

# SCIENTIFIC REPORTS



OPEN

## Heritability and genome-wide association study of benign prostatic hyperplasia (BPH) in the eMERGE network

Jacklyn N. Hellwege<sup>1,2</sup>, Sarah Stallings<sup>3</sup>, Eric S. Torstenson<sup>1,2</sup>, Robert Carroll<sup>4</sup>, Kenneth M. Borthwick<sup>5</sup>, Murray H. Brilliant<sup>6</sup>, David Crosslin<sup>7</sup>, Adam Gordon<sup>8</sup>, George Hripcsak<sup>9,10</sup>, Gail P. Jarvik<sup>11</sup>, James G. Linneman<sup>12</sup>, Parimala Devi<sup>7</sup>, Peggy L. Peissig<sup>13</sup>, Patrick A. M. Sleiman<sup>14,15</sup>, Hakon Hakonarson<sup>14,15</sup>, Marylyn D. Ritchie<sup>16</sup>, Shefali Setia Verma<sup>16</sup>, Ning Shang<sup>9</sup>, Josh C. Denny<sup>4,17</sup>, Dan M. Roden<sup>4,17,18</sup>, Digna R. Velez Edwards<sup>2,4,19</sup> & Todd L. Edwards<sup>1,2</sup>

Benign prostatic hyperplasia (BPH) results in a significant public health burden due to the morbidity caused by the disease and many of the available remedies. As much as 70% of men over 70 will develop BPH. Few studies have been conducted to discover the genetic determinants of BPH risk. Understanding the biological basis for this condition may provide necessary insight for development of novel pharmaceutical therapies or risk prediction. We have evaluated SNP-based heritability of BPH in two cohorts and conducted a genome-wide association study (GWAS) of BPH risk using 2,656 cases and 7,763 controls identified from the Electronic Medical Records and Genomics (eMERGE) network. SNP-based heritability estimates suggest that roughly 60% of the phenotypic variation in BPH is accounted for by genetic factors. We used logistic regression to model BPH risk as a function of principal components of ancestry, age, and imputed genotype data, with meta-analysis performed using METAL. The top result was on chromosome 22 in *SYN3* at rs2710383 ( $p$ -value =  $4.6 \times 10^{-7}$ ; Odds Ratio = 0.69, 95% confidence interval = 0.55–0.83). Other suggestive signals were near genes *GLGC*, *UNCA13*, *SORCS1* and between *BTBD3* and *SPTLC3*. We also evaluated genetically-predicted gene expression in prostate tissue. The most significant result was with increasing predicted expression of *ETV4* (chr17;  $p$ -value = 0.0015). Overexpression of this gene has been associated with poor prognosis in prostate cancer. In conclusion, although there were no genome-wide significant variants identified for

<sup>1</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.

<sup>2</sup>Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>3</sup>Division of Geriatric Medicine, Meharry-Vanderbilt Alliance, Nashville, TN, USA. <sup>4</sup>Department of Biomedical Informatics Vanderbilt University, Nashville, TN, USA. <sup>5</sup>Hood Center for Health Research, Geisinger Health System, Danville, PA, USA.

<sup>6</sup>Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI, USA. <sup>7</sup>Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, USA. <sup>8</sup>Division of Medical Genetics, University of Washington, Seattle, WA, USA. <sup>9</sup>Department of Biomedical Informatics, Columbia University, New York, NY, USA. <sup>10</sup>Medical Informatics Services, New York-Presbyterian Hospital, New York, NY, USA.

<sup>11</sup>Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA, USA. <sup>12</sup>Office of Research Computing and Analytics/Marshfield Clinic Research Institute, Marshfield, WI, USA. <sup>13</sup>Center for Computational and Biomedical Informatics, Marshfield Clinic Research Institute, Marshfield, WI, USA. <sup>14</sup>Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>15</sup>Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

<sup>16</sup>Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA. <sup>17</sup>Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>18</sup>Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>19</sup>Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA.

Correspondence and requests for materials should be addressed to D.R.V.E. (email: [digna.r.velez.edwards@vumc.org](mailto:digna.r.velez.edwards@vumc.org)) or T.L.E. (email: [todd.l.edwards@vumc.org](mailto:todd.l.edwards@vumc.org))

**BPH susceptibility, we present evidence supporting the heritability of this phenotype, have identified suggestive signals, and evaluated the association between BPH and genetically-predicted gene expression in prostate.**

Benign prostatic hyperplasia (BPH) is characterized by an enlarged prostate that affects a moderate proportion of middle-aged men and a large proportion of elderly men and can result in significant discomfort and reproductive and urinary tract dysfunction<sup>1–4</sup>. Lower urinary tract symptoms (LUTS) are commonly attributed to BPH in the absence of other causes<sup>5–7</sup>. Very severe cases can result in urinary tract infections and bleeding, bladder stones, and kidney damage from failing to void<sup>7–10</sup>. Pharmaceutical treatments for BPH include alpha blockers to relax muscles and treat some LUTS symptoms, and 5-alpha reductase inhibitors which can shrink the prostate in some patients but may increase risk for prostate cancer<sup>11–16</sup>. Additionally, the available surgical remedies can present additional risks and have considerable potential consequences for reproductive and urinary tract health<sup>17–21</sup>.

Heritability of LUTS scores in twins has been estimated at 20–40%<sup>22</sup>, with some estimates as high as 83%<sup>23</sup>, while heritability of benign prostate disease has been estimated at 49% from twin studies<sup>24</sup>. The presence of racial disparities also supports a genetic contribution to BPH risk<sup>25,26</sup>. Evaluation of the SNP-based additive genetic heritability has not yet been published.

The genetic factors underlying BPH risk remain unclear. To date there have been many BPH candidate gene studies, often evaluating the effect of prostate cancer susceptibility variants, with mixed success<sup>27–47</sup>. Three larger-scale studies have been performed, a MetaboChip analysis of prostate volume<sup>48</sup> and two recent genome-wide association studies (GWAS) of BPH<sup>49,50</sup>. In the present study, we evaluated genetic heritability of clinically reported BPH and conducted a GWAS using cases and controls identified from the Electronic Medical Records and Genomic (eMERGE) network, and evaluated the contribution of the genetic associations to gene expression in prostate tissue.

## Results

**Heritability.** SNP-based additive heritability among common variants was assessed in the 5 sites of the eMERGE-1 network and one of the Geisinger datasets (CoreExome) as they had the largest number of cases assessed on a common genotyping array (Table 1). After stringent filters to remove residual population stratification, there were 755 cases and 899 controls included from eMERGE-1 and 423 cases and 1278 controls included from Geisinger CoreExome. Heritability results were consistent between the two groups, with an estimated heritability of 0.65 ( $\pm 0.30$ ) in eMERGE-1 (p-value = 0.011) and 0.56 ( $\pm 0.38$ ) in Geisinger (p-value = 0.070) (Table 2). Results were largely consistent across inclusion of increasing PCs (Supplementary Table 1). Results across chromosomes varied substantially (Supplementary Fig. 1), however, chromosomes 6 and 7 were present among the top 5 results for both eMERGE-1 and Geisinger, suggesting the likelihood of one or more BPH-susceptibility loci being located on these chromosomes.

