Cancer Genome Evolutionary Trajectories in Metastasis

Nicolai J Birkbak[#1,2], and Nicholas McGranahan[#3,4]

1. Department of Molecular Medicine, Aarhus University, Aarhus, Denmark
2. Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark
3. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, Paul O'Gorman Building, 72 Huntley Street, London, WC1E 6BT, United Kingdom
4. Cancer Genome Evolution Research Group, University College London Cancer Institute, University College London, London, UK

[#]correspondence should be addressed to: nbirkbak@clin.au.dk or nicolas.mcgranahan.10@ucl.ac.uk

Abstract
Metastatic cancer is a major cause of death and remains largely incurable. A better understanding of metastasis is therefore desperately needed to improve prognosis for late stage disease. Here we survey the landscape of studies exploring the genomics of metastatic cancer. We consider evidence for genomic drivers of metastasis and explore studies investigating modes of metastatic spread.

Introduction
Metastasis involves the dispersal and subsequent colonization of cancer cells from a primary tumor to a distant organ in the body. Each metastatic cell represents an evolutionary offshoot of its parental primary tumor, sharing genomic scars and key driver alterations, yet necessarily able to overcome the steps involved in the metastatic cascade. Despite remarkable progress in understanding and treating primary tumors, survival related to metastatic cancer remains poor, with 90% of cancer-related deaths linked to metastasis (Lambert et al., 2017).

Metastasis represents an evolutionary process (Turajlic and Swanton, 2016) and involves, in its most basic form, a living cancer cell entering circulation, surviving in and exiting the bloodstream, embedding into a remote tissue with a potentially hostile microenvironment, and forming a new metastatic tumor. Experimental work in mice has demonstrated this process is inherently inefficient; the vast majority of cancer cells either perish in circulation, getting stuck in capillaries, or undergo apoptosis within 24 hours of exiting the bloodstream(Lambert et al., 2017). And of the cancer cells that successfully colonise distant organs, evidence suggests that only a subset of these ever grow into macrometastatic tumours(Massague and Obenauf, 2016).

Cancer cells themselves are not under positive selection to metastasize. Rather, there are likely a set of key adaptations or hallmarks that may be selected which concomitantly increase the likelihood of a cancer cell acquiring the traits necessary to obtain metastatic potential, including motility, immune evasion and the ability to survive in circulation and proliferate at distant sites. Metastatic potential is likely not limited to a single or select few cell-autonomous traits but is dependent on the complex interaction of the cancer cell and the host stroma. The term “metastatic potential” as used below may thus cover any combination of cancer phenotypes that enable metastatic dissemination, whether primarily driven by rapid proliferation and cell shedding, increasing the probability of metastasis by sheer numbers of cells, or, by increased ability to survive in circulation and at sites of distant organs.

Traditionally metastatic dissemination has been thought of as the end-product of cancer development, however, with emerging data this linear view of metastatic evolution has become more nuanced. Recent work has revealed both early and late metastatic

dissemination (Klein, 2009; Yates et al., 2017), and metastatic seeding patterns consistent with monoclonal and polyclonal dispersal (Gundem et al., 2015).

Here we review studies in this important and burgeoning field. We focus on metastasis from the perspective of the cancer genome and consider differences in genomic events between cancer cells found in metastatic and primary tumor samples and explore whether common metastatic evolutionary trajectories can be identified within and between cancer types. We explore evidence for different modes and timing of metastasis and consider the evolutionary metastatic trajectories observed from primary tumors of different origins.

Defining the genomic landscape of metastatic disease

The somatic alterations found in a metastatic cancer do not simply reflect metastatic mutations. Rather, each metastatic cancer cells contains an archeological record of its past; some alterations will likely have been acquired by an ancestral 'normal' cell, some by an ancestral primary tumour cell and others after metastatic dissemination.

While all alterations in a given metastatic cancer cell reflect its lineage, the majority will be 'passengers' and only a subset can be considered 'drivers', conferring a fitness advantage upon the cancer cell (Stratton et al., 2009). The fitness effect will be dependent on the specific conditions of the cell upon the time of somatic acquisition, including the background of somatic and germline alterations that define the cell (epistatic interactions) as well as the specific microenvironment and selection pressures (Burrell et al., 2013; Lawrence et al., 2014; Vogelstein et al., 2013). Genes which harbour driver alterations can, by definition, be considered 'cancer genes' (Stratton et al., 2009).

Using statistical inference, large scale sequencing efforts have successfully identified scores of cancer-associated genes, as well as large-scale somatic copy number events, across different cancer types (Bailey et al., 2018; Carter et al., 2012; Lawrence et al., 2014; Martincorena et al., 2017; Zack et al., 2013). A deeper knowledge of the cancer genome and driver events has also heralded a rise in genome-driven cancer treatment, whereby patient's tumors are subject to sequencing in the hope of identifying genomic driver alterations that can be targeted with matched drugs (Mateo et al., 2018). The clinical relevance of such cancer driver alterations is clear; targeted therapies directed against identified driver events are often successful in inducing tumour regression. However, unfortunately, in the metastatic setting such responses are almost invariably temporary.

To date, studies to identify cancer driver alterations have predominantly focussed on primary disease, leaving the critical question of whether or not metastatic-specific cancer genes and drivers exist unanswered.

Two broad approaches for defining the genomic features of metastatic disease and the identification of metastasis-associated genes and metastatic driver alterations have been explored.

Large-scale cohort analysis of metastatic samples

Genes mutated at significant frequencies in metastatic cohorts will also include cancer-associated genes which are mutated at significant frequencies in the primary tumor. As such, to identify metastatic specific cancer genes based on statistical inference using unpaired cohorts of metastatic samples, frequency and effect sizes must be compared with cohorts of primary tumors (Figure 1).

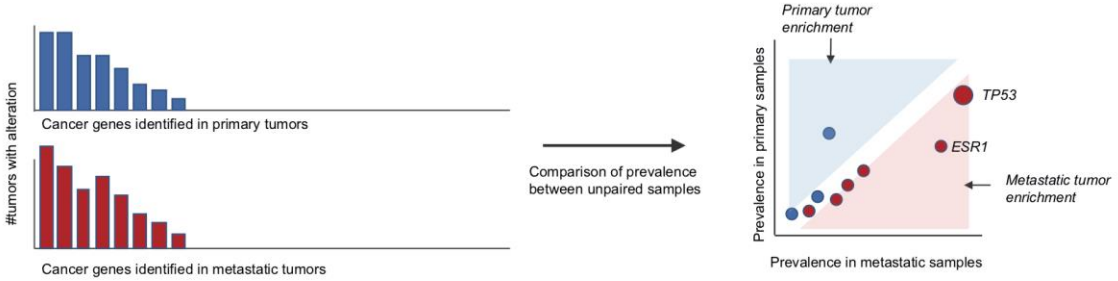

Figure 1 Large scale identification of metastatic drivers
Cancer genes can be identified independently in unmatched primary and metastatic using cancer gene tools such as MutSigCV2 (Lawrence et al., 2014) or dNdScv (Martincorena et al., 2017). Their prevalence and effect size can then be compared to identify cancer genes which are enriched or depleted in the context of metastatic disease.

Several studies have undertaken this approach in an attempt to identify drivers of metastasis (Bertucci et al., 2019; Priestley et al., 2019; Robinson et al., 2017). Yet to date, a conspicuous absence of cancer genes exclusively mutated in the context of metastatic disease have been identified.

A pan-cancer analysis of metastatic tumors, encompassing solid tumors from 30 tissues from 500 cancer patients, revealed a long-tailed distribution of cancer driver alterations, yet none were unique to metastatic disease (Robinson et al., 2017). Similar observations were made in larger studies of metastatic samples from mixed cancer types using both a panel based approach in a study of more than 10,000 advanced or metastatic cancer samples(Zehir et al., 2017) and in a recent pan-cancer analysis of 2,520 whole genome sequenced (WGS) metastatic tumors from the Hartwig Medical Foundation (Priestley et al., 2019). Indeed, the analysis based off Hartwig WGS data yielded only six novel cancer genes with high-confidence driver mutations not observed in previous analysis of more than 11,000 primary tumours (Bailey et al., 2018) (Figure 2A). Of these, four genes were previously labeled as cancer genes through either statistical analysis of a large pan-cancer cohort of primary tumor data (Martincorena et al., 2017) or reported in the cancer gene census, a list of genes causally implicated in cancer, published and maintained by the Catalogue Of Somatic Mutations In Cancer (COSMIC)(Sondka et al., 2018). A caveat to this analysis is that the datasets were not balanced in numbers or in cancer types, potentially more cancer genes firmly associated with metastatic disease may be found through analysis of larger cohorts of metastatic samples with specific histologies.

