

Whole genome sequencing refines stratification and therapy of patients with clear cell renal cell carcinoma

Received: 28 November 2023

Accepted: 17 June 2024

Published online: 15 July 2024

 Check for updates

Richard Culliford ^{1,27}, Samuel E. D. Lawrence^{1,27}, Charlie Mills ^{1,27}, Zayd Tippu^{2,3,4,27}, Daniel Chubb¹, Alex J. Cornish ¹, Lisa Browning⁵, Ben Kinnersley ^{1,6}, Robert Bentham⁶, Amit Sud ¹, Husayn Pallikonda ^{2,3,4}, The Renal Cancer Genomics England Consortium*, Anna Frangou ^{7,8,9}, Andreas J. Gruber ¹⁰, Kevin Litchfield ¹¹, David Wedge ^{12,13}, James Larkin^{2,3}, Samra Turajlic ^{2,3,4} & Richard S. Houlston ¹ ✉

Clear cell renal cell carcinoma (ccRCC) is the most common form of kidney cancer, but a comprehensive description of its genomic landscape is lacking. We report the whole genome sequencing of 778 ccRCC patients enrolled in the 100,000 Genomes Project, providing for a detailed description of the somatic mutational landscape of ccRCC. We identify candidate driver genes, which as well as emphasising the major role of epigenetic regulation in ccRCC highlight additional biological pathways extending opportunities for therapeutic interventions. Genomic characterisation identified patients with divergent clinical outcome; higher number of structural copy number alterations associated with poorer prognosis, whereas VHL mutations were independently associated with a better prognosis. The observations that higher T-cell infiltration is associated with better overall survival and that genetically predicted immune evasion is not common supports the rationale for immunotherapy. These findings should inform personalised surveillance and treatment strategies for ccRCC patients.

Renal cell carcinoma (RCC) is an increasing global health problem with 431,000 new diagnoses each year, set to increase to 666,000 by 2040¹. Around 75% of RCCs are clear cell RCC (ccRCC) tumours. These cancers have a variable clinical course and while 75–80% of patients present with apparently localised disease and are offered curative intent treatment 30% will subsequently relapse². There is therefore a pressing

need for more accurate risk stratification, to guide clinical decisions relative to therapy and surveillance.

While therapeutic advances in the treatment of metastatic ccRCC have been made with the advent of antiangiogenic targeted therapies and immune checkpoint inhibitors (ICIs) only a subset of patients experience durable clinical benefit. Importantly, clinical biomarkers

¹Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK. ²Renal and Skin Units, The Royal Marsden NHS Foundation Trust, London, UK. ³Melanoma and Kidney Cancer Team, The Institute of Cancer Research, London, UK. ⁴Cancer Dynamics Laboratory, The Francis Crick Institute, London, UK. ⁵Department of Cellular Pathology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. ⁶Department of Oncology, University College London Cancer Institute, London, UK. ⁷Nuffield Department of Medicine, Big Data Institute, University of Oxford, Oxford, UK. ⁸Algebraic Systems Biology, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. ⁹Algebraic Systems Biology, Centre for Systems Biology Dresden, Dresden, Germany. ¹⁰Department of Biology, University of Konstanz, Konstanz, Germany. ¹¹Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK. ¹²Manchester Cancer Research Centre, University of Manchester, Manchester, UK. ¹³NIHR Manchester Biomedical Research Centre, Manchester, UK. ²⁷These authors contributed equally: Richard Culliford, Samuel E. D. Lawrence, Charlie Mills, Zayd Tippu. *A list of authors and their affiliations appears at the end of the paper. ✉ e-mail: richard.houlston@icr.ac.uk

fail to reconcile the variable disease course following surgery^{3–5} in patients receiving adjuvant PD1-therapy.

The need to understand ccRCC biology, to inform development of novel therapies and better predict patient outcomes, has been a major motivation in sequencing studies. While these projects have identified recurrent gene mutations and chromosomal rearrangements, analyses have primarily been based on whole-exome sequencing or panel testing of cancer-associated genes, hence the full complement of drivers is incomplete. Correspondingly, studies of the relationship between clinical parameters and genomic alterations have been limited^{6–9}.

To advance our understanding of ccRCC, we analyse whole genome sequencing (WGS) data from 778 ccRCC patients recruited to the UK Genomics England (Gel) 100,000 Genomes Project (100kGP)^{10,11}.

Results

The Gel cohort

The analysed cohort (100kGP, release v14) comprised tumour-normal (T/N) sample pairs from 778 patients (mean age 63 years, range 25–88 years) with primary ccRCC recruited to 100kGP through 13 Genomic Medicine Centres across England (Fig. 1). Comprehensive clinicopathology information on the patients is provided in Supplementary Data 1. We restricted our WGS analysis to samples with high-quality data from polymerase-chain-reaction (PCR) free (allowing for accurate uniform coverage without sequencing bias from PCR duplicates), flash-frozen fresh tumour samples ('Methods'). For 29 of the patients, WGS data on multi-regional sampling of tumours was available (2–4 samples per tumour, 94 samples in total). In addition to using variant calls from the 100kGP analysis pipeline we: (i) removed alignment bias introduced by ISAAC soft clipping of semi-aligned reads¹²; (ii) called tumour copy number using Battenberg¹³; (iii) called structural variants (SVs) from a consensus of Manta¹⁴, LUMPY¹⁵, and DELLY¹⁶; (iv) removed insertion-deletions (indels) within 10 base pairs (bp) of a common germline indel. Complete details on sample curation, somatic variant calling, and annotation of mutations are provided in the "Supplementary Methods".

Restricting our analysis to WGS data on one sample per patient ('Methods'), we identified 4,267,943 single nucleotide variants (SNVs), 699,100 indels, and 19,756 chromosomal rearrangements or structural

variants (Supplementary Data 2). While the median tumour mutational burden (TMB) was 2.07/Mb, three tumours displayed a hypermutated phenotype: i.e., excessively high SNV/indel mutation burden (maximal SNV/Mb = 33.65, maximal indel/Mb = 21.77). Twenty-two of the patients (2.8%) were carriers of pathogenic germline variants in one of the well-established RCC susceptibility genes (*CHEK2*, *FH*, *MITF*, *SDHA*, *VHL*)¹⁷ and 10 (1.2%) in a pan-cancer susceptibility gene (*ATM*, *BRCA2*, *BRIPI*, *FANCM*, *MSH6*, *PALB2*, *PMS2*; Supplementary Data 3). Four of the 32 patients had a prior history of non-RCC cancer (thyroid, prostate, testicular and chronic lymphocytic leukaemia).

Driver mutations

Protein-coding driver gene identification at the base pair level was performed using IntOGen¹⁸, which incorporates seven complementary algorithms. A total of 38 genes were identified as driver genes, including 25 well-recognized drivers and 13 which either had not been reported previously or have frequencies <1% in landmark genomics studies of ccRCC^{19–23} (Fig. 2). Of the mutations annotated by AlphaMissense²⁴, 72.4% of driver gene missense SNVs (436/602) were predicted to be pathogenic as compared to 29.8% (11,112/37,297) of the missense SNVs in non-driver genes ($P = 4.2 \times 10^{-112}$). The major known ccRCC driver genes were mutated at close to reported frequencies^{19–21,23,25–27}: *VHL* (80.2%), *PBRM1* (49.9%), *SETD2* (17.7%) and *BAP1* (11.8%). Subclonal drivers, such as *TSC1* were, however seen with lower frequency, which is likely to be the consequence of our reliance on a single biopsy rather than multi-regional sampling as per studies such as TRACERx⁸ (Supplementary Data 4). All of the candidate coding drivers were detected at low frequencies, as expected given the scale of previous exome sequencing studies (0.3–2.4%; Supplementary Fig. 1a–m, Supplementary Data 4). The increased power of our study has enabled us to assign driver status to a number of genes recurrently mutated at low frequency in other studies (Supplementary Fig. 2).

Mutations in the candidate drivers were frequently accompanied by loss of heterozygosity (LOH) (see 'Recurrent structural and copy number alterations'), implying loss of function (Supplementary Fig. 3). Furthermore, leveraging TCGA²⁸ and GTEx²⁹ expression data as well as DepMap gene perturbation screening^{30,31} in addition to referencing literature provide supporting evidence that 6/13 of the candidate genes we identify have relevance to the biology of RCC

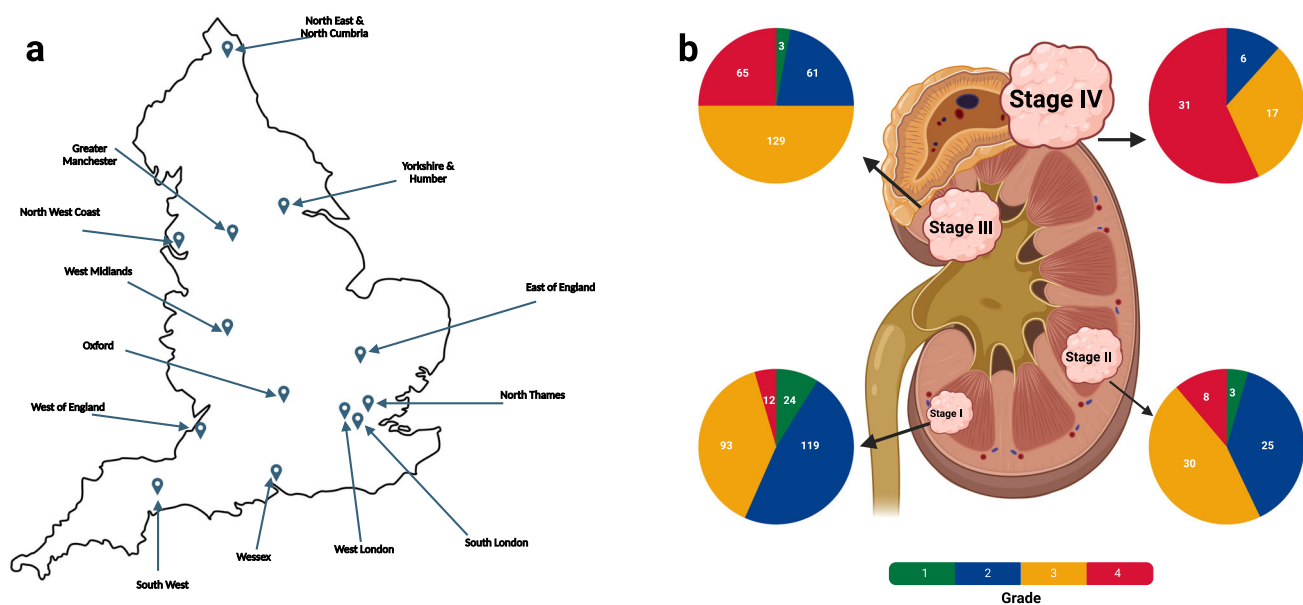


Fig. 1 | Overview of the Gel cohort of ccRCC patients. **a** The location of the 13 Genomic Medicine Centers (GMCs) across England from which patients were recruited. **b** The breakdown of the cohort by tumour grade and stage. Figure 1