**Genome-wide association.** In total, 2,656 cases and 7,763 controls were included across eight eMERGE sites (including the data used in the heritability analysis; Table 1) for the analysis of common (minor allele frequency [MAF] > 0.05) genetic association at 10,973,920 SNPs. Overall, the samples were predominantly identified as white (85%) and cases were slightly older than controls (average age of 68.88 years in cases vs 61.45 years in controls). The most statistically significant results from single SNP GWAS analyses was on chromosome 22 in synapsin 3 (*SYN3*) at rs2710383 (allele frequency = 0.12, p-value =  $4.56 \times 10^{-7}$ ; Odds Ratio [OR] = 0.69, 95% confidence interval [CI] = 0.55–0.83; Table 3; Fig. 1). Other suggestive signals were within genes glutamate-cysteine ligase catalytic subunit (*GCLC*; chromosome 6), unc-13 homolog A (*UNC13A*; chromosome 19), and ELOVL [elongation of very long chain fatty acids] fatty acid elongase 6 (*ELOVL6*; chromosome 4), near the long intergenic non-protein coding RNA 1919 (*LINC01919*; chromosome 18), and in an intergenic region on chromosome 20 between BTB domain containing 3 (*BTBD3*) and serine palmitoyltransferase long chain base subunit 3 (*SPTLC3*). Secondary analysis restricting to whites yielded consistent results for top SNPs (Table 3), but identified the top variants as rs10786938 in *SORCSI* (p-value =  $3.84 \times 10^{-7}$ , OR = 1.23; Supplementary Table 2).

We also evaluated the previously identified variants from recent GWAS to determine whether variants replicated across studies (Table 4)<sup>49,50</sup>. None of the variants reported in those studies were significant or suggestively associated with BPH in this analysis, although effect estimates were largely consistent in both direction and magnitude. Restricting to whites only to more closely match the papers<sup>49,50</sup> did not yield significant results.

**Gene expression.** We also evaluated genetically predicted gene expression (GPGE) in prostate tissue using S-PrediXcan<sup>51</sup> and models constructed in GTEx samples<sup>52</sup> (Table 5; Fig. 1). The top result did not reach statistical significance (Bonferroni threshold for number of genes, p-value <  $1.93 \times 10^{-5}$ ) was with increased predicted expression of ETS variant 4 (*ETV4*; chromosome 17; p-value = 0.0015). Other nominally significant genes were identified on chromosomes 6, 20, 3, 14, 1, and 7. It is noteworthy that neither of the genes on chromosome 6 (histone cluster 1 H3 family member e [*HIST1H3E*]) and 20 (gonadotropin-releasing hormone 2 [*GNRH2*]), were near the top signals implicated in the GWAS results (*GLGC* and the intergenic region between *BTBD3* and *SPTLC3*), instead these suggestive GPGE results arose from secondary signals in other regions.

## Discussion

We have performed the first SNP-based heritability assessments of BPH followed by a trans-ethnic GWAS and evaluation of genetically predicted gene expression in prostate tissue. Our results indicate that BPH is likely to be substantially heritable, with consistent point estimates near 60% across two comparable EMR-based cohorts, which is somewhat higher than the 49% reported previously from twin studies<sup>24</sup>. The LUTS symptom score

Study Population	% White	Cases		Controls	
		N	Mean Age (Years)(SD)	N	Mean Age (Years)(SD)
eMERGE-1	94.8	912	67 (10)	1,077	69 (13)
Geisinger (eMERGE-2)	99.5	242	66 (8)	807	61 (13)
Mayo (eMERGE-2)	94.2	157	65 (9)	1,166	62 (11)
Mt Sinai (eMERGE-2)	15.9 <sup>a</sup>	283	64 (9)	1,037	56 (10)
BioVU (eMERGE-2)	100	156	65 (10)	106	62 (12)
Harvard	88.9	94	64 (10)	943	60 (11)
Columbia	30.4 <sup>b</sup>	34	73 (9)	242	60 (7) <sup>c</sup>
Geisinger CoreExome	99.4	574	66. (7)	1,685	61 (9)
Mayo Clinic	95.6 <sup>d</sup>	204	65 (7)	700	63 (10)
TOTAL		2,656	69	7,763	61

**Table 1.** Study Characteristics. <sup>a</sup>Mt Sinai analyzed as 4 separate groups by race/ethnicity. <sup>b</sup>Columbia was a mixed sample, with many samples lacking reported race information. Excluded from whites-only analysis. <sup>c</sup>Only 27 controls for Columbia had age information. <sup>d</sup>Mayo Clinic was composed of 4 separate chip-sets, each one was >93% white.

Cohort	N Cases/N Controls	h <sup>2</sup>	SE	P-value
eMERGE-1	755/899	0.65	0.30	0.011
Geisinger CoreExome	423/1278	0.56	0.38	0.070

**Table 2.** SNP-based heritability of BPH in two cohorts. Adjusted for age and PCs 1–5.

heritability however has been reported to be variable, with estimates ranging from 20 to 83%. The cases in this study likely have overt symptoms of BPH that lead to their clinical diagnoses and treatments, and may represent a more severe phenotype than from some cohort studies.

In this first GWAS of EHR-assessed BPH, we identified previously unreported suggestive SNPs. The gene containing the top SNP from the GWAS, *SYN3* is a neuronal protein<sup>53,54</sup> which has been implicated in GWAS of many diverse phenotypes including age-related macular degeneration<sup>55–57</sup>, height<sup>58–60</sup>, and uric acid levels<sup>61</sup>. Expression of *SYN3* in GTEx is highest in testis, followed by several brain regions, but is low in prostate and no predictive model was constructed for *SYN3* expression in that tissue<sup>52,62</sup>. Another neuronal protein *UNC13A* (unc-13 homolog A) was also implicated from these GWAS results. Variants near this gene have been consistently associated with amyotrophic lateral sclerosis in several genome-wide studies<sup>63–65</sup>.

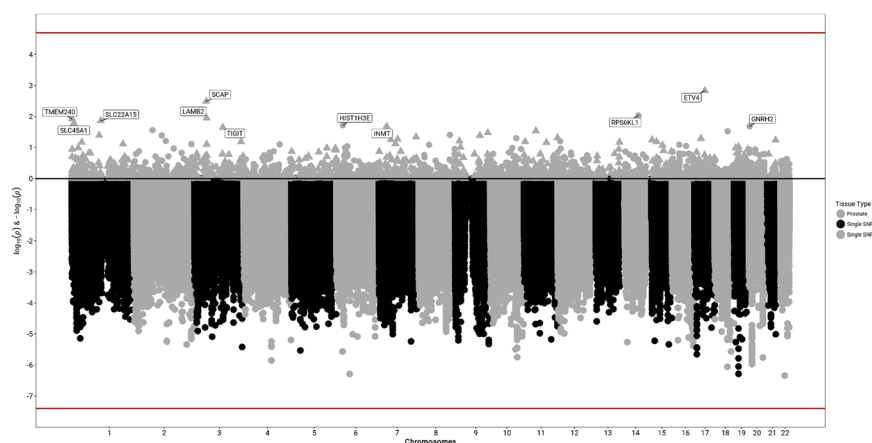
The second suggestive signal from GWAS, in the gene *GCLC* is also interesting, due in part to the localization of the SNP-based heritability on chromosome 6. Also relevant is the finding of modest association with the lead variant (rs534957) from our study which also demonstrated a weak association with prostatitis in the UK Biobank data (p-value =  $6 \times 10^{-3}$ , as viewed in the Global Biobank Engine<sup>66</sup> [<https://biobankengine.stanford.edu/>]). This suggests a consistent finding with another EHR-defined data set despite differences in case/control classification. Additionally, the identification of suggestive GPGE on chromosome 6 apart from the *GCLC* locus provides modest support for the heritability analysis, suggesting that the relevant SNPs have yet to be detected, perhaps due to a lack of power in the present studies.

