

RESEARCH ARTICLE

Open Access



Overcoming intratumoural heterogeneity for reproducible molecular risk stratification: a case study in advanced kidney cancer

Alexander L. R. Lubbock^{1,7}, Grant D. Stewart^{1,2,8}, Fiach C. O'Mahony^{1,2}, Alexander Laird^{1,2}, Peter Mullen³, Marie O'Donnell^{2,4}, Thomas Powles⁵, David J. Harrison^{2,3} and Ian M. Overton^{1,6*}

Abstract

Background: Metastatic clear cell renal cell cancer (mccRCC) portends a poor prognosis and urgently requires better clinical tools for prognostication as well as for prediction of response to treatment. Considerable investment in molecular risk stratification has sought to overcome the performance ceiling encountered by methods restricted to traditional clinical parameters. However, replication of results has proven challenging, and intratumoural heterogeneity (ITH) may confound attempts at tissue-based stratification.

Methods: We investigated the influence of confounding ITH on the performance of a novel molecular prognostic model, enabled by pathologist-guided multiregion sampling ($n = 183$) of geographically separated mccRCC cohorts from the SuMR trial (development, $n = 22$) and the SCOTRRCC study (validation, $n = 22$). Tumour protein levels quantified by reverse phase protein array (RPPA) were investigated alongside clinical variables. Regularised wrapper selection identified features for Cox multivariate analysis with overall survival as the primary endpoint.

Results: The optimal subset of variables in the final stratification model consisted of N-cadherin, EPCAM, Age, mTOR (NEAT). Risk groups from NEAT had a markedly different prognosis in the validation cohort (log-rank $p = 7.62 \times 10^{-7}$; hazard ratio (HR) 37.9, 95% confidence interval 4.1–353.8) and 2-year survival rates (accuracy = 82%, Matthews correlation coefficient = 0.62). Comparisons with established clinico-pathological scores suggest favourable performance for NEAT (Net reclassification improvement 7.1% vs International Metastatic Database Consortium score, 25.4% vs Memorial Sloan Kettering Cancer Center score). Limitations include the relatively small cohorts and associated wide confidence intervals on predictive performance. Our multiregion sampling approach enabled investigation of NEAT validation when limiting the number of samples analysed per tumour, which significantly degraded performance. Indeed, sample selection could change risk group assignment for 64% of patients, and prognostication with one sample per patient performed only slightly better than random expectation (median logHR = 0.109). Low grade tissue was associated with 3.5-fold greater variation in predicted risk than high grade ($p = 0.044$).

(Continued on next page)

* Correspondence: ian.overton@ed.ac.uk

Alexander L. R. Lubbock and Grant D. Stewart share joint first authorship.

David J. Harrison and Ian M. Overton share joint senior authorship.

¹MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

⁶Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH16 4UX, UK

Full list of author information is available at the end of the article



(Continued from previous page)

Conclusions: This case study in mcrRCC quantitatively demonstrates the critical importance of tumour sampling for the success of molecular biomarker studies research where ITH is a factor. The NEAT model shows promise for mcrRCC prognostication and warrants follow-up in larger cohorts. Our work evidences actionable parameters to guide sample collection (tumour coverage, size, grade) to inform the development of reproducible molecular risk stratification methods.

Keywords: Cancer, Tumour heterogeneity, Prognostic markers, Renal cell carcinoma, Tumour biomarkers

Background

There is a great unmet need for better treatment and diagnosis of kidney cancer, which remains the most lethal of all genitourinary malignancies. Five-year survival in renal cell cancer (RCC) is approximately 40% overall, 10% in metastatic disease [1, 2]. Clear cell RCC (ccRCC) represents about 80% of cases, and around one-third of patients present with metastasis. Current risk stratification of advanced ccRCC uses clinico-pathological scoring systems, for example, the International Metastatic Database Consortium (IMDC) [3] and Memorial Sloan Kettering Cancer Center (MSKCC) [4] scores. Molecular markers promise to overcome the performance plateau encountered by clinico-pathological variables; however, success rates have historically been low [5–8].

Sunitinib is a first-line treatment for metastatic ccRCC (mcrRCC), doubling median progression-free survival compared with older immunotherapies such as IL-2 and interferon- α [9, 10]. Sunitinib targets tumour, endothelial cells and pericytes, where the mechanism of action includes competitive inhibition of multiple receptor tyrosine kinases (RTKs) [11, 12]. Up to 70% of patients treated with sunitinib show little or no tumour response [10], although they may derive a survival benefit, despite incurring significant toxicity. Improved algorithms are critically needed to guide treatment decisions for current and emerging modalities [6, 7, 13].

Advances in prediction of treatment response and prognostication may be severely hindered by intratumoural heterogeneity (ITH) [14–16]. Indeed, percutaneous biopsy of mcrRCC is a poor guide for pathological assessment of prognostic features [17]. Development of tumour sampling approaches to capture ITH is key for discovery and validation of candidate molecular risk stratification algorithms [6, 7, 13, 15]. We studied protein expression ITH in the context of mcrRCC risk stratification, controlling for clinical variables, and developed a novel prognostic model (NEAT, for N-cadherin, EPCAM, Age, mTOR) that compares well with established clinico-pathological scores. The variables selected in NEAT inform mcrRCC biology and suggest sunitinib action directly on tumour growth signaling. We quantitatively show a dramatic effect of tumour sampling on NEAT performance in a validation cohort receiving current standard treatment and demonstrate parameters pertinent to the development of molecular

diagnostic tools for cancer medicine. We present recommendations that guide tumour sample selection for biomarker research in order to overcome variability in the presence of ITH. Indeed, sampling protocols may determine the success or failure of attempts to validate molecular biomarkers where ITH is a factor.

Methods

Cohorts and tissue samples

This study examined two geographically separated cohorts of mcrRCC patients with multiregion tumour sampling (Table 1). Excluding necrotic tissue, 108 and 75 fresh-frozen samples respectively were analysed from development and validation cohorts. The development cohort was drawn from the SuMR phase II clinical trial of upfront sunitinib (NCT01024205, $n = 22$, London [18]). The validation cohort were cytoreductive nephrectomy patients from the SCOTRRCC study and received standard of care treatment (validation, $n = 22$, Scotland [1, 19]). The development cohort received three cycles of sunitinib 50 mg (4 weeks on, 2 weeks off) prior to nephrectomy; following nephrectomy, the validation cohort received either sunitinib ($n = 8$), similar targeted agents ($n = 3$) or no drug ($n = 11$). These cohorts were enriched for patients with a poor or intermediate prognosis, in line with the SuMR trial selection criteria [18]. Median follow-up time, defined as time of entry to death or last contact, was 22.0, 12.3 months respectively for the development, validation cohorts. Univariate Cox regression for mTOR and overall survival analysed an overlapping cohort ($n = 45$) which included an additional patient [20]. Comparisons of cohort characteristics used Mann–Whitney, Fisher or binomial tests as appropriate; p values were two-tailed and corrected for multiple hypothesis testing [21]. Net reclassification improvement (NRI) confidence intervals were calculated using bootstrapping [22, 23].

