The impact of personal genomics on risk perceptions and medical decision-making

To the Editor:

The question of whether direct-to-consumer (DTC) personal genomic testing (PGT) will lead to effective preventative medicine instead of a "raiding of the medical commons"[1] or inappropriate changes to medical treatment has spurred debate about regulations for PGT. But despite scientific progress and a high-profile regulatory engagement, there is a dearth of empirical evidence on how such services might influence participant decision-making. On one hand, better information about genomic disease risks seems likely to empower consumers and add to the physician's toolkit, encouraging rational screening behavior and monitoring of disease symptoms. On the other hand, however, this information might mislead or alarm consumers, resulting in overuse of medical resources.

Understanding how genomic test results affect individuals' healthcare use is a two-part problem that requires establishing the casual relationships between genomic risk information and personal risk perception, and then tracing out how changes in risk perception translate into action. Such decisions involve complex factors such as baseline knowledge of genetic risk, learning processes and cognitive biases. Here we model risk perception as a combination of baseline beliefs and learning in response to genetic news while addressing healthcare use as a function of changes in these risk perceptions. We draw on two lines of research: the medical and policy literature surrounding PGT and research in cognitive psychology on risk perception learning. To our knowledge, this is the first study to evaluate risk perception changes and subsequent medical actions in a real-world setting with direct-to-consumer PGT customers.

The PGT literature has evaluated how test results influence psychological states, health behavior and follow-up screening tests[2–9]. Related work has also explored how hypothetical test results influence risk perceptions[10]. However, these studies do not establish links from specific genetic results to risk perception or from perception to action.

The nature of the particular genetic conditions is also important for evaluating the impact of PGT. Genetic tests for autosomal dominant conditions such as Huntington's disease are more certain and have stronger implications for lifestyle and financial decisions[11,12]. In contrast, PGT for common complex genetic disorders carries an uncertain message because it involves conditions (e.g., diabetes) that are associated with several mutations as well as lifestyle choices. Furthermore, the interpretations of these common complex conditions may change over time as knowledge of genotype–phenotype associations improves. The analysis below focuses on a set of eight conditions that fall into the category of common complex conditions. The sample under study includes consumers who individually sought out and purchased PGT services from 23andMe before the US Food and Drug Administration (FDA) banned DTC health reports.

By linking participants' test results to their perception changes and follow-up medical actions, we can examine how the participants responded to different results across a set of diverse medical conditions. Understanding how consumers learn from and react to the multi-condition set of risk results requires outcome measures that capture both perceptions and related actions. Our approach takes advantage of the variation in disease and participant characteristics to evaluate these causal relationships for risk perception changes and follow-up medical use.

Cognitive psychology and, to a lesser extent, behavioral economics have a rich literature investigating how individuals learn, or fail to learn, from new information[13–18]. The 'good news–bad news effect' describes the phenomena of valence-dependent learning, where individuals are more prone to update their beliefs after receiving favorable news than after receiving unfavorable news. The news may be about general risk information (e.g., likelihood of being robbed) or intrinsic qualities (e.g., intelligence and attractiveness)[13–16]. Though some lab studies have considered the good news–bad news effect in the context of population health information[14–16], we are not aware of any that address good news–bad news asymmetric learning from personal genetic risks across multiple health conditions. Because most such studies are conducted in laboratory settings, they are also limited to measuring short-term perception (e.g., on the same day as the information treatment) and cannot measure real-world behavioral changes that result from longer-term risk perception changes.

The information treatment in our study is novel and personalized, as participants in the sample did not know their genotypes before they undertook DTC testing. This specialized setting allows us to gain new insight into risk information processing and good news–bad news belief updating over a meaningful period of time.

Our sample includes 617 23andMe customers enrolled in the Impact of Personal Genomics (PGen) study, a longitudinal study of real DTC genomics customers involving both detailed surveys and genetic test results, and funded by the National Human Genome Research Institute (NHGRI). Participants reported their risk perceptions at baseline and again 6 months after receiving results (see **Supplementary Methods** section for more detail about the PGen Study's survey design). All participants reported on their risk perceptions for nine conditions (Alzheimer's disease, Parkinson's disease, breast cancer, lung cancer, colon cancer, prostate cancer, type II diabetes, coronary heart disease, and obesity), as well as any

Table 1 Participant baseline risk perceptions in eight conditions (5-point Likert scale)

	Count	Mean	s.d.	Percentage below average	Percentage above average
Lung cancer	614	2.27	1.03	58%	12%
Parkinson's disease	554	2.37	0.99	49%	8%
Colorectal cancer	615	2.68	0.98	38%	17%
Alzheimer's disease	538	2.68	1.02	39%	17%
Type 2 diabetes	580	2.69	1.16	45%	26%
Breast cancer	323	2.73	1.01	37%	19%
Prostate cancer	260	2.80	0.92	29%	17%
Coronary heart disease	591	2.95	1.07	32%	33%
Total	4,075	2.63	1.06	42%	19%

Summary statistics for the eight condition risk perceptions in our sample, measured at baseline (before receiving PGT results). In cases where a participant reported that he or she had been diagnosed with a condition, the corresponding participant–condition observation was excluded. The remaining condition risk perceptions were drawn from 617 participants and for the eight above conditions of 4,075 participant–conditions from 23andMe customers who participated in the PGen Study. Conditions are listed in ascending order of average baseline risk perception. The minimum and maximum baseline risk perception for each condition was 1 and 5, respectively.

other conditions that they reported special interest in learning about. Breast cancer and prostate cancer responses were limited to women and men, respectively. We excluded obesity because most participants were old enough to know whether their risk had been realized and because the genetic risk reports had very little variation for this condition.

