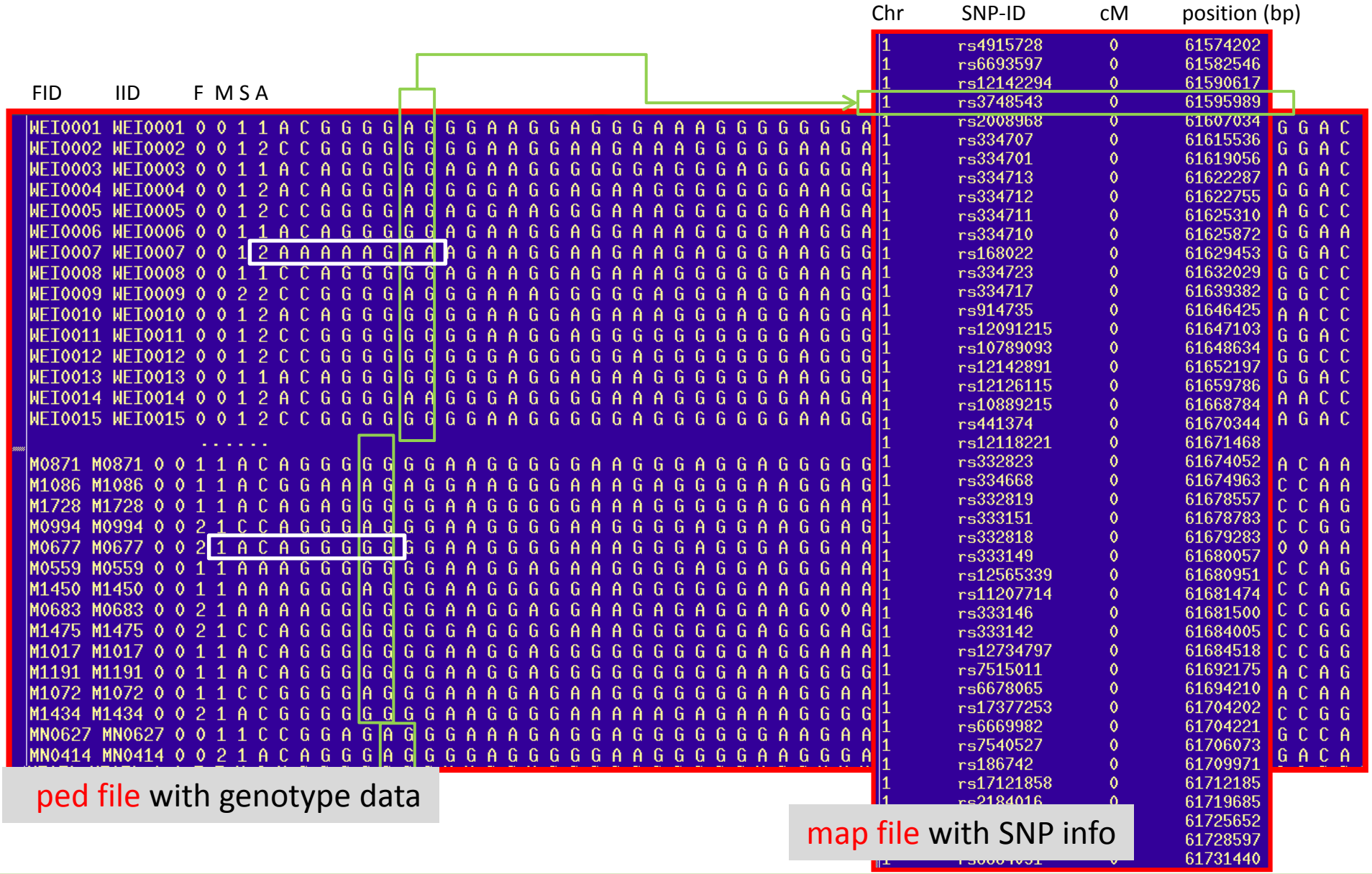
GWAS: population stratification using IBS

Robert Yu

GWAS data



ped file with genotype data

map file with SNP info

GWAS data

Population Diversity

ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	Genotype Detail A	Genotype Detail A/A	Genotype Detail A/G	Genotype Detail G/G	Genotype Detail HWP	Alleles A	Alleles G
ss118732970	YRI		2	IG		1.000				1.000	
ss137962797	ENSEMBL_Watson		2	IG		1.000				1.000	
ss138926948	ENSEMBL_Venter		2	IG		1.000				1.000	
ss163434191	YRI	Sub-Saharan African	2	IG		1.000				1.000	
ss164383830	CEU	European	2	IG		1.000				1.000	
ss166587487	PGP		2	IG		1.000				1.000	
ss198514210	BUSHMAN_POP2		1	IG	1.000					1.000	
	BANTU		1	IG	1.000					1.000	
ss218409404	pilot_1_YRI_low_coverage_panel		118	AF						0.907	0.093
ss230552416	pilot_1_CEU_low_coverage_panel		120	AF						0.950	0.050
ss238243769	pilot_1_CHB+JPT_low_coverage_panel		120	AF						0.933	0.067
ss43879764	HapMap-CEU	European	226	IG		0.903	0.097		1.000	0.951	0.049
	HapMap-HCB	Asian	86	IG		0.884	0.093	0.023	0.100	0.930	0.070
	HapMap-JPT	Asian	172	IG		0.919	0.081		1.000	0.959	0.041
	HapMap-YRI	Sub-Saharan African	226	IG		0.814	0.186		0.655	0.907	0.093
	HAPMAP-ASW		98	IG		0.816	0.184		1.000	0.908	0.092
	HAPMAP-CHB	Asian	82	IG		0.829	0.171		1.000	0.915	0.085
	HAPMAP-CHD		170	IG		0.953	0.047		1.000	0.976	0.024
	HAPMAP-GIH		176	IG		0.818	0.182		0.752	0.909	0.091
	HAPMAP-LWK		180	IG		0.744	0.222	0.033	0.343	0.856	0.144
	HAPMAP-MEX		100	IG		0.900	0.100		1.000	0.950	0.050
	HAPMAP-MKK		284	IG		0.620	0.310	0.070	0.200	0.775	0.225
	HAPMAP-TSI		176	IG		0.920	0.080		1.000	0.960	0.040
ss66403709	HapMap-CEU	European	118	IG		0.915	0.085		1.000	0.958	0.042
	HapMap-HCB	Asian	90	IG		0.889	0.089	0.022	0.100	0.933	0.067
	HapMap-JPT	Asian	90	IG		0.889	0.111		1.000	0.944	0.056
	HapMap-YRI	Sub-Saharan African	120	IG		0.833	0.167		1.000	0.917	0.083
ss76155077	ICMHP		6	IG		1.000				1.000	
ss97941075	J. Craig Venter		2	IG		1.000				1.000	

```
1    rs4915728     0    61574202
1    rs6693597     0    61582546
1    rs12142294    0    61590617
1    rs3748543     0    61595989
1    rs2008968     0    61607034
1    rs334707      0    61615536
1    rs334701      0    61619056
1    rs334713      0    61622287
1    rs334712      0    61622755
```

NCBI Resources How To

dbSNP SNP rs3748543

Save search Limits Advanced

Display Settings:

rs3748543 *[Homo sapiens]*

1.

AAAGAGTAAGATGACATCTGTCTGTT[A/G]TACATTTGGTATCTCTGATGCCGTA

MapView No VarVu No PubMed GeneView SeqView No 3D No OMIM

MAF/MinorAlleleCount: C=0.0586/128

HGVS Names: [NC_000001.10:g.61595989C>T] [NG_011787.1:g.58044C>T] [NM_001134673.3:c.559+41637C>T] [NM_

```
1    rs332823      0    61674052
1    rs334668      0    61674963
1    rs332819      0    61678557
1    rs333151      0    61678783
1    rs332818      0    61679283
1    rs333149      0    61680057
1    rs12565339    0    61680951
1    rs11207714    0    61681474
1    rs333146      0    61681500
1    rs333142      0    61684005
1    rs12734797    0    61684518
1    rs7515011     0    61692175
1    rs6678065     0    61694210
1    rs17377253    0    61704202
1    rs6669982     0    61704221
1    rs7540527     0    61706073
1    rs186742      0    61709971
1    rs17121858    0    61712185
1    rs2184016     0    61719685
                       61725652
                       61728597
                       61731440
```

map file with SNP info

GWAS data

label	population sample	number of samples
ASW	African ancestry in Southwest USA	90
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection	180
CHB	Han Chinese in Beijing, China	90
CHD	Chinese in Metropolitan Denver, Colorado	100
GIH	Gujarati Indians in Houston, Texas	100
JPT	Japanese in Tokyo, Japan	91
LWK	Luhya in Webuye, Kenya	100
MEX	Mexican ancestry in Los Angeles, California	90
MKK	Maasai in Kinyawa, Kenya	180
TSI	Toscans in Italy	100
YRI	Yoruba in Ibadan, Nigeria	180

GWAS data

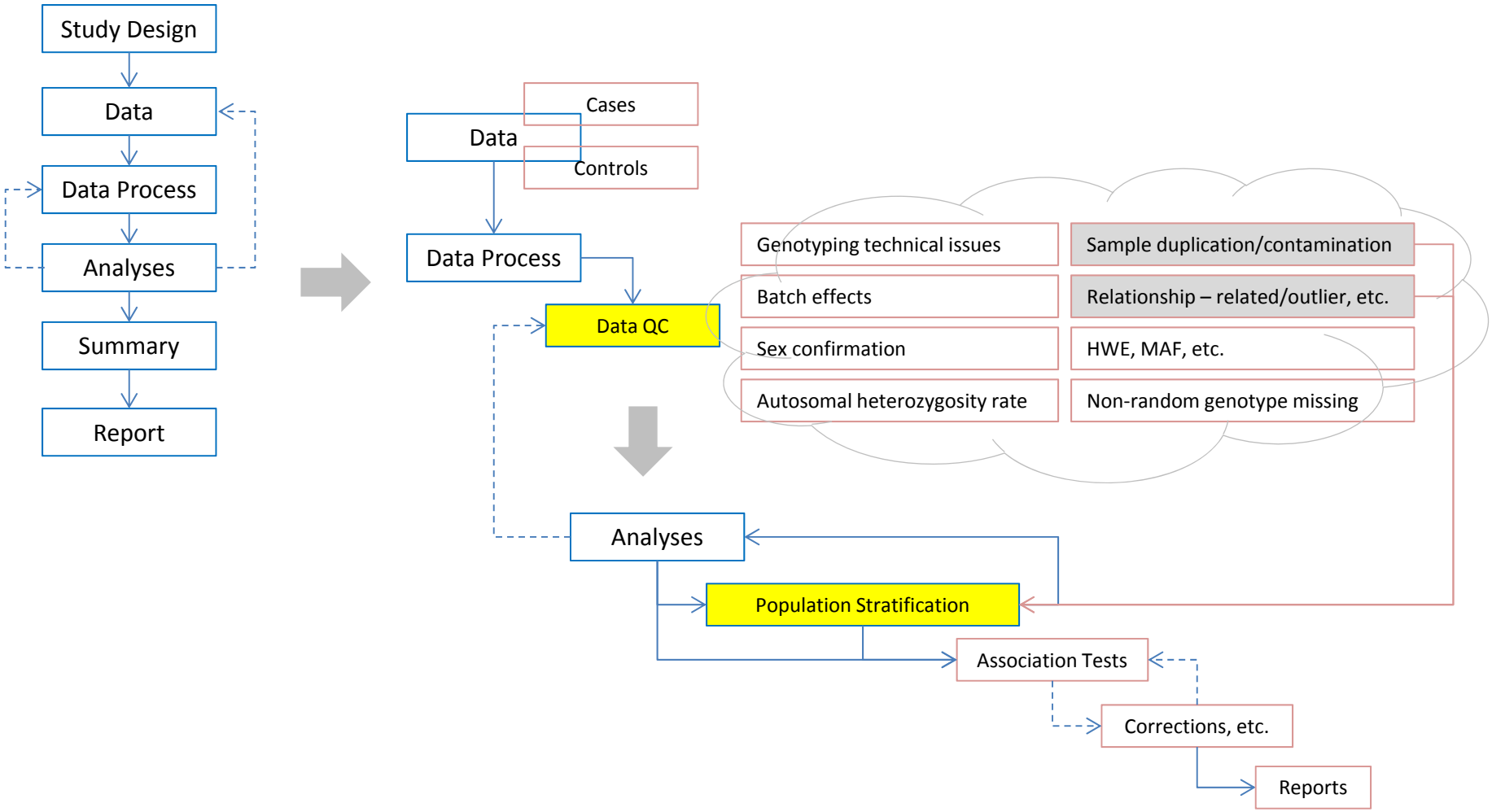
Population Diversity

ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	Genotype Detail C/C	C/T	T/T	HWP	Alleles C	T
ss118733406	HapMap-CEU	European	226	IG	0.009	0.159	0.832	1.000	0.088	0.912
	HapMap-HCB	Asian	86	IG			1.000			1.000
	HapMap-JPT	Asian	88	IG			1.000			1.000
	HapMap-YRI	Sub-Saharan African	226	IG	0.062	0.301	0.637	0.294	0.212	0.788
	HAPMAP-ASW		98	IG	0.082	0.286	0.633	0.251	0.224	0.776
	HAPMAP-CHB	Asian	82	IG		0.024	0.976	1.000	0.012	0.988
	HAPMAP-CHD		170	IG		0.035	0.965	1.000	0.018	0.982
	HAPMAP-GIH		176	IG		0.080	0.920	1.000	0.040	0.960
	HAPMAP-LWK		180	IG	0.122	0.344	0.533	0.150	0.294	0.706
	HAPMAP-MEX		100	IG	0.020	0.300	0.680	0.752	0.170	0.830
	HAPMAP-MKK		286	IG	0.014	0.175	0.811	0.655	0.101	0.899
	HAPMAP-TSI		176	IG		0.068	0.932	1.000	0.034	0.966
	ENSEMBL_Watson		2	IG			1.000			1.000
	YRI		2	IG	1.000				0.500	0.500
	ENSEMBL_Venter		2	IG			1.000			1.000
ss198515012	BANTU		2	IG	1.000				0.500	0.500
ss218409640	pilot_1_YRI_low_coverage_panel		118	AF					0.203	0.797
ss230552659	pilot_1_CEU_low_coverage_panel		120	AF					0.100	0.900

1	rs4915728	0	61574202
1	rs6693597	0	61582546
1	rs12142294	0	61590617
1	rs3748543	0	61595989
1	rs2008968	0	61607034
1	rs334707	0	61615536
1	rs334701	0	61619056
1	rs334713	0	61622287
1	rs334712	0	61622755
1	rs334711	0	61625310
1	rs334710	0	61625872
1	rs168022	0	61629453
1	rs334723	0	61632029
1	rs334717	0	61639382
1	rs914735	0	61646425
1	rs12091215	0	61647103
1	rs10789093	0	61648634
1	rs12142891	0	61652197
1	rs12126115	0	61659786
1	rs10889215	0	61668784
1	rs441374	0	61670344
1	rs12118221	0	61671468
1	rs332823	0	61674052
1	rs334668	0	61674963
1	rs332819	0	61678557
1	rs333151	0	61678783
1	rs332818	0	61679283
1	rs333149	0	61680057
1	rs12565339	0	61680951
1	rs11207714	0	61681474
1	rs333146	0	61681500
1	rs333142	0	61684005
1	rs12734797	0	61684518
1	rs7515011	0	61692175
1	rs6678065	0	61694210
1	rs17377253	0	61704202
1	rs6669982	0	61704221
1	rs7540527	0	61706073
1	rs186742	0	61709971
1	rs17121858	0	61712185
1	rs2184016	0	61719685
			61725652
			61728597
			61731440

map file with SNP info

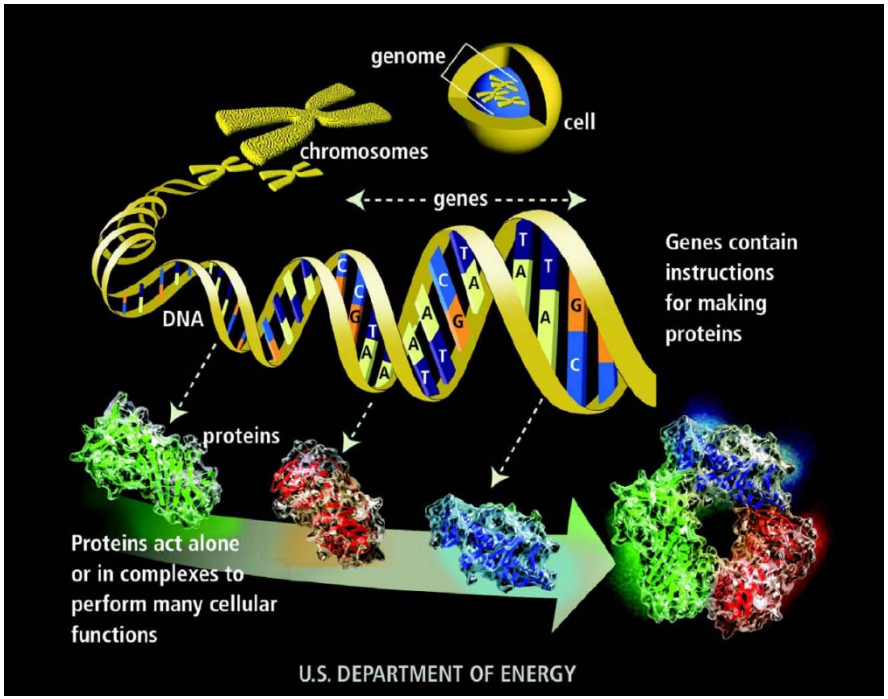
general workflow in GWAS

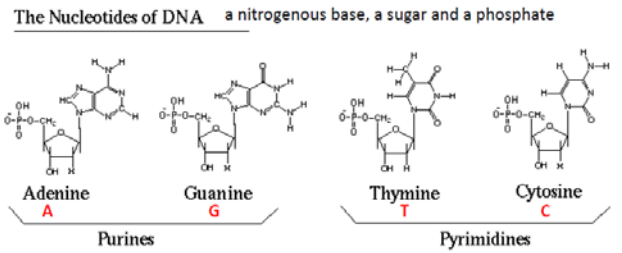


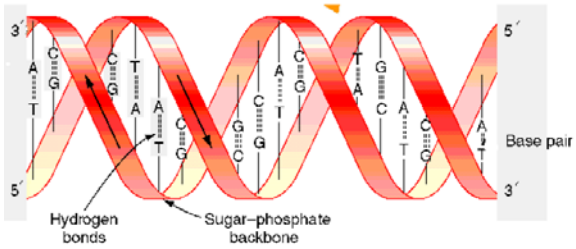
population-based GWAS

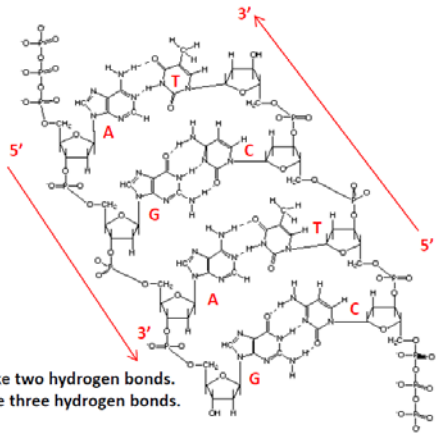
- Population-based GWAS will yield spurious association test results if population confounding factors are not eliminated.
- Allelic frequency of a locus in genome could be significantly different among individuals representing distant different populations.
- Stratification of population structure within the data (e.g. between cases and controls or within cases/controls) is crucial.
- Detection and removal of relatedness or outlier in the sample are another vital step.
- Using IBS to estimate IBD from dense SNP data set can achieve the above goal.
- What are IBS and IBD?

