Genetic variants associated with motion sickness point to roles for
inner ear development, neurological processes, and glucose
homeostasis

Bethann S. Hromatka¹, Joyce Y. Tung¹, Amy K. Kiefer¹, Chuong B. Do¹, David A.
Hinds¹, and Nicholas Eriksson¹*

¹23andMe, Inc., Mountain View, CA, USA
*nick@23andme.com

February 4, 2014

Abstract

Roughly one in three individuals is highly susceptible to motion sickness and yet the underlying causes of this condition are not well understood. Despite high heritability, no associated genetic factors have been discovered to date. Here, we conducted the first genome-wide association study on motion sickness in 80,494 individuals from the 23andMe database who were surveyed about car sickness. Thirty-five single-nucleotide polymorphisms (SNPs) were associated with motion sickness at a genome-wide-significant level ($p < 5 \times 10^{-8}$). Many of these SNPs are near genes involved in balance, and eye, ear, and cranial development (e.g., $PVRL3$, $TSHZ1$, $MUTED$, $HOXB3$, $HOXD3$). Other SNPs may affect motion sickness through nearby genes with roles in the nervous system, glucose homeostasis, or hypoxia. We show that several of these SNPs display sex-specific effects, with as much as three times stronger effects in women. We searched for comorbid phenotypes with motion sickness, confirming associations with known comorbidities including migraines, postoperative nausea and vomiting (PONV), vertigo, and morning sickness, and observing new associations with altitude sickness and many gastrointestinal conditions. We also show that two of these related phenotypes (PONV and migraines) share underlying genetic factors with motion sickness. These results point to the importance of the nervous system in motion sickness and suggest a role for glucose levels in motion-induced nausea and vomiting, a finding that may provide insight into other nausea-related phenotypes such as PONV. They also highlight personal characteristics (e.g., being a poor sleeper) that correlate with motion sickness, findings that could help identify risk factors or treatments.
Introduction

Motion sickness is provoked by exposure to a variety of motions (e.g., traveling in cars, boats, or planes; amusement park rides; skiing; and riding on camels) [1]. Simulators and virtual reality environments can also induce motion sickness [2]. Symptoms of motion sickness include dizziness, nausea, vomiting, headache, and pallor [3]. Sweating, drowsiness, increased salivation, hyperventilation, and emotional distress may also occur. Motion sickness is associated with other conditions including migraines, vertigo, postoperative nausea and vomiting (PONV), and chemotherapy-induced nausea and vomiting (CINV) [1,4].

Roughly one in three individuals is highly susceptible to motion sickness and the rest of the population may experience motion sickness under extreme conditions [5]. The underlying etiology of motion sickness, however, is not well understood. One theory suggests that motion sickness results from contradictory information the brain receives during motion [1,5]. The vestibular system of the inner ear, which senses motion and body position and influences balance, signals “moving” to the brain, while the eye signals “stationary” because the car or boat appears stationary relative to the viewer. The vestibular system is also thought to serve as a sensor of disequilibrium-causing neurotoxins (i.e., a toxin detector) and is believed to trigger the emetic response in order to rid the body of toxins. Thus, motion sickness may be an aberrant trigger of the emetic response. Evidence for the involvement of the vestibular system comes from the observation that individuals with complete loss of the vestibular apparatus, a component of the vestibular system, are immune to motion sickness [1].

A variety of factors influence risk for motion sickness. Women are more susceptible than men [6–9] and younger individuals are at increased risk [8,9]. Ancestry may also play a role; there is some evidence that motion sickness occurs more frequently in individuals with Asian ancestry compared to European ancestry [10,11]. Some variables are situational and/or behavioral. For instance, one study showed that passengers without a view of the road ahead were about three times more likely to experience illness [8] and another report suggested that adopting a wider stance may reduce motion sickness [12]. There is also evidence that diet and eating behavior influence risk [7].

