How subclonal modelling is changing the metastatic paradigm

Geoff Macintyre[1], Peter Van Loo[2,3], Niall M. Corcoran[5,6], David C. Wedge[4], Florian Markowetz[1,*], Christopher M. Hovens[5,6,*]

1. Cancer Research UK Cambridge Institute, University of Cambridge, United Kingdom
2. The Francis Crick Institute, London, United Kingdom
3. Department of Human Genetics, University of Leuven, Leuven, Belgium
4. Oxford Big Data Institute, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom
5. Department of Surgery, The University of Melbourne, The Royal Melbourne Hospital, Parkville, Australia
6. Australian Prostate Cancer Research Centre at Epworth Hospital, Australia

*** Corresponding authors:**
Professor Christopher M Hovens
Department of Surgery
The University of Melbourne
The Royal Melbourne Hospital, Parkville, Australia
Phone: +(61 3) 9342 7703
Fax: +(61 3) 9035 5511
chovens@unimelb.edu.au

Dr Florian Markowetz
Cancer Research UK Cambridge Institute
University of Cambridge
Cambridge, United Kingdom
Phone: +44 (0) 1223 769 628
Fax: +44 (0) 1223 769 510
florian.markowetz@cruk.cam.ac.uk

Running title: How subclonal modelling is changing the metastatic paradigm

Conflicts of interest: none

Abstract

A concerted effort to sequence matched primary and metastatic tumours is vastly improving our ability to understand metastasis in humans. Compelling evidence has emerged that supports the existence of diverse and surprising metastatic patterns. Enhancing these efforts is a new class of algorithms that facilitate high-resolution subclonal modelling of metastatic spread. Here we summarise how subclonal models of metastasis are influencing the metastatic paradigm.

Introduction

For the vast majority of patients that die from solid malignancies, lethality can be directly traced to the propensity of their tumour cells to metastasize. Paget's seminal *seed and soil* hypothesis proposed that the colonization of distant sites by primary *seed* tumour cells is dependent on a compatible environment in the secondary *soil* site (1). Developments of this central idea over the years has led to the prevailing view that metastases are founded by rare single cells that escape from the primary site. A key advantage of this view is that it provides an explanation for the relative rarity of clinical metastasis formation in the general cancer population.

A body of evidence has subsequently accumulated that supports this model of tumour dissemination. Some of the early works include a study of spontaneously arising lung metastases in mouse models of melanoma, where cells uniquely tagged with random irradiation induced karyotypic markers unequivocally indicated that metastases originated from a single progenitor cell (2). Follow-on experiments showed that when mixtures of two distinct melanoma cell lines were injected intravenously, subsequent lung metastases were derived from only one line and not admixtures of the two cell lines (3).

More recently, models of human metastasis have been updated, especially with regard to timing of spread (4). Largely responsible for this shift is the application of next-generation sequencing to *matched* primary and metastatic samples. By identifying sets of shared and private mutations, sample relatedness can be observed and an approximate evolutionary relationship determined. Studies of human colorectal cancer (5), pancreatic cancer (6), melanoma (7) and neuroblastoma (8) have shown that spread can occur late in the evolution of the primary disease, revealing a *linear* evolutionary relationship between primary and metastasis. Conversely, in renal cancer (9), metastatic progression has been shown to occur early, with both primary tumour and metastasis having private mutations and thus evolving in *parallel*. However, follow-up studies in pancreatic (10) and other cancers (11) show examples of both early and late spread, suggesting that timing patterns are not necessarily tumour specific.

Studies across *multiple* metastases from the same patient have also revealed that asynchronous spread can occur from primary to multiple distant metastatic sites in colorectal cancer (11) as well as seeding from metastatic to secondary metastatic site in a *cascading* manner in prostate cancer (12), ovarian cancer (13) and pancreatic cancer (10).

One limitation of these studies is that the clonal composition of each sample is determined using the presence or absence of private and shared mutations. This type of modelling does not allow estimation of clonal frequencies – vital for accurate evolutionary reconstruction and identification

of more than two clones per sample. In an attempt to adopt a more detailed modelling strategy, algorithms have been developed that model the clonal composition *within* a tumour using mutation variant allele frequencies. These algorithms have vastly improved our ability to model and understand metastatic spread. The first use of such an algorithm appeared in a study of primary breast cancers where it was used to accurately identify the clonal makeup of a tumour and infer the evolutionary history of its clones (genetically distinct populations of tumour cells) (14). Since then, a rapidly developing field has emerged that uses high-coverage exome, capture, amplicon and/or whole-genome tumour sequence data to trace clone lineages and infer phylogenetic relationships within and between lesions from individual patients (15, 16).

