White Paper 23-05

Neanderthal Ancestry Inference

Authors:
Robin P. Smith
Aaron Kleinman
Kasia Bryc
Joanna Mountain
Eric Y.Durand

First Published: October 20, 2015
Updated: December 14, 2015
(This updated white paper replaces White Paper 23-05, titled "Neanderthal Ancestry Estimator" published on December 5, 2011 and last edited on January 8, 2012)

Summary

This white paper serves as a companion to the Neanderthal report offered to 23andMe customers as part of the 23andMe® Personal Genome Service. It offers more details on the methods used in calculating Neanderthal variant counts and in detecting human traits that are associated with Neanderthal variants. We hope that

this white paper may also serve the scientific community to understand how genetic variation inherited from Neanderthal ancestors may affect modern-day phenotypes.

Introduction

A fascinating topic in human evolution is the relationship between present-day humans and our closest evolutionary relatives, the now extinct *Homo neanderthalensis*, or Neanderthals. One of the most debated aspects of this relationship was whether and how much Neanderthals and modern humans interbred during the tens of thousands of years we cohabited in Europe and Western Asia.

Analysis of the first draft of the Neanderthal genome [1] led to the discovery that Neanderthals did indeed interbreed with modern humans. Researchers found that 1-4% of the genomes of all modern humans outside of Africa are of Neanderthal origin.

Subsequent research identified specific Neanderthal-derived alleles (variants) that are likely to have been introduced into modern humans [2]. Researchers detected these variants using rigorous statistical modeling, and it is believed that they are seen in present-day populations only because they were introduced by interbreeding with Neanderthals. Therefore, the observation of any of these variants in a living individual is indicative of the presence of DNA that came from Neanderthals.

Scientists at 23andMe further noticed that several of these Neanderthal variants are associated with traits in 23andMe customers. This white paper details not only our method for estimating the number of Neanderthal variants that a 23andMe customer carries, but also our approach to identifying the physical traits associated with those variants that are likely to have entered modern human populations through Neanderthals.

Methods

Counts of Neanderthal variants

We obtained, from the authors of [2], a list of 135,171 SNPs (Single Nucleotide Polymorphisms) for which the derived allele (the "Neanderthal variant") is believed to have originated in Neanderthals and later entered the modern human population via interbreeding. Of these, 1,436 SNPs are assayed via our genotyping chip and also pass our genotyping quality criteria. For each SNP, an individual carries 0, 1 or 2 Neanderthal variant copies. We report the total number of copies (the Neanderthal variant count) that the individual carries out of a possible 2,872. Although it is theoretically possible for someone to carry all 2,872 Neanderthal variant copies, most individuals carry only a subset. The average 23andMe customer carries 267.4 ± 47.3 Neanderthal variant copies (mean + SEM). The minimum and maximum Neanderthal variant copies that we've observed are 0 and 387, respectively.

A person's Neanderthal variant count is a signal that is likely to be correlated with the amount of Neanderthal DNA in their genome. That variant count can be compared with other 23andMe customers to determine what proportion carries more or less. The proportion of customers that has fewer Neanderthal variant copies is displayed in the report as a percentile and is reported along with the variant count.

Phenotypes associated with Neanderthal variants

By combining and analyzing data from our large database of customers who have consented to participate in research, web-based surveys, and custom genome-wide association study (GWAS) tools, 23andMe scientists routinely identify genetic associations for a variety of human traits. These analyses have been used to replicate associations reported in the literature and to publish novel associations in peer reviewed journals [3-13].

We examined our database of GWAS associations across 1,436 Neanderthal-associated SNPs genotyped by 23andMe. We focused on 310 phenotypes in European customers, and found five SNPs that are lead "hits" associated with four different traits and achieve genome-wide significance with a p-value cutoff of 5e-8 (see **Table 1**). Furthermore, by examining the sign of the estimated effect of the variant on the trait, we inferred the impact of carrying the Neanderthal variant on the phenotype in present-day individuals. In our Neanderthal report, we describe which of these Neanderthal variants a customer carries (if any), and the expected effect of each variant on the corresponding trait.

Table 1. Neanderthal variants significantly associated with traits

Phenotype	SNP	Neanderthal allele	Signed effect
Back hair	rs4849721	T	-0.066
Hair curl	rs12458349	G	-0.07
Height	rs7544462	C	-0.117
Height	rs1877547	A	0.074
Dark chocolate sneeze	rs11213819	T	-0.387

Discussion

The 23andMe Neanderthal report now offers 23andMe customers a detailed look at their Neanderthal ancestry by looking at a set of 1,436 Neanderthal-derived variants. While the Neanderthal report no longer provides a customer's overall "percent Neanderthal", it instead provides a detailed map of Neanderthal ancestry and GWAS-based clues regarding how this ancestry might influence traits. Although we are confident that the variants tested for in the Neanderthal report represent signals of Neanderthal ancestry, it is important to note that they represent only a fraction of the total amount of Neanderthal ancestry an individual carries.