First, we evaluated how risk perceptions for eight common conditions (**Table 1**) changed in response to PGT results. After adjusting for baseline risk perception levels and individual trends across conditions, we tested whether the good news–bad news effect occurred in this setting. Next, we compared how perception changes differed across the eight medical conditions. Finally, we assessed how shifts in risk perception affected an individual's propensity to make or plan follow-up medical appointments, exams or procedures related to their PGT results. Risk perception data are provided in **Supplementary Figures 1–4**.

Risk perception results. Initially, we evaluated baseline risk perception levels for each participant across the eight conditions of interest. Our results show that participants had a slight optimistic bias in baseline risk perceptions, with variation in the average level of optimism across the conditions (**Table 1** and **Supplementary Fig. 1a**). After agreeing to participate but before receiving their test results, participants ranked their risk perception for each applicable condition on a 5-point Likert scale (1 = much lower than average, 2 = lower than average, 3 = average, 4 = higher than average, 5 = much higher than average).

For the 4,075 participant–condition observations in the sample, the mean risk Likert scale risk perception was 2.63 (s.d. = 1.06). In the overall sample, 42% of all perceptions reflected below-average risk perception, compared with only 19% for above-average risk perception. The mean baseline risk perception by condition ranged from 2.27, for lung cancer (58% below average, 12% above average), to 2.95, for coronary heart disease (32% below average, 33% above average). The ordering of conditions, in terms of baseline risk, did not appear to follow a pattern based on disease mortality, morbidity, population frequency or heritability. The analysis data set for participant–condition risk perceptions can be viewed in the **Supplementary Data** file.

Our analysis of the test results and risk perceptions revealed that, on average, participants updated their risk perceptions after they learned about their test results. However, the magnitude of this updating was asymmetric across different risk result groups.

We evaluated how participants changed their risk perceptions between baseline and 6 months after viewing test results for the pooled sample of 4,075 participant conditions. Risk perceptions at both time points were recorded in the PGen Study's online survey on the same 5-point Likert scale. In our first risk perception analysis, we used 23andMe's risk result groups to categorize participant–condition results. 23andMe divides risk results into three categories on the basis of risk relative to the population: decreased risk (<0.8× population risk ratio), typical risk (0.8–1.2× population risk ratio) or elevated risk (>1.2× population risk ratio) (**Supplementary Fig. 5**). The outcome variable was change in risk perception.

In our primary empirical specification, we used ordinary least-squares (OLS) regression to evaluate the average risk updating in response to each type of test result. This regression also adjusted for baseline risk perceptions by including indicator variables for a participant's initial risk perception for a given condition, and for individual trends across all conditions by including participant fixed effects (see **Supplementary Methods**).

Because the risk perception questions elicited categorical responses (1 = much lower than average, 2 = lower than average, 3 = average, 4 = higher than average, 5 = much higher than average), we also employed ordered logistic regression models to analyze how different PGT results led to perception changes. These alternative specifications account for the potential nonlinear nature of the five risk perception categories and estimate how different PGT risk results influence the odds of responding with a higher level of risk perception change. Since these ordinal logit regressions produce the same qualitative results as the OLS regressions, and the OLS regressions have the additional benefit of intuitive interpretation (i.e., coefficient magnitudes can be interpreted as conditional means), we used the OLS versions as our primary specifications. Both sets of regressions are detailed in **Supplementary Note 1**, and the regression coefficients for all models are reported in **Supplementary Tables 1–4**. **Supplementary Figures 1–4** provide additional graphical depictions of the regression data and risk perception changes.

The OLS regressions show that after adjusting for baseline risk perceptions and individual trends, participants reported, on average, a 0.49-point drop ($P < 0.01$) in risk perception after a decreased risk result, a 0.23-point drop ($P < 0.01$) after a typical risk result, and a 0.37-point ($P < 0.01$) increase after an elevated risk result (**Fig. 1a**). To evaluate asymmetry in the reaction to risk result groups, we performed Wald tests of linear hypotheses to test whether the magnitude of risk perception changes after a decreased risk result was different from the magnitude of changes following an increased risk result. These tests revealed that the magnitude of risk perception updating in response to decreased risk results was statistically greater ($P < 0.05$) than the updating for elevated risk results.

