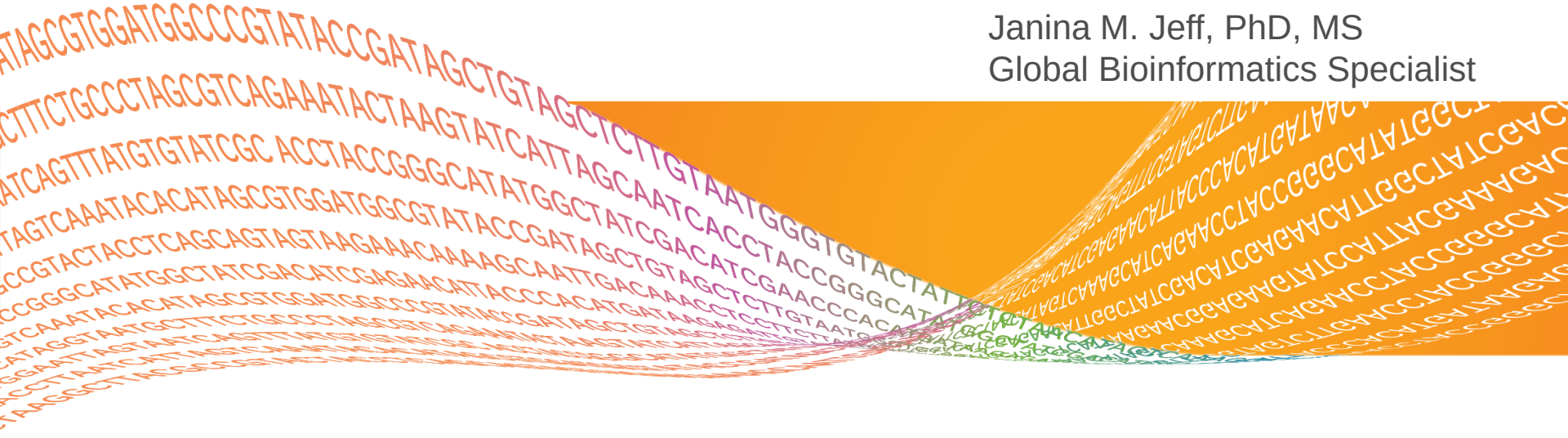
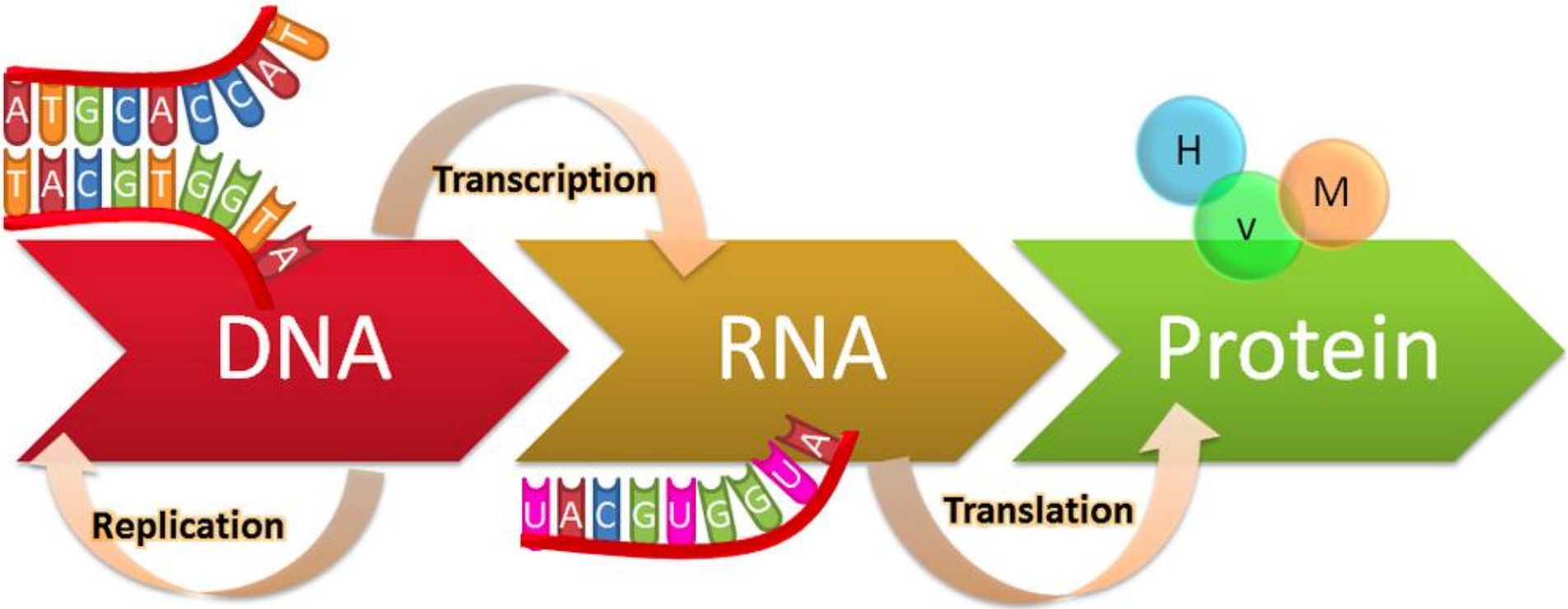
Emerging Resources for Genomic Assays That Power Discoveries in Diverse Populations

Janina M. Jeff, PhD, MS
Global Bioinformatics Specialist



illumina®

The Central Dogma



Central Dogma Career Path for Scientist

Central Dogma Career Path for Scientist

PhD

VANDERBILT UNIVERSITY

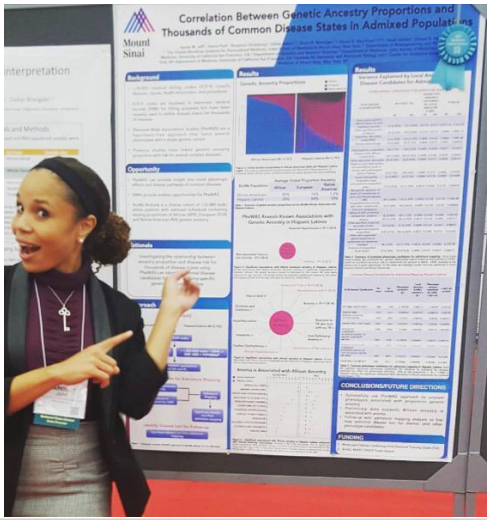
Interdisciplinary Graduate Program
~80-100 PhD students
~stipend, no teaching
~3-4 rotations

Dana Crawford, PhD

Crawford Lab
Genetic Epidemiology and Statistical Genetics
PhD Human Genetics
MS Applied Statistics

Pushed to my deepest limits

1. Imposter syndrome → building confidence (publications and oral presentations)
2. Fear of failing → understanding success cannot exist without failure (graduating and getting post doc offers)



illumina

Central Dogma Career Path for Scientist

Post-Doc

Icahn School of Medicine at Mount Sinai

Erwin Bottinger, MD

Original post-doc advisor
New York, NY
most diverse biobank

Communal post-doc

Adopted post-doc advisors

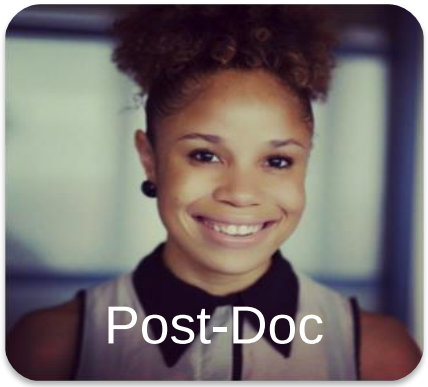
Eimear Kenny, PhD

Ruth Loos, PhD

Pushed to my deepest limits

1. Competitive climate → Defining your own success (forming new career opportunities: mentorship, teaching)
2. Different mentorship styles → Lead to independence (independently submitting abstracts and grants)

Central Dogma Career Path for Scientist



Assistant Pro.

Big Pharma

Direct to Consumer

Government

Biotechnology

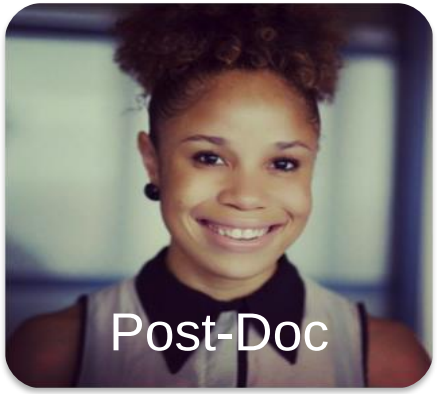
Science

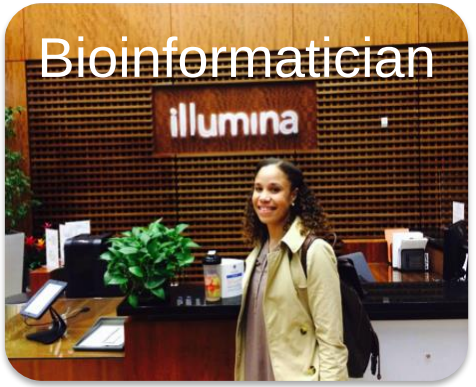
Lifestyle

Teaching and Mentorship

Research

Central Dogma Career Path for Scientist Does Not Exists





Biotechnology

illumina®

Science

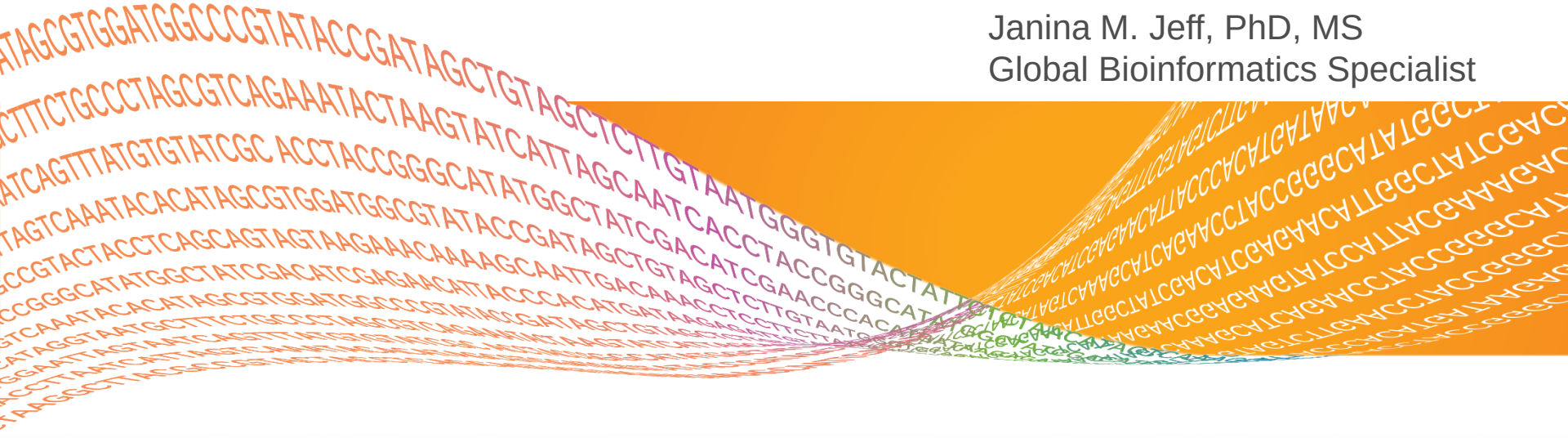
Lifestyle

Teaching and Mentorship

Research

Emerging Resources for Genomic Assays That Power Discoveries in Diverse Populations

