Identifying Novel Genes and Pathways Correlated with Group 3 and Group 4 Medulloblastoma Metastasis to the Spine

by

Stuart Aaron Lithwick (Matan-Lithwick)

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Laboratory Medicine & Pathobiology
University of Toronto

Identifying Novel Genes and Pathways Correlated with Group 3 and Group 4 Medulloblastoma Metastasis to the Spine

Stuart Aaron Lithwick (Matan-Lithwick)

Doctor of Philosophy

Laboratory Medicine & Pathobiology
University of Toronto

2021

Abstract

Medulloblastoma (MB) is a cancer of the cerebellum and the most common childhood brain malignancy. For children with high-risk MB, mortality is nearly always the result of the primary tumour having metastasized to the leptomeninges. This is due to the fact that current therapy for metastatic MB is less effective than that for primary disease. This is a product of the fact that most MB research has focused on primary tumours rather than metastases, due to the low availability of metastatic MB tissues for research, and, until recently, the absence of mouse models of metastatic MB. With greater understanding of primary MB has come a more rounded foundation upon which to build a better therapy. Thus, it is exciting to note that mouse models of MB dissemination now exist, and with these models has come first glimpses of mechanisms that may be driving MB metastasis. In an effort to build upon this new body of knowledge our lab engaged in a process of repeated selection for increased metastasis propensity among spine metastases from mouse patient-derived xenograft (PDX) models of human Group 3 and Group 4 MB. By phenotypically selecting for increased metastasis propensity, we would necessarily also select for increased activity of the genes and pathways necessary and sufficient for increased metastasis propensity, some of which might represent valuable new therapeutic targets. In this manner, we have identified several long non-coding RNAs (lncRNAs) and Cancer Testis

Antigen (CTA) genes increasingly transcribed in correlation with metastasis. Further, we have found that their increased transcription is followed closely by the activation of several well-known metastasis pathways. So lncRNAs and CTA genes may represent novel new components of the complex systems regulating MB metastasis.

Acknowledgments

I am supported by the Vanier-Canada Doctoral Scholarship Program, affiliated with the Canadian Institutes of Health Research

I would like to thank my supervisors, Drs. Michael Taylor and Gary Bader, and the members of my supervisory committee Drs. Anna Goldenberg and Ken Aldape, for their invaluable mentorship throughout my doctoral studies.

I would also like to thank all of the members of the Taylor and Bader labs, in particular Sachin Kumar and Drs. Noriyuki Kijima, Laura Donovan, Sorana Morrissy, Florence Cavalli, Paquito de Antonellis, Shraddha Pai, and Ruth Isserlin for their significant support as wet-bench collaborators, mentors, and friends throughout the duration of my doctoral studies.

I also offer my deepest gratitudes to Meera Kamal for her invaluable administrative support, mentorship, and friendship throughout my time spent as a PhD student.

I would like to especially thank all of my friends and family, in particular Annie Matan and Grace Matan, for their kindness, support, and encouragement throughout my doctoral studies.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: Introduction

Chapter 1: Introduction

1.1 Medulloblastoma (MB): A cerebellar cancer of childhood that affects patients world-wide

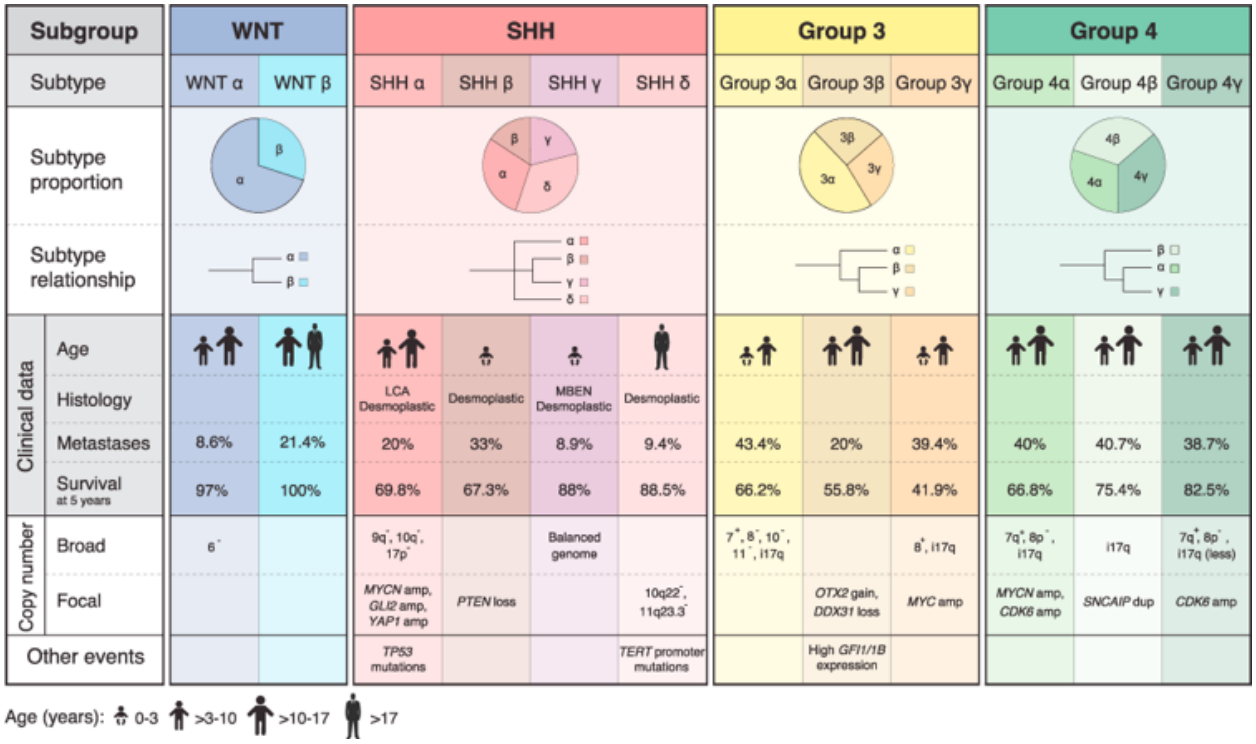
Medulloblastoma (MB), a cancer of the cerebellum, is the most common childhood brain malignancy. Globally, MB accounts for 20 – 25% of all paediatric brain tumours in high-income countries [1, 2], and 6 – 49% of all diagnosed brain tumours in low and middle-income countries [2]. Furthermore, 5-year survival rates have increased, due to improved treatment methods, now ranging from 70%-85% in high-income countries [1]. In contrast, survival is more variable in low and middle-income countries with rates ranging from as low as 33% to as high as 73% [1]. So, MB is a world-wide disease worthy of world-wide attention.

1.2 Characteristics of MB subgroups and subtypes

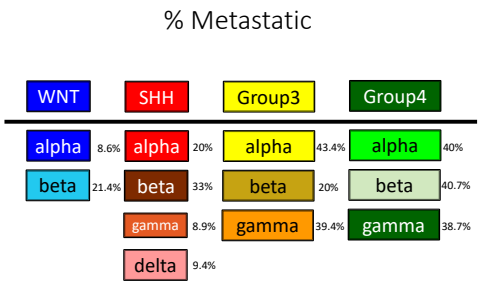
MB does not present as a single disease; rather, MB is a paediatric cancer with several subgroups. Consolidating the work of several research teams over many years, a consensus was first reached in 2012, that 4 subgroups of MB exist, given the names WNT, SHH, Group 3, and Group 4 [3]. Subgroups were defined transcriptionally, through the clustering of MB expression data. Subgroups were named based on the signaling networks most frequently disrupted if known. Building upon this foundation, by including both transcription and epigenetic (methylation) data in clustering, Cavalli and colleagues demonstrated that the MB subgroups could be further divided into subtypes [4] (Figure 1a). In particular, the percent of patients with metastatic MB at the time of clinical presentation (% Metastatic) (Figure 1b), the percent of patients, who survive for at least 5 additional years after clinical presentation (% Survival) (Figure 1c), and the specific genes mutated or misexpressed (Figure 1d) vary between subgroups and subtypes. Early clues as to the mechanisms driving MB metastasis further suggest that within each subgroup, the genes and pathways driving metastatic dissemination frequently differ from those driving primary tumour growth and development [5]. Due to a limited availability of MB metastatic tissues and a low number of pre-clinical animal models of MB metastasis, research into mechanisms of MB metastasis has been similarly limited [5]. Thus, in-depth investigation of these mechanisms, using readily available tools, is needed to deepen our understanding of the subgroup-specific mechanisms driving metastasis, in order to support the

development of subgroup-specific therapeutics that, in their specificity, hold promise to be more effective and safer than current therapeutic approaches

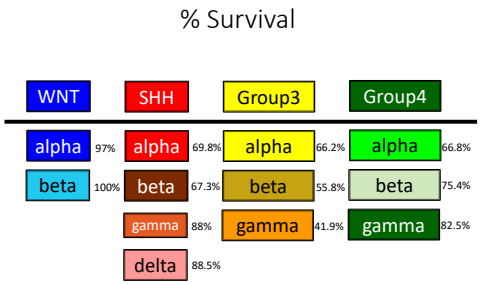
a)



b)



c)



d)

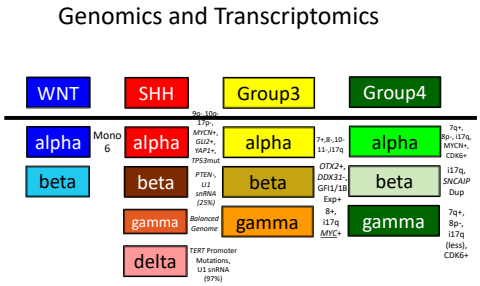


Figure 1. Medulloblastoma: One family of cancers with 12 distinct subtypes. a) Graphical depiction of the 4 MB subgroups and 12 subgroup-specific subtypes b) % metastatic – % of MB patients who present with metastatic disease c) % survival – % of patients who survive for at least 5 years post-presentation d) Genomics and transcriptomics – Genes mutated or misexpressed within each subgroup and subtype. + = Amplification; - = Deletion; Dup = Duplication; Mono = Monosomy; i = isochromosome.

1.2.1 Characteristics of MB Subgroups and Subtypes: % Survival is Correlated with Age of Onset and % Metastatic

Across MB subgroups and subtypes, the age of onset of MB symptoms is correlated with the percent of patients that survive at least 5 years after clinical presentation (% Survival) [4] (Figure 1c). MB symptoms first appear in WNT patients between the ages of 3 and 17 years for WNT-alpha, and from 10 years to adulthood for WNT-beta. Comparatively, no Group 4 MB patients ever reach adulthood without symptoms, usually first presenting between 3 and 17 years. Significantly greater heterogeneity exists between SHH subtypes, with symptoms first appearing in infancy (beta, gamma), adolescence (alpha), or adulthood (delta). Group 3 subtypes exhibit similar heterogeneity with alpha and gamma patients exhibiting symptoms as early as infancy and up to the age of 10 years, and symptoms first appearing in beta patients during adolescence.

The percent of MB patients with metastatic MB at clinical presentation (% Metastatic) also varies across subgroups and subtypes [4]. Metastatic MB is most commonly observed in Group 3 and Group 4 patients at presentation, followed by SHH, and specifically WNT-beta (Figure 1b).

Generally, % Survival is positively correlated with age of onset and negatively correlated with % Metastatic [4]. Demonstrating this in practice, among the highest-risk MB patients, mortality is nearly always attributable to the primary tumour having metastasized from its site of origin in the posterior fossa to the leptomeninges [5]. Indeed, across SHH subtypes increasing % Survival is correlated with decreasing % Metastatic [4] (Figure 1a). However, across Group 3 subtypes, % Survival correlates with slightly increasing % Metastatic. Similarly, in SHH-beta and SHH-gamma subtypes, MB appears in infancy yet % Survival is high (Figure 1a). So, this suggests that, while % Metastatic does correlate with % Survival, additional factors also influence % Survival.

1.2.2 Characteristics of MB Subgroups and Subtypes: Symptoms

Patient symptoms at presentation also vary by subgroup [6]. These include headache, vomiting, ataxia, motor challenges, gaze palsy, behavioral changes/irritability, extracranial pain, vertigo, hearing loss, lethargy, and acute intracranial pressure elevation. Symptoms are observed preferentially in association with specific subgroups. For example, headaches are reported more often by WNT and Group 4 patients than SHH or Group 3 patients, while vomiting is more frequently characteristic of Group 3 and Group 4 patients [6]. Furthermore, the pre-diagnostic

interval, namely the time between symptom appearance and disease diagnosis, is subgroup-specific; WNT and Group 4 patients have a much longer pre-diagnostic interval than patients with SHH or Group 3 MB tumours [6].

1.2.3 Characteristics of MB Subgroups and Subtypes : Location & Histology

The subgroup specificity observed among patients is similarly present at the cellular level, considering tumour location, histology, and frequency of metastasis [4]. WNT, Group 3, and Group 4 tumours form along the midline while SHH tumours form peripherally. This difference between SHH and other subgroups is also evident with respect to histology; WNT, Group 3, and Group 4 tumours exhibit classic histology while SHH tumours tend to be desmoplastic. Further variation in histology is also observed within the SHH subgroup, with some subtype alpha tumours exhibiting large cell anaplastic (LCA) histology, and MB extensive nodularity (MBEN) observed for some subtype gamma tumours.

1.2.4 Characteristics of MB Subgroups and Subtypes: Genomics & Transcriptomics

Although rare, certain genomic and transcriptomic changes recur in specific MB subtypes [4] (Figure 1a,d). Copy number gains or losses, affecting single genes or large chromosome regions, are observed only in tumours from a specific subgroup or subtype. Similarly, mutations or changes in the expression of single genes, and by extension, changes in the activity of specific pathways occur in a subgroup and subtype-specific fashion.

Within the WNT subgroup, WNT-alpha tumours have only a single copy of chromosome 6, while WNT-beta tumours are frequently diploid for chromosome 6 [4]. Prior to the definition of MB subtypes, it was thought that monosomy 6 was a characteristic of all WNT tumours. So, the knowledge of MB subtypes will prevent WNT-beta patients from being misdiagnosed. Considering pathway activation, only the WNT signaling pathway is overrepresented in WNT tumours [3]. However, it is not yet known whether that is subtype-specific or in all WNT tumours.

Significant genomic variation is also observed across the subtypes of the SHH subgroup. Local amplification of *MYCN*, *GLI2*, and *YAP1*, mutation of *TP53*, and loss of 9q, 10q, and 17p are

commonly observed in SHH-alpha tumours. Comparatively, SHH-beta tumours harbor *PTEN* deletions, and a variety of other focal amplifications, the largest being on 2p and 2q. In contrast to the first two subtypes, chromosome gains or losses for SHH-gamma are limited to focal deletions on 9q and 10q. Again, different from the three previous subtypes, the most frequently recurrent events for SHH-delta tumours are mutations within the *TERT* promoter. At the pathway level, DNA repair and cell cycle related pathways are enriched in SHH-alpha tumours. Furthermore, developmental pathways are enriched in both SHH-beta and SHH-gamma tumours, with the magnitude of enrichment being greater in SHH-gamma than in SHH-beta. Of even greater interest are enriched pathways that are actionable, including sumoylation and DNA repair pathways in SHH-alpha, receptor tyrosine kinase signaling and ion channels in SHH-beta and SHH-gamma, and telomere maintenance in SHH-alpha and SHH-delta tumours.

For Group 3 tumours, the *MYC* locus on the long arm of chromosome 8 (8q24) is subject to change in both Group 3-alpha and Group 3-gamma but in Group 3-alpha it is deleted, while in Group 3-gamma it is amplified [4]. In contrast, activation of *GFI1* and *GFI1B*, and inactivation of *DDX3I*, all likely through enhancer hijacking, are observed for Group 3-beta tumours, as well as amplification of the *OTX2* locus. In terms of pathway enrichment, Group 3-alpha exhibit enrichment of photoreceptor pathways, mechanisms of muscle contraction, and pathways related to the primary cilium. Furthermore, the actionable pathways of protein translation are enriched in Group 3-beta and Group 3-gamma, and the similarly actionable pathway of telomere maintenance is enriched in Group 3-gamma [4].

Finally, Group 4 tumours show similar diversity. *MYCN* amplification is observed in only Group 4-alpha tumours. In addition, both 8p loss and 7q gain are specific to both Group 4-alpha and Group 4-gamma, as is focal amplification of *CDK6*. Furthermore, it is interesting to note that *SNCAIP* duplication, and isochromosome 17q, originally identified across Group 4 tumours without definitions of subtype [7] , are actually specific to only Group 4-beta. With respect to pathways, cell migration is enriched in Group 4-alpha, not unexpected given its frequency of metastasis of 40%. The actionable MAPK and FGFR1 pathways are enriched in Group 4-beta tumours, and the equally actionable PI3K-AKT and ERBB4-mediated nuclear signaling pathways are enriched in Group 4-gamma tumours.

Taken together, the significant diversity in genomic changes and pathway enrichment that exists across the subtypes of any of the 4 MB subgroups reinforces how imperative it will be to consider both subgroup and subtype affiliation rather than subgroup alone when designing new therapeutic approaches. Only by considering both subgroup and subtype will it be possible to maximize the frequency of treatments being both effective and safe.

1.2.5 Characteristics of MB Subgroups and subtypes : Sex-Bias

Males are at higher risk of having MB than females, but only in specific subgroups [8]. For the WNT and SHH subgroups, no difference in risk exists between males and females. In contrast, males are at 2-fold greater risk for Group 3 MB, and 3-fold greater risk than females for Group 4 MB [8]. Survival among MB patients similarly is worse for males than females. It has not yet been definitively determined whether the effect of sex bias on survival varies across subtypes [9]. Furthermore, the underlying mechanism conferring sex-specific risk for MB is still not understood. However, recent work has suggested that IL6-STAT3 signaling may, in part, account for the worse survival that has been observed for male MB patients relative to females in multiple MB subgroups [10, 11]. Biallelic deletion of *STAT3* in the granule cell precursors of SHH $Ptch^{lacZ/+}$ model mouse embryos prevented SHH tumour formation and increased survival in males but had no effect on either tumour formation or survival in females [10]. Furthermore, biallelic inactivation of *STAT3* led to activated SHH signaling in females and suppressed SHH signaling in males. This suggested that STAT3p might have been serving opposite roles in male and female mice, namely as an activator of SHH signaling in males and as an inhibitor of SHH signalling in females. Even more interesting, in human MB patients increased *STAT3* transcription was associated with decreased survival in males and increased survival in females. Building upon these findings, expression of cytokine Interleukin 6 (*IL6*), normally expressed in only male mice and MB patients, promoted SHH signaling in cultured mouse embryonic fibroblasts. With respect to Group 3 MB, intact IL6-STAT3 signaling was necessary and sufficient for cultured Met8A cells (Group 3 MB) to become resistant to chemotherapeutic agents vincristine, cisplatin, mitoxantrone, and idarubicin used alone through repeated training using increasing vincristine concentrations [11]. However, the cell line remained chemosensitive to vincristine used in combination with cisplatin or STAT3 inhibitor niclosamide. *STAT3* transcription was also shown to vary across the subtypes of Group 3 MB with Group 3-beta and Group 3-gamma subtypes exhibiting increased *STAT3* relative to Group 3-alpha. This is

especially interesting since Group 3-gamma patients have % survival lower than would be expected based on their % Metastatic alone. So could higher *STAT3* expression in Group 3-gamma MB patients be an additional factor bringing about the lower than expected rate of survival observed? Taking one step further, among SHH, Group 3, and Group 4 MB, *STAT3* transcription was lowest in SHH, higher in Group 3, and highest in Group 4. This is in direct alignment with the magnitudes of sexual dimorphism that are observed across these subgroups [8]. Thus, this association serves as a clue that dysregulation of *STAT3* may be, at least in part, responsible for the increased risk and decreased survival that is observed for male MB patients relative to females.

1.2.6 Subgroup/Subtype Specificity: A Product of Differences in Cellular Ancestry?

For a long time the origins of the transcriptional and epigenetic differences distinguishing MB subgroups and subtypes remained unknown. However, transcriptional programs in MB tumours from different subgroups are now known to closely resemble conserved transcriptional programs characteristic of specific neural cell progenitors, each active during specific stages of brain development [12]. WNT tumours most closely resemble mossy fiber precursor cells from the lower rhombic lip [13]. In contrast, SHH tumours are most similar to granule cell precursors, Group 3 to Nestin+ stem cells, and Group 4 tumours to unipolar brush cell progenitors. all three from the upper rhombic lip [12]. Knowledge of the developmental programs associated with these specific cell lineages could yield insights into the developmental points of origin for each of the MB tumour types. This could serve as a foundation for the development of new therapeutic options taking this into consideration.

1.3 The Mutational Spectrum of MB

1.3.1 Sequence Changes

Changes in the transcription of specific genes mediate MB tumourigenesis. This misexpression can be brought about through the mutation of the genomic sequence. This can take the form of single-nucleotide variations (SNVs), somatic copy number aberrations (SCNAs), larger-scale insertions and deletions of large chromosome regions, or complete loss of chromosomes. In all of these cases what results is a significant change in the expression of a gene or genes, leading to tumourigenesis, and tumour progression.

1.3.2 Changes in the Regulation of Gene Expression

Alternatively, the aberrant changes in transcription which drive tumour progression, can arise through aberrant changes to epigenetic mechanisms, which normally serve to strictly regulate the transcription of the genes in question. . Enhancer hijacking is an example of one such mechanism. Through a chromosome rearrangement, an enhancer cis-regulatory element or region is placed upstream of a gene with which it is not usually associated; as a result, the transcription of that gene is increased several-fold leading to tumourigenesis [14]. There is no change to the wild-type sequence of the gene; only the regulatory mechanisms controlling expression are subject to mutation. In the context of Medulloblastoma, aberrant transcription is also achieved through the aberrant activity of other epigenetic mechanisms, including but not limited to the aberrant methylation of lysine residues within histones [15], the upregulation of chromatin remodeling proteins [16], and aberrant CpG island methylation [17]. So, taken together, transcriptional changes, which drive MB tumour progression, are achieved both through mutation of gene sequences, and aberrant regulation of transcription.

1.4 MB Treatment

1.4.1 Current MB Treatment Regimen

Currently, MB patients are treated with surgery, cerebrospinal irradiation, or chemotherapy either alone or in combination depending on the age of the patient [18] (Personal communication with Dr. V. Ramaswamy). Specifically, MB patients 5 years of age or older undergo surgery, followed by cerebrospinal radiation (23.4 gray (gy) for standard-risk patients, or 36-39 gy for high-risk patients), and subsequently are administered 4-9 rounds of chemotherapy consisting of cisplatin (cumulative dose of at least 300 mg/m^2), an alkylating agent (cyclophosphamide or 1-(2-Chloroethyl)-3-Cyclohexyl-1-Nitrosourea (CCNU)), and vincristine. Patients younger than 5 years of age are not given radiation. Instead, they undergo surgery followed by chemotherapy consisting of 3 rounds of cisplatin, cyclophosphamide, etoposide, vincristine and methotrexate, with a total cisplatin dosage of 300 mg/m^2 (for infants) followed by three rounds of high-dose carboplatin and thiotepa.

These intense therapeutic approaches are effective at treating tumours for a subset of MB patients. However, for patients with recurrent MB, these treatments frequently prove to be

ineffective at treating the tumour. To make matters worse, these treatments often leave patients with severe, debilitating side effects, and a deteriorating quality of life that persists through the rest of their lives [18]. It is thought that this low efficacy and these side effects may be the result of heterogeneity among standard risk and high-risk patients not being considered in the design of therapy. What better motivation could there be to seek a better understanding of the mechanisms of MB metastasis, to develop new patient-centred therapies rather than a one-size-fits-all approach.

1.4.2 Next-Generation MB Therapy

1.4.2.1 Overview

In order for new approaches to MB therapy to be both safer and no less effective than current approaches, therapeutics have tended to arise within discrete categories, namely patient-specific tailoring of existing therapies, subgroup-specific therapies, immunotherapy, the development of new vehicles for drug delivery across the blood-brain barrier, and targeting of MB stem cells.

1.4.2.2 Patient-specific Tailoring of Existing Therapies

It has been observed that the statically-defined therapeutic regimens described above are associated with heterogeneous outcomes among MB patients [18]. Trials have been run to see if patient-specific modifications to therapy might offer better outcomes relative to existing regimens. In the phase III Children's Oncology group trial (ACNS0331), a 5.4 Gy reduction in cerebrospinal irradiation (CSI) was attempted for patients 3-7 years of age, to see if this would be associated with similar treatment efficacy and fewer side effects [19]. Unfortunately, reduced CSI led to worse overall survival (OS) relative to patients given standard therapy. In contrast, in the HIT'2000 trial, strengthened chemotherapy combined with hyperfractionated CSI (increased number of CSI treatments but each at a lower dose) led to OS of 74% specifically for M2/3 MB patients with dissemination to the leptomeninges, significantly improved from a preceding trial, HIT'91. The reason for the improved outcomes in the HIT'2000 is thought to have been, in part, due to the hyperfractionated CSI protocol, but also significantly due to advances in imaging, surgical practice, and supportive care since HIT'91 was completed [20]. So, tailoring of existing therapies to best suit specific groups of patients can yield outcomes better than what is possible with one-size-fits-all therapy, but not always.

1.4.2.3 Subgroup-Specific Therapy

With the knowledge that MB subgroups and subtypes differ significantly, one new therapeutic approach has been to create therapies that are subgroup or subtype-specific [18]. In practice, this means the targeting of subgroup-specific molecules or adapting of therapies to be as safe and effective as possible for a specific subgroup.

WNT tumours have a leaky BBB rendering tumours inherently chemosensitive, which is in part why WNT tumors have such positive outcomes using existing therapies [21]. In addition, WNT-specific pathways are critical for bone regeneration and tissue formation in non-cancerous tissues [22]. So rather than risk harmful side effects, no WNT pathway-specific therapies are being developed. Instead, ongoing clinical trials are exploring whether current therapies can be de-escalated for low-risk WNT patients without sacrificing efficacy. If this is possible, this would also mean less severe side-effects.

Inhibitors of SHH-specific protein smoothened (SMO) are the class of molecules most frequently tested as a potential SHH-specific therapy [18, 23]. Inhibition of SMO using small molecules such as vismodegib (GDC-0449) [24, 25], blocks signaling downstream of SMO, translocation of Gli, and ultimately the activation of Hedgehog target genes [18, 23]. Of SHH-MB patients, 80% carry mutations in the Patched *PTCH1* or *SMO* genes [18]. Thus, it is not surprising that Vismodegib is an effective treatment for recurrent SHH MB. However, SMO inhibition also interrupts the development of normal bones and teeth. So, also being considered are bromodomain inhibitors, inhibitors of G2/M proteins AURK and PLK, and inhibitors of cMET [26].

For Group 3 and Group 4 knowledge of the biology of Group 3 and Group 4 is still limited so subgroup-specific primary therapies do not yet exist, and current approaches involve the testing of new forms of nonspecific chemotherapy [18]. However, for recurrent Group 3 MB, inhibitors of Pi3K and mTOR signaling are being tested [27].

Taken together, treating MB patients with inhibitors to proteins or pathways specific to their subgroup, relative to the use of more generic inhibitors, by its specificity the likelihood that the therapeutic agent will effectively inhibit tumour growth will be greater, and the likelihood that the therapeutic agent will affect proteins and pathways other than those intended will be less. So,

subgroup and subtype-specific therapies do hold the promise of greater efficacy and safety as a product of their specificity to the subgroup being treated.

1.4.2.4 Immunotherapy

A wide variety of forms of immunotherapy are also being considered as alternative approaches, separate from targeted therapies, to the treatment of MB, especially at advanced stages.

CAR T-based immunotherapy is an innovative approach in which T- cells are engineered ex-vivo to express chimeric T-cell receptors, which recognize tumour-specific antigens, and are able to bind T-cells [28]. Upon being reintroduced into mouse models or human tumours, these engineered cells render the tumours visible to the resident immune system, enabling the tumour to be eliminated like any other foreign invader. In practice, Car-T cells engineered to recognize Her2 were able to clear out Her2-expressing tumour cells implanted into the posterior fossa of preclinical mouse models of human MB [29]. Also, in the BrainChild-01 clinical trial (NCT03500991) CD4+ and CD8+ cells transduced with lentivirus expressing chimeric Her2 receptor and a truncated Epidermal Growth Factor Receptor (EGFR), and are being introduced into the tumour resection cavity of the ventricular system of MB patients with Her-2 expressing recurrent tumours.

A second form of immunotherapy being used to treat MB tumours is the use of immune checkpoint inhibitors [18]. Cancer cells overexpress immune checkpoint proteins, to eliminate T-cell responses, which might otherwise lead to their detection and removal. Thus, through the use of inhibitors to immune checkpoint proteins, the immune response can once again be activated in response to the presence of tumours. Checkpoint proteins currently being targeted for inhibition include CTLA4, PD-1, and PD-L1.

A third therapeutic approach, which functions independently, and in conjunction with the immune system, is the use of oncolytic viruses (OVs) [18]. OVs are made such that they replicate only in cancer cells. As a result, neighboring normal cells are unaffected by the presence of OVs, and once OVs infect a small number of tumour cells, they rapidly spread throughout the tumour cell population. Of even greater benefit, lysis of tumour cells by OVs leads to the release of neoantigens that can then be recognized by the immune system, which can

further act to remove any tumour cells that remain. Several different viruses are being used in clinical trials as potential treatments for recurrent MB.