Even after adjusting for baseline perceptions, a typical risk result led to a downward shift in risk perception. Though smaller in

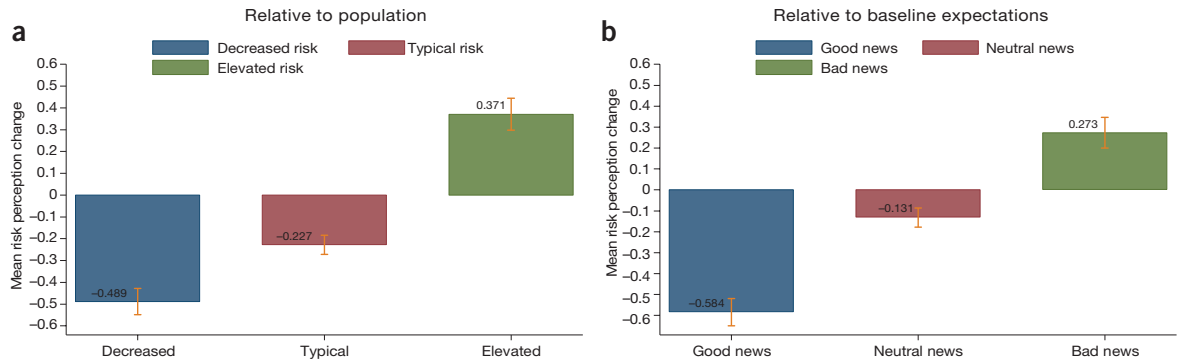

Figure 1 Risk perception changes by risk result group. (**a**,**b**) Adjusted mean risk perception changes corresponding with decreased, typical and elevated risk results (**a**) or good news, neutral news and bad news (**b**). We calculated adjusted mean risk perception change using OLS regressions of risk perception change (measured using a 5-point Likert scale) on risk result categories. Error bars represent the 95% confidence intervals. The sample includes 4,075 participant–condition observations (1,029 decreased, 2,269 typical and 777 elevated results in **a**; 1,056 good news, 1,382 neutral news and 1,637 bad news results in **b**) from 617 23andMe customers. The regressions adjust for baseline risk perceptions and include participant fixed effects. Standard errors were clustered at the participant level. All groups' adjusted perception changes were significantly different from 0 ($P < 0.01$). Using random effects instead of individual fixed effects yielded nearly identical regression results.

magnitude than the other results' updating responses, this statistically significant ($P < 0.01$) shift suggests that 6 months after viewing a typical risk result, participants interpreted it as a positive signal.

To consider results relative to individuals' expectations (rather than the overall population), and to provide a more direct test for the good news–bad news effect in our sample, we recategorized the results as 'good news,' 'neutral news' and 'bad news' (**Fig. 1b**). Here, good news reflects test results that were better than expectations (e.g., a 'decreased' risk result with an 'average' baseline risk score), neutral news reflects expectations that match up with risk results (e.g., an elevated risk result with a higher-than-average baseline risk score) and bad news reflects risk results worse than expectations (e.g., a typical risk result with a lower-than-average baseline risk score).

These alternative result categories show an even greater asymmetry between result groups and confirm the presence of a good news–bad news effect. After adjusting for baseline risk perceptions and individual fixed effects in the regression analysis, good news led to a 0.58-point drop ($P < 0.01$) in risk perception, neutral news resulted in a 0.13-point drop ($P < 0.01$) and bad news provided a 0.27-point ($P < 0.01$) increase. The Wald test confirmed that the good news–bad news asymmetry is significant (at the 1% level). Under these alternative result categories, the risk perception drop for neutral news was >40% less than it was for typical results under the population risk categories. **Supplementary Tables 1** and **2** report the estimates of both the OLS and ordered logit regressions that estimate the good, neutral and bad news effects.

Because the information content of test results differs by both result category and condition (see **Supplementary Fig. 5** for an example of 23andMe's disease risk report), we also set out to examine how perception changes differed across conditions. We used the same OLS regression framework but further parsed the result groups by condition (**Fig. 2**). Although we analyzed these effects with both population risk result groups (decreased, typical and elevated) and relative to baseline expectations groups (good, neutral and bad news), for brevity, we report only the latter here.

The direction and statistical significance of risk perception changes remained mostly the same as in the pooled sample when we divided results by condition, but the extent of the updating and good news–bad news asymmetry varied across conditions (**Fig. 2** and **Supplementary Table 3**). After adjusting for baseline expectations and participant fixed effects, the perception changes for good news results ranged from −0.71 points ($P < 0.01$), for Alzheimer's disease, to −0.17 points ($P < 0.05$), for coronary heart disease. Risk perception updating after neutral news results ranged from changes statistically indistinguishable from 0 (for Alzheimer's disease and coronary heart disease) to −0.43 points (for breast cancer). Bad news results produced average risk perception changes that ranged from statistically indistinguishable from 0 (breast cancer) to a rise of 0.55 points ($P < 0.01$, Alzheimer's disease). Factors such as condition population risk, mortality and genetic heritability showed no clear correlation pattern with the magnitude and asymmetry of perception changes across conditions (data not shown).

