Treatment represents a key driver of metastatic cancer evolution

Christensen, D.S.[1,2,3*], Ahrenfeldt, J.[1,2*]., Sokač, M.[1,2], Kisistók, J.[1,2], Thomsen, M[4]., Maretty, L.[1,2,7], McGranahan, N.[5,6$] , Birkbak, N. J.[1,2,7$]

[1]Department of Molecular Medicine, Aarhus University Hospital, Denmark
[2]Department of Clinical Medicine, Aarhus University, Denmark
[3]Department of Clinical Oncology, Aarhus University Hospital, Denmark
[4]Department of Biomedicine, Aarhus University, Denmark
[5]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, Paul O'Gorman Building, London, UK
[6]Cancer Genome Evolution Research Group, University College London Cancer Institute, University College London, London, UK
[7]Bioinformatics Research Center, Aarhus University, Aarhus, Denmark
*These authors contributed equally
$*Correspondence: nicolas.mcgranahan.10@ucl.ac.uk (N.M.), nbirkbak@clin.au.dk (N.J.B.)

Conflict of interest

N.M. has received consultancy fees and has stock options in Achilles Therapeutics. N.M.. holds European patents relating to targeting neoantigens (PCT/EP2016/ 059401), identifying patient response to immune checkpoint blockade (PCT/ EP2016/071471), determining HLA LOH (PCT/GB2018/052004), predicting survival rates of patients with cancer (PCT/GB2020/050221). The other authors declare no potential conflicts of interest.

Abstract

Metastasis is the main cause of cancer death, yet to this day the evolutionary processes behind remain largely unknown. Here, through analysis of large panel-based genomic datasets from the AACR GENIE project including 40,979 primary and metastatic tumors across 25 distinct cancer types, we explore how the evolutionary pressure of cancer metastasis shapes the selection of genomic drivers of cancer. The most commonly affected genes were *TP53, MYC and CDKN2*A with no specific pattern associated with metastatic disease suggesting that on a driver mutation level the selective pressure operating in primary and metastatic tumors is similar. The most highly enriched individual driver mutations in metastatic tumors were known resistance mutations, to hormone therapies in breast and prostate cancer affecting *ESR1* and *AR*, to anti-EGFR therapy in non-small cell lung cancer (EGFR T790M) and to imatinib in gastrointestinal cancer (KIT V654A). We also observed specific mutational signatures associated with treatment in three cancer types, supporting clonal selection following anti-cancer therapy. Overall, this implies that acquisition of driver mutations are predominantly shaped by the tissue of origin where specific mutations define the developing primary tumor and drives growth, immune escape and tolerance to chromosomal instability.

Significance

The genomic drivers of metastatic cancer remain largely unknown. We show how the driver landscape mirrors primary disease, with the main genomic drivers of metastatic cancer evolution associating with resistance to therapy.

Introduction

Metastasis is the process where cancer cells from a primary site colonize to distant organs(1), and is usually considered the terminal step in the evolution of lethal cancer. Given that a significant proportion of primary cancers are cured entirely through surgery, metastatic dissemination must be a relatively late event in many if not most cases. It has been hypothesized that the ability to metastasise is not inherent to primary cancers, but must be acquired during cancer evolution(2). As more than 90 percent of all cancer-related deaths are caused by metastatic cancer(3), understanding the process of how the primary tumor achieves metastatic potential is of critical importance if cancer survival is to be substantially improved.

Cancer is a molecular disease driven by accumulation of somatic alterations to the genome. Most of our current understanding of cancer is derived from studies investigating primary tumors, while there has been considerably less analysis of metastatic tumors. The metastatic process is a multi-step process and consists of local invasion, intravasation, survival in the circulation, extravasation and colonization to distant tissues(1,2). Studies in mice have suggested that the metastatic process is generally highly inefficient, with the vast majority of cancer cells dying in circulation(4). The established association between tumor size and risk of metastasis suggests that the metastatic process may be a stochastic combination of cell proliferation and tumor size, with larger and/or more proliferative tumors shedding more cells into circulation, ultimately increasing the chance of successful colonization of distant metastasis. However, phylogenetic analysis of metastatic lesions across multiple cancer types have recently shown that a monophyletic relationship commonly exists between metastatic lesions(5). While this observation may be limited by the number of metastatic tumors sampled, this might suggest that some aggressive phenotypic traits either dramatically increase the chance of metastatic dissemination, or are potentially acting as gate-keeper events required for successful dissemination.

