2016

Characterization of Mechanisms That Mediate Cancer Metastatic Colonization

Alexander Huan Nguyen

CHARACTERIZATION OF MECHANISMS THAT MEDIATE CANCER METASTATIC COLONIZATION

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Alexander Huan Nguyen

June 2016

CHARACTERIZATION OF MECHANISMS THAT MEDIATE CANCER CELL METASTATIC COLONIZATION

Alexander Huan Nguyen, Ph.D.

The Rockefeller University 2016

Metastatic disease is the major cause of death in all solid tumor cancers. Current therapeutic strategies fail to target metastasis as the genes and mechanisms that regulate this process remain poorly understood. Metastatic colonization is the final step of the metastatic cascade whereby cancer cells form a tumor at a distant site. This final step is the culmination of clonal evolution of cancer populations that results in a highly aggressive population with enhanced metastatic capacity and often presents clinically as numerous inoperable tumor nodules that lead to mortality. Characterization of the mechanisms that govern metastatic colonization at cellular and molecular levels is necessary for the prevention and treatment of metastatic disease in patients.

The first half of this thesis presents work towards understanding mechanisms that mediate colorectal cancer colonization of the liver in order to guide novel therapeutic strategies. An *in vivo* large-scale RNA-interference screen was performed to identify genes required for liver colonization. Liver and red blood cell pyruvate kinase (PKLR) was identified as a driver of liver metastasis in experimental models. In patients, PKLR was found to be expressed at higher levels in liver metastases relative to primary colorectal cancer tumors and also over-expressed in the primary tumors of patients with metastatic disease. PKLR

was found to promote cell survival in the tumor core and enhance survival during conditions of concurrent high cell density and low oxygen availability. Molecular studies revealed that PKL negatively regulates pyruvate kinase M2 (PKM2) enzymatic activity. By inhibiting cellular pyruvate kinase activity, PKLR allows for the diversion of metabolites towards glutathione generation—allowing for the maintenance of glutathione levels. Adequate glutathione levels appears critical for metastatic colonization as GCLC, the catalytic subunit of glutamate-cysteine ligase and the rate-limiting enzyme for glutathione synthesis, was found to be similarly required for effective metastasis, associated in its expression with human liver metastatic progression, and could be therapeutically targeted to reduce metastatic colonization. These findings highlight the impact of metabolic regulation on cancer cell adaptation within the metastatic niche. The robust effects on liver metastatic colonization observed upon modulating this metabolic pathway suggest clinical potential for therapeutic targeting of PKLR or cellular glutathione synthesis in colorectal cancer.

The second half of this thesis presents work towards an understanding of diversity generation in clonal populations as it benefits cancer evolution and metastatic colonization. Clonal human breast cancer subpopulations were isolated to allow for the identification of subpopulations that exhibit population-level phenotypic diversity. These high variability clonal subpopulations were found to be more proficient at metastatic colonization—consistent with a positive role for diversification capacity in cancer progression. Through single-cell RNA-sequencing, cell-

to-cell transcript expression variability was identified as a defining feature of these subpopulations, extending to protein-level variability. Furthermore, spliceosomal machinery was identified as a gene set with high expression variability, suggesting a means by which variation could be transmitted to a global level. Engineered variable expression of the spliceosomal gene SNRNP40 promoted metastatic fitness, and this metastatic capacity was attributable to cells with low SNRNP40 expression. Clinically, low SNRNP40 expression is associated with metastatic relapse. These findings reveal that transcriptomic variability generation may serve as a mechanism by which cancer subpopulations achieve diversification of gene expression states, which allows for enhanced fitness under changing environmental pressures encountered during metastatic progression.

For Ba và Mẹ

ACKNOWLEDGEMENTS

First, I would like to thank Sohail for the support, mentorship and guidance. I appreciate the confidence you had in me, even before I had joined this program and your lab, and the independence to develop as a scientist. I have learned a great deal and am truly grateful for the experience.

I would like to thank members of the Tavazoie lab from whom I have learned so much and with whom I was fortunate to have worked. Thanks to Jia Min for experimental guidance and for being a great colleague/baymate; to Ethan for being a supportive colleague and always being generous with your time and energy; to Hani for support and discussions, both intellectual and trivial; to Claudio for your wisdom and guidance; to Lisa for your continued assistance; to Rohit for your enthusiasm and hard work; to Steven for your youthful spirit; to Fung Ying for your efforts and positivity; to Mitsu for your experimental contributions; and Paul and Jason for passing on wisdom on the path ahead. I also want to thank the rotation students who have worked with me, Dan, Sohn, Yetis, Phil, Amy, and Robbie, as your efforts have aided this work.

I would like to thank Dr. Brian Chait, Dr. Sandy Simon, Dr. Jochen Buck for providing guidance and feedback as my faculty advisory committee. Thank you to Dr. Siavash Kurdistani for serving as my external examiner. I would like to thank the Tri-Institutional MD/PhD Program, especially Dr. Olaf Anderson, Dr. Mark Pecker, and Mrs. Ruth Gotian for the academic guidance and personal support, and the

Rockefeller Office of Graduate Studies. I would like to thank the core laboratories at Rockefeller University that this work would not have been possible without their facilities and technical support: Dr. Svetlana Mazel of the Flow Cytometry Resource Center, Dr. Connie Zhou of the Genomics Resource Center, Dr. Allison North of the Bio-Imaging Resource Center, Dr. Henrik Molina of the Proteomics Resource Center, and Dr. Fraser Glickman of the High Throughput & Spectroscopy Resource Center. I would also like to thank my former mentor Iswar Hariharan, who introduced me to science, the joy of discovery, and the pleasure in pursuing one's passion.

Finally, I'd like to thank my family and friends. I especially want to thank Carrie who has always been supportive. I want to thank my siblings and my parents for their continued guidance, love, faith and support.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

Metastasis as a major cause of cancer death

Cancer is the second most common cause of death in the United States, accounting for an estimated 589,430 deaths in 2015 (Siegel et al., 2015). It is estimated that 90% of these deaths is a result of metastasis, when cancer cells leave the site of the primary tumor and develop a tumor in a distant organ (Mehlen and Puisieux, 2006; Weigelt et al., 2005). Despite the significant clinical impact, current therapies fail to affect metastatic outcomes, which is highlighted by the drastically low 5-year survival rates of metastatic breast and colorectal cancer, two of the most prevalent and deadly cancers in the US (Figure 1.1)(Siegel et al., 2015). There currently does not exist any therapy with broad metastasis-suppressive activity as the mechanisms that regulate metastasis remain poorly characterized (Chiang and Massague, 2008).

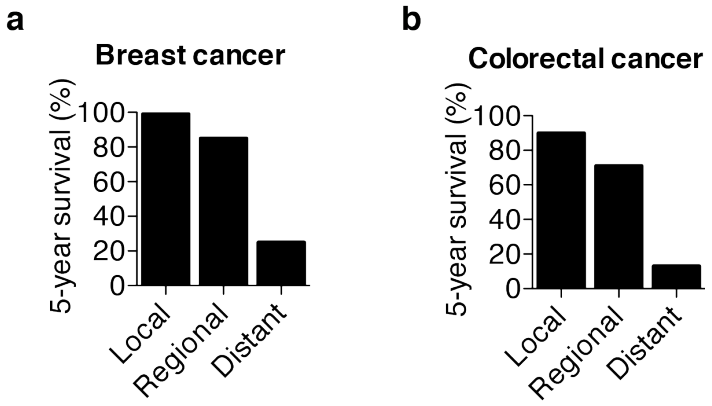

Figure 1.1: 5-year survival for breast and colorectal cancer
Survival rates as calculated by Siegal et al., 2015 for breast (**a**) and colorectal (**b**) cancers. Local stage is defined as no spread to lymph nodes or nearby structures. Regional stage is defined as spread to nearby tissues or lymph nodes. Distant stage is defined as spread to distant lymph nodes or organs.

The metastatic cascade and organ tropism

Metastasis is a complex, multi-step process whereby cancer cell populations survive and expand through a series of different selective pressures before a tumor can form in a distant organ (Illustration 1.1). During early stages of metastasis, cancer must first locally invade through the basement membrane of the tissue wall, allowing for access to local lymphatic and hematogenous circulation (Gupta and Massague, 2006; Valastyan and Weinberg, 2011). Cancer populations must survive in circulation until arrival at a distant organ, where successful extravasation, adaptation to microenvironmental pressures, and cancer population growth must occur to allow for continued cancer progression (Talmadge and Fidler, 2010). Metastatic colonization is the final step of this process, which is defined as the capacity of cancer cells that have arrested at a distant organ to develop into macroscopic metastatic tumor nodules.

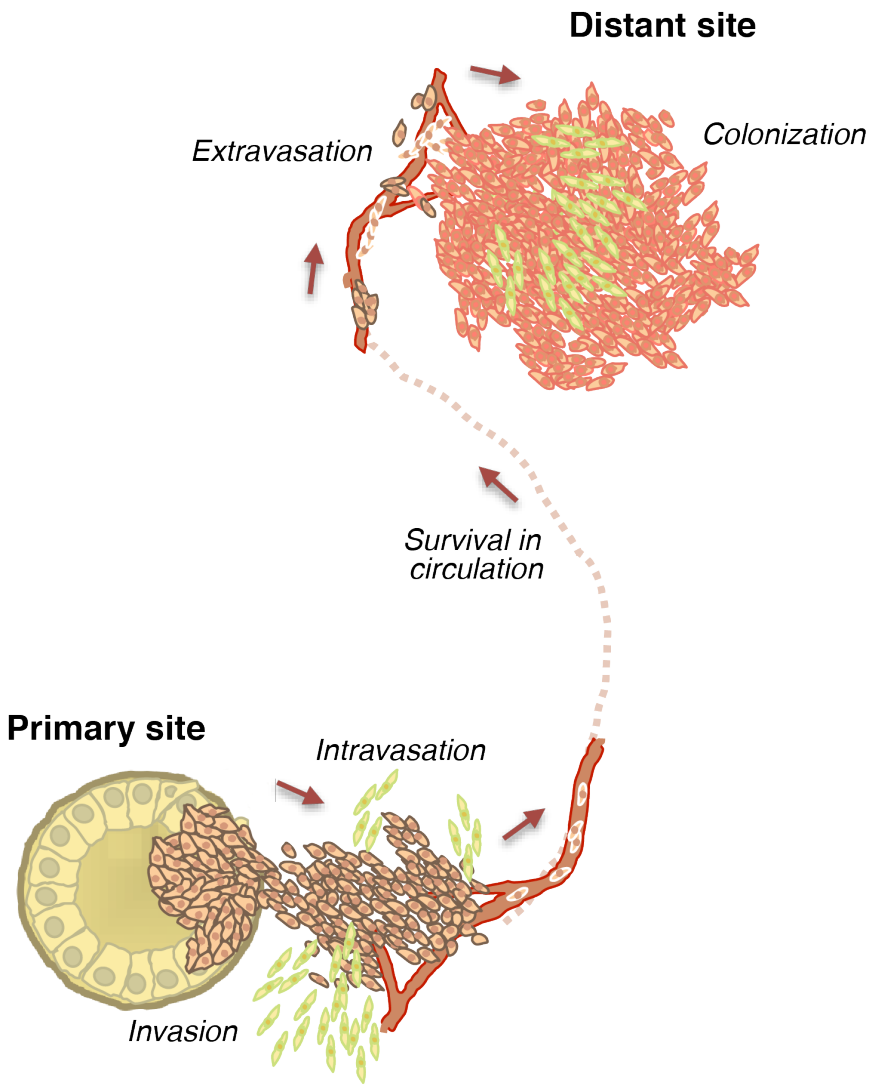

Illustration 1.1: The metastatic cascade

Cancer populations that develop in the primary site must invade through the basement membrane, intravasate into circulation, survive in transit until arresting at a distant organ. Populations must then successfully extravasate into the distant organ where colonization must occur before a macroscopic metastasis is formed.

Much of the difficulty in broadly impacting metastasis in the clinic can be attributed to the heterogeneity of metastasis. While primary tumors from cancers of different types harbor similar oncogenic and tumor-suppressive mutations that enables deregulation of pathways essential to tumorigenesis (Hanahan and Weinberg, 2000), metastatic cancer varies in its presentation clinically and in its phenotypic capabilities depending on the cancer type. Clinicians have long been aware that cancer types differ in their propensity to colonize distant organs—metastatic breast cancer most commonly occurs in the lung, liver, bone and brain, whereas metastatic colorectal cancer most commonly occurs in the liver (Fidler, 2003). This inherent propensity was originally attributed to anatomy as it was proposed by Virchow in 1858 that cancer cells dislodged from a primary tumor follow circulation to form emboli in the vasculature of distant organs (Virchow, 1989). This theory was further advanced by Ewing in 1928 who proposed that circulatory architecture and mechanical forces dictated organ specificity (Ewing, 1928). The countering theory of metastasis was developed by Paget in 1889, who noted from autopsy records that in addition to the lungs, metastatic breast cancer preferentially colonized bones, liver and ovaries—organs that do not immediately follow the breast in the circulatory system (Paget, 1989). This preferential metastatic colonization led Paget to generate the "seed-and-soil" hypothesis in which he posited that cancer cells disseminate to many organs, but that macroscopic metastases develop only at the sites permissive to cancer cell survival and growth. Studies by Fidler nearly a century later revealed that while cancer cells arrive and can arrest in

nearly all organs, metastatic colonization only occurs in select organs, confirming Paget's hypothesis and indicating that biological factors at the metastatic niche can dictate organ specificity (Hart and Fidler, 1980). Fider and colleagues also demonstrated that subpopulations in a primary tumor contribute to the metastatic population, indicating that cancer cells possess inherent biological features that enable metastatic potential (Fidler and Kripke, 1977). Together, it is now understood that molecular drivers from both the cancer cells as well as the metastatic niche contribute to the development of macroscopic metastases.

Organ tropism has been validated at the molecular level where functional studies have revealed distinct molecular drivers in breast cancer populations enable proclivities to colonize the bone, brain, lung, and liver (Bos et al., 2009; Kang et al., 2003; Minn et al., 2005; Tabaries et al., 2011). Importantly, metastatic cancer cells require distinct phenotypic abilities to colonize different organs. Breast and melanoma cancer cells that colonize the lung depend heavily on the ability to migrate through endothelium, invade through extracellular matrix, and recruit endothelial cells for continued tumor growth (Pencheva et al., 2012; Png et al., 2012; Tavazoie et al., 2008). Colorectal cancer appears more dependent on adequate cancer cell energetic supply to survive in the hypoxic liver microenvironment during initial colonization of the liver (Loo et al., 2015). Molecular and cellular characterization of these driver genes and pathways in various cancer types and organs have allowed for the potential development of novel strategies to therapeutically target metastasis (Loo et al., 2015; Pencheva et al., 2014; Pencheva et al., 2012).

The heterogeneity between cancer types illustrates how focused characterization of cancer type-specific mechanisms that govern metastatic colonization can be translated into clinical therapies.

Clonal evolution during the metastatic cascade

The understanding of the metastatic cascade as a step-wise progression of a population with increasing aggressiveness is based on the model of clonal evolution of cancer populations first proposed by Nowell in 1976 (Nowell, 1976). Based on Darwinian natural selection, this model proposes that cancer populations progress via mutational events that provide a fitness advantage through rounds of selective pressures (Illustration 1.2). This model was delineated in patients most notably by Vogelstein who observed an association between specific driver mutations with the aggressiveness of early stage of colorectal cancer (Vogelstein et al., 1988). More recently, whole-genome sequencing has revealed that within individual tumors, subclonal expansion of cancer populations during cancer progression can be attributed to the effect of specific driver mutations (Ding et al., 2012; Gerlinger et al., 2012; Juric et al., 2015), consistent with Nowell's model.

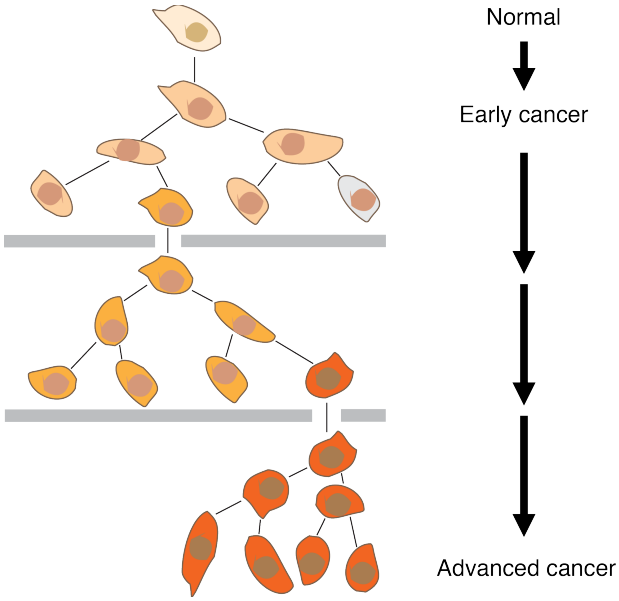

Illustration 1.2: Clonal evolution of cancer populations
Cancer progresses by generating mutational diversity, which impacts phenotypic differences in a population, depicted by varying cell color. Variation in phenotypes leads to increased fitness in a subset of the population, enabling survival when faced with a selective pressure, depicted by the gray bar, and subsequent clonal expansion. This process is repeated through multiple rounds until a highly aggressive cancer population is generated.

The original theory behind the evolution of cancer populations can be extended to current areas of interest in the cancer biology field. First, the inherent variation in the population originally defined as genetic in nature now includes epigenetic mutational events. Many cancers are can be driven by epigenetic processes, where variation in histone modifications can predict tumor recurrence (Seligson et al., 2005). Furthermore, the clonal evolution theory can be applied to the metastatic cascade as each step poses a significant impediment to cell survival and

division. Importantly, the final step of metastatic colonization is a highly selective step with an estimation that less than 0.01% of cells that arrive at a distant organ are able to form metastases (Fidler, 1970). Cancer populations at this step would greatly benefit from an increase in phenotypic variation, leading to higher success at this evolutionary bottleneck (Merlo et al., 2006). Characterization of mechanisms that generate phenotypic diversity in a cancer population has the ability to impact understanding and the prevention of metastasis formation broadly.

Benefit of diversity to cancer evolution

Generation of phenotypic diversity is beneficial if the phenotype serves to increase survival and reproduction. Acquisition of the hallmarks of cancer, such as self-sufficient growth, limitless replication potential and evasion of apoptosis, translate into an increase in population fitness (Hanahan and Weinberg, 2000). The initial acquisition of these features and subsequent clonal expansion is best understood in the context of genetic alterations. During DNA replication, spontaneous mutations can occur in normal cells at a rate of 10^{-6} to 10^{-7} mutations per cell division (Araten et al., 2005), which may lead to a rare yet significant cancer-initiating event. However, with an estimated 3-12 mutations required to initiate cancer (Renan, 1993), it was argued that this background mutation rate is insufficient to cause cancer (Loeb, 1991). The discovery of mutations in DNA repair genes in hereditary nonpolyposis colorectal

cancer revealed a mutator phenotype that can enhance cancer progression by increasing the rate at which mutations occur, which increases the probability that oncogenes and tumor suppressors become genetically altered (Fishel et al., 1993; Leach et al., 1993; Rampino et al., 1997; Renault et al., 1996). These cancer-promoting mutations in individual cells lead to their clonal expansion into a significant proportion of a given tumor; high-throughput sequencing of different regions within individual tumors have revealed spatial genetic heterogeneity and evolutionary history of a cancer as it progressed (de Bruin et al., 2014; Gerlinger et al., 2012; Navin et al., 2011; Park et al., 2010). After iterative rounds of clonal selection and expansion, a large number of genetic variants can exist in a population. It is estimated that there may be approximately 1,000-3,000 mutations per cancer cell and on the order of 10^{12} mutations in a tumor (Bielas et al., 2006), although the majority are in non-coding regions or have neutral effect on fitness. This number of mutations observed corresponds to a 200-fold higher mutational rate in cancers as compared to normal cells (Klein, 2006). This capacity to generate genetic diversity in cancer population allows for cancer progression at a rate much higher than chance alone.

Nevertheless, any heritable change in subclonal populations that provides a fitness advantage will enable clonal expansion and cancer progression. Most notably, epigenetic mutations and modifications are understood to occur more readily as compared to genetic mutations and thus, significantly contribute to cancer progression (Baylin and Herman, 2000; Feinberg et al., 2006). Heterogeneity of histone modification patterns has been shown to be of predictive utility for prostate cancer recurrence

(Seligson et al., 2005). One mechanism by which epigenetic mutations contribute to population diversity is through inactivating hypermethylation of DNA-damage response and repair genes, leading to genetic instability (Horie-Inoue and Inoue, 2006; Weisenberger et al., 2006). However, additional mechanisms by which epigenetic alterations contribute to phenotypic diversity remain poorly characterized and warrant further investigation.

From an ecological perspective, metastasis is perceived as the colonization of a new habitat, which is dependent on the invasive ability of the cancer cells as well as features of the new ecosystem (Merlo et al., 2006). Ecological studies have indicated the benefit of increased species complexity for invaders of a new habitat (Shea and Chesson, 2002), while studies in bacteria similarly indicated that population fitness is dependent on population diversity (Imhof and Schlotterer, 2006). Importantly, metastasis poses numerous selective pressures that the generation of phenotypic diversity would provide a significant fitness advantage. Furthermore, analysis of metastases from breast, colorectal and pancreatic cancers have revealed limited genetic alterations as compared to primary tumors (Ding et al., 2010; Jones et al., 2008; Yachida et al., 2010), suggesting that genetic instability may be insufficient for the acquisition of metastatic capacity. Thus, the acquisition of metastatic capacity may in part be driven by non-genetic mechanisms.

Colonization at end organ is the clinically relevant event

While metastasis involves numerous sequential steps, end organ colonization is a critical event as it is the point at which metastatic disease can first be clinically diagnosed. Biologically, as compared to all the phenotypic selections required throughout metastatic cascade, end organ colonization has experimentally been demonstrated to be one of the most rate-limiting selective pressures. Xenograft inoculation of human cancer cells into immune-deficient mice was the first *in vivo* experimental method that enabled study of end organ colonization (Fidler, 1986; Rygaard and Povlsen, 1969). Initial studies by Fidler observed that upon intravenous inoculation of melanoma cells, while less than 0.1% of cells could survive circulation, less than 0.01% of these surviving cells could produce experimental metastases (Fidler, 1970). More recently, intermediate steps of the metastatic cascade that include survival in circulation, arrest at a distant organ, and extravasation were observed to occur efficiently—80% of intravenously inoculated cells successfully extravasated. However, the ability to form macroscopic metastases was calculated to be less than 0.02% (Luzzi et al., 1998).

This selectivity is observed analogously in patients. Among ovarian cancer patients who received peritoneovenous shunts to treat peritoneal ascites, effectively releasing tens of millions of cancer cells into circulation, many never developed metastases (Tarin et al., 1984). Furthermore, among thousands of breast cancer patients who exhibited micrometastases in their bone marrow at diagnosis, only 50% developed overt metastases within 10 years (Braun et al., 2005). The development of

macroscopic metastases is ultimately the cause of mortality. Given the selectivity, clinical intervention to target metastatic colonization would reduce the likelihood of successful therapeutic resistance and escape. The characterization of genes and mechanisms that regulate this process would provide significant benefit towards improving patient survival.

Objectives

The goal of this thesis is to characterize mechanisms involved in metastatic colonization. Two approaches were performed. The first portion of this thesis was to characterize key signaling molecules that drive colorectal cancer metastatic liver colonization in order to develop novel therapeutic strategies. Using an *in vivo* large-scale RNAi screen, liver and red blood cell pyruvate kinase (PKLR) was identified as a driver of liver metastatic colonization. In patients, PKLR expression was increased in liver metastases as well as in primary tumors of patients with metastatic disease. PKLR promoted cell survival in the tumor core during conditions of high cell-density and oxygen deprivation through increasing glutathione levels. PKLR was found to negatively regulate the glycolytic activity of PKM2, a master metabolic regulator involved in controlling glycolytic flux as well cellular glutathione levels. Glutathione appeared to be critical for metastasis as GCLC, the rate-limiting enzyme for glutathione synthesis, was similarly required for liver colonization and could be therapeutically inhibited to reduce metastatic colonization. These findings highlighted the impact of metabolic re-programming within the

niche as metastases progress and suggested clinical potential for targeting this pathway in colorectal cancer.

The second portion of this thesis was to characterize mechanisms by which breast cancer cells may generate diversity as an enhancement for clonal evolution and metastatic fitness. These mechanisms could potentially be relevant in many cancer types as clonal evolution pertains to all contexts of metastasis. Clonal human breast cancer subpopulations were isolated that displayed substantial population-level phenotypic diversity. These high variability clonal subpopulations were more proficient at metastatic colonization—consistent with a positive role for diversification capacity in cancer progression. Through single-cell RNA-sequencing, cell-to-cell transcript expression variation was identified as a defining feature of these subpopulations, extending to protein-level variation. Furthermore, spliceosomal machinery was identified as a gene set with high expression variability, suggesting a means by which variation could be transmitted to a global level. Engineered variable expression of spliceosomal gene SNRNP40 promoted metastatic fitness, attributable to cells with low SNRNP40 expression. Clinically, low SNRNP40 expression was found to be associated with metastatic relapse. The experimental model established here could be applied to various cancers to better understand non-genetic contributions to heterogeneity and to study the impact of such deregulation amongst cancer populations and their progeny.

CHAPTER 2: PKLR Promotes Colorectal Cancer Liver Colonization by Increasing Glutathione Levels

2.1 Introduction

Clinical need for therapies against metastatic colorectal cancer

One cancer type that would greatly benefit from improved therapeutic options is colorectal cancer, the third most common cancer and the third leading cause of cancer death in America (Siegel et al., 2015). In 2015, 132,700 new cases are expected to be diagnosed, while 49,700 individuals are expected to die from colorectal cancer. In the past few decades, incidence and death rates have decreased largely due to improved prevention and detection through screening. The adoption of colonoscopy as a highly sensitive screening method has enabled both the detection of adenomatous polyps, which can be removed to prevent cancer development (Corley et al., 2014; Zauber et al., 2012), as well as the detection of early-stage cancer, which allows for potentially curative surgery and greatly improved outcomes (Nishihara et al., 2013). When colorectal cancer is localized to the intestine, the five-year relative survival rate is nearly 90%.

Despite the effectiveness of colorectal cancer screening, only 59% of the US population aged fifty years and older participates in screening measures. Because early-stage disease is often asymptomatic, those who do not participate in these screening measures can develop undetected disease that slowly progresses to advanced stages over the course of ten to twenty years (Winawer and Zauber, 2002). As a result, distant disease is found in 25% of all initially diagnosed patients and develops in 40-50% of newly diagnosed patients (Van Cutsem et al., 2009). This translates to a

five-year survival of 13% for those with metastatic disease (Davies and Goldberg, 2011; Ferlay et al., 2014).

While localized disease can be cured with surgery in the majority of cases, metastatic disease is associated with the worst prognosis with the liver being the most common site of distant disease. Curative surgical resection of liver metastases may be achieved only in a minority of cases. However, relapse after resection occurs in 75% of these patients (Nordlinger et al., 1996), and the five-year survival rate for those who relapse remains low at only 36% (Rees et al., 2008). Importantly, for the majority of cases, surgery is not possible, and chemotherapy remains the only therapeutic option to extend life.

Current chemotherapy regimens are based on 5-fluorouracil (5-FU), a pyrimidine analog that was introduced and adopted for clinical use in the 1950s (Heidelberger et al., 1957). 5-fluorouracil was found to incorporate into DNA during replication and inhibit RNA synthesis, leading to decreased cell growth and proliferation. This growth-inhibitory effect was first observed in solid tumors implanted in mice, and rapidly was translated to clinical responses in patients with a variety of cancer types (Curreri et al., 1958). Leucovorin (LV), a vitamin that inhibits thymidylate synthase, was added to enhance 5-fluorouracil efficacy (Poon et al., 1989). While 5-FU/LV showed significant efficacy in prolonging life, this combination exhibited no major impact on overall survival (1992).

Alkylating agents such as oxaliplatin and irinotecan were included more recently, forming the current cytotoxic chemotherapy combination regimens, FOLFOX (5-FU/LV/oxaliplatin) (de Gramont et al., 2000) and

FOLFIRI (5-FU/LV/irinotecan) (Saltz et al., 2000). Despite significant effects on progression-free survival time, modest effects on overall survival time, and nuanced changes to timing and delivery strategies, only 10% of patients live beyond five years since metastases ultimately lose responsiveness to these therapies (Davies and Goldberg, 2011). Molecularly targeting therapies such as the VEGF inhibitor bevacizumab and EGFR inhibitor cetuximab have marginally improved survival by delaying progression, though patients eventually do succumb to the disease (Davies and Goldberg, 2011). At this point, the current chemotherapy regimen mainly serves a palliative role to extend survival and improve quality of life (Glimelius et al., 1994). For a therapy to have curative potential, strategies specific against metastasis-promoting phenotypes must be developed to limit progression at rate-limiting steps where resistance will be less readily acquired. With metastatic disease in the liver being the life-limiting aspect for late-stage colon cancer patients, characterization of metastatic colonization and progression would provide significant advances towards the development of more effective therapies.

