IJC
International Journal of Cancer

The genomic imprint of cancer therapies helps timing the formation of metastases

Eszter Németh[1], Marcin Krzystanek[2], Lilla Reiniger[3,4], Dezső Ribli[5], Orsolya Pipek[5], Zsófia Sztupinszki[2], Tibor Glasz[6], István Csabai[5], Judit Moldvay[4,7], Zoltan Szallasi[2,4,8] and Dávid Szüts[1]

[1]Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
[2]Translational Cancer Genomics, Danish Cancer Society Research Center, Copenhagen, Denmark
[3]1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary
[4]SE-NAP Brain Metastasis Research group, 2nd Department of Pathology, Semmelweis University, Budapest, Hungary
[5]Department of Physics of Complex Systems, Eötvös University, Budapest, Hungary
[6]2nd Department of Pathology, Semmelweis University, Budapest, Hungary
[7]Department of Tumor Biology, National Korányi Institute of Pulmonology—Semmelweis University, Budapest, Hungary
[8]Computational Health Informatics Program, Boston Children's Hospital, Harvard Medical School, Boston, MA

A retrospective determination of the time of metastasis formation is essential for a better understanding of the evolution of oligometastatic cancer. This study was based on the hypothesis that genomic alterations induced by cancer therapies could be used to determine the temporal order of the treatment and the formation of metastases. We analysed the whole genome sequence of a primary tumour sample and three metastatic sites derived from autopsy samples from a young never-smoker lung adenocarcinoma patient with an activating EGFR mutation. Mutation detection methods were refined to accurately detect and distinguish clonal and subclonal mutations. In comparison to a panel of samples from untreated smoker or never-smoker patients, we showed that the mutagenic effect of cisplatin treatment could be specifically detected from the base substitution mutations. Metastases that arose before or after chemotherapeutic treatment could be distinguished based on the allele frequency of cisplatin-induced dinucleotide mutations. In addition, genomic rearrangements and late amplification of the EGFR gene likely induced by afatinib treatment following the acquisition of a T790M gefitinib resistance mutation provided further evidence to tie the time of metastasis formation to treatment history. The established analysis pipeline for the detection of treatment-derived mutations allows the drawing of tumour evolutionary paths based on genomic data, showing that metastases may be seeded well before they become detectable by clinical imaging.

Introduction

Metastatic progression makes an important contribution to the mortality of multiple cancer types, but clinical imaging provides limited information on when metastases first arise, and whether they are already present at the time of the first treatment. Genomic analyses can retrospectively help understanding the evolution of a tumour and distinguishing between early or late metastasis formation.[1,2] Our aim was to use the example of lung adenocarcinoma to tie metastasis formation to timed events in the patient's treatment history. An improved knowledge of the typical times of metastatic seeding could help with weighing surgical and medical treatment options.

Multiregion sequencing has been used for mapping evolution within the primary tumour in lung adenocarcinoma,[3,4]

Key words: lung adenocarcinoma, cisplatin, whole genome sequencing, mutagenic chemotherapy, EGFR amplification, tumour phylogeny, EGFR T790M, afatinib
Additional Supporting Information may be found in the online version of this article.
Disclosure declaration: The authors declare that they have no conflicts of interest.
Grant sponsor: Breast Cancer Research Foundation; **Grant number:** BCRF-17-156; **Grant sponsor:** Kræftens Bekæmpelse; **Grant number:** R90-A6213 ; **Grant sponsor:** Novo Nordisk; **Grant number:** NNF15OC0016584; **Grant sponsor:** Research and Technology Innovation Fund of Hungary; **Grant numbers:** NAP2-2017-1.2.1-NKP-0002, KTIA_NAP_13-2014-0021, FIEK_16-1-2016-0005
DOI: 10.1002/ijc.32159

History: Received 3 Oct 2018; Accepted 16 Jan 2019; Online 29 Jan 2019
Correspondence to: Zoltan Szallasi, Boston Children's Hospital, 300 Longwood Ave, Boston, MA, 02215, E-mail: zoltan.szallasi@childrens.harvard.edu; or Dávid Szüts, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudosok krt 2, Budapest, H-1117, Hungary, Tel.: +36-13826708, E-mail: szuts.david@ttk.mta.hu

What's new?

Cancer treatment induces mutations either through direct DNA damage or through the evolutionary selection of resistance mutations. The authors exploited this effect to study temporal tumor phylogeny and the formation of metastases. Using whole-genome sequences isolated from a patient with lung adenocarcinoma, they identify metastases that arose before or after onset of treatment with platinum and tyrosine kinase inhibitors, underscoring that metastases develop and need to be treated much earlier than they become clinically apparent.

and post-mortem sequencing of metastases can elucidate what further genomic changes take place during metastasis formation.[5,6] However, multiregion whole genome sequencing (WGS) studies, which provide the broadest and most reliable view on somatic mutagenesis and gene copy number changes,[7,8] have not been performed in lung cancer. Even with available WGS data, it is difficult to obtain timing information for the evolutionary steps from contemporaneous samples. In the current study we reasoned that if treatments leave recognisable marks on the genome, the treatment-related mutations could provide extra information for the timing of evolutionary branchpoints and metastasis formation.

We have recently demonstrated the mutagenicity of the cytotoxic agents cisplatin and cyclophosphamide in cell line-based studies,[9] and an experimentally derived cisplatin mutation spectrum was subsequently found in cisplatin-treated esophageal and liver tumours.[10] If certain somatic mutations in tumour samples can be assigned to a treatment agent that the patient received, then the branching time of two samples can be determined relative to the time of the treatment based on whether the treatment-derived mutations in two samples are unique or shared. Accurate mutation detection is critical for such analyses, confounded by varying tumour content and the varying allele frequency (AF) of subclonal mutations, necessitating a careful bioinformatics approach.

Treatments may also select for known, identifiable resistance-causing mutations,[11] whose appearance can help understanding evolutionary paths.

In order to study the evolution of lung adenocarcinoma, we performed whole genome sequencing on the primary tumour and multiple metastases of a young non-smoker patient with *EGFR* mutant cancer. In addition to kinase activating mutations, *EGFR* may also be subjected to genomic amplification. Treatment with tyrosine kinase inhibitors almost invariably leads to resistance, in about half the cases through the acquisition of an *EGFR* T790M mutation.[12,13] We were able to detect ongoing *EGFR* amplification following the acquisition of gefitinib resistance through T790M mutation, and importantly also showed that mutations caused by cisplatin treatment can be used to time the formation of metastases. The demonstrated analysis pipeline allows the drawing of tumour evolutionary paths, which in this case distinguish metastases formed before and after the cisplatin treatment, and also show the progress of cells through the same lymph node to distinct distant metastases.

Materials and Methods

DNA isolation

Written informed consent was obtained from the patient to perform genomic analyses of the tumour and peripheral blood samples. Permissions to use the archived tissue have been obtained from the Regional Ethical Committee (No: 510/2013, 86/2015).

Peripheral blood taken during routine diagnostic tests was collected and frozen. Primary tumour and metastasis samples were collected and frozen during autopsy. The formalin-fixed paraffin embedded bronchoscopy biopsy sample of the primary tumour was also used in this study. Genomic DNAs were extracted with the High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's recommendations.

DNA sequencing

One hundred and fifty basepairs paired end whole genome DNA sequencing to a mean coverage of 60× (peripheral blood) or 61×-86× (tumour samples) was performed on Illumina HiSeq X Ten instruments at Edinburgh Genomics (Edinburgh, UK). For detailed coverage information, see Supporting Information Methods.

