DECODING THE PATTERNS OF CLONAL DYNAMICS IN BREAST CANCER METASTASIS USING SINGLE-CELL SEQUENCING IN PATIENT-DERIVED XENOGRAFT MODELS

by

Hakwoo Lee

MD, Yonsei University, Wonju College of Medicine, South Korea, 2007

M.Sc., Yonsei University, South Korea, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Decoding the patterns of clonal dynamics in breast cancer metastasis using single-cell sequencing in patient-derived xenograft models

submitted by Hakwoo Lee in partial fulfillment of the requirements for

the degree of MASTER OF SCIENCE

in Pathology and Laboratory Medicine

Examining Committee:

Samuel Aparicio, Pathology and Laboratory Medicine, UBC
Supervisor

Michael Underhill, Pathology and Laboratory Medicine, UBC
Supervisory Committee Member

Christian Steidl, Pathology and Laboratory Medicine, UBC
Additional Examiner

Peter Watson, Pathology and Laboratory Medicine, UBC
Additional Examiner

Additional Supervisory Committee Members:

Will Lockwood, Pathology and Laboratory Medicine, UBC
Supervisory Committee Member

Poul Sorensen, Pathology and Laboratory Medicine, UBC
Supervisory Committee Member

Abstract

Introduction: Selection and evolution of tumour cells occurs during cancer progression and metastasis. Understanding the mechanism of clonal dynamics and evolution is an important key to develop new therapeutic strategies for cancer metastasis. This thesis summarizes the development of breast cancer patient-derived xenograft (PDX) metastasis models and measurement of clonal dynamics during metastasis by identifying genomic changes in single cells from primary tumours and metastases.

Methods: Tumour cells from untreated primary breast cancer patients were used to develop PDX tumours in immunodeficient mice. Tumours were removed when they reached maximum allowed endpoint size (1,000mm^3) during a survival surgery and mice were monitored for metastasis. Immunohistochemical (IHC) staining was performed with 10 markers to characterize tumours. Single cell whole-genome sequencing (scWGS) was used to analyse primary and metastatic tumour cells and copy number alterations (CNAs) were identified which allow us to cluster cells and identify clones. Phylogenetic analysis was performed to identify clonal relationship between primary and metastatic tumour cells.

Results: Nine different triple-negative breast cancer PDX lines were tested and 5 developed metastases (SA919, SA535, SA1142, SA605, SA609). We observed that protein marker expression was similar between primary tumour and metastases. Metastatic sites were reproducible over multiple passages in both SA919 and SA535. We also observed that metastatic potential increased with passage number in SA919 while 4 different passages of SA535 showed similar metastases development. From single cell analysis, we observed that the ability to metastasize of primary tumour increases with passage number due to the evolution of clonal population in SA919 and metastatic potential is a property distributed across CNA-defined

clones in both SA919 and SA535. We also observed that metastasis to specific anatomical site was not associated with genomic clones and CNA induced genotype and LOH are potential factors that can affect metastatic potential of clones.

Conclusion: We established breast cancer metastasis mouse model using patient-derived tissues and were able to capture different patterns of metastases in several PDXs. From the two transplant systems studied in detail, we observed metastatic potential was distributed across many genomic clones and CNAs have potential impact on metastatic potential of tumour cells.

Lay Summary

Cancer is the leading cause of death and the most of cancer patients die from spread of cancer cells to other parts of the body, which is called metastasis. However, the mechanism of metastasis is still poorly understood and therefore, it is important to understand the biology and behavior of cancer cells during metastasis. We developed breast cancer metastasis animal models using tissues from patients' tumour implanted into mice and analysed genome or DNA information of individual cancer cells from tumours and metastases developed in mice. We observed the difference of cell population between cells from primary tumours and metastases based on their genomic structure and revealed that genomic properties of tumour cells affected the competitiveness in metastases. Further investigation using various tumour materials will shed more light to understand the mechanism of cancer metastasis.

Preface

This thesis is an original intellectual product of the author, Hakwoo Lee. I developed the hypothesis, the aims of this study and the experimental design under the guidance from Dr. Samuel Aparicio.

I have received training in surgical technique, anesthesia, euthanasia, and proper handling of equipment from Animal Resource Centre of BC Cancer Research Institute. I performed all mammary fat pad transplant, primary tumour removal, monitoring of mice (with assistance of animal technician), palpation of tumour, detection of metastasis, euthanasia, collection of distant organs and processing tumour tissues from mice.

PET-CT scan was performed by technicians in Dr. Francois Benard lab at BC Cancer Research Institute.

Single-cell spotting with DLP+ was performed by Jazmine Brimhall, Justina Biele and Biexi Wang (lab technicians).

Tissue microarray (TMA) construction was done by Brian Chang and Dr. Takako Kono.

Immunohistochemistry scoring for antibody staining on TMA was done by Dr. Takako Kono.

Bioinformatics analysis of genomic data from single cell sequencing was performed by Dr. Daniel Lai, Dr. Sohrab Salehi, Michael Yeun, Patricia Ye and Dr. Hoa Tran.

The University of British Columbia Animal Care Committee approved the animal studies under the ethics certificate H16-01625 and H20-00170.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BAF	B-allele frequency
CN	Copy number
CNA	Copy number alteration
CNV	Copy number variance
DLP+	Direct library preparation
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
H&E	Hematoxylin and eosin
HER2	Human epidermal growth factor 2
IHC	Immunohistochemistry
ITH	Intra-tumoural heterogeneity
LOH	Loss of heterozygosity
MAP	Maximum a posteriori
MFP	Mammary fat pad
OSEM	Ordered subset expectation maximization
PDX	Patient derived xenograft
PET-CT	Positron emission tomography – computed tomography
PR	Progesterone receptor
scWGS	Single-cell whole genome sequencing
SERM	Selective estrogen receptor modulator
SNV	Single nucleotide variant
SUV	Standardized uptake values

TMA	Tissue microarray
TNBC	Triple negative breast cancer
WGS	Whole genome sequencing

Acknowledgements

First and foremost, I would like to thank Dr. Samuel Aparicio for accepting me as a graduate student and giving me the opportunity and supports to learn molecular oncology in one of the top laboratories in the field. I am grateful for his guidance and support for both academic and personal aspects. I thank my supervisory committee members for their precious input and feedback on my thesis. I would also like to thank all the members of Aparicio lab for their kindness and all the support, making my experience in the lab enjoyable. My special thanks to Farhia Kabeer and Dr. Ciara O'Flanagan for our valuable discussion over the project. My sincere gratitude to Dr. Hoa Tran for her tremendous work on bioinformatics analyses.

I would also like to express my sincere gratitude to Dr. Joon Jeong who lead me to the breast cancer field and taught me everything about breast cancer from diagnosis to treatment, both surgical and medical, as well as the basics of patient care.

I should acknowledge the Severance Research Initiative scholarship from Yonsei University which made it possible for me to study at UBC. I also thank Laurel L. Watters Research Fellowship for their support on my study.

I've always appreciated the love and supports from my parents and parents-in-law. Most importantly, my wife Christine Choi, my daughter Grace and my son Ben, I couldn't have done anything without your existence.

Dedication

This work is dedicated to my dearest wife Christine, and the reasons of my life, Grace and Ben.

Chapter 1: Introduction

1.1 Cancer

Cancer is a group of diseases that involve abnormal proliferation of cells and invasion or spread to other parts of the body. Accumulated knowledge from extensive cancer research in recent decades allowed a greater understanding of the underlying mechanism of cancer development and led to development of new therapeutic strategies. Cancer cells acquire distinct capabilities, known as Hallmarks of Cancer, during their development from normal cells to achieve the traits that enable them to become tumorigenic and malignant [1, 2]. The Hallmarks of Cancer include self-sufficiency in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis, abnormal metabolic pathways, evading the immune system, genome instability and inflammation.

Despite advanced understanding of cancer, it is still the leading cause of death in Canada and worldwide (Figure 1-1 A). In 2018, 18.1 million new cases were developed and 9.6 million deaths occurred. One in 5 men and one in 6 women worldwide develop cancer and one in 8 men and one in 11 women die from the disease during their lifetime [3]. In Canada, nearly 1 in 2 Canadians is expected to be diagnosed with cancer during their lifetime and 1 in 4 Canadians is expected to die from Cancer [4] (Figure 1-1 B).

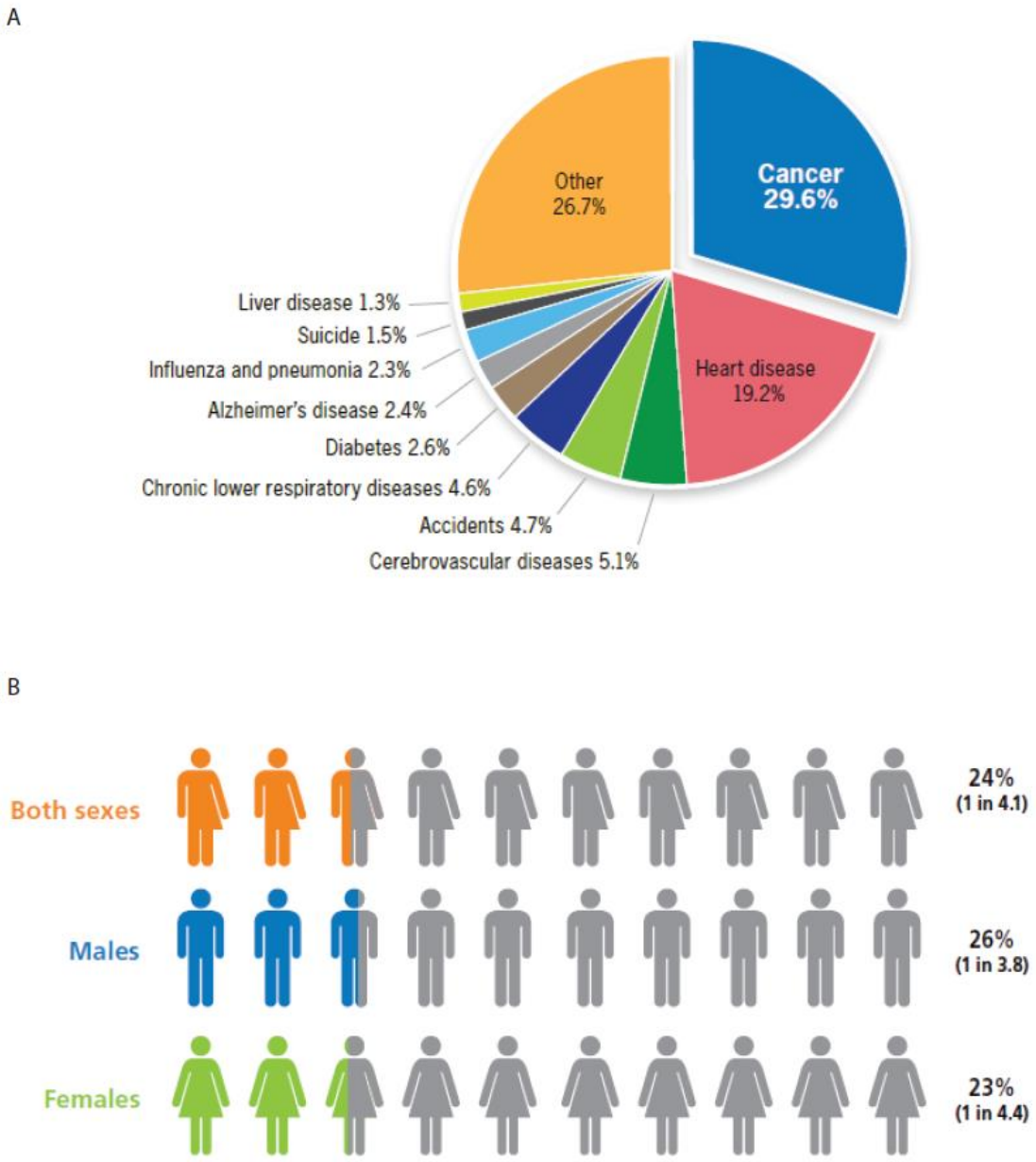


Figure 1-1 Cancer statistics.

A. Proportion of death due to cancer and other causes, Canada, 2016, B. Lifetime probability of dying from cancer, Canada, 2015. Adapted from Canadian Cancer Statistics 2019.

1.1.1 Breast cancer

Breast cancer is the most common cancer in women and also the most common cause of cancer death worldwide [3]. In Canada, breast cancer is the most commonly diagnosed cancer in female, accounting for 25% of all new cases. One in 8 females is expected to be diagnosed with breast cancer in their lifetime [4].

1.1.2 Tumour heterogeneity

Human tumours are shown to have substantial heterogeneity [5, 6]. Tumour heterogeneity consists of inter-tumoural heterogeneity and intra-tumoural heterogeneity. Inter-tumoural heterogeneity refers to tumours from different patients that show distinct genotype and phenotype. Intra-tumoural heterogeneity refers to phenotypic and genomic variations within a tumour across time and/or space. At the time of diagnosis, many of tumours present heterogeneity in morphological and physiological features [7]. Cancer cells within the same tumour also display phenotypic and functional heterogeneity because of genetic and non-genetic sources.

1.1.2.1 Inter-tumoural heterogeneity in breast cancer.

Clinically, breast cancer can be divided into three subtypes: 1) hormone receptor, either estrogen receptor (ER) or progesterone receptor (PR) positive (HR+), 2) amplification and/or overexpression of ERBB2 (human epidermal growth factor receptor, HER2+), and 3) triple negative breast cancer (TNBC) which does not express any of these three markers [8, 9]. While targeted therapies for HR+ (e.g. Selective Estrogen Receptor Modulators (SERMs), aromatase inhibitor (AI)) and HER2+ (e.g. trastuzumab) provide significant survival benefit to patients [10, 11], TNBC does not have druggable targets and there are currently no approved and/or effective targeted therapies for TNBC.

Breast cancer can also be classified into molecular subtypes using RNA expression analysis. In a landmark study, five intrinsic molecular subtypes of breast cancer were proposed: luminal A (ER+, with lower proliferation signatures), luminal B (ER+, with elevated proliferation signatures), basal-like (predominantly TNBC), HER2-enriched, and normal breast-like [12], and claudin-low subtype is added in a subsequent study [13]. Dawson et al. identified 10 integrative clusters of breast cancer that show distinct molecular and clinical features using both genomic and transcriptomic data [14].

1.1.2.2 Intra-tumoural heterogeneity (ITH) in breast cancer

The clinical evolution of tumour characteristics within patients and the variability in responses between patients has brought a new appreciation of intra-tumoural cell heterogeneity as an underpinning factor. For example, breast cancer can be diagnosed as ER+ even when only 1% of cells express detectable levels of ER in immunohistochemistry (IHC) which means rest of cells consisting a tumour may not be ER+. Similarly, there can be variation of HER2 expression among tumour cells of HER2+ breast cancer. Moreover, TNBC has been also shown to exhibit broad spectrum of ITH based on genomic analysis of patient samples [15]. Heterogeneous tumour cells within one tumour exhibit various traits related to tumorigenesis, such as angiogenic, invasive and metastatic potential, and therapeutic response [7]. Thus, ITH can lead to inconsistent response to treatment and is a significant challenge to make effective treatment decisions [16].

1.1.3 Triple-negative breast cancer (TNBC)

TNBC comprises approximately 11-13% of all breast cancers [17, 18]. TNBC lacks both hormone receptors and HER2 and is the subtype that shows the worst prognosis with high rate of recurrence and metastasis [19]. However, recent studies are showing that TNBC is a

heterogeneous group of different diseases and several attempts to classify them into further subtypes have been done [20–22]. Lehmann et al. analyzed gene expression profiles of TNBCs and identified 6 TNBC subtypes with specific drug targets [20]. Similarly, Burstein et al. identified 4 distinct molecular TNBC subtypes using mRNA profiling [21].

However, transcript expression subtyping is not routinely used in clinical diagnosis in TNBC. IHC staining is a basic method established in the pathological diagnosis of breast cancer and a number of proteins assessed by IHC can be correlated with molecular sub-types of breast cancer. Basal-like breast cancer, which is characterized by high histologic grade, high mitotic index and low differentiation, consists 75-80% of TNBC. The common morphologic features of basal-like breast cancers include marked cellular pleomorphism, high nuclear-cytoplasmic ratio, vesicular chromatin, prominent nucleoli, lack of tubule formation, high mitotic index, frequent apoptotic cells, scant stromal content, pushing invasion borders, central geographic or comedo-type necrosis and stromal lymphocytic infiltrate. It also expresses both EGFR and cyokeratins (CK5, 6, 14) [23]. Indeed, basal-like breast cancer can be identified by a variety of IHC markers, such as triple-negative markers (ER, PR, HER2), positive expression of cytokeratins (CK 5, 6, 7, 14), EGFR, SMA, P-cadherin, p53, c-kit, fascin, moesin, vimentin, nestin, laminin and Ki-67 [24, 25]. INPP4B, which is a tumour suppressor, has been shown to be lost in basal-like breast cancer [26]. Unfortunately, there is no single marker that can identify basal-like breast cancer and many panels of markers have been suggested [25]. The panel of four markers (ER, EGFR, HER2 and CK5/6) suggested by Nielsen et al. shows 100% specificity and about 76% sensitivity. Despite many efforts to classify TNBC, there is no standard classification so far nor effective targeted therapy for TNBC.

1.2 Metastasis

Metastasis is spread of tumour cells to distant sites with the development of secondary tumour. Despite the recent advances in cancer research, metastasis is still responsible for >90% of cancer related mortality [27, 28] and there is no cure for metastasis once it has developed.

1.2.1 Metastatic process

Metastasis progression can be simplified into several sequential steps which has helped to understand general biological properties required for cancer cells to develop metastatic disease [29]. The basic steps of metastasis include invasion into surrounding tissue, intravasation, survival in circulation, extravasation and colonization. Metastasis is considered as a highly inefficient process as most cancer cells released from primary tumour will die and only a small number of cells succeed in developing metastatic lesion [30]. A discovery by Stephan Paget suggested that metastasis is an interactive process between 'seeds' (or the cancer cells) and the 'soil' (or the host microenvironment) [31]. Studies have shown that primary tumour preconditions specific host organ sites to form pre-metastatic niches which are the microenvironments favorable for tumour cells to colonize and grow [32]. The changes in host microenvironments involve increased vascular leakiness, remodeling of stroma and extracellular matrix, followed by systemic effects on the immune system [33-35]. In terms of intrinsic factors of cancer cells, several metastasis-related genes [36-38], copy number alterations (CNAs) of the genome [39] as well as alteration of cellular metabolism [40] have been identified to affect the metastatic process. However, underlying mechanism of how those disseminated cells survive and colonize in distant organs is largely unknown.

1.2.2 Metastasis in breast cancer

The common sites for metastatic spread in breast cancer are bone, lung and liver [41]. Traditional prognostic markers for breast cancer metastasis include tumour size, axillary lymph node status, histological grade, hormone receptor and HER2 status [27]. Several metastasis related genes have been also identified over the years. For example, acquiring mutations in ANGPTL4 in breast tumours enhances metastatic potential to lungs [36], whereas the expression of ST6GALNAC5 enhances metastasis to brain [37]. There are several other genes identified to mediate metastasis to lung [38], or brain [37].

However, despite the advancement of our understanding of metastasis related genes, there is no method to predict metastatic potential of primary tumour at the time of diagnosis. One of the limitations of gene expression studies can be attributed to the reversible cancer cell plasticity [42]. Cancer related gene expression from a single time point biopsy can only describe the snapshot of cell state at certain time point during the evolution of cancer cells. Thus, it is important to understand the evolutionary process during cancer progression.

1.3 Cancer evolution

1.3.1 Clonal theory

Darwinian theory of evolution describes the evolution of species through natural selection of inherited traits that increase the individual's ability to compete, survive and proliferate. The concept of evolution in cancer was synthesized by Peter Nowell in 1976 [43], where he described a clone, which can be defined as a group of cells related to each other by descent from a unitary origin, clonal structure and evolution in cancer. The heritable markers could be anything that can be transmitted faithfully from cell to cell, such as SNVs, CNAs, or

DNA barcodes in experimental systems. Clonal relationship between sub-populations arises as a function of cell division and mutations and selection operating on cells of different fitness induces clonal dynamics [44]. In this clonal evolution setting, cancer cells are undergoing selection pressure caused by genetic or non-genetic processes.

1.3.2 Clonal evolution in cancer

Genetic and epigenetic alterations that affect the fitness of clones will lead to expansion or shrinkage of specific clones. Clonal selection operates on cell populations that undergo evolution and result in dynamical behaviour which may increase or decrease a sub-population. The clonal evolution, that involves cancer cells and their microenvironment, results in ITH which provides the substrate of phenotypic variation on which selection process can occur [45]. Dynamical behaviour can also arise from random effects and thus requires quantitative methods to differentiate genetic drift or stochastic dynamics, from fitness differences resulting in stable selection. Evolution continues through cancer's lifetime as a mean to adapt to its environment and ITH is a consequence of this evolutionary process. Linear evolution describes that fitter clone arises, replaces lesser fit lineages and become a dominant clone of the population. Branched (parallel) evolution is defined as the development of multiple subclones from same ancestor, each with selective advantages and distinct phenotype. Clonal relationship can also be neutral wherein genomic differences are derived from mutations over time (genetic drift) without selection and therefore confers neutral phenotypic consequences. Williams et al. have provided the evidence of neutral growth pattern of subclones [46] as well as Ling et al. who showed neutral evolution pattern based on the degree of heterogeneity in colorectal cancer cells [47]. Punctuation evolution is described as sudden changes in cancer genome in a single catastrophic event that induces a radical change in phenotype [48]. However, not a single model of evolution

can explain the whole process of cancer evolution. The picture of evolutionary process depends on the nature of time and space of sampling as well as resolution of the assay. For example, the evolution may appear as linear if the sample is taken after the fitter clone has replaced the population or only the dominant clone has been sampled. Limited resolution of the assay could result in missing clones and over-representing others.

1.3.3 Tumour heterogeneity and metastasis

ITH has been shown to be associated with metastatic potential. Ahmed et al. showed that heterogeneity within tumour as well as within circulating tumour cell clusters contribute to metastasis indicating the importance intratumour heterogeneity in cancer metastasis [49]. Yang et al. measured ITH by calculating Shannon Index and showed that high ITH is associated with worse metastasis free survival [50]. It has been shown that metastatic tumour demonstrates less heterogeneity of IHC markers compared to its primary tumour [51] suggesting that selected cells from primary tumour contribute to formation of metastatic lesion. It is plausible that tumours with high ITH will have higher chance to have clones with metastatic potential. However, it has not been shown to what extent ITH of primary tumour contributes to metastasis. Moreover, the cellular heterogeneity in cancers requires methods for identifying and studying the properties of single cells, across space and time.

1.3.4 Metastasis as an evolutionary process

The cells of primary tumour may undergo evolutionary processes and acquire traits that can lead to metastasis development. There are two general models of the evolution of metastatic disease which are distinguished based on; (i) the relative timing of the emergence of metastatic clones and; (ii) the expected genetic divergence between primary tumour and its metastasis. Common assumption in both models is that primary tumour and its metastases are derived from a

common ancestral cell, thus clonally related [52]. The linear progression model describes metastasis as a later stage of primary tumour progression where the most advanced primary clones disseminate, so the degree of genetic divergence between primary and metastasis is expected to be small. Clinical observations that the association of larger tumour size (higher stage) with higher frequency of metastasis support the concept of late dissemination of tumour cells during primary tumour progression. The linear progression model assumes that metastases are from fully malignant cells that arise mostly in advanced cancers [53]. Predicting response to therapy using the primary tumour is based on linear progression model as the molecular characteristics of disseminated tumour cells are derived from primary tumour cells. However, the linear progression model cannot explain the metastases that are diagnosed at early-stage cancers or the high frequency of distinct genetic alterations in primary tumours and metastasis. The parallel progression model describes that dissemination of metastatic clones occur at earlier stage of primary tumour development and both primary tumour and metastasis evolve in parallel, resulting in substantial genetic divergence between them [52]. In parallel model, tumour cells of primary and metastasis undergo independent accumulation of genetic and epigenetic alterations and evolve under different selection pressures from different microenvironment [53]. Several studies have shown the evidence of early dissemination of tumour cells from primary tumour and parallel evolution [54, 55].

1.3.5 Identifying clonal structure of cancer

Cell populations of cancer with common features can be identified by several means including defining the clonal structure, which is a function of cell division and heritability of mutations. Cell sub-populations can be defined by the heritable markers, such as mutations or CNAs, with repeated observation of sub-populations in time or space. The heritable markers may

or may not be directly linked to a phenotype. The other method of identifying cell sub-populations is by identifying single-cell gene expression characters and grouping them. These approaches can be combined, however in this thesis we are focusing on understanding sub-population structure using copy number features of tumour cells since CNAs are common features in breast cancer [56, 57].

1.3.6 Phylogeny and cancer evolution

The application of evolutionary theory and modern genomic technologies in cancer research have led to the development of computational methods to reconstruct evolutionary processes that is, phylogenetics [58]. Phylogeny, also known as phylogenetic tree, shows the lines of evolutionary descent from common ancestors. Phylogenetic methods define certain traits or markers on observed taxonomic units to infer their relationship. The use of phylogenetics in cancer has advanced our understandings of tumour progression. Cell division and inheritance describes a branched process where a tree structures can naturally show the relationship between elements of the population. Tumour cells can be grouped into clones that share same traits, such as SNVs or CNVs, and clonal lineage can be identified by placing clones to a tree based on their relationship to each other. The labels at the tips of a phylogeny can correspond to individual clones. A clade is a piece of a phylogeny that includes an ancestral lineage and all the descendants of that ancestor which can be referred to as monophyletic group (Figure 1-2 A). Clones in the same clade share a portion of history that is common to all members of the clade and to no other clones in different clades [59, 60]. A monophyletic group can be separated from the root with a single cut, whereas a non-monophyletic group needs two or more cuts. For example, clone A and B in Figure1-2 A are in the same clade and can be considered as a monophyletic group whereas clone D and E are from different clade of a tree and thus can be

considered as non-monophyletic or polyphyletic group. By building a phylogenetic tree, we can infer evolutionary history of tumour cells during cancer progression and it would give us insights to understand the processes of emergence of drug resistance and metastasis.

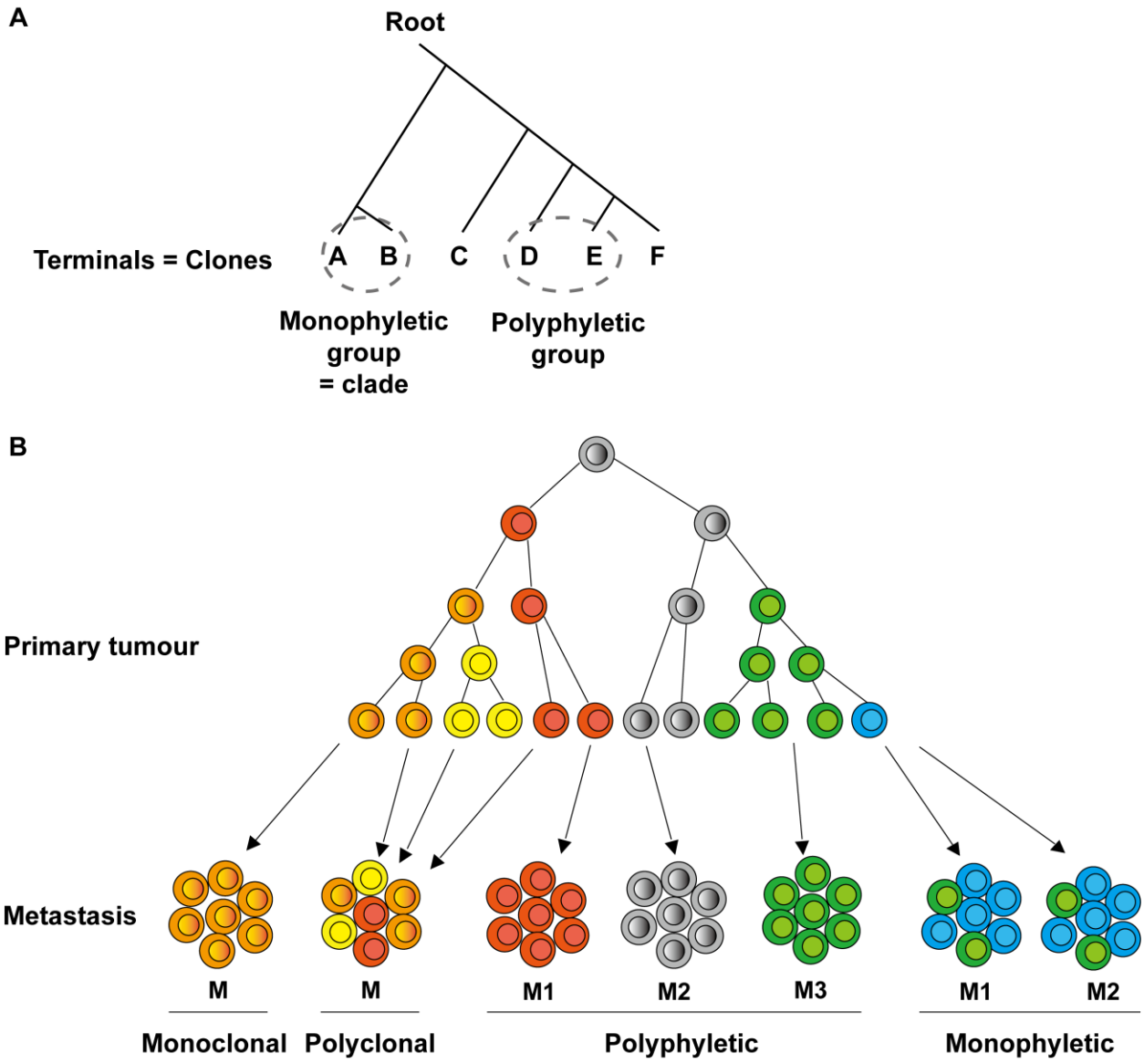


Figure 1-2 Patterns and routes of metastasis.

A. A root is the ancestral population from which all the other tumour cells originate. Terminals of branches are labeled as clones. A monophyletic group or a clade can be separated from the root with a single cut, whereas a non-monophyletic group needs two or more cuts. B. Metastasis can be monoclonal or polyclonal. Multiple metastasis to different organs may derive from the same subclone (monophyletic) or different subclones (polyphyletic).

1.3.7 Patterns of metastasis

Metastasis can develop via multiple routes and pattern of metastasis may vary among tumours (Figure 1-2 B). Metastatic lesions from a primary tumour can be both monoclonal or polyclonal [61]. Monoclonal metastasis can be resulted from a single clone that survived the metastatic process and managed to proliferate at distant site. Polyclonal metastasis can result from multiple clonal cooperation during metastatic process. Evidence for polyclonal seeding of metastases was observed in a common mouse model of breast cancer [62]. Multiple metastases to different organs from the same subclone of the primary tumour are termed monophyletic whereas distinct subclones in the primary tumour result in different metastases at different times are termed polyphyletic [52]. McPherson et al. observed both monophyletic and polyphyletic metastasis pattern in high-grade serous ovarian cancer patients and pointed out the importance of obtaining samples from multiple metastases to better understand the whole view of clonal evolution in metastasis [63]. Whether a certain tumour type or characteristic develops specific pattern of metastasis remains unknown.

1.4 Single cell analysis

Identifying the molecular basis for cancer progression and metastasis at single-cell resolution is a leading field in cancer research. Traditional bulk approaches are sequencing DNA

from a mixture of cells which can sequence at high coverage depth and identify major tumour subclones. Several methods have been developed to infer clonal population structure using bulk sequencing data, such as PyClone which determines the fraction of tumour cells and clusters them to infer clonal population [64]. The prevalence of a particular clone measured over time and space is an indirect inference of population structure. Multi-region sampling and sequencing is another approach to measure clonal population and infer clonal evolution [6, 65, 66]. However, bulk approaches are limited to resolve minor populations by sequencing error rates [67]. Single-cell analysis can provide new perspectives to our understanding of cancer biology in unprecedented resolution and is required for studying clonal evolution and rare cell populations as well as sub-clonal mutational patterns as bulk sequencing obscures signals from small subsets of cells. Wang et al. studied clonal evolution in breast cancer using single nucleus genome sequencing and revealed different evolution pattern between ER+ and triple negative tumour cells [68]. Eirew et al. identified dynamics of genomically defined clonal population in serial propagation of primary and metastatic human breast cancers in immunodeficient mice using single-cell sequencing methods [69]. Furthermore, Gao et al. revealed that the majority of CNAs are acquired at early stage of tumour evolution in TNBC using single nucleus sequencing method [70]. It has also been shown in colorectal cancer that primary and metastatic tumours can be distinguished by copy number substructure from single-cell sequencing [71]. As shown in previous studies, analysis of tumour cells in single-cell resolution is imperative to understand the clonal dynamics during metastasis in detail.

One of the technical issues for single-cell sequencing technology arise from the low quantity of DNA extracted from single cells, leading to requirement of whole-genome amplification. Several amplification-based methods have been described [68, 72–75] however,

amplification induces technical artifacts such as nonuniform coverage and polymerase bias [76, 77]. We used DLP+ method, which is a scalable single-cell whole-genome sequencing platform without whole-genome amplification, to identify clonal populations and their genomic features of tumour cells [77].

