Nat Genet. Author manuscript; available in PMC 2015 November 06.

Published in final edited form as:
Nat Genet. 2014 March ; 46(3): 225–233. doi:10.1038/ng.2891.

Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing

Marco Gerlinger[#1], **Stuart Horswell**[#2], **James Larkin**[#3], **Andrew J Rowan**[#1], **Max P Salm**[#2], **Ignacio Varela**[4], **Rosalie Fisher**[3], **Nicholas McGranahan**[1], **Nicholas Matthews**[5], **Claudio R Santos**[1], **Pierre Martinez**[1], **Benjamin Phillimore**[5], **Sharmin Begum**[5], **Adam Rabinowitz**[5], **Bradley Spencer-Dene**[6], **Sakshi Gulati**[7], **Paul A Bates**[7], **Gordon Stamp**[6], **Lisa Pickering**[3], **Martin Gore**[3], **David L Nicol**[8], **Steven Hazell**[9], **P Andrew Futreal**[10], **Aengus Stewart**[2], and **Charles Swanton**[1,11]

[1]Translational Cancer Therapeutics Laboratory, Cancer Research UK London Research Institute, London, UK

[2]Bioinformatics and Biostatistics, Cancer Research UK London Research Institute, London, UK

[3]Department of Medicine, Royal Marsden Hospital, London, UK

[4]Instituto de Biomedicina y Biotecnología de Cantabria (CSIC-UC-Sodercan), Departamento de Biología Molecular, Universidad de Cantabria, Santander, Spain

[5]Advanced Sequencing Facility, Cancer Research UK London Research Institute, London, UK

[6]Experimental Histopathology, Cancer Research UK London Research Institute, London, UK

[7]Biomolecular Modelling, Cancer Research UK London Research Institute, London, UK

[8]Department of Urology, Royal Marsden Hospital, London, UK

[9]Department of Pathology, Royal Marsden Hospital, London, UK

[10]Department of Genomic Medicine, MD Anderson Cancer Center, Houston, Texas, USA

[11]University College London Cancer Institute, University College London, London, UK

[#] These authors contributed equally to this work.

Correspondence should be addressed to C.S. (charles.swanton@cancer.org.uk).

AUTHOR CONTRIBUTIONS

M. Gerlinger, J.L. and C.S. designed the study. R.F., L.P., M. Gore, D.L.N. and J.L. provided clinical specimens. M. Gerlinger, A.J.R. and R.F. processed the samples. G.S., B.S.-D. and S. Hazell performed histopathological analyses. N. Matthews, B.P., S.B., A.J.R. and A.R. sequenced the samples. S. Horswell, I.V., N. McGranahan, M.P.S., P.M., S.G., P.A.B., A.S. and M. Gerlinger performed bioinformatics analyses. B.S.-D. processed histological samples, which were analyzed by G.S. and S. Horswell. M. Gerlinger, N. McGranahan and C.R.S. analyzed all data. M. Gerlinger, N. McGranahan, C.R.S., P.A.F., J.L. and C.S. interpreted the data. M. Gerlinger, N. McGranahan, C.R.S. and C.S. wrote the manuscript. All authors read and approved the final manuscript.

Accession codes. Gene expression microarray data have been deposited in the Gene Expression Omnibus (GEO) archive under accession GSE53000. Exome sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001000667.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Abstract

Clear cell renal carcinomas (ccRCCs) can display intratumor heterogeneity (ITH). We applied multiregion exome sequencing (M-seq) to resolve the genetic architecture and evolutionary histories of ten ccRCCs. Ultra-deep sequencing identified ITH in all cases. We found that 73–75% of identified ccRCC driver aberrations were subclonal, confounding estimates of driver mutation prevalence. ITH increased with the number of biopsies analyzed, without evidence of saturation in most tumors. Chromosome 3p loss and *VHL* aberrations were the only ubiquitous events. The proportion of C>T transitions at CpG sites increased during tumor progression. M-seq permits the temporal resolution of ccRCC evolution and refines mutational signatures occurring during tumor development.

Solid tumor therapeutic and clinical outcomes can differ substantially between individuals with similar clinical and pathological characteristics such as tumor type, grade and stage. Despite major efforts to identify prognostic and predictive molecular biomarkers with superior precision compared to clinical and pathological predictors, only a small number of molecular tests have been introduced into routine oncological practice[1]. We recently found evidence, using M-seq of two ccRCCs, of multiple genetically distinct subclones within primary tumors or in primary tumors and their metastases[2]. Our data further showed that both ccRCCs had evolved in a branched rather than linear fashion, similar to the branched evolution described in leukemia[3], glioblastoma[4] and colon[5], pancreatic[6], ovarian[7] and breast cancers[8].

ITH may lead to an underestimation of the genetic complexity of a tumor when single-biopsy procedures are used. Furthermore, subclonal driver mutations have been shown to contribute to the acquisition of drug resistance[9]. Thus, branched tumor evolution and the spatial separation of subclones might contribute to difficulties in efforts to validate oncological biomarkers. The presence of subclonal tumor drivers also has implications for drug development strategies; targeting of founder events, located on the trunk of a tumor's phylogenetic tree, may prove more effective than targeting of subclonal driver events[10].

To resolve ccRCC subclonal architectures, discriminate early from late driver events and decipher mutational spectra that occur during the course of ccRCC evolution, we applied M-seq to eight sporadic ccRCCs and combined these data with those for two previously reported cases[2]. We find that most driver aberrations are confined to spatially separated subclones and that the number of driver aberrations identified in individual ccRCCs increases with the number of biopsies sampled. Spatial separation of tumor subclones results in underestimation of the prevalence of driver mutations on a per-case basis if only a single biopsy is considered. Comparison of the mutation spectra of trunk versus branch mutations in phylogenetic trees of tumor subclones indicates that mutational processes change during tumor evolution.

RESULTS

Intratumor heterogeneity and spatial separation of subclones

We collected tissue from multiple tumor regions, representing the spatial extent and macroscopic heterogeneity of the primary tumors, from each of ten individuals with stage T2 (n = 2), T3 (n = 7) or T4 (n = 1) ccRCC obtained by the London Renal Cancer Consortium. All except two of these cases (RMH008 and RK26) presented with metastatic disease (Supplementary Table 1). Three cases (RMH004, RMH008 and RK26) were untreated. Six cases (EV001–EV003 and EV005–EV007) had received treatment with the mTOR inhibitor everolimus for 6 weeks and one case (RMH002) received treatment with the anti-angiogenic drug sunitinib for 16 weeks before nephrectomy. As previously described, samples from a perinephric metastasis (M1) and from a chest wall metastasis (M2a and M2b) were available for EV001, and a biopsy from a metastasis obtained at the time of disease progression on everolimus treatment was available for EV002 (M)[2]. Samples from a lymph node metastasis for EV006 (LN1a and LN1b) and a tumor thrombus from the renal vein of RMH004 (VT) were also included. All subjects provided written informed consent for the presented analyses.

EV001 and EV002 have previously been sequenced[2], and whole-exome M-seq was performed on tumor and normal DNA samples from the other eight cases with a target depth of >70× per sample (Supplementary Table 2). Overall, 679 nonsynonymous nucleotide substitutions and insertions and deletions (indels) were detected in at least 1 region and were subjected to ultra-deep amplicon sequencing (average depth of >400×) for validation, with the exception of 28 mutations for which PCR amplicons could not be generated. We successfully validated 92.5% (602/651) of mutations (for mutation details, see Supplementary Table 3), and 49 mutations that failed validation were removed from the analysis. We identified somatic mutations in the von Hippel–Lindau gene (*VHL*), which is inactivated in up to 90% of ccRCCs[11], in all tumors except EV003. Bisulfite treatment identified *VHL* DNA methylation, described previously to result in gene inactivation[12], in all tumor regions of EV003 (Supplementary Fig. 1).

We determined the regional distribution of nonsynonymous mutations on the basis of data from ultra-deep amplicon sequencing. We called a mutation as being present in a tumor region if a nucleotide substitution was detected in ≥0.5% of reads or an indel was detected in ≥1% of reads. We chose these thresholds on the basis of the error rate of the sequencing platform[13]. The regional distribution of 28 mutations for which ultra-deep sequencing data were not available was inferred from the exome sequencing data. Exome sequencing of EV001 and EV002 has previously been reported[2] and was included in this analysis. On average, 67% (range of 28–92%) of the nonsynonymous somatic mutations were heterogeneous and not detectable across all sampled regions of an individual tumor (Fig. 1). The presence of somatic mutational heterogeneity in all (10/10) treatment-naive or pretreated cases indicates that ITH, characterized by the spatial separation of subclones, is a common feature in stage T2–T4 ccRCCs.

To identify the optimal number of biopsies that can reliably detect the majority of nonsynonymous somatic mutations in a tumor, we calculated the number of mutations that

would have been detected from each tumor by sampling one to *n* biopsies (where *n* was the number of biopsies sampled from that tumor), taking into account all permutations of the available biopsies (Supplementary Fig. 2). Although mutation detection seemed to plateau off with increasing biopsy number in some cases (EV001, EV003 and EV007), a persistent increase in the number of detected mutations was apparent with each additional biopsy in the other seven cases. Thus, a further increase in the number of sampled tumor regions might result in the identification of additional subclones. Therefore, our M-seq approach probably still underestimates the number of mutations present in many ccRCCs.

