

Technical University of Denmark



## Parallel evolution of tumor subclones mimics diversity between tumors

**Martinez, Pierre; Birkbak, Nicolai Juul; Gerlinger, Marco; McGranahan, Nicholas; Burrell, Rebecca; Rowan, Andrew; Joshi, Tejal; Fisher, Rosalie; Larkin, James; Szallasi, Zoltan Imre; Swanton, Charles**

*Published in:*  
Journal of Pathology

*Link to article, DOI:*  
[10.1002/path.4214](https://doi.org/10.1002/path.4214)

*Publication date:*  
2013

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Martinez, P., Birkbak, N. J., Gerlinger, M., McGranahan, N., Burrell, R., Rowan, A., ... Swanton, C. (2013). Parallel evolution of tumor subclones mimics diversity between tumors. *Journal of Pathology*, 230(4), 356-364. DOI: 10.1002/path.4214

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Parallel evolution of tumour subclones mimics diversity between tumours

Pierre Martinez,<sup>1#</sup> Nicolai Juul Birkbak,<sup>2#</sup> Marco Gerlinger,<sup>1,3#</sup> Nicholas McGranahan,<sup>1,4</sup> Rebecca A Burrell,<sup>1</sup> Andrew J Rowan,<sup>1</sup> Tejal Joshi,<sup>2</sup> Rosalie Fisher,<sup>5</sup> James Larkin,<sup>5</sup> Zoltan Szallasi<sup>2,6</sup> and Charles Swanton<sup>1,7\*</sup>

<sup>1</sup> Cancer Research UK London Research Institute, London, UK

<sup>2</sup> Centre for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark

<sup>3</sup> Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, London, UK

<sup>4</sup> Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, UK

<sup>5</sup> Department of Medicine, Royal Marsden Hospital, London, UK

<sup>6</sup> Children's Hospital Informatics Program at the Harvard–MIT Division of Health Sciences and Technology (CHIP@HST), Harvard Medical School, Boston, MA, USA

<sup>7</sup> UCL Cancer Institute, London, UK

\*Correspondence to: C Swanton, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK.

E-mail: Charles.swanton@cancer.org.uk

#These authors contributed equally to this study.

## Abstract

Intratour heterogeneity (ITH) may foster tumour adaptation and compromise the efficacy of personalized medicine approaches. The scale of heterogeneity within a tumour (intratumour heterogeneity) relative to genetic differences between tumours (intertour heterogeneity) is unknown. To address this, we obtained 48 biopsies from eight stage III and IV clear cell renal cell carcinomas (ccRCCs) and used DNA copy-number analyses to compare biopsies from the same tumour with 440 single tumour biopsies from the Cancer Genome Atlas (TCGA). Unsupervised hierarchical clustering of TCGA and multi-region ccRCC samples revealed segregation of samples from the same tumour into unrelated clusters; 25% of multi-region samples appeared more similar to unrelated samples than to any other sample originating from the same tumour. We found that the majority of recurrent DNA copy number driver aberrations in single biopsies were not present ubiquitously in late-stage ccRCCs and were likely to represent subclonal events acquired during tumour progression. Such heterogeneous subclonal genetic alterations within individual tumours may impair the identification of robust ccRCC molecular subtypes classified by distinct copy number alterations and clinical outcomes. The co-existence of distinct subclonal copy number events in different regions of individual tumours reflects the diversification of individual ccRCCs through multiple evolutionary routes and may contribute to tumour sampling bias and impact upon tumour progression and clinical outcome.

© 2013 The Authors. *Journal of Pathology* published by John Wiley & Sons Ltd on behalf of Pathological Society of Great Britain and Ireland.

**Keywords:** intratumour heterogeneity; copy number; tumour evolution; chromosomal instability; biomarker; personalized medicine

Received 7 September 2012; Revised 17 May 2013; Accepted 22 May 2013

No conflicts of interest were declared.

## Introduction

Over the last 5 years, tumour deep sequencing studies have revealed the scale of diversity between tumours (intertour heterogeneity), with few somatic alterations shared between tumours of the same histopathological subtype [1–5]. Increasingly, evidence is emerging of extensive variation within individual tumours, termed 'intratumour heterogeneity', with reports of diversity in DNA ploidy, somatic

mutation and DNA copy number status in both haematological and solid tumours [6–15].

Morphological intratumour heterogeneity has been recognized for several decades and forms the basis of many pathological grading tools, where variation in nuclear size (nuclear pleomorphism), a poor prognostic marker, provides an estimate of diversity between cells of the same tumour [16]. Furthermore, mechanisms generating genetic diversity within tumours, such as chromosomal instability, are associated with poor patient prognosis [17–21].

© 2013 The Authors. *Journal of Pathology* published by John Wiley & Sons Ltd on behalf of Pathological Society of Great Britain and Ireland. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Diversity within individual tumours may contribute to the acquisition of drug resistance or progression to invasive disease by facilitating subclonal selection [22]. This has been witnessed in lung cancer treatment; patients on epidermal growth factor receptor tyrosine kinase inhibitors have a poorer progression-free survival if the tumour harbours a low-frequency gatekeeper resistance mutation, T790M, prior to therapy initiation [23]. Clonal diversity is also associated with increased risk of progression from premalignant Barrett's oesophagus to invasive oesophageal adenocarcinoma [21]. Intratumour heterogeneity may also confound efforts to predict long-term outcome and personalize cancer medicine, if treatment-resistant tumour subclones are only present at low frequency and evade detection at first presentation [24]. Somatic mutational heterogeneity has also been shown to result in heterogeneous signal transduction pathway dependencies [15]. Such diversity may impact upon stratified medicine approaches, where the entire somatic mutational landscape of the tumour is not represented by a single tumour biopsy, and transcriptomic signatures of outcome vary between good and poor outcome within a primary tumour. Due to this scale of heterogeneity, ubiquitous 'tumour trunk' driver events present in multiple regions of a tumour may prove more effective therapeutic targets than heterogeneous 'tumour branch' events, that may evade detection in a single biopsy [25].

