Uncovering genetic mechanisms of hypertension through multi-omic analysis of the kidney

James M. Eales[1,37], Xiao Jiang[1,37], Xiaoguang Xu[1,37], Sushant Saluja[1], Artur Akbarov[1], Eddie Cano-Gamez[2], Michelle T. McNulty[3,4], Christopher Finan[5], Hui Guo[6], Wojciech Wystrychowski[7], Monika Szulinska[8], Huw B. Thomas[9], Sanjeev Pramanik[1], Sandesh Chopade[5], Priscilla R. Prestes[10], Ingrid Wise[11], Evangelos Evangelou[12,13], Mahan Salehi[1], Yusif Shakanti[1], Mikael Ekholm[1,†], Matthew Denniff[14], Alicja Nazgiewicz[1], Felix Eichinger[15], Bradley Godfrey[16], Andrzej Antczak[16], Maciej Glyda[17], Robert Król[7], Stephen Eyre[18], Jason Brown[19], Carlo Berzuini[6], John Bowes[18], Mark Caulfield[20,21], Ewa Zukowska-Szczechowska[22], Joanna Zywiec[23], Pawel Bogdanski[8], Matthias Kretzler[15], Adrian S. Woolf[24,25], David Talavera[1], Bernard Keavney[1,26], Pasquale Maffia[27,28,29], Tomasz J. Guzik[28,30], Raymond T. O'Keefe[9], Gosia Trynka[2,31], Nilesh J. Samani[14,32], Aroon Hingorani[5], Matthew G. Sampson[3,4,33], Andrew P. Morris[18,34], Fadi J. Charchar[10,14,35] and Maciej Tomaszewski[1,36*]

[1]Division of Cardiovascular Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK.
[2]Department of Human Genetics, Wellcome Sanger Institute, Cambridge, UK.
[3]Division of Nephrology, Boston Children's Hospital, Boston, MA, USA.
[4]The Broad Institute, Cambridge, MA, USA.
[5]Institute of Cardiovascular Science, University College London, London, UK.
[6]Centre for Biostatistics, School of Health Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK.
[7]Department of General, Vascular and Transplant Surgery, Medical University of Silesia, Katowice, Poland.
[8]Department of Obesity, Metabolic Disorders Treatment and Clinical Dietetics, Karol Marcinkowski University of Medical Sciences, Poznan, Poland.
[9]Division of Evolution and Genomic Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK.
[10]Health Innovation and Transformation Centre, School of Science, Psychology and Sport, Federation University Australia, Ballarat, Victoria, Australia
[11]Australian Institute of Tropical Health & Medicine, James Cook University, Cairns, Australia.
[12]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK.
[13]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.
[14]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK.
[15]Department of Medicine, University of Michigan, Ann Arbor, MI, USA.
[16]Department of Urology and Uro-oncology, Karol Marcinkowski University of Medical Sciences, Poznan, Poland.
[17]Department of Transplantology and General Surgery Poznan, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland.
[18]Division of Musculoskeletal & Dermatological Sciences, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK.
[19]Division of Research and Innovation, Manchester University NHS Foundation Trust, Manchester, UK.
[20]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
[21]National Institute for Health Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK.
[22]Department of Health Care, Silesian Medical College, Katowice, Poland.
[23]Department of Internal Medicine, Diabetology and Nephrology, Zabrze, Medical University of Silesia, Katowice, Poland.

[24]Division of Cell Matrix Biology and Regenerative Medicine, Faculty of Biology Medicine and Health, University of Manchester, Manchester, UK.
[25]Royal Manchester Children's Hospital and Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, Manchester, UK.
[26]Division of Cardiology and Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust Manchester, Manchester, UK.
[27]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.
[28]Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.
[29]Department of Pharmacy, University of Naples Federico II, Naples, Italy.
[30]Department of Internal and Agricultural Medicine, Jagiellonian University College of Medicine, Kraków, Poland.
[31]Open Targets, Wellcome Genome Campus, Cambridge, UK.
[32]National Institute for Health Research, Leicester Biomedical Research Centre, Leicester, UK.
[33]Harvard Medical School, Boston, MA, USA.
[34]Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, Liverpool, UK.
[35]Department of Physiology, University of Melbourne, Parkville, VIC, Australia.
[36]Division of Medicine and Manchester Heart Centre, Manchester University NHS Foundation Trust Manchester, Manchester, UK.
[37]These authors contributed equally to this work.

†Current address (if different from affiliation above)
Mikael Ekholm: Karolinska Institutet, Department of Clinical Sciences, Danderyd Hospital, Stockholm, Sweden
Sanjeev Pramanik: East Lancashire Hospitals NHS Trust, Blackburn, UK

*e-mail: maciej.tomaszewski@manchester.ac.uk

The kidney is an organ of key relevance to blood pressure (BP) regulation, hypertension and antihypertensive treatment. However, genetically mediated renal mechanisms underlying susceptibility to hypertension remain poorly understood. We integrated genotype, gene expression, alternative splicing and DNA methylation profiles of up to 430 human kidneys to characterize the effects of BP index variants from genome-wide association studies (GWAS) on renal transcriptome and epigenome. We uncovered kidney targets for 479 (58.3%) BP-GWAS variants and paired 49 BP-GWAS kidney genes with 210 licensed drugs. Our colocalization and Mendelian randomization analyses identified 179 unique kidney genes with evidence of putatively causal effects on BP. Through Mendelian randomization, we also uncovered effects of BP on renal outcomes commonly affecting hypertensive patients. Collectively, our studies identified genetic variants, kidney genes, molecular mechanisms and biological pathways of key relevance to the genetic regulation of BP and inherited susceptibility to hypertension.

Persistently elevated blood pressure (BP)—hypertension—is one of the most common complex human diseases[1]. Hypertension is a key driver of coronary heart disease (CHD) and stroke and the single most important cause of disability and premature death worldwide[2]. The individual predisposition to hypertension and its main cardiovascular complications have a strong genetic component encapsulated (at least in part) in variants uncovered by genome-wide association studies (GWAS)[3-6]. While the chromosomal landscape of >800 BP-GWAS loci is very well characterized, the identity of mediator genes and the downstream biological pathways responsible for these associations remain elusive. The majority of single nucleotide polymorphisms (SNPs) uncovered by GWAS map to non-coding segments of DNA of no apparent biological relevance to BP regulation[6]. Some of these variants operate as expression quantitative trait loci (eQTLs)[4,5], but it is becoming increasingly clear that other molecular mechanisms (i.e. DNA methylation or alternative splicing) not captured by *cis*-eQTL effects may explain a number of GWAS signals[7-11]. However, the relevance of these mechanisms to BP and hypertension has not been extensively examined within the most relevant human tissues, i.e. the kidney.

Here we have established one of the largest repositories of human kidneys using samples collected after elective nephrectomies or prior to renal transplantation[12-14]. Through the integrated analysis of their genome, transcriptome, and DNA methylome, we uncover connections between 1,038 kidney genes and 479 signals of associations to BP from previous GWAS. Using variants implicated in these analyses, we also demonstrate effects of BP on several kidney outcomes known as complications of human hypertension. We further show that some kidney gene targets of BP-GWAS variants are druggable, and we highlight those of potential relevance to treatment of hypertension. Finally, our analyses identify 179 kidney genes with putatively causal effect on BP, and map many of these genes onto novel regulatory pathways and biological processes.

Results

Kidney *cis*-expression quantitative trait locus analysis. Through integration of matched kidney genomes and transcriptomes from 430 white-European individuals (Supplementary Table 1), we generated information on 6,461,055 SNPs and 18,201 kidney genes. After correction for multiple testing, 903,870 genetic variants had at least one expression target in *cis* and 7,348 kidney genes had at least one SNP partner (kidney eSNPs and kidney eGenes, respectively) with 1,464,131 eSNP-eGene pairs (kidney eQTLs) (Fig. 1a and Supplementary Table 2). Further adjustment for inter-individual differences in cellular heterogeneity between samples (using cell-type proportions de-convolved from single-cell map of human kidney[15-17] in *cis*-eQTL analysis) had very little effect on *cis*-eQTL analysis—there was 95% consistency in kidney eGenes identity with the baseline analysis (Supplementary Table 3).

Next, we examined how 1,037 kidney eGenes identified by the Genotype-Tissue Expression (GTEx) project[18] in 65 renal cortex samples map onto our kidney eGene repository. We discovered that 787 GTEx kidney eGenes were available for a look-up in our dataset and that 669 (85%) of them were kidney eGenes in our *cis*-eQTL analysis.

We then quantified the degree of overlap in *cis*-eQTL output between our dataset and 244 glomerular and 314 tubulointerstitial transcriptomes from the NEPTUNE study[19]. Of 6,120 available kidney eGenes identified in our discovery analysis, 4,368 (71.4%) showed an overlap with eGenes in at least one of the two histological compartments of the NEPTUNE resource (Fig. 1b and Supplementary Table 4).

These data show an abundance of genetic effects on transcriptional programs operating in the human kidney. We also demonstrate the robustness of our eGene discovery in the context of cellular heterogeneity (inherent to transcriptome profiling of bulk tissue samples) and a high degree of replication in independent kidney datasets.

BP-GWAS variants and kidney gene expression programs. A total of 885 SNPs were associated with at least one of BP-defining traits (systolic BP (SBP), diastolic BP (DBP), pulse pressure (PP)) in previous GWAS[5,20-26] (Supplementary Table 5). Of those, 821

independent sentinel BP-GWAS SNPs were available in our dataset (Supplementary Table 6).

We determined that 252 (30.7%) independent BP-GWAS loci contain kidney eSNPs (Fig. 1c). We identified 418 targets of BP-GWAS kidney eSNPs (BP-GWAS kidney eGenes) forming 424 unique BP-GWAS kidney eSNP-eGene combinations (BP-GWAS kidney *cis*-eQTLs) (Fig. 1c and Supplementary Table 7). Several of the identified kidney eGenes mapped onto the classical pathways of BP regulation (*AGT*, *REN*, *ACE*, *UMOD*, *DDC*, *ADRA2B*, *GUCY1A3, PDE5A*), but a majority had no prior biological connection with hypertension.

We then compared the enrichment for kidney eSNPs in BP-GWAS SNPs versus SNPs identified in non-BP-GWAS. We reasoned that BP-GWAS variants should be enriched for kidney eSNPs given a well-established role of the kidney in hypertension[27]. We found that SBP and DBP ranked as the top traits with the most significant enrichment among 24 GWAS phenotypes examined (Fig. 1d).

Next, we explored tissue-specificity of the identified BP-GWAS kidney e-signals taking advantage of *cis*-eQTL repositories for 48 non-renal GTEx tissues. Of 251 BP-GWAS kidney eSNPs available for analysis in GTEx, 81 (32.3%) showed kidney-specific effects on the target genes (Fig. 1e and Supplementary Table 8).

We then assessed whether BP-GWAS *cis*-eQTL were specifically expressed in major kidney cell types using 41,778 cells and 27,240 genes from a single-cell dataset of human kidney[15]. We first defined 13 different cell types (Fig. 1f), largely consistent with the annotations provided in the original study[15] and clustered by histological location (Fig. 1f). We confirmed that 389 (93%) BP-GWAS kidney eGenes identified in our study were expressed in the single-cell dataset (Supplementary Table 9). Of those, 69 (18%) showed cell-type specific expression[28,29] (Supplementary Table 10). For example, *GUCY1A3* showed highest expression in juxtaglomerular apparatus cells, while *UMOD* was expressed predominantly within the loop of Henle cells (Fig. 1g and Supplementary Note).

Taken together, our results demonstrate a particular significance of kidney *cis*-eQTLs as the genetic component of BP regulation. We uncover gene expression targets for approximately one third of BP-GWAS variants and characterize their tissue specificity. Finally, through single-cell transcriptomics, we map many of these genes to specific cell types and functional compartments of the human nephron.

