Emerge III

Sequence!

Why?

Rare Variation Taking Center Stage

Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants

Yingrui Li[1,19], Nicolas Vinckenbosch[2,19], Geng Tian[1,19], Emilia Huerta-Sanchez[2,3,19], Tao Jiang[1,19], Hui Jiang[1], Anders Albrechtsen[4], Gitte Andersen[5], Hongzhi Cao[1], Thorfinn Korneliussen[4], Niels Grarup[5], Yiran Guo[1], Ines Hellman[6], Xin Jin[1,7], Qibin Li[1], Jiangtao Liu[1], Xiao Liu[1], Thomas Sparsø[5], Meifang Tang[1], Honglong Wu[1], Renhua Wu[1], Chang Yu[1], Hancheng Zheng[1,7], Arne Astrup[8], Lars Bolund[1,9,10], Johan Holmkvist[5], Torben Jørgensen[11,12], Karsten Kristiansen[1,4], Ole Schmitz[13,14], Thue W Schwartz[15], Xiuqing Zhang[1], Ruiqiang Li[1], Huanming Yang[1], Jian Wang[1], Torben Hansen[5,16], Oluf Pedersen[5,17,18], Rasmus Nielsen[2–4] & Jun Wang[1,4]

Sciencexpress Research Articles

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes

Jacob A. Tennessen,[1]* Abigail W. Bigham,[2]*† Timothy D. O'Connor,[1]* Wenqing Fu,[1] Eimear E. Kenny,[3] Simon Gravel,[3] Sean McGee,[1] Ron Do,[4,5] Xiaoming Liu,[6] Goo Jun,[7] Hyun Min Kang,[7] Daniel Jordan,[8] Suzanne M. Leal,[9] Stacey Gabriel,[4] Mark J. Rieder,[1] Goncalo Abecasis,[7] David Altshuler,[4] Deborah A. Nickerson,[1] Eric Boerwinkle,[6,10] Shamil Sunyaev,[4,8] Carlos D. Bustamante,[3] Michael J. Bamshad,[1,2]‡ Joshua M. Akey,[1]‡ Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [2]Department of Pediatrics, University of Washington, Seattle, Washington, USA. [3]Department of Genetics, Stanford University, Stanford, California, USA. [4]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [5]The Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [6]Human Genetics Center, University of Texas Health Sciences Center at Houston, Houston, Texas, USA. [7]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. [8]Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA. [9]Department of Molecular and

ing rare variation and has facilitated the genetic dissection of unsolved Mendelian disorders and studying human evolutionary history (*9–14*). Rare and low frequency (MAF between 0.5%-1%) variants have been hypothesized to explain a substantial fraction of the heritability of common, complex diseases (*15*). Since common variants explain only a modest fraction of the heritability of most traits (*16, 17*), NHLBI recently sponsored the multicenter Exome Sequencing Project (ESP), to identify novel genes and molecular mechanisms underlying complex heart, lung, and blood disorders by sequencing the exomes of a large number of individuals measured for phenotypic traits of substantial public health significance (e.g., early-onset myocardial infarction, stroke, body mass index).

18, 2012

Sciencexpress Reports

An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People

Matthew R. Nelson,1*† Daniel Wegmann,2* Margaret G. Ehm,1 Darren Kessner,2 Pamela St. Jean,1 Claudio Verzilli,1 Judong Shen,1 Zhengzheng Tang,3 Silviu-Alin Bacanu,1 Dana Fraser,1 Liling Warren,1 Jennifer Aponte,1 Matthew Zawistowski,6 Xiao Liu,4, Hao Zhang,4 Yong Zhang,4 Jun Li,5 Yun Li,3 Li Li,1 Peter Woollard,1 Simon Topp,1 Matthew D. Hall,1 Keith Nangle,1 Jun Wang,4,6 Gonçalo Abecasis,7 Lon R. Cardon,1 Sebastian Zöllner,7,8 John C. Whittaker,1 Stephanie L. Chissoe,1 John Novembre,2†‡ Vincent Mooser1‡

[1]Quantitative Sciences, GlaxoSmithKline, RTP, NC, USA; Upper Merion, PA, USA; and Stevenage, UK. [2]Ecology and Evolutionary Biology, University of California–Los Angeles, Los Angeles, CA, USA. [3]Genetics, Biostatistics, University of North Carolina–Chapel Hill, Chapel Hill, NC, USA. [4]BGI, Shenzhen, China.

mately 7% of genes considered current or potential drug targets (*12*), enriched for cell signaling proteins and membrane-bound transporters (table S2). A total of 864 kb were targeted, including 351 kb of coding and 323 kb of untranslated (UTR) exon regions (database S1). Over 93% of target bases were successfully sequenced at a median depth of 27 reads per site (*13*). Because rare variant discovery can easily be confounded with sequencing errors, we performed numerous experiments to demonstrate high data quality (table S3) (*13*). The sequenced subjects include two population samples (n = 1,322 and 2,059) and 12 disease collections (n = 125–1,125 cases, table S4). The self-reported ancestry of the sample was predominantly European (12,514), African American (594) and