The two tumors with the lowest extent of mutational heterogeneity (EV003 and EV006) had been pretreated with everolimus for 6 weeks before nephrectomy. One possible explanation for this finding might be the elimination of subclones with high sensitivity to drug. However, everolimus is a cytostatic agent[14] that is unlikely to lead to substantial depletion of cancer cells. In addition, the largest diameters of these tumors only showed minor shrinkage over the treatment period (EV003, −1%; EV006, −8%), similar to decreases in diameter observed in the more heterogeneous cases with everolimus pretreatment (Supplementary Table 4). Thus, everolimus therapy is unlikely to have led to a dramatic depletion of subclones and a reduction in heterogeneity specifically in EV003 and EV006. No other clinical or pathological characteristic seemed to correlate with mutational ITH, and larger series will be required to determine the biological basis for the diversity in ccRCC phylogenetic structures.

Identification of intraregional subclones

Variant allele frequencies (VAFs) for aberrations from all eight cases that had been validated by ultra-deep sequencing were analyzed to investigate whether multiple subclones could be found in individual regions. Mutations at positions with a sequencing depth of less than 100× in one or more tumor regions were excluded from the analysis to ensure accurate VAFs. Regional VAFs for the remaining 587 mutations (Supplementary Table 5) are shown (Fig. 2).

Six of 62 regions showed clear evidence of multiple subclones (EV005 R6, EV007 R3 and R9, RMH008 R4 and R6, and RK26 R5). These subclones were identifiable by clusters of mutations with consistently low VAFs compared to other mutations in the same region (Fig. 2) and by the presence of two distinct large peaks in density plots (Supplementary Fig. 3). Mutation clusters defining these minority clones were present in a dominant clone in another tumor region in all cases. For example, the minority clone detected in EV007 R3 was present as a dominant clone in R4 and R9.

Resolving intraregional subclonal structures has implications for phylogenetic analysis. For example, 13 of 38 mutations found ubiquitously by M-seq in all regions of EV005 showed subclonal VAFs in R6, defining an intraregional minority clone. Hence, although these 13 mutations were detected in all regions, they were not located on the trunk of the phylogenetic tree. Similarly, a mutation in the tumor suppressor gene *BAP1* that was detected ubiquitously in EV007 appeared to be subclonal in R4 and R9, also indicating that it was not a truncal mutation.

Additional regional mutations demonstrated low VAFs compared to ubiquitous mutations in the same tumor region, suggesting the presence of further intraregional subclones, similar to previous reports in ccRCC[15]. However, these subclonal structures could not be further resolved because of the limited number of mutations available for analysis from exome sequencing data and the lack of integer copy number data, precluding the application of more refined methods for the reconstruction of subclonal structures[16,17].

ccRCC life histories mapped by phylogenetic analysis

We constructed phylogenetic trees for these tumors by maximum parsimony (Fig. 3). Mutations in the polyclonal regions described above were separated into mutations likely to be present in the dominant or minority subclones (Supplementary Table 6), and both classes of mutations were plotted. Similar to our previous report[2] of EV001 and EV002, all tumors showed a branched evolutionary pattern, indicating that a linear model inadequately portrays ccRCC evolution.

Identification of truncal and branched driver mutations

Next, we mapped driver mutations onto the phylogenetic trees to address whether specific driver genes were predominantly altered on trunks or branches. Nonsynonymous mutations were classified into three categories: (i) category 1 ‘high confidence driver mutations’ (amino acid substitutions identical to those found previously in cancer driver genes or disrupting mutations (frameshift, nonsense and splice-site mutations) in tumor suppressor genes as defined by large-scale sequencing efforts in ccRCC[15,18–21] or by the Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census[22]); (ii) category 2 ‘probable driver mutations’ (mutations leading to amino acid changes in ccRCC driver or COSMIC Cancer Gene Census genes that have not previously been described but affect amino acids that have been previously found to be altered based on COSMIC or one up to five amino acids away from an amino acid affected by a mutation in a ccRCC driver gene); and (iii) category 3 ‘mutations of unknown relevance’ (all other nonsynonymous somatic mutations). This approach identified 45 category 1 driver mutations and 4 category 2 driver mutations in our 10 cases (Supplementary Table 7). Methylation of the *VHL* gene in EV003 was also classified as a category 1 driver event as it has been described to result in *VHL* inactivation[12]. Mutated driver genes are highlighted in Figures 1 and 2 and were mapped onto the phylogenetic trees (Fig. 3). Most regions harboring *PBRM1* or *BAP1* mutations displayed gene expression profiles concordant with those described for each of these driver aberrations[23] (Supplementary Fig. 4), and regions harboring a *BAP1* or *TP53* mutation showed a trend toward higher grade ($P = 0.056$) as described[24] (Supplementary Fig. 5), suggesting that the identified driver mutations have functional consequences.

VHL-inactivating alterations were present on the trunks of the phylogenetic trees for all ten tumors, consistent with the role of inactivation of the *VHL* tumor suppressor gene as a critical founder event in the majority of ccRCCs[25]. Mutations of the *PBRM1* tumor suppressor gene, the second most commonly mutated gene in ccRCC[19], mapped onto the trunks of the phylogenetic trees for three of the six tumors that harbored a *PBRM1* mutation. Thus, *PBRM1* inactivation occurs early in a subgroup of tumors. All other driver mutations were located on the branches of the phylogenetic trees. In total, 36 driver mutations (73%) in

14 genes were subclonal, and only 13 somatic driver alterations (27%) in 2 genes were truncal. Thus, the majority of driver mutations detected in these ccRCCs appear to be spatially separated and subclonal. We further examined density plots (Supplementary Fig. 3) for 52 tumor regions harboring at least 1 driver mutation to determine whether driver mutations showed evidence of intraregional subclonality (Supplementary Note). As expected, none of the 48 truncal driver mutations were subclonal. However, although driver mutations on branches were heterogeneous in all tumors where they had been detected, only 16 of 56 (29%) of these driver mutations also appeared subclonal within individual regions, and the remaining 40 (71%) appeared clonally dominant in individual regions. These findings show that the detection of a clonally dominant driver in one tumor region does not predict clonal dominance in the entire tumor.

Parallel evolution of subclones

We previously demonstrated[2] parallel evolution involving multiple independent mutations converging on the tumor suppressor genes *SETD2*, *KDM5C* and *PTEN* in EV001 and EV002. Here we identified parallel evolution in four additional tumors—EV005 harbored two distinct and spatially separated activating mutations[26] in *PIK3CA* (encoding p.Glu545Lys and p.His1047Arg), EV007 carried two spatially separated mutations in the *SETD2* histone methyltransferase (encoding p.Tyr545* and p.Thr1652Ile), and RMH008 harbored three spatially separated mutations in the histone deubiquitinase *BAP1* (encoding p.Gln277*, p.Asn411fs and p.Pro519fs). RMH004 harbored two distinct mutations in *PBRM1* (encoding p.Lys1282fs and p.Leu207fs) and one mutation each in *ARID1A* (encoding p.Arg1020Lys) and *SMARCA4* (encoding p.Thr910Met), all components of the SWI/SNF chromatin-remodeling complex. Parallel evolution events converging on distinct genes or on multiple genes encoding subunits of multiprotein complexes indicate that there is strong selection for these mutations in these particular tumors.

Identification of heterogeneous copy number driver events

We previously demonstrated ITH of somatic copy number aberrations (SCNAs) by SNP array profiling in ccRCC[2,27]. We now sought to investigate whether predominantly truncal (analogous to *VHL* mutations) or branched (analogous to *SETD2* or *BAP1* mutations) driver copy number events could be identified in SCNA profiles reconstructed from the exome sequencing data. We excluded regions with low tumor content, defined by a median VAF for truncal mutations of <15% (EV003 R6; EV006 R3 and N1b; EV007 R2, R5 and R7; RMH002 R2; RMH008 R1 and R3; and RK26 R6, R7 and R9) and regions showing flat SCNA and B-allele frequency profiles, most likely indicating technical failure (EV001 R8; EV002 R1 and R7; EV003 R5; EV007 R4; and RMH002 R1), leaving 61 regions for analysis. We identified putative driver events by overlapping SCNA profiles with recurrent regions of DNA copy number gain or loss, previously defined as likely driver aberrations in ccRCC[28] (Supplementary Fig. 6). Overall, 75% (57/76) of all driver SCNAs were heterogeneous and spatially separated (Fig. 4). Driver SCNA heterogeneity positively correlated with nonsynonymous ($r = 0.553$) and driver ($r = 0.684$) mutation heterogeneity (Supplementary Fig. 7), indicating that mutational and SCNA diversity were not mutually exclusive but often occurred together in these ccRCC cases.

