New mechanistic insights from alignment of accessible chromatin and renal lineage factor DNA-binding with genetics of kidney function

Johannes Schödel

johannes.schoedel@uk-erlangen.de

University Erlangen https://orcid.org/0000-0002-3587-8839

René Krüger

University Erlangen

Victoria Lauer

Uniklinikum Erlangen https://orcid.org/0009-0000-7102-3048

Nora Feldker

Uniklinikum Erlangen

Klaus Stark

University Regensburg

Mathias Gorski

University of Regensburg https://orcid.org/0000-0002-9103-5860

Thomas Winkler

University Regensburg

Fulvia Ferrazzi

University Hospital Erlangen https://orcid.org/0000-0003-4011-4638

Mario Schiffer

University Hospital Erlangen https://orcid.org/0000-0002-8414-1470

Stephanie Naas

Uniklinikum Erlangen https://orcid.org/0000-0001-9301-6631

Iris Heid

University Regensburg

Article

Keywords:

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8000899/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

Genome-wide association studies for kidney function identified hundreds of genomic signals overlapping non-coding regions. Prioritizing regulated target genes is key to clarify mechanisms and discover novel therapeutic routes into kidney function preservation.

Comprehensive alignment of variants credibly associated with estimated glomerular filtration rate with accessible chromatin across 29 human tissues revealed substantial overlap with regulatory DNA-elements in the liver and the kidney. Variants were enriched in regulatory elements bound by the lineage-transcription factor HNF-1β in renal tubules, resolving several association signals to their likely regulated genes. Our work revealed a polymorphic *NDRG1* enhancer that differentially interacted with HNF-1β and regulated cell growth and migration of human tubule cells.

Our study follows genetic association to the cellular mechanisms in humans and underscores the utility of intersecting genetic information of common traits with lineage-specific transcription factor binding in a tissue-specific way for determining regulatory mechanisms.

Introduction

Kidney function depends on multiple factors including age, comorbidities, and genetic predisposition. Recent genome-wide association studies (GWAS) discovered hundreds of genetic signals linked to estimated glomerular filtration rate (eGFR), a common approach to assess kidney function[1–5]. Most genetic association signals for eGFR include multiple variants inherited jointly generating the need for prioritizing variants within a given signal to help identify the underlying molecular mechanism. Although Bayesian fine-mapping approaches quantify the statistical evidence for a variant to be likely causal for a signal, biological data are needed to help reveal mechanisms of action. While functional follow-up of variants with clear effects on protein-coding DNA-sequences is comparably well established, the majority of signals overlap with non-coding genomic regions. Together with the presence of an immense number of expression quantitative trait loci (eQTL), this implicates that variants exert effects on gene transcription through modulating the accessibility or activity of regulatory DNA-elements (RE)[1,6]. Unraveling their mechanisms remains a substantial challenge. Accessibility of DNA-elements and thus transcriptional regulation of genes depend on the underlying tissue and cell-type[7]. In principle, it is possible that genetic variation at open chromatin in one or more tissues outside the kidney have an influence on kidney function. However, recent GWAS reports have particularly underlined the importance of tubule cells within the kidney for renal function[6,8]: genes at loci harboring eGFR-associated variants are more expressed in tubule cells compared to other cell types in the kidney[8]. Furthermore, recent work attributed more than 50% of germline genetic effects on kidney function to the tubular system, mediated by affecting chromatin accessibility and gene regulation[6].

The clustering of eGFR-associated genetic variation and cell-lineage specificity may modify tubule-specific expression programs governed by lineage transcription factors. One such potential factor is

hepatocyte-nuclear factor 1β (HNF-1β) which is expressed predominantly in tubule cells of the kidney and cells of the liver, the gastrointestinal tract, and the pancreas[9]. HNF-1β is known to control the expression of genes necessary for proper nephrogenesis, electrolyte handling and renal metabolism[10,11]. In line with this, patients with mutations in *HNF1B* present with severe kidney phenotypes including autosomal-dominant tubule-interstitial kidney disease, and polymorphisms at the *HNF1B* gene associate with kidney function[1,12–14].

Here, we aimed to reveal underlying regulatory mechanisms of genetic associations relevant for kidney function. Since many REs are shared between different tissues and cell-types[15,16], we were interested in how heritability of kidney function intersected with chromatin accessibility in tissues across the human body. Furthermore, we aimed to understand the gain from statistical prioritization of eGFR-associated variants by intersecting them with accessible chromatin of the kidney and renal tubule cells. Finally, we sought mechanistic explanations for eGFR-associated variants in REs by alignment with direct DNA-binding of the lineage-transcription factor HNF-1β in tubule cells.

Results

Cross-tissue analysis revealed overlap of accessible chromatin and eGFRcrea-associated variants across organs

We hypothesized that non-coding genetic variants require open chromatin for functional effects on gene regulation. We defined accessible chromatin in 29 human tissues across organs using publicly available ATAC-seq data sets and applying the summit method including fixed-width peaks according to Corces et al.[17] This resulted in a range of 28,779 (spleen) to 61,840 (right lobe of liver) accessible sites per tissue (**Supplementary Table 1**)[5,15,18]. Overlap with ENCODE defined candidate cis regulatory elements (cCRE) was ≥ 95% for each tissue (**Supplementary Table 1**). We intersected these sites with eGFR-associated variants based on creatinine measurements (eGFRcrea) in 95%-credible sets derived from a recent GWAS meta-analysis including Bayesian fine-mapping of more than 1.2 million individuals of European ancestry[1,19]. The 35,881 credible variants contained in the 424 loci reflected 594 independent signals associated with eGFRcrea. In total, 3,304 variants co-incided with 2,369 potential REs across tissues which corresponded to 471 of the 594 signals overlapping with accessible chromatin in at least one tissue (**Supplementary Fig. 1a**). We determined a median of 593 variant carrying RE per tissue (range from 334 in the spleen to 877 in the right lobe of liver, **Supplementary Table 2**).

More than 95% of the eGFRcrea-associated variants are located in non-protein coding regions. In a first step, we tested whether the presence of variants in accessible chromatin is linked to their statistical evidence to be potentially causal for the association signal as determined by the posterior probability of association (PPA) from fine-mapping analysis. To do this, we first prioritized variants with a PPA > = 2.5% from signals with < 1000 variants [19]. Of these 4,250 PPA-filtered eGFRcrea-associated variants (12% of

credible set variants), 698 variants from 307 signals overlapped with REs across tissues (**Main** Fig. 1a). From these 698 variants, 272 overlapped with REs in only one tissue and 9 variants overlapped REs in all of the 29 tissues (**Main** Fig. 1b, **Supplementary Table 3**). Of note, PPA-filtered variants more frequently overlapped REs compared to variants with low PPA (< 2.5%) or from signals with > 1,000 variants, with the highest proportion being detected for liver, kidney and muscle; variants with low PPA (< 2.5%) or from signals with > 1,000 variants overlapped more uniformly across the tissues (**Main** Fig. 1c). Clustering analysis revealed that REs with PPA-filtered variants were grouped in data sets from tissues of similar origin (e.g. gastrocnemius vs. psoas muscle or lower lobe of left vs. upper lobe of right lung, **Main** Fig. 1d). Enrichment of PPA-filtered variants was strongest in accessible chromatin of liver, kidney and muscle (**Main** Fig. 1e, **Supplementary Fig. 1b**). The number of variants located in accessible chromatin (score per million (SPM)-value > 3.5) in only one tissue was highest for kidney and liver (**Main** Fig. 1f, **Supplementary Fig. 1c, Supplementary Table 3**). For example, the variant rs2364368 next to the *ALDH1L1* gene was located in open chromatin of the liver and rs11624512 at an intronic site in *RIN3* was in accessible chromatin solely in the adrenal gland (**Main** Fig. 1g). We repeated the analyses using open chromatin data sets comprising the top 50,000 REs according to the SPM value for each tissue. This resulted in comparable findings regarding the preferential overlap of PPA-filtered variants with REs in liver, kidney and muscle, supporting the robustness of findings (**Supplementary Fig. 2**).

Thus, our analysis showed that eGFRcrea-associated genetic signals overlap with accessible chromatin across various tissues. Top enrichment was detected in tissues with a clear link to kidney function or creatinine metabolism. Filtering for variants with increased statistical probability of being a causal variant sharpened the enrichment in accessible chromatin in relevant tissues. We thus restricted to PPA-filtered variants in the following unless stated otherwise.

Additional biomarkers for kidney function refine detection of candidate regulatory elements in the kidney

The previous analyses intersected accessible chromatin with all eGFRcrea-defined variants. Since creatinine metabolism in tissues such as liver and muscle may bias serum measurements of creatinine irrespective of kidney function, we refined our analyses by incorporating available genetic association with alternative kidney function biomarkers cystatin C based eGFR (eGFRcys) and blood urea nitrogen (BUN) to screen for kidney function relevant genomic sites[1]. When focusing on eGFRcys and/or BUN-validated variants and signals, liver and kidney were the tissues with the highest number of variants and signals in REs in general or tissue-specific (**Main** Fig. 2a, b, **Supplementary Fig. 3**). These findings would be in line with an intrinsic effect of kidney function genetics by affecting chromatin in kidney tissue.

To validate our results regarding the kidney, we analyzed enrichment of kidney REs with PPA-filtered variants within open chromatin compartments defined by DNase I hypersensitive sites in 733 biosamples[20]. Of the 16 organ compartments of open chromatin, the tissue invariant, primitive/embryonic, organ development/renal, renal/cancer as well as digestive compartments

enriched most for kidney REs with PPA-filtered variants (**Main** Fig. 2c). Next, we annotated kidney REs with PPA-filtered variants to the nearest gene and performed an overrepresentation-analysis (ORA) for gene expression using cell type signature gene sets (C8) from MsigDB. This revealed that genes neighboring kidney REs with PPA-filtered variants were expressed predominantly in kidney epithelial cells, specifically in the thick ascending limb of the loop of Henle or proximal tubules (**Main** Fig. 2d). Thus, kidney REs with PPA-filtered variants align with accessible chromatin in the kidney and gene expression programs from tubule cells.

Regulatory elements in kidney tubule cells enrich for eGFR GWAS variants with elevated PPA

Tubule cells represent the dominating cell-type in the kidney. Proximal tubule cells account likely for more than 50% of total kidney mass[21]. Our findings above and recent work suggest that genetic effects on kidney function are localized to a great part in the renal epithelium[6,8]. To further explore enrichment of eGFR-associated variants in tubule cell REs, we generated ATAC-seq from freshly isolated primary tubule cells (PTC) from un-diseased parts of kidneys of seven patients undergoing tumor nephrectomy. We defined accessible chromatin as outlined above and compiled a landscape of accessible chromatin in PTC (68,405 REs, **Supplementary Table 4**). These PTC REs clustered closely with REs in kidney tissue indicating similar sites of accessible chromatin (**Main** Fig. 2e). The PTC REs contained 188 PPA-filtered variants (858 variants among all 35,881 eGFRcrea variants). 44% of the corresponding REs were also accessible in kidney tissue (**Main** Fig. 2f, **Supplementary Table 5**).

Again, we were interested the gain by prioritizing eGFRcrea-associated variants by statistical fine-mapping evidence (PPA) and alternative biomarker validation regarding the overlap with REs in kidney tissue or PTC. For this, we removed 1,004 variants predicted to affect protein-coding sequences or splicing (CADD, **Supplementary Table 6**)[19,22]. When evaluating sets of eGFRcrea-associated variants with different thresholds of PPA-filters, we observed an increase of the proportion of RE-positive variants to total variants by increasing PPA-threshold (**Main** Fig. 2g). This also highlighted the 100% PPA variant rs77924615 in *PDILT* and known for association with *UMOD* expression, that resided in accessible chromatin in kidney tissue and PTC (**Main** Fig. 2g). When further separating eGFRcrea-associated variants into those with or without eGFRcys-/BUN-validation the proportion of variants in REs in kidney tissue or PTC to total number of variants increased by PPA substantially for eGFRcrea/eGFRcys/BUN-validated variants; the proportion was near zero for eGFRcrea-only variants (**Supplementary Fig. 4**). Thus, the higher the PPA the more likely are kidney function relevant variants located in REs of kidney tissue and PTC.

