Title: A multi-layered systems approach for renal cell carcinoma

Lindsay S. Cooley[1,2]*, Justine Rudewicz[1,2,3]*, Wilfried Souleyreau[1,2]*, Kim Clarke[4], Francesco Falciani[4], Maeva Dufies[5,6], Stephanie Verbeke [1,2], Andrea Emanuelli [1,2], Sebastien Benzekry[7]. Cécile Taing[7], Emeline Julie Ribot[8], Sylvain Miraux[8], Diether Lambrechts[9], Elodie Modave[9], Raphael Pineau[10], Marie-Alix Derieppe[10], Damien Ambrosetti[11], Jean-Christophe Bernhard[12], Alain Ravaud[13], Sylvie Négrier[14], Jean-Marc Ferrero[15], Gilles Pagès[5,6], Macha Nikolski[3,16], Andreas Bikfalvi[1,2]

1. University of Bordeaux, LAMC, Pessac, France
2. INSERM U1029, Pessac, France
3. Bordeaux Bioinformatics Center, CBiB, University of Bordeaux, France
4. Institute of Integrative Biology, University Liverpool, UK
5. Centre Scientifique de Monaco, Biomedical Department, Principality of Monaco
6. University Côte d'Azur, Institute for Research on Cancer and Aging of Nice (IRCAN), CNRS UMR 7284; INSERM U1081, Centre Antoine Lacassagne, France
7. MONC, Centre de Recherche Inria, Bordeaux, France
8. Centre de Résonance Magnétique des Systèmes Biologiques (RMSB),UMR 5536 CNRS, Bordeaux
9. VIB-KU Leuven Center for Cancer Biology, Leuven, Belgium
10. University of Bordeaux, «Service Commun des Animaleries», Bordeaux, France
11. Centre Hospitalier Universitaire (CHU) de Nice, Hôpital Pasteur, Central laboratory of Pathology, France
12. Centre Hospitalier Universitaire (CHU) de Bordeaux, service d'urologie
13. Centre Hospitalier Universitaire (CHU) de Bordeaux, service d'oncologie médicale
14. Université de Lyon, Centre Léon Bérard, Lyon, France
15. Centre Antoine Lacassagne, Clinical Research Department, Nice, France
16. University of Bordeaux, IBGC, Bordeaux, France

* These authors have equally contributed to this work

Key words: Renal cell carcinoma, tumor model, metastasis, systems biology approach
Running Title: Renal cell carcinoma model

Correspondence: Andreas Bikfalvi MD PhD
LAMC INSERM U1029
Université Bordeaux
Allée Geoffroy St Hilaire
33615 Pessac
France
E-mail: andreas.bikfalvi@u-bordeaux.fr

Acknowledgments: This work was supported by grants from the PlanCancer («Systems Biology of Renal Cell Carcinoma using an Experimental RCC model» (C18005GS, SystemsRCC)) and from the university Bordeaux (G2P project) to AB. LC and AE were supported by post-doctoral fellowships from the Region "Nouvelle Aquitaine". The Crispr/cas9 construct was gratefully provided by Michel Tremblay (Goodman Cancer Center, McGill, University). The authors would like to thank the lentiviral production plateform Vect'UB for EGFP lentivirus, the "service commun des animaleries" for animal production and housing, and the Bordeaux Imaging Center (BIC).

Abstract

In order to discover molecular pathways and players in renal cancer development and metastasis, we developed a mouse model to generate sequentially more aggressive and specialized cell lines. Multiple cell lines for primary tumor growth, survival in the blood circulation and lung metastasis or metastatic spread from the primary tumor were generated and analyzed using a multi-layered approach which includes large-scale transcriptome, genome and methylome analyses. Transcriptome and methylome analyses demonstrated distinct clustering in three different groups. Remarkably, DNA sequencing did not show significant genomic variations in the different groups which indicates absence of clonal selection during the *in vivo* amplification process. Transcriptome analysis revealed several markers and signatures of tumor aggressiveness which were validated, at the mRNA and protein level, in patient cohorts from TCGA, local biobanks and clinical trials. This also includes soluble markers. In particular, SAA2 and CFB were highly predictive for survival and tumor progression. Methylome analysis of full-length DNA allowed clustering of the same groups and revealed clinically predictive signatures. We also uncovered IL34 as a key regulator of renal cell carcinoma (RCC) progression which was also functionally validated *in vivo*, and a mathematical model of IL34-dependent primary tumor growth and metastasis development was provided. These results indicate that such multilayered analysis in a RCC animal model leads to meaningful results that are of translational significance.

INTRODUCTION

Renal Cell Carcinoma (RCC) encompasses a heterogeneous group of cancers derived from renal tubular epithelial cells, including multiple histological and molecular sub-types, of which clear cell RCC (ccRCC) is the most common[1]. The incidence and prevalence of RCC are rising[2,3], along with increases in related risk factors such as hypertension, diabetes and obesity[4]. Gandaglia et al[5] reported a continuing upward trend in both incidence and mortality even in patients with localized disease. When disease is localized to the kidney, surgical resection is the preferred option and the prognosis is favorable. However, therapeutic options for metastatic disease are limited. ccRCC metastasizes primarily to the lungs (secondarily to liver and bone), and 5-year survival ranges 10-12 %[6–8]. Furthermore, 40% of patients with seemingly localized disease later relapse with localized or metastatic disease. Localized recurrence is also difficult to treat and difficult to predict, and has a poor prognosis[9,10].

Challenges associated with treatment of RCC include high levels of resistance to traditional chemotherapeutic drugs[4]. The majority of currently available targeted therapies focus on inhibiting angiogenesis driven by the VEGF/VEGFR axis[11]. More recently, immunotherapeutic agents (nivolumab and ipilimimab) have also been added as first line treatment[12,13]. However, such therapies are rarely curative, and eventual drug resistance is almost inevitable. Furthermore, clinical treatment of RCC is hampered by a lack of relevant biomarkers. Currently, no fully validated molecular biomarkers for RCC are used in clinical practice. Response to available treatments and long-term disease-free survival is highly variable and unpredictable. Patient diagnosis, prognosis and clinical decisions are currently based on histological information such as Fuhrman grade and tumor stage, and are included in the SSIGN or UISS classification for primary tumor risk of recurrence[14–17], and in the MSKCC and/or IMDC classification for metastatic disease [18–20]. In addition, therapy selection

is based on limited guidelines and response to previous treatments. In this respect, clinical treatment of RCC lags behind other cancers for which molecular knowledge is invaluable in guiding clinical decisions (e.g. hormone receptor status in breast cancer).

Tumor progression from initiation to full metastasis is a multi-step process and occurs via a series of overlapping stages. Tumor progression can be seen as an evolutionary process whereby tumor cells within the primary tumor adapt to varied selection pressures. Cells must detach from the primary tumor, gain access to and survive in the circulation, exit the vasculature, and survive and proliferate in the environment of the secondary organ. Thus, different mechanisms come into play at different stages, and overall tumor progression is the sum of these processes. A better understanding of the molecular changes that enable cancer cells to overcome the barriers imposed during the metastatic process could aid in the diagnosis, prevention and treatment of metastatic cancer.

The process of serial implantation of a cell line *in vivo* is a known method to render a cell line increasingly aggressive[21,22]. In this study, we leveraged this technique to generate increasingly aggressive sub-lines of RENCA cells, a mouse renal carcinoma cell line. To this purpose, we used different implantation strategies to dissect the different stages of tumor progression and metastasis. Cell-line derived RNA and DNA were used to analyze transcriptional, DNA and methylome profiles of the cell types, and to generate signatures able to predict global, disease-free and progression-free survival time in human patients. Furthermore, we identified and investigated a number of genes with the potential to be biomarkers and therapeutic targets and characterized interleukin-34 (IL34) as an important universal driver in RCC.

RESULTS

Generation and characterization of the mouse model

We used a syngeneic (and thus immunocompetent) orthotopic murine model in which the mouse renal carcinoma cell line RENCA[23], expressing GFP for visualization purposes, were implanted orthotopically under the kidney capsule in female BALB/c mice. Under these conditions, metastasis occurred specifically to the lungs after 14-16 days, with ethical sacrifice at 26 days following implantation (Fig. 1a). We adapted this model to produce new sub-cell lines with progressively enhanced aggressiveness and metastatic potential with differing characteristics. We serially passaged the cell lines using different implantation strategies designed to replicate different aspects of tumor development and metastasis (Fig. 1a). After each passage, primary tumor cells were purified from organ explant cultures to generate new sub-cell lines. The three injections modes were as follows:

(1) Orthotopic injection under the renal capsule leading to formation of a primary tumor (“Kidney” group). The primary tumor was explanted and cell lines purified for re-implantation into the kidney. The “Kidney” cell lines thus derived have undergone repeated cycles of primary tumor formation and are expected to reveal enhancement of mechanisms relevant to primary tumor growth and evolution.

(2) Intravenous injection (“Tail vein” group). Tumor cells were injected directly into the blood stream leading to formation of tumors in the lungs in the absence of a tumor in the kidney. This injection mode is expected to recapitulate key aspects of metastasis such as survival in the blood stream, evasion of host immune response, colonization and growth in the lungs.

(3) Orthotopic injection under the renal capsule followed by metastasis to the lungs (“Lung” group). Tumor cells were purified from the lung metastases and re-implanted under the

kidney capsule for the subsequent passage. These “Lung” cell lines have undergone the full cycle of primary tumor formation, dissemination in the blood, and secondary tumor formation at a distant site, and were, thus, enhanced for all steps required for tumor progression.

The sub-cell lines derived as described above were then sequentially passaged 6 times *in vivo* using multiple mice per injection mode and per passage to finally give a total of 67 newly derived sub-cell lines. The generated tumors became increasingly aggressive with each passage, with a reduced survival time for the mice of all groups (Fig. 1b), from 26 to 15 days after 6 *in vivo* cycles. Furthermore, we assessed primary tumor and lung tissue at Passage 6 (Fig. 1c, d). “Kidney” cell lines generated primary tumors that were significantly larger than tumors generated with either parental or “Lung” cell lines (Fig. 1c, d). The lungs of mice with orthotopic implant of parental cells showed no macroscopic metastases, whereas visible metastases were detected in 55% of “Kidney” and 77% of “Lung” cell implanted mice (Fig. 1d, e) indicating that metastatic ability was enhanced in passaged cell lines. In the intravenous injection experimental metastasis model (i.e. “Tail” cell line), the number and size of metastases were drastically increased (Fig. 1f), as a consequence of enhanced aggressiveness and specialization of serial passaged cells compared to the parental cells. Thus, serial passaging resulted in generation of cell lines that were more aggressive and specialized when compared to control cells.

Functional genomics of amplified mouse cell lines

Transcriptomic analysis

We performed full genome transcriptomic analysis of the 67 cell lines. The P1 cell line was excluded from the analysis because of insufficient number of animals (<3). Therefore, data acquired from the different lines were labeled S0 to S5 (S0 being the parental cells and S1 to S5 representing P2 to P6). We used Principal Component Analysis (PCA) in order to

summarize the information contained in our data sets for sample series S2 and S5. Principal Component Analyses showed that, while the subtypes had similar expression profiles in the early passages, the transcriptional profiles became different between “Kidney” (K), “Tail” (T), and “Lung” (L) groups in later passages and, thus, clustered into distinct groups (Fig. 2a, b). Tight clustering of the biological replicates (one mouse per data point) also showed that the profiles were stable. When all genes that were major contributors to the Principal Component 1 (PC1) and 2 (PC2) were pooled, the heat map of the transcriptomic profiles revealed a gradual change in expression (Fig. 2c).

Enrichment analysis of these most contributing genes (see Methods for details) demonstrated several highly enriched categories. These include the following GO terms: extracellular matrix, angiogenesis, cell proliferation, cell adhesion, cell migration, immune process and inflammation and apoptosis. Fig. 2d, e and f represent enriched GO terms in common for PC1 and PC2 or specific to PC1 or PC2, respectively.

Next, we investigated whether transcriptional signatures derived from the differentially expressed genes in the K, T and L groups could predict outcome for patients using the Clear Cell Renal Cell Carcinoma dataset (KIRC) from The Cancer Genome Atlas (TCGA).

The general strategy is outlined in Supplementary Fig. 1. We compared genes that changed their expression between the parental and S5 cell line and named these genes as progressively regulated (up or down) (Supplementary Table 1). We included in our analysis only genes having their expression consistently increasing in the different series.

The analysis has been performed by comparison to the parental cell line for K, T and L groups alone. The additional group “Kidney-Tail-Lung” (KTL) is represented by pooling the different groups in a unique group. The KTL group is also compared to the parental cell line

(see Methods for details). The Supplementary Table 1 lists genes differentially expressed in each group (K, T, L, KTL).

A total of 131 genes for all groups satisfied the inclusion criteria. Among these, 26 were in the K, 45 in the T, 59 in the L and 31 in the KTL group (Supplementary Table 1). Among those, 123 genes were upregulated and only a small fraction was downregulated.

We next investigated the predictive value of these 4 gene signatures using TCGA (KIRC cohort). Since only 8 genes were downregulated, we chose only to include the upregulated genes in the validation of our signatures. Supplementary Table 2 depicts the list of genes for each group. For each gene, we fitted a Cox-proportional hazard regression model based on overall survival (OS) or disease-free survival (DFS). A gene was conserved if its false-discovery rate (FDR) adjusted p-value of its log-rank test was lower than 0.01 and if the hazard ratio was in agreement with the differential expression. We have further created a 5th "All Merged" signature composed of genes that are members of all of the previously defined 4 signatures.

Table 1 depicts the signatures for each group (K, T, L, KTL) and their significance for OS and DFS. In Fig. 3 the results of "Lung" subgroup are depicted. We analyzed OS (Fig. 3a) of all patients and DFS (Fig. 3b) of M0 patients. We validated our signatures by computing an empirical p-value and by testing our signature against 1000 random signatures of equivalent size (Fig. 3c, d). Furthermore, we performed multivariate Cox regression analysis of our signature (Fig. 3e, f). After adjusting for clinical variables (TNM stage and Fuhrman grade), the Lung signature remained an independent prognostic factor for predicting both OS and DFS. The results for the other signatures are shown in Supplementary Fig. 2-5. The Lung signature remains the most significant for OS and DFS when compared to the others, except for the "All Merged" signature. The performances of the signatures were the following for the

different groups when ranked according to the hazard ratio: K<T<KTL<L<All Merged for both OS and DFS.

The identification of signatures predictive of patient outcome also validates our experimental approach and shows that the strategy of generation of increasingly specialized mouse cell lines revealed novel genes and signatures with relevance to human RCC.

DNA sequencing analysis

We subjected both the parental RCC line and those that were passaged 6 times (series S5) to low-coverage whole-genome sequencing to assess whether copy number variability could possibly underlie the change in phenotype (Supplementary Fig. 6). We failed however to observe significant differences in copy numbers between parental and passaged samples, both at the level of the number of breakpoints detected (45 for parental versus 41.75±6.32 for passaged lines) and the percentage of the genome with a copy number different than 2 (19% for parental versus 18%±1% for passaged lines). Visually, we also failed to observe new copy number events. When using ABSOLUTE to estimate purity (proportion of cancer cells present in the sample) and mean ploïdy of the samples, no differences were noted (purity and ploïdy in parental line were of 63% and 2.69, respectively versus 64%±2% and 2.65±3% for passaged lines). This suggests the altered profile of the “Kidney”, “Tail” and “Lung” lines was not due to clonal selection.

Methylome sequencing analysis

Since DNA sequencing did not provide conclusive evidence for clonal selection, we conducted a methylome analysis by full methylome sequencing of the S5 series samples for

the K, T and L groups as well as of the parental control cells. Similarly to the transcriptional data, Principal Component Analysis (PCA) showed clustering of the global methylome in 4 clusters corresponding to the K, T, L and parental cell line samples (Fig. 4a). Enrichment analysis demonstrated several highly enriched categories (Fig. 4b). These include cell cytokine activity, nervous system development, regulation of cell population/proliferation, extracellular matrix, transcription and gene regulation, angiogenesis, regulation of gene expression among the most significantly enriched categories.

We focused on CpGs that were located in the gene body with methylation equal or higher than 20 % and with a q-value of 0.05. We found 33 095 CpGs differentially methylated, out of which 9 731 were associated with a known gene. They were confronted with the transcriptome results (Fig. 4c). Differentially expressed and methylated genes were included in this analysis if they contained at least one differentially methylated CpG in their gene-body. Several genes contained in the transcriptome signatures were found in this analysis. For progressive signatures (Supplemental Table 1): *Gm2012* (K), *B3galt1* (T), *Krt20* (T), *Thbd* (T), *Adam22* (L), *Apbb1ip* (L), *Arap3* (L), *Wscd2* (L). For final signatures (Table 1 and Supplemental Table 2, corresponding to red dots): *IL34* (KTL), *Rik/MAGEC2/3* (T), *Krt17* (L), *Samsn1* (L).

We then investigated the clinical relevance of CpG methylation in the gene body of these genes in the KIRC TCGA cohort. Two CpGs (cg01782798 and cg26831220) in the gene body of *IL34* were predictive for OS (Supplementary Fig. 7). In our data sets, the differentially methylated cytosine (DMC) was located in intron 1 of *IL34* (Supplementary Fig. 7a), as were cg01782798 and cg26831220 (Supplementary Fig. 7b). According to ENSEMBL for the GRCm38 mouse genome and GeneHancer for the GRCh37 Human genome, these three CpGs are part of enhancer regions. Correlations of *IL34* transcription levels and cg01782798

or cg26831220 are depicted in Supplementary Fig. 7c and d. Kaplan-Meier and log-rank test for OS were done by separating samples into 2 groups by the median value for each CpG resulting in a p-value of 1.62e-05 and 0.00057 and a Hazard Ratio (HR) of 2.422 (1.651-3.551 95% CI) and 1.970 (1.342-2.894 95% CI) for CpG cg01782798 and cg26831220, respectively (Supplementary Fig. 7e-g). In the same manner as for the transcriptomics data, the clinical relevance and prognostic performance of the 2 CpGs were tested against 1000 random 1-CpG signatures resulting in empirical p-values of 0.015 and 0.033 for CpG cg01782798 and cg26831220, respectively (Supplementary Fig. 7g).