Janina M. Jeff, PhD, MS
Global Bioinformatics Specialist

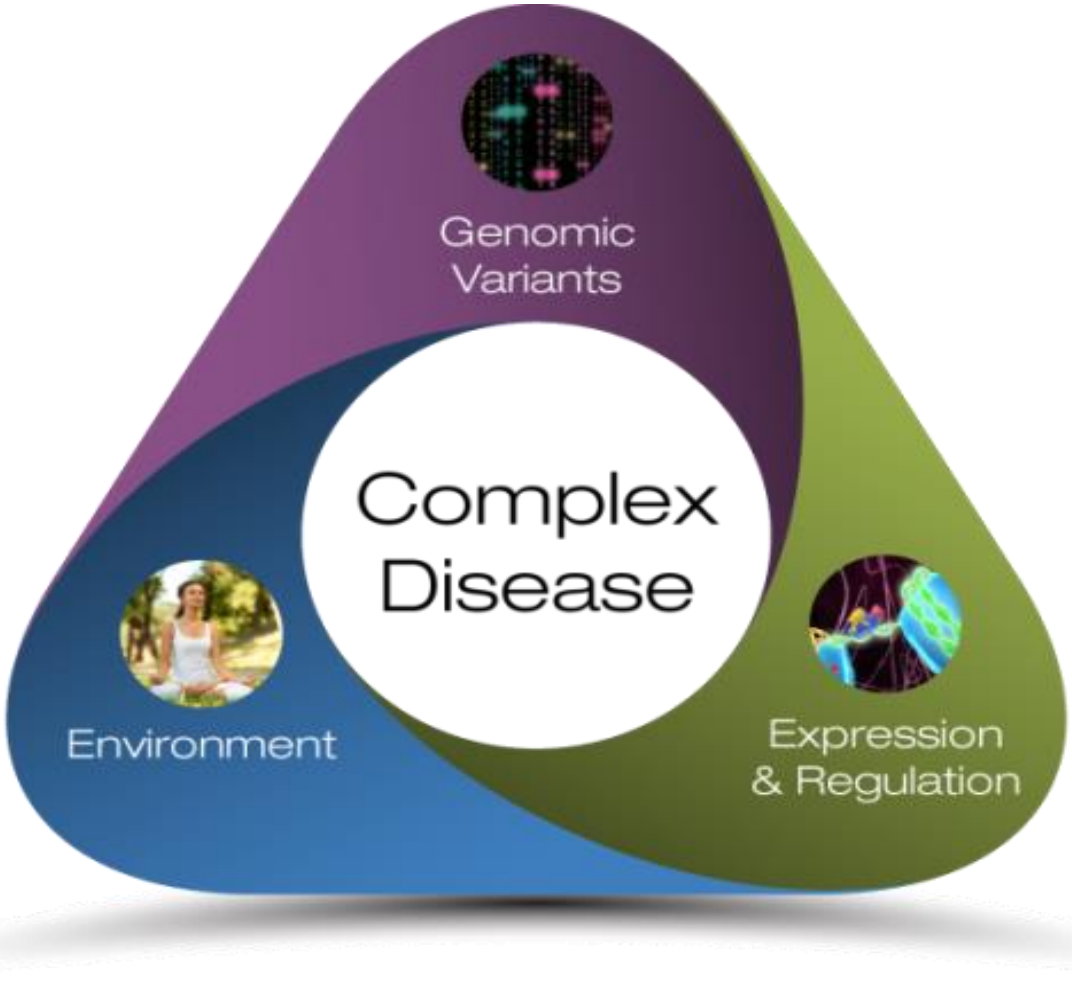


illumina®

An Integrated View of Complex Disease

A combination of genetic and environmental factors contributes to disease

Arrays



Clinical Database

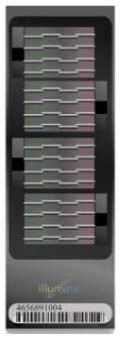
Sequencing

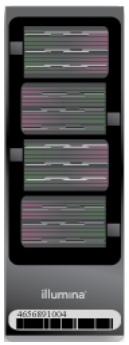
The Evolution of Large-Scale Genomics

10 years of Genome-Wide Association Studies (GWAS)

2004 | **2009** | **2011** | **2013** | **2014** | **2015** | **2016**

Common variants	Common and rare variants	Multi-ethnic / imputation Common and rare variants	Population based studies Common and rare variants
• HumanHap-300 • 550,660,610, 1M	• Omni family • Core family	• Multi-Ethnic family	• **Screening Array Program (GSA, ASA, MEA etc)**





For Research Use Only. Not for use in diagnostic procedures.

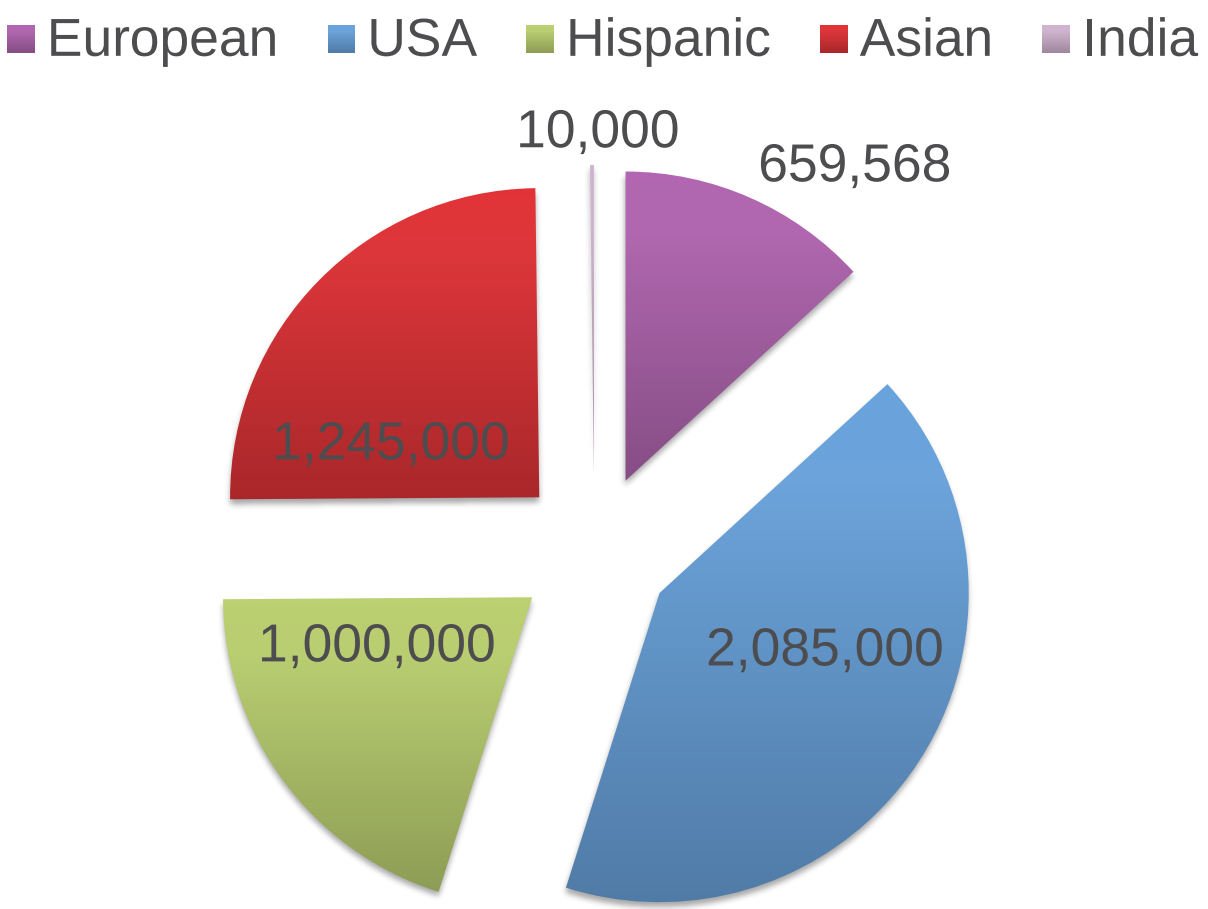
illumina®

Global Screening Array Consortium Members

- Global participation
- Large population cohorts, Biobanks, DTC, Researchers, Pharmaceutical companies, Health care providers, Service providers
- Millions of samples available within the next 2-3 years
- Total of ~200 consortia members from 55 institutions

Global Commitment to Population Studies

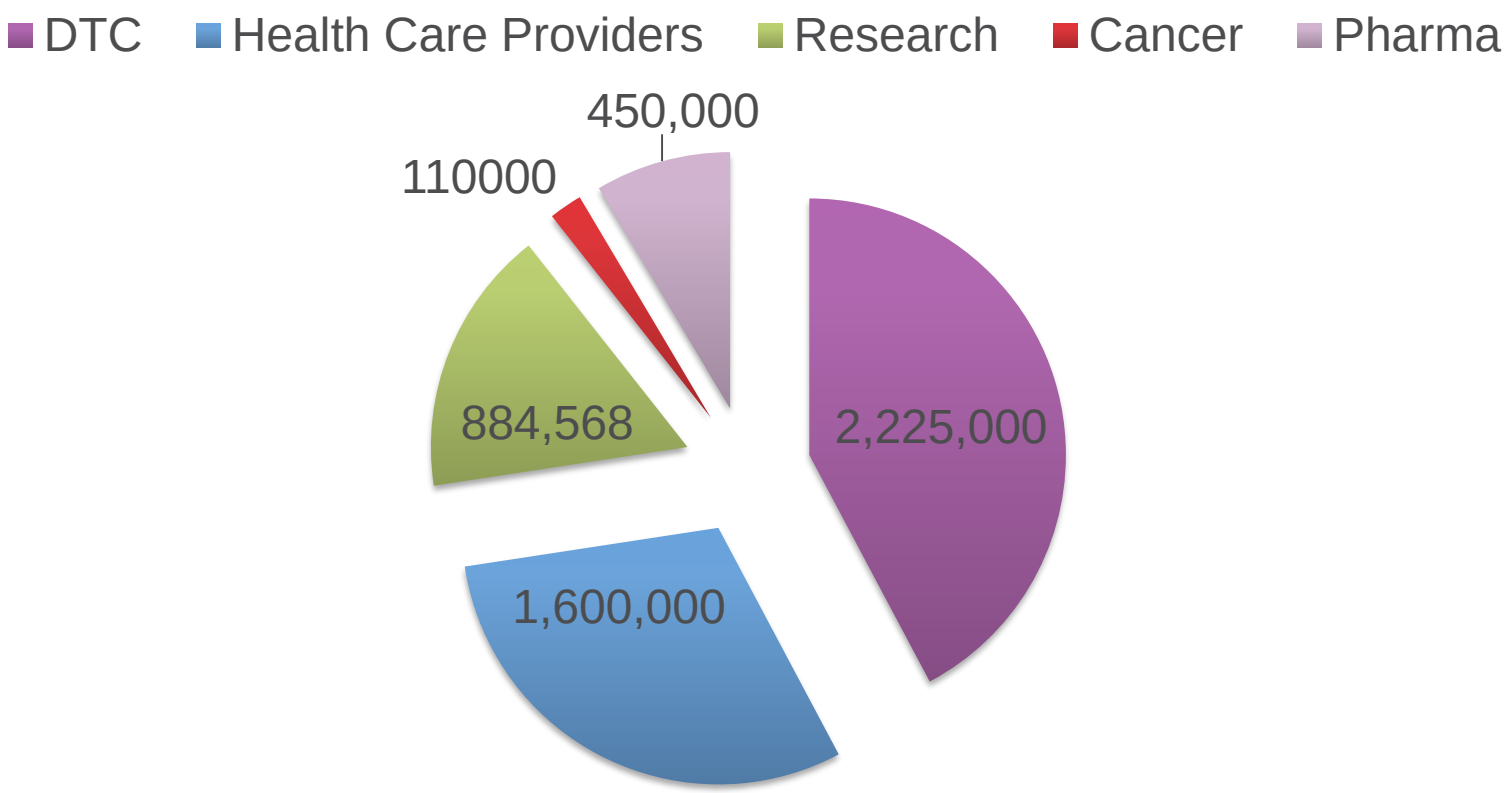
5 Million Samples Globally Sold



illumina®

New Applications on a Global Scale

Increasing Application Breadth



illumina®

Multi-Ethnic-GWAS backbone

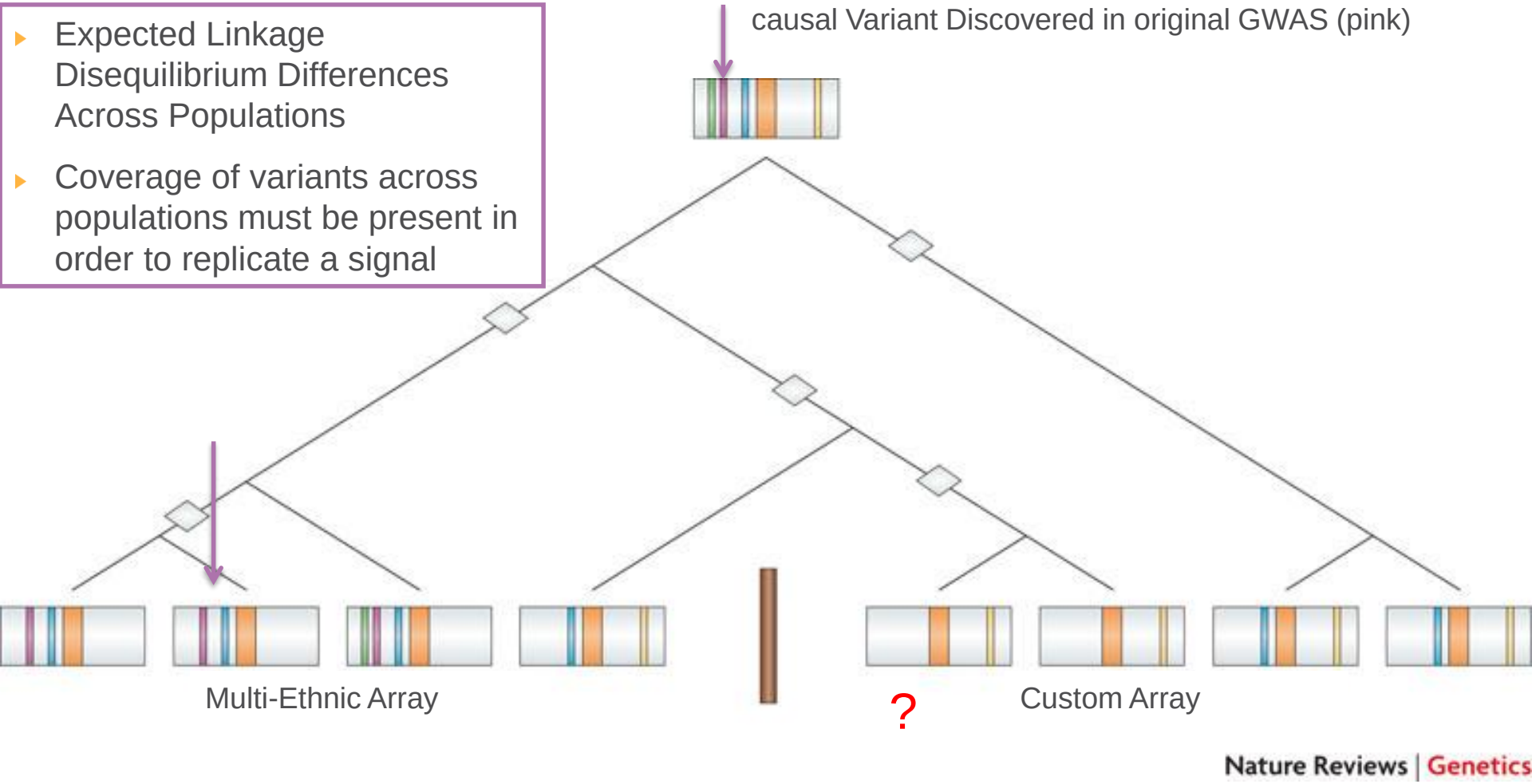
Two reasons to embrace multi-ethnic genomics studies

