OXFORD

Human Molecular Genetics, 2022, Vol. 31, 13, 2279–2293
https://doi.org/10.1093/hmg/ddac003
Advance access publication date: 12 January 2022
Association Studies Article

Ancestry- and sex-specific effects underlying inguinal hernia susceptibility identified in a multiethnic genome-wide association study meta-analysis

Hélène Choquet[1,*], Weiyu Li[2], Jie Yin[1], Rachael Bradley[2], Thomas J. Hoffmann[3,4], Priyanka Nandakumar[5], 23 and Me Research Team[5], Rouzbeh Mostaedi[6], Chao Tian[5], Nadav Ahituv[2,3,†] and Eric Jorgenson[7,†]

[1]Kaiser Permanente Northern California (KPNC), Division of Research, Oakland, CA 94612, USA
[2]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco (UCSF), San Francisco, CA 94158, USA
[3]Institute for Human Genetics, UCSF, San Francisco, CA 94143, USA
[4]Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA 94158, USA
[5]23andMe Inc, Sunnyvale, CA 94086, USA
[6]KPNC, Richmond Medical Center, Richmond, CA 94801, USA
[7]Regeneron Genetics Center, Tarrytown, NY 10591, USA

*To whom correspondence should be addressed at: KPNC, Division of Research, 2000 Broadway, Oakland, CA 94612, USA. Tel: +1 5108915972; Fax: +1 5108913508; Email: Helene.Choquet@kp.org

†These authors jointly supervised this work: Nadav Ahituv, Eric Jorgenson.

A list of 23andMe research members and their affiliations appears in the Supplementary Material.

Abstract

Inguinal hernias are some of the most frequently diagnosed conditions in clinical practice and inguinal hernia repair is the most common procedure performed by general surgeons. Studies of inguinal hernias in non-European populations are lacking, though it is expected that such studies could identify novel loci. Further, the cumulative lifetime incidence of inguinal hernia is nine times greater in men than women, however, it is not clear why this difference exists. We conducted a genome-wide association meta-analysis of inguinal hernia risk across 513 120 individuals (35 774 cases and 477 346 controls) of Hispanic/Latino, African, Asian and European descent, with replication in 728 418 participants (33 491 cases and 694 927 controls) from the 23andMe, Inc dataset. We identified 63 genome-wide significant loci ($P < 5 \times 10^{-8}$), including 41 novel. Ancestry-specific analyses identified two loci (*LYPLAL1-AS1/SLC30A10* and *STXBP6-NOVA1*) in African ancestry individuals. Sex-stratified analyses identified two loci (*MYO1D* and *ZBTB7C*) that are specific to women, and four (*EBF2*, *EMX2/RAB11FIP2*, *VCL* and *FAM9A/FAM9B*) that are specific to men. Functional experiments demonstrated that several of the associated regions (*EFEMP1* and *LYPLAL1-SLC30A10*) function as enhancers and show differential activity between risk and reference alleles. Our study highlights the importance of large-scale genomic studies in ancestrally diverse populations for identifying ancestry-specific inguinal hernia susceptibility loci and provides novel biological insights into inguinal hernia etiology.

Introduction

Inguinal hernias are characterized by an opening in the myofascial plane of the oblique and transversalis tissues of the abdominal wall. Inguinal hernias account for 75% of all abdominal wall hernias and can display with a wide range of symptoms, including asymptomatic bulge, severe pain or intestinal obstruction caused by incarceration or strangulation (1,2). Men have a much greater cumulative lifetime incidence of inguinal hernias (20–27%) compared to women (3–6%) (3), and African American men have a lower incidence of inguinal hernia compared to non-Hispanic white men (3), however it is not clear why these differences exist.

Patients with a known family history of an inguinal hernia are more likely to develop an inguinal hernia than patients with no known family history (4). Family study and array-based heritability (array-h^2) estimates found a stronger contribution of genetic risk factors in women (sibling standardized incidence ratio (SIR) = 2.38, 95% CI (2.30–2.47); array-h^2 = 20.8–25.5%) compared to men (SIR = 1.91, 95%CI (1.89–1.94); array-h^2 = 13.2–18.3%) (4,5), suggesting that sex-specific genetic effects may underlie some of the difference in risk.

We have previously conducted the first genome-wide association study (GWAS) of adult-onset inguinal hernia, using the Kaiser Permanente Northern California (KPNC) Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, and identified four genetic loci (*EFEMP1*, *ADAMTS6*, *EBF2* and *WT1*) associated at a genome-wide level of significance ($P < 5 \times 10^{-8}$) with inguinal hernia risk in individuals of European ancestry (5). A recent genetic study identified 24 loci associated with inguinal hernia susceptibility at a genome-wide level of significance by conducting a

Received: September 14, 2021. **Revised:** December 6, 2021. **Accepted:** January 4, 2022

transethnic meta-analysis using BioBank Japan and UK Biobank data (6). However, none of these loci has been independently replicated in an external cohort and GWAS of inguinal hernia in Africa ancestry individuals and Hispanic/Latinos are lacking. Finally, to our knowledge, no studies have yet conducted in-depth functional studies of inguinal hernia-associated loci to provide important biological insights of inguinal hernia etiology.

To overcome these limitations and better understand the genetics of inguinal hernia, we conducted the largest multiethnic meta-analysis of GWAS of inguinal hernia to date, with a total of 513 120 subjects, including, 35 774 inguinal hernia cases and 477 346 hernia-free controls from two multiethnic cohorts: the GERA (7) and the UK Biobank (UKB) (8,9). We tested the top independently associated single nucleotide polymorphisms (SNPs) ($P < 5.0 \times 10^{-8}$) in 728 418 participants (33 491 inguinal hernia cases and 694 927 controls) from 23andMe research cohort. Cohort summary details are presented in Supplementary Material, Data S1. Conditional, ancestry-, and sex-specific analyses were also conducted (Supplementary Material, Fig. S1), as well as genetic correlation between inguinal hernia risk and many disorders and complex traits (10) and Mendelian randomization (MR) analysis to assess the nature of the relationship between body mass index (BMI) and inguinal hernia. Finally, we prioritize inguinal hernia-associated genomic regions using *in silico* annotation tools (11–13), and characterize the functionality of these regions using RNA-sequencing (RNA-seq), chromatin immunoprecipitation (ChIP)-seq experiments and differential enhancer assays.

Results

Multiethnic meta-analysis of GERA and UKB

We first undertook a GWAS analysis of inguinal hernia risk stratified by sex and ethnic group, followed by a meta-analysis across all strata. In the meta-analysis, we identified 63 loci associated with inguinal hernia ($P < 5 \times 10^{-8}$), of which 41 were novel (Table 1 and Supplementary Material, Figs S2 and S3). The effect estimates of 62 lead SNPs were consistent across the two studies (Table 1, Supplementary Material, Data S2, and Supplementary Material, Fig. S4), with the lead SNP rs6478957 at *HMCN2* being the one exception.

Replication in the 23andMe research cohort

For replication, we used the 23andMe research cohort, where 45 out of 63 lead SNPs replicated at a Bonferroni corrected significance threshold of 7.94×10^{-4} (P-value $< 0.05/63$) and additional 11 lead SNPs were associated at a nominal level ($P < 0.05$) with a consistent direction of effect (Table 1, Supplementary Material, Data S2, and Supplementary Material, Fig. S5). Taken together, 56 of the 63 SNPs tested (89%) were validated in the 23andMe Research cohort, highlighting the reliability of our results.

Replication of previous inguinal hernia GWAS results

We also investigated in GERA the lead SNPs within 24 loci associated with inguinal hernia at a genome-wide significance level from the most recent and exhaustive GWAS of inguinal hernia conducted to date (6). Ten of the 23 available SNPs replicated at Bonferroni significance ($P < 0.05/23 = 2.17 \times 10^{-3}$) in our GERA multiethnic meta-analysis (including *LYPLAL1-AS1* rs2820465, *ADAMTS6* rs7702887 and *LIMK1* rs75566398) (Supplementary Material, Data S3). Further, five additional SNPs showed nominal evidence of association.

Conditional analyses

To identify independent signals within the 63 identified genomic regions, we performed a multi-SNP-based conditional & joint association analysis (COJO) (14), which revealed 14 additional independent SNPs within 10 of the identified genomic regions, including at known loci (*TGFB2/LYPLAL1*, *EFEMP1*, *ERC2*, *LY86/RREB1*, *ELN*, *EBF2*, *BNC2* and *WT1-AS*) and at newly identified loci (near *ADAMTS16*, and *DMRT2/SMARCA2*) (Supplementary Material, Data S4).

Ancestry-specific analyses

For ethnic groups represented in each cohort, we conducted ancestry-specific meta-analyses of each group. In the non-Hispanic white/European groups, we identified two additional loci: *CMPK1* (rs150464441, $P = 3.89 \times 10^{-8}$), and *PLEKHM3* (rs6435401, $P = 3.86 \times 10^{-8}$) (Supplementary Material, Fig. S6 and Supplementary Material, Data S5). In the African American/African British groups, we identified two genome-wide significant loci: *LYPLAL1-AS1/SLC30A10* (rs184568680, $P = 6.49 \times 10^{-9}$) and *STXBP6-NOVA1* (rs148423010, $P = 4.60 \times 10^{-8}$). Regional association plots of the association signals are presented in Supplementary Material, Figure S7. The meta-analysis of East Asian groups did not result in the identification of genome-wide significant findings. Similarly, while conducting a GWAS of Hispanic/Latinos in the GERA cohort did not result in the identification of genome-wide significant findings, we found that two loci (i.e. *WT1-AS* and *MIR222/AK098783*—identified in the multiethnic meta-analysis of GERA and UKB) were associated with inguinal hernia at Bonferroni significance ($P < 0.05/63 = 7.94 \times 10^{-4}$) (Supplementary Material, Data S2).