1.5 Copy number alteration in cancer

Chromosomal instability is a hallmark of cancer [1, 2] and leads to CNAs that can affect a large portion of the genome and impact gene expression in several human cancers [78–80]. In breast cancer, CNA is associated with distinct clinicopathological features, including clinical outcome and gene-expression subtypes [81]. It has been also shown that CNA was associated with nodal metastasis in breast cancer patients [82] and difference of CNA between primary tumour and brain metastasis has been identified [83]. However, we understand very little regarding how copy number of the genome affects the clonal evolution during tumour progression and metastatic potential of tumour cells. The role of CNA during metastasis development and its impact on metastatic potential of tumour cells are yet to be discovered. Although many forms of genomic variation could be used as heritable markers for clonal analysis, we are focusing on CNAs of breast cancer to understand their impact on sub-population structure and metastatic potential.

1.6 Murine models of breast cancer and metastasis

A number of immortalized cell lines have been established and characterized over the last several decades, but there are limitations to this *in vitro* model to understand the cancer biology. For example, most established human breast cancer cell lines have been maintained *in vitro* over

decades and this long-term culture has been associated to extensive clonal selection and loss of heterogeneity that these cell lines do not represent clinical samples anymore [84]. Also, cell-line culture does not involve the microenvironment which plays an important role in development of cancer. To overcome the limitation of cell line models, *in vivo* models are developed and being used for cancer research. Immortalized cancer cell lines can be injected into animal and grown as xenografts to explore underlying molecular mechanism of cancer [85]. Metastasis can be studied by using experimental, forced metastasis or spontaneous metastasis model [85–88]. Tail vein injection is used to study lung metastasis and intracardiac injection of tumour cells is used for systemic dissemination. In spontaneous metastasis model, tumour cells can be injected subcutaneously or orthotopically which can mimic the metastatic processes in human cancer. However, xenografts using cancer cell lines cannot investigate the complexity of human tumours due to lack of heterogeneity of tumour cells and stromal components.

On the other hand, patient-derived xenograft (PDX) models can represent the full spectrum of heterogeneity of breast cancer, theoretically. The PDX model approach is based on the transplantation of primary or metastatic human tumours directly into highly immunocompromised mice. PDX models are established and maintained *in vivo* and have been shown to retain critical biological, histological, genomic, transcriptomic and proteomic features of their tumour of origin [89–92]. Although PDX models cannot fully recapitulate human microenvironment due to lack of immune and stromal component [92], it can be used as a proxy to study the biology of cancer under various experimental conditions that cannot be done in patients.

To establish human tumour growth in mice, immunocompromised, or immunodeficient, mouse model should be used. There are a variety of immunocompromised, or immunodeficient

mouse models that can be used to develop breast cancer PDX models. The choice of immunocompromised hosts largely depends on researcher's preference however, NSG (NOD/SCID gamma) mice, which lacks T cells, B cells and natural killer cells, are currently the most popular choice for development of breast cancer PDX [90].

1.6.1 PDX models of breast cancer heterogeneity and evolutionary dynamics

Efforts to develop PDX lines have been going from several laboratories and 537 individual PDX lines representing over 500 patients have been reported. From the report, 56% of patients yielding PDX lines had TNBC, 36% of patients had ER+ cancer and 8% of patients had HER2+ cancers [90]. PDX models of breast cancers also have been shown to reproduce the molecular heterogeneity of patients [93]. Recent studies have shown that PDX models can be successfully used to investigate the heterogeneous nature of human breast cancers and their evolution [15, 69].

1.6.2 PDX models for breast cancer metastasis

With respect to metastasis, PDX have been shown to produce circulating tumour cells as well as metastatic lesions to several distant sites [89, 91]. It has been shown that PDX tumours follow metastatic site specificity of tumour of origin to some extent however, whether PDX tumours show fidelity with the metastatic behaviour of the tumour of origin needs to be investigated further. Recent study using spontaneous breast cancer metastasis PDX model showed that metastatic cells showed substantial transcriptional diversity [88]. Primary tumours and metastases of PDXs harbored different proportions of clusters that showed distinct transcript expression. It highlights the importance of heterogeneity of PDX tumours to identify the mechanisms involved in human cancer metastasis. Thus, it is crucial to establish a PDX models

that retain heterogeneous population of tumour cells to understand the complex nature of human cancer metastasis as an evolutionary process.

1.7 Hypothesis and aims

The universal mechanism of metastasis has not been identified and individual tumours may use different mechanism to develop metastasis. All tumours have different clonal populations in a different environment, thus mechanisms used to develop metastasis may differ for each tumour in terms of evolutionary process. It has not been shown how TNBC tumours evolve to develop metastasis in a single-cell level. Questions that this thesis aims to address are What is the clonal relationship based on copy number states between primary and metastatic tumour cells? How does the copy number alteration play a role in metastasis? Is there a metastatic specific clone among heterogeneous population of primary tumour?

We hypothesize that metastasis develops from pre-metastatic genomic clones, that is defined by copy number states of the genome, and CNAs in tumour cells can affect their metastatic potential.

Aim 1: Establish breast cancer metastasis PDX models and characterise PDX tumours and metastases.

Aim 2: Identify clonal population/structure of primary and metastatic tumour cells and determine the relationship between them using single-cell whole-genome sequencing.

Chapter 2: Methods

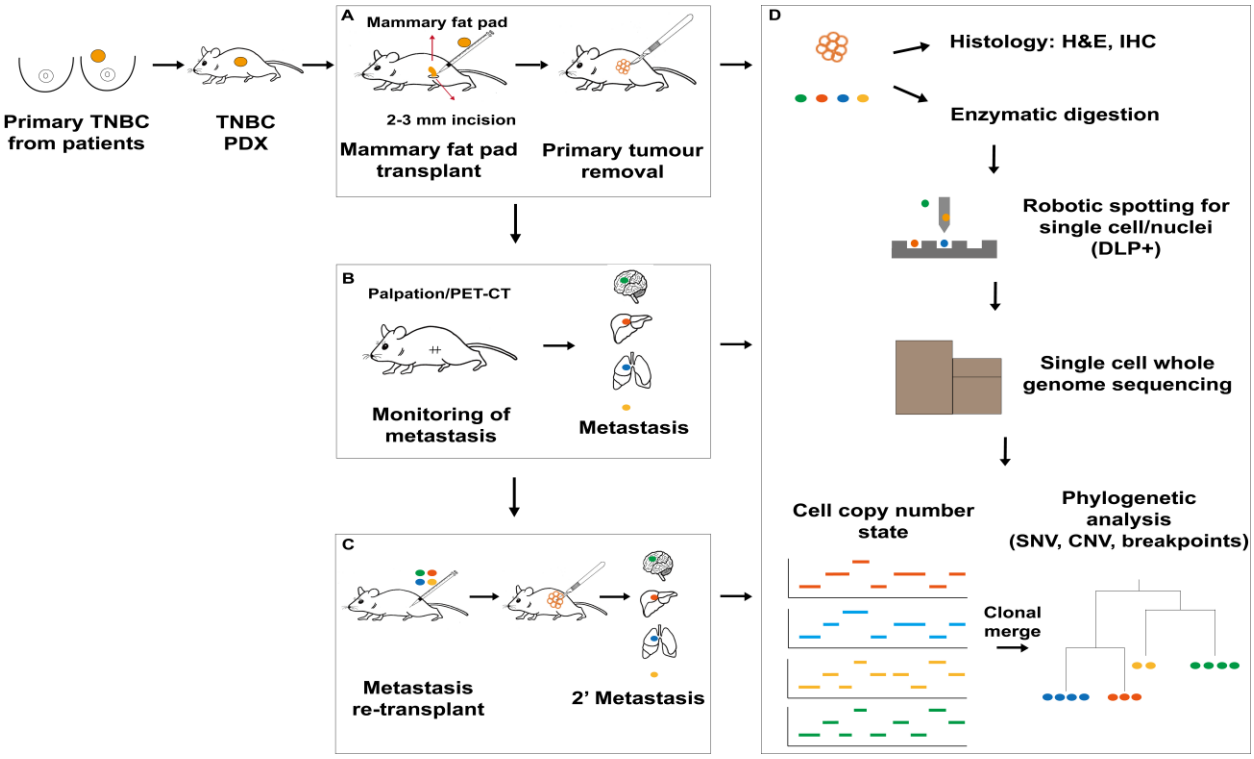


Figure 2-1 Study design.

Primary breast tumour tissues from TNBC patients have been implanted into mice to develop TNBC PDX model. TNBC PDX tumours were then used to develop breast cancer metastasis PDX model. A. Development of breast cancer patient-derived xenograft; mammary fat pad transplant, primary tumour removal and surgical technique optimization. B. Monitoring and detection of metastasis development. C. Re-transplant of metastatic cells into mammary fat pad experiment. D. Analysis of tissues and tumour cells from primary tumour and metastases using histology (H&E, IHC) and single cell whole genome sequencing.

2.1 Xenografting

2.1.1 Animals

Female NOD/SCID interleukin-2 receptor gamma null (NSG) and NOD Rag-1 null interleukin-2 receptor gamma null (NRG) mice were bred and housed at the Animal Resource

Centre at the British Columbia Cancer Research Centre. Surgery was carried out on mice between the ages of 8-14 weeks. All experimental procedures were approved by the University of British Columbia Animal Care Committee and the University of Cambridge Animal Welfare and Ethical Review Committee (H16-01625 and H20-00170).

2.1.2 Mammary fat pad transplant

Mice were anesthetized using isoflurane. Lidocaine (0.25%) up to 60 ul was injected subcutaneously for local anesthesia. Ophthalmic ointment on the eyes were applied to prevent the eyes from drying out. Meloxicam (company) was administered at 5mg/ml subcutaneously before surgery for analgesia and 0.9% sterile warm saline for prevention of dehydration. The hair on the flank was shaved and a sterile skin preparation was made by hibitaine soap followed by 70 isopropanol. Skin incision of 2-3 mm was made on the left flank of the mouse over mammary fat pad, cranio-caudal direction. Mammary fat pad was exposed with forceps and tumour cells suspended in 60ul volume of 50:50 v/v mixture of cold Matrigel: DMEM/F12 were injected into mammary fat pad using a 16-gaugesyringe. Skin incision was closed using Vicryl 5-0 suture with subcutaneous and simple interrupted suture. Tumour cells were transplanted into four mice per each PDX line, unless otherwise specified (Figure 2-1 A).

2.1.3 Surgical removal of primary tumour

Mice were anesthetized and prepared for surgery as described above. Skin incision approximately 1-1.5cm was made over the tumour depending on the tumour size, cranio-caudal direction. Tumour was dissected from surrounding tissue using Metzenbaum scissors to minimize tissue damage. Mammary fat pad attached to tumour was removed along with the tumour. Overlying skin was removed with the tumour when the tumour was firmly attached to

the skin. Skin incision was closed using Vicryl 5-0 suture with subcutaneous and simple interrupted suture.

2.1.4 Serial passaging of PDX

Selected tumours were serially passaged as previously described. Briefly, tumours removed during survival surgery were minced finely with scalpels then mechanically disaggregated for one minute using a Stomacher 80 Biomaster (Seward Limited, Worthing, UK) in 1 to 2 ml cold DMEM-F12 medium. Serial passaging was performed with aliquots directly from processed tumour cells or cryopreserved materials from previous passage.

2.1.5 Organ harvesting

All mice were euthanized when there was a recurred tumour, palpable metastatic mass, symptom suggestive of metastasis or evidence of metastasis from imaging study. Palpable tumours that developed under the skin were removed first. Then a long, vertical midline incision was made on the abdomen with scissors. Any suspicious organs or masses were removed and transferred to a tube with medium (DMEM-F12) and kept in ice. Thoracic cage was opened by removing rib cage. The heart was infused with saline and lungs were infused with saline through airway before collecting into a tube. Other parts of mouse body were fully evaluated for metastasis including axillary area, intracranial, intraperitoneal, retroperitoneal and thoracic space. Brain, lungs, and liver were routinely harvested for histopathological review for the evidence of metastasis.

2.2 Tissue processing for PDX tumours

A small piece of tumour tissue was removed using scalpels and fixed in 10% formalin buffered saline (Fisher Scientific, Kalamazoo, MI, USA) for histological analysis. Additional

small fragments from different portions of the tissue, dissociated with scalpel, were collected together, flash frozen in liquid nitrogen and stored at -80℃ for nucleic acid extraction. The remaining tissue was minced finely with scalpels then mechanically disaggregated for one minute using a Stomacher 80 Biomaster (Seward Limited, Worthing, UK) in 2 mL cold DMEM-F12 medium. Aliquots from the resulting suspension of cells and clumps were used for xenotransplantation or cryopreserved for single-cell analysis in DMEM-F12 medium with 40% FBS and 10% DMSO.

2.2.1 Digestion of tumour cells for single cell sequencing

Cryopreserved stomached cells/organoids were thawed rapidly in a 37℃ water bath, topped up to 1.5 mL with DMEM (Sigma) and centrifuged (1100 rpm, 5 minutes), discarding the supernatant to remove DMSO from freeze media. 0.5mL collagenase/hyaluronidase (StemCell) was added to the tissue and topped up to 1.5 mL with DMEM, pipetting up and down to dislodge tissue pellet. The tissue was incubated at 37℃ for two hours, pipetting up and down the sample for 1 minute every 30 minutes during the first hour, and every 15-20 minutes for the second hour, before centrifuging (1100 rpm, 5 minutes) and removing the supernatant. The tissue pellet was resuspended in 0.5mL trypsin, pipetted up and down for 1 minute, topped up with FBS to 1.5mL and centrifuged (1100 rpm 5 minutes), discarding the supernatant. 1mL dispase (StemCell) was added to the tissue pellet and pipetted up and down 1 minute, and centrifuged for 5 minutes at 1050-1100 rpm, discarding the supernatant. Digested cells were resuspended in PBS + 0.04% BSA inappropriate volume to achieve a concentration of 1 million cells/mL). Cells were passed twice through a 70um filter to remove remaining undigested tissue and this single-cell suspension was used for DLP+.

2.3 Detection of metastasis

Metastasis was detected either by palpation, PET/CT or gross inspection after euthanasia and confirmed by histologic evaluation.

2.3.1 Palpation

Mice palpation was done weekly until detection of primary tumour and then twice a week after detection for a measurement. Measurement of tumour size was performed by two technicians to reduce the variability in measurement. Tumours were measured in two dimensions using a digital caliper and expressed as tumour volume in mm^3; defined as: [volume = 0.52 x (Length) x (Width)]. After removal of primary tumour, mice were monitored for development of metastasis. When mice developed palpable metastatic lesion, measurement was done twice a week until euthanasia.

2.3.2 Histopathological evaluation

Tumours and organs from mice were processed into formalin-fixed paraffin-embedded blocks. All samples were stained with hematoxylin and eosin (H&E) and slides were evaluated for the presence of tumour cells by a pathologist (Dr. Takako Kono). Tissue microarrays were constructed using duplicate 0.6mm cores extracted from formalin-fixed paraffin-embedded blocks. From each tissue microarray block, 4um thick sections were cut and immunostained on a Ventana Discovery XT staining system (Ventana Medical Systems, AZ, USA). Sections were deparaffinized in xylene, dehydrated through three alcohol changes, and transferred to Ventana Wash solution. Endogenous peroxidase activity was blocked in 3% hydrogen peroxide. The slides were reviewed by a pathologist (Dr. Takako Kono) and scoring method for each antibody is described in Appendix Table A-2. The positivity of biomarkers was determined as described in

the table. In general, protein expression was scored visually based on the determination of staining intensity (0 –negative, 1 – weak, 2 – moderate, 3 – strong) and percentage of cells with nuclear, cytoplasmic or membranous staining (0-100%). H-score for EGFR was calculated as: [H-score = intensity x percentage of positive cells]. Biomarker information was considered uninterpretable if there were no tumour cells in the cores or the cores were missing.

2.3.3 PET/CT

PET imaging experiments were conducted using a Siemens Inveon micro-PET/CT scanner. The mice were sedated with 2% isoflurane inhalation and positioned in the scanner. A baseline CT scan was obtained for localization and attenuation correction before radiotracer injection, using 80 kV X-rays at 500 uA, three sequential bed position with 34% overlap, and 220-degree continuous rotation. This was followed by a 10 min static PET scan. For dynamic PET scans, a 60 min list-mode acquisition was started at the time of intravenous injection with 4-6 MBq of 18F-FDG following a baseline CT scan. The mice were kept warm by a heating pad during acquisition. The PET images were reconstructed using the ordered subset expectation maximization and maximum a posteriori algorithm (OSEM3D/MAP), using 2 OSEM3D iterations followed by 18 MAP iterations, with a requested resolution of 1.5 mm. Inveon Research Workplace was used for image viewing and analysis.

2.4 Quantification and analysis of single-cell whole-genome sequencing (scWGS) data

2.4.1 Single-cell library construction and data processing

After processing tumour cells as described above, single-cell suspensions or nuclei were isolated and spotted, and libraries were prepared through optimized DLP+ method as described previously [77, 94]. Briefly, single cell suspension was loaded into a contactless piezoelectric

dispenser (sciFLEXARRAYER S3, Scienion) and spotted into the open nanowell arrays (SmartChip, TakaraBio). Cell dispensing was followed by enzymatic and heat lysis and then each well was dispensed with tagmentation mix (14.335 nL TD Buffer, 3.5 nL TDE1, and 0.165 nL 10% Tween-20) in PCR water followed by incubation and neutralization. Final recovery and purification of single cell libraries was done after 8 cycles of PCR. Cleaned-up pooled single-cell libraries were analyzed using the Agilent Bioanalyzer 2100 HS kit. Libraries were sequenced at UBC Biomedical Research Centre (BRC) in Vancouver, British Columbia on the Illumina NextSeq 550 (mid- or high-output, paired-end 150-bp reads), or at the GSC on Illumina HiSeq2500 (paired-end 125-bp reads) and Illumina HiSeqX (paired-end 150-bp reads).

FASTQ pre-processing, sequence alignment, quality control, copy number calling and S-phase classification and filtering was performed on all libraries as detailed in [77]. Briefly, cells were assigned a QS score for data quality based on a 13 feature Random forest classifier fitted and applied. The quality score ranges from 0 to 1, with 1 indicating a high probability that a library is high quality. We used a threshold score of greater than or equal to 0.75 and cells or nuclei with quality score equal or greater than 0.75 threshold were considered as good quality data and considered in downstream analysis.

Copy number alterations on a per cell basis were determined using a hidden Markov Model (HMM) approach [95] using the HMM Copy package with parameterizations [77]. S-phase cells were identified using an automated classifier trained using cell-cycle sorted cells. Features from HMM output were used to identify cells most probably in early or late phase replication of their genomes and these S-phase cells are remove from the analysis as they can induce strong GC bias and interfere with downstream analysis [77]. We also filtered cells with low coverage. Libraries vary in sequencing depth so there is no standard threshold. We used

threshold of 500,000 reads per library for SA919 and 10,000 reads per library for SA535 to filter cells with low coverage.

To further enable phylogenetic inference, 10-15% of cells with highest average copy number state jumps were removed. Upon inspection, these cells included early and late dividing cells that were not captured by the s-phase classifier. Mouse cells were filtered from xenograft libraries using fastqscreen [96].

2.4.2 Phylogenetic analysis

Phylogenetic analysis plays an important role in characterization of ITH and measuring cell sub-population dynamics. Phylogenetic tree provides the information regarding evolutionary relationships among tumour cells based on the branching structure of ancestry of sequenced cells. Phylogenetic methods are based on defining certain traits or markers on observed individual cells. The majority of single cell phylogenetic tree inference methods focus on point mutations [58, 96–98]. These methods are not suitable for scWGS data from DLP+ which needs scalable methods based on copy number variation. In DLP+, entire tumour cell populations in the range 1000–10,000 or more cells are sequenced, each cell measured in a genome-wide but low coverage fashion. Low coverage limits the inference of phylogenetic tree based on point mutations and an alternative computational method is required to infer a phylogenetic tree from single cell data composed of a large number of cells with shallow coverage. A novel computational method called BREAKTREE has been developed to perform phylogenetic tree inference which is based on single cell copy number matrices retaining only presence or absence of changes in copy number profiles, a change point indicator variable [99]. It is a single-cell Bayesian tree reconstruction method based on copy number binary change point variables that

can reduce the computational cost of tree exploration compared to traditional phylogenetic moves and has better computational scalability.

A clone is defined as a monophyletic clade that has sufficient genomic homogeneity. The degree of homogeneity can be tuned by limiting the number loci and the difference in copy number of sub-clades in a clone. Clones were constructed by identifying connected components in the phylogenetic tree reconstruction and the tree-cutting algorithm (LUMBERJACK) is used to select major subpopulations, i.e., clones. The objective of the tree-cutting algorithm is to identify clones by grouping subsets of cells that are genomically distinct based on both the topology of the tree and the copy number states of cells [100]. The algorithm works by first finding the coarse structure, that is dividing the tree into major clades. The coarse structure was obtained by a two-step procedure. (i) First, identify monophyletic clades. (ii) Then remove the cells comprising the clades found in step one from the tree and repeat algorithm. Then, the fine structures within the initial clades were identified by traversing the tree in a bottom-up manner and merging loci that are sufficiently similar.

2.5 Statistical analysis

All statistical analyses were conducted using GraphPad Prism. Shapiro-Wilk test was used to determine normal distribution of data. Student's two-tailed t-tests (parametric) were used for normally distributed data and Mann-Whitney test was used to compare non-normally distributed data (non-parametric).

Chapter 3: Establishment and characterization of metastatic breast cancer PDX models

3.1 Chapter 3 summary

We established metastatic breast cancer PDX models using from 9 different patient-derived primary human breast cancer tissues in 68 immunodeficient mice. In this chapter, we will explore following findings (Figure 3-1):

1. Not all primary tumour tissues from breast cancer patients developed metastasis in mice. Among 9 different PDX lines, 5 of them developed metastases to loco-regional as well as distant sites. Tumour growth rate and metastatic sites were distinct for each PDX tumour. Metastasis sites included axillary, inguinal, paraspinal, lung, liver, intra-abdominal, and tumour recurrence at transplant site.

2. Protein marker expression was similar between primary tumour and metastases. We assessed primary and metastatic tumour samples with IHC using several protein markers, including ER, PR, HER2, Ki-67, CK5/6, EGFR, nestin, vimentin, E-cadherin, and INPP4B. We observed that expression of above protein markers was similar between primary tumour and corresponding metastases.

3. Metastatic potential appeared to increase with passage number in SA919. We observed that the frequency of metastasis development increased with the passage number in SA919. Metastasis development in passage 3 replicates was none (0/3), 2 out of 4 in passage 4 (50%) and 4 out of 4 in passage 7 (100%).

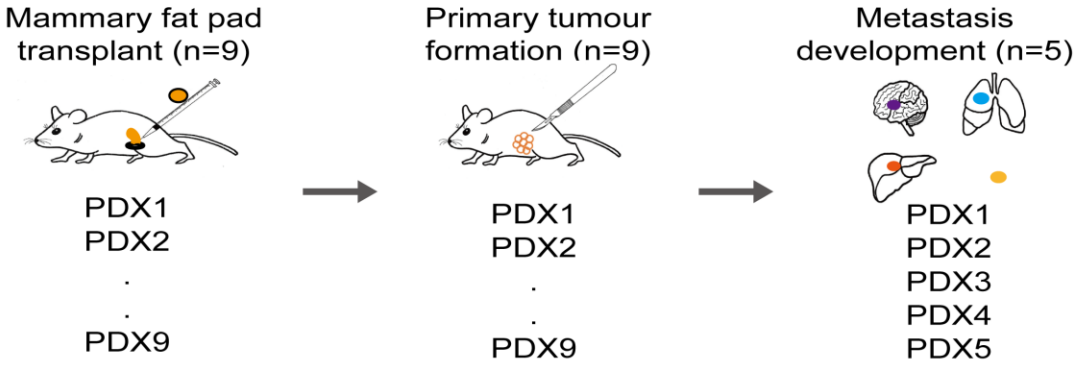
4. Sites of metastases were reproducible over multiple passages. In 2 PDX lines that developed metastases (SA919 and SA535), sites of metastasis were reproducible when multiple

passages were tested. In passage 4 and 7 of SA919, paraspinal metastasis was repeatedly identified in all metastasizing replicates. In SA535, axillary and lung metastases as well as tumour recurrence at transplant site were repeatedly observed in most of replicates of 4 passages (axillary = 12/15, lung = 13/15, tumour recurrence = 13/15 from 15 replicates of passage 4, 5, 6 and 7).

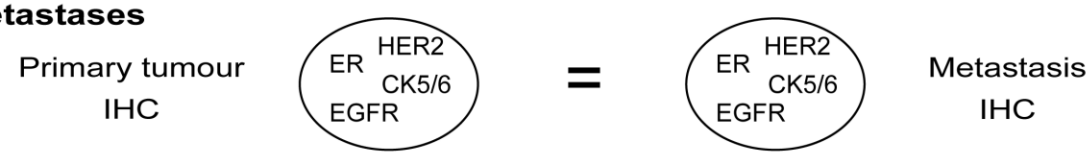
5. Phenotype change from non-basal-like to basal-like occurred upon serial propagation of SA919 tumours. We defined basal-like phenotype based on CK5/6, EGFR, nestin, and INPP4B. We identified that phenotype change from non-basal-like to basal-like occurred in SA919 primary tumours upon serial propagation (passage 3, 4 = non-basal-like, passage 7 = basal-like).

6. Metastatic potential was variably maintained when metastatic PDX tumour cells were re-transplanted into mice. We re-transplanted metastatic cells from SA919 and SA535 into mammary fat pad of mice to see whether metastatic activity is maintained. Re-transplanted cells developed primary tumour as well as metastases in all replicates of both SA919 and SA535 however, SA919 showed increased metastatic activity while SA535 showed reduced activity in terms of the number of metastasis.

1. Not all patient derived tumour tissue developed metastasis



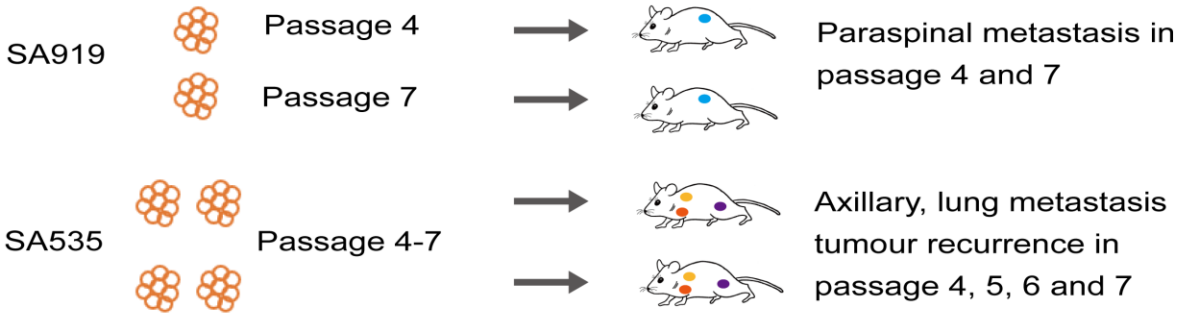
2. Protein marker expression was similar between primary tumour and metastases



3. Metastatic potential appeared to increased with passage number in SA919

Passage 3 -> 0/3 metastasis (0%)
Passage 4 -> 2/4 metastasis (50%)
Passage 7 -> 4/4 metastasis (100%)

4. Sites of metastasis were reproducible over multiple passages



5. Phenotype change from non-basal-like to basal-like occurred upon serial propagation of SA919 tumours

6. Metastatic potential was variably maintained when metastatic PDX tumour cells were re-transplanted into mice

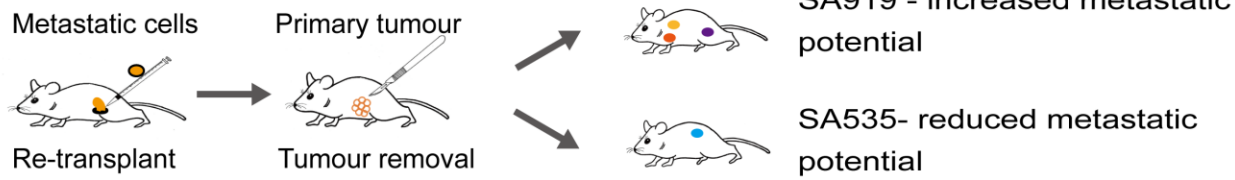


Figure 3-1 Chapter 3 summary.

Chapter 3 describes the findings from the establishment and characterization of metastatic breast cancer PDX models using 9 different patient-derived primary human breast cancer tissues. We identified 5 metastasizing PDX lines and reported tumour growth and metastasis pattern as well as characterization of primary and metastatic tumours with immunohistochemistry (1-2). We also noted that metastatic potential of primary tumour appeared to increase with the passage number in SA919 (3) and metastatic sites were reproducible over multiple passages in both SA919 and SA535 (4). SA919 tumours also showed phenotypic change from non-basal-like to basal-like over the passages (5) and metastatic potential was maintained when metastatic tumour cells were re-transplanted into mice (6).

3.2 Establishment of metastatic breast cancer models from primary TNBC tumour with modified MFP transplant-resection

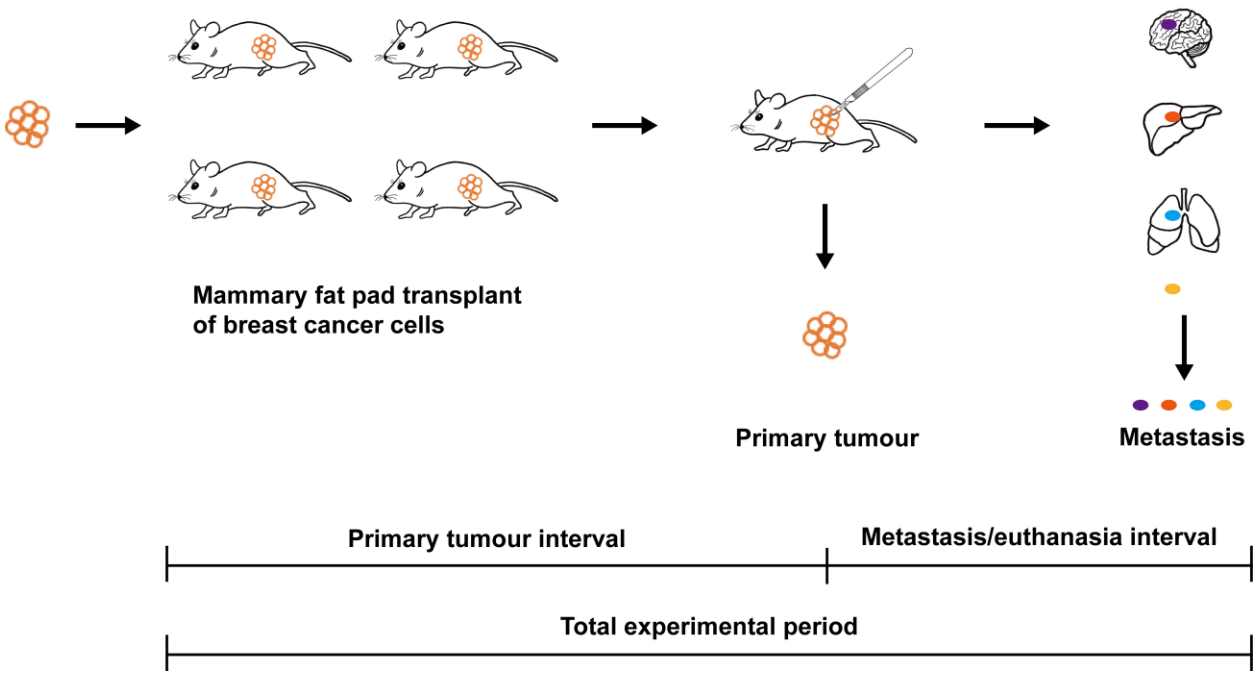


Figure 3-2 Schematic diagram of development of breast cancer metastasis patient-derived xenograft model with orthotopic mammary fat pad transplant.

Primary tumour is removed when it reached maximum allowed end-point size. Mice were monitored and allowed to develop metastasis. Development of metastasis is evaluated in distant organs and collected for subsequent analysis.

To develop metastasis in PDX model, debulking surgery of primary tumour at orthotopic transplant site is required as the time taken for metastasis to develop would be well beyond the endpoint growth of implanted tumour cells [101]. Tumour cells from established primary TNBC PDXs in the lab were transplanted into MFP of mice. Primary tumour was removed when it reached a maximum allowed endpoint size (1,000mm^3) and mice were monitored and allowed to grow metastasis (Figure 3-2). Mice were euthanized to collect distant organs when they

developed clinical signs of distress from metastasis, evidence of metastasis from palpation or PET-CT scan, or reached humane endpoint.

For the purpose of this experiment, mice needed to receive two survival surgeries, MFP transplant and primary tumour removal, each with surgical recovery for a period of weeks. Surgical technique for MFP transplant was adapted and modified from previously reported methods [91, 102]. A small incision was made on left flank area above the 4th mammary fat pad, instead of making a conventional wide midline incision extended to hind-leg [91], to minimize the wound and surgical trauma (Figure 2-1 A).

