Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation

Thomas J Hoffmann1,2, Georg B Ehret3,4, Priyanka Nandakumar3, Dilrini Ranatunga5, Catherine Schaefer5, Pui-Yan Kwok2, Carlos Iribarren5, Aravinda Chakravarti3 & Neil Risch1,2,5

Longitudinal electronic health records on 99,785 Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort individuals provided 1,342,814 systolic and diastolic blood pressure measurements for a genome-wide association study on long-term average systolic, diastolic, and pulse pressure. We identified 39 new loci among 75 genome-wide significant loci \( (P \leq 5 \times 10^{-8}) \), with most replicating in the combined International Consortium for Blood Pressure (ICBP; \( n = 69,396 \)) and UK Biobank (UKB; \( n = 152,081 \)) studies. Combining GERA with ICBP yielded 36 additional new loci, with most replicating in UKB. Combining all three studies \( (n = 321,262) \) yielded 241 additional genome-wide significant loci, although no replication sample was available for these. All associated loci explained 2.9\%, 2.5\%, and 3.1\% of variation in systolic, diastolic, and pulse pressure, respectively, in GERA non-Hispanic whites. Using multiple blood pressure measurements in GERA doubled the variance explained. A normalized risk score was associated with time to onset of hypertension (hazards ratio = 1.18, \( P = 8.2 \times 10^{-45} \)). Expression quantitative trait locus analysis of blood pressure loci showed enrichment in aorta and tibial artery.

Blood pressure is an important cardiovascular risk factor\(^1\), with estimated 30–50\% heritability\(^2,3\). Over the past several years, genome-wide association studies (GWAS) have identified 85 SNPs associated with blood pressure\(^4–22\). However, the heritability explained remains less than that for other quantitative cardiovascular traits, for example, lipids\(^23\). Three strategies to identify additional variants are the use of larger sample sizes, more precise measurements, and more extensive imputation panels. Thus far, all large studies have used measurements from research protocols rather than clinical records. There is little doubt that the phenotype seen in observational research or randomized trials is similar to that in a clinical encounter, but clinical measures may be influenced by somewhat different circumstances and measurements may be obtained under a less stringent protocol\(^24\). However, studies using clinical measurements from electronic health records (EHRs) permit not only very large sample sizes but also a long-term average of multiple independent clinical measurements from many different clinical visits, yielding reduced phenotype variance (as shown by simulation and experimental data)\(^7\). We therefore reasoned that a large-sample blood pressure GWAS with longitudinal EHR-based measures would provide improved statistical power and understanding of blood pressure genomic architecture, which we show theoretically (Online Methods) and through data application.

RESULTS
GERA cohort

We conducted primary discovery in the GERA cohort \( (n = 99,785 \) for this study), which is composed of self-described non-Hispanic whites \( (81\%; 80,792) \), Latinos \( (8\%; 8,231) \), East Asians \( (7\%; 7,243) \), African Americans \( (3\%; 3,058) \), and South Asians \( (1\%; 461) \) (Table 1). GERA is part of the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), whose participants are members of an integrated health care delivery system. The average follow-up time was 4 years, beginning at age 60.9 years, leading to high prevalence of hypertension and antihypertensive therapy. Figure 1 describes EHR extraction and study design (Online Methods). Multiple blood pressure measurements \( (1,342,814 \) in total) were available for many participants: 46.4\% had at least one untreated measurement and 62.6\% had at least one treated measurement. We included all individuals who had at least one (untreated or treated) blood pressure measurement. The availability of multiple measurements enabled use of a long-term average to increase accuracy\(^7\). There were differences in anthropometric and blood pressure values at the first visit among the ancestry groups (Table 1): African Americans and Latinos had the highest body mass index (BMI) values, whereas South Asians had the lowest values, although this group was on average the youngest. Untreated systolic blood pressure (SBP) and diastolic blood pressure

1Department of Epidemiology and Biostatistics, University of California at San Francisco, San Francisco, California, USA. 2Institute for Human Genetics, University of California at San Francisco, San Francisco, California, USA. 3Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Baltimore, Maryland, USA. 4Cardiology, Department of Specialties of Internal Medicine, University of Geneva, Geneva, Switzerland. 5Kaiser Permanente Northern California Division of Research, Oakland, California, USA. Correspondence should be addressed to T.J.H. (thomas.hoffmann@ucsf.edu) or N.R. (neil.risch@ucsf.edu).

Received 8 April; accepted 11 October; published online 14 November 2016; doi:10.1038/ng.3715
Table 1 Characteristics of the GERA cohort

|                      | Non-Hispanic whites | Latinos | East Asians | African Americans | South Asians |
|----------------------|---------------------|---------|-------------|------------------|-------------|
| n (mean) (%)         | 80.792 81.0%        | 8,231 8.2% | 7,243 7.3%  | 3,058 3.1%      | 461 0.5%    |
| n female (%)         | 46,771 57.9%       | 4,960 60.2% | 4,190 57.9% | 1,819 59.5%    | 184 39.9%   |
| Average BMI (at first measure, mm Hg) | 13.2 | 11.1 | 15.6 | 10.7 | 7.5 |
| Age (at first measure, years) | 56.8 14.6 | 53.9 14.6 | 54.0 15.5 | 51,221 129.2 | 4,261 125.2 |
| BMI (at first measure) | 25.1 | 15.9 | 17.6 | 13.7 | 7.5 |
| DBP (at first measure, mm Hg) | 4.5 | 4.0 | 4.5 | 4.5 | 6.5 |
| Untreated female mean (SE) | 3.9 | 3.0 | 3.0 | 3.0 | 3.8 |
| Untreated male mean (SE) | 7.7 | 6.6 | 6.6 | 6.6 | 6.6 |
| DBP (at first measure, mm Hg) | 128.2 15.5 | 128.2 15.5 | 127.4 14.8 | 130.4 15.4 | 125.0 16.2 |
| Untreated male mean (SE) | 125.2 13.0 | 124.7 13.0 | 122.4 12.8 | 127.0 14.5 | 120.4 12.5 |
| Untreated female mean (SE) | 129.7 15.6 | 129.2 15.9 | 128.1 15.5 | 130.9 15.9 | 123.1 14.8 |
| PP (at first measure, mm Hg) | 7.5 | 7.5 | 7.5 | 7.5 | 7.5 |
| Age (at first measure, years) | 63.9 12.1 | 59.1 13.7 | 59.2 13.6 | 61.5 11.7 | 54.8 14.0 |
| BMI (at first measure) | 28.0 6.0 | 28.6 6.3 | 24.5 4.5 | 30.8 6.9 | 25.1 4.2 |
| SBP (at first measure, mm Hg) | 128.2 15.5 | 128.2 15.5 | 127.4 14.8 | 130.4 15.4 | 125.0 16.2 |
| Untreated male mean (SE) | 125.2 13.0 | 124.7 13.0 | 122.4 12.8 | 127.0 14.5 | 120.4 12.5 |
| Untreated female mean (SE) | 129.7 15.6 | 129.2 15.9 | 128.1 15.5 | 130.9 15.9 | 123.1 14.8 |
| Untreated male mean (SE) | 121.2 14.4 | 118.7 14.1 | 117.3 14.3 | 122.5 13.4 | 113.8 13.3 |
| PP (at first measure, mm Hg) | 73.9 10.1 | 75.0 10.3 | 74.8 10.2 | 76.9 10.1 | 73.2 10.5 |
| Untreated male mean (SE) | 75.3 8.8 | 75.5 9.0 | 75.0 9.1 | 76.8 9.2 | 73.2 8.9 |
| Untreated female mean (SE) | 73.8 9.9 | 74.5 10.1 | 74.4 10.5 | 75.9 10.4 | 73.3 9.8 |
| Untreated male mean (SE) | 72.4 9.2 | 72.0 9.3 | 71.2 9.5 | 75.0 8.8 | 69.7 8.8 |
| Untreated female mean (SE) | 54.4 12.5 | 53.2 12.3 | 52.6 12.4 | 54.0 12.5 | 51.8 11.8 |
| Untreated male mean (SE) | 49.9 10.3 | 49.2 10.2 | 47.4 9.6 | 50.1 10.9 | 47.2 9.5 |
| Untreated female mean (SE) | 55.9 13.8 | 54.8 13.6 | 53.7 13.2 | 55.0 13.7 | 49.8 11.4 |
| Untreated female mean (SE) | 54.0 17.4 | 51.8 17.3 | 52.1 18.0 | 51.6 17.3 | 43.0 14.6 |

