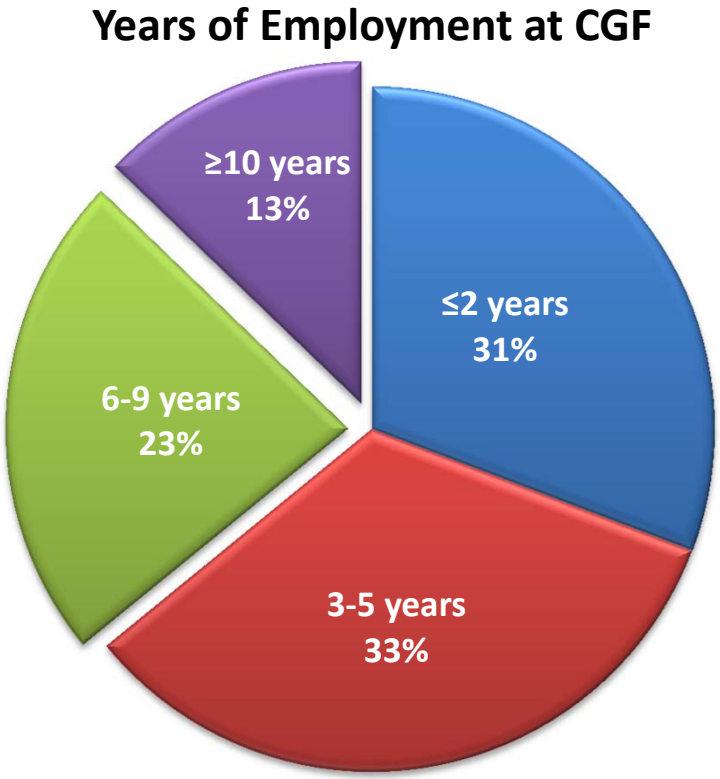
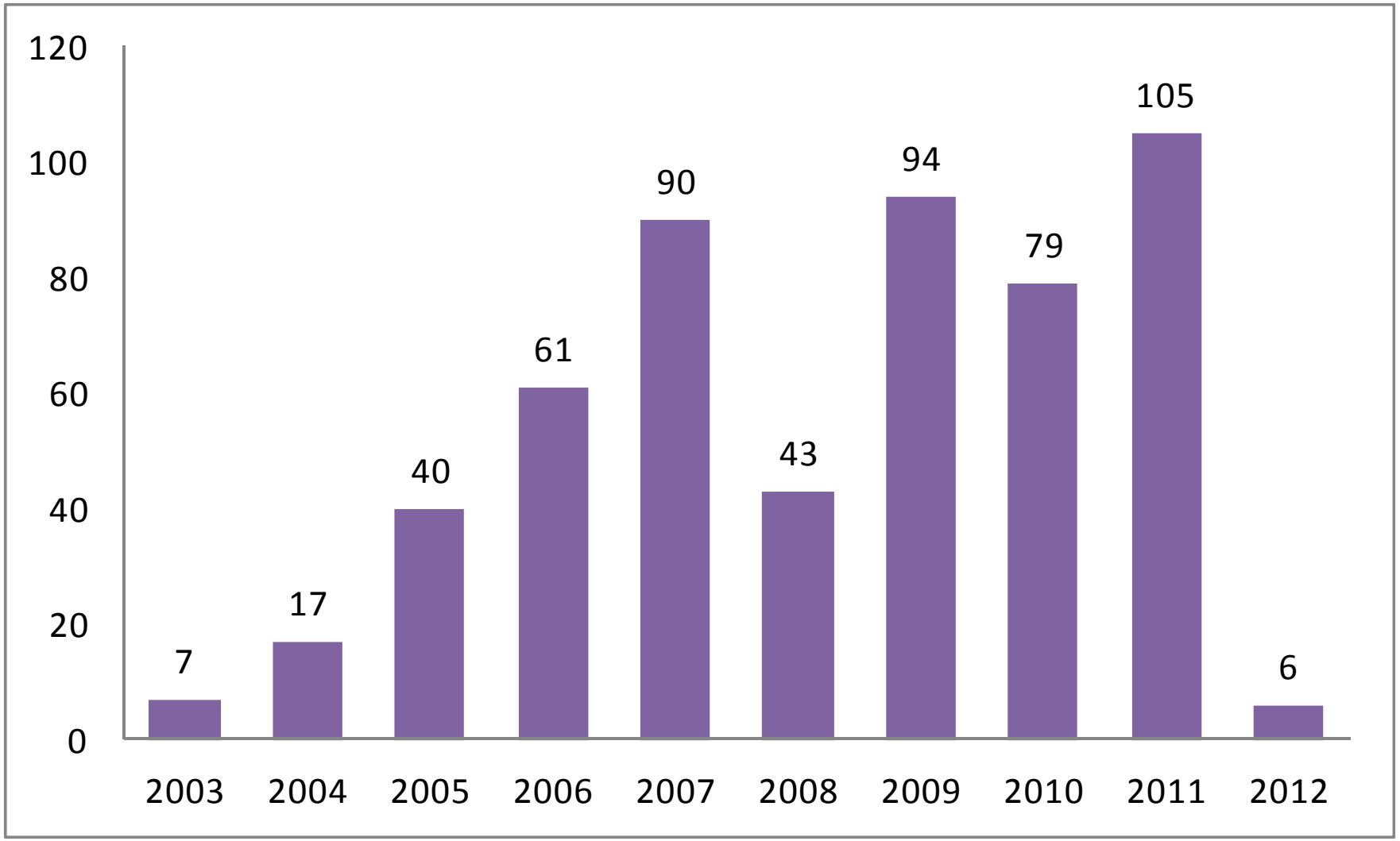
Possible Questions

Stability of CGF Staff



Number of CGF-coauthored publications per year



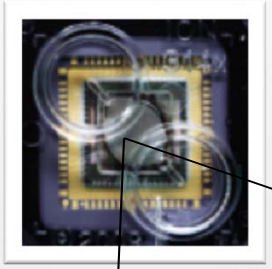
Technology Assessment

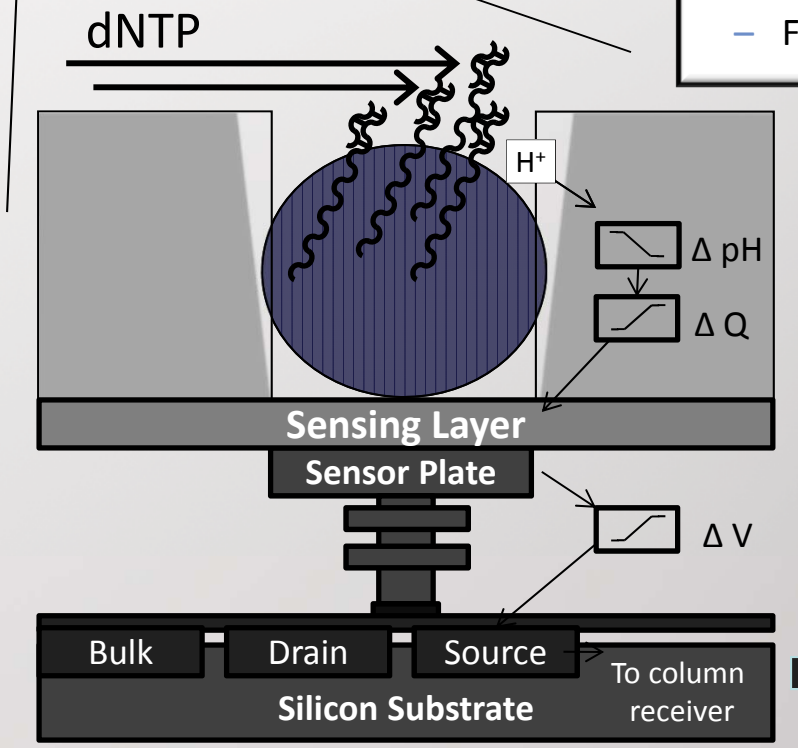
- Collaboration with Academic and Commercial Laboratories
  - Early Access
  - Rapid Evaluation of Emerging Technologies
    - 15 Projects
  - Assist in DCEG PIs in Application& Study Design
  - Translation to Production Capacity
  - Prevent Waste of Biospecimens and Resources

R & D Projects @ CGF

- WGA Kits
- Fluidigm
  - Biomark
  - Access Array
- Illumina β testing
  - Infinium/Omni
  - Methylation
- Exome Capture
- Raindance
- Ion Torrent
- ABI
  - SNapShot, SNPlex
  - DME Panels
- EPOCH
- Sequenom* (1st gen)
- 454- Exome
- Affymetrix*
- Illumina-HiScan