One of the more biologically interesting candidates identified in this study is *GNRH2* (gonadotropin-releasing hormone 2). GPGE analysis indicated that reduced expression of this gene in the prostate was associated with increased risk of BPH (p-value = 0.021). *GNRH2* is expressed in the prostate<sup>67–70</sup> and its expression is regulated by several reproductive hormones<sup>71</sup>. Both gonadotropin releasing hormone (GnRH) antagonists and agonists have been investigated as treatments for BPH and prostate cancer<sup>12,16,72–77</sup>, however, the side effects have made many of these impractical as therapeutic options. There is currently a Phase 3 trial underway to evaluate whether a GnRH antagonist in combination with radiation can improve progression of prostate cancer. This therapeutic was previously part of a phase 2 trial for efficacy in BPH, however the trial was stopped early due to not meeting primary efficacy endpoints. This is potentially consistent with the results observed here in which reduced expression levels of *GNRH2* are associated with increased risk of BPH. A genetic variant in *GNRH2* (rs6051545) was observed to impact testosterone levels during androgen deprivation therapy to treat metastatic prostate cancer<sup>78</sup>. It has been suggested that this may lead to a negative effect of the therapy on prognosis<sup>78</sup>.

Of the 11 genes included in Table 5, more than half of them have been previously reported such that expression changes have been associated in prostate tissue, often with various stages of prostate cancer. The top result from the GPGE analysis, *ETV4* (ETS variant 4) has been previously found in studies of prostate cancer to have significantly higher relative expression in the tumor tissues than in benign samples<sup>79</sup>, as well as an association with poor prognosis<sup>80,81</sup>. We found that increased predicted expression of *ETV4* is associated with increased risk of BPH in this study (p-value = 0.0015). Laminin subunit beta 2 (*LAMB2*) has been identified as being downregulated in the transition from prostate intraepithelial neoplasia to invasive prostate cancer from differential expression analysis<sup>82</sup>. Our results suggest that increasing *LAMB2* expression is associated with increased risk of BPH.

Lead rsID	CHR:BP (hg19)	Annotation	A1	Freq. A1	All Samples			Whites Only	
					OR (95% CI)	P-value	HetPVal	P-value	OR
rs2710383	22:32950969	SYN3, intron	C	0.12	0.69 (0.55–0.83)	$4.56 \times 10^{-7}$	0.20	$4.56 \times 10^{-7}$	0.69
rs534957	6:53406351	GCLC, intron	C	0.35	1.26 (1.17–1.36)	$5.17 \times 10^{-7}$	0.29	$4.78 \times 10^{-6}$	1.24
rs4239633	19:17742469	UNC13A, intron	T	0.31	0.80 (0.71–0.88)	$5.19 \times 10^{-7}$	0.57	$6.25 \times 10^{-7}$	0.79
rs141179786	18:51188544	DCC, 126 kb downstream	A	0.98	0.40 (0.04–0.77)	$8.72 \times 10^{-7}$	0.72	$8.72 \times 10^{-7}$	0.40
rs71162163	19:17744075	UNC13A, intron	G	0.69	1.25 (1.16–1.34)	$8.95 \times 10^{-7}$	0.65	$1.29 \times 10^{-6}$	1.25
rs6078585	20:12428260	BTBD3 – SPTLC3	T	0.54	1.20 (1.12–1.27)	$1.07 \times 10^{-6}$	0.35	$1.84 \times 10^{-6}$	1.20
rs2423669	20:12431381	BTBD3 – SPTLC3	T	0.45	0.84 (0.76–0.91)	$1.19 \times 10^{-6}$	0.25	$2.46 \times 10^{-6}$	0.84
rs34163230	20:12429780	BTBD3 – SPTLC3	T	0.43	0.83 (0.75–0.90)	$1.20 \times 10^{-6}$	0.88	$1.41 \times 10^{-6}$	0.83
rs6131339	20:12427947	BTBD3 – SPTLC3	C	0.45	0.84 (0.76–0.91)	$1.30 \times 10^{-6}$	0.34	$2.28 \times 10^{-6}$	0.84
rs2423668	20:12430673	BTBD3 – SPTLC3	T	0.46	0.84 (0.76–0.91)	$1.38 \times 10^{-6}$	0.32	$2.32 \times 10^{-6}$	0.84
rs2423667	20:12430414	BTBD3 – SPTLC3	A	0.46	0.84 (0.76–0.91)	$1.38 \times 10^{-6}$	0.32	$2.32 \times 10^{-6}$	0.84
rs36067435	4:111101546	ELOVL6, intron	G	0.65	0.83 (0.75–0.91)	$1.40 \times 10^{-6}$	0.36	$2.50 \times 10^{-5}$	0.84
rs2423658	20:12418229	BTBD3 – SPTLC3	A	0.45	0.84 (0.77–0.91)	$1.41 \times 10^{-6}$	0.41	$2.24 \times 10^{-6}$	0.84
rs2423660	20:12418517	BTBD3 – SPTLC3	T	0.55	1.19 (1.12–1.27)	$1.47 \times 10^{-6}$	0.44	$2.13 \times 10^{-6}$	1.19

**Table 3.** Top GWAS Results.



**Figure 1.** Genome-wide genetically-predicted gene expression and SNP association meta-analysis results with BPH.

SCAP, which encodes SREBP cleavage-activating protein, has also been identified to show expression changes in prostate cancer<sup>83,84</sup>, and has been specifically noted to be regulated by androgens<sup>83,85</sup>. Recently, TIGIT expression has been implicated in failures of prostate cancer checkpoint inhibition<sup>86,87</sup>. Together, these results suggest that though these results did not achieve statistical significance, germline genetic associations with BPH may alter gene expression in prostate tissue, and that those genes without a presently documented role may yet be identified as important in studies of prostate gene expression implicated in disease.

There have been two recent GWAS of BPH in whites which have identified many significant and suggestive associations, though none were identified by both studies. Evaluation of these reported signals in the eMERGE data revealed modest associations at only five loci, including *BCL11A*, *TERT*, *CLPTMIL*, *GATA6*, and *FGFR2* (Table 4). It is notable, that although none of the variants reported were significant in this study, effect estimates were largely consistent in both direction and magnitude. The lack of replication may be due in part to differences across studies in disease definition (varied use of IPSS scores, prostate volume, history of transurethral resection of the prostate, etc), participant recruitment from clinical trials, community cohorts, and hospital-based populations, or differences in age. Evaluation of associated variation reported in candidate gene analyses of BPH<sup>28,30–32,34,37,38,88</sup> and an evaluation of prostate volume<sup>48</sup> did not yield any suggestive results in the present study (Supplementary Table 3).

Previous studies have shown adequate positive and negative predictive values based on electronic diagnoses (International Classification of Diseases, Ninth Revision (ICD9) codes and problem list) for BPH<sup>89</sup>, however, the phenotyping of BPH in the medical record likely reflects the presence of symptoms. Studies of care-seeking behavior with respect to BPH and LUTS have consistently shown that those seeking medical care tend to have higher symptom scores/more severe symptoms, but that reasons for not seeking treatment include diverse social and treatment concerns, even among those experiencing symptoms<sup>90–94</sup>. This suggests the possibility that some