Mutations that increase cancer cell fitness are referred to as driver mutations. In the primary tumor, driver mutations are selected for as they generally lead to improved cancer cell fitness through either increased cell proliferation or decreased cell death(6). While selection in the primary setting is independent of metastasis, it is plausible that certain traits that increase fitness in the primary tumor also promote survival in circulation and the ability to colonize distant tissue. These would in effect act as gate-keeper events, where the risk of metastasis is vastly increased following acquisition. While the existence of such gatekeeper mutations or alterations have long been hypothesized(2), so far, they have eluded detection. Several recent studies have analyzed genomic data from metastatic tumors in smaller cohorts. Robinson et al analyzed data from more than 500 metastatic cancer samples from 20 different cancer types and found an increased mutation burden relative to primary samples using The Cancer Genome Atlas (TCGA)(7), and increased global dysregulation of gene transcription. However, on both gene and pathway level, they were unable to find any defining characteristic that facilitated development of metastatic potential. Similar results indicating limited recurrent evolution of the cancer genome in the metastatic setting was reported by the Hartwig Medical Foundation (HMF), where the authors demonstrated a conspicuous lack of metastasis-specific driver mutations in a cohort of 2,520 metastatic patients analyzed by whole genome sequencing(8). A more recent follow-up study by the HMF analyzed paired samples, either primary-metastatic or metastatic-metastatic, across 250 individuals analyzed with whole genome sequencing(9). Here, de Haar and colleagues found that when focusing on clinically relevant genomic biomarkers, full concordance was observed for 99% of patients between paired biopsies. Thus, to this day, no evidence has been found supporting a clear genomic basis for metastatic potential, and the existence of gatekeeper events for metastatic dissemination remains a hypothesis.

With this work we aim to leverage the power of large datasets to investigate if cancer gatekeeper mutations exist by comparing genomic panel data from primary and metastatic tumor samples from the GENIE project. We define a common gene set of 174 established cancer genes with sequence data from 40,979 samples. With these we investigate whether metastatic tumors are preferentially enriched in driver mutations or copy number alterations in cancer genes. Through gene and pathway-based analysis we decipher whether certain alteration patterns may be necessary for achieving metastatic potential. While this gene set only represents a small fraction of the genome, together these genes represent the targets of more than half of all cancer driver mutations reported in the to-date largest analysis of metastatic cancers using whole genome sequencing, performed by the HMF(8). Thus, with this geneset we are able to investigate shifts to the evolutionary landscape within established cancer genes, but as a limitation of the panel-based approach we are unable to discern metastasis-driven evolution to other parts of the genome.

Methods

Data acquisition

Basic patient information along with mutation and copy number data 112,935 sequenced tumor samples was acquired from the AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE) consortium(10), version 9.0. The data includes primary or metastatic tumor genomic data from 104,125 of cancer patients treated at 18 institutions worldwide. Genomic data was based on 92 different gene panels, containing between 11 and 760 genes. The public GENIE repository includes restricted clinical annotations, limited to primary/metastatic status, gender, ethnicity, age, and cancer type.

Determining an optimal geneset

To identify the largest number of genes assessed across the largest number of patients, we defined an optimal gene-set based on genes shared by the most panels assessed in the largest number of tumors. First cancer types with less than 100 samples in both primary and metastatic cancer were excluded. All panels were sorted by number of included samples, the largest was compared to all other panels to identify the panel with the largest gene overlap, this panel was kept, and the overlapping genes were compared to the remaining panels. This was repeated until all panels had been ranked. The final gene set was chosen by weighing the number of genes versus the number of patients and included 174 genes assayed by 29 different gene panels in 40,979 tumors across 25 different cancer types (Figure S1, Table S1).

Summary genomic scores

The somatic tumor mutation burden (TMB) was defined as the number of mutations per megabase, for each sample calculated based on the full panel size. The weighted genome integrity index (wGII) was calculated from the available segmented copy number data, as previously described(11).

Annotation of driver events

All somatic mutations in both TCGA and Genie were annotated to genes by ANNOVAR(12) using the hg19 reference genome. Driver mutations were defined for frameshifts as Frameshift indels in tumor suppressor genes (TSG), and as Non-frameshift indels with an occurrence in the COSMIC v90(13) database of at least 3 in oncogenes. SNVs were defined as drivers for TSGs if either predicted "deleterious" by SIFT, "probably damaging" by PolyPhen or if it was a stop gain or splice mutation. For oncogenes, specific mutations had to occur at least 3 times on COSMIC. Finally, any specific mutation with a Cosmic count > 10 was included in the definition of driver mutations. Data on somatic copy number alterations (SCNA) as defined by the GENIE pipeline were available for a subset of the samples (Table S2). Here, the GENIE data set was annotated as either 2, 1, -1 or - 2, specifying deletions, losses, gains and amplifications. Genes annotated as deletions or amplifications were classified as driver events.

