White Paper 23-03

Guidelines on Vetting Genetic Associations

Authors:

Andro Hsu
Brian Naughton
Shirley Wu

Created: November 14, 2007
Revised: February 14, 2008
Revised: June 10, 2010
(see end of document for summary of major changes)

Summary:

23andMe's goal is to bridge the gap between the scientific community and consumers interested in what their personal genetic information means. Although scientific consensus can and does change, we consider it our duty to present customers with information that is reliable and relatively stable (U.S. Government Accountability Office, 2006; Human Genetics Commission, 2003). To this end, we have prepared guidelines for choosing phenotypes and associations by

reviewing the scientific literature and consulting with expert advisors. Associations in the Health product that fully meet these criteria are labeled Established Research. Associations not meeting these criteria are labeled Preliminary Research, though the designation may change.

0 Executive Summary

1. **Choosing Phenotypes of Interest**
    - **Objective:** To choose phenotypes of scientific value and interest to report to the customer.
    - **Guidelines:**
        - Focus on common traits of broad interest to customers.
        - Preferentially report phenotypes having a prevalence of 1% or more.
        - Report on conditions inherited in a Mendelian manner when technologically possible.
        - Focus on genetic variants affecting responses to drugs and other molecular compounds.
2. **Avoiding False Discoveries**
    - **Objective:** To report genetic associations that are highly likely to be genuine, minimizing the possibility that future work will justify removal of data in topics designated as Established Research.
    - **Guidelines:**
        - Preferentially consider studies of sufficient statistical power to detect modest effects; generally, sample size should exceed 750 cases for discovery studies using common genome-wide single nucleotide polymorphism (SNP) arrays.
        - After correction for multiple hypothesis testing, p-values for associations should be less than 0.01.
        - Associations must be replicated in at least one independently published study.
            * P-value should be less than 0.01 after correction for multiple hypothesis testing.
            * 95% confidence intervals for the odds ratios (ORs) reported in each study must overlap.
            * Preferentially consider studies published in high-impact journals, especially if they are genome-wide association studies (GWASs).
            * Also consider replication studies and meta-analyses published in journals of lower impact, provided that they meet the above guidelines.

        * In cases of rare disease where multiple large independent studies are not expected a single, large consortium-based study or meta-analysis may be considered sufficient.

3. **Using Data From Multiple Studies**

    - **Objective:** To choose a single study's OR for use in calculating genotype-specific incidence, while minimizing "winner's curse."
    - **Guidelines:**
        - Use the OR from the study with the highest statistical power to detect the observed effect (typically the study with the largest sample size), with a preference towards more recent studies.

4. **Missing SNPs**

    - **Objective:** To take advantage of linkage disequilibrium (LD) structure in customer data to impute genotype at loci for which data are not available.
    - **Guidelines:**
        - Allow use of correlated SNP only if it is in complete LD with the absent SNP of interest.
        - Correlated SNP must be in complete LD in the HapMap population most closely corresponding to the subpopulation in which the association was confirmed.

5. **Associations and Ethnicity**

    - **Objective:** To report scientifically valid associations to customers in the context of customer-chosen ethnicity.
    - **Guidelines:**
        - Use the same criteria for associations in any population, with regard to sample size, significance, and replication, and OR confidence interval overlap.
            * When replicating an association in a population other than the one in which the association was originally confirmed, a single study satisfying the above criteria is sufficient.
        - Add or remove SNPs in calculation of genotype-specific incidence as appropriate for ethnicity selected by user.

6. **Reporting Preliminary vs. Established Research**

   - **Objective:** To provide customers with more information by reporting statistically significant associations from peer-reviewed research that has not yet passed sample size or replication criteria.
   - **Guidelines:**
     - Use standard criteria for statistical significance ($p < 0.05$).
     - Clearly designate the size of the sample and the population in which the association was observed.
     - Report odds with respect to the most common genotype for the SNP associated with the condition.
     - Consider associations of general scientific interest (for example, those found in non-European populations or in conditions that are not as widely studied).
     - Provide clear rankings as a proxy for the reliability of associations reported as Preliminary Research.