A subset of recent studies have used these algorithms to infer the evolutionary relationship of clones in *matched* primary and metastatic samples (17-26), revealing patterns of metastasis only observable using this type of quantitative analysis. A recent review has outlined the implications of these studies on treatment, including a summary of the potential underlying genetic determinants of spread (27). Here we focus specifically on how *subclonal modelling* of multiple samples from the same individual has shaped our understanding of metastasis in humans.

Subclonal modelling of metastasis

By comparing the constituent subclonal mutations between pairs of primary and metastatic samples it is possible to derive the ancestral relationships between tumour *clones* rather than between tumour samples. This type of modelling has allowed confirmation of existing patterns of metastasismetastasic spread at increased (subclonal) resolution, and has yielded new insights into the patterns and timing of tumour cell spread which we articulate below.

Timing of spread

Seeding from an ancestral clone early during disease development (Figure 1a), results in a branched evolution pattern, where primary and metastasis evolve in a “parallel” manner (28). This has been shown at subclonal resolution in two lung cancer cases (22), two glioblastoma cases (26), one ovarian case (17), seven prostate cases (18, 29), as well as in mouse models, where evolutionary analysis of skin cancer demonstrated that the majority of tumours adopt a parallel mode of evolution (30). Much debate exists, however, whether particular tumour types have a dominant mode of evolution in humans. Spread occurring late in the evolution of the primary tumour in a *linear* fashion (Figure 1b) has been observed in one oral cancer case (25), eight melanoma cases (24), and four glioblastoma cases (in these cases from residual tumour cells) (26). While sample sizes across these studies are not yet sufficient to determine whether certain tumour types are enriched for late or early spread, examples of both have been seen in a study of eighty-two patients with brain metastases originating from various primaries (23), as well as across eleven cases of head and neck cancer (19).

Seed composition

All tumour types studied at subclonal resolution mentioned in this review showed at least one example of *monoclonal seeding* where a single clone escapes the primary to found a metastatic deposit (Figure 1c).

New data in mouse models of cancer metastasis have challenged the predominant *monoclonal* model of how metastases are constituted positing that some metastases are comprised of mixtures of distinct tumour clones seeded in a *polyclonal* manner (31-33). Furthermore, it has also been argued that clones present in polyclonal mixtures are not necessarily indifferent to one another, but may actually cooperate to seed a secondary lesion, suggesting that mutual interclonal cooperation between distinct clones exists (34). The evidence for such oncogenic cooperation in different model systems has recently been extensively covered in an excellent review (35).

The key distinguishing feature required to confirm the existence of *polyclonal seeding* in bulk sequencing of human samples is the presence of *subclonal clusters* of mutations across *multiple* tumours from *distinct* locations. A mutation is considered *subclonal* if it appears in only a fraction of the tumour cells in a sample. Sets of mutations appearing subclonally in two or more metastases can arise under two potential scenarios: (1) the same sets of mutations occur independently in each sample; (2) two distinct founder cells containing the sets of mutations spread to each location together. While convergent evolution could give weight to scenario 1, it is extremely unlikely statistically given the sizeable sets of subclonal mutations observed in the studies discussed here. Therefore, scenario 2 can be the only real explanation for these subclonal clusters. It is this reasoning that has allowed the determination of the existence of polyclonal seeding in humans.

Many of the studies discussed here have gone a step beyond subclonal clustering and inferred the evolutionary relationship between clones. This process facilitates finer understanding of polyclonal seeding and begins to help us determine if the polyclonal spread occurs *synchronously* with both cells transiting in unison, or *asynchronously* with multiple waves of spread to the same location. Although evidence is yet to accumulate to unequivocally determine synchronicity, the clonal evolution trees determined from multiple studies tend to favour one or the other.