Perhaps the most important and least understood variable is the underlying physiological susceptibility of the individual. In women, increased cortisol levels are predictive of motion sickness [13] and susceptibility to motion sickness changes as a function of the menstrual cycle, suggesting that levels of estrogen and other hormones might play a role [14]. In both sexes, hyperglycemia is implicated in motion-induced nausea and vomiting [15]. There is also some evidence that lower baseline levels of adrenocorticotropic hormone (ACTH) [16], also known as corticotropin, and low sympathetic nervous system activity [17] increases susceptibility. Finally, since antihistamines (e.g., Dramamine), anticholinergics (e.g., scopolamine), and sympathomimetics (e.g., d-amphetamine and ephedrine) are effective treatments, altered baseline activity of the receptors these drugs bind to might influence risk for motion sickness.

Although heritability estimates for motion sickness range from 57-70% [18], genome-wide association studies (GWAS) on this phenotype have not been reported. Here, we describe a large GWAS in which we find 35 regions significantly associated with motion sickness.

Results

Genome-wide Association Study of Motion Sickness

We performed a GWAS in 80,494 individuals from the customer base of 23andMe, Inc., a personal genetics company. Participants were of primarily European ancestry and were at most distantly related to each other (i.e., first cousins and closer were excluded). Motion sickness was assessed using online self-report. Participants responded to questions about their degree of car sickness and questions were combined onto a scale of 0 (never motion sick), 1 (occasionally), 2 (sometimes), or 3 (frequently). Details about the cohort can be found in Table 1 and in the Methods. All
analyses were controlled for age, sex, and five principal components of genetic ancestry. Manhattan and quantile-quantile plots are provided in Figures 1 and S1.

Lead SNPs with p-values under $5 \times 10^{-8}$ for motion sickness are shown in Table 2; 35 regions were significant (Figure 2A). We created a genetic propensity score based on the number of risk alleles for the 35 index SNPs. Individuals in the top five percent of the distribution (allele dosage of 40.25 or more risk alleles) had an average motion-sickness score 0.546 units higher than those in the bottom five percent (28.37 or fewer risk alleles). The top five percent had 6.37 times increased odds of being “frequently” motion sick as opposed to “never” motion sick as compared to the bottom five percent. The variance in motion sickness explained by the propensity score (which may be inflated as it was assessed in the discovery population) was 0.029.

A few associated SNPs are in regions implicated in eye and ear development or balance. For example, our most significant association is with rs66800491 ($p = 4.2 \times 10^{-44}$), located roughly one Mb upstream of PVRL3, which encodes the cell adhesion protein Nectin-3. Loss of PVRL3 expression in both humans and mice results in ocular defects [19]. The SNP rs10514168 ($p = 2.7 \times 10^{-9}$) is located downstream of TSHZ1, a gene involved in inner ear development in the mouse [20]. Another association is with rs2153535 ($p = 2.7 \times 10^{-18}$), located upstream of MUTED, which is implicated in balance [21]. Three additional associated SNPs are near genes with major roles in early development: rs2551802 ($p = 2 \times 10^{-12}$) between HOXD3 and HOXD4; rs9906289 ($p = 6.4 \times 10^{-11}$) in HOXB3; and rs149951341 ($p = 3.4 \times 10^{-12}$) near TLE4. The HOXD SNP is in LD ($r^2 \approx 0.9$) with rs2072590, which is associated with ovarian cancer [22].

Several other associated SNPs are located near genes involved in neurological processes including synapse development and function: rs11713169 ($p = 5.9 \times 10^{-13}$) in NLGN1 encoding neuroligin; rs6069325 ($p = 7.2 \times 10^{-21}$) upstream of CBLN4 encoding a member of the cerebellin precursor protein family; rs62018380 ($p = 1.7 \times 10^{-9}$) downstream of MCTP2, a gene involved in intercellular signal transduction and synapse function; rs7957589 in PDZRN4 ($p = 7.9 \times 10^{-10}$) near CNTN1 (contactin 1), which plays a role in axon guidance during neural development [23]; and two independent SNPs, rs4343996 and rs34912216 ($p = 8.7 \times 10^{-11}$ and $2.7 \times 10^{-8}$, respectively) in SDK1 encoding sidekick-1, a cell adhesion molecule that localizes to synapses. The SNP rs2150864 ($p = 6.3 \times 10^{-15}$) is located about 1.5 Mb upstream of LINGO2, a gene implicated in essential tremor [24]. Additional associated SNPs in or near genes in neurological pathways include: rs9834560 ($p = 9.7 \times 10^{-15}$) in CPNE4 encoding copine-4, and two independent SNPs in or near AUTS2 (rs1195218 and rs6946099 ($p = 4.5 \times 10^{-22}$ and $1.9 \times 10^{-9}$, respectively).