Sex-specific analyses identified additional loci

Next, we conducted separate meta-analyses by sex. We identified two additional loci that were significantly associated with inguinal hernia susceptibility in women but not in men (Fig. 1). These include *MYO1D* (rs138335232, $P = 5.36 \times 10^{-9}$ in women, and $P = 0.84$ in men), and *ZBTB7C* (rs150461228, $P = 2.11 \times 10^{-8}$ in

Table 1. Inguinal hernia loci identified in the combined (GERA+UKB) GWAS multiethnic meta-analysis and replication in the multiethnic 23andMe research cohort

					Combined (GERA + UKB) meta-analysis		Replication in 23andMe multiethnic meta-analysis		Direction of effect (GERA-UKB-23andMe)
SNP	**Chr**	**Pos**	**Locus**	**EA/OA**	**OR (SE)**	**P**	**OR (SE)**	**P**	
rs9435183	1	9343528	*H6PD/SPSB1*	C/T	1.09 (0.013)	4.7×10^{-10}	1.07 (0.0118)	6.24×10^{-8}	+++
rs2365498	**1**	**62312654**	***INADL***	**A/G**	**1.06 (0.0097)**	$\mathbf{7.3 \times 10^{-9}}$	**1.02 (0.0098)**	**0.0849**	+++
rs34641909	**1**	**113191585**	***CAPZA1***	**CA/C**	**1.07 (0.011)**	$\mathbf{6.8 \times 10^{-10}}$	**1.06 (0.0106)**	$\mathbf{1.78 \times 10^{-7}}$	+++
rs2799098	1	218521609	*TGFB2*	A/G	1.09 (0.011)	3.3×10^{-14}	1.03 (0.011)	0.00988	+++
rs2820441	1	219734960	*Near LYPLAL1-AS1*	C/A	1.12 (0.0087)	1.5×10^{-37}	1.08 (0.0089)	1.17×10^{-19}	+++
rs6749170	2	25110962	*ADCY3*	G/A	0.95 (0.0082)	4.7×10^{-9}	0.97 (0.0083)	0.000756	—
rs77972916	**2**	**43762112**	***THADA***	**A/G**	**0.90 (0.016)**	$\mathbf{4.4 \times 10^{-11}}$	**0.94 (0.0161)**	$\mathbf{8.04 \times 10^{-5}}$	—
rs966003	**2**	**54884870**	***SPTBN1***	**T/C**	**0.95 (0.0084)**	$\mathbf{3.9 \times 10^{-10}}$	**0.97 (0.0084)**	**0.000277**	—
rs3791679	2	56096892	*EFEMP1*	G/A	0.82 (0.010)	2.5×10^{-87}	0.85 (0.01)	1.64×10^{-55}	—
rs80225179	**2**	**239937457**	***Near HDAC4***	**C/T**	**1.10 (0.017)**	$\mathbf{5.6 \times 10^{-9}}$	**1.05 (0.0187)**	**0.00804**	+++
rs165177	**3**	**8603243**	***LMCD1***	**C/T**	**0.95 (0.0094)**	$\mathbf{4.8 \times 10^{-8}}$	**0.96 (0.0095)**	$\mathbf{1.19 \times 10^{-6}}$	—
rs4974167	3	56139250	*ERC2*	C/A	0.93 (0.0091)	3.4×10^{-15}	0.94 (0.0093)	2.84×10^{-10}	—
rs113371581	**3**	**78816577**	***ROBO1***	**A/G**	**0.95 (0.0091)**	$\mathbf{1.1 \times 10^{-8}}$	**0.99 (0.0092)**	**0.46**	—
rs6805055	**3**	**99430360**	***COL8A1***	**A/T**	**1.09 (0.012)**	$\mathbf{1.3 \times 10^{-12}}$	**1.07 (0.0118)**	$\mathbf{2.13 \times 10^{-9}}$	+++
rs7625122	**3**	**169362673**	***MECOM***	**A/T**	**1.05 (0.0083)**	$\mathbf{1.0 \times 10^{-8}}$	**1.02 (0.0085)**	**0.0198**	+++
rs10535280	4	4913796	*Near LOC101928306*	A/AACACACACAC	1.06 (0.0091)	2.7×10^{-11}	1.03 (0.0091)	0.000197	+++
rs11722597	4	174595677	*Near HAND2-AS1*	T/C	1.07 (0.0083)	2.8×10^{-16}	1.06 (0.0085)	6.37×10^{-11}	+++
rs7715383	**5**	**5350637**	***Near ADAMTS16***	**C/G**	**1.11 (0.014)**	$\mathbf{3.7 \times 10^{-14}}$	**1.14 (0.0139)**	$\mathbf{5.06 \times 10^{-21}}$	+++
rs370763	5	64355060	*CWC27/ADAMTS6*	A/T	1.11 (0.0087)	1.0×10^{-33}	1.08 (0.0089)	1.15×10^{-18}	+++
rs10519694	5	121407219	*LOX*	T/C	1.07 (0.0093)	3.9×10^{-13}	1.03 (0.0095)	0.000304	+++
rs31209	**5**	**134360431**	***Near PITX1***	**T/A**	**0.95 (0.0085)**	$\mathbf{4.4 \times 10^{-9}}$	**0.97 (0.0086)**	**0.00160**	—
rs1294415	6	6740633	*LY86/RREB1*	A/G	0.93 (0.0084)	3.1×10^{-18}	0.95 (0.0085)	2.93×10^{-8}	—
rs115661362	**6**	**35510650**	***TULP1/FKBP5***	**C/T**	**1.15 (0.022)**	$\mathbf{5.5 \times 10^{-10}}$	**1.01 (0.0251)**	**0.75**	+++
rs62400367	**6**	**45481873**	***RUNX2***	**G/A**	**0.93 (0.012)**	$\mathbf{2.7 \times 10^{-10}}$	**0.94 (0.0118)**	$\mathbf{1.34 \times 10^{-6}}$	—
rs3065790	**6**	**55619121**	***BMP5***	**G/GAC**	**0.94 (0.010)**	$\mathbf{7.8 \times 10^{-9}}$	**0.94 (0.0101)**	$\mathbf{7.23 \times 10^{-9}}$	—
rs147247917	6	117479236	*Near VGLL2*	TGTTAGTCACT-GAGGGCACAA/T	0.94 (0.0084)	1.2×10^{-14}	0.95 (0.0086)	1.35×10^{-9}	—
rs6901152	6	143659012	*AIG1*	C/T	0.92 (0.0083)	2.5×10^{-25}	0.95 (0.0084)	1.7×10^{-9}	—
rs36057798	7	25700324	*LOC646588*	GT/G	0.93 (0.010)	2.9×10^{-12}	0.97 (0.0099)	0.00117	—
rs17855988	7	73474825	*ELN*	C/G	0.87 (0.015)	1.1×10^{-19}	0.92 (0.015)	1.66×10^{-7}	—
rs35088033	**7**	**83818276**	***SEMA3A***	**A/G**	**1.10 (0.017)**	$\mathbf{2.0 \times 10^{-8}}$	**1.01 (0.0177)**	**0.67**	+++
rs9640666	**7**	**101032269**	***COL26A1***	**A/G**	**0.95 (0.0082)**	$\mathbf{5.6 \times 10^{-9}}$	**0.97 (0.0083)**	**0.000242**	—
rs34545179	**7**	**116142532**	***CAV2***	**CT/C**	**0.95 (0.0086)**	$\mathbf{1.2 \times 10^{-8}}$	**0.96 (0.0087)**	$\mathbf{4.83 \times 10^{-6}}$	—
rs4618702	8	25708820	*EBF2*	T/G	1.17 (0.0082)	5.3×10^{-83}	1.11 (0.0083)	1.77×10^{-39}	+++