Alternative splicing of kidney genes and BP. Using Leafcutter and our collection of 430 renal transcriptomes, we identified a total of 241,390 intron excision isoforms (IEIs) within 50,982 intron excision clusters (IECs) (Supplementary Note). We then combined the uncovered kidney IEIs with genotypes of 6,461,055 SNPs in *cis*-sQTL analysis and detected 17,673 unique IEIs mapping onto 5,365 genes (sGenes) as kidney targets for 724,178 SNPs (sSNPs) (Fig. 2a and Supplementary Table 11). A comparable magnitude of *cis*-sQTL discovery was reported in studies of other human tissues with similar sample size[30] (Supplementary Note and Supplementary Figs. 1-3). We then overlapped the catalog of 724,178 kidney sSNPs with all BP-GWAS SNPs and their proxies to identify which of them may operate through alternative splicing mechanisms in the kidney. We found that 189 (23%) of BP-GWAS kidney loci contain sSNPs and that these BP-GWAS kidney sSNPs targeted 829 kidney IEIs (with 47,974 unique BP-GWAS kidney sQTL pairs) (Fig. 2b). A total of 828 BP-GWAS kidney sIEIs were annotated to 318 genes (BP-GWAS kidney sGenes) (Fig. 2b and Supplementary Table 12). Some of the uncovered BP-GWAS kidney sGenes had prior connection to BP regulation (*FGFR1*)[31] or kidney disease (*AKR1B10*)[32] or their alternatively spliced isoforms were linked to disease (*NDUFAF6*)[33]. However, the majority of these genes had no established physiological role in BP regulation, and 50 of them were non-coding. We noted that protein-coding BP-GWAS kidney sGenes expressed in the kidney produce an average of 6.5 protein-coding transcripts per gene—a 1.38-fold (95%CI: 1.26-1.50, $P < 1 \times 10^{-6}$) excess when compared to random sets of protein-coding genes expressed in the kidney (4.7 protein-coding transcripts per gene on average). This suggests that an average BP-GWAS kidney sGene is a potentially stronger contributor to protein

diversity in the kidney than an average gene expressed in renal tissue. We then conducted a biological annotation analysis on all BP-GWAS kidney sGenes using DAVID and found enrichment for 10 different functional categories. As expected, the top two enriched categories were directly related to alternative splicing, while others (i.e. nucleotide-binding, magnesium) replicated the categories enriched for GWAS *cis*-sQTLs in other human tissues (Fig. 2c)[34]. The enrichment for mitochondria revealed through this analysis (Fig. 2c) is in line with the increasingly recognized role for this organelle in shaping the mRNA expression programs and alternative splicing[35].

Next, we investigated the extent to which kidney sSNPs operate separately from kidney eSNPs within the BP-GWAS loci. We observed that 63 BP-GWAS loci (33.3% of all BP-GWAS loci with a splicing signature) contain variants operating exclusively as kidney sSNPs and not as eSNPs. This indicates that approximately 8% of all BP-GWAS loci associated with changes in splicing do not display a concomitant change in total gene expression in the kidney; this is slightly higher than the percentage of blood GWAS sSNPs that are not eSNPs[34]. For example, BP-GWAS variant, rs4750358, identified as the best sSNP for one of the IEIs of *BEND7,* was not associated with the total renal expression of the gene (Extended Data Fig. 1 and Supplementary Note). In line with previous studies, only a very small proportion (0.2%) of the BP-GWAS kidney sSNPs mapped directly onto known splicing sites[30,34]. However, we determined that BP-GWAS kidney sSNPs showed approximately 2.9-fold (P = 0.0086) enrichment for intron branch point locations when compared to one million permuted samples from our collection of autosomal imputed SNPs used in QTL analysis.

In summary, these data reveal that approximately one in four BP-GWAS signals is associated with alternative splicing mechanisms operating in the kidney and that a significant proportion of the relevant *cis*-sQTLs are not captured through the *cis*-eQTLs. We also demonstrate that while BP-GWAS kidney sGenes and sSNPs collectively exhibit some typical molecular characteristics expected of those involved in alternative splicing, they do not usually map in proximity to BP-GWAS sentinel variants or the classical splice sites. This suggests that a majority of genetic associations between BP and renal alternative splicing operate through more complex, subtle and distant regulatory elements and networks.

Kidney DNA methylome and BP. Our previous studies showed that the transcriptomic footprint of apparently healthy tissue collected from kidneys after cancer nephrectomies is similar to that of renal tissue samples from individuals who did not have cancer[36]. We conducted a similar type of a comparative experiment at the kidney epigenome level and found that the kidney DNA methylation profiles of samples collected from healthy (cancer-unaffected) parts of the kidney after cancer nephrectomies cluster with those from non-cancer kidney biopsies (conducted prior to transplantation) (Fig. 2d). Through the analysis of 195 available kidney DNA-wide methylation profiles, we identified 374,826 CpG sites. Of those, 32.6% mapped onto known CpG islands. Our *cis*-mQTL analysis uncovered 1,556,997 SNPs (mSNPs) as partners of 34,913 CpG sites with a total of 4,600,245 mSNP-CpG pairs (Fig. 2e and Supplementary Table 13). On average, each kidney mSNP was associated with three CpG sites, while each renal m-target (CpG) was partnered with an average of 132 mSNPs; this is most likely a reflection of LD in examined regions[37]. The median distance between the best mSNP to its CpG target was calculated at 8,648 bp, and 38.4% of mCpG sites were located within the promoter sequences of the respective genes (Fig. 2f). The best kidney mSNPs were mapped more frequently to the transcriptional start site (TSS) proximity than non-mSNPs (Fig. 2g). We also detected that a majority of kidney mCpG sites mapping onto the promoter sequences were associated with a negative effect on the respective gene's expression (Fig. 2h). This is consistent with the known role of DNA methylation in the initiation of transcription and the repressive effect of hyper-methylated promoters on mRNA expression[14,37]. We detected an overlap between kidney mSNPs and variants acting on transcriptional activity in the renal tissue—34.6% and 27.6% of kidney mSNPs were identified as kidney eSNPs and sSNPs, respectively. These values are

comparable to data from other human tissues[38]. A total of 19.8% of kidney mSNPs received both additional e- and s- annotation.

The joint analysis of all kidney mSNPs and the catalog of BP-GWAS SNPs (and their proxies) uncovered that 391 (47.6%) independent BP-GWAS loci show at least one kidney *cis*-mQTL signature (Fig. 2i and Supplementary Table 14). Of 1,204 BP-GWAS kidney CpG sites partnering with BP-GWAS mSNPs, 920 (76.4%) were annotated to 578 known genes (BP-GWAS kidney mGenes) (Fig. 2i). Several of these genes (i.e. *EDN1*)[39] are well-known contributors to BP regulation and the development of hypertension. Others were implicated in syndromes that affect the function/structure of the kidney (*WDR73*) or diseases co-existing with hypertension, i.e. diabetes (*KCNJ11*)[40]. Some were also mapped to potential drug targets; for example, cg04510874 (the CpG partner of rs4932373BP-GWAS kidney mSNP) is within the CpG island and the promoter sequence of *FES* (a gene target for naproxen—a non-steroidal anti-inflammatory medication with high BP known as one of its side effects)[6] (Extended Data Fig. 2 and Supplementary Note).

A large proportion of BP-GWAS kidney mSNPs overlapped with both BP-GWAS kidney eSNPs, sSNPs or both (Supplementary Fig. 4). Further analysis centered on common target/common variant revealed that, within 87, 63 and 30 loci, the BP-GWAS kidney mSNP shared the same identity and the same target with the BP-GWAS kidney eSNP, BP-GWAS kidney sSNP or both, respectively (Supplementary Note and Supplementary Fig. 5).

In summary, we detect an abundance of kidney DNA methylation signatures within loci associated with BP in previous GWAS and demonstrate that approximately half of them map onto regions in the TSS vicinity. We also show clustering of kidney DNA methylation with gene expression and alternative splicing across a large number of BP-GWAS loci, suggesting that these interact together in regulatory processes shaping up the cellular and clinical BP phenotype.

Insights from Mendelian randomization analyses with kidney SNPs. Previous Mendelian randomization (MR) studies reported causal effects of BP on several cardiovascular and kidney outcomes, including CHD, stroke and urinary albumin to creatinine ratio (UACR)[41]. We sought to investigate how genetic instruments for BP, partitioned on presence/absence of kidney eSNPs, sSNPs and mSNPs, would influence causal associations with these outcomes using two-sample MR[42] (Supplementary Table 15). We observed significant effects of higher SBP, DBP and PP on increased risk of CHD, ischemic stroke and heart failure, which were consistent when using genetic instruments derived from kidney SNPs or not. We also observed causal association between higher SBP and increased UACR as well as increased risk of microalbuminuria when using genetic instruments derived from kidney SNPs. Kidney-derived causal effects on these outcomes were stronger ($P < 0.05$) than those observed when considering genetic instruments derived from SNPs with no apparent functional relevance to the kidney (Supplementary Fig. 6). The effects of DBP and PP on UACR and microalbuminuria were less significant but were also stronger when using genetic instruments derived from kidney SNPs. We also detected statistically significant effects of DBP and SBP on hypertensive renal disease and of DBP on chronic kidney disease (CKD) when using the genetic instruments selected from kidney SNPs. However, for hypertensive renal disease and CKD, there was no significant difference in the causal effect size between genetic instruments derived from kidney SNPs or not. Taken together, these results suggest putatively causal effects of BP on a range of cardiovascular and renal outcomes commonly affecting hypertensive patients and the importance of BP-GWAS kidney SNPs in detecting these relationships.

BP-GWAS kidney eGenes, mGenes and sGenes as therapeutic targets. We then sought to investigate new therapeutic opportunities for hypertension through exploiting the concept of the 'druggable genome'[43] by pairing BP-GWAS kidney genes (renal targets of BP-GWAS kidney eSNPs, sSNPs and mSNPs) with pharmacological agents and chemical compounds. We identified 210 unique licensed drugs that target 49 kidney gene products

(Supplementary Tables 16-19). Each drug was then assigned to one of three tiers corresponding to positions in the drug development pipeline (Supplementary Tables 18 and 19)[44]. Several of these associations were concordant with the clinical indications. For example, *KCNJ11* was identified as a target for minoxidil and diazoxide (antihypertensive medications), while *GUCY1A3* was identified as a target for nitrates and riociguat (medications with known BP lowering effect) (Supplementary Tables 16-19).

Three of the detected gene-drug associations were clinically discordant (Supplementary Table 19). Of those, bepotastine and entinostat have observational evidence of putative effects on BP regulatory systems from pre-clinical models[45,46]. Topiramate (approved for treatment of tonic-clonic seizures and migraine prophylaxis[47]) was reported to reduce BP concomitantly with body weight in clinical studies in hypertensive and diabetic patients[48,49].

Taken together, these data show that an integration of pharmacological resources with genome, transcriptome and epigenome from tissues of key relevance to BP regulation can verify the expected gene-drug associations and generate therapeutically relevant observations on potential drug repurposing.

Colocalization of BP-GWAS and kidney QTL signals. We further deployed colocalization to explore whether the overlapping BP-GWAS signals and kidney *cis*-QTL indeed track the same genetic variant. We detected colocalization between BP and kidney *cis*-QTL in 221 (27%) of independent GWAS loci, with 358 genes implicated as potential drivers of these signals (Supplementary Fig. 7 and Supplementary Tables 20-22). In 193 GWAS loci, the signals of colocalization between BP and kidney *cis*-QTL pointed to single genes (as exemplified in Fig. 3a-c).

Through summarized chromatin states of adult human kidney derived from RoadMap Epigenomics, we then explored whether BP-GWAS kidney e/s/m-SNPs with evidence of colocalization are over-represented in functionally relevant genomic regions. When compared to 1,000 randomly selected sets of matched autosomal SNPs, the colocalized BP-GWAS kidney SNPs showed approximately 3.0-, 2.5- and 3.6-fold enrichment for transcription start sites, transcribed regions, and enhancers, respectively, while being generally depleted in transcriptionally quiescent states (Fig. 3d). A very similar pattern of enrichment emerged from the comparison of the colocalized variants to non-BP-GWAS SNPs and all BP-GWAS SNPs (Fig. 3d). We then used Combined Annotation-Dependent Depletion (CADD) scores to see whether the colocalized SNPs are enriched for deleteriousness when compared to randomly selected sets of matched autosomal SNPs and GWAS variants. This analysis revealed a statistically significantly increase in median phred-scaled CADD scores for the colocalized variants than those for the comparators (Fig. 3e).

In summary, our data demonstrate that approximately one in three BP-GWAS loci contain a colocalized kidney QTL and BP-GWAS association and that these signals are enriched for functional chromatin annotations and genetic deleteriousness.

MR analysis of kidney eGenes, sGenes and mGenes and BP. We then employed MR to further investigate putatively causal effects of renal gene expression, alternative splicing and methylation signals on BP (Supplementary Note). We uncovered 309 molecular (expression, splicing or methylation) targets (mapping to 179 kidney genes) with causal evidence of association with BP after correction for multiple testing (Fig. 3f). These MR signals were apparent in 125 BP-GWAS loci (Supplementary Tables 23-25, Supplementary Note, and Supplementary Fig. 8).

A total of 91.6% of the kidney genes showing causal association with BP were protein-coding (Supplementary Table 26). Of these, 7.8% had prior connection to BP regulation or human hypertension (Fig. 3f), and for the vast majority (78.6%), the directional effect on BP in MR was consistent with that expected from their biological function (Supplementary Table 26). For example, the reduced kidney expression of *GUCY1A3* (a mediator of nitric oxide effects on the target cells and tissues) was causally associated with increased BP in MR (Supplementary Table 26). A total of 11.2% of the genes had an

established role/association with renal physiology or pathology (Supplementary Table 26). Others represented different biological themes underpinning basic housekeeping biological processes and cellular functions, intracellular degradation of proteins, mitochondrial energy balance and different dimensions of metabolism (Fig. 3f and Supplementary Table 26).