1 Choosing Phenotypes of Interest

Common diseases are a subject of intense study today. The productiveness of this field promises to increase the value of a customer's data more and more as scientific knowledge advances. Thus, a large part of our focus is on **common and complex diseases** caused by genetic variants that can be genotyped on our high-throughput platform. Complex, multifactorial diseases tend to be the type where genotype **influences the probability of an outcome**, instead of acting in a deterministic way (Pritchard, 2001).

Other types of phenotypes are of potential interest to customers, including genetically-influenced responses to drugs and conditions inherited in a direct Mendelian manner, where more concrete conclusions can be drawn from the data.

Summary

- **Objective:** To choose phenotypes of scientific value and interest to report to the customer as Established Research.
- **Guidelines:**

- Focus on common traits of broad interest to customers.
- Preferentially report phenotypes having a prevalence of 1% or more.
- Report on conditions inherited in a Mendelian manner when technologically possible.
- Focus on genetic variants affecting responses to drugs and other molecular compounds.

2 Avoiding False Discoveries

Several researchers have discussed the importance of adequate sample sizes for genome-wide association studies (GWASs) and sufficiently stringent thresholds for declaring genetic associations to be statistically significant ([Editorial], 2005; Ioannidis *et al.*, 2001; Freimer & Sabatti, 2004). These guidelines are now a part of the standard initial design of such studies (Altshuler & Daly, 2007). Even so, examples of false positive associations discovered in GWASs may still be expected (Herbert *et al.*, 2006; Thomas *et al.*, 2005).

Faced with what may be a changing landscape of methods for assessing the significance of genetic associations detected using GWASs, 23andMe's challenge is how to decide when a genetic association is significant enough to report to customers as broadly accepted by scientific consensus, minimizing the possibility that it will later be retracted. These associations are given the label Established Research.

Since many associations discovered by GWASs are likely to be of modest effect, 23andMe only reports results of studies of sufficient statistical power to detect these associations. The size of any given real association and the number of expected associations are not known a priori, which makes it difficult to set concrete thresholds of statistical power and sample size, but a good rule of thumb is that GWASs should have **at least 750 cases and appropriately chosen controls** to be considered. Statistical power to detect an association may also be calculated retroactively for a GWAS and used to judge whether the sample size was appropriately large.

Associations should reach high statistical significance to be included for consideration ($P < 0.01$ after correction). The nature of GWASs requires that p-values be corrected for multiple hypothesis testing. Bonferroni correction is a conservative way to correct p-values in widespread use in GWASs (Thomas *et al.*, 2005). We would also accept other, less conservative methods of correction with well-

described methodology (Benjamini & Hochberg, 1995; Storey & Tibshirani, 2003; Manly *et al.*, 2004).

In consultation with our scientific advisory board, we would also consider the use of the Bayes factor (BF) as a method of assessing evidence for the believability of an association. The Wellcome Trust Case Control Consortium's (WTCCC) landmark study was remarkable for including both p-values and BFs for associations (Wellcome Trust Case Control Consortium, 2007). For studies providing Bayes factors, using the WTCCC's estimate of prior odds at 100,000:1 against association suggests that log(BF) must reach 6-7 to result in a 90-99% probability that the association is real.

Even so, we do require associations to be **replicated in an independent study**, satisfying the same significance criteria in each. Some groups screen for weak associations in initial discovery sample and confirm associations in larger replication samples within the same study. However, the discovery sample may be statistically underpowered, so that replication and subsequent pooling to achieve statistical power should really only be counted as a single observation.

In deference to expert opinion, we generally limit reporting to GWASs **published in high-impact journals**, including: *Science*, *Nature*, *Nature Genetics*, *PLoS Genetics*, *PLoS Biology*, *PLoS Medicine*, *PNAS*, the *American Journal of Human Genetics*, the *New England Journal of Medicine*, the *Journal of the American Medical Association*, and *The Lancet*. Publication in any of these journals shows that the studies have passed high standards of design methodology.

