

### Associating Mitochondrial DNA Variation with Complex Traits

*To the Editor:*

We read with great interest the recent paper by Saxena and colleagues describing the “comprehensive association testing of common mitochondrial DNA [mtDNA] variation in metabolic disease.”<sup>1</sup> We strongly support the use of large cohorts of carefully phenotyped cases and controls in studies of this type, large cohorts being essential to provide adequate power to detect small or moderate associations between mtDNA sequence variants and disease.<sup>2</sup> We do, however, have concerns that the study described by Saxena et al. is neither sufficiently accurate nor sufficiently comprehensive for exploring all possible mechanisms of disease association with common genetic variants of mtDNA. We therefore urge caution when applying their approach to other disease states, because it is not well grounded in the fundamental aspects of mtDNA phylogenetics. As a result, their approach may lead to spurious associations, as well as to false negative results.

Mitochondrial DNA is inherited strictly down the maternal line and therefore undergoes negligible intermolecular recombination, and the signature of recombination has not been detected at the population level.<sup>3,4</sup> Mutations acquired throughout human history have subdivided the human population into a number of discrete clades or haplogroups.<sup>5</sup> The major haplogroups have been identified by a search for common genetic variants in the population, initially through the use of restriction site analysis and subsequently through direct sequencing of both coding and noncoding regions of mtDNA.<sup>6</sup> In constructing the most parsimonious phylogenetic structure of human mtDNA evolution, it has become clear that many substitutions have arisen more than once on different parts of the phylogenetic tree (i.e., homoplasies; see table 2 in the work of Herrnstadt et al.<sup>7</sup> for the most complete list). Homoplasia adds complexity to the population structure of mtDNA, and it is a serious confounding phenomenon in the association approach used by Saxena et al.

Saxena and colleagues used publicly available human mtDNA coding-sequence data sets to identify mtDNA polymorphisms that occur at a frequency of >1% in the European population. To derive this key information, the authors state that they used only sequences of European origin to compile their list of polymorphisms. Thus, they report that they used 536 sequences from the MitoKor data set released by Herrnstadt et al.<sup>7</sup> However, this data set contained a total of 560 sequences, of which 56 were African, 69 Asian, and only 435 European. On the basis of their report, it appears that Saxena et al. included at least 101 non-European sequences in their core data set,

some ~8% of the total. The inclusion of sequences from non-European subjects could have resulted in the identification of non-European SNPs in their initial data set or in the exclusion of specific European variants, because, in their set of 536 sequences, some rare European markers fell below the 1% threshold for inclusion. As an example, site 13105 is most often associated with African lineages, and, in the MitoKor data, it is present in 23 of 56 sequences of African origin but in only 3 of 435 sequences of European origin. As a result, this site should not have been used in the analysis of Saxena et al.

Even the strict use of European sequences is not without problems. Of the GenBank mtDNA sequences used by Saxena et al., 192 were of Finnish origin. It is now recognized that some mtDNA clades are considerably overrepresented in the Finnish phylogeny, presumably because of a recent founder event.<sup>8,9</sup> For example, site m.4928 is absent in the 435 MitoKor European mtDNA sequences but is reported in 13 Finns. In a similar fashion, site 5495 is seen in only 2 of 435 MitoKor sequences but was observed in 23 Finns.<sup>9</sup> As a result, there is another reason to doubt that the authors were able to analyze a representative set of mtDNA SNPs. At best, it appears that their approach is reliable only for robust disease associations with common markers.

Perhaps more importantly, although many of the U.S. and U.K. samples from the MitoKor data set were from the population controls, Herrnstadt et al. explicitly noted that the data set also included mtDNA sequences from patients with various disorders. In fact, the MitoKor set of 435 sequences of European origin included patients with type 2 diabetes ( $n = 42$  [MIM 125853]), with Alzheimer disease ( $n = 60$  [MIM 104300]), and with Parkinson disease ( $n = 37$  [MIM 168600]), thereby accounting for at least 11% of the core data set. This is not a problem if the sequences are being used simply to draw general conclusions about the pattern of mtDNA evolution in the population. However, given the reported association between these diseases and mtDNA sequence variants,<sup>10–13</sup> the use of these sequences potentially alters the spectrum of polymorphic change identified in the first phase of the study by Saxena et al. and subsequently influencing the final list of “tag” SNPs.