Since the differential analyses of methylation presents only few similarities with the transcriptomic results, we checked whether these modifications/dis-regulations are related to the same processes. To this aim, we compared enrichment results (GO terms) obtained in the transcriptome and methylome analyses (Fig. 4d). This comparison revealed that in both data sets, a limited number of GO terms are shared (26 GO terms in common of a total of 160 for the transcriptome and 112 for the methylome). This suggests that gene expression and methylation analysis converge only partially to a common set of processes (26/160 = 16% and 26/112=23%).

Novel RCC biomarkers and clinical relevance

We used the molecular data from the mouse model to select novel potential diagnostic/prognostic/therapeutic targets. Potential targets were chosen in terms of their possible relevance in different aspects of tumor progression, their upregulated expression in the transcriptomic data (in one or multiple cell subgroups), and their correlation with reduced OS/PFS in the TCGA KIRC dataset. Some of the genes (*SAA2*, *CFB*) selected using this approach, were included in the signatures above determined with our stringent method.

Others (*GRPC5a*, *PODNL1*) could not be included in the signatures, because they did not strictly full-fill the stringency criteria applied for signature determination. We also obtained tissue samples and clinical data from the UroCCR Kidney Cancer tissue bank in Bordeaux as a confirmation cohort for the selected genes.

Serum Amyloid A2 (SAA2)

SAA2 is an acute phase protein related to SAA1, which was previously linked to metastasis[24]. Its expression was strongly upregulated with passage in the Lung cell lines (Supplementary Fig. 8a). *In silico* analysis of the TCGA KIRC database *SAA2* was a very strong predictor of OS (Supplementary Fig. 8b) and DFS (Supplementary Fig. 8c). Furthermore, the analysis was also done for the M0 and M1 subgroups (Supplementary Fig. 8d-g). Analysis of the UroCCR patient cohort confirmed the effect on OS and DFS (Supplementary Fig. 8h, i). Tumors from patients with the highest Fuhrman Grade, had a significantly increased SAA2 expression compared to all other grades (Supplementary Fig. 8j). We used grade-matched plasma samples from patients with and without metastases, collected before primary tumor surgery (Supplementary Fig. 8k). Patients with metastases had higher plasma levels of SAA2. When patients were divided into two groups of equivalent size, the group with higher SAA2 levels had a significantly shorter DFS (Supplementary Fig. 8l). A second set of plasma samples, collected in the weeks following surgery for removal of the primary tumor, was tested for SAA2 (Supplementary Fig. 8m). In this case, patients with higher expression had shorter OS. Hence, circulating SAA2 levels appear as an indicator of metastatic progression that deserves to be evaluated at diagnosis.

We next used plasma samples from metastatic patients before receiving a first cycle of sunitinib or bevacizumab (SUVEGIL and TORAVA clinical trials). Patients treated with

sunitinib only and stratified according to low and high SAA2 levels, had a spectacular better OS and progression-free survival (PFS) when belonging to the SAA2 low group (cut-off of 269 µg/ml) (Fig. 5 a, b). When patients treated with sunitinib and bevacizumab were analyzed together, the PFS was of limited significance (borderline p-value of 0.0507) (Fig. 5c). The median of PFS for SAA2high patients was of 5.35 month versus 16.17 month for the SAA2low group. Thus, determining SAA2 plasma levels could be a useful measure for deciding a treatment strategy in RCC.

Complement factor-B (CFB)

CFB was most strongly upregulated in the “Lung” and to a lesser extent in the “Tail” group, both considered to recapitulate features of metastasis (Supplementary Fig. 9a, b). TCGA analysis in ccRCC showed that *CFB* expression is correlated in primary tumors with shortened DFS and OS (Supplementary Fig. 9c, d). We also performed the analysis in the M0 and M1 subgroups (Supplementary Fig. 9e-h). Using samples and data from the UroCCR cohort, we demonstrated that *CFB* was overexpressed in the tumor tissue versus the adjacent kidney at the mRNA level (Supplementary Fig. 9i), and that increased expression correlated with reduced DFS and OS, consistent with the results obtained with the TCGA cohort (Supplementary Fig. 9j, k). As for SAA2, CFB can be measured in the blood. For this purpose, we used UroCCR plasma samples collected from patients either before surgery (primary tumor intact) or in the following weeks after surgery (no primary tumor present but metastases *in situ* possible). Before surgery, a trend was observed without reaching significance whereas after surgery patients with metastases had higher plasma CFB levels compared to patients without metastases (Supplementary Fig. 9l, m). This suggests that circulating CFB measurement may be useful as a blood-born marker of metastasis in the

follow-up after surgical tumor removal. As for SAA2, CFB plasma levels were tested in patients with metastases before the first cycle treatment with sunitinib or bevacizumab (SUVEGIL and TORAVA clinical trials). Patients whose levels were high (cut-off 310 µg/ml) had faster disease progression compared to patients whose levels were low (high CFB, 3.58 month; low CFB, 18.7month, p=0.0004) (Fig. 5d).

We then grouped the significance of testing SAA2 and CFB plasmatic levels (Fig.5 e, f). Three different groups with different survival can be identified: group 1 (CFB low+SAA2 low, PFS: 19.37 months, OS: NR), group 2 (CFB high SAA2 low or CFB low SAA2 high, PFS: 9.87 months, OS: 20.9 months), group 3 (CFB high Saa2 high, PFS: 2.8 month, OS: 8.33 months). Group 1 had the best survival rate while group 3 had the worst. Group 2 had intermediate survival outcome. Thus, the combined analysis of these two markers is a powerful predictor of patient outcome following anti-angiogenic treatment with sunitinib or bevacizumab.

Podocan Like Protein-1 (PODNL1)

PODNL1 is a member of the small leucine-rich proteoglycan (SLRP) family of 17 genes. It is secreted extracellularly and its function is currently unknown. High expression has previously been linked with poor outcome in ovarian cancer[25]and glioblastoma[26]. *PODNL1* expression was upregulated in our mouse cell lines in the “Kidney” subgroup (Supplementary Fig. 10a), although the increase was relatively modest. However, this gene showed a very strong link with reduced DFS and OS in the TCGA KIRC database (Supplementary Fig. 10b, c). We have also performed this analysis in M0 and M1 patients (Supplementary Fig. 10d-g). When using UroCCR samples, *PODNL1* was overexpressed at the mRNA level in the tumor versus healthy tissue (Supplementary Fig. 10h). In the UroCCR biobank, DFS also showed different

trends depending on *PODNL1* expression albeit statistically not significant, the latter was also the case for OS (Fig. 10i, j). This largely unknown and interesting gene may play a key role in RCC and further studies are required to investigate this possibility.

G Protein-Coupled Receptor Class C Group 5 Member A (GPRC5a)

GPRC5a is involved in epithelial cancers[27] and is proposed to be an emerging biomarker in several human cancers[28]. It was upregulated in all 3 conditions (K, L, T) (Supplementary Fig. 11a). In the TCGA, *GPRC5a* expression was correlated with shortened OS and DFS (Supplementary Fig. 11b, c). This analysis was also conducted in M0 and M1 patient subgroups (Supplementary Fig. 11d-g). Using UroCCR samples, we confirmed the link with DFS but not with OS (Supplementary Fig. 11h, i). Furthermore, Fuhrman grade 4 samples showed higher expression compared to the lower grades (Supplementary Fig. 11j).

Expression, clinical relevance and functional analysis of Interleukin-34

In our mouse cell lines, *IL34* mRNA expression was strongly upregulated with increasing passage in all cell lines groups, suggesting that it may play a role in multiple stages of tumor progression (Fig. 6a). This was confirmed in an independent experiment where mRNA expression in primary tumor cells and lung metastatic cells was compared (Fig. 6b). We confirmed by ELISA that higher amounts of IL34 protein was produced and secreted into the culture medium by Passage 6 in “Kidney”, “Tail” and “Lungs” cells when compared to parental cells (Fig. 6c). *In silico* analysis in TCGA demonstrated that higher IL34 mRNA expression correlated with shortened OS and DFS (Fig. 6 d, e). This was also analyzed in M0 and M1 patients where IL34 expression was significantly correlated to OS and DFS in M0 patients (Supplementary Fig. 12a-d). Furthermore, *IL34* expression was also correlated to

tumor grade in TCGA (Fig. 6f). In the UroCCR cohort, IL34 mRNA expression was higher in tumors compared to healthy tissue (Fig. 6g). A Tissue MicroArray (TMA) from the same set of samples was stained for IL34 and graded and scored by a pathologist. In the tissue sections, expression was detected in tumor cells but also in normal tubules (Fig. 6h). Immunohistochemical staining correlated with Fuhrman grade, and was predictive of DFS (Fig. 6i-j). We also tested a limited number of plasma samples from a different cohort of patients undergoing treatment for metastases with the receptor tyrosine kinase inhibitor sunitinib (SUVEGIL an TORAVA clinical trial) (Supplementary Fig. 12e). In these conditions, a subset of patients (6/20) showed a marked spike in plasma IL34 following the first cycle of treatment. The sample size is too small to conclude a link between this spike and clinical parameters such as treatment response, and requires further investigation. However, in a xenograft mouse model (human 786-0 ccRCC cell line implanted s.c. in immunodeficient mice), treated with sunitinib, *IL34* mRNA expression was increased in tumor cells (Supplementary Fig. 12f, left panel). Stromal mouse *IL34* showed a similar trend towards increased expression indicative of a similar, though weaker, host response (Supplementary Fig. 12f, right panel).

To further study the role of IL34 in RCC, we overexpressed IL34 (variant 1) in RENCA-GFP cells, using a lentiviral vector and analyzed *in vivo* tumor development (Supplementary Fig. 13a-c). Tumors generated with IL34-overexpression versus control cells did not show significant altered size, and lifespan of mice was not changed (Supplementary Fig. 13d). IL34 expression was validated in tumors at the mRNA and protein level (Supplementary Fig. 13e, f). We next hypothesized that IL34 overexpression could alter the balance of Type 1 and 2 macrophages within the tumor and analyzed the levels of various macrophage markers in IL34-overexpressing versus control primary tumor lysates. We observed that IL34-

overexpressing tumors contained higher levels of the total macrophage markers *F4/80* and *CFSR1* (Fig. 7a) compared to control tumors, suggesting an increase in total macrophage number. No significant difference was detected in levels of the Type 1 markers *IL1β* and *IL12a* (Fig. 7b), whereas the Type 2 markers *MMR* and *ARG1* were increased in IL34-overexpressing tumors (Fig. 7c). Furthermore, *MMR* mRNA levels correlated positively with the level of IL34 in primary tumors (Fig. 7d) as well as the level of *CSF1R* (Supplementary Fig. 13g). Increased MMR expression in IL34 overexpressing tumors was validated at a protein level by Western Blot (Fig. 7e). Importantly, the level of CSF1, a twin cytokine to IL34 and also involved in macrophage attraction and differentiation, was not altered (Supplementary Fig. 13h). Furthermore, an analysis of MMR-stained tumor sections revealed an increased density of MMR-positive cells in the primary tumor (Fig. 7f).

Analyses of expression in metastases are limited by the relatively low number of metastases using the orthotopic model. Using the intravenous tail vein injection of IL34 overexpressing tumor cells and immunohistochemical analysis, we did not observe a significant increase in the metastasis area (Supplementary Fig. 13i). Nevertheless, we did detect an increased density of MMR-positive macrophages in IL34-overexpressing metastases (Fig. 9g), similarly to orthotopic primary tumor results.

Taken together, these data suggest that IL34 overexpression mediates an influx of macrophages with a polarization towards the type 2 phenotype. Furthermore, our data suggests that, in our mouse model, IL34 overexpression alone is insufficient to increase tumor size or speed of progression. It is important to note that parental RENCA cells do not express the CSFR1 receptor, and, thus, an autocrine role of the IL34-CSFR1 axis was not examined.

We next deleted IL34 in RENCA cells using two different Crispr/cas9 constructs targeting IL34 (CrisprIL34_1a, CrisprIL34-1c). We used a Crispr-Cas9 vector targeting LacZ as a control (Crispr_LacZ). Targeting IL34 by Crispr-Cas9 strongly decreased IL34 release in the medium when analyzed by ELISA (Supplementary Fig. 14a). Control and IL34-deleted cell lines were then implanted in the kidney (sub-capsular implantation) and mice were sacrificed at 26 days. Primary tumor growth was strongly inhibited when using the IL34-deleted cell lines (Fig. 8a). *In vitro* cell proliferation was however not affected (Supplementary Fig. 14b). We used magnetic resonance imaging (MRI) to track primary tumor volume dynamics, and appearance of metastases in the orthotopic model for one of the IL34-deleted cell lines (≥5 mice/group). Survival was significantly increased in the IL34-deleted cell line (Fig. 8b) and metastasis appearance was delayed (Fig. 8c). Primary tumor growth and metastasis formation was monitored in IL34 KO tumor and control tumor mice by Dynamic MRI. Tumor development and spread was mathematically modeled using ordinary differential equations (ODE, see Methods) (Fig. 8d). In the IL34 Crispr group, 3 mice did not develop primary tumors and were therefore not included in the ODE modeling. Furthermore, two of these mice did not develop metastasis and were also excluded. Consequently, the prediction of primary tumor volume and of metastasis number for the IL34 Crispr group shown in Fig. 8d is overestimated, but nevertheless demonstrated slower primary tumor growth and metastatic spread. The data were fitted using a population approach by means of a non-linear mixed effect model (see Methods). This approach allows individual modeling of growth and dissemination kinetics, while also quantifying inter-animal variability and group effects. It is particularly adapted to sparse data and strengthens the statistical power. The data fitted reasonably well the mathematical model, both at the population (Fig. 8d) and individual level (Supplementary Fig. 14c) which is composed of two parameters: a growth coefficient α and a

dissemination coefficient μ. A negative group effect for IL34 KO was observed for both α (p=0.0615) and μ (p=0.053) (Supplementary Table 3).

To examine macrophage density in the lungs after IL34 deletion, we used the tail vein metastasis model since this model is more efficient in generating lung metastasis. In the tail vein model, metastasis formation was drastically reduced using two of the CrisprIL34 cell lines (Fig. 8e). We used lung sections stained for GFP to calculate tumor area/Lung area, and MMR staining to calculate macrophage density. Within the metastases, MMR+ cell density was reduced in IL34-deleted tumors (Fig. 8f).

DISCUSSION

The usual strategy in translational research is to start from human samples, to identify molecular markers and gene networks and then to functionally validate them *in vitro* and in animal models. We devised herein a completely opposite strategy (from “mouse to man”) starting from the animal model and used functional genomics, imaging, clinical data and computational approaches. We took advantage of a syngeneic mouse model and developed it further to unravel mechanisms of RCC development and metastasis formation, and to identify novel therapeutic targets and biomarkers.

The strengths of our model are the following. The model is based on iterative cycles of injection using 3 different injection modes capturing local tumor development (“Kidney” group), metastasis formation from the primary tumor involving all steps of metastasis formation (“Lung” group; local invasion, transmigration into vessels, survival in the blood stream, extravasation into the lung tissue and seeding, metastatic tumor expansion), or only the last steps of metastasis formation (“Tail” group; presence in the blood circulation, extravasation). With increasing cycles, a shortening of the mice’s lifespan was also observed. The model is, furthermore, a syngeneic mouse model which leaves the adaptive immune system intact. Another important strength of our model is that the 3 different groups yield very distinct transcriptomic and methylome clustering and signatures, which are clinically relevant. These strengths outweigh the inconvenience of the model. The RENCA mouse model is indeed not a ccRCC model and does not have the VHL mutation. The advantages, however, are that the model is robust and reproducible, develops in an immunocompetent context in a relative short time frame and leads to the formation of lung metastasis after orthotopic implantation[29]. Our aim was to uncover the molecular changes that are occurring with increasing aggressiveness and tumor specialization and not the genetic characteristics of

the tumor itself, for which this RENCA model is well suited. This is different from human ccRCC models including implantation of cell lines or patient derived xenograft (PDX) models which require longer time to develop *in vivo* in an immunocompromised context and are mainly focused on primary tumor growth. In addition, no suitable genetically engineered mouse model (GEMM) for RCC that recapitulates with fidelity the metastatic process exists to date[30–33]. It is true that around 10 different GEMMs for RCC have been developed[34] such as the MYC activation, Vhl del Cdkn2adel, Ink4a/Arf del model[35], the Bap1 and Pbrm1 inactivation model[36]. Only the Vhl del Ink4a/Arf del model the produce some metastasis but only in the liver, which is not suitable for our study. Thus, even if our model does not strictly reflect the exact characteristics of a ccRCC tumor, it, nevertheless, recapitulates critical steps of tumor development and spread of a renal carcinoma.

In the sequentially amplified cell lines, RNA expression and methylation analysis demonstrated distinct clustering for the 3 different injection modes. DNA sequencing did not show clonal variations based on chromosomal variability, which indicate that the phenotypic changes were epigenetically regulated. Transcriptomic analysis led to the identification of specific gene signatures for each injection cycle which were predictors of overall, disease-free or progression-free survival in RCC as based on TCGA analysis and on our RCC cohorts. Importantly, some of the signatures, especially the Lung signature, are stronger predictors than current predictors in clinical use such as Fuhrman Grade or clinical stage. Beuselink et al determined four ccRCC signatures (ccrcc1-4 signatures) from the transcriptome of metastatic patient treated sunitinib [37]. The HR reported in their article is lower than in our study. In a follow-up study by this team, HR were calculated in a patient cohort treated with pazopanib which is different to our study[38]. Recently, Ricketts et al[39] reported a comprehensive molecular characterization using the TCGA database where

they compared the 3 types of RCC (ccRCC, papillary RCC and chromophobe RCC). For ccRCC the signature were related to increased ribose metabolism pathway and to Th2 immune profile. However our study is very different, because starting from an animal model we specifically focused our comparative translational analysis on the ccRCC subtype and, thus, the results cannot be compared, albeit their study also revealed immunology-related gene expression. Furthermore, a recent study reported tracking of ccRCC evolution at the genomic level and demonstrated that metastatic competence was afforded by chromosomal complexity with loss of 9p as a selective event for metastasis and patient survival[40]. Our study did not include an analysis of the chromosomal events, since our animal model did not reveal chromosomal alterations and, thus, we specifically focused on modifications of the transcriptome and the methylome.

