Next Generation Sequencing

Epidemiology 217: Lecture 10

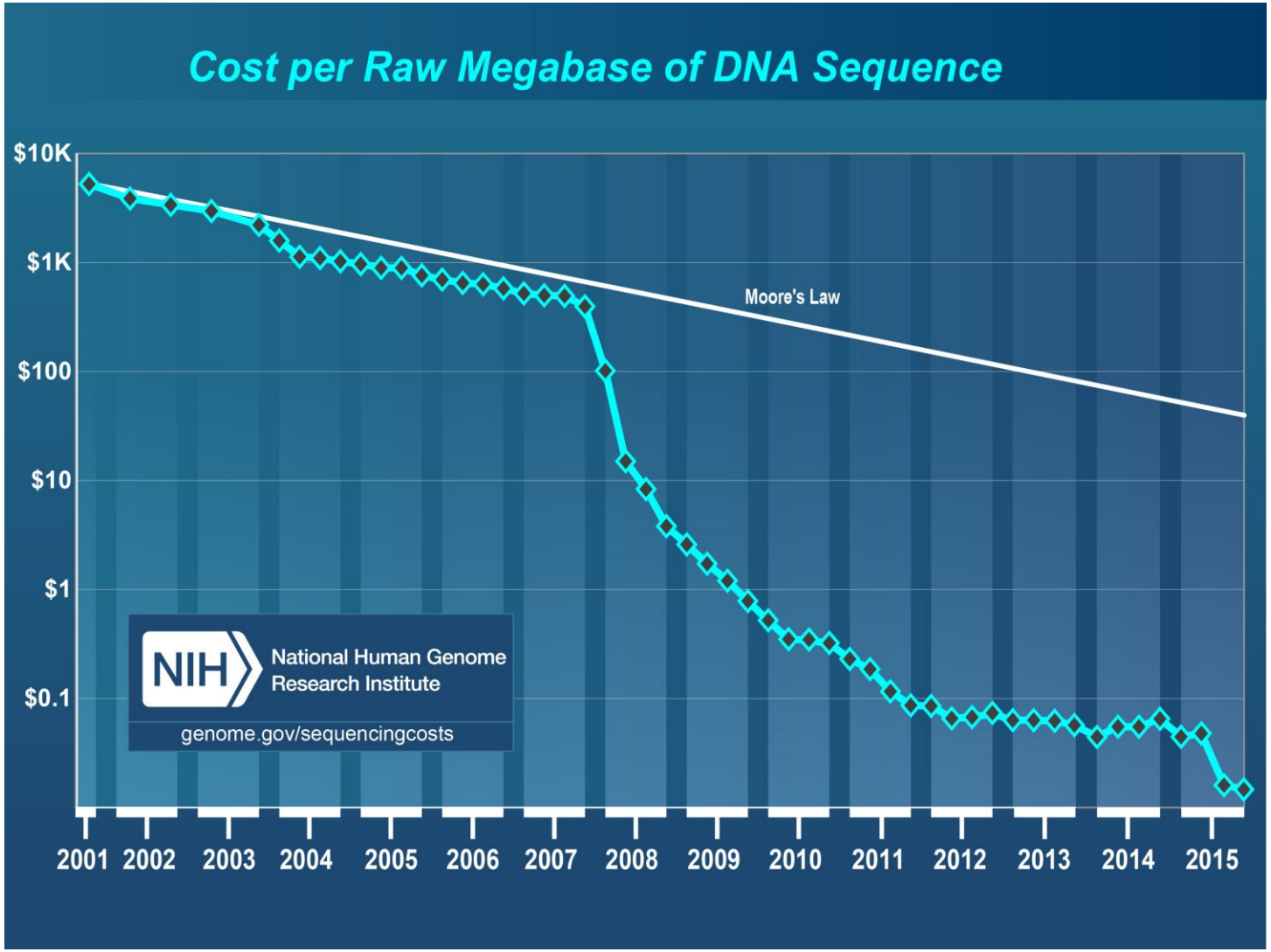
Eric Jorgenson, PhD

March 8th, 2016

Outline

- Overview
- Examples of Next Generation Sequencing Studies: Whole Genome, Exome, Families (IBD), Cancer
- PTC Taste Sensitivity
- Implications

Sequencing Cost Have Fallen



KAISER PERMANENTE®

Number of Genetic Markers Needed for Genetic Studies

Genome-wide Linkage Studies

300-400 Microsatellite Markers

Genome-wide Association Studies

100,000-2,500,000 SNPs

Exome Sequencing Studies

30,000,000 Basepairs

Exome Array Studies

>240,000 exonic variants

Whole Genome Sequencing Studies

3,200,000,000 Basepairs

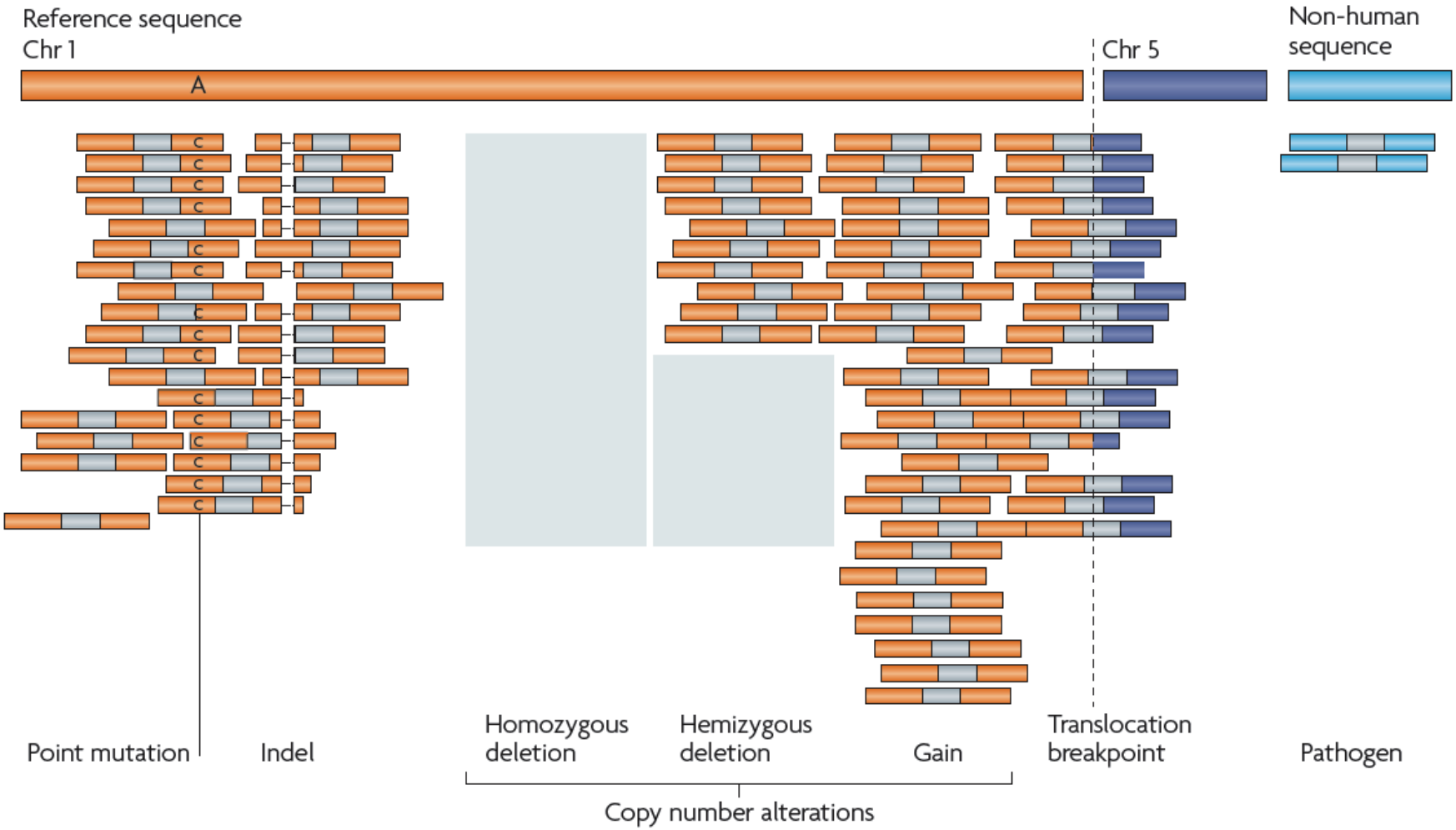
Links to Sequencing Projects

1000 Genomes:
http://browser.1000genomes.org/Homo_sapiens/Info/Index

Exome Sequencing Project:
https://esp.gs.washington.edu/drupal/

Exome Array Design:
http://genome.sph.umich.edu/wiki/Exome_Chip_Design#Second_Generation_Arrays

Variant Detection through Next Generation Sequencing



Meyerson et al. NRG 2010

KAISER PERMANENTE®

Outline

- Overview
- **Examples of Next Generation Sequencing Studies: Whole Genome, Exome, Families (IBD), Cancer**
- PTC Taste Sensitivity
- Implications

Sequencing of a Single Individual with Family Data

ORIGINAL ARTICLE

Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy

James R. Lupski, M.D., Ph.D., Jeffrey G. Reid, Ph.D., Claudia Gonzaga-Jauregui, B.S., David Rio Deiros, B.S., David C.Y. Chen, M.Sc., Lynne Nazareth, Ph.D., Matthew Bainbridge, M.Sc., Huyen Dinh, B.S., Chyn Jing, M.Sc., David A. Wheeler, Ph.D., Amy L. McGuire, J.D., Ph.D., Feng Zhang, Ph.D., Pawel Stankiewicz, M.D., Ph.D., John J. Halperin, M.D., Chengyong Yang, Ph.D., Curtis Gehman, Ph.D., Danwei Guo, M.Sc., Rola K. Irikat, B.S., Warren Tom, B.S., Nick J. Fantin, B.S., Donna M. Muzny, M.Sc., and Richard A. Gibbs, Ph.D.

Lupski et al. NEJM 2010

KAISER PERMANENTE®

The First 8 Human Genomes

Table 3. Individual Human Genomes Sequenced to Date.*

Genome†	Technology Used	Average Depth of Coverage	SNPs		
			Total	Known	Novel
				$\times 10^{-6}$	
Venter	Sanger method	7.5	3.21	2.80	0.74
Watson	454 Sequencing System (Roche)	7.4	3.32	2.71	0.61
Chinese (YH)	Genome Analyzer (Illumina)	36	3.07	2.67	0.39
African (NA18507)	Genome Analyzer (Illumina)	40.6	3.61	2.72	0.88
African (NA18507)	SOLiD system (Applied Biosystems)	17.9	3.86	3.13	0.73
Korean (SJK)	Genome Analyzer (Illumina)	28.95	3.43	3.01	0.42
Korean (AK1)	Genome Analyzer (Illumina)	27.8	3.45	2.88	0.57
Proband in this study	SOLiD system (Applied Biosystems)	29.9	3.42	2.85	0.56

Lupski et al. NEJM 2010

KAISER PERMANENTE®

SNP Distribution in Proband

Table 2. SNPs Identified through Whole-Genome Sequencing of DNA from the Proband.*

SNP Type	No. of SNPs
Nongene	2,255,102
Gene	1,165,204
Intron	1,064,655
Promoter	60,075
3′ UTR	16,350
5′ UTR	3,517
Splice regulatory site	2,089
Splice site	112
Synonymous	9,337
Stop→stop	17
Nonsynonymous	9,069
Stop→gain	121
Stop→loss	27
Total	3,420,306

Lupski et al. NEJM 2010

KAISER PERMANENTE®

Nonsynonymous SNPs in Known Disease Genes

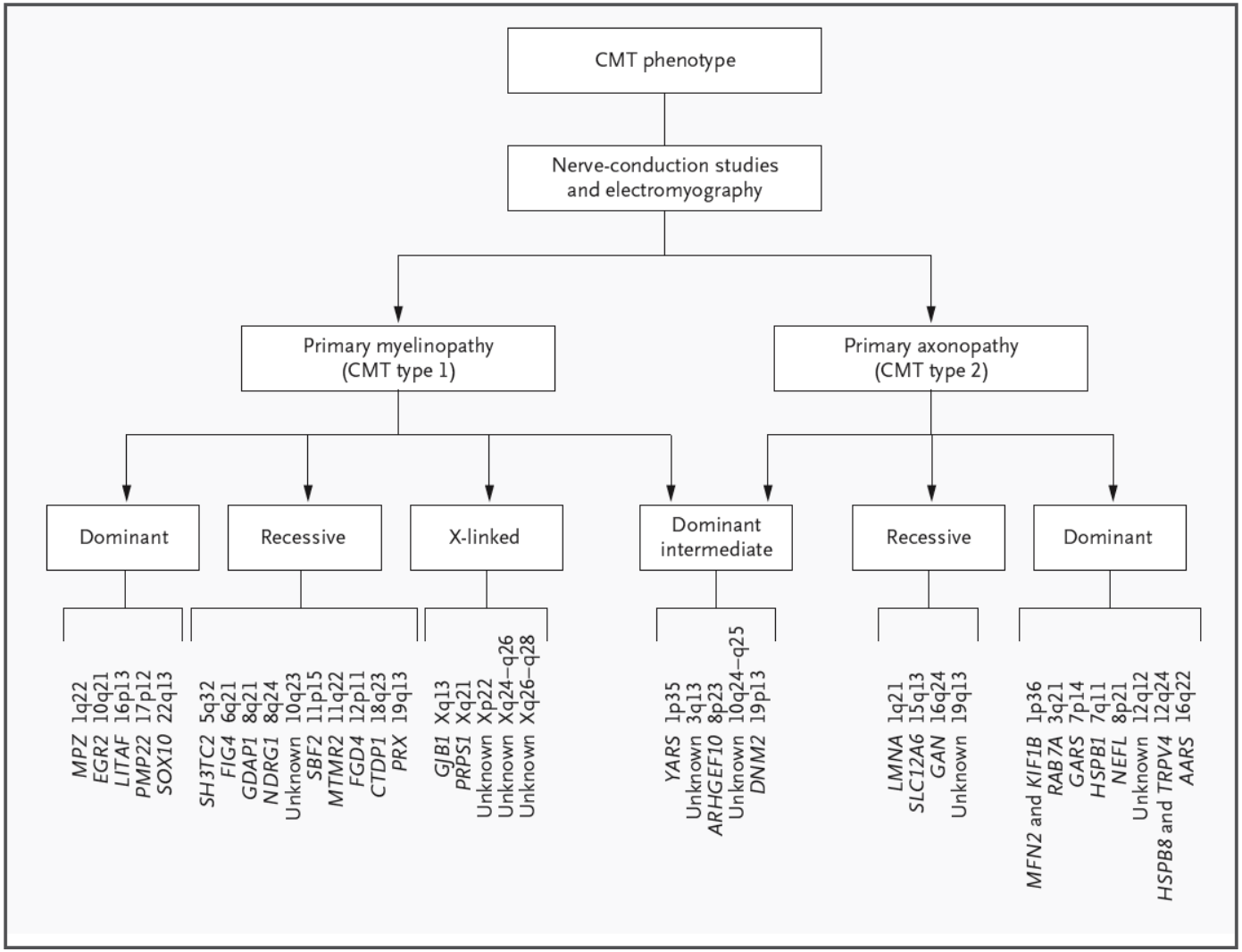
Table 4. Disease and Trait Associations of Nonsynonymous SNPs Identified in the Proband, According to the Human Gene Mutation Database.*

Disease or Trait Associated with Mutation	SNPs
	no. (%)
Total	159 (100)
Behavioral disorder	6 (4)
Cancer	33 (21)
Association	7
Increased risk	9
Reduced risk	3
Susceptibility	14
Complex disease	48 (30)
Mendelian disease	21 (13)
Metabolic trait	17 (11)
Pharmacogenetic trait	14 (9)
Other traits	20 (13)