In summary, we found that 15% of BP-GWAS signals may be driven by a kidney gene at a different level of molecular regulation (expression, alternative splicing or DNA methylation). The vast majority of these genes had no established physiological connection to BP, human hypertension or the kidney, but many play essential biological roles of yet unknown relevance to cardiovascular and renal systems. This signifies the existence of novel uncharacterized regulatory networks connecting BP and the kidney with the molecular determinants of human health and disease.

Discussion

GWAS have uncovered hundreds of loci associated with different complex traits and diseases. Yet, despite discovery of new loci, the investments in these studies have not fully matched the high expectations of the research community and the general public, partly because the causal genes and the downstream molecular pathways through which GWAS signals operate have not been uncovered. Indeed, only a few of >800 variants associated with BP in GWAS have been mapped to coding regions of the human genome and/or assigned a putative molecular mechanism[5,20-26]. By exploring the effects of BP-associated genetic variants on molecular targets operating at the intersection of the kidney transcriptome and the epigenome, our data help to bridge the existing knowledge gap between the “sequence” (discoveries from GWAS) and “consequence” (human hypertension).

We document an enrichment of transcriptionally active variants operating in the kidney for SNPs associated with BP in GWAS and the abundance of kidney genes whose expression, splicing or methylation shows putatively causal relationship with BP. We further demonstrate that many of these kidney genes are targets for drugs with proven therapeutic potential. Finally, through utilizing kidney SNPs as genetic instruments in MR, we reveal causal effects of BP on clinical kidney outcomes, including UACR, microalbuminuria and CKD. To this end, our data fill an important information vacuum driven by the absence of human renal samples in post-GWAS analyses[50], and provide evidence for the role of the kidney as the tissue mediator of common genetic effects on BP and potentially causal role of BP in the development of renal disease (Supplementary Note).

Our results demonstrate the importance of characterizing multiple molecular layers within existing repositories of human tissue for post-GWAS discoveries. Indeed, by integrating kidney gene expression data with alternative splicing and renal DNA methylation, we increased the discovery of kidney genes with evidence of a causal effect on BP by approximately two-fold. The genome-wide evidence for the role of DNA methylation in BP regulation and human hypertension has only recently started to emerge through studies in the most accessible tissues (blood), but there has been no comprehensive analysis of the kidney DNA methylome in post-GWAS BP analyses[51,52]. We also demonstrate that genetically mediated changes in kidney expression of alternatively spliced isoforms may have potentially causal contributions to BP regulation across loci identified by GWAS. Several of these BP-associated splicing isoforms overlapped with signals of expression and/or DNA methylation within the same kidney genes. This not only strengthens the evidence for the associations between these genes and BP but also indicates the existence of interactions between the kidney epigenome and transcriptome in shaping the molecular, cellular and clinical BP phenotype (Supplementary Note).

Although the vast majority of the novel kidney genes showing causal association with BP were protein-coding, 84.4% of them had no prior connection to hypertension or kidney physiology or pathology. However, many of them have very well established biological roles in human health and diseases yet without immediately obvious direct involvement of cardiovascular or urogenital systems (Supplementary Note).

In conclusion, our study has provided translation of the loci identified by GWAS of BP into specific renal RNA isoforms, methylation sites, genes and biological themes that represent plausible molecular mechanisms through which individual GWAS variants may act on the risk of hypertension. This functional interpretation of GWAS is a critical step for genomics to converge with medicine, more precise definition and classification of the disease, and the development of new therapeutic avenues for hypertension. By elucidating the molecular mechanisms of hypertension embedded in the kidney, our study will ultimately lead to advancements in patient-centered diagnostic accuracy in hypertension and new targeted strategies to BP lowering, thereby accelerating progress in precision medicine.

Acknowledgements

This work was supported by British Heart Foundation project grants PG/17/35/33001 and PG/19/16/34270 and Kidney Research UK grants RP_017_20180302 and RP_013_20190305 to M.T., National Institutes of Health grant R01 DK117445-01A1 to A.P.J., National Institutes of Health (USA) NIDDK grants R01 DK108805 and R01 DK119380 to M.G.S., British Heart Foundation Personal Chair CH/13/2/30154 and Manchester Academic Health Science Centre: Tissue Bank Grant to B.K., Medical University of Silesia grants KNW-1-152/N/7/K to J.Z. and KNW-1-171/N/6/K to W.W. T.G. acknowledges support from European Research Council - ERC-CoG-Inflammatension 726318 and European Commission/Narodowe Centrum Badan i Rozwoju, Poland - EraNet-CVD-Plaquefight. P.M. acknowledges support of British Heart Foundation grant PG/19/84/34771. D.T. acknowledges support of Medical Research Council New Investigator Research Grant MR/R010900/1. B.K. is supported by a British Heart Foundation Personal Chair. G.T. is supported by the Wellcome Trust (grant WT206194) and Open Targets. E.C.-G. is supported by a Gates Cambridge Scholarship (OPP1144). The Nephrotic Syndrome Study Network Consortium (NEPTUNE), U54-DK-083912, is a part of the National Institutes of Health (NIH) Rare Disease Clinical Research Network (RDCRN), supported through a collaboration between the NIH Office of Rare Diseases Research, National Center for Advancing Translational Sciences, and the National Institute of Diabetes, Digestive, and Kidney Diseases. Additional funding and/or programmatic support for this project has also been provided by the University of Michigan, NephCure Kidney International and the Halpin Foundation. Access to TCGA kidney and GTEx data has been granted by the NIH [approvals 50804-2 and 50805-2]. The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics funded by Wellcome Trust [grant reference 203141/Z/16/Z] for the generation and initial processing of sequencing data.

Author Contributions

These authors performed the main analytical and experimental tasks: J.M.E., X.J., X.X., S.S., and A. Akbarov. The following authors provided additional analyses and data: E.C.-G., M.T.M., C.F., H.G., H.B.T., S.P., S.C., P.R.P., I.W., E.E., M.S., Y.S., M.E., M.D., F.E., B.G., S.E., C.B., J.B., M.C., M.K., A.S.W., D.T., B.K., P.M., T.J.G., R.T.-O'K., G.T., N.J.S., A.H., M.G.S., A.P.M., and F.J.C. The following authors were involved in the collection of kidney resources: W.W., M.S., A. Antczak, M.G., R.K., J.B., E.Z.-S., J.Z., and P.B. The following authors were involved in additional sample processing and sequencing: M.D., A.N., P.R.P., I.W., and F.J.C. The following authors contributed to drafting the manuscript: J.M.E., X.J., X.X., F.J.C., A.P.M., and M.T. Overall supervision of the project: M.T. All authors reviewed and approved the accepted version of the manuscript.

Competing Interests

M.K. reports grants from NIH and Goldfinch Bio during the conduct of the study, and grants from AstraZeneca, Gilead, NovoNordisc, Eli Lilly, Janssen, Merck, Elipidera, Certa and Boehringer Ingelheim, outside the submitted work.

References

1. Beaney, T. et al. May Measurement Month 2017: an analysis of blood pressure screening results worldwide. *Lancet Glob. Heal.* **6**, e736–e743 (2018).
2. Lim, S. S. et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2224–2260 (2012).
3. Doris, P. A. The genetics of blood pressure and hypertension: the role of rare variation. *Cardiovasc. Ther.* **29**, 37–45 (2011).
4. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
5. Warren, H. R. et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**, 403–415 (2017).
6. Giri, A. et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**, 51–62 (2019).
7. Do, C. et al. Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am. J. Hum. Genet.* **98**, 934–955 (2016).
8. Do, C. et al. Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol.* **18**, 120 (2017).
9. Park, E., Pan, Z., Zhang, Z., Lin, L. & Xing, Y. The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* **102**, 11–26 (2018).
10. Smith, A. K. et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics* **15**, 145 (2014).
11. Raj, T. et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* **50**, 1584–1592 (2018).
12. Jiang, X. et al. Hypertension and renin-angiotensin system blockers are not associated with expression of angiotensin-converting enzyme 2 (ACE2) in the kidney. *Eur. Heart J.* **41**, 4580–4588 (2020)
13. Xu, X. et al. Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nat. Commun.* **9**, 4800 (2018).
14. Rowland, J. et al. Uncovering genetic mechanisms of kidney aging through transcriptomics, genomics, and epigenomics. *Kidney Int.* **95**, 624–635 (2019).
15. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
16. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
17. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).
18. Ardlie, K. G. et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
19. Gillies, C. E. et al. An eQTL landscape of kidney tissue in human nephrotic syndrome. *Am. J. Hum. Genet.* **103**, 232–244 (2018).
20. Cabrera, C. P. et al. Exploring hypertension genome-wide association studies findings and impact on pathophysiology, pathways, and pharmacogenetics. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **7**, 73–90 (2015).
21. Ehret, G. B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).

22. 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).
23. 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).
24. 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).
25. Hoffmann, T. J. et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**, 54–64 (2017).
26. Wain, L. V. et al. Novel blood pressure locus and gene discovery using genome-wide association study and expression data sets from blood and the kidney. *Hypertension* **70**, e4–e19 (2017).
27. Crowley, S. D. & Coffman, T. M. The inextricable role of the kidney in hypertension. *J. Clin. Invest.* **124**, 2341–2347 (2014).
28. Morris, A. P. et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat. Commun.* **10**, 29 (2019).
29. Wu, H. et al. Comparative analysis and refinement of human PSC-derived kidney organoid differentiation with single-cell transcriptomics. *Cell Stem Cell* **23**, 869-881.e8 (2018).
30. Saferali, A. et al. Analysis of genetically driven alternative splicing identifies FBXO38 as a novel COPD susceptibility gene. *PLOS Genet.* **15**, e1008229 (2019).
31. Han, X. et al. Cardiovascular effects of renal distal tubule deletion of the FGF receptor 1 gene. *J. Am. Soc. Nephrol.* **29**, 69–80 (2018).
32. Shaw, N., Yang, B., Millward, A., Demaine, A. & Hodgkinson, A. AKR1B10 is induced by hyperglycaemia and lipopolysaccharide in patients with diabetic nephropathy. *Cell Stress Chaperones* **19**, 281–287 (2014).
33. Hartmannová, H. et al. Acadian variant of Fanconi syndrome is caused by mitochondrial respiratory chain complex I deficiency due to a non-coding mutation in complex I assembly factor NDUFAF6. *Hum. Mol. Genet.* **25**, 4062–4079 (2016).
34. Zhang, X. et al. Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.* **47**, 345–352 (2015).
35. Guantes, R. et al. Global variability in gene expression and alternative splicing is modulated by mitochondrial content. *Genome Res.* **125**, 633–644 (2015).
36. Marques, F. Z. et al. Signatures of miR-181a on the renal transcriptome and blood pressure. *Mol. Med.* **21**, 739–748 (2015).
37. Hannon, E. et al. Leveraging DNA-methylation quantitative-trait loci to characterize the relationship between methylomic variation, gene expression, and complex traits. *Am. J. Hum. Genet.* **103**, 654–665 (2018).
38. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
39. Schiffrin, E. L., Deng, L. Y., Sventek, P. & Day, R. Enhanced expression of endothelin-1 gene in resistance arteries in severe human essential hypertension. *J. Hypertens.* **15**, 57–63 (1997).
40. Barbetti, F. & D'Annunzio, G. Genetic causes and treatment of neonatal diabetes and early childhood diabetes. *Best Pract. Res. Clin. Endocrinol. Metab.* **32**, 575–591 (2018).
41. Haas, M. E. et al. Genetic association of albuminuria with cardiometabolic disease and blood pressure. *Am. J. Hum. Gen.* **103**, 461-473 (2018).
42. Pierce, B. L. & Burgess, S. Efficient design for mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* **178**, 1177–1184 (2013).
43. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).

44. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
45. Kitaba, S., Murota, H., Yahata, Y., Azukizawa, H. & Katayama, I. Novel functional aspect of antihistamines: the impact of bepotastine besilate on substance P-induced events. *J. Allergy* **2009**, 853687 (2009).
46. Ryu, Y. et al. Class I histone deacetylase inhibitor MS-275 attenuates vasoconstriction and inflammation in angiotensin II-induced hypertension. *PLoS One* **14**, e0213186 (2019).
47. Spritzer, S. D., Bravo, T. P. & Drazkowski, J. F. Topiramate for Treatment in Patients with Migraine and Epilepsy. *Headache* **56**, 1081–1085 (2016).
48. Tonstad, S. et al. Efficacy and safety of topiramate in the treatment of obese subjects with essential hypertension. *Am. J. Cardiol.* **96**, 243–251 (2005).
49. Moradi, S., Kerman, S. R. J. & Mollabashi, M. The effect of topiramate on weight loss in patients with type 2 diabetes. *J. Res. Med. Sci.* **18**, 297–302 (2013).
50. Gamazon, E. R. et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018).
51. 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).
52. Richard, M. A. et al. DNA methylation analysis identifies loci for blood pressure regulation. *Am. J. Hum. Genet.* **101**, 888–902 (2017).