Somatic mutation calling

Alignment of the sequencing reads to the reference genome GRCh38/hg38 was performed with the Burrows-Wheeler alignment algorithm,[14] followed by post-processing with the IndelRealigner tool of the Genome analysis Toolkit (GATK, version 3.4).[15] Somatic mutations, insertions and deletions were obtained using the GATK MuTect2 mutation caller.[16] The merged list of mutation positions identified by MuTect2 from all samples was used to detect subclonal mutations. This was done by extracting the AFs directly from the binary alignment (BAM) files. The supplemented data set was subjected to carefully optimised post-filtering (see Supporting Information Methods). In the final filter, one mutation-containing read was allowed in the control blood sample to allow for index switching noise.[17] For the tumour samples, 0 or at least 3 mutation containing reads were allowed, considering the positions with 1 or 2 mutated reads as potential noise.

Germline and somatic DNA analysis

GATK HaplotypeCaller was used to detect all germline variations in the genome. The determined polymorphisms were analysed and annotated with InterVar.[18] Mutations classified as 'likely pathogenic' or 'pathogenic' were considered. The somatic

mutations determined by MuTect2 and the post-processing were also annotated with InterVar.

Structural variations

Structural variants were determined by the CREST algorithm.[19] Post-filtering was performed with the condition of minimum 3 reads at both the left and right clipped chromosome. Chromosome translocations were visualised with the circlize R package.[20]

Estimation of tumour content and EGFR allele numbers

Tumour content of the samples was estimated based on the heterozygous polymorphisms (SNPs) in the *TP53* gene, located in a region with loss of heterozygosity and normal coverage. AFs at these SNPs were used to determine tumour content with high precision as an average of the estimated tumour content in each of these positions.

The genomic region around the *EGFR* gene presented an increased coverage in the tumour samples because of amplification. To determine the number of alleles containing the exon 19 deletion (e19del) mutation we used the AF at this site, corrected with the deletion-related drop in coverage and the tumour content determined from *TP53* polymorphisms. The possibility of the amplification of WT allele in tumour cells was rejected based on estimations from the nearby polymorphisms. Based on the tumour content and EGFR e19del alleles the number of alleles simultaneously containing the T790M mutation could be estimated in a similar manner (see Supporting Information Methods for more details).

Decomposition of mutational spectra

The sequence context of the preceding and following base was determined for each mutation. The resulting triplet SNV spectra, were analysed for contributions of known mutational signatures in the COSMIC cancer mutation database.[21] The deconstructSigs R package[22] was applied using a restricted set of COSMIC signatures that included ageing and tissue-specific cancer signatures[1,2,4–6,12,13,15–17,23] supplemented with the signature drawn from cisplatin-treated human cell lines,[10] and with a minimum signature contribution of 6%.

Analysis of TCGA data

WGS BAM files of treatment naïve lung cancer cases were downloaded from The Cancer Genome Atlas (TCGA). Somatic mutations were called by MuTect2 and were used to construct the 96 category-based triplet SNV spectra. These were deconstructed using COSMIC and cisplatin signatures as described above.

Model drawing

The qualitative model of tumour evolution was built using figures constructed with the fishplot R package.[23]

Data access

Binary alignment map (BAM) files of the primary tumour, the bone, liver, lymph node metastases, and the peripheral blood sample, as well as the mutational data including SNVs, DNVs, short insertions and deletions and chromosome translocations are accessible at the European Genome-phenome Archive under study ID EGAS00001003416.

Results

The collection of post-chemotherapy metastatic tumour samples

WGS data deposited in databases usually originates from the time before chemotherapy, hampering the analysis of the genetic effects of the treatments. For this study numerous samples were collected from a complex metastatic lung adenocarcinoma case, in which cisplatin chemotherapy was applied prior to the clinically detectable emergence of the sampled metastases. The course of the disease and the treatment history are shown on Figure 1.

Gefitinib therapy was started on the exon 19 deletion (e19del) positive *EGFR* mutant tumour, and subsequent chest CTs showed partial response (Supporting Information Fig. S1). Sixteen months later an *EGFR* T790M resistance mutation was detected in a bronchoscopy biopsy sample of the primary tumour. Further treatments included one cycle of pemetrexed-cisplatin chemotherapy, followed by pemetrexed monotherapy, and a switch to the second generation irreversible EGFR inhibitor afatinib (Fig. 1 and Supporting Information Fig. S1). Samples were taken from the primary tumour and the vertebral, liver and paratracheal lymph node metastases at autopsy. None of these metastases were observable during the preceding clinical investigations. A detailed description of the clinical observations and sample collection can be found in the Supporting Information Methods.

Accurate detection of somatic mutations in metastatic cancer

An initial attempt at detecting somatic single nucleotide variations (SNVs) and small insertions and deletions (indels) was made using MuTect2[16] on each sample *versus* a sequenced peripheral blood (PB) DNA sample. Using this pipeline, we detected a total of 17,415 SNVs and 4,412 indels (Supporting Information Fig. S2). We found fewer mutations in the samples with lower tumour content (lung and lymph node), suggesting that a fraction of low AF mutations were missed in these samples. In addition, we found a considerable number of mutations common to any combination of tumour samples, which is incompatible with drawing a phylogenetic tree. Based on the common origin of the tumour sites, we were able to extract further low AF mutations from the raw data. After subsequent filtering, we increased the number of detected mutations common to all samples and reduced the number of paradoxical mutations common to various subsets of the samples (Supporting Information Figs. S2 and S3, more details in

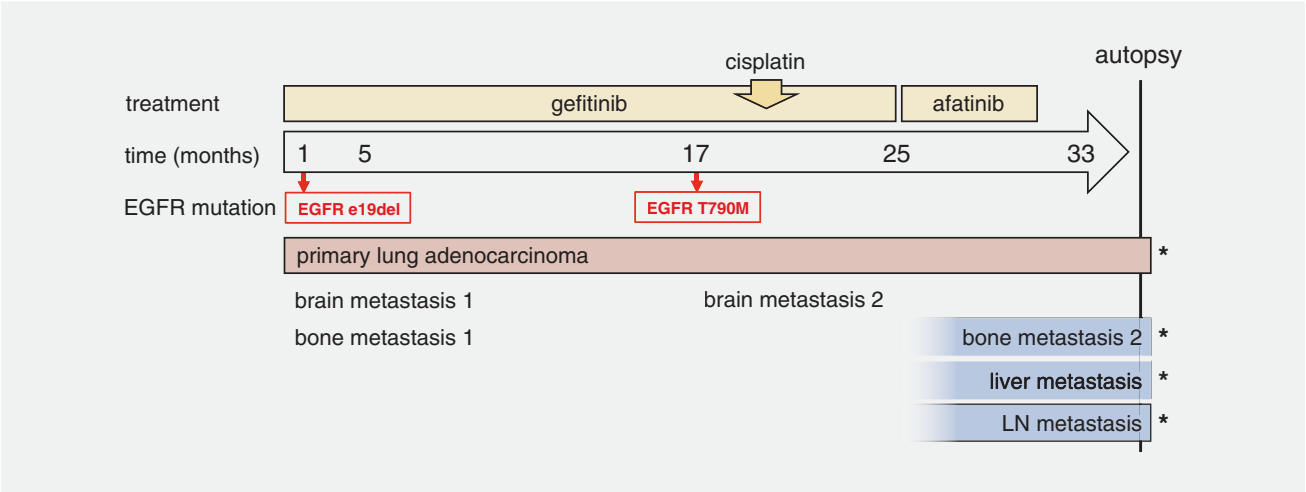

Figure 1. Disease history of the young never-smoker patient. The main diagnostic events and relevant treatments are shown over a timeline in months. Collected samples are indicated with asterisks. [Color figure can be viewed at wileyonlinelibrary.com]

the Supporting Information Methods). Similar post-filtering steps were used for indels. The resulting final data set contained a total of 9,832 SNVs, 330 insertions and 485 deletions (Figs. 2*a*–2*d*). The distribution of the SNV AFs suggested higher tumour content in the bone and liver metastasis samples, but double AF peaks in the same samples indicate that subclonal mutations were also found (Fig. 2*e*).