Lupski et al. NEJM 2010

KAISER PERMANENTE®

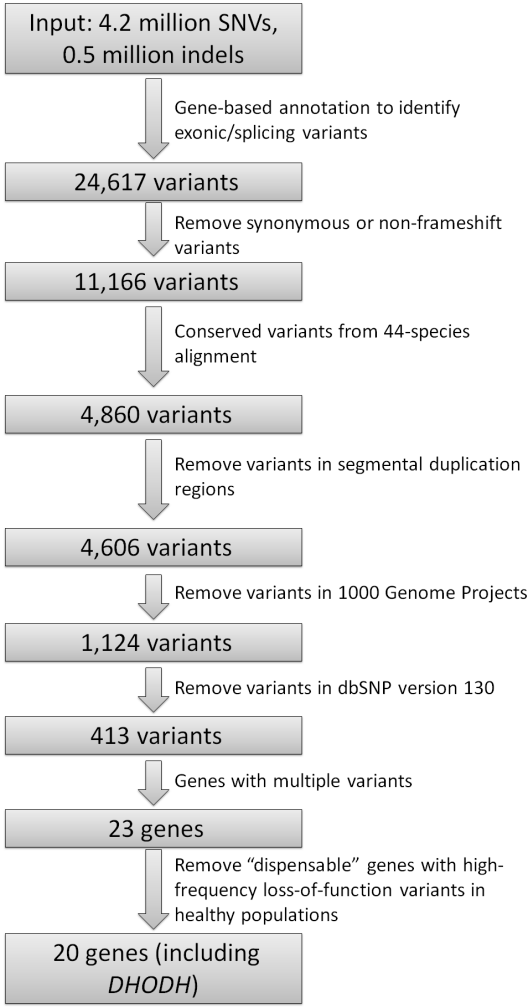
CMT Subtypes: Many Genes



Lupski et al. NEJM 2010

KAISER PERMANENTE®

ANNOVAR: Using Annotation to Narrow the Search Space



openbioinformatics.org/annovar

KAISER PERMANENTE®

Phenotypes in Unseqeunced Family Members

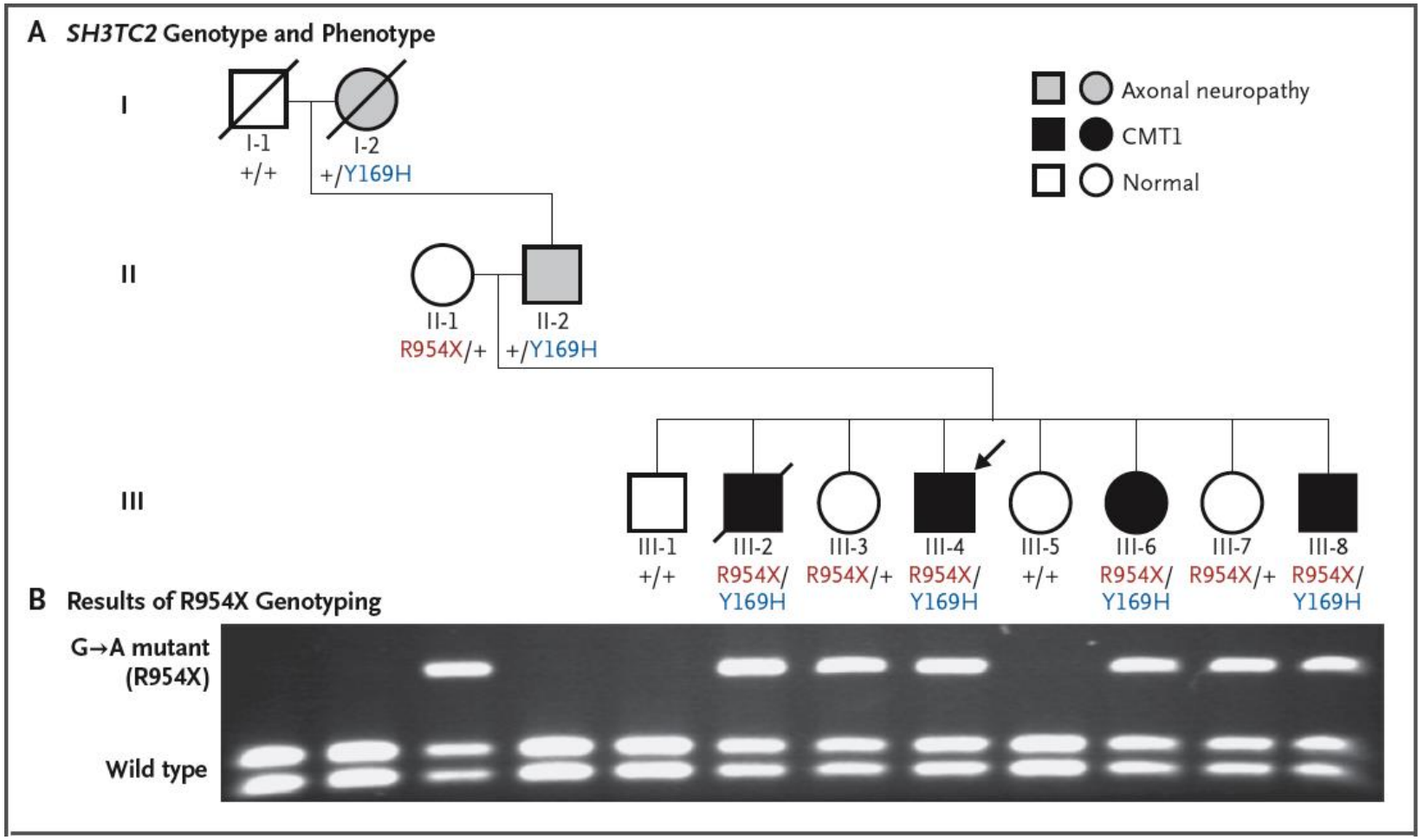
Table 1. Neurophysiological Findings in the Study Family.*

Subject No.	Age	Sex	Demyelination	Axonopathy	Median-Nerve Entrapment	Diagnosis
	yr					
I-1	80	M	No	Motor: peroneal, 2.2 mV	Yes	
I-2	77	F	No	Sural: absent; motor: peroneal, 0.5 mV; tibial, 2.8 mV	Yes	Axonal neuropathy
Maternal grandmother of proband	90	F	No	Normal	Yes; SCV, 43 m/sec	MMM
II-I	58	F	No	Normal	Yes; SCV, 46 m/sec	MMM
II-2	57	M	No	Sural: absent; motor: peroneal, 0.2 mV; tibial, 1.4 mV; H-reflexes: 38 msec	Yes	Axonal neuropathy
III-1	37	M	No	Normal	No	
III-2	35	M	Yes	No	Yes; median-nerve motor TL, 14.9 msec; ulnar-nerve motor TL, 8.1 msec	CMT
III-3	34	F	No	No	Yes; SCV, 42 m/sec; median-nerve motor TL, 4.4 msec	MMM
III-4 (proband)	32	M	Yes	No	Probably; median-nerve motor TL, 10.2 msec; ulnar-nerve motor TL, 7.5 msec	CMT
III-5	31	F	No	No	No	
III-6	29	F	Yes	No	Yes; median-nerve motor TL, 11.6 msec; ulnar-nerve motor TL, 6.2 msec	CMT
III-7	26	F	No	Motor: peroneal, 36 m/sec; H-reflexes: 35 msec	Yes; SCV, 36 m/sec; median-nerve motor TL, 4.8 msec	MMM
III-8	25	M	Yes	No	Yes; median-nerve motor TL, 9.2 msec; ulnar-nerve motor TL, 6.2 msec	CMT

Lupski et al. NEJM 2010

KAISER PERMANENTE®

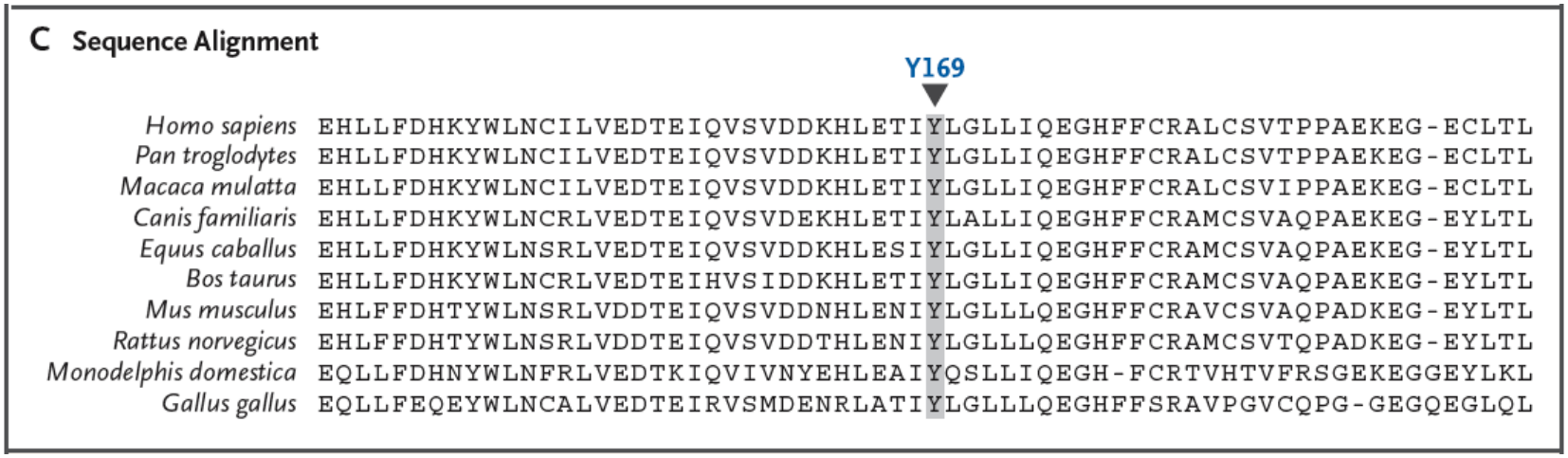
Family Pedigree



Lupski et al. NEJM 2010

KAISER PERMANENTE®

Putative Causal Variant at a Conserved Amino Acid



Lupski et al. NEJM 2010

KAISER PERMANENTE®

Narcissome

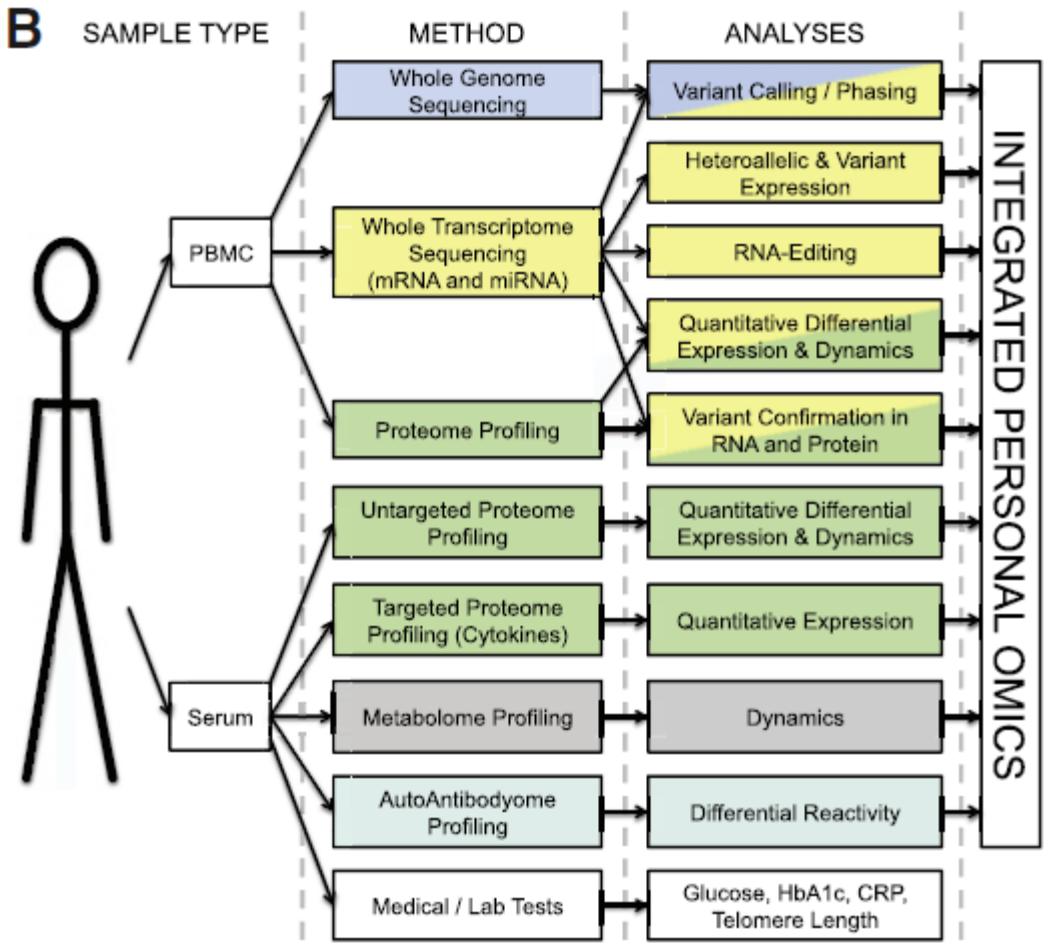
Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes

Rui Chen,[1,11] George I. Mias,[1,11] Jennifer Li-Pook-Than,[1,11] Lihua Jiang,[1,11] Hugo Y.K. Lam,[1,12] Rong Chen,[2,12] Elana Miriami,[1] Konrad J. Karczewski,[1] Manoj Hariharan,[1] Frederick E. Dewey,[3] Yong Cheng,[1] Michael J. Clark,[1] Hogune Im,[1] Lukas Habegger,[6,7] Suganthi Balasubramanian,[6,7] Maeve O'Huallachain,[1] Joel T. Dudley,[2] Sara Hillenmeyer,[1] Rajini Haraksingh,[1] Donald Sharon,[1] Ghia Euskirchen,[1] Phil Lacroute,[1] Keith Bettinger,[1] Alan P. Boyle,[1] Maya Kasowski,[1] Fabian Grubert,[1] Scott Seki,[2] Marco Garcia,[2] Michelle Whirl-Carrillo,[1] Mercedes Gallardo,[9,10] Maria A. Blasco,[9] Peter L. Greenberg,[4] Phyllis Snyder,[1] Teri E. Klein,[1] Russ B. Altman,[1,5] Atul J. Butte,[2] Euan A. Ashley,[3] Mark Gerstein,[6,7,8] Kari C. Nadeau,[2] Hua Tang,[1] and Michael Snyder[1,*]

