Origins and timing of emerging lesions in advanced renal cell carcinoma

Andrew Wallace[1*], Sima P. Porten[2*], Amy A. Lo[3], Daniel Oreper[1], Nicolas Lounsbury[1], Charles Havnar[3], Ximo Pechuan-Jorge[4], Oliver A. Zill[1*], Maxwell V. Meng[2*]

[1]*Department of Oncology Bioinformatics, Genentech, Inc*
[2]*Department of Urology, University of California, San Francisco*
[3]*Department of Research Pathology, Genentech, Inc*
[4]*Department of Cancer Immunology, Genentech, Inc*
*These authors made equal contributions

Running title: Evolutionary genomics of RCC with VTT and metastasis

Keywords: multi-region sequencing, renal cell carcinoma, metastasis, phylogenetic analysis, tumor thrombus

Corresponding authors:

Oliver A. Zill

501 DNA Way, 44-2210, South San Francisco, CA 94080

zill.oliver@gmail.com

(650)302-5171

Maxwell V. Meng

1825 Fourth St Sixth Floor, San Francisco, CA 94158

max.meng@ucsf.edu

(415)353-7096

Conflict of interest statement: All authors except Dr. Meng and Dr. Porten are paid employees and/or stockholders of Roche.

Word count: 5,404

Total number of figures and tables: 6

Statement of Translational Relevance

Renal cell carcinoma (RCC) is a deadly and relatively common malignancy, which often presents as or progresses to metastatic disease. We used multi-region sequencing of RCC patients with venous tumor thrombus (VTT) and metastasis to ask how and when new lesions arise from the primary tumor, and what genomic factors contribute to their spread. Phylogenetic analysis of patients with VTT and co-presenting metastases suggested that in most cases, the VTT and metastases derive from distinct tumor clones. Moreover, metastatic clones often appear many years prior to diagnosis. We found that local TNFα inflammation may contribute to VTT formation, whereas MTOR signaling is associated with metastases. Our study sheds light on the relationship of VTT and metastases, suggests therapeutic and biomarker strategies for RCC, and points to the need for early detection studies in RCC to better understand when metastases emerge and to identify at-risk patients.

Abstract

Purpose: Renal cell carcinoma (RCC) with venous tumor thrombus (VTT) arising from the primary tumor occurs in 4-10% of cases and is associated with advanced disease. RCC with VTT and distant metastasis represents a unique clinical entity, and provides opportunities to examine the origins and relative timing of tumor lesion emergence and to identify molecular correlates with disease state.

Experimental Design: We performed genomic and evolutionary analyses on 16 RCC patients with VTT, with eight also having metastases, using multi-region exome and RNA sequencing.

Results: No genomic alterations were specifically associated with the VTT or metastasis lesions; each tumor had multiple hallmark driver alterations, consistent with advanced disease state. We found that 21% (3/14) of clear-cell RCC cases could be assigned a previously defined "evolutionary subtype". Somatic mutation signatures were largely consistent with previously established RCC signatures, and showed low heterogeneity across regions of each tumor. Mismatch repair and homologous recombination ("BRCA-ness") deficiency signatures consistently co-occurred across most tumors, suggesting a pervasive role for intracellular DNA damage in RCC and the potential for related treatment strategies. Phylogenetic timing analysis of metastatic cases suggested that in most tumors, metastases branched from the primary tumor prior to formation of VTT and in some cases before diversification of the primary tumor. Both VTT and the earliest metastases were predicted to emerge many years prior to diagnosis. Transcriptional landscape analysis identified key differences distinguishing each lesion type from primary tumor: VTT upregulated TNFα signaling and associated inflammatory pathways, whereas metastases upregulated *MTOR* signaling.

Conclusions: Our results provide a map of how RCC tumors can evolve, with metastatic clones typically emerging early in RCC development and taking hold via *MTOR* signaling, and later formation of VTT via local inflammatory processes.

Introduction

Renal cell carcinoma (RCC) is the most common type of kidney cancer, and is one of the most common neoplasms worldwide (1). RCC exhibits local vascular invasion in up to 10% of cases, and the resulting venous tumor thrombus (VTT) impacts treatment options and may present challenges for surgical intervention (2,3). Recent work has made strides towards genomic characterization of clear cell RCC (ccRCC) tumors with VTT, and has begun to delineate the genomic features of VTT tumors from non-VTT tumors (4,5). For example, ccRCC tumors with VTT exhibit greater tumor mutational burden, higher genomic instability, and higher tumor proliferative indices than tumors without VTT (4,6).

Several studies have employed multi-region sequencing (MR-seq) to explore patterns of tumor evolution in RCC for cases both with and without VTT (5–9). RCC tumors exhibit branched evolution leading to high intratumoral heterogeneity (ITH) and substantial subclonal diversity. MR-seq has also shed some light on the evolutionary patterns that give rise to VTT. Warsow et al. (5) demonstrated that while multiple VTT regions typically derive from a single common ancestor, they likely originate from subclones common to primary tumor regions. Turajlic et al. (6) found similar patterns, with the VTT seemingly derived directly from the most recent common ancestor of all tumor clones in the greatest proportion of assayed cases. However, the relationship between VTT and metastases remains unclear. Many patients with VTT present with metastasis (20-50%), but VTT formation and metastatic spread may not necessarily be linked (10,11). Broadly, it is possible that either metastases could be seeded from the VTT clone via hematogenous spread, or that metastases are derived from separate clones that may originate earlier or later than the VTT clone. As VTT physically emerge from the primary tumor, their genomic composition provides a point of comparison for the relative timing and subclonal origin of distant metastasis from the primary tumor.

The timing of metastatic spread is an issue of substantial clinical significance (12), and has been explored in other cancers such as CRC (13,14). Many patients with RCC present with metastases at the time of diagnosis (20-30%), however in some patients metastases only become evident over time (15). Knowledge of the relevant timelines is critical to properly inform screening regimens, as patients at risk for early metastasis should be subject to more regular screenings, and may impact treatment decisions. Therefore, quantitative approaches for estimating the timing of metastasis in RCC from genomic evidence are increasingly critical for improved prognosis, treatment algorithms, and ultimately clinical outcomes (16). Similarly, understanding which biological pathways contribute to the formation of new lesions and their spread from the primary tumor is important to inform therapeutic development and treatment

decisions. Toward these ends, we performed MR-seq analysis of RCC with VTT and metastases--including both mutation and gene expression characterization--to estimate the timing of their clonal emergence and to identify biological pathways associated with each type of lesion. We further hypothesized that the resulting phylogenomic patterns might shed light on the clinical pictures of these patients and inform RCC treatment more broadly.

Materials and Methods

Clinical sample identification and pathology analysis

Institutional Review Board approval was obtained from the University of California San Francisco. Cases clinically diagnosed as renal cell carcinoma with venous tumor thrombus that contained banked tissue for research were identified. All patient identifiers were reassigned to protect anonymity. A surgical pathologist reviewed all slides associated with each case, established that primary tumor, metastatic tumor, and thrombus regions were available based on the gross description and confirmed diagnoses based on morphologic and immunohistochemical findings. Viable tumor content (% viable tumor/total epithelial surface area), stromal abundance (low, moderate or high) and presence of cystic change were estimated for each case by an independent pathologist. Retrospective chart review was performed to identify and capture relevant clinical and demographic information (Figure S1). A full cohort description and patient metadata are available in Tables S1 and S2.

IHC analysis

IHC was performed on 4um thick FFPE tissue sections mounted on glass slides. IHC for PD-L1 clone SP263 (Roche Tissue Diagnostics, Tuscan, AZ, cat 790-4905) was performed on the Ventana Benchmark XT platform. The slides were pretreated with CC1 for 64 min followed by primary antibody incubated for 16 minutes at 37℃. The antibody was detected with the OptiView DAB IHC Detection Kit. PanCK and CD8 duplex chromogenic IHC was performed using established methods on the Ventana Discovery Ultra. The fraction of viable tumor cells (%) that express membrane PD-L1 were quantified. The overall immune phenotype was classified as "desert", "inflamed", "excluded" (two types: invasive margin or intratumoral) based on the predominant (>10%) location of CD8 positive cells in relation to the tumor. If no single pattern occurred in the majority of the tumor, then the slide was categorized as "heterogeneous". Automated slide assessment was performed quantitatively using Visiopharm

analysis software. Tissue area that was positive for the panCK was used to generate an epithelial tumor mask and the relative surface area of CD8 positive cells within stromal and epithelial tumor compartments was determined using Visiopharm software applications.

Generation and processing of sequencing data

Exome sequencing libraries were created with Agilent SureSelectXT. SureSelect All Exon V6 bait probes were used to perform targeted capture at Q^2 Solutions (Valencia, CA). Exome sequencing coverage was approximately 75 million 100bp paired-end reads, yielding an average depth (before removing duplicate reads) of 150X per sample. RNA-seq libraries were generated using the RNA Access platform. Sequencing coverage was approximately 50 million 50bp paired-end reads per sample.

Exome FASTQ files were trimmed to remove adapter contamination and low-quality trailing sequences, and then mapped to the GRCh38 reference genome with GSNAP (17). Mutations were called using both Lofreq2 (18) and Strelka (19), with the union of both outputs retained prior to further filtering. The filtering required mutations to have a minimum VAF of 0.05 and coverage of 20 in at least one sample in which they were called. Further, they were required to exhibit a maximum VAF of 0.01 in the normal sample, and a maximum ExAC (20) global frequency of 0.01. Somatic copy number alterations (SNCA) were identified using TitanCNA (21). RNA-seq FASTQ files were trimmed to remove adapter contamination and low-quality trailing sequences, and then mapped to GRCh38 using GSNAP (17).

Driver alterations and mutational signature analysis

Driver alterations were identified as putative loss-of-function mutations in previously reported ccRCC and pRCC driver genes, including *VHL, PBRM1, SETD2, PIK3CA, MTOR, PTEN, KDM5C, CSMD3, BAP1, TP53, TSC1, TSC2, ARID1A, TCEB1, MET, NF2, KDM6A, SMARCB1, FAT1, STAG2, NFE2L2* and *CDK2NA,* as well as previously established arm-level driver SCNAs in chromosomes 3p loss, 5q gain, 4q loss, 8p loss, and 14q loss (7,22–24). Loss-of-function mutations were defined as indels, nonsense SNVs, splice site dinucleotide SNVs, and putatively deleterious missense SNVs according to SIFT (25), PolyPhen (26), or CONDEL (27).

The sample frequencies of SNV trinucleotide motifs were extracted using the SomaticSignatures R package (28). Signatures previously identified by Alexandrov et al (29)

were then fit to the resulting motif spectra using non-negative least squares with the MutationalPatterns R package (30).

Differential gene expression analysis

RNA-seq reads within gene coding regions were quantified using HTseq (31), and differential expression was assessed using DESeq2 (32). Genes were considered differentially expressed if the comparison had $p_{adjusted} \leq 0.05$, and the sign of the log-fold change was consistent among individual regions in more than half of the patients. The relevant pairwise comparisons were performed while controlling for patient-specific expression. Gene set overlap analysis of differentially expressed genes was performed using Enrichr (33), which performs a hypergeometric test, and gene set enrichment analysis (GSEA) was performed using the R package fGSEA (34). For the latter, all genes included in the differential expression testing were ranked by Wald statistic of the comparison.

Phylogenetic analysis and lesion emergence timing

Phylogenetic trees were constructed using the method of maximum parsimony, on the basis of binary presence or absence of each mutation per sample. Trees were reconstructed 100 additional times on nonparametric bootstrap samples of the input mutation sets, according to the method of Felsenstein (35). We used the phangorn (36) and the APE R packages (37) for tree construction and bootstrapping respectively. To assess the relative order of region emergence, the sums of the internal branch lengths from the root to the region parent nodes were compared in each bootstrap replicate. In order to estimate absolute timing of lesion parent clone emergence, we adapted the linear mixed modeling approach of Mitchell et al (38), using the lme4 package (39). (See Supplementary Methods for further details.)

Results

Genomic alteration patterns suggest advanced disease, minimal overlap with previously described evolutionary subtypes

We investigated a cohort consisting of 16 total patients, 14 of whom had ccRCC, and 2 had pRCC (Figure S1, Table S1). A total of 78 distinct tumor regions from these patients were assayed for this study, including multiple primary regions and at least one VTT region from all patients, as well as at least one metastasis from eight of the 11 metastatic cases (Table S2).

Five patients had more than one metastasis sampled. Exome sequencing was performed on all patients and tumor regions along with one matched normal sample per patient. RNA sequencing (RNA-seq) was performed on 56 tumor regions from 12 patients along with one matched normal sample per patient. Somatic point mutation analysis identified a range of 3.6-25.2 unique mutations/Mb per patient, and a range of 1.8±0.24 mutations/Mb to 8±2.4 mutations/Mb per sample (Table S3). Tumors in this cohort exhibited a wide range of global clonality--the percentage of mutations in any given sample that were detected in all samples from the same tumor--with values as low as 6.5% and as high as 81% (Table S3). The ccRCC cases exhibited a median average of 48% globally clonal mutations, higher than previous reports have indicated (8). Both pRCC patients had an average of 73% globally clonal mutations, higher than all but two non-metastatic ccRCC cases (73% in PT6 and 81% in PT13). Lower intratumoral heterogeneity and higher global clonality have been previously reported for pRCC, when compared with ccRCC (40).

We analyzed the driver alteration landscape of all patients and samples, accounting for both somatic genomic alterations and gene expression of known driver genes in RCC (for details see Methods). All patients exhibited clonal, putatively inactivating mutations in at least one previously identified RCC driver gene (Figure 1A). Expression of canonical RCC driver genes was generally consistent with the expected behavior for tumor suppressors (downregulated in tumor relative to matched normal), with the exception of *FAT1*, *MET*, *SMARCB1*, and *TP53* (Figure S3). Driver mutations with at least 5% prevalence in this cohort displayed a surprising level of homogeneity relative to other RCC cohorts. In six patients these mutations were entirely clonal (i.e., were observed in all regions; PT7, PT8, PT10, PT11, PT13, PT16), and in an additional four patients there was only one region displaying heterogeneity (mutation presence/absence; PT5, PT6, PT14, PT15). There were several examples of parallel mutations in separate tumor regions of six patents (Figure 1B), consistent with previous studies, and suggesting selective pressure for loss of function in these genes (8).

