Single Nucleotide Polymorphisms and Copy Number Variation

Francesca Demichelis, PhD
Dep. of Pathology and Laboratory Medicine
Institute for Computational Biomedicine
Weill Cornell Medical College

6 billion nucleotides organized in two sets of 23 chromosomes

(22 autosomes, 2 sex chromosomes)

DNA encodes 30,000 genes

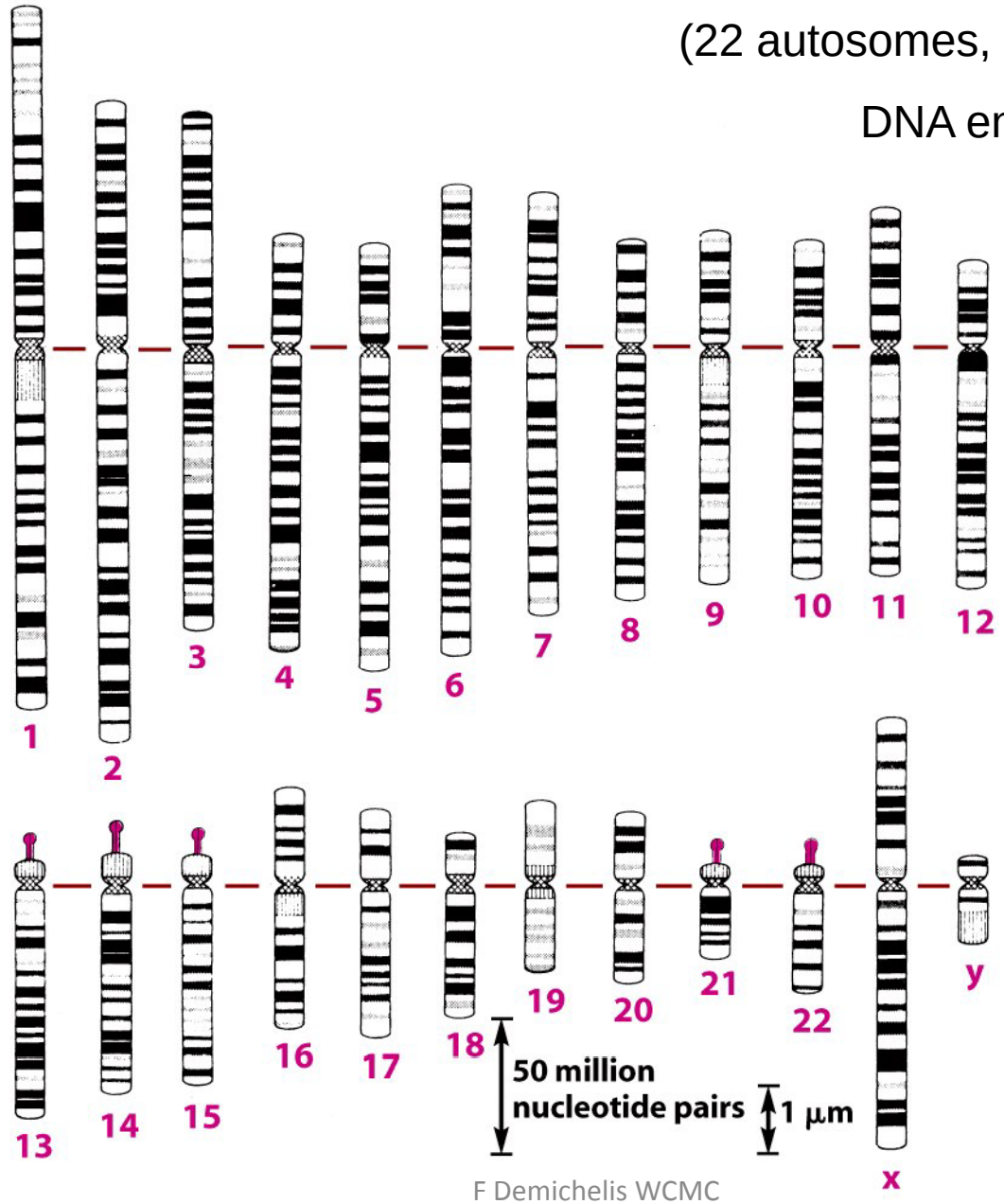


Figure 1-11a The Biology of Cancer (© Garland Science 2007)

1 page == 1 stretch of DNA (gene)

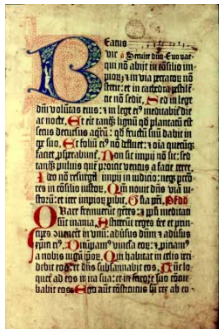
DIVERSITY

Single Nucleotide Polymorphism (SNP)

Mouse

Moose

Copy Number Variants (CNV)



1 copy

5 copies

Outline

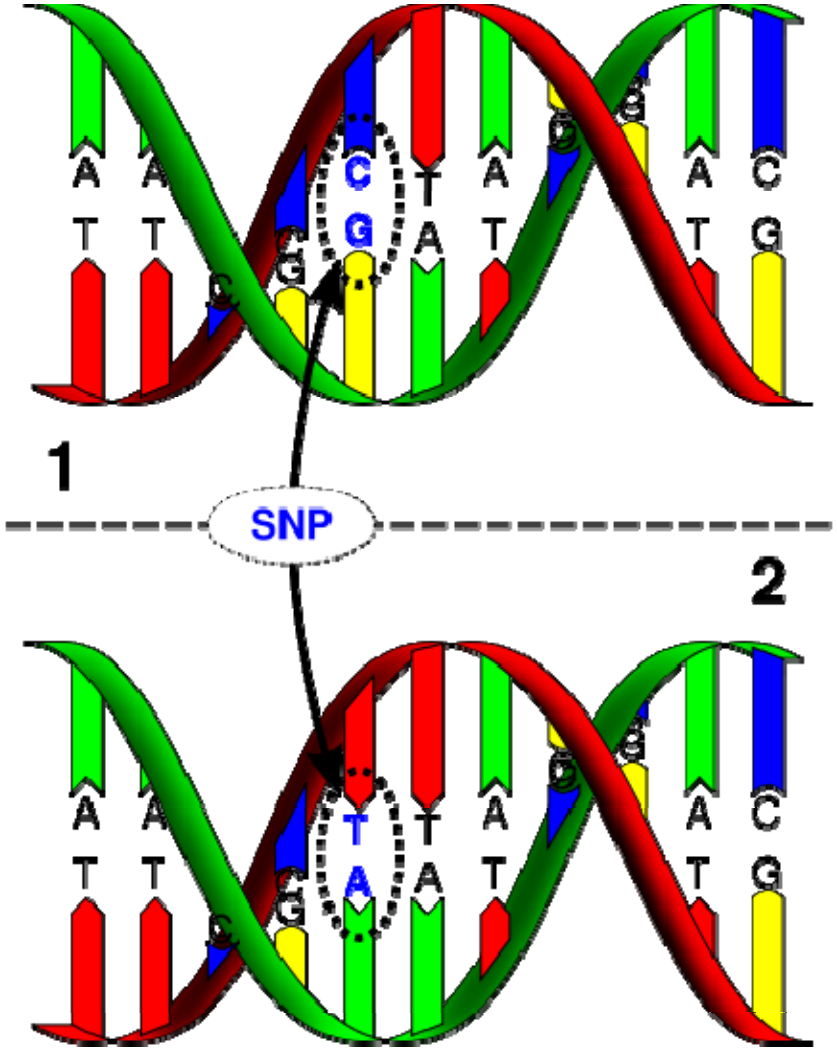
- ✓ Genome diversity
- ✓ Copy Number Variant – data evaluation
- ✓ SNP Panel Identification Assay

GENETIC VARIATION (1)

Point mutations

(sequence variation affecting single amino acid - **Single Nucleotide Polymorphism** SNP)

A- Adenine; T – Thymine;
C- Cytosine; G – Guanine.



~0.1% difference in the genomes of 2 unrelated individuals

GENETIC VARIATION (1)

- Non-coding region (SNPs that are not in protein-coding regions may still have consequences for gene splicing, transcription factor binding, or the sequence of non-coding RNA).

- Coding region:

  - o Synonymous (both forms lead to the same polypeptide sequence is termed synonymous (sometimes called a silent mutation));

  - o Nonsynonymous (if a different polypeptide sequence is produced);

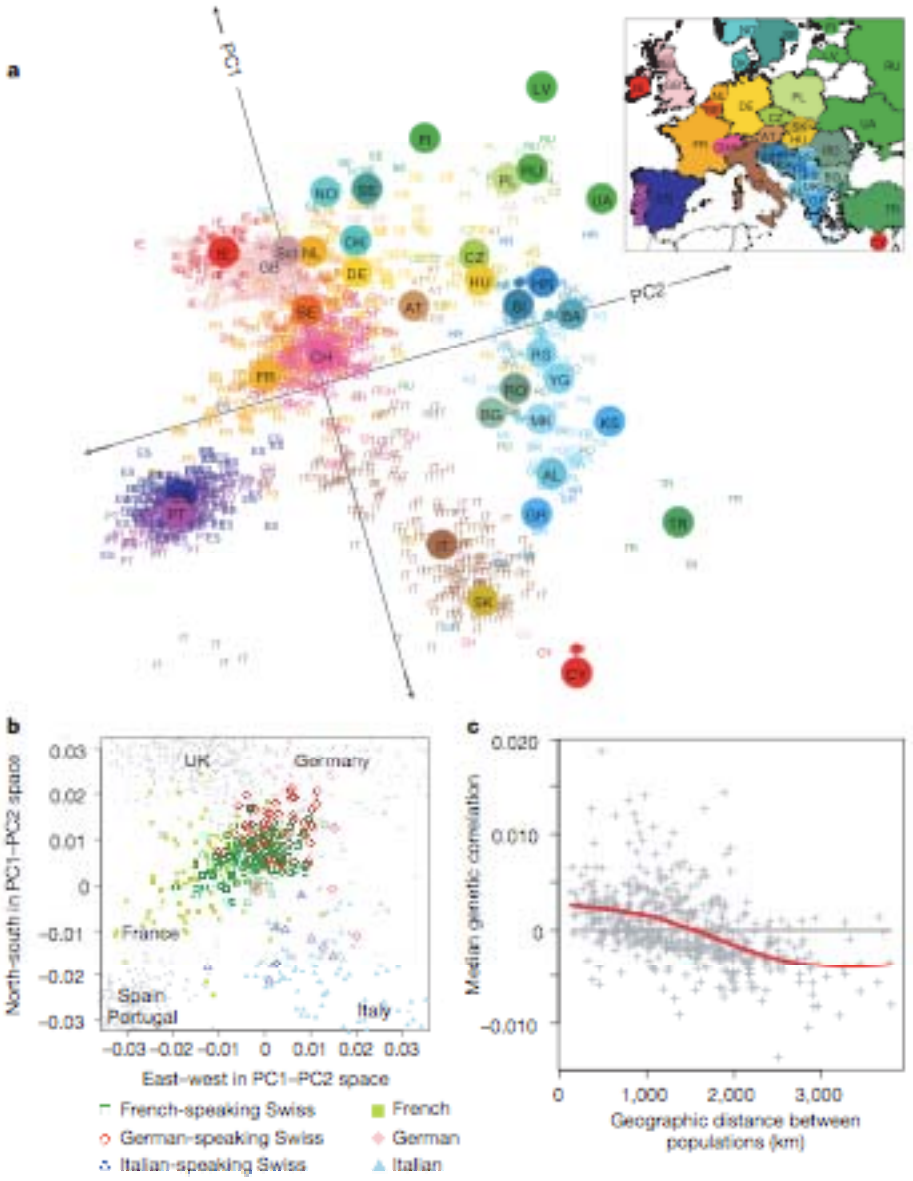
Variations in the DNA sequences of humans can affect how humans develop diseases and respond to pathogens, chemicals, drugs, vaccines, and other agents.

Variations in the DNA sequences of humans help understanding the genetic structure of human populations

Example 1: genetic ancestry

Nature 2008, Novembre et al: Genes mirror geography within Europe

[…*The results emphasize that when mapping the genetic basis of a disease phenotype, spurious associations can arise if genetic structure is not properly accounted for. In addition, the results are relevant to the prospects of genetic ancestry testing; an individual's DNA can be used to infer their geographic origin with surprising accuracy— often to within a few hundred kilometres*…]



Principal component analysis

Vol 456|6 November 2008|doi:10.1038/nature07331

Example 2: disease susceptibility

Prostate cancer has strong genetic component

The New England Journal of Medicine

© Copyright, 2000, by the Massachusetts Medical Society

VOLUME 343 JULY 13, 2000 NUMBER 2

ENVIRONMENTAL AND HERITABLE FACTORS IN THE CAUSATION OF CANCER

Analyses of Cohorts of Twins from Sweden, Denmark, and Finland

PAUL LICHTENSTEIN, PH.D., NIELS V. HOLM, M.D., PH.D., PIA K. VERKASALO, M.D., PH.D., ANASTASIA ILIADOU, M.SC., JAAKKO KAPRIO, M.D., PH.D., MARKKU KOSKENVUO, M.D., PH.D., EERO PUKKALA, PH.D., AXEL SKYTTHE, M.SC., AND KARI HEMMINKI, M.D., PH.D.