Fig. 1 | *cis*-eQTL analysis of the human kidney and genetic variants identified in genome-wide association studies of blood pressure (BP-GWAS SNPs). a, The discoveries from *cis*-eQTL analysis of 430 human kidneys. SNPs, single nucleotide polymorphisms; eSNPs, genetic variants with at least one renal expression target (eGene); eQTLs, pairs of kidney eSNPs and their target eGenes. **b**, The extent of replication for kidney eGenes identified in the analysis of 430 kidneys in an independent resource of 244 glomerular and 314 tubulointerstitial transcriptomes (top row, available for analysis; bottom row, replicated genes), data are counts and percentages (in brackets). **c**, The extent of overlap between 821 BP-GWAS sentinel SNPs (and their 26,197 proxies) and kidney *cis*-eQTL. **d**, GWAS trait enrichment in the human kidney. The enrichment estimate is expressed as a log-odds ratio (with 95% confidence intervals shown as error bars) for the association between trait-associated variants in GWAS and kidney eSNPs. Points are colored by log-odds ratio from white (lowest enrichment) to blue (highest enrichment). **e**, The proportions of BP-GWAS kidney eSNPs showing tissue-specificity; data are counts and percentages (in brackets); 1, kidney-exclusive BP-GWAS eSNPs (with expression targets operating as eGenes only in the kidney tissue); 2, kidney-exclusive BP-GWAS eSNP-eGene pairs (eGene targets are different between renal and non-renal tissues); 3, BP-GWAS eSNP-eGene pairs with kidney specific allelic direction (the direction of the association with eGene is different between renal and non-renal tissues); 4, BP-GWAS eSNPs showing no kidney specificity. **f**, t-SNE representation of cells from normal kidney tissue. Cells are colored and labelled by cell-type, as identified from gene expression of canonical markers and single cell clustering; t-SNE, t-distributed stochastic neighbor embedding. **g**, Heatmap of cell-type specific expression of BP-GWAS eGenes. Normalized gene expression is shown in shades of blue and red, from lowest expression (dark blue, 3 standard deviations below the mean) through mean expression (white) to highest expression (dark red, 3 standard deviations above the mean).

Fig. 2 | *cis*-sQTL and *cis*-mQTL analysis of the human kidney and variants identified in genome-wide association studies of blood pressure (BP-GWAS SNPs). **a**, The discoveries from *cis*-sQTL analysis of 430 human kidneys. IEIs, intron excision isoforms; sQTL pairs, pairs of single nucleotide variants and partnering renal IEIs (sIEIs); sSNPs, single nucleotide variants with at least one target IEIs in the kidney. **b**, The flowchart and outcomes from the analysis of overlap between BP-GWAS SNPs and kidney sSNPs. **c**, Network representation of 10 DAVID categories statistically enriched in BP-GWAS kidney sGenes (Fisher's exact test, all results $P_{B\text{-}H} < 0.05$). Nodes are sized according to the number of BP-GWAS sGenes that fall into that category; larger nodes contain more genes. Edges represent an overlapping percentage of genes greater than 70% between two categories, thicker edges represent greater overlap. **d**, Hierarchical clustering of Euclidean distances derived from 90 human kidney DNA methylome profiles. Blue, TRANSLATE nephrectomy samples; green, non-cancer kidney biopsies prior to transplantation; red, kidney cancer samples. **e**, The discoveries from *cis*-mQTL analysis of 195 human kidneys. CpGs, DNA methylation probes; mQTLs, pairs of single nucleotide variants and their target kidney CpGs (mCpGs); mSNPs, single nucleotide variants with at least one target CpG in the kidney. **f**, Distribution of kidney mCpGs in relation to distance to the nearest TSS. mCpGs, DNA methylation probes associated with at least one single nucleotide variant; TSS, transcription start site; blue, gene body; yellow, 3' untranslated region; green, 5' untranslated region; light blue, first exon; light orange, promoter (≤ 200bp); orange, promoter (200-1,500 bp). **g**, Comparison of distance from TSS between kidney best mSNPs and non-mSNPs (chi-squared test). mSNPs, single nucleotide variants with at least one target CpG in the kidney; non-mSNPs, single nucleotide variants with no target CpG in kidney *cis*-mQTL analysis; light blue, > 100 kb; dark blue, 10-100 kb; very light blue, 5-10 kb; blue, 3-5 kb; light green, 1-3 kb; green, 0-1 kb. **h**, Comparison between the observed and expected percentage of kidney mCpG sites mapping onto the promoter sequences associated with the negative effect on the respective gene expression (binomial test). **i**, BP-GWAS SNPs as kidney mSNPs and their DNA methylation target genes (mGenes) in renal tissue.

Fig. 3 | Colocalization and Mendelian randomization analyses. **a-c**, Examples of loci where a *cis*-eQTL, *cis*-sQTL and *cis*-mQTL signal colocalizes with the BP-GWAS signal, pointing to single genes. The *x*-axis shows the negative $\log_{10}$ *P*-value of the GWAS association (linear mixed model), and the negative $\log_{10}$ *P*-value for the QTL signal (linear regression) is shown on the *y*-axis. The most significant colocalizing SNP is shown in purple and is labelled with its refsnp ID. Other SNPs are colored by their r^2 with the most significant SNP. **d**, Heatmap of statistically significant enrichment of "best" colocalizing SNPs in summarized chromatin states from adult human kidney tissue (permutation test). Enrichment is shown in shades of red, from the least significant enrichment (pale red) to the most significant enrichment (dark red). Depletion is shown in shades of blue from the least significant depletion (pale blue) to the most significant depletion (dark blue). Non-significant results are shown in dark grey. Matched SNPs, autosomal variants matched for proximity to genes, number of LD buddies and proximal SNP density; non-BP-GWAS SNPs, a collection of non-BP-GWAS SNPs from GWAS Catalog; BP-GWAS SNPs, the catalog of BP-GWAS SNPs and its proxies. **e**, Permutation-based enrichment analysis (permutation test) for phred-scaled CADD scores (numeric measure positively correlated with the deleteriousness of a SNP). The median CADD score for the "best" colocalizing SNPs is shown as a red line. Density plots summarize 1 million permutations of the equally sized sets of SNPs drawn from Matched SNPs (yellow), non-BP-GWAS SNPs (red), and BP-GWAS SNPs (blue). **f**, Circular representation of information on 179 putative causal genes for BP. Genes are grouped by their biological theme, and this grouping is shown as colored regions. From outermost to innermost data circle: *cis*-eQTL, *cis*-sQTL and *cis*-mQTL causal association signals (colored blue, green and red, respectively), established role in the kidney (black), established role in BP regulation (black), and specific expression in one or more kidney single cell types (black if specific to one or more cell-type).

Methods

Repository of variants associated with BP in previous GWAS. We used information provided by Evangelou et al.[4] as the resource with most comprehensive and up-to-date information on variants associated with BP-defining traits in GWAS at the time the current project commenced. From this resource (generated using information from 1,006,863 participants), we selected a set of 885 sentinel single nucleotide polymorphisms (SNPs) showing association with at least one of BP-defining traits (SBP, DBP, PP) at the GWAS threshold of statistical significance ($P < 5 \times 10^{-8}$) in either one of the previous GWAS (discovery or meta-analysis) or the combined meta-analysis conducted by Evangelou et al.[4] (Supplementary Table 5). Out of the 885 sentinel variants, 822 were available for analysis in the panel of SNPs used for the purpose this project. These 822 variants were then pruned to 821 independent SNPs (based on an r^2 threshold of 0.2). Using LD information provided by Google Genomics (https://googlegenomics.readthedocs.io/en/latest/use_cases/linkage_disequilibrium/public_ld_datasets.html?highlight=linkage), we then obtained 26,197 proxies for the 821 sentinel BP-GWAS variants under the threshold of $r^2 \geq 0.80$ (Supplementary Table 6).

Populations. *The TRANScriptome of renaL humAn TissuE Study* (*TRANSLATE*) study recruited patients of white European ethnicity diagnosed with unilateral non-invasive renal cancer, eligible for elective nephrectomy and with no previous personal history of primary nephropathy[36,53]. Every participant underwent standardized phenotyping, which consisted of taking personal history (using anonymized questionnaires), height, weight, waist circumference and BP measurements[36,53]. Different types of biological material (including blood, urine and kidney samples) were secured for further biochemical and molecular analysis[36,53]. The renal tissue specimens were taken directly from the healthy (unaffected by cancer) pole of the kidney immediately after nephrectomy and immersed in RNA*later* and formalin (for the purpose of renal histology)[14]. A recent extension of the TRANSLATE study (*TRANSLATE-T*) conducted "zero time" pre-implantation biopsy from deceased donors' kidneys prior to transplantation[54], as reported before[13]. Needle biopsy samples were collected within 6–28 h after the extraction time (donation after brain death)[54] and the material from each kidney biopsy sample was then used for further molecular processing[13]. The basic clinical information about the donors was collected from available hospital documentation[13].

MoleculAr analysis of human kiDney-Manchester renal tIssue pRojEct (ADMIRE) is a resource of human kidney tissue developed in Manchester Biomedical Research Centre Biobank (MBRCB). Eligible for inclusion were individuals undergoing elective surgical removal of the kidney because of kidney cancer. Similar to the TRANSLATE Study, specimens were secured immediately after nephrectomy from healthy (unaffected by cancer) part of the kidney through immersion into RNA*later*, formalin and/or snap-freezing. Additional (blood, urine) biological materials were also collected for the purpose of biochemical analyses. The demographic and clinical information was retrieved retrospectively from the clinical files.

Renal gEne expreSsion and PredispOsition to cardiovascular and kidNey Disease (RESPOND) is an on-going project that recruits patients diagnosed with kidney cancer referred for an elective nephrectomy. The renal tissue samples were taken from the healthy (unaffected by cancer) pole of the kidney after surgical removal of the organ and processed through the pipeline developed earlier in the TRANSLATE study. Each individual provided a blood and urine sample for the downstream analyses, including standard blood biochemistry. The demographic and clinical information was collected using purpose-designed anonymized questionnaires; all patients were of white European ancestry. Basic anthropometry and blood pressure measurements were also conducted in line with the guidelines[55].

In *Molecular analysis of mechanisms regulating gene expression in post-ischemic*

injury to renal allograft (REPAIR), kidney specimens were collected from renal grafts prior to the organ transplantation. All grafts were retrieved from deceased donors after brain death and cold-stored (8.2 h to 26.8 h of cold ischemia time). Two core needle (16Gx10mm) biopsy specimens were then taken from the upper pole of the kidney and immediately fixed in RNA*later* or formalin for further analyses. Each donor was of white European ancestry; their medical history and clinical data were obtained from available medical documentation.

The Cancer Genome Atlas (TCGA) is a National Institute of Health (NIH)-funded resource of human tissues (including the kidney) collected from over 10,000 individuals with cancer. The tissue samples were collected after elective surgical procedures for the purpose of molecular studies[56]. Apart from tissue taken from the kidney tumor, a healthy (cancer-unaffected) sample was also secured as "companion normal tissue specimen" from adjacent normal renal tissue at the time of nephrectomy[57]. The biological materials were flash-frozen prior to further processing[57]. Only basic demographic information (ethnicity, age and sex) is available for patients recruited in TCGA. These samples have been used as a source of information on normal kidney transcriptome in both our and others' studies[13,36,58].

Ethical compliance. The studies adhered to the Declaration of Helsinki and were approved/ratified by the Bioethics Committee of the Medical University of Silesia (Katowice, Poland), Bioethics Committee of Karol Marcinkowski Medical University (Poznan, Poland), Ethics Committee of University of Leicester (Leicester, UK), University of Manchester Research Ethics Committee (Manchester, UK) and National Research Ethics Service Committee North West (Manchester, UK). Informed written consents were obtained from all individuals recruited (for the deceased donors, the consent was obtained in line with the local governance; e.g. from the family members).

DNA analysis – extraction, genotyping, imputation, quality control, and principal components. In TRANSLATE, TRANSLATE-T, ADMIRE, RESPOND and REPAIR studies, DNA was extracted from the secured apparently normal kidney samples (upon prior homogenization) using Qiagen DNeasyBlood and Tissue Kit. The extracted DNA was then hybridized to the Infinium® HumanCoreExome-24 beadchip array composed of 547,644 variants. Genotype calls were made using GenomeStudio[13,14].