All ccRCC patients exhibited either truncal *VHL* mutation or truncal loss of chromosome 3p, with co-occurrence of these truncal driver alterations in 10 patients. Consistent with previous observations (22), *PBRM1* was the second most commonly mutated of the assessed established driver genes, harboring globally clonal mutations in three ccRCC patients, and subclonal mutations in a further five. *PBRM1* mutation has been associated with better prognosis, except when *BAP1* is co-mutated, which is associated with aggressive disease and poor prognosis (41,42). However, PT3 had *PBRM1, BAP1* co-mutation but exhibited no evidence of disease on final follow-up (Figure S1). For the two papillary cases, PT11 exhibited

mutations in a number of ccRCC driver genes (*PBRM1*, *SETD2*, and *ARID1A*), whereas patient PT4 exhibited only subclonal mutation of *ARID1A*. In metastatic patients, we observed an enrichment of SCNA losses in chr14q, 4q, and 8p relative to non-metastatic patients, consistent with the more advanced disease state. We did not identify any recurrently mutated genes associated with the VTT samples, aside from genes known to be frequent false positives due to long coding sequences or a large number of paralogs (Table S4).

To investigate whether these tumors followed prescribed patterns of evolution, we used canonical driver alterations to assign the ccRCC patients to the seven evolutionary subtypes described by Turajlic et al (7). We note that we were unable to consider the methylation status of driver genes; only loss-of-function mutations in driver genes and driver SCNAs were included. We were able to assign three of the 14 ccRCC cases confidently to evolutionary subtypes: PT6, PT9, and PT10 (Table 1). Four additional cases exhibited similarities to the established subtypes, but could not be clearly assigned. The remaining seven cases were inconsistent with the pattern of driver mutations characteristic of any of the seven subtypes.

Prevalence and co-occurrence of two types of DNA repair deficiency mutational signatures

We next asked whether specific mutational signatures were identified in patients with VTT. We examined somatic SNV substitution patterns, along with their trinucleotide context, and compared the resulting spectra to previously identified mutational signatures (29). Determination of the relative contribution of known signatures to the observed spectra revealed broad similarities within patients and heterogeneity among patients (Figure 2A). Extraction of the signatures predominantly contributing to each sample's mutational spectra revealed an enrichment of signatures associated with age (S1A and S1B), mismatch repair deficiency (S6), and *BRCA1/2* mutation (S3) across a majority of patients (Figure 2B).

The most prevalent signatures across our patients were signatures S3 and S1A. S3 is typically associated with *BRCA1/2* mutation and has been previously observed in RCC (5). Almost all patients (15/16) exhibited some degree of signature S3 (53/78 regions exhibited >10% contribution), but there were only two patients with potential deleterious *BRCA1/2* alterations. Patient PT8 exhibited a truncal frameshift variant in *BRCA1*, while patient PT4 exhibited a truncal variant in a splice region of the gene *ZAR1L*, which occurs immediately upstream of *BRCA2*. The presence of signature S3 in the absence of BRCA mutations is a described phenomenon (43,44). The only case lacking the S3 "BRCA-ness" signature was patient PT2, who instead displayed a strong signature S6 (associated with DNA mismatch repair

deficiency). Strikingly, this patient also harbored a truncal nonsense mutation in *MLH1*, which is known to cause mismatch repair defects but is rare in RCC (45). Moreover, this patient exhibited by far the most mutations per sample (7.9 mutations/Mb compared to an average of 2.8 mutations/Mb for the other patients), and by far the most total unique mutations (25.2 mutations/Mb compared to an average of 5.7 for the other patients). Signature S6 was also identified as exhibiting >10% contribution to at least one tumor region in nine additional patients. This observation is consistent with previous reports, but is unexpected in light of the relatively low mutational burden of RCC (29). Interestingly, S3 and S6 co-occurred in at least one region in 14/16 patients, and co-occurred at >10% contribution each in eight patients. Signature S1A is associated with age, and also occurred in most patients (14/16) and in most tumor regions (>10% contribution in 55/78 regions). Presence of age-associated signatures is consistent with previous assessments of RCC mutational processes (29).

Phylogenetic and timing analyses suggest relatively early emergence of metastasizing clones in some patients

We next asked whether the VTT acts as a reservoir for hematogenous metastases. If this were the case, VTT subclones should typically emerge prior to metastases and the metastases should share a clonal ancestry with the VTT. To test this hypothesis, we performed phylogenetic analysis using binarized somatic mutation calls for the eight patients that had primary, VTT, and metastasis lesions (Figure 3). In seven of eight patients, the VTT had a distinct clonal ancestry from the earliest metastasis (PT10, PT2, PT3, PT4, PT11, PT1, PT12). In PT14, the VTT and LN metastasis shared an immediate ancestor. PT10 was a more complex case, wherein the VTT shared an immediate ancestor with the LN metastasis, but the two adrenal gland metastases arose earlier from a distinct ancestor. To establish the certainty of the phylogenetic reconstruction, and in particular to establish the confidence in the placement of specific nodes, we performed non-parametric bootstrapping of the mutational tree (13). Overall across the eight patients, the prior assertions of distinct clonal origins for VTT and metastasis were well supported, with nodal bootstrap values consistently above 70 (top panels of Figure 3A-C). However, PT4 had lower bootstrap values for the two key nodes (44 and 41), suggesting less certainty in the phylogenetic placement of the clonal ancestors for the AG, BN, P1, P2, and VTT regions. These results suggest that in most patients the VTT likely did not act as a reservoir for hematogenous metastases.

The relative timing of a given phylogenetic node can be approximated by the sum of the internal branch lengths leading up to that node. However, a single "representative" tree does not fully capture the inherent uncertainty in these branch lengths. To make our nodal confidence analysis more quantitative, we developed a novel approach for quantifying total internal branch lengths that also captured the uncertainty in the phylogenetic reconstruction of the clonal ancestors of the VTT or of the metastasis. For each tumor, we quantified the total internal branch lengths across 100 bootstrap trees, and plotted the resulting empirical distribution of distances from the root node to branch points immediately preceding either VTT or metastasis. These root-node distance distributions then provide a measure of certainty regarding which region emerged first (Figure 3A-C, bottom panels). The frequency with which one branch-point precedes another in the bootstrap sampled trees provides an estimated p-value (Figure 3D). For example, in PT10 the AG1 and AG2 metastases showed clearly lower root-node distance distributions than the VTT and LN metastasis, suggesting that AG1 and AG2 emerged significantly earlier in mutational time. Similarly, in PT4 the AG and BN metastases showed consistently lower root-node distances than the VTT, with 96/100 bootstrap trees supporting the earlier emergence of the metastases (Figure 3A and D). Taken together, these results suggest that not only does the VTT typically not give rise to metastases, but many metastases actually emerge earlier than does the VTT.

This analysis revealed three categories of RCC tumors: (1) early emergence of at least one metastasis prior to VTT emergence, (2) late emergence of the earliest metastases after VTT emergence, and (3) similar timing of VTT and the earliest metastasis emergence due to their divergence from the same branch. Patients PT2, PT3, PT4, and PT10 fell clearly into the early metastasis class ($p < 0.05$, Figure 3A). Patient PT14 exhibited similar timing of relative emergence due to the VTT and metastasis sharing the same immediate ancestor ($p < 0.05$, Figure 3B). Patient PT12 fell clearly into the late metastasis class ($p < 0.01$, Figure 3C). Patients PT1 and PT11 were most consistent with the early metastasis class, but there was lower certainty ($p \cong 0.26$ and $p \cong 0.11$, respectively) in distinguishing the root-node distance distributions for VTT and metastasis owing to the small numbers of mutations distinguishing their immediate ancestors (Figure 3D). We note that in three patients (PT2, PT10, PT11), the earliest metastatic clones emerged prior to diversification of the earliest primary tumor clones.

Estimating the absolute timing of metastatic clone emergence could have important implications for clinical care of RCC. Having established the relatively early emergence of metastatic clones relative to VTT clones, we sought to determine the age of the patients at which metastases emerged, as well as the number of years between metastasis emergence

and clinical diagnosis. We adapted the modeling approach of Mitchell et al (16) by generalizing it to encompass any node on the tumor phylogeny. This allowed us to infer the absolute timing of metastatic subclone emergence relative to the root (most recent common ancestor, or "MRCA"), to the VTT subclone, and to age at diagnosis. Using the average of the root-node branch lengths from each patient's bootstrap sampled trees, we predicted the patient ages at which the earliest metastasis and VTT branching events occurred (Figure 3E). This timing analysis suggested that metastases emerged 11.5-30.3 years prior to diagnosis (Table S5) in the seven assessed patients (PT2 was excluded due to the *MLH1* mutation leading to violation of model assumptions). This analysis largely recapitulated the observations from the phylogenetic trees, with the early-metastasis patients also exhibiting earlier absolute estimates of metastasis emergence as compared to VTT emergence. Three of these patients appeared to show nearly concomitant emergence of the VTT and metastasis clonal ancestors (PT1, PT4, PT11), and the other two patients had significant, multi-year lags between the emergence of the VTT and metastasis clonal ancestors (PT3, PT10). In PT12, the sole patient in whom the emergence of the VTT preceded the earliest metastasis, the VTT clonal ancestor emerged a striking 33.4 years prior to the emergence of the adrenal gland metastases.

TNF-α signaling is associated with VTT emergence, and MTOR signaling is associated with metastasis

To explore whether distinct molecular programs govern the outgrowth of VTT from primary tumor, as well as metastatic clone outgrowth, we performed paired differential gene expression analysis on 12 patients (RNA-seq was not available for PT1, PT2, PT3, and PT4). We first compared primary tumor regions to VTT regions while accounting for patient-specific variability in gene expression. We identified 30 genes as significantly upregulated in VTT relative to primary tumor, and 12 as significantly downregulated (Figure 4A and B).

Gene set overlap analysis of the differentially expressed genes using Enrichr (33) revealed striking overlap with the TNFα signaling Hallmark pathway (Figure S6A). Out of the 42 differentially expressed genes, 17 are associated with TNFα signaling (adjusted p-value < 10^{-6}). Gene set enrichment analysis (GSEA) with the full set of detected genes (see Methods) revealed a number of significantly upregulated Hallmark pathways (46), including inflammatory response (Figure 4C). Intriguingly, the sole significantly downregulated pathway was epithelial to mesenchymal transition (EMT). GSEA further revealed a number of significantly enriched gene ontology biological process (47) (GOBP) terms (Figure 4D). The upregulated GOBP

terms were concordant with the observed enrichment of the Hallmark inflammatory response pathway, and included many immune-related processes, as well as some related to the cell cycle. An overlap of our differentially expressed gene list with genes identified by a previous study (4) resulted in three genes in common: *MARCO*, *PF4*, and *PPBP*. These genes as well as others that are upregulated in VTT (e.g., *APOE*) are associated with activated macrophages and platelets.

To address the possibility that these differences in gene expression were driven by specific cell types, we performed histology analysis and gene expression deconvolution analysis. The histology analysis confirmed that significant neutrophilic infiltrate was not detected in VTT or primary tumors, and IHC analysis suggested no increase in CD8 immune infiltration of VTT as compared to the primary tumor (Figure 4B). We then used xCell (48) to assess the enrichment of gene expression signatures associated with 67 immune and stromal cell types. We normalized the scores within each cell type and patient, and compared them across tumor region types (Figure S7). No cell type was significantly different across regions after correction for multiple testing. Together, these results suggest that TNFα signaling and inflammation-like changes in genes expression are associated with outgrowth of VTT from primary tumor, but the TNFα likely does not arise from lymphocytic infiltration of tumor.

We next compared primary tumor regions to metastases, and identified 19 genes as significantly downregulated in the metastases relative to the primary tumors, and five genes as upregulated (Figure 5A and B). Notably, *MTOR* was consistently upregulated in metastases relative to primary tumor regions. Although no terms were enriched in these 24 genes alone, GSEA with the full set of detected genes revealed an upregulation of MTORC1 signaling in metastases, along with *MYC* targets and oxidative phosphorylation (Figure 5B). These data are consistent with previous observations that mutations that increase mTOR activity are associated with metastatic RCC (49).

Finally, we compared gene expression in the metastases to that in VTT, and identified four genes as significantly upregulated in the metastases relative to VTT, and five genes as downregulated (Figure S8A). This analysis was restricted to the four patients for which we had both metastatic samples and RNA-seq data, and so may be under-powered.

Discussion

In this study we analyzed 16 RCC patients by MR-seq to characterize the genomic and evolutionary landscape of tumors with VTT. The unique nature of these cases--with a distinct

lesion extending from the primary tumor and in some cases with separate metastases--also enabled us to investigate how and when new lesions arise from the primary tumor. Although only eight patients had both VTT and metastasis, our study surveyed a diverse variety of metastatic sites across these cases. Among metastatic cases, genomic alterations were generally consistent with advanced disease state. Given that RCC is known for frequent subclonal driver mutations, it was somewhat surprising that most tumors tended to share the more prevalent driver mutations across all regions (Figure 1A). Interestingly, the six tumors with higher regional heterogeneity among driver mutations also tended to have parallel mutations in separate tumor regions (Figure 1B). This may point to early RCC clones tending to grow in "fits and starts", and only when a critical set of hallmark pathways become altered can full-blown tumor lesions emerge. We found that while 3/16 cases exhibit patterns consistent with the seven evolutionary subtypes proposed by Turajlic et al (7), others appear to follow distinct evolutionary modes (Table 1). Notably, 37% of the TRACERx RCC cases were also considered unclassifiable. We did not identify candidates for recurrently mutated genes restricted to the VTT, suggesting that tumor-intrinsic epigenetic factors or tumor-extrinsic factors may lead to VTT formation.