Mechanisms of colorectal cancer progression

Colorectal cancer arises through a progression of mutational and transformative events on epithelial tissue that allows for increasing aggressiveness. The initiating event is often the acquisition of genomic instability. Genomic instability can arise by DNA hypermethylation in fifteen percent of sporadic cases, while germ-line mutations in DNA

repair machinery predisposes patients to a high likelihood of colorectal cancer (Markowitz and Bertagnolli, 2009). The most common mode of genomic instability is through chromosomal instability. This early genetic deregulation allows for hyperplastic growth and proliferation, characteristic of adenomatous polyps. Uncontrolled growth and continued genetic instability leads to inactivating mutations in critical tumor-suppressor genes such as APC, p53, and SMAD4 and activating mutations in oncogenes such as K-RAS, BRAF, and PIK3CA (Markowitz and Bertagnolli, 2009)(Illustration 3.1). These genes regulate cellular processes that when deregulated, help define fundamental aspects of cancer regardless of cancer type, which include self-sufficiency in growth signaling, resistance to anti-growth signals, evasion of apoptosis, and limitless replicative potential (Hanahan and Weinberg, 2000). Continued evolution of cancer populations led to even more aggressive phenotypes, eventually progressing into invasive and disseminated disease. The model of progressive acquisition of cancer-promoting mutations is supported by the observed genetic profile of primary tumors of different grades (Fearon and Vogelstein, 1990; Vogelstein et al., 1988). Progression into metastasis is understood to be an evolutionary process whereby clonal selection of cancer cells with the highest fitness at a phenotypic level allows for population advancement (Nowell, 1976; Talmadge and Fidler, 2010).

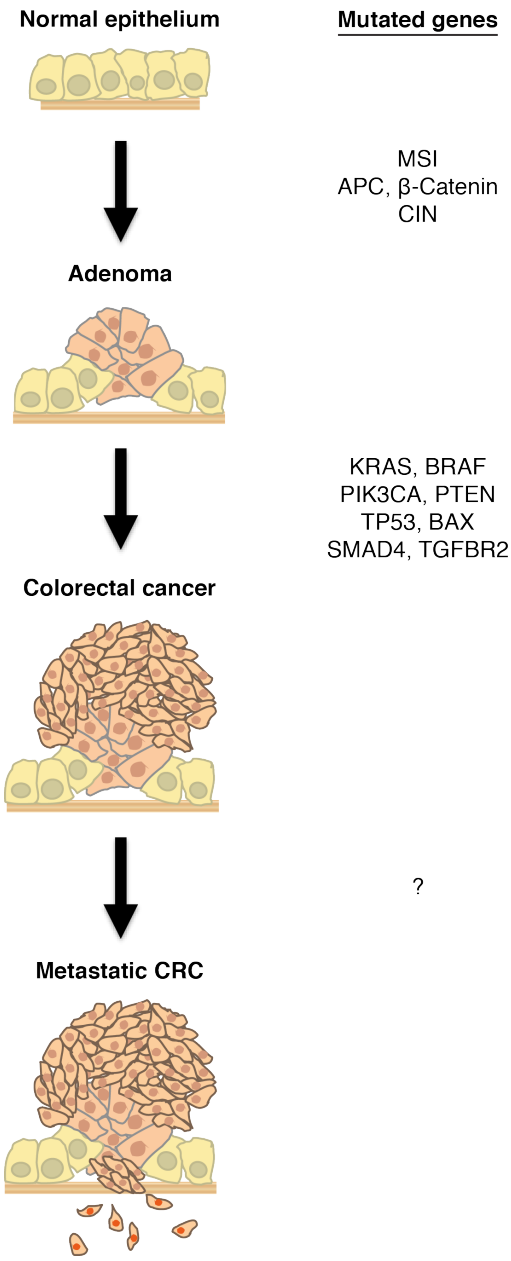

Illustration 2.1: Genetic progression of colorectal cancer

Progression of colorectal cancer is shown with known genetic alterations that contribute to cancer development. MSI = microsatellite instability, CIN = chromosomal instability

Metastatic colonization of the liver

To clinically impact colorectal cancer, the process of liver colonization must be better characterized (Talmadge and Fidler, 2010; Valastyan and Weinberg, 2011). Metastatic colonization, the formation of macroscopic metastases, has been observed to be a significant rate-limiting step during metastatic liver growth (Luzzi et al., 1998). This selective pressure may be in part attributable to the liver microenvironment, which is marked by hypoxic regions, a unique circulatory architecture, and significant metabolic activity (Jungermann and Kietzmann, 1996). The genes and pathways most suitable for therapeutic inhibition would be those whose expression is necessary for colonization events. Furthermore, it remains unclear whether cancer cells utilize different mechanisms for early or late colonization events (Sullivan and Christofk, 2015). Early metastatic colonization can be achieved through utilization of liver-derived energetic molecules to enhance survival upon arrival in the hypoxic, nutrient-poor liver microenvironment (Loo et al., 2015). However, therapeutics that target early colonization are only useful for prevention of metastatic nodule formation. Late colonization events, which contribute to the continued progression of macro-metastases, represent a stage with significant clinical need given that patients who develop liver metastases exhibit a poor prognosis and that approximately twenty percent of patients initially present with distant metastatic disease (Siegel et al., 2014). Nevertheless, both early and late metastatic colonization remain poorly characterized and must be systematically

approached for the development of biologically relevant therapeutic strategies.

Functional genomics in cancer

The identification of the key regulators of metastatic colonization can benefit from a systematic, functional approach. Classical genetic screening approaches utilized chemical and transposon-based mutagenesis, allowing for the unbiased identification of genes responsible for the biological process of interest (Boutros and Ahringer, 2008). However, the process of identification and characterization of the genes of interest is time-consuming and laborious. Additionally, complex phenotypes and the large size of the mammalian genome render classical screening approaches infeasible (Boutros and Ahringer, 2008). The discovery and application of RNAi for screening purposes has enabled large-scale screening in mammalian systems (Elbashir et al., 2001) (Root et al., 2006). RNAi screens in cancer were first applied to identify essential genes for cancer cell proliferation and survival (Luo et al., 2008; Schlabach et al., 2008; Silva et al., 2008; Whitehurst et al., 2007). More recent applications have brought this screening technique to cancer progression *in vivo,* enabling the characterization of pathological and developmental processes under more physiological conditions (Beronja et al., 2013; Zuber et al., 2011). Utilization of large-scale loss-of-function screens has enabled the identification and characterization of novel regulators of disease processes.

2.2 Identification of kinases that promote liver colonization

Large-scale in vivo shRNA screen for promoters of liver colonization

To molecularly study colorectal cancer colonization in the liver, a large-scale *in vivo* shRNA screen was performed to identify key regulatory genes (Figure 2.1). Three human colon cancer cell lines, LS174T, SW620 and WiDr, representing the mutational spectrum commonly seen in patients (Kras wild-type/mutant, Braf wild-type/mutant, MSS/MSI), were transduced with shRNA-encoding lentiviral particles. Ten subpools were used in total that included 54,591 total hairpins targeting 14,095 human genes. Cells were transduced at a low titer to ensure only a single shRNA would be incorporated into a cell. These transduced cell populations were then divided into three fractions. The first fraction was saved as a reference population to account for shRNA inserts initially present. A second fraction was inoculated into immuno-deficient mice by direct liver injection for the selection of cells capable of colonizing the liver. The third fraction was grown in culture to select for cells capable of basal growth and survival. Once tumors had developed in the liver, these tumors were harvested, and all three fractions were processed to isolate genomic DNA and amplify the incorporated shRNA inserts for high-throughput sequencing. Sequencing allowed for quantification of shRNA inserts and calculation of relative depletion of shRNAs from tumors and cultured cells. Loss of shRNA representation from the tumors would suggest that depletion of the targeted genes suppressed the ability of colon cancer cells to colonize the liver.

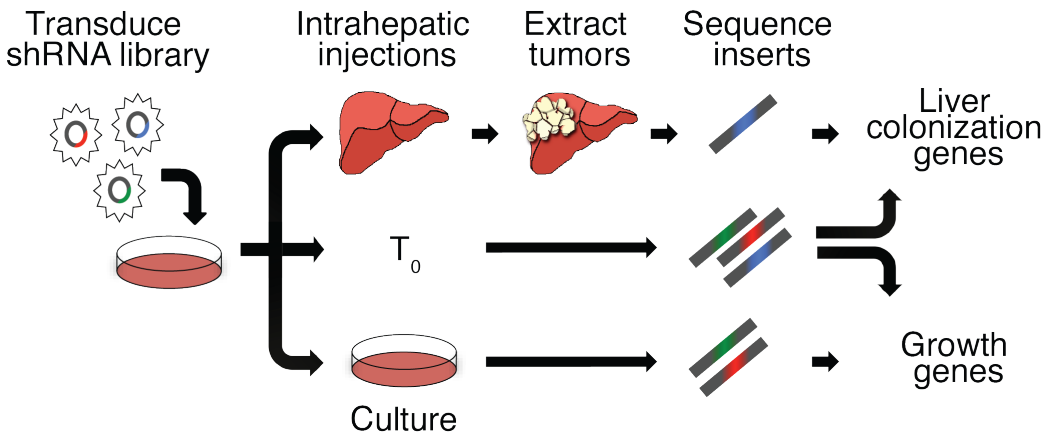

Figure 2.1: Large-scale shRNA screen for promoters of liver metastasis
Schematic of pooled shRNA library dropout screen. Colon cancer cell lines were transduced with an shRNA library and subject to direct liver injections (top), growth in culture (bottom), and reference sample (T_0, middle). Samples were processed to allow for sequencing of shRNA inserts to allow for identification of genes that promote liver colonization as well as genes that promote population growth.

Analysis of shRNA profiling revealed significant depletion of shRNAs across all cell lines in both tumors and cultured cells (Figure 2.2). A correlation matrix for all profiled samples revealed that for the majority of samples, liver tumor profiles displayed greater similarity despite differing parental cell line origins (Figure 2.3). To identify broadly relevant metastasis regulators, genes were considered as putative promoters of liver colonization if at least two shRNAs targeting a given gene were depleted in tumors derived from all three cell lines and both independent transductions. This analysis resulted in 556 candidate promoters of liver colonization (Figure 2.4). Using similar criteria, 719 genes appeared to be required for survival in cultured cells, while 187 genes were in common to both cultured cells and tumor cells (Figure 2.4).

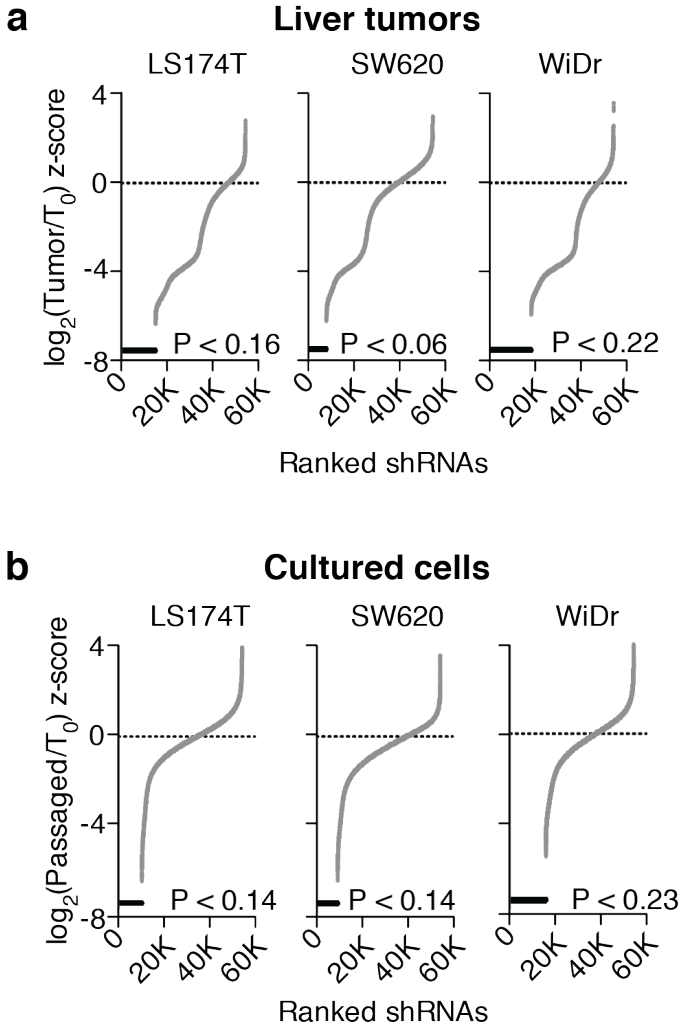

Figure 2.2: Depletion of shRNAs from large-scale shRNA screen for promoters of liver metastasis

Overview of shRNA depletion in liver tumors (**a**) and cultured cells (**b**). shRNAs were ranked based on peak median absolute deviation normalized z-score of $\log_2(\text{Tumor}/T_0)$ or $\log_2(\text{Passaged}/T_0)$. shRNAs absent in both tumor replicates or cultured cells are in black. P-values for absent shRNAs were derived by bootstrapping with 1000 random samplings.

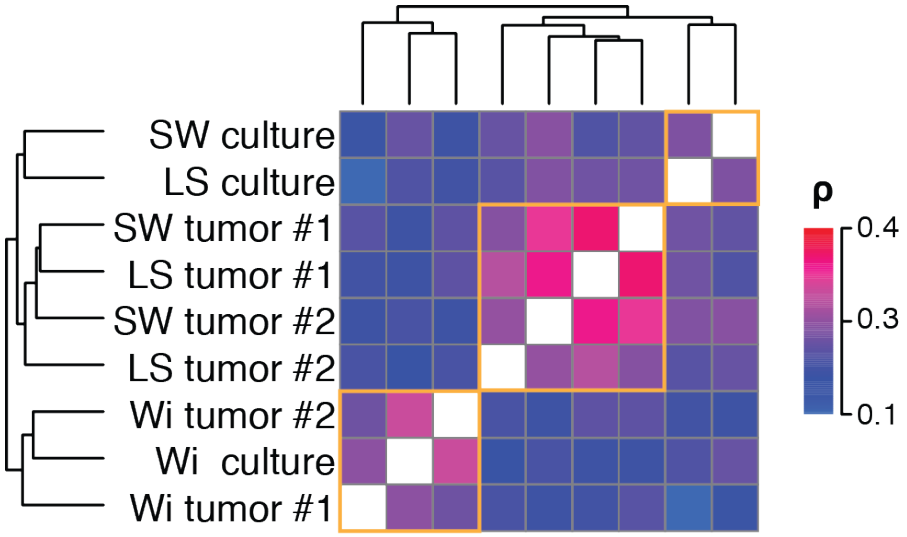

Figure 2.3: Correlation matrix of shRNA screen samples

Correlation matrix of shRNA-depletion profiles of samples using Spearman's correlation coefficient. Clustering was performed in R using Euclidean distance and complete agglomeration method. SW = SW620, LS = LS174T, Wi = WiDr.

a

	shRNAs absent in tumors	Genes with 2+ shRNAs absent
LS174T	15,199 (27.8%)	3,682 (26.1%)
WiDr	18,432 (33.8%)	4,770 (33.8%)
SW620	8,111 (14.9%)	1,465 (10.4%)
All		556 (3.9%)

b

	shRNAs absent in culture	Genes with 2+ shRNAs absent
LS174T	10,548 (19.3%)	2,186 (15.5%)
WiDr	15,994 (29.3%)	3,973 (28.2%)
SW620	9,534 (17.5%)	1,891 (13.4%)
All		719 (5.1%)

c

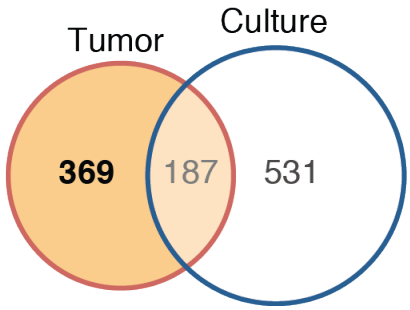

Figure 2.4: Genes with two or more shRNAs absent from screen samples
Summary of absent shRNAs and gene hits from tumor (**a**) and cultured cells (**b**) in each cell line and in common to all cell lines. Percentage of total population is indicated in parentheses. (**c**) Overlap between gene hits scored in tumors and those scored in culture.

To confirm that the analyzed shRNAs were effective in suppressing liver colonization, a secondary library was generated using top-scoring shRNAs that targeted the 556 genes. These top-scoring shRNAs as well as 20 control shRNAs were cloned into the plko.1 expression vector and pooled to allow for equal representation in the library. This library was transduced into a highly metastatic *in vivo*-selected colon cell line, LS-LVM3b (Loo et al., 2015). These populations were then subjected to the experimental procedures performed for the primary large-scale screen. This experiment revealed that these shRNAs were significantly depleted from tumors, confirming the technical efficacy (Figure 2.5). To better rank the candidate hits from the screen, genes were scored from the primary large-scale liver colonization screen by applying the RIGER algorithm, which ranks genes based on the number and relative depletion of targeting shRNAs and thus accounts for differences in the number of shRNAs per gene (Luo et al., 2008). These scores were used to generate gene percentile scores, providing a continuous scale to assess relative importance of genes. As expected, the top 556 candidate genes were significantly enriched among the top percentile scores, providing confidence in the candidate liver colonization gene list (Figure 2.5).

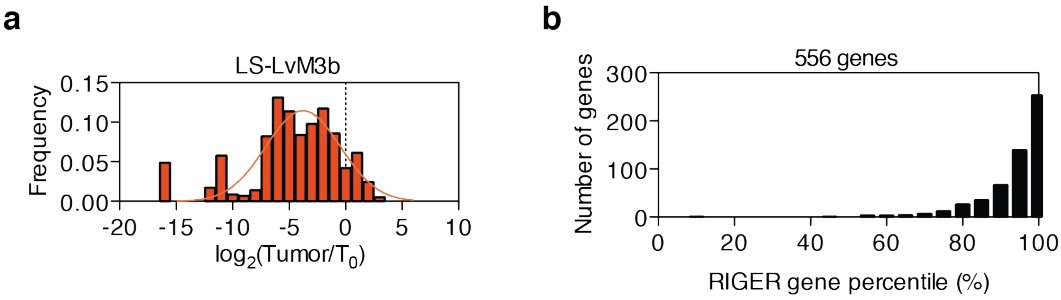

Figure 2.5: Validation of scoring shRNAs and final scoring method
(**a**) Distribution of shRNA depletion in tumors from secondary library using a scoring shRNA targeting each of the 556 top genes. (**b**) Distribution of 556 top genes as scored by RIGER and ranked by gene percentile.

Identification of PKLR as a kinase required for liver colonization

Of the top-scoring 556 genes from the primary screen, kinases, which are known to mediate most of the signal transduction in eukaryotic cells (Manning et al., 2002) were focused on to allow for the identification of critical regulatory mechanisms. The top-scoring kinases from the RIGER gene ranking included many known regulators of colorectal cancer progression including IGF1R (Reinmuth et al., 2002) and DDR2 (Badiola et al., 2011), which, consistent with previous studies, displayed high gene percentile scores in the screen performed in cultured cells, indicative of roles in *in vitro* population growth (Figure 2.6). Surprisingly, one of the top-scoring genes was liver and red blood cell pyruvate kinase (PKLR), which had been commonly believed to be expressed only in liver, kidney, and red blood cells (Luo and Semenza, 2012). Interestingly, PKLR displayed a low gene percentile score under culture conditions, suggesting a specific requirement for this gene during liver colonization.

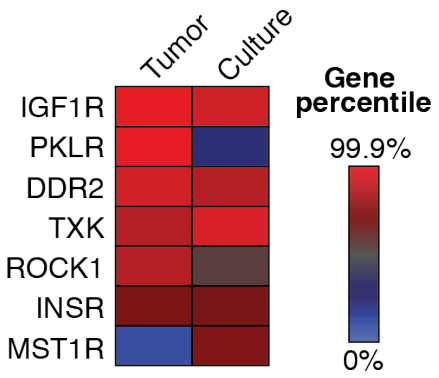

Figure 2.6: Gene percentile scores of top kinases from large-scale shRNA screen

Kinases from top 556 candidate promoters were ranked by RIGER gene percentile score in liver tumor and cell culture large-scale screens.

Fuctional validation of PKLR as promoter of metastatic liver colonization

To confirm these findings and account for potential off-target effects from the targeting shRNAs in the primary screen, LS174T cell lines were generated with additional independent PKLR-targeting shRNAs not previously used in the primary screen (Figure 2.7). PKLR expression was adequately depleted to the levels of 13% & 17% of control. Importantly, in concordance with the *in vivo* screen results, PKLR-depleted cells were significantly less effective than control cells in colonizing the liver following direct liver inoculation (Figure 2.7). Additionally, *in vitro* proliferation assays confirmed that PKLR depletion did not reduce population growth in culture and also did not reduce subcutaneous tumor growth (Figure 2.8).

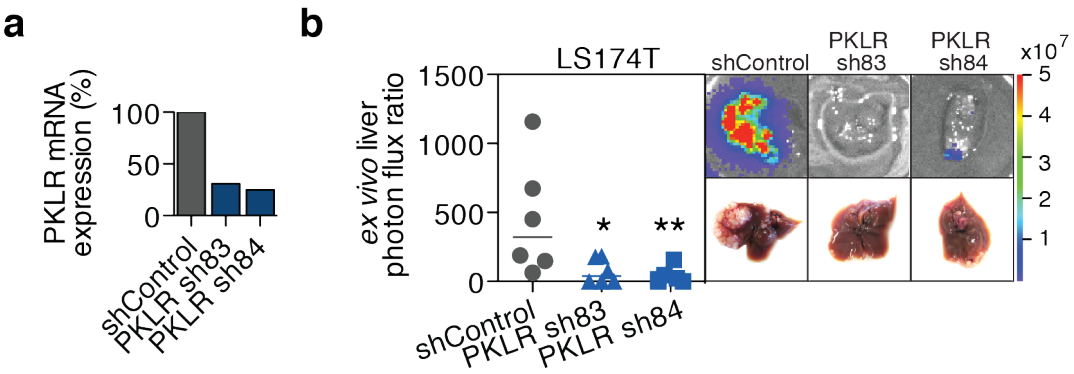

Figure 2.7: PKLR promotes liver colonization

(**a**) PKLR depletion as measured by quantitative RT-PCR in LS174T cells. (**b**) 5x10^5 LS174T cells were inoculated by direct liver injection, and liver colonization was measured by e*x vivo* bioluminescence after 20 days. Representative liver bioluminescence and gross histology are shown (n=6). P-values were derived using one-sided Mann Whitney test.

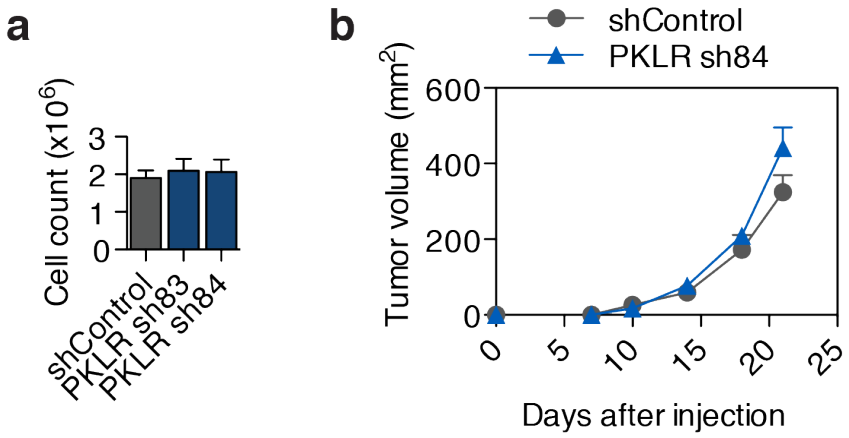

Figure 2.8: PKLR does not promote population growth in culture or subcutaneous tumor growth

(**a**) 5x10^4 LS174T cells were seeded in triplicate and viable cells were counted by trypan blue exclusion after 5 days under recommended cell culture conditions. Data shown is from three independent experiments. (**b**) 10^6 LS174T cells were injected into mice subcutaneously and tumor volume was measured over time (n=6).

To validate this colonization defect in the context of metastasis, additional cell lines were generated with additional shRNAs targeting PKLR. In the highly metastatic LS-LVM3b cell line, PKLR expression was depleted to 9% of control, while in the SW620 cell line, PKLR expression was depleted to 58% and 42% of control (Figure 2.9). These cells were then inoculated via portal circulation to assess metastatic capacity upon hematogenous arrival in the liver. PKLR-depleted cells exhibited significantly decreased liver metastatic colonization capacity in metastasis assays (Figure 2.9). These results revealed PKLR to promote metastatic liver colonization of colon cancer cells.

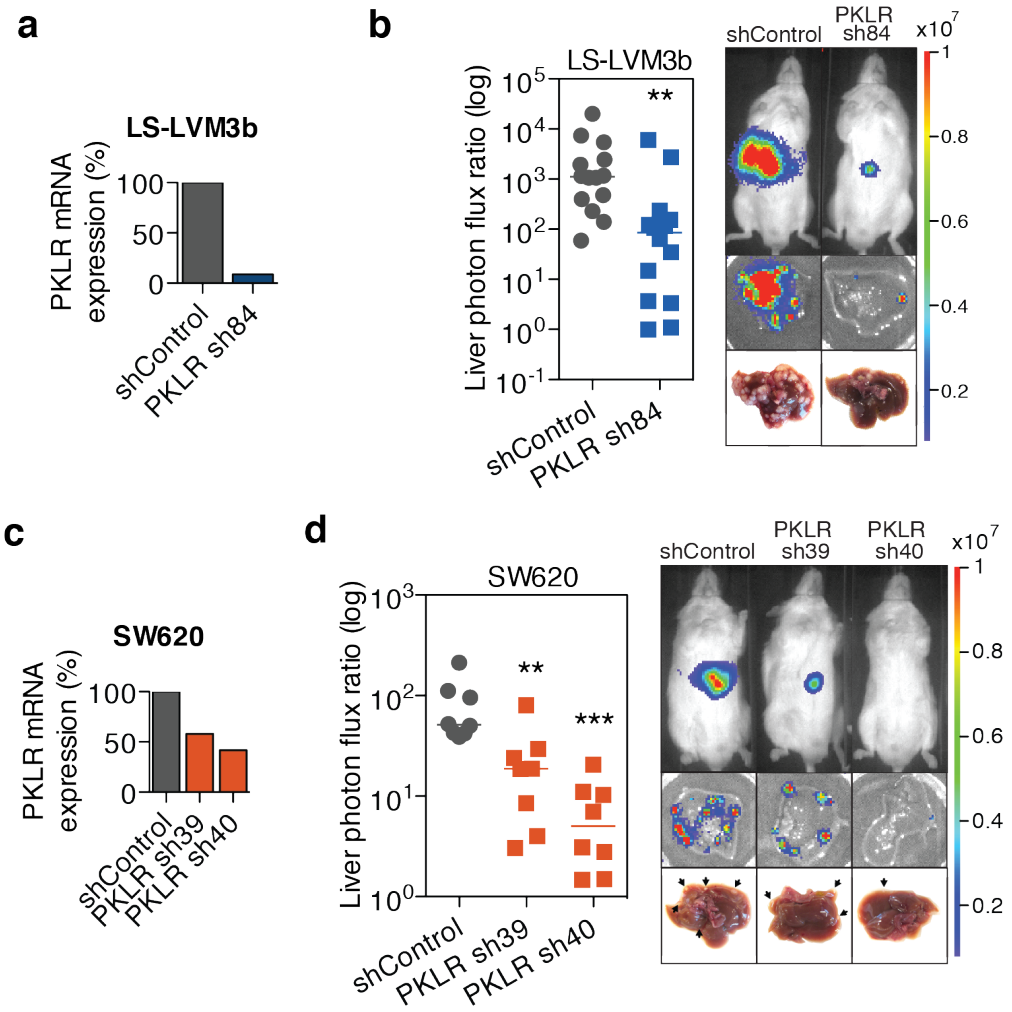

Figure 2.9: PKLR promotes metastatic liver colonization

PKLR depletion as measured by quantitative RT-PCR in LS-LVM3b (**a**) and SW620 cells (**c**). 5x10^5 LS-LVM3b (**b**) and SW620 (**d**) cells were inoculated by portal circulation injection and metastatic colonization was measured by liver bioluminescence after 20 days (n=14) for LS-LVM3b and 35 days (n=8) for SW620. Representative mouse bioluminescence, liver bioluminescence and gross histology are shown. P-values were derived using one-sided Mann Whitney test.

To further characterize the role of PKLR in colonization, PKLR was assessed for whether it promoted colonization in the setting of the primary tumor before cancer cell dissemination. Cells were inoculated into the cecal wall of mice to model primary colorectal cancer tumor growth. PKLR depletion did not affect tumor growth under these conditions (Figure 2.10), indicating that PKLR provides an advantage selectively to metastatic growth.

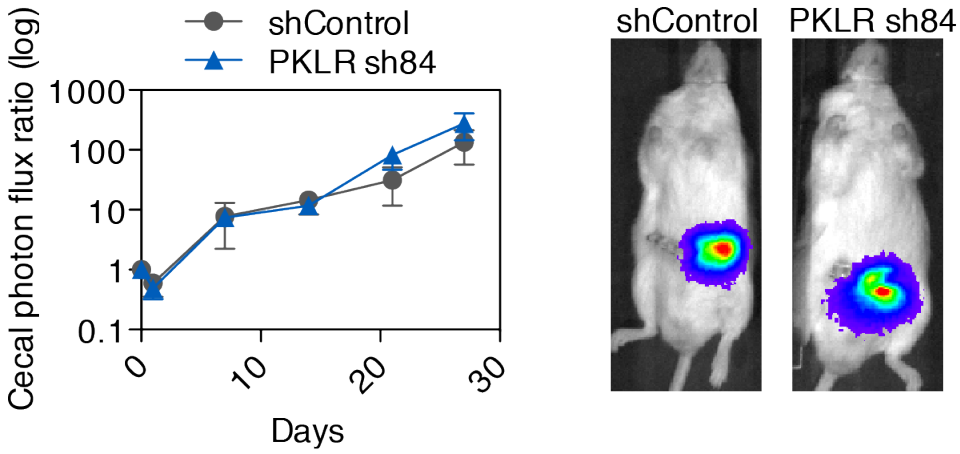

Figure 2.10: PKLR does not promote primary tumor growth

10^6 LS174T cells were inoculated by cecal injection and tumor growth was monitored by *in vivo* bioluminescence over time (n=5). Representative mouse bioluminescence at final time point are shown on the right.