Lastly, trials are investigating whether Natural Killer (NK) cells represent another viable tool for immunotherapy. NK cells participate in the endogenous response of the immune system to the presence of malignant tumours by recognizing tumours in an antigen-independent fashion [30]. NK cells are controlled by stimulatory and inhibitory receptors; so several trails have involved the creation and testing of modified NK cells expressing receptors such as dominant-negative TGF-beta II (DNRII) [31],

1.4.2.5 Conjugates to Facilitate Delivery of Therapeutics Across the Blood Brain Barrier (BBB)

One of the greatest challenges associated with current approaches to MB therapy has been the extent to which therapeutics are able to cross the Blood Brain Barrier (BBB) [32]. Several new delivery approaches have been developed and have been tested to see if they increase the effectiveness of drug treatment. Angiopeps, peptides based on the sequence of peptide aprotinin, have been designed and used for drug delivery given their demonstrated ability to bind to low-density lipoprotein receptor-related protein (LRP), and mediate transport across the BBB. Chemotherapeutic agents etoposide and doxorubicin, conjugated to Angiopep-2, a 19-aa Angiopep [33], via glutaric acid and succinic acid linkers, respectively, were less toxic, and better distributed in the brains of mouse models than the chemotherapeutics not conjugated to Angiopep-2 [32]. Also, Bovine Serum Albumin (BSA) nanoparticles containing cisplatin (CPT), capable of glutathione-sensitive CPT release, are readily taken up into MB cells in culture (Daoy), and are cytotoxic [34].

Taken together, targeted therapy, immunotherapy, and new methods of drug delivery are all being investigated as potential alternative approaches to therapy for MB, either independently or in conjunction with existing therapies. The major advantage that these approaches offer in contrast to existing therapies is specificity, which can enable therapeutic approaches to be both safer and more effective.

1.5 The Biology of Metastasis

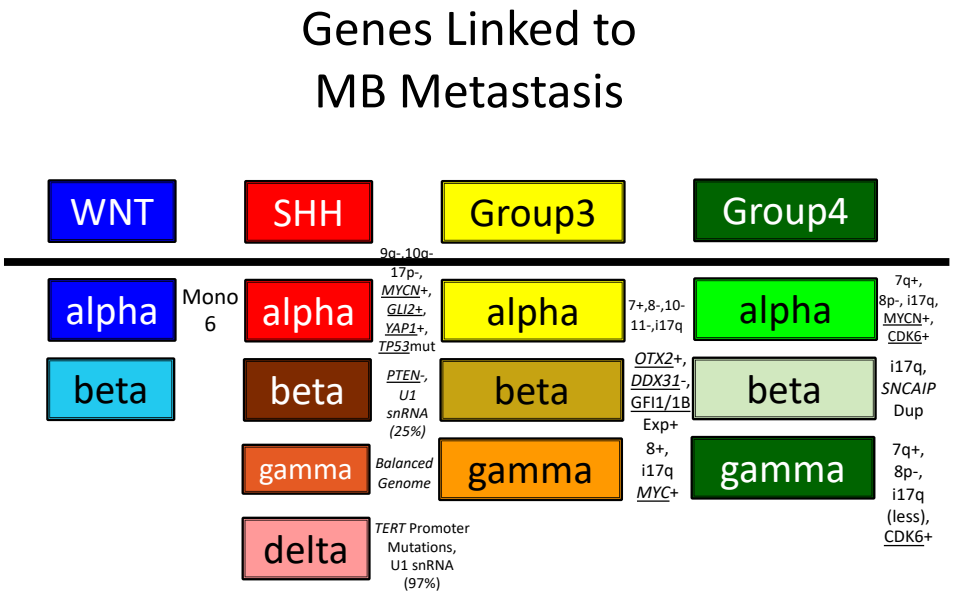
Tumourigenesis first takes place in the cerebellum, and through rapid cell division tumour cells become a genetically heterogeneous malignant tumour. Additional sequence changes then occur in specific cells of the primary tumour; these changes confer upon those cells the ability to detach and migrate away from the primary tumour, and enter circulation, specifically either the cerebrospinal fluid (CSF) or the blood. These environments serve as means for cells to migrate locally to the leptomeningeal space (SHH), or distally to the spinal cord, lymph nodes, liver, or spleen (Group3 , Group 4) [35]. Arriving at the secondary site, cells invade and divide to form micro metastases; with additional genomic changes, micro metastases further develop to form macro metastases. The molecular biology of tumour detachment, migration, invasion, growth and transition from micro to macro metastases are still unclear. However, these are all active areas of research.

1.6 Subtype-specific Genomics/Transcriptomics Define Subtype-specific Mechanisms of Metastasis

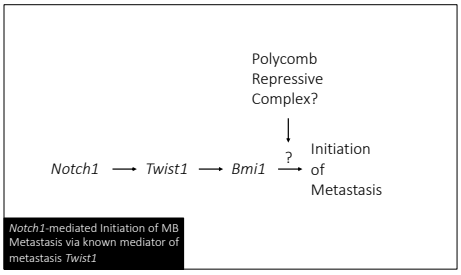
Increases in the activity of specific genes, and pathways, and the existence of specific chromosomal rearrangements are correlated with tumour progression and metastasis of MB tumours from all subgroups and subtypes (Figure 2a); however the understanding of how these and other genomic changes mediate the metastasis of tumours from each subtype is still rudimentary, lagging significantly behind our comprehensive understanding of the mechanisms of tumourigenesis. As a result, significant scientific attention is being paid to the biology of MB metastasis. This has begun to yield first glimpses of what may be taking place. *Notch1* has been linked to the initiation of metastasis for Group 3 MB tumours (Figure 2b); *Notch1* increases the transcription of known premetastatic transcription factor *TWIST1*, which then activates *Bmi1*, which initiates metastasis either alone or as a subunit of the Polycomb Repressive Complex 1 (PRC1) of which it is known to be a component [36, 37]. Comparatively, increased *MYC* transcription in response to Thrombospondin-1 (*TSP1)* inactivation is sufficient to promote cell migration, and invasion [38]. From a different perspective, epigenetic mechanisms have recently been implicated in the driving of MB metastasis; long non-coding RNA (lncRNA) *LOXL1-AS1* promotes MB tumour progression and metastasis by activating the PI3 Kinase (PI3K)/AKT pathway by promoting the phosphorylation of PI3K, and AKT (Figure 2c). Knockdown of

LOXL1-A1 leads to decreased levels of phosphorylated PI3K and AKT, while the level of non-phosphorylated protein remains the same [39]. Similarly, lncRNA *CCAT1* promotes MB metastasis through the activation of the MAPK pathway [39]. Taken together, the first clues of mechanisms mediating MB metastasis are coming to light, and this will facilitate the in-depth characterization of the mechanisms driving metastasis in each of the subgroup-specific subtypes of MB.

a)

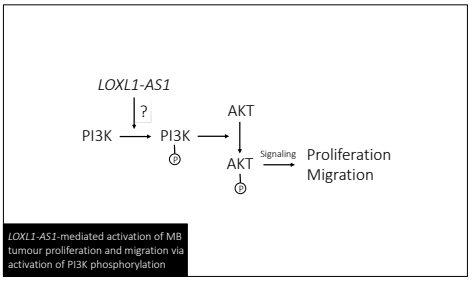


b)



Described in Kahn *et al.*, 2018

c)



Described in Gao *et al.*, 2018

Figure 2. Mechanisms of MB metastasis: First clues. a) Subtype-specific genomics and transcriptomics of MB to date with links to metastasis noted. Underline = Literature links to metastasis; + = Amplification; - = Deletion; Dup = Duplication; Mono = Monosomy; i = isochromosome b) Notch1-mediated initiation of Group 3 MB Metastasis c) LncRNA *LOXL1-AS1* promotes PI3K phosphorylation leading to AKT phosphorylation, increased cell proliferation, and increased cell migration.

1.7 Using Bioinformatics to Identify from Within Bulk RNA Sequencing (RNA-Seq) Data Prometastatic Genes Activated in the Spine Metastases of Mouse Patient-Derived Xenograft (PDX) Models of Human MB Subject to Iterative Selection for Metastasis Propensity

1.7.1 Overview

Throughout my doctoral studies, four powerful technologies, namely Patient-derived Xenograft Modeling of human MB tumours, In vivo Iterative Selection, Bulk Paired-end RNA-Seq, and Bioinformatics analysis, have been used by bench-based collaborators and myself in order to identify genes and pathways, which are activated in metastases to the spine relative to primary tumours. These novel correlates of metastasis could represent new insights to further our understanding of MB metastasis, and new options to consider in the development of next-generation MB therapeutics.

1.7.2 Model Human MB Without the Limitations of Tissue Availability Using Patient-derived Xenograft (PDX) Mice

The amount of metastatic MB tumour tissue available for research purposes is very small [5]. This is largely due to the fact that the accepted treatment regimen for metastatic MB is radiation and chemotherapy, and not surgery. With the use of this treatment regimen metastases are rarely removed. An effective alternative has become patient-derived xenograft (PDX) mouse models [40]. Human tumour cells are introduced into a mouse, usually through injection, tumor cells engraft among existing tissues, undergo rapid cell division, and develop into a primary tumour. In some cases, genomic changes, genetic or epigenetic, occur in cells of the primary tumour, which confer upon those cells the ability to separate from the primary tumour, migrate away from the site of tumor formation to local or distal secondary sites within the body, where they establish new tumors known as metastases. Multiple mice are injected in this manner to produce sufficient tumour material, both primary and metastatic, for study. This approach thus represents a way in which to study human tumour biology in vivo without the limits associated with the use of patient tissues.

Modeling of human MB tumours using mouse PDX models does however have limitations. Most importantly, the microenvironment of the surrogate mouse must be significantly modified to

permit the growth of human tumours. Specifically, for human MB to grow inside a mouse the mouse must be deficient in both innate and acquired immunity, lacking B cells, T cells, and Natural Killer cells [40]. The NSG mouse, produced by The Jackson Laboratory, is one such mouse strain frequently used for this purpose [41]. This better recapitulates the human microenvironment but is still in a mouse. Overall, even with their limitations, PDX models are powerful tools to rapidly and cost-effectively identify and characterize new models of human tumours that can then be validated and elucidated by other means.

1.7.3 Iteratively Select for Characteristics of Interest and their Underlying Genomics and Transcriptomics in Mouse PDX Models of Human Disease

In vivo iterative selection is an effective method to select for the presence of specific characteristics and/or behavior in cells. In vivo iterative selection was first designed and used by Fidler and colleagues in 1973, to select for characteristics of melanoma cells grown subcutaneously in C57B1/6J mice, which metastasized to the lungs after intravenous injection into C57B1/6J mice [42]. Fidler and colleagues removed metastatic nodules from the lungs of injected mice, grew the cells in culture, and upon reaching confluence intravenously injected the cells into a new mouse. This process was repeated five times to optimally select for cells with the greatest propensity for metastasis, namely increased ability to survive, and form secondary tumours. This work proved that in vivo iterative selection could be used to identify specific characteristics of tumour cells correlated with an increasing propensity for metastasis. Using this approach in combination with differential expression analysis, Clark and colleagues were able to show that increased transcription of *RhoC*, a small GTPase, in melanoma metastases to the lungs relative to cells in the primary tumour bed, enhanced metastasis, specifically the invasion of tumour cells into the lung [43]. Most recently, using in vivo iterative selection combined with differential expression analysis, Dr. Adrienne Boire identified the gene for Complement C3 as being a gene for which transcriptional activation was necessary and sufficient for the metastasis of breast and lung carcinoma cells from the CSF to the leptomeningeal space [44]. To select for cells that could survive the harsh conditions of the cerebrospinal fluid (CSF), Dr. Boire directly inoculated the CSF through the Cisterna Magna then collected cells that survived and metastasized to the leptomeninges. Boire cultured these cells, and then used them to re-inoculate a second mouse in an identical fashion. After 3 rounds of selection in this manner, Boire cultured

the cells from the leptomeninges, considering them now to be “Intermediate” cells (INT), now comprising cells with propensity for growth in the CSF. Boire then introduced INT cells into the arterial circulation of new mice. Cells that metastasized from the blood to the leptomeninges were isolated, and analyzed using RNA-Seq. It was in these cells that *C3* transcription was significantly increased. Boire then went even further to demonstrate that *C3* from the tumour cells was released and interacted with C3 receptors embedded in the surface of blood vessels. This served to make the blood-brain barrier permeable, facilitating both the release of compounds into the CSF, rendering it capable of supporting tumour cell growth, and the migration of tumour cells across the blood-brain barrier into the CSF. Lastly, to strengthen the link between *C3* and leptomeningeal metastasis, Dr. Boire demonstrated that interference in *C3* signaling led to inhibition of metastasis. So, Dr. Boire’s thorough characterization of *C3* signaling in the context of metastasis beautifully demonstrates how powerful iterative selection and transcriptional profiling can be to identify and characterize novel mechanisms of metastasis.

1.7.4 Use Bulk RNA Sequencing (RNA-Seq) to Accurately Quantify Transcription Across the Genome

The development of next generation sequencing (NGS), first named Massively Parallel Signature Sequencing (MPSS), by Dr. Sydney Brenner and colleagues in 2000 [45], came the ability to sequence at rates never before possible. Applying NGS to transcriptomics, Bainbridge and colleagues created RNA Sequencing (RNA-Seq), and used it to sequence the complete transcriptome of the prostate cancer cell line LNCaP [46]. From this point forward, RNA-Seq became, and continues to be the preferred method for large-scale expression analysis. Relative to microarray-based approaches variation between replicates was significantly less, and expression of all transcripts could be measured not just those specifically included on a microarray. So, RNA-Seq is a powerful way to quickly and comprehensively profile genome-wide transcription. However, to assess changes in transcription between datasets requires the tools of bioinformatics.

1.7.5 Conduct Differential Expression Analyses and Pathway Analyses In silico to Identify Genes and Pathways Significantly Activated in Tissues of Interest

Bioinformatic identification of genes and pathways that are activated in a specific cellular or developmental context is an especially powerful way in which to rapidly identify genes that might be worthy of further study to determine what role their activation plays in establishing the

context in which they have been found to be activated. Differential expression analysis of RNA-Seq data is in extensive use within the scientific community to identify genes that are transcriptionally activated in a specific cellular context relative to a second context [47]. Several bioinformatics tools for differential expression analysis now exist, but the shared purpose is to distinguish those genes that are differentially expressed from those that are not.

Similarly, several approaches to pathway analysis have been developed; however, all types fall into one of three categories namely enrichment analyses, functional class scoring, or assessment of pathway topology [48]. Enrichment analyses assess which pathways are overrepresented within lists of differentially expressed genes ranked by a defined characteristic. One example, that I used extensively throughout my studies, is Gene Set Enrichment Analysis (GSEA) [49]. In a version of this method known as GSEA prerankd, defined sets of genes, formed based upon knowledge that all genes belong to a shared pathway, gene family, or other attribute, are compared to lists of differentially expressed genes sorted by decreasing probability of activation. GSEA will identify those gene sets that are enriched among genes specifically at the top or bottom of the list to an extent that is significantly higher than would be expected by random chance. Those are the gene sets that are ultimately identified as being enriched within the expression datasets subject to analysis.

In contrast, functional class sorting examines, across lists of differentially expressed genes, the overall average magnitude of differential expression across all genes in the pathway. GSEA is also capable of running this type of an analysis. Finally, some approaches assess pathway topology and significance based upon integrated sets of data from several pathway databases. One example of this method implemented is Signaling Pathway Impact Analysis (SPIA) [50].

Taken together, differential expression analysis and pathway analysis offer a rapid means to discover previously unidentified associations between specific genes and pathways, and phenotypes of interest, that through further investigation, can potentially serve as valuable insights towards the better understanding of biological systems, and ultimately the development of therapies to use when those systems fail.

1.8 Rationale: A Deeper Understanding of MB Metastasis is Needed

MB tumourigenesis is much better understood than MB metastasis [5]. This deeper understanding has informed therapeutic development and thus has led to the current state in which therapies for primary tumors are more effective than those for metastases. To develop therapies for metastatic MB that are safer and more effective than those that exist will require a deeper understanding of the mechanisms that drive metastasis within each MB subgroup and subtype.

In the context of Medulloblastoma, the mechanisms of MB primary tumour initiation and progression have been studied extensively; the resulting insights into the biology of MB tumourigenesis have served as a strong foundation for the improvement of current therapies for primary MB and the development of new approaches that are both safer and more effective than anything that had previously existed. In contrast to our deep knowledge of the biology of MB primary tumours, our understanding of MB metastasis remains rudimentary.

The progress in the understanding of MB primary tumour biology has led to the current reality in which only a small proportion of patients die due to the action of the primary tumour and the majority of MB patient deaths are attributable to the primary tumour having metastasized from the leptomeninges.

1.9 Hypothesis

Specific coding and non-coding RNA transcripts are transcriptionally activated, and specific pathways are activated in MB metastases to the spine relative to primary tumours, and for a subset of those activated transcripts and pathways the increased activity is correlated with metastasis.

Chapter 2: Materials & Methods

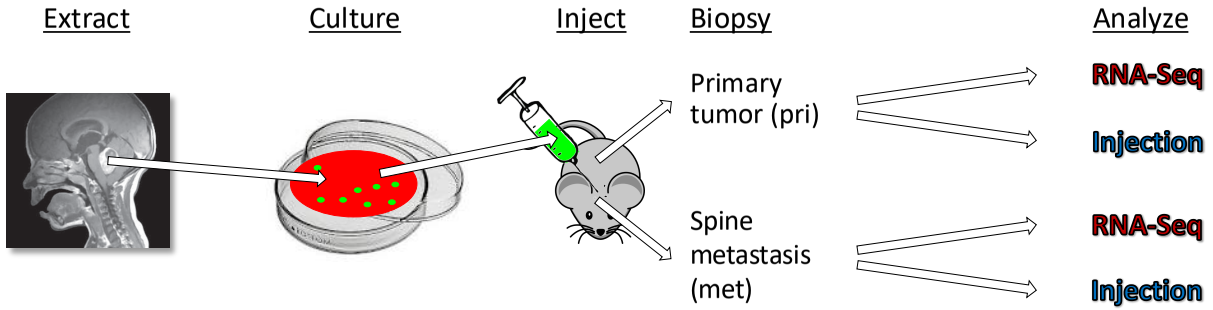
Chapter 2: Materials & Methods

2.1 Cell Culture

GFP-tagged D425 cells, originally from a 5-year old male with Group 3 MB harboring a *c-Myc* amplification, and a 10q deletion [46,XY,i(17q),del(10)(q222), +DMs] [51], were obtained from Dr. Yoon-Jae Cho's lab at Stanford, and grown in 10% FBS in Iscove's Modified Eagle Medium (IMEM) + 5g NaBi , and 1x HEPES. MB411FH cells were obtained from Dr. Jim Olson's lab at the Fred Hutchinson Cancer Research Center. These cells could not be grown in vitro so frozen stocks were thawed and used directly for intracranial injections. Rcmb06 cells (Group 4 MB), characterized by a duplication of the *SNCAIP* gene, were obtained from Dr. Weschler-Reya's lab at the Sanford Burnham Prebys Medical Discovery Institute) and grown in 10% FBS in Iscove's Modified Eagle Medium (IMEM) + 5g NaBi , and 1x HEPES.

2.2 Mouse PDX Models of Human Group 3 and Group 4 MB

To create mouse PDX models of human metastatic MB, MB cells grown in culture (or in the case of MB411FH taken directly from frozen stocks since MB411FH cells could not be successfully cultured in vitro) were intracranially injected into NSG mice. In mice in which primary tumours formed, and migrated to the spine to form distal metastases, cells were extracted and used for either transcriptome analyses by bulk RNA-Seq or injection into additional mice as a part of in vivo iterative selection (Figure 3).



MB Subgroup:		Group 3				Group 4	
MB Cell Line:		D425		MB411FH		SNCAIPDup	
Tissue:		pri	spi	pri	spi	pri	spi
Round of Injection:	1	3 3	4 3	3 3	3 3	2	2
	2	1 3	1 3	1	1		
	3	1 3	1 3				
	4	3	3				
	5	2	2				

REPLICATES
RNA-SEQ
INJECTION

Figure 3. Detailed description of In vivo Iterative Selection. MB tumour cells are cultured (if possible), intracranially injected into NSG mice to create PDX models. Cells from primary tumours and spine metastasis are used for RNA-Seq or Iterative Selection. Blue circles = # of mouse replicates used for RNA-Seq; Red circles = # of mouse replicates used for In vivo Iterative Selection; pri = primary tumour; spi = spine metastasis; Refer to Table 1 for numbers of cells intracranially injected in each round of selection.

25

2.3 In vivo Iterative Selection

The in vivo iterative selection and RNA-Seq analysis described here was performed by my wet bench collaborator, Dr. Noriyuki Kijima, in the Taylor lab, the data from which I then used for my doctoral work in bioinformatics. PDX mice in which MB primary tumours and spine metastases formed were sacrificed and tumour tissues were dissected. From those tissues cells were isolated by FACS sorting based on GFP expression. Purified cells allocated for in vivo iterative selection were then injected intracranially into additional NSG mice to complete the first round of iterative selection. This process was then repeated four additional times to further select for metastasis propensity among spine metastases (Figure 3). Numbers of cells injected varied based upon the number of cells that could be harvested in each round (Table 1a). Similarly, the total number of mice injected varied between rounds of selection and the numbers of metastases that formed per mouse oscillated across rounds of selection (Table 1b). However, all mice intracranially injected with Group 3 MB cells (either D425 or MB411FH) formed spinal metastases (Figure 1b).

a)

MB Subgroup:	Group 3				Group 4	
MB Cell Line:	D425		MB411FH		Rcmb06	
Tissue:	pri	spi	pri	spi	pri	spi
Round of Selection: 1	1000	8250	4000	4000	10000	10000
2	100	100	100	100		
3	100	100				
4	100	100				
5	100	100				

b)

Cell Line	Round of Selection	Total Number of Mice Injected	Number of Injected Mice That Formed Spine Metastases	Number of Injected Mice that Formed Extraneural Metastases	Average Number of Spine Metastases Per Mouse (of those that Formed Spine Metastases)	Average Percentage of Spine Cells that were Metastases
D425	1	6	6	0/2	278	0.002893%
	2	2	2	0/2	401.5	0.017088%
	3	2	2	N/A	371	0.002014%
	4	3	3	N/A	254.5	0.002231%
	5	3	3	N/A	100	0.001019%
MB411FH	1	5	5	1/5	21	0.001556%
	2	2	2	1/2	1285.5	0.050719%
Rcmb06	1	3	2	N/A	53	0.000186%

Table 1. Numbers of cells intracranially injected into xenograft mice, and the frequencies of metastasis in xenograft mice across rounds of in-vivo iterative selection. a) Numbers of cells intracranially injected at each round of selection b) Numbers of injected mice that formed spine metastases, and frequency of metastases across rounds of selection.

2.4 Paired-end Bulk RNA-Seq

In this work, bulk RNA-Seq was used to profile transcriptome-wide expression across samples of primary tumours and spine metastases from PDX mice. Primary tumours and spine metastases were dissected from PDX mice, and tumour cells were isolated by sorting based on tumour cell specific GFP expression. Cells were then lysed, total RNA was extracted, RNA Integrity Number (RIN) values, indicative of the quality of the RNA samples, were calculated, and with RIN values judged to be reasonable, 10 ng of RNA from each sample was then used as input for RT-PCR to generate cDNA libraries. These libraries were then used to conduct paired-end bulk RNA-Seq.

2.5 Bioinformatic Identification of Genes and Pathways Activated in Mouse PDX and Human Spine Metastases Relative to Primary Tumours

2.5.1 Bioinformatics Analysis Overview

I used a multi-step bioinformatics pipeline to identify genes and pathways that were most significantly activated in metastases to the spine relative to primary tumours (Figure 4, Table 2).

Pairs of read files from paired-end bulk RNA-Seq analysis were first subject to FastQC quality control analysis, to verify that all read files were of sufficiently high quality to be included in further analyses. Paired read files were then aligned to the human genome using the RNA-Seq aligner Spliced Transcript Aligned to Reference (STAR). RSEM-Calculate-Expression, a component of the RNA-Seq by Expectation Maximization (RSEM) bioinformatics suite for gene expression quantification and differential expression analysis, was first used to quantify gene expression across the genome using RSEM-Calculate Gene Expression. Output data for all RNA-Seq datasets, measured in Fragments Per Kilobase of transcript sequenced per Million reads (FPKM), was then consolidated into a single file, which was used as input for calculations of pairwise Pearson correlation coefficients for all datasets, and a Principal Component Analyses (PCA).

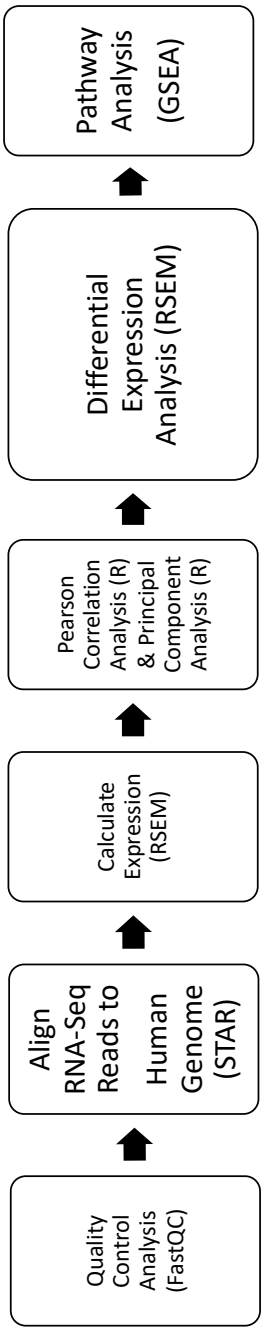


Figure 4. Bioinformatics analysis used to identify and characterize genes and pathways activated in mouse PDX spine metastases relative to primary tumours.

Analysis Type	Tool	Application	Reference
Quality Control	FastQC	Assess the quality of RNA-Seq datasets using several metrics	http://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqcx
Quality Control	ArrayQualityMetrics	Assess the quality of Microarray data using several metrics	Kauffmann A, Gentleman R, Huber W.2009. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics*. 25(3):415-416.
Quality Control	R	Identify and remove outlier RNA-Seq datasets	https://www.r-project.org
Differential Expression Analysis	STAR	Align RNA-Seq reads to the human genome	Dobin, A., Davis, C.A., Schlesinger, F., et al. 2013. STAR: universal ultrafast RNA-seq aligner. Bioinformatics. 29(1):15-21.
Differential Expression Analysis	RSEM	Quantify xenograft RNA-Seq gene expression and identify differentially expressed genes	Li, B., and Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323.
Differential Expression Analysis	LIMMA	Quantify human microarray gene expression and identify differentially expressed genes	Phipson, B, Lee, S, Majewski, IJ, Alexander, WS, and Smyth, GK (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect di erential expression. Annals of Applied Statistics 10(2), 946{963.
Pathway Analysis	GSEA	Identify pathways represented by ranked lists of differentially expressed genes	http://software.broadinstitute.org/gsea/index.jsp

Table 2. Comprehensive list of the in silico tools comprising the bioinformatics analysis pipeline that was used to identify and characterize genes and pathways activated in mouse PDX spine metastases relative to primary tumours. Note: ArrayQualityMetrics was used to assess the quality of microarray expression data from human primary tumours and spine metastases from the Medulloblastoma Advanced Genomics International Consortium (MAGIC) tissue bank.

30

Separate expression data files were also used as input for RSEM-Generate-Data-Matrix, which incorporated specific expression data into consolidated matrices for differential expression analysis with replicates in adjacent columns. These data matrices, one for each differential expression analysis, were passed to RSEM-Run-EBSeq, the R script, which served as a streamlined, user-friendly interface for Empirical Bayes Seq (EBSeq) [52], a powerful yet complicated tool for differential expression analysis, which would have been significantly more challenging to work with directly. Lists of differentially expressed genes and associated statistics were then consolidated into summary tables, and further annotation was automatically downloaded and added to the tables using proprietary BASH and R scripts that I wrote. Microarray expression data from human Group 3 MB primary tumours and spine metastases from the Medulloblastoma Advanced Genomics International Consortium (MAGIC) were then subject to quality control analysis using the R package arrayQualityMetrics, and Limma differential expression analysis. Lists of differentially expressed genes were then compared with those from xenografts.

Lists of activated genes from xenograft and human samples, sorted by decreasing probability of differential expression, were used as input for Gene Set Enrichment Analysis (GSEA). Lists of enriched pathways and associated statistics were then consolidated into summary tables, and pathway-specific genes were added for significantly enriched pathways, using proprietary BASH and R scripts that I wrote.

Other proprietary scripts that I wrote were a variety of BASH and R scripts to generate subset tables from the original summary tables, and scripts to all analyses described here in high-throughput.

2.5.2 Quality Control Analysis Using FastQC

To assess the quality of the expression data obtained from paired-end bulk RNA-Seq, all datasets were analyzed using FastQC [53] (Table 2). Characteristics assessed were overall, per-sequence, and per-base sequence quality, per-base sequence content and N content, per-sequence GC content, the sequence length distribution, the levels of duplication per sequence, overrepresented sequences, and adapter content.

