *WDR62* in siblings with recurrent polymicrogyria†

David R. Murdock[1], Gary D. Clark[2,3], Matthew N. Bainbridge[1], Irene Newsham[1], Yuan-Qing Wu[1], Donna M. Muzny[1], Sau Wai Cheung[4], Richard A. Gibbs[1], Melissa B. Ramocki[2,3,*]

Article first published online: 10 AUG 2011

DOI: 10.1002/ajmg.a.34165

Issue

American Journal of Medical Genetics Part A

Volume 155, Issue 9, pages 2071–2077, September 2011

Mendelian Disease Score Card:

(Approximately; n~30):
10% ‘new’ genes
50% ‘retrospective ‘insight’
40% unsolved so far!

AJMG AMERICAN JOURNAL OF medical genetics

Invited Comment

Editorial comment on “Whole Exome Sequencing Identifies Compound Heterozygous Mutations in WDR62 in Siblings With Recurrent Polymicrogyria”†

Leslie G. Biesecker

Article first published online: 10 AUG 2011
DOI: 10.1002/ajmg.a.34183

Issue

American Journal of Medical Genetics Part A
Volume 155, Issue 9, pages 2069–2070, September 2011

Nice Editorial: Les Biesecker

1. Unambiguous
2. Less expensive
3. Tackles *de novo’s*
4. Better use of physician time…
5. ***SCALABLE!!!***

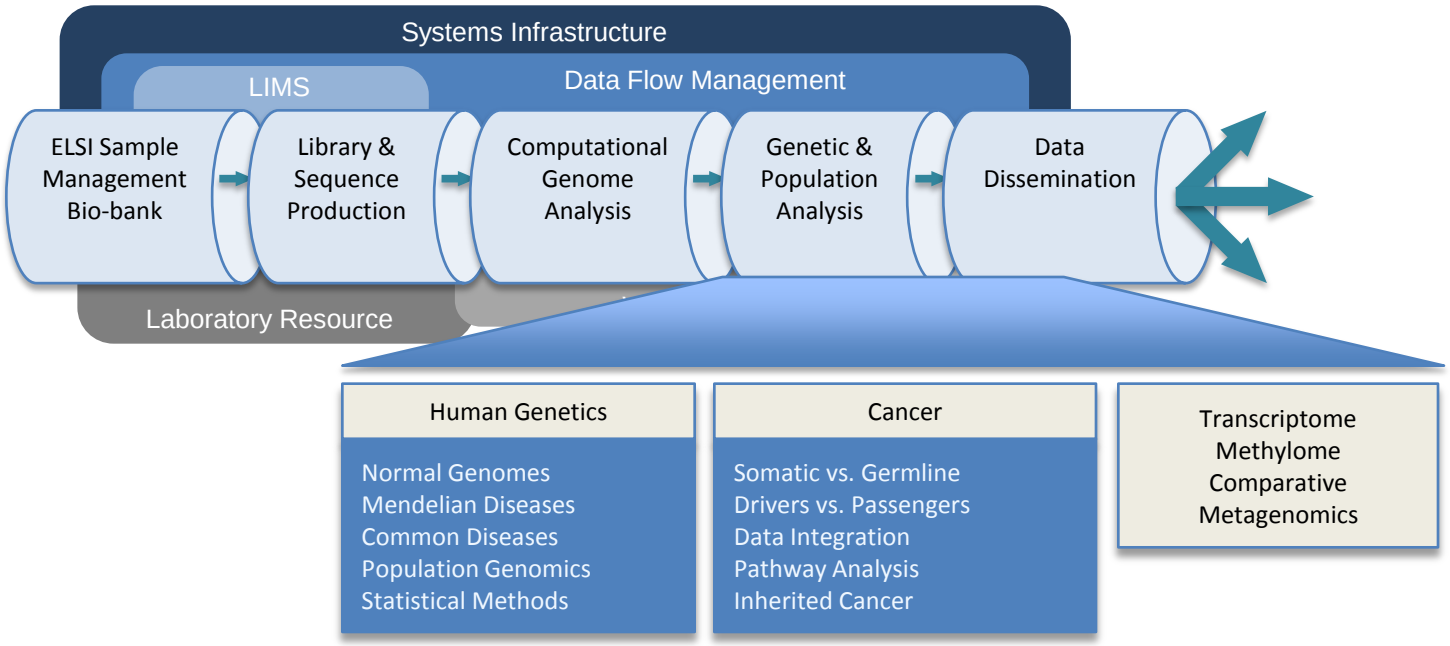
Conclusions from Genetic Discovery so far:

New methods make simple genetic problems easy to solve!!

