Racial Differences and the Probability of C2orf16 rs191912 to be the Major Gene Locus of General Cognitive Ability

Volkmar Weiss[1]
Leipzig

By applying the logic of genetics and available knowledge on allele frequencies it is possible to put forward hypotheses on genes underlying IQ by data mining. At the present state of knowledge the probability of a false positive finding that C2orf16 rs1919128 is the major gene of IQ is about 0.000005

Key Words: General intelligence; IQ; HapMap database; SNPlogic; Personal Genome Project.

The "Statement on 'Race'" of the Executive Board of the American Anthropological Association (1968) claims: "At the end of the 20th century, we now understand that human cultural behaviour is learned, conditioned into infants beginning at birth, and always subject to modification. … Our temperaments, dispositions, and personalities, regardless of genetic propensities, are developed within sets of meanings and values that we call 'culture'. … Given what we know about the capacity of normal humans to achieve and function within any culture, we conclude that present-day inequalities between so-called 'racial' groups are not consequences of their biological inheritance."

Nevertheless, despite this general capacity of healthy human beings, blacks are black and whites are white, among a number of additional differing physical and mental traits (Rushton, 2000). Therefore, to Mark D. Shriver the idea that a gene responsible for skin color should differ in its allele frequencies among blacks and whites seemed not to be illogical. And indeed, if we rank the known list of millions of single nucleotide polymorphisms (SNPs) of the human genome according to the magnitude of their differences in allele frequencies between blacks and whites, we see on the second place of this list the gene which is essential to make an European man white: SLC24A5. On position rs1426654 of the 111th amino acid of this gene two alleles differ in only

[1] Address for communication: volkmar-weiss@t-online-de

one nucleotide, changing from alanine to threonine. In the database of the International HapMap Project of genetic variation in human populations the threonine allele was present in 100% of the European sample, while the ancestral alanine allele was found in 98% of the Yoruba sample from Nigeria (Lamason et al., 2005).

If we try to apply this logic of data mining within genetic databases to discover the genes underlying IQ we have to overcome one taboo (Pearson, 1997) and one mainstream dogma. A lot of people, who will not deny that there could be genes of skin color, will be outraged to the outmost when someone claims (read, for example, Malloy, 2008, on the James Watson Affair) that there should also be genes underlying IQ differences between populations and races. Even the author of the German bestseller *Deutschland schafft sich ab* ("Germany Does Away With Itself") Sarrazin (2010) had to add in subsequent editions a page, inserted before the title, on which he had to assure that he never claimed that IQ differences between populations had any genetic background.

But if we break this taboo and take seriously the results of more than a century of research of psychometrics (Lynn & Vanhanen, 2002; Rindermann, 2007) and on a possible genetic background of general intelligence (Weiss, 1992), then a different distribution of allele frequencies in different populations and social strata should be expected.

If all individuals within a population share the same gene, as for example all humans share a gene to develop four limbs, then 100 percent of relatives of all degrees have four limbs. However, if there is a genetic polymorphism and one allele that is very rare and its frequency in the overall population near zero, only homozygote carriers of such a rare allele may exhibit the characteristic, for example, to be an albino. In this case, the frequency of albinism among relatives decreases very rapidly with each degree of decreasing consanguinity to the proband. In other words, from the slope of the decrease the allele frequency underlying the character can be estimated. By applying the method of stochastic Mendelian matrices (Li and Sacks, 1954) Weiss (see Weiss, 1992, 2000) put forward the hypothesis that a major gene could explain the frequency of

high giftedness (all subjects with an IQ above 123) among the consanguine kin of the highly gifted (see Weiss 2009a, Tabl. 4).

With this method, in a population with a mean IQ of about 100, the frequency q of this hypothetical allele G was estimated to be about 0.20. From the Hardy-Weinberg-law of population genetics follows that the frequency q of the hypothetical major gene G is $(1-q)^2 + 2q(1-q) + q^2 = 1$. From $q = 0.20$ follows that $2q(1-q) + q^2 = 0.36$. The percentile rank of this frequency of 0.36 corresponds to an IQ of 105 (in a population with a mean IQ of 100), and the percentage of individuals with an IQ beyond or below the threshold of IQ 105 can be used to estimate the allele frequency q of G (see Weiss 2009a, Table 5) in different countries.

Further, in the general population from Mendelian segregation of IQ within families the conclusion was drawn that healthy individuals with a genotypical IQ above 123 are homozygous and those above IQ 104 are heterozygous for this allele G. This means that children, whose both parents have an IQ above 123, will also have an IQ above 123. Children, whose both parents have an IQ below 105, will also have an IQ below 105. The offspring of parents with an IQ between IQ 104 and 124 segregates in accordance with the Mendelian rules: 25 percent of children have an IQ below 105, 50 percent between IQ 104 and 124, and 25 percent an IQ above 123. The data in favor of this major gene hypothesis have already been published, with the necessary extent and detail in books (Weiss, Lehrl, & Frank, 1986; Weiss, 2000) and in refeered journals (Weiss, 1992, 1994, 1995), where its pro and contra was discussed, and this will not be repeated here.