SNP	CHR:BP (hg19)	Gene	PMID	P-value	OR	EA <sup>ab</sup>	OA <sup>ac</sup>	eMERGE P-value	OR (95% CI)	eMERGE Effect Allele
rs6537704	1:112052105	ADORA3	28656603	$1.1 \times 10^{-5}$	1.34	G	A	0.17	1.08 (0.97–1.19)	T
rs10915971	1:226508645	PARP1	28656603	$3.9 \times 10^{-5}$	1.21	A	G	0.33	0.96 (0.88–1.04)	T
rs10202096	2:28596317	FOSL2	28656603	$1.1 \times 10^{-5}$	0.81	A	G	0.2	1.05 (0.97–1.13)	A
rs2556378	2:60762502	BCL11A	30410027	$3.4 \times 10^{-12}$	1.12	T	G	0.020	1.13(1.03–1.22)	T
rs12490648	3:27782163	EOMES	28656603	$7.8 \times 10^{-6}$	1.48	G	A	0.53	1.05 (0.89–1.22)	A
rs7630221	3:37192199	LRRFIP2	28656603	$1.4 \times 10^{-5}$	1.22	A	G	0.7	0.99 (0.92–1.06)	A
rs3922844	3:38624253	SCN5A	28656603	$1.8 \times 10^{-5}$	1.23	A	G	0.37	0.97 (0.89–1.04)	T
rs2853677	5:1287194	TERT	30410027	$1.7 \times 10^{-12}$	1.09	G	A	0.021	1.09 (1.02–1.17)	G
rs381949	5:1322468	CLPTM1L	30410027	$4.9 \times 10^{-19}$	0.9	A	G	0.033	0.92 (0.85–1.00)	A
rs2676245	5:13628062	DNAH5	28656603	$7.3 \times 10^{-6}$	0.81	A	G	0.95	1.00 (0.93–1.08)	A
rs10054105	5:110909333	STARD4	30410027	$3.5 \times 10^{-12}$	0.91	G	T	0.80	0.99 (0.90–1.08)	G
rs677394	5:134607559	Csof66, H2AFY	30410027	$2.9 \times 10^{-11}$	0.88	G	C	0.50	0.96 (0.82–1.09)	G
rs200476	6:27768348	HIST1H2BL	30410027	$3.9 \times 10^{-17}$	0.88	T	A	0.076	0.92 (0.82–1.01)	T
rs17144046	10:8606014	GATA3	28656603	$8.9 \times 10^{-7}$	1.41	G	A	0.068	1.08 (1.00–1.16)	A
rs7906649	10:22310298	EBLN1	30410027	$2.1 \times 10^{-7}$	1.07	G	A	0.74	1.01 (0.93–1.10)	G
rs148678804	10:22427289	DNAJC1	30410027	$3.0 \times 10^{-14}$	1.27	A	G	0.52	0.91 (0.63–1.19)	A
rs7911634	10:56441151	PCDH15	28656603	$1.5 \times 10^{-5}$	0.67	A	C	0.87	1.01 (0.87–1.16)	T
rs4548546	10:122629579	WDR11	30410027	$2.0 \times 10^{-16}$	1.11	T	C	0.099	1.07 (0.99–1.14)	T
rs11199879	10:123045212	FGFR2	30410027	$5.7 \times 10^{-23}$	1.14	C	T	0.0065	1.13 (1.04–1.21)	C
rs72878024	11:199492	ODF3	30410027	$1.4 \times 10^{-12}$	0.85	A	G	0.43	0.94 (0.79–1.09)	A
rs428215	11:37126189	RAG2	28656603	$4.2 \times 10^{-5}$	0.76	G	A	0.8	0.98 (0.87–1.10)	T
rs10894022	11:129191124	BARX2	28656603	$2.6 \times 10^{-5}$	0.79	A	G	0.48	0.97 (0.87–1.06)	T
rs11613206	12:40609581	LRRK2	28656603	$4.9 \times 10^{-5}$	1.15	A	G	0.81	0.98 (0.81–1.15)	A
rs6538357	12:93150937	PLEKHG7	28656603	$4.9 \times 10^{-5}$	0.84	C	A	0.61	0.98 (0.91–1.05)	A
rs2555019	12:114668618	TBX5	30410027	$2.4 \times 10^{-11}$	0.93	T	C	0.66	0.98 (0.91–1.06)	T
rs8853	12:115108907	TBX3	30410027	$1.4 \times 10^{-9}$	1.07	C	T	0.47	1.03 (0.95–1.10)	C
rs1638703	13:51088356	DLEU1	30410027	$1.1 \times 10^{-13}$	1.1	C	G	0.19	1.06 (0.98–1.14)	C
rs6561599	13:51478918	RNASEH2B	30410027	$1.8 \times 10^{-7}$	0.94	C	G	0.26	0.96 (0.89–1.03)	C
rs9302078	13:96171939	CLDN10	28656603	$3.0 \times 10^{-5}$	0.83	A	G	0.67	1.02 (0.94–1.09)	A
rs7152796	14:89924520	FOXP3	28656603	$4.0 \times 10^{-5}$	1.46	A	G	0.94	0.99 (0.84–1.15)	T
rs943587	14:97636701	VRK1	28656603	0.80	1.11	G	A	0.96	1.00 (0.91–1.10)	A
rs11651052	17:36102381	HNF1B	30410027	$3.2 \times 10^{-10}$	0.93	A	G	0.77	0.99 (0.92–1.06)	A
rs9958656	18:19904144	GATA6	30410027	$4.3 \times 10^{-19}$	1.11	T	C	0.031	1.08 (1.01–1.16)	T
rs17670370	18:20034017	CTAGE1	30410027	$1.6 \times 10^{-7}$	1.07	G	T	0.31	0.96 (0.87–1.04)	G
rs11084596	19:32104979	THEG5	30410027	$2.1 \times 10^{-24}$	0.88	C	T	0.33	1.04 (0.96–1.12)	C
rs6061244	20:61041653	GATA5	30410027	$5.7 \times 10^{-8}$	0.94	C	G	0.27	0.96 (0.88–1.03)	C
rs200383755	20:61050522	GATA5	30410027	$3.2 \times 10^{-9}$	0.67	C	G	NA	NA	NA

**Table 4.** Replication of suggestive index SNPs reported in recent GWAS of BPH. <sup>a</sup>Na *et al.* (PMID 28656603) contained no indication for which was the modeled effect allele. For the purpose of this table, their A1 is listed as EA, and A2 is under OA. <sup>b</sup>Effect Allele. <sup>c</sup>Other Allele.

portion of the controls in our study may have experienced (or will experience in the future) symptoms of BPH but have not (yet) sought treatment for the condition. This is a limitation of the present study.

Based on these results, wherein BPH was shown to be heritable but no significant susceptibility loci were detected, it seems that BPH is a complex disease made up of many physiological symptoms and the genetic underpinnings of this trait are likely to consist of a multitude of variants of small effect. This makes large sample sizes crucial for detecting genetic loci associated with BPH as has been demonstrated<sup>50</sup>. In conclusion, we have shown that BPH is heritable, identified suggestive association signals, and are the first to evaluate the association between BPH and genetically-predicted gene expression in prostate.

## Methods

**Study Populations.** The eMERGE Network is a consortium of several EHR-linked biorepositories formed with the goal of developing approaches for the use of the EHR in genomic research<sup>95,96</sup>. Consortium membership has evolved over eMERGE's 11 year history, with many sites contributing data including Group Health/University of Washington, Marshfield Clinic, Mayo Clinic, Northwestern University, Vanderbilt University (Phase 1 sites), Children's Hospital of Philadelphia (CHOP), Boston Children's Hospital (BCH), Cincinnati Children's Hospital Medical Center (CCHMC), Geisinger Health System, Mount Sinai School of Medicine (sites added in Phase 2), Harvard University and Columbia University (sites added in Phase 3). The eMERGE study was approved by



Gene	Chromosome	Effect Size	P-value	Variance	Prediction-Performance R <sup>2</sup>	N SNPs used	N SNPs in model
<i>ETV4</i>	17	1.83	0.0015	0.002	0.061	3	3
<i>SCAP</i>	3	-0.42	0.0032	0.034	0.070	5	5
<i>RPS6KL1</i>	14	-0.16	0.0093	0.179	0.112	46	47
<i>LAMB2</i>	3	0.13	0.011	0.251	0.074	32	47
<i>TMEM240</i>	1	0.35	0.012	0.048	0.062	16	18
<i>SLC22A15</i>	1	-0.22	0.014	0.086	0.164	12	12
<i>SLC45A1</i>	1	-0.81	0.016	0.006	0.069	2	3
<i>HIST1H3E</i>	6	-0.13	0.019	0.244	0.116	57	61
<i>GNRH2</i>	20	-0.36	0.021	0.031	0.118	2	2
<i>INMT</i>	7	-0.13	0.021	0.214	0.059	46	48
<i>TIGIT</i>	3	-5.21	0.022	0.000	0.066	2	2

**Table 5.** Suggestive results from predicted gene expression in prostate.

the Ethical Committee/Institutional Review Board at each site (Vanderbilt University Medical Center, Group Health/University of Washington, Marshfield Clinic, Mayo Clinic, Northwestern University, Children's Hospital of Philadelphia, Boston Children's Hospital, Cincinnati Children's Hospital Medical Center, Geisinger Health System, Mount Sinai School of Medicine, Harvard University and Columbia University) and all methods were performed in accordance with the relevant guidelines and regulations. In this study of BPH, data from the eMERGE pediatric study sites (CHOP, BCH, CCHMC) were not included. Participants at all study sites provided written, informed consent, and for participants under the age of 18 years (who were not included in the analyses presented herein), informed consent was obtained from a parent and/or legal guardian.