2.5.3 STAR Alignment of RNA-Seq Reads to the Genome

Paired-end reads from bulk RNA-Seq analyses in FASTQ-formatted data files were mapped to specific loci within the human genome (Ensembl GRCh37 [54]) using Spliced Transcript Aligned to Reference (STAR) [55] (Table 2). Proprietary scripts were written in BASH and R to automate alignments.

2.5.4 Measurement of Genome-wide Transcription

Once mapped to specific loci, RNA-Seq by Expectation Maximization (RSEM) was used to quantify gene expression [56] (Table 2). Specifically, the RSEM-Calculate-Expression function measured the transcription of all sequences included in the RNA-Seq data in the unit FPKM, fragments per Kb sequenced per million reads. This specific unit was used in order to avoid underreporting of expression levels, and to account for differences in gene length. With expression levels calculated, the RSEM-Generate-Data-Matrix function was used to construct the data matrix used by RSEM for differential expression analysis.

2.5.5 Pearson Correlation Between RNA-Seq Datasets

To verify that datasets being grouped as biological replicates exhibited significant similarity, and datasets from different tissues exhibited significant difference, Pearson correlation coefficients were calculated for pairwise comparisons between all RNA-Seq datasets being considered. Percent correlation coefficients were grouped into categories: 0-0.50, 0.50-0.75, 0.75-0.85, 0.85-0.99, and 1.00.

2.5.6 Principal Component Analysis (PCA)

A principal component analysis was performed to examine the clustering of expression data before performing differential expression analyses. Coordinates along the primary axis of variation (X-axis) were combined with the complete set of values from each expression dataset for specific scale variables connected to specific steps in the preparation of RNA samples for RNA-Seq analysis. Pairwise Pearson correlation assays were then performed between the X-axis coordinates and the complete set of values for each of the scale variables to determine whether the variation associated with any of these scale variables corresponded to any part of the variation represented by the principal component of variation. Scale variables included in this analysis were i) Number of Sorted Cells, ii) Sorted Cell RNA Concentration (ng/ul), iii) RNA

Integrity Number (RIN), iv) RNA Input for Amplification (ng), v) Post-PCR cDNA Concentration (ng/ul), and vi) Final Library Concentration (ng/ul).

2.5.7 Differential Expression Analysis Between Xenograft Spine Metastases and Primary Tumours

The RSEM-Run-EBSeq function was used to conduct an Empirical Bayes Seq (EBSeq) differential expression analysis of genome-wide transcription in Group 3 and Group 4 MB spine metastases relative to primary tumours. EBSeq is a tool for differential expression analysis, which, built upon a foundation of Empirical Bayes (EB) methodologies, generates posterior probabilities for every gene reflecting the likelihood that a gene is differentially expressed [52]. In addition to detecting differentially expressed genes, EBSeq is also capable of conducting differential expression of isoforms for a single gene. This extensive functionality makes working with EBSeq directly quite challenging. To provide a more user-friendly interface for EBSeq, RSEM-Run-EBSeq was created. So, I used this script rather than EBSeq directly to conduct my differential expression analyses. Comparisons were made between primary tumours from each round of selection, and between primary tumours and spine metastases from each round of selection, to serve as controls to identify changes in transcription that were a product of the process of in-vivo iterative selection itself rather than increased propensity for metastasis. In contrast, comparisons were also made between spine metastases from each round of selection, and primary tumours from the first round of selection to identify genes that were becoming activated specifically in correlation with increasing metastasis (Table 3). Included with EBSeq results were posterior probabilities to aid in the assessment of the confidence with which to consider genes identified as being differentially expressed. Specifically, statistics comprised the posterior probability of differential expression (PPDE) and the posterior probability of equal expression (PPEE), as well as a posterior fold-change in transcription (PostFC), and a normalized fold change in transcription (RealFC). From this point forward, P will be used to represent PPEE, and FC to represent RealFC.

Subgroup	Comparison:	pri1 vs. spi1	pri1 vs. spi2	pri1 vs. spi3	pri1 vs. spi5
Group 3	D425	✓	✓	✓	✓
Group 3	MB411FH	✓	✓		
Group 4	Rcmb06	✓			

Table 3. Summary of all comparisons made between RNA-Seq datasets. Note: Comparisons were also made between primary tumours and spine metastases from each round of selection as a control, to verify that changes in transcription observed were not changes attributable to the process of in vivo iterative selection itself independent of propensity for metastasis. pri = primary tumour; spi = spine metastasis; spi1 = round 1 of selection; spi2 = round 2 of selection; spi3 = round 3 of selection; spi5 = round 5 of selection. Note: There were only enough cells in round 4 of selection to inject mice but not to perform RNA-Seq. It is for that reason that a comparison with round 4 spine metastases was not possible.

34

2.5.8 Generation of Summary Tables of Genes Transcriptionally Activated in Spine Metastases Relative to Primary Tumours with Genes Sorted by Significance of Activation

I generated summary tables of differentially expressed genes using proprietary scripts that I wrote in BASH, and R (v.3.1.1). Genes were ordered by score (-log10[P] x sign of log2[FC]). Top gene candidates had P=0, which made it not possible to calculate a value for -log10[P] (infinite solution). To overcome this, for all genes with P=0, I made their score equivalent to the value of the top score across the rest of the dataset plus 1.

2.6 Differential Expression and Pathway Analysis of Microarray Data from Human Group 3 MB patients and Comparisons with Genes and Pathways Activated in Mouse PDX Models

2.6.1 Bioinformatics Analysis Overview

I used a multi-step bioinformatics pipeline to identify genes and pathways that were most significantly activated in metastases to the spine relative to primary tumours (Figure 5, Table 2).

2.6.2 Source of Human MB Expression Data: The MAGIC tumour bank

The microarray expression data used for differential expression analyses was derived from MB primary tumor and spine metastasis tissues managed by the Medulloblastoma Advanced Genomics International Consortium (MAGIC) tumor bank, which contains more than 2000 frozen MB tumors from 90 neuro-oncology institutions around the world [1]. Microarray expression analysis had previously been performed for a subset of MAGIC primary tumours and spine metastases. So, I selected 4 Group 3 primary tumors and 4 spine metastases with available microarray expression data to use in differential expression analyses.

2.6.3 Quality Control Analysis Using ArrayQualityMetrics

Microarray expression data from the 4 primary tumour samples and 4 spine metastasis samples selected were subjected to quality control analyses using the R Package ArrayQualityMetrics (Table 2) [57].

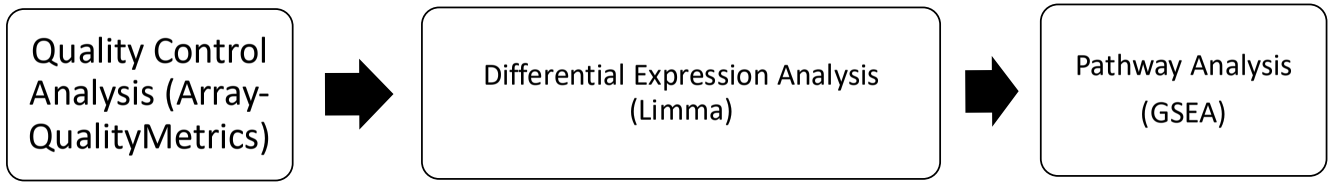


Figure 5. Bioinformatics analysis used to identify genes and pathways activated in human spine metastases relative to primary tumours from the Medulloblastoma Advanced Genomics International Consortium (MAGIC) tissue bank.

36

2.6.4 Comparison of Genome-wide Transcription Between Primary Tumours and Spine Metastases from Human Group 3 MB Patients

Microarray expression data from human Group 3 spine metastases and primary tumours were compared by differential expression analysis using Limma run within R (v3.3.2) (Table 2) [58]. Differentially expressed transcripts were ranked by -log10[P] X the sign of log2[FC], with P representing the probability of equal expression, and FC being the fold change in transcription. Comparisons between lists of genes transcriptionally activated in spine metastases from mouse xenograft models and lists of genes transcriptionally activated in spine metastases from human Group 3 MB patients were carried out using custom BASH scripts.

2.7 Gene Set Enrichment Analysis (GSEA) of Ranked Lists of Genes Transcriptionally Activated in Spine Metastases from PDX Mice, and Human Group 3 MB Patients Relative to Primary Tumours

I used lists of differentially expressed genes, ordered by -log10[P] X the sign of log2[FC], as input for GSEA Preranked. I considered gene sets to be significantly enriched if the False Discovery Rate (FDR) was 0.15 or less. I then wrote proprietary scripts in BASH and R (v.3.1.1/v.3.3.2) to generate summary tables of enriched pathways sorted by increasing FDR (decreasing significance). I then extracted lists of pathway-specific genes from GSEA output files and wrote additional scripts in BASH and R to add these lists to existing summary tables.

2.8 Bench-based Validation of Genes Transcriptionally Activated in Mouse PDX Spine Metastases Relative to Primary Tumours

2.8.1 Immunohistochemical Staining to Determine Whether Protein Expression in Mouse PDX Spine Metastases is Reflective of the Activation Observed at the Transcript Level

For four of the most significantly transcriptionally activated genes, Immunohistochemistry was used to determine whether protein levels were increased in a manner reflective of the increased transcription that was observed in spine metastases relative to primary tumours. Brain and spinal cord tissue sections from intracranially-injected PDX mice were stained with Hematoxylin and Eosin (H & E staining) to emphasize the location of tumour cells within the brain and spinal cord

(Figure 6a). Adjacent sections were then stained with polyclonal antibodies against PAGE1 (Atlas Antibodies - HPA003473), MAGEC2 (Atlas Antibodies - HPA062230), MAGEA10 (Merck - SAB1300318), and CT45A1 (Atlas Antibodies - HPA046987) (Figure 6b). As a positive control, all of these antibodies were used to stain testis tissues.

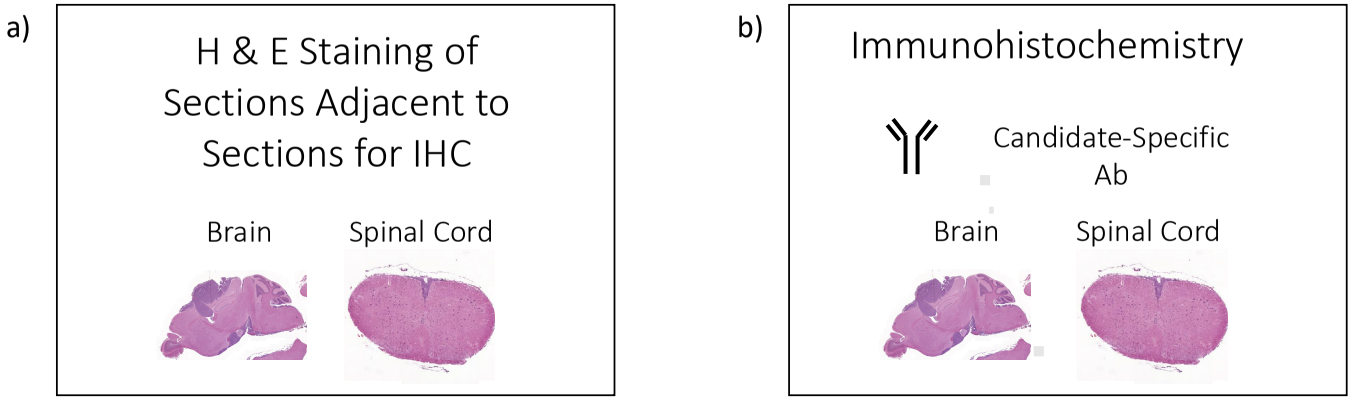


Figure 6. Immunohistochemical (IHC) staining of tissue sections of PDX mouse primary tumours and spine metastases to determine if protein expression in spine metastases resembles transcription. a) Tissue sections adjacent to those to be used in IHC are H & E stained, to define tumour location b) Primary tumour and spine metastasis tissue sections are stained with antibodies. Increased protein levels in spine metastases would validate the differences observed at the transcript level.

39

Chapter 3: Results

Chapter 3: Results

3.1 Overview

For my doctoral studies, I sought to draw upon the significant biological resources available at the Hospital for Sick Children, and the power of bioinformatics analysis to deepen the understanding of the specific molecular mechanisms, which drive the metastasis of Medulloblastoma (MB) from the primary tumour bed to the spine. Working in collaboration with wet bench researchers, I made use of differential expression analyses and pathway analyses to identify and characterize genes and pathways that had become more active in metastases to the spine relative to primary tumours. Such genes and pathways, activated in correlation with the activation of metastases, would represent potential new leads in the ongoing effort to map the molecular means by which MB tumours gain the ability to separate from the primary tumour, leave the primary tumour bed, and migrate, via the cerebrospinal fluid (CSF) or the bloodstream, to target sites throughout the body. This is valuable because metastatic MB is strongly correlated with poor outcomes among MB patients and genes and pathways activated in correlation with metastasis could represent new therapeutic targets that if studied further, could serve as origins of new patient-tailored therapies that could be both safer and more effective than what currently exists. It is exciting to report that this work implicated both Cancer Testis Antigen (CTA) genes, and long non-coding RNAs (lncRNAs) as being RNAs for which increased transcription is correlated with metastasis, or possibly even RNAs that play a role in mediating MB metastasis. Only through further research will it become possible to elucidate the specific roles played by these RNAs in the metastasis of Group 3 MB tumours from the cerebellum to the spine.

3.2 Differential Expression Analysis of RNA-Seq Data from Patient-Derived Xenograft (PDX) Models of Human Group 3 and Group 4 MB Tumours

3.2.1 Introduction

To identify novel genes and pathways activated in correlation with MB metastasis to the spine, bioinformatics approaches were used to explore RNA-Seq data from the spine metastases and

primary tumours of mouse PDX models of human MB that had been subjected to iterative selection for an increased propensity for metastasis to the spine.

3.3 Exploratory Analysis of Xenograft-derived RNA-Seq Data

3.3.1 Quality Control (QC) Analyses Confirm that All RNA-Seq Data is of Sufficient Quality to be Included in Differential Expression Analyses

Prior to conducting differential expression analyses, it was necessary to verify that all RNA-Seq data did not show signs of bias, contamination, or inaccuracy, factors that would each lead to error in RNA-Seq analyses. I used FastQC [53] (Table 2), an established tool for the assessment of RNA-Seq data quality, for this purpose.

FastQC used a specific set of metrics to assess overall dataset quality, namely sequence quality (per base, per sequence, per complete file), and sequence content (%GC, frequency/length of duplications, presence of adaptor sequences) (Appendix QC). Across all datasets, sequence quality at all resolutions was very high, with no sequences needing removal due to poor quality. In addition, GC content and sequence length across all sequences followed a normal distribution, duplications were very rare, and when they would occur, would be only 1 or 2 bases in length. Also, adapter sequences were completely absent. So overall, sequences were of very high quality and thus could all be included in subsequent analyses.

3.3.2 Pairwise Pearson Correlation Analyses Confirm Transcriptional Similarity Correlates with Tissue Similarity Among RNA-Seq Datasets

With data quality confirmed, I evaluated whether the expression data derived from biological replicates of primary tumours were more strongly correlated with one another or with the expression data derived from biological replicates of spine metastases. Pearson correlation coefficients (PCCs) were calculated for all possible pairwise comparisons between RNA-Seq data from the primary tumours and spine metastases of PDX mice derived from D425, MB411FH, and Rcmb06 cell lines. PCCs only are included in the main text while detailed results, including both PCCs and corresponding P-values, are included in Appendix 1. The P-values, indicative of the significance of the PCCs calculated, were 0.00 for all comparisons. This indicated that all PCCs were statistically significant, a finding that is not surprising given the fact that expression of 57391 transcripts was available for comparison between datasets. These

claims, that the 0.00 P-values were indicative of significance, and were so low, at least in part, due to the large sizes of the datasets, are supported by the fact that PCCs calculated for a subsampling of datasets were associated with significant non-zero P-values. A good example of this is the P-value that was reached for a comparison of the expression of the first 900 transcripts, ordered by increasing EnsemblID, from the D425pri1 and D425spi1 datasets (pri=primary tumour; spi=spine metastasis; pri/spi1= first round of selection) (PCC = 0.89; $P = 3.47x10^{-302}$, data not shown). For control comparisons between identical datasets, in which PCCs were, by necessity, 1.00, P-values could not be calculated.

Comparing all of the RNA-Seq datasets, several observations could be drawn (Appendix 1a). Across data from different cell lines, Group 3 MB cell lines D425 and MB411FH were transcriptionally more similar to one another than to the Group 4 cell line Rcmb06. However, correlation between tissue-specific replicates within cell lines was stronger than correlation between replicates of the same tissues from different cell lines, even if both lines were models of the same MB subgroup. So, taken together, this data demonstrated that although the Group 3 cell lines are more similar to one another than to Group 4, they are still sufficiently different to merit separation by cell line for all further analyses rather than combining into a single expression set divided only by tissue and round of selection.

Building upon these findings further Pearson correlation analyses were then performed separately for the datasets from each cell line. In D425 RNA-Seq data strong correlation was observed between biological replicates from corresponding rounds of selection for both primary tumours and spine metastases (Table 4). The only exception to this was the fourth primary tumour replicate, for which expression was only weakly correlated with that of the other primary tumour datasets, and those from spine metastases. This primary tumour dataset was, as a result, omitted from future analyses. In addition, in the first round, the difference in the magnitude of expression between primary tumours and spine metastases was small; however, this difference grew with increasing rounds of selection.

Correlation between expression datasets from corresponding tissues, whether primary tumour or spine metastasis, was much more variable in MB411FH-derived xenograft mice relative to those derived from D425 (Table 5). In the majority of cases, correlation between same tissues was greater than correlation between primary tumour and spine datasets. However, there were also

several occasions in which correlation was strongest between a primary tumour dataset, and a spine metastasis dataset rather than other primary tumours. This was observed both before selection, and across the rounds of selection. Taken together, correlation in expression between tissue-specific replicates was present in both D425 and MB411FH-derived xenograft models of human Group 3 MB but was stronger for D425 than for MB411FH.

With respect to the expression datasets from primary tumours and spine metastases of the Rcmb06-derived xenograft mouse (Group 4 MB), tissue-specific biological replicates were strongly correlated with one another, as was observed for the D425 and MB411FH – derived xenograft models of Group 3 MB (Table 6). So, taken together, this data served to validate that the transcriptomes of spine metastases were significantly different, albeit to differing extents, from those of primary tumours in each of the MB models studied.

Dataset	D425_pri_1_R1	D425_spi_1_R1	D425_pri_1_R2	D425_spi_1_R2	D425_pri_1_R3	D425_spi_1_R3	D425_pri_1_R4	D425_spi_1_R4	D425_spi_2_R1	D425_spi_3_R1	D425_spi_5_R1	D425_spi_5_R2
D425_pri_1_R1	1.00	0.90	0.96	0.92	0.97	0.88	0.14	0.95	0.83	0.88	0.57	0.90
D425_spi_1_R1	0.90	1.00	0.86	0.80	0.88	0.91	0.09	0.88	0.92	0.92	0.77	0.93
D425_pri_1_R2	0.96	0.86	1.00	0.95	0.97	0.85	0.11	0.99	0.81	0.84	0.59	0.90
D425_spi_1_R2	0.92	0.80	0.95	1.00	0.95	0.86	0.11	0.96	0.79	0.82	0.60	0.91
D425_pri_1_R3	0.97	0.88	0.97	0.95	1.00	0.89	0.12	0.97	0.83	0.89	0.60	0.92
D425_spi_1_R3	0.88	0.91	0.85	0.86	0.89	1.00	0.15	0.88	0.96	0.97	0.82	0.97
D425_pri_1_R4	0.14	0.09	0.11	0.11	0.12	0.15	1.00	0.11	0.13	0.14	0.09	0.11
D425_spi_1_R4	0.95	0.88	0.99	0.96	0.97	0.88	0.11	1.00	0.85	0.87	0.63	0.92
D425_spi_2_R1	0.83	0.92	0.81	0.79	0.83	0.96	0.13	0.85	1.00	0.97	0.90	0.96
D425_spi_3_R1	0.88	0.92	0.84	0.82	0.89	0.97	0.14	0.87	0.97	1.00	0.80	0.95
D425_spi_5_R1	0.57	0.77	0.59	0.60	0.60	0.82	0.09	0.63	0.90	0.80	1.00	0.84
D425_spi_5_R2	0.90	0.93	0.90	0.91	0.92	0.97	0.11	0.92	0.96	0.95	0.84	1.00

Pearson Correlation Coefficient	Colour
0.00-0.50	
0.50-0.75	
0.75-0.85	
0.85-0.99	
1.00	

Table 4. Pairwise correlation coefficients for comparisons between D425 RNA-Seq datasets. pri = primary tumour; spi = spine metastasis; spi 1 = round 1 of selection; spi 2 = round 2 of selection; spi 3 = round 3 of selection; spi 5 = round 5 of selection; spi 1 R1 = replicate 1; spi 1 R2 = replicate 2; spi 1 R3 = replicate 3; spi 1 R4 = replicate 4; Red = 0.00-0.50; Orange = 0.50-0.75; Yellow = 0.75-0.85; Green = 0.85-0.99; White = 1.00.

45

Dataset	MB411FH_pri_1_R1	MB411FH_spi_1_R1	MB411FH_pri_1_R2	MB411FH_spi_1_R2	MB411FH_pri_1_R3	MB411FH_spi_1_R3	MB411FH_spi_2_R1
MB411FH_pri_1_R1	1.00	0.82	0.90	0.85	0.90	0.83	0.88
MB411FH_spi_1_R1	0.82	1.00	0.92	0.86	0.71	0.63	0.94
MB411FH_pri_1_R2	0.90	0.92	1.00	0.95	0.83	0.77	0.96
MB411FH_spi_1_R2	0.85		0.95	1.00	0.82	0.86	0.95
MB411FH_pri_1_R3	0.90	0.71	0.83	0.82	1.00	0.89	0.81
MB411FH_spi_1_R3	0.83	0.63	0.77	0.86	0.89	1.00	0.79
MB411FH_spi_2_R1	0.88	0.94	0.96	0.95	0.81	0.79	1.00

Pearson Correlation Coefficient	Colour
0.00-0.50	
0.50-0.75	
0.75-0.85	
0.85-0.99	
1.00	

Table 5. Pairwise correlation coefficients for comparisons between MB411FH RNA-Seq datasets. pri = primary tumour; spi = spine metastasis; spi 1 = round 1 of selection; spi 2 = round 2 of selection; spi 3 = round 3 of selection; spi 5 = round 5 of selection; spi 1 R1 = replicate 1; spi 1 R2 = replicate 2; spi 1 R3 = replicate 3; spi 1 R4 = replicate 4; Red = 0.00-0.50; Orange = 0.50-0.75; Yellow = 0.75-0.85; Green = 0.85-0.99; White = 1.00.

46

Dataset	Rcmb06_pri_1_R1	Rcmb06_spi_1_R1	Rcmb06_pri_1_R2	Rcmb06_spi_1_R2
Rcmb06_pri_1_R1	1.00	0.85	0.84	0.70
Rcmb06_spi_1_R1	0.85	1.00	0.93	0.87
Rcmb06_pri_1_R2	0.84	0.93	1.00	0.86
Rcmb06_spi_1_R2	0.70	0.87	0.86	1.00

Pearson Correlation Coefficient	Colour
0.00-0.50	
0.50-0.75	
0.75-0.85	
0.85-0.99	
1.00	

Table 6. Pairwise correlation coefficients for comparisons between Rcmb06 RNA-Seq datasets. pri = primary tumour; spi = spine metastasis; spi 1 = round 1 of selection; spi 2 = round 2 of selection; spi 3 = round 3 of selection; spi 5 = round 5 of selection; spi 1 R1 = replicate 1; spi 1 R2 = replicate 2; spi 1 R3 = replicate 3; spi 1 R4 = replicate 4; Red = 0.00-0.50; Orange = 0.50-0.75; Yellow = 0.75-0.85; Green = 0.85-0.99; White = 1.00.

47

3.3.3 Principal Components of Variation are In Part Attributable to Variation in Scale Variables Associated with Specific Steps of Bulk RNA-Seq Protocol

To qualitatively and quantitatively explore the variation present in the RNA-Seq data, I performed principal component analyses (PCA) for a combined dataset including RNA-Seq data from all datasets, and for subsets of data specifically from D425, MB411FH, and Rcmb06 (referred to as VJDup on PCA plots). In the combined dataset, the top two components of variation (PC1 and PC2) accounted for 81% of the total variation within the dataset (Appendix 2). Similarly, PC1 and PC2 together accounted for 79%, 91%, and 87% of the total variation in D425, MB411FH, and Rcmb06 datasets respectively (Appendix 2). We then sought to find specific characteristics of the expression datasets to which this variation could be attributed. Points on the PCA plot were assigned different colours to reflect different discrete possibilities for those characteristics of the datasets, predominantly linked to specific stages of the bulk RNA-Seq procedure. Clustering of a single colour would indicate association of that characteristic with the variation within the RNA-Seq data. No clear clustering was observed qualitatively for any of the characteristics considered, including tissue (Appendix 2). To investigate quantitatively whether associations exist, Pearson correlation analysis was performed between the x coordinates for PC1 (pc1_x) from the PCA plots for D425, MB411FH, and Rcmb06, and the specific values for the 6 scale variables among those considered qualitatively in PCA plots. These were specifically 1) the number of cells obtained from xenograft tissues through sorting (Number_of_Sorted_Cells), 2) the concentration of RNA isolated from sorted cells (Sorted_Cell_RNA_Concentration_ngperul), 3) the RNA Integrity Number (RIN) (RNA_Integrity_Number_RIN; a measure of RNA quality), 4) the mass of RNA used as input for PCR amplification (RNA_Input_for_Amplification_ng), 5) the cDNA concentration after PCR amplification (Post_PCR_cDNA_Concentration_ngperul), and 6) the final cDNA library concentration (Final_Library_Concentration_ngperul). For D425 the principal component of variation represented by pc1_x was moderately correlated with the number of sorted cells (Correlation Coefficient (CC)=-0.57), and strongly correlated with the RNA input for amplification (CC=-0.75) (Table 7). Similarly, for Rcmb06 pc1_x was moderately corelated with the RIN (CC=-0.51), and the post-PCR cDNA concentration (CC=-0.62) (Table 8).

Scale Variable	PC1_x	Number_of_Sorted_Cells	Sorted_Cell_RNA_Concentration_ngperul	RNA_Integrity_Number_RIN	RNA_Input_for_Amplification_ng	Post_PCR_cDNA_Concentration_ngperul	Final_Library_Concentration_ngperul
PC1_x	1.00	-0.57	-0.18	-0.01	-0.75	-0.26	0.12
Number_of_Sorted_Cells	-0.57	1.00	0.45	0.54	0.84	0.70	-0.13
Sorted_Cell_RNA_Concentration_ngperul	-0.18	0.45	1.00	0.09	0.50	0.12	-0.08
RNA_Integrity_Number_RIN	-0.01	0.54	0.09	1.00	0.35	0.39	0.43
RNA_Input_for_Amplification_ng	-0.75	0.84	0.50	0.35	1.00	0.46	-0.14
Post_PCR_cDNA_Concentration_ngperul	-0.26	0.70	0.12	0.39	0.46	1.00	-0.12
Final_Library_Concentration_ngperul	0.12	-0.13	-0.08	0.43	-0.14	-0.12	1.00

Pearson Correlation Coefficients	Colour
-1.00	
-0.99 - -0.85	
-0.85 - -0.75	
-0.75 - -0.50	
-0.50 - 0.50	
0.50-0.75	
0.75-0.85	
0.85-0.99	
1.00	

Table 7. Pairwise Pearson correlations between the coordinates of the first principal component of variation along the x-axis of the PCA plot (pc1_x) for D425 and the values of specific scale variables across the axis. White=-1.00 or 1.00; Green=-0.99 to -0.85 or 0.85 to 0.99; Yellow= -0.85 to -0.75 or 0.75 to 0.85; Orange=-0.75 to -0.50 or 0.50 to 0.75; Red =-0.50 to 0.50.

49

Scale Variable	PC1_x	Number_of_Sorted_Cells	Sorted_Cell_RNA_Concentration_ngperul	RNA_Integrity_Number_RIN	RNA_Input_for_Amplification_ng	Post_PCR_cDNA_Concentration_ngperul	Final_Library_Concentration_ngperul
PC1_x	1.00	-0.41	-0.37	-0.51	-0.42	0.62	-0.34
Number_of_Sorted_Cells	-0.41	1.00	0.97	0.22	0.66	-0.81	-0.42
Sorted_Cell_RNA_Concentration_ngperul	-0.37	0.97	1.00	0.00	0.81	-0.89	-0.26
RNA_Integrity_Number_RIN	-0.51	0.22	0.00	1.00	-0.41	0.10	-0.44
RNA_Input_for_Amplification_ng	-0.42	0.66	0.81	-0.41	1.00	-0.95	0.34
Post_PCR_cDNA_Concentration_ngperul	0.62	-0.81	-0.89	0.10	-0.95	1.00	-0.19
Final_Library_Concentration_ngperul	-0.34	-0.42	-0.26	-0.44	0.34	-0.19	1.00

Pearson Correlation Coefficients	Colour
-1.00	
-0.99 - -0.85	
-0.85 - -0.75	
-0.75 - -0.50	
-0.50 - 0.50	
0.50-0.75	
0.75-0.85	
0.85-0.99	
1.00	

Table 8. Pairwise Pearson correlations between the coordinates of the first principal component of variation along the x-axis of the PCA plot (pc1_x) for Rcmb06 and the values of specific scale variables across the axis. White=-1.00 or 1.00; Green=-0.99 to -0.85 or 0.85 to 0.99; Yellow= -0.85 to -0.75 or 0.75 to 0.85; Orange=-0.75 to -0.50 or 0.50 to 0.75; Red =-0.50 to 0.50.