Multiregion tumour sampling

Details of multiregion tissue mapping and sample preparation are given in [24]. Briefly, samples taken forward for reverse phase protein array (RPPA) analysis were spatially separated and selected to represent morphological diversity across the tumour. Fresh-frozen tumours were divided into spatially mapped 1 cm³ pieces; cryostat sections of each piece were examined to confirm ccRCC

Table 1 Clinical characteristics of cohorts studied

Cohort	Total	Development (SuMR)	Validation (SCOTRRCC)
Number of patients	44	22	22
Age, median (range) ^a	66.5 (38.2–79.3)	64.5 (44–78)	67.5 (38.2–79.3)
Overall survival (months): ^b			
Median	16.2	23.5	12.3
Interquartile range	9.1–25.9	14.8–29.8	7.6–18.8
Number censored	22	10	12
Modes (bimodal model)	10.6, 27.3	10.2, 27.4	11.6, 27.5
Number of samples per tumour ^a median (range)	4 (1–10)	4 (1–10)	4 (2–8)
Male gender ^b	29	15	14
Fuhrman grade: ^b			
4	9	1	8
3	22	12	10
2	13	9	4
Stage: ^b			
T4	4	4	0
T3	33	13	20
T2	6	4	2
T1	1	1	0
Performance status: ^c			
KPS >80 (unavailable)	34 (1)	13 (0)	21 (1)
Anaemia ^{b,+} (unavailable)	22 (2)	12 (0)	10 (2)
Raised calcium ^b (unavailable)	8 (7)	5 (0)	3 (7)
Raised LDH ^c (unavailable)	10 (15)	5 (0)	5 (15)
Neutrophil count >70% upper limit of normal ^b (unavailable)	9 (2)	2 (0)	7 (2)
Platelet count >400 ^b (unavailable)	12 (2)	6 (0)	6 (2)
<i>VHL</i> mutation ^b	31	14	17
Renal response at surgery:			
Partial response	–	2	–
Stable disease	–	20	–
Number of metastatic sites: ^b			
1	19	6	13
2	18	11	7
3	7	5	2
IMDC class: ^b			
Intermediate	26	12	14
Poor	16	10	6
Unavailable	2	0	2
MSKCC class: ^b			
Favourable/intermediate:	26	13	13
Poor	12	9	3
Unavailable	6	0	6

^aHb <130 (M), <110 (F); ^a*p* < 0.05; ^b*p* > 0.05; ^c*p* < 10⁻⁴

status and for morphological classification. Up to four samples per morphologically distinct region in each tumour were selected for protein extraction; each of these samples reflected *circa* 50–75 mm³ of tissue.

Intratumoural protein expression variance in sunitinib-exposed and sunitinib-naïve cancers

Fifty-five protein targets were investigated by RPPA, selected according to prior knowledge and validated antibody availability [20]. Each tumour sample analysed by RPPA reflected 50–75 mg of lysed tissue taken from a 1 cm³ spatially mapped region [24]. Protein extraction, RPPA slide spotting, immunofluorescence data acquisition, data processing and identification of four markers that had increased variance associated with sunitinib treatment ($p < 0.05$) were described previously [20, 25]. Briefly, 1 mg/ml lysates were spotted onto nitrocellulose slides using a robotic spotter, and immunofluorescence imaging was performed with an Odyssey scanner (Li-Cor Biosciences, Lincoln, NB, USA). Image processing and logistic curve fitting to the RPPA dilution series employed MicroVigene software (VigeneTech, Carlisle, MA, USA). Protein variance per tumour was estimated using batch-corrected, normalised RPPA expression values from multiregion sampling, comparing the ratio of mean-squared errors between sunitinib-exposed and sunitinib-naïve cohorts per protein marker in an analysis of variance (ANOVA) framework. Statistical significance of variance differences was assessed using the F test only when relevant assumptions held, assessed by the Lillefors and Fligner-Kileen tests [20]. Ranking by the protein expression variance log-ratio between sunitinib-exposed and sunitinib-naïve tumours identified a further two proteins of potential interest where variance was greater than at least one of the four significant markers; these proteins did not meet F -test assumptions and so were not assessed in our previous work using the ANOVA framework. Therefore, six proteins (CA9, N-cadherin (CDH2), EPCAM, mTOR (MTOR), MLH1, BCL2) were candidate molecular variables input into feature selection (described in the following section). The antibodies used for these candidate variables are listed in (supplementary) Table S1 of Additional file 1.

Selection of variables and multivariate modelling

Variables were selected for Cox proportional hazards regression to overall survival on the development cohort using wrapper feature selection with backward elimination regularised by Bayesian information criterion (BIC) [26, 27]. Backward elimination iteratively removed a single feature (i.e. protein expression or a clinical parameter) at each step, selecting for the greatest improvement in BIC value. BIC regularisation seeks to balance the model complexity (number of parameters, including

candidate features) against the model likelihood (fit to the data); therefore, this approach removes features with the smallest contribution to model likelihood while penalising redundancy. The selection procedure terminated with a final model when removing any single feature did not improve BIC. The 'coxph' and 'stepAIC' functions were used respectively from the 'survival' and 'MASS' R libraries (with model complexity penalty specified for BIC) [28].

Comparison with established clinico-pathological scores

IMDC and MSKCC scores were calculated according to the relevant clinical parameters [3, 4]. Sufficient data were available to calculate the IMDC score for 20/22 patients in the validation cohort, all of whom fell into the 'intermediate' or 'poor' categories. MSKCC score was used to group patients into (1) favourable/intermediate and (2) poor prognosis; sufficient data were available to classify 14/22 patients. A further two patients were on the borderline of intermediate or poor prognosis with MSKCC parameters due to missing data, but had short survival times and were assigned to the poor prognosis group. Therefore, two ambiguous values were resolved in favour of the MSKCC score performance, making comparison with NEAT more stringent; hence 16/22 patients were assigned MSKCC scores. All patients in the development cohort had sufficient data for IMDC and MSKCC scoring. The reported hazard ratio (HR) for NEAT reflects stratification into either better or worse than average risk groups (i.e. classification threshold of $\log\text{HR} = 0$); this threshold was predetermined and not derived from exploratory data analysis. HR reported for IMDC, MSKCC follows the groupings described above.

Investigating stratification performance with reduced number of samples per tumour

In order to evaluate tumour sampling effects on NEAT performance, a subsampling procedure produced datasets taking a maximum number of tumour samples (MNTS) of 1, 2 or 3 per tumour (and thus per patient). This approach employed Sobol sampling [29]; see supplementary methods in Additional file 1 for further details. The selected tumour samples were used to calculate median protein expression per patient as input for the NEAT algorithm. Patient age was unchanged. The HR and log-rank p value for stratification into 'high' and 'low' risk groups defined by NEAT $\log\text{HR} = 0$ were calculated. This analysis was performed on 10^6 datasets per MNTS examined, where each dataset represented a unique combination of samples across all patients in the validation cohort. Therefore, every patient was represented in each of the 10^6 datasets; thus, 10^6 NEAT HR and log-rank p values were generated for each

MNTS, representing predictive performance distributions across the different tumour sample combinations.