created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

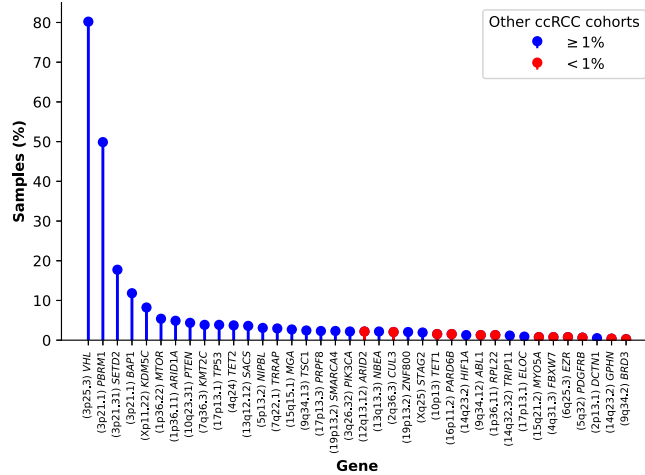


Fig. 2 | Frequency of patients with nonsynonymous mutations in driver genes. The cohort frequency of a driver gene reported as being above (blue) or below (red) 1% in other ccRCC cohorts is indicated.

(Supplementary Fig. 4a, b, Supplementary Data 4). Candidate drivers emphasise the central importance of epigenetic modification in the development of ccRCC through SWI/SNF mediated chromatin remodelling (*BRD3*, *ARID2*), histone deubiquitination (*CUL3*, *FBXW7*) as well as the role of methylcytosine dioxygenase activity (*TET1*). Biological mechanisms highlighted by the candidate drivers are shown in Fig. 3a–e. These included new vessel formation (*PDGFR-β*), ribosomal activity (*RPL22*), cytoskeletal interactions (*EZR*, *GPHN*, *NBEA*), cell division and cell polarisation (*PARD6B*). Mutations of *ARID2*, *CUL3*, *TET1*, *FBXW7*, *EZR* and *GPHN* were frequently accompanied by copy number loss consistent with their predicted/documentated tumour suppressor roles (Supplementary Data 4).

As previously reported³², mutations in *VHL* and *ELOC*, which both play a key role in oxygen sensing and degradation of hypoxia-inducible factors, were mutually exclusive ($Q = 0.002$). Co-occurrence analysis also supported mutually exclusive relationships: *BAP1* and *PBRM1* ($Q = 4.36 \times 10^{-12}$), SWI/SNF mediated chromatin modifiers *PBRM1* and *ARID2* ($Q = 0.04$), and *SETD2* and *BAP1* ($Q = 0.049$). In contrast, mutations of *ARID2* and *ELOC* ($Q = 0.001$), and *PBRM1* and *SETD2* ($Q = 1.33 \times 10^{-5}$), tended to co-occur (Supplementary Data 5). In five tumours, *BAP1* mutations were the sole driver.

To search for non-coding drivers in gene promoters, untranslated and non-canonical splice regions, we used OncodriveFML³³, ActiveDriverWGS³⁴, and negative binomial regression adjusting for trinucleotide mutational context³⁵. By collating and post processing of these data we nominated nine significant non-coding elements as potential candidate cancer drivers (Supplementary Methods).

The *TERT* promoter associations were primarily driven by the canonical mutations 5:1295113G > A, and 5:1295135G > A which have also been documented in TRACERx³⁶ and are early drivers in bladder cancer^{37–39}. The 5'UTR of *TERT* was also recurrently mutated with 5:1295046T > G, which has previously been implicated in glioma and bladder cancer accounting for 23% of mutations within the region^{39,40}.

Our analysis also implicates the distal promoter of the *BRD2* gene, which together with other bromodomain proteins plays a role in histone acetylation modification in RCC. Other potential non-coding drivers included the distal promoter regions of *RSU1*, *FANK1*, and *BCAT2*^{41,42} (Supplementary Data 6).

Systematic analyses of cancer genomes provide an opportunity of estimating the number of patients eligible for a targeted therapy and identify opportunities for therapeutic interventions. We assessed the clinical actionability of driver gene mutations by referencing OncoKB Knowledge Base⁴³ (version 3.11), and found 60 unique alterations were

targetable (OncoKB Level 1–4), and were all at least Level 4 (compelling biological evidence supporting the biomarker being predictive of drug response). We also examined the COSMIC Mutation Actionability in Precision Oncology⁴⁴ database highlighting an additional 708 unique alterations which are potentially targetable (Supplementary Fig. 5a, b, Supplementary Data 7).

Recurrent structural and copy number alterations

In addition to the previously reported common 3p loss, 5q gain and 14q loss^{7,36} we identified 25 other arm-level alterations that occurred more frequently than expected (Fig. 4a, Supplementary Data 8). We used GISTIC2⁴⁵ to identify genomic regions recurrently affected by focal amplifications and deletions ($Q < 0.05$; Fig. 4a, Supplementary Data 9). Aside from the previously reported CNAs⁷, including del9p21.3 (*CDKN2A*), del3p12.2 (*GBE1*), amp5q35.3 (*SQSTM1*) and amp 8q24.21 (*MYC*) we identified four candidate CNAs: amp2q31.1 (*ACVR2A*, *CASP8*, *NFE2L2*, *PMS1*, *SF3B1*), amp13q34 (*ERCC5*), amp12p11.21 and del22q11.23. Of the genes implicated by these candidate focal amplifications, *NFE2L2* and *SF3B1* are documented to be oncogenic^{46–49}. 16.6% of tumours showed whole genome duplication (WGD), a finding almost identical to the 15% reported by TracerX⁵⁰ (Supplementary Data 2). Complex chromosomal rearrangements were a feature of 60% of tumours with two-thirds of these displaying hallmarks of chromothripsis. As previously documented^{7,36} the most frequent pattern on 3p loss was from rearrangement between 3p and 5q, ascribable to chromothripsis.

We identified 37 hotspots of recurrent simple SVs ($FDR < 0.05$) by piecewise constant fitting adjusting for local genomic features known to influence rearrangement density (chromatin accessibility, repeated elements, GC content, replication timing, gene density and expression). Fragile sites are prone to rearrangement (possibly due to replication error) and tend to co-occur with large, late-replicating genes. SVs occurring at fragile sites are hence likely to be the consequence of mechanistic rather than selective factors. After excluding 14 SV hotspots mapping to potential fragile sites, we identified 23 SV hotspots (Fig. 4b, Supplementary Figs. 6a–t, 7a–l, 8a–e, Supplementary Data 10). We identified a total of 66 breakpoints within 5p15.33, spanning *TERT*. These included, a deletion breakpoint and an unclassified event 2 kb downstream of the *TERT* promoter, and tandem duplications overlapping *TERT* ($n = 5$). In tumours from the 34 patients with a *TERT* 5' UTR mutations, there were no overlapping unclassified/tandem-duplication events or a SV deletion/unclassified promoter breakpoints; an observation consistent with earlier findings³⁶.

Mutational signatures

To gain insight into mutational processes in ccRCC, we extracted single-base substitution (SBS), double-base-substitution (DBS) and indel (ID) signatures de novo and related those to known COSMIC signatures (v3.2) using SigProfilerExtractor^{51,52}. In the majority of cancers, single base substitutions could be assigned to signatures SBS5/SBS40 and SBS1 (nomenclature as per COSMIC) resulting from clock-like mutagenic processes. Other signatures recovered with known specific underlying aetiology include those associated with oxidative damage (SBS18), defective base excision repair (SBS30), APOBEC (SBS2, SBS13), tobacco smoking (SBS4, DBS2, ID3) and aristolochic acid (SBS22) (Fig. 5, Supplementary Fig. 9a, b, Supplementary Data 11). Three of the 7 tumours with SBS2 or SBS13 harboured a somatic mutation in an APOBEC gene, however, none of them are pathogenic. While the incidence of renal cancer has been linked to aristolochic acid exposure in residents of Danube river countries⁵³, 88% of patients with SBS22 tumour activity in the Gel cohort were self-reported to be white British. SBS31 and SBS35 have been attributable to platinum chemotherapy. We recovered SBS35 in four cases, but none had a reported past history of platinum chemotherapy. In contrast the tumours from five patients, which had a past history of a non-RCC cancer and had