Saxena et al. used mathematical methods designed to measure linkage disequilibrium to identify SNPs that tag common mtDNA haplotypes as a way to identify haplogroup clusters. In a comparison of the haplotype tags described by Saxena et al. with a phylogenetic analysis of European mtDNA sequences, a number of patterns emerge. The  $r^2$  approach reliably detects some common haplogroup-defining polymorphisms but—and this is an important point—not others. For example, mtDNA haplogroup T is reliably tagged by two tags, m.12633 and m.11812, that capture site m.8697 shared by all haplo-

group T sequences (shown in blue in fig. 1). However, their technique also identifies common homoplasies, which can account for up to 25% of substitutions in some data sets.<sup>9</sup> As one example, in the MitoKor set of 560 sequences, 497 haplogroup-associated polymorphisms were identified, but 174 (35%) of these were associated with two or more haplogroups. The majority of established mtDNA haplotypes are defined by a combination of SNPs, and, by their very definition, a single homoplasy cannot be used to reliably “tag” a single mtDNA haplotype. The approach of Saxena et al. leads to the association analysis of polyphyletic groups (e.g., as shown in fig. 1, m.709 is found independently on haplogroups U, K, T, and W). On the basis of the phylogeny shown in the figure, at least 53% of the tag SNPs identified by Saxena et al. are associated not with a single mtDNA haplotype but with two or more. Thus, their comparison of the frequencies of a common homoplasy in cases and controls is fraught with difficulty. If one of the distal branches of the phylogenetic tree does harbor a functionally relevant disease-associated polymorphism, that association with the tagging homoplasic marker will be “diluted” by the co-occurrence of the same mtDNA substitution on other, nonassociated clades. An example is shown in red in figure 1, site m.12795 and m.709 capture m.5046 of haplogroup W. However, because of the occurrence of m.709 at multiple places in the phylogeny, many other unrelated sequences are also captured. On the other hand, if the homoplasy itself contributes to pathogenicity, then it might be possible to use the approach developed by Saxena et al. for the identification of a disease association. However, the reported aim of this approach was to identify tag SNPs and not the functional variants themselves.

Another problem with the approach of Saxena et al. is that, by ignoring what is known about mtDNA evolutionary history, they tested a number of sites that were providing the same information on association. For example, both m.11674 and m.12414 “tag” mtDNA haplogroup W (fig. 1). On the basis of an updated reduced median network analysis of the MitoKor data set, 12 (19%) of the 64 tag SNPs studied by Saxena et al. occur in exactly the same place on the phylogeny as other tag SNPs (green boxes in fig. 1). Thus, their approach is inefficient, because of the increased superfluous genotyping, tagging the same haplotype on more than one occasion. This is an instance where less can be more. That is, by using fewer SNPs—but ones that are specifically chosen on the basis of knowledge of mtDNA phylogeny—it is possible to cover all major clades of mtDNA. This approach would reduce the number of statistical tests being performed and thus would increase the power of the study, potentially revealing a significant and biologically relevant association.

The conventional first step in studying the association between mtDNA and complex disease involves the sequential genotyping of specific SNPs that, in European populations, define the major European haplogroups. The results are interpreted in a stepwise manner, with reference