Chen et al. Cell 2012

KAISER PERMANENTE®

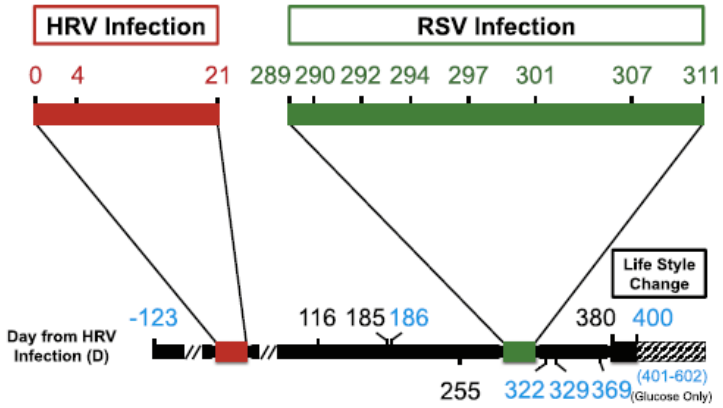
Narcissome

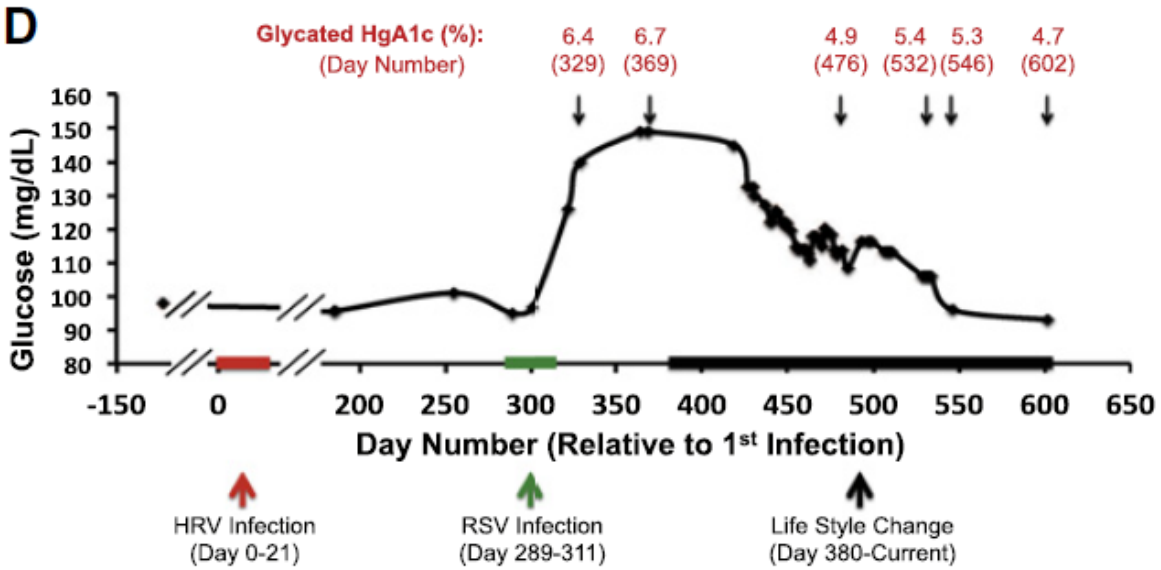


Chen et al. Cell 2012

KAISER PERMANENTE®

Narcissome





Chen et al. Cell 2012

KAISER PERMANENTE®

Exome Sequencing Identifies a Tibetan Adaptation

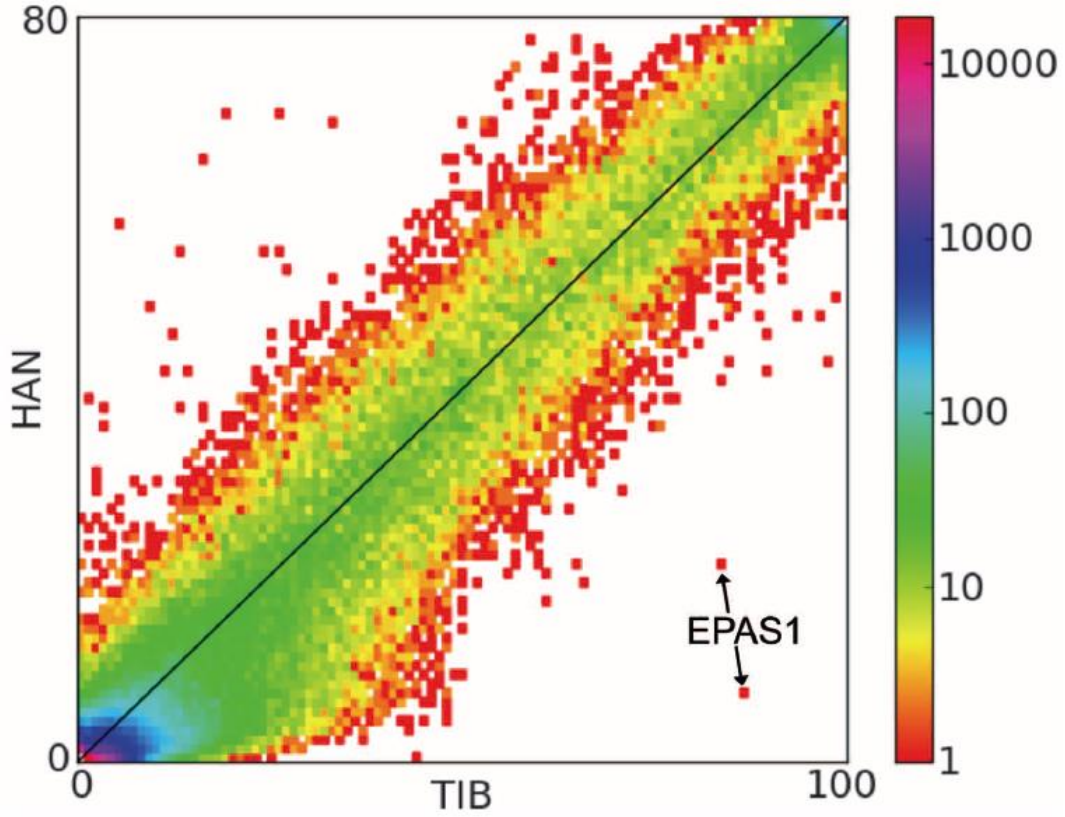
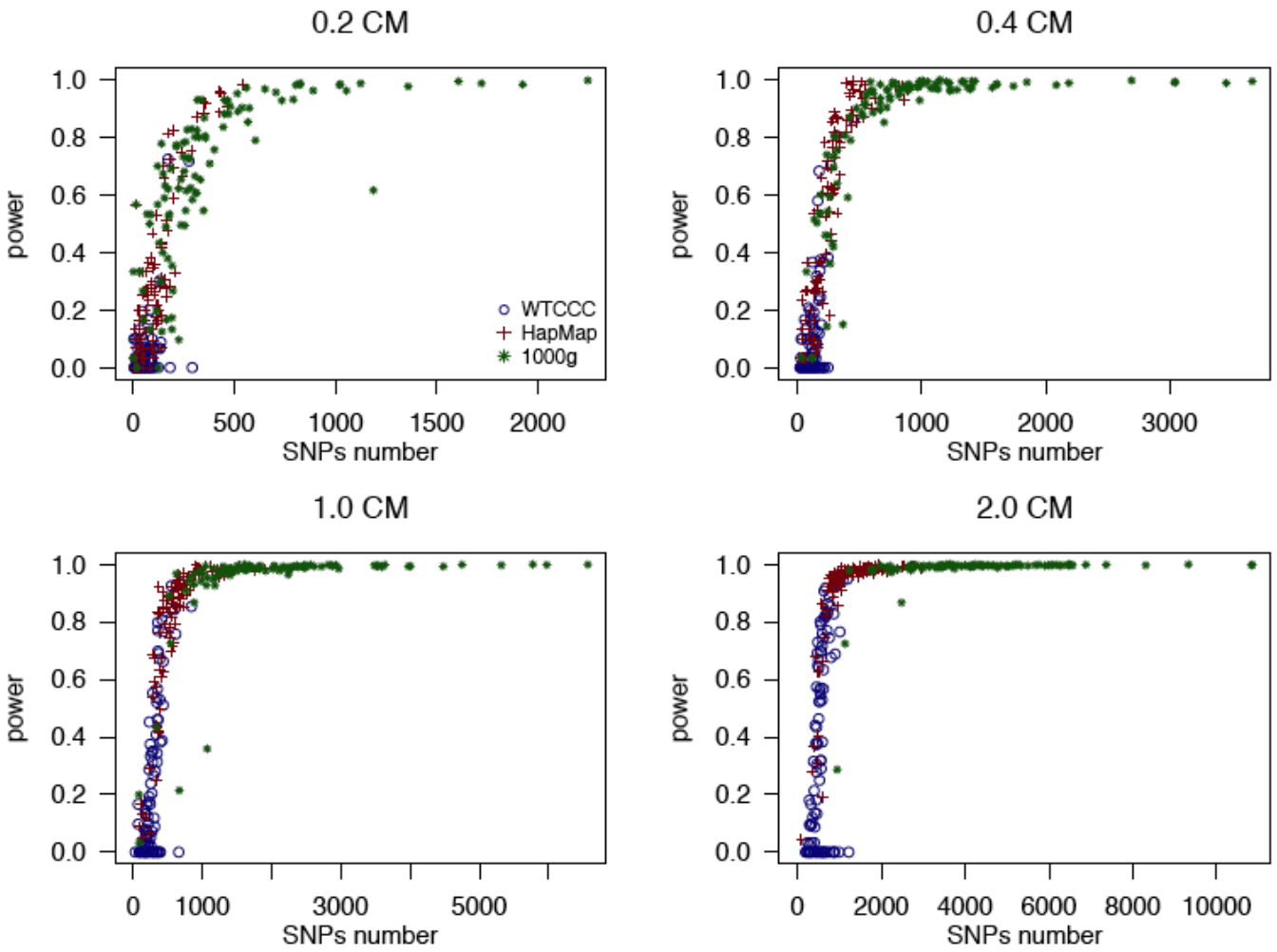


Fig. 1. Two-dimensional unfolded site frequency spectrum for SNPs in Tibetan (*x* axis) and Han (*y* axis) population samples. The number of SNPs detected is color-coded according to the logarithmic scale plotted on the right. Arrows indicate a pair of intronic SNPs from the *EPAS1* gene that show strongly elevated derived allele frequencies in the Tibetan sample compared with the Han sample.

Yi et al. Science 2010

KAISER PERMANENTE®

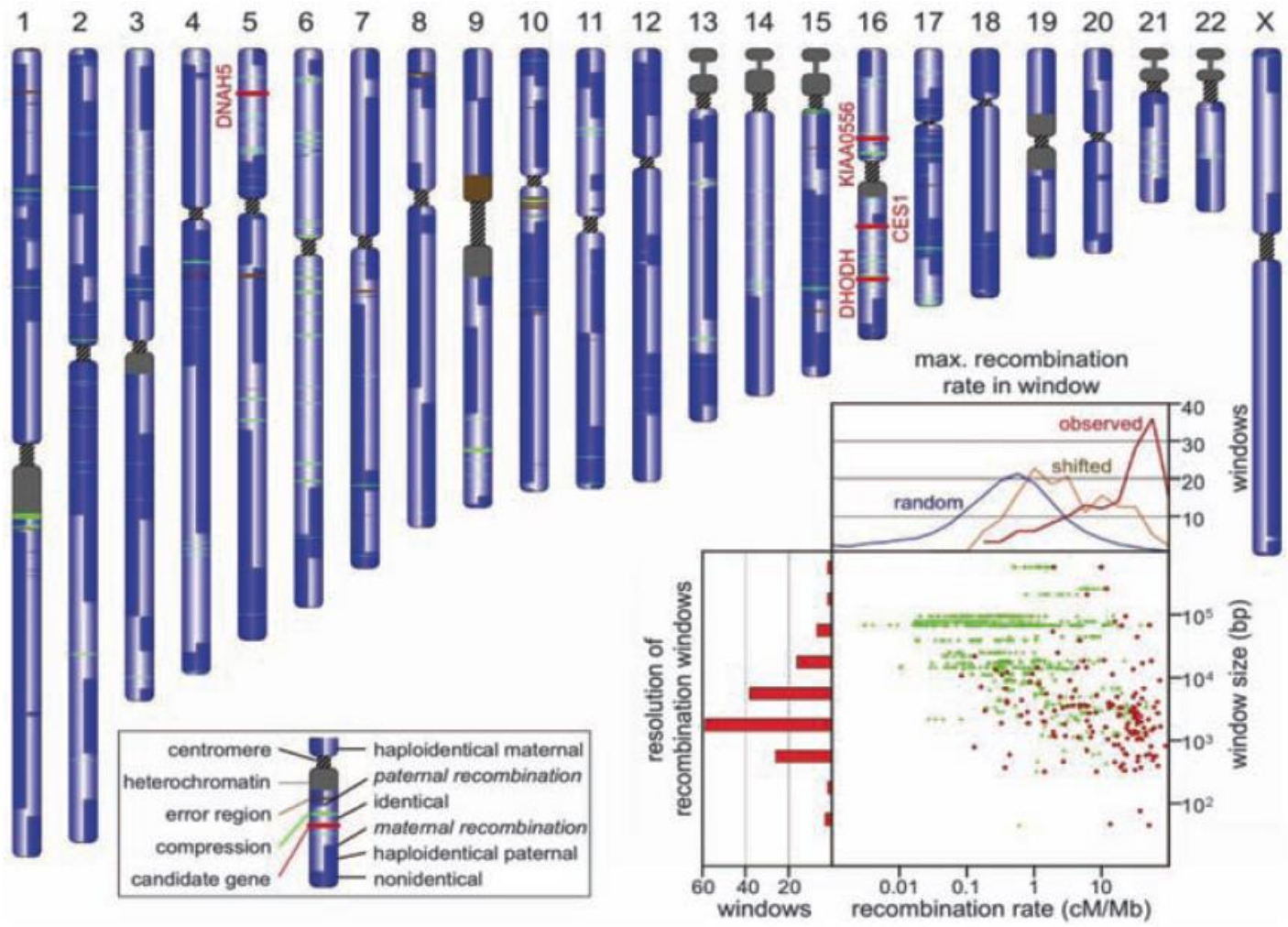
Sequence Data Improves Identity by Descent Resolution