However, we actively consider and take into account replication studies published in journals of lower impact, since these studies may be performed to high standards but are considered less novel. Replications that reach appropriately corrected significance and that are confirmed in samples of sufficient statistical power can often be found in journals specializing in the disease or phenotype in question—in the case of replication, we do not hold the reputation of the journal against the quality of the evidence. We may also choose to report meta-analyses published in high quality epidemiology journals such as the *American Journal of Epidemiology*.

Summary

- **Objective:** To report genetic associations that are highly likely to be genuine, minimizing the possibility that future work will justify removal of data in topics designated as Established Research.

- **Guidelines:**
  - Preferentially consider studies of sufficient statistical power to detect modest effects; generally, sample size should exceed 750 cases for discovery studies using common genome-wide single nucleotide polymorphism arrays.
  - After correction for multiple hypothesis testing, p-values for associations should be less than 0.01.
  - Associations must be replicated in at least one independently published study.
    * P-value should also be less than 0.01 after correction for multiple hypothesis testing.
    * 95% confidence intervals for the odds ratios (ORs) reported in each study must overlap.
    * Preferentially consider studies published in high-impact journals, especially if they are GWASs.
    * Also consider replication studies and meta-analyses published in journals of lower impact, provided that they meet the above guidelines.
    * In cases of rare disease where multiple large independent studies are not expected a single, large consortium-based study or meta-analysis may be considered sufficient.

3 Using Data From Multiple Studies

Our method of calculating genotype-specific incidence for a disease phenotype requires allele- or genotype-specific odds ratios (ORs) for each SNP known to be associated with risk of disease. (The method is fully detailed in 23andMe White Paper 23-01.) Ideally, the OR reported by any given study would approximate the"true" OR for the broader population of which the sample population is part.

There are several types of bias that may result in significant differences between a study's reported OR and the true effect size. These include publication bias due to the fact that negative results are less likely to be published, and the "winner's curse," a tendency for the initial study on an association to overstate its effect (Ioannidis *et al.*, 2001; Lohmueller *et al.*, 2003; Zollner & Pritchard, 2007). Additionally, the sizes of effects discovered in GWASs are often modest, and associations may have different true effects in the populations used in each study.

In practice, we use the OR from a single study instead of conducting meta-analysis. The challenge facing 23andMe is how to decide which study's OR to use when calculating genotype-specific incidence.

The winner's curse is more pronounced in studies with very low statistical power to detect a modest effect. Because of our requirement that associations we report be confirmed in large samples, we expect that winner's curse will be minimized. Since we will generally choose an OR from two or more studies, we preferentially use the OR of the study with the **most power to detect the observed effect** (in general, the study with the largest sample size), with a preference towards more recent studies.

Summary

- **Objective:** To choose a single study's OR for use in calculating genotype-specific incidence, while minimizing "winner's curse."
- **Guidelines:**
  - Use the OR from the study with the highest statistical power to detect the observed effect (typically the study with the largest sample size), with a preference towards more recent studies.

4 Missing SNPs

The genotyping panel we currently use covers much of the common variation in the genome (Eberle *et al.*, 2007). Occasionally, SNPs found to be significantly associated with a phenotype will not be present on our genotyping platform. Additionally, a small fraction of SNPs will have poor data quality, preventing us from determining a customer's genotype at those loci. If this type of error occurs at a SNP we report as associated with a phenotype, we would like to be able to use information from linked SNPs to impute the customer's genotype at the original locus (and thus the contribution of that locus to the customer's genotype-specific incidence).

We are able to estimate linkage disequilibrium (LD) between SNPs for which HapMap data are available. If the original SNP reported in an association study is unavailable to us for whatever reason, our policy is to **use correlated SNPs that are in complete LD** with the associated SNP **in the population in which the as-**

sociation was confirmed. This procedure allows robust reporting of associations to the customer while assuring validity.