(DBP) were highest in African Americans, followed by non-Hispanic whites; South Asians had lower values (Fig. 2). Untreated blood pressure measures were higher in males than in females across the groups, as also found previously. To further investigate covariate effects, we assessed the effects of age, sex, BMI, and genetic ancestry on SBP, DBP, and pulse pressure (PP) within each ancestry group (Supplementary Table 1). Age and age² accounted for substantial variation in SBP, as expected, with the
Of the 75 identified loci, 36 replicated previous GWAS findings. Of the remaining 39 new loci (Fig. 3), 25 were strictly replicated ($P \leq 0.00067$, Bonferroni correction for $75 = 39 + 36$ SNPs; Online Methods) in 221,477 individuals from ICBP (HapMap summary statistics augmented with 1000 Genomes Project data; Online Methods and Supplementary Fig. 5)\(^{17}\) and UKB (imputed additionally using UK10K\(^{27}\)). Of the 14 remaining loci, 8 had suggestive significance ($P \leq 0.01$) and 1 X-chromosome SNP was unavailable for replication. All SNPs with at least suggestive significance ($P \leq 0.05$) had effects in the same direction as in GERA and had no significant heterogeneity among the GERA ancestry groups or between GERA, ICBP, and/or UKB (Fig. 3 and Supplementary Table 3), giving further credibility to the notion that these loci are also true positive findings. Of note, ICBP alone poorly replicated new SNPs (only three SNPs met significance with Bonferroni correction in ICBP alone), although the SNPs were highly enriched for small $P$ values. These results emphasize the importance of large replication cohorts.

Expanding our discovery to a meta-analysis of GERA and ICBP also did not indicate significant inflation (average $\lambda = 1.042$; Supplementary Table 2); this $\lambda$ value is slightly smaller than for GERA alone, likely owing to the slightly conservative nature of extending the ICBP summary statistics (Online Methods). Thirty-six additional new loci reached genome-wide significance for at least one blood pressure phenotype. Using 152,081 individuals from UKB for replication, 22 loci replicated at $P \leq 0.00067$ (Bonferroni correction for 75 SNPs), 7 were suggestive with $P < 0.01$, and 2 reached nominal significance ($P < 0.05$). As before, all SNPs with at least nominal significance ($P < 0.05$) had the same direction of effect in UKB, arguing for a low rate of false positive findings (Fig. 4 and Supplementary Table 3). We did not detect significant heterogeneity for any lead SNP.

Finally, to maximize discovery power, we combined all three studies (GERA, ICBP, and UKB; $n = 321,262$). Our genome-wide meta-analyses of SBP, DBP, and PP had $\lambda$ values of 1.069, 1.076, and 1.076, respectively. We identified 241 additional new genome-wide significant loci (Supplementary Fig. 6 and Supplementary Table 3), although replication was not possible. Only rs139491786 showed evidence of heterogeneity ($P^2 = 88, P = 1.5 \times 10^{-5}$).

**Conditional analysis**

We first searched for additional genome-wide significant SNPs within a 1-Mb window ($\pm 0.5$ Mb with respect to the lead SNP) involving each previously described or new locus in GERA, testing for replication in UKB. We first identified an additional new SNP, rs1322640, 129 kb from rs13197550 (lead GERA SNP) that replicated in UKB ($P = 8.3 \times 10^{-6}$; Table 2 and Supplementary Table 5). We next identified a new indel (chromosome 20, 10,573,001 (Build 37)) located 396 kb from rs2104574 (lead GERA SNP) that replicated in UKB ($P = 0.012$; Table 2 and Supplementary Table 5).

We further combined GERA and UKB in a discovery conditional meta-analysis, identifying four additional independent signals (Table 2 and Supplementary Table 5). No replication was possible for these.

**Replication of previous GWAS results**

We also investigated replication of previously described blood pressure loci in GERA (Supplementary Table 6, which also reports the GERA lead SNP when it differs from the previously described lead SNP at the locus\(^{4-22}\). Of the 85 previously described lead SNPs (or a proxy with $r^2 = 1.00$ for one SNP), 62.4% (53/85) were significantly associated with at least one GERA blood pressure phenotype.
| rs9729719 A/G   | rs35410524 T/C   | rs12050260 T/C   |
|-----------------|-----------------|-----------------|
| MTF1            | FUT9            | HOMEOZ          |
|                 | UFLT            | PPPRH3E         |
| rs783621 A/G    | rs13192978 T/A  | rs2244643 C/A   |
| HIVEP           | ESR1            | FBLN9           |
|                 |                 |                 |
| rs147696085 G/A | rs1322639 A/G   | rs1475130 C/T   |
| FAF1            | SMDC2           | CYP4A1          |
|                 | THBS2           | EML7            |
| rs60199046 A/G  | rs937951 A/G    | rs62011052 C/T  |
| HSD52           | AIAA1465        | ADAMTS7         |
| MIR4117         |                 | MORN4L          |
| rs2761436 T/C   | rs34872471 C/G  | rs12596053 G/A  |
| CRY1            | TCF7L2          |                 |
| CD46            |                 |                 |
| rs642942 G/A    | rs7927515 A/G   | rs750044 A/G    |
| SDCCAG8         | PRKIR           | CDH13           |
| MIR4677         | C11orf36        |                 |
| rs791058 T/C    | rs6288125 C/A   |                 |
| CDDC149         | NOX4            |                 |
|                 |                 |                 |
| rs1200247 C/G   | rs1261744 C/G   | rs12605620 G/T  |
| FN5             | CEP162          | SYT4            |
|                 |                 | SETBP1          |
| rs76217164 T/A  | rs7977589 T/C   | rs2193635 T/C   |
| C3orf58         | FAM1869         | SLC14A2         |
| EPHA3           |                 |                 |
| rs7530534 T/A   | rs10741750 A/G  | rs10427021 T/C  |
| RYK             | COX1A           | INSR            |
|                 |                 |                 |
| rs2178452 G/A   | rs14311816 T/A  | rs10418303 G/C  |
| SCARNA7         | SARNR           | NOTCH3          |
| ARL14           |                 |                 |
| rs258494 C/G    | rs76765025 C/T  | rs8105753 A/G   |
| POC5            | LOC144486       | TSHZ3           |
| SIV20           | MIR5109         | THREG           |
| rs4475259 G/A   | rs3418562 C/T   | rs141216868 C/T |
| KCNQ2           | SLC7A1          | ACTRT1          |
| TRIM36          |                 | SMARCA1         |