a review of molecular biology

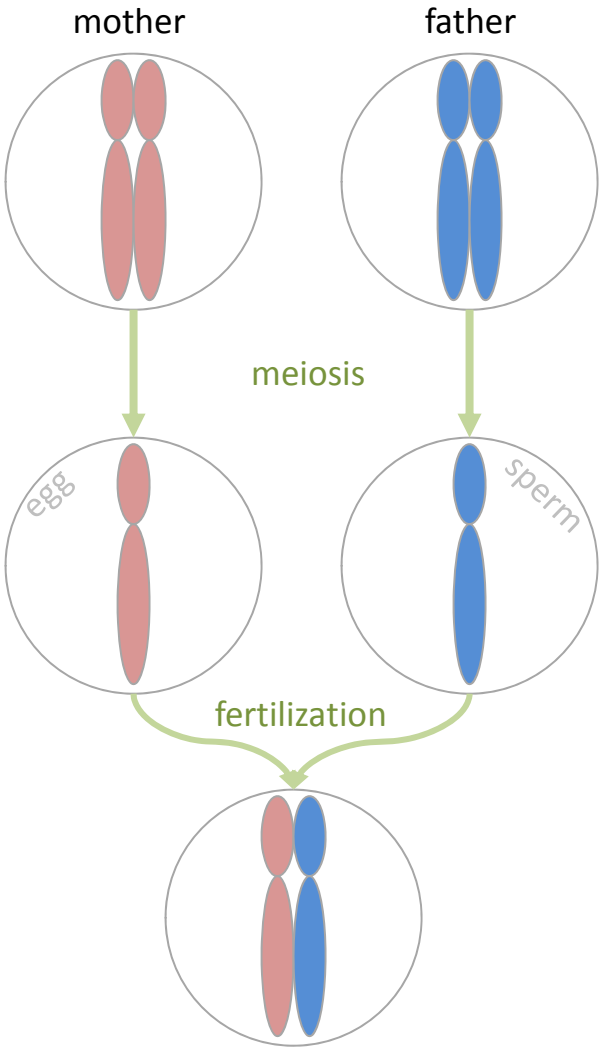








Gamete – a haploid cell during meiosis

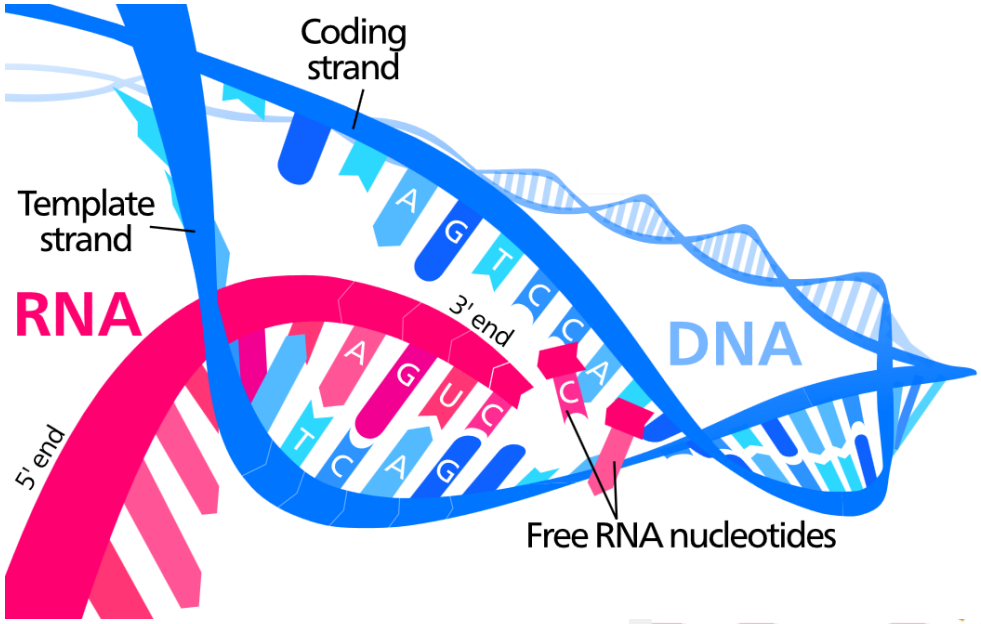


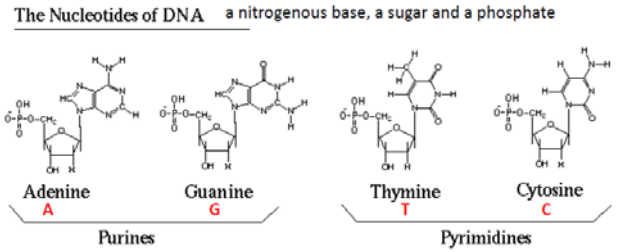
diploid cells

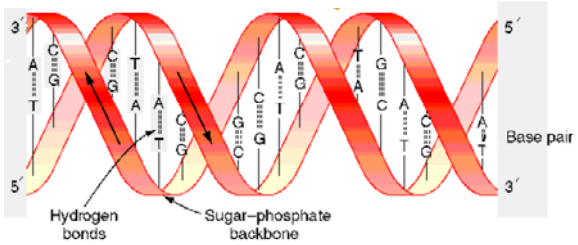
haploid gametes

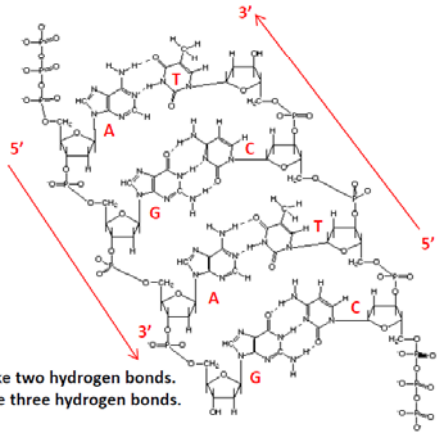
a diploid zygote
(child)

a review of molecular biology









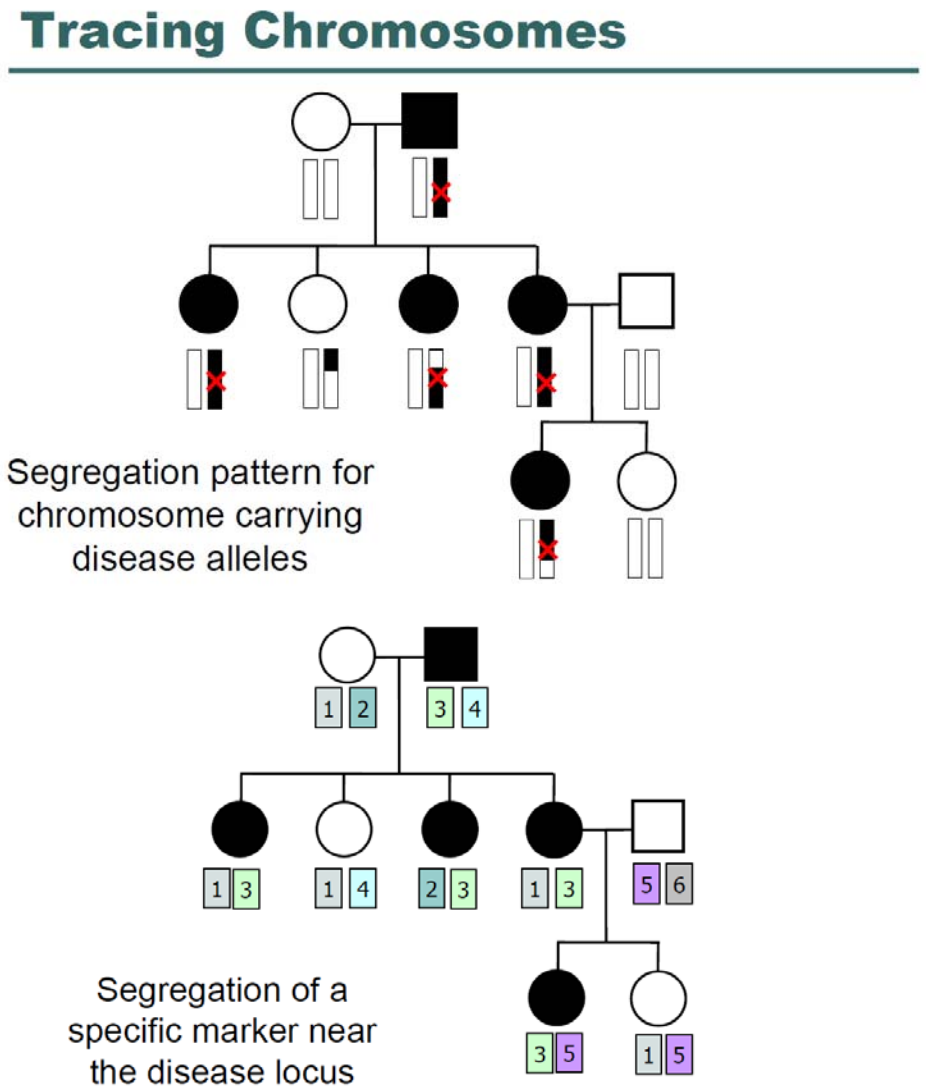
Adenine pairs with **Thymine** because they make **two hydrogen bonds.**
Guanine pairs with **Cytosine** because they make **three hydrogen bonds.**

As in DNA replication, DNA is read **from 3'UTR → 5'UTR during transcription**. Meanwhile, the complementary RNA is created from the 5'UTR → 3'UTR direction. Although DNA is arranged as two antiparallel strands in a double helix, only one of the two DNA strands, called the **template strand**, is used for transcription. This is because RNA is only single-stranded, as opposed to double-stranded DNA. The other DNA strand is called the **coding (lagging) strand**, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine).

Reference

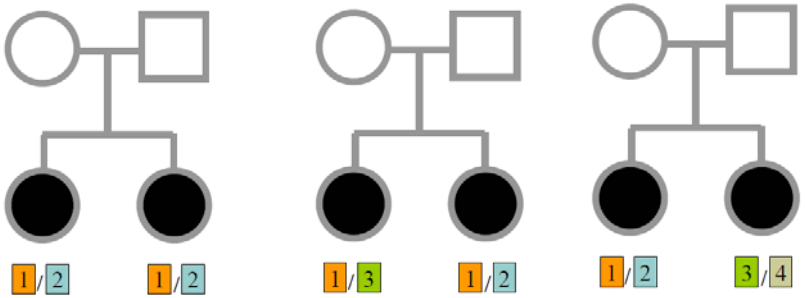
"Transcription (genetics)" - http://en.wikipedia.org/wiki/Transcription_(genetics)

IBS Methods in linkage analysis



Reference Gonçalo Abecasis's Lecture Notes, Biostat 666, "IBS Methods for Affected Pairs Linkage"

IBS and IBD



IBS = 2
IBD = 2

IBS = 1
IBD = 1

IBS = 0
IBD = 0

IBS – Identity By State

- At a locus, two individuals have the same allele(s).

IBD – Identity By Descent

- At a locus, two individuals have the same allele(s), and the allele(s) was "copied" from the same parents/ancestry.

Distinction of IBD and IBS

- Alleles that have identical nucleotide sequences but have descended from dierent ancestors in the reference population are IBS but not IBD.
- Alleles that are IBD are necessarily IBS provided there is no mutation of the inherited allele.

Reference Gonçalo Abecasis's Lecture Notes, Biostat 666, "IBS Methods for Affected Pairs Linkage"

IBS Methods in linkage analysis

Test for Independence

$$\chi^2_{2df} = \sum_i \frac{[N_{IBS=i} - E(N_{IBS=i})]^2}{E(N_{IBS=i})}$$ (general test, for sibling pairs)

$$\chi^2_{1df} = \frac{[N_{IBS=0} - E(N_{IBS=0})]^2}{E(N_{IBS=0})} + \frac{[N_{IBS>0} - E(N_{IBS>0})]^2}{E(N_{IBS>0})}$$ (grouping often preferable for other relatives)

- Assuming all counts are relatively large
- If counts are small, use binomial or trinomial distribution

Modeling IBS Sharing

- For any relative pair, calculate:
  - Probability of IBD sharing
    - 0, 1 or 2 alleles
  - Conditional probability of IBS sharing
    - 0, 1, 2 alleles
  - IBS sharing >= IBD sharing
    - Why?

Reference Gonçalo Abecasis's Lecture Notes, Biostat 666, "IBS Methods for Affected Pairs Linkage"

IBS Methods in linkage analysis

IBD

- The underlying sharing of chromosomes segregating within a family
- Siblings share 0, 1 or 2 alleles
  - Probabilities ¼, ½ and ¼
- Unilineal relatives share 0 or 1 alleles
  - Probability of sharing is kinship coefficient ϕ * 4

P(Marker Genotype|IBD State)

Relative		IBD		
I	II	0	1	2
(a,b)	(c,d)	$4p_a p_b p_c p_d$	0	0
(a,a)	(b,c)	$2p_a^2 p_b p_c$	0	0
(a,a)	(b,b)	$p_a^2 p_b^2$	0	0
(a,b)	(a,c)	$4p_a^2 p_b p_c$	$p_a p_b p_c$	0
(a,a)	(a,b)	$2p_a^3 p_b$	$p_a^2 p_b$	0
(a,b)	(a,b)	$4p_a^2 p_b^2$	$(p_a p_b^2 + p_a^2 p_b)$	$2p_a p_b$
(a,a)	(a,a)	p_a^4	p_a^3	p_a^2
Prior Probability		¼	½	¼

Glossary:

Unilineal descent
Descent links are traced only through ancestors of one gender.

Kinship
Culturally defined relationships between individuals, usually based on marriage, descent, etc.

Kinship coefficient θ
a measurement of relatedness between two individuals. It's useful predictors of covariance and correlation between relatives.
The probability that 2 alleles are IBD is defined to be **coefficient of coancestry** or **kinship coefficient** and is often represented as θ.

In non-inbred pedigrees, kinship coefficients can be derived from IBD probabilities:

$$\theta = \frac{1}{4}P(IBD = 1) + \frac{1}{2}P(IBD = 2)$$

Reference Gonçalo Abecasis's Lecture Notes, Biostat 666, "IBS Methods for Affected Pairs Linkage"

IBS Methods in linkage analysis

P(IBS = i | IBD = j)

$$P(IBS=2 \mid IBD=0) = 2\sum_{i\neq j} p_i^2 p_j^2 + \sum_i p_i^4$$

$$P(IBS=1 \mid IBD=0) = 4\sum_{i\neq j} p_i^2 p_j (1-p_i-p_j) + 4\sum_i p_i^3 (1-p_i)$$

$$P(IBS=0 \mid IBD=0) = \sum_{i\neq j} p_i p_j (1-p_i-p_j)^2 + \sum_i p_i^2 (1-p_i)^2$$

$$P(IBS=2 \mid IBD=0) = 2\sum_{i\neq j} p_i^2 p_j^2 + \sum_i p_i^4$$

$$P(IBS=1 \mid IBD=0) = 4\sum_{i\neq j} p_i^2 p_j (1-p_i-p_j) + 4\sum_i p_i^3 (1-p_i)$$

$$P(IBS=0 \mid IBD=0) = \sum_{i\neq j} p_i p_j (1-p_i-p_j)^2 + \sum_i p_i^2 (1-p_i)^2$$

$$P(IBS=2 \mid IBD=2) = 1$$
$$P(IBS=1 \mid IBD=2) = 0$$
$$P(IBS=0 \mid IBD=2) = 0$$

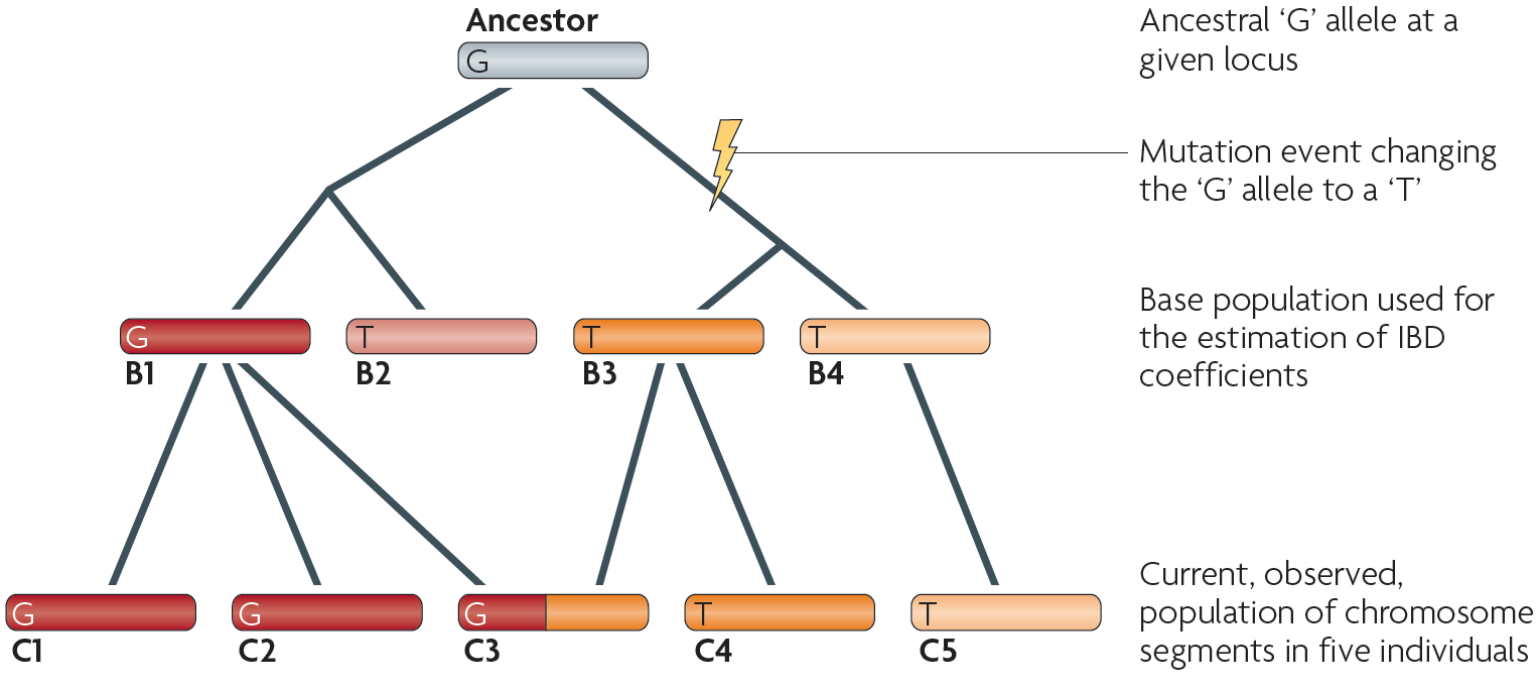
$$P(IBS=2 \mid IBD=1) = \sum_i p_i^2$$

$$P(IBS=1 \mid IBD=1) = \sum_{i\neq j} p_i p_j$$

$$P(IBS=0 \mid IBD=1) = 0$$

Reference Gonçalo Abecasis's Lecture Notes, Biostat 666, “IBS Methods for Affected Pairs Linkage”

IBD, IBS and coalescence



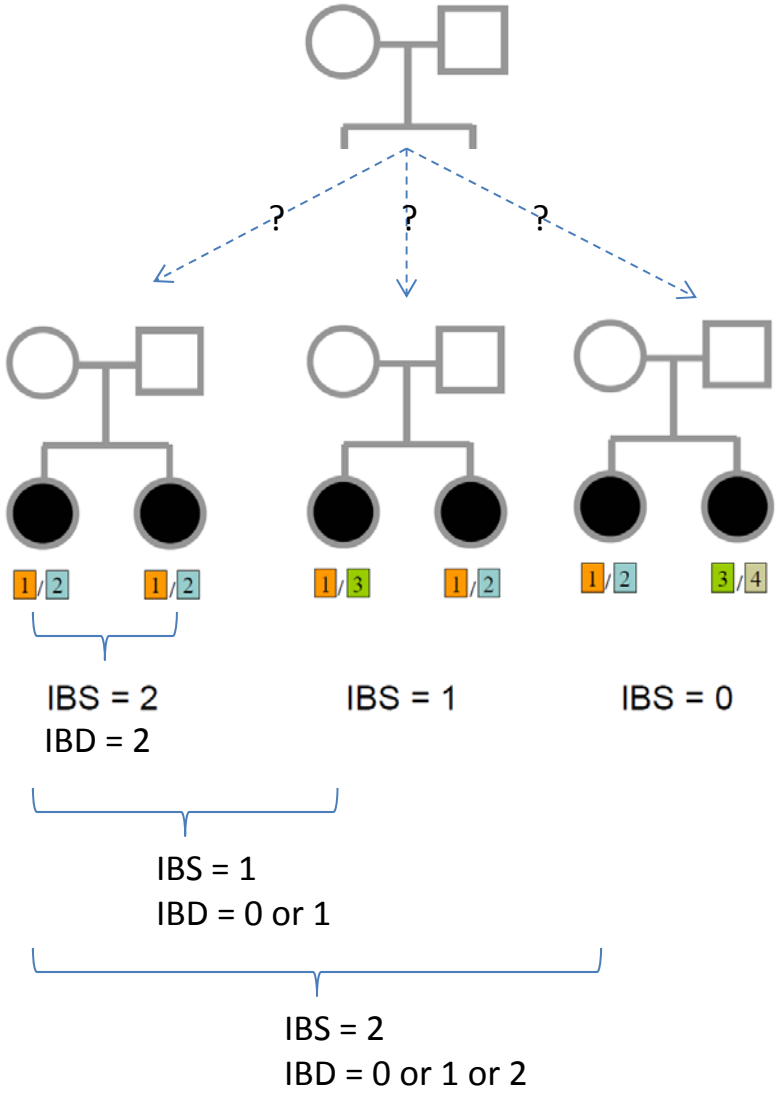
The figure depicts an ancestral allele at a locus, representing the point of coalescence for alleles in the current population (C1–C5). At the point of coalescence (the most recent common ancestor) this locus carries a copy of a G allele that is subject to a mutation event (G→T; lightning symbol) leading to a G/T polymorphism.

IBD at the polymorphic locus among individuals (C1–C5) can be defined with respect to a base population (B1–B4) in which individuals are assumed to be unrelated (shown by the differently coloured chromosome segments). Then the G alleles in C1, C2 and C3 are IBD to each other as all three descend from the G allele in B1. The T alleles in C4 and C5 are IBS but not IBD as they descend from different alleles in the base population.

The whole chromosome segments C1 and C2 are IBD because they descend from a common ancestor (B1) without recombination, but chromosome segment C3 is not IBD to C1 and C2.