Synchronous polyclonal seeding is a plausible explanation for the patterns of spread observed in six separate studies across five tumour types: oral, breast, glioblastoma, melanoma and prostate (Figure 1d). In these studies, similar mixes of clones were detected in multiple samples from the same individual: Wood *et al*. reconstructed the clonal evolution of a matched oral primary and metastasis in patient PG030, showing that the same mix of clones was present in both samples (25); Murtaza et al observed two subclonal mutation clusters present at varying frequencies across five distant metastatic sites from a single breast cancer patient (21); two patients (C and E) showed evidence of polyclonal seeding in a study of melanoma (24); two cases of glioblastoma revealed clusters of mutations present at subclonal fractions in both primary and recurrent disease (26) and, two separate studies into prostate cancer revealed multiple cases of polyclonal seeding (18, 29). While it is feasible that the mix of clones observed

across these cases could have arisen asynchronously, evidence seen in studies of circulating tumour cell clusters lends weight to a synchronous model of spread: for example, a recent study of clusters of circulating tumour cells versus single cells showed that cell clusters had up to 50-fold increased metastatic potential compared to single cells (36). Interestingly, however, in a study of 86 brain metastasis cases arising from various primary tumours, no evidence of polyclonal seeding was found (23) even though the authors explicitly searched for it. These differences could be attributed to the metastatic niche, whereby the blood-brain barrier prevented clusters of cells transiting but allowed single cell spread. This suggests that the ability for multiple clones to colonise a site could be heavily dependent on the metastatic niche.

Despite the preference for a *synchronous* model of spread, *asynchronous polyclonal seeding* has been shown to be a more likely explanation for at least two patients from the aforementioned prostate studies (Figure 1d) (18, 29). In patient 177 from Hong *et al.* (29), a combination of unusual mutant allele frequency patterns combined with structural variant allele frequencies lead to the most likely explanation of the polyclonal makeup of a metastasis being an early spread from the primary tumour, followed by a late spread of a further evolved clone (Figure 1c). In patient A32 from Gundem *et al.*, a left supraclavicular lymph node was seeded twice from the primary tumour. In the first wave of metastatic seeding, all 4 of the metastatic sites in this patient were seeded with a particular clone, however in a subsequent round of spread, a second distinct metastasizing clone spread to the left supraclavicular lymph node only and not the other three metastatic sites. These findings raise important questions as to whether some tumour clones act as *pathfinders*' colonizing distant sites, which then act as beacons to attract subsequent waves of metastatic colonization in the nascent metastatic niche. Properties of the metastatic niche itself are also likely to contribute to metastatic subclonal seeding and expansion, as evidenced by patient A32. They also clearly suggest that for some patients, at least, removal of the primary tumour even after distant metastases have already been detected may still be clinically warranted as the primary tumour may continue to serve as an *incubator* of further metastatic tumour cell dissemination. This concept is now supported by a growing body of clinical evidence suggesting that treatment of primary tumours in patients with synchronous metastases can provide clinical benefits, including improvements in overall survival (37-40). Further along these lines, one could postulate that polyclonal seeding may occur more often at terminal disease stages where natural defence mechanisms are strained, facilitating easier colonization by multiple tumour clones.

Seed source

Subclonal modelling of multiregional of multiregional primary prostate tumour samples, allowed Hong *et al.* (29) to precisely pinpoint the clone that gave rise to a distant metastasis (Figure 1d). Furthermore, by defining each clone in the primary, they were able to interrogate its presence in circulating tumour DNA and found that in addition to the (expected) detection of metastatic clones, clones (presumed) exclusive to the primary tumour were also detected, despite the primary tumour being removed two years prior. These clones had not seeded any clinically obvious metastases, strongly implying that all clones had colonised distant sites, some occult.

As well as seeding from the primary, cells from one metastasis can seed another metastasis, resulting in what is known as an evolutionary *cascade* (Figure 1d). This phenomenon has been seen at subclonal resolution from lymph node to distant metastasis in mouse models of skin cancer (30), single cases of human breast cancer (21) and melanoma (24), and multiple prostate cases (18, 29). In one of these prostate cases, cross-metastatic site seeding appeared to occur directly in response to the onset of targeted treatment, with marked remodelling of the original subclonal composition at an iliac crest metastatic site within 12 weeks of the patient starting androgen deprivation therapy. Similar subclonal remodelling has also been shown in response to chemotherapy in ovarian cancer (17) and leukemia (41).