Results

Cohort characteristics

The two mcrRCC cohorts were similar across many characteristics (Table 1), although statistically significant differences were identified for Karnofsky performance status, elevated lactate dehydrogenase and age. Clustering analysis of overall survival (OS) using regularised Gaussian mixture modelling for unsupervised cardinality selection identified two modes (clusters) in the combined cohorts ($n = 44$, Fig. 1). The longer survival cluster had a median OS (mOS) of 27.3 months, matching the favourable or intermediate prognosis subgroups defined in pivotal studies. For example, the favourable subgroup reported for the MSKCC score had mOS of 30 months [4], mOS for the IMDC score intermediate subgroup was 27 months [3] and a further independent study reported mOS of 26 months for the favourable subgroup [30]. The shorter survival cluster had mOS of 10.6 months, which is similar to reported mOS values across poor and intermediate prognosis subgroups in the preceding studies [3, 4, 30]. Greater representation of the shorter survival cluster in the validation cohort was partly due to censoring and also arose from the drug response selection

criterion for the development cohort [18]. However, survival times for the validation and development cohorts were not significantly different. Therefore, the population studied ($n = 44$) has a bimodal OS distribution that aligns with that of subgroups identified in larger mcrRCC cohorts [3, 4, 30].

The NEAT algorithm for risk stratification of patients with metastatic renal cancer

A machine learning approach using regularised wrapper selection [27] with Cox multivariate analysis [26] on the development cohort identified a novel model for mcrRCC patient risk stratification by overall survival. We hypothesised that proteins with increased intratumoural variance following therapy may function as markers of resistance or aggressiveness and so enable prognostication. Indeed, factors underlying changes in tumour composition with treatment include clonal selection and proteomic diversity across isogenic cell populations [16, 31, 32]. Twelve variables were examined, including six key clinical parameters (grade, gender, age, neutrophils, haemoglobin, IMDC score [3]) and values for six proteins where intratumoural variance was greater in sunitinib-exposed mcrRCC. Prognostic variables automatically identified by machine learning were N-cadherin, EPCAM, Age and mTOR (NEAT), controlling for the above clinical parameters. Protein expression values for these markers in the development and validation cohorts are shown in Fig. 2. The resulting multivariate Cox proportional hazards model for the development cohort had likelihood ratio test $p = 1.18 \times 10^{-4}$, and all selected variables were individually significant in the multivariate model (Table 2).

The interesting positive relationship of mTOR with survival was followed up in an overlapping cohort and was significant in univariate Cox regression ($p = 0.034$). The proportional hazards assumption was not invalidated (Grambsch-Therneau test [33], (supplementary) Table S2 of Additional file 1). The HR was calculated from relative protein expression values and age in years at diagnosis as follows:

$$\text{Hazard ratio} = \exp(8.927 \text{ N-cadherin} + 3.800 \text{ EPCAM} + 0.129 \text{ Age} - 18.385 \text{ mTOR})$$

NEAT performed well on the geographically separated validation and development cohorts (Fig. 3). This work reflects evidence level IB [34], where development used prospective clinical trial data and validation was performed with patients who received current standard therapy. Concordance index (C index) [35] values for the NEAT, IMDC and MSKCC score risk groups in the validation cohort were respectively 0.77 (95% CI 0.66–0.88), 0.76 (95% CI 0.60–0.92) and 0.64 (95% CI 0.54–0.75). Net reclassification improvement [22] values for NEAT on the validation cohort were 7.1% vs IMDC (95% CI –24.8%,

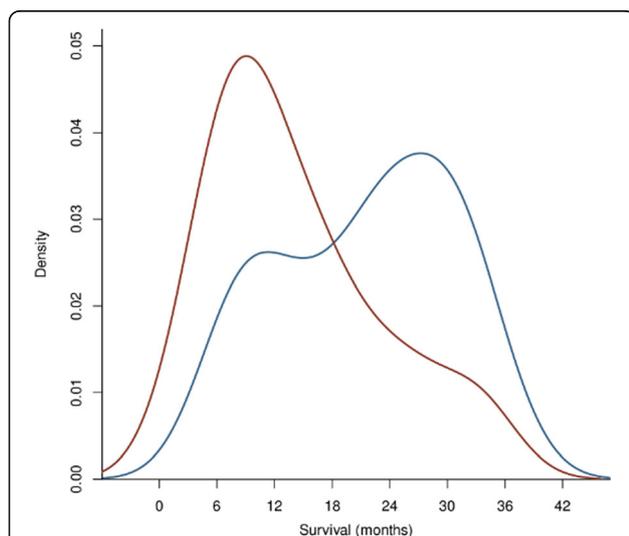
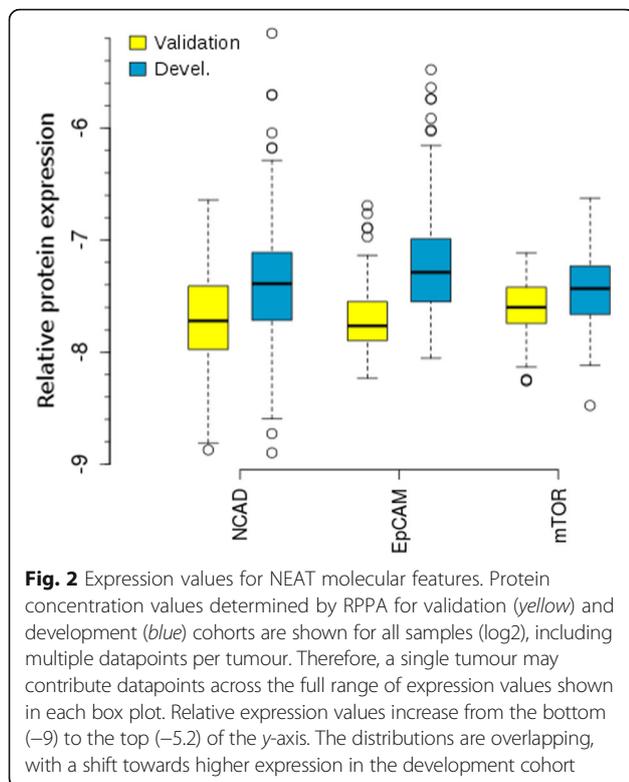


Fig. 1 Overall survival distributions for the development (SuMR) and validation (SCOTRRCC) cohorts. Kernel density plots are shown for all survival data, including censored data. The above distributions indicate bimodality for both cohorts studied, with similar mode positions around 11 and 27 months. These survival modes align with survival subgroups reported in pivotal studies [3, 4, 30]. The development cohort (blue) had the greatest proportion of patients in the mode centred around 27 months, reaching a density value of 0.037. The majority of patients in the validation cohort (red) are in the survival mode around 11 months (reaching a density value of 0.049), partly due to greater censoring in this cohort



39.0%) and 25.4% vs MSKCC score (95% CI -25.7%, 76.5%), shown in Table 3.