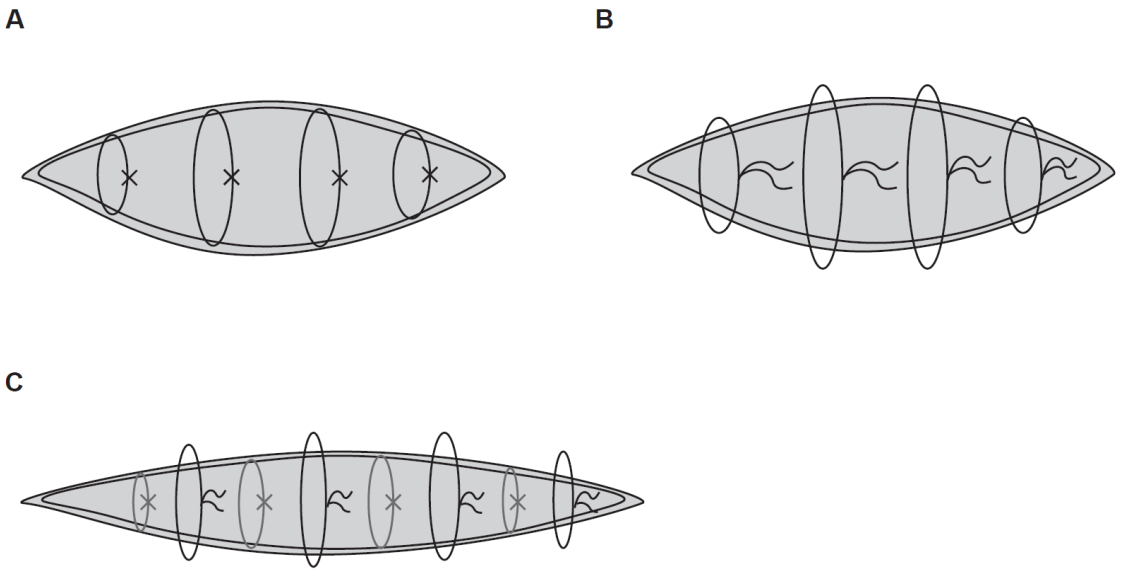


Figure 3-3 Optimization of wound closure surgical technique.

A. wound closure using subcutaneous suture with buried knots. B. wound closure using regular suture. C. Double layered wound closure with both subcutaneous and regular suture.

The published wound closure methods resulted in unacceptably high wound dehiscence (3 out of 12, Figure 3-3 A, B). We explored the use of longer suture materials and tissue glue, neither of which were effective, however double layer closure reduced the wound dehiscence to 0 out of 8 and was adopted as the standard procedure (Figure 3-3 C). The same suture technique was used for wound closure after primary tumour removal without major complications.

3.3 Identification of metastases by PET-CT and palpation

Development of primary tumour is anticipated in nearly all mice, whereas metastasis may occur in only some. Unlike primary tumour development, distant metastasis will often not be palpable and we therefore used cross-sectional tumour metabolic enhanced imaging (PET-CT), combined with general health evaluation (body weight, appearance, etc), and physical palpation. To determine endpoints, if mice reached more than 10 months post-transplant with no change in health or imaging evidence of metastasis or reached 12months of age, they were sacrificed.

Nine different patient TNBC primary tumours were implanted, all of which exhibited primary site recurrence, as expected with re-transplantable primary PDX. Among 9 PDXs, 5 developed metastases (SA919, SA535, SA1142, SA605, SA609, Table 3-1). Metastasis was detected either by palpation, PET-CT or gross inspection of internal organs after euthanasia and confirmed by histologic evaluation. All tumour recurrence and some axillary metastases were detected by palpation of mass. For example, SA535X4 developed palpable tumour recurrence during monitoring and some axillary/inguinal metastases were also palpable. Metastatic masses in supraspinal area of SA919 were detected by palpation.

Table 3-1 Summary of metastasis development in patient-derived xenograft.

PDX ID	Primary tumour interval (weeks)	Metastasis/ Euthanasia interval (weeks)	Metastasis site	Frequency	Patient metastasis
SA919 (X3)	12	14-18	No metastasis	0/3	Bone (spine)
SA919 (X4)	9-12	16-20	Paraspinal (ventral), abdominal, axillary	2/4	
SA919 (X7)	12-17	10-15	Paraspinal (supra, ventral)	4/4	
SA535 (X4)	16-28	7-18	Axillary, inguinal, lung, tumour recur	4/4	Not detected
SA535 (X5)	19-22	13-15	Axillary, lung, liver, tumour recur	3/3	
SA535 (X6)	11-13	10-15	Axillary, lung, liver, abdominal, tumour recur	4/4	
SA535 (X7)	7	9-16	Axillary, lung, abdominal, tumour recur	4/4	
SA1142 (X2)	21-28	8	Lung, liver	3/4	Not detected
SA605 (X4)	19-26	15-19	Lung	2/3	Bone, liver
SA609 (X3)	7	21-43	Liver	1/4	Bone, pelvic mass, tumour recur
SA609 (X4)	7-11	16-33	Axillary, lung, tumour recur	1/4	
SA609 (X11)	5-8	9-33	Tumour recur (1/4)	0/4	
SA604 (X10)	8-9	22-28	No metastasis	0/4	Bone
SA501 (X5)	6-7	15-17	No metastasis	0/4	Lung, tumour recur
SA1139 (X4)	9-13	20	No metastasis	0/3	Not detected
SA1146 (X2)	10-14	14-40	Tumour recur (2/4)	0/4	Not detected

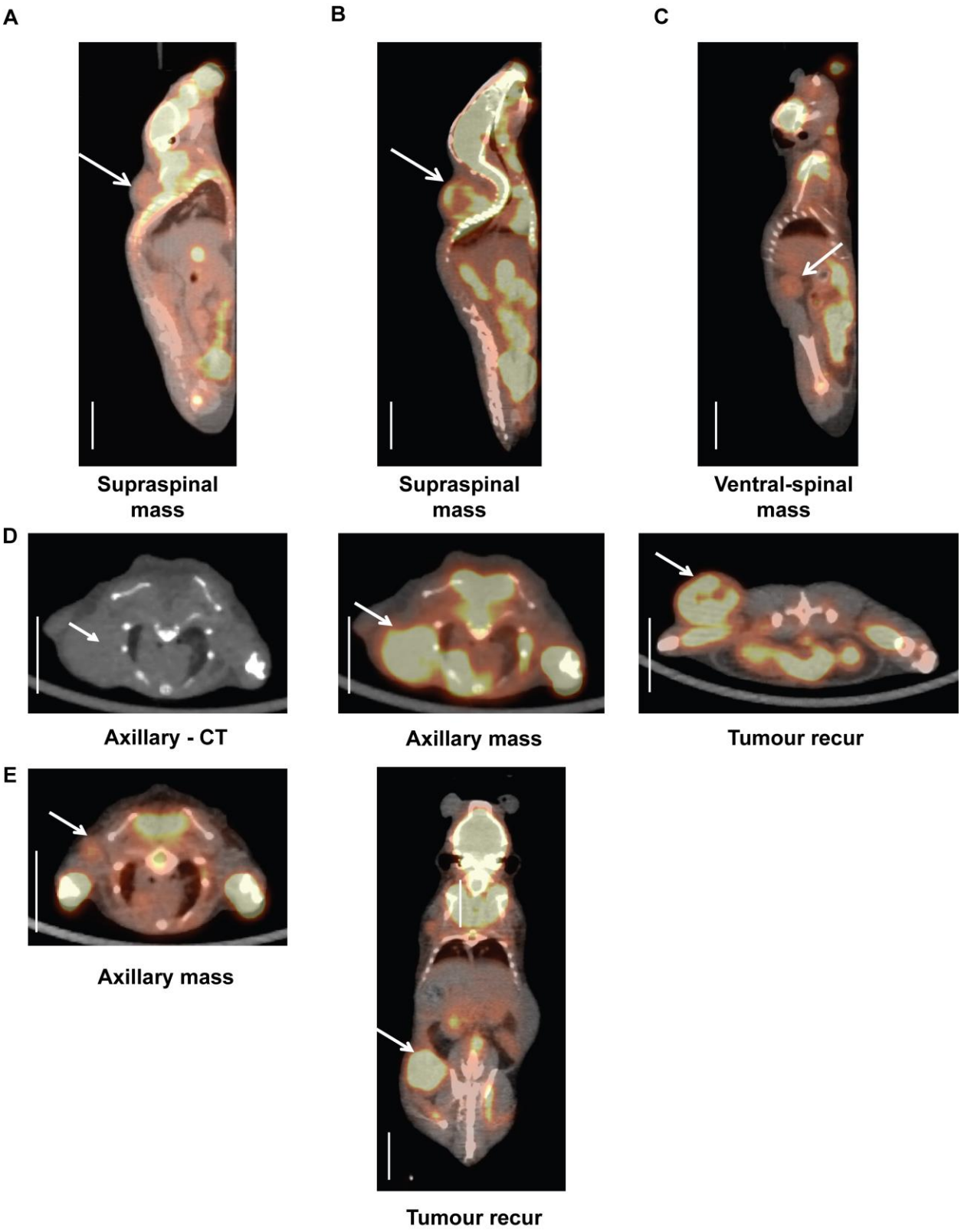


Figure 3-4 PET-CT scan of SA919X7 and SA535X4.

Sagittal view from PET-CT scan of A. SA919X7-1, B. SA919X7-2, and C. SA919X7-4. Paraspinal masses (arrow) at supraspinal (A and B) and ventral spinal (C) region is identified in each scan. PET-CT was not performed on SA919X7-3. D. In SA535X4-1, CT scan revealed a large mass in left axillary area which exhibited high SUV uptake in PET-CT (arrow). Tumour recurrence was identified at left 4th mammary fat pad area (arrow). E. In SA535X4-2, small mass with mild SUV uptake in left axillary area and large mass with high SUV uptake in left transplant site were identified (arrow). PET-CT was negative for SA535X4-2 and was not performed on SA535X4-4. Scale = 10mm.

3.3.1 Imaging based metastasis evaluation

To establish primary site tumour recurrence and metastatic growth, PET-CT imaging was conducted in all 9 lines. A total of 39 scans of PET-CT was performed from 9 PDX lines. Purpose of PET-CT scan was to determine relative time point of metastasis development. PET-CT scan at 10 weeks after tumour removal in SA919X7-4 showed negative finding however, ventral spinal mass was detected from another scan at 15 weeks after tumour removal suggesting that the formation of detectable metastasis occurred between 10 and 15 weeks after tumour removal (Figure 3-4 C). SA919X7-2 also showed metastatic mass at supraspinal area at 13 weeks after tumour removal (Figure 3-4 B). However, in SA919X7-1 supraspinal tumour was detected on PET-CT at 9 weeks after tumour removal (Figure 3-4 A). It should be noted that primary tumour of SA919X7-1 was removed 3 weeks and 5 weeks later than SA919X7-4 and SA919X7-2, respectively, due to slower growth rate. PET-CT was not performed on SA919X7-3 due to the presence of metastasis related symptoms (paralysis of hind legs) before the scan. The relative time point of primary tumour removal for SA919 was at least 12 weeks post-transplant and metastasis development was at least 25 weeks post-transplant.

For SA535, PET-CT was performed 16 weeks after tumour removal in SA535X4-1 and large metastasis to left axillary area with tumour recurrence was detected (Figure 3-4 D). PET-CT of SA535X4-3 revealed small metastasis at left axillary area in addition to tumour recurrence at MFP site (Figure 3-4 E). Tumour recurrence was detected by palpation at 6 weeks after tumour removal and PET-CT was performed 7 weeks after tumour removal. In SA535X4-2, primary tumour was removed 28 weeks after transplant, which took at least 9 more weeks compared to other replicates, because of slow growth rate. PET-CT in SA535X4-2 was negative for metastasis at 9 weeks after tumour removal however, mice developed axillary and inguinal metastasis afterwards which was detected by palpation and confirmed by necropsy at 18 weeks after tumour removal. Additional PET-CT was not taken for SA535X4-2 before euthanasia. PET-CT was not performed on SA535X4-4.

In addition to passage 4 of SA535, we scanned other passages of SA535 to evaluate the validity of PET-CT. SA535X5-2 was scanned with PET-CT 12 weeks after tumour removal, which revealed tumour recurrence and axillary metastasis. The mouse was euthanized 4.5 weeks after PET-CT and small liver metastasis was identified in addition to axillary metastasis and tumour recurrence. Liver metastasis may have been too small to be detected by PET-CT or it developed during 4.5-week period between PET-CT and necropsy. We also scanned SA535X6-2 when tumour recurrence developed 4 weeks after tumour removal. There was no evidence of metastasis other than tumour recurrence from PET-CT. However, axillary and liver metastases were identified during necropsy, 5.5 weeks after PET-CT. It is plausible that these metastases developed during the window period between PET-CT and necropsy or were too small to be detected at the time of PET-CT. Similarly, SA535X7-1 only had tumour recurrence when PET-

CT was performed 10 weeks after tumour removal, however, axillary metastasis was identified 1.5 weeks after PET-CT during necropsy.

PET-CT can be useful to determine the relative time point to euthanize the mouse and harvest metastatic lesions. In most of SA535 PDXs, development of metastasis required more time compared to development of tumour recurrence. Moreover, PET-CT can provide information regarding small metastatic lesions, such as small axillary mass in SA535X4-3 (Figure 3-4 E), that can be missed during necropsy. It was difficult to identify relative time point of metastasis development with PET-CT, since time to develop metastasis in each individual mouse may vary even within the same passage that were transplanted with the same material based on our findings from SA919 and SA535. However, PET-CT can be used as a guide to determine when to evaluate for metastasis and whether to keep or sacrifice mice.

3.3.2 Metastasis evaluation in organs by histology

Palpable lesions from SA919 and SA535 were removed and sent for histology and metastasis was confirmed when tumour cells were present (Figure 3-5, 3-6). SA919 developed metastatic lesion in paraspinal area and SA535 developed axillary or inguinal metastasis as well as tumour recurrence at MFP transplant site. Histologic features of tumour cells, such as pleomorphic cells with high nuclear-cytoplasmic ratio, were similar between primary and metastatic tumour cells in both SA919 and SA535.

Internal organs including lung and liver were grossly evaluated and suspicious findings were confirmed by histologic evaluation. Metastatic involvement of lung was evaluated by histology in all cases since there was no gross lung metastasis among study group. Lung metastasis was observed in 3 of 9 transplants (SA535, SA1142 and SA605, Table 3-1) and the pattern was tumour line dependent. Larger group of cells with multiple lesions were observed in

SA535 (Figure 3-6) and small number of micrometastasis was observed in SA1142 (Figure 3-7) and SA605 (not shown). Lung metastasis in PDX was never detected by PET-CT in this study. Liver metastasis was identified in SA1142 (2 of 3 replicates) which was grossly positive during necropsy and confirmed with histologic evaluation (Figure 3-7). Metastatic cells in liver formed a mass-like lesion invading into liver parenchyme which was different from metastatic cells in lung where a small number of tumour cells were clustered in groups forming several micrometastasis.

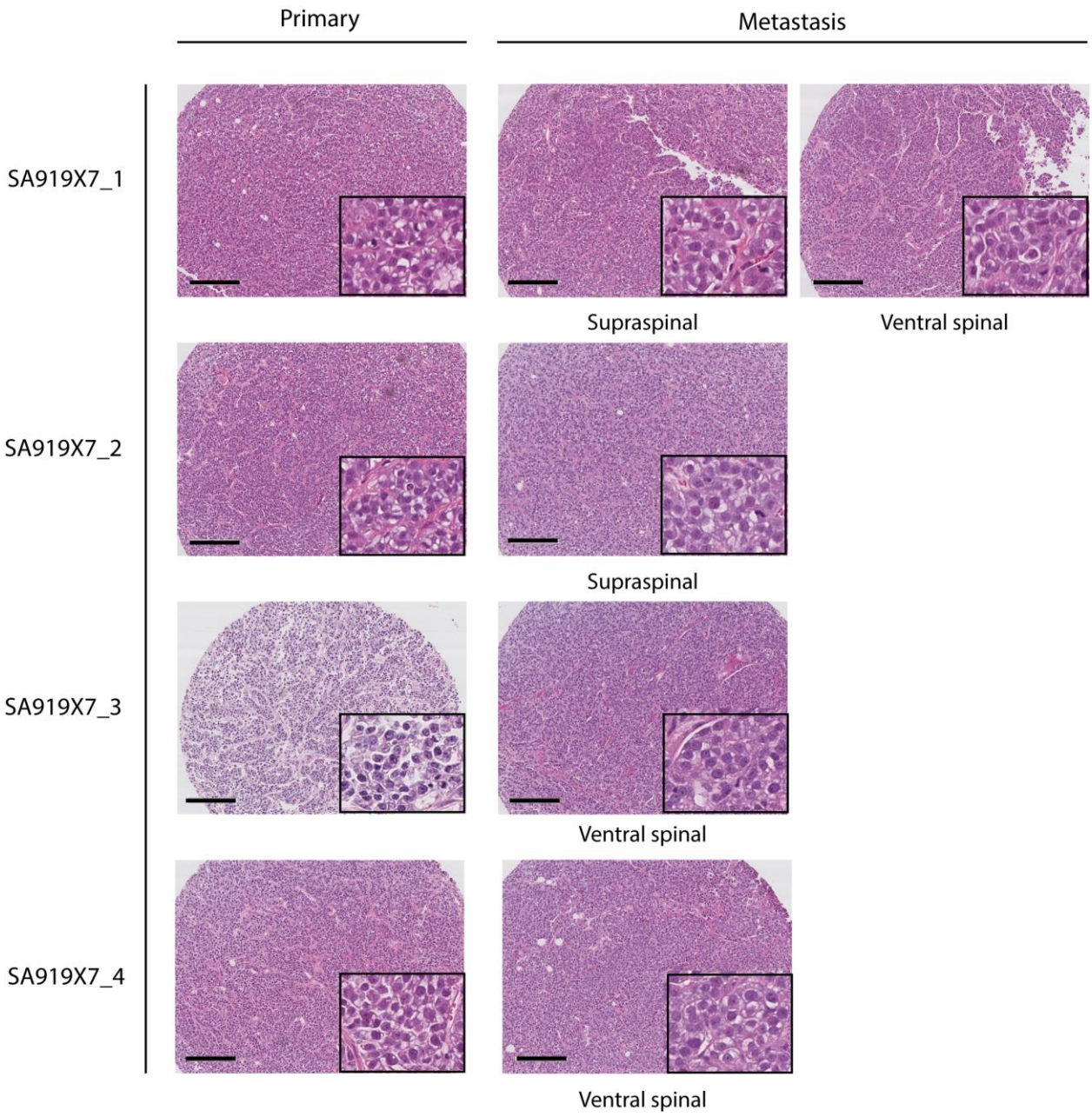


Figure 3-5 Histology images of primary tumours and metastases in SA919X7.

Hematoxylin and eosin staining images showing primary tumours and metastasis from SA919X7 with magnified view at right bottom in each image. Scale = 200 um.

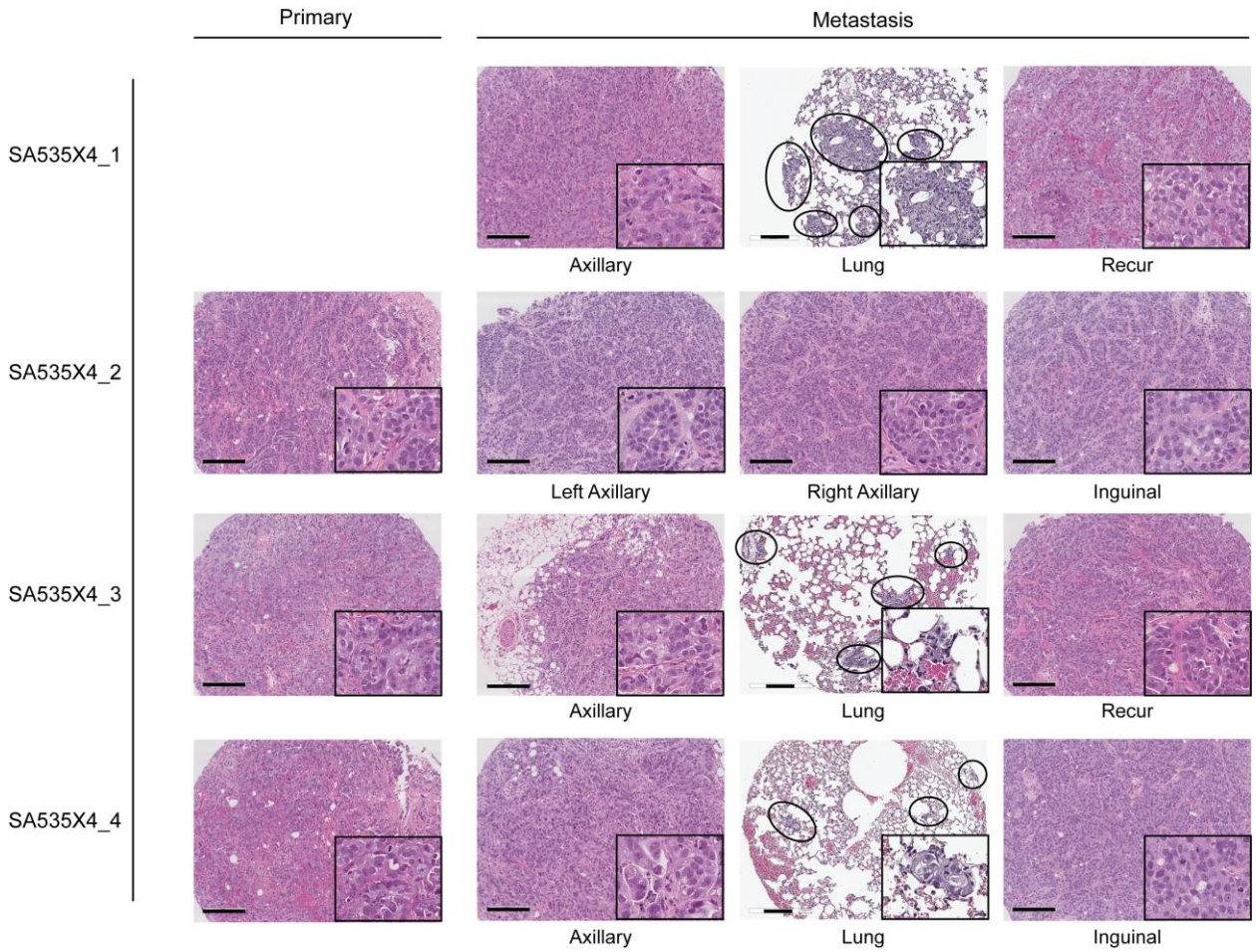


Figure 3-6 Histology images of primary tumours and metastases in SA535X4.

Hematoxylin and eosin staining images showing primary tumours and metastasis from SA535X4 with magnified view at right bottom in each image. Metastatic deposits identified in lung are circled. Scale = 200 um.

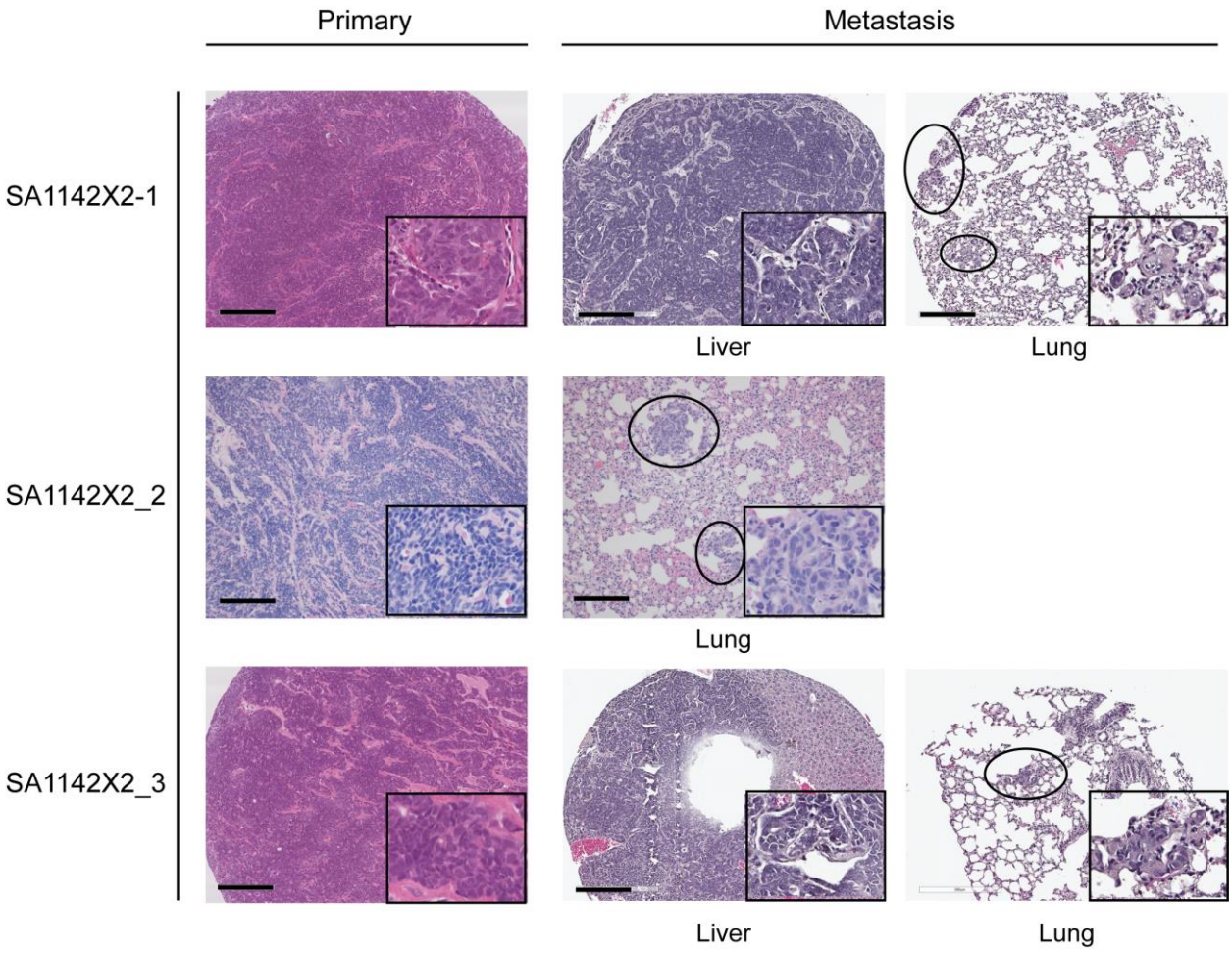


Figure 3-7 Histology images of primary tumours and metastases in SA535X4.

Hematoxylin and eosin staining images showing primary tumours and metastasis from SA1142X2 with magnified view at right bottom in each image. Metastatic deposits identified in lung are circled. Scale = 200um.

3.4 Histopathologic features of primary tumours are maintained in metastases

We constructed tissue microarray of primary tumours of all PDX lines and metastases from three metastatic PDX lines (SA919, SA535 and SA1142). We assessed samples via immunohistochemistry (IHC) using antibodies for ER, PR, HER2, Ki-67, CK5/6, EGFR, nestin and INPP4B. E-cadherin and vimentin were added for SA919, SA535 and SA1142. The percentage of positive cells and the staining intensity were scored, as described in the methods

and Table A-2, for each of the markers by a pathologist Dr. Takako Kono. In lung samples of SA535X4-4 and SA1142X2-3, staining was not available for PR, HER2 and nestin due to lack of tumour cells.

We observed distinct patterns of immunoreactivity from different PDXs (Figure 3-9 A, B). All tumours were negative for ER, PR and HER2 except for SA609 which showed weakly positive (1+) in 1% of cells for PR. SA609 showed negative EGFR while all other tumours were EGFR positive. Interestingly, Ki67 levels (percentage of positive cells) were significantly low (range 1-45%, mean = 30.05%) in tumours that developed solid metastasis (SA919, SA535 and SA1142) than non-metastasizing tumours (SA609, SA604, SA501, SA1139 and SA1146) that showed higher levels (range 50-70%, mean = 60%) of Ki67 (Mann-Whitney test, $p < 0.0001$).

All primary tumours and metastases were negative for ER, PR and HER2, consistent with patients' diagnostic tumour marker profile, except for one tumour recurrence from SA535 (SA535X4-3) which showed a slight increase of ER (intensity/percentage = 1/20, Tables 3-6, A-4). IHC profiles of primary tumours were similar to those of corresponding metastases in all three PDXs (Tables 3.5, 3.6). We evaluated markers associated with epithelial-mesenchymal transition (EMT), such as E-cadherin, vimentin and nestin, and basal-like phenotype markers to identify differentially expressed markers between primary tumour and metastases. E-cadherin was expressed in all primary tumours and metastases of 3 PDXs while vimentin and nestin showed different expressions between PDXs but not between primary tumours and metastases. Vimentin expression has been shown to contribute to epithelial cell migration and is considered as a hallmark of EMT which is a well-known mechanism for cancer cell invasiveness and metastasis [103–105]. However, there was no difference of vimentin expression between primary and metastatic tumour cells. All samples of SA919 showed completely negative vimentin

staining (Table 3-5) while SA535 showed high intensity and percentage of vimentin (intensity = 3, percentage range = 20-95) and SA1142 exhibited low (1-10%) percentage of positively stained cells with intensity 1-3 (Tables 3-6, A-2). Nestin is expressed in the basal/myoepithelial cells of normal mammary gland and nestin expression has been shown to be associated with a basal phenotype [106, 107], poor prognosis [108, 109] and metastasis [110]. Similar to vimentin, there was no difference of nestin expression between primary and metastatic tumours in SA919 and SA535. Most of SA919 tumours (17/19) showed negative staining for nestin while most of SA535 tumours (21/22) showed positive. In SA1142, 2 primary tumours and 1 lung metastasis were positive for nestin while 2 liver metastases were negative (Table 3-6).

Lung metastasis showed distinct marker expressions compared to other organs in PDXs. In SA535, lung metastasis showed lower level of intensity and percentage for EGFR compared to primary tumour and other metastases in passage 4 (H-score comparison, Mann-Whitney test, p=0.0005, Figure 3-11 C, Table A-4). Lung metastasis in SA1142 showed higher percentage of vimentin compared to primary tumour and liver metastasis (Lung = 20, 70%, Primary tumour = 5, 10%, Liver metastasis = 1, 5%, Table A-4).

In summary, diagnostic tumour marker expression obtained by IHC from metastatic deposits were not different from those of primary tumours except for lung metastases where lower EGFR or higher vimentin expression was observed.

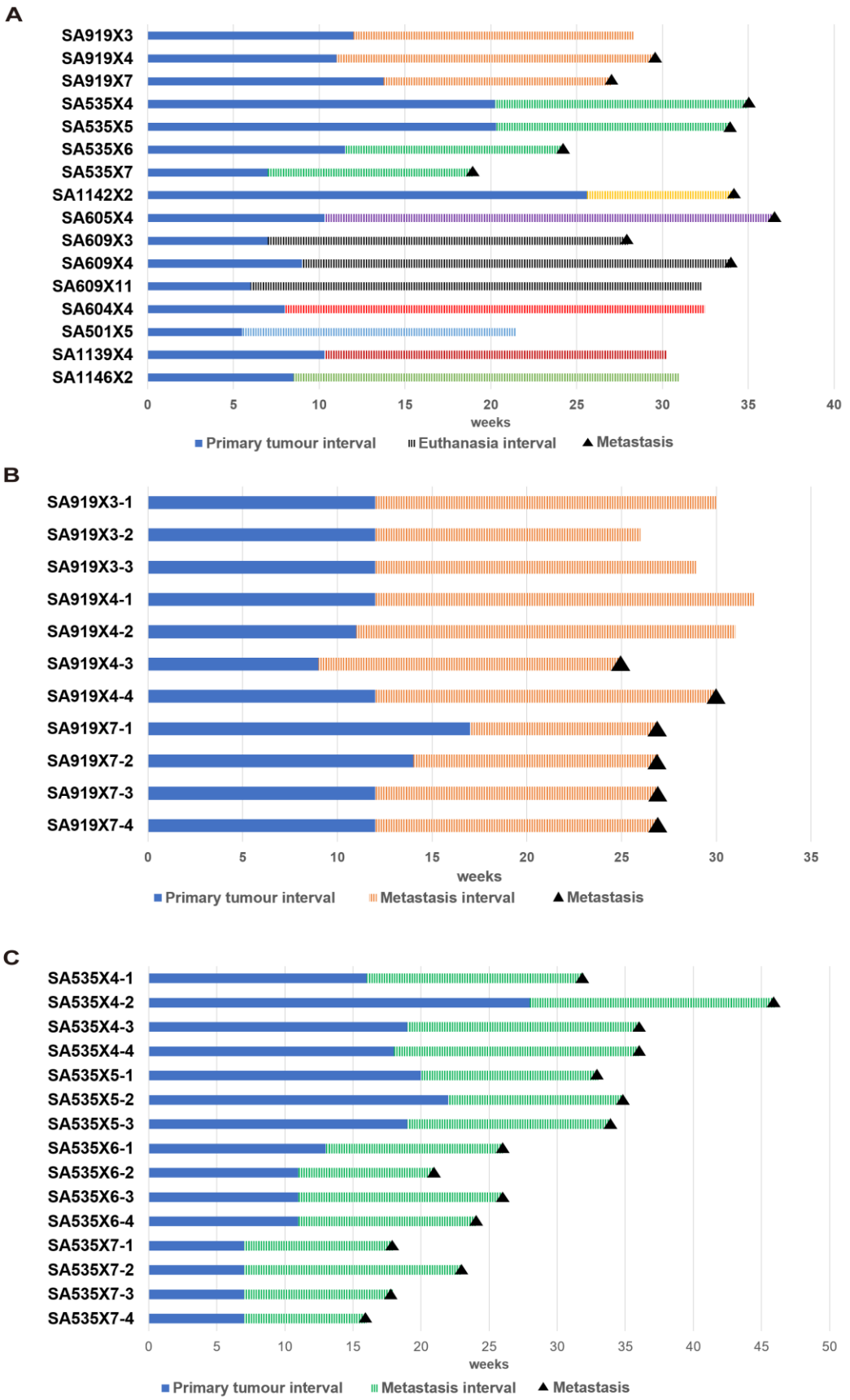

A
SA919X3
SA919X4
SA919X7
SA535X4
SA535X5
SA535X6
SA535X7
SA1142X2
SA605X4
SA609X3
SA609X4
SA609X11
SA604X4
SA501X5
SA1139X4
SA1146X2
0 5 10 15 20 25 30 35 40
weeks
Primary tumour interval
Euthanasia interval
Metastasis
B
SA919X3-1
SA919X3-2
SA919X3-3
SA919X4-1
SA919X4-2
SA919X4-3
SA919X4-4
SA919X7-1
SA919X7-2
SA919X7-3
SA919X7-4
0 5 10 15 20 25 30 35
weeks
Primary tumour interval
Metastasis interval
Metastasis
C
SA535X4-1
SA535X4-2
SA535X4-3
SA535X4-4
SA535X5-1
SA535X5-2
SA535X5-3
SA535X6-1
SA535X6-2
SA535X6-3
SA535X6-4
SA535X7-1
SA535X7-2
SA535X7-3
SA535X7-4
0 5 10 15 20 25 30 35 40 45 50
weeks
Primary tumour interval
Metastasis interval
Metastasis

Figure 3-8 Time to develop primary tumour and metastasis.

