






ARTICLE

<https://doi.org/10.1038/s42003-019-0568-2>

OPEN

Genome-wide association study of knee pain identifies associations with *GDF5* and *COL27A1* in UK Biobank

Weihua Meng ^{1,9}, Mark J. Adams ^{2,9}, Colin N.A. Palmer¹, The 23andMe Research Team, Jingchunzi Shi³, Adam Auton³, Kathleen A. Ryan⁴, Joanne M. Jordan ⁵, Braxton D. Mitchell^{4,6}, Rebecca D. Jackson⁷, Michelle S. Yau⁸, Andrew M. McIntosh ² & Blair H. Smith ¹

Knee pain is one of the most common musculoskeletal complaints that brings people to medical attention. Approximately 50% of individuals over the age of 50 report an experience of knee pain within the past 12 months. We sought to identify the genetic variants associated with knee pain in 171,516 subjects from the UK Biobank cohort and seek supporting evidence in cohorts from 23andMe, the Osteoarthritis Initiative, and the Johnston County Osteoarthritis Project. We identified two loci that reached genome-wide significance in the UK Biobank: rs143384, located in *GDF5* ($P = 1.32 \times 10^{-12}$), a gene previously implicated in osteoarthritis; and rs2808772, located near *COL27A1* ($P = 1.49 \times 10^{-8}$). These findings were supported in cohorts with self-reported osteoarthritis/radiographic knee osteoarthritis without pain information. In this report on genome-wide association of knee pain, we identified two loci in or near *GDF5* and *COL27A1* that are associated with knee pain.

¹Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, UK. ²Division of Psychiatry, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK. ³23andMe, Inc., Mountain View, CA, USA. ⁴Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. ⁵Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, USA. ⁶Geriatric Research, Education and Clinical Center, Veterans Affairs Medical Center, Baltimore, MD, USA. ⁷Division of Endocrinology, Diabetes and Metabolism, The Ohio State University, Columbus, OH, USA. ⁸Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Boston, MA, USA. ⁹These authors contributed equally: Weihua Meng and Mark J. Adams. A full list of the 23andMe Research Team members appears at the end of the paper. Correspondence and requests for materials should be addressed to W.M. (email: w.meng@dundee.ac.uk)

The knee supports body weight when walking, standing upright and bending. Knee pain describes a specific area of pain inside the knee or diffuse pain around knee area¹. It is one of the most common musculoskeletal complaints that bring people to medical attention². The knee pain experience varies from person to person and can present as a dull ache to a sharp, stabbing pain and from intermittent weight bearing pain to persistent pain³.

Knee pain is highly prevalent in older individuals, with ~50% of individuals over the age of 50 reporting an experience of knee pain within the past 12 months⁴. In one US general population cohort, knee pain prevalence has increased from 15.7 to 32.9% in females and from 8.7 to 27.7% in males between 1983 and 2005, regardless of knee osteoarthritis status⁵. In another estimate, the prevalence of knee osteoarthritis in the USA increased from 8% in 1950s to 16% currently⁶. There are over eight million patients suffering from knee osteoarthritis in the UK⁷. According to the Global Burden of Diseases 2016, osteoarthritis including knee osteoarthritis is the twelfth leading cause of years of life lived with disability globally⁸. It is estimated that ~50% of all people with knee osteoarthritis have reported knee pain symptoms and of those without knee osteoarthritis, 20% have reported knee pain⁵. There are many underlying mechanisms that can cause knee pain, including injuries, gout and infection, as well as arthritis. Among these, osteoarthritis is the most common cause, particularly in people over the age of 50⁵. People with knee pain will experience progressive loss of knee function and declining quality of daily life, and display increasing dependence in daily activities⁹. Further, knee pain caused by osteoarthritis frequently accompanies pain in other joints, such as hips and hands, which further reduces quality of life¹⁰. The disease has generated huge economic burdens to the health care systems across the world. For example, although no figures exist specifically on knee osteoarthritis, the total direct cost of osteoarthritis as a whole in the UK in 2010 was around £1 billion and the corresponding total indirect cost of osteoarthritis in 2010 was over £3.2 billion¹¹.

Epidemiological studies have suggested multiple risk factors for knee pain, including female sex, age, obesity, previous knee injuries, knee-straining work and smoking¹². Similar risk factors are reported in studies of knee osteoarthritis specifically, which also included kneeling and squatting as further risk factors^{12,13}. With aging populations and increasing rates of obesity, the prevalence of knee pain is likely to increase. Psychological factors are also important risk factors of knee pain¹⁴. These environmental and lifestyle factors are likely to interact with genetic factors, and are important to understand in genetic association studies.

Genetic studies to date have focused on knee osteoarthritis, but not knee pain more generally. Studies in siblings have reported heritabilities for knee osteoarthritis as high as 0.62¹⁵. In a recent large twin study, 45% of the respective variation for severe knee osteoarthritis requiring joint replacement could be explained by genetic factors¹⁶. The genetic architecture of knee osteoarthritis was considered to follow an additive genetic model, involving multiple genes or loci but each with small effect size¹⁷. Candidate genes, including *GDF5*, *COL9A1*, *IL1B*, *IL1RN*, *LRCH1*, *CLIP*, *TNA* and *BMP2*, have been reported to be associated with knee osteoarthritis^{18–22}. Genome-wide association studies (GWAS) have also reported that the *GDF5*, *DVWA*, *HLA-DQB1*, *BTNL2*, *COG5*, *MCF2L*, *TP63*, *FTO*, *SUPT3H/RUNX2*, *GLN3/GLT8D1* and *LSP1P3* genes contribute specifically to knee osteoarthritis^{23–29}. Recently, Zengini et al. reported nine novel genetic loci associated with osteoarthritis based on five different osteoarthritis definitions according to the self-reported status questionnaire and the Hospital Episode Statistics data from the UK Biobank cohort³⁰. However, these analyses did not specifically focus on the knee area.