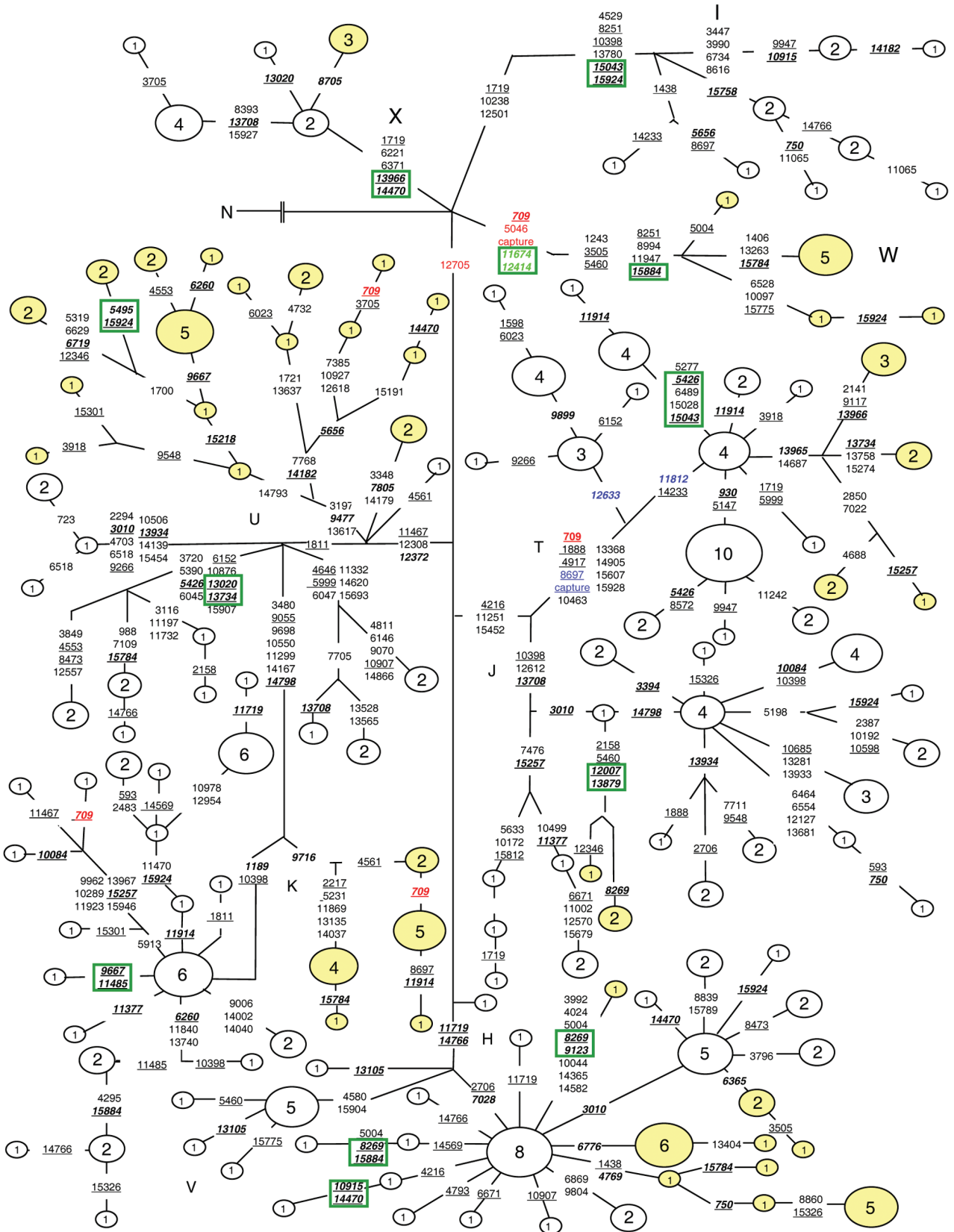
to the phylogenetic tree, to prevent misinterpretations due to the occurrence of homoplasies. For example, the presence of m.4216 defines the JT cluster, which is separated into J and T by m.13708. However, genotyping m.13708 on its own does not define J, because, among European mtDNAs, the same substitution occurs on a subgroup of haplogroup K and X.

By its very nature, the approach of Saxena et al. identifies some common haplogroup markers, but how much additional “tagging” information was acquired by their 64 SNPs? After removal of the nonunique sites ( $n = 35$ ), which occur more than once on the phylogeny and thus do not tag a specific haplotype, followed by removal of the unique haplogroup markers ( $n = 5$ ), 24 unique sites remain (38% of the original total). On the basis of the MitoKor data set, these 24 SNPs tag only 142 of 435 European sequences and, by extrapolation, only 33% of the European population (shown in shaded yellow in fig. 1).