The likelihood of a somatic alteration being subject to positive selection in metastatic cancer will not only depend on the origin of the primary tumor, but also on the metastatic site itself and often on treatment-induced selection pressures. Exploration of 617 metastatic breast cancers uncovered nine established cancer genes (*TP53*, *ESR1*, *GATA3*, *KMT2C*, *NCOR1*, *AKT1*, *NF1*, *RIC8A* and *RB1*) that were more frequently mutated in the metastatic setting compared to early breast cancers from the TCGA (Bertucci et al., 2019). Mutations in *ESR1* are rarely found in primary tumors of ER positive breast cancer, but are likely induced by endocrine therapy and are found enriched specifically at sites of liver metastasis(Jeselsohn et al., 2018; Merenbakh-Lamin et al., 2013). Likewise, systematic comparisons of unmatched primary and metastatic prostate cancers, has revealed an enrichment of alterations in *TP53, PTEN, RB1, FOXA1, APC, BRCA2* and *AR* (Armenia et al., 2018). In the case of *AR* alterations, these are likely selected in response to androgen-deprivation therapy, enabling the AR pathway to be bypassed. Selective pressure from therapy influencing metastatic evolutionary trajectories is also evidenced in breast cancer where treatment of a patient harbouring a *PIK3CA* mutant primary tumor resulted in parallel evolution of six distinct resistance mutations in *PTEN* (Juric et al., 2015), and in NSCLC where phylogenetic analysis revealed how treatment of a patient harbouring an EGFR exon 19 deletion resulted in multiple

independent EGFR T790M resistance mutations arising in metastatic lesions(Blakely et al., 2017).

Further large-scale panel-based approaches such as the GENIE (Genomics, Evidence, Neoplasia, Information, Exchange) initiative (Consortium, 2017) aim to assemble and uniformly analyse thousands of cancer samples, both primary and metastatic, and make the data easily available for research. Panel-based approaches do not allow for identification of new cancer genes, but cohort-specific analysis provides the opportunity to investigate enrichment and depletion of a fixed set of known cancer genes across large numbers of cases. As exemplified in Figure 2 B-C for 1,897 primary and 1,133 metastatic samples from NSCLC patients from the GENIE cohort, direct comparison between primary and metastatic tumors reveal largely overlapping frequencies of commonly mutated genes. Despite the large number of cases included in this analysis, only TP53 remains significant following adjustment for false discovery rate (FDR). A limitation is that some primary samples are from patients with metastatic disease, thus these tumors may already harbor mutations that facilitate metastatic spread. Nevertheless, expanding the analysis to 11 cancer types with at least 150 metastatic and 150 primary samples, most genes show no significant depletion or enrichment in metastatic setting following FDR correction (Figure 2D). Of 13 unique genes, the two most highly significantly enriched are *ESR1* and *AR*, in breast and prostate cancer, respectively, reflecting their known role in endocrine therapy resistance (see above). Breast cancer shows enrichment of six additional genes. While these may play a role in therapy resistance, they may also be subtype-specific and reflect the increased risk of relapse of HER2 amplified and triple negative breast cancers. TP53 is the only gene significant across multiple cohorts. The role of TP53 in carcinogenesis is traditionally linked to defective cell cycle regulation, faulty apoptotic programs and aneuploidy(Bieging et al., 2014). Whether the enrichment in metastatic samples may be related to a crucial role of chromosomal instability in advancing metastatic potential, or if TP53 inactivation is associated with drug resistance remains to be elucidated. Overall, most genes with large proportional changes in frequency between primary and metastatic disease involve rare events, potentially supporting a diverse path towards metastatic potential.

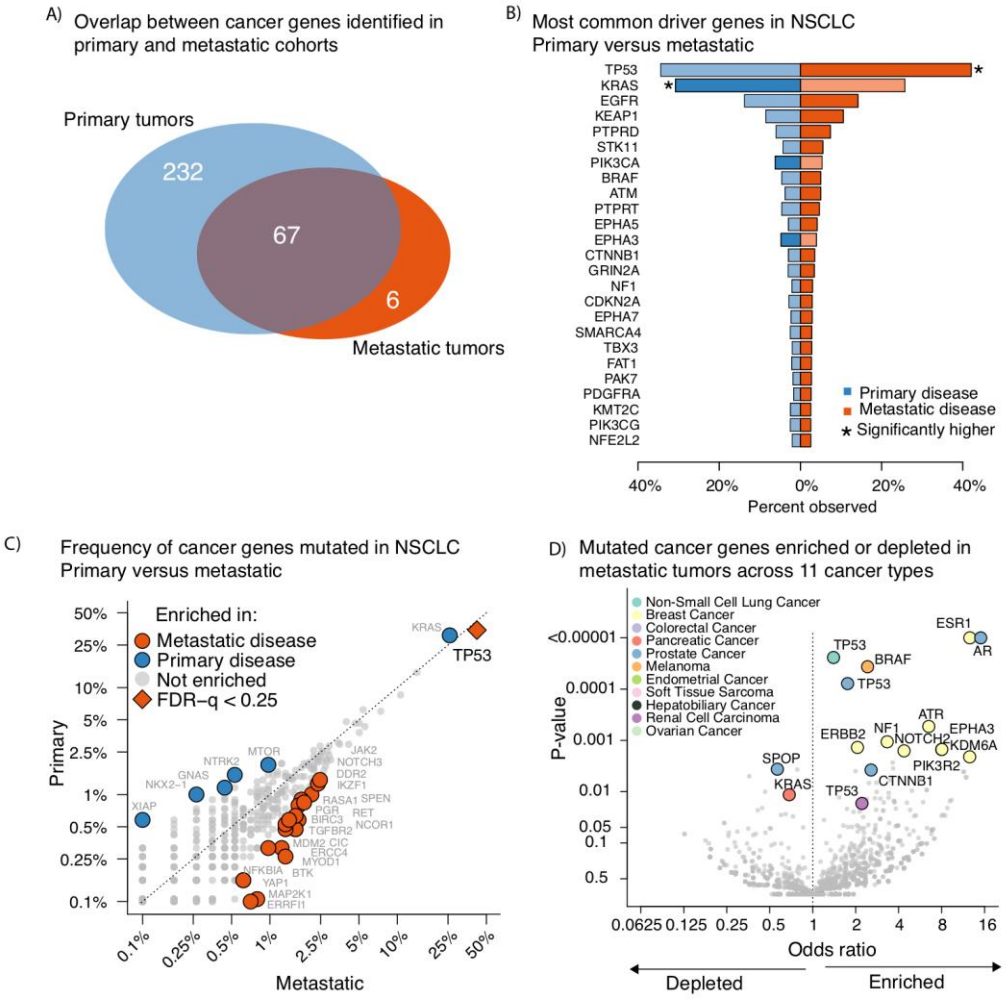

Figure 2 Comparison of NSCLC drivers for primary and metastatic cohorts

A) Overlap between driver genes as defined based on primary tumor cohorts(Bailey et al., 2018) and based on analysis of 2,520 metastatic tumor samples from 2399 patients(Priestley et al., 2019). B-D, analysis of mutations in driver genes reported in primary and metastatic samples in GENIE(Consortium, 2017) cohort, using intersection of 410 genes on panels MSK-468 and MSK-410. Driver mutation defined as “detrimental” by SIFT or “probably damaging” by PolyPhen or “pathogenic” by ClinVar. B) Top 25 most commonly mutated genes in metastatic NSCLC, comparing primary to metastatic disease based on 1,897 primary tumors and 1,133 metastatic tumors. Dark red indicates gene more commonly mutated in metastatic disease, dark blue indicates gene more commonly mutated in primary disease. Star indicates significance based on two-sided Fisher’s exact test without adjustment. C) Scatter plot of observed frequencies in primary versus metastatic NSCLC. Grey indicates non-significant difference based on uncorrected p-value < 0.05, diamond indicates p-value corrected for false discovery rate (FDR q-value) below 0.25 . D) Volcano plot showing enrichment or depletion in metastatic disease of 410 MSK genes across 11 cancers represented in GENIE with at least 150 primary and 150 metastatic samples. Y-axis shows uncorrected p-value. Grey indicates non-significant genes following FDR correction. Colored circles indicates FDR q-value < 0.25 in specific cancers.