**Figure 3** New blood pressure loci detected in GERA and tested for replication in ICBP + UKB meta-analysis. SNPs rs76217164 and rs143118162 failed to impute in ICBP (owing to low allele frequency), and rs141216868 was on the X chromosome and not available in ICBP or UKB. We used an additive model. The effect allele is the allele to the left (for example, A in A/G). Effect sizes are indicated in mm Hg. Each line represents the effect size and its 95% confidence interval for each measure and group, with the text beside each line denoting the group tested: G, GERA (n = 99,785); IU, meta-analysis of ICBP and UKB (n = 221,477); and GIU, meta-analysis of GERA, ICBP, and UKB (n = 321,262). The color of each line corresponds to the statistical significance of the test: red, P ≤ 1 × 10⁻³; orange, 1 × 10⁻³ < P ≤ 5 × 10⁻²; green, 5 × 10⁻² < P ≤ 0.00067 (Bonferroni correction for 39 + 36 = 75 SNPs); blue, 0.00066 < P ≤ 0.05; black, P > 0.05. If a SNP is in a gene, that gene is given; otherwise, the two surrounding genes are given.
at $P < 0.00059$ (with Bonferroni adjustment for 85 tests) and had the same direction of effect; 78.8% (67/85) were nominally significant; and 95.3% (81/85) had effects in the same direction. Replication was stronger in UKB, with 77.6% (66/85) replicating at Bonferroni significance, 89.4% (76/85) replicating at nominal significance, and 96.5% (82/85) having effects in the same direction. Replication was

Figure 4 New blood pressure loci identified in the GERA + ICBP meta-analysis and tested for replication in UKB. We used an additive model. The effect allele is the allele to the left (for example, A in A/G). Effect sizes are indicated in mm Hg. Each line represents the effect size and its 95% confidence interval for each measure and group, with the test beside each line denoting the group tested: GI, meta-analysis of GERA and ICBP ($n = 169,181$); U, UKB ($n = 152,081$); and GIU, meta-analysis of GERA, ICBP, and UKB ($n = 321,262$). The color of each line corresponds to the statistical significance, 89.4% (76/85) replicating at nominal significance, and 96.5% (82/85) having effects in the same direction. Replication was

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
### Table 2 Conditional/stepwise regression models to test for additional independent blood pressure SNPs at each locus

| Chr. Trait | SNP | Position (bp) | Alleles | GERA meta-analysis (n = 99,785) | UKB meta-analysis (n = 152,081) | GERA + UKB meta-analysis (n = 251,866) |
|------------|-----|---------------|---------|-----------------------------|-------------------------------|-------------------------------------|
|            |     |               |         | Univariate                  | Joint                         | Univariate                          |
|            |     |               |         | Eff. | P       | Eff. | P       | Eff. | P       | Eff. | P       | Eff. | P       |
| SNPs discovered in GERA |      |            |        |      |        |      |        |      |        |      |        |      |        |
| 6          | PP  | rs1322640    | 16,958,688 | T/C  | −0.27 | 8.9 × 10^−11 | −0.28 | 4.7 × 10^−11 | −0.32 | 1.1 × 10^−13 | −0.33 | 2.1 × 10^−13 | 7.7 × 10^−32 | 0.073 |
|            |     | rs13197550   | 7,169,716,025 | C/A  | −0.23 | 2.1 × 10^−9  | −0.23 | 1.4 × 10^−9  | −0.22 | 2.5 × 10^−14 | −0.22 | 5.8 × 10^−14 | 3.2 × 10^−17 | 0.00005 |
| 20^a       | PP  | rs2104574    | 10,968,891 | C/T  | −0.10 | 0.012 | 0.024 | −0.091 | 1.0 × 10^−8 | −0.18 | 0.000025 | −0.18 | 0.00005 | 0.37 | 1.1 × 10^−5 |
|            |     | rs10573001:1 | 10,573,001 | C/CG | 0.45  | 6.2 × 10^−10 | 0.44  | 1.0 × 10^−8  | 0.24  | 0.012 | 0.22 | 0.023 | 0.37 | 1.1 × 10^−9 |
| 20^a       | DBP | rs2104574    | 10,968,891 | C/T  | −0.25 | 2.6 × 10^−10 | −0.25 | 2.4 × 10^−8  | −0.21 | 2.2 × 10^−7 | −0.22 | 7.1 × 10^−8  | −0.23 | 4.0 × 10^−16 |
|            |     | rs10573001:1 | 10,573,001 | C/CG | 0.0026 | 0.97 | −0.015 | 0.84 | −0.25 | 0.00089 | −0.28 | 0.000024 | −0.12 | 0.021 | −0.15 | 0.0067 |

Effect sizes (in mm Hg) and P values are given from a univariate fit of each SNP one at a time (univariate) and a joint fit of all of the SNPs at each locus (joint). Chr, chromosome; position, Build 37 position; Eff., effect.

^aSNPs on chromosome 20 show an independent phenotype trait association for different SNPs. rs2104574 is the lead GERA SNP near a previously identified SNP.

### Further Discussion

Further improved in meta-analysis of the GERA and UKB cohorts, where 84.7% (72/85) met Bonferroni-adjusted significance, 89.4% (76/85) were nominally significant, and 96.5% (82/85) had effects in the same direction.

In addition, testing an aggregate, weighted genetic risk score (GRS) using all 85 previously described SNPs for each blood pressure trait led to highly significant associations in all GERA ancestry groups, with P < 1 × 10^-16 (whites), P < 1 × 10^-12 (Latinos), P < 1 × 10^-9 (East Asians), P < 0.002 (African Americans), and P < 1 × 10^-50 in UKB whites, for all blood pressure traits (Table 3). In GERA, Latinos had a larger mean SBP GRS than whites (P = 0.053), whereas African Americans had a lower one (P = 0.032). When GERA African Americans were stratified by European ancestry, SBP GRSs were lower in individuals with 0–50% European ancestry (coefficient = 0.65, 95% confidence interval (CI) = 0.18–1.13) than they were in those with 50–100% European ancestry (coefficient = 1.04, 95% CI = 0.56–1.51), although these confidence intervals overlap. For SBP, DBP, and PP (Table 3). There was also a very high degree of concordance for the estimated regression coefficients for SBP, DBP, and PP among the non-Hispanic whites in GERA, ICBP, and UKB (Supplementary Fig. 7).

When examining the effects of individual SNPs, we found that SNPs discovered in ICBP typically had weaker effects than those discovered in GERA, likely owing to winner's curse. The opposite was also the case: SNPs discovered in GERA had weaker effects in ICBP. UKB comparisons were similar, with SNPs having stronger effect sizes in the discovery cohort (GERA or GERA + ICBP) than in the replication cohort (UKB). Seven SNPs exhibited significant heterogeneity among studies (P < 0.00059, with Bonferroni correction for 85 SNPs) at the lead trait (Supplementary Table 6).

### Variance explained and gain using multiple blood pressure measurements

The variance explained in an additive linear model by the 75 genome-wide significant loci identified in our GERA discovery cohort was 1.4%, 1.2%, and 1.8% for SBP, DBP, and PP, respectively, in GERA non-Hispanic whites; note that the same individuals were used for the discovery cohort (GERA or GERA + ICBP) than in the replication cohort (UKB). Seven SNPs exhibited significant heterogeneity among studies (P < 0.00059, with Bonferroni correction for 85 SNPs) at the lead trait (Supplementary Table 6).
for GERA: for example, 2.7%, 2.5%, and 3.0% for UKB whites. Adding dominance terms to the linear regression model did not increase the variance explained (none significant after multiple-comparison correction).

