.

A way to analyse GWAS data using exons

daniel.rovera@gmail.com supervised by chloe-agathe.azencott@mines-paristech.fr
Institut Curie and Mines Paris Tech, May 27, 2021

Abstract

Genome Wide Association Studies allow to analyse the link between frequency of single nucleotide polymorphism and phenotype by comparing to a reference population. The final goal is to find relevant biomarkers to predict predisposition to disease.

The first step of this analyse is evaluating how the p-values of SNPs resulting from comparison are transferred to genes. The most of methods consider genes in totality. In the reality of the genome, genes are divided into exons and are built by a mechanism where splicing is a key step.

Here method is based on the base position of exons and the statistical relation between positions of SNPs and positions of exons. So a continuous function is established, a function of distance from SNPs to exons, which gives the p-values transferred to genes.

Introduction

The aim of here study is to find a continuous function giving p-values of genes from p-values of SNPs typifying a phenotype using GWAS data. To carry to the end, the used principles are:

- the linear consistency of the genome, base positions as reference coordinates ;
- the mechanism of gene expression which transforms precursor messenger RNA (pre-mRNA) transcript into a mature messenger RNA (mRNA).

The here process is different from usually used softwares such as VEGAS2 [6].

The Genome

The elements of human genome on which are focused are the coding sequences (a tiny part less than 2%) and, between them, the introns, interfering in gene expression by splicing (about 30% of genome). Are not taken in considération: regulatory DNA sequences, repetitive DNA sequences and mobile genetic elements.

Base position of exons are extracted by chromosomes from ***ncbi.nlm.nih.gov***(acces by [11]). The computed number end length are globally:

- 20,108 Genes divided in 1 to 190 exons (repartition fig 1)
- 193,532 Exons covering 1,1 % of genome
- Exons and Introns covering 34 % of genome

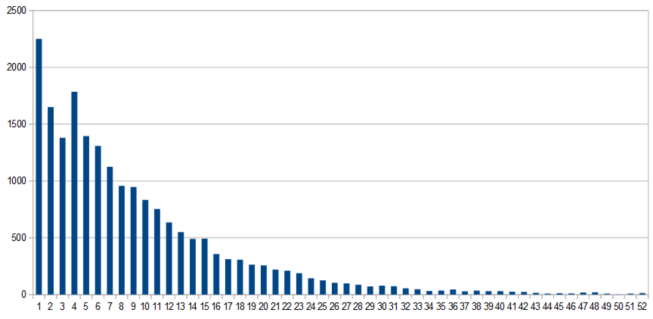

Figure 1: Number of exons by genes, until 52 exons

Cellular machinery builds proteins from messenger RNA. A primary transcript of DNA is transformed in mRNA by removing introns between exons and binding exons. This process is splicing. An exon can be used to make several genes.

Several works show the effect of punctual mutations on the mechanism of splicing as [5] [7] and [1]. Here the analysis is not about the detailed mechanism but is based on statistics about relative position of mutations and genes.

Single nucleotide polymorphism

Single Nucleotide Polymorphism frequency between two populations is evaluated by Genome Wide Association Studies. A particular genotype (disease, susceptibility ..) is compared to a reference population.

SNPs are punctual and some studies are about repartition of SNPs along the genome [3] and [4], but they do not approach the other components of genome.

The relative position of SNPs in respect of exons are listed, so the direct effects by SNPs on genes are of different types according to their positions:

1. inside an exon: effect on thin structure of the gene;
2. between 2 exons of the same gene: effect on global structure of the gene by disturbance of splicing;

3. between 2 exons of 2 different genes and included in the 2 genes: effect on structure of 2 genes by splicing;

4. between 2 exons of 2 genes and included only in 1 gene: effect on structure of the gene by splicing, no effect on the other;

5. between 2 exons and outside any gene: no effect;

6. before the first exon and after the last exon: no effect.

The effect of SNPs to gene is different according to the relative position, confirmed by some publication of effect by SNPs on splicing ([5] [7] and [1]). For the case 'no direct effect', other effects are possible by the regulation of expression by regulatory DNA sequences.

Statistical Results

Here, the SNPs data are from GWAS about predisposition of breast cancer in abbreviated form BCAC [9]. The 194,957 SNPs are distributed thus according to their relative position in respect of exons and genes:

Position	Part	$\frac{Number}{Length}$ *
inside exons	1.2%	72.9
inside genes outside exons	37.2%	72.2
outside genes	61.6%	60.2

* *unit: number by million base positions*

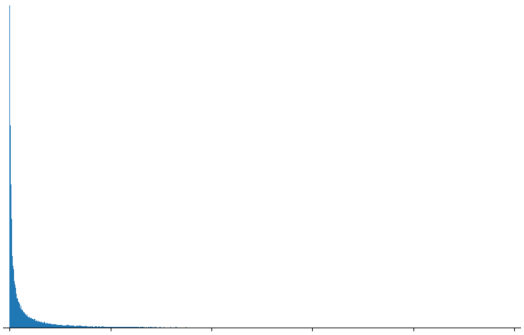

Figure 2: Histogram showing the variation of the number of SNPs by bin of minimal distance from SNPs to exons

The average number by length of SNPs is not uniform in the genome. It is different according to position and is specified by the histogram 2.

The proximity of exons plays a role in its variations: more SNPs are near exons, more they are numerous. So, the chosen variable for evaluating these variations is the minimal distance from SNPs to exons. Exons are considered as anchors in the genome.

The profile of this relation is specified by drawing the normalised cumulative number of SNPs in function of the minimal distance from SNPs to exons (fig 3). Obviously, these inside genes are nearer to exons than these outside genes.

Result from these observations

This distribution can be fitted by a γ distribution (app About γ distribution) whose probability density function (pdf) is:

$$\gamma.pdf(x, shape, scale) = \frac{1}{\Gamma(shape) * scale^{shape}} * x^{shape\ -\ 1} * e^{-\frac{x}{scale}}$$

The fitted curve is the cumulative density function (cdf) of γ law (fig 4) .

This distribution can be interpreted as a heterogeneous spatial Poisson process. The intensity λ, average number of points by length, depends on the location. It is shown (app Intensity of inhomogeneous Poisson process):

$$\lambda(d) = \gamma.pdf(d).$$

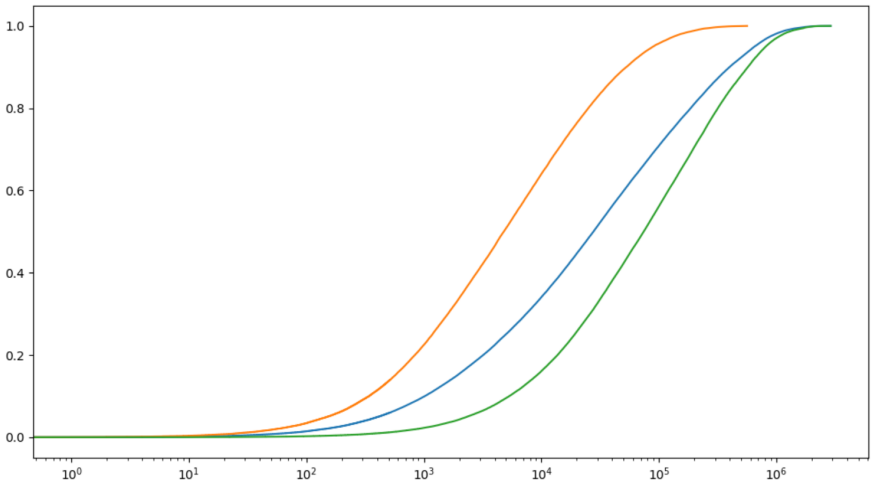

Figure 3: Position of SNPs relative to exons (orange in genes, green out genes, blue all), normalized cumulative number in function of minimal distance of SNPs to exons, logarithmic scale on the abscissa

The intensity λ is the parameter of Poisson law which is not constant contrary to the well known law. This can be pictured by cars driving along a straight line where there are villages which slow the traffic (traffic light, speed limit). Exons are villages, SNPs are cars at a moment and intensity in the instantaneous car flow.