Comparisons of cohorts of unmatched primary and metastatic tumors has revealed that metastatic tumors are frequently characterized by a higher burden of somatic alterations, including a higher frequency of driver mutations (Armenia et al., 2018; Robinson et al., 2017) and elevated levels of somatic copy number aberrations (Brown et al., 2017; Kawamata et al., 2018). This increase in alteration burden does not necessarily reflect elevated mutation rates in metastatic disease, but may be caused by metastatic seeding by a limited number of cancer cells. All cancer cells carry lineage-specific mutations, shared by only a minor population of cells, present at very low frequency in the tumor bulk. When the metastatic process is driven by a single or few cells seeding to distant location, this process effectively acts as a bottle-necking event, unmasking low frequency mutations, which then become clonal.

Given the extensive levels of aneuploidy observed in metastatic cancer, it has been suggested that somatic copy number events may play a critical role in enabling metastatic potential, equivalent to a macro-evolutionary speciation event (Gerlinger et al., 2014; Turajlic and Swanton, 2016). Consistent with this notion, one of the few genomic events that has been consistently linked with metastatic potential across cancer types is whole-genome doubling (WGD) (Bielski et al., 2018; Priestley et al., 2019). WGD may also be associated with persistent chromosomal instability, and may facilitate more rapid cancer genome evolution (Dewhurst et al., 2014; Selmecki et al., 2015). Consistent with this, WGD is associated with elevated copy number heterogeneity in both non-small cell lung cancer (NSCLC) and melanoma (Birkeland et al., 2018; Jamal-Hanjani et al., 2017). Conceivably, WGD may also act as a buffering event against excessive levels of somatic alterations by virtue of providing additional wild-type copies (López et al., 2019).

Taken together, large-scale genomic studies on metastatic cancer appears to show continued evolution along an axis of increased malignant potential, already identified in primary disease setting and even observed in clonally expanded normal cell populations (Brunner et al., 2019; Lee-Six et al., 2019; Martincorena et al., 2018; Martincorena et al., 2015). Metastatic samples generally show a higher mutation burden, more cancer driver events, and greater levels of aneuploidy than their primary cancer cell counterparts. However, despite extensive efforts, the holy grail of cancer research remains elusive; to this day we have been unable to identify any common events that may act as gate-keepers facilitating metastatic potential. Nevertheless, it remains possible that specific drivers that triggers metastatic growth may still be found through novel statistical and analytical approaches that also considers system-level changes such as replication timing, epigenomic landscape and transcriptional alterations.

Paired analysis of primary-metastatic samples

The success of precision oncology is contingent upon the assumption that knowing the genomic underpinning of a patient's cancer can guide the choice of therapy. Paired analysis of primary-metastatic samples has the potential to test whether a single biopsy from the primary tumor or metastasis is sufficient to inform treatment choices. Moreover, paired analysis can reveal whether specific events are enriched in metastatic sites within individual tumors, disentangling events which occurred prior to metastatic dissemination from those that may be directly involved, or those occurring after dissemination (Figure 3).

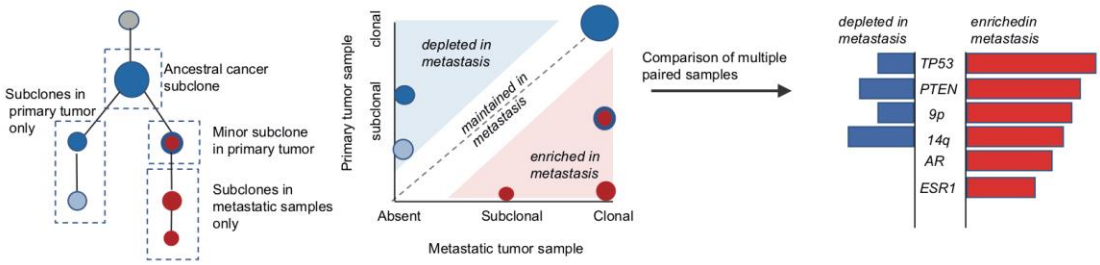

Figure 3 Paired analysis of primary and metastatic tumors

Paired analysis of primary and metastatic tumors permits identification of which clones are present at both primary and metastatic sites, and which are specific. Comparison of their prevalence can reveal which events may be putatively 'selected'. Comparisons of multiple paired samples can be used to infer whether any events are recurrently selected in metastatic clones.

Consistent with analysis of primary tumors (Dentro et al., 2018; McGranahan et al., 2015), exploration of driver heterogeneity between primary tumors and their matched metastases,

has revealed that the majority of driver alterations tend to be clonal, occurring on the trunk of a tumors phylogenetic tree and thus necessarily in every metastatic cancer cell (Makohon-Moore et al., 2017; Reiter et al., 2018; Turajlic et al., 2018b; Yates et al., 2017). However, while most driver alterations may be clonal, metastatic specific drivers have also been identified, some of which are clinically relevant. Evaluation of differences in drivers between primary tumors and matched brain metastasis found that in 53% of 86 cases, potentially clinically informative alterations were identified in the metastatic samples which were absent or not identified in the primary tumor (Brastianos et al., 2015). In particular, these included alterations involving the PI3K pathway, and proliferation signalling through the EGFR/ERBB2 receptors. However, more recent studies in breast(De Mattos-Arruda et al., 2019) and colorectal(Hu et al., 2019) cancer have shown that cancer clones metastasising to the brain may be clonally distinct from other sites, and exhibit early clonal divergence with increased numbers of private mutations relative to other metastatic clones. It is therefore possible that the high number of clinically relevant alterations observed in metastasis to the brain are not generally applicable to metastasis elsewhere.

Looking beyond currently clinically relevant genes, a recurring theme across a number of different studies, including metastatic breast (Yates et al., 2017), metastatic endometrial (Gibson et al., 2016), locally progressive hepatocellular carcinoma (He et al., 2015), as well as early stage lung cancer (Jamal-Hanjani et al., 2017), is for inactivation of genes involved in the SWI/SNF signalling, such as *ARID1A,* late in disease evolution. Chromatin structure imposes tissue-specific restrictions to gene transcription(Ziller et al., 2013). It is plausible reduced efficiency of epigenetic regulator complexes expand the potential for transcription-based phenotypic diversification, with potential for both increased proliferation and improved metastatic capacity through heightened ability to dynamically adapt to microenvironments across distant locations.

Intriguingly, in clear cell renal cell carcinoma (ccRCC) cancer clones that were enriched at metastatic sites compared to their matched primary tumors did not exhibit increased drivers as identified by single nucleotide variants or indels (Turajlic et al., 2018a). However, 9p loss, encompassing the p16 tumor suppressor gene, was significantly enriched in metastatic subclones, and together with loss of 14q was identified in 36/38 (95%) of cases with metastasis, yet only 22/62 (35%) of cases without metastatic disease at presentation. Both 9p and 14q loss were predominantly subclonal in the primary tumor, consistent with a bottle-necking event. Curiously, cancer clones metastasizing to the pancreas generally lacked 9p loss and also exhibited a much greater latency period compared to other metastatic sites (Turajlic et al., 2018a). This observation is consistent with the seed and soil hypothesis (Paget, 1989), suggesting clonal and phenotypic divergence depending on metastatic site, highlighting the need to consider evolutionary trajectories as highly dependent on both their site of origin and their destination.

Overall, the data support a cancer progression model where the majority of cancer-specific high-frequency driver events are acquired early in cancer evolution and are generally found clonally in primary tumors(Dentro et al., 2018; Jamal-Hanjani et al., 2017; McGranahan et al., 2015). These are well characterised and represents the bulk of clinically actionable targets (Mateo et al., 2018). Given their clonal nature, such events can reliably be identified in single biopsies from either primary or metastatic tissue. However, it also seems clear that most solid tumors undergo subclonal divergence in the primary, with subclonal drivers often sampling the tail of putative cancer genes, mutated at low-frequency. Thus, while our knowledge of high-frequency clonal drivers may be reaching saturation (Lawrence et al., 2014), exploration of the long tail of subclonal drivers requires further attention, from both a biological and clinical perspective.

Moreover, data from these studies suggest it is increasingly likely that the key to metastatic proliferation lies not simply with the sum of all parts but rather depends strongly on context,

including order of events and the clonal composition (Rogers et al., 2018; Turajlic et al., 2018a) that may define disease state, dictate metastatic potential, influence location of distant metastasis, and, from a therapeutic perspective, determine response to therapy. Indeed, the genomic features that may be selected and confer successful colonization at one site may be drastically different to those at another site, such that exploration of driver alterations must consider the specific fitness landscapes of different disease sites, and moreover, how this is influenced by current treatment strategies.