TABLE 3. EFFECTS OF HERITABLE AND ENVIRONMENTAL FACTORS IN CANCERS AT VARIOUS SITES, ACCORDING TO DATA FROM THE SWEDISH, DANISH, AND FINNISH TWIN REGISTRIES.

SITE OR TYPE	PROPORTION OF VARIANCE (95% CI)*			FIT OF MODEL	
	HERITABLE FACTORS	SHARED ENVIRONMENTAL FACTORS	NONSHARED ENVIRONMENTAL FACTORS	χ^2 (df)	P VALUE
Stomach	0.28 (0–0.51)	0.10 (0–0.34)	0.62 (0.49–0.76)	8.9 (38)	1.0
Colorectum	0.35 (0.10–0.48)	0.05 (0–0.23)	0.60 (0.52–0.70)	25.8 (38)	0.93
Pancreas†	0.36 (0–0.53)	0 (0–0.35)	0.64 (0.47–0.86)	0.5 (3)	0.92
Lung	0.26 (0–0.49)	0.12 (0–0.34)	0.62 (0.51–0.73)	28.1 (38)	0.88
Breast‡	0.27 (0.04–0.41)	0.06 (0–0.22)	0.67 (0.59–0.76)	10.1 (18)	0.93
Cervix uteri†‡	0 (0–0.42)	0.20 (0–0.35)	0.80 (0.57–0.97)	0.3 (3)	0.96
Corpus uteri‡	0 (0–0.35)	0.17 (0–0.31)	0.82 (0.64–0.98)	6.6 (18)	0.99
Ovary‡	0.22 (0–0.41)	0 (0–0.24)	0.78 (0.59–0.99)	6.0 (18)	1.0
Prostate§	0.42 (0.29–0.50)	0 (0–0.09)	0.58 (0.50–0.67)	26.5 (18)	0.09
Bladder†	0.31 (0–0.45)	0 (0–0.28)	0.69 (0.53–0.86)	1.7 (3)	0.64
Leukemia†	0.21 (0–0.54)	0.12 (0–0.41)	0.66 (0.45–0.88)	0.0 (3)	0.99

Polymorphisms and risk for PCA

Table 2. Association of SNPs at Five Chromosomal Regions with Prostate Cancer.*

SNP	Chromosomal Region	Position†	Alternative Alleles	Associated Allele§	Allelic Tests: Frequency	Allelic Tests: Odds Ratio (95% CI)¶	Allelic Tests: P Value	Best-Fitting Genetic Model‡: Model	Best-Fitting Genetic Model‡: Genotype‖	Best-Fitting Genetic Model‡: Odds Ratio (95% CI)	Best-Fitting Genetic Model‡: P Value**
rs4430796	17q									1.40 (1.23–1.59)	2.68×10^{-7}
rs7501939	17q									1.33 (1.17–1.50)	5.54×10^{-6}
rs3760511	17q									1.42 (1.20–1.68)	4.47×10^{-5}
rs1859962	17q2									1.28 (1.12–1.46)	3.54×10^{-4}
rs7214479	17q2									1.15 (1.00–1.32)	0.06
rs6501455	17q2									1.13 (0.99–1.29)	0.06
rs983085	17q2									1.11 (0.97–1.26)	0.12
rs6983561	8q24 (re									CC 1.60 (1.28–2.00)	2.14×10^{-5}
rs16901979	8q24 (re									CA 1.60 (1.28–2.01)	2.14×10^{-5}
rs6983267	8q24 (re									GG 1.38 (1.19–1.59)	1.74×10^{-5}
rs7000448	8q24 (re									TT 1.18 (1.04–1.33)	1.21×10^{-2}
rs1447295	8q24 (re									AA 1.26 (1.10–1.44)	8.27×10^{-4}
rs4242382	8q24 (re									AA 1.29 (1.12–1.47)	2.53×10^{-4}
rs7017300	8q24 (re									CC 1.20 (1.05–1.36)	6.20×10^{-3}
rs10090154	8q24 (re									TT 1.31 (1.14–1.50)	1.03×10^{-4}
rs7837688	8q24 (re									TT 1.21 (1.06–1.39)	5.87×10^{-3}
rs3760511	17q12			Recessive	AA or CA	CC	1.42 (1.20–1.68)	4.47×10^{-5}			
rs16901979	8q24 (region 2)			Dominant	CC	AA or CA	1.60 (1.28–2.01)	2.14×10^{-5}			

* CI denotes confidence interval, and SNP single-nucleotide polymorphism.
† The position is based on the National Center for Biotechnology Information database, build 35.
‡ The best-fitting model for each SNP was determined after testing associations of a series of genetic models, including dominant and recessive models, with prostate cancer.
§ These alleles were reported to be associated with prostate cancer in studies published previously.[4-6,10]
¶ Allelic odds ratios are based on the multiplicative model.
‖ Reference genotypes and those associated with prostate cancer for each SNP were defined on the basis of the best-fitting genetic model.
** P values are two-sided and were calculated by the likelihood-ratio test with one degree of freedom, adjusted for age and geographic region.

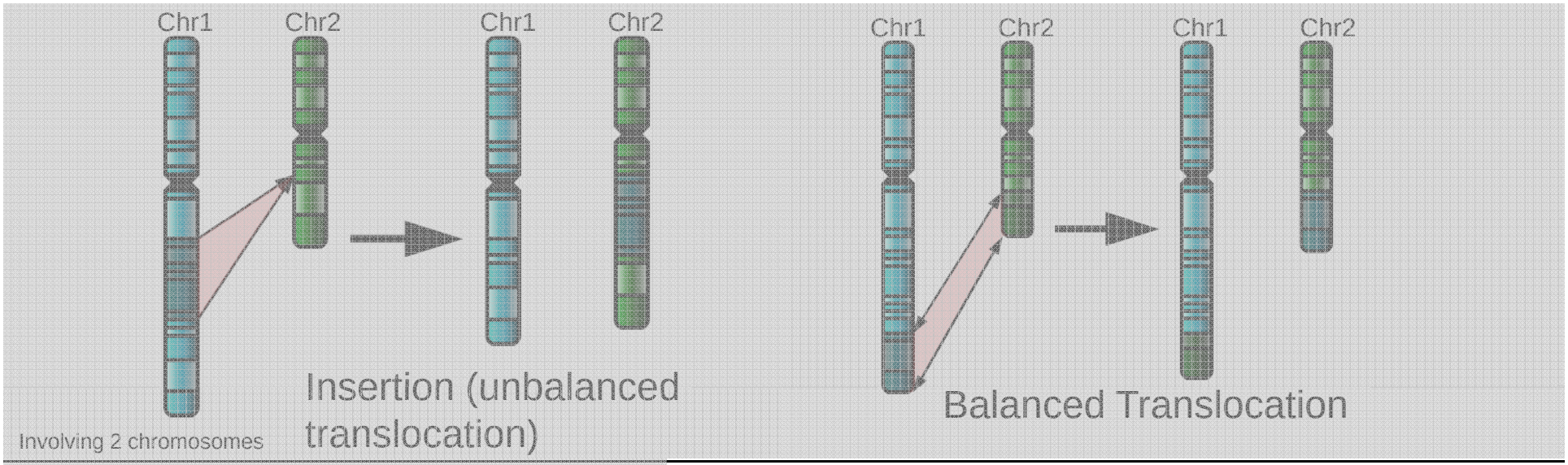
Cumulative Association of Five Genetic Variants with Prostate Cancer

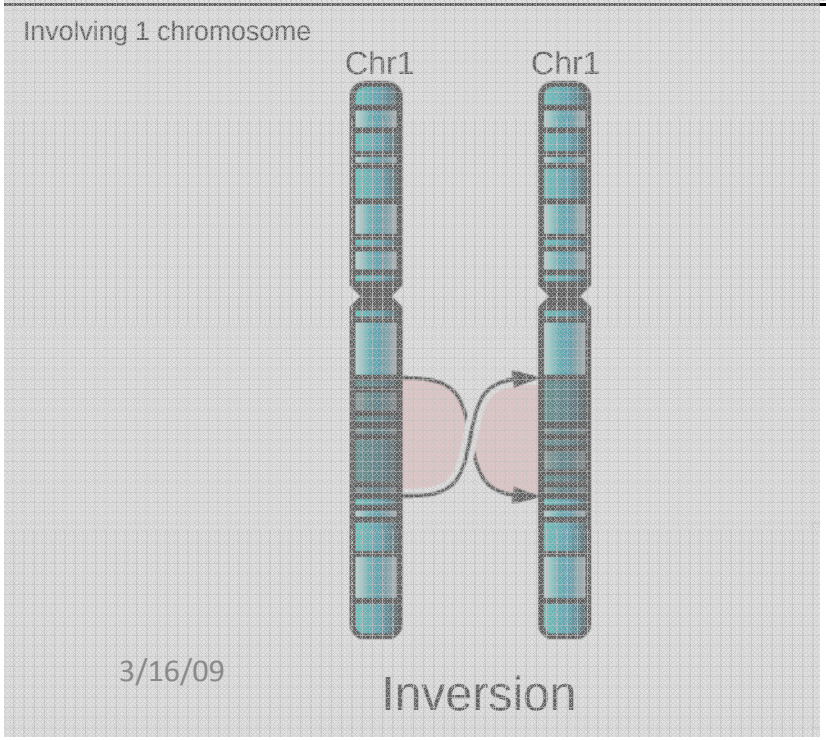
Zheng et al

N ENGL J MED 10.1056/NEJMoa075819

Contingency tables, odds ratios, Fisher test, p-value, CI

GENETIC VARIATION (2)





Variation of DNA 'quantity'