**Phenotyping.** Among men of at least age 40, without prostate or bladder cancers (defined via ICD9 codes [233.4, 233.7 or 233.9], tumor registries [Primary site = C619] and problem lists [containing keywords e.g. "prostate cancer", "malignant tumor of the prostate", "bladder cancer", "bladder CA"], we included all cases of BPH with at least two ICD9 codes indicating a BPH diagnosis (600, 600.0, 600.0\*, 600.2, 600.2\*, 600.9, 600.9\*), in addition to either receiving medications for the treatment of BPH or 1 or more procedure (Current Procedural Terminology [CPT]) codes for BPH-related surgeries (52450, 52601, 52630, 52648, 53850, 53852). Controls were males of at least age 40, with at least 3 outpatient visits within any 2-year period after the age of 40, without prostate or bladder cancer or instances of BPH ICD9 codes, medications or BPH-related surgical CPT codes. The algorithm is available on PheKB.org (<https://phekb.org/phenotype/216>).

**Genotyping and Quality Control.** Genotyping was performed for eMERGE-1 study sites using one of two Illumina arrays across two genotyping centers. Individuals of self-identified or administratively-assigned European-descent were genotyped on the Illumina 660W-Quad, while individuals of self-identified or administratively-assigned African-descent were genotyped on the Illumina 1 M. For the majority of patients, genotyping was performed at one of two centers: the Center for Inherited Disease Research (CIDR) at Johns Hopkins University and the Center for Genotyping and Analysis at the Broad Institute as previously described<sup>95,96</sup>. Existing genotype data available for eMERGE-2 and -3 study sites included data from the Illumina 550, Illumina 610, Illumina HumanOmni Express, Illumina MultiEthnic Genotyping Array, Illumina CoreExome, and Affymetrix 6.0 arrays.

Genotype quality control (QC) was performed within each study population, and a uniform protocol was implemented. QC for all studies was performed using PLINK<sup>97</sup>, including a 95% single nucleotide polymorphism (SNP) and individual call rate threshold, removal of first-degree related individuals, sex checks, alignment of alleles to the genomic '+' strand. Visualization of ancestry by principal components analysis was performed by study using either Eigenstrat<sup>98</sup> or flashPCA<sup>99</sup>.

**Statistical Analysis.** Restricted maximum likelihood estimation as implemented in GCTA<sup>100</sup> was used to determine the proportion of phenotypic variance explained by common additive genetic variants in two cohorts, eMERGE-1 and Geisinger CoreExome. Data was filtered to include only common variants (MAF > 0.05), and samples with IBD probabilities < 0.025, as well as restricted using principal components to retain only EA samples. Disease prevalence was based on average age within cohort and set at 0.67 for eMERGE-1 and 0.63 for Geisinger CoreExome.

Genotype data was imputed from the 1000 Genomes Project haplotypes using SHAPEIT2<sup>101</sup> and IMPUTE2<sup>102</sup> by site or study and analyzed for associations separately. We used logistic regression to model BPH risk as a function of genotype, age, and principal components of ancestry with the SNPTEST software package, with subsequent meta-analysis performed using METAL<sup>103</sup>. There was no substantial genomic inflation observed, with a meta-analysis lambda of 1.017 (Supplementary Fig. 2).

To further evaluate the genetic association results in the context of gene expression, we employed the novel method S-PrediXcan<sup>51</sup>, an extension of the PrediXcan method<sup>62</sup>. PrediXcan conducts a test of association between phenotypes and gene expression levels predicted by genetic variants in a library of tissues from the Genotype-Tissue Expression (GTEx) project<sup>52,104</sup>. S-PrediXcan is a meta-analysis approach that conducts the

PrediXcan test using genotype association summary statistics, rather than performing the tests in individual-level data. We utilized covariance matrices built for prostate tissue from GTEx to annotate SNP association signals as well as to provide information about likely tissue expression patterns and relevant biological information.

## Data Availability

The datasets generated during and/or analysed during the current study are available in the eMERGE dbGAP repository (accession phs000888.v1.p1; [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000888.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000888.v1.p1)).