Modes of metastatic dissemination

The extensive heterogeneity and branched tumor evolution documented within primary tumors (de Bruin et al., 2014; Gerlinger et al., 2012; McPherson et al., 2016; Nik-Zainal et al., 2012; Yates et al., 2017) leads to questions as to whether metastasis is seeded by a single subclone, from one branch of the phylogenetic tree, or multiple subclones, conceivably on separate branches of the tumor's phylogenetic tree (Figure 4).

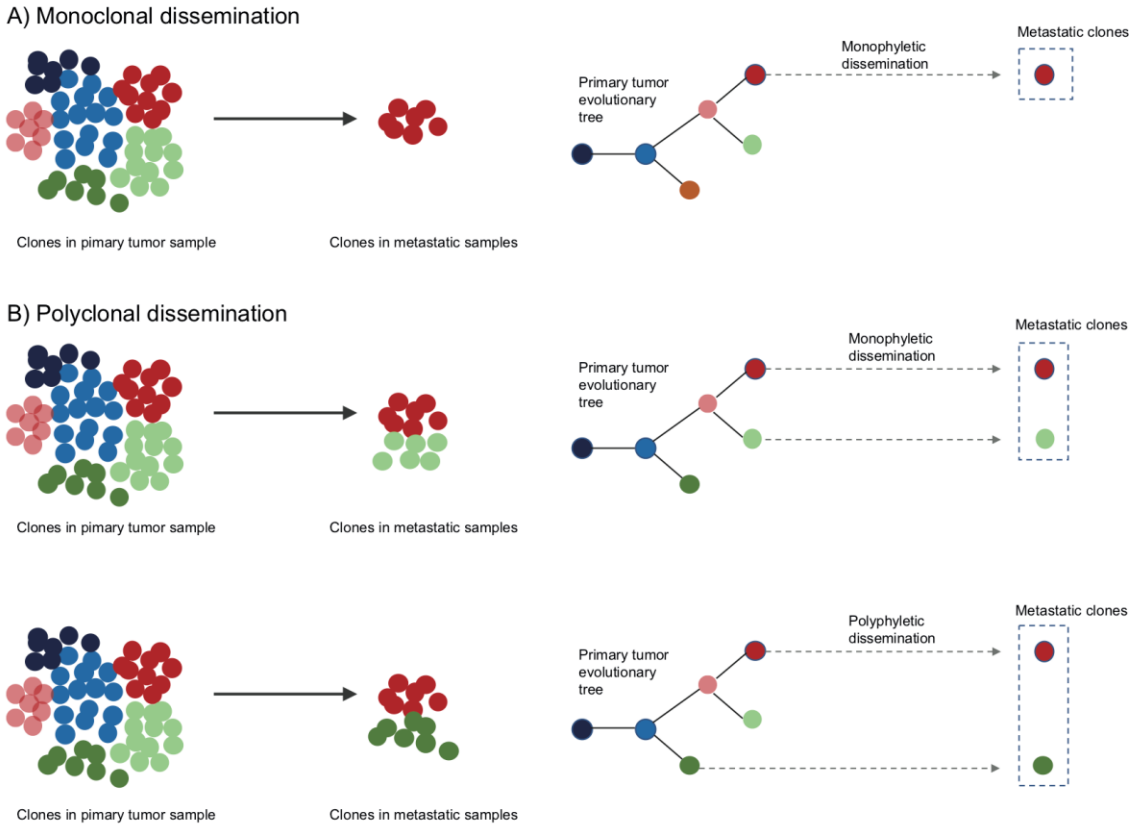

Figure 4 Modes of metastatic dissemination from the primary tumor

A) Monoclonal dissemination involves only one clone in the primary tumor. Monoclonal dissemination is necessarily monophyletic, involving only one branch of the primary tumor's phylogenetic tree. B) Polyclonal dissemination involves more than one clone in the primary tumor. If disseminating clones share a recent common ancestor, thus descending from one branch of the tumor's phylogenetic tree, this may reflect monophyletic dissemination. Conversely, if disseminating clones are derived from distinct branches and their most recent common ancestor represents the trunk of the primary tumor's phylogenetic tree, this will reflect polyphyletic dissemination.

Monoclonal dissemination from primary tumor

Monoclonal dissemination is defined as a single subclone within the primary tumor acquiring metastatic potential and seeding all metastatic lesions. This model is consistent with a bottle-necking event such that all metastatic cells necessarily share a recent common ancestor, and thereby, conceivably, may respond similarly to first-line therapy. Notably, due to the bottle-

necking event, monoclonal dissemination will almost necessarily lead to an increase in detectable mutation burden between primary and metastatic disease.

Importantly, although further metastasis-to-metastasis dispersal may involve multiple clones, under this definition, it is still monoclonal dissemination from the primary tumor given only a single primary tumor clone is involved. Monoclonal dissemination is necessarily *monophyletic* in origin, reflecting only one branch of a tumor's phylogenetic tree.

Polyclonal dissemination from primary tumor

By contrast, polyclonal dissemination is defined as one or more metastatic lesions seeded by multiple distinct subclones derived from the primary tumor. In this model, metastatic potential has either been acquired once in the evolution of the primary tumor, ensuring the capability to metastasise is inherited by any clone that may subsequently arise, or metastatic potential has evolved in parallel across multiple independent clones. In the former case, where metastatic potential is acquired once, polyclonal dissemination may have a *monophyletic* origin such that all metastatic clones share a common ancestor. However, in the latter case, polyclonal seeding will also be *polyphyletic*, whereby distinct metastatic clones may be more similar to clones primary than to each other. Distinguishing whether metastatic potential is acquired once or multiple times will often require further experimental analysis. In practice, polyphyletic seeding is often assumed if distinct metastasizing clones are more similar to other subclones in the primary tumour than they are to each other.

Polyclonal seeding to a single metastatic tumor site may result from either multiple waves of migrating cells or, alternatively, simultaneous migration of clusters of cells composed of genetically distinct clones.

Importantly, monoclonal, polyclonal, monophyletic and polyphyletic dissemination can be considered at both the patient-level and also the site-level. For instance, a patient may harbor metastatic disease at multiple different sites. Considered individually, each site may exhibit monoclonal, monophyletic, dissemination. However, taken as a whole, at the patient-level, this may be classified as polyclonal, polyphyletic, given that dissemination at each metastasis may derive from a separate branch of the primary tumor's phylogenetic tree (this is also referred to as *parallel metastatic dissemination*).

Following metastatic dissemination, cancer evolution and therapeutic selection pressures will further influence the clonal composition at metastatic sites. Indeed, to complicate matters, while monoclonal dissemination will result in an initial homogeneous population of metastatic cells, continued evolution will necessarily result in further heterogeneity and potentially a complex, polyclonal, subclonal structure. Moreover, multi-source seeding and reseeding between the primary tumor and metastasis can further complicate inference of metastatic evolutionary trajectories. For simplicity, here we exclusively focus on modes of metastatic dissemination from the primary tumor.

What are the main modes of metastatic dissemination observed in human tumors?

To address whether a seeding event is monophyletic or polyphyletic from the primary tumor requires clonal decomposition and thereby multi-region sampling of both the primary tumor and metastatic biopsies. As such, to date, the number of studies with samples capable of addressing this issue directly is limited. Nevertheless, the literature holds evidence for multiple modes of metastatic spread, with emerging data suggesting the relative likelihood may be influenced both by the primary tumor and the site of metastasis.

In the case of brain metastasis deriving from primary tumors from a range of sites, including breast (n=38), lung (n=21) and ccRCC (n=10), an absence of polyclonal or polyphyletic metastasis has been observed (Brastianos et al., 2015), with all metastatic samples clonal in origin and more similar to each other than the primary tumor samples. Likewise, thrombus

metastasis from ccRCC all appear to be monoclonal and monophyletic, reflecting rapid outgrowth of a single clone (Turajlic et al., 2018a). However, pervasive polyclonal and polyphyletic seeding of independent metastatic lesions has been observed from primary tumors derived from esophageal cancer (Noorani et al., 2018). In this case, the authors suggest metastasis occurs rapidly across multiple sites. In ccRCC, while 19/24 tumors exhibited monophyletic disease, 4/5 cases with polyphyletic disease were classified as rapid progressors, indicating early onset, with a simple clonal architecture and limited phylogenetic diversity between separate branches of the phylogenetic tree.