Within each result category, we can explore the differences in the magnitude of perception updating across diseases. After adjusting for multiple comparisons (see **Supplementary Methods**), we were able to detect a number of statistically significant pairwise differences within the result groups. For example, Alzheimer's elicited the greatest amount of risk perception updating after good news and bad news, so we can use Alzheimer's as an example to highlight these pairwise differences. Within the good news category, we found significantly smaller decreases in risk perception for colorectal cancer ($P < 0.05$), diabetes ($P < 0.05$) and coronary heart disease ($P < 0.10$) than for Alzheimer's disease. Among the neutral news results, we found more negative-perception changes in the breast cancer and lung cancer groups than in the Alzheimer's disease group ($P < 0.05$). For the bad news results group, the Alzheimer's disease group saw greater perception changes than the breast cancer, lung cancer, colorectal cancer, diabetes and Parkinson's disease groups ($P < 0.05$). **Supplementary Table 3** provides a more comprehensive breakdown of the adjusted pairwise differences; we address the overall patterns across diseases below.

As an exercise in testing at least one dimension of generalizability, we compared updating results of participants who were primarily interested in their ancestry results with those of participants who were primarily interested in their results for other information. Our motivation for this analysis was the possibility that risk perception results are driven by participants most interested in health risks, whereas the health reports would

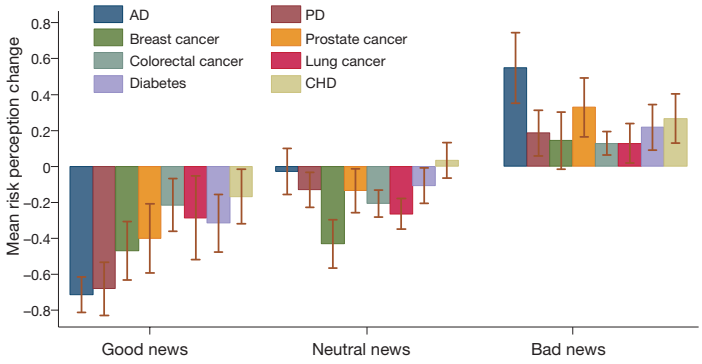

Figure 2 Risk perception changes by risk result group and medical condition. Mean risk perception changes corresponding with good news, neutral news and bad news risk results, split by medical condition groups. We calculated mean risk perception change using OLS regression of risk perception change (measured using a 5-point Likert scale) on risk result groups, using interaction variables to estimate the differential effects for specific diseases. Estimates include control variables for baseline risk perception. Red bars represent 95% confidence intervals. Fixed effects for the 617 participants are included in the regression model and standard errors are clustered at the participant level. AD, Alzheimer's disease, CHD, coronary heart disease; PD, Parkinson's disease.

be less salient for participants who seek testing for non-health reasons. We found that participants motivated by ancestry still significantly updated their disease-risk perceptions for each result, though the magnitude of their perception changes was slightly smaller than for other participants (**Fig. 3**). Ancestry-motivated participants responded to good news with a –0.53-point decrease in their adjusted risk perception, whereas other participants showed an average –0.62-point decrease. The response to neutral news was even more similar across groups: –0.13 points for ancestry-motivated participants, and –0.14 points for other participants. The difference across motivation groups was most pronounced for bad news results, where ancestry-motivated participants had a 0.13-point increase in risk perception, whereas other participants had a 0.33-point increase. The pairwise difference between the magnitude of results was statistically significant ($P < 0.05$) across groups only for bad news results (corrected for multiple comparisons). In other words, overall risk updating was partially limited by motivation for testing, but the results imply that even those not primarily interested in health results still changed their risk perceptions similarly to those whose primary motivation was health related. Furthermore, the good news–bad news asymmetry was robust to individuals' motivation for testing.

Medical utilization results. To explore the role of PGT on individuals' healthcare choices, we analyzed how risk perception changes influenced reported follow-up tests, exams, procedures and appointments related to PGT results. To connect these reports with our perception analyses, we linked the 6-month survey results about healthcare utilization to participants' risk perception changes from baseline to 6 months after viewing results. Whereas the risk perception analyses were done at the participant–condition level, follow-up action survey questions were asked at the participant level and not linked to a specific condition. Therefore, our analysis in this section is for the 617 individuals in our 23andMe sample. We used two binary measures of medical utilization. The first measure, exams and procedures, took on a value of 1 for participants who said they had undergone exams, procedures, body scans or additional genetic tests as a result of seeing their 23andMe results. The second measure was participants' responses to the question of whether their 23andMe results prompted them to make an appointment with a medical professional(s). The analysis data sample for medical utilization survey responses can be found in the **Supplementary Data** file.

We found a significant ($P < 0.01$) correlation between participants' maximum risk perception change and their likelihood of making appointments or undergoing related exams or procedures but no significant correlation between median risk perception change and those same actions (**Fig. 4a**). This relationship was noted in both logistic and OLS regressions of medical follow-ups on maximum risk perception (**Supplementary Tables 5–8**). This result also held when we adjusted for average baseline risk perceptions across all eight conditions (**Supplementary Tables 5–8**). When adjusted for average baseline risk perception, both models suggested that a 1-point increase in maximum risk perception change (on a 5-point Likert scale) was associated with a 5% ($P < 0.01$) increase in likelihood of reporting having had a follow-up appointment and a 4% ($P < 0.01$) increase in the likelihood of reporting that exams or procedures were ordered after and related to PGT.