Only two driver mutations in metastatic lung adenocarcinoma

To understand the genetic drivers of the investigated primary tumour and its metastases, we first looked for somatic mutations that are classified as pathogenic or likely pathogenic by the InterVar tool.[18] We found in all samples the *EGFR* kinase-activating e19del mutation, and the *EGFR* T790M mutation responsible for gefitinib resistance (Supporting Information Table S1). In addition, a homozygous *TP53* R280T mutation was present in all samples, resulting from a loss-of-heterozygosity event on chromosome 17. The same *TP53* mutation has been repeatedly observed in lung adenocarcinoma (LUAD) as well as other cancer types.[24] No further somatic drivers common to all samples were found, suggesting that *EGFR* activation and *TP53* inactivation were sufficient to trigger lung adenocarcinoma.

Potentially pathogenic somatic mutations were found in one allele of the *XRN1*, *EVC2* and *FGA* genes in subsets of the samples (Supporting Information Table S1). *XRN1* has been suggested as a candidate tumour suppressor gene in osteosarcoma,[25] but none of these three genes have been confirmed as oncogenic drivers in lung cancer. Therefore, it appears that the formation of metastases was not induced by the acquisition of extra driver mutations.

Six genes contained potentially pathogenic heterozygous germline mutations (Supporting Information Table S1). The *RET* T791Y mutation was most interesting as it can trigger the ligand-independent activation of the RET receptor, thereby contributing to the activation of the JAK/STAT3 pathway.[26] Although *RET* activating mutations are widely found in thyroid cancers, only rearrangements of the *RET* gene are common in NSCLC.[27] Indeed, while the *RET* Y791F mutation was initially considered a low risk mutation, it was later reclassified as a likely non-pathogenic polymorphism.[28,29]

Somatic mutagenesis across multiple metastases

Patterns of somatic mutations in cancer genomes provide important information concerning tumour aetiology.[30] When viewed as triplet mutation patterns in the context of the neighbouring bases, the SNV patterns in the four tumour samples were very similar (Fig. 2*f*). All categories of base substitutions were present, with a slight dominance of C > T and C > A mutations. When compared to common cancer mutational signatures[21,31] the SNV patterns of the analysed lung adenocarcinoma samples bore closest resemblance to the similarly broad-spectrum signatures 5, 8 and 3, and little resemblance to the smoking-related signature 4, in agreement with the non-smoking history of the patient (Fig. 2*g*). The spectrum of common mutations, indicative of mutagenesis early in the life of the tumour, was also similar to the SNV spectra of the individual metastatic samples (Fig. 2*g* and Supporting Information Fig. S4). The lack of a marked difference between the mutagenic processes in the early tumour and the different metastases is in agreement with the lack of novel driver mutations in metastases, and also suggests that the changed tissue environment following metastasis formation had no influence on mutagenesis in tumour cells.

Detection of cisplatin treatment-derived mutations

The main aim of this study was to detect and utilise treatment-induced mutations for the purpose of understanding the history

