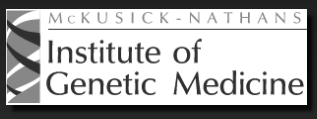
The Search for Mendelian Disease Genes: Opportunities afforded; lessons learned

David Valle, MD

McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University
30 April 2014



dvalle@jhmi.edu

Current Topics in Genome Analysis 2014

David Valle, M.D.

No Relevant Financial Relationships with Commercial Interests

CME Disclosures

- I am fired up about genetics !

The Fascination:

Gregor Mendel (1822 – 1884)

- Fascination with variation and its relationship to the heritable material
- Making "the connection"

⟷ "factors"

The Excitement:

TH Morgan and the origins of fly genetics

SPECIAL ARTICLES

SEX LIMITED INHERITANCE IN DROSOPHILA

In a pedigree culture of *Drosophila* which had been running for nearly a year through a considerable number of generations, a male appeared with white eyes. The normal flies have brilliant red eyes.....

T. H. MORGAN

WOODS HOLE, MASS.,
July 7, 1910

Science, 32: 120, 1910

The Burden of Mendelian Disorders

- About 8,000 Mendelian disorders now known
- Inheritance
  - ✓ ~ 65% autosomal dominant
  - ✓ ~ 30% autosomal recessive
  - ✓ ~ 6% X-linked disorders
- Most present in the pediatric age range
- Incidence ~0.4% liveborn infants
- Account for 6 - 10% hospitalized children

Pre-Human Genome Project: Disease gene identification a slow process

Nucleotide Sequence of a Full-Length Complementary DNA Clone and Amino Acid Sequence of Human Phenylalanine Hydroxylase†

Simon C. M. Kwok, Fred D. Ledley, Anthony G. DiLella, Kathryn J. H. Robson,‡ and Savio L. C. Woo*

Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030
Received November 13, 1984

Biochemistry 24: 556, 1985

cDNA Cloning of the Type 1 Neurofibromatosis Gene: Complete Sequence of the *NF1* Gene Product

DOUGLAS A. MARCHUK, ANN M. SAULINO, ROXANNE TAVAKKOL, MANJU SWAROOP, MARGARET R. WALLACE, LONE B. ANDERSEN, ANNA L. MITCHELL, DAVID H. GUTMANN, MARK BOGUSKI,* AND FRANCIS S. COLLINS

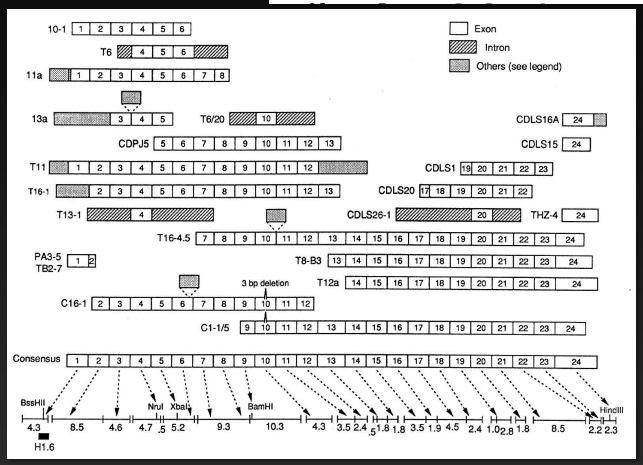
Genomics 11: 931, 1991

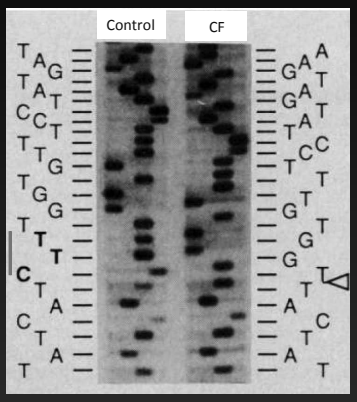
2-3 years/ disease gene

CFTR as an example

Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA

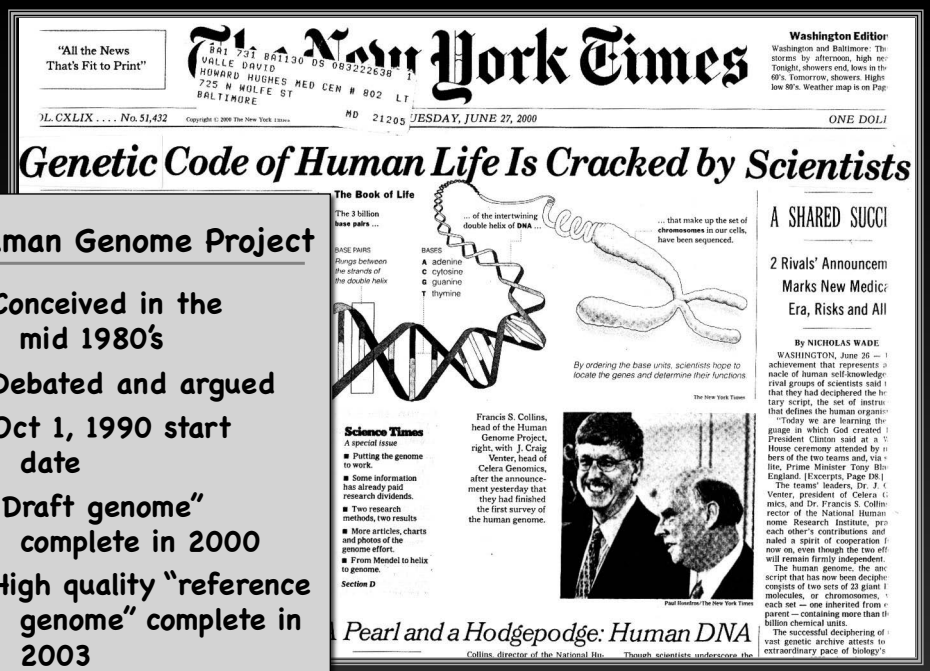
JOHN R. RIORDAN, JOHANNA M. ROMMENS, BAT-SHEVA KEREM, NOA ALON, RICHARD ROZMAHEL, ZBYSZKO GRZELCZAK, JULIAN ZIELENSKI, SI LOK, NATASA PLAVSIC, JIA-LING CHOU, MITCHELL L. DRUMM, MICHAEL C. IANNUZZI, FRANCIS S. COLLINS, LAP-CHEE TSUI Science 245: 1066, 1989





Human Genome Project

- Conceived in the mid 1980's
- Debated and argued
- Oct 1, 1990 start date
- "Draft genome" complete in 2000
- High quality "reference genome" complete in 2003



"All the News That's Fit to Print"

The New York Times

Washington Edition

VOL. CXLIX .... No. 51,432 TUESDAY, JUNE 27, 2000 ONE DOLLAR

Genetic Code of Human Life Is Cracked by Scientists

The Book of Life

The 3 billion base pairs ...

... of the intertwining double helix of DNA ...

... that make up the set of chromosomes in our cells, have been sequenced.

BASE PAIRS
Rungs between the strands of the double helix

BASES
A adenine
C cytosine
G guanine
T thymine

By ordering the base units, scientists hope to locate the genes and determine their functions.

The New York Times

Science Times
A special issue
- Putting the genome to work.
- Some information has already paid research dividends.
- Two research methods, two results
- More articles, charts and photos of the genome effort.
- From Mendel to helix to genome.