Other associated SNPs are in regions involved in glucose and insulin homeostasis. For example, the second most significant association we found is with rs56051278 ($p = 1.5 \times 10^{-29}$) in GPD2 that encodes glycerol-3-phosphate dehydrogenase 2, an enzyme implicated in glucose homeostasis. This SNP is in high LD ($r^2 \approx 0.8$) with rs2116665 (the non-synonymous substitution H264R in GPD2) that was previously associated with free fatty acid and glycerol levels [25]. The SNP rs11129078 ($p = 3.4 \times 10^{-21}$) is located downstream of UBE2E2, which encodes a component of the ubiquitin-proteasome system. This system is implicated in the autophagy of pancreatic beta-cells that produce insulin and plays important roles in insulin homeostasis [26]. In addition, rs705145 ($p = 1.4 \times 10^{-21}$) is located just upstream of GPR26, encoding a G protein-coupled receptor. Mice with a deletion of the GPR26 gene develop hyperphagia and diet-induced obesity, which leads to metabolic complications linked to obesity including glucose intolerance, hyperinsulenia and dyslipidemia [27]. The SNP rs4076764 ($p = 2.9 \times 10^{-9}$) is located upstream of RGS5, a regulator of G protein signaling. Loss of RGS5 in the mouse is also associated with hyperphagia [28]. Finally, rs7170668 ($p = 1 \times 10^{-10}$) is located upstream of NR2F2 encoding COUP-TFI (chicken ovalbumin upstream promoter transcription factor II), a protein with roles in glucose homeostasis and energy metabolism [29].

The remaining associated SNPs are in regions implicated in hypoxia (rs1858111 near RWDD3, $p =$
Figure 1: Manhattan plot. The 35 genome-wide significant regions are listed with the proposed candidate gene; regions that are close together share a label.

Table 1: Cohort statistics for motion sickness GWAS. Degree of motion sickness stratified by sex and age. Females and younger people tend to be more motion sick.

<table>
<thead>
<tr>
<th>Group</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
<th>≤ 30</th>
<th>31–45</th>
<th>46–60</th>
<th>≥ 61</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>40,042</td>
<td>25,137</td>
<td>14,905</td>
<td>5,510</td>
<td>10,707</td>
<td>10,592</td>
<td>13,233</td>
</tr>
<tr>
<td>Occasionally</td>
<td>24,902</td>
<td>11,855</td>
<td>13,047</td>
<td>4,597</td>
<td>8,451</td>
<td>6,459</td>
<td>5,395</td>
</tr>
<tr>
<td>Sometimes</td>
<td>6,723</td>
<td>3,067</td>
<td>3,656</td>
<td>1,175</td>
<td>2,015</td>
<td>1,759</td>
<td>1,774</td>
</tr>
<tr>
<td>Frequently</td>
<td>8,827</td>
<td>3,011</td>
<td>5,816</td>
<td>1,784</td>
<td>3,173</td>
<td>2,310</td>
<td>1,560</td>
</tr>
<tr>
<td>Total</td>
<td>80,494</td>
<td>43,070</td>
<td>37,424</td>
<td>13,066</td>
<td>24,346</td>
<td>21,120</td>
<td>21,962</td>
</tr>
</tbody>
</table>