==

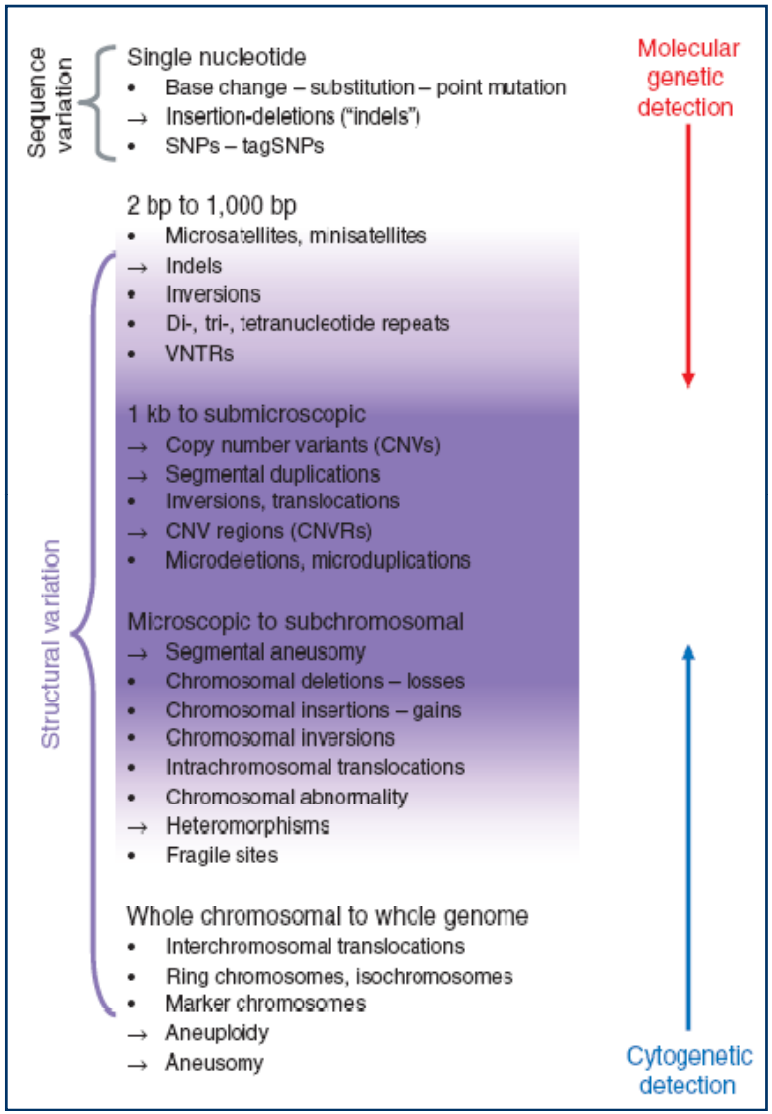
Copy Number Variants

Deletion

Duplication

BACKGROUND

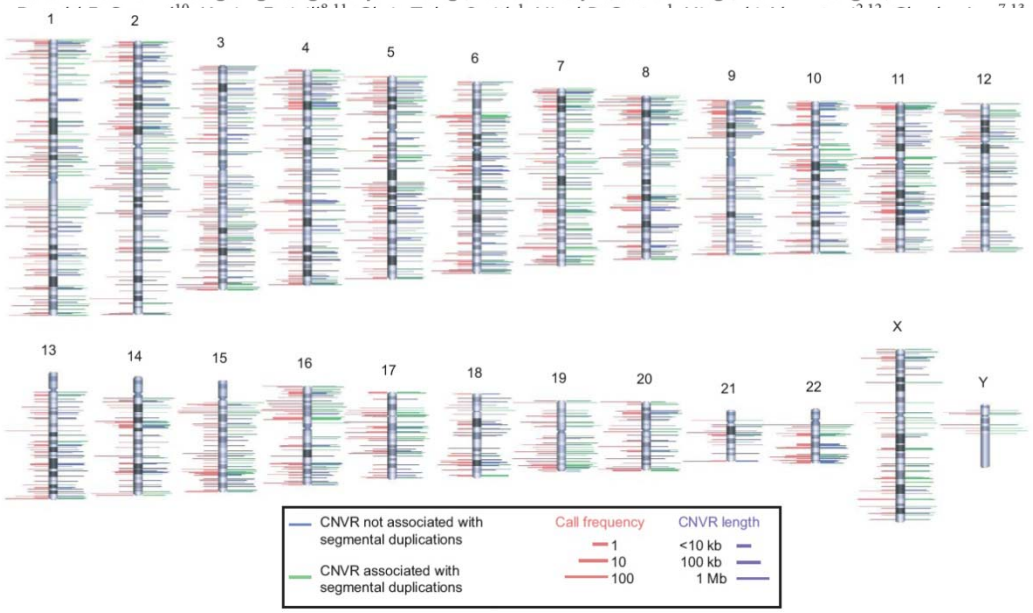
LEXICON - CYTOGENOMICS



Global variation in copy number in the human genome

Nature 2006

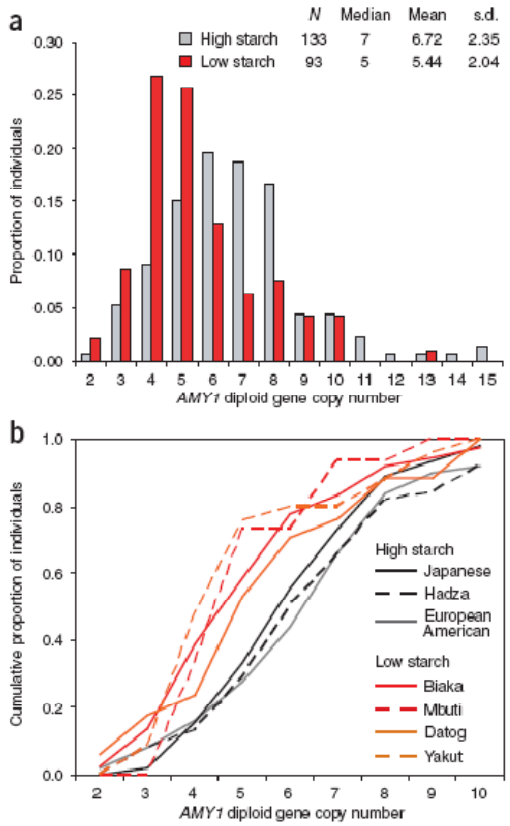
Richard Redon[1], Shumpei Ishikawa[2,3], Karen R. Fitch[4], Lars Feuk[5,6], George H. Perry[7], T. Daniel Andrews[1], Heike Fiegler[1], Michael H. Shapero[4], Andrew R. Carson[5,6], Wenwei Chen[4], Eun Kyung Cho[7], Stephanie Dallaire[7], Jennifer L. Freeman[7], Juan R. González[8], Mònica Gratacòs[8], Jing Huang[4], Dimitrios Kalaitzopoulos[1], Daisuke Komura[3], Jeffrey R. MacDonald[5], Christian R. Marshall[5,6], Rui Mei[4], Lyndal Montgomery[1], Kunihiro Nishimura[2], Kohji Okamura[5,6], Fan Shen[4], Martin J. Somerville[9], Joelle Tchinda[7], Armand Valsesia[1], Cara Woodwark[1], Fengtang Yang[1], Junjun Zhang[5], Tatiana Zerjal[1], Jane Zhang[4], Lluis Armengol[8],



Scherer S et al

VOLUME 39 | JULY 2007 | NATURE GENETICS SUPPLEMENT
nature genetics

Dosage variable CNVs - example



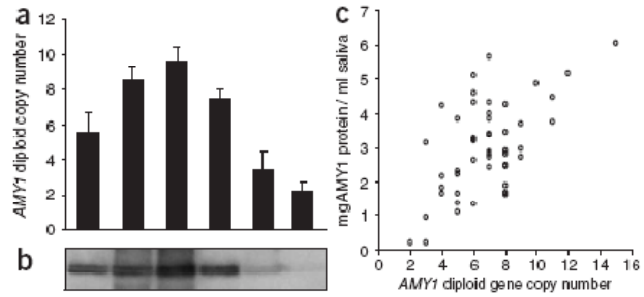


Figure 1 *AMY1* copy number variation and salivary amylase protein expression. (**a**,**b**) For the same European American individuals, we estimated diploid *AMY1* gene copy number with qPCR (**a**) and estimated amylase protein levels in saliva by protein blot (**b**). Error bars indicate s.d. (**c**) Relationship between *AMY1* diploid copy number and salivary amylase protein level ($n = 50$ European Americans). A considerable amount of variation in AMY1 protein expression is not explained by copy number ($R^2 = 0.351$), which may reflect other genetic influences on *AMY1* expression, such as regulatory region SNPs or nongenetic factors that may include individual hydration status, stress level and short-term dietary habits.

Perry GH et al, Nat Gen 2007

- CN of the salivary amylase gene (AMY1) correlates with salivary amylase protein level

- Individuals from high-starch diet populations have higher number of AMY1 copies

- <u>CNV formation event maybe sufficient to specifically promote gene expression modification;</u>

<u>thus gene copy number changes may facilitate evolutionary adaptation involving protein</u>

<u>abundance change.</u>

Why are copy number variants important/interesting?

- Hundreds of CNVs per individual. 20% potentially affecting protein-coding genes.

- CNV genesis occurs at higher rate than point mutation (1e-4/1e-6 vs 2e-8 per generation). They carry different information.

- DNA variation extension: CNVs more likely to affect coding sequence than point mutation.

- Relationship bw gene dosage and mRNA expression is basis for phenotypic traits (disease susceptibility and dietary preferences).

- CNVs tend to affect specific gene functional categories (such as environmental response related and not basic cellular processes related)

- Duplication of genetic material is common cause of protein birth (protein evolution).

? Characterization is not complete

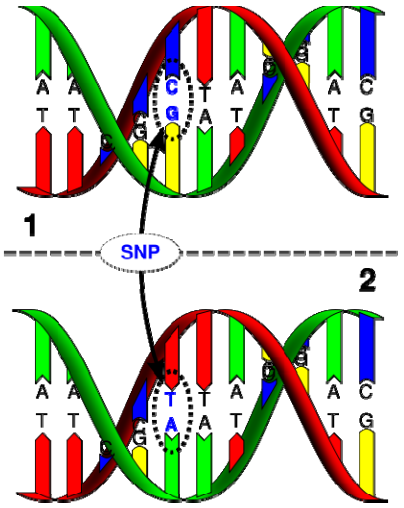
? Mechanism behind formation of CNVs is not clear. Formation bias: non uniform distribution along chromosome.

Korbel J, et al, review 2008

GENETIC VARIATION

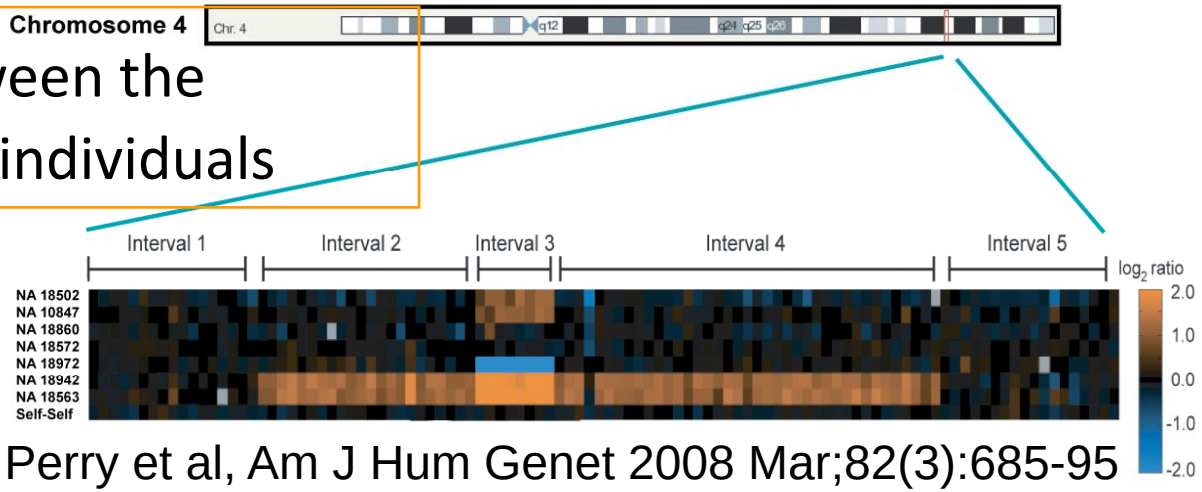
Single Nucleotide Polymorphism (SNP)

~ 0.1% difference between the genomes of 2 unrelated individuals



Copy Number Variants (CNV)

~0.4% difference between the genomes of 2 unrelated individuals



Perry et al, Am J Hum Genet 2008 Mar;82(3):685-95

Importance of polymorphisms

- Population study/Evolution
- Phenotype/traits
- Disease risk

How to evaluate Copy Number Variants in high-throughput fashion

Arrays – aiming at high resolution AND high sensitivity

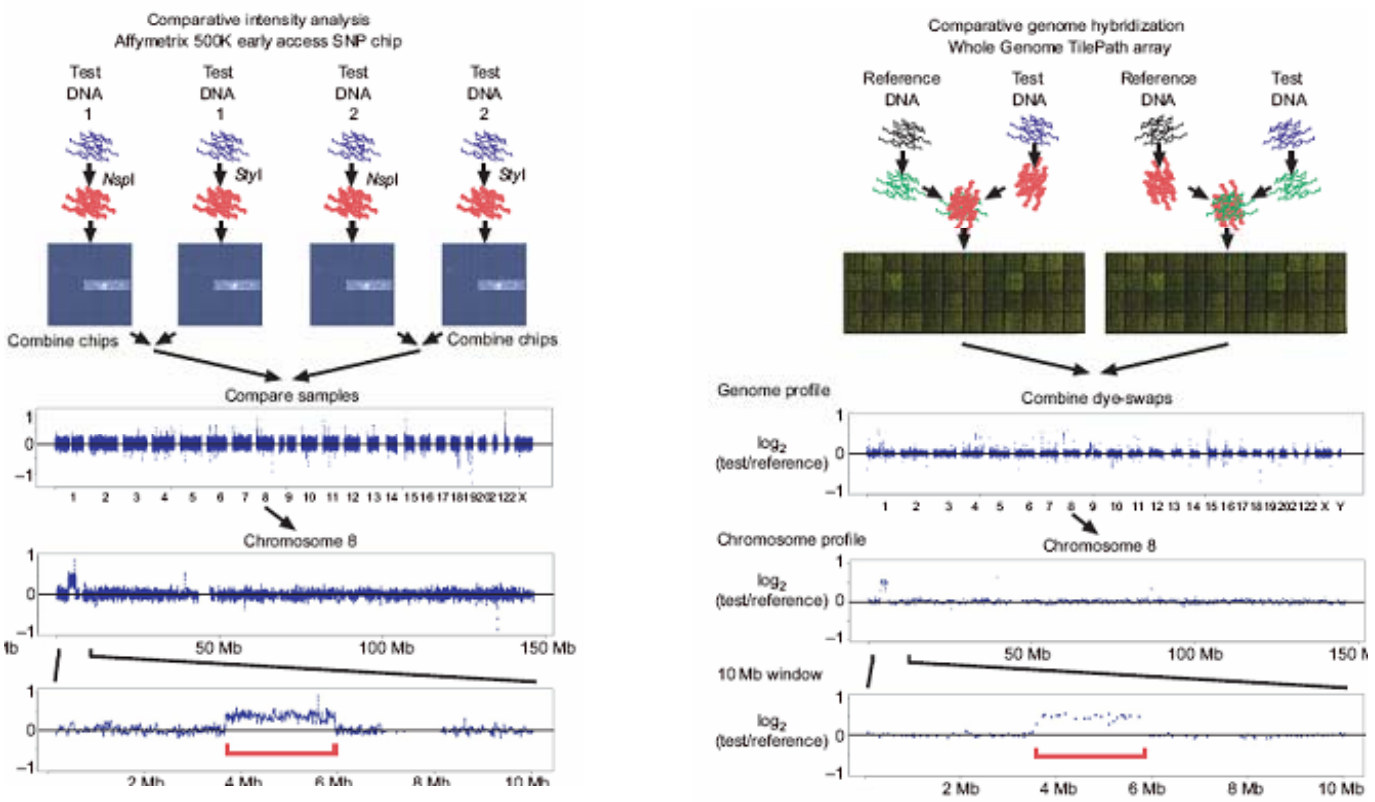
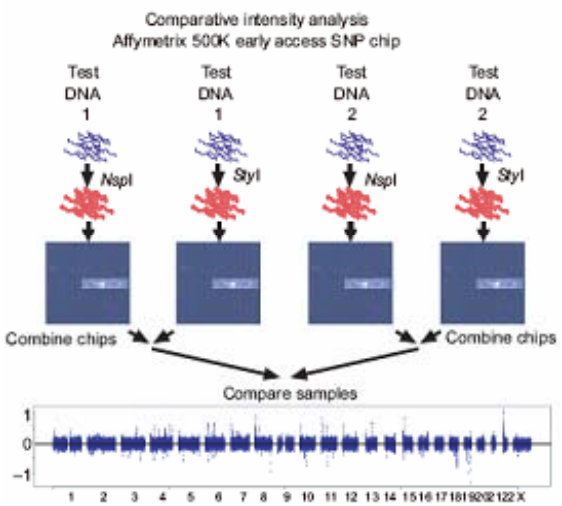