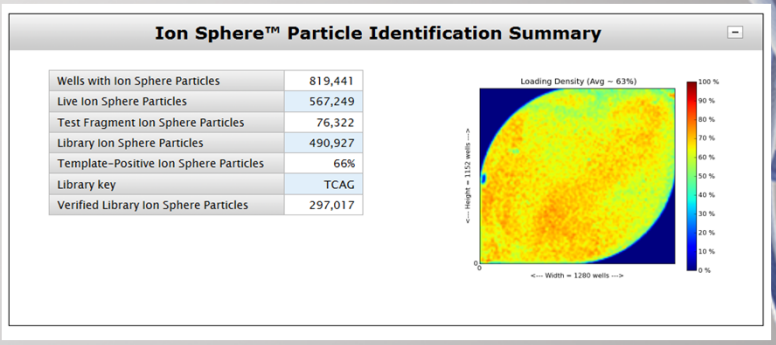
Ion Torrent Technology

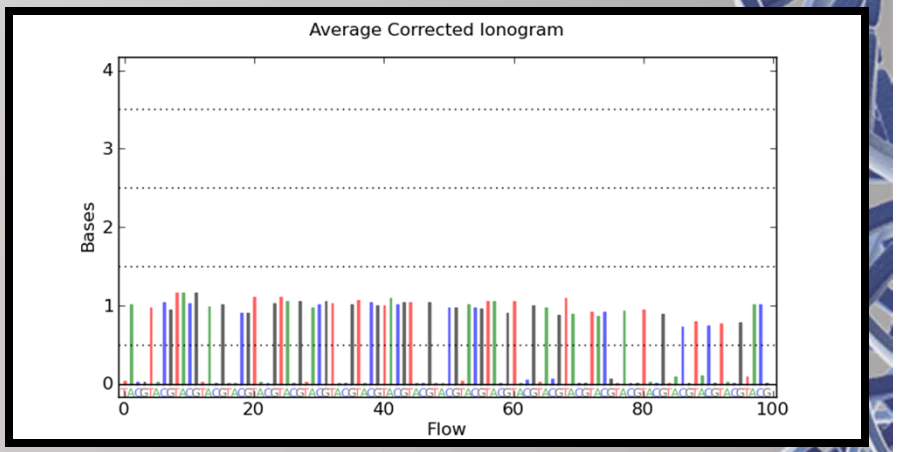




DNA → Ions → Sequence

- Nucleotides flow sequentially over Ion semiconductor chip
- One sensor per well per sequencing reaction
- Direct detection of natural DNA extension
- Millions of sequencing reactions per chip
- Fast cycle time, real time detection

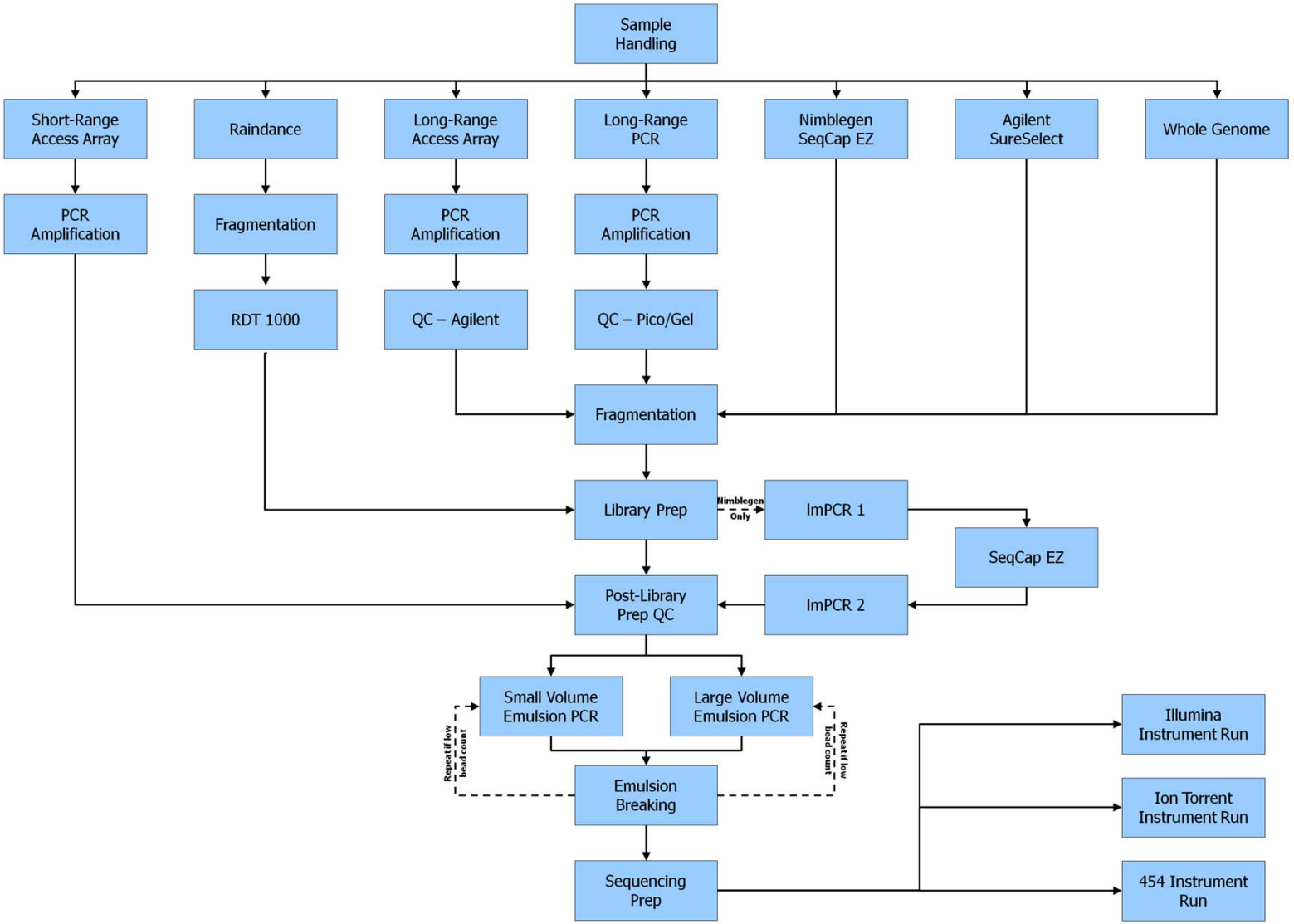




Ion Torrent Investment: ‘Small Job Shop’

- Confirmation of exome and targeted variant sequencing
- Reduced cost
- Rapid turnaround: 2 week total @ 12-18 chips/day
- Custom activities
    1. Large amplicon / highly-multiplex sample studies
    2. RNAseq studies
        a. whole transcriptome
        b. small RNA
    3. Fixed or custom amplicon panels for preclinical sample and tumor profiling
    4. FFPE sample sequencing
    5. Rapid exome sequencing and supplementation
    6. Methylseq (RainDance and/or Ion reagents)