Su and Jorgenson 2012

KAISER PERMANENTE®

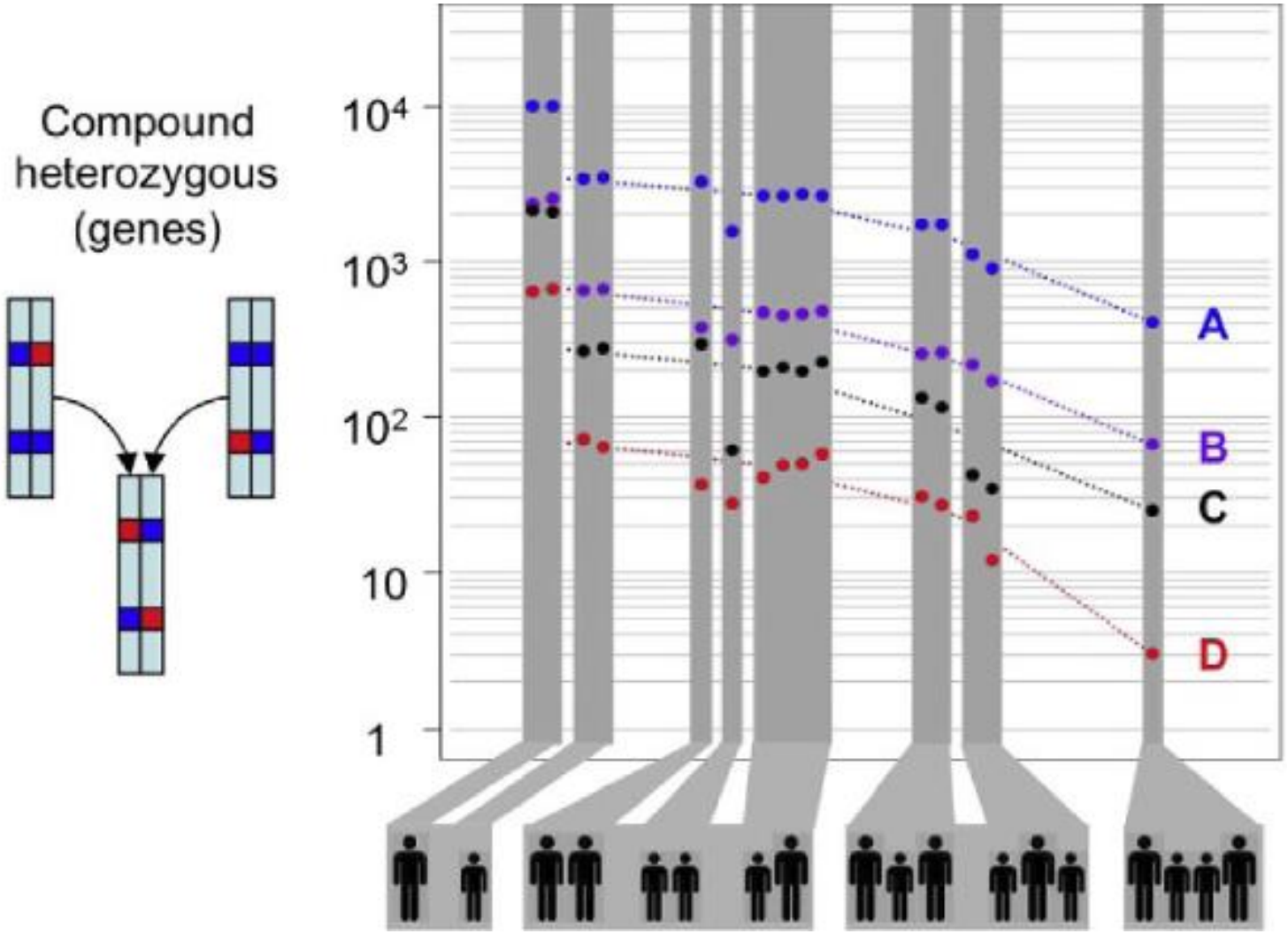
Family Sequencing for Rare Diseases



Roach et al. Science 2010

KAISER PERMANENTE®

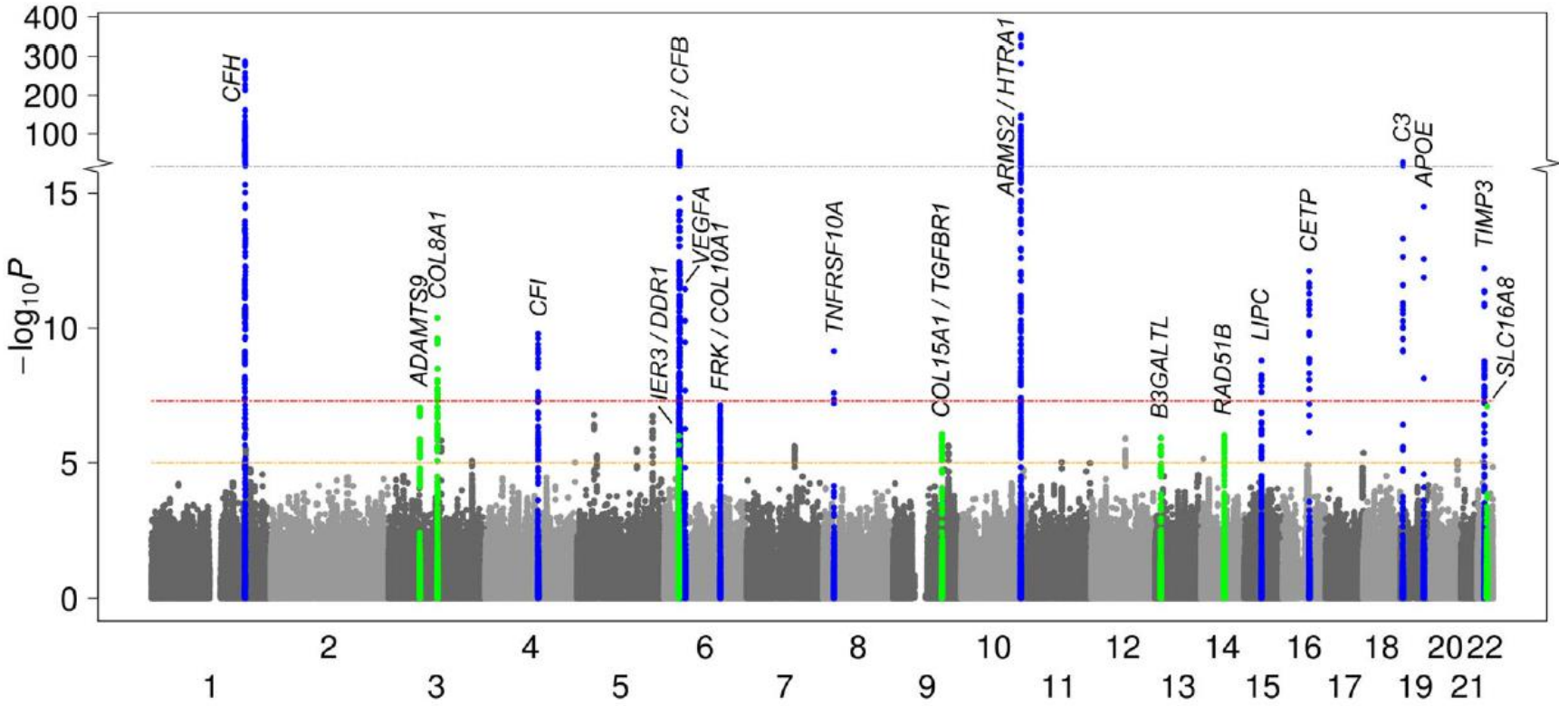
Families Can Reduce Error Rate



Roach et al. Science 2010

KAISER PERMANENTE®

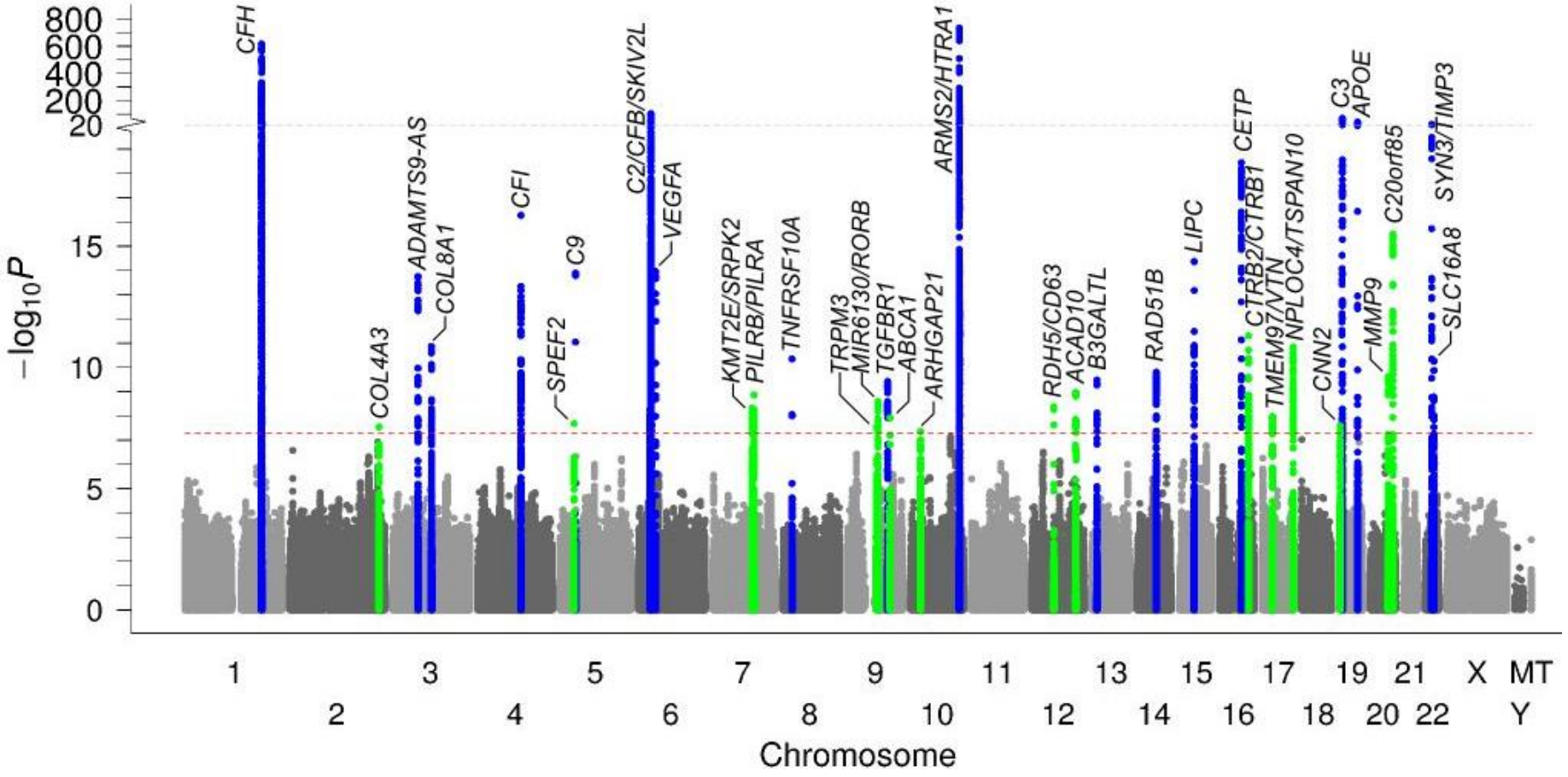
AMD GWAS Findings, 2013



 | © 2011 Kaiser Foundation Health Plan, Inc. For internal use only.

KAISER PERMANENTE®

AMD GWAS Findings, 2015

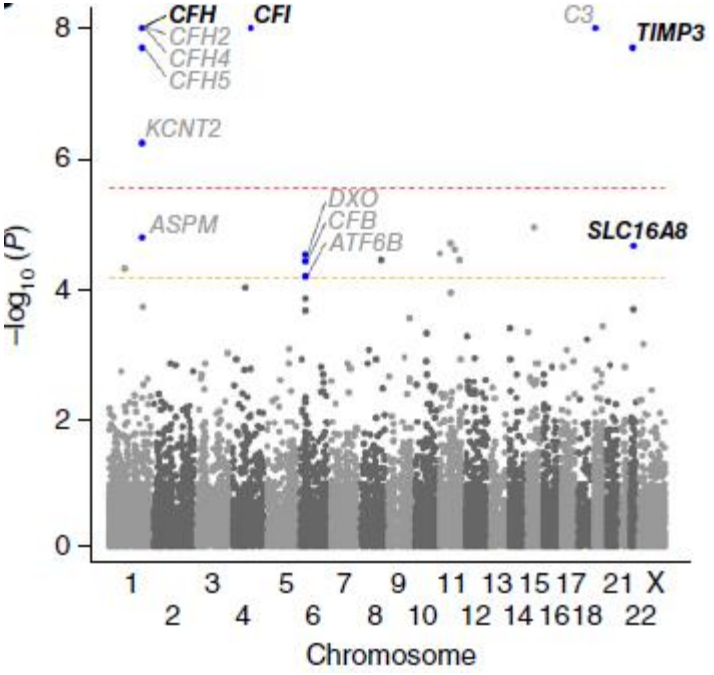


 | © 2011 Kaiser Foundation Health Plan, Inc. For internal use only.

KAISER PERMANENTE®

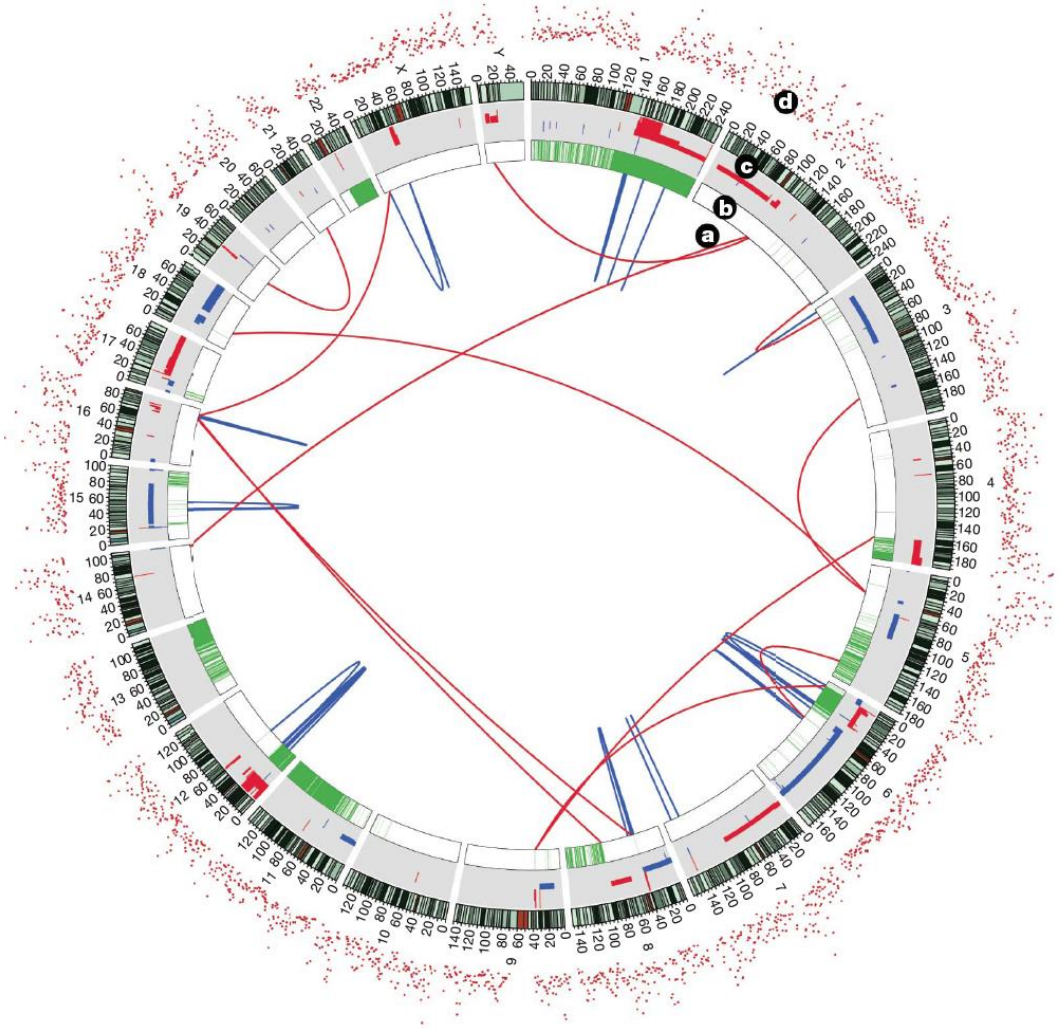
AMD Rare Variant Findings, 2015

	Optimal RAC threshold	Number of variants below optimal RAC	Summed RAC (frequency as %)			
Gene	Count (%)	Total (exome chip base + custom)	Cases n = 16,144	Controls n = 17,832	P[a]	OR
CFH	10 (0.015)	37 (9 + 28)	88 (0.273)	38 (0.107)	1.2×10^{-6}	2.94
CFI	46 (0.068)	43 (17 + 26)	213 (0.660)	82 (0.230)	1.0×10^{-8}	2.95
TIMP3	14 (0.021)	9 (1 + 8)	29 (0.0898)	1 (0.00280)	9.0×10^{-8}	31.21
SLC16A8	648 (0.954)	9 (7 + 2)	487 (1.51)	392 (1.10)	3.1×10^{-6}	1.40



© 2011 Kaiser Foundation Health Plan, Inc. For internal use only.
KAISER PERMANENTE®