As has previously been reported, we observed extensive, early branching evolution in the ccRCC cases (Figure 3A-C and Figure S4). The two pRCC cases exhibited a greater degree of truncality (and later branching) than the metastatic ccRCC cases. At this time, too few pRCC cases have been analyzed by MR-seq to generalize this observation. Despite the genetic heterogeneity, mutational signatures were relatively concordant across regions, with occasional exceptions such as region P3 from patient PT12, which was the only region to exhibit substantial contributions from signature S5. We observed signature S6, which is associated with mismatch repair deficiency, consistently across most patients. The presence of this signature, albeit at low levels in most cases, is seemingly at odds with the relatively low mutation burden broadly observed in RCC (29). Given that many RCC tumors respond to immunotherapy agents, future studies should explore whether the level of S6 represents a potential biomarker. We also observe signature S3, or "BRCA-ness", in a substantial proportion of patients, consistent with previous observations (5). Genes required for homologous recombination and mismatch repair are downregulated by *VHL* inactivation, presumably via the resulting stabilization of HIF-1α (50). The related presumed deficiency in double-strand break repair has generated recent interest in the use of PARP inhibitors to treat RCC (51), and the level of signature S3 may also represent a potential biomarker.

Our phylogenetic analysis of eight metastatic cases with VTT ruled out the possibility of direct hematogenous seeding of metastases from the VTT, with the possible exception of PT14, in which the VTT and LN metastasis share a clonal ancestor. Among these eight cases, we identified three primary modes of interplay between VTT and metastasis: early metastasis, late metastasis, and concomitant metastasis arising from the same parent clone. The latter two modes are consistent with what was described for the six metastatic cases with VTT in the TRACERx cohort. However, most tumors in our cohort (6/8) displayed a distinct early emergence of metastases relative to VTT. Indeed, in PT10 the two adrenal gland metastases were the first to emerge from the root node. A caveat of this analysis is that we are reliant on the use of the parent node in the phylogenetic tree as a proxy for the emergence of each tumor region or lesion. We cannot be certain that VTT or metastatic competence was attained at the time of this branching event, as it could have occurred later via subsequent mutations or epigenetic processes. Our findings that VTT development may be separate from the development of metastatic disease is consistent with clinical experience where aggressive surgical intervention despite VTT can result in long-term survival. Moreover, this apparent independence is consistent with the fact many patients with VTT present without metastatic disease, and the fact that probability of metastatic presentation is proportional to primary tumor size, while probability of VTT exhibition is not (52).

Our timing analysis enabled us to estimate absolute timing of metastatic clone emergence, as well as VTT emergence. This approach is contingent on the assumption that the tumor mutation rate is not substantially different from the background somatic mutation rate, and that mutation rates do not vary substantially across tumor regions. In our patients, the presence of a number of non-age associated mutation signatures suggests that the former assumption may be violated. Indeed, two of the mutational signatures we observed, S3 and S6, have been associated with higher SNV mutation rates (53,54). However, upon fitting the model, the patient-specific slope estimates were relatively similar to one-another, with the exception of PT2 (Figure S5). The truncal inactivating *MLH1* mutation likely led to an increased mutation rate in this tumor, and we therefore excluded PT2 from this analysis.

The results of the timing analysis suggested that in all patients the earliest metastases (i.e., the clones that ultimately constituted the earliest metastases) emerged well before diagnosis--as long as three decades in one patient. This early metastatic clone emergence is consistent with recent findings that metastases in colorectal, breast, and lung cancer are disseminated early relative to time of diagnosis (13,14). A caveat is that our timing analysis provides an estimate of the emergence time for a metastatic subclone--it does not establish

when that subclone seeded a distal site. However, given that a large number of RCC cases present with metastases on diagnosis, early detection studies are needed to examine whether RCC metastases are formed well before the disease is clinically evident or detected. A better understanding of the patterns and timelines of RCC evolution has relevance not only for small renal masses but also for more advanced disease. The increased application of percutaneous tumor biopsy prior to treatment may allow personalized management based on molecular characterization of the cancer.

Differential expression analysis revealed 42 genes that were consistently differentially expressed between primary tumor and VTT. VTT regions showed upregulation of TNFα signaling (17 of 42 genes) and related inflammatory processes relative to primary tumor regions. There was no evidence from IHC analysis or immune-cell gene expression signatures that these differences were due to lymphocytic infiltration of VTT. It is possible that the source of TNFα in these RCC tumors was myeloid, stromal, or endothelial cells. RCC tumors are known to have an angiogenic phenotype and substantial myeloid involvement, leading us to speculate that intratumoral vascular endothelial cells and/or activated macrophages produce TNFα, and potentially other cytokines, leading the tumor to invade the renal vein.

The sole significantly depleted Hallmark pathway in the VTT identified by GSEA was epithelial-to-mesenchymal transition (EMT). This could indicate that the VTT is in general less competent to generate metastases than the primary tumor, rendering it a less frequent source of hematogenous metastases than might otherwise be expected. Although the genes we identified as differentially expressed between primary and VTT regions had minimal overlap (only three genes) with those from a previous study (4), the two analyses did identify similar pathways as upregulated (e.g., inflammatory response) and downregulated (e.g., EMT). A smaller number of genes were observed to be differentially expressed between primary and metastatic sites, but *MTOR* notably featured among them, with mTORC1 targets further featuring in GSEA results. *MTOR* signaling is thought to drive invasiveness in RCC and other cancers, and the use of mTORC1 inhibitors has seen some clinical success (49).

Taken together, our data reveal new dimensions of RCC VTT and metastasis biology. They highlight the need for earlier detection of RCC, and suggest that metastatic clones may emerge and remain largely dormant for years before becoming clinically detectable. They also suggest potential expression and mutational biomarkers of RCC patient prognosis and treatment response.

Acknowledgments

References

1. Padala SA, Barsouk A, Thandra KC, Saginala K, Mohammed A, Vakiti A, et al. Epidemiology of Renal Cell Carcinoma. World J Oncol. 2020;11:79-87–87.

2. Quencer KB, Friedman T, Sheth R, Oklu R. Tumor thrombus: incidence, imaging, prognosis and treatment. Cardiovasc Diagn Ther. 2017;7:S165–77.

3. Psutka SP, Leibovich BC. Management of inferior vena cava tumor thrombus in locally advanced renal cell carcinoma. Ther Adv Urol. 2015;7:216–29.

4. Wang X-M, Lu Y, Song Y-M, Dong J, Li R-Y, Wang G-L, et al. Integrative genomic study of Chinese clear cell renal cell carcinoma reveals features associated with thrombus. Nat Commun. 2020;11:739.

5. Warsow G, Hübschmann D, Kleinheinz K, Nientiedt C, Heller M, Van Coile L, et al. Genomic features of renal cell carcinoma with venous tumor thrombus. Sci Rep. 2018;8:7477.

6. Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, et al. Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. Cell 2018;173:581-594.e12.

7. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. Cell 2018;173:595-610.e11.

8. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. 2014;46:225–33.

9. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366:883–92.

10. Whitson JM, Reese AC, Meng MV. Population based analysis of survival in patients with renal cell carcinoma and venous tumor thrombus. Urol Oncol. 2013;31:259–63.

11. Abel E. Jason, Spiess Philippe E., Margulis Vitaly, Master Viraj A., Mann Michael, Zargar-Shoshtari Kamran, et al. Cytoreductive Nephrectomy for Renal Cell Carcinoma with Venous Tumor Thrombus. J Urol. 2017;198:281–8.

12. Gong J, Maia MC, Dizman N, Govindarajan A, Pal SK. Metastasis in renal cell carcinoma: Biology and implications for therapy. Asian J Urol. 2016;3:286–92.

13. Hu Z, Ding J, Ma Z, Sun R, Seoane JA, Scott Shaffer J, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. Nat Genet. 2019;51:1113–22.

14. Hu Z, Li Z, Ma Z, Curtis C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. Nat Genet. 2020;52:701–8.

15. Brufau BP, Cerqueda CS, Villalba LB, Izquierdo RS, González BM, Molina CN. Metastatic Renal Cell Carcinoma: Radiologic Findings and Assessment of Response to Targeted Antiangiogenic Therapy by Using Multidetector CT. RadioGraphics. 2013;33:1691–716.

16. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018;173:611-623.e17.

17. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010;26:873–81.

18. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-

quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40:11189–201.

19. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics. 2012;28:1811–7.

20. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

21. Ha G, Roth A, Khattra J, Ho J, Yap D, Prentice LM, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. Genome Res. 2014;24:1881–93.

22. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep. 2018;23:313-326.e5.

23. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med. 2016;374:135–45.

24. Creighton CJ, Morgan M, Gunaratne PH, Wheeler DA, Gibbs RA, Gordon Robertson A, et al. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499:43–9.

25. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812–4.

26. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Curr Protoc Hum Genet. 2013;76:7.20.1-7.20.41.

27. Yuan X, Bai J, Zhang J, Yang L, Duan J, Li Y, et al. CONDEL: Detecting Copy Number Variation and Genotyping Deletion Zygosity from Single Tumor Samples Using Sequence Data. IEEE/ACM Trans Comput Biol Bioinform. 2020;17:1141–53.

28. Gehring JS, Fischer B, Lawrence M, Huber W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics. 2015;31:3673–5.

29. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.

30. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10:33.

31. Pau G, Reeder J. HTSeqGenie: A NGS analysis pipeline. [Internet]. Bioconductor version: Release (3.13); 2021 [cited 2021 Jun 6]. Available from: https://bioconductor.org/packages/HTSeqGenie/

32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

33. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90–7.

34. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021;060012.

35. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evol Int J Org Evol. 1985;39:783–91.

36. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3.

37. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20:289–90.

38. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018;173:611-623.e17.

39. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67:1–48.

40. Zhu B, Poeta ML, Costantini M, Zhang T, Shi J, Sentinelli S, et al. The genomic and epigenomic evolutionary history of papillary renal cell carcinomas. Nat Commun. 2020;11:3096.

41. Dizman N, Philip EJ, Pal SK. Genomic profiling in renal cell carcinoma. Nat Rev Nephrol. 2020;16:435–51.

42. Joseph Richard W., Kapur Payal, Serie Daniel J., Parasramka Mansi, Ho Thai H., Cheville John C., et al. Clear Cell Renal Cell Carcinoma Subtypes Identified by BAP1 and PBRM1 Expression. J Urol. 2016;195:180–7.

43. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017;49:1476–86.

44. Bodily WR, Shirts BH, Walsh T, Gulsuner S, King M-C, Parker A, et al. Effects of germline and somatic events in candidate BRCA-like genes on breast-tumor signatures. PLOS ONE. 2020;15:e0239197.

45. Leach FS, Koh MS, Sharma K, McWilliams G, Talifero-Smith L, Codd A, et al. Mismatch Repair Gene Mutations in Renal Cell Carcinoma. Cancer Biol Ther. 2002;1:530–6.

46. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. 2015;1:417–25.

47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

48. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.

49. Lieberthal W, Levine JS. The Role of the Mammalian Target Of Rapamycin (mTOR) in Renal Disease. J Am Soc Nephrol. 2009;20:2493–502.

50. Scanlon SE, Hegan DC, Sulkowski PL, Glazer PM. Suppression of homology-dependent DNA double-strand break repair induces PARP inhibitor sensitivity in VHL-deficient human renal cell carcinoma. Oncotarget. 2017;9:4647–60.

51. Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins. Phase II Study of Olaparib in Metastatic Renal Cell Carcinoma Patients Harboring a BAP-1 or Other DNA Repair Gene Mutations (ORCHID) [Internet]. clinicaltrials.gov; 2021 Mar. Report No.: NCT03786796. Available from: https://clinicaltrials.gov/ct2/show/NCT03786796

52. Thompson RH, Hill JR, Babayev Y, Cronin A, Kaag M, Kundu S, et al. Metastatic renal cell carcinoma risk according to tumor size. J Urol. 2009;182:41–5.

53. Póti Á, Gyergyák H, Németh E, Rusz O, Tóth S, Kovácsházi C, et al. Correlation of homologous recombination deficiency induced mutational signatures with sensitivity to PARP inhibitors and cytotoxic agents. Genome Biol. 2019;20:240.

54. Zámborszky J, Szikriszt B, Gervai JZ, Pipek O, Póti Á, Krzystanek M, et al. Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions. Oncogene. 2017;36:746–55.