Tumour sampling is a critical limiting factor for validation of molecular stratification approaches

The overall approach to investigate the effects of tumour sampling on predictive performance is summarised in Fig. 4. Three distributions of NEAT hazard ratio and log-rank *p* value were generated to reflect sampling 1, 2 or 3 regions per tumour in the validation cohort; these distributions capture NEAT performance for different sample combinations taken across tumours and patients. For example, consider three patients, each with RPPA data from four different tumour samples. If a single sample is taken from each patient for NEAT analysis, there would be 4³ (i.e. 64) unique combinations of tumour samples across the three patients. Validation power rose significantly at each increase in the number of tumour samples taken per patient, and the full dataset with a

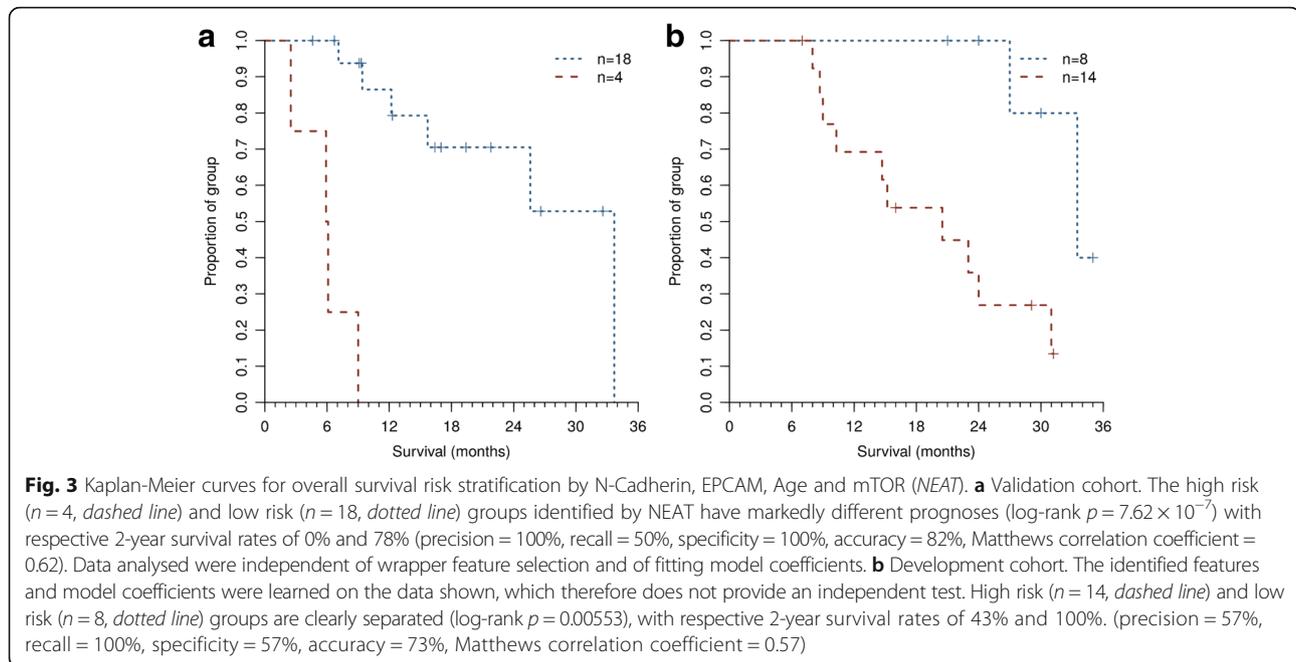
Table 2 Multivariate Cox proportional hazards model for overall survival, fitted on the development cohort

Feature	HR (95% confidence interval)	<i>p</i>
mTOR	1.04 × 10 ⁻³ (6.09 × 10 ⁻⁴ – 1.71 × 10 ⁻³)	0.001
EPCAM	44.7 (845.6–2.36)	0.011
N-Cadherin	7.53 × 10 ³ (2.76 × 10 ⁶ – 20.71)	0.003
Age	1.14 (1.27–1.02)	0.018

median of four spatially separated samples per tumour appeared adequate, conferring good predictive power. NEAT overall performance on the validation cohort was poor when limited to a single tumour sample per patient, and was significantly impaired with two samples per patient (Fig. 5a). In the single sample regime, stratification into good and poor prognosis groups was only just better than random expectation (median logHR = 0.109, binomial *p* < 10⁻³²²); strong statistical significance is due to the large datasets studied. Taking two samples per tumour gave improved stratification performance over a single sample (median logHR = 1.614, Mann–Whitney *p* < 10⁻³²⁴), and substantial further improvement was found when taking three samples (median logHR = 3.030, Mann–Whitney *p* < 10⁻³²⁴). Application of NEAT to different subsets of tumour samples per individual patient changed risk group assignment for 64% of the validation cohort (Fig. 5b). Interestingly, the median variance in per-patient HR was 3.5-fold greater in low grade samples than high grade samples (Mann–Whitney *p* = 0.044). In order to further investigate the independent prognostic power of individual tumour regions, we compared prediction using expression values averaged across all available samples for each individual against the best possible results obtained using only one sample per tumour. Validation using all of the available samples per tumour outperformed even the most predictive single sample taken (*p* < 10⁻⁶).

Discussion

This study examines the effect of sampling on the performance of a novel molecular prognostic approach, NEAT, using protein measurements from 183 regions across 44 mcrRCC tumours. The unique development cohort from the SuMR trial allowed for selection of proteins that had increased intratumoural expression variance with treatment; we hypothesised that these proteins may be markers of aggressiveness and therefore useful in prognostication. Although the cohorts are relatively small, NEAT gave statistically robust stratification of the independent validation cohort by overall survival (Fig. 3a). The trend for favourable NEAT performance relative to the IMDC, MSKCC scores would benefit from investigation in a larger cohort, and the good performance of IMDC relative to the MSKCC score aligns with previous work [3]. To our knowledge, the mcrRCC cohorts analysed here are the largest available with RPPA data from pathologist-guided, multiregion tumour sampling. Our approach to capture grade diversity is likely to better represent ITH than standard sampling methods. Furthermore, each sample analysed by RPPA reflects a large tissue volume (circa 50–75 mm³) relative to standard approaches based on tissue sections from formalin-fixed paraffin embedded material such as tissue microarray analysis



(<0.2 mm³ per region). Therefore, the RPPA data analysed cover a higher proportion of the overall tumour volume relative to standard approaches. The sampling approaches may be an important enabling factor in NEAT reproducibility and hence good validation performance, despite the relatively small cohorts studied. The RPPA technique offers potential as a quantitative alternative to IHC and has already been applied in a clinical setting through the Clinical Laboratory Improvement Amendments (CLIA) facility certification process [36, 37]. The NEAT model might ultimately be applied to inform decision making and patient management in several areas: (1) monitoring and follow-up, (2) recruitment into clinical trials with new agents, (3) treatment decisions, for example, for patients on the borderline of receiving drug due to other factors and (4) patient counselling.