Detecting polyclonal seeding

Patterns of polyclonal seeding can only be detected using algorithms that identify the subclonal makeup of multiple tumour samples from a given patient (42-46). While there are a number of different computational techniques for inferring subclonal structure, the majority of studies covered in this review have used a statistical clustering algorithm known as a Bayesian Dirichlet Mixture Model. Therefore, to illustrate how polyclonal seeding is detected, we adapt an example from Gundem *et al.* (18), see Figure 2. We look at two samples from patient A22, a bladder metastasis (G) and a pelvic lymph node metastasis (H). Firstly, using copy-number, tumour purity, and tumour ploidy, the mutant allele fraction of each mutation is converted to the fraction of tumour cells harbouring the mutation, represented as black dots in Fig. 2a, also known as the cancer cell fraction (for conversion details see Nik-Zainal *et al.* (14)). A Bayesian Dirichlet Mixture Model is then used to group mutations into clusters based on their frequencies in both samples (red shading, Fig. 2a). These clusters subsequently help define the distinct populations of cells that arose from clonal expansions during the evolution of the tumour. The cluster of mutations present in all tumour cells in both samples represents the founding clone (dark blue circle Fig. 2a). Clusters of mutations that are in tumour cells across both samples, represent founding cells of the metastases (dark blue and purple circles, Fig. 2a). Clusters that are unique to one of the two samples represent the clones that are emerging at each site (orange, light blue, and green circles, Fig. 2a, for simplicity two clones belonging to the same metastasis with the same ancestor are coloured green). The frequencies of the clusters combined with the pigeon-hole principle (14) can then be used reconstruct the most likely clonal evolution tree (Fig. 2b). As the purple cluster is present at subclonal frequencies in both samples, both cells from this clone and cells from the ancestral clone (dark blue circle) must have founded the metastatic site G in a *polyclonal* manner. The resulting clonal makeup can be represented by colour coded nested ovals reflecting the evolutionary relationship between clones (Fig. 2c, white space represents normal cell admixture). Finally an overall schematic of the clonal spread can be derived (Fig. 2d).

Discussion

The application of whole-genome sequencing and new computational methods to multiple metastatic samples has enabled exciting insights into the process of metastatic seeding, the presence of polyclonal seeding being the most significant. However, there are now many open questions around the underlying mechanisms behind this observation. Do clones transit as

polyclonal clusters or as single cells? If as clusters, are they cooperating within the cluster to survive blood transit and eventual seeding of distant sites? Do they form clusters within the blood or within the primary tumour site?

Some headway has been made through animal models of breast cancer, with a recent study showing that clusters of tumour cells have a much higher capacity to induce metastasis formation, despite being present at much lower frequency than single cells (36). Furthermore, tumour cell clusters did not form in the blood but rather appeared to form within the site of tumour cell inoculation. Another important question is whether and to what extent specific clones may be involved in establishing pre-metastatic niches conducive to subsequent waves of tumour cell inoculation. Evidence in favour of this is the observed extracellular vesicles secreted by tumour cells that can be sequestered by bone-marrow derived cells, enhancing their capacity to form a metastatic niche (47-49). Following-on, specific clones might also be able to modify the metastatic potential of surrounding less metastatic clones through transfer of metastatic extracellular vesicles, as has been recently demonstrated in animal models of breast cancer (50). Further application of subclonal modelling to this question in humans is likely to yield greater insight.

The polyclonal seeding observed in multiple sites across the cases discussed in this review may be indicative of intimate crosstalk occurring between metastatic clones and suggests that targeted disruption of these interactions might be productive in obstructing metastasis formation. Certain patterns of metastasis may be targeted by particular treatment regimes. However, these insights are currently limited by the availability of samples, so predicting which pattern is likely to occur in a given tumour subtype is not yet feasible. Further studies incorporating the subclonal analysis of multiple primary and multiple metastases from individual patients are required to not only answer fundamental questions as to how tumour cells metastasize but also provide insights in to how this process may be disrupted.

Acknowledgements

This work was supported by: a Federal grant for the Australian Prostate Cancer Research Centres and by NHMRC grants 1047581 and 1104010 to CMH and NMC; the Ovarian Cancer Action [BriTROC project grant: G.M]; the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001202), the UK Medical Research Council (FC001202), and the Wellcome Trust (FC001202), PVL; The University of Cambridge, Cancer Research UK, Hutchinson Whampoa Limited, CRUK core grant C14303/A17197 (CRUK CI institute core grant, FM and GM) and A19274 (FM lab grant). DCW is funded by the Li Ka Shing foundation.