Figure 1.
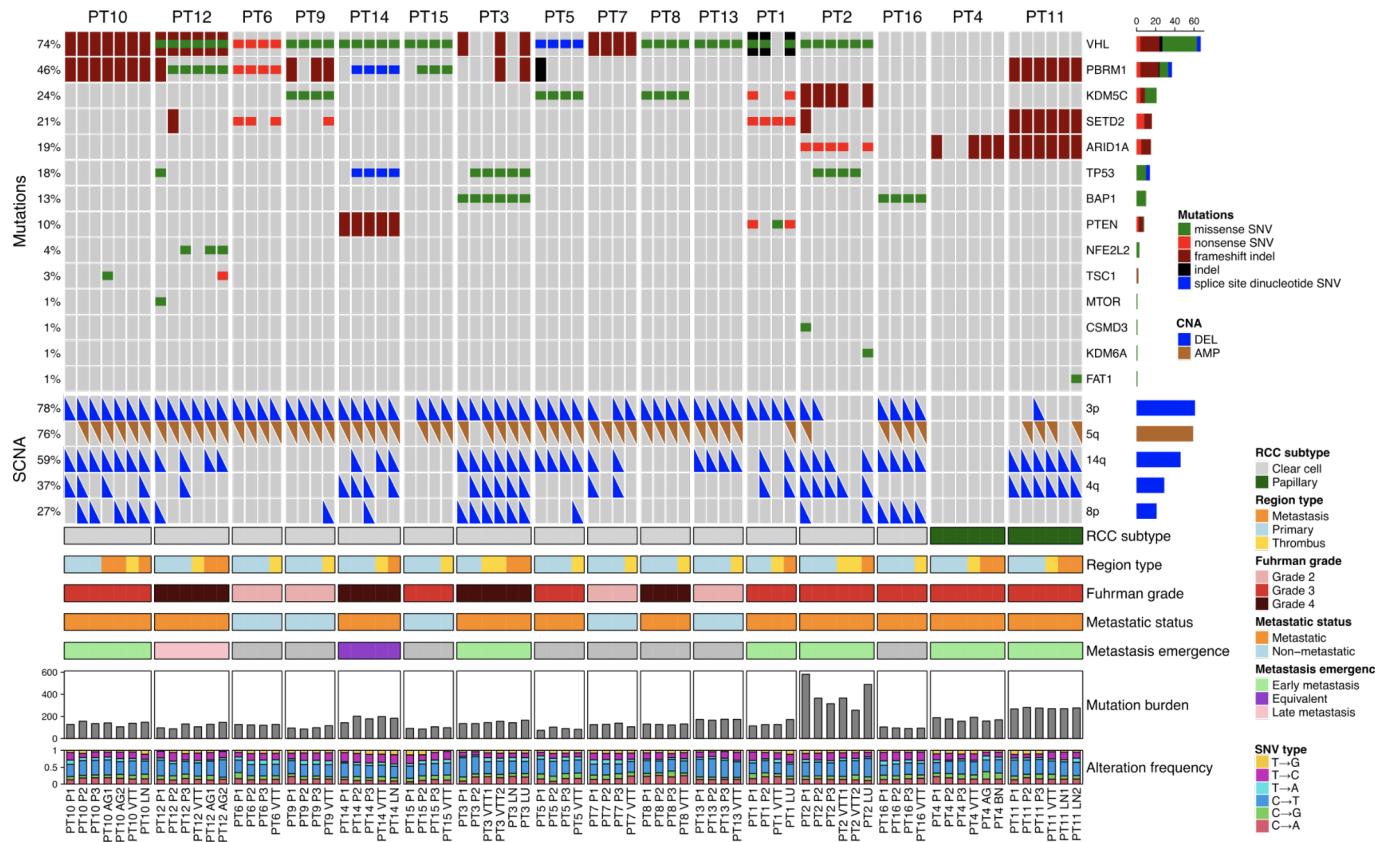
Patient	Mutation	Regions
PT1	VHL H110del	P1, P2
	VHL S111del	LU
	PTEN R130*	P1, LU
	PTEN F104V	VTT
PT2	ARID1A Q1512*	P1, P2, P3, VTT1
	ARID1A R1276*	LU
PT3	BAP1 L14F	P1
	BAP1 H141Y	P2, VTT1, VTT2, LN, LU
PT6	SETD2 L214*	P1, P2
	SETD2 Q2484*	VTT
PT9	KDM5C C730S	P1, P3
	KDM5C S1110*	P2
	KDM5C E189*	P3
	KDM5C D517Y	VTT
PT12	PBRM1 1019fs	P1
	PBRM1 R876P	P2, P3, VTT, AG1, AG2

Figure 1. Genomic alteration landscape across tumor regions in 16 RCC patients with VTT. (A) An oncoprint showing the most prevalent somatic mutations and SCNAs found in this cohort. Prevalences of each alteration are indicated at left. Tumor lesion labels: P = primary tumor; VTT = venous tumor thrombus; AG = adrenal gland metastasis; LN = lymph node metastasis; LU = lung metastasis; BN = brain metastasis. (B) Parallel mutations found in distinct regions of six tumors.

Figure 2.
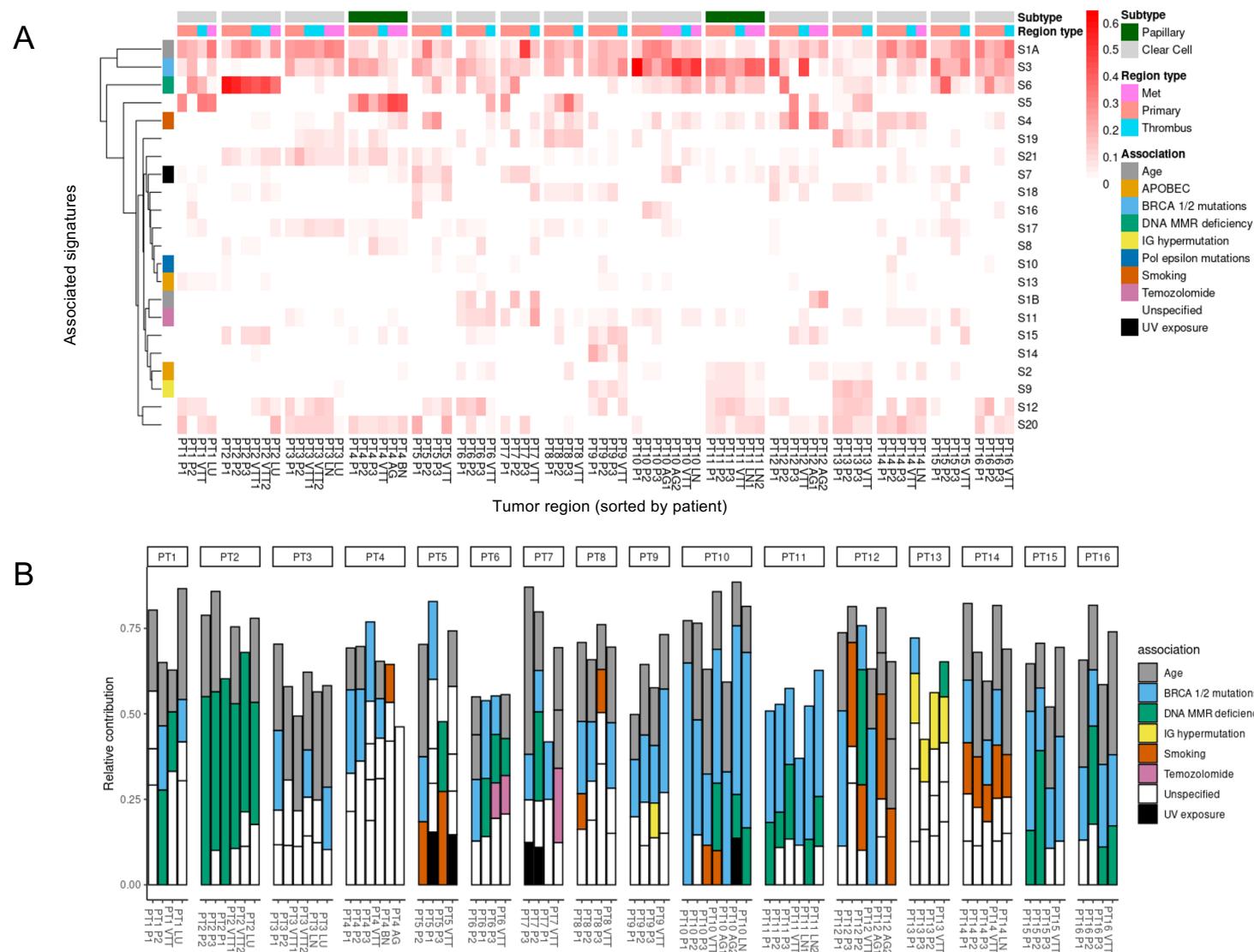
Figure 2. SNV mutational signature analysis. (A) Heatmap of previously established SNV mutational signatures to the mutational spectra of each tumor region across all patients. (B) Contribution of previously established SNV mutational signatures to the mutational spectra of each region across all patients. Included signatures are restricted to those with >10% contribution in a sample.

Figure 3.
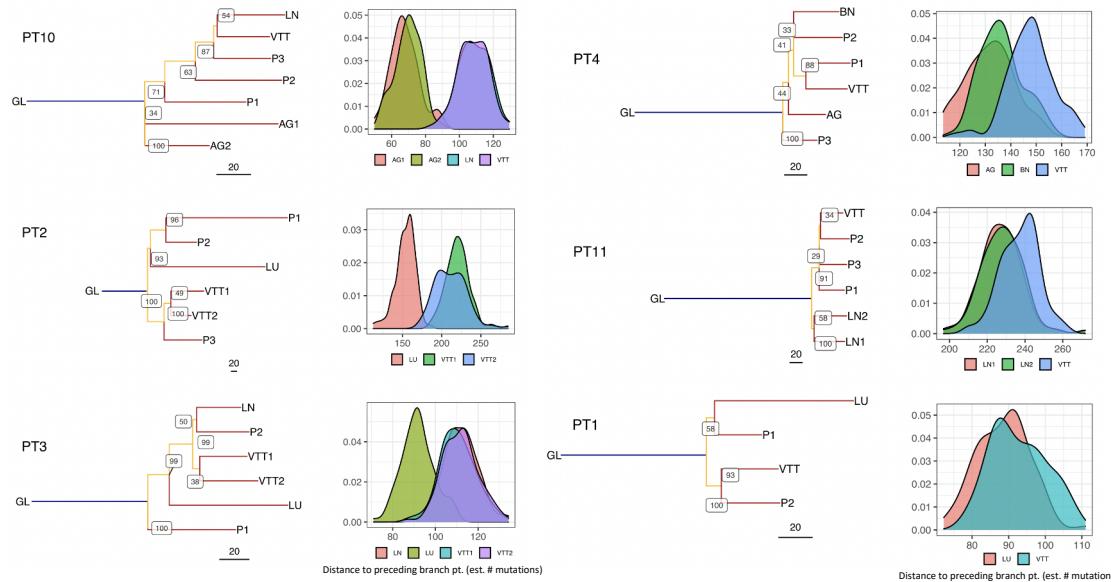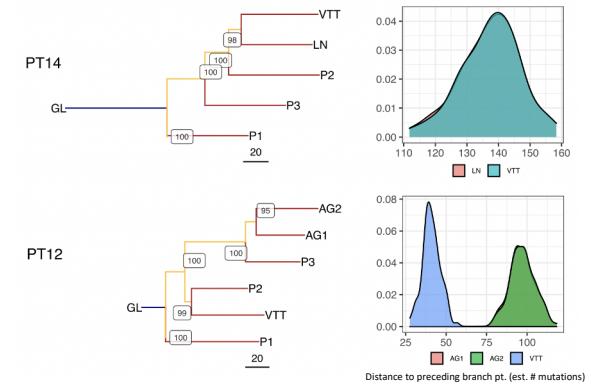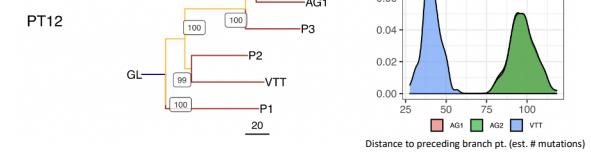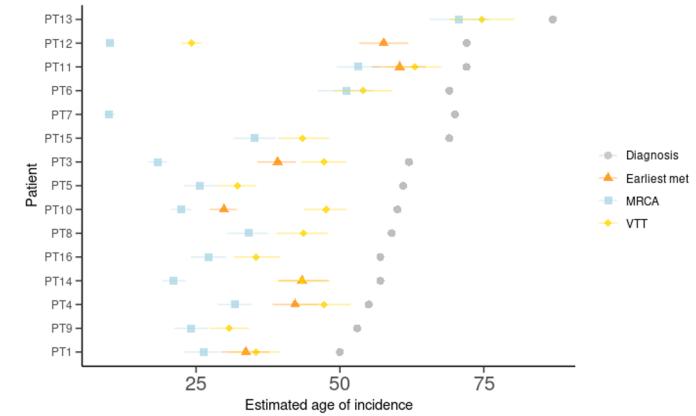
Patient	# rep in which earliest metastasis precedes VTT	# rep in which earliest metastasis and VTT occur simultaneously	# rep in which VTT precedes earliest metastasis	Class
PT10	100	0	0	Earlier metastasis
PT2	100	0	0	Earlier metastasis
PT3	100	0	0	Earlier metastasis
PT4	96	0	4	Earlier metastasis
PT11	89	0	11	Earlier metastasis
PT1	74	3	23	Earlier metastasis
PT14	2	98	0	Simultaneous emergence
PT12	0	0	100	Later metastasis

Figure 3. Phylogenetic and timing analyses of eight metastatic RCC tumors. (A-C) Mutational phylogenies and reconstruction robustness for eight RCC tumors with VTT and distant metastasis. Internal node labels on the phylogenetic trees display the number of bootstrap replicates supporting each node (i.e., the certainty out of 100 that the placement of a particular node is supported). Density plots illustrate the distribution of path lengths to the parent nodes of all non-primary regions derived from 100 bootstrap replicates of the tree building process. Region labels: GL = germline; P = primary tumor; VTT = venous tumor thrombus; AG = adrenal gland metastasis; LN = lymph node metastasis; LU = lung metastasis; BN = brain metastasis. (A) Patients with early emergence of metastases relative to VTT. (B) Patient with near-simultaneous emergence of VTT and the earliest assayed metastases, with both emerging from the same branch. (C) Patient with late emergence of metastases relative to VTT. (D) Table showing the number of bootstrap replicates in which phylogenetic tree reconstruction resulted in earlier, simultaneous, or later emergence of the earliest metastasis relative to the VTT. (E) Timing estimates generated with linear mixed effects models of the emergence of MRCA and earliest metastasizing clones, and the VTT clone, relative to the age at diagnosis. Error bars indicate 95% confidence intervals generated by parametric bootstrapping of the i.i.d errors.