AND

Family studies most tractable….

LESSON IV: Analysis Networks:
Many participants, group efforts:



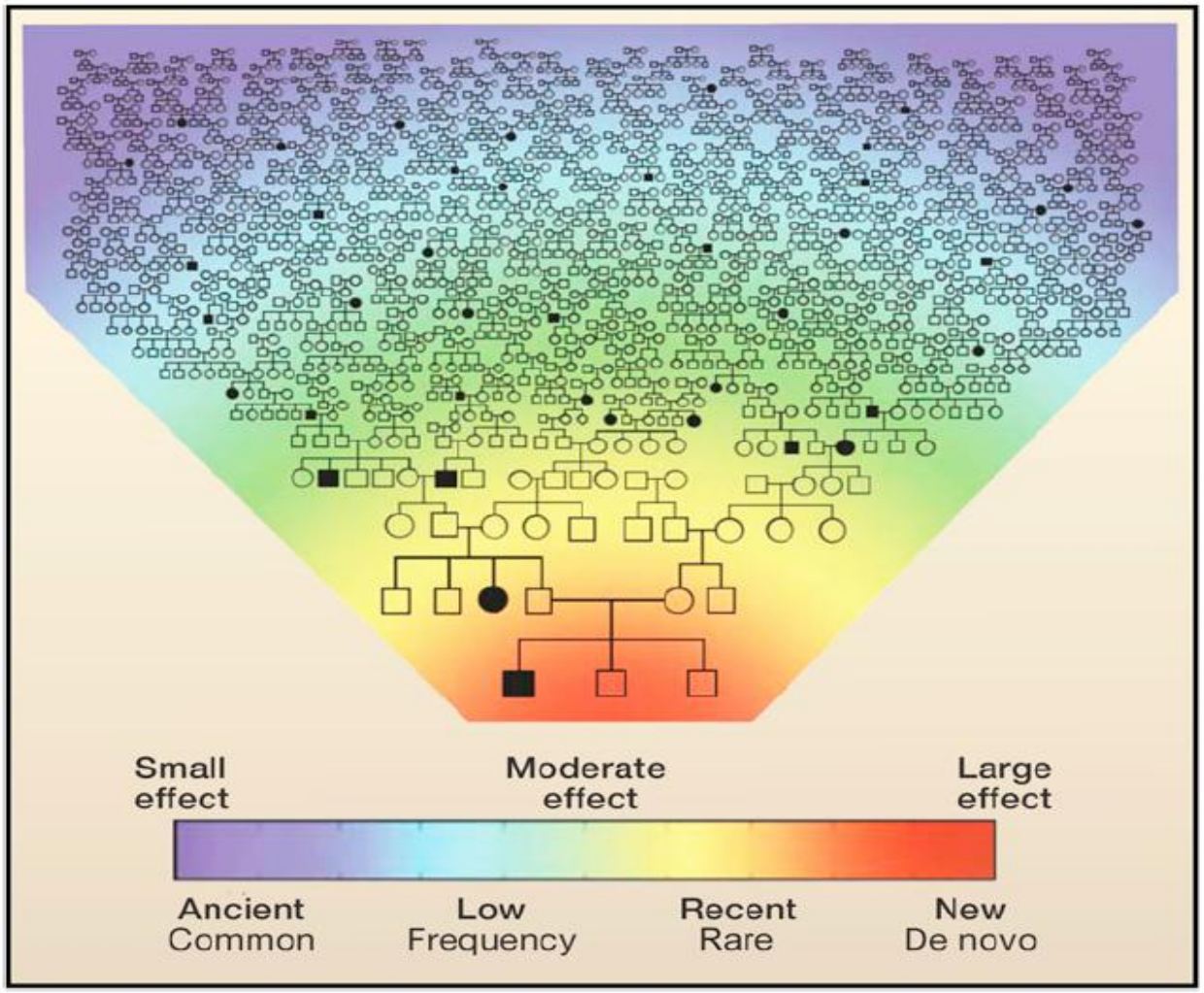
Dozens of conference calls, multiple centers and individuals, redundancy as well as single dedicated individuals.