References

1. Paget S. The distribution of secondary growths in cancer of the breast. . Lancet 1889; 1:571–3.
2. Talmadge JE, Wolman SR, Fidler IJ. Evidence for the clonal origin of spontaneous metastases. Science. 1982;217(4557):361-3.
3. Fidler IJ, Talmadge JE. Evidence that intravenously derived murine pulmonary melanoma metastases can originate from the expansion of a single tumor cell. Cancer Res. 1986;46(10):5167-71.

4. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. Nat Rev Clin Oncol. 2015;12(5):258-72. doi: 10.1038/nrclinonc.2014.238..
5. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, et al. Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci U S A. 2008;105(11):4283-8. doi: 10.1073/pnas.0712345105.
6. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010;467(7319):1114-7. doi: 10.1038/nature09515.
7. Turajlic S, Furney SJ, Lambros MB, Mitsopoulos C, Kozarewa I, Geyer FC, et al. Whole genome sequencing of matched primary and metastatic acral melanomas. Genome Res. 2012;22(2):196-207. doi: 10.1101/gr.125591.111.
8. Schramm A, Koster J, Assenov Y, Althoff K, Peifer M, Mahlow E, et al. Mutational dynamics between primary and relapse neuroblastomas. Nat Genet. 2015;47(8):872-7. doi: 10.1038/ng.3349.
9. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883-92. doi: 10.1056/NEJMoa1113205.
10. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010;467(7319):1109-13. doi: 10.1038/nature09460.
11. Zhao ZM, Zhao B, Bai Y, Iamarino A, Gaffney SG, Schlessinger J, et al. Early and multiple origins of metastatic lineages within primary tumors. Proc Natl Acad Sci U S A. 2016;113(8):2140-5. doi: 10.1073/pnas.1525677113.
12. Haffner MC, Mosbruger T, Esopi DM, Fedor H, Heaphy CM, Walker DA, et al. Tracking the clonal origin of lethal prostate cancer. J Clin Invest. 2013;123(11):4918-22. doi: 10.1172/JCI70354.
13. Schwarz RF, Ng CK, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med. 2015;12(2):e1001789. doi: 10.1371/journal.pmed.1001789.
14. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994-1007. doi: 10.1016/j.cell.2012.04.023.
15. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014;346(6206):251-6. doi: 10.1126/science.1253462.
16. Greaves M. Evolutionary determinants of cancer. Cancer Discov. 2015;5(8):806-20. doi: 10.1158/2159-8290.CD-15-0439.
17. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J Pathol. 2013;231(1):21-34. doi: 10.1002/path.4230.
18. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JM, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015;520(7547):353-7. doi: 10.1038/nature14347.
19. Hedberg ML, Goh G, Chiosea SI, Bauman JE, Freilino ML, Zeng Y, et al. Genetic landscape of metastatic and recurrent head and neck squamous cell carcinoma. J Clin Invest. 2016;126(1):169-80. doi: 10.1172/JCI82066.
20. Hong WS, Shpak M, Townsend JP. Inferring the origin of metastases from cancer phylogenies. Cancer Res. 2015. doi: 10.1158/0008-5472.CAN-15-1889.
21. Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nat Commun. 2015;6:8760. doi: 10.1038/ncomms9760.