Only a fraction of accessible chromatin is actively involved in gene regulation. To pinpoint potential target genes regulated from active REs more precisely, we extended our analyses by ChIP-seq for H3K27ac, a marker of active chromatin, and RNA-seq experiments in three of the seven PTC isolates. We proceeded with an activity-by-contact analysis of putative promoter-enhancer interactions (ABC)[23]. The

ABC model integrates available transcriptomic and epigenetic data such as H3K27ac with background information on chromatin interactions from a large set of cell-types to define putative promoter-enhancer contacts. In the three data sets from PTC we retrieved 66,595, 64,203, and 64,511 putative promoter-enhancer interactions, of which 80% overlapped between the three individual data sets (Naas S. et al. manuscript in revision). The 4.7 to 4.9 enhancer contacts per gene were in the range of the initial ABC analyses by Fulco et al.[23] In the consensus of all three ABC data sets, we retrieved 196 eGFRcrea-associated variants across 29 GWAS-signals corresponding to 273 putative target genes(**Supplementary Table 7, 8**). For example, on chromosome 20 the variants rs6062343 resided in an ABC-positive site that has putative interactions with promoters of the two genes *TCEA2* and *RGS19* (**Main** Fig. 2h, **Supplementary Fig. 5**).

Our results underscore that, integrating epigenetic and transcriptomic data from PTC further helped to narrow down the number of candidate REs and putative target genes relevant for kidney function.

Open chromatin from the kidney and PTC enriches for tubulointerstitial eQTLs

In a first approach to screen for transcriptional dependencies of eGFRcrea-associated variants, we aligned the variants within open chromatin from different tissues with eQTL data from the GTEx consortium[24]. We identified 566 out of the 698 variants from PPA-filtered set of variants in any tissue RE as eQTL in any tissue (81%, total 2,839 of all 3,304 variants in RE, **Supplementary Table 9**). However, for kidney, this analysis was limited by a low number of eQTLs in GTEx.

We therefore used published eQTL information of tubulointerstitial and glomerular compartments for exploring transcriptional effects of the PPA-filtered eGFRcrea-associated variants in REs of kidney tissue and PTC[25]. Within the tubulointerstitial compartment, we identified 87 eQTLs mapping to 133 variant-gene pairs in kidney RE and 81 eQTLs with 152 variant-gene pairs for PTC REs, with 45 variants in the overlap (including rs77924615 near *UMOD*, **Supplementary Table 10**). For the glomerular compartment, we identified 25 eQTLs with 28 eQTL variant-gene pairs for kidney REs and 26 eQTLs with 33 variant-gene pairs for PTC REs, with 8 variants in the overlap (**Supplementary Table 10**). This emphasizes that accessible chromatin in tubule cells preferentially contains eGFRcrea-associated tubulointerstitial eQTLs rather than glomerular eQTLs. This is not only supportive that accessible chromatin and eQTLs are linked in the same compartment but also that the mechanism of the eGFRcrea-associated variant is regulatory of the mapped target gene. A prominent example is again, the rs77924615, an eGFRcrea-associated variant being known to modify *UMOD* expression, that presents here as eQTL for *UMOD* in a region of accessible chromatin in the kidney and PTC.

We also intersected the eQTL information from the tubulointerstitial compartment with the ABC-information from the PTC above. Across the three ABC datasets, we identified nine variants within consensus ABC regions for each of which target genes were supported by an eQTL in tubulointerstitium

(**Main Table 1**). For example, for rs6062343, located in the intronic region of the *TCEA2* gene, consensus ABC signals and tubulointerstitial eQTL was detected for the downstream gene *RGS19* (**Main** Fig. 2h, **Supplementary Fig. 5**).

We integrated the epigenetic and associated eQTL data for the different tissues in an updated version of the Kidney GenePriorisations (KidneyGPS) web application (https://kidneygps.ur.de/gps/)[19].

REs from kidney tissue and PTC with eGFR-associated variants preferentially contain the HNF-1β motif

Having defined candidate regions with accessible chromatin in kidney tissue and PTC for mediating genetic effects on kidney function, we next sought to discover transcription factors involved in differential regulation of putative target genes. In kidney REs with PPA-filtered eGFRcrea-associated variants, we determined strong enrichment for motifs of hepatocyte-nuclear factors such as HNF-1β and HNF-4α using HOMER (Hypergeometric Optimization of Motif EnRichment) motif discovery (**Main** Fig. 3a, **Supplementary Table 11**). HNF-1β is predominantly expressed in tubule cells of the kidney, hepatocytes, cells of the pancreas, and epithelial cells of the gastrointestinal tract[9]. In line with this, we found the HNF-1β motif also significantly enriched in REs with PPA-filtered variants from other tissues such as liver, colon, or pancreas (**Main** Fig. 3b). In the kidney, HNF-1β is necessary for accurate nephrogenesis and tubule function[11]. Mutations in *HNF1B* can cause autosomal-dominant tubulointerstitial kidney disease and maturity onset diabetes of the young 5 (MODY5) underlining the crucial function of this transcription factor for the detected tissues[9,13,14]. To substantiate our finding from the motif search in kidney tissue, we also analyzed motifs REs from the PTC ATAC-seq data set harboring PPA-filtered eGFRcrea-associated variants. Again, the HNF-1β motif was significantly enriched (**Supplementary Table 11**). Moreover, eGFR variant-positive REs derived from a published kidney snATAC-seq also enriched for the HNF-1β motif, especially in cells from proximal tubule (**Supplementary Table 12**)[26]. Taken together, our analyses revealed that HNF-1β motifs were overrepresented at DNA elements harbouring eGFRcrea-associated variants. This gives rise to the hypothesis that genetic variation governs activity of this transcription factor at certain kidney REs, and by this affects kidney function.

eGFRcrea-associated variants enrich at HNF-1β DNA-bindings sites in human tubule cells

Given the significant concentration of HNF-1β motifs in the above-mentioned data sets, we set out to validate a regulatory mechanism by defining true HNF-1β DNA-interactions in human tubule cells. We used cultures of PTC from five individuals and employed CUT&Tag technology using a specific HNF-1β antibody[27]. This revealed robust binding to the known HNF-1β target *FXYD2*, which is involved in renal magnesium handling, across the five samples (**Main** Fig. 3c)[28,29]. Peak calling revealed 3,691 HNF-1β

binding sites overlapping across all five experiments (between 7,429 to 23,954 sites per PTC culture, **Supplementary Fig. 6a, Supplementary Table 13**). We used this PTC data for further analyses. First, we validated the binding-sites by generating independent data for HNF-1β DNA-interactions using chromatin-immunoprecipitation (ChIP-seq). This resulted in 3,539 binding sites with a robust overlap > 50% to the peaks from the individual CUT&Tag-seq experiments, e.g. at the *FXYD2* locus (**Main** Fig. 3c **and data not shown**). Second, motif analyses confirmed significant enrichment of the HNF-1β consensus motif in HNF-1β CUT&Tag sites (**Main** Fig. 3d). Third, we used ATAC-seq data from PTC and different tissues to evaluate overlap with HNF-1β binding sites. As expected, HNF-1β binding sites matched well (98%) with accessible chromatin in PTC (**Main** Fig. 3e). Furthermore, among the different tissues, the most robust overlap with accessible chromatin was seen with kidney (59%), followed by liver (41%) and pancreas (33%) (**Main** Fig. 3f). This overlap appears plausible given the restricted expression pattern of HNF-1β in the body (kidney, liver, pancreas). Fourth, we annotated HNF-1β binding sites to the nearest gene and performed 16 compartments enrichment and ORA analysis using cell-type signature gene sets from MSigDB. This revealed strongest enrichment of HNF-1β binding sites in accessible chromatin from the renal/cancer compartment and at genes that reflect expression profiles of renal tubules, validating the relevance of the corresponding genes for tubular function (**Main** Fig. 3g, h). Given the strong association of HNF-1β motifs in kidney REs harboring PPA-filtered eGFR-associated variants, we hypothesized that some of these variants align with true binding of HNF-1β. Among the 3,691 HNF-1β binding sites we observed 64 sites harboring 112 eGFRcrea-associated variants, of which 42 were also contained in the PPA-filtered set of variants (**Supplementary Table 14**). Notably, REs with HNF-1β binding included the variant rs77924615 at the *PDILT/UMOD* locus, rs10783193 at the *ABLIM* locus, and multiple variants at the *KLHDC7A* locus (**Main** Fig. 3i, **Supplementary Table 14**). Regulation of *UMOD* by HNF-1β has been described in mouse models of polycystic kidney disease[30]. Of note, the rs77924615-associated site was the only relevant signal for HNF-1β DNA-interactions within this genomic region rendering it an excellent candidate to confer transcriptional regulation of *UMOD* by HNF-1β (**Supplementary Fig. 6b**).

In general, we could determine a significant 4.5-fold overrepresentation of PPA-filtered variants at HNF-1β binding sites when compared to variants with lower PPA and at loci with > 1000 variants (χ^2 test, $p = 4x10^{-17}$). REs overlapping HNF-1β binding sites significantly enriched for PPA-filtered eGFRcrea-associated variants to approximately 2.2-fold (χ^2 test, $p < 0.001$) compared to REs with no HNF-1β binding signal (**Main** Fig. 3j). Although enrichment of variants at REs with HNF-1β binding was only slightly stronger to enrichment at REs containing an HNF-1β motif, the presence of both, HNF-1β motif and DNA-binding, increased the probability of an eGFR variant within the respective RE substantially. Similar to the results from the accessible chromatin analysis (Fig. 3f) and confirming the relevance of these sites for specific tissues, variant-carrying and HNF-1β-associated REs from PTC co-incided preferentially with accessible chromatin in the kidney (**Supplemental Fig. 6c**).

In summary, we have generated an HNF-1β DNA-binding map in human renal tubule cells confirming direct DNA-binding at plausible target genes such as *FXYD2*. Furthermore, we detected significant

enrichment of DNA-binding of HNF-1β to REs harboring eGFR variants, e.g. at the *PDILT/UMOD* locus.

A polymorphic DNA-element containing variant rs10283362 guides HNF-1β mediated NDRG1 expression in tubule cells

To unravel the mechanistic role of HNF-1β in regulating genes mapped to eGFR-associated variants, we further inspected the relation of eGFRcrea-associated variants with HNF-1β binding sites and HNF-1β motifs. One variant was very prominent across evaluations: the variant rs10283362 at the *NDRG1* locus (i) resided in HNF-1β binding sites across all five samples, (ii) directly interfered with a putative HNF-1β motif, (iii) had the highest PPA (15%) within the corresponding association signal and (iv) was the only variant out of nine with PPA > 2.5% in this signal which overlapped with accessible chromatin and HNF-1β binding in PTC (**Main** Fig. 4a)[1]. The T-allele of rs10283362 is associated with higher eGFR at $p < 5x10^{-8}$ (eGFRcrea and eGFRcys-/BUN-validated; 16% allelic frequency)[19] and corresponds to a DNA sequence reflecting the HNF-1β consensus motif (**Main** Fig. 4b). An snATAC-seq analysis linked the associated element to expression of *NDRG1*[6]. Similarly, our ABC analysis suggested an interaction of the *NDRG1* promoter with this site in PTC from all three individuals (**Main** Fig. 4a, **Supplementary Fig. 7**). *NDRG1* is highly expressed in proximal tubule cells and its upregulation is involved in kidney protective pathways[31,32]. In our data, HNF-1β binding to the RE was robust across the five different CUT&Tag experiments and the ChIP-seq experiment (**Supplementary Fig. 8a**). In addition to accessible chromatin in PTC, the kidney and the proximal tubules from a snATAC-seq experiment (**Supplementary Fig. 8b**)[33], chromatin at this site was also open in other tissues (e.g. liver and muscle), but SPM values indicated a kidney-preference of accessibility (**Main** Fig. 4c).

Next, we confirmed an involvement of HNF-1β in regulating endogenous *NDRG1* expression by performing siRNA experiments against HNF-1β in the renal proximal tubule cell lines HKC-8 and HK-2, which led to a downregulation of NDRG1 mRNA and protein (**Main** Fig. 4d-g). To explore the role of the putative enhancer further, we performed reporter assays using DNA sequences PCR-amplified from the different alleles. Luciferase activity was highest in HK-2 and HKC-8 cells transfected with constructs harboring the T-allele compared to the C-allele and vector alone transfections (**Main** Fig. 4h, i). To confirm an involvement of the transcription factor in this assay, we used siRNA to knock-down HNF-1β in HKC-8 cells transfected with the different reporter constructs. Depleting cells of HNF-1β led to a significant reduction of reporter activity preferentially in cells transfected with constructs harboring the T-allele (**Main** Fig. 4j). From this, we hypothesized that the switch from C to T may affect HNF-1β DNA interactions at this enhancer leading to an imbalance in *NDRG1* expression. To follow this up, we analyzed the allelic frequency of rs10283362 within the reads of the different CUT&Tag-seq experiments from cells of four individuals heterozygous for rs10283362 (**Main** Fig. 4k). This revealed a preference for the T-allele within the sequences interacting with HNF-1β. Performing allele-specific qPCR analyses on CUT&Tag samples confirmed this preference for HNF-1β interactions and the active chromatin marker H3K27ac (**Main** Fig. 4l, m). Since the T-allele associates with higher eGFR in GWAS and higher NDRG1 mRNA levels in tubulointerstitium (NEPTUNE;p = 0.00016, KidneyGPS)[25], we resorted to our 141 PTC

isolates, genotyped DNA for rs10283362, and measured NDRG1 mRNA levels for genotype-expression analysis. We determined 97 C/C, 41 C/T and 3 T/T genotypes which correspond well to the frequencies determined in European populations (1000 Genomes Project Phase 3; C/C: 0.658, C/T: 0.306, T/T: 0.036)[34]. Stratifying *NDRG1* expression levels by rs10283362 genotype revealed that cells harboring at least one T-allele have significant higher NDRG1 mRNA levels, confirming a previous study (**Main** Fig. 4n)[6]. Thus, our data give a plausible explanation of how rs10283362 interferes with gene regulation and align this variant to one of the most prominent lineage pathways in the kidney, HNF-1β.