4.1 × 10^{-14}; iron homeostasis (rs10970305 near ACO1, \( p = 1 \times 10^{-27} \)); brown adipose tissue (rs1759167 in PRDM16, \( p = 3.5 \times 10^{-13} \)); and other less characterized processes: rs2360806 (\( p = 7.2 \times 10^{-11} \)) in ST18, rs2318131 (\( p = 3.3 \times 10^{-12} \)) near COP58, rs60464047 (\( p = 6.7 \times 10^{-9} \)) in POU6F2, rs10752212 (\( p = 1.1 \times 10^{-10} \)) near CELF2, rs6833641 (\( p = 1.8 \times 10^{-9} \)) near ARAP2, rs17515225 and rs4311235 (independently associated, \( p = 2.5 \times 10^{-9} \) and \( 7.9 \times 10^{-9} \), respectively) in LRP1B, rs13788552 (\( p = 4.3 \times 10^{-9} \)) in a gene desert on 4q34.3; and rs1782032 (\( p = 9 \times 10^{-9} \)) near TUSC1, and rs34311235 (independently associated, \( p = 2.5 \times 10^{-9} \) and \( 7.9 \times 10^{-9} \), respectively) in LRP1B, rs13788552 (\( p = 4.3 \times 10^{-9} \)) in a gene desert on 4q34.3; and rs1782032 (\( p = 9 \times 10^{-9} \)) near TUSC1, and rs1847202 near SHQ1 and GXYLT2. Finally, rs997295 in MAP2K5 (\( p = 3.3 \times 10^{-9} \)) is in LD with rs2241423 (\( r^2 \approx 0.36 \)), which is associated with BMI [30].

### Enrichment

Analysis of all regions with \( p < 10^{-5} \) using GREAT [31] showed a significant enrichment in regions containing genes involved in fusion of atlas and occipital bones (FDR=0.002) and abnormal arcus anterior morphology (FDR=0.038) in mouse. The genes annotated with one or both of these processes were HOXB, HOXD, TSHZ1, and RARB regions (the SNP near RARB is rs2067120, \( p = 8.2 \times 10^{-6} \)).
to the positive strand of build 37 of the human genome. The effect is the change per copy of the second
morbidities (migraines, vertigo, PONV, and morning
sex (where applicable), and 5 principal components.
We investigated comorbidities with motion sickness
rs34311235 2q22.2 142,767,433 C/T 7
rs10514168 18q22.3 73,098,949 A/C 2
rs66800491 3q13.2 54,139,486 G/T 7
rs2315355 2p13.2 8,369,679 C/G 2
rs2510864 9p21.1 29,363,265 A/G 6
rs9834560 3q22.1 131,716,105 A/G 9
rs1858111 1p21.3 96,089,731 A/G 4
rs61759167 1p36.32 3,091,587 C/T 3
rs11713169 3q26.31 173,384,589 A/C 5
rs2551802 2q31.1 177,022,158 C/G 2
rs2318131 2q37.3 237,933,966 C/T 3
rs149513141 9q21.31 81,268,149 A/C 4
rs9006289 17q21.32 46,644,677 C/T 4
rs2368006 8q11.23 53,125,734 A/C 7
rs4434996 7p22.2 3,362,642 A/G 8
rs7170668 15q26.2 96,014,134 C/T 1
rs10552212 10p14 10,917,121 A/G 1
rs7957589 12q12 41,874,292 A/T 7
rs60218380 15q26.2 95,275,917 A/C 10
rs6833641 4p15.1 35,563,786 C/G 10
rs6946969 7q11.22 70,211,027 A/G 9
rs15715225 22q11.21 141,545,755 C/T 2
rs10514168 18q22.3 73,098,949 A/C 2
rs4076764 1q21.3 163,441,286 C/T 2
rs997295 15q23 68,016,343 C/T 3
rs1378552 4q34.3 180,356,846 C/T 3
rs60464047 7p11.22 39,418,538 A/T 7
rs34311235 2q22.2 142,767,433 C/T 7
rs1782032 9q21.2 25,804,285 A/G 9
rs1847203 3p13 72,934,371 C/T 2
rs34912216 7p22.2 4,118,377 A/G 2

Table 2: Genome-wide significant index SNPs. Alleles are reported in alphabetical order with respect to the positive strand of build 37 of the human genome. The effect is the change per copy of the second allele on a four point scale of increasing motion sickness. Frequency is the frequency of the alphabetically second allele in the cohort. Quality is imputation r^2 for imputed SNPs, call rate for genotyped SNPs (those marked with a *). Gene is a proposed candidate gene in the region.