Continued

Table 1. Continued

					Combined (GERA + UKB) meta-analysis		Replication in 23andMe multiethnic meta-analysis		Direction of effect (GERA-UKB-23andMe)
SNP	Chr	Pos	Locus	EA/OA	OR (SE)	P	OR (SE)	P	
rs4292648	**8**	**30324803**	***RBPMS***	**T/C**	**1.05 (0.0082)**	**1.7×10^{-8}**	**1.01 (0.0082)**	**0.50**	**+++**
rs112367264	**8**	**74534528**	***STAU2***	**AT/A**	**0.93 (0.010)**	**6.2×10^{-12}**	**0.95 (0.0098)**	**2.83×10^{-7}**	**—**
rs58846390	**8**	**75566127**	***MIR2052HG***	**C/G**	**0.91 (0.016)**	**1.0×10^{-8}**	**0.92 (0.0162)**	**1.34×10^{-7}**	**—**
rs4741240	**9**	**1280060**	***DMRT2/SMARCA2***	**A/G**	**0.92 (0.012)**	**2.8×10^{-11}**	**0.98 (0.0128)**	**0.0683**	**—**
rs2039187	9	16741049	*BNC2*	C/T	0.90 (0.014)	2.7×10^{-14}	0.96 (0.0146)	0.00944	—
rs6478957	9	133048501	*HMCN2*	A/G	0.94 (0.0093)	2.5×10^{-10}	0.97 (0.0083)	0.000875	+−
rs1417569	**10**	**31101743**	***Near ZNF438***	**A/G**	**0.95 (0.0087)**	**5.1×10^{-9}**	**0.98 (0.0088)**	**0.0143**	**—**
rs2505559	**10**	**43663327**	***CSGALNACT2***	**T/A**	**1.06 (0.0094)**	**7.7×10^{-11}**	**1.06 (0.0095)**	**9.53×10^{-11}**	**+++**
rs11187786	**10**	**95852314**	***PLCE1***	**G/A**	**1.05 (0.0094)**	**2.9×10^{-8}**	**1.03 (0.0097)**	**0.00217**	**+++**
rs374237077	**10**	**119539990**	***EMX2/RAB11FIP2***	**T/TTATCCATCCATC-CATCCCTCCATC-CATC**	**1.12 (0.015)**	**5.9×10^{-14}**	**1.08 (0.0148)**	**6.05×10^{-8}**	**+++**
rs4140413	11	32459228	*WT1-AS*	T/G	0.88 (0.0086)	8.9×10^{-54}	0.92 (0.0087)	7.03×10^{-20}	—
rs10878346	**12**	**66320873**	***HMGA2***	**A/G**	**1.07 (0.0096)**	**2.1×10^{-11}**	**1.05 (0.0097)**	**1.99×10^{-7}**	**+++**
rs11616527	**13**	**32368419**	***RXFP2***	**C/G**	**0.94 (0.0088)**	**9.4×10^{-12}**	**0.98 (0.0085)**	**0.00586**	**—**
rs573666	13	51194405	*BCMS*	T/C	1.07 (0.0085)	6.2×10^{-14}	1.07 (0.0085)	9.01×10^{-14}	+++
rs11315136	**15**	**67478645**	***SMAD3***	**C/CA**	**0.95 (0.0097)**	**7.0×10^{-9}**	**0.95 (0.0097)**	**1.2×10^{-6}**	**—**
rs12442790	**15**	**84487890**	***ADAMTSL3***	**T/C**	**1.05 (0.0084)**	**2.6×10^{-8}**	**1.03 (0.0086)**	**0.000177**	**+++**
rs2076435	**16**	**1607574**	***IFT140***	**T/C**	**0.94 (0.0098)**	**2.2×10^{-9}**	**0.96 (0.0088)**	**2.18×10^{-7}**	**—**
rs4238714	16	84856552	*CRISPLD2*	C/T	1.08 (0.0084)	3.5×10^{-20}	1.04 (0.0084)	2.34×10^{-7}	+++
rs12453693	17	12191339	*MAP2K4/MYOCD*	T/C	1.09 (0.0088)	1.2×10^{-21}	1.07 (0.0089)	3.11×10^{-14}	+++
rs139356332	**17**	**19289286**	***MFAP4***	**C/G**	**0.82 (0.035)**	**9.4×10^{-9}**	**0.88 (0.0353)**	**0.000378**	**—**
rs59640213	**17**	**65892280**	***BPTF***	**A/G**	**0.95 (0.0097)**	**4.3×10^{-8}**	**0.95 (0.0097)**	**3.19×10^{-7}**	**—**
rs299237	**18**	**20307194**	***LOC101927571***	**G/A**	**0.94 (0.011)**	**1.9×10^{-8}**	**0.96 (0.0111)**	**6.86×10^{-5}**	**—**
rs11083561	**19**	**41101981**	***LTBP4***	**C/T**	**1.05 (0.0084)**	**5.6×10^{-9}**	**1.02 (0.0084)**	**0.00360**	**+++**
rs6123685	**20**	**55836040**	***BMP7***	**A/G**	**0.94 (0.0094)**	**1.0×10^{-12}**	**0.93 (0.0098)**	**8.2×10^{-12}**	**—**
rs35318931	**23**	**38009121**	***SRPX***	**A/G**	**0.88 (0.012)**	**7.7×10^{-29}**	**0.93 (0.0125)**	**1.61×10^{-9}**	**—**
rs56976399	**23**	**45634577**	***MIR222/AK098783***	**A/C**	**1.06 (0.0064)**	**3.3×10^{-21}**	**1.05 (0.0067)**	**4.01×10^{-12}**	**+++**
rs12847546	**23**	**109730477**	***RGAG1/ TDGF1P3***	**A/G**	**0.96 (0.0063)**	**2.4×10^{-12}**	**0.98 (0.0068)**	**0.000293**	**—**
rs140000208	**23**	**115181931**	***DANT2/AGTR2***	**G/A**	**0.94 (0.0068)**	**7.3×10^{-20}**	**0.94 (0.0066)**	**1.14×10^{-20}**	**—**
rs71806106	**23**	**133781440**	***PLAC1***	**C/CTG**	**1.04 (0.0072)**	**9.4×10^{-9}**	**1.04 (0.0079)**	**1.2×10^{-6}**	**+++**
rs5905042	**23**	**146444527**	***MIR514A2/FMR1***	**A/C**	**0.96 (0.0066)**	**4.1×10^{-8}**	**0.99 (0.0067)**	**0.131**	**—**

Note: Table 1 reports the fixed effects summary estimates for an additive model and the full results, including random effects summary estimates (along with heterogeneity index, I^2 (0–100%) as well as *P*-value for Cochrane's Q statistic among groups/cohorts) are reported in Supplementary Material, Data S2. SNP, single nucleotide polymorphism; Chr, chromosome; Pos, position; EA, effect allele; OA, other allele; SE, standard error. SNPs with novel associations for inguinal hernia identified here highlighted in bold are novel

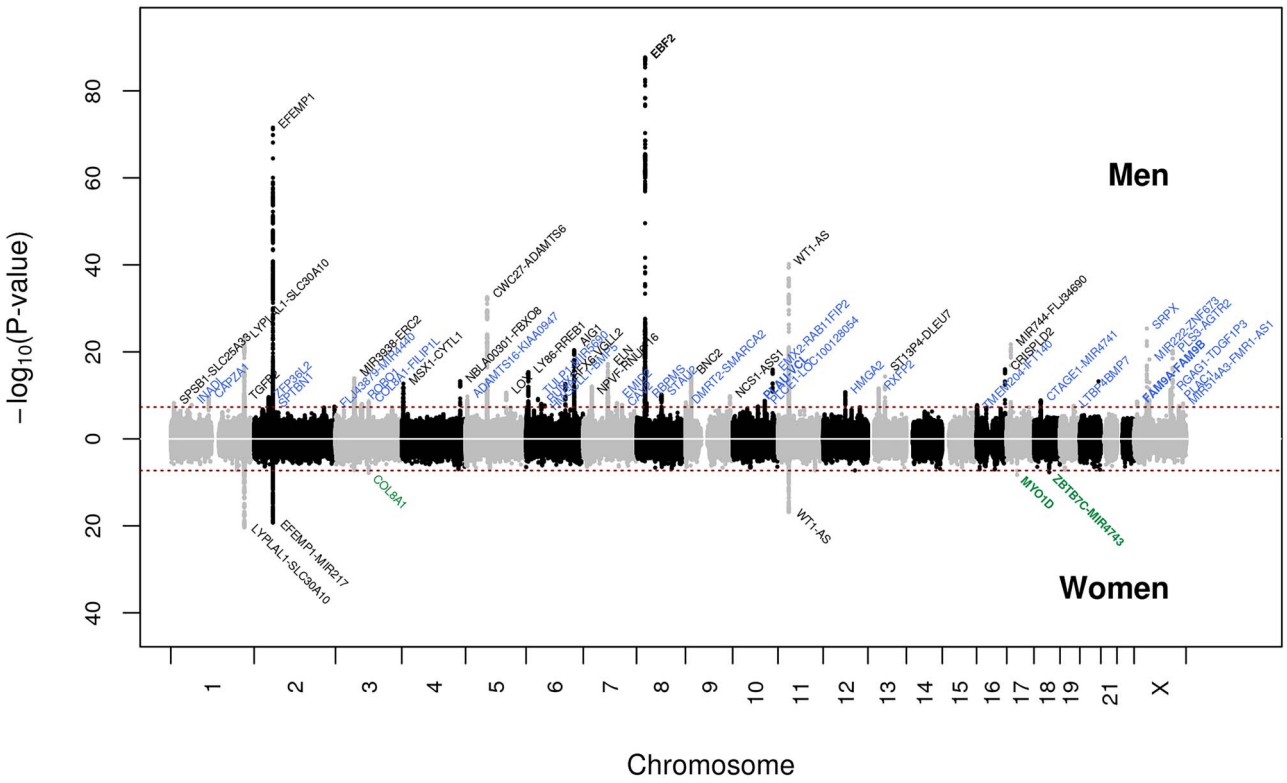

Figure 1. Chicago plot of the sex-stratified multiethnic GWAS meta-analyses of inguinal hernia. Results from the meta-analysis combining men from GERA and UKB are presented on upper panel, whereas results from the meta-analysis combining women from GERA and UKB are presented on the lower panel. The *y*-axis represents the -$\log_{10}$(*P*-value); all *P*-values derived from logistic regression model are two-sided. The red dotted line represents the threshold of $P = 5 \times 10^{-8}$ which is the commonly accepted threshold of adjustments for multiple comparisons in GWAS. Locus names in black are for those previously reported. Locus names in bold (*MYO1D*, *ZBTB7C*, near *VCL* and *FAM9A/FAM9B*) are for the additional novel loci specific to women or men (compared to the multiethnic meta-analysis (GERA+UKB)). Although novel loci significantly associated ($P < 5 \times 10^{-8}$) with inguinal hernia in women are highlighted in green, those significantly associated with inguinal hernia in men are highlighted in blue.

women, and $P = 0.94$ in men) (Fig. 2 and Supplementary Material, Data S6). We also identified two additional loci, near *VCL* and *FAM9A/FAM9B*, that were significantly associated with inguinal hernia susceptibility in men (*VCL* rs3955312, $P = 3.60 \times 10^{-8}$; *FAM9A/FAM9B* rs56355307, $P = 2.69 \times 10^{-9}$) but not in women (rs3955312, $P = 0.52$; rs56355307, $P = 0.20$). We then tested the 63 inguinal hernia lead variants identified in the combined multiethnic meta-analysis for significant differences in effects between men and women (Fig. 3). We observed that four loci, *LYPLAL1-AS1/SLC30A10*, *COL8A1*, *EBF2* and *EMX2/RAB11FIP2*, were significantly differently associated with inguinal hernia susceptibility in men and women (Supplementary Material, Data S7 and Supplementary Material, Fig. S8).

To further evaluate the shared genetic basis of inguinal hernia between women and men, we compared the GWAS results from the two sex-specific meta-analyses for each ethnic group by performing a LD score regression (LDSC). We observed a positive genetic correlation (r_g) between women and men for inguinal hernia among European ancestry individuals ($r_g = 0.63$, $P = 4.70 \times 10^{-18}$), which is the largest group of individuals in the current study. However, we were unable to assess the shared genetic basis of inguinal hernia between women and men in other ethnic groups (i.e. Hispanic/Latinos, East Asians and African ancestry individuals), due to the limited sample size (and the lack or limited of significant signals) in these groups.