A. Time interval (weeks) from transplant to primary tumour resection (blue) and to euthanasia (other colours) in all PDXs. The mean of replicates (3-4) in each PDX line are presented. B. Time interval (weeks) in multiple passages of SA919. C. Time interval (weeks) in multiple passages of SA535. Horizontal axis = time (weeks), vertical axis = replicates, metastasis events = black triangle.

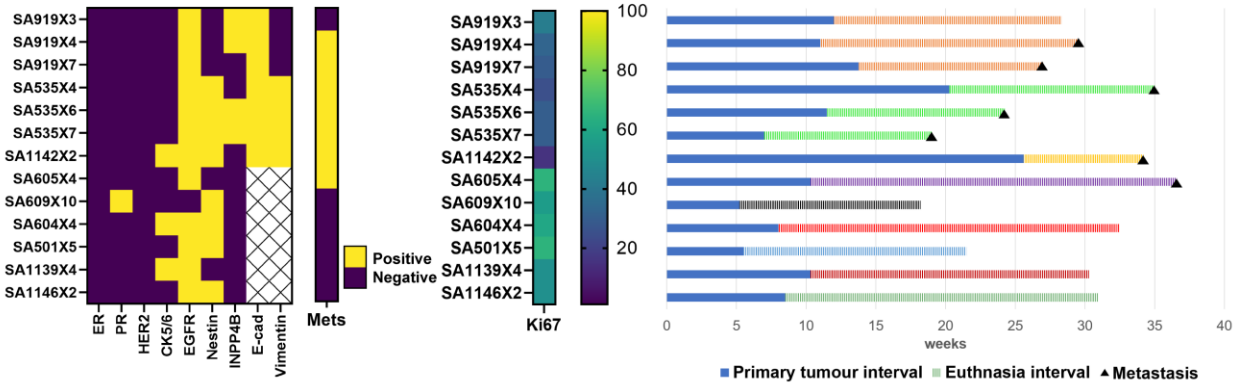


Figure 3-9 Immunohistochemistry staining results for PDXs and interval from transplants to tumour removal and euthanasia.

Left panel shows IHC results (positive or negative) for biomarkers in primary tumours of PDXs. Middle panel shows the mean Ki67 levels (percentage of positive cells). Right panel shows the interval between transplant, primary resection (blue) and endpoint (other colours). Horizontal axis = time (weeks), vertical axis = PDX line/passage, metastasis events = black triangle.

3.5 Tumour growth rate and metastatic sites are distinct for each PDX tumour

We defined primary tumour interval as time from transplant to primary tumour removal, metastasis interval as time from tumour removal to metastasis detection and euthanasia interval as time from tumour removal to euthanasia (Figure 3-2). Each PDX line showed different growth rate of primary tumour and each metastatic line exhibited distinct metastasis interval (Table 3-1, Figure 3-8 A). Primary tumour interval for all PDX tumours ranged from 5-28 weeks depending on the growth rate of each tumour.

Metastasis interval of 5 PDXs ranged from 5 to 18 weeks. We observed that metastasis interval varied between different PDXs and also within same PDXs. Since it is impossible to identify the exact time of metastasis development, we calculated time from primary tumour removal to metastasis detection for each PDXs. In SA919, metastasis interval was 16-18 weeks for passage 4 and 10-15 weeks for passage 7 (Table 3-2). In SA535, we observed metastasis interval ranging from 7 to 18 weeks in 4 different passages (Table 3-3). Metastases of both SA919 and SA535 were detected by palpation of masses with or without PET-CT. SA1142 developed general illness and weight loss during observation period and metastases to lung and liver were identified after histologic evaluation of organs. Metastasis interval was 8 weeks in all 3 replicates of SA1142 (Table 3-4). SA605 replicates were found to have micrometastasis to lung without any related symptoms and metastasis interval was 14-18 weeks. In SA609, one replicate from passage 3 were found to have micrometastasis to lung with metastasis interval of 21 weeks whereas passage 4 replicate with palpable mass in axillary and mammary fat pad area only had 5 weeks of metastasis interval.

Considering loco-regional and distant sites of metastasis, we noted that each tumour tended to metastasize to the same anatomical sites in repeated transplants. Loco-regional

metastasis occurred in SA535 (inguinal, tumour recurrence). Nodal and soft tissue metastasis occurred in SA919 (paraspinal metastasis) and SA535 (axillary). Visceral organ metastasis occurred in SA535 (lung), SA605 (lung) and SA1142 (lung, liver). There were also metastasis developments that occurred in low frequency in SA609X3 (liver; 1/4), SA609X4 (axillary, lung; 1/4), SA535X5 (liver; 1/3) and SA535X7 (intra-abdominal; 1/4) that needs more investigation with further replication (Table 3-1).

A natural question which arises is the possible relationship between organotypic metastasis in the PDX system, in comparison with TNBC patients, some of whom relapsed with metastatic disease. We observed that the development of metastasis in PDX was independent of patients' metastasis status in 4 lines (Table 3-1). SA535 and SA1142 developed metastasis in PDXs when originating patients showed no evidence of metastasis during the follow up. On the other hand, tumours of SA605 and SA501 showed metastasis in patients but not in PDXs. Three PDX lines showed metastasis development in both PDXs and patients with different organ involvement (SA919, SA605, SA609). Two lines showed no distant metastasis in both PDXs and patients (SA1139, SA1146) however, SA1146 developed tumour recurrence in 2 of 4 replicates. We noted that patients' tumour characteristics, such as cancer stage, histologic grade or lymphovascular invasion, did not affect PDX metastasis development (Table A-1).

Metastasis was not identified in 4 PDX lines; SA604, SA501, SA1139 and SA1146. Mice were observed at least 16 weeks after primary tumour removal, which is longer than the observation period in other study in the field [89]. The average interval between primary tumour removal and euthanasia were 24.5, 16, 20 and 22.5 weeks in SA604, SA501, SA1139 and SA1146, respectively. SA1146 developed tumour recurrence at mammary fat pad site in 2 of 4

mice, however there was no evidence of distant metastasis in other organs. Thus, these tumours are considered as non-metastasizing tumours.

Table 3-2 Metastasis development in multiple passages of SA919

PDX ID	Primary tumour interval (weeks)	Metastasis interval (weeks)	Number of metastasis	Metastasis site
SA919X3-1	12	18	0	None
SA919X3-2	12	14	0	None
SA919X3-3	12	17	0	None
SA919X3-4	NA	NA	NA	NA (lymphocytic tumour)
SA919X4-1	12	20	0	None
SA919X4-2	11	20	0	None
SA919X4-3	9	16	2	Abdominal, paraspinal (ventral)
SA919X4-4	12	18	2	Axillary, paraspinal (ventral)
SA919X7-1	17	10	2	Paraspinal (supra, ventral)
SA919X7-2	14	13	1	Paraspinal (supra)
SA919X7-3	12	15	1	Paraspinal (ventral)
SA919X7-4	12	15	1	Paraspinal (ventral)

NA = Not available

Table 3-3 Metastasis development in multiple passages of SA535

PDX ID	Primary tumour interval (weeks)	Metastasis interval (weeks)	Metastasis site					
			axillary	inguinal	lung	liver	abdominal	recur
SA535X4-1	16	16	+		+			+
SA535X4-2	28	18	++	+	+			
SA535X4-3	19	7	+		+			+
SA535X4-4	18	18	+	+	+			
SA535X5-1	20	13	+		+			+
SA535X5-2	22	13	+		+	+		+
SA535X5-3	19	15			+			+
SA535X5-4	NA	NA	No tumour development					
SA535X6-1	13	13	+					+
SA535X6-2	11	10	+		+	+		+
SA535X6-3	11	15	+				+	+
SA535X6-4	11	13			+			+
SA535X7-1	7	11	+		+		+	+
SA535X7-2	7	16	+		+			+
SA535X7-3	7	12			+			+
SA535X7-4	7	9	+		+			+

+ = 1 metastasis, ++ = 2 metastases

Table 3-4 Metastasis development in SA1142, SA605 and SA609

PDX ID	Primary tumour interval (weeks)	Metastasis interval (weeks)	Metastasis detection	Metastasis site
SA1142X2-1	28	8	General illness/weight loss	Lung
SA1142X2-2	28	8	General illness/weight loss	Lung, liver
SA1142X2-3	21	8	General illness/weight loss	Lung, liver
SA605X4-3	10	18	Histologic evaluation	Lung
SA605X4-4	12	14	Histologic evaluation	Lung
SA609X3-3	7	21	Histologic evaluation	Liver
SA609X4-1	11	5	Palpable mass	Axillary, lung, tumour recur

Table 3-5 IHC results for primary tumours and metastases from SA919

PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA919X3-1	Primary	-	-	-	3/40	-	+	+	-	-	+
SA919X3-3	Primary	-	-	-	3/45	-	+	+	-	-	+
SA919X4-1	Primary	-	-	-	3/25	-	-	+	-	-	+
SA919X4-2	Primary	-	-	-	3/30	-	-	+	-	-	+
SA919X4-3	Primary	-	-	-	3/35	-	+	+	-	-	+
SA919X4-3	Abdominal	-	-	-	3/40	-	+	+	-	-	+
SA919X4-3	Ventral spinal	-	-	-	3/35	-	+	+	-	-	+
SA919X4-4	Primary	-	-	-	3/40	-	+	+	-	-	+
SA919X4-4	Axillary	-	-	-	3/35	-	+	+	-	-	+
SA919X4-4	Ventral-spinal	-	-	-	3/25	-	-	+	-	-	-
SA919X7-1	Primary	-	-	-	3/35	-	+	+	-	-	-
SA919X7-1	Supra-spinal	-	-	-	3/35	-	+	+	-	+	-
SA919X7-1	Ventral-spinal	-	-	-	3/35	-	+	+	-	-	-
SA919X7-2	Primary	-	-	-	3/30	-	+	+	-	-	-
SA919X7-2	Supra-spinal	-	-	-	3/40	-	+	+	-	-	-
SA919X7-3	Primary	-	-	-	2/1	+	-	+	-	-	-
SA919X7-3	Ventral-spinal	-	-	-	3/35	-	+	+	-	-	-
SA919X7-4	Primary	-	-	-	3/50	-	+	+	-	-	-
SA919X7-4	Ventral-spinal	-	-	-	3/40	-	+	+	-	+	-

Table 3-6 IHC results for primary tumours and metastases from SA535 and SA1142

PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA535X4-1	Recur	-	-	-	3/15	-	+	+	+	-	-
SA535X4-1	Axillary	-	-	-	3/25	-	+	+	+	+	+
SA535X4-1	Lung	-	-	-	3/25	-	+	+	+	+	+
SA535X4-2	Primary	-	-	-	3/25	-	+	+	+	+	-
SA535X4-2	Axillary (R)	-	-	-	3/30	-	+	+	+	+	-
SA535X4-2	Axillary (L)	-	-	-	3/40	-	+	+	+	+	-
SA535X4-2	Inguinal	-	-	-	3/25	+	+	+	+	+	-
SA535X4-2	Lung	-	-	-	3/30	-	+	+	+	+	-
SA535X4-3	Primary	-	-	-	3/25	-	+	+	+	+	-
SA535X4-3	Recur	+	-	-	3/30	-	+	+	+	+	-
SA535X4-3	Axillary	-	-	-	3/25	-	+	+	+	+	-
SA535X4-3	Lung	-	-	-	NA	-	+	+	+	+	-
SA535X4-4	Primary	-	-	-	3/30	-	+	+	+	+	-
SA535X4-4	Axillary	-	-	-	3/35	-	+	+	+	+	-
SA535X4-4	Inguinal	-	-	-	3/30	-	+	+	+	+	-
SA535X4-4	Lung	-	NA	NA	NA	-	+	+	+	NA	+
SA535X6-1	Primary	-	-	-	3/30	-	+	+	+	+	+
SA535X6-2	Primary	-	-	-	3/35	-	+	+	+	+	+
SA535X6-3	Primary	-	-	-	3/35	-	+	+	+	+	+
SA535X6-4	Primary	-	-	-	3/35	-	+	+	+	+	+
SA535X7-4	Primary	-	-	-	3/30	-	+	+	+	+	+
SA535X7-4	Recur	-	-	-	3/50	-	+	+	+	+	+

PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA535X7-4	Axillary	-	-	-	3/50	-	+	+	+	+	+
SA1142X2-1	Primary	-	-	-	3/15	+	+	+	+	+	-
SA1142X2-1	Lung	-	-	-	3/20	+	+	+	+	+	-
SA1142X2-1	Liver	-	-	-	3/20	+	+	+	-	-	-
SA1142X2-3	Primary	-	-	-	3/15	+	+	+	+	+	-
SA1142X2-3	Lung	-	NA	-	NA	+	+	+	+	NA	+
SA1142X2-3	Liver	-	-	-	3/40	+	-	+	+	-	-

NA = Not available

3.6 Relationship between passage number, growth rate of PDX tumours and metastatic potential

Previous study has shown that serially propagated PDXs exhibited different clonal population in primary tumour [69]. We hypothesized that the different passage number of PDX tumour affects metastatic potential. To address this, we transplanted increasing chronological passages of primary tumours (SA919 and SA535) and evaluated the appearance of loco-regional and distant metastases. We observed that increasing number of passages affected metastatic potential in SA919 but not in SA535.

In SA919, metastatic potential appeared to increase with passage number (Table 3-2). Passage 3 did not show any metastasis, passage 4 developed metastasis in 2 of 4 replicates and passage 7 showed metastasis in all 4 replicates.

All 4 passages of SA535 showed metastatic potential that spread to both loco-regional and distant site. Metastases were repeatedly observed across replicates in the same sites of each passage. Most frequent metastatic sites were lung (13/16) and axillary (12/16). Tumour recurrence at mammary fat pad area was also observed in most of cases (13/16) (Table 3-3).

We set out to investigate tumour growth rate of each passage to identify whether it contributed to metastatic potential. We observed primary tumour and metastasis interval between and within several passages of SA919 and SA535. We noted that primary tumour intervals were relatively constant throughout the chronologically increasing passage in SA919, ranging from 9 to 17 weeks (Figure 3-8 B, Table 3-2) and did not affect metastatic potential of tumour cells. On the other hand, primary tumour interval decreased as chronological passage number increased in SA535 (Figure 3-8 C, Table 3-3). Primary tumour interval of passage 4 and 5 was 16 to 28 weeks while it was 11 to 13 weeks in passage 6 and 7 weeks in passage 7. However, primary

tumour interval did not affect metastatic potential of tumour cells in SA535 since all passages showed similar metastatic activity in terms of the number of metastasis (range 1-4) and metastasis sites (Table 3-3).

3.7 Phenotype change from non-basal-like to basal-like breast cancer occurred upon serial propagation of SA919 tumour but not in SA535

In addition to triple-negative status of ER, PR and HER2, several markers have been suggested to identify basal-like breast cancer [25, 113, 114]. We defined basal-like and non-basal-like breast cancer based on IHC staining results of CK5/6, EGFR, nestin and INPP4B. Basal-like breast cancer was determined when any of CK5/6, EGFR or nestin showed positive staining with negative staining for INPP4B [25, 26, 115]. Non-basal-like breast cancer was determined when all three markers of CK5/6, EGFR and nestin were negative and INPP4B was positive.

We observed the phenotype changes upon serial propagation in SA919. In earlier passage 3 and 4, primary and metastatic tumour cells were negative for CK5/6 and nestin and positive for INPP4B (Table 3-5). EGFR showed some level of staining in passage 3 and 4 (intensity 0-2, percentage of positive cells 0-25%, Table A-3) however, IHC scoring was significantly low compared to passage 7 (H-score calculated as intensity x percentage of positive cells, see Methods, Mann-Whitney test, $p = 0.002$, Figure 3-10 A). Thus, earlier passages of SA919 tumours are considered as non-basal-like breast cancer. Interestingly, passage 7 tumour cells showed high level of EGFR and negative INPP4B, indicating a phenotype change toward basal-like breast cancer which may have affected metastatic potential.

In SA535, passage 4 tumours showed minimal staining for INPP4B but passage 6 and 7 tumours showed positive staining with 100% of cells in all 7 samples (Table A-4). There was no EGFR positivity difference between passages (Table 3-6), although passage 6 and 7 showed higher H-score compared to passage 4 (Student's two-tailed t-test, p=0.0044, Figure 3-10 B). However, it should be noted that passage 6 and 7 had a small sample size (n=7). Moreover, since lung metastasis showed lower level of EGFR H-score (Figure 3-10 C) and we did not have lung metastasis staining result for passage 6 and 7, we compared H-score between passage 4 and passage 6, 7 without lung metastasis. EGFR H-score from passage 4 still showed statistically significantly lower level compared to passage 6 and 7, however to less extent (Mann-Whitney test, p=0.0189, Figure 3-10 D). It cannot be concluded that there is a difference of EGFR expression between earlier and later passage in SA535 because of small sample size. There was no difference of other IHC markers in multiple passages of SA535. It appears that the change in INPP4B expression, from negative to positive over the passages, did not affect the metastatic activity of tumour cells since different passages of SA535 showed similar pattern of metastasis.

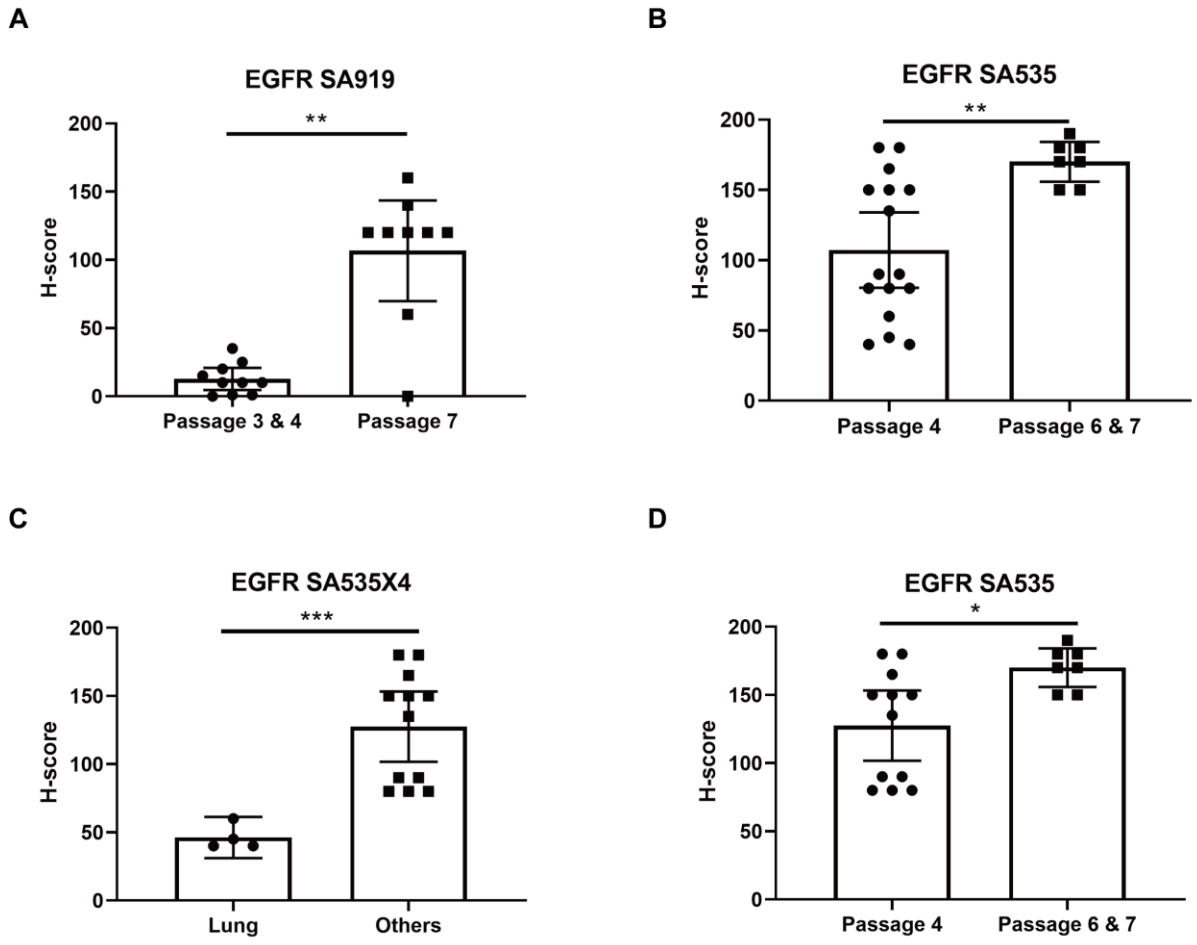


Figure 3-10 EGFR expression in SA919 and SA535 (H-score).

A. H-scores of passages 3 and 4 (n=10) are compared to that of passage 7 (n=9) in SA919 (Mann-Whitney test, p=0.002). B. H-scores of passage 4 tumours (n=16) are compared to that of passage 6 and 7 tumours (n=7) in SA535 (Student's two-tailed t-test, p=0.0044). C. H-scores of lung metastases are compared to that of other metastases and primary tumours from passage 4 in SA535 (Mann-Whitney test, p=0.0005). D. H-scores of passage 4 tumours without lung (n=12) are compared to that of passage 6 and 7 tumours (n=7) in SA535 (Mann-Whitney test, p=0.0189). H-score is calculated as [intensity x percentage of positive cells] from IHC staining. Student's two-tailed t-test or Mann-Whitney test for comparison: ***p<0.001; **p<0.005; *p<0.05.

3.8 Metastatic potential is maintained when metastatic PDX tumour cells are re-transplanted into MFP

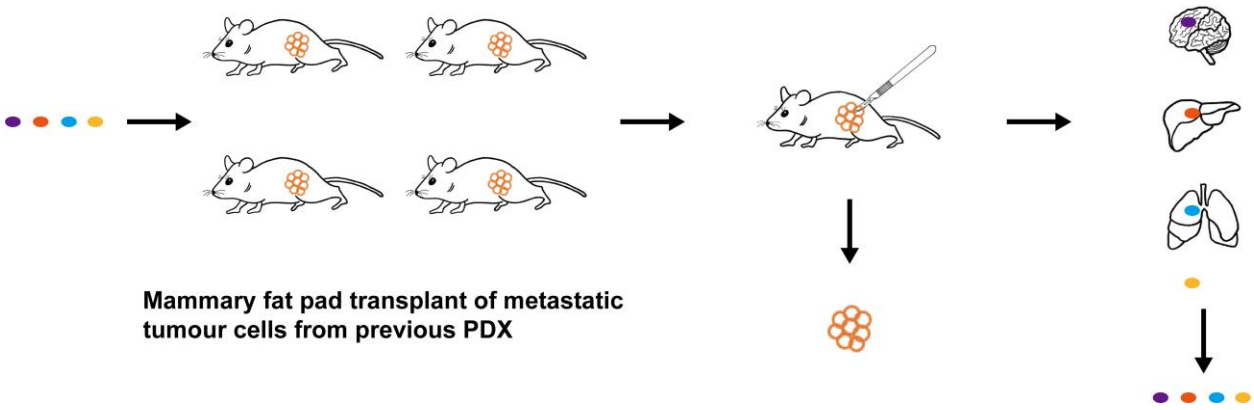


Figure 3-11 Schematic diagram of re-transplant experiment.

Metastatic tumour cells from previous patient-derived xenograft were re-transplanted into mammary fat pad and observed for development of primary tumour and metastasis.

Previous studies have shown that genomic and transcriptomic changes occur during metastatic progression [88, 116] and clonal diversity is reduced in metastasis as a result of selection [61]. We sought to investigate whether once formed metastatic cells, which underwent selection process during metastatic progression, maintain metastatic potential when transplanted back into MFP. Metastatic cells from SA919 and SA535 were re-transplanted into mammary fat pad to see whether metastatic activity is maintained in those cells (Figure 3-11).

In SA919, re-transplanted materials showed more aggressive behavior compared to primary tumour in terms of growth rate of tumour and metastasis (Figure 3-12 A, Table 3-7). Two of four re-transplanted mice developed metastasis to multiple sites including cervical, axillary and intra-abdominal area. One mouse died from sickness and only lung metastasis was identified. Another mouse showed substantial weight loss and paraspinal and lung metastases

were identified. Tumour growth rate of re-transplanted material was higher (primary tumour interval = 8-11 weeks) compared to original primary PDX tumour (primary tumour interval = 12-17 weeks).

In SA535, reduced metastatic activity was observed in re-transplanted cells. Metastatic cells from each of tumour recurrence and axillary metastasis were re-transplanted into MFP in four replicates. In tumour recurrence re-transplant, three mice developed tumour recurrence and lung metastasis. Only one mouse developed axillary metastasis. One mouse was excluded from analysis due to development of lymphocytic tumour. Interestingly, three of four axillary re-transplants developed tumour recurrence and no axillary metastasis. One mouse developed axillary metastasis only. Tumour growth rate and metastasis development rate were higher in tumour recurrence re-transplant (primary tumour interval = 10 weeks, metastasis interval = 6-14 weeks) compared to axillary re-transplant (primary tumour interval = 12-13 weeks, metastasis interval 20-25 weeks, Figure 3-12 B, Table 3-7). Aggressive behavior of re-transplanted metastatic cells was not observed in SA535.

We observed that metastatic potential was maintained when metastatic tumour cells were re-transplanted into mammary fat pad. However, metastatic behaviour in terms of number of metastasis and fidelity to original metastatic sites varied depending on PDX tumours.

Table 3-7 Metastasis in re-transplant experiment

Source	PDX ID	Primary tumour interval (weeks)	Metastasis interval (weeks)	Metastasis site
Paraspinal metastasis (from SA919X7-3)	SA919M-1	8	11	Lung
	SA919M-2	11	10	Paraspinal (supra), axillary, left kidney, lung
	SA919M-3	9	12	Paraspinal (supra), axillary
	SA919M-4	9	10	Paraspinal (supra), lung
Tumour recur (from SA535X4-1)	SA535R-1	10	14	Paraspinal (supra), lung, tumour recur
	SA535R-2	10	13	Axillary, lung, tumour recur
	SA535R-3	NA	NA	NA (lymphocytic tumour)
	SA535R-4	10	6	Lung, tumour recur
Axillary metastasis (from SA535X4-1)	SA535M-1	13	20	Lung, tumour recur
	SA535M-2	13	21	Lung, tumour recur
	SA535M-3	12	22	Lung, tumour recur
	SA535M-4	12	25	Axillary, lung

NA = Not available

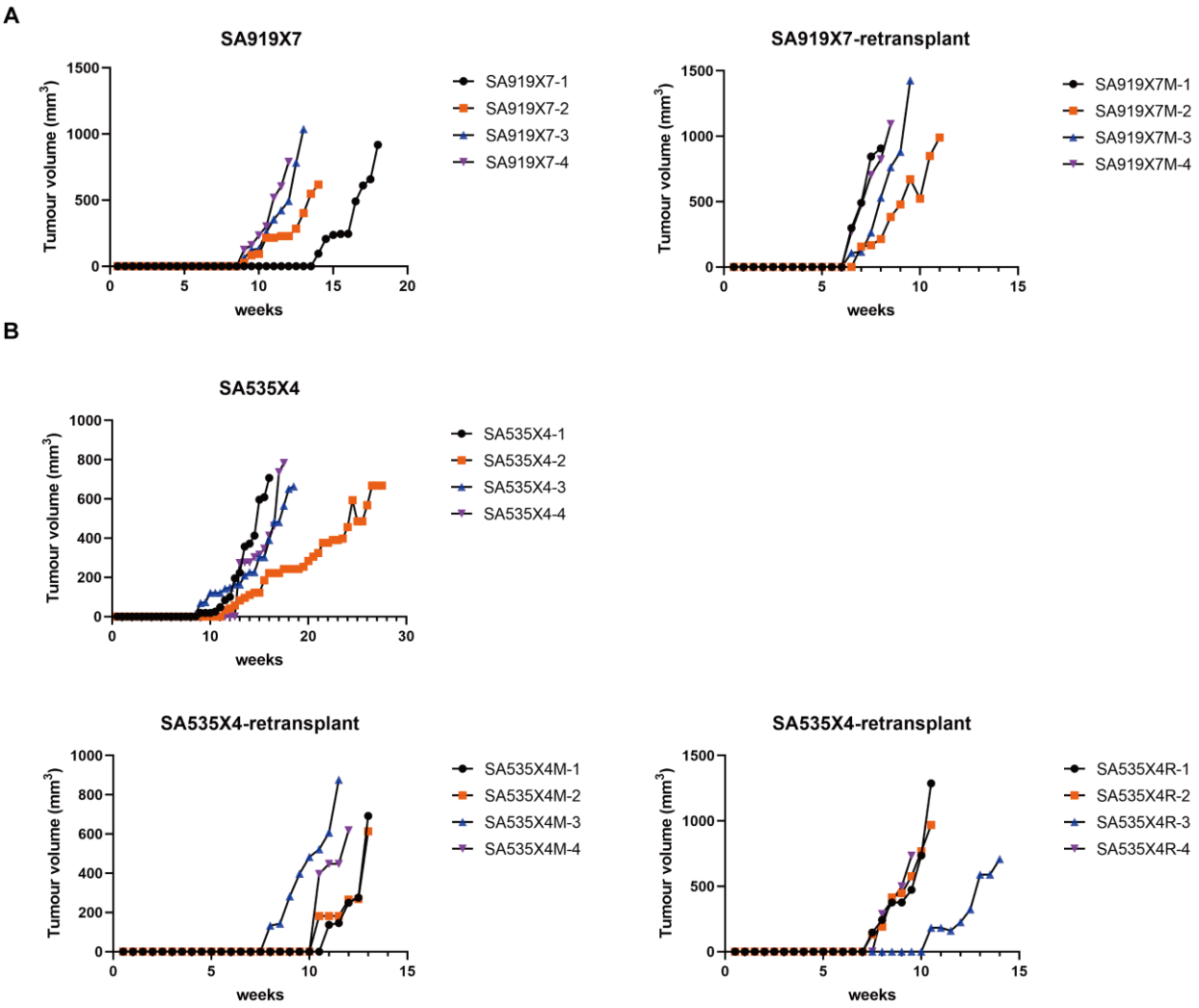


Figure 3-12 Growth curve for tumours from initial transplant of tumour cells and re-transplanted metastatic cells.

A. Growth curve for primary tumours of SA919X7 and re-transplant tumours (SA919X7-M). B. Growth curve for primary tumours (SA535X4) and re-transplant tumours from axillary metastasis (SA535X4-M) and tumour recurrence (SA535X4-R). Horizontal axis = time (weeks), vertical axis = tumour volume (mm^3), Colours = replicates.

Chapter 4: Clonal analysis of primary and metastases

Copy number aberration (CNA) form a major component of genomic instability in cancer and contribute to the somatic mutation landscape which underpins cancer ontogeny [78-80]. Since CNA are heritable, they may be used to infer clonal population structure in analysis of clonal dynamics during cancer progression [69]. Recently, the Aparicio lab has published scalable methods for WGS of single cells, DLP+ [77]. We measured copy number states of single primary and metastatic tumour cells (> 9,000 cells) in 2 PDXs (SA919 and SA535) using DLP+ to identify clonal relationship and evolutionary history. In this chapter, we will explore following findings (Figure 4-1):

1. **Metastasis pattern can be identified using single cell CNAs and phylogenetic analysis.** We observed polyclonal/monophyletic metastasis in SA919 and polyclonal/polyphyletic metastasis in SA535.
2. **Ability to metastasize is a property distributed across clones in both SA919 and SA535.** We observed that several clones in SA919 and SA535 primary tumours exhibited metastatic potential. We also showed that clones may have different degree of metastatic potential in SA919 (clone C appears to dominate metastatic potential when in competition with clones A or B).
3. **Ability to metastasize increased with passage number as CNA-determined clones evolve in SA919.** Passage 7 tumours showed higher metastatic potential (4/4) compared to passage 3 (0/3) and 4 (2/4).
4. **Metastatic clones derived from the subset of polyclonal primary tumour population in SA535.**

5. **Metastasis to specific anatomical site is not associated with genomic clones.** Different clones were observed in metastases to the same anatomical site in SA535.

6. **Genotype/LOH have potential impact metastatic potential of clones.** We observed genotype difference between clones in SA919. Also, genotype/LOH difference between primary and metastatic clones was observed in SA535.