We subsequently investigated the impact of multiple blood pressure measurements in an analysis restricted to individuals who had ≥5 measurements (Supplementary Fig. 8). Using all measurements, in comparison to just one, reduced the regression coefficient standard deviation.
error (SE) by 25%; the regression coefficient estimate itself did not change significantly. With a large number of measurements, the GRS approximately doubled the variance explained for SBP and DBP but explained over threefold greater variance for PP, owing to the greater error in measurement for the latter (Supplementary Table 7). We estimated the blood pressure variance due to measurement error (Online Methods) as 56.5% (SBP), 47.5% (DBP), and 71.5% (PP). Lastly, the number of genome-wide significant variants that would have been found when using 1, 2, 3, 4, and all measurements (in a fixed subset of non-Hispanic white individuals with ≥5 measurements and using genotyped SNPs only) was 2, 3, 3, 7, and 7 SNPs for SBP, 2, 4, 7, 7, and 11 SNPs for DBP, and 4, 7, 15, 14, and 23 SNPs for PP, respectively, demonstrating a large increase in the number of SNPs identified with more measurements included. However, when not fixing the sample size and using all individuals with at least 1, 2, 3, 4, and 5+ measurements, we found 12, 10, 11, 10, and 7 genome-wide significant SNPs for SBP, 14, 14, 14, 13, and 11 genome-wide significant SNPs for DBP, and 20, 21, 23, 21, and 23 significant loci for PP, using a total of 80,792, 78,372, 75,446, 71,834, and 67,547 individuals, respectively, reflecting the loss of statistical power with decreasing sample size. Consequently, it is difficult to determine the optimal minimum number of measurements for subject inclusion, owing to the precision versus sample size tradeoff.

### Blood pressure risk scores and onset of hypertension

We tested the association of the GRSs (described above for SBP, DBP, and PP) with time to onset of hypertension. The predictive value of the GRSs increased with the number of blood pressure–associated SNPs included (Table 3), as expected. When including SNPs from the meta-analysis of all three cohorts, the SBP GRS was the strongest predictor of hypertension, with a hazards ratio (HR) in non-Hispanic whites of 1.18 ($P = 8.2 \times 10^{-25}$); the DBP GRS was slightly less significant with HR = 1.14 ($P = 1.3 \times 10^{-30}$), as was the GRS for PP with HR = 1.15 ($P = 1.5 \times 10^{-33}$). The GRSs were also predictive in other ancestry groups: for example, the significance for the SBP GRS was $P = 1.4 \times 10^{-6}$ in Latinos, $P = 0.0021$ in East Asians, and $P = 0.00024$ in African Americans.

### Sex differences

We tested for differences in SNP effect size by sex (heterogeneity test; Supplementary Table 8; coefficients plot in Supplementary Fig. 9). After Bonferroni correction ($\alpha = 0.00013$, including all 386 new and previously described SNPs), no SNP showed a significantly difference in effect size between the sexes. However, 25 SNPs had a nominally significant ($P < 0.05$) difference for the lead trait, which is in slight excess of the 19.3 SNPs expected; of the SNPs with the same direction of effect in males and females, 17 of 20 (85.0%, 95% CI = 61.1–96.0%) had effects of stronger magnitude in females than in males.

### Differences in SBP, DBP, and PP effects

We tested whether the normalized effect size of each SNP was greater for SBP or DBP (Online Methods and Supplementary Table 9). We found that 26.2% of the SNPs had significantly different normalized effect sizes for SBP and DBP ($P < 0.00013$, Bonferroni correction for 386 SNPs); for 57.4% of these SNPs the normalized effect was greater for SBP than for DBP.

### Heritability from all genotyped and imputed SNPs

Array heritability estimates derived from genotyped SNPs on the basis of PC-Relate kinship estimates, to account for population stratification in the kinship estimates, obtained using GEAR$^{36}$ in non-Hispanic whites were 15.5% (95% CI = 13.9–17.1%) for SBP, 15.1% (95% CI = 13.5–16.7%) for DBP, and 14.5% (95% CI = 12.7–16.2%) for PP, increasing only modestly when adding imputed SNPs to 16.1% (95% CI = 14.5–17.7%) for SBP, 17.0% (95% CI = 15.6–18.4%) for DBP, and 15.6% (95% CI = 14.0–17.2%) for PP. These estimates were similar to those obtained when not accounting for population stratification in the kinship estimates but adjusting for it in the phenotypic model instead using GCTA$^{33}$ (SBP $h^2 = 16.8%$, 95% CI = 15.1–18.6%); this may be because the ancestry effect in non-Hispanic whites is modest. Sample sizes were too small to evaluate the other GERA groups.

### eQTL analysis in different tissues

We investigated whether the previously identified and all new loci colocalized with expression quantitative trait loci (eQTLs). We used eQTLs from 44 Genotype-Tissue Expression (GTEx) tissues and kidney$^{32,33}$. Across all tissues, 186 of the 367 sentinel SNPs were eQTLs in at least one tissue; at least one SNP in 213 of the same 367 loci was an eQTL. We determined for each tissue whether the number of eQTLs (identified by either sentinel SNP or locus) was greater than expected by chance, where expectation was derived from a random sampling of SNPs and loci (Online Methods). We ranked the tissues by eQTL value, for both the sentinel SNP and locus analyses. We generally expect tissues with more eQTLs to overlap more SNP sets and enrichment to be greater simply because of chance GWAS set overlap, especially when eQTLs in tissues relevant to the phenotype are also found in these tissues. To observe whether the enrichment visible for a given tissue was greater than expected relative to the total number of eQTLs the tissue contained, we examined the relationship between $P$ value and total eQTL count for each tissue (Fig. 5). The aorta and tibial artery were clear outliers in comparison to other tissues, even when accounting for the total number of eQTLs.

### Enrichment analysis for functional elements

We subsequently investigated whether genes near sentinel variants were enriched for certain functional pathways. We included genes within ±0.5 Mb of the 390 sentinel variants with a significant eQTL in either aorta or tibial artery. We identified 2,013 genes near all
390 sentinel variants (Online Methods) and tested for enrichment of functional annotation. Using DAVID 6.8 (refs. 34,35), 1,480 of these genes had annotations, yielding 26 significant annotation terms (Benjamini–Hochberg $P < 0.05$; Supplementary Table 10), but no clear functional pathway emerged.

**DISCUSSION**

In this large, ancestrally diverse GERA cohort with EHR-derived blood pressure measures, we discovered 39 new genome-wide significant blood pressure loci, most of which replicated in ICBP and UKB. Merging GERA and ICBP identified 36 additional new genome-wide significant loci, most of which replicated in UKB. Finally, merging all three cohorts identified 241 additional genome-wide significant loci, although no replication was available. Conversely, we were able to replicate almost all 85 previously described blood pressure–associated SNPs. We also showed that using multiple EHR blood pressure measurements almost doubled the variance explained, although the total variance explained remains small (for example, 2.9% for SBP in non-Hispanic whites). We also showed that blood pressure signals are enriched in two large arteries, the aorta and tibial artery.

Our study used a large general population sample with EHR-derived data for the first time, to our knowledge, in blood pressure GWAS. The consistency and generalizability of blood pressure genomics findings from one-time research-protocol-based assessments to purely clinical measures recorded in an EHR have been questioned. We were able to replicate most previously identified loci from many cohorts using research-based assessments, demonstrating that blood pressure genetic findings are not significantly different between studies using research assessments and those using clinical, EHR-derived ones. This is important because clinical measures recorded in the EHR are the basis for clinical decisions in general, real-world clinical practice. Moreover, this extends the reach of GWAS to numerous clinical samples.