2.3 Clinical relevance of PKLR expression

PKLR expression in normal colon and colon cancer tissue

In mammalian cells, pyruvate kinase is encoded by four isozymes: M1 (PKM1), M2 (PKM2), liver (PKL) and red blood cell (PKR). While the M1, L and R isozymes are described to exhibit tissue-specific expression, the pyruvate kinase M2 isoform is highly expressed across cancer types (Mazurek et al., 2005). Pyruvate kinase L and R have only been described in their respective tissues to promote glycolysis, catalyzing the final rate-limiting step, which involves the transfer of a phosphate group from phosphoenolpyruvate (PEP) to ADP, generating pyruvate and ATP. Although PKLR had originally been described to be expressed solely in liver, kidney and red blood cells (Imamura and Tanaka, 1972), PKL expression has also been observed in epithelial cells in the intestinal epithelium of rats (Domingo et al., 1992; Osterman and Fritz, 1974). Analysis of the Human Protein Atlas data (Uhlen et al., 2010) revealed PKL mRNA expression in normal colon tissue and PKLR protein expression in normal colonic glandular tissue and colorectal cancer tumors (http://www.proteinatlas.org, Figure 2.11).

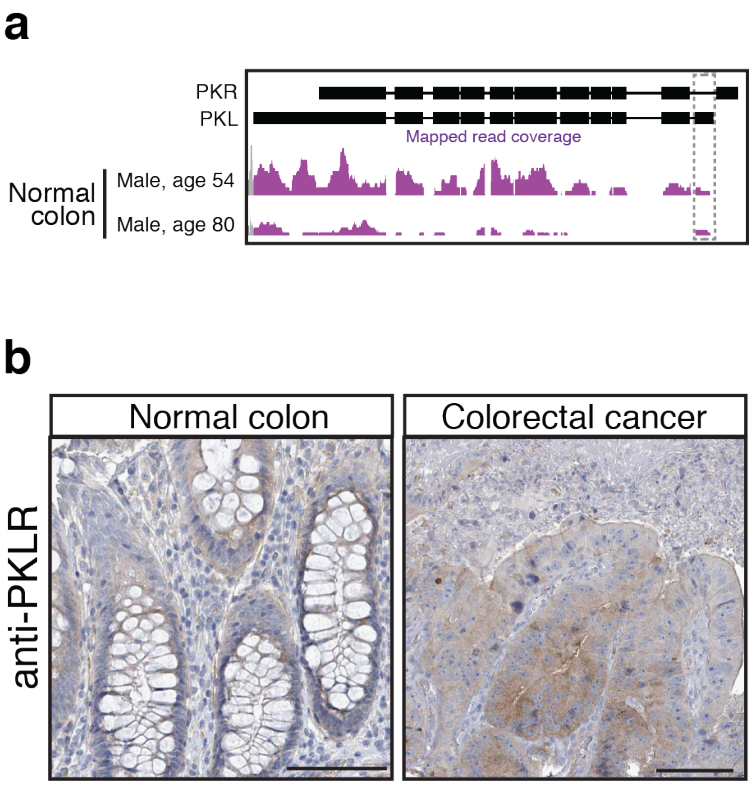

Figure 2.11: PKLR is expressed in normal colon and colon cancer tissue
(**a**) RNA-sequencing read density at PKLR locus from normal colon tissue. PKL-specific exon is indicated with dashed box. (**b**) PKLR protein expression as measured by immunohistochemistry in normal colon and colorectal cancer tissue. Scale bar is 100um.

PKLR expression is associated with clinical liver metastasis

To determine whether PKLR expression levels are associated with liver metastasis, PKLR transcript levels were compared in patient samples of primary tumors and liver metastases. Samples were obtained from patients at MSKCC to compare PKLR transcript levels in patient primary tumor and liver metastasis samples. In 70 samples, PKLR mRNA was significantly up-regulated in liver metastases relative to primary tumors (Figure 2.12c). This up-regulation was also observed in multiple independent datasets of unmatched (Figure 2.12a-b) and matched (Figure 2.12d-e) colorectal cancer patient samples (Ki et al., 2007; Kikuchi et al., 2013; Kim et al., 2014; Matsuyama et al., 2010; Sheffer et al., 2009; Stange et al., 2010; Watanabe et al., 2011). These samples were from a total of 311 patients. Additionally, in primary tumors, increased PKLR expression was associated with the presence of metastatic disease as well as the development of liver metastases (Figure 2.13). These tumors totaled 284 in sum. Taken together, PKLR expression in human samples is consistent with a positive role for PKLR in the development of liver metastasis and in agreement with functional experimental data.

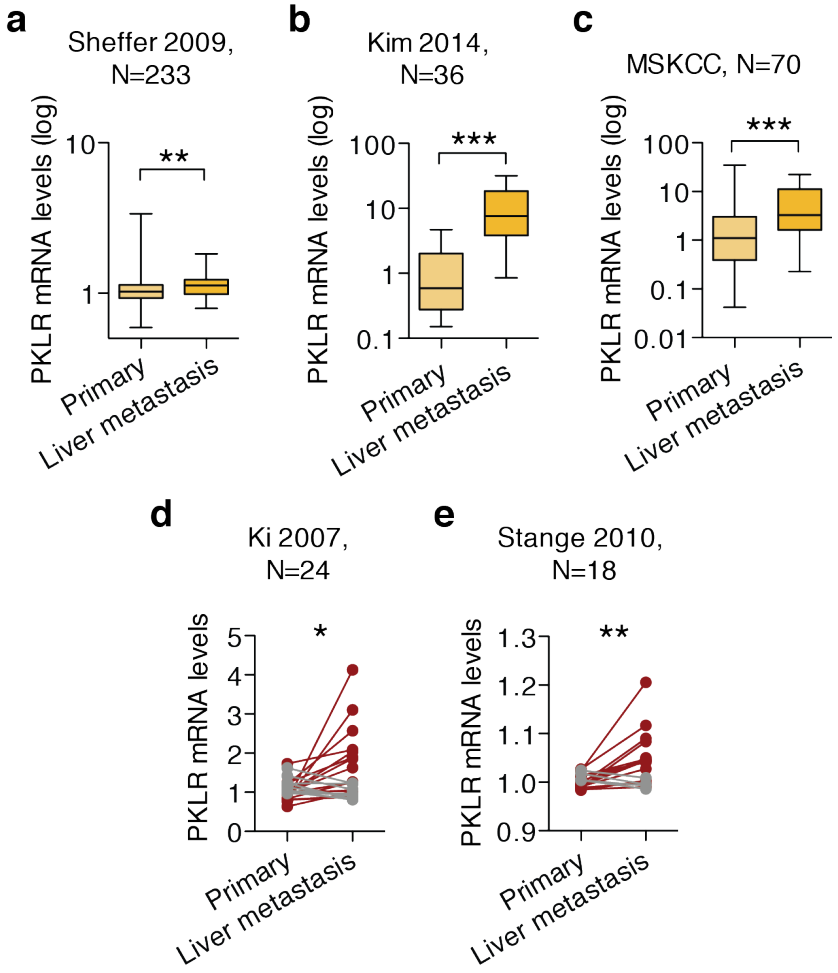

Figure 2.12: PKLR expression is increased in human liver metastases

PKLR expression as measured by microarray (**a**), RNA-sequencing (**b**), and quantitative RT-PCR (**c**) in unmatched primary and liver metastasis tumors as shown by box plots. P-values were derived using one-sided Mann Whitney test unless otherwise indicated. Error bars represent SEM. (**d**)(**e**) PKLR expression as measured by microarray in matched primary and liver metastasis tumors. Increase in PKLR expression is indicated in red; decreased expression is indicated in gray. P-values were derived using paired one-sided t-test. *=P<0.05, **=P<0.01, ***=P<0.001

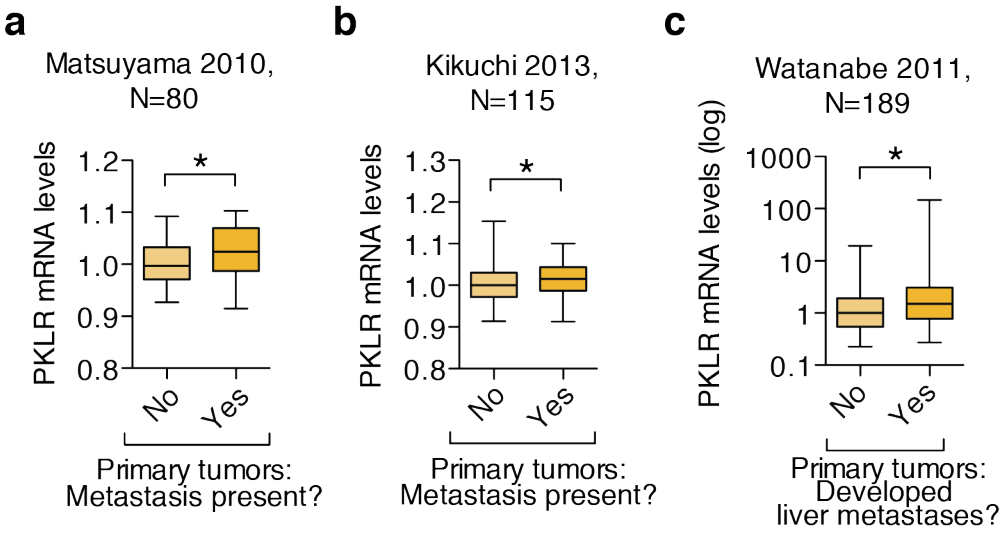

Figure 2.13: PKLR expression is increased in primary tumor samples from patients with metastases

(**a**)(**b**) PKLR expression as measured by microarray in primary tumors of patients with or without metastases present at time of sampling. (**c**) PKLR expression as measured by microarray in primary tumors of patients who were monitored over time for the development of liver metastases. P-values were derived using one-sided Mann Whitney test unless otherwise indicated. Error bars represent SEM. *=P<0.05

2.4 Cellular characterization of PKLR function

PKLR promotes cell survival during colonization in vivo

To characterize how PKLR promotes metastasis at the cellular level, two possibilities were considered. PKLR could function to promote cell growth and proliferation, or PKLR could function to prevent cell death. Since PKLR depletion had no effect on subcutaneous tumor growth or primary tumor growth (Figures 2.8, 2.10), PKLR knockdown cells were assessed for cell death *in vivo*. To assess cell death, an *in vivo* caspase-3/7-dependent bioluminescence reporter (Hickson et al., 2010) was used to measure apoptotic cell burden during metastatic liver growth. At early as well as at late time points, apoptosis was significantly increased in PKLR knockdown cells (Figure 2.14). Calculation of cumulative apoptotic burden over time revealed a significantly elevated rate of apoptosis (Figure 2.14), suggesting that PKLR is required for continuous metastatic cell survival.

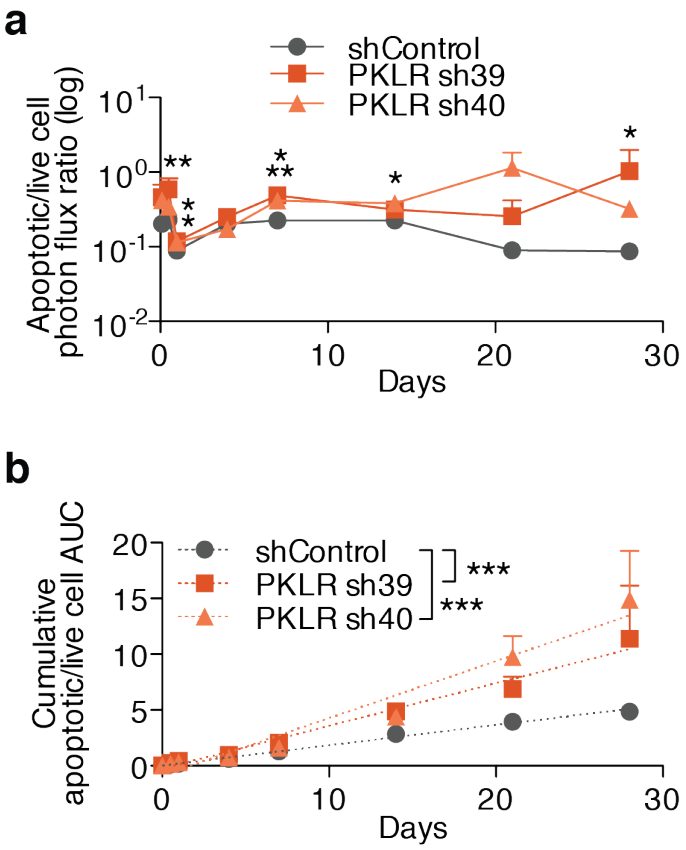

Figure 2.14: PKLR knockdown leads to increased cancer apoptosis

(**a**) 5x10^5 SW620 cells were inoculated by portal circulation injection and apoptotic cell burden in the liver was monitored using DEVD-luciferin bioluminescence relative to live cell bioluminescence over time (n=5). P-values were derived using one-sided Student's t test. (**b**) Cumulative apoptotic/live cell burden over time was determined by calculating the area under the curve for each mouse. Linear regression lines are shown. P-values were derived using ANCOVA testing for difference in slope.

To confirm these findings, livers were extracted and processed to visualize apoptotic markers in tumor nodules. PKLR-depleted cells exhibited significantly fewer tumor nodules per liver (Figure 2.15), confirming our original assessments by gross histology. Interestingly, while tumor nodules generated from control cells appeared to be highly cellular, the majority of PKLR-depleted tumor nodule cores appeared devoid of cells and composed of extracellular cleaved caspase-3 (Figure 2.15). While the reduced number of tumor nodules indicates the importance of PKLR in early colonization, enhanced apoptosis in tumor cores suggested a pathophysiological requirement for PKLR function during later stages of metastatic colonization.

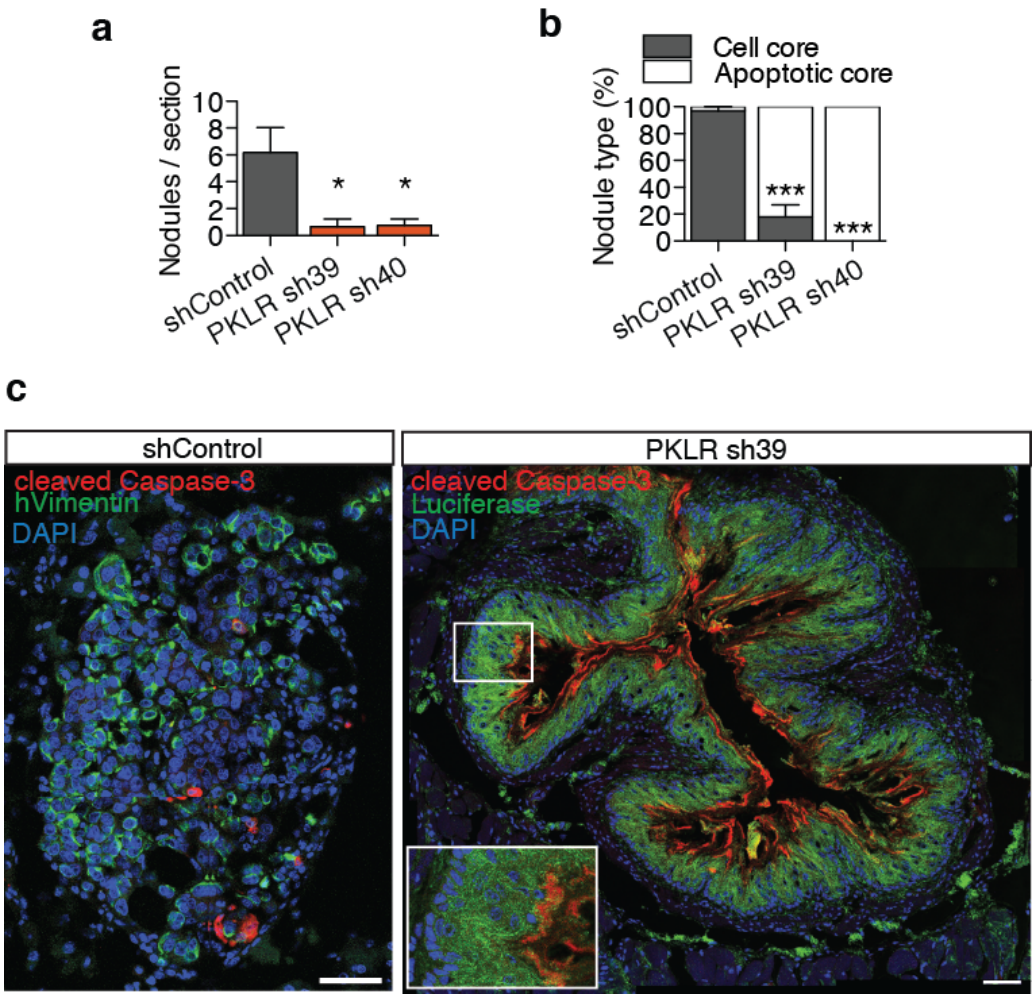

Figure 2.15: PKLR knockdown metastatic nodules display increased apoptosis in tumor core

(**a**) 5x10^5 SW620 cells were inoculated by portal circulation injection; after 28 days, livers were resected, and nodules per liver section was quantified (n=4). The average of three liver sections for each mouse was used. (**b**) For each liver, proportion of the indicated nodule type is shown. P-values were derived using two-sided Fisher's exact test on total nodule count for each PKLR shRNA compared to shControl. (**c**) Representative images of nodule types. Luciferase and human Vimentin were used to label cancer cells. Inset shows higher magnification. Scale bars = 50um.

PKLR promotes cell survival under hypoxia and high cell density in vitro

The tumor core represents a distinct microenvironment with limited nutrient supply, unique stromal composition, and altered cell-to-cell architecture (Quail and Joyce, 2013). Liver metastases are susceptible to stress-induced apoptosis (Bao et al., 2004) and demonstrate variable angiogenic capacity (Vermeulen et al., 2001), suggesting that cancer cells farther from vasculature may be especially sensitive to nutrient and oxygen deprivation. PKLR may be required for survival when cancer cells are subject to cellular stressors specific to the tumor core. To identify potential stressors, PKLR knockdown cells were grown under various conditions *in vitro* and were found to exhibit no survival defects under physiologic levels of glucose, hypoxia (1% O_2), anoxia (<0.01% O_2), high cell density, or low pH (Figure 2.16). Moreover, PKLR-depleted cells did not exhibit defects in well-established pro-metastatic phenotypes including Matrigel invasion capacity, anchorage-independent growth, and migration (Figure 2.16).

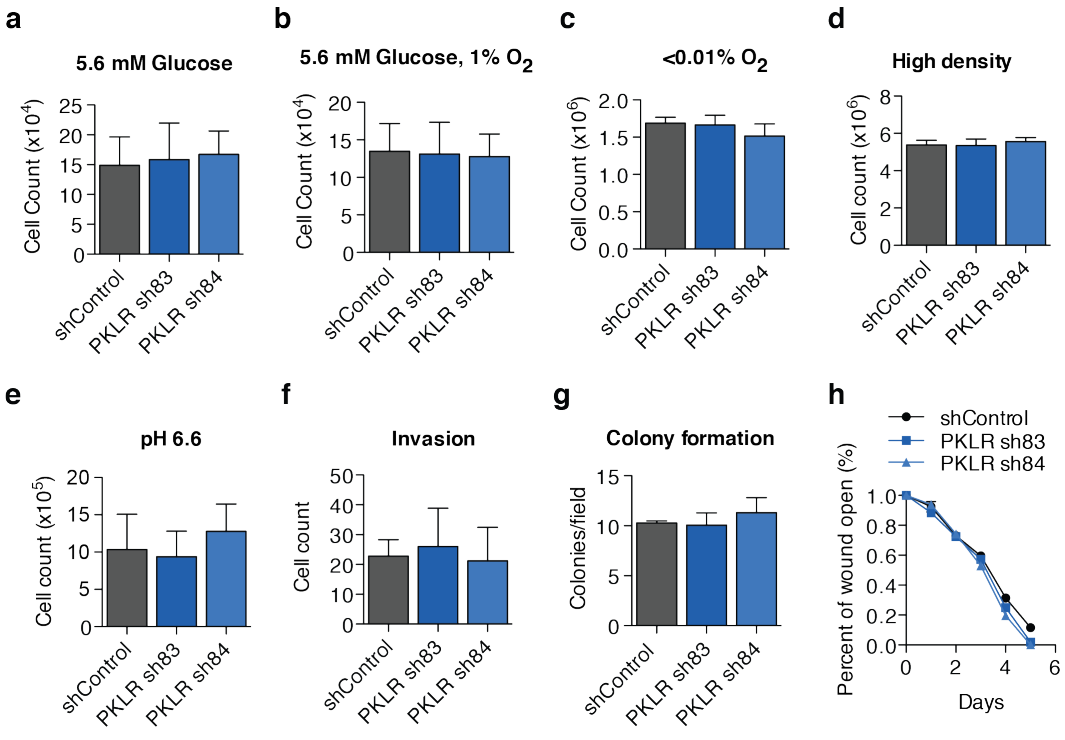

Figure 2.16: PKLR does not promote survival under various *in vitro* conditions

Cells were seeded in triplicate under the following conditions: (**a**) 10^5 cells in 5.6mM glucose for 5 days, (**b**) 10^5 cells in 5.6mM glucose & hypoxia for 5 days, (**c**) 10^6 cells in anoxia for 3 days, (**d**) 10^6 cells seeded at a density of 1000 cells/mm^2 for 3 days, (**e**) 10^5 cells in media buffered at pH 6.6 for 5 days, (**f**) $25x10^3$ cells in quadruplicate in invasion chambers and were allowed to invade for 48 hours, (**g**) $5x10^3$ cells seeded in soft agar for 14 days. Data shown is from three independent experiments. (**h**) 10^6 cells seeded overnight and scratched using a pipette tip the next day. Cells were seeded in duplicate and imaged daily.

While individual stressors elicited no phenotypic defect alone, the tumor core features multiple stressors that simultaneously impact cancer cells. To better recapitulate the microenvironment in the tumor core, cells were concurrently seeded at high cell density and in hypoxia. Under these conditions, PKLR knockdown populations displayed reduced survival—decreased number of live cells and concurrently increased number of dead cells (Figure 2.17). To characterize whether the survival defect occurs continuously or whether nutrient depletion led to a population collapse, cells were assayed daily for viability under high cell density and hypoxia. Cells with reduced PKLR expression began to display a survival defect after as early as one day under these conditions, suggesting that the effect is continuous (Figure 2.17). Given the increased apoptosis rate observed in PKLR-depleted tumors, cells were assessed for apoptosis and necrosis by flow cytometry after one day under concurrent high cell-density and hypoxia. PKLR depleted cells exhibited fewer live cells, greater proportion of early apoptotic cells, and significantly greater proportion of late apoptotic cells under this condition (Figure 2.17). These results indicated that PKLR functions to enhance cell survival under cell-dense, hypoxic conditions, appropriately simulating pathophysiological features pertaining to metastatic growth in the liver.

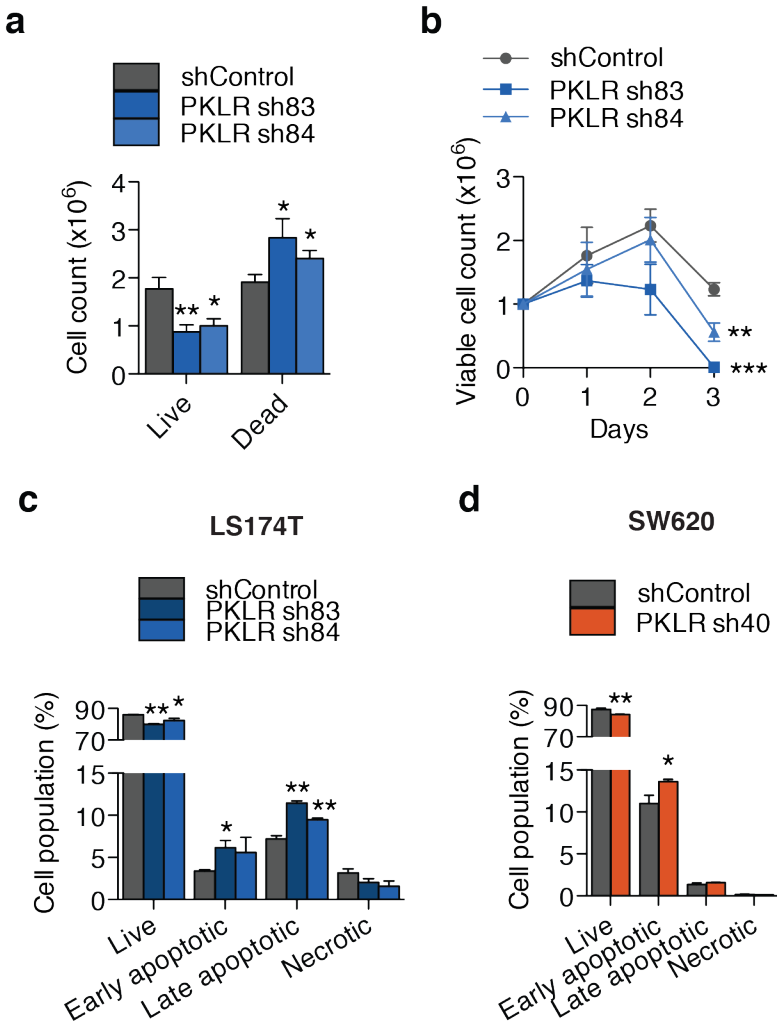

Figure 2.17: PKLR promotes survival under conditions of hypoxia and high cell density

(**a**) 10^6 LS174T cells were seeded in triplicate at a density of 1000 cells/mm^2 and were allowed to grow for 5 days in 1% O_2. Live and dead cells discriminated using trypan blue. Data shown is from four independent experiments. (**b**) 10^6 LS174T cells were seeded in triplicate at a density of 1000 cells/mm^2 and were allowed to grow in 1% O_2. Live cells were quantified using trypan blue exclusion. Data shown is from three independent experiments. 10^6 LS174T (**c**) and SW620 (**d**) cells were seeded at a density of 1000 cells/mm^2 and were assessed for apoptosis after 24 hours in 1% O_2. P-values were derived using one-sided Student's t test.

2.5 Molecular characterization of PKLR function

Identification of binding partners of PKL and PKR

Interestingly, the *in vitro* survival defect of PKLR-depleted cells was observed using cell culture media supplemented with excess pyruvate, which can be readily imported into cells (Garcia et al., 1994)—arguing against the canonical glycolytic function of PKLR in these colon cancer cells. To determine whether PKLR could be functioning non-canonically, independent of its pro-glycolytic role, PKL and PKR binding partners were identified in colorectal cancer cells using tagged proteins for co-immunoprecipitation experiments since commercial and custom antibodies were inadequate for the recognition of endogenous PKLR proteins in cell lysates. Immunoprecipitation of PKL and PKR followed by mass spectrometry revealed similar protein binding profiles and binding to many glycolytic enzymes (Figure 2.18, Table 2.1, Table 2.2), consistent with the role of pyruvate kinase in a glycolytic enzyme complex that allows for efficient lactate production (Mazurek, 2011).

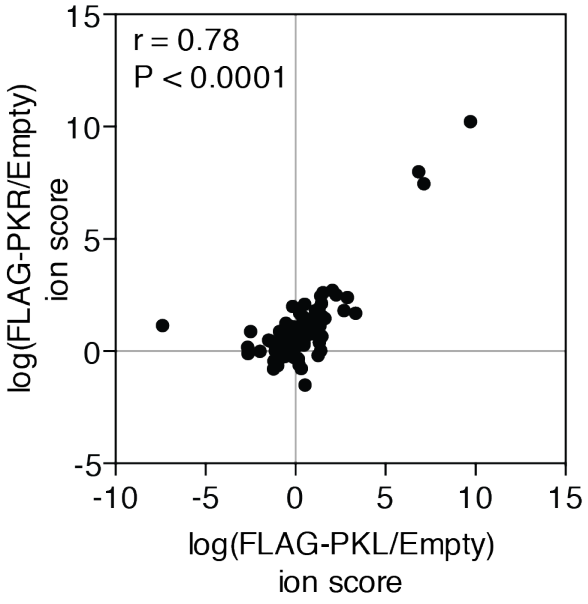

Figure 2.18: Mass spectrometry analysis of PKL & PKR co-immunoprecipitated proteins

LC-MS/MS ion scores of FLAG-PKL and FLAG-PKR co-immunoprecipitated proteins relative to empty vector in LS-LVM3b. Pearson's correlation coefficient is shown.

