

OPINION

An agenda for personalized medicine

Pauline C. Ng, Sarah S. Murray, Samuel Levy and J. Craig Venter find differences in results from two direct-to-consumer genetics-testing companies. They therefore give nine recommendations to improve predictions.

More than 1,000 DNA variants associated with diseases and traits have been identified^{1,2}. Direct-to-consumer (DTC) companies are harnessing these discoveries by offering DNA tests that provide insights into personal genetic traits and disease risks. Genetic testing can improve lifestyle choices and increase preventive screening³. However, understanding of the genetic contribution to human disease is far from complete.

There is debate in the genetics community as to the usefulness of DTC testing. Therefore, we compared results from two DTC companies (test kits provided by genomics companies 23andMe in Mountain View, California, and Navigenics in Foster City, California) on 13 diseases for 5 individuals. Despite this limited data set we find potential implications for personalized medicine. Here we provide recommendations to improve predictions and support the continued growth of this nascent industry.

DTC genome scans are easy to get. Users order tests online, provide saliva or a cheek swab, and within a few weeks 500,000–1,000,000 of their DNA variant markers are scanned. The service provider then calculates a set of disease risks based on the customer's specific combination of markers, and presents the results to the user online (see graphic).

The accuracy of DTC genome-scan tests has been questioned. It is our assessment that the accuracy of the raw data is high. We found that the genotypes, or particular DNA bases observed, of an individual's markers from 23andMe and Navigenics agreed more than 99.7% of the time. This is similar to accuracies reported by the genotyping companies.

Two other major concerns are whether the predicted disease risks have any clinical validity, and how well a genetic variant correlates with a specific disease or condition⁴. A few individuals have alluded to getting different predictions from different DTC companies for the same disease^{5,6}. We compared the consistency of disease-risk predictions between the two DTC companies to see where differences may arise (see Table 1).

Both companies report absolute risk, which is the probability that an individual will develop a disease. Absolute risk is derived from two parameters: 'relative risk' and 'average population disease risk'. Relative risk is modelled from an individual's genetics. Average population disease risk varies depending on how one defines the population. For example, Navigenics distinguishes population disease risk between men and women (for example, men are more likely to have heart attacks than women), whereas 23andMe primarily takes into account age (for example, incidence of rheumatoid arthritis increases with age). This ambiguity in the definition of a 'population' underscores the caution one must exercise when interpreting absolute risk results.

Even after we removed the average population risk variable we still found that only two-thirds of relative risk predictions qualitatively agree between 23andMe and Navigenics when averaged across our five individuals (see Table 1). Certain diseases have better prediction agreement than others. For four diseases, the predictions between the two companies completely agree for all individuals. In contrast, for seven diseases, 50% or less of the predictions

SUMMARY

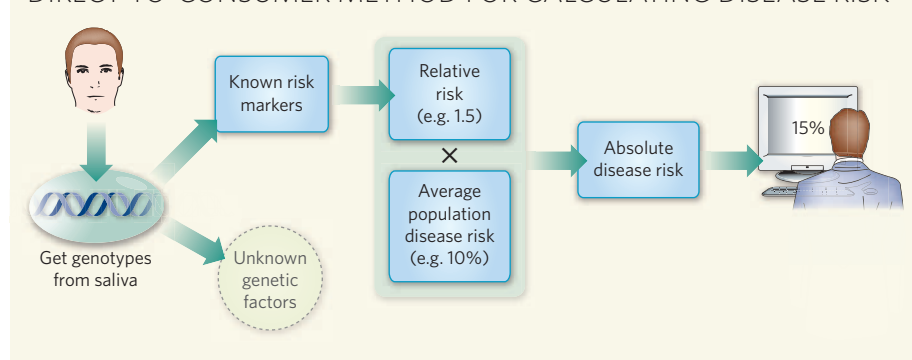
- For seven diseases, 50% or less of the predictions of two companies agreed across five individuals
- Companies should communicate high risks better and test for drug response markers
- Community should study markers in all ethnicities and look at behaviour after tests

agree between the two companies across the individuals.

A major contributor to the discrepancies in disease-risk predictions is the set of markers that each service chooses to use in calculating relative risk. Risk markers are determined from genome-wide association studies, which survey hundreds of thousands or millions of markers across control and disease patients⁷. Each marker has different possible alleles. Alleles that occur more frequently in disease patients are designated as risk alleles and have odds ratios greater than 1. For example, in Alzheimer's disease patients, 38% of *ApoE* alleles are the *ApoE4* risk allele; this allele's frequency is only 14% in normal controls⁸. The odds ratio for the *ApoE4* risk allele is 3.7 (odds of exposure in cases, divided by odds of exposure in controls is $(0.38/0.62)/(0.14/0.86)$). The greater the frequency disparity between disease patients and normal controls, the higher the odds ratio associated with the allele. Conversely, alleles conferring protection against disease are observed less frequently in disease patients and have odds ratios less than 1.

DTC companies harness the same publicly available research to decide which markers to include, and for the most part, could use the same or similar markers. Yet no disease has an identical set of markers between the two DTC companies because each company has its own criteria for accepting a genome-wide association result into its relative risk calculation^{9,10}. Some markers are used by both companies for a particular disease. For identical markers and correlated markers, the odds ratios are similar between the two DTC companies ($r=0.98$ for identical markers; $r=0.89$ for correlated markers). In other words, once DTC companies agree that a marker is predictive of disease, they tend to agree on its genetic contribution to disease predictions.

DIRECT-TO-CONSUMER METHOD FOR CALCULATING DISEASE RISK



What gives two similar DTC predictions? Coeliac disease is one condition for which predictions agreed between the two companies for all five individuals in our analysis. For coeliac disease, both companies have one strong-effect marker with a high odds ratio in common; Navigenics also reports on seven additional markers that 23andMe does not use. Thus, the number of markers in common does not necessarily correlate with better prediction agreement. However, the one strong-effect marker in common between both companies occurs in >90% of people with coeliac disease¹¹ and its risk allele has an odds ratio of 7 (ref. 12). The seven markers unique to Navigenics have modest effects¹² and therefore do not affect the overall relative risk prediction for this disease as much. Generally predictions tended to agree when there was consensus on the strong-effect markers for a disease.

When the DTC companies did not use the same strong-effect markers, we saw large differences in prediction. A clear example is the predicted disease risk for psoriasis. In one individual, 23andMe reports a relative risk of 4.02, whereas Navigenics reports a relative risk of 1.25, more than a threefold difference. The difference is attributable to a marker unique to 23andMe whose risk allele has an odds ratio of 2.8 (ref. 13). This marker is not included in the Navigenics analysis because the result does not seem to pass Navigenics' publication requirements for marker inclusion.

Another concern is the use of markers that have uncertain odds ratios estimates. A marker for type 2 diabetes that Navigenics uses has the highest odds ratio among all of Navigenics' type 2 diabetes markers as reported in the literature. It therefore makes the strongest contribution to the overall disease prediction. However, Navigenics warns that the marker's effect is statistically insignificant and may not contribute to disease. The average consumer is unlikely to appreciate the significance, or lack thereof, of this result.

These findings lead us to propose the following recommendations for a personalized medicine research agenda.