- Seamless replication
- Improved Understanding of Disease/Phenotypes

Causal Variant Must be Covered in Replication Population

Replication is not possible if SNP is not covered in replication sample

- Expected Linkage Disequilibrium Differences Across Populations
- Coverage of variants across populations must be present in order to replicate a signal



Successfully replicate causal variant

Fail to replicate causal variant

Rosenburg et al (2010) Nat Rev Genet 11(5):356-66

illumina®

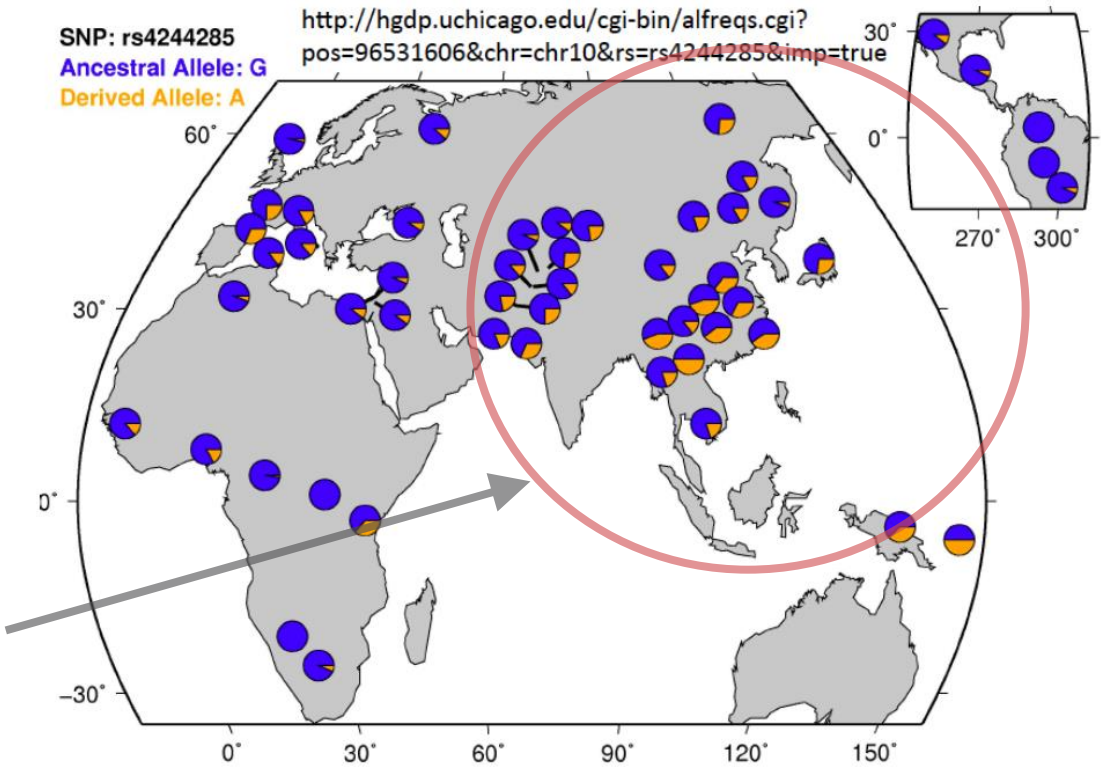
Improved Understanding of Disease/Phenotypes

Variants with large functional effects can be largely population specific*

High prevalence in East Asian populations of CYP2C19 genotype leads to lower Plavix efficacy

Blue = good metabolizers

Orange: poor metabolizers

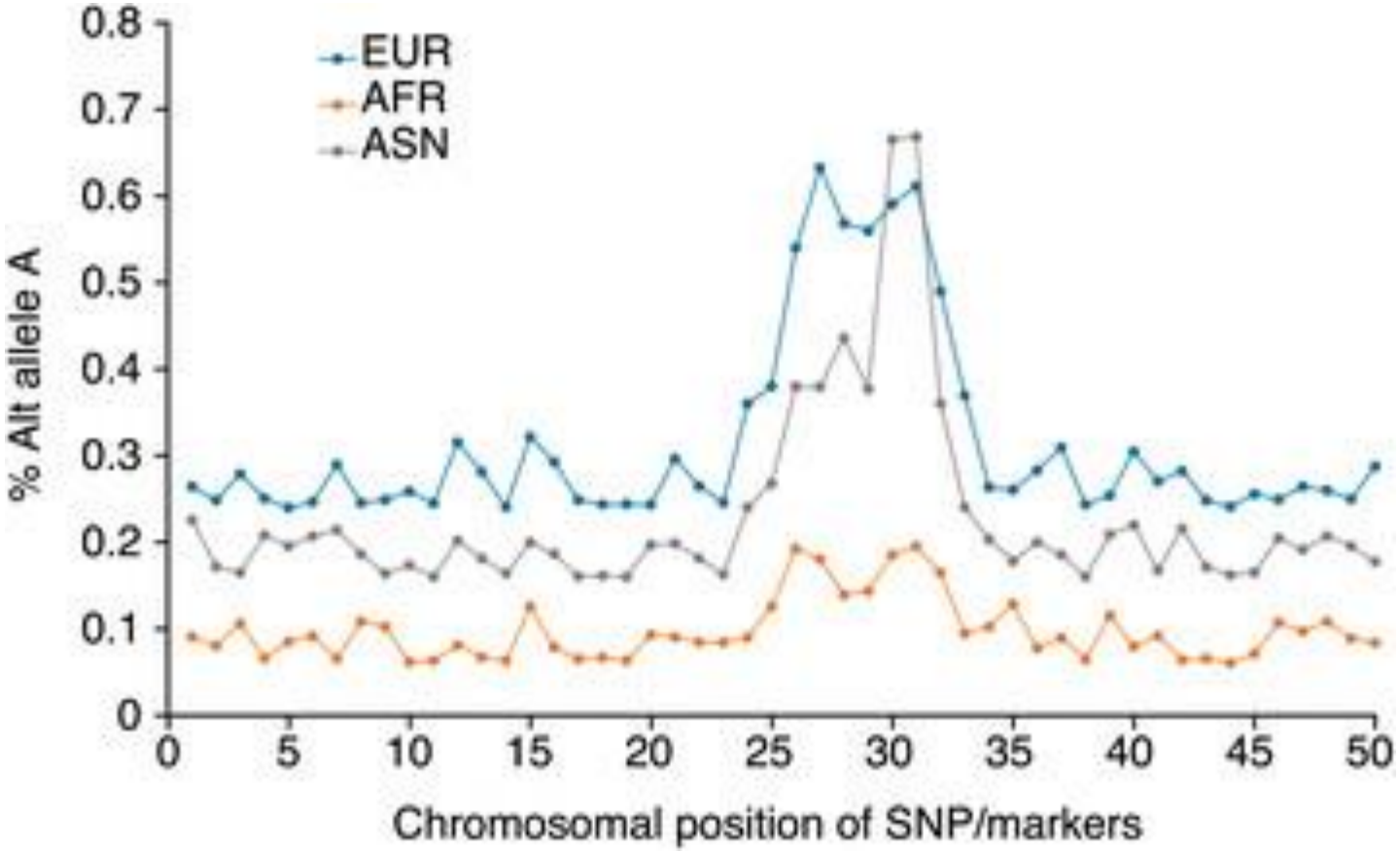


Large percentage of East Asian population with variant for poor metabolism of Plavix

* “Both PON1 Q192R and CYP2C19*2 influence platelet response to clopidogrel and ischemic events in Chinese patients undergoing percutaneous coronary intervention,” Yu Chen, Xiaohong Huang, Yong Tang, Yuguan Xie and Yachen Zhang, Int J Clin Exp Med. 2015; 8(6): 9266–9274

Opportunity for fine-mapping

Leveraging differential LD patterns enables fine- mapping of the causal SNP variants



Genome Med. 2014 Oct 31;6(10):91. doi: 10.1186/s13073-014-0091-5. eCollection 2014.

illumina®

Illumina Global Screening Array

A high powered, economical tool for population scale genomics

Universal Genome Wide Association Study (GWAS) Array delivering power across multiple populations

- Global content enriched for population specific and cross population variation
- High imputation accuracy across the entire allelic frequency spectrum

Comprehensive selection of clinical research variants

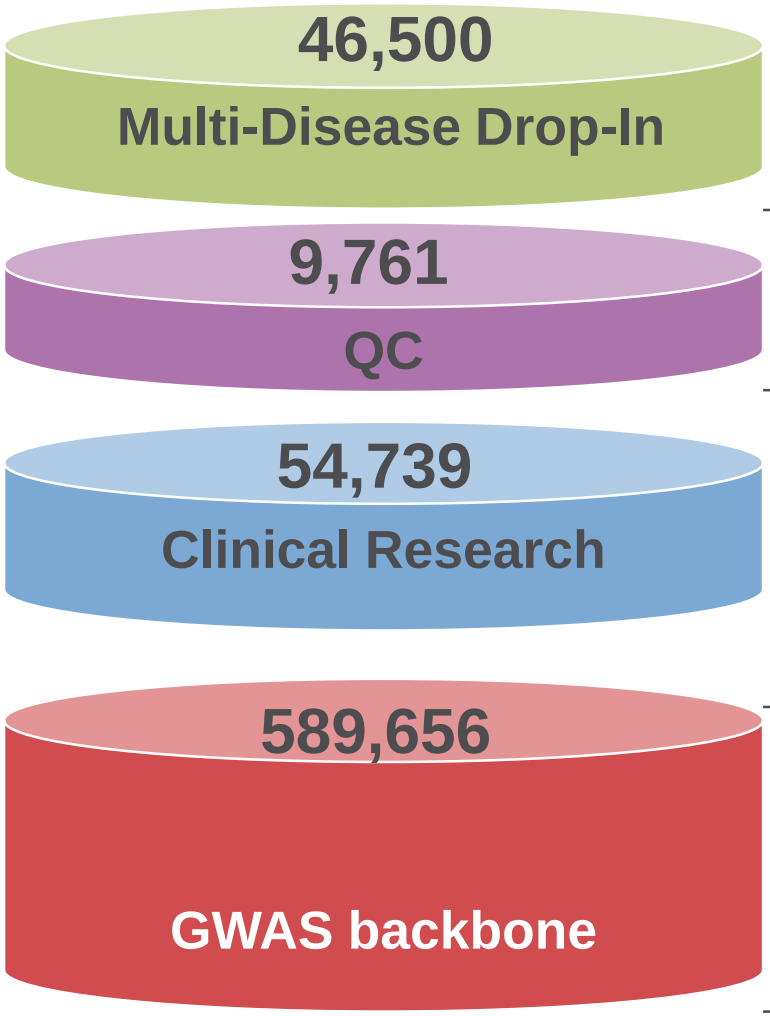
- >50K clinical research variants spanning breadth of functional variation
- Up to date content reviewed and validated by experts in medical genomics

Developed for high throughput population screening

- QC and high value content for sample identification, tracking, and stratification
- Designed on the Infinium® 24 sample format, add up to 50K in custom content

Global Screening Array(GSA) Manifest variant count

Predictive, Clinical Research, and QC



- User defined custom content or
- Pre-designed multi-disease drop in

- Sample tracking and stratification

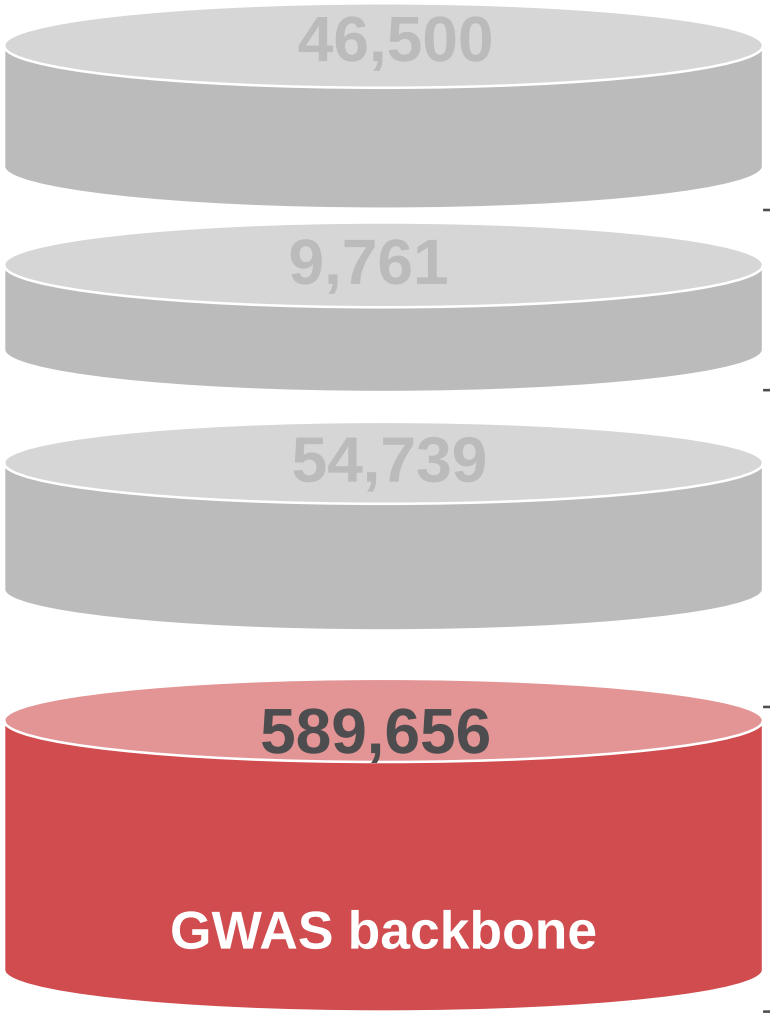
- Up-to-date known clinical associations
- Pharmacogenomics
- Well-curated exome content
- NHGRI-GWAS and HLA content