The only driver SCNA found in all regions of all ten tumors was chromosome 3p loss, which evolved into copy-neutral loss of heterozygosity (LOH) in some subclones (Supplementary Fig. 8), underscoring the importance of this aberration together with *VHL* somatic mutations as founder events in ccRCC tumorigenesis. Losses on chromosomes 4q, 8p and 14q and gains of chromosome 5q were ubiquitously detected in some tumors but also occurred as heterogeneous aberrations in other cases, indicating that they are not obligatory truncal events in every tumor in which they occur. We could map 63 of 76 driver SCNAs onto the trunk or a single branch of the phylogenetic trees generated from point mutation data, indicating that the evolutionary histories identified using mutation and SCNA data were highly concordant (Supplementary Fig. 9). The regional distribution of the remaining 13 driver SCNAs could not be explained by a single event. These aberrations might have evolved independently on multiple occasions within individual tumors. Alternatively, SCNAs acquired in a single event might have been reversed through subsequent events in some subclones, or minor inaccuracies in the phylogenetic trees generated from point mutation data might account for these differences. Furthermore, ubiquitous chromosomal events might have been acquired early as truncal alterations; however, they could also be the consequence of parallel evolution of SCNAs that appeared to be similar but originated independently in distinct subclones within individual tumors. This scenario is likely for the chromosome 8p loss that was found ubiquitously in EV006 but displayed two different SCNA profiles (Supplementary Fig. 6). Exact breakpoint mapping using higher resolution techniques would be necessary to determine whether similar SCNA driver events in different regions are related and to unequivocally map them onto phylogenetic trees.

Our data suggest that driver mutations and SCNAs can be classified into (i) predominant truncal drivers, such as chromosome 3p loss and *VHL* mutations, which are likely to be necessary to initiate tumor growth in the majority of ccRCCs; (ii) facultative truncal drivers, such as *PBRM1* mutation, which in some cases may be acquired early but which can also occur on branches (such events were located on the trunk in up to 60% of ccRCCs; Fig. 5a); and (iii) subclonal drivers, representing the largest group in our data, which were always located on branches, suggesting that they may be important for adaptation during disease progression.

Driver mutation prevalence

Single-biopsy approaches may underestimate the prevalence of driver mutations in heterogeneous tumors. To investigate this possibility, we compared the prevalence of driver mutations across the 79 sequenced ccRCC samples to the prevalence of driver mutations in 102 *VHL*-deficient ccRCCs in stages T2–T4 analyzed by whole-exome sequencing of single biopsies by The Cancer Genome Atlas (TCGA)[21]. The prevalence of driver mutations in these 102 cases was strikingly similar to the prevalence we identified on a per-biopsy basis across our 79 samples (Table 1), indicating that the mutational diversity in a large cohort is similar to that of a small cohort of cases analyzed through multiple biopsies for each tumor.

However, when we determined the prevalence of driver mutations per case rather than per biopsy, the prevalence for most driver genes was higher than that based on individual samples sequenced by us or TCGA (Table 1). Most strikingly, *TP53* mutations, which were

only found in 6% of individual biopsies, were detected in 40% of cases when all tumor regions were considered for each. Similarly, mutations in the phosphoinositide 3-kinase (PI3K)-mTOR pathway were detected in 28% of individual biopsies; however, 60% of cases in our study harbored at least one mutation in this pathway. These data suggest that single-biopsy approaches underestimate the prevalence of driver mutations in individuals with ccRCC.

Mutation spectra of truncal and branch mutations

Somatic cells acquire mutations over the lifetime of an individual. Assuming that most truncal mutations occurred before transformation, phylogenetic trees allow us to explore whether the mutational spectrum changes as a result of transformation and to gain insights into the underlying mutation and DNA repair mechanisms[29]. The proportion of C>T transitions at CpG sites increased significantly in branches (q = 0.007; Fig. 5b). This increase was observed in eight of the ten cases (Fig. 5c). The proportion of C>T transitions increased in all XpCpG contexts, where X represents any nucleotide (Supplementary Fig. 10). We also observed a significant decrease in the proportion of A>G transitions (q = 0.007) among the branch mutations in each case analyzed. This finding indicates that specific mutational processes foster the generation of genetic diversity in ccRCC cells.

DISCUSSION

We have identified ITH of somatic mutations and SCNAs in all 10 ccRCCs analyzed by M-seq and demonstrated that 73–75% (36/49 driver mutations and 57/76 driver SCNAs) of genetic driver events were subclonal. ITH increased with the number of biopsies analyzed without evidence of saturation in seven of the ten tumors, suggesting that our M-seq approach still underestimates ITH. Inactivation of the *VHL* tumor suppressor gene through mutation or DNA methylation and LOH at chromosome 3p were truncal events in all cases. This is consistent with the role of these alterations as founding events in most ccRCCs. Within the boundaries of current knowledge of ccRCC biology, these two drivers may be sufficient to initiate tumor proliferation in cases that did not display any additional truncal driver aberrations. We observed truncal *PBRM1* mutations and ubiquitous SCNAs on chromosomes 4q, 5q, 8p and 14q in up to 60% of affected tumors, demonstrating that these alterations can be acquired early, potentially together with *VHL* and 3p loss, in the founding cell of the tumor. We only analyzed ten tumors, and larger series will be necessary to determine whether mutations in other driver genes such as *TCEB1*, which occur recurrently in ccRCCs with wild-type *VHL*[15], are also predominantly localized on the trunk.

All other ccRCC driver gene mutations and driver SCNAs were restricted to tumor subclones. These alterations were acquired during tumor progression and are likely drivers of the expansion of tumor clades, comprising all subclones emerging distally on the branches. Regional separation of subclones was the predominant pattern of ITH in all cases. This separation may result from spatial barriers limiting subclonal intermixing in solid tumors. The parallels with allopatric speciation, resulting from geographic separation fostering the evolution of new species, are striking. Driver mutations with relatively small selective advantage may also render selective sweeps unlikely, fostering the evolution of

spatially separated subclones. The coexistence of subclones with mutations in different driver genes in non-overlapping pathways within some tumors (for example, *BAP1*, *PBRM1* and *TP53* mutations in distinct subclones in RK26) suggests that these mutations do not define intrinsic molecular subgroups but that individual ccRCCs can traverse different evolutionary routes simultaneously.

We previously demonstrated parallel evolution in individual tumors through the identification of distinct inactivating mutations in the tumor suppressor genes *SETD2* and *KDM5C* or *PTEN*[2]. Here we identified four further tumors harboring multiple driver mutations converging on *PIK3CA*, *BAP1*, *SETD2* or *PBRM1*. Mutational convergence affecting different SWI/SNF complex members was seen in one case. Parallel evolution of *SETD2* mutations occurred in three of the ten tumors, and no *SETD2* mutations were identified in the remaining seven tumors, suggesting that different tumors may follow distinct evolutionary trajectories. It will be important to investigate whether these trajectories are determined by selection under specific microenvironmental conditions or through epistatic interactions and whether they are predictable and targetable in advance of their occurrence.

According to our small data set, truncal driver events are similar across different ccRCCs; however, clinical outcomes can differ substantially for individuals with ccRCC. Thus, it is likely that the set of subclonal drivers that is acquired by individual tumors during progression and that shows high variability between different cases influences clinical outcome. The relevance of subclonal drivers for outcome prediction is further illustrated by chromosome 9p losses and *BAP1* mutations, both markers of poor prognosis in ccRCC[24,30], which were heterogeneous in all cases where they were identified. The reliable detection of such prognostically relevant subclonal drivers is likely to be of paramount importance in improving the accuracy of prognostic and predictive biomarkers[31].

We found that 71% of driver mutations that were heterogeneous between tumor regions appeared clonally dominant within individual regions. This illusion of clonal dominance calls into question the ability of single biopsy–based VAF analysis to accurately measure heterogeneity and to identify driver alterations that are clonally dominant across an entire ccRCC tumor. This finding suggests a need for caution when attempting to define clonally dominant driver events from single biopsies for treatment stratification with targeted therapeutics.

As the majority of ccRCC driver mutations are heterogeneous, therapeutic targeting of truncal drivers such as *VHL* inactivation or chromosome 3p loss might be an effective strategy to improve clinical outcomes[10], assuming that tumors remain dependent on these drivers throughout their life histories. Both truncal driver events inactivate tumor suppressor genes that are not directly targetable. However, these alterations may be exploitable through synthetic lethal approaches, as demonstrated for *VHL*[32–34]. New treatment personalization strategies may also be necessary to improve outcomes in ccRCC. For example, we previously suggested that systemic therapy for cancer be readjusted after a partial response has been achieved through performing new biopsies of residual tumor masses and specific targeting of the molecular drivers of surviving clones[9].

The relative increase in the proportion of C>T transitions in a CpG context and the decrease in the proportion of A>G transitions in the mutation spectrum of branch versus truncal mutations indicate that specific processes may foster the generation of genetic diversity in ccRCC. C>T transitions at CpG sites have been proposed to occur through spontaneous deamination of methylated cytosines[35], in the same process that dominates the background mutation rate in mammalian genomes[36]. It will be important to investigate whether *VHL* or chromosome 3p alterations contribute to these processes and whether such interactions can be exploited for therapeutic purposes. The proportion of C>T transitions at CpG sites decreases during breast cancer progression[16], highlighting the fact that dynamics of mutational processes can vary between different cancer types.