The NDRG1 enhancer promotes cell proliferation and directional migration of tubule cells

The ABC data, eQTL analyses, and HNF-1β knock-down experiments suggested *NDRG1* as the target of the variant-carrying and HNF-1β binding enhancer. To proof this, we targeted the HNF-1β motif within the RE using CRISPR/Cas9 technology in pools of PTC, which resulted in decreased NDRG1 mRNA and protein levels (**Supplementary Fig. 9**). Using the proximal tubule cell line HK-2, we generated single clones of enhancer-knock out and control cells (**Main** Fig. 5a-c). We confirmed efficient DNA knock-out by Sanger sequencing (**Supplementary Fig. 10**). qPCR and immunoblot experiments indicated significantly reduced NDRG1 mRNA and protein levels in the enhancer knock-out clones of HK-2 cells (**Main** Fig. 5b, c). HNF-1β and H3K27ac CUT&Tag-seq experiments confirmed reduced interactions of HNF-1β and diminished chromatin activity at the variant-associated site in enhancer knock-out clones compared to control cells (**Main** Fig. 5d, **Supplementary Fig. 11**). Interactions of HNF-1β and chromatin activity at the independent site on chromosome 11 (*FXYD2*) was not affected (**Main** Fig. 5d, **Supplementary Fig. 11**). We also performed RNA-seq experiments from lysates of cells from the two enhancer knock-out clones and the two control clones of cells to screen for downstream transcriptional effects (**Main** Fig. 5e, **Supplementary Table 15**). Of note, maneuvers at the *NDRG1* enhancer resulted in downregulation of *NDRG1* but no other gene within 500kb upstream or downstream indicating selectivity of the enhancer-promoter interaction which was also suggested by the ABC-model (**Main** Fig. 4a, **Supplementary Table 16**). Metascape analysis of significantly regulated transcripts (p-adjust < 0.05, log2fold change ≥ 0.5/≤0.5) revealed enrichment of pathways related to cell proliferation, cell-cell contacts as well as pathways implicated in organ development (e.g. kidney, heart, lung) and epithelial tube morphogenesis (**Main** Fig. 5f, **Supplementary Table 17**)[35].

From these results, we speculated that manipulation of the enhancer might affect tubule cell properties related to proliferation and migration. First, we performed cell proliferation assays which revealed that enhancer knock-out clones of cells grew significantly slower compared to control clones of cells over a period of 72h (**Main** Fig. 5g). Second, we conducted a barrier assay testing cell migration as a surrogate of wound repair capacity. Enhancer knock-out clones revealed a significant delay in migration compared to control clones of cells (**Main** Fig. 5h, I, **Supplementary Fig. 12a**). Using a monitoring system, we followed the tracks of 60 single cells per clone until the 12h time point at which differences in gap closing were greatest (**Main** Fig. 5i). We observed that accumulated and Euclidian distance as well as velocity of the enhancer knock-out clones of cells were reduced compared to control cells (**Main** Fig. 5j-m, **Supplementary Fig. 12b**). Furthermore, the knock-out clones displayed a significant defect in

directional migration (**Main** Fig. 5n). Confirming the overall results from enhancer knock-out clones and highlighting *NDRG1* as the effector gene, we observed comparable effects in a clone of cells, in which we had targeted the exon 2 of *NDRG1* by CRISPR/Cas9 (**Supplementary Fig. 13**). Thus, our results from the functional assays indicate defective proliferation and directional migration in the knock-out clones of cells and support findings from the pathway analyses. In addition to regulation of *NDRG1* expression, we can ascribe pivotal roles of cell growth and motility to this enhancer and, by this, were able to follow human genetic association evidence to a plausible cellular mechanism and its impact on the kidney.

Discussion

We provided one of the first cross-tissue alignments of genetic associations of kidney function with accessible chromatin. This supported the predominant role of kidney tissue and tubule cells, but also liver and pancreas in kidney-relevant gene regulation, particularly when genetic variants were prioritized by their statistical evidence to be causal for the association. Our analyses and experiments revealed HNF1-β as a key driver of regulatory effects among kidney function genetics. These also enabled the follow-up of genetic association to mechanism in human kidney cells for the NDRG1 gene.

Our cross-tissue analyses revealed that, within a great number of genetic signals defined by variants credibly associated with eGFR, at least one variant overlapped with accessible chromatin in at least one tissue. Thus, our analyses may help to prioritize the relevant tissue or cells for detailed functional analyses of associated variants. This also opens the possibility that some variants act on kidney function by regulating the gene through interference with REs in one or even multiple tissues outside the kidney. Given that some of the genetic effects on eGFR cannot be explained by interference with kidney REs, our results present a route into functional analysis of the cross-talk between tissues and kidney function[6]. We detected highest enrichment of variants and signals in kidney and liver when focusing on eGFRcrea-associated genetic variants that were validated by eGFRcys or BUN. This is in line with the kidney's function for waste excretion.

Apart from the detection of candidate REs related to kidney function in and outside the kidney, we demonstrated that statistical prioritization of genetic variants derived by a Bayesian fine-mapping can help narrow down to regulatory regions with accessible chromatin and transcription factor binding sites in kidney tissue[1]: variants with a posterior probability of association of > 2.5% were more likely to reside in accessible chromatin and more likely to co-incide with HNF-1β motifs and HNF-1β DNA-binding. For kidney REs, this effect of variant enrichment was solely observed for eGFRcys and BUN-validated variants, which rather reflect true kidney function than creatinine metabolism, further increasing the confidence in our approach.

Regulatory mechanisms and corresponding genes are difficult to be mapped for GWAS-identified variants. Variant-to-gene mapping using eQTL data often leads to multiple genes mapped by one variant. Our ABC experiment on accessible chromatin in PTC together with gene expression information allowed us to reveal potential regulatory mechanisms of nine candidate variants credibly associated with

eGFRcrea. For example, we linked rs72629024 as a potential regulatory variant to expression of pancreatic progenitor cell differentiation and proliferation factor (*PPDPF*). This gene has recently been shown in experimental models to protect proximal tubules in chronic kidney disease by preserving NAD^+ and mitochondrial homeostasis[36].

The success of our approach and the significance of our findings was exemplified at the signal around the *NDRG1* gene. From the nine variants within the GWAS signal near this gene, only rs10283362, which has the highest PPA value of the nine variants, resided in open chromatin in our data. The variant-associated RE was open preferentially in the kidney and it is an eQTL for the *NDRG1* gene in the tubulointerstitial compartment of the kidney[25]. During the course of this work, an allele-specific analysis of accessible chromatin has linked this variant to regulation of *NDRG1* expression in the kidney[6,24]. Our genotype-expression correlation in PTC from 141 individuals confirmed the observed eQTL and revealed the role of tubule cells for this signal[25]. This was further supported by the knock-out of the element that led to a marked reduction of *NDRG1* expression to levels comparable to those observed using a CRISPRi approach in the previous study [6].

Through alignment of transcription factor DNA-binding, we detected a direct role of the variant rs10283362 in modulating HNF-1β mediated transactivation of *NDRG1* expression. The T-allele, which is associated with higher eGFR values and higher NDRG1 expression, creates an HNF-1β motif. From our experimental data on differential cell growth and migration in enhancer knock-out cells, we speculate that this regulation may be relevant for kidney intrinsic functions such as nephrogenesis and tubule repair - all of which could impact kidney function during development and lifetime[11,32]. Imbalanced expression of *NDRG1* through the polymorphic enhancer might compromise the potential of cells to regenerate after acute injury. *NDRG1* expressing proximal tubule cells proliferate massively after ischemia/reperfusion injury in the mouse model and ndrg1a allows for cell survival and preserved kidney function under prolonged hypoxia in zebrafish[31,32]. This is in line with a role of *NDRG1* in tubule repair.

In addition to the *NDRG1* locus, our analyses revealed that several eGFR variants are located at HNF-1β binding sites in tubule cells. Given the permanent expression and function of HNF-1β in kidney tubules throughout lifetime, variants within the binding site may create a “genetic pressure” on this pathway and thus modulate kidney function in the long term. In this regard, rs77924615 (PPA 100%) at the *PDILT/UMOD* locus is of special interest, since this variant has been identified in multiple GWAS for kidney function and chronic kidney disease progression[1,3,37]. The variants in this region harbor the largest effect size on eGFR and chronic kidney disease[2,3,38,39]. Recent work has linked this variant to circulating levels of uromodulin and suggested the regulation of *UMOD* expression, which is mainly restricted to the cells of the thick ascending limb of the loop of Henle, in which the regulatory element is also preferentially accessible[40]. *UMOD* expression has also been reported to depend on the presence of HNF-1β[30]. Since we did not detect substantial other HNF-1β signals within this region, the variant-carrying enhancer reported in this study would give a plausible explanation for the reported transactivation of *UMOD* by HNF-1β. Whether rs77924615 directly interferes with this regulation remains

to be determined. Interestingly, mutations in both genes, *UMOD* and *HNF1B*, provoke autosomal-dominant tubulointerstitial kidney disease further highlighting the tight connection between these two factors in determining kidney function.

Our study has several strengths and some limitations. While we demonstrate liver and pancreas as target of relevant regulatory mechanisms related to kidney function genetics, we focused on tubule-specific resolution of variant enrichment within accessible chromatin in the kidney. To yield robust results, we applied rather strict filtering criteria to the variants and ATAC-seq data sets and strict criteria for calling HNF-1β peaks. This opens the possibility that variants in accessible chromatin or interactions of the transcription factor with eGFRcrea-associated variants might have escaped detection in our analyses. An important strength is that we followed *NDRG1* with numerous experiments from the genetic association to the cellular mechanism in human kidney cells.

Our work highlights the diversity of putative eGFRcrea-associated variant-chromatin interactions across human tissues. This cross-organ view reveals liver and pancreas as relevant for kidney function genetics but also underscores the predominance of the kidney, especially the tubules, as target organ of kidney function genetics. It provides first evidence that alignment of renal-lineage factors with genetic variability of kidney function results in the discovery of regulatory and relevant transcriptional relationships. Given the substantial number of lineage factors relevant for the kidney, we anticipate this as a promising route to unveil a great part of the direct connections between eGFRcrea-associated genetic variation and kidney gene transcription. Further unraveling regulatory mechanisms of GWAS-associated signals bares ample opportunities to identify actionable genes to preserve or restore kidney function.

Methods

Definition of accessible chromatin in 29 tissues and PTC

Published ATAC-seq data (BAM files) from 28 tissues were obtained from the ENCODE database (https://www.encodeproject.org/, accessed 30/04/2024)[18]. Quality control included determination of the total number of mapped reads to the genome after the filtering processes, the distribution of fragment sizes using the CollectInsertSizeMetrics function from Picard (v2.25.2). The Fraction of Reads in Peak (FRiP) score and the Transcription Start Site (TSS) enrichment score were evaluated via the $TSSEsc$ or e function from the ATACseqQC (v1.18.1) package[41]. Only high-quality samples according to current ENCODE criteria were selected from ENCODE data set to define the open chromatin landscapes. Peaks were identified for each sample and tissue using MACS2 (v2.2.7.1) by applying the MACS2 callpeak command, with parameters set to "--shift − 75 --extsize 150 --nomodel --call-summits --nolambda --keep-dup all -p 0.1."[42]. Additionally, peak files from ATAC-seq data related to kidneys were sourced from the Susztak laboratory's website (accessed April 10, 2022)[5]. Subsequently, the summit method, as detailed by Corces et al.[17], was applied to merge multiple peak files into a single peak file

across 29 tissues. We used a similar approach to define accessible chromatin in 7 PTC ATAC-seq experiments (sequenced in duplicates) and to create a landscape of open chromatin in these cells.