Enriched and depleted genes and pathways

For the enrichment analyses genes were considered altered if they harbored either a somatic mutation considered a likely driver, or a copy number change called as a deletion or an amplification as defined by the GENIE processing pipeline. A two-sided Fisher's exact test was used to compare primary to metastatic disease, on the number of patients with and without altered genes, per cancer type. P-values were corrected by false discovery rate (FDR) and considered significant if the corrected p-values were below 0.05. A similar enrichment analysis was performed on a pathway level, where all gene driver events were mapped to the cancer specific pathways from Sanchez-Vega et al.(14) before further analysis on pathway level, including only pathways where at least 50% of the genes were found in the GENIE gene set (Table S3). Enrichment or depletion of co-occurring alteration in pathways was tested using a two-sided Fisher's exact test to compare primary to metastatic disease, on the number of patients with and without altered pathway pairs, per cancer type. P-values were corrected by false discovery rate (FDR) and considered significant if the corrected p-values were below 0.25.

Mutational signature analysis

All mutations were annotated relative to their trinucleotide context, as described(15). For each cancer type, within primary and metastatic tumors separately, the frequencies of 96 potential trinucleotide mutations were determined, and COSMIC mutational signatures v3.0 were inferred using the deconstructSigs R package(15,16), version 1.9.0. To determine the number of mutations assigned to each signature, we multiplied the signature proportion with the average tumor mutation burden of the cancer type, primary and metastatic separately.

Statistical analysis

All analysis was performed in R version 3.6.2 (17), using Tidyverse (18) and ggpubr(19), scales(20), ggrepel(21) for visualizations. For significance testing Wilcoxon test was used unless otherwise mentioned.

Results

Samples and genes

The public GENIE version 9 dataset contains a total of 112,935 tumors from 104,125 patients, profiled using 92 different gene panels, containing 11-760 genes. The clinical annotations are limited to basic information, such as age, gender, cancer type and primary or metastatic biopsy. While based on gene panels and thus restricted in the genomic content covered, this very large cohort provides a unique opportunity to study the evolution of primary to metastatic cancer in a selected set of well-defined cancer genes. To perform this analysis, we defined a common sequenced gene set that contained the maximum number of tumors covered by the highest number of genes (see methods, Figure S1). The final gene set included 174 established cancer genes which were assayed in a total of 40,979 tumors, 24,333 primary, 16,546 metastatic, across 25 cancer types (Figure 1A-B). While the common gene set is relatively small compared to whole exome or whole genome sequence data, these 174 cancer genes are commonly mutated in cancer, with 145 listed in the COSMIC cancer gene census. Indeed, in 2,520 metastatic tumors analyzed by the Hartwig Medical Foundation(8), 20,070 driver mutations were reported. Of these, 10,889 (54.3%) were found within genes covered by the GENIE common gene set (Figure 1C), making this gene set a valid reference point for analysis based on driver alterations (somatic driver mutations and copy number events).

Metastatic cancer harbors more somatic mutations and increased levels of genomic instability

To investigate the summary genomic differences between metastatic and primary cancer, we determined the tumor somatic mutation burden (TMB), the mutation variant allele frequency

(VAF), and the overall level of chromosomal alterations defined using the weighted genome integrity index (wGII)(11). We found that TMB trended higher in metastatic samples compared to primary samples in 17 of 25 cancer types, being significantly higher in 10 (Figure 2A). In only one cancer type did we find a significantly higher TMB in primary tumors (bladder cancer). VAF, a proxy measure of tumor purity and clonality, was generally found to be higher in metastatic samples, being significantly higher in 10/25 (Figure S2), but significantly lower in five cancer types (uterine sarcoma, glioma, bone, bladder and renal cancer). Similarly, we observed that wGII trended higher in metastatic tumors in 20 of 25 cancer types, being significantly higher in 13 (Figure 2B). In none of the cancer types did we observe higher wGII in primary relative to metastatic tumors. When comparing TMB to wGII directly, as previously reported(22) we found an inverse association, with tumors harboring high levels of mutations showing lower levels of chromosomal alterations (Figure S3, A-B). To investigate if the observed increase in wGII in metastatic relative to primary samples only applied to samples with low TMB, we divided samples into bins based on their TMB (Figure S3C). We observed a significantly higher level of wGII in metastatic samples across all bins, indicating that the increased levels of chromosomal alterations found in the metastatic setting is independent of the overall mutation burden (Figure S3D).