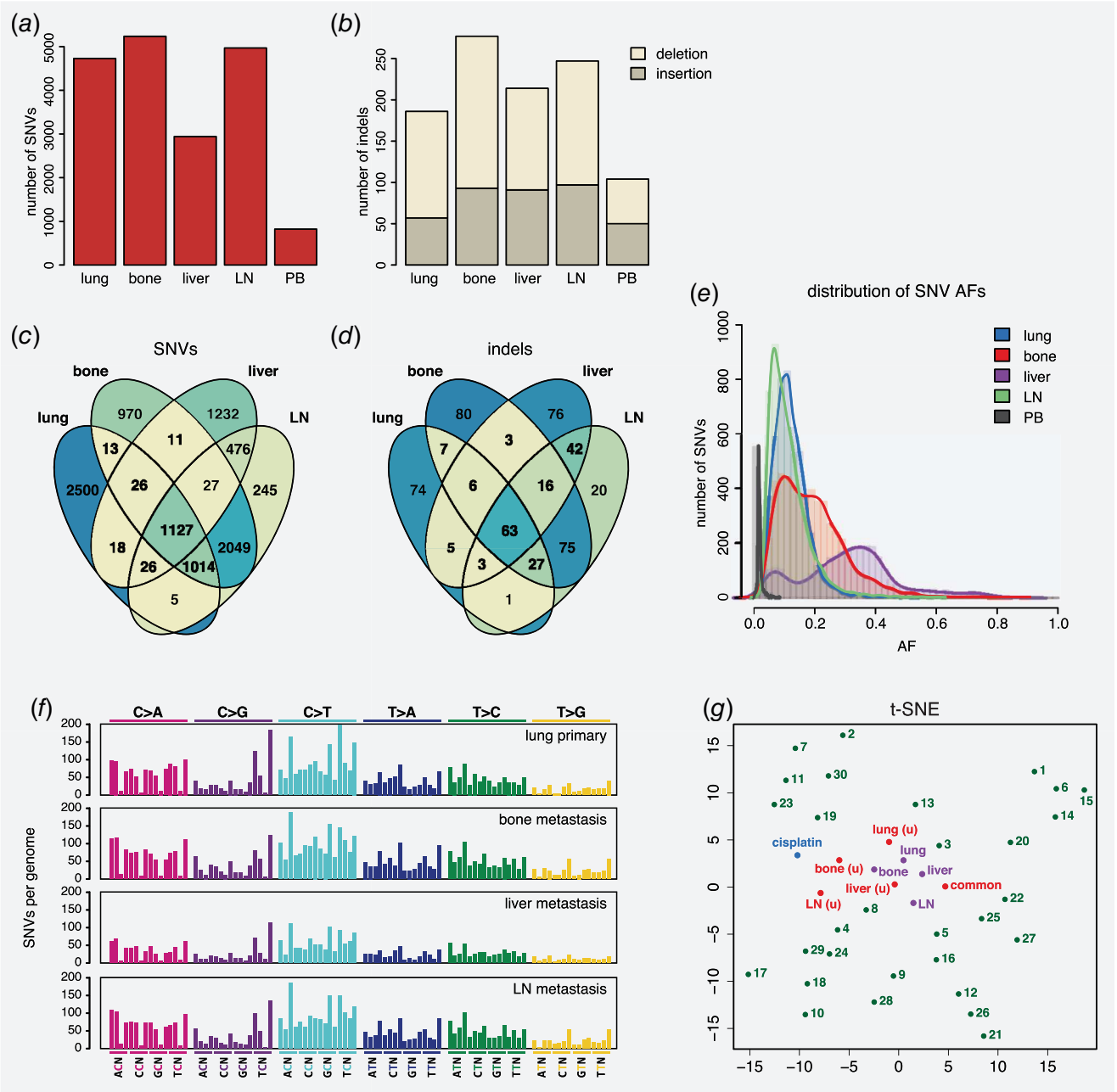

Figure 2. Characterisation of mutations in whole genome sequences. (*a*) Number of SNVs in the lung, bone, liver and the lymph node (LN) samples. The PB column accounts for those mutations that are present in the peripheral blood sample with one sequence read, originating mainly from noise. (*b*) Number of insertions and deletions. (*c*, *d*) Venn diagrams showing the distribution of SNVs and indels amongst the samples, respectively. (*e*) Distribution of allele frequencies of the detected SNVs. (*f*) Triplet SNV spectra of the tumour samples. Each mutation class, as indicated at the top of the panel, is separated into 16 categories based on the identity of the preceding and following nucleotide as shown below. The order of the following nucleotides, not shown due to lack of space, is alphabetical. (*g*) Similarity of the determined triplet spectra (purple) to COSMIC cancer signatures (numbered) as shown on a t-distributed stochastic neighbour embedding (t-SNE) plot. The spectrum of mutations that are unique (u) to individual samples, common to all samples, and the cell line-derived cisplatin spectrum are also shown. [Color figure can be viewed at wileyonlinelibrary.com]

of metastatic tumours. Our expectation was that treatment-induced mutations would likely be subclonal if the treatment hit an existing tumour site or metastasis, but clonal if a metastasis was formed from a single seeding cell following the treatment. We could indeed detect mutations with different AFs (Fig. 3*a*).

SNVs common to all tumour sites had pronouncedly higher AF in the lung, bone and lymph node sample than

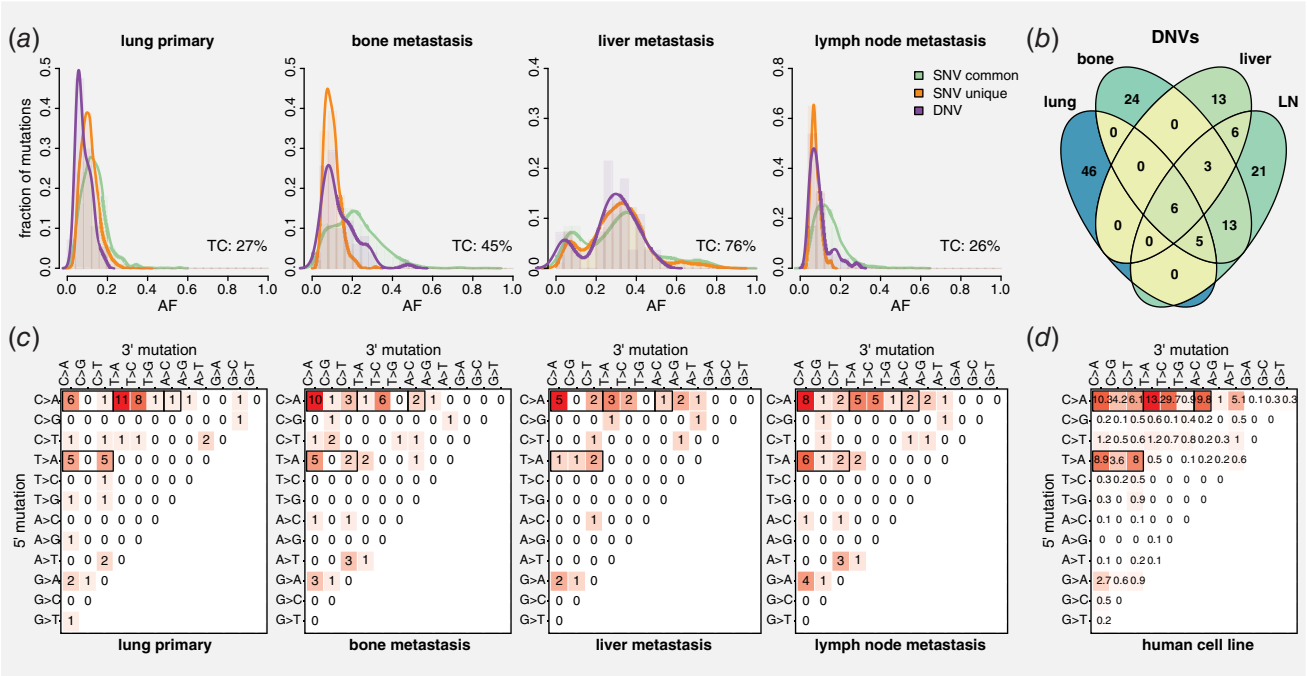

Figure 3. Characterisation of DNVs. (*a*) Allele frequency distribution of DNVs as compared to the SNVs, classified based on the Venn diagram (unique if found only in one sample and common if found in all four tumour samples). (*b*) Venn diagram of DNVs. (*c*, *d*) Sequence categories of DNVs. The categories most predominant in the cisplatin treated human cell lines[10] (*d*) are shown in black boxes. [Color figure can be viewed at wileyonlinelibrary.com]

unique SNVs, suggesting that we could detect subclonal mutations in the unique SNV pool (Supporting Information Fig. S5). Dinucleotide mutations (DNVs) had a similar AF distribution to unique SNVs in all samples, and we could confirm that the majority of DNVs were unique to individual tumour samples (Fig. 3*b* and Supporting Information Fig. S6). DNVs associate with certain SNV signatures,[32] and are otherwise rare in cancer genomes. However, we showed earlier that cisplatin induces specific types of dinucleotide mutations at sites of intrastrand crosslinks[9] and the cisplatin treatment of human cell lines led to a closely related pattern of DNV mutations to what we found in chicken DT40 cells.[10] Remarkably, we found very similar types of DNVs in the sequenced samples to those induced by cisplatin in cell lines (Figs. 3*c* and 3*d*). The most common types of DNVs were CC > AA and CT > AA, consistent with mutations forming at G ̂G and A ̂G cisplatin intrastrand adducts.

To further explore whether the single cycle of cisplatin treatment administered to the patient was mutagenic, we defined a cisplatin-induced triplet SNV signature based on the mutation spectrum of cisplatin-treated human cell lines[10] and looked for its contribution to the detected SNV pools. A decomposition into all triplet signatures implicated in cancers of the lung and the sampled metastatic sites in COSMIC,[21] plus the cisplatin signature, revealed a contribution of the cisplatin SNV signature to unique mutations in the lung primary tumour sample as well as in the bone and lymph node metastases (Fig. 4*a*), but not to mutations detected in TCGA-derived whole genome LUAD or LUSC sequences of either non-smoker or smoker patients who did not receive prior cisplatin chemotherapy (Fig. 4*b*), suggesting that we specifically detected cisplatin-induced SNVs in the samples from the treated patient.

The DNV spectra of the TCGA samples were also markedly different from those of our sequenced samples, with the samples from smokers dominated by CC > AA mutations (Figs. 4*c* and 4*d* and Supporting Information Fig. S7). Also, the DNV spectrum of our four sequenced samples was significantly more similar to the cell line-derived cisplatin spectrum than that from the TCGA samples from either lifelong non-smokers or current smokers (Fig. 4*e*; $p = 1.963 \times 10^{-4}$ and $p = 2.416 \times 10^{-7}$, respectively, *t*-test). Our results conclusively prove that even a single cisplatin treatment induces detectable SNV and DNV mutagenesis in lung cancer genomes.