50

In contrast, pc1_x from MB411FH was not correlated with any of the scale variables considered (Table 9). Taken together, this indicated that transcriptional differences identified between primary tumours and spine metastases for D425 might be due to genuine differences but might also be attributable to other confounding factors. So, for any genes found to be transcriptionally activated, validation would be needed, to verify that transcriptional activation was in correlation with metastasis and not other variables such as the scale variables described here.

Scale Variables	PC1_x	Number_of_Sorted_Cells	Sorted_Cell_RNA_Concentration_ngperul	RNA_Integrity_Number_RIN	RNA_Input_for_Amplification_ng	Post_PCR_cDNA_Concentration_ngperul	Final_Library_Concentration_ngperul
PC1_x	1.00	-0.10	-0.10	0.06	-0.04	0.24	-0.45
Number_of_Sorted_Cells	-0.10	1.00	0.49	0.31	0.57	0.18	-0.18
Sorted_Cell_RNA_Concentration_ngperul	-0.10	0.49	1.00	0.04	0.85	0.46	0.09
RNA_Integrity_Number_RIN	0.06	0.31	0.04	1.00	0.28	-0.12	-0.59
RNA_Input_for_Amplification_ng	-0.04	0.57	0.85	0.28	1.00	0.14	-0.04
Post_PCR_cDNA_Concentration_ngperul	0.24	0.18	0.46	-0.12	0.14	1.00	0.14
Final_Library_Concentration_ngperul	-0.45	-0.18	0.09	-0.59	-0.04	0.14	1.00

Pearson Correlation Coefficients	Colour
-1.00	
-0.99 - -0.85	
-0.85 - -0.75	
-0.75 - -0.50	
-0.50 - 0.50	
0.50-0.75	
0.75-0.85	
0.85-0.99	
1.00	

Table 9. Pairwise Pearson correlations between the coordinates of the first principal component of variation along the x-axis of the PCA plot (pc1_x) for MB411FH and the values of specific scale variables across the axis. White=-1.00 or 1.00; Green=-0.99 to -0.85 or 0.85 to 0.99; Yellow= -0.85 to -0.75 or 0.75 to 0.85; Orange=-0.75 to -0.50 or 0.50 to 0.75; Red =-0.50 to 0.50.

52

In summary, exploratory analysis of the expression datasets generated from bulk RNA-Seq confirmed that data was of high quality. Furthermore, replicates of expression data from primary tumours and spine metastases exhibited patterns of correlation specific to tissue type. However, PCA demonstrated that the principal component of variation for D425 was correlated with the quantities of cells and RNA used as input for bulk RNA-Seq, and the principal component of variation for Rcmb6 was correlated with the RIN. This reinforced the fact that bench-based validation would be needed for any observations made using bioinformatics.

3.4 Differential Expression Analysis (DEA) Identifies Genes Transcriptionally Activated in Xenograft Spine Metastases Relative to Primary Tumours Prior to Selection

In mouse PDX models of human Group 3 and Group 4 MB, several genes were transcriptionally activated in spine metastases relative to primary tumours. In most cases activated genes were unique to either D425, MB411FH, or Rcmb06 mouse models (Appendix 3); however, a small number of activated genes were shared between datasets (Figure 7). Of greatest interest would be any transcriptionally activated genes shared between xenograft models. Two genes, specifically the *PLP1* gene coding for proteolipid protein 1, and the *RPL21P75* gene coding for a pseudogene of ribosomal protein 21, were activated in all three types of xenograft mouse models (Table 10). In addition, two activated RNA species, specifically the long non-coding RNA (lncRNA) *RP11-282O18.6,* and the enolase pseudogene *ENO1P1*, were shared among the xenograft mouse models of specifically human Group 3 MB (Table 11). Although not statistically testable due to the low number of genes shared in this manner, qualitatively an apparent consistency in the magnitudes of P and FC was especially striking for *RP11-282O18.6, RPL21P75,* and *ENO1P1* (Table 10, Table 11). So, taken together some genes were found to be transcriptionally activated in a recurrent fashion, and further even, in the case of *PLP1*, in a non-subgroup-specific fashion. However, this overlap in transcriptionally activated genes between groups was very small, and might have been due purely to chance, resulting from variability introduced into the dataset by sources other than tissue differences. Further, while 2 activated pathways were shared between D425 and MB411FH in the first round of selection, the majority of activated pathways were unique to each cell line (Appendix 8). So further clarification will be needed to determine whether this finding is genuine or artifactual.

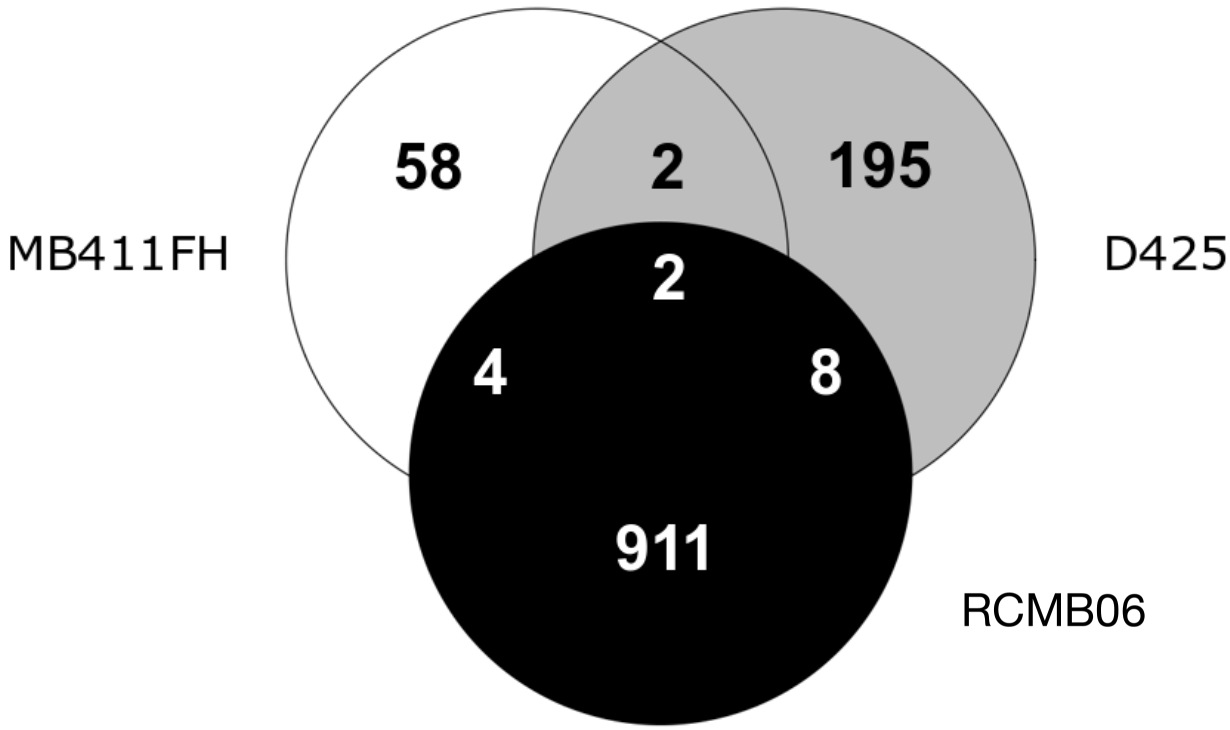


Figure 7. Genes are transcriptionally activated (-log10[P] X sign of log2[FC] > 1.30103) in spine metastases of mouse PDX models of Group 3 (D425, MB411FH), and Group 4 (RCMB06) MB. Of transcriptionally activated genes a select few are shared, but most are unique to D425, MB411FH, or RCMB06.

54

Cell Line	-log10[P]			log2[FC]			Mean Transcription (FPKM)					
Statistic	D425	MB411FH	Rcmb06	D425	MB411FH	Rcmb06	D425		MB411FH		Rcmb06	
Comparison	pri1spi1	pri1spi1	pri1spi1	pri1spi1	pri1spi1	pri1spi1	pri1	spi1	pri1	spi1	pri1	spi1
Candidate Gene												
PLP1	6.36	2.06	15.65	7.52	4.54	5.26	0.86	159.14	8.95	208.38	8.29	317.34
RPL21P75	2.69	4.35	1.63	0.77	0.84	2.14	322.84	548.63	366.57	654.59	27.46	120.98

-log10[P]>1.30103 -log2[FC] > 1 FPKM > 5

Table 10. Two transcriptionally activated genes are shared between D425, MB411FH, and RCMB06. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 ($FC > 2$); black shading = Transcription > 5 FPKM.

55

Cell Line	-log10[P]		log2[FC]		Mean Transcription (FPKM)			
Statistic	D425	MB411FH	D425	MB411FH	D425		MB411FH	
Comparison	pri1spi1	pri1spi1	pri1spi1	pri1spi1	pri1	spi1	pri1	spi1
Candidate Gene								
RP11-282O18.6	2.26	4.70	5.30	5.19	0.86	34.11	0.98	35.90
ENO1P1	1.71	1.61	4.87	5.14	0.86	25.47	0.98	34.68

-log10[P]>1.30103 | -log2[FC] > 1 | FPKM > 5

Table 11. Two transcriptionally activated genes are shared between D425 and MB411FH. pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

56

Shifting focus to genes activated in only one of the mouse models, 195 genes were transcriptionally activated (P<0.05) in D425 spine metastases relative to primary tumours (Appendix 3a). The five genes most significantly activated were the gene encoding the inhibitor of DNA-binding *ID1*, and the Nudix Hydrolase gene *NUDT11,* a specific variant of which is weakly associated with the appearance of tumour-like characteristics in prostate cancer cells *[59, 60]*. In addition, of particular relevance and interest to this study, the protein phosphatase *PPP1CC* was also transcriptionally activated. *PPP1CC* transcription has previously been shown to be reduced by more than half in response to the siRNA-mediated knockdown of *LASP1*, a mediator of MB metastatic dissemination [61] . Also, increased *PPP1CC* transcription has been observed in proliferating astrocytes and microglia following an induced traumatic spinal cord injury in rats [62]. Furthermore, also included was the gene *SEPHS*, which codes for selenophosphate synthetase 1. While *SEPHS* has not yet been directly linked to metastasis, *SEPHS* overexpression further enhances the increased radiosensitivity that is associated with increased levels of p53 [63]. In addition, transcriptional activation of the gene coding for signal transduction protein STAT3 was observed. Increased *STAT3* transcription in MB stem cells (MBSCs) drives tumorigenesis via the regulation of *c-Myc [64]*. Furthermore, as previously discussed *STAT3* expression is necessary and sufficient to render Group 3 MB cell line Med8A resistant to multiple chemotherapeutic agents used alone. However, the cell line remains sensitive to vincristine used in combination with cisplatin or STAT3 inhibitor niclosamide. Furthermore, the chemoresistance might be sexually dimorphic, since the chemoresistance is dependent on the co-expression of male-specific signaling factor IL6 [11]. In addition, the gene encoding complement 3 (*C3)*, previously linked to metastasis by Boire and colleagues [65], was transcriptionally activated. As well, *PRSS54*, a serine protease and member of the Cancer Testis Antigen (CTA) gene family, exhibited more than 16X increased transcription (log2FC=4.52) in spine metastases; CTA genes are frequently associated with metastatic cancer upon mutation or overexpression, most frequently as biomarkers. However, increased transcription of *PRSS54* has been shown to not be necessary for male fertility in mice so might similarly have a lesser role to play in mediating metastasis [66].

Comparatively, only 58 genes were transcriptionally activated in spine metastases from MB411FH xenografts (Appendix 3b). The 5 most significantly activated genes were *DNAJC19*, a mitochondrial translocator correlated with cerebellar atrophy in the context of paediatric

cardiomyopathy [67, 68], an X-linked gene, *BEX2*, which interestingly has been shown to promote the metastasis of colorectal cancer through aberrant Hedgehog signaling when silenced [69], *ACYP2*, a muscle and biomarker associated with hearing loss in pediatric cancer patients treated with Cisplatin, an uncharacterized noncoding RNA *RP11-82L2.1*, and the protein-coding RNA *CMC1*. It is also worth noting that *SSX2*, a CTA gene and known regulator of focal adhesion [70], was also transcriptionally activated in the MB411FH spine metastases, although again less significantly than the top 5 genes. In contrast, for Rcmb06 (Group 4 MB), 915 genes were transcriptionally activated in spine metastases relative to primary tumours with the 5 most significantly activated genes being the cannabinoid receptor interacting protein 1 gene *CNRIP1*, the peripheral clock protein gene *NCKAP5*, an uncharacterized non-coding RNA *RP1-228P16.1*, deubiquitinating enzyme with Josephin domain *JOSD1*, and a regulator of G-protein signaling *RGS17* (Appendix 3c).

Taken together, genes transcriptionally activated in the spine metastases from all xenograft models relative to primary tumours, largely did not show strong connections to known mechanisms of metastasis, with a few exceptions. *PPP1CC*, for example, is a candidate that would be worth exploring further given its earlier identified link to MB metastatic dissemination. Comparatively, *BEX2*, while correlated with metastasis when silenced, would not be a target easily pursued; is more difficult to definitely show that the removal of a gene product drives tumorigenesis or metastasis relative to the introduction of increased levels of a gene product. Furthermore, in practice, there are many fewer confounding variables associated with linking gene activation to metastasis than gene silencing. In addition, *ACYP2* could represent a valuable new biomarker of risk for side-effects associated with Cisplatin chemotherapy for paediatric MB patients. The identification of these genes does serve as proof that in some cases increased expression of metastasis-associated genes is being selected for. However, at this point it is the exception not the rule. Rather it appears that other influences are more generally driving increases in gene expression.

3.5 Iteratively Selecting for Increased Metastasis, Genes Exhibit Several Patterns of Increased Transcription in Spine Metastases Relative to Primary Tumours

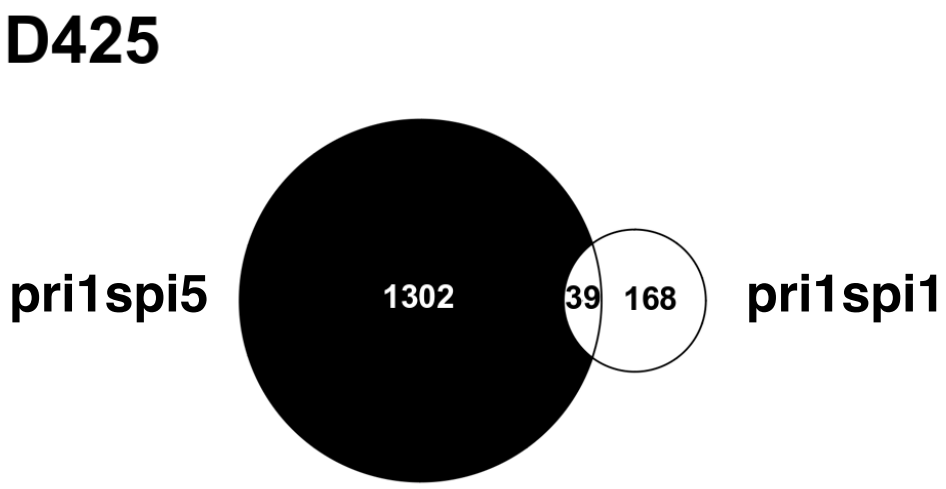
To enrich for genes with transcription increasing in correlation with metastasis, spine metastases were isolated from xenograft mice, and intracranially injected into the cerebella of new mice,

serving as the seeds for a next generation of xenograft models. Five rounds of in vivo iterative selection of this sort were performed, in order to produce spine metastases with increased propensity for metastasis; at the genetic level these super-metastases would also, by necessity, exhibit increased transcription of genes necessary and/or sufficient for metastasis, and unchanged or decreased expression of genes irrelevant to metastasis relative to the primary tumour.

3.5.1 Genes Exhibit a Wide Range of Modes of Increased Transcription Across 5 Rounds of Selection

Across the 5 rounds of selection carried out for the D425 PDX mouse model, multiple patterns of change in transcription were observed (Appendix 4). Some genes exhibited increased transcription in single rounds of selection, but ultimately did not remain activated once all rounds of selection were complete (Appendix 4). These genes were likely among those for which their increased transcription was irrelevant to metastasis or genes for which promotion of metastasis was less than that of other genes and thus rendered irrelevant. In contrast, of significant interest were 39 genes transcriptionally activated in spine metastases in both the first and fifth rounds of selection (Figure 8). Included among these genes were both *PLP1*, earlier found to be shared among all three PDX models of human MB, and *PPP1CC*, previously linked to MB metastasis. The latter especially served as valuable evidence that iterative selection was selecting for the increased transcription of a gene linked to metastasis. Other genes were transcriptionally activated in the first round of selection and were then maintained at the same level across the remaining rounds of selection. (Appendix 4). It can be hypothesized that these genes may have actively promoted metastasis in the first round of selection but then became neither beneficial nor detrimental to metastasis through the remaining rounds of selection. As well, other genes were transcriptionally activated in only the fifth and final round of selection. This pattern would suggest that other changes in previous rounds set the stage for these genes to become active correlates of metastasis in round 5 (Appendix 4). This pattern was observed for a large proportion of the genes of the Cancer Testis Antigen (CTA) gene family. Taken together, iterative selection for increased metastasis propensity in spine metastases affected transcription of genes in several ways, suggesting that their connections to metastasis may be just as heterogeneous.

a)



b)

39 Genes Activated in Rounds 1 and 5

NUDT11	*RP11-184M15.1*	*DPH7*
PPP1CC	*KRT8*	*SOSTDC1*
GGH	*SEC16A*	*GOLM1*
FOXN4	*RP11-282O18.6*	*CKAP4*
PLP1	*PARD3*	*AC131180.1*
MIR4435-1HG	*U47924.30*	*NDN*
PATZ1	*S1PR3*	*EIF4A3*
NT5C3A	*RB1CC1*	*AC119673.1*
SEPT8	*C1orf198*	*MRPL9*
TXNDC5	*ARHGAP17*	*UPK3BL*
GSPT2	*EPS8*	*WDR72*
PRELP	*PLK2*	*RP11-254F7.3*
ADAM20P1	*MALL*	*USP32P2*

Figure 8. Transcriptionally activated genes shared between rounds 1 and 5 of selection. a) Among the genes transcriptionally activated in round 1 of selection, 39 are also transcriptionally activated in round 5 of selection b) Names of the 39 transcriptionally activated genes shared between rounds 1 and 5 of selection. Underlined genes are of particular interest due to shared identification as a candidate in the first round of selection (*PLP1*), or previous links to MB metastasis (*PPP1CC*).

3.5.2 Transcription of a Subset of Genes Increases At an Increasing Rate Across All Rounds of Selection

Of greatest interest were a subset of genes for which transcription increased at a rate, which progressively increased across all rounds of selection (Table 12, Appendix 5). Genes exhibiting this pattern of increasing transcription included genes with a wide range of functions; however, by far the gene with the greatest magnitude of increase in both P ad FC was *XIST* a long non-coding RNA (lncRNA) most frequently associated with X-chromosome inactivation [71] (Table 13).

Cell Line	D425 (Group 3)												
Statistic	-log10[P]				log2[FC]				Mean Transcription (FPKM)				
Comparison	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1	spi1	SPI2	spi3	spi5
Candidate Gene													
FLOT1	1.29	1.83	2.01	15.95	1.25	0.80	0.82	1.01	182.45	433.74	375.54	378.65	423.65
MAGEF1	0.11	0.23	5.53	16.95	0.89	0.59	1.10	1.46	180.60	334.23	321.17	453.75	573.28
RP1-37N7.5	0.05	1.00	3.26	16.95	1.15	2.97	3.12	4.06	6.18	13.77	56.97	62.99	119.73
FAM107B	0.03	0.04	0.05	16.95	0.58	0.30	-0.85	3.40	21.10	31.62	30.20	13.56	255.16
MAGEA10	0.03	0.04	12.23	16.95	1.54	1.29	2.86	2.36	342.21	993.10	969.56	2883.15	2016.80
MXRA8	0.02	1.80	2.95	16.95	0.43	1.43	1.61	2.50	51.68	69.79	164.11	185.67	333.76
PIAS2	0.02	0.02	0.10	16.95	0.62	0.23	-0.44	0.85	276.26	423.50	382.59	238.87	573.84
PYGL	0.02	0.04	0.34	16.95	0.37	0.26	0.39	1.28	620.87	803.83	870.89	954.44	1740.81
AL358781.1	0.01	0.02	0.02	16.95	0.37	0.10	0.33	1.35	132.14	170.75	167.13	196.10	388.55
EMILIN3	0.01	0.10	0.23	16.95	0.15	1.05	1.28	2.37	60.16	66.98	146.99	172.11	355.48

-log10[P]>1.30103 | -log2[FC] > 1 | FPKM > 5

Table 12. Transcription of a subset of genes increases progressively across some or all rounds of selection. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

62

Cell Line	D425 (Group 3)												
Statistic	-log10[P]				log2[FC]				Mean Transcription (FPKM)				
Comparison	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1	spi1	spi2	spi3	spi5
Candidate Gene													
XIST	2.14	16.65	16.95	16.95	-4.22	4.97	5.70	8.58	26.05	1.39	975.60	1615.78	11652.14

-log10[P]>1.30103 -log2[FC] > 1 FPKM > 5

Table 13. The gene by far exhibiting the greatest magnitudes of increase in both P and FC across rounds of selection is lncRNA *XIST*. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

63

3.5.3 CTA Genes Predominate Among Transcriptionally Activated Genes in General, and are Present Specifically Among Those Genes with Progressively Increasing Transcription Across Rounds of Selection

Included among the 1966 genes with significant transcriptional activation in at least one round of selection (Appendix 4) were 46 CTA genes (Table 14), representing 5% of the total list of 1966 genes. Furthermore, among these CTA genes were 11, which also exhibited progressively increasing transcription across rounds of selection (Table 15). Looking more closely at the transcriptionally activated CTA genes, many transitioned from not being activated in round 1 of selection to being significantly activated in round 5 of selection, evidenced by significant increases in both P and FC (Figure 9). So, CTA genes are predominant among transcriptionally activated genes, and are clearly present among genes exhibiting progressively increasing transcription across rounds of selection.

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5				
Cell Line	**D425 (Group 3)**												
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**				
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**`spi5**
Candidate Gene													
MAGEA10	0.03	0.04	12.23	16.95	1.54	1.29	2.86	2.36	342.21	993.10	969.56	2883.15	2016.80
MAGEC2	0.01	16.65	16.95	16.95	0.73	4.38	2.84	4.63	120.81	201.03	2955.00	1014.94	3440.64
CT45A3	0.01	0.00	15.95	16.95	0.94	1.01	4.79	9.42	2.17	4.18	5.13	70.74	1723.95
XAGE1D	0.00	3.51	0.88	16.95	0.23	0.97	0.76	1.34	1113.52	1304.18	2570.33	2214.22	3232.35
XAGE1B	0.00	1.30	1.79	16.95	0.11	0.73	0.79	1.25	4224.11	4562.24	8226.64	8598.31	11523.18
PAGE4	0.00	0.00	0.00	16.95	-0.50	1.12	0.18	7.94	1.56	1.10	4.03	2.09	446.86
IL13RA2	0.00	0.00	2.77	16.95	-3.49	1.09	2.45	5.37	12.49	1.10	31.21	80.32	596.60
CT45A1	0.00	0.00	0.00	16.95	1.10	3.07	4.24	8.30	0.94	2.04	9.41	21.11	343.92
HORMAD1	0.00	0.00	0.00	16.95	0.31	-0.35	-0.30	6.96	1.09	1.36	1.01	1.04	158.60
GAGE12I	0.01	15.35	5.47	16.95	-0.13	1.85	1.42	2.36	78.34	71.56	334.10	246.51	466.52
GAGE1	0.02	1.47	16.95	16.95	-0.28	2.17	3.86	3.24	23.58	19.47	122.99	396.05	256.00
CT45A4	0.00	0.00	0.00	13.82	0.02	3.00	3.64	8.52	1.09	1.10	10.36	16.06	466.07
XAGE2B	1.24	13.54	3.95	12.01	4.01	4.76	4.32	5.54	2.03	32.90	63.80	46.94	108.17
TEX15	0.03	0.53	0.11	12.28	-0.31	2.05	-2.21	5.47	8.40	6.76	40.27	2.09	428.93
HORMAD2	0.32	0.07	0.14	8.02	-2.86	0.84	1.26	2.14	8.05	1.10	17.12	22.95	40.76
TEX14	1.14	1.58	0.00	8.70	-2.51	2.52	1.67	4.05	10.67	1.87	71.48	39.64	203.04
PTPN20A	0.04	3.97	0.07	5.33	-1.15	3.05	1.29	5.68	4.43	1.99	42.05	12.39	259.41
CSAG2	0.02	0.04	0.02	5.14	0.40	-0.46	-0.40	0.80	131.60	173.74	112.61	117.47	263.28
ATAD2	0.00	0.01	0.02	5.14	0.23	0.41	-0.75	1.30	783.29	917.00	1221.26	544.50	2220.99
MAGEB1	0.04	0.19	0.04	4.46	0.93	1.31	0.14	2.05	9.54	18.18	28.19	12.52	45.75
CT45A5	0.00	0.00	0.00	4.43	0.87	0.73	1.46	6.70	1.16	2.12	2.25	3.72	139.32
GPAT2	0.00	0.01	0.00	4.39	0.69	4.55	3.97	6.57	1.15	1.86	31.74	21.13	126.41
NOL4	0.00	0.00	0.00	4.24	-1.28	1.79	-0.03	3.78	2.69	1.10	11.07	3.13	42.94
TMEFF2	0.55	0.00	0.03	3.44	2.80	2.94	0.37	4.18	4.86	33.85	43.29	7.30	100.92
XAGE5	0.02	0.08	0.14	3.87	-0.20	-1.44	1.78	2.29	11.89	10.35	5.03	46.94	66.46
TPTE	1.60	2.21	0.00	5.29	-0.82	-1.06	0.49	0.75	96.63	54.88	54.37	158.55	186.40
SSX2	0.00	16.65	16.95	3.64	0.35	9.37	6.86	6.22	0.86	1.10	675.68	118.40	74.48
XAGE1C	0.01	0.01	0.01	3.50	0.04	0.44	0.43	1.95	342.71	353.50	548.80	544.38	1522.72
CT45A2	0.00	0.00	0.00	3.21	-0.52	3.92	-0.81	5.40	1.59	1.10	27.95	1.04	77.13
CT45A6	0.00	0.00	0.00	3.12	0.15	0.61	3.07	5.58	1.26	1.40	2.27	12.50	69.73
TDRD1	0.00	0.00	0.00	3.04	0.00	0.00	0.00	5.68	0.00	0.00	0.00	0.00	51.22
DPPA2	0.00	0.00	0.00	2.59	-0.43	2.53	-0.73	4.94	1.49	1.10	10.07	1.04	52.33
TAF7L	0.03	0.07	0.00	2.35	0.40	-1.36	2.38	4.51	2.19	2.89	1.01	13.56	57.34
MAGEA1	0.01	0.05	0.01	2.31	-0.15	0.88	0.53	1.41	433.34	390.57	945.40	741.65	1333.19
NXF2B	0.16	0.16	0.14	1.97	1.87	-3.11	-3.06	4.11	7.69	28.16	1.01	1.04	151.43
DCAF12	0.01	0.01	0.01	1.78	0.48	-0.18	0.07	1.04	723.69	1007.20	752.09	896.03	1715.00
SPAG17	0.03	0.01	0.01	1.74	1.33	0.30	-0.19	1.25	87.52	219.38	126.86	89.71	240.43
RBM46	0.00	0.00	0.00	1.72	0.00	0.00	0.00	4.37	0.00	0.00	1.01	0.00	20.60
NXF2	0.16	0.15	0.06	1.43	1.88	-2.97	-0.61	3.53	6.95	25.54	1.01	5.22	91.49
PRSS54	1.34	0.00	0.00	0.00	4.52	0.00	0.05	0.18	0.86	19.96	1.01	1.04	1.12
CTAG1B	0.04	0.00	16.95	0.00	1.08	1.07	4.61	-0.97	8.93	18.84	22.12	255.97	5.25
CTAG2	0.02	0.29	1.58	0.05	0.84	0.69	0.91	-0.23	172.95	308.97	329.26	382.41	170.69
SPAG4	0.83	0.08	16.95	2.26	-2.64	0.83	3.16	-3.74	13.15	2.10	27.18	136.65	1.12