For other cancer types, different studies have reported differences in the prevalence of distinct seeding patterns. For example, in breast cancer, two autopsy studies each of 10 patients observed a polyclonal seeding pattern in either a subset (Brown et al., 2017) or in most patients (De Mattos-Arruda et al., 2019). Both studies report heterogeneous metastatic tumors, but while Brown and colleagues(Brown et al., 2017) infer a predominant model of monoclonal seeding followed by metastasis-to-metastasis seeding, Mattos-Arruda and colleagues (De Mattos-Arruda et al., 2019) conclude that metastases are initiated and maintained as groups of cellular clones, suggesting a polyclonal seeding pattern. Likewise, Ullah et al. (2018) identified a polyclonal seeding pattern in eleven of fifteen patients based on presence of shared subclonal mutations across sites. In this case, all four of the patients with monoclonal seeding had primary cancers of luminal subtype, while the remaining tumors were a mix of non-luminal (8/11) and luminal (3/11) (Ullah et al., 2018). Polyclonal seeding was also reported in an analysis of two metastatic triple negative breast cancers (Hoadley et al., 2016). In one case, the authors suggest the lung is seeded by clones from the breast and rib, while polyclonal seeding occurs from lung to brain and from liver to kidney.

While many studies have reported polyclonal seeding, a recent study has emphasized the need for caution when interpreting a polyclonal seeding patterns based on clustering mutations from bulk sequenced tumours (El-Kebir et al., 2018). There is considerable uncertainty in the estimation of each mutation's variant allele frequency (VAF) - which will be influenced by the sequencing depth, tumor purity and mappability of the genomic region where the mutation resides. As such not all mutation clusters, which are often assumed to reflect subclones, will necessarily be correct. If this uncertainty is ignored, the inferred seeding pattern may be erroneously interpreted. Re-analysis of the data from Hoadley et al. (2016) using a novel tool, which combines subclonal clustering with inference of migration histories, suggests monoclonal rather than polyclonal spread to be the most parsimonious explanation (El-Kebir et al., 2018). Likewise, without rigorous filtering approaches, low-frequency artifact mutations, present in all regions, may give an illusion of polyclonal spread. Conversely, undersampling of the primary and/or metastasis is more likely to lead to monoclonal inferences. Indeed, without comprehensively sampling the tumour mass it often difficult to unequivocally exclude the possibility of polyclonal spread.

In support of the suggestion that difference in the modes of metastasis does not necessarily simply reflect underlying biology, different studies in colon cancer have reported differences in the prevalence of mono versus polyclonal seeding. Wei and colleagues(2017) analysed four colorectal cancers and identified polyclonal seeding in every case. However, a more recent analysis of 118 samples from 23 patients found most metastatic lesions were monoclonal (Hu et al., 2019). A detailed longitudinal analysis of 36 samples from two colon cancer patients collected over 11 years showed multiple metastatic waves with polyclonal dissemination where single clones either from the primary tumor or from established metastatic tumors seeded new individual metastatic tumors(Angelova et al., 2018). In metastatic prostate cancer, an autopsy study using whole-genome sequencing of 51 tumors from 10 patients revealed that for half the patients, multiple distinct clones were involved in metastatic spread. Intriguingly, in these five polyclonal cases, re-seeding at multiple sites was also observed, consistent with the notion that clones with a significant survival advantage can often spread to and reseed other sites(Gundem et al., 2015). While likely polyclonal in origin,

further sequencing of the primary tumor would be required to elucidate whether these metastases were monophyletic. An analysis of seven patients with ovarian cancer revealed that five patients exhibited monoclonal seeding from ovary to intraperitoneal sites, while for two patients there was evidence of polyclonal spread, emerging from distinct branches of each tumor's phylogenetic tree. A similar pattern is observed in bladder cancer, where 3/6 patients demonstrated monophyletic spread (Faltas et al., 2016). In the TRACERx non-small cell lung cancer study, metastatic dissemination was of monoclonal origin in three out of five NSCLC cases, with one patient showing polyphyletic spread to lymph nodes, and one patient showing polyclonal and polyphyletic seeding to a single metastatic lesion (Abbosh et al., 2017).

Documented differences in prevalence of patient-level monoclonal versus polyclonal seeding is often contingent upon whether lymph nodes are included as part of the analysis. Given their prognostic value, the prevailing model of metastatic dissemination is that lymph node metastases subsequently give rise to distant metastasis. However, as documented above for NSCLC, and also in breast cancer (Ullah et al., 2018), esophageal cancer (Noorani et al., 2018) and colorectal cancer (Naxerova et al., 2017), divergent evolutionary trajectories between lymph node and distant metastasis appear to be the norm, with a possible exception in ccRCC (Turajlic et al., 2018a)

Overall, it is clear that both mono- and polyclonal dissemination occurs in cancer. Nonetheless, the still nascent literature on metastatic disease subjected to multiple sampling indicates a potential preference for a bottle-necking event resulting in monoclonal metastatic dissemination to individual metastatic sites (site-level monoclonal seeding). However, as exemplified from the re-analysis of data described above (El-Kebir et al., 2018), there are uncertainties inherent in inferring metastatic evolutionary trajectories from sequencing data Ultimately, additional studies on larger cohorts with extensively sampled metastatic disease, which may be obtained through analysis of post-mortem samples (Iacobuzio-Donahue et al., 2019), will be required to address the prevalence and clinical relevance of different modes of metastasis, and moreover the key factors, such as therapeutic selection pressures, that may dictate the likelihood of different evolutionary trajectories.

Timing of metastasis

The likelihood of different modes of metastasis and also the extent of metastatic heterogeneity is intimately linked to the timing of metastasis. Traditionally, metastasis has been viewed as the end-product of tumor development. However, recent reports have suggested that dissemination can occur early in a tumor's evolutionary history, prior to primary tumor resection, and that primary and metastatic tumors evolve in parallel. Such early dissemination of cancer cells has been supported by findings that even in early stage breast cancer patients (M0 stage), disseminated cancer cells (DCCs) from the bone marrow can be detected (Klein et al., 2002). These results were also supported in a HER2 driven mouse model of breast cancer where early DCC were found to exhibit elevated metastatic ability compared to cells that disseminated late (Hosseini et al., 2016) and recent modelling of colorectal cancer development which suggested that metastatic dissemination occurs before the primary tumor is clinically detectable in the majority of cases (Hu et al., 2019).

However, distinguishing between early and late dissemination models is not straightforward (Turajlic and Swanton, 2016). Given that a primary tumor is seldom exhaustively sampled, genomic divergence between primary tumor and metastasis does not necessarily equate with early dissemination. Moreover, recent analysis of copy number and somatic single nucleotide variants has suggested that only a fraction of the cells commonly assumed to be DCCs - derived from bone marrow aspirates in breast cancer - actually derive from the same lineage as the tumor clones identified in the primary tumor (Demeulemeester et al., 2016). Such a late dissemination model is also supported by observations in a cohort of 170 patients with locally relapsed or metastatic breast cancer where metastatic divergence was found to occur at 87%

of molecular time within the primary tumor and the majority of driver alterations found in the primary cancer were also present in the relapsing clone and (Yates et al., 2017).

Given the significant clinical implications of early versus late metastatic spread, developing methods to detect occult disease at the time of diagnosis is of great importance. A current frontrunner to overcome limitations of imaging-based diagnostics is analysis of circulating tumor DNA (ctDNA)(Abbosh et al., 2018; Wan et al., 2017). Approaches utilising ctDNA for early diagnosis of relapse has recently proved successful across multiple cancer types(Abbosh et al., 2017; Christensen et al., 2019; Phallen et al., 2017; Reinert et al., 2019) where the presence of ctDNA in liquid biopsies at resection or during follow-up predicts for recurrence with near-perfect specificity. However, sensitivity of ctDNA approaches remains an issue. Not all lesions release detectable amounts of ctDNA despite considerable tumour mass(Abbosh et al., 2017). Rather, the association between tumour burden and ctDNA release may depend on lesion-specific biology such as fast versus slow proliferation, and on tumour location, where particularly brain tumours may require alternate sampling strategies incorporating cerebrospinal fluids(Miller et al., 2019). Nevertheless, as the technology matures and enters clinical practise, increased sensitivity may present ctDNA assays as an excellent tool to supplement diagnostic imaging to detect early metastatic spread and guide treatment choices.

The immune system and the cancer genome

Recent work has begun to shed light on how the immune microenvironment may shape cancer genome evolution, and metastatic trajectories. In a detailed study of 31 metastases from two cases of stage IV CRC with exceptionally long survival, Angelova and colleagues explored the temporal dynamics of immunoediting. Echoing a previous autopsy study documenting multiple metastatic sites of ovarian cancer (Jimenez-Sanchez et al., 2017) and recent findings in breast cancer (De Mattos-Arruda et al., 2019), multiple distinct immune microenvironments were found at different metastatic sites. Strikingly, in this small case study, the likelihood of clonal persistence was found to be directly related to immunoediting, with only clones that were unedited or non-immunogenic found to survive (Angelova et al., 2018).