CGF Sequencing Workflows

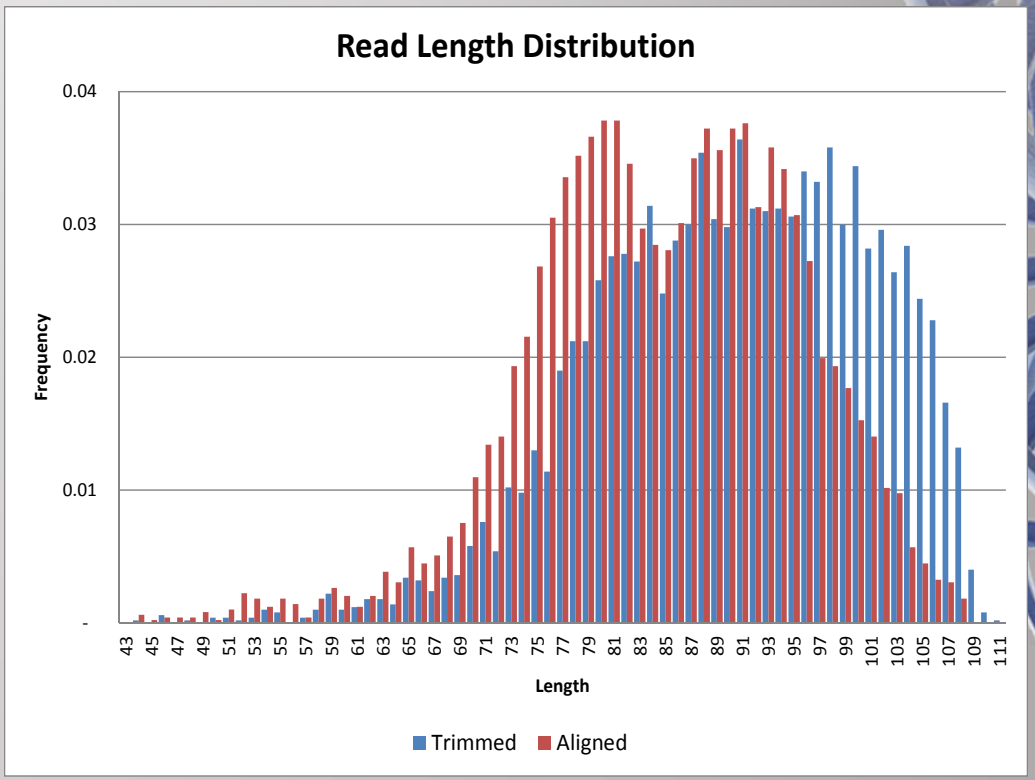


CGF Bioinformatics & Scientific Operations

Sequencing Informatics

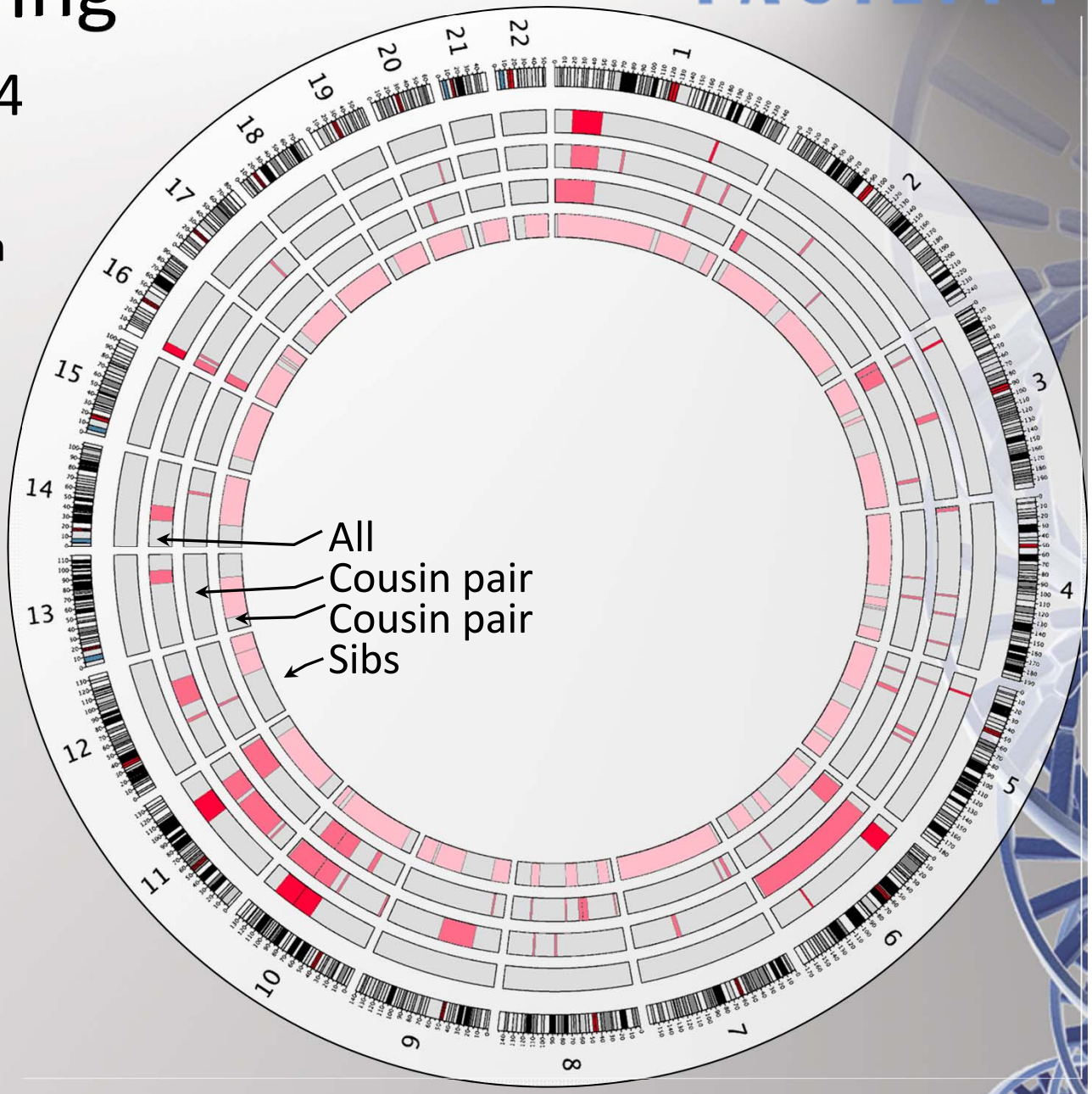
2 staff

- **Sequencing Data Analysis**
- **Working on development of standard analysis pipeline**
- **Feedback metrics and reports to the laboratory**



IBD Sharing

- Based on 3,277,154 autosomal SNPs
  - Biallelic, polymorphic, in dbSNP, Mendelian consistent
- Methods
  - Phasing: BEAGLE
  - IBD sharing: Germline
  - Plot: Circos



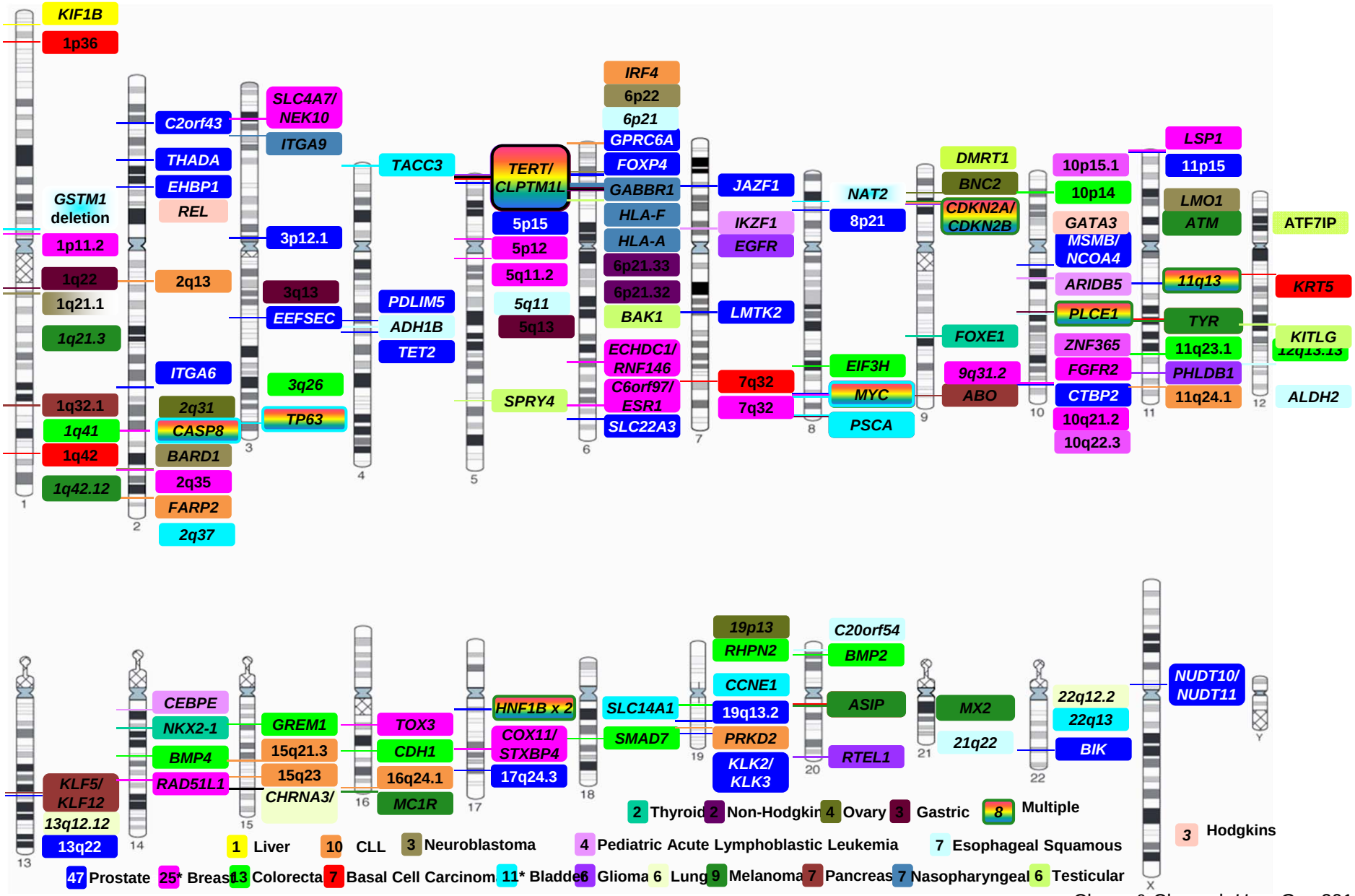
DCEG GWAS Available on dbGaP
Shift from caBIG to dbGaP in 2009