EHR-based studies offer additional benefits. Our identification of new variants takes advantage of multiple independent measurements in the EHR to increase statistical power. Our study increased the standardization and reduced the variability of the EHR-derived blood pressure measures by excluding measures obtained in clinical settings with increased measurement variability, for example, emergency rooms, and retaining measures obtained in visits to primary care/internal medicine departments.

The new blood pressure–associated SNPs identified have similar genomic contexts to those previously described, of which 8.2%, 20.0%, 32.9%, and 38.8% were located in exons, UTRs, introns, and intergenic regions, respectively; the new SNPs identified in GERA were distributed with 2.6%, 23.1%, 33.3%, and 41.0% in these regions, those from the GERA and ICBP meta-analysis had a distribution of 0%, 2.8%, 55.6%, and 41.7%, and those from the GERA, ICBP, and UKB meta-analysis had a distribution of 2.5%, 14.9%, 41.5%, and 41.1% (Supplementary Table 4). The frequencies and variant types of the lead SNPs are also similar to those previously described; for individuals of European ancestry, 85.9% of previously described SNPs have minor allele frequencies (MAFs) $>0.10$, in comparison to 89.7% of GERA-identified SNPs, 94.4% of SNPs identified by GERA and ICBP meta-analysis, and 82.2% of SNPs identified by GERA, ICBP, and UKB meta-analysis. When comparing results across traits within GERA, the leading trait locus was more often associated with PP for new loci than before (24.7% of previously described SNPs were associated with PP versus 59.0%, 58.3%, and 41.9% in GERA, GERA + ICBP, and GERA + ICBP + UKB, respectively); this may reflect the fact that earlier blood pressure studies tested SBP and DBP, but not PP. We additionally demonstrate the significant effect of summary blood pressure SNP scores on time to onset of hypertension, enabled by GERA longitudinal EHR data. We note that a GERA hypertension GWAS produced no additional novel results (and yielded results much less significant than those for the continuous blood pressure traits, as expected).

One limitation was that 1000 Genomes Project–imputed results were unavailable in ICBP; however, the much larger UKB replication cohort did not have this limitation. For ICBP, we therefore relied on imputation of summary test statistics using HapMap. Use of these approximated results, and the fact that all test statistics from ICBP were based on SNP results imperfectly imputed to HapMap data, likely led to diminished effect sizes in ICBP. Overall, we needed a very large number of individuals for replication, in both the replication of our new GERA results, which improved greatly when adding UKB to ICBP, and the replication of previously described results, which improved when adding UKB to GERA.

Another advantage of a single large cohort, such as GERA, is the ability to directly assess additional local SNPs by conditional analysis. The absence of individual-level data requires assumptions about linkage disequilibrium (LD) from other studies. Nevertheless, we only found two additional variants in GERA that were ultimately not explained by nearby previously described SNPs, and an additional four when combining GERA and UKB. We further note that these additional conditional hits were located at a substantial distance from the sentinel SNP of the locus, likely indicating an independent gene and/or mechanism involved. The lack of identification of additional SNPs close to sentinel SNPs is quite distinct from what is observed for serum lipids, for example, and suggests that lower-frequency variants with larger effects within the same loci identified here are uncommon. A similar conclusion was recently obtained in a sequencing study of type 2 diabetes.

Although our sample sizes were smaller in the other ancestry groups than they were for non-Hispanic whites, we noticed that Latinos had the highest standardized GRs, followed closely by non-Hispanic whites and then by East Asians and African Americans. In African Americans, European ancestry was associated with lower blood pressure, but individuals with more European ancestry had higher standardized blood pressure GRs (created from previously described SNPs); this observation is counterintuitive but may reflect the fact that the GWAS discovery occurred primarily in European-ancestry individuals, and it suggests that there may be other SNPs in African Americans remaining to be identified.

We also looked for a pattern in terms of which loci replicated. Logically, the largest replication indicator was discovery $P$ value, as stronger associations likely require a smaller sample size for replication than weaker ones. In GERA, loci with $P \leq 1 \times 10^{-9}$ replicated at a Bonferroni level at a rate of 76.5% (13/17) versus a rate of 54.5% (12/22) for loci with $5 \times 10^{-8} \leq P < 1 \times 10^{-9}$; all of the ICBP SNPs with $P \leq 1 \times 10^{-9}$ replicated at a Bonferroni level in GERA + UKB; however, this pattern was not seen in the GERA and ICBP meta-analysis, where 57.1% (4/7) of the loci with $P \leq 1 \times 10^{-9}$ versus 62.1% (18/29) of the loci with $5 \times 10^{-8} \leq P < 1 \times 10^{-9}$ replicated, although numbers were small. Perhaps also of note, the two SNPs in GERA in MAF $<0.001$ failed to replicate in UKB ($P > 0.05$).

We also searched for eQTL enrichment in a variety of tissues. Both aorta and tibial artery were clear outliers in comparison to other tissues, suggesting that genetic factors influencing vascular elasticity and/or stiffness are important determinants of blood pressure and hypertension.

There are several reasons for the enhanced discovery in our study: increased sample size, multiple blood pressure measures (reducing...
Genome-wide association analysis of blood-pressure traits