Methylome analysis from the S5 series revealed that only 29 % of 33095 CpGs differentially methylated genes were associated with a known gene. Only a restricted number of genes found in this analysis was contained in the transcriptome signatures which include *IL34*, *Krt17*, *RIK/MAGEC2/3*, and *SAMSN1*.

Thus, despite the excellent clustering only a restricted number of genes are in common. When GO terms are crossed between transcriptomics and methylome data sets, they also, only, converge partially to a common set of processes. The reason for this resides in the fact that many methylation sites are found within intronic sequences, the role of which is not really understood, but which may have major regulatory functions. For IL34, three CpGs were identified as part of enhancer regions with two of those found in the gene body and predictive for OS which points to a possible epigenetic regulation of IL34 (see for more on IL34 later in discussion).

We further selected, for a more detailed investigation, a number of individual genes having no defined roles in RCC, which we showed to be predictive of patient outcome in online databases and verified in a second local patient cohort.

In this context, we selected SAA2, an acute phase protein related to SAA1, which was found to be a strong predictor of OS and DFS in the TCGA KIRC database globally and also for the M0 and M1 subgroups. In patients with the highest Fuhrman Grade, significantly increased SAA2 expression was observed. Furthermore, analysis of plasma samples from patients with metastases before or after surgery showed higher plasma levels of SAA2 and worse OS. Analysis of samples from the clinical trials evidences SAA2 as an excellent predictive biomarker especially in the sunitinib treated patient cohort. Serum amyloid A has been previously reported to be prognostic marker in RCC[41–43]. In the Vermaat et al study[42,43], there is no distinction between the four SAA variants (SAA1-4) contrary to our study where we clearly identified SAA2. In addition, in the first article[42], most of the patients included in their study were treated with interferon as first line treatment and not with anti-angiogenic drugs currently in clinical use and in this case PFS was not analyzed and only the combination with apolipoprotein A2 (ApoA2) was predictive for OS. The second study[43], which demonstrated highly significant predictive value for SAA in RCC metastatic patients treated by tyrosine-kinase inhibitors (TKI), is in agreement with our data, albeit they did not specify the SAA variant.

Another molecule of interest is complement factor B (CFB) which was upregulated in “Lung” and “Tail” cell lines. Complement factor B expression correlated with survival and metastasis in the TCGA data set and in the UroCCR cohort. Plasma measurements showed that, similarly to SAA2, patients with metastases had higher CFB plasma levels compared to patients without metastases. This remained the case whether the samples were taken before or

after surgery, suggesting that CFB may be a useful blood borne marker of metastasis. Like SAA2, CFB plasma levels were tested in patients with metastasis treated with antiangiogenic drugs, before and after first cycle treatment with sunitinib. Patients whose CFB levels were increased following treatment had faster disease progression and shortened survival compared to patients whose level were decreased. Interestingly, the combination of SAA2 and CFB plasma level measurements produced a more powerful analysis than either marker alone. This suggests that a combined analysis of these two markers represent a powerful tool for predicting patient outcome following sunitinib treatment. Interestingly, CFB is a prognostic preoperative marker in pancreatic carcinoma which outperformed CA19-9 and CEA[44]. However, the combination of CFB and SAA2 was not tested in this study.

PodocanLike-1(PODNL1) which was upregulated in cell lines specialized for primary tumor development was one of the strongest predictors of survival and metastasis in the TCGA and UroCCR cohort. Immunohistochemistry showed strong expression in patient tumor samples. The function of PODNL1 is completely unknown and further studies are required.

GPRC5a is a potential therapeutic target and biomarker in several solid tumors including prostate cancer, lung cancer or colon cancer[45,46]. It has been identified in a differential gene expression analysis comparing T1 and T3 ccRCC stages[47]. In our study, increased GPRC5A expression was associated with a shorter survival. This is in agreement with results from prostate cancer, where GPRC5A expression also predicts a worse OS and was shown to be involved in metastasis since knock-down of this molecule limited bone metastasis[27]. Interestingly, GPRC5A has been identified as a novel hypoxia-induced protein that functions to protect cancer cells from apoptosis during oxygen deprivation and was found to be upregulated in colorectal cancers. Mechanistically, GPRC5A acts via the Hippo pathway effector YAP[48]. However, in non-small cell lung carcinoma (NSCLC), GPRC5A is expressed

as much lower levels than in the corresponding noncancerous tissues and was associated to a worse prognosis[46]. Thus, the role of GPRC5A seems context-dependent and either associated with a pro-tumor or anti-tumor context.

One of the most interesting genes in our study was Interleukin-34 (IL34), a secreted cytokine, which binds to CSFR1[49], receptor-type protein-tyrosine phosphatase ζ (PTP-ζ)[50] and Syndecan-1[51]. It is expressed in several organs, including in tubular cells of the kidney[52]. IL34 supports the growth and survival of primary human monocytes and induces their differentiation into type 2 macrophages[53] IL34 has been linked with autoimmune and inflammatory diseases such as rheumatoid arthritis[53,54], hepatitis C infection[55], lupus nephritis[56], inflammatory bowel disease[57]. It has been linked to chronic kidney injury and kidney graft rejection[52], and to cancers including osteosarcoma[58] or lung cancer[59].

In our study, interleukin-34 was upregulated in all 3 groups of aggressive mouse cell lines, suggesting that it could contribute to all stages of RCC progression, acting as a key driver/regulator of aggressiveness. We showed that its expression in RCC correlated with increased Fuhrman grade, and shortened survival and faster metastatic progression, using two cohorts of patients, and RNA and immunohistochemistry data. We also provided evidence that IL34 is detected in the plasma of a subset of patients treated with sunitinib, the first line treatment for metastatic RCC.

For functional analysis, we focused on IL34 since IL34 was a common driver to all experimental groups and since our data also indicated a potential epigenetic regulation of this gene with two potential methylation sites predictive for OS and PFS. We first overexpressed IL34 in RCC and determined the effect on macrophages and tumor development in the mouse model. Our studies showed that mouse tumors overexpressing IL34 had an altered pool of

tumor associated macrophages. Markers of total macrophages were increased relative to control tumors indicating an increased number of macrophages. A further examination of macrophages markers showed that this was due to an increase in type 2 pro-tumorigenic macrophages while type 1 macrophages were not significantly altered. We further confirmed the increase in type 2 macrophage density in the tail vein metastasis assay. It is known that Type 2 macrophages are associated with worsened progression in multiple cancer types[59]. However, IL34 overexpression alone was not sufficient to further increase tumor progression in our model, but CrisprCas9-mediated deletion of IL34 in RENCA cells strongly inhibited primary tumor formation and metastasis after both, sub-capsular implantation and tail vein injection. *In vitro* proliferation in IL34 KO cells was not modified, suggesting that this effect is microenvironment-dependent. Mouse survival time was increased, and MRI analysis showed reduced numbers and slowed development of metastases which was also mathematically modeled. Importantly, RENCA cells do express CSF1, but our experiments showed that the loss of IL34 cannot be compensated by the presence of CSF1. Our data also point to the fact that IL34 is already required at an early stage of tumor development. Taken together, these studies demonstrate a previously unknown role for IL34 in progression of RCC and suggest it as an important potential target. IL34 has been also shown to promote tumor progression and metastasis in other tumor types such as osteosarcoma[60–66]. The effect of IL34 and CSF-1 blockade was recently explored using specific blocking antibodies[58]. Mouse colon carcinoma (MC38 cell line) development was only weakly inhibited by both IL34 and CSF-1 blockade but only CSF1 blockade had an effect on TAM homeostasis and $CD4^+$ T cell accumulation which is very different from our data.

All in all, our work demonstrates that a multilayered systems approach, from mouse to man as described in this work, generates many meaningful results that are not only important for the understanding of tumor biology but which are also highly relevant for clinical translation.

Methods

Mice and cell lines. All animal experiments were approved by the "Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI)" (authorizations numbers 2016072015478042; 2015110618597936 and 2015070315335217), and were carried out in accordance with the approved protocols.

Mouse tumor and metastasis models. The Renca murine renal cancer cells, and the human 786-0 cancer cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% feral bovine serum (FBS) and 1% penicillin/streptomycin (Complete Medium), and were incubated at 37°C, 5% CO_2 in an incubator.
For sub-capsular implantations of Renca cells, $1*10^5$ cells were injected under the left kidney capsule of 6-week-old female BALB/c mice (Charles River Laboratories).
For intravenous injections of Renca cells, $5*10^6$ cells were injected into the caudal vein (tail vein) of 6-week-old female BALB/c mice. When the endpoints defined by the approved protocols were reached, mice were sacrificed and tumor tissues and lungs were collected. When a first mouse from a group was determined as having reached an endpoint, all mice from that group were sacrificed. Mice were housed in the animal facility of Bordeaux University (Animalerie Mutualisée, Université de Bordeaux, France).
For sub-cutaneous implantations of 786-O cells, $5*10^6$ cells were injected into the flank of 5-week-old nude (nu/nu) female mice (Janvier, France). The tumor volume was determined with a caliper (v = L*l2*0.5). When the tumor reached 100 mm^3, mice were treated five a week for 4 weeks, by gavage with placebo (dextrose water vehicle) or sunitinib (40 mg/kg).

Tissue dissociation and tumor cells purification. For tumor cell purification, tissues were cut into small pieces with a scalpel and digested with Collagenase I and Collagenase II (Liberase TL, Roche, 05401020001) for 1 hour at 37°C. To further improve the dissociation, digested tissues were filtered in cell strainers (100µm, 70µm and 40µm) and seeded in complete medium, and incubated at 37°C, 5% CO_2 in incubator. Cell cultures were checked daily and passaged as necessary. Tumor cell outgrowth and primary cell death resulted in tumor cell only cultures, verified by visualization of GFP using fluorescence microscopy When no GFP-negative cells could be visually detected, cell cultures were considered sufficiently pure. RENCA-GFP cells were collected for analysis or re-implanted into mice for the next *in-vivo* passage.

Tissue treatments. For immunochemistry, tissue were fixed in paraformaldehyde 4% (PFA 4%, Santa Cruz Biotechnology, sc-281692) for 2 hours and then incubated for 72 hours in 30% sucrose. Tissues were frozen in OCT Compound (Tissue-Tek OCT compound, Sakura, 4583). Prior to embedding, lungs were inflated with 1mL of diluted OCT (1:1 PBS/OCT dilution). Frozen tissues were preserved at -80°C. For protein, DNA and RNA analysis, tissues were snap-frozen in liquid nitrogen and preserved at -80°C.

Low-coverage whole-genome sequencing. Genomic DNA was isolated use the Dneasy Blood and Tissue Kit (Qiagen, 69504), according to the manufacturer's protocols. 13 samples were subjected to low-coverage whole-genome sequencing. Libraries were created using the KAPA LTP library preparation kit for Illumina platforms (Kapa Biosystems) following the manufacturers recommendations and sequenced on an Illumina HiSeq4000 sequencer in 51bp single-end mode. Raw reads were mapped to the mouse reference genome (mm10/GRCm38)

using the Burrows-Wheeler Aligner[67]. On average, 6,882,693 reads were mapped per sample. PCR duplicates were removed using Picard (v1.43) resulting in, on average, 5,966,979 uniquely mapped reads per sample. These reads were further processed with the Bioconductor package QDNASeq.mm10. Problematic regions were excluded and read counts were corrected for mappability using LOESS regression. The number of reads was counted in bins of 50kb along the genome, log2-transformed and normalized by the median. These resulted in logR values per bin, that were subsequently segmented using the ASCAT algorithm[68].These segmented values, as well as the individual logR values per bin, were used to plot the individual copy-number profile of each sample. Subclonal tumor fractions of the samples were estimated by ABSOLUTE[69].

Transcriptomic data generation and analysis. Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, #74134), according to the manufacturer's protocols. Analyses of transcription sets, signature computation and enrichment analysis were done using R studio (R v3.5.2[70], R studio v1.1.463[71]).
Sample count by group and series:

Group	**S0**	**S1**	**S2**	**S3**	**S4**	**S5**
Parental cell line	2	NA	NA	NA	NA	NA
Kidney	NA	2	2	3	5	5
Tail	NA	2	3	4	6	4
Lung	NA	1	5	3	4	3

From the log2 scale normalized data set, Principal Component Analysis was performed on series S0 and S2 as well as S0 and S5 (function prcomp of stats R package (v3.5.2) with the parameter center = T). Genes with the most important association where selected by keeping genes whose contribution is above the mean of all contributions for PC1 and/or PC2 for both S0 and S5 PCA analysis. This resulted in a set of 5140 genes. Expression values of these genes were further centered and reduced. For each of the series S0 through S5, a mean value was computed. We further built a heatmap of these mean values (pheatmap R package; v1.0.12).

Biomarker discovery. ***(1) Differential expression between parental cell line (series 0, S0) and series 5 (S5):*** To select genes that present the strongest difference in expression between parental cell line and series 5 experiments, we used a z-score approach. First, we compute the log fold (logFC) change for each gene. Second, we compute the mean and standard deviation of the logFC and a z-score for all genes. A gene was considered as differentially expressed if the absolute value of its z-score was ≥ 2.58 and if its logFC is ≤ -1 (down expressed) or ≥ 2 (up expressed). ***(2) Progressive expression pattern through series:*** To capture genes that present a progressive expression through series, we compute mean values of expression for each series. A gene will be considered progressive if its mean values between replicates for a given series are strictly increasing (up) or decreasing (down) through series 0 to 5. As the S1 of Lung group had only one measure, it was not usedfor the determination of the progressive quality of genes. ***(3) Differentially expressed and progressive genes.*** A gene was considered differentially progressively expressed through series if it was in the intersection of the gene sets selected in steps (1) and (2) with the additional consistency constraint that the expression should be in the same direction (up or down).

Biomarker validation. Only genes whose overexpression was predictive of a bad prognostic were conserved. ***(1) Selection of clinically relevant genes:*** In order to conserve genes that are relevant in human kidney cancer, we used the TCGA KIRC cohort. For each gene, we fitted a Cox proportional hazard regression model based on overall survival (OS) time and disease-free survival (DFS). A gene was conserved if its cox proportional hazard ratio (HR) was in accordance with the differential expression (HR ≥ 1 for up regulated genes and HR ≤ 1 for down regulated genes) and if the FDR adjusted p-value of its log-rank test was ≤ 0.01 for OS and/or DFS. ***(2) Relevance of the association of selected genes:*** To measure the clinical relevance of the resulting signature and check their prognostic performance against random signatures, we used SigCheck R package (v 2.14.0). To separate samples into groups, we computed a score corresponding to the mean value of all the expression values of genes for each sample (scoreMethod="High" in the sigCheck function). Patients were then classified in accordance to their score and splited in 3 groups using a log-rank test followed by computation of the associated log-rank HR (survival v2.44-1.1 and survminer v0.4.3 R packages). We then compared the performance of our previously defined signatures with 1000 signatures composed of the same number of randomly selected genes. TCGA Kidney Renal Clear Cell Carcinoma (KIRC) HiSeqV2 data were downloaded from XenaBrowser[72]. We chose log2(x+1) transformed RSEM normalized count (version 2017-10-13) as recommended on its web site (https://xenabrowser.net/; page: «dataset:gene expression RNAseq-IlluminaHiSeq percentile»). We removed genes where there were more than 90% of samples that had null values resulting in 19186 genes. Complementary associated clinical data were download from cbioportal (https://www.cbioportal.org/). Conversion of *Mus Musculus* to Human gene name was done by downloading conversion table from the Biomart website (https://www.ensembl.org[73]).