CGEMS
Cancer Genetic Markers of Susceptibility

GWAS Site	# Approved
Breast*	134
Prostate*	92
Pancreas	88
Lung	118
Bladder	16
Renal	15
UGI (China)	12
Imputation	7

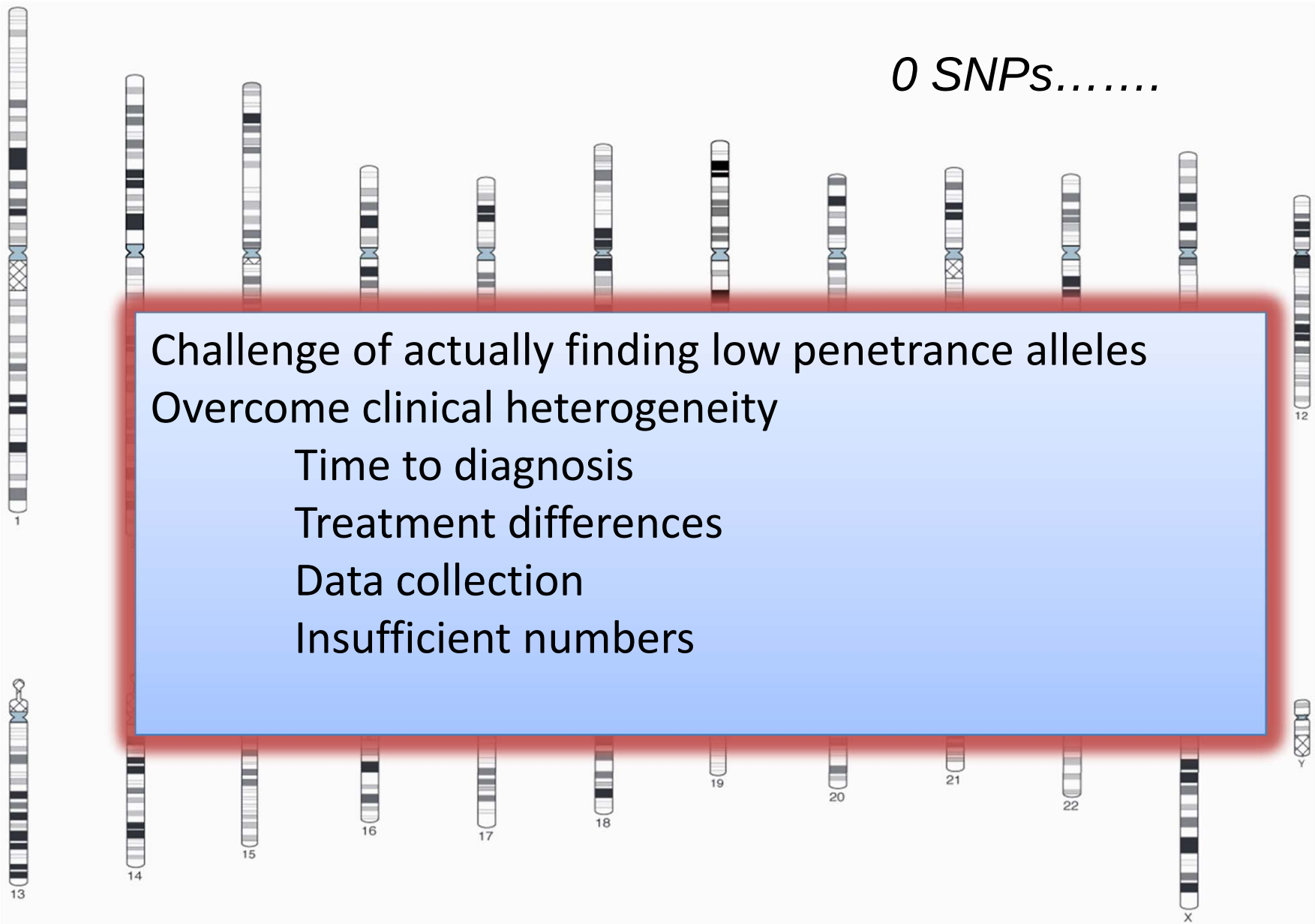
Previously on CGEMS Site with > 100 for each