The 64 SNPs described by Saxena et al. can be used to test two hypotheses *in part*: (1) association of a disease with deep-rooted, ancient mtDNA polymorphisms (an association occurs *either* because the specific variant has direct functional consequences *or* because it tags a clade containing functionally relevant variants); and (2) association of the disease with ancestrally young genetic variants that are present on multiple different haplogroup backgrounds. Of the 24 unique sites that do not tag major haplogroups, only 6 could have direct functional consequences (affecting tRNA genes, rRNA genes, or the protein sequence), with the majority ( $n = 15$ ) being synonymous changes within structural genes. Interpreting the results of the analysis is therefore ambiguous—whether the outcome is positive or negative. The approach of Saxena et al. seems beautifully simple, with both statistical significance and power calculated by simulation and random permutation of the data. However, the examples presented here demonstrate the complexity and limitations of the data structure and show that it becomes very difficult to test a disease association without accounting for phylogenetic history. That is, their statistical tests are sound, but they are applied to flawed data in this instance.

Although we acknowledge that their approach will detect a robust disease association, little comfort can be taken because such associations are detected by other approaches. More importantly, we are learning that the reverse seems to be case for disease associations. That is, there is accumulating evidence that such associations will be weak and/or genetically complex.

In terms of what is emerging from other analyses, it must be noted that the approach of Saxena et al. does not test the equally plausible hypothesis that *multiple* mtDNA variants interact and subtly compromise respiratory chain function. This type of association is not without precedent, because suppressor mutations have been described in patients with mitochondrial disorders and possibly have occurred throughout mtDNA evolution.<sup>14</sup> In terms of a specific example, an association has been found be-



tween longevity and a combination of three common polymorphisms m.150T, m.489C, and m.10398G.<sup>16</sup> Such a disease process fits with the basic biology of the mitochondrial electron-transfer chain, in which all of the genes encoded by the mitochondrial genome contribute to the same basic biochemical process of oxidative phosphorylation.<sup>15</sup> Mitochondrial structure and function are guided by the interaction of hundreds of nuclear- and mitochondrially-encoded subunits. It will be a major challenge to test the hypothesis that multiple mtDNA variants interact to alter disease risk—but without testing this hypothesis, no approach can be said to be “comprehensive.”

Finally, mitochondrial disorders are metabolic diseases, affecting the final common pathway of energy metabolism. Leber hereditary optic neuropathy (LHON [MIM 535000]) is the most common mtDNA disease, and >95% of cases are caused by one of three point mutations of mtDNA affecting genes that code for complex I subunits of the respiratory chain: m.3460G→A, m.11778G→A, and m.14484T→C.<sup>17</sup> These mutations have arisen many times over human history, and they are associated with a metabolic defect of respiratory chain complex I activity. Intriguingly, only m.11778G→A and m.14484T→C preferentially occur on a specific mtDNA haplogroup background (J).<sup>18,19</sup> The most likely explanation for these results is that the penetrance (or expressivity) of these pathogenic mutations is modified by the mitochondrial genetic “background.” Recent work suggests that specific subhaplogroups (J1c and J2b) are responsible for this association, possibly through polymorphic variation in the mtDNA cytochrome *b* gene.<sup>20</sup> Iranian pedigrees affected by LHON do not show the same the association with haplogroup J.<sup>22</sup> However, the conspicuous absence of both J1c and J2b from the Iranian population probably explains this observation, illustrating the important point that a clear knowledge of the local population genetic structure, derived from phylogenetics, informs our understanding of relevant disease associations (or the lack thereof).<sup>20</sup> Although there is evidence that genetic variation of mtDNA is associated with some mitochondrial disorders, this is not always the case.<sup>21</sup>