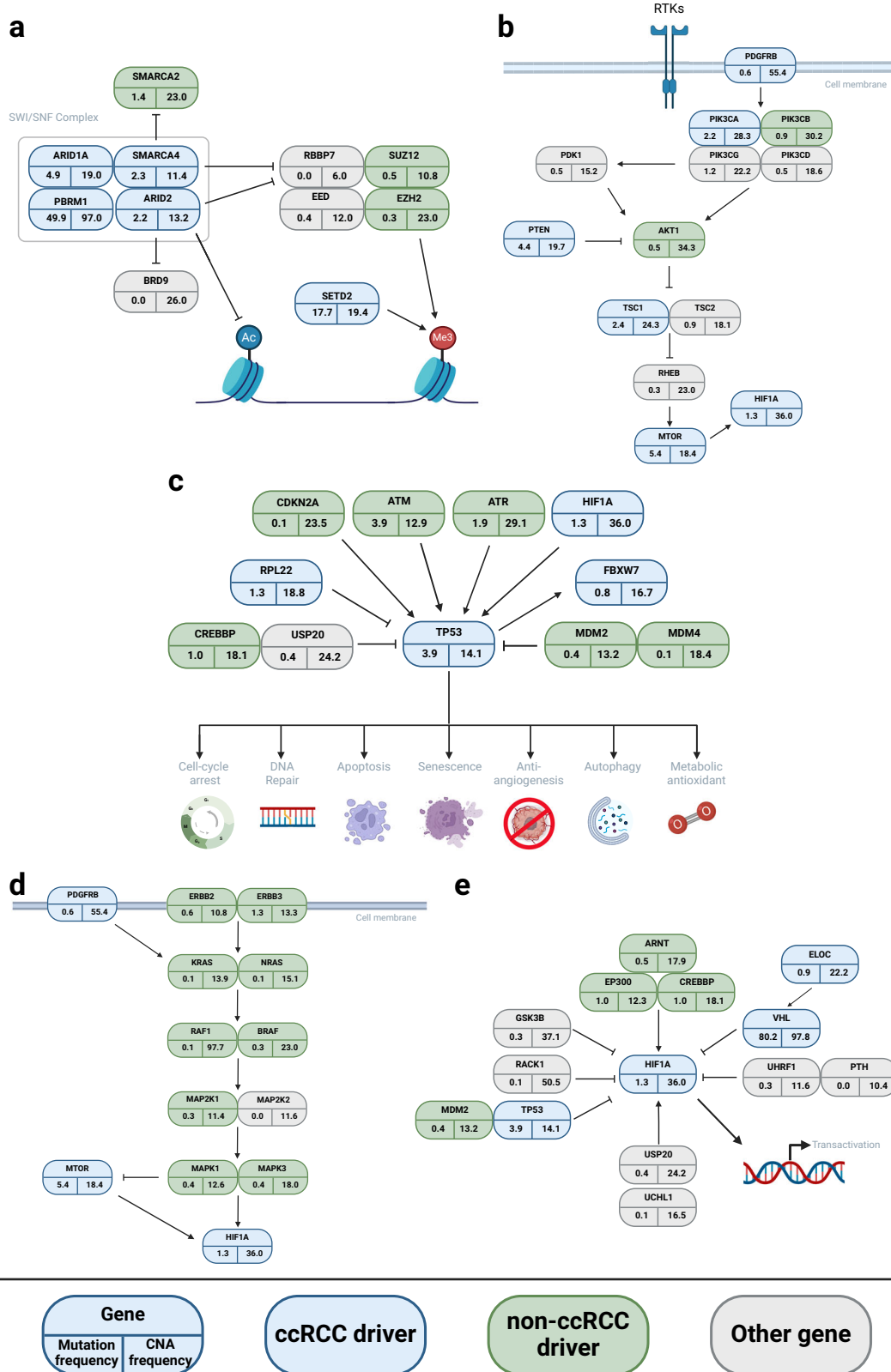


Fig. 3 | Biological pathways in ccRCC. a The *SWI/SNF* pathway. **b** The *PI3K/AKT/MTOR* signalling pathway. **c** The *TP53* pathway. **d** The *RAS/ERK* pathway. **e** The *VHL/HIF1A* and hypoxia pathway. Driver genes identified shown in blue, non-ccRCC driver genes in green and other pathway genes in grey. Non-ccRCC driver genes are defined as those identified in any other cancer. The number in the bottom left is the

nonsynonymous mutational frequency and the number in the bottom right the copy number alteration (CNA) frequency. Figure 3 created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

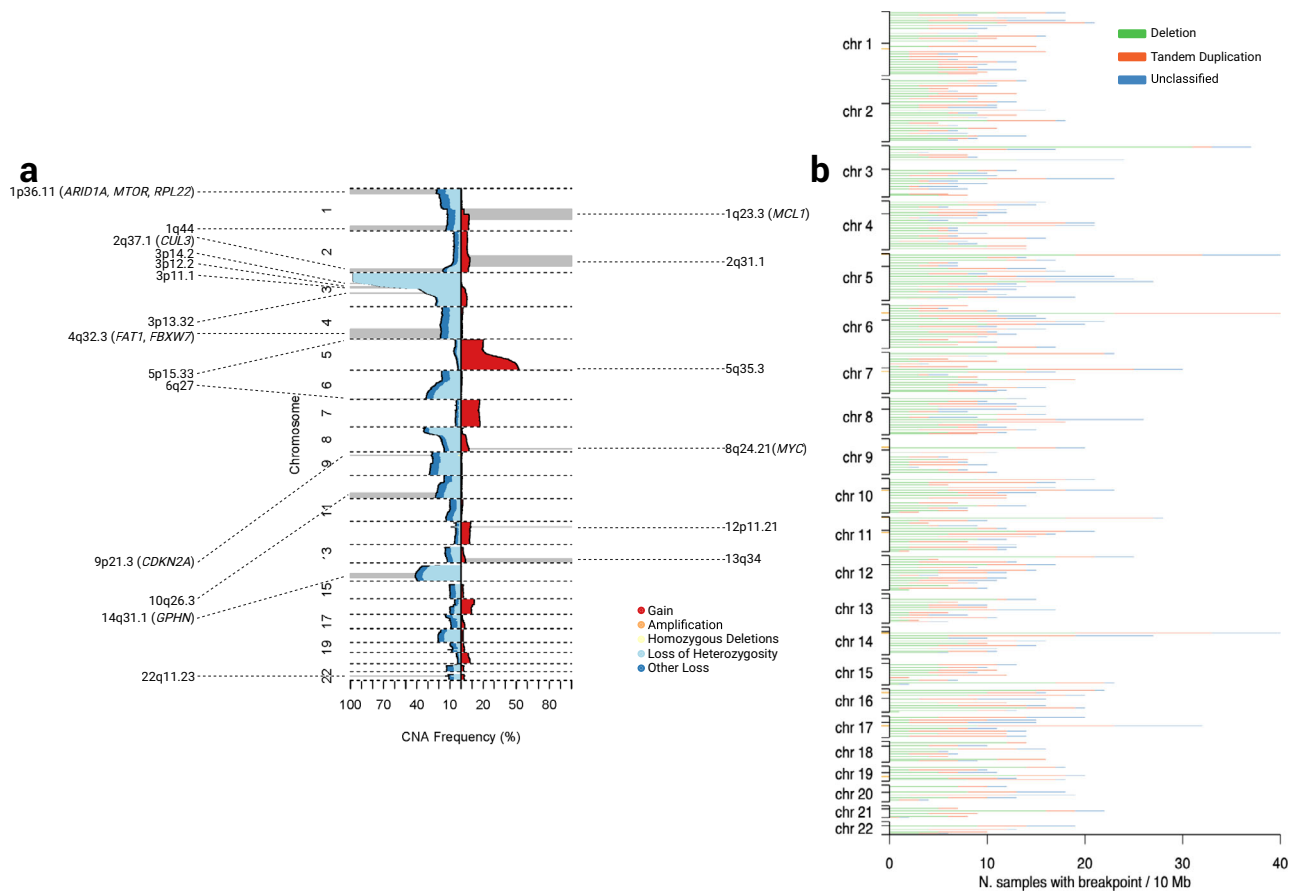


Fig. 4 | Copy number alterations and structural variants. **a** Frequency of copy number alterations across the ccRCC cohort. Copy number losses are coded in blue shades and copy number gains shown in red. The focal copy number alterations, as identified by GISTIC, are annotated along with predicted target gene. **b** The distribution of structural variants across the ccRCC cohort. Structural variants are

classified as deletions, tandem duplications or unclassified structural variants. The black ticks on the y axis correspond to the chromosome start, centromere and end position while the orange ticks represent identified structural variant hotspot regions.

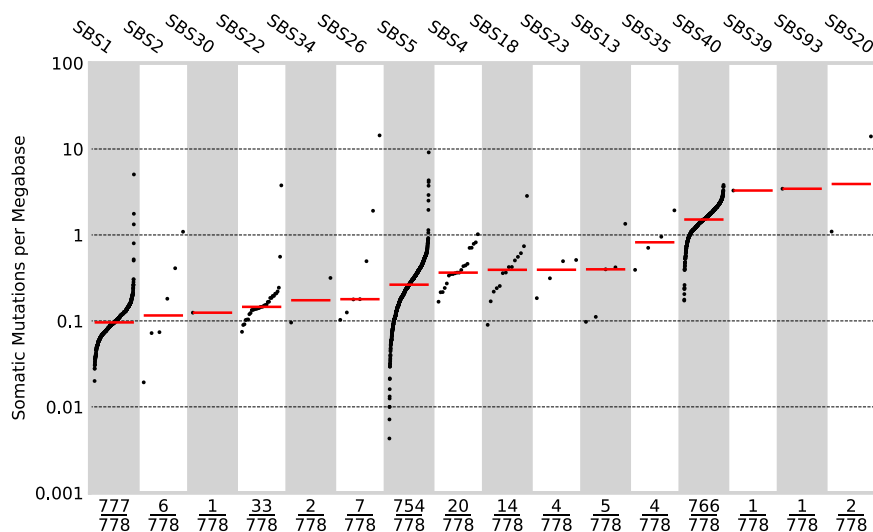


Fig. 5 | Mutational signatures. The mutational burden of single base substitution signatures ($n = 778$), as defined in COSMIC. The x axis indicates the number of samples for which each signature was active. The y axis is the mutational burden of

the given signature. SBS1, SBS5 and SBS40, present in the majority of samples, are clock-like signatures. The higher average mutational burden of SBS40 signifies that it is the predominant signature active in the majority of samples.