Summary

- **Objective:** To take advantage of LD structure in customer data to impute genotypes at loci for which data are not available.
- **Guidelines:**
  - Allow use of correlated SNP only if it is in complete LD with the absent SNP of interest.
  - Correlated SNP must be in complete LD in the HapMap population most closely corresponding to the subpopulation in which the association was confirmed.

5 Associations and Ethnicity

An association must be verified in every population for which we report it. Since linkage patterns are not identical between populations of different ethnicity, a highly associated SNP linked to an unknown causal variant in one population may not be linked to the causal variant in another. In a world with unlimited research funding, GWASs would identify a genetic association in a sample of one ethnicity and attempt to confirm the association in samples of a different ethnicity or in large multi-ethnic cohorts.

We use the **same vetting criteria for associations in all populations**, specifically with regard to minimum sample size, statistical significance, and confirmation in an independent study. We do not consider evidence of association in one population as evidence for the association in a different population. Once an association has been confirmed in one population, however, a single study may be sufficient to confirm the association in a different ethnic population providing it satisfies the criteria above with regard to sample size and statistical significance, and that the confidence intervals overlap.

23andMe's customers are allowed to select ethnicity when viewing genotype-specific incidence. (A full discussion of ethnicity and estimating genotype-specific incidence can be found in 23andMe White Paper 23-02.) A population is only presented as a choice if there is at least one SNP that passes our criteria for association with a phenotype (and if population-specific data on disease incidence is

available). When selecting an ethnicity, **only SNPs independently passing our criteria for association in samples of that ethnicity are included in the calculation** of genotype-specific incidence. The customer is notified which SNPs are included in the calculation.

Summary

- **Objective:** To report scientifically valid associations to customers in the context of customer-chosen ethnicity.
- **Guidelines:**
  - Use the same criteria for associations in any population, with regard to sample size, significance, and replication, and OR confidence interval overlap.
    - When replicating an association in a population other than the one in which the association was originally confirmed, a single study satisfying the above criteria is sufficient.
  - Add or remove SNPs in calculation of genotype-specific incidence as appropriate for ethnicity selected by user.

6 Reporting Preliminary vs. Established Research

Scientific knowledge about associations between genetic variants and human health and traits is constantly evolving. Our strict criteria can prevent us from reporting on new information as Established Research, but the findings are often valid within the context of the published studies and of general scientific interest. It is also more difficult for associations in non-European populations and rarer phenotypes to become Established Research as the studies tend to be smaller and are not as common. This leaves many gaps in our coverage of genetically influenced conditions.

We therefore report on associations that have not yet passed all of the Established Research criteria, and designate these as Preliminary Research. Each association must still pass the requirement for statistical significance after correction for multiple hypothesis testing, but studies may be smaller and are not yet independently replicated. For each association, 23andMe provides information on the relevant genetic marker, the study size, the population used, and the odds of the

condition associated with each genotype in the context of the study. Odds are reported with respect to the most common genotype, which can be distinct from the lowest risk genotype often used in published research.

Preliminary Research reports group together associations pertaining to a phenotype and are ranked according to the sample size of the largest study included in the report. We use sample size as a proxy for the reliability of the research contained in the report. A report containing an association derived from a sample size of at least 750 cases is given three gray stars, 100 to 750 cases is given two gray stars, and fewer than 100 cases is given one gray star. The associations in Preliminary Research reports are not combined into a single odds estimate, but are presented as individual pieces of information. Established Research reports, which typically contain associations that have been independently replicated in large studies, are given a ranking of four gold stars.

The inclusion of Preliminary Research allows us to provide even more information to our customers while making clear that the associations have not yet been confirmed in large studies or replicated independently. It also allows us to report on conditions that are not as widely studied as the common diseases and to report on more research relevant to non-European populations.