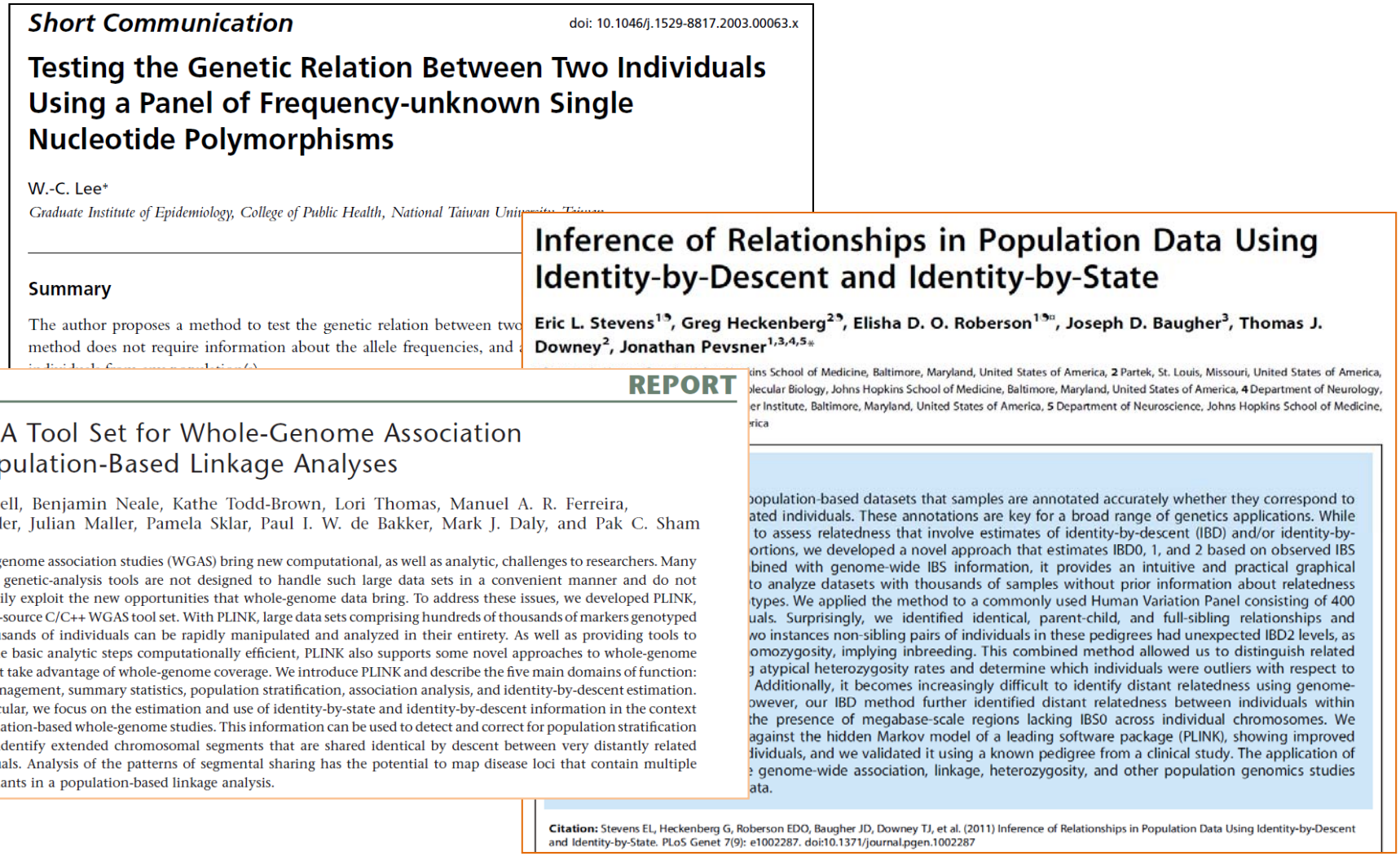
Reference Powell, JE, Visscher, PM, and Goddard, ME, Nature Reviews | GENETICS, vol 11, Nov. 2010, pp800-5

IBS Methods in GWAS



Reference The diagram was modified and based on Gonçalo Abecasis's Lecture Notes, Biostat 666

IBS Methods in GWAS



Short Communication

doi: 10.1046/j.1529-8817.2003.00063.x

Testing the Genetic Relation Between Two Individuals Using a Panel of Frequency-unknown Single Nucleotide Polymorphisms

W.-C. Lee*

Graduate Institute of Epidemiology, College of Public Health, National Taiwan Univ

Summary

The author proposes a method to test the genetic relation between two
method does not require information about the allele frequencies, and

Inference of Relationships in Population Data Using Identity-by-Descent and Identity-by-State

Eric L. Stevens[1¤], Greg Heckenberg[2¤], Elisha D. O. Roberson[1¤], Joseph D. Baugher[3], Thomas J. Downey[2], Jonathan Pevsner[1,3,4,5]*

ins School of Medicine, Baltimore, Maryland, United States of America, **2** Partek, St. Louis, Missouri, United States of America, olecular Biology, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, **4** Department of Neurology, er Institute, Baltimore, Maryland, United States of America, **5** Department of Neuroscience, Johns Hopkins School of Medicine, rica

opulation-based datasets that samples are annotated accurately whether they correspond to ated individuals. These annotations are key for a broad range of genetics applications. While to assess relatedness that involve estimates of identity-by-descent (IBD) and/or identity-by- ortions, we developed a novel approach that estimates IBD0, 1, and 2 based on observed IBS bined with genome-wide IBS information, it provides an intuitive and practical graphical to analyze datasets with thousands of samples without prior information about relatedness types. We applied the method to a commonly used Human Variation Panel consisting of 400 uals. Surprisingly, we identified identical, parent-child, and full-sibling relationships and wo instances non-sibling pairs of individuals in these pedigrees had unexpected IBD2 levels, as omozygosity, implying inbreeding. This combined method allowed us to distinguish related g atypical heterozygosity rates and determine which individuals were outliers with respect to Additionally, it becomes increasingly difficult to identify distant relatedness using genome- owever, our IBD method further identified distant relatedness between individuals within the presence of megabase-scale regions lacking IBS0 across individual chromosomes. We against the hidden Markov model of a leading software package (PLINK), showing improved dividuals, and we validated it using a known pedigree from a clinical study. The application of e genome-wide association, linkage, heterozygosity, and other population genomics studies ata.

Citation: Stevens EL, Heckenberg G, Roberson EDO, Baugher JD, Downey TJ, et al. (2011) Inference of Relationships in Population Data Using Identity-by-Descent and Identity-by-State. PLoS Genet 7(9): e1002287. doi:10.1371/journal.pgen.1002287

REPORT

PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses

Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham

Whole-genome association studies (WGAS) bring new computational, as well as analytic, challenges to researchers. Many existing genetic-analysis tools are not designed to handle such large data sets in a convenient manner and do not necessarily exploit the new opportunities that whole-genome data bring. To address these issues, we developed PLINK, an open-source C/C++ WGAS tool set. With PLINK, large data sets comprising hundreds of thousands of markers genotyped for thousands of individuals can be rapidly manipulated and analyzed in their entirety. As well as providing tools to make the basic analytic steps computationally efficient, PLINK also supports some novel approaches to whole-genome data that take advantage of whole-genome coverage. We introduce PLINK and describe the five main domains of function: data management, summary statistics, population stratification, association analysis, and identity-by-descent estimation. In particular, we focus on the estimation and use of identity-by-state and identity-by-descent information in the context of population-based whole-genome studies. This information can be used to detect and correct for population stratification and to identify extended chromosomal segments that are shared identical by descent between very distantly related individuals. Analysis of the patterns of segmental sharing has the potential to map disease loci that contain multiple rare variants in a population-based linkage analysis.

IBS Methods in GWAS

Short Communication doi: 10.1046/j.1529-8817.2003.00063.x

Testing the Genetic Relation Between Two Individuals Using a Panel of Frequency-unknown Single Nucleotide Polymorphisms

W.-C. Lee*
Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taiwan

Summary

The author proposes a method to test the genetic relation between two individuals using a panel of SNPs. The method does not require information about the allele frequencies, and as such it can be used to test any pair of individuals from any population(s).

Testing 3 possibilities of relationship between 2 individuals being

1) from the same random-mating population and genetically unrelated (H_0)
2) genetically related (H_{a1})
3) from different random mating populations (H_{a2})

At a locus for a SNP, the 'discordant homozygotes' (Dh, e.g. AA vs BB) and the 'concordant heterozygotes' (Ch, AB vs AB), the conditional probabilities for concordance under H_0 are

$$\pi_i^{H_0} = \Pr(\text{Ch})/[\Pr(\text{Ch}) + \Pr(\text{Dh})] = 4p_i^2q_i^2/(4p_i^2q_i^2 + 2p_i^2q_i^2) = 2/3$$

The probabilities are equal for each and every locus and do not depend on allele frequency p_i (or $q_i = 1 - p_i$). Thus, the test statistic T_1 has

$$E^{H0}(T_1)=2/3,$$

where $T_1 = m^{-1}\sum_{i=1}^{m} X_i, i = 1,2,\ldots,m(m \le n), X_i = 1\ for\ Ch\ or\ 0\ for\ Dh.$

And, Pr(Ch) = 2 Pr(Dh), or IBS2* = 2 x IBS0.

$\mathrm{Var}^{H_0}(T_1) = m^{-2} \cdot \sum_{i=1}^{m} \mathrm{Var}^{H_0}(X_i) = 2/(9m)$. And the statistic, $Z_1 = (T_1 - 2/3)/\sqrt{2/(9m)}$, is asymptotically the standard normal distribution under H_0.

One constraint PLINK applies is called the "pairwise population concordance" (PPC) test, similar to a method used by Lee,[17] such that for any putative new cluster, all pairs of individuals pass this test. For a given pair, we expect to see autosomal SNPs with two copies of each allele occur in a 2:1 ratio of IBS 2 {Aa,Aa} to IBS 0 {AA,aa} SNP pairs if both members of the pair come from the same random-mating population. For SNPs selected far enough

Reference

Lee W (2003). Ann Hum Genet. Pp 618–619.

IBS Methods in GWAS

Short Communication

doi: 10.1046/j.1529-8817.2003.00063.x

Testing the Genetic Relation Between Two Individuals Using a Panel of Frequency-unknown Single Nucleotide Polymorphisms

W.-C. Lee*

Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taiwan

Summary

The author proposes a method to test the genetic relation between two individuals using a panel of SNPs. The method does not require information about the allele frequencies, and as such it can be used to test any pair of individuals from any population(s).

1. Under H_{a1}, the conditional concordance probabilities are $\pi_i^{H_{a1}} = (4k_0p_i^2q_i^2 + k_1p_iq_i + 2k_2p_iq_i)/(4k_0p_i^2q_i^2 + k_1p_iq_i + 2k_2p_iq_i + 2k_0p_i^2q_i^2) = [2k_0 + 2\psi \cdot (p_iq_i)^{-1}]/[3k_0 + 2\psi \cdot (p_iq_i)^{-1}] \geq 2/3$, where k_0, k_1, and k_2 denote, respectively, the probabilities of 0, 1, and 2 genes identical by decent (IBD) ($k_0 + k_1 + k_2 = 1$), and the $\psi = k_1/4 + k_2/2$ is the 'kinship coefficient' between the two individuals (Thompson, 1986). Since the probabilities are greater than or equal to the null value of 2/3 (equality holds when $k_0 = 1$ or equivalently $\psi = 0$), one can perform a one-sided test based on Z_1 for concordance excess to test H_0 against H_{a1}.

2. Under H_{a2}, we have $\pi_i^{H_{a2}} = 4p_ip_i'q_iq_i'/[4p_ip_i'q_iq_i' + p_i^2(q_i')^2 + (p_i')^2q_i^2] = 2/[3 + (p_i - p_i')^2 \cdot (2p_ip_i'q_iq_i')^{-1}] \leq 2/3$, where the p_i and the p_i' ($q_i = 1 - p_i$, $q_i' = 1 - p_i'$) represent the allele frequencies in the two source populations. This time, the probabilities are in the opposite direction from null (equality holds when $p_i = p_i'$). Thus, a one-sided test for concordance deficiency can be used to test H_0 against H_{a2}.

3. To relax the assumption of linkage equilibrium between the SNP markers, we first select a total of J widely spaced (and thus independent) 'localities' along the genome (indexed by j). Next, we type n_j SNP markers at/around each locality. The informative markers in the jth locality are indexed by i, $i = 1, \ldots, m_j (m_j \leq n_j)$. Let X_{ij} be the indicator function for the ith informative marker in the jth locality, with value defined as before. Let $D_j = \sum_{i=1}^{m_j}(X_{ij} - 2/3)$. As before, $E^{H_0}(D_j) = 0$. Since D_j's are independent of one another, we have $\mathrm{Var}^{H_0}(\sum_{j=1}^{J} D_j/J) = \sum_{j=1}^{J} \mathrm{Var}^{H_0}(D_j)/J^2 \approx \sum_{j=1}^{J} D_j^2/J^2$. Hence, the statistic, $Z_2 = \sum_{j=1}^{J} D_j/\sqrt{\sum_{j=1}^{J} D_j^2}$, is asymptotically the standard normal distribution under H_0.

Reference

Lee W (2003). Ann Hum Genet. Pp 618–619.

“pairwise population concordance” (PPC) test

- PPC assumes that in a random-mating population, for a given pair of autosomal SNPs, the ratio IBS2 (Aa, Aa) over IBS0 (AA, aa) = 2:1
- For SNPs selected far enough apart to be approximately independent (e.g. 500 kb), a test of binomial proportion can suggest concordant or discordant ancestry for each pair of individuals in the test.
- A pair from different populations is expected to show relatively more IBS0 SNPs; a one-sided test for the departure from a 2:1 ratio is given by the normal approximation to the binomial: (L is the total number of informative, independent SNP pairs and L_2 is the IBS2 subset)

$$Z = \frac{\frac{L_2}{L} - \frac{2}{3}}{\sqrt{\frac{2}{3} \times \frac{1}{3} \times \frac{1}{L}}} .$$

- A threshold, e.g. 1e-3, of testing significance provides the clustering criterion.

Reference Purcell, S, et al. “PLINK”, Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

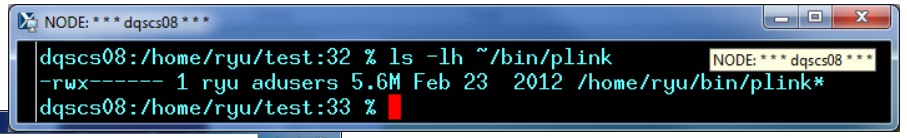
population stratification in PLINK

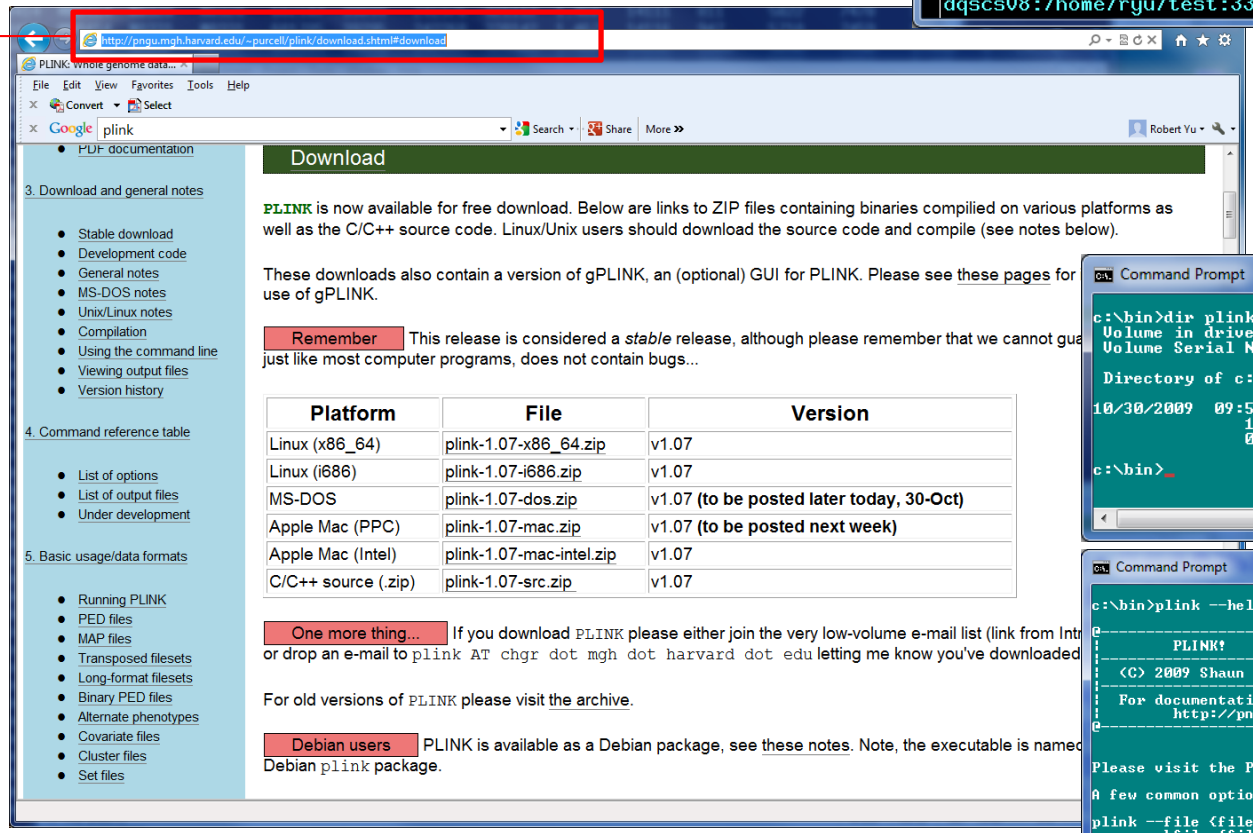
- PLINK is one of the most powerful tools for GWAS
- PLINK deals with the confounding effect in the population-based GWAS data sets
  - ❖ Population stratification
    - ▪ Heterogeneity in cases
    - ▪ Heterogeneity in cases with controls
  - ❖ Non-random genotyping failure
- PLINK uses approach of a population-based linkage analyses by estimating IBD (segment) between seemingly unrelated individuals.

Reference Purcell, S, et al. "PLINK", Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

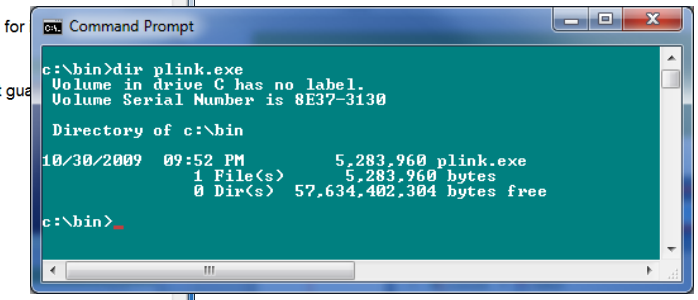
PLINK

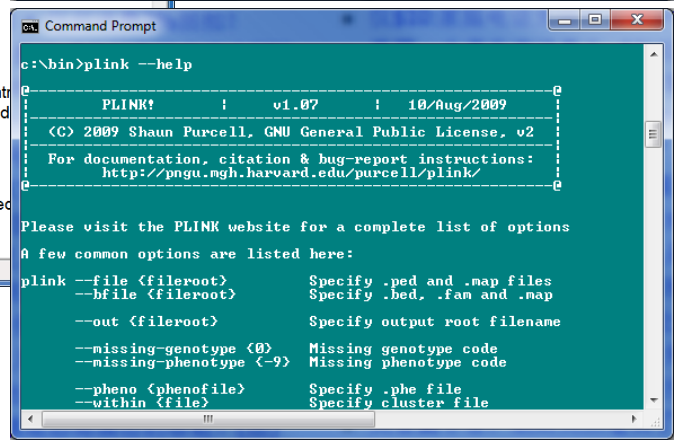
Linux version





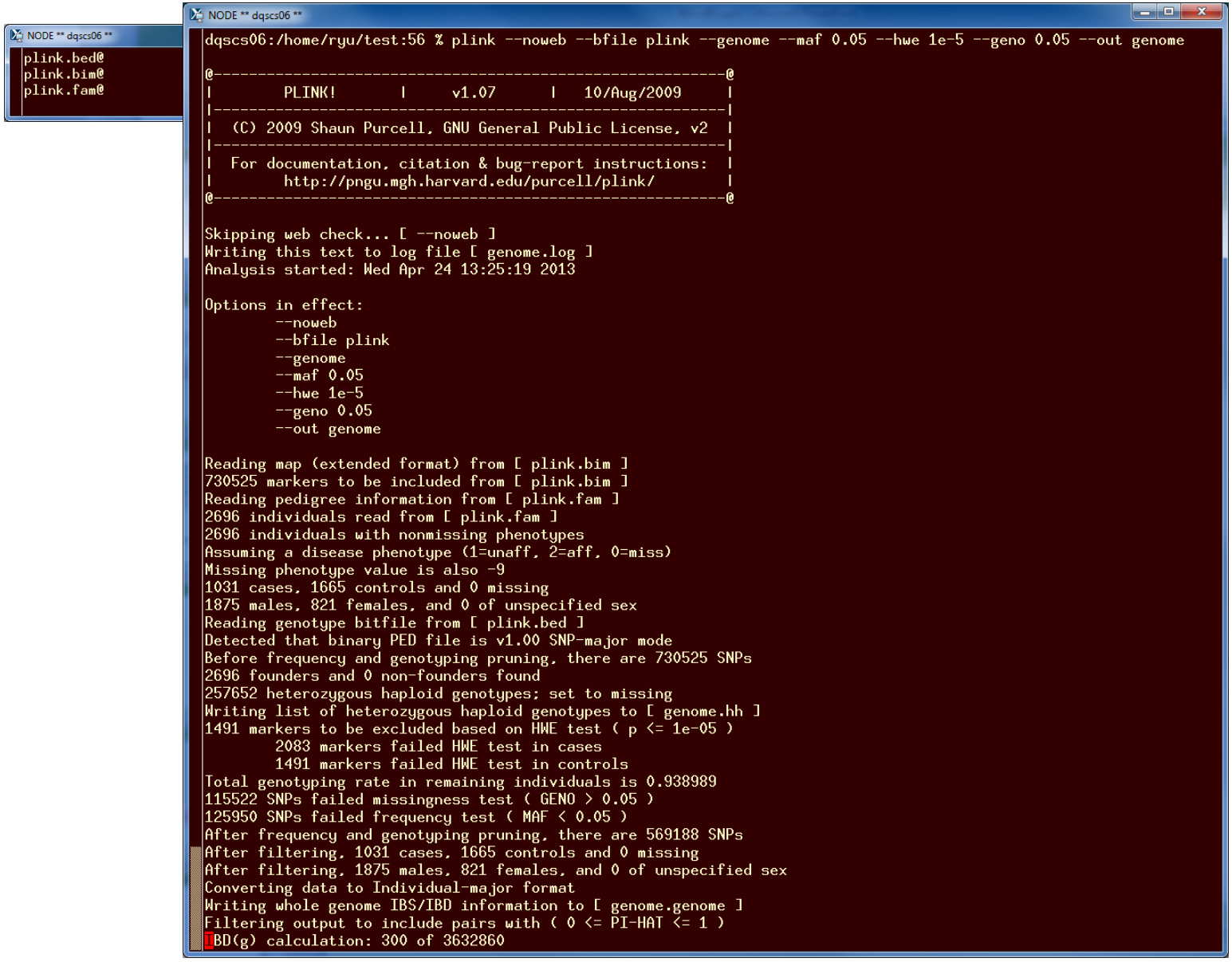
Windows version
Running under CMD window