Gene and pathways prioritization and tissue-enrichment analysis

Data-driven expression prioritized integration for complex traits (DEPICT) (11) gene prioritization analysis identified 12 genes after false-discovery rate (FDR) correction, of which five (i.e. *WNT2B*, *LMCD1*, *COL8A1*, *ADAMTS16* and *RBPMS*) were within novel inguinal hernia-associated loci (Supplementary Material, Data S8). Prioritized genes at identified loci included genes involved in crosslinking of collagens, elastin and elastic fibers (*COL8A1*, *ELN*, *LOX*), transforming growth factor-β (TGF-β) signaling pathway (*SPSB1*, *WNT2B*), protein–protein interactions (*LMCD1*, *EFEMP1*, *SPSB1*, *ADAMTS16*). DEPICT (11) tissue-enrichment analysis highlighted 17 significantly associated (FDR < 0.05) tissues or cell type annotations; four annotations pertained to the musculoskeletal system such as joint capsule, joints, synovial membrane and cartilage (Supplementary Material, Data S9). Additional annotations included mesenchymal stem cells, chondrocytes, stromal cells, fibroblasts and adipocytes/adipose tissue; or involved the urogenital or digestive system, including, myometrium, genitalia or pancreas. DEPICT (11) gene-set enrichment

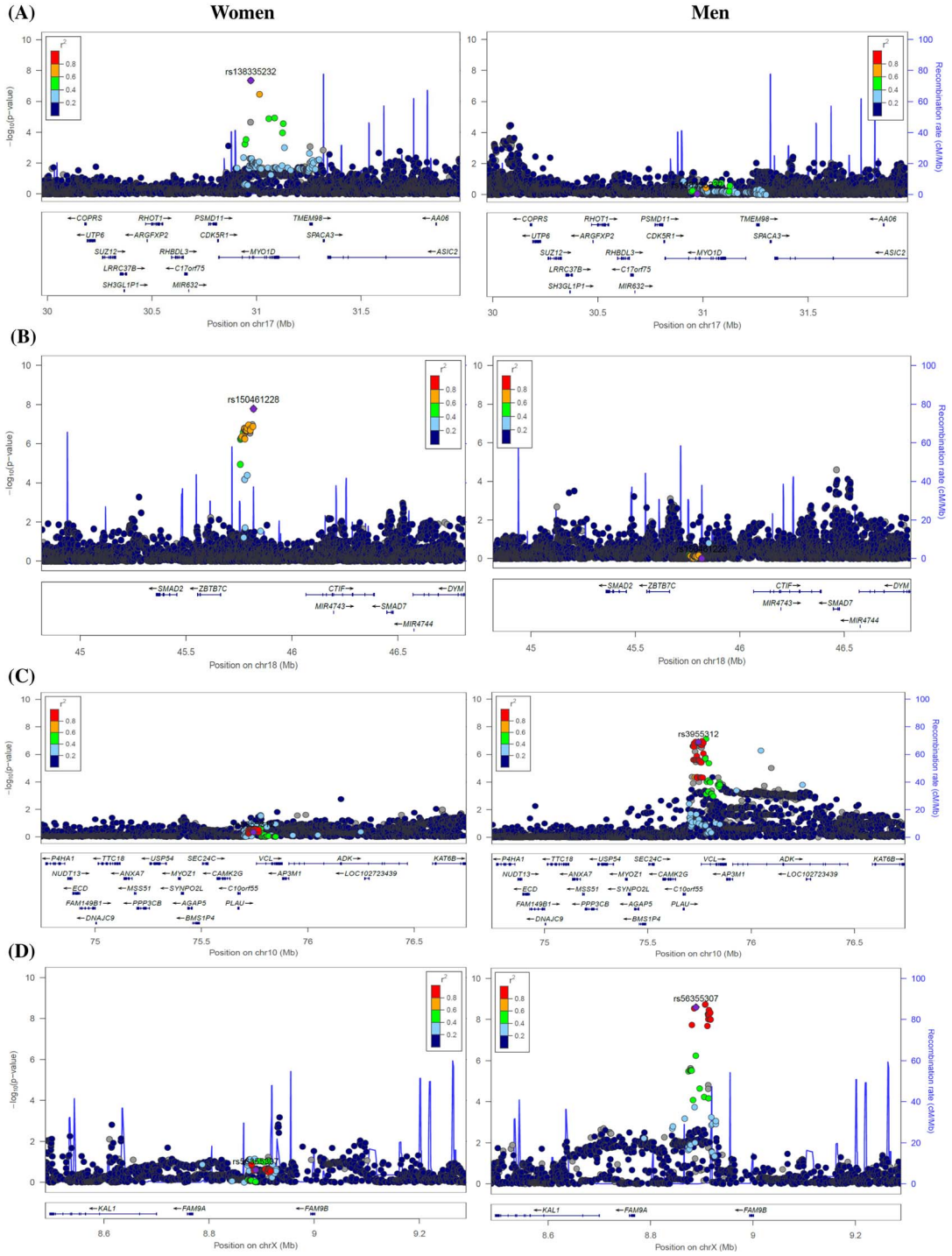

Figure 2. Locus Zoom plots of regions showing differential association with inguinal hernia across women and men. Multiethnic combined (GERA + UKB) meta-analysis stratified by sex identified: two regions significant in women ($P < 5 \times 10^{-8}$) but not significant ($P > 0.05$) in men: (**A**) *MYO1D*, and (**B**) near (or within) *ZBTB7C* (according to the *ZBTB7C* transcript); and two regions significant in men ($P < 5 \times 10^{-8}$) but not significant ($P > 0.05$) in women: (**C**) near *VCL*, and (**D**) *FAM9A*/*FAM9B*

analysis detected nine pathways to prioritize after FDR correction, including those involved in the morphogenesis of epithelial tube and tissue, and cellular response to nutrient (Supplementary Material, Data S10). To provide tissue-enrichment visualization, we used FUMA (12) integrative web-based platform that accommodates expression quantitative trait loci (eQTL) data for 53 tissues from the Genotype Tissue Expression Project (GTEx) v7 (15). FUMA (12) tissue eQTL specificity analysis highlighted the esophagus gastroesophageal junction, esophagus muscularis and uterus, as the main tissues for which expression were affected by inguinal hernia-associated variants (Bonferroni significance $P < 9.43 \times 10^{-4}$) (Supplementary Material, Fig. S9).

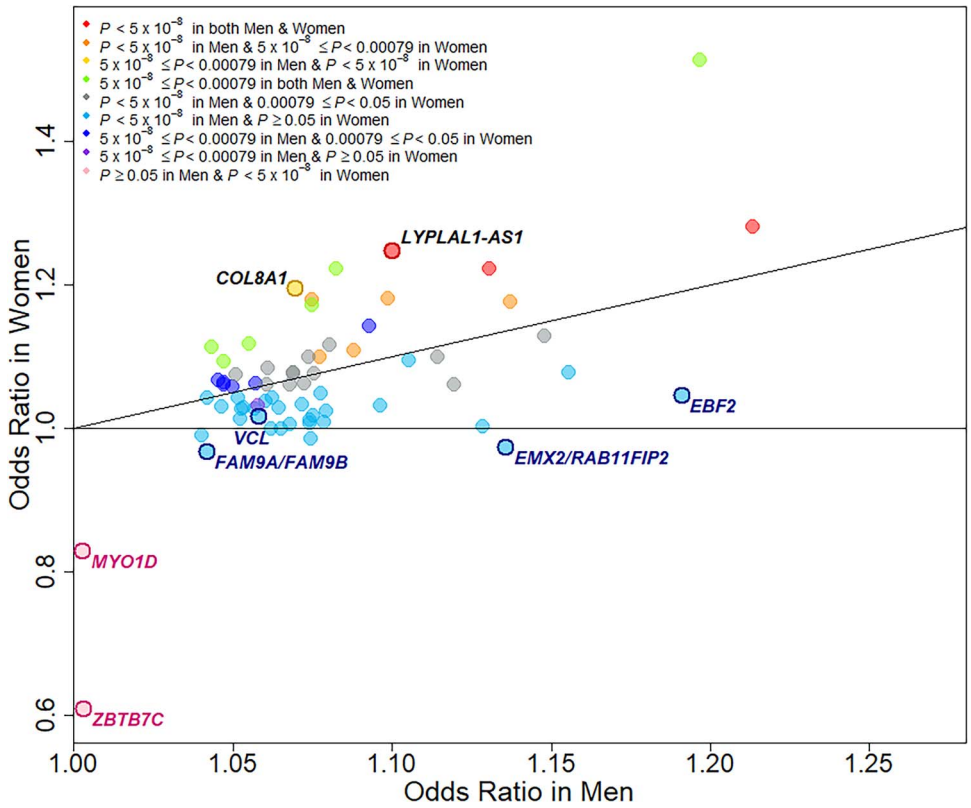

Figure 3. Correlation of effect sizes for inguinal hernia between women and men for the lead 63 SNPs identified in the combined (GERA + UKB) GWAS multiethnic analysis. Here we report the locus names for the lead SNPs that have a significantly different OR between the women- and men-specific analysis. The effect sizes were compared using a correlation test ($R^2 = 0.35$, two-sided P-value = 1.17×10^{-7}). Locus names in dark pink color are for the loci specific to women and the ones in navy color are for the loci specific to men.

To prioritize which cell types are more relevant for inguinal hernia, we also conducted a cell type-specific analysis using chromatin data from multiple tissues (16) and using stratified LD score regression (S-LDSC) with the baseline-LD model v2.0 (17). We found that fetal muscle and fetal stomach were the most relevant cell types for inguinal hernia based on statistical significance ($P < 1.02 \times 10^{-4}$ corresponding to 0.05/489 cell type-specific chromatin annotation tested) (Supplementary Material, Data S11). We then used S-LDSC with the baseline-LD model to partition the heritability of inguinal hernia in order to evaluate the contribution of the cell type-specific chromatin annotations to this condition. Notably, we conducted cell type-specific analyses using 'GastroIntestinal system'-, 'SkeletalMuscle'- and 'Connective/Bone'- specific chromatin annotations, as those were the most relevant for inguinal hernia. We found that inguinal hernia had significant enrichment for several annotations (Supplementary Material, Data S12–S14).

Genetic correlation between inguinal hernia and other traits

To estimate the pairwise genetic correlations (r_g) between inguinal hernia and >700 diseases/traits from different publicly available resources/consortia, we used the LD Hub web interface (10), which performs automated LD score regression. We detected significant genetic correlations between inguinal hernia and 26 other diseases/traits after Bonferroni correction. In particular, we found a negative genetic correlation between inguinal hernia and BMI ($r_g = -0.14$, $P = 1.99 \times 10^{-8}$) and a positive genetic correlation between inguinal hernia and moderate physical activity ($r_g = 0.15$, $P = 5.55 \times 10^{-5}$) (Supplementary Material, Data S15 and Supplementary Material, Fig. S10), both of which have previously been identified as inguinal hernia risk factors in observations studies (18–20). We found an additional 137 nominal genetic correlations ($P < 0.05$) with inguinal hernia, including for tobacco smoking status, and diverticular disease of intestine.