Section D

Francis S. Collins, head of the Human Genome Project, right, with J. Craig Venter, head of Celera Genomics, after the announcement yesterday that they had finished the first survey of the human genome.

Paul Sancya/The New York Times

A SHARED SUCCI

2 Rivals' Announcem Marks New Medica Era, Risks and All

By NICHOLAS WADE

WASHINGTON, June 26 —

"Today we are learning the guage in which God created

President Clinton said at a

House ceremony attended by

bers of the two teams and, via

lite, Prime Minister Tony Bla

England. [Excerpts, Page D8.]

The teams' leaders, Dr. J.

Venter, president of Celera

mics, and Dr. Francis S. Collin

rector of the National Human

nome Research Institute, pra

each other's contributions and

naled a spirit of cooperation

now on, even though the two eff

will remain firmly independent.

The human genome, the anc

script that has now been decipher

consists of two sets of 23 giant

molecules, or chromosomes,

each set — one inherited from

parent — containing more than

billion chemical units.

The successful deciphering of

vast genetic archive attests to

extraordinary pace of biology's

A Pearl and a Hodgepodge: Human DNA

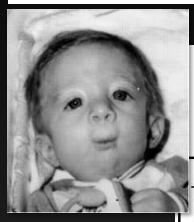
Collins, director of the National Hu... Though scientists underscore the

An inconvenient reality: each of us varies from the "reference sequence" by about 3 million SNPs plus a large and variable number of CNVs; how can we filter the neutral variants away to find the causative variant??

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[4], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

Nat Genet Sept 09



Freeman- Sheldon syndrome mutations in *MYH3*

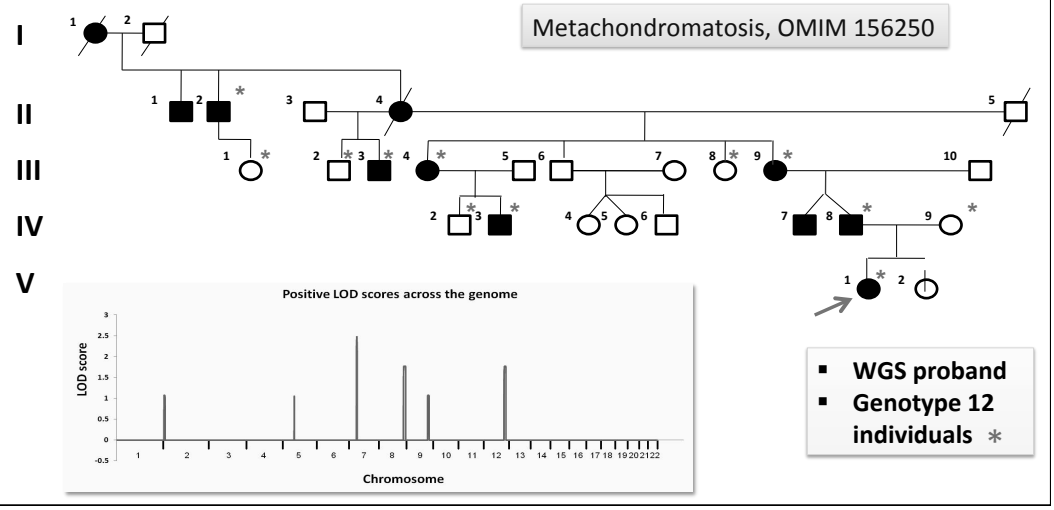
		FSS24895	FSS24895 FSS10208	FSS24895 FSS10208 FSS10066	FSS24895 FSS10208 FSS10066 FSS22194	Any 3 of 4: FSS24895 FSS10208 FSS10066 FSS22194
Number of genes in which each affected has at least one…	Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)	4,510	3,284	2,765	2,479	3,768
	NS/SS/I not in dbSNP	513	128	71	53	119
	NS/SS/I not in eight HapMap exomes	799	168	53	21	160
	Both NS/SS/I neither in dbSNP nor eight HapMap exomes	360	38	8	1 (*MYH3*)	22
	…And predicted to be damaging	160	10	2	1 (*MYH3*)	3

NS, ns cSNP; SS, splice site variant; I, indel

Whole-Genome Sequencing of a Single Proband Together with Linkage Analysis Identifies a Mendelian Disease Gene

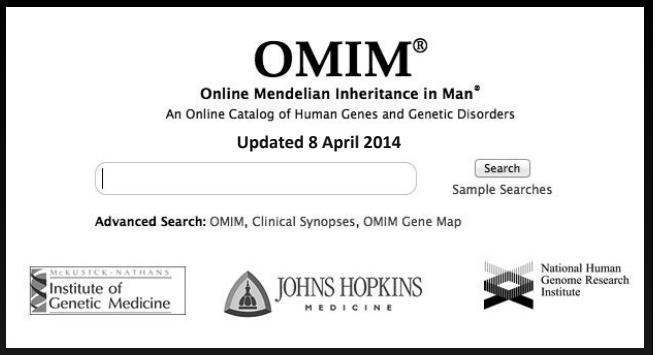
Nara L. M. Sobreira[1,2¶], Elizabeth T. Cirulli[3¶], Dimitrios Avramopoulos[1,4¶], Elizabeth Wohler[5], Gretchen L. Oswald[1], Eric L. Stevens[1,2], Dongliang Ge[3], Kevin V. Shianna[3], Jason P. Smith[3], Jessica M. Maia[3], Curtis E. Gumbs[3], Jonathan Pevsner[6,7], George Thomas[1,5], David Valle[1,8¶], Julie E. Hoover-Fong[1,8,9¶], David B. Goldstein[2¶*]