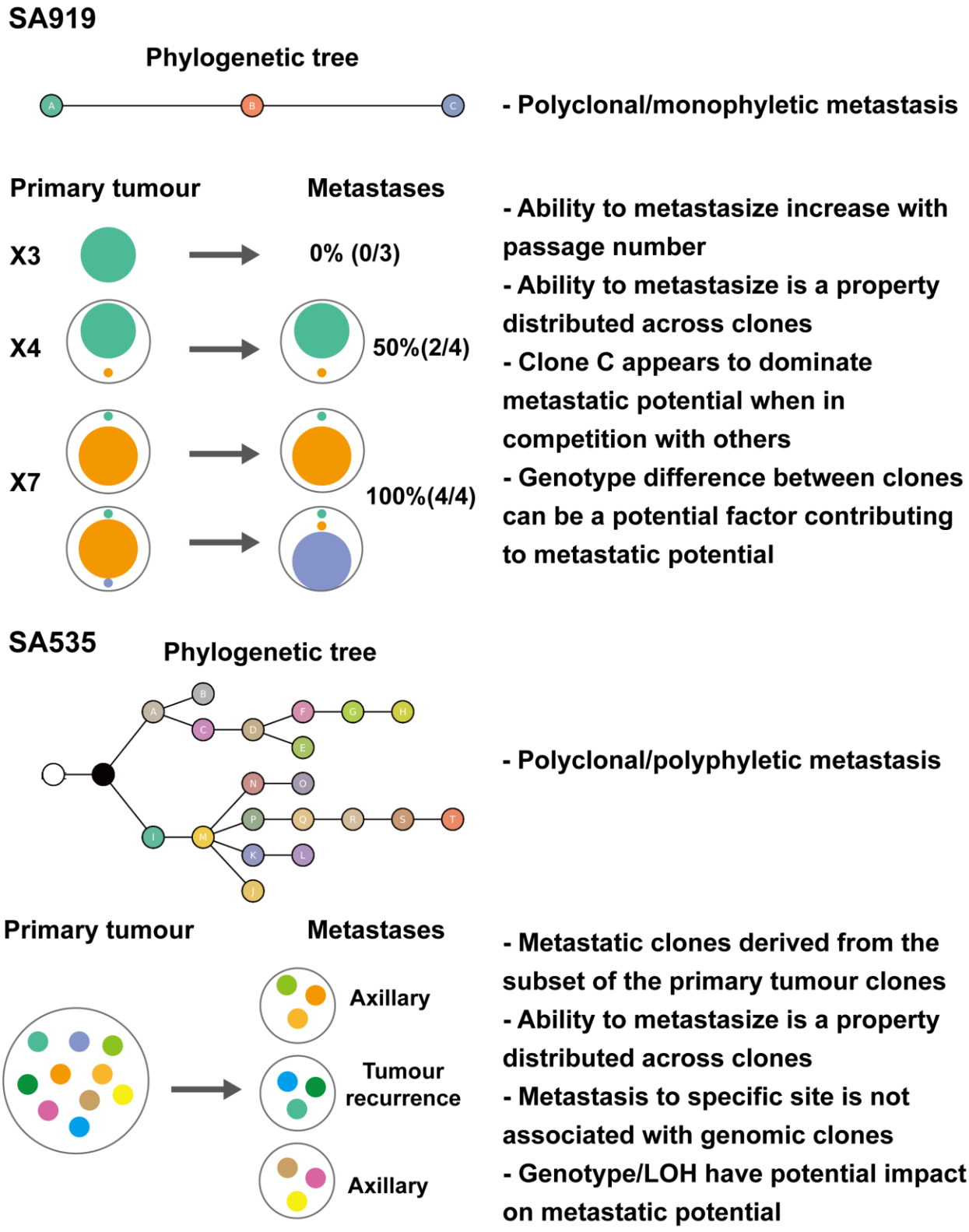


Figure 4-1 Chapter 4 summary

Chapter 4 describes the findings from clonal analysis of primary and metastatic tumour cells using single cell analysis. We used 2 PDXs (SA919 and SA535) to identify clonal relationship and evolutionary history. In SA919, metastasis pattern was polyclonal/monophyletic and the ability to metastasize of primary tumour increased with passage number. We also observed that when primary tumour clones are in competition, certain clones appear to show higher metastatic potential. In SA535, metastasis pattern was polyclonal/polyphyletic and metastatic clones derived from the subset of primary tumour clones. We observed that metastasis to specific anatomical site was not associated with genomic clones. In both PDXs, we observed that metastatic potential was distributed across clones and CNAs have potential impact on metastatic potential of clones.

4.1 Single-cell analysis revealed clonal population structure and clonal dynamics between primary tumour and metastases in SA919

To determine clonal population structure in primary and metastatic cancers, we used DLP+ to sequence and determine the genomic copy number states of single genome. From the DLP+ sequencing, individual CNA states were determined as previously described [77]. In order to determine population structure, a phylogeny was inferred using BREAKTREE [99], a novel scalable Bayesian tree inference method. We also measured allele-specific copy number of tumour cells. Allele-specific copy number, which indicates the number of copies of each homolog of chromosome, can provide more accurate information regarding copy number states since CNA affects one allele of genomic region located on either of the two homologous chromosomes in diploid human genome. Allele-specific analysis assigns a state of copy number (A, B) where A and B are the copy numbers of the two alleles. Allelic imbalance or B-allele frequency (BAF) is calculated as B/(A+B) where 0 or 1 means loss of one allele or LOH. Analysis of allele-specific copy numbers can identify copy-neutral loss of heterozygosity (LOH), where one allele is lost and the other is duplicated, that would otherwise remain as diploid in total copy number.

From 7 primary tumours and 7 metastases samples of passage 3, 4, and 7 SA919, we sequenced 19,289 cells (9,733 primary tumour cells, 9,556 metastatic cells) and 7,057 cells (2,537 primary tumour cells, 4,520 metastatic cells) were used for subsequent analysis after filtering out low quality cells (methods). For phylogenetic analysis, additional filtering removed cells with highest copy number jumps resulting in 5,471 cells (2,014 primary tumour cells, 3,457 metastatic cells). DLP+ sequencing identified 3 major clones based on copy number of the genome in SA919 tumour cells (Figure 4-2, 4-3 A). Clones differed from each other by copy

number of chromosomes 7, 5p or 10p. Clone A had diploid chromosome 7 while clone B and C had a mean copy number of 3 in chromosome 7. Clone C had additional amplifications on chromosomes 5p and 10p with mean copy number of 4 in each region (Figures 4-2, 4-3 A). Allele-specific copy number of tumour cells showed that clones shared allele-specific CNA features and allele-specific CNA regions were entirely concordant with the total CNA regions (Figure 4-3 E).

4.1.1 Increasing metastatic potential over the passages can be caused by CNAs

Next, we sought to compare the clonal population structure of primary tumours from multiple passages of SA919 since we observed different metastatic activity from each passage. We analysed 2,014 primary tumour cells from all 3 passages, including passage 3 (1 replicate), which did not show any metastasis, passage 4 primary tumour cells from 2 replicates, one with metastasis and the other without metastasis and 4 replicates of passage 7 from which all replicates developed metastasis.

The majority of primary cells from passages 3 and 4 were identified as clone A (>99%), whereas passage 7 primary cells were mostly identified as clones B (92.5%), indicating that clonal dynamics had occurred during passaging (Figure 4-2). Copy number states of the single-cell genome from each clone showed that clone B acquired additional amplification in chromosome 7 compared to clone A. Clone C acquired further additional amplification in chromosome 5p and 10p compared to clone B (Figure 4-2, red arrows). Clone B was the dominant clone of passage 7 primary tumours (SA919X7-1 = 95.3%, SA919X7-2 = 98.3%, SA919X7-3 = 93.9%, SA919X7-4 = 65.8%) and clones A (4.7%, 1.1%, 2.0%, 33.2%, respectively) and C (0%, 0.6%, 4.1%, 1.1%, respectively) were minor population (Figure 4-5).

A phylogenetic tree of all data (Figure 4-3 B) consists of primary tumour cells from 3 different time points (passage 3, 4, 7, Figure 4-3 C) and metastatic cells from 2 different time points (passage 4, 7, Figure 4-3 D). We noted that the proportion of clones B and C were different between primary tumours and metastases in the branches of the tree from the phylogeny inferred jointly over all primary and metastatic cells and some cells were only observed in metastases (Figure 4-3 C, D, shaded area). We observed that most of metastatic cells from passage 4 were assigned to clone A (99.8%) and most of passage 7 metastatic cells were found in clones B and C (54.3%, 45%, respectively, Figures 4-3 A, 4-4 C-G). Allele-specific copy number of tumour cells showed that cells in each clone shared allele-specific CNA features and allele-specific CNA regions were entirely concordant with the total CNA regions (Figure 4-3 A, E). This indicates that the difference of total copy number between clones was due to allele-specific CNA events. The mean percentage of LOH per cell in all cells of SA919 was 23.5% and mean percentage for each clone was 23.3%, 24.0%, 22.6% for clones A, B, C, respectively.

Next, we measured clonal proportion for each replicate and plotted cells into an existing phylogenetic tree that was already inferred jointly over all cells. Primary tumours of passage 3 (SA919X3) and 4 (SA919X4-1) that did not develop metastasis were comprised entirely of clone A cells (Figure 4-4 A, B). Passage 4 primary tumour that developed metastasis (SA919X4-4) was composed of clone A as a dominant clone (98.8%) with clone B as a minor clone (1.2%) (Figure 4-4 C). Axillary metastasis also had clone A as a dominant clone (99.7%) with minor clone B (0.3%) while ventral-spinal metastasis had only clone A. In SA919X7-1 and SA919X7-4, primary tumours were comprised largely of clone B (95.3% (n=142), 65.8% (n=121), respectively) with minor clone A (4.7% (n=7), 33.2% (n=61), respectively) and metastases were dominated by clone B (99.6% (n=761) and 98.4% (n=305) in SA919X7-1, 100% (n=30) in

SA919X7-4, Figure 4-4 D, G). SA919X7-4 also has a small portion of clone C in primary tumour (1.1%, n=2), but it was not present in metastasis. However, it should be noted that only a small number of tumour cells (primary = 184 cells, metastasis = 30 cells) were analyzed from SA919X7-4 (Figure 4-5). In ventral-spinal metastasis of SA919X7-1, clone C was detected at low prevalence (0.3%, n=1) which was not present in primary tumour. It is possible that clone C appeared as a new clone in this metastasis or was present in primary tumour as a rare cell population but not detected in the analysis. Primary tumours of SA919X7-2 and SA919X7-3 had clones A, B and C with clone B comprising the majority of primary tumour (proportion of clones A, B, C in the primary tumour of SA919X7-2, 4.7% (n=7), 98.3% (n=648), 0.6% (n=4); SA919X7-3, 0.2% (n=11), 93.9% (n=504), 4.1% (n=22), respectively). However, we observed that clone C became the major clone in metastases while clones A and B contributed to a lesser degree and became minor clones (proportion of clones A, B, C in metastasis of SA919X7-2, 2.2% (n=6), 9.7% (n=27), 88.2% (n=246); SA919X7-3, 0.3% (n=2), 3.4% (n=25), 96.3% (n=706), respectively, Figure 4-4 E, F).

We noted that primary tumour that developed metastases in earlier passage (SA919X4-4) had a small number of clone B (1.2%, n=1), which was not present in tumour without metastasis (SA919X4-1). In all primary tumours of passage 7, clone B was present as a dominant clone with clones A and C being the minor clones. In SA919X4-4, clone A dominated the metastatic clonal population while in passage 7, clone A appeared in metastasis as a minor clone. We also noted that clone C that was present in the primary tumours as a very minor clone, a few cells only, became the dominant clone in the metastases in 2 replicates of passage 7 (SA919X7-2, SA919X7-3).

Taken together, we observed that clonal populations of primary tumours evolve over the passages as new clones emerge in later passages. Moreover, a minor clone (clone C) in primary tumours of the later passage, that emerges alongside with clone B, appears to have increased metastasis potential over clone A as clone A is present only as a minor population in the later passage metastases. Increasing metastatic potential of primary tumours over chronological passages can be explained by higher proportion of clone B and emergence of clone C in later passage. However, more replicates are needed to identify fitness advantage difference between clones.

4.1.2 Copy number genotype and LOH difference between clones may affect metastatic potential

Next, we compared the genotype of clones by measuring median copy number difference in chromosome positions between clones to identify the genes that may have affected metastatic potential (Figure 4-7). First, we detected genome positions with copy number difference between two clones and then mapped copy number genome positions with ensemble gene ID. Then, we identified genes that were described in the hallmark gene sets to explore associated pathways [116]. We compared clone A (n=570) and B (n=1,476) of primary tumours and identified 37 genes were associated with CNV differences. Among them, 8 genes (HDAC9, AUTS2, MAGI2, CD36, RELM, DOCK4, GRM8, EXOC4) were identified in hallmark gene sets (Figure 4-6 A). When we compared clone A and B of metastases, AKT3 (at chromosome 1) was identified in addition to those 8 genes (Figure 4-6 B). In clone C, where additional amplification of chromosome 5p and 10p were observed compared to clone B, 5 genes (AKT3 at chromosome 1, ADCY2, TRIO at chromosome 5, CELF2, CAMK1D at chromosome 10) were identified in amplified region (Figure 4-6 C). We also compared primary (n=570) and metastatic (n=1,354)

clone A but did not identify any different hallmark genes (Appendix figure B-1 A). We observed 3 hallmark genes (AKT3, KIF26B, SMYD3) that were associated with copy number difference between primary and metastatic clone B (Appendix figure B-1 B). However, it should be noted that copy number change does not always correlate with gene expressions and further analysis of transcriptomic data is required to identify the relationship between copy number changes and gene expressions as well as their impact on metastatic potential of clones.

We also measured LOH events across the genome that contributed to copy number difference between clones to identify whether LOH affected metastatic potential of tumour cells. First, we detected LOH events that are present in both clones and then identified the number of LOH events accompanied with copy number difference (loss or gain) between two clones. LOH events were defined as follows: deletion LOH = LOH with copy number < 2, amplification LOH = LOH with copy number > 2, neutral LOH = LOH with copy number =2. When we compared clones A (n=570) and B (n=1416) from primary tumours, we identified 4 LOH events (0.45%) with CNAs (in chromosome 7) in clone B from the total of 891 LOH events observed in both clones (Figure 4-7 A). Compared to metastatic clone A (n=1354), we observed only 1 LOH events with copy number changes (0.12%) in metastatic clone B (n=1150) from total of 825 LOH events observed in both clones (Figure 4-7 B). Metastatic clone B showed 1 copy neutral LOH in chromosome 3 where metastatic clone A had deletion LOH with total copy number of 1. Similarly, metastatic clones B (n=1,150) and C (n=953) were almost identical except for 1 copy neutral LOH (0.12%) with total copy number of 1 observed in chromosome 3 in metastatic clone B. Since these LOH events are negligible, there is no difference in terms of LOH between metastatic clones A, B and C. Overall, there was only a small number of LOH events that contributed to copy number difference between clones in SA919.

4.1.3 Metastatic potential is distributed across genomic clones

We also observed that in earlier passages of SA919 metastasis developed from primary tumour clones without any evidence of copy number evolution underpinning the cell populations. Clone A was identified as a major clone in both primary tumour and metastases in SA919X4-4 (Figure 4-4 C). However, it should be noted that metastasis developed in the presence of clone B in SA919X4-4 primary tumour whereas SA919X4-1 primary tumour, which was comprised of only clone A, did not develop metastasis. It is not clear whether or how the presence of clone B affected the metastatic potential of tumour cells and it needs to be investigated further.

In SA919X7-1, we observed that both clones A and B of primary tumour were present in metastases, although clonal proportion of clone A was very low (0.004% and 0.013% in each metastasis). In SA919X7-2 and SA919X7-3, all 3 clones were present in both primary tumours and metastases indicating that metastatic potential is distributed across tumour cells/clones. However, growth advantage or fitness of clones in primary tumour or metastasis may differ as shown in different clonal proportions between primary tumour and metastasis.

From SA919 tumours, we discovered that (i) the ability to metastasize in primary tumour increases with the passage number due to the evolution of clonal population, (ii) the ability to metastasize is a property distributed across all observed clones, (iii) when clones are in competition in primary tumour, clone C appears to strongly dominate the metastatic potential, (iv) the genotype difference between clones can be a potential factor contributing to metastatic potential.

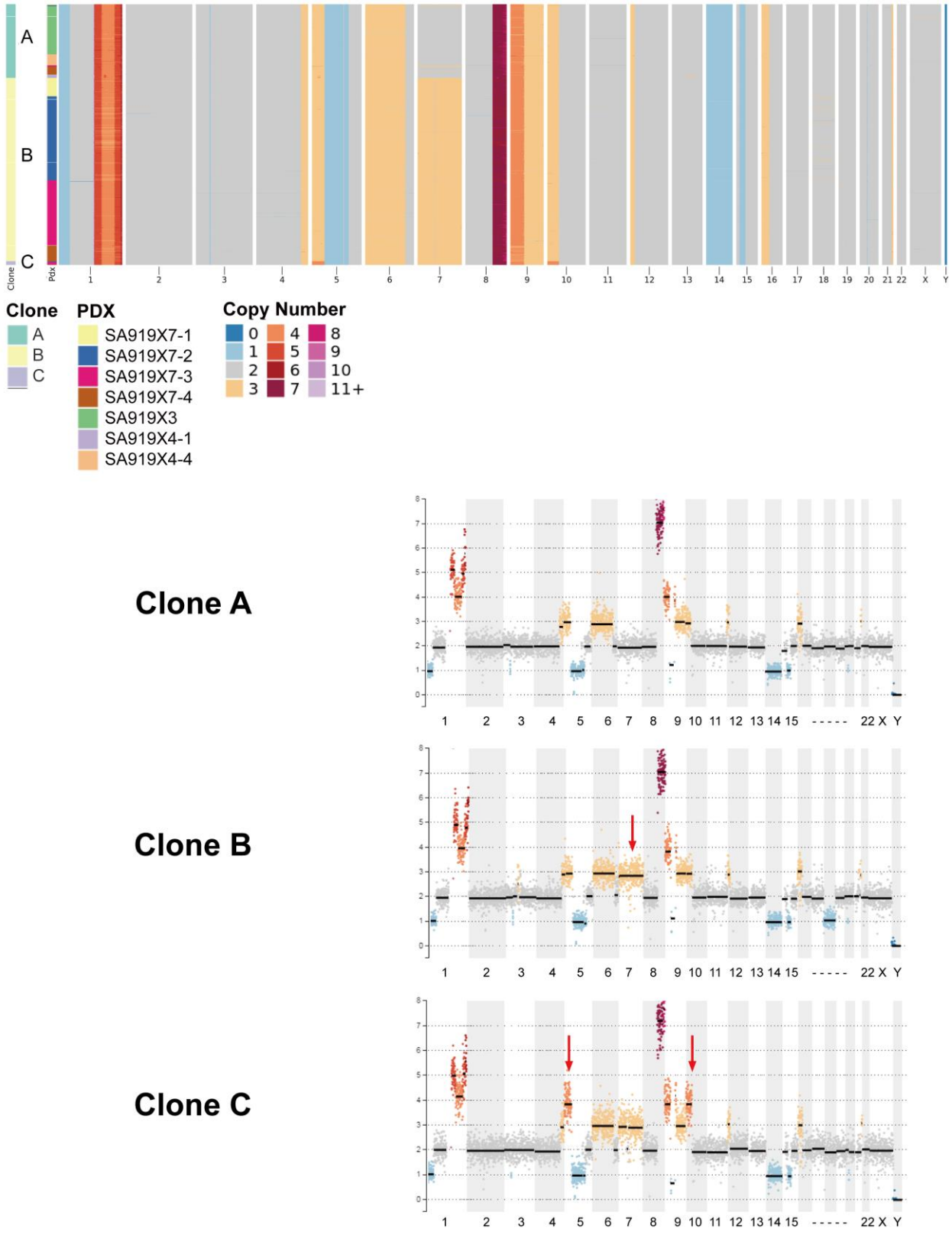


Figure 4-2 Copy number states of primary tumours of different passages from SA919.

Heatmap of single cell copy number states, grey = diploid, copy gain/loss heat colours as per key, chromosome/genome position horizontal axis. Cells are labeled according to clone (A-C) and originating PDX which is displayed on left side of the heatmap. Single cell copy number profiles for clones are shown in lower panel. Black lines = the median. Red arrows indicate the regions of copy number difference between clones.

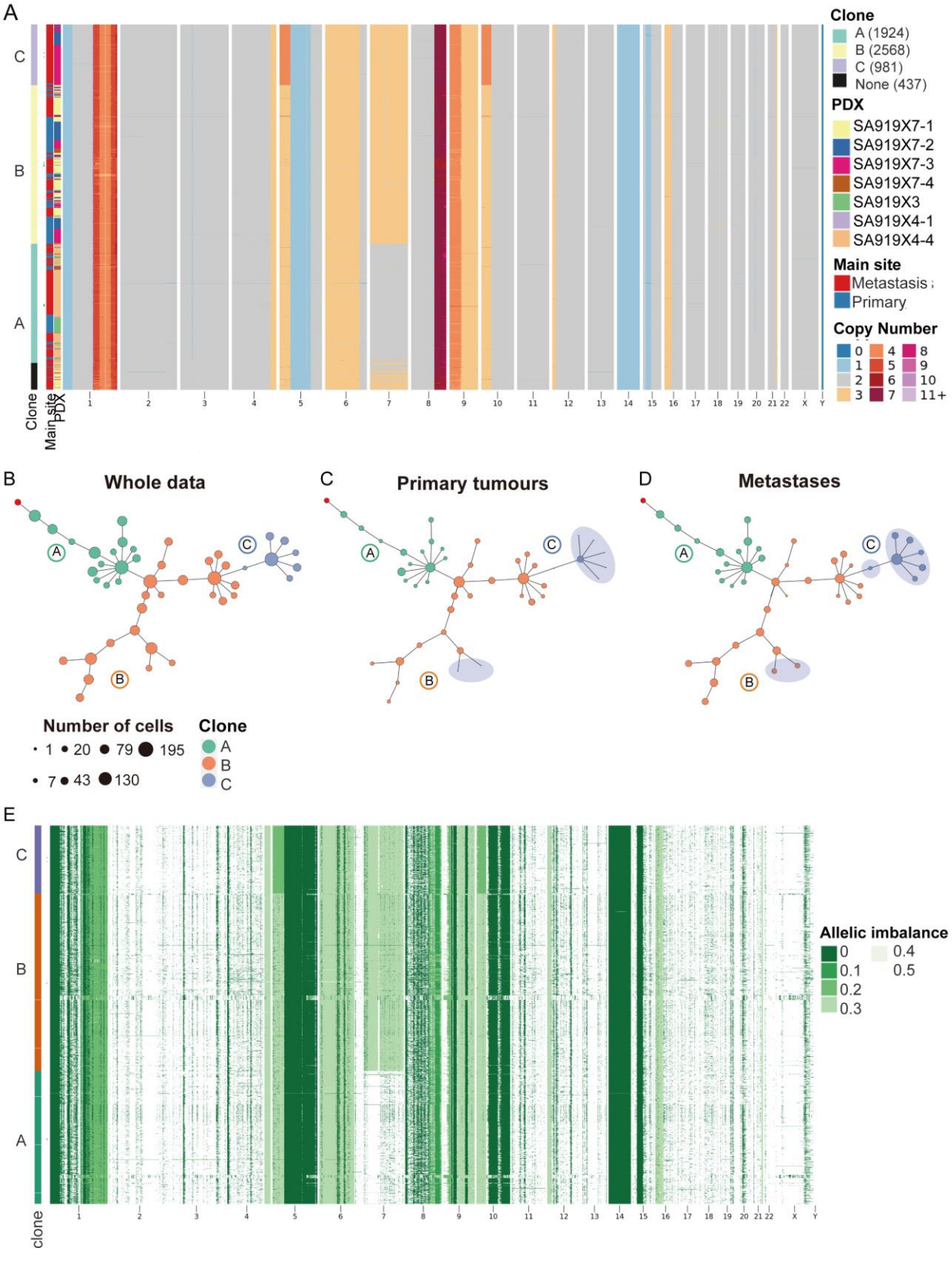

A
C
B
A
Clone
Main site
PDX
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Clone
A (1924)
B (2568)
C (981)
None (437)
PDX
SA919X7-1
SA919X7-2
SA919X7-3
SA919X7-4
SA919X3
SA919X4-1
SA919X4-4
Main site
Metastasis
Primary
Copy Number
0 1 2 3 4 5 6 7 8 9 10 11+
B
Whole data
C
Primary tumours
D
Metastases
A
B
C
Number of cells
1 20 79 195
7 43 130
Clone
A
B
C
E
C
B
A
clone
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
Allelic imbalance
0
0.1
0.2
0.3
0.4
0.5

Figure 4-3 Phylogenetic analysis of primary and metastatic tumour cells from combined data of SA919

A. Heatmap of single cell copy number states, grey = diploid, copy gain/loss heat colours as per key, chromosome/genome position horizontal axis. A phylogenetic tree was jointly inferred by BREAKTREE over all primary and metastatic tumour single-cell data from passage 3 (SA919X3), 4 (SA919X4) and 7 (SA919X7) replicates. The origins of tumour cells (primary vs. metastasis, replicates) and clones are separately labeled. A phylogenetic tree, which identified 3 different clones, is presented in B-D. B. whole data, C. primary tumours, D. metastases. Each dot (nodes of trees) represents the cells (size of dots showing the number of cells) assigned to each clone (colours). Shaded area highlights where the clonal proportion is different between primary tumours and metastases. E. Heatmap representation of allelic imbalance score. Allelic imbalance score is calculated as b/(a+b) where a and b are copy number of each allele. 0 = loss of one allele or LOH. 0.5 = equal number of copies in each allele. Allelic imbalance score as per key, chromosome/genome position horizontal axis.

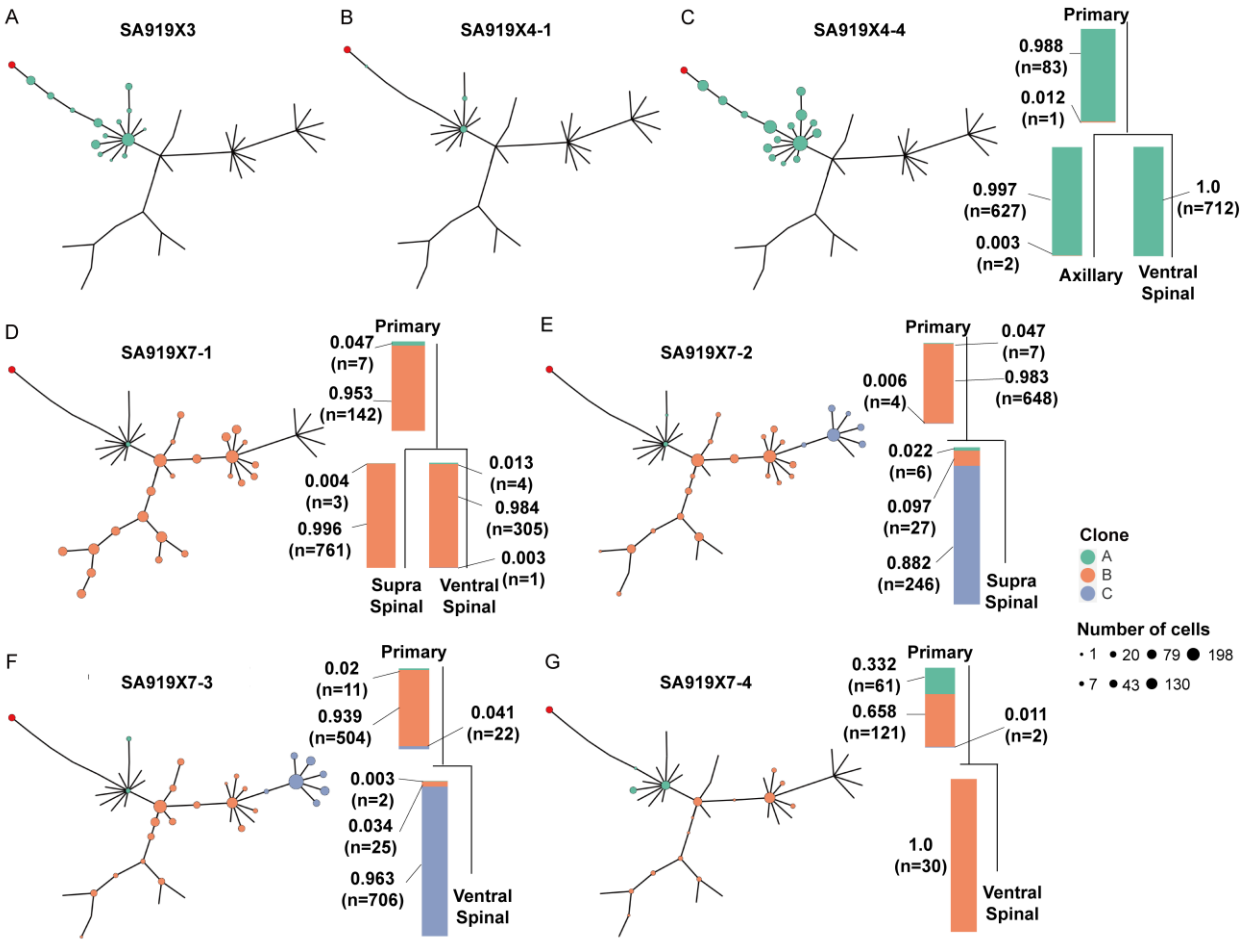


Figure 4-4 Phylogeny and clonal prevalence of primary and metastatic tumour cells of SA919

Phylogenetic trees with number of cells for A. SA919X3 (no metastasis), B. SA919X4-1 (no metastasis). Phylogenetic trees and clonal prevalence (bar-plot) of primary tumour and metastasis in C. SA919X4-4, D. SA919X7-1, E. SA919X7-2, F. SA919X7-3, and G. SA919X7-4. Clonal fractions are described for each clone in bar-plots.

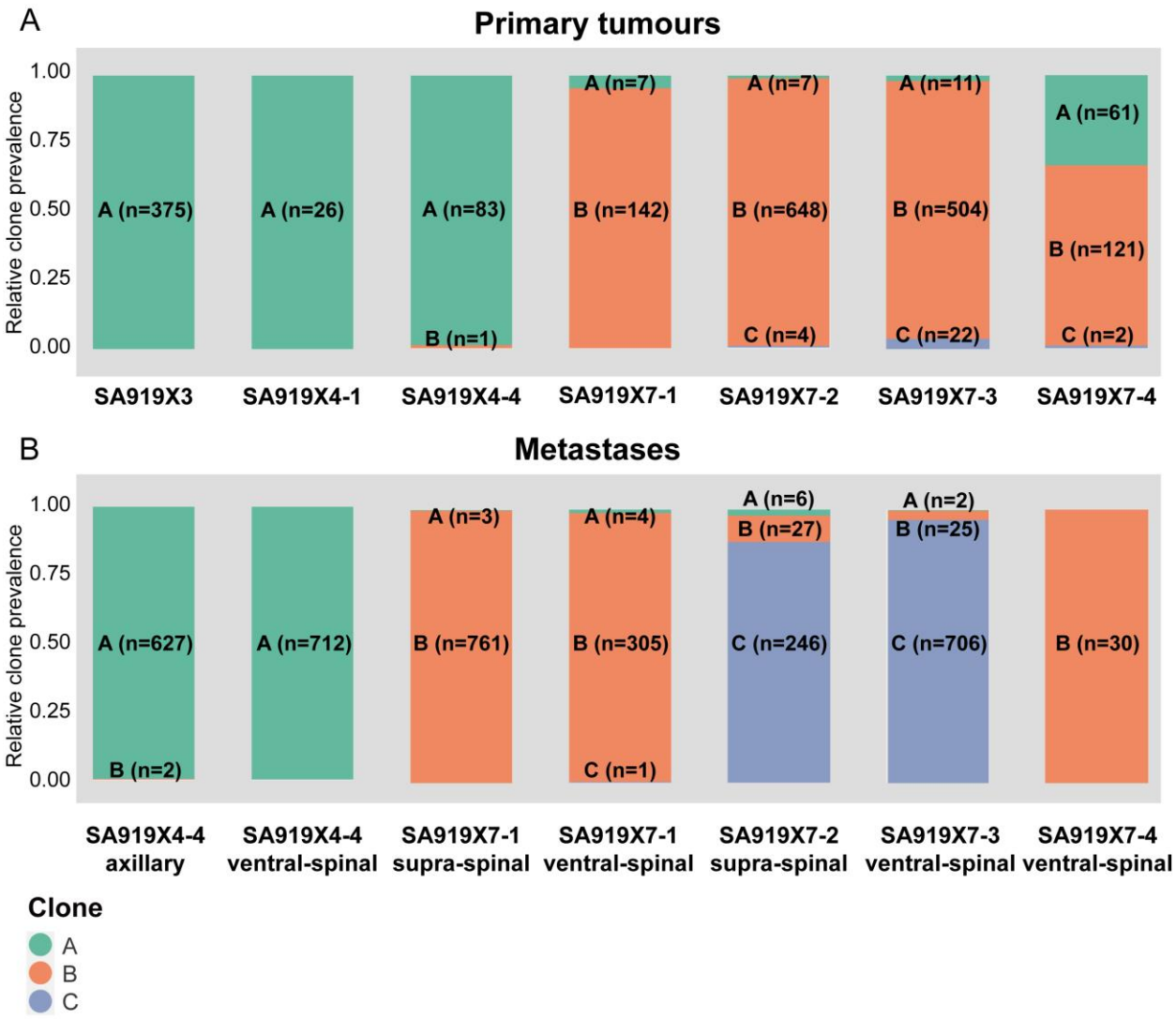


Figure 4-5 Absolute and proportional clonal abundance in SA919 primary and metastatic cancers

Relative clone prevalence (y-axis) of each replicate (x-axis) is described in bar-plot and the number of cells for each clone is shown in each bar. A. Primary tumours of multiple passages of SA919 are shown. SA919X3 and SA919X4-1 did not develop metastasis. B. Metastases in each replicate of multiple passages are shown. Colours = clones.