Metastatic tumors harbor fewer drivers per mutation

Across the full GENIE cohort, we found 138,253 driver mutations and 193,708 driver copy number alterations, with an average of 3 driver events observed per sample (range 0-152). To investigate if cancer driver mutations might be more prevalent in metastatic samples, we compared the total number in each sample between primary and metastatic tumors. Based on the gene panels in this study, we found a higher overall number of driver mutations per sample in metastatic versus primary tumors, with 8/25 cancer types showing slight but significant increase in driver mutation count in metastatic samples, and 4/25 cancer types showing a slight but significant decrease (Figure 2C), overall this indicates that metastatic biology is not driven solely by the sheer number of established cancer drivers. To determine whether an increase in driver mutations might be driven by a general increase in mutation burden, we calculated the ratio of driver mutations to TMB. Intriguingly, in 12 of 25 cancer types we found a significantly lower number of drivers when corrected for the total mutation burden. In none of the cancertypes did we observe the inverse (Figure 2D). This suggests that for the core set of cancer driver genes analyzed in this work, selection for specific drivers is lower in the metastatic setting relative to the rate of mutation accumulation. To further investigate this, we determined the percentage of samples with cancer driver events (defined as either individual mutations or SCNAs causing gene alterations with a likely role in driving cancer) within metastatic and primary disease, and determined the delta change in the percentage of tumors with a driver mutation in a given gene, when combining all cases across cancer types (Figure S4). We found that overall, most genes showed a delta value of less than 2 when subtracting the percent mutated primary tumors from the percent mutated metastatic tumors. No gene showed a delta value higher than 4, indicating that acquisition of driver mutations is mostly shaped by the tissue of origin.

Metastatic tumors are enriched in resistance mutations

In order to identify specific genes with driver events that are either depleted or enriched in metastasis, we compared the occurrence of driver events between primary and metastatic cancer directly within individual cancer types. First, we calculated the fraction of patients with driver events for each gene. Excluding cancer types with neither altered genes in primary and metastatic samples (Figure 3A). We found overall a correlation between primary and metastatic disease (r = 0.96, P < 0.0001), supporting that most cancer driver mutations are early events, and thus shared by both primary and metastatic disease across all cancer types(23). However, although the observed differences between primary and metastatic disease are small, we do find that certain genes consistently trends towards higher fractions in the metastatic setting. To further investigate this, we determined the relative ratio of affected genes between primary and metastatic disease (Figure 3B, Table S4). This analysis revealed that the most commonly affected driver genes enriched in metastatic cancer across all cancer types were *MYC* and *CDKN2A*, well established cancer genes with a known role in controlling apoptosis and genomic stability, proliferation and cell cycle. Overall, while a total of 68 genes were significantly enriched or depleted in at least one cancer type, most showed limited difference in mutation rate between primary and metastatic tumors (Figure 3C). The genes showing the most dramatic enrichment in the metastatic setting were *ESR1* in breast cancer (2.4% vs 16.7%, primary/metastatic) and *AR* in prostate cancer (2.5% vs 36.1%, primary/metastatic). These alterations were almost certainly selected for through development of acquired resistance to anti-hormone treatment, demonstrating how treatment is a major driver of cancer evolution in the metastatic setting. Notably, we also identified mutations in several genes as significantly depleted in metastatic cancer. These include *ARID1A* (endometrial, colorectal and ovarian) and *PTEN* (endometrial and colorectal), both observed as depleted in more than one cancer type (Figure 3C). *PIK3CA* is particularly of interest as it is one of the most commonly mutated cancer genes(24), and the target of several precision therapies. Interestingly, while found as one of the most commonly enriched genes overall, we also observed *TP53* as significantly depleted in metastatic Head and Neck Cancer (OR = 0.51, P = 0.0068).

Metastatic disease is dominated by p53 and cell cycle pathways

Except for *TP53*, which represents 15.7% of all GENIE driver events, most genes are rarely affected by somatic driver alterations in any individual tumor. Thus, even in a cohort like GENIE with samples reaching more than 40,000, most genes are found affected only a few times and therefore difficult to assess for metastatic relevance. However, as genes act in concert, we investigated if patterns might emerge when individual events are summarized to pathways. For this we relied on the pathway definitions as described by Sanchez-Vega et al(14), including in the analysis all pathways where at least 50% of the assigned genes were found in the GENIE common geneset (Table S3). As on gene-level, we calculated the fraction of patients with altered pathways in both primary and metastatic tumors, excluding cancer types with neither altered pathways in primary and metastatic samples Again, we found overall a correlation between primary and metastatic disease ($r = 0.97$, $P < 0.0001$, Figure S5A), indicating that also on pathway level, most cancer driver events happen in the primary tumor, and are maintained in the metastatic setting. However, while the difference in frequency between primary and metastatic was minor for most cancer types, we did observe either slight increase or slight decrease in the frequency of certain pathways in metastatic versus primary tumors. Again, an enrichment analysis was performed (Figure S5B, Table S5). Here we found that of the 5 pathways assessed, all were significantly enriched in at least one cancer type, with a total of 17 significant hits across all cohorts (Figure S5C). Consistent with literature(25), p53 pathway was one of the most affected pathways across all tumors, affected in approximately 50% of all tumors. Together with the Cell Cycle pathway (affecting 26.9% of all tumors), it was also pathway most enriched in metastatic disease, both enriched in 6/25 different cancer types. The p53 pathway is dominated by mutations in *TP53* itself and the *CDKN2A* gene. While both *TP53* and *CDKN2A* are typically clonally dominant, this suggests that when they do occur subclonally, they are more likely to be present in the metastasising clone. Hits to both p53 and Cell Cycle pathways are well established as strong drivers of cancer through analysis of primary tumors. That we also observe continued selection of these events in the metastatic setting indicates that on a pathway level, the drivers of high stage aggressive disease are also the drivers of metastatic disease.