Table 2.1: Top gene sets scored by Gene Set Enrichment Analysis of FLAG-PKL binding partners as measured by mass spectrometric quantification

Gene set	*NES*	*NOM p-val*	*FDR q-val*
REACTOME_METABOLISM_OF_CARBOHYDRATES	1.72	0.012	0.539
REACTOME_GLUCOSE_METABOLISM	1.72	0.003	0.274
REACTOME_GLYCOLYSIS	1.70	0.004	0.278
BANDRES_RESPONSE_TO_CARMUSTIN_MGMT_48HR_DN	1.59	0.012	1

Table 2.2: Top gene sets scored by Gene Set Enrichment Analysis of FLAG-PKR binding partners as measured by mass spectrometric quantification

Gene set	*NES*	*NOM p-val*	*FDR q-val*
REACTOME_GLUCOSE_METABOLISM	1.82	0.011	0.333
REACTOME_GLYCOLYSIS	1.81	0.007	0.214
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM	1.74	0	0.359
REACTOME_METABOLISM_OF_CARBOHYDRATES	1.74	0.011	0.276

Interestingly, one of the most abundant binding partners identified was PKM, with multiple peptides matching specifically to the M2 isoform (Table 2.3, Figure 2.19)—the predominant isoform observed in colorectal cancer tissue and the only isoform observed to be expressed in the cell lines used (Bluemlein et al., 2011; Christofk et al., 2008a; Thangaraju et al., 2009). Immunoprecipitation of tagged PKL or PKR revealed an association with endogenous PKM2 (Figure 2.19). Moreover, immunoprecipitation of endogenous PKM2 revealed PKM2 to associate with endogenous PKL (Figure 2.19). This suggested the formation of a pyruvate kinase hybrid complex (Saheki et al., 1978).

Table 2.3: PKM2-specific peptides as measured by mass spectrometric quantification

	peak area		
PKM2-specific peptides	Empty	FLAG-PKL	FLAG-PKR
LAPITSDPTEATAVGAVEASFK	0	3.688E7	1.846E7
RLAPITSDPTEATAVGAVEASFK	0	2.185E6	7.746E5
EAEAAIYHLQLFEELRR	0	9.152E6	3.930E6

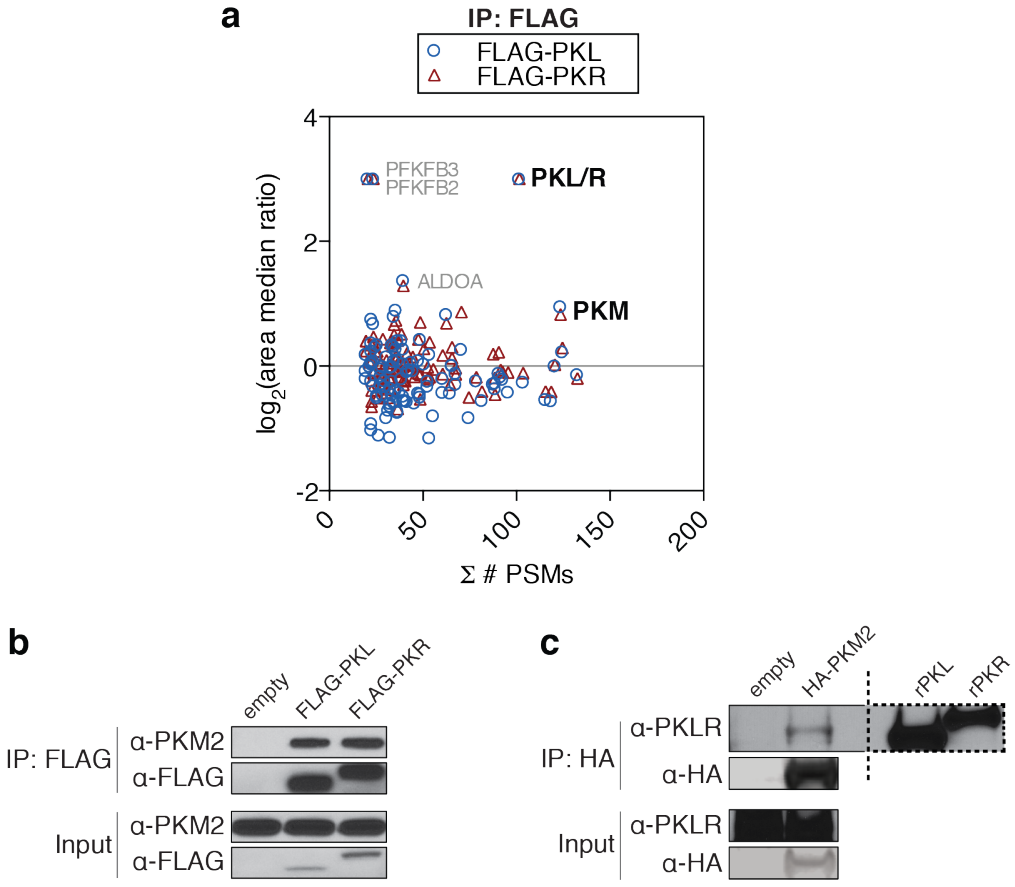

Figure 2.19: PKL binds to PKM2 in colon cancer cells

(**a**) Cell lysates from LS-LVM3b cells with indicated overexpression constructs were subjected to immunoprecipitation using anti-FLAG beads followed by LC-MS/MS. Protein peptide spectrum matches as compared to label free quantitation protein enrichment (relative to empty vector immunoprecipitation) is shown. (**b**) Cell lysates from LS-LVM3b cells were subjected to immunoprecipitation using anti-FLAG beads followed by western blot analysis. (**c**) Cell lysates from LS174T cells were subjected to immunoprecipitation using anti-HA beads followed by western blot analysis. 20 ng of recombinant PKL (rPKL) and PKR (rPKR) were analyzed as controls (dashed border).

PKLR negatively regulates PKM2 pyruvate kinase activity

Given the differing biochemical properties of pyruvate kinase isozymes and the tetrameric structure required for maximal enzymatic activity (Mazurek, 2011), PKL could function to alter PKM2 pyruvate kinase activity. To test this hypothesis, PKLR-depleted cells were grown in cell-dense, hypoxic conditions and were subject to immunoprecipitation of PKM2 complexes. PKM2 complexes isolated from PKLR-depleted cells exhibited enhanced pyruvate kinase activity relative to control cells (Figure 2.20), indicating that PKLR negatively regulates PKM2 pyruvate kinase activity. Additionally, whole cell lysates from PKLR-depleted cells displayed an increase in total pyruvate kinase activity as well as decreased phosphoenolpyruvate-to-pyruvate ratio, consistent with decreased reaction substrate and increased reaction product driven by increased catalytic activity, following incubation under hypoxic, cell-dense conditions (Figure 2.20). These findings revealed that in colon cancer cells, PKLR expression can be utilized to negatively regulate PKM2 pyruvate kinase activity and deregulate metabolism to allow for cancer progression.

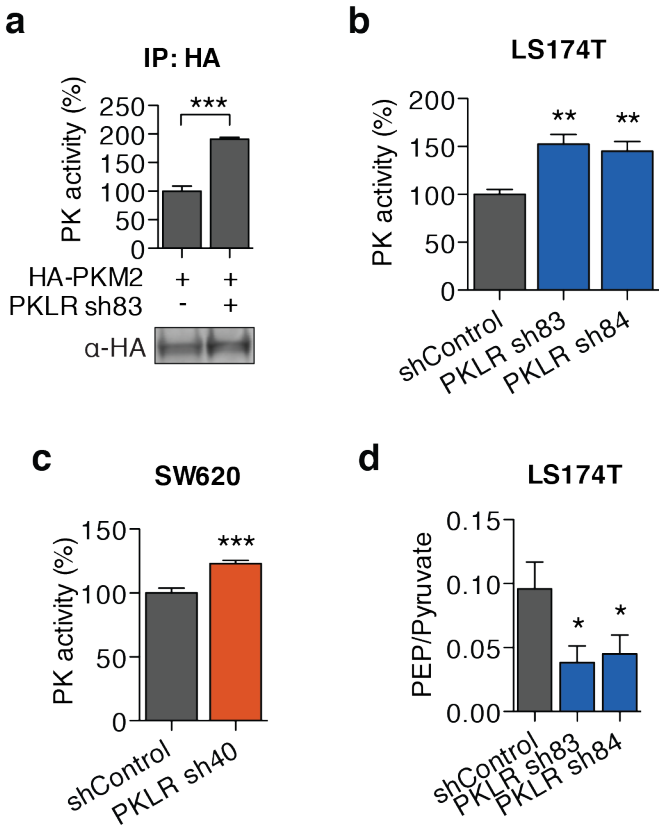

Figure 2.20: PKLR negatively regulates PKM2 pyruvate kinase activity
(**a**) Immunoprecipitated PKM2 complexes from LS174T cells, subjected to 16 hours in high cell-density and hypoxia, were assayed for pyruvate kinase activity. Activity was normalized to immunoprecipitated HA-PKM2 as measured by western blotting. Data shown is from three biological replicates. P-value was derived using two-sided Student's t test. Pyruvate kinase activity was measured from LS174T (**b**) and SW620 (**c**) cell lysates after 24 hours under hypoxic, cell-dense conditions. Activity was normalized to live cells. P-values were derived using one-sided Student's t test. Data shown is from six biological replicates and two independent experiments. (**d**) Pyruvate levels and phosphoenolpyruvate (PEP) levels were measured from LS174T cell lysates after 24 hours under hypoxic, cell-dense conditions. P-value was derived using one-sided t-test. Each experiment was performed in triplicate. Data shown is from four independent experiments.

PKLR increases glutathione levels

PKM2 is essential for aerobic glycolysis observed in cancer cells, which leads to preferential lactate production and decreased mitochondrial metabolism (Christofk et al., 2008a). Previous studies have demonstrated an advantage conferred by decreasing PKM2 activity in cancer cells, namely the diversion of metabolic flux towards anabolic pathways, which promote biosynthetic processes (Christofk et al., 2008b; Lunt et al., 2014) and intracellular reducing power under oxidizing conditions through the pentose phosphate pathway (Anastasiou et al., 2011). Because PKLR knockdown cells demonstrated no proliferative defect under oxygenated conditions (Figure 2.8) as seen with impaired biosynthetic processes, PKLR knockdown cells were assessed for depletion of intracellular glutathione, the main cellular antioxidant (Franklin et al., 2009), under hypoxic, cell-dense conditions. Indeed, PKLR knockdown cells demonstrated decreased glutathione levels (Figure 2.21).

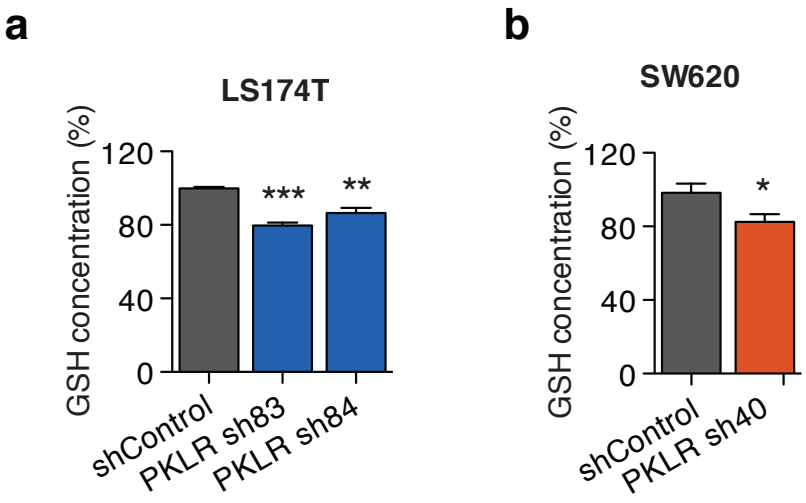

Figure 2.21: PKLR increases glutathione levels

Glutathione levels were measured using ThiolTracker from LS174T (**a**) and SW620 (**b**) cell lysates after 24 hours under hypoxic, cell-dense conditions. P-values were derived using one-sided t-test. Data shown is from three biological replicates.

Restoration of glutathione levels in PKLR-depleted cells rescues metastatic survival

To confirm glutathione depletion is responsible for the cell survival defect, supplementation with molecules that can restore glutathione levels should rescue cell survival under hypoxic and cell-dense conditions as well as metastatic colonization. PKLR knockdown cells incubated in hypoxic and cell-dense conditions were supplemented with the antioxidant *N*-acetyl cysteine (NAC), a pro-drug to a glutathione precursor, and glutathione-monoethylester (GSH-MEE), a membrane-permeable glutathione analog as glutathione does not readily cross the cell membrane. Treatment with these compounds rescued cell viability and decreased apoptosis in PKLR-depleted cells *in vitro* (Figure 2.22). Consistent with these *in vitro* findings, treatment of mice bearing PKLR knockdown cells with NAC in the drinking water partially rescued metastatic liver colonization and nearly completely restored the apoptosis rate to the levels of control cells (Figure 2.23).

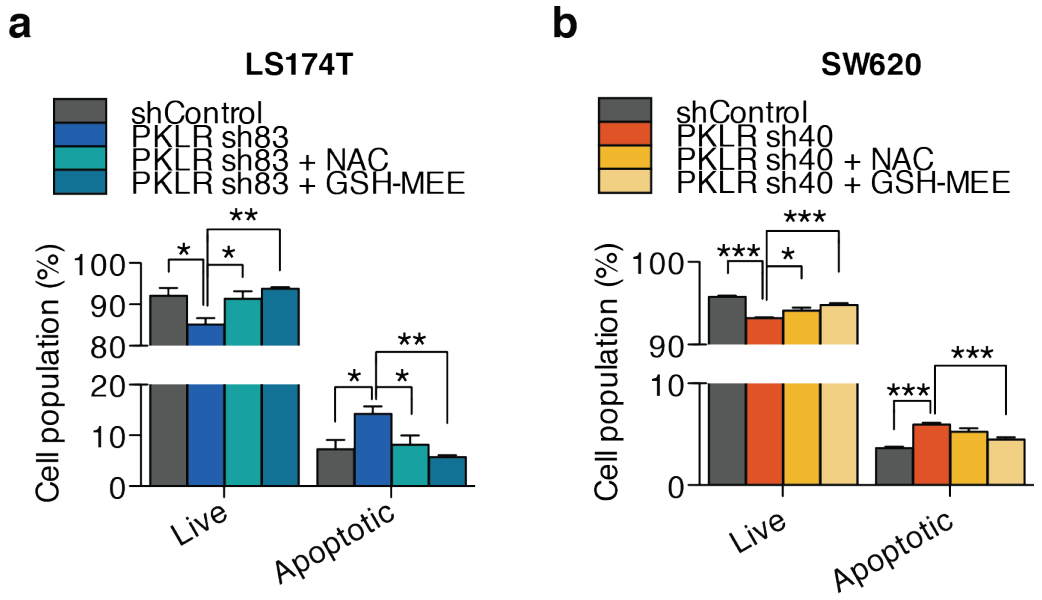

Figure 2.22: Restoration of glutathione levels rescues cell survival *in vitro*

LS174T (**a**) and SW620 (**b**) cells were assayed after 24 hours under hypoxic, cell-dense conditions for live and apoptotic cells by flow cytomentry. N-acetyl cysteine (NAC) and glutathione-monoethylester (GSH-MEE) were added to culture medium. P-values were derived using one-sided t-test. Data shown is from three biological replicates.

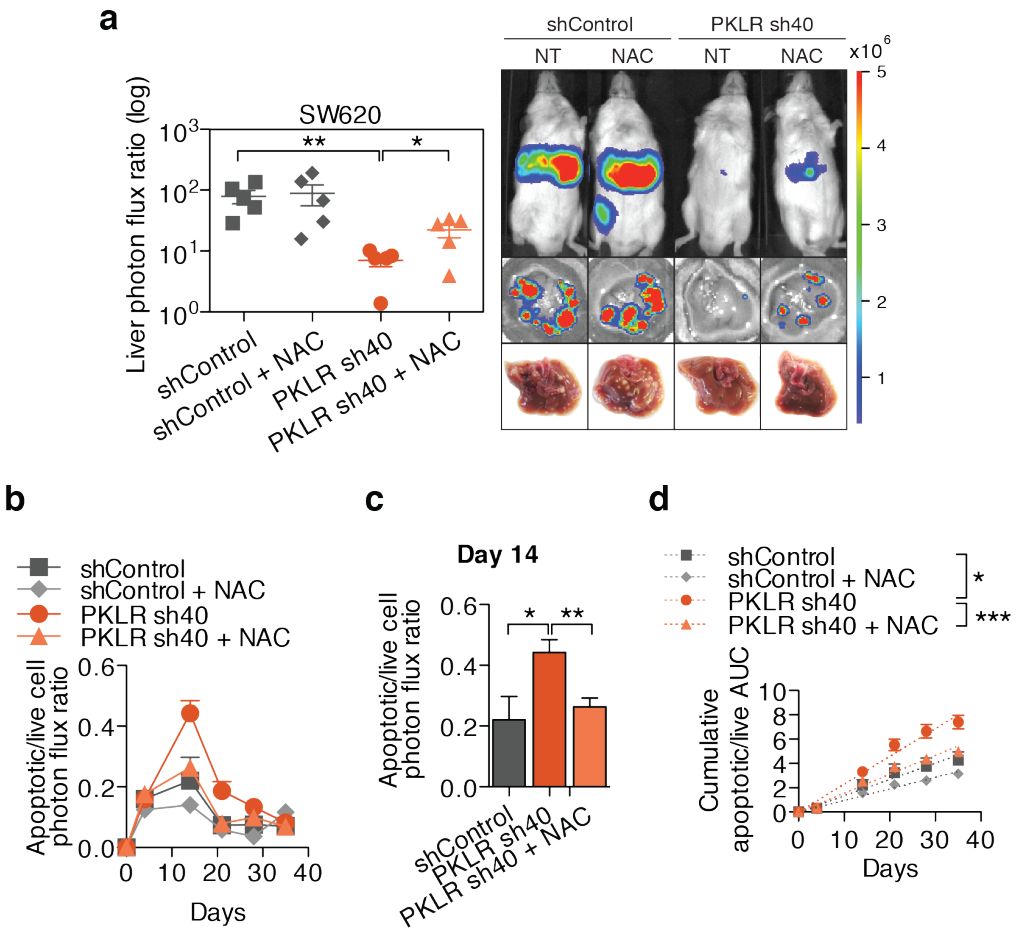

Figure 2.23: Restoration of glutathione levels rescues metastatic survival

(**a**) 5x10^5 SW620 cells were inoculated by portal circulation injection, mice were treated with NAC or control drinking water, and metastatic colonization was measured by liver bioluminescence after 35 days (n=5). P-values were derived using one-sided Mann Whitney test. Representative mouse bioluminescence, liver bioluminescence and gross histology are shown. (**b**)(**c**) Apoptotic cell burden was monitored by DEVD-luciferin liver bioluminescence relative to total live cell bioluminescence over time. Day 14 is shown in (**c**). (**d**) Cumulative apoptotic/live cell burden over time was determined by calculating the area under the curve for each mouse. Linear regression lines are shown. P-values were derived using ANCOVA testing for difference in slope.

2.6 Therapeutic inhibition of glutathione synthesis

GCLC expression is associated with liver metastasis in patients

Given the susceptibility of liver metastases to glutathione depletion, glutathione synthesis might serve as a critical pathway utilized for metastatic progression. GCLC, the catalytic subunit of glutamate cysteine ligase, is responsible for the rate-limiting reaction of glutathione synthesis (Griffith, 1999) and thus, would be an appropriate molecular target for therapeutic inhibition. To determine whether GCLC may play a role in colorectal cancer metastasis to the liver, transcript expression levels were analyzed in patient samples. GCLC expression levels were found to be significantly higher in human colorectal cancer liver metastases relative to primary tumors (Figure 2.24).

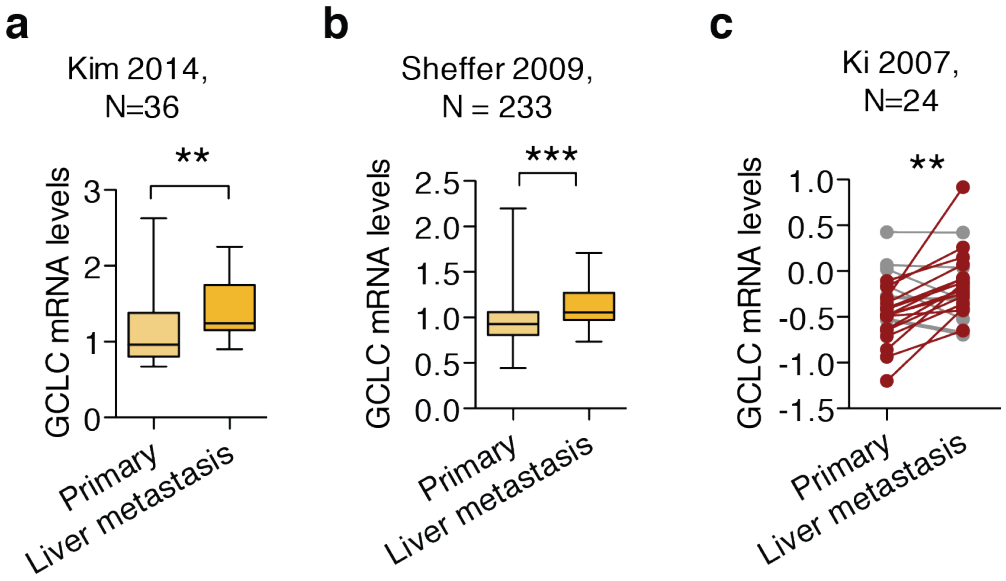

Figure 2.24: GCLC expression is increased in human liver metastases
GCLC expression as measured by RNA-sequencing (**a**) and microarray (**b**) in unmatched primary and liver metastasis tumors as shown by box plots. P-values were derived using one-sided Mann Whitney test. (**c**) GCLC expression as measured by microarray in matched primary and liver metastasis tumors. Increase in GCLC expression is indicated in red; decreased expression is indicated in gray. P-values were derived using paired one-sided t-test.

Functional validation of GCLC as promoter of liver metastasis

To determine if GCLC is functionally required for metastasis in a manner consistent with the cell survival function of PKLR, functional studies were performed using cells depleted of GCLC by RNAi. GCLC expression was depleted to 27% & 13% of control in LS174T cells and 41% & 26% of control in SW480 cells (Figure 2.25). Functional studies *in vitro* revealed GCLC to be required for cell survival under hypoxia and cell-dense conditions (Figure 2.25). Functional studies *in vivo* revealed GCLC to be required for metastatic liver colonization and to suppress cancer cell apoptosis (Figure 2.26), consistent with the observed phenotypes seen with PKLR knockdown.

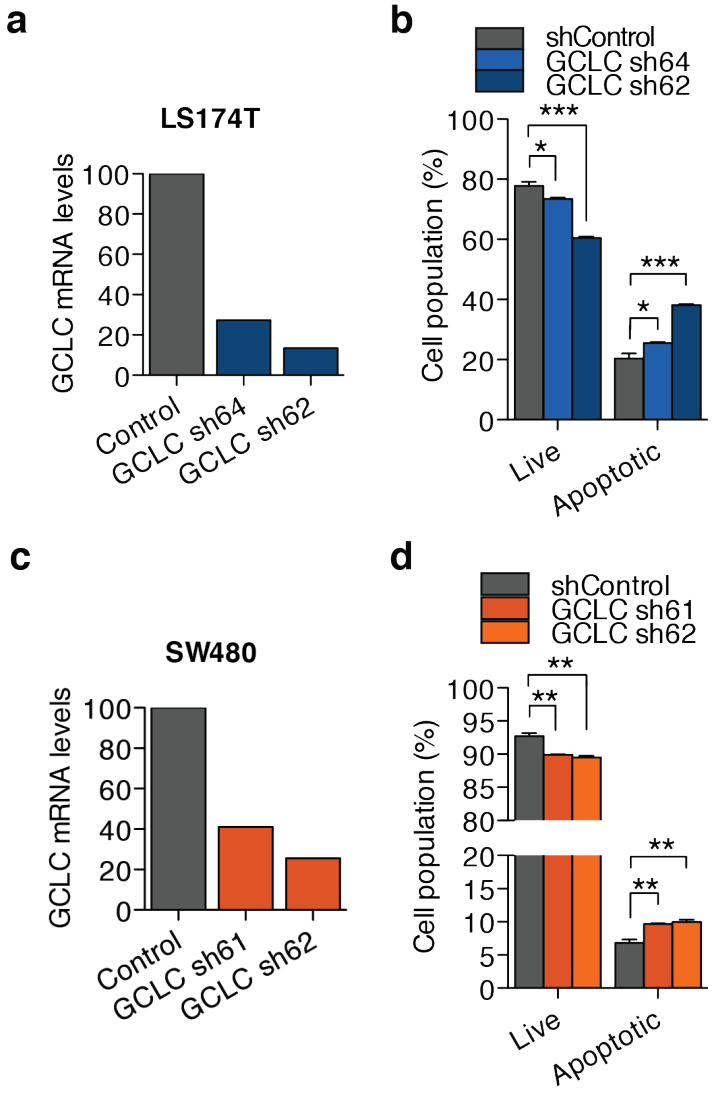

Figure 2.25: GCLC promotes survival under conditions of hypoxia and high cell density

GCLC knockdown as measured by qRT-PCR in LS174T cells (**a**) and SW620 (**c**). 10^6 LS174T (**b**) and SW480 (**d**) cells were seeded at a density of 1000 cells/mm^2 and were assessed for apoptosis after 24 hours in 1% O_2. P-values were derived using one-sided Student's t test.

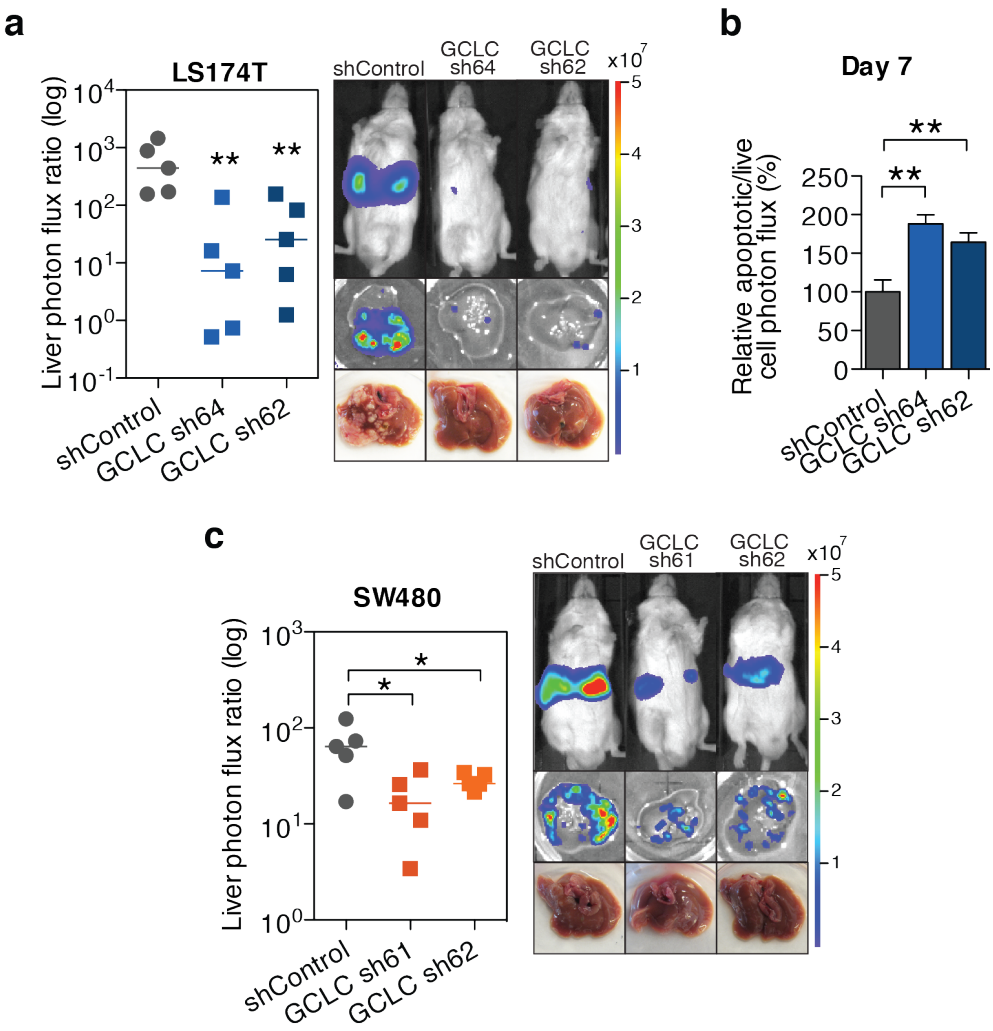

Figure 2.26: GCLC promotes metastatic liver colonization

5x10^5 LS174T (**a**) and SW480 (**c**) cells were inoculated by portal circulation injection and metastatic colonization was measured by liver bioluminescence after 21 days (LS174T) or 30 days (SW480). Representative mouse bioluminescence, liver bioluminescence and gross histology are shown. n=5. (**b**) DEVD-luciferin bioluminescence relative to live cell bioluminescence was measured in the liver at day 7 (n=5). P-values were derived using one-sided Student's t-Test.

Small molecule inhibition of GCLC suppresses metastasis

The findings that GCLC promotes metastatic colonization suggested that small molecules that block glutathione synthesis may exhibit metastasis-suppressive effects (Schulze and Harris, 2012). To test this hypothesis, colon cancer cells were first inoculated into the portal circulation of mice, and then L-buthionine-(S,R)-sulfoximine (BSO), a small molecule inhibitor of GCLC (Griffith, 1982), was added to the drinking water. Therapeutic delivery of BSO suppressed metastatic colonization and increased cancer cell apoptosis in colon cancer cell lines (Figure 2.27). Additionally, BSO suppressed metastasis of a primary cell line recently obtained from a patient's colorectal cancer tumor and passaged in mice (Figure 2.28). These results highlight the importance of cancer cell glutathione levels for metastatic survival in the liver and identify this metabolic pathway as a viable therapeutic target.