- Cross-Population and population-specific
- Enriched for low-frequency variants (1-5%)
- High imputation accuracy for ALL populations

FINAL Product : 700, 656

GSA Manifest variant count

Predictive, Clinical Research, and QC



- User defined custom content OR
- Pre-designed multi-disease drop in

- Sample tracking and stratification

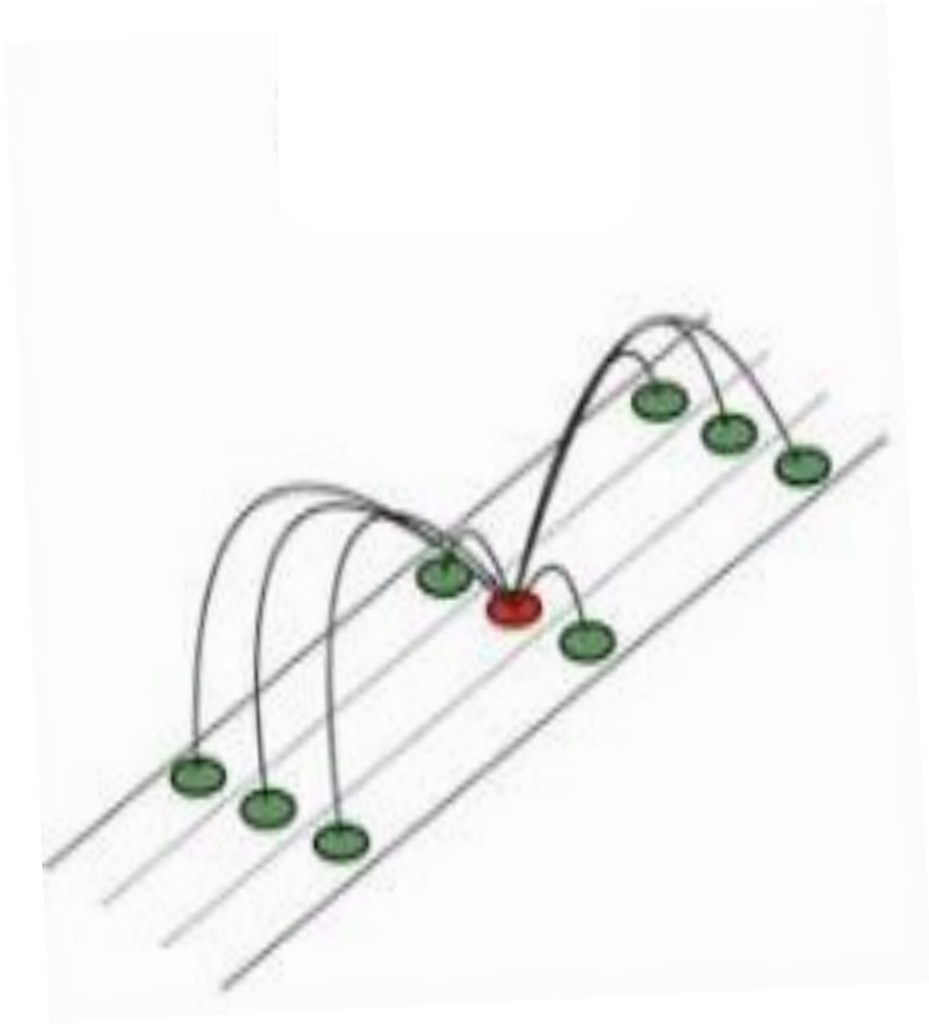
- Up-to-date known clinical associations
- Pharmacogenomics
- Well-curated exome content
- NHGRI-GWAS and HLA content

- Cross-Population and population-specific
- Enriched for low-frequency variants (1-5%)
- High imputation accuracy for ALL populations

Imputation

A New Paradigm Emerges

- Imputation is the New Gold Standard in Array Analysis
- Imputed SNPs have more power than pairwise LD (r2)
  - Clark et. al., NATURE GENETICS, 2007
- Imputation increases resolution of the association peak
- Key SNP criteria:
  - Imputation accuracy: how accurate is the imputed SNP
  - Imputation Efficiency: how many SNPs can you impute
  - Minor Allele Frequency*
    - **Low MAF 1-5% - Key Focus of GSA**
    - Common >5%



*Genomes Project C, Auton A, Brooks LD *et al.* A global reference for human genetic variation. Nature 2015; 526:68-74

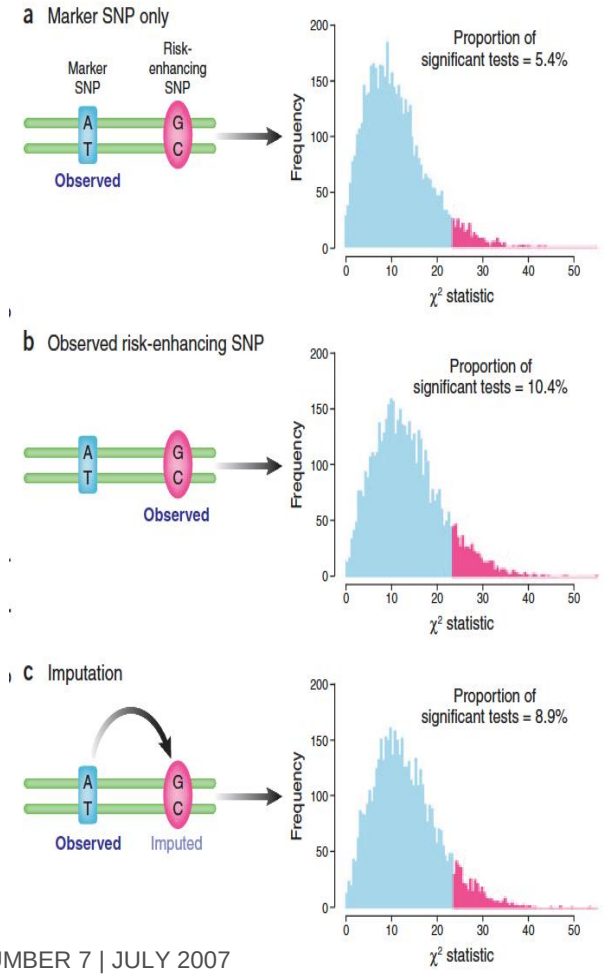
Gain more power with Imputed SNPs

New Paradigm in GWAS Analysis

nature genetics

- **Traditional r2 pairwise LD**
- **Direct Risk SNP Typing**
- **Imputed SNPs**

Proportion of significant tests
5.4%
10.4%
8.9%

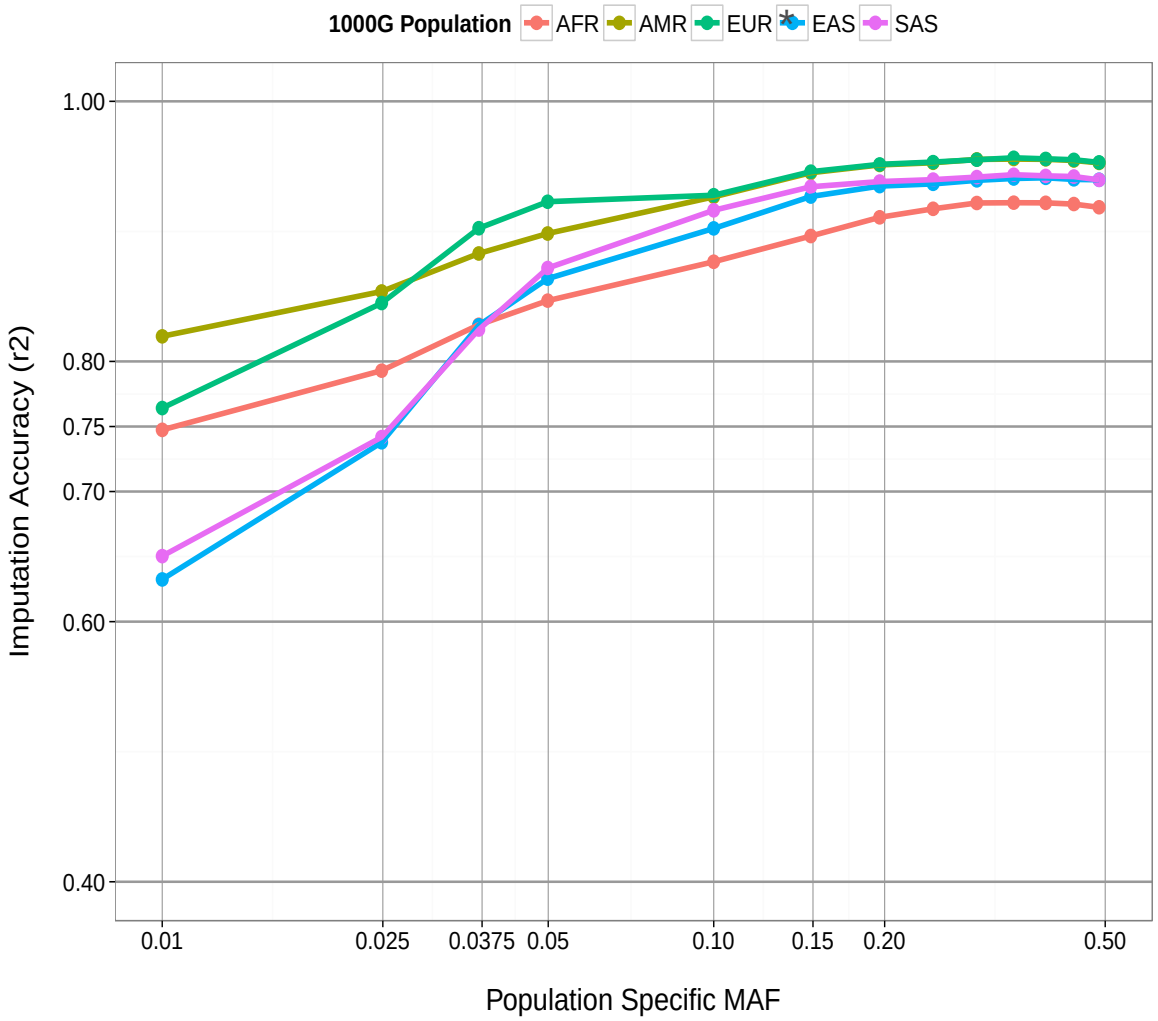


From: Conjuring SNPs to detect associations. Clark and J. Li NATURE GENETICS | VOLUME 39 | NUMBER 7 | JULY 2007