Additional support for the major gene hypothesis comes from the fact that raw test scores of elementary cognitive tasks can exhibit multimodal distributions where the means of the allelotypes GG, GA and AA are simple manifolds with corresponding variances (Lehrl and Frank, 1982: Weiss 1992, 1995).

The present state of knowledge allows a search for this major gene locus of IQ by data mining.

From the total of autosomal 76690 nonsynonymously coding SNPs in the HapMap database (Genecards, 2000), we

have used the database SNPLogic (2009) to filter out 204 SNPs fitting within the expected ranges of allele frequencies in samples of European (CEU), Chinese (CHB), Japanese (JPT) and black Sub-Saharan Yoruba (YRI) populations. However, one should be aware that especially the Asian HapMap samples drawn from Beijing and Tokyo are not socially representative in any way. The Chinese sample comes from a Beijing university resident academic population. Therefore, for this and the Japanese population the threshold should be set far above 0.20, for the black population below 0.10.

Because among representative European samples the frequency of the rare allele underlying high IQ in the homozygous state should not exceed 0.30, homozygosity by pure chance (0.30 x 0.30) can be expected in less than 0.10 of cases. Assuming (without IQ testing) that Craig Venter is a proband of high IQ, we used the published *Craig Venter Genome* (Venter, 2008) by looking for homozygosity of this allele, and we reduced in this way the list of candidate SNPs from 204 to 22.

By looking at a second high-IQ proband with decoded genome data of similar quality as that of Craig Venter's, theoretically, this list could be further reduced by a factor of 0.10 to about 2, including the hypothetical major gene locus. However, for 10 candidate SNPs there are no data in the James Watson (Venter, 2008) and George Church (2010) genome databases. The Personal Genome Project and 23andMe are using the 500 K Affymetrix chip for genotyping. Also from the Steven Pinker data follows that a high IQ gene cannot be on this chip, which analyzes 500 000 SNPs per proband on one run. By data mining we replicated in this way the completely negative results of Butcher et al. (2008), who found no replicable correlation between any SNP on the 500 K Affymetrix chip and IQ. This kind of genetic investigation has always a similarity with a lottery. By analyzing 500 000 loci even in large samples a small number of SNPs will be associated with the target criterion by pure chance and assumed to be minor genes of IQ, but turn out to be false positives in attempts to replicate the results, as was the case in many studies done be Plomin and his coworkers. This non-replicability of many results, because they are

nothing else than false positives, is a general problem of contemporary genetics and not specific for the genetical background of IQ.

After the exclusion of all the SNPs which are on the 500 K Affymetrix chip, there remained 11 SNPs as candidates for a major gene locus of IQ (Weiss, 2009b). By looking at the published data of the Personal Genome Project (Church, 2010), we could reduce this number to two SNPs: C2orf16 rs1919128 Venter GG CEU 0.24 CHB 0.56 JPT 0.62 YRI 0.04 and tcag7956 rs6961834 Venter TT CEU 0.37 CHB 0.48 JPT 0.44 YRI 0.06. Until now, for the latter intergenic SNP there are no additional data available.

C2orf16 codes for a still uncharacterized protein, thought to be involved in phosphorylation and signaling (Genecards, 2010), until now has never been the target of any association or linkage study. If we hypothesize C2orf16 I774V to be the major gene locus underlying IQ, we can calculate the following probabilities – this to be a false positive finding:

(1) 204 of 76690 SNPs are within the expected range of allele frequencies: 204/76690 = 0.0027

(2) Given the frequency for G as 0.25, the probability of Craig Venter to be homozygous GG is 0.0625

(3) The frequency of the ancestral allele A is 0.75, the probability to be homozygous AA 0.5625. From 6 known probands of the Personal Genome Project with above average cognitive ability none has the reference genome type AA. The probability of this is 0.0317

The product of the probabilities (1), (2) and (3) is: 0.0027 x 0.0625 x 0.0317 = 0.0000053

Without doubt, under certain circumstances (see, for example, Payton et al., 2010) the IQ is influenced by some hundreds of minor genes (Meisenberg, 2005) and even more by measurement error and environmental effects greater than any of such a single minor effect. Further, if we exclude theoretically any major gene effect on IQ, we have the advantage to share the mainstream dogma which stresses the overwhelming effects of the culture. But the data of segregation of IQ within families (Weiss, 1992), the

percentages of highly gifted relatives (IQ above 123) among highly gifted probands (Weiss, 1994) and the overall data of social mobility, all never discussed or explained by the adherents of the mainstream position in accord with Occam's razor, speak another language.

Genotyping only 12 additional probands with an IQ above 105 (or better IQ 115, because, of course, the phenotypes are overlapping to a certain degree) and confirming their non-AA status would add a probability factor of about 0.001 and cost no more than 300 Dollars in an experienced laboratory. However, at present, in Germany we have already a law that no laboratory is allowed to investigate such a hypothesis without the consent of an ethics committee. And no such committee will give its consent to test a hypothesis which is put forward by starting with a taboo break. In a science fiction novel Weiss (2007) predicted the discovery of the major gene locus of IQ in 2013 and the worldwide prohibition of such research in 2018. In fact, a prohibition in an indirect way of preventing any such discovery from being made. Maybe, there are still some places in the world, where scientists live, who can do such research and can dare to publish it.