The fitting normalized cumulative number in function of minimal distance of SNPs to exons (fig 4) gives these values of parameters: shape = 0.454746696, scale = 2,9702.7070. Only the distances of SNPs whose p-values are not outlier (app Outliers p-values) inside genes are taken into account.

Shape is the critical parameter for the link between intensity and distance (confer γ distribution for different values of shape [10]). Its value is less than one and so the curve shows a hyperbolic form. It leads to notice accumulation of SNPs near exons as the histogram 2 shows.

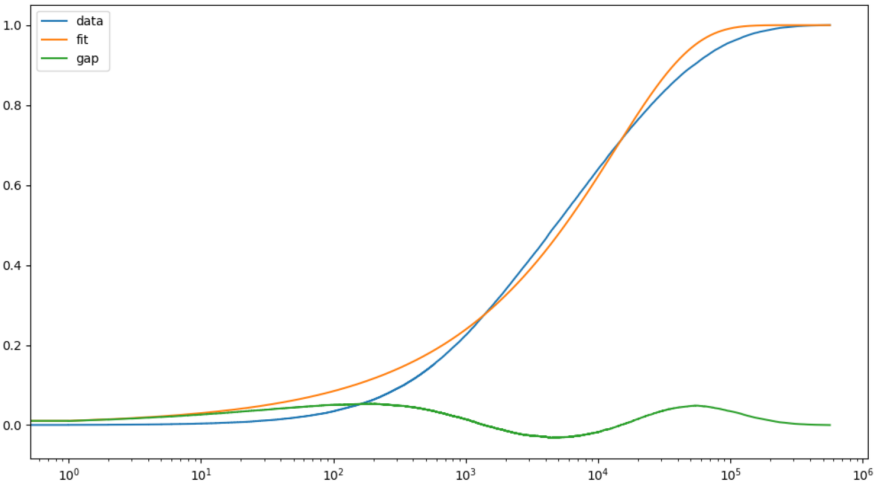

Figure 4: fitting normalized cumulative number in function of minimal distance of SNPs to exons, , logarithmic scale on the abscissa

First Results and Choices to Taking Account P-Values

These observations lead to these intermediary results. The effects by SNPs on genes are roughly identified. SNPs are located preferentially inside genes and near exons according to a probability law. these considerations suggest that effect of SNPs decreases when the distances from SNPs to

exons increases.

From these intermediary results, the elements defining the quantitative link between SNPs and genes are chosen:

- only the direct effect is kept for SNPs inside genes, the indirect régulation is neglected
- the scope of the effect by SNPs is limited to the neighboring exons and does not jump over exon
- the resulting p-values of genes are computed by the Z-score method
- the weights used in Z-score formula are the intensity

Stouffer's Z-score method ([13]) (Z-score is the number of standard deviations between value and the mean value):

$$Z = \frac{\sum_i w_i * z_i}{\sqrt{\sum_i w_i^2 + 2 * \sum_{i<j} r_{ij} * w_i * w_j}}$$

The Z-score of every gene is computed from the Z-score of SNP_i in the scope. The weights w_i are the intensity at the distance between SNP_i and exons of this gene. They are one or two according to the case because the effect of SNPs does not jump over genes.

The simple formula of Z-score does not containt the term $2 * \sum_{i<j} r_{ij} * w_i * w_j$. This formula is commonly used in meta analysis. But in this case this term must be taken in account. Indeed r_{ij} is the correlation coefficient between the data used for computing the p-values of SNP_i and SNP_j. The SNPs contribute to the same phenotype, so the correlation coefficient is not null and must be evaluated.

Computing every r_{ij} is out of reach. Only one value is assigned to the correlation coefficient r_{ij}. It is computed by linear adjustment of log(p-values) and log(cumulative number), what is justified by the beta uniform mixture model (app Correlation and distribution of p-values).
Its value is 0.8478 (fig 5).

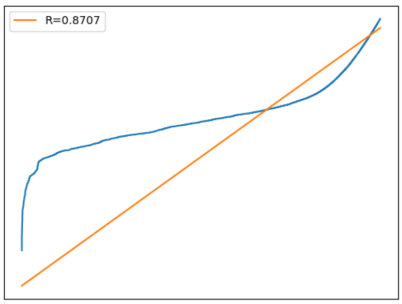

Figure 5: Linear log log fitting of normalized cumulative number of SNPs and their p-values

To evaluate the effect of SNPs inside exons, the weights are taken infinite, which cancels the effect of any SNPs outside exons (app Formula of Z-score). The effects of SNPs outside exons and inside genes are computed separately.

Biological Results

So the p-values of genes resulting from effect of BCAC are computed. Three cases are distinguished: SNPs inside exons ('in'), SNPs inside genes and outside exons and SNPs ('out') and outside genes which are considered having no effect.

It is to highlight that one gene can have two transferred p-values by the two effects in cases 'in' and 'out'. The Z-scores from these two effects are not added except for sorting.

The result of computing according to here method gives:

- the classical file containing chromosomes, genes, begin and end of genes in base positions, resulting Z-score, associated p-value and a flag for the two cases (a gene can be present twice);
- three lists of genes sorted by decreasing computed Z-scores in case 'in', in case 'out' and the artificial sums of Z-score in the two case; these lists show the decomposition of effect by SNPs;
- similarly this decomposition can be obtained for a list of genes
- two Manhattan Plots for same two cases, fig 6 and fig 7