Phenotypic Study of Motion Sickness

We investigated comorbidities with motion sickness within the 23andMe database. Briefly, we looked at partial correlations between each of 695 different phenotypes and motion sickness, controlling for age, sex (where applicable), and 5 principal components. Table 3 shows selected large correlations. Some of the associated phenotypes are known symptoms of motion sickness (e.g., headache) or established co-morbidities (migraines, vertigo, PONV, and morning sickness). In addition to PONV, other gastrointestinal (GI) phenotypes were also associated with motion sickness (e.g., irritable bowel syndrome (IBS); acid reflux; stomach upset with antidepressants, codeine and nonsteroidal anti-inflammatory drugs (NSAIDs); and indigestion with dairy products). Other associations include poor sleep, poor circulation, altitude sickness, hay fever, and neuroticism. Phenotypes associated with lower risk for motion sickness include a history of tobacco use, a good sense of direction, higher BMI, being single, and a better ability to han-
dle stress.

### Genetic Correlations Between Motion Sickness and Related Phenotypes

We determined if any of the 35 SNPs associated with motion sickness were also associated with six correlated and clinically important phenotypes (PONV, migraines, hay fever, altitude sickness, morning sickness, and vertigo). Table 4 shows SNPs associated with these phenotypes with a (Bonferroni-corrected) \( p \)-value under 0.05/35 \( \approx 0.0002 \) (under a more stringent threshold of 0.05/(35 \times 6) \( \approx 0.0002 \) only the first two are significant). One motion sickness-associated SNP was significantly associated with migraines: rs61759167 in PRDM16 (\( p = 1.1 \times 10^{-6} \)). A previous study [32] reported an association between migraines and another SNP in PRDM16, rs2651899, which is in weak LD with rs61759167 (\( r^2 \approx 0.44 \)). Three motion sickness-associated SNPs were also significantly associated with PONV: rs6833641 near ARAP2, rs1195218 near AUTS2, and rs6069325 near CBLN4. For all four examples, the higher risk allele for migraines or PONV is also the higher risk allele for motion sickness. We did not detect significant associations between motion sickness-associated SNPs and altitude sickness, hay fever, morning sickness, or vertigo. While this data suggests some shared etiology for motion sickness and PONV or migraines, it is difficult to assess whether or not this is due to shared causal SNPs.

### Sex-specific effects

Motion sickness is much more common in women than in men (Table 1) and several of our SNPs show much stronger effects in women than in men. The SNP rs66800491 has a 1.5x larger effect in women (-0.097 versus -0.062) and rs1847202 has a 3x larger effect in women (0.048 versus 0.016) (both SNPs \( p < 0.05 \) for interaction, corrected for 35 tests). Overall 26 of the 35 SNPs have estimated larger effects in women than men (binomial \( p < 0.003 \); Table S1).

### Discussion

Here we report 35 novel, genome-wide significant associations for motion sickness (Table 2). Genes in regions associated with motion sickness appear to play roles in eye and ear development, balance and other neurological processes, and glucose homeostasis. Two of the genome-wide significant regions contain hypoxia-inducible genes. We also provide evidence that motion sickness is phenotypically associated with numerous conditions and traits (Table 3).