Figure 4.
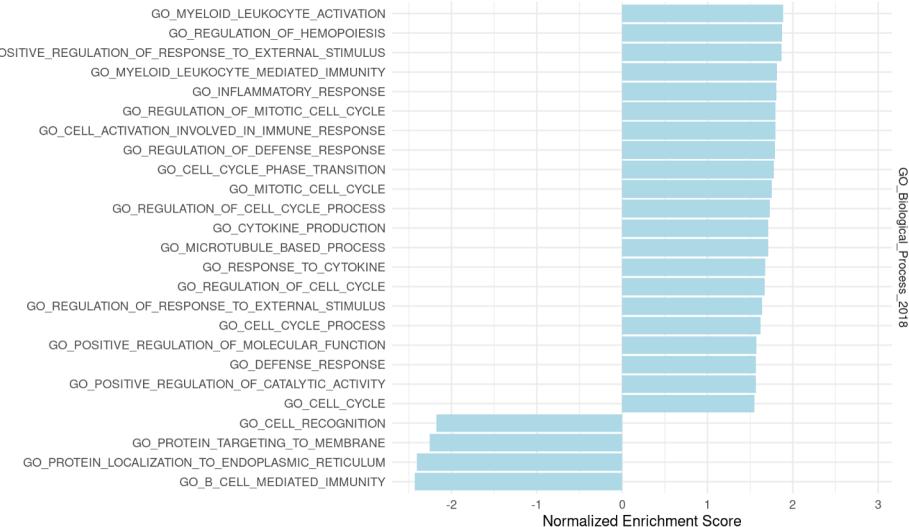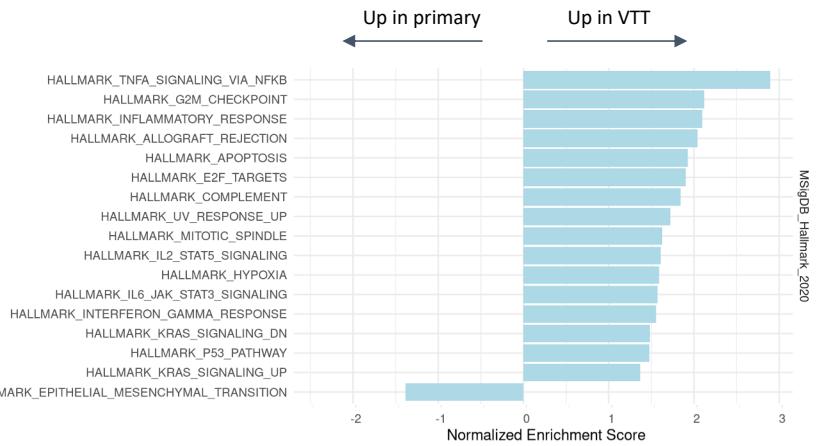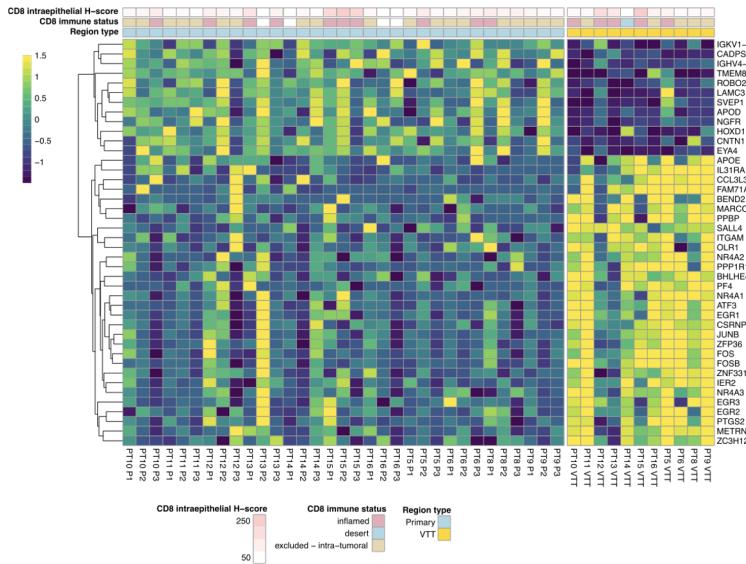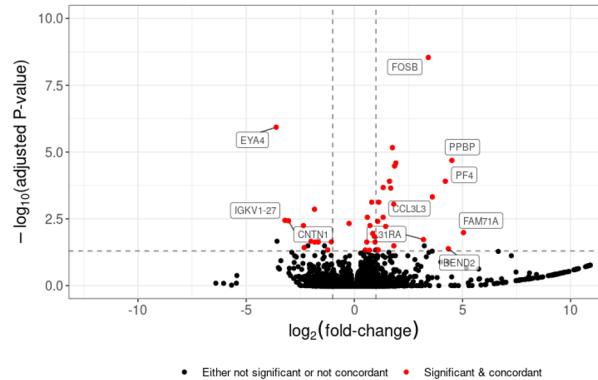
Figure 4. Differential gene expression analysis of VTT versus primary tumor regions. (A) Volcano plot showing log-fold changes and adjusted p-values for differentially expressed genes. Genes highlighted in red were both significantly different and exhibited concordance in the directionality of their changes across regions and patients. A few non-concordant outliers have been removed for clarity of plotting. (B) Heatmap showing expression values of differentially expressed genes between primary and VTT regions. Expression values are normalized per gene and per patient. (C) Significantly enriched or depleted Hallmark pathways in VTT relative to primary regions as determined by GSEA. (D) Significantly enriched GO Biological Process terms enriched or depleted in VTT relative to primary regions as determined by GSEA. Significance is defined as adjusted p-value <= 0.05 (Benjamini-Hochberg).

Figure 5.

A

- Either not significant or not concordant
- Significant & concordant

B

C

Up in primary ← → Up in metastasis

D
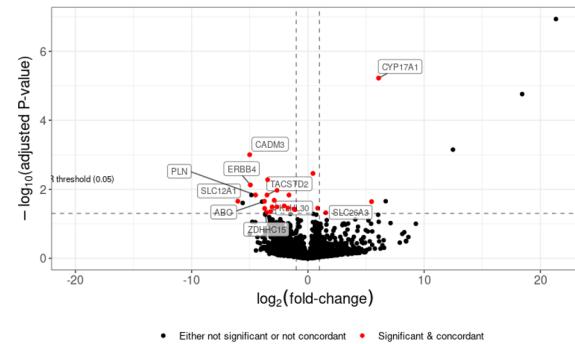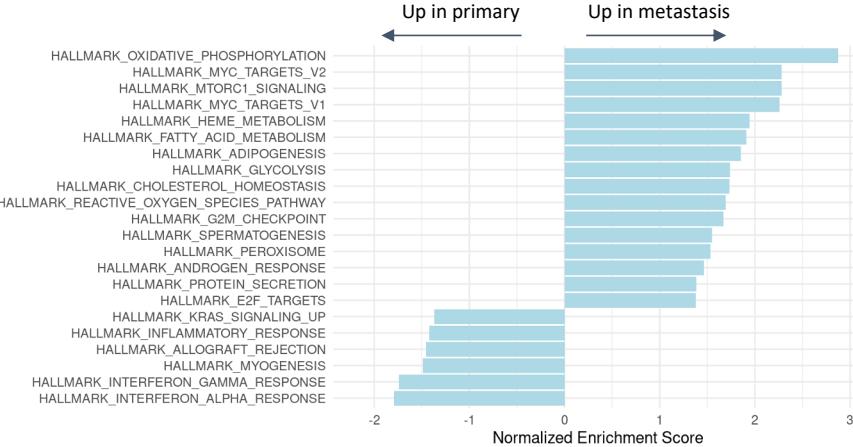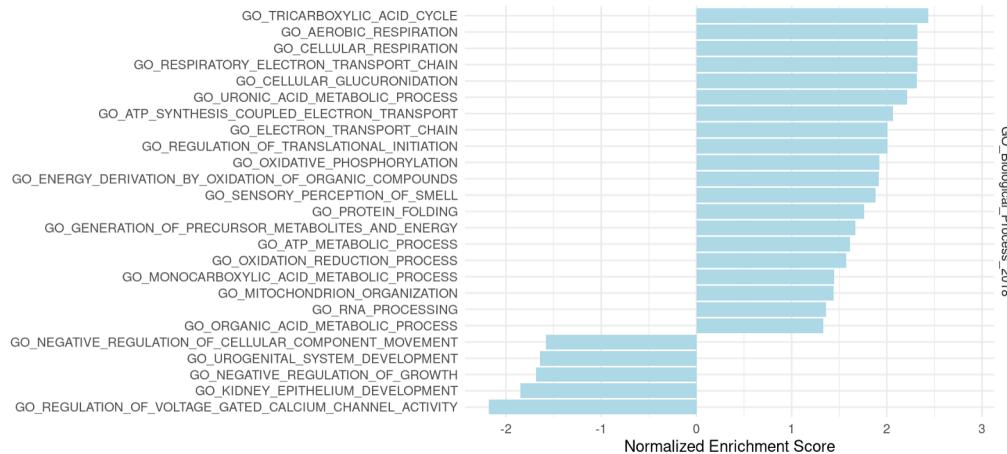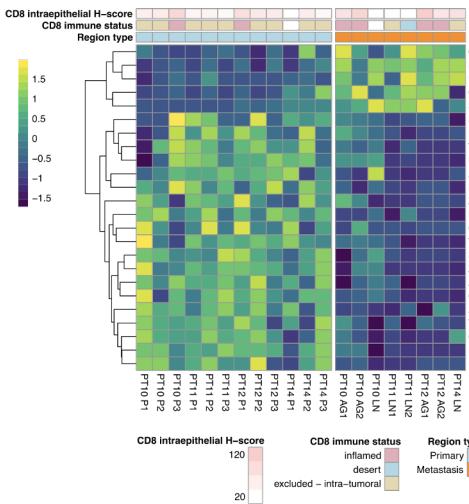
Figure 5. Differential gene expression analysis of metastases versus primary tumor regions. (A) Volcano plot showing log-fold changes and adjusted p-values for differentially expressed genes. (B) Heatmap showing expression values of differentially expressed genes between primary and metastatic regions. Expression values are normalized per gene and per patient. (C) Significantly enriched or depleted Hallmark pathways in metastases relative to primary regions as determined by GSEA. (D) Significantly enriched GO Biological Process terms enriched or depleted in metastases relative to primary regions as determined by GSEA. Significance is defined as adjusted p-value <= 0.05 (Benjamini-Hochberg).

Table 1.

Patient	Evolutionary subtype	trunk/early	branch/late
PT1	undetermined	SETD2, 3p loss	KDM5C, PTEN, VHL, 5q gain, 4q loss, 14q loss
PT2	undetermined	VHL	KDM5C, ARID1A, TP53, KDM6A, SETD2, CSMD3, 3p loss, 5q gain, 4q loss, 8p loss, 14q loss
PT3	undetermined	3p loss, 14q loss	TP53,BAP1,PBRM1,VHL, 5q gain, 4q loss, 8p loss
PT4	NA (papillary RCC)		ARID1A
PT5	undetermined	VHL,KDM5C, 3p loss, 5q gain	PBRM1, 8p loss, 14q loss
PT6	subtype 4 (PBRM1 -> SETD2)	VHL, PBRM1, 3p loss, 5q gain	SETD2
PT7	subtype 7 (VHL monodriver)	VHL, 5q gain	3p loss, 14q loss
PT8	undetermined	VHL, KDM5C, 3p loss, 5q gain	
PT9	undetermined	VHL, KDM5C, 3p loss, 5q gain	PBRM1, SETD2, 8p loss
PT10	subtype 5 (PBRM1 -> PI3K)	VHL, PBRM1, 3p loss	TSC1, 5q gain, 4q loss, 8p loss, 14q loss
PT11	NA (papillary RCC)	PBRM1, SETD2, ARID1A, 4q loss, 14q loss	FAT1, 3p loss, 5q gain
PT12	undetermined	VHL, 3p loss, 5q gain	PBRM1, NFE2L2, SETD2, MTOR, TSC1, TP53, 4q loss, 8p loss, 14q loss
PT13	undetermined	VHL, 3p loss, 5q gain, 14q loss	
PT14	undetermined	VHL, PTEN, 3p loss, 5q gain	TP53, PBRM1, 4q loss, 8p loss, 14q loss
PT15	undetermined	VHL	PBRM1, 3p loss, 5q gain
PT16	undetermined	BAP1, 3p loss, 5q gain, 8p loss, 14q loss	

Table 1. Phylogenetic placement of driver alterations (early/late) and concordance of each tumor with previously described evolutionary subtypes (6).

Supplementary case description

All 16 patients had VTT extending at least into the renal vein, and in nine it extended into the inferior vena cava. Seven of the 16 patients exhibited metastases on diagnosis (VTT+met), and metastases were later discovered in an additional four. The primary tumor regions and VTT samples were all derived from the initial VTT thrombectomy, which was performed shortly after diagnosis (in all but one case). Metastatic regions were sampled both from metastasectomies concomitant to the initial surgery, and from recurrences. In three of the six cases where recurrences were sampled, one or more non-surgical interventions preceded the sampling (PT1, PT4, PT11; see Figure S1, Table S1). Semi-quantitative histologic analysis of the 78 tumor regions revealed variable stromal abundance within tumors that was relatively consistent across tumor regions in any one patient. Very little intrastromal inflammatory response or desmoplasia was detected (Figure S2).

Supplementary Methods

AVENIO Millisect tissue harvest

AVENIO Millisect automated dissection for tumor enrichment was performed on all cases. Formalin Fixed Paraffin Embedded (FFPE) tissue blocks were serially sectioned with one section at 4μm, followed by 7 sections at 10μm, followed by 3 sections at 4μm, collected onto Superfrost Plus positively charged slides (Thermo Scientific, Runcorn, UK) and allowed to dry at room temperature overnight. Serial sections 1 and 9 (4μm) were baked at 60℃ for 30 minutes and stained with Hematoxylin and Eosin (H&E) on an automated Leica Autostainer XL using a routine protocol. H&E stained slides were scanned on a NanoZoomer 2.0 HT whole slide imager (Hamamatsu, Bridgewater NJ) at 20X magnification. Scanned slide images were annotated by a pathologist for tumor regions of interest, percent tumor area necrosis (% necrosis/total tumor area) was captured and digital masks were created as a dissection reference.

Tissue sections were dissected using the reference mask image from serial section 1 to collect regions of interest using medium or large AVENIO Millisect milling tips (Roche Sequencing Solutions, Pleasanton, CA), collected with Molecular Grade Mineral Oil (Sigma-Aldrich, St. Louis, MO) as dissection fluid and dispensed into nuclease-free 1.5mL Eppendorf tubes. Dissections from slides 2 through 5 and 6 through 8 were centrifuged for 10 minutes at 20,000rpm to pellet tissue. Portions of mineral oil were removed from the tissue pellets. Pellets from slides 2 through 5 were pooled in a 1.5mL Eppendorf tube and held for DNA extraction and pellets from slides 6 through 8 were pooled in a separate 1.5mL Eppendorf tube and held for

RNA extraction. Post AVENIO Millisect dissected tissue slides were baked at 60℃ for 30 minutes and stained with Hematoxylin and Eosin (H&E) on an automated Leica Autostainer XL using routine protocols and scanned on a NanoZoomer 2.0 HT whole slide imager (Hamamatsu, Bridgewater NJ) at 20X magnification. DNA extraction was performed using the Qiagen AllPrep DNA/RNA tissue kit (Qiagen, Germantown, MD) at Q^2 Solutions (Valencia, CA).