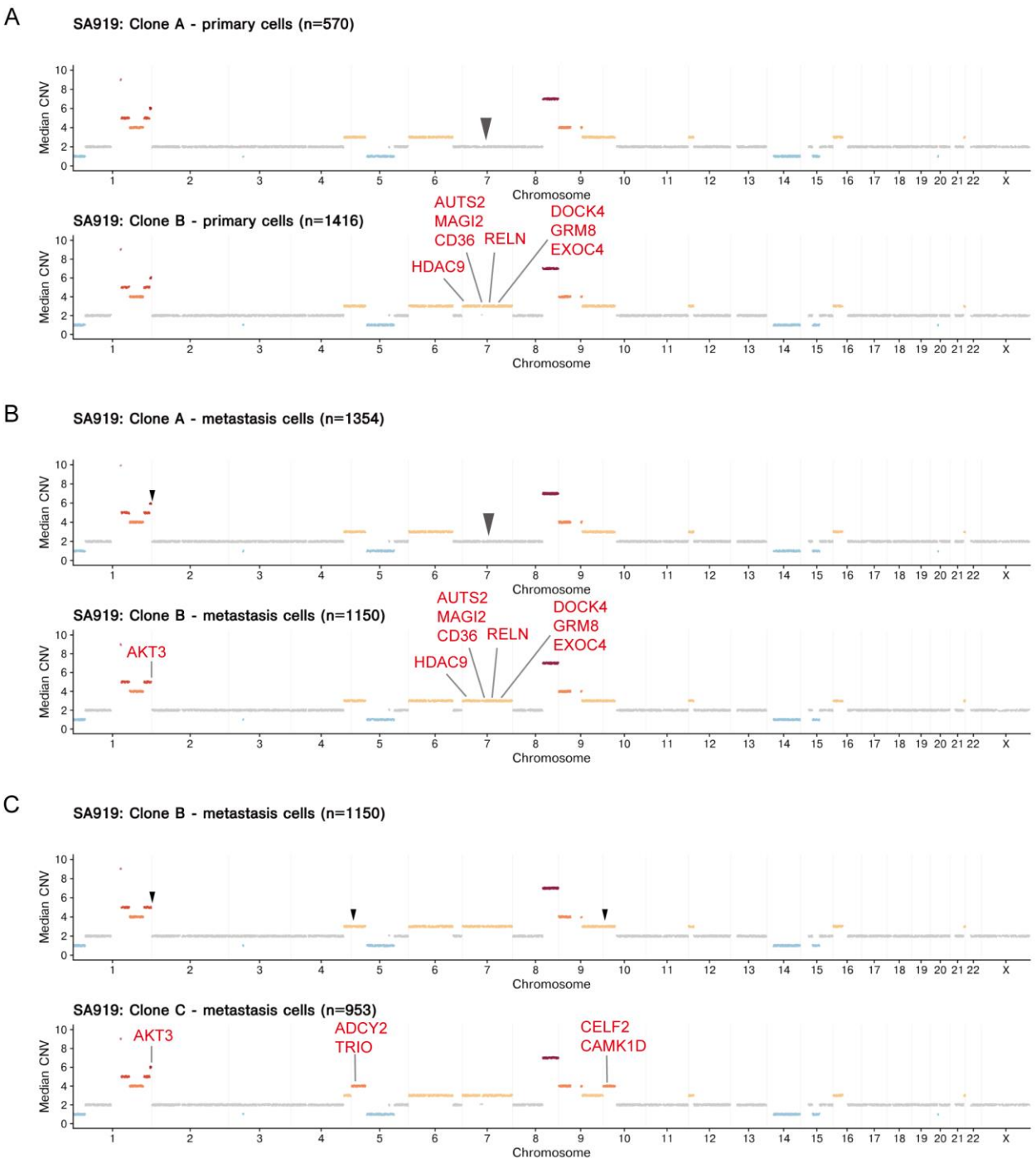


Figure 4-6 Copy number genotype comparison between clones in SA919.

Median copy number state (y-axis, median CNV) across the genome (x-axis) is compared between clones. A. Genotype of clones A and B from primary tumours. B. Genotype of clones A and B from metastases. C. Genotype

of clones B and C from metastases. Colours = copy number state, arrowhead = regions of copy number difference between clones, CNV = copy number variance, genes described in red.

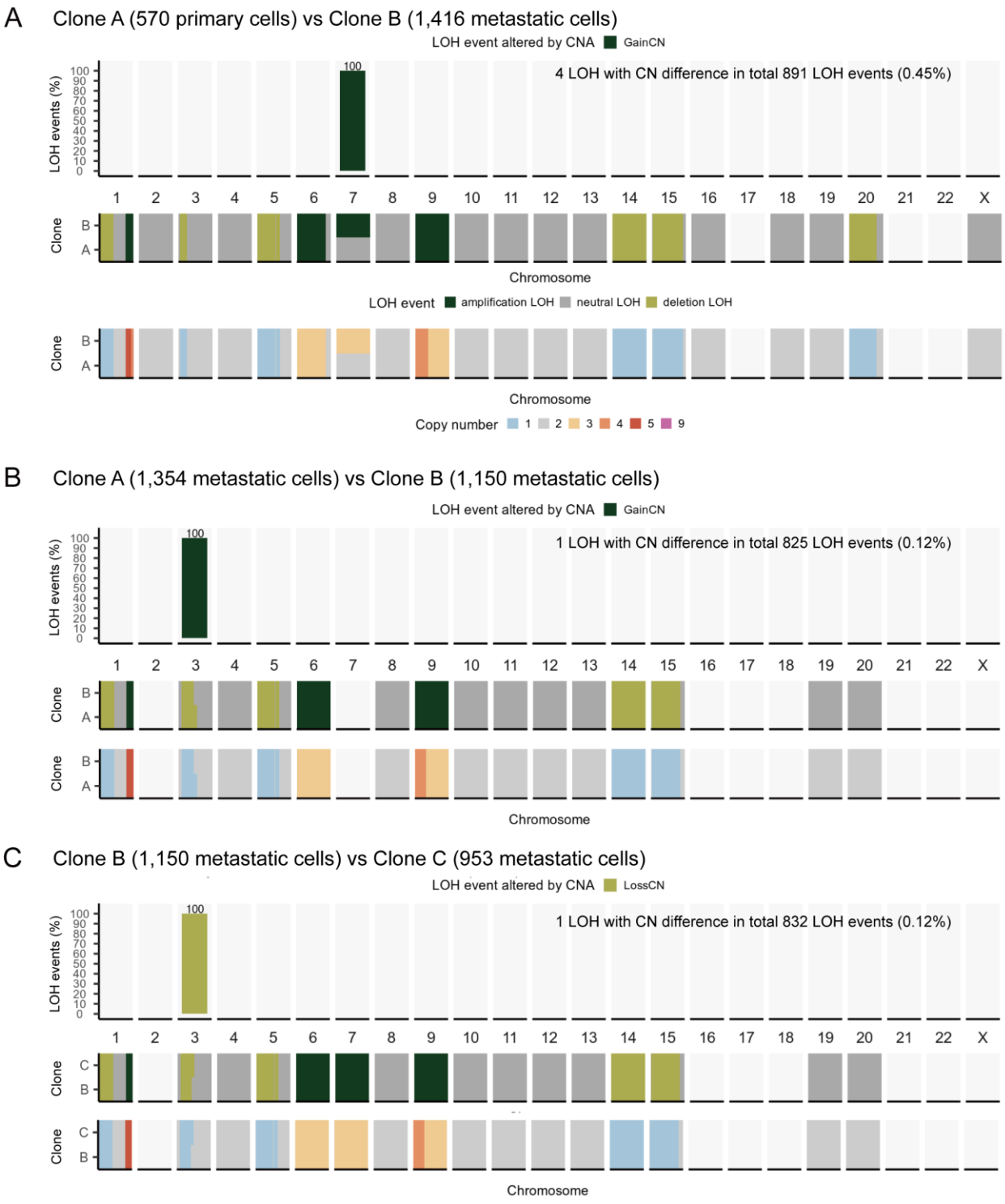


Figure 4-7 Loss of heterozygosity (LOH) events with copy number differences between clones in SA919.

Upper panel shows the proportion of LOH events (y-axis, percentage of LOH) accompanied by copy number difference across the genome (x-axis, chromosomes) between clones from total LOH events identified in both clones

Middle panel describes LOH events (amplification LOH, neutral LOH, deletion LOH as per key) in each chromosome region of clones in comparison. Lower panel shows the median copy number of cells analyzed for LOH (x-axis = chromosomes, colours = copy number as per key). A. LOH events with copy number loss or gain in primary clone B compared to primary clone A. B. LOH events with copy number loss or gain in metastatic clone B compared to metastatic clone A. C. LOH events with copy number loss or gain in metastatic clone C compared to metastatic clone B. CN = copy number, amplification LOH = LOH with CN > 2, neutral LOH = LOH with CN =2, deletion LOH = LOH with CN < 2.

4.2 Metastasis is driven by expansion of the subset of primary tumour clones in SA535

4.2.1 Genome and clonal structure of SA535 tumours

From 4 primary tumours and 9 metastases samples of passage 4 SA535, we sequenced 25,614 cells (12,405 primary tumour cells, 13,209 metastatic cells) and 4,375 cells (2,253 primary tumour cells, 2,122 metastatic cells) were used for subsequent analysis after filtering out low quality cells (methods).

We measured copy number states of SA535 primary and metastatic tumour cells which showed, unlike SA919, much higher degree of genomic instability which manifests as a large number of distinct clones (Figure 4-8). Figure 4-8 A shows the heatmap of total copy number from all primary and metastatic tumour cells from SA535 where a total of 20 clones were identified. Allele-specific copy number was measured with allelic imbalance score throughout the genome and allele-specific CNA regions were concordant with the total CNA regions (Figure 4-8 B). We observed several regions of LOH: whole chromosome imbalance (in chromosome 4, 9, 14, 15, 17), partial chromosome imbalance (in chromosome 1, 2, 3, 5, 6, 8, 10,11, 12, 13). These changes are shared in all cells indicating they were earlier founder events during tumour progression. Tumour cells gained copy number in these regions to different extent which contribute to distinct genome structure development and subsequently evolving into several clones. We noted that allele-specific copy number events are distinct for each clone supporting the phylogeny algorithm used to identify clones. The mean percentage of the genome LOH per cell in all cells of SA535 was 47.9% and mean percentage for each clone ranged from 46.8% to 51%. Copy-neutral LOH was identified in several regions, such as in chromosome 4, where allelic imbalance = 0 and mean percentages of copy-neutral LOH clone ranged from 27.9% to 37.1% (Figure 4-8 B, C).

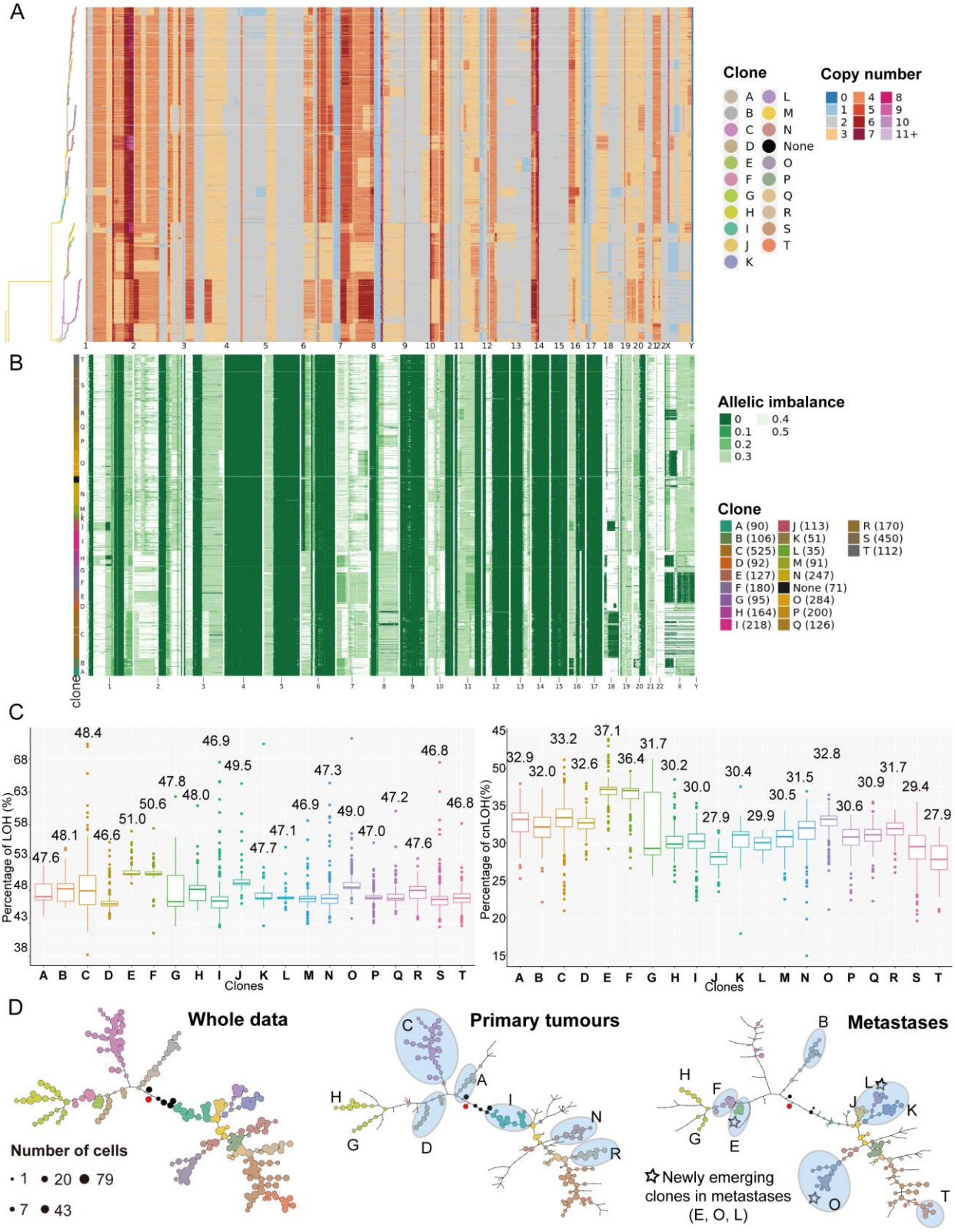

A
Clone
A B C D E F G H I J K L M N None O P Q R S T
Copy number
0 1 2 3 4 5 6 7 8 9 10 11+
B
Allelic imbalance
0 0.1 0.2 0.3 0.4 0.5
Clone
A (90) B (106) C (525) D (92) E (127) F (180) G (95) H (164) I (218) J (113) K (51) L (35) M (91) N (247) None (71) O (284) P (200) Q (126) R (170) S (450) T (112)
clone
C
Percentage of LOH (%)
Clones
Percentage of cnLOH(%)
Clones
D
Whole data
Number of cells
1 20 79
7 43
Primary tumours
Metastases
Newly emerging clones in metastases (E, O, L)

Figure 4-8 Clonal population structure and phylogeny of SA535 primary and metastatic tumour cells

A. Heatmap of single cell copy number states of primary and metastatic tumour cells from SA535X4, grey = diploid, copy gain/loss heat colours as per key, horizontal axis = chromosome/genome position. BREAKTREE inferred 20 clones as per key. B. Heatmap representation of allelic imbalance score. Allelic imbalance score is calculated as b/(a+b) where a and b are copy number of each allele. 0 = loss of one allele or LOH. 0.5 = equal number of copies in each allele. Allelic imbalance score as per key, chromosome/genome position horizontal axis. C. The mean percentages of loss of heterozygosity (LOH) per cell and copy-neutral LOH (cnLOH) per cell for each clone are showed in bar-plots with percentages described above each bar. D. On the phylogenetic tree inferred jointly from all primary and metastatic cells (whole data), clones for primary tumours and metastases are separately plotted. Each dot (nodes) represents tumour cells (size of dots showing the number of cells) assigned to each clone (colours) in a tree. Clones that are dominant in either primary tumours or metastases were highlighted in circled area (shades). Star = newly emerging clones in metastases.

4.2.2 Metastatic clones are derived from the subset of primary tumour clones

We then performed phylogenetic analysis using combined data of all primary and metastatic tumour cells. Single cell genomes from 3,937 cells (2,004 primary tumour cells, 1,933 metastatic cells) were analyzed for phylogeny after filtering out cells with highest copy number jumps (methods). A phylogenetic tree was inferred from all of the cells sequenced in primary and metastatic tumours and each cell was labeled for clone representation (Figure 4-8 D). Heatmap of copy number states of tumour cells was generated and grouped according to clones. The BREAKTREE algorithm identified 20 clones comprising a phylogenetic tree, which had 2 main branches. We visualized a tree where the size of dots represents the number of cells in clades with colours indicating clones (Figure 4-8 D). On the backbone of the phylogenetic tree inferred from all primary and metastatic cells (Figure 4-8 D, whole data), we separately plotted cells from primary tumours and metastases. We noted that all clones comprising primary tumours were also

found in metastases with different prevalence of each clone (Figure 4-8 D, primary tumours, metastases). Clones A, C, D, I, R, and N were more prevalent in primary tumour while clones B, F, K, J, and T showed higher prevalence in metastases. Clones H and G were present in both primary tumour and metastases however, in different locations/branches of the tree. Metastases also had clones that were not present in primary tumours (clone E, O, L).

Next, we measured clonal prevalence and number of cells of primary tumour and metastases of each replicate. The relative clonal prevalence showed distinct clonal population of each primary tumour and several clones were observed independently as prevalent metastatic clones, indicating that the ability to metastasize is distributed across clones (Figure 4-9). The pattern of metastases in SA535 was polyclonal/polyphyletic. Clonal composition of primary tumours were as follows (Table 4-1) : SA535X4-1 (A, 3.1%, n=10; C, 17.1%, n=55; D, 23% (n=74); F, 2.2%, n=7; G, 8.4%, n=27; H, 0.6%, n=2; I, 2.5%, n=8; K, 0.6%, n=3; M, 1.9%, n=6; N, 1.9%, n=6; P, 1.2%, n=4; R, 32.3%, n=104; S, 3.7%, n=12; T, 1.2%, n=4); SA535X4-2 (I, 14.6%, n=105; J, 0.6%, n=4; K, 0.7%, n=5; L, 0.4%, n=3; M, 2.6%, n=19; N, 20.9%, n=150; P, 1.8%, n=13; Q, 12.3%, n=88; R, 2.5%, n=18; S, 39.5%, n=284; T, 4%, n=29); SA535X4-3 (A, 25.4%, n=69; B, 1.8%, n=5; C, 7.4%, n=20; D, 0.4%, n=1; G, 0.4%, n=1; H, 40.1%, n=109; I, 6.6%, n=18; L, 0.3%, n=1; M, 5.1%, n=14; N, 9.9%, n=27; P, 0.4%, n=1; S, 2.2%, n=6); SA535X4-4 (A, 0.5%, n=3; C, 62.4%, n=396; D, 1.4%, n=9; G, 5.7%, n=36; H, 0.9%, n=6; I, 11.3%, n=72; J, 0.5%, n=3; K, 0.9%, n=6; M, 3.8%, n=24; N, 2.8%, n=18; P, 0.8%, n=5; Q, 0.2%, n=1; R, 5%, n=32; S, 3.5%, n=22; T, 0.3%, n=2). Primary tumours of SA535X4-1, SA535X4-3 and SA535X4-4 had clones from both of 2 main branches whereas SA535X4-2 had clones I-N and P-T which were from one main branch of the phylogenetic tree (Figure 4-9).

We observed that metastatic clones were derived from the subset of primary tumour clones and clonal diversity was reduced in metastases compared to originating primary tumour in all PDXs. In SA535X4-1, 14 clones comprised the primary tumour from which 5 clones developed metastases (Figure 4-10 A, B, C, Table 4-1). Clone E was a new clone emerged in metastasis. Each metastasis had 5 different clones (axillary metastasis: A, E, F, G, H, tumour recur: D, E, F, G, H, Figure 4-10 C). SA535X4-2 developed inguinal and axillary (ipsilateral and contralateral) metastasis. Twelve different metastatic clones were identified of which 11 clones were derived from the primary tumour. Metastasis to inguinal and contralateral axillary metastasis had newly emerging clone (clone O) that was not present in primary tumour. All 12 clones were identified in inguinal metastasis while axillary and contralateral axillary metastasis had 9 and 8 clones, respectively. In SA535X4-3, 12 clones formed the primary tumour and 3 clones were identified in metastasis. From 14 primary tumour clones of SA535X4-4, 7 clones contributed to metastasis and 1 new clone emerged in metastasis (clone L). Interestingly, metastatic clones in one replicate were from either one side of 2 main branches. In SA535X4-1 and SA535X4-3, metastatic clones were from 1 main branch of the tree that is comprised of clones A-H. On the other hand, metastatic clones in SA535X4-2 an SA535X4-4 were from the other main branch that is comprised of clones I-T (Figures 4-10 C, 4-11 B Phylogenetic tree).

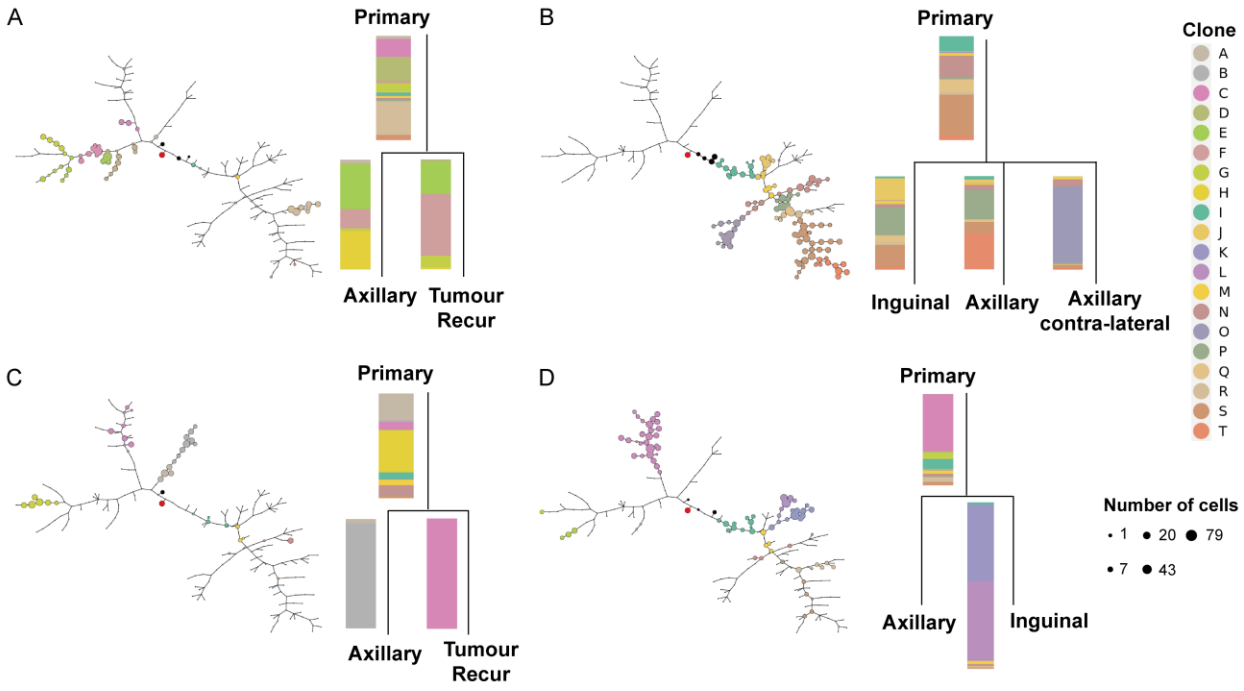


Figure 4-9 Phylogeny and clonal prevalence of SA535 primary tumours and metastases for each replicate

Phylogenetic tree and clonal prevalence (bar-plot) for primary tumours and metastases for A. SA535X4-1, B. SA535X4-2, C. SA535X4-3, D. SA535X4-4. Each dot (nodes) represents tumour cells (size of dots showing the number of cells) assigned to each clone (colours) in a tree.

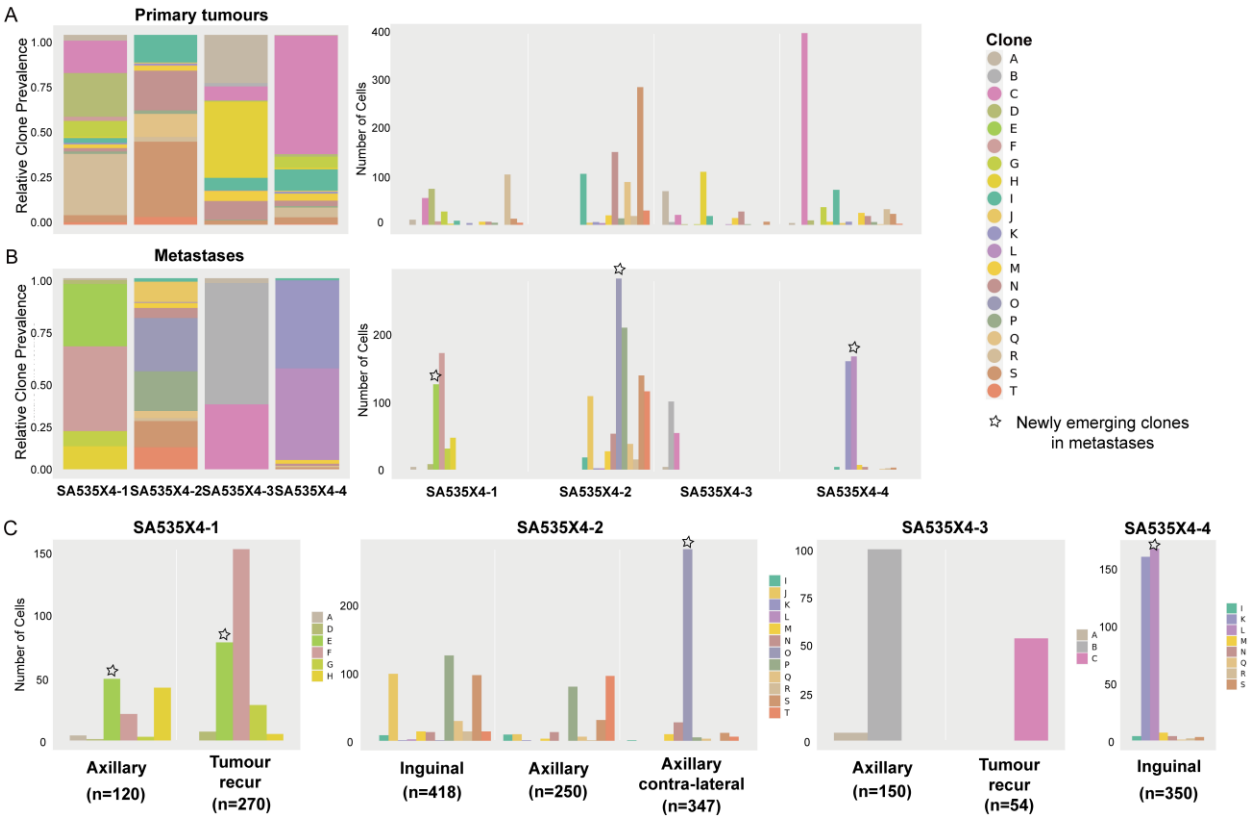


Figure 4-10 Clonal prevalence of SA535 primary tumours and metastases

The relative clone prevalence and number of cells for clones (y-axis) for primary tumours (A), metastases (B) and metastatic sites (C) of SA535 are shown. Colours represent clones as per keys. Newly emerging clones (star, clones E, O, L) were identified in metastases of SA535X4-1, SA535X4-2, and SA535X4-4.

Table 4-1. Clonal fraction of primary tumours and metastases for SA535.

	SA535X4-1			SA535X4-2				SA535X4-3			SA535X4-4	
Clone	**Prim**	**Ax**	**TR**	**Prim**	**Ax**	**Ax(cont)**	**Ing**	**Prim**	**Ax**	**TR**	**Prim**	**Ing**
A	0.031	0.033	0	0	0	0	0	0.254	0.038	0	0.005	0
B	0	0	0	0	0	0	0	0.018	0.962	0	0	0
C	0.171	0	0	0	0	0	0	0.074	0	1	0.624	0
D	0.23	0.008	0.026	0	0	0	0	0.004	0	0	0.014	0
E	0	0.408	0.289	0	0	0	0	0	0	0	0	0
F	0.022	0.175	0.563	0	0	0	0	0	0	0	0	0
G	0.084	0.025	0.104	0	0	0	0	0.004	0	0	0.057	0
H	0.006	0.35	0.019	0	0	0	0	0.401	0	0	0.009	0
I	0.025	0	0	0.146	0.036	0.003	0.019	0.066	0	0	0.113	0.011
J	0	0	0	0.006	0.04	0	0.237	0	0	0	0.005	0
K	0.009	0	0	0.007	0.004	0	0.002	0	0	0	0.009	0.46
L	0	0	0	0.004	0	0	0.005	0.003	0	0	0	0.48
M	0.019	0	0	0.026	0.012	0.029	0.033	0.051	0	0	0.038	0.02
N	0.019	0	0	0.209	0.052	0.078	0.031	0.099	0	0	0.028	0.011
O	0	0	0	0	0	0.816	0.002	0	0	0	0	0
P	0.012	0	0	0.018	0.32	0.014	0.301	0.004	0	0	0.008	0
Q	0	0	0	0.123	0.024	0.009	0.069	0	0	0	0.002	0.003
R	0.323	0	0	0.025	0.004	0	0.033	0	0	0	0.05	0.006
S	0.037	0	0	0.395	0.124	0.034	0.232	0.022	0	0	0.035	0.009
T	0.012	0	0	0.04	0.384	0.017	0.033	0	0	0	0.003	0

Prim = primary, Ax = axillary, TR = Tumour recurrence, cont = contralateral, Ing = Inguinal

4.2.3 Copy number genotype and LOH differences between clones may affect metastatic potential

A phylogenetic tree inferred by BREAKTREE shows the relationship between clones however, it does not provide branch lengths meaning that it is difficult to identify how close/far clones are related to each other. We measured genomic distance between clones by computing the Hamming distance, which is a metric for comparing two binary data strings, between clones. We measured the number of different chromosome positions between 2 clones and calculated the percentage of differences. For example, distance between clone A and B of SA919 is 0.06 meaning that there is a 6% of difference between the genome of clone A and B (Figure 4-11 A). Clone B is closer to clone C than clone A because genomic distance between clone B and C is 0.03 while it is 0.06 between clone B and A.

In SA535, clones A-H are more closely related to each other (Hamming distance range 0.05-0.37) and different from clones I-T (Hamming distance range 0.23-0.49) by genomic distance which supports the phylogenetic tree that has 2 main branches, each consisting of clones A-H and clones I-T (Figure 4-11 B). We compared the genomic distance between metastatic clones to identify if or how metastatic clones are related to each other. Metastatic clones of each replicate are as follows; SA535X4-1 (D, E, F, G, H), SA535X4-2 (I, J, M, N, O, P, Q, R, S, T), SA535X4-3 (A, B, C), SA535X4-4 (I, J, K, L, M, N). We noted that metastatic clones in each replicate are from one of 2 main branches (either A-H or I-T) and closely related to each other however, all clones in phylogenetic tree appeared at least in one of metastases indicating that metastatic potential is distributed across clones.

We also noted that metastases from each replicate showed distinct clonal population even in same metastatic sites. SA535X4-1 and SA535X4-3 both had axillary metastasis and tumour

recur, however clonal population for same metastatic sites in replicates were different from each other. Axillary metastasis in SA535X4-1 were consisted of clones A, E, F, G, and H while SA535X4-3 had clone A and B in axillary metastasis. Tumour recurrence in SA535X4-1 had clones D, E, F, G and H while in SA535X4-3, tumour recurrence had clone C. This suggests that metastasis to specific anatomical site is at least not governed by genomic clones that are defined by copy number states.

Next, we compared the genotype of clones that were prevalent in either primary tumour or metastases (Figure 4-12). From the phylogenetic tree (Figure 4-8 D), we noted that clone B (prevalent in metastasis) evolved from clone A (prevalent in primary tumour). We compared clones A and B to identify genes that might be associated with metastatic potential. We identified 55 genes that were associated with CNV difference between clone A (n=82) and B (n=101) and 10 genes (CDC42BPA, AKT3, RYR2, NTM, CADM1, PLCB1, PLCB4, SLC24A3, CDH4, DMD) were identified in hallmark gene sets (Figure 4-12 A). We also observed that clone G is present in primary tumours of SA535X4-1, 3, 4, however it only appeared in metastases of SA535X4-1 (Table 4-1). In a phylogenetic tree, clone G was comprised of 2 branches, one prevalent with primary tumour cells and the other with metastatic cells. We compared the genotype of primary (n=64) and metastatic (n=31) tumour cells of clone G and identified 30 hallmark genes (RASAL2, AKT3, RYR2, CDC42BPA, KCNH1, TRIO, ADCY2, CDKAL1, CD36, AUTS2, RELN, MAGI2, DOCK4, EXOC4, GRM8, ITPR2, CNTN1, TSPAN8, RPTOR, L3MBTL4, PTPRM, PLCB1, PLCB4, SLC24A3, CDH4, TIAM1, RUNX1, DMD, GPC3) from 182 CNV related genes (Figure 4-12 B). Similarly, clone H was present in primary tumours of SA535X4-1, 3, 4, however, only appeared in metastases of SA535X4-1 (Table 4-1). Clone H in phylogenetic tree was also comprised of 2 branches, one prevalent with primary tumour cells and

the other with metastatic cells. We identified 22 hallmark genes (PDE4B, RASAL2, VAV3, NTNG1, DPYD, CDC42BPA, KCNH1, AKT3, RYR2, HDAC9, CELF2, CAMK1D, ITPR2, CNTN1, L3MBTL4, PTRPM, PLCB1, PLCB4, SLC24A3, DMD, GPC3) that showed copy number difference between primary (n=117) and metastatic (n=47) cells of clone H (Figure 4-12 C). Finally, we compared primary tumour cells of clone S (n=324) and metastatic cells of clone T (n=116) because clone T evolved from clone S in a phylogenetic tree (Figure 4-8 D). Among 103 CNV related genes, 18 genes (CDC42BPA, PRKCE, GRID2, BANK1, UNC5C, ARHGAP10, INPP4B, PALLD, NR3CS, SORBS2, HDAC9, PTK2, TIAM1, RUNX1, GPC3, SLC4A4, BMPR1B, MARCH1) were identified as hallmark genes (Figure 4-12 D). CDC42BPA, which encodes serine/threonine protein kinase and is associated with mitotic spindle hallmark, appeared in 4 metastatic clones (B, G, H, T). AKT3, which encodes AKT kinase, appeared in 3 metastatic clones (B, G, H) as well as RYR2, which is associated with KRAS signaling. GPC3 also appeared in 3 metastatic clones (B, G, H) and is associated with hypoxia, inflammatory response and glycolysis. Interestingly, all genes that were associated with CNAs in SA919 (AKT3, HDAC9, AUTS2, MAGI2, CD36, RELN, DOCK4, EXOC4, GRM8, ADCY2, TRIO, CELF2, CAMK1D) were also identified in SA535.