2012

nature COMMUNICATIONS OPEN

ARTICLE

Received 30 Mar 2010 | Accepted 3 Nov 2010 | Published 30 Nov 2010 DOI: 10.1038/ncomms1130

Deep resequencing reveals excess rare recent variants consistent with explosive population growth

Alex Coventry[1,*], Lara M. Bull-Otterson[2,*], Xiaoming Liu[3], Andrew G. Clark[1], Taylor J. Maxwell[3], Jacy Crosby[3], James E. Hixson[3], Thomas J. Rea[4], Donna M. Muzny[2], Lora R. Lewis[2], David A. Wheeler[2], Aniko Sabo[2], Christine Lusk[4], Kenneth G. Weiss[4], Humeira Akbar[2], Andrew Cree[2], Alicia C. Hawes[2], Irene Newsham[2], Robin T. Varghese[2], Donna Villasana[2], Shannon Gross[2], Vandita Joshi[2], Jireh Santibanez[2], Margaret Morgan[2], Kyle Chang[2], Walker Hale IV[2], Alan R. Templeton[5], Eric Boerwinkle[3], Richard Gibbs[2] & Charles F. Sing[4]

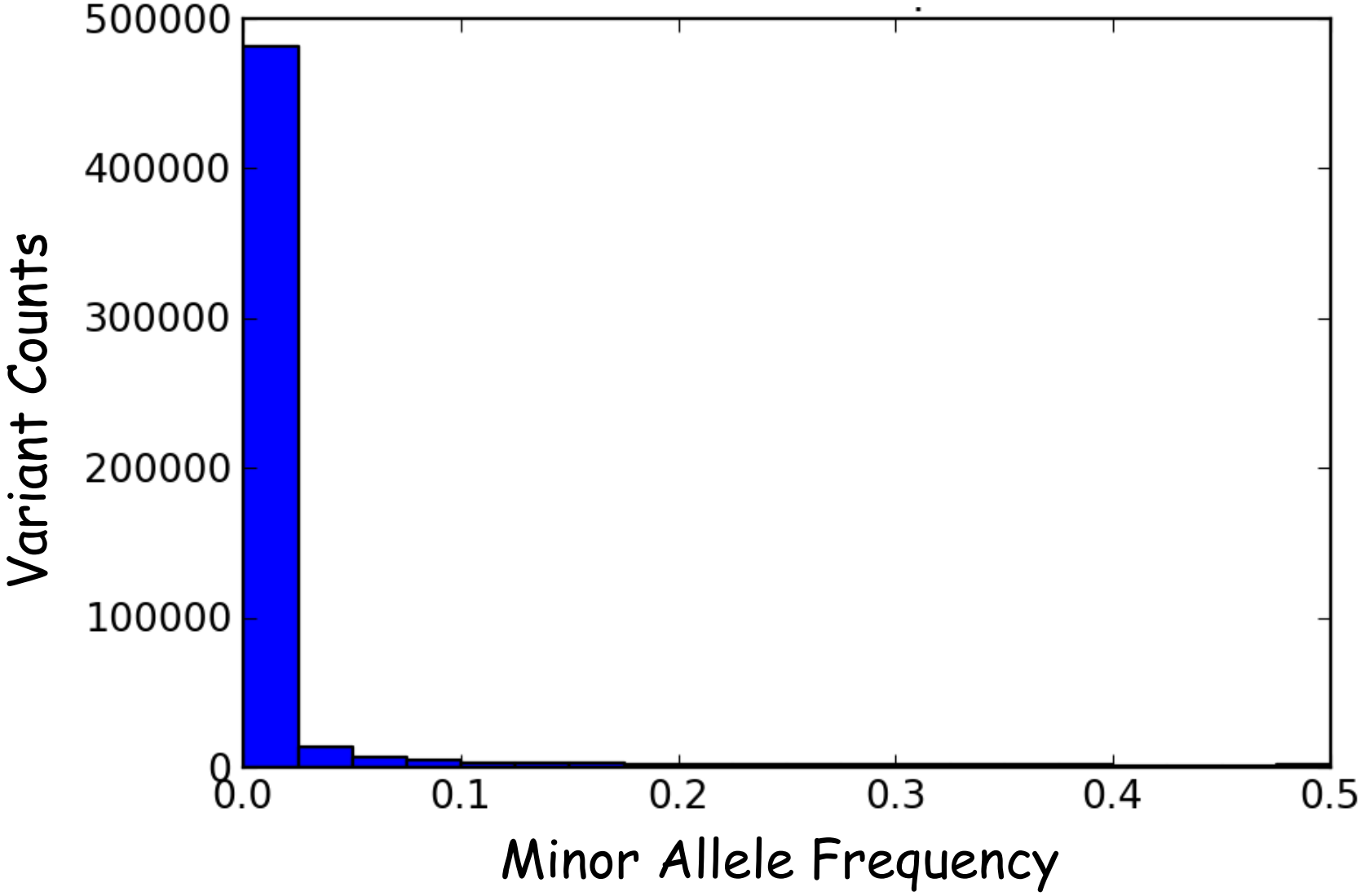
LETTER

doi:10.1038/nature11690

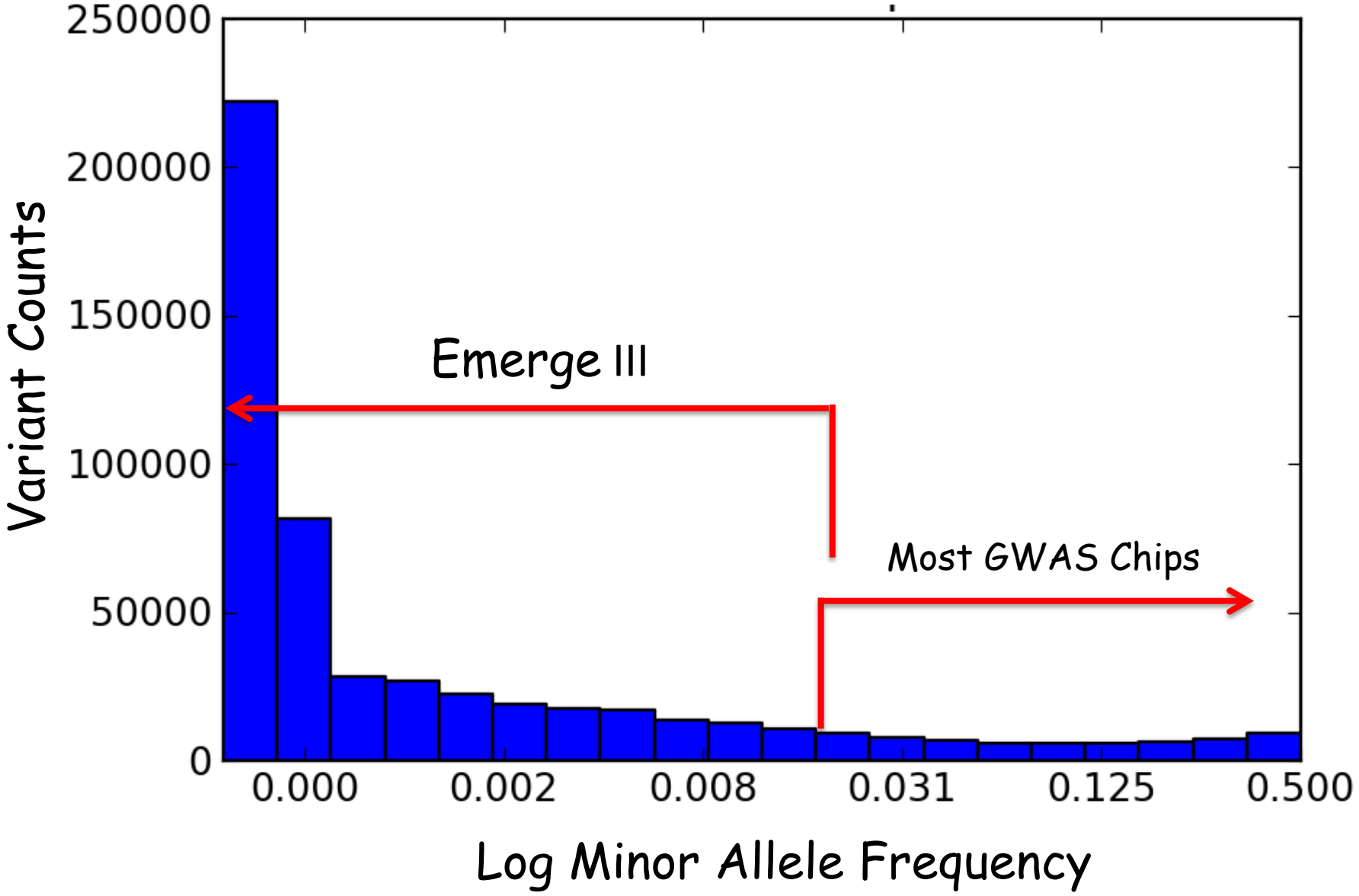
Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants

Wenqing Fu[1], Timothy D. O'Connor[1], Goo Jun[2], Hyun Min Kang[2], Goncalo Abecasis[2], Suzanne M. Leal[3], Stacey Gabriel[4], David Altshuler[4], Jay Shendure[1], Deborah A. Nickerson[1], Michael J. Bamshad[1,5], NHLBI Exome Sequencing Project* & Joshua M. Akey[1]

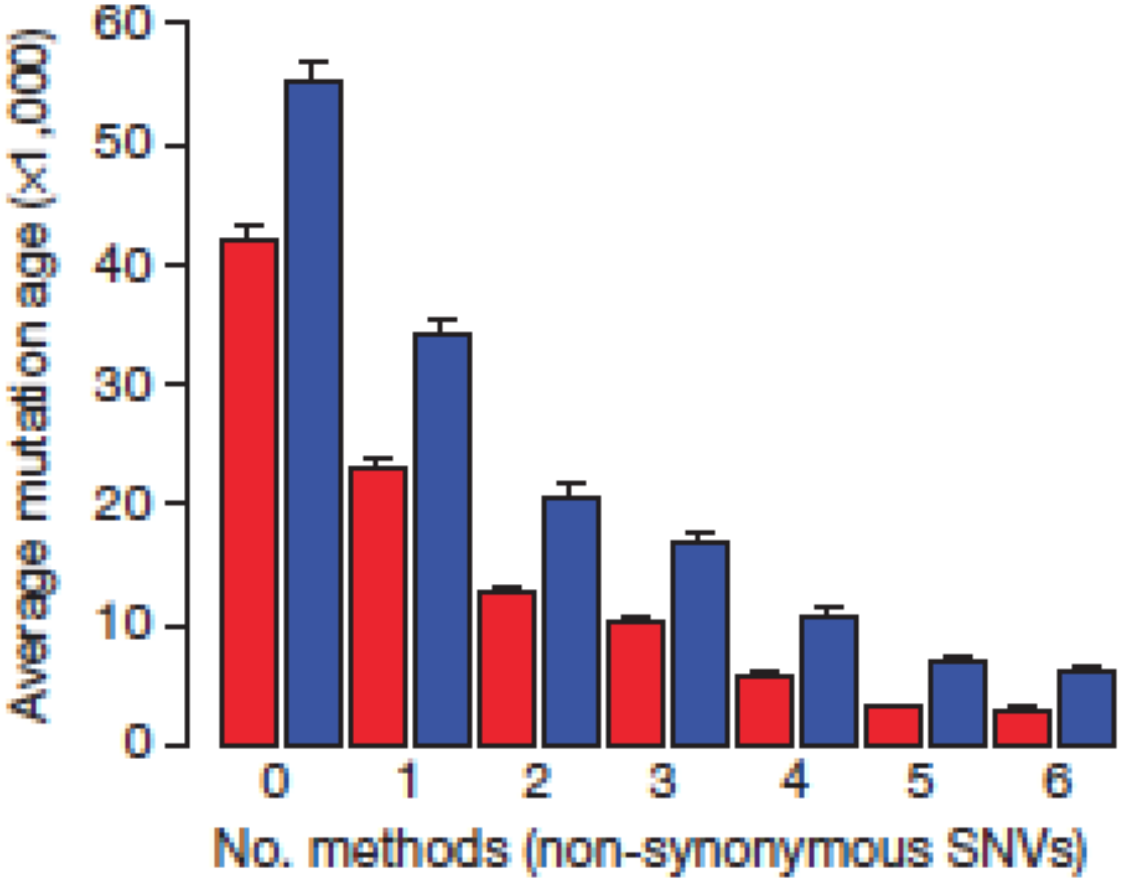
Most genetic variation in the population is rare



Most genetic variation in the population is rare



Most deleterious variation is rare and young in the population



Fu et al Nature 2013

Deleterious variants are individually rare collectively common

7 to 12% of the population carry a potentially major damaging *rare* variant in a common drug metabolizing *CYP*

PGRN

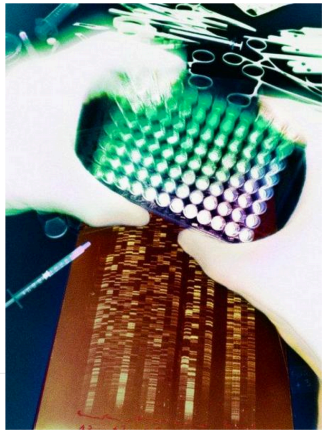
eMERGE Network
electronic medical records & genomics

Login | Register

Home | For Researchers | Phenotypes on PheKB | eMERGE RecordCounter | SPHINX | Publications | Contact

The eMERGE Network

The mapping of the human genome has enabled new exploration of how genetic variations contribute to health and disease. To better realize this promise, researchers must now determine ways in which genetic make-up gives some individuals a greater chance of becoming sick with chronic conditions such as diabetes, Alzheimer's, or heart disease. The goal of gaining this knowledge is to translate it to bedside practice and ultimately improve patient care.



