Description of the "bins" used in this analytic framework

The bin categories can be briefly summarized as follows:

- Bin 1 contains genes in which the discovery of a mutation would trigger specific medical action and provide definable medical benefit (i.e., clinical utility or defined clinical actionability). Mutations in Bin 1 genes would be returned to patients by default, since the benefits of learning this information would be expected to outweigh the potential harms of the revelation of such a disorder.
- Bin 2 contains genes known to be associated with human disease or disease risk but for which there are no specific guidelines or management changes that could mitigate the risk or manifestations of the disease (i.e., Bin 2 genes demonstrate clinical validity but not clinical utility). Mutations in these genes vary widely in terms of their potential to cause psychosocial harm if revealed unexpectedly to an inadequately prepared individual. Therefore, Bin 2 is further stratified into Bin 2a, Bin 2b, and Bin 2c based on a likely increasing potential for harm. In the absence of definable medical benefits that could be gained by the revelation of mutations associated with such disorders, we propose that the return of Bin 2 findings be carefully guided by an appropriately qualified clinician, in the context of that individual's personal situation, or not reported at all.
- Bin 3 contains all other genes whose roles in human disease are undefined. Thus, incidental variants in these genes would not be utilized in a medical context.

Manual review of variants identified by the "binning" algorithm

A central argument for the proposed binning strategy is to reduce the number of variants that require human analysis to a tractable number, due to the complexities of such manual deliberations. The computational "binning" analysis selected 1391 variants among the 80 genomes that met the criteria of having <5% allele frequency and being either truncating or present in the Human Gene Mutation Database. These variants were manually curated to assess their likely pathogenicity. Some of the variants are likely to be true positive findings, since a small number of mutations identified among the 19 participants in the hereditary cancer susceptibility research study appear to explain some of the family histories of cancer (manuscript in preparation). These mutations, which would also be detected in a targeted diagnostic analysis, demonstrate the ability of this screening algorithm to identify likely disease-causing mutations. However, there were also findings among the 61 presumably healthy individuals that would be highly suggestive of an unexpected Mendelian disorder.

- For example, the frameshifting insertion in the *MSH2* gene (c.592dupG) detected in subject NA12883 is highly suggestive of Lynch syndrome, and indeed this mutation was previously reported in a patient with hallmarks of the Muir-Torre subgroup of this cancer syndrome [1].
- On the other hand, some truncating mutations may have less obvious clinical consequences. For example, a frameshifting insertion in the *COL3A1* gene was detected in subject NA18956. Heterozygous mutations in *COL3A1* are known to cause Ehlers-Danlos syndrome type IV (the "vascular" subtype), which is associated with a high risk for vascular dissection and rupture of visceral organs [2]. However 95% of disease causing mutations are missense alterations of glycine residues comprising the triple helical domain, and the implications of this predicted truncating mutation are likely to be somewhat less severe and less predictable [3].

- The HGMD query also facilitated the identification of reported disease-causing missense mutations. For example, subject NA19648 (who is of Mexican-American origin) was found to be a heterozygous carrier of the Q188R mutation in *GALT* gene, which results in classical galactosemia when present in a homozygous state [4]. The Q188R allele accounts for a substantial fraction of mutant alleles in Caucasians and was the most frequently observed mutation in a sample of 19 Mexican galactosemia patients [5].

There were also several scenarios that prompted reclassification of variants, as described below and in Table 2. Some cases involved two variants identified in an individual that comprised either a single complex substitution allele that was not resolved by the variant calling algorithm, or two variants in *cis* that comprise a single common haplotype:

- One example of variants comprising a single complex substitution was found in subject NA12878, in whom two single base indels were found in the *KIF1B* gene associated with a form of autosomal dominant Charcot-Marie-Tooth disease. The first of these variants was an insertion and the second was a deletion several nucleotides later that was most likely in *cis* and restored the reading frame. The translated protein sequence would result in two missense amino acid substitutions, which would be interpreted as a variant of uncertain significance and therefore would not be reported as an incidental finding. This example depicts the difficulty inherent in calling certain types of variants informatically.
- In some cases, two variants were identified in a gene associated with an autosomal recessive condition, which would predict that the individual would be affected; however, on further review the two variants were considered likely to be present in *cis*. For example, the P369S and R408Q variants in the *MEFV* gene associated with Familial Mediterranean Fever are reported as “DM” in HGMD and were found together in NA18504, NA18526, and NA18947; these variants were

previously reported to always be in *cis* and are associated with a highly variable phenotype [6].

- In other cases, different types of mutations in the same gene cause different conditions with distinct modes of inheritance. For example, subject NA18947 was found to have a predicted truncating mutation in the *SLC34A1* gene. It has been previously demonstrated that homozygous loss of function mutations in this gene cause Fanconi renotubular syndrome [7], whereas heterozygous missense mutations are the cause of autosomal dominant nephrolithiasis [8]. Thus, the finding of a single predicted truncating mutation would imply only carrier status for Fanconi renotubular syndrome.