MR analyses

To investigate whether BMI causally influences inguinal hernia risk, we conducted a two-sample MR analysis using established genetic variants from a GWAS of BMI conducted in the UK Biobank European sample (21), to proxy the BMI exposure (see Methods). Using 444 independent genetic variants previously reported as genome-wide significant ($P < 5.0 \times 10^{-8}$) as genetic instruments for BMI (Supplementary Material, Data S16), we found evidence for a causal effect of BMI on the risk of inguinal hernia, such as lower BMI was associated with an increased risk of inguinal hernia (Inverse variance-weighted (IVW) model: OR [95% CI] 0.78 [0.69–0.89],

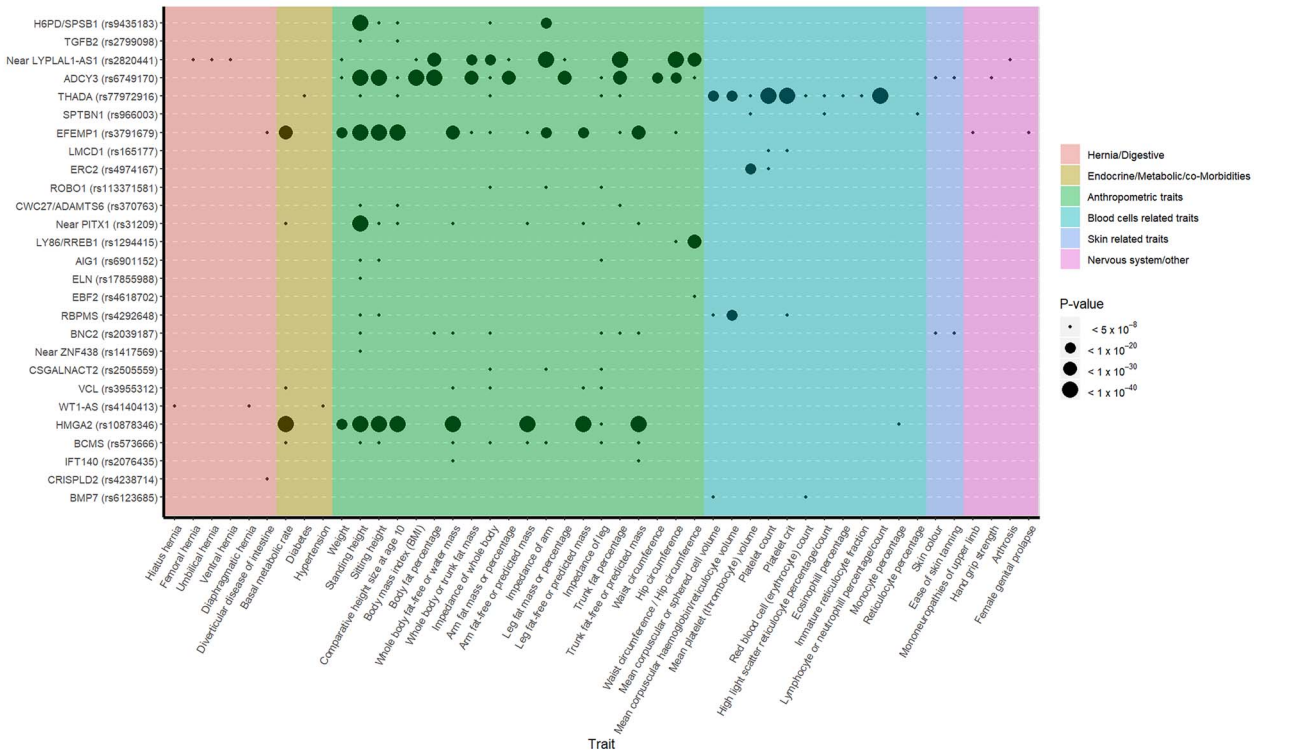

Figure 4. Phenome-wide association matrix of inguinal hernia top variants. PheWAS was carried out for the 48 lead SNPs in our loci of interest identified in the combined (GERA + UKB) multiethnic analysis. SNPs were queried against 778 traits ascertained for UKB participants and reported in the Roslin Gene Atlas (21), including other hernia types, diverticular disease of the intestine, anthropometric traits, hematologic laboratory values and skin related traits. Among the 71 lead SNPs in our loci of interest (63 from the combined multiethnic analysis + 4 ethnic-specific + 4 sex-specific), 48 were available in Gene Atlas database. We reported SNPs showing genome-wide significant association with at least one trait (in addition to inguinal hernia). As a note, a few lead SNPs were located in intergenic regions, and for those, we reported the nearest gene as the locus name preceded with 'near'.

$P = 1.65 \times 10^{-4}$) (Supplementary Material, Data S17–Supplementary Material, Fig. S11). These 444-SNP genetic instruments explained close to 3.5% of the phenotypic variation in BMI in the GERA non-Hispanic white sample.

Pleiotropic analyses

We next performed a phenome-wide association study (PheWAS), which can determine whether a genetic variant is associated with other phenotypes, by testing associations between 48 lead SNPs (out of 71 lead SNPs in our loci of interest: 63 from the combined multiethnic analysis + 4 ethnic-specific + 4 sex-specific) that were available in GeneATLAS, and 778 traits (21). We found that 16 of the top inguinal-associated variants were significantly associated ($P < 5.0 \times 10^{-8}$) with additional traits (Fig. 4). Although variant rs2820441, near *LYPLAL1-AS1*, was associated with femoral, umbilical, and ventral hernias, variant rs4140413 at *WT1-AS* was associated with diaphragmatic and hiatus hernias (Supplementary Material, Data S18). Moreover, variants at *EFEMP1* and *HMGA2* were significantly associated with endocrine/metabolic traits, such as basal metabolic rate, and anthropometric traits such as whole body fat-free. Variants at *THADA* and *ERC2*, and *RBPMS*, were significantly associated with blood cells traits, such as platelet crit and mean corpuscular hemoglobin.

Variants prioritization and annotations

To prioritize genes for follow-up functional evaluation based on causal variants, we used a Bayesian approach (CAVIARBF) (13). For each of the associated signals, we computed each variant's capacity to explain the identified signal within a 2 Mb window (±1.0 Mb with respect to the original top variant) and derived the smallest set of variants that included the causal variant with 95% probability. Seven sets included a unique variant (Supplementary Material, Data S19). These included *PNPT1* rs7584120, *LMCD1* rs165177, *ST13P4-DLEU7* rs573666, *MFAP4* rs139356332, *BMP7* rs6123685, *SRPX* rs35318931 and *MIR222-ZNF673* rs56976399 with >95.0% posterior probability of being the causal variants, suggesting that those variants could be causal.

Functional characterization

As many of the hernia-associated genes have connective tissue roles, we analyzed H3K27ac ChIP-seq, a marker for active promoter and enhancer regions (22,23), in mouse connective tissue generated by our lab (24) to identify putative regulatory elements that overlap with our hernia-associated SNPs (Supplementary Material, Data S20). We mapped these elements to hernia risk loci identified in the CAVIARBF analysis and selected segments of DNA containing both associated genetic

variants and H3K27ac ChIP-seq peaks from our connective tissue H3K27ac ChIP-seq or from ENCODE (25) H3K27ac from a variety of human cell lines for enhancer assays.

We next selected 30 ChIP-seq peaks that had CAVIRABF SNPs in them to test for enhancer activity using a luciferase reporter assays. We cloned all 30 sequences in front of a minimal promoter and luciferase reporter gene and transfected them into human male fibroblast cells (BJ cells), as fibroblasts are the most common cell type in connective tissue, and this is an easy to transfect cell line. We found 15 (50%) to be functional enhancers (Fig. 5A). We then cloned the alternate allele for eight of these functional enhancer regions, chosen by the strength of their associations with inguinal hernia in the GWAS analyses, and found that six of them, including at *EFEMP1* and *LYPLAL1-SLC30A10*, show significant differential enhancer activity between the reference and risk allele (Fig. 5B).

Discussion

In summary, we reported a large multiethnic meta-analysis GWAS for inguinal hernia that identified 41 novel loci as contributing to the pathophysiology of this common disease. Importantly, we reported, for the first time of our knowledge, two genetic loci (*LYPLAL1-SLC30A10* and *STXBP6-NOVA1*) associated at a genome-wide level of significance with inguinal hernia risk in African ancestry individuals. Additionally, eight loci showed sex-specific effects on inguinal hernia susceptibility. Finally, the functional characterization of inguinal hernia-associated regions (i.e. at *EFEMP1* and *LYPLAL1-SLC30A10*) supports effects of genetic variants on gene regulation.

Our study also reported, for the first time to our knowledge, sex-specific loci associated with inguinal hernia susceptibility. Although intronic variants at *MYO1D* and *ZBTB7C* were associated with inguinal hernia risk in women but not in men, intergenic variants near *VCL* and at *FAM9A/FAM9B* were associated with inguinal hernia in men but not in women. The *MYO1D* gene on chromosome 17 encodes a member of the class I myosin family which is produced in the intestinal epithelium. In mice, *MYO1D* has been shown to maintain epithelial integrity and protect against intestinal homeostasis abnormalities such as colitis (26). The *ZBTB7C* on chromosome 18 encodes the zinc finger and BTB domain containing 7C protein and is broadly expressed in the esophagus. Zbtb7c is involved in the regulation of fatty acid biosynthesis, gluconeogenesis and adipocyte differentiation (27,28). The *VCL* (vinculin) gene on chromosome 10 encodes a cytoskeletal protein associated with cell–cell and cell-matrix junctions and is crucial for the regulation of force transduction in cells (29). Thus, even though our findings provide important insights into the biological mechanisms underlying inguinal hernia susceptibility, future studies will help to elucidate whether those genes are causal and how they contribute to this condition.