PLoS Genetics 6:1, 2010



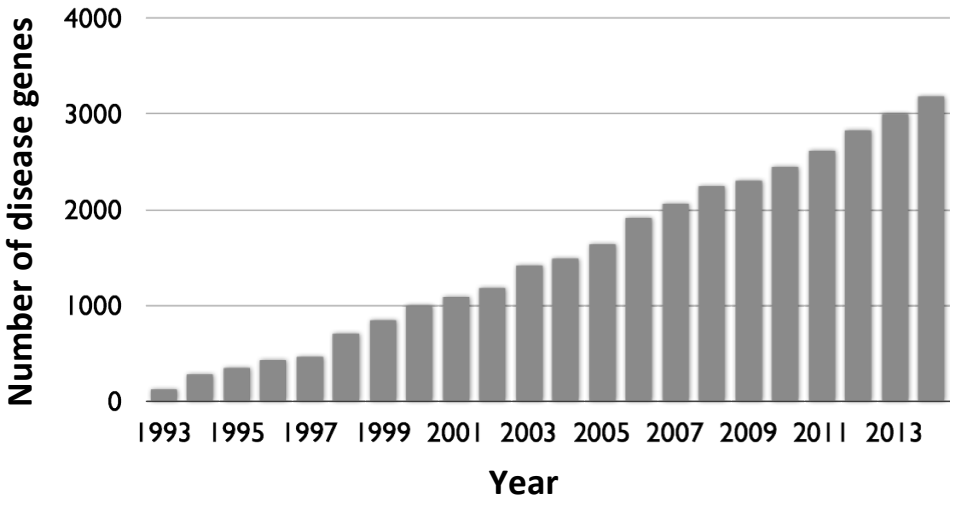
Current status

OMIM.org

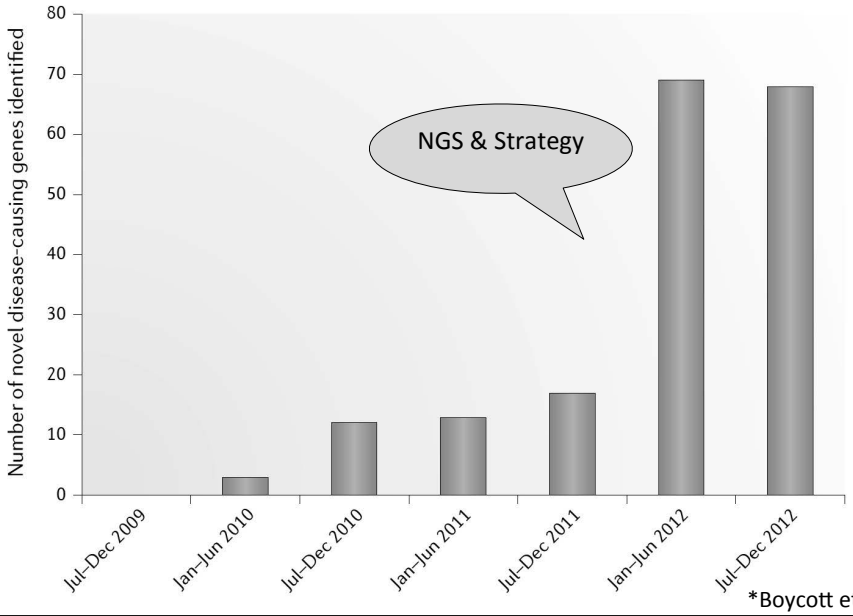


- Mendelian phenotypes -- ~8,700
- Disease genes -- 3,181 (~14% of total)
- Explained phenotypes --- 5,181
- Unexplained phenotypes --- ~3,600

Increase in Identified Disease Genes over Time



*Increase in Identified Disease Genes over Time**

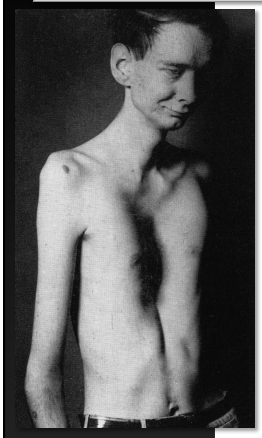


*Boycott et al, NRG, 2013

Finding disease genes: Some immediate consequences

- Connects genes to phenotypes
- Connects phenotype to biological system, normal and perturbed
- Unravels locus heterogeneity
- Enables precise diagnosis and counseling
- First step in path towards informed treatment
- Research stimulus, bench to bedside

Marfan syndrome and related disorders
Hal Dietz and colleagues



Mendelian diseases predict drugable nodes in biologic systems

DISORDER	GENE	Rx	Diseases
FH	*LDLR*	Statins	Common varieties of hypercholesterolemia
Marfan syndrome	*FBN1*	Losartan	MFS, sarcopenia, etc
Familial amyloidosis	*TTR*	Tafamidis	Other disorders of protein folding?
CF	*CFTR*	Kalydeco VX809	CF, other disorders of protein folding?

Discovery of a selective $Na_V1.7$ inhibitor from centipede venom with analgesic efficacy exceeding morphine in rodent pain models PNAS 2013

Shilong Yang[a,b,1], Yao Xiao[a,b,1], Di Kang[a,b,1], Jie Liu[a,b], Yuan Li[a,b], Eivind A. B. Undheim[c], Julie K. Klint[c], Mingqiang Rong[a,2], Ren Lai[a,2], and Glenn F. King[c,2]

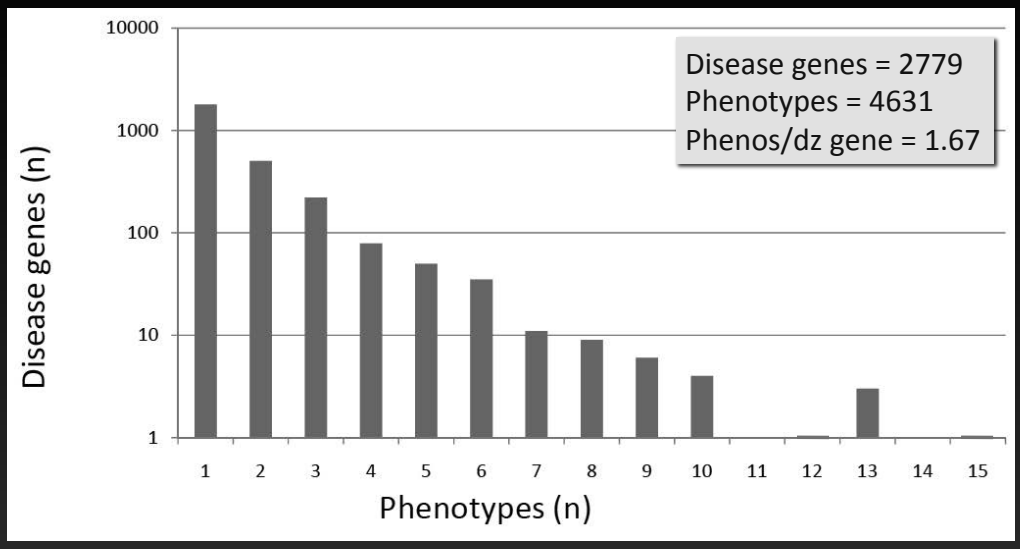
Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes

Jason Flannick[1–3], Gudmar Thorleifsson[4], Nicola L Beer[1,5], Suzanne B R Jacobs[1], Niels Grarup[6], Noël P Burtt[1], Anubha Mahajan[7], Christian Fuchsberger[8], Gil Atzmon[9,10], Rafn Benediktsson[11], John Blangero[12], Don W Bowden[13–16], Ivan Brandslund[17,18], Julia Brosnan[19], Frank Burslem[20], John Chambers[21–23], Yoon Shin Cho[24], Cramer Christensen[25], Desirée A Douglas[26], Ravindranath Duggirala[12], Zachary Dymek[1], Yossi Farjoun[1], Timothy Fennell[1], Pierre Fontanillas[1], Tom Forsén[27,28], Stacey Gabriel[1], Benjamin Glaser[29,30], Daniel F Gudbjartsson[4], Craig Hanis[31], Torben Hansen[6,32], Astradur B Hreidarsson[11], Kristian Hveem[33], Erik Ingelsson[7,34], Bo Isomaa[35,36], Stefan Johansson[37–39], Torben Jørgensen[40–42], Marit Eika Jørgensen[43], Sekar Kathiresan[1,44–46], Augustine Kong[4], Jaspal Kooner[22,23,47], Jasmina Kravic[48], Markku Laakso[49], Jong-Young Lee[50], Lars Lind[51], Cecilia M Lindgren[1,7], Allan Linneberg[40,41,52], Gisli Masson[4], Thomas Meitinger[53], Karen L Mohlke[54], Anders Molven[37,55,56], Andrew P Morris[7,57], Shobha Potluri[58], Rainer Rauramaa[59,60], Rasmus Ribel-Madsen[6], Ann-Marie Richard[19], Tim Rolph[19], Veikko Salomaa[61], Ayellet V Segrè[1,2], Hanna Skärstrand[26], Valgerdur Steinthorsdottir[4], Heather M Stringham[8], Patrick Sulem[4], E Shyong Tai[62–64], Yik Ying Teo[62,65–68], Tanya Teslovich[8], Unnur Thorsteinsdottir[4,69], Jeff K Trimmer[19], Tiinamaija Tuomi[27,35], Jaakko Tuomilehto[70–72], Fariba Vaziri-Sani[26], Benjamin F Voight[1,73,74], James G Wilson[75], Michael Boehnke[8], Mark I McCarthy[5,7,76], Pål R Njølstad[1,37,77], Oluf Pedersen[6], Go-T2D Consortium[78], T2D-GENES Consortium[78], Leif Groop[48,79], David R Cox[58], Kari Stefansson[4,69] & David Altshuler[1–3,44,45,80,81]

Nat Genet 2 March 2014

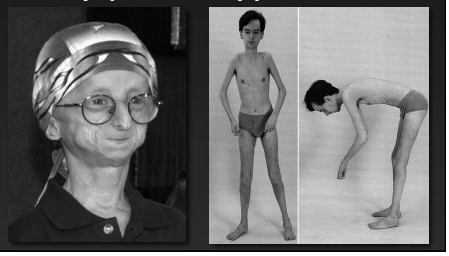
What questions could we ask if we had phenotypes for > 50% of our gene complement ??