In ovarian cancer evolution, Zhang and colleagues found evidence those tumor regions harbouring the highest levels of immune infiltrate exhibiting the lowest cancer cell clonal diversity, coupled with neoantigen depletion and loss of heterozygosity (LOH) at the human leukocyte antigen (HLA) loci. Consistent with this, HLA LOH, which restricts the number of neoantigens presented by the cancer cell, has previously been linked to brain metastasis (McGranahan et al., 2017). Moreover, in a cohort of breast cancers, HLA LOH was found to be enriched in lymph node samples compared to their matched primary tumor regions (Messaoudene et al., 2019). These results are also consistent with findings that down-regulation of interferon regulatory factor 7 (Irf7) in breast cancer cells can decrease MHC expression and promote bone metastasis (Bidwell et al., 2012).

The prevalence of HLA LOH during metastatic transition highlights the importance of immune-mediated escape from neoantigen presentation. Interestingly, while shared truncal mutations, and by extensions neoantigens, are expected to be found across different metastatic sites, recent work suggests reactive T-cells towards such neoantigens can also be found at all sites. T-cell receptor (TCR) sequencing in breast and ovarian cancer has revealed the presence of shared TCRs across distinct metastasis, implying immune surveillance between metastatic sites (De Mattos-Arruda et al., 2019; Zhang et al., 2018). Analogous to these observations, in primary NSCLC subjected to multi-region sampling, the number of expanded TCRs found ubiquitously across all tumor samples correlated with the number of ubiquitous mutations, but not with the number of regional mutations(Joshi et al., 2019). Together this implies that some level of immune surveillance directed against clonal neoantigens may be initiated early and maintained through all levels of cancer development, including metastatic progression.

However, as is evident through the lack of cancer cell elimination, immune escape mechanisms are able to prevent T-cell mediated death.

Taken together, these studies point towards a next phase of cancer evolution analysis, moving beyond genomic characterisation towards a more integrated ecosystem-level appreciation of cancer development. Indeed, while this review has focussed on genomic evolutionary trajectories from the point of view of the cancer genome, a more holistic view will be required to fully understand metastatic disease. This should include considerations of pre-metastatic niches (Peinado et al., 2017), the mechanistic basis of metastatic dissemination(Chaffer and Weinberg, 2011), hypoxia and the extracellular matrix (Barkan et al., 2010; Gilkes et al., 2014), and the role of epithelial-mesenchymal transition(David et al., 2016; Esposito et al., 2019; Kalluri and Weinberg, 2009)

Conclusions

While it has long been established that the metastatic process itself is highly non-random (Paget, 1989), our knowledge of the cancer genome evolutionary trajectories of metastasis is only in its infancy.

Accumulating evidence points to an absence of universal metastatic specific driver alterations, which are exclusive to metastatic disease. Indeed, most studies have found that the majority of established driver alterations occur prior to metastatic dissemination (Dentro et al., 2018; Reiter et al., 2018; Turajlic et al., 2018a). Nevertheless, recent data suggests certain alterations are enriched at metastatic sites, and analysis of primary tumors and their matched metastasis has revealed drivers specific to metastatic clones (Armenia et al., 2018; Bertucci et al., 2019; Turajlic et al., 2018a; Yates et al., 2017). While some of these can be linked to therapeutic selection pressures (e.g. *ESR1* mutations in breast cancer), others appear to confer a more aggressive phenotype (e.g. 9p loss in ccRCC).

Across cancer-types, whole-genome doubling has emerged as a common feature enriched in metastatic samples (Bielski et al., 2018). Whether genome doubling itself engenders metastatic potential, or whether it reflects a proxy for elevated chromosomal instability, which may facilitate more rapid cancer genome evolution, permitting the tumor to explore a greater evolutionary space, remains unclear. A related recurring theme across cancers involves disruption to chromatin modification later in the disease course (Jamal-Hanjani et al., 2017; Yates et al., 2017), which may also play a role in relaxing genomic constraints imposed upon the cancer genome.

Mapping of the clonal composition of tumors has revealed a spectrum of different metastatic seeding patterns, which depend on both the site of the primary tumor and destination of metastasis. In certain cases, such as metastasis to the brain, a monoclonal seeding pattern appears pervasive (Brastianos et al., 2015). In other cases, such as metastasis derived from esophageal adenocarcinomas, a polyclonal seeding pattern dominates, with multiple distinct clones within the primary tumor harboring metastatic potential (Noorani et al., 2018). Emerging data suggests metastatic cancer clones within a given organ are more likely to share a recent common ancestor, compared to clones metastasizing to different organs in the same patient. Indeed, it is becoming clear that we must consider both the fitness landscape of the primary tumor and each site metastasis independently to begin to shed light on the complexities of metastatic evolutionary trajectories.

However, our evolutionary maps of metastatic disease remain far from complete and are currently riddled with uncertainties (El-Kebir et al., 2018). Building a more comprehensive picture of the landscape of metastatic disease, characterising the prevalence and importance of different modes of metastasis, will require comprehensive sampling of primary and matched tumors which can often only be achieved through autopsy studies, coupled with investigation of the immune microenvironment which plays a role in sculpting their evolution. Such detailed

cartography of metastasis is required to inform clinical decision making and create new therapeutic opportunities.

Acknowledgements
N.J.B. is a fellow of the Lundbeck Foundation and acknowledges funding from the Aarhus University Research Foundation and the Danish Cancer Society. 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 CRUK, Rosetrees, and the NIHR BRC at University College London Hospitals. 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. Finally, we also wish to thank Maise Al Bakir for carefully reading the manuscript and providing invaluable input.

Declaration of Interests
N.M. has received consultancy fees from Achilles Therapeutics.

References

Abbosh, C., Birkbak, N. J., and Swanton, C. (2018). Early stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection. Nat Rev Clin Oncol *15*, 577-586.
Abbosh, C., Birkbak, N. J., Wilson, G. A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D. A., Veeriah, S., Rosenthal, R.*, et al.* (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature *545*, 446-451.
Angelova, M., Mlecnik, B., Vasaturo, A., Bindea, G., Fredriksen, T., Lafontaine, L., Buttard, B., Morgand, E., Bruni, D., Jouret-Mourin, A.*, et al.* (2018). Evolution of Metastases in Space and Time under Immune Selection. Cell *175*, 751-765 e716.
Armenia, J., Wankowicz, S. A. M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W. K., Chakravarty, D., Han, G. C., Coleman, I.*, et al.* (2018). The long tail of oncogenic drivers in prostate cancer. Nature genetics *50*, 645-651.
Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B.*, et al.* (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385 e318.
Barkan, D., Green, J. E., and Chambers, A. F. (2010). Extracellular matrix: a gatekeeper in the transition from dormancy to metastatic growth. Eur J Cancer *46*, 1181-1188.
Bertucci, F., Ng, C. K. Y., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N., Soria, J. C., Dien, A. T., Adnani, Y., Kamal, M.*, et al.* (2019). Genomic characterization of metastatic breast cancers. Nature *569*, 560-564.
Bidwell, B. N., Slaney, C. Y., Withana, N. P., Forster, S., Cao, Y., Loi, S., Andrews, D., Mikeska, T., Mangan, N. E., Samarajiwa, S. A.*, et al.* (2012). Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape. Nature medicine *18*, 1224-1231.
Bieging, K. T., Mello, S. S., and Attardi, L. D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. Nature reviews Cancer *14*, 359-370.
Bielski, C. M., Zehir, A., Penson, A. V., Donoghue, M. T. A., Chatila, W., Armenia, J., Chang, M. T., Schram, A. M., Jonsson, P., Bandlamudi, C.*, et al.* (2018). Genome doubling shapes the evolution and prognosis of advanced cancers. Nature genetics *50*, 1189-1195.
Birkeland, E., Zhang, S., Poduval, D., Geisler, J., Nakken, S., Vodak, D., Meza-Zepeda, L. A., Hovig, E., Myklebost, O., Knappskog, S., and Lonning, P. E. (2018). Patterns of genomic evolution in advanced melanoma. Nature communications *9*, 2665.
Blakely, C. M., Watkins, T. B. K., Wu, W., Gini, B., Chabon, J. J., McCoach, C. E., McGranahan, N., Wilson, G. A., Birkbak, N. J., Olivas, V. R.*, et al.* (2017). Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. Nat Genet *49*, 1693-1704.