Company recommendations

Report the genetic contribution for the markers tested. Currently, the markers that have been discovered by genome-wide association studies do not explain the majority of the genetic heritability of disease. For example,

TABLE 1: PREDICTIONS FOR DISEASE RELATIVE RISKS FOR FIVE INDIVIDUALS

Disease	Female A	Female B	Female C	Male D	Male E
Breast cancer	↑↑	↑↑	↓↓		
Coeliac disease	↓↓	↓↓	↓↓	↓↓	↓↓
Colon cancer	==	==	=↓	↑↑	=↓
Crohn's disease	↓↑	↓↑	↓↓	↓↓	↓=
Heart attack	↓↓	=↓	=↓	=↓	↑↑
Lupus	↑↓	↓↓	↓↓	↑=	↑=
Macular degeneration	↓↓	↓↓	↑=	↓↓	↓↓
Multiple sclerosis	↑↑		↓↓	↓↓	↓↓
Prostate cancer				↑↑	↓↑
Psoriasis	↓↑		↑↓	↑↑	↓↓
Restless legs syndrome	=↓	↑↑	↓=	↓↑	↑↑
Rheumatoid arthritis	↑↑	↑↑	↓↓	↓↓	↑↑
Type 2 diabetes	↓↓	=↓	↓↓	↑↓	=↓

↑ increased risk (RR > 1.05), ↓ decreased risk (relative risk (RR) < 0.95), = average risk (0.95 ≤ RR ≤ 1.05). First prediction is from 23andMe; second prediction is from Navigenics. Different predictions are highlighted in beige.

current literature indicates that approximately 60–65% of the heritability of coeliac disease is still unaccounted for. Therefore, the marker set used to screen for disease can miss unknown genetic factors, leading to false negatives. We recommend that DTC companies report the proportion of the genetic contribution of a disease that can be attributed to the markers used in their test, and the proportion of the genetic contribution that is still unknown. This is different from reporting the genetic contribution versus the environmental contribution, which DTC companies emphasize on their websites.

Focus on high-risk predictions. Most of the diseases predicted in the DTC reports imply only a modest risk compared with the average population (approximately 80% of reported relative risks lie between 0.5 and 1.5). We recommend that DTC companies structure their communications with users around diseases and traits that have high-risk predictions. Customers could focus their lifestyle changes based on these. However, if there is a low-risk prediction for disease, a sense of security should not be assumed because much of the genetic contribution to disease risk has yet to be understood.

Directly genotype risk markers. If the risk marker in the published literature is not directly assayed by the DTC company, DTC companies currently use linkage disequilibrium (the non-random association of alleles) to choose a surrogate risk marker. We

estimate that 1% of markers in an individual will violate the assumption of perfect linkage disequilibrium, and in these cases, using a surrogate marker would be misleading. Although this percentage seems small, hundreds of markers are tested so there is likely to be at least one error. Instead, direct genotyping of the disease-associated marker would improve accuracy of risk genotypes. Some DTC companies already target specific key markers not on the whole genome arrays or specific markers that have failed on the whole genome arrays¹⁰.

Test pharmacogenomic markers.

An estimated 100,000 people die annually in the United States from adverse drug reactions^{14,15}. Although few drugs are labelled to require or recommend genetic testing, consumers could find specific variants useful. Variants in drug metabolism genes or recommended for test-

ing by drug labels are informative and could greatly affect an individual's treatment. Examples include variants affecting the efficacy of clopidogrel (used to reduce the risk of stroke or heart attack) or tamoxifen (used to treat breast cancer)¹⁶. Most of the DTC companies are testing for some pharmacogenomic markers^{17,18}; we encourage inclusion of as many of these markers as possible.

Agree on strong-effect markers. DTC companies have agreed to use clinically validated markers for prediction, but not necessarily the same markers or number of markers¹⁹. This lack of consensus leads to inconsistent results between DTC companies. As studies are replicated, the number of markers and better estimates of their odds ratios should converge so that there is consensus to include a marker. Because these studies will take time, a stopgap solution is for DTC companies to agree on using a core set of strong-effect markers to achieve better prediction consensus and consistent reporting to the consumer.

Community recommendations

Monitor behavioural outcomes. One of the fundamental questions with DTC tests is whether they modify consumers' behaviour long term, and hence benefit lifestyle and health²⁰. More public studies need to be funded to monitor behaviour resulting from DTC testing to identify the best strategies for using personal genomic data to improve an individual's health. Studies are currently under

way, and applying the findings will be central to the success of DTC genome tests and credibility of the field²⁰.

Carry out prospective studies. Agreement on risk predictions by DTC companies does not necessarily imply that the predictions are accurate or meaningful, and at this point in time, we cannot determine who has the 'best' predictions. To effectively assess the clinical validity of these genetic tests the community needs more prospective studies with tens or hundreds of thousands of individuals that measure the predictive value of known markers^{21–23}. Such studies are useful because they consider risk markers simultaneously, measure the interaction between different markers and do not assume a risk model. It may be practical to prioritize common diseases with significant health impact because of the large numbers of individuals and the expense associated with prospective studies.

Replicate associated markers in other ethnicities. Genome-wide association studies have been conducted primarily on populations with European ancestry¹. Disease-associated markers may not transfer from one population to another — allele frequencies or linkage disequilibrium patterns may differ^{1,24}. Therefore, we strongly recommend the validation of these markers and the surrounding pattern of genetic variation in other ethnicities.

Sequence rather than genotype. Eventually, sequencing an individual's genome will become economically feasible. Sequencing has an advantage over genotyping because it captures the full spectrum of an individual's variation and determines, rather than infers, a higher resolution of variants. However, identification of variants should not be confused with their interpretation, and pinpointing the causative disease variant will still be challenging²⁵. Our ability to identify variants from comprehensive sequence data will far outstrip our ability to characterize their biological effect. However, accurate and complete reporting is a necessary predecessor to a precise functional understanding of genomic data for the consumer. ■

Pauline C. Ng and **J. Craig Venter** are at the J. Craig Venter Institute, Science Center Drive, San Diego, California 92121, USA. **Sarah S. Murray** is at Scripps Genomic Medicine, Scripps Translational Science Institute, North Torrey Pines Court, La Jolla, California 92037, USA. **Samuel Levy** is at the J. Craig Venter Institute, Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: png@jcv.org

1. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. *Nature Rev. Genet.* **10**, 241–251 (2009).
2. Mailman, M. D. et al. *Nature Genet.* **39**, 1181–1186 (2007).
3. Heshka, J. T., Palleschi, C., Howley, H., Wilson, B. & Wells, P. S. *Genet. Med.* **10**, 19–32 (2008).
4. Hogarth, S., Javitt, G. & Melzer, D. *Annu. Rev. Genomics Hum. Genet.* **9**, 161–182 (2008).
5. Duncan, D. E. *Experimental Man: What One Man's Body*

Reveals About His Future, Your Health, and Our Toxic World (John Wiley & Sons, 2009).