To identify the genetic variants associated with knee pain, we conducted a GWAS using the large UK Biobank cohort. We defined knee pain as ‘knee pain in the last month interfering with usual activity’, based on the information available from the study questionnaire. Since there are no knee pain GWAS summary statistics available, we chose to support our genome-wide findings on knee pain using three independent cohorts that defined osteoarthritis using either questionnaire data (i.e. 23andMe) or radiographic criteria (i.e. the Osteoarthritis Initiative (OAI) and the Johnston County Osteoarthritis Project (JoCo)).

Here, we use GWAS on knee pain to screen for genetic variants; similar approaches have been taken for headache and back pain using the UK Biobank cohort^{31,32}.

Results

GWAS results. A total of 501,708 UK Biobank participants were invited to respond to the pain questionnaire during the initial assessment visit (2006–2010). Among those who responded, 29,995 participants selected the ‘Knee pain’ option (cases), and 197,149 participants selected the ‘None of the above’ option (controls). After removing samples from non-British participants, those who were related with another individual in the cohort and those who failed quality control (QC), we identified 22,204 cases (12,062 males and 10,142 females) and 149,312 controls (71,480 males and 77,832 females) for the GWAS association analysis and there were 15,377,520 single nucleotide polymorphism (SNPs) available for the GWAS analysis. The genomic control value (λ) was 1.06.

Table 1 summarises the clinical characteristics of these cases and controls. There were statistical differences ($P < 0.001$) in sex, age and body mass index (BMI) between cases and controls in the UK Biobank samples.

We identified two SNP clusters that were associated with knee pain, with genome-wide significance ($P < 5 \times 10^{-8}$, Fig. 1, Table 2). Four independent significantly associated SNPs within two clusters are shown in Table 2. All significantly associated SNPs ($N = 107$) in the discovery stage are shown in Supplementary Data 1.

The most significantly associated SNP cluster was in the *GDF5* gene in the chromosome 20q11.22 region with a P value of 1.32×10^{-12} for rs143384 (A allele, beta: -0.008). The second most significantly associated cluster was in the *LOC105376225* gene (near the *COL27A1* gene) in chromosome 9 with a lowest P value of 1.49×10^{-8} for rs2808772 (A allele, beta: 0.006). The regional plots for loci in *GDF5* and *COL27A1* are shown in Figs. 2 and 3.

The Q–Q plot of the GWAS in the discovery stage is shown in Fig. 4. The SNP-based heritability of knee pain was 0.08 (standard error = 0.03).

In the independent cohort 1 (23andMe, self-reported osteoarthritis), the P value for rs143384 was 2.44×10^{-9} in the 23andMe

Table 1 Clinical characteristics of knee pain and controls in the UK Biobank

Covariates	UK Biobank		P
	Cases	Controls	
Sex (male: female)	12,062;10,142	71,480;77,832	<0.001
Age (years)	58.3 (7.64)	56.9 (7.97)	<0.001
BMI (kg m ⁻²)	28.6 (4.88)	26.7 (5.00)	<0.001

A χ^2 test was used to test the difference of gender frequency between cases and controls and an independent t test was used for other covariates. Continuous covariates were presented as mean (standard deviation). BMI body mass index

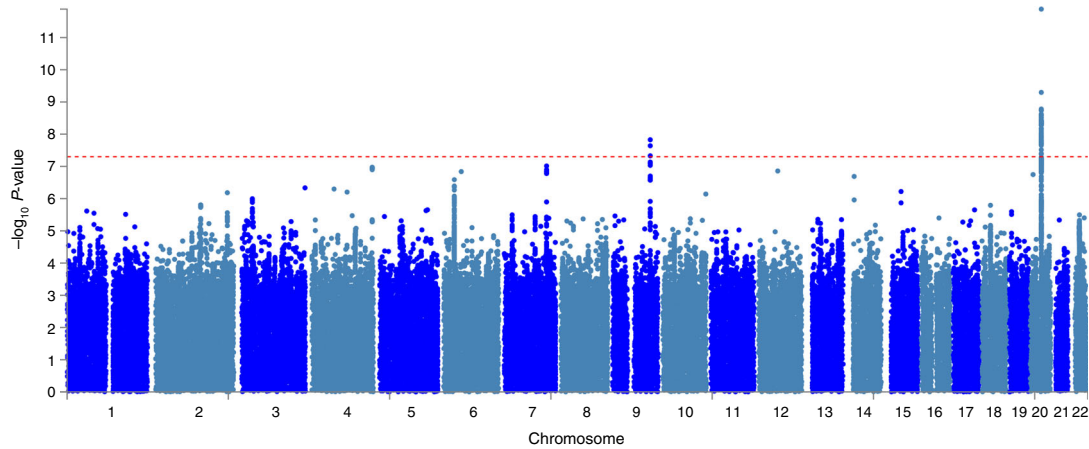


Fig. 1 Manhattan plot of the GWAS on knee pain using the UK Biobank cohort

Table 2 Summary of the four independent and significant SNPs associated with knee pain in the *GDF5* and *COL27A1* regions

Chromosome	Gene	SNPID	Effective allele in all cohorts	Effective allele frequency discovery cohort	P value (beta) UK Biobank discovery cohort	P value (beta) 23andMe independent cohort 1	P value (beta) OAI-JoCo independent cohorts 2
9	<i>LOC105376225</i> (near <i>COL27A1</i>)	rs919642	A	73.2%	2.29×10^{-8} (0.007)	4.84×10^{-12} (0.028)	0.88 (-0.008)
9	<i>LOC105376225</i> (near <i>COL27A1</i>)	rs2808772	A	52.5%	1.49×10^{-8} (0.006)	4.43×10^{-5} (0.014)	0.36 (0.041)
20	<i>GDF5</i>	rs143384	A	60.0%	1.32×10^{-12} (-0.008)	2.44×10^{-9} (-0.021)	0.01 (-0.12)
20	<i>GDF5</i>	rs6120946	A	78.2%	6.81×10^{-9} (-0.008)	0.00071 (-0.014)	0.0074 (-0.16)