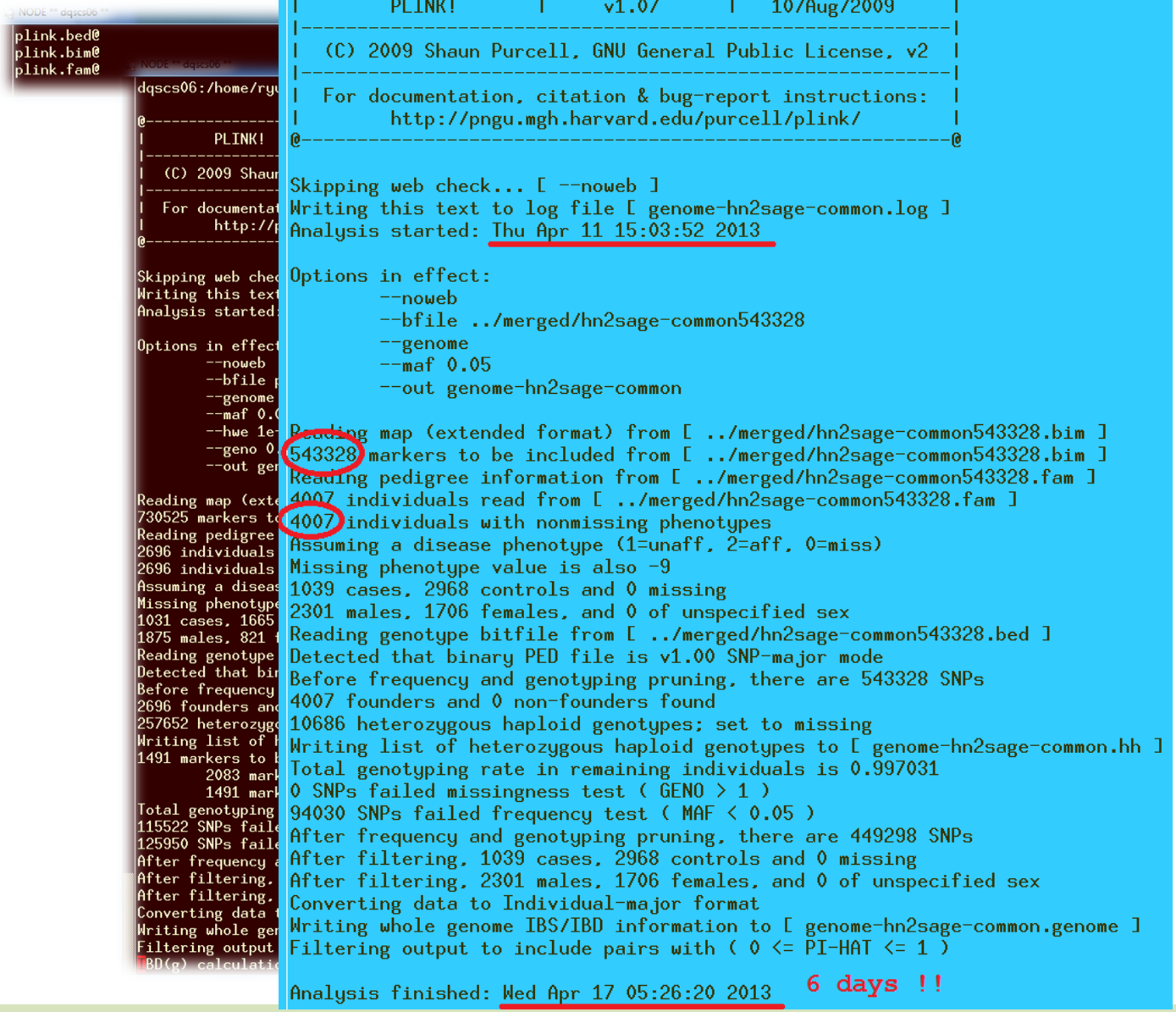
Reference http://pngu.mgh.harvard.edu/~purcell/plink/download.shtml#download

running PLINK



Reference Purcell, S, et al. “PLINK”, Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

running PLINK

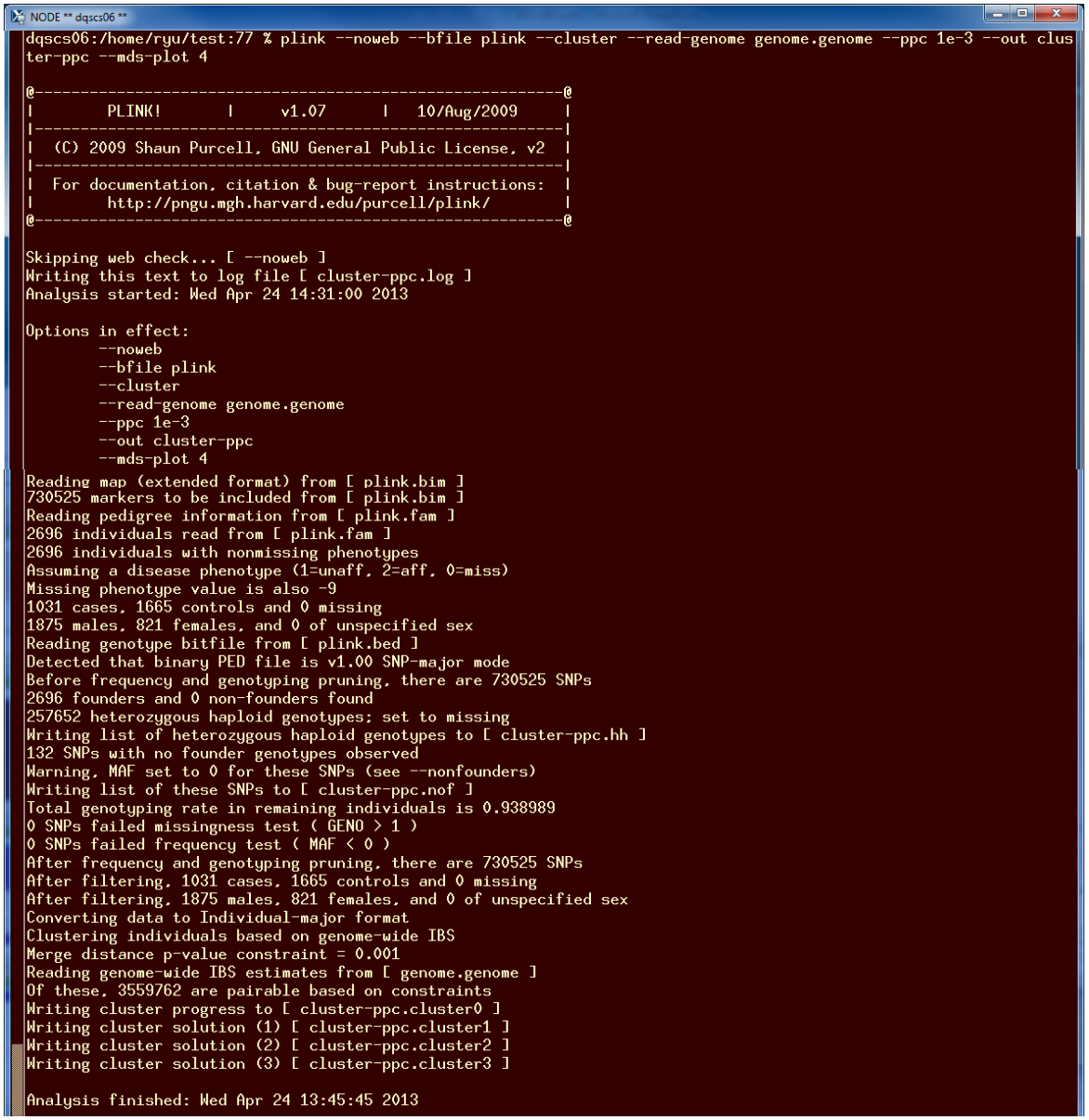


```
plink.bed@
plink.bim@
plink.fam@
```

```
@----------------------------------------------------------@
|        PLINK!       |     v1.07      |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Skipping web check... [ --noweb ]
Writing this text to log file [ genome-hn2sage-common.log ]
Analysis started: Thu Apr 11 15:03:52 2013

Options in effect:
        --noweb
        --bfile ../merged/hn2sage-common543328
        --genome
        --maf 0.05
        --out genome-hn2sage-common

Reading map (extended format) from [ ../merged/hn2sage-common543328.bim ]
543328 markers to be included from [ ../merged/hn2sage-common543328.bim ]
Reading pedigree information from [ ../merged/hn2sage-common543328.fam ]
4007 individuals read from [ ../merged/hn2sage-common543328.fam ]
4007 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
1039 cases, 2968 controls and 0 missing
2301 males, 1706 females, and 0 of unspecified sex
Reading genotype bitfile from [ ../merged/hn2sage-common543328.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 543328 SNPs
4007 founders and 0 non-founders found
10686 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ genome-hn2sage-common.hh ]
Total genotyping rate in remaining individuals is 0.997031
0 SNPs failed missingness test ( GENO > 1 )
94030 SNPs failed frequency test ( MAF < 0.05 )
After frequency and genotyping pruning, there are 449298 SNPs
After filtering, 1039 cases, 2968 controls and 0 missing
After filtering, 2301 males, 1706 females, and 0 of unspecified sex
Converting data to Individual-major format
Writing whole genome IBS/IBD information to [ genome-hn2sage-common.genome ]
Filtering output to include pairs with ( 0 <= PI-HAT <= 1 )

Analysis finished: Wed Apr 17 05:26:20 2013   6 days !!
```

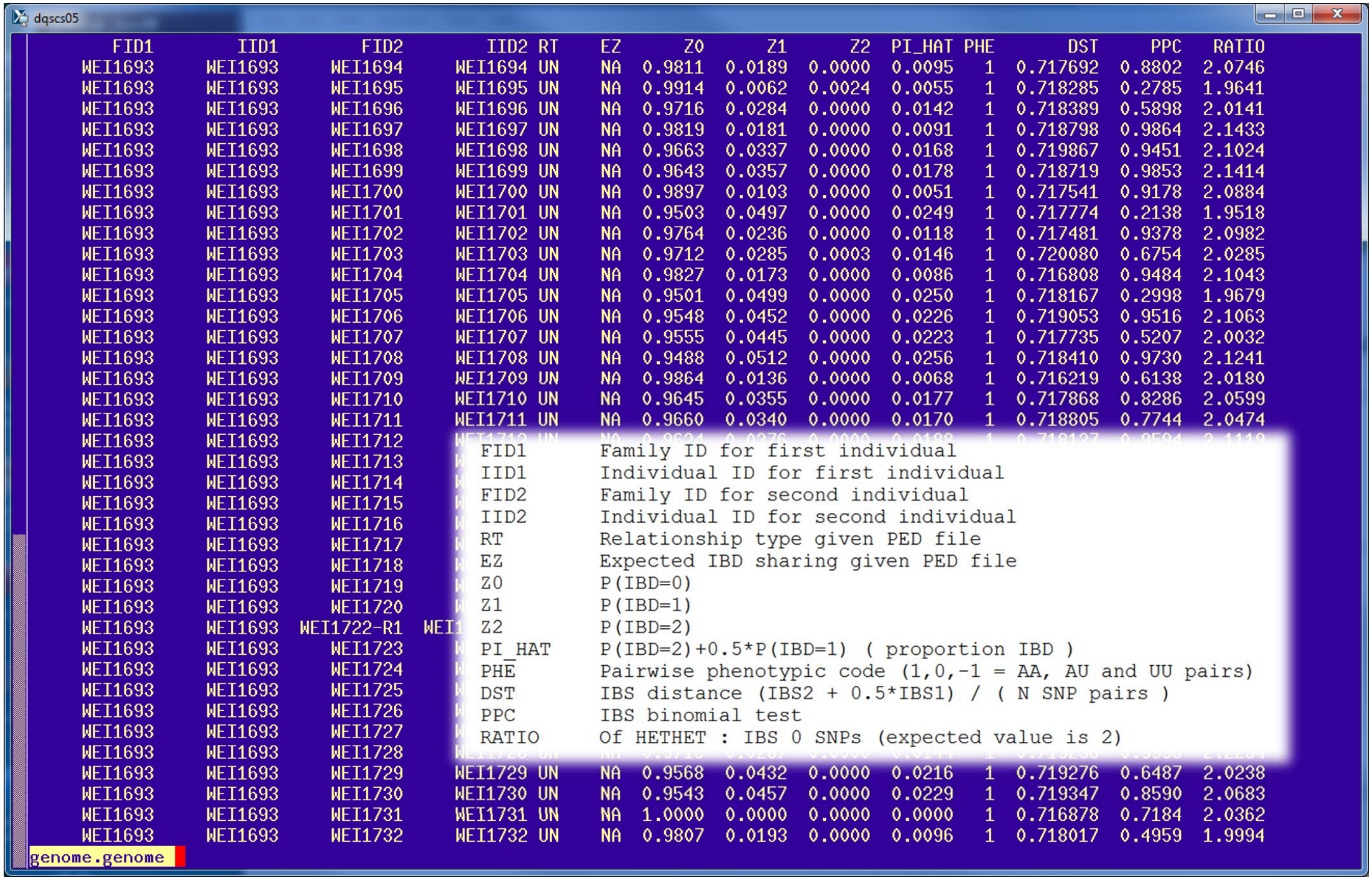
Reference Purcell, S, et al. “PLINK”, Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

running PLINK



Reference Purcell, S, et al. “PLINK”, Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

running PLINK



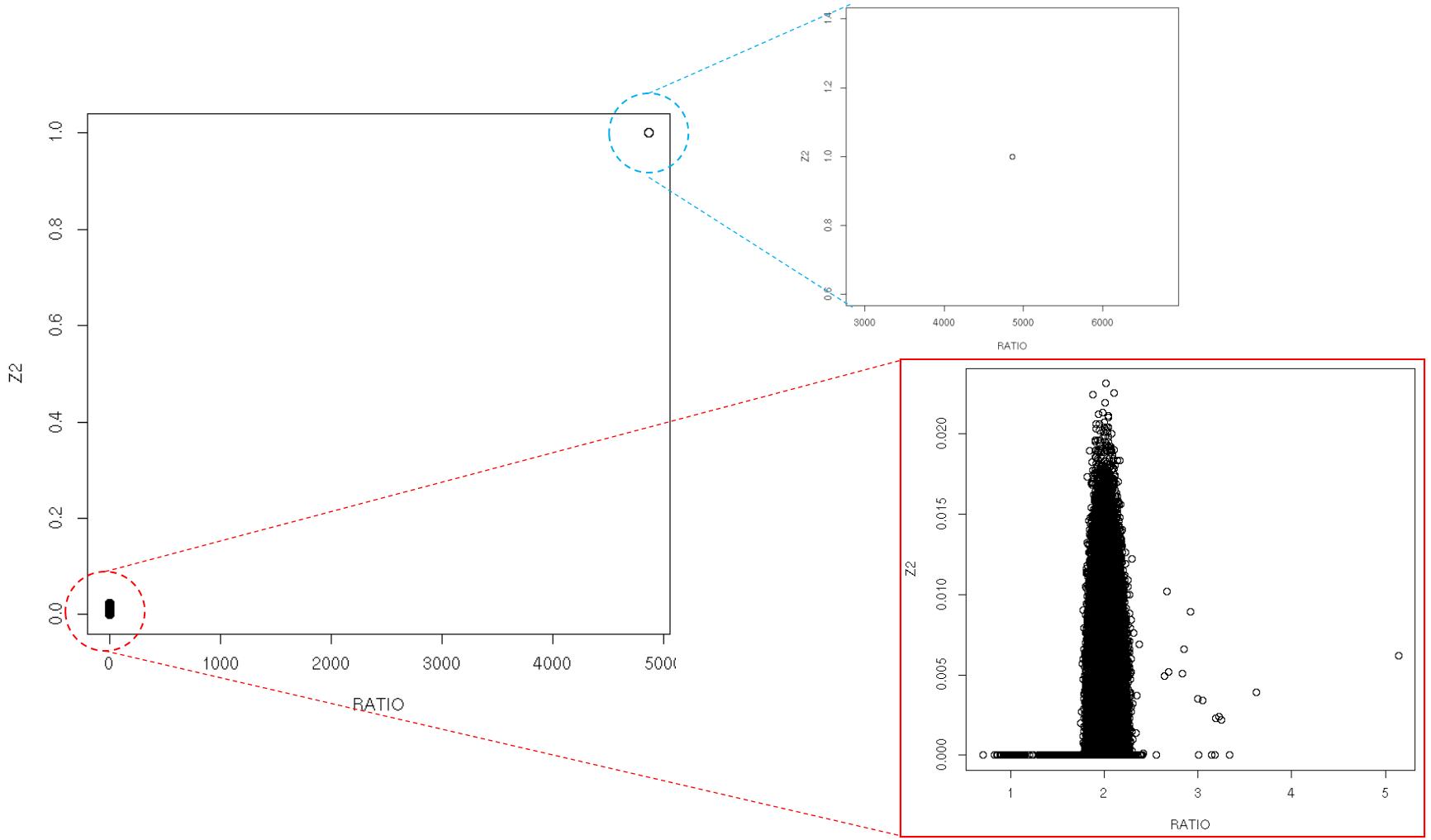
FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	PHE	DST	PPC	RATIO
WEI1693	WEI1693	WEI1694	WEI1694	UN	NA	0.9811	0.0189	0.0000	0.0095	1	0.717692	0.8802	2.0746
WEI1693	WEI1693	WEI1695	WEI1695	UN	NA	0.9914	0.0062	0.0024	0.0055	1	0.718285	0.2785	1.9641
WEI1693	WEI1693	WEI1696	WEI1696	UN	NA	0.9716	0.0284	0.0000	0.0142	1	0.718389	0.5898	2.0141
WEI1693	WEI1693	WEI1697	WEI1697	UN	NA	0.9819	0.0181	0.0000	0.0091	1	0.718798	0.9864	2.1433
WEI1693	WEI1693	WEI1698	WEI1698	UN	NA	0.9663	0.0337	0.0000	0.0168	1	0.719867	0.9451	2.1024
WEI1693	WEI1693	WEI1699	WEI1699	UN	NA	0.9643	0.0357	0.0000	0.0178	1	0.718719	0.9853	2.1414
WEI1693	WEI1693	WEI1700	WEI1700	UN	NA	0.9897	0.0103	0.0000	0.0051	1	0.717541	0.9178	2.0884
WEI1693	WEI1693	WEI1701	WEI1701	UN	NA	0.9503	0.0497	0.0000	0.0249	1	0.717774	0.2138	1.9518
WEI1693	WEI1693	WEI1702	WEI1702	UN	NA	0.9764	0.0236	0.0000	0.0118	1	0.717481	0.9378	2.0982
WEI1693	WEI1693	WEI1703	WEI1703	UN	NA	0.9712	0.0285	0.0003	0.0146	1	0.720080	0.6754	2.0285
WEI1693	WEI1693	WEI1704	WEI1704	UN	NA	0.9827	0.0173	0.0000	0.0086	1	0.716808	0.9484	2.1043
WEI1693	WEI1693	WEI1705	WEI1705	UN	NA	0.9501	0.0499	0.0000	0.0250	1	0.718167	0.2998	1.9679
WEI1693	WEI1693	WEI1706	WEI1706	UN	NA	0.9548	0.0452	0.0000	0.0226	1	0.719053	0.9516	2.1063
WEI1693	WEI1693	WEI1707	WEI1707	UN	NA	0.9555	0.0445	0.0000	0.0223	1	0.717735	0.5207	2.0032
WEI1693	WEI1693	WEI1708	WEI1708	UN	NA	0.9488	0.0512	0.0000	0.0256	1	0.718410	0.9730	2.1241
WEI1693	WEI1693	WEI1709	WEI1709	UN	NA	0.9864	0.0136	0.0000	0.0068	1	0.716219	0.6138	2.0180
WEI1693	WEI1693	WEI1710	WEI1710	UN	NA	0.9645	0.0355	0.0000	0.0177	1	0.717868	0.8286	2.0599
WEI1693	WEI1693	WEI1711	WEI1711	UN	NA	0.9660	0.0340	0.0000	0.0170	1	0.718805	0.7744	2.0474
WEI1693	WEI1693	WEI1712	[illegible]										
WEI1693	WEI1693	WEI1713											
WEI1693	WEI1693	WEI1714											
WEI1693	WEI1693	WEI1715											
WEI1693	WEI1693	WEI1716											
WEI1693	WEI1693	WEI1717											
WEI1693	WEI1693	WEI1718											
WEI1693	WEI1693	WEI1719											
WEI1693	WEI1693	WEI1720											
WEI1693	WEI1693	WEI1722-R1	WEI1										
WEI1693	WEI1693	WEI1723											
WEI1693	WEI1693	WEI1724											
WEI1693	WEI1693	WEI1725											
WEI1693	WEI1693	WEI1726											
WEI1693	WEI1693	WEI1727											
WEI1693	WEI1693	WEI1728	[illegible]										
WEI1693	WEI1693	WEI1729	WEI1729	UN	NA	0.9568	0.0432	0.0000	0.0216	1	0.719276	0.6487	2.0238
WEI1693	WEI1693	WEI1730	WEI1730	UN	NA	0.9543	0.0457	0.0000	0.0229	1	0.719347	0.8590	2.0683
WEI1693	WEI1693	WEI1731	WEI1731	UN	NA	1.0000	0.0000	0.0000	0.0000	1	0.716878	0.7184	2.0362
WEI1693	WEI1693	WEI1732	WEI1732	UN	NA	0.9807	0.0193	0.0000	0.0096	1	0.718017	0.4959	1.9994