22. Paik PK, Shen R, Won H, Rekhtman N, Wang L, Sima CS, et al. Next-Generation Sequencing of Stage IV Squamous Cell Lung Cancers Reveals an Association of PI3K Aberrations and Evidence of Clonal Heterogeneity in Patients with Brain Metastases. Cancer Discov. 2015;5(6):610-21. doi: 10.1158/2159-8290.CD-14-1129.
23. Brastianos PK, Carter SL, Santagata S, Cahill DP, Taylor-Weiner A, Jones RT, et al. Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. Cancer Discov. 2015;5(11):1164-77. doi: 10.1158/2159-8290.CD-15-0369.
24. Sanborn JZ, Chung J, Purdom E, Wang NJ, Kakavand H, Wilmott JS, et al. Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination. Proc Natl Acad Sci U S A. 2015;112(35):10995-1000. doi: 10.1073/pnas.1508074112.
25. Wood HM, Conway C, Daly C, Chalkley R, Berri S, Senguven B, et al. The clonal relationships between pre-cancer and cancer revealed by ultra-deep sequencing. J Pathol. 2015;237(3):296-306. doi: 10.1002/path.4576.
26. Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. 2015;25(3):316-27. doi: 10.1101/gr.180612.114.
27. Turajlic S, Swanton C. Metastasis as an evolutionary process. Science. 2016;352(6282):169-75. doi: 10.1126/science.aaf2784.
28. Klein CA. Parallel progression of primary tumours and metastases. Nat Rev Cancer. 2009;9(4):302-12. doi: 10.1038/nrc2627.
29. Hong MK, Macintyre G, Wedge DC, Van Loo P, Patel K, Lunke S, et al. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. Nat Commun. 2015;6:6605. doi: 10.1038/ncomms7605.
30. McCreery MQ, Halliwill KD, Chin D, Delrosario R, Hirst G, Vuong P, et al. Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers. Nat Med. 2015;21(12):1514-20. doi: 10.1038/nm.3979.
31. McFadden DG, Papagiannakopoulos T, Taylor-Weiner A, Stewart C, Carter SL, Cibulskis K, et al. Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell. 2014;156(6):1298-311. doi: 10.1016/j.cell.2014.02.031.
32. Cleary AS, Leonard TL, Gestl SA, Gunther EJ. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. Nature. 2014;508(7494):113-7. doi: 10.1038/nature13187.
33. Letouze E, Allory Y, Bollet MA, Radvanyi F, Guyon F. Analysis of the copy number profiles of several tumor samples from the same patient reveals the successive steps in tumorigenesis. Genome Biol. 2010;11(7):R76. doi: 10.1186/gb-2010-11-7-r76.
34. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature. 2014;514(7520):54-8. doi: 10.1038/nature13556.
35. Tabassum DP, Polyak K. Tumorigenesis: it takes a village. Nat Rev Cancer. 2015;15(8):473-83. doi: 10.1038/nrc3971.
36. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell. 2014;158(5):1110-22. doi: 10.1016/j.cell.2014.07.013.
37. Loppenberg B, Dalela D, Karabon P, Sood A, Sammon JD, Meyer CP, et al. The Impact of Local Treatment on Overall Survival in Patients with Metastatic Prostate Cancer on Diagnosis: A National Cancer Data Base Analysis. Eur Urol. 2016. doi: 10.1016/j.eururo.2016.04.031.
38. Culp SH, Schellhammer PF, Williams MB. Might men diagnosed with metastatic prostate cancer benefit from definitive treatment of the primary tumor? A SEER-based study. Eur Urol. 2014;65(6):1058-66. doi: 10.1016/j.eururo.2013.11.012.

39. Fleckenstein J, Petroff A, Schafers HJ, Wehler T, Schope J, Rube C. Long-term outcomes in radically treated synchronous vs. metachronous oligometastatic non-small-cell lung cancer. BMC Cancer. 2016;16(1):348. doi: 10.1186/s12885-016-2379-x.
40. Griffioen GH, Toguri D, Dahele M, Warner A, de Haan PF, Rodrigues GB, et al. Radical treatment of synchronous oligometastatic non-small cell lung carcinoma (NSCLC): patient outcomes and prognostic factors. Lung Cancer. 2013;82(1):95-102. doi: 10.1016/j.lungcan.2013.07.023.
41. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152(4):714-26. doi: 10.1016/j.cell.2013.01.019.
42. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat Commun. 2014;5:2997. doi: 10.1038/ncomms3997.
43. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome Biol. 2015;16:35. doi: 10.1186/s13059-015-0602-8.
44. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods. 2014;11(4):396-8. doi: 10.1038/nmeth.2883.
45. Zare H, Wang J, Hu A, Weber K, Smith J, Nickerson D, et al. Inferring clonal composition from multiple sections of a breast cancer. PLoS computational biology. 2014;10(7):e1003703. Epub 2014/07/11. doi: 10.1371/journal.pcbi.1003703.
46. El-Kebir M, Oesper L, Acheson-Field H, Raphael BJ. Reconstruction of clonal trees and tumor composition from multi-sample sequencing data. Bioinformatics (Oxford, England). 2015;31(12):i62-70. Epub 2015/06/15. doi: 10.1093/bioinformatics/btv261.
47. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015;17(6):816-26. doi: 10.1038/ncb3169.
48. Peinado H, Aleckovic M, Lavotshkin S, Matei I, Costa-Silva B, Moreno-Bueno G, et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat Med. 2012;18(6):883-91. doi: 10.1038/nm.2753.
49. Zhang Y, Wang XF. A niche role for cancer exosomes in metastasis. Nat Cell Biol. 2015;17(6):709-11. doi: 10.1038/ncb3181.
50. Zomer A, Maynard C, Verweij FJ, Kamermans A, Schafer R, Beerling E, et al. In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. Cell. 2015;161(5):1046-57. doi: 10.1016/j.cell.2015.04.042.