In many cases, variants annotated as "DM" in HGMD were found to be VUS or likely polymorphisms:

- A *BRCA2* missense variant (c.6347A>G; p.H2116R) was identified in subject NA19025, as well as two members of a Yoruban trio (NA19239 and NA19240). This variant is annotated as "DM" in HGMD, but is unlikely to be pathogenic based on review of the literature [9].
- Four subjects (NA18502, NA18504, NA19704 and NA21767) were found to have a *TSC2* missense variant (c. 3986G>A; p. R1329H; rs45517323) that is classified as "DM" in HGMD. This variant was present in >1% of all 1000 Genomes Population alleles (and in higher percentages in those of African ancestry) and was previously considered to be of uncertain significance [10].

Finally, we found examples in which the type of variant or its location within a specific transcript was inconsistent with a pathogenic effect:

- For example, subject NA12878 was found to have a frameshifting single base deletion in the *HTT* gene, which is of uncertain clinical significance since Huntington's disease is known to be caused by trinucleotide repeat expansion [11] and not loss of function mutations.

- Simlarly, subject NA19700 was found to have a frameshifting single base insertion in the *APP* gene. Since the reported pathogenic mutations in the *APP* gene are missense mutations causing toxic gain of function [12], this putative loss of function variant is of uncertain clinical significance.

Supplemental References:

1. Mangold E, Pagenstecher C, Leister M, Mathiak M, Rütten A, Friedl W, Propping P, Ruzicka T, Kruse R: **A genotype-phenotype correlation in HNPCC: strong predominance of MSH2 mutations in 41 patients with Muir-Torre syndrome.** *J Med Genet* 2004, **41**:567-572.
2. Pepin M, Schwarze U, Superti-Furga A, Byers PH: **Clinical and genetic features of Ehlers-Danlos syndrome type IV, the vascular type.** *N Engl J Med* 2000, **342**:673-680.
3. Leistritz DF, Pepin MG, Schwarze U, Byers PH: **COL3A1 haploinsufficiency results in a variety of Ehlers-Danlos syndrome type IV with delayed onset of complications and longer life expectancy.** *Genet Med* 2011, **13**:717-722.
4. Tyfield L, Reichardt J, Fridovich-Keil J, Croke DT, Elsas LJ 2nd, Strobl W, Kozak L, Coskun T, Novelli G, Okano Y, Zekanowski C, Shin Y, Boleda MD: **Classical galactosemia and mutations at the galactose-1-phosphate uridyl transferase (GALT) gene.** *Hum Mutat* 1999, **13**:417-430.
5. Velázquez-Aragón J, Alcántara-Ortigoza MA, Vela-Amieva M, Monroy S, Martínez-Cruz V, Todd-Quiñones C, González-del Angel A: **Low allelic heterogeneity in a sample of Mexican patients with classical galactosaemia.** *J Inherit Metab Dis* 2008, **31**:S333-337.
6. Ryan JG, Masters SL, Booty MG, Habal N, Alexander JD, Barham BK, Remmers EF, Barron KS, Kastner DL, Aksentijevich I: **Clinical features and functional significance of the P369S/R408Q variant in pyrin, the familial Mediterranean fever protein.** *Ann Rheum Dis* 2010, **69**:1383-1388.

7. Magen D, Berger L, Coady MJ, Ilivitzki A, Militianu D, Tieder M, Selig S, Lapointe JY, Zelikovic I, Skorecki K: **A loss-of-function mutation in NaPi-IIa and renal Fanconi's syndrome.** *N Engl J Med* 2010, **362**:1102-1109.
8. Prié D, Huart V, Bakouh N, Planelles G, Dellis O, Gérard B, Hulin P, Benqué-Blanchet F, Silve C, Grandchamp B, Friedlander G: **Nephrolithiasis and osteoporosis associated with hypophosphatemia caused by mutations in the type 2a sodium-phosphate cotransporter.** *N Engl J Med* 2002, **347**:983-991.
9. Haffty BG, Silber A, Matloff E, Chung J, Lannin D: **Racial differences in the incidence of BRCA1 and BRCA2 mutations in a cohort of early onset breast cancer patients: African American compared to white women.** *J Med Genet* 2006, **43**:133-137.
10. Rendtorff ND, Bjerregaard B, Frödin M, Kjaergaard S, Hove H, Skovby F, Brøndum-Nielsen K, Schwartz M; Danish Tuberous Sclerosis Group: **Analysis of 65 tuberous sclerosis complex (TSC) patients by TSC2 DGGE, TSC1/TSC2 MLPA, and TSC1 long-range PCR sequencing, and report of 28 novel mutations.** *Hum Mutat* 2005, **26**:374-383.
11. Huntington's Disease Collaborative Research Group: **A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.** *Cell* 1993, **72**:971-983.
12. Goate AM: **Monogenetic determinants of Alzheimer's disease: APP mutations.** *Cell Mol Life Sci* 1998, **54**:897-901.