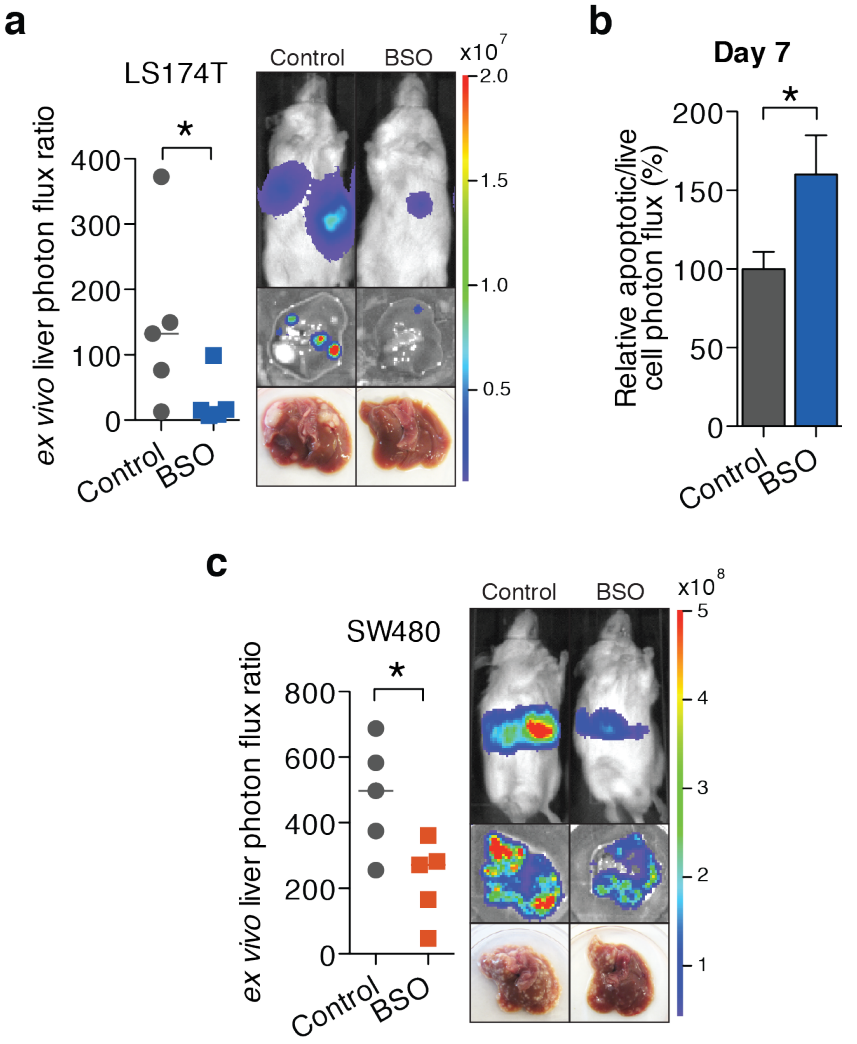

Figure 2.27: BSO suppresses metastasis in colon cancer cell lines

$5x10^5$ LS174T (**a**) and SW480 (**c**) cells were inoculated by portal circulation injection, and mice were provided BSO in drinking water. Metastatic colonization was measured by liver bioluminescence after 28 days (LS174T) or 35 days (SW480). Representative mouse bioluminescence, liver bioluminescence and gross histology are shown. n=5. (**b**) DEVD-luciferin bioluminescence relative to live cell bioluminescence was measured in the liver at day 7 (n=5). P-values were derived using one-sided Student's t-Test.

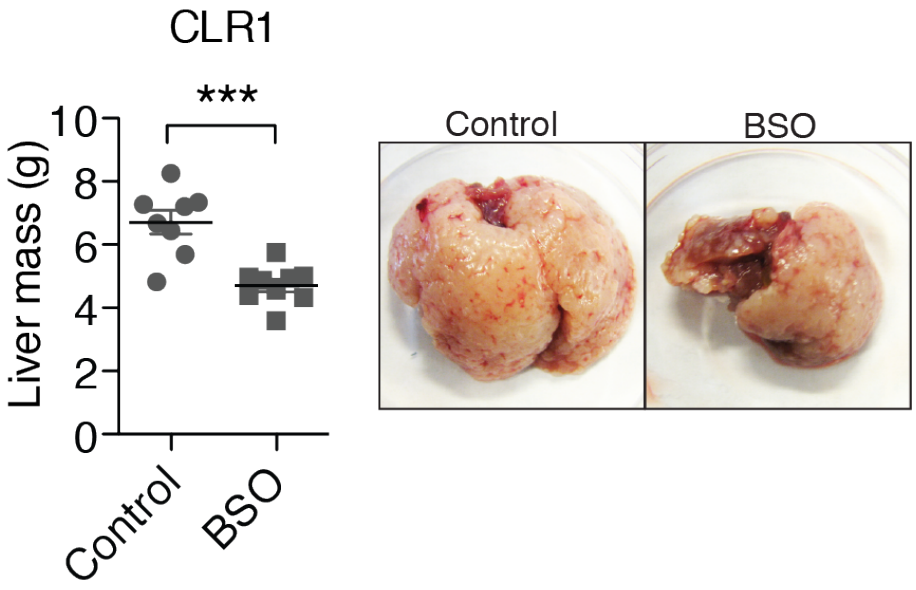

Figure 2.28: BSO suppresses metastasis in primary colon cancer cell line
$7.5x10^5$ CLR1 cells were inoculated by portal circulation injection, and mice were provided BSO in drinking water. Metastatic colonization was measured by liver mass after 28 days. Representative gross histology is shown. n=8.

2.7 Discussion

Glycolytic deregulation in cancer

Metabolic deregulation has recently emerged as a hallmark of cancer (Cairns et al., 2011; Hanahan and Weinberg, 2011; Kaelin and McKnight, 2013). Following Warburg's observation of cancer cell dependence on aerobic glycolysis (Warburg, 1956), molecular dissection of glycolytic control has revealed the proliferative benefits of gene expression changes under hypoxia as well as metabolite diversion (DeBerardinis et al., 2008; Semenza, 2007; Vander Heiden et al., 2009). The pyruvate kinase M2 isoform enzymatic catalysis is a key regulatory event for glycolysis, controlling both glycolytic flux for energy demands and flux towards biosynthetic pathways of nucleotides, amino acids, and reducing power for cancer cell proliferation and antioxidant generation (Anastasiou et al., 2011; Chaneton et al., 2012; Christofk et al., 2008a; Christofk et al., 2008b). Cancer cells take advantage of this regulation—PKM2 pyruvate kinase activity can be inhibited through destabilized subunit interactions and resultant loss of homotetramer structure, which is required for maximum enzymatic activity (Vander Heiden et al., 2009). Our findings reveal that colon cancer cells utilize an additional means of PKM2 regulation, namely expression of an additional isozyme PKL.

Utilization of pathways endogenous to metastatic niche

The advantageous expression of liver-specific pyruvate kinase in cancer cells to allow for metastatic growth in the liver suggests that Paget's 'seed and soil' hypothesis can be explained in part by cancer cell utilization of pathways endogenous to cells at the metastatic niche (Fidler, 2003). The PKL isozyme is essential for hepatocytes as the sole enzyme for glycolytic pyruvate production and regulated flux through alternative pathways. This importance is highlighted by regulation of hepatic PKL activity and expression in response to glucagon, insulin and dietary carbohydrates (Yamada and Noguchi, 1999). Since hepatocytes experience significant oxidative stress as a result of hypoxemic blood supply and xenobiotic metabolism (Jungermann and Kietzmann, 1996), antioxidant supply and controlled flux towards glutathione regeneration is likely critical for hepatocyte survival and function. Future studies into the regulatory role of PKL in hepatocytes will clarify this enzyme's function in normal physiology. While early metastatic colonization can be achieved through utilization of organ-specific nutrients (Loo et al., 2015), cancer cell activation of pathways endogenous to the metastatic niche appears to contribute to continued progression of metastatic colonies.

Whereas the liver actively promotes high pentose phosphate pathway activity in mammals (Cabezas et al., 1999), colon cancer cells modulate glycolysis through expression of PKLR, consistent with the enhancement of pentose phosphate pathway flux and regeneration of glutathione required for survival in the liver. The importance of glutathione levels is highlighted by further, independent promotion of

glutathione levels: increased expression of GCLC, the rate-limiting reaction in glutathione synthesis. While GCLC and PKLR function appears to be critical regulatory steps that are broadly utilized by colon cancer metastases as suggested by clinical expression association, other mechanisms may be additionally utilized to ensure sustained glutathione levels such as activation of hepatocyte glutathione efflux to allow for the import of glutathione precursors from plasma (Obrador et al., 2011). Nevertheless, glutathione levels appear critical for metastatic cell survival and represent a potential clinical target.

Therapeutic potential of metabolic modulation

Therapeutic inhibition of glutathione synthesis can suppress liver colonization through increasing cancer cell apoptosis, suggesting that targeting this pathway has potential to induce cell death of already-formed metastatic nodules. BSO has been previously tested in a phase I clinical trial on a variety of advanced cancers as an adjuvant and has demonstrated both significant glutathione suppressive activity as well as minimal toxicity (Bailey et al., 1997). Additionally, the development of more potent inhibitors of GCLC may allow for enhanced efficacy (Tokutake et al., 1998). Given that small-molecule activation of PKM2 is being investigated as a strategy to suppress tumor growth (Anastasiou et al., 2012), the findings presented here suggest that glycolytic activation might display additional apoptotic effects in liver metastasis. The poor survival of patients with metastatic disease and the need for therapies specific to metastasis warrants further development of metabolic modulation as a therapeutic approach.

2.8 Materials and Methods

Cell culture

LS174T, LS-LVM3b, SW620, SW480, and WiDr cell lines were propagated as previously described (Loo et al., 2015) and were obtained from ATCC except for *in vivo*-selected LS-LVM3b derived in the laboratory. All cell lines were labeled with luciferase reporter construct for *in vivo* bioluminescent monitoring (Hickson et al., 2010). LS174T, LS-LVM3b, and WiDr cell lines were maintained in DMEM media supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin and Fungizone. SW620 and SW480 were maintained in McCoy's media supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin and Fungizone. The cell lines used were selected to represent the mutational spectrum most commonly seen in patients (Kras wild-type/mutant, Braf wild-type/mutant, MSS/MSI). Cells in culture were routinely tested for mycoplasma contamination. Standard cell culture conditions for these cell lines are considered to be cell seeding at a density of approximately 50-100 cells/mm^2, growth at 37°C and 5% CO_2, and dissociation and dilution of cells before reaching confluency.

shRNA screening

10 subpools of lentiviral particles from the TRC1 human shRNA pooled library (Sigma), consisting of 71,444 shRNA clones representing 14,523 genes, were transduced into luciferase-expressing LS174T, WiDr and

SW620 colon cancer cell lines. Two transductions for each cell line were performed as biological replicates at a low titer (MOI < 1) to reduce the likelihood of multiple shRNAs in a single cell. Transductions were performed with 8ug/ml of Polybrene (Millipore) to enhance efficiency. 48 hours after transductions, puromycin was added at a concentration of 2ug/ml until un-transfected cells died to select for cells with stable integrants. After antibiotic selection, the remaining cells were allowed to recover for a week prior to subsequent experiments. A portion of the selected cells were set aside and genomic DNA extracted. This is the reference pool of genomic DNA prior to the selective pressure of liver colonization. A second population of the cells was used for *in vivo* experiments, while a third was kept in culture for the duration of the *in vivo* experiment. For direct liver injections, 800K cells from each transduced population containing a single subpool was inoculated into an average of 7 mice, summing to 204 mice in total for all subpools and cell lines and allowing for a 783x depth of coverage. After 3 to 4 weeks when tumors had developed as measured by bioluminescence (approximately 10^9-10^{10} photons/s), mice were sacrificed, and tumors were resected. Cells from the various conditions were processed by isolation of genomic DNA using DNeasy kit (Qiagen). A total of 5ug of genomic DNA isolated from tumors from different mice were pooled and PCR amplification of shRNA inserts was performed using 10 reaction tubes and pooled. A first amplification was performed using touchdown PCR (Fwd_rd1: TGGACTATCATATGCTTACCGTAACT;Rev_rd1:
AAAGAGGATCTCTGTCCCTGT. The PCR product from all 10 reactions

(~350 bp) were gel-purified and pooled. Subsequently, a second round of amplification was performed with Illumina-sequencing specific primers (Fwd_rd2: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCTGTATTCTTGGCTTTATATATCTTGTGGA AAGGAC,
Rev_rd2: CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGGATG AATACTGCCATTTGTCTCGAGGTCGA). The subsequent PCR product was gel purified (~300bp) in preparation for high-throughput sequencing on Illumina HiSeq2000. Once the sequencing results were obtained, the data was filtered to normalize samples by total reads and to remove shRNAs that were not included in all reference samples. 54,591 shRNA clones representing 14,095 genes were used in the final data analysis. For each sample, shRNA ratios to the reference cell line were calculated, and z-scores were calculated using peak median absolute deviation (PMAD) (Luo et al., 2008) to normalize for the global loss of shRNAs under these experimental conditions. Given the effects of off-target RNAi silencing, genes were scored as hits only if at least two shRNAs in each cell line were absent from the final tumor samples in both independent transduction replicates. The secondary library was generated by cloning a top-scoring shRNA from the large-scale screen into plko.1 and producing a viral library that was subjected to the experimental procedures listed above. Gene scoring by RIGER algorithm was performed using weighted-sum method on shRNAs in all cell lines. Gene ranking was averaged across cell lines, which was then used to calculate gene percentile. To identify kinases, genes from the top 556 hits were selected if annotated as a kinase

by GO annotation KINASE_ACTIVITY (GO: 0016301) and mean depletion in the secondary screen was greater than 2 z-scores. These kinases were then assessed for their gene percentile score by RIGER.

Stable cell lines

Virus was generated using ViraSafe lentiviral packaging system (Cell Biolabs). shRNA plasmids used were obtained from Sigma TRC library. Indicated shRNAs (Sigma) are as follows: shControl (SHC002), PKLR sh83 (TRCN0000006383), PKLR sh84 (TRCN0000006384), PKLR sh39 (TRCN0000199139), PKLR sh40 (TRCN0000194740), GCLC sh64 (TRCN0000333564), and GCLC sh62 (TRCN0000344862). Because shRNAs displayed varying efficacy depending on the cell line used, shRNAs used for experiments were selected among multiple for the best-depleted efficiency. To generate overexpression, N-terminal FLAG tag or HA tag was added to cDNA of PKL/PKR or PKM2, respectively, and cloned into pBabe-Puro or pBabe-Hygro expression vector. Transductions were performed with 8ug/ml Polybrene (Millipore).

Animal studies

Animal experiments were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee at The Rockefeller University. NOD-SCID male mice, aged 6 weeks, were used in experiments involving direct liver injections and LS-LVM3b cells. NOD-SCID gamma male mice, aged 6-8 weeks, were used in all other

experiments. For direct liver injections, cells in PBS were mixed with Matrigel (Corning) at a 1:1 ratio and injected with a total volume of 20ul into the most anterior lobe. For subcutaneous injections, cells in PBS were mixed with Matrigel (Corning) at a 1:1 ratio and injected with a total volume of 100ul bilaterally near hind limbs. For portal circulation injections, cells in PBS were injected into the spleen with the direction of circulatory flow in a volume of 50ul. After 30 seconds to allow for complete exit of cells from spleen, the spleen was removed by cautery. For cecal injections, cells in PBS were mixed with Matrigel (Corning) at a 1:1 ratio and injected with a total volume of 20ul into the cecal wall. To monitor tumor burden, mice were inoculated with 1.67mg D-luciferin (Perkin Elmer) in 100ul by suborbital injection, and photon flux was immediately measured in IVIS Lumina System (Perkin Elmer). Bioluminescence was performed immediately following cell inoculation with an exposure time of 5 min and weekly thereafter unless otherwise noted with an exposure time of 1 min. Tumor burden over time between groups was performed by normalizing to matched bioluminescence signal immediately following cell inoculation. To monitor apoptosis, mice were inoculated with 0.83mg of Z-DEVD-aminoluciferin (VivoGlo Caspase 3/7 Substrate, Promega) in 100ul by suborbital injection, and photon flux was immediately measured with an exposure time of 5 minutes in IVIS Lumina System as a measure of early apoptotic cells. 3 hours later, total tumor burden was measured using D-luciferin as described above. Relative apoptotic cell burden at a given time point was calculated by taking the ratio of bioluminescence signal obtained using Z-DEVD-

aminoluciferin over the bioluminescence signal obtained with D-luciferin. To calculate an apoptosis rate, for each mouse, the relative apoptotic cell burden was plotted over time, and the area-under-the-curve (AUC) was calculated between t=0 and each time point. The calculated AUC data points were plotted over time, and a linear regression was performed. For N-acetyl cysteine treatments, mice were providing 40mM NAC drinking water with the pH matched to control water and the solution replaced twice weekly. BSO treatment studies were performed by supplying BSO in drinking water at 20mM, a dose previously demonstrated to display no significant toxicities in mice (Watanabe et al., 2003), and the solution was replaced twice weekly. Mice were randomized following cancer cell inoculation and prior to treatment.

Patient-derived primary colon cancer graft

CLR1 was derived from the liver metastasis of a subject with recurrent rectal adenocarcinoma with a KRAS G12 mutation. Within two hours of surgical resection, tumor tissue not needed for diagnostic purposes was implanted subcutaneously into two NSG mice at the Antitumor Assessment Core at Memorial Sloan-Kettering Cancer Center. When the tumor reached 1,000 mm^3, mice were euthanized, tumor was excised, and a 100 mm^3 piece of tumor tissue was reimplanted into another set of mice at Rockefeller University. To maintain the xenograft line, the subcutaneous tumor was passaged each time the tumor reached 1,000 mm^3. The cells used for the splenic injection of CLR1 had been passaged

subcutaneously six times prior. To obtain a single-cell suspension for injection, mice were euthanized, and the subcutaneous tumor was excised and subjected to a series of mechanical and enzymatic separations to obtain a single-cell suspension. Cells of mouse origin were depleted from the single-cell suspension via magnetic-activated cell sorting using the mouse depletion kit (Miltenyi Biotec), leaving only cells derived from human colorectal cancer.

Quantitative RT-PCR

RNA was extracted using total RNA isolation kit as indicated by manufacturer (Norgen Biotek). cDNA was generated using Superscript III (Invitrogen). Fast SYBR Green Master Mix (Life Technologies) was used to analyze samples on Applied Biosystems 7900HT. Expression was normalized to HPRT expression. Primers sequences are as follows: HPRT-F: GACCAGTCAACAGGGGACAT, HPRT-R: CCTGACCAAGGAAAGCAAAG, PKLR-F: TGGGAAAACTGGGTGGGATGGATG,
PKLR-R: GAAGGAAGCAGCCGGGGATTTGAC, GCLC-F: GACCCATGGAGGTGCAATTA , GCLC-R: AACCTTTGACAGTGGAATGAGA

Clinical analysis

The following datasets were used for analysis: GSE41258, GSE50760, GSE6988, GSE14297, GSE18105, GSE27854, and GSE14095. RNA-

sequencing data was normalized to total reads per sample. Affymetrix microarrays were analyzed for PKLR expression using probe 210451_at, which displayed the highest counts.

Immunofluorescence of frozen sections

Livers were harvested from mice, washed with PBS, and fixed in 4% paraformaldehyde overnight. The following day, livers were washed with PBS, submerged in 30% sucrose until fully saturated, and then frozen in OCT solution on dry ice and stored at -80 degrees Celsius. Frozen liver blocks were sliced on tissue microtome in 10um sections and treated with methanol:acetone on slides. Slides were stained with the following antibodies: anti-Luciferase (1:100, Pierce), anti-cleaved Capase-3 (1:1000, Cell Signaling), and anti-hVimentin (1:40, Vector laboratories). Fluorescent secondary antibodies (1:250, Invitrogen) were applied for labeling of primary antibodies. DAPI was added prior to mounting. Slides were imaged on Leica TCS SP5 system.

Hypoxia and high cell density assay

Hypoxia was performed using 1% O_2 5% CO_2, 94% N_2 mixed gas (Praxair) in a Modular Incubator Chamber (Billups-Rothernberg). High cell-density was defined as a seeding density of 1000 cells/mm^2. Experiments were performed in 6-well or 12-well plates using 2mL and 1mL of media, respectively.

Flow cytometry

All experiments were done after 24 hours under conditions of high cell-density and hypoxia. Apoptosis was assessed using GFP-certified Apoptosis/Necrosis detection kit (Enzo Life Sciences). AnnexinV$^+$ cells were considered apoptotic, and AnnexinV$^-$, 7-AAD$^-$ cells were considered live. For rescue experiments, N-acetyl cysteine was pH adjusted to 7.4 and used fresh at 2.5mM. GSH-MEE (Sigma) was used at 1mM. For glutathione measurements, ThiolTracker Violet (Life Technologies) was used according to manufacturer's instructions at 10uM and analyzed in combination with 7-AAD to exclude dead cells. Cells were analyzed on LSRII (BD). Analysis was performed on FloJo. Experiments were performed in biological triplicate and were repeated at least twice.

Co-immunoprecipitation & western blotting

Cells were lysed using 50mM Tris, pH 7.5, 150mM NaCl, 1mM EDTA, 1% NP40, protease inhibitor (Roche), and phosphoSTOP (Roche). Immunoprecipitation was performed using anti-FLAG M2 magnetic beads (Sigma) or anti-HA magnetic beads (Pierce) for 1 hour at 4 degrees Celsius. For FLAG-based immunoprecipitation, beads were washed 2x with wash buffer containing 1M NaCl followed by a wash with wash buffer containing 50mM NaCl. For HA-based immunoprecipitation, beads were washed 2x with lysis buffer. Elution was performed using either 3x FLAG peptide (Sigma) or HA peptide (Pierce). Samples were reduced with reducing agent (Life Technologies), denatured, separated by SDS-

PAGE, transferred to PVDF membrane (Pierce or Millipore), blocked with 5% milk or 5% BSA, and probed with primary antibody. The following antibodies were used: anti-FLAG (1:1,000; Sigma), anti-PKM2 (1:1,000; Cell Signaling), anti-PKLR (1:75; Santa Cruz), and anti-HA (1:1,000; Cell Signaling). For PKM2 and HA detection, fluorescent secondary antibodies (1:10,000; Li-Cor) were used and detected on Odyssey SA Imaging System (Li-Cor). For PKLR and FLAG detection, antibodies were chemiluminescently detected using horseradish peroxidase–conjugated secondary antibodies (1:10,000), ECL2 Western Blotting Substrate (Pierce) and the SRX-101A (Konica Minolta) developer, according to the manufacturer's instructions.

Label Free Quantitation LC-MS/MS and analysis

Co-immunoprecipitated and 3xFLAG eluted proteins were trypsinized in-solution overnight. Peptides were desalted using home-made Empore C18 columns prior to being analyzed by LC-MS/MS (Dionex 3000 HPLC coupled to Orbitrap XL, Thermo Scientific). Peptides were separated at 300nL/min using a gradient increasing from 10% B to 45% B in 120 minutes (A: 0.1% Formic Acid, B: Acetonitrile/0.1% Formic Acid). Generated LC-MS/MS data were queried against Uniprots complete Human Proteome (August 2013) and quantitated using MaxQuant 1.5.0.30. In short, Peptide Spectrum Match false discovery rate was set to 1% while protein false discovery rate was set to 1%. A total of 1,140 proteins were matched. Match between runs were used for the label free

quantitation. Generated label free quantitation values were analyzed using Perseus 1.5.0.9. All LFQ values were log2 transformed and filtered, requiring that a given protein displayed at least five matching peptides and three times as many peptide spectrum matches (PSMs). This resulted in 109 proteins. Absent LFQ values were replaced by a value one order of magnitude less.

Pyruvate kinase assay

10^6 LS174T cells were seeded in triplicate in 6-well plates in 1% O_2 for 16 hours. Immunoprecipitation was performed as described above for HA-tagged proteins. Following elution of immunoprecipitated pyruvate kinase complex, PK activity was measured using lactate dehydrogenase (LDH)-coupled assay (Christofk et al., 2008b). Each reaction was performed under the following conditions: 50mM Tris, pH 7.5, 100mM KCl, 10mM MgCl2, 1mM PEP, 1mM ADP, 200uM NADH, 8 units LDH in 50% glycerol, and 100uM FBP. Activity was normalized to immunoprecipitated HA-tagged PKM2 levels as measured by Li-Cor quantitative western blotting. Assay was monitored for pyruvate-dependent conversion of NADH to NAD+ by fluorescence on Biotek Synergy Neo. For whole cell lysates, pyruvate kinase activity assay (Biovision) was used according to manufacturer's instructions by fluorescent measurements. Activity was normalized to relative live cell proportion as measured by cell counting with trypan blue dead cell exclusion.

Pathway analysis

Gene Set Enrichment Analysis (Subramanian et al., 2005) was performed on a list ranked by enrichment of proteins in FLAG-PKL/FLAG-PKR immuno-precipitated samples using the most abundant proteins relative to control based on the sum of peptide spectrum matches per protein. Curated C2 gene sets from version 4.0 of the MSigDB were used for analysis.

Pyurvate assay

$3x10^6$ cells were seeded in 60mm plate in triplicate and grown under 1% O_2 for 24 hours in recommended media without supplemental sodium pyruvate. Cells were washed with PBS containing 100uM phloretin and lysed and deproteinized using perchloric acid precipitation (Biovision). Pyruvate and PEP assays (Biovision) were used by fluorometric methods to quantify intracellular metabolite levels.

Recombinant protein production

PKL and PKR cDNAs were cloned into pGEX-6P-L vector, followed by transformation in BL21 (DE3) Gold (Agilent). 0.1mM IPTG was used to induce expression and after 3 hours, cells were lysed using 50mM Tris-HCl, pH 7.5, 5mM EDTA, 200mM NaCl, 0.1% NP-40, 1mM DTT, 5mg/ml lysozyme and sonicated (Misonix Sonicator S-4000). Purification was

performed using glutathione-agarose beads (Sigma) and eluted using 5mM reduced glutathione (Sigma). Cleavage of GST-tag was performed using Prescission protease (GE).

Additional in vitro cell growth assays

Anoxia was generated using the AnaeroPack System (Mitsubishi Gas Chemical). Cell culture under acidic pH was obtained by using MES buffer and adjusting to pH 6.6. Anchorage-independent growth was performed as follows: cells were suspended in 0.4% agar in cell culture media and layered on top of 0.6% agar in 24-well plates. Colonies formed with feed media changed every 4 days and were counted and imaged by light microscopy at 14 days. Invasion assays were performed by seeding cells on top of Matrigel invasion chambers (BD) and using light microscopy to image and quantify the number of cells that passed through to bottom well.

Statistics

Bootstrapping was performed by sampling with replacement in R. ANCOVA testing was performed using aov() command in R. Sample size in mice experiments was chosen based on biological variability observed with a given genotype. Non-parametric tests were used when normality could not be assumed.

CHAPTER 3: Highly Variable Cancer Subpopulations that Exhibit Metastatic Fitness and Transcriptomic Variability

3.1 Introduction

Non-genetic cell-to-cell variability in cancer

While the majority of intra-tumoral heterogeneity described to date focuses on genetic diversity, the mechanisms that govern non-genetic diversity remain poorly characterized. Much of the challenge is due to the inherent genetic instability of cancer populations. As an individual tumor progresses over time, the genetic background changes with many beneficial as well as neutral mutations incorporated into the population. Additionally, the variation from patient-to-patient hampers clinical analyses and makes controlling for mutational differences nearly impossible.

However, there exist precedence for non-genetic mechanisms to contribute to cancer progression and metastatic growth. While chemotherapy resistance has been well-characterized to be mediated by genetic mutations in a number of cases (Gerlinger and Swanton, 2010), colorectal cancer clonal populations can display significant variability in the ability to form tumors and respond to chemotherapy despite stable genetic lineages over time (Kreso et al., 2013), suggesting non-genetic mechanisms as the source of molecular variability generation (Singer et al., 2014). Other studies have observed non-genetic cell-to-cell heterogeneity that enhances fitness of cells under cytotoxic conditions with chemotherapies (Gupta et al., 2011; Kreso et al., 2013; Sharma et al., 2010; Spencer et al., 2009). Furthermore, molecular variability in the

dynamic response to chemotherapies can influence survival of cancer populations (Cohen et al., 2008).

It remains unclear whether non-genetic phenotypic diversity in cancer is molecularly regulated or whether it is simply biological noise (Marusyk and Polyak, 2010). Stochastic noise may be derived from inherently noisy processes such as transcription and translation (Brock et al., 2009). On the other hand, population heterogeneity in tumors can be driven by specific molecules that mediate subclonal interactions, balancing cell proliferation and cell competition (Marusyk et al., 2014). Further characterization of the mechanisms that govern phenotypic diversity will enable distinction between stochastic and deterministic noise. Nevertheless, non-genetic heterogeneity is an inherent yet currently poorly understood aspect in cancer that has the potential to enhance Darwinian evolution through increasing variability of a beneficial trait (Brock et al., 2009).