6. Singer, E. *Technology Review* (MIT, 15 June, 2009); available at www.technologyreview.com/blog/editors/23680
7. Topol, E. J., Murray, S. S. & Frazer, K. A. *J. Am. Med. Assoc.* **298**, 218–221 (2007).
8. Bertram, L., McQueen, M. B., Mullin, K., Blacker, D. & Tanzi, R. E. *Nature Genet.* **39**, 17–23 (2007).
9. Hsu, A. & Naughton, B. (eds) *Guidelines on Vetting Genetic Associations* (23andMe; 2008); available at https://23andme.com/res/4400/pdf/23-03_Vetting_Genetic_Associations.pdf
10. *Applying Preventive Genomic Medicine in Clinical Practice* (Navigenics, 2008); available at www.navigenics.com/static/pdf/physician/physician-whitepaper.pdf
11. Green, P. H. R. & Jabri, B. *Lancet* **362**, 383–391 (2003).
12. Hunt, K. A. et al. *Nature Genet.* **40**, 395–402 (2008).
13. Liu, Y. et al. *PLoS Genet.* **4**, e1000041 (2008).
14. Lazarou, J., Pomeranz, B. H. & Corey, P. N. *J. Am. Med. Assoc.* **279**, 1200–1205 (1998).
15. Pirmohamed, M. et al. *Br. Med. J.* **329**, 15–19 (2004).
16. Sangkuhl, K., Berlin, D. S., Altman, R. B. & Klein, T. E. *Drug Metab. Rev.* **40**, 539–551 (2008).
17. www.23andme.com/health/all
18. www.pathway.com/more_info/drug_responses
19. Personalized Medicine Coalition *Personal Genomics and Industry Standards: Scientific Validity* (2008); available at <http://cancercontrol.cancer.gov/od/phg/docs/PMCSiValid.pdf>
20. Prainsack, B. et al. *Nature* **456**, 34–35 (2008).
21. Aulchenko, Y. S. et al. *Nature Genet.* **41**, 47–55 (2009).
22. Lyssenko, V. et al. *N. Engl. J. Med.* **359**, 2220–2232 (2008).
23. Meigs, J. B. et al. *N. Engl. J. Med.* **359**, 2208–2219 (2008).
24. Adeyemo, A. & Rotimi, C. *Public Health Genomics* doi: 10.1159/000218711 (2009).
25. Easton, D. F. et al. *Am. J. Hum. Genet.* **81**, 873–883 (2007).

Supplementary information accompanies this article online. The authors have disclosed a statement of conflicting interests that is available online.

See Editorial page 697, and online at <http://go.nature.com/VqPUE2>.

Let's celebrate human genetic diversity

Science is finding evidence of genetic diversity among groups of people as well as among individuals. This discovery should be embraced, not feared, say **Bruce T. Lahn** and **Lanny Ebenstein**.

A growing body of data is revealing the nature of human genetic diversity at increasingly finer resolution^{1,2}. It is now recognized that despite the high degree of genetic similarities that bind humanity together as a species, considerable diversity exists at both individual and group levels (see box, page 728). The biological significance of these variations remains to be explored fully. But enough evidence has come to the fore to warrant the question: what if scientific data ultimately demonstrate that genetically based biological variation exists at non-trivial levels not only among individuals but also among groups? In our view, the scientific community and society at large are ill-prepared for such a possibility. We need a moral response to this question that is robust

irrespective of what research uncovers about human diversity. Here, we argue for the moral position that genetic diversity, from within or among groups, should be embraced and celebrated as one of humanity's chief assets.

The current moral position is a sort of 'biological egalitarianism'. This dominant position emerged in recent decades largely to correct grave historical injustices, including genocide, that were committed with the support of pseudoscientific understandings of group diversity. The racial-hygiene theory promoted by German geneticists Fritz Lenz, Eugene Fischer and others during the Nazi era is one notorious example of such pseudoscience. Biological egalitarianism is the view that no or almost no meaningful genetically

based biological differences exist among human groups, with the exception of a few superficial traits such as skin colour³. Proponents of this view seem to hope that, by promoting biological sameness, discrimination against groups or individuals will become groundless.

We believe that this position, although well-intentioned, is illogical and even dangerous, as it implies that if significant group diversity were established, discrimination might thereby be justified. We reject this position. Equality of opportunity and respect for human dignity should be humankind's common aspirations, notwithstanding human differences no matter how big or small. We also think that biological egalitarianism may not remain viable in light of the growing body of empirical data (see box).

Genotype Concordance

We examined the analytical validity of the DTC genetic tests by assessing whether genotypes are consistently called. We identified 166,135 SNPs genotyped by both 23andMe and Navigenics, of which >99.7% have concordant genotypes in one individual (Supplemental Table 1). We also compared the genotypes to those obtained in the research laboratory setting (Supplemental Table 1)¹. In summary, we estimate that less than 3 in 1000 genotypes are inaccurate.

Prediction Comparison

If the DTC tests have high clinical validity, i.e., the genetic variant correlates with the specific disease or condition², then one would expect the results between the two DTC companies to be consistent. Although consistent risk predictions between the two companies do not necessarily mean the prediction is accurate, disagreement between the predictions suggests that either 1) at least one of the DTC companies is incorrect, or 2) there is enough uncertainty that a reliable prediction cannot be made. Thirteen diseases were in common between the two DTC companies and had sufficient background technical information (markers and their odds ratios) to enable comparison. For these thirteen diseases, we compare the consistency of risk predictions between the two DTC companies.

Prediction Comparison: Relative Risk

To compare disease risk predictions, we compared the relative risk (RR) of disease in each individual (Supplemental Figure 1, Methods) and compared the qualitative agreement of predictions. Averaged among the 5 individuals, 64% of

predictions qualitatively agree between 23andme and Navigenics (Supplemental Figure 2A). We also assessed whether certain diseases have better prediction agreement between the DTC companies (Supplemental Figure 2B). For four diseases, the predictions between the two DTC companies completely agree.

The relative risk values for the five individuals' diseases from 23andMe and Navigenics are correlated ($r=0.54$) (Supplemental Figure 3). There were diseases in individuals that have increased risk in one company, but decreased in the other (Supplemental Figure 3, shaded areas). Most risks range between 0.5 and 1.5 conferring only modest risk. There can be some utility if a test predicts a high risk enabling an individual to take preventive/precautionary measures for that disease. The caveat, however, is in the case of a low risk prediction for disease that a false sense of security is not assumed since much of the genetic contribution to disease risk is still not understood.

The disease risk estimates used by both DTC companies are calculated from a risk model that utilizes odds ratios for specific genotypes of individual markers associated with disease. Supplemental Figure 4 shows the number of markers used to calculate disease risks for both DTC companies. No disease has markers that are completely shared by the two companies. The risk markers reported by the two DTC companies can differ because each company has different criteria for including a disease risk marker^{3,4}. For Navigenics, the following criteria applies: 1) the marker is detected in at least two independent replications, 2) the studies typically have at least 250 cases and 250 controls for modest effects⁴. Meanwhile 23andMe

utilizes a distinct set of criteria to consider disease risk markers. These are: 1) studies identifying the genetic marker have at least 1,000 cases and 1,000 controls for modest effects ³, and 2) association of the marker with disease must be replicated in at least one independently published study.

Marker differences were mainly due to selection criteria and not due to differences between the marker sets. For the 25 risk markers unique to Navigenics, 24 are identical to, or in strong linkage disequilibrium with markers that were genotyped by 23andMe. Similarly, 9 out of the 10 risk markers unique to 23andMe are identical to or in strong linkage disequilibrium with markers genotyped by Navigenics. Therefore, it is possible to achieve general marker consensus between companies.