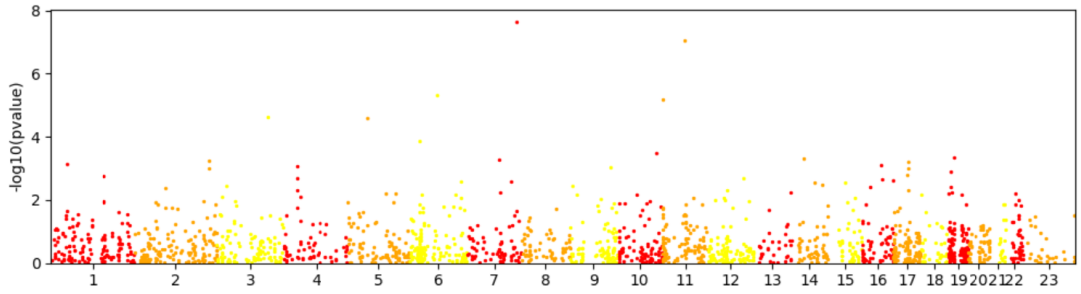

Figure 6: Manhattan plot for SNPs inside exons

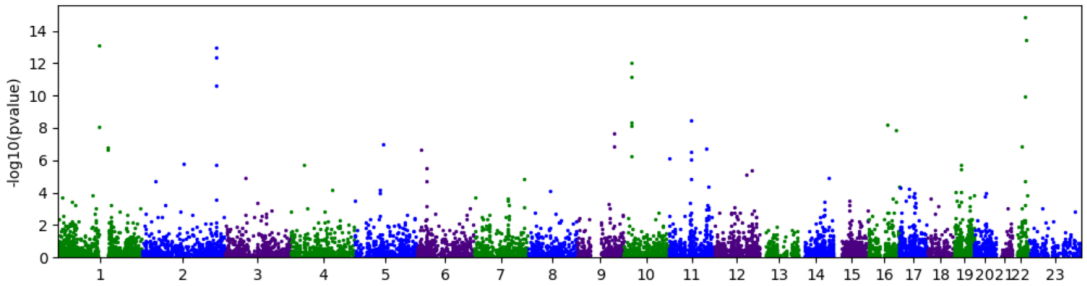

Figure 7: Manhattan plot for SNPs outside exons and inside genes

the inter-classification of these lists in cases 'in' and 'out' must be confusing, the modes of action being different.

The contributions of every SNPs are detailed. It is the product of intensity from distance SNPs to exons and Z-scores of SNPs divided by the denominator computed for exons with effect on the gene. This detailed result shows how a set of SNPs can reinforce its effects.

To illustrate this analysis, the array Result of two effects shows the top 10 of genes sorted by the sum of Z-score. Among the first three, EHBP1L1 and TNS1 are compared.

Their functions by [12] are: EHBP1L1 acts as Rab effector protein and play a role in vesicle trafficking and TNS1 is involved in fibrillar adhesion formation, cell migration, cartilage development and in linking signal transduction pathways to the cytoskeleton.

TNS1 is influenced by many SNPs, which reinforces inklink of role by TNS1 in breast cancer, taking account uncertainty about p-values. In view of their respective functions, this interpretation is plausible and must be confirmed by other analysis.

Discussion

in comparison, the order of genes by decreasing p-value is very far from this got by VEGAS2 [6].

This method has some avantages. The Z-scores or p-values of genes are computed from Z-scores or p-values of SNPs by a simple continuous function. All parameters are computed from available data been rid of outlier values. No hyper-parameter is necessary.

But it ignores some important phenomenons. The indirect mechanisms of regulation by regulatory DNA are neglected. SNPs in these areas probably disturb the expression of genes. Here method is based on the linear distances in the genome as 23 segments and so it obscures the spatial storage of folded DNA.

The effect on every SNPs on genes can be modeled as a network which can be merged in a gene gene interaction network. This approach must show how some SNPs can strengthen their effect.

Here method is another way to explore the relationships between SNPs and genes and must be usable for other GWAS data.

Appendices

Outliers p-values

Python cannot compute the Z-score below $5.55 * 10^{-17}$. So the p-value between $6.7 * 10^{-133}$ and $3.7 * 10^{-17}$ can be taken in account. They are considered as outlier values. Other outlier values are also eliminated in tails by keeping only 99% of p-values. The fork becomes $[1.38 * 10^{-15}, 0.99]$.

Intensity of inhomogeneous Poisson process

The inhomogeneous Poisson process is a counting process n(t) in a time interval with:

$$E(n(t+h) - n(t)) = \int_t^{t+h} \lambda(\tau) d\tau$$

This expression with h $\to$ 0 is brought closer to the law from observation (N number of SNPs):

$$\frac{dn(t)}{dt} = \frac{1}{N} * \gamma.pdf(t)$$

Considering t as the minimal distances from SNPs to exons, the expression of λ is:

$$\lambda(t) = \frac{1}{N} * \gamma.pdf(t)$$

This function is used only as weight, the term $\frac{1}{N}$ can be removed without changing the result.

Correlation and distribution of p-values

The distribution of p-values can be described by a $\beta(a, 1)$ law to which the noise is added (beta uniform mixture model [8]). The expression of β is:

$$\beta.pdf(x) = a * x^{a-1} \; and \; \beta.cdf(x) = x^a$$

The formula with noise λ is:

$$bum.pdf(x) = \lambda + (1 - \lambda) * a * x^{a-1}$$

The noise ruins the correlation between p-values. So, the linear fitting log-log on cumulative normalized values give the correlation coefficient between p-values.

Formula of Z-score

The formula of Z-score method for n Z-score to be weighted with R as the common correlation coefficient:

$$Z = \frac{\sum_i w_i * z_i}{\sqrt{\sum_i w_i^2 + 2R * \sum_{i<j} w_i * \ w_j}}$$

When the weights are infinity, it becomes, :

$$Z = \frac{\sum_i w_i * z_i}{\sqrt{n + R * n * (n \ - \ 1)}}$$

About γ distribution

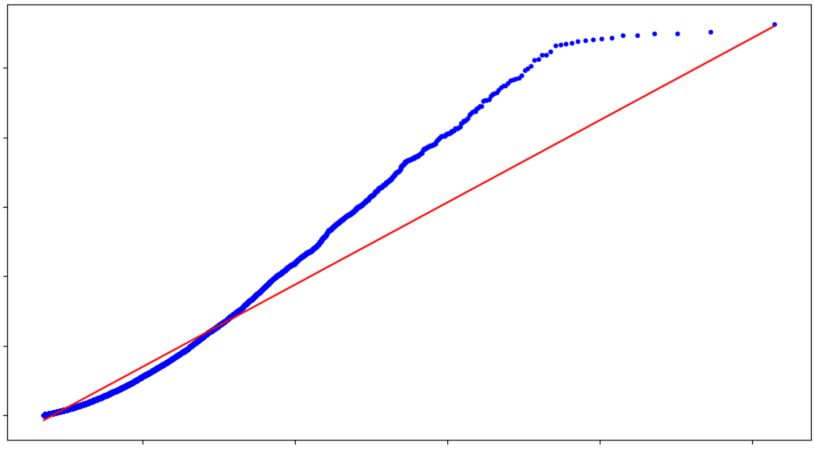

Figure 8: Q-Q plot minimal distance from SNPs inside genes to exon - γ distribution

A Q-Q plot precises how the minimal distance from SNPs inside genes to exons follows a γ distribution with shape = 0.454746696, scale = 2,9702.7070 (fig 8).

Finding a γ distribution is not surprising, it is common in nature [2]. γ distribution is the maximum entropy distribution for a positive variable, given its mean value and the mean value of its logarithm.

Even distances between successive SNPs follow a γ distribution with shape = 0.502497962 and scale = 21,513.9128 (fig 9).

In short, SNPs tend to agglomerate preferably near exons.