Metastatic tissues were only available for three cases, which is a clear limitation of our study. Larger numbers of metastatic sites per case will need to be analyzed to determine whether heterogeneity between metastases mirrors the heterogeneity seen within primary tumors or whether metastases are relatively homogenous, which might foster the development of biomarkers and therapeutics for metastatic ccRCCs.

In conclusion, M-seq identified heterogeneous genomic landscapes and branched evolution for each of ten sporadic *VHL*-deficient ccRCCs with spatially separated subclones, creating an illusion of clonal dominance when assessed by single biopsies. The presence of subclonal driver events in solid tumors may provide an explanation for the inevitable acquisition of resistance to targeted therapeutics in advanced disease. The regional isolation of tumor clades in ccRCC, together with extensive evidence of parallel evolution of tumor subclones, raises important questions regarding how tumors can be optimally sampled and whether future evolutionary tumor branches might be predictable and ultimately targetable.

ONLINE METHODS

Subjects samples and consent

Samples were collected from individuals enrolled in the “Response and Resistance to Targeted Therapy in Renal Cell Carcinoma” tissue collection protocol of the London Renal Cancer Consortium (ethics approval reference 11/LO/1996) or into the E-PREDICT translational clinical trial (EUDRACT number 2009-013381-54; principal investigator J.L.). Tissue collection and trial protocols had been approved by an ethics committee, and all individuals included in this study had provided written informed consent for the analysis presented. EV001–EV007 were consecutively enrolled into the E-PREDICT trial and received 6 weeks of preoperative first-line treatment with the mTOR inhibitor everolimus before nephrectomy. Samples from EV004 were included in an ongoing study using whole-genome sequencing in ccRCC and will be published separately. The other four cases were enrolled in the renal cell carcinoma tissue collection protocol, and their tumors were chosen for M-seq analysis on the basis of ccRCC histology and primary tumor stage of T2–T4. Samples were sequenced consecutively. RMH004, RMH008 and RK26 were treatment-naive cases with a new diagnosis of ccRCC, whereas RMH002 had been treated with the anti-angiogenic drug sunitinib for 14 weeks before nephrectomy. Pathological analysis confirmed renal cell carcinoma of the clear cell type in all cases (see Supplementary Table 1 for clinical and pathological characteristics).

Tumor specimen processing

We isolated between 8 and 20 samples of 10 × 5 × 5 mm, representing the spatial extent and macroscopic heterogeneity of the primary tumor, from each nephrectomy specimen. Samples were macrodissected to minimize stromal contamination, and half of each sample was snap frozen in liquid nitrogen within 1 h of clamping of the renal artery. Sample collection was performed according to strict standard operating procedures in all cases and was documented by photography. DNA and RNA were extracted using the Qiagen AllPrep kit following the manufacturer's instructions. Nucleic acid yields were determined by Qubit (Invitrogen), and quality and integrity of DNA was examined by agarose gel electrophoresis. We excluded regions that did not contain tumor cells on histopathological examination, and we selected up to 12 regions per tumor from which sufficient high-quality DNA was available for exome-capture library preparation for exome sequencing.

M-seq multiregion exome sequencing

Exome capture was performed on 3–6 μg of genomic DNA per sample with the Agilent SureSelect Human All Exon V4 kit according to the manufacturer's instructions, and paired-end multiplex sequencing of samples on the Illumina HiSeq platform was performed at the Beijing Genome Institute (BGI) (EV003, EV005, EV006 and EV007) or on the Illumina Genome Analyzer II and HiSeq platforms at the Cancer Research UK London Research Institute (LRI) (RMH002, RMH004, RMH008 and RK26). In brief, genomic DNA sequenced at LRI was randomly fragmented by Covaris to obtain fragments mainly distributed between 250 and 300 bp in length. Adaptors were ligated to both ends of the fragments, adaptor-ligated templates were purified using Agencourt AMPure SPRI beads, and fragments with an insert size of about 250 bp were excised. Extracted DNA was amplified by ligation-mediated PCR (LM-PCR), purified and hybridized to the SureSelect Biotinylated RNA Library (BAITS) for enrichment. Hybridized fragments were bound to streptavidin beads, whereas non-hybridized fragments were washed out after 24 h. Each captured library was loaded on the Illumina platform, and paired-end sequencing was performed to the desired median sequencing depth (>70×). Raw image files were processed using HCS1.4.8 for base calling with default parameters.

M-seq bioinformatics analyses

Paired-end reads in FastQ format generated by the Illumina pipeline were aligned to the "full" build of hg19 (including unknown contigs), using the Burrows-Wheeler Aligner (BWA) 0–5.9 with a seed length of 90 bp for data sequenced at BGI and 100 bp for data from LRI. All other settings were left as default (in particular, up to 3 or 4 mismatches were allowed per read for data from BGI or LRI, respectively). Variant calling was performed using CAVEMAN as described[19]. The following filtering criteria were applied to the called variants: (i) only nucleotides with Phred quality of 20 or greater were considered; (ii) only reads mapping uniquely to the genome were considered; (iii) a minimum of 10× coverage in both germline and tumor samples was required; (iv) a minimum of two instances of the variant in the tumor region were required; and (v) variants in positions listed in dbSNP132 were removed. Somatic mutations present in at least 5% of the reads on the basis of exome sequencing in at least one tumor region were further analyzed. Filtering was performed

using R 2.15.0, the Bioconductor packages ShortRead and deepSNV, and in-house parsers written in C. Variants were annotated as synonymous or nonsynonymous using both dbNSFP and Annovar. Small indels were identified using Pindel[37] version 0.2.437 in paired tumor-normal mode. In-house filter software was used to extract high-quality indels: considering the high sequence coverage obtained in these samples, only indels with minimum coverage by 20 reads in both tumor and normal samples, a minimum frequency of 10% of the reads, a minimum of 5 independent reads supporting the event in the tumor sample and no evidence of the event in the normal sample were considered. All nonsynonymous candidate mutations were reviewed on the Integrated Genomics Viewer[38] (IGV), and variants only present in poorly aligned reads and those showing an Illumina-specific error profile[39] were removed. The known ccRCC driver genes *VHL*, *PBRM1*, *SETD2*, *BAP1*, *KDM5C*, *MTOR*, *TP53*, *PIK3CA* and *PTEN* were also reviewed manually on IGV to reduce the false negative rate.

Bisulfite sequencing of the *VHL* gene

Identification of methylation of the *VHL* gene of EV003 was carried out after bisulfite treatment of 500 ng of DNA from each tumor region and normal kidney sample using the EZ DNA Methylation kit (Zymo Research). Oligonucleotides were designed using MethPrimer to amplify a 321-bp PCR fragment spanning exon 1 and continuing into intron 1 of *VHL* from nucleotide positions chr. 3: 10,183,739–10,184,059. PCR products were sequenced using BigDye Terminator technology on an ABI 3730, and the presence or absence of guanine conversion to uracil at CpG sites in the sequence was used to determine methylation.

Mutation validation and variant frequency analysis

For all tumors except EV001 and EV002, an AmpliSeq (Lifetech) validation panel was designed by entering the genomic positions of all nonsynonymous somatic mutations called in at least one region into the Ion AmpliSeq Designer (150-bp amplicon size option selected). Multiplex PCR was performed according to the manufacturer's instructions with the tumor-specific primer pool on DNA from each region of the relevant tumor. Amplicon pools were used for the construction of barcoded sequencing libraries, and these were multiplex sequenced with 200-bp read length on the Ion Torrent PGM sequencer (Lifetech). Validation data were reviewed on IGV, and VAF (total reads at the position carrying the variant/read depth at the position) values were extracted. Twenty-eight nonsynonymous mutations failed amplification, either because the AmpliSeq Designer could not generate suitable primers or because primer pairs failed in PCR. Four driver gene mutations were affected by this (EV005, *VHL*; EV006, *VHL*; RMH002, *VHL*; and RK26, *BAP1*), but all of these were successfully validated by conventional capillary sequencing. The remaining 24 mutations were included in the analysis, as the high validation rate of >90% indicated that the majority of these were likely to be correct calls. A tumor region was called positive for a specific nucleotide substitution if the mutation was present in 0.5% or more of the reads covering the position in this region or if an indel was present in at least 1% of reads in the Ion Torrent sequencing data. Regional mutation heat maps were constructed on the basis of these regional presence/absence calls.