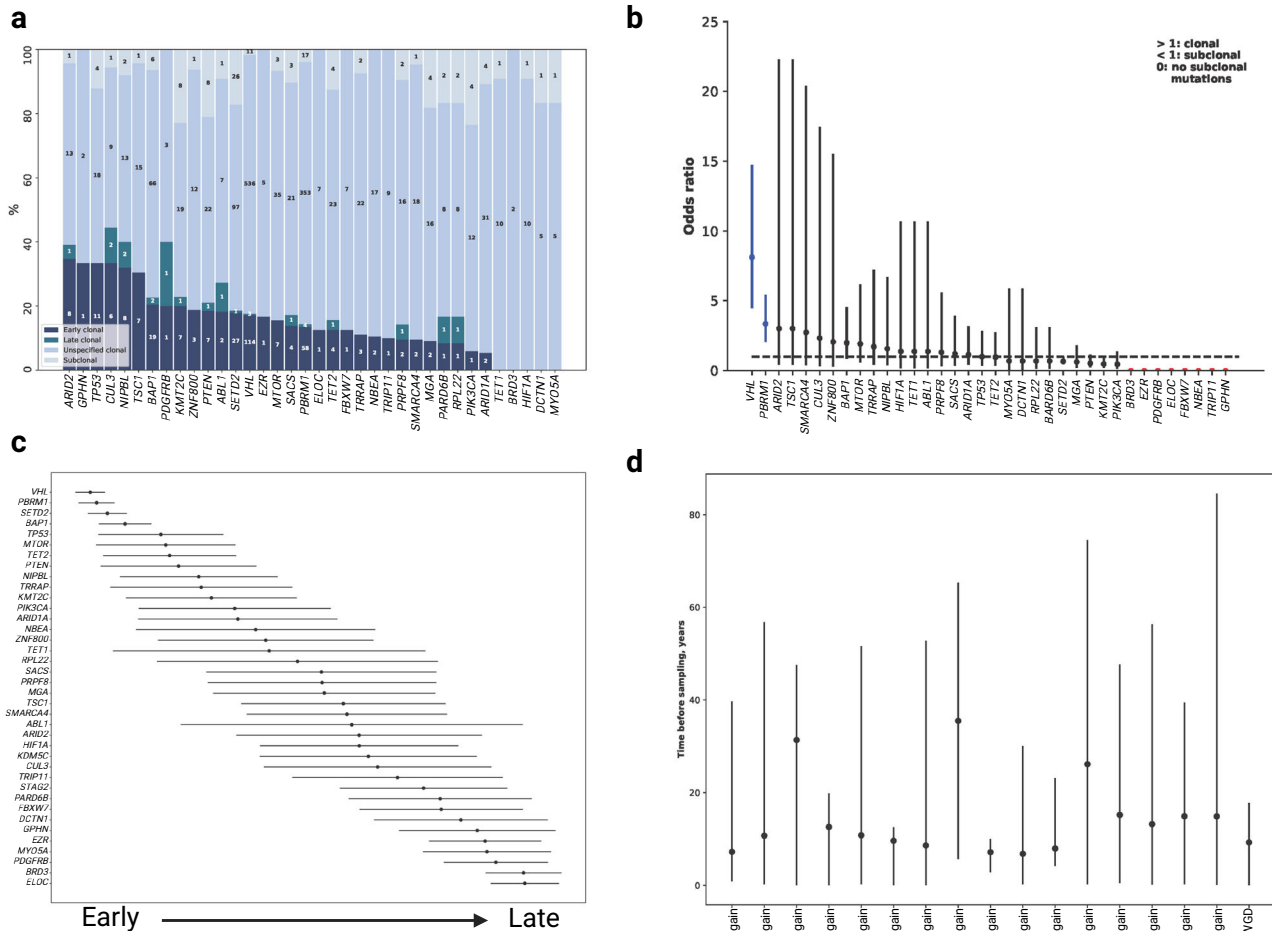


Fig. 6 | Mutational timing. **a** The proportion of clonal and subclonal nonsynonymous mutations ($n = 775$) in driver genes. **b** Odds ratio (OR) with 95% confidence intervals that a mutation in a driver gene is clonal relative to mutations in the other driver genes; $OR > 1.0$ indicates a mutation is more likely to be clonal. The genes in blue are significantly more likely to be clonal or subclonal. Genes in red have no subclonal mutations. **c** The relative ordering of nonsynonymous mutations in

driver genes. The error bars correspond to the 95% confidence intervals of the average finishing position in the league model. The ordering was determined based on the distributions in (a) (Supplementary Methods). **d** Estimates, with 95% confidence intervals, of the real time at which copy number gains occur during tumour evolution.

received carboplatin or oxaliplatin did not display SBS31 or SBS35 (Fig. 5, Supplementary Data 11). To complement SigProfilerExtractor we searched for mutational signatures associated with defective mismatch repair (dMMR) and defective homologous recombination (dHR) using mSINGS⁵⁴ and HRdetect⁵⁵. Two of the 3 tumours with evidence of dMMR (2 with SBS20, SBS26 and hypermutated phenotype) harboured *MLH1* somatic mutations which were accompanied by LOH, while the other carried a *POLE* mutation of unknown significance. None of these 3 cases carried germline pathogenic MMR variants and during the period of the study none developed metastatic RCC or were treated with ICIs. No case showed evidence of dHR. Considering mutational signature activity between clonal and subclonal mutations we found no significant enrichment or depletion of any SBS signatures between clonal and subclonal mutations.

Ordering of mutational events

Using PhyloGicNTD⁵⁶ in conjunction with MutationTimer⁵⁷, we reconstructed the chronological ordering of focal CNAs and driver mutations. These methods estimate the time at which copy number gains occur by considering the fraction of mutations before and after the gain. In the regions of copy number gain individual mutations can therefore be estimated to be early or late events; subclonal mutations will, by definition, have occurred late in tumorigenesis. Across all

tumours gain of 5q were consistently earlier alterations. As expected, mutations in *VHL*, *PBRM1*, *SETD2* and *BAP1* were predicted to be early events, with higher odds of harbouring clonal mutations, generally occurring before corresponding CNAs. In contrast, mutations in *KDM5C*, *STAG2* and *BRD3* were late events (Fig. 6a–c). Estimating the chronological timing of CNAs under varying mutational rates and tumour initiation time (Supplementary Methods) implies WGD occurred on average 9.2 years before tumour sampling and gain of 5q, 35.5 years before sampling (Fig. 6d). Moreover, the estimated lead time of 5q gain and WGD were both correlated with age at presentation (adjusting for grade and stage $P = 9.4 \times 10^{-13}$ and 0.02, respectively).

Immune profile and evasion

Increased T-cell receptor alpha (TCRA) T-cell fraction⁵⁸ (i.e. fraction of T-cells present in the WGS sample), reflecting immune infiltration, was associated with increased tumour grade ($P = 4.0 \times 10^{-4}$; Supplementary Data 12). Considering TCRA T-cell fraction as a proxy for CD8+ infiltration, higher immune cell infiltration has previously been reported in a subset of inflamed *BAP1* or *PBRM1* mutated ccRCC tumours⁵⁹. Consistent with this observation in our study *BAP1* mutated tumours were associated with higher TCRA T-cell fraction (OR = 3.43, 95% CI: 1.91–6.15), but no similar relationship was shown for *PBRM1* (OR = 0.73, 95% CI: 0.53–1.01). We also noted that TCRA T-cell fraction was lower in

MTOR-mutated tumours (OR = 0.44, 95% CI: 0.2–0.97; Supplementary Data 13).

Using pVAC-Seq⁶⁰, we predicted 24,893 class I neoantigens across the 778 tumours (1–327 per tumour, median 26), resulting from: 66.5% missense mutations, 32.0% frameshift variants, 1.3% inframe deletions and 0.25% inframe insertions (Fig. 7a). Neoantigen burden was not correlated with stage or tumour grade (including presence of sarcomatoid features) but was positively associated with TMB (OR = 1.21, 95% CI: 1.18–1.23; Supplementary Data 12 and 14).

Examining evidence of immune evasion we LOH or mutation of HLA class I genes (*HLA-A*, *HLA-B*, *HLA-C*) and immune escape genes (Supplementary Methods). Using LOHHLA, we detected LOH of HLA in only 5.9% of tumours. It has been reported that LOH on HLA class I genes and 9p21.3 loss tend to co-occur⁶¹, suggesting a potential mechanism for immune escape. After adjusting for stage and grade we found no evidence to support such a relationship (OR = 1.32, 95% CI: 0.58–2.99; Supplementary Data 15). We also found no association between indel burden and HLA allele status (OR = 0.92, 95% CI: 0.82–1.03; Supplementary Data 16). Nonsynonymous mutations of HLA genes were rare (0.5%). An inactivating mutation in at least one of the 22 antigen presenting genes^{62,63} (APG) was seen in only 3.1% (24/778) of tumours (Fig. 7b). None of the APGs displayed a propensity for mutation. Collectively, on the basis of alteration of these escape pathways (HLA LOH, APG inactivation or HLA somatic mutation) only 9.0% (70/778) of tumours were predicted to exhibit some form of genetically-driven immune evasion.

Clinico-pathological relationships

Increased tumour grade was associated with necrosis ($P = 9.4 \times 10^{-15}$), increased TMB ($P = 0.001$) and mutational sub-clonality ($P = 0.02$). High SV count ($P = 1.5 \times 10^{-11}$), WGD ($P = 1.2 \times 10^{-6}$) and weighted genome instability index ($P = 1.7 \times 10^{-11}$) were associated with higher tumour grade. Consistent with previous literature⁶⁴, tumours with mutations in *BAP1* and *TP53* were more likely to present as high grade ($P = 1.8 \times 10^{-5}$, $P = 2.1 \times 10^{-6}$, respectively; Supplementary Data 12). The genomic landscape of several cancers have been shown to differ by age at diagnosis⁶⁵. To explore this possibility with respect to ccRCC we compared cases diagnosed younger than age 46 and older. This stratification provided no evidence for significant differences in genomic features other than TMB (Supplementary Data 12).

Information was available on 605 patients for us to examine the relationship between genomic features with overall survival (OS) and cancer specific survival (CSS) (Supplementary Figs. 10a–d, 11, 12a–f, 13, Supplementary Data 1, 17, 18). Tumour grade and stage and sarcomatoid features were strong predictors of OS and CSS (Supplementary Data 17 and 18). After adjusting for co-variants using Cox regression increased OS and CSS was associated with *VHL* (Hazard Ratio (HR) = 0.60, 95% CI: 0.36–0.98; HR = 0.60, 95% CI: 0.3–1.19) and *PBRMI* (HR = 0.64, 95% CI: 0.42–0.97; HR = 0.55, 95% CI: 0.32–0.95 respectively) mutation status. Given the co-occurrence of *VHL* and *PBRMI* mutations (84%, 325/388 of *PBRMI*-positive tumours were also *VHL* mutated), after adjusting for *VHL* status, the association of *PBRMI* mutational status was no longer statistically significant for either OS or CSS (HR = 0.68, 95% CI: 0.44–1.03; HR = 0.58, 95% CI = 0.33–1.02). Aside from *VHL*, mutation of no other driver gene showed an independent association with patient survival; acknowledging we had limited statistical power to demonstrate a relationship with less frequently mutated genes (Fig. 8a–e, Supplementary Fig. 11). After adjusting for *VHL* status, higher SV count was, however, associated with worse OS (HR = 1.01, 95% CI: 1.00–1.10) and CSS (HR = 1.01, 95% CI: 1.01–1.02; Fig. 8c, Supplementary Figs. 11, 12e, 13, Supplementary Data 18).