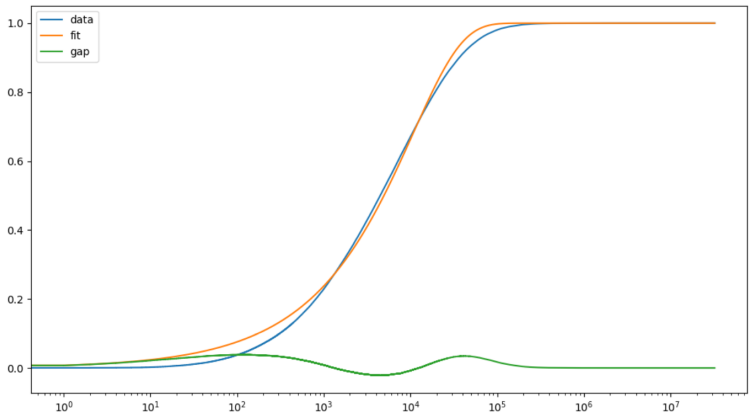

Figure 9: Normalized cumulative number of SNPs in function of distances between SNPs

Result of two effects

Genes are sorted artificially by added Z-scores. Internal effect is in first and splicing effect in second

Only internal effect below

R=0.8478	EHBP1L1	PIDD1	TNS1	IBTK	NEK10	COL1A2	ANKRD55	FAM114A1	CYP4B1	COPG2	
gene Z	5.2715	4.3635	2.5289	4.4305	2.7065	3.2741	4.0465	3.128	3.1747	1.5301	SNP_Z
rs17833842	0	0	0.9041	0	0	0	0	0	0	0	5.9016
rs3903072	1.8958	0	0	0	0	0	0	0	0	0	5.3912
rs7114014	1.8017	0	0	0	0	0	0	0	0	0	5.1234
rs11227311	1.574	0	0	0	0	0	0	0	0	0	4.4762
rs1124000	0	0	0	4.4305	0	0	0	0	0	0	4.4305
rs10902221	0	4.3635	0	0	0	0	0	0	0	0	4.3635
rs78394680	0	0	0	0	1.4731	0	0	0	0	0	4.1892
rs832574	0	0	0	0	0	0	4.0465	0	0	0	4.0465
rs2542204	0	0	0.5106	0	0	0	0	0	0	0	3.3327
rs10281073	0	0	0	0	0	3.2741	0	0	0	0	3.2741
rs5013330	0	0	0	0	0	0	0	0	3.1747	0	3.1747
rs721653	0	0	0	0	0	0	0	3.128	0	0	3.128
rs1921987	0	0	0.4409	0	0	0	0	0	0	0	2.8782
rs1882425	0	0	0.3814	0	0	0	0	0	0	0	2.4893
rs9874914	0	0	0	0	0.6953	0	0	0	0	0	1.9774
rs115091469	0	0	0	0	0.538	0	0	0	0	0	1.5301
rs157893	0	0	0	0	0	0	0	0	0	1.5301	1.5301
rs116301296	0	0	0.1655	0	0	0	0	0	0	0	1.0803
rs78876460	0	0	0.0632	0	0	0	0	0	0	0	0.4125
rs62179537	0	0	0.0632	0	0	0	0	0	0	0	0.4125

Only splicing effect below

R=0.8478	EHBP1L1	PIDD1	TNS1	IBTK	NEK10	COL1A2	ANKRD55	FAM114A1	CYP4B1	COPG2	
gene Z	6.1848	5.5271	7.3453	5.1015	5.0879	3.7957	2.8481	3.5583	3.4848	4.9103	SNP_Z
rs13399995	0	0	0.149	0	0	0	0	0	0	0	7.8865
rs6721811	0	0	0.181	0	0	0	0	0	0	0	7.8482
rs6705818	0	0	0.1084	0	0	0	0	0	0	0	7.8418
rs72759712	0	0	0	0	0	0	0.0323	0	0	0	7.8298
rs13411492	0	0	0.0075	0	0	0	0	0	0	0	7.8189
rs6735298	0	0	0.2765	0	0	0	0	0	0	0	7.7866
rs10199394	0	0	0.1019	0	0	0	0	0	0	0	7.7787
rs11709858	0	0	0	0	0.0255	0	0	0	0	0	7.7712
rs13426946	0	0	0.0077	0	0	0	0	0	0	0	7.7422
rs1882419	0	0	0.0181	0	0	0	0	0	0	0	7.7006
rs34909633	0	0	0.1356	0	0	0	0	0	0	0	7.6794
rs11915704	0	0	0	0	0.2538	0	0	0	0	0	7.6776
rs17778299	0	0	0.0784	0	0	0	0	0	0	0	7.6741
rs11712293	0	0	0	0	0.0564	0	0	0	0	0	7.6741
rs12465515	0	0	0.1295	0	0	0	0	0	0	0	7.6708
rs10932689	0	0	0.0143	0	0	0	0	0	0	0	7.6691
rs62255172	0	0	0	0	0.0839	0	0	0	0	0	7.6659
rs1402974	0	0	0	0	0.0723	0	0	0	0	0	7.6596
rs2272075	0	0	0	0	0.059	0	0	0	0	0	7.6536
rs9857897	0	0	0	0	0.1801	0	0	0	0	0	7.6536
rs12655019	0	0	0	0	0	0	0.0406	0	0	0	7.6507
rs10932690	0	0	0.0142	0	0	0	0	0	0	0	7.6385
rs11709516	0	0	0	0	0.1028	0	0	0	0	0	7.6273
rs7595393	0	0	0.0162	0	0	0	0	0	0	0	7.5901
rs9870969	0	0	0	0	0.1375	0	0	0	0	0	7.5822
rs7600279	0	0	0.0154	0	0	0	0	0	0	0	7.5611
rs10195963	0	0	0.0143	0	0	0	0	0	0	0	7.5611
rs10192415	0	0	0.0156	0	0	0	0	0	0	0	7.5269
rs4571035	0	0	0.0144	0	0	0	0	0	0	0	7.488
rs12495557	0	0	0	0	0.6121	0	0	0	0	0	7.4807
rs13063368	0	0	0	0	0.182	0	0	0	0	0	7.4609
rs6708579	0	0	0.0332	0	0	0	0	0	0	0	7.4492
rs1522131	0	0	0	0	0.0502	0	0	0	0	0	7.4464
rs6735174	0	0	0.1648	0	0	0	0	0	0	0	7.4384
rs11917805	0	0	0	0	0.1272	0	0	0	0	0	7.4358
rs6723013	0	0	0.0347	0	0	0	0	0	0	0	7.4147
rs16886496	0	0	0	0	0	0	0.0338	0	0	0	7.3658
rs7572103	0	0	0.0595	0	0	0	0	0	0	0	7.3599
rs73053821	0	0	0	0	0.1574	0	0	0	0	0	7.2775
rs59298018	0	0	0.0924	0	0	0	0	0	0	0	7.2625
rs13417588	0	0	0.763	0	0	0	0	0	0	0	7.2556
rs4621152	0	0	0.0457	0	0	0	0	0	0	0	7.2366
rs17833945	0	0	0.1281	0	0	0	0	0	0	0	7.2308
rs76973702	0	0	0.125	0	0	0	0	0	0	0	7.2253
rs6435958	0	0	0.0839	0	0	0	0	0	0	0	7.2253
rs17434131	0	0	0	0	0.1773	0	0	0	0	0	7.2005
rs13392238	0	0	0.4023	0	0	0	0	0	0	0	7.1682
rs62174839	0	0	0.1074	0	0	0	0	0	0	0	7.145
rs10191184	0	0	0.0146	0	0	0	0	0	0	0	7.142
rs10183545	0	0	0.0489	0	0	0	0	0	0	0	7.1199
rs4583440	0	0	0.0864	0	0	0	0	0	0	0	7.0943
rs55999136	0	0	0.2369	0	0	0	0	0	0	0	7.0901
rs10211289	0	0	0.1083	0	0	0	0	0	0	0	7.0571
rs13022815	0	0	0.0081	0	0	0	0	0	0	0	7.0539
rs2372938	0	0	0.2757	0	0	0	0	0	0	0	7.0212
rs12478570	0	0	0.0076	0	0	0	0	0	0	0	7.009
rs4491709	0	0	0.149	0	0	0	0	0	0	0	6.9777
rs12329133	0	0	0.0084	0	0	0	0	0	0	0	6.9686
rs28369659	0	0	0.0069	0	0	0	0	0	0	0	6.9372
rs4674132	0	0	0.0072	0	0	0	0	0	0	0	6.8946
rs10932691	0	0	0.0452	0	0	0	0	0	0	0	6.8316
rs2888449	0	0	0.0132	0	0	0	0	0	0	0	6.8185
rs2372932	0	0	0.0127	0	0	0	0	0	0	0	6.706
rs6435957	0	0	0.0283	0	0	0	0	0	0	0	6.6676