An alternative approach to that of Saxena et al., as mentioned above, is to use our knowledge of mtDNA population structure to reduce the number of redundant sta-

tistical tests and thus to increase the power of the study. For example, one such hypothesis-driven approach would involve testing for an association between nonsynonymous mtDNA homoplasies and disease on different mtDNA haplogroup backgrounds. This hierarchical strategy would provide a means of independently testing a disease association multiple times within the same data set. Careful selection of specific SNPs on the basis of prior knowledge of their evolutionary conservation and likely biochemical effect might also reduce the number of statistical tests being performed. We believe that such hypothesis-driven approaches are less likely to identify a false-positive association. They will also maximize the potential of a given sample size, revealing associations that are biologically plausible and that may not be apparent after less-discriminate genotyping followed by a correction for multiple significance testing or by a simulation-based significance test not reflecting the true structure of genome evolution. In the end, we should expect that a “one size fits all” approach is simply not optimal and that analysis of large data sets with multiple hypothesis-based tests will be necessary.

JOANNA L. ELSON, KARI MAJAMAA, NEIL HOWELL,  
AND PATRICK F. CHINNERY

### Web Resource

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for Alzheimer disease, LHON, Parkinson disease, and type 2 diabetes)

### References

1. Saxena R, Bakker PIW, Singer K, Mootha VK, Burt N, Hirschorn JN, Gaudet D, Isomaa B, Daly MJ, Groop L, et al (2006) Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet* 79:54–61
2. Samuels DC, Carothers AD, Horton R, Chinnery PF (2006) The power to detect disease associations with mitochondrial DNA haplogroups. *Am J Hum Genet* 78:713–720
3. Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (2001) Analysis of European mtDNAs for recombination. *Am J Hum Genet* 68:145–153

---

**Figure 1.** Revised and updated reduced median network analysis of European mtDNA sequences from the MitoKor data set, adapted from the work of Herrnstadt et al.<sup>7</sup> A reticulation at the haplogroups W, I, and X has been resolved. An example of haplogroup H4 has been added. Tag SNPs described in table 4 of Saxena et al.<sup>1</sup> are shown in bold italics (not all are present on the network). Underlined sites are homoplasies in the 435 European sequences of the MitoKor data set. In a larger data set, it is likely that the number of homoplasies would be greater; thus, the data set used by Saxena et al. will have fewer unique sites. *Blue*, An example of good tagging of a cluster of related mtDNA sequences. *Red*, An example of a tag SNP that is a common homoplasy, present at many different sites on the phylogeny. *Green*, An example of two SNPs that uniquely occur at exactly the same point on the phylogeny and thus provide the same information regarding that particular clade. *Green boxes*, Other examples of SNPs that occur at the same point on the phylogeny. *Yellow bubbles*, The areas of the phylogeny captured by the unique sites of Saxena et al. The total number of sequences captured (142) is 66 greater than shown on the network, which does not show all the H sequences in the data set of 435 European mtDNA sequences (because of space limitations).