Peaks located within ENCODE blacklisted regions were excluded from further analysis. To normalize the peak scores, we calculated a “score per million” (SPM) value for each peak, as outlined by Corces et al. [17]. We used an SPM value of 3.5 or higher to define accessible chromatin for each tissue or the PTC. Alternatively, we used the top 50,000 peaks based on their calculated SPM values for further analysis as indicated.

Open chromatin landscapes derived from the tissues (≥ 3.5 SPM or top 50,000 peaks) were compared with well-characterized cis-regulatory elements (cCREs) obtained from https://downloads.wenglab.org/V3/GRCh38-cCREs.bed (accessed on 12 February 2021)[15]. The open chromatin landscapes from PTC were integrated with the 29 tissue landscapes to create a comprehensive matrix, utilizing the "reduce" function from the GenomicRanges package[43]. The resulting matrix was analyzed using the prcomp function from the stats package (v 3.6.2) to perform PCA [44]. PCA results were visualized using ggplot2 and ggrepel (v0.9.6)[45,46].

eGFR Meta-GWAS Data

We used data from a GWAS that focused on kidney function (eGFR) and included 1.2 million participants[1]. This GWAS identified 35,881 eGFRcrea-associated credible variants by fine-mapping. We used a refined set of variants compared to the original study[19]. GWAS data in hg19 was converted to the hg38 coordinates using CrossMap.py (v0.6.1)[47].

GWAS variants overlapping open chromatin in the 29 different tissues and PTC were identified using the $\subset ByOverlaps$ function from the GenomicRanges package[48]. To analyze the tissue-specificity of REs with eGFR variants, we calculated the fraction of common REs between each tissue. Heatmap was generated using the pheatmap package (v1.0.13)[49].

Overlap of REs with the 16 compartments of accessible chromatin regions

The 16 compartments of accessible chromatin regions, as defined by Meuleman et al.[20], were utilized to characterize regulatory elements (REs). We employed the enrichPeakOverlap function from the ChIPseeker R package to assess the statistical significance of the enrichment[50]. P-values were calculated through a permutation test (N = 1,000 random permutations of genomic regions).

Overrepresentation analysis

Regulatory DNA elements were annotated using the “annotatePeak” function with TxDb.Hg38.KnownGene (v3.14.0) to the nearest TSS window[51]. The TSS window was defined as a -1000 to + 1000 bp range from the TSS. Overrepresentation analysis (ORA) was performed on the genes using

the “enricher” function from the clusterProfiler package, utilizing cell type signature gene sets (C8) from MsigDB[52,53]. The results of the ORA were visualized using the dotplot function from the clusterProfiler package[52].

Healthy kidney samples

Healthy human kidney cortical tissue from patients undergoing tumor nephrectomy was provided by the Comprehensive Cancer Center Erlangen-EMN (CCC) at the Uniklinikum Erlangen. The local ethics committee at the University of Erlangen-Nürnberg approved use of the tissue (3755, 271_Bc) and each patient gave informed consent. Specimens were collected in accordance with the World Medical Association Declaration of Helsinki.

Cell Culture

Healthy human kidney cortical tissue from patients undergoing tumor nephrectomy was used for primary cell isolation as described previously[54]. Samples were minced, digested with collagenase II (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), and sieved through 100 µm and 70 µm filters. Primary human tubular cell cultures were maintained in Dulbecco’s modified Eagle’s medium (DMEM)/ F-12 (1:1) (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 2.5% (day 1) or 0% (day 2 onwards) fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin, 5 µg/ml insulin, 5 µg/ml transferrin, and 5 ng/ml selenium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), triiodothyronine (T3, Sigma Aldrich, St. Louis, MO, USA) 10 ng/ml, hydrocortisone 1 mg/ml, epidermal growth factor 100 µg/ml (Peprotech, Hamburg, Germany). Epithelial origin was confirmed by immunocytochemistry for N- and E-Cadherin. Cell cultures of primary cells were used up to passage 4. The renal proximal tubule cell lines HK-2 (ATCC CRL-2190) and HKC-8 (kindly provided by L. Racusen, Baltimore, MD, USA) were cultured in DMEM/F-12 (1:1) (Gibco) supplemented with 5% FCS, 100 U/ml penicillin and 100 µg/ml streptomycin.

Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq)

ATAC experiments were performed as previously described[55]. Primary renal tubule cells were grown under standard culture conditions and harvested at 80% confluency. Cells were trypsinized, manually counted and 60,000 cells were directly subjected to the Omni-ATAC protocol as described by Corces et al.[56] Fragments of accessible DNA were generated using the Illumina Tagment DNA TDE1 Enzyme and Buffer Kit (Illumina, San Diego, CA, USA) and purified using the DNA Clean and Concentrator-5 Kit (Zymo Research, Freiburg, Germany) according to the manufacturer’s recommendations. Following purification, library fragments were amplified by using the NEB Next® High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, UK) and custom made barcoded Nextera PCR primers (Sigma Aldrich, St. Louis, MO, USA). Libraries were amplified for a total of 8–12 cycles. Library clean-up and double-sided size selection was performed by using AMPure XP beads (Beckman Coulter GmbH, Krefeld, Germany) according to the manufacturer’s protocol. Fragment size and DNA concentration of the libraries were

determined by electrophoresis using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Equimolar concentrations of barcoded libraries were pooled prior to sequencing.

ATAC-seq analysis

Barcoded amplicons of the ATAC-libraries were sequenced on an Illumina Novaseq 6000 platform to a 2 × 150 pair-ended format by Novogene (Cambridge, UK). Reads were quality filtered according to the standard Illumina pipeline and de-multiplexed. Fastq files were generated. Illumina adapter sequences were trimmed using Trim Galore (0.6.6)[57]. Reads were aligned against the human reference genome (hg38) using Bowtie2 (2.3.4.1)[58]. SAMtools (1.8) removed aligned reads with a mapping quality (MAPQ) below 30 and those mapped to ChrM[59]. Duplicated reads were excluded via Picard (http://broadinstitute.github.io/picard/) (2.25.2). ENCODE blacklisted regions (access 20/10/2020) and chromosome Y were removed using GenomicRanges (v1.46.1)[48]. To check the sequencing quality of the samples, we used the total number of mapped reads to the genome after filtering processes, the fraction of reads in peaks (FRiP) score, and the transcription start site (TSS) score according to ENCODE criteria (https://www.encodeproject.org/atac-seq/).

Chromatin immunoprecipitation (ChIP)

Primary tubule cells were cultured in standard cell culture conditions in 15 cm cell culture dishes until sub-confluency. Chromatin immunoprecipitation experiments were performed using the ChIP-IT high sensitivity kit (Active Motif, Carlsbad, CA, USA) according the manufacturer's instructions. In brief, for immunoprecipitations, 70 µg of isolated chromatin and 3 µl of rabbit antibodies directed against H3K27ac (ab4729, Abcam, Cambridge, UK) or HNF-1β (HPA002083, Atlas Antibodies, Sigma Aldrich, St. Louis, MO, USA) were used. Purified normal rabbit IgG (12–370, Merck Millipore, Burlington, MA, USA) served as a negative control. Antibody-chromatin complexes were pulled down by proteinase A agarose beads. After reversal of the crosslinking by heating, DNA was isolated via DNA purification columns.

ChIP-seq analysis

Library preparation of immuno-precipitated DNA and ChIP-sequencing were performed by Novogene (Cambridge, UK). Amplicons from the ChIP library were sequenced on an Illumina Novaseq 6000 platform to a 2 × 150 pair-ended format. Reads were quality filtered according to the standard Illumina pipeline and de-multiplexed. Fastq files were generated. Illumina adapter sequences were trimmed using Trim Galore (v0.6.6)[57]. Quality control was conducted with FastQC (v0.11.8)[60]. Subsequently, reads were mapped to the hg38 version of the human genome using Burrows-Wheeler Aligner (BWA-mem; v0.7.17-r1188). SAMtools (v1.8) removed aligned reads with a Mapping Quality (MAPQ) below 30 and those mapped to ChrM[61]. Duplicated reads were excluded via Picard (http://broadinstitute.github.io/picard/) (v2.25.2). ENCODE blacklisted regions from merged files were excluded. We generated normalized BigWigs for visualization using BamCoverage (v3.3.2)[62]. Files were visualized using the Integrative Genomics Viewer (IGV,_v2.8.4)[63].

Cleavage Under Targets and Tagmentation (CUT&Tag)

CUT&Tag was performed as described in Kaya-Okur et al. with modifications based on the protocol published by Bartosovic et al.[27,64] We used 250,000 to 350,000 cells per condition and antibody. Cells were washed three times in 200 µl of antibody buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM EDTA, 0.5 mM spermidine, 0.05% digitonin, 0.01% NP-40, 1× protease inhibitors and 1% BSA). Cells were centrifuged for 5 min at 600*g*, resuspended in 100 µl antibody buffer with 1:50 diluted primary antibody (monoclonal rabbit anti-H3K27ac antibody [ab177178, Abcam, Cambridge, UK] or polyclonal rabbit anti-HNF-1β antibody [HPA002083, Atlas Antibodies, Sigma Aldrich, St. Louis, MO, USA]) and incubated overnight. The next day cells were centrifuged for 3 min at 600*g*, washed once with 200 µl of Dig-wash buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 0.05% digitonin, 0.01% NP-40, 1× protease inhibitors, 1% BSA), resuspended in 200 µl of Dig-wash buffer with 1:50 diluted secondary antibody (ABIN101961, Guinea Pig anti-Rabbit IgG antibody, Antibodies-Online, Aachen, Germany) and incubated for 1h at room temperature with slow rotation. Afterwards cells were centrifuged for 3 min at 600*g*, washed three times with 200 µl of Dig wash buffer, resuspended in 200 µl of Dig-300 buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.05% digitonin, 0.01% NP-40, 1× protease inhibitors, 1% BSA) with 1:100 diluted protein A–Tn5 transposase (C01070002, Diagenode, Seraing, Belgium, loaded with mosaic adapters according to manufacturer's instructions) and incubated for 1h rotating at room temperature. Following pA-Tn5 loading, cells were washed three times with Dig-300 buffer, resuspended in 200 µl of tagmentation buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM spermidine, 0.05% digitonin, 0.01% NP-40, 1× protease inhibitors, 10 mM MgCl2) without BSA and incubated for 1h at 37°C. Tagmentation was stopped by addition of 7 µl of 500 mM EDTA. Subsequently 4 µl SDS 10% and 2 µl Proteinase K were added and the sample was incubated for 1h at 55°C. DNA isolation and clean-up were performed using the ZYMO DNA Clean & Concentrator-5 kit with a 1:5 ratio of binding buffer and 25 µl of elution buffer (Zymo Research, Freiburg, Germany). Following purification, DNA fragments were amplified by using the NEB Next High-Fidelity 2X PCR Master Mix (632 New England Biolabs, Ipswich, UK) and custom made barcoded Nextera PCR primers (Sigma Aldrich, St. Louis, MO, USA) as published by Kaya-Okur et al.[27] Libraries were amplified for a total of 15 cycles. SPRIselect bead-based reagent (B23317, Beckman Coulter GmbH, Krefeld, Germany) was used for library clean-up according to the manufacturer's protocol.

CUT&Tag-seq analysis

CUT&Tag libraries were sequenced by Novogene (Cambridge, UK) on an Illumina Novaseq XPlus platform to a 2 x 150 paired-end format (PE). For H3K27ac CUT&Tag samples 5 to 40 Mio mapped reads were generated. For CUT&Tag with antibodies directed against HNF-1β 1.3 to 21.4 Mio mapped reads were acquired. FastQC (v0.11.8) was applied for quality control and sequencing adapters were removed using Trim Galore (v0.3.3)[57,60]. Bowtie2 (v2.3.4.1) was used for read alignment against the human reference genome (hg38)[58]. Low quality reads (MAPQ < 30) and reads mapping to mitochondrial DNA or ENCODE blacklisted regions were eliminated using SAMtools (v.1.8)[59]. Picard (http://broadinstitute.github.io/picard) (v.2.25.2) removed duplicated reads. BamCoverage (3.3.2) was

used to create normalized bigwig tracks for visualization[62]. Peaks were identified for each sample using MACS2 (v2.2.7.1) through the MACS2 callpeak command with parameters set to "-f BAM -q 0.05." [65].

Activity-By-Contact (ABC) model analysis

The Activity-By-Contact (ABC) model of enhancer-gene specificity (v1.0) was applied to predict enhancer-promoter interactions regulating NDRG1 gene expression in PTC[23]. Inputs for the individual-specific analysis were BAM files for ATAC-seq and H3K27ac ChIP-seq as well as TPM-normalized RNA-seq data from the corresponding PTC of three different individuals. For quantification of the contact frequency publicly available average Hi-C profiles (https://github.com/charlesfulco/ABC-Enhancer-Gene-Prediction) were used. Only enhancer predictions with Powerlaw scores ≥ 0.015 were kept. ABC-data for PTC will be published within a different study (Naas et al. manuscript under revision).