R=0.8478	EHBP1L1	PIDD1	TNS1	IBTK	NEK10	COL1A2	ANKRD55	FAM114A1	CYP4B1	COPG2	
rs6734010	0	0	0.0078	0	0	0	0	0	0	0	6.5833
rs78282939	0	0	0.0284	0	0	0	0	0	0	0	6.4731
rs66538660	0	0	0.0717	0	0	0	0	0	0	0	6.3994
rs17835044	0	0	0.0218	0	0	0	0	0	0	0	6.3845
rs35120201	0	0	0.0246	0	0	0	0	0	0	0	6.3676
rs17778427	0	0	0.0246	0	0	0	0	0	0	0	6.3676
rs17778329	0	0	0.0829	0	0	0	0	0	0	0	6.3333
rs2372939	0	0	0.0788	0	0	0	0	0	0	0	6.3209
rs735361	0	0	0.0169	0	0	0	0	0	0	0	6.2888
rs12622764	0	0	0.0333	0	0	0	0	0	0	0	6.2793
rs10207736	0	0	0.0117	0	0	0	0	0	0	0	6.2793
rs12621130	0	0	0.0871	0	0	0	0	0	0	0	6.262
rs56372055	0	0	0.0266	0	0	0	0	0	0	0	6.254
rs12615418	0	0	0.054	0	0	0	0	0	0	0	6.2391
rs7421376	0	0	0.0483	0	0	0	0	0	0	0	6.2255
rs4315498	0	0	0.2086	0	0	0	0	0	0	0	6.2191
rs17778091	0	0	0.03	0	0	0	0	0	0	0	6.2191
rs67188612	0	0	0.1163	0	0	0	0	0	0	0	6.2129
rs58844941	0	0	0.026	0	0	0	0	0	0	0	6.1853
rs12613030	0	0	0.0113	0	0	0	0	0	0	0	6.0422
rs11957276	0	0	0	0	0	0	0.0347	0	0	0	6.0113
rs12614773	0	0	0.1099	0	0	0	0	0	0	0	5.8406
rs16856812	0	0	0.0108	0	0	0	0	0	0	0	5.8344
rs2372934	0	0	0.0141	0	0	0	0	0	0	0	5.7909
rs10932687	0	0	0.0113	0	0	0	0	0	0	0	5.7241
rs12614767	0	0	0.011	0	0	0	0	0	0	0	5.612
rs702681	0	0	0	0	0	0	0.0262	0	0	0	5.5955
rs34851859	0	0	0.0089	0	0	0	0	0	0	0	5.5804
rs33320	0	0	0	0	0	0	0.0214	0	0	0	5.5414
rs2272525	0	0	0.0118	0	0	0	0	0	0	0	5.5195
rs55844925	0	0	0.0108	0	0	0	0	0	0	0	5.4999
rs76402955	0	0	0.1116	0	0	0	0	0	0	0	5.4999
rs41521045	0	0	0.0055	0	0	0	0	0	0	0	5.474
rs62176305	0	0	0.0055	0	0	0	0	0	0	0	5.4586
rs75511623	0	0	0.0054	0	0	0	0	0	0	0	5.4586
rs4706896	0	0	0	0.4996	0	0	0	0	0	0	5.4443
rs62176307	0	0	0.0054	0	0	0	0	0	0	0	5.4376
rs1321761	0	0	0	0.3033	0	0	0	0	0	0	5.4249
rs12494536	0	0	0	0	0.025	0	0	0	0	0	5.4249
rs55791210	0	0	0.0117	0	0	0	0	0	0	0	5.4017
rs7731700	0	0	0	0	0	0	0.4365	0	0	0	5.3671
rs832538	0	0	0	0	0	0	0.0235	0	0	0	5.3231
rs832539	0	0	0	0	0	0	0.024	0	0	0	5.3196
rs72949860	0	0	0.0241	0	0	0	0	0	0	0	5.2905
rs173764	0	0	0	0	0	0	0.0242	0	0	0	5.2652
rs832567	0	0	0	0	0	0	0.0656	0	0	0	5.2453
rs832573	0	0	0	0	0	0	0.2933	0	0	0	5.2295
rs2548663	0	0	0	0	0	0	0.1457	0	0	0	5.2168
rs702689	0	0	0	0	0	0	0.1566	0	0	0	5.2031
rs13403428	0	0	0.0095	0	0	0	0	0	0	0	5.0787
rs10198784	0	0	0.0731	0	0	0	0	0	0	0	5.0508
rs62174848	0	0	0.0544	0	0	0	0	0	0	0	5.0341
rs10896050	2.805	0	0	0	0	0	0	0	0	0	5.0341
rs2372972	0	0	0.01	0	0	0	0	0	0	0	5.0263
rs62174799	0	0	0.0094	0	0	0	0	0	0	0	5.0045
rs13406843	0	0	0.1465	0	0	0	0	0	0	0	5.0045
rs62174797	0	0	0.0093	0	0	0	0	0	0	0	4.9978
rs9449340	0	0	0	0.7109	0	0	0	0	0	0	4.9978
rs13386831	0	0	0.0561	0	0	0	0	0	0	0	4.9978
rs12213217	0	0	0	0.8592	0	0	0	0	0	0	4.9559
rs13430080	0	0	0.0111	0	0	0	0	0	0	0	4.9454
rs12165173	0	0	0.0137	0	0	0	0	0	0	0	4.9258
rs9677455	0	0	0.0109	0	0	0	0	0	0	0	4.9123
rs9443924	0	0	0	0.4942	0	0	0	0	0	0	4.9038
rs9449341	0	0	0	0.9036	0	0	0	0	0	0	4.8308
rs9344191	0	0	0	0.4747	0	0	0	0	0	0	4.8034
rs17530068	0	0	0	0.5864	0	0	0	0	0	0	4.7724
rs10932688	0	0	0.0087	0	0	0	0	0	0	0	4.6575
rs11227306	3.3798	0	0	0	0	0	0	0	0	0	4.5486
rs4593472	0	0	0	0	0	0	0	0	0	0.6231	4.5127
rs72947775	0	0	0.0084	0	0	0	0	0	0	0	4.5062
rs9443923	0	0	0	0.4274	0	0	0	0	0	0	4.4305
rs6597981	0	4.5251	0	0	0	0	0	0	0	0	4.3965
rs2257505	0	0	0	0	0	0	0.0169	0	0	0	4.374
rs79680734	0	0	0	0	0	0	0.0254	0	0	0	4.3535
rs13389571	0	0	0.0118	0	0	0	0	0	0	0	4.3346
rs77549302	0	0	0.0182	0	0	0	0	0	0	0	4.2835
rs11246316	0	1.002	0	0	0	0	0	0	0	0	4.2787
rs3843337	0	0	0.0073	0	0	0	0	0	0	0	4.2436
rs75237043	0	0	0.0641	0	0	0	0	0	0	0	4.2436
rs10075381	0	0	0	0	0	0	0.2053	0	0	0	4.224
rs6973318	0	0	0	0	0	0	0	0	0	0.4398	4.1735
rs2372958	0	0	0.0089	0	0	0	0	0	0	0	4.1449
rs205739	0	0	0	0	0	0	0	0	0	0.1871	4.1318
rs205744	0	0	0	0	0	0	0	0	0	0.1825	4.1193
rs4141794	0	0	0	0	0	0	0	0	0	0.554	4.1193
rs2372957	0	0	0.0088	0	0	0	0	0	0	0	4.1193
rs16856877	0	0	0.0084	0	0	0	0	0	0	0	4.1193
rs36029265	0	0	0.0563	0	0	0	0	0	0	0	4.0854
rs863839	0	0	0	0	0	0	0.0512	0	0	0	4.0751
rs832581	0	0	0	0	0	0	0.1166	0	0	0	4.0376
rs832531	0	0	0	0	0	0	0.0415	0	0	0	4.0208
rs33330	0	0	0	0	0	0	0.0786	0	0	0	4.0051
rs12329135	0	0	0.0046	0	0	0	0	0	0	0	3.8303
rs205752	0	0	0	0	0	0	0	0	0	0.5492	3.8082
rs17688449	0	0	0	0	0	0	0	0	0	0.7865	3.7719
rs4672819	0	0	0.0038	0	0	0	0	0	0	0	3.6522
rs2372960	0	0	0.0033	0	0	0	0	0	0	0	3.6153
rs6957511	0	0	0	0	0	0	0	0	0	1.3387	3.6153
rs16856890	0	0	0.0052	0	0	0	0	0	0	0	3.5401
rs1882421	0	0	0.0222	0	0	0	0	0	0	0	3.5401
rs17317135	0	0	0	0	0.0305	0	0	0	0	0	3.4917
rs115619116	0	0	0.0154	0	0	0	0	0	0	0	3.4917
rs7598926	0	0	0.0049	0	0	0	0	0	0	0	3.4808
rs34422916	0	0	0.0039	0	0	0	0	0	0	0	3.4316
rs13433970	0	0	0	0	0.0284	0	0	0	0	0	3.3975
rs7618713	0	0	0	0	0.0852	0	0	0	0	0	3.3743
rs2372928	0	0	0.018	0	0	0	0	0	0	0	3.346