Since motion sickness is thought to stem from the brain receiving contradictory signals from the inner ear, the inner ear may be involved in one of the SNPs that are also associated with motion sickness.
ear versus the eye (e.g., the inner ear signals “moving” while the eye signals “stationary”), it is interesting that a region implicated in eye development (rs56100358 near PVRL3) is our most significant association. Chromosomal rearrangements that lead to loss of PVRL3 expression have been associated with ocular defects in humans and the PVRL3 knockout mouse exhibits lens and other vision problems [19]. The associations with regions involved in the inner ear (rs12111385 near MUTED and rs1435985 near TSHZ1) are also interesting since disturbances in the vestibular system of the inner ear, which senses motion and body position and influences balance, are thought to play a central role in motion sickness. It has been suggested that the mouse homolog of MUTED controls the synthesis of otoliths of the vestibular labyrinth of the inner ear [21]. Otoliths are sensitive to gravity and linear acceleration and play a role in balance. Mutations in TSHZ1 and deletions in the 18q22.3 region that includes TSHZ1 are associated with congenital aural atresia (CAA) [33, 34], a spectrum of ear deformities that involve malformation of the external auditory canal (EAC). More generally, our enrichment analysis suggests that genes involved in certain aspects of cranial developmental may play an important role in motion sickness. Associations with SNPs in or near genes involved in synapse formation and function (NLGN1, CBLN4, MCTP2, PDZRN4, CNTN1, and SDK1) and other neurological pathways (LINGO2, CPNE4, AUTS2) point to the importance of the brain in motion sickness.

Five associated SNPs are in or near genes implicated in glucose and insulin homeostasis or BMI. Although these SNPs are not in linkage disequilibrium (LD) with SNPs reported in GWAS of type 2 diabetes (T2D) [35–40], rs6051278 is in high LD ($r^2 \approx 0.8$) with rs2116665, a nonsynonymous substitution (H264R) in the GPD2 gene. H264R has been associated with increased plasma glycerol and free fatty acid concentrations in a French Canadian population [25]. Increased free fatty acid levels are indicative of glucose intolerance and hyperinsulinemia. Although it is unclear why genes involved in glucose and insulin regulation might also play a role in motion sickness, one study suggested that hyperglycemia may be related to the gastrointestinal symptoms of motion sickness [15]. In this study, individuals who experienced motion-induced nausea and vomiting had lower levels of insulin than people who did not experience gastrointestinal symptoms. The study further suggested that stable glucose levels might help to relieve motion-induced gastrointestinal upset.

At least two of our associated SNPs are near hypoxia-inducible genes: RGS5 and RWDD3 (encoding the RSUME protein). RSUME promotes the activity of hypoxia-inducible factor 1 (HIF-1alpha), a master regulator of the hypoxic response [41]; RGS5 is an apoptotic stimulator induced by hypoxia in endothelial cells [42]. This data suggests a potential relationship between motion sickness and hypoxia. Motion sickness might lead to hypoxia or individuals predisposed to hypoxia might also be more susceptible to motion sickness. Both possibilities are intriguing since our phenotypic analysis suggested an association between motion sickness and altitude sickness, which occurs when individuals become hypoxic at higher altitudes (Table 3).

Among the regional association plots (Supplement), one SNP in particular stands out: rs1195218

### Table 4: Significant associations between motion sickness-associated SNPs and other phenotypes.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>SNP</th>
<th>N</th>
<th>pvalue</th>
<th>effect</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migraine</td>
<td>rs61759167</td>
<td>72,901</td>
<td>$1.1 \times 10^{-6}$</td>
<td>0.08</td>
<td>(0.051, 0.119)</td>
</tr>
<tr>
<td>PONV</td>
<td>rs1195218</td>
<td>25,223</td>
<td>0.00012</td>
<td>-0.14</td>
<td>(-0.213, -0.069)</td>
</tr>
<tr>
<td>PONV</td>
<td>rs6069325</td>
<td>25,223</td>
<td>0.00079</td>
<td>0.09</td>
<td>(0.038, 0.143)</td>
</tr>
<tr>
<td>PONV</td>
<td>rs6833641</td>
<td>25,223</td>
<td>0.00101</td>
<td>0.09</td>
<td>(0.037, 0.148)</td>
</tr>
</tbody>
</table>

$N$ is the number of people with data for motion sickness and migraines or PONV.
near AUTS2. This genotyped SNP has a p-value under $10^{-20}$ and no other SNPs in the region have $p < 10^{-6}$. This lack of signal from LD is not terribly surprising, as none of the three proxy SNPs ($r^2 > 0.2$) for this SNP in 1000 Genomes pass imputation quality control in our data. As the clusters for this SNP look excellent and the call rate is 99.98%, we believe this is a true signal.