5 years of cancer GWAS – 216 signals for 24 cancer types

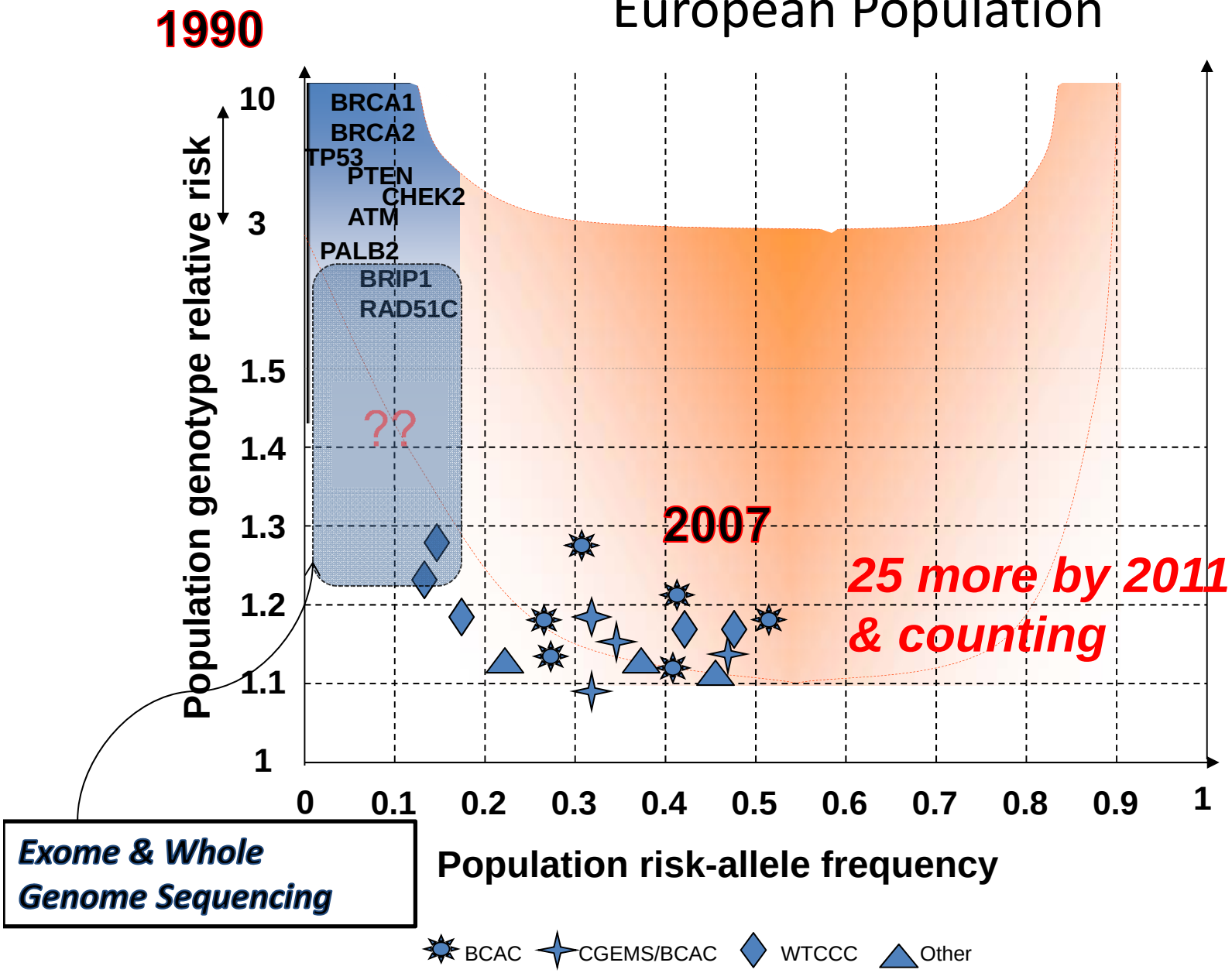


Chung & Chanock *Hum Gen* 2011

CANCER GWAS Hits for SURVIVAL or OUTCOME



Genetic Predisposition to Breast Cancer
European Population



Theoretical Limits of Risk Prediction

Crohn's Disease
Sibling relative-risk=20-35

Common cancers (Br or Pr)
Sibling relative risk=2-3

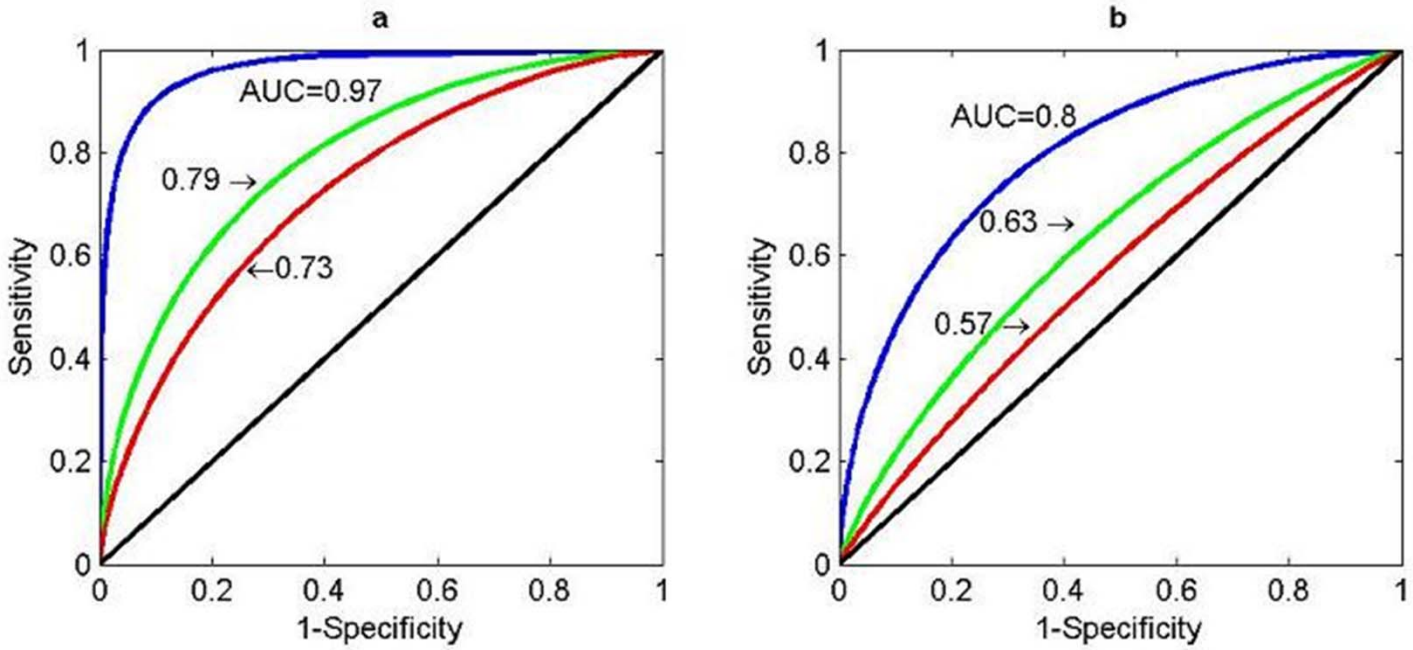


Figure 2 Receiver operator characteristics (ROC) curves for genetics risk models.

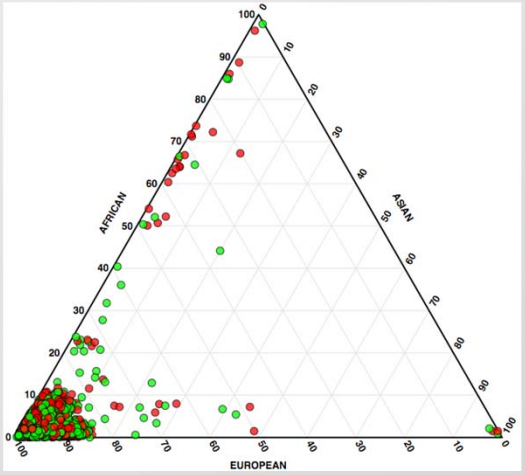
- **Random**
- **Using known loci**
- **Using all estimated loci**
- **Ideal (if we could explain all heritability)**