R=0.8478	EHBP1L1	PIDD1	TNS1	IBTK	NEK10	COL1A2	ANKRD55	FAM114A1	CYP4B1	COPG2	
rs9863368	0	0	0	0	0.0268	0	0	0	0	0	3.302
rs480646	0	0	0	0	0.0232	0	0	0	0	0	3.2585
rs40497	0	0	0	0	0	0	0.0126	0	0	0	3.2585
rs80233204	0	0	0	0	0	0	0.165	0	0	0	3.2389
rs674798	0	0	0	0	0.0181	0	0	0	0	0	3.2295
rs551621	0	0	0	0	0.0191	0	0	0	0	0	3.2295
rs515402	0	0	0	0	0.0186	0	0	0	0	0	3.2295
rs6790344	0	0	0	0	0.0192	0	0	0	0	0	3.2295
rs514668	0	0	0	0	0.0186	0	0	0	0	0	3.2249
rs820617	0	0	0	0	0.0186	0	0	0	0	0	3.2249
rs10490444	0	0	0.0239	0	0	0	0	0	0	0	3.2204
rs864241	0	0	0	0	0.018	0	0	0	0	0	3.216
rs653076	0	0	0	0	0.0678	0	0	0	0	0	3.2116
rs653886	0	0	0	0	0.0604	0	0	0	0	0	3.2116
rs6551192	0	0	0	0	0.054	0	0	0	0	0	3.2116
rs608374	0	0	0	0	0.1031	0	0	0	0	0	3.2073
rs2053696	0	0	0	0	0	0	0.0686	0	0	0	3.203
rs4973760	0	0	0	0	0.0466	0	0	0	0	0	3.203
rs569960	0	0	0	0	0.0194	0	0	0	0	0	3.1988
rs3103745	0	0	0	0	0.0435	0	0	0	0	0	3.1988
rs6531677	0	0	0	0	0	0	0	0.6846	0	0	3.1906
rs578501	0	0	0	0	0.1052	0	0	0	0	0	3.1865
rs480238	0	0	0	0	0.0555	0	0	0	0	0	3.1865
rs642570	0	0	0	0	0.0384	0	0	0	0	0	3.167
rs574955	0	0	0	0	0.0527	0	0	0	0	0	3.0993
rs12186035	0	0	0	0	0.0481	0	0	0	0	0	3.0618
rs646577	0	0	0	0	0.113	0	0	0	0	0	3.0618
rs6976724	0	0	0	0	0	2.5083	0	0	0	0	3.0357
rs4376436	0	0	0	0	0	1.2874	0	0	0	0	2.9889
rs10489769	0	0	0	0	0	0	0	0	3.4848	0	2.9677
rs115160760	0	0	0.0297	0	0	0	0	0	0	0	2.9677
rs7626646	0	0	0	0	0.0132	0	0	0	0	0	2.9478
rs7639475	0	0	0	0	0.0291	0	0	0	0	0	2.9112
rs7628297	0	0	0	0	0.0185	0	0	0	0	0	2.8943
rs10510590	0	0	0	0	0.022	0	0	0	0	0	2.8782
rs6551174	0	0	0	0	0.0195	0	0	0	0	0	2.8782
rs1522166	0	0	0	0	0.0252	0	0	0	0	0	2.8782
rs2888447	0	0	0.0136	0	0	0	0	0	0	0	2.8782
rs13420604	0	0	0.0204	0	0	0	0	0	0	0	2.8338
rs3950256	0	0	0.0403	0	0	0	0	0	0	0	2.8202
rs34684700	0	0	0	0	0.0124	0	0	0	0	0	2.7944
rs7635836	0	0	0	0	0.0798	0	0	0	0	0	2.7589
rs1522134	0	0	0	0	0.0358	0	0	0	0	0	2.7478
rs13092134	0	0	0	0	0.1007	0	0	0	0	0	2.7478
rs2141817	0	0	0.0075	0	0	0	0	0	0	0	2.7478
rs13315261	0	0	0	0	0.0302	0	0	0	0	0	2.7266
rs62174841	0	0	0.0199	0	0	0	0	0	0	0	2.6783
rs13433904	0	0	0	0	0.0202	0	0	0	0	0	2.6783
rs2141818	0	0	0.0073	0	0	0	0	0	0	0	2.6693
rs2542197	0	0	0.0065	0	0	0	0	0	0	0	2.6437
rs513546	0	0	0	0	0.0386	0	0	0	0	0	2.6437
rs1604834	0	0	0	0	0	0	0	0.557	0	0	2.6276
rs10228040	0	0	0	0	0	0	0	0	0	0.1238	2.6045
rs17616765	0	0	0	0	0	0	0	1.8039	0	0	2.6045
rs7627508	0	0	0	0	0.0174	0	0	0	0	0	2.5972
rs1922010	0	0	0.03	0	0	0	0	0	0	0	2.569
rs73038733	0	0	0	0	0.0137	0	0	0	0	0	2.5364
rs7589722	0	0	0.0212	0	0	0	0	0	0	0	2.5302
rs1882423	0	0	0.0291	0	0	0	0	0	0	0	2.4949
rs10194193	0	0	0.0251	0	0	0	0	0	0	0	2.4838
rs1882422	0	0	0.0162	0	0	0	0	0	0	0	2.4573
rs73149363	0	0	0	0	0.012	0	0	0	0	0	2.4573
rs7688418	0	0	0	0	0	0	0	0.5129	0	0	2.4228
rs10195960	0	0	0.0258	0	0	0	0	0	0	0	2.4089
rs873907	0	0	0.0144	0	0	0	0	0	0	0	2.3867
rs16856807	0	0	0.0046	0	0	0	0	0	0	0	2.3656
rs1882417	0	0	0.0465	0	0	0	0	0	0	0	2.3615
rs1921984	0	0	0.0225	0	0	0	0	0	0	0	2.3575
rs6714809	0	0	0.0532	0	0	0	0	0	0	0	2.3378
rs1011502	0	0	0.0869	0	0	0	0	0	0	0	2.3339
rs17165295	0	0	0	0	0	0	0	0	0	0.0296	2.3263
rs10171745	0	0	0.0304	0	0	0	0	0	0	0	2.3263
rs10191536	0	0	0.0257	0	0	0	0	0	0	0	2.2904
rs6778395	0	0	0	0	0.0179	0	0	0	0	0	2.2904
rs7612312	0	0	0	0	0.0046	0	0	0	0	0	2.2904
rs832586	0	0	0	0	0	0	0.0364	0	0	0	2.2904
rs683503	0	0	0	0	0.035	0	0	0	0	0	2.2262
rs610118	0	0	0	0	0.0131	0	0	0	0	0	2.2262
rs78140845	0	0	0	0	0.0052	0	0	0	0	0	2.2262
rs113510705	0	0	0.0218	0	0	0	0	0	0	0	2.1701
rs73036939	0	0	0	0	0.0105	0	0	0	0	0	2.1201
rs2253540	0	0	0.0212	0	0	0	0	0	0	0	2.1201
rs17019306	0	0	0	0	0.0952	0	0	0	0	0	2.1201
rs6957613	0	0	0	0	0	0	0	0	0	0.0182	2.1201
rs860579	0	0	0	0	0	0	0.0257	0	0	0	2.1201
rs1851357	0	0	0	0	0.0141	0	0	0	0	0	2.0969
rs1879769	0	0	0	0	0.0552	0	0	0	0	0	2.0749
rs55891220	0	0	0.0183	0	0	0	0	0	0	0	2.0537
rs13414201	0	0	0.0234	0	0	0	0	0	0	0	2.0537
rs73036988	0	0	0	0	0.0044	0	0	0	0	0	2.0335
rs34946790	0	0	0	0	0.0632	0	0	0	0	0	1.9954
rs1445114	0	0	0	0	0.0646	0	0	0	0	0	1.96
rs66472045	0	0	0	0	0.0055	0	0	0	0	0	1.9431
rs9809173	0	0	0	0	0.0322	0	0	0	0	0	1.9431
rs17317435	0	0	0	0	0.0138	0	0	0	0	0	1.9431
rs17019335	0	0	0	0	0.1593	0	0	0	0	0	1.9268
rs34453788	0	0	0	0	0.0113	0	0	0	0	0	1.911
rs17681498	0	0	0	0	0.0109	0	0	0	0	0	1.911
rs17019305	0	0	0	0	0.2497	0	0	0	0	0	1.8808
rs17737947	0	0	0	0	0	0	0	0	0	0.0118	1.8808
rs17019287	0	0	0	0	0.03	0	0	0	0	0	1.8384
rs34641383	0	0	0	0	0.0264	0	0	0	0	0	1.8384
rs73053866	0	0	0	0	0.0228	0	0	0	0	0	1.8384
rs77850809	0	0	0	0	0	0	0.0378	0	0	0	1.825
rs13091109	0	0	0	0	0.0035	0	0	0	0	0	1.7507
rs13005158	0	0	0.018	0	0	0	0	0	0	0	1.7392
rs1522153	0	0	0	0	0.0309	0	0	0	0	0	1.7392