Tumor content ranged from 5 to 99% in analyzed tissue regions. Tumor enrichment was performed using AVENIO Millisect for semi-automated dissection, resulting in tumor input of 5.9-1439.81mm^2 (Table S2) that excluded the majority of surrounding normal tissue and necrotic regions from capture and analysis.

Estimating lesion emergence timing

In order to estimate absolute timing of lesion parent clone emergence, we adapted the linear mixed modeling approach of Mitchell et al (38), using the lme4 package (39). Patient age was modeled as a function of sample mutation counts as both a fixed and random slope, with intercept set to zero:

Eq (1) $y_{p,s} = \beta \times x_{p,s} + b_p \times x_{p,s}$

where $y_{p,s}$ is the age of patient p upon collection of sample s, β a fixed slope term, $x_{p,s}$ is the number of mutations detected in sample s from patient p, and b_p is a random slope term for patient p.

Lesion parent clone emergence timing was estimated by predicting the patient age given the estimated number of mutations present at the branch point immediately upstream of the lesion on the phylogenetic trees. The number of mutations was estimated by calculating the sum of the branch lengths leading to that branch point in each bootstrapped tree, and then taking the mean over all the resulting sums. For example, the number of mutations at VTT emergence in PT4 was estimated by summing over the branch lengths leading to the node immediately upstream of the VTT (i.e. the branch point leading to both P1 and VTT in the representative tree in Figure 3A) in each of the 100 bootstrapped trees generated for that patient, and then calculating the mean of the resulting sums.

The timing of clone emergence was estimated for both the VTT (in all patients) and for the earliest metastasis (i.e. the metastasis with the fewest estimated mutations according to the above approach) in all patients from which a metastatic sample was collected. Confidence

intervals were estimated using parametric bootstrapping of the model residuals with the *bootMer* function from the lme4 package.

Supplementary Figure Legends

Figure S1. Timelines of patient diagnosis, surgeries, and other treatment.

Figure S2: Semi-quantitative analysis of stromal abundance captured 3 levels (low, moderate or high) of stromal presence within tumors with little inflammatory response.

Figure S3. Expression status of putative driver genes across all regions of all patients, excepting CSMD3, which had no uniquely mapped counts in any region. Expression values were normalized by variance-stabilizing transformation, then turned into z-scores normalized within each gene and patient prior to visualization.

Figure S4. Phylogenetic analysis of eight RCC tumors, which are either non-metastatic or are missing metastasis samples. Internal node labels on the phylogenetic trees display the number of bootstrap replicates supporting each node (i.e., the certainty out of 100 that the placement of a particular node is supported). Region labels: GL = germline; P = primary tumor; VTT = venous tumor thrombus.

Figure S5. LME mutation rate model fits (A) with and (B) without PT2.

Figure S6. Enrichr set enrichment of differentially expressed genes between primary and VTT regions. (A) Significantly enriched terms from MSigDB Hallmark pathways. (B) Significantly enriched terms from GO Biological Processes. Significance is defined as adjusted P-value <= 0.05.

Figure S7. xCell enrichment scores for 67 cell types across all regions of all patients. Scores were normalized within each cell type and patient prior to visualization.

Figure S8. Differential expression between VTT and metastatic regions. (A) Volcano plot showing log-fold changes and adjusted P-values for differentially expressed genes. (B) Heatmap showing expression values of genes determined to be differential between VTT and

metastatic regions. Expression values are normalized per gene and per patient. (C) Significantly enriched or depleted Hallmark pathways in metastases relative to VTT regions as determined by GSEA. (D) Significantly enriched GO Biological Process terms enriched or depleted in metastases relative to VTT regions as determined by GSEA. Significance is defined as adjusted p-value <= 0.05 (Benjamini-Hochberg).

Table S1. Extended characteristics of cases in this study.

Table S2. Characteristics of individual samples in this study.

Table S3. Mutation burden and across patients.

Table S4. Mutated genes restricted to VTT and occurring in more than one patient.

Table S5. Timing estimates of metastases, VTT, and diagnosis.

Figure S1.
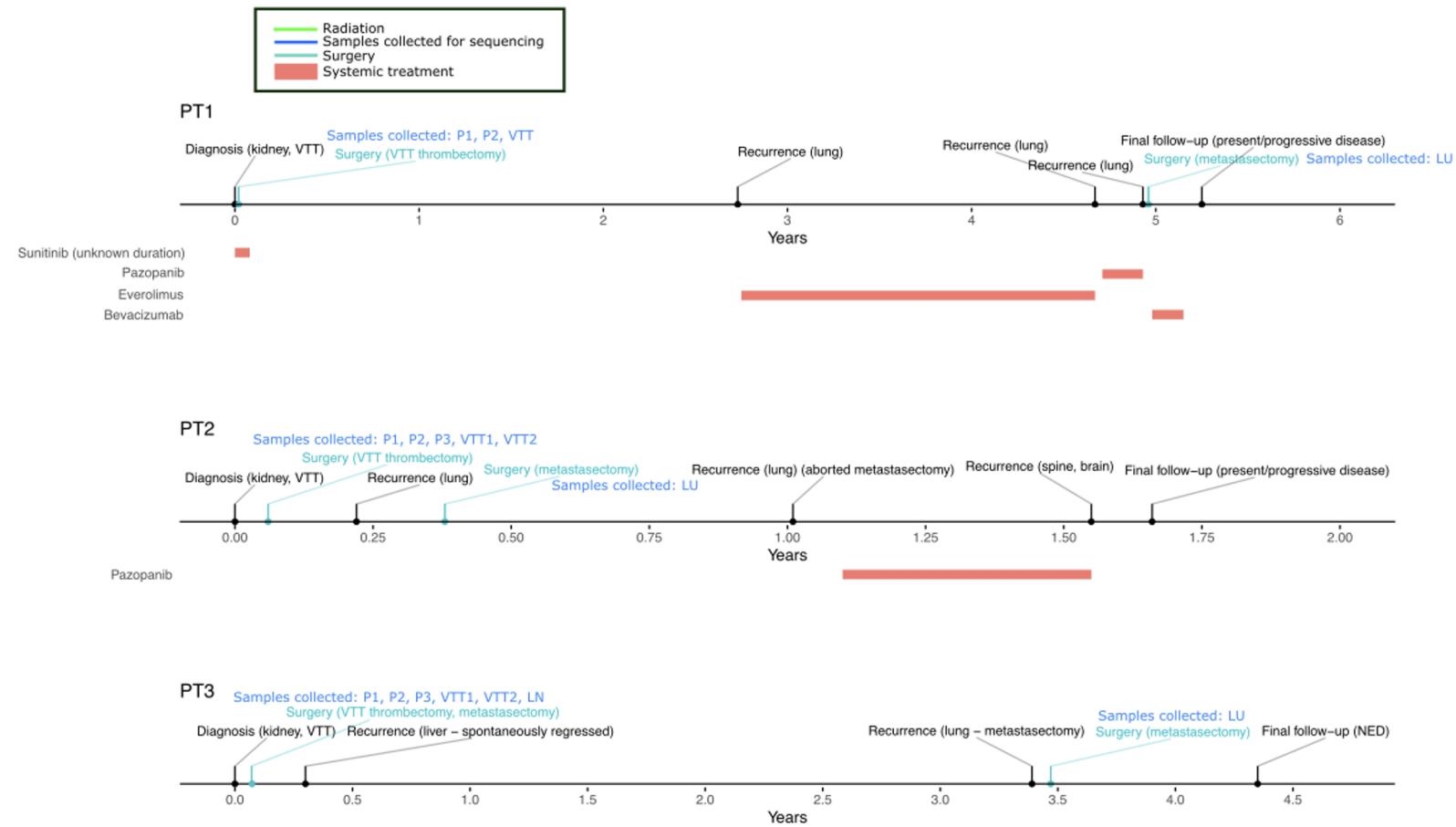
Figure S1.
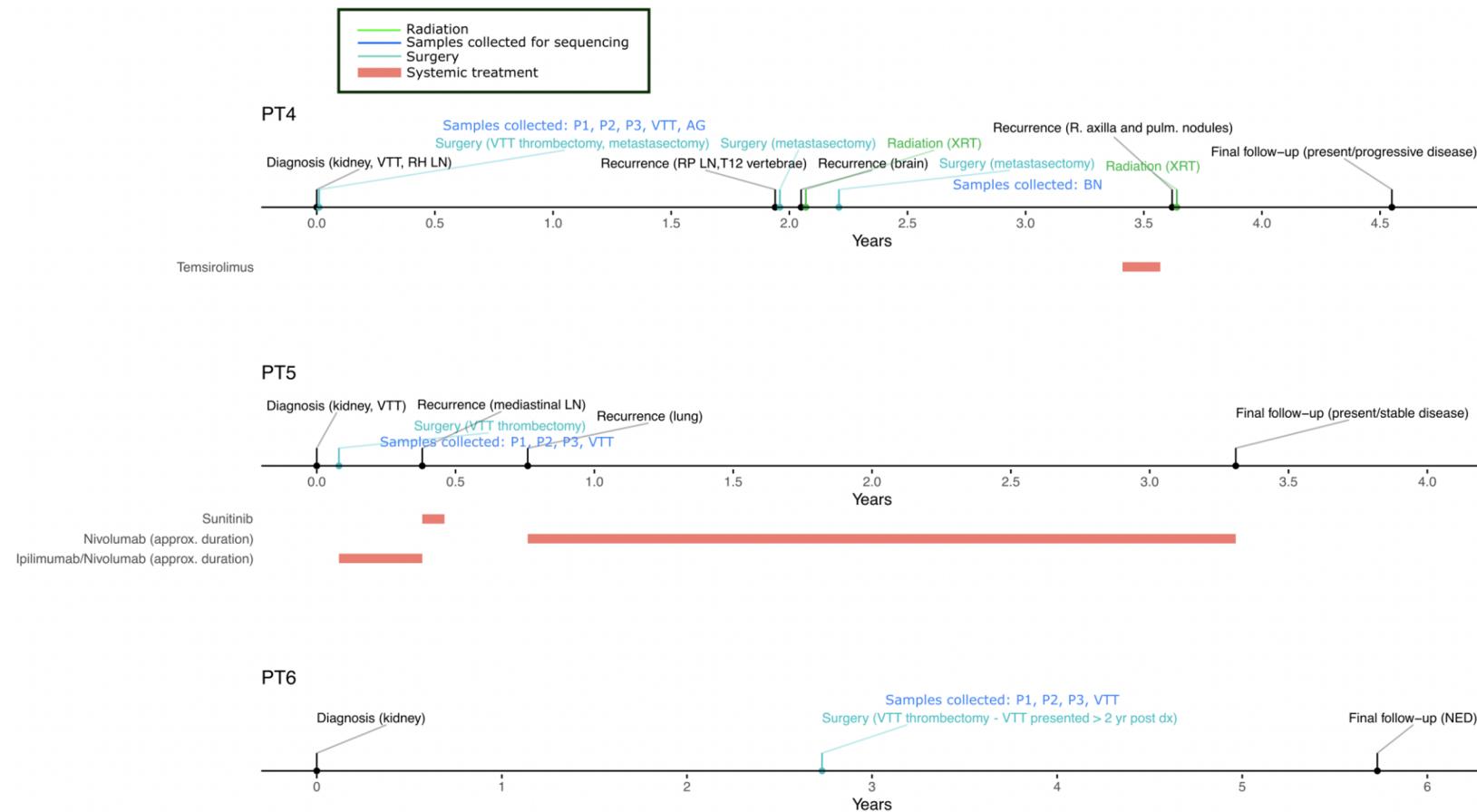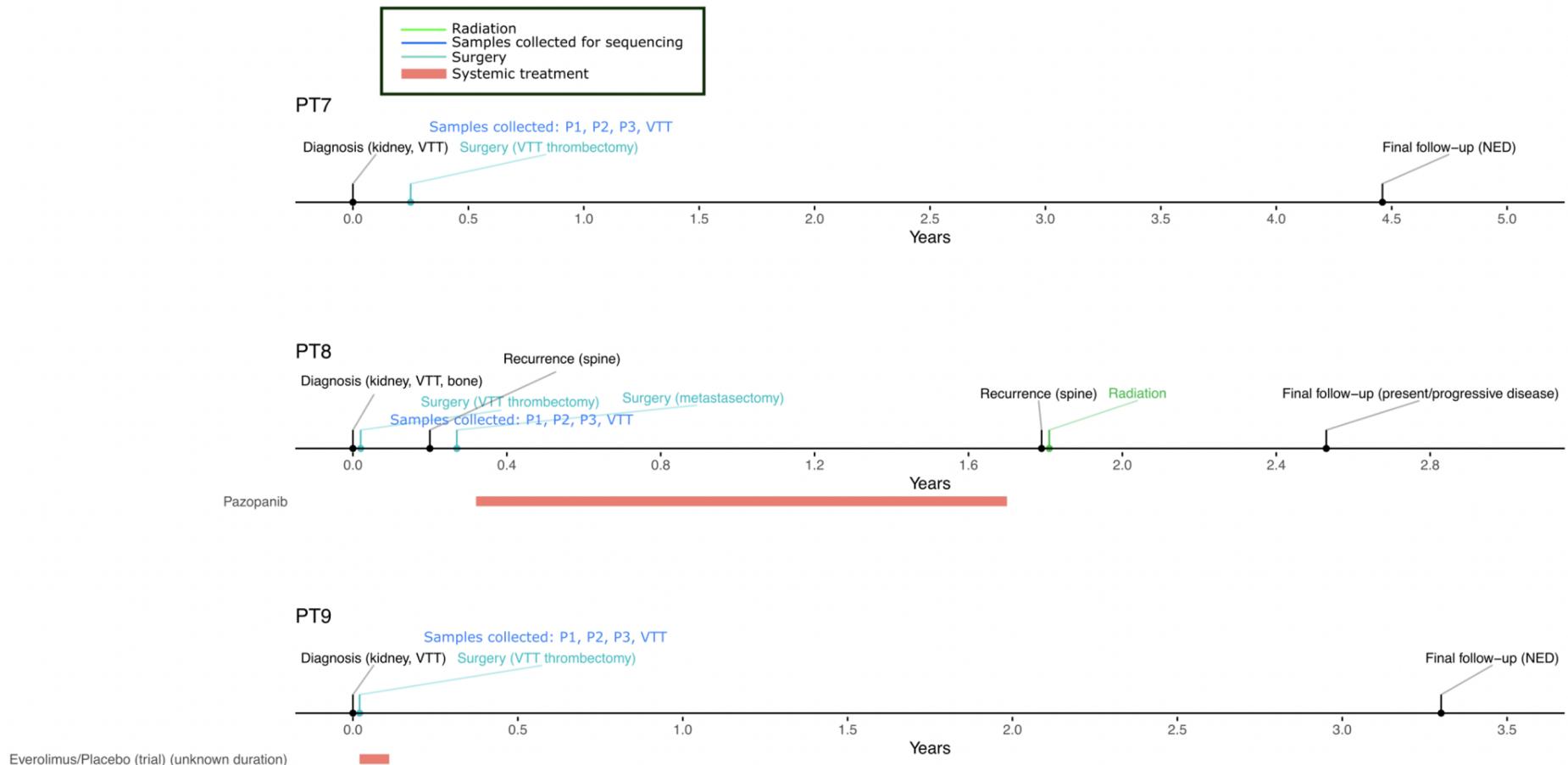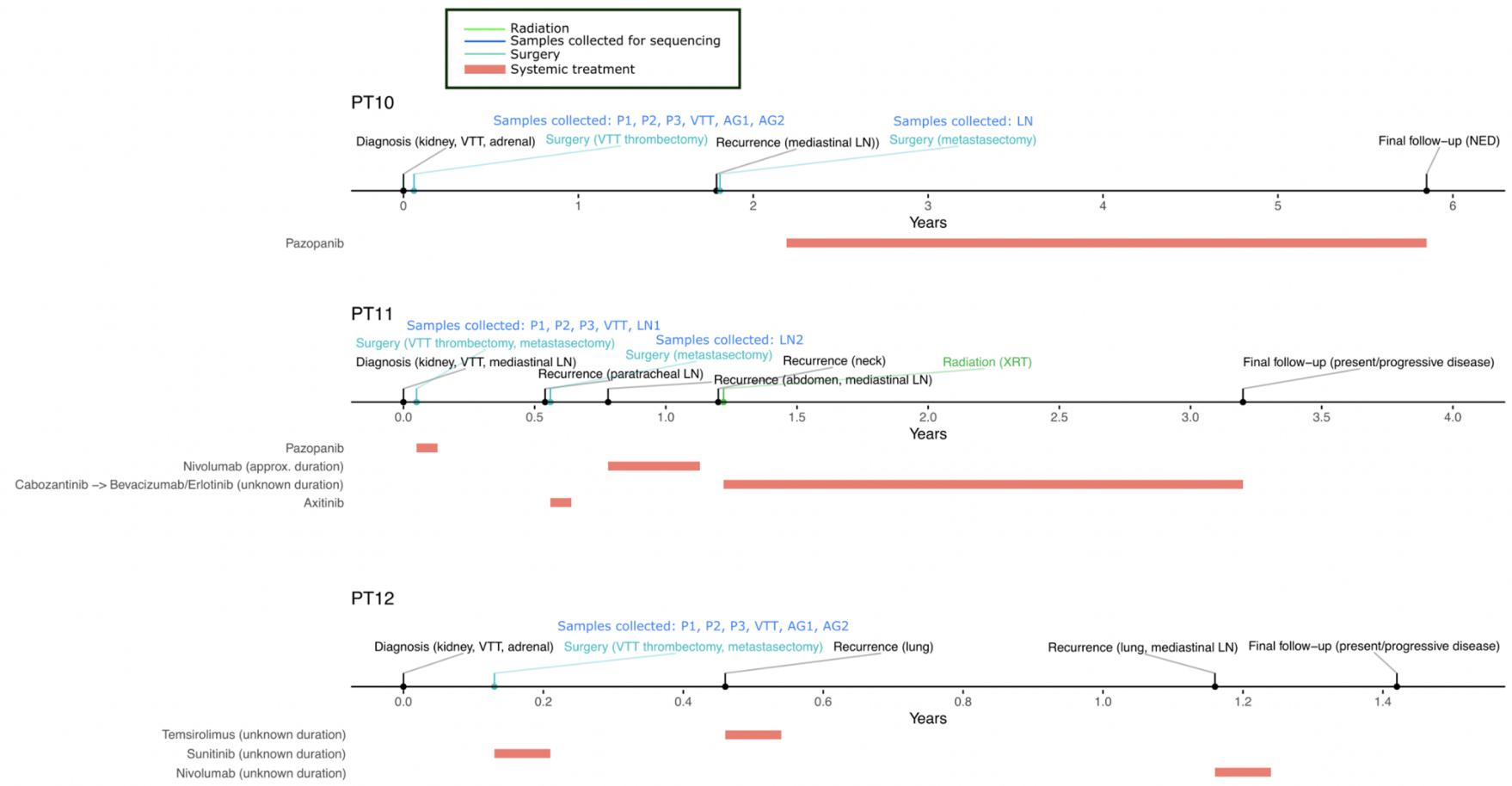
Figure S1.