Despite these observations of intratumour genetic, transcriptomic and phenotypic variation from biopsy to biopsy, the scale of intratumour genetic heterogeneity relative to diversity between tumours of the same histopathological subtype in advanced metastatic tumours remains unclear. Minimal intratumour genomic variation relative to the scale of diversity between patients might suggest a more modest impact of intratumour heterogeneity on personalized medicine approaches derived from single tumour biopsies. In contrast, if the scale of genomic change between biopsies from the same tumour can, in some cases, tend toward a similar magnitude as the variation between tumours from different patients, this might suggest a greater challenge to biomarker discovery and validation initiatives.

We applied SNP array analysis to multiple biopsies from each of eight clear cell renal carcinomas (ccRCCs) and compared the copy number profiles to those occurring in tumours of the same histopathological subtype. We found that the scale of genetic heterogeneity within individual tumours mimicked diversity between tumours of the same histopathological subtype. The observation that multiple copy number profiles can co-exist within individual tumours indicates that discrete and uniform copy number-defined ccRCC subgroups may not exist in advanced disease. We conclude that intratumour diversity in copy number states, together with somatic mutational heterogeneity, provide multiple parallel routes for tumour progression and metastasis.

## Materials and methods

### Patient samples and consent

Samples from patients P5, P6 and P7 were collected from nephrectomy specimens acquired from patients on the E-PREDICT translational clinical trial (EU DRAC T No. 2009-013381-54; <http://public.ukcrn.org.uk/Search/StudyDetail.aspx?StudyID=10710>, open to recruitment since January 2010; Principal Investigator JL) of pre- and post-nephrectomy Everolimus treatment in patients presenting with metastatic renal cell carcinoma, the details of which were previously reported [15]. Tumours from patients RK21, RK23, RK29, RMH3 and RMH8 were also collected at the Royal Marsden Hospital, London, but not as part of the PREDICT trial, and received no therapeutic treatment prior to nephrectomy. The study and the translational work were approved by the Royal Marsden Hospital Research Ethics Committee. The eight patients whose data are presented gave written informed consent for study participation and for the translational analyses. Samples from patient 5 were processed on Affymetrix SNP6 arrays in a first batch, samples from patients 6 and 7 in a second and all samples from the remaining five patients in a third batch. Data were deposited in the GEO database (GSE47077). Affymetrix SNP6 arrays for 450 ccRCC patients were obtained from TCGA (<https://tcga-data.nci.nih.gov/tcga/>) on 14 March 2012.

### SNP array data normalization and copy number data

The aroma R package (CRMA v2, CalMaTe 'v1' algorithm and TumourBoost) [26–28] was used to obtain logR and BAF values from the 450 TCGA tumour samples, using normal samples as references and hg19 coordinates. logR and BAF values for 63 tumour samples from the eight patients were obtained with the same method, using a normal sample from each of the eight patients and all normal TCGA samples as references. Sex chromosomes were excluded from the analysis. ASCAT was run on all samples to obtain copy number profiles and ploidy using paired normal samples [29]. The output was further filtered by selecting segments consisting of more than 10 probes. Copy number profiling failed for eight samples (P6-R12, P6-R2, P6-R3, P6-R5, RK21-R2, RMH3-R1, RMH3-R4, RMH8-R7) and sample P6-R1 was excluded because of no discernible tumour content, leaving 54 valid copy number profiles. Cyto-band coordinates were retrieved from the UCSC Genome Browser database (<http://genome.ucsc.edu/>) [30]. Differences in tumour content can potentially account for uneven detection of ubiquitous aberrations across related samples. Samples with very low variability ( $wGII < 0.01$ ; see below) were thus discarded to avoid undetected normal contamination and ensure minimum variability for the computation of sample–sample correlations. Samples P7-R5, P7-R7,

P7-R15, RK29-R2, RMH8-R1 and RMH8-R3 and 10 TCGA samples were thus discarded, leaving a total of 48 multi-region and 440 TCGA samples. Samples were represented as vectors of the mean copy number within each cytoband. Loss and gains were defined as copy numbers deviating from the ploidy, as estimated by ASCAT, by more than 0.6, similarly to the original ASCAT publication [29].

#### Weighted Genomic Instability Index (wGII)

The wGII score [31] is computed on each chromosome as the proportion of bases whose copy number deviates from the ploidy value of the sample given by ASCAT by more than 0.6. The sample score is the sum of the chromosomal scores divided by the number of analysed chromosomes ( $n = 22$ ).

#### Unsupervised clustering of samples

The *pvclust* R package [32] was used to cluster the 48 multi-region and 440 TCGA samples having passed all quality control checks, using a Spearman correlation coefficient-based distance,  $D$ , such that  $D_{ab} = [1 - \text{SCC}(a, b)]/2$  where  $\text{SCC}(a, b)$  is the Spearman's rank correlation coefficient between samples  $a$  and  $b$ .  $D_{ab} = 0$  if  $a$  and  $b$  are perfectly correlated;  $D_{ab} = 0.5$  if the correlation is null; and  $D_{ab} = 1$  if the samples are perfectly anticorrelated. The number of bootstrap replications was set to 1000. Significant clusters were identified using a confidence level of 0.05.