Comparison of Markers' Adjusted Odds Ratios

There are three possible genotypes for each disease marker (homozygous for the common allele, heterozygous, and homozygous for the rare allele) and each genotype has an odds ratio (OR) for disease. We compared the adjusted odds ratio for genotypes at identical markers or markers in strong linkage disequilibrium (LD) with each other (Methods). For markers that are identically shared between the two DTC companies, the adjusted ORs for identical genotypes are very similar and 96% (23/24) have values within 10% of the other (Supplemental Figure 5A). For markers that are in strong linkage disequilibrium with each other, 77% (24/31) have odds ratios for the observed genotypes that are within 10% of each other (Supplemental Figure 5B). For those marker pairs that do not have adjusted ORs

within 10% of each other, most of the discrepancies can be explained by the DTC company preferentially taking the OR from the publication that directly assayed their marker. In conclusion, similar ORs are observed for identical markers and markers in strong LD and the genetic contribution to disease predictions from identical and overlapping markers are similar between these two DTC companies.

Errors in Genotype Imputation

Because DTC companies rely on LD to impute genotypes when the disease-associated marker is not on their specific microarray platform, we wanted to quantify how well the LD pattern from the HapMap Project corresponds to what is observed in an individual. Our motivation for this was because we observed one pair of markers in the 8q24 region associated with prostate cancer where the adjusted OR is 1.35 in 23andMe but 0.95 in Navigenics for two individuals (Supplemental Figure 5B). Although the markers in the 8q24 region are in perfect linkage disequilibrium ($r^2=1$, $D'=1$) according to a European population (CEU HapMap data ⁵), the two markers are not in linkage disequilibrium in two males from the same family resulting in discordant genotypes between the two markers, which suggests that a recombination event has occurred in this family. We wanted to see how well LD structure correlated between an individual and the HapMap reference population. We looked at SNPs unique to Navigenics that were in complete linkage disequilibrium ($r^2=1$) with SNPs unique to 23andMe according to the CEU HapMap data. For these pairs of markers, we assessed how often one subject's genotypes did not perfectly correlate with the predicted genotype based on LD observed in the CEU

population, and observed 1% genotype discordance (3,686/367,931 marker pairs). While this number is relatively small, differences in LD structure between an individual and the HapMap reference population may provide a source of error.

Environment

The environment can be a strong component in determining whether a person gets disease. Both DTC companies display the fraction of disease that is attributable to genetics and environment on their websites. The DTC companies for the most part agree on the contribution of environment to disease (Supplemental Figure 6). However, they do not fully discuss how much of the genetic component is understood and represented in their predictions.

Materials and Methods

Human Samples

We compare the DTC profiles for five individuals, designated Subjects A-E. Subjects A-C are females; Subjects D and E are males. Subjects B-E are from the same family. All individuals in this study were provided a study outline and a detailed informed consent to enable their participation. Study participants provided saliva samples as directed in the sample kits provided by 23andMe and Navigenics. Genotyping was conducted at both companies prior to the requirement that the companies comply to the CLIA laboratory standards. Individual DNA samples in this study were genotyped for ~580,000 SNPs (23andMe) and ~925,000 SNPs (Navigenics). Participants provided access to their DTC disease reports for this study. We

compared the disease risks from the DTC profiles provided between January 2009 and April 2009.

Qualitative Prediction

Both companies report the absolute risk of disease, which is the probability that an individual will develop the disease. Absolute risk is calculated by: Absolute risk = Relative risk * Average population disease risk. When comparing the absolute disease risks between different companies, one reason for the observed differences is because the DTC companies use different estimates for the "average population disease risk" (Supplemental Table 2). Navigenics takes into account gender, while 23andMe can take into account age. Thus, differences in "average population disease risk" can cause the absolute risk to differ between the two DTC companies. If instead we focus on relative risk, then the confounding variable "average population disease risk" is removed.

We define a relative risk (RR) < 0.95 as a "decreased risk" in disease, $0.95 \leq \text{RR} \leq 1.05$ as an "average risk" in disease, and $\text{RR} > 1.05$ as an "increased risk" in disease. If both DTC companies predict an increased risk of disease, then the predictions qualitatively agree. If one DTC company predicts a decreased risk while the other predicts an increased risk, then the predictions disagree.

Comparing Odds Ratios Between Platforms

For each marker genotype, we compared the adjusted odds ratio (OR) between the two DTC companies rather than the standard OR. 23andMe reports an

“adjusted odds ratio” by normalizing to genotype frequencies, so that the average OR is 1 and there are genotypes that have reduced disease risk compared to the average population. Navigenics reports the odds ratio in the classical sense where 1.0 is baseline (lowest risk), and risk alleles have ORs greater than 1. By dividing the odds ratio with the average odds ratio reported by Navigenics, we calculated Navigenic’s adjusted odds ratio so that it could be directly compared to 23andMe. Both 23andMe and Navigenics generally apply a multiplicative model when combining multiple markers⁶.

In some cases, the reported associated risk markers are not directly assayed on the microarrays used by the DTC company. In this case, the DTC companies rely on linkage disequilibrium to choose a proxy SNP that correlates with the marker reported in the published literature. In this study, we define SNPs with $r^2 \geq 0.8$ and $D' \geq 0.9$ from a Caucasian population (HapMap CEU⁵) data as proxy markers.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure 1. The procedure that DTC companies use to generate disease risk predictions.

Supplemental Figure 2A. Agreement between qualitative disease risk predictions from Navigenics and 23andMe for five individuals. Thirteen diseases were compared, but because breast cancer predictions are exclusive to females and prostate cancer predictions are exclusive to males, twelve disease comparisons are shown for each individual. For Subject B, 23andMe did not provide predictions for two diseases due to DNA quality issues.

Supplemental Figure 2B. Agreement between qualitative disease risk predictions from Navigenics and 23andMe for thirteen diseases. Breast cancer predictions were provided for the three females only; prostate cancer predictions were provided for the two males only.

Supplemental Figure 3. The predicted relative risks for diseases in five individuals. Shaded areas indicate where one DTC test predicts an increased disease risk while the other predicts a decreased risk. 81% and 78% of the disease relative risk predictions range between 0.5 and 1.5 for 23andMe and Navigenics, respectively.

Supplemental Figure 4. Overlap between disease marker sets used by Navigenics and 23andMe. “Both Navigenics and 23andMe” refers to identical markers or markers that are in strong linkage disequilibrium ($r^2 \geq 0.8$, $D' \geq 0.9$). In light purple are markers that refer to the same locus, but are not in strong LD.

"Ambiguous" refers to markers in the same chromosomal region, but for which there was no HapMap data to elucidate LD properties. "Navigenics only" (red) refers to markers uniquely used by Navigenics. "23andMe only" (blue) refers to markers uniquely used by 23andMe.

Supplemental Figure 5A. Adjusted odds ratios for identical SNPs used by both 23andMe and Navigenics for disease prediction.

Supplemental Figure 5B. Adjusted odds ratios for pairs of SNPs in strong linkage disequilibrium ($r^2 \geq 0.8$, $D' \geq 0.9$) used by 23andMe and Navigenics for disease prediction.

Supplemental Figure 6. The percentage of the disease that is attributable to the environment, as stated by DTC Companies. For type 2 diabetes, Navigenics and 23andMe show an environmental contribution of 36% and 74%, respectively. For Crohn's disease, Navigenics and 23andMe show an environmental contribution of 20% and 40-50%, respectively. Presumably both companies are using different studies to estimate the genetic and environmental components for the diseases whose environmental contributions differ.

Supplemental Table 1. Concordance of genotypes between 23andMe and Navigenics (first row), and concordance of genotypes between the DTC companies and a research laboratory setting (second and third rows).