Methylomics data generation and analysis. Whole Genome Bisulfite Sequencing (WGBS) was performed for 3 replicates of the S5 series for K, T, and L groups and 1 for the parental cell line at the GeT-PlaGe core facility, INRA Toulouse. WGBS libraries have been prepared according to Biooscientific's protocol using the Biooscientific NEXTflex™ Bisulfite Library Prep Kit for Illumina Sequencing. Briefly, DNA was fragmented by sonication, size selection was performed using Agencourt AMPure beads XP and adaptators were ligated to be sequenced. Then, bisulfite treatment was performed for 2.5 hours using the EZ DNA Methylation-Gold™ Kit from Zymo Research, and 12 cycles of PCR were performed. Library quality was assessed using an Advanced Analytical Fragment Analyser and libraries were quantified by QPCR using the Kapa Library Quantification Kit. WGBS experiments have been performed on an Illumina HiSeq3000 using a paired end read length of 2x150 pb with the Illumina HiSeq3000 Reagent Kits. To determine conversion efficiency, fastq files were trimmed for adapters and low quality bases with Trim Galore (v0.4.4, calling cutadapt 1.3[74]) then mapped to the pUC19 reference genome (pUC19.fa) with Bismark (v0.13.0[75]). Samtools (v0.1.19-44428cd[76]) was used to remove duplicated reads. Then methylation calling was performed with Bismark_methylation_extractor. As methyled and non-methylated cytosine positions are known on the pUC19 reference genome, over and under-conversion could be assessed. Filtered fasq file were generated by CASAVA 2.17. Fastq files were aligned with Bismark (v0.17.1_dev[77]) against the GRCm38.p5 *mus musculus* genome (download from http://www.ensembl.org/, release 89[76]) with following parameters -N 0 and --maxins 800. Bismark use Bowtie 2 (v2.3.4.3[74]) and samtools (v1.9[78]). Incomplete bisulfite conversion filtering was done on Bismark BAM files in order to remove reads that exceed a certain threshold of methylated calls in non-CG context. Then, deduplication was applied

(deduplicate_bismark) followed by extraction of methylated positions (bismark_methylation_extractor). Next analysis were done using R studio (R v3.5.2[71], R studio v1.1.463[72]). We then considered CG positions with at least 10 reads of coverage resulting to 106303 CGs on chromosomes 1-19 and X. CGs were then annotated with RefSeq (download from UCSC website[77]) using the biomaRt R package (v2.38.0[79]) resulting in 6606 unique gene names. Principal Component Analysis (PCA) was performed on S0 and S5 for methylation frequencies of all CGs (function prcomp of stats R package (v3.5.2) with the parameter center = T). CGs with the most important association were selected by keeping CGs whose contribution is up to 5 times the mean of all contributions for PC1 or PC2 of the PCA. The resulting 2063 CGs were localized in the gene body of unique 1427 genes used to perform enrichment analysis.
We further used the methylKit R package (v1.9.4) to calculate differential methylation at base resolution[80]. After normalization of coverage values between samples by the median, reads covering both strands of a CpG dinucleotide were merged. Fisher's exact test was applied to identify differentially methylated cytosines (DMCs). Resulting DMCs in gene body were then confronted with genes that were differentially expressed between S0 and S5. TCGA methylation data sets (illuminaMethyl450_hg19_GPL16304_TCGAlegacy) of the KIRC cohort were download from XenaBrowser. Refseq (download for GRCh37 from UCSC website) was used to conserve only CGs that were in accord with the locus of the DMCs and only for genes that were in transcriptomics signatures. The resulting 5 CGs were located the intron 1 of IL34 and all had a positive correlation with IL34 expression. For each of these 5 CGs, we fit a Cox proportional hazard regression model based on overall survival (OS) and disease-free survival (DFS). Two CGs had an HR > 1 and a p-value of its log-rank test ≤ 0.01 for OS. As for transcriptomics data, the clinical relevance and prognostic performance of the 2 CGs were test against 1-CG random signatures using SigCheck R package. To separate samples into groups we computed a score corresponding to the mean value over all the methylation values for each sample (scoreMethod="High" in the sigCheck function). Patients were then split into 2 groups (low and high) to perform a log-rank test and compute associated HR (survival v2.44-1.1 and survminer v0.4.3 R packages). The Genome Browser screen shots were produced by using http://genome.ucsc.edu webtool (GRCm38 and GRCh37[81]) and ensembl regulation data sets (mus_musculus.GRCm38.Regulatory_Build.regulatory_features) for GRCm38 and GeneHancer for GRCh37[82].

Enrichment analyses. To perform enrichment analyses we used hypergeometric test (enricher function of Cluster Profiler R package[83]; v3.10.1) with go_terms.mgi download on Mouse Genome Database (MGD) at the Mouse Genome Informatics website (URL: http://www.informatics.jax.org[84] (04, 2019). Enrichment analysis was done for following gene sets: (i) transcriptomics data: gene set related to PC1, gene set related to PC2 and gene sets related to PC1 and PC2 (universe was setting to the set of genes which are measured in transcriptomic experiments) (ii) methylomics data: gene sets related to PC1 and PC2.

Data availability. Microarray gene expression data is available via Gene Expression Omnibus using the accession GSE142109, Access: Upon request.
DNA sequencing was deposited in ArrayExpress:
http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8645, Access:Upon request.
Methylation data is available via Gene Expression Omnibus using the accession GSE139338 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139338; Access: upon request.

Immunochemistry and immunofluorescence. For frozen tissues, 10µm sections were performed with a cryostat (Leica CM1900).For paraffine tissues, 10µm sections were performed with a microtome. For frozen tissue immunofluorescence, sections were incubated 1 hour with a blocking buffer (5% BSA in PBS). Slides were incubated overnight with primary antibody (MMR R&D Systems, AF2535), and then with secondary fluorescent antibody (Interchim, FP547H) and DAPI (ThermoFisher Scientific, 62247). For paraffin tissues immunohistochemistry, slides were deparaffinized, re-hydrated and heated in Antigen Retrieval Solution pH6 (HIER Sodium Citrate Buffer, pH6; 10mM Sodium Citrate, 0,05% Tween 20, pH 6,0). To block endogenous peroxidase activity, slices were treated with 0,3% hydrogen peroxide. After 1 hour of blocking in PBS 5% BSA, slides were incubated overnight with primary antibody (see table), and then incubated with biotinylated secondary antibody for 1h (see table). Secondary antibodies were HRP-conjugated using the “ABC” technique (Vectastain PK-6100) and then revealed with a peroxidase substrate kit (DAB, Vector Laboratories, SK-4100).

Target	Antibody dilution	Reference	Secondary
Mouse MMR	1/25	R&D Systems, AF2535	Interchim, FP-SA2110
Human IL34	1/200	Abcam, ab101443	Dako, E0433

Bright-field microscopy was performed with a Nikon Eclipse E4000 microscope (Nikon) and NIS-Elements F 3.2 software (Nikon). Fluorescent microscopy was performed with a Nikon Eclipse i90 microscope (Nikon) and NIS-Elements AR 4.30 software (Nikon). A slide scanner (Hamamatsu, Nanozoomer 2.0HT) from the Bordeaux Imaging Center was used for whole slide imaging using NDP.scan software (Hamamatsu). Analysis were performed using Fiji. Cell counting was performed with “Cell Counter” plugin (Kurt De Vos).

Generation of GFP, IL34, overexpressing cells and crispr/cas9 induced IL34 knock-down cells.

EGFP expressing lentivirus (pRRLsin-MND-eGFP-WPRE) was obtained from the vectorology plateform of the University of Bordeaux (Vect’UB).

The IL34 lentiviral vector was obtained the in following way. Total RNA was extracted from Renca cell and reverse-transcribed into cDNA using the high-capacity cDNA reverse transcription kit (Applied Biosystems). Mouse IL34 variant 1 cDNA were then amplified by PCR using specific primers containing attB site : fwd : 5'-GGGG-ACA-AGT-TTG-TAC-AAA-AAA-GCA-GGC-TTC-ATG-CCC-TGG-GGA-CTC-GCC-TGG-CTA-3’; rev: 5’-GGGG-AC-CAC-TTT-GTA-CAA-GAA-AGC-TGG-GTC-TCA-GGG-CAA-CGA-GCC-ATG-GCT-TGA-3’. attB-PCR product was cloned into pDONR221 (ThermoFisher, #12536017) by BP gateway cloning and then transferred into pLenti CMV/TO Puro DEST (Addgene #17293) by LR gateway cloning to give pLenti CMV/TO-mIL34v1-Puro lentiviral vector. For the IL34 lentiviral cripr/cas9 construct, lentiCRISPRv2 plasmid containing spCas9 cassette and expressing guide RNA targeted LacZ or mouse IL34 exon 1 were kindly provided by Michel Tremblay (McGill University) with the following targeted sequence: mIL34 ex1a: 5’-gatcctacttgacgtggctttgg-3’ and mIL34 ex1c: 5’-gaccttacaggctaccttcgggg-3’. Viral particles were produced by calcium phosphate co-transfection of HEK293T cells with the packaging plasmids pVSVg (Addgene 8454), psPAX2 (Addgene 12260) and a transfer plasmid (e.g. EGFP- or mIL34v1-plasmid or pLentiCRISPRv2-LacZ, -mIL34_1a, -

mIL34_1c). Renca cells were infected with lentiviruses at a multiplicity of 25 infectious units per cell and selected for 4 days in media supplemented with 1µg/ml puromycin.

qPCR gene expression analysis. Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen, #74134), according to the manufacturer's protocols. 1µg of total RNA was reverse-transcribed into complementary DNA (cDNA) using the high-capacity cDNA reverse transcription kit (Applied Biosystems, 4368814). The resulting cDNA were amplified using specific primers for the genes of interest. For RT-qPCR, Taqman (Eurobio Scientific, EurobioProbe) or SybrGreen (Eurobio Scientific, EurobioGreen) technologies were used. HPRT was used as internal control.

Gene name	Primer Reference / sequences
Mouse HPRT	ThermoFisher Scientific, TaqMan Assay Mm01545399_m1
Mouse IL34	ThermoFisher Scientific, TaqMan Assay Mm01243248_m1
Mouse MMR	SybrGreen, fwd : 5'-TGCCGGCGTTGCAGCCTATT ; rev : 5'-GCTCATTCTGCTCGATGTTGCCCA
Mouse CSF1R	ThermoFisher Scientific, TaqMan Assay Mm01266652_m1
Mouse CSF1	ThermoFisher Scientific, TaqMan Assay Mm00432686_m1
Mouse ARG1	ThermoFisher Scientific, TaqMan Assay Mm00475988_m1
Mouse IL1β	ThermoFisher Scientific, TaqMan Assay Mm00434228_m1
Mouse IL12a	SybrGreen, fwd : 5'-CATCGATGAGCTGATGCAGT ; rev : 5'-CAGATAGCCCATCACCCTGT
Mouse F4/80	SybrGreen, fwd : 5'-TGAGCAGATACAGCAATGCCA ; rev : 5'-AGCTGCACTCTGTAAGAACAC
Mouse CFB	ThermoFisher Scientific, TaqMan Assay Mm00433909_m1
Mouse PODNL1	SybrGreen, fwd : 5'-CACTCCGGTTCTGTGTACCTC ; rev : 5'-GTGGAAGGTGTTAGGTGGCA
Mouse GPRC5a	SybrGreen, fwd : 5'-AGTCGCAAGTTCTCGCTTCA ; rev : 5'-TCATTCTGGTTGGTCCTGGG
Mouse SAA2	SybrGreen, fwd : 5'-TTCAGAAGGCTGTGTTGGGG ; rev : 5'-CCCTCTCCTCCTCAAGCAGT
Mouse IL12a	SybrGreen, fwd : 5'-CATCGATGAGCTGATGCAGT ; rev : 5'-CAGATAGCCCATCACCCTGT
Human IL34	ThermoFisher Scientific, TaqMan Assay Hs00380959_m1
Human CFB	SybrGreen, fwd : 5'-CAAAGCAAGCCAGGACACAC ; rev : 5'-GTGGTCACACCTCCAGACAA
Human GPRC5A	SybrGreen, fwd : 5'-GGCACTAGGGTCCAGAATGG ; rev : 5'-CAGTCCGATGATGAAGGCGA
Human PODNL1	SybrGreen, fwd : 5'-CACCTTCAGCAAGCRGCATA ; rev : 5'-CTGCAGCAACAAATAGCGCA
Human SAA2	SybrGreen, fwd : 5'-ACCCCAATCACTTCCGACC ; rev : 5'-GCTTCTCTGGACATAGACCTCACTA
Human HPRT	ThermoFisher Scientific, TaqMan Assay Hs99999909_m1

Western Blot analysis. Cells were washed with PBS and dissolved in lysis buffer (50mM Tris-HCl pH 7,4, 150mM NaCl, 1% TritonX-100, 0,1% SDS, 0,1% sodium deoxycholate) supplemented with protease and phosphatase inhibitors cocktails (Roche, Complete Protease Inhibitor Cocktail; PhosSTOP). For tissues, pieces of approximately 5mm^3 were dissolved in lysis buffer. Protein concentration was quantified by Bradford assay (Euromedex). Protein lysates were diluted in Laemmli buffer (62,5mM Tris-HCl pH 6,8, 10% glycerol, 2,5% SDS, 2,5% β-mercaptoethanol), boiled for 5 minutes at 95°C and resolved by SDS-PAGE. Protein were then transferred onto a nitrocellulose membrane (BioTrace, 66485). Membranes were incubated with blocking buffer (see table) for 1 hour, and then probed with primary antibody (see table). A fluorescent secondary fluorescent antibody (see table) was used as detection system. Membranes were scanned with an infrared imaging system (Odyssey, Li-Cor Biosciences, Nebraska, US). Densitometry analysis were performed using Image Studio Lite 4,0 Software (Li-Cor).

Protein of interest	Associated Blocking buffer	Concentration, Buffer	Antibody Reference	Secondary antibody reference
Mouse IL-34	TBS-T 3% BSA	1/250, TBS-T 0,5% BSA	R&D systems, F5195	IRDye 800CW 925-32214 (Li-Cor Biosciences) (also react with sheep IgG)
Mouse MMR	TBS-T 5% BSA	1/200, TBS-T 5% BSA	R&D Systems, AF2535	IRDye 800CW 925-32214 (Li-Cor Biosciences)
Mouse vinculin	Odyssey Blocking Buffer (Li-Cor Biosciences)	1/500, Odyssey Blocking Buffer (Li-Cor Biosciences)	Sigma, V9131	IRDye 800CW 926-32212 (Li-Cor Biosciences)

ELISA. ELISA were performed according to the manufacturer's protocols: human IL34 (R&D systems, D3400), human SAA2 (DLDEVELOP, DL-SAA2-Hu-96T), Human CFB (abcam, ab137973).

Human patient samples. Patient samples (tumor tissue and plasma) from the UroCCR cohort were used with associated clinical data (clinicaltrial.gov, NCT03293563). Eligible patients for SUVEGIL and TORAVA trials were at least 18 years of age and had metastatic ccRCC histologically confirmed, with the presence of measurable disease according to Response Evaluation Criteria in Solid Tumors v1.1. Patients had not received previous systemic therapy for RCC and were eligible for sunitinib or bevacizumab treatment in the first-line setting. Patients were ineligible if they had symptomatic or uncontrolled brain metastases, an estimated lifetime less than 3 months, uncontrolled hypertension or clinically significant cardiovascular events (heart failure, prolongation of the QT interval), history of other primary cancer. All patients gave written informed consent. Tumors were assessed at baseline and then every 12 weeks by thoracic, abdominal, pelvic and bone CT scans. Brain CT scans were performed in case of symptoms. This cohort includes patients from the SUVEGIL (24 patients) and TORAVA (35 patients) trials. The SUVEGIL trial (clinicaltrial.gov, NCT00943839) was a multi-center prospective single-arm study. The goal of the trial is to determine whether a link exists between the effectiveness of therapy with

sunitinib malate and development of blood biomarkers in patients with kidney cancer. Patients received oral sunitinib (50 mg per day) once daily for 4 weeks (on days 1 to 28), followed by 2 weeks without treatment. Courses repeat every 6 weeks in the absence of disease progression or unacceptable toxicity. The TORAVA trial (clinicaltrial.gov, NCT00619268) was a randomized prospective study. Patient characteristics and results have been previously described[85]. Briefly, patients aged 18 years or older with untreated metastatic ccRCC were randomly assigned (2 : 1 : 1) to receive the combination of bevacizumab (10 mg kg^{-1} iv every 2 weeks) and temsirolimus (25 mg iv weekly) IFN-α (9 mIU i.v. three times per week), or one of the standard treatments: sunitinib (50 mg per day orally for 4 weeks followed by 2 weeks off)[86]. These studies were approved by the ethic committee at each participating center and run in agreement with the International Conference on Harmonization of Good Clinical Practice Guideline. Blood samples were collected during the inclusion visit (baseline).

Magnetic Resonance Imaging for in vivo mouse monitoring. Experiments were performed on a 7T Bruker BioSpec system equipped with a gradient coil of 660 mT.m-1 maximum strength and 110 µs rise time. A volume resonator operating in quadrature mode was used for excitation (75.4mm inner diameter, 70mm active length), and a proton phased array (RAPID Biomedical GmbH) was used for signal reception (4 elements of 30 mm long around an elliptic cylinder housing: 19 × 25.5 mm). A Self-Gating balanced Steady State Free Precession sequence (SG-bSSP[87]) was employed to image the mouse lungs in order to count the pulmonary metastases. The following parameters were used: Field-Of-View (FOV): 25x22x20mm, spatial resolution: 195x172x156um, TE/TR=2/4ms, reception bandwidth (BW): 100kHz, flip angle: 30°, 4 repetitions, 4 phase offsets, total acquisition time: 22min. Thereafter, the motion correction technique was used as already explained in Ribot EJ et al[87]. A standard bSSFP sequence was also applied to image the mouse abdomen and measure the primary tumor volumes. The following parameters were used: FOV: 25x22x20mm, spatial resolution: 195x172x156um, TE/TR=2/4ms, BW: 75kHz, flip angle: 30°, 4 averages, 4 phase offsets, total acquisition time: 22min.

Mathematical modeling methods. The data consisted of longitudinal observations of total volume of the kidney and the primary tumor (PT) coupled with lung metastases counts, for two groups of six mice, the control group LacZ and the IL34 KO CRISPR group. These data were obtained from MRI images of subcapsular implanted mice. The volumes of the mice kidneys were measured and were approximately equal to 187 mm^3, enabling then to deduce the PT volume for each mouse. We considered a model that included two ordinary differential equations: one for the primary tumor growth, and the other for metastatic invasion (counts of metastases) as a result of dissemination from the PT. We denoted by V_p and N the volume of the PT and the number of lung metastases, respectively. The initial time t^0=0 is the time of tumor cells implantation. The initial volume of the primary tumor $V_p{}^0$ is set at the volume converted from the number of injected cells (using the conversion rule 1 mm^3 ≈10^6 cells[88], thus $Vp0$= 10^5 cells here. The model writes

$$\begin{cases} \frac{dV_p}{dt}(t) = \alpha V_p{}^{\beta}, & V_p(t^0) = V_p^0, \\ \frac{dN}{dt}(t) = \mu V_p{}^{\beta}, & N(t^0) = 0. \end{cases}$$

The first equation describes the growth kinetics of the PT with a growth rate that depends on the volume of the PT. It is assumed to follow a power law model[89] with parameters α

(proliferation rate) and β (fractal dimension of the PT vasculature). The second equation features the metastatic development through a PT volume-dependent rate, defined by the parameters μ and β. The parameter μ is the per cell probability of successful emission of metastases from the primary tumor to the lungs. We refer to this parameter as the intrinsic dissemination coefficient. The same fractal dimension β is used following the assumption that cells able to leave are cells close to the vasculature[90]. In order to calibrate the mathematical models with the data, we used a statistical population approach by means of the non-linear mixed effects formalism[91]. We assumed log-normal distributions for the parameters (α^i,μ^i) with i the animal index, with group effect considered as a categorical on α and μ. Parameter β was found to exhibit small variability and no random effects were considered for this parameter.

$$\begin{cases} \ln \alpha^i = \ln \alpha_{pop} + \text{group_IL34}(\alpha) x^i + \eta_\alpha^i, & \eta_\alpha{}^i \sim \mathcal{N}(0, \omega_\alpha) \\ \ln \beta^i = \ln \beta_{pop} \\ \ln \mu^i = \ln \mu_{pop} + \text{group_IL34}(\mu) x^i + \eta_\mu^i, & \eta_\mu^i \sim \mathcal{N}(0, \omega_\mu) \end{cases}$$

where x^i=1 if animal i is in the IL34 KO group and 0 if not.We assumed a proportional error model for the PT volume observations and a constant error model for the metastases counts. To compute the parameters of interest, we maximized the likelihood of the pooled data using the stochastic approximation expectation maximization (SAEM) algorithm implemented within the Monolix 2019R1 software (Lixoft SAS).