Genome-wide association study identifies six new loci supervised the creation of genotype data. D.R., in collaboration with C.I., C.S., T.J.H., G.B.E., C.I., A.C., and N.R. conceived and designed the study. P.-Y.K. collected and analyzed the data, performed statistical analyses. T.J.H., G.R.E., P.N., C.I., A.C., and N.R. interpreted the results of analyses. All authors contributed to the drafting and critical review of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Yang, Q. et al. Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults. J. Am. Med. Assoc. 307, 1273–1283 (2012).
2. Levy, D. et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC Med. Genet. 8 (Suppl. 1), S3 (2007).
3. Mital, W.E. & Oldham, P.D. The hereditary factor in arterial blood-pressure. BMJ 1, 75–80 (1963).
4. Fox, E.R. et al. Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. Hum. Mol. Genet. 20, 2273–2284 (2011).
5. Franceschini, N. et al. Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. Am. J. Hum. Genet. 93, 545–554 (2013).
6. Ganesh, S.K. et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum. Mol. Genet. 22, 1663–1678 (2013).
7. Ganesh, S.K. et al. Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. Am. J. Hum. Genet. 95, 49–65 (2014).
8. Ho, J.E. et al. Discovery and replication of novel blood pressure genetic loci in the Women’s Genome Health Study. J. Hypertens. 29, 62–69 (2011).
9. Johnson, T. et al. Blood pressure loci identified with a genome-wide array. Am. J. Hum. Genet. 89, 686–700 (2011).
10. Kato, N. et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nat. Genet. 43, 531–538 (2011).
11. Kato, N. et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat. Genet. 47, 1282–1293 (2015).
12. Kelly, T.N. et al. Genome-wide association study meta-analysis reveals transethnic replication of mean arterial and pulse pressure loci. Hypertension 62, 853–859 (2013).
13. Levy, D. et al. Genome-wide association study of blood pressure and hypertension. Nat. Genet. 41, 677–687 (2009).
14. Newton-Cheh, C. et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat. Genet. 41, 666–676 (2009).
15. Padmanabhan, S. et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. PLoS Genet. 6, e1000117 (2010).
16. Siminoj, J. et al. Gene–age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. Am. J. Hum. Genet. 95, 24–38 (2014).
17. Ehteh, G.B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103–109 (2011).
18. Takeuchi, F. et al. Blood pressure and hypertension are associated with 7 loci in the Japanese population. Circulation 121, 2302–2309 (2010).
19. Tragante, V. et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. Am. J. Hum. Genet. 94, 349–360 (2014).
20. Wain, L.V. et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat. Genet. 43, 1005–1011 (2011).
21. Wang, Y. et al. Whole-genome association study identifies STK39 as a hypertension susceptibility gene. Proc. Natl. Acad. Sci. USA 106, 226–231 (2009).
22. Zhu, X. et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am. J. Hum. Genet. 95, 21–36 (2015).
23. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283 (2013).
24. Sewell, K. et al. Blood pressure measurement biases in clinical settings. Am. J. Hum. Genet. 93, E01 (2016).
25. Nwanwko, T., Yoon, S.S., Burt, V. & Gu, Q. Hypertension among adults in the United States: National Health and Nutrition Examination Survey, 2011–2012. NCHS Data Brief 133, 1–8 (2013).
26. Devlin, B. & Roeder, K. Genomic control for association studies. Biometrics 55, 997–1004 (1999).
27. Sudlow, C. et al. 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 (2015).
28. Kraft, P. Curses—winner’s and otherwise—in genetic epidemiology. Epidemiology 19, 649–651 discussion 657–658 (2008).
29. Conomos, M.P., Reiner, A.P., Weir, B.S. & Thornton, T.A. Model-free estimation of recent genetic relatedness. Am. J. Hum. Genet. 98, 127–148 (2016).
30. Chen, G.-B. Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman–Elston regression. Stat. Genet. Methodol. 5, 107 (2014).
31. 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 (2011).
32. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015).
33. Zhang, X. et al. Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. BMC Genomics 15, 532 (2014).
34. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
35. Huang, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13 (2009).
36. Kho, A.N. et al. Electronic medical records for genetic research: results of the eMERGE Consortium. Sci. Transl. Med. 3, 79re1 (2011).
37. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934 (2012).
38. Tada, H. et al. Multiple associated variants increase the heritability explained for plasma lipids and coronary artery disease. Circ Cardiovasc Genet 7, 583–587 (2014).
39. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. Nature 536, 41–47 (2016).
40. Hoffmann, T.J. et al. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. Genomics 98, 79–89 (2011).
41. Hoffmann, T.J. et al. 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 (2011).
42. Liu, C. et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat. Genet. 48, 1162–1170 (2016).
43. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat. Genet. 48, 1151–1161 (2016).
44. Ehret, G.B. et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat. Genet. 48, 1171–1184 (2016).
ONLINE METHODS

All statistical tests were two-sided.

Participants, phenotype, and genotyping. Our primary analysis used individuals from the RPGEH GERA cohort, which has been described previously.45,46 We used three trait outcomes: SBP, DBP, and PP, where PP = SBP – DBP. We began with 3,197,317 GERA EHR blood pressure measurements. In KPNC, blood pressure is measured and recorded in the EHR at the beginning of each clinic visit, regardless of the visit reason. Examination of mean blood pressure measurements by medical specialty showed that, in comparison to internal medicine, average blood pressure measurements obtained in the following departments were significantly higher (P < 0.0001): anesthesiology, chemical and alcohol dependency, health education, emergency room, hospital care, ophthalmology, physical therapy, rehabilitation, transplant, urgent care, and urology. Higher average blood pressure measurements in these specialties likely indicate effects of acute illnesses or other effects on blood pressure, and we excluded all blood pressure measurements obtained in these specialty visits; 3,046,609 blood pressure measurements (95%) remained after these exclusions. We further excluded 1,127,077 measurements recorded as binned into five systolic and seven diastolic blood pressure ranges (for example, systolic blood pressure recorded in the range 140–159 mm Hg); this was an early recording method before the full EHR implementation in 2006. After noting that 75.6% of the 1,919,532 remaining measurements were from visits to internal medicine departments, we excluded the 188,173 obstetrics/gynecology and 280,501 other departmental measurements to obtain the most homogeneous blood pressure phenotype, resulting in 1,450,858 measurements from IM visits on 107,196 individuals. Finally, after excluding those failing genotyping, 1,342,814 independent SBP and DBP internal medicine visit measurements from different days (345,031 untreated and 997,783 treated) on 99,785 individuals obtained from the beginning of 2006 to the end of 2011 remained for analysis. Treatment with antihypertensive medication was assessed via EHR prescription filling information; once an individual started a drug, they were considered to be treated for all subsequent measurements. We added 15 mm Hg to treated SBP values and 10 mm Hg to treated DBP values, similar to previous blood pressure GWAS, to correct for treatment effect.

Individuals were genotyped at over 650,000 SNPs on one of four ancestry-specific Affymetrix Axiom arrays optimized for individuals of European (EUR), Latino (LAT), East Asian (EAS), and African-American (AFR) ancestry.43,44 We analyzed 80,792 non-Hispanic whites, 8,231 Latinos/other, 7,243 East Asians, 3,058 African Americans, and 461 South Asians (genotyped on the EUR array). The Kaiser Foundation Research Institute and University of California at San Francisco institutional review boards approved this project. Written informed consent was obtained from all subjects.

Genotype quality control and imputation. Initial genotype quality control was performed across ancestry-specific array, as described.46 In addition, we required an array per-SNP call rate ≥90%, resulting in 665,350 (EUR), 777,927 (LAT), 704,105 (EAS), 864,905 (AFR), and 663,783 South Asian (SAS) SNPs. We excluded SNPs with a minor allele count (MAC) <20, resulting in 665,350 (EUR), 3,058 (LAT), 7,243 (EAS), 3,058 (AFR), and 663,783 (SAS) SNPs, respectively.

Imputation was performed on an array-wise basis. We first prephased the genotypes with SHAPEIT v2.2.7219 (ref. 48). We then imputed variants from the 1000 Genomes Project (phase I integrated release, March 2012, with Aug 2012 chromosome-X update, a cosmopolitan reference panel with singletons removed) with IMPUTE2 v2.30 (refs. 49–51). The estimated quality control metric, r̄2, used in this study is the Info metric from IMPUTE2, which is an estimate of the correlation of the imputed genotype with the true genotype. Poorly imputed (r̄2 < 0.3) SNPs and those with MAC <20 were removed, resulting in 24,149,855 (EUR), 20,828,585 (LAT), 15,248,462 (EAS), 21,485,958 (AFR), and 8,607,429 (SAS) SNPs (28,613,428 unique SNPs) available for analyses.

GWAS analysis and covariate adjustment. We first analyzed each of the five ancestry groups separately. Data from each SNP were modeled using additive dosages accounting for imputation uncertainty.53 For each quantitative trait (treatment-adjusted SBP, DBP, and PP), for computational efficiency, we first ran a mixed model of the blood pressure measurement adjusted for age, age2, BMI, and sex using all blood pressure measurements for each individual. We then constructed a long-term average residual for each individual as the dependent variable in a linear mixed model using estimated kinship matrices with leave one chromosome out (LOCO) to account for population substructure and cryptic relatedness with Bolt-LMM.54 Finally, we undertook a fixed-effects meta-analysis to combine the results of the five groups using Metasoft v2.0 (ref. 55). We considered as new loci that were at a physical distance >0.5 Mb from any previously described locus (and visual inspection for stretches of longer LD).