A second method was applied by running the *hmm-Mix* software [33] to identify three clusters with default parameters. Integer copy numbers were converted to  $\log_2(\text{copy number}/\text{ploidy})$  values for each cytoband as input for *hmm-Mix*. The operation was repeated to produce any number from three to 10 clusters.

#### Network analysis

The *cytoscape* software [34] was used to produce a graph view of the fully interconnected network of multi-region and stage III-IV TCGA samples. Nodes represent samples and a force-directed layout was applied using  $1 - \text{normalized}(D)$  as the force between them, using a default spring length of 200. Significant edges were determined as those connecting samples for which  $D$ , the Spearman correlation-based distance between them, was significantly smaller than expected, using all pairwise  $D$ s between stage III and IV TCGA samples as background distribution ( $p < 0.05$ , one-tailed test).

## Results

#### Intratour heterogeneity in copy number profiles

Cytoband-based copy number profiles were derived from Affymetrix SNP 6 array analysis for biopsies from 48 regions of eight individual tumours and 440 TCGA

samples. The 488 profiles were then summarized into a matrix of average copy-number per cytoband. Three of the multi-region sampled tumours received Everolimus treatment (P5, P6, P7), while the other five (RK21, RK23, RK29, RMH3, RMH8) were untreated. The clinicopathological features of each patient are described in Table S1 (see Supplementary material). All regions in all multi-region sampled tumours, except those from patient RK21, displayed loss of genomic material on chromosome 3p (Figure 1). However, all four samples from patient RK21 displayed loss of heterozygosity on the entire chromosome 3 (see Supplementary material, Figure S1). Variability was observed amongst profiles from each individual patient, indicating the presence of ITH of chromosomal gains and losses in all analysed ccRCCs. For instance, patient RK29 displayed two distinct subclonal populations defined by events observed only in a fraction of all RK29 samples; nine profiles showed losses on chromosome 9 and 14q (regions R1, R4, R6, R7, R8, R10, R11, R12, R13), while three profiles did not (regions R3, R5 and R9). Patient P5 appeared relatively homogeneous, with regions R4 (9q loss) R6 (16q loss) and R2 (5q gain, loss of 6q, 9q and 14q) displaying additional broad copy number alterations compared to the other five regions (loss of 3p and 4q).

#### Hierarchical clustering reveals segregation of copy number profiles from the same tumour in several incomplete clusters

The previously generated matrix of average copy number per cytoband was used to compare DNA copy number alterations between our 48 biopsies taken from multiple regions of eight stage III and IV ccRCC tumours with the 440 single biopsies taken from the stage I-IV ccRCC cohort from TCGA. Unsupervised clustering using Spearman's rank correlation coefficient was applied to these samples to reveal whether single tumour biopsies clustered more closely with biopsies from the same tumour or with biopsies from other tumours (Figure 2). If biopsies from the same tumour were more similar to each other than to biopsies from different tumours, we would have expected the emergence of eight robust clusters representing the biopsies derived from the same patients. One thousand bootstrapping iterations produced a total of 39 distinct statistically significant clusters ( $p < 0.05$ ), consisting of samples recurrently clustering together throughout the simulation. Tumour biopsies from patient RK21 were the only case of the eight tumours analysed in which all multi-region samples of the same tumour exclusively and significantly clustered together. Samples from the same tumour were segregated into two or more clusters in five cases (patients P6, P7, RK23, RK29 and RMH8) and regions from three of the multi-region sampled tumours formed significant clusters with at least one unrelated TCGA sample (patients P5, RK23 and RMH3). This lack of robust tumour-specific clusters suggests that, on a copy-number level, intratumour