Ongoing structural rearrangements affect the EGFR gene

A further genomic consequence of treatment was the emergence of the *EGFR* T790M mutation responsible for gefitinib resistance. Irreversible inhibition of T790M-EGFR can induce the amplification of the mutation containing allele,[33] and afatinib resistance has been shown to be accompanied by elevated EGFR expression in cell lines.[34] In order to investigate whether the afatinib treatment induced variations in the *EGFR* gene, we performed a copy number analysis. The sequence coverage at the *EGFR* locus and the whole chromosome 7 revealed that the *EGFR* gene also underwent amplification

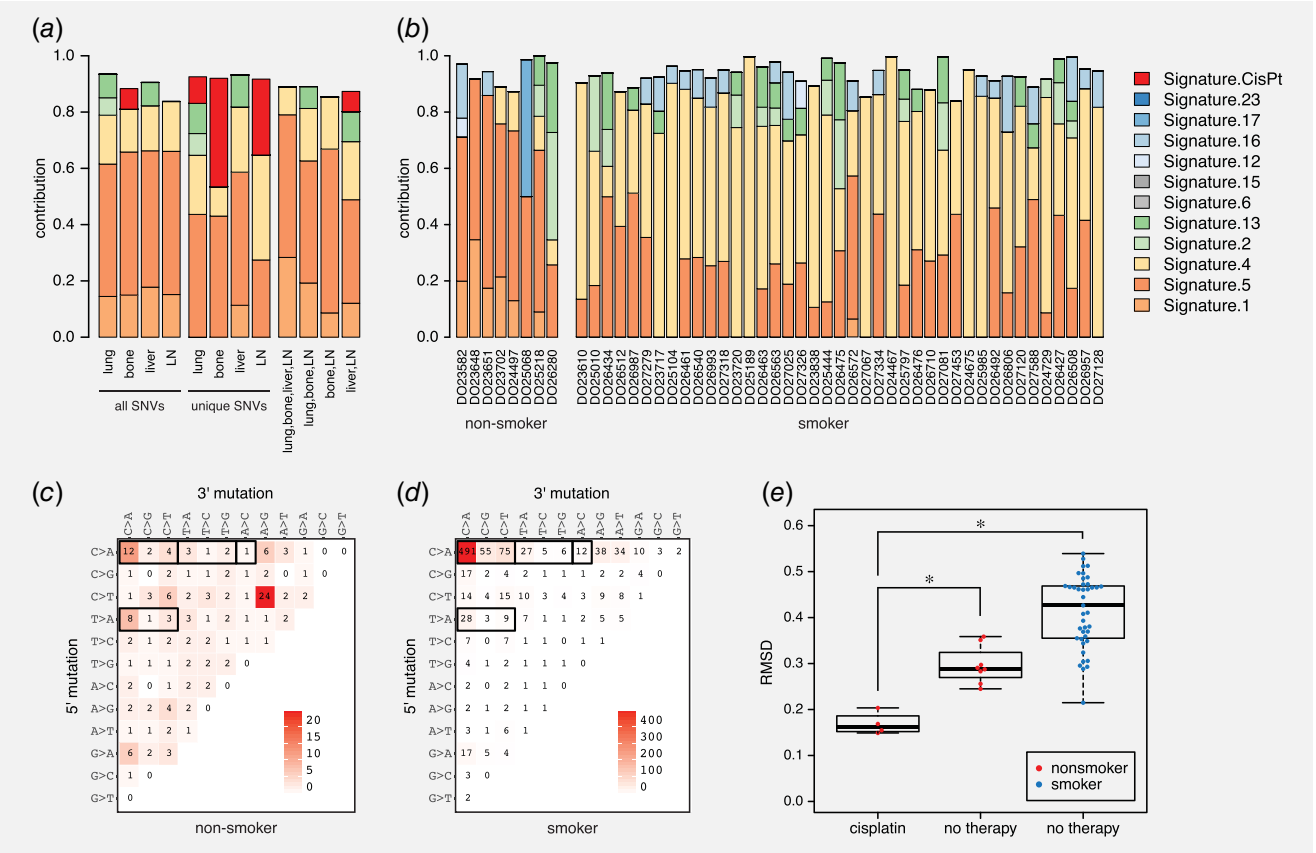

Figure 4. The mutagenic effect of cisplatin treatment. (*a*) Decomposition of SNV spectra to a selection of COSMIC signatures supplemented with the cisplatin spectrum as defined in Boot *et al.*[10] (*b*) A similar SNV spectrum decomposition on TCGA-derived WGS LUAD and LUSC samples of never-smoker and smoker patients, not treated with cisplatin. (*c*, *d*) DNV sequence categories in case of never-smoker and smoker patients, with black boxes indicating the cisplatin type mutations. (*e*) RMSD of the DNV spectra from the cell line-derived DNV-spectrum. The four samples from the cisplatin-treated patient are compared to smoker and non-smoker patient data from TCGA. [Color figure can be viewed at wileyonlinelibrary.com]

(Figs. 5*a* and Supporting Information Fig. S8), a phenomenon that is commonly observed in *EGFR*-mutated LUAD.

To estimate the timing of the amplification events, we determined the copy number of each *EGFR* allele using AF information and tumour content data precisely calculated from germline SNP AFs at the *TP53* locus that underwent loss of heterozygosity (Figs. 3*a* and Supporting Information Fig. S9) and assuming that the T790M mutation occurred on the e19del EGFR bearing allele.[35] Each sample was different: in addition to one normal allele, the primary tumour contained one e19del allele and ten additional e19del+T790M alleles, while the metastases contained two e19del alleles and one or two e19del+T790M alleles (Fig. 5*b*). In the lymph node sample, the calculation showed very close to 1.5 copies of the e19del+T790M allele; we suspect this is due to a mixture of two cell populations in this sample (see below). The last common ancestor of the primary tumour and the three sampled metastases had at most three copies of *EGFR*: a wild type, an e19del and an e19del+T790M allele, though our results cannot exclude the possibility that the T790M mutation arose several times independently. These data show that either the deletion-activated or the TKI-resistant *EGFR* allele may undergo amplification, and that such amplification events continue after metastasis formation.

Searching for a mechanism for *EGFR* amplification, we looked for genome-wide patterns of structural variations (SVs). Uneven sequence coverage across the whole genome suggested the presence of many rearrangements (Supporting Information Fig. S10), and we detected numerous translocation breakpoints using the CREST algorithm. Interestingly, most translocations were unique to the lung and liver samples, though a set of the liver translocations were also present in the lymph node sample at low AF (Figs. 5*c* and 5*e*). There was an especially high number of low AF subclonal breakpoints in the lung sample (Supporting Information Fig. S11). We selected two breakpoints downstream of the *EGFR* gene (Fig. 5*a*), and confirmed by PCR across the breakpoints that these two translocations between chromosomes 2 and 7 were not present in the bronchoscopy biopsy sample taken at the time of the emergence of the T790M mutation (Figs. 1*a* and 5*d*). All these observations