Figure S1.

Figure S1.

Figure S2

Low

Moderate

High
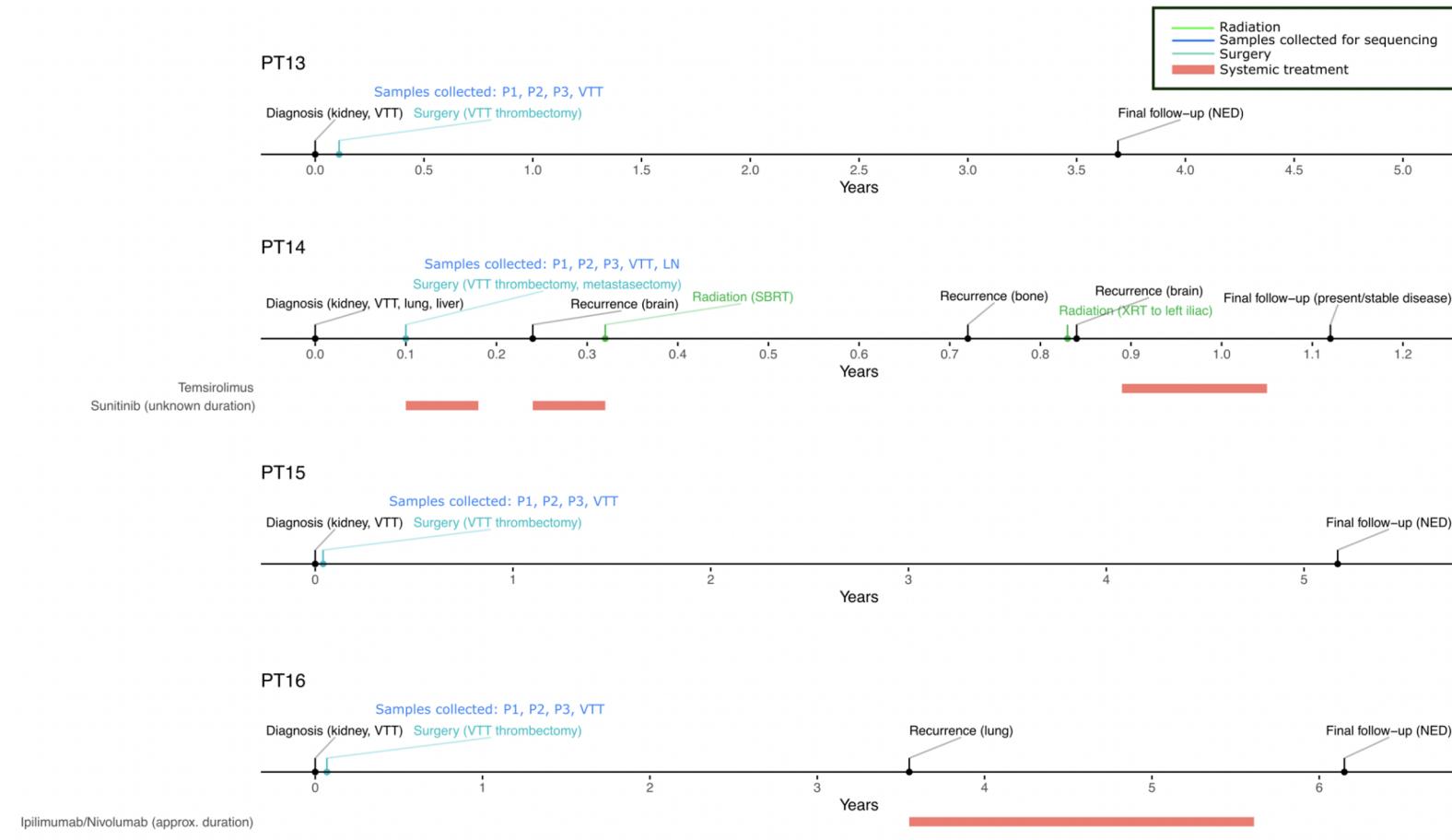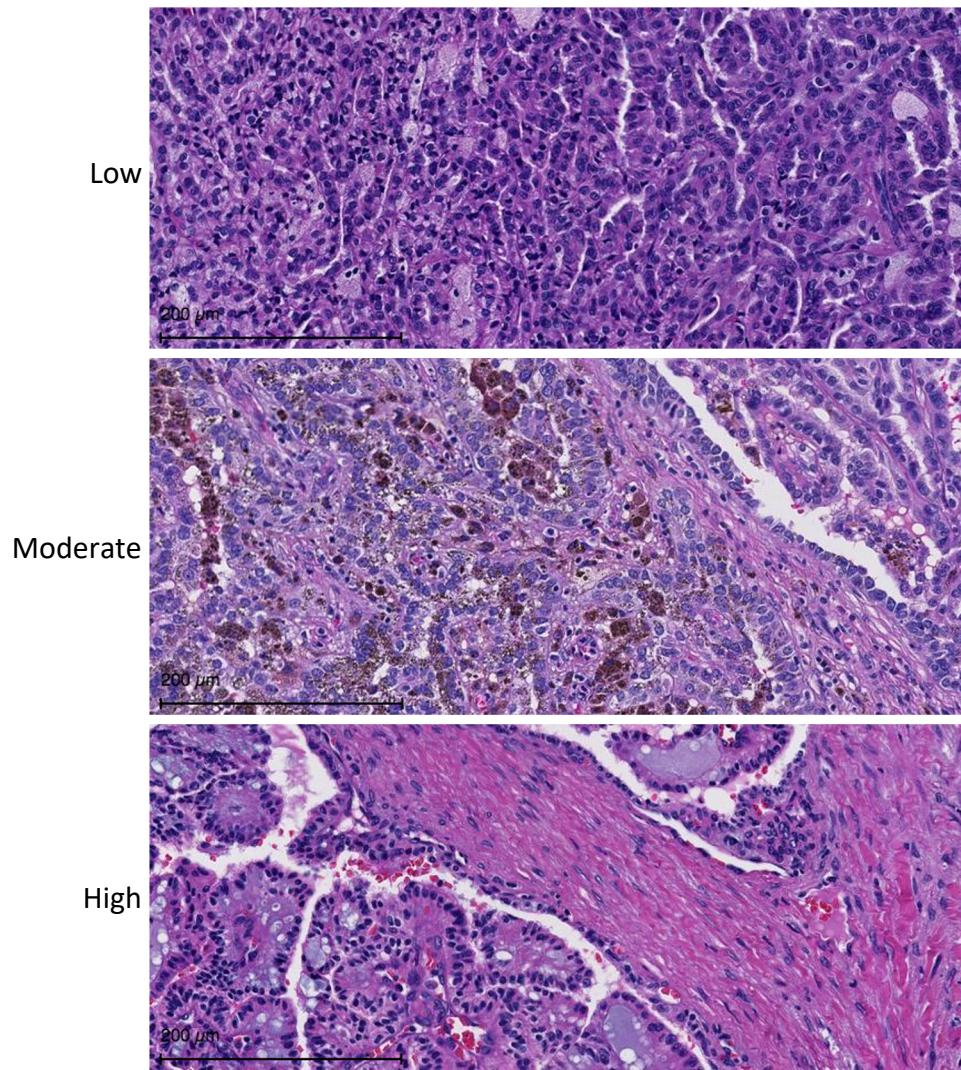
Figure S3.

Figure S4.

Figure S5.

A.

Mutation Rate Model Fit

B.

Mutation Rate Model Fit

Figure S6.

A

MSigDB_Hallmark_2020

	Combined Enrichment Score
TNF-alpha Signaling via NF-kB	
UV Response Up	
Angiogenesis	
KRAS Signaling Up	
Hypoxia	
TGF-beta Signaling	
p53 Pathway	

B

GO_Biological_Process_2018

	Combined Enrichment Score
negative regulation of synapse organization	
negative regulation of macrophage activation	
regulation of brown fat cell differentiation	
negative regulation of cytokine production involved in inflammatory response	
endothelial cell chemotaxis	
regulation of platelet-derived growth factor receptor signaling pathway	
3'-UTR-mediated mRNA destabilization	
negative regulation of leukocyte activation	
fat cell differentiation	
positive regulation of fat cell differentiation	
cellular response to arsenic-containing substance	
positive regulation of cell death	
positive regulation of neuron death	
regulation of neuron death	
positive regulation of chemotaxis	
negative regulation of inflammatory response	
response to lipopolysaccharide	
regulation of endothelial cell proliferation	
regulation of axonogenesis	
response to reactive oxygen species	
response to molecule of bacterial origin	
regulation of tumor necrosis factor production	
cellular response to cytokine stimulus	
positive regulation of epithelial cell proliferation	
positive regulation of transcription from RNA polymerase II promoter	

Figure S7.