FID1	Family ID for first individual
IID1	Individual ID for first individual
FID2	Family ID for second individual
IID2	Individual ID for second individual
RT	Relationship type given PED file
EZ	Expected IBD sharing given PED file
Z0	P(IBD=0)
Z1	P(IBD=1)
Z2	P(IBD=2)
PI_HAT	P(IBD=2)+0.5*P(IBD=1) (proportion IBD)
PHE	Pairwise phenotypic code (1,0,-1 = AA, AU and UU pairs)
DST	IBS distance (IBS2 + 0.5*IBS1) / (N SNP pairs)
PPC	IBS binomial test
RATIO	Of HETHET : IBS 0 SNPs (expected value is 2)

Reference Purcell, S, et al. "PLINK", Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

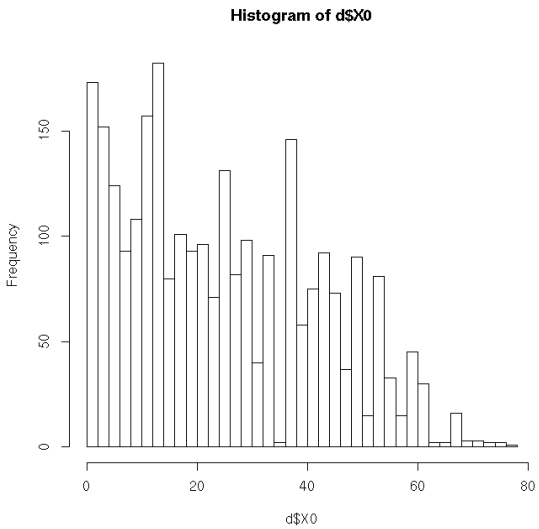
running PLINK

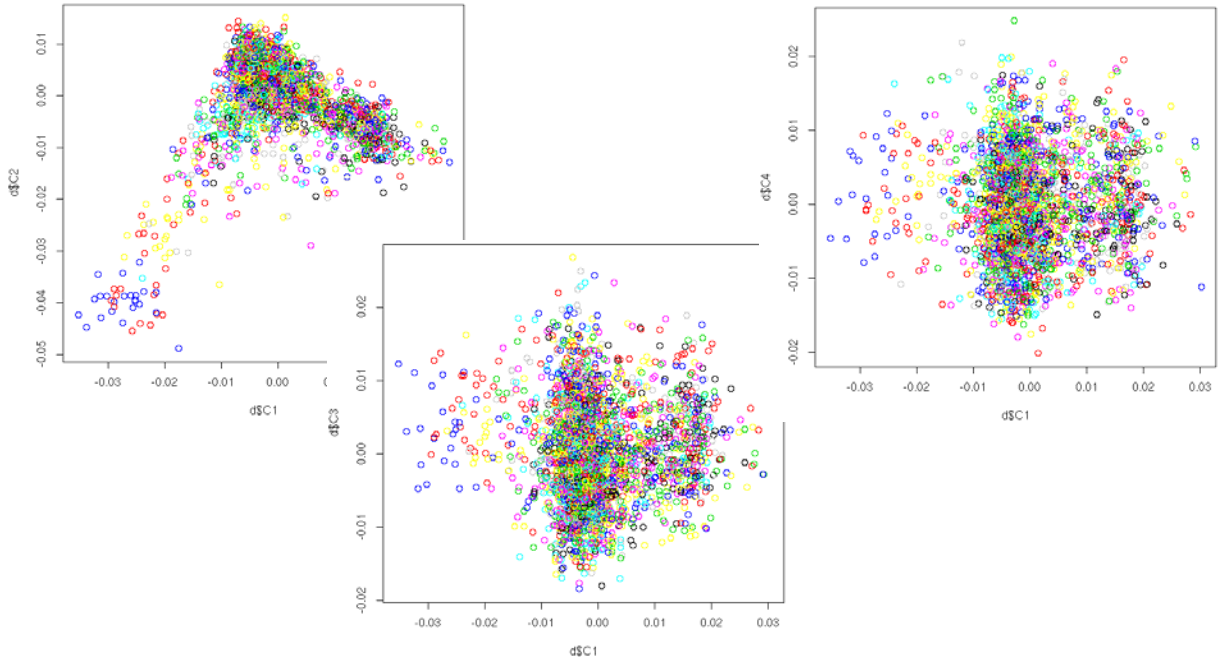
```
dqscs05
     FID1        IID1        FID2        IID2 RT    EZ      Z0      Z1      Z2  PI_HAT PHE       DST     PPC   RATIO
  WEI3155     WEI3155      M1571     M1571 UN    NA  0.0002  0.0000  0.9998  0.9998  -1  0.999958  1.0000 4863.0000
dqscs05:/home/ryu/test:8 %
```



Reference Purcell, S, et al. "PLINK", Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

running PLINK





Reference Purcell, S, et al. "PLINK", Am. J. Human Genetics, Vol 81., pp 559-75 (Sept 2007)

running KING

BIOINFORMATICS **ORIGINAL PAPER** Vol. 26 no. 22 2010, pages 2867–2873
doi:10.1093/bioinformatics/btq559

Genetics and population analysis Advance Access publication October 5, 2010

Robust relationship inference in genome-wide association studies

Ani Manichaikul[1,2], Josyf C. Mychaleckyj[1], Stephen S. Rich[1], Kathy Daly[3], Michèle Sale[1,4,5] and Wei-Min Chen[1,2,*]

[1]Center for Public Health Genomics, [2]Department of Public Health Sciences, Division of Biostatistics a Epidemiology, University of Virginia, Charlottesville, VA, [3]Department of Otolaryngology, University of Mi Minneapolis, MN, [4]Department of Medicine and [5]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA

Availability: Our robust relationship inference algorithm is implemented in a freely available software package, KING, available for download at http://people.virginia.edu/~wc9c/KING.

Contact: wmchen@virginia.edu

We present a novel framework for relationship inference, Kinship-based INference for Genome-wide association studies (KING),

Consider two individuals, indexed by i and j. Let ϕ_{ij} denote the kinship coefficient, defined as the probability that two alleles sampled at random from two individuals are identical by descent, and π_{0ij}, π_{1ij} and π_{2ij} denote the probability that the two individuals share zero, one and two alleles identical by descent, respectively. Table 1 lists values of ϕ_{ij} and π_{0ij} for relative

Table 1. Relationship inference criteria based on estimating kinship coefficients (ϕ) and probability of zero IBD sharing (π_0)

Relationship	ϕ	Inference criteria	π_0	Inference criteria
Monozygotic twin	$\frac{1}{2}$	$> \frac{1}{2^{3/2}}$	0	< 0.1
Parent–offspring	$\frac{1}{4}$	$(\frac{1}{2^{5/2}}, \frac{1}{2^{3/2}})$	0	< 0.1
Full sib	$\frac{1}{4}$	$(\frac{1}{2^{5/2}}, \frac{1}{2^{3/2}})$	$\frac{1}{4}$	(0.1,0.365)
2nd Degree	$\frac{1}{8}$	$(\frac{1}{2^{7/2}}, \frac{1}{2^{5/2}})$	$\frac{1}{2}$	$(0.365, 1-\frac{1}{2^{3/2}})$
3rd Degree	$\frac{1}{16}$	$(\frac{1}{2^{9/2}}, \frac{1}{2^{7/2}})$	$\frac{3}{4}$	$(1-\frac{1}{2^{3/2}}, 1-\frac{1}{2^{5/2}})$
Unrelated	0	$< \frac{1}{2^{9/2}}$	1	$> 1-\frac{1}{2^{5/2}}$

Robust relationship inference in the presence of population substructure

$$\hat{\phi}_{ij} = \frac{1}{2} - \frac{1}{4}\frac{\sum_m \left(X_m^{(i)} - X_m^{(j)}\right)^2}{N_{Aa}^{(i)}} = \frac{N_{Aa,Aa} - 2N_{AA,aa}}{2N_{Aa}^{(i)}} + \frac{1}{2} - \frac{1}{4}\frac{N_{Aa}^{(i)} + N_{Aa}^{(j)}}{N_{Aa}^{(i)}}$$

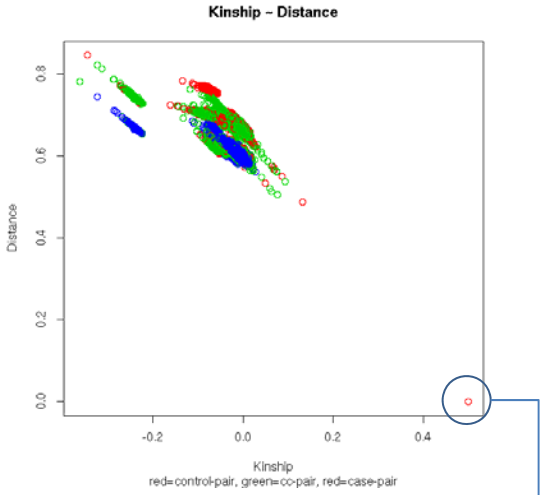
Table 2. Computation time of two software implementations to estimate kinship coefficients in three sets of GWAS SNP data

Summary of genome scan data				Computing time	
Index	No. of SNPs	No. of samples	No. of pairs	KING	PLINK
1	3 079 857	269	36 046	2 min	2 h 9 min
2	324 748	602	180 901	1 min	1 h 13 min
3	549 338	2 454	3 009 832	25 min	28 h 30 min

Reference *Manichaikul A,…, **Chen WM** (2010) Robust relationship inference in genome-wide association studies. **Bioinformatics** 26(22):2867-2873*

running KING

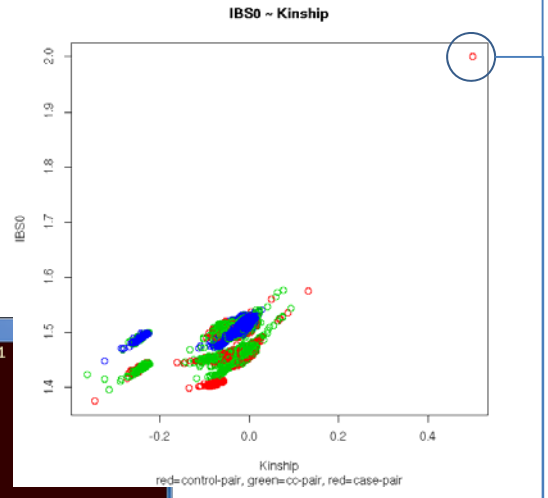
```
dqscs06:/home/ryu/test:85 % king -b plink.bed --kinship --ibs --prefix king-plink
KING 1.4 - (c) 2010-2011 Wei-Min Chen

The following parameters are in effect:
                    Data File :                (-dname)
                Pedigree File :                (-pname)
                     Map File :                (-mname)
                  Binary File :       plink.bed (-bname)

Additional Options
        Family Relationship : --kinship [ON], --ibs [ON]
    Relationship Parameter : --related, --degree
  Relationship Application : --unrelated
   Pedigree Reconstruction : --cluster, --build, --errorrate
    Homogeneous Population : --homo, --minMAF, --showIBD
      Population Structure : --individual, --pca, --mds
                    Output : --prefix [king-plink]
         Rewrite In Format : --binary, --merlin, --plink

Loading genotype data in PLINK binary format...
Read in PLINK fam file plink.fam...
  PLINK pedigrees loaded: 2696 samples
Read in PLINK bim file plink.bim...
  Genotype data consist of 711061 autosome SNPs (including 473 XY SNPs), 18055 X-chromosome SNPs, 1409 Y-chromosome
 SNPs
  PLINK maps loaded: 730525 SNPs
Read in PLINK bed file plink.bed...
  PLINK binary genotypes loaded and converted to KING binary.
Both kinship and IBS statistics will be analyzed.
Genotypes stored in 44442 integers for each of 2696 individuals.
Each family consists of one individual.
Relationship inference across families starts at Wed Apr 24 15:20:05 2013
                                          ends at Wed Apr 24 15:39:35 2013
Between-family kinship data saved in file king-plink.ibs0
```



NODE ** dqscs06 **

FID1	ID1	FID2	ID2	N_IBS0	N_IBS1	N_IBS2	IBS	SE_IBS	N_HetHet		N_Het1	N_Het2	Distance	SE_Dist	Kinship
MN0414	MN0414	MN0627	MN0627	39769	242577	321036	1.466	0.617	76961	198925	197574	0.666	1.0067	-0.0082	
MN0414	MN0414	M0214	M0214	40630	241779	319836	1.464	0.619	76197	198620	195553	0.671	1.0150	-0.0169	
MN0414	MN0414	M0217	M0217	38885	242774	322209	1.469	0.615	78570	199107	200807	0.660	0.9986	-0.0001	
MN0414	MN0414	M0218	M0218	38174	242306	321277	1.470	0.613	77394	198381	198713	0.656	0.9934	0.0022	
MN0414	MN0414	M0220	M0220	39143	244228	320514	1.466	0.615	79054	199094	203242	0.664	1.0005	-0.0033	
MN0414	MN0414	M0327	M0327	39454	243228	321114	1.466	0.616	78065	199066	200292	0.664	1.0036	-0.0037	
MN0414	MN0414	M0328	M0328	39099	241976	321112	1.468	0.616	78214	198523	199881	0.662	1.0015	-0.0017	
MN0414	MN0414	M0331	M0331	39295	242268	320142	1.467	0.616	77030	198356	197972	0.664	1.0034	-0.0044	
MN0414	MN0414	M0337	M0337	39278	242437	321469	1.468	0.616	78917	198881	201390	0.662	1.0025	-0.0022	
MN0414	MN0414	M0348	M0348	38258	243509	320740	1.469	0.613	77585	198617	200062	0.658	0.9935	0.0009	
MN0414	MN0414	M0350	M0350	39756	242822	321228	1.466	0.617	78304	199084	200346	0.666	1.0063	-0.0046	
MN0414	MN0414	M0356	M0356	38871	242928	321494	1.468	0.615	77357	198900	198742	0.660	0.9987	-0.0012	
MN0414	MN0414	M0370	M0370	39664	244258	319323	1.464	0.617	77837	198874	201058	0.668	1.0054	-0.0065	
MN0414	MN0414	M0371	M0371	39435	243830	320505	1.466	0.616	77790	199059	200351	0.665	1.0033	-0.0043	

king-plink.ibs0



NODE ** dqscs06 **

FID1	ID1	FID2	ID2	N_IBS0	N_IBS1	N_IBS2	IBS	SE_IBS	N_HetHet		N_Het1	N_Het2	Distance	SE_Dist	Kinship	Sex1	Pheno1
MN0414	MN0414	MN0627	MN0627	39769	242577	321036	1.466	0.617	76961	198925	197574	0.666	1.0067	-0.0082	2	1	1
MN0414	MN0414	M0214	M0214	40630	241779	319836	1.464	0.619	76197	198620	195553	0.671	1.0150	-0.0169	2	1	1
MN0414	MN0414	M0217	M0217	38885	242774	322209	1.469	0.615	78570	199107	200807	0.660	0.9986	-0.0001	2	1	2
MN0414	MN0414	M0218	M0218	38174	242306	321277	1.470	0.613	77394	198381	198713	0.656	0.9934	0.0022	2	1	2
MN0414	MN0414	M0220	M0220	39143	244228	320514	1.466	0.615	79054	199094	203242	0.664	1.0005	-0.0033	2	1	2
MN0414	MN0414	M0327	M0327	39454	243228	321114	1.466	0.616	78065	199066	200292	0.664	1.0036	-0.0037	2	1	1
MN0414	MN0414	M0328	M0328	39099	241976	321112	1.468	0.616	78214	198523	199881	0.662	1.0015	-0.0017	2	1	1
MN0414	MN0414	M0331	M0331	39295	242268	320142	1.467	0.616	77030	198356	197972	0.664	1.0034	-0.0044	2	1	2

added-sex-pheno-king-plink.ibs0

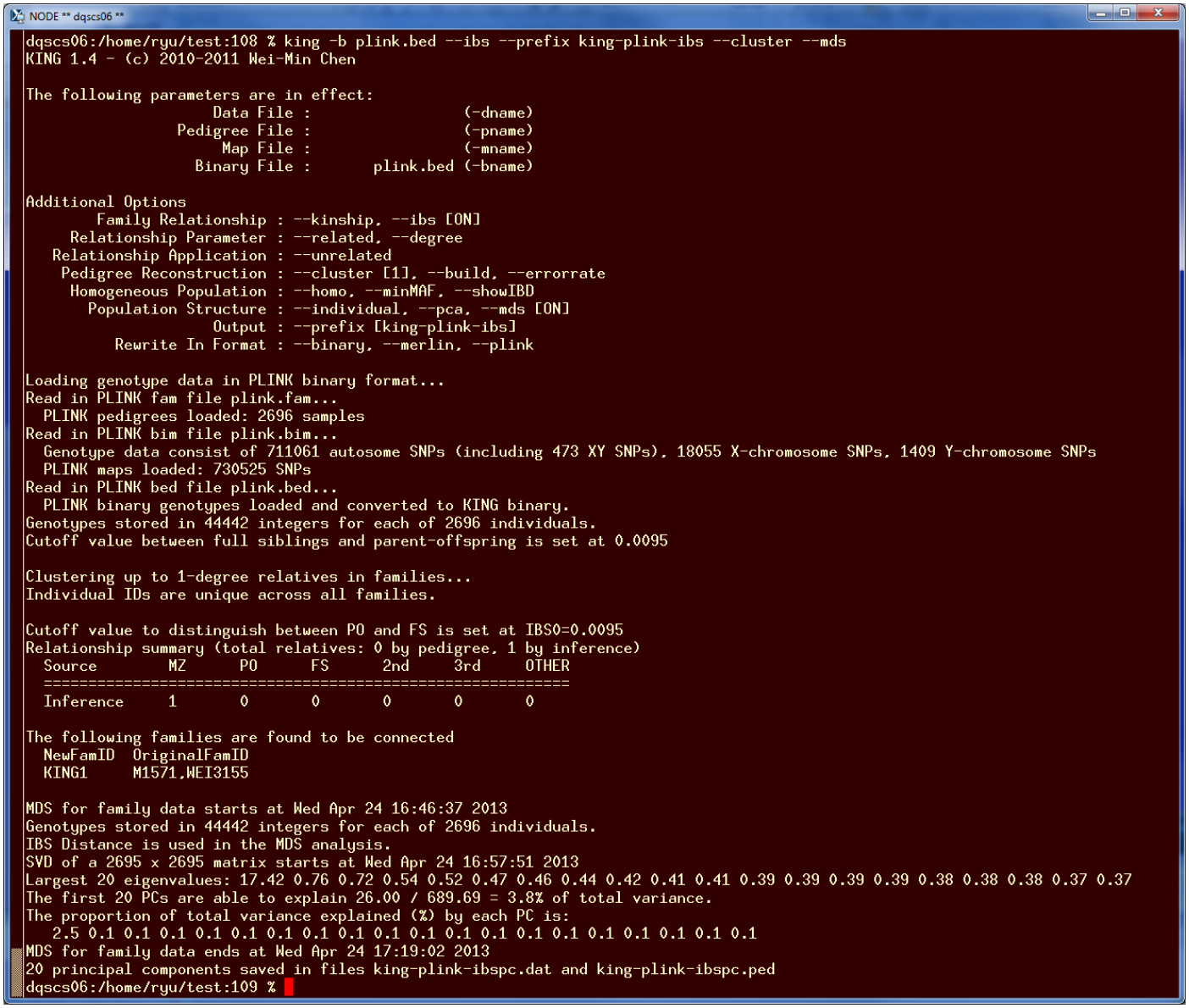
NODE ** dqscs06 **

```
dqscs06:/home/ryu/test:104 % awk '{if($15>0.2) print}' added-sex-pheno-king-plink.ibs0
FID1    ID1     FID2    ID2     N_IBS0  N_IBS1  N_IBS2  IBS     SE_IBS  N_HetHet        N_Het1  N_Het2  Distance        SE_Dist Kinship Sex1    Pheno1  Sex2    Pheno2
M1571   M1571   WEI3155 WEI3155 10      35      601842  2.000   0.011   203651  203667  203670  0.000   0.0180  0.4999  1       1       1       1
dqscs06:/home/ryu/test:105 %
```