Genes and phenotypes:



Genes and phenotypes:

- One gene / many variants / one phenotype?
  - ✓ Many inborn errors of metabolism
  - ✓ Is this biology or more precise diagnostic methods?
- One gene / many variants / many phenotypes?
  - ✓ *LAMA with 13 "discrete" phenotypes*



Marfan Syndrome and Stiff Skin Syndrome: Allelic Disorders at FBN1 locus

- Tall stature
- Long limbs
- Joint hyper-extensibility
- *FBN1* LOF muts

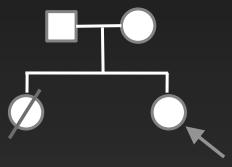
- Short stature
- Progressive fibrosis
- Joint contractures
- *FBN1* missense muts in 4th 8 Cys domain

Loeys, Gerber et al Sci Trans Med, 2010

Phenotypic "expansion"

- Adding additional features to a known phenotype or adding additional phenotypes to a known disease gene
- History shows we find the phenotypes that we know
- New technologies, new ways of looking expands the "classical" phenotype
- New understanding from a gene based view

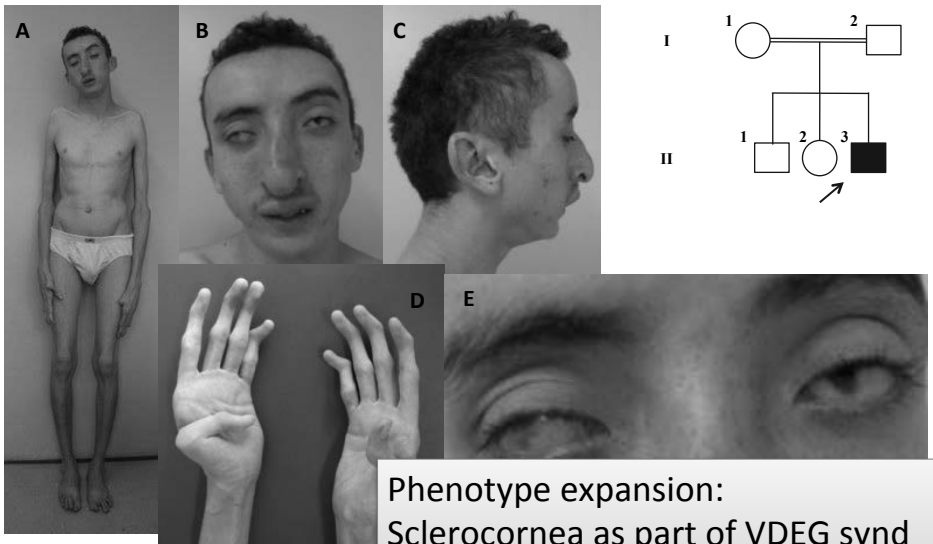
Expanding phenotypes



- 2 mo female with dilated cardiomyopathy
- FHx positive for sib who died of DCM at 8 mo
- Neg molecular work up for DCM panel
- Cardiac transplant at 7 mo
- WES – cpd het for *ALMS1* LOF muts

Phenotype expansion:
early cardiomyopathy in Alstrom synd

Sclerocornea and the van den Ende-Gupta syndrome (OMIM 600920)



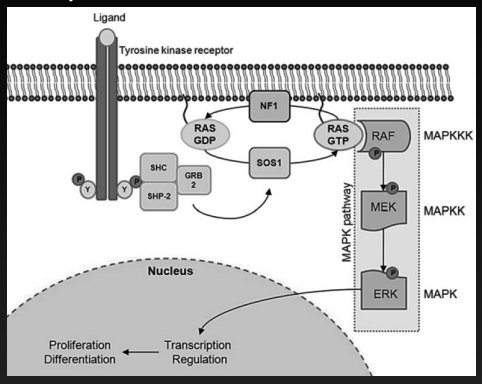
Phenotype expansion:
Sclerocornea as part of VDEG synd

Biological networks and disease: some questions

- Are all networks equally vulnerable; if not what are the rules?
- Are all components of a system equally vulnerable; if not what are the rules?
- Can we predict the consequences of variation in one component on the behavior of a system?

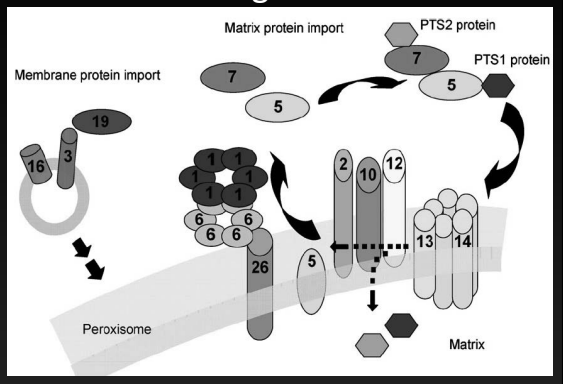
Are all components equally vulnerable?

Ras/MAPK



~30 genes
>15 phenotypes
No one gene predominates

Peroxisome biogenesis



~30 genes
1-2 phenotypes
~65% *PEX1*

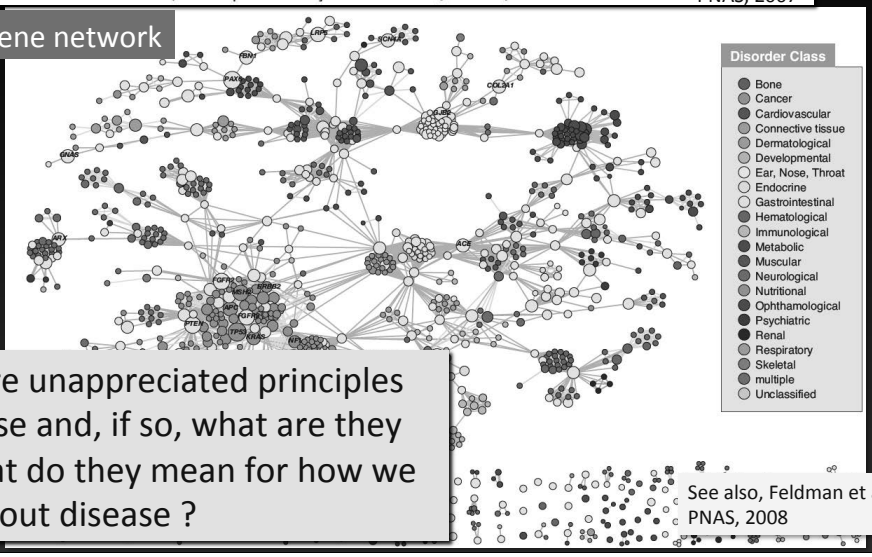
The human disease network

Kwang-Il Goh*†‡§, Michael E. Cusick†‡¶, David Valle‖, Barton Childs‖, Marc Vidal†‡¶**, and Albert-László Barabási*†‡**

*Center for Complex Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556; †Center for Cancer Systems Biology (CCSB) and ¶Department of Cancer Biology, Dana–Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; ‡Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115; §Department of Physics, Korea University, Seoul 136-713, Korea; and ‖Department of Pediatrics and the McKusick–Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205

PNAS, 2007

Disease gene network



Are there unappreciated principles of disease and, if so, what are they and what do they mean for how we think about disease ?