The NEAT development and validation cohorts were relatively small ($n = 44$ total), which is associated with increased risk of type II error and wide confidence intervals on predictive performance. Cytoreductive nephrectomy is standard clinical practice, and the use of upfront tyrosine kinase inhibitor (TKI) treatment is variable,

limiting recruitment of a uniform cohort (as was obtained from the SuMR clinical trial) for NEAT development. A further limiting factor on the size of cohorts in our study was the availability of appropriately consented fresh-frozen material with multiregion sampling and pathology assessment for RPPA analysis. Our approach to discover resistance biomarkers required multiregion sampling of tumour tissue from patients treated with upfront sunitinib in order to enable comparison of candidate marker variance in sunitinib-exposed and sunitinib-naïve material. Therefore, the cohorts received different treatment regimens and also had significant differences in some clinical characteristics. NEAT performed well on both cohorts despite these differences, and so might be broadly useful for prognostication of mRCC. Further study of NEAT performance on an independent upfront sunitinib cohort would be of interest to further explore potential clinical utility, such as to inform decision making about performing a cytoreductive nephrectomy [38].

Subsampling of the multiregion RPPA data showed that validation of the NEAT prognostic model was critically dependent on the number of samples analysed per

Table 3 Performance characteristics of NEAT and clinico-pathological scores

Cohort	Percentage NRI (95% CI)		HR (95% CI)		
	NEAT vs IMDC	NEAT vs MSKCC ^a	NEAT	IMDC	MSKCC
Validation	7.1 ^a (-24.8, 39.0)	25.4 ^b (-25.7, 76.5)	37.89 (4.1, 353.8)	7.317 ^a (1.54, 34.69)	14.39 ^b (1.24, 165.7)
Development	25.0 (-34.8, 84.8)	23.3 (-28.3, 74.9)	10.649 (1.35, 84.08)	1.378 (0.44, 4.36)	3.577 (1.00, 12.85)

NRI net reclassification improvement, HR Hazard ratio, NEAT N-cadherin EPCAM Age mTOR multivariate model, IMDC International Metastatic Database Consortium, MSKCC Memorial Sloan Kettering Cancer Center

Values reflect assignment of either high risk or low risk status (methods)

^a $n = 20$ due to data availability

^b $n = 16$ due to data availability

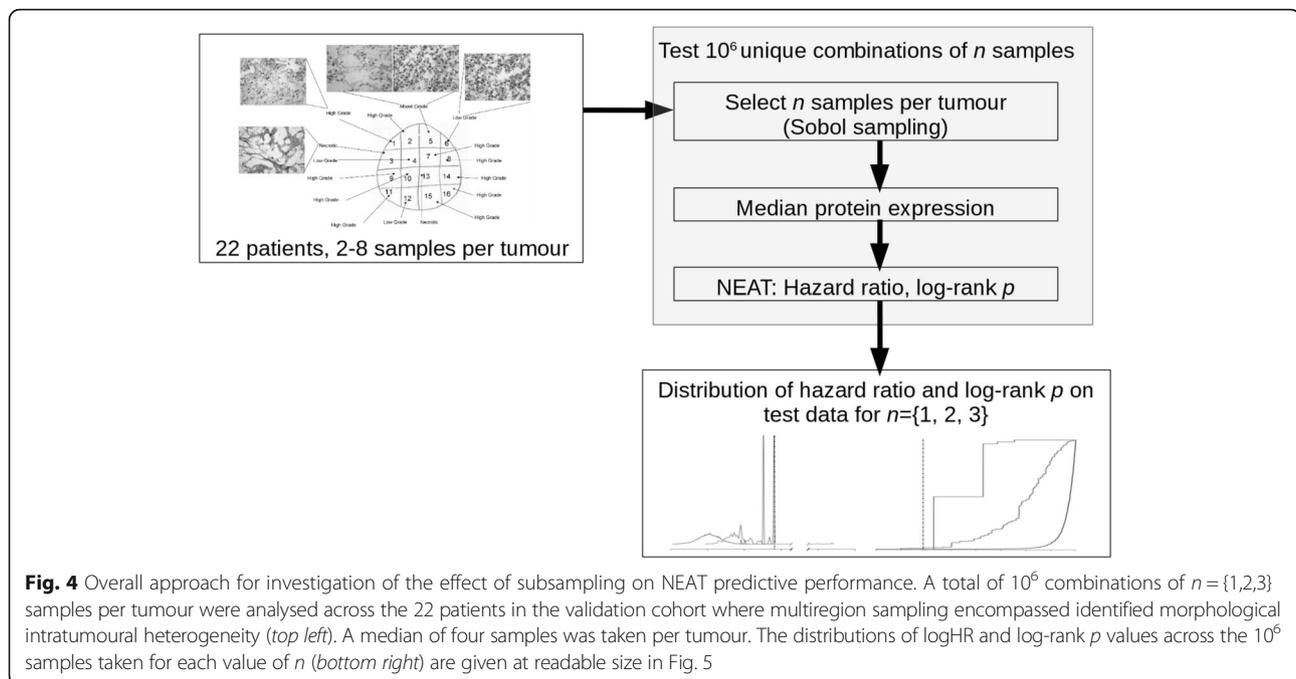
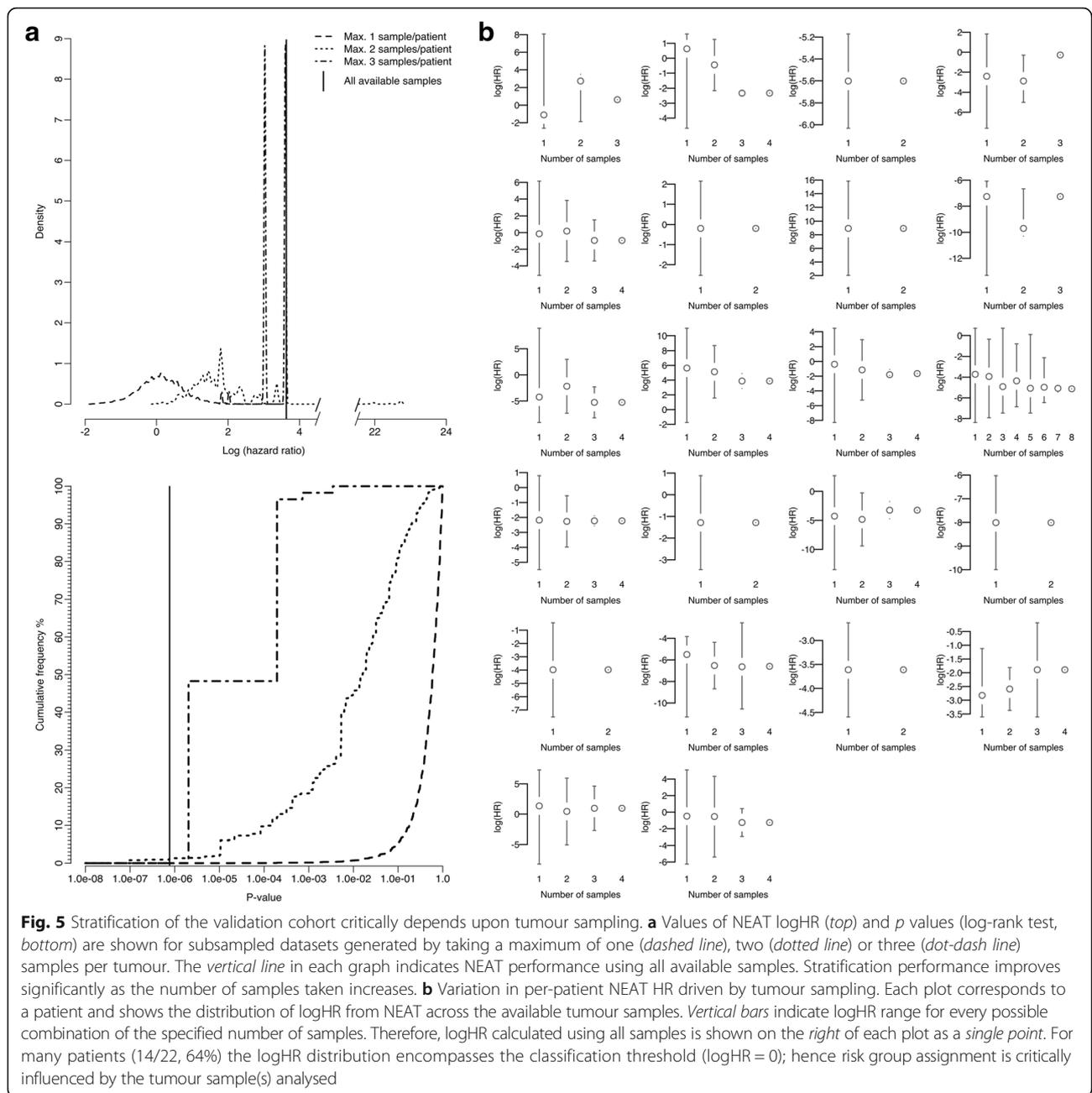