Figure 1 | Protocol outline for two CNV detection platforms. The experimental procedures for comparative genome hybridization on the WGTP array and comparative intensity analysis on the 500K EA platform are shown schematically (see Supplementary Methods for details), for a comparison of two male genomes (NA10851 and NA19007). The genome profile shows the $\log_2$ ratio of copy number in these two genomes chromosome-by-chromosome. The 500K EA data are smoothed over a five-probe window. Below the genome profiles are expanded plots of chromosome 8, and a 10-Mb window containing a large duplication in NA19007 identified on both platforms (indicated by the red bracket).

Nature 2006

SNP arrays



1) Noisier data

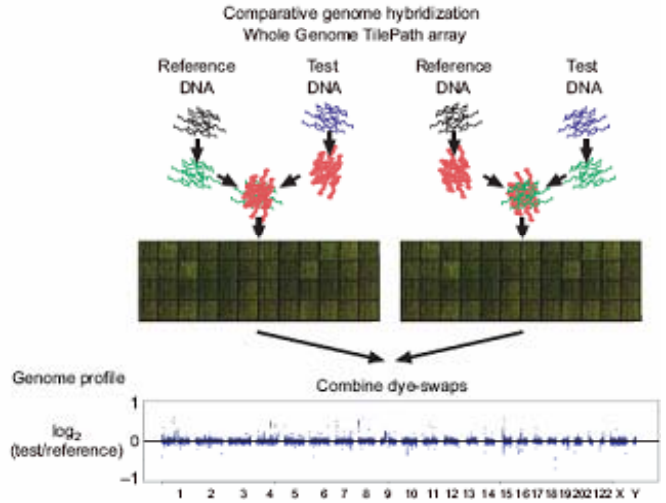
2) Genotype call for SNP markers

Allele specific information:

S1_A1 : 1 copy S1_A2 : 1 copy	S1 : 2 copies
S1_A1 : 0 copy S1_A2 : 2 copies	S1 : 2 copies

3) Reference sample at data level

aCGH



1) High signal to noise ratio

2) No SNP information

--

3) Reference sample at experimental level

Data Signal Processing

A. Preprocessing (normalization)

B. Quality Control (exclusion of bad sample data based on predefined measures)

B. Evaluation of ratio of target/reference signal on a ‘marker’ basis

C. Analysis of signal along the genome (consider signal values of neighboring markers to control for noise and to define breakpoints of copy number variations – segmentation)

Very easy in theory: for each locus/marker *i*:

*log2(*CN2_target/CN2_reference*)* = *0* *(*Normal)

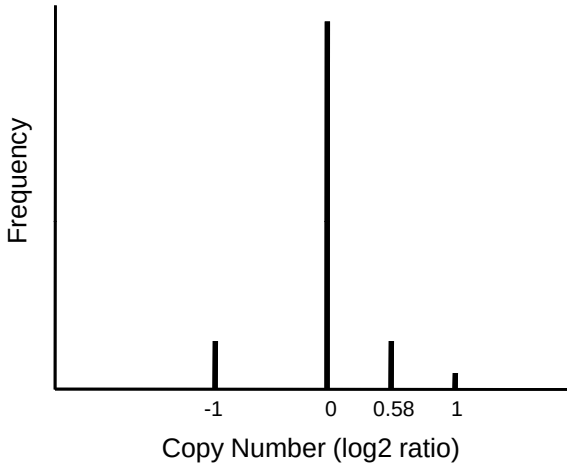
*log2(*CN1_target/CN2_reference*)* = *-1* *(*Hemizygous del)

*log2(*CN0_target/CN2_reference*)* = -∞ *(*Homozygous del)

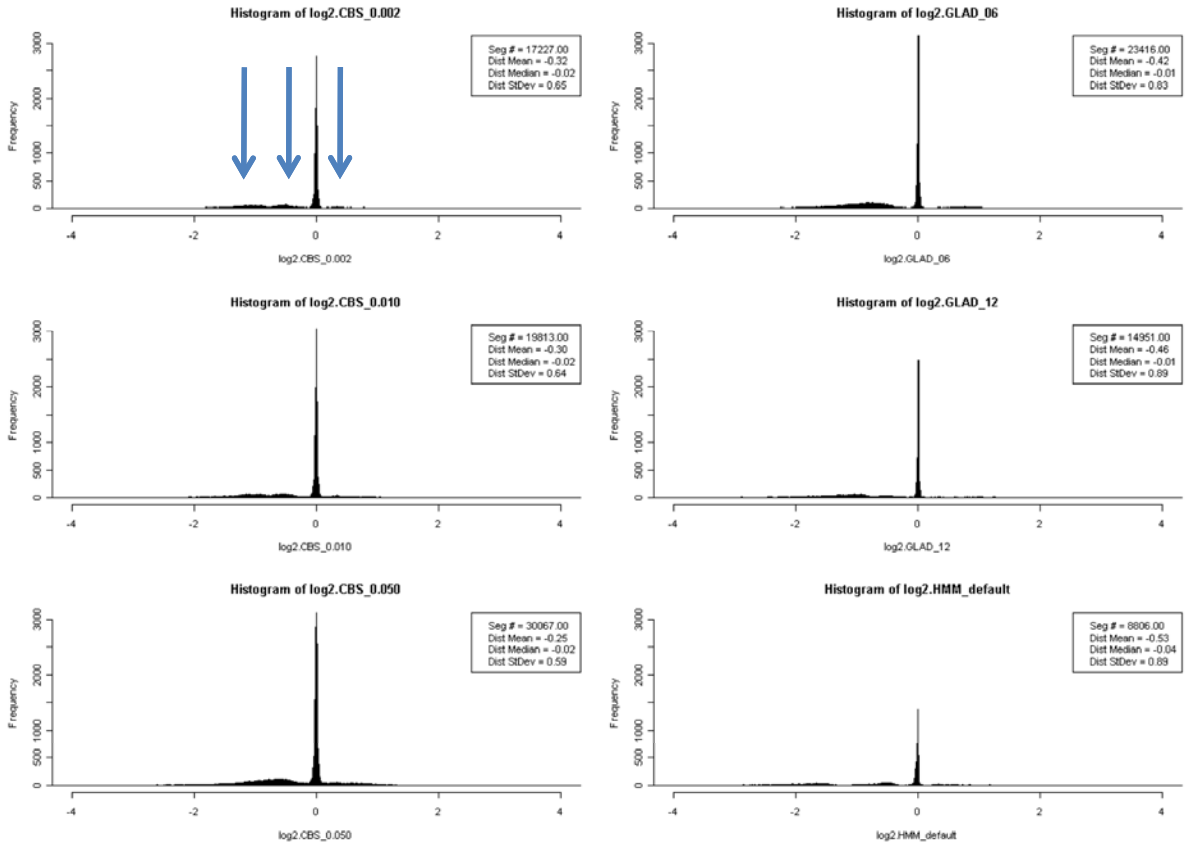
*log2(*CN3_target/CN2_reference*)* = *0.58* *(*Gain 1 copy)

Data Signal Processing (2) – efficiency/dynamic range

Expected data



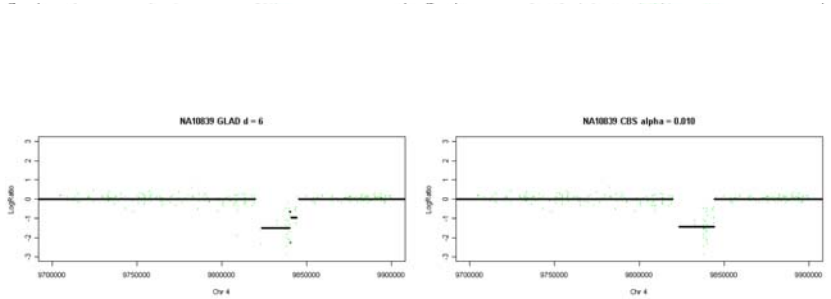
Real data

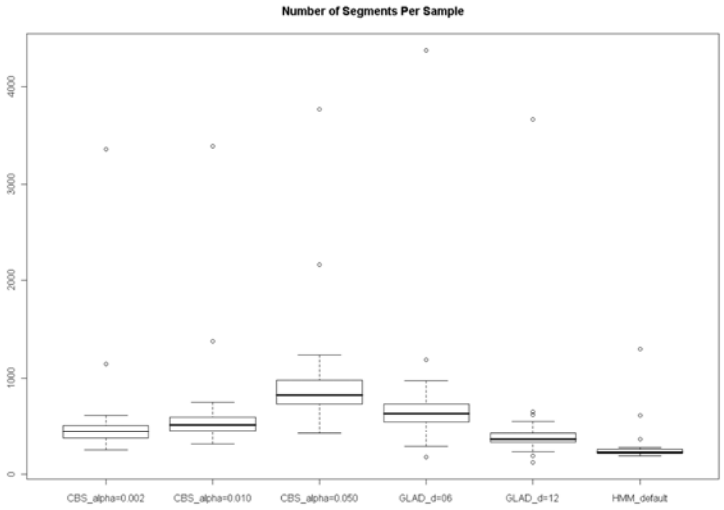


<u>To set thresholds for CN states, one needs</u>:

- gold standard
- previous experimentally validated data;
- ad hoc experiment (FISH, PCR absolute curve)

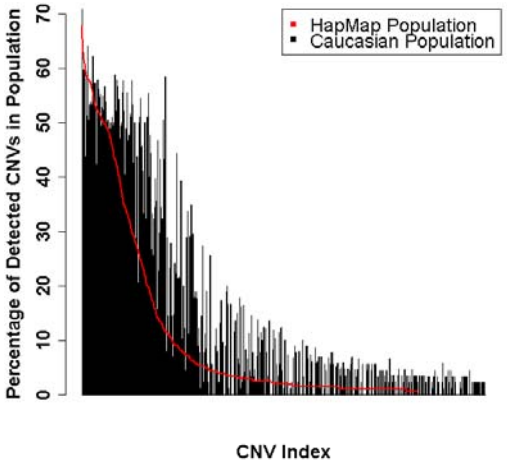
Data Signal Processing (3) – parameter tuning





Images removed

Data Signal Processing (4) – reference model



Images removed

No perfect data / no perfect analysis

Image removed

Validation of interesting variants/results by different experimental procedure

Room for improvement for statistical approaches to process CN data (analytical approaches for threshold setting, breakpoint identification, correction for reference)