For studies on alzheimer, dyslexia, autism, schizophrenia and so on social status and hence IQ are confounding variables of such importance that research workers in these fields should have an interest to partial out the influence of general cognitive ability and to clarify its genetic background (Payton, 2006). It should not be impossible that this argument might even convince an enlightened ethics committee (Reiss, 2000).

If it could be confirmed that C2orf16 I774V is the major gene locus of IQ, it would be a major breakthrough. And if not, this paper remains an example of the possibilities of data mining, which in the future could be applied to copy number variations and the genetics of microRNA, on which population data are entirely lacking in the current databases which still contain a lot of deficiencies, inconsistencies and simply errors.

Acknowledgement

Without the help of Ivan Smirnov (SNPLogic), Steven Pinker and Andrew Walsh this data mining would not have been possible.

References

American Anthropological Association
(1998) Statement on Race. http://www.aaanet.org/stmts/racepp.htm

Butcher, L. M., Davis, O. S., Craig, I. Q. & Plomin R.
(2008) "Genom-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K singe nucleotide polymorphism microarrays." *Genes, Brain, and Behavior* 7: 435-446.

Church, G.
(2010) Personal Genome Project: http://evidence.personalgenomes.org/C2ORF16-I774V

Genecards
(2010) http://www.genecards.org/cgi-bin/carddisp.pl?gene=C2orf16&search=C2orf16

Lamason, R. L. et al.
(2005) "SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans." *Science* 310: 1782–17866.

Lehrl, S. & Frank, H. G.
(1982) „Zur humangenetischen Erklärung der Kurzspeicherkapazität als der zentralen individuellen Determinante von Spearmans Generalfaktor der Intelligenz„. *Grundlagenstudien der Kybernetik und Geisteswissenschaften/Humankybernetik* 23: 177-187.

Li, C. C. & Sacks, L.
(1954) "The derivation of joint distribution and correlation between relatives by the use of stochastic matrices." *Biometrics* 10: 347-360.

Lynn, R. & Vanhanen, T.
(2002) *IQ and the Wealth of Nations*. Westport: Praeger.

Malloy, J.
(2008) "First a hero of science and now a martyr to science: The James Watson Affair – Political correctness crushes free scientific communication." *Medical Hypotheses* 70: 1077-1080.

Meisenberg, G.
(2005) "Genes for intelligence. A review of recent progress." *Mankind Quarterly 36*: 139-164.

Payton, A.
(2006) "Investigating cognitive genetics and its implications for the treatment of cognitive deficit." *Genes, Brain, and Behavior* 5 Suppl. 1: 44-53.

Payton, A., Miyajima, F., Ollier, W., Rabbitt, P., Pickles, A., Weiss, V., Pendleton, N., & Horan, M.
(2010) "Investigation of a functional quinine oxidoreductase (NQO2) polymorphism and cognitive decline." *Neurobiology of Aging* 31: 351-352.

Pearson, R.
(1997) Race, Intelligence and Bias in Academe. Washington, DC: Scott-Townsend.

Reiss, M. J.
(2000) "The ethics of genetic research on intelligence." *Bioethics* 14: 1-15

Rindermann, H.
(2007) "The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests." *European Journal of Personality* 21: 667-706: http://www-classic.uni-graz.at/pslgcwww/rindermann/publikationen/07EJPall.pdf

Rushton, J. P.
(2000) Race, Evolution, and Behavior: A Life History Perspective. Port Huron, MI: Charles Darwin Research Institute.

Sarrazin, T.
(2010) Deutschland schafft sich ab. Munich: Deutsche Verlags-Anstalt.

SNPLogic
(2009) http://www.snplogic.org/

Venter, C.
(2008) http://jimwatsonsequence.cshl.edu/cgi-perl/gbrowse/cvsequence/

Weiss, V.
(1992) "Major genes of general intelligence." *Personality and individual Differences* 13: 1115-1134. http://www.v-weiss.de/majgenes.html

Weiss, V.
(1994) "Mathematical giftedness and family relationship." *European Journal of High Ability* 5: 58-67: http://www.v-weiss.de/ability.html

Weiss, V.
(1995) "The advent of a molecular genetics of general intelligence." *Intelligence* 20: 115-124.

Weiss, V.
(2000) Die IQ-Falle: Intelligenz, Sozialstruktur und Politik. Graz: Leopold Stocker.

Weiss, V.
(2007) Das Reich Artam. Die alternative Geschichte. Leipzig: Engelsdorfer Verlag.

Weiss, V.
(2009a) "National IQ means transformed from Programme for International Student Assessment (PISA) scores, and their underlying gene frequencies." *Journal of Social, Political, and Economic Studies* 34: 71-94.

Weiss V.
(2009b) "IQ genetics by data mining." *Nature News and Opinion:* http://network.nature.com/groups/naturenewsandopinion/forum/topics/3871?page=9

Weiss, V., Lehrl, S. & Frank, H.
(1986) Psychogenetik der Intelligenz. Dortmund: Modernes Lernen.