## References

- Lee, S. W. H., Chan, E. M. C. & Lai, Y. K. The global burden of lower urinary tract symptoms suggestive of benign prostatic hyperplasia: A systematic review and meta-analysis. *Scientific reports* **7**, 7984, <https://doi.org/10.1038/s41598-017-06628-8> (2017).
- Foster, S. A., Shorridge, E. F., DiBonaventura, M. & Viktrup, L. Predictors of self-reported benign prostatic hyperplasia in European men: analysis of the European National Health and Wellness Survey. *World journal of urology* **33**, 639–647, <https://doi.org/10.1007/s00345-014-1366-6> (2015).
- Chokkalingam, A. P. *et al.* Prevalence of BPH and lower urinary tract symptoms in West Africans. *Prostate cancer and prostatic diseases* **15**, 170–176, <https://doi.org/10.1038/pcan.2011.43> (2012).
- Foo, K. T. Pathophysiology of clinical benign prostatic hyperplasia. *Asian journal of urology* **4**, 152–157, <https://doi.org/10.1016/j.ajur.2017.06.003> (2017).
- Egan, K. B. The Epidemiology of Benign Prostatic Hyperplasia Associated with Lower Urinary Tract Symptoms: Prevalence and Incident Rates. *The Urologic clinics of North America* **43**, 289–297, <https://doi.org/10.1016/j.ucl.2016.04.001> (2016).
- Speakman, M., Kirby, R., Doyle, S. & Ioannou, C. Burden of male lower urinary tract symptoms (LUTS) suggestive of benign prostatic hyperplasia (BPH) - focus on the UK. *BJU international* **115**, 508–519, <https://doi.org/10.1111/bju.12745> (2015).
- Rodriguez-Nieves, J. A. & Macoska, J. A. Prostatic fibrosis, lower urinary tract symptoms, and BPH. *Nature reviews. Urology* **10**, 546–550, <https://doi.org/10.1038/nrurol.2013.149> (2013).
- Jung, J. H. *et al.* The association of benign prostatic hyperplasia with lower urinary tract stones in adult men: A retrospective multicenter study. *Asian journal of urology* **5**, 118–121, <https://doi.org/10.1016/j.ajur.2017.06.008> (2018).
- Huang, W. *et al.* Risk factors for bladder calculi in patients with benign prostatic hyperplasia. *Medicine* **96**, e7728, <https://doi.org/10.1097/md.00000000000007728> (2017).
- Ko, Y. H. *et al.* Clinical Implications of Residual Urine in Korean Benign Prostatic Hyperplasia (BPH) Patients: A Prognostic Factor for BPH-Related Clinical Events. *International neurourology journal* **14**, 238–244, <https://doi.org/10.5213/inj.2010.14.4.238> (2010).
- Bishr, M. *et al.* Medical management of benign prostatic hyperplasia: Results from a population-based study. *Canadian Urological Association journal = Journal de l'Association des urologues du Canada* **10**, 55–59, <https://doi.org/10.5489/cuaj.3058> (2016).
- Kumar, R., Malla, P. & Kumar, M. Advances in the design and discovery of drugs for the treatment of prostatic hyperplasia. *Expert opinion on drug discovery* **8**, 1013–1027, <https://doi.org/10.1517/17460441.2013.797960> (2013).
- Kim, E. H., Brockman, J. A. & Andriole, G. L. The use of 5-alpha reductase inhibitors in the treatment of benign prostatic hyperplasia. *Asian journal of urology* **5**, 28–32, <https://doi.org/10.1016/j.ajur.2017.11.005> (2018).
- Wallner, L. P. *et al.* 5-Alpha Reductase Inhibitors and the Risk of Prostate Cancer Mortality in Men Treated for Benign Prostatic Hyperplasia. *Mayo Clinic proceedings* **91**, 1717–1726, <https://doi.org/10.1016/j.mayocp.2016.07.023> (2016).
- Hirshburg, J. M., Kelsey, P. A., Therrien, C. A., Gavino, A. C. & Reichenberg, J. S. Adverse Effects and Safety of 5-alpha Reductase Inhibitors (Finasteride, Dutasteride): A Systematic Review. *The Journal of clinical and aesthetic dermatology* **9**, 56–62 (2016).
- Russo, A. *et al.* Latest pharmacotherapy options for benign prostatic hyperplasia. *Expert opinion on pharmacotherapy* **15**, 2319–2328, <https://doi.org/10.1517/14656566.2014.955470> (2014).
- Malaeb, B. S., Yu, X., McBean, A. M. & Elliott, S. P. National trends in surgical therapy for benign prostatic hyperplasia in the United States (2000–2008). *Urology* **79**, 1111–1116, <https://doi.org/10.1016/j.urology.2011.11.084> (2012).
- Rieken, M., Antunes-Lopes, T., Geavlete, B. & Marcelissen, T. What Is New with Sexual Side Effects After Transurethral Male Lower Urinary Tract Symptom Surgery? *European urology focus* **4**, 43–45, <https://doi.org/10.1016/j.euf.2018.05.001> (2018).
- Gacci, M. *et al.* Best practice in the management of storage symptoms in male lower urinary tract symptoms: a review of the evidence base. *Therapeutic advances in urology* **10**, 79–92, <https://doi.org/10.1177/1756287217742837> (2018).
- Moreira, A. M. *et al.* A Review of Adverse Events Related to Prostatic Artery Embolization for Treatment of Bladder Outlet Obstruction Due to BPH. *Cardiovascular and interventional radiology* **40**, 1490–1500, <https://doi.org/10.1007/s00270-017-1765-3> (2017).
- Qian, X. *et al.* Functional outcomes and complications following B-TURP versus HoLEP for the treatment of benign prostatic hyperplasia: a review of the literature and Meta-analysis. *The aging male: the official journal of the International Society for the Study of the Aging Male* **20**, 184–191, <https://doi.org/10.1080/13685538.2017.1295436> (2017).
- Afari, N. *et al.* Heritability of Lower Urinary Tract Symptoms in Men: A Twin Study. *The Journal of urology* **196**, 1486–1492, <https://doi.org/10.1016/j.juro.2016.06.018> (2016).
- Meikle, A. W., Bansal, A., Murray, D. K., Stephenson, R. A. & Middleton, R. G. Heritability of the symptoms of benign prostatic hyperplasia and the roles of age and zonal prostate volumes in twins. *Urology* **53**, 701–706 (1999).
- Partin, A. W. *et al.* Concordance rates for benign prostatic disease among twins suggest hereditary influence. *Urology* **44**, 646–650 (1994).
- Hoke, G. P. & McWilliams, G. W. Epidemiology of benign prostatic hyperplasia and comorbidities in racial and ethnic minority populations. *The American journal of medicine* **121**, S3–10, <https://doi.org/10.1016/j.amjmed.2008.05.021> (2008).
- Colon, I. & Payne, R. E. Benign prostatic hyperplasia and lower urinary tract symptoms in African Americans and Latinos: treatment in the context of common comorbidities. *The American journal of medicine* **121**, S18–26, <https://doi.org/10.1016/j.amjmed.2008.05.023> (2008).
- Lee, C. L., Lee, J., Na, Y. G. & Song, K. H. Combined effect of polymorphisms in type III 5-alpha reductase and androgen receptor gene with the risk of benign prostatic hyperplasia in Korea. *Journal of exercise rehabilitation* **12**, 504–508, <https://doi.org/10.12965/jer.1632802.401> (2016).
- Kim, S. K. *et al.* Association between polymorphisms of estrogen receptor 2 and benign prostatic hyperplasia. *Experimental and therapeutic medicine* **10**, 1990–1994, <https://doi.org/10.3892/etm.2015.2755> (2015).
- Choubey, V. K. *et al.* SRD5A2 gene polymorphisms and the risk of benign prostatic hyperplasia but not prostate cancer. *Asian Pacific journal of cancer prevention: APJCP* **16**, 1033–1036 (2015).
- Ruan, L. *et al.* Association between single nucleotide polymorphism of vitamin D receptor gene FokI polymorphism and clinical progress of benign prostatic hyperplasia. *TheScientificWorldJournal* **2015**, 235895, <https://doi.org/10.1155/2015/235895> (2015).
- Winchester, D. *et al.* SPINK1 Promoter Variants Are Associated with Prostate Cancer Predisposing Alterations in Benign Prostatic Hyperplasia Patients. *Anticancer research* **35**, 3811–3819 (2015).