While we found no association between OS or CSS and either neoantigen burden or immune escape, a higher TCRA T-cell fraction was associated with a better OS (HR = 0.65, 95% CI: 0.43–0.99; Fig. 8d).

Aside from an association between DBS4 and CSS, we found no other significant associations between additional mutational signatures as independent predictors of either OS or CSS. We also did not find evidence to support a relationship between either OS or CSS with intra-tumour heterogeneity or wGII, both features which have previously been purported to influence prognosis^{25,66} (Supplementary Figs. 11 and 13, Supplementary Data 18).

We were able to examine the relationship between molecular features and progression-free survival (PFS) in 167 of the patients ascertained on the basis of being at intermediate-high risk of tumour recurrence on the basis of their Leibovich score⁶⁷ (Supplementary Data 1). While *VHL* status was not associated with better PFS we observed that *KDM5C* mutation was independently associated with worse outcome (HR = 1.98, 95% CI: 1.00–3.91; Supplementary Figs. 14a–h and 15, Supplementary Data 19) and a higher incidence of necrosis (OR = 4.81, 95% CI: 1.20–19.11; Supplementary Data 20). Thirty-seven of the 167 patients had received ICI therapy as a first or second line treatment and in 21 of these there was documented evidence of clinical benefit. Restricting our analysis to these 37 patients, we found no evidence of a difference in ICI response in patients with sarcomatoid tumours features and only deletion of 6q was associated with clinical benefit (OR = 7.66, 95% CI: 1.11–52.55). Although not statistically significant, but consistent with other reports⁶⁸, sarcomatoid histology was associated with better response to ICI (Supplementary Data 21).

Discussion

This study represents a comprehensive description of the genomic landscape of ccRCC by utilising a large cohort of WGS samples. We acknowledge that there are limitations to our analysis. Specifically, our reliance on short-read sequencing and lack of transcriptomic information. During the period of this study the classification of the renal tumours was conducted in accordance with the WHO Classification of Urinary and Male Genital Tumours (4th and 3rd Editions). Although there was no change to the diagnostic criteria for ccRCC between WHO 3 and 4, the classification of RCC has evolved with new entities recognised within the (current) WHO Classification (5th Ed, 2022) that were not previously ‘recognised’/accepted entities. Some of these mimic ccRCC and there is therefore a chance that a small number of cases within our study cohort might be re-classified if these were subject to contemporary review. Such considerations will also pertain to previously published studies conducted by TCGA and PCAWG. Accepting the caveats, as well as confirming established ccRCC driver genes we identify candidate drivers further highlighting oncogenic metabolism and epigenetic reprogramming as being central to ccRCC biology. Additionally, we confirm *pTERT* mutations as drivers, thereby further substantiating telomerase dysfunction in the development of ccRCC. Mutational signature analysis provides a mechanistic basis for known lifestyle and exposure risk factors as well as potentially indirectly suggesting additional ones. While we did not identify any new mutational signatures, our analysis provides further support for tobacco smoking being a risk factor for ccRCC⁶⁹.

The large size of our study, coupled with the standardised management protocols for ccRCC patients within the UK National Health System, has enabled us to investigate the correlation between molecular features and patient prognosis. The clinical course for many ccRCC patients with apparent same stage disease can be highly variable. Upfront identification of patients who are likely to relapse early offers the prospect of intervening preemptively to maintain remission. Furthermore, since metastatic ccRCCs are chemotherapy and radiotherapy resistant, identifying tumour sub-groups with targetable molecular dependencies has the potential to inform on biologically driven therapies.

The relationship between mutations in the major clonal driver genes and patient survival has been the subject of a number of previous studies, but findings have been inconsistent^{6,7,70–76}

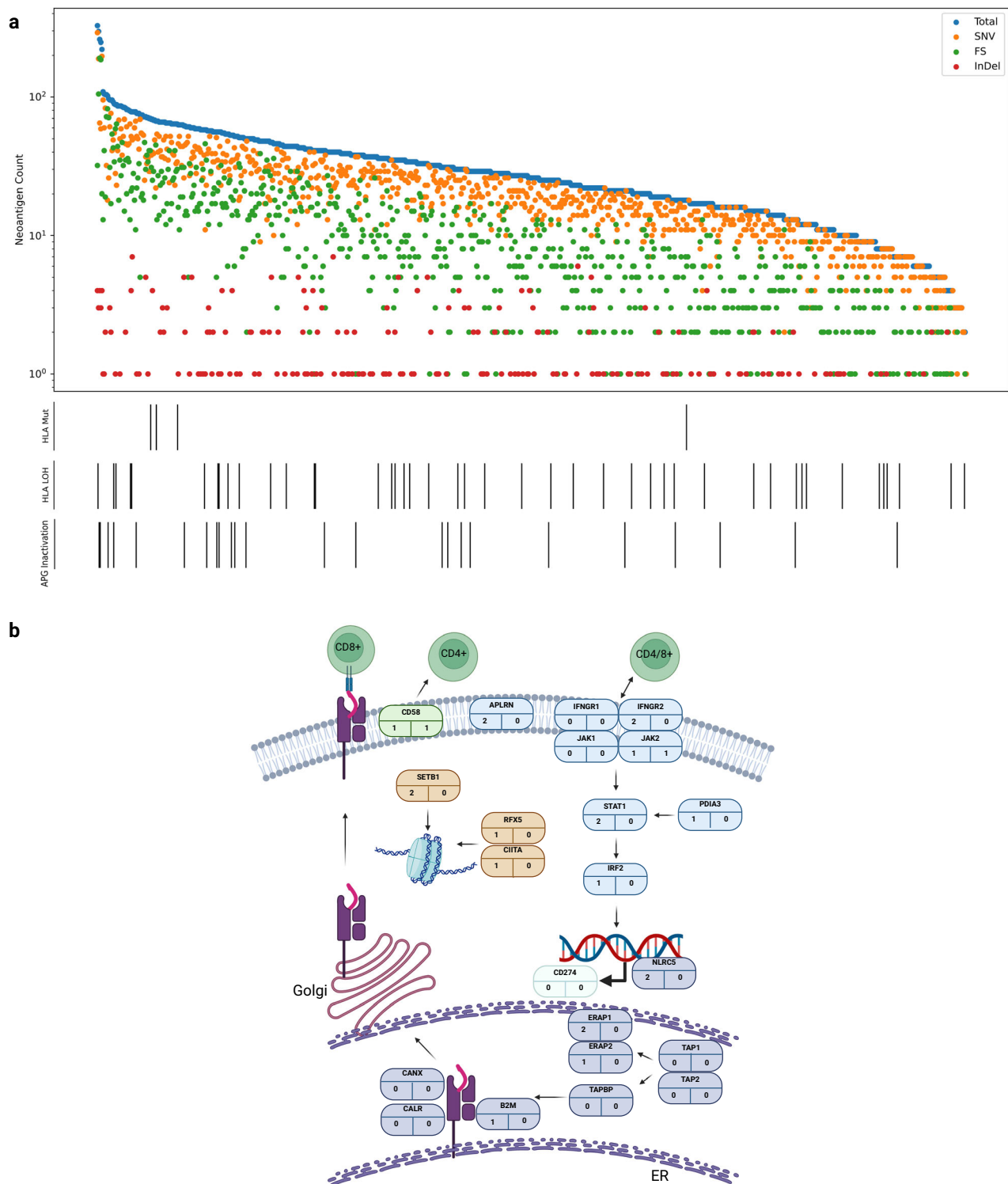


Fig. 7 | Immune landscape of ccRCC. a Neoantigen burden and immune escape mutations. Lower bars show antigen processing genes (APG) and human leucocyte antigens (HLA) alterations present in each cancer. **b** Somatic mutations in each of the antigen presentation pathway genes. The number in the bottom left is the truncating mutation count and the number in the bottom right is the number of

biallelic nonsynonymous mutations. APGs in purple, IFN- γ pathway genes in blue, epigenetic modifier genes in brown, CD274 comprises the PD-L1 receptor and CD58 receptor is encoded by CD58. Figure 7b created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

(Supplementary Data 22). While some studies^{7,9} have reported *BAP1* mutations being associated with a worse clinical outcome, other studies^{71,75} have failed to demonstrate any relationship. As previously documented^{9,77}, and herein, *BAP1* mutations are strongly associated with increased grade and after adjustment we failed to show support

for an independent relationship. In our study, we, however, show *VHL* mutation status was independently associated with an improved OS, consistent with a recent study⁷⁴. *VHL* mutations are early events of ccRCC development whereas other mutated genes are acquired later therefore they might be assumed to play more of a role in disease