Acknowledgements

We are grateful to Sriram Sankararaman, David Reich, and colleagues for providing the list of Neanderthal variants. We thank the customers of 23andMe for consenting to participate in this research and for answering surveys. We also thank all the employees of 23andMe, who together have made this research possible.

References

1. Green, R. E, Krause, J, Briggs, A. W, Maricic, T, Stenzel, U, Kircher, M, Patterson, N, Li, H, Zhai, W, Fritz, M. H. Y, Hansen, N. F, Durand, Eric Y, Malaspinas, A. S, Jensen, J. D, Marques-Bonet, T, Alkan, C, Prufer, K, Meyer, M, Burbano, H. A, Good, J. M, Schultz, R, Aximu-Petri, A, Butthof, A, Hober, B, Hoffner, B, Siegemund, M, Weihmann, A, Nusbaum, C, Lander, E. S, Russ, C, Novod, N, Affourtit, J, Egholm, M, Verna, C, Rudan, P, Brajkovic, Dejana, Kucan, Z, Gusic, I, Doronichev, V. B, Golovanova, L. V, Lalueza-Fox, C, Rasilla, M De La, Fortea, J, Rosas, A, Schmitz, R. W, Johnson, P. L. F, Eichler, E. E, Falush, D, Birney, E, Mullikin, J. C, Slatkin, M, Nielsen, R, Kelso, J, Lachmann, M, Reich, D, & Paabo, S. 2010. A Draft Sequence of the Neandertal Genome. Science, 328(5979), 710– 722.
2. Sriram Sankararaman, Swapan Mallick, Michael Dannemann, Kay Prüfer, Janet Kelso, Svante Pääbo, Nick Patterson, David Reich. The genomic landscape of Neanderthal ancestry in present-day humans. Nature (2014)
3. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, et al. Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. PLoS Genet. 2010;6: e1000993. doi:10.1371/journal.pgen.1000993
4. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, et al. Web-Based Genome-Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for Parkinson's Disease. PLoS Genet. 2011;7: e1002141. doi:10.1371/journal.pgen.1002141
5. Tung JY, Do CB, Hinds DA, Kiefer AK, Macpherson JM, Chowdry AB, et al. Efficient Replication of over 180 Genetic Associations with Self-Reported Medical Data. PLoS ONE. 2011;6: e23473. doi:10.1371/journal.pone.0023473

6. Eriksson N, Benton GM, Do CB, Kiefer AK, Mountain JL, Hinds DA, et al. Genetic variants associated with breast size also influence breast cancer risk. BMC Med Genet. 2012;13: 53. doi:10.1186/1471-2350-13-53
7. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, et al. Novel Associations for Hypothyroidism Include Known Autoimmune Risk Loci. PLoS ONE. 2012;7: e34442. doi:10.1371/journal.pone.0034442
8. Hinds DA, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet. 2013;45: 907–911. doi:10.1038/ng.2686
9. Kiefer AK, Tung JY, Do CB, Hinds DA, Mountain JL, Francke U, et al. Genome-Wide Analysis Points to Roles for Extracellular Matrix Remodeling, the Visual Cycle, and Neuronal Development in Myopia. PLoS Genet. 2013;9: e1003299. doi:10.1371/journal.pgen.1003299
10. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet. 2014;46: 989–993. doi:10.1038/ng.3043
11. Ferreira MAR, Matheson MC, Tang CS, Granell R, Ang W, Hui J, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol. 2014;133: 1564–1571. doi:10.1016/j.jaci.2013.10.030
12. Tung JY, Kiefer AK, Mullins M, Francke U, Eriksson N. Genome-Wide Association Analysis Implicates Elastic Microfibrils in the Development of Nonsyndromic Striae Distensae. J Invest Dermatol. 2013;133: 2628–2631. doi:10.1038/jid.2013.196
13. Heilmann S, Kiefer AK, Fricker N, Drichel D, Hillmer AM, Herold C, et al. Androgenetic Alopecia: Identification of Four Genetic Risk Loci and Evidence for the Contribution of WNT Signaling to Its Etiology. J Invest Dermatol. 2013;133: 1489–1496. doi:10.1038/jid.2013.43

Revision History

October 20, 2015

- This white paper reflects a fundamental change in the presentation and methodology for inferring Neanderthal ancestry. It replaces the previous 23andMe Neanderthal white paper published on December 5, 2011 (last edited on January 8, 2012), which addressed the calculation of a customer's percentage of Neanderthal DNA. The new method for inferring Neanderthal ancestry reflected in this white paper counts the number of genetic variants of Neanderthal origin and associations between these variants and a select group of traits.

December 8, 2015

- This minor revision clarifies several aspects of the methodology employed by the new Neanderthal ancestry report.