Type of Test	Platform 1	Platform 2	Genotype Concordance
Between two technologies	Navigenics (using Affymetrix 6.0)	23andMe (using Illumina HumanHap550 + 50,000 custom SNPs)	99.75%
Within the same technology	23andMe using Illumina HumanHap550 + 50,000 custom SNPs	Illumina HumanHap650Y BeadChip ¹	99.99%
Within the same technology	Navigenics (using Affymetrix 6.0)	Affymetrix Mapping 500K Array Set Set ¹	99.71%

¹ Research laboratory setting

¹Research setting

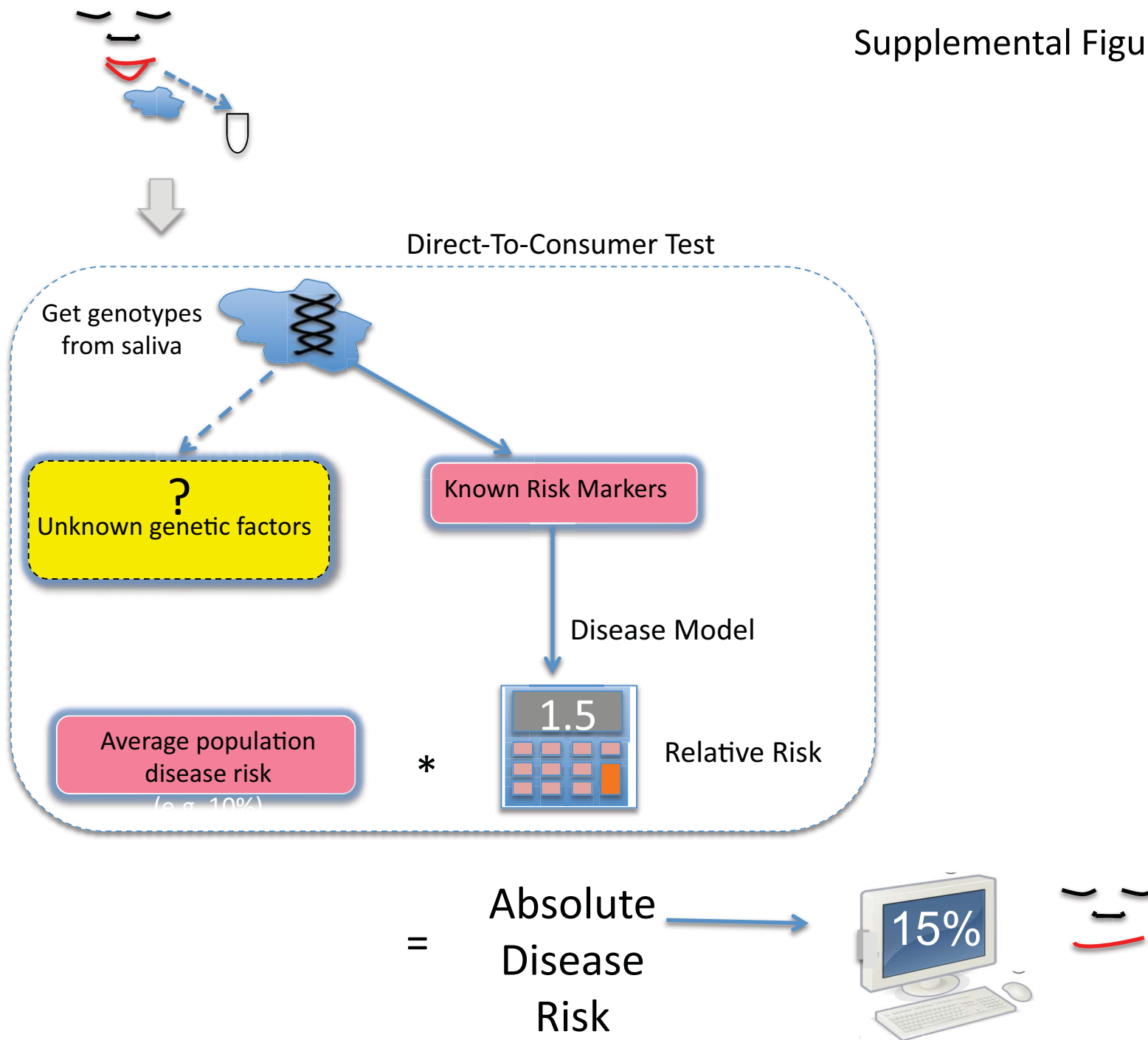
Supplemental Table 2. Population disease risk. M=male; F=female. The risk for 23andMe is shown for the entire age range.

	Navigenics	23andMe
Breast cancer	13% (F)	16.2% (F)
Celiac Disease	0.06% (M) / 0.11% (F)	0.18%
Colon cancer	6% (M) / 5% (F)	8.7%
Crohn's Disease	0.58% (M) / 0.54% (F)	0.49%
Heart Attack	42% (M) / 25% (F)	17.7%
Lupus	0.03% (M) / 0.26% (F)	0.25% (F)
Macular Degeneration	3.1% (MF)	7%
Multiple Sclerosis	0.30% (M) / 0.77% (F)	0.52%
Prostate Cancer	17% (M)	17.8% (M)
Psoriasis	4% (MF)	10.7%
Restless Legs Syndrome	4% (MF)	4%
Rheumatoid Arthritis	1.6% (M) / 3.3% (F)	3.4%
Type 2 diabetes	25% (M) / 30% (F)	21.9%