Additional statistical analysis. Additional statistical analysis (Besides R studio, see above) was performed using GraphPad Prism version 6.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com. For the statistical tests used, see figure legends.

Figure Legends

Fig. 1 The design and development of the experimental model. **a** Time scale of tumor metastasis in mice implanted orthotopically with RENCA cells. Three different injection modalities were used to develop increasingly aggressive cell lines: Kidney (1), expected to reveal enhancement of mechanisms related to primary tumor formation; Tail (2), expected to recapitulate key aspects that precede metastasis formation (i.e. survival in the blood stream, evasion of host immune response, and distant organ colonization); Lung (3), representing cells which underwent the full cycle of primary tumor formation, dissemination, and secondary tumor formation at a distant site. At the bottom, scheme of different tumor progression phases and mechanisms which are expected to be enhanced in the relevant cell line. **b** the graph shows reduced mice survival, represented as time passed since the implantation until the day of sacrifice, after 6 cycles of RENCA injection. Data are represented as mean ± SEM. **c** Dots graph shows the weight of primary tumors generated by parental (P0), or late passaged Kidney P6 or Lung P6 RENCA cells, after implantation into kidney. Data are represented as mean ± SEM. Stats ANOVA followed with Tukey's multiple comparisons. *$p<0.05$; **$p<0.01$. **d** Representative pictures of primary tumors and lung metastasis generated by orthotopic implant into kidney of either parental (P0) or late passaged cell lines (Kidney P6 or Lung P6). Bar = 1cm. **e** Bars graph showing the percentage of mice presenting visible metastases in the lungs, upon injection of late passaged RENCA (either Kidney P6 or Lung P6) compared to parental (P0) cell line. **f** representative pictures of lung metastases formed, at day 15, by late passaged RENCA cells (Tail P6), compared to parental cells (P0), which were tail injected. To note, upon tail injection, parental cell line P0 requires a wider window of time (i.e. 26 days) to form visible lung metastases.

Fig. 2 Analysis of transcription data sets. **a, b** Projection of samples onto principal component (PC) 1 and 2 for series 2 and 5 respectively. Samples of series 2 did not cluster whereas series 5 did. **c** Heatmap for PC1- et PC2-associated gene expression from the PCA analysis through S0 to S5. A progressive pattern is displayed. **d, e, f** Each enriched GO term is represented in function of the -log10 of its adjusted p-value. Count: number of genes in the GO term. Intersection of enriched GO terms for PC1-andPC2-associated genes from the PCA analysis for S5 (d); Unique enriched GO terms for PC1 (e) or PC2 (f) associated genes from PCA analysis for S5.

Fig. 3 Validation of the "Lung" signature. **a ,b** Kaplan-Meier for overall (OS) and disease-free (DFS) survival analysis stratified in 3 groups of equivalent size. Signature "low": patient group with low score; signature "medium": patient group with a medium score; signature "high": patient group with high score. **c, d** Density plot of p values (log-rank test) from 1000 random signatures of equal size for OS and DFS. p = empirical p-value. **e, f** Forest plot of multivariate Cox proportional hazards model for the Lung signature in the KIRC cohort adjusted for clinical variables (Fuhrman grade and TNM stage).

Fig. 4 Analysis of methylome data sets. **a** Principal component of full-genome methylation analysis for S5 series. **b** Enriched GO terms for PC1/PC2-associated genes from the PCA analysis of the S5 methylation series. Each enriched GO term is represented in function of the -log10 of its adjusted p-value. Count: number of genes in the GO term; **c** Dot-plot of differentially expressed genes (DEG) corresponding to differentially methylated CpGs in the

gene body or the promoter. **d** Intersection between transcription and methylation of enriched GO terms for PC1 and PC2 -associated genes.

Fig. 5 Clinical relevance of SAA2 and CFB after anti-angiogenic treatment (SUVEGIL-TORAVA cohorts). **a, b** Correlation between plasmatic SAA2 levels at diagnosis and survival (OS and PFS) in patients after sunitinib treatment (plasmatic level at the diagnosis less or greater than a cut-off for SAA2 (269 µg/ml)[OS: HR(log-rank)=5.557; PFS: HR(log-rank)=7.669. **c** Correlation between plasmatic SAA2 levels at diagnosis and PFS in patients after sunitinib or bevacizumab treatment (plasmatic level at the diagnosis less or greater than a third quartile cut-off for SAA2 (269 µg/ml; HR(log-rank)=1.987). **d** Correlation between plasmatic CFB levels at diagnosis and PFS in patients after sunitinib or bevacizumab treatment (plasmatic level at the diagnosis less or greater than a third quartile cut-off for CFB (310µg/ml; HR(log-rank)=3.113) **e, f** PFS (**e**) and OS (**f**) patients treated with either Sunitinib of bevacizumab and stratified according to plasma levels of both SAA2 and CFB. Three subgroups were identified *i)* CFB low and SAA2 low, *ii)* CFB low and SAA2 high or CFB high and SAA2 low, *iii)* CFB high and SAA2 high (Low-low vs high-high: OS HR(log-rank)=5.086; PFS HR(log-rank)=4.196).

Fig. 6 Analysis of IL34 expression and its significance for RCC progression. **a** Expression analyses of IL34 gene expression in amplified RENCA cell lines (Kidney, Tail and Lung) where IL34 was strongly upregulated compared to Parental cells (Series 0). Of note, the increase of expression is cell-passage dependent and, thus, related to progression toward a more aggressive phenotype. Graph shows mean ± SEM. **b** IL34 expression in primary tumor cells and in tumor cells from lung metastasis compared to the parental cell line. **c** ELISA assay of IL34 in supernatants in late passaged (S5) cell lines compared to control (Parental S0) of 1 million cells. Graph shows mean ± SEM. **d** Correlation of IL34 expression to overall survival (OS) in patients from TCGA KIRC cohort ((HR(log-rank)=1.857 (1.38-2.498)). **e** IL 34 expression correlated to Disease Free Survival (DFS) in patients from TCGA KIRC cohort. ((HR(log-rank)=2.056 (1.315-3.214)) **f** Correlation of Fuhrman grade with IL34 gene expression in the TCGA KIRC cohort. g. qPCR analyses of IL34 expression in healthy kidney and in tumors from patients (UroCCR). Results as mean and SEM. Analysis by two-tailed paired t-test. P=0.0006. **h** Histological of IL34 in healthy kidneys and in tumors from patients (UroCCR). Scale bar = 50µm. **i** Evaluations of IL34 expression in TMA from RCC patients (UroCCR cohort) (Scoring by a pathologist, D. Ambrosetti). ANOVA followed with tukey's multiple comparisons. **p<0.01 ; ****p<0.0001 **j** Disease Free Survival (DFS) from the TMA IL34 scoring (UroCCR Cohort)HR(log-rank)=1.878 (1.056-3.342).

Fig. 7 Analysis of IL34 overexpression on macrophages. Expression analysis conducted in orthotopic implanted tumors generated from IL34-overexpressing RENCA cells. Expression of markers in IL34-overexpressing tumors for (**a**) total macrophages (F4/80 and CSF1R), (**b**) M1-type macrophages (IL12a and IL1β) and (**c**) M2-type macrophages (MMR and ARG1) compared to control tumors. Graph show mean ± SEM. T-test. *p<0.05; ***p<0.001. **d** Correlation relationship between IL34 and MMR (type 2 macrophage marker). Statistical test: Spearman correlation test. **e** Semi-quantitative analysis of MMR protein expression in mouse tumors generated from IL34-overexpressing and control RENCA cells from western blots (Quantification of band intensities compared to vinculin and using a reference sample). On the right, a representative WB of MMR in IL34-overexpressing and control tumors. T-test. ****p<0.0001. **f** Quantification of MMR-positive cells in the kidney tumors. On the

right, representative images of primary tumors stained by IF using an anti-MMR fluorescent antibody (red) and DAPI (blue) for DNA counterstaining. Scale bar, 100 µm. T-test. ****p<0.0001. **g** Quantification of MMR-positive cells in lung metastasis after tail vein injection. On the right, representative images of lung metastases stained by IF using an anti-MMR fluorescent antibody (red) and DAPI (blue) for DNA counterstaining. Scale bar, 100 µm. T-test. ****p<0.0001.

Fig. 8: Effect of IL34 silencing *in vivo*. **a** Weight from kidney bearing tumors after orthotopic implanted IL34 knock-out (Crispr-Cas9) or control (LacZ) cells represented as mean+-SEM. The pictures on the right shows images of implanted tumors. ANOVA followed by Dunnett's multiple comparisons test. ***p>0.001; ****p<0.0001. **b** Mouse survival (HR(log-rank)=8.09 (1.7-38.497)) and **c** number of mice without lung metastasis (HR(log-rank)=3.839 (0.967-15.244)). **d** Mathematical modeling issued of tumors from sub-capsular implant of IL 34 Crispr or control (LacZ) cells. Volume of primary tumors and the number of lung metastasis were quantified by MRI and fitted to generate simulations for tumor growth and metastasis formation, using a population mixed-effects approach. **e** Metastasis density after tail-vein injections of IL34-silenced and control (LacZ) cells. The graph shows the means of the metastasis area in lung sections relative to the tissue area ± SEM. Examples of macroscopic images of lungs, and lung metastasis from GFP+ RENCA cells are shown . Statistical analysis: ANOVA followed by Dunnett's multiple comparisons test. *p<0.05 ; **p<0.01. **f** Quantification of MMR-positive cells in lung metastasis after the tail vein injection. Representative images of IF staining for MMR (red) and DAPI(blue) are shown. Scale bar = 100µm. Statistics: ANOVA followed by Dunnett's multiple comparisons test. ****p<0.0001

Table 1 Summary table of the signatures and their predictive value in the KIRC TCGA cohort

Supplementary Figure Legends

Supplementary Fig. 1 The figure depicts the general strategy used to determine clinically-relevant signatures from the transcriptomic data.

Supplementary Fig. 2 Validation of K signature in the KIRC TCGA cohort. **a, b** Kaplan-Meier for overall (OS) and disease-free (DFS) survival analysis stratified in 3 groups of equivalent size. Signature "low": patient group with low score; signature "medium": patient group with a medium score; signature "high": patient group with high score. **c, d** Density plot of p values (log-rank test) from 1000 random signatures of equal size for OS and DFS. p = empirical p-value. **e, f** Forest plot of multivariate Cox proportional hazards model for the Lung signature in the KIRC cohort adjusted for clinical variables (Fuhrman grade and TNM stage).

Supplementary Fig. 3 Validation of T signature in the KIRC TCGA cohort. **a, b** Kaplan-Meier for overall (OS) and disease-free (DFS) survival analysis stratified in 3 groups of equivalent size. Signature "low": patient group with low score; signature "medium": patient group with a medium score; signature "high": patient group with high score. **c, d** Density plot of p values (log-rank test) from 1000 random signatures of equal size for OS and DFS. p = empirical p-value. **e, f** Forest plot of multivariate Cox proportional hazards model for the Lung signature in the KIRC cohort adjusted for clinical variables (Fuhrman grade and TNM stage).

Supplementary Fig. 4 Validation of KTL signature in the KIRC TCGA cohort. **a, b** Kaplan-Meier for overall (OS) and disease-free (DFS) survival analysis stratified in 3 groups of equivalent size. Signature "low": patient group with low score; signature "medium": patient group with a medium score; signature "high": patient group with high score. **c, d** Density plot of p values (log-rank test) from 1000 random signatures of equal size for OS and DFS. p = empirical p-value. **e, f** Forest plot of multivariate Cox proportional hazards model for the Lung signature in the KIRC cohort adjusted for clinical variables (Fuhrman grade and TNM stage).

Supplementary Fig. 5 Validation of AllMerged signature in the KIRC TCGA cohort. **a, b** Kaplan-Meier for overall (OS) and disease-free (DFS) survival analysis stratified in 3 groups of equivalent size. Signature "low": patient group with low score; signature "medium": patient group with a medium score; signature "high": patient group with high score. **c, d** Density plot of p values (log-rank test) from 1000 random signatures of equal size for OS and DFS. p = empirical p-value. **e, f** Forest plot of multivariate Cox proportional hazards model for the Lung signature in the KIRC cohort adjusted for clinical variables (Fuhrman grade and TNM stage).

Supplementary Fig. 6 Low-coverage shallow DNA sequencing of the S5 series versus parental cells. 5 Kidney, 3 Lungs and 4 tail samples were sequenced. There is no significant copy number variation in these different groups.

Supplementary Fig. 7 IL34 CpG methylation and clinical relevance. **a** Locus of the identified DMC in intron 1 of IL34 (GRCm38) overlapping an enhancer (ENSEMBL) **b** Locus of the CpGs cg01782798 and cg26831220 in the intron 1 of IL34 (GRCh37)

overlapping an enhancer (GeneHancer); **c, d** Scatterplot comparing the IL34 transcription and methylation level of the CpGs X (f) and Y (g) respectively; **e, f** Kaplan-Meier plots for overall survival stratified in 2 groups (TCGA KIRC cohort). low: patient group with low methylation level; high: patient group with high methylation level. **g** Summary of results. HR (95% CI): Hazard Ratio with 95% Confidence Interval; p value: log-rank test p-value; empirical p value: prognostic performance of the 2 CGs were test against 1-CG random signatures.

Supplementary Fig. 8 SAA2 and clinical relevance. **a** SAA2 expression in Lung cell lines with increasing passages compared to parental cells (Series 0). graph shows means ± SEM. **b** Overall Survival (OS) of patients stratified according to SAA2 expression (TCGA KIRC Cohort). (HR(log-rank)=2.472 (1.835-3.329)) **c** Disease Free Survival (DFS) of patients stratified according to SAA2 expression (TCGA KIRC Cohort)(HR(log-rank)=3.562 (2.278-5.57)). **d, e, f, g** OS and DFS of patients stratified according to SAA2 expression (TCGA KIRC Cohort) and their metastatic status(OS: M0 HR(log-rank)=2.334 (1.591-3.423); M1[NS] HR(log-rank)=1.62 (0.955-2.749) – DFS: M0 HR(log-rank)=3.511 (1.851-6.658); M1[NS] HR(log-rank)=2.055 (1.038-4.07)). **h** Overall Survival (OS) of patients stratified according to SAA2 expression in the UroCCR cohort (HR(log-rank)=2.901 (1.526-5.517)). **i** Disease Free Survival (DFS) of patients stratified according to SAA2 expression in the UroCCR cohort (HR(log-rank)=2.342 (1.211-4.529)). **j** Correlation between SAA2 expression in patients and the Fuhrman grade (UroCCR cohort). **k** Plasma SAA2 levels in non-metastatic (M0) and metastatic (M1) patients at diagnosis in the UroCCR cohort. **l** DFS and correlation with pre-surgery SAA2 plasma levels (UroCCR cohort)(HR(log-rank)=8.191 (2.04-32.891)). **m** OS and correlation with post-surgical SAA2 plasma levels (UroCCR cohort)(HR(log-rank)=Inf (Inf-Inf)[not calculable]).

Supplementary Fig. 9 CFB and Clinical Relevance. **a** CFB expression in Lung and **b**Tail cell lines with increasing passage compared to parental cells (Series 0). graph shows means ± SEM**. c** Correlation of CFB to OS in patients from the TCGA KIRC Cohort.(HR(log-rank)=1.684 (1.251-2.266)) **d** Correlation of CFB to DFS in patients from the TCGA KIRC Cohort (HR(log-rank)=1.935 (1.238-3.025)). **e, f, g, h** Correlation of CFB to OS and DFS in metastatic and non-metastatic patients patients from the TCGA KIRC Cohort (OS: M0 HR(log-rank)=1.616 (1.105-2.363); M1[NS] HR(log-rank)=0.883 (0.514-1.518) – DFS: M0 HR(log-rank)=2.605 (1.379-4.921); M1[NS] HR(log-rank)=0.771 (0.388-1.532)). **i** CFB expression in tumor and healthy tissues from UroCCR cohort. **j** Correlation of CFB to OS in patients from the UroCCR cohort (HR(log-rank)=2.556 (1.24-5.267)). **k** Correlation of CFB to DFS in patients from the UroCCR cohort.(HR(log-rank)=2.846 (1.323-6.123)). **l**, **m**.CFB plasma level in patients from the UroCCR cohort (**l**) before and after (**m**) surgery.