Park et al Nature Genetics 2010

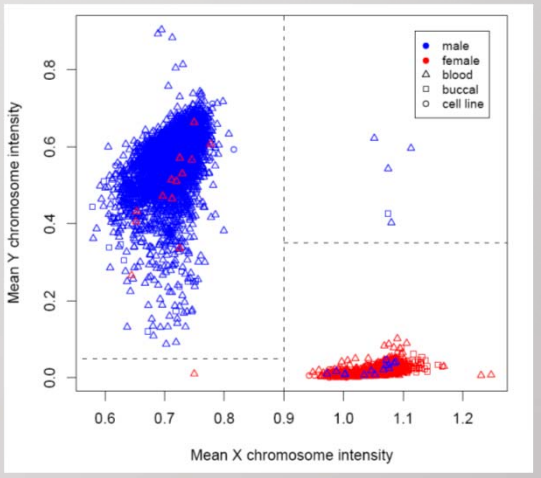
Quality Control

Genome-wide association studies

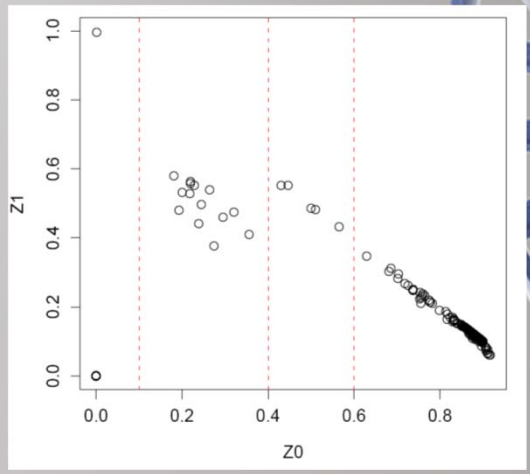
Population genetics and ancestry

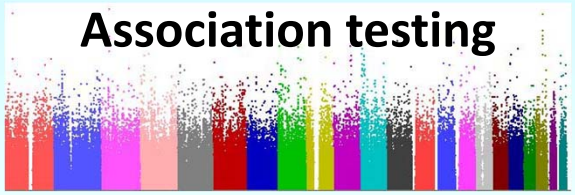


Sex verification



Relationship testing



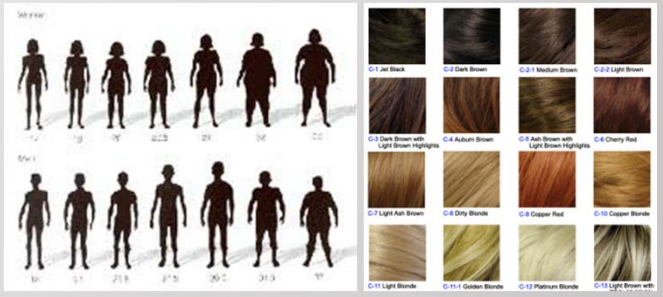


Genome-wide association studies

Behavioral traits

Tobacco,
Caffeine,
Alcohol

Biometrics



Height, Weight, BMI,
Menarche/Menopause
Hair and eye color

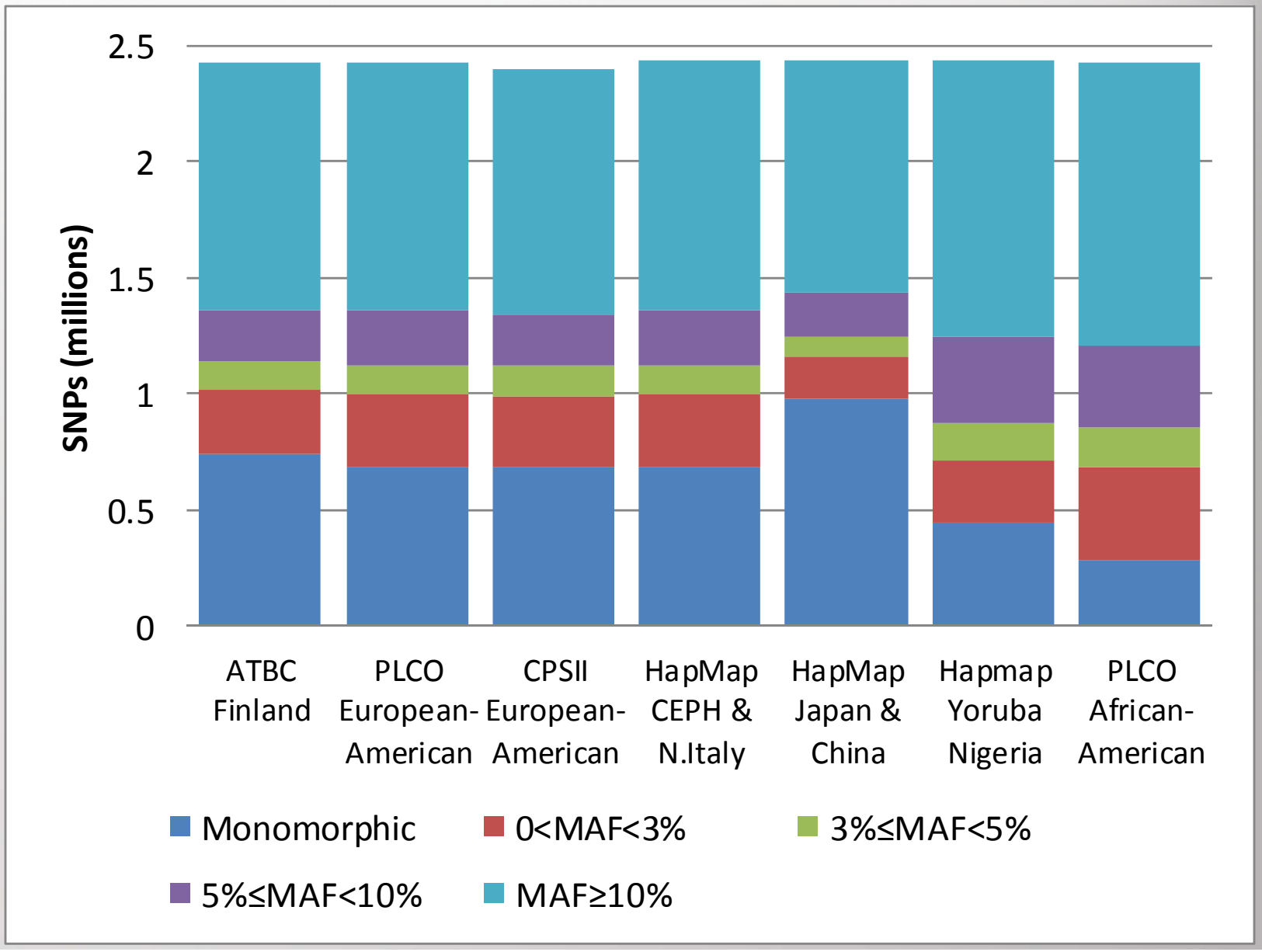
Nutrient levels

Vitamins D,B_{12}
Carotene, etc.

…and much more!

DCEG & International Consortia (e.g., GIANT, SUNLIGHT)

Omni2.5 allele frequency distribution



OPEN ACCESS Freely available online

PLoS BIOLOGY

Rare Variants Create Synthetic Genome-Wide Associations

Samuel P. Dickson[1,2], Kai Wang[3], Ian Krantz[3,4,5], Hakon Hakonarson[3,4,5], David B. Goldstein[1]*

1 Institute for Genome Sciences and Policy, Center for Human Genome Variation, Duke University, Durham, North Carolina, United States of America, **2** Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, United States of America, **3** Center for Applied Genomics, Children's Hospital of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **4** Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **5** Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, United States of America

Really?

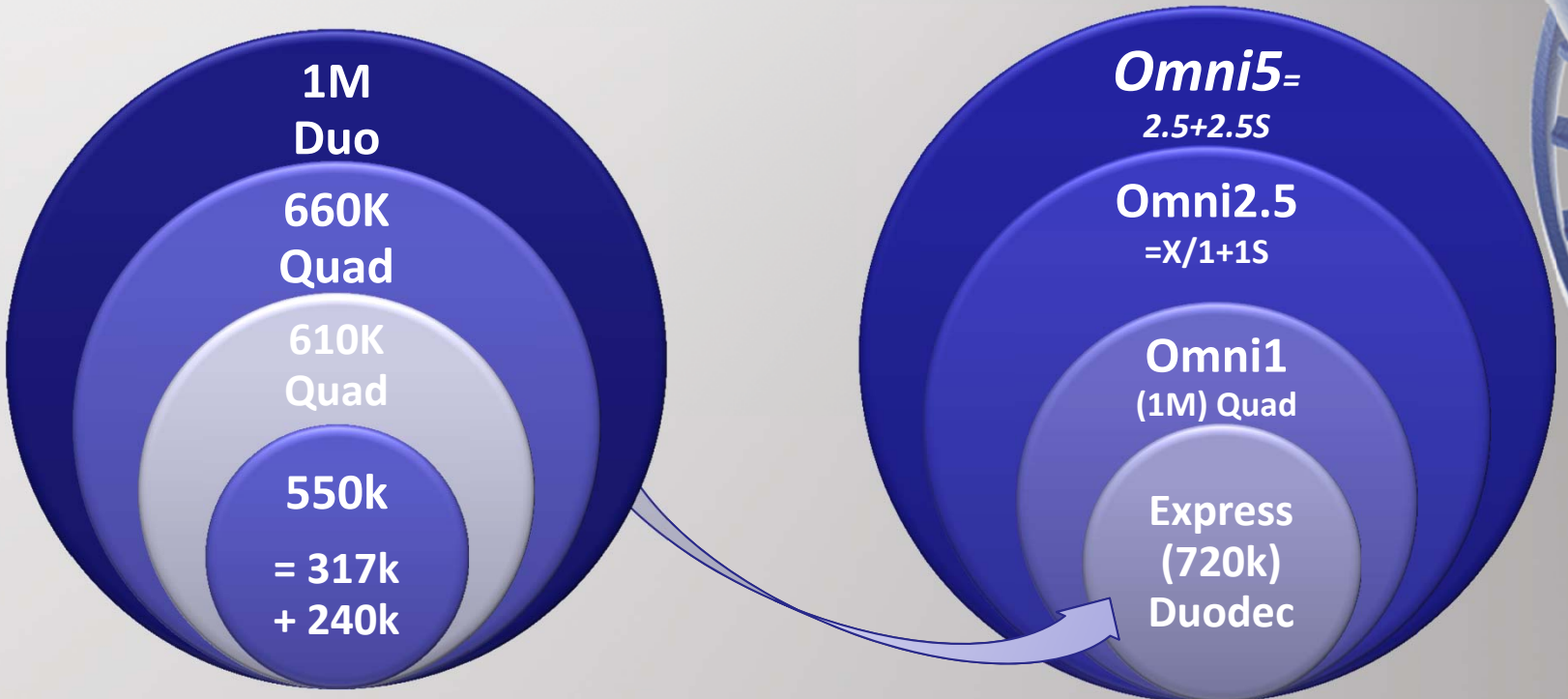
A single nucleotide polymorphism tags variation in the arylamine *N*-acetyltransferase 2 phenotype in populations of European background

Montserrat García-Closas[a,l], David W. Hein[d], Debra Silverman[a], Núria Malats[k], Meredith Yeager[a,b], Kevin Jacobs[a,b], Mark A. Doll[d], Jonine D. Figueroa[a], Dalsu Baris[a], Molly Schwenn[e], Manolis Kogevinas[g,h,i,m], Alison Johnson[f], Nilanjan Chatterjee[a], Lee E. Moore[a], Timothy Moeller[c], Francisco X. Real[l,j], Stephen Chanock[a,b] and Nathaniel Rothman[a]

Pharmacogenet Genomics. 2011 Apr;21(4):231-6.

Illumina GWAS Capacity

- Current capacity is ~432 Infinium arrays/week
  - 1,728 samples/week for quad arrays (660k, Omni1, Omni5)
  - 3,456 samples/week for octo arrays (Omni1S, 2.5, 2.5S)
  - 5,184 samples/week for duodec arrays (iSelect/OmniX)