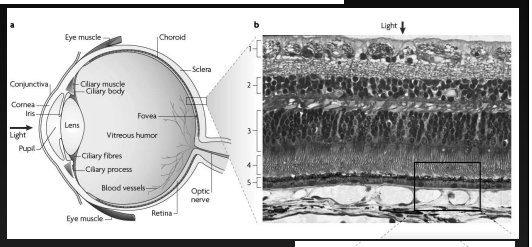
See also, Feldman et al, PNAS, 2008

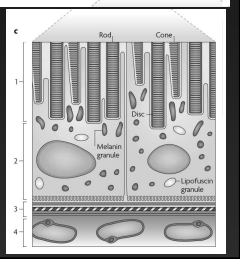
Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait

Nat Rev Genet, 11: 273, 2010

Alan F. Wright, Christina F. Chakarova‡, Mai M. Abd El-Aziz‡ and Shomi S. Bhattacharya‡*

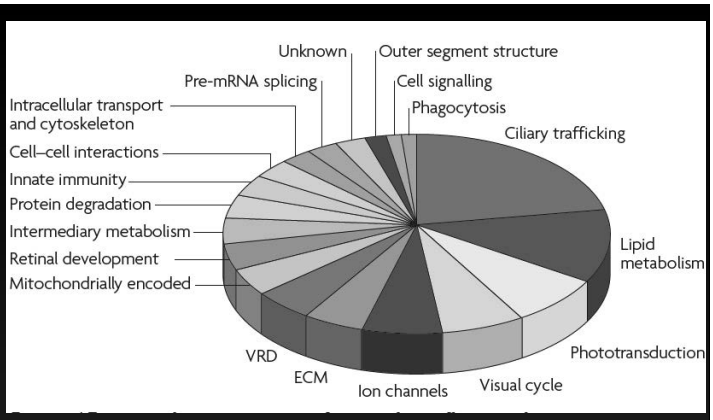
- Defined system
- Easily recognized phenotype – loss of visual function secondary to loss +/or dysfunction of photoreceptors
- 184 loci, 146 identified at the time of this review





*Cellular function, 146 PR degen genes **

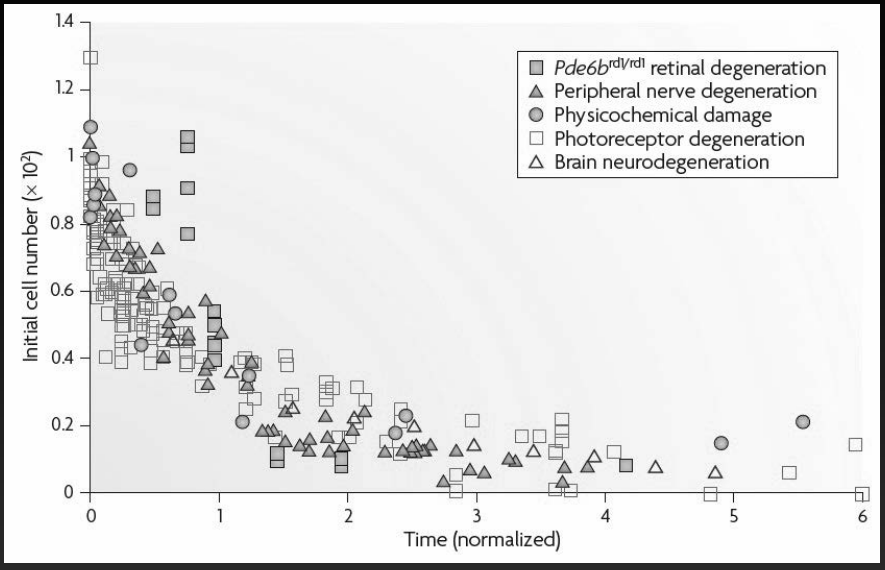
- Encoded proteins involved in nearly all aspects of cellular function



- Most show widespread rather than PR specific expression
- Most, but not all, Mendelian alleles rare (MAF < 0.01)
- The locus and allelic spectra quite different for Mendelian vs. complex trait (AMD) forms

* Wright et al., 2010

Photoreceptor death follows an exponential decay model



Wright et al., 2010

Cancer Genome Landscapes

Bert Vogelstein, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz Jr., Kenneth W. Kinzler*

Science 339: 1546, 2013

- 3,284 tumors WES sequenced
- 18, 306 genes with somatic muts
- 125 "driver" mutations: 71 tumor suppressors, 54 oncogenes
- Only 12 pathways regulating 3 cellular processes

Contribution of rare Mendelian disease to common complex traits

- + FHx/linkage
- Early onset
- Extremes of distributions

Linkage Studies in a Large Kindred with Familial Hypercholesterolemia

Jurg Ott,[1] Helmut G. Schrott,[1, 2] Joseph L. Goldstein,[1, 3] William R. Hazzard,[4] F. H. Allen, Jr.,[5] C. T. Falk,[5] and Arno G. Motulsky[6]

AJHG, 1974

Hyperlipidemia in Coronary Heart Disease

I. LIPID LEVELS IN 500 SURVIVORS OF MYOCARDIAL INFARCTION

Joseph L. Goldstein, William R. Hazzard, Helmut G. Schrott, Edwin L. Bierman, and Arno G. Motulsky with the assistance of Mary Jo Levinski and Ellen D. Campbell

JCI, 1973

Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol

Jonathan C. Cohen,[1,2,3*†] Robert S. Kiss,[5*] Alexander Pertsemlidis,[1] Yves L. Marcel,[5†] Ruth McPherson,[5] Helen H. Hobbs[1,3,4]