Figure S8.

A
B
C
D

Table S1

Case	Age	Sex	Subtype	Fuhrman grade	AJCC stage	Tumor size	VTT level	Metastatic	Metastatic samples collected	Multiple surgical timepoints	Intervening treatment	Data types
PT1	50	m	Clear Cell	3	pT3bN0MX	12.0	1	Yes	Yes	Yes	sunitinib,everolimus,pazopanib	WES
PT2	56	m	Clear Cell	3	pT3bN0	16.7	2	Yes	Yes	Yes		WES
PT3	62	m	Clear Cell	4	pT3cN1	19.0	3	Yes	Yes	Yes		WES
PT4	55	m	Papillary	3	pT3bN2MX	16.0	2	Yes	Yes	Yes	radiation	WES
PT5	61	m	Clear Cell	3	pT3a(m)N0	11.5	1	Yes	No	No		WES+RNAseq
PT6	71	m	Clear Cell	2	pT3aNx	3.3	1	No	No	No		WES+RNAseq
PT7	70	m	Clear Cell	2	pT3aNX	6.2	1	No	No	No		WES+RNAseq
PT8	59	m	Clear Cell	4	pT3aNX	8.1	1	Yes	No	No		WES+RNAseq
PT9	53	m	Clear Cell	2	pT3a	8.0	1	No	No	No		WES+RNAseq
PT10	60	m	Clear Cell	3	pT3bNXM1	9.8	2	Yes	Yes	Yes		WES+RNAseq
PT11	72	m	Papillary	3	pT3bN1	12.1	2	Yes	Yes	Yes	pazopanib	WES+RNAseq
PT12	72	m	Clear Cell	4	pT3aN0M1	11.5	2	Yes	Yes	No		WES+RNAseq
PT13	87	f	Clear Cell	2	pT3aNX	5.5	2	No	No	No		WES+RNAseq
PT14	57	f	Clear Cell	4	pT3bN1MX	13.3	2	Yes	Yes	No		WES+RNAseq
PT15	69	m	Clear Cell	3	pT3aNx	13.0	1	No	No	No		WES+RNAseq
PT16	57	m	Clear Cell	3	pT3aN0	9.3	2	Yes	No	No		WES+RNAseq

Table S2

Patient	Anatomic site	Region type	Tissue abbreviation	Num. mutations	Fraction globally clonal mutations	% Tumor (viable tumor/total tissue area)	% Necrosis in harvested tumor area	RNAseq	WES	WES Harvest Area (mm2)	RNA Harvest Area (mm2)
PT1	Lung tumor	met	LU	172	0.40	NA	NA	No	Yes	NA	NA
PT1	Kidney tumor	primary	P1	114	0.61	NA	NA	No	Yes	NA	NA
PT1	Kidney tumor	primary	P2	126	0.55	NA	NA	No	Yes	NA	NA
PT1	Vein tumor	thrombus	VTT	127	0.54	NA	NA	No	Yes	NA	NA
PT2	Lung tumor	met	LU	490	0.14	NA	NA	No	Yes	NA	NA
PT2	Kidney tumor	primary	P1	582	0.12	NA	NA	No	Yes	NA	NA
PT2	Kidney tumor	primary	P2	365	0.19	NA	NA	No	Yes	NA	NA
PT2	Kidney tumor	primary	P3	315	0.22	NA	NA	No	Yes	NA	NA
PT2	Inferior vena cava	thrombus	VTT1	366	0.19	NA	NA	No	Yes	NA	NA
PT2	Inferior vena cava	thrombus	VTT2	256	0.27	NA	NA	No	Yes	NA	NA
PT3	Lymph node	met	LN	143	0.30	NA	NA	No	Yes	NA	NA
PT3	Lung tumor	met	LU	166	0.26	NA	NA	No	Yes	NA	NA
PT3	Kidney tumor	primary	P1	135	0.32	NA	NA	No	Yes	NA	NA
PT3	Kidney tumor	primary	P2	135	0.32	NA	NA	No	Yes	NA	NA
PT3	Inferior vena cava	thrombus	VTT1	144	0.30	NA	NA	No	Yes	NA	NA
PT3	Inferior vena cava	thrombus	VTT2	156	0.28	NA	NA	No	Yes	NA	NA
PT4	Adrenal tumor	met	AG	159	0.62	NA	NA	No	Yes	NA	NA
PT4	Brain tumor	met	BN	170	0.58	NA	NA	No	Yes	NA	NA
PT4	Inferior vena cava	primary	P1	189	0.52	NA	NA	No	Yes	NA	NA
PT4	Inferior vena cava	primary	P2	177	0.56	NA	NA	No	Yes	NA	NA
PT4	Inferior vena cava	primary	P3	157	0.63	NA	NA	No	Yes	NA	NA
PT4	Vein tumor	thrombus	VTT	193	0.51	NA	NA	No	Yes	NA	NA

Table S2

PT5	Kidney	primary	P1	74	0.53	97	0	Yes	Yes	747.25	1019.4
PT5	Kidney	primary	P2	103	0.38	75	0	Yes	Yes	536.50	652.12
PT5	Kidney	primary	P3	90	0.43	75	0	Yes	Yes	460.01	655.3
PT5	Inferior vena cava	thrombus	VTT	83	0.47	40	0	Yes	Yes	134.50	169.2
PT6	Kidney	primary	P1	127	0.72	80	0	Yes	Yes	813.30	1081.5
PT6	Kidney	primary	P2	123	0.74	85	0	Yes	Yes	479.70	659.91
PT6	Kidney	primary	P3	120	0.76	80	0	Yes	Yes	521.40	691.9
PT6	Inferior vena cava	thrombus	VTT	128	0.71	80	0	Yes	Yes	223.80	297.1
PT7	Kidney	primary	P1	126	0.06	45	0	Yes	Yes	526.50	586.85
PT7	Kidney	primary	P2	129	0.06	85	3	Yes	Yes	707.30	948.8
PT7	Kidney	primary	P3	139	0.06	85	0	Yes	Yes	780.60	776.3
PT7	Inferior vena cava	thrombus	VTT	106	0.08	85	0	No	Yes	41.70	—
PT8	Kidney	primary	P1	131	0.56	95	0	Yes	Yes	713.42	539.29
PT8	Kidney	primary	P2	128	0.58	90	0	Yes	Yes	820.57	462.1
PT8	Kidney	primary	P3	123	0.60	97	0	Yes	Yes	1319.08	1025.6
PT8	Inferior vena cava	thrombus	VTT	132	0.56	90	0	Yes	Yes	425.90	320.2
PT9	Kidney	primary	P1	95	0.48	90	0	Yes	Yes	711.69	526.5
PT9	Kidney	primary	P2	86	0.53	85	0	Yes	Yes	632.35	477.06
PT9	Kidney	primary	P3	98	0.47	95	0	Yes	Yes	771.86	565.09
PT9	Inferior vena cava	thrombus	VTT	116	0.40	75	0	Yes	Yes	185.50	142
PT10	Adrenal gland	met	AG1	141	0.36	30	1	Yes	Yes	116.85	84.4
PT10	Adrenal gland	met	AG2	106	0.48	60	0	Yes	Yes	994.65	713.86
PT10	Inguinal lymph node	met	LN	147	0.35	98	0	Yes	Yes	47.50	42.1
PT10	Kidney	primary	P1	128	0.40	80	0	Yes	Yes	580.10	334.9
PT10	Kidney	primary	P2	157	0.32	80	2	Yes	Yes	721.30	571.1
PT10	Kidney	primary	P3	135	0.38	80	0	Yes	Yes	759.23	440.72
PT10	Inferior vena cava	thrombus	VTT	138	0.37	80	10	Yes	Yes	856.07	645.01

Table S2

PT11	Super retrocaval lymph node	met	LN1	270	0.74	85		0	Yes	Yes	1269.34 914.51
PT11	Paratracheal lymph node	met	LN2	277	0.72	99		0	Yes	Yes	116.86 65.41
PT11	Kidney	primary	P1	268	0.75	50		1	Yes	Yes	567.26 428.36
PT11	Kidney	primary	P2	281	0.71	75		3	Yes	Yes	707.36 549.86
PT11	Kidney	primary	P3	276	0.72	80		2	Yes	Yes	636.05 451.66
PT11	Inferior vena cava	thrombus	VTT	270	0.74	95		5	Yes	Yes	601.14 424.98
PT12	Adrenal gland	met	AG1	129	0.13	85		0	Yes	Yes	801.87 558.88
PT12	Adrenal gland	met	AG2	146	0.12	90		2	Yes	Yes	998.32 789.33
PT12	Kidney	primary	P1	97	0.18	85		0	Yes	Yes	503.32 370.11
PT12	Kidney	primary	P2	88	0.19	75		0	Yes	Yes	304.62 202.05
PT12	Kidney	primary	P3	132	0.13	20		10	Yes	Yes	128.15 100.7
PT12	Inferior vena cava	thrombus	VTT	106	0.16	97		0	Yes	Yes	1074.96 830.19
PT13	Kidney	primary	P1	173	0.81	90		0	Yes	Yes	771.98 535.22
PT13	Kidney	primary	P2	166	0.84	75		0	Yes	Yes	232.70 185.58
PT13	Kidney	primary	P3	175	0.80	85		0	Yes	Yes	301.25 220.4
PT13	Inferior vena cava	thrombus	VTT	174	0.80	80		0	Yes	Yes	497.13 375.84
PT14	Lymph node	met	LN	183	0.37	85		10	Yes	Yes	789.38 590.95
PT14	Kidney	primary	P1	143	0.47	80		1	Yes	Yes	1161.80 866.47
PT14	Kidney	primary	P2	202	0.33	40		5	Yes	Yes	652.79 518.35
PT14	Kidney	primary	P3	178	0.38	75		3	Yes	Yes	874.71 689.63
PT14	Inferior vena cava	thrombus	VTT	199	0.34	97		2	Yes	Yes	1125.65 836.71

Table S2

PT15	Kidney	primary	P1	92	0.55	50	0	Yes	Yes	668.46	472.9
PT15	Kidney	primary	P2	86	0.59	30	0	Yes	Yes	264.00	188.64
PT15	Kidney	primary	P3	106	0.48	75	0	Yes	Yes	1096.29	794.04
PT15	Inferior vena cava	thrombus	VTT	98	0.52	75	0	Yes	Yes	1238.49	929.88
PT16	Kidney	primary	P1	104	0.45	75	0	Yes	Yes	1189.13	887.97
PT16	Kidney	primary	P2	96	0.49	85	0	Yes	Yes	658.98	506.8
PT16	Kidney	primary	P3	91	0.52	75	0	Yes	Yes	740.93	533.5
PT16	Inferior vena cava	thrombus	VTT	94	0.50	85	0	Yes	Yes	322.01	232.84

Table S3

Patient	Total unique mutations	Globally clonal mutations	Mean mutations/sample	Mean mutations/sample/Mb	Mean fraction globally clonal
PT1	249	69	130 ± 26	2.7 ± 0.51	0.52 ± 0.087
PT2	1260	69	400 ± 120	7.9 ± 2.4	0.19 ± 0.054
PT3	306	43	150 ± 12	2.9 ± 0.25	0.3 ± 0.024
PT4	288	99	170 ± 15	3.5 ± 0.3	0.57 ± 0.049
PT5	193	39	88 ± 12	1.8 ± 0.24	0.45 ± 0.062
PT6	181	91	120 ± 3.7	2.5 ± 0.074	0.73 ± 0.022
PT7	380	8	120 ± 14	2.5 ± 0.28	0.065 ± 0.0077
PT8	234	74	130 ± 4	2.6 ± 0.081	0.58 ± 0.018
PT9	207	46	99 ± 13	2 ± 0.25	0.47 ± 0.057
PT10	354	51	140 ± 16	2.7 ± 0.32	0.38 ± 0.05
PT11	438	200	270 ± 5.1	5.5 ± 0.1	0.73 ± 0.014
PT12	396	17	120 ± 23	2.3 ± 0.45	0.15 ± 0.03
PT13	233	140	170 ± 4.1	3.4 ± 0.082	0.81 ± 0.02
PT14	428	67	180 ± 24	3.6 ± 0.47	0.38 ± 0.055
PT15	180	51	96 ± 8.5	1.9 ± 0.17	0.54 ± 0.048
PT16	198	47	96 ± 5.6	1.9 ± 0.11	0.49 ± 0.027

Table S4.

symbol	patients
MUC6	PT11,PT15,PT6
TTN	PT12,PT2,PT7
CFAP54	PT5,PT8
COL4A3	PT9,PT7
EIF3FP1	PT3,PT11
FHAD1	PT7,PT13
FRG1DP	PT8,PT5
GOLGA6L2	PT3,PT15
GRIK4	PT9,PT11
IGFN1	PT16,PT11
KRTAP5-8	PT15,PT3
MUC4	PT7,PT3
NBPF12	PT2,PT3
PCDH18	PT3,PT12
PCDHGA3	PT7,PT14
PRB4	PT9,PT8
SETD2	PT6,PT9
SOX9	PT13,PT12
TSPEAR	PT4,PT2
TYW1	PT2,PT3

Table S5.

Patient	Time from initial metastasis emergence to diagnosis (years - lower bound)	Time from initial metastasis emergence to diagnosis (years - fit)	Time from initial metastasis emergence to diagnosis (years - upper bound)	Time from initial metastasis emergence to VTT emergence (years - lower bound)	Time from initial metastasis emergence to VTT emergence (years - fit)	Time from initial metastasis emergence to VTT emergence (years - upper bound)
PT10	27.8	30.1	32.4	11.2	17.6	23.5
PT3	19.9	22.9	26.0	1.0	8.1	15.2
PT1	12.1	16.3	20.6	-6.4	1.6	9.9
PT12	10.2	14.5	18.6	-39.3	-33.2	-27.4
PT14	9.3	13.8	18.3	-9.2	0.1	8.8
PT4	8.9	12.9	16.6	-3.0	5.3	13.9
PT11	7.2	11.6	16.0	-6.9	2.6	11.9