The clinical presentation of metastases suggests that metastatic progression may be driven in part by non-genetic mechanisms. The mutational status of metastatic tumor nodules fails to lend credence to genetic instability as the source of molecular diversity. Metastatic tumor nodules from breast, colorectal, and pancreatic cancers display a limited number of mutations, and of those found, none have been causally linked to the development of metastases (Ding et al., 2010; Jones et al., 2008; Yachida et al., 2010). Additionally, the amount of time needed to develop metastasis can be rapid, being as little as a year in the twenty-year lifespan of colorectal cancer (Jones et al., 2008). With multiple, independent, highly

selective steps in the metastatic cascade, it is unclear how a sufficient number of cancer-promoting mutations, which so far have not been detected by sequencing analyses, could contribute to this rapid cancer evolution.

Clonal heterogeneity as a model for non-genetic variation

The study of non-genetic variability was pioneered in bacterial systems because of the genetic homogeneity of strains. In 1976, Koshland demonstrated that individual *E. coli* cells displayed variable phenotypic responses to stimuli despite being genetically identical (Spudich and Koshland, 1976). Furthermore, clonal populations of *E. coli* allowed for quantification of biological noise at the molecular level in genetic lineages (Elowitz et al., 2002; Taniguchi et al., 2010). To study non-genetic variability in clonal cancer populations, genetic variation must be accounted for. This can be accomplished in part by limiting the number of cells divisions, thus reducing the potential contribution of functional genetic mutations to the population. Additionally, genomic analyses of cancer populations following clonal expansion would allow for an assessment of genomic integrity. This approach has been performed to characterize phenotypic variability of clonal colorectal cancer populations to form tumors and respond to chemotherapy (Kreso et al., 2013). Utilization of clonal analyses in cancer populations can allow for better understanding of the non-genetic variability that may be contributing to metastatic colonization.

3.2 Isolation of clonal subpopulations with enhanced diversification capacity

Morphologic diversification

To generate an experimental model wherein genetic variation between cells is minimized so that non-genetic contributions to heterogeneity generation can be assessed, clonal subpopulations were derived from human cancer populations. The breast cancer cell lines MDA-MB-231 (MDA) and the minimally passaged CN34 breast cancer line (CN), both originally derived from the pleural effusion of patients with metastatic breast cancer (Bos et al., 2009; Cailleau et al., 1978) were seeded sparsely to allow individual cells to proliferate into isolated colonies. Once colonies could be visualized, these subpopulations were grown individually until cells could be seeded for microscopy. Nearly two hundred clonal subpopulations from these two cell lines were isolated.

Next, each subpopulation was assessed by microscopy to determine whether clonal populations could generate morphological variability following clonal expansion. Microscopic imaging was performed on cell populations labeled with CellMask stain to label entire cells and DAPI dye to label nuclei, and images were analyzed by automated image analysis to measure size parameters of individual cells, which include cell area, cytoplasmic area, nucleus size, perimeter, major axis length, and minor axis length. When cell size parameters were compared between subpopulations, there was significant inter-clonal variation (Figure 3.1). Namely, certain subpopulations consisted of large cells, while others consisted of primarily small cells.

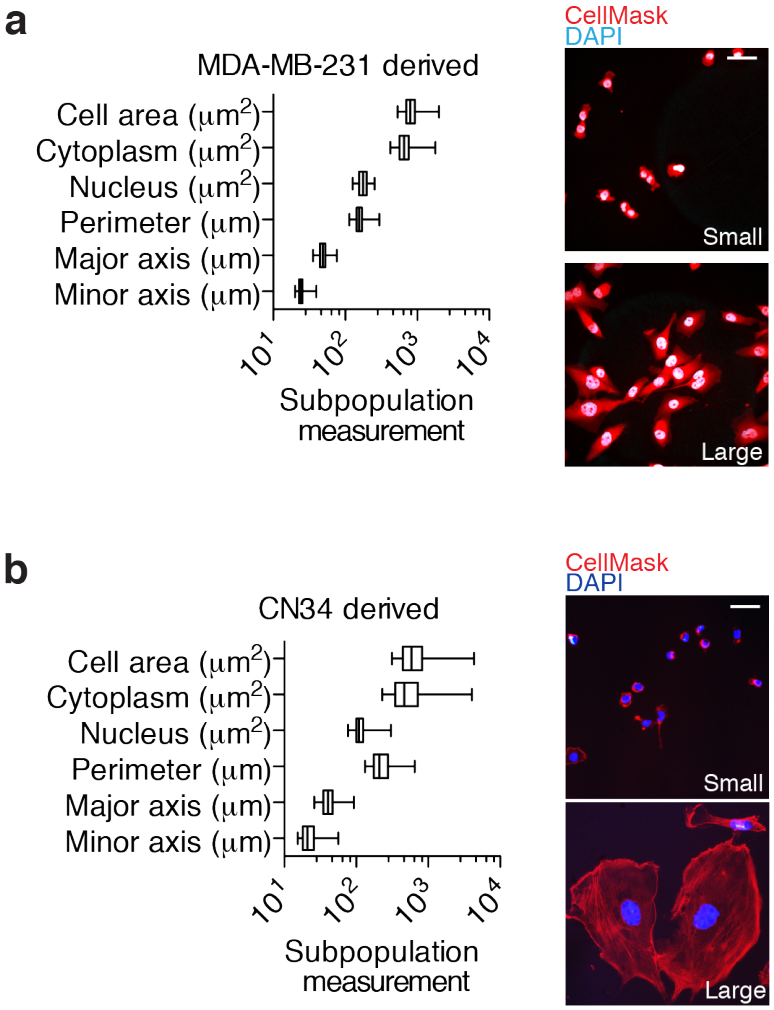

Figure 3.1: Clonal subpopulations display inter-clonal variation

Size parameters were measured for clonal subpopulations generated from MDA-MB-231 (**a;** n=98) and CN34 (**b;** n=97) breast cancer cell lines. Median, interquartile range, minimum, and maximum are depicted by box plots. Representative populations of small-sized and large-sized cell populations are shown; scale bars are 50um. CellMask red was used to label entire cells, while DAPI dye was used to label nuclei.

To identify subpopulations that generate morphological variability, the distribution of cell size within a population was assessed by calculating the coefficient of variation for each size parameter. Furthermore, principal component analysis on the six coefficient of variation measurements was performed to generate a single linear measurement of intra-clonal variability. The majority of clonal subpopulations displayed a range of variability as assessed by using the first principal component—consistent with a single peak distribution (Figure 3.2). Notably, a few subpopulations demonstrated exceptionally high intra-clonal cell size variability without exhibiting significant differences in their population-level means (Figures 3.2, 3.3). These subpopulations served as candidates that have the capacity to generate phenotypic diversity.

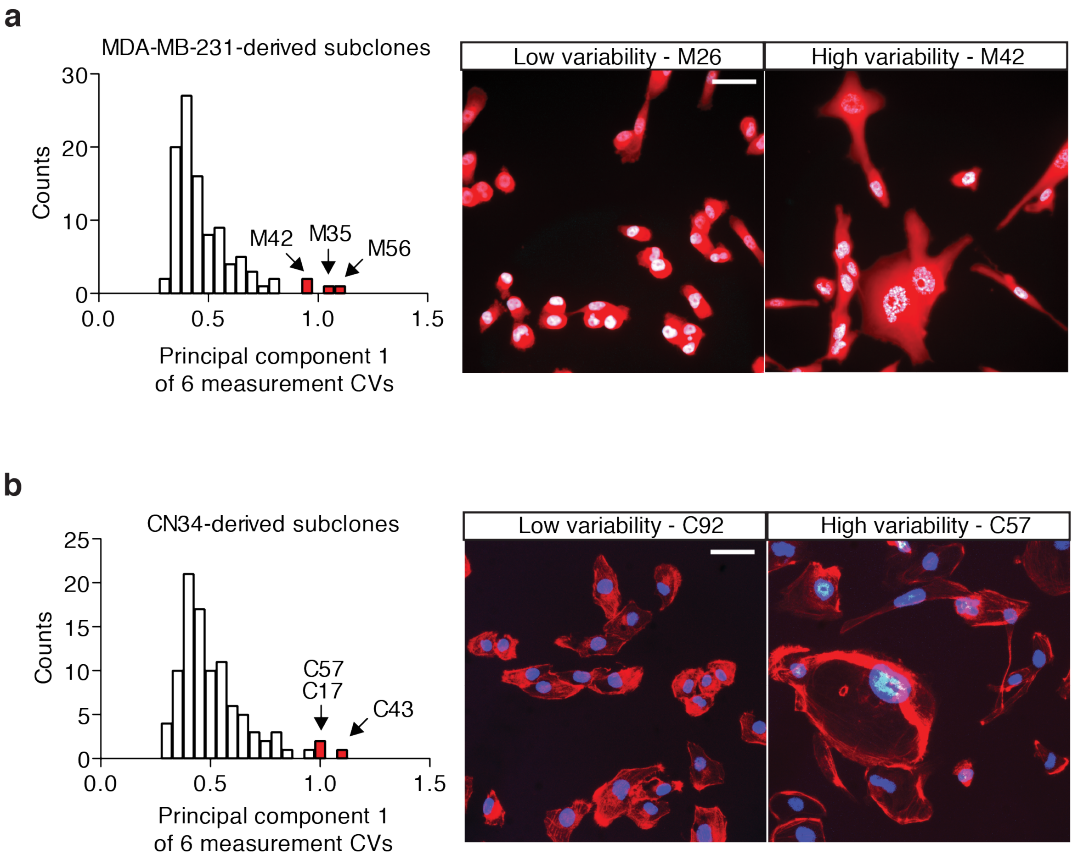

Figure 3.2: Identification of subpopulations with high intra-clonal variation

Histograms of the first principal component of size coefficient of variation (CV) was used to assess subpopulation size variability of MDA-MB-231-derived clones (**a**) and CN34-derived clones (**b**). Candidate high variability populations are indicated in red. Representative images of high and low variability subpopulations are shown on the right. Scale bars = 50um.

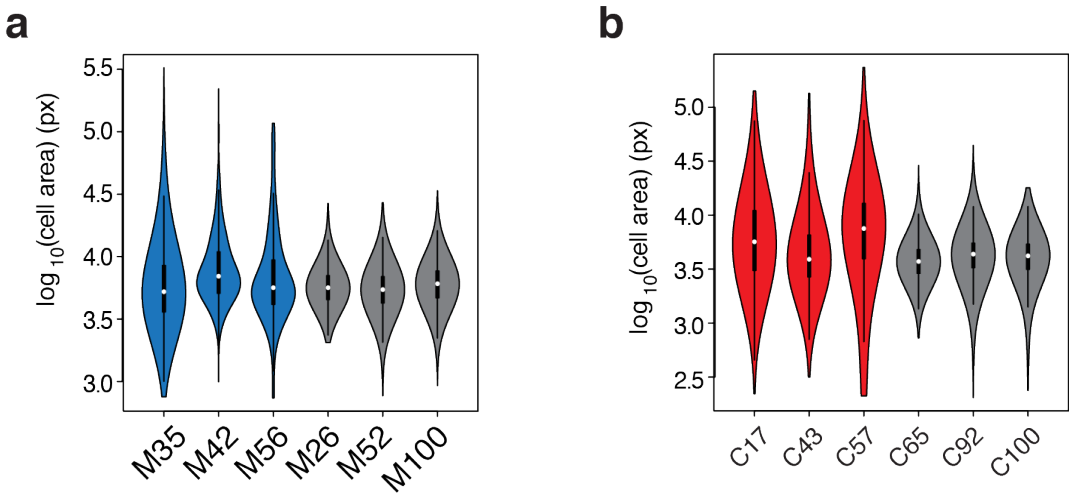

Figure 3.3: Variability in cell area in candidate high and low variability populations

Distribution of cell area as shown from candidate high and low variability populations in MDA-MB-231-derived clones (**a**) and CN34-derived clones (**b**). Median, interquartile range, minimum, and maximum are depicted by violin plots.

Phenotypic diversification

To determine if these subpopulations could give rise to and maintain phenotypic diversity beyond cell size, proliferative variability was assessed as an independent functional measure by colony formation assays. Cells from each subpopulation were seeded sparsely, allowed to grow into colonies, and then were stained with crystal violet to allow for macroscopic measurements. A macroscopically small colony would be indicative of a slower colony growth rate, whereas a large colony would indicate a faster colony growth rate. Thus, proliferative variability would manifest as both small and large colonies derived from the same subpopulation. Cell size and cell density within colonies were first assessed to confirm that these factors were not contributing to colony area differences (Figure 3.4).

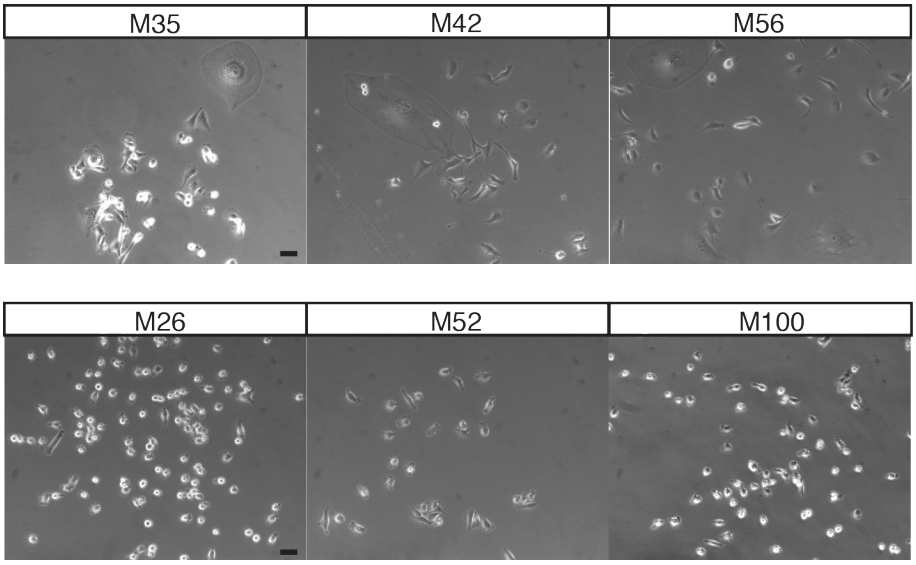

Figure 3.4: Cell size and cell density are not responsible for colony area differences

Representative images of cells in a colony from colony formation assay are shown; scale bar is 100um.

While bulk population growth in culture was observed to be similar between subpopulations (Figure 3.5), MDA-derived subpopulations that exhibited high intra-clonal size variability also displayed high variability in proliferative capacity, which could be visualized as high variability in colony sizes (Figure 3.6a). Among the CN-derived populations, only subpopulation C57 displayed a modest increase in proliferative variability (Figure 3.6b), suggesting that the other morphologically variable subpopulations from this parental line may not exhibit variation of additional phenotypes beyond morphology.

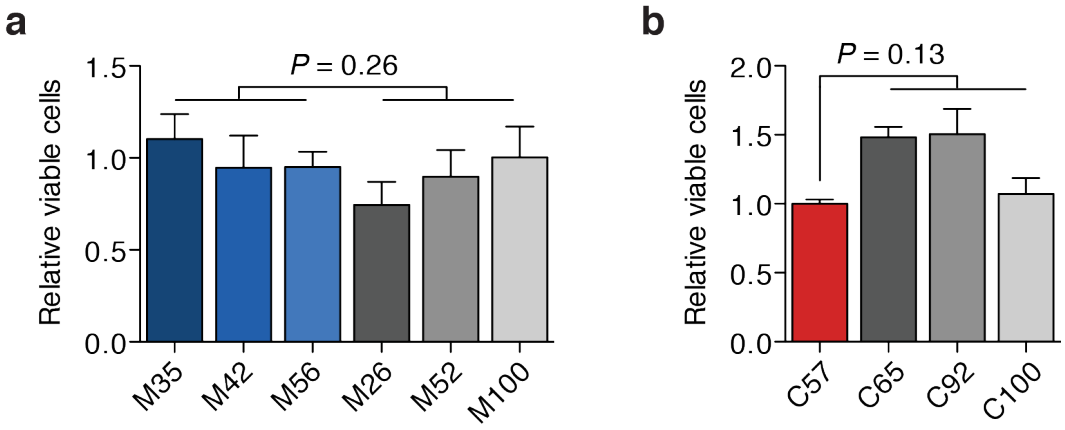

Figure 3.5: Population growth in culture is not different between high and low variability subpopulations

$5x10^3$ cells were seeded and assessed 72 hours later for viable cells using WST-1 reagent for MDA-MB-231-derived subpopulations (**a**) and CN34-derived subpopulations (**b**). P-values were derived from two-sided student's *t*-test (two-sample for MDA, one-sample for CN).

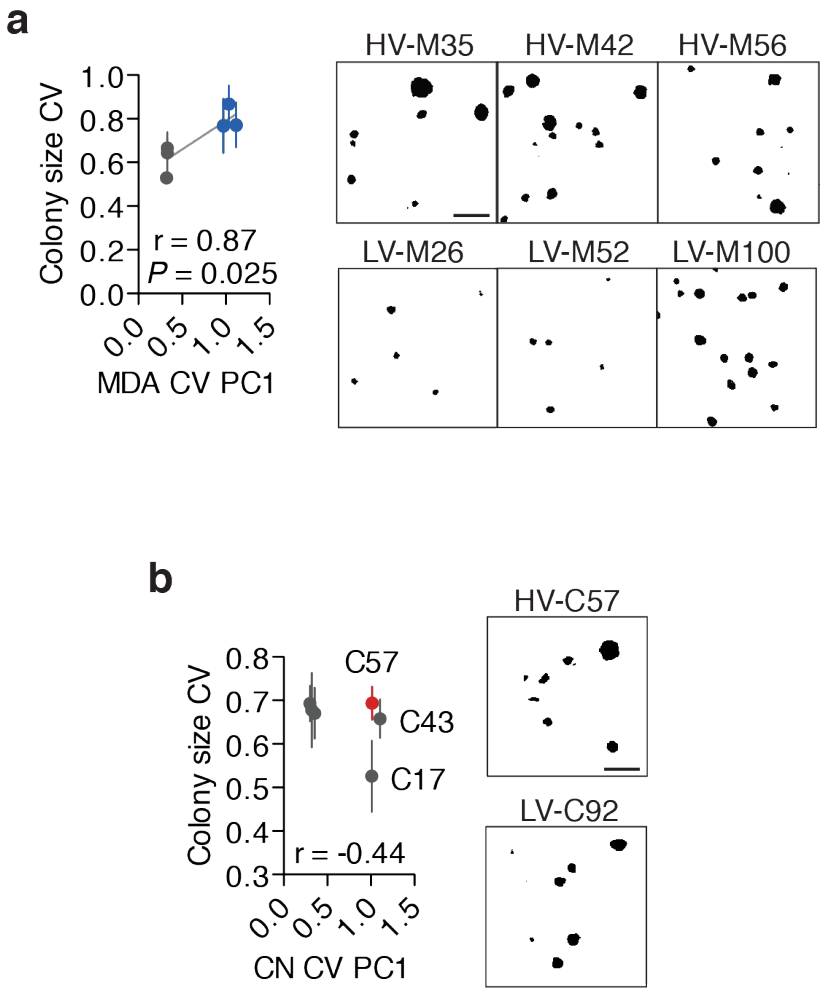

Figure 3.6: High variability subpopulations exhibit proliferative variability

MDA-MB-231-derived colonies (**a**) and CN34-derived colonies (**b**) were grown in triplicate plates and measured for variability in colony area. Error bars indicate s.e.m. of three independent experiments. P-values were generated by testing Pearson's correlation coefficient with two-sides. Representative images of colonies stained by with crystal violet and thresholded in ImageJ are shown on right; scale bar is 5mm.

These subpopulations, herein referred to as high variability subpopulations, were next assessed for whether the observed variability was heritable. High variability (HV) and low variability (LV) subpopulations were seeded sparsely to allow for isolation of clonal progeny. These clonal progeny were expanded to allow for microscopic imaging of population cell sizes. High variability subpopulations maintained high cell size variability in progeny clonal subpopulations as compared to low variability subpopulations (Figure 3.7).

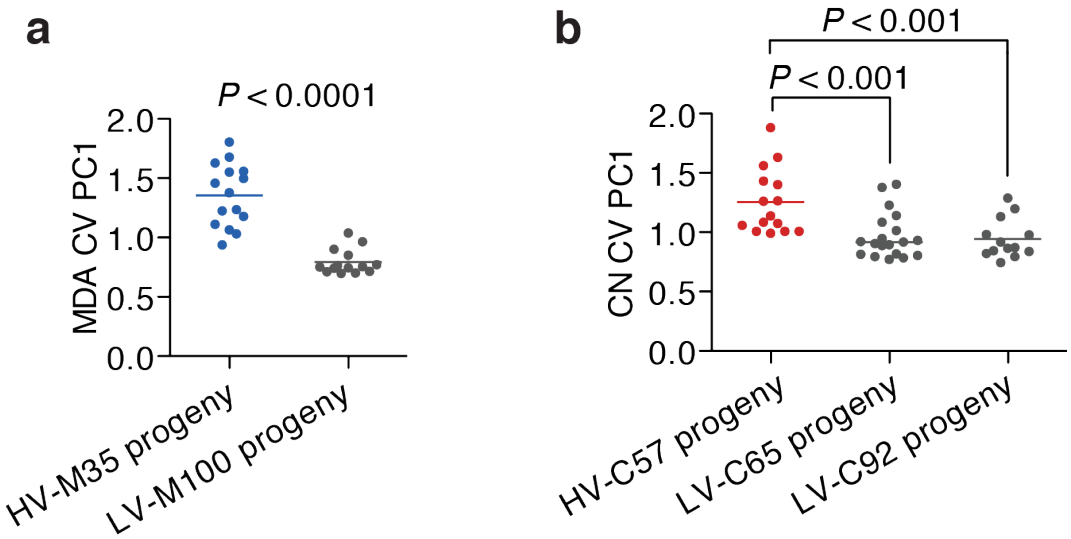

Figure 3.7: Progeny from high variability subpopulations maintain size variability after clonal isolation and expansion

Single cells isolated from indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations were expanded into clonal populations and were assessed for cell size variability. P-values were derived using two-sided Mann Whitney U test.

To further characterize cell size variability, a number of additional features were assessed. First, since size variability measured in two-dimensions on cell culture plates could be due to either morphological shape or cell volume, high variability subpopulations were assessed for size variability in three-dimensions using ImageStream flow cytometry with high content image analysis. Indeed, high variability subpopulations exhibited size variability in three-dimensions (Figure 3.8), indicating variation in cell mass. Next, it is possible that cell size variability could be also attributed to altered cell cycle phasing as cells must grow in size prior to mitosis. Cell cycle phasing as measured by flow cytometric analysis of DNA content revealed no difference between high and low variability subpopulations (Figure 3.9). Finally, cell size variability was assessed for its robustness following passage in culture and growth at varying densities. Increased cell size variability of high variability subpopulations was maintained after passage in culture and regardless of seeding density (Figures 3.10, 3.11). Taken together, these high variability clonal subpopulations display an enhanced and heritable ability to generate diversity across multiple phenotypic dimensions.

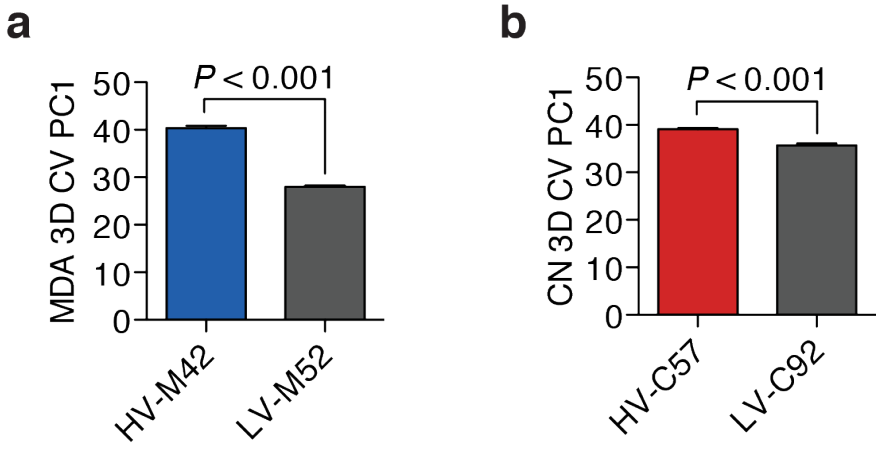

Figure 3.8: High variability subpopulations exhibit three-dimensional size variability

The first principal component of 3D size coefficient of variation was used to assess three-dimensional size variability from indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations. Error bars indicate s.e.m. of three independent experiments. P-values were derived using two-sided *t*-test.

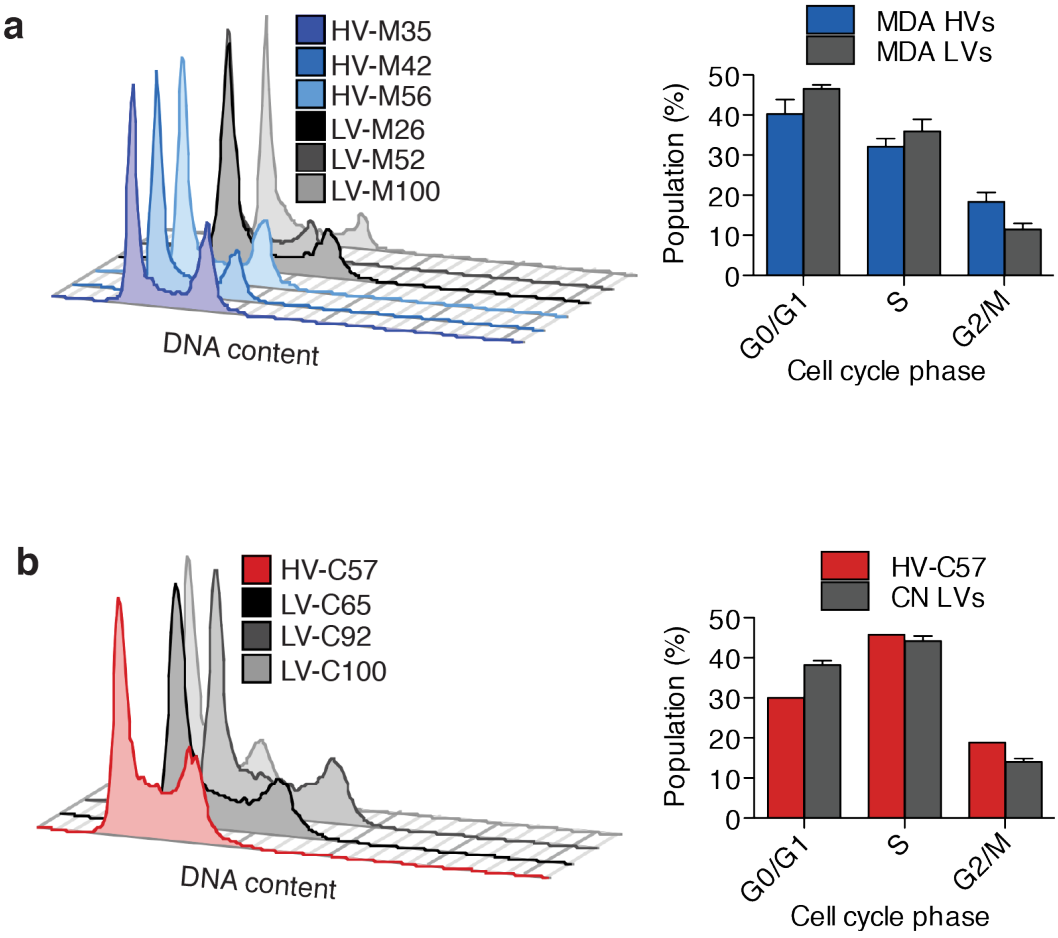

Figure 3.9: High variability subpopulations exhibit no significant difference in cell cycle phasing

Fixed cells from indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations were stained with DAPI and analyzed by flow cytometry to measure DNA content. Histograms of each population are shown (left). Watson pragmatic cell cycle analysis was performed in FlowJo to assess cell cycle phasing (right).

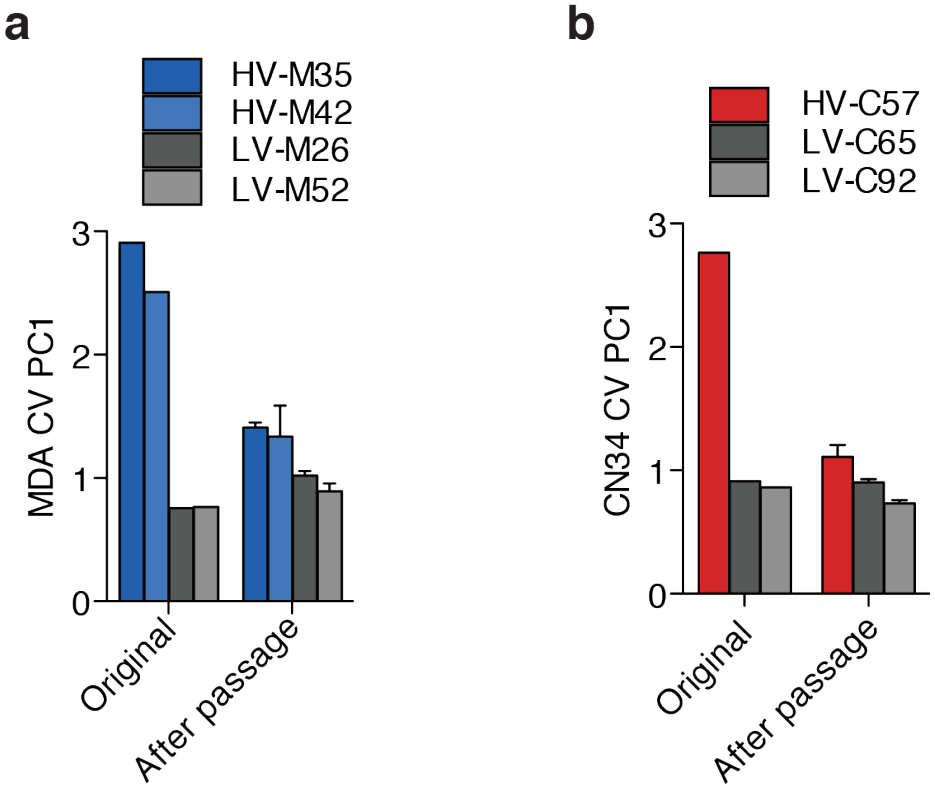

Figure 3.10: High variability subpopulations maintain relative size variability after passage in culture

Subpopulations from indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations were passaged five times in culture under standard cell culture conditions and then imaged and analyzed for cell size coefficient of variation.