OAI-JoCo, the joint Osteoarthritis Initiative-Johnston County Osteoarthritis cohorts

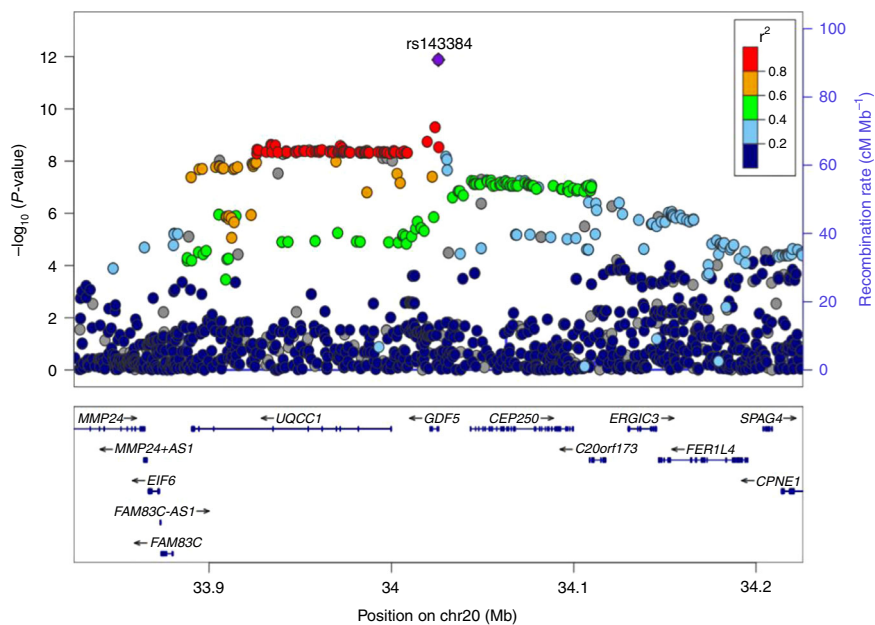


Fig. 2 Regional plot of the *GDF5* gene region based on the GWAS on knee pain using the UK Biobank cohort

cohort and in the independent cohorts 2 (OAI and JoCo cohorts, radiographic knee osteoarthritis), the *P* value for rs143384 was 0.01 in the combined OAI and JoCo cohorts. The *P* values for rs2808772 were 4.43×10^{-5} in the 23andMe (self-reported osteoarthritis) and 0.36 in the combined OAI and JoCo cohorts (radiographic knee osteoarthritis) (Table 2).

Gene, gene-set and tissue expression analysis by FUMA. In the gene analysis, all the SNPs that are located within genes were mapped to 19,436 protein coding genes. *GDF5* demonstrated the strongest association, with a *P* value of 1.09×10^{-11} . The eleven associated genes with *P* values $< 3 \times 10^{-6}$ (0.05/19436) were *GDF5*, *UQC1*, *CEP250*, *PODXL*, *C20orf173*, *SPAG4*, *MTMR3*,

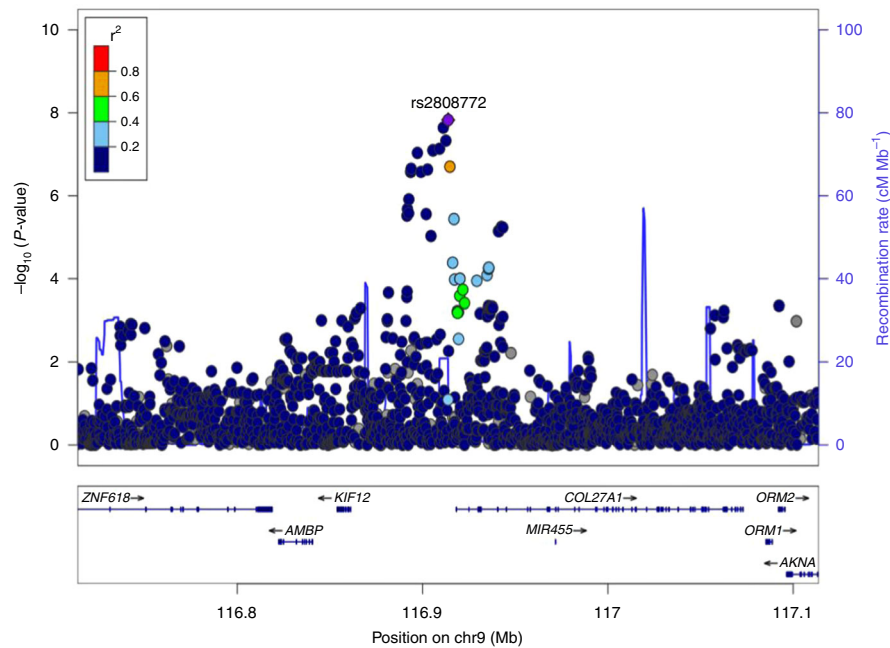


Fig. 3 Regional plot of the *COL27A1* gene region based on the GWAS on knee pain using the UK Biobank cohort

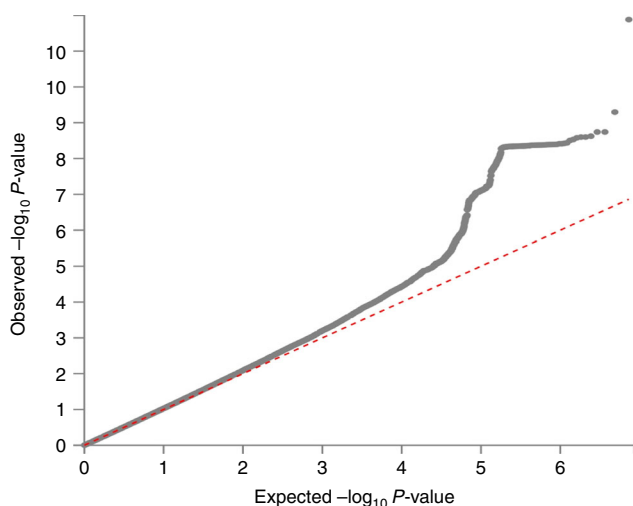


Fig. 4 Q-Q plot of the GWAS on knee pain using the UK Biobank

ERGIC3, *FBLN2*, *CPNE1* and *CDC42SE2*. The results are included in Supplementary Table 1.