e.g. the ‘1000 Phone Calls Project’

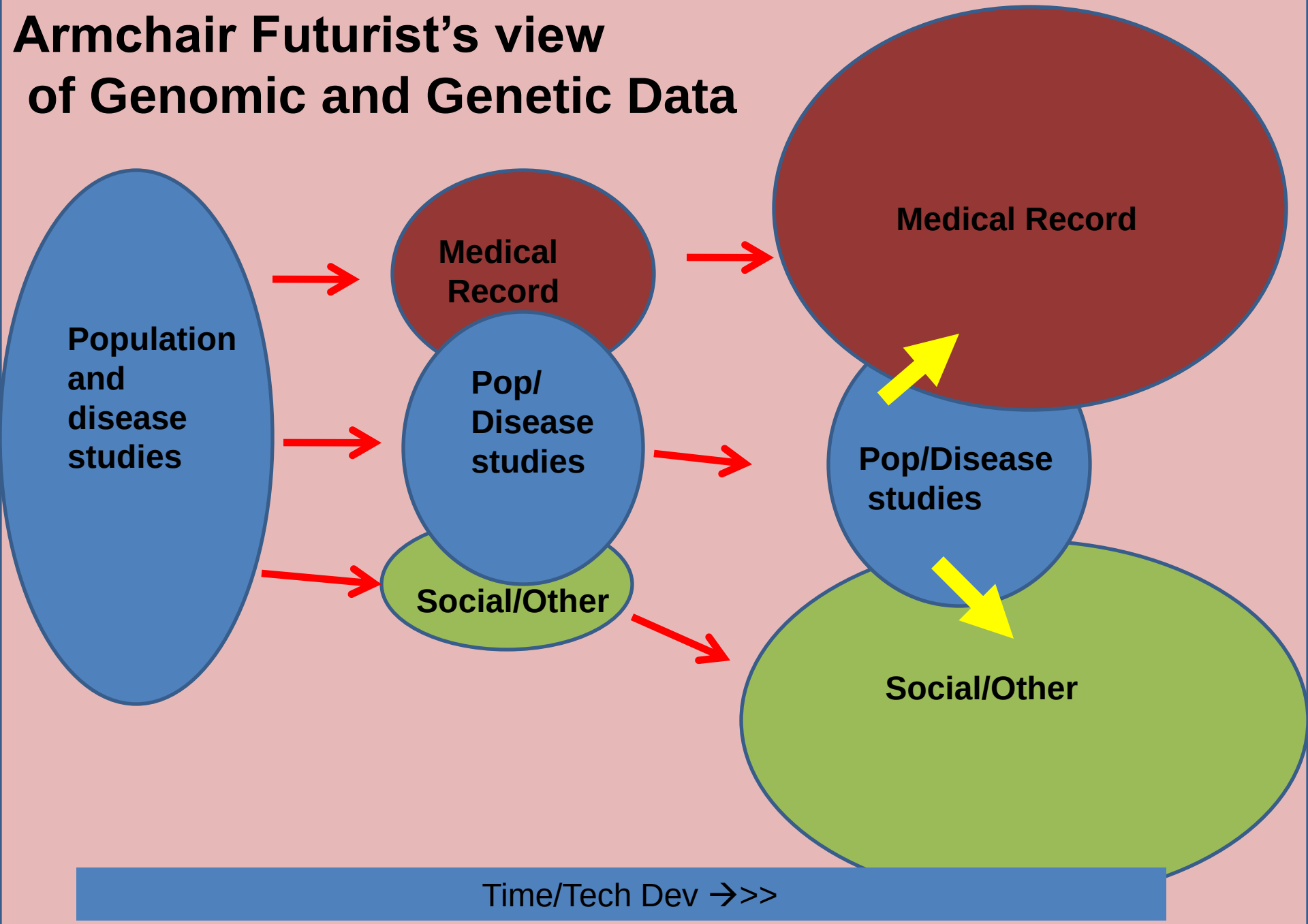
Approximately equal resources are needed post-data accumulation, as needed for sequencing.

LESSON V: UBIQUITY

Clan Genomics: Interest in Family Genetic Health will drive diagnostics:



Armchair Futurist's view of Genomic and Genetic Data



How might this project look:

Thought ‘project’: $5M:

- 5,000 genomes? / 20,000 exomes?