R=0.8478	EHBP1L1	PIDD1	TNS1	IBTK	NEK10	COL1A2	ANKRD55	FAM114A1	CYP4B1	COPG2	
rs78489857	0	0	0.0142	0	0	0	0	0	0	0	1.5382
rs73036929	0	0	0	0	0.0198	0	0	0	0	0	1.5301
rs2705593	0	0	0.0032	0	0	0	0	0	0	0	1.4909
rs75882503	0	0	0	0	0.0101	0	0	0	0	0	1.4684
rs17317379	0	0	0	0	0.0094	0	0	0	0	0	1.4395
rs13083746	0	0	0	0	0.0251	0	0	0	0	0	1.4051
rs1882420	0	0	0.0063	0	0	0	0	0	0	0	1.3984
rs73053826	0	0	0	0	0.0149	0	0	0	0	0	1.3984
rs114818822	0	0	0	0	0.0243	0	0	0	0	0	1.3787
rs832575	0	0	0	0	0	0	0.0287	0	0	0	1.2816
rs860581	0	0	0	0	0	0	0.0156	0	0	0	1.2265
rs80214772	0	0	0.0026	0	0	0	0	0	0	0	1.2265
rs79376126	0	0	0	0	0.0052	0	0	0	0	0	1.175
rs115365220	0	0	0	0	0.0238	0	0	0	0	0	1.175
rs17833456	0	0	0.0021	0	0	0	0	0	0	0	1.1264
rs76744836	0	0	0.0049	0	0	0	0	0	0	0	1.1264
rs11892687	0	0	0.005	0	0	0	0	0	0	0	1.1264
rs17688079	0	0	0	0	0	0	0	0	0	0.0246	0.9542
rs76114469	0	0	0.0019	0	0	0	0	0	0	0	0.9542
rs866222	0	0	0	0	0	0	0.0258	0	0	0	0.9154
rs860580	0	0	0	0	0	0	0.0121	0	0	0	0.9154
rs10954290	0	0	0	0	0	0	0	0	0	0.0062	0.9154
rs832548	0	0	0	0	0	0	0.5495	0	0	0	0.9154
rs17833328	0	0	0.0018	0	0	0	0	0	0	0	0.8416
rs114571087	0	0	0.0129	0	0	0	0	0	0	0	0.8064
rs77680993	0	0	0	0	0	0	0.0131	0	0	0	0.7722
rs13029188	0	0	0.0017	0	0	0	0	0	0	0	0.7722
rs252895	0	0	0	0	0	0	0.0051	0	0	0	0.7722
rs116303454	0	0	0	0	0.0216	0	0	0	0	0	0.7388
rs13158067	0	0	0	0	0	0	0.0029	0	0	0	0.7388
rs115863720	0	0	0	0	0.0029	0	0	0	0	0	0.7063
rs2662033	0	0	0	0	0	0	0.0029	0	0	0	0.7063
rs78649797	0	0	0	0	0	0	0.0124	0	0	0	0.6745
rs2293746	0	0	0	0	0	0	0	0	0	0.0122	0.5828
rs77561695	0	0	0	0	0.0136	0	0	0	0	0	0.5828
rs17832752	0	0	0.0038	0	0	0	0	0	0	0	0.5828
rs115189646	0	0	0	0	0	0	0.0103	0	0	0	0.5534
rs72644086	0	0	0	0	0	0	0.0094	0	0	0	0.5244
rs12694402	0	0	0.0024	0	0	0	0	0	0	0	0.4399
rs547311	0	0	0	0	0	0	0	0	0	0.0031	0.4399
rs79790279	0	0	0.0013	0	0	0	0	0	0	0	0.4399
rs205730	0	0	0	0	0	0	0	0	0	0.0115	0.4399
rs79337924	0	0	0.0087	0	0	0	0	0	0	0	0.3853
rs74622910	0	0	0	0	0.0018	0	0	0	0	0	0.3585
rs17628423	0	0	0	0	0.0007	0	0	0	0	0	0.3319
rs72947750	0	0	0.0018	0	0	0	0	0	0	0	0.3055
rs62255593	0	0	0	0	0.0016	0	0	0	0	0	0.2533
rs62255255	0	0	0	0	0.0005	0	0	0	0	0	0.2275
rs12328709	0	0	0.0002	0	0	0	0	0	0	0	0.2019
rs72951824	0	0	0.0002	0	0	0	0	0	0	0	0.2019
rs7798603	0	0	0	0	0	0	0	0	0	0.0103	0.2019
rs12328842	0	0	0.0002	0	0	0	0	0	0	0	0.1764
rs80216790	0	0	0.0006	0	0	0	0	0	0	0	0.1764
rs11690464	0	0	0.0002	0	0	0	0	0	0	0	0.151
rs2542200	0	0	0.0009	0	0	0	0	0	0	0	0.1257
rs13091103	0	0	0	0	0.0002	0	0	0	0	0	0.1004
rs10048828	0	0	0.0001	0	0	0	0	0	0	0	0.0753
rs1321765	0	0	0	0.02	0	0	0	0	0	0	0.0753
rs1921983	0	0	0.0002	0	0	0	0	0	0	0	0.0251
rs12477625	0	0	0	0	0	0	0	0	0	0	0
rs115169587	0	0	-0.0007	0	0	0	0	0	0	0	-0.1004
rs115193441	0	0	0	0	0	0	-0.0005	0	0	0	-0.1257
rs13405034	0	0	-0.0002	0	0	0	0	0	0	0	-0.1257
rs2895195	0	0	0	0	0	0	0	0	0	-0.0012	-0.2275
rs13436820	0	0	0	0	0	0	-0.0015	0	0	0	-0.2275
rs157935	0	0	0	0	0	0	0	0	0	-0.0013	-0.2275
rs205763	0	0	0	0	0	0	0	0	0	-0.002	-0.3319
rs13422596	0	0	-0.0011	0	0	0	0	0	0	0	-0.3853
rs80053539	0	0	0	0	0	0	-0.0056	0	0	0	-0.4677
rs157925	0	0	0	0	0	0	0	0	0	-0.0034	-0.5534
rs13247947	0	0	0	0	0	0	0	0	0	-0.0032	-0.5534
rs207210	0	0	0	0	0	0	0	0	0	-0.0043	-0.5534
rs2372946	0	0	-0.0049	0	0	0	0	0	0	0	-0.8779
rs4146566	0	0	0	0	0	0	-0.047	0	0	0	-0.9154
rs76142591	0	0	-0.0197	0	0	0	0	0	0	0	-0.9154
rs6435962	0	0	-0.0009	0	0	0	0	0	0	0	-1.0364
rs79509403	0	0	-0.003	0	0	0	0	0	0	0	-1.0364
rs2705594	0	0	-0.0028	0	0	0	0	0	0	0	-1.1264
rs12193653	0	0	0	-0.1777	0	0	0	0	0	0	-1.2816
rs16886445	0	0	0	0	0	0	-0.025	0	0	0	-1.4758