In the gene-set analysis, a total of 10,894 gene sets were tested. The regulation pathway of *breast_cancer_20q11_amplicon* demonstrated a P value of 2.59×10^{-8} and this was the only gene set with a statistically significant association ($P < 5 \times 10^{-6}$ (0.05/10,894)). The top ten gene sets from this analysis are shown in Supplementary Table 2.

In the tissue expression analysis, none of the tissue types demonstrated statistically significant associations ($P < 0.001$), either in the expression analysis of 30 general tissue types from multiple organs or in the 53 specific tissue types within some of these organs. See Supplementary Figs. 1 and 2.

Genetic correlation analysis by LD Hub. We identified multiple significant and negative genetic correlations for knee pain with other traits (Supplementary Data 2). The genetic correlations (r_g) surviving multiple testing correction were: Years of schooling

2016 ($r_g = -0.29$, $P = 4.97 \times 10^{-8}$), college completion ($r_g = -0.36$, $P = 6.55 \times 10^{-6}$), age of having first baby ($r_g = -0.30$, $P = 1.92 \times 10^{-5}$).

Discussion

In the first reported GWAS of knee pain using the UK Biobank resource, we identified variants in or near *GDF5* and *COL27A1*, which were subsequently supported in osteoarthritis cohorts from the 23andMe, OAI and JoCo cohorts. In addition, we found that knee pain was genetically and negatively correlated with a number of socioeconomic factors, such as years of schooling and college completion.

The generic pain question used by the UK Biobank is useful as a screening tool and a useful step to test whether heterogeneous pain phenotypes (such as knee pain) have genetic components at all. The same question has been used to identify the genetic variants of broadly defined headache, and the findings were similar to those for well-defined migraine phenotypes^{31,33}. The benefit of using UK Biobank on heterogeneous phenotypes will allow researchers to overcome potential issues with reduced power due to heterogeneity by using very large numbers to cut through the statistical noise.

In this GWAS, we have identified two loci for knee pain. The top locus was in the *GDF5* gene in chromosome 20q11.2 with a lowest P value of 1.32×10^{-12} for rs143384, while the locus itself was 140 kb long spanning from the *UQCC1* gene to the *GDF5* gene containing 104 genome-wide positive SNPs (Supplementary Data 1). The *GDF5* gene encodes a secreted ligand of the transforming growth factor-beta (TGF-beta) superfamily of proteins³⁴. This protein not only regulates the development of numerous tissue and cell types, but also promotes the maintenance and repair of synovial joint tissues, particularly cartilage and bones^{34,35}. Mutations in the gene can cause cartilage or bone related disorders, such as chondrodysplasia, acromesomelic dysplasia and brachydactyly, suggesting a protective role in skeletal development³⁴. The *GDF5* gene has been repeatedly reported to be associated with osteoarthritis through genetic studies^{18,28}. Functional studies have suggested that knee morphology is profoundly affected by *Gdf5* absence in mice models, and

downstream regulatory sequences mediate its effects by controlling *Gdf5* expression in knee tissues³⁶. It was also suggested that osteoarthritis susceptibility mediated by variants in the *GDF5* gene was not restricted to cartilage but joint wide³⁷. Recently, Capellini et al. combined the transgenic mice model with population genetic analyses in humans to identify a *GDF5* enhancer that influences human growth and osteoarthritis risk³⁸. Overall, there is sufficient and solid biological evidence relating the *GDF5* gene with knee osteoarthritis, and we assume that this finding is due to detection of knee pain caused by osteoarthritis, rather than other pathologies.

The second SNP cluster was in the *LOC105376225* area (which is next to the *COL27A1* gene) in chromosome 9 with a lowest *P* value of 1.49×10^{-8} for rs2808772. There have been no specific studies published about *LOC105376225* and its relationship with knee pain or knee osteoarthritis. However, the neighbouring *COL27A1* gene is clearly a good candidate gene. This gene encodes a member of the fibrillar collagen family, and plays a role during the calcification of cartilage and the transition of cartilage to bone³⁹. Mutations in the *COL27A1* gene have been reported to be associated with the Steel syndrome. This syndrome is characterised by bone changes, such as bilateral hip and radial head dislocations, short stature, characteristic facies, fusion of carpal bones, scoliosis, *pes cavus* and cervical spine anomalies⁴⁰. Further, the gene was reported to be associated with knee osteoarthritis in the first stage, but did not replicate in the second stage in a recent GWAS study on knee osteoarthritis²⁹. Thus, our large study on knee pain has suggested that the *COL27A1* gene might play a contributing role for lesions in the knee area. Importantly, polymorphisms in the gene have been associated with tendinopathy around the ankle joint⁴¹. The concordance between radiographically defined knee osteoarthritis and knee pain is quite poor, with between 15 and 81% of patients diagnosed by radiographic methods having pain symptoms⁴². It is therefore likely that many people reporting knee pain have pain that is not bone or cartilage related, but tendon related. It is possible that variants in the *COL27A1* gene could eventually affect the quality of the tendons such as strength or flexibility, which might lead to poor tendon function, or their stability of the attachment to the patella, which might also be knee pain related. A recent study has suggested that a collagen gene *COL11A2* play a role in pain sensitisation after the development of osteoarthritis⁴³. Interestingly, the *COL27A1* gene was not mentioned in GWAS on osteoarthritis using the UK Biobank dataset⁴⁴.