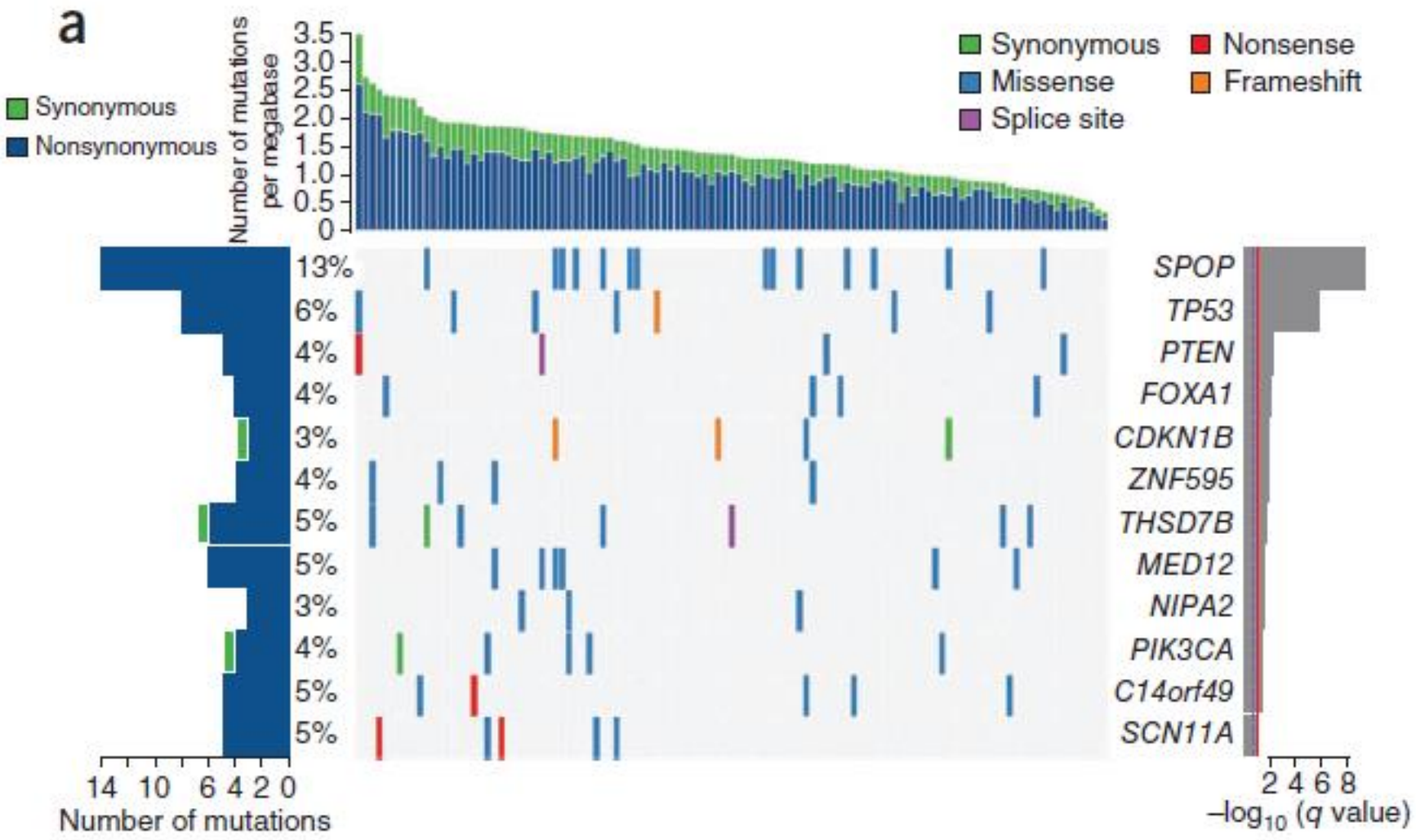
Cancer: Tumor vs. Normal



Lee et al. Nature 2010

KAISER PERMANENTE®

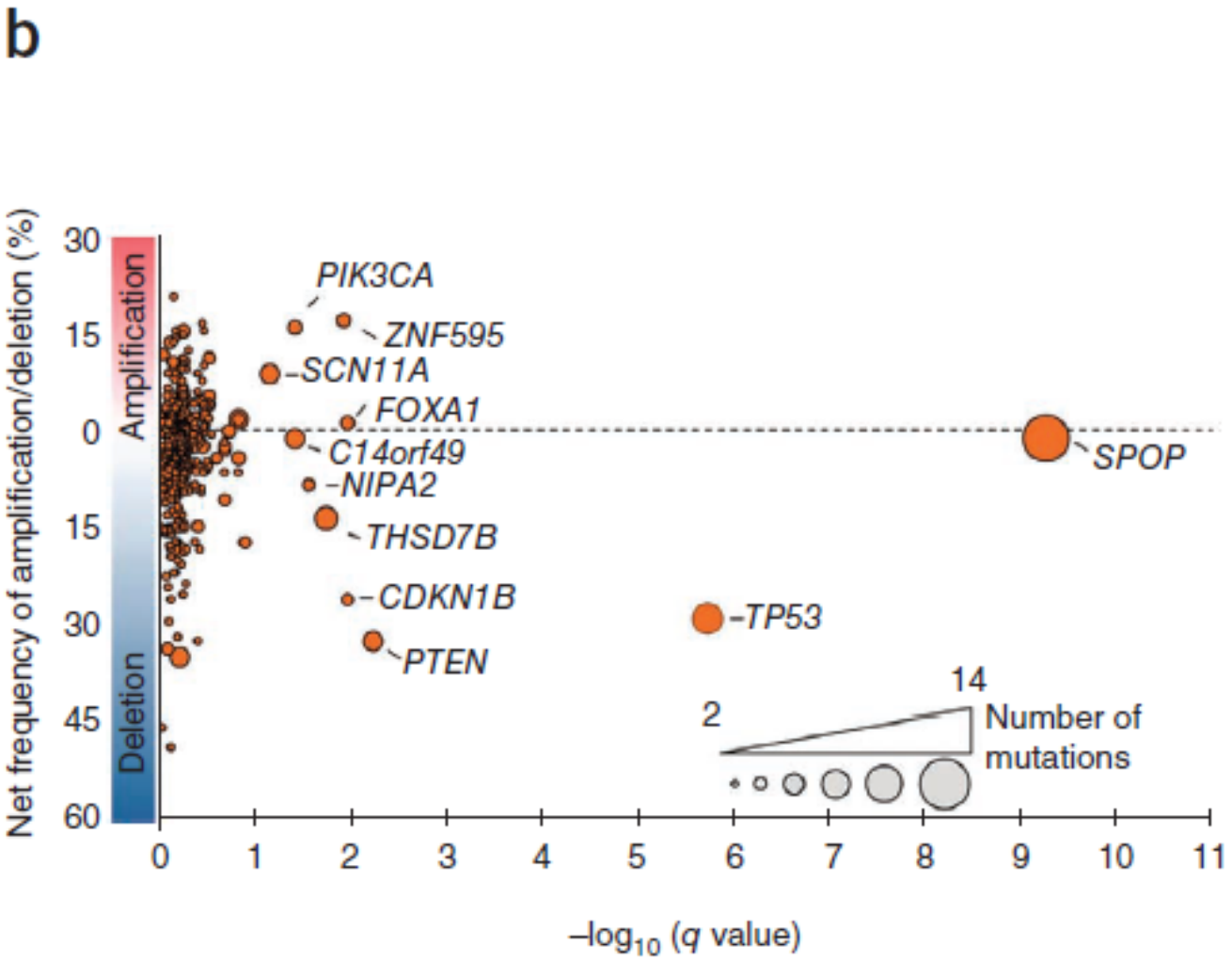
Exome Sequencing in Prostate Cancer



Barbieri et al. Nature Genetics 2012

KAISER PERMANENTE®

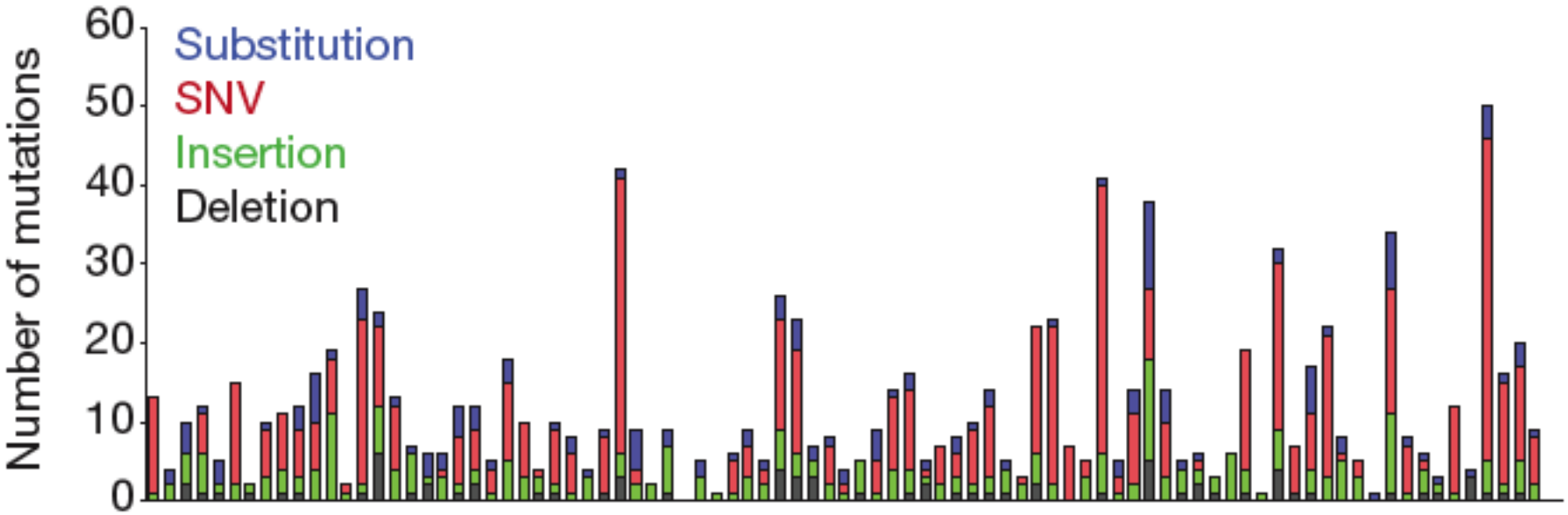
Exome Sequencing in Prostate Cancer



Barbieri et al. Nature Genetics 2012

KAISER PERMANENTE®

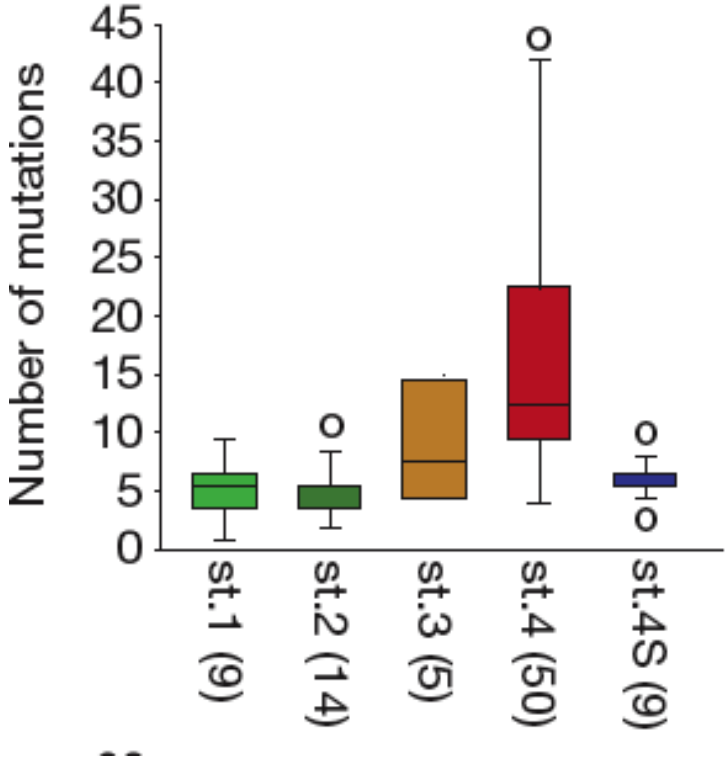
Nonsynonymous Somatic Mutations in Neuroblastoma