The Electronic Medical Records and Genomics (eMERGE) Network is a national consortium organized by NHGRI to develop, disseminate, and apply approaches to research. It combines DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput genetic research with the ultimate goal of returning genomic testing results to patients in a clinical care setting. The Network is currently exploring more than a dozen phenotypes (with 13 additional electronic algorithms having already been published). Various models of returning clinical results have been implemented or planned for pilot at sites across the Network. Themes of bioinformatics, genomic medicine, privacy and community engagement are of particular relevance to eMERGE.

What makes eMERGE unique?

Each center participating in the Network is studying the relationship between genome-wide genetic variation and a common human trait. Such studies commonly involve testing hundreds of thousands of genetic variants called single nucleotide polymorphisms (SNPs) throughout the genome in people with and without the trait. A number of such studies are routinely conducted to uncover the association between disease and a person's genetic make-up, but those studies are typically costly and take a long time to complete.

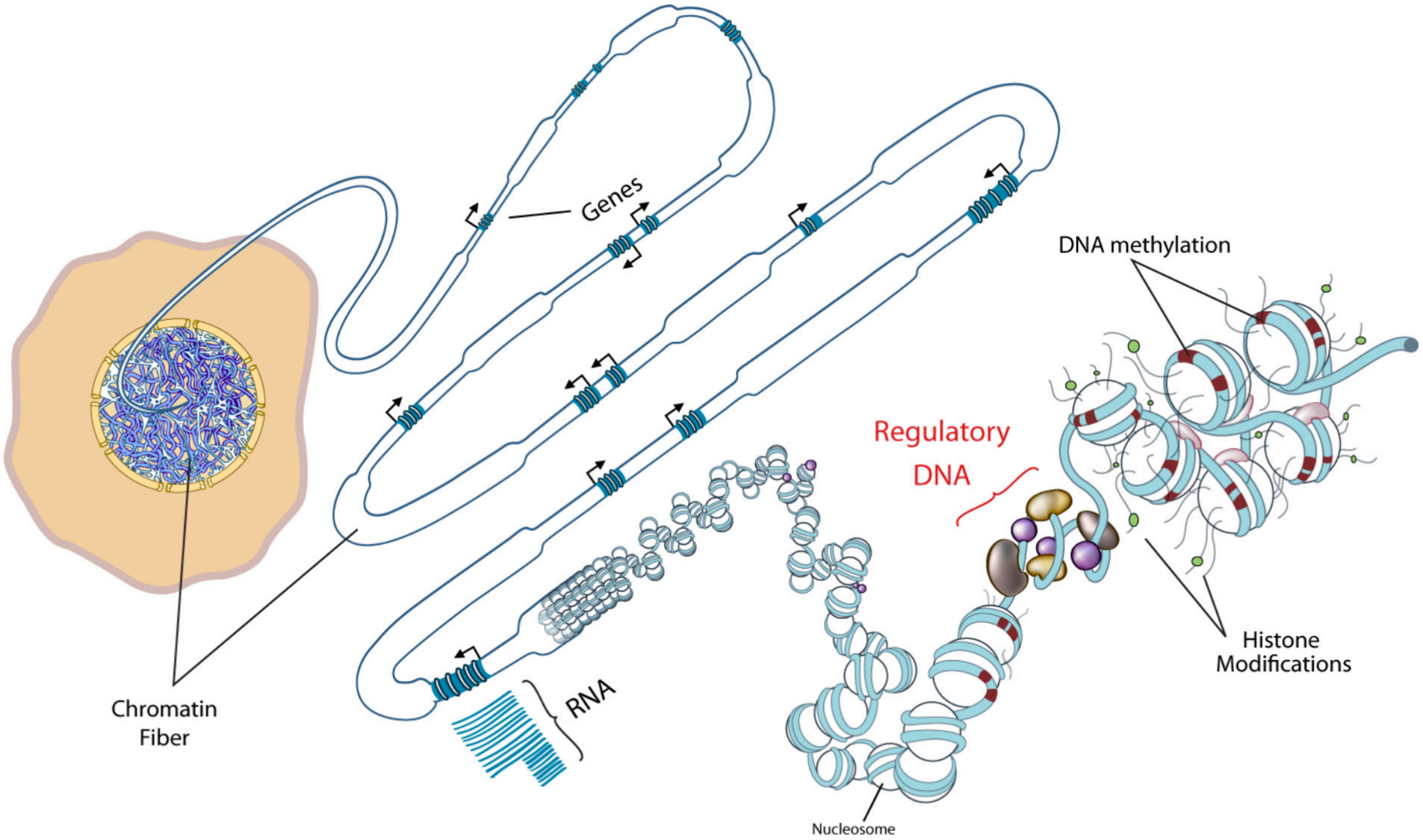
The case for sequencing

The vast majority of variation is rare and previously unknown. Although individually rare, collectively common.