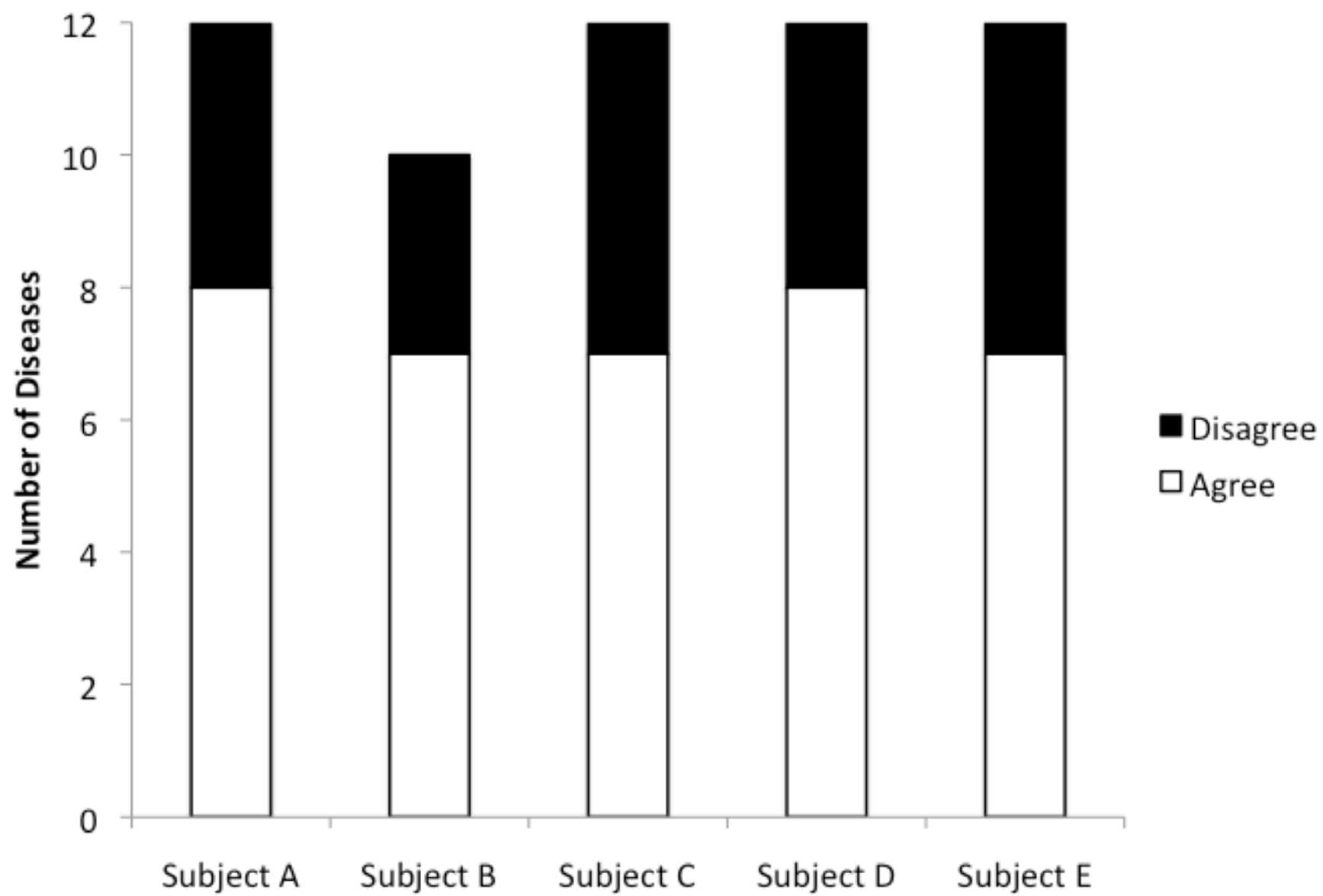
REFERENCES

1. Levy, S., *et al.* The diploid genome sequence of an individual human. *PLoS Biol* **5**, e254 (2007).
2. Hogarth, S., Javitt, G. & Melzer, D. The current landscape for direct-to-consumer genetic testing: legal, ethical, and policy issues. *Annu Rev Genomics Hum Genet* **9**, 161-182 (2008).
3. 23andme. Guidelines on Vetting Genetic Associations.
https://23andme.https.internapcdn.net/res/3987/pdf/23-03_Vetting_Genetic_Associations.pdf. (ed. Hsu, A., Naughton, B.) (23andme, 2008).
4. Navigenics. Applying Preventive Genomic Medicine in Clinical Practice
http://www.navigenics.com/static/pdf/physician/physician_whitepaper.pdf. (2008).
5. Frazer, K.A., *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
6. Personalized Medicine Coalition. Personal Genomics and Industry Standards: Scientific Validity.
<http://cancercontrol.cancer.gov/od/phg/docs/PMCSiValid.pdf>. (2008).