Reference

Manichaikul A,…, ***Chen WM*** *(2010) Robust relationship inference in genome-wide association studies.* ***Bioinformatics*** *26(22):2867-2873*

running KING



Reference *Manichaikul A,…,* ***Chen WM*** *(2010) Robust relationship inference in genome-wide association studies.* ***Bioinformatics*** *26(22):2867-2873*

chromosomal IBS patterns

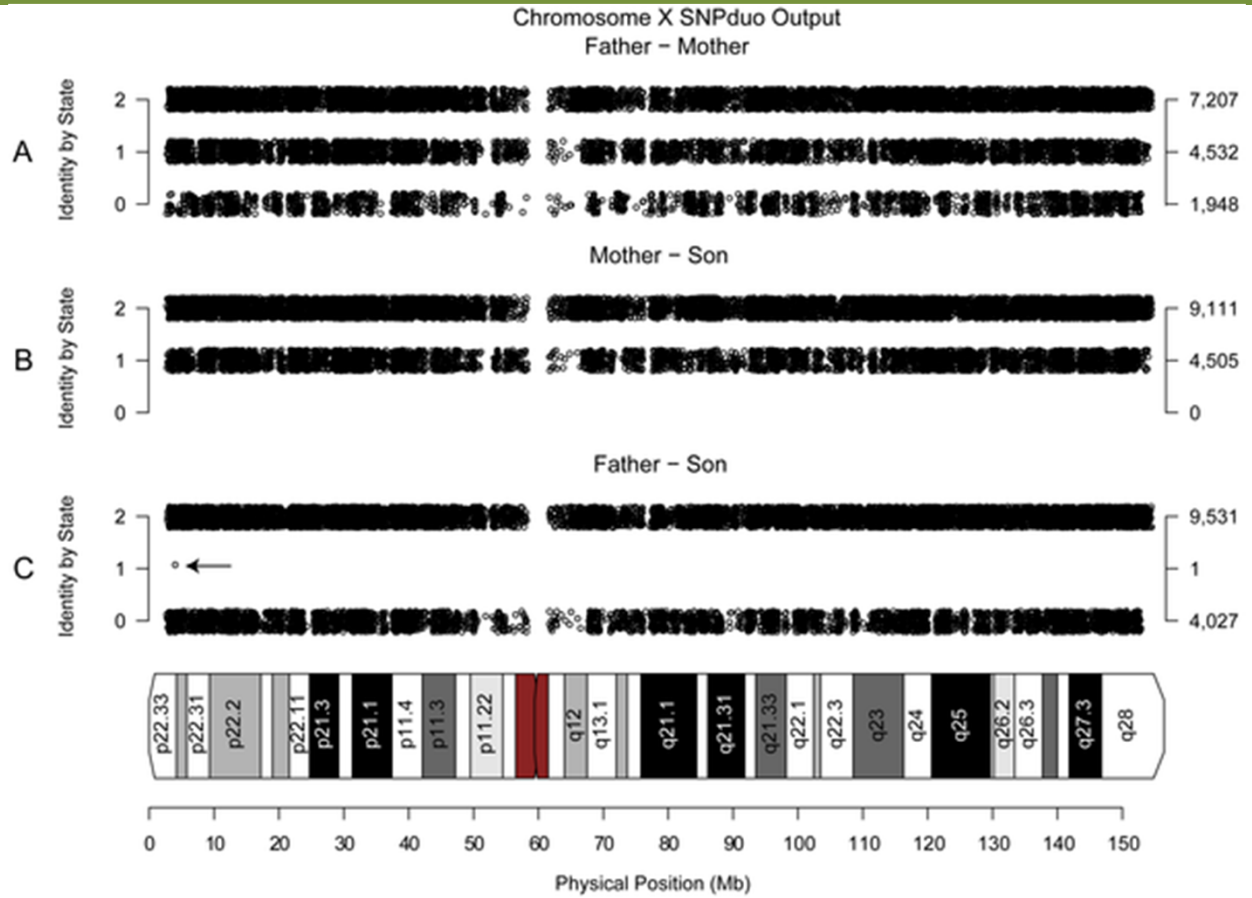
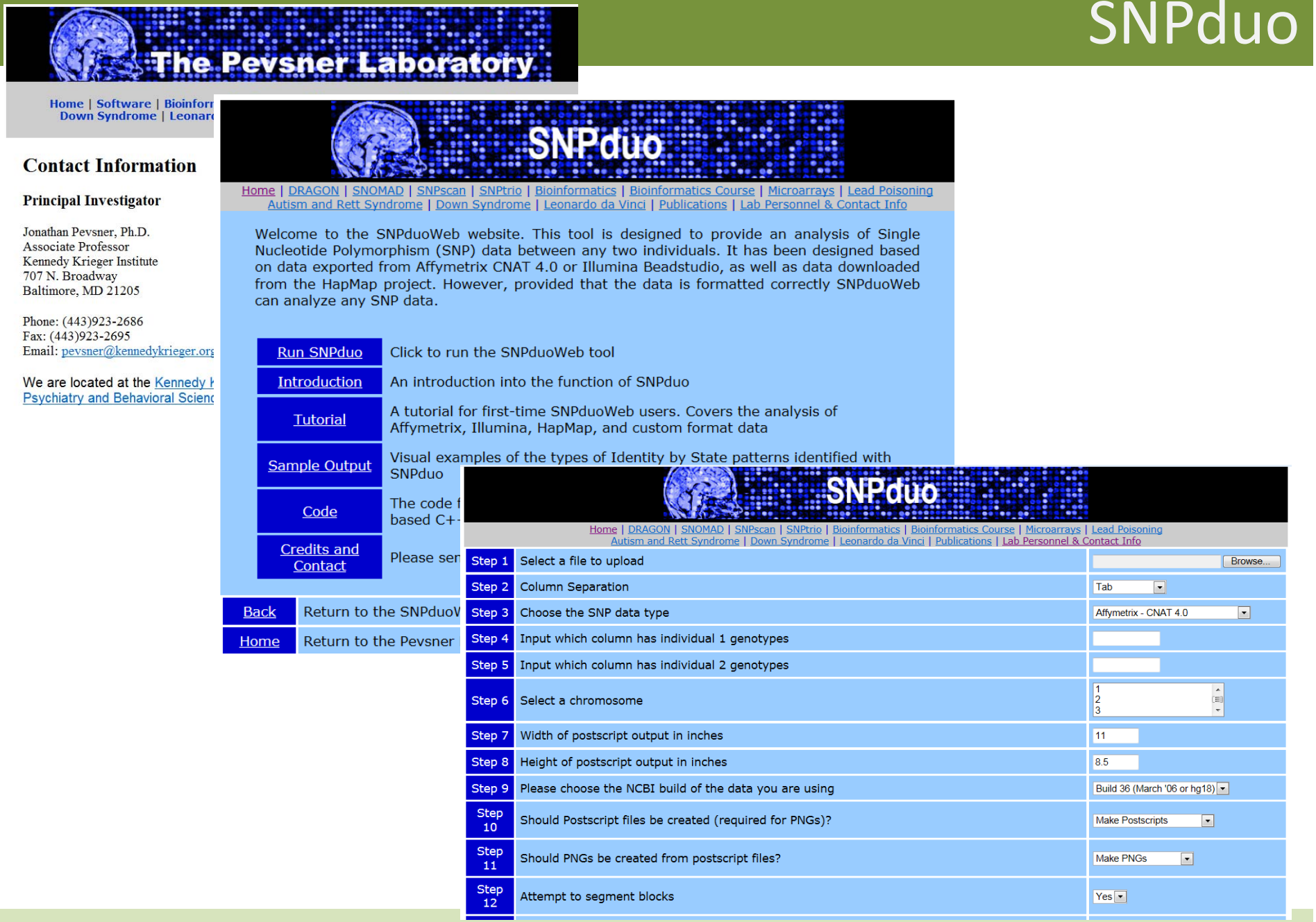


Figure 1. IBS patterns for father, mother, and son on chromosome X.
A portion of the SNPduo output for three pairwise comparisons of the X chromosome of father/mother (A), mother/son (B), and father/son (C) genotyped on the Illumina HumanHap 550K platform. In the unrelated parents, there were many instances of no shared alleles (e.g. AA to BB; panel A). In the mother-son comparison, there were no IBS-0 SNPs because the son inherited a copy of the maternal X. In the father/son comparison, each chromosome was hemizygous (either A or B genotypes, interpreted as AA or BB) and in the absence of heterozygous calls no IBS-1 SNPs were expected to occur since the X chromosomes were non-identical (both IBS-2 and IBS-0 SNPs were apparent). Thus, the one call of an IBS-1 SNP (arrow) was likely a genotyping error.

Reference " Roberson EDO, Pevsner J (2009)Visualization of Shared Genomic Regions and Meiotic Recombination in High-Density SNP Data.PLoS ONE 4(8)

SNPduo



Reference " Roberson EDO, Pevsner J (2009)Visualization of Shared Genomic Regions and Meiotic Recombination in High-Density SNP Data.PLoS ONE 4(8)

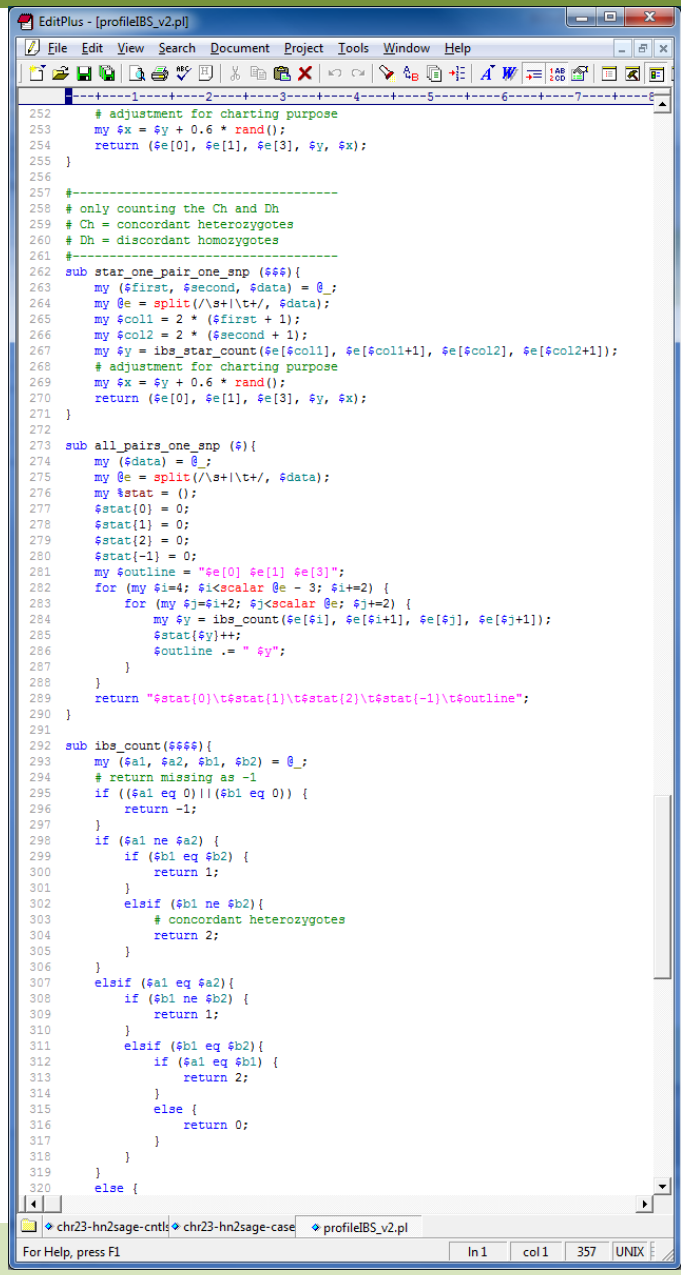
Program to explore IBS patterns

Algorithm:

Sample-pair-loop for pairwise or single process for one pair

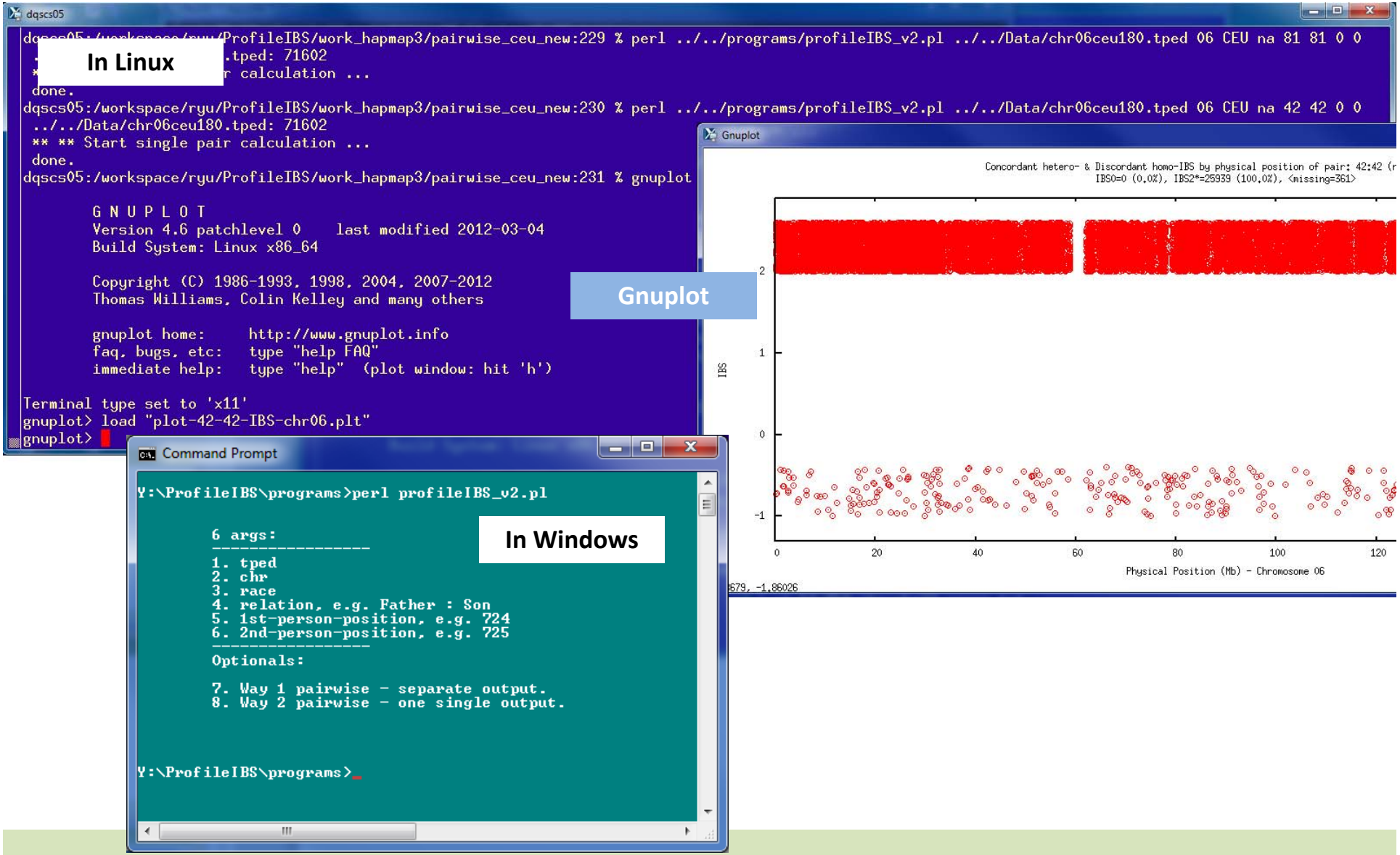
1. Read in tped genotype file from PLINK output
2. Choose pair of individuals
   Marker-loop from SNP_1 -> SNP_N:
   1) Compare alleles, one SNP a time
   2) Save results
      case 0: any missing => missing
      case 1: AA : aa => IBS0*
              => IBS0
      case 2: AA : Aa => IBS1
              Aa : AA => IBS1
      case 3: Aa : Aa => IBS2
              => IBS2*
      case 4: AA : AA => IBS2
              aa : aa => IBS2
   3) Attach SNP info to the result, e.g. chr, bp, rs#
   4) Back to **Marker-loop**
3. Output results
   1) Total IBS counts using {IBS0, IBS1, IBS2, missing}
   2) IBS* (for relationship) using {IBS0*, IBS2*, missing}

   Optional: back to **sample-pair-loop** if pairwise comparison is set.
4. Result summary
   1) Profile plotting using GNUPLOT
   2) Statistics of various counting
   3) Pattern study, e.g. fragments search, etc.
5. Optional: back to sample-pair-loop if looping is activated.