References

[1] ang I. Li, Bryce van de Geijn, Anil Raj, David A. Knowles, Allegra A. Petti, David Golan, Yoav Gilad, , and Jonathan K. Pritchard. Rna splicing is a primary link between genetic variation and disease. *Science*, April 2016.

[2] Steven A. Frank. The common patterns of nature. *J Evol Biol*, August 2009.

[3] Norio Gouda, Yuh Shiwa, Motohiro Akashi, Hirofumi Yoshikawa, Ken Kasahara, and Mitsuru Furusawa. Distribution of human single-nucleotide polymorphisms is approximated by the power law and represents a fractal structure. *Genes to Cells*, March 2016.

[4] Chang-Yong Lee. A model for the clustered distribution of snps in the human genome. *arXiv*, May 2016.

[5] Younghee Lee, Seonggyun Han, Dongwook Kim, Emrin Horgousluoglu, Shannon L. Risacher, Andrew J Saykin, and Kwangsik Nho. Genetic variation affecting exon skipping contributes to brain structural atrophy in alzheimer's disease. *AMIA joint Summits on Tanslational Science*, May 2018.

[6] Aniket Mishra and Stuart Macgregor. Vegas2: Software for more flexible gene-based testing. *Cambridge University Press*, February 2015.

[7] Eliseos J. Mucaki, Ben C. Shirley, and Peter K. Rogan. Expression changes confirm genomic variants predicted to result in allele-specific, alternative mrna splicing. *frontiers in Genetics*, March 2020.

[8] Stan Pounds and Stephan W. Morris. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *BIOINFORMATICS*, January 2003.

[9] web site. bcac.ccge.medschl.cam.ac.uk/.

[10] web site. en.wikipedia.org/wiki/gamma distribution.

[11] web site. www.ncbi.nlm.nih.gov/genome/51. in column refseq click on chromosome. in box send to choose coding sequence, file, format fasta nucleotide.

[12] web site. www.uniprot.org.

[13] Dmitri V. Zaykin. Optimally weighted z-test is a powerful method for combining probabilities in meta-analysis. *Journal of Evolutionary Biology*, August 2011.