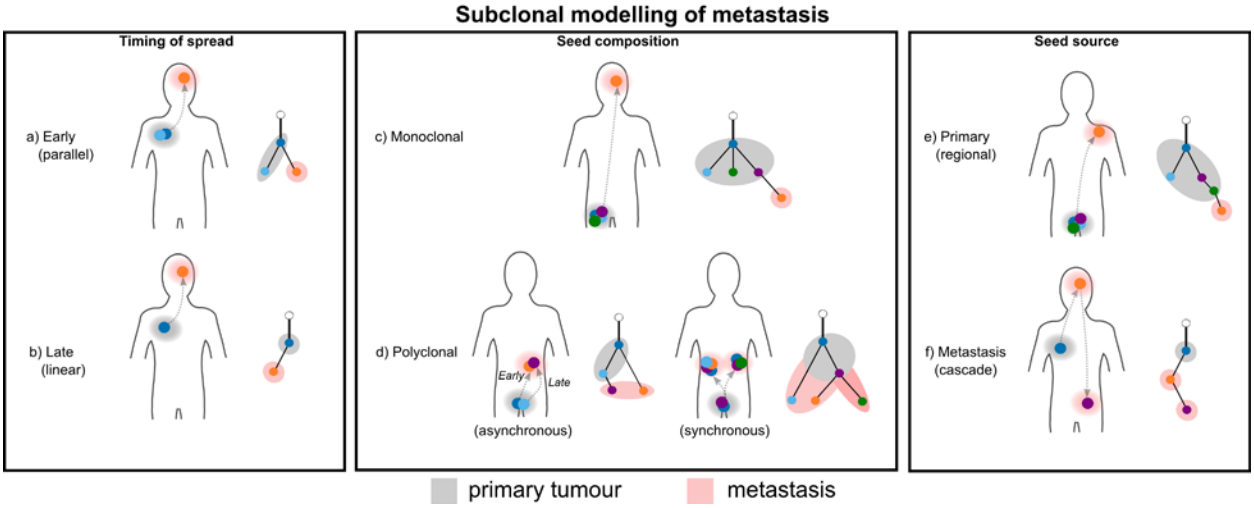

Figure 1 – Examples of subclonal modelling of metastasis in human tumours. This figure summarises the findings of several recent studies that sequenced the DNA of matched primary and metastatic tissue from lung, breast, melanoma, and prostate cancer patients. In each case, mutations were used to infer the evolutionary history of the disease as a *clonal* tree, where each node in the tree represents a genetically distinct population of cells, or clone. A schematic of the clonal spread in each patient is shown, along with a simplified version of the clonal evolution tree reported in the original studies. a) A lung cancer patient (PP4) from the Paik *et al. (22)* that showed early metastasis to the brain, resulting in a branched clonal tree with parallel evolution of both primary and metastasis. b) A lung cancer patient (308) from Brastianos *et al. (23)* that showed late spread to the brain resulting in a linear clonal tree. c) A melanoma patient (H) from Sanborn *et al.* that showed a distant brain metastasis seeded from a single clone present in a leg lesion. d) left – a prostate cancer patient (177) from Hong *et al. (29)* showing an early and late spread of two clones to the ilium in an asynchronous polyclonal manner. Right – a prostate cancer patient (A32) from Gundem *et al. (18)* showing the spread of two clones to two separate metastatic locations in a synchronous manner. e) A prostate cancer patient (299) from Hong *et al. (29)* with multiple regions of the primary sequenced showing a single, extraprostatic clone as the source of the shoulder metastasis. f) A breast cancer patient from Murtaza *et al.* (21) showing a cascade of spread from primary to brain metastasis, then brain to ovary.