Variant frequencies extracted from the AmpliSeq validation data were plotted as bar charts for all mutations with a minimum coverage of 100× across all sequenced regions of a tumor using the R software package. These charts were converted into three-dimensional bar charts using Adobe Illustrator. Three-dimensional charts were examined for evidence of intraregional subclones. A mutation that displayed low VAF compared to the majority of ubiquitous mutations in the same region could indicate the presence of an intraregional subclone. However, low VAFs could also be caused by an SCNA at the position of a somatic mutation. Thus, intraregional subclones were only defined if a group of three or more mutations could be identified in which mutations all had similar VAFs that were clearly below the VAFs of the majority of ubiquitous mutations in the same region and if these mutations were spread across different chromosomes, reducing the probability that the low VAFs were merely a result of SCNAs. The minority and dominant clones that were defined through this process were separated in the binary heat maps demonstrating the regional distribution of mutations before phylogenetic tree reconstruction (Supplementary Table 6). Private mutations that were present in a region with a dominant and a minority clone were assigned to the subclone to which the VAFs of the private mutations were most similar. For those regions where the median VAFs of two intraregional subclones were similar and private mutations could not be assigned to one or the other, private mutations were equally distributed between the two subclones. The intraregional subclones in EV001 R4 have previously been described[2] and were included in Supplementary Table 6.

Individual mutations displaying low VAFs were generally not used to define intraregional subclones, as we were not able to reconstruct copy number profiles for all sequenced regions and could therefore not discriminate whether such single outliers were the result of subclonal structures or copy number changes. The single exception was the category 1 driver mutation detected in the *BAP1* tumor suppressor gene in EV007, which we identified as a subclonal mutation in R4 and R9. A SCNA event as the underlying reason for the low VAFs of this mutation in R4 and R9 appeared unlikely, as the mutation was located within the ubiquitous chromosome 3p LOH region in this tumor. Thus, if this mutation had been acquired on the trunk of the phylogenetic tree and was followed by a ubiquitous chromosome 3p LOH event that eliminated the wild-type allele, then a reduction in VAF could only be explained by a further loss affecting the chromosome 3p allele that harbors the *BAP1* mutation. In this case, no significant allelic imbalance should be detectable in EV007 R9. As allelic imbalance was clearly present in R9 (Supplementary Fig. 9), the most parsimonious explanation was that the *BAP1* mutation was acquired after LOH at chromosome 3p and defined a subclone that was present at very low frequency in R4 and R9.

Clonality analysis via kernel density plots

Observed variant frequencies were transformed by dividing all frequencies within a region by the median frequency within that region of the ubiquitous variants for that case. For these transformed frequencies, a density plot was constructed using the R function density(), with a “cosine” kernel (strictly speaking, $(1 + \cos(\pi x))/2$ for x in $[-1,1]$) and 128 interpolation points.

Mutation numbers detected by a fixed number of biopsies

For each tumor, we calculated the total number of mutations detected in each possible subset of regions of size 1, 2, …, *n*, where *n* was the total number of biopsies taken from that tumor. For example, in a tumor for which three regions were available, we examined the following subgroups: R1, R2, R3, R1 + R2, R1 + R3, R2 + R3, and R1 + R2 + R3. Then, for each fixed subset size, we calculated the mean number of mutations detected.

Phylogenetic analysis

The heat maps in Supplementary Table 6 showing binary distribution of mutations were used for phylogenetic tree reconstruction. Maximum-parsimony trees were inferred using the max-mini branch- and-bound algorithm[40], calculating branch lengths with the average pathway method, as implemented in the MEGA5 package[41]. The germline sample was designated as the outgroup, and, in the event of multiple optimal topologies, a consensus tree was constructed by collapsing branches reproduced in less than 50% of the trees. Phylogenetic trees were constructed with these algorithms and were redrawn in Adobe Illustrator with trunk and branch lengths proportional to the average number of nonsynonymous mutations that were acquired on these branches on the consensus tree. Angles between branches were chosen only for convenience of display.

Identification of driver mutations

Mutations that failed validation were removed, and all remaining genes carrying nonsynonymous somatic mutations were compared with lists of potential driver genes in ccRCC. These included the genes found to be recurrently mutated in ccRCC in previous studies[15,18–21]. Genes mutated in our ten cases were also compared to the COSMIC Cancer Gene Census (downloaded 7 August 2013), a list of genes assumed to acquire driver function though genetic alterations in a variety of human cancers. Genes that were mutated in the ccRCCs analyzed by M-seq and were also found in one of these databases were analyzed in detail. A list of all identified category 1 and category 2 driver mutations can be found in Supplementary Table 7. Two mutations in *KDM5C* identified in EV001 R1, R2, R3, R5, R8 and R9 were always detected together in close proximity on the same allele and were counted as a single driver event.

Analysis of driver mutation prevalence

We downloaded the supplementary tables containing somatic mutation and clinical data for the ccRCC cases analyzed by TCGA[21]. We identified 102 *VHL*-mutant cases with a primary tumor stage of T2–T4, similar to our 10 cases, and determined the prevalence of nonsynonymous somatic mutations in *PBRM1*, *SETD2*, *BAP1*, *KDM5C*, *TP53*, *ATM*, *ARID1A*, *PTEN*, *MTOR*, *PIK3CA* and *TSC2* in this cohort.

mRNA gene expression profiling

Samples from EV001 have previously been analyzed (GSE31610 archive in the GEO repository)[2]. RNA was extracted from all other tumor specimens and from normal kidney specimens obtained from the same individuals and quantified on an Agilent Bioanalyzer. mRNA expression profiling was performed on the Affymetrix HuGene-1_0-st-v1 platform

by an external service provider for 56 tumor regions and for 6 specimens from normal kidney from which sufficient high-quality RNA was available, and data were deposited in the GSE53000 archive in the GEO repository. Samples were normalized using the oligo R package and the RMA algorithm. A list of the genes differentially expressed in tumor samples with *BAP1* and *PBRM1* mutations was obtained from ref. 23. The expression of these genes was assessed for all regions that harbored *BAP1* or *PBRM1* mutations.

Identification of driver SCNAs

To generate raw copy number estimates, the VarScan2 (v2.2.11) algorithm was employed using default parameters except for the data-ratio parameter, which was calculated as described[42]. We excluded regions with low mappability (segmental duplications and Encyclopedia of DNA Elements (ENCODE) “DAC blacklisted” regions, as annotated by the UCSC Genome Browser). Raw $\log_2$-ratio (logR) calls were adjusted for GC content and recentered to 0 using the modal logR value determined using kernel density estimates. Outliers were detected and modified using Median Absolute Deviation Winsorization before case-specific joint segmentation ($\gamma = 300$)[43], to identify genomic segments of constant logR. Adjacent segments of similar logR were merged by an internally developed R script. The peak locations of 14 recurrent copy number changes that are likely to be driver copy number events in ccRCC were retrieved from Beroukhim *et al.*[28], and these regions were highlighted in the logR plots by vertical bars. Blue bars denote recurrently lost peaks (1p36.11, 3p25.3, 4q34.3, 6q22.33, 8p23.2, 9q21.3 and 14q31.1), and red bars denote recurrently gained peaks (1q25.1, 2q14.3, 5q35.3, 7q22.3, 8q24.21, 12p11.21 and 20p13.33). LogR plots were reviewed independently by two researchers to determine regional copy number gains and losses. Consensus was obtained for diverging results by reviewing LogR plots and B-allele frequency plots. Some but not all tumor regions in EV007, RMH002 and RMH004 showed copy number losses of chromosome 3p, and all other tumors showed ubiquitous chromosome 3p losses. Inspection of the B-allele plots for EV007, RMH002 and RMH004 showed abnormal B-allele frequencies of chromosome 3p in all regions, suggesting that initial LOH at 3p had evolved into copy-neutral LOH by duplication of the remaining allele in some regions.

Mutation spectrum analysis

All single nonsynonymous point mutations from all samples were combined in a table and defined as occurring on the trunks or branches of phylogenetic trees. A χ^2 test was used to examine the overall difference in the distribution of trunk and branch mutations, using the mutation types listed. Fisher’s exact tests were implemented to assess the difference between specific mutation types. For each mutation type, the complementary mutation was also assessed; for example, XpCpG to XpTpG and the complementary CpGpX to CpApX. Each mutation type was also classified on the basis of its trinucleotide context. *P* values calculated for each of the seven mutation types were corrected for multiple testing using the Benjamini-Hochberg method.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Europe PMC Funders Author Manuscripts

ACKNOWLEDGMENTS

We thank the patients, the research nurses at the Royal Marsden Hospital, and Lifetech and Westminster Genomic Services at the University of Westminster, London, for their assistance with validation. C.S. and M. Gerlinger are supported by grants from Cancer Research UK Biomarkers and Imaging Discovery and Development Committee (BIDD), the Medical Research Council and the Seventh European Union Framework Programme, and C.S. is supported by the Breast Cancer Research Foundation and the Rosetrees Trust. We acknowledge the Ramón y Cajal program of the Ministerio de Economía y Competitividad, Spain, and Novartis for funding support for E-PREDICT clinical trials. This study was supported by researchers at the National Institute for Health Research Biomedical Research Centres at University College London Hospitals and at the Royal Marsden Hospital.

URLs

E-PREDICT translational clinical trial, http://public.ukcrn.org.uk/Search/StudyDetail.aspx?StudyID=10710; MethPrimer, http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi; Ion AmpliSeq Designer, http://www.ampliseq.com/; COSMIC Cancer Gene Census, http://cancer.sanger.ac.uk/cancergenome/projects/census/.