Identifying motifs using HOMER

To identify enriched motifs within DNA elements, a motif discovery analysis was conducted using HOMER2 (Hypergeometric Optimization of Motif EnRichment, v5.1), as described by Heinz et al. (2010)[66].

Consensus HNF-1β-binding sites

We utilized data from HNF-1β CUT&Tag-seq experiments from five PTC to identify consensus HNF-1β binding sites. The peaks were consolidated using the "reduce" function from the GenomicRanges package, retaining only those peaks present in all five samples, resulting in a total of 3,691 peaks. The overlap of these peaks was visualized using a Venn diagram via eulerr (v7.0.2)[67]. Additionally, we explored the association between peaks and gene features. Peaks were annotated using the annotatePeak function with TxDb.Hg38.KnownGene (v3.14.0) with a TSS annotation in a -1000 to + 1000 bp window from the TSS[51]. Overlap of consensus HNF-1β binding sites with ATAC-sites was analyzed using the “subsetByOverlaps” function from the GenomicRanges package[48]. The percentage of HNF-1β binding sites overlapping ATAC-sites at at least one base pair was visualized using the pheatmap package (v1.0.13)[49].

RNA isolation and qPCR

RNA isolation, cDNA transcription and qPCRs were performed as described earlier[68]. Briefly, RNA was isolated using the peqGold total RNA kit (VWR Peqlab, Erlangen, Germany) and 100 ng of total RNA was transcribed into cDNA using the High capacity cDNA reverse transcription kit (Thermo Fisher Scientific, Waltham, MA, USA). qPCRs were performed on a StepOnePlus real-time PCR system (Thermo Fisher Scientific, Waltham, MA, USA) using the Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA). Expression values were normalized to values for the housekeeping gene HPRT (for expression primers please see **Supplementary Table 18**).

RNA-seq analysis

Adequate quality of isolated RNA was confirmed using the RNA Integrity Number (RIN) determined by a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). RNA was sequenced in technical duplicates by Novogene (Cambridge, UK) on an Illumina Novaseq 6000 or Illumina NovaSeq XPlus platform to a 2 x 150 paired-end format with an average number of 20–30 million reads per sample. After quality control using FastQC (v0.11.8) reads were aligned to the human reference genome (hg38) using the STAR alignment software (v2.7.10)[60,69]. A count table of a mapped reads was generated with the feature-counts software (v1.6.1)[70]. Raw reads were normalized in R (v4.1.1) using the parameter transcript per million mapped reads (TPM)[44]. Transcripts with fewer than 10 reads were discarded. The DESeq2 package version 1.32.0 was used for logarithmic transformation of the data and for data exploration[71]. Differential expression analysis was performed by using the DESeq2 lfc Shrink approach. Adjusted p-values were calculated using the Benjamini–Hochberg (BH) method within DESeq2. Gene annotations were added to the result files using biomaRT (v2.48.3)[72]. Differentially expressed genes (DEG) were visualized with a Volcano plot. We compared the two control clones of cells to the two *NDRG1* enhancer knock-out clones of cells or the exon knock-out clone of cells, respectively.

Functional annotation of differentially expressed genes via Metascape

To explore functional associations of differentially expressed genes (DEGs) derived from *NDRG1* enhancer or *NDRG1* exon knock-out clones of HK-2 cells compared to control HK-2 cells, we utilized the Metascape platform[35]. We used DEGs based on an adjusted p-value threshold of < 0.05. We considered genes to be upregulated if their $\log_2$(fold change) was ≥ $\log_2$(0.5) and downregulated if ≤ $\log_2$(-0.5). The Metascape Platform https://metascape.org/ (access 07/24/2025) was used for GO enrichment analysis. Results were visualized using bar plots.

siRNA transfection

siRNAs targeting HNF-1β and non-targeting control siRNA (nt) have been previously described[68] (**Supplementary Table 18**). siRNAs were transfected at a final concentration of 40nM using Lipofectamin3000® transfection reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol.

CRISPR/Cas9 mediated genome editing

For CRISPR/Cas9 experiments we used the TrueCut Cas9 protein (Themo Fisher Scientific, Waltham, MA, USA) and single guide RNA (sgRNA) directed against the variant rs10283362, *NDRG1* exon 2 (#A35534, Themo Fisher Scientific, Waltham, MA, USA) or a non-targeting sgRNA (#A35526, Thermo Scientifc, Waltham, MA, USA) (**Supplementary Table 18**). sgRNA and Cas9 protein were transfected in HK-2 cells or PTC using Lipofectamin CRISPRMAX® Cas9 transfection reagent (HK-2 cells, Thermo Fisher Scientific, Waltham, MA, USA) or electroporation (PTC) at a final concentration of 450 ng/ml for sgRNAs and 2.5 µg/ml for Cas9 protein. Single clones of cells were generated by dilution. For mutation screens,

genomic DNA of each clone was isolated by phenol-chloroform extraction. The locus targeted by CRISPR/Cas9 was amplified by PCR and analyzed by Sanger sequencing. To evaluate efficient knock-out the “EditCo’s ICE Analysis Tool” method (ICE CRISPR Analysis. 2025. v3.0. EditCo Bio; [accessed 05/07/2025]) was used.

Genotyping and allele-specific assays

Genomic DNA from individual PTC samples was isolated by phenol/chloroform extraction. A commercially available TaqMan® assay (Thermo Fisher Scientific, Waltham, MA, USA) was used to genotype for rs10283362 (C_9086181_10). Data were analyzed using the TaqMan® Genotyper Software V1.3 (Thermo Fisher Scientific, Waltham, MA, USA). For assessment of allele-specific interactions DNA from CUT&Tag experiments was analyzed. Allelic ratios of different alleles were calculated as described earlier[73].

Immunoblotting

Cells were lysed in UREA/SDS buffer and proteins were resolved by SDS-PAGE. After transferring the proteins onto PVDF membranes proteins were detected using a rabbit polyclonal anti-NDRG1 antibody (1:1000; #5196, Cell Signaling, Danvers, MA, USA) and a rabbit polyclonal anti-HNF-1β antibody (1:2000; HPA002083, Atlas Antibodies, Sigma Aldrich, St. Louis, MO, USA). We further used a mouse monoclonal anti-β-actin antibody (1:60.000; A3854, Sigma Aldrich, St. Louis, MO, USA) or a mouse monoclonal anti-GAPDH antibody (1:5000; sc-32233; Santa Cruz, Dallas, TX, USA). Horseradish peroxidase-conjugated secondary antibodies were used as applicable (Dako, Agilent Technologies, Santa Clara, CA, USA). Signal detection and densitometry of bands was performed on an Amersham Imager 600 (GE Healthcare, Amersham, UK).

Reporter assays

A 438 bp sequence of the cis-regulatory element including the variant rs10283362 (hg38; chr8:133320472–133320909) was PCR amplified from genomic DNA and cloned into the pGL3 promoter vector (Promega, Madison, WI, USA) using KPNI and NHEI restriction sites (**Supplementary Table 18**). Sequences were verified using Sanger sequencing. Transfections of plasmids and a control β-galactosidase reporter plasmid were performed in HK-2 and HKC-8 cells using XtremeGENE HP DNA transfection reagent (Roche Diagnostics, Mannheim, Germany). Luciferase activity in extracts was measured by adding the Steady-Glo luciferase reporter assay system (Promega, Madison, WI, USA) and was analyzed via the GloMAX Detection System (Promega, Madison, WI, USA). Values were normalized to β-galactosidase activity in the same cell lysates.

Proliferation assay

For cell proliferation analysis, the CellTiter 96 AQueous One Solution Cell Proliferation Assay (MTS) (Promega, Madison, WI, USA) was used according to the manufacturer’s protocol. Briefly, 500 HK-2 cells were seeded per well in 100 µl medium on a 96-well plate (Greiner Bio-One, Frickenhausen, Germany).

After 4 hours CellTiter 96 AQueous One Solution Reagent was added to the adhered cells. Absorbance at 490 nm was measured at the indicated time points using the Infinite F50 Plus detection system (Tecan Trading AG, Switzerland). Values were normalized to background and the respective control sample.

Wound healing and cell migration assays

For cell migration studies of HK-2 *NDRG1* enhancer KO clones, 2-well culture inserts (ibidi GmbH, Gräfelfing, Germany) were set at the bottom of a 24-well plate. 20,000 cells per well were seeded in 100 µl cell culture medium. After 24h culture inserts were removed and 250 µl cell culture medium per well was added. Cell migration was monitored over a duration of 24h with pictures every 30 minutes using a TCS-SPE microscope (Leica Microsystems GmbH, Wetzlar, Germany). Migration analyses were performed in ImageJ Software with following plugins: wound area calculations (MRI Wound Healing Tool, Montpellier Ressoures Imagerie, France), cell tracking (Manual Tracking; Fabrice Cordeliéres, Institut Curie, Orsay, France) and migration calculations (Chemotaxis and Migration Tool for ImageJ, ibidi, Gräfelfing, Germany).

Statistical analysis

All statistical analyses excluding sequencing data were performed using GraphPad Prism v9.0.2 (GraphPad Software Inc., San Diego, CA, USA). Results are presented as mean +/- SD if not indicated otherwise. Statistical tests applied to determine significance are provided in the figure legends. The one-sample t-test comparing the mean with the hypothetical value of 1 was used if the values of the control conditions were set to 1. X^2-test was used to determine significance of enrichment of variant carrying REs in REs with an HNF-1β motif or with HNF-1β binding. Statistical testing of NGS data sets is described in the respective methods section.

Visualization

BigWig files were visualized using the Integrative Genomics Viewer (v2.16.2)[63]. GraphPad Prism v9.0.2 (GraphPad Software Inc., San Diego, CA, USA) and Adobe Photoshop Elements (Adobe Inc., San Jose, CA USA) were used for creating figures. Schematics were designed with BioRender.com

Declarations

Author contributions

René Krüger: Investigation, Software, Data Curation, Writing - Original Draft; Victoria Lauer: Investigation, Data Curation, Writing - Original Draft; Nora Feldker: Data Curation; Klaus Stark: Software, Data Curation, Writing - Original Draft; Mathias Gorski: Software, Data Curation; Thomas Winkler: Software, Data Curation; Fulvia Ferrazzi: Resources; Mario Schiffer: Resources; Stephanie Naas: Investigation, Software, Data Curation; Iris Heid: Conceptualization, Supervision, Funding acquisition, Resources, Writing -

Original Draft; Johannes Schödel: Conceptualization, Writing - Original Draft, Supervision, Funding acquisition, Data Curation.

Acknowledgment

We thank Stephanie Palffy, Pia Klanke, Astrid Ebenau-Eggers, Johanna Stöckert, Laura Bernad, and Margot Rehm for excellent technical assistance. We thank the Department of Urology and Pediatric Urology and the Comprehensive Cancer Center Erlangen-EMN at the Uniklinikum Erlangen for their support. Parts of this work were performed by RK in fulfillment of the requirements for obtaining the degree Dr. rer. biol. hum. This study was funded by the German Research Foundation (Project-ID 509149993; TRR 374, TP C5, C6 and SCHO 1598/2 – 1). SN was supported by the Interdisciplinary Center for Clinical Research (IZKF) at the University Hospital of the University of Erlangen-Nuremberg (Junior Project J114). We acknowledge financial support by the Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg within the funding programme “Open Access Publication Funding”. All authors approved the final version of this manuscript.

Data availability

All DNA-sequencing data (ATAC-seq, ChIP-seq, CUT&Tag-seq, RNA-seq) generated for this study is available at GEO (ATAC-seq: GSE256001) and zenodo (10.5281/zenodo.17454418). Raw ATAC-seq data was downloaded from ENCODE or https://cmdga.org/ (access 10/04/2021)[5]. We used eQTL data from the GTEx project and NEPTUNE v2.0[24,25]. We utilized snATAC-seq data obtained from the adult human kidney cortex (GSE151302)[26,33].

Code availability

No customized code was used for data analyses in this study.