In this study, we also found evidence for shared genetic influences between BMI and inguinal hernia, as well as potential causal effects of BMI on inguinal hernia risk. Future investigations could benefit from genetic instrumental variable analyses for mechanism-specific BMI risk. Indeed, numerous biological processes underlying BMI variation have been reported (e.g. linked to adipose cell impairment, including the adipogenesis and insulin signaling pathways) and may have distinct consequences on inguinal hernias development. Thus, BMI genetic subscores related to each of the biological processes could be used to elucidate aspects of BMI physiology that may influence risk of inguinal hernias development.

In this study, we also found evidence for shared genetic influences between physical activity and inguinal hernia. In observational studies, the impact of occupation, heavy lifting, exercise and physical activity is controversial (30–33). Future studies would be needed to clarify further the nature of the relationship between physical activity and inguinal hernia risk.

Interestingly, our genetic correlation results also indicate that inguinal hernia was significantly correlated with diverticular disease of intestine. Some of the inguinal hernia-associated loci reported here were previously associated with diverticular disease (i.e. *LYPLAL1*, *EFEMP1*, *CWC27/ADAMTS6*, *ELN* and *CRISPLD2*) (34). *ELN* encodes elastin which confers elasticity to tissues; altered *ELN* can lead to structural changes of the colonic wall observed in diverticular disease (35). In parallel, our PheWAS findings demonstrate that inguinal hernia-associated variants at *LYPLAL1-AS1* and *WT1-AS* are also associated with other subtypes of hernia (i.e. femoral, umbilical or ventral). Our PheWAS findings are consistent with a recent study (36), which reported a high-level of genetic correlation among hernia subtypes. Future large and ethnically diverse studies will determine whether the identified loci contribute to different hernia subtypes (i.e. femoral, umbilical, ventral, diaphragmatic or hiatus) and the extent to which these loci display shared effects across subtypes.

Our study should be interpreted within the context of its limitations. First, although all of the inguinal hernia cases in GERA and 99.6% of the UKB cases were based on diagnosis or procedure codes (e.g. ICD-10 diagnosis or CPT-4 procedure codes), inguinal hernia cases in 23andMe research cohort were based on self-reported data. This may result in phenotype misclassification, however, the associations identified in our meta-analysis combining GERA and UKB were well validated in the 23andMe research cohort. Second, we recognize that the analysis of large cohorts, such as GERA, for which phenotypes are mainly derived from electronic health records (EHR) could lead to substantial case–control imbalance that could result in elevated Type 1 error rates (false positives) (37,38). However, when we applied REGENIE

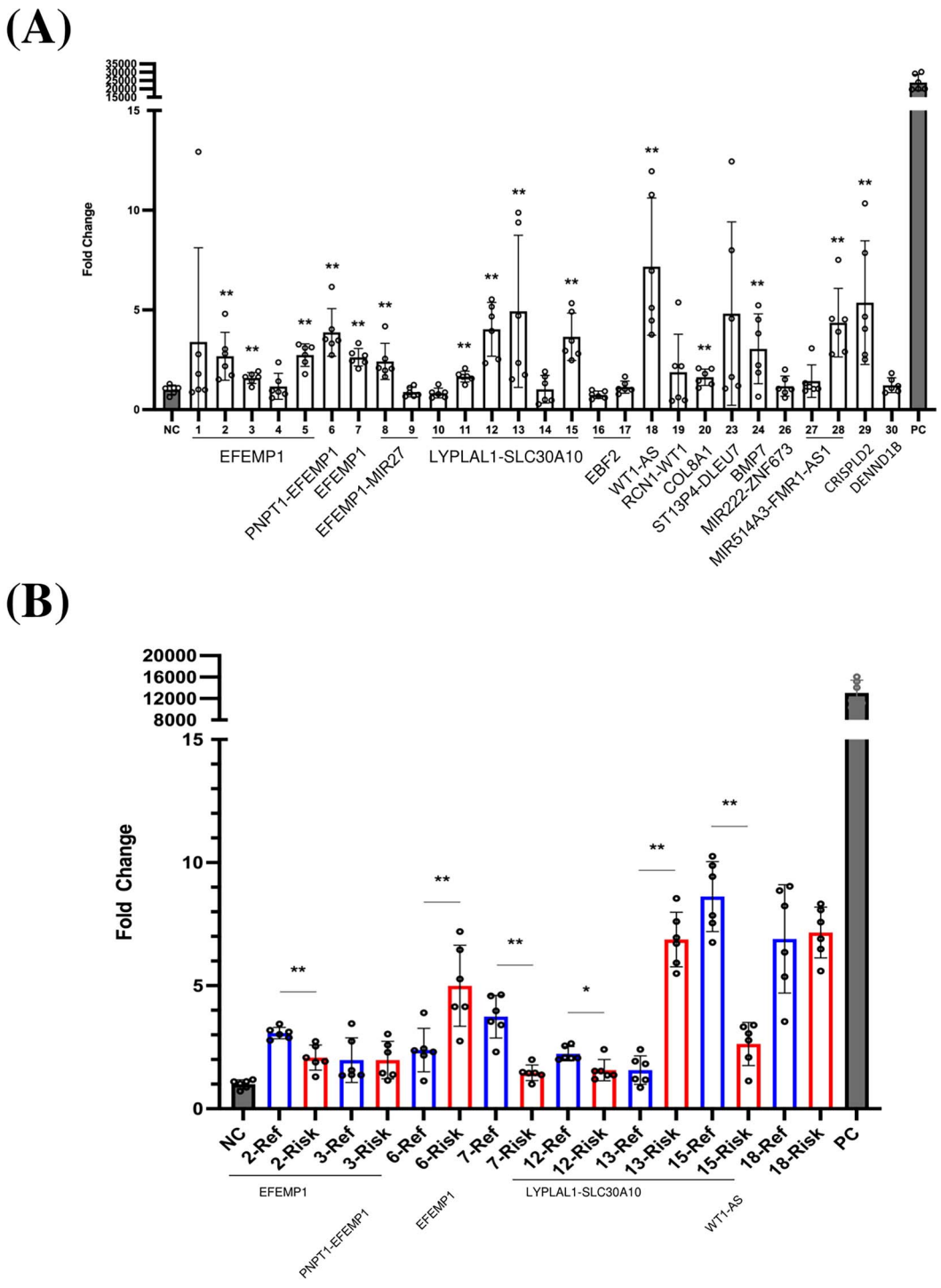

Figure 5. Luciferase enhancer assays for hernia-associated sequences. (**A**) Relative luciferase activity after 72 h post transfection for 27 hernia-associated sequences, normalized for transfection efficiency with Renilla. Fold changes were calculated compared to negative control (NC). Loci names are written below. PC = positive control, * = *P*-value < 0.05 and ** = *P*-value < 0.01 for a Student T-test. (**B**) Comparison of differential enhancer activity between the reference allele (Ref) and hernia risk allele (Risk). Loci names are written below. Negative control (NC), PC = positive control, * = *P*-value < 0.05 and ** = *P*-value < 0.01 for a Student T-test comparing between both alleles.

and an approximate Firth regression approach, which has been shown to efficiently control for case–control imbalance (38), we found almost identical genetic associations with inguinal hernia, compared to standards approaches. Third, the difference in study participation (i.e. relatively active participant engagement in UKB and 23andMe versus more passive participant involvement in GERA) and sex-differential participation have the potential to impact our genetic results due to study participation bias (39) and could be considered as a study limitation. Finally, while our functional experiments demonstrate that several of the associated regions function as enhancers and show differential activity between alleles, further research is needed to fully

understand the causal variant at those loci and their underlying mechanism.

Our findings provide a biological foundation for understanding the etiology of ancestry- and sex-differences in inguinal hernia susceptibility, and, more generally, identify potential targets for the development of non-surgical treatment of inguinal hernias.

Materials and Methods

Populations and participants

GERA cohort

The GERA cohort contains genome-wide genotype, clinical and demographic data of over 110 000 adult members from mainly four ethnic groups (non-Hispanic white, Hispanic/Latino, East Asian and African American) of the KPNC Medical Care Plan (7,40). The Institutional Review Board of the Kaiser Foundation Research Institute has approved all study procedures. Patients eligible for inclusion were identified from clinical diagnoses captured in the KPNC EHR system. These clinical diagnoses and procedures were recorded in the EHR system as International Classification of Diseases, Ninth or Tenth Revision (ICD-9 or ICD-10) codes, or as Current Procedural Terminology, 4th Edition (CPT-4) procedure codes. We defined inguinal hernia cases as having any evidence of inguinal hernia, based on diagnosis codes (ICD-9: 550.X; ICD-10: K40.X), procedure codes (ICD-9: 53.0X, 53.1X, 17.1X, 17.2X; CPT-4: 49491, 49492, 49495, 49496, 49500, 49501, 49505, 49507, 49520, 49521, 49525, 49650, 49651, 49659), and post-operative diagnosis. After excluding subjects who have any evidence of any type of hernia, our control group included all the non-cases. In total, 9861 inguinal hernia cases and 74 249 controls from GERA were included in this study. Protocols for participant genotyping, data collection and quality control have been described in detail (40). Briefly GERA participants' DNA samples were extracted from Oragene kits (DNA Genotek Inc., Ottawa, Ontario, Canada) at KPNC and genotyped at the Genomics Core Facility of UCSF. DNA samples were genotyped at over 665 000 genetic markers on four ethnic-specific Affymetrix Axiom arrays (Affymetrix, Santa Clara, CA, USA) optimized for European, Latino, East Asian and African American individuals (41,42). Genotype quality control (QC) procedures and imputation were conducted on an array-wise basis (40). For imputation, we additionally removed variants with call rates <90%, by array. Genotypes were then pre-phased with Eagle (43) v2.3.2, and then imputed with Minimac3 (44) v2.0.1, using two reference panels. Variants were preferred if present in the EGA release of the Haplotype Reference Consortium (HRC; $n = 27\,165$) reference panel (45), and from the 1000 Genomes Project Phase III release if not ($n = 2504$; i.e. indels) (46).