Diversity makes each individual unique

- DNA test for paternity
- legal issues
- identity check of biological material

SNP Panel Identification Assay (SPIA):

a genetic-based assay for the identification of samples

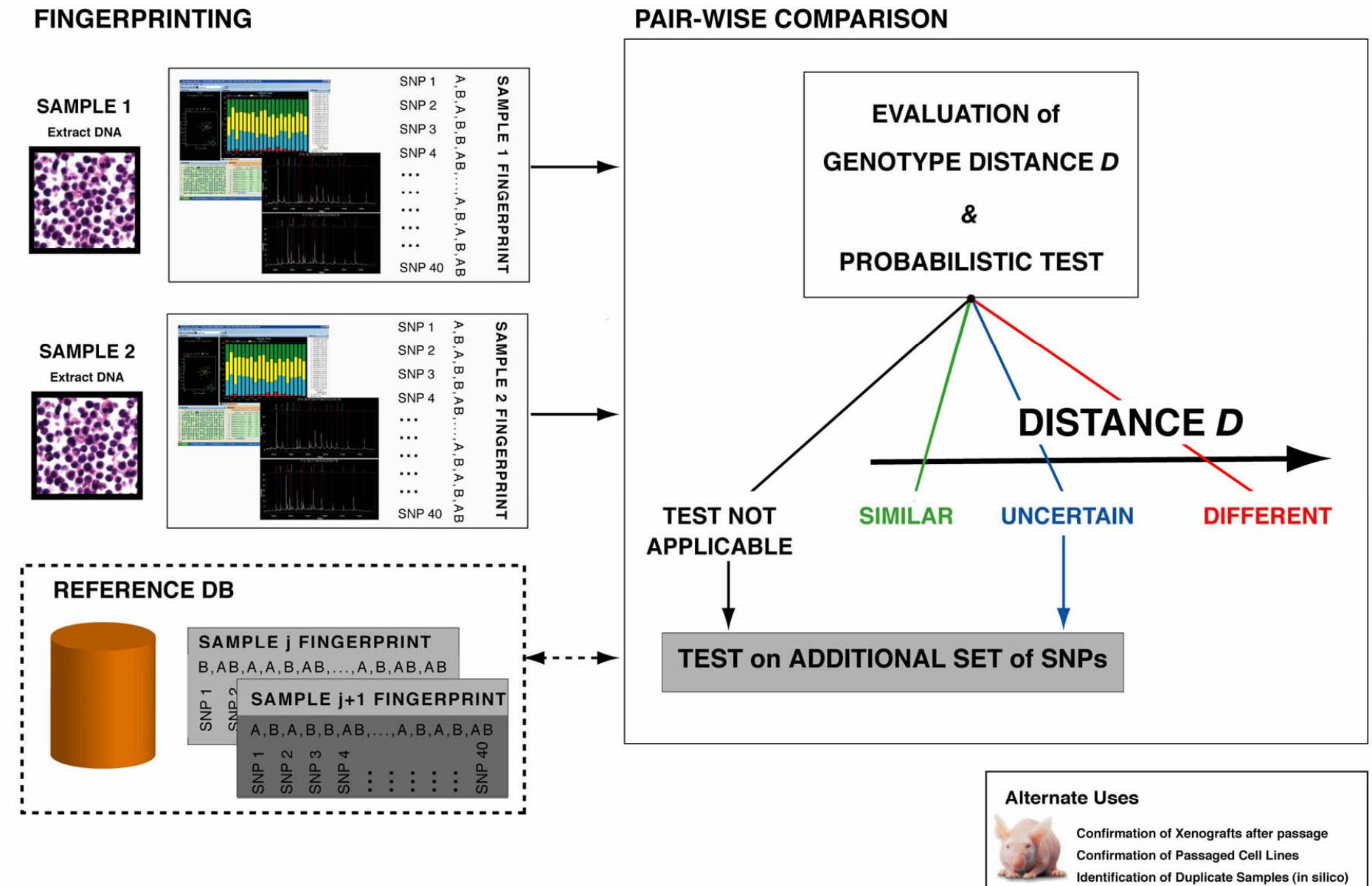
16 FEBRUARY 2007 VOL 315 **SCIENCE**

IN THE 1980S, WHEN HE WAS A postdoctoral fellow at the Scripps Research Institute in San Diego, California, Reinhard Kofler received what was supposed to be a human cancer cell line from a collaborator. "We cultured it, we cloned genes into it," he recalls, then "[we] genotyped it and realized it was 100% mouse."

- Risk in cell line maintenance is human error, either by mislabeling or cross-contamination. Potential common problem (MedLine search of 'cell line' and 'cancer human' identified 96,758 studies on cancer biology.)

- Cancer genomics involves the accurate verification of sample provenance, as well as continual tracking to ensure accurate identity.
- Population studies

SNP PANEL IDENTIFICATION ASSAY (SPIA)



Nucleic Acids Research, 2008

PROBLEM STATEMENT

Normal	A A B AB AB B B A A A AB AB A A B B AB B AB AB AB A A B
Tumor	A A B AB A B B A A A AB AB A A B B AB B AB AB AB A A B

1 mismatches out of 24 SNPs (concordance 95.8%)

Normal	A A B AB AB B B A A A AB AB A A B B AB B AB **na** AB A A B
Tumor	A A B AB A B B A A A AB AB A A B B AB B AB AB AB A A **na**

1 mismatches out of 22 SNPs (concordance 95.4%)

Normal	A A B AB AB B B A A A AB AB **na na na na na na na na na na na na**
Tumor	A A B AB AB B B A A A AB AB **na na na na na na na na na na na na**

0 mismatches out of 12 (100%)

How many markers to use? Confidence in calling identity?

TYPES of MISMATCHES

($CALL_{on\_NORMAL}$, $CALL_{on\_TUMOR}$)

BIOLOGICAL MISMATCHES

LOH:	(AB,A)	P(AB,A) = P(AB)* P(A|AB)
GOH:	(A,AB)	P(A,AB) = P(A)* P(AB|A)
Doub. Mut:	(A,B)	P(A,B) = P(A)* P(B|A)

EXPERIMENTAL MISMATCHES

GENOTYPE CALL ERROR RATE (PLATFORM and ALGORITHM DEPENDENT)

Mismatches are ‘possible’ and need to be considered

Probabilities can be SNP specific and/or tissue specific

The joint probability of two events E1 and E2, P(E1,E2) or P(E1 AND E2) is P(E1,E2) = P(E1) x P(E2|E1) and P(E2|E1) is called the conditional probability of E2 given E1.

APPROACH

To identify the ideal SNP panel, **which maximizes the probability of obtaining distinct genotype calls on different samples with 'reasonable confidence'**.

MULTI-STEP COMPUTATION to BUILD AND VALIDATE SPIA

1. definition of a *genotype distance* to compare samples,
2. filtering,
3. iterative procedure of training and test steps (with bootstrap) to identify best SNPs,
4. implementation of a probabilistic test (*different*, *uncertain*, or *similar*),
5. *in silico* validation on independent dataset,
6. *lab* validation on cell lines genotyped on independent platform (Sequenom).

DATASETS

50K genotype data of 155 cancer cell lines (CLs) derived from different organs.
SPIA panel genotype data of 93 CLs generated on on a mass spectrometer system (Sequenom).

COMPARING GENOTYPES

To count the number of loci where the two samples do not match and normalize on the total number of loci. This value is the ‘distance’ between the two samples.

This distance is proportional to the number of discordant calls.

$$d(s1_i, s2_i) = \begin{cases} 1 \text{ if } & s1_i \neq s2_i \\ 0 \text{ if } & s1_i = s2_i \end{cases}$$

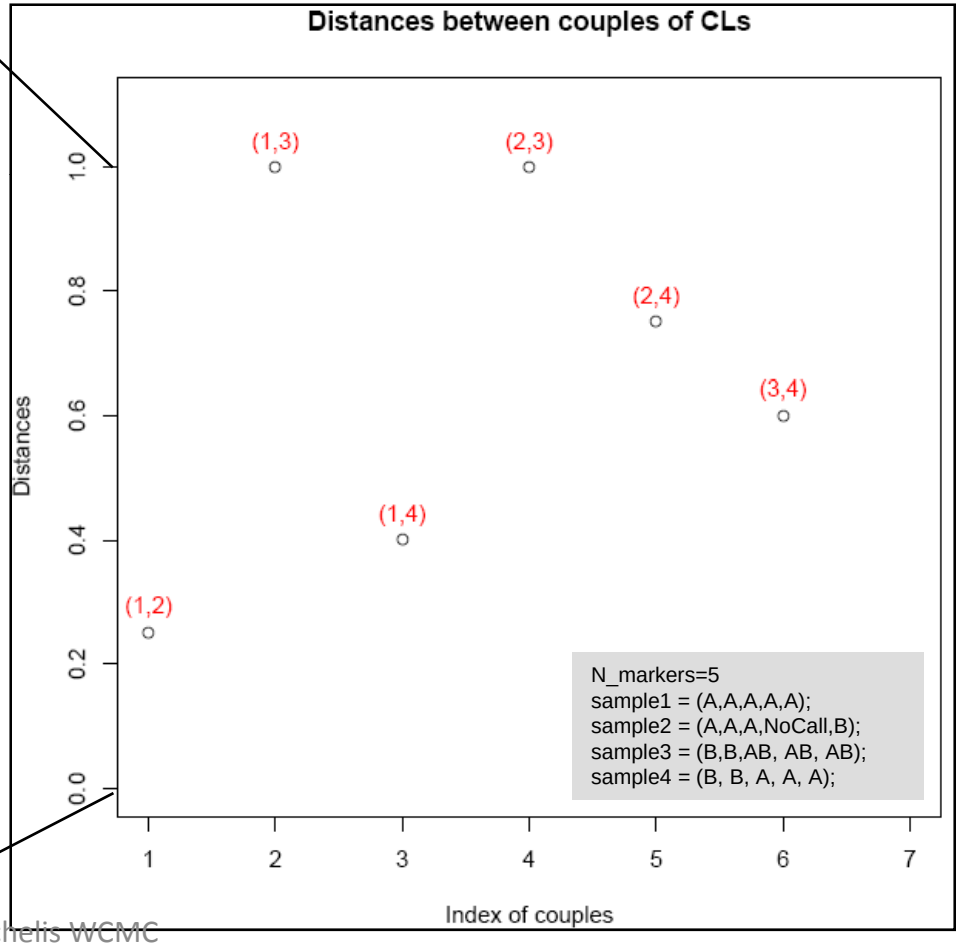
$$D(s1, s2) = \frac{\sum_{i=1..vNSNPs} w_i(j)\, d(s1_i, s2_i)}{\sum_{i=1..vNSNPs} w_i(j)}$$

$$vNSNPs <= NSNPs$$

w_i depends on type j (match, LOH, GOH, DM)