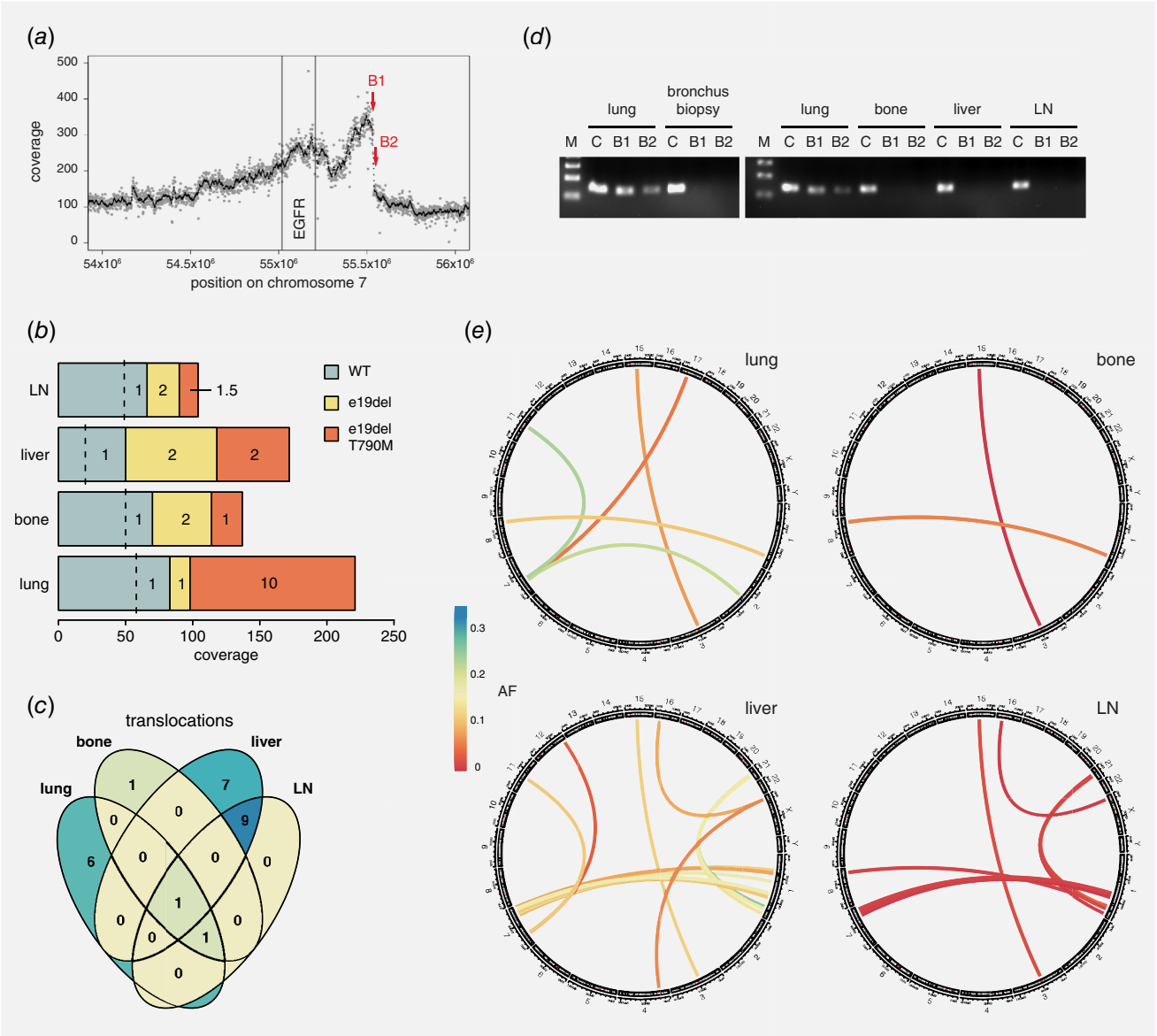

Figure 5. Structural variations identified with the CREST algorithm. (*a*) Amplification of the EGFR gene demonstrated by the increased coverage of the aligned sequence. B1 and B2 indicate two mapped breakpoints. (*b*) Copy numbers of the WT, exon19del and T790M containing EGFR alleles, estimated based on the calculated tumour content and the coverage at the EGFR positions, as well as the allele frequencies of heterozygous polymorphisms, the coverage of the T790 position and allele frequency of the T790M mutation. (*c*) Venn diagram of identified chromosomal translocations. (*d*) PCR from the autopsy samples and the bronchus biopsy sample confirms the late formation of the B1 and B2 breakpoints. (*e*) Chromosomal translocations shown on genomic chord diagrams. The translocations are coloured based on their allele frequencies. [Color figure can be viewed at wileyonlinelibrary.com]

point to ongoing large-scale genome instability late in the history of the tumour, which may provide the selectable events leading to *EGFR* amplification.

A phylogenetic model for tumour growth and spread

The distribution of somatic mutations amongst the samples allows the construction of simple tumour phylogeny.[36] Our aim was to use the distribution and AF of treatment-induced mutations to augment a phylogenetic tree with timing information. The Venn diagram of high AF somatic mutations with the strictest filter suggests a simple phylogeny whereby the bone metastasis is an almost direct descendant of the lymph node metastasis, while the liver metastasis evolved separately (Supporting Information Fig. S2E). However, the inclusion of low AF mutations in the final filtering revealed two exclusive groups of SNVs in the lymph node that are shared with either the bone or the liver metastasis. Both of these groups are subclonal, with a mean AF of 0.11 and 0.06, respectively, while the clonal common mutations have a mean AF of 0.14 (Supporting Information Fig. S5), in agreement