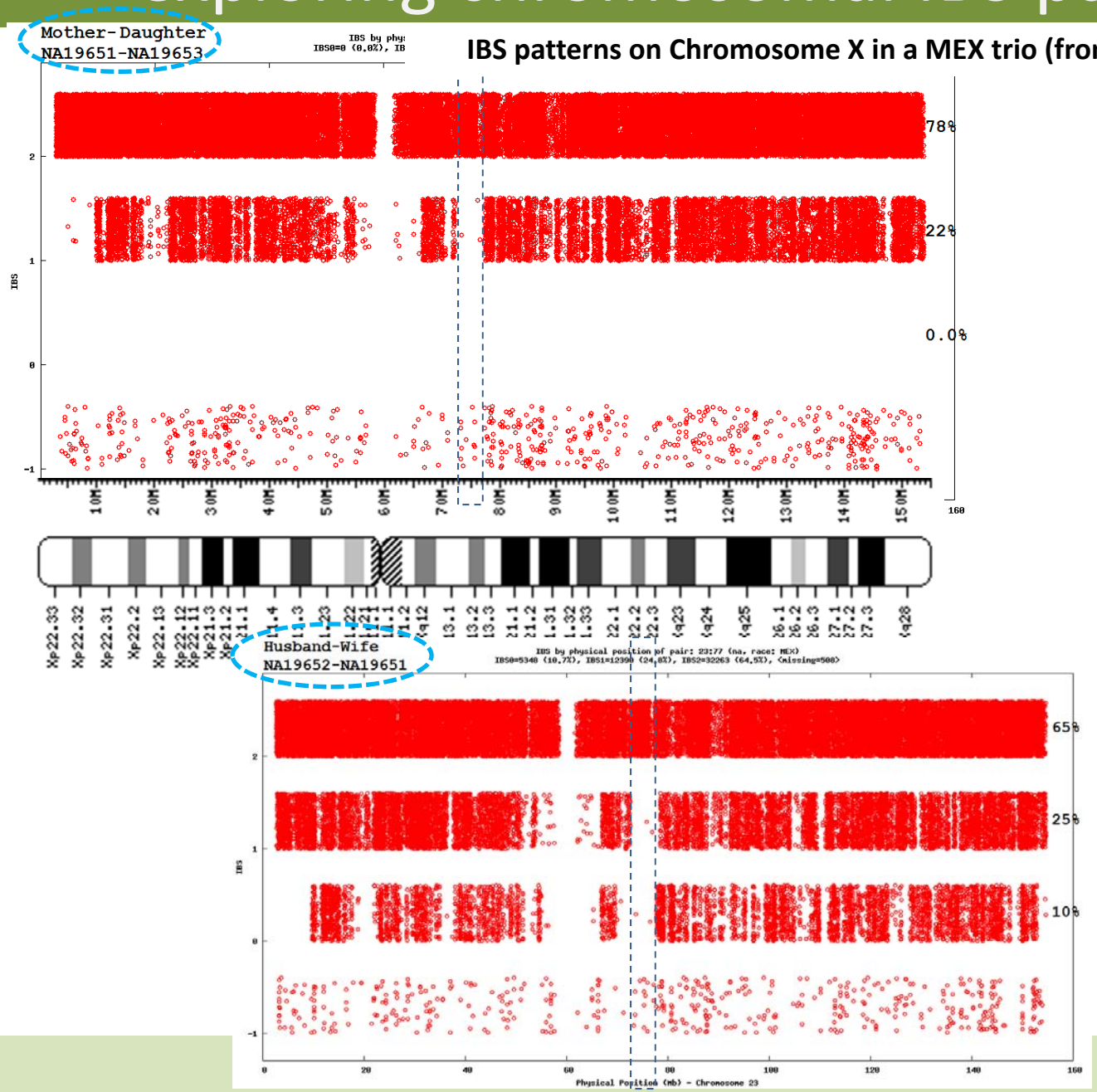
Program to explore IBS patterns

The PERL program can be run either in Linux or in Windows environment



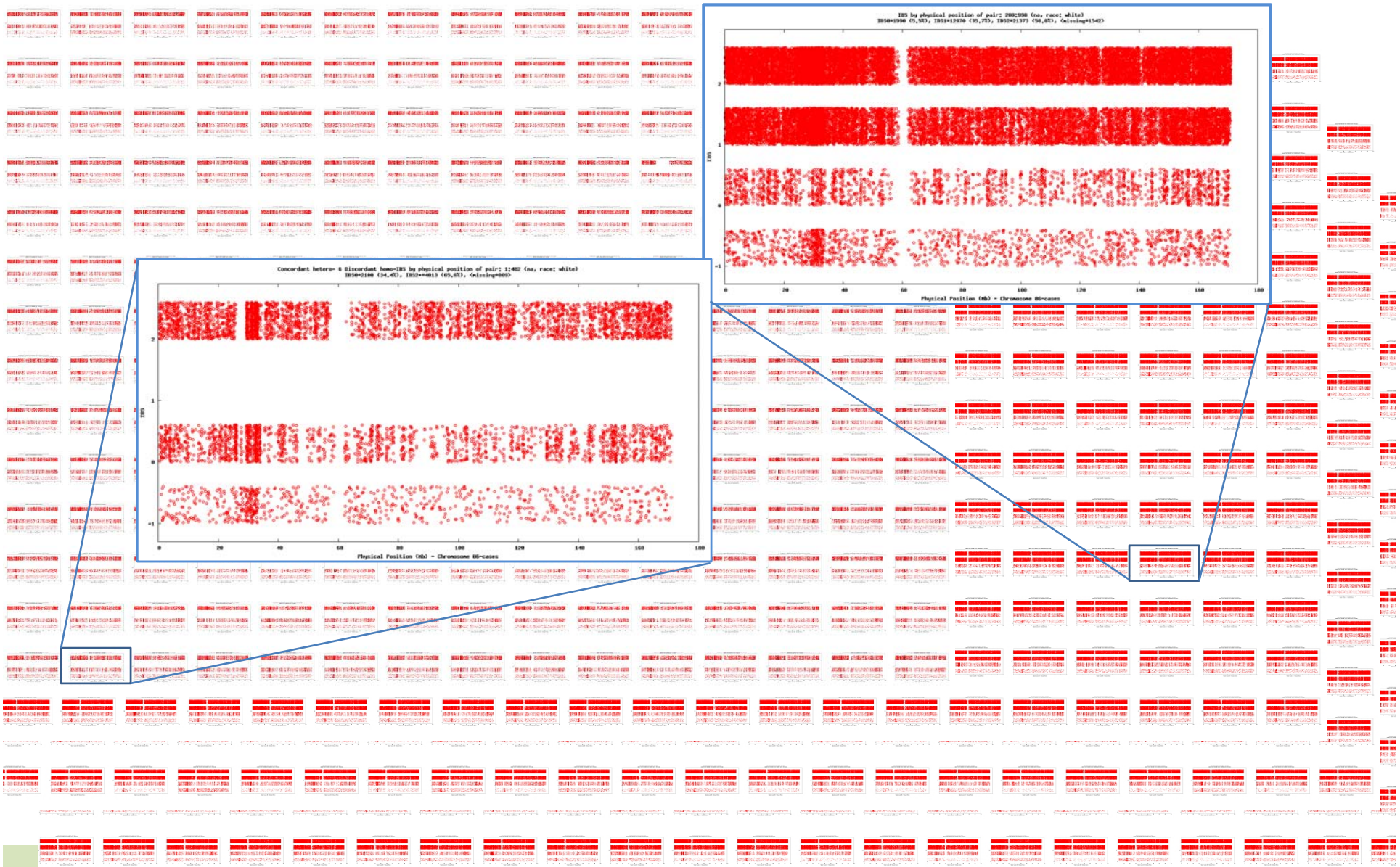
exploring chromosomal IBS patterns

IBS patterns on Chromosome X in a MEX trio (from HapMap3)



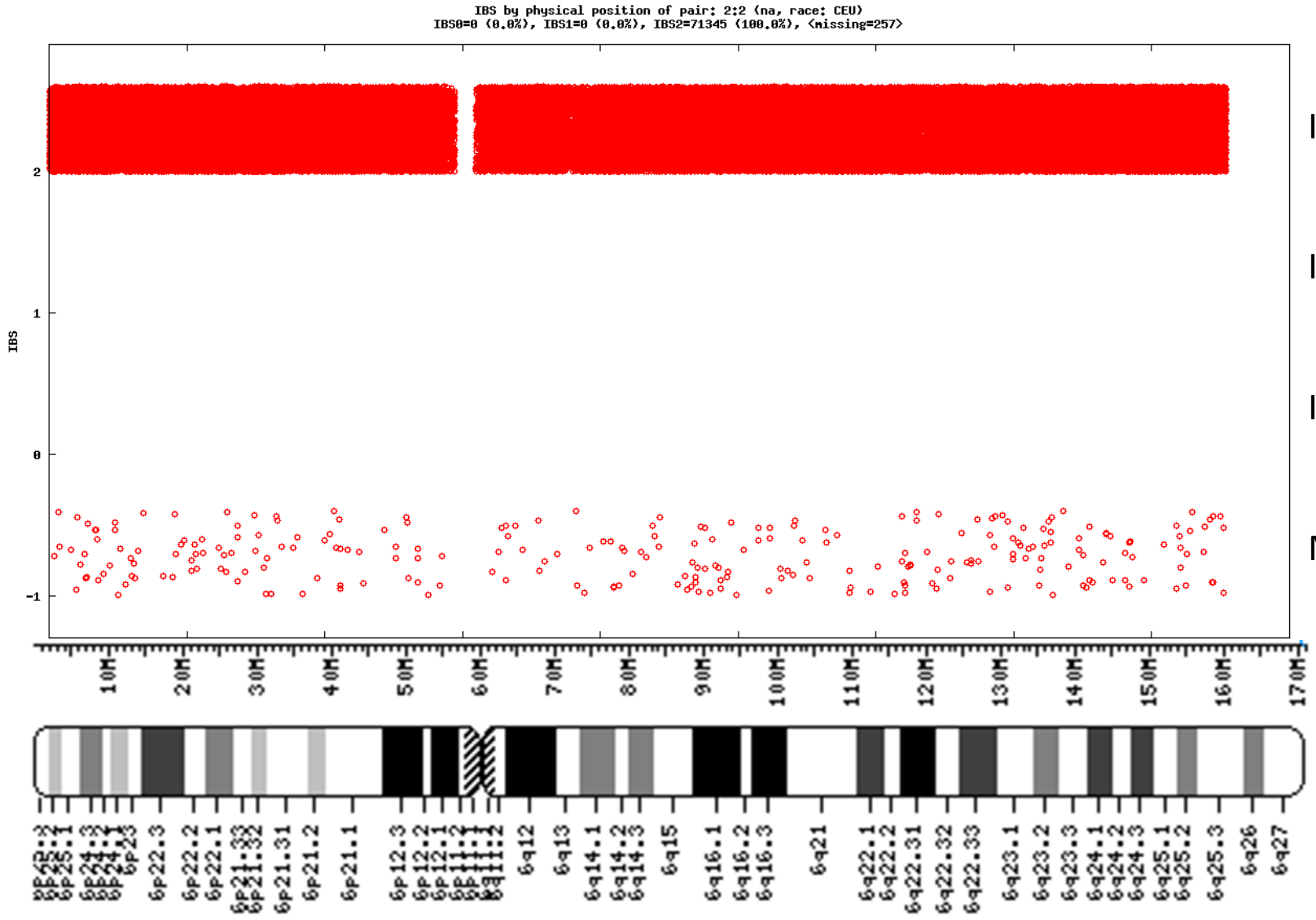
exploring chromosomal IBS patterns

Pairwise IBS patterns on Chromosome 6 in 1,031 cases data (HN)



exploring chromosomal IBS patterns

IBS patterns on Chromosome 6, a **self-pairing**, NA11891 (male, CEU)



IBS2 (71,345)

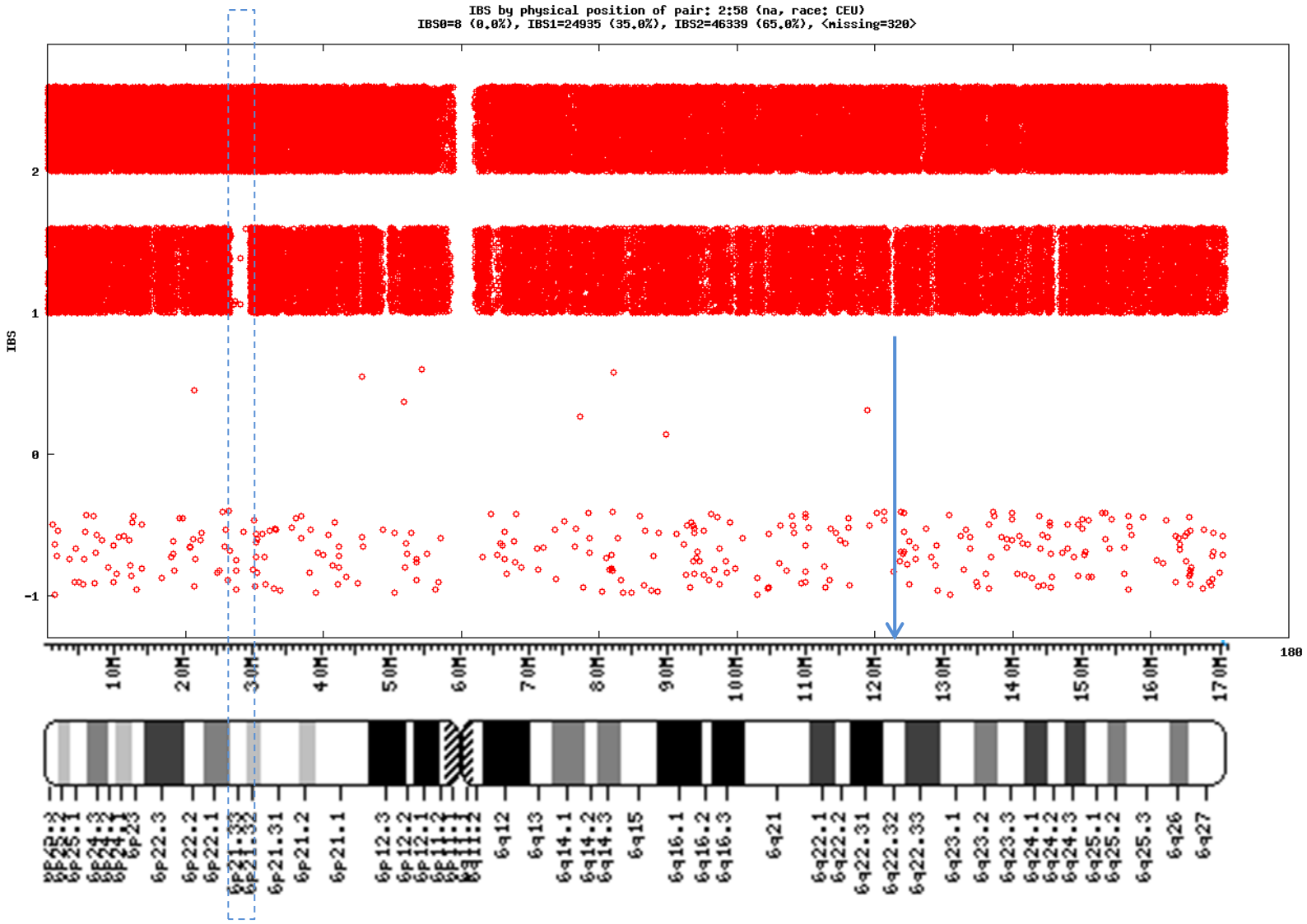
IBS1 (0)

IBS0 (0)

Missing (257)

exploring chromosomal IBS patterns

IBS patterns on Chromosome 6, **a father-son pairing**, NA11891-NA10865 (male, CEU)



IBS2 (46,339)

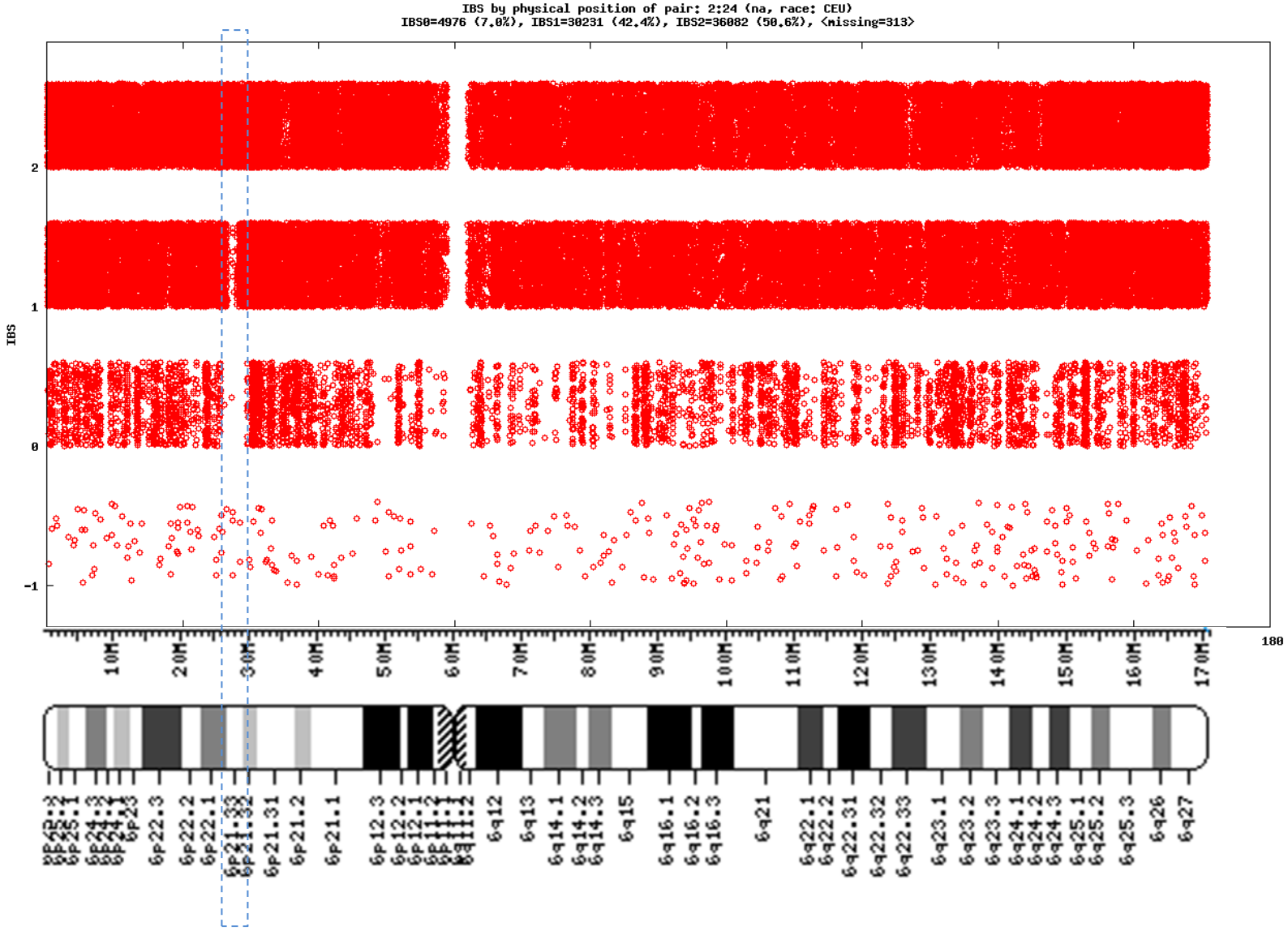
IBS1 (24,935)

IBS0 (8)

Missing (329)

exploring chromosomal IBS patterns

Total IBS patterns on Chromosome 6, a **husband-wife pairing**, NA11891 (male, CEU) - NA11892 (female, CEU)



IBS2 (36,082)

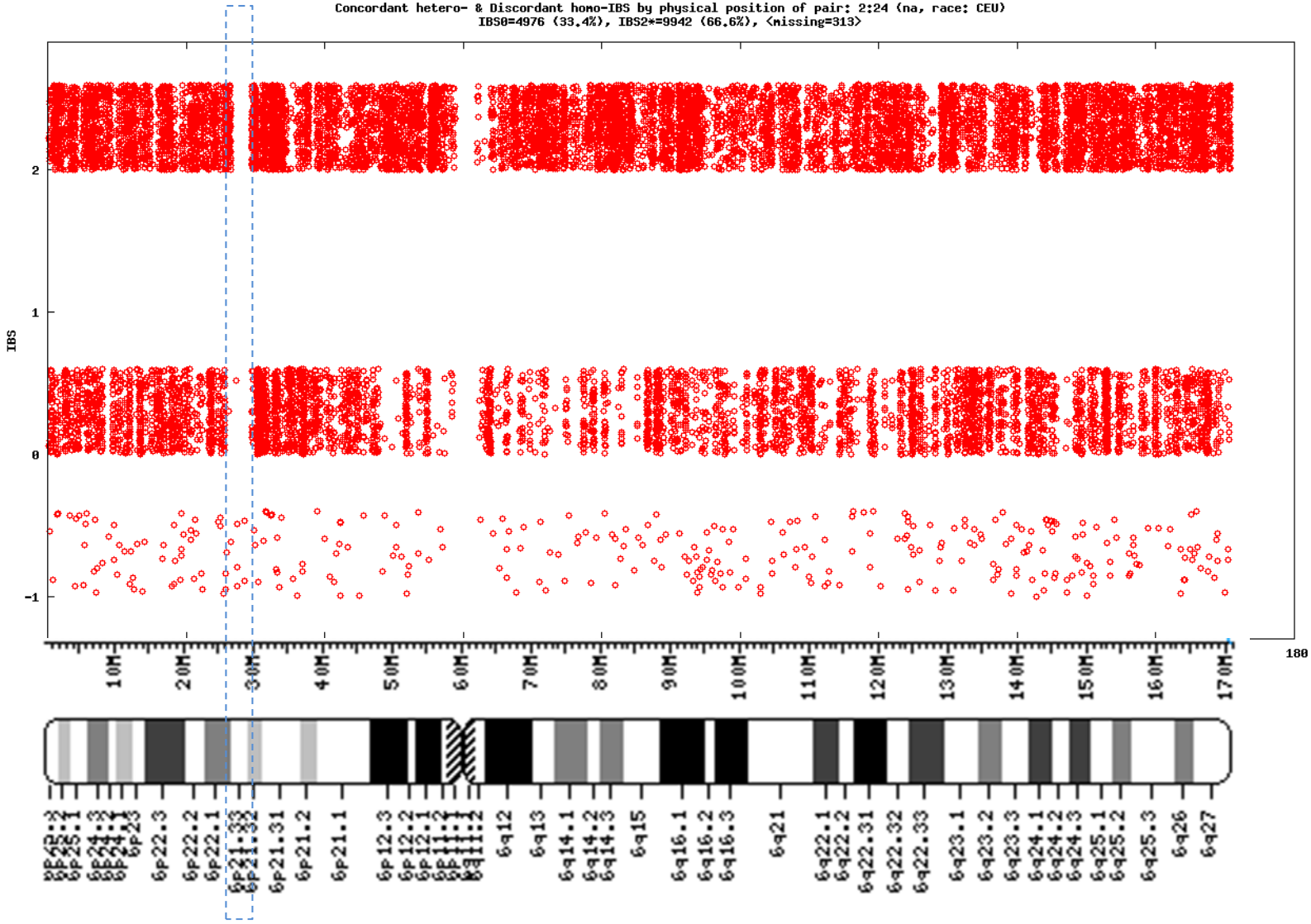
IBS1 (30,231)

IBS0 (4,976)

Missing (313)

exploring chromosomal IBS patterns

Concord het & Discord homo IBS on Chr. 6, a **husband-wife pairing**, NA11891 (male, CEU) - NA11892 (female, CEU)