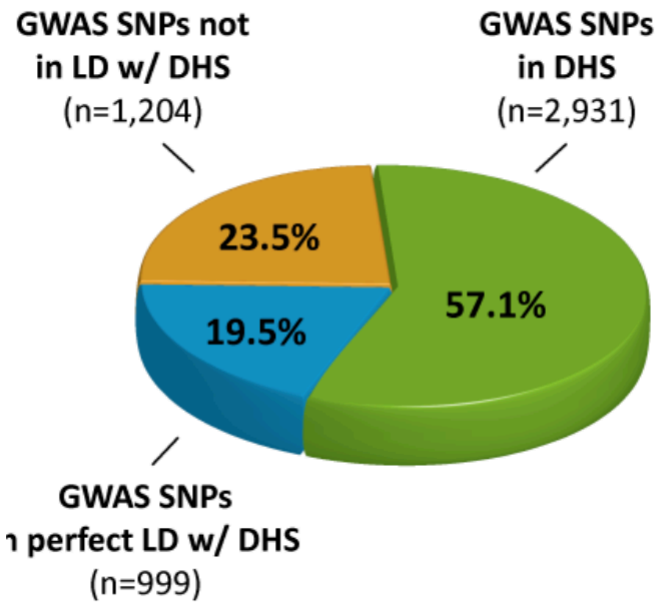
The most impactful variants are not only rare but young in ancestry – family specific? – thoughts on families?

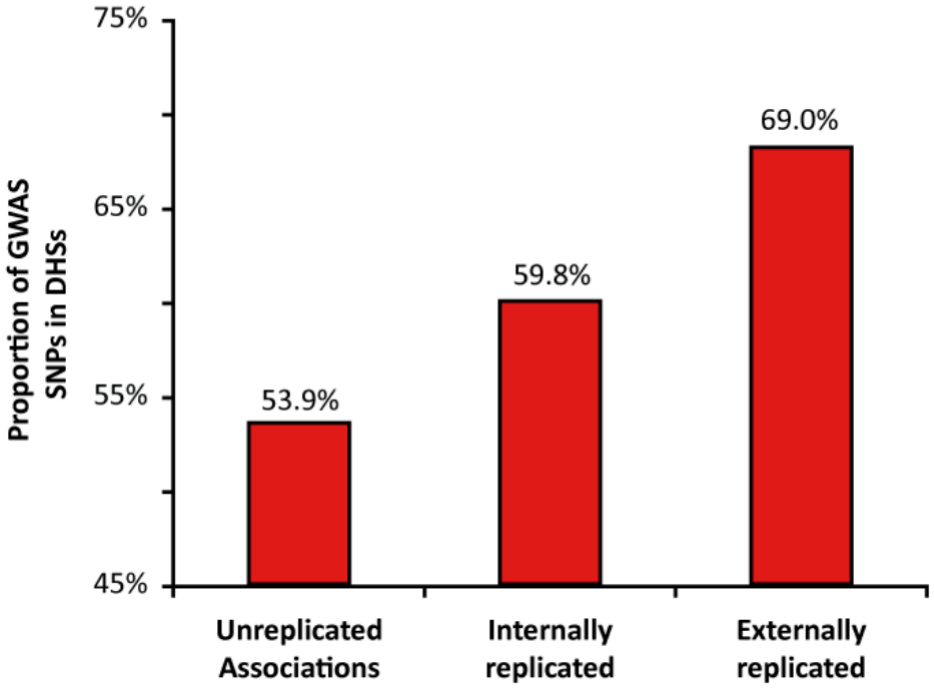
Seque from GWAS from EMERGE I & II

ENCODE – enables an understanding of non-coding variants



Disease- and trait-associated SNPs are concentrated in regulatory DNA





~1.8-fold for all replicated variants in all disorders

>10-fold for specific disese-cell type pairings

Maurano et al., Science 2012

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

CELL 155: 70, 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

Uncovered thousands of associations between Mendelian and complex disease

Explore how mendelian disease gene variants interact to contribute to common complex diseases/traits

ACMG POLICY STATEMENT | Genetics inMedicine

ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing

Robert C. Green, MD, MPH[1,2], Jonathan S. Berg, MD, PhD[3], Wayne W. Grody, MD, PhD[4–6], Sarah S. Kalia, ScM, CGC[1], Bruce R. Korf, MD, PhD[7], Christa L. Martin, PhD, FACMG[8], Amy L. McGuire, JD, PhD[9], Robert L. Nussbaum, MD[10], Julianne M. O'Daniel, MS, CGC[3], Kelly E. Ormond, MS, CGC[11], Heidi L. Rehm, PhD, FACMG[2,12], Michael S. Watson, PhD, FACMG[13], Marc S. Williams, MD, FACMG[14] and Leslie G. Biesecker, MD[15]