UK Biobank

The UKB is a large prospective study following the health of ~500 000 participants from five ethnic groups (European, East Asian, South Asian, African British and mixed ancestries) resident in the UK aged between 40- and 69 years old at the baseline recruitment visit (9,47). Demographic information and medical history were ascertained through touch-screen questionnaires. Participants also underwent a wide range of physical and cognitive assessments, including blood sampling. The inguinal hernia phenotype was assessed through diagnosis codes (ICD-10: K40.X), procedure codes (code 1563 in operation data field 20004, and codes T19, T20, T21 in procedure data field 41200) or self-report data (code 1513 in self-reported data field 20002), and cases were defined as having any evidence of inguinal hernia. The control group included the non-cases and excluded individuals with any evidence of any type of hernia. Phenotyping, genotyping and imputation were carried out by members of the UK Biobank team. Imputation to the Haplotype Reference Consortium reference panel has been described (www.ukbiobank.ac.uk). Following QC, over 10 million variants in 429 010 individuals were tested for adjusting for age, and genetic ancestry principal components. The analyses presented in this paper were carried out under UK Biobank Resource project #14105.

23andMe research cohort

Replication analysis of 63 loci identified in the combined (GERA+UKB) meta-analysis was conducted using self-reported data from a GWAS including 33 491 inguinal hernia cases and 694 927 controls of five ethnic groups (i.e. European, Latino, East Asian, South Asian and African American), from 23andMe, Inc., customer database. All individuals included in the analyses provided informed consent and answered surveys online according to the 23andMe human subject protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com). 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). Participants were included in the analysis on the basis of consent status as checked at the time data analyses were initiated. Cases were defined as those who reported having inguinal hernia; controls were defined as those who reported not having no hernia. Those with other hernia types are excluded from the control definition.

Statistical analyses

GWAS and adjustment in GERA. We first analyzed men and women separately for each ethnic group (non-Hispanic white, Hispanic/Latino, East Asian and African American). We ran a logistic regression of inguinal hernia and each SNP using PLINK (48) v1.9 (www.cog-genomics.org/plink/1.9/) adjusting for age, and ancestry principal components (PCs), which were previously (7) assessed within each ethnic group using Eigenstrat (49) v4.2. We included as covariates the top 10 ancestry PCs for the non-Hispanic whites, whereas we included the top six

ancestry PCs for the three other ethnic groups. To adjust for genetic ancestry, we also included the percentage of Ashkenazi (ASHK) ancestry as a covariate for the non-Hispanic white sample analyses (7). As a sensitivity analysis, we have also conducted a GWAS of inguinal hernia adjusted for BMI and this analysis produced relatively similar results compared to the analysis without adjusting for BMI (Supplementary Material, Data S21–Supplementary Material, Fig. S12). The GWAS analyses were also conducted using a recent approach accounting for relatedness that fits a whole-genome regression model, implemented in REGENIEv2.0.2 (38) (https://rgcgithub.github.io/regenie/). The GWAS results generated using REGENIE were similar compared to the results generated using PLINK (Supplementary Material, Data S22–Supplementary Material, Fig. S13).

GWAS meta-analyses. First, a meta-analysis of inguinal hernia was conducted in GERA to combine the results of men and women and the results of the four ethnic groups using the R (50) (https://www.R-project.org) package 'meta'. Similarly, a meta-analysis was conducted in UKB to combine the results of men and women and the results of the five ethnic groups. Three ethnic-specific meta-analyses were also performed: (1) combining European-specific samples (i.e. GERA non-Hispanic whites and UKB Europeans); (2) combining Asian-specific samples (i.e. GERA and UKB East Asians); and (3) combining African-specific samples (i.e. GERA African Americans and UKB Africans). Two sex-specific meta-analyses were also performed: (1) combining women from GERA and UKB; and (2) combining men from GERA and UKB. A last meta-analysis was conducted to combine the results from GERA and UKB. Fixed effects summary estimates were calculated for an additive model. We also estimated heterogeneity index, I^2 (0–100%) and *P*-value for Cochrane's Q statistic among different groups, and studies. For each locus, we defined the top SNP as the most significant variant within a 2 Mb window. Novel loci were defined as those that were located over 1 Mb apart from any previously reported locus (5).

COJO analysis. A multi-SNP-based COJO analysis (14) was performed on the combined European-specific (GERA non-Hispanic whites + UKB Europeans) meta-analysis results to potentially identify independent signals within the 63 identified genomic regions. To calculate linkage disequilibrium (LD) patterns, we used 10 000 randomly selected samples from GERA non-Hispanic white ethnic group as a reference panel. A *P*-value $< 5.0 \times 10^{-8}$ was considered as the significance threshold for this COJO analysis.

Post-GWAS analyses

DEPICT prioritization

To prioritize genes and highlight gene-set and tissue/-cell enrichments within the 63 inguinal hernia genomic regions identified in the multiethnic combined (GERA + UKB) meta-analysis, we used DEPICT (11). This integrative tool considers multiple lines of complementary evidence to systematically prioritize the most likely causal genes at associated loci, highlight enriched pathways, and identify tissues/cell types in which genes from associated loci are highly expressed. Genes, gene-sets, tissue/-cell annotations that achieved a nominal significance level of 0.05 after FDR correction were subsequently prioritized.

FUMA tissue eQTL specificity

To highlight and visualize tissue eQTL enrichments within the 63 inguinal hernia-associated genomic regions identified in the combined multiethnic (GERA + UKB) meta-analysis, we used FUMA (12) integrative tool. FUMA is an integrative web-based platform (https://fuma.ctglab.nl/) that accommodates eQTL, and provides tissue-enrichment results for each of 53 tissue types based on the genotype-tissue expression (GTEx) v6 RNA-seq data (15).

Genetic correlations

To estimate the genetic correlation of EF phenotype with >700 diseases/traits, including hernia phenotypes, from different publicly available resources/consortia, we used the LD Hub web interface (http://ldsc.broadinstitute.org/) (10), which performs automated LD score regression. In the LD Score regressions, we included only HapMap3 SNPs (51) with MAF > 0.01. Genetic correlations were considered significant after Bonferroni adjustment for multiple testing ($P < 6.5 \times 10^{-5}$ which corresponds to 0.05/770 phenotypes tested).

Genetic instruments for BMI

Genetic variants as instrumental variables for BMI (exposure) were extracted from one GWAS conducted in 499 421 UK Biobank participants of European ancestry from GeneATLAS (http://geneatlas.roslin.ed.ac.uk/) (21). In the current study, we used the following set of genetic instruments: lead SNPs previously reported as genome-wide significant ($P < 5.0 \times 10^{-8}$). Genetic instruments were then clumped using a window of 10 Mb and maximal LD of $r^2 = 0.001$ between instruments to ensure that genetic variants were independent. After clumping, a total of 444 genetic instruments for BMI ($P < 5.0 \times 10^{-8}$) were used for the MR analyses (Supplementary Material, Data S16). The proportion of phenotypic variance in BMI explained by those variants was calculated in the GERA non-Hispanic white sample to assess the strength of genetic instruments.

Two-sample MR analyses

All two-sample MR analyses were conducted in the R computing environment V.4.0.1. using the 'TwoSampleMR' package. This package makes causal inference about an exposure on an outcome using GWAS summary statistics, generates LD pruning of exposure SNPs and harmonizes exposure and outcome data sets (e.g.

direction of association effects). We used the IVW method as our primary source of MR estimates. This IVW method essentially translates to a weighted regression of SNP outcome effects on SNP-exposure effects where the intercept is constrained to zero.

PheWAS analyses

PheWAS was carried out for the 71 lead SNPs in our loci of interest (63 from the combined multiethnic analysis + 4 ethnic-specific + 4 sex-specific). SNPs were queried against 778 traits ascertained for UKB participants and reported in the Roslin Gene Atlas (21), including anthropometric traits, hematologic laboratory values, ICD-10 clinical diagnoses and self-reported conditions. Among the 71 lead SNPs, 48 were available in Gene Atlas database. We reported SNPs showing genome-wide significant association with at least one trait (in addition to inguinal hernia).

Variants prioritization

To prioritize variants within the identified genomic regions for follow-up functional evaluation, a Bayesian approach (CAVIARBF) (13) was used, which is available publicly at https://bitbucket.org/Wenan/caviarbf. Each variant's capacity to explain the identified signal within a 2 Mb window (±1.0 Mb with respect to the original top variant) was computed for each identified genomic region. Then, the smallest set of variants that included the causal variant with 95% probability (95% credible set) was derived. Out of the 2198 total variants, 48 variants had >20% probability of being causal (including 34 lead SNPs), prioritizing 23 genes.

Functional experiments

Luciferase assays

All sequences were PCR amplified (for primers see Supplementary Material, Data S14) and cloned into the pGL4.23 enhancer assay plasmid (Promega) using the NEBuilder HIFI DNA Assembly Master Mix (New England Biolabs). Inserts were Sanger sequence verified for having the proper sequence and allele and were purified using the QIAprep Spin Miniprep Kit (Qiagen). BJ (ATCC, CRL-2522) cells were cultured using Eagle's Minimum Essential Medium (ATCC) supplemented with 10% Fetal Bovine Serum and 1% Penstrep. Cells were subcultured ∼every 3 days. Cells were plated on 24 well plates at 3.0×10^4/ml. One thousand three hundred and fifty nanograms of plasmid was transfected into cells along with 150 ng Renilla luciferase vecto to correct for transfection efficiency, at a 10:1 ratio for each triplicate (pGL4.73; Promega) using X-tremeGENE HP DNA Transfection Reagent (Roche) according to the manufacturer's protocol. Three independent replicate cultures were carried out for each plasmid and two independent biological replicates. After 72 h, the cells were washed with PBS and lysed in buffer PLB (Promega). Firefly and Renilla luciferase activities were measured on a Glomax microplate reader (Promega) using the Dual-Luciferase Reporter Assay System (Promega). Enhancer activity was calculated as the fold change of each plasmid's firefly luciferase activity normalized to Renilla luciferase activity.

Supplementary Material

Supplementary Material is available at *HMG* online

Acknowledgements

We are grateful to the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment, and Health. We would also like to thank the research participants and employees of 23andMe for making this work possible.

Conflict of Interest statement. The 23andMe authors (PN and CT) are employees of and own stock or stock options in 23andMe, Inc. All other authors declare they have no competing interests.