illumina®

High Imputation Accuracy Across All Populations

Accurate, economical coverage of the genome



Mean Imputation Accuracy

>1 % Minor allele frequency

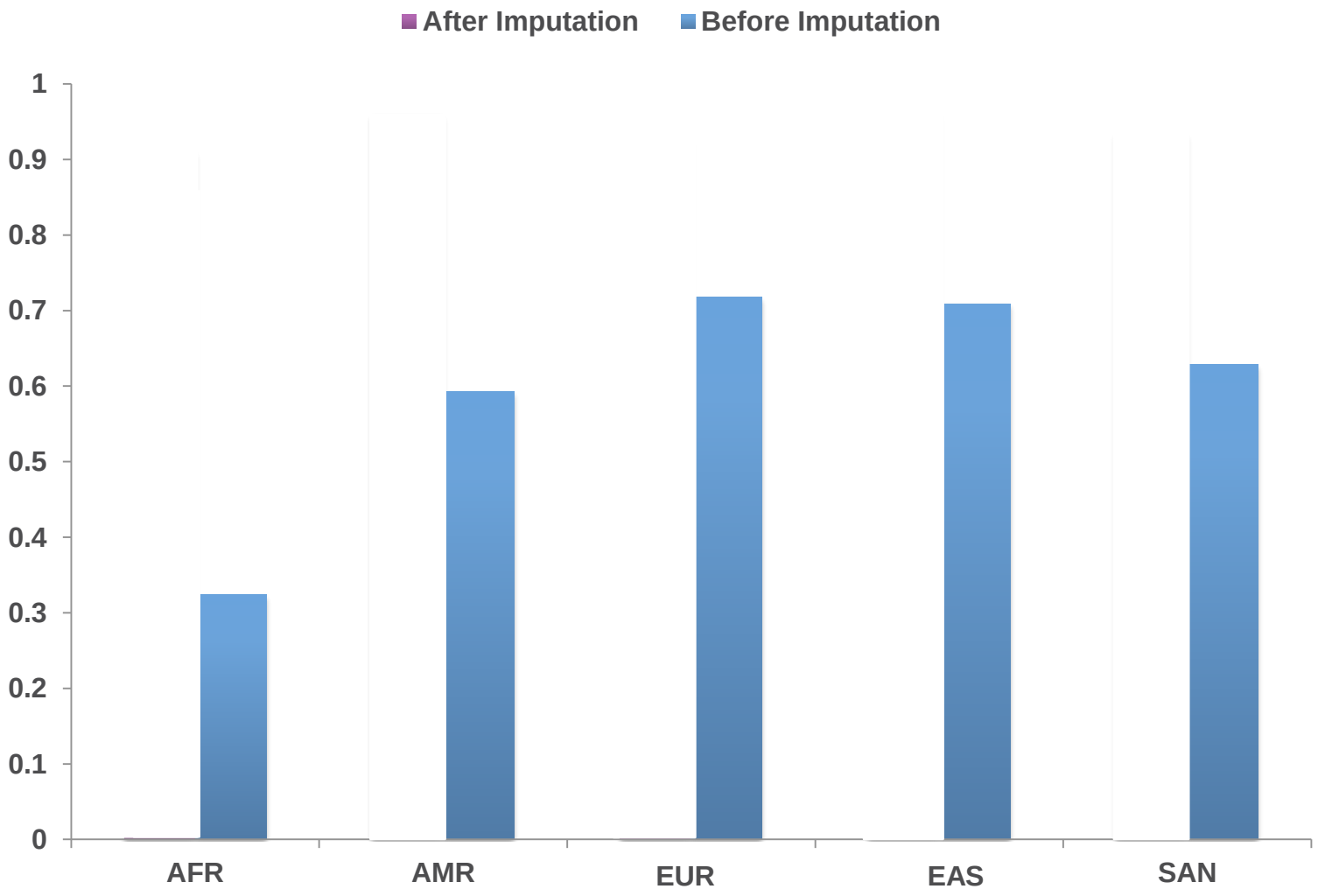
Population	GSA
AFR	0.95
AMR	0.97
EUR	0.97
EAS	0.96
SAS	0.96

Data calculations on file. Illumina, Inc. 2016

illumina®

Imputation Increases Genomic Coverage

GSA has greater than 90% coverage in all populations after imputation

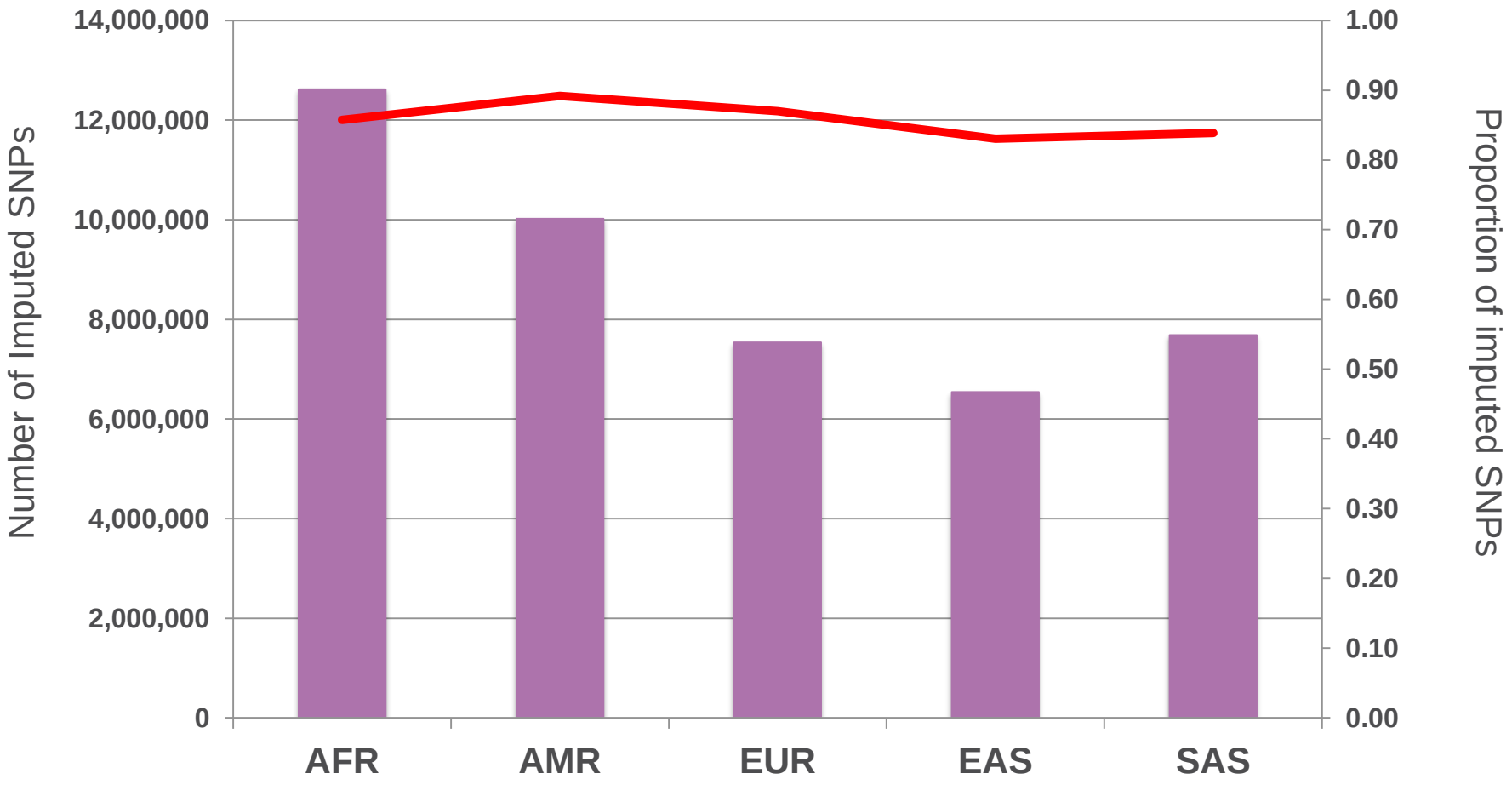


Data calculations on file. Illumina, Inc. 2016

illumina®

Universal Imputation Power Across Populations

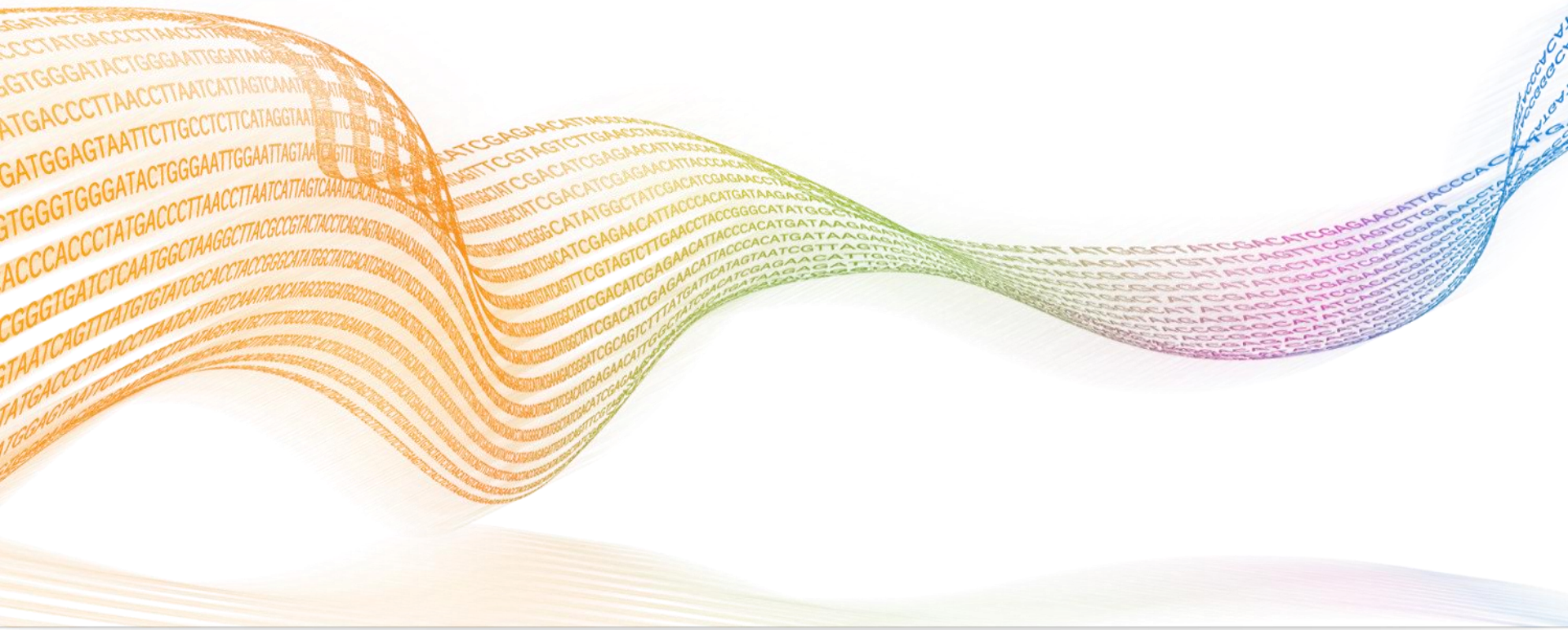
all SNPs in 1kGP >1% MAF and >80% imputation accuracy