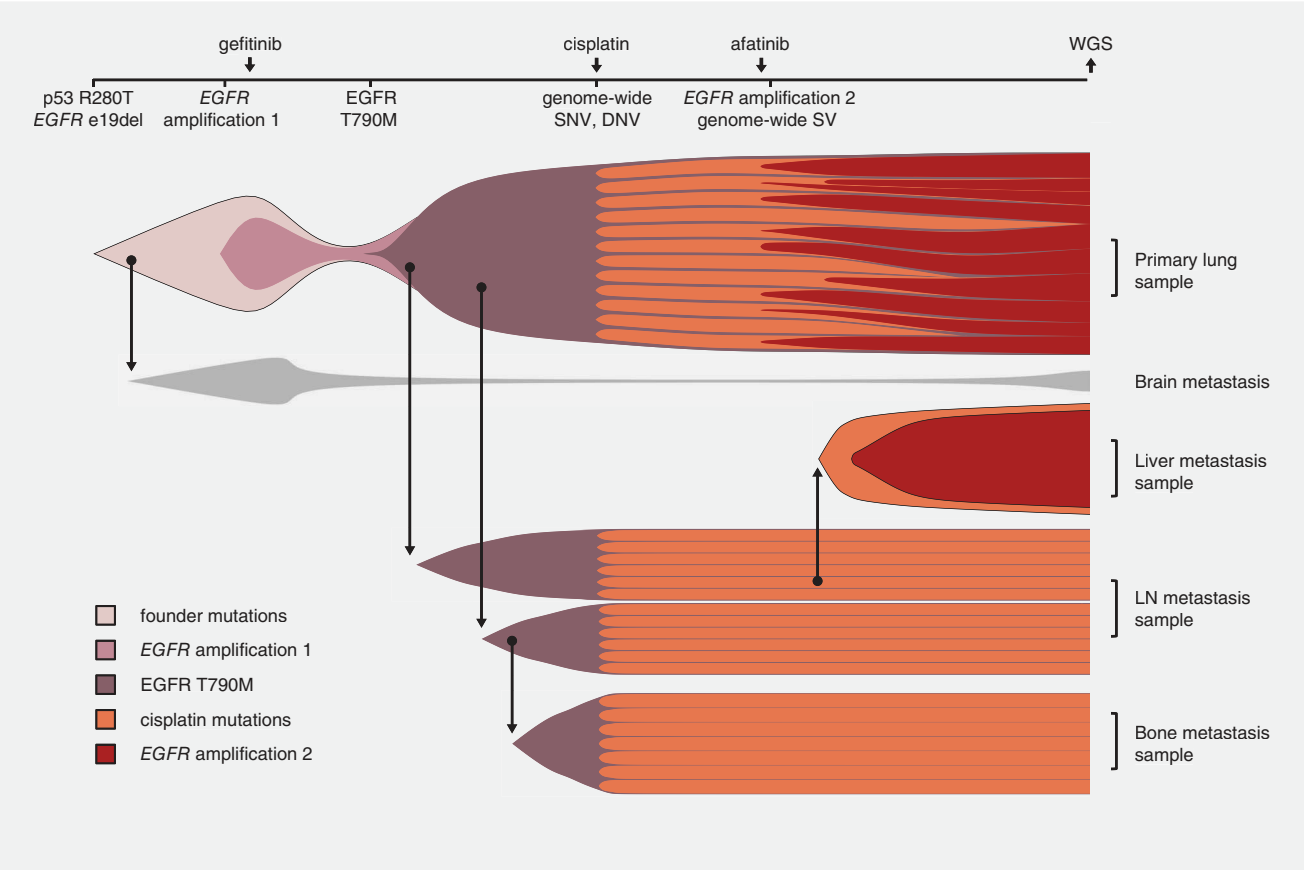

Figure 6. A qualitative model showing the relationship between the primary tumour and metastases. Cisplatin-induced mutant subclones are depicted. [Color figure can be viewed at wileyonlinelibrary.com]

with the calculated tumour content of the lymph node sample (26%). This suggests that the lymph node sample contained two distinct groups of tumour cells related to either the bone or the liver metastasis, implying that the seeding route for both these distant metastases led through the same parathracheal lymph node (Fig. 6). Thus the improved filtering of low AF mutations allows the refinement of tumour phylogeny.

The AF distribution of DNVs, which are mostly cisplatin-induced, closely matches that of common SNVs in the liver metastasis, thus they must be clonal. We can conclude that the ancestor of the liver metastasis sample was a single cell at the time of the cisplatin treatment, therefore this metastasis likely arose after the treatment. The contribution of the cisplatin SNV signature to mutations shared by the liver and lymph node but not to unique liver mutations also confirms the clonality of these mutations in the liver sample, and places the origin of the liver metastasis in the lymph node. In contrast, cisplatin-induced DNVs are subclonal in the bone and lymph node metastases, which must therefore have existed before the treatment despite being undetectable by clinical imaging. With the caveat that a sampled site cannot fully reflect the clonality of the full metastatic site, these data allow the drawing of a timed phylogenetic tree (Fig. 6); also showing early and late *EGFR* amplification events gleaned from the study of rearrangements.

Discussion

In this study, we were able to directly demonstrate and measure the mutagenic effect of cisplatin treatment on lung adenocarcinoma, and use this information together with a detailed analysis of genome sequences of primary and metastatic tumour sites to build a timed model of tumour development.

The mutagenicity of cytotoxic therapy is a concern due to the potential of induced mutations to accelerate the evolution of resistance in cancer cells, and to trigger carcinogenic changes in normal tissue. In a DT40 cell culture model we previously estimated the mutagenicity of cisplatin treatment at IC_{50} concentration as 200 base substitutions per gigabase per treatment cycle,[9] and not entirely comparable weekly treatments of MCF10A and HepG2 cells caused 150–450 base substitutions per gigabase per treatment.[10] To find the mutagenicity of cisplatin per tumour cell, we must consider only clonal mutations in the sequenced samples. Eight percent of the 476 SNVs common to the liver and lymph node samples belong to the cisplatin signature, and these mutations are

clonal in the liver metastasis, suggesting that at least 13 SNVs per gigabase were induced by cisplatin in its single ancestral cell. An alternative estimate can be obtained from the number of DNVs, which were present at about 7.5% of the number of SNVs in both of the above studies. In the liver sample there were 17 clonal DNVs attributable to the cisplatin effect, leading to an estimate of 75 SNVs per gigabase. These estimates are lower than the cell line-based observations, suggesting more limited access of the drug to tumour cells. LUAD genomes from smokers or never-smokers contain a mean of 12,100 and 2,600 mutations per gigabase respectively,[37] but these mutations accumulated over decades, therefore repeated cycles of cisplatin treatment probably significantly increase the mutation rate and contribute to the evolution of resistance.

The dominance of cisplatin-induced mutations amongst all DNVs in cell lines and tumours from a never-smoker suggests that a DNV spectrum can be used for more sensitive detection of cisplatin-induced mutagenesis than an SNV spectrum. Smoking also induces a high number of mainly CC > AA DNVs, but we were able to distinguish the smoking DNV spectrum from the cisplatin-induced DNVs. Finding clonal cisplatin-induced DNVs in a metastatic tumour sample is therefore a simple method for ascertaining that the given metastasis arose after cisplatin treatment of the patient.

In the investigated LUAD case the liver metastasis appeared late, and the data also proves that it was seeded from a single cell or cell clone. In contrast, the subclonality of cisplatin mutations in the bone and lymph node metastases suggests that the metastatic process that initiated their formation had begun prior to the treatment, even though the sampled metastases were not observed at that time. Subclonality of somatic mutations in metastases can also suggest polyclonal seeding.[38,39] We observed evidence for this in the lymph node metastasis, similar to recently reported data from colon cancer.[40] The clonal state of the same mutations at two other sites suggests that this subclonality was due to the anatomical location and seeding routes of the respective metastases rather than a cooperative interaction of the two cell populations.[41] The shared seeding routes through the same proximal lymph node metastasis to distant metastases may be a trait of oligometastatic LUAD.

EGFR amplification is associated with more aggressive tumours, and often occurs at a late stage of the disease.[42] The presence of extreme *EGFR* amplification only in the primary site suggests that it was indeed a late event in this case, also supported by the subclonality of the amplification breakpoints upstream of the gene. The therapy was switched from gefitinib to afatinib following the seeding of metastases, raising the possibility that the amplification was a response to increase the afatinib resistance of the tumour that already contained the *EGFR* T790M mutation. The amplification of the *EGFR* gene occurred *via* chromosomal translocations, and other late stage structural variations were also identified in the primary tumour and the liver metastasis. A number of the liver-specific clonal chromosome translocations were found as subclonal events in the lymph node, supporting the conclusion that the liver metastasis was seeded at the last stages of the disease, after the switch to afatinib.

The detection of clonal tumour evolution from high-coverage WGS samples has been described,[43] and various strategies have been reported for analysing clonal evolution from WGS data.[44,45] The analysis of clonal composition of tumours has showed that subclones can contribute to metastatic seeding and establish resistance to treatments.[46,47] Evolutionary studies of a glioblastoma case have demonstrated the importance of understanding of the subclonal events for personalised therapy.[48] Treatments can also induce subclonal events, but few WGS studies have addressed this question. The effect of aromatase inhibitors on clonal architecture was shown by WGS analysis of matched tumour-normal pairs before and after neoadjuvant therapy in oestrogen-receptor-positive breast cancers,[49] and resistance and clonal advantage after therapy was identified in circulating tumour DNA in chronic lymphocytic leukaemia.[50] Here we showed for the first time that the analysis of subclonal events in matched metastatic tumour samples can also be used for a retrospective analysis of the timing metastatic events, even using standard-coverage WGS data, contributing to our understanding of the evolution of oligometastatic disease, as well as of the genomic imprint of chemotherapeutic treatments.