4. Piganeau G, Eyre-Walker A (2004) A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity* 92:282–288
5. Wallace DC (1994) Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci USA* 91: 8739–8746
6. Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232–249
7. Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson S, Ghosh SS, Olefsky J, Beal MF, Davis RE, et al (2002) Reduced median network analysis of complete mtDNA coding region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70:1152–1171
8. Finnila S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet* 68:1475–8144
9. Moilanen JS, Majamaa K (2003) Phylogenetic network and physicochemical properties of nonsynonymous mutations in the protein-coding genes of human mitochondrial DNA. *Mol Biol Evol* 20:1195–1210
10. Mohlke KL, Jackson AU, Scott LJ, Peck EC, Suh YD, Chines PS, Watanabe RM, Buchanan TA, Conneely KN, Erdos MR, et al (2005) Mitochondrial polymorphisms and susceptibility to type 2 diabetes-related traits in Finns. *Hum Genet* 118:245–254
11. Elson JL, Herrnstadt C, Preston G, Thal L, Morris CM, Edwardson JA, Beal MF, Turnbull DM, Howell N (2006) Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? *Hum Genet* 119:241–254
12. van der Walt JM, Nicodemus KK, Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Jonathan L. Haines, William C. Koller, Lyons K, et al (2003) Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet* 72:804–811
13. van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, Welsh-Bohmer KA, Saunders AM, Roses AD, Small GW, et al (2004) Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett* 365:28–32
14. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303:223–226
15. Lertrit P, Kapsa RM, Jean-Francois MJ, Thyagarajan D, Noer AS, Marzuki S, Byrne E (1994) Mitochondrial DNA polymorphism in disease: a possible contributor to respiratory dysfunction. *Hum Mol Genet* 3:1973–1981
16. Niemi AK, Moilanen JS, Tanaka M, Hervonen A, Hurme M, Lehtimaki T, Arai Y, Hirose N, Majamaa K (2005) A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. *Eur J Hum Genet* 13:166–170
17. Mackey DA, Oostra RJ, Rosenberg T, Nikoskelainen E, Bronte-Stewart J, Poulton J, Harding AE, Govan G, Bolhuis PA, Norby S (1996) Primary pathogenic mtDNA mutations in multigeneration pedigrees with Leber hereditary optic neuropathy. *Am J Hum Genet* 59:481–485
18. Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F, Carducci C, Leuzzi V, Carelli V, Barboni P, et al (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am J Hum Genet* 60:1107–1121
19. Man PY, Howell N, Mackey DA, Norby S, Rosenberg T, Turnbull DM, Chinnery PF (2004) Mitochondrial DNA haplogroup distribution within Leber hereditary optic neuropathy pedigrees. *J Med Genet* 41:e41
20. Carelli V, Achilli A, Valentino ML, Rengo C, Semino O, Pala M, Olivieri A, Mattiazi M, Pallotti F, Carrara F, et al (2006) Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. *Am J Hum Genet* 78:564–574
21. Torroni A, Campos Y, Rengo C, Sellitto D, Achilli A, Magri C, Semino O, Garcia A, Jara P, Arenas J, et al (2003). Mitochondrial DNA haplogroups do not play a role in the variable phenotypic presentation of the A3243G mutation. *Am J Hum Genet* 72:1005–1012
22. Houshmand M, Sharifpanah F, Tabasi A, Sanati MH, Vakilian M, Lavasani SH, Joughedoust S (2004) Leber's hereditary optic neuropathy: the spectrum of mitochondrial DNA mutations in Iranian patients. *Ann N Y Acad Sci* 1011:345–349

From the Mitochondrial Research Group (J.L.E.; P.F.C.) and Institute of Human Genetics (P.F.C.), The University of Newcastle upon Tyne, Newcastle upon Tyne; Department of Neurology, University of Turku, Turku, Finland (K.M.); and MIGENIX Corporation, San Diego (N.H.)

Address for correspondence and reprints: Dr. Joanna L. Elson, Department of Neurology, The Medical School, Newcastle University, Newcastle-upon-Tyne NE2 4HH, United Kingdom. E-mail: J.L.elson@ncl.ac.uk

*Am. J. Hum. Genet.* 2007;80:378. © 2006 by The American Society of Human Genetics. All rights reserved.  
0002-9297/2007/8002-0019\$15.00

## Reply to Elson et al.

### To the Editor:

We recently reported an association study of common mtDNA variation with risk of type 2 diabetes and metabolic traits.<sup>1</sup> We tested a simple model with few assumptions: that each common DNA sequence variant might, irrespective of its place in the mtDNA phylogeny, be associated with disease. Although Elson et al.<sup>2</sup> agree that our “approach will detect a robust disease association,” they have three concerns: (i) that the catalogue of common variation is not representative of European mtDNAs, (ii) that our tag SNPs do not adequately capture mtDNA haplogroups, and (iii) that we failed to consider epistasis. We believe these views to be misinterpretations of our article, as well as a reflection of a difference in philosophy on limiting association studies on the basis of prior assumptions.

We developed a catalogue of common mtDNA variation on the basis of publicly available mtDNA sequences. Perhaps we were insufficiently precise in the text, as Elson et al. misinterpret which samples were used. In the European alignment, we studied 928 mtDNAs of only European origin: 429 from MitoKor, 240 from the Armed Forces Institute of Pathology, 192 from Finland, and 67 from smaller studies. (Although we used 536 MitoKor sequences for