Table 14: Summary table of Cancer Testis Antigen (CTA) genes transcriptionally activated in D425 spine metastases relative to primary tumours in at least one round of selection. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 (P < 0.05); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5				
Cell Line	**D425 (Group 3)**												
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**				
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**spi5**
Candidate Gene													
MAGEA10	0.03	0.04	12.23	16.95	1.54	1.29	2.86	2.36	342.21	993.10	969.56	2883.15	2016.80
XAGE1B	1.85E-03	1.30	1.79	16.95	0.11	0.73	0.79	1.25	4224.11	4562.24	8226.64	8598.31	11523.18
IL13RA2	1.30E-15	1.87E-14	2.77	16.95	-3.49	1.09	2.45	5.37	12.49	1.10	31.21	80.32	596.60
CT45A1	0.00	2.51E-13	6.71E-08	16.95	1.10	3.07	4.24	8.30	0.94	2.04	9.41	21.11	343.92
CT45A4	0.00	3.24E-13	8.24E-11	13.82	0.02	3.00	3.64	8.52	1.09	1.10	10.36	16.06	466.07
ATAD2	3.99E-03	8.95E-03	0.02	5.14	0.23	0.41	-0.75	1.30	783.29	917.00	1221.26	544.50	2220.99
CT45A5	0.00	4.34E-16	8.68E-16	4.43	0.87	0.73	1.46	6.70	1.16	2.12	2.25	3.72	139.32
XAGE5	0.02	0.08	0.14	3.87	-0.20	-1.44	1.78	2.29	11.89	10.35	5.03	46.94	66.46
CT45A6	0.00	4.34E-16	1.06E-12	3.12	0.15	0.61	3.07	5.58	1.26	1.40	2.27	12.50	69.73
SSX3	0.00	4.34E-16	1.30E-15	0.04	0.63	-0.48	1.56	1.29	1.19	1.85	1.01	4.17	3.36
PRAME	9.00E-04	1.25E-03	1.28E-03	4.62E-03	0.03	-0.12	-0.21	0.19	13523.18	13836.39	14690.41	13827.45	17798.26
ODF2	0.01	0.01	0.01	0.02	-0.26	0.19	-0.21	-0.03	421.02	352.30	565.83	427.67	474.92
MAGEA8	0.00	1.30E-15	1.74E-15	0.05	-2.31	-2.68	1.17	-0.34	5.50	1.10	1.01	14.60	5.03
TSGA10	0.03	0.04	0.07	0.09	0.77	0.57	1.01	-0.35	9.17	15.62	16.11	21.91	8.36
TMEM108	8.87E-03	0.03	0.05	0.08	0.01	0.56	0.75	-0.42	122.80	123.91	213.44	243.04	106.15
RQCD1	0.01	0.02	0.03	0.32	-0.21	-0.26	-0.96	-1.31	169.21	146.45	166.12	102.22	79.20
TEKT5	0.03	0.05	0.30	0.42	1.02	-0.25	-3.77	-1.66	12.18	24.69	12.08	1.04	4.47
ACTL8	5.08E-03	5.54E-03	0.08	2.07	-0.56	-0.10	-2.59	-3.63	776.71	527.36	850.76	150.21	72.63
PLAC1	0.02	0.02	0.24	7.34	-0.38	-0.09	-1.62	-5.07	33.02	25.45	36.25	12.52	1.12

Table 15: Summary table of Cancer Testis Antigen (CTA) genes with progressively increasing transcriptional activation in D425 spine metastases relative to primary tumours. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

99

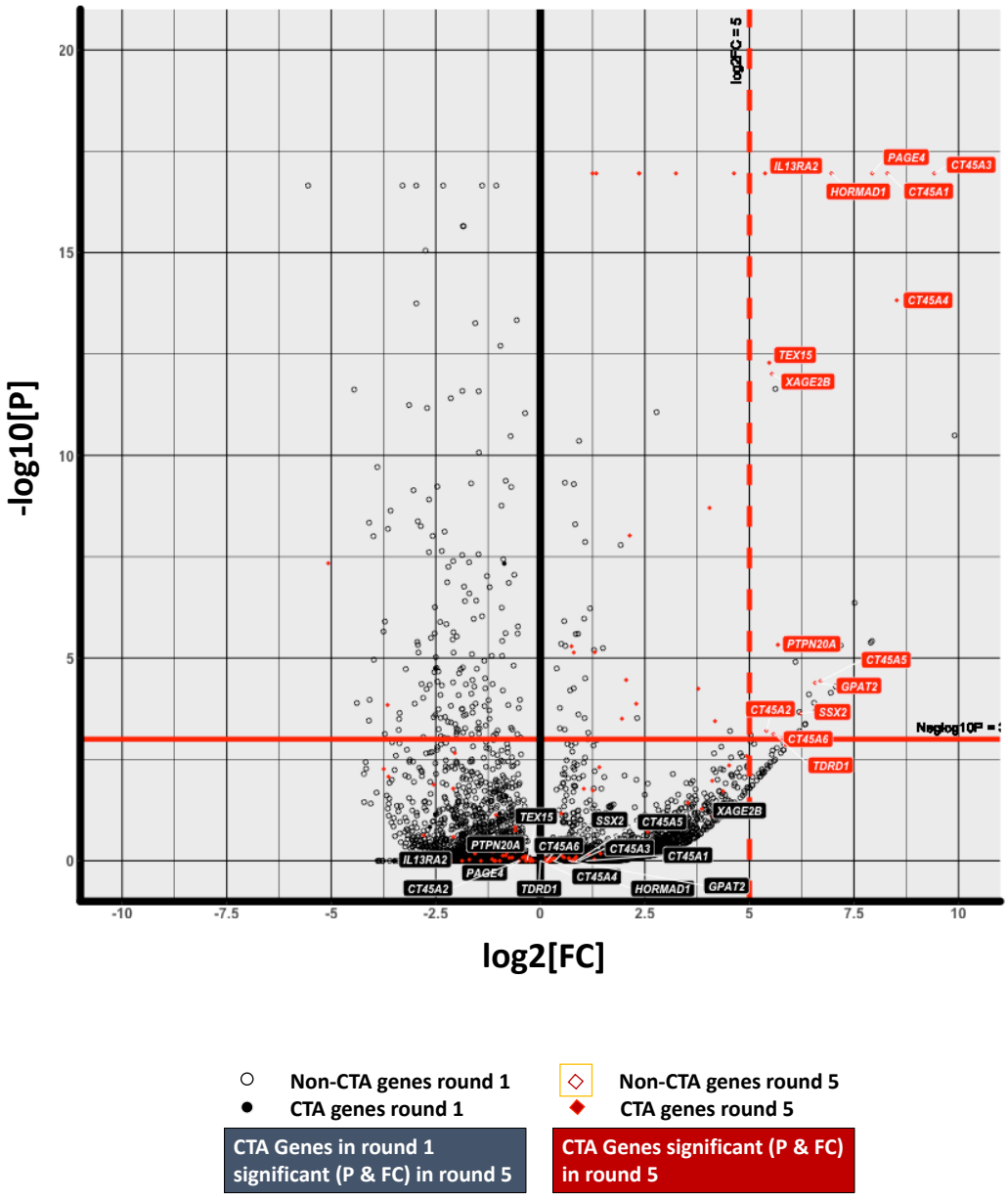


Figure 9. Scatter plot of –log10[P] as a function of log2[FC] in rounds 1 and 5 of selection. It is striking the magnitude of increase in both –log10[P] and log2[FC] that occurs between rounds 1 and 5 of selection for CTA genes.

Looking more closely at the 46 transcriptionally activated CTA genes (Table 14) , the CTA genes exhibiting this pattern were mainly from three specific CTA gene subfamilies namely the *MAGE*, *XAGE*, and *CT45* gene families (Tables 16-18). This progressive increase in transcription with every round of selection suggested that the increasing transcription of these genes favored metastasis of these cells across rounds of selection.

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5				
Cell Line	**D425 (Group 3)**												
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**				
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**spi5**
Candidate Gene													
MAGEF1	0.11	0.23	5.53	16.95	0.89	0.59	1.10	1.46	180.60	334.23	321.17	453.75	573.28
MAGEA10	0.03	0.04	12.23	16.95	1.54	1.29	2.86	2.36	342.21	993.10	969.56	2883.15	2016.80
MAGEC2	0.01	16.65	16.95	16.95	0.73	4.38	2.84	4.63	120.81	201.03	2955.00	1014.94	3440.64
NDN	1.70	7.78	3.06	4.88	0.79	1.69	1.36	1.26	116.53	201.61	445.01	352.57	322.53
MAGEB1	0.04	0.19	0.04	4.46	0.93	1.31	0.14	2.05	9.54	18.18	28.19	12.52	45.75
MAGEA1	0.01	0.05	0.01	2.31	-0.15	0.88	0.53	1.41	433.34	390.57	945.40	741.65	1333.19
MAGEE2	1.56	0.00	0.00	0.00	3.98	0.00	0.05	0.18	0.86	13.71	1.01	1.04	1.12
MAGEC3	0.00	0.00	0.00	1.28	0.00	0.00	0.00	3.86	0.00	0.00	0.00	0.00	14.48
MAGEH1	0.04	3.96	16.95	1.18	0.93	-3.83	2.48	2.11	24.37	46.34	2.01	160.64	121.27
MAGEA4	0.00	0.00	0.00	0.00	-0.07	-0.41	-0.36	4.03	1.15	1.10	1.01	1.04	21.78
MAGEB17	0.00	0.00	0.00	0.00	0.01	-0.35	-0.30	-0.18	1.09	1.10	1.01	1.04	1.12
MAGEA13P	0.00	0.00	0.00	0.00	0.35	0.00	0.05	0.18	0.86	1.10	1.01	1.04	1.12
MAGEB16	0.00	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.86	1.10	0.00	0.00	0.00
MAGEA8-AS1	0.00	0.00	0.00	0.00	0.35	0.00	0.05	0.18	0.86	1.10	1.01	1.04	1.12
MAGEB6P1	0.00	0.00	0.00	0.00	0.35	0.00	0.05	0.18	0.86	1.10	1.01	1.04	1.12
MAGEB18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAGEA5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAGEB3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAGEB4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAGEB5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAGEA2	0.03	0.00	0.00	0.04	0.54	-0.09	0.08	-0.31	1568.67	2276.33	1743.34	1952.23	1462.07
MAGEA3	0.00	0.00	0.00	0.02	0.81	-0.10	0.11	-0.31	1557.37	2725.24	1716.74	1972.72	1457.10
MAGEA9B	0.00	0.00	0.00	0.02	-0.35	0.06	0.29	-0.48	2049.41	1607.07	2528.49	2963.20	1710.94
MAGED1	0.01	0.01	0.01	0.03	0.29	0.54	0.37	-0.21	427.59	521.01	736.99	650.90	427.84
MAGEA8	0.00	0.00	0.00	0.05	-2.31	-2.68	1.17	-0.34	5.50	1.10	1.01	14.60	5.03
MAGED2	0.01	0.03	0.00	0.05	-0.18	-0.55	-0.18	-0.59	3854.01	3405.44	3105.01	4002.41	2944.97
MAGEB2	0.01	0.02	0.17	0.07	-0.50	-0.78	-3.00	-0.74	168.88	119.47	116.79	25.03	116.63
MAGED4B	0.02	0.02	0.01	0.09	-0.69	-0.20	0.15	-0.55	38.49	23.90	39.27	49.82	30.27
MAGEA9	0.03	0.03	0.03	0.21	-0.58	0.31	-0.21	-0.79	25.79	17.26	37.88	26.35	17.29
MAGEC1	0.03	0.03	0.02	0.24	-0.77	-0.06	0.25	-1.16	49.61	29.09	55.37	68.85	25.66
MAGEA6	0.02	0.03	0.03	0.32	0.25	0.13	0.38	-1.23	24.17	28.67	31.07	36.76	11.99
MAGED4	0.04	0.06	0.07	0.32	-0.80	-0.49	1.09	-1.56	8.43	4.84	7.05	21.11	3.31
MAGEA12	0.01	0.09	0.01	0.75	0.19	-0.52	-0.16	-0.60	1262.62	1438.05	1039.03	1330.44	959.79
MAGEA2B	0.03	0.02	0.02	1.13	0.58	0.03	-0.18	-1.06	49.94	74.60	59.89	51.70	27.61
MAGEA11	0.02	0.03	0.03	1.78	-0.33	0.33	-0.28	-2.08	47.14	37.57	70.48	45.90	12.84
MAGEE1	0.02	0.88	0.22	2.97	-0.19	-4.24	-2.19	-4.06	16.42	14.35	1.01	4.17	1.12

Table 16: Summary table of *MAGE* Cancer Testis Antigen (CTA) genes. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5				
Cell Line	**D425 (Group 3)**												
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**				
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**spi5**
Candidate Gene													
XAGE1D	0.00	3.51	0.88	16.95	0.23	0.97	0.76	1.34	1113.52	1304.18	2570.33	2214.22	3232.35
XAGE1B	0.00	1.30	1.79	16.95	0.11	0.73	0.79	1.25	4224.11	4562.24	8226.64	8598.31	11523.18
XAGE2B	1.24	13.54	3.95	12.01	4.01	4.76	4.32	5.54	2.03	32.90	63.80	46.94	108.17
XAGE5	0.02	0.08	0.14	3.87	-0.20	-1.44	1.78	2.29	11.89	10.35	5.03	46.94	66.46
XAGE1C	0.01	0.01	0.01	3.50	0.04	0.44	0.43	1.95	342.71	353.50	548.80	544.38	1522.72
XAGE1A	2.88	0.00	0.00	0.00	5.11	-0.49	-0.44	-0.32	1.22	42.50	1.01	1.04	1.12
XAGE2	0.63	0.00	0.00	1.10	3.78	4.48	0.05	4.14	0.86	11.96	22.78	1.04	17.54
XAGE1E	0.09	0.64	0.02	0.18	1.83	-5.71	0.93	1.44	46.67	166.39	1.01	101.31	144.17
XAGE3	0.04	0.02	0.02	0.05	-1.40	0.34	1.19	0.36	67.41	25.62	97.66	175.24	98.57

Table 17: Summary table of *XAGE* Cancer Testis Antigen (CTA) genes. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

70

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5				
Cell Line	**D425 (Group 3)**												
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**				
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**spi5**
Candidate Gene													
CT45A3	0.01	0.00	15.95	16.95	0.94	1.01	4.79	9.42	2.17	4.18	5.13	70.74	1723.95
CT45A1	0.00	0.00	0.00	16.95	1.10	3.07	4.24	8.30	0.94	2.04	9.41	21.11	343.92
CT45A4	0.00	0.00	0.00	13.82	0.02	3.00	3.64	8.52	1.09	1.10	10.36	16.06	466.07
CT45A5	0.00	0.00	0.00	4.43	0.87	0.73	1.46	6.70	1.16	2.12	2.25	3.72	139.32
CT45A2	0.00	0.00	0.00	3.21	-0.52	3.92	-0.81	5.40	1.59	1.10	27.95	1.04	77.13
CT45A6	0.00	0.00	0.00	3.12	0.15	0.61	3.07	5.58	1.26	1.40	2.27	12.50	69.73

Table 18: Summary table of *CT45* Cancer Testis Antigen (CTA) genes. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

71

In summary, iterative selection for a propensity for metastasis in the spine metastases of mouse xenograft models of human MB led to the identification of a variety of genes exhibiting a variety of patterns of changing transcription across the rounds of selection. While transcription of some genes was clearly selected against, a large subset exhibited patterns of progressively increasing selection across rounds of selection, suggestive of an ongoing association with metastasis, and making them significant genes of interest.

3.5.4 Long non-coding RNAs Exhibit Progressively Increasing Transcription Resembling that of CTA Genes

Among the genes exhibiting progressively increasing transcription across rounds of selection, the most significant was *XIST*, a long non-coding RNA (lncRNA). Given this finding, the P and FC of all lncRNAs were examined. Interestingly, a subset of lncRNAs exhibited increasing transcription with increasing magnitude of change similar to that of CTA genes (Table 19). Other lncRNAs also exhibited some of the other alternative patterns of changes in expression across rounds of selection that had been exhibited by CTA genes (Appendix 6). So, it can be hypothesized that lncRNAs are in a regulatory network that is shared with CTA genes, either upstream or downstream, ultimately promoting metastasis.

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5				
Cell Line	**D425 (Group 3)**												
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**				
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**spi5**
Candidate Gene													
XIST	2.14	16.65	16.95	16.95	-4.22	4.97	5.70	8.58	26.05	1.39	975.60	1615.78	11652.14
LINC01139	0.00	0.00	9.02	15.65	0.35	0.00	5.62	5.64	0.86	1.10	1.01	50.07	49.68
LINC00657	0.00	0.01	0.13	11.38	-0.29	0.42	0.69	0.99	2445.63	1996.32	3848.04	4616.80	5591.29
LINC00958	0.01	0.53	3.17	4.82	0.42	0.82	1.16	1.10	475.34	635.91	982.65	1242.34	1169.05
LINC00960	0.00	0.00	0.00	3.22	-2.28	1.03	1.63	3.00	5.40	1.10	13.09	19.82	50.23
LINC01106	0.00	0.00	0.00	3.19	-0.07	-0.41	3.95	4.97	1.15	1.10	1.01	20.86	41.79
LINC01152	0.00	0.00	0.00	2.44	0.89	0.00	2.36	4.75	0.86	1.60	1.01	5.22	26.74
LINC00355	0.00	0.00	0.00	1.28	0.65	1.57	2.04	3.69	0.86	1.35	3.02	4.17	12.82
LINC01123	0.00	0.00	0.00	0.89	0.35	2.31	3.50	3.33	0.86	1.10	5.03	11.47	10.02
LINC00683	0.00	0.00	0.00	0.84	0.35	1.57	2.36	3.26	0.86	1.10	3.02	5.22	9.51
SNHG7	0.01	0.07	0.08	0.81	0.12	0.86	0.94	1.09	58.05	63.26	125.11	131.43	143.26
TCF7L1-IT1	0.00	0.00	0.00	0.79	-1.25	0.76	-2.50	2.70	5.05	2.12	10.07	1.04	37.86
LINC00941	0.02	0.05	0.07	0.44	-1.32	0.62	-1.64	1.96	2.76	1.10	5.03	1.04	12.31
TRIM36-IT1	0.03	0.07	0.14	0.22	-0.34	-0.92	-3.18	1.80	8.25	6.50	5.03	1.04	32.86
FAM155A-IT1	0.04	0.05	0.09	0.12	0.69	0.38	-1.65	0.58	11.37	18.31	17.12	4.17	19.51
LINC00339	0.01	0.02	0.06	0.06	0.06	-0.21	0.42	0.44	111.21	115.86	112.76	174.20	173.24
SNHG15	0.00	0.00	0.01	0.01	0.05	-0.10	0.32	0.13	2364.42	2452.74	2587.51	3454.78	2969.31
LINC00665	0.00	0.01	0.01	0.03	0.17	0.11	0.59	-0.35	771.43	870.65	982.65	1367.81	700.99
JPX	0.01	0.01	0.01	0.05	0.52	-0.29	-0.52	-0.38	340.63	490.10	330.23	280.60	300.84
MIR181A1HG	0.01	0.01	0.02	0.05	-0.34	-0.19	-1.22	-0.61	548.94	432.68	559.79	274.34	413.73
SOS1-IT1	0.03	0.03	0.06	0.12	0.45	0.39	-1.00	-0.52	12.56	17.14	19.13	7.30	10.07
MAGI1-IT1	0.06	0.08	0.09	0.25	1.04	-1.36	-2.31	-2.19	4.49	9.27	2.01	1.04	1.12
SNHG3	0.00	0.03	0.06	0.26	0.10	-0.68	-0.84	-0.63	1038.19	1112.01	760.24	682.09	769.63
LINC00115	0.02	0.04	0.15	0.33	0.52	0.56	1.21	-1.35	25.57	36.57	44.29	69.02	11.58
LINC00221	0.01	0.02	0.02	0.63	-0.05	-0.01	0.37	-0.87	81.13	78.20	95.65	124.13	51.38
PTCSC3	0.08	0.11	0.18	0.67	-1.74	-1.98	-3.50	-3.40	10.40	3.10	3.02	1.04	1.12
DLEU1	0.00	0.01	0.03	1.13	-0.21	-0.39	-0.62	-0.78	863.65	748.13	770.21	658.20	578.89
LINC00910	0.01	0.02	0.07	1.37	-0.36	-0.20	-0.92	-1.44	128.30	99.79	130.89	79.28	54.22
LINC00839	0.01	0.02	0.11	1.46	-0.24	0.40	-1.29	-1.74	67.54	57.17	104.71	32.34	23.42
SNHG8	0.01	0.01	0.12	4.29	0.35	-0.14	0.69	-1.30	975.49	1242.26	1044.07	1841.09	456.46
LINC00951	0.02	0.05	0.41	4.37	-0.51	-0.62	-1.52	-2.41	43.68	30.58	33.22	17.73	9.48
LINC00882	0.02	0.58	4.33	5.09	-0.47	-1.05	-2.35	-2.22	49.81	36.06	28.19	11.47	12.31
LINC01024	0.02	0.04	1.50	5.11	-0.03	-0.68	1.34	-4.08	16.43	16.13	12.08	49.03	1.12
LINC00312	0.04	0.11	0.16	6.06	1.23	0.58	0.63	-4.50	44.16	103.96	77.52	80.32	2.24
LINC00883	0.01	0.15	2.45	9.02	-0.24	-0.47	-0.97	-1.12	265.87	225.61	225.53	159.60	140.66

Table 19: Summary table of lncRNAs with progressively increasing transcription in D425 PDX models. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 (P < 0.05); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

3.6 Transcription of CREB Pathway Components, Known Downstream Effectors of CTA Gene *CT45A1*, are Unaffected by Iterative Selection

The mechanism by which CTA gene *CT45A1* regulates metastasis has been characterized in detail and is thus a great model to use for how other CTA genes at the same locus, or other CTA genes in general, might regulate metastasis. Specifically, *CT45A1* activates ERK/CREB in order to drive metastatic dissemination, evidenced by the fact that siRNA-mediated silencing of *CT45A1* inhibits metastasis and at the molecular level leads to significantly reduced *CT45A1* translation, significantly lower levels of proteins BCL2, BIRC5 (survivin), and MMP9, moderately lower levels of MMP2, and increased levels of BAX [72]. By extension, it is possible that increased *CT45A1* transcription would similarly lead to corresponding changes in the transcription of CREB pathway components. In reality, changes in the expression of all CREB genes was negligible across all rounds of selection (Table 20). So, while the CREB pathway is a downstream effector of *CT45A1*, it does not appear to be the case here. It is very likely that that activation of CREB signaling is context-specific, and so a lack of activation in this context does not necessarily negate the hypothesis that CREB signaling might act downstream of CTA genes other than *CT45A1*, just under different conditions. An effective way to validate that the CREB pathway is not activated as a result of CTA activation would be to perform another GSEA with just the genes of the CREB pathway used as the sole gene set.

With transcriptional activation now identified, it was necessary to examine what were the effects of selection at the level of pathways.

Cell Line	D425 (Group 3)												
Statistic	-log10[P]				log2[FC]				Mean Transcription (FPKM)				
Comparison	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1	spi1	spi2	spi3	spi5
Candidate Gene													
CT45A1	0.00	-2.51E-13	-6.71E-08	-16.95	1.10	3.07	4.24	8.30	0.94	2.04	9.41	21.11	343.92
BCL2	-4.32E-03	-4.34E-16	0.00	0.00	1.42	-4.56E-03	0.00	0.00	0.86	2.32	1.01	0.00	0.00
BAX	-0.01	-0.05	-0.02	-0.08	0.50	-1.07	-0.18	-0.48	112.86	159.24	63.43	116.83	93.80
MMP2	-9.11E-03	-0.04	-0.01	-0.03	-0.33	0.69	-0.16	0.11	214.61	170.88	409.77	225.31	268.41
BIRC5	-4.48E-03	-0.03	-9.23E-03	-4.88	-0.16	-0.65	-0.27	-1.98	644.49	576.28	482.46	628.80	189.55

Table 20. Transcription of components of the CREB pathway, known to act downstream of CT45A1, is unaffected by iterative selection. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 (P < 0.05); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

75

3.7 *Vimentin* (*VIM*), a Metastasis Regulator Known to Function Downstream of CTA Genes, is Activated in Spine Metastases Relative to Primary Tumours

CTA genes have been found to regulate metastasis indirectly, modulating the activity of downstream effector genes and proteins that promote or impede metastasis pathways [72]. To investigate whether intermediate effectors were similarly involved in this study, the transcription of four known CTA downstream effector genes *TWIST1* (Twist1), *VIM* (Vimentin), *CDH1* (E-Cadherin), and *CDH2* (N-Cadherin) was examined across rounds of selection (Table 21). Transcription of *VIM* was 430.05 and 439.98 in spine metastases from rounds 1 and 2, respectively, in both cases more than 2X higher than the level of *VIM* expression observed in primary tumours from round 1, specifically 176.95. The difference reached statistical significance in round 5 (-log10[P]=16.95). In contrast, while *TWIST1* transcription was 19.13 and 6.14 in rounds 2 and 5, respectively, greater than the 0.86 observed in primary tumors, the difference was never statistically significant. Furthermore, *TWIST1* expression was always very low in both primary and spine tissues, increasing the chance that the difference in transcription was artifactual. Finally, cadherin genes E-Cadherin (*CDH1*) and N-Cadherin (*CDH2*) exhibited contrasting changes in transcription across rounds of selection; *CDH1* transcript was nearly undetectable across all rounds of selection; in contrast, *CDH2* transcription increased from 7.52 in primary tumours to 137.46 in spine metastases in the first round of selection, a difference that was statistically significant (-log10P=2.19). *CDH2* transcription decreased significantly in subsequent rounds, to 20.14, 2.09, and 21.26 in rounds 2, 3, and 5, respectively. But even with these decreases in transcription, in rounds 2 and 5, levels of *CDH2* still remained higher in spine metastases than in primary tumours although not to an extent that was statistically significant. Taken together, changes in *VIM* transcription across rounds of selection resembled the progressive transcriptional increases observed for CTA genes; thus, this led us to hypothesize that *VIM* might be a part of a regulatory pathway, shared with CTA genes, which serves to regulate Group 3 MB metastasis.

Cell Line	D425 (Group 3)												
Statistic	-log10[P]				log2[FC]				Mean Transcription (FPKM)				
Comparison	ri1spi	pri1spi2	pri1spi3	pri1spi5	ri1spi	pri1spi2	ori1spi	ri1spi	pri1	spi1		spi3	spi5
Candidate Gene													
VIM	0.30	0.06	0.01	16.95	1.28	1.10	-0.42	2.34	176.95	430.05	439.98	153.34	1028.00
TWIST1	0.00	1.68E-08	4.34E-16	1.30E-15	0.35	4.23	0.05	2.63	0.86	1.10	19.13	1.04	6.14
CDH1	0.00	4.34E-16	0.00	0.00	0.35	-4.56E-03	0.00	0.00	0.86	1.10	1.01	0.00	0.00
CDH2	2.19	4.77E-15	0.04	0.17	4.19	1.19	-2.07	1.29	7.52	137.46	20.14	2.09	21.26

-log10[P]>1.30103 | -log2[FC] > 1 | FPKM > 5

Table 21. Transcription of known CTA downstream effector *VIM* is activated in round 5 of selection but not that of several other known metastasis effectors. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

77

3.8 CTA Genes with Increased Transcription in Spine Metastases Localize in Clusters on the X Chromosome

Two subfamilies of CTA genes exist, which differ with respect to where they are localized in the human genome. The CT-X subfamily of CTA genes are clustered within regions of the X-chromosome, specifically Xp11, and Xq26-28 [73, 74]. All other CTA genes are spread across the autosomes. The most relevant difference between these groups is that increased transcription of CT-X genes is associated with poor prognosis for a variety of cancer types including but not limited to breast cancer [75], multiple myeloma [76], and non-small-cell lung cancer [77]. In my studies, CTA genes with increased transcription in MB spine metastases relative to primary tumours clustered in specific regions of X (Table 22). Specifically, the *CT45A* gene cluster, *MAGEC2*, and several *MAGEA* genes cluster within the terminal 23 Mb of the X chromosome [78]. It is interesting to note that a majority of the genes most significantly activated mapped to the negative strand. This is made especially interesting given that *XIST* also maps to the negative strand, albeit more than 68 Mb (68,217,543 bp) upstream of these gene clusters. However, the *XIST* locus is only 21 Mb (21,299848 bp) downstream of significantly activated members of both the *XAGE* and *GAGE* CTA gene families, and *MAGE* gene *MAGEB1*. Taken together, activated CTA genes clustered on the X-chromosome, the most significant of which were on the negative strand, in relatively close proximity to *XIST*.