Supplementary Fig. 10 PODNL1 and clinical relevance. **a** PODNL1 expression in kidney cell lines with increasing passages compared to parental cells (Series 0). graph shows means ± SEM. **b** Overall Survival (OS) of patients stratified according to PODNL1 expression (TCGA KIRC Cohort) (HR(log-rank)=2.566 (1.904-3.458))**. c**. Disease Free Survival (DFS) of patients stratified according to PODNL1 expression (TCGA KIRC Cohort)(HR(log-rank)=3.731 (2.385-5.838)). **d, e, f, g** OS and DFS of patients stratified according to PODNL1 expression (TCGA KIRC Cohort) and their metastatic status (OS: M0 HR(log-rank)=2.621 (1.788-3.844); M1[NS] HR(log-rank)=1.633 (0.983-2.711) – DFS: M0 HR(log-rank)=6.528 (3.445-12.37); M1[NS] HR(log-rank)=1.606 (0.844-3.055)). **h** Expression of

PODNL1 in healthy and tumor tissue in the UroCCR cohort **i** Overall survival (OS) of patients stratified according to PODNL1 expression in the UroCCR cohort.(HR(log-rank)=0.764 (0.384-1.516)). **j** Disease Free Survival (DFS) of patients stratified according to PODNL1 expression in the UroCCR cohort (HR(log-rank)=1.933 (0.967-3.866)).

Supplementary Fig. 11 GPRC5a and Clinical Correlations. **a** GPRC5a expression in all cell lines with increasing passages compared to parental cells (Series 0). graph shows means ± SEM. **b** Overall Survival (OS) of patients stratified according to GPRC5a expression (TCGA KIRC Cohort) (HR(log-rank)=2.183 (1.621-2.939)). **c** Disease Free Survival (DFS) of patients stratified according to GPRC5a expression (TCGA KIRC Cohort) (HR(log-rank)=2.884 (1.845-4.508)). **d, e, f, g** OS and DFS of patients stratified according to GPRC5a expression (TCGA KIRC Cohort) and their metastatic status (OS: M0 HR(log-rank)=2.305 (1.577-3.369); M1[NS] HR(log-rank)=1.875 (1.147-3.065) – DFS: M0 HR(log-rank)=5.492 (2.908-10.374); M1[NS] HR(log-rank)=1.908 (1.01-3.604)). **h** Overall survival (OS) of patients stratified according to GPRC5a expression in the UroCCR cohort (HR(log-rank)=0.787 (0.333-1.86)) **i** Disease Free Survival (DFS) of patients stratified according to GPRC5a expression in the UroCCR cohort (HR(log-rank)=2.451 (1.062-5.658)). **j** Correlation of GPRC5a expression in patient tumors and correlation to Fuhrmann grade (UroCCR cohort).

Supplementary Fig. 12 IL34 and clinical correlations. **a, b, c, d.** OS and DFS of patients stratified according to IL34 expression (TCGA KIRC Cohort) and their metastatic status (OS: M0 HR(log-rank)=1.985 (1.359-2.9); M1[NS] HR(log-rank)=1.368 (0.834-2.242) – DFS: M0 HR(log-rank)=3.542 (1.874-6.694); M1[NS] HR(log-rank)=1.38 (0.731-2.606)). **e** IL34 plasmatic levels in patients before and after sunitinib treatment (SUVEGIL-TORAVA cohorts) **f** IL34 mRNA expression in mice implanted with 786-O cells and treated with sunitinib. Specific primers were used for human (left panel) and mouse (right panel) IL34. Results presented as mean± SEM. *p<0.05. ns :Non significant. Unpaired t test.

Supplementary Fig. 13 IL34 overexpression in renal carcinoma cells. **a, b.** Validation of IL34 expression by qPCR (**a**) and western-blot (**b**) in RENCA cells transduced with lentiviral vector (pLenti CMV/TO-mIL34v1-Puro) encoding for mouse IL34. In **c**, The IL34 ELISA assay from supernatants of transduced cells is shown. **d** Tumor weights of IL34 overexpressing and control tumors. **e, f** Validation of IL34 expression in IL34 overexpressing tumors by qPCR (e) (T-test. ***p<0.001) and Western Blot (f). g. Correlation between IL34 mRNA and CSF1R mRNA (global macrophage marker). And Spearman correlation test. **h** CSF1 mRNA expression in IL34 overexpressing and control tumors. **i** Quantification of the metastasis area in lung sections after tail vein injection of IL34 overexpressing or control cells. T-test (ns).

Supplementary Fig. 14 Validation of IL34 knock-out and mathematical modeling. **a** validation of IL34 knock-down by ELISA in conditioned media of RENCA cells. **b** Incucyte proliferation assay with crispr IL34-ko (IL341a, IL341b) and control cells (LacZ) with no significant differences (Dunnett's multiple comparison tests on all timepoints). **c** mathematical modeling of primary tumor growth kinetics for 7 different animals.

Supplementary Table 1 Summary table of up and downregulated genes from the first biomarker discovery step.

Supplementary Table 2 List of genes present for each signature with details of the predictive value

Supplementary Table 3 Parameter values of the mathematical model of primary tumor growth and metastatic dissemination. A mixed-effects approach was used and associated likelihood maximization for the pooled data (all animals from all groups). Belonging to the IL34 KO group was considered by means of covariates on parameters α and μ. S.D. = standard deviation of the random effects. R.S.E = relative standard error of the estimate. p-value refers to a Wald test for statistical significance of the covariate.

References

1. Hsieh, J. J. *et al.* Renal cell carcinoma. *Nat. Rev. Dis. Prim.* **3**, 17009 (2017).

2. Ljungberg, B. *et al.* The epidemiology of renal cell carcinoma. *European Urology* (2011). doi:10.1016/j.eururo.2011.06.049

3. American Cancer Society. Key Statistics about kidney cancer. (2016). Available at: http://www.cancer.org/cancer/kidney-cancer.html. (Accessed: 16th December 2019)

4. Chow, W. H., Dong, L. M. & Devesa, S. S. Epidemiology and risk factors for kidney cancer. *Nat Rev Urol* **7**, 245–257 (2010).

5. Gandaglia, G. *et al.* Contemporary incidence and mortality rates of kidney cancer in the united states. *J. Can. Urol. Assoc.* (2014). doi:10.5489/cuaj.1760

6. K.-H., T. *et al.* Prognostic indicators for renal cell carcinoma: A multivariate analysis of 643 patients using the revised 1997 TNM staging criteria. *J. Urol.* (2000).

7. Lam, J. S. *et al.* Postoperative surveillance protocol for patients with localized and locally advanced renal cell carcinoma based on a validated prognostic nomogram and risk group stratification system. *J. Urol.* (2005). doi:10.1097/01.ju.0000165572.38887.da

8. Antonelli, A. *et al.* The follow-up management of non-metastatic renal cell carcinoma: Definition of a surveillance protocol. *BJU Int.* (2007). doi:10.1111/j.1464-410X.2006.06616.x

9. Ö., A. & Ö., Ş. Surgical Management of Local Recurrences of Renal Cell Carcinoma. *Surg. Res. Pract.* (2016). doi:10.1155/2016/2394942 LK

10. Chin, A. I., Lam, J. S., Figlin, R. A. & Belldegrun, A. S. Surveillance strategies for renal cell carcinoma patients following nephrectomy. *Rev. Urol.* (2006).

11. Molina, A. M. & Motzer, R. J. Clinical Practice Guidelines for the Treatment of Metastatic Renal Cell Carcinoma: Today and Tomorrow. *Oncologist* (2011). doi:10.1634/theoncologist.2011-s2-45

12. Motzer, R. J. *et al.* Nivolumab plus Ipilimumab versus Sunitinib in advanced renal-cell carcinoma. *N. Engl. J. Med.* (2018). doi:10.1056/NEJMoa1712126

13. Escudier, B. *et al.* Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **30**, 706–720 (2019).

14. Zisman, A. *et al.* Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected renal cell carcinoma. *J. Clin. Oncol.* (2002). doi:10.1200/JCO.2002.05.111

15. Patard, J. J. *et al.* Use of the University of California Los Angeles integrated staging system to predict survival in renal cell carcinoma: An international multicenter study. *J. Clin. Oncol.* (2004). doi:10.1200/JCO.2004.09.104

16. Ficarra, V. *et al.* External validation of the mayo clinic stage, size, grade and necrosis (SSIGN) score to predict cancer specific survival using a European series of conventional renal cell carcinoma. *J. Urol.* (2006). doi:10.1016/S0022-5347(05)00684-1

17. Zigeuner, R. *et al.* External Validation of the Mayo Clinic Stage, Size, Grade, and Necrosis (SSIGN) Score for Clear-Cell Renal Cell Carcinoma in a Single European Centre Applying Routine Pathology. *Eur. Urol.* (2010). doi:10.1016/j.eururo.2008.11.033

18. Heng, D. Y. C. *et al.* Consortium prognostic model : a population-based study. *Lancet. Oncol.* (2014). doi:10.1016/S1470-2045(12)70559-4.External

19. Bensalah, K. *et al.* Recommandations en onco-urologie 2016-2018 du CCAFU : Cancer du rein. *Progrès en Urol.* **27**, S27–S51 (2016).

20. Bensalah, K. *et al.* Recommandations françaises du Comité de Cancérologie de l'AFU – Actualisation 2018–2020 : prise en charge du cancer du rein. *Progrès en Urol.* **28**, R5–R33 (2018).

21. Fidler, I. J. Selection of successive tumour lines for metastasis. *Nat. New Biol.* (1973). doi:10.1038/newbio242148a0

22. Bruns, C. J., Harbison, M. T., Kuniyasu, H., Eue, I. & Fidler, I. J. In vivo selection and characterization of metastatic variants from human pancreatic adenocarcinoma by using orthotopic implantation in nude mice. *Neoplasia* (1999). doi:10.1038/sj.neo.7900005

23. Murphy, G. P. & Hrushesky, W. J. A murine renal cell carcinoma. *J. Natl. Cancer Inst.* (1973). doi:10.1093/jnci/50.4.1013

24. Hansen, M. T. *et al.* A link between inflammation and metastasis: serum amyloid A1 and A3 induce metastasis, and are targets of metastasis-inducing S100A4. *Oncogene* **34**, 424–435 (2015).

25. Teng, C. & Zheng, H. Low expression of microRNA-1908 predicts a poor prognosis for patients with ovarian cancer. *Oncol. Lett.* **14**, 4277–4281 (2017).

26. Yan, Y. *et al.* SAMSN1 is highly expressed and associated with a poor survival in glioblastoma multiforme. *PLoS One* **8**, e81905 (2013).

27. Sawada, Y. *et al.* GPRC5A facilitates cell proliferation through cell cycle regulation and correlates with bone metastasis in prostate cancer. *Int. J. Cancer* ijc.32554 (2019). doi:10.1002/ijc.32554

28. Jiang, X. *et al.* GPRC5A: An Emerging Biomarker in Human Cancer. *Biomed Res. Int.* **2018**, 1823726 (2018).

29. Tracz, A., Mastri, M., Lee, C. R., Pili, R. & Ebos, J. M. L. Modeling spontaneous metastatic renal cell carcinoma (mRCC) in mice following nephrectomy. *J. Vis. Exp.* (2014). doi:10.3791/51485

30. Di Martino, S. *et al.* Renal cancer: New models and approach for personalizing therapy. *J. Exp. Clin. Cancer Res.* (2018). doi:10.1186/s13046-018-0874-4

31. Hillman, G. G. Experimental Animal Models for Renal Cell Carcinoma. 493–505 (2002). doi:10.1007/978-1-59259-100-8_26

32. Schueler, J. *et al.* Patient derived renal cell carcinoma xenografts exhibit distinct sensitivity patterns in response to antiangiogenic therapy and constitute a suitable tool for biomarker development. *Oncotarget* **9**, 30946–30961 (2018).

33. Patel, A. *et al.* Patient-derived xenograft models to optimize kidney cancer therapies. *Translational Andrology and Urology* (2019). doi:10.21037/tau.2018.11.04

34. Hou, W. & Ji, Z. Generation of autochthonous mouse models of clear cell renal cell carcinoma: mouse models of renal cell carcinoma. *Exp. Mol. Med.* **50**, 30 (2018).

35. Bailey, S. T. *et al.* MYC activation cooperates with Vhl and Ink4a/Arf loss to induce clear cell renal cell carcinoma. *Nat. Commun.* **8**, 15770 (2017).

36. Gu, Y.-F. *et al.* Modeling Renal Cell Carcinoma in Mice: *Bap1* and *Pbrm1* Inactivation Drive Tumor Grade. *Cancer Discov.* **7**, 900–917 (2017).

37. Beuselinck, B. *et al.* Molecular Subtypes of Clear Cell Renal Cell Carcinoma Are Associated with Sunitinib Response in the Metastatic Setting. *Clin. Cancer Res.* **21**, 1329–1339 (2015).

38. Verbiest, A. *et al.* Molecular Subtypes of Clear Cell Renal Cell Carcinoma Are Associated With Outcome During Pazopanib Therapy in the Metastatic Setting. *Clin. Genitourin. Cancer* **16**, e605–e612 (2018).

39. Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.* **23**, 313-326.e5 (2018).

40. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173**, 581-594.e12 (2018).

41. Wood, S. L. *et al.* Association of serum amyloid A protein and peptide fragments with prognosis in renal cancer. *Br. J. Cancer* **103**, 101–111 (2010).

42. Vermaat, J. S. *et al.* Two-protein signature of novel serological markers apolipoprotein-A2 and serum amyloid alpha predicts prognosis in patients with metastatic renal

cell cancer and improves the currently used prognostic survival models. *Ann. Oncol.* (2009). doi:10.1093/annonc/mdp559

43. Vermaat, J. S. *et al.* Validation of Serum Amyloid α as an Independent Biomarker for Progression-Free and Overall Survival in Metastatic Renal Cell Cancer Patients. *Eur. Urol.* **62**, 685–695 (2012).

44. Kim, S. H. *et al.* Prognostic potential of the preoperative plasma complement factor B in resected pancreatic cancer: A pilot study. *Cancer Biomarkers* **24**, 335–342 (2019).

45. Zhou, H. & Rigoutsos, I. The emerging roles of GPRC5A in diseases. *Oncoscience* **1**, 765 (2014).

46. Jin, E. *et al.* Clinical significance of reduced GPRC5A expression in surgically resected non-small cell lung cancer. *Oncol. Lett.* (2018). doi:10.3892/ol.2018.9537

47. Borys, A. M. *et al.* Patterns of gene expression characterize T1 and T3 clear cell renal cell carcinoma subtypes. *PLoS One* **14**, e0216793 (2019).

48. Greenhough, A. *et al.* Cancer cell adaptation to hypoxia involves a HIF-GPRC5A-YAP axis. *EMBO Mol. Med.* **10**, (2018).

49. Lin, H. *et al.* Discovery of a cytokine and its receptor by functional screening of the extracellular proteome. *Science (80-.).* (2008). doi:10.1126/science.1154370

50. Nandi, S. *et al.* Receptor-type Protein-tyrosine Phosphatase ζ Is a Functional Receptor for Interleukin-34. *J. Biol. Chem.* **288**, 21972–21986 (2013).

51. Segaliny, A. I. *et al.* Syndecan-1 regulates the biological activities of interleukin-34. *Biochim. Biophys. Acta - Mol. Cell Res.* (2015). doi:10.1016/j.bbamcr.2015.01.023

52. Baek, J.-H. H. *et al.* IL-34 mediates acute kidney injury and worsens subsequent chronic kidney disease. *J. Clin. Invest.* **125**, 3198–214 (2015).

53. Foucher, E. D. *et al.* IL-34 Induces the Differentiation of Human Monocytes into Immunosuppressive Macrophages. Antagonistic Effects of GM-CSF and IFNγ. *PLoS One* **8**, (2013).

54. Boulakirba, S. *et al.* IL-34 and CSF-1 display an equivalent macrophage differentiation ability but a different polarization potential. *Sci. Rep.* **8**, 256 (2018).

55. Chang, S. H. *et al.* Baseline serum interleukin-34 levels independently predict radiographic progression in patients with rheumatoid arthritis. *Rheumatol. Int.* (2015). doi:10.1007/s00296-014-3056-5

56. Preisser, L. *et al.* IL-34 and macrophage colony-stimulating factor are overexpressed in hepatitis C virus fibrosis and induce profibrotic macrophages that promote collagen synthesis by hepatic stellate cells. *Hepatology* (2014). doi:10.1002/hep.27328

57. Wada, Y. *et al.* IL-34-Dependent intrarenal and systemic mechanisms promote lupus nephritis in MRL-Faslpr mice. *J. Am. Soc. Nephrol.* (2019). doi:10.1681/ASN.2018090901

58. Ségaliny, A. I. *et al.* Interleukin-34 promotes tumor progression and metastatic process in osteosarcoma through induction of angiogenesis and macrophage recruitment. *Int. J. Cancer* **137**, 73–85 (2015).

59. Baghdadi, M. *et al.* High co-expression of IL-34 and M-CSF correlates with tumor progression and poor survival in lung cancers. *Sci. Rep.* **8**, 418 (2018).

60. Guadagno, E. *et al.* Role of Macrophages in Brain Tumor Growth and Progression. *Int. J. Mol. Sci.* **19**, 1005 (2018).

61. Franklin, R. A. *et al.* The cellular and molecular origin of tumor-associated macrophages. *Science (80-.).* **344**, 921–925 (2014).

62. Sica, A., Invernizzi, P. & Mantovani, A. Macrophage plasticity and polarization in liver homeostasis and pathology. *Hepatology* **59**, 2034–2042 (2014).

63. Comito, G. *et al.* Cancer-associated fibroblasts and M2-polarized macrophages synergize during prostate carcinoma progression. *Oncogene* **33**, 2423–2431 (2014).

64. Funada, Y. *et al.* Prognostic significance of CD8+ T cell and macrophage peritumoral infiltration in colorectal cancer. *Oncol. Rep.* **10**, 309–13

65. Kurahara, H. *et al.* Significance of M2-Polarized Tumor-Associated Macrophage in Pancreatic Cancer. *J. Surg. Res.* **167**, e211–e219 (2011).

66. Ryder, M., Ghossein, R. A., Ricarte-Filho, J. C. M., Knauf, J. A. & Fagin, J. A. Increased density of tumor-associated macrophages is associated with decreased survival in advanced thyroid cancer. *Endocr. Relat. Cancer* **15**, 1069–1074 (2008).