References

1. Stanzick KJ et al (2021) Discovery and prioritization of variants and genes for kidney function in > 1.2 million individuals. Nat Commun 12:4350
2. Kottgen A et al (2009) Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet 41:712–717
3. Wuttke M et al (2019) A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat Genet 51:957–972
4. Hellwege JN et al (2019) Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. Nat Commun 10:3842
5. Liu H et al (2025) Kidney multiome-based genetic scorecard reveals convergent coding and regulatory variants. Science 387:eadp4753

6. Loeb GB et al (2024) Variants in tubule epithelial regulatory elements mediate most heritable differences in human kidney function. Nat Genet 56:2078–2092
7. Dunham I et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
8. Li Y et al (2020) Integration of GWAS Summary Statistics and Gene Expression Reveals Target Cell Types Underlying Kidney Function Traits. J Am Soc Nephrol 31:2326–2340
9. Clissold RL, Hamilton AJ, Hattersley AT, Ellard S, Bingham C (2015) HNF1B-associated renal and extra-renal disease-an expanding clinical spectrum. Nat Rev Nephrol 11:102–112
10. Tholen LE, Hoenderop JGJ, de Baaij JHF (2022) Mechanisms of ion transport regulation by HNF1beta in the kidney: beyond transcriptional regulation of channels and transporters. Pflugers Arch 474:901–916
11. Ferre S, Igarashi P (2019) New insights into the role of HNF-1beta in kidney (patho)physiology. Pediatr Nephrol 34:1325–1335
12. Buffin-Meyer B et al (2024) Renal and Extrarenal Phenotypes in Patients With HNF1B Variants and Chromosome 17q12 Microdeletions. Kidney Int Rep 9:2514–2526
13. Devuyst O et al (2019) Autosomal dominant tubulointerstitial kidney disease. Nat Rev Dis Primers 5:60
14. Econimo L et al (2022) Autosomal Dominant Tubulointerstitial Kidney Disease: An Emerging Cause of Genetic CKD. Kidney Int Rep 7:2332–2344
15. Encode Project Consortium (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710
16. Donohue LKH et al (2022) A cis-regulatory lexicon of DNA motif combinations mediating cell-type-specific gene regulation. Cell Genom 2
17. Corces MR et al (2018) The chromatin accessibility landscape of primary human cancers. Science 362
18. Luo Y et al (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res 48:D882–D889
19. Stanzick KJ et al (2023) KidneyGPS: a user-friendly web application to help prioritize kidney function genes and variants based on evidence from genome-wide association studies. BMC Bioinformatics 24:355
20. Meuleman W et al (2020) Index and biological spectrum of human DNase I hypersensitive sites. Nature 584:244–251
21. Clark JZ et al (2019) Representation and relative abundance of cell-type selective markers in whole-kidney RNA-Seq data. Kidney Int 95:787–796
22. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2019) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47:D886–D894

23. Fulco CP et al (2019) Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat Genet 51:1664–1669
24. Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348:648–660
25. Han SK et al (2023) Mapping genomic regulation of kidney disease and traits through high-resolution and interpretable eQTLs. Nat Commun 14:2229
26. Eun M et al (2024) Chromatin accessibility analysis and architectural profiling of human kidneys reveal key cell types and a regulator of diabetic kidney disease. Kidney Int 105:150–164
27. Kaya-Okur HS et al (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 10:1930
28. Ferre S, Veenstra GJ, Bouwmeester R, Hoenderop JG, Bindels RJ (2011) HNF-1B specifically regulates the transcription of the gammaa-subunit of the Na+/K+-ATPase. Biochem Biophys Res Commun 404:284–290
29. Adalat S et al (2009) HNF1B mutations associate with hypomagnesemia and renal magnesium wasting. J Am Soc Nephrol 20:1123–1131
30. Gresh L et al (2004) A transcriptional network in polycystic kidney disease. EMBO J 23:1657–1668
31. Endo T et al (2015) Exploring the origin and limitations of kidney regeneration. J Pathol 236:251–263
32. Park JS et al (2022) N-myc downstream regulated gene 1 (ndrg1) functions as a molecular switch for cellular adaptation to hypoxia. Elife 11
33. Muto Y et al (2021) Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat Commun 12:2190
34. Genomes Project C et al (2015) A global reference for human genetic variation. Nature 526:68–74
35. Zhou Y et al (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10:1523
36. Fang X et al (2025) PPDPF preserves integrity of proximal tubule by modulating NMNAT activity in chronic kidney diseases. Sci Adv 11:eadr8648
37. Morris AP et al (2019) Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. Nat Commun 10:29
38. Devuyst O, Bochud M, Olinger E (2022) UMOD and the architecture of kidney disease. Pflugers Arch 474:771–781
39. Devuyst O, Pattaro C (2018) The UMOD Locus: Insights into the Pathogenesis and Prognosis of Kidney Disease. J Am Soc Nephrol 29:713–726
40. Li Y et al (2022) Genome-wide studies reveal factors associated with circulating uromodulin and its relationships to complex diseases. JCI Insight 7
41. Ou J et al (2018) ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. BMC Genomics 19:169

42. Zhang Y et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137
43. Lawrence M et al (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118
44. R Development Core Team (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing
45. Slowikowski K, Schep A, Hughes S, Lukauskas S, Irisson J-O, Kamvar ZN, Ryan T, Christophe D, Hiroaki Y, Gramme P (2018) Package ggrepel. in *Automatically Position Non-Overlapping Text Labels with 'Ggplot2*.
46. Wickham H (2011) ggplot2. *Wiley interdisciplinary reviews: computational statistics* 3, 180–185
47. Zhao H et al (2014) CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 30:1006–1007
48. Lawrence M et al (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118
49. Kolde R, Kolde MR (2015) Package 'pheatmap'. R package 1:790
50. Yu G, Wang LG, He QY (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31:2382–2383
51. Bioconductor Core Team and Bioconductor Package Maintainer. _TxDb.Hsapiens.UCSC.hg38.knownGene: Annotation package for TxDb object(s)_. in R package version (2024)
52. Wu T et al (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (Camb) 2:100141
53. I D (2025) msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. in *R package version 25.1.1*
54. Kroening S, Neubauer E, Wullich B, Aten J, Goppelt-Struebe M (2010) Characterization of connective tissue growth factor expression in primary cultures of human tubular epithelial cells: modulation by hypoxia. Am J Physiology-Renal Physiol 298:F796–F806
55. Protze J et al (2022) The renal cancer risk allele at 14q24.2 activates a novel hypoxia-inducible transcription factor-binding enhancer of DPF3 expression. J Biol Chem, 101699
56. Corces MR et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959–962
57. Krueger F (2015) Trim galore. A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files
58. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
59. Li H et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079
60. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data
61. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM

62. Ramirez F et al (2016) deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:W160–W165
63. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP (2023) igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinformatics 39
64. Bartosovic M, Kabbe M, Castelo-Branco G (2021) Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat Biotechnol 39:825–835
65. Carroll TS, Liang Z, Salama R, Stark R, de Santiago I (2014) Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. Front Genet 5:75
66. Heinz S et al (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576–589
67. Larsson Jeulerr (2024) Area-Proportional Euler and Venn Diagrams with Ellipses. in *R package version 7.0.2*
68. Lauer V et al (2020) Hypoxia drives glucose transporter 3 expression through hypoxia-inducible transcription factor (HIF)-mediated induction of the long noncoding RNA NICI. J Biol Chem 295:4065–4078
69. Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
70. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923 – 30
71. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550
72. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4:1184–1191
73. Grampp S et al (2017) Multiple renal cancer susceptibility polymorphisms modulate the HIF pathway. PLoS Genet 13:e1006872

Figures

Figure 1

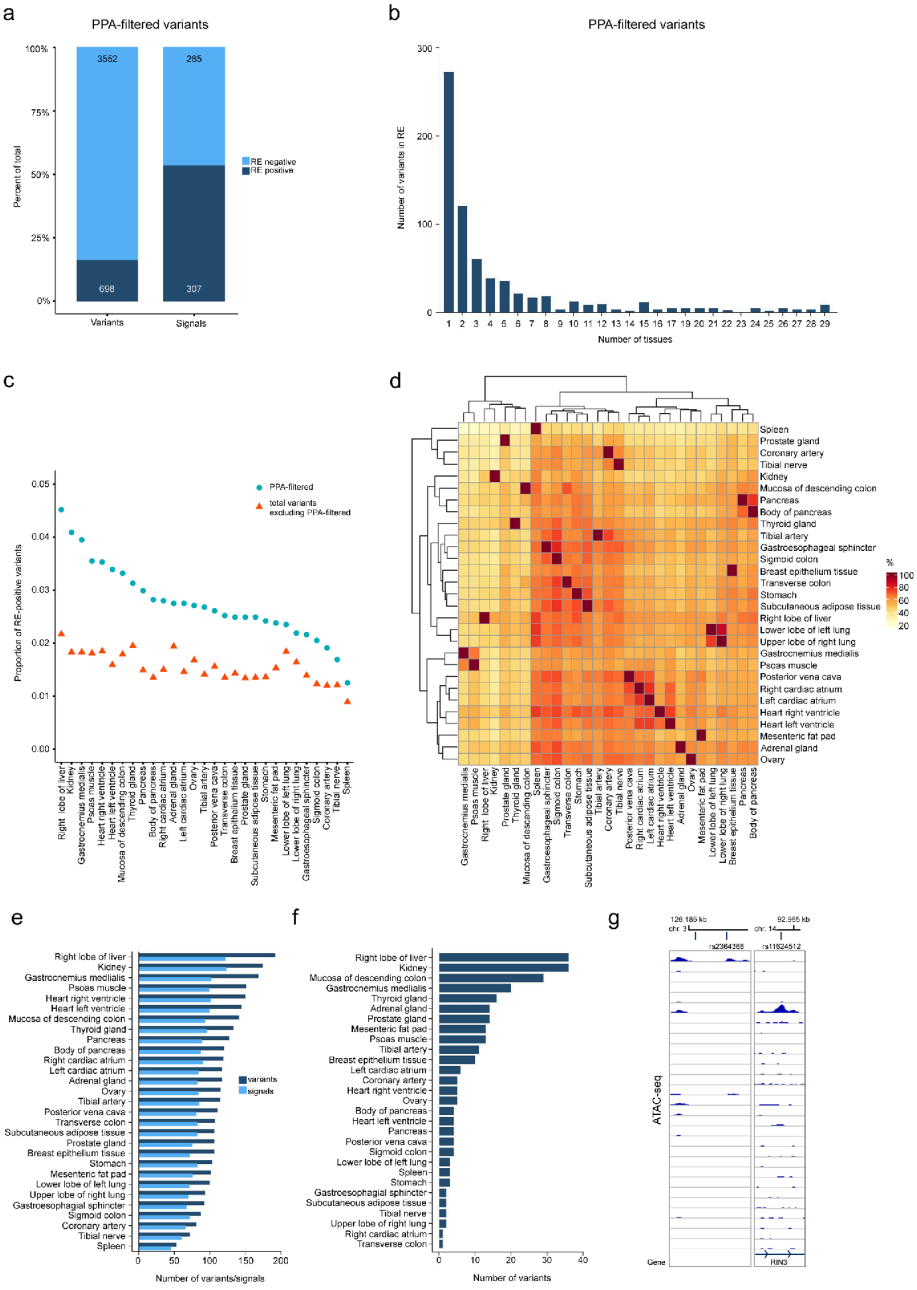

Figure 1

Cross-tissue analysis of accessible chromatin and eGFRcrea-associated signals and variants

a) Fraction of overlap of 4,250 eGFRcrea-associated variants (posterior probability of association (PPA)>2.5%, no locus with more than 1,000 variants) with accessible chromatin across 29 tissues. **b**) Number of PPA-filtered eGFRcrea-associated variants in open chromatin according to the number of

tissues with open chromatin at the respective position. **c**) Proportion of eGFRcrea-associated variants within open chromatin stratified according to tissue type and PPA. Blue indicates eGFRcrea variants with a PPA>2.5% at loci with <1,000 variants (PPA-filtered). Red indicates eGFRcrea variants with a PPA of less than 2.5% and includes variants from loci with more than 1,000 variants. **d**) Cluster analysis of REs with PPA-filtered eGFRcrea-associated variants overlapping across the 29 tissues. The porportion of overlapping REs with variants is colour-coded with dark red indicating highest overlap. **e**) Cumulative number of PPA-filtered eGFR-associated variants and signals overlapping accessible chromatin in the respective tissue. **f**) Number of PPA-filtered eGFRcrea-associated variants per tissue which co-incide with chromatin uniquely accessible in the respective tissue. **g**) Tissue ATAC-seq tracks for eGFRcrea-associated variants with tissue-specific chromatin accessibility. rs2364638, located close to the *ALDH1L1* gene, revealed tissue-specific chromatin accessibility in the right lobe of the liver; rs11624512 located in an intronic region of *RIN3* overlaps open chromatin specifically in the adrenal gland. Tracks of the tissues are in the same order as in f).