Thresholds	-log10[P] > 1.30103				log2[FC] > 1				FPKM > 5					Gene on X			
Cell Line	**D425 (Group 3)**																
Statistic	**-log10[P]**				**log2[FC]**				**Mean Transcription (FPKM)**					**Genome Location**			
Comparison	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1spi1**	**pri1spi2**	**pri1spi3**	**pri1spi5**	**pri1**	**spi1**	**spi2**	**spi3**	**spi5**	**Chromosome**	**Strand**	**Start**	**End**
Candidate Gene																	
MAGEA10	0.03	0.04	12.23	16.95	1.54	1.29	2.86	2.36	342.21	993.10	969.56	2883.15	2016.80	**X**	**-1**	**151301782**	**151307033**
MAGEC2	0.01	16.65	16.95	16.95	0.73	4.38	2.84	4.63	120.81	201.03	2955.00	1014.94	3440.64	**X**	**-1**	**141290131**	**141293076**
CT45A3	5.99E-03	8.68E-16	15.95	16.95	0.94	1.01	4.79	9.42	2.17	4.18	5.13	70.74	1723.95	**X**	**1**	**134883488**	**134891885**
XAGE1D	4.63E-03	3.51	0.88	16.95	0.23	0.97	0.76	1.34	1113.52	1304.18	2570.33	2214.22	3232.35	X	**-1**	**52528159**	**52533303**
XAGE1B	1.85E-03	1.30	1.79	16.95	0.11	0.73	0.79	1.25	4224.11	4562.24	8226.64	8598.31	11523.18	X	**1**	**52235228**	**52243954**
PAGE4	0.00	8.68E-16	4.34E-16	16.95	-0.50	1.12	0.18	7.94	1.56	1.10	4.03	2.09	446.86	X	**1**	**49593863**	**49598576**
IL13RA2	1.30E-15	1.87E-14	2.77	16.95	-3.49	1.09	2.45	5.37	12.49	1.10	31.21	80.32	596.60	X	**-1**	**114238538**	**114254540**
CT45A1	0.00	2.51E-13	6.71E-08	16.95	1.10	3.07	4.24	8.30	0.94	2.04	9.41	21.11	343.92	X	**1**	**134847185**	**134857354**
HORMAD1	0.00	4.34E-16	4.34E-16	16.95	0.31	-0.35	-0.30	6.96	1.09	1.36	1.01	1.04	158.60	1	**-1**	**150670536**	**150693364**
GAGE12I	0.01	15.35	5.47	16.95	-0.13	1.85	1.42	2.36	78.34	71.56	334.10	246.51	466.52	X	**1**	**49216659**	**49223943**
GAGE1	0.02	1.47	16.95	16.95	-0.28	2.17	3.86	3.24	23.58	19.47	122.99	396.05	256.00	X	**1**	**49363628**	**49373139**
CT45A4	0.00	3.24E-13	8.24E-11	13.82	0.02	3.00	3.64	8.52	1.09	1.10	10.36	16.06	466.07	X	**-1**	**134928331**	**134953994**
XAGE2B	1.24	13.54	3.95	12.01	4.01	4.76	4.32	5.54	2.03	32.90	63.80	46.94	108.17	X	**1**	**52112173**	**52118823**

Table 22. Summary table of expression data and chromosome localization for most significantly activated CTA Genes. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM; white shading = Chromosome localization

79

3.9 Five CTA Genes Were Selected For Further Study

Based on the findings up to this point, CTA genes, specifically *PAGE1, MAGEA10, MAGEC2, CT45A3,* and *CT45A1* were selected for further study based on their significant transcriptional activation across rounds of selection, with FC also becoming progressively larger, and previous links to metastasis existing in the literature (Table 23, Figure 10). The protein products of these genes were subject to validation by Immunohistochemistry, to determine whether the transcriptional changes observed were leading to changes at the protein level.

Cell Line	D425 (Group 3)												
Statistic	-log10[P]				log2[FC]				Mean Transcription (FPKM)				
Comparison	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1	spi1	spi2	spi3	spi5
Candidate Gene													
PAGE1	0.00	16.65	0.00	1.16	0.06	0.60	0.36	0.50	9390.16	9814.96	16708.07	14111.17	15271.94
MAGEA10	0.03	0.04	12.23	16.95	1.54	1.29	2.86	2.36	342.21	993.10	969.56	2883.15	2016.80
MAGEC2	0.01	16.65	16.95	16.95	0.73	4.38	2.84	4.63	120.81	201.03	2955.00	1014.94	3440.64
CT45A3	0.01	0.00	15.95	16.95	0.94	1.01	4.79	9.42	2.17	4.18	5.13	70.74	1723.95
CT45A1	0.00	0.00	0.00	16.95	1.10	3.07	4.24	8.30	0.94	2.04	9.41	21.11	343.92

-log10[P]>1.30103 -log2[FC] > 1 FPKM > 5

Table 23: Summary table with five CTA genes selected for further study. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 ($P < 0.05$); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM.

81

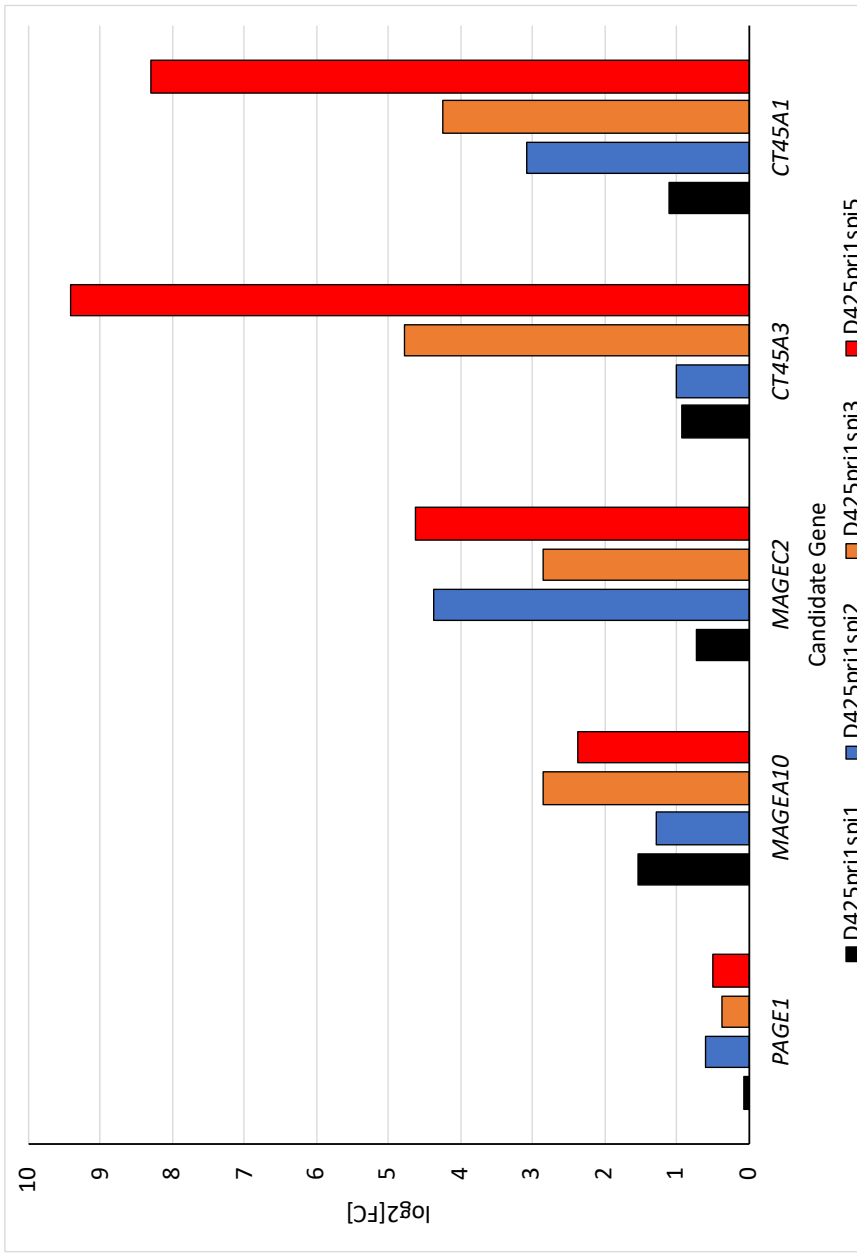


Figure 10. Histogram depicting progressively increasing change in log2[FC] with increasing rounds of selection for the five CTA genes selected for further study. pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection.

3.10 Enrichment of CTA Gene Family Precedes Enrichment of Metastasis Pathways

Gene Set Enrichment Analysis (GSEA) was performed using lists of the most significantly activated genes to determine whether specific pathways of metastasis were represented by those most activated genes. For this investigation the CTA gene family was treated as a gene set. After one round of selection, the CTA gene family was not significantly enriched in spine metastases from any of the PDX mice (Appendix 8). However, two pathways, specifically nonsense-mediated decay and eukaryotic translation termination, were enriched in both D425 and MB411FH. But these pathways were not known to be associated with metastasis. More generally, pathways significantly enriched in the first round of selection did not include any commonly associated with metastasis. After the second round of selection however, for D425 CTA genes did become enriched (FDR=0.06) (Appendix 8a). Interestingly, for MB411FH FDR values corresponding to the CTA gene family decreased significantly from round 1 to round 2, although not reaching significance in either case (0.42 in round 1, 0.19 in round 2). This suggested that with additional replicates or rounds of selection CTA genes might have also reached significance in MB411FH (Appendix 8b). In subsequent rounds of selection several well established pathways of metastasis also reached significance, especially in round 5 (Figure 11, Appendix 8a). So, this suggested that transcriptional activation of CTA genes might have been a necessary predecessor to the activation of metastasis pathways.

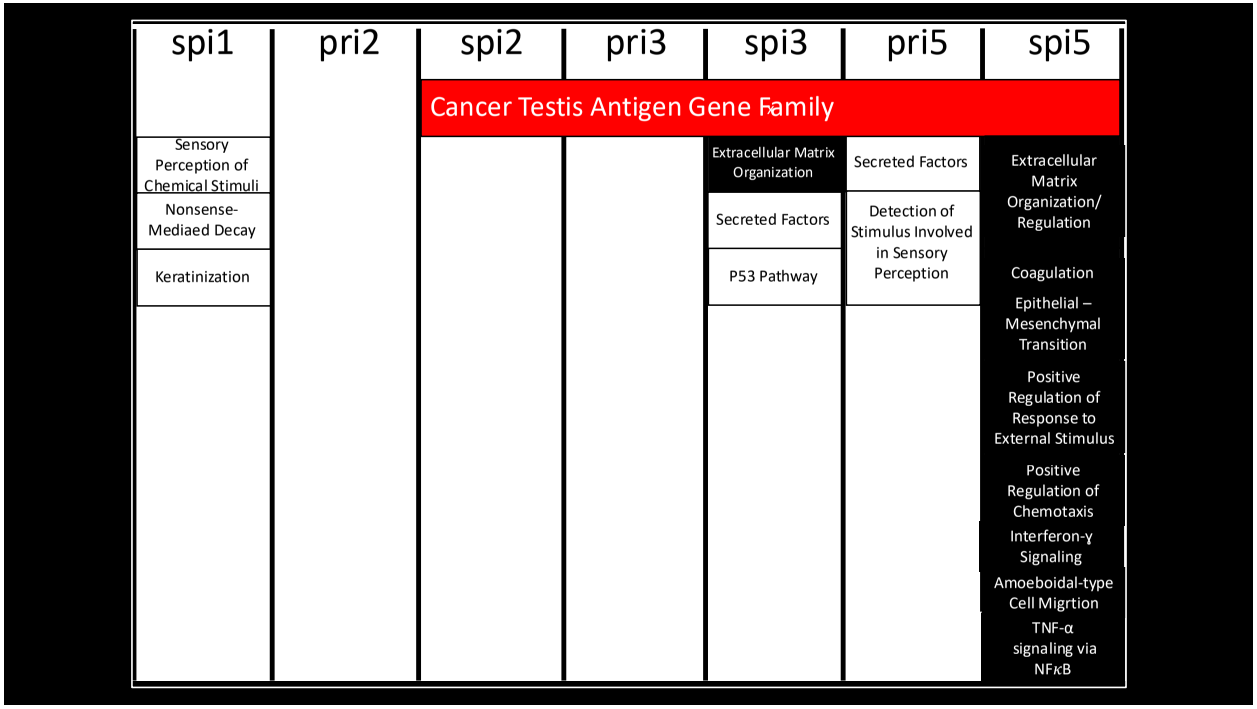


Figure 11. Visual representation of the relative timing of pathways reaching significance across rounds of selection for metastasis propensity among spine metastases of D425 PDX mice. Significance threshold – FDR < 0.15. Red shading = CTA gene family; Black shading = Metastasis pathways; White shading = Non-metastasis pathways.

84

3.11 Metastatic Pathways Activated in Correlation with CTA Genes Model Pathways Activated in Human Group 3 MB Metastases

Now that the genes transcriptionally activated in MB spine metastases, and pathways represented by those genes had been identified, the degree to which these were reflective of actual MB in human patients was assessed. I conducted a differential expression analysis of 4 spine metastases relative to 4 primary tumours from human Group 3 MB patients. The list of genes significantly activated in spine metastases relative to primary tumours I compared to a corresponding list of genes from D425 PDX mice significantly activated in at least one round of selection. I also performed GSEA using a ranked list of the genes significantly activated in human spine metastases, and compared the list of activated pathways (FDR < 0.15) and pathway-specific genes with those from the D425 PDX mice in the fifth round of selection.

Out of 1966 genes significantly activated (-log10[P] X sign of log2[FC] > 1.30103) in D425 PDX mouse spine metastases in at least one round of selection, 3 genes, specifically *ROBO2*, *NMNAT3*, and *SV2A*, were shared with human MB patients (Table 24). Interestingly, *ROBO2* is a tumor suppressor gene, linked to luteolysis in the ovary [79], which has also been implicated in the metastasis of breast cancer [80]; *ROBO2* would normally suppress tumour development by inactivating β-catenin/LEF/TCF and PI3K/Akt signaling so upon deletion, *ROBO2*-mediated inhibition of signaling would be relieved leading to tumourigenesis. In addition, *ROBO2* is deleted in more than half of 19 patients with metastatic colorectal cancer [81]. However, in contrast to these cases metastasis was associated with *ROBO2* activation. So, mechanisms, including *ROBO2* but different than those described in previous studies may be at work. Alternatively, the transcriptional activation of *ROBO2* observed might have been in correlation with something other than metastasis. In support of this argument, while *ROBO2* became significantly activated in the third round of selection, the activation was no longer significant in round 5.

blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM; white shading = Chromosome localization

Thresholds	-log10[P] > 1.30103	log2[FC] > 1	log2Exp > 5
Source	Human MAGIC Group 3		
Statistic	-log10[P]	log2[FC]	Mean log2 Transcription
Comparison	prispi	prispi	prispi
Candidate Gene			
ROBO2	1.86	2.75	6.22
NMNAT3	1.51	-2.04	6.11
SV2A	1.51	1.91	7.51

Thresholds	-log10[P] > 1.30103	log2[FC] > 1	FPKM > 5										
Source	D425 (Group 3)												
Statistic	-log10[P]				log2[FC]				Mean Transcription (FPKM)				
Comparison	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1spi1	pri1spi2	pri1spi3	pri1spi5	pri1	spi1	spi2	spi3	spi5
Candidate Gene													
ROBO2	0.03	######	7.43	0.25	0.02	-3.27	3.00	1.44	8.38	8.48	1.01	78.23	26.17
NMNAT3	0.06	0.05	0.13	1.43	1.21	-0.54	1.25	1.83	11.36	26.37	9.06	31.29	46.29
SV2A	0.02	0.02	0.21	16.95	-0.79	0.34	0.64	1.38	185.09	106.82	275.87	337.97	556.05

Table 24. Comparison of genes activated in PDX mice and human patients from the Medulloblastoma Advanced Genomics International Consortium (MAGIC) tissue bank. pri = primary tumour; spi = spine metastasis; pri1/spi1 = first round of selection; spi2 = second round of selection; spi3 = third round of selection; spi5 = fifth round of selection; red shading = -log10[P] > 1.30103 (P < 0.05); blue shading = log2[FC] > 1 (FC > 2); black shading = Transcription > 5 FPKM; white shading = Chromosome localization

98

In contrast to *ROBO2*, no links to metastasis have been made for *NMNAT3*. Instead, it is a gene, which encodes an enzyme, which catalyzes an important step in the production of NAD from ATP [82]. Overexpression has further been shown to lessen age-related insulin resistance [83]. However, it is interesting to note that *NMNAT3* was significantly activated by round 5. An identical pattern was also observed for *SV2A*, a gene encoding a synaptic vesicle glycoprotein which plays an important role in the regulation of neurotransmitter release and has anti-epileptic properties. *SV2A* has not yet been linked to cancer or metastasis. However, in this study -log10[P] for *SV2A* increased across rounds of selection, reaching maximum possible significance in round 5. So taken together, while there is overlap in the genes that are transcriptionally activated in the spine metastases of mouse PDX models of human Group 3 MB, and those of human Group 3 MB patients, the overlap is small, and those that do overlap, and are significantly activated in round 5 of selection do not show clear associations with metastasis.

Interestingly, much more overlap existed between mouse PDX models and human patients at the level of pathways (Table 25). Extracellular matrix organization, and pathways regulating cell migration (regulation of chemotaxis, positive regulation of leukocyte migration) were enriched in both PDX models and human patients. In addition, the epithelial – mesenchymal transition was enriched in both PDX mice, and humans. Finally, specific signaling pathways were jointly enriched (Interferon- ɣ Signaling, G-proteins coupled to cyclic nucleotide second messengers, and tumor necrosis factor alpha (TNF-α) signaling via Nuclear Factor Kappa-light-chain-enhancer of B-cells (NFκB).

Thresholds	FDR < 0.05		Note: Pathway specific genes are given for all pathways with FDR < 0.15		
	Comparison	Mouse PDX Models pri-1 vs. spi-5		Human Patients pri vs. spi	
Category	Feature	-log10[FDR]	Core Genes	-log10[FDR]	Core Genes
ECM	EXTRACELLULAR_MATRIX_ORGANIZATION--REACTOME	0.04	*FN1,KDR,ADAMTS3,CAPN2,LAMA1,BMP7,SDC1,FBLN5,MMP16,ITGA2B,MMP1,ITGB1,FBN2,SDC3,SPOCK3,COL4A6,COL2A1,ITGA2,COL11A1,EFEMP1,FMOD,COL4A1,SPARC,ITGA3,NID1,ITGB5,ADAMTS18,COL5A2,COL3A1,COL7A1,ITGAE,TLL2,DCN,COL24A1,CAPN3,LAMA4,ADAMTS4*	0.10	*ACOT7,PCYOX1L,OGN,CHAC1,CHAC2,OMD,FMOD,NCAN,HEXB,AGXT,GLB1,CTH*
CELL MIGRATION	REGULATION_OF_CHEMOTAXIS--GOBP	0.04	*CD74,PDGFRB,KDR,FGFR1,CXCL11,VEGFC,SCG2,SEMA4D,NTRK3,SEMA6A,SEMA3E,SEMA5B,SEMA4C,BST1,F3,DAPK2,C1QBP,SLIT2,FIGF*	0.11	*MFI2,RHBDD1,F12,CCBE1,MYH9,HPN,C19orf80,GSN,NKD2*
CELL MIGRATION	POSITIVE_REGULATION_OF_LEUKOCYTE_MIGRATION--GOBP	0.14	*CD74,CXCL11,VEGFC,DAPK2,C1QBP,FIGF,AIF1,WNT5A*	0.06	*PDE2A,GUCY1B3,PDE3A,ITPR1,GUCY1A2,PDE10A,GUCY1A3,NOS3,PDE11A,NOS1,PDE1A*
METASTASIS	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION--MSIGDB_C2	0.01	*VIM,DPYSL3,PDGFRB,FN1,RHOB,GEM,TPM2,LAMA1,SGCB,SDC1,VEGFC,FBLN5,SCG2,SERPINE2,PFN2,MMP1,ITGB1,FBN2,MYLK,FSTL1,ACTA2,EDIL3,ITGA2,TGM2,COL11A1,FMOD,COL4A1,SPARC,SNTB1,WIPF1,ITGB5,COL5A2,COL3A1,COL7A1,LRRC15,DCN,MYL9*	0.00	*MSX1,PDLIM4,COL11A1,GPX7,IGFBP2,SDC4,THY1,ID2,LOXL1,TPM2,CD59,TNFRSF12A,PMP22,LAMA3,TIMP3,SLIT3,LAMA1,NT5E,CXCL12,GADD45B,FBLN5,LOX,RHOB,COLGALT1,MAGEE1,ITGA2,BDNF,ELN,CRLF1,ITGAV,TNFAIP3,SFRP1,CAP2,GLIPR1,LAMC2,TIMP1,FMOD,RGS4,PDGFRB,FERMT2,ITGA5,AREG,PVR,SERPINH1,SERPINE1,MFAP5,CALU,GREM1,COL12A1,IL32,NTM,FBN2,PRRX1,COL8A2,NNMT,PCOLCE,MEST,MYLK,TGM2,GADD45A,TGFB1,LGALS1,IL6,TAGLN,PLAUR,VIM,FBN1,CDH11,LRRC15,SNAI2,CXCL6,VCAM1,IL15,COL16A1,FOXC2,ADAM12,NOTCH2,EMP3,FSTL1,GAS1,MMP2,FZD8,FSTL3,COL4A2,CXCL1,FN1*
SIGNALING	INTERFERON_GAMMA_SIGNALING--REACTOME	0.05	*JAK1,B2M,TRIM45,TRIM26,TRIM21,TRIM34,HLA-DPA1,MID1,MT2A,GBP1,HLA-DRB5,TRIM29,HLA-C,HLA-DPB1*	0.13	*FGFR3,INPPL1,MEF2C*
SIGNALING	G-PROTEIN_COUPLED_RECEPTOR_SIGNALING_PATHWAY,_COUPLED_TO_CYCLIC_NUCLEOTIDE_SECOND_MESSENGER--GOBP	0.08	*GNAQ,MC4R,ADCY6,AGT,PDE4D,ADCY3,CACNA1D,GNAS,GRM4,CNR1,GRM3*	0.12	*SIGMAR1,PNPT1,MGST1,MLKL*
SIGNALING	HALLMARK_TNFA_SIGNALING_VIA_NFKB--MSIGDB_C2	0.10	*RHOB,GEM,NFIL3,TLR2,CXCL11,PLK2,IRS2,IFIH1,NFKBIA,KLF6,GCH1*	0.02	*LIF,B4GALT5,SLC2A6,PER1,F2RL1,PPAP2B,CD83,CXCL2,SOCS3,SDC4,GPR183,ID2,KLF9,DUSP2,ICAM1,SPSB1,MAFF,GADD45B,G0S2,RHOB,MAP2K3,ZC3H12A,ZFP36,PTGS2,BHLHE40,FOSL1,F3,SLC2A3,NFKB2,MSC,SPHK1,DRAM1,TNFAIP3,KLF10,SMAD3,CEBPB,CSF1,AREG,KLF2,GCH1,SERPINE1,HES1,PHLDA2,CD80,TLR2,SERPINB8,CXCL3,KLF4,PTPRE,MYC,KYNU,BMP2,GADD45A,IL6,PLAUR,DUSP1,NINJ1,TRIB1,IRS2,PPP1R15A,PTGER4,NFKBIA,TNFAIP2,ETS2,TNIP2,FOS,LITAF,IL15RA,CXCL6,CDKN1A,ATF3,HBEGF,TNF,JUNB,FOSL2,JAG1,SLC16A6,EFNA1,KLF6,CXCL1,IL18,IFIT2,BCL2A1,LDLR,OLR1,SOD2,DUSP4*

Table 25. Summary table of pathways identified by GSEA as being enriched that are shared between D425 PDX mice and human patients. For both the mouse and human data, FDR < 0.15 was used as the criterion for inclusion in the summary table, and for displaying pathway-specific genes.

Many of these shared pathways and mechanisms have been shown to play essential roles in the transition of tumours from being localized to being metastatic. With respect to signaling, it is interesting that Interferon-ɣ signaling is enriched, since in the context of cancer development and metastasis Interferon-ɣ is most often observed to be expressed in Natural Killer cells, and actually inhibits tumour growth and metastasis by several means, including but not limited to modifying the tumour architecture to be less conducive to metastasis [84]. It is possible that the Interferon-ɣ signaling activity that was detected in samples from human patients could have come from Natural Killer cells that had infiltrated the tumour rather than tumour cells. But this could not have also been applied to the samples from xenografts, since the mice were NSG, and thus, by necessity, lacked a functional immune system. However, in a separate study overexpression of Interferon-ɣ did promote increased metastasis [85]. It is proposed that the increased Interferon-ɣ expression might have conferred upon those cells resistance to Natural Killer cells, thereby rendering them able to evade immune detection and elimination.

With respect to the signaling axis including TNF-α and NFκB, this axis serves to promote cancer cell survival, and metastasis via downstream signaling activated by the activity of these pathways. Increased TNF-α signaling activates NFκB signaling [86]. This, in turn, increases production of CXCL1/2. This signal then serves to recruit myeloid cells to the tumour, via interactions of CXCL1/2 in the tumour cells with CXCR2 receptor on the surface of myeloid cells . Once interspersed with the tumour, myeloid cells express S100A8/9, which promotes tumour cell survival and metastasis. TNF-α can bring about morphological changes associated with the epithelial-mesenchymal transition [87]. This is a mechanism in which cells undergo major morphological changes, and make significant changes to their microenvironment in order to transition to a state, similar to mesenchymal cells, in which they can migrate away from the primary tumour bed [88]. Along this line, the shared pathways associated with cell migration also have established links to metastasis. It has now been established that metastatic tumors migrate by chemotaxis, migrating towards chemical cues, which in the case of migrating tumour cells, are secreted exosome-like extracellular vesicles [89]. Secreted vesicles form a gradient from their point of secretion, and so cancer cells regulate their movement based upon their position in the gradient. It is also known that white blood cells or leukocytes of the immune system such as neutrophils also migrate using similar mechanisms [89] and play critical roles in mediating metastasis [90].

Examining the pathway-specific genes for which activation led to the pathway being identified as being enriched, what became apparent was that, while pathways were shared, only 3 genes were shared. This suggested that metastasis was being driven in both cases by the malfunctioning of common pathways, but activation of different genes from the same pathways were responsible for the pathway malfunction in PDX models and human patients. No matter where the mutation happened the effect on the pathway would be the same. So, this could represent an opportunity to identify previously uncharacterized gene candidates for further study based on their implication in pathways already linked to human MB.

The limited number of transcriptionally activated genes shared between human MB patients and PDX models might have been, in part, due to the current low availability of human metastatic tissues for research. With greater numbers of samples will come greater power to define profiles of MB metastasis.

3.12 Validation by IHC: PAGE1 Protein is Expressed in Both the Primary Tumour and Spine Metastases but a Difference in Protein Levels Could Not be Detected

To validate whether the increased transcription observed for the top 5 gene candidates led to increased protein levels, immunohistochemistry was performed using frozen sections of primary tumours and spine metastases from mice subject to in vivo iterative selection. Staining of testis tissue with antibodies against each of the protein targets being studied served as positive controls to verify that the IHC procedure and reagents would be able to detect PAGE1, MAGEA10, MAGEC2, and CT45 proteins (Figure 12). Using this validated approach PAGE1 protein was then shown to be clearly present in both primary tumours and spine metastases from the first round evidenced by the aligning of PAGE1 IHC (Figure 13c,d) with H & E staining (Figure 13a,b). The same was not true, however, for MAGEA10 (Figure 14), MAGEC2 (Figure 15), or CT45A1 (Figure 16), for which negligible expression was observed in either primary tumours or spine metastases from each of the rounds of selection. Also, it was not possible to get separate antibodies for CT45A1 and CT45A3 so one staining was done for CT45 proteins that could have been either CT45A1 or CT45A3. So, taken together this has shown that PAGE1 is expressed in both MB primary tumours and spine metastases, but not whether protein levels are as different as the levels of *PAGE1* transcript that were observed.

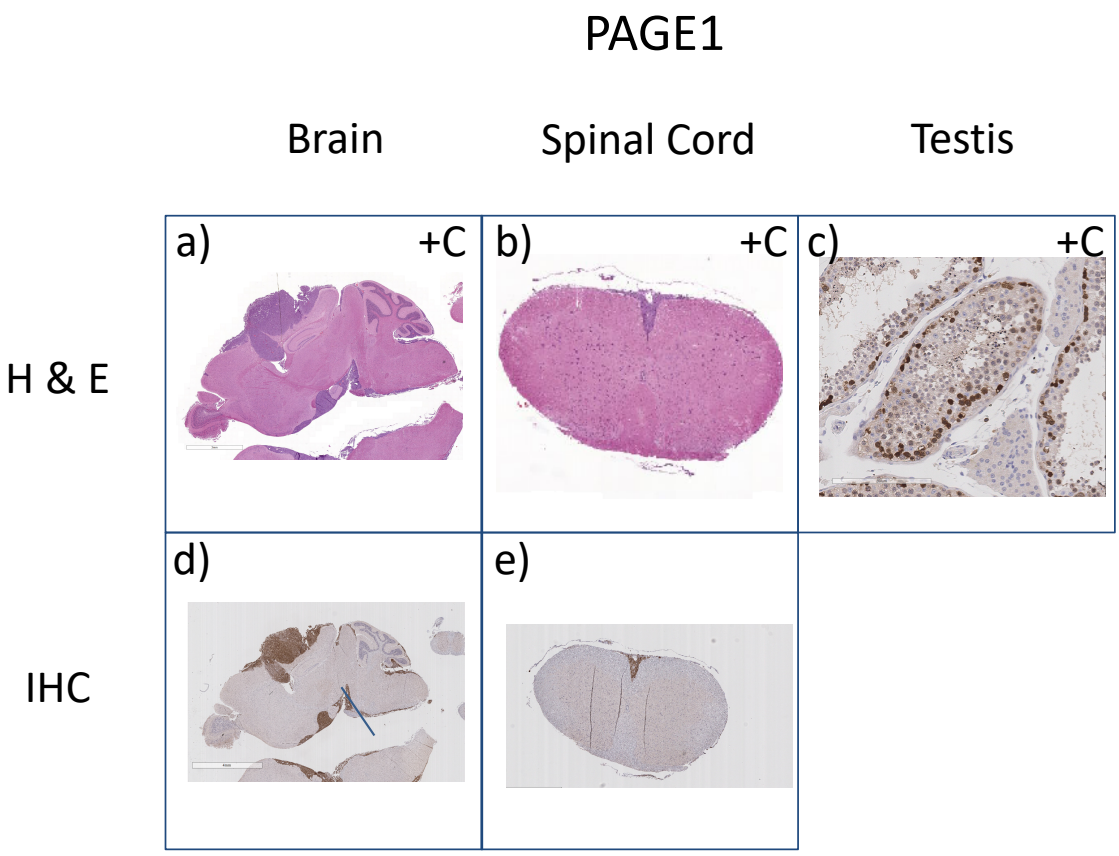


Figure 12. Validation of PAGE1 expression in xenograft primary tumour and spine metastases by immunohistochemical (IHC) detection of PAGE1 protein. a), b) Negative control (-C) H & E staining of brain and spinal cord sections adjacent to the sections to be used for IHC to identify the location of tumours prior to IHC c) Positive control (+C) staining of testis tissues to validate that PAGE1 antibody can be used successfully d), e) IHC for PAGE1 in sections of brain (d)) and spinal cord (e)). Staining is evident in both tissues, but the number of stained cells was insufficient to identify differential expression of PAGE1 protein.