References

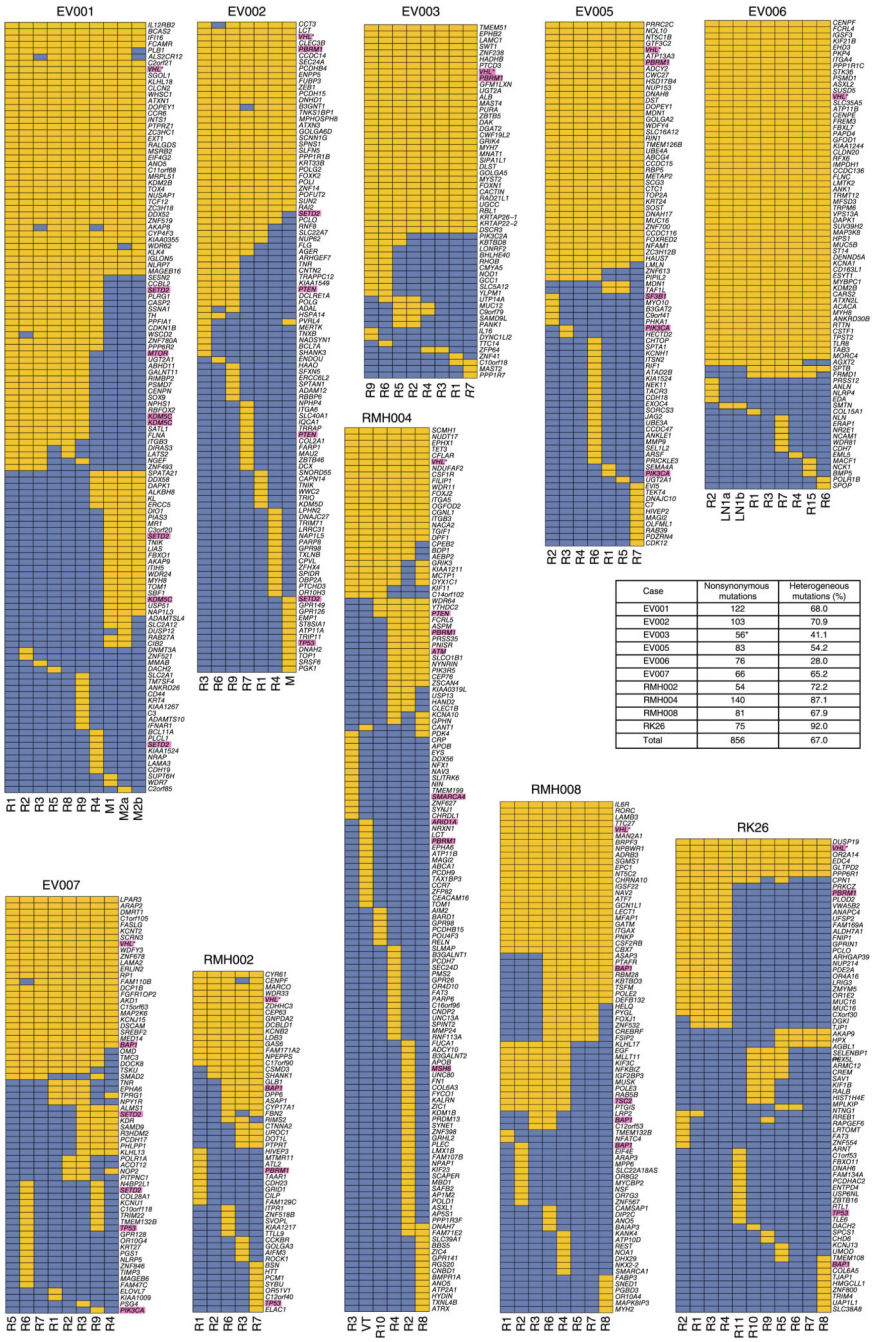
1. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. Cancer Res. 2012; 72:6097–6101. [PubMed: 23172309]
2. Gerlinger M, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. 2012; 366:883–892. [PubMed: 22397650]
3. Anderson K, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature. 2011; 469:356–361. [PubMed: 21160474]
4. Sottoriva A, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc. Natl. Acad. Sci. USA. 2013; 110:4009–4014. [PubMed: 23412337]
5. Thirlwell C, et al. Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. Gastroenterology. 2010; 138:1441–1454. [PubMed: 20102718]
6. Campbell PJ, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010; 467:1109–1113. [PubMed: 20981101]
7. Bashashati A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J. Pathol. 2013; 231:21–34. [PubMed: 23780408]
8. Navin N, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011; 472:90–94. [PubMed: 21399628]
9. Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. Br. J. Cancer. 2010; 103:1139–1143. [PubMed: 20877357]
10. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. Sci. Transl. Med. 2012; 4:127ps10.
11. Nickerson ML, et al. Improved identification of von Hippel–Lindau gene alterations in clear cell renal tumors. Clin. Cancer Res. 2008; 14:4726–4734. [PubMed: 18676741]
12. Herman JG, et al. Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma. Proc. Natl. Acad. Sci. USA. 1994; 91:9700–9704. [PubMed: 7937876]
13. Rothberg JM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011; 475:348–352. [PubMed: 21776081]
14. Houghton PJ. Everolimus. Clin. Cancer Res. 2010; 16:1368–1372. [PubMed: 20179227]
15. Sato Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat. Genet. 2013; 45:860–867. [PubMed: 23797736]
16. Nik-Zainal S, et al. The life history of 21 breast cancers. Cell. 2012; 149:994–1007. [PubMed: 22608083]

17. Shah SP, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012; 486:395–399. [PubMed: 22495314]
18. Dalgliesh GL, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010; 463:360–363. [PubMed: 20054297]
19. Varela I, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. Nature. 2011; 469:539–542. [PubMed: 21248752]
20. Guo G, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. Nat. Genet. 2012; 44:17–19. [PubMed: 22138691]
21. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013; 499:43–49. [PubMed: 23792563]
22. Forbes SA, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39:D945–D950. [PubMed: 20952405]
23. 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. 2013; 14:159–167. [PubMed: 23333114]
24. Peña-Llopis S, et al. BAP1 loss defines a new class of renal cell carcinoma. Nat. Genet. 2012; 44:751–759. [PubMed: 22683710]
25. Foster K, et al. Somatic mutations of the von Hippel–Lindau disease tumour suppressor gene in non-familial clear cell renal carcinoma. Hum. Mol. Genet. 1994; 3:2169–2173. [PubMed: 7881415]
26. Kang S, Bader AG, Vogt PK. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. Proc. Natl. Acad. Sci. USA. 2005; 102:802–807. [PubMed: 15647370]
27. Martinez P, et al. Parallel evolution of tumor subclones mimics diversity between tumors. J. Pathol. 2013; 230:356–364. [PubMed: 23716380]
28. Beroukhim R, et al. Patterns of gene expression and copy-number alterations in von-Hippel Lindau disease–associated and sporadic clear cell carcinoma of the kidney. Cancer Res. 2009; 69:4674–4681. [PubMed: 19470766]
29. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013; 3:246–259. [PubMed: 23318258]
30. La Rochelle J, et al. Chromosome 9p deletions identify an aggressive phenotype of clear cell renal cell carcinoma. Cancer. 2010; 116:4696–4702. [PubMed: 20629029]
31. Swanton C. Intratumor heterogeneity: evolution through space and time. Cancer Res. 2012; 72:4875–4882. [PubMed: 23002210]
32. Turcotte S, et al. A molecule targeting VHL-deficient renal cell carcinoma that induces autophagy. Cancer Cell. 2008; 14:90–102. [PubMed: 18598947]
33. Bommi-Reddy A, et al. Kinase requirements in human cells: III. altered kinase requirements in $VHL^{-/-}$ cancer cells detected in a pilot synthetic lethal screen. Proc. Natl. Acad. Sci. USA. 2008; 105:16484–16489. [PubMed: 18948595]
34. Chan DA, et al. Targeting GLUT1 and the Warburg effect in renal cell carcinoma by chemical synthetic lethality. Sci. Transl. Med. 2011; 3:94ra70.
35. Rubin AF, Green P. Mutation patterns in cancer genomes. Proc. Natl. Acad. Sci. USA. 2009; 106:21766–21770. [PubMed: 19995982]
36. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. Nat. Rev. Genet. 2011; 12:756–766. [PubMed: 21969038]
37. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009; 25:2865–2871. [PubMed: 19561018]
38. Robinson JT, et al. Integrative genomics viewer. Nat. Biotechnol. 2011; 29:24–26. [PubMed: 21221095]
39. Nakamura K, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 2011; 39:e90. [PubMed: 21576222]

40. Purdom PW Jr. Bradford PG, Tamura K, Kumar S. Single column discrepancy and dynamic max-mini optimizations for quickly finding the most parsimonious evolutionary trees. Bioinformatics. 2000; 16:140–151. [PubMed: 10842736]
41. Tamura K, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 2011; 28:2731–2739. [PubMed: 21546353]
42. Koboldt DC, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568–576. [PubMed: 22300766]
43. Nilsen G, et al. Copy number: efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics. 2012; 13:591. [PubMed: 23442169]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Case	Nonsynonymous mutations	Heterogeneous mutations (%)
EV001	122	68.0
EV002	103	70.9
EV003	56*	41.1
EV005	83	54.2
EV006	76	28.0
EV007	66	65.2
RMH002	54	72.2
RMH004	140	87.1
RMH008	81	67.9
RK26	75	92.0
Total	856	67.0

Figure 1.