ALL **MISMATCHES**

ALL **MATCHES**



Pair-wise distance after SNP filtering

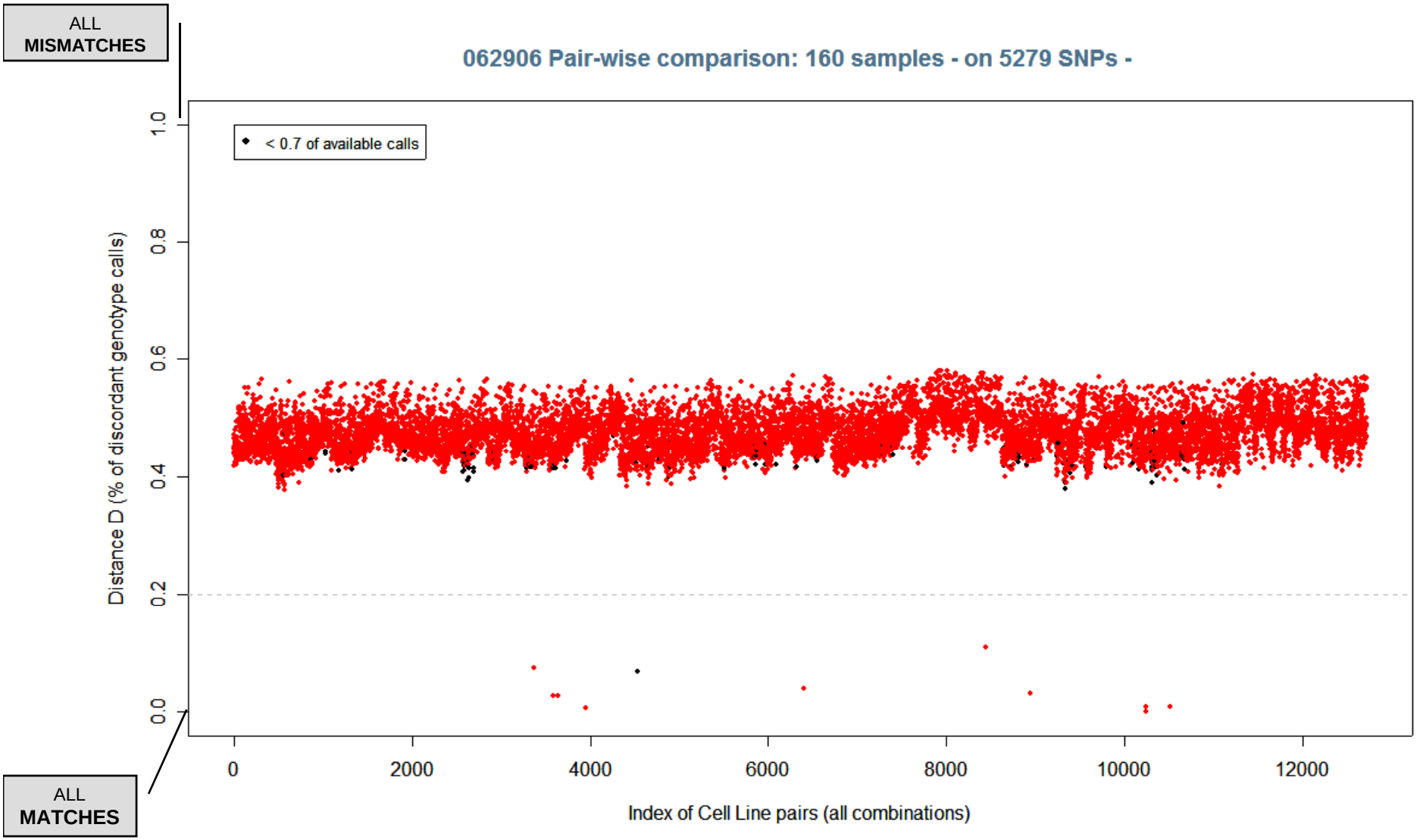
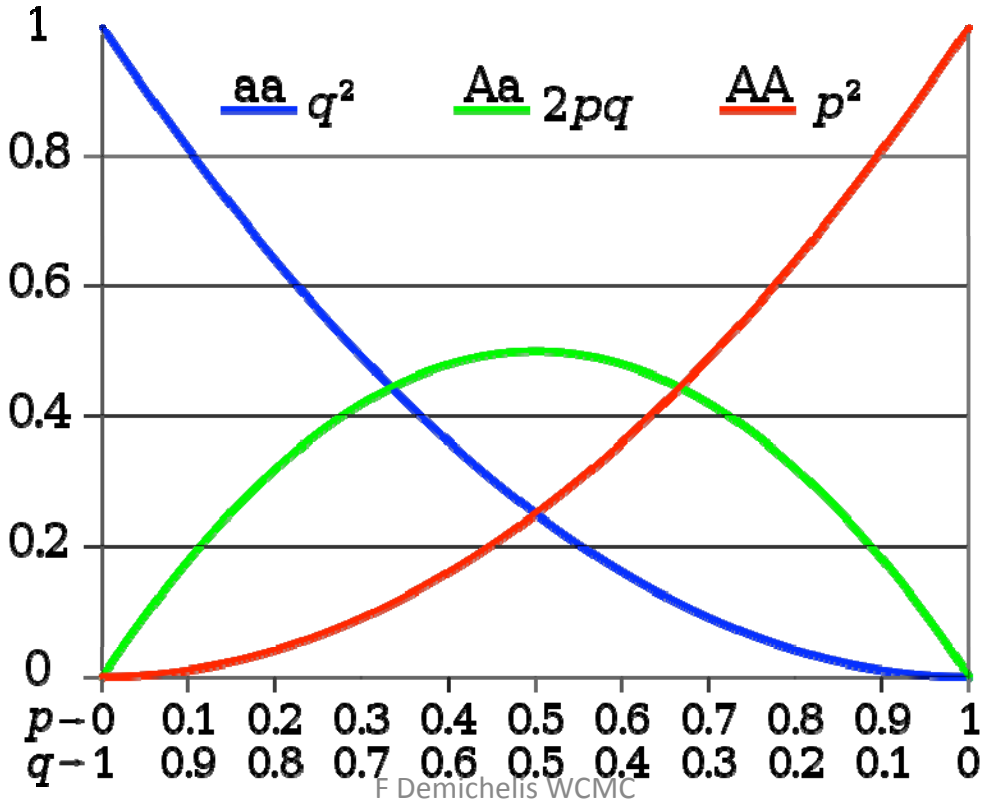


Table 1: List of cell lines which are detected to have very similar genotype profile evaluated on a set of 5.3K SNPs.

				Mismatches	
CL1 Name	**CL2 Name**	***D***	**% Valid Calls**	**Ho-Ho**	**Ho-Het**
M14	MDA.MB435	0.0747	0.794	2	311
MCF7	BT.20	0.0279	0.781	0	115
MCF7	KPL.1	0.0271	0.797	0	114
NCI.ADR.RES	OVCAR.8	0.0076	0.874	0	35
NCI.H460	H2195	0.0680	0.696	0	250
SNB.19	U251	0.0394	0.866	0	180
184A1	184B5	0.1084	0.978	37	523
BT.20	KPL.1	0.0308	0.831	1	134
H1450	H2141	0.0092	0.866	0	42
H1450	H220	0.0000	0.861	0	0
H2141	H220	0.0088	0.857	0	40

Hardy-Weinberg equilibrium

The Hardy–Weinberg principle: both allele and genotype frequencies in a population remain constant from generation to generation unless specific disturbing influences are introduced (as non-random mating, mutations, selection, limited population size, random genetic drift and gene flow).
Genetic equilibrium is an ideal state that provides a baseline to measure genetic change against.



From Wikipedia

Table 2: summary of pair-wise distances/differences varying the number of selected SNPs

	Set of CLs (133) used for the SNP **selection** process		Set of CLs (13) used for independent **validation**	
Number of SNPs	Mean *D* (stdev)	*D* Min-Max	Mean *D* (stdev)	*D* Min-Max
~58000(*)	0.3832	0.2653-0.4855	0.4227 (0.0330)	0.3613- 0.4962
5279 (**)	0.4723 (0.0328)	0.3774 - 0.5765	0.4967 (0.0337)	0.4274 - 0.5699
80	0.66 (0.06)	0.44 - 0.86	0.65 (0.06)	0.49 – 0.78
60	0.66 (0.07)	0.37 – 0.90	0.65 (0.06)	0.50 – 0.77
40	0.66 (0.09)	0.28 – 0.94	0.65 (0.08)	0.46 – 0.85
20	0.66 (0.12)	0.20 - 1	0.64 (0.11)	0.40 – 0.90

LEGEND: CL = cell line; (*) set of SNPs represented on the 50K Xba chip; (**) set of filtered SNPs, used for the selection of the best SNPs.

COMPARING GENOTYPES

QUESTIONS:

1. How close (far) two samples need to be to be called 'similar' ('different')? How *confident* we are?

2. What is the *minimum number of loci* we need to make a decision?

PROBABILISTIC APPROACH

To evaluate the number of mismatches (matches) and to compare with ***expectations (gold standard)***

PROBABILISTIC APPROACH

UNDER THE ASSUMPTION THAT

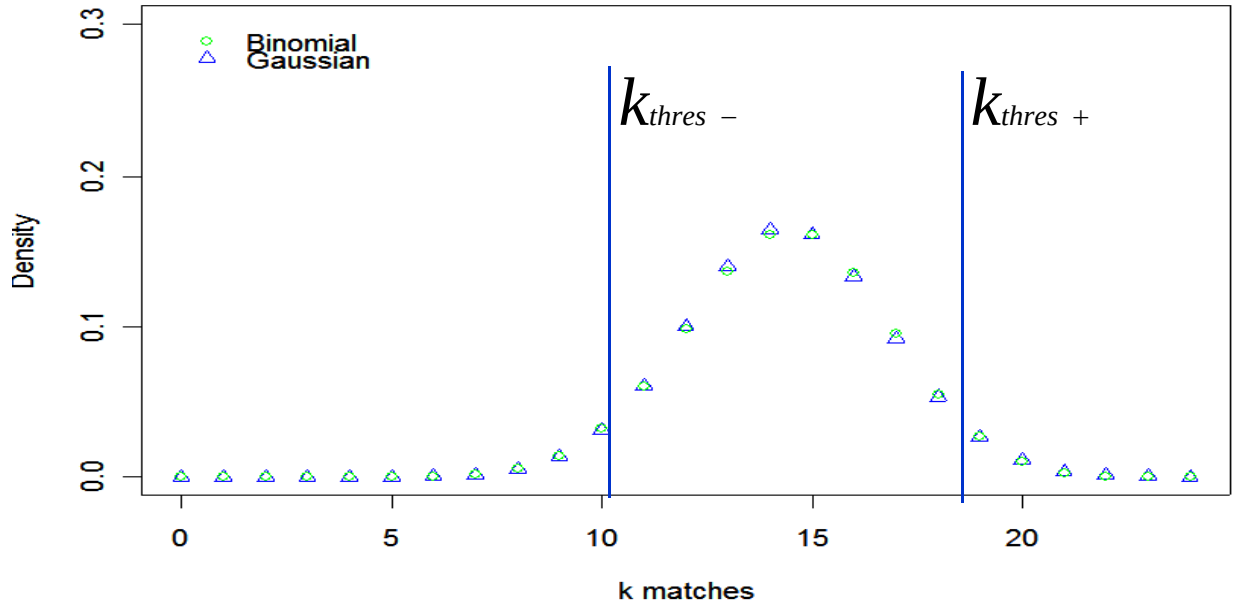
SNP calls are independent, e.g. call at locus *i* does not depend on call at locus *j, for each j≠i*

The probability of having *k* matches (successes) out of N SNPs (trials) follows the binomial distribution.

$$p_k = \binom{N}{k} p^k q^{N-k} = \frac{N!}{k!(N-k)!} p^k q^{N-k} \qquad \sum_{k=0}^{N} p_k = 1$$

where $p = P\_pm$, $q = P\_mm$ $\qquad (q + p = 1)$

Gaussian approximation of Binomial distribution



Given p and N:

$$k_{mean} = Np$$

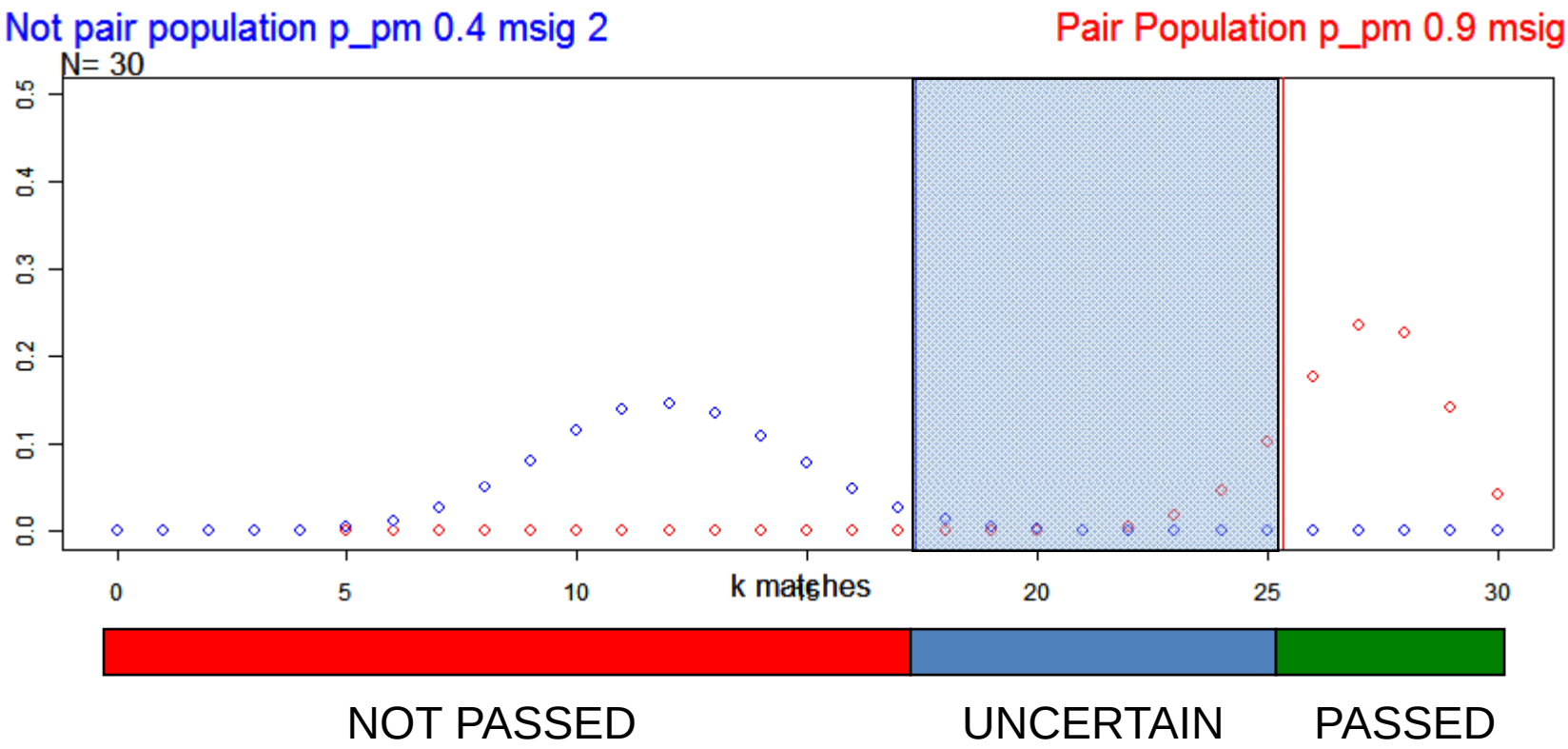
$$sd_{kmean} = sqrt(Np(1-p))$$

Thresholds can be set as:
(m *confidence*)