Our study focused on knee pain as a broad phenotype and the genes that we identified are suggested to be related to knee osteoarthritis. This suggests that the phenotype we chose was genetically similar to the phenotype of knee osteoarthritis. The relationship between knee pain and knee osteoarthritis deserves further investigation. Studies have shown that people with end-stage knee osteoarthritis all presented with knee pain⁴⁵, but this might not be the case for early stage knee osteoarthritis. As described above, 20% of knee pain was not caused by knee osteoarthritis and only 50% of knee osteoarthritis patients with radiographic evidence had knee pain symptoms⁵. In addition, a study has reported that 86% of people reporting knee pain will develop knee osteoarthritis over 12 years⁴⁶. Interestingly, the hospital-diagnosed ($N = 10,083$ cases) osteoarthritis and self-reported ($N = 12,658$ cases) osteoarthritis in the UK Biobank were only half the size of those reported with knee pain in this study ($N = 22,204$)³⁰. This indicates the importance of treat knee pain as a single entitle (like back pain) instead of solely being a symptom of knee disorders. The Neale lab (<http://www.nealelab.is/uk-biobank>) has performed over 2000 traits using the UK Biobank dataset treating knee pain and knee osteoarthritis as individual phenotypes.

The gene analysis by FUMA also supports our finding that *GDF5* was the strongest gene for knee pain. The gene-set analysis by FUMA revealed that the regulation pathway of Nikolsky_breast_cancer_20q11_amplicon signalling was associated with the phenotype we use. We noticed that *GDF5* and this amplicon were both located in chromosome 20q11 area and it was reported that *GDF5* protein regulates TGF-beta dependent angiogenesis in breast carcinoma MCF-7 Cells⁴⁷.

The SNP-based heritability for knee pain was 0.08 in our study, which is the first report of its kind. This low heritability suggests the greater role of environmental than genetic factors in development of knee pain, though this method of estimating heritability is likely to miss some of the genetic contribution. We identified that knee pain was genetically and negatively correlated with a number of phenotypes such as years of schooling, college completion and age of having first baby. This means that those with more years of schooling, those with completed college education, and those who were older when they had their first baby were less likely to report current troublesome knee pain. These factors could be related to lifestyle and occupation.

Using the CaTS power calculator (<http://csg.sph.umich.edu/abecasis/cats/>), we had 80% power to identify SNP associations with a significance level of 5×10^{-8} , based on 22,204 cases and 149,312 controls, assuming an additive model, a minor disease allele frequency of 0.20, a genotypic relative risk of 1.06 and an estimated prevalence of knee pain in the general population of 0.2.

The major limitation of this study was that we used different (though similar) phenotypes in discovery and supporting stages. This was because of the phenotypic information that was available in the relevant and available cohorts, and the lack of any existing relevant cohorts or datasets examining knee pain as a phenotype. We defined knee pain cases and controls based on the responses by UK Biobank participants to a specific pain question. This question focused on knee pain occurrence, sufficient to cause interference with activities, during the previous month. The question does not ask information of the severity and frequency and the exact area of knee pain. Therefore, our phenotyping should be considered as widely defined. The situation was similar for the 23andMe cohort, in which disease status was also self-reported via survey and self-reported osteoarthritis in the 23andMe cohort was not specific to the knee. Self-reported knee pain has been widely used in other studies as well, though not for studies of genetic associations^{48,49}. The OAI and JoCo assessed radiographic evidence of knee osteoarthritis but with limited sample size. Although not true replications, these independent cohorts suggest possible overlapping risk alleles among knee pain, general osteoarthritis and knee osteoarthritis.

In conclusion, we have identified two loci (*GDF5* and *COL27A1*) for knee pain in a GWAS using the UK Biobank resource and found evidence to support them in the 23andMe, OAI and JoCo cohorts. In addition, we found several significant and negative genetic correlations between knee pain and a number of educational phenotypes, suggesting that the genetic aetiology of knee pain may also be related to these traits.

Methods

The participants and genetic information of all cohorts. *Discovery cohort—UK Biobank:* Over 500,000 people aged between 40 and 69 years were recruited by the UK Biobank cohort in 2006–2010 across England, Scotland and Wales. All participants provided informed consent that their health records could be accessed for research purposes. Further information about the UK Biobank cohort can be found at www.ukbiobank.ac.uk. Ethical approval was granted by the National Health Service National Research Ethics Service (reference 11/NW/0382).

DNA extraction and QC were standardised and the detailed methods can be found at <http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/DNA-Extraction-at-UK-Biobank-October-2014.pdf>. The Wellcome Trust Centre for

Human Genetics at Oxford University was in charge of standard QC procedures for genotyping results. The detailed QC steps can be found at <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580>.

In July 2017, the UK Biobank released the genetic information (including directly genotyped genotypes and imputed genotypes) of 501,708 samples to all approved researchers. The detailed QC steps of imputation were described by Bycroft et al.⁵⁰

Independent cohort 1: 23andMe Inc: The 23andMe company is a privately held personal genomics and biotechnology company based in the USA. It includes more than 1,500,000 genotyped subjects who have consented to participate in research. All participants included in the analyses provided informed consent and answered surveys online according to 23andMe's human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board. The DNA extraction from saliva and the QC of the genotyping and imputation results were all based on the company's standardised procedures. Further methodological details can be found in the supplementary file of a previous publication⁵¹.

Independent cohorts 2: (1) The OAI is a prospective longitudinal study designed to identify risk factors for the incidence and progression of symptomatic tibiofemoral knee osteoarthritis. Participants aged between 45 and 79 years were recruited at four different clinical sites in the USA. Details of the study protocol, including recruitment procedures and eligibility criteria are available on the OAI web site. (<http://oai.epi-ucsf.org/dataset/docs/StudyDesignProtocol.pdf>). (2) The JoCo is an ongoing, community-based study of the occurrence of knee and hip osteoarthritis in African American and Caucasian residents, aged 45 years and above from Johnston County, North Carolina in the USA. A total of 3068 individuals were recruited at baseline. A detailed description of the cohort has been reported⁵². All participants provided informed consent.