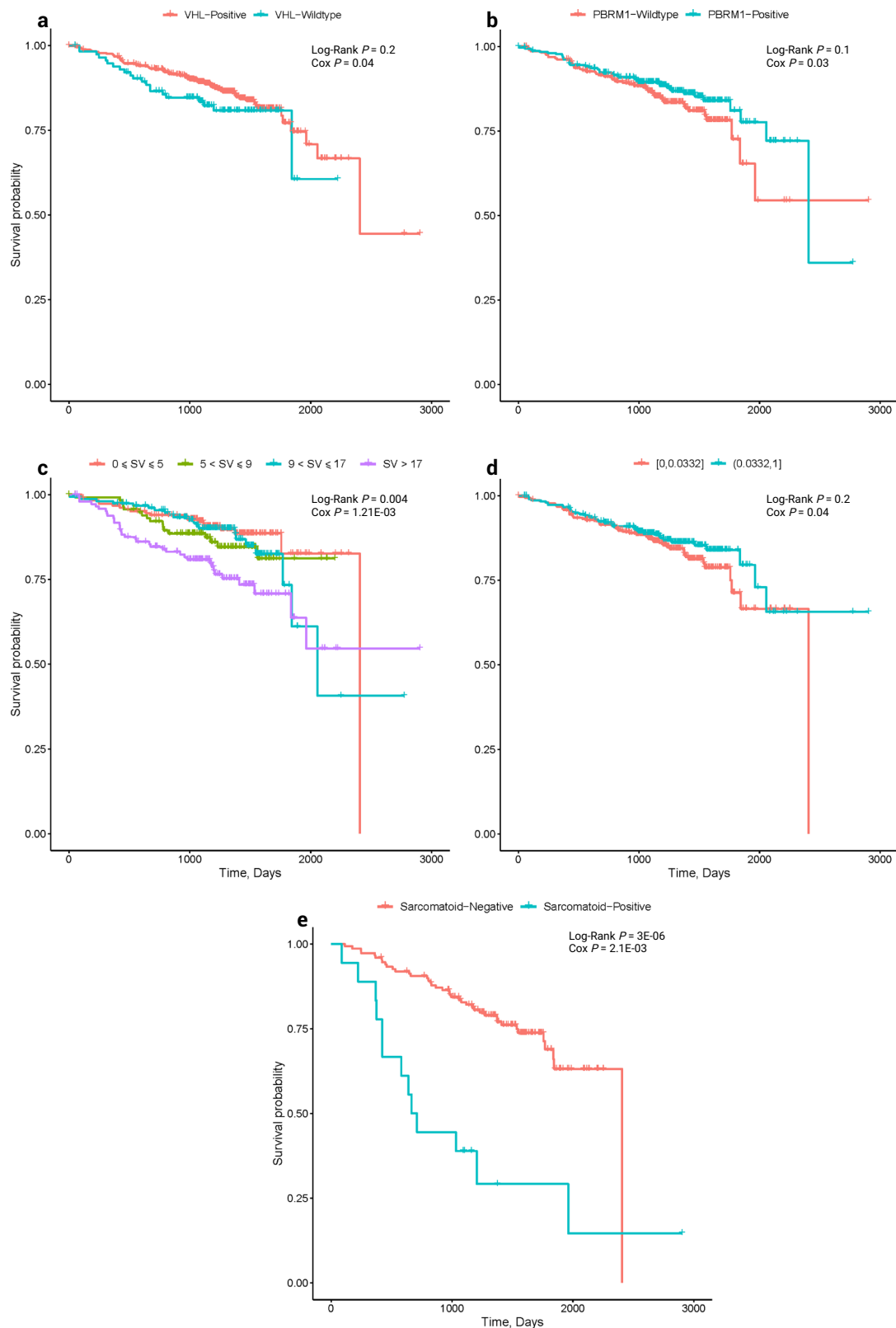


Fig. 8 | Kaplan–Meier survival curves of overall survival (OS). Relationship between OS and **a** *VHL* mutation status ($n = 605$); **b** PBRM1 mutation status ($n = 605$); **c** structural variant count ($n = 605$); **d** TCRA T-cell fraction ($n = 605$);

e sarcomatoid ($n = 167$). Log-Rank P and Cox P refers to the Log Rank test and Cox Regression (two-sided z-test), respectively.

progression. Hence it is unclear why *VHL*-positive ccRCC tumours might have a more favourable outcome than *VHL*-wildtype ccRCC. Distinct evolutionary subtypes of ccRCC have, however, been proposed that appear biologically and clinically distinct, with subtypes

defined being by *VHL*-wildtype, *VHL*-monodrivens, and those with multiple clonal drivers⁶. After adjusting for *VHL* status, we did not detect a statistically significant independent association between other driver mutations and survival. Amongst the strongest

relationships we identified was between increased copy number with increased survival, which was independent of tumour grade, presumably reflecting tumour heterogeneity. We did not find support for the purported relationship between intra-tumour heterogeneity and prognosis²⁵, however, our analysis did not benefit from multi-region sampling.

Although current drug treatment paradigms for ccRCC exploit targeted therapies they are primarily not directed against any specific genomic feature. To investigate the prospect of targeting specific driver mutations we queried OncoKB⁴³, which is regularly curated by an expert panel and therefore generally considered to reflect the current state of knowledge. Since other investigators have reported a higher targetable variant detection rate by applying multiple tools to annotate variants we also made use of The COSMIC Mutation Actionability in Precision Oncology resource⁴⁴. The majority of the alterations we describe as being actionable are based on clinical evidence from other cancers or biological plausibility. As per previous reports, the majority of the targetable alterations we identified are within PI3K/mTOR pathway genes. Randomised clinical trials showing clinical benefit of the mTOR inhibitors temsirolimus and everolimus in RCC have already led to their regulatory agency approval. Other targets have not been specifically studied in the context of ccRCC, hence results cannot be interpreted as definitive proof of response prediction. Examples of drugs where the population of interest could be expanded to treat ccRCC include: Temsirolimus, which is undergoing ongoing trials as a treatment for *FBXW7*-positive solid tumours^{78,79}, nilotinib for *ABL1* mutations⁸⁰, niraparib for *BAP1* mutations⁸¹, Tazemetostat hydrobromide for *SMARCA4* mutated cancers^{82,83}, olaparib with pembrolizumab for *ARID2*-positive melanoma⁸⁴, and alpelisib for *PIK3CA* in ER-positive metastatic breast cancer^{85,86}. An important caveat to our analysis is that the genetic profiles we derived are of a single region, which has potentially limited our ability to detect clinically important subclonal targetable alterations.

In many other cancers, a high mutational and neoantigen burden have been linked to better overall survival and responsiveness to checkpoint inhibitors presumably reflecting native immune responsiveness⁸⁷. In our study, there was no association between neoantigen burden and OS. In contrast, there was a strong relationship between increased T-cell infiltration and better prognosis. While this might seem counterintuitive, however, this finding may be explained by the poor accuracy (6%) of current HLA-affinity-based neoantigen prediction algorithms⁸⁸. Accepting these limitations, the twin observations of higher T-cell infiltration being associated with better outcome and genetically predicted immune evasion is uncommon and supports the rationale for immunotherapy.

There is interest in the prospect of population screening for RCC, given the rising incidence of the disease, the high proportion of asymptomatic individuals at diagnosis and associated high mortality rate. Our analysis supports previous work suggesting that ccRCC driver mutations often precede diagnosis by many years, if not decades³⁶, information relevant to the design of any screening programme.

Although some cancers have reaped demonstrable benefits from the current genomic revolution, the same benefits have not been yet observed in RCC, and further efforts should be directed to identify the precise role of genomic tumour profiling in the clinical setting.

Methods

Ethics statement

This study was conducted as part of the 100,000 Genomes Projects and approved by the East of England – Cambridge South Research Ethics Committee (REC reference: 14/EE/1112). All patients provided written informed consent and the study was conducted in accordance with the Declaration of Helsinki (see <https://www.genomicsengland.co.uk/initiatives/100000-genomes-project/documentation> for further information on patient consent and withdrawal). Sex-stratified analysis

was not considered for this investigation, and there are no results that are only applicable to a single sex.

The 100kGP cohort

The analysed cohort comprised tumour-normal sample pairs from patients with primary RCC recruited to 100kGP (v14 release) through 13 Genomic Medicine Centres across England^{10,11} (Fig. 1). The renal tumour cases included in this study were all routine surgical cases reported by diagnostic histopathologists at contributing centres. Histology of RCC was as per WHO Classification of Urinary and Male Genital tumours 3rd/4th edition⁸⁹, which have matching clear cell renal cell carcinoma (ccRCC) diagnostic criteria. The collection of tissue and the preparation, extraction and quantification of DNA was undertaken locally, followed by transfer of DNA to a central biorepository. We restricted our analysis to high-quality sequencing data derived from PCR-free, flash-frozen primary tumour samples from 10,470 adults (Supplementary Data 23). Illumina conducted whole genome sequencing of paired tumour/normal DNA. Processed BAM files were sent to Genomics England, who performed additional quality checks and managed data storage. We corrected for reference bias in calling of variants using FixVAF¹². We selected one tumour/normal pairing per patient. We used Strelka to call somatic variants⁹⁰, a four-stage pipeline incorporating Battenberg¹³ for copy number calling (Supplementary Data 25) and a consensus approach based on Delly¹⁶, Lumpy¹⁵ and Manta¹⁴ for calling somatic structural variants (SVs). To assess the clinical relevance of individual mutations, in addition to Ensembl Variant Effect Prediction⁹¹, all unique missense SNVs were annotated using AlphaMissense²⁴ to compare pathogenicity between driver and non-driver mutations. Comprehensive clinico-pathology information on the patients is provided in Supplementary Data 1. Complete details on sample curation, tumour purity estimation, WGS, somatic variant calling, mutation annotation, copy number alteration calling/annotation, somatic structural variant calling/annotation, whole genome duplication annotation and tumour/germline telomere length estimation provided in “Supplementary Methods”.

Identification of drivers and driver annotation

We used IntOGen to identify coding drivers¹⁸ (Supplementary Methods). The relative evolutionary timings of candidate driver mutations were obtained using MutationTimeR (Supplementary Methods). We considered gene perturbation screening data from DepMap to determine the functional basis of candidate driver genes^{30,31}. To complement this approach, we also considered gene expression data from TCGA²⁸ and GTEx²⁹, accessed through GEPIA⁹². We assessed the clinical actionability of driver gene mutations by interrogating OncoKB Knowledge Base⁴³ (version 3.11) and the COSMIC Mutation Actionability in Precision Oncology database⁴⁴. We searched for non-coding drivers within core promoters, distal promoters, 5' and 3' UTRs of canonical protein-coding transcripts, non-canonical splice regions, and lincRNAs using OncoDriveFML³³, ActiveDriverWGS³⁴ and Negative binomial regression modelling³⁵. We used Empirical Brown's method to combine *P*-values from methods and adjusted for multiple-testing using the Benjamini-Hochberg procedure (Supplementary Methods).

Cellular pathways containing driver genes (Supplementary Data 4), identified from both the IntoGen and non-coding pipelines involved in tumourigenesis, were referenced to the literature using PubMed. Pathways were also interrogated using ActivePathways⁹³ and MSigDB^{94,95} (v7.5.1).