Programs:

1. Rare non-HbS cases?
2. Variable HU responders
3. Other ‘variant phenotype’ issues?

How might this project look:

Thought ‘project’: $5M:

- 5,000 genomes? / 20,000 exomes?

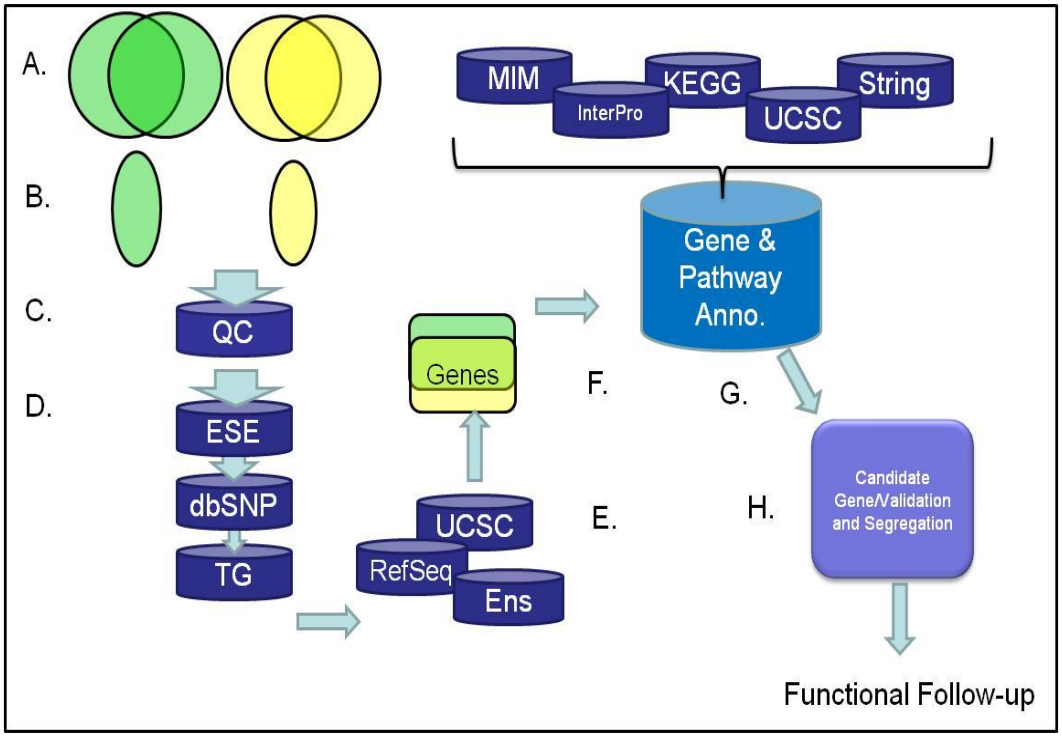
Programs:

1. Rare non-HbS cases? (25%)
2. Variable HU responders (25%?)
3. Other ‘variant phenotype’ issues? (50%)

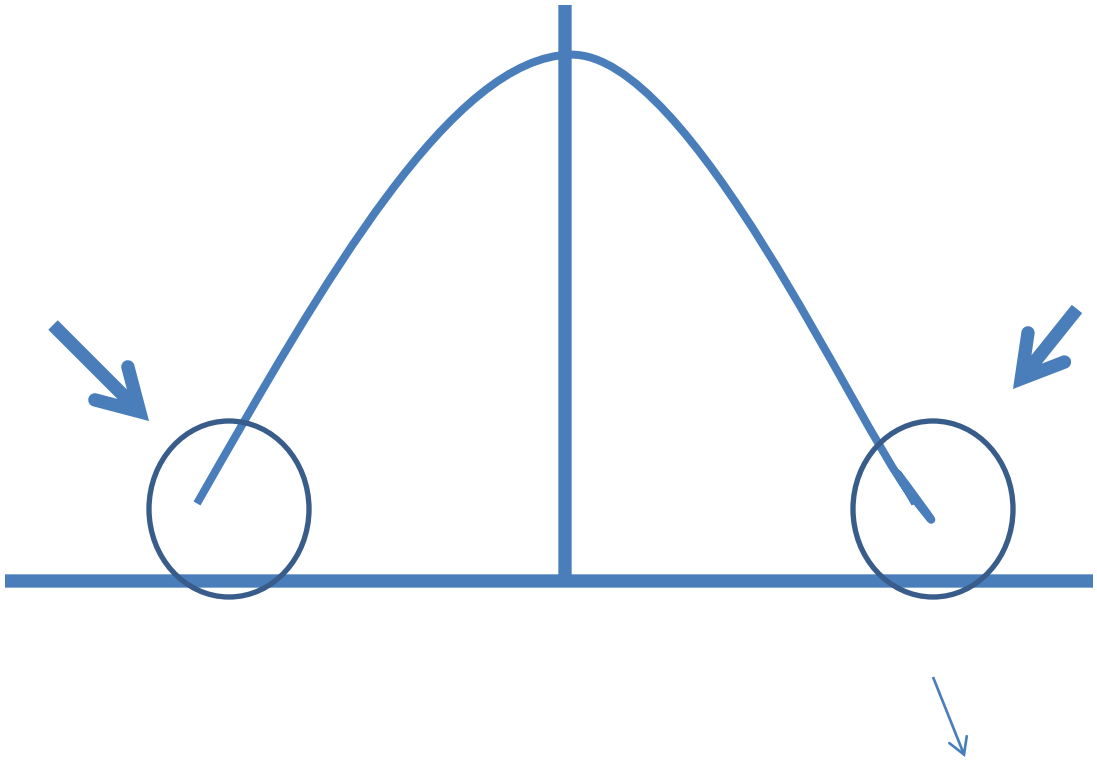
DESIGN:

1. Rare non-HbS cases? (25%)

 - Sequence probands, parents, sibs etc



DESIGN: Variable therapeutic response?



Whole exomes of low and high segments of distribution?

DESIGN: Other ‘variant phenotype’ issues?

- Multiple designs
- Phenotyping critical
- Family studies optional
- Power calculations needed (>1,000 cases)
- GWAS data to be considered

WORK PLAN:

- Complete design
- Identify samples, data generators, analysts
- Revisit ELSI/Consent issues
- Power calculations needed (>1,000 cases)
- GWAS data to be considered
- BUT……..
- Devise central data management/sharing protocols
- Establish management structure, mission milestones
- Manage timelines, integrate and respond to other efforts
- *etc*

For Meeting:

- Do we need a ‘centralized effort’
- Can we establish community-led ‘buy in’ to a mission objective.
- Will this compromise distributed science approach
- Is it the time to think BIG, challenge the *status quo?*

Acknowledgements

James Lupski*
Jeff Reid
David Deiros
Donna Muzny
Lynne Nazareth
Christie Kovar
Aniko Sabo
Matthew Bainbridge
Claudia Gonzaga-Jauregui
David Wheeler

Zogbhi Lab:
Huda Zoghbi
Christian Schaaf

454/James Watson –
Watson Project

LIfe
Linh Hoang*
Kevin McKernan
Francisco de la Vega
Tim Hunkapiller
Michael Rodes
Gina Costa
Quynh Dong
Fiona Hyland
Heather Peckham
Yutao Fu

All HGSC Staff

Micortex Team
Eric Boerwinkle
Charlie Sing
Aleks Coventry
Andy Clark
Lara Bull

ENCODE/HapMap3
Fuli Yu

Other HGSC
Steve Scherer
Lynne Nazareth
Mike Metzker
Rui Chen

MHG
David Nelson
Art Beaudet

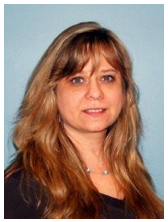
1000 Genomes
Lisa Brooks
Richard Durbin
David Altshuler
Etc

African Genomes
Stephan Schuster
Vanessa Hayes
Desmond Tutu

Sequencing Discovery
Huda Zoghbi
Jeff Noebels
Melissa Ramocki
Rui Chen
Shalini Jhangiani

ARRA AUTISM
Joe Buxbaum
Bernie Devlin
Jerry Schellenberg
Jim Sutcliffe
Kathryn Roeder
Benjamin Neale
Eric Boerwinkle
Aniko Sabo


Jeff Reid

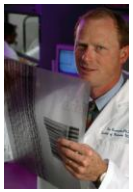

Donna Muzny


Yuan-Qing Wu


Fuli Yu


Aniko Sabo


Lisa Trevino


Eric Boerwinkle


Claudia Gonzaga-Jauregui


James Lupski