Standard procedures of imputation and QC were applied when genotyping OAI and JoCo samples. The detailed description of the cohorts has been previously published²⁹.

Phenotypic information of all cohorts in this study. *Discovery cohort—UK Biobank:* We used a bespoke pain-related questionnaire adapted by the UK Biobank, which included the question: 'in the last month have you experienced any of the following that interfered with your usual activities?' The options were: (1) Headache; (2) Facial pain; (3) Neck or shoulder pain; (4) Back pain; (5) Stomach or abdominal pain; (6) Hip pain; (7) Knee pain; (8) Pain all over the body; (9) None of the above; (10) Prefer not to say. More than one option could be selected. (UK Biobank Questionnaire field ID: 6159)

The knee pain cases in this study were those who selected the 'knee pain' option for the above question, regardless of whether they had selected other options.

The controls in this study were those who selected the 'None of the above' option.

Independent cohort 1—23andMe, Inc: The 23andMe cohort used an online survey to determine the phenotypic status of osteoarthritis of all participants.

Cases were defined as those self-reported having been diagnosed or treated for osteoarthritis.

Controls were defined as those self-reporting as having not been diagnosed or treated for osteoarthritis.

The cohort included 253,880 cases and 1,286,245 controls.

Independent cohorts 2—OAI and JoCo:

Cases. Knee osteoarthritis was evaluated with fixed-flexion posteroanterior radiographs for OAI samples and for JoCo participants, weight-bearing anteroposterior extended radiographs were taken during initial recruitments and fixed-flexion posteroanterior radiographs were taken during follow-up. Cases were those with definitive knee osteoarthritis, defined as radiographic evidence of the presence of definite osteophytes and possible joint space narrowing (Kellgren-Lawrence grade ≥ 2) or total joint replacement in one or both knees. Controls were those having no or doubtful evidence of OA (Kellgren-Lawrence grade = 0 or 1) in both knees at all available time points.

These definitions were previously used by Yau et al.²⁹. In the current study, there were 2672 cases (2014 from OAI and 658 from JoCo) and 1776 controls (953 from OAI and 823 from JoCo).

Statistical analysis. *GWAS analysis:* In the discovery stage, the BGENIE (<https://jmarchini.org/bgenie/>) was used as the main GWAS software, which was designed for analysing the UK Biobank genetic datasets. Routine QC steps included removal of SNPs with INFO scores < 0.1 , SNPs with minor allele frequency $< 0.5\%$ or SNPs that failed Hardy-Weinberg tests ($P < 10^{-6}$). SNPs on the X and Y chromosomes and mitochondrial SNPs were also removed. We further removed data from individuals whose ancestry was not white British based on the principal component analysis, those who were related to at least one other participant in the cohort (a cutoff value of 0.025 in the generation of the genetic relationship matrix), and those who failed QC. Association tests based on the linear association were performed using BGENIE adjusting for age, sex, BMI, nine population principal components, genotyping arrays and assessment centres. A χ^2 test was used to test for gender differences between cases and controls. Age and BMI were compared using independent t-test in IBM SPSS 22 (IBM Corporation, New York). A P value $< 5 \times 10^{-8}$ was considered to indicate a genome-wide association. Independent SNPs were

defined as those that were not correlated ($r^2 < 0.6$) with any other associated SNP. GCTA (<https://cns.genomics.com/software/gcta/#Overview>) was used to calculate the narrow-sense heritability using a genomic relationship matrix calculated from genotyped autosomal SNPs.

In the supporting stage (using independent cohorts), details of the identified genome-wide positive and independent SNPs associated with knee pain from the discovery stage were sent to 23andMe Inc and the combined OAI and JoCo cohorts. The genome-wide positive and independent SNPs were defined as those with P value $< 5 \times 10^{-8}$ and with linkage disequilibrium value $r^2 < 0.6$. The 23andMe and the combined OAI and JoCo cohorts then extracted the summary statistics of these SNPs from their GWAS results, correspondingly.

23andMe performed GWAS on self-reported osteoarthritis in any joint using the logistic regression method assuming an additive genetic model for allelic effects adjusting for age, sex, five principal components and four DNA chip platforms. Participants were restricted to a set of individuals who had $> 97\%$ European ancestry, as determined through an analysis of local ancestry. A maximal set of unrelated individuals was chosen for the GWAS analysis using a segmental identity-by-descent estimation algorithm. Further details can be found in the supplementary file of a previous publication⁵¹.

The OAI and JoCo performed GWAS on radiographic knee osteoarthritis using logistic regression assuming an additive genetic model for allelic effects adjusting for age, sex, study site and principal components. Summary statistics from both cohorts were then combined in a meta-analysis. Only participants of Caucasian origin were included in the GWAS study. Standard procedures were used to remove data from non-Caucasian individuals and related individuals. Further details can be found in ref. ²⁹.

GWAS-associated analysis: The FUMA web application was used as the main annotation tool, and a Manhattan plot and a Q-Q plot were also generated by this⁵³. LocusZoom (<http://locuszoom.org/>) was used to provide regional visualisation.

FUMA mainly provides three types of analysis: the gene analysis, the gene-set analysis and the tissue expression analysis. In gene analysis, summary statistics of SNPs were aggregated to the level of whole genes to test the associations between genes with the phenotype. In gene-set analysis, groups of genes sharing certain biological, functional or other characteristics were tested together to provide insight into the involvement of specific biological pathways or cellular functions in the genetic aetiology of a phenotype. The tissue expression analysis was based on GTEx (<https://www.gtexportal.org/home/>), which is integrated into FUMA. Average gene expression per tissue type was used as gene covariate to test for relationships between gene expression in a specific tissue type and genetic associations with knee pain.