Disclaimer: These recommendations are designed primarily as an educational resource for medical geneticists and other health-care providers to help them provide quality medical genetic services. Adherence to these recommendations does not necessarily ensure a successful medical outcome. These recommendations should not be considered inclusive of all proper procedures and tests or exclusive of other procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure or test, geneticists and other clinicians should apply their own professional judgment to the specific clinical circumstances presented by the individual patient or specimen. It may be prudent, however, to document in the patient's record the rationale for any significant deviation from these recommendations.

Genetics in Medicine 15: 565-574,2013

Exploring the spectrum of actionable variants in the sequence will help to explore disease associations

Sequence!

Selected Targets, such as ROR targets

Exome – coding regions

Genome $1,000 - coding plus non-coding
perfect seque for EMERGE

The case for sequencing

The vast majority of variation is rare and previously unknown. Although individually rare, collectively common.

The most impactful variants are not only rare but young in ancestry – family specific? – thoughts on families?

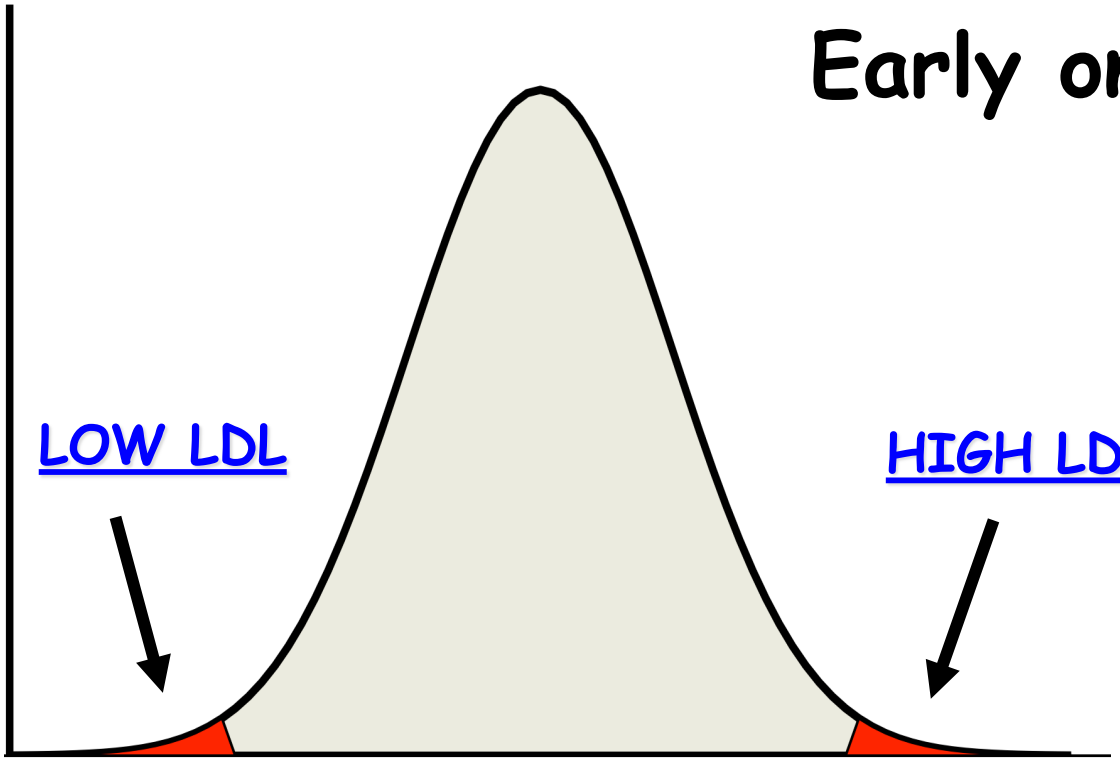
Careful selection of phenotypes will be key to uncovering new insights

Sequencing the tails?