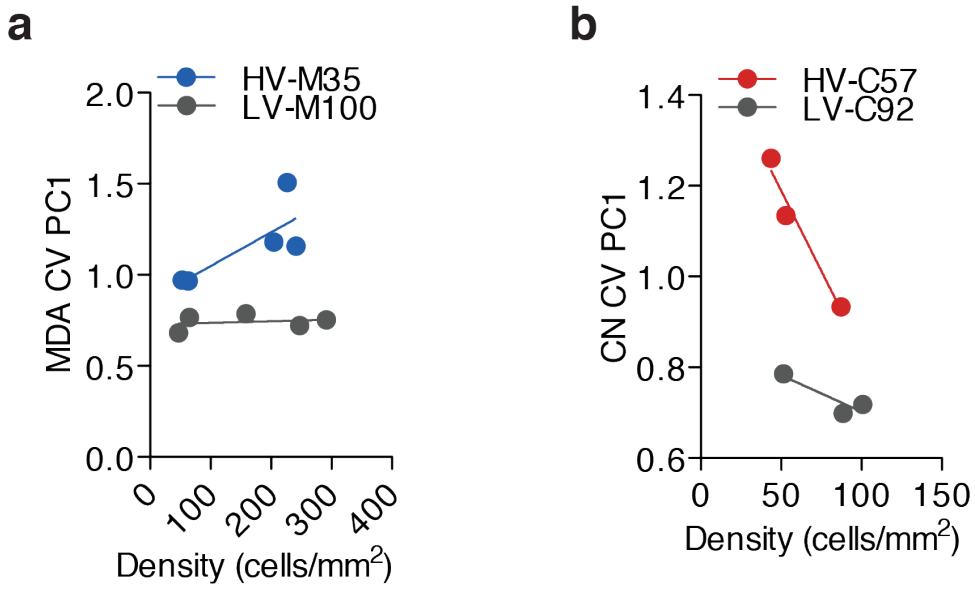

Figure 3.11: High variability subpopulations exhibit increased cell size variability at different seeding densities

Indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations were seeded at various densities and then imaged and analyzed for cell size coefficient of variation.

3.3 High variability subpopulations exhibit metastatic fitness

Systemic metastasis

Metastatic colonization of an end organ represents a major bottleneck during cancer evolution that would greatly benefit from diversity generation and is clinically responsible for the majority of cancer deaths (Aktipis et al., 2013; Gupta and Massague, 2006; Klein, 2013). The identification of subpopulations with high versus low diversification potential from the same individual patients' cancer populations allowed for testing the impact of intra-clonal diversification capacity on metastatic colonization. Because diversity generation should broadly enhance cancer evolution, in the context of the metastasis, diversity is likely to benefit colonization of multiple organs by increasing the likelihood of successful metastasis. To assess broad colonization capacity, cells were inoculated into arterial circulation of mice by intracardiac injection, which allows cells to arrest in the capillary beds of numerous organs. High variability subpopulations exhibited enhanced systemic metastatic capacity as determined by total tumor burden (Figure 3.12). Additionally, this increased tumor burden was in large part attributed to an increased number of systemic metastatic foci (Figure 3.13), indicating an increased frequency to colonize systemic sites as opposed to an increased growth rate of the formed metastases.

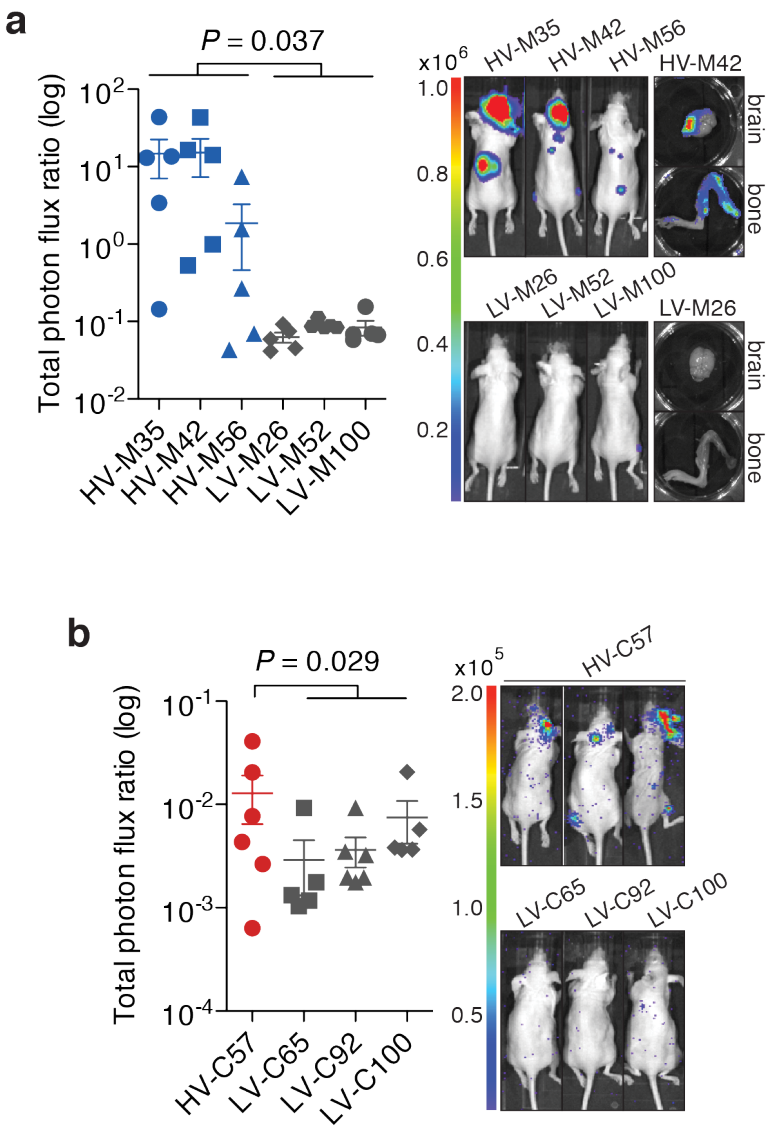

Figure 3.12: High variability subpopulations exhibit enhanced systemic metastatic capacity

Bioluminescence quantification of total body metastatic burden by $4x10^4$ MDA-derived cell populations after 42 days (**a**) and $2x10^5$ CN-derived cell populations after 72 days (**b**) following inoculation into arterial circulation via intracardiac injection. P-values were derived using two-sample (**a**) and one-sample (**b**) one-sided t-test; n=5-6. Representative mouse and organ bioluminescence are shown.

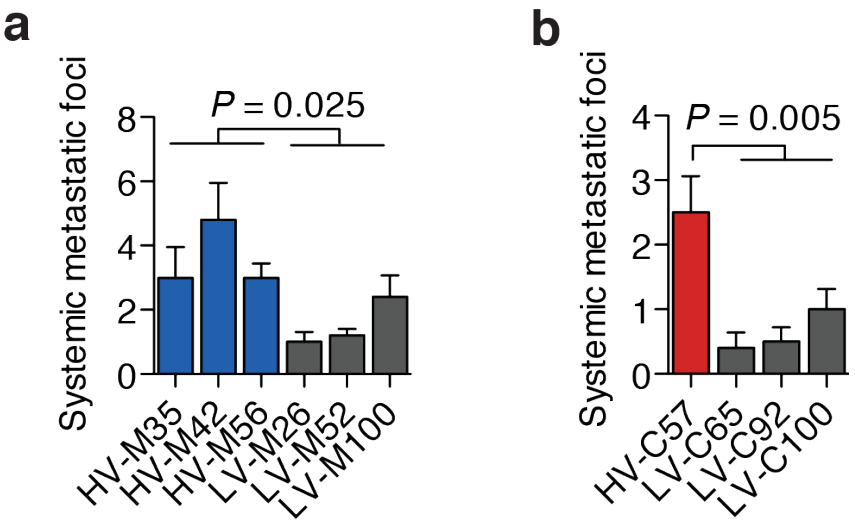

Figure 3.13: High variability subpopulations exhibit increased systemic colonization frequency

Systemic metastatic foci per mice were counted as distinct, minimal bioluminescence signals cells from indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations. P-values were derived using two-sample (**a**) and one-sample (**b**) one-sided *t*-test.

Organ metastasis

To confirm the increased colonization capacity of high variability subpopulations in a more clinically relevant context, these populations were assessed for the capacity to colonize the lung and liver, two of the most common sites of breast cancer metastasis (Siegel et al., 2015). High variability subpopulations derived from both human lines more efficiently colonized the lung upon tail vein injection (Figure 3.14) and the liver upon portal circulation injection (Figure 3.15), indicating that the enhanced metastatic capacity is broadly applicable to multiple organs posing diverse selective barriers. Additionally, to confirm that high variability populations contribute to metastasis to a greater degree in a mixed population with numerous subclones, a mixed population consisting of an equal number of high and low variability cells labeled with barcodes was generated. Portal circulation inoculation of this mixed population revealed high variability cells to contribute between 81.2% to 99.6% of formed liver metastases as measured by quantitative PCR of DNA barcodes (Figure 3.16), demonstrating that high variability cells maintained an increased metastatic capacity in a mixed population. The ability of high variability subpopulations to both generate phenotypic diversity and metastasize more efficiently is consistent with a positive role for phenotypic diversification in cancer progression.

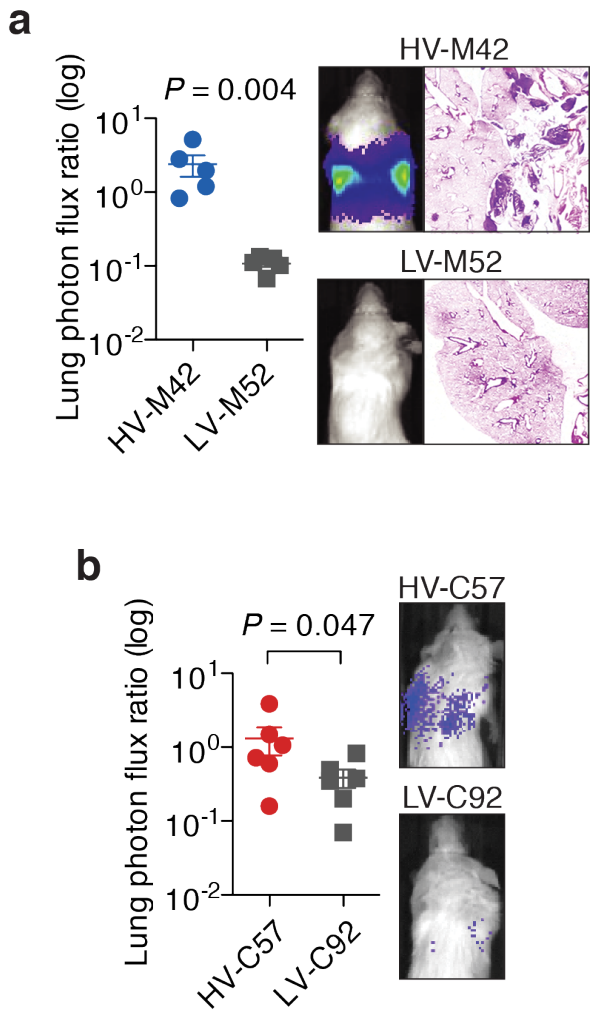

Figure 3.14: High variability subpopulations exhibit increased lung colonization capacity

Bioluminescence quantification of lung metastases by $4x10^4$ MDA-derived cell populations after 70 days (**a**) and $2x10^5$ CN-derived cell populations after 112 days (**b**) following inoculation into the tail vein. P-values were derived using one-sided Mann Whitney U test; n=5. Representative bioluminescence and lung histology is shown.

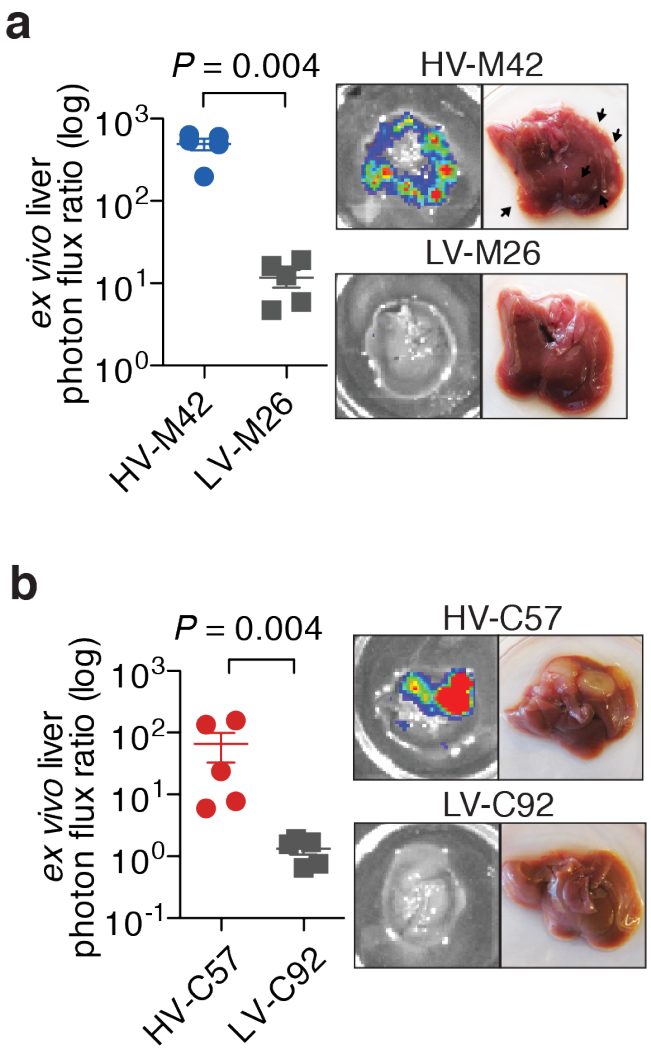

Figure 3.15: High variability subpopulations exhibit increased liver colonization capacity

Ex vivo bioluminescence quantification of liver metastases by 4x10^4 MDA-derived cell populations after 39 days (**a**) and 2x10^5 CN-derived cell populations after 91 days (**b**) following inoculation into the portal circulation. P-values were derived using one-sided Mann Whitney U test; n= 5. Representative bioluminescence and gross histology are shown.

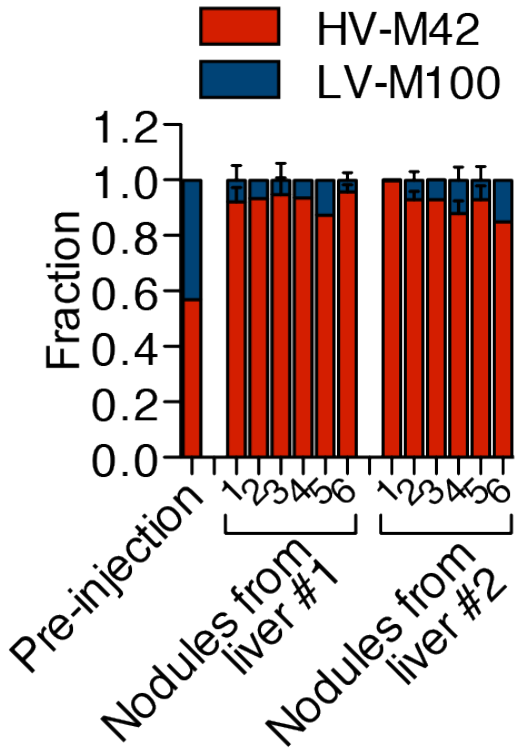

Figure 3.16: High variability subpopulations exhibit increased liver colonization capacity in a mixed population

$5x10^5$ cells consisting of equal parts of HV-M42 cells, labeled with blasticidin resistance gene, and LV-M100 cells, labeled with puromycin resistance gene, were inoculated into the portal circulation. Liver metastases were extracted after 28 days and processed for quantitative PCR measurement of genomic DNA resistance genes.

Clinical metastasis

In order to corroborate the role of phenotypic diversification in metastatic progression, the next goal was to determine whether phenotypic diversity was associated with clinical outcomes. Given the observation of increased nuclear size variability in high variability subpopulations (Figure 3.17) as well as the ability to distinctly label nuclei with DNA stains, this tractably quantifiable parameter of nuclear size was used as a readout of phenotypic variability in human invasive breast cancer tumor core biopsies. A tissue microarray containing 168 useable core biopsies of primary breast cancer tumors was stained with DAPI to label nuclei, imaged by confocal microscopy, and analyzed by automated image analysis to measure cancer cell nuclei sizes within each core. Consistent with the findings in the clonal populations, nuclear area variability of cancer cells was significantly increased in tumors of patients with more advanced stage disease (Figure 3.18). Primary tumors that progressed to lymph nodes displayed significantly higher variation in nuclear area than lymph node negative tumors. Furthermore, primary tumors that progressed to distant metastases exhibited significantly higher nuclear area variation than those tumors that did not metastasize. Mean nuclear size and mitotic index did not significantly correlate with disease stage (Figure 3.19). These clinical correlations are consistent with our findings that cancer populations with greater diversification potential positively contribute to metastatic disease.

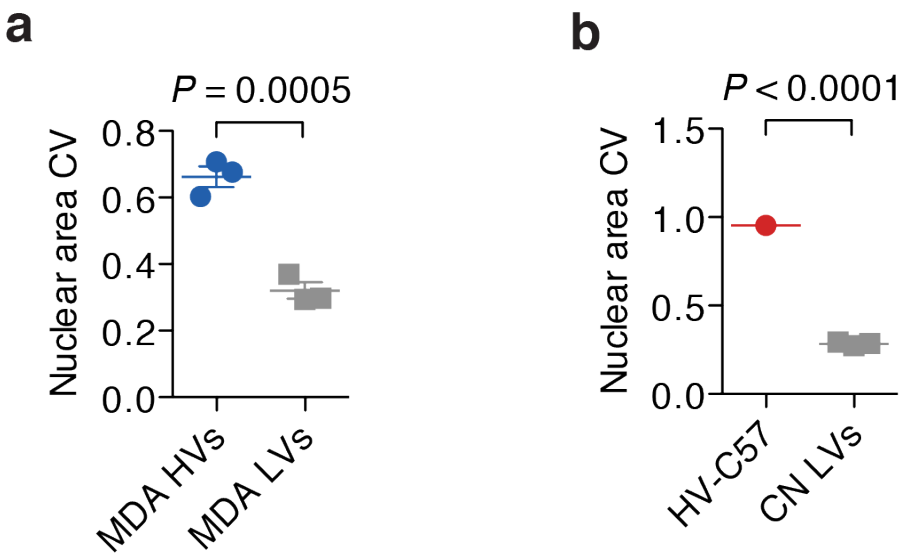

Figure 3.17: High variability subpopulations exhibit increased nuclear area variability

Nuclear area variability as measured by automated image analysis of indicated MDA-MB-231-derived (**a**) and CN34-derived (**b**) subpopulations.

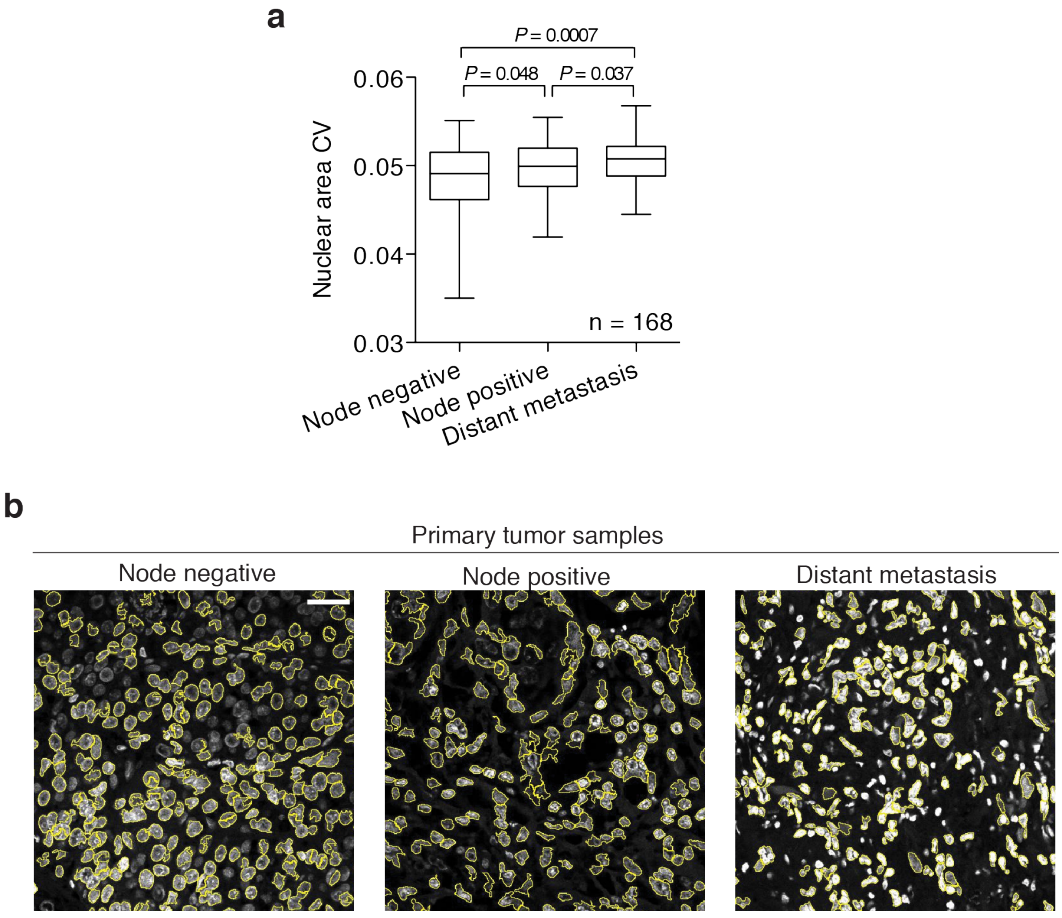

Figure 3.18: Nuclear area variability associates with human breast cancer metastasis

(**a**) NCI CDP Breast Cancer Progression Tissue Microarray was stained with DAPI, imaged by confocal microscopy, and analyzed for nuclear area coefficient of variation for each breast cancer core. *P*-values were derived using one-sided *t*-test. (**b**) Representative images with measured cancer cell nuclei (yellow boundary) are shown; scale bar = 50um.

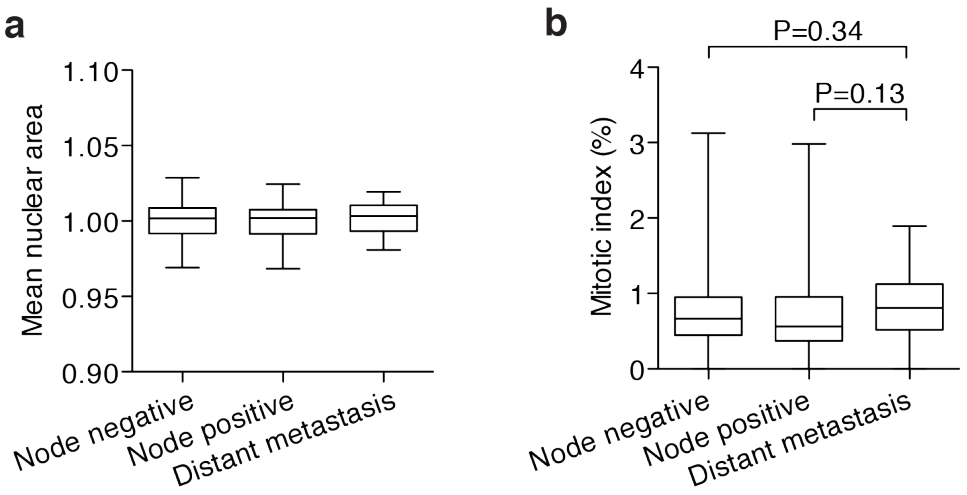

Figure 3.19: Nuclear area variability in human breast cancer is not dependent on mean nuclear area or mitotic index

NCI CDP Breast Cancer Progression Tissue Microarray was stained with DAPI, imaged by confocal microscopy, and analyzed for mean nuclear area (**a**) and mitotic index (**b**) in each breast cancer core. *P*-values were derived using one-sided *t*-test.

3.4 Molecular characterization of high variability subpopulations

Genomic analysis

These clonal subpopulations of cells derived from isogenic backgrounds offered the potential for studies into the molecular basis of cancer variability. Given the contribution of genomic instability to tumor heterogeneity (de Bruin et al., 2014; Greaves and Maley, 2012), high variability subpopulations were assessed for mutational burden. Genomic DNA was isolated from subpopulations and processed for whole exome sequencing to allow for quantification of single-nucleotide variant and insertion-deletion mutation frequencies. The mutational profiles allowed for reconstruction of phylogenic relationships as subpopulations diverged from an ancestral parent population (Gerlinger et al., 2012). Exome sequencing revealed that high variability subpopulations diverged from a common high variability ancestor and exhibited genetic similarity to the parental MDA-MB-231 population (Figure 3.20), indicating that substantial genetic mutational changes did not accumulate as these subpopulations clonally expanded. Furthermore, high variability subpopulations did not differ in the number of population-specific nucleic acid variants relative to low variability subpopulations or the parental population (Figure 3.20). While it is difficult to exclude the contribution of nucleic acid mutations to subpopulation heterogeneity, the comparable genetic phenotypes and relative genetic stability in the isogenic high variability subpopulations suggests that enhanced genetic diversification may not be the primary source of the observed phenotypic diversity.

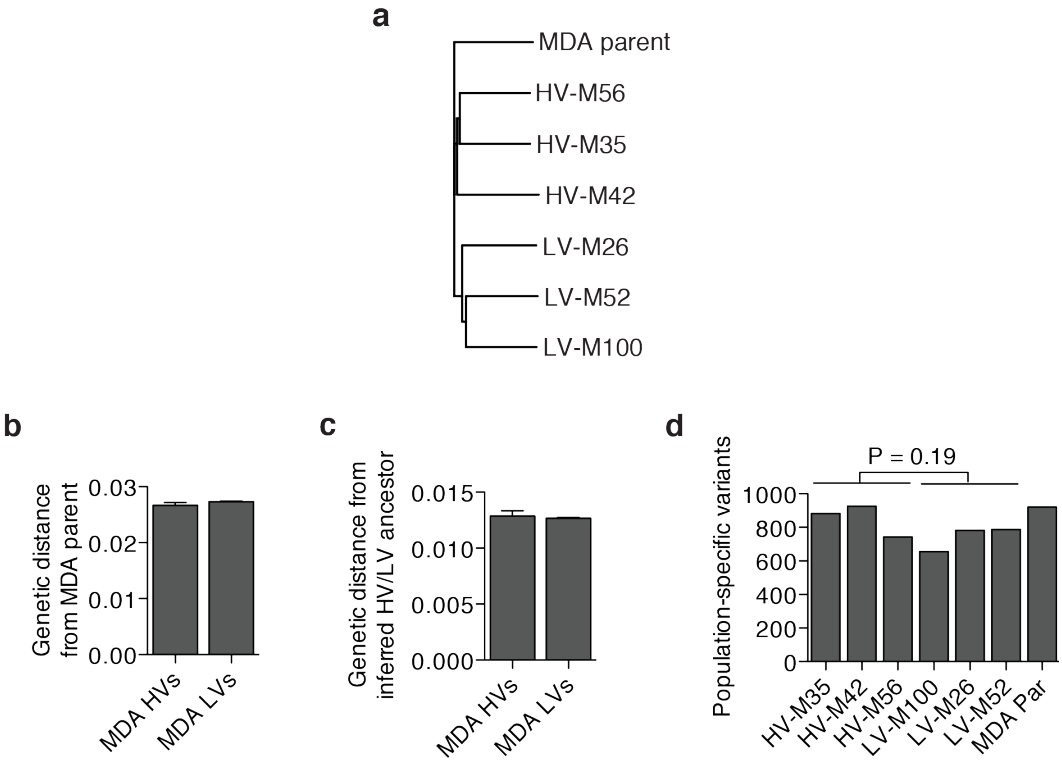

Figure 3.20: No evidence of genomic instability as a factor associated with high variability subpopulations

(**a**) Phylogenic tree based on genomic variant allele frequencies as measured by exome sequencing from cell populations. Genetic distance of subpopulations relative to MDA-MB-231 parent cell line (**b**) or from inferred HV or LV common ancestor (**c**) as determined by Nei's genetic distance. (**d**) Variants found only in the indicated subpopulations at any frequency greater than zero were counted.

