



WHITEHEAD INSTITUTE

To: Paul Ferrara, George Sensabaugh, Tom Caskey
From: Eric Lander
Re: Proposed meeting to discuss population genetics
Date: November 24, 1991
cc: All Committee Members

Although we all enjoy visiting our nation's capital and enjoy seeing one another, I am writing to explore whether we truly need another meeting to reach consensus on the population genetic issues.

After discussing your concerns with several committee members, I think that there is no actual disagreement about substance. Rather, there seems to be confusion over what the current version of the report actually recommends.

If so, it should only be necessary to clarify the text. In any case, it seems worth investigating this possibility before convening a meeting that may be unnecessary. Let me try to address the issues raised. Please review these points and let me know whether you agree with them in substance. If so, we need only discover where the report fails to make these points clearly.

The central issue seems to be the admissibility of calculations based on the multiplication rule. Apparently, you read the report as calling for the 1/N rule instead of the multiplication rule.

In fact, this is not the report's intention. Rather, the text (p. 3-30,31) specifically calls for the admissibility of both calculations: specifically, an expert should first state the result of comparing a complete pattern to the database (i.e., the 1/N rule) and then may state the result of using the multiplication rule. Thus, both calculations are explicitly endorsed as admissible and scientifically acceptable.

The only caveats are that the expert should (1) state that the multiplication rule is based on the assumption that alleles are statistically independent, which assumption is subject of ongoing studies but which the expert considers to be a reasonable and valid assumption; (2) use conservative calculations (specifically, not place too much weight on any one allele); and (3) avoid claims of absolute uniqueness, at least at this point in time.

These recommendations seem to accomplish what you want, while maintaining some appropriate cautions.

Current court battles concern the admissibility of population statistics based on the multiplication rule. Several jurisdictions--including Washington DC, Massachusetts, Minnesota, Chicago--exclude DNA evidence because of the multiplication rule.

In this report, we are explicitly endorsing the admissibility of the multiplication rule. The immediate effect should be to end challenges to the admissibility of DNA based on theoretical issues of population genetics. This would seem to be your major goal. Once the multiplication rule is admissible as a matter of law, it gets to the jury. Once it gets to the jury, the battle is over. (Of course, regardless of what we say, the defense always has the constitutional right to try to rebut any particular application as a matter of fact, but this typically carries little weight. Specifically, while judges have excluded DNA as a matter of law in a pre-trial hearing, no defendant has ever convinced a jury to ignore the evidence at trial.)

In return for this endorsement, the report asks for three caveats:

(1) The statistics be reported in two-steps: (i) the result of simple comparison to the database (because it requires no assumptions about population structure), followed by (ii) the result of the multiplication rule (which requires assumptions about population structure). Because we all agree that there is continued debate and ongoing population studies, the expert is required to acknowledge such debate exists but is free to state that, in his opinion, the multiplication rule appears to be well justified based on the available evidence.

I think this accomplishes everyone's goals. It explicitly declares the multiplication rule admissible, while recognizing that the underlying assumptions are still being debated in some quarters.

Realistically, I don't see any other position that could be win unanimous endorsement on the committee. And, for my own part, I think that unanimity is crucial. If we have anything but a unanimous position, then we are explicitly telling the courts that there is not general agreement in the scientific community--which will be tantamount to declaring the population genetics must be ruled inadmissible under the Frye standard.

(2) The statistics should be somewhat conservative. Here, the matter is purely technical. If one observes an allele frequency of p in a database, what conservative frequency should be used in calculations? One must consider two types of error: (i) error due to sampling, and (ii) error due to genetic drift in ethnic groups. Both terms are given by classical textbook formulas, with the second being the larger (see page 3- 18). The key point is this: the error term is large compared to p for allele frequencies near 1%, but is small compared to p for allele frequencies in the range 5 - 10%. Thus, even if one estimate that the allele frequency is only 1%, one has no confidence that it is not actually several-fold higher in some subgroups. However, if you estimate

the allele frequency at 5%, you can be pretty sure that it cannot be much higher in any ethnic group. Estimates that are swamped by the error term are meaningless. Thus, the estimate of 5% or so comes from solving the equation stating that *signal* should be greater than the *noise*. Accordingly, it makes sense to use a lower bound of 5%.

Realistically, this lower bound amounts to no serious limitation. With four loci, one can obtain population frequencies of up to 6,400,000,000. Indeed, most allele frequencies used in practice exceed 5% given the existing match windows. The rule simply prevents placing too much weight on any one allele. A maximum weight of 1 in 400 per locus or 1 in 6.4 billion per four-locus genotype seems just fine: it is justified both in terms of rigorous mathematics and common sense.

(3) The expert should avoid claiming absolute uniqueness, at the present time. This point seems uncontroversial.

Finally, there seems to be some concern that we are calling for the re-opening of past cases. In fact, the report calls for just the opposite (3-32). Moreover, this seems unlikely because we are explicitly endorsing the multiplication rule.

As a constructive step, I think we should follow a suggestion of Paul's that we explicitly provide a sample statistical report that we would all find acceptable. If we can agree on this, the rest is mere commentary. Toward this end, let me offer a draft sample statistical report to insert in the chapter:

Sample Statistical Testimony

"Your honor, we first compared the DNA pattern to the 2000 samples in our population database. We found that the pattern matched none of these 2000 samples, showing that the frequency of the pattern certainly cannot be much greater than 1 in 2000.

In fact, the DNA pattern is much rarer than this. If we assume that each component of the DNA pattern is statistically independent, the frequency of the pattern would be estimated to be about 1 in 20 million. Although there are ongoing scientific studies to evaluate the assumption of statistical independence, it is my scientific opinion that the evidence reasonably supports this assumption. Accordingly, it is my expert opinion that the frequency of the pattern can be estimated to be approximately 1 in 20 million in the general population. In short, the DNA pattern is extremely rare."

Based on extensive conversations with committee members over two years, it seems to me that this statement strikes a balance that is acceptable to all, while a statement that goes much further in either direction will engender dissent.

If you find the Sample Statistical Testimony essentially acceptable, then I think that we simply need a few phone calls or faxes to edit the text.

If the basic tenor of this statement is fundamentally unacceptable, however, then we can surely meet but there is a good chance that the group will not reach consensus. This would be unfortunate, because I think that courts will take conflicting views as proof of lack of scientific consensus. If anything, this will result in exclusions of DNA evidence.