There was no significant association between median risk perception changes and follow-up medical actions (**Fig. 4b** and **Supplementary Tables 5–8**). Thus, participants' overall risk perception changes did not affect their propensity to seek related medical care.

Discussion of results and their implications. Our findings suggest that participants significantly altered their risk perceptions in response to their PGT results and that large risk perception changes were more likely to lead to follow-up medical actions. Assessing the ways in which individuals update their risk perception is difficult owing to the complex interactions among genetics, behavior and environment that influence one's true risk. Before our analyses, the salience of the test results was not obvious, because the conditions in the 23andMe tests had a wide range of heritability and behavioral components that were still new to the DTC market. Therefore, it was not clear what type of impact this set of information would have on individuals' risk perceptions.

However, the significance and modest magnitude of observed risk updating suggests that neither excessive overreaction nor complete disregard for the test results was prevalent in our sample. This result was also robust to different motivations for purchasing the tests. Furthermore, the perception updating results are novel, as prior studies on the impact of PGT did not test risk perception changes for actual PGT customers by condition-specific results and risk result groups.

The significant decrease in risk perception after typical and neutral results is of interest (though small in magnitude). A standard model of Bayesian learning would not predict such a decrease after controlling for baseline risk perception (as we do in the regression analysis). Therefore, one must ask why an individual's risk perception would shift downward

after a result that matched their baseline expectation. One possible explanation is ambiguity aversion[19], where reduction in uncertainty brings positive utility and is interpreted as a good news signal. Under this reasoning, an individual worried about being surprised by an elevated risk result would be relieved to find out that their risk is at the level expected, and this relief translates to decreased risk perception. Other explanations might include faulty memory of baseline perceptions after viewing results or natural increases in optimism over time. Further experimental evidence would be needed to adjudicate between these potential explanations for the typical and neutral news updating results.

The good news–bad news asymmetry in our overall risk updating results suggests that optimism bias is present even when information is very personalized. This is consistent with prior work showing similar results with nongenomic treatments in laboratory settings[13–16]. In addition to showing that the good news–bad news asymmetry effects exist in this real-world DTC setting, our results show that the effects persist a full 6 months after the information disclosure. Another key difference from the laboratory experiments of the good news–bad news effect is that the participants in our study were customers who sought the information on their own and thus may have had greater personal interest in the results. Unlike in other settings in which individuals receive information about disease risk, PGT results present novel information that requires genetic material (e.g., DNA from saliva) and cannot be replaced by knowledge of family medical history or general population risk figures. Whereas past experience might reasonably inform reactions to personalized information about other types of characteristics (e.g., intelligence and beauty), genetic makeup and risk for unrealized medical conditions are harder to infer from prior knowledge. Despite the novelty and specificity of information, the good news–bad news effect is clearly present in our sample.

One potential catalyst of the good news–bad news effect in the PGT setting is an individual's presumed ability to influence the risk of some of medical conditions. For some conditions (e.g., diabetes, coronary heart disease and lung cancer), the negative surprise of an unexpected elevated risk result might be mitigated by an intention to make lifestyle changes. For example, increasing exercise, changing one's diet or quitting smoking in the months after seeing test results might counterbalance the perception effects of the elevated risk result. Of course, less healthy behavior after decreased (good news) risk results could have the opposite effect, but mitigation of behavioral reactions seems more likely after elevated (bad news) results.

The variation in risk perception updating across conditions shows that the risk result group is not the only information content that matters. For some conditions, such as Parkinson's disease, we observed significant risk changes in response to each type of news, as well as significant good news–bad news asymmetry. For other conditions, such as coronary heart disease, our results show more modest levels of perception changes and no good news–bad news asymmetry. Though we tried to uncover the drivers of this variation using measures of mortality, population risk and heritability, no simple pattern emerges (**Supplementary Table 9**). Good news results for the neurological diseases, which are the least preventable and treatable of the eight conditions studied here, led to the largest decline in average risk perception.

Given the exceptional attention that breast cancer testing receives in the media, we should not be surprised that our breast cancer results follow a slightly different pattern from that of other conditions. Our analyses show a strong decline in risk perception (of similar magnitude) after both good news and neutral news, but a small and statistically insignificant rise after bad news. Our estimation precision is relatively reduced for breast cancer because it is sex specific in our sample (323 participants were female); however, the point estimates for average updating suggest that participants were skeptical of bad news and equally relieved by neutral and good news. One explanation of this response pattern is that decreased and typical risk results have the same practical implications (no change in screening and prevention behaviors). This condition-level result suggests that within the overall cautious and asymmetric perception updating pattern, the nature of the condition altered the magnitude of the response.