In TCGA, DNA was extracted from blood samples using QiAAmp Blood Midi Kit (CGARN, 2016) and hybridized with probes on the Affymetrix SNP 6.0 array (composed of 906,600 probes); genotype calls were made using the Birdseed algorithm (https://www.broadinstitute.org/birdsuite/birdsuite-analysis). The TCGA genotype data were downloaded from the GDC Portal's legacy archive. A total of 525 cases/files were initially identified using the following query criteria: "project name"—"TCGA", "primary site"—"kidney", "sample type"—"solid tissue normal", "race"—"white", "data category"—"simple nucleotide variation", "data type"—"genotypes", "experimental strategy"—"genotyping array" and "access"—"controlled". We downloaded the data for 110 individuals who had matching RNA-seq-derived information on the transcriptome of normal kidney tissue.

We applied the same set of quality control filters to genotyped variants and individuals in all datasets using the following packages/software: PLINK[59], KING (cryptic relatedness)[60] as well as SNPWeights[61] and EIGENSTRAT[62] (genetic ancestry). We excluded individuals with: (i) genotyping rate < 95%, (ii) heterozygosity rate outside ±3 standard deviations from the mean, (iii) cryptic relatedness to other individuals, (iv) discordant sex information (mismatch between reported sex and genotyped sex), and (v) genetic ancestry other than white-European, as reported before. At the variant level, we excluded genetic variants with: (i) genotyping rate < 95%, (ii) genomic location mapped to the sex chromosomes or mitochondrial DNA, (iii) ambiguous genomic location, (iv) Hardy-Weinberg equilibrium (HWE) $P < 1 \times 10^{-3}$, and (v) minor allele frequency (MAF) < 5%. After quality control, there were 660,208 variants remaining for TCGA and 268,549 variants in the populations genotyped using Infinium® HumanCoreExome-24 BeadChip array.

Genotype imputations were carried out separately on the Michigan Imputation Server (MIS)[63] using 1000 Genomes Project Phase 3 data as the reference panel applied to all

genotyped variants that passed quality control. MIS uses minimac3[63] to perform imputations with the default phasing software Eagle v2.3. At the post-imputation quality control level, we excluded variants with: (i) duplicate genomic locations, (ii) imputation score < 0.40, (iii) MAF < 5%, and (iv) HWE $P < 1 \times 10^{-6}$. A total of 6,461,055 variants common for all populations remained after the imputation quality control.

Genotype principal components were derived from genotyped autosomal variants that passed all genotyping quality control filters using EIGENSTRAT[62] and SNPWeights[61].

RNA analysis – extraction, RNA-sequencing, data processing and gene expression normalization. In 76% of TRANSLATE, 16% of TRANSLATE-T, and 68% of ADMIRE samples, RNA was extracted from kidney tissue using RNeasy Kits (Qiagen). The remaining samples from these studies together with all RESPOND and REPAIR samples were subjected to an RNA extraction method using miRNeasy Mini Kit (Qiagen). Upon checking of RNA purity and integrity, 1 µg of kidney RNA extracted from each sample was processed through Illumina TruSeq RNA Sample Preparation protocol with poly-A selection. All the libraries were then sequenced using either 100-bp reads (on an Illumina HiSeq 2000) or 75-bp paired-end reads (on an Illumina NextSeq or HiSeq 4000), producing an average of 32 million paired reads and 5.5 Gb per sample. The base call and sequence quality across all generated FASTQ data was evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

In TCGA, kidney RNA was extracted from snap-frozen samples using a modification of the DNA/RNA AllPrep Kit (Qiagen, https://brd.nci.nih.gov/brd/sop/show/1450). The mRNA libraries were sequenced with 50-bp reads on a HiSeq 2000, yielding an average of 80.6 million paired reads and 7.9 Gb per sample. Data were downloaded from the GDC Portal using the following query criteria: “project name”—“TCGA”, “primary site”—“kidney”, “sample type”—“solid tissue normal”, “race”—“white”, “data category”—“raw sequencing data”, “data type”—“aligned reads” and “experimental strategy”—“RNA-Seq”. In total, 112 cases/files were identified; 103 of them had matching array-based DNA information that passed all DNA quality control filters.

In all studies, the input library complexity was assessed using RNA-SeQC[64]. The pre-processing of reads for adapter trimming was conducted by Trimmomatic[65]. The reads were then pseudoaligned to the GRCh38 Ensembl transcriptome reference (Ensembl release 83) using Kallisto[66]. The expression of genes was quantified in transcripts per million (TPM) at a transcript level using Kallisto[66]. Transcript expression values were then summed to give gene-level expression values. A gene was selected for downstream analyses if its expression in at least 20% of kidney samples within each population was > TPM of 0.1 and the read count was ≥6. Genes not meeting the above expression criteria or those with either location on sex chromosomes or whose expression interquartile range was equal to zero were excluded from further analyses. After applying these quality control filters, 18,201 renal genes (common for all datasets) were identified as suitable for further analyses. All RNA-sequenced samples were examined using several quality control filters including: (i) number of total reads (> 10 million reads), (ii) D-statistic test (a normalized measure of within tissue sample inter-correlation, D > 0.75)[67], (iii) sex compatibility check (consistency between the reported sex and gene expression sex, determined based on XIST and male-specific region of the Y-chromosome gene expression), (iv) verification of sample code based on comparing variant calls obtained from RNA-seq using GATK and DNA genotype calls, and (v) visual inspection of principal component plots of processed TPM data. The final number of samples that passed all RNA-seq sample quality control filters and had matching genotype data was 430 (332 in the non-TCGA resources and 98 in TCGA).

Gene expression data from all samples that passed quality control underwent normalization prior to any statistical analysis. The gene expression values underwent $\log_2$-transformation (logarithm to base 2 of TPM (plus an offset of 1)) followed by quantile normalization (first within each study or sequencing batch and then across studies). We used robust quantile normalization, which uses the medians of quantiles rather than the means[68]. The quantile-normalized data were then standardized using rank-based inverse

normal transformation. To account for unmeasured variation in the RNA-seq data, we used probabilistic estimation of expression residuals (PEER)[69] as reported[13]. Briefly, PEER estimates hidden factors that account for global variation in the RNA-seq, datasets[70]. The optimal number of hidden factors (119) was determined based on the smallest number that maximizes the number of eGenes in the eQTL results, a strategy recommended by GTEx[18,70-72].

Identification, quantification and normalization of intron usage. All 430 kidney samples that passed quality control filters for RNA-seq-derived gene expression quantification and had matching genotype information were included in this analysis. Of these, 332 were from non-TCGA resources and 98 from TCGA.

We used LeafCutter[73] to quantify intron usage levels. Briefly, this method defines intron excision clusters using information from intron spanning reads aligned to the human genome reference sequence. First, we used STAR[74] to align reads to the GRCh38 genomic reference sequence producing a bam file for each sample. The output bam files were then converted into intron junction files (.junc) using CIGAR string data. Intron excision clusters were selected for further analysis if within introns that: (i) contained at least 10 aligned reads, (ii) were below 500 kb in size, (iii) were located on autosomes, and (iv) were present in more than 40% of all samples. Individual introns with a standard deviation of their intron excision ratios (across all samples) of < 0.005 were removed.

To obtain the normalized intron usage ratio as the phenotype data for the *cis*-splicing QTL-mapping (*cis*-sQTL) analysis, we used the phenotype preparation script provided by LeafCutter, which computed the intron usage ratio by standardizing the intron reads across samples and quantile normalizing the standardized values across all introns. The locus of each intron was converted from GRCh38 to GRCh37 using CrossMap[75]. To capture the unmeasured confounders, we calculated the hidden factors from the normalized intron usage ratios using PEER[69]. The optimal number of PEER factors was calculated at 65, using the strategy proposed in GTEx[70]. We also applied the above computational pipeline on RNA-seq data from 355 blood samples available in GTEx for the purpose of comparing the outputs from Leafcutter-based analysis of the kidney with those conducted in another human tissue.

DNA methylation – sample and array processing, quantification, quality controls and normalization. In non-TCGA studies (TRANSLATE, TRANSLATE-T, and REPAIR), we used DNA extracted from homogenized renal tissue samples. A total of 192 available kidney DNA samples underwent first bisulphite treatment to convert cytosine to uracil while leaving 5-methylcytosine (5-mC) intact and differentiate unmethylated from methylated cytosines. Briefly, 750 ng of high-quality kidney DNA underwent bisulphite conversion using the Zymo EZ DNA Methylation Kit (Zymo Research, USA). Conversion was conducted according to the manufacturer's guidelines using the alternative incubation conditions for the Illumina methylation arrays. Bisulphite-converted kidney DNA was then hybridized with the Infinium HumanMethylation450 BeadChip (Illumina) (96 samples) or MethylationEPIC BeadChip arrays (96 samples) as per the manufacturer's guidelines. The Infinium Methylation450 BeadChip provides a comprehensive DNA methylation status of over 485,000 CpG sites spanning 99.9% of Refseq genes (with an average of 17 CpG sites per gene region distributed across the promoter, 5' untranslated region, first exon, gene body and 3' untranslated region). Apart from a vast majority of CpG islands (covered in 96%), the array provides coverage for island shores and the regions flanking them. MethylationEPIC BeadChip arrays provide coverage for over 850,000 methylation sites, including CpG islands of RefSeq genes, promoter regions of microRNAs, FANTOM5 promoters, and ENCODE enhancers as well as transcription factor binding sites. The content of the array covers 90% of CpG sites of Infinium Methylation450 BeadChip, with an additional targeting of regulatory regions characterized by FANTOM5[76] and ENCODE projects[77]. All samples were then processed through Illumina iScan to deliver rapid, high-resolution imaging of the data.

All TCGA kidney DNA samples were processed using Methylation450 BeadChip arrays. The data were downloaded from GDC's legacy portal (https://portal.gdc.cancer.gov/) using the following criteria: "project name" – "TCGA", "primary site" – "kidney", "sample type" – "solid tissue normal", "race" – "white", "ethnicity" – "not Hispanic or Latino", "data category" – "raw microarray data", "data type" – "raw intensities", "experimental strategy" – "methylation array", "data format" – "idat", "platform" – "Illumina Human Methylation 450".

The data from all populations were merged using R package minfi[78]. The extent of regional methylation was quantified in M-values; the latter corresponds to the ratio of methylated to unmethylated intensity and have a statistical advantage over β-values[79]. Samples were excluded based on the following criteria: (i) sample label mix-ups, (ii) unavailability of the matching genotype information that passed all quality control filters, (iii) the overall methylation rate < 95% (sample contains > 5% CpG probes that failed signal to noise detection, determined by detection *P*-values with a threshold of 1×10^{-16}). Of 192 non-TCGA samples (96 processed using Illumina Methylation450 BeadChip array and 96 using EPIC array), 177 samples (149 from TRANSLATE, 10 from TRANSLATE-T, 17 from REPAIR and 1 from RESPOND) passed quality control and were included in further analyses. Of 126 TCGA samples, two were excluded because of label mix-ups and 106 because their corresponding genotype data were either unavailable or failed quality control filters. The final combined dataset consisted of 195 samples (177 from non-TCGA studies and 18 from TCGA).

Of 452,567 CpG probes common for both types of arrays, excluded were those: (i) with a high detection *P*-value (an indicator of poor signal intensity and general measure of poor probe performance)[80], (ii) mapping onto the sex chromosomes, (iii) showing cross-reactivity to more than one genomic location, and (iv) overlapping with SNPs (MAF > 1%). After these quality control filters, 374,826 CpG sites were available for further analysis. All M-values were normalized using the "dasen" method from the wateRmelon R package[81]. This corrects for differences between Type I and Type II probes and applies quantile normalization within each probe type separately.

***Cis*-expression quantitative trait locus (*cis*-eQTL) analysis – discovery.** We combined 430 samples with informative genotype and transcriptome information from all studies in *cis*-eQTL analysis to investigate the effects of common genetic variation on kidney gene expression. This analysis brought together 6,461,055 genetic variants and 18,201 kidney genes common for all studies. The normalized expression of each kidney gene was regressed against alternative allele dosage, age, sex, source of tissue indicator (nephrectomy/kidney biopsy), the top three principal components derived from genotyped autosomal variants, genotyping array and 119 hidden factors estimated using PEER. Only variants within 1 Mb from the transcription start site of a gene were considered, and the analysis was carried out using *FastQTL*[82].

The correction for multiple testing was conducted in two stages. First, correction for testing each gene against its *cis* variants was conducted based on the analysis of permuted datasets the number of permutations was determined by FastQTL and was permitted to vary between 1,000 and 10,000. For each gene g, permutations were used to derive an approximate empirical distribution of the smallest *P*-value under the null hypothesis of no association modelled using a beta distribution: $P_{g,min} \sim Beta(\hat{a}_g, \hat{b}_g)$. The adjusted *P*-value for a gene g was then generated as $P_{g,adjusted} = F_{g,min}(P_{g,min} \leq P_{g,nominal}$. Secondly, correction for simultaneously testing thousands of genes was performed by applying Storey's method[83] (http://www.bioconductor.org/packages/release/bioc/html/qvalue.html) to the adjusted *P*-values to calculate false discovery rates (FDR). Genes with FDR $\leq 5\%$ were declared as eGenes (genes whose expression is regulated by at least one genetic variant). Since more than one variant can be associated with the expression of a particular gene, nominal *P*-value thresholds were derived for each eGene as $P_{g,threshold} = F_{g,min}^{-1}(P_t)$, where P_t is the adjusted *P*-value closest to FDR of 5%. All variants with nominal *P*-values below the threshold for a given gene were declared as transcriptionally active variants (eSNPs).