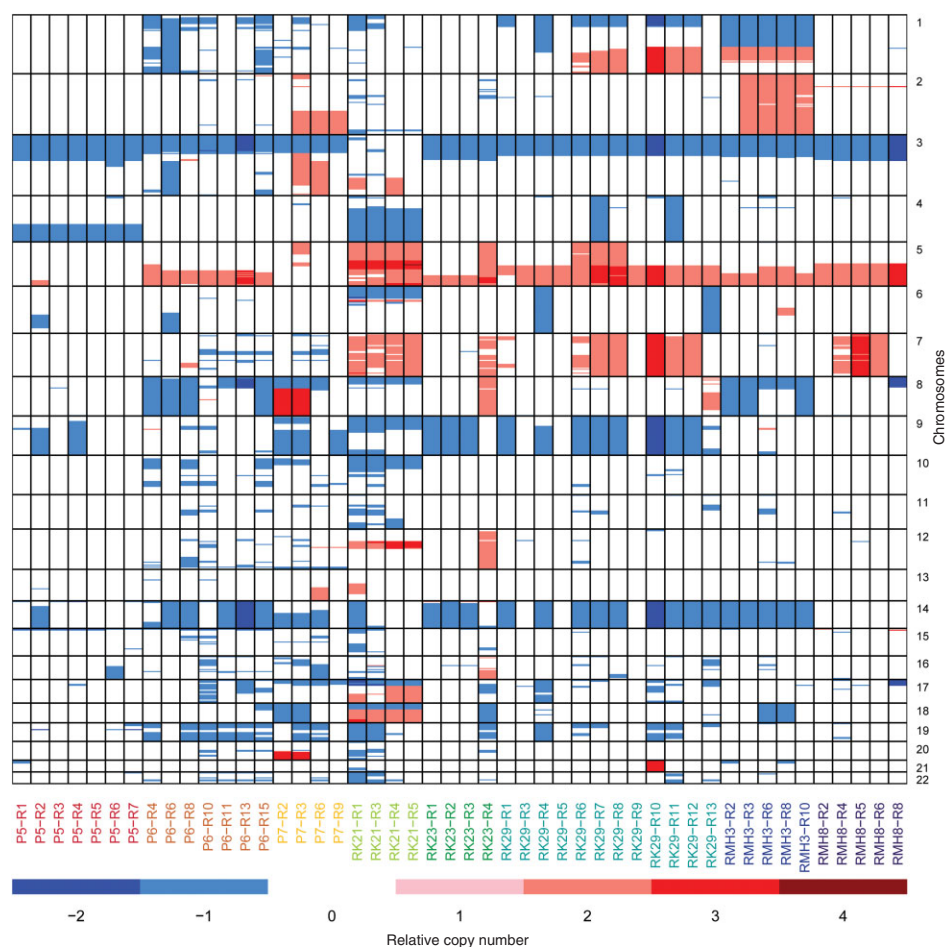


Figure 1. Copy number profiles of the 48 multi-region samples. Columns represent samples and rows represent cytobands. The colour coding indicates the relative copy number of each cytoband in each sample (the copy number observed in a cytoband minus the sample's ploidy). Sample labels are coloured to identify the patient of origin.

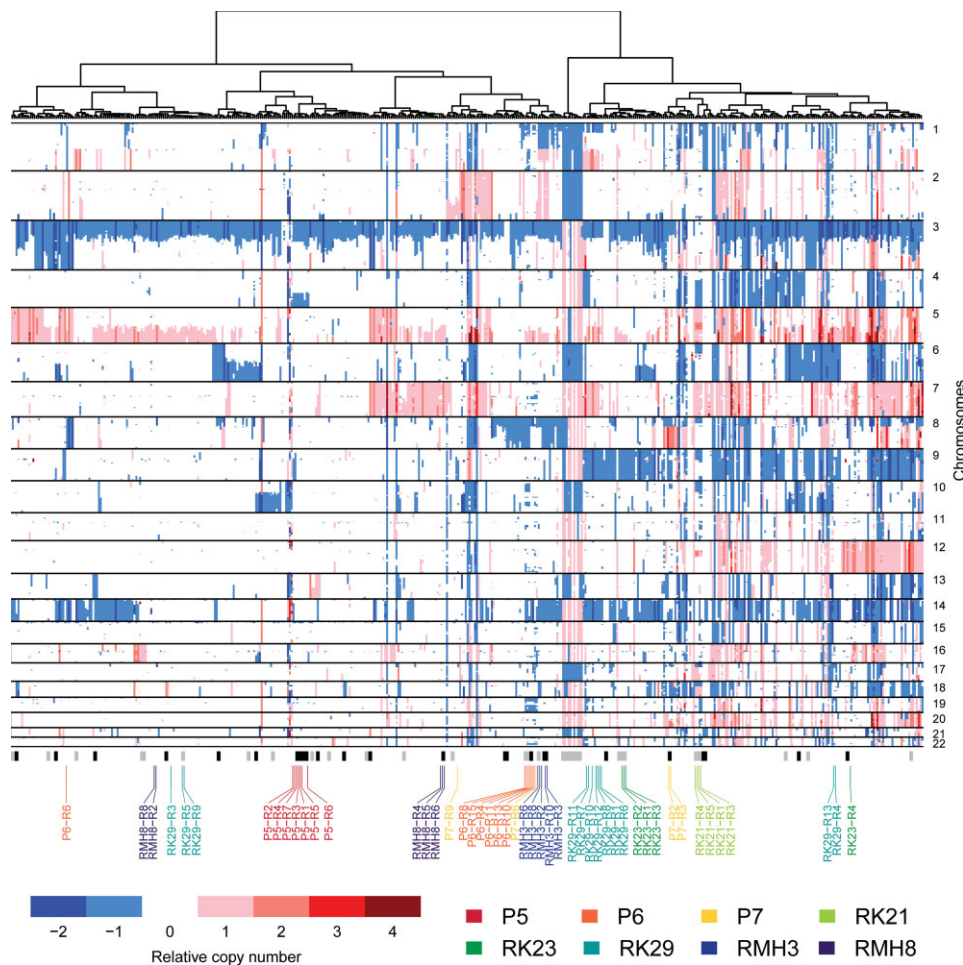
heterogeneity in stage III–IV ccRCCs may result in DNA copy number profiles that are more similar to copy number profiles of other tumours, including early-stage tumours, than to spatially separated biopsies from the same tumour. Similar results were obtained when the analysis was restricted to include only the subset of stage III and IV samples from the TCGA cohort (see Supplementary material, Figure S2), or using Euclidean rather than correlation-based distances (see Supplementary material, Figure S3).

We repeated the analysis using a model-based clustering method (hmmMix [33]; see Materials and methods) to assess whether an independent, non-hierarchical approach method would produce comparable results. Analyses in which the algorithm was instructed to group samples into three to 10 clusters consistently revealed that spatially separated biopsies from the same tumour resided in different DNA copy number clusters. Results are shown for four clusters (see Supplementary material, Figure S4), in which case five of the eight multi-region sampled tumours harboured biopsies that resided in two or more clusters. Only the samples from patients P6, RMH3 and RMH8 were located in the same cluster.

### Scale of intratumour heterogeneity mimics intertumour heterogeneity

To further assess how often a given biopsy resembled an unrelated sample more than any other biopsy from the same tumour, we identified the copy-number profile with the shortest sample-to-sample distance to each of the multi-region sample profiles (see Supplementary material, Table S2). Depending on the distance metrics used, at least one sample in up to six of the eight tumours was more similar to an unrelated sample rather than to any other sample from the same tumour. In total, 25% of the 48 biopsies from the eight tumours appeared to be more similar to an unrelated sample than to any related sample. This indicates that the similarity between one or more biopsies from the same tumour with unrelated samples is observed independently of clustering methods.