Copy number and structural variant hotspots

Recurrent arm-level copy number events, focal amplifications and deletions, were identified using Genomic Identification of Significant Targets in Cancer⁴⁵ (GISTIC, v2.0.2.3). SVs were classified into simple and complex SVs using ClusterSV⁹⁶ (Supplementary Methods) and

structural variation hotspots for deletion, tandem duplication and simple unclassified SVs were identified using a permutation-based approach as per as Glodzik et al.⁹⁷. Additional filtering was applied to identify if the hotspot regions were predicted fragile regions in the genome (Supplementary Methods).

Mutational signatures

De novo extraction of single-base-substitution (SBS), doublet-base-substitution (DBS) and insertion and deletion (ID) signatures, including decomposition to known COSMIC signatures⁴⁴ (v3.2), was performed using SigProfilerExtractor⁵¹. We complemented this using mSINGS⁵⁴ and HRDetect^{55,98} to specifically identify mismatch repair deficient and homologous recombination deficient tumours respectively (Supplementary Methods).

Immune escape

All genomes were HLA-typed using POLYmorphic loci reSOLVER⁹⁹ (POLYSOLVER). Neoantigens were predicted using personalised Variant Antigens by Cancer Sequencing⁶⁰ (pVAC-Seq). We classed tumours as exhibiting immune escape on the basis of a non-synonymous mutation, or loss-of-heterozygosity in any one of three HLA Class-I genes, or an inactivating mutation in an antigen presenting gene (Supplementary Methods).

Clinical correlations and survival analysis

Correlations between clinical and mutational properties were identified using logistic, linear and negative binomial regression. Both univariate and multivariate regression adjusting for sex, age of sampling, stage and grade was considered (Supplementary Methods). Overall survival and cancer-specific survival was defined as the time from the date of sampling to death from any cause. Kaplan–Meier survival curves were generated and the homogeneity between groups was evaluated with the log-rank test. Progression-free survival was defined as the time from the date of sampling to radiological progression. Cox regression analysis was used to estimate hazard ratios and respective 95% confidence intervals, and adjustment for clinical variables was performed by multivariable analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data supporting the findings of this study have been deposited in the National Genomic Research Library and can be accessed via the Genomics England Research Environment, a secure cloud workspace. The raw data, including patient profiles and corresponding genomic sequencing data, are only available under restricted access for patient privacy reasons. Access can be obtained by first applying to become a member of either the Genomics England Research Network (<https://www.genomicsengland.co.uk/research/academic>) or the Discovery Forum (industry partners <https://www.genomicsengland.co.uk/research/research-environment>). The process for joining the network is described at <https://www.genomicsengland.co.uk/research/academic/join-gecip> and consists of the following steps: (1) Your institution will need to sign a participation agreement available at <https://files.genomicsengland.co.uk/documents/Genomics-England-GeCIP-Participation-Agreement-v2.0.pdf> and email the signed version to gecip-help@genomicsengland.co.uk. (2) Once you have confirmed your institution is registered and have found a domain of interest, you can apply through the online form at <https://www.genomicsengland.co.uk/research/academic/join-gecip> where you can specify the reason for access and expected timeframe that you wish to have access. Once your Research Portal account is created you will be able to login and track your application. (3) Your application will be reviewed within 10

working days. (4) Your institution will validate your affiliation. (5) You will need to complete the online Information Governance training and will be granted access to the Research Environment within 2 days of passing the online training. The processed clinical and genomic data applied to the investigation are available in the Research Environment within the folder `/re_gecip/shared_allGeCIPs/rculliford/ccRCC_landscape`. At present, there is no proposed end date for data access within the research environment. All other public/private datasets used in the study, including corresponding download links and version numbers, can be found in Supplementary Data 24.

Code availability

After completion of the instructions given by the Data availability statement, code to allow for reproducibility of results and figures are available in the research environment within the folder `/re_gecip/shared_allGeCIPs/rculliford/ccRCC_landscape`. Sources of each software package and externally downloaded data are shown in Supplementary Data 24.

References

1. Bukavina, L. et al. Epidemiology of renal cell carcinoma: 2022 update. *Eur. Urol.* **82**, 529–542 (2022).
2. Post nephrectomy management of localized renal cell carcinoma. From risk stratification to therapeutic evidence in an evolving clinical scenario. *Cancer Treat. Rev.* **115**, 102528 (2023).
3. Choueiri, T. K. et al. Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. *N. Engl. J. Med.* **385**, 683–694 (2021).
4. Pal, S. K. et al. Adjuvant atezolizumab versus placebo for patients with renal cell carcinoma at increased risk of recurrence following resection (IMmotion010): a multicentre, randomised, double-blind, phase 3 trial. *Lancet* **400**, 1103–1116 (2022).
5. Motzer, R. J. et al. Adjuvant nivolumab plus ipilimumab versus placebo for localised renal cell carcinoma after nephrectomy (CheckMate 914): a double-blind, randomised, phase 3 trial. *Lancet* **401**, 821–832 (2023).
6. Kapur, P. et al. Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *Lancet Oncol.* **14**, 159–167 (2013).
7. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
8. Turajlic, S. et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. *Cell* **173**, 581–594.e12 (2018).
9. Hakimi, A. A. et al. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. *Clin. Cancer Res.* **19**, 3259–3267 (2013).
10. Turnbull, C. et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).
11. Turnbull, C. Introducing whole-genome sequencing into routine cancer care: the Genomics England 100 000 Genomes Project. *Ann. Oncol.* **29**, 784–787 (2018).
12. Cornish, A. J. et al. Reference bias in the Illumina Isaac aligner. *Bioinformatics* **36**, 4671–4672 (2020).
13. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
14. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
15. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
16. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

17. Yngvadottir, B. et al. Frequency of pathogenic germline variants in cancer susceptibility genes in 1336 renal cell carcinoma cases. *Hum. Mol. Genet.* **31**, 3001–3011 (2022).
18. Martínez-Jiménez, F. et al. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).
19. Miao, D. et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science* **359**, 801–806 (2018).
20. Guo, G. et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat. Genet.* **44**, 17–19 (2011).
21. Gerlinger, M. et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
22. Liu, J. et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416.e11 (2018).
23. Sato, Y. et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
24. Cheng, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).
25. Turajlic, S. et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* **173**, 595–610.e11 (2018).
26. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
27. Hoadley, K. A. et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018).
28. Weinstein, J. N., Collisson, E. A., Mills, G. B. & Shaw, K. R. The cancer genome atlas pan-cancer analysis project. *Nature* **45**, 1113–1120 (2013).
29. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R. & Lo, E. The genotype-tissue expression (GTEx) project. *Nature* **45**, 580–585 (2013).
30. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576.e16 (2017).
31. Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516 (2019).
32. Andreou, A. et al. Elongin C (ELOC/TCEB1)-associated von Hippel-Lindau disease. *Hum. Mol. Genet.* **31**, 2728–2737 (2022).
33. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* **17**, 128 (2016).
34. Zhu, H. et al. Candidate cancer driver mutations in distal regulatory elements and long-range chromatin interaction networks. *Mol. Cell* **77**, 1307–1321.e10 (2020).
35. Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111 (2020).
36. Mitchell, T. J. et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx renal. *Cell* **173**, 611–623.e17 (2018).
37. Vandekerckhove, G. et al. Circulating tumor DNA reveals clinically actionable somatic genome of metastatic bladder cancer. *Clin. Cancer Res.* **23**, 6487–6497 (2017).
38. Allory, Y. et al. Telomerase reverse transcriptase promoter mutations in bladder cancer: high frequency across stages, detection in urine, and lack of association with outcome. *Eur. Urol.* **65**, 360–366 (2014).
39. Zvereva, M. et al. Activating telomerase promoter mutations and their application for the detection of bladder cancer. *Int. J. Mol. Sci.* **21**, 6034 (2020).
40. Kircher, M. et al. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat. Commun.* **10**, 3583 (2019).
41. Lee, J. H. et al. Branched-chain amino acids sustain pancreatic cancer growth by regulating lipid metabolism. *Exp. Mol. Med.* **51**, 1–11 (2019).
42. Cai, Z. et al. BCAT2 shapes a noninflamed tumor microenvironment and induces resistance to Anti-PD-1/PD-L1 immunotherapy by negatively regulating proinflammatory chemokines and anticancer immunity. *Adv. Sci.* **10**, e2207155 (2023).
43. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* **1**, 1–16 (2017).
44. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
45. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
46. Masayuki, T. & Yamamoto, M. The KEAP1-NRF2 system in cancer. *Front. Oncol.* **7**, 85 (2017).
47. Lister, A. et al. Nrf2 is overexpressed in pancreatic cancer: implications for cell proliferation and therapy. *Mol. Cancer* **10**, 37 (2011).
48. Jiménez-Vacas, J. M. et al. Spliceosome component SF3B1 as novel prognostic biomarker and therapeutic target for prostate cancer. *Transl. Res.* **212**, 89–103 (2019).
49. López-Cánovas, J. L. et al. Splicing factor SF3B1 is overexpressed and implicated in the aggressiveness and survival of hepatocellular carcinoma. *Cancer Lett.* **496**, 72–83 (2021).
50. Watkins, T. B. K. et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126–132 (2020).
51. Islam, S. M. A. et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genom.* **2**, None (2022).
52. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
53. Turesky, R. J. et al. Aristolochic acid exposure in Romania and implications for renal cell carcinoma. *Br. J. Cancer* **114**, 76–80 (2016).
54. Salipante, S. J., Scroggins, S. M., Hampel, H. L., Turner, E. H. & Pritchard, C. C. Microsatellite instability detection by next generation sequencing. *Clin. Chem.* **60**, 1192–1199 (2014).
55. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **23**, 517–525 (2017).
56. Leshchiner, I. et al. Inferring early genetic progression in cancers with unobtainable premalignant disease. *Nat Cancer* **4**, 550–563 (2023).
57. Gerstung, M. et al. The evolutionary history of 2658 cancers. *Nature* **578**, 122–128 (2020).
58. Bentham, R. et al. Using DNA sequencing data to quantify T cell fraction and therapy response. *Nature* **597**, 555–560 (2021).
59. Friedhoff, J. et al. BAP1 and PTEN mutations shape the immunological landscape of clear cell renal cell carcinoma and reveal the intertumoral heterogeneity of T cell suppression: a proof-of-concept study. *Cancer Immunol. Immunother.* **72**, 1603–1618 (2023).
60. Hundal, J. et al. pVAC-Seq: a genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **8**, 11 (2016).
61. Golkaram, M. et al. Spatiotemporal evolution of the clear cell renal cell carcinoma microenvironment links intra-tumoral heterogeneity to immune escape. *Genome Med.* **14**, 143 (2022).
62. Martínez-Jiménez, F. et al. Genetic immune escape landscape in primary and metastatic cancer. *Nat. Genet.* **55**, 820–831 (2023).
63. Kelly, A. & Trowsdale, J. Genetics of antigen processing and presentation. *Immunogenetics* **71**, 161–170 (2019).