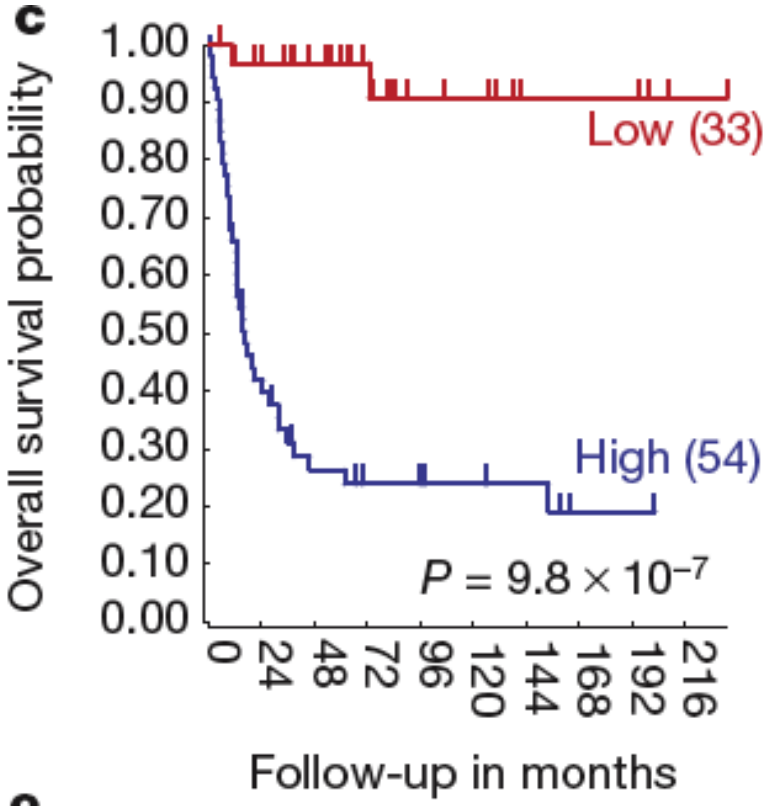


Molenaar et al. Nature 2012

KAISER PERMANENTE®

Mutation Count Associated with Age, Stage, and Survival





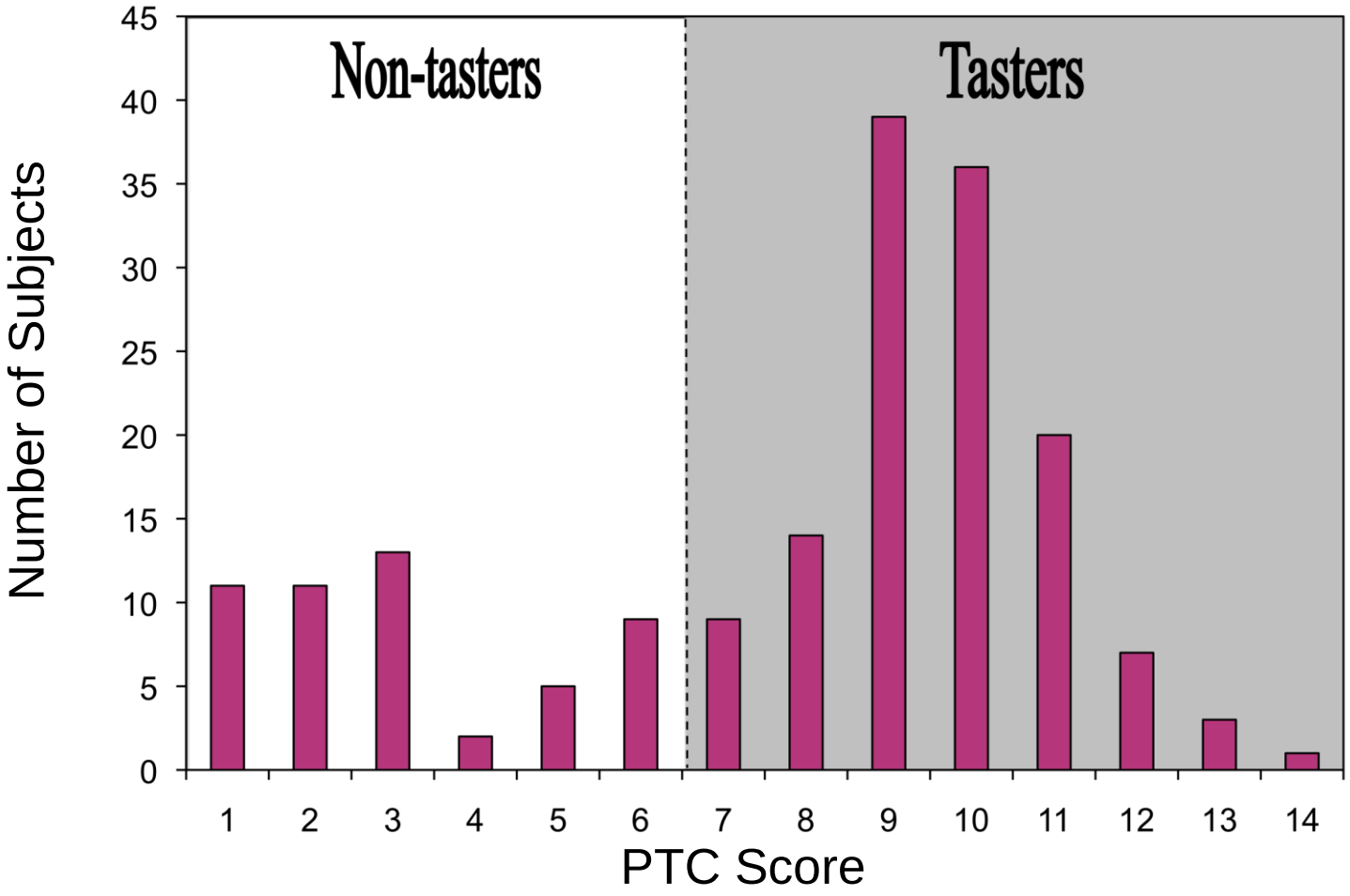
Molenaar et al. Nature 2012

KAISER PERMANENTE®

Outline

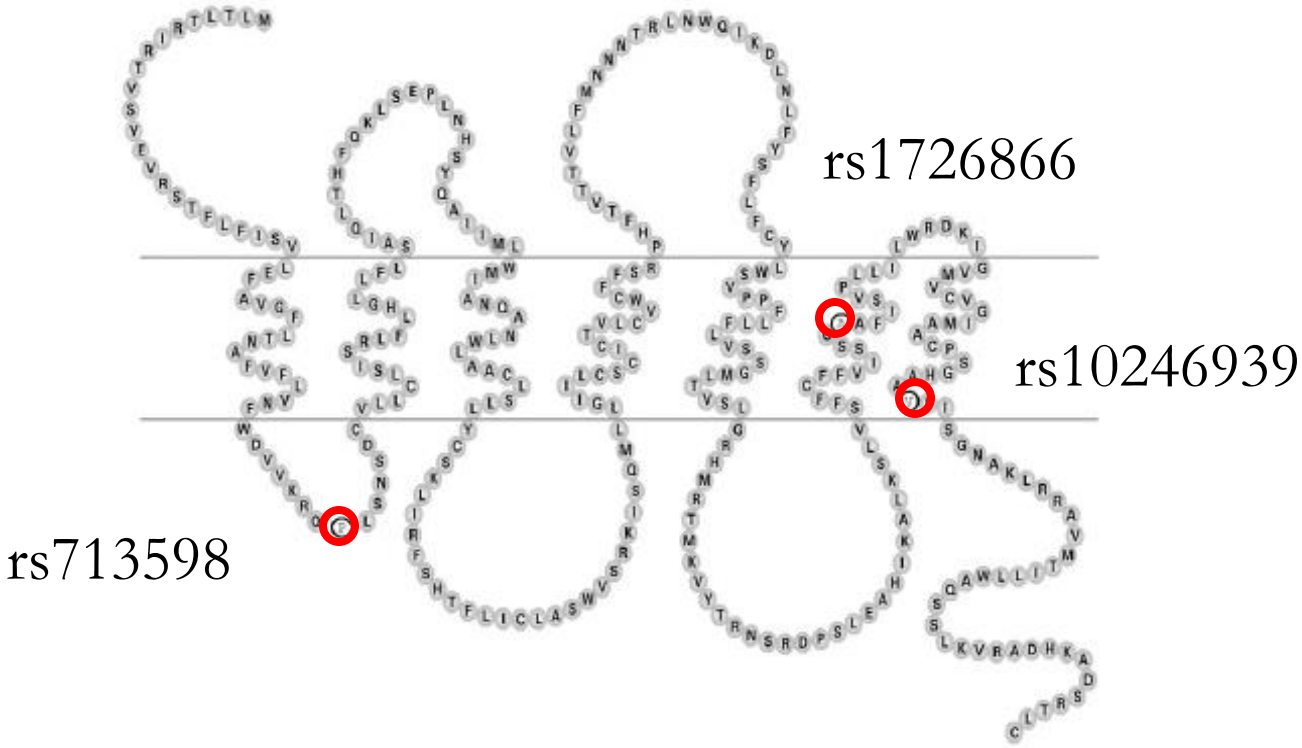
- Overview
- Examples of Next Generation Sequencing Studies: Whole Genome, Exome, Families (IBD), Cancer
- **PTC Taste Sensitivity**
- Implications