Fig. 4 Overall approach for investigation of the effect of subsampling on NEAT predictive performance. A total of 10^6 combinations of $n = \{1, 2, 3\}$ samples per tumour were analysed across the 22 patients in the validation cohort where multiregion sampling encompassed identified morphological intratumoural heterogeneity (*top left*). A median of four samples was taken per tumour. The distributions of logHR and log-rank p values across the 10^6 samples taken for each value of n (*bottom right*) are given at readable size in Fig. 5

tumour. Indeed, the model's performance in risk stratification improved significantly at each increase in the number of tumour regions analysed (Fig. 5a). These results therefore evidence the benefit of more extensive tumour sampling both for biomarker development and also in validation studies where the sampling protocol may contribute to a reported lack of reproducibility. The efficacy of even the most promising tissue-based biomarkers is diminished by ITH [39], and identification of molecular predictors that are unaffected by ITH may be very challenging. Indeed, cancer biomarkers have historically suffered from a high attrition rate [8]. The available data provided for subsampling analysis of one, two and three samples per tumour; however, analysis with the full dataset (median of four samples) performed best. In principle, even higher sampling rates may be beneficial; several patients where >3 samples were taken, reflecting larger tumours, show considerable variation in HR even when large numbers of samples are analysed (Fig. 5b). One patient where eight tumour regions were examined had substantial variation in NEAT HR even across subsets containing six samples. Therefore, the influence of tumour sampling on predicted risk is clear for individual patients. These results also evidence benefit of sampling in proportion to tumour volume for molecular diagnostics. We found considerably greater variance in HR for low grade over high grade samples; thus, tumour biomarker studies would benefit from performing more extensive sampling of low grade regions. This result also underlines the additional information provided by NEAT. Indeed, the automatic feature selection process

deprioritised grade relative to molecular variables. Prognostication using all of the multiple tumour samples gave better risk stratification than provided by analysis of any single sample in isolation. Therefore, NEAT analysis with multiple tumour regions captures information unavailable in any single sample; this information may reflect the adaptive potential arising from ITH [40] and also might include aspects of disease progression such as the degree of vascularisation or the length of time since initial dissemination competence.

With regard to the individual components of the NEAT model, the positive association of mTOR with overall survival was the strongest, most significant feature and was also found in univariate analysis of an overlapping cohort. The mTOR pathway is an important mediator of RTK growth signalling [41]. Improved prognosis associated with elevated mTOR in NEAT suggests that tumours dependent upon mTOR have enhanced sensitivity to sunitinib. Therefore, sunitinib may act directly on tumour cells to inhibit mTOR growth, consistent with results in ovarian cancer that VEGF stimulates the mTOR pathway [42]. Additionally, the mTORC1 complex, which includes mTOR, exerts negative feedback on RTKs to suppress proliferation and survival [41]; this negative feedback could enhance therapeutic RTK inhibition by sunitinib. Notably, mTOR inhibitors are currently in clinical use (for example, everolimus), possibly in conjunction with sunitinib or similar agents. Our results suggest caution in co-treating with mTOR inhibitors and sunitinib, resonating with the poor performance of everolimus followed by sunitinib in the



RECORD-3 trial [43]. Consistent with previous results, for example [44, 45], a significant negative association with survival was identified for N-cadherin, a canonical marker of epithelial to mesenchymal transition. Additionally, N-cadherin is expressed by endothelial cells and so may also represent a surrogate for vascularisation [46]. Age is a known RCC prognostic factor that was not selected for the IMDC score [3, 47, 48]. Our analysis took age as continuous values, which may partly explain selection of this variable for the NEAT model and not in the IMDC analysis, which dichotomised age at 60 years [49]. The IMDC score was not selected by our machine learning

approach which implies that, in the development cohort, prognostic information captured by the IMDC score overlaps with that provided by the NEAT variables. High EPCAM expression is also associated with poor prognosis in NEAT and multiple cancers [50, 51], although reports link EPCAM with better prognosis in localised RCC; see, for example, [52, 53]. The contrasting association with survival for EPCAM in NEAT may be due to differences between advanced and localised ccRCC, technologies used and context-specific function, for example, in signal transduction by nuclear localisation of the cleaved intracellular domain [54].

Conclusions

Multiregion sampling to capture mcrRCC grade diversity enabled investigation of ITH impact on risk stratification with a novel protein-based prognostic model, NEAT (N-Cadherin, EPCAM, Age, mTOR). NEAT compares well with established clinico-pathological scores on a geographically separate independent validation cohort who received current standard therapy. Results show that evaluation or attempted use of any molecular prognostic and predictive methods with few tumour samples will lead to variable performance and low reproducibility. We demonstrate parameters (tumour coverage, size, grade) that may be used to inform sampling in order to enhance biomarker reproducibility, and results underline the critical importance of addressing heterogeneity to realise the promise of molecular stratification approaches. Through studies such as TRACERx [55], we anticipate that extensive multiregion sampling will become standard procedure for discovery and validation of molecular diagnostics across a range of cancer types.

Recommendations arising from our research include the following: (1) biomarker validation studies should implement tumour sampling protocols that match as closely as possible to the discovery work; (2) clinical biomarker research and ultimately front-line diagnostic approaches may benefit from greater tumour sampling rates; (3) clinical parameters (including tumour grade, size, coverage) can guide sample selection, and investigation of additional parameters to inform sampling may be useful; (4) optimisation of tumour sampling rate and sample selection protocols are important research areas to enable advances in stratified cancer medicine.