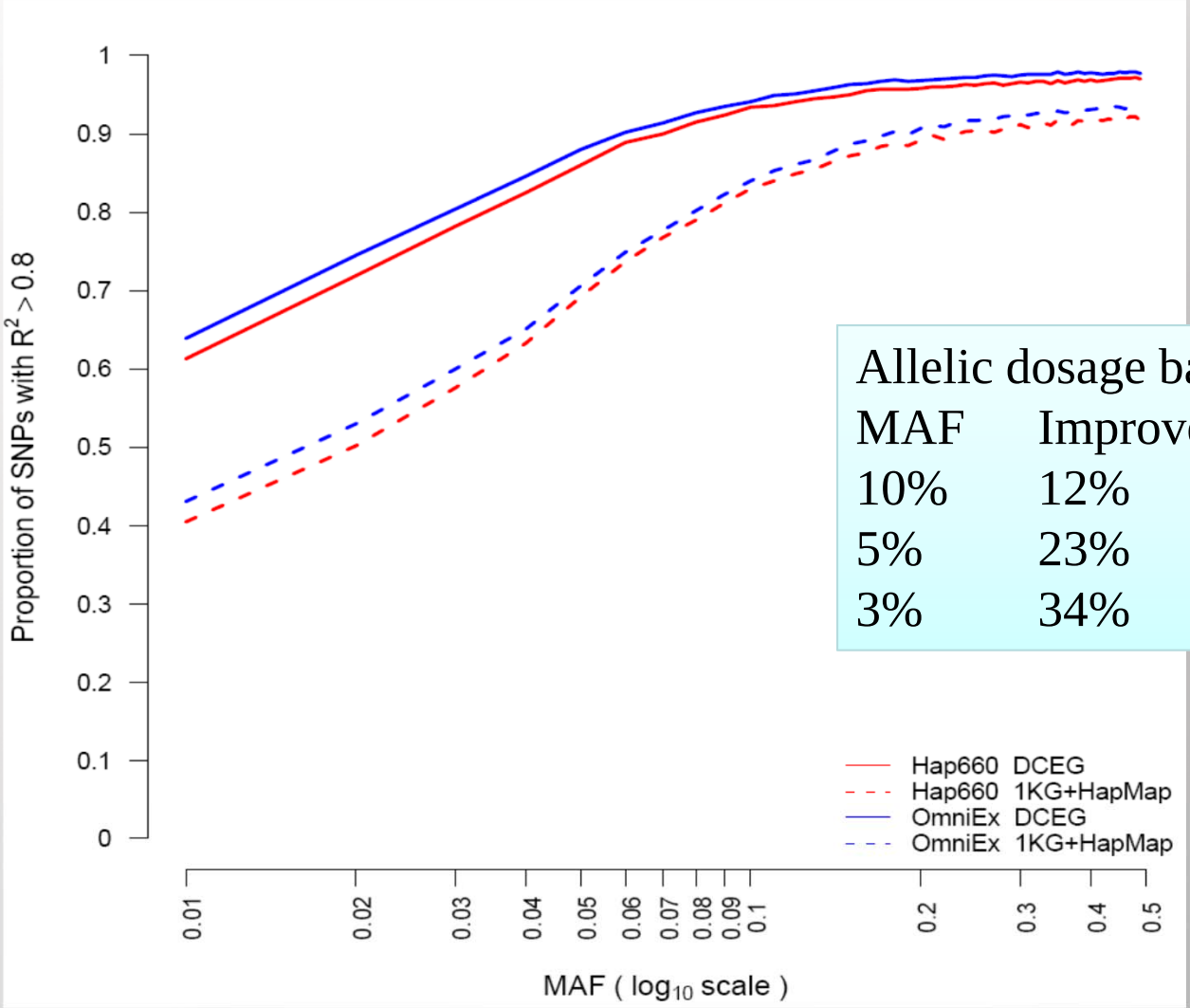


Samples included in Build 1 of DCEG Imputation Reference Set

	Populations				Illumina Array			
Group	**European American**	**African American**	**African**	**Asian**	**Hap660**	**Hap1**	**Omni1**	**Omni2.5**
ATBC	246					✓	✓	✓
CPSII	227					✓	✓	✓
PLCO	255					✓	✓	✓
PLCO		98				✓		✓
SHNX				74	✓			✓
HapMap								
CEU	116							✓
CHB				44				✓
JPT				44				✓
TSI	86							✓
YRI			59					✓
Total	930	98	59	162				

Available in dbGaP in October 2011

Imputation accuracy for European-American data with DCEG and public reference set

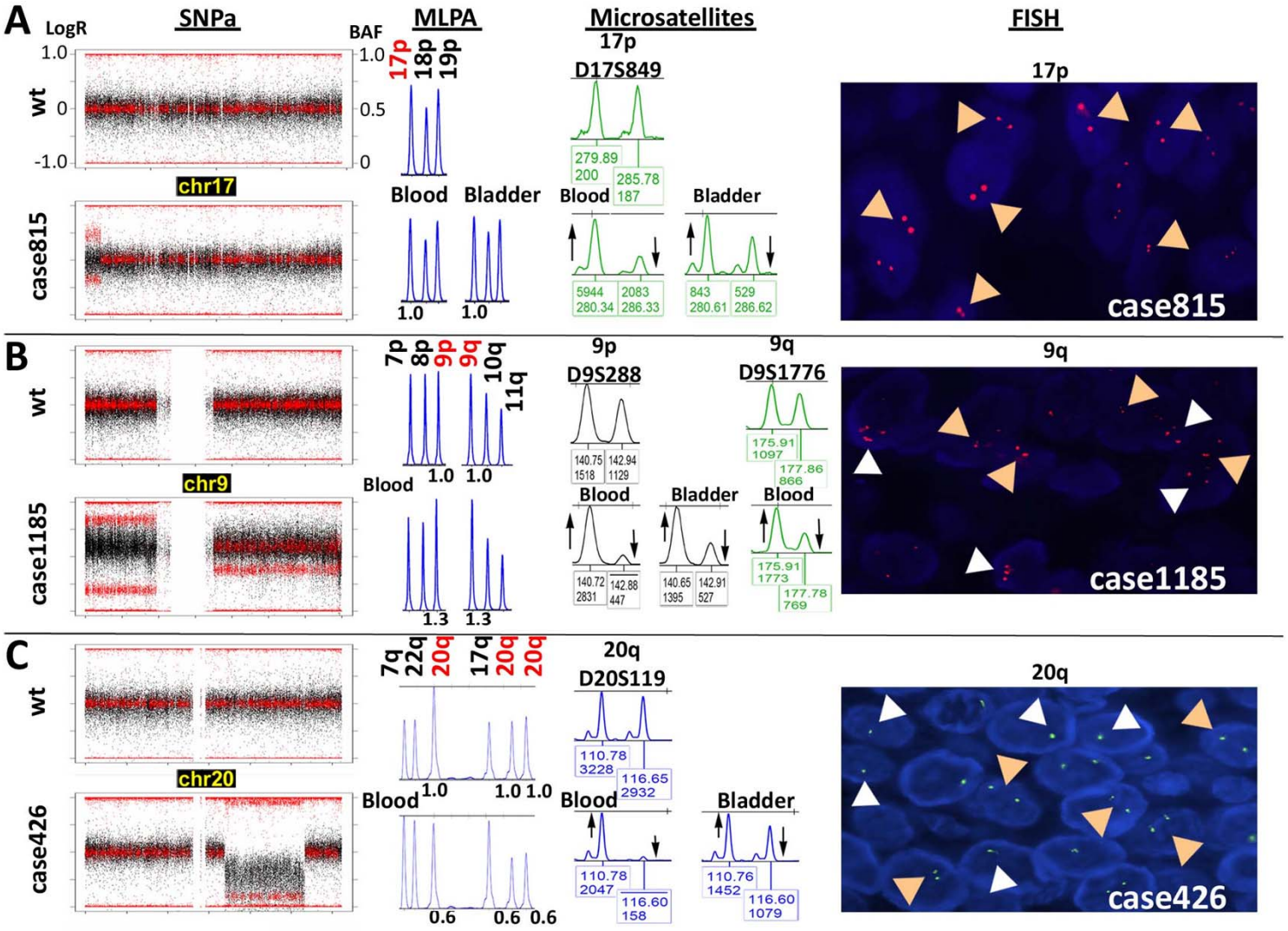


Allelic dosage based $R^2 > 0.8$

MAF	Improvement
10%	12%
5%	23%
3%	34%

3 sets of 60 samples using IMPUTE2 (Confirmed with BEAGLE)
R2 (pearson) for correlation