Funding

The National Institute of Diabetes and Digestive and Kidney Diseases [R01 DK116738 to H.C. N.A. and E.J.]. The National Eye Institute [R01 EY027004 to H.C.] and the National Cancer Institute [R01 CA241623 to H.C.] and the National Institute of Arthritis and Musculoskeletal and Skin Diseases [R21 AR076009 to H.C.]. Genotyping of the GERA cohort was funded by a grant from the National Institute on Aging, National Institute of Mental Health, and National Institute of Health Common Fund [RC2 AG036607]. Support for GERA participant enrollment, survey completion, and biospecimen collection for the Research Program on Genes, Environment and Health was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente Community Benefit Programs.

References

1. Fitzgibbons, Jr., Giobbie-Hurder, A., Gibbs, J.O., Dunlop, D.D., Reda, D.J., McCarthy, M., Jr., Neumayer, L.A., Barkun, J.S., Hoehn, J.L., Murphy, J.T. *et al.* (2006) Watchful waiting vs repair of inguinal hernia in minimally symptomatic men: a randomized clinical trial. *JAMA*, **295**, 285–292.
2. Nilsson, H., Nilsson, E., Angeras, U. and Nordin, P. (2011) Mortality after groin hernia surgery: delay of treatment and cause of death. *Hernia*, **15**, 301–307.
3. Ruhl, C.E. and Everhart, J.E. (2007) Risk factors for inguinal hernia among adults in the US population. *Am. J. Epidemiol.*, **165**, 1154–1161.
4. Zoller, B., Ji, J., Sundquist, J. and Sundquist, K. (2013) Shared and nonshared familial susceptibility to surgically treated inguinal

hernia, femoral hernia, incisional hernia, epigastric hernia, and umbilical hernia. *J. Am. Coll. Surg.*, **217**, 289, e281–299.
5. Jorgenson, E., Makki, N., Shen, L., Chen, D.C., Tian, C., Eckalbar, W.L., Hinds, D., Ahituv, N. and Avins, A. (2015) A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. *Nat. Commun.*, **6**, 10130.
6. Hikino, K., Koido, M., Tomizuka, K., Liu, X., Momozawa, Y., Morisaki, T., Murakami, Y., The Biobank Japan, P, Mushiroda, T. and Terao, C. (2021) Susceptibility loci and polygenic architecture highlight population specific and common genetic features in inguinal hernias: genetics in inguinal hernias. *EBioMedicine*, **70**, 103532.
7. Banda, Y., Kvale, M.N., Hoffmann, T.J., Hesselson, S.E., Ranatunga, D., Tang, H., Sabatti, C., Croen, L.A., Dispensa, B.P., Henderson, M. *et al.* (2015) Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*, **200**, 1285–1295.
8. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.
9. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. *et al.* (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, **12**, e1001779.
10. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Early, G. *et al.* (2017) 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.
11. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T. *et al.* (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.*, **6**, 5890.
12. Watanabe, K., Taskesen, E., van Bochoven, A. and Posthuma, D. (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.*, **8**, 1826.
13. Chen, W., Larrabee, B.R., Ovsyannikova, I.G., Kennedy, R.B., Haralambieva, I.H., Poland, G.A. and Schaid, D.J. (2015) Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics*, **200**, 719–736.
14. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication, D.I.G., Meta-analysis, C, Madden, P.A., Heath, A.C., Martin, N.G. *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369, S361–375, 363.
15. Consortium, G.T (2015) Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
16. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.R., Lareau, C., Shoresh, N. *et al.* (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.*, **50**, 621–629.
17. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K. *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.
18. Lau, H., Fang, C., Yuen, W.K. and Patil, N.G. (2007) Risk factors for inguinal hernia in adult males: a case-control study. *Surgery*, **141**, 262–266.
19. Liem, M.S., van der Graaf, Y., Zwart, R.C., Geurts, I. and van Vroonhoven, T.J. (1997) Risk factors for inguinal hernia in women: a case-control study. The Coala trial group. *Am. J. Epidemiol.*, **146**, 721–726.
20. Rosemar, A., Angeras, U. and Rosengren, A. (2008) Body mass index and groin hernia: a 34-year follow-up study in Swedish men. *Ann. Surg.*, **247**, 1064–1068.
21. Canela-Xandri, O., Rawlik, K. and Tenesa, A. (2018) An atlas of genetic associations in UK Biobank. *Nat. Genet.*, **50**, 1593–1599.
22. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21931–21936.
23. Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T. and Wysocka, J. (2012) Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell*, **11**, 633–648.
24. Makki, N., Zhao, J., Liu, Z., Eckalbar, W.L., Ushiki, A., Khanshour, A.M., Wu, J., Rios, J., Gray, R.S., Wise, C.A. *et al.* (2021) Genomic characterization of the adolescent idiopathic scoliosis-associated transcriptome and regulome. *Hum. Mol. Genet.*, **29**, 3606–3615.
25. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
26. McAlpine, W., Wang, K.W., Choi, J.H., San Miguel, M., McAlpine, S.G., Russell, J., Ludwig, S., Li, X., Tang, M., Zhan, X. *et al.* (2018) The class I myosin MYO1D binds to lipid and protects against colitis. *Dis. Model. Mech.*, **11**, dmm035923.
27. Choi, W.I., Yoon, J.H., Choi, S.H., Jeon, B.N., Kim, H. and Hur, M.W. (2021) Proto-oncoprotein Zbtb7c and SIRT1 repression: implications in high-fat diet-induced and age-dependent obesity. *Exp. Mol. Med.*, **53**, 917–932.
28. Choi, W.I., Yoon, J.H., Song, J.Y., Jeon, B.N., Park, J.M., Koh, D.I., Ahn, Y.H., Kim, K.S., Lee, I.K. and Hur, M.W. (2019) Zbtb7c is a critical gluconeogenic transcription factor that induces glucose-6-phosphatase and phosphoenylpyruvate carboxykinase 1 genes expression during mice fasting. *Biochim Biophys Acta Gene Regul Mech*, **1862**, 643–656.
29. Kanoldt, V., Kluger, C., Barz, C., Schweizer, A.L., Ramanujam, D., Windgasse, L., Engelhardt, S., Chrostek-Grashoff, A. and Grashoff, C. (2020) Metavinculin modulates force transduction in cell adhesion sites. *Nat. Commun.*, **11**, 6403.
30. Patterson, T., Currie, P., Spence, R., McNally, S. and Spence, G. (2018) A systematic review of the association between a single strenuous event and the development of an inguinal hernia: a medicolegal grey area. *Surgeon*, **16**, 309–314.
31. Vad, M.V., Frost, P., Bay-Nielsen, M. and Svendsen, S.W. (2012) Impact of occupational mechanical exposures on risk of lateral and medial inguinal hernia requiring surgical repair. *Occup. Environ. Med.*, **69**, 802–809.
32. Svendsen, S.W., Frost, P., Vad, M.V. and Andersen, J.H. (2013) Risk and prognosis of inguinal hernia in relation to occupational mechanical exposures–a systematic review of the epidemiologic evidence. *Scand. J. Work Environ. Health*, **39**, 5–26.
33. Mitura, K., Smietanski, M., Koziel, S., Garnysz, K. and Michalek, I. (2018) Factors influencing inguinal hernia symptoms and preoperative evaluation of symptoms by patients: results of a prospective study including 1647 patients. *Hernia*, **22**, 585–591.

34. Maguire, L.H., Handelman, S.K., Du, X., Chen, Y., Pers, T.H. and Speliotes, E.K. (2018) Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease. *Nat. Genet.*, **50**, 1359–1365.
35. Golder, M., Burleigh, D.E., Ghali, L., Feakins, R.M., Lunniss, P.J., Williams, N.S. and Navsaria, H.A. (2007) Longitudinal muscle shows abnormal relaxation responses to nitric oxide and contains altered levels of NOS1 and elastin in uncomplicated diverticular disease. *Color. Dis.*, **9**, 218–228.
36. Wei, J., Attaar, M., Shi, Z., Na, R., Resurreccion, W.K., Haggerty, S.P., Zheng, S.L., Helfand, B.T., Ujiki, M.B. and Xu, J. (2021) Identification of fifty-seven novel loci for abdominal wall hernia development and their biological and clinical implications: results from the UK Biobank. *Hernia*, https://doi.org/10.1007/s10029-021-02450-4. Online ahead of print.
37. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A. *et al.* (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.*, **50**, 1335–1341.
38. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B. *et al.* (2021) Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.*, **53**, 1097–1103.
39. Pirastu, N., Cordioli, M., Nandakumar, P., Mignogna, G., Abdellaoui, A., Hollis, B., Kanai, M., Rajagopal, V.M., Parolo, P.D.B., Baya, N. *et al.* (2021) Genetic analyses identify widespread sex-differential participation bias. *Nat. Genet.*, **53**, 663–671.
40. Kvale, M.N., Hesselson, S., Hoffmann, T.J., Cao, Y., Chan, D., Connell, S., Croen, L.A., Dispensa, B.P., Eshragh, J., Finn, A. *et al.* (2015) Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*, **200**, 1051–1060.
41. Hoffmann, T.J., Kvale, M.N., Hesselson, S.E., Zhan, Y., Aquino, C., Cao, Y., Cawley, S., Chung, E., Connell, S., Eshragh, J. *et al.* (2011) Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics*, **98**, 79–89.
42. Hoffmann, T.J., Zhan, Y., Kvale, M.N., Hesselson, S.E., Gollub, J., Iribarren, C., Lu, Y., Mei, G., Purdy, M.M., Quesenberry, C. *et al.* (2011) Design and coverage of high throughput genotyping arrays optimized for individuals of east Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics*, **98**, 422–430.
43. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., Reshef, Y.A., Finucane, H.K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R. *et al.* (2016) Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.*, **48**, 1443–1448.
44. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M. *et al.* (2016) Next-generation genotype imputation service and methods. *Nat. Genet.*, **48**, 1284–1287.
45. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K. *et al.* (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
46. Birney, E. and Soranzo, N. (2015) Human genomics: the end of the start for population sequencing. *Nature*, **526**, 52–53.
47. Allen, N.E., Sudlow, C., Peakman, T., Collins, R. and Biobank, U.K. (2014) UK biobank data: come and get it. *Sci. Transl. Med.*, **6**, 224ed224.
48. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.
49. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
50. (2014) R: a language and environment for statistical computing. *R Found Stat Comput*.
51. International HapMap, C, Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58.