Brastianos, P. K., Carter, S. L., Santagata, S., Cahill, D. P., Taylor-Weiner, A., Jones, R. T., Van Allen, E. M., Lawrence, M. S., Horowitz, P. M., Cibulskis, K.*, et al.* (2015). Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. Cancer discovery *5*, 1164-1177.
Brown, D., Smeets, D., Szekely, B., Larsimont, D., Szasz, A. M., Adnet, P. Y., Rothe, F., Rouas, G., Nagy, Z. I., Farago, Z.*, et al.* (2017). Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. Nature communications *8*, 14944.
Brunner, S. F., Roberts, N. D., Wylie, L. A., Moore, L., Aitken, S. J., Davies, S. E., Sanders, M. A., Ellis, P., Alder, C., Hooks, Y.*, et al.* (2019). Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. Nature *574*, 538-542.
Burrell, R. A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature *501*, 338-345.
Carter, S. L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P. W., Onofrio, R. C., Winckler, W., Weir, B. A.*, et al.* (2012). Absolute quantification of somatic DNA alterations in human cancer. Nature biotechnology *30*, 413-421.
Chaffer, C. L., and Weinberg, R. A. (2011). A perspective on cancer cell metastasis. Science (New York, NY *331*, 1559-1564.
Christensen, E., Birkenkamp-Demtroder, K., Sethi, H., Shchegrova, S., Salari, R., Nordentoft, I., Wu, H. T., Knudsen, M., Lamy, P., Lindskrog, S. V.*, et al.* (2019). Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *37*, 1547-1557.
Consortium, A. P. G. (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer discovery *7*, 818-831.
David, C. J., Huang, Y. H., Chen, M., Su, J., Zou, Y., Bardeesy, N., Iacobuzio-Donahue, C. A., and Massague, J. (2016). TGF-beta Tumor Suppression through a Lethal EMT. Cell *164*, 1015-1030.
de Bruin, E. C., McGranahan, N., Mitter, R., Salm, M., Wedge, D. C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A. J.*, et al.* (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science (New York, NY *346*, 251-256.
De Mattos-Arruda, L., Sammut, S. J., Ross, E. M., Bashford-Rogers, R., Greenstein, E., Markus, H., Morganella, S., Teng, Y., Maruvka, Y., Pereira, B.*, et al.* (2019). The Genomic and Immune Landscapes of Lethal Metastatic Breast Cancer. Cell reports *27*, 2690-2708 e2610.
Demeulemeester, J., Kumar, P., Moller, E. K., Nord, S., Wedge, D. C., Peterson, A., Mathiesen, R. R., Fjelldal, R., Zamani Esteki, M., Theunis, K.*, et al.* (2016). Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. Genome biology *17*, 250.
Dentro, S. C., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., Deshwar, A. G., Yu, K., Rubanova, Y., Macintyre, G., Vázquez-García, I.*, et al.* (2018). Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. bioRxiv, 312041.
Dewhurst, S. M., McGranahan, N., Burrell, R. A., Rowan, A. J., Gronroos, E., Endesfelder, D., Joshi, T., Mouradov, D., Gibbs, P., Ward, R. L.*, et al.* (2014). Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. Cancer discovery *4*, 175-185.
El-Kebir, M., Satas, G., and Raphael, B. J. (2018). Inferring parsimonious migration histories for metastatic cancers. Nature genetics *50*, 718-726.
Esposito, M., Mondal, N., Greco, T. M., Wei, Y., Spadazzi, C., Lin, S. C., Zheng, H., Cheung, C., Magnani, J. L., Lin, S. H.*, et al.* (2019). Bone vascular niche E-selectin induces mesenchymal-epithelial transition and Wnt activation in cancer cells to promote bone metastasis. Nat Cell Biol *21*, 627-639.

Faltas, B. M., Prandi, D., Tagawa, S. T., Molina, A. M., Nanus, D. M., Sternberg, C., Rosenberg, J., Mosquera, J. M., Robinson, B., Elemento, O*., et al.* (2016). Clonal evolution of chemotherapy-resistant urothelial carcinoma. Nature genetics.
Gerlinger, M., McGranahan, N., Dewhurst, S. M., Burrell, R. A., Tomlinson, I., and Swanton, C. (2014). Cancer: evolution within a lifetime. Annual review of genetics *48*, 215-236.
Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P*., et al.* (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine *366*, 883-892.
Gibson, W. J., Hoivik, E. A., Halle, M. K., Taylor-Weiner, A., Cherniack, A. D., Berg, A., Holst, F., Zack, T. I., Werner, H. M., Staby, K. M*., et al.* (2016). The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. Nature genetics *48*, 848-855.
Gilkes, D. M., Semenza, G. L., and Wirtz, D. (2014). Hypoxia and the extracellular matrix: drivers of tumour metastasis. Nature reviews Cancer *14*, 430-439.
Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L. B., Tubio, J. M., Papaemmanuil, E., Brewer, D. S., Kallio, H. M., Hognas, G., Annala, M*., et al.* (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353-357.
He, F., Li, J., Xu, J., Zhang, S., Xu, Y., Zhao, W., Yin, Z., and Wang, X. (2015). Decreased expression of ARID1A associates with poor prognosis and promotes metastases of hepatocellular carcinoma. J Exp Clin Cancer Res *34*, 47.
Hoadley, K. A., Siegel, M. B., Kanchi, K. L., Miller, C. A., Ding, L., Zhao, W., He, X., Parker, J. S., Wendl, M. C., Fulton, R. S*., et al.* (2016). Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases. PLoS Med *13*, e1002174.
Hosseini, H., Obradovic, M. M. S., Hoffmann, M., Harper, K. L., Sosa, M. S., Werner-Klein, M., Nanduri, L. K., Werno, C., Ehrl, C., Maneck, M*., et al.* (2016). Early dissemination seeds metastasis in breast cancer. Nature *540*, 552-558.
Hu, Z., Ding, J., Ma, Z., Sun, R., Seoane, J. A., Scott Shaffer, J., Suarez, C. J., Berghoff, A. S., Cremolini, C., Falcone, A*., et al.* (2019). Quantitative evidence for early metastatic seeding in colorectal cancer. Nature genetics *51*, 1113-1122.
Iacobuzio-Donahue, C. A., Michael, C., Baez, P., Kappagantula, R., Hooper, J. E., and Hollman, T. J. (2019). Cancer biology as revealed by the research autopsy. Nature reviews Cancer.
Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., Shafi, S., Johnson, D. H., Mitter, R., Rosenthal, R*., et al.* (2017). Tracking the Evolution of Non-Small-Cell Lung Cancer. The New England journal of medicine *376*, 2109-2121.
Jeselsohn, R., Bergholz, J. S., Pun, M., Cornwell, M., Liu, W., Nardone, A., Xiao, T., Li, W., Qiu, X., Buchwalter, G*., et al.* (2018). Allele-Specific Chromatin Recruitment and Therapeutic Vulnerabilities of ESR1 Activating Mutations. Cancer cell *33*, 173-186 e175.
Jimenez-Sanchez, A., Memon, D., Pourpe, S., Veeraraghavan, H., Li, Y., Vargas, H. A., Gill, M. B., Park, K. J., Zivanovic, O., Konner, J*., et al.* (2017). Heterogeneous Tumor-Immune Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. Cell *170*, 927-938 e920.
Joshi, K., Robert de Massy, M., Ismail, M., Reading, J. L., Uddin, I., Woolston, A., Hatipoglu, E., Oakes, T., Rosenthal, R., Peacock, T*., et al.* (2019). Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. Nature medicine.
Juric, D., Castel, P., Griffith, M., Griffith, O. L., Won, H. H., Ellis, H., Ebbesen, S. H., Ainscough, B. J., Ramu, A., Iyer, G*., et al.* (2015). Convergent loss of PTEN leads to clinical resistance to a PI(3)Kalpha inhibitor. Nature *518*, 240-244.
Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. J Clin Invest *119*, 1420-1428.
Kawamata, F., Patch, A. M., Nones, K., Bond, C., McKeone, D., Pearson, S. A., Homma, S., Liu, C., Fennell, L., Dumenil, T*., et al.* (2018). Copy number profiles of paired primary and metastatic colorectal cancers. Oncotarget *9*, 3394-3405.