Figure 2

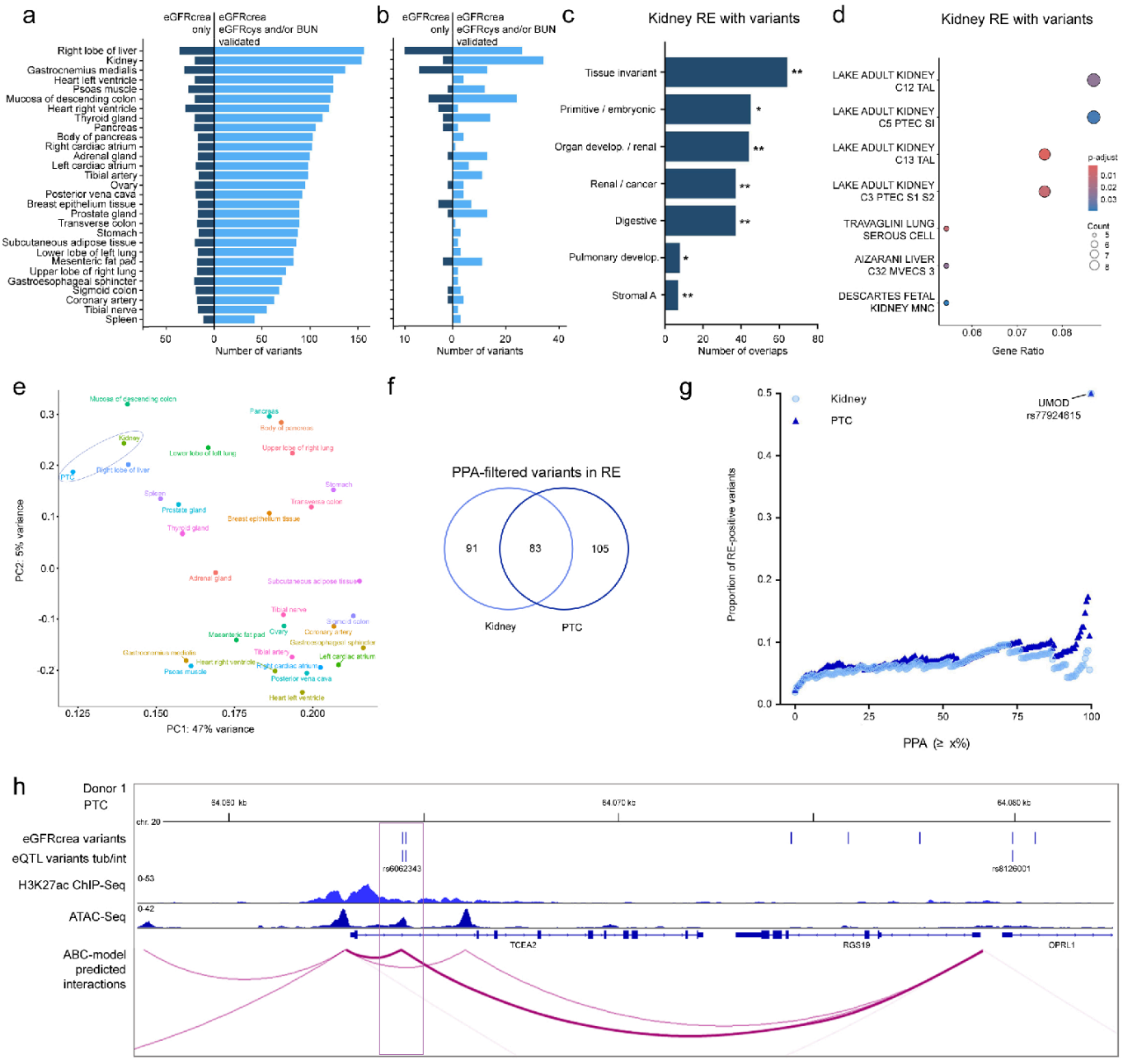

Figure 2

Refinement of genetic eGFRcrea associations at accessible chromatin in the kidney

a) Distribution of PPA-filtered eGFRcrea-associated variants across accessible chromatin in the 29 tissues stratified according to their association with eGFRcrea alone (eGFRcrea only) or eGFRcrea validated by cystatin C and/or blood urea nitrogen (eGFRcrea/eGFRCys and/or BUN validated). **b**) Number of PPA-filtered eGFRcrea-associated variants per tissue located in accessible chromatin uniquely in the respective tissue. Variants were stratified according to their association with eGFRcrea

alone or additional validation by eGFRcys and/or BUN. **c**) Enrichment of REs from the kidney harbouring PPA-filtered eGFRcrea-associated variants in the 16 compartments defined by Meuleman et al.[20] Permutation-based peak overlap test, *, p_adjust<0.05; **, p_adjust<0.01. **d**) Overrepresentation analysis of genes annotated to the REs with PPA-filtered eGFR-associated variants in the cell type signature gene sets (c8) from MsigDB. **e**) Principal component analysis of accessible chromatin defined in 29 tissues and primary human tubular cells (PTC). Positions of PTC and kidney are highlighted in light blue. **f**) Overlap of PPA-filtered eGFRcrea-associated variants within accessible chromatin of the kidney (light blue) and isolated tubular cells (dark blue). **g**) Proportion of variants in REs of the kidney (light blue circles) or PTC (dark blue triangle) present in the set of variants defined by a PPA value equal or higher than the value on the x-axis. **h**) Genomic locus of the *TCEA2* and *RGS19* genes. Indicated are eGFRcrea-associated variants, eQTL variants (tub/int: tubulointerstitial compartment of the kidney from Han et al.[25]), H3K27ac ChIP-seq as well as ATAC-seq from PTC derived from the same individual (donor 1). The activity-by-contact (ABC) model suggests interactions of the variant carrying RE with both promoters (dark magenta). The variants rs6062343 and rs6062344 are eQTLs for *RGS19* in the tubulointerstitial compartment.

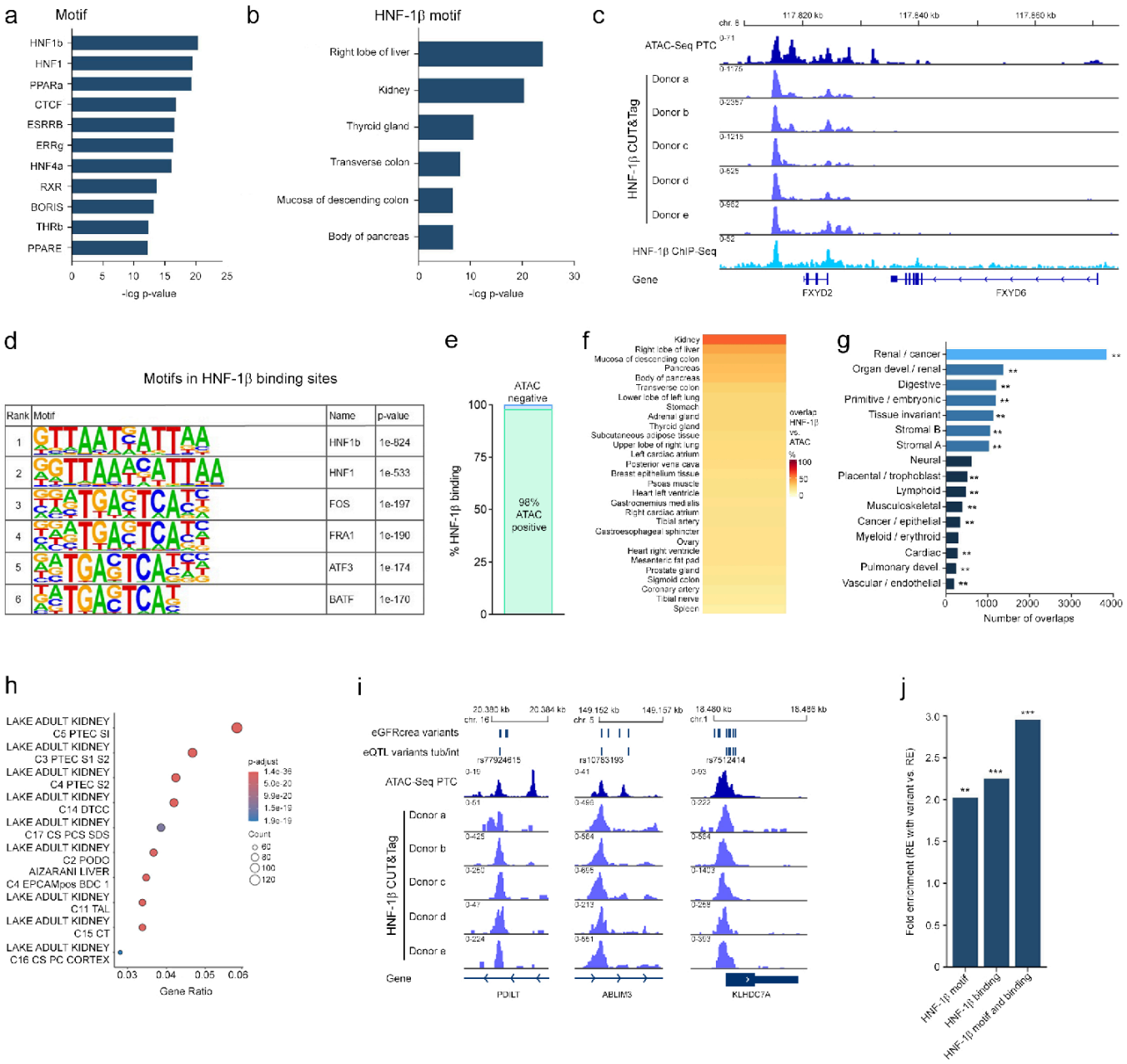

Figure 3

eGFR variants enrich at HNF-1β DNA-binding sites in human tubule cells

a) Motif enrichment analysis of REs with PPA-filtered eGFRcrea-associated variants in kidney tissue. **b**) Tissue with significant enrichment of the HNF-1β motif in REs with PPA-filtered eGFRcrea-associated variants in the respective tissue. **c**) HNF-1β CUT&Tag-seq tracks at the *FXYD2* gene locus generated from PTC from five individual donors. ATAC-seq from one PTC isolate and HNF-1β ChIP-seq from one isolate are shown for comparison. **d**) HOMER Motif enrichment analysis in the consensus set of 3,691

HNF-1β binding sites in PTC. **e**) Fraction of HNF-1β binding sites in accessible chromatin from PTC. **f**) Fraction of HNF-1β binding sites in PTC in accessible chromatin from the 29 tissues. **g**) Enrichment of HNF-1β binding sites in accessible chromatin from the 16-compartment model. Note: the number of overlaps is greater in the renal / cancer compartment than the number of binding sites, because one binding site can coincide with more than one site of accessible chromatin. Permutation-based peak overlap test, **, p_adjust<0.01. **h**) Overrepresentation analysis of genes annotated to the HNF-1β binding sites and genes expressed in the defined cell types (C8, MsigDB). **i**) Examples of variants in HNF-1β binding sites the *PDILT*, the *ABLIM3* and the *KLHDC7A* gene loci. Shown are five independent CUT&Tag-seq experiments from PTC, eGFR variants, eQTL variants from the tubulointerstitium (tub/int), and an ATAC-seq track from one PTC experiment for comparison. **j**) Enrichment of PPA-filtered eGFRcrea-associated variant carrying REs in regulatory elements harboring HNF-1β motifs or overlapping HNF-1β binding sites. χ^2-test, *p<0.001.