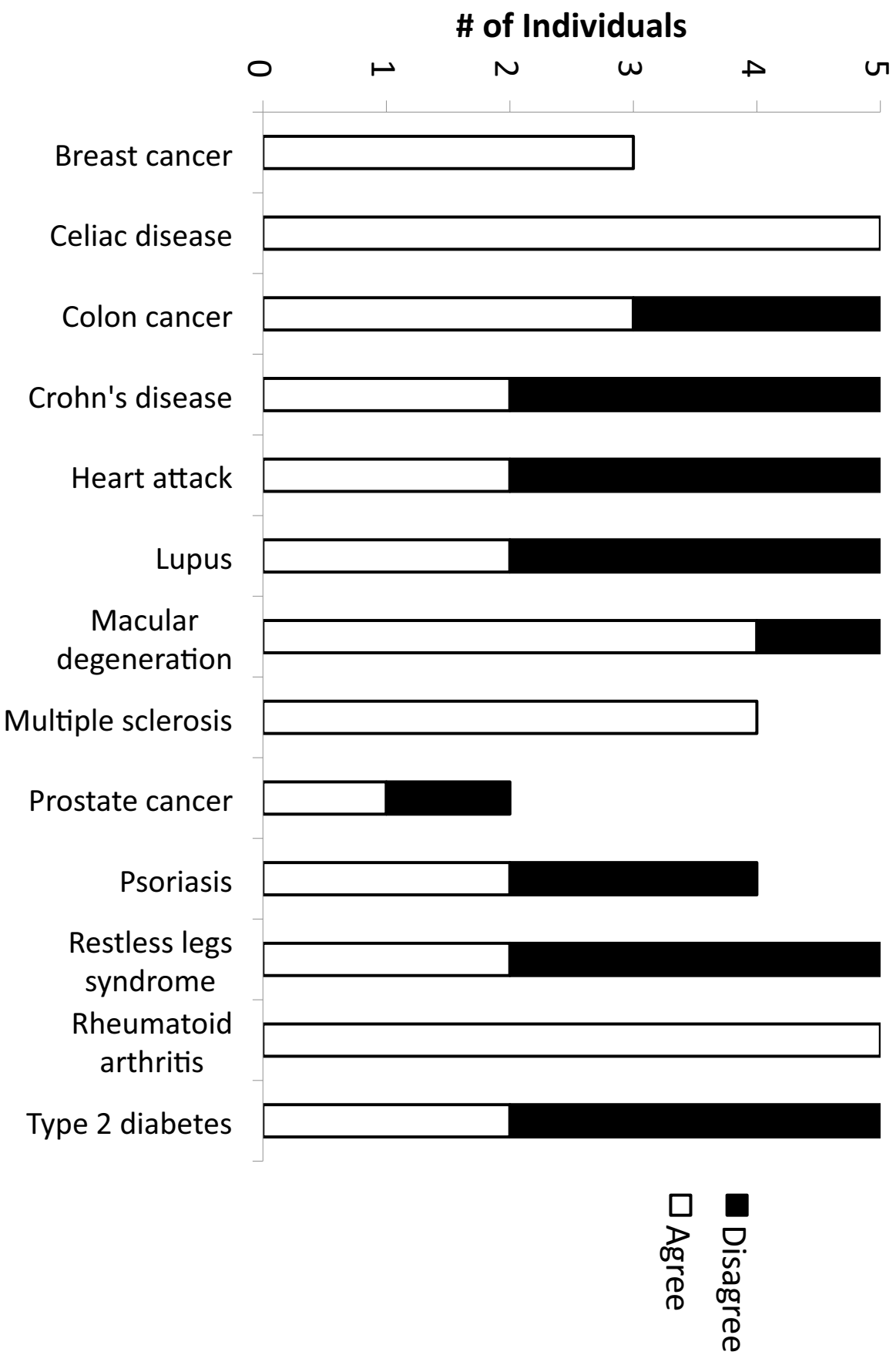
Supplemental Figure 1



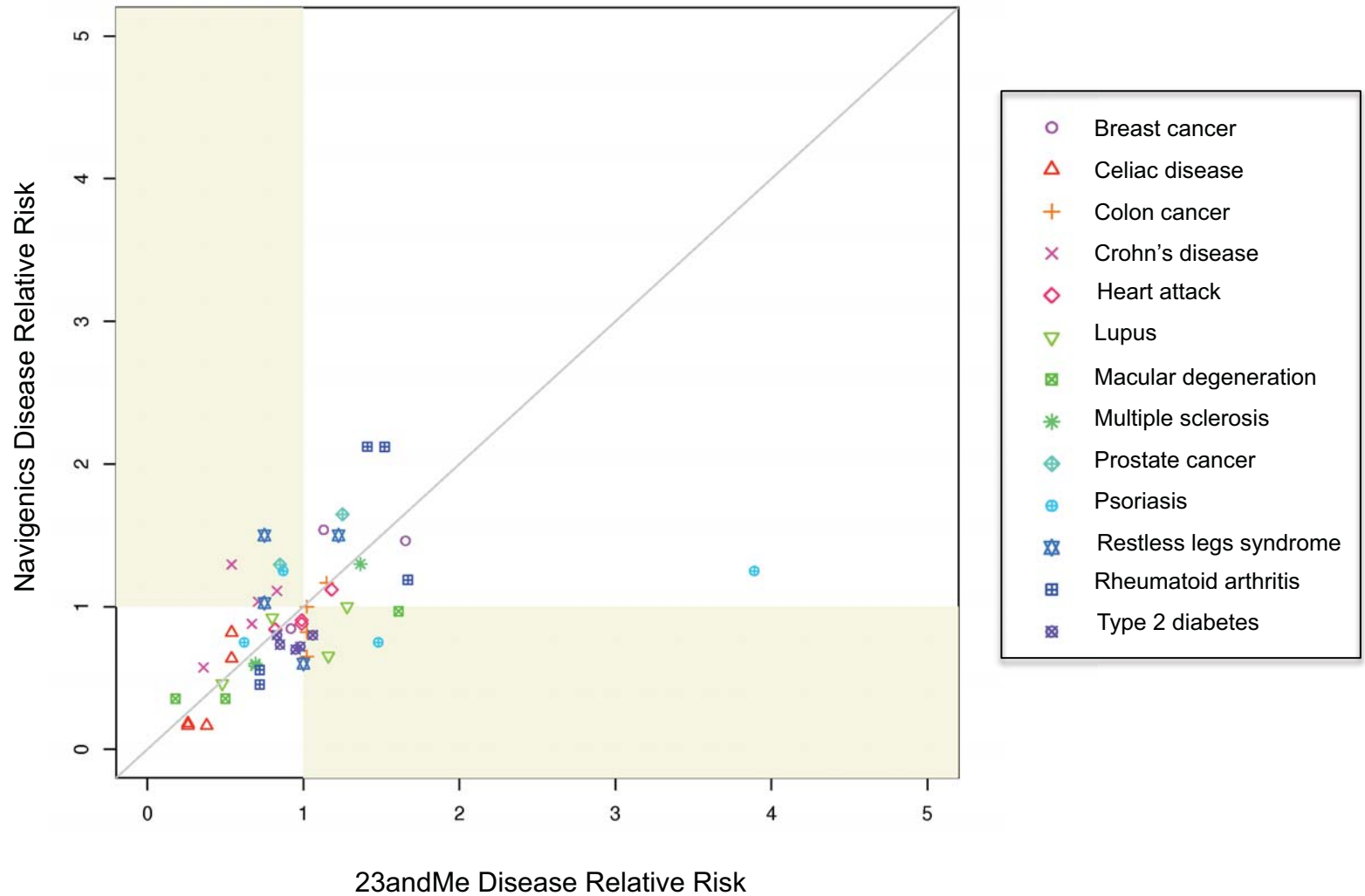
Supplemental Figure 2A



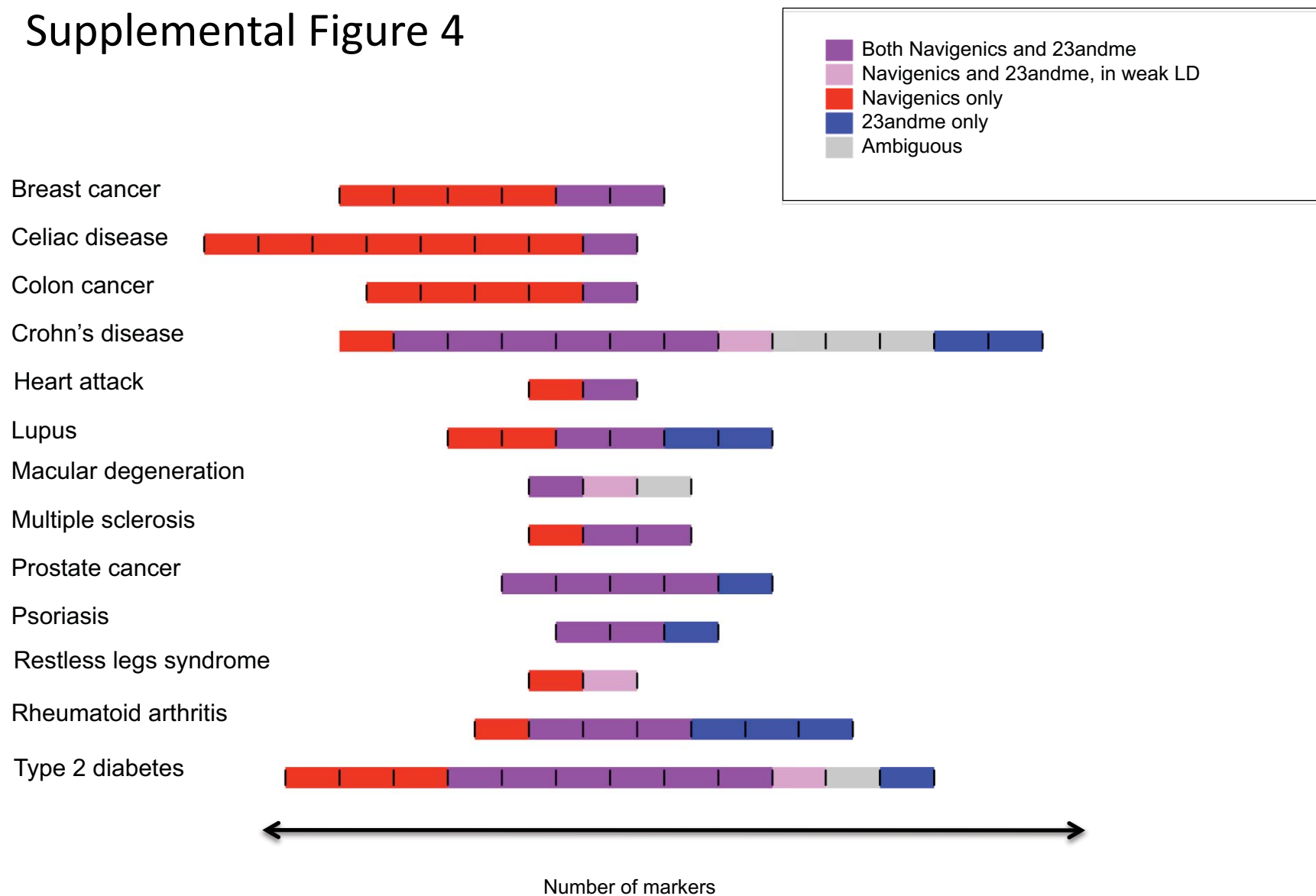
Supplemental Figure 2B