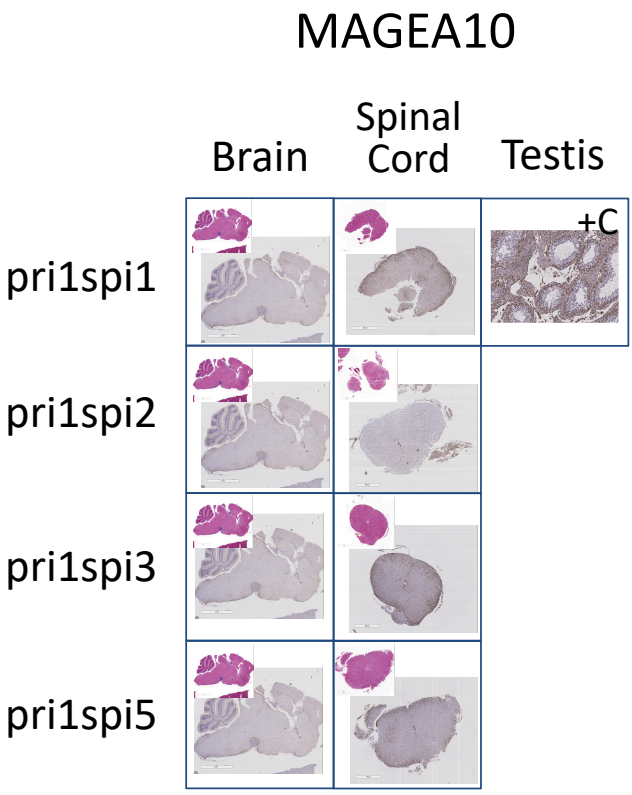


Figure 13. Immunohistochemical validation of MAGEA10 expression in xenograft primary tumours from round 1 of selection and spine metastases from rounds 1, 2, 3, and 5 of selection. Inset in the top left corner of every IHC panel: H & E staining of a section adjacent to the section that was used for IHC. Top rightmost panel: Positive control – Staining of testis with anti-MAGEA10.

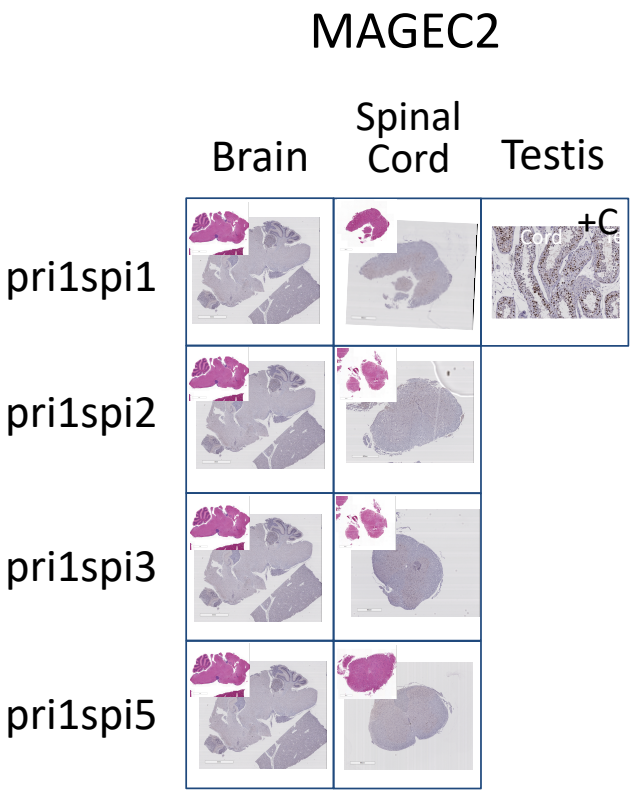


Figure 14. Immunohistochemical validation of MAGEC2 expression in xenograft primary tumours from round 1 of selection and spine metastases from rounds 1, 2, 3, and 5 of selection. Inset in the top left corner of every IHC panel: H & E staining of a section adjacent to the section that was used for IHC. Top rightmost panel: Positive control – Staining of testis with anti-MAGEC2.

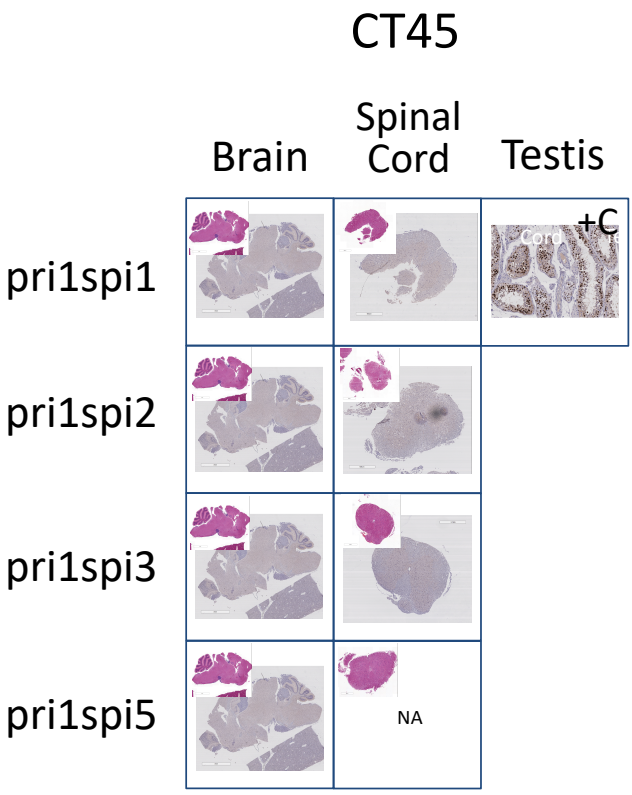


Figure 15. Immunohistochemical validation of CT45 expression in xenograft primary tumours from round 1 of selection and spine metastases from rounds 1, 2, 3, and 5 of selection. Inset in the top left corner of every IHC panel: H & E staining of a section adjacent to the section that was used for IHC. Top rightmost panel: Positive control – Staining of testis with anti-CT45. The tissue section from the spine for pri1spi5 was lost during processing.

Taken together, while PAGE1 protein expression was identified in both primary tumours and spine metastases it was still unclear whether CTA protein expression was increased in correlation with what was observed at the transcript level.

In summary, iteratively selecting for increased metastasis propensity among the spine metastases of mouse PDX models of human MB led to the increased transcription of a wide range of genes, some of which exhibited progressively increasing transcription with every round of selection, including CTA genes and lncRNAs. These transcriptional changes were also accompanied by increased activity of metastasis pathways, many also activated in the context of human MB, observed once the CTA gene family, treated as a pathway, had become activated. So iterative selection has identified specific subfamilies of CTA genes, namely *MAGE*, *XAGE*, and *CT45* gene families, as well as lncRNAs, as being potential new correlates of human Group 3 MB metastasis. With this new association, these genes are potential new targets for anti-cancer therapy.

Chapter 4: Discussion

Chapter 4: Discussion

4.1 In vivo Iterative Selection: A Powerful Approach to Define the Transcriptomics of MB Metastasis

In this work, in vivo iterative selection was used to identify genes and pathways activated in metastases to the spine borne by mouse xenograft models of human MB. As successfully applied by several other groups, repeatedly selecting for increased propensity for metastasis among spine metastases by repeated use of spine metastases as seeds for successive xenograft mice, the proportion of genes transcriptionally activated specifically in correlation with metastasis would grow with every round of selection. This was what was observed, with increasing presence of metastasis-related genes among those transcriptionally activated with every round of selection, ultimately culminating in the appearance of metastasis pathways by the fifth round of selection.

But is this approach involving repeated tumourigenesis and metastasis a model that is representative of endogenous tumourigenesis and metastasis? Indeed, tumourigenesis and metastasis are processes of ongoing clonal evolution [91, 92]. New mutations occur in cells of a tumour, and each cell is then either selected for or against depending on whether the mutation has increased or decreased its fitness within the population of cells that are making up the tumour. Similarly, if selection pressure within a tumour becomes high, mutations, which enable tumour cells to migrate away from this restrictive environment to a permissive environment (i.e. undergo metastasis), will, in that moment, become highly selectively favourable. Thus, cells with the mutation will migrate away from the primary tumour bed and form metastases. This model was first proposed in 1976 by Dr. Peter Nowell [91], then expanded upon by Dr. Isaiah Fidler in 1978 to include more details regarding evolution as it applies to metastasis [92]. From that point forward, the model has been shown to be an apt description of tumourigenesis and metastasis for a wide range of cancer types, including but not limited to colorectal cancer [93], breast cancer [94], and medulloblastoma [95]. In groundbreaking work by Dr. Sorana Morrissy in 2016, tumour evolution leading to clonal divergence has even been shown to extend to MB recurrence [96].

So, the repeated rounds of selection characteristic of in vivo iterative selection do represent a useful model for the endogenous mechanisms of mutation and selection that drive endogenous tumourigenesis and metastasis. In vivo iterative selection just makes the process of generating highly metastatic cells faster and more efficient than in endogenous tumour development [97] by conducting a form of artificial selection, explicitly selecting metastatic cells for continued selection, rather than natural selection, in which selection is dependent only on fitness.

4.2 Genes are Transcriptionally Activated in PDX Spine Metastases Relative to Primary Tumours in the First Round of Selection

Before beginning to select for increased metastasis propensity in the spine metastases of xenograft mice, an initial xenograft mouse was created by intracranially injecting cultured D425 cells that had not yet undergone metastasis. In a subset of the successfully created xenograft mice, in which injected D425 cells engrafted within the cerebellum and formed a MB primary tumour, cells migrated to the spine and formed metastases. In these cells that disseminated three transcriptionally activated genes, *BEX2*, *PPP1CC*, and *NUDT11*, had previously been linked to cancer, metastasis, or in the case of *PPP1CC*, Medulloblastoma metastasis [61]. However, these genes were in the minority, with 204 additional genes that were transcriptionally activated genes yet had no previous link to metastasis. So, it appears that transcriptional activation of some genes was in correlation with metastasis while for other genes was in correlation with a characteristic of the cells other than their propensity for metastasis.

4.3 Selecting for Increased Metastasis Increasingly Selected for Cancer Testis Antigen (CTA) Genes

Iterative selection of spine metastases from mouse PDX models of human Group 3 MB to favor those with increased propensity for metastasis led to progressively increasing transcription of cancer testis antigen (CTA) genes (predominantly from the X chromosome) and long non-coding RNAs (lncRNAs).

CTA genes vary significantly with respect to their structure, regulation, and function [98]; however, the characteristics shared by all CTA genes are transcription that is confined, both spatially and temporally, to germ tissues during periods of rapid cell division, and, although

having different functions, the ultimate shared function of promoting a stem-like state in the cells in which they are expressed.

Progressively increasing transcription across rounds of selection was observed for *MAGE*, *XAGE*, and CT45 subfamilies of CTA genes more than any others. *MAGE* and *CT45* genes in combination with other CTA genes have been found to promote epithelial-mesenchymal transition, and the production of tumour cells with stem-like properties [99]. *MAGEA* genes function as modulators of ubiquitin ligases. MAGE proteins complex with E3 RING ubiquitin ligases such as MDM2 to form MAGE-RING ubiquitin ligases; in the complexes MAGE proteins then regulate ligase activity, target specificity, and location in the cell. In the context of MB, expression of MAGE proteins inhibits chemosensitivity [100]. In contrast, overexpression of miR34a increases MB chemosensitivity by inhibiting the transcription of MAGEA genes, which leads to increased levels of P53, and its transcriptional targets [101]. This suggests that MAGEA genes regulate P53. It has been established that P53 loss in tumour cells leads to reduced cell senescence, reduced apoptosis, and malfunctioning cell checkpoints [102]. Specifically, MB germline and somatic mutations in P53 are characteristic of SHH subgroup MB and patients with P53 mutations are at very high risk (% Survival to 5y post-presentation = 50-75%) [103]. It can be further hypothesized that the functional changes associated with reduced P53 expression would support metastasis in any subgroup.

In comparison, the function of members of the CT45 CTA subfamily is poorly understood. siRNA-mediated inhibition of *CT45A1* transcription correlates with decreased cell migration, invasion, and cell viability of lung cancer cells, possibly through reduced ERK/CREB signaling, which was also downregulated as a result of the siRNA treatment. So, it appears that the regulation of metastasis by CTA genes, observed in much greater detail for other cancers than for MB, may truly also be taking place in the context of MB.

XAGE-1b and *XAGE-1d*, two of the four transcripts produced from the *XAGE-1* gene, exhibited progressively increasing transcription across selection. Increased expression of *XAGE-1b* has been shown to either correlate with metastasis in the context of melanoma [104], or promote metastasis in the context of adenoid cystic carcinoma [105]. In contrast, *XAGE-1d* has yet to be linked to metastasis but has been identified as a biomarker of squamous cell carcinoma [106].

4.4 Vimentin: Is it Acting to Support Metastasis in Association with CTA Genes?

CTA genes regulate metastases indirectly, through intermediate effectors or in association with other molecules such as the CREB pathway for *CT45A1* [72]. In this study, the CREB pathway components were not transcriptionally activated, nor were effectors *TWIST1, E-Cadherin* (*CDH1*), or *N-Cadherin* (*CDH2*). Significantly increased transcription was observed, however for effector *Vimentin* (*VIM*), one of 71 protein components that make up filaments, that support the migration of mesenchymal cells [107], and the metastasis of cancer cells [108]. Specifically, that support comes in multiple forms. Networks of Vimentin fibres associate with networks of microtubules and given the greater stability of Vimentin fibres relative to microtubules, guide the building of microtubules from exposed "plus" ends in order to maintain polarity within the cell, which is required for cell migration. In addition, Vimentin fibres interact with actin fibres directly via the Vimentin tail domain, and indirectly via a linker protein, plectin, and a regulatory protein CAMIl2, to support the formation of lamellipodia, cellular protrusions, which serve as the leading edge as cells migrate along a surface. Furthermore, Vimentin supports changes in cell shape, reductions in cell-cell contacts and focal adhesions, and supports cell elasticity, something that is especially important given the biophysical stress that is associated with the processes of migration and invasion. So, Vimentin is a critical component of the molecular machinery, which mediates cell migration in the contexts of both health and disease. Thus, its transcriptional activation in the spine metastases of PDX mice relative to primary tumours suggests that it may be participating in the process of MB tumour dissemination. It is interesting to note that in pathway analyses, *VIM* was the top gene contributing to the enrichment of the Epithelial-Mesenchymal Transition pathway. So, this further supports the hypothesis that activated *VIM* is promoting metastasis. However, the question still remains whether *VIM* might be acting upstream or downstream of CTA genes.

4.5 Enrichment of CTA Genes Comes Before Metastasis Pathways: Correlative or Causative?

To assess whether the observed changes in transcription were leading to changes in pathway activity, Gene Set Enrichment Analysis was performed. The CTA gene family, treated as a pathway, became enriched after the second round of selection (Figure 11). Subsequent to this, in

the 5th round of selection metastasis pathways also became enriched. So, this suggested that activation of the CTA gene family might have been sufficient for metastasis.

That being said, it is interesting to note that no CTA genes were included in the gene sets representing metastasis pathways. As a result, based on the data in hand the relationship between CTA gene activation and metastasis must be judged as one that is correlative.

This is supported by the existing model in which CTA genes indirectly regulate metastasis pathways rather than be a part of the pathways themselves, as is observed for *CT45A1,* which promotes metastasis via aberrant CREB signaling [72]. Also, pathways activated in human MB patients were shared with the xenograft models, yet the genes differed. Thus, an assessment of the pathways affected by mutation provided a valuable view complementing the assessment of genes. Taken together, it appears that activated CTA transcription may be sufficient for the activation of metastasis pathways in human Group 3 MB tumours.

4.6 CTA Genes Are Identified as Biomarkers not Drivers in Cancer: Evidence of Absence or Absence of Evidence?

Mutations in CTA genes have been identified as biomarkers predictive of worse prognosis in a wide range of cancer types [109]; however, CTA genes have not yet been identified as drivers, for which mutation is necessary and sufficient for tumourigenesis or metastasis. So, in what ways might CTA genes play roles in promoting metastasis?

CTA genes participate in the process of metastasis; inhibition of *CT45A1* transcription leads to decreased invasion, migration, and cell viability [72]. Rather than drive metastasis directly CTA genes might play important roles in making intracellular or extracellular conditions permissive to tumour formation, growth, or spread. For example, *MAGEA* genes regulate TP53; by reducing TP53 levels and inhibiting apoptosis. Under these conditions, in the presence of other driver mutations tumours would be more easily able to survive, grow, and spread.

It is also possible that CTA genes might not be metastasis driver genes, but instead effectors of metastasis upstream or downstream of the genes driving metastasis. In such an instance, increased transcription of CTA genes would correlate with metastasis but mutations would only ever be identified in the driver genes themselves.

It is also possible that driver mutations in CTA genes have not yet been identified because they occur after cells have spread and are thus only found in metastases in patients. For MB, little metastatic tumour tissue is available for research purposes; this is due to the fact that the treatment regimen for patients with metastatic MB is chemotherapy and radiation, and so metastatic tumours are only rarely ever resected. Further, until recently, the genomes of metastases were assumed to be identical to that of the primary tumour, and so, it was believed that there would be no need to also sequence metastases. So, it is possible that CTA mutations mediating metastasis are have evaded detection by remaining confined to just metastases.

Long non-coding RNAs (lncRNAs) could be regulators of CTA genes. The gene most significantly activated across rounds of selection was *XIST*, a lncRNA. Furthermore, other lncRNAs exhibited increasing transcription similar to that of *XIST* and CTA genes. So, this leads to a model in which lncRNAs might serve as regulatory RNAs promoting metastasis through downstream effectors including CTA genes.

4.7 Model: lncRNAs Bind and Sequester miRNAs Away From Targets, Including CTA Genes, Leading to Target Derepression and Increased Metastasis

Among several functional modes, lncRNAs have been shown to serve as endogenous micro-RNA inhibitors (Figure 16) [110]. LncRNAs bind to inhibitory micro-RNAs (miRNAs), sequestering the miRNAs away from their targets. As a result, target expression is no longer inhibited, and thus increases significantly leading to disease.

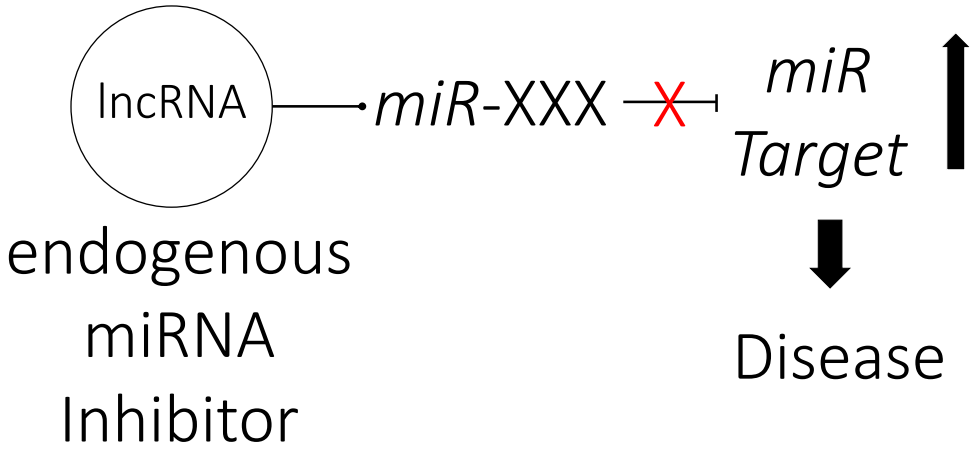


Figure 16. LncRNAs function as endogenous miRNA inhibitors in a variety of disease contexts. LncRNAs bind to specific miRNAs, and sequester them away from their intended targets, leading to target overexpression.

Since being first hypothesized and subsequently identified in 2011 [110], this mode of function for lncRNAs as miRNA sponges has been studied in depth [111]. This has led to this mode being identified for a wide range of lncRNAs, across a wide range of cancer contexts including most recently glioblastoma [112], hepatocellular carcinoma [113], thyroid cancer [114], metastatic lung cancer produced through aberrant TGF-beta signaling [115], and a variety of hematological cancers [116]. This regulatory mode has also been implicated in healthy mammalian brain development [117].

Furthermore, there is a precedent for a lncRNA serving as a miRNA sponge in the context of MB. *Linc-NeD125* binds to miRNAs miR-19a-3p, miR-19b-3p, and miR-106a-5p, which leads to the de-repression of their targets, the mRNAs encoding well-established drivers of Group 4 MB tumourigenesis CDK6, MYCN, Rcmb05, and KDM6A [118]. With their inhibitory miRNAs bound by *Linc-NeD125*, these proteins can be translated at high levels.

Of similar relevance to this study, *XIST,* the lncRNA which exhibited the greatest increase in transcription across rounds of selection, has been found to serve as a miRNA sponge in a wide variety of cancer contexts. In colorectal cancer *XIST* regulates metastasis by serving as a miRNA sponge specifically for target *ZEB1* [119]. Increased *XIST* transcription correlated with worse prognosis, and *XIST* knockdown reduced cell proliferation, invasion, and epithelial-mesenchymal transition in vitro, and reduced tumour growth and metastasis in vivo. Expression of *ZEB1* did not increase in correlation with increases in *XIST* in the current study, but that does not eliminate the possibility that *XIST* might have been regulating other transcription factors in the context of MB. One such transcription factor could be ZEB family member *ZEB2*, which was significantly activated in round 5 of selection. Furthermore, published evidence demonstrates that, in the context of non-small-cell lung cancer, *XIST* regulates epithelial-mesenchymal transition mediated by TGF-beta signaling, itself known to be upregulated in Group 3 MB [120], through a complex regulatory network including *ZEB2* and miRNAs miR-367 and miR-141 [121]. Complementing these findings and making them especially relevant to the current study, overexpression of miR-367 in MB cell lines grown in culture promotes cell proliferation and stem-like properties [122]. Also, resembling the downstream portion of this mechanism, published evidence demonstrates that miR-34a inhibits translation from MAGEA mRNAs in MB cells, and increased miR-34a expression leads to decreased MAGEA protein levels, increased P53 protein levels, and increased chemosensitivity [101].

XIST has also been shown to function as an endogenous miRNA inhibitor in a variety of other cancers, including lung adenocarcinoma [123], laryngeal squamous cell carcinoma [124], and nasopharyngeal carcinoma [125].

So, in the context of this study and the model proposed, aberrantly increased transcription of lncRNAs could lead to aberrantly increased sequestering of inhibitory miRNAs, leading to the aberrantly increased transcription of their targets, such as CTA genes ultimately leading to disease (Figure 17).

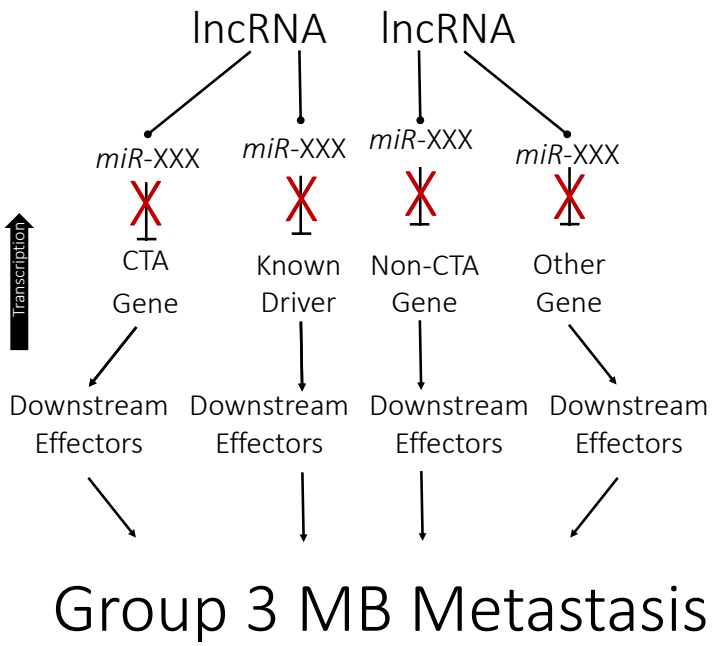


Figure 17. Graphical description of lncRNA-mediated dysregulation of transcription through the competitive inhibition of miRNAs. Presented here is a systems description of the proposed model including multiple lncRNAs, miRNAs, miRNA targets, and downstream effectors.

Taken together lncRNAs and CTA mRNAs have been identified that, at minimum, exhibit increased transcription in correlation with metastasis but may, in fact, be regulating metastasis. Increased lncRNA transcription could lead to excessive sequestering of inhibitory miRNAs away from their regular targets. This ultimately would lead to those targets, such as CTA genes, being also transcribed in excess. This further transcriptional activation could then be passed on further to downstream effector genes, like *VIM*, that ultimately drive metastasis.

It will now be necessary to test whether the preliminary patterns in this study do represent genuine regulatory relationships between lncRNAs, miRNAs, and mRNAs including all those for CTA genes. This will need to be evaluated in multiple cellular contexts namely in cerebellar precursor cells during development, in mature healthy cerebellum, in Medulloblastoma primary tumours, and in Medulloblastoma metastases. This would require the profiling of transcription for each of these categories. Clustering and differential expression analyses could then be used to specifically identify all lncRNAs and mRNAs that exhibit similar patterns of expression across categories, which would be suggestive of those genes being in a regulatory relationship. An ideal first model system to study would be the regulatory axis including miR-34a, *MAGEA* mRNA, and P53 [101]. In this system, no lncRNA regulating miR-34a has yet been identified. But the authors did report the existence of cell lines with distinctly low miR-34a transcription suggesting that some form of upstream inhibition was taking place.

With respect to the goal of this study to discover new, more effective, and safer ways to treat metastatic medulloblastoma, the results reported are especially valuable because CTA genes represent unusually suitable candidate targets for immunotherapy given their highly restricted pattern of endogenous expression. So, with a better understanding of complex regulatory networks, which may be defining when, where, and to what extent CTA genes are transcribed, it may offer access to a wide range of new, and highly immunogenic targets for study as potential new treatments for metastatic Group 3 MB.

In summary, selection for metastases to the spine in mouse PDX models of human Group 3 MB led to progressive increases in the transcription of lncRNAs and CTA genes. This ultimately resulted in the increased activity of several fundamental metastasis pathways. So, a regulatory network of lncRNAs may play an important role in mediating Group 3 MB metastasis to the spine.

Chapter 5: Conclusions

Chapter 5: Conclusions

5.1 Summary

Through my years of doctoral study, I drew on the power of bioinformatics approaches to profile the transcriptomes of primary MB tumours, and MB metastases to the spine from mouse PDX models of human MB iteratively selected to favor increased metastasis to the spine. The ultimate goal of the work was both to deepen the understanding within the scientific community of the mechanisms driving metastasis to the spine, and to identify new possible therapeutic targets for the safe and effective treatment of metastatic MB. Through my work, for the first time ever to this level, I demonstrated that transcription of many members of the Cancer Testis Antigen Gene family are correlated with metastasis. Furthermore, I identified several long non-coding RNAs, most significantly *XIST*, for which their levels of transcription are also correlated with metastasis, in a manner very similar to that observed for the CTA genes.

This has led me to ask an interesting scientific question; could these and other regulatory RNAs be regulating the transcription of a network of genes that, when functioning properly, delicately guide the development of complex multicellular organisms, but when dysregulated drive tumourigenesis and metastasis? My data seems to suggest that this might be the case since a subset of lncRNAs and CTA genes are transcriptionally activated at discrete stages in the process of in vivo iterative selection, while others exhibit progressively increasing transcription spanning all of the rounds of selection. Furthermore, extensive evidence has been, and continues to be published, describing lncRNAs, which are serving as endogenous miRNA inhibitors, sequestering inhibitory miRNAs away from their targets and thereby causing cancer and disease. So, networks of RNAs may be playing an important part in regulating the metastasis of MB in a subgroup-specific fashion. If this is the case, it will be essential for networks such as these to be characterized in detail, in order to gain a comprehensive understanding of how metastasis occurs, and by extension how metastatic disease could be most effectively treated, within each MB subgroup and subtype.