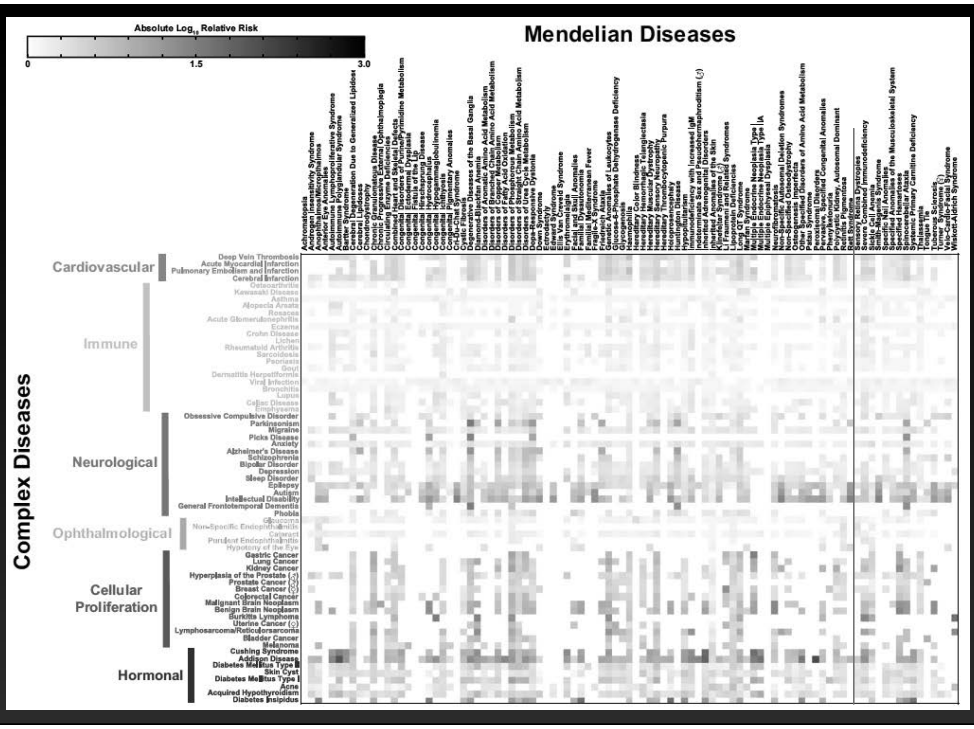
Science, 2004

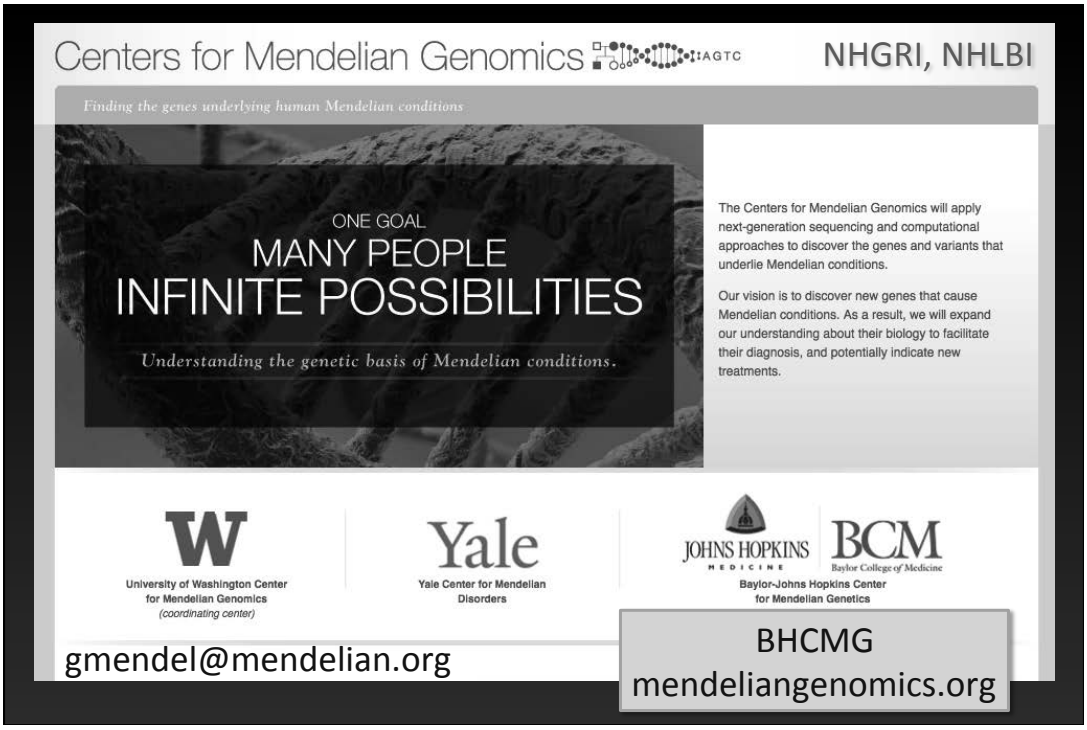
A Nondegenerate Code of Deleterious Variants in Mendelian Loci Contributes to Complex Disease Risk

CELL, 2013

David R. Blair,[1] Christopher S. Lyttle,[2] Jonathan M. Mortensen,[7] Charles F. Bearden,[8] Anders Boeck Jensen,[9] Hossein Khiabanian,[10] Rachel Melamed,[10] Raul Rabadan,[10] Elmer V. Bernstam,[8] Søren Brunak,[9,11] Lars Juhl Jensen,[9,11] Dan Nicolae,[3,4,5] Nigam H. Shah,[7] Robert L. Grossman,[4,6] Nancy J. Cox,[4,5] Kevin P. White,[4,5,6,*] and Andrey Rzhetsky[4,5,6,*]

- Surveyed 110 M medical records looking for connections between Mendelian disorders and complex traits
- Do genes responsible for Mendelian disorders interact with one another to contribute to common complex traits?

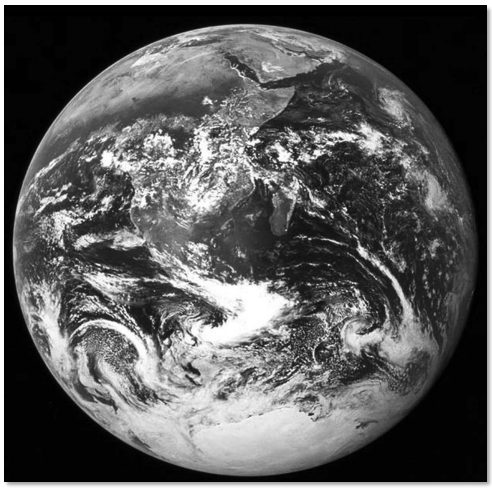


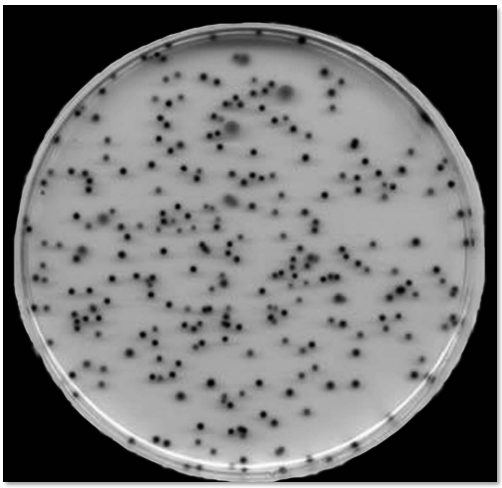


CMGs: Some goals

- Identify Mendelian phenotypes associated by >50% of our genes
- Improve diagnosis; inform pathophysiology
- Increase opportunities for informed treatment
- Increase our understanding of disease principles
- Education