Additional files

Additional file 1: Supplementary methods, tables and summary of the Additional file 2 zip archive (PDF). (PDF 61 kb)

Additional file 2: NEAT model, custom computer code and relevant datasets for investigation of limiting sample number on NEAT performance (ZIP archive). (ZIP 13 kb)

Additional file 3: Reverse phase protein array data for multiregion tumour sampling (comma-separated values, CSV). (CSV 141 kb)

Additional file 4: Clinical data; columns are: anonymised patient identifier, age, grade, event, overall survival, Heng class, MSKCC class. Anonymised identifiers correspond with those given in Additional file 3 (comma-separated values, CSV). (CSV 962 bytes)

Abbreviations

BIC: Bayesian information criterion; ccRCC: Clear cell renal cell cancer; HR: Hazard ratio; IMDC: International Metastatic Database Consortium; ITH: Intratumoural heterogeneity; mcrRCC: Metastatic clear cell renal cell cancer; MNTS: Maximum number of tumour samples; mOS: Median overall survival; MSKCC: Memorial Sloan Kettering Cancer Center; NEAT: N-cadherin EPCAM Age mTOR multivariate model; OS: Overall survival; RCC: Renal cell cancer; RPPA: Reverse phase protein array; RTK: Receptor tyrosine kinase; SCOTRRCC: Scottish Collaboration On Translational Research into Renal Cell Cancer

Acknowledgements

Thanks to Professor Fei Ye, Department of Biostatistics, Vanderbilt University, Nashville, TN, USA for critical reading of the manuscript.

Funding

We acknowledge financial support from the Royal Society of Edinburgh Scottish Government Fellowship co-funded by Marie Curie Actions (IMO), Carnegie Trust (50115; IMO, DJH, GDS), IGMM DTF (IMO, GDS), Medical Research Council (MC_UU_12018/25; IMO), Chief Scientist Office Scotland (ETM37; GDS, DJH), Cancer Research UK (Experimental Medicine Centre; TP, DJH), Renal Cancer Research Fund (GDS), Kidney Cancer Scotland (GDS), MRC Clinical Training Fellowship (AL), RCSEd Robertson Trust (AL) and Melville Trust (AL).

Availability of data and materials

Data generated or analysed during this study are included in this published article and the Additional files, as follows. The NEAT model, custom computer code and datasets for investigation of stratification performance using limited tumour samples per patient are provided in Additional file 2, under the Creative Commons CC-BY-NC-SA license. Please see the 'README.md' data file within the Additional file 2 zip archive. Normalised multiregion sampling RPPA data, representing 183 regions and 55 markers, are provided in Additional file 3. Key clinical data are provided in Additional file 4; all clinical data are anonymised. Any further data that may be required are available from the corresponding author upon reasonable request.

Authors' contributions

IMO conceived and designed the data analysis, led the writing of the manuscript and was scientific lead for this study. IMO supervised ALRL, who co-designed and implemented the data analysis. DJH, GDS and TP conceived, designed and implemented clinical aspects and also obtained ethical approval. DJH developed the RPPA discovery platform. GDS and DJH contributed samples and clinical data and supervised FM, AL and PM, who collected protein data (RPPA). AL also provided clinical data, and MO'D contributed pathology review. TP provided clinical data and samples for analysis by RPPA. IMO, GDS, ALRL and DJH drafted the manuscript and interpreted results. IMO, DJH, GDS, AL and TP obtained funding. All authors read and approved the final manuscript.

Competing interests

IMO, GDS, ALRL, DJH and TP have an ownership interest in a patent pending relating to the NEAT algorithm. TP has received speakers bureau honoraria from and is a consultant/advisory board member for GlaxoSmithKline and Pfizer and is additionally a recipient of commercial research grants from Pfizer. GDS has received speakers bureau honoraria from Pfizer.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Ethical approval was obtained prior to commencing work. REC references: 07/Q0603/58 (East London and the City Research Ethics Committee 1), 10/S1102/68 (South East Scotland Research Ethics Committee 02) and 10/S1402/33 (East of Scotland Research Ethics Service).

Publisher's Note

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

Author details

¹MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. ²Scottish Collaboration On Translational Research into Renal Cell Cancer (SCOTRRCC), Scotland, UK. ³School of Medicine, University of St Andrews, St Andrews, Fife KY16 9TF, UK. ⁴Department of Pathology, Western General Hospital, Edinburgh EH4 2XU, UK. ⁵Barts Cancer Institute, Experimental Cancer Medicine Centre, Queen Mary University of London, London EC1M 6BQ, UK. ⁶Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH16 4UX, UK. ⁷Present Address: Vanderbilt University School of Medicine, Vanderbilt University, Nashville, Tennessee, USA. ⁸Present Address: Academic Urology Group, University of Cambridge, Box 43, Addenbrooke's Hospital, Cambridge Biomedical Campus, Hill's Road, Cambridge CB2 0QQ, UK.