67. Lin, W. *et al.* Function of CSF1 and IL34 in Macrophage Homeostasis, Inflammation, and Cancer. *Front. Immunol.* **10**, 2019 (2019).

68. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).

69. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci.* **107**, 16910–16915 (2010).

70. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

71. R Foundation for Statistical Computing. *R: a Language and Environment for Statistical Computing. http://www.R-project.org/* (2018).

72. Rstudio Team. RStudio: Integrated development for R. RStudio, Inc., Boston MA. *RStudio* (2016). doi:10.1007/978-3-642-20966-6

73. Goldman, M. *et al.* The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv* (2019). doi:10.1101/326470

74. Aken, B. L. *et al.* Ensembl 2017. *Nucleic Acids Res.* **45**, D635–D642 (2017).

75. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011). doi:10.14806/ej.17.1.200

76. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* (2011). doi:10.1093/bioinformatics/btr167

77. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp352

78. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Scaling read aligners to hundreds threads Gen. Process.* (2017). doi:10.1101/205328

79. Karolchik, D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* (2004). doi:10.1093/nar/gkh103

80. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* (2009). doi:10.1038/nprot.2009.97

81. Akalin, A. *et al.* MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* (2012). doi:10.1186/gb-2012-13-10-R87

82. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* (2002). doi:10.1101/gr.229102

83. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford).* (2017). doi:10.1093/database/bax028

84. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* (2012). doi:10.1089/omi.2011.0118

85. Bult, C. J. *et al.* Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gky1056

86. Négrier, S. *et al.* Temsirolimus and bevacizumab, or sunitinib, or interferon alfa and bevacizumab for patients with advanced renal cell carcinoma (TORAVA): a randomised phase 2 trial. *Lancet Oncol.* **12**, 673–680 (2011).

87. Ribot, E. J. *et al.* Self-gated bSSFP sequences to detect iron-labeled cancer cells and/or metastases in vivo in mouse liver at 7 Tesla. *J. Magn. Reson. Imaging* (2015). doi:10.1002/jmri.24688

88. Spratt, J. S., Meyer, J. S. & Spratt, J. A. Rates of growth of human solid neoplasms: Part I. *Journal of Surgical Oncology* (1995). doi:10.1002/jso.2930600216

89. Benzekry, S. *et al.* Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth. *PLoS Comput. Biol.* (2014). doi:10.1371/journal.pcbi.1003800

90. Iwata, K., Kawasaki, K. & Shigesada, N. A dynamical model for the growth and size distribution of multiple metastatic tumors. *J. Theor. Biol.* (2000). doi:10.1006/jtbi.2000.1075

91. Lavielle, M. *Mixed Effects Models for the Population Approach*. *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools* (Chapman and Hall/CRC, 2014). doi:10.1201/b17203

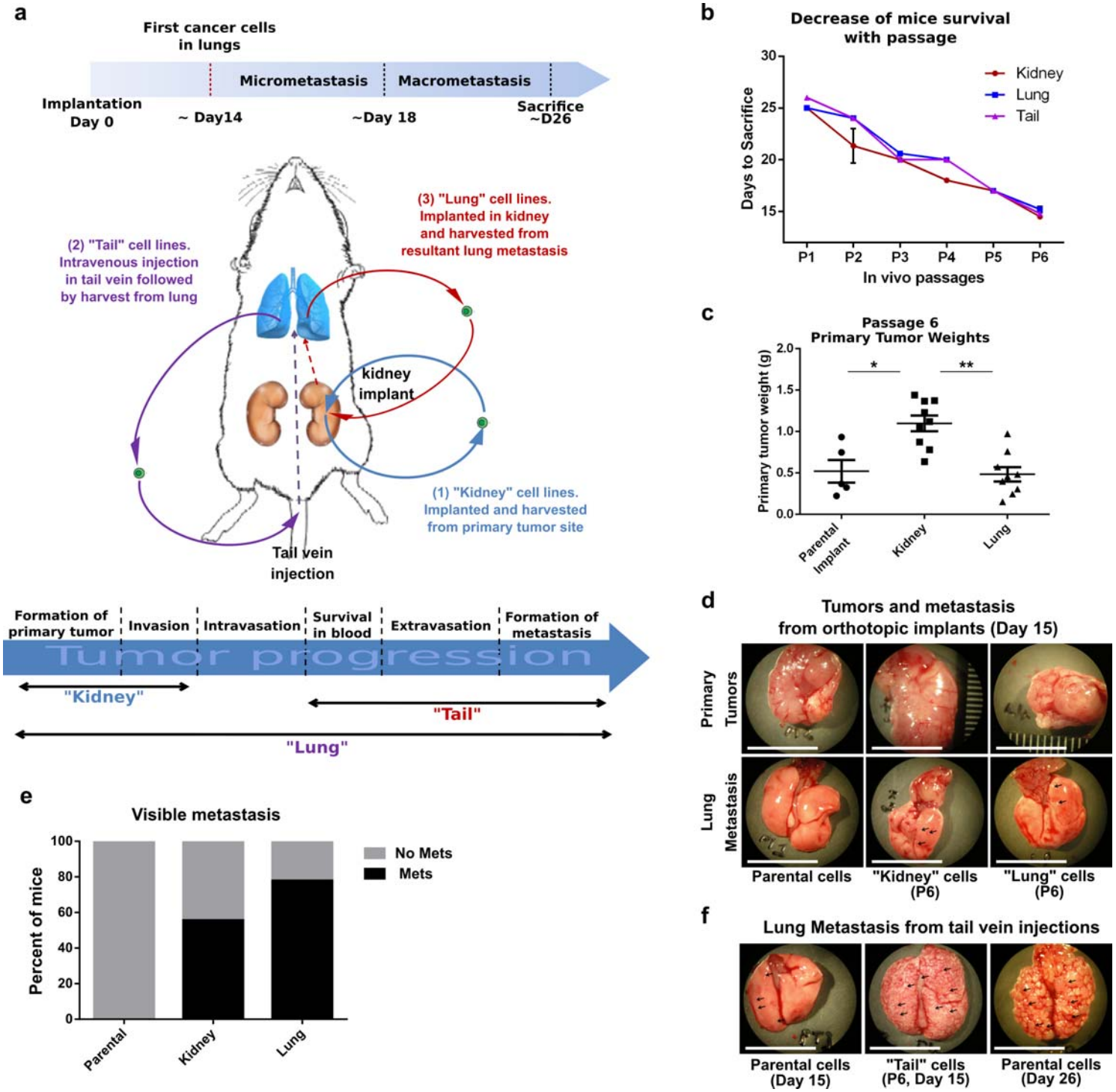

Fig. 1 The design and development of the experimental model.

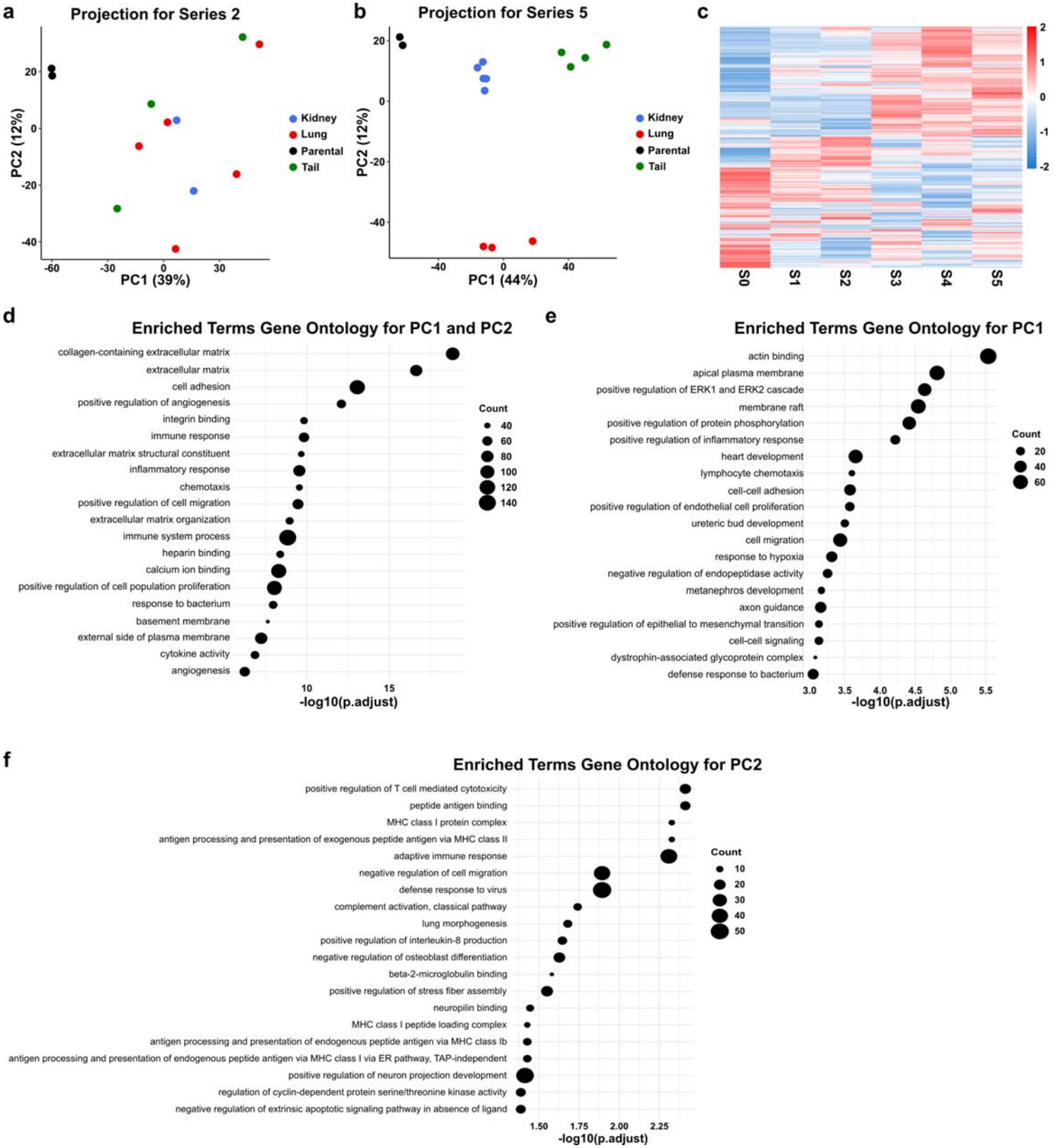

Fig. 2 Analysis of transcription data sets.

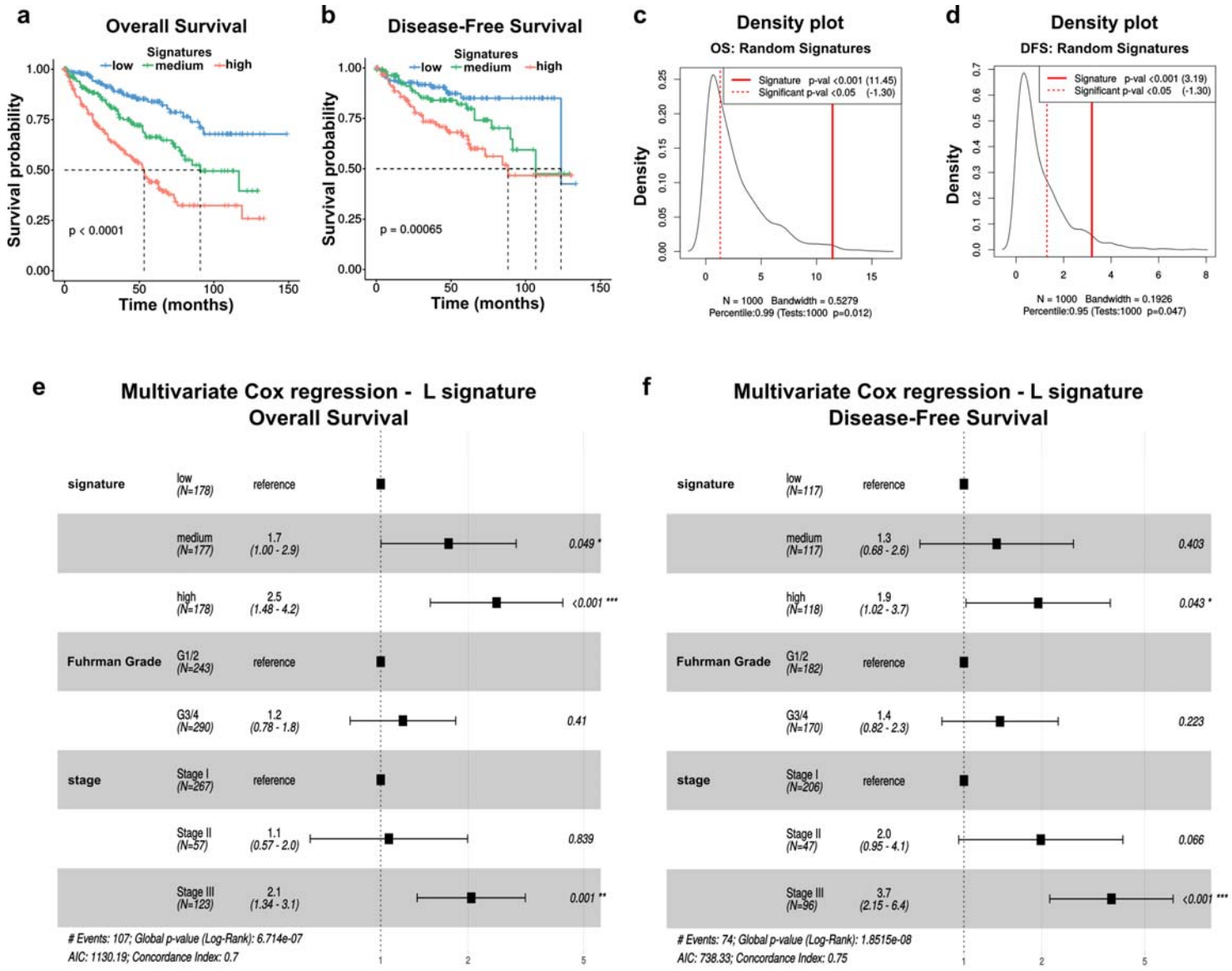

Fig. 3 Validation of the "Lung" signature.

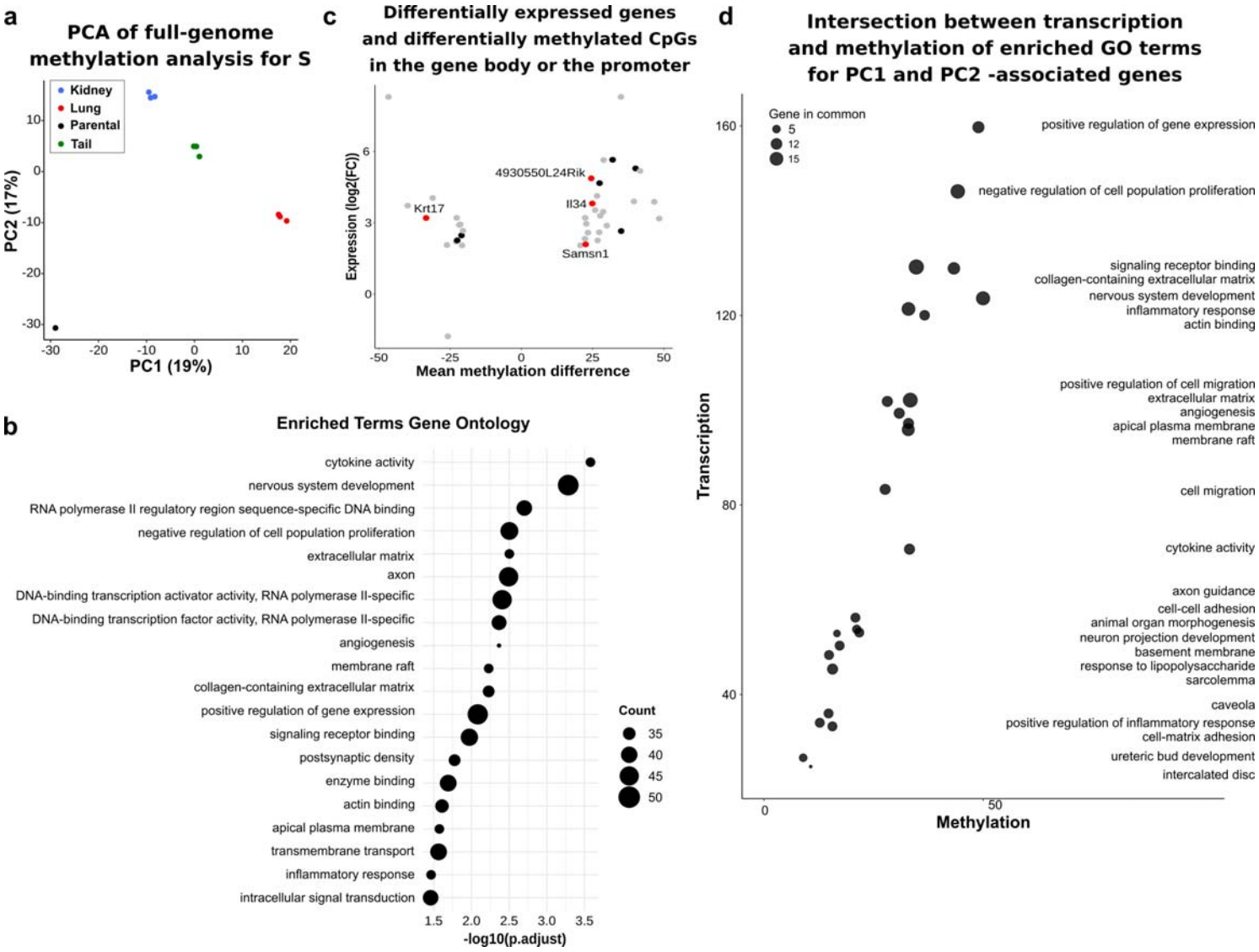

Fig. 4 Analysis of methylome data sets.