When considering the regulation of PGT, an important question is how individual cognitive shifts translate into medical decisions and behaviors. The PGen Study's survey design provided a unique opportunity to link actions to genetic-testing results. Since participants received a bundle of test results rather

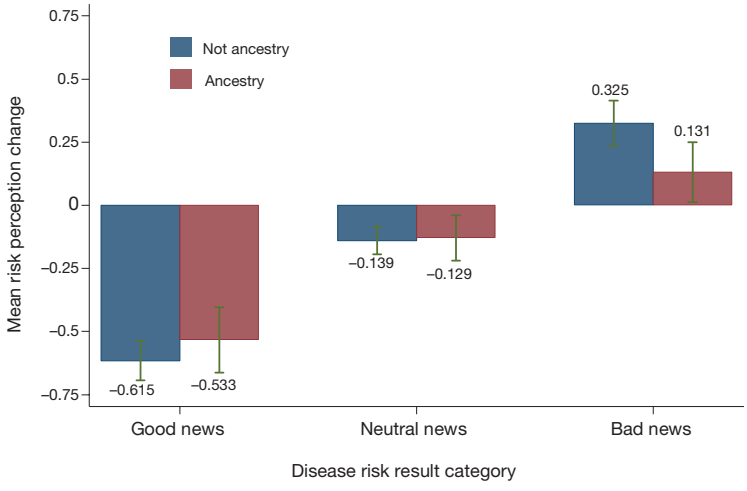

Figure 3 Risk perception changes by risk result group and motivation for testing. Adjusted mean risk perception changes corresponding with good news, neutral news and bad news risk results. The ancestry group consisted of the participants who indicated that ancestry was their primary motivation for purchasing PGT. We calculated adjusted mean risk perception change using OLS regressions of risk perception change (measured using a 5-point Likert scale) on risk result categories, using interaction variables to estimate the differential effects by motivation group. Error bars represent the 95% confidence intervals. The sample included 4,075 participant–condition observations (1,056 good news, 1,382 neutral news and 1,637 bad news results) from 617 23andMe customers. 182 (29.5%) of the participants indicated in the survey that ancestry information was their primary motivation for purchasing PGT. Estimates include control variables for baseline risk perception. Standard errors are clustered at the participant level. All group responses were statistically different from 0 ($P < 0.05$), and perception changes in response to elevated test results were significantly lower than risk perception changes for the ancestry group ($P < 0.05$).

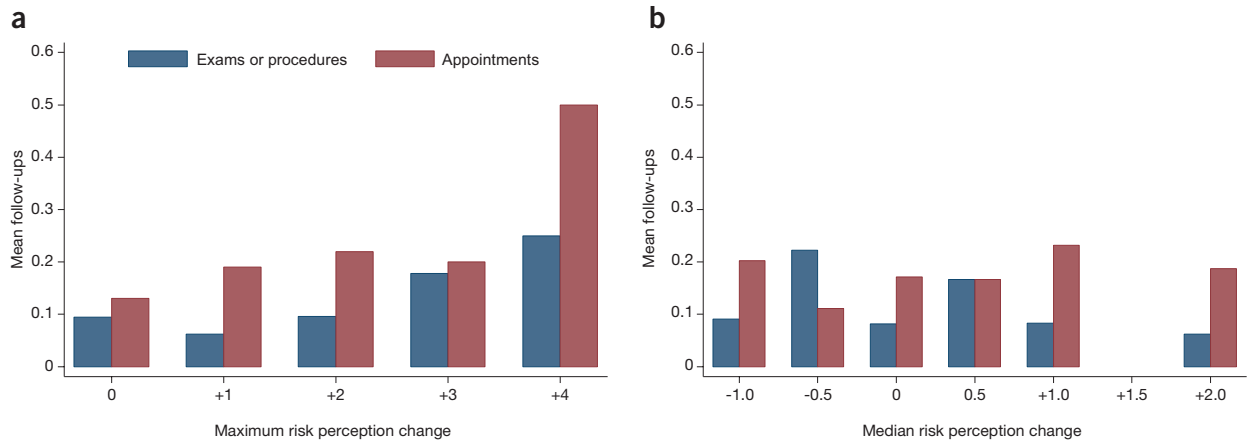

Figure 4 Medical utilization according to maximum and median risk perception change. (**a**,**b**) Likelihood of medical follow-up action by maximum (**a**) and median (**b**) risk perception change. The two outcome variables were 'exams or procedures' (if the participant self-reported undergoing an exam or procedure as a result of their results) and 'appointments' (if the participant self-reported having made a medical appointment related to their results). In **b**, median risk values of –2 and –1.5 were excluded because only six participants showed those values (each median risk perception change bar shown includes at least 10 participants). OLS and logistic regressions confirm a positive and statistically significant correlation ($P < 0.01$) between maximum risk perception change and follow-up appointments, as well as between maximum risk perception change and exams or procedures (see **Supplementary Methods** for regression details).