To identify genetic correlations between knee pain and all other 234 complex traits, we used linkage disequilibrium score regression through LD Hub v1.9.0 (available at <http://ldsc.broadinstitute.org/ldhub/>)⁵⁴. The LD Hub estimates the bivariate genetic correlations of a phenotype with 234 traits using individual SNP allele effect sizes and the average linkage disequilibrium in a region. Those with P values less than 2.1×10^{-4} (0.05/234) were considered significant surviving Bonferroni correction for multiple testing.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The summary statistics of the UK Biobank results on knee pain can be accessed through <https://figshare.com/articles/kneepain/gwas/9611198>. Data from 23andMe were obtained under a data transfer agreement. Further information about obtaining access to the 23andMe Inc. summary statistics is available from: <https://research.23andme.com/collaborate/>. Any other data relevant to the study that are not included in the article or its supplementary materials are available from the authors upon reasonable request.

Received: 15 April 2019 Accepted: 2 August 2019

Published online: 28 August 2019

References

- Thompson, L. R. et al. The knee pain map: reliability of a method to identify knee pain location and pattern. *Arthritis Rheumatol.* **61**, 725–731 (2009).
- Neogi, T. The epidemiology and impact of pain in osteoarthritis. *Osteoarthr. Cartil.* **21**, 1145–1153 (2013).
- Hawker, G. A. et al. Understanding the pain experience in hip and knee osteoarthritis—an OARSI/OMERACT initiative. *Osteoarthr. Cartil.* **16**, 415–422 (2018).
- Jinks, C., Jordan, K., Ong, B. N. & Croft, P. A brief screening tool for knee pain in primary care (KNEST). 2. Results from a survey in the general population aged 50 and over. *Rheumatology* **43**, 55–61 (2004).
- Nguyen, U. S. et al. Increasing prevalence of knee pain and symptomatic knee osteoarthritis: survey and cohort data. *Ann. Intern. Med.* **155**, 725–732 (2011).

6. Wallace, I. J. et al. Knee osteoarthritis has doubled in prevalence since the mid-20th century. *Proc. Natl Acad. Sci. USA* **114**, 9332–9336 (2017).
7. NICE. *Clinical guideline CG177. Osteoarthritis: care and management in adults*. <https://www.nice.org.uk/guidance/cg177> (2014.)
8. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1211–1259 (2017).
9. Bindawas, S. M., Vennu, V. & Al Snih, S. Differences in health-related quality of life among subjects with frequent bilateral or unilateral knee pain: data from the Osteoarthritis Initiative study. *J. Orthop. Sports Phys. Ther.* **45**, 128–136 (2015).
10. Prieto-Alhambra, D. et al. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. *Ann. Rheum. Dis.* **73**, 1659–1664 (2014).
11. Chen, A., Gupte, C., Akhtar, K., Smith, P. & Cobb, J. The global economic cost of osteoarthritis: how the UK compares. *Arthritis* **2012**, 698709 (2012).
12. Miranda, H., Viikari-Juntura, E., Martikainen, R. & Riihimäki, H. A prospective study on knee pain and its risk factors. *Osteoarthr. Cartil.* **10**, 623–630 (2002).
13. Heidari, B. Knee osteoarthritis prevalence, risk factors, pathogenesis and features: part I. *Casp. J. Intern. Med.* **2**, 205–212 (2011).
14. Iijima, H. et al. Psychological health is associated with knee pain and physical function in patients with knee osteoarthritis: an exploratory cross-sectional study. *BMC Psychol.* **6**, 19 (2018).
15. Neame, R. L., Muir, K., Doherty, S. & Doherty, M. Genetic risk of knee osteoarthritis: a sibling study. *Ann. Rheum. Dis.* **63**, 1022–1027 (2004).
16. Magnusson, K. et al. Genetic factors contribute more to hip than knee surgery due to osteoarthritis - a population-based twin registry study of joint arthroplasty. *Osteoarthr. Cartil.* **25**, 878–884 (2017).
17. Ikegawa, S. New gene associations in osteoarthritis: what do they provide, and where are we going? *Curr. Opin. Rheumatol.* **19**, 429–434 (2007).
18. Miyamoto, Y. et al. A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat. Genet.* **39**, 529–533 (2007).
19. Loughlin, J. et al. Finer linkage mapping of a primary hip osteoarthritis susceptibility locus on chromosome 6. *Eur. J. Hum. Genet.* **10**, 562–568 (2002).
20. Meulenbelt, I. et al. Association of the interleukin-1 gene cluster with radiographic signs of osteoarthritis of the hip. *Arthritis Rheumatol.* **50**, 1179–1186 (2004).
21. Spector, T. D. et al. Association between a variation in LRCH1 and knee osteoarthritis. *Arthritis Rheumatol.* **54**, 524–532 (2006).
22. Valdes, A. M. et al. Reproducible genetic associations between candidate genes and clinical knee osteoarthritis in men and women. *Arthritis Rheumatol.* **54**, 533–539 (2006).
23. Nakajima, M. et al. New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS. ONE* **5**, e9723 (2010).
24. Miyamoto, Y. et al. Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat. Genet.* **40**, 994–998 (2008).
25. Evangelou, E. et al. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann. Rheum. Dis.* **70**, 349–355 (2011).
26. Day-Williams, A. G. et al. A variant in MCF2L is associated with osteoarthritis. *Am. J. Hum. Genet.* **89**, 446–450 (2011).
27. arcOGEN Consortium. et al. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815–823 (2012).
28. Valdes, A. M. et al. The GDF5rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Ann. Rheum. Dis.* **70**, 873–875 (2011).
29. Yau, M. S. et al. Genome-wide association study of radiographic knee osteoarthritis in North American Caucasians. *Arthritis Rheumatol.* **69**, 343–351 (2017).
30. Zengini, E. et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nat. Genet.* **50**, 549–558 (2018).
31. Meng, W. et al. A genome-wide association study finds genetic associations with broadly-defined headache in UK biobank ($N = 223,773$). *EBioMedicine* **28**, 180–186 (2018).
32. Suri, P. et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. *PLoS Genet.* **14**, e1007601 (2018).
33. Gormley, P. et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48**, 856–866 (2016).
34. *GDF5 growth differentiation factor 5 [Homo sapiens (human)]* (NCBI, 2019); <https://www.ncbi.nlm.nih.gov/gene/8200>.
35. Mikic, B. Multiple effects of GDF-5 deficiency on skeletal tissues: implications for therapeutic bioengineering. *Ann. Biomed. Eng.* **32**, 466–476 (2004).
36. Pregizer, S. K. et al. Impact of broad regulatory regions on Gdf5 expression and function in knee development and susceptibility to osteoarthritis. *Ann. Rheum. Dis.* **77**, 450 (2018).
37. Eglı, R. J. et al. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheumatol.* **60**, 2055–2064 (2009).
38. Capellini, T. D. et al. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. *Nat. Genet.* **49**, 1202–1210 (2017).
39. *COL27A1 collagen type XXVII alpha 1 chain [Homo sapiens (human)]* (NCBI, 2019); <https://www.ncbi.nlm.nih.gov/gene/85301>.
40. Gonzaga-Jauregui, C. et al. Mutations in COL27A1 cause steel syndrome and suggest a founder mutation effect in the Puerto Rican population. *Eur. J. Hum. Genet.* **23**, 342–346 (2015).
41. Saunders, C. J. et al. Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations. *J. Orthop. Res.* **31**, 632–637 (2013).
42. Bedson, J. & Croft, P. R. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. *BMC Musculoskelet. Disord.* **9**, 116 (2008).
43. Ho, K. W. D. et al. Single nucleotide polymorphism in the COL11A2 gene associated with heat pain sensitivity in knee osteoarthritis. *Mol. Pain* **13**, 1744806917724259 (2017).
44. Tachmazidou, I. et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet.* **51**, 230–236 (2019).
45. Zeni, J. A. Jr., Axe, M. J. & Snyder-Mackler, L. Clinical predictors of elective total joint replacement in persons with end-stage knee osteoarthritis. *BMC Musculoskelet. Disord.* **11**, 86 (2010).
46. Thorstensson, C. A., Andersson, M. L., Jönsson, H., Saxne, T. & Petersson, I. F. Natural course of knee osteoarthritis in middle-aged subjects with knee pain: 12-year follow-up using clinical and radiographic criteria. *Ann. Rheum. Dis.* **68**, 1890–1893 (2009).
47. Margheri, F. et al. GDF5 regulates TGF β -dependent angiogenesis in breast carcinoma MCF-7 cells: in vitro and in vivo control by anti-TGF β peptides. *PLoS. ONE* **7**, e50342 (2012).
48. Baldwin, J. N. et al. Self-reported knee pain and disability among healthy individuals: reference data and factors associated with the Knee injury and Osteoarthritis Outcome Score (KOOS) and KOOS-Child. *Osteoarthr. Cartil.* **25**, 1282–1290 (2017).
49. Ho-Pham, L. T. et al. Prevalence of radiographic osteoarthritis of the knee and its relationship to self-reported pain. *PLoS. ONE* **9**, e94563 (2014).
50. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
51. Warrington, N. M. et al. Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Hum. Mol. Genet.* **27**, 2025–2038 (2018).
52. Jordan, J. M. et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. *J. Rheumatol.* **34**, 172–180 (2007).
53. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
54. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