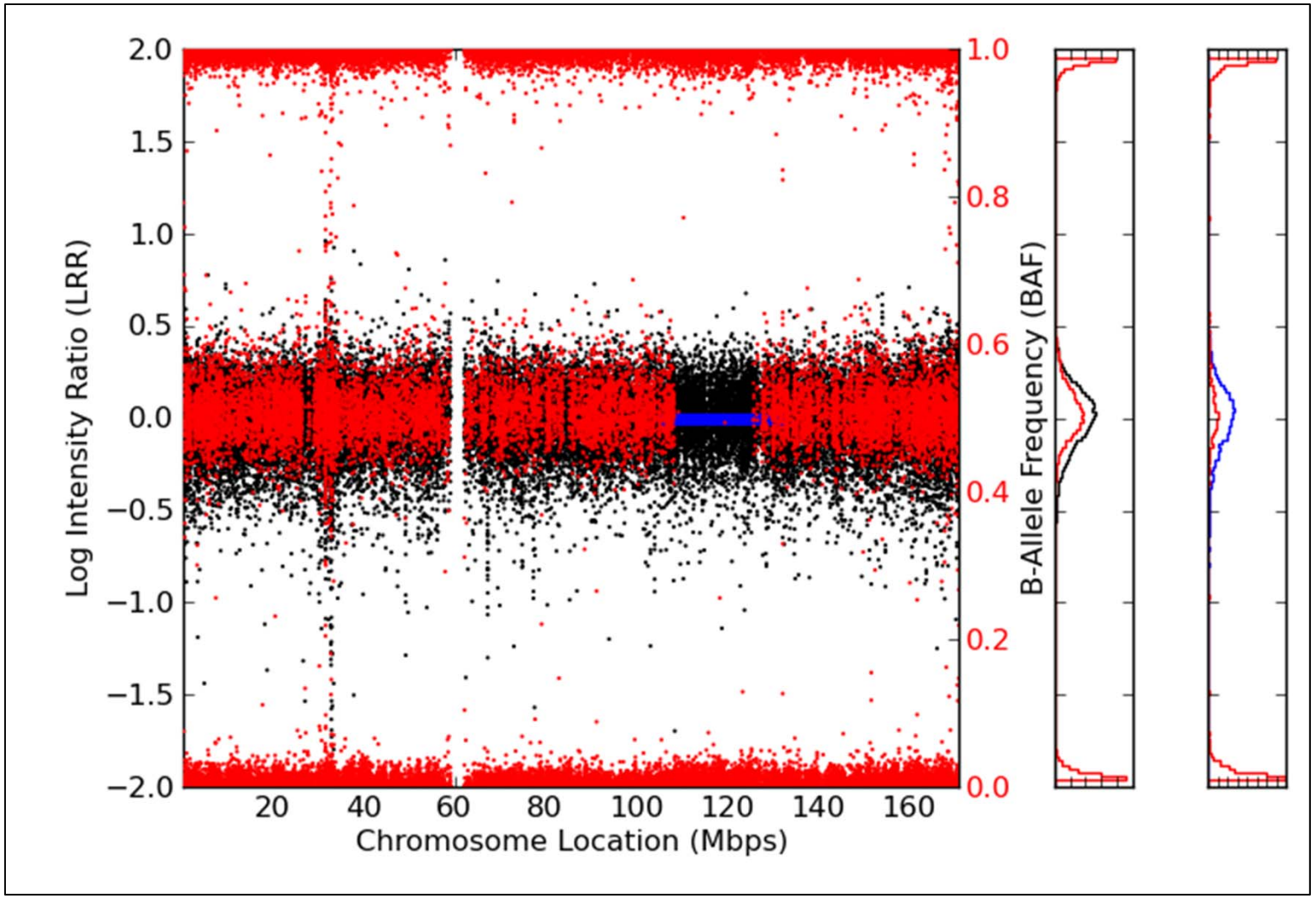
Validation for 42 events: 100% Validation



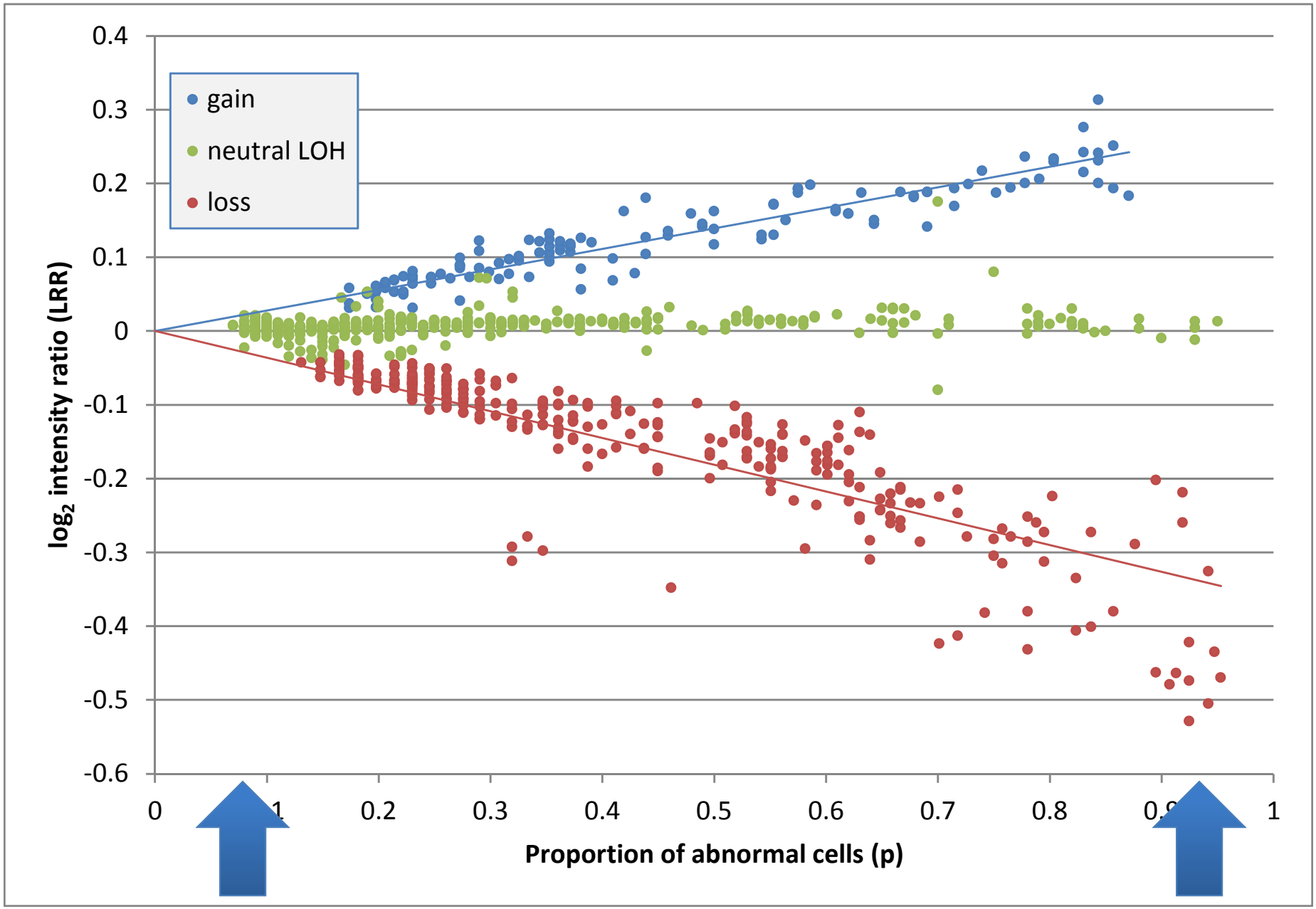
Rodríguez-Santiago et al. *Am J Hum Genet*. 2010;87:129-38

Segment inherited identical by decent

(probably not *de novo* LOH)



Genetic Mosaic Events



Frequency of mosaic events by type & location

	Mosaic Chromosome Count					Mosaic Chromosome Frequency (%)				
Event Location	**gain**	**loss**	**cnloh**	**mixed**	**Total**	**gain**	**loss**	**cnloh**	**mixed**	**Total**
chromosome	62	11	42	5	**120**	9.7	1.7	6.6	0.8	**18.7**
telomeric P	11	13	114	1	**139**	1.7	2.0	17.8	0.2	**21.7**
telomeric Q	9	10	149	0	**168**	1.4	1.6	23.2	0.0	**26.2**
interstitial	14	185	2	1	**202**	2.2	28.9	0.3	0.2	**31.5**
span centromere	1	1	2	0	**4**	0.2	0.2	0.3	0.0	**0.6**
complex	0	3	0	5	**8**	0.0	0.5	0.0	0.8	**1.2**
Total	**97**	**223**	**309**	**12**	**641**	**15.1**	**34.8**	**48.2**	**1.9**	

Adjusted Analysis of Association Between Genetic Mosaicism and Cancer in 49 studies

	All Cancer Cases			Likely Untreated			Possibly Treated		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P Value
Non-heme Cancers	1.27	1.05-1.52	0.012	1.45	1.18-1.80	5.4E-04	1.03	0.81-1.30	0.804

Preliminary evidence for
Lung & kidney cancer

Early Detection of Hematological Cancers as Genetic Mosaicism

	Mosaic Counts			Non-Mosaic Counts			Mosaic Frequency (%)		
	Untreated	Possibly Treated	Total	Untreated	Possibly Treated	Total	Untreated	Possibly Treated	Overall
hematologic cancer	**9**	**9**	**18**	**34**	**62**	**96**	**20.93**	**12.68**	**15.79**
leukemia	**9**	**8**	**17**	**34**	**11**	**45**	**20.93**	**42.11**	27.42
lymphocytic	5	4	9	14	5	19	26.32	44.44	32.14
myeloid	3	4	7	16	5	21	15.79	44.44	25.00
other/nos	1	0	1	4	1	5	20.00	0.00	16.67
lymphoma	0	1	1	0	42	42		2.33	2.33
multiple myeloma	0	0	0	0	9	9		0.00	0.00

- For untreated leukemia vs. cancer-free controls
  - DNA collected at least one year prior to diagnosis
  - OR=35.4 (14.7-76.6 95% CI), $p=3.8 \times 10^{-11}$
- DNA was obtained >5 years prior to diagnosis for 6 mosaic individuals, with the longest interval being 14 years