Certain phenotypic associations are interesting given what is known about motion sickness. PONV is an established comorbidity of motion sickness [43] and is thought to stem from the anaesthetics that are administered for surgery. It may not, therefore, be surprising that motion sickness was also associated with vomiting and/or nausea with use of codeine, antidepressants and nonsteroidal anti-inflammatory drugs (NSAIDs). Additional gastrointestinal phenotypes (e.g., IBS, acid reflux, and indigestion with dairy products) as well as other drug-related phenotypes like being drowsy when taking Benadryl and feeling jittery when taking Sudafed were also associated with motion sickness. Interestingly, our findings suggest shared genetic susceptibility for both motion sickness and PONV (Table 4).

Some phenotypic associations might provide clues about the etiology of motion sickness (e.g., poor circulation and becoming light headed with exercise) or they might suggest simple remedies for motion sickness such as improving sleep quality. A number of associated phenotypes were related to personality (e.g., neuroticism) or behavior (e.g., smoking). We note, however, that it is difficult to assess causality for these phenotype-phenotype associations. For example, does being a sound sleeper make one less susceptible to being motion sick, or vice versa, or are both related to a third condition? The validity of these novel phenotypic findings is bolstered by the fact that we also detected associations with known symptoms (dizziness and headache) and established comorbidities (PONV, migraines, vertigo and morning sickness) of motion sickness. In some cases we even identified shared genetic factors for motion sickness and related comorbidities (e.g., PONV and migraines). Some of the correlated phenotypes are not established comorbidities or symptoms of motion sickness, however, and do not have an obvious biological relationship to motion sickness.

Our web-based method of capturing phenotypic information allows us to build a very large cohort (e.g., 80,494 individuals in our study), but we may not have obtained a complete picture of an individual’s susceptibility to motion sickness. For finding SNPs, the gain in power from having a large sample more than makes up for the reduction in power due to possible misclassification. An additional potential limitation is that we only surveyed individuals about car sickness; future studies should investigate these SNPs in populations phenotyped for other forms of motion sickness. An advantage of our web-based phenotypic collection method is that we can easily investigate whether seemingly related traits have shared underlying genetics. We identified four SNPs simultaneously associated with motion sickness plus PONV or migraines. These findings may provide clues into the etiology of all three conditions and may point to overlapping risk factors or treatments.

**Methods**

**Human Subjects**

All participants were drawn from the customer base of 23andMe, Inc., a consumer genetics company. This cohort has been described in detail previously [44,45]. Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services (E&I Review).

**Phenotype collection**

Participants invited to fill out web-based questionnaires, which included four questions about motion sickness during road travel, whenever they logged into their 23andMe accounts. The responses to each question were translated into a motion sickness score of 0 (never), 1 (occasionally), 2 (sometimes), or 3 (frequently). Responses of “I’m not sure” or “Don’t know” were excluded from the analysis. The questions were prioritized (1-4) and a participant’s final
score was based on the question they answered that had the highest priority. The questions were:

1. “How often do you experience motion sickness while in a car? (Never / Occasionally / Sometimes / Frequently / I’m not sure)”

2. “Have you experienced motion sickness while riding in a car (car sickness)? (Yes, I do now frequently / Yes, I did frequently, but only as a child / Yes, occasionally / No / Don’t know)”

3. “As a child, how often did you experience motion sickness while in a car? (Never / Occasionally / Sometimes / Frequently / I’m not sure)”

4. “Can you read in a moving car without becoming nauseated? (Never / Sometimes / Always / I’m not sure)”

Responses to the final question were scored as Never=3, Sometimes=1, and Always=0.