We used a network approach to illustrate how copy number profiles from the same tumour correlate with each other and those of unrelated stage III and IV samples (Figure 3). In this network, nodes represent samples (diamonds for stage III tumours, circles for stage IV) and the edges (lines) connecting them are weighted by the Spearman correlation-based



**Figure 2.** Hierarchical clustering of 440 TCGA samples and 48 multi-region samples from eight patients. Heat map columns represent samples and are ordered to match the dendrogram, while rows represent cytobands. The hierarchical clustering is based on Spearman's rank correlation coefficient between samples. The colour coding indicates the relative copy number of each cytoband in each sample. Significant clusters obtained using 1000 bootstrapping iterations are highlighted by bars below the copy number profiles of the relevant samples, using alternating grey and black for clarity. Multi-region samples are highlighted and reported below the significant clusters bar.

distances used for clustering in Figure 2 (see Materials and methods). Thick lines highlight pairs of samples with significant similarity to each other, as defined by a sample-to-sample distance which is smaller than expected by chance ( $p < 0.05$ ). In total, 36 of 48 (75%) of the multi-region samples are only significantly similar to some, but not all, other samples from their tumour of origin (see Supplementary material, Table S3). Of these 36 samples, 28 are also significantly similar to samples from unrelated high-stage tumours. Furthermore, one sample (RK23-R4) is only significantly similar to samples from different tumours and three samples (P6-R10, P7-R6 and P7-R9) are not significantly similar to any other sample. Network representation provides a deeper insight into the similarities between advanced ccRCC copy number profiles from common and unrelated origins. For instance, while the similarities of samples P5-R2 and P5-R4 with other samples from patient 5 are smaller than expected by chance, they are more similar and located closer to TCGA samples. The two subclonal populations observed in patient RK29 are also identifiable as two separate, highly interconnected groups.

Comparable results were obtained using Euclidean instead of correlation-based distances (see Supplementary material, Figure S5). These results suggest that intratumour heterogeneity in advanced ccRCCs can result in divergence between subclonal populations of the same tumour that is on a similar scale to that seen between different tumours. In addition, such a high divergence in copy number profiles originating from the same tumour can be observed in both untreated and Everolimus-treated samples.

#### ccRCC frequently altered loci demonstrate ITH occurring following 3p loss

Copy number gains and losses that are detected frequently in ccRCCs have been suggested to be driver events as they are recurrently selected [35]. Unsupervised clustering of genome-wide copy number profiles may favour the detection of complex aberration patterns but may fail to detect ccRCC subgroups that, from a biological point of view, might be better classified by such single recurrent genomic events. To assess this possibility, we identified chromosomal gains or