Data calculations on file. Illumina, Inc. 2016

illumina®

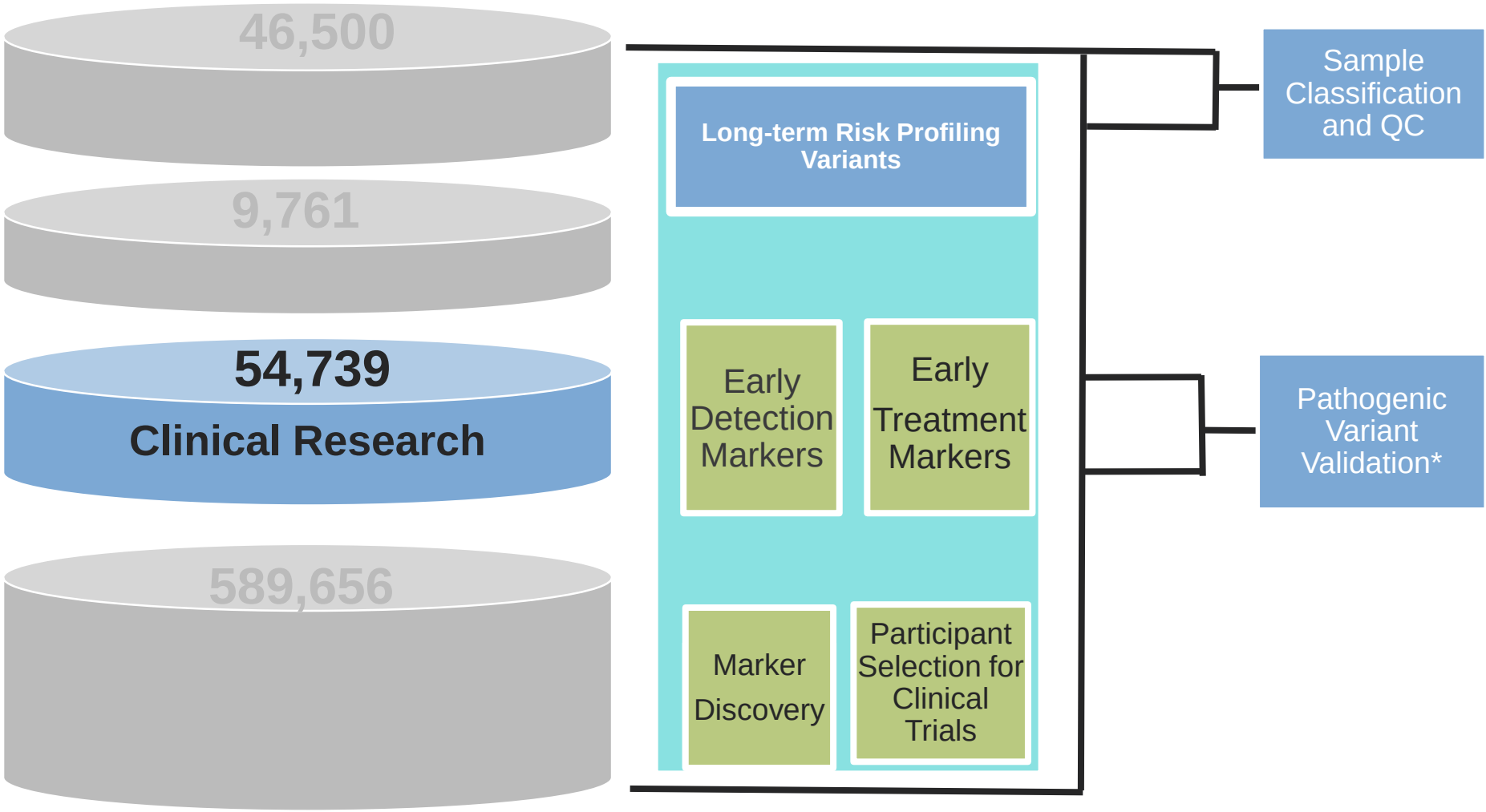
Thank you and Questions



illumina®

GSA Manifest variant count

Predictive, Clinical Research, and QC



*ClinVar, ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/

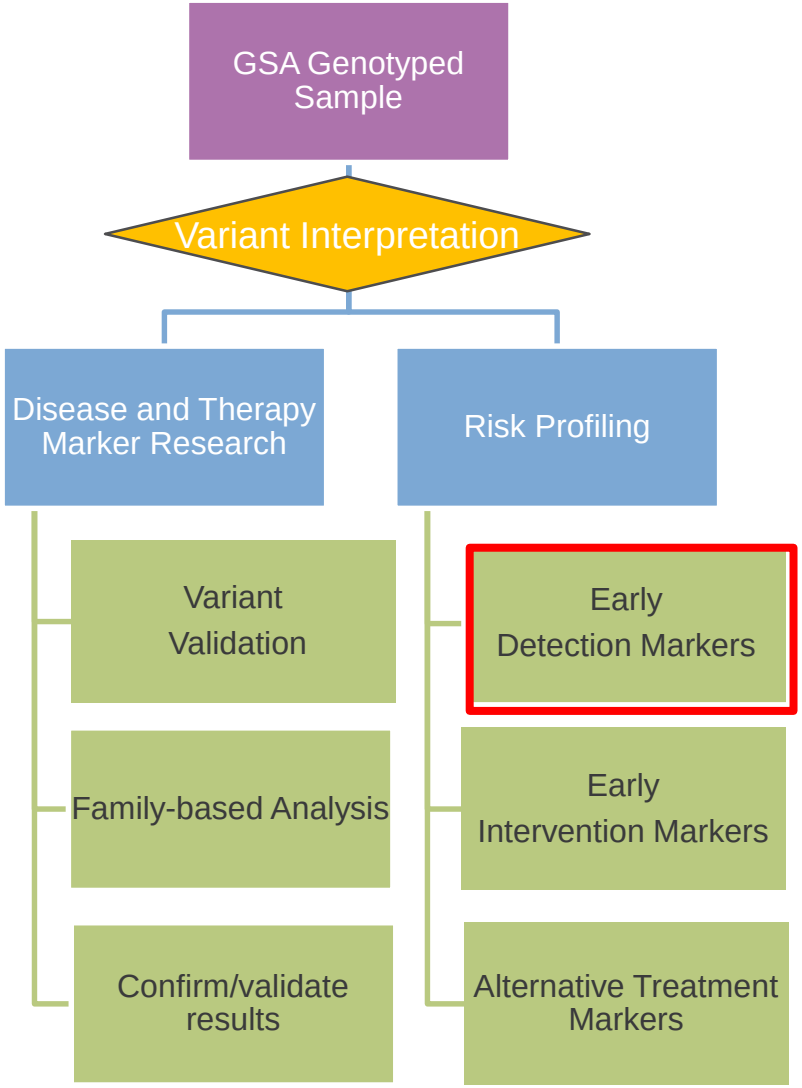
illumina®

Databases Overview

	Web page
ClinVar (135K): – Reports of the relationships among human variations and phenotypes, with supporting evidence.	www.ncbi.nlm.nih.gov/clinvar/
ExAC: – The Exome Aggregation Consortium: exome sequencing data from a variety of large-scale sequencing projects. The data set spans 60706 unrelated individuals sequenced as part of various disease-specific and population genetic studies.	http://exac.broadinstitute.org
NHGRI: – A Catalog of Published Genome-Wide Association Studies	https://www.genome.gov/gwastudies/
eMERGE: – Electronic Medical Records and Genomics (eMERGE) Network Combines biorepositories with electronic medical record (EMR) systems for genomic discovery and genomic medicine	https://www.genome.gov/27540473/electronic-medical-records-and-genomics-emerge-network/
PharmGKB: – Pharmacogenomics knowledge resource with clinical information including dosing guidelines and drug labels, potentially clinically actionable gene-drug associations and genotype-phenotype relationships	https://www.pharmgkb.org/

GSA Clinical Research Applications

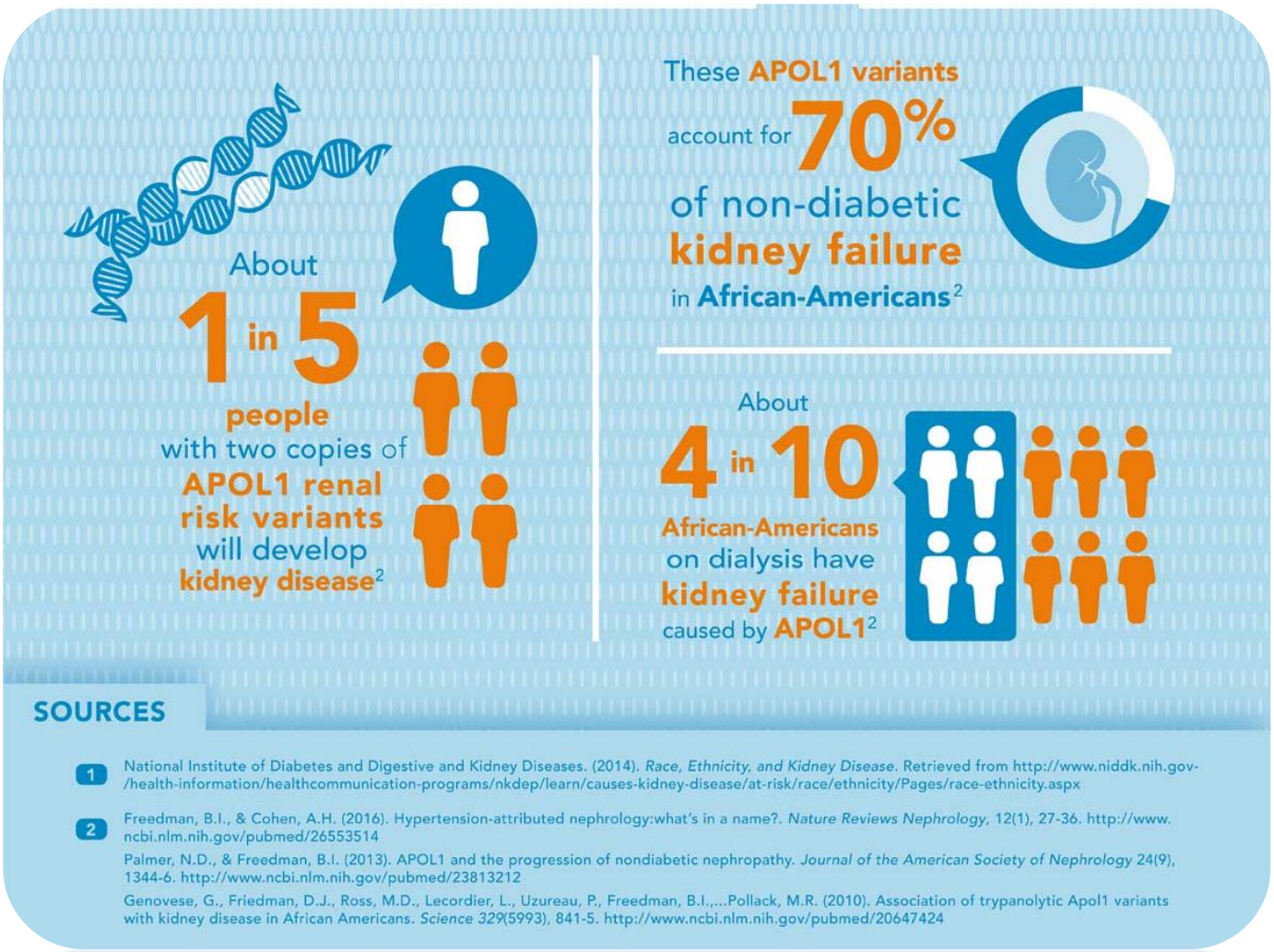
Variant Validation



illumina®

GSA Pathogenic Variant: *APOL1*

May Predict Kidney Failure:



illumina®

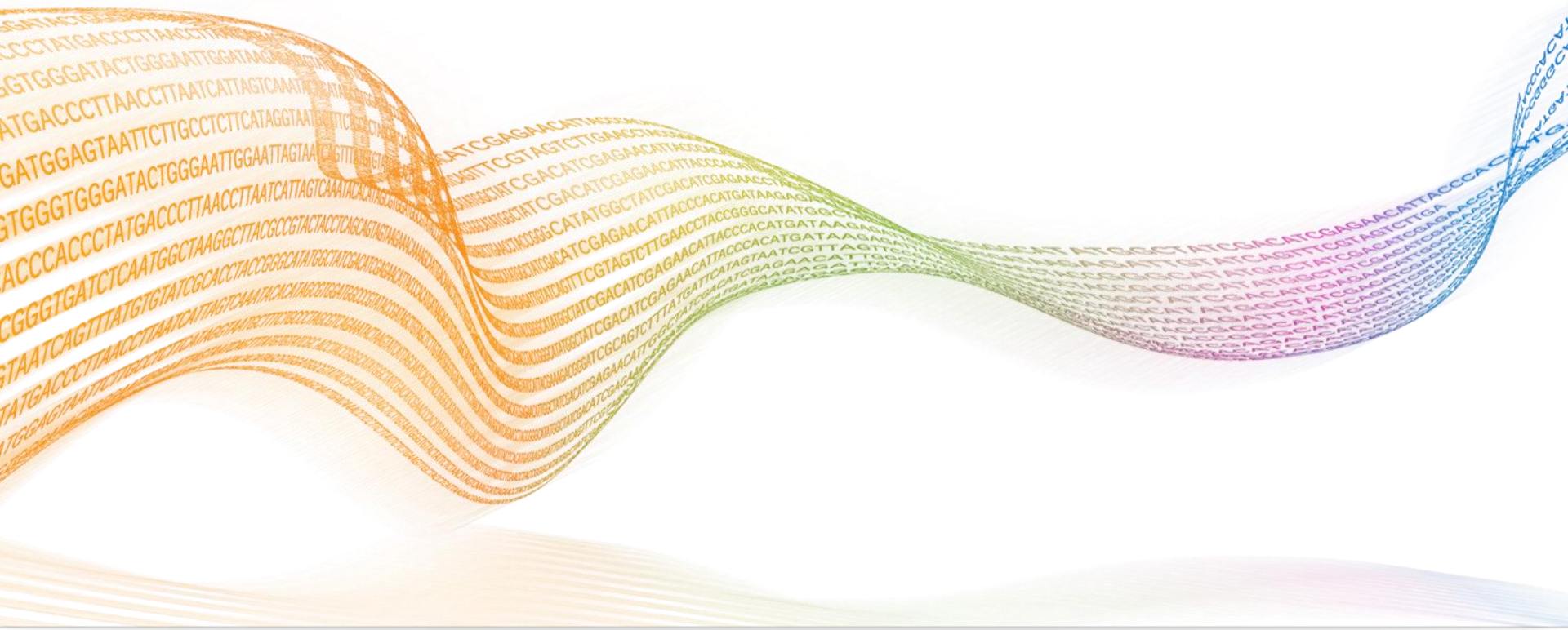
Genetic Ancestry Testing

Identification of Y, Mitochondrial and Autosomal variants in a single assay

- Y chromosome testing:
  - passed exclusively from father to son
  - can be used to explore ancestry in the direct male line (males only)
- Mitochondrial DNA testing:
  - mitochondrial DNA is passed on from mothers,
  - can be used by either sex
  - provides information about the direct female ancestral line
- Autosomal Single nucleotide polymorphism (SNP) testing:
  - capture the overall ethnic background of an individual
  - provide an estimate of a person's ethnic background.
  - SNPs can indicate that a person's mixed ancestry
  - 50% African, 25% European, 20% Asian, and 5% unknown

http://www.istockphoto.com/photo/a-pile-of-old-black-and-white-photographs-gm153079734-21383461?st=_p_photography%20family%20tree%20retro%20revival%20family

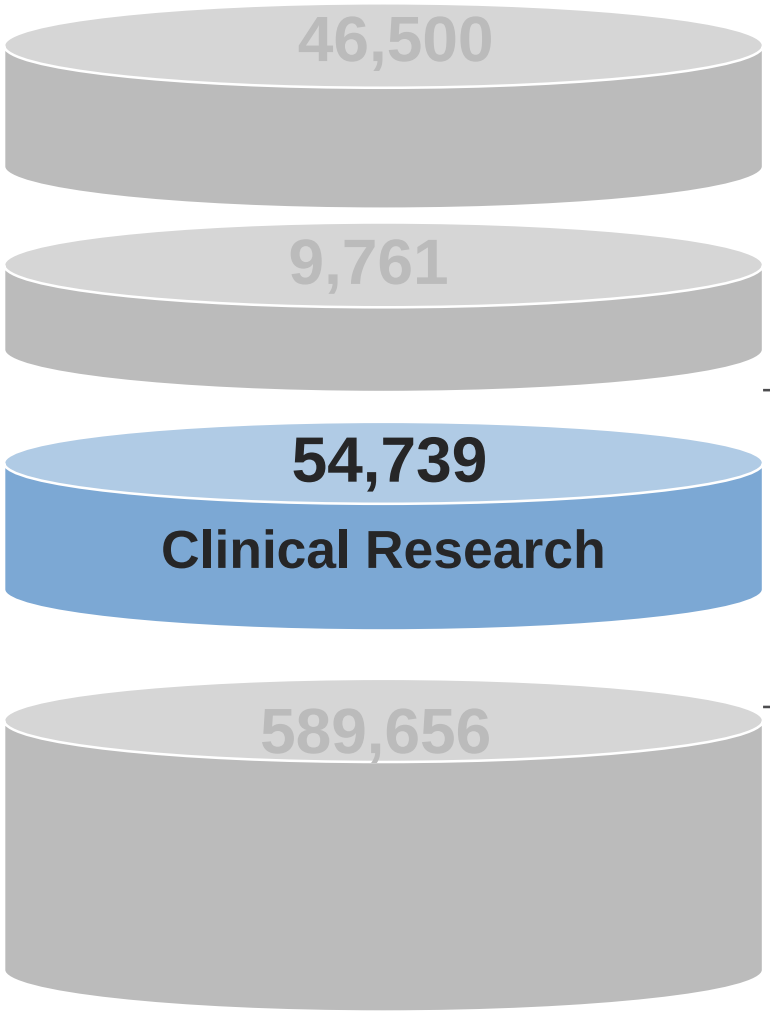
Thank you and Questions



illumina®

GSA Manifest variant count

Predictive, Clinical Research, and QC



- Up-to-date known clinical associations
- Pharmacogenomics
- Well-curated exome content
- NHGRI-GWAS and HLA content

illumina®

Expert Selected Clinical Research Content

Comprehensive set of known and putative clinical variants

Tier 1	Tier 2
Known clinical association	**Strong evidence for clinical association**
▸ Validate Disease Associations ▸ Risk profiling ▸ Pre-emptive screening ▸ PGx Studies (CPIC +)	▸ Establish functionality ▸ Support putative clinical associations ▸ Discover novel associations