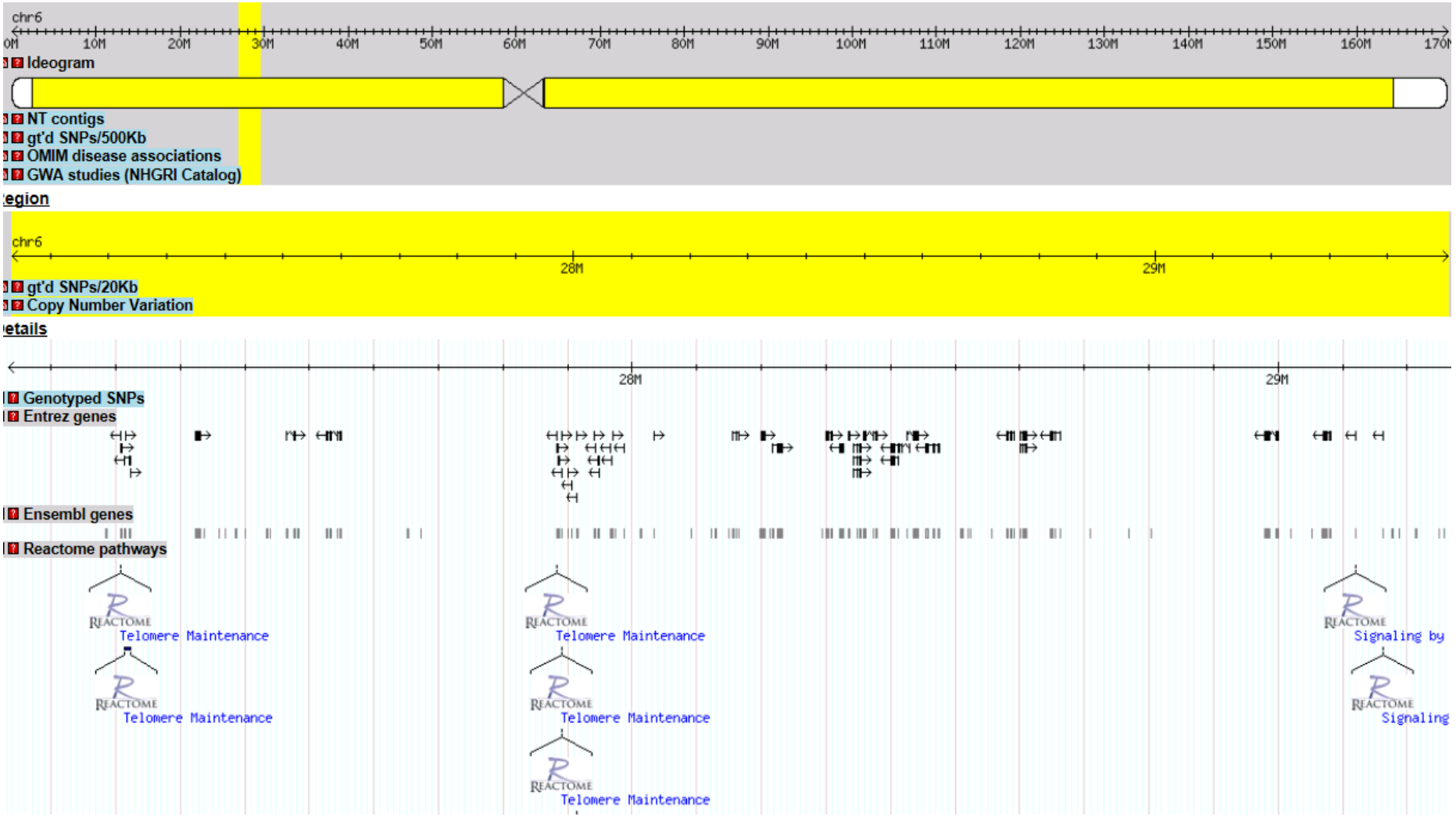
IBS2* (hetero)
66.6%(9,942)

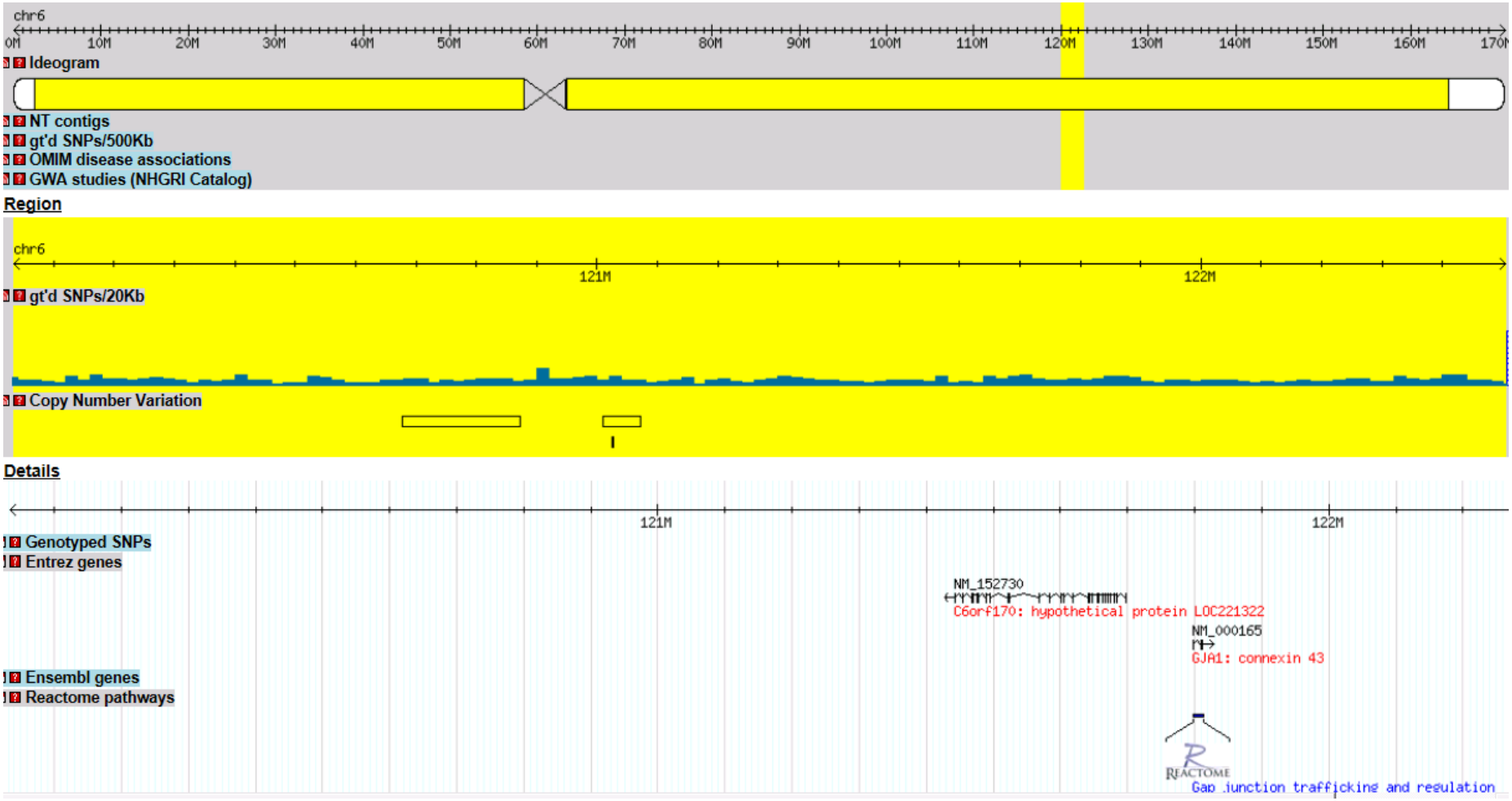
IBS1 (na)

IBS0 (homo)
33.4%(4,976)

Missing (313)

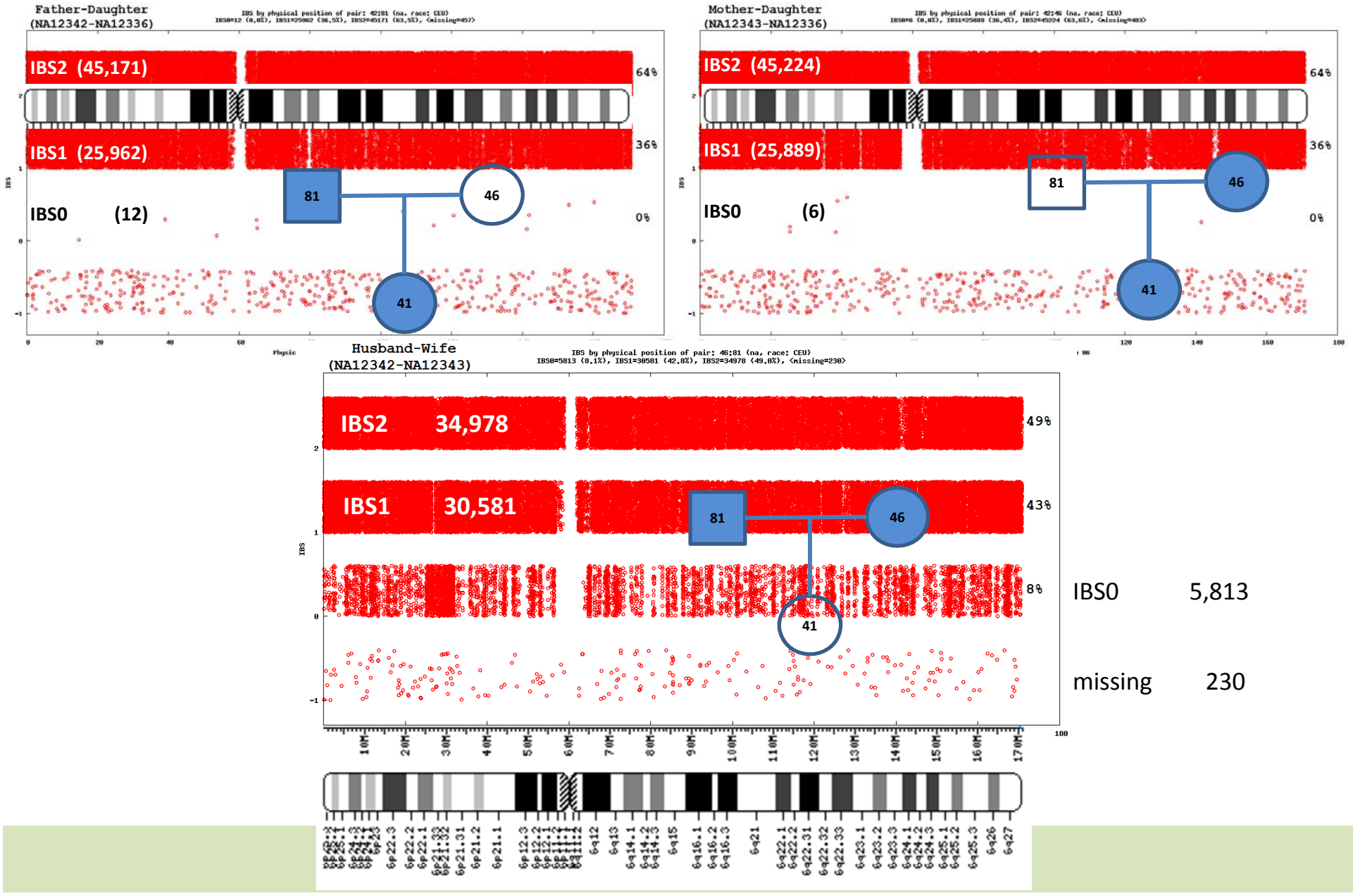
IBS2*/IBS0=2


chr6
0M
10M
20M
30M
40M
50M
60M
70M
80M
90M
100M
110M
120M
130M
140M
150M
160M
170M
Ideogram
NT contigs
gt'd SNPs/500Kb
OMIM disease associations
GWA studies (NHGRI Catalog)
egion
chr6
28M
29M
gt'd SNPs/20Kb
Copy Number Variation
etails
28M
29M
Genotyped SNPs
Entrez genes
Ensembl genes
Reactome pathways
REACTOME
Telomere Maintenance
REACTOME
Telomere Maintenance
REACTOME
Telomere Maintenance
REACTOME
Telomere Maintenance
REACTOME
Telomere Maintenance
REACTOME
Signaling by
REACTOME
Signaling


chr6
0M
10M
20M
30M
40M
50M
60M
70M
80M
90M
100M
110M
120M
130M
140M
150M
160M
170M
Ideogram
NT contigs
gt'd SNPs/500Kb
OMIM disease associations
GWA studies (NHGRI Catalog)
Region
chr6
121M
122M
gt'd SNPs/20Kb
Copy Number Variation
Details
121M
122M
Genotyped SNPs
Entrez genes
NM_152730
C6orf170: hypothetical protein LOC221322
NM_000165
GJA1: connexin 43
Ensembl genes
Reactome pathways
REACTOME
Gap junction trafficking and regulation

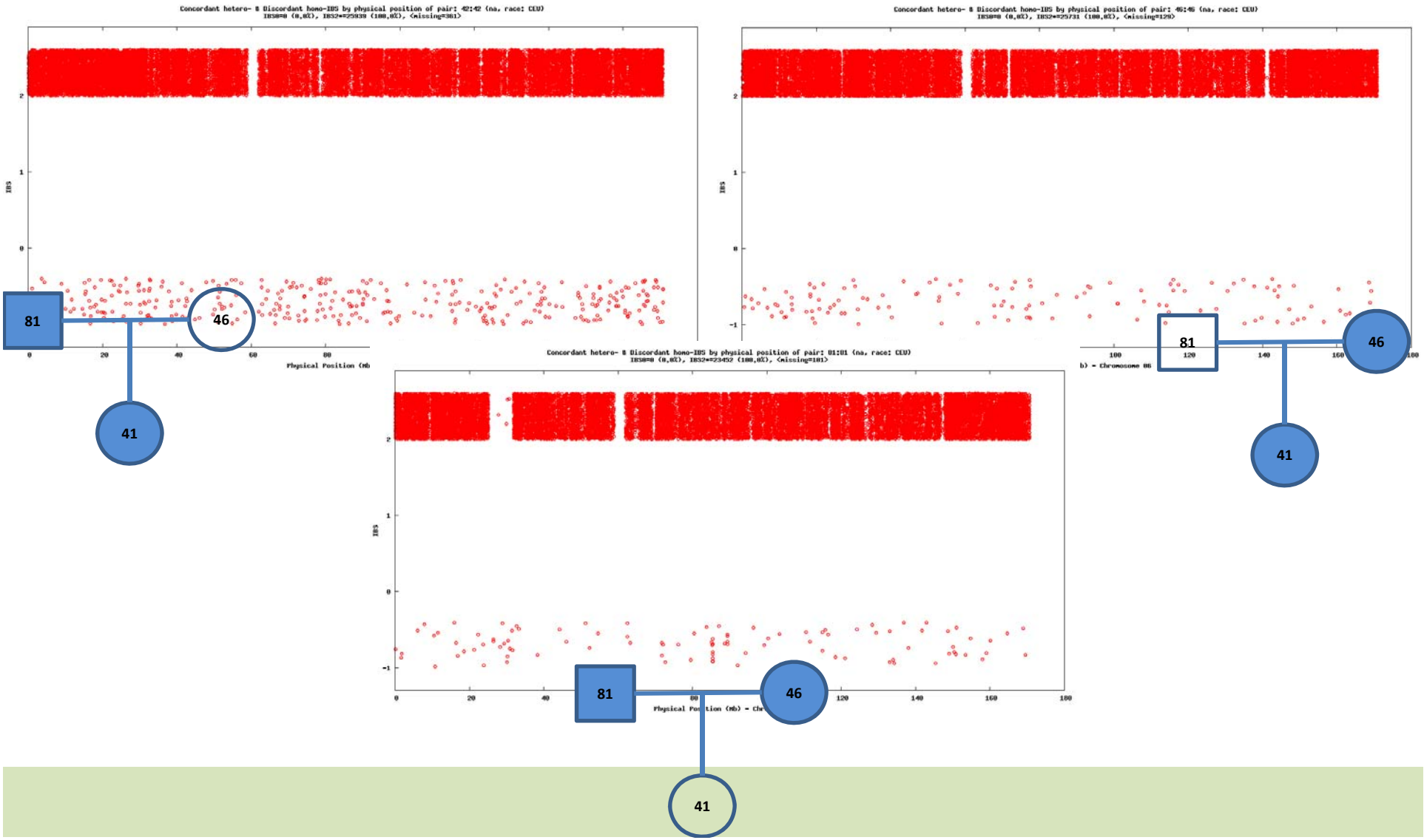
exploring chromosomal IBS patterns

Total IBS patterns on Chromosome 6 in a CEU trio(from HapMap3)



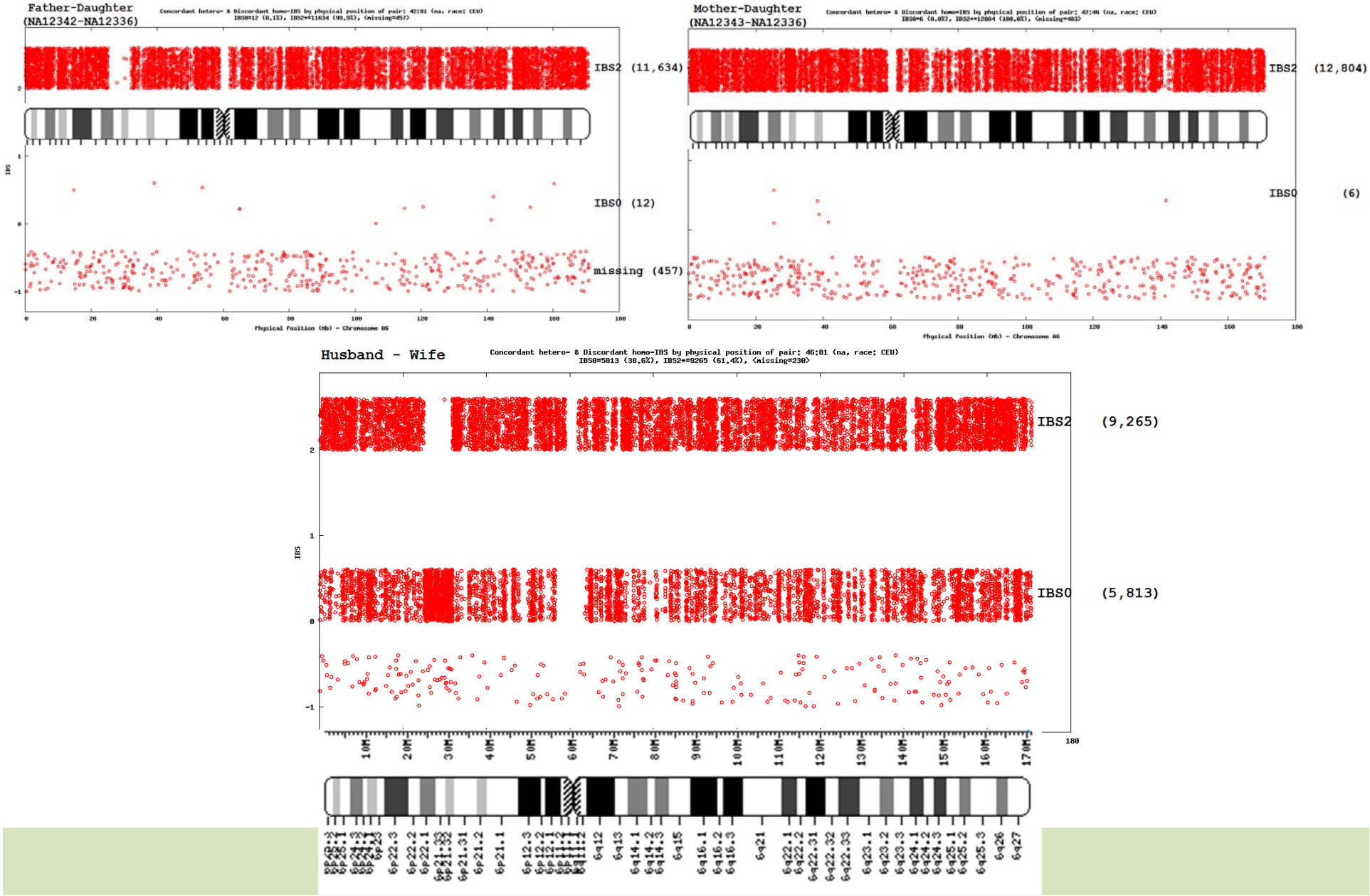
exploring chromosomal IBS patterns

A Self-pairing: Total IBS patterns on Chromosome 6 in a CEU trio (from HapMap3)



exploring chromosomal IBS patterns

Concordant heterozygotes and **Discordant homozygotes** IBS patterns on Chromosome 6 in a CEU trio (from HapMap3)



Glossary

Coalescence theory

A population genetics model of inheritance relationships among alleles at a given locus. The coalescence of two alleles is the most recent point (going back in time) at which they shared a common ancestor.

Cryptic relatedness

The presence of close relatives in a sample of ostensibly unrelated individuals. It is characterized by a recent common ancestry that can be revealed from marker-based relatedness coefficients.

Genome-wide association study

Analysis of the entire genome using association models to identify regions of the genome that contribute to genetic variation in a phenotype. These studies typically analyse data from high-density SNP arrays.

Heritability

The proportion of phenotypic variation in a population that is attributable to genetic variation among individuals. Statistical methods are used to estimate the relative contributions of differences in genetic and non-genetic factors to the total phenotypic variation in a population.

Identity by descent

(IBD). Two or more alleles are IBD if they are identical copies of the same ancestral allele in a base population. IBD can be estimated for alleles at single loci in a diploid individual or between individuals.

Identity by state

(IBS). Refers to two or more alleles that 'look' the same. For example, if two individuals both carry a 'G' allele at a specific locus.

Pedigree

A population of individuals in which the mating records for multiple generations are known. Pedigree information is typically available for livestock populations, in which controlled breeding has been implemented to maximize the response to genetic selection.

Reference Powell, JE, Visscher, PM, and Goddard, ME, Nature Reviews | GENETICS, vol 11, Nov. 2010, pp800-5