Distribution of the PTC Phenotype



KAISER PERMANENTE®

TAS2R38 Receptor Structure



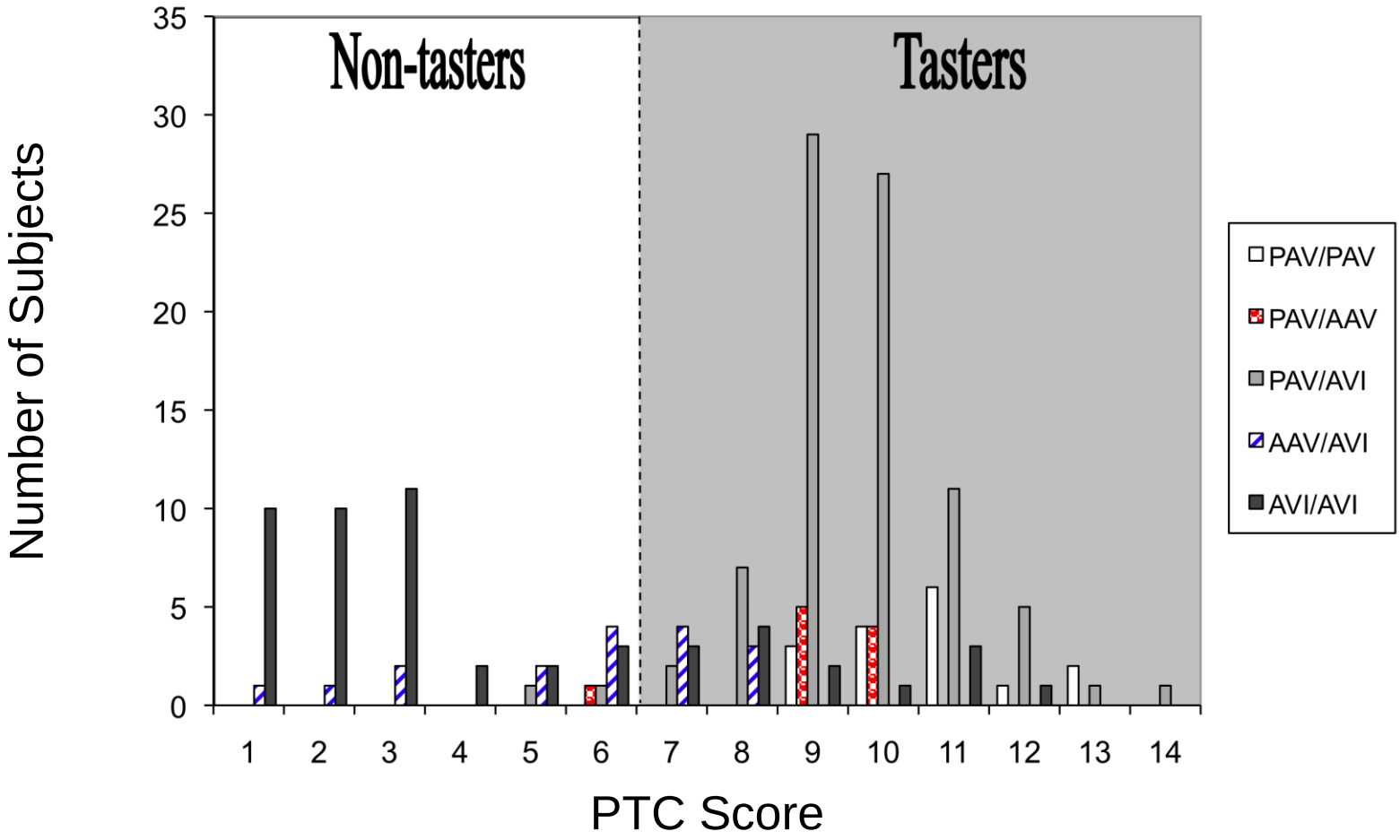
Kim et al. J Dent Res 2004

KAISER PERMANENTE®

3 SNPs Form 3 Haplotypes

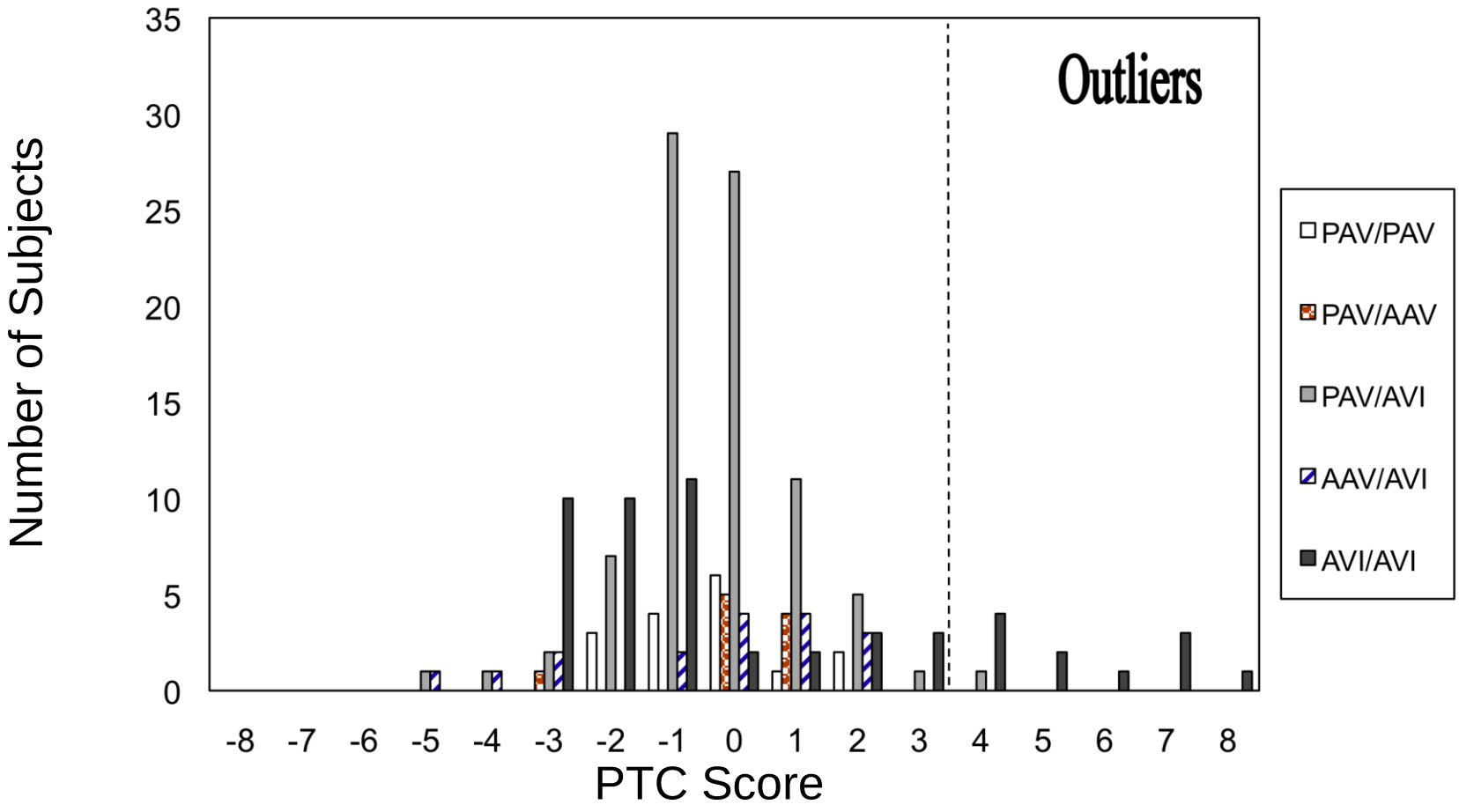
KAISER PERMANENTE®

PTC Phenotype by *TAS2R38* Diplotype



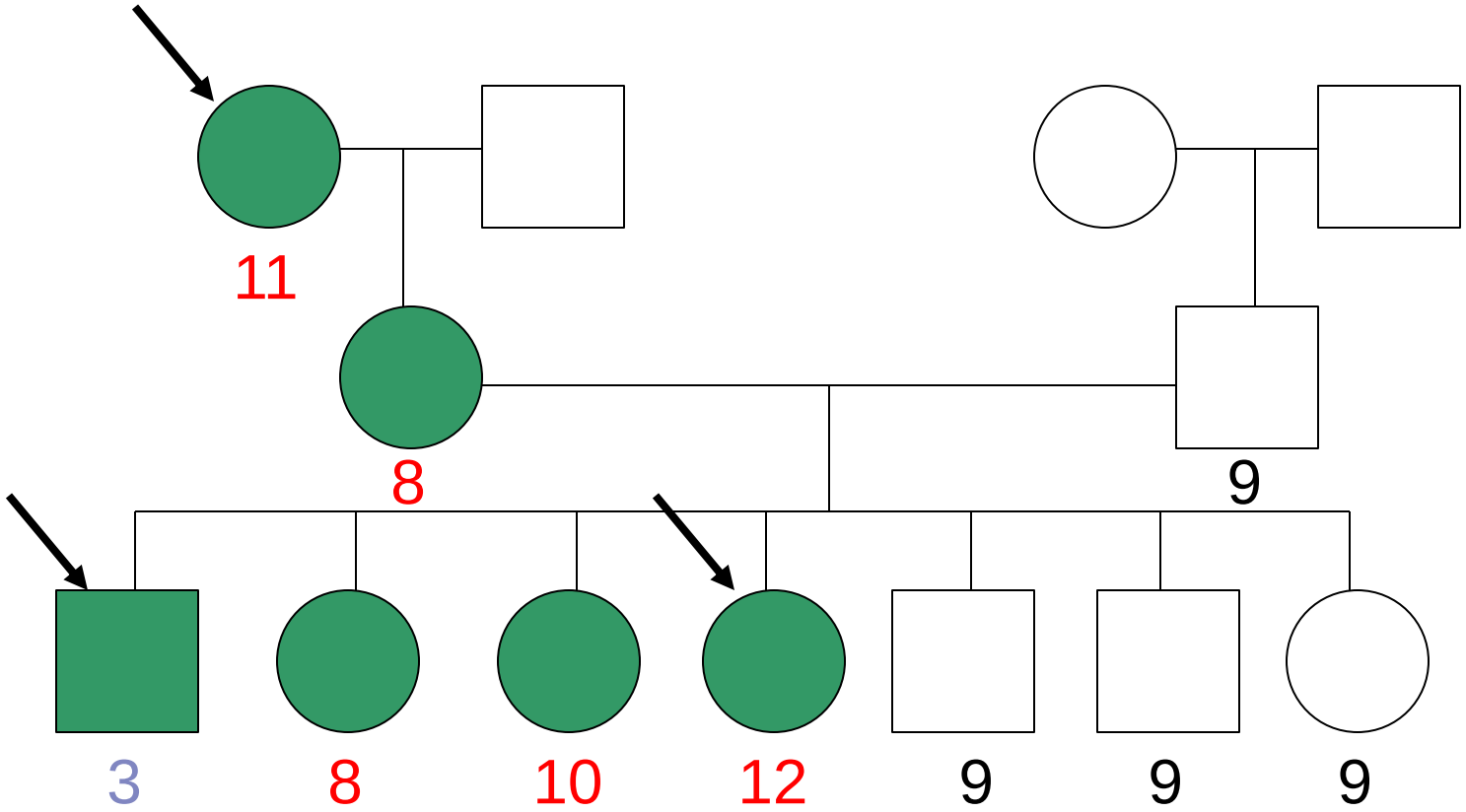
KAISER PERMANENTE®

Outliers after Adjusting for *TAS2R38* Diplotype



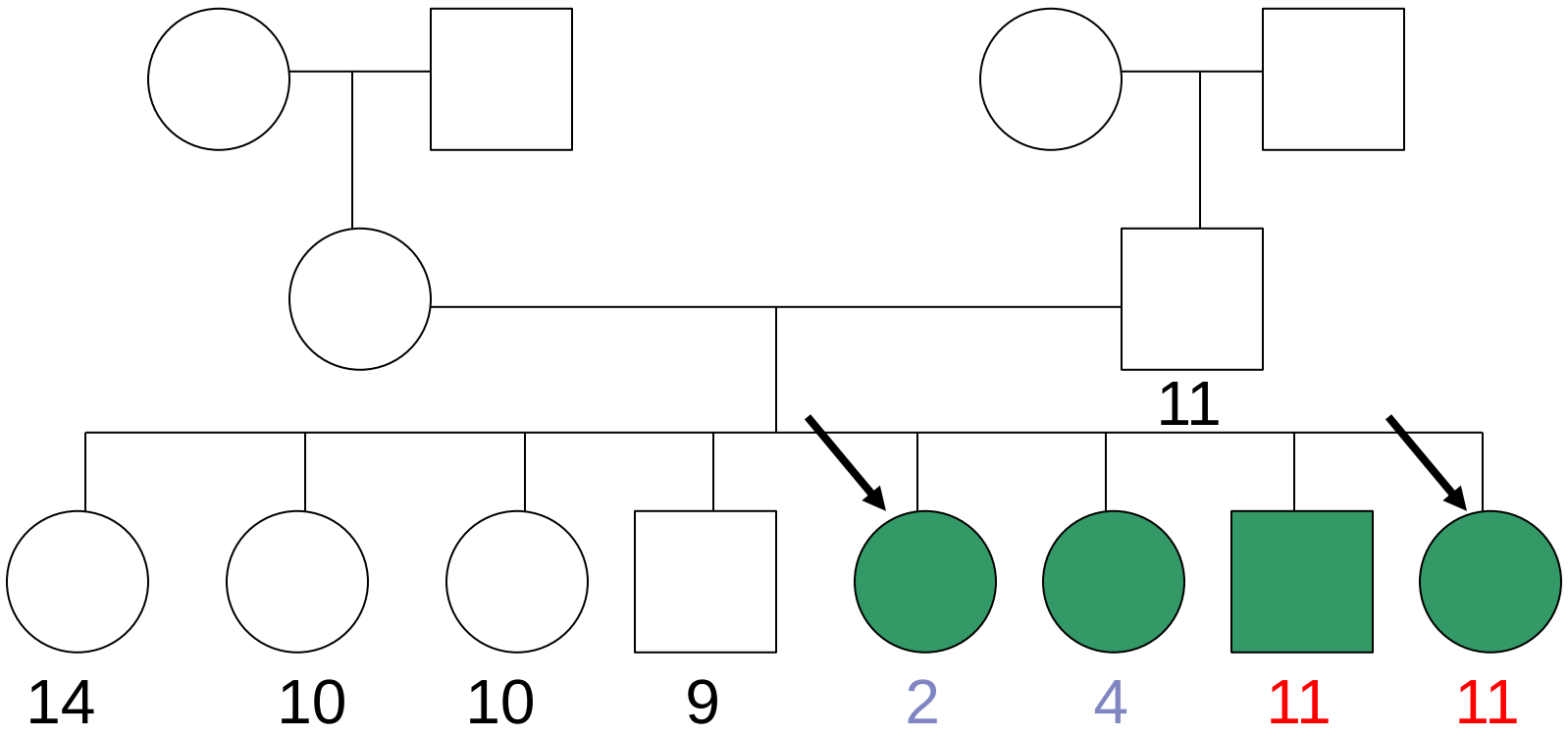
KAISER PERMANENTE®

Unusual PTC Phenotypes (AVI Homozygotes in Green)



KAISER PERMANENTE®

Unusual PTC Phenotypes (AVI Homozygotes in Green)



KAISER PERMANENTE®

10 Genomes, 5 Hard Drivers

KAISER PERMANENTE®

Summary of Variation

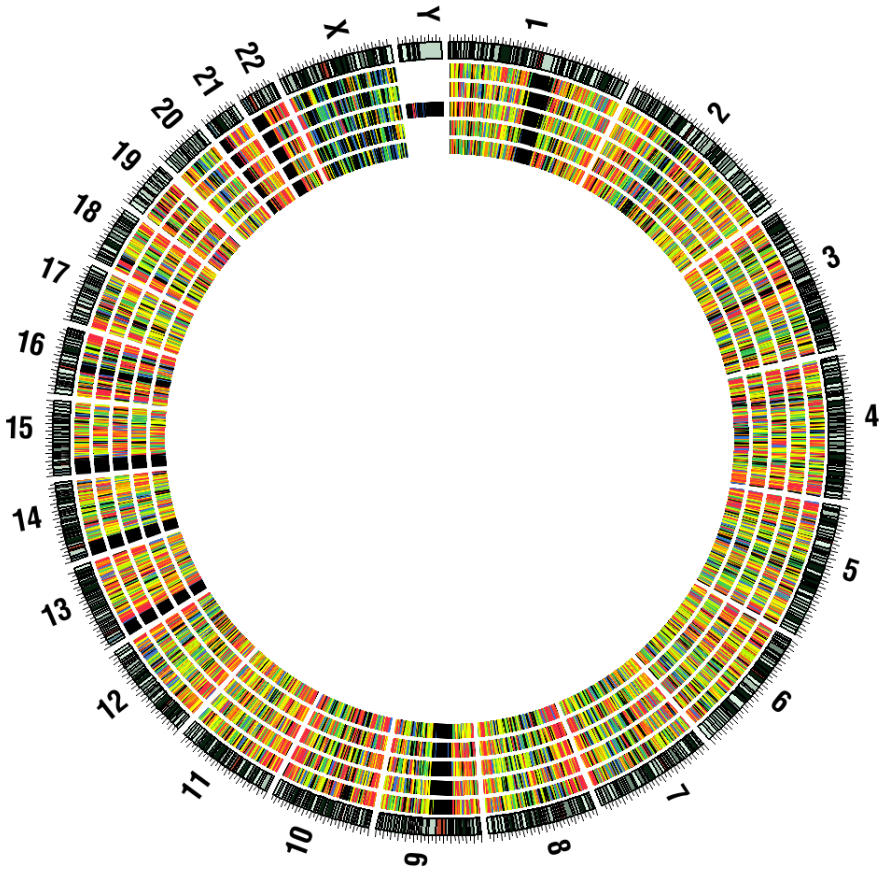
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Gender	Female	Female	Male	Female	Female
Total Sequence (Gb)	214	220	218	243	219
Percent fully called	0.95	0.96	0.96	0.97	0.96
Coverage (X fold)	53	55	53	63	54
SNPs	3,270,920	3,269,487	3,278,557	3,355,266	3,341,154
Insertions	184,763	190,633	197,830	210,805	206,120
Deletions	195,419	200,495	208,031	221,532	216,578
Synonymous SNPs	9,666	9,547	9,808	10,004	9,981
Missense SNPs	9,253	9,135	9,350	9,486	9,581
Nonsense SNPs	90	97	82	88	92
Frameshift Insertions	103	102	97	112	127
Frameshift Deletions	99	101	91	108	116
Novel SNPs	0.04	0.04	0.04	0.04	0.04
Novel Insertions	0.18	0.18	0.19	0.20	0.19
Novel Deltions	0.22	0.22	0.23	0.23	0.23

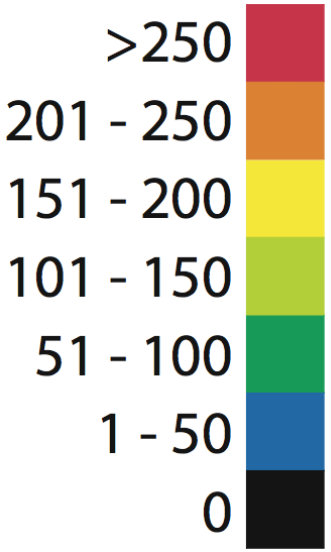
KAISER PERMANENTE®

Quality Control: 99.8% Concordance

Sample 1	Genotyping		
Sequencing	Homozygous Reference	Heterozygous	Homozygous Variant
Homozygous Reference	479,773	429	422
Heterozygous	426	234,156	293
Homozygous Variant	65	168	172,479