32. Ban, J. Y. & Yoo, K. H. Promoter Polymorphism (rs12770170, -184C/T) of Microseminoprotein, Beta as a Risk Factor for Benign Prostatic Hyperplasia in Korean Population. *International neurourology journal* **18**, 63–67, <https://doi.org/10.5213/inj.2014.18.2.63> (2014).
33. Kumar, V. *et al.* Association of CYP1A1, CYP1B1 and CYP17 gene polymorphisms and organochlorine pesticides with benign prostatic hyperplasia. *Chemosphere* **108**, 40–45, <https://doi.org/10.1016/j.chemosphere.2014.02.081> (2014).
34. Seok, H., Yoo, K. H., Kim, Y. O. & Chung, J. H. Association of a Missense ALDH2 Single Nucleotide Polymorphism (Glu504Lys) With Benign Prostate Hyperplasia in a Korean Population. *International neurourology journal* **17**, 168–173, <https://doi.org/10.5213/inj.2013.17.4.168> (2013).
35. Karatzas, A. *et al.* Lack of association between the UDP-glucuronosyltransferase 1A1 (UGT1A1) gene polymorphism and the risk of benign prostatic hyperplasia in Caucasian men. *Molecular biology reports* **40**, 6665–6669, <https://doi.org/10.1007/s11033-013-2781-2> (2013).
36. Zambra, F. M., Biolchi, V., Brum, I. S. & Chies, J. A. CCR2 and CCR5 genes polymorphisms in benign prostatic hyperplasia and prostate cancer. *Human immunology* **74**, 1003–1008, <https://doi.org/10.1016/j.humimm.2013.04.031> (2013).
37. Qi, J. *et al.* Genetic variants in 2q31 and 5p15 are associated with aggressive benign prostatic hyperplasia in a Chinese population. *Prostate* **73**, 1182–1190, <https://doi.org/10.1002/pros.22666> (2013).
38. Jiao, Y. *et al.* LILRA3 is associated with benign prostatic hyperplasia risk in a Chinese Population. *International journal of molecular sciences* **14**, 8832–8840, <https://doi.org/10.3390/ijms14058832> (2013).
39. Biolchi, V. *et al.* Androgen receptor GGC polymorphism and testosterone levels associated with high risk of prostate cancer and benign prostatic hyperplasia. *Molecular biology reports* **40**, 2749–2756, <https://doi.org/10.1007/s11033-012-2293-5> (2013).
40. Choubey, V. K. *et al.* Null genotypes at the GSTM1 and GSTT1 genes and the risk of benign prostatic hyperplasia: a case-control study and a meta-analysis. *Prostate* **73**, 146–152, <https://doi.org/10.1002/pros.22549> (2013).
41. Biolchi, V., Silva Neto, B., Koff, W. & Brum, I. S. Androgen receptor CAG polymorphism and the risk of benign prostatic hyperplasia in a Brazilian population. *International braz j urol: official journal of the Brazilian Society of Urology* **38**, 373–379 (2012).
42. Izmirli, M., Arıkan, B., Bayazit, Y. & Alptekin, D. Associations of polymorphisms in HPC2/ELAC2 and SRD5A2 genes with benign prostate hyperplasia in Turkish men. *Asian Pacific journal of cancer prevention: APJCP* **12**, 731–733 (2011).
43. Konwar, R. *et al.* Glutathione S-transferase (GST) gene variants and risk of benign prostatic hyperplasia: a report in a North Indian population. *Asian Pacific journal of cancer prevention: APJCP* **11**, 1067–1072 (2010).
44. Mittal, R. D., Kesarwani, P., Singh, R., Ahirwar, D. & Mandhani, A. GSTM1, GSTM3 and GSTT1 gene variants and risk of benign prostate hyperplasia in North India. *Disease markers* **26**, 85–91, <https://doi.org/10.3233/dma-2009-0611> (2009).
45. Ma, Z. *et al.* Polymorphisms of fibroblast growth factor receptor 4 have association with the development of prostate cancer and benign prostatic hyperplasia and the progression of prostate cancer in a Japanese population. *International journal of cancer. Journal international du cancer* **123**, 2574–2579, <https://doi.org/10.1002/ijc.23578> (2008).
46. Das, K. *et al.* Shorter CAG repeats in androgen receptor and non-GG genotypes in prostate-specific antigen loci are associated with decreased risk of benign prostatic hyperplasia and prostate cancer. *Cancer Lett* **268**, 340–347, <https://doi.org/10.1016/j.canlet.2008.04.009> (2008).
47. Cartwright, R. *et al.* Systematic review and meta-analysis of candidate gene association studies of lower urinary tract symptoms in men. *European urology* **66**, 752–768, <https://doi.org/10.1016/j.eururo.2014.01.007> (2014).
48. Giri, A., Edwards, T. L., Motley, S. S., Byerly, S. H. & Fowke, J. H. Genetic Determinants of Metabolism and Benign Prostate Enlargement: Associations with Prostate Volume. *PLoS One* **10**, e0132028, <https://doi.org/10.1371/journal.pone.0132028> (2015).
49. Na, R. *et al.* A genetic variant near GATA3 implicated in inherited susceptibility and etiology of benign prostatic hyperplasia (BPH) and lower urinary tract symptoms (LUTS). *Prostate* **77**, 1213–1220, <https://doi.org/10.1002/pros.23380> (2017).
50. Gudmundsson, J. *et al.* Genome-wide associations for benign prostatic hyperplasia reveal a genetic correlation with serum levels of PSA. *Nature communications* **9**, 4568, <https://doi.org/10.1038/s41467-018-06920-9> (2018).
51. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature communications* **9**, 1825, <https://doi.org/10.1038/s41467-018-03621-1> (2018).
52. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (New York, N.Y.)* **348**, 648–660, <https://doi.org/10.1126/science.1262110> (2015).
53. Kao, H. T. *et al.* A third member of the synapsin gene family. *Proc Natl Acad Sci USA* **95**, 4667–4672 (1998).
54. Hosaka, M. & Sudhof, T. C. Synapsin III, a novel synapsin with an unusual regulation by Ca<sup>2+</sup>. *J Biol Chem* **273**, 13371–13374 (1998).
55. Fritsche, L. G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat Genet* **45**(433–439), 439e431–432, <https://doi.org/10.1038/ng.2578> (2013).
56. Neale, B. M. *et al.* Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc Natl Acad Sci USA* **107**, 7395–7400, <https://doi.org/10.1073/pnas.0912019107> (2010).
57. Chen, W. *et al.* Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc Natl Acad Sci USA* **107**, 7401–7406, <https://doi.org/10.1073/pnas.0912702107> (2010).
58. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173–1186, <https://doi.org/10.1038/ng.3097> (2014).
59. He, M. *et al.* Meta-analysis of genome-wide association studies of adult height in East Asians identifies 17 novel loci. *Hum Mol Genet* **24**, 1791–1800, <https://doi.org/10.1093/hmg/ddu583> (2015).
60. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838, <https://doi.org/10.1038/nature09410> (2010).
61. Charles, B. A. *et al.* A genome-wide association study of serum uric acid in African Americans. *BMC medical genomics* **4**, 17, <https://doi.org/10.1186/1755-8794-4-17> (2011).
62. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091–1098, <https://doi.org/10.1038/ng.3367> (2015).
63. van Es, M. A. *et al.* Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet* **41**, 1083–1087, <https://doi.org/10.1038/ng.442> (2009).
64. Benyamin, B. *et al.* Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nature communications* **8**, 611, <https://doi.org/10.1038/s41467-017-00471-1> (2017).
65. van Rheenen, W. *et al.* Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* **48**, 1043–1048, <https://doi.org/10.1038/ng.3622> (2016).
66. McInnes, G. *et al.* Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *bioRxiv*, <https://doi.org/10.1101/304188> (2018).
67. Desaulniers, A. T., Cederberg, R. A., Lents, C. A. & White, B. R. Expression and Role of Gonadotropin-Releasing Hormone 2 and Its Receptor in Mammals. *Frontiers in endocrinology* **8**, 269, <https://doi.org/10.3389/fendo.2017.00269> (2017).
68. Millar, R. *et al.* A novel mammalian receptor for the evolutionarily conserved type II GnRH. *Proc Natl Acad Sci USA* **98**, 9636–9641, <https://doi.org/10.1073/pnas.141048498> (2001).
69. Neill, J. D., Duck, L. W., Sellers, J. C. & Musgrove, L. C. A gonadotropin-releasing hormone (GnRH) receptor specific for GnRH II in primates. *Biochemical and biophysical research communications* **282**, 1012–1018, <https://doi.org/10.1006/bbrc.2001.4678> (2001).