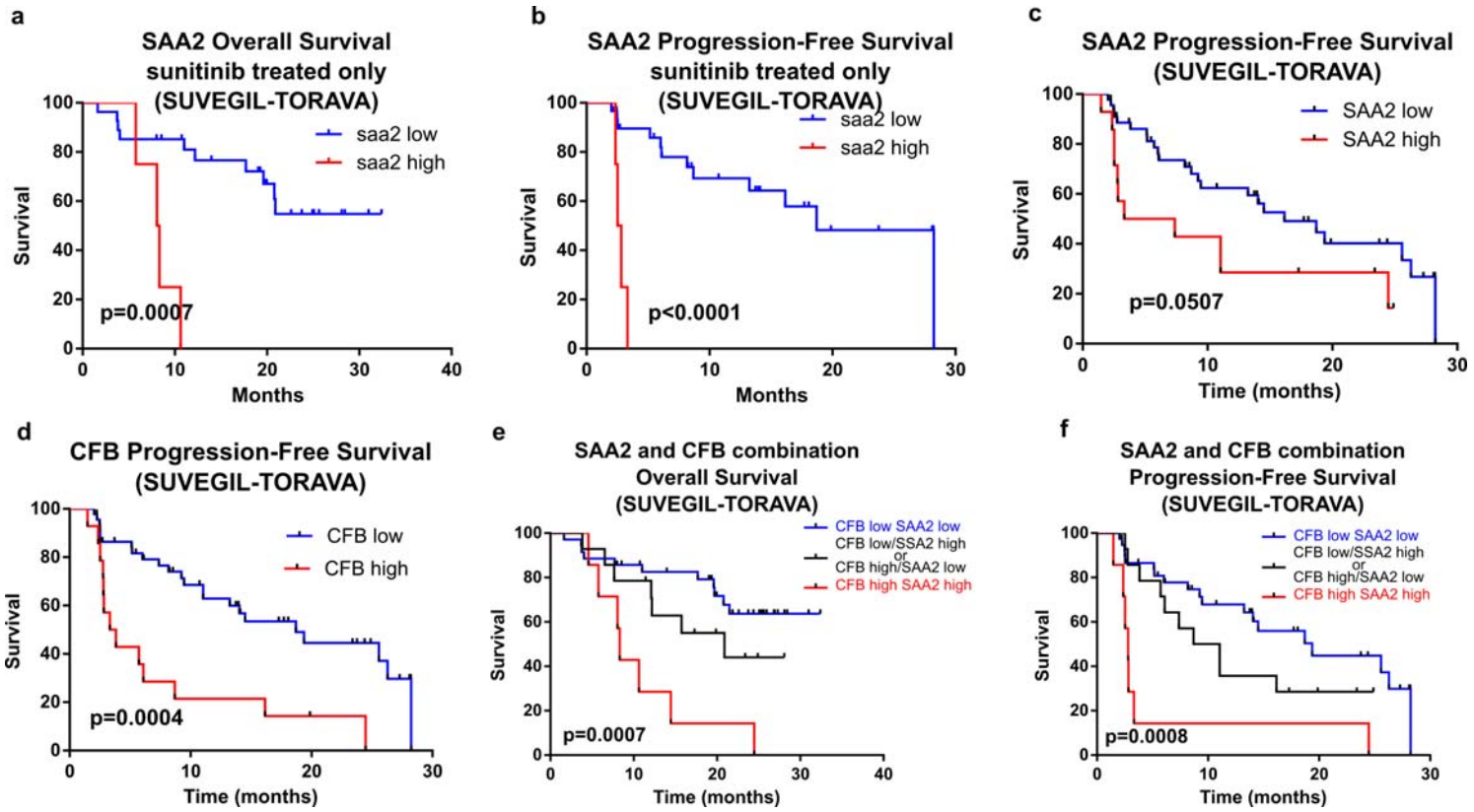

Fig. 5 Clinical relevance of SAA2 and CFB after anti-angiogenic treatment (SUVEGIL-TORAVA cohorts).

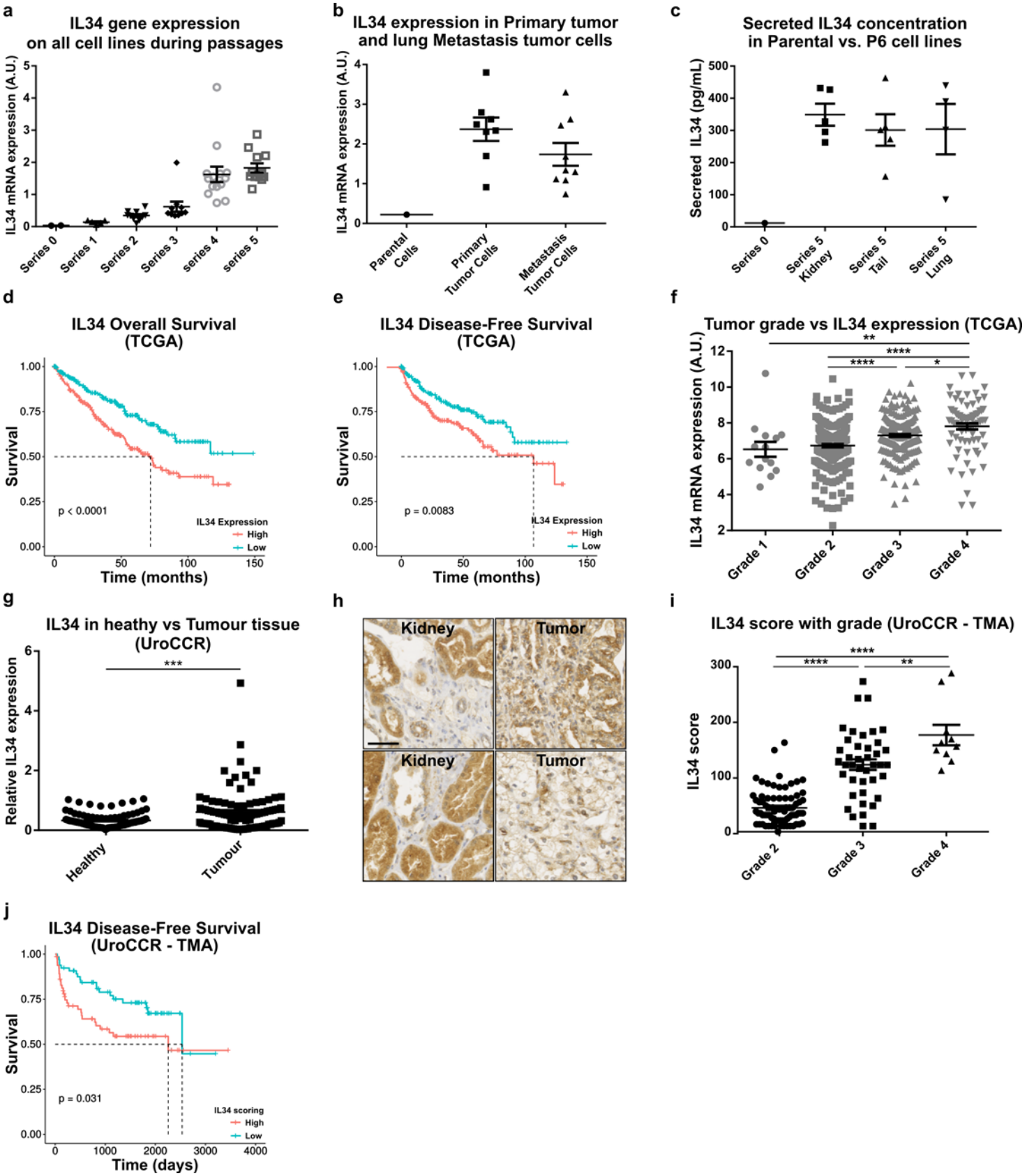

Fig. 6 Analysis of IL34 expression and its significance for RCC progression.

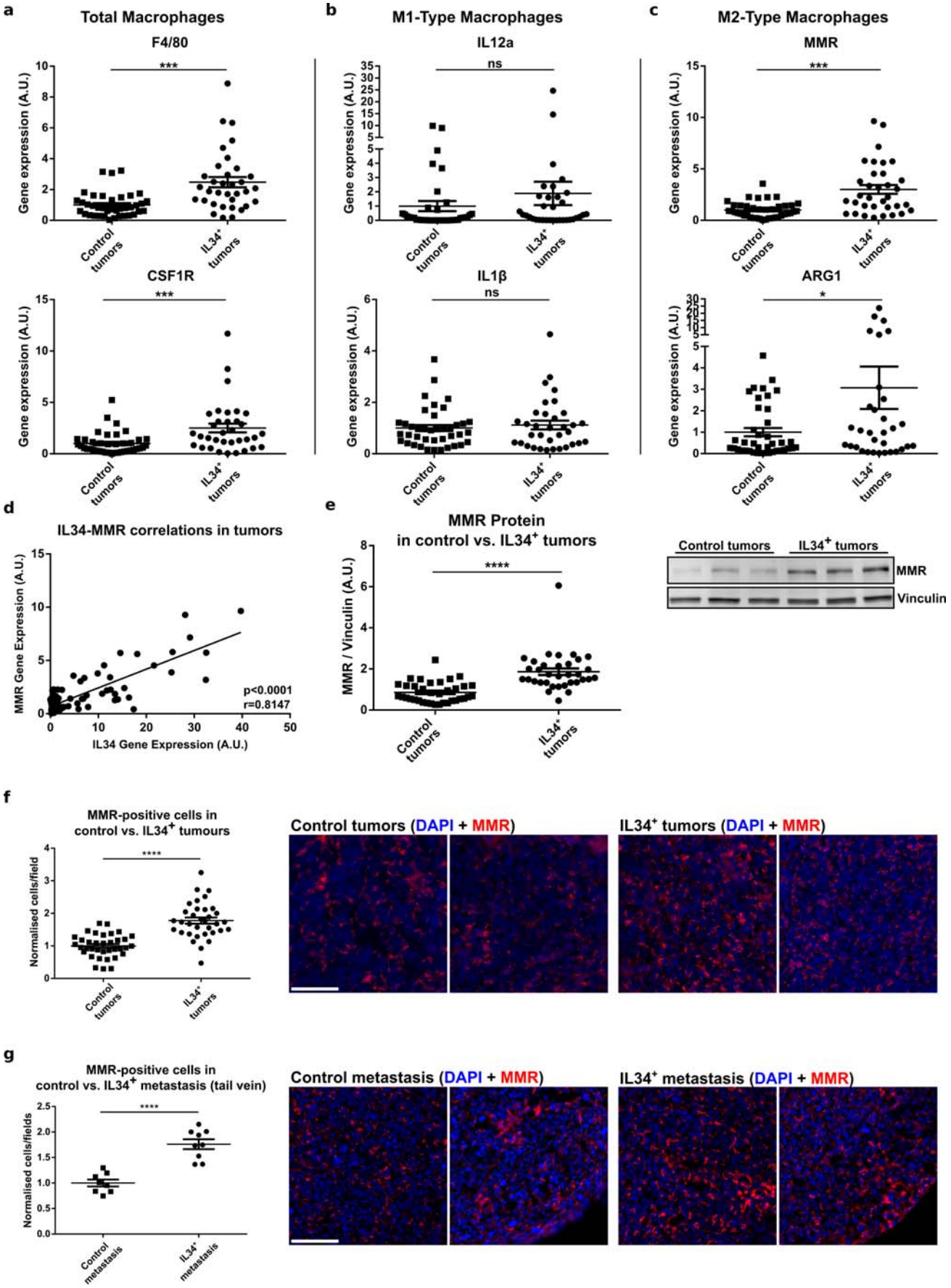

Fig. 7 Analysis of IL34 overexpression on macrophages.

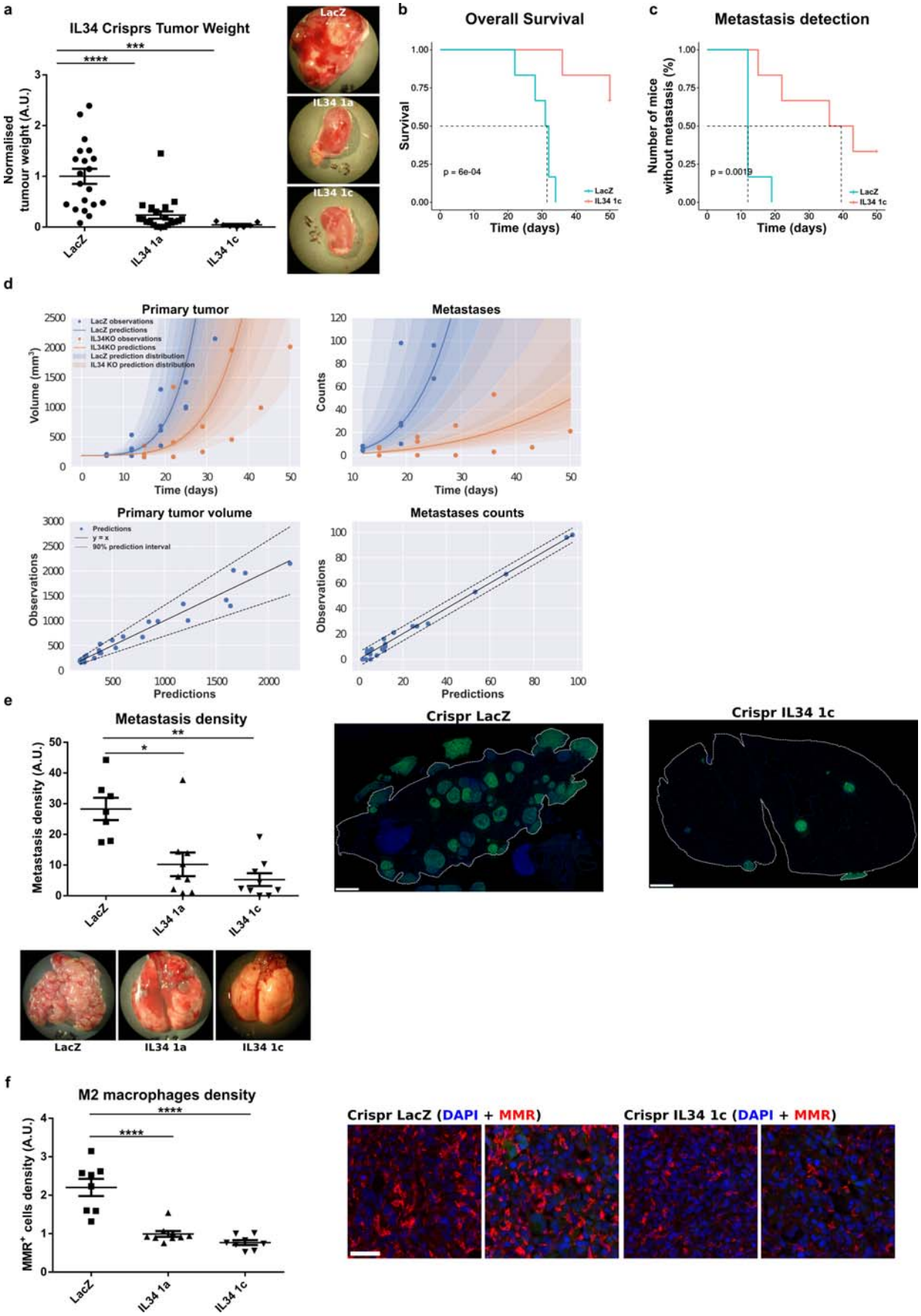

Fig. 8: Effect of IL34 silencing *in vivo*.

Signature	Gene	Count	OS				DFS			
			HR-logRank (95% CI)	p value 2 arms	p value 3 arms	Empirical P value	HR-logRank (95% CI)	p value 2 arms	p value 3 arms	Empirical P value
K	IL34, MAGEB2, TSPAN32, WAS	4	2.457 (1.74-3.471)	5.139E-07	1.250E-07	0.075	1.746 (1.024-2.979)	4.166E-02	6.937E-02	0.323
T	BASP1, CRABP2, DNAJC12, DPYSL3, F3, GLIPR2, HMGA2, MAGEC2, MAGEC3, PLAC8, PPEF1, RAB15, SLC11A1, SNAI2	14	3.467 (2.45-4.906)	2.188E-11	1.100E-12	0.006	2.683 (1.586-4.537)	3.682E-04	3.682E-04	0.033
L	ATP1A3, CCL8, CFB, IL2RG, KRT13, KRT17, PCBP3, PLAC8, RAC2, SAA1, SAA2, SAMSN1, SERPINA3, SLC2A3, UCK2, WT1	16	4.011 (2.809-5.727)	2.346E-12	3.568E-12	0.012	2.94 (1.716-5.039)	2.756E-04	6.453E-04	0.047
KTL	CCL8, CFB, DNAJC12, IL2RG, IL34, MAGEC2, MAGEC3, PLAC8, SERPINA3, CCL8, CFB, DNAJC12, IL2RG, IL34, MAGEC2, MAGEC3, PLAC8, SERPINA3	9	3.597 (2.542-5.088)	4.882E-11	6.730E-12	0.016	2.894 (1.689-4.958)	3.452E-04	1.071E-03	0.089
allMerged	ATP1A3, BASP1, CCL8, CFB, CRABP2, DNAJC12, DPYSL3, F3, GLIPR2, HMGA2, IL2RG, IL34, KRT13, KRT17, MAGEB2, MAGEC2, MAGEC3, PCBP3, PLAC8, PPEF1, RAB15, RAC2, SAA1, SAA2, SAMSN1, SERPINA3, SLC11A1, SLC2A3, SNAI2, TSPAN32, UCK2, WAS, WT1	33	4.583 (3.199-6.565)	7.350E-14	1.213E-13	0.004	2.975 (1.77-5.001)	1.364E-04	3.886E-05	0.012

Table 1 Summary table of the signatures and their predictive value in the KIRC TCGA cohort.