We also analyzed LOH events accompanied by copy number difference in clones as previously done in SA919. In primary clone A (n=82) and metastatic clone B (n=101), 121 LOH events with copy number difference (5.85%) were identified from total of 2,069 LOH events observed in both clones (Figure 4-13 A). The regions that showed genotype difference were involved in regions of LOH with copy number differences (chromosomes 11, 20, Figure 4-12 A). In clone G from primary tumours (n=64) and metastases (n=31), 274 LOH events with copy number difference (13.45%) were identified from total of 2,037 LOH events in both clones.

Chromosome regions involved in both genotype and LOH copy number difference were chromosomes 6, 12, 17 and 20 (Figure 4-12 B, 4-13 B). In clone H of primary tumours (n=117) and metastases (n=47), 9 LOH events with copy number difference (0.44%) were identified in total of 2,034 LOH observed in both clones. Chromosome regions involved in both genotype and LOH difference was chromosome 12 (Figure 4-12 C, 4-13 C). From primary tumour clone S (n=324) and metastatic clone T (n=77), we identified 292 LOH events with copy number difference (14.06%) in total of 2,077 LOH observed in both clones. Chromosome 4 was involved in both genotype and LOH difference (Figure 4-12 D, 4-13 D). Unlike SA919, we observed some level of LOH events accompanied by copy number difference between clones. However, it should be noted that we measured only the abundance of LOH events in the genome and whether LOH affected the genotype or gene expression changes in clones needs to be further investigated.

Taken together, we observed differences of genotype and LOH induced by CNAs between clones that are predominant in either primary tumours or metastases. It is possible that these differences contributed to metastatic potential of clones leading to distinct clonal population of metastases. However, whether and how these changes affected metastatic potential of clones need to be studied further.

In summary, SA535 tumours showed (i) polyclonal/polyphyletic metastasis pattern and we discovered that (ii) metastatic clones were derived from the subset of the primary tumour clones, (iii) metastatic potential was distributed across clones, (iv) metastasis to specific site is not associated with genomic clones and (v) genotype and LOH in clones have potential impact on metastatic potential.

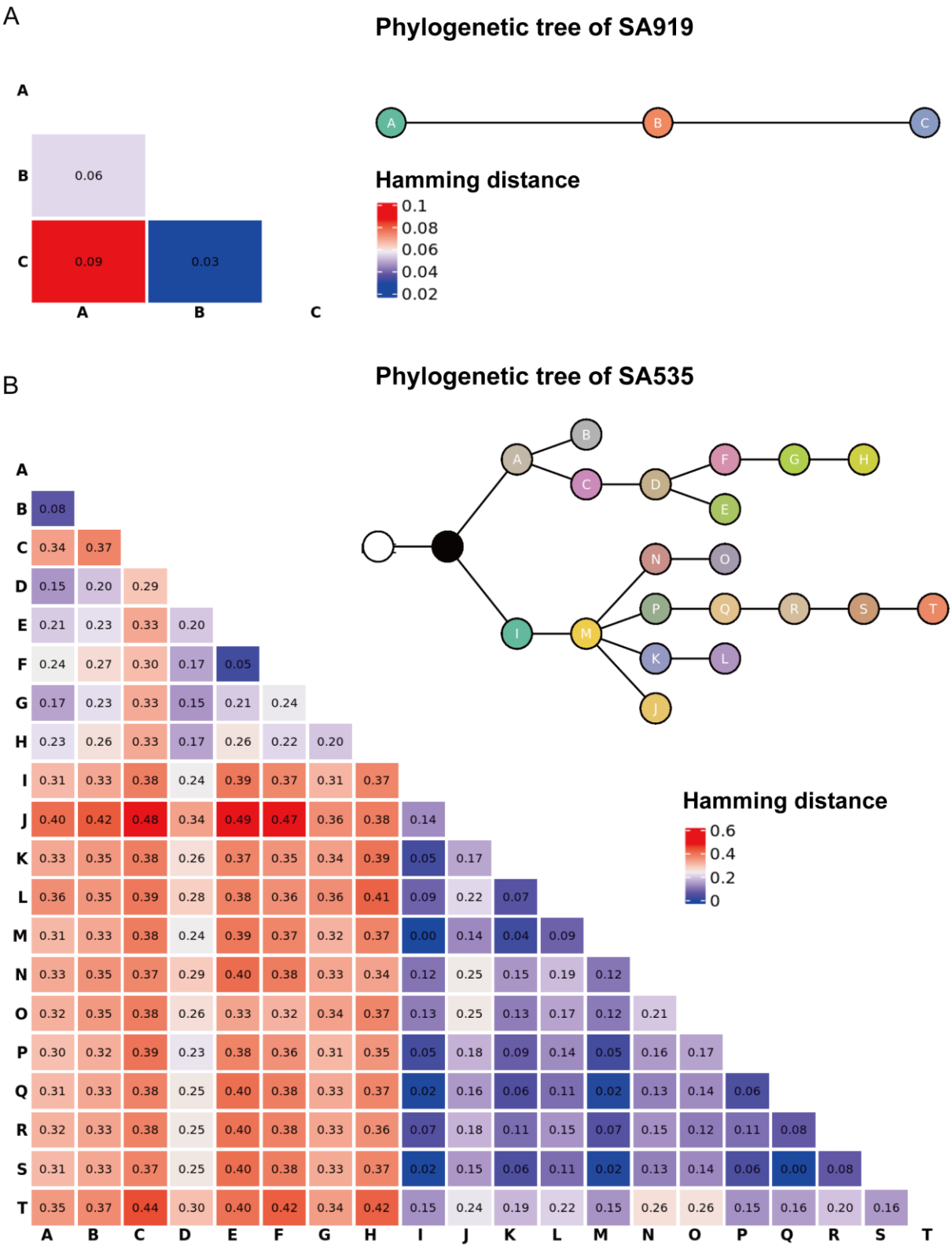


Figure 4-11 Genomic distance between clones in SA919 and SA535

Genomic distance between clones is computed using Hamming distance and presented as a fraction of difference for SA919 (A) and SA535 (B). A. Genomic distances between clones A, B and C of SA919 are described and summarized phylogenetic tree for SA919 is presented. B. Genomic distances between clones A-T of SA535 are described and summarized phylogenetic tree for SA535 is presented. Colours and numbers indicate genomic distance between clones.

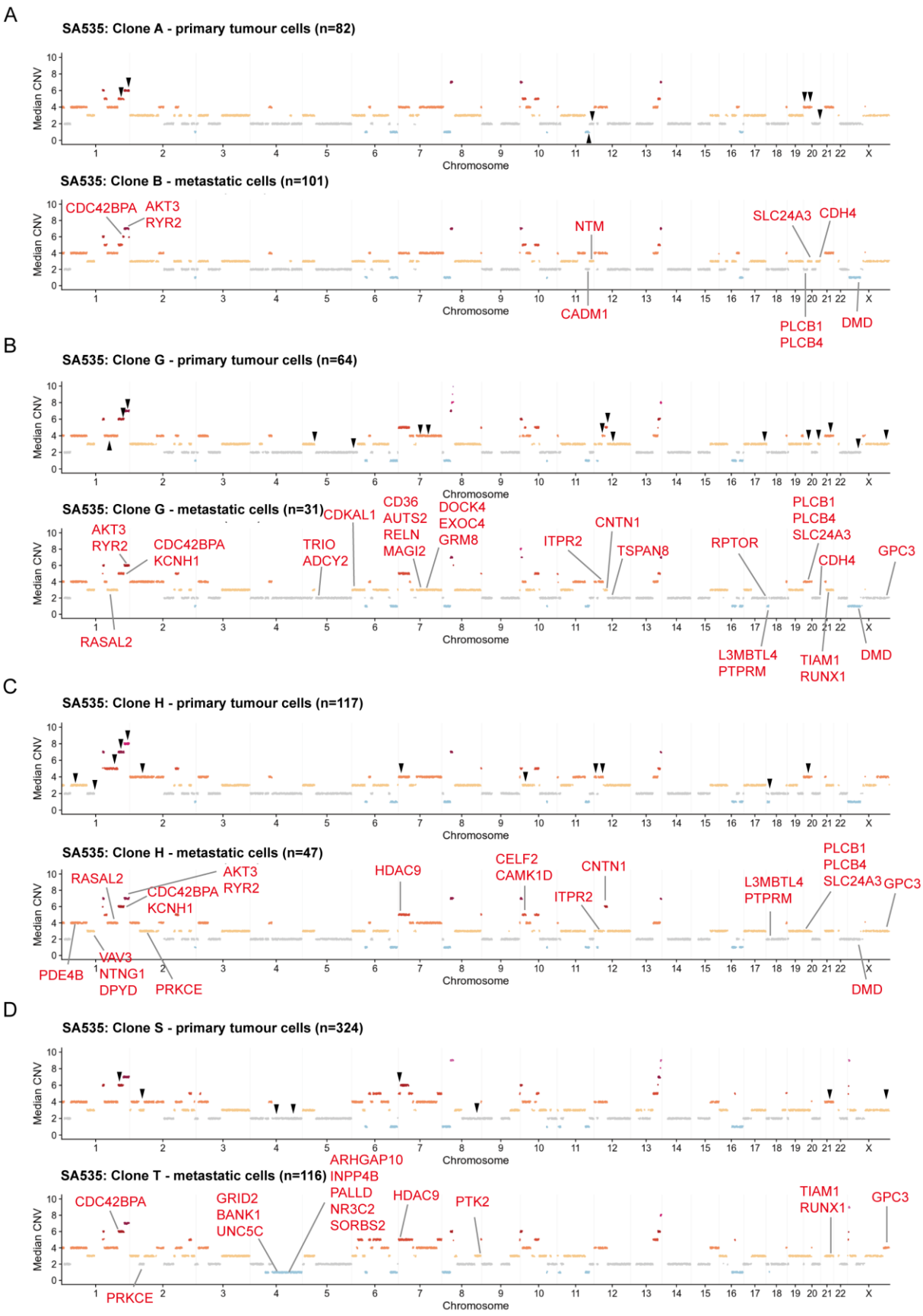

A
SA535: Clone A - primary tumour cells (n=82)
Median CNV
Chromosome
SA535: Clone B - metastatic cells (n=101)
CDC42BPA
AKT3
RYR2
NTM
SLC24A3
CDH4
CADM1
PLCB1
PLCB4
DMD
B
SA535: Clone G - primary tumour cells (n=64)
SA535: Clone G - metastatic cells (n=31)
AKT3
RYR2
CDC42BPA
KCNH1
CDKAL1
TRIO
ADCY2
CD36
AUTS2
RELN
MAGI2
DOCK4
EXOC4
GRM8
ITPR2
CNTN1
TSPAN8
RPTOR
PLCB1
PLCB4
SLC24A3
CDH4
GPC3
RASAL2
L3MBTL4
PTPRM
TIAM1
RUNX1
DMD
C
SA535: Clone H - primary tumour cells (n=117)
SA535: Clone H - metastatic cells (n=47)
RASAL2
AKT3
RYR2
CDC42BPA
KCNH1
HDAC9
CELF2
CAMK1D
CNTN1
ITPR2
L3MBTL4
PTPRM
PLCB1
PLCB4
SLC24A3
GPC3
PDE4B
VAV3
NTNG1
DPYD
PRKCE
DMD
D
SA535: Clone S - primary tumour cells (n=324)
SA535: Clone T - metastatic cells (n=116)
ARHGAP10
INPP4B
PALLD
NR3C2
SORBS2
CDC42BPA
GRID2
BANK1
UNC5C
HDAC9
PTK2
TIAM1
RUNX1
GPC3
PRKCE

Figure 4-12 Copy number genotype comparison between clones in SA535

Median copy number state (y-axis, median CNV) across the genome (x-axis) is compared between clones. A. Genotype of clone A from primary tumour and clone B from metastases. B. Genotype of clone G from primary tumours and metastases. C. Genotype of clone H from primary tumours and metastases. D. Genotype of clone S from primary tumour and clone T from metastases. Colours = copy number state, arrowhead = regions of copy number difference between clones, CNV = copy number variance, genes described in red.


A
Clone A (82 primary cells) vs Clone B (101 metastatic cells)
LOH event altered by CNA LossCN GainCN
121 LOH with CN difference in total 2,069 LOH events (5.85%)
LOH events (%)
45.5
26.4
24
3.3
0.8
Clone B A
Chromosome
LOH event amplification LOH neutral LOH deletion LOH
Copy number 1 2 3 4 5
B
Clone G (64 primary cells) vs Clone G (31 metastatic cells)
274 LOH with CN difference in total 2,037 LOH events (13.45%)
50.4
30.3
6.6
8
0.4
0.71.1
2.6
G_metastasis
G_primary
C
Clone H (117 primary cells) vs Clone H (47 metastatic cells)
9 LOH with CN difference in total 2,034 LOH events (0.44%)
77.8
22.2
H_metastasis
H_primary
D
Clone S (324 primary cells) vs Clone T (77 metastatic cells)
292 LOH with CN difference in total 2,077 LOH events (14.06%)
69.2
30.5
0.3
Clone T S
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X

Figure 4-13 Loss of heterozygosity (LOH) with copy number difference between clones in SA535.

Upper panel shows the proportion of LOH events (y-axis, percentage of LOH) accompanied by copy number difference across the genome (x-axis, chromosomes) between clones from total LOH events identified in both clones. Middle panel describes LOH events (amplification LOH, neutral LOH, deletion LOH as per key) in each chromosome region of clones in comparison. Lower panel shows the median copy number of cells analyzed for LOH (x-axis = chromosomes, colours = copy number as per key). A. LOH events with copy number loss or gain in metastatic clone B compared to primary clone A. B. LOH events with copy number loss or gain in metastatic clone G compared to primary clone G. C. LOH events with copy number loss or gain in metastatic clone H compared to primary clone H. D. LOH events with copy number loss or gain in metastatic clone T compared to primary clone S. CN = copy number, amplification LOH = LOH with CN > 2, neutral LOH = LOH with CN =2, deletion LOH = LOH with CN < 2.

Chapter 5: Discussion

5.1 Metastasis study using PDX models

5.1.1 Metastasis development in patients and PDX models

Breast cancer is a heterogeneous disease and metastasis pattern in individual patients are all different from each other, even if tumours exhibit same pathologic or molecular subtype. We have tested metastatic activity of breast cancer cells from 9 different patient-derived tumour tissues and 5 PDX lines developed metastasis. It has been shown in previous study that PDX recapitulates the metastasis pattern of a patient where transplanted materials originated from [89]. However, metastasis in PDXs did not match with metastatic sites from originating patients in this study. Among 5 PDX lines that developed metastasis, 3 lines showed metastasis in both PDX and patient (SA919, paraspinal in PDX vs. bone metastasis in patient; SA605, lung metastasis in PDX vs. bone and liver metastasis in patient; SA609, liver, axillary or lung metastasis in PDX vs. bone, pelvic metastasis in patient). Two PDX lines developed metastasis while corresponding patient did not (SA535, SA1142). Among 4 PDX lines that did not develop metastasis, only 2 lines showed metastasis in patients (SA604, SA501). It should be noted that transplant materials used in this study are not directly from patients. Instead, we used tumour cells from previously established PDXs that were further propagated in the lab as transplant materials. It has been shown that clonal populations change over serial passages [69], thus it is likely that tumour cells used for transplant may not represent the clonal population of patient's tumour from the outset. Moreover, PDX studies are limited by sampling bias, thus cannot fully represent the patients' tumour by its nature. Another important point is that we used immunocompromised mice that lack immune components. NSG and NRG strain mice are commonly used in cancer research since they lack immune cells, such as T cells, B cells,

macrophages, NK and NKT cells which enables engraftment of human tissues into mice [117]. PDX studies can provide insights to understand intrinsic properties of cancer cells, however it may not accurately simulate the physiology of cancer patients in part due to lack of immune system, leading to different outcome between PDX and patient.

To study metastasis *in vivo*, it is important to choose a PDX line that develop reproducible and sufficient number of metastases. We observed that metastasis development was reproducible over multiple passages of PDX models. SA919 repeatedly developed paraspinal metastases in 2 different passages and SA535 developed axillary, lung metastases as well as tumour recurrence at primary transplant site repeatedly in all 4 passages. These 2 PDX lines developed metastases that are large enough to provide sufficient number of cells and used for single cell analysis. SA1142 developed liver and lung metastases and is a potential PDX that can be further passaged for future studies. On the other hand, SA605 and SA609 developed metastases that are either not reproducible (SA609, liver metastasis in 1 of 4 replicates in passage 3, axillary and lung metastasis in 1 of 4 replicates in passage 4, no metastasis in passage 11) or sufficient (SA605, only micrometastasis in lung in 2 of 3 replicates). It should be also noted that time to develop primary tumour as well as metastasis often takes several weeks. For example, SA919X7 developed primary tumour in 12-17 weeks and metastasis in 10-15 weeks after tumour removal resulting in 27 weeks in total (Table 3-2). Total amount of time from transplant to metastasis in SA535X4 was 26-46 weeks although it was shorter in later passage (16-19 weeks in SA535X7) (Table 3-3). Thus, it is important to acknowledge these factors when planning metastasis study using PDX models.

5.1.2 Further optimization of tumour cell isolation method from organs is required

For single cell analysis, especially for DLP+, a sufficient number of viable cells is required [77]. However, it was difficult to retrieve viable cells from metastasis in organs (SA535X4-2, lung metastasis = 2 cells; SA1142-1, lung metastasis = 2 cells, liver metastasis = 5 cells). It should be noted that all lung metastasis developed in this study were micrometastasis without grossly visible lesions, which may have attributed to low yield of tumour cells. We used MACS Miltenyi Biotec Mouse Depletion Kit which removed more than 90% of cells from dissociated lung tissue indicating the low tumour burden of lung metastasis. Also, foreign microenvironment in distant site suppresses and inhibits the survival of metastatic cells [118, 119] which could be the cause of low viability of these micrometastatic cells. Since the formation of metastases has several rate-limiting steps, metastatic cells in organs are in a vulnerable state until they overcome all the steps and proliferate [118]. On the other hand, primary tumour cells as well as cells from large metastatic mass formed outside the organ structure showed relatively high viability. It requires at least 50 cells from DLP+ to define a clone and much more cells to infer clonal population structure of a tumour. We need to explore more PDX lines to identify ones that develop large enough organ metastasis (macrometastasis) which will enable us to analyze clonal relationship between primary tumour and metastatic cells. Moreover, optimizing the method for isolating and dissociating metastatic tumour cells from organs, such as lung and liver, for subsequent single cell studies should be further investigated in the future. Thus, it is preferable to use PDX lines that develop solid metastatic lesions with sufficient size at the moment, which is the reason we used SA919 and SA535.

5.2 Relationship between CNA determined clonal evolution and metastatic potential

We observed that clonal population of primary tumours of SA919 evolved over the passages and metastatic potential increased with increasing passage number. Clonal population of passage 3 primary tumour was comprised entirely of clone A and clones B and C appeared as the primary tumour was propagated into passage 4 and 7. It is likely that the appearance of clones B and C increased the metastatic potential of primary tumour. Unlike SA919, SA535 tumours showed similar metastatic potential over 4 different passages. All passage 4 to 7 tumours developed metastases with highly reproducible metastatic sites. We analyzed passage 4 primary and metastatic tumour cells where we identified polyclonal and polyphyletic metastases. It is possible that other passage primary tumours have similar clonal population structure and developed reproducible metastases however, it should be investigated further.

From re-transplant metastatic cell experiment, we observed different metastatic activity between 2 PDX lines. In SA919X7, metastatic cells from SA919X7-3 were re-transplanted into MFP and developed aggressive behavior resulting in multiple metastasis. Tumour cells selected during metastasis process may have achieved traits that provided cells the better fitness to survive and proliferate in mammary fat pad and distant microenvironments. We observed that clonal population of metastasis in SA919X7-3 was dominated by clone C which was minor clone of primary tumour (Figure 4-5 F) and therefore, re-transplanted tumour cells had different clonal population compared to its original primary tumour. It is possible that we selected tumour cells with higher metastatic potential (clone C) and re-transplanted into mammary fat pad resulting in aggressive metastasis development.

On the other hand, metastatic cells from SA535X4 showed more modest behaviour compared to its original primary tumour cells. Metastatic tumour cells from both axillary

metastasis and tumour recurrence, collected from SA535X4-1, exhibited faster primary tumour growth. Development of tumour recurrence was observed in most of re-transplanted mice (6/7). All 4 mice transplanted with cells from tumour recurrence developed second tumour recurrence. However, axillary metastasis was observed in only one mouse. It is likely that MFP, which is the same tumour microenvironment of recurrent tumour, may have favored re-development of tumour recurrence whereas it may have been a foreign environment for cells from axillary metastasis. Interestingly, mice transplanted with cells from axillary metastasis also developed tumour recurrence in 3 of 4 mice and the development of axillary metastasis was observed only in 1 mouse. We observed that clonal population of tumour recurrence and axillary metastasis were similar to each other with different clonal proportions (Figure 4-10 A). Similar metastatic behaviour of tumour cells from axillary metastasis and tumour recurrence can be explained by their similar clonal population. However, it is still not clear why tumour cells from axillary metastasis did not develop axillary metastasis. We observed that metastasis to specific anatomical site is not clone-dependent. Perhaps, metastasis to axillary area is simply more difficult than to develop tumour recurrence. It is also possible that metastatic cells underwent selection process during the tumour development in mammary fat pad upon re-transplant leading to reduced metastatic activity. Further analysis of samples from re-transplant experiment is required to enhance our understanding of the behaviour of metastatic tumour cells.

5.3 Contribution of CNAs to metastatic potential

Clonal dynamics during serial passage in PDXs were previously shown [69]. Eirew et al. showed that clonal population of engrafted tumour changes over serial passage, suggesting the evolution of clones to survive and adapt to forced microenvironment changes. In multiple passages of SA919, development of metastasis was different for each passage. Passage 3 showed

no metastasis development while passage 4 developed metastasis in 50% (2/4). Passage 3 and 4 tumours that did not develop metastasis showed identical clonal population, which had one dominant clone comprising the whole population (clone A). Passage 4 tumour that developed metastases had a minor clone with additional amplification of chromosome 7 (clone B). Moreover, all primary tumours of passage 7 also had a clone with additional amplification of chromosome 7 (clone B), which was the dominant clone of these primary tumours. It is possible that primary tumour cells have achieved amplification of chromosome 7 during the evolution upon serial passage. Interestingly, the frequency of metastasis was increased in later passage indicating that additional amplification in chromosome 7 may have provided higher metastatic potential to clones in passage 7 since it resulted in metastasis in all replicates. We also observed that re-transplanted metastatic tumour cells, mostly consisted of clones with additional CNAs (clone C), showed more aggressive metastatic behavior suggesting the impact of CNAs to metastatic potential. Genotype difference between clones is a potential factor that may contributed to the difference of metastatic potential in clones. We identified several genes that were involved in the regions of chromosome copy number changes. However, further investigation with integrated analysis using both copy number and transcriptomic data is required to identify the relationship between copy number changes and gene expressions as well as their impact on metastatic potential.

It should be also noted that there was a phenotype change (non-basal to basal) from earlier passage 3 and 4 to later passage 7. EGFR, which is located at chromosome 7, showed low to no protein expression from IHC staining in tumour cells from passage 3 and 4 while passage 7 tumour cells showed high EGFR protein expression. It is likely that increase in EGFR expression in later passage resulted from amplification of chromosome 7. INPP4B also changed from

positive in earlier passage to negative in later passage, however it was not associated with copy number alteration. INPP4B is located in chromosome 4 and there was no alteration of copy number in chromosome 4 in any of tumour cells from primary tumour or metastasis. Protein expression of INPP4B is likely to have resulted from the mechanism other than CNA.

The contribution of CNA to metastatic potential was also suggested in metastasis of passage 4. Amplification of chromosome 7 was identified in a minor population of tumour cells of primary tumour in passage 4 that developed metastasis. Clonal population of passage 4 metastasis consisted of 2 clones, pre-existing clone (clone A) of primary tumour being the dominant clone with newly emerging clone (clone B) as a minor population. It is not clear how the appearance of new clone in passage 4 tumour affected the metastatic potential of pre-existing clone and it needs further investigation with more replicates.

In passage 7 of SA919, primary tumours had another newly emerging clone with additional amplification in chromosomes 5 and 10 (clone C). Clone B was the major clone in primary tumours of passage 7 with clones A and C as minor clones however, clone C became the major clone in metastases indicating the higher metastatic potential of clone C. Clones A and B also appeared in metastases, suggesting that these clones also had metastatic potential but with lower magnitude compared to clone C. From genotype analysis, we identified 13 genes that were associated with CNAs and pathways that these genes mediate are as follows: AKT3 = Apical junction, UV response; HDAC9 = Allograft rejection, KRAS signaling; AUTS2 = Interferon gamma response; MAGI2 = UV response; CD36 = IL6_JAK_STAT3_Signaling; RELN = KRAS signaling; DOCK4 = Mitotic spindle; EXOC4 = Apical junction; GRM8 = Spermatogenesis; ADCY2 = PI3K_AKT_mTOR signaling; TRIO = Mitotic spindle; CELF2 = UV response; CAMK1D = KRAS signaling.

In SA535, we showed that metastasis contributing clones are different between each replicate although clonal population of primary tumours were similar to each other. Clones that were found in metastases of one replicate did not contribute to metastases in other replicates (clones G, H were present in primary tumours of SA535X4-1, 3, 4, but only appeared in metastases of SA535X4-1). We also observed that all clones of primary tumours appeared at least in one of metastases in replicates. These results suggest that the metastatic potential was distributed across clones, however, their contribution to metastases is determined by other factors. We identified the genotype difference between primary and metastatic counterparts of clones G and H. Interestingly, several genes (RASAL2, AKT3, RYR2, CDC42BPA, KCNH1, L3MBTL4, PTPRM, PLCB1, PLCB4, SLC24A3, DMD, GPC3) were identified in metastatic cells of both clones G and H indicating their potential impact on metastatic potential. Similarly, we observed that clone B, which was evolved from clone A in a phylogenetic tree, was only present in primary tumour of SA535X4-3 as a minor clone and contributed to metastases as a major clone. Clone B was not present in primary tumours of SA535X4-1 and SA535X4-4, both of which involved clone A, as well as in their metastases. We identified genotype difference between clones A and B and several genes were also present in metastatic clones G and H (CDC42BPA, AKT3, RYR2, SLC24A3, PLCB1, PLCB4, DMD). Pathways that these genes medicate are: CDC42BPA = Mitotic spindle, UV response; AKT3 = Apical junction, UV response; RYR2 = KRAS signaling; SLC24A3 = Estrogen response; PLCB1 = PI3K_AKT_mTOR signaling; PLCB4 = UV response; DMD = G2M checkpoint, myogenesis. Interestingly, all genes identified in SA919 genotype analysis were also present in SA535. The common pathways identified in both PDXs were Mitotic spindle, Apical junction, UV response,

KRAS signaling and PI3K_AKT_mTOR signaling. It is possible that these pathways are involved in mechanism which control metastatic potential of clones.

These results indicate that CNAs have impact on metastatic potential of clones. It is not clear why some clones evolved to acquire additional CNAs and appeared in metastases in one replicate but not in others in SA535. One possible explanation is that interplay between clones of primary tumour resulted in distinct clonal population structure of metastases leading to variance in replicates, however, more investigation is required. Furthermore, metastases to same anatomical site in replicates were comprised of different clones meaning that the CNA-determined clones did not decide the metastatic site. Previous studies have identified several metastasis related genes that potentially affects the specific site of metastasis [36-38]. Thus, it is more likely that factors other than CNAs, such as specific gene expressions, determine the organotropism in metastasis.

In addition to genotype difference between clones, we also observed the difference of LOH events with CNAs between clones in SA535, suggesting that LOH may play a role in metastatic potential of clones. In SA919, we only observed minimal difference in LOH events between primary and metastatic clones and no difference between metastatic clones which can be attributed to relatively lower genomic instability of the tumour.

CNAs appeared to affect the metastatic potential of tumour cells based on our observation of two PDX metastasis model systems. It is difficult to analyze the effect of CNAs on metastatic potential in complex polyclonal tumours, such as SA535, where interplay between several clones occur that can also contribute to metastatic potential. However, it provides chances to study how clones interact and what determines the ability to metastasize in polyclonal tumours. On the other hand, tumours that have simple clonal population structure, such as

SA919, are ideal to study the impact of specific CNAs on metastatic potential. Breast cancer metastasis PDX models that we established in this thesis will contribute to future studies to better understand the mechanism of metastasis in breast cancer.

5.4 Limitation of the study

We were limited by the number of viable tumour cells from metastasis to perform single cell analysis. To isolate single cells using DLP+, a sufficient number of viable cells are required from tumour samples. SA919 and SA535 developed metastasis that formed solid mass, such as paraspinal or axillary metastasis, which is the reason that we selected these PDXs for single cell analysis. Lung metastasis was identified in 3 PDX lines including SA535 and SA1142, although they were all micrometastasis. SA1142 also developed metastatic mass in liver. However, isolating viable tumour cells from lung or liver for single cell analysis was not successful. For example, only 5 tumour cells, accompanied by many mouse cells, were successfully isolated and analyzed from lung metastasis from SA535X4-2 (Appendix Figure B-2). Similarly, we were able to isolate only 2 tumour cells from SA1142 lung metastasis and 5 cells from liver metastasis while 899 cells were retrieved from primary tumour sample (Appendix Figure B-2). Thus, further experiments to optimize the dissociation method as well as enriching metastatic tumour cells from mouse organ are required.

Also, we only studied 2 PDX lines (SA919 and SA535) for single cell analysis, each of which showed distinct pattern of metastases and clonal population structure. We need to increase the number of PDX lines to identify various patterns of metastases and to study the underlying mechanism of breast cancer metastasis using single cell analysis.

Another limitation of this study is the application of BREAKTREE algorithm to infer phylogeny of metastatic samples. We used copy number states of single cells to infer evolutionary history of tumour cells during metastatic process. However, we could not infer the migration history of tumour cells and identify the timing of metastasis. That is because history of CNAs in tumour cells does not directly model the history of cellular migration between anatomical sites. Application of algorithm that can model and infer clonal lineages as well as migration histories of cancer metastasis [117] or using lineage tracing methods [118] may advance our understanding of clonal evolution during metastasis.

5.5 Future directions

To understand how CNAs can affect metastatic behaviour of tumour cells, further replications of PDX metastasis experiments and generations of more sequencing libraries alongside with RNA analysis will be required. First, we will conduct single cell RNA sequencing from primary and metastatic tumour samples to identify the difference of gene expressions between primary tumour and metastases as well as in multiple passages of primary tumours. We will also identify pathways associated with differential gene expressions. Transcriptomic analysis will provide additional information regarding metastasis development. Second, we will repeat transplant experiment of multiple passages in SA919 to increase the number of replicates to support that previous results did not derive from stochastic event or random chance and to achieve more power. We will also conduct an experiment to see whether additional amplification in chromosome 7 provides better metastatic fitness by mixing clonal population of primary tumour cells from passage 3 (without chromosome amplification 7) and passage 7 (with chromosome amplification 7). Moreover, we will conduct a mixing experiment using passage 4

and 7. Analysis of samples from re-transplant experiments will also provide insights to understand the mechanism of tumour evolution during metastasis cascade. By combining the results of copy number status and transcript changes, we will be in better position to understand the landscape of clonal evolution during metastasis and underlying mechanisms.

5.6 Conclusions

We developed breast cancer metastasis mouse model using patient-derived tissues. Each PDX line showed different growth rate of primary tumour and each metastatic PDX exhibited unique pattern of metastasis. Expression of protein markers of primary tumour was maintained in metastatic tissues. We re-transplanted metastatic tumour cells into mammary fat pad and observed that the behavior of re-transplanted tumour cells varied depending on the originating material either being more aggressive and metastatic or remained similar.