Acknowledgements

We would like to thank all participants of the UK Biobank, 23andMe and the OAI and JoCo cohorts who have provided necessary genetic and phenotypic information. The current study was conducted under approved UK Biobank data application number 4844. This work was supported by the STRADL project (Wellcome Trust, grant number: 104036/Z/14/Z), the DOLORisk project (Horizon 2020, grant number 633491) and the GCRF academic exchange visits programme by the University of Dundee. The Osteoarthritis Initiative (OAI) was public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the NIH. The Johnston County Osteoarthritis Study (JoCo) was supported in part by S043, S1734, & S3486 from the CDC/Association of Schools of Public Health; 5-P60-AR30701 & 5-P60-AR49465-03 from NIAMS/NIH; genotyping was supported by Algnomics Inc. Additional support was obtained from NIH grant P30-DK072488.

Author contributions

W.M. organised project, drafted the paper and contributed to the analysis. M.A. performed the main UK Biobank GWAS analysis. C.P., B.M. and M.Y. provided essential comments. K.R., J.J., B.M., R.J. and M.Y. provided the OAI and JoCo GWAS summary statistics on knee osteoarthritis. J.S. and A.A. provided results in the 23andMe cohort. A.M. and B.S. organised the project and provided comments.

Additional information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s42003-019-0568-2>.

Competing interests: J.S., A.A. and members of the 23andMe Research Team are employees of 23andMe, Inc., and hold stock or stock options in 23andMe. Other authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

The 23andMe Research Team

Michelle Agee³, Babak Alipanahi³, Robert K. Bell³, Katarzyna Bryc³, Sarah L. Elson³, Pierre Fontanillas³, Nicholas A. Furlotte³, Barry Hicks³, David A. Hinds³, Karen E. Huber³, Ethan M. Jewett³, Yunxuan Jiang³, Aaron Kleinman³, Keng-Han Lin³, Nadia K. Litterman³, Jennifer C. McCreight³, Matthew H. McIntyre³, Kimberly F. McManus³, Joanna L. Mountain³, Elizabeth S. Noblin³, Carrie A. M Northover³, Steven J. Pitts³, G. David Poznik³, J. Fah Sathirapongsasuti³, Janie F. Shelton³, Suyash Shringarpure³, Chao Tian³, Joyce Y. Tung³, Vladimir Vacic³, Xin Wang³ & Catherine H. Wilson³