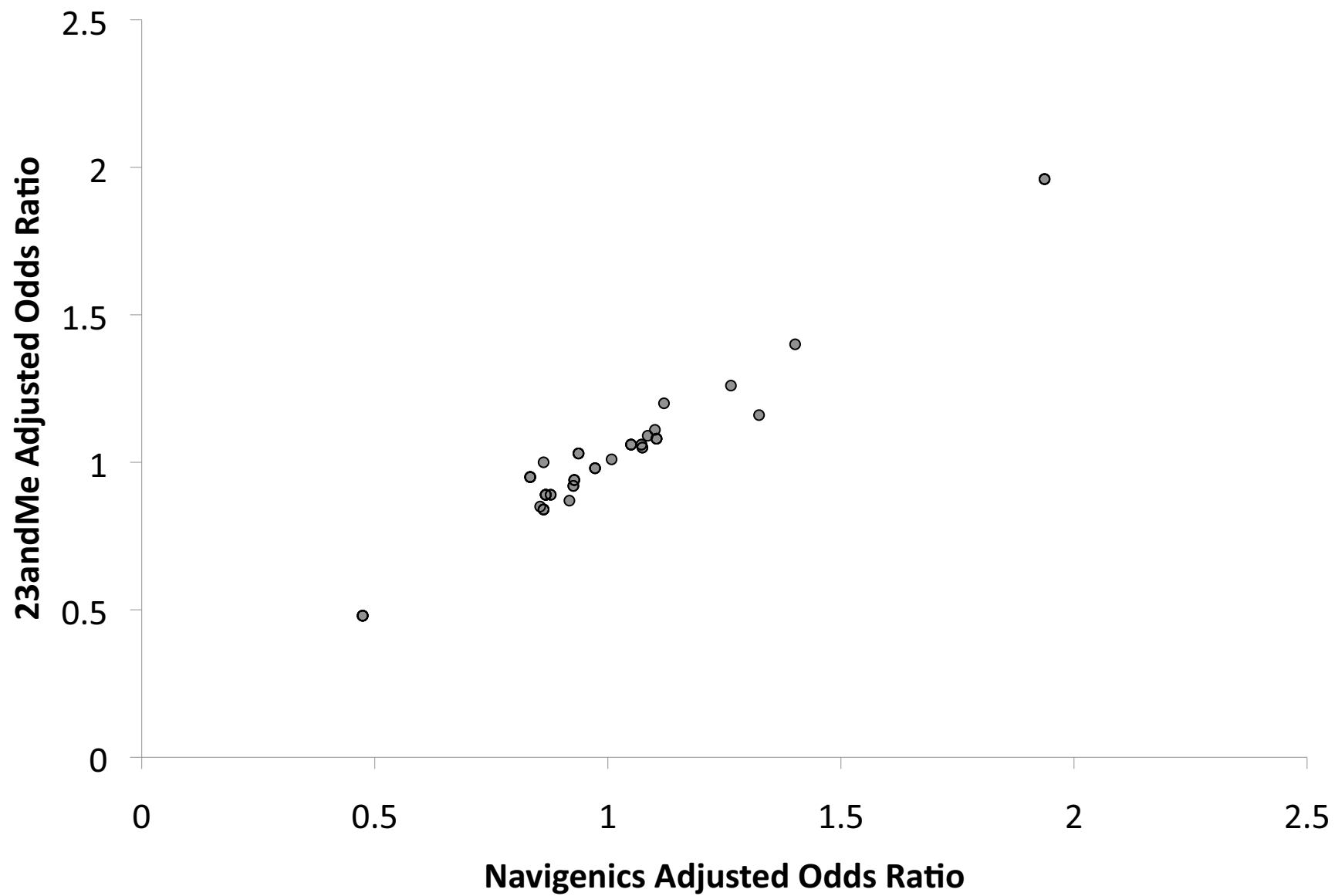
Supplemental Figure 3



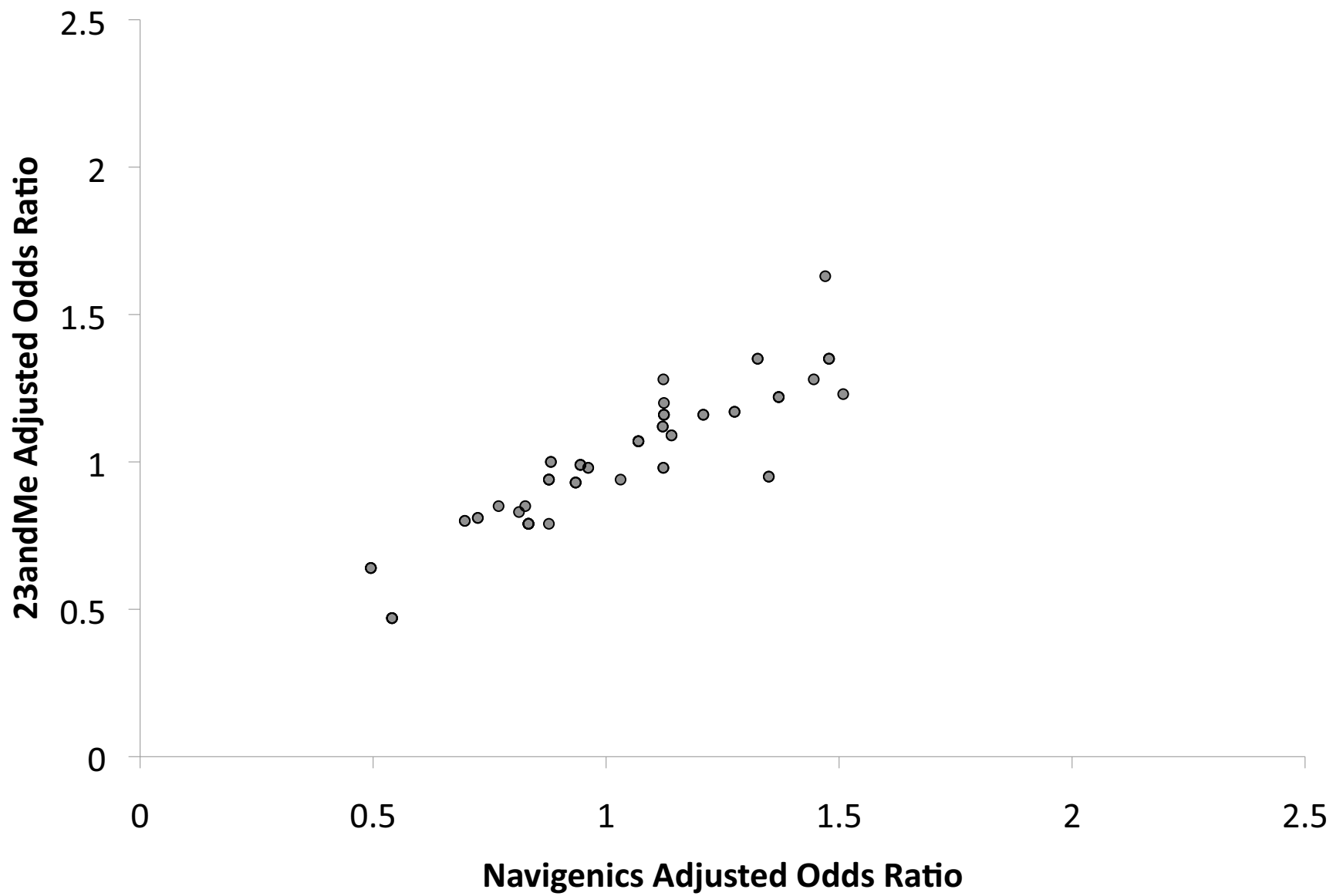
Supplemental Figure 4



Supplemental Figure 5A



Supplemental Figure 5B



Supplemental Figure 6