Received: 8 November 2016 Accepted: 15 May 2017

Published online: 26 June 2017

References

- Stewart GD, O'Mahony FC, Powles T, Riddick ACP, Harrison DJ, Faratian D. What can molecular pathology contribute to the management of renal cell carcinoma? *Nat Rev Urol*. 2011;8:255–65.
- Sun M, Thuret R, Abdollah F, Lughezzani G, Schmitges J, Tian Z, et al. Age-adjusted incidence, mortality, and survival rates of stage-specific renal cell carcinoma in North America: a trend analysis. *Eur Urol*. 2011;59:135–41.
- Heng DY, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, et al. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. *Lancet Oncol*. 2013;14:141–8.
- Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon- α as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. *J Clin Oncol*. 2002;20:289–96.
- Kim HL, Seligson D, Liu X, Janzen N, Bui MHT, Yu H, et al. Using protein expressions to predict survival in clear cell renal carcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res*. 2004;10:5464–71.
- Galsky MD. A prognostic model for metastatic renal-cell carcinoma. *Lancet Oncol*. 2013;14:102–3.
- Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, et al. EAU guidelines on renal cell carcinoma: 2014 Update. *Eur Urol*. 2015;67:913–24.
- Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res*. 2012;72:6097–101.
- Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Oudard S, et al. Overall survival and updated results for sunitinib compared with interferon α in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2009;27:3584–90.
- Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Sunitinib versus interferon α in metastatic renal-cell carcinoma. *N Engl J Med*. 2007;356:115–24.
- Mendel DB, Laird AD, Xin X, Louie SG, Christensen JG, Li G, et al. In vivo antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. *Clin Cancer Res*. 2003;9:327–37.
- Vázquez S, León L, Fernández O, Lázaro M, Grande E, Aparicio L. Sunitinib: the first to arrive at first-line metastatic renal cell carcinoma. *Adv Ther*. 2012;29:202–17.
- Weinstock M, McDermott D. Targeting PD-1/PD-L1 in the treatment of metastatic renal cell carcinoma. *Ther Adv Urol*. 2015;7:365. doi:10.1177/1756287215597647.
- Heppner G. Tumor heterogeneity. *Cancer Res*. 1984;44:2259–65.
- Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet*. 2014;46:225–33.
- Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012;12:323–34.
- Abel EJ, Culp SH, Matin SF, Tamboli P, Wallace MJ, Jonasch E, et al. Percutaneous biopsy of primary tumor in metastatic renal cell carcinoma to predict high risk pathological features: comparison with nephrectomy assessment. *J Urol*. 2010;184:1877–81.
- Powles T, Blank C, Chowdhury S, Horenblas S, Peters J, Shamash J, et al. The outcome of patients treated with sunitinib prior to planned nephrectomy in metastatic clear cell renal cancer. *Eur Urol*. 2011;60:448–54.
- Stewart GD, Riddick ACP, Rae F, Marshall C, MacLeod L, O'Mahony FC, et al. Translational research will fail without surgical leadership: SCOTTRCC a successful surgeon-led Nationwide translational research infrastructure in renal cancer. *Surgeon*. 2015;13:181–6.
- Stewart GD, O'Mahony FC, Laird A, Rashid S, Martin SA, Eory L, et al. Carbonic anhydrase 9 expression increases with vascular endothelial growth factor-targeted therapy and is predictive of outcome in metastatic clear cell renal cancer. *Eur Urol*. 2014;66:956–63.
- Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29:1165–88.
- Pencina MJ, Steyerberg EW, D'Agostino RB. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30:11–21.
- Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology*. 2014;25:114–21.
- O'Mahony FC, Nanda J, Laird A, Mullen P, Caldwell H, Overton IM, et al. The use of reverse phase protein arrays (RPPA) to explore protein expression variation within individual renal cell cancers. *J Vis Exp*. 2013;22. doi: 10.3791/50221
- Stewart GD, O'Mahony FC, Laird A, Eory L, Lubbock ALR, Mackay A, et al. Sunitinib treatment exacerbates intratumoral heterogeneity in metastatic renal cancer. *Clin Cancer Res*. 2015;21:4212–23.
- Cox D. Regression models and life tables. *J R Stat Soc B*. 1972;34:187–220.
- Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell*. 1997;97:273–324.
- Venables WN, Ripley BD. *Modern applied statistics with S*. New York: Springer; 2010.
- Press WH, Teukolsky SA. Quasi (that is, sub) random numbers. *Comput Phys*. 1989;3:76–9.
- Mekhail TM, Abou-Jawde RM, Boumerhi G, Malhi S, Wood L, Elson P, et al. Validation and extension of the Memorial Sloan-Kettering prognostic factors model for survival in patients with previously untreated metastatic renal cell carcinoma. *J Clin Oncol*. 2005;23:832–41.
- Pisco AO, Brock A, Zhou J, Moor A, Mojtahedi M, Jackson D, et al. Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat Commun*. 2013;4:2467.
- Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol*. 2015;16:64.
- Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81:515–26.
- Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*. 2009;101:1446–52.
- Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004;23:2109–23.
- Negm OH, Muftah AA, Aleskandarany MA, Hamed MR, Ahmad DAJ, Nolan CC, et al. Clinical utility of reverse phase protein array for molecular classification of breast cancer. *Breast Cancer Res Treat*. 2016;155:25–35.
- Deeken JF, Wang H, Subramaniam D, He AR, Hwang J, Marshall JL, et al. A phase 1 study of cetuximab and lapatinib in patients with advanced solid tumor malignancies. *Cancer*. 2015;121:1645–53.
- Lane BR, Derweesh IH, Kim HL, O Malley R, Klink J, Ercole CE, et al. Presurgical sunitinib reduces tumor size and may facilitate partial nephrectomy in patients with renal cell carcinoma. *Urol Oncol Semin Orig Investig*. 2015;33:112.e15–21.
- Gulati S, Martinez P, Joshi T, Birkbak NJ, Santos CR, Rowan AJ, et al. Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. *Eur Urol*. 2014;66:936–48.
- Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer*. 2013;108:479–85.
- Laplanche M, Sabatini DM. mTOR signaling in growth control and disease. *Cell*. 2012;149:274–93.
- Trinh XB, Tjalma WA, Vermeulen PB, Van den Eynden G, Van der Auwera I, Van Laere SJ, et al. The VEGF pathway and the AKT/mTOR/p70S6K1 signalling pathway in human epithelial ovarian cancer. *Br J Cancer*. 2009;100:971–8.
- Motzer RJ, Barrios CH, Kim TM, Falcon S, Cosgriff T, Harker WG, et al. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2014;32:2765–72.
- Shimazui T, Kojima T, Onozawa M, Suzuki M, Asano T, Akaza H. Expression profile of N-cadherin differs from other classical cadherins as a prognostic marker in renal cell carcinoma. *Oncol Rep*. 2006;15:1181–4.
- Pantuck AJ, An J, Liu H, Rettig MB. NF- κ B-Dependent plasticity of the epithelial to mesenchymal transition induced by Von Hippel-Lindau inactivation in renal cell carcinomas. *Cancer Res*. 2010;70:752–61.
- Cavallaro U, Liebner S, Dejana E. Endothelial cadherins and tumor angiogenesis. *Exp Cell Res*. 2006;312:659–67.
- Taccox X, Valeri A, Descotes J-L, Morin V, Stindel E, Doucet L, et al. Renal cell carcinoma in adults 40 years old or less: young age is an independent prognostic factor for cancer-specific survival. *Eur Urol*. 2007;51:980–7.
- Sánchez-Ortiz RF, Rosser CJ, Madsen LT, Swanson DA, Wood CG. Young age is an independent prognostic factor for survival of sporadic renal cell carcinoma. *J Urol*. 2004;171:2160–5.

49. Heng DYC, Xie W, Regan MM, Warren MA, Golshayan AR, Sahi C, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. *J Clin Oncol.* 2009;27:5794–9.
50. Spizzo G, Fong D, Wurm M, Ensinger C, Obrist P, Hofer C, et al. EpCAM expression in primary tumour tissues and metastases: an immunohistochemical analysis. *J Clin Pathol.* 2011;64:415–20.
51. Trzpis M, McLaughlin PMJ, de Leij LMFH, Harmsen MC. Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *Am J Pathol.* 2007;171:386–95.
52. Seligson DB, Pantuck AJ, Liu X, Huang Y, Horvath S, Bui MHT, et al. Epithelial cell adhesion molecule (KSA) expression pathobiology and its role as an independent predictor of survival in renal cell carcinoma. *Clin Cancer Res.* 2004;10:2659–69.
53. Eichelberg C, Chun FK, Bedke J, Heuer R, Adam M, Moch H, et al. Epithelial cell adhesion molecule is an independent prognostic marker in clear cell renal carcinoma. *Int J Cancer J Int Cancer.* 2013;132:2948–55.
54. Maetzel D, Denzel S, Mack B, Canis M, Went P, Benk M, et al. Nuclear signalling by tumour-associated antigen EpCAM. *Nat Cell Biol.* 2009;11:162–71.
55. Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol.* 2014;12:e1001906.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