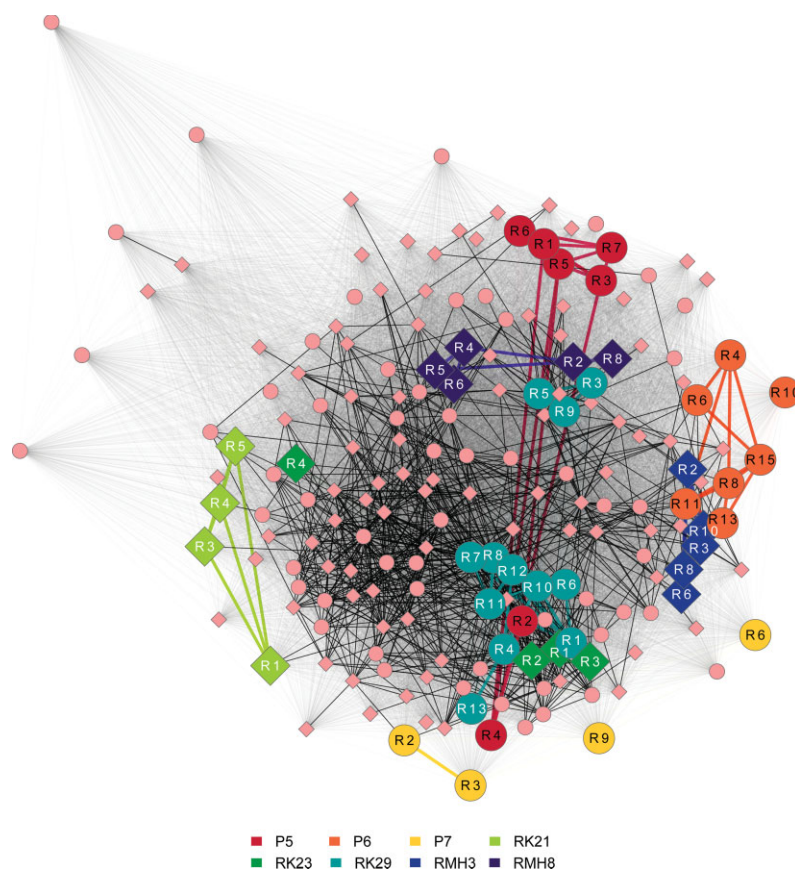


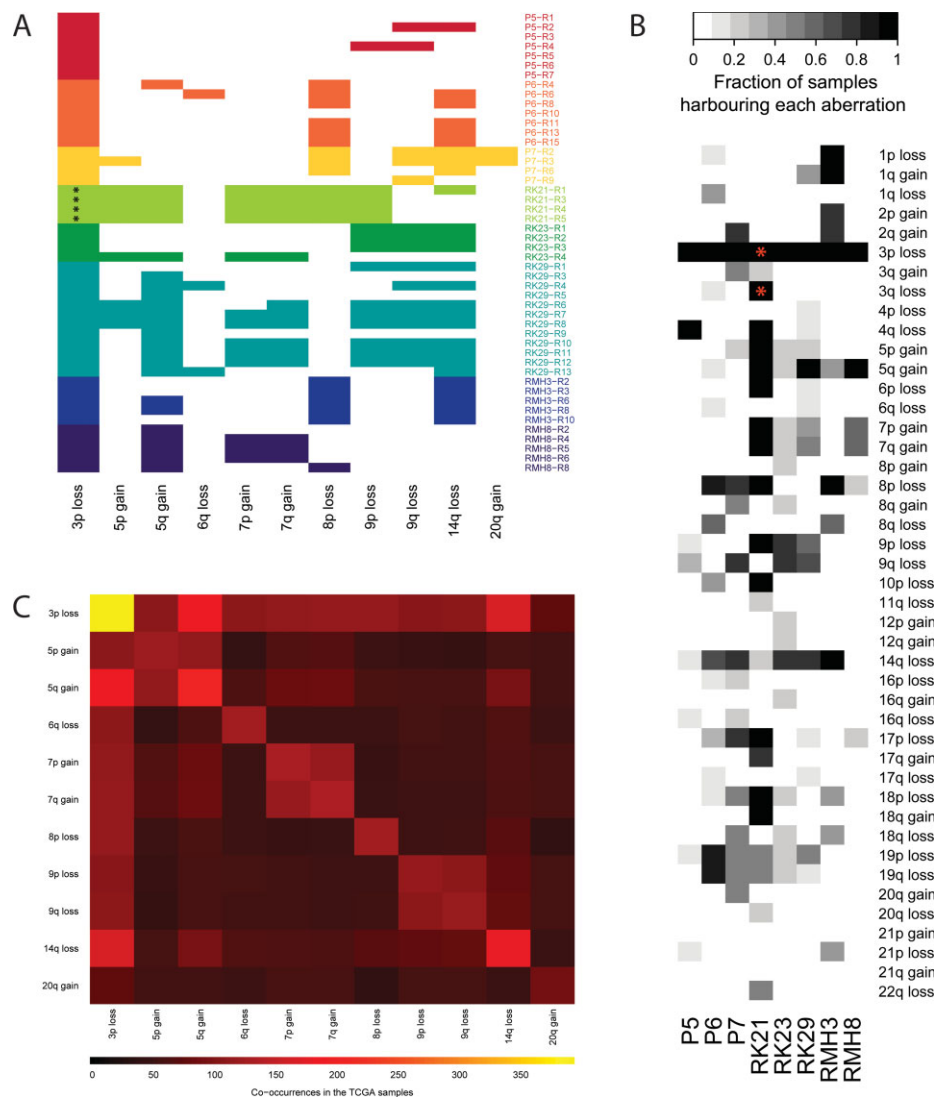
Figure 3. Network representation of relationships between multi-region and stage IV TCGA samples. Samples are represented as nodes (diamonds for stage III samples, circles for stage IV) and correlation-based distances are used to weight the edges connecting them. The network is fully interconnected and all edges are used to create the layout. Only edges corresponding to distances significantly smaller than expected (including the bottom 5% of the unrelated samples distance distribution) are drawn with thick lines. Coloured lines highlight connections between samples from the same tumour; black lines are used for unrelated samples. White node labels indicate regions from untreated tumours; black node labels indicate regions from treated tumours.

losses present in at least 20% of samples in the TCGA cohort. Recurrent chromosomal aberrations in kidney cancer are usually broad, involving most or the whole of a chromosome arm [35]. Similarly to a previous study [36], we defined such broad chromosome arm aberrations as gains or losses of at least half of the arm's cytobands in order to ensure high sensitivity and specificity (see Supplementary material, Figure S6). Losses were detected on chromosome arms 3p, 6q, 8p, 9p, 9q and 14q, and gains on 5p, 5q, 7p, 7q and 20q. Analysis of each of these events revealed concordant changes in gene expression in 400 TCGA samples for which RNA-seq data were also available (see Supplementary material, Figure S7), indicating that copy number changes likely impact upon gene expression in these regions of recurrent gain or loss.

Analysis of these recurrent chromosomal aberrations in the six multi-region sampled cases revealed that chromosome 3p was the only copy number event ubiquitously lost in all regions of all tumours (Figure 4A), supporting its known role as an early founder genetic event, important for tumour initiation. All other recurrent chromosomal aberrations were found to be heterogeneous in at least two patients. Thus, apart from the early loss of chromosome 3p, recurrent copy number

changes in advanced ccRCCs in this small cohort were confined to tumour subclones, as opposed to their ubiquitous alteration in all tumour cells (Figure 4A) [25]. This supports a parallel progression model in which novel driver events are unlikely to eradicate progenitor clones through a selective sweep, contributing to branched evolution and the polyclonal nature of ccRCCs.

Next we analysed the frequencies of all aberrations affecting at least half of a chromosome arm detected in at least one tumour sample from the multi-region copy number analysis (Figure 4B). The number of aberrations varied greatly both between different tumours and between samples from the same tumour (see Supplementary material, Table S4). These data suggest that 38–96% of the large chromosomal aberrations in an individual tumour are heterogeneous and that, on average, only 40% of the chromosomal aberrations in a single tumour biopsy derived from a late-stage ccRCC will be present ubiquitously across the tumour mass. Thus, the majority of somatic copy number gains and losses present in a single ccRCC biopsy, including recurrent and potential driver aberrations, were most likely acquired as subclonal events in tumour progression.