Example

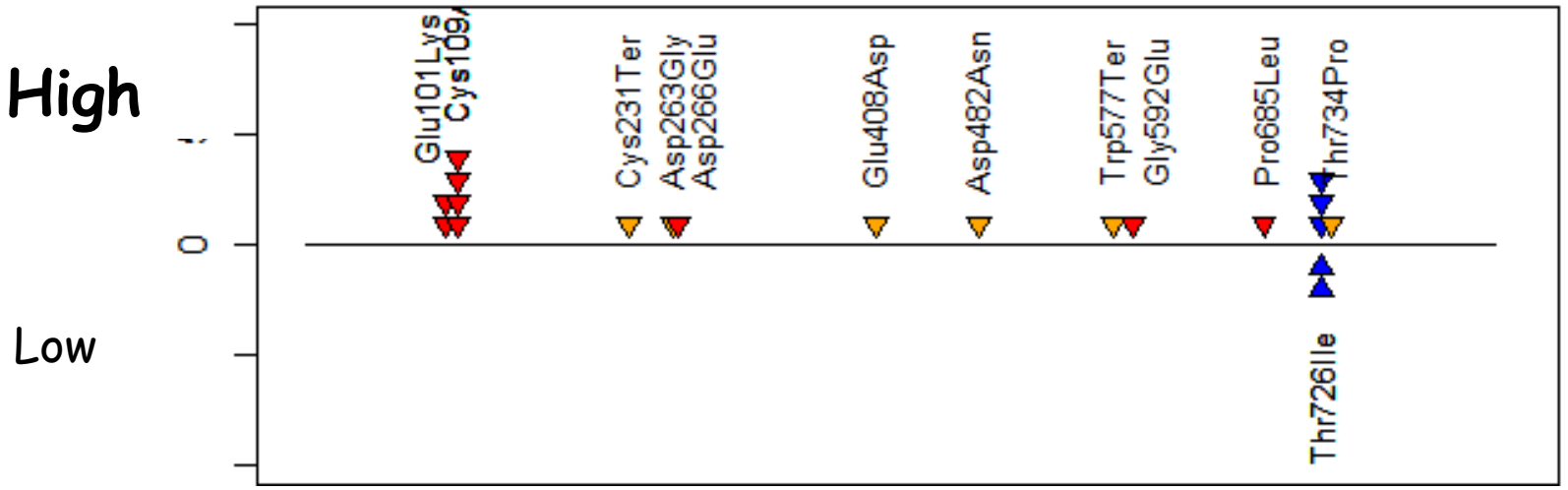
Extremes of a Trait

Early onset disease

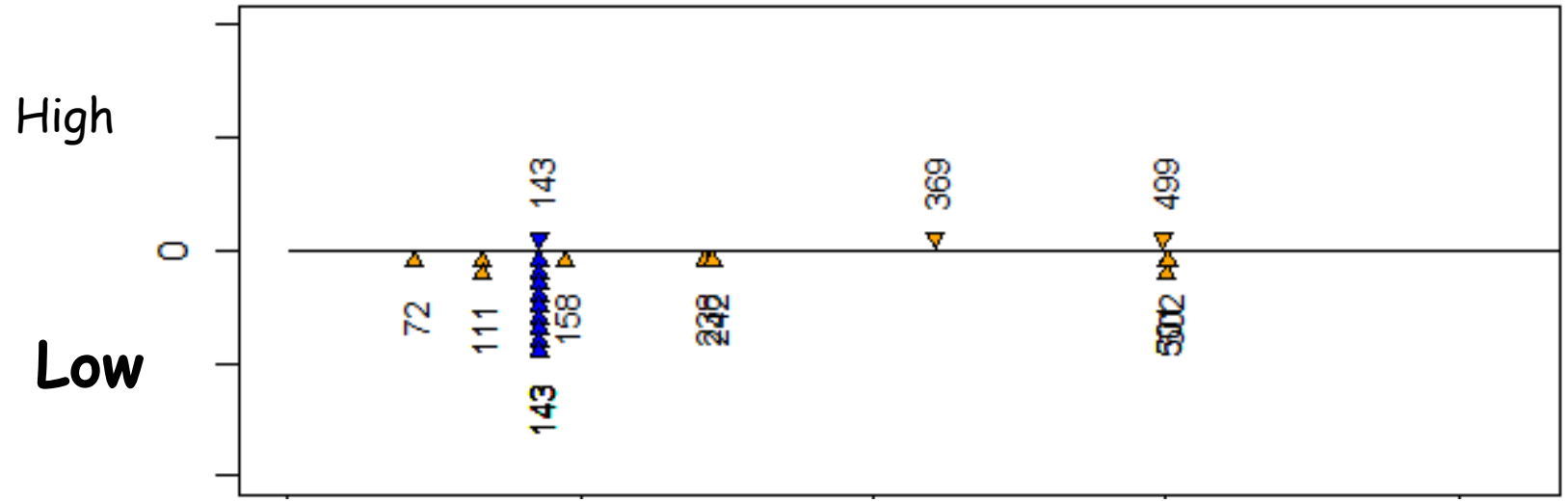


LDLR

Tails of LDL (0.1%)



PCSK9



Cristen Willer (UM) and Leslie Lange (UNC)

Discussion?