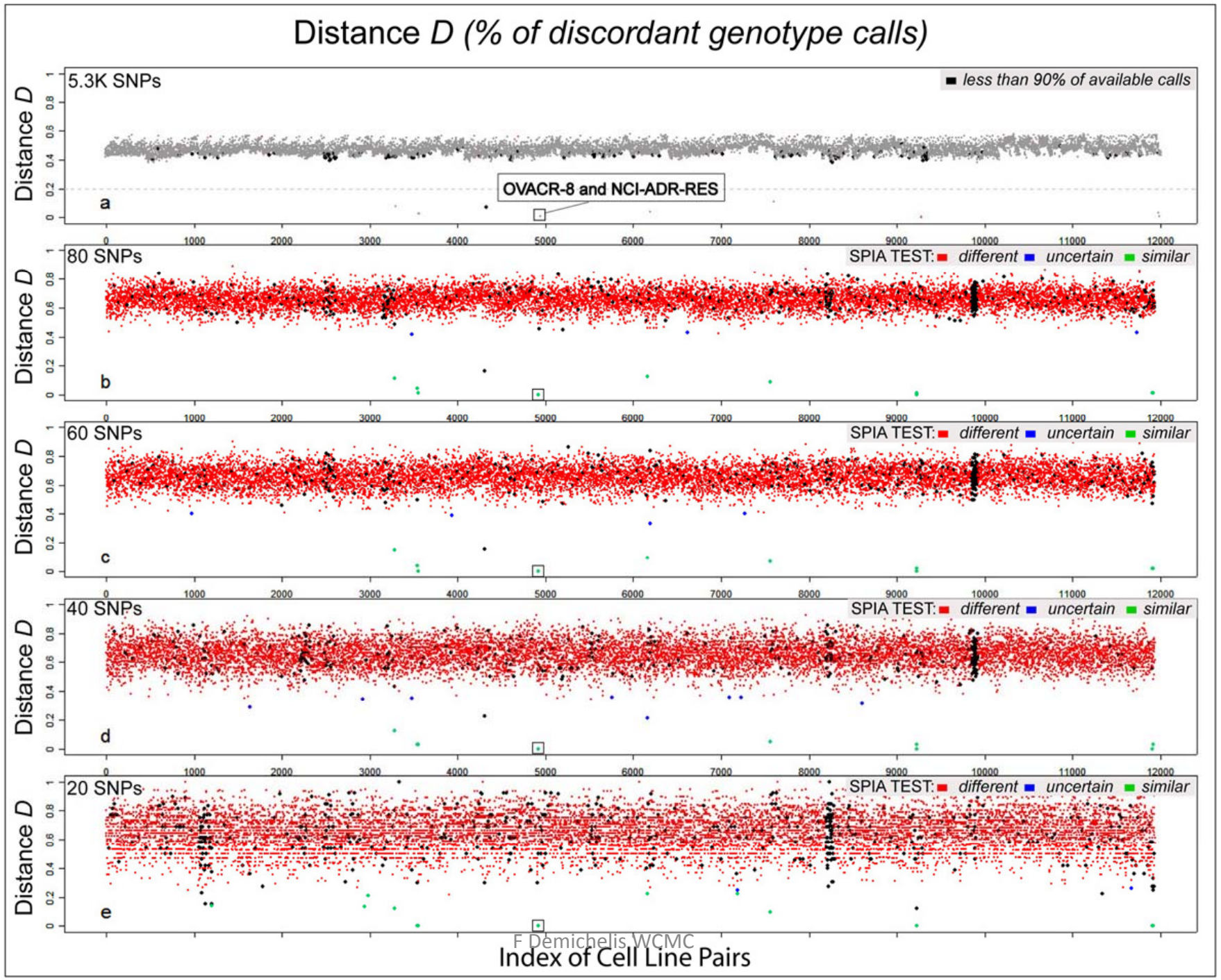
$$k_{thres} = k_{mean} +/- \mathrm{m}\ sd_{kmean}$$

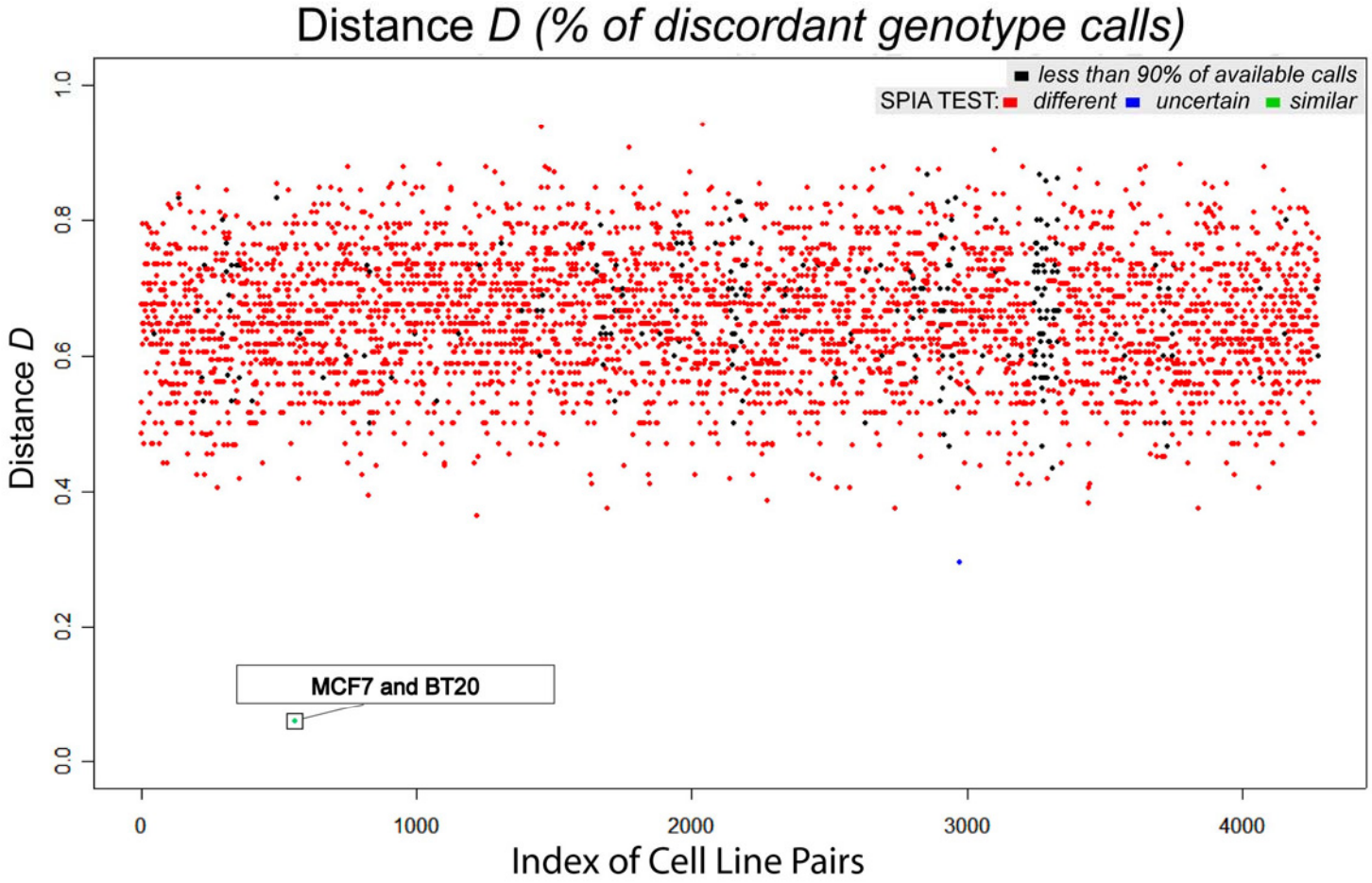
PROBABILISTIC APPROACH
DOUBLE TEST

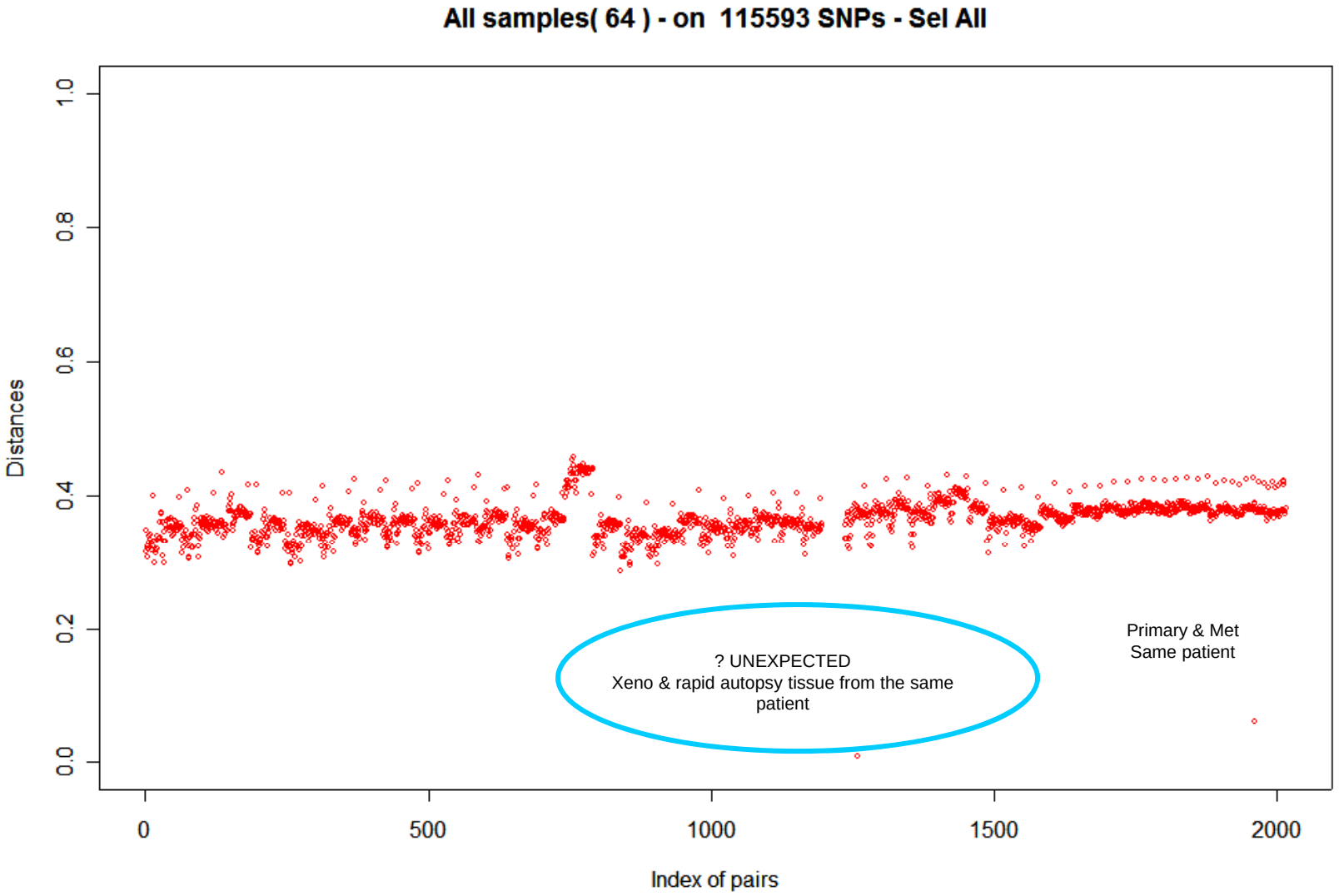
GOLD STANDARD POPULATIONS



Varying the m_{PAIR} *and* $m_{NON\text{-}PAIR}$ we set how CONSERVATIVE the test is.