To find additional independent SNPs at each locus, we ran a conditional stepwise regression analysis at all SNPs with r2 >0.8 in the GERA meta-analysis, around each previously described and new GERA SNP. We looked for additional genome-wide significant SNPs within a 1-Mb window (30.5 Mb) centered on the lead SNP. Although this generally worked well, certain portions of the genome have stronger LD (we noted particularly at the ends of chromosomes and centromeres, where recombination is suppressed), which we assessed via visual inspection of the Manhattan plots to form an expanded window size, and repeated the stepwise regression on the expanded window. In these analyses, we adjusted for ancestry principal components instead of the mixed-model approach, both for simplicity and computational efficiency.

To adjust for genetic ancestry/population stratification when not using Bolt-LMM LOCO, we performed a principal-component analysis, as described.53 The first ten eigenvectors for non-Hispanic whites and the first six eigenvectors for all other ancestry groups were included as covariates in the regression model described above. When we tested European versus African ancestry percentages in African Americans, we used PC1 as a surrogate for European admixture.

Replication of new GERA SNPs using ICBP and UKB. To test the 39 new GERA genome-wide significant SNPs for replication, we evaluated associations using a fixed-effects meta-analysis of ICBP and UKB. We also tested the 36 new SNPs found in the meta-analysis of GERA and ICBP for replication in UKB. We report associations that replicate at a strict Bonferroni threshold (P < 0.000067, to account for a total of 75 new SNPs tested), as well as suggestive (P < 0.01) and nominally suggestive (P < 0.05) findings with effects in the same direction as the original.

ICBP ICBP GWAS summary statistics from 69,396 individuals at 2,696,785 SNPs were obtained from dbGaP. As only summary statistics were available, we did not use these data to replicate conditional SNPs.

As the ICBP has been imputed to HapMap v22, a smaller reference panel than used here for GERA, we used ImpG v1.0 (ref. 56) to estimate the summary statistics for the 1000 Genomes Project reference panel SNPs used for GERA imputation. To solve for the effect size β of the additive coded genotype Xij (i indexing N individuals, j indexing SNPs) from the summary statistics imputed to 1000 Genomes Project from ImpG, we assumed that the ICBP had the same allele frequency as in the 1000 Genomes Project European-ancestry individuals and Hardy–Weinberg equilibrium. Let qj be the MAF and p1 = 1 – qj. Assuming Hardy–Weinberg equilibrium, Np2 j individuals have Xij = 0, Np1 qj individuals have Xij = 1, and Nq2 j individuals have Xij = 2. It is known that SE(βj) = r̄2j/σε, where r̄2j is the residual of the phenotype regressed on the SNP genotype Xij and σε2 = σX2 – mean(X)$^2$. It can be shown that σX2 = 2np2 j. Although σX2 is unknown in ICBP, a reasonable approximation is obtained by assuming that each SNP individually explains little variation of the trait variance and thus σX2 is constant and does not depend on j, that is, σX2 = σ2. We solve for this quantity using the existing effect size estimate of βj from the available HapMap SNPs. Using ImpG assumes that all HapMap SNPs were imputed without error; such error likely damps the results.

UKB. The UKB cohort has previously been described57. Of note, genotypes were imputed using a larger number of individuals from the UK10K combined with 1000 Genomes Project as a reference panel (n = 6,285). SBP measures were taken from manual (variable 93.0–2.0–1) and automatic (4,080.0–2.0–1) readings, as were DBP measures (94.0–2.0–1 and 4,079.0–2.0–1, respectively). Age was reported as the age at measurement (34.0.0). Antihypertensive use was assessed by self-report (6,153.0–1.0 and 6,177.0–2.0), and blood pressure measurements were corrected as in GERA. BMI was calculated from measured
weight and height (21,001.0). Sex was determined genetically (22,001.0). Analysis was carried out as in GERA, a meta-analysis of each self-reported ancestry group (21,000.0–2.0): we identified 145,341 individuals who reported any white group and with global ancestry PC1 ≤ 50 and PC2 ≤ 50, where global PC1 and PC2 were calculated from the entire cohort (22,009.0.1–2), including 2,274 South Asians, 2,029 African British, 1,979 mixed/other, and 458 East Asians, totaling 152,081 individuals. Ancestry principal components within whites were calculated using 50,000 random white individuals with the remaining subjects projected, which has been shown to work well15, and then within each other group. We analyzed 35,893,267, 12,078,001, 19,866,667, 15,820,020, and 7,298,789 SNPs with \( r^2 \) > 0.05 and MAF \( \geq 0.001 \) and 0.005, 0.005, and 0.025, in whites, South Asians, African Europeans, mixed/other, and East Asians, respectively (42,521,712 unique SNPs).

GERA meta-analysis with ICBP and with UKB. We additionally performed meta-analysis of the GERA and ICBP results for genome-wide discovery using a fixed-effects meta-analysis, with UKB for replication. We further performed a discovery meta-analysis of GERA, ICBP, and UKB for maximal discovery size but with no replication sample available. In this analysis, we reviewed the locus plots, manually merging the windows located ±0.5 Mb with respect to the SNPs when necessary. Specifically, after assessing SNPs in the GERA + ICBP + UKB meta-analysis, we checked whether the SNPs appeared independent in a meta-analysis of GERA and UKB, as both had individual-level data. Most regions were either obviously correlated with high \( r^2 \) or were obviously not correlated with \( r^2 < 0.05 \); however, to formalize the conditional analysis and retain a SNP as independent, we required that the reduction in \( P \) values from univariate to joint in the GERA + UKB meta-analysis be less than tenfold and additionally that translating an equivalent reduction in \( P \) values to the GERA + ICBP + UKB meta-analysis still led to a genome-wide significant result (that is, if we assumed that \( P_{\text{joint,GERA+ICBP+UKB}} / P_{\text{univariate,GERA+ICBP+UKB}} = P_{\text{joint,GERA+UKB}} / P_{\text{univariate,GERA+UKB}} \), the approximated \( P_{\text{joint,GERA+ICBP+UKB}} \) would still need to be genome-wide significant). This may have been slightly conservative.

Replication analysis of previously described SNPs in GERA. To determine how many of the 85 previously described loci from ICBP and other GWAS replicated in this study, we tested the sentinel SNPs from those studies in our data set4–12. Frequently, multiple blood pressure phenotypes are reported for the same loci. We used Bonferroni correction for replication (85 SNPs, \( \alpha = 0.00059 \)). The SNP rs2446849 was not in our reference panel, so we used the closest proxy, rs2513758, at a physical distance of 876 bp and \( r^2 = 0.98 \) in Europeans.

GRS construction. We constructed a GRS for each of the three blood pressure traits for each individual by summing the additive coding of each set of SNPs associated with the particular blood pressure trait weighted by the previously described effect size from ICBP and then standardized the distribution of all groups simultaneously by the mean and standard deviation (to a standard normal distribution) for interpretability. We used the leading SNP from each locus.

Multiple measurements. To assess the impact of multiple blood pressure measurements, we compared the \( P \) value and effect size estimates for the previously described GWAS significant SNPs using one, two, three, four, and all measurements from each individual. We used a set of 67,547 non-Hispanic white individuals, all with 25 blood pressure measurements available for this analysis, to keep the sample size identical among comparisons. We also examined the variance explained by the GRS of the previously described hits assuming previous effect sizes as a function of the number of blood pressure measurements.