**Figure 4.** Large aberrations in the TCGA and multi-region samples. Large aberrations involve at least half of the length of a chromosome arm. (A) Occurrences of recurrent aberrations (observed in at least 20% of the TCGA samples) in the multi-region samples. Samples and chromosomal aberrations are represented by columns and rows, respectively. Colours differ for each patient. Colours indicate the presence of an aberration in a sample and white indicates their absence. (B) Frequencies of large aberrations in multi-region samples summarized for each patient. Aberrations that did not occur in the multi-region samples were not plotted. Darker colours indicate higher frequency of an aberration in samples from the same patient. (C) Co-occurrences of recurrent large aberrations in TCGA samples. \*Regions showing ubiquitous loss of heterozygosity without overall copy number loss.

Finally, we asked whether any of these recurrent aberrations were mutually exclusive in the TCGA cohort. Any two of the recurrent copy number gains or losses occurred together within single biopsies (Figure 4C). Thus, none of these putative driver aberrations appears to restrict evolutionary possibilities of ccRCC subclones in terms of further acquisition of additional recurrent copy number events.

## Discussion

Genomic assays are increasingly used for prognostic or predictive purposes to identify patients at high risk of recurrence for treatment stratification or to identify patient cohorts that may be preferentially treated with specific targeted approaches. Accumulating evidence

suggests that intratumour heterogeneity impacts upon disease outcome [19,21], therapeutic response [23,37] and biomarker analysis [15].

Understanding the scale of genetic ITH and its impact on personalized medicine approaches, patient outcome and therapeutic resistance is of major clinical importance. Here, we extend our recent work in ccRCCs, where we characterized ITH by examining different biopsies from the same tumour [15]. We now examined DNA copy number characteristics derived from multiple tumour regions from each of eight patients and their relationships to single tumour samples derived from 440 ccRCCs from TCGA. Our data indicate that copy number aberrations are likely to contribute to both inter- and intratumour heterogeneity in ccRCCs and reveal that different tumour biopsies from the same patient may cluster more closely



with other tumours. Our findings further suggest that, in late stage renal cancer, the extent of intratumour copy number heterogeneity may closely resemble the differences observed across large panels of individual tumours. These results may have relevance across other tumour types; in particular, Teixeira and colleagues have previously reported similar observations in primary breast cancer tumours [38].

These data support a parallel progression model in which subclones in individual tumours acquire diverse copy number profiles early after the loss of the p arm of chromosome 3. Parallel subclone progression may then lead to subclonal driver aberrations that contribute to intratumour heterogeneity and tumour sampling bias. Similarly to evidence of subclonal mutations impacting upon drug treatment and survival [39], one or more of these subclonal copy number events may contribute to tumour progression. Chromosome 9p deletion, for instance, has been associated with worse prognosis and higher risk of disease recurrence [40]. In our study, 9p alterations are present as heterogeneous copy number aberrations in two of eight multi-region sampled tumours. Intratumour diversity, where biopsies from the same tumour may resemble other tumours more than the tumour from which they were derived, illustrates the scale of the genetic dynamics within a single tumour mass. The identification of recurrent ubiquitous genetic events, which are always located in the tumour 'trunk', and of functional dependencies and clinical characteristics associated with such events, may serve to mitigate the challenge of tumour sampling bias in ccRCCs and provide more robust targets for therapeutic intervention.

Consistent with our previous demonstration that good and poor prognostic gene expression signatures can co-exist within individual ccRCCs [15], this analysis indicates that the diversity amongst individual primary ccRCCs may complicate the development of copy number-based biomarkers and their therapeutic application. Further studies are required to assess whether the predictive power of genetic tests can be increased by multi-region assessment. The development of non-invasive methods to comprehensively profile the genetics of heterogeneous tumours, for instance exploiting circulating tumour DNA, may provide a further step towards understanding the clinical implications of intratumour heterogeneity and help to resolve genetic diversity present within individual tumours. Such techniques would allow the dynamics of subclonal events to be profiled within tumours at multiple time points during the disease course in order to elucidate the contribution of these copy number aberrations to tumour evolution.

### Acknowledgements

This work was funded by the European Union Framework 7 programme PREDICT Consortium, Villum Fonden and Cancer Research UK. ZS and CS were

supported by the Breast Cancer Research Foundation. CS and MG were funded by the Prostate Cancer Foundation and the UK Medical Research Council. JL was supported by the NIHR RMH/ICR Biomedical Research Centre for Cancer. The results published here are in part based upon data generated by the Cancer Genome Atlas pilot project established by the NCI and NHGRI (information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>). The data were retrieved through dbGaP authorization (Accession No. phs000178.v5.p5). We thank the Royal Marsden Hospital trials staff, all patients and their relatives who participated in this study; Martin Gore and Lisa Pickering for their help in the sample acquisition; and David Endesfelder, Max Salm and Gavin Kelly for their help with the bioinformatics analysis.

### Author contributions

PM, NJB, MG, RAB, NM, ZS, and CS designed the experiments; PM, NJB and TJ performed the bioinformatics analysis; PM, NJB, MG and CS interpreted the data and wrote the article; RF and JL provided the clinical samples; AJR processed all biological samples; and ZS and CS supervised the work.