Tier 1: Clinical Research Content Selection

Comprehensive collection of high value variants

Total Tier 1

Variants from GSA consortia

- eMERGE
- PharmGKB*
- Consortia provided

Variants from ClinVar**

- Include Pathogenic and likely pathogenic
- Exclude all somatic mutations
- Exclude variants missing clinical annotations
- Exclude variants missing review status
- Exclude benign and likely benign

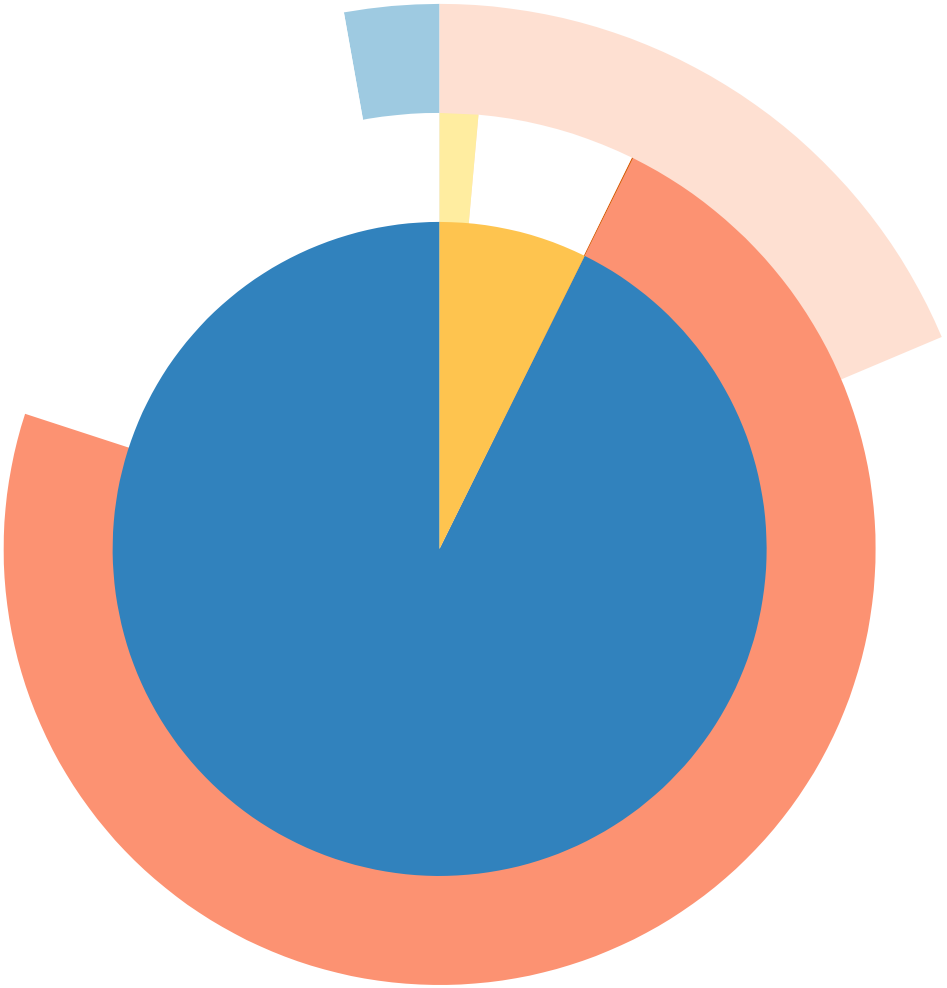
*https://www.pharmgkb.org
**http://www.ncbi.nlm.nih.gov/clinvar/

GSA: More Than Just a SNP Array

Ability to design breadth of functional variation

Illumina technology enables:

- Ability to design based on clinically-relevant CNVs
- >95% of selected variants designed

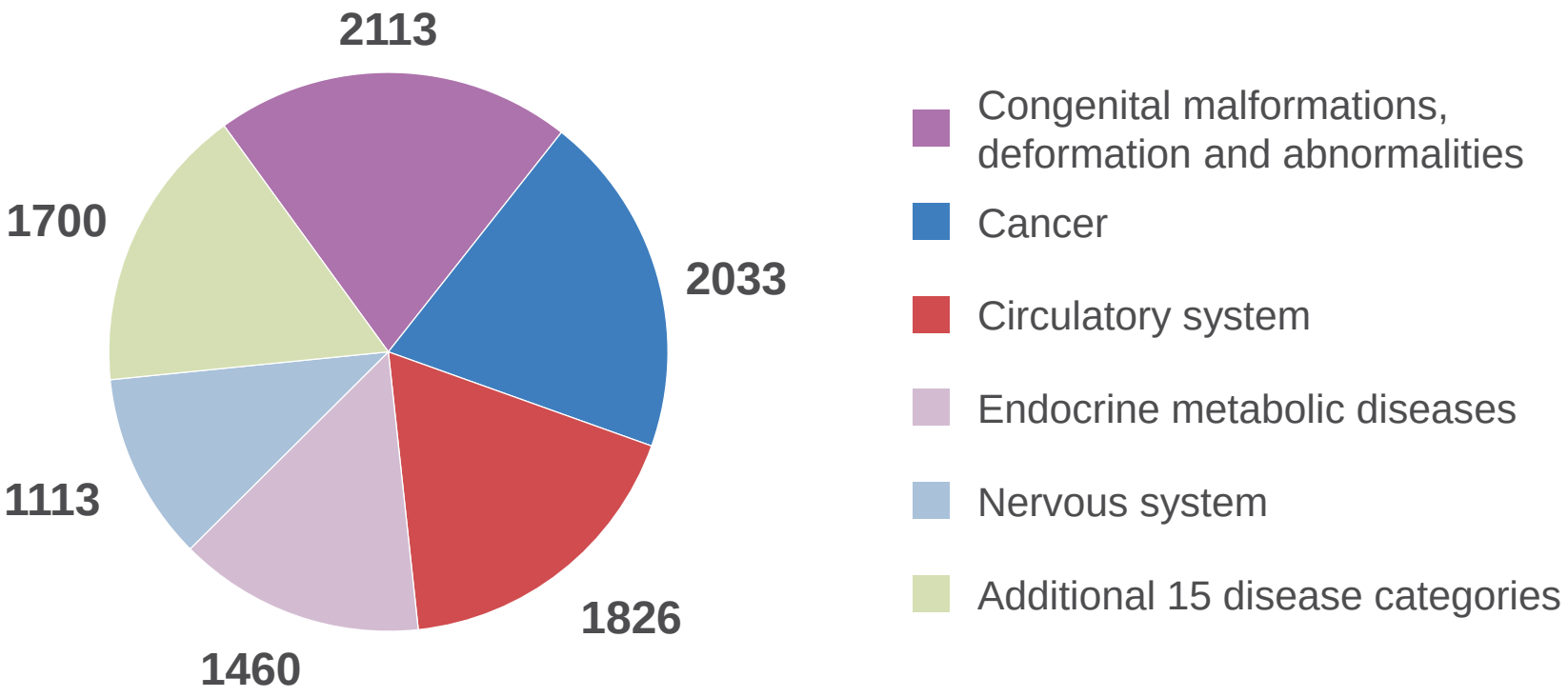


Data calculations on file. Illumina, Inc. 2016

illumina®

GSA: Wide Breadth of Highly Penetrant Variants

*Potential to detect >10,000 markers across 20 disease categories**



*ClinVar (135K): Reports of the relationships among human variations and phenotypes, with supporting evidence, www.ncbi.nlm.nih.gov/clinvar/

GSA: Discover Associations and Establish Functionality

>27K markers with strong evidence of association

Total Tier 2: ~39 K Variants

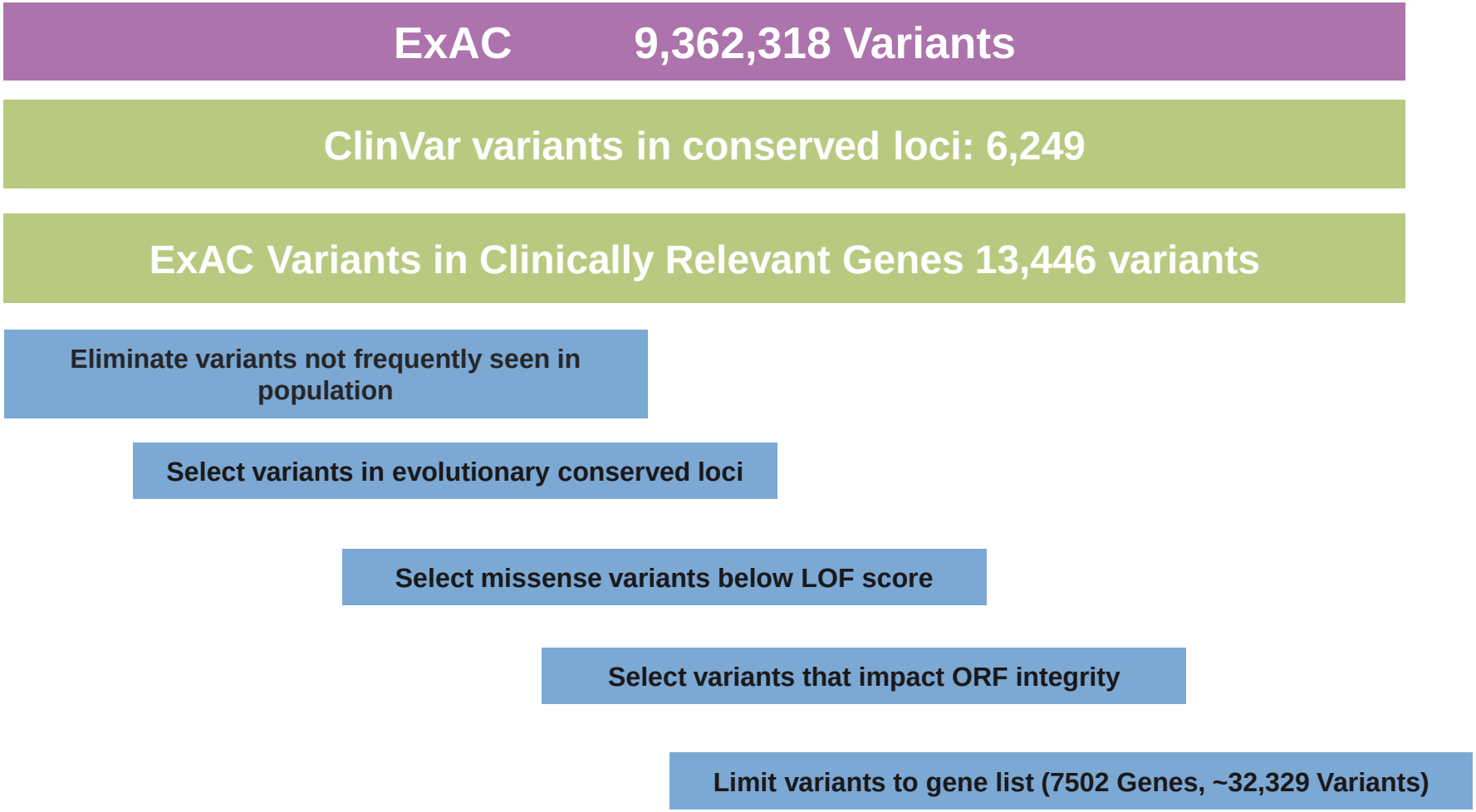
Putative disease & phenotype associations	Functionally interesting genes
~7,000	~32K
NHGRI-GWAS	**ExAC***
• Genome-wide associations with $p \leq 5.0E^{-8}$ • >700 diseases and traits	• All ClinVar variants in conserved loci • All non-tumor samples • MAF >1% • Conserved loci with loss of function in functionally interesting genes