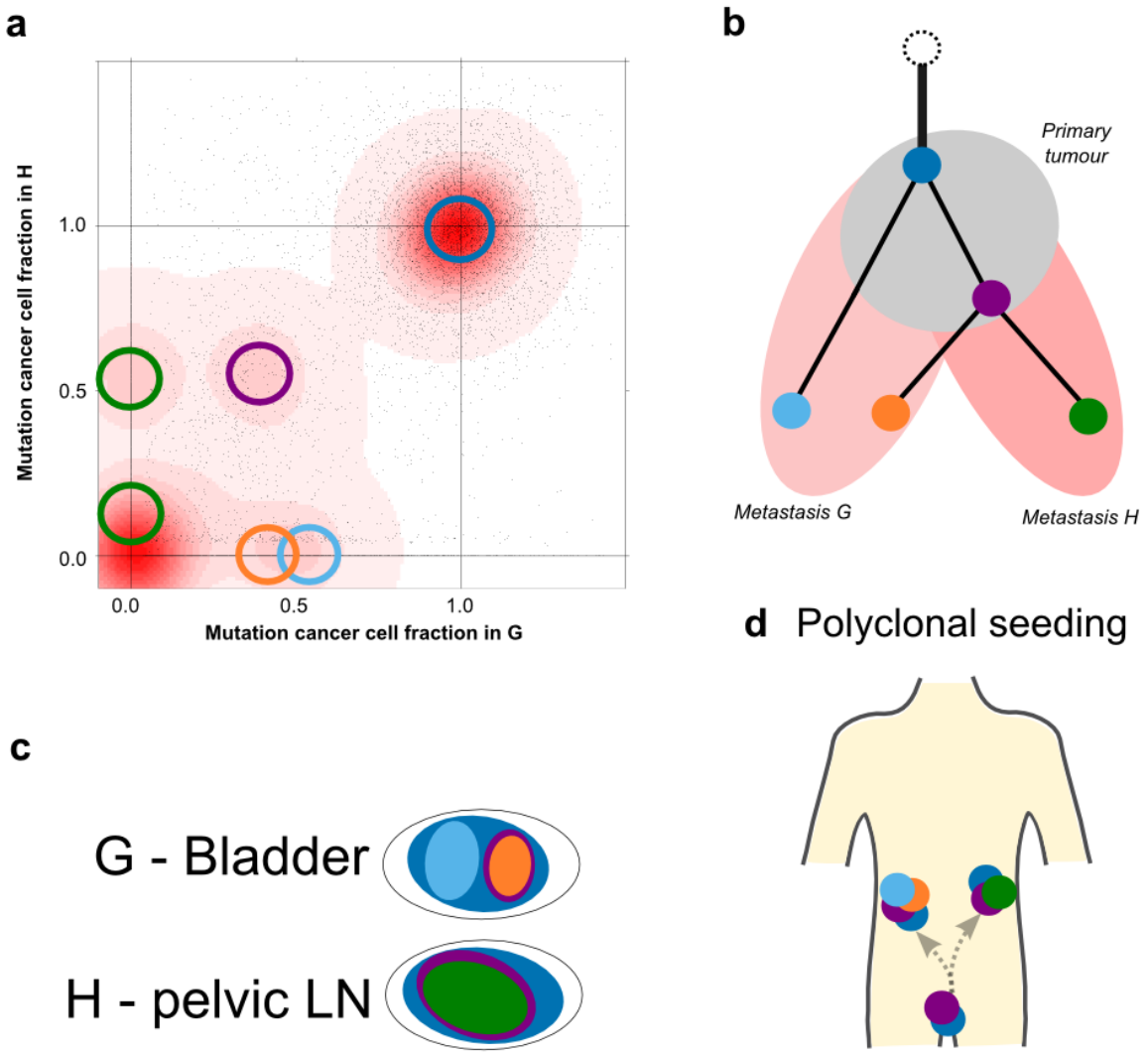

Figure 2 – Detecting polyclonal seeding. This figure illustrates how polyclonal seeding can be detected using a Dirichlet Mixture modelling approach (see main text for a detailed description). a) A density plot showing the cancer cell fractions of mutations (black dots) in two metastatic samples of a prostate patient. The red shading represents the posterior probability of a cluster as determined using a Dirichlet Mixture Model. The coloured circles show the defined mutation clusters. b) A clone tree where each node represents a tumour clone with a distinct genotype. The shaded ellipses show the clone membership for the samples from this patient. c) An "easter egg" plot showing clone membership and ancestry as a series of embedded ellipses. The size of the ellipses is approximately proportional to the number of cells in the sample from that clone. d) A schematic showing the clonal composition of the primary tumour and metastases.