Regional distribution of nonsynonymous mutations in ten ccRCC tumors. Mutations that failed validation were not included. Heat maps indicate the presence of a mutation (yellow) or its absence (blue) in each region. Category 1 high-confidence driver mutations and category 2 probable driver mutations are highlighted in magenta. The table shows the number of nonsynonymous mutations and the ratio of heterogeneous mutations per tumor. An asterisk indicates where *VHL* methylation was included in the analysis.

Europe PMC Funders Author Manuscripts

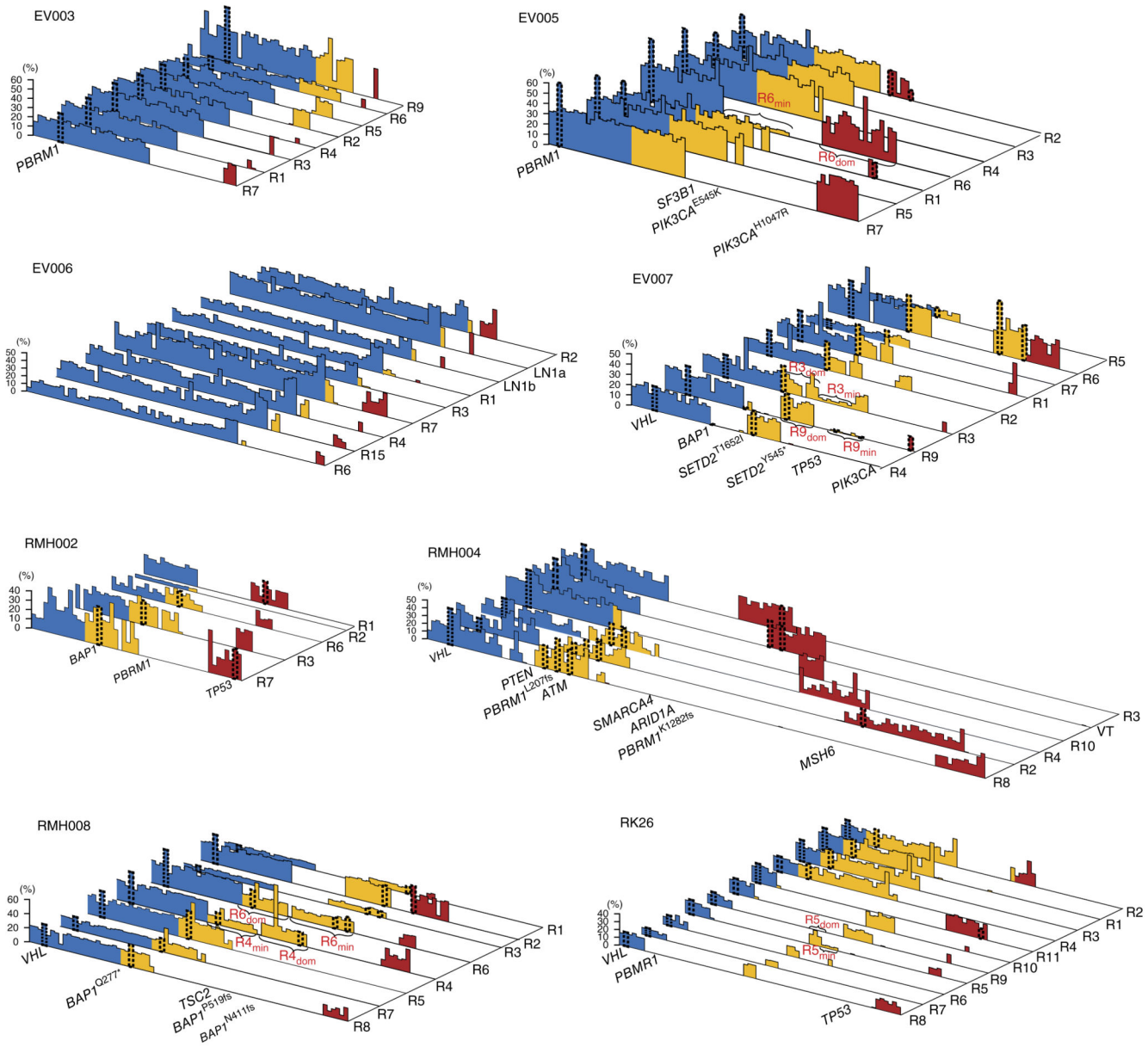


Figure 2.

Variant frequencies for the nonsynonymous somatic mutations in eight ccRCC tumors based on ultra-deep amplicon sequencing. Individual mutations are shown on the *x* axis, and variant frequencies are plotted on the *y* axis. Category 1 high-confidence driver mutations and category 2 probable driver mutations are labeled on the *x* axis and are highlighted by dashed boxes. Amplicon generation for deep sequencing failed for *VHL* mutations in EV005, EV006 and RMH002 and for a *BAP1* mutation in RK26, and thus these could not be included. Mutations that define dominant (dom) and minority (min) clones in cases EV005, EV007, RMH008 and RK26 are indicated.

Europe PMC Funders Author Manuscripts

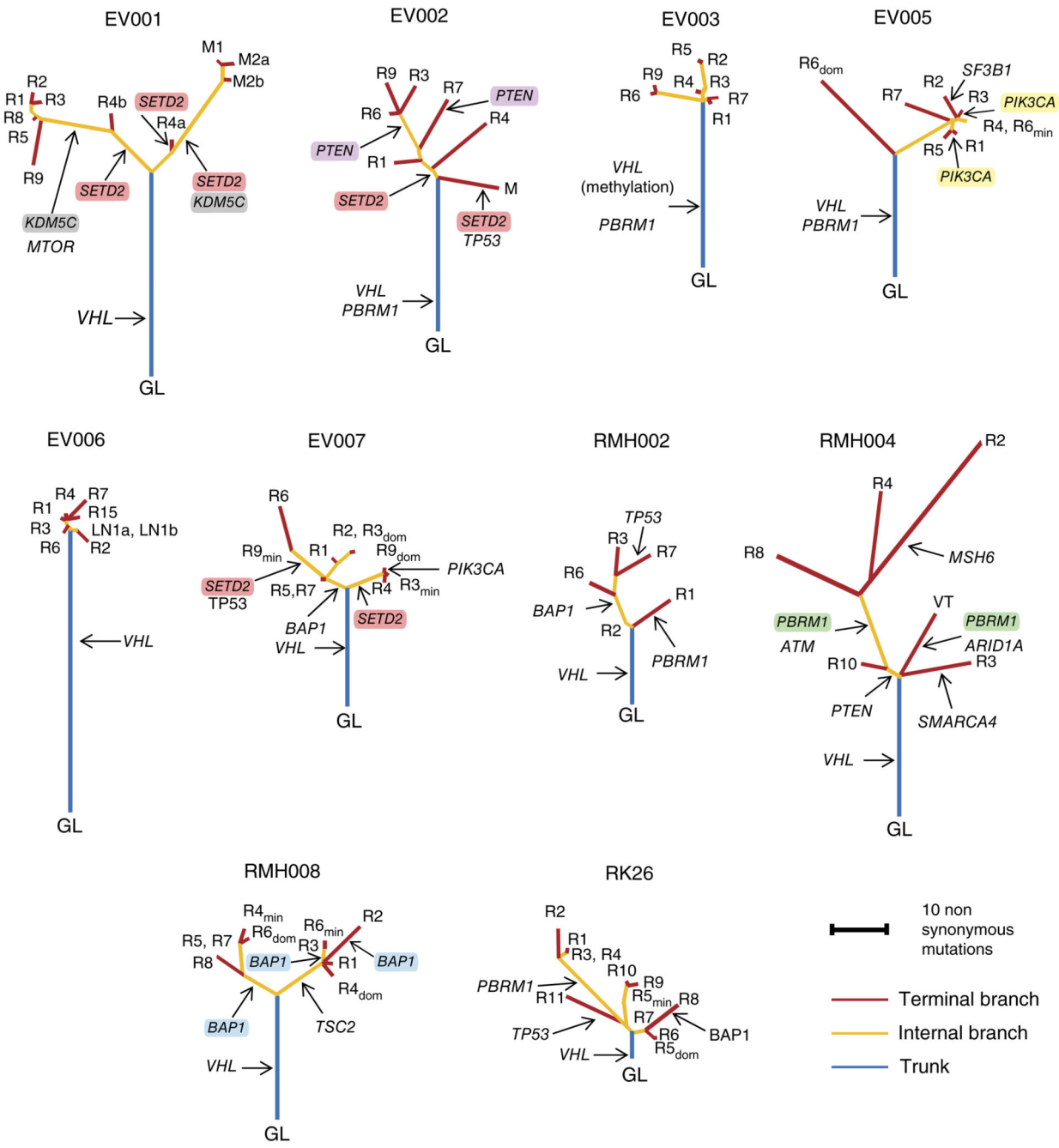
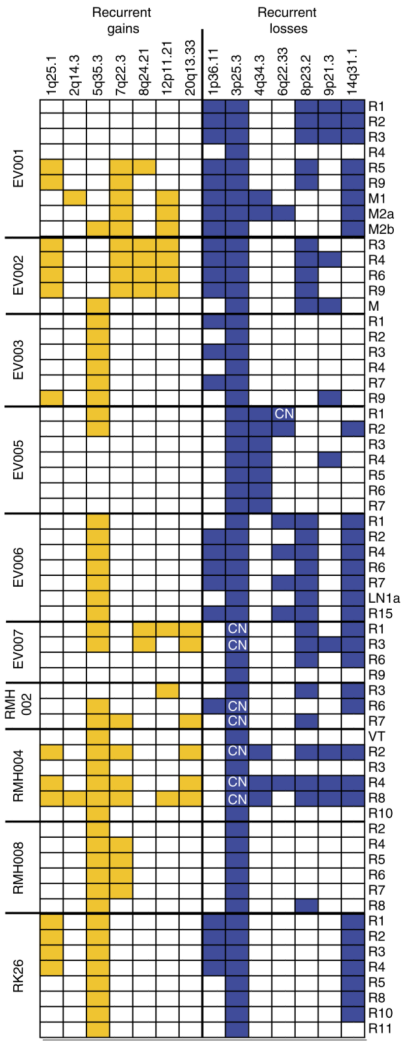