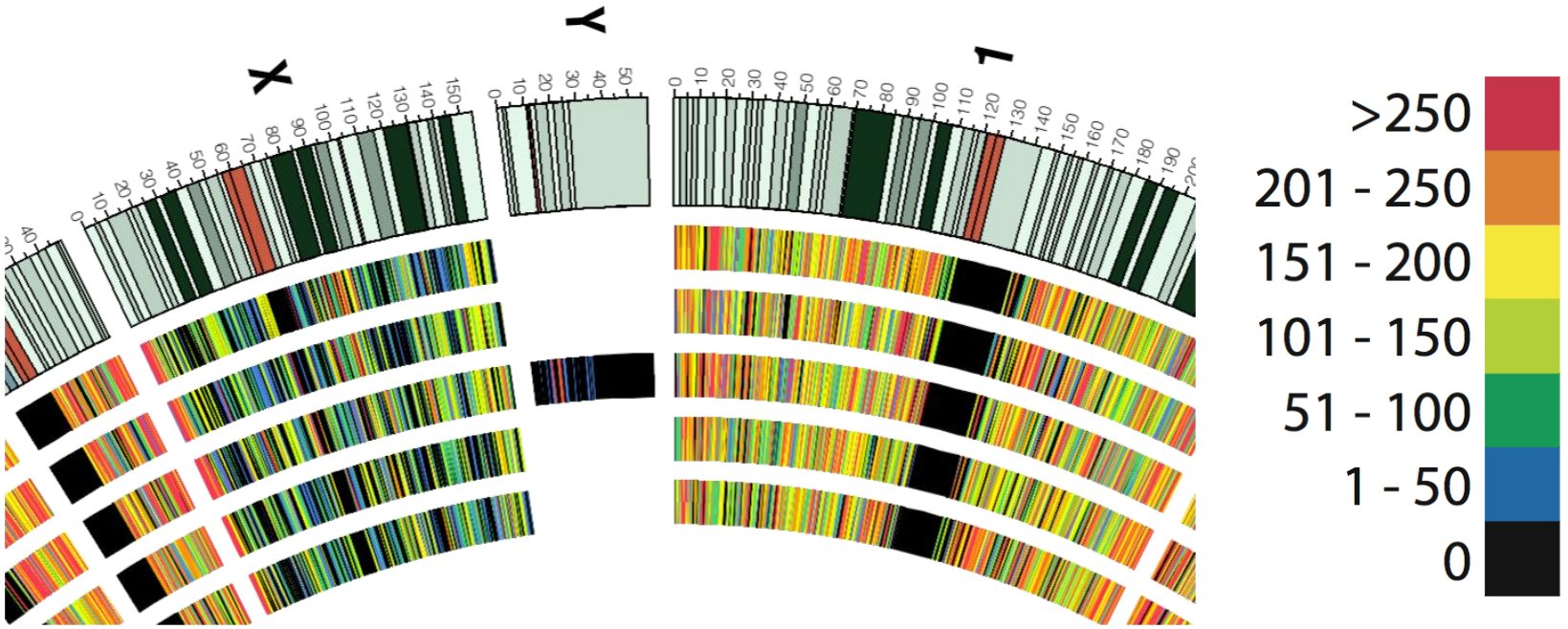
Variant Distribution in PTC Subjects





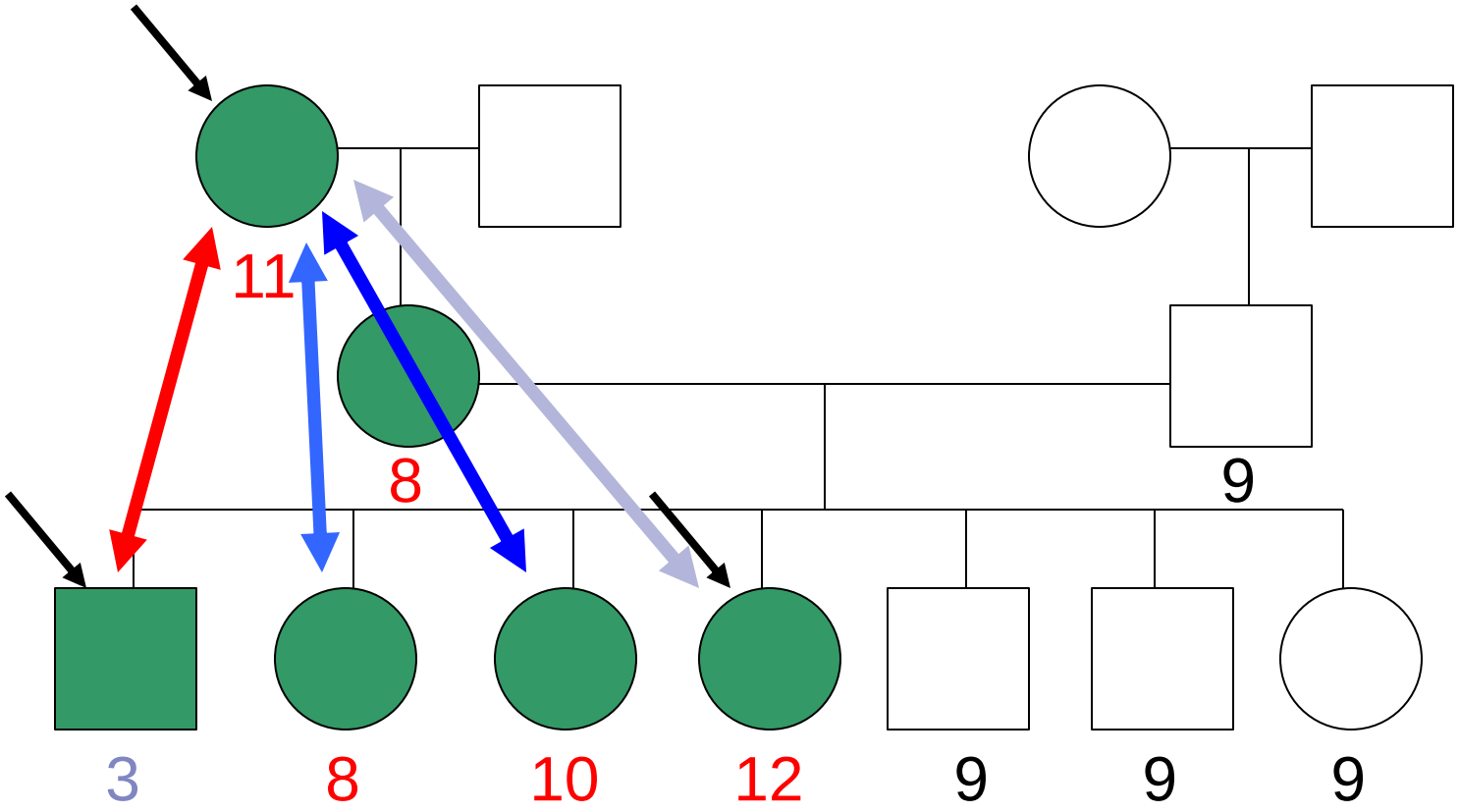
KAISER PERMANENTE®

Variant Distribution in PTC Subjects

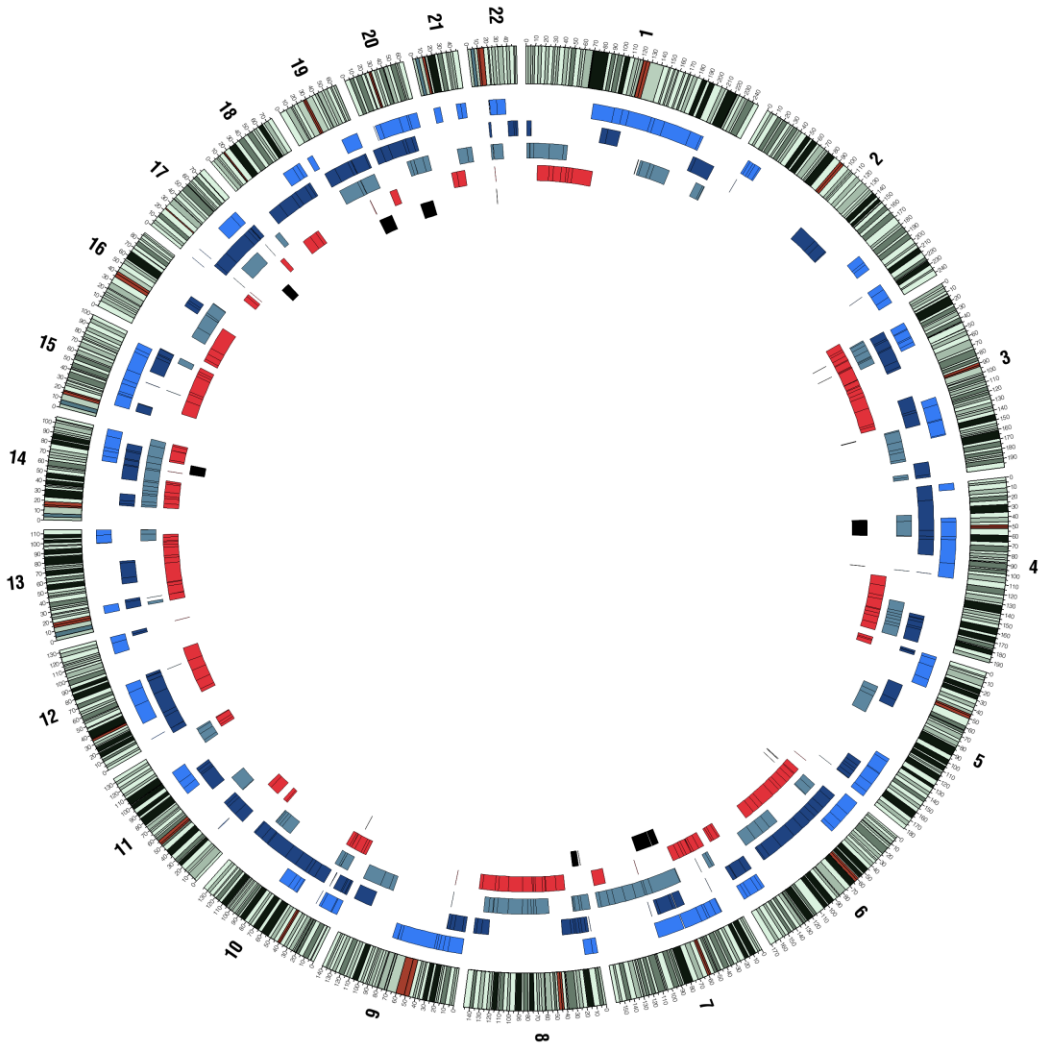


KAISER PERMANENTE®

Using Relatedness

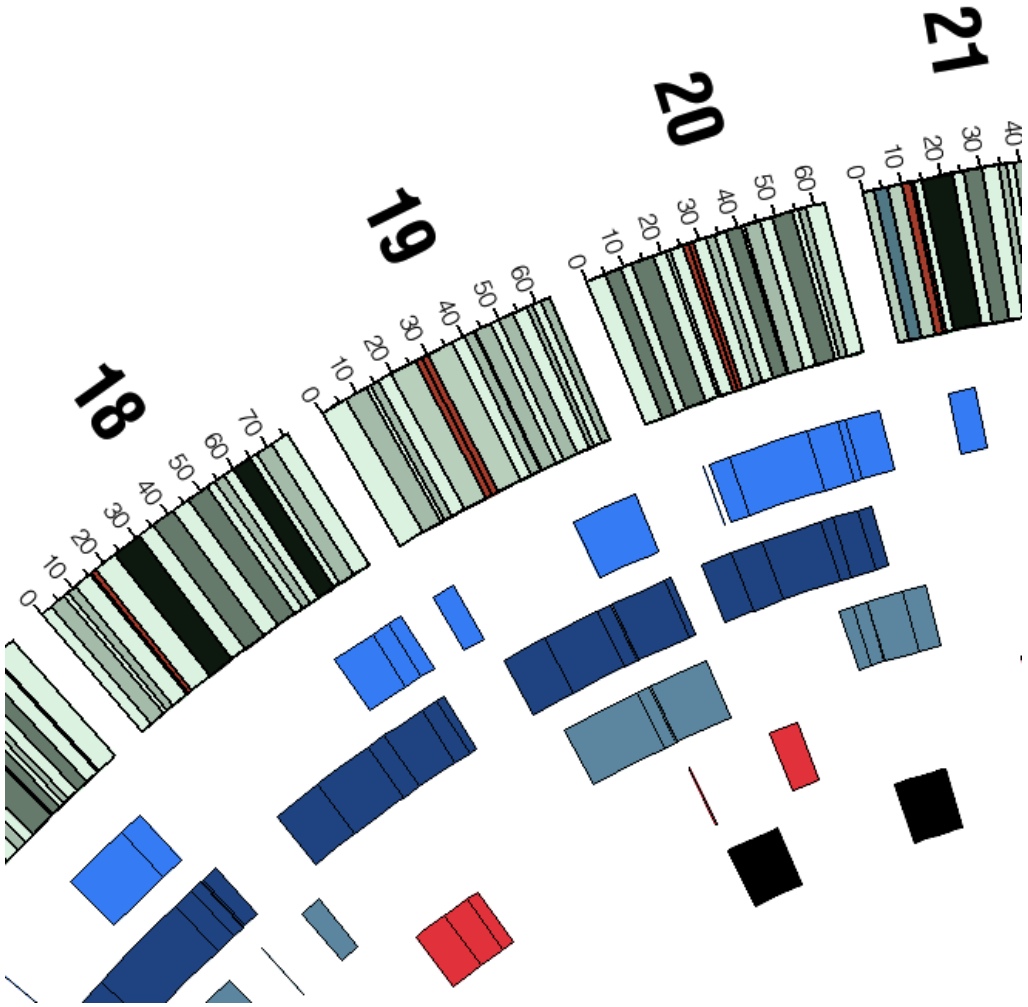


Identity by Descent



KAISER PERMANENTE®

Identity by Descent



KAISER PERMANENTE®

Nonsynonymous Variants

Chromosome	Gene	Start	Stop
1	ANXA9	150958836	150958836
1	S100A7A	153391729	153391729
5	PCDHB5	140517174	140517174
6	HLA-A	29910604	29910604
6	HLA-A	29912856	29912856
7	ZAN	100371474	100371474
9	BAG1	33264540	33264540
10	WDR96	105957714	105957714
10	GSTO1	106027059	106027059
11	MUC6	1018116	1018116
11	OR4C45	48367311	48367311
11	TRIM64	89701844	89701844
12	NANOGNB	7917936	7917936
14	ZBTB1	64988830	64988830
19	PSG3	43243238	43243238
X	HDHD1	6975782	6975782
X	ZCCHC16	111698036	111698036
X	SAGE1	134994005	134994005
X	GPR101	136112707	136112707

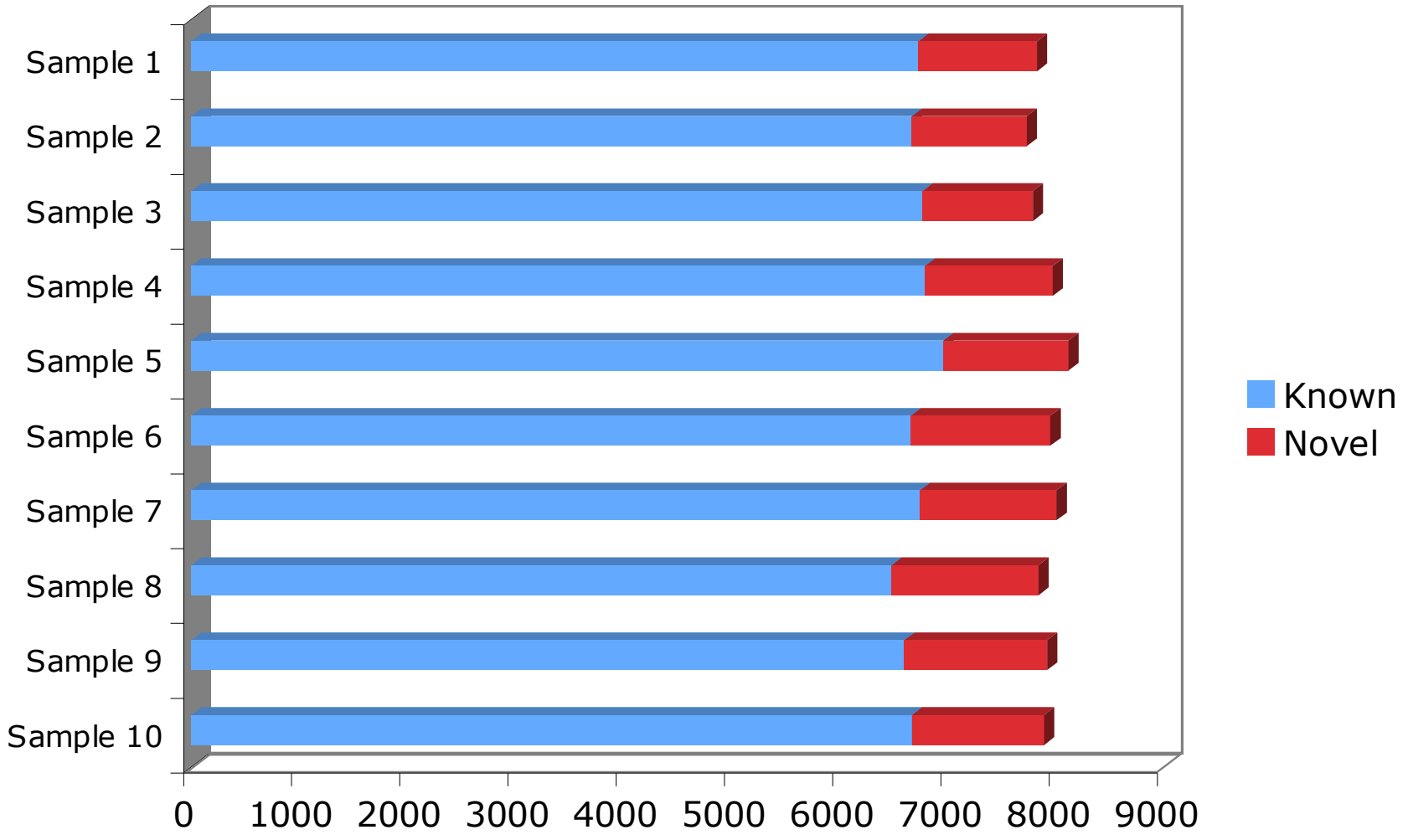
Outline

- Overview
- Examples of Next Generation Sequencing Studies: Whole Genome, Exome, Families (IBD), Cancer
- PTC Taste Sensitivity
- **Implications**

How Can Whole Genome Sequencing Influence Treatment?

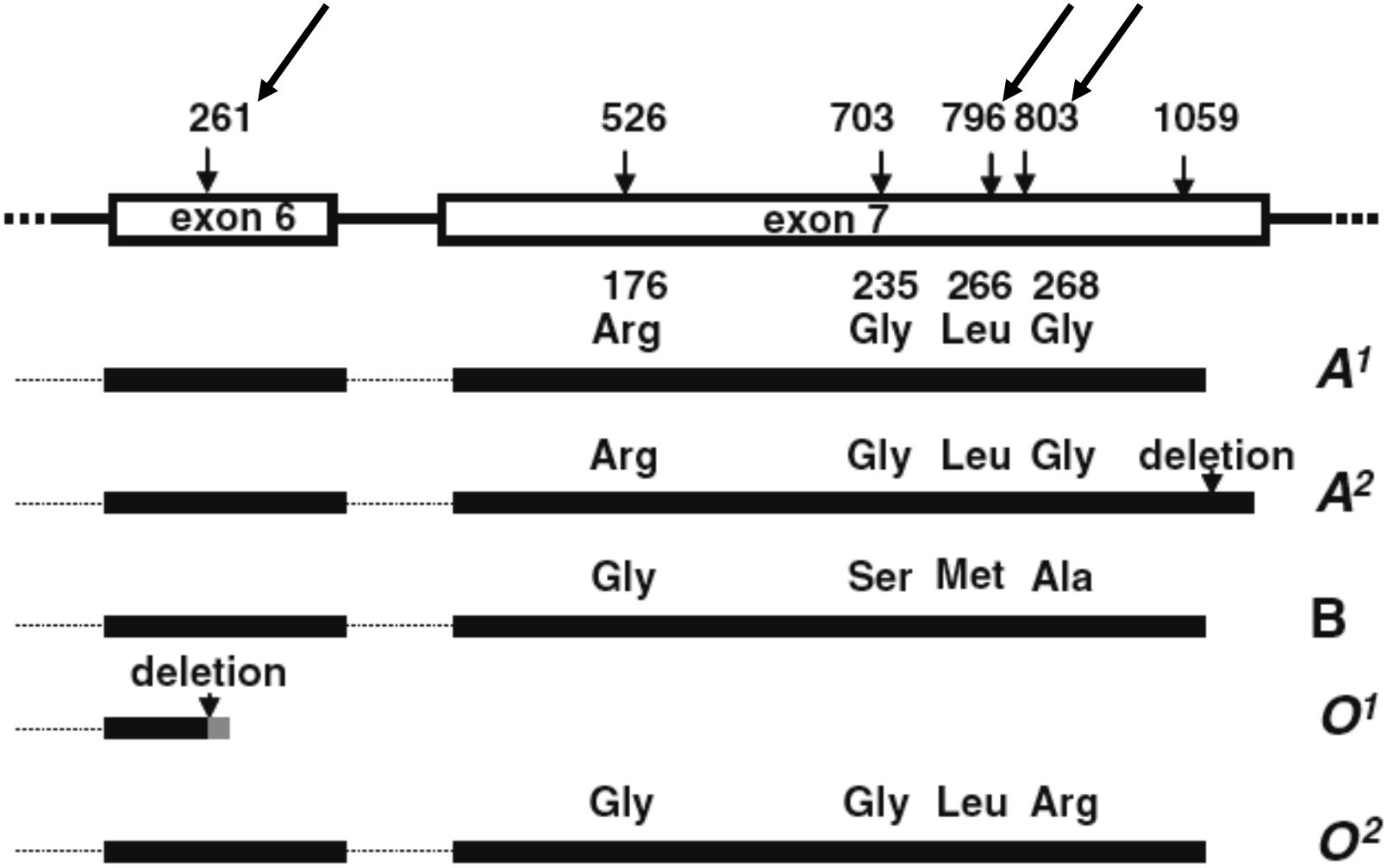
- Identify Genes with Protein Altering Mutations
- Determine Variation in Specific Genes

Genes with Protein Altering Variants



KAISER PERMANENTE®

Genetic Variation Determines ABO Blood Type



KAISER PERMANENTE®

Determination of ABO Blood Type

Population	Position			Genotype	Blood Type
	261	796	803		
Utah					
Sample 1	-/-	G/G	C/C	OO	O
Sample 2	-/-	G/G	C/C	OO	O
Sample 3	-/-	G/G	C/C	OO	O
Sample 4	C/-	G/G	C/C	AO	A
Sample 5	C/-	G/G	C/C	AO	A
Costa Rica					
Sample 6	C/-	G/G	C/C	AO	A
Sample 7	C/-	G/G	C/C	AO	A
Sample 8	-/-	G/G	C/C	OO	O
Sample 9	C/?	G/G	C/C	AO or AA	A
Sample 10	-/-	G/G	C/C	OO	O

KAISER PERMANENTE®

Huntington's Disease Testing Has Conseqeunces

Direction of Predictive-Testing Results and Frequency of CEs

Test and Result	No. (%) of Participants		
	Without a CE[a]	With a CE[b]	*P*[c]
Linkage analysis:			
Increased risk	251 (34.3)	6 (66.7)	Not significant
Decreased risk	451 (61.6)	3 (33.3)	
Uninformative result	30 (4.1)	0 (0)	
Total	732 (100)	9 (100)	
Direct test:			
Mutation carrier	1,529 (40.8)	31 (88.6)	<.0001
Normal CAG length	2,143 (57.1)	4 (11.4)	
Intermediate allele	79 (2.1)	0 (0)	
Total	3,751 (100)	35 (100)	
Totals for both linkage analysis and direct testing:			
Increased risk/mutation carrier	1,780 (39.7)	37 (84.1)	<.0001
Decreased risk/normal CAG length	2,594 (57.9)	7 (15.9)	
Uninformative result by linkage analysis	30 (.7)	0 (0)	
Intermediate allele	79 (1.8)	0 (0)	
Grand total	4,483 (100)	44 (100)	

Almqvist AJHG 1999

KAISER PERMANENTE®

Links to Videos and Articles on Seqeuncing

http://topics.bloomberg.com/the-dna-conundrum/

http://www.bloomberg.com/video/84364498-genetic-counselor-charts-reporter-s-family-history.html/

http://www.bloomberg.com/video/86406762-bloomberg-news-reporter-learns-genome-results.html/

http://www.bloomberg.com/video/bloomberg-reporter-gets-second-opinion-on-dna-test-a~bxIq0BTPyoD_iEV2EcwA.html/

Next Week

- Putting it all Together: Incorporating Molecular and Genetic Measures into Your Research