70. White, R. B., Eisen, J. A., Kasten, T. L. & Fernald, R. D. Second gene for gonadotropin-releasing hormone in humans. *Proc Natl Acad Sci USA* **95**, 305–309 (1998).
71. Darby, S. *et al.* Expression of GnRH type II is regulated by the androgen receptor in prostate cancer. *Endocrine-related cancer* **14**, 613–624, <https://doi.org/10.1677/erc-07-0041> (2007).
72. Sakai, M., Elhilali, M. & Papadopoulos, V. The GnRH Antagonist Degarelix Directly Inhibits Benign Prostate Hyperplasia Cell Growth. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme* **47**, 925–931, <https://doi.org/10.1055/s-0035-1555899> (2015).
73. Raja, A., Hori, S. & Armitage, J. N. Hormonal manipulation of lower urinary tract symptoms secondary to benign prostatic obstruction. *Indian journal of urology: IJU: journal of the Urological Society of India* **30**, 189–193, <https://doi.org/10.4103/0970-1591.126904> (2014).
74. Rick, F. G. *et al.* Hormonal manipulation of benign prostatic hyperplasia. *Current opinion in urology* **23**, 17–24, <https://doi.org/10.1097/MOU.0b013e32835abd18> (2013).
75. Polisca, A. *et al.* Clinical efficacy of the GnRH agonist (deslorelin) in dogs affected by benign prostatic hyperplasia and evaluation of prostatic blood flow by Doppler ultrasound. *Reproduction in domestic animals = Zuchthygiene* **48**, 673–680, <https://doi.org/10.1111/rda.12143> (2013).
76. Soler, R. *et al.* Future direction in pharmacotherapy for non-neurogenic male lower urinary tract symptoms. *European urology* **64**, 610–621, <https://doi.org/10.1016/j.eururo.2013.04.042> (2013).
77. Lepor, H. The role of gonadotropin-releasing hormone antagonists for the treatment of benign prostatic hyperplasia. *Reviews in urology* **8**, 183–189 (2006).
78. Shiota, M. *et al.* The Association of Polymorphisms in the Gene Encoding Gonadotropin-Releasing Hormone with Serum Testosterone Level during Androgen Deprivation Therapy and Prognosis of Metastatic Prostate Cancer. *The Journal of urology*, <https://doi.org/10.1016/j.juro.2017.09.076> (2017).
79. Cavazzola, L. R. *et al.* Relative mRNA expression of prostate-derived E-twenty-six factor and E-twenty-six variant 4 transcription factors, and of uridine phosphorylase-1 and thymidine phosphorylase enzymes, in benign and malignant prostatic tissue. *Oncology letters* **9**, 2886–2894, <https://doi.org/10.3892/ol.2015.3093> (2015).
80. Qi, M. *et al.* Overexpression of ETV4 is associated with poor prognosis in prostate cancer: involvement of uPA/uPAR and MMPs. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **36**, 3565–3572, <https://doi.org/10.1007/s13277-014-2993-7> (2015).
81. Mesquita, D. *et al.* Specific and redundant activities of ETV1 and ETV4 in prostate cancer aggressiveness revealed by co-overexpression cellular contexts. *Oncotarget* **6**, 5217–5236, <https://doi.org/10.18632/oncotarget.2847> (2015).
82. Ashida, S. *et al.* Molecular features of the transition from prostatic intraepithelial neoplasia (PIN) to prostate cancer: genome-wide gene-expression profiles of prostate cancers and PINs. *Cancer research* **64**, 5963–5972, <https://doi.org/10.1158/0008-5472.can-04-0020> (2004).
83. Heemers, H. *et al.* Androgens stimulate lipogenic gene expression in prostate cancer cells by activation of the sterol regulatory element-binding protein cleavage activating protein/sterol regulatory element-binding protein pathway. *Molecular endocrinology (Baltimore, Md.)* **15**, 1817–1828, <https://doi.org/10.1210/mend.15.10.0703> (2001).
84. Prabhu, A. V., Krycer, J. R. & Brown, A. J. Overexpression of a key regulator of lipid homeostasis, Scap, promotes respiration in prostate cancer cells. *FEBS letters* **587**, 983–988, <https://doi.org/10.1016/j.febslet.2013.02.040> (2013).
85. Heemers, H. *et al.* Identification of an androgen response element in intron 8 of the sterol regulatory element-binding protein cleavage-activating protein gene allowing direct regulation by the androgen receptor. *J Biol Chem* **279**, 30880–30887, <https://doi.org/10.1074/jbc.M401615200> (2004).
86. Papanicolaou-Sengos, A. *et al.* Identification of targets for prostate cancer immunotherapy. *Prostate* **79**, 498–505, <https://doi.org/10.1002/pros.23756> (2019).
87. Nava Rodrigues, D. *et al.* Immunogenomic analyses associate immunological alterations with mismatch repair defects in prostate cancer. *J Clin Invest* **128**, 4441–4453, <https://doi.org/10.1172/jci121924> (2018).
88. Gu, X. *et al.* SRD5A1 and SRD5A2 are associated with treatment for benign prostatic hyperplasia with the combination of 5 $\alpha$ -reductase inhibitors and alpha-adrenergic receptor antagonists. *The Journal of urology* **190**, 615–619, <https://doi.org/10.1016/j.juro.2013.03.024> (2013).
89. Szeto, H. C., Coleman, R. K., Gholami, P., Hoffman, B. B. & Goldstein, M. K. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic. *The American journal of managed care* **8**, 37–43 (2002).
90. Kok, E. T. *et al.* Simple case definition of clinical benign prostatic hyperplasia, based on International Prostate Symptom Score, predicts general practitioner consultation rates. *Urology* **68**, 784–789, <https://doi.org/10.1016/j.urology.2006.04.008> (2006).
91. Roberts, R. O. *et al.* Natural history of prostatism: worry and embarrassment from urinary symptoms and health care-seeking behavior. *Urology* **43**, 621–628 (1994).
92. Jacobsen, S. J. *et al.* Natural history of prostatism: factors associated with discordance between frequency and bother of urinary symptoms. *Urology* **42**, 663–671 (1993).
93. Sarma, A. V., Wallner, L., Jacobsen, S. J., Dunn, R. L. & Wei, J. T. Health seeking behavior for lower urinary tract symptoms in black men. *The Journal of urology* **180**, 227–232, <https://doi.org/10.1016/j.juro.2008.03.032> (2008).
94. Wolters, R., Wensing, M., van Weel, C., van der Wilt, G. J. & Grob, R. P. Lower urinary tract symptoms: social influence is more important than symptoms in seeking medical care. *BJU international* **90**, 655–661 (2002).
95. Gottesman, O. *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genetics in medicine: official journal of the American College of Medical Genetics* **15**, 761–771, <https://doi.org/10.1038/gim.2013.72> (2013).
96. McCarty, C. A. *et al.* The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC medical genomics* **4**, 13, <https://doi.org/10.1186/1755-8794-4-13> (2011).
97. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575, <https://doi.org/10.1086/519795> (2007).
98. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909, <https://doi.org/10.1038/ng1847> (2006).
99. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766, <https://doi.org/10.1371/journal.pone.0093766> (2014).
100. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76–82, <https://doi.org/10.1016/j.ajhg.2010.11.011> (2011).
101. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181, <https://doi.org/10.1038/nmeth.1785> (2012).
102. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529, <https://doi.org/10.1371/journal.pgen.1000529> (2009).
103. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191, <https://doi.org/10.1093/bioinformatics/btq340> (2010).
104. Mele, M. *et al.* Human genomics. *The human transcriptome across tissues and individuals. Science (New York, N.Y.)* **348**, 660–665, <https://doi.org/10.1126/science.aaa0355> (2015).

## Acknowledgements

J.N.H. was supported by T32CA160056 (P.I. X-O Shu). The eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG8657 (Kaiser Washington/University of Washington); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG8676 (Partners Healthcare/Broad Institute); and U01HG8664 (Baylor College of Medicine). Some of the dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding, the 1S10RR025141-01 instrumentation award, and by the CTSA grant UL1TR000445 from NCATS/NIH. The authors would also like to acknowledge NCATS grant UL1TR002373 (P.I. Marc Drezner).

## Author Contributions

Study design/conception: D.M.R., J.C.D., T.L.E.; Data collection/Sample recruitment: K.M.B., D.M.R., J.C.D., M.B., J.G.L., G.J., N.S., G.H., D.C., A.G., P.D., P.L.P., P.A.M.S., H.H.; Data analysis: J.N.H., E.S.T., T.L.E.; Methods/Algorithm Development: J.C.D., R.C., S.S.; Manuscript writing: J.N.H., D.R.V.E., T.L.E.; Manuscript editing/comments: M.D.R., S.S.V., D.M.R., J.C.D., J.G.L.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-42427-z>.

**Competing Interests:** The authors declare no competing interests.

**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