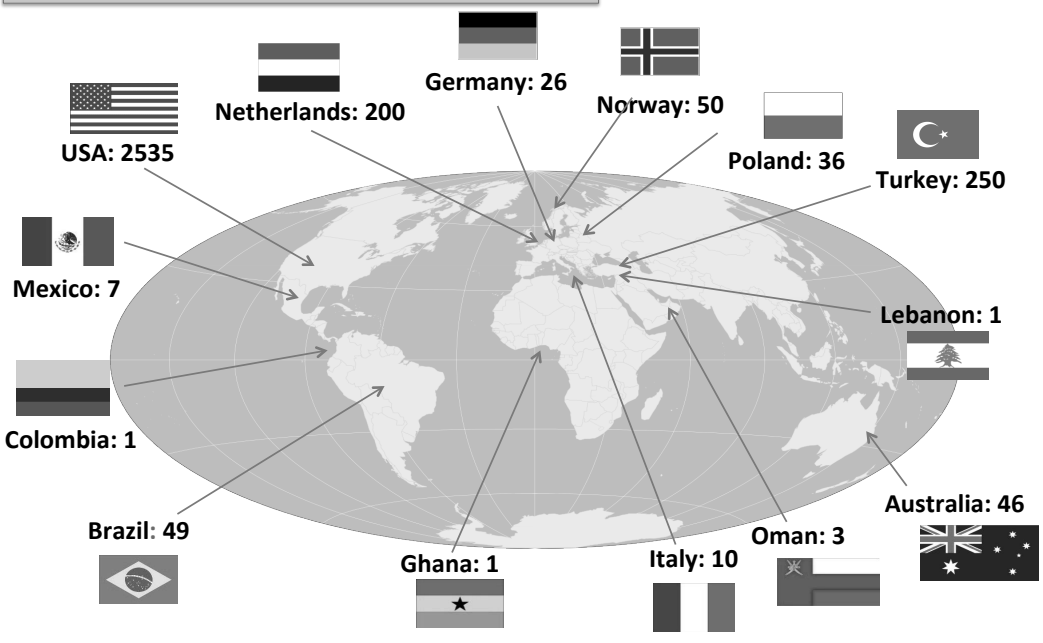
CMGs: Searching for disease genes





BHCMG: Sample recruitment

March 2013, N = 3199 total



CMGs: Overall strategy

- Find well-phenotyped cases and families
- Perform whole exome sequencing on relevant family members
- Use family relationships, allele frequency data, functional predictions, model organism results and functional studies to identify the responsible genes and variants
- Return the information to submitter for publication

BHCMG
mendeliangenomics.org

PhenoDB: A New Web-Based Tool for the Collection, Storage, and Analysis of Phenotypic Features

Ada Hamosh,[1]* Nara Sobreira,[1] Julie Hoover-Fong,[1] V. Reid Sutton,[2] Corinne Boehm,[1] François Schiettecatte,[3] and David Valle[1]

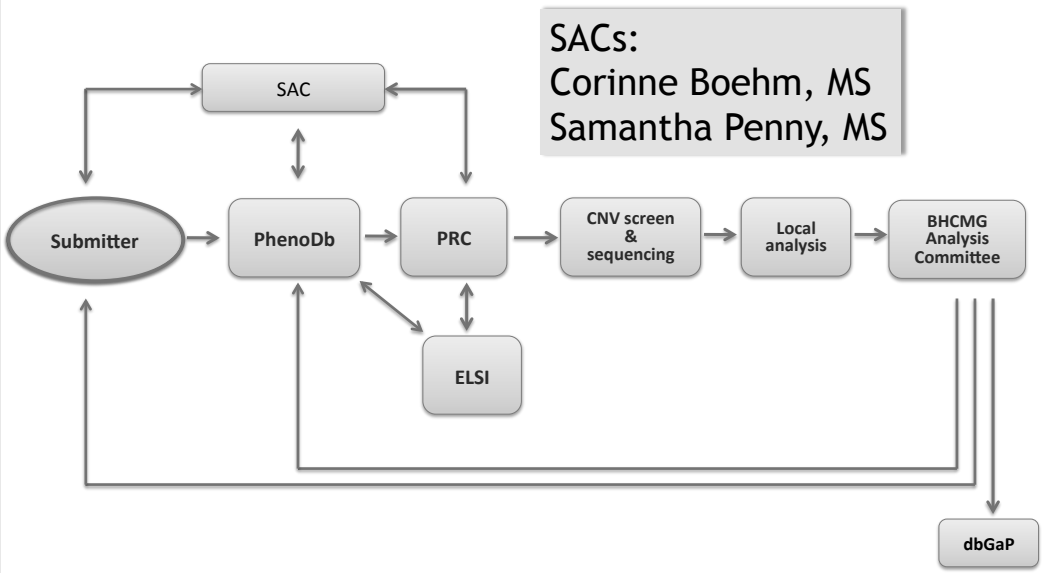
[1]McKusick-Nathans Institute of Genetic Medicine Johns Hopkins University, Baltimore, Maryland; [2]Department of Molecular & Human Genetics Baylor College of Medicine, Houston, Texas; [3]FS Consulting, Salem, Massachusetts

Hum Mut 34:561, 2013

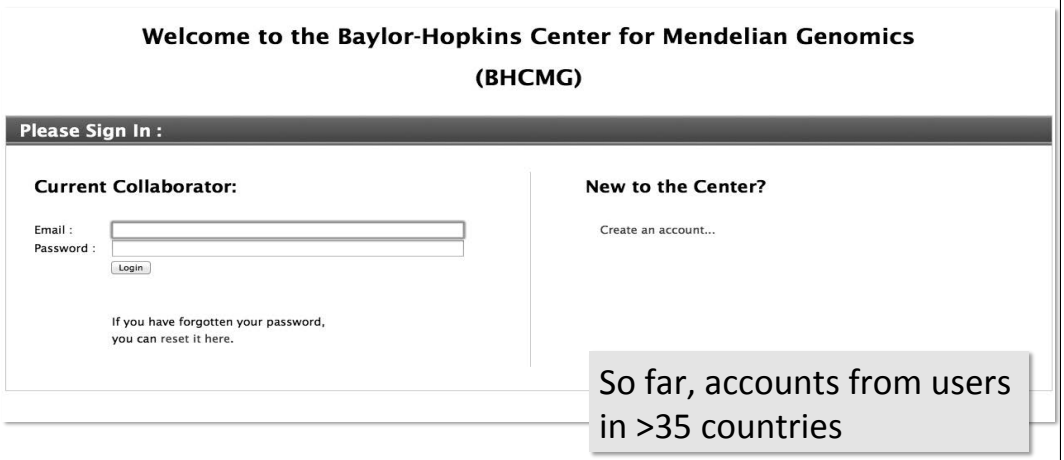
- Rapid and efficient entry of families or cohorts
- Provides unique identifiers
- Clinical features based on OMIM Clinical Synopses
- Accepts image data
- Searchable
- Organizes phenotypic features in standard format for easy review
- Added and Analysis module

mendeliangenomics.org

BHCMG: Pipeline



BHCMG: PhenoDb – Submitter view



So far, accounts from users in >35 countries

mendeliangenomics.org

Analysis

- A work in progress
- Requires individual attention, family by family
- Success depends on mode of inheritance, samples available and other variables
- Deeper and smarter levels of data mining will improve success rate
- Analysis tool now built for PhenoDB (Sobreira)

Analysis example

- CIDR Anovar file on an ~51 MB capture
- Filtering variants: AR, cpd het model, proband only
  - ✓ All - ~85 K
  - ✓ Only heterozygotes ~54 K
  - ✓ Coding & Splice Sites ~12 K
  - ✓ Exclude synonymous ~6 K
  - ✓ Exclude DbSNP 126, 129, 131 ~750
  - ✓ Exclude MAF $\geq$ 0.01 in EVS, 1KG ~450
  - ✓ Exclude CIDR db ~260
  - ✓ Genes with 2 hits ~20
  - ✓ Links to OMIM, MGI, PubMed, Expression, Networks

Mendelian disease: Burden of proof

- Value of another unrelated case
  - ✓ Identify another family
  - ✓ Literature
- Statistical approaches – 1KG, dbSNP, EVS, Genic tolerance, etc…..
- Biological studies
- Variant may perturb function of protein product but does this really explain the phenotype?