Recurring mutations in metastatic tumors are associated with treatment resistance

Given the size of the GENIE dataset, we have the power to investigate if specific variants are selected for in metastatic tumors. For this analysis, we included all non-synonymous mutations observed in at least 1% of metastatic tumors within each cancer type, and asked if a given amino acid change was enriched in the metastatic setting. Nine variants were found significantly enriched in the metastatic setting (Figure 4A). Four enriched mutations were in *ESR1*, found in breast cancer, three mutations were in *AR*, found in prostate cancer, both likely representing acquired resistance to anti-hormone therapies. Additionally, the KIT V654A mutation was found significantly enriched in metastatic Gastrointestinal Stromal tumors, previously associated with acquired resistance to imatinib(26), and the EGFR T790M mutation was found significantly enriched in metastatic non-small cell lung cancer, a well documented resistance mutation to anti-EGFR therapies(27) (Figure 4B). Taken together, this analysis of the driver landscape of metastatic cancer shows that acquisition of specific driver mutations mostly reflects tissue of origin where specific events define the developing primary tumor and likely promotes growth and immune escape. It is not evident that any further acquisition of driver mutations are needed to specifically contribute to metastatic dissemination. Rather, while metastatic tumors are slightly enriched in known drivers of aggressive biology such as cell proliferation and chromosomal instability, a key dominant driver of metastatic cancer evolution is anti-cancer therapy. Overall, this supports that in the absence of treatment, the evolutionary pressures that act on established cancer driver genes such as those investigated in this work are indistinguishable in the evolution of both primary and metastatic disease.

Mutational signature analysis reveals a treatment footprint and evidence of APOBEC activity in metastatic tumors

To explore if the processes generating mutations may change during metastatic progression, we performed mutational signature analysis based on the framework published by Alexandrov et al(15,28). We did not have sufficient power to perform this analysis on an individual sample level. Instead, we combined all observed mutations in primary and in metastatic tumors, respectively, within each cancer type. The dominating mutational signatures across the cohort were SBS1 and SBS5, both of which were found in most cancer types and are associated with age and mitosis(29), reflecting that the primary driver of somatic mutations in both primary and metastatic cancer is associated with cell proliferation (Figure S6). In three cancer types, we observed clear evidence of treatment leaving a mutational footprint on the metastatic tumors, indicating selection for drug resistant subclones. In Glioma and Pancreatic cancer, a significant increase in the number and proportion of SBS11 mutations were observed, representing temozolomide treatment or alkylating chemotherapy (Figure 4C). In germ cell tumors, we observed a significant increase in SBS35, representing platinum chemotherapy. Five cancer types showed a significant increase in SBS2 and SBS13 representing APOBEC induced mutations, with cervical, head and neck, non-small cell lung, breast and thyroid cancer showing significant enrichment while a slight and non-significant decrease was observed for bladder cancer (Figure 4D). SBS4 represents mutations induced by tobacco smoke(30) and were found in small cell and non-small cell lung cancer. Small cell lung cancer showed a significant decrease in the number of SBS4 mutations observed in metastatic tissue, while non-small cell lung cancer showed a non-significant decrease (Figure 4E). This is consistent with tobacco-induced mutations mostly playing a role in carcinogenesis and less in metastatic cancer development. Excluding SBS1 and SBS5, no signature was found in more than six cancer types, and these rarer signatures combined showed a significant increase in metastatic tumors (Figure 4F P = 0.0034, paired Wilcoxon test). This supports either a shift in mutation processes in metastatic tumors, or clonal bottlenecking potentially through monoclonal seeding occurring during metastatic dissemination.

Discussion

With this study we demonstrated how large genomic datasets, even with limited genomic coverage and sparse clinical information, can be used to make novel insights into cancer biology and contribute to our understanding of metastatic cancer biology. With more than 40,000 tumors spanning 25 different cancer types, this study is to our knowledge the largest to date that directly compares molecular data from metastatic and primary tumors. Consistent with previous studies in smaller cohorts, we found that metastatic tumors harbor a higher TMB, more driver mutations, and more genomic alterations compared to primary tumors(5,7,8,31,32). While we found similar levels of driver mutations in metastatic and primary tumors, the number of drivers per mutation was higher in primary tumors. This indicates that for the majority of tumors no additional canonical driver mutations are required for metastatic transition and thus not selected for. The increased mutation burden overall may reflect that metastatic dissemination is mostly monoclonal, driven by few specific cancer driver events that are already selected for in the primary tumor. Conversely, the primary tumor is often heterogeneous, composed of multiple independent subclones(33). Here, tissue biopsies may sample more heterogeneous tissue containing multiple subclones, each defined by both clonal and subclonal mutations and driver events(5). Cancer cells acquire mutations both through intrinsic and extrinsic factors, but individual mutations are private to a small number of cells sharing the same lineage. Monoclonal metastatic dissemination would result in a strong bottleneck, where all lineage-specific mutations acquired by the metastasising subclone throughout the life-history of the cancer are suddenly unmasked. This may result in a higher VAF in metastatic tumors due to higher clonality, and an increased observed TMB.