From this analysis, we can also estimate both the variance due to measurement error and variance explained by the GRS in the absence of measurement error, as follows. Let \( B \) be the observed blood pressure measurement, \( G \) be the GRS, \( E \) be the residual genetic and environmental effect on blood pressure, \( M \) be the component of blood pressure due to measurement error, and \( k \) be the number of blood pressure measurements. We assume that the measurement error is independent across multiple measurements within an individual, and the additive model \( B = G + E + M_k \) for the average of \( k \) blood pressure measurements. Let \( V_E = Var(E) \), \( V_G = Var(G) \), and \( V_M = Var(M) \). For \( k \) blood pressure measurements with independent measurement error, \( V_M = V/\sigma^2 \). The proportion \( H \) of variance in blood pressure attributable to the GRS is \( \beta_0 V_G + V_E + V_M \). Then, \( 1/\beta_0 = (1 + V_M/V_G) + (V_G/V_E) + (V_E/V_M) = k/a + \beta_0/k \), where \( \beta_0 = 1 + V_M/V_G \) and \( \beta_0 = V_E/V_M \). We thus have a linear model of \( 1/\beta_0 \) in terms of 1/k, and \( 1/\beta_0 = V/G + V + V_M \) is the proportion of variance due to the GRS in the absence of measurement error and \( \beta_0 (1 + \beta) \) is the proportion of variance in blood pressure due to measurement error. Fitting a linear regression model to \( 1/\beta_0 \) as a function of \( 1/k \), we can then use the estimated intercept (\( \alpha \)) and regression coefficient (\( \beta \)) to estimate the error variance and variance due to the GRS in the absence of measurement error.

Blood pressure risk scores and onset of hypertension. We additionally tested GRS constructed by weighting different subsets of identified blood pressure–associated SNPs (identified for SBP, for DBP, and for PP, constructed as described above). Hypertension onset here was defined as the first hypertension treatment time or the first time either SBP ≥ 140 or DBP ≥ 90 occurred in an individual and was maintained for the next subsequent blood pressure measurement. Individuals were left censored at their first measurement (and not included if already meeting the hypertension diagnosis criterion) and right censored at their latest measurement if not hypertensive.

Differences in SBP, DBP, and PP effects. We also tested whether the normalized effect size of each SNP was different for SBP versus DBP. Suppose that \( Y \) is SBP normalized to a standard normal (mean centered and then divided by the standard deviation) and \( Z \) is normalized DBP and that \( X \) is the SNP dosage. Then, we model \( Y = ax + E \) and \( Z = bx + F \), where \( a \) is the regression coefficient for \( Y \) on \( X \) and, similarly, \( b \) is the coefficient for \( Z \); \( E \) and \( F \) are the residual errors, respectively. Since \( Var(Y) = Var(Z) = 1 \), assuming \( a \) and \( b \) have the same sign (which is generally the case because the phenotypes are correlated), testing the equality of \( a \) and \( b \) is also a test of effect difference between SBP and DBP. Now, consider the difference \( Y = Z = (a - b)x + (E - F) \). Regressing \( Y - Z \) on \( X \) tests the difference between \( a \) and \( b \); in this analysis, we additionally adjust for the same covariates as discussed previously.

GWAS heritability from all measured SNPs. We estimated the additive array heritability of each individuals long-term average age- and BMI-adjusted blood pressure residual pressures using GEAR v0.7.7 (ref. 30). Array heritability estimates may be more sensitive to artifacts than GWAS results30, so we restricted our analysis to the largest group of individuals, non-Hispanic whites, that were run with the same reagent kit and type of microarray (\( n = 73,133 \))46. We used only autosomal data, a common practice in array heritability estimation, and also LD filtered our data so no two pairwise SNPs had \( r^2 > 0.8 \) with a standard greedy algorithm in PLINK v1.07 (ref. 58). This resulted in 547,922 genotyped SNPs and 3,796,606 imputed SNPs restricted to \( r^2 < 0.8 \). Because of population stratification, we used PC-Relate29 to estimate kinship coefficients rather than the standard GCTA estimates31, which assume a homogeneous population; we also compared the results to those obtained using the standard GCTA kinship estimates with principal-component adjustment. We used GEAR rather than GCTA to estimate heritability because the PC-Relate kinship matrix estimate was not positive definite; this can be explained by the fact that the matrix entries are computed on the basis of different allele frequencies, that is, those depending on ancestry from the principal-component analysis. In all analyses, we removed individuals so that no two remaining individuals had a kinship estimate >0.025; sample size was maximized with PLINK v1.9 (ref. 59), leaving us with 62,133 individuals.

eQTL enrichment analysis. To carry out tissue-specific eQTL enrichment analysis, we used 44 tissue types with at least 70 samples available from GTEx25 in addition to 7 kidney eQTLs33. We used 367 sentinel variants from previously identified SNPs and the three discovery stages presented here with MAF >0.001 and in eQTL databases. Next, 100 sets of 367 random pseudo-sentinel variants were selected matching the MAF to that of the original 367 SNPs (within ±0.5%). Within each set, the selection was done without replacement; the match for each variant was selected one at a time, and selection of the
subsequent variant excluded all previously selected variants, as well as all variants within ±0.5 Mb of all previously selected variants.

Enrichment was tested at both the sentinel SNP level and locus level, conceptually similar to Nicolae et al.\(^1\). At the sentinel SNP level, the number of variants that were also eQTLs in any of the 45 tissues was counted. At the locus level, variants in high LD (r\(^2\) > 0.8) with any of the 367 sentinel variants were examined for overlap with eQTLs, and if at least one variant within the locus was also an eQTL the locus was counted. Subsequently, this was repeated for 100 randomly generated sets to observe whether an eQTL enrichment was visible in the GWAS set. To assess which of the 45 tissues were driving the enrichment, counts were also computed per tissue. For each tissue, an upper-tailed P value for enrichment of the GWAS count was calculated with a z score computed using the mean and standard deviation of the null distribution for that tissue.

**DAVID analysis.** Annotation of genes surrounding sentinel variants was conducted with DAVID 6.8 beta (the non-beta version was 6 years old)\(^2\).\(^3\).\(^4\). Genes within ±0.5 Mb of each of the 390 sentinel variants were selected, as defined by GENCODE v19 GTF\(^5\). Subsequently, those with at least one significant eQTL in tissues identified from the previous enrichment analysis were included in the final list for analysis. Functional annotation analysis was run on the *Homo sapiens* background with default annotations in the categories of disease, functional categories, gene ontology, pathways, and protein domains, as well as with default parameters, retaining terms with at least two assigned genes. Annotation terms meeting Benjamini–Hochberg default parameters, retaining terms with at least two assigned genes. Annotation terms meeting Benjamini–Hochberg P < 0.05 (adjusting for the number of terms) were considered significant.

**Data availability.** Data, including all genotype data and information on hypertension status, are available on approximately 78% of GERA participants from the database of Genotypes and Phenotypes (dbGaP) under accession phs000674.v2.p2. This includes individuals who consented to having their data shared with dbGaP. The complete GERA data are available upon application to the KP Research Bank Portal. The ICBP summary statistics are available from dbGaP under accession phs000585.v1.p1. The UK Biobank data are available upon application to the UK Biobank.

45. Banda, Y. et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* **200**, 1285–1299 (2015).
46. Kvale, M.N. et al. 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 (2015).
47. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. & Burton, P.R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935 (2005).
48. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
49. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011).
50. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
51. 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 (2009).
52. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
53. Huang, L., Wang, C. & Rosenberg, N.A. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am. J. Hum. Genet.* **85**, 692–698 (2009).
54. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
55. Han, B. & Esko, T. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
56. Pasaniuc, B. et al. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* **30**, 2906–2914 (2014).
57. Lee, S.H., Way, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
58. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
59. Chang, C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
60. Nicolae, D.L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
61. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).