5.2 Future Directions

A powerful first step towards characterizing networks of RNAs would be to cluster all mRNA, miRNA, and lncRNA based upon their levels of transcription in MB cells of origin, primary tumours, and metastases. Complex regulatory networks linking lncRNAs, miRNAs, and mRNAs could be delineated on a systemic level, rather than on a gene by gene basis.

It would be of equal importance, beyond transcriptional profiling, to validate that transcriptional activation of specific genes is occurring specifically in correlation with metastasis. This would require the inhibition and forced overexpression of target genes to demonstrate that expression of that gene was both necessary and sufficient for metastasis to occur.

Both overexpression and inhibition could be achieved using CRISPR, a technology for the sequence-specific editing of DNA. In traditional CRISPR-Cas9, Cas9, a sequence-specific endonuclease, binds to synthesized guide RNAs, with sequence complementary to the site where DNA is to be edited. Cas9 then migrates to the site of interest using the guide RNA and cuts at the sequence specified enabling the insertion of new DNA, or repair by the cell machinery. Traditional CRISPR-Cas9 could be used for inhibition. However, forced expression would require the use of a modified version of CRIPSR known as CRISPR-dCas9-VPR. In this form, a dCas9-VPR is used, in which the Cas9 endonuclease function is inactivated, and fused to the active domains of three other transcriptional activators, Vp64, p65, and Rta. This modified dCas9-VPR can then be recruited to a gene of interest based on the guide RNA, and bring about its overexpression.

In our approach, cultured tumour cells would be infected with lentivirus carrying the coding sequence for dCas9-VPR and lentivirus carrying the coding sequence for either Blue Fluorescent Protein labeled case guide RNAs or Red mCherry-labeled scrambled guide RNAs (Figure 18). Lentiviral cargo would be incorporated into the genome of the cultured cells, and transcribed leading to the forced overexpression of the gene of interest by dCas9-VPR in cells with the case guide RNA. Both populations of cells (now Blue and Red) would be mixed and intracranially injected into mice. Spine metastases significantly more blue than red would indicate that forced overexpression was sufficient to promote metastasis. Conversely, metastases with equal parts blue and red would indicate that forced overexpression was not sufficient to promote metastasis. Finally, metastases significantly more red than blue, or entirely red would indicate that forced

overexpression was not only insufficient but actually detrimental to metastasis. It should be noted that this experiment is actually in progress, being carried out by Taylor lab collaborators Sachin Kumar, Paquito de Antonellis, Kyle Juraschka, and Olga Sirbu, with results imminent.

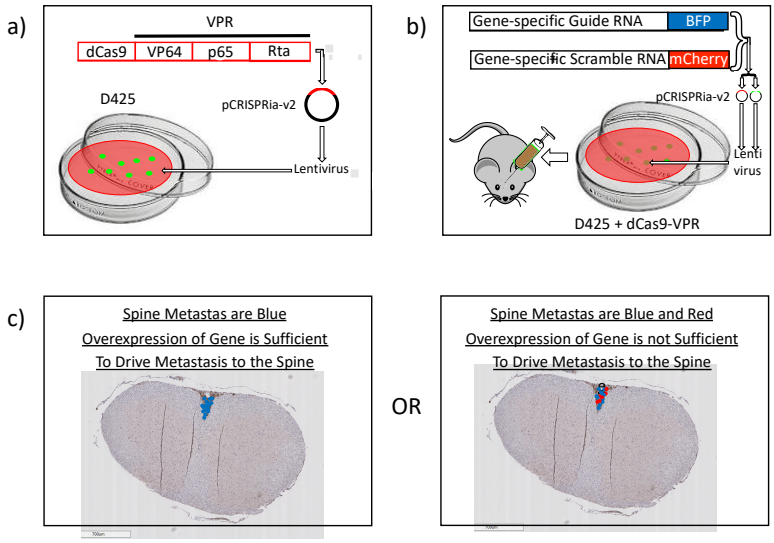


Figure 18. CRISPR-dCas9-VPR-mediated overexpression of genes of interest in D425 Group 3 MB cells to validate whether overexpression of a gene of interest confers an increased propensity for metastasis upon the D425 cells in which the gene is being overexpressed. a) Introduction of Cas9-VPR into GFP-labeled D425 cells through lentiviral infection b) Introduction of BFP-labeled gene-specific guide RNAs or mCherry-labeled scrambled guide RNAs into two populations of modified D425 cells through lentiviral infection followed by intracranial injection of mixed cells into NSG mice c) Blue metastases would indicate gene overexpression promotes metastasis. Metastases of mixed colour would indicate that gene overexpression promotes metastasis only weakly or not at all. Red metastases would indicate that gene overexpression inhibits metastasis (not shown).

There are also other mysteries begging further investigation. CTA genes have been readily identified as cancer biomarkers and even effective drug targets for immunotherapy but have yet to ever be implicated as drivers, necessary for tumourigenesis or metastasis. They have frequently been shown to promote increased rates of metastasis when overexpressed but have yet to be identified as a causative driver. A deeper understanding of the exact roles that are being played by CTA genes in metastasis would serve as the best means to identify CTA genes that would be most useful as therapeutic targets.

So, what appears to be the case, as seems to be the case so often within the scientific community, is that as much as we do now know and understand about MB tumourigenesis and metastasis, we have only scratched the surface and still have much to learn.

Bibliography

1. de Medeiros, C.B., et al., *Medulloblastoma has a global impact on health related quality of life: Findings from an international cohort.* Cancer Med, 2020. **9**(2): p. 447-459.

2. Parkes, J., et al., *SIOP PODC adapted treatment recommendations for standard-risk medulloblastoma in low and middle income settings.* Pediatr Blood Cancer, 2015. **62**(4): p. 553-64.

3. Taylor, M.D., et al., *Molecular subgroups of medulloblastoma: the current consensus.* Acta Neuropathol, 2012. **123**(4): p. 465-72.

4. Cavalli, F.M.G., et al., *Intertumoral Heterogeneity within Medulloblastoma Subgroups.* Cancer Cell, 2017. **31**(6): p. 737-754 e6.

5. Van Ommeren, R., et al., *The molecular biology of medulloblastoma metastasis.* Brain Pathol, 2020. **30**(3): p. 691-702.

6. Ramaswamy, V., et al., *Duration of the pre-diagnostic interval in medulloblastoma is subgroup dependent.* Pediatr Blood Cancer, 2014. **61**(7): p. 1190-4.

7. Northcott, P.A., et al., *Subgroup-specific structural variation across 1,000 medulloblastoma genomes.* Nature, 2012. **488**(7409): p. 49-56.

8. Juraschka, K. and M.D. Taylor, *Medulloblastoma in the age of molecular subgroups: a review.* J Neurosurg Pediatr, 2019. **24**(4): p. 353-363.

9. Dressler, E.V., et al., *Demographics, patterns of care, and survival in pediatric medulloblastoma.* J Neurooncol, 2017. **132**(3): p. 497-506.

10. White, et al., *A Sexually Dimorphic Role for STAT3 in Sonic Hedgehog Medulloblastoma.* Cancers, 2019. **11**(11).

11. Sreenivasan, L., et al., *Autocrine IL-6/STAT3 signaling aids development of acquired drug resistance in Group 3 medulloblastoma.* Cell Death Dis, 2020. **11**(12): p. 1035.

12. Vladoiu, M.C., et al., *Childhood cerebellar tumours mirror conserved fetal transcriptional programs.* Nature, 2019. **572**(7767): p. 67-73.

13. Jessa, S., et al., *Stalled developmental programs at the root of pediatric brain tumors.* Nat Genet, 2019. **51**(12): p. 1702-1713.

14. Northcott, P.A., et al., *Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma.* Nature, 2014. **511**(7510): p. 428-34.

15. Dubuc, A.M., et al., *Aberrant patterns of H3K4 and H3K27 histone lysine methylation occur across subgroups in medulloblastoma.* Acta Neuropathol, 2013. **125**(3): p. 373-84.

16. Robinson, M.H., et al., *Upregulation of the chromatin remodeler HELLS is mediated by YAP1 in Sonic Hedgehog Medulloblastoma.* Sci Rep, 2019. **9**(1): p. 13611.

17. Gonzalez-Gomez, P., et al., *Deletion and aberrant CpG island methylation of Caspase 8 gene in medulloblastoma.* . Oncology Reports, 2004. **12**(3): p. 663-6.

18. Menyhart, O. and B. Gyorffy, *Molecular stratifications, biomarker candidates and new therapeutic options in current medulloblastoma treatment approaches.* Cancer Metastasis Rev, 2020. **39**(1): p. 211-233.

19. Michalski, J., et al., *Mb-109preliminary Results of Cog Acns0331: A Phase Iii Trial of Involved Field Radiotherapy (Ifrt) and Low Dose Craniospinal Irradiation (Ld-Csi) with Chemotherapy in Average Risk Medulloblastoma: A Report from the Children's Oncology Group.* Neuro-Oncology, 2016. **18**(suppl 3): p. iii122-iii122.

20. von Bueren, A.O., et al., *Treatment of Children and Adolescents With Metastatic Medulloblastoma and Prognostic Relevance of Clinical and Biologic Parameters.* J Clin Oncol, 2016. **34**(34): p. 4151-4160.

21. Phoenix, T.N., et al., *Medulloblastoma Genotype Dictates Blood Brain Barrier Phenotype.* Cancer Cell, 2016. **29**(4): p. 508-522.

22. Houschyar, K.S., et al., *Wnt Pathway in Bone Repair and Regeneration - What Do We Know So Far.* Front Cell Dev Biol, 2018. **6**: p. 170.

23. Svard, J., et al., *Genetic elimination of Suppressor of fused reveals an essential repressor function in the mammalian Hedgehog signaling pathway.* Dev Cell, 2006. **10**(2): p. 187-97.

24. Rudin, C.M., et al., *Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449.* . N Engl J Med. , 2009. **361**(12): p. 1173-8.

25. Robinson, G.W., et al., *Vismodegib Exerts Targeted Efficacy Against Recurrent Sonic Hedgehog-Subgroup Medulloblastoma: Results From Phase II Pediatric Brain Tumor Consortium Studies PBTC-025B and PBTC-032.* . J Clin Oncol. , 2015 **33**(24): p. 2646-54.

26. Menyhart, O. and B. Gyorffy, *Principles of tumorigenesis and emerging molecular drivers of SHH-activated medulloblastomas.* Ann Clin Transl Neurol, 2019. **6**(5): p. 990-1005.

27. Menyhart, O., F. Giangaspero, and B. Gyorffy, *Molecular markers and potential therapeutic targets in non-WNT/non-SHH (group 3 and group 4) medulloblastomas.* J Hematol Oncol, 2019. **12**(1): p. 29.

28. Martinez, M. and E.K. Moon, *CAR T Cells for Solid Tumors: New Strategies for Finding, Infiltrating, and Surviving in the Tumor Microenvironment.* Front Immunol, 2019. **10**: p. 128.

29. Nellan, A., et al., *Durable regression of Medulloblastoma after regional and intravenous delivery of anti-HER2 chimeric antigen receptor T cells.* J Immunother Cancer, 2018. **6**(1): p. 30.

30. Perez-Martinez, A., L. Fernandez, and M.A. Diaz, *The therapeutic potential of natural killer cells to target medulloblastoma.* Expert Rev Anticancer Ther, 2016. **16**(6): p. 573-6.

31. Powell, A.B., et al., *Medulloblastoma rendered susceptible to NK-cell attack by TGFbeta neutralization.* J Transl Med, 2019. **17**(1): p. 321.

32. Kumar, V., et al., *Challenges and Recent Advances in Medulloblastoma Therapy.* Trends Pharmacol Sci, 2017. **38**(12): p. 1061-1084.

33. Gao, H., et al., *Angiopep-2 and activatable cell-penetrating peptide dual-functionalized nanoparticles for systemic glioma-targeting delivery.* Mol Pharm, 2014. **11**(8): p. 2755-63.

34. Catanzaro, G., et al., *Albumin nanoparticles for glutathione-responsive release of cisplatin: New opportunities for medulloblastoma.* Int J Pharm, 2017. **517**(1-2): p. 168-174.

35. Rochkind, S., et al., *Extracranial metastases of medulloblastoma in adults: literature review.* . J Neurol Neurosurg Psychiatry., 1991. **54**(1): p. 80-86.

36. Kahn, S.A., et al., *Notch1 regulates the initiation of metastasis and self-renewal of Group 3 medulloblastoma.* Nat Commun, 2018. **9**(1): p. 4121.

37. Lessard, J. and G. Sauvageau, *Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells.* . Nature, 2003. **423**(6937): p. 255-60.

38. Zhou, L., et al., *Silencing of thrombospondin-1 is critical for myc-induced metastatic phenotypes in medulloblastoma.* Cancer Res, 2010. **70**(20): p. 8199-210.

39. Gao, R., et al., *LncRNA LOXL1-AS1 Promotes the Proliferation and Metastasis of Medulloblastoma by Activating the PI3K/AKT Pathway.* Anal Cell Pathol (Amst), 2018. **2018**: p. 9275685.

40. Hidalgo, M., et al., *Patient-derived xenograft models: an emerging platform for translational cancer research.* Cancer Discov, 2014. **4**(9): p. 998-1013.

41. *NSG(TM) Models*. 2019.

42. Fidler, I.J., *Selection of successive tumor lines for metastasis.* Nature new Biology, 1973. **242**(118): p. 148-9.

43. Clark, E.A., et al., *Genomic analysis of metastasis reveals an essential role for RhoC.* Nature, 2000. **406**(6795): p. 532-5.

44. Boire, A., et al., *Complement Component 3 Adapts the Cerebrospinal Fluid for Leptomeningeal Metastasis.* Cell, 2017. **168**(6): p. 1101-1113 e13.

45. Brenner, S., et al., *Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.* Nature Biotechnology, 2000. **18**(6): p. 630-4.

46. Bainbridge, M.N., et al., *Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach.* BMC Genomics, 2006. **7**: p. 246.

47. Soneson, C. and M. Delorenzi, *A comparison of methods for differential expression analysis of RNA-seq data. .* BMC Bioinformatics., 2013. **14**: p. 91.

48. Khatri, P., M. Sirota, and A.J. Butte, *Ten years of pathway analysis: current approaches and outstanding challenges.* PLoS Comput Biol, 2012. **8**(2): p. e1002375.

49. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.

50. Fang, H., et al., *Signaling pathway impact analysis by incorporating the importance and specificity of genes (SPIA-IS).* Comput Biol Chem, 2017. **71**: p. 236-244.

51. Bigner, S.H., et al., *Amplification of the c-myc Gene in Human Medulloblastoma Cell Lines and Xenografts.* Cancer Research, 1990. **50**: p. 2347-50.

52. Leng, N., et al., *EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments.* Bioinformatics, 2013. **29**(8): p. 1035-43.

53. *FastQC website* Available from: http://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc.

54. Ensembl. *GRCh37 (Ensembl Release 75)*. 2014; Available from: http://grch37.ensembl.org/index.html.

55. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

56. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.

57. Kauffmann, A., R. Gentleman, and W. Huber, *arrayQualityMetrics--a bioconductor package for quality assessment of microarray data.* Bioinformatics, 2009. **25**(3): p. 415-6.

58. Phipson, B., ***ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION*** S. Lee, Editor. 2016. p. 946-963.

59. *Ensembl GRCh37 record for NUDT11*. Available from: http://grch37.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000196368;r=X:51232863-51239448;t=ENST00000375992.

60. Grisanzio, C., et al., *Genetic and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate cancer pathogenesis.* Proc Natl Acad Sci U S A, 2012. **109**(28): p. 11252-7.

61. Traenka, C., et al., *Role of LIM and SH3 protein 1 (LASP1) in the metastatic dissemination of medulloblastoma.* Cancer Res, 2010. **70**(20): p. 8003-14.

62. Liu, X., et al., *PPP1CC is associated with astrocyte and microglia proliferation after traumatic spinal cord injury in rats.* Pathol Res Pract, 2017. **213**(11): p. 1355-1364.

63. Chung, H.J., et al., *p53-Mediated enhancement of radiosensitivity by selenophosphate synthetase 1 overexpression.* J Cell Physiol, 2006. **209**(1): p. 131-41.

64. Garg, N., et al., *CD133(+) brain tumor-initiating cells are dependent on STAT3 signaling to drive medulloblastoma recurrence.* Oncogene, 2017. **36**(5): p. 606-617.

65. Boire, A., *Co**mplement Component 3 Adapts the Cerebrospinal Fluid for Leptomeningeal Metastasis***, Y. Zou, Editor. 2017. p. 1101-1113.

66. Holcomb, R.J., et al., *The testis-specific serine proteases PRSS44, PRSS46, and PRSS54 are dispensable for male mouse fertilitydagger.* Biol Reprod, 2020. **102**(1): p. 84-91.

67. Heinemeyer, T., et al., *Underappreciated Roles of the Translocase of the Outer and Inner Mitochondrial Membrane Protein Complexes in Human Disease.* DNA Cell Biol, 2019. **38**(1): p. 23-40.

68. Al Teneiji, A., et al., *Progressive Cerebellar Atrophy and a Novel Homozygous Pathogenic DNAJC19 Variant as a Cause of Dilated Cardiomyopathy Ataxia Syndrome.* Paediatric Neurology, 2016. **62**: p. 58-61.

69. Tan, Y., et al., *Silencing of brain-expressed X-linked 2 (BEX2) promotes colorectal cancer metastasis through the Hedgehog signaling pathway.* Int J Biol Sci, 2020. **16**(2): p. 228-238.

70. Bloom, J.E. and D.G. McNeel, *SSX2 regulates focal adhesion but does not drive the epithelial to mesenchymal transition in prostate cancer. .* Oncotarget, 2016. **7**(32): p. 50997-51011.

71. Brockdorff, N., *Localized accumulation of Xist RNA in X chromosome inactivation.* Open Biol, 2019. **9**(12): p. 190213.

72. Tang, F., et al., *CT45A1 siRNA silencing suppresses the proliferation, metastasis and invasion of lung cancer cells by downregulating the ERK/CREB signaling pathway.* Mol Med Rep, 2017. **16**(5): p. 6708-6714.

73. Salmaninejad, A., et al., *Cancer/Testis Antigens: Expression, Regulation, Tumor Invasion, and Use in Immunotherapy of Cancers.* Immunol Invest, 2016. **45**(7): p. 619-40.

74. Ross, M.T., et al., *The DNA sequence of the human X chromosome.* Nature, 2005. **434**(7031): p. 325-37.

75. Grigoriadis, A., et al., *CT-X antigen expression in human breast cancer.* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13493-8.

76. Andrade, V.C., et al., *Prognostic impact of cancer/testis antigen expression in advanced stage multiple myeloma patients. .* Cancer Immun. , 2008. **8**(2): p. 1-8.

77. Gure, A.O., et al., *Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer.* Clin Cancer Res, 2005. **11**(22): p. 8055-62.

78. Bredenbeck, A., et al., *Coordinated expression of clustered cancer/testis genes encoded in a large inverted repeat DNA structure.* Gene, 2008. **415**(1-2): p. 68-73.

79. Dickinson, R.E., M. Myers, and W.C. Duncan, *Novel regulated expression of the SLIT/ROBO pathway in the ovary: possible role during luteolysis in women.* Endocrinology, 2008. **149**(10): p. 5024-34.

80. Gu, F., et al., *Function of Slit/Robo signaling in breast cancer.* Front Med, 2015. **9**(4): p. 431-6.

81. Skuja, E., et al., *Deletions in metastatic colorectal cancer with chromothripsis.* Exp Oncol, 2019. **41**(4): p. 323-327.

82. Felici, R., et al., *Insight into molecular and functional properties of NMNAT3 reveals new hints of NAD homeostasis within human mitochondria.* PLoS One, 2013. **8**(10): p. e76938.

83. Gulshan, M., et al., *Overexpression of Nmnat3 efficiently increases NAD and NGD levels and ameliorates age-associated insulin resistance.* Aging Cell, 2018. **17**(4): p. e12798.

84. Glasner, A., et al., *NKp46 Receptor-Mediated Interferon-gamma Production by Natural Killer Cells Increases Fibronectin 1 to Alter Tumor Architecture and Control Metastasis.* Immunity, 2018. **48**(2): p. 396-398.

85. Lollini, P.L., et al., *Inhibition of tumor growth and enhancement of metastasis after transfection of the gamma-interferon gene. .* Int J Cancer. , 1993 **55**(2): p. 320-9.

86. Acharyya, S., et al., *A CXCL1 paracrine network links cancer chemoresistance and metastasis.* Cell, 2012. **150**(1): p. 165-78.

87. Zhang, Z., et al., *Ponicidin inhibits pro-inflammatory cytokine TNF-alpha-induced epithelial-mesenchymal transition and metastasis of colorectal cancer cells via suppressing the AKT/GSK-3beta/Snail pathway.* Inflammopharmacology, 2019. **27**(3): p. 627-638.

88. Smith, B.N. and N.A. Bhowmick, *Role of EMT in Metastasis and Therapy Resistance.* J Clin Med, 2016. **5**(2).

89. Sung, B.H. and A.M. Weaver, *Exosome secretion promotes chemotaxis of cancer cells.* Cell Adh Migr, 2017. **11**(2): p. 187-195.

90. Coussens, L.M. and Z. Werb, *Inflammation and cancer.* Nature, 2002. **420**(6917): p. 860-7.

91. Nowell, P.C., *The clonal evolution of tumor cell populations.* Science, 1976. **194**(4260): p. 23-8.

92. I.J., F., *Tumor heterogeneity and the biology of cancer invasion and metastasis. .* Cancer Res. , 1978. **38**(9): p. 2651-60.

93. Vogelstein, B., et al., *Genetic alterations during colorectal-tumor development. .* N Engl J Med. , 1988. **319**(9): p. 525-32. .

94. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. **472**(7341): p. 90-4.

95. Wu, X., et al., *Clonal selection drives genetic divergence of metastatic medulloblastoma.* Nature, 2012. **482**(7386): p. 529-33.

96. Morrissy, A.S., et al., *Divergent clonal selection dominates medulloblastoma at recurrence.* Nature, 2016. **529**(7586): p. 351-7.

97. Massague, J. and A.C. Obenauf, *Metastatic colonization by circulating tumour cells.* Nature, 2016. **529**(7586): p. 298-306.

98. Salmaninejad, A., M. Reza Zamani, and M.G.f. Pourvahedi e, Z. Bereshneh f, A. H/ Rezaei , N., *Cancer/Testis Antigens Expression, Regulation, TumorInvasion, and Use in Immunotherapy of Cancers.* Immunological Investigations, 2016. **45**(7): p. 619-40.

99. Yang, P., et al., *Cancer/testis antigens trigger epithelial-mesenchymal transition and genesis of cancer stem-like cells.* Current Pharmaceutical Design, 2015. **21**(10): p. 1292-300.

100. Kasuga, C., et al., *Expression of MAGE and GAGE genes in medulloblastoma and modulation of resistance to chemotherapy. Laboratory investigation.* J Neurosurg Pediatr, 2008. **1**(4): p. 305-13.

101. Weeraratne, S.D., et al., *miR-34a confers chemosensitivity through modulation of MAGE-A and p53 in medulloblastoma.* Neuro Oncol, 2011. **13**(2): p. 165-75.

102. Powell, E., D. Piwnica-Worms, and H. Piwnica-Worms, *Contribution of p53 to metastasis.* Cancer Discov, 2014. **4**(4): p. 405-14.

103. Ramaswamy, V., et al., *Risk stratification of childhood medulloblastoma in the molecular era: the current consensus.* Acta Neuropathol, 2016. **131**(6): p. 821-31.

104. Zendman, A.J.W., et al., *Characterization of XAGE-1b, a short major transcript of cancer/testis-associated gene XAGE-1, induced in melanoma metastasis.* International Journal of Cancer 2002. **97**(2): p. 195-204.

105. Zhou, B., et al., *Promoting effects on the proliferation and metastasis of ACC tumor cell with XAGE-1b overexpression.* Oncol Rep, 2013. **30**(5): p. 2323-35.

106. Ha, J.S., et al., *XAGE-1a and XAGE-1d are potential biomarkers of lung squamous cell carcinoma.* Clin Chim Acta, 2012. **413**(15-16): p. 1226-31.

107. Battaglia, R.A., et al., *Vimentin on the move: new developments in cell migration.* F1000Res, 2018. **7**.

108. Sharma, P., et al., *Intermediate Filaments as Effectors of Cancer Development and Metastasis: A Focus on Keratins, Vimentin, and Nestin.* Cells, 2019. **8**(5).

109. Sammut, S.J., et al., *A novel cohort of cancer-testis biomarker genes revealed through meta-analysis of clinical data sets.* Oncoscience, 2014. **1**(5): p. 349-59.

110. Salmena, L., et al., *A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?* Cell, 2011. **146**(3): p. 353-8.

111. Bak, R.O. and J.G. Mikkelsen, *miRNA sponges: soaking up miRNAs for regulation of gene expression.* Wiley Interdiscip Rev RNA, 2014. **5**(3): p. 317-33.

112. Baspinar, Y., et al., *Long non-coding RNA MALAT1 as a key target in pathogenesis of glioblastoma. Janus faces or Achilles' heal?* Gene, 2020. **739**: p. 144518.

113. Li, D., J. Zhang, and J. Li, *Role of miRNA sponges in hepatocellular carcinoma.* Clin Chim Acta, 2020. **500**: p. 10-19.

114. Ghafouri-Fard, S., H. Mohammad-Rahimi, and M. Taheri, *The role of long non-coding RNAs in the pathogenesis of thyroid cancer.* Exp Mol Pathol, 2020. **112**: p. 104332.

115. Lai, X.N., et al., *MiRNAs and LncRNAs: Dual Roles in TGF-beta Signaling-Regulated Metastasis in Lung Cancer.* Int J Mol Sci, 2020. **21**(4).

116. Ghafouri-Fard, S., M. Esmaeili, and M. Taheri, *Expression of non-coding RNAs in hematological malignancies.* Eur J Pharmacol, 2020. **875**: p. 172976.

117. Ang, C.E., A.E. Trevino, and H.Y. Chang, *Diverse lncRNA mechanisms in brain development and disease.* Current Opinions in Genetics and Development, 2020. **65**: p. 42-6.

118. Laneve, P., et al., *The long noncoding RNA linc-NeD125 controls the expression of medulloblastoma driver genes by microRNA sponge activity.* Oncotarget, 2017. **8**(19): p. 31003-15.

119. Chen, D.L., et al., *Long noncoding RNA XIST expedites metastasis and modulates epithelial-mesenchymal transition in colorectal cancer.* Cell Death Dis, 2017. **8**(8): p. e3011.

120. Ferrucci, V., et al., *Metastatic group 3 medulloblastoma is driven by PRUNE1 targeting NME1-TGF-beta-OTX2-SNAIL via PTEN inhibition.* Brain, 2018. **141**(5): p. 1300-1319.

121. Li, C., et al., *Long non-coding RNA XIST promotes TGF-beta-induced epithelial-mesenchymal transition by regulating miR-367/141-ZEB2 axis in non-small-cell lung cancer.* Cancer Lett, 2018. **418**: p. 185-195.

122. Kaid, C., et al., *miR-367 promotes proliferation and stem-like traits in medulloblastoma cells.* Cancer Sci, 2015. **106**(9): p. 1188-95.

123. Rong, H., et al., *Long non-coding RNA XIST expedites lung adenocarcinoma progression through upregulating MDM2 expression via binding to miR-363-3p.* Thorac Cancer, 2020.

124. Cui, C.L., et al., *lncRNA XIST promotes the progression of laryngeal squamous cell carcinoma by sponging miR144 to regulate IRS1 expression.* Oncol Rep, 2020. **43**(2): p. 525-535.

125. Shi, J., et al., *LncRNA XIST knockdown suppresses the malignancy of human nasopharyngeal carcinoma through XIST/miRNA-148a-3p/ADAM17 pathway in vitro and in vivo.* Biomed Pharmacother, 2020. **121**: p. 109620.

Appendix 1: Pairwise Pearson Correlations Between All RNA-Seq Datasets (12 pages)

Appendix 2: Principal Component Analysis (PCA) (35 pages)

Appendix 3: Transcriptionally Activated Genes Unique to D425, MB411FH, and Rcmb06 (24 pages)

Appendix 4: Summary Table of Genes Transcriptionally Activated in D425 Spine Metastases Relative to Primary Tumours Across Rounds of In vivo Iterative Selection (25 pages)

Appendix 5: Summary Table of Genes With Progressively Increasing Transcription in D425 Spine Metastases Relative to Priimary Tumours Across All Rounds of Selection (28 pages)

Appendix 6: Summary Table of Expression Data for All Long Non-coding RNAs (lncRNAs) From D425 PDX Mice Across All Rounds of Selection (8 pages)

Appendix 7: Summary Table of Expression Data and Chromosome Localization for All CTA Genes From D425 PDX Mice Across All Rounds of Selection (5 pages)

Appendix 8: Summary Tables of Pathways Enriched in GSEA Analyses from D425, MB411FH, and Rcmb06 PDX Mouse Models (10 pages)