BHCG: Two year scorecard

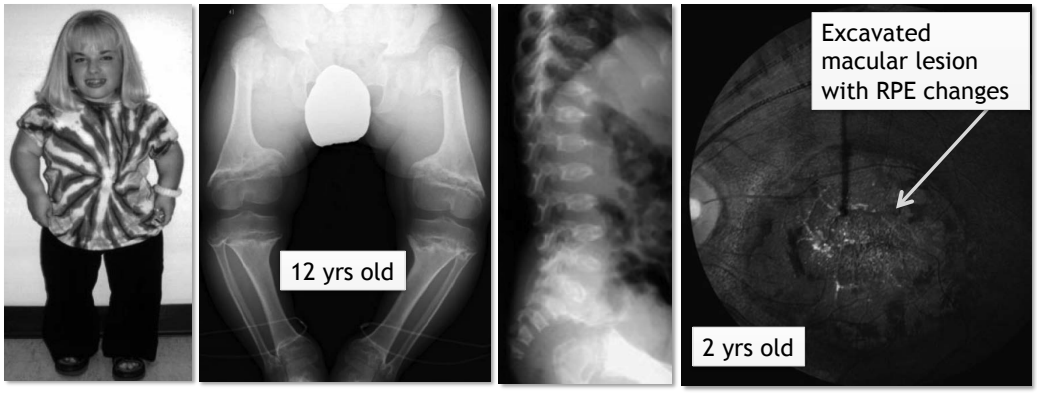
- 4326 samples collected
- 609 probands with complete analysis
- Newly recognized disease genes 189
- Known disease genes 121
- Known disease genes, phenotype expansion 85

Explaining Principles of Mendelian disease

- Pleiotropy – one variant, multiple, apparently unrelated, phenotypic consequences
- Penetrance – the probability that someone with the genotype will manifest the phenotype
- Variable expressivity – affected individuals with same genotype show different manifestations

Predictive power of Mendelian disease

- Spondylometaphyseal dysplasia – cone/rod dystrophy
  - » Postnatal short stature and loss of visual function
  - » Rare autosomal recessive trait

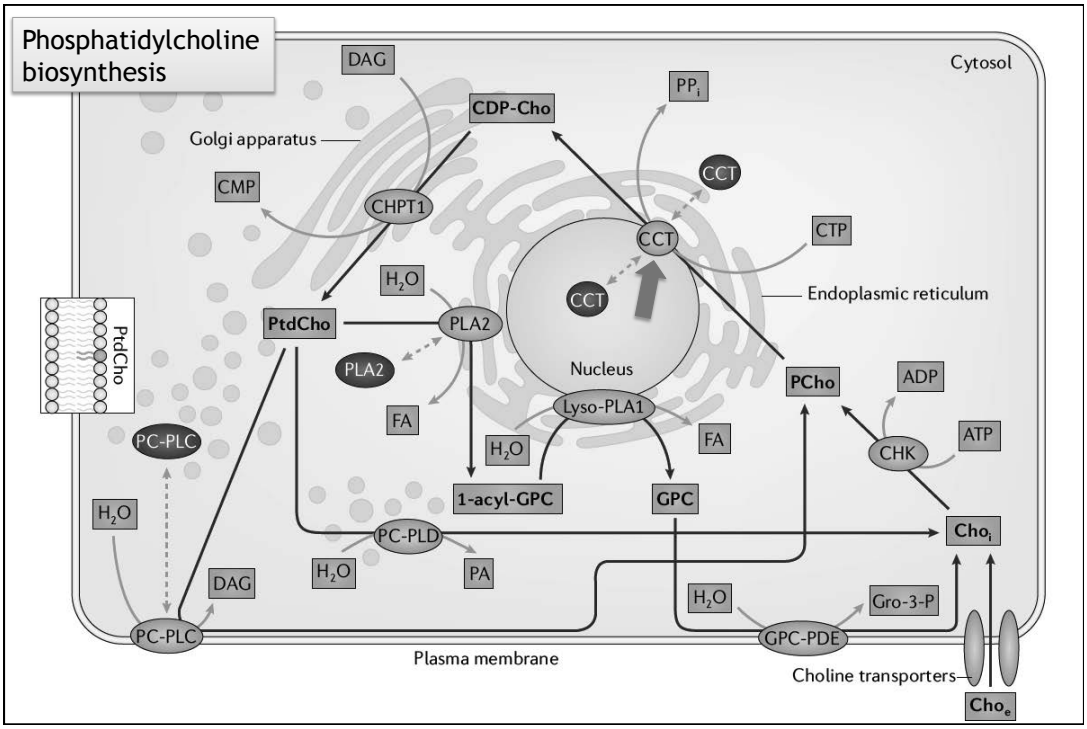


* Julie Hoover-Fong, Nara Sobreira, Julie Juergens et al

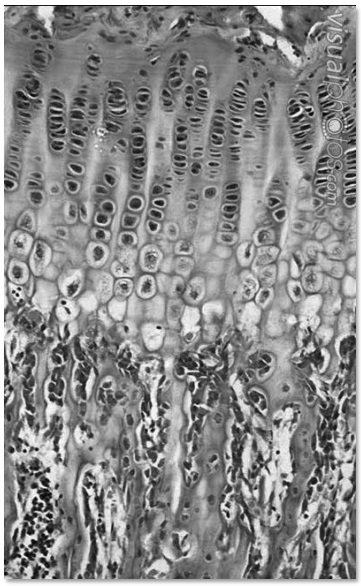
PCYT1A and SMD-CRD

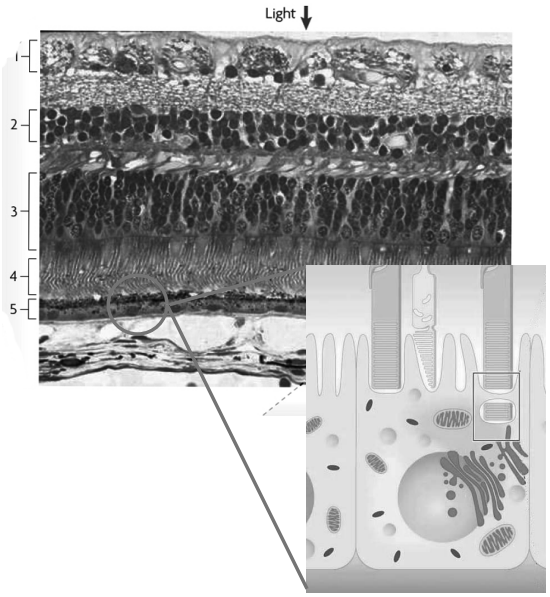
- Three unrelated SMD-CRD pedigrees segregating two missense mutations, A99V and P150A, in *PCYT1A* at 3q29
- Encodes Phosphocholine cytidylytransferase
- Both residues conserved to fish
- Catalyzes synthesis of phosphatidylcholine, a major membrane structural lipid

CTP + Choline phosphate ⟶ CDP-choline + PPi



Linking metaphysis and retina?





Thanks for your attention!
Acknowledgements

- The CMGS and especially the Baylor-Hopkins CMG team
- Ada Hamosh, Julie Hoover-Fong, Reid Sutton, Jim Lupski, Nara Sobreira & others for PhenoDb
- Dan Judge, Kim Doheny & CIDR team for the Alstrom project