### References

1. Parsons DW, Jones S, Zhang X, *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008; **321**: 1807–1812.
2. Sjoblom T, Jones S, Wood LD, *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* 2006; **314**: 268–274.
3. Stephens PJ, McBride DJ, Lin ML, *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009; **462**: 1005–1010.
4. Varela I, Tarpey P, Raine K, *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* 2011; **469**: 539–542.
5. Bell D, Berchuck A, Birrer M, *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* 2011; **474**: 609–615.
6. Ding L, Ley TJ, Larson DE, *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012; **481**: 506–510.
7. Anderson K, Lutz C, van Delft FW, *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 2011; **469**: 356–361.
8. Wu X, Northcott PA, Dubuc A, *et al.* Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature* 2012; **482**: 529–533.
9. Navin N, Kendall J, Troge J, *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* 2011; **472**: 90–94.
10. Navin N, Krasnitz A, Rodgers L, *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res* 2010; **20**: 68–80.
11. Shah SP, Morin RD, Khattra J, *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 2009; **461**: 809–813.
12. Shah SP, Roth A, Goya R, *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012; **486**: 395–399.

13. Nik-Zainal S, Alexandrov LB, Wedge DC, *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012; **149**: 979–993.
14. Nik-Zainal S, Van Loo P, Wedge DC, *et al.* The life history of 21 breast cancers. *Cell* 2012; **149**: 994–1007.
15. Gerlinger M, Rowan AJ, Horswell S, *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; **366**: 883–892.
16. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer* 1957; **11**: 359–377.
17. Walther A, Houlston R, Tomlinson I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut* 2008; **57**: 941–950.
18. Carter S, Eklund A, Kohane I, *et al.* A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006; **38**: 1043–1048.
19. McGranahan N, Burrell RA, Endesfelder D, *et al.* Cancer chromosomal instability: therapeutic and diagnostic challenges. 'Exploring aneuploidy: the significance of chromosomal imbalance' review series. *EMBO Rep* 2012; **13**: 528–538.
20. Roylance R, Endesfelder D, Gorman P, *et al.* Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. *Cancer Epidemiol Biomarkers Prev* 2011; **20**: 2183–2194.
21. Maley CC, Galipeau PC, Finley JC, *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet* 2006; **38**: 468–473.
22. Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br J Cancer* 2010; **103**: 1139–1143.
23. Su KY, Chen HY, Li KC, *et al.* Pretreatment epidermal growth factor receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *J Clin Oncol* 2012; **30**: 433–440.
24. Keats JJ, Chesi M, Egan JB, *et al.* Clonal competition with alternating dominance in multiple myeloma. *Blood* 2012; **120**: 1067–1076.
25. Yap T, Gerlinger M, Futreal A, *et al.* Intratumour heterogeneity: seeing the wood for the trees. *Sci Translat Med* 2012; **4**: 127ps10.
26. Bengtsson H, Wirapati P, Speed TP. A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. *Bioinformatics (Oxford, England)* 2009; **25**: 2149–2156.
27. Bengtsson H, Neuvial P, Speed T. TumorBoost: normalization of allele-specific tumor copy numbers from a single pair of tumor-normal genotyping microarrays. *BMC Bioinform* 2010; **11**: 245.
28. Ortiz-Estevéz M, Aramburu A, Bengtsson H, *et al.* CalMaTe: a method and software to improve allele-specific copy number of SNP arrays for downstream segmentation. *Bioinformatics* 2012; **28**: 1793–1794.
29. Van Loo P, Nordgard S, Lingjærde OC, *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA* 2010; **107**: 16910–16915.
30. Fujita PA, Rhead B, Zweig AS, *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 2011; **39**: D876–882.
31. Burrell RA, McClelland SE, Endesfelder D, *et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* 2013; **494**: 492–496.
32. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 2006; **22**: 1540–1542.
33. Shah SP, Cheung KJ, Jr., Johnson NA, *et al.* Model-based clustering of array CGH data. *Bioinformatics* 2009; **25**: i30–38.
34. Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–2504.
35. Beroukhi R, Brunet J-P, Di Napoli A, *et al.* Patterns of gene expression and copy-number alterations in von Hippel–Lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res* 2009; **69**: 4674–4681.
36. Weir BA, Woo MS, Getz G, *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007; **450**: 893–898.
37. Lee A, Endesfelder D, Rowan A, *et al.* Chromosomal instability confers intrinsic multidrug resistance. *Cancer Res* 2011; **71**: 1858–1870.
38. Torres L, Ribeiro F, Pandis N, *et al.* Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. *Breast Cancer Res Treat* 2007; **102**: 143–155.
39. Landau DA, Carter SL, Stojanov P, *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 2013; **152**: 714–726.
40. La Rochelle J, Klatte T, Dastane A, *et al.* Chromosome 9p deletions identify an aggressive phenotype of clear cell renal cell carcinoma. *Cancer* 2010; **116**: 4696–4702.

## SUPPLEMENTARY MATERIAL ON THE INTERNET

The following supplementary material may be found in the online version of this article:

**Figure S1.** Copy-neutral loss of heterozygosity in patient RK21

**Figure S2.** Hierarchical clustering of 143 stage III and IV TCGA samples and 48 multi-region samples from eight patients

**Figure S3.** Hierarchical clustering of 440 TCGA samples and 48 multi-region samples from eight patients

**Figure S4.** Model-based clustering of 440 TCGA samples and 48 multi-region samples from eight patients with hmmMix

**Figure S5.** Network representation of relationships between multi-region and stage III and IV

**Figure S6.** Overall copy-number profiles of all samples used in the hierarchical clustering analysis

**Figure S7.** Expression of genes impacted by recurrent, large copy number aberrations

**Table S1.** Patient clinical and pathological data

**Table S2.** List of the samples most similar to each multi-region profile, as defined by smallest correlation-based and Euclidean distances

**Table S3.** Number of significantly similar samples

**Table S4.** Number of large aberrations (involving at least half of the length of a chromosome arm) per multi-region sample and number of average vs ubiquitous number of aberrations per patient