*Most novel content not captured in ClinVar will come from ExAC

Data calculations on file. Illumina, Inc. 2016

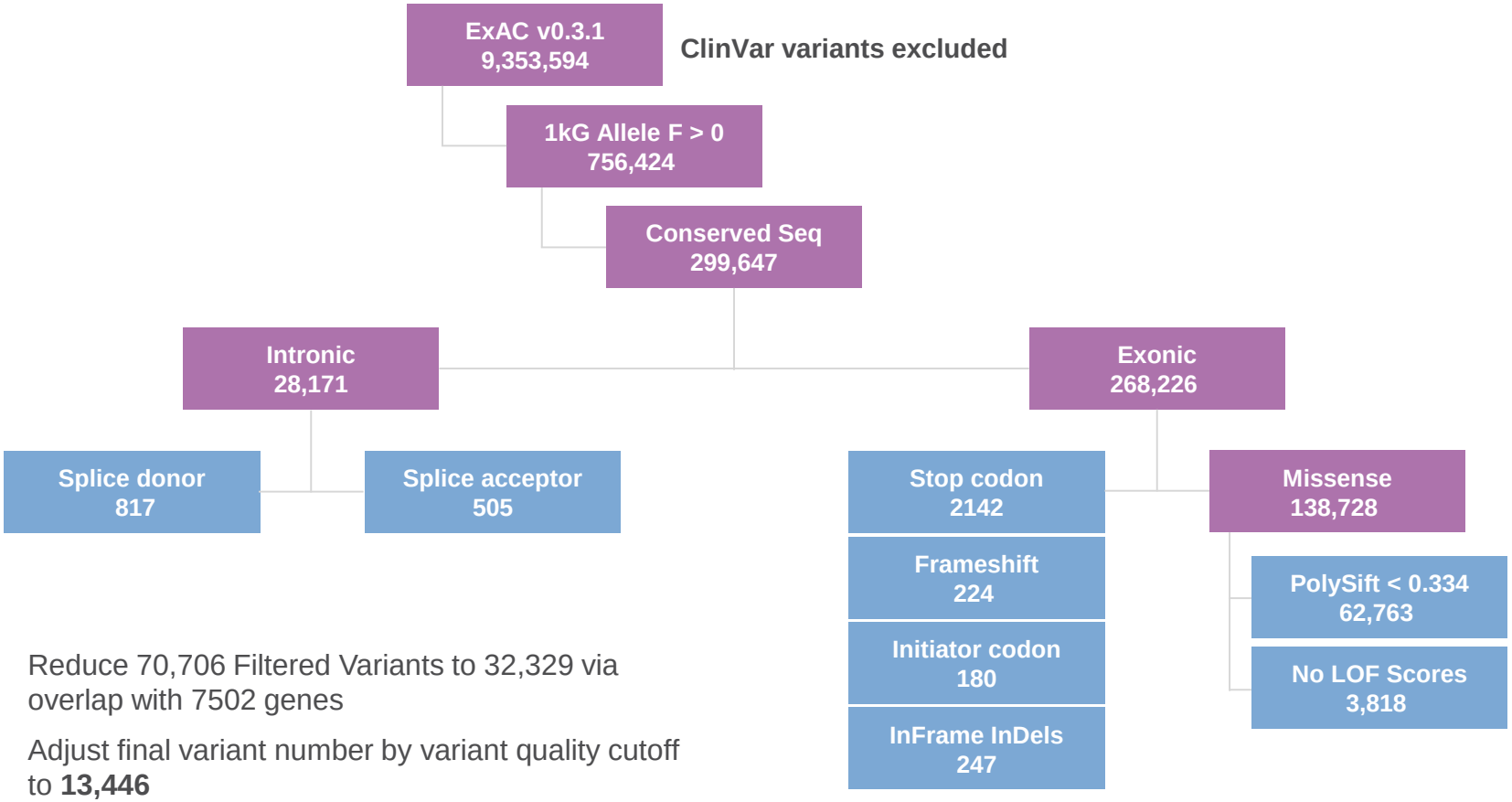
Selection of Highest Value Exome Variants

Potential for clinical associations



Filtering the Non-ClinVar ExAC Variants

Remove rare alleles & select conserved loci with probable LOF



- Reduce 70,706 Filtered Variants to 32,329 via overlap with 7502 genes
- Adjust final variant number by variant quality cutoff to **13,446**

Global Representation of ExAC Variants in GSA Clinical

Novel Exome variants for clinical and predictive research

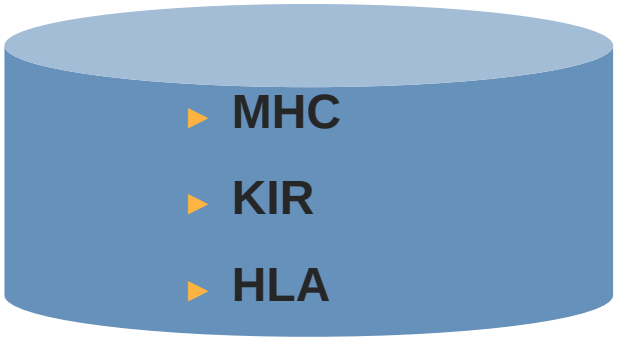
- A total > 30,000 ExAC variants were targeted on GSA
- Equal distribution of exclusive clinical variants across all populations
- Asian and African populations contribute highest number of exclusive variants

	ExAC variants in GSA Clinical Content					
1000 Genome Population	None	AMR	ASN	AFR	EUR	All 4
Exclusive variants detected only in	**4,855**	2,839	5,807	6,231	5,786	2,326
Total Variants present in selected Population	4,855	9,213	9,127	11,685	11,074	**27,991**

Data calculations on file. Illumina, Inc. 2016

HLA Content Optimized for Imputation

- 4,000 variants
- Best tag SNP's based on HLA imputation in available reference populations (European and Japanese)
- HLA collaborators report > 0.80 imputation accuracy for > 96% variants in Europeans

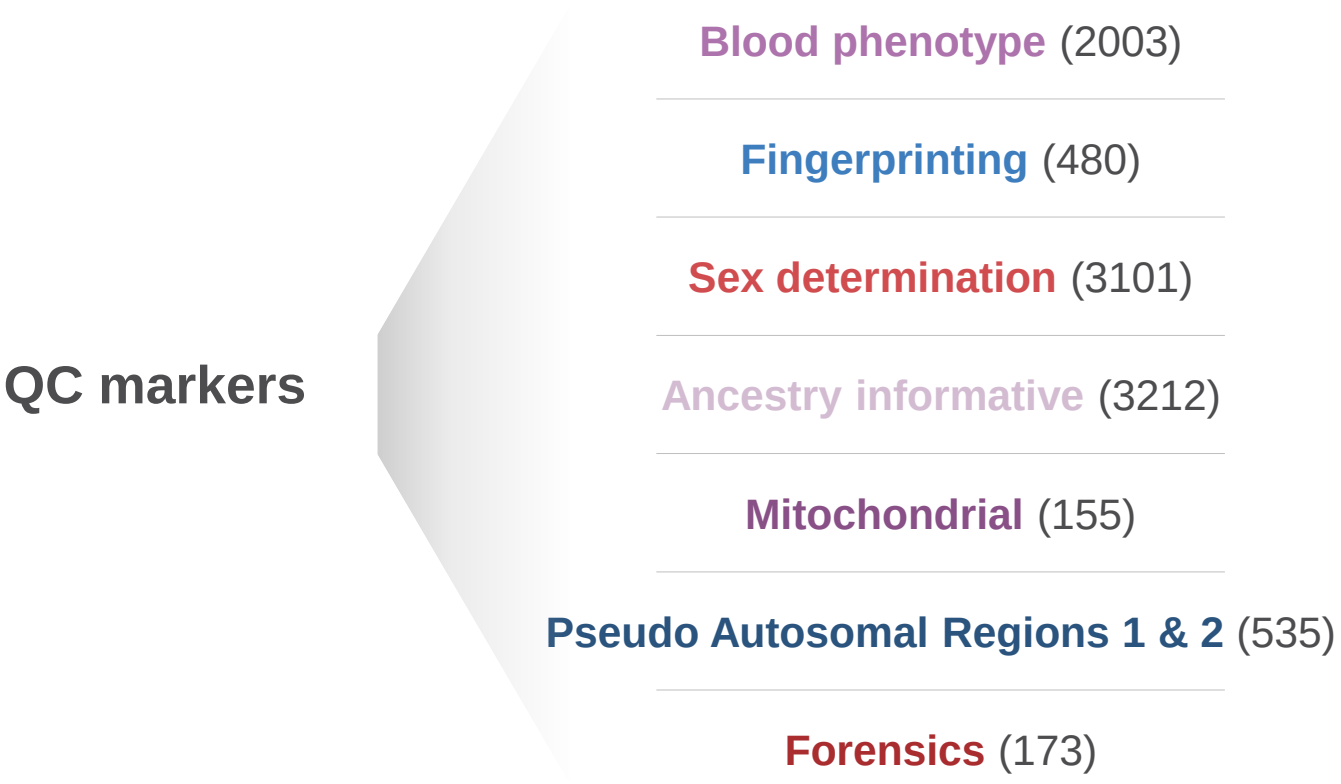


HLA Imputation Accuracy		
	Europeans	Asians
HLA-A	99.1%	98.1%
HLA-B	96.8%	65.6%
HLA-C	99.1%	68.8%
HLA-DQA1	98.5%	96.3%
HLA-DQB1	99.1%	96.5%
HLA-DRB1	96.9%	92.3%
All loci	98.3%	86.4%

**Adapted from Jia et. al., *Plos One* 2013

Infinium® GSA: Seamless Sample Tracking and Stratification

>9500 variants for sample tracking, QC, and stratification



Data calculations on file. Illumina, Inc. 2016

illumina®

GSA: Predefined Multi-Disease Content

Fine mapping of genome-wide significant loci

Derived from large-scale meta-analysis and exome content

Predefined 50K bead types* in lieu of custom content

Leverage established phenotype-specific consortia

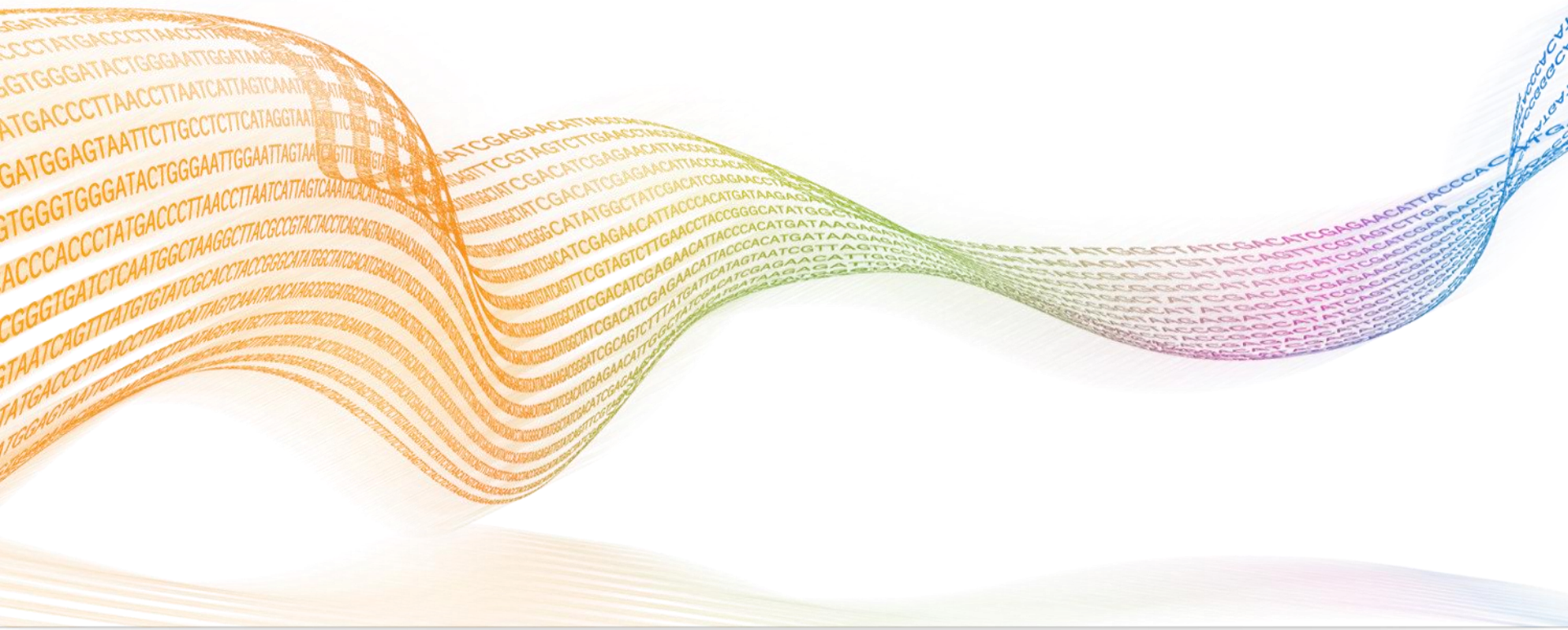
- Cardiometabolic
- Autoimmune
- Psychiatric
- Neurological
- Cancer
- Anthropometric

*Available at additional cost

Acknowledgements

Ben Neale Ph.D.	Broad Institute
Eimear Kenny Ph.D.	Mount Sinai
Stephen Chanock Ph.D.	NCI
Mitch Machiela Ph.D.	NCI
Heidi Rehm Ph.D.	Harvard Partners
Jouke Hottenga Ph.D.	University of Amsterdam
Gail Jarvik M.D., Ph.D.	University of Washington
All Consortia Members	

Thank you and Questions



illumina®

Preliminary Content List Disclaimer

- The data and content in these slides is considered preliminary. We want to ensure that everyone is fully aware that this will not be representative of the ‘final’ marker list that will be in the released product but rather contains all genomic locations which have been selected for inclusion on the BeadChip. However, due to additional testing and validation, some of these locations might not be available for typing on the final product.

- The information being shared is considered confidential. The marker list will be available for review under CDA.