Klein, C. A. (2009). Parallel progression of primary tumours and metastases. Nature reviews Cancer *9*, 302-312.
Klein, C. A., Blankenstein, T. J., Schmidt-Kittler, O., Petronio, M., Polzer, B., Stoecklein, N. H., and Riethmuller, G. (2002). Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. Lancet *360*, 683-689.
Lambert, A. W., Pattabiraman, D. R., and Weinberg, R. A. (2017). Emerging Biological Principles of Metastasis. Cell *168*, 670-691.
Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495-501.
Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R. J., Sanders, M. A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M.*, et al.* (2019). The landscape of somatic mutation in normal colorectal epithelial cells. Nature *574*, 532-537.
López, S., Lim, E., Huebner, A., Dietzen, M., Mourikis, T., Watkins, T. B. K., Rowan, A., Dewhurst, S. M., Birkbak, N. J., Wilson, G. A.*, et al.* (2019). Whole Genome Doubling mitigates Muller's Ratchet in Cancer Evolution. bioRxiv, 513457.
Makohon-Moore, A. P., Zhang, M., Reiter, J. G., Bozic, I., Allen, B., Kundu, D., Chatterjee, K., Wong, F., Jiao, Y., Kohutek, Z. A.*, et al.* (2017). Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. Nature genetics *49*, 358-366.
Martincorena, I., Fowler, J. C., Wabik, A., Lawson, A. R. J., Abascal, F., Hall, M. W. J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M. R.*, et al.* (2018). Somatic mutant clones colonize the human esophagus with age. Science (New York, NY *362*, 911-917.
Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., and Campbell, P. J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell.
Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M.*, et al.* (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science (New York, NY *348*, 880-886.
Massague, J., and Obenauf, A. C. (2016). Metastatic colonization by circulating tumour cells. Nature *529*, 298-306.
Mateo, J., Chakravarty, D., Dienstmann, R., Jezdic, S., Gonzalez-Perez, A., Lopez-Bigas, N., Ng, C. K. Y., Bedard, P. L., Tortora, G., Douillard, J. Y.*, et al.* (2018). A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). Annals of oncology : official journal of the European Society for Medical Oncology / ESMO *29*, 1895-1902.
McGranahan, N., Favero, F., de Bruin, E. C., Birkbak, N. J., Szallasi, Z., and Swanton, C. (2015). Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med *7*, 283ra254.
McGranahan, N., Rosenthal, R., Hiley, C. T., Rowan, A. J., Watkins, T. B. K., Wilson, G. A., Birkbak, N. J., Veeriah, S., Van Loo, P., Herrero, J.*, et al.* (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell.
McPherson, A., Roth, A., Laks, E., Masud, T., Bashashati, A., Zhang, A. W., Ha, G., Biele, J., Yap, D., Wan, A.*, et al.* (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. Nature genetics *48*, 758-767.
Merenbakh-Lamin, K., Ben-Baruch, N., Yeheskel, A., Dvir, A., Soussan-Gutman, L., Jeselsohn, R., Yelensky, R., Brown, M., Miller, V. A., Sarid, D.*, et al.* (2013). D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. Cancer research *73*, 6856-6864.
Messaoudene, M., Mourikis, T. P., Michels, J., Fu, Y., Bonvalet, M., Lacroix-Trikki, M., Routy, B., Fluckiger, A., Rusakiewicz, S., Roberti, M. P.*, et al.* (2019). T cell Bispecific Antibodies in Node-Positive Breast Cancer: Novel Therapeutic Avenue for MHC class I Loss Variants. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO.

Miller, A. M., Shah, R. H., Pentsova, E. I., Pourmaleki, M., Briggs, S., Distefano, N., Zheng, Y., Skakodub, A., Mehta, S. A., Campos, C.*, et al.* (2019). Tracking tumour evolution in glioma through liquid biopsies of cerebrospinal fluid. Nature *565*, 654-658.
Naxerova, K., Reiter, J. G., Brachtel, E., Lennerz, J. K., van de Wetering, M., Rowan, A., Cai, T., Clevers, H., Swanton, C., Nowak, M. A.*, et al.* (2017). Origins of lymphatic and distant metastases in human colorectal cancer. Science (New York, NY *357*, 55-60.
Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M.*, et al.* (2012). The life history of 21 breast cancers. Cell *149*, 994-1007.
Noorani, A., Goddard, M., Crawte, J., Alexandrov, L. B., Li, X., Secrier, M., Eldridge, M. D., Bower, L., Weaver, J., Lao-Sirieix, P.*, et al.* (2018). Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma. bioRxiv, 454306.
Paget, S. (1989). The distribution of secondary growths in cancer of the breast. 1889. Cancer Metastasis Rev *8*, 98-101.
Peinado, H., Zhang, H., Matei, I. R., Costa-Silva, B., Hoshino, A., Rodrigues, G., Psaila, B., Kaplan, R. N., Bromberg, J. F., Kang, Y.*, et al.* (2017). Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer *17*, 302-317.
Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E.*, et al.* (2017). Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med *9*.
Priestley, P., Baber, J., Lolkema, M. P., Steeghs, N., de Bruijn, E., Shale, C., Duyvesteyn, K., Haidari, S., van Hoeck, A., Onstenk, W.*, et al.* (2019). Pan-cancer whole-genome analyses of metastatic solid tumours. Nature.
Reinert, T., Henriksen, T. V., Christensen, E., Sharma, S., Salari, R., Sethi, H., Knudsen, M., Nordentoft, I., Wu, H. T., Tin, A. S.*, et al.* (2019). Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. JAMA Oncol.
Reiter, J. G., Makohon-Moore, A. P., Gerold, J. M., Heyde, A., Attiyeh, M. A., Kohutek, Z. A., Tokheim, C. J., Brown, A., DeBlasio, R. M., Niyazov, J.*, et al.* (2018). Minimal functional driver gene heterogeneity among untreated metastases. Science (New York, NY *361*, 1033-1037.
Robinson, D. R., Wu, Y. M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V.*, et al.* (2017). Integrative clinical genomics of metastatic cancer. Nature *548*, 297-303.
Rogers, Z. N., McFarland, C. D., Winters, I. P., Seoane, J. A., Brady, J. J., Yoon, S., Curtis, C., Petrov, D. A., and Winslow, M. M. (2018). Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. Nature genetics *50*, 483-486.
Selmecki, A. M., Maruvka, Y. E., Richmond, P. A., Guillet, M., Shoresh, N., Sorenson, A. L., De, S., Kishony, R., Michor, F., Dowell, R., and Pellman, D. (2015). Polyploidy can drive rapid adaptation in yeast. Nature *519*, 349-352.
Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., and Forbes, S. A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nature reviews Cancer *18*, 696-705.
Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The cancer genome. Nature *458*, 719-724.
Turajlic, S., and Swanton, C. (2016). Metastasis as an evolutionary process. Science (New York, NY *352*, 169-175.
Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., Lopez, J. I., Nicol, D., O'Brien, T., Larkin, J., Horswell, S.*, et al.* (2018a). Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. Cell *173*, 581-594 e512.
Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J. I., Watkins, T. B. K., Nicol, D.*, et al.* (2018b). Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. Cell *173*, 595-610 e511.
Ullah, I., Karthik, G. M., Alkodsi, A., Kjallquist, U., Stalhammar, G., Lovrot, J., Martinez, N. F., Lagergren, J., Hautaniemi, S., Hartman, J., and Bergh, J. (2018). Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. The Journal of clinical investigation *128*, 1355-1370.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., and Kinzler, K. W. (2013). Cancer genome landscapes. Science (New York, NY *339*, 1546-1558.
Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R., and Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nature reviews Cancer *17*, 223-238.
Wei, Q., Ye, Z., Zhong, X., Li, L., Wang, C., Myers, R. E., Palazzo, J. P., Fortuna, D., Yan, A., Waldman, S. A*., et al.* (2017). Multiregion whole-exome sequencing of matched primary and metastatic tumors revealed genomic heterogeneity and suggested polyclonal seeding in colorectal cancer metastasis. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO *28*, 2135-2141.
Yates, L. R., Knappskog, S., Wedge, D., Farmery, J. H. R., Gonzalez, S., Martincorena, I., Alexandrov, L. B., Van Loo, P., Haugland, H. K., Lilleng, P. K*., et al.* (2017). Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer cell *32*, 169-184 e167.
Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., Lawrence, M. S., Zhang, C. Z., Wala, J., Mermel, C. H*., et al.* (2013). Pan-cancer patterns of somatic copy number alteration. Nature genetics *45*, 1134-1140.
Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S. M*., et al.* (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nature medicine *23*, 703-713.
Zhang, A. W., McPherson, A., Milne, K., Kroeger, D. R., Hamilton, P. T., Miranda, A., Funnell, T., Little, N., de Souza, C. P. E., Laan, S*., et al.* (2018). Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. Cell *173*, 1755-1769 e1722.
Ziller, M. J., Gu, H., Muller, F., Donaghey, J., Tsai, L. T., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E*., et al.* (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477-481.