Figure 3.

Phylogenetic trees generated by maximum parsimony from M-seq data for ten ccRCC tumors. Trees for EV001 and EV002 are adapted from Gerlinger *et al.*[2]. Branch and trunk lengths are proportional to the number of nonsynonymous mutations acquired on the corresponding branch or trunk. Driver mutations were acquired by the indicated genes in the branches the arrows indicate. Driver mutations defining parallel evolution events are highlighted by color. Trees are rooted at the germline (GL) DNA sequence, determined by exome sequencing of DNA from peripheral blood.



Case	Driver CNVs	Heterogeneous driver CNVs (%)
EV001	13	12 (92.3)
EV002	9	7 (77.8)
EV003	5	3 (60.0)
EV005	6	4 (66.7)
EV006	6	2 (33.3)
EV007	8	7 (87.5)
RMH002	8	7 (87.5)
RMH004	12	10 (83.3)
RMH008	4	2 (50.0)
RK26	5	3 (60.0)
Total	76	57 (75)

Figure 4.

Regional distribution of somatic driver copy number aberrations in ten ccRCC tumors. The heat map indicates copy number gains (yellow) or losses (blue) for all tumor regions for which copy number profiles could be reconstructed from exome sequencing data. CN, regions displaying copy-neutral LOH of chromosome 3p25.3.

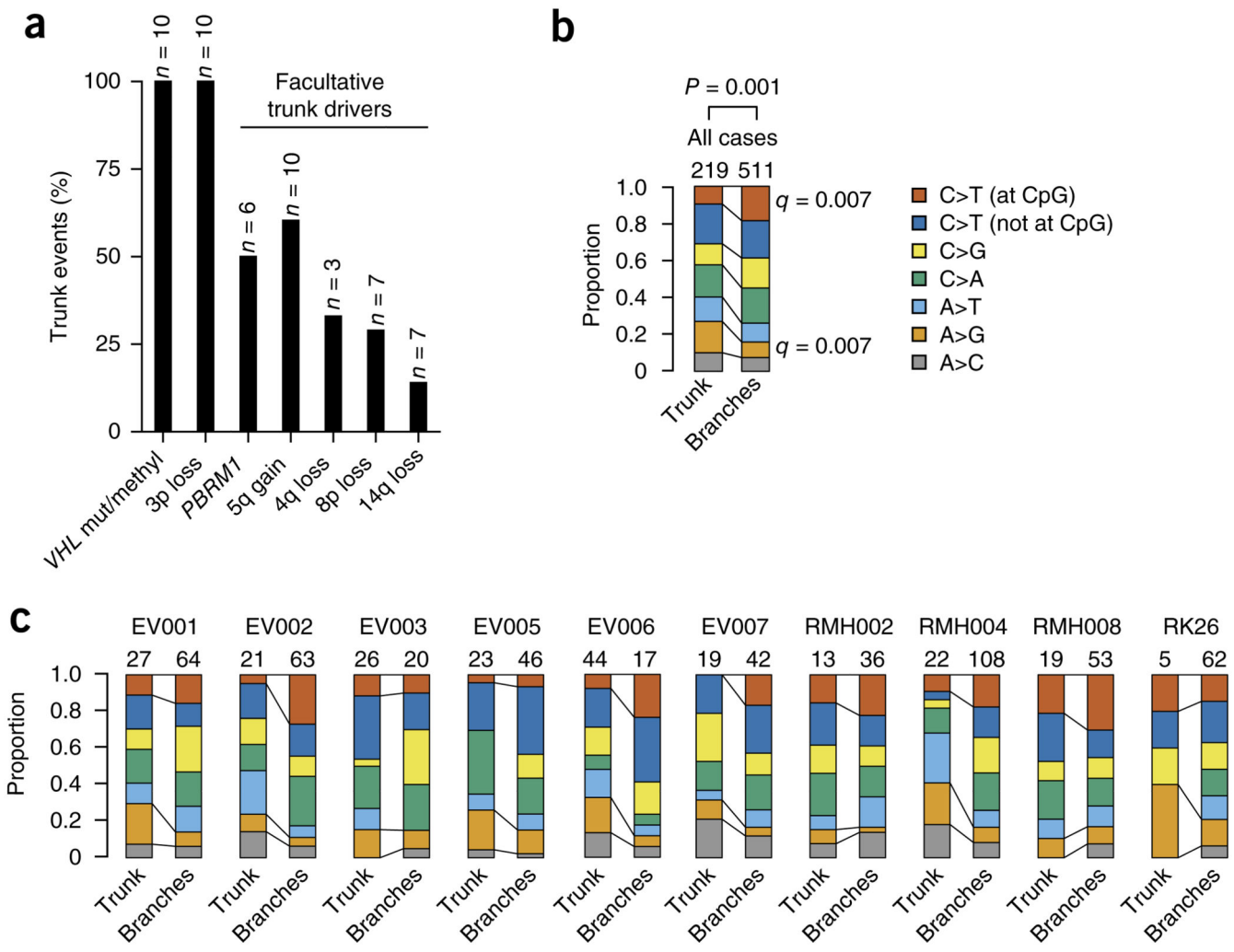


Figure 5.

Truncal location of driver aberrations and mutation spectrum in ten ccRCC tumors. (**a**) Percentage of cases in which a specific driver aberration was located on the trunk of the phylogenetic tree. *n* indicates the number of cases in which each event was found. Mut, mutation; methyl, methylation. (**b,c**) Mutation spectrum of nonsynonymous trunk versus branch mutations combined across all cases (**b**) and per case (**c**). The number of mutations analyzed is displayed on top of each bar. The difference between the spectra for trunk and branch mutations across all cases was assessed using a χ^2 test. For specific mutation types, a Fisher's exact test was used, and significant *P* values, corrected for multiple testing using the Benjamini-Hochberg method where appropriate (*q* values), are shown.

Europe PMC Funders Author Manuscripts

Table 1
Comparison of driver mutation prevalence in ccRCC samples

	Prevalence in TCGA samples (*n* = 102 samples)	Prevalence in all M-seq samples (*n* = 79 samples)	Prevalence in cases based on M-seq (*n* = 10 cases)	Prevalence cases/ prevalence M-seq samples
PBRMl	42%	39%	60%	1.5
SETD2	18%	27%	30%	1.1
BAPl	21%	23%	40%	1.7
KDM5C	7%	11%	10%	0.9
TP53	5%	6%	40%	6.7
ATM	3%	4%	10%	2.5
ARIDlA	6%	1%	10%	10.0
PTEN	5%	10%	20%	2.0
MTOR	9%	8%	10%	1.3
PIK3CA	3%	4%	20%	5.0
TSC2	2%	6%	10%	1.7
PI3K-mTOR pathway	18%	28%	60%	2.1

Comparison of the prevalence of driver mutations in a cohort of 102 *VHL*-mutant ccRCC samples with stages T2–T4 from TCGA[21] and across all 79 individual samples for the 10 cases sequenced by M-seq in this study. The last column displays the fold increase if prevalence is analyzed on the basis of individual cases instead of individual samples. The combined prevalence of the indicated PI3K-mTOR pathway genes (*PTEN*, *PIK3CA*, *TSC2* and *MTOR*) is shown in the last row.