In conclusion, we have demonstrated the use of genomics to map the progress of metastatic cancer, making use in particular of information derived from the mutagenic effect of cisplatin therapy. As in related studies, the conclusions on clonality are limited by the size of the tissue sample in relation to the tumour site. Nevertheless, we used the obtained information to show that early and late metastases may seed through the same proximal lymph node, and that independent late *EGFR* amplification events at different sites contribute to the ongoing evolution of the lung adenocarcinoma. Tumour type specific studies on the timing of metastasis formation relative to established treatment practices will be valuable for shaping future therapy regimens.

Acknowledgements

The authors wish to thank Gyula Ostoros for sharing clinical data. The results shown here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. This work was funded by The Danish Cancer Society (grant R90-A6213 to MK), the Research and Technology Innovation Fund of Hungary (FIEK_16-1-2016-0005 to DS and IC, KTIA_NAP_13-2014-0021 and NAP2-2017-1.2.1-NKP-0002 to ZS); Breast Cancer Research Foundation (BCRF-17-156 to ZS) and the Novo Nordisk Foundation Interdisciplinary Synergy Programme Grant (NNF15OC0016584 to ZS and IC).

References

1. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nat Rev Clin Oncol* 2015;12:258–72.
2. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* 2016;352:169–75.
3. Zhang J, Fujimoto J, Wedge DC, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256–9.
4. Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* 2017;376:2109–21.
5. Awad MM, Katayama R, McTigue M, et al. Acquired resistance to crizotinib from a mutation in CD74-ROS1. *N Engl J Med* 2013;368:2395–401.
6. Niederst MJ, Sequist LV, Poirier JT, et al. RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nat Commun* 2015;6:6377.
7. Yates LR, Gerstung M, Knappskog S, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 2015;21:751–9.
8. Cooper CS, Eeles R, Wedge DC, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet* 2015;47:367–72.
9. Szikriszt B, Poti A, Pipek O, et al. A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol* 2016;17:99.
10. Boot A, Huang MN, Ng AWT, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res* 2018;28:654–65.
11. Holohan C, Van Schaeybroeck S, Longley DB, et al. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer* 2013;13:714–26.
12. Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat Rev Clin Oncol* 2014;11:473–81.
13. Blakely CM, Watkins TBK, Wu W, et al. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. *Nat Genet* 2017;49:1693–704.
14. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009;25:1754–60.
15. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
16. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–9.
17. Costello M, Fleharty M, Abreu J, et al. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics* 2018;19:332.
18. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet* 2017;100:267–80.
19. Wang J, Mullighan CG, Easton J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods* 2011;8:652–4.
20. Gu Z, Gu L, Eils R, et al. Circlize implements and enhances circular visualization in R. *Bioinformatics* 2014;30:2811–2.
21. COSMIC. COSMIC: Signatures of mutational processes in human cancer. Available from: http://cancer.sanger.ac.uk/cosmic/signatures. Accessed 20 August 2018.
22. Rosenthal R, McGranahan N, Herrero J, et al. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016;17:31.
23. Miller CA, McMichael J, Dang HX, et al. Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* 2016;17:880.
24. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017;23:703–13.
25. Zhang K, Dion N, Fuchs B, et al. The human homolog of yeast SEP1 is a novel candidate tumor suppressor gene in osteogenic sarcoma. *Gene* 2002;298:121–7.
26. Plaza Menacho I, Koster R, van der Sloot AM, et al. RET-familial medullary thyroid carcinoma mutants Y791F and S891A activate a Src/JAK/STAT3 pathway, independent of glial cell line-derived neurotrophic factor. *Cancer Res* 2005;65:1729–37.
27. Drilon A, Hu ZI, Lai GGY, et al. Targeting RET-driven cancers: lessons from evolving preclinical and clinical landscapes. *Nat Rev Clin Oncol* 2018;15:151–67.
28. Toledo RA, Hatakana R, Lourenco DM Jr, et al. Comprehensive assessment of the disputed RET Y791F variant shows no association with medullary thyroid carcinoma susceptibility. *Endocr Relat Cancer* 2015;22:65–76.
29. Dong YU, Ren W, Qi J, et al. EGFR, ALK, RET, KRAS and BRAF alterations in never-smokers with non-small cell lung cancer. *Oncol Lett* 2016;11:2371–8.
30. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 2014;15:585–98.
31. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 2013;3:246–59.
32. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
33. Ercan D, Zejnullahu K, Yonesaka K, et al. Amplification of EGFR T790M causes resistance to an irreversible EGFR inhibitor. *Oncogene* 2010;29:2346–56.
34. Yamaoka T, Ohmori T, Ohba M, et al. Distinct Afatinib resistance mechanisms identified in lung adenocarcinoma harboring an EGFR mutation. *Mol Cancer Res* 2017;15:915–28.
35. Hidaka N, Iwama E, Kubo N, et al. Most T790M mutations are present on the same EGFR allele as activating mutations in patients with non-small cell lung cancer. *Lung Cancer* 2017;108:75–82.
36. Schwartz R, Schaffer AA. The evolution of tumour phylogenetics: principles and practice. *Nat Rev Genet* 2017;18:213–29.
37. Alexandrov LB, Ju YS, Haase K, et al. Mutational signatures associated with tobacco smoking in human cancer. *Science* 2016;354:618–22.
38. Gundem G, Van Loo P, Kremeyer B, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature* 2015;520:353–7.
39. Hoadley KA, Siegel MB, Kanchi KL, et al. Tumor evolution in two patients with basal-like breast cancer: a retrospective genomics study of multiple metastases. *PLoS Med* 2016;13:e1002174.
40. Ulintz PJ, Greenson JK, Wu R, et al. Lymph node metastases in colon cancer are polyclonal. *Clin Cancer Res* 2018;24:2214–24.
41. Cheung KJ, Padmanaban V, Silvestri V, et al. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc Natl Acad Sci U S A* 2016;113:E854–63.
42. Yatabe Y, Takahashi T, Mitsudomi T. Epidermal growth factor receptor gene amplification is acquired in association with tumor progression of EGFR-mutated lung cancer. *Cancer Res* 2008;68:2106–11.
43. Griffith M, Miller CA, Griffith OL, et al. Optimizing cancer genome sequencing and analysis. *Cell Syst* 2015;1:210–23.
44. Miller CA, White BS, Dees ND, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol* 2014;10:e1003665.
45. Deshwar AG, Vembu S, Yung CK, et al. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol* 2015;16:35.
46. Patch AM, Christie EL, Etemadmoghadam D, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 2015;521:489–94.
47. Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol* 2014;8:1095–111.
48. Favero F, McGranahan N, Salm M, et al. Glioblastoma adaptation traced through decline of an IDH1 clonal driver and macro-evolution of a double-minute chromosome. *Ann Oncol* 2015;26:880–7.
49. Miller CA, Gindin Y, Lu C, et al. Aromatase inhibition remodels the clonal architecture of estrogen-receptor-positive breast cancers. *Nat Commun* 2016;7:12498.
50. Yeh P, Hunter T, Sinha D, et al. Circulating tumour DNA reflects treatment response and clonal evolution in chronic lymphocytic leukaemia. *Nat Commun* 2017;8:14756.