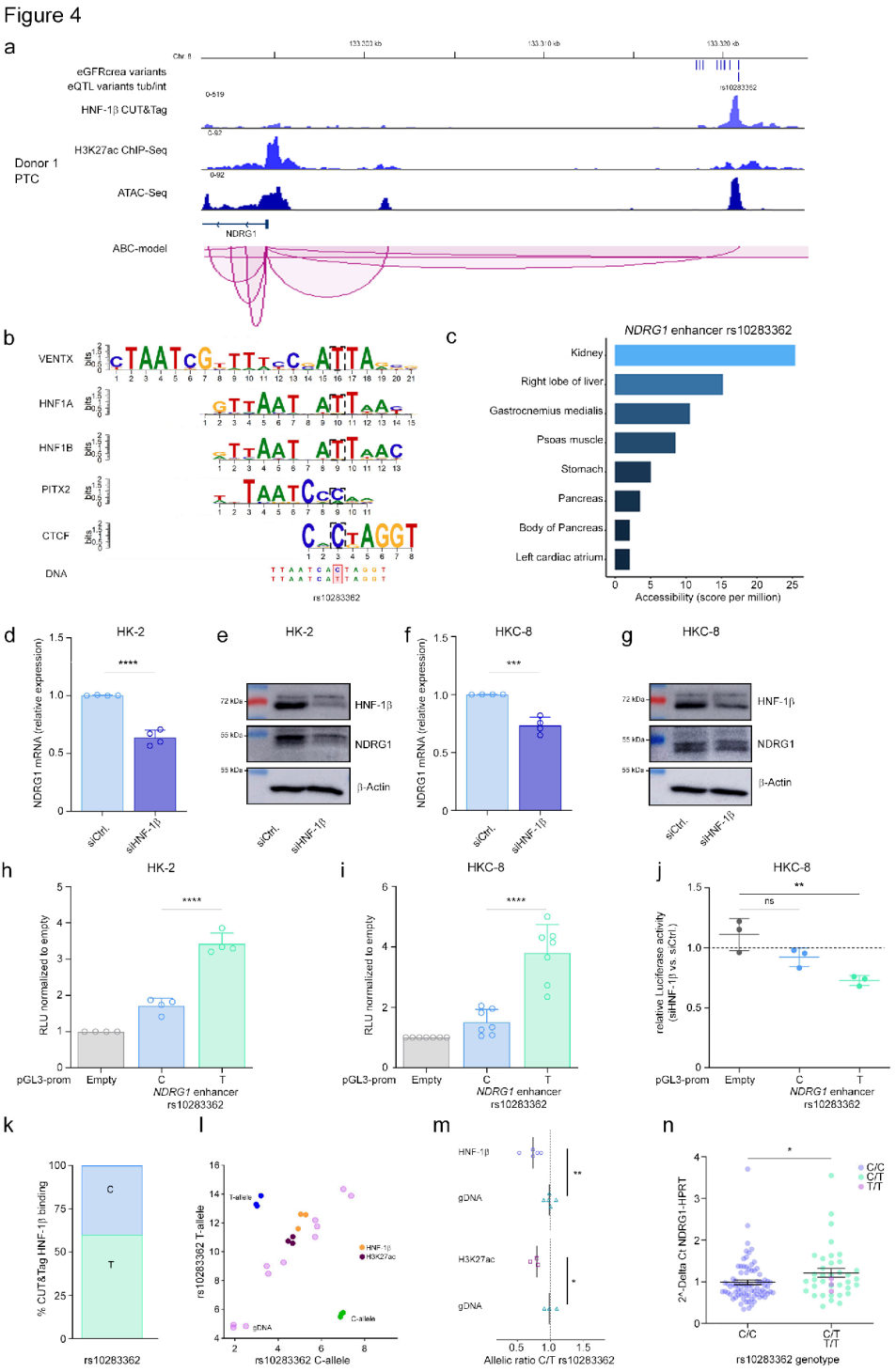

Figure 4

Allele specific HNF-1β binding at a polymorphic enhancer drives *NDRG1* gene regulation

a) Sequencing tracks from HNF-1β and H3K27ac CUT&Tag-seq as well as ATAC-seq performed in PTC at the rs10283362-*NDRG1* locus. Activity-by-contact (ABC) analysis was used to predict promoter-enhancer interactions. eGFRcrea-associated variants and eQTL variants from the tubulointerstitial compartment

(tub/int) are indicated. **b**) Motif breaker analysis at the rs10283362 position indicates interference with an HNF-1β motif. **c**) Accessibility at the rs10283362 associated RE in different tissues. **d**) NDRG1 mRNA expression in HK-2 cells after treatment with siRNA against HNF-1β (siHNF-1β) or non-targeting control siRNA (siCtrl.). n=4 independent experiments **e**) Immunoblot analyzes for HNF-1β and NDRG1 protein in samples from HK-2 cells after treatment with siRNA against HNF-1β (siHNF-1β) or non-targeting control siRNA (siCtrl.). **f**) NDRG1 mRNA expression in HKC-8 cells after treatment with siRNA against HNF-1β (siHNF-1β) or non-targeting control siRNA (siCtrl.). n=4 independent experiments. **g**) Immunoblot analyzes for HNF-1βand NDRG1 in samples from HKC-8 cells after treatment with siRNA against HNF-1β (siHNF-1β) or non-targeting control siRNA (siCtrl.). **h,i**) Reporter assay using sequences covering the regulatory element including variant rs10283362. Plasmids containing either the C- or the T-allele for rs1028332 or pGL3-promoter vector backbone (empty) were transfected into HK-2 and HKC-8 cells. Reporter luciferase activity was measured after 24 hours. Luciferase values were from four (HK-2, h) or seven (HKC-8, i) independent transfections and normalized to activity of co-transfected β-galactosidase. Values of C- or T- constructs were normalized to values of empty pGL3-promoter vector transfections. **j**) HNF-1βknock down in HKC-8 was followed by a reporter assay using the pGL3-promoter vector constructs from h). 24 hours after siRNA transfection cells were transfected with pGL3-promoter vectors. Luciferase values were from three independent transfections and normalized to activity of co-transfected β-galactosidase and values of control siRNA transfected cells. Results are from three independent experiments. **k**) Proportions of sequences covering rs10283362 captured by HNF-1β CUT&Tag in four PTC isolates heterozygous for the variant. **l**) Allele-specific genotyping of HNF-1β -(orange) and H3K27ac-CUT&Tag samples (violet) of heterozygous PTC with Taqman probes against rs10283362. HNF-1β and H3K27ac samples revealed a shift to the T allele of rs10283362, which indicates an enrichment of bound DNA fragments with a T allele. DNA from individuals homozygous for the T- or the C-allele are shown as control. Serial dilution of genomic DNA of the same individual is in light pink. Data is from cells of one PTC culture. **m**) Allelic ratio of rs10283362 in HNF-1β and H3K27ac CUT&Tag samples from four individuals heterozygous for the variant. **n**) NDRG1 mRNA expression normalized to house-keeping gene HPRT in PTC from individuals with homozygous (C/C; n=97 (blue) or T/T (light pink); n=3) or heterozygous (C/T (light turquoise); n=41) rs10283362 genotype. Data shown in mean +/- SD. Unpaired, two-tailed t-test was performed; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$

Figure 5

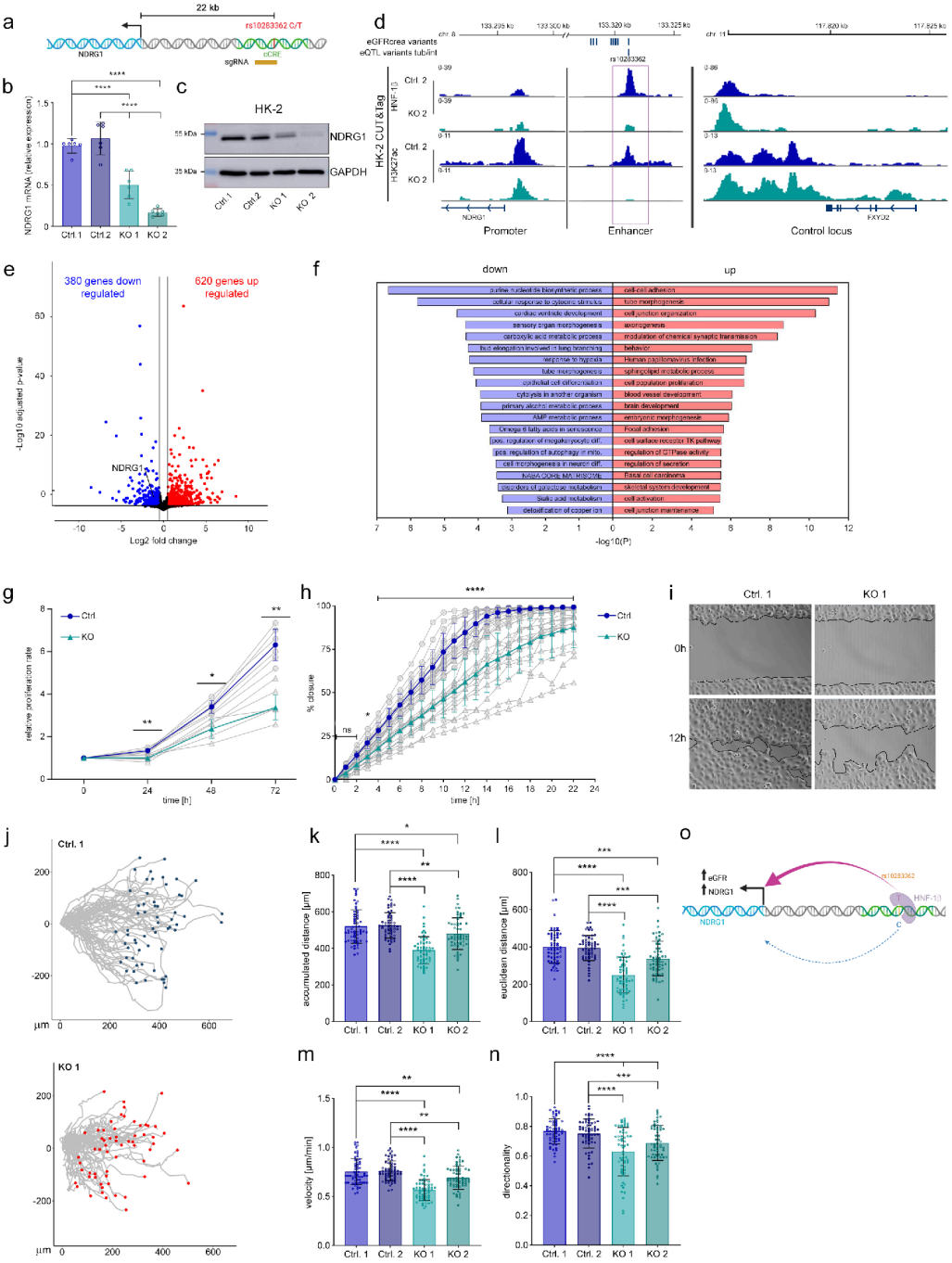

Figure 5

The rs10283362 carrying *NDRG1* enhancer promotes cell proliferation and directional migration of tubule cells

a) Schematic of the *NDRG1* locus, the candidate cis-regulatory element (cCRE) with variant rs10283362 and the position of the single guide RNA (sgRNA) used for CRISPR/Cas9 experiments. **b**) NDRG1 mRNA

expression in HK-2 clones of cells after CRISPR/Cas9-mediated knock out of the rs10283362-carrying element (KO) compared to non-targeting sgRNA transfected clones of cells (Ctrl.). Data was from 6 independent RNA-isolations, was normalized on values from one sample from Ctrl.1, and is shown in mean +/- SD. Ordinary one-way ANOVA for multiple comparisons was performed; ****, p<0.0001. **c**) NDRG1 protein expression in the two control and *NDRG1* enhancer knock-out clones of cells. **d**) HNF-1β and H3K27ac CUT&Tag-seq tracks from control (Ctrl. 2, blue) and *NDRG1* enhancer KO cells (KO2, turquoise). Shown are the *NDRG1* promoter, the *NDRG1* enhancer and the HNF-1β binding control locus *FXYD2*. **e**) Volcano plot from RNA-seq experiments in HK-2 *NDRG1* enhancer knock out vs. intact enhancer control clones of cells. **f**) GO Hallmark analysis of down- or upregulated genes from the RNA-seq experiment (KO vs Ctrl.) using Metascape platform. DEGs were based on adjusted p-value < 0.05 and $\log_2$(fold change) ≥ $\log_2$(0.5) or ≤ $\log_2$(-0.5). **g**) Proliferation assay of control (blue) and *NDRG1* enhancer KO (turquoise) clones of cells using an MTS assay. Data was generated in three independent experiments per clone of cells and is shown in mean +/- SD. Multiple, unpaired t-test was performed; *, p< 0.05; **, p<0.01. **h**) Wound healing assay of control and *NDRG1* enhancer KO clones of cells imaged in a live cell imaging experiment. Depicted is the percentage of closure of the gap over time [h]. Data was from n=6 independent experiments per clone of cells and is shown in mean +/- SD. Two-way ANOVA for multiple comparisons was performed; *, p<0.05; ****, p<0.0001. **i**) Representative pictures of control vs. *NDRG1* enhancer KO clones of cells from the wound healing assay at the timepoints 0 and 12 hours. **j**) Cell tracking from wound healing assay of 60 individual cells of clones (Ctrl. 1 and KO 1) shows the route migrated by each cell over a duration of 12 hours. **k, l**) Accumulated distance (k) and Euclidean distance (l) for the 60 cells per condition (Ctrl. 1, Ctrl. 2, KO 1 and KO 2) over 12 hours in wound healing assay. **m**) Velocity [µm/min] for the 60 cells within the wound healing assay. **n**) Computation of directional migration of the cells of clones (Ctrl. and KO). Whereby a value of 1 represents a linear movement, a value less than 1 indicates an undirected migration. Data shown in mean +/- SD. Ordinary one-way ANOVA for multiple comparisons was performed; *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001. **o**) Schematic how rs10283362 interferes with HNF-1β binding to the *NDRG1* enhancer and differentially regulates *NDRG1* expression and eGFR.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.xlsx
- SupplementaryMaterial.pdf