Single-cell whole-genome sequencing of primary and metastatic tumour cells revealed clonal population structure and the relationship between these cells. Analysis of primary and metastatic cells of multiple passages in SA919 identified that metastatic potential of clones appeared to be affected by CNAs. We observed that competitiveness of individual tumour cells was affected by CNAs leading to survival and growth advantage in either primary tumour or metastasis. In SA535, we observed that metastatic clones derived from the subset of primary tumour clones and newly emerging clones were identified in some metastases. Metastasis pattern of SA535 was polyclonal and polyphyletic. We also noted that the ability to metastasize was distributed across clones and metastasis to specific anatomical site is not clone-dependent. We observed the genotype/LOH difference between clones in primary tumours and metastases suggesting the potential impact of CNAs on metastatic potential.

In conclusion, from the two transplant systems studied in detail, metastatic potential is distributed across many CNA-defined clones and CNAs have potential impact on metastatic potential of tumour cells.

Bibliography

[1] Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* 100, 57−70 (2000).

[2] Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* 144, 646−674 (2011).

[3] Bray, F.et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries.CA: *A cancer journal for clinicians* 68, 394−424 (2018).

[4] Canadian Cancer Statistics Advisory Committee. *Canadian Cancer Statistics 2019* (September, 2019). URL https://www.cancer.ca/Canadian-Cancer-Statistics-2019-EN. Toronto, ON:Canadian Cancer Society. Accessed: November, 2019.

[5] Navin, N.et al. Inferring tumor progression from genomic heterogeneity. *Genome research* 20, 68−80 (2010).

[6] McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 168, 613−628 (2017).

[7] Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochimica et Biophysica Acta (BBA)* - Reviews on Cancer 1805, 105−117 (2010).

[8] Shipitsin, M.et al. Molecular definition of breast tumor heterogeneity. *Cancer cell* 11, 259−273 (2007).

[9] Allred, D., Harvey, J. M., Berardo, M. & Clark, G. M. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Modern pathology*: an official journal of the United States and Canadian Academy of Pathology, Inc11,155−168 (1998).

[10] Burstein, H. J.et al. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: Asco clinical practice guideline focused update. *Journal of Clinical Oncology* 37, 423–438 (2019).

[11] Giordano, S. H.et al. Systemic therapy for patients with advanced human epidermal growth factor receptor 2−positive breast cancer: Asco clinical practice guideline update. *Journal of Clinical Oncology* 36, 2736−2740 (2018).

[12] Sørlie, T.et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences* 98, 10869−10874 (2001).

[13] Prat, A.et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research* 12, R68 (2010).

[14] Dawson, S.-J., Rueda, O. M., Aparicio, S. & Caldas, C. A new genome-driven integrated classification of breast cancer and its implications. *The EMBO journal* 32, 617−628 (2013).

[15] Shah, S. P.et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486, 395 (2012).

[16] Beca, F. & Polyak, K. Intratumor heterogeneity in breast cancer. In *Novel Biomarkers in the Continuum of Breast Cancer*, 169−189 (Springer, 2016).

[17] Kwan, M. L.et al. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast cancer research* 11, R31 (2009).

[18] Kohler, B. A.et al. Annual report to the nation on the status of cancer, 1975-2011, featuring incidence of breast cancer subtypes by race/ethnicity, poverty, and state. *Journal of the National Cancer Institute* 107(2015).

[19] Li, X.et al. Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast cancer research and treatment* 161, 279–287 (2017).

[20] Lehmann, B. D.et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation* 121, 2750–2767 (2011).

[21] Burstein, M. D.et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clinical cancer research* 21,1688–1698 (2015).

[22] Pareja, F.et al. Triple-negative breast cancer: the importance of molecular and histologic subtyping, and recognition of low-grade variants. *NPJ breast cancer* 2,16036 (2016).

[23] Nielsen, T. O.et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clinical cancer research* 10,5367–5374 (2004)

[24] Rakha, E. A., El-Sayed, M. E., Reis-Filho, J. & Ellis, I. O. Patho-biological aspects of basal-like breast cancer. *Breast cancer research and treatment* 113, 411–422 (2009).

[25] Badowska-Kozakiewicz, A. M. & Budzik, M. P. Immunohistochemical characteristics of basal-like breast cancer. *Contemporary Oncology* 20, 436(2016).

[26] Fedele, C. G.et al. Inositol polyphosphate 4-phosphatase ii regulates pi3k/akt signaling and is lost in human basal-like breast cancers. *Proceedings of the National Academy of Sciences* 107, 22231–22236 (2010).

[27] Weigelt, B., Peterse, J. L. & Vanʾt Veer, L. J. Breast cancer metastasis: markers and models. *Nature reviews cancer* 5, 591 (2005).

[28] Gupta, G. P. & Massagué, J. Cancer metastasis: building a framework. *Cell* 127,679−695 (2006).

[29] Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nature reviews cancer* 9, 274 (2009).

[30] Vanharanta, S. & Massagué, J. Origins of metastatic traits. Cancer cell 24,410−421 (2013).

[31] Paget, S. The distribution of secondary growths in cancer of the breast. *Lancet* 571−573 (1889).

[32] Peinado, H.et al. Pre-metastatic niches: organ-specific homes for metastases. *Nature reviews cancer* 17, 302−317 (2017).

[33] Huang, Y.et al. Pulmonary vascular destabilization in the premetastatic phase facilitates lung metastasis. *Cancer research* 69, 7529−7537 (2009).

[34] Hiratsuka, S.et al. Primary tumours modulate innate immune signaling to create pre-metastatic vascular hyperpermeability foci. *Nature communications* 4, 1−10 (2013).

[35] Giles, A. J.et al. Activation of hematopoietic stem/progenitor cells promotes immunosuppression within the pre−metastatic niche. *Cancer research* 76, 1335−1347 (2016).

[36] Padua, D.et al. Tgfβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* 133, 66−77 (2008).

[37] Bos, P. D.et al. Genes that mediate breast cancer metastasis to the brain. *Nature* 459, 1005 (2009).

[38] Minn, A. J.et al. Genes that mediate breast cancer metastasis to lung. *Nature* 436, 518 (2005).

[39] Li, F., Sun, L. & Zhang, S. Acquirement of dna copy number variations in non-small cell lung cancer metastasis to the brain. *Oncology reports* 34, 1701−1707 (2015).

[40] Faubert, B., Solmonson, A. & DeBerardinis, R. J. Metabolic reprogramming and cancer progression. *Science* 368 (2020).

[41] Lee, Y.-T. N. Breast carcinoma: pattern of metastasis at autopsy. *Journal of surgical oncology* 23, 175−180 (1983).

[42] Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* 501, 328−337 (2013).

[43] Nowell, P. C. The clonal evolution of tumor cell populations. *Science* 194, 23−28 (1976).

[44] Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *New England journal of medicine* 368, 842−851 (2013).

[45] Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501, 338−345 (2013).

[46] Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A. Identification of neutral tumor evolution across cancer types. *Nature genetics* 48, 238 (2016).

[47] Ling, S.et al. Extremely high genetic diversity in a single tumor points top revalence of non-darwinian cell evolution. *Proceedings of the National Academy of Sciences* 112, E6496−E6505 (2015).

[48] Williams, M. J., Sottoriva, A. & Graham, T. A. Measuring clonal evolution in cancer with genomics. *Annual review of genomics and human genetics* 20 (2019).

[49] Ahmed, Z. & Gravel, S. Intratumor heterogeneity and circulating tumor cell clusters. *Molecular biology and evolution* 35, 2135−2144 (2017).

[50] Yang, F.et al. Intratumor heterogeneity predicts metastasis of triple-negative breast cancer. *Carcinogenesis* 38, 900–909 (2017).

[51] Shi, Y.-J., Tsang, J. Y., Ni, Y.-B. & Gary, M. T. Intratumoral heterogeneity in breast cancer: a comparison of primary and metastatic breast cancers. *The oncologist* 22, 487–490 (2017).

[52] Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* 352, 169–175 (2016).

[53] Klein, C. A. Parallel progression of primary tumours and metastases. *Nature reviews cancer* 9, 302 (2009)

[54] Juric, D.et al. Convergent loss of pten leads to clinical resistance to a pi (3) kαinhibitor. *Nature* 518, 240 (2015).

[55] Schmidt-Kittler, O.et al. From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proceedings of the National Academy of Sciences* 100, 7737–7742 (2003).

[56] Zhang, Y.et al. Copy number alterations that predict metastatic capability of human breast cancer. *Cancer research* 69, 3795–3801 (2009).

[57] Pollack, J. R.et al. Microarray analysis reveals a major direct role of dna copy number alteration in the transcriptional program of human breast tumors. *Proceedings of the National Academy of Sciences* 99, 12963–12968 (2002).

[58] Schwartz, R. & Schäffer, A. A. The evolution of tumour phylogenetics: principles and practice. *Nature reviews genetics* 18, 213–229 (2017).

[59] Baum, D. Reading a phylogenetic tree: the meaning of monophyletic groups. *Nature Education* 1, 190 (2008).

[60] Baum, D. A. & Offner, S. Phylogenies & tree-thinking. *The American Biology Teacher* 70, 222−229 (2008).

[61] Echeverria, G. V.et al. High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. *Nature communications* 9, 5079 (2018).

[62] Cheung, K. J.et al. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proceedings of the National Academy of Sciences* 113, E854−E863 (2016).

[63] McPherson, A.et al. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nature genetics* 48, 758 (2016).

[64] Roth, A.et al. Pyclone: statistical inference of clonal population structure in cancer. *Nature methods* 11, 396−398 (2014).

[65] Gerlinger, M.et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature genetics* 46, 225 (2014).

[66] Caravagna, G.et al.Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature methods* 15, 707−714 (2018).

[67] Gerstung, M.et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nature communications* 3, 811 (2012).

[68] Wang, Y.et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512, 155–160 (2014).

[69] Eirew, P.et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 518, 422–426 (2015).

[70] Gao, R.et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nature genetics* 48, 1119 (2016).

[71] Leung, M. L.et al. Single-cell dna sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome research* 27, 1287–1299 (2017).

[72] Navin, N.et al. Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90 (2011).

[73] Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626 (2012).

[74] Hou, Y.et al. Single-cell exome sequencing and monoclonal evolution of ajak2-negative myeloproliferative neoplasm. *Cell* 148, 873–885 (2012).

[75] Ni, X.et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proceedings of the National Academy of Sciences* 110, 21083–21088 (2013).

[76] Davis, A. & Navin, N. E. Computing tumor trees from single cells. *Genome biology* 17, 113 (2016).

[77] Laks, E.et al. Clonal decomposition and dna replication states defined by scaled single-cell genome sequencing. *Cell* 179, 1207–1221 (2019).

[78] Beroukhim, R.et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905 (2010).

[79] Zack, T. I.et al. Pan-cancer patterns of somatic copy number alteration. *Nature genetics* 45, 1134–1140 (2013).

[80] Alaei-Mahabadi, B., Bhadury, J., Karlsson, J. W., Nilsson, J. A. & Larsson, E. Global analysis of somatic structural genomic alterations and their impact on gene expression in diverse human cancers. Proceedings of the National Academy of Sciences 113, 13768−13773 (2016).

[81] Bergamaschi, A.et al. Distinct patterns of dna copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes, chromosomes and cancer* 45, 1033–1040 (2006).

[82] Behring, M.et al. Integrated landscape of copy number variation and rna expression associated with nodal metastasis in invasive ductal breast carcinoma. *Oncotarget* 9, 36836 (2018).

[83] Tyran, M.et al. A comparison of dna mutation and copy number profiles of primary breast cancers and paired brain metastases for identifying clinically relevant genetic alterations in brain metastases. *Cancers* 11, 665 (2019).

[84] Gillet, J.-P.et al. Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. *Proceedings of the National Academy of Sciences* 108, 18708−18713 (2011).

[85] Lin, Q.et al. Asph-notch axis guided exosomal delivery of prometastatic secretome renders breast cancer multi-organ metastasis. *Molecular cancer* 18,1−17 (2019).

[86] Eyre, R.et al. Microenvironmental il1βpromotes breast cancer metastatic colonisation in the bone via activation of wnt signalling. *Nature communications* 10, 1−15 (2019).

[87] Vishnoi, M.et al. The identification of a tnbc liver metastasis gene signature by sequential ctc-xenograft modeling. *Molecular oncology* 13, 1913–1926 (2019).

[88] Davis, R. T.et al. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell rna sequencing. *Nature cell biology* 22, 310–320 (2020).

[89] DeRose, Y. S.et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nature medicine* 17, 1514 (2011).

[90] Dobrolecki, L. E.et al. Patient-derived xenograft (pdx) models in basic and translational breast cancer research. *Cancer and metastasis reviews* 35, 547–573 (2016).

[91] DeRose, Y. S.et al. Patient-derived models of human breast cancer: protocols for in vitro and in vivo applications in tumor biology and translational medicine. *Current protocols in pharmacology* 60, 14–23 (2013).

[92] Cassidy, J. W., Caldas, C. & Bruna, A. Maintaining tumor heterogeneity inpatient-derived tumor xenografts. *Cancer research* 75, 2963–2968 (2015).

[93] Coussy, F.et al. A large collection of integrated genomically characterized patient-derived xenografts highlighting the heterogeneity of triple-negative breast cancer. *International journal of cancer* 145, 1902–1912 (2019).

[94] Zahn, H.et al. Scalable whole-genome single-cell library preparation without preamplification. *Nature methods* 14, 167 (2017).

[95] Ha, G.et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome research* 22, 1995–2007 (2012).

[96] Wingett, S. W. & Andrews, S. Fastq screen: A tool for multi-genome mapping and quality control. *F1000 Research* 7 (2018).

[97] Singer, J., Kuipers, J., Jahn, K. & Beerenwinkel, N. Single-cell mutation identification via phylogenetic inference. *Nature communications* 9, 1–8 (2018).

[98] Zafar, H., Tzen, A., Navin, N., Chen, K. & Nakhleh, L. Sifit: inferring tumor trees from single-cell sequencing data under finite-sites models. *Genome biology* 18,1–20 (2017).

[99] Ross, E. M. & Markowetz, F. Onconem: inferring tumor evolution from single-cell sequencing data. *Genome biology* 17, 1–14 (2016).

[100] Dorri, F.et al. Efficient bayesian inference of phylogenetic trees from large scale, low-depth genome-wide single-cell data. *bioRxiv* (2020).

[101] Salehi, S.et al. Single cell fitness landscapes induced by genetic and pharmacologic perturbations in cancer. *bioRxiv* (2020).

[102] Razmara, A. M.et al. Tumor shedding and metastatic progression after tumor excision in patient-derived orthotopic xenograft models of triple-negative breast cancer. *Clinical & experimental metastasis* 37, 413–424 (2020).

[103] Brill, B., Boecher, N., Groner, B. & Shemanko, C. A sparing procedure to clear the mouse mammary fat pad of epithelial components for transplantation analysis. *Laboratory animals* 42, 104–110 (2008).

[104] Gilles, C.et al. Vimentin contributes to human mammary epithelial cell migration. *Journal of cell science* 112, 4615–4625 (1999).

[105] Liu, C.-Y., Lin, H.-H., Tang, M.-J. & Wang, Y.-K. Vimentin contributes to epithelial-mesenchymal transition cancer cell mechanics by mediating cytoskeletal organization and focal adhesion maturation. *Oncotarget* 6, 15966 (2015).

[106] Kidd, M. E., Shumaker, D. K. & Ridge, K. M. The role of vimentin intermediate filaments in the progression of lung cancer. *American journal of respiratory cell and molecular biology* 50, 1−6 (2014).

[107] Li, H.et al. Nestin is expressed in the basal/myoepithelial layer of the mammary gland and is a selective marker of basal epithelial breast tumors. *Cancer research* 67, 501–510 (2007).

[108] Krüger, K.et al. Expression of nestin associates with brca1 mutations, a basal-like phenotype and aggressive breast cancer. *Scientific reports* 7, 1−12(2017).

[109] Asleh, K., Won, J. R., Gao, D., Voduc, K. D. & Nielsen, T. O. Nestin expressionin breast cancer: association with prognosis and subtype on 3641 cases with long-term follow-up. *Breast cancer research and treatment* 168, 107−115 (2018).

[110] Piras, F.et al. Nestin expression associates with poor prognosis and triple negative phenotype in locally advanced (t4) breast cancer. *European journal of histochemistry*: EJH55(2011).

[111] De Lara, S.et al. The prognostic relevance of foxa1 and nestin expression in breast cancer metastases: a retrospective study of 164 cases during a 10-year period (2004-2014). *BMC Cancer* 19, 187−187 (2019).

[112] Rakha, E. A.et al. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clinical cancer research* 15, 2302−2310 (2009).

[113] Cheang, M. C.et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clinical cancer research* 14, 1368−1376 (2008).

[114] Won, J. R.et al. A survey of immunohistochemical biomarkers for basal-like breast cancer against a gene expression profile gold standard. *Modern pathology* 26, 1438−1450 (2013).

[115] Malek, J. A.et al. Copy number variation analysis of matched ovarian primary tumors and peritoneal metastasis. *PloS one* 6, e28561 (2011).

[116] Liberzon, A.et al. The molecular signatures database hallmark gene set collection. *Cell systems* 1, 417−425 (2015).

[117] Hidalgo, M.et al. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer discovery* 4, 998−1013 (2014).

[118] Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. *Nature reviews cancer* 9, 239−252 (2009).

[119] Suzuki, M., Mose, E. S., Montel, V. & Tarin, D. Dormant cancer cells retrieved from metastasis-free organs regain tumorigenic and metastatic potency. *The American journal of pathology* 169, 673−681 (2006).

[120] El-Kebir, M., Satas, G. & Raphael, B. J. Inferring parsimonious migration histories for metastatic cancers. *Nature genetics* 50, 718−726 (2018).

[121] Quinn, J. J.et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* (2021).

Appendices

Appendix A

Supplementary tables.

A.1 Table. Pathologic features of breast cancer patients

Patient ID	Age at diagnosis	TNM	Stage	Grade	LVI	Subtype
SA919	67	NA	NA	NA	NA	TNBC
SA535	44	T2N1M0	IIb	3	(-)	TNBC
SA1142	74	T2N0M0	IIa	2	(-)	TNBC
SA605	64	T2N1M0	IIb	3	(-)	TNBC
SA609	43	T2N0M0	IIa	3	(-)	TNBC
SA604	40	T2N1M0	IIb	3	(-)	TNBC
SA501	35	T2N1M0	IIb	3	(-)	TNBC
SA1139	84	T2N0M0	IIa	3	NA	TNBC
SA1146	48	T2N1M0	IIb	3	(+)	TNBC

A.2 Table. Antibodies and scoring information for immunochemistry staining

Target	Clone	Vendor	Dilution	Staining pattern	Scoring method
ER	EP1	Dako	RTU*	Nuclear	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 1%
PR	1294	Dako	RTU	Nuclear	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 1%
E-cadherin	NCH-38	Dako	RTU	Membranous	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 1%
EGFR	EP22	ESBE (Cell Marque)	1:200	Membranous	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 10%
HER2	4B5	Roche	1:8, RTU	Membranous	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when 3+, ≥ 10%
Ki-67	MIB1	Dako	RTU	Nuclear	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained
CK5/6	D5/16 B4	Dako	RTU	Membranous and cytoplasmic	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 10%
Vimentin	V9	Dako	RTU	Cytoplasmic	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 1%
Nestin	10C2	Santa Cruz Biotechnology	1:50	Cytoplasmic	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 1%
INPP4B	EPR3108Y	Epitomics	1:50	Cytoplasmic	Intensity (1=weak, 2=intermediate, 3=strong), % cells positively stained, positive when ≥ 5%

A.3 Table. IHC scoring for primary tumours and metastases from SA919 (Intensity/percentage)

PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA919X3-1	Primary	0/0	0/0	0/0	3/40	1/1	1/25	3/95	0/0	0/0	1/100
SA919X3-3	Primary	0/0	0/0	0/0	3/45	1/1	1/15	3/100	0/0	0/0	1/100
SA919X4-1	Primary	0/0	0/0	0/0	3/25	1/1	1/1	3/100	0/0	0/0	1/100
SA919X4-2	Primary	0/0	0/0	0/0	3/30	1/1	1/1	3/100	0/0	0/0	1/100
SA919X4-3	Primary	0/0	0/0	0/0	3/35	1/1	1/10	3/100	0/0	0/0	1/100
SA919X4-3	Abdominal	0/0	0/0	0/0	3/40	1/1	2/10	3/100	0/0	0/0	1/100
SA919X4-3	Ventral spinal	0/0	0/0	0/0	3/35	1/1	1/10	3/100	0/0	0/0	1/100
SA919X4-4	Primary	0/0	0/0	0/0	3/40	1/1	1/10	3/100	0/0	0/0	1/100
SA919X4-4	Axillary	0/0	0/0	0/0	3/35	1/5	1/35	3/95	0/0	0/0	1/100
SA919X4-4	Ventral-spinal	0/0	0/0	0/0	3/25	1/1	0/0	2/80	0/0	0/0	1/1
SA919X7-1	Primary	0/0	0/0	0/0	3/35	1/1	2/60	3/100	0/0	0/0	0/0
SA919X7-1	Supra-spinal	0/0	0/0	0/0	3/35	1/1	2/60	3/100	0/0	2/1	0/0
SA919X7-1	Ventral-spinal	0/0	0/0	0/0	3/35	1/1	2/60	3/100	0/0	1/1	0/0
SA919X7-2	Primary	0/0	0/0	0/0	3/30	1/1	2/60	3/100	0/0	0/0	0/0
SA919X7-2	Supra-spinal	0/0	0/0	0/0	3/40	1/1	2/80	3/100	0/0	1/1	0/0
SA919X7-3	Primary	0/0	0/0	0/0	2/1	3/65	0/0	3/100	0/0	0/0	0/0

PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA919X7-3	Ventral-spinal	0/0	0/0	0/0	3/35	2/1	1/60	3/100	0/0	3/1	0/0
SA919X7-4	Primary	0/0	0/0	0/0	3/50	1/1	2/60	3/100	0/0	0/0	0/0
SA919X7-4	Ventral-spinal	0/0	0/0	0/0	3/40	1/1	2/70	3/100	0/0	1/1	0/0

A.4 IHC scoring for primary tumours and metastases from SA535 and SA1142 (Intensity/percentage)

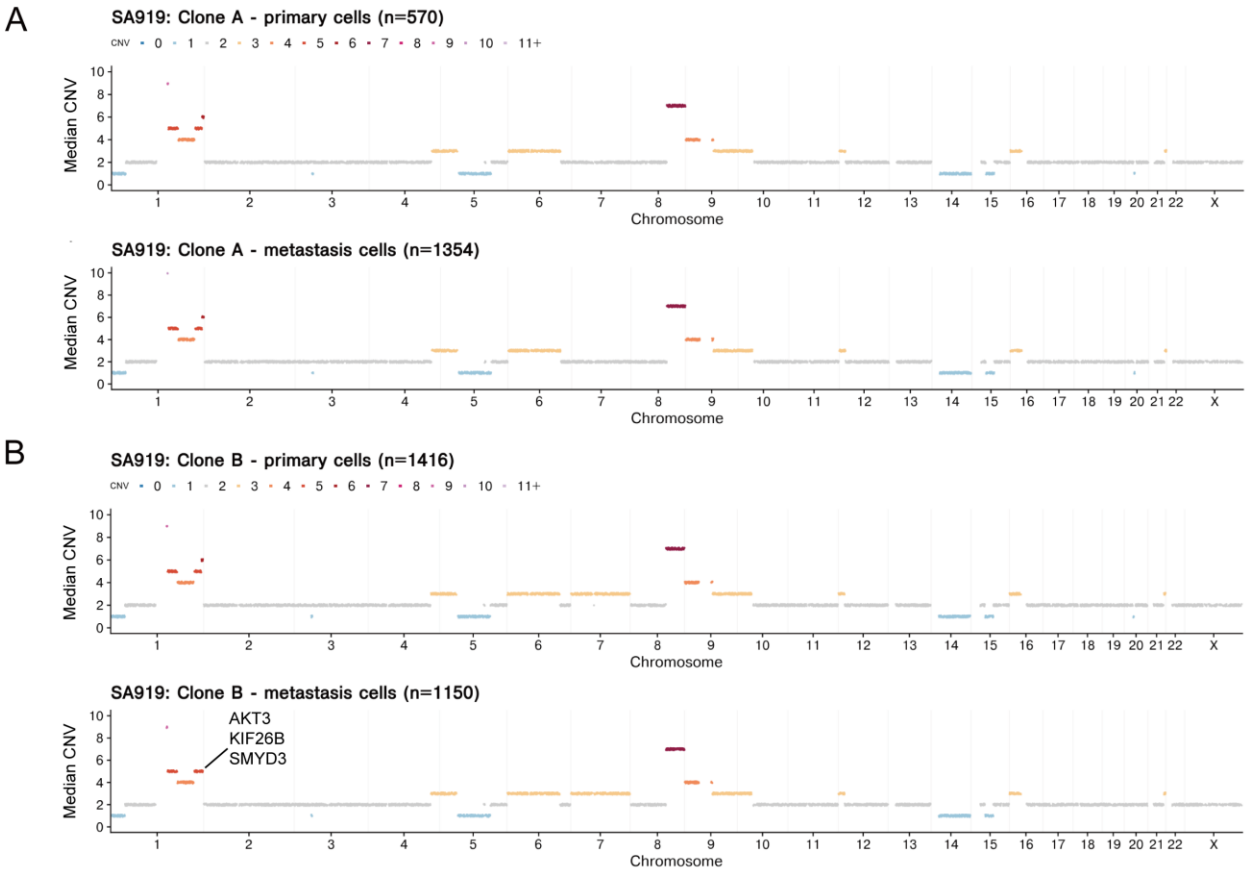
PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA535X4-1	Recur	0/0	0/0	0/0	3/15	0/0	2/40	3/100	3/40	1/1	1/5
SA535X4-1	Axillary	0/0	0/0	0/0	3/25	2/1	2/40	3/100	3/90	2/15	1/10
SA535X4-1	Lung	0/0	0/0	0/0	3/25	0/0	1/45	1/60	3/55	2/35	1/10
SA535X4-2	Primary	0/0	0/0	0/0	3/25	2/1	3/60	3/100	3/90	2/15	0/0
SA535X4-2	Axillary (R)	0/0	0/0	0/0	3/30	2/1	3/50	3/100	3/80	2/30	0/0
SA535X4-2	Axillary (L)	0/0	0/0	0/0	3/40	2/1	3/50	3/100	3/95	2/80	0/0
SA535X4-2	Inguinal	0/0	0/0	0/0	3/25	2/30	3/60	3/100	3/75	2/20	0/0
SA535X4-2	Lung	0/0	0/0	0/0	3/30	1/1	1/40	2/80	3/80	3/95	0/0
SA535X4-3	Primary	0/0	0/0	0/0	3/25	1/1	3/30	3/100	3/75	2/15	1/1
SA535X4-3	Recur	1/20	0/0	0/0	3/30	2/5	3/55	3/100	3/20	1/1	1/1
SA535X4-3	Axillary	0/0	0/0	0/0	3/25	2/1	2/75	3/100	3/95	2/20	1/1
SA535X4-3	Lung	0/0	0/0	0/0	NA	1/1	1/60	1/30	3/80	2/10	0/0
SA535X4-4	Primary	0/0	0/0	0/0	3/30	2/5	3/45	3/95	3/95	2/20	1/5
SA535X4-4	Axillary	0/0	0/0	0/0	3/35	0/0	2/40	3/85	3/95	2/20	1/5
SA535X4-4	Inguinal	0/0	0/0	0/0	3/30	1/1	2/45	2/60	3/95	2/40	1/1
SA535X4-4	Lung	0/0	NA	NA	NA	0/0	1/40	1/80	3/90	NA	1/100

PDX ID	Source	ER	PR	HER2	Ki67	CK5/6	EGFR	E-cadherin	Vimentin	Nestin	INPP4B
SA535X6-1	Primary	0/0	0/0	0/0	3/30	1/1	2/75	3/100	3/80	2/5	2/100
SA535X6-2	Primary	0/0	0/0	0/0	3/35	1/5	2/75	3/100	3/85	2/15	2/100
SA535X6-3	Primary	0/0	0/0	0/0	3/35	1/5	2/85	3/100	3/95	2/5	2/100
SA535X6-4	Primary	0/0	0/0	0/0	3/35	1/5	2/90	3/100	3/90	2/5	2/100
SA535X7-4	Primary	0/0	0/0	0/0	3/30	1/1	2/85	3/100	3/55	2/20	1/100
SA535X7-4	Recur	0/0	0/0	0/0	3/50	1/1	2/90	3/100	3/90	2/25	1/100
SA535X7-4	Axillary	0/0	0/0	0/0	3/50	1/1	2/95	3/100	3/80	2/20	2/100
SA1142X2-1	Primary	0/0	0/0	0/0	3/15	3/95	2/15	3/100	3/10	2/1	0/0
SA1142X2-1	Lung	0/0	0/0	0/0	3/20	3/90	1/40	2/80	3/70	2/5	1/5
SA1142X2-1	Liver	0/0	0/0	0/0	3/20	3/75	1/10	2/50	1/1	0/0	0/0
SA1142X2-3	Primary	0/0	0/0	0/0	3/15	3/65	2/20	3/90	3/5	2/1	0/0
SA1142X2-3	Lung	0/0	NA	0/0	NA	3/90	1/50	1/70	2/20	NA	1/40
SA1142X2-3	Liver	0/0	0/0	0/0	3/40	2/60	1/5	3/80	2/5	0/0	0/0

Appendix B

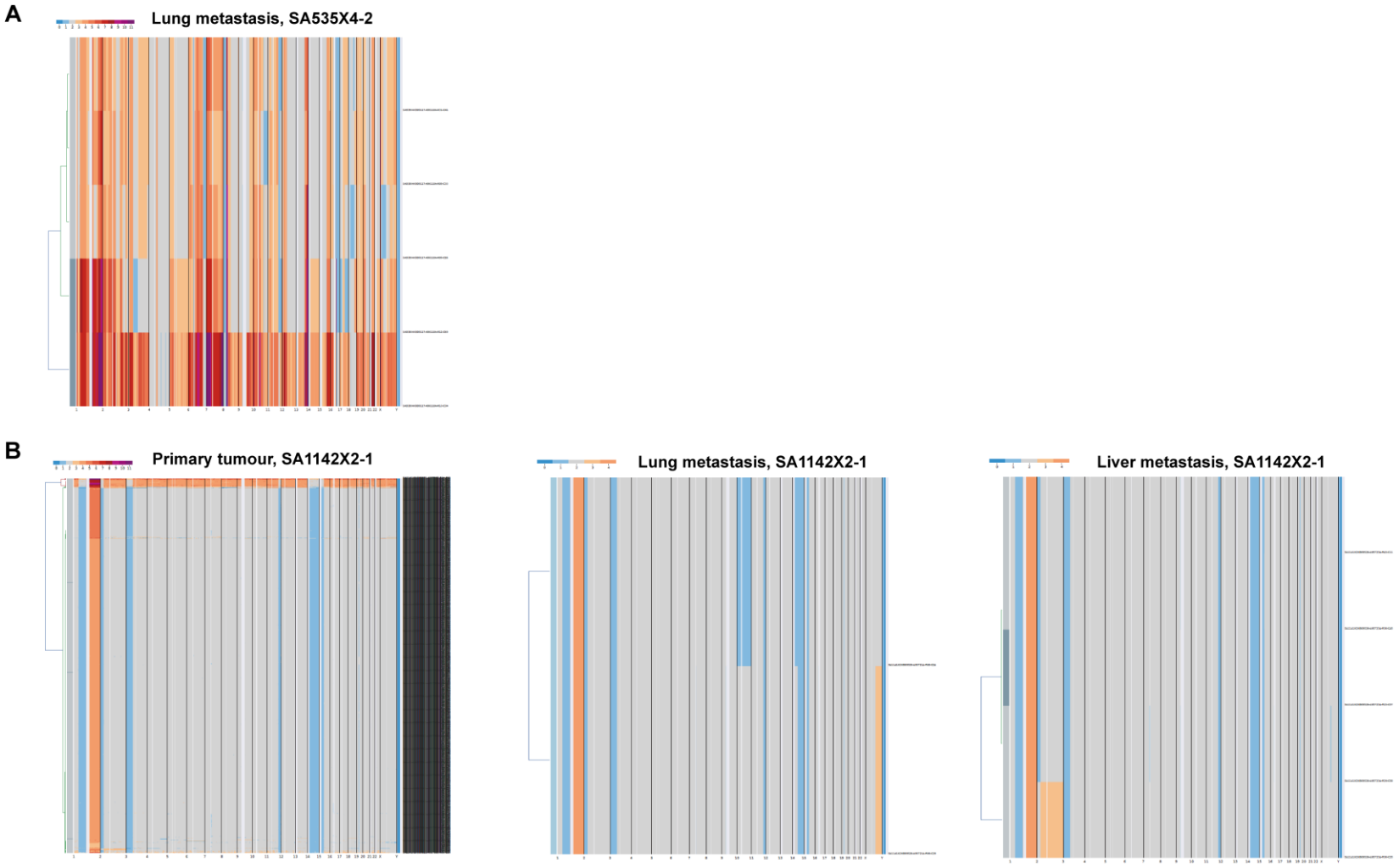
Supplementary figures.

B.1 Copy number genotype comparison between primary and metastatic clones in SA919.



Median copy number state (y-axis, median CNV) across the genome (x-axis) is compared between clones. A. Copy number genotype comparison between primary and metastatic clone A in SA919. B. Copy number genotype comparison between primary and metastatic clone B in SA919.

B.2 Copy number states of tumour cells from lung (SA535, SA1142) and liver (SA1142).



Heatmap of copy number status (colours) for single cells (y-axis) across the genome (x-axis). A. Lung metastasis (n=5) from SA535X4-2, B. Primary tumour (n=677), lung metastasis (n=2) and liver metastasis (n=5) from SA1142X2-1.