than a single risk ratio, the data allowed us to evaluate which sets of results and risk updating corresponded to follow-up medical action. Furthermore, we were able to test whether risk perception changes led to more follow-up medical actions; and, if so, whether such action(s) were driven by extreme perception changes in response to a single risk result, or by average perception change trends across the bundle of risk results. Our analyses indicated that large changes in risk perception for single conditions, rather than trends across all conditions, were associated with reports of medical appointments, tests, exams or procedures in response to PGT results. A pattern of general and moderately increased concern across several conditions did not appear to spark decisions to use medical services, whereas a single higher-amplitude risk perception change increased the odds of engaging in follow-up medical action. This is consistent with previous work that reported no significant associations between composite measures of risk and health behavior among DTC genetic-testing recipients who were recruited specifically for the purposes of research[2], and it goes further by directly linking follow-up actions to the test results and revealing how maximum risk increases correspond to more appointments, tests, exams and procedures. Given the time frame of the study (6 months), some participants might eventually seek related appointments, exams or procedures after the study period. Given the possibility that some patients will wait until routine medical-checkup appointments (unrelated to their PGT results) to bring up questions, we consider our results for follow-up medical action to be conservative estimates.

The medical usage results also help to validate the risk perception results by confirming that increased risk perceptions have real consequences. Whereas a survey response about risk beliefs has little cost to the research subject, follow-up appointments, exams or procedures require investments in time and even psychological anticipation[20]. Further work is needed to better understand the specific thresholds for acting on risk perception change and how the level and variation of an individual's good news–bad news learning asymmetry affects these thresholds.

Limitations. Our study's limitations fall into four general areas: sample constraints, risk category tradeoffs, self-reporting issues and clinical utility measurement. First, because the DTC genomics industry was in its early stages at the time of the PGen Study, the composition of the sample reflects the participants' 'early-adopter' status and as such is not representative of the broader population in terms of observable variables (e.g., they are predominantly white and well educated) and unobservable factors (e.g., level of curiosity and personal value of the PGT info). Additionally, study participants were volunteers willing to contribute their information and give their time filling out surveys. This sample limitation is a challenge for many studies involving early-stage information technologies, and we cannot claim that our study sample is representative of the general population.

Although the nature of the sample limits the generalizability of our results, our risk perception results comparing ancestry-motivated to health risk–motivated participants mitigate generalizability concerns. These results showed that consumers with relatively less interest in health risk results still significantly updated their beliefs after viewing their PGT outcomes. Early adopters motivated by interest in their ancestry might still differ from late adopters along other important dimensions (e.g., socioeconomic status, education and ethnic background), but interest in health results does not appear to be a major factor limiting generalizability.

Second, the categorical nature of the survey questions (Likert scale responses) constrains the precision of our risk perception results. The advantage of the 5-point scale is that it is straightforward for the survey respondents. However, these simple risk groups can lead to censoring in the extremes of the distribution (e.g., participants cannot change their risk beliefs further downward from '1, much lower than average'), which could lead to under-measurement of risk perception changes. These categories are inherently nonlinear, with underlying risk levels that are not equally spaced along a continuum. We do not believe that such nonlinearity threatens our main risk updating and good news–bad news results, both because this nonlinearity would be symmetric (same for below- and above-average perceptions) and because the OLS and ordinal logit regressions produce the same qualitative result. However, future

studies might benefit from continuous, rather than ordinal, risk response designs.

Self-reporting error is another limitation of our study. This issue is a well-known challenge in survey design. We limited our medical usage outcomes to survey questions asking about discrete events that clearly linked action to PGT results to minimize vague interpretations. But self-reporting error cannot be eliminated under this research design. Furthermore, the 6-month risk perception changes and medical usage choices are only a snapshot of an individual's beliefs and behavior. Longer-term tracking and longer follow-up surveys would be needed to understand how these perceptions and actions evolve over time.

Finally, the study can measure only the increases in healthcare usage as a result of increased risk perceptions and does not provide conclusions about the overall health value of PGT. Decreased risk or good news PGT results might give a consumer false reassurance and lead to a reduction in healthcare usage or health-enhancing behaviors. An ideal study design might include an individual's entire medical history as well as measures of health behaviors both before and after PGT. However, the time horizon and survey limitations of the PGen Study did not allow this. Whereas the study provides a step toward better empirical understanding of the psychological and behavioral impact of PGT, the data do not allow us to measure the long-term health benefits (or costs) of PGT. We hope this study will encourage future attempts for linking information interventions with medical records and long-term behavioral tracking, as well as qualitative data on risk perception.

A frequently cited concern regarding the regulation of DTC genomics is a lack of understanding about how individuals respond to the information presented in these tests. Our results provide early evidence of how customers adjust their perceptions and engage with their health providers as a result of different types of PGT results. Though we found good news–bad news asymmetry in risk perception changes, these changes appeared to be moderate and congruent with test results. Furthermore, extreme perception changes drove much of the follow-up medical appointments and procedures. Taken together, our results suggest that DTC consumers learn from their PGT results and update their beliefs, but they primarily seek additional medical actions in response to large and unexpected risks.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.3661).

Editor's note: This article has been peer-reviewed.

AUTHOR CONTRIBUTIONS
J.L.K. formulated the research question, designed and performed the analyses, prepared the figures and wrote the manuscript. F.M. helped with the research design, analyses, and manuscript writing. R.C.G. and J.S.R. designed and implemented the surveys, in their roles as the primary investigators of the Impact of Personal Genomics (PGen) study, and advised on the preparation of the figures and the manuscript. All authors edited and approved the manuscript.