Overlap with independent kidney eQTL datasets. We first obtained summary statistics from *cis*-eQTL analysis conducted by GTEx in kidney cortex samples from individuals of white-European ancestry (https://storage.googleapis.com/gtex_analysis_v8/single_tissue_qtl_data/GTEx_Analysis_v8_eQTL_EUR.tar). Briefly, their *cis*-eQTL analysis was conducted using information on expression of 24,049 autosomal genes, >46 million SNPs and was adjusted for sex, sequencing protocol, sequencing platform, first 5 PCs and 15 PEER factors[18]. The correction for multiple testing was conducted by FDR with a corrected threshold of statistical significance of qval ≤ 0.05).

We then used a collection of kidney tissue from *Nephrotic Syndrome Study Network (NEPTUNE)* cohort. Briefly, kidney biopsy samples collected from patients with nephrotic syndrome were microdissected into glomeruli and tubulointerstitium[19] and underwent whole genome sequencing (WGS) and RNA-sequencing.

WGS (30x) was conducted using the Illumina HiSeq system. Alignment and variant calling was performed using default settings of GotCloud with the GRCh37 human genome reference[84]. Using VCFtools, PLINK and the HardyWeinberg R (v3.5.1) package, the following quality control filters were applied on variants[85-87]: (i) multi-allelic variants were converted to bi-allelic, (ii) variants with GQ < 20 and AB < 0.2 or > 0.8 were set to missing, (iii) variants with genotyping rate < 0.85, MAF < 0.01 and inbreeding coefficient < -0.3 were removed, and (iv) variants failing HWE ($P < 1 \times 10^{-6}$), in either European or African subsamples were removed. After quality controls, a total of 12,481,386 and 11,956,449 variants remained in glomeruli and tubulointerstitial analyses, respectively.

Total RNA from glomerular and tubulointerstitial biopsies were extracted, and libraries were prepared using the Clontech SMARTSeq v4 kit. Samples underwent sequencing using Illumina HiSeq 2500, resulting in 150-bp paired-end reads. Reads found in fastq files underwent quality control filtering and trimming using fastQC, fastQScreen (https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen), and picardtools (https://broadinstitute.github.io/picard). Trimmed reads were then aligned to the human genome (GRCh37) with STAR 2.6.0a[74]. Read counts for each sample were obtained using HTSeq version 0.9.1[88]. Read counts for each sample were initially normalized and converted to counts per million (CPM) then $\log_2$-transformed. Normalization was performed using the trimmed mean of M-values (TMM) method of the edgeR package[89]. Genes with 0.3 CPM in at least 20% of the samples were kept. This filtering resulted in 19,912 and 19,922 genes in glomeruli and tubulointerstitial compartment, respectively. Since sample normalization depends on the genes included, read counts for these selected genes were extracted and underwent normalization again using the TMM method. Finally, $\log_2$ CPM expression values were normalized across genes using rank-based inverse normalization in R. Only protein-coding RNAs and lincRNAs were considered for downstream analyses, resulting in 16,481 glomeruli and 16,435 tubulointerstitial expressed genes.

After all quality controls, a total of 244 and 314 of glomeruli and tubulointerstitial samples with matched informative genome and transcriptome were included in the *cis*-eQTL analysis, respectively. Matrix eQTL was used with a window of 500 kb from the gene boundaries adjusting for age, sex, four PCs, and PEER factors (30 in glomeruli, 45 in tubulointerstitial)[90,91], correcting for multiple testing using FDR. PCs for genetic ancestry were calculated on LD-pruned WGS data using PLINK. PEER factors were calculated with the PEER framework adjusting for age, sex, and batch. We used Bayesian fine mapping and eGene discovery through the TORUS/DAP pipeline described previously[19], accounting for distance to the genes TSS and LD. A total of 4,324 eGenes (those with at least one variant in *cis* with FDR < 0.05) were identified in the glomerulus and 6,951 in the tubulointerstitium. We then examined (i) proportion of kidney eGenes identified in the discovery resource that is expressed in the replication dataset and (ii) the overlap between the kidney eGenes uncovered in the discovery sample (available for lookup) and the eGenes identified in the replication datasets.

Analysis of cellular heterogeneity in the human kidney transcriptome and its effect on the *cis*-eQTL analysis. To account for heterogeneity resulting from variance in cell-type proportions between kidney samples, we used computational deconvolution combined with single-cell renal gene expression profiling. Briefly, single-cell gene expressions from 3,448 normal kidney cells were first obtained from Young et al.[15]. These profiles were generated from FACS-sorted single-cell suspensions of 30 mm^3 kidney tissue samples; cDNA libraries were created by the 10X Genomics Chromium single-cell platform and sequenced on an Illumina HiSeq4000. The expression profiles were then normalized, clustered and cell types identified using the Seurat R package[16]. For the purpose of this analysis, we identified 10 distinct cell-type clusters in these cells. We then used non-negative least squares multivariate regression on the single-cell data to determine cell-type specific gene weightings (by prioritizing genes that show stable expression across individuals as well as within cell-type clusters) using MuSiC R package[17]. We applied these gene weightings to all our bulk tissue sample gene expression profiles and de-convolved the cell-type proportions in each sample. We conducted the sensitivity *cis*-eQTL analysis using the parameters specified above plus 10 variables illustrating proportions of cell types in each of 430 kidney samples as additional covariates in the regression models. We then estimated an effect of adjustment for inter-individual differences in cellular heterogeneity on the outputs from our *cis*-eQTL analysis (through measuring the percentage of initially identified kidney eGenes that retained their significant associations with the partner eSNPs after the adjustment).

Kidney eQTL GWAS traits enrichment analysis. To investigate whether the identified kidney eQTLs are enriched for variants identified before in BP-GWAS, we used a computational framework known as TORUS[92]. Briefly, the enrichment calculated by TORUS is the logistic regression coefficient for association between absolute *Z* scores from a given GWAS and the probability of genetic variants being causal eQTLs. The coefficient is expressed on log-odds scale and is used to rank GWAS traits within a specific tissue; it is not sensitive to the sample sizes of the compared GWAS traits[92]. The positive enrichment indicates that variants with stronger evidence for association in a GWAS (higher absolute *Z* scores) are more likely to be causal eQTLs in the tissue of interest. The sources of information on 25 GWAS used in this analysis are listed in Supplementary Table 27.

Analysis of tissue-specificity of the kidney BP-GWAS eQTLs. We compared the results of our kidney BP-GWAS *cis*-eQTL studies against the outputs generated by *cis*-eQTL analyses across 48 non-kidney GTEx tissues to explore the renal tissue specificity of the uncovered BP-GWAS kidney eQTL signals[18,70-72,93]. A BP-GWAS kidney eQTL was considered as tissue-specific if it fulfilled one the following criteria: (i) kidney-exclusive BP-GWAS eSNPs (variants with the expression targets operating as eGenes only in the kidney tissue), (ii) kidney-exclusive BP-GWAS eSNP-eGene pairs (variants whose eGene targets are different between renal and non-renal tissues), and (iii) BP-GWAS eSNP-eGene pairs with kidney specific allelic direction (variants whose eGene partner overlap between the kidney and other tissues but the allelic direction of association with a given eGene is different between renal and non-renal tissues).

Single-cell analysis of BP-GWAS kidney eGenes. Single-cell RNA-seq data for the human kidney was obtained from Young et al.[15] and imported into R. This study profiled 72,501 cells from kidney samples, including those of cancer origin, but for the purpose of this analysis only cells derived from normal kidney tissue were retained in the dataset (41,778 cells). Cell-level quality control included: removing cells with less than 200 or more than 8,000 unique molecular identifiers (UMIs), as well as those with more than 20% of UMIs coming from mitochondrial genes. Gene expression values were next normalized to the total number of UMIs per cell, log-transformed, and scaled using standard Seurat parameters[16]. This was followed by PCA using the highest variable genes in the dataset. Based on the percentage of variance explained, we determined that the first 30 PCs captured most of the

variability in the data. Thus, we embedded the first 30 PCs in two dimensions using t-distributed stochastic neighbour embedding (tSNE) with a perplexity value of 30[94].

In order to identify distinct kidney cell types, we performed unsupervised clustering using Seurat's sNN algorithm. We used the first 30 PCs as input to sNN, and a resolution of 0.6. Based on this approach, we identified 23 distinct cell clusters. We then combined the cluster specific gene expression with known markers of immune and kidney cell types reported in the literature. This led to collapsing the 23 clusters into 13 cell types of known renal and immune identity and expressing canonical cell type markers. We then verified that these cell types also clustered by histological location in the tSNE space.

Next, we calculated the proportions of kidney eGenes and BP-GWAS kidney eGenes (identified at previous stages of the project) in the single cell-type dataset. For BP-GWAS kidney eGenes with confirmed presence in the single cell dataset, we calculated the average expression in each cell-type (normalized UMIs per cell) (Supplementary Table 9). We then examined if any of these eGenes showed cell-type specific expression. Thus, we intersected our kidney eGenes with the computational markers identified for each cell-type. To define these markers, we used Seurat with previously verified criteria: (i) expression in > 25% of the total cells in the given cell-type; and (ii) log-fold change in expression > 0.25 in the specified cell-type when compared to all other cell-types[28,29].

To determine if there was enrichment for cell type-specific genes in any particular cell type, we compared all kidney cell type-specific BP-GWAS kidney eGenes to 1,000 numerically equivalent random sets of non-BP-GWAS kidney genes and determined the proportion of cell type-specific genes across all 13 different renal cell types. The statistical significance of this difference was examined using the chi-squared test.

***Cis*-splicing quantitative trait locus (*cis*-sQTL) analysis in kidney and blood.** The *cis*-sQTL analysis was performed in FastQTL using normalized intron usage ratios calculated by LeafCutter[73] from 430 kidney samples and a common panel of 6,461,055 genetic variants. The *cis*-region was specified as ±1 Mb from the middle point of each intron. All the regression analyses (whereby normalized intron usage ratio was regressed against alternative allele dosage) were conducted under additive model and adjusted for age, sex, source of tissue indicator (nephrectomy/kidney biopsy), genotyping array, the top three genotype PCs from autosomal DNA, and 65 hidden factors estimated using PEER[69]. The multiple testing correction followed the principles of the two-stage strategy used in *cis*-eQTL studies. Briefly, at the first stage we calculated the corrected *P*-value by applying the 'qvalue' method[83] to the Beta-approximated permuted *P*-value from FastQTL (in agreement with that used in *cis*-eQTL analysis). Introns with adjusted FDR < 5% were declared *cis*-splicing intron excision isoforms (sIEIs, introns whose usage ratios is regulated by at least one variant in *cis*). Clusters with at least one sIEIs were declared *cis*-splicing intron excision cluster (sIECs), and genes overlapping at least one sIEC were declared *cis*-sGenes (sGenes). The genetic variants associated with sIEIs at the level of significance below the nominal *P*-value threshold (calculated by the inverse quantile Beta-approximation) were declared as sSNPs.

In the absence of publicly available *cis*-sQTL resources using LeafCutter from other human tissues, we conducted *de novo* analysis of 355 GTEx blood samples using the strategy reported above. Included in this analysis were 39,459,042 common genetic variants derived from WGS and 41,933 IECs containing 198,715 IEIs generated by LeafCutter from RNA-seq reads provided as BAM files. The *cis*-sQTL analysis was conducted in FastQTL with the window parameters used in the kidney resource and adjusted for age, sex, top three genotype PCs, and 54 hidden factors estimated using PEER[69]. The correction for multiple testing was conducted in agreement with that used in *cis*-sQTL analysis of the kidneys.

Functional characterisation of kidney sSNPs and sGenes. *Comparative analysis of BP-GWAS kidney sGenes with expressed kidney genes.* We examined the average number of protein-coding transcripts per gene for 254 protein-coding BP-GWAS sGenes and 10 million randomly selected sets of 254 expressed protein-coding kidney genes. We quantified the latter through counting the number of times the mean number of protein-coding transcripts

from each permutation exceeded the mean for the BP-GWAS sGenes. The statistical significance of this difference was calculated by dividing the number of times the BP-GWAS sGenes mean was exceeded by the number of permutations performed. This analysis was performed in R using the "future.apply" package (https://CRAN.R-project.org/package=future.apply). The 95% confidence interval of the observed enrichment was calculated as the 2.5% and 97.5% percentiles of the distribution of enrichment values calculated from 1 million bootstrap resamplings of the original data.