64. Gu, Y.-F. et al. Modeling renal cell carcinoma in mice: Bap1 and Pbrm1 inactivation drive tumor grade. *Cancer Discov.* **7**, 900–917 (2017).
65. Chatsirisupachai, K., Lagger, C. & de Magalhães, J. P. Age-associated differences in the cancer molecular landscape. *Trends Cancer Res.* **8**, 962–971 (2022).
66. Morris, L. G. T. et al. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget* **7**, 10051–10063 (2016).
67. Leibovich, B. C. et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer* **97**, 1663–1671 (2003).
68. Tannir, N. M. et al. Efficacy and safety of nivolumab plus ipilimumab versus sunitinib in first-line treatment of patients with advanced sarcomatoid renal cell carcinoma. *Clin. Cancer Res.* **27**, 78–86 (2021).
69. Liu, X. et al. Dose-response relationships between cigarette smoking and kidney cancer: a systematic review and meta-analysis. *Crit. Rev. Oncol. Hematol.* **142**, 86–93 (2019).
70. Ravaud, A. et al. Update on the medical treatment of metastatic renal cell carcinoma. *Eur. Urol.* **54**, 315–325 (2008).
71. Manley, B. J. et al. Integration of recurrent somatic mutations with clinical outcomes: a pooled analysis of 1049 patients with clear cell renal cell carcinoma. *Eur. Urol. Focus* **3**, 421–427 (2017).
72. Hakimi, A. A., Pham, C. G. & Hsieh, J. J. A clear picture of renal cell carcinoma. *Nat. Genet.* **45**, 849–850 (2013).
73. Smits, K. M. et al. Genetic and epigenetic alterations in the von Hippel-lindau gene: the influence on renal cancer prognosis. *Clin. Cancer Res.* **14**, 782–787 (2008).
74. Patard, J.-J. et al. Absence of VHL gene alteration and high VEGF expression are associated with tumour aggressiveness and poor survival of renal-cell carcinoma. *Br. J. Cancer* **101**, 1417–1424 (2009).
75. Hakimi, A. A. et al. Impact of recurrent copy number alterations and cancer gene mutations on the predictive accuracy of prognostic models in clear cell renal cell carcinoma. *J. Urol.* **192**, 24–29 (2014).
76. Patard, J.-J. et al. Low CAIX expression and absence of VHL gene mutation are associated with tumor aggressiveness and poor survival of clear cell renal cell carcinoma. *Int. J. Cancer* **123**, 395–400 (2008).
77. Peña-Llopis, S. et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759 (2012).
78. Skamene, T. et al. Canadian profiling and targeted agent utilization trial (CAPTUR/PM.1): a phase II basket precision medicine trial. *J. Clin. Orthod.* **36**, TPS12127–TPS12127 (2018).
79. Dancey, J. *Canadian Profiling and Targeted Agent Utilization Trial (CAPTUR) (CAPTUR)*. ClinicalTrials.gov. <https://clinicaltrials.gov/study/NCT03297606> (2023).
80. Blay, J.-Y. & Tredan, O. *Adapting Treatment to the Tumor Molecular Alterations for Patients With Advanced Solid Tumors: MyOwnSpecificTreatment (MOST plus)*. ClinicalTrials.gov. <https://clinicaltrials.gov/study/NCT02029001> (2022).
81. George, T. *A Trial of Niraparib in BAP1 and Other DNA Damage Response (DDR) Deficient Neoplasms (UF-STO-ETI-001)*. ClinicalTrials.gov. <https://clinicaltrials.gov/study/NCT03207347> (2023).
82. Parsons, D. W. et al. Actionable tumor alterations and treatment protocol enrollment of pediatric and young adult patients with refractory cancers in the National Cancer Institute-Children’s Oncology Group Pediatric MATCH trial. *J. Clin. Oncol.* **40**, 2224–2234 (2022).
83. Parsons D. W. *Targeted Therapy Directed by Genetic Testing in Treating Pediatric Patients With Relapsed or Refractory Advanced Solid Tumors, Non-Hodgkin Lymphomas, or Histiocytic Disorders (The Pediatric MATCH Screening Trial)*. ClinicalTrials.gov. <https://clinicaltrials.gov/study/NCT03155620> (2023).
84. Kim, K. B. *Phase II Study of Olaparib and Pembrolizumab in Advanced Melanoma With Homologous Recombination (HR) Mutation*. ClinicalTrials.gov. <https://clinicaltrials.gov/study/NCT04633902> (2022).
85. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 1004 (2017).
86. Solit, D. & Stadler, Z. *Genomic Profiling in Cancer Patients*. ClinicalTrials.gov. <https://clinicaltrials.gov/study/NCT01775072> (2023).
87. Wang, P., Chen, Y. & Wang, C. Beyond tumor mutation burden: tumor neoantigen burden as a biomarker for immunotherapy and other types of therapy. *Front. Oncol.* **11**, 672677 (2021).
88. Wells, D. K. et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818–834.e13 (2020).
89. Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO classification of tumours of the urinary system and male genital organs—Part A: renal, penile, and testicular tumours. *Eur. Urol.* **70**, 93–105 (2016).
90. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
91. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
92. Tang, Z. et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).
93. Paczkowska, M. et al. Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.* **11**, 735 (2020).
94. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
95. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
96. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2023).
97. Glodzik, D. et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet.* **49**, 341–348 (2017).
98. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
99. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).

Acknowledgements

R.S.H. acknowledges grant support from Cancer Research UK (C1298/A8362), the Wellcome Trust (214388) and the Medical Research Council. S.T. is funded by Cancer Research UK (A29911); the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC10988), the UK Medical Research Council (FC10988), and the Wellcome Trust (FC10988); the National Institute for Health Research (NIHR) Biomedical Research Centre at the Royal Marsden Hospital and Institute of Cancer Research (grant reference number A109), the Royal Marsden Cancer Charity, The Rosetrees Trust (reference A2204), Ventana Medical Systems Inc (references 10467 and 10530), the National Institute of Health (U01 CA247439) and Melanoma Research Alliance (reference 686061), and the US Department of Defense (award W81XWH-22-1-0652) and VHL Alliance. This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for

Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Author contributions

A.J.C., D.C., Z.T., H.P., J.L. and K.L. processed clinical data. L.B. and D.C. reviewed histology; A.J.C. performed sequencing data quality control. D.C. and A.J.C. performed quality control of simple somatic mutations. A.J.C. called copy number alterations. A.J.C., R.C. called structural variants and identified recurrent events. B.K. and S.E.D.L. identified recurrent copy number alterations. R.C. and B.K. identified driver mutations. A.G., D.W., A.F. and C.M. extracted mutational signatures. S.E.D.L. carried out HLA estimation, neoantigen prediction and immune escape prediction. R.B. provided TCRA T Cell fractions. B.K., C.M. estimated telomere content. B.K. and C.M. inferred driver mutation clonality and timing. C.M., R.C. and S.E.D.L. correlated genomic and clinical features. R.C. carried out survival analysis. C.M., R.C., S.E.D.L. A.S. and B.K. performed clinical actionability analyses. R.S.H. and S.T. supervised the study. R.S.H., R.C., C.M., S.E.D.L., Z.T., H.P. and S.T. wrote the manuscript. All authors read and approved the final version.

Competing interests

S.T. has received speaking fees from Roche, AstraZeneca, Novartis and Ipsen. S.T. has the following patents filed: Indel mutations as a therapeutic target and predictive biomarker PCTGB2018/051892 and PCTGB2018/051893 and Clear Cell Renal Cell Carcinoma Biomarkers P113326GB. None of the other authors have a financial or non-financial conflict of interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-49692-1>.

Correspondence and requests for materials should be addressed to Richard S. Houlston.

Peer review information *Nature Communications* thanks Hao Zeng, Weizhou Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

The Renal Cancer Genomics England Consortium

Mehran Afshar¹⁴, Oyeyemi Akala¹⁵, Janet Brown¹⁶, Guy Faust¹⁵, Kate Fife¹⁷, Victoria Foy¹⁸, Styliani Germanou¹⁴, Megan Giles¹⁹, Charlotte Grieco²⁰, Simon Grummet²¹, Ankit Jain²², Anuradha Kanwar¹⁵, Andrew Protheroe²³, Iwan Raza²³, Ahmed Rehan¹⁶, Sarah Rudman²⁴, Joseph Santiapillai²⁵, Naveed Sarwar²⁶, Pavetha Seeva²³, Amy Strong¹⁷, Maria Toki²⁴, Maxine Tran²⁵, Rippie Tutika²¹, Tom Waddell¹⁸ & Matthew Wheeler¹⁹

¹⁴St George's University Hospitals NHS Foundation Trust, London, UK. ¹⁵University Hospitals of Leicester NHS Trust, Leicester, UK. ¹⁶Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK. ¹⁷Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. ¹⁸Christie NHS Foundation Trust, Manchester, UK. ¹⁹University Hospital Southampton NHS Foundation Trust, Southampton, UK. ²⁰The Royal Marsden NHS Foundation Trust, London, UK. ²¹University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. ²²The Royal Wolverhampton NHS Trust, Wolverhampton, UK. ²³Oxford University Hospitals NHS Foundation Trust, Oxford, UK. ²⁴Guy's and St Thomas' NHS Foundation Trust, London, UK. ²⁵Royal Free London NHS Foundation Trust, London, UK. ²⁶Imperial College Healthcare NHS Trust, London, UK.