In our work we found that 10/25 cancer types showed significantly higher VAF in metastatic tumors. The observed increase in VAF in the metastatic setting is consistent with previously observed increased clonality of metastatic tumors(34), and may in these cancer types support a predominantly monoclonal model of metastatic dissemination. In metastatic small cell lung cancer, we observed a higher VAF in metastatic settings, no change in TMB, but a significantly lower aging signature (SBS1) supporting early metastatic dissemination, as previously described(35). Interestingly, metastatic small cell lung cancer harbors a lower level of smoking SBS4 mutations, suggesting that while smoking-induced mutations might be causal of the primary tumor, they do not significantly contribute to metastatic tumor development. Thyroid and cervical cancer both showed increased SBS2 and SBS13, indicating late activation of APOBEC enzymes. APOBEC has previously been found to activate late in cancer evolution where it may contribute to cancer development through increased mutation burden(36).

Several cancer types showed a clear footprint of treatment-induced mutations in metastatic tumors. In glioma and pancreatic cancer, we observed a significant increase in SBS11 mutations, induced by alkylating agents such as temozolomide(28,37). In germ cell tumors, a strong increase in SBS35 was observed, representing platinum-induced mutations(28,37). The observed increase in mutation signatures in metastatic tumors is likely indicative of clonal bottlenecking occurring during treatment as the cancer responds to therapy. We also found that mutations in both *ESR1* and *AR* were highly significantly enriched in metastatic breast and prostate cancers, respectively, while essentially absent from primary tumors. These are almost certainly treatment induced, as patients with breast cancer and prostate cancer are routinely treated with adjuvant anti-hormone therapy. We also observed specific enrichment of known resistance mutations KIT V654A and EGFR T790M in gastrointestinal tumors and non-small cell lung cancer, again almost certainly treatment induced. Together these findings demonstrate how treatment alters the evolutionary pressures acting on cancer, induces clonal bottlenecking of the disease, and causes selection for resistance-associated drivers. However, beyond *ESR1, AR,* KIT and *EGFR* we did not find any metastasis-specific driver events. Our work thus supports a model for cancer evolution where in the absence of treatment, the selective pressure towards increased malignancy is similar in primary and metastatic disease, with no specific cancer driver events associated specifically with metastatic dissemination. Consistent with this, most of the genes identified as enriched or depleted in the driver enrichment analysis were already well established cancer genes commonly accepted as drivers of cancer from studies of primary tumors. This suggests that metastasis may not be caused by acquisition of new metastatic features or specific driver events, at least not among the gene set analyzed here. It may be that metastatic potential primarily depends less on proliferation-associated oncogenes, but rather on acquisition of other phenotypes, such as immune-evasion, allowing cancer cells to disseminate more widely and grow without triggering local immune responses(31). This conclusion is supported by several recent publications, including work by de Haar et al(9), where analysis of 250 patients with paired samples using whole genome sequencing demonstrated limited evolution of the actionable cancer genome in the metastatic setting. Likewise, in an analysis of genomic characteristics of 25,000 primary and metastatic tumors from Memorial Sloan Kettering(32) using panel based sequencing, including on a subset of the samples analyzed in this work, the authors reported a higher fraction of *TP53* mutations in metastatic tumors, and higher levels of chromosomal instability, as well as an increase in *AR* and *ESR1* mutations in prostate and breast cancer, respectively.

From a clinical perspective the limited genomic differences found between primary and metastatic tumors are of particular interest. It indicates that in the absence of prior therapy, biopsies from primary tumors can be safely used as a treatment guide and that the need for additional primary or metastatic biopsies may not be necessary. An important caveat to our study is that all analysis is performed upon 174 shared genes. However, all of the 174 genes investigated in this study are previously reported cancer genes representing the bulk of known cancer driver events, making this a valid study of cancer driver landscape. A major limitation is that the primary and metastatic data is not from paired samples but from individual patients. Additionally, as a significant fraction of primary tumor patients eventually will develop metastatic disease, it is possible that metastasis-causing driver events may be present among the genes analyzed, but are not reaching statistical significance due to inaccurate classification of primary tumors. Unfortunately, to date only few and limited studies exist with paired data available. In contrast, the results of this study are derived from an exceptionally large number of patients, which increases the power of the results.