Distance D (% of discordant genotype calls)
5.3K SNPs
less than 90% of available calls
OVACR-8 and NCI-ADR-RES
a
80 SNPs
SPIA TEST: different uncertain similar
b
60 SNPs
SPIA TEST: different uncertain similar
c
40 SNPs
SPIA TEST: different uncertain similar
d
20 SNPs
SPIA TEST: different uncertain similar
e
Distance D
Index of Cell Line Pairs
F Demichelis, WCMC


Distance D (% of discordant genotype calls)
less than 90% of available calls
SPIA TEST: different uncertain similar
MCF7 and BT20
Distance D
Index of Cell Line Pairs
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000


All samples(64) - on 115593 SNPs - Sel All
Distances
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
Index of pairs
? UNEXPECTED
Xeno & rapid autopsy tissue from the same patient
Primary & Met
Same patient

PROSTATE CELL LINES (from Jill Macoska)

At DIFFERENT PASSAGES

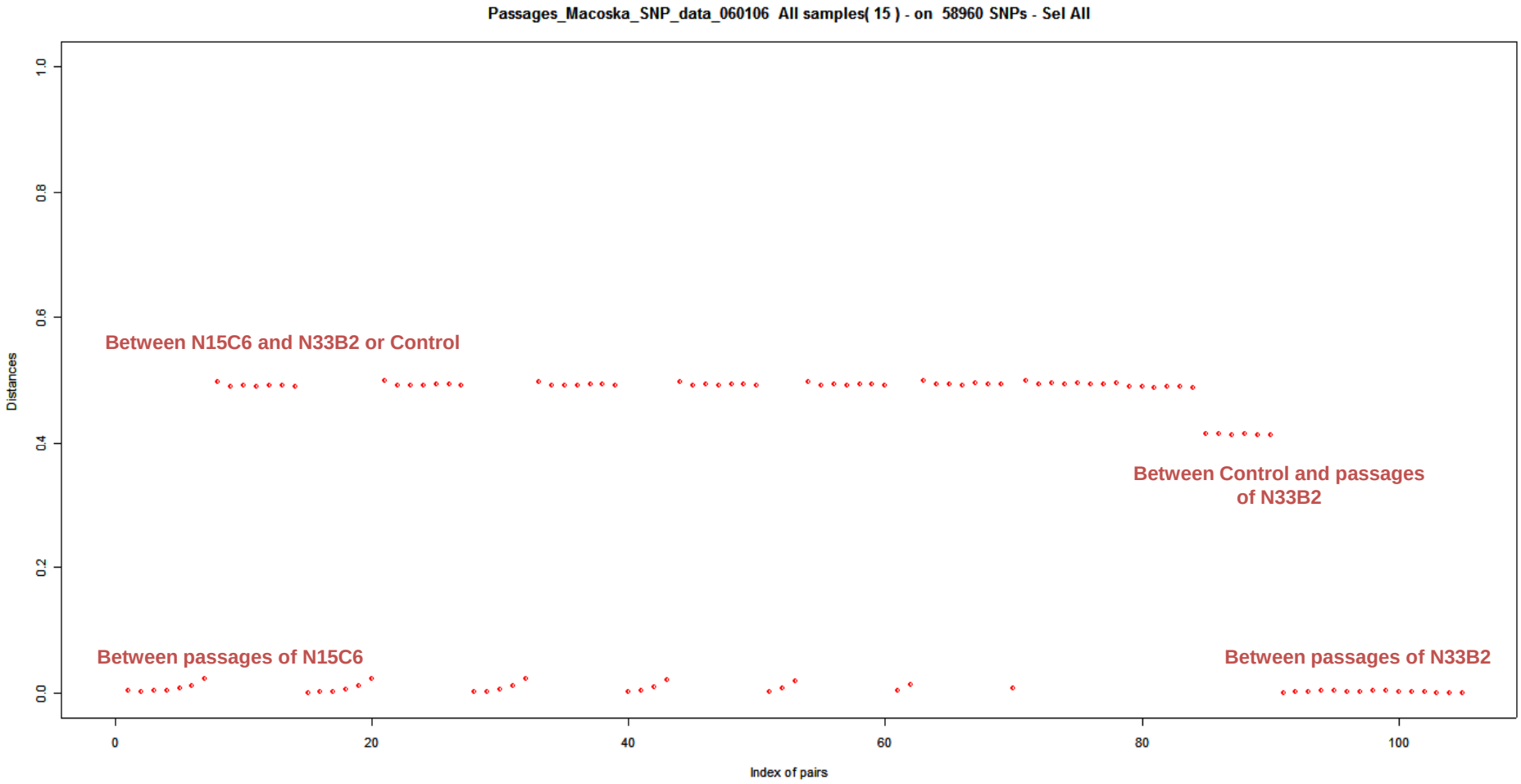
	Name	#_Passage	#_Tubes	Remaining Aliquote in lab	50KXba
1	N15C6	48	5	2.3	60106
3	N15C6	50	7	17	60106
5	N15C6	52	3	2.5	60106
7	N15C6	54	26	1.7	60106
9	N15C6	56	9	1.25	60106
11	N15C6	58	21	7.7	60106
13	N15C6	60	19	1	60106
15	N15C6	63	24	0.9	60106
1	N33B2	21	2	2.14	60106
4	N33B2	27	4	2.2	60106
8	N33B2	33	1	1.25	60106
10	N33B2	35	11	2	60106
12	N33B2	37	20	19	60106
13	N33B2	39	14	1	60106

SPIA – Allelotype distance

PROSTATE CELL LINES (from Jill Macoska)

At DIFFERENT PASSAGES

50K chip – 58960 SNPs

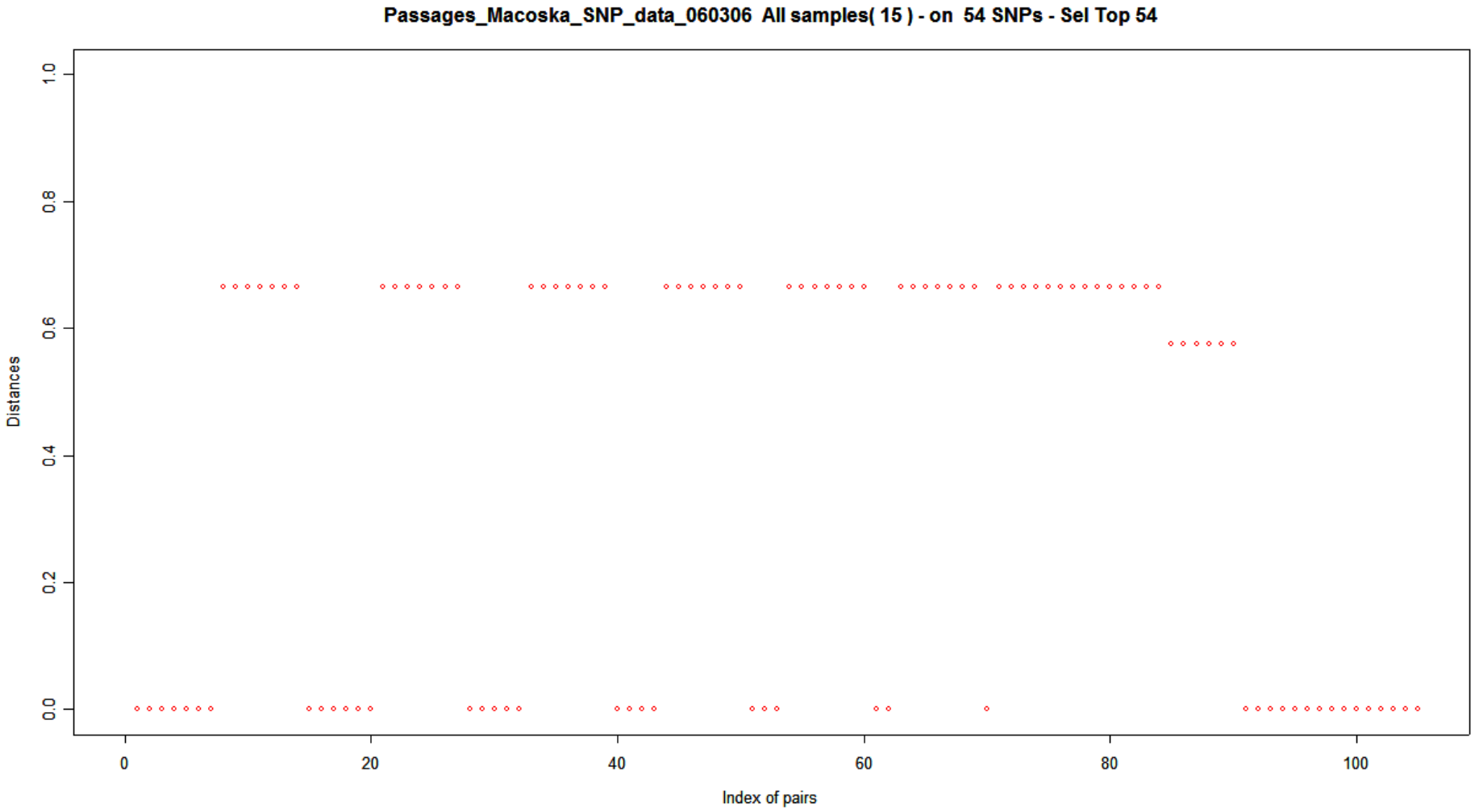


SPIA – Allelotype distance

PROSTATE CELL LINES (from Jill Macoska)

At DIFFERENT PASSAGES

SPIA top 54 selected SNPs

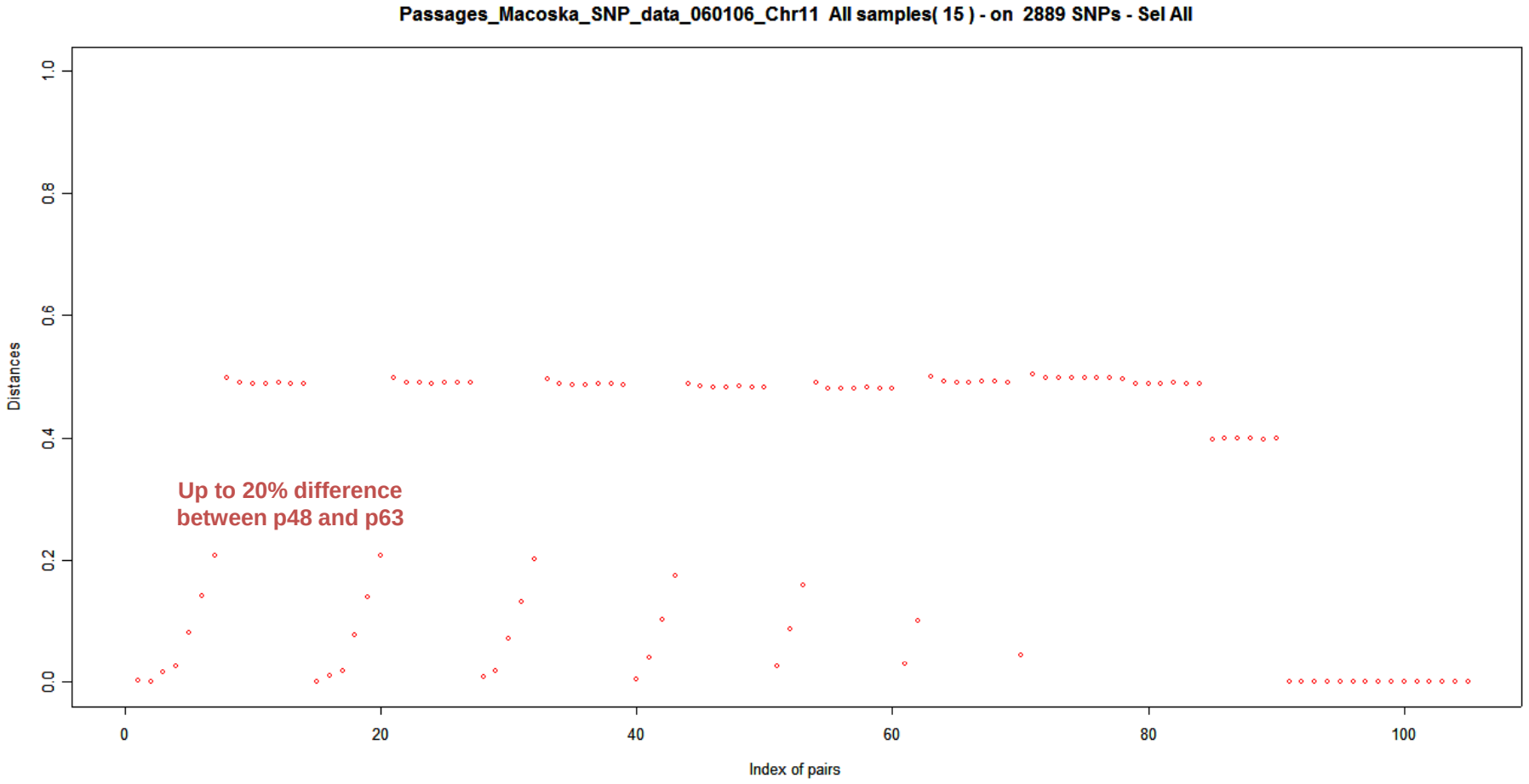


SPIA – Allelotype distance

PROSTATE CELL LINES (from Jill Macoska)

At DIFFERENT PASSAGES

Chromosome 11 -2889 SNPs



Our interest is in studying genome polymorphisms with respect **to cancer susceptibility and characterization**, by applying quantitative methods to genome-wide data.

http://icb.med.cornell.edu/faculty/demichelis/lab/