Functional categories annotation to BP-GWAS kidney sGenes. We used DAVID to examine whether BP-GWAS kidney sGenes are enriched for functional annotations[95]. The latter is a web-based interface that stores and calculates enrichment for a wide variety of gene annotations. We submitted 318 BP-GWAS kidney sGene symbols (276 were matched by DAVID) and used all human genes as the background comparator. The fold-enrichment were calculated by a Fisher's exact test and the correction for multiple testing was calculated using the Benjamini-Hochberg method.

BP-GWAS kidney sSNPs and intron branch points – enrichment analysis. We compared 9,326 BP-GWAS kidney sSNPs with 1 million permuted samples of 9,326 SNPs drawn from our set of imputed autosomal SNPs (as used in the QTL analyses). We counted the number of SNPs that overlapped with 63,734 branch point locations from Supplementary Table 1 of Taggart et al.[96] in each of our SNP groups. Intron branch points are one of the key sequence-based regulators of splicing outside of intron-exon junctions[96]. The statistical significance of the difference in branch point overlap was calculated as the proportion of permuted samples that contained more branch point overlaps than the BP-GWAS kidney sSNPs divided by the number of permutations performed (1 million). This analysis was performed in R using the "future.apply" package (https://CRAN.R-project.org/package=future.apply).

BEND7 mRNA isoform quantification. In order to accurately quantify the BEND-002a and BEND7-002b transcripts, we created a custom Kallisto[66] transcriptome based on the standard Ensembl v83 transcriptome but with a single manual addition of the BEND7-002b cDNA sequence. The Kallisto[66] gene expression quantification process was then repeated exactly as previously described for the standard RNA-sequencing but using the custom transcriptome instead of the standard Ensembl v83 transcriptome. BEND7 transcript abundances were then extracted and summarised for all samples.

BEND7 analyses in silico. BEND7-2a and BEND7-2b exons were extracted from the Ensembl[97] release 97 ENSG00000165626 entry. Both sequences of concatenated exons were computationally translated using all possible translation frames (https://web.expasy.org/translate/)[98], and the most likely ORFs were selected. The translated BEND7-2a protein sequence was confirmed to be identical to the Ensembl ENSP00000367868.3 sequence. BEND7-2a and BEND7-2b protein sequences were analysed with InterProScan (https://www.ebi.ac.uk/interpro/beta/)[99,100] in order to identify functional domains. PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/)[101], Quick2D (https://toolkit.tuebingen.mpg.de/tools/quick2d)[102] and NetSurfP (http://www.cbs.dtu.dk/services/NetSurfP/)[103] were used to analyze the physicochemical properties of the protein sequences, including their propensities for secondary structure elements and disordered regions. The absence of signal peptides and transmembrane helices was confirmed with SignalP (http://www.cbs.dtu.dk/services/SignalP/)[104], TargetP (http://www.cbs.dtu.dk/services/TargetP/)[105], TMHMM (http://www.cbs.dtu.dk/services/TMHMM/)[106] and TOPCONS (http://topcons.cbr.su.se/pred/)[107]. Finally, NetPhos (http://www.cbs.dtu.dk/services/NetPhos/)[108,109] was used to identify possible phosphorylation sites in the protein sequences as well as the protein kinases responsible for the attachment of the phosphoryl groups.

Hierarchical clustering of Euclidean distances derived from human kidney DNA methylome profiles of different origin. We examined kidney methylome of 90 samples from three different sources of human renal tissue: (i) apparently normal tissue collected

from cancer-unaffected pole of the kidney after cancer nephrectomy (nephrectomy group, *n* = 57), (ii) non-cancer kidney samples from biopsies conducted prior to the transplantation (biopsy group, *n* = 27) and (iii) kidney cancer (cancer group, *n* = 6). Kidney DNA extracted from these samples was processed using the biochemical and bioinformatic pipelines established for methylome profiling and reported above. We then calculated Euclidean distances between all kidney DNA methylation profiles across all three groups of samples using the "dist" function in R. Hierarchical clustering of these values was performed by R (https://www.r-project.org/) and visualized using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree).

***Cis*-methylation trait locus (*cis*-mQTL) analysis in the kidney.** Our *cis*-mQTL analysis was performed on 195 samples kidney DNA samples by FastQTL[82]. We used normalized M-values (as the dependent variable) and genotype information for all genotyped and imputed variants surviving the quality control filters as an independent covariate under an additive mode of inheritance. Included in the regression model as additional covariates were age, sex, genotyping array, source of tissue indicator (nephrectomy/kidney biopsy), the top three PCs derived from genotyped autosomal variants (genotype PCs), and six PCs derived from the methylation array control probes (methylation PCs) as the independent variables. The latter set of PCs was derived from the PCA of 220 Illumina HumanMethylation450 array control probes based on the approach of Lehne et al.[110] and Schulz et al.[111] (to account for unmeasured confounding). The FastQTL *cis*-region was specified as ±1 Mb from each tested CpG position. The correction for multiple testing was conducted using the principles similar to those used in the *cis*-eQTL analysis. We defined a significant CpG as the DNA methylation site associated with at least one SNP in *cis* after the correction for multiple testing (FDR < 0.05). Given that more than one variant can be associated with a particular CpG, nominal *P*-value thresholds were derived for each significant CpG using the same method as for identifying eGenes. All variants with nominal *P*-values below the threshold for a given CpG were declared as mSNPs. We mapped all significant CpG sites to the closest gene using annotation from the IlluminaHumanMethylation450kanno.ilmn12.hg19 R package (https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylation450kanno.ilmn12.hg19.html) and defined genes that the significant CpG was mapped to as kidney mGenes.

Functional characterization of kidney DNA methylation targets. *Positional characteristics of kidney mSNPs and CpG sites under genetic control of mSNPs.* We characterized the distribution of CpGs partnering with mSNPs across six categories of location in relation to the closest gene (gene body, 3' untranslated region, 5' untranslated region, first exon, proximal promoter (≤ 200 bp), distal promoter (200 bp-1,500 bp)) using information from the Illumina annotation file. We compared the differences in the categories pertaining to the distance from transcription start site (> 100 kb, 10-100 kb, 5-10 kb, 3-5 kb, 1-3 kb, 0-1 kb) between kidney best mSNPs and non-mSNPs. The statistical significance of the difference was assessed by Fisher's exact test.

Analysis of the effect of kidney mCpG sites in promoter sequences on renal gene expression. We examined how the location of mCpG sites in promoter sequences affects the expression of kidney genes. First, we generated residuals of methylation and expression of the mapped gene from regression models adjusted for age, sex, source of tissue indicator (nephrectomy/kidney biopsy), top three PCs derived from genotyped autosomal variants, genotyping array, and corresponding hidden factors from the respective *cis*-QTL analysis. Second, the residuals of gene expression were regressed on the residuals of methylation. We only examined genes whose association between methylation and expression survived the statistical significance after correction for multiple testing (FDR < 0.05). Then, we applied the binominal test to compare the statistical significance of the distribution in directionality of association between promoter methylation and kidney gene expression (positive versus negative).

Analysis of potentially causal effects of BP on clinical outcomes using kidney SNPs as instruments. We obtained DBP, SBP and PP association summary statistics for lead BP SNPs from the ICBP Genetics Consortium[4]. We performed a lookup of association summary statistics for these SNPs across a range of clinically relevant cardiovascular and kidney outcomes from public resources. Cardiovascular outcomes included: coronary heart disease (60,801 cases and 123,504 controls, published data from the CardiogramplusC4D Consortium[112]), heart failure (ICD10 I50, 5,901 cases and 446,363 controls, extracted from UK Biobank using GeneATLAS[113]) and ischemic stroke (10,307 cases and 19,326 controls, published data from the MEGASTROKE Consortium[114]). Kidney outcomes included: CKD (12,385 cases and 104,780 controls, published data from the CKDGen Consortium[115]), eGFR from creatinine (133,814 individuals, published data from the CKDGen Consortium[115]), eGFR from cystatin C (33,152 individuals, published data from the CKDGen Consortium[115]), hypertensive renal disease (ICD10 I12, 1,663 cases and 450,601 controls, extracted from UK Biobank using GeneATLAS[113]), UACR (46,061 individuals without diabetes mellitus, published data from the CKDGen Consortium[116]) and microalbuminuria (55,390 individuals, published data from the CKDGen Consortium[116]).

We performed two-sample MR to identify putatively causal effects of each BP trait on each cardiovascular and kidney outcome. The lead BP SNPs were not in LD with each other, so that their effects on BP traits and cardiovascular and kidney outcomes were uncorrelated. Analyses were performed separately for two sets of genetic instruments; (i) all kidney eSNPs, mSNPs and sSNPs, and (ii) all other SNPs with no apparent functional relevance to the kidney. For each analysis, we first accounted for heterogeneity in causal effects via modified Q-statistics[117], implemented in the R package RadialMR, which identified outlying SNPs that were removed from the genetic instruments to reduce the impact of pleiotropy. Our primary MR analysis was then performed using inverse variance weighted (IVW) regression[118], implemented in the R package TwoSampleMR[119]. We also assessed the evidence for causal association using two additional approaches that are less sensitive to heterogeneity (although less powerful) and implemented in the R package TwoSampleMR[119]: weighted median regression[120] and MR-EGGER[121]. Differences in causal effects from the IVW meta-analysis between instruments with or without functional relevance in kidney were assessed using a *Z*-test. The significance threshold was set at $P < 0.00093$ based on Bonferroni correction for three exposures, nine outcomes and two instruments (i.e. 54 tests).

Druggability analyses. We identified 479 informative genetic variants partnering with 918 protein-coding kidney e-, s- and mGenes. We then mapped the proteins encoded by these genes onto a set of licensed drugs or compounds with bioactivities against these targets *in silico*. Briefly, gene annotations from Ensembl version 79 were extracted and filtered followed by identification of drug targets from ChEMBL 20 and DGIdb (accessed June 2019). We stratified druggable genes into three tiers corresponding to position in the drug development pipeline, as per Finan et al.[44]. Tier 1 included efficacy targets of approved small molecule and bio-therapeutic drugs as well as clinical-phase candidates (e.g. *KCNJ11*, *EDNRA* and *ADRA2B*). Tier 2 comprised genes encoding targets with known bioactive drug-like small molecules binding partners, as well as those with ≥ 50% identity (over ≥ 75% of the sequence) with approved drug targets (e.g. *PDE1A*, *CDK14* and *TERT*). Tier 3 contained genes encoding secreted or extracellular proteins, proteins with more distant similarity to approved drug targets, and members of key druggable gene families not already included in Tier 1 and Tier 2 (GPCRs, nuclear hormone receptors, ion channels, kinases and phosphodiesterases). Any drugs not falling into one of the three tiers were excluded from further analysis. We conducted manual classification of all drugs with an *in silico* evidence for targeting kidney eGenes/mGenes/SGenes into the following categories: (i) disease association and treatment indication precisely concordant; (ii) disease association and treatment indication concordant within the same disease area; (iii) disease association corresponding to a mechanism-based adverse effect; (iv) disease association with a known

biomarker of therapeutic efficacy that can also be responsible for mechanism-based side-effects; (v) discordant disease association and target indication considered to imply a potential repurposing opportunity; (vi) discordant disease association and target indication considered to imply either a repurposing opportunity or mechanism-based side effect depending on the direction; and (vii) unclassified (no clear relation between gene expression and blood pressure). Information on drug indication and adverse effects were obtained from Drug Bank[122], CheMBL 21[123], first drug bank database (http://www.fdbhealth.co.uk) and Side Effect Resource (SIDER) database[124]. The analyses of directionality were conducted when there was uncertainty about whether discordance represented a repurposing opportunity, or an unrecognized mechanism-based adverse effect. Directionality plots of gene expression against BP were plotted and the effect of drug on the gene was determined using DGIdB database[125]. If a drug inhibited a gene associated positively with BP then it represented a repurposing opportunity, while if a drug was an agonist for such a gene it was considered to represent a mechanism-based adverse effect, and *vice-versa*.

We also conducted two-sample MR (IVW method) to determine whether discordant drugged genes in categories v and vi may potentially have a causal influence on BP. Each MR analysis of the putative causal effect of a drugged renal gene on BP was based on a selected set of independent ($r^2 < 0.2$) instrumental SNPs showing a significant ($P < 0.05$) association with the renal gene in a respective *cis*-eQTL analysis. The summary statistics for outcome (BP) were derived from the meta-analysis summary of the associations between the instrumental SNPs and SBP/DBP, calculated from UK Biobank and ICBP data across 750k individuals. Nominal $P < 0.05$ was considered as the level of statistical significance in two-sample MR results.