Taken together, our results suggest that acquisition of cancer driver mutations are initially mostly shaped by the tissue of origin where specific cancer driver mutations define the developing primary tumor, while acquisition of driver mutations that contribute to metastatic disease are less specific. This might suggest that the metastatic process is driven less by newly acquired metastatic features, but more by non-cancer features such as inflammation of the tumor microenvironment and in the surrounding tissue. These findings shed new light upon the mystery of metastatic cancer dissemination, the critical step that typically defines operable from inoperable disease and is generally lethal. Our findings here expands on the current understanding of metastatic cancer biology, but considering the limitations of the gene panel examined, they must be further investigated and confirmed in additional cancer cohorts using more inclusive genomic platforms, such as whole exome or whole genome sequencing.

Figure legends

Figure 1. Cohort characteristics. (A) Schematic overview of the Genie cohort and the analytical workflow. (B) Distribution of primary and metastatic samples within the 25 cancer types. (C) Number of driver events reported in the HMF 2,520 patient metastatic cohort(8) and number of these overlapping with the GENIE common gene set of 174 genes.

Figure 2. Summary mutations and copy number alterations. Primary samples on the x-axis and metastatic samples on the y-axis. Each cancer type is individually color coded. The dot size indicates if a given measure is significantly different between primary and metastatic within each cancer type. Only the significant cancer types are labeled. (A) Mean TMB across all cancer types. TMB is significantly higher in metastatic samples in 9/25 cancer types. (B) Mean wGII across all cancer types. wGII is higher in metastatic samples in 20/25 cancer types and significantly higher in 13/25 cancer types. (C) Mean number of driver mutations across all cancer types. Higher numbers of driver mutations in metastatic cancer are found in 8/25 cancer types, while for 4/25 cancer types more drivers are found in primary

samples. (D) The calculated ratio of driver mutations corrected for total mutational burden. For 12/25 cancer types the level of driver mutation/TMB ratio are found significantly higher in metastatic samples.

Figure 3. Enrichment of individual driver events. (A) Fraction of patients with driver events for each gene per cancer type. Only genes with values above 0.25 are labeled (B) Volcano plot showing enrichment or depletion in metastatic disease. FDR-adjusted p-value on the y-axis. (C) Summary overview, top panel showing for each gene the number of cancer types it is enriched, either in metastatic (orange) or in primary (blue) tumors. Bottom panel shows metastatic tumor percentage minus primary tumor percentage for each gene, in the cancer types showing significant enrichment based on the analysis in (B).

Figure 4. Metastatic tumors are enriched in specific variants and mutation signatures. (A) Using a one-sided Fisher's test performed within cancer types to evaluate enrichment in metastatic tumors, 9 specific gene variants are found significantly enriched. (B) The incidence of the significantly enriched variants and the cancer types they are found in are shown. (C-E) Mutation signatures in primary versus metastatic tumors. The Y-axis represents the numeric contribution of a specific mutation signature to the total TMB of each cancer type. Open circles indicate a signature is found but is not significantly different between primary and metastatic. (C) Treatment-related SBS11 (alkylating agents, temozolomide) and SBS35 (platinum chemotherapy). (D) The sum of APOBEC-related signatures, SBS2 and SBS13. (E) Smoking-related signature, SBS4. (F) Showing for each cancer type the contribution of individual mutation signatures, excluding the very common mitotic-related signatures SBS1 and SBS5. P-value based on a paired Wilcoxon test.

References

1. Turajlic S, Swanton C. Metastasis as an evolutionary process. Science. 2016;352:169–75.

2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.

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

4. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. Cell. 2017;168:670–91.

5. Birkbak NJ, McGranahan N. Cancer Genome Evolutionary Trajectories in Metastasis. Cancer Cell. 2020;37:8–19.

6. Reiter JG, Bozic I, Allen B, Chatterjee K, Nowak MA. The effect of one additional driver mutation

on tumor progression. Evol Appl. 2013;6:34–45.

7. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548:297–303.

8. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575:210–6.

9. van de Haar J, Hoes LR, Roepman P, Lolkema MP, Verheul HMW, Gelderblom H, et al. Limited evolution of the actionable metastatic cancer genome under therapeutic pressure. Nat Med. 2021;27:1553–63.

10. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov. 2017;7:818–31.

11. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013;501:338–45.

12. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

13. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47:D941–7.

14. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173:321–37.e10.

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

16. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.

17. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

18. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. The Open Journal; 2019;4:1686.

19. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots [Internet]. 2020. Available from: https://rpkgs.datanovia.com/ggpubr/

20. Hadley W, Seidel D. scales: Scale functions for visualization [Internet]. GitHub San Francisco:; 2019. Available from: https://CRAN.R-project.org/package=scales

21. Slowikowski K. ggrepel: Automatically Position Non-Overlapping Text Labels with “ggplot2” [Internet]. 2021. Available from: https://CRAN.R-project.org/package=ggrepel

22. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45:1127–33.

23. Reiter JG, Baretti M, Gerold JM, Makohon-Moore AP, Daud A, Iacobuzio-Donahue CA, et al. An analysis of genetic heterogeneity in untreated cancers. Nat Rev Cancer. 2019;19:639–50.

24. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333–9.

25. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013;339:1546–58.

26. Roberts KG, Odell AF, Byrnes EM, Baleato RM, Griffith R, Lyons AB, et al. Resistance to c-KIT kinase inhibitors conferred by V654A mutation. Mol Cancer Ther. 2007;6:1159–66.

27. Yun C-H, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong K-K, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A. 2008;105:2070–5.

28. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

29. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47:1402–7.

30. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354:618–22.

31. De Mattos-Arruda L, Sammut S-J, Ross EM, Bashford-Rogers R, Greenstein E, Markus H, et al. The Genomic and Immune Landscapes of Lethal Metastatic Breast Cancer. Cell Rep. 2019;27:2690–708.e10.

32. Nguyen B, Fong C, Luthra A, Smith SA, DiNatale RG, Nandakumar S, et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. Cell. 2022;185:563–75.e11.

33. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168:613–28.

34. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015;520:353–7.

35. Tariq S, Kim SY, Monteiro de Oliveira Novaes J, Cheng H. Update 2021: Management of Small Cell Lung Cancer. Lung. 2021;199:579–87.

36. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med. 2015;7:283ra54.

37. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell. 2019;176:1282–94.e20.

Acknowledgements

The authors would like to acknowledge the American Association for Cancer Research and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of study authors. NJB is a fellow of the Lundbeck Foundation (R272-2017-4040), and acknowledges funding from Aarhus University Research Foundation (AUFF-E-2018-7-14) and the Danish Cancer Society (R230-A13715). N.M is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant Number 211179/Z/18/Z), and also receives funding from Cancer Research UK Lung Cancer Centre of Excellence, Rosetrees, and the NIHR BRC at University College London Hospitals

Legends for the Supplementary figures

Figure S1. Selecting an optimal gene set. Number of shared genes on the left y-axis and number of patients on the right y-axis. Number of panels on the x-axis. The vertical line indicates selection cut-off.

Figure S2. Variant allele frequency in primary and metastatic tumors across cancer types. Mean VAF across all cancer types. VAF is significantly higher in metastatic samples in 10/25 cancer types. Primary samples on the x-axis and metastatic samples on the y-axis. Each cancer type is individually color coded. The dot size indicates if a given measure is significantly different between primary and metastatic within each cancer type. Only the significant cancer types are labeled.

Figure S3. Comparison of TMB and wGII. Primary samples are blue and metastatic samples orange. (A) TMB relative to wGII in primary tumors. Tumors harboring high levels of mutations showing lower levels of chromosomal alterations. (B) TMB relative to wGII in metastatic tumors. Tumors harboring high levels of mutations showing lower levels of chromosomal alterations. (C) TMB divided into bins. A significantly higher level of wGII in metastatic samples is found across all bins (D) wGII divided into bins. A significantly higher level of wGII in metastatic samples is found across 3 bins.

Figure S4. Limited differences observed between primary and metastatic tumors. Difference in percentage of driver mutations, comparing all primary to all metastatic tumors across the GENIE cohort. Y-axis shows metastatic tumor percentage minus primary tumor percentage for each gene. Genes with higher percentage in metastatic tumors are colored orange, while genes with higher percentage in primary samples are colored blue. The analysis is performed only for pathways where at least 50% of the assigned genes were found in the GENIE common geneset.

Figure S5. Enrichment of pathways with driver events. (A) fraction of patients with altered pathways in both primary and metastatic tumors. (B) The calculated ratio of affected pathways in both primary and metastatic samples. FDR-adjusted p-value on the y-axis. (C) Summary overview, top panel showing for each pathway the number of cancer types it is enriched, either in metastatic (orange) or in primary (blue) tumors. Bottom panel shows the percentage point difference between metastatic and primary tumors for each pathway, in the cancer types showing significant enrichment based on the analysis in (B).

Figure S6. Mutational signatures. Mutational signatures were determined separately within cancer types, by combining separately all primary mutations and all metastatic mutations and determining the contribution of individual mutational processes, as defined by Alexandrov and colleagues(15,28), using the deconstructSigs(16) R package (A) Contribution of individual mutational signatures to the total mutation burden within primary tumors. (B) Contribution of individual mutation signatures to the total mutation burden within metastatic tumors.