Summary

- **Objective:** To provide customers with more information by reporting statistically significant associations from peer-reviewed research that has not yet passed sample size or replication criteria.
- **Guidelines:**
  - Use standard criteria for statistical significance ($p < 0.05$).
  - Clearly designate the size of the sample and the population in which the association was observed.
  - Report odds with respect to the most common genotype for the SNP associated with the condition.
  - Consider associations of general scientific interest (for example, those found in non-European populations or in conditions that are not as widely studied).
  - Provide clear rankings as a proxy for the reliability of associations reported as Preliminary Research.

7 Summary of major changes

Last revision: June 10, 2010.
Previous revision: Feb 14, 2008.

- Added drug response and common traits to Phenotypes of Interest. (Section 1)
- Minimum sample size criteria for Established Research lowered from 1,000 cases and 1,000 controls to 750 cases. (Section 2)
- P-value criteria modified from more general "genome-wide significance" to more specific "0.01 after correction for multiple hypothesis testing." (Section 2)
- Added criteria regarding overlap of confidence intervals for ORs in replication studies. (Section 2)
- Added exception to independent replication for rare diseases where a single, large, consortium-based study or meta-analysis is likely to be the only study of sufficient size. (Section 2)
- An association may be considered replicated in additional populations other than the one in which the association was originally confirmed using a single large study. (Section 5)
- Added section describing the inclusion of Preliminary Research and rankings for Preliminary Research reports. (Section 6)

References

Altshuler, D, & Daly, M. 2007. Guilt beyond a reasonable doubt. *Nat Genet*, **39**(7), 813–815.

Benjamini, Y, & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B*, **57**(1), 289–300.

Eberle, M A, Ng, P C, Kuhn, K, Zhou, L, Peiffer, D A, Galver, L, Viaud-Martinez, K A, Lawley, C T, Gunderson, K L, Shen, R, & Murray, S S. 2007. Power to detect risk alleles using genome-wide tag SNP panels. *PLoS Genet*, **3**(10), 1827–1837.

[Editorial]. 2005. Framework for a fully powered risk engine. *Nat Genet*, **37**(11), 1153–1153.

Freimer, N, & Sabatti, C. 2004. The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nat Genet*, **36**(10), 1045–1051.

Herbert, A, Gerry, N P, McQueen, M B, Heid, I M, Pfeufer, A, Illig, T, Wichmann, H E, Meitinger, T, Hunter, D, Hu, F B, Colditz, G, Hinney, A, Hebebrand, J, Koberwitz, K, Zhu, X, Cooper, R, Ardlie, K, Lyon, H, Hirschhorn, J N, Laird, N M, Lenburg, M E, Lange, C, & Christman, M F. 2006. A common genetic variant is associated with adult and childhood obesity. *Science*, **312**(5771), 279–283.

Human Genetics Commission. 2003. *Genes Direct: Ensuring the Effective Oversight of Genetic Tests Supplied Directly to the Public*. London, U.K.: Department of Health.

Ioannidis, J P, Ntzani, E E, Trikalinos, T A, & Contopoulos-Ioannidis, D G. 2001. Replication validity of genetic association studies. *Nat Genet*, **29**(3), 306–309.

Lohmueller, K E, Pearce, C L, Pike, M, Lander, E S, & Hirschhorn, J N. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*, **33**(2), 177–182.

Manly, K F, Nettleton, D, & Hwang, J T. 2004. Genomics, prior probability, and statistical tests of multiple hypotheses. *Genome Res*, **14**(6), 997–1001.

Pritchard, J K. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*, **69**(1), 124–137.

Storey, J D, & Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*, **100**(16), 9440–9445.

Thomas, D C, Haile, R W, & Duggan, D. 2005. Recent developments in genomewide association scans: a workshop summary and review. *Am J Hum Genet*, **77**(3), 337–345.

U.S. Government Accountability Office. 2006. *Nutrigenetic Testing: Tests Purchased from Four Web Sites Mislead Consumers*. Washington, D.C.: United States Government Accountability Office. GAO-06-977T.

Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**(7145), 661–678.

Zollner, S, & Pritchard, J K. 2007. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*, **80**(4), 605–615.