ACKNOWLEDGMENTS
We thank Tali Sharot, Alec Brandon, Jean Tirole, Alexandra Rodman, Daniel Fehder, Abhishek Nagaraj, Rebecca Grunberg, Roberto Fernandez, Sarah Kalia and members of the PGen Study Group (particularly Kurt Christensen, Deanna Alexis Carere, and Joanna L. Mountain). Members of the PGen Study Group are listed in **Supplementary Note 2**. The PGen Study is supported by the US National Institutes of Health (NIH) National Human Genome Research Institute (R01-HG005092). R.C.G. was additionally supported by the following NIH grants: U01-HG006500, U19-HD077671, U01-HG008685 and U41-HG006834.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.3661).

Joshua L Krieger[1], Fiona Murray[1], J Scott Roberts[2] & Robert C Green[3]

[1]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [2]Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, Michigan, USA. [3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Partners HealthCare Personalized Medicine, the Broad Institute and Harvard Medical School, Boston, Massachusetts, USA.
e-mail: jkrieger@mit.edu

1. McGuire, A.L. & Burke, W. *J. Am. Med. Assoc.* **300**, 2669–2671 (2008).
2. Bloss, C.S., Schork, N.J. & Topol, E.J. *N. Engl. J. Med.* **364**, 524–534 (2011).
3. Bloss, C.S., Topol, E.J. & Schork, N.J. *Genet. Epidemiol.* **36**, 66–70 (2012).
4. Christensen, K.D., Roberts, J.S., Uhlmann, W.R. & Green, R.C. *Genet. Med.* **13**, 409–414 (2011).
5. Green, R.C. *et al. N. Engl. J. Med.* **361**, 245–254 (2009).
6. Kaufman, D.J., Bollinger, J.M., Dvoskin, R.L. & Scott, J.A. *J. Genet. Couns.* **21**, 413–422 (2012).
7. Ostergren, J.E. *et al. Public Health Genomics* **18**, 216–224 (2015).
8. Roberts, J.S. & Ostergren, J. *Curr. Genet. Med. Rep.* **1**, 182–200 (2013).
9. Shiloh, S. *et al. Clin. Genet.* **87**, 117–123 (2015).
10. Bansback, N., Sizto, S., Guh, D. & Anis, A.H. *Genet. Test. Mol. Biomarkers* **16**, 1165–1171 (2012).
11. Oster, E., Shoulson, I. & Ray Dorsey, E. *Am. Econ. Rev.* **103**, 804–830 (2013).
12. Oster, E., Shoulson, I., Quaid, K. & Ray Dorsey, E. *J. Public Econ.* **94**, 1041–1050 (2010).
13. Eil, D. & Rao, J.M. *Am. Econ. J. Microecon.* **3**, 114–138 (2011).
14. Moutsiana, C. *et al. Proc. Natl. Acad. Sci. USA* **110**, 16396–16401 (2013).
15. Sharot, T. *et al. Proc. Natl. Acad. Sci. USA* **109**, 17058–17062 (2012).
16. Sharot, T., Korn, C.W. & Dolan, R.J. *Nat. Neurosci.* **14**, 1475–1479 (2011).
17. Tversky, A. & Kahneman, D. *Science* **185**, 1124–1131 (1974).
18. Kuhnen, C.M. *J. Finance* **70**, 2029–2062 (2015).
19. Ellsberg, D. *Q. J. Econ.* **75**, 643–669 (1961).
20. Köszegi, B. *J. Health Econ.* **22**, 1073–1084 (2003).

Containment of transgenic trees by suppression of *LEAFY*

To the Editor:

Field studies and commercial use of genetically engineered (GE) trees have been limited, in large part owing to concerns over transgene flow into wild or feral tree populations[1–4]. Unlike other crops, trees are long-lived, weakly domesticated and their propagules can spread over several kilometers[5]. Although male sterility has been engineered in pine, poplar, and eucalyptus trees grown under field conditions by expression of the barnase RNase gene in anther tapetal cells[6,7], barnase can reduce rates of genetic transformation and vegetative growth[6]. Furthermore, barnase expression may not be fully stable[8]. Bisexual sterility would allay concerns over seed dispersal, could be used to control invasive exotic trees, and might increase wood production[9]. We report the use of RNA interference (RNAi) to suppress expression of the single-copy *LEAFY* (*LFY*) gene to produce sterility in poplar.

RNAi has been used to reduce gene expression in many plant species[10,11], and the reduction in gene expression that RNAi confers is highly stable in trees under field conditions[12]. *LFY* is required for the early stages of male and female floral organ formation in plants, and encodes a transcription factor that promotes floral meristem identity[13,14]. In *Arabidopsis thaliana*, loss of *LFY* function results in the formation of vegetative structures instead of floral meristems, whereas reduction of *LFY* expression decreases floral abundance and results in partial conversion of floral organs to leaf-like structures[13,14]. We selected *LFY*