Genotyping and imputation

Participants were genotyped and additional SNP genotypes were imputed against the August 2010 release of the 1000 Genomes data as described previously [46]. Briefly, they were genotyped on at least one of three genotyping platforms, two based on the Illumina HumanHap550+ BeadChip, the third based on the Illumina Human OmniExpress+ BeadChip. The platforms included assays for 586,916, 584,942, and 1,008,948 SNPs, respectively. Genotypes for a total of 11,914,767 SNPs were imputed in batches of roughly 10,000 individuals, grouped by genotyping platform. Imputation was performed as in [46]. Prior to the imputation we discarded genotyped SNPs that were not present in the imputation panel. For the GWAS, we added such SNPs back (if they passed quality control), for a total of 7,428,049 SNPs (7,378,897 imputed and 49,152 genotyped). To filter SNPs whose imputation results had changed over time, we performed an ANOVA test for frequency differences across batches. The quality control criteria for imputed SNPs were batch effects $p$-value at least $10^{-50}$, average $r^2$ across batches of at least 0.5, and minimum $r^2$ across batches of at least 0.3. For genotyped SNPs, we required a MAF of at least 0.001, a Hardy-Weinburg $p$-value of at least $10^{-20}$, and a call rate at least 0.9.

Statistical analysis

In order to minimize population substructure while maximizing statistical power, the study was limited to individuals with European ancestry. Ancestry was inferred from the genome-wide genotype data and principal component analysis was performed as in [44,47]. The cohort was filtered by relatedness to remove participants at a first cousin or closer relationship. More precisely, no two participants shared more than 700 cM of DNA identical by descent (IBD; approximately the lower end of sharing between a pair of first cousins). IBD was calculated using the methods described in [48]. The genomic control inflation factor was 1.156; all p-values are adjusted for this inflation factor.

The GWAS was performed using likelihood ratio tests for the linear regression

$$C \sim G + A + S + \sum_{i=1}^{5} P_i$$

of carsickness on genotype, age, sex and 5 principal components of genetic ancestry. Genotypes were coded as a dosage from 0–2 (counting the estimated number of minor alleles present for imputed SNPs) or as a count 0, 1, or 2 (also number of minor alleles, for genotyped SNPs). Significant SNPs were grouped into regions with at least 500kb between pairs of significant SNPs; the SNP with the lowest $p$-value in each region was chosen to be the index SNP. As some of the regions were under 1Mb apart, a joint regression with all index SNPs was run to make sure that they all represented independent signals.

Partial correlation between carsickness $C$ and a phenotype $Y$ were computed by computing the correlation between the residuals produced by regressing both $C$ and $Y$ on age, sex, and five principal components, using linear regression even if $Y$ was a binary trait. We did not attempt to quantify the significance of these regressions nor any causality.
Tests of SNPs associated with motion sickness against other correlated traits were done using logistic or linear regressions as appropriate with the same covariates as in the GWAS (except for morning sickness, which dropped sex). The phenotypes studied (PONV, migraines, hay fever, altitude sickness, morning sickness, and vertigo) were all case control except for morning sickness, which was scored on a 5 point scale: None; Mild (occasional bouts of queasiness or nausea, did not require treatment); Moderate (nausea and some vomiting, but did not require treatment); Severe (Severe nausea and vomiting that required treatment); Very severe (requiring hospitalization and intravenous fluid (IV) therapy).

Enrichment analysis using GREAT was conducted on all regions with an index SNP with $p < 10^{-5}$, where regions were enforced to be 500kb apart. Windows of 500kb on either side of each index SNP were uploaded into GREAT using default settings.

Acknowledgements

We thank the customers of 23andMe who answered surveys and participated in this research. We also thank all the employees of 23andMe, who together have made this research possible.

References


Table S1: **Sex-specific effects for significant SNPs.** Estimated effect sizes for females, males, and their ratio. \( p \)-values are for a non-zero interaction term between sex and genotype. SNPs are sorted by increasing \( p \)-value.
Figure S1: Quantile-quantile plot. Observed versus expected $p$-values under the null.
(19) rs4343996

(20) rs7170668

(21) rs10752212

(22) rs7957589

(23) rs62018380

(24) rs6833641
Figure S2: **Region plots for genome-wide significant regions.** SNPs are colored by $r^2$ with the index SNP (which is labeled). Circles are imputed SNPs, plus signs are genotyped SNPs.