Colocalization analysis. Using the R package *moloc*[126], we examined pairwise colocalization between BP defining traits and kidney *cis*-QTLs. We defined the test regions as genetic regions with a distance of 200 kb on each side of the sentinel BP-GWAS SNPs. In each of these test regions, as an input into the analysis we used summary statistics (estimated SNP effects and their standard errors) for all overlapping SNPs included in both BP-GWAS and kidney *cis*-QTL (eQTL, sQTL or mQTL) analyses. As the method requires dense genotype information, regions with sparse (≤ 30) SNPs were excluded from the analysis. The prior probability of a SNP being causal to BP and to a kidney QTL were both set to 1×10^{-4} by default. The prior probability of this SNP being a shared causal variant for both BP and a QTL trait was set to 1×10^{-5}. Therefore, the overall probability of a SNP that is causal to BP also being causal to a QTL trait is 0.1; this is also true for the inverse situation. On the basis of the posterior probabilities (PP), regions with PP(1 shared or 2 distinct causal SNP(s)) > 0.9 and PP(1 shared causal SNP):PP(2 distinct causal SNPs) > 3[127] were considered as showing evidence for colocalization if the best causal SNPs were also the BP-GWAS SNPs or their proxies ($r^2 > 0.8$). Within each BP-GWAS locus, we then identified and selected best independent colocalization signals through filtering out those that are a result of correlation between the tested clinical phenotypes (between SBP, DBP and PP) or molecular traits (i.e. between CpG sites or sIEIs) or those that do not map to any known coding or non-coding genes. We then classified the colocalization signals as those pointing out to single or numerous genes. We visualized the colocalization using an R package LocusCompareR (https://github.com/boxiangliu/locuscomparer). For each test region, in addition to a Manhattan plot for each trait, this software also provides $-\log_{10}(P\text{-value})$ plots for each pair of traits. SNPs were marked in different colours depicting the strength of LD (r^2) between the lead SNP and other SNPs.

***In silico* functional annotations of variants implicated in colocalization studies.** Gene biotype, location and description data for all expressed kidney genes was obtained from Ensembl BioMart release 83 (https://www.ensembl.org/biomart). To functionally characterize colocalized BP-GWAS kidney QTL SNPs, we first generated the 15-state chromatin segmentation in adult kidney tissue using ChIP-seq signal data for four different histone modifications (H3K4me1, H3K4me3, H3K36me3, H3K9me3) from Roadmap Epigenomics

GEO Series GSE19465. The input bed files were binarized (using the background input signal) and combined into a single chromatin state segmentation using ChromHMM[128] following the standard Roadmap Epigenomics protocol[129] for the 15-state segmentation. The 15-state model file from Roadmap (http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state) was used for the final segmentation by ChromHMM. We grouped the 15 chromatin states into 6 summarized states using a combination of histone mark specificity (Supplementary Table 28) and the 15-state clustering as reported by Roadmap Epigenomics in Extended Data Figure 2[129]. The chromatin state annotations for colocalized BP-GWAS kidney SNPs were compared to (i) 1,000 sets of matched autosomal SNPs, (ii) all SNPs from our catalogs of BP-GWAS SNPs and their proxies, (iii) a collection of 202,511 GWAS SNPs for non-BP traits from GWAS catalog. Each summarized state was tested individually for each of the comparators. The enrichment fold-change and level of statistical significance were determined using Fisher's exact test, and *P*-values were corrected for multiple testing by FDR[130]. We used phred-scaled CADD scores extracted from the v1.4 GRCh37 CADD release to assess enrichment of functional potential in our colocalizing BP-GWAS SNPs. CADD score enrichment was calculated by permutation; first, we calculated the median phred-scaled CADD score for all colocalized BP-GWAS SNPs; second, we calculated the number of times this value was exceeded in 1 million randomly sampled and equally sized sets of SNPs drawn from each of our three comparators (1,000 sets of matched autosomal SNPs, BP-GWAS SNPs and non-BP-GWAS SNPs). The *P*-value was calculated as the number of times the colocalized BP-GWAS SNP median was exceeded by the comparator median, divided by the number of random permutations performed (1 million). This analysis was performed in R using the "future.apply" package (https://CRAN.R-project.org/package=future.apply).

Analysis of causal effect of BP-GWAS kidney genes on BP using MR. We used two-sample MR to identify BP-GWAS eGenes, sGenes and mGenes (collectively referred to as xGenes) with putatively causal effect on BP. MR is a common approach to causal inference; the analytical design mimics a randomized controlled trial, where exposure associated SNPs are used as instruments to investigate if the exposure is causally associated with an outcome of interest[131]. In our MR analysis, kidney molecular targets (those emerging from renal expression, splicing and methylation analyses) were exposures of interest. Mutually independent SNPs (eSNPs, sSNPs and mSNPs) that were associated with the respective kidney molecular targets were selected as instruments. BP-defining traits (SBP and DBP) were used as the outcomes. Included in these analyses were genes that fulfilled the following criteria: (i) evidence of colocalization between BP and kidney *cis*-QTL analyses and (ii) molecular targets mapped to known protein-coding or non-coding kidney genes.

We applied several MR methods; the IVW method and penalized weighted median method[132] were used to quantify putatively causal effects of kidney molecular targets on SBP and DBP. Each MR analysis of the putative causal effect of a kidney molecular target on BP was based on a selected set of independent ($r^2 < 0.2$) instrumental SNPs showing a significant ($P < 0.05$) association in a respective *cis*-QTL analysis. We excluded kidney targets with fewer than 15 available and valid instrumental SNPs. As an input to each MR analysis we used: (i) meta-analysis summary of the associations between the instrumental SNPs and SBP/DBP, calculated from UK Biobank and ICBP data across 750K individuals[4]; and (ii) summary of the associations between the same SNPs and the respective kidney molecular targets generated in kidney *cis*-QTL studies. Point estimates and standard errors were calculated for each method separately. MR-Egger regression was then used to detect horizontal pleiotropy[132]. As an additional sensitivity analysis, we evaluated individual instruments (e-, s- and mSNPs) for potential violation of the assumptions of our MR analyses by using the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) method. MR-PRESSO is capable of detecting and removing instruments that show evidence of horizontal pleiotropy or may otherwise be acting as outliers[133]. The

correction for multiple testing was calculated using the Benjamini-Hochberg FDR, and the threshold of corrected statistical significance was established at FDR < 0.05.

We set the following criteria for the indication of a positive finding of causality: (i) causal effect estimates from the two main MR methods (the IVW method and penalized weighted median) must be significant after a correction for multiple testing, (ii) there is no evidence of horizontal pleiotropy in MR-Egger and (iii) signals identified through (i) and (ii) must be significant in MR-PRESSO after a correction for multiple testing.

To examine the relationships between kidney DNA methylation, alternative splicing and gene expression in a specific locus with multiple signals of causality mapping onto one gene (i.e. FES), we further used MR Steiger tests[134] to infer the direction of putatively causal effect between molecular traits (gene expression, CpGs and sIEIs) and then conducted robust IVW to quantify a potentially causal effect on signals identified by MR Steiger tests. Each MR Steiger test was based on independent (r^2 < 0.2) SNPs significantly associated (P < 0.001) with a kidney molecular target and each robust IVW test was based on a set of independent SNPs significantly associated (P < 0.05) with a kidney molecular target. Bonferroni-adjusted P-value was used to correct for multiple testing. MR Steiger tests were performed using the R package TwoSampleMR[119]. All other MR analyses were conducted using the R package MendelianRandomization[135].

Biological characterisation of kidney genes with a causal effect on BP. All genes with evidence of a causal effect on BP underwent initial functional characterization via manual inspection of data available on PubMed, OMIM and GeneCards. Based on the revision of the collected information, each gene was assigned into one of the key biological themes (acid-base balance, cell cycle, cell growth and proliferation, cellular architecture, cell-cell contact, ciliogenesis, intracellular signalling, intracellular processing and degradation of proteins, immunity, metabolism, mitochondria and energy metabolism, sympathetic nervous system, vascular tone/sodium reabsorption, transcription). Each gene was also classified using the following categories: (i) prior evidence of connection to blood pressure/hypertension, (ii) prior evidence of role in kidney physiology/pathology, (iii) evidence of specificity to single kidney cells. For genes with prior evidence of relevance to BP regulation/hypertension, we further examined whether the direction of expected effect on BP/hypertension based on the available information is consistent with that observed in our eQTL dataset.

Data availability

The data supporting the results presented in this article are available either in the Supplementary Information (Supplementary Tables, Supplementary Note) or can be obtained from the authors upon reasonable request. The normalized gene expression, splice junction usage and DNA methylation data are archived at the Dryad digital repository (https://doi.org/10.5061/dryad.15dv41nvx). The e-, s- and mQTL summary statistics are available in the Supplementary Tables.

Code availability

Our studies make use of well-established computational and statistical analysis software and these are fully referenced in the Methods. All custom code used to orchestrate these analyses are available on request.

Methods-only references

53. Tomaszewski, M. et al. Renal mechanisms of association between Fibroblast Growth Factor 1 and blood pressure. *J. Am. Soc. Nephrol.* **26**, 3151–3160 (2015).
54. Naesens, M. Zero-time renal transplant biopsies. *Transplantation* **100**, 1425–1439 (2016).
55. Mancia, G. et al. 2013 ESH/ESC guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Eur. Heart J.* **34**, 2159–219 (2013).
56. Chang, K. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
57. Creighton, C. J. et al. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
58. Ko, Y. A. et al. Genetic-variation-driven gene-expression changes highlight genes with important functions for kidney disease. *Am. J. Hum. Genet.* **100**, 940–953 (2017).
59. 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).
60. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
61. Chen, C. Y. et al. Improved ancestry inference using weights from external reference panels. *Bioinformatics* **29**, 1399–1406 (2013).
62. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
63. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
64. Deluca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
67. Wright, F. A. et al. Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–437 (2014).
68. Bengtsson, H., Neuvial, P. & Lun, A. Aroma-light: light-weight methods for normalization and visualization of microarray data using only basic R data types. doi: 10.18129/B9.bioc.aroma.light
69. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
70. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
71. Saha, A. et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* **27**, 1843-1858 (2017)
72. Chiang, C. et al. The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692-699 (2017)
73. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
74. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
75. Zhao, H. et al. CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).

76. Lizio, M. et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).
77. Siggens, L. & Ekwall, K. Epigenetics, chromatin and genome organization: Recent advances from the ENCODE project. *J. Intern. Med.* **276**, 201–214 (2014).
78. Fortin, J. P., Triche, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**, 558–560 (2017).
79. Du, P. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
80. Aryee, M. J. et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
81. Pidsley, R. et al. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).
82. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
83. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
84. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).
85. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
86. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
87. Graffelman, J. Exploring diallelic genetic markers: the HardyWeinberg package. *J. Stat. Softw.* **64**, 1–23 (2015).
88. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
89. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
90. Stegle, O., Parts, L., Durbin, R. & Winn, J. A bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput. Biol.* **6**, 1–11 (2010).
91. Parts, L., Stegle, O., Winn, J. & Durbin, R. Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genet.* **7**, e1001276 (2011).
92. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann. Appl. Stat.* **10**, 1619–1638 (2016).
93. Carithers, L. J. et al. A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv. Biobank.* **13**, 311–319 (2015).
94. van der Maaten, L. J. P. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15,** 3221-3245 (2014).
95. Huang, D. W. et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8**, R183 (2007).
96. Taggart, A. J. et al. Large-scale analysis of branchpoint usage across species and cell lines. *Genome Res.* **27**, 639–649 (2017).
97. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
98. Artimo, P. et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–W603 (2012).
99. Jones, P. et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

100. Mitchell, A. L. et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).
101. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
102. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
103. Klausen, M. S. et al. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins Struct. Funct. Bioinforma.* **87**, 520–527 (2019).
104. Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
105. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).
106. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).
107. Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407 (2015).
108. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).
109. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633-1649 (2004).
110. Lehne, B. et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* **16**, 37 (2015).
111. Schulz, H. et al. Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat. Commun.* **8**, 1511 (2017).
112. Nikpay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
113. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
114. Malik, R. et al. Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. *Neurology* **87**, 1306–1306 (2016).
115. Pattaro, C. et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.* **7**, 10023 (2016).
116. Teumer, A. et al. Genome-wide association studies identify genetic loci associated with Albuminuria in diabetes. *Diabetes* **65**, 803–817 (2016).
117. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* **47**, 1264–1278 (2018).
118. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
119. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
120. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
121. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
122. Wishart, D. S. et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

123. Gaulton, A. et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).
124. Kuhn, M., Letunic, I., Jensen, L. J. & Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–D1079 (2016).
125. Cotto, K. C. et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* **46**, D1068–D1073 (2017).
126. Giambartolomei, C. et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018)
127. Guo, H. et al. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum. Mol. Genet.* **24**, 3305–3313 (2015).
128. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
129. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
130. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser.* B **57**, 289–300 (1995).
131. Smith, G. D. & Ebrahim, S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *Br. Med. J.* **330**, 1076–1079 (2005).
132. Burgess, S., Bowden, J., Dudbridge, F. & Thompson, S. G. Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization. arXiv:1606.03729v2 (2016).
133. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
134. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
135. Yavorska, O. O. & Burgess, S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).