Transcriptomic variability

In order to identify additional molecular mechanisms that could contribute to phenotypic diversity, single-cell RNA-sequencing from high and low variability subpopulations was performed to assess gene expression variability between cells (Figure 3.21), which could contribute to the functional variability between cells in high variability subpopulations. Cells were sorted by flow cytometry into individual wells with lysis buffer to expose RNA for immediate generation of barcoded cDNA. Barcoded cDNA was then pooled and processed for generation of sequencing libraries.

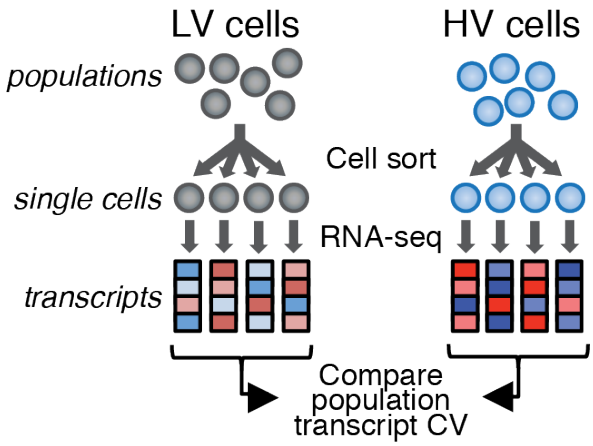

Figure 3.21: Single-cell RNA-sequencing to compare cell-to-cell gene expression variability

Schematic for single-cell RNA sequencing and transcriptomic variability analysis.

Assessment of single-cell RNA sequencing quality and fidelity demonstrated appropriate spike-in expression and no difference in spike-in variability between wells (Figure 3.22). Total transcript abundance per cell showed no significant differences between high and low variability subpopulations (Figure 3.23). Moreover, global mean transcript abundance was also not significantly different between high and low variability subpopulations (P = 0.91 for HV-C57 vs LV-C92 & P = 0.62 for HV-M42 vs LV-M26 by two-sided paired *t*-tests).

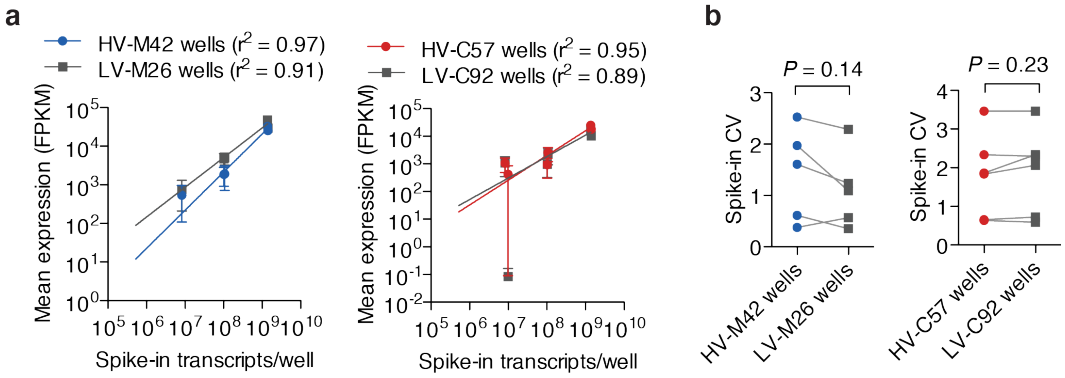

Figure 3.22: Spike-in expression and variability in single cell wells indicates appropriate single-cell sequencing fidelity

(**a**) Mean expression of spike-in transcripts per well as assessed by single cell sequencing was plotted against known spike-in concentrations. (**b**) Coefficient of variation of spike-in transcripts between wells containing the indicated population of cells was calculated. P-values were derived using two-sided paired *t*-test.

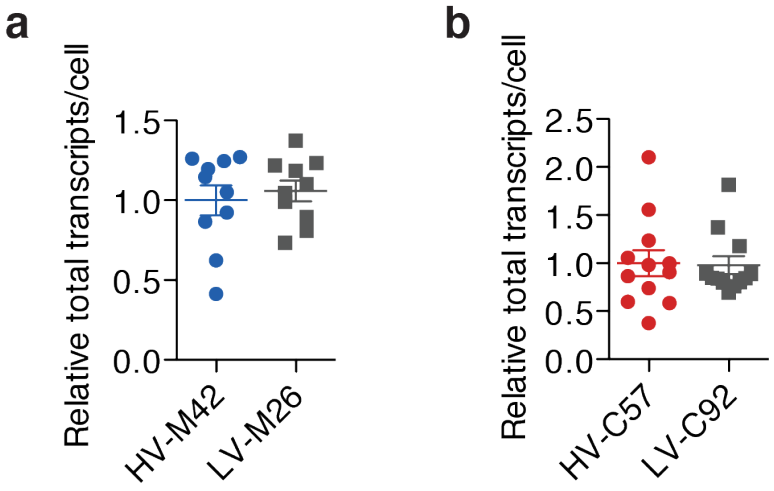

Figure 3.23: Single-cell total transcripts was similar between high and low variability cells

Relative transcript abundance per cell from indicated MDA-derived subpopulations (**a**) and CN-derived subpopulations (**b**).

Interestingly, linear and nonlinear clustering (Amir el et al., 2013) of single-cell transcript expression profiles could not classify high or low variability subpopulations (Fig. 3.24), suggesting that averaged over-expression or repression of a gene or gene set may not be the primary cause of phenotypic variation recurrently observed in high variability subpopulations (Altschuler and Wu, 2010). However, global cell-to-cell transcript expression variability, as assessed by quantification of 8,218 transcripts from single MDA-MB-231-derived high variability cells, was significantly elevated relative to expression variability between single low variability cells (53.9% transcripts with higher CV in HV-M42; P = 7e-8 by two-sided paired *t*-test; Figure 2.25). The same was observed for single cells from the HV-C57 subpopulation (5,826 transcripts; 55.0% transcripts with higher CV in HV-C57; P = 3e-10 by two-sided paired *t*-test; Figure 3.25).

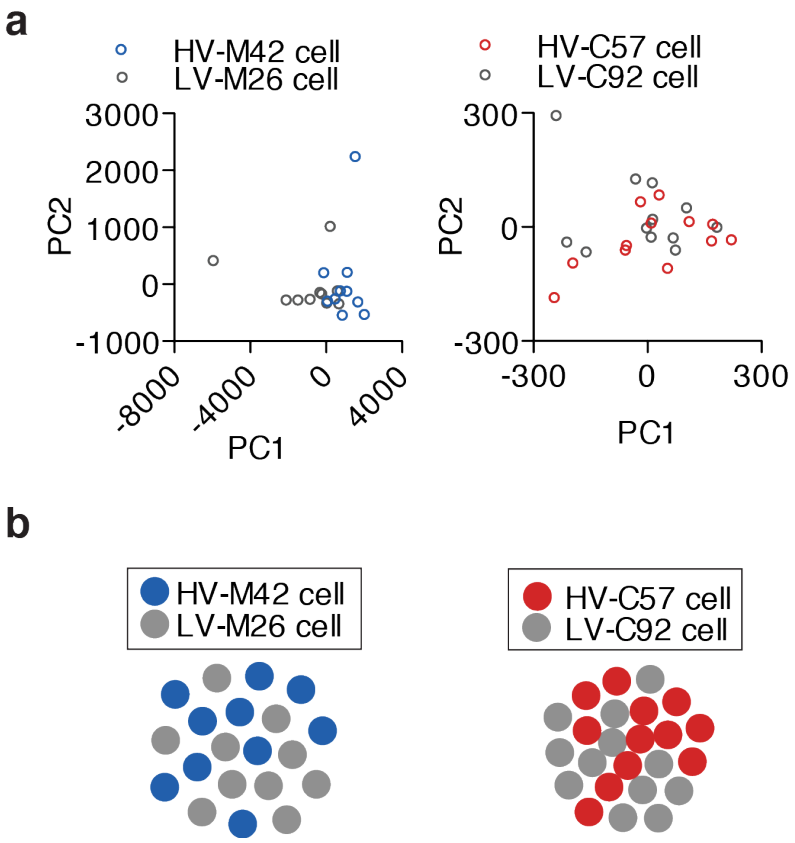

Figure 3.24: Gene expression profiles of cells from high and low variability subpopulations are similar

(**a**) Principal component analysis was performed on single-cell transcript expression profiles from MDA-derived subpopulations (left) and CN-derived subpopulations (right). (**b**) Visualization of transcript expression profiles of 10 cells from selected MDA-derived subpopulations (left) and 12 cells from selected CN-derived subpopulations using t-distributed stochastic neighbor embedding (t-SNE).

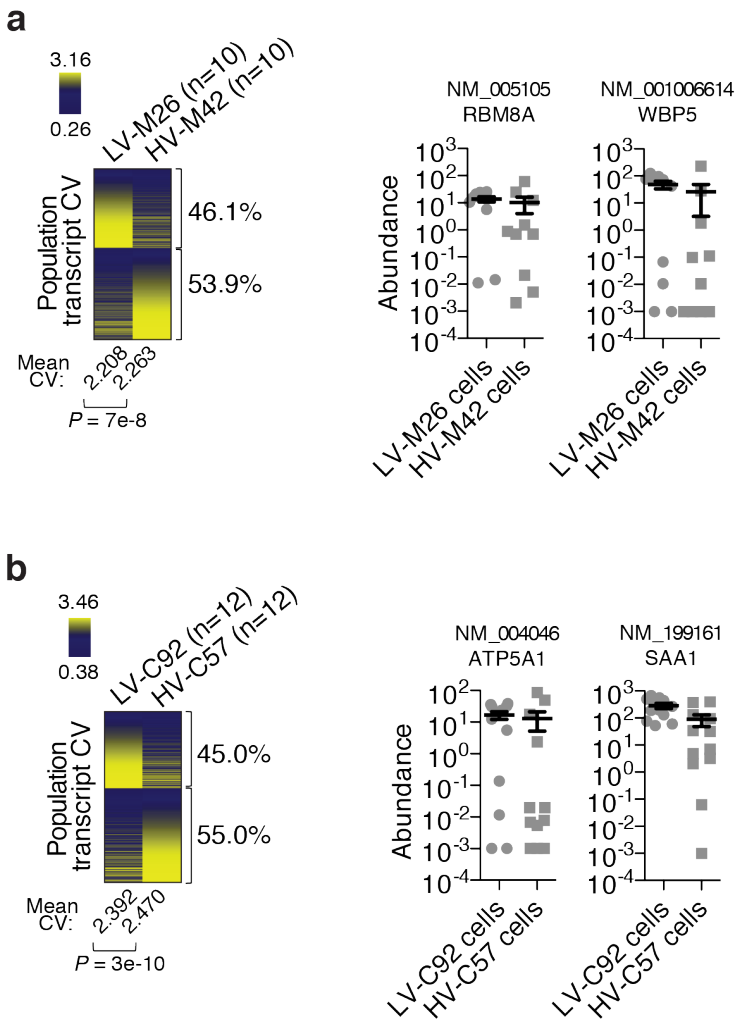

Figure 3.25: High variability populations exhibit increased cell-to-cell transcript expression variability

Transcriptomic variability analysis was performed by calculating expression coefficient of variation for each transcript between cells from MDA-derived subpopulations (**a**; n=10 single cells per population) and from CN-derived subpopulations (**b**; n=12 single cells per population). Mean transcript coefficient of variation was calculated for all transcripts. P-value was derived from two-sided paired t-test. Percentages of transcripts with higher CV in each population are shown. Representative high variability transcripts in single cells are shown (right); n=10 single cells for MDA-derived populations and n=12 single cells for CN-derived populations.

Characterization of transcriptomic variability

To assess the robustness of this molecular phenotype, sampling analyses of the single-cell sequencing experiments were performed. While more abundant transcripts are known to be detected with higher accuracy, elevated transcriptomic variability was observed regardless of transcript abundance (Figure 3.26). To determine whether sequencing from limited number of cells affected the outcome, random sampling of single cells from subpopulations was performed to determine if the enhanced transcriptomic variability could be detected with fewer cells. Indeed, transcriptomic variability was observed when setting the analysis to as few as five cells per population (Figure 3.27). Additionally, to determine if a unique cell was responsible for the molecular variability, sampling was performed to assess population transcriptomic variability following removal of each single high variability cell sequenced. Transcriptomic variability was consistently detected regardless of which high variability cell was excluded from the analysis (Figure 3.28), indicative of robust and indiscriminate population-level variability. While the increased transcriptomic variability appears modest, this effect is consistently observed under different analysis parameters and in two subpopulations derived from independent breast cancer populations. Taken together, these findings indicate that phenotypically diverse, metastatic cancer subpopulations maintain enhanced intra-clonal transcriptomic variability generation capacity.

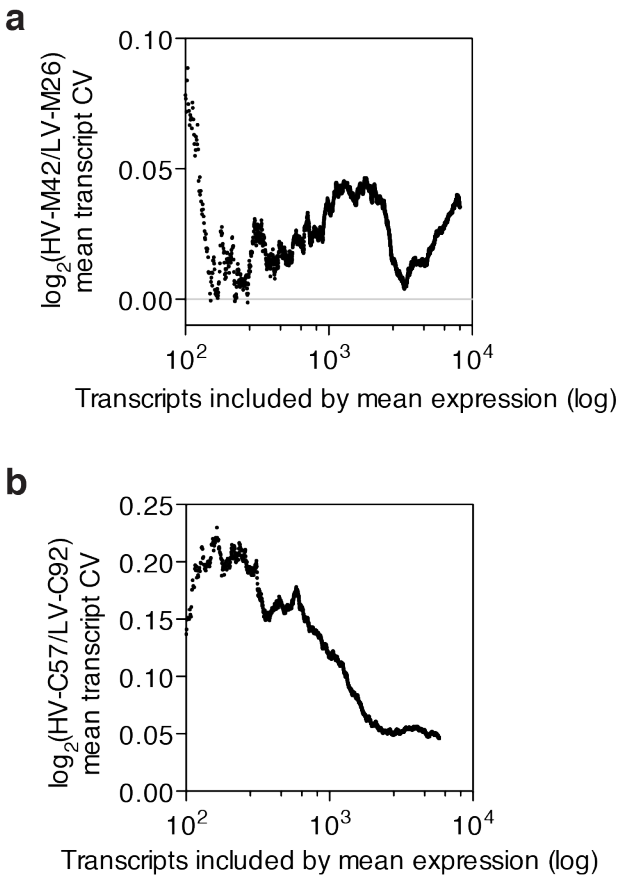

Figure 3.26: Increased cell-to-cell transcript expression variability is not dependent on transcript abundance

Log-ratios of transcript CV mean were calculated by using varying number of transcripts, ordered by descending mean single cell expression. Analysis was performed using a minimum of 100 transcripts. All cells, 10 for each MDA-derived HV and LV subpopulation (**a**) and 12 for each CN-derived HV and LV subpopulation (**b**), were used for analysis.

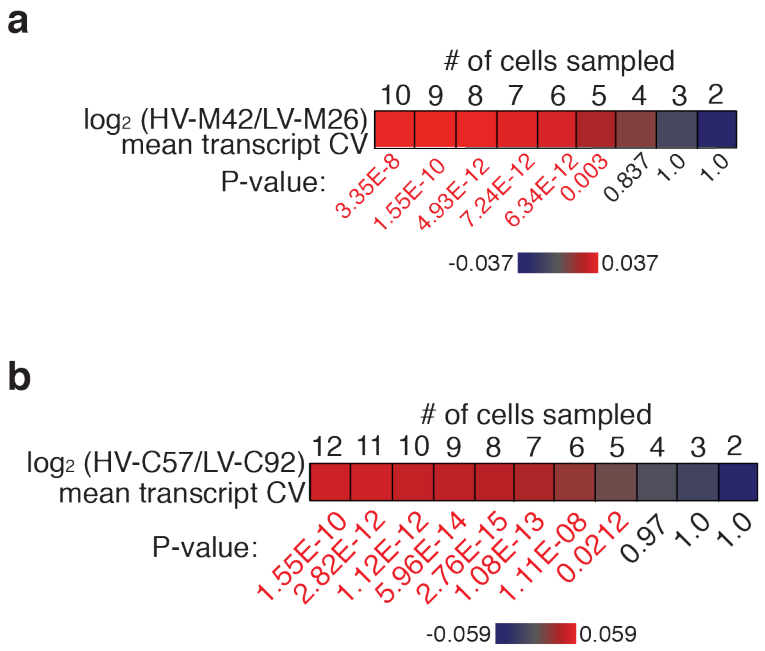

Figure 3.27: Increased cell-to-cell transcript expression variability is detected with as few as five cells per population

Log-ratios of transcript CV mean were calculated by sampling MDA-derived (**a**) and CN-derived (**b**) HV or LV subpopulations 10 times using different set sizes and calculating the mean transcript CV for HV and LV subpopulations. P-values were derived using one-sided paired *t*-test. All transcripts were used for analysis. Significant p-values are in red. Color scales are shown.

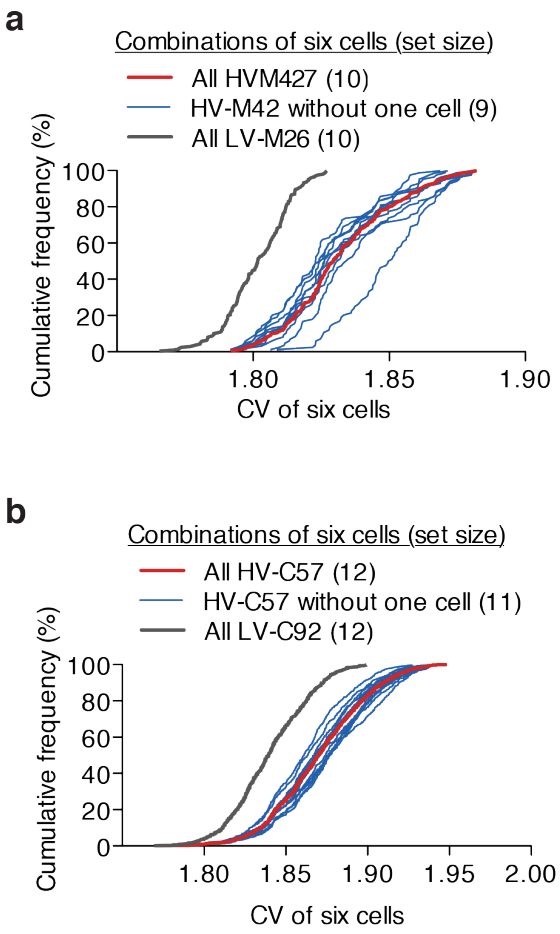

Figure 3.28: Increased cell-to-cell transcript expression variability is detected regardless of which cell is excluded from analysis

Cumulative frequency distribution is shown from sampling analysis. (**a**) All possible combinations of six cells were sampled from the 10 MDA-derived HV cells with one specific cell removed (set size of 9) and distribution of mean CV is plotted (blue, 10 possible combinations to exclude each of the 10 cells). (**b**) All possible combinations of six cells were sampled from the 12 CN-derived HV cells with one specific cell removed (set size of 11) and distribution of mean CV is plotted (blue, 12 possible combinations to exclude each of the 12 cells). Distributions of mean CV from samplings of six cells from all HV cells (red, 1 possible combination to include all cells) and all LV cells (grey, 1 possible combination to include all cells) are shown.

Transmission of variability to protein expression

To validate the relevance of transcriptomic variability as it pertains to biological function, flow cytometry was used to assess protein-level variation. XPNPEP3 and UPF2, two genes that displayed high transcript-level variability in both high variability subpopulations (Table 3.1) and whose protein expression per cell could be readily quantified in a high-throughput manner by flow cytometry, demonstrated a consistent level of increased protein expression variability in high variability subpopulations (Figure 3.29). These findings were extended to the following additional five proteins that could be readily quantified by flow cytometry: ALDOA, PABPC1, HNRNPA1, CD110, and HNRNPA0. In all high and low variability subpopulations, we observed a significant correlation between transcript-level variation and protein-level variation for the seven genes tested (Figure 3.30) with no consistent correlation in mean protein abundance (Figure 3.31). These findings reveal that transcript variation is transmitted to the protein level and highlight molecular variability as a principal feature of high variability subpopulations.

Table 3.1: Top 20 most variable genes in common to MDA-derived and CN-derived high variability subpopulations

RefSeq	Gene	MDA HV/LV log2-ratio	CN HV/LV log2-ratio
NM_020464	NHSL1	2.885	1.965
NM_015542	UPF2	1.460	1.993
NM_005837	POP7	2.141	1.085
NM_000967	RPL3	2.838	0.310
NM_145796	POGZ	3.074	0.058
NM_032307	C9orf64	1.812	1.265
NM_006990	WASF2	1.920	1.083
NM_199287	CCDC137	1.461	1.531
NM_001031804	MAF	1.416	1.537
NM_001242312	FAM124A	1.757	1.190
NM_001127175	MRO	1.876	1.054
NM_001077203	SENP7	1.921	0.980
NR_033702	MYNN	2.874	0.002
NM_173510	CCDC117	1.796	1.014
NM_022098	XPNPEP3	1.326	1.473
NM_001198530	TCF7L2	1.444	1.336
NM_139135	ARID1A	1.821	0.956
NM_021033	RAP2A	2.126	0.642
NM_021130	PPIA	1.176	1.576
NM_003111	SP3	1.182	1.569

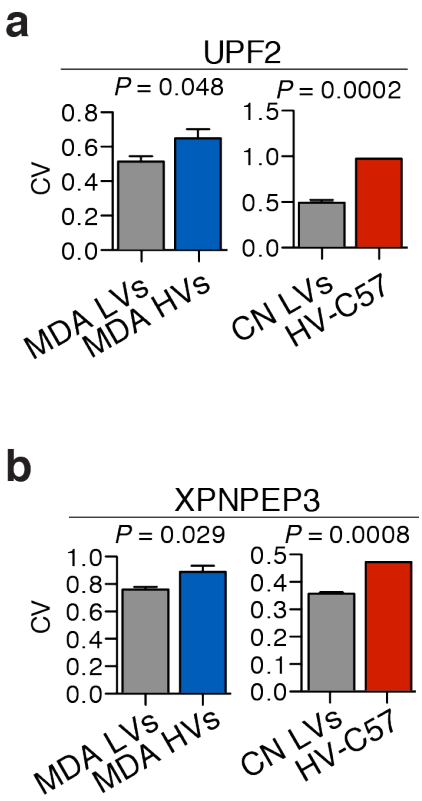

Figure 3.29: Cell-to-cell transcript expression variability is transmitted to the protein-level

Flow cytometry was used to assess UPF2 (**a**) and XPNPEP3 (**b**) protein coefficient of variation in subpopulations by assaying over 2.5x10^4 cells. P-values were derived using two-sample (MDA) or one-sample (CN) one-sided *t*-test.

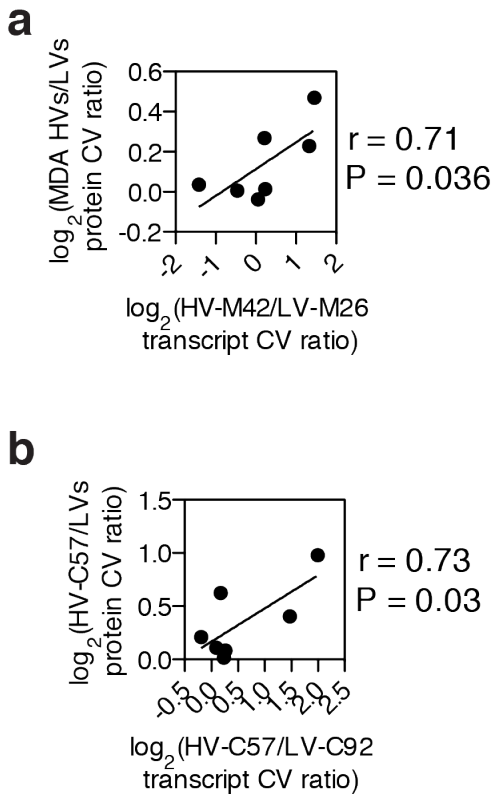

Figure 3.30: Cell-to-cell transcript expression variability is correlated with protein expression variability

Protein CV log-ratios of seven proteins as measured by flow cytometry from MDA-derived (**a**) and CN-derived (**b**) subpopulations was compared to transcript CV log-ratios measured by single-cell RNA sequencing. P-value was derived from testing Pearson's correlation coefficient with one-side.

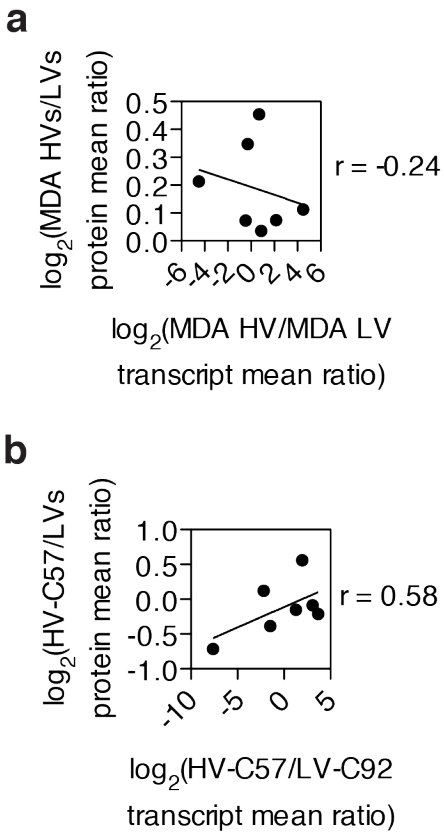

Figure 3.31: Mean transcript expression is not consistently correlated with mean protein expression

Protein mean expression log-ratios as measured by flow cytometry from MDA-derived (a) or CN-derived (b) was compared to transcript mean expression log-ratios as measured by single cell RNA sequencing.

3.5 Spliceosome-related gene transcripts are high variability

Validation of spliceosome-related gene variability

The next goal was to identify a potential mechanism that could contribute to transcriptomic variation in high variability subpopulations. MDA-derived and CN-derived subpopulations displayed similarity in high variability genes (Table 2.1), many of which alone could contribute to global transcriptomic variability such as chromatin modifiers (SENP7, ARID1A), transcription factors (TCF7L2, SP3), and regulators of nonsense-mediated decay (UPF2). While individual genes are likely to contribute to the effect seen in high variability populations, any coordinated variation of a common subset of transcripts may reveal a major contribution to cell-to-cell variability. To identify regulatory networks that might be responsible for transcriptomic variation, variable transcripts were searched for functional gene sets that might contribute to molecular variability. Transcripts were binned into four categories depending on whether relative transcript variability was increased in either or both MDA-derived and CN-derived high variability subpopulations. Pathway discovery analysis using iPAGE (Goodarzi et al., 2009) revealed spliceosome machinery and myeloid cell differentiation gene transcripts as the only two gene sets to exhibit higher expression variability in high variability subpopulations derived from both patients' cancer populations (Figure 3.32). This raised the possibility that cell-to-cell variation in the expression levels of splicing genes and the resultant mRNA processing

activity may represent a conceivable mechanism through which population-level heterogeneity of mature transcripts may be achieved at a global scale (Munsky et al., 2012).

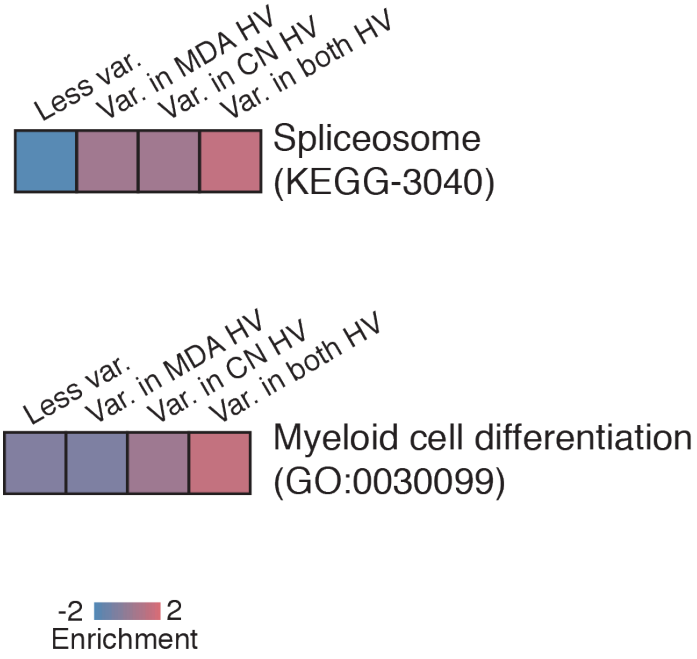

Figure 3.32: Pathway analysis reveals spliceosome machinery and myeloid cell differentiation gene transcripts as gene sets with variable expression

Pathway analysis was used to identify gene sets that are enriched in transcripts with high expression variability in either or both HV subpopulations. Enrichment scale is shown below.

Intron retention variability

Intron retention, a product of improper splicing, is known to lead to transcript degradation via non-sense mediated decay (Wong et al., 2013). Increased intron retention could thus reduce the expression of a large number of mature transcripts. Given the expression variation of spliceosome machinery components, variation in constitutive spliceosome activity in HV cells could cause enhanced variation in global intron retention. Variation in intron retention could then contribute to transcript abundance variation. Indeed, high variability subpopulations demonstrated higher cell-to-cell intron retention variation, as determined by analysis of 1,132 matched introns from MDA-derived cells (54.8% retained introns with higher CV in HV-M42, P = 1e-4 by one-sided paired *t*-test; Figure 2.33) and 1,666 matched introns from CN-derived cells (57.6% retained introns with higher CV in HV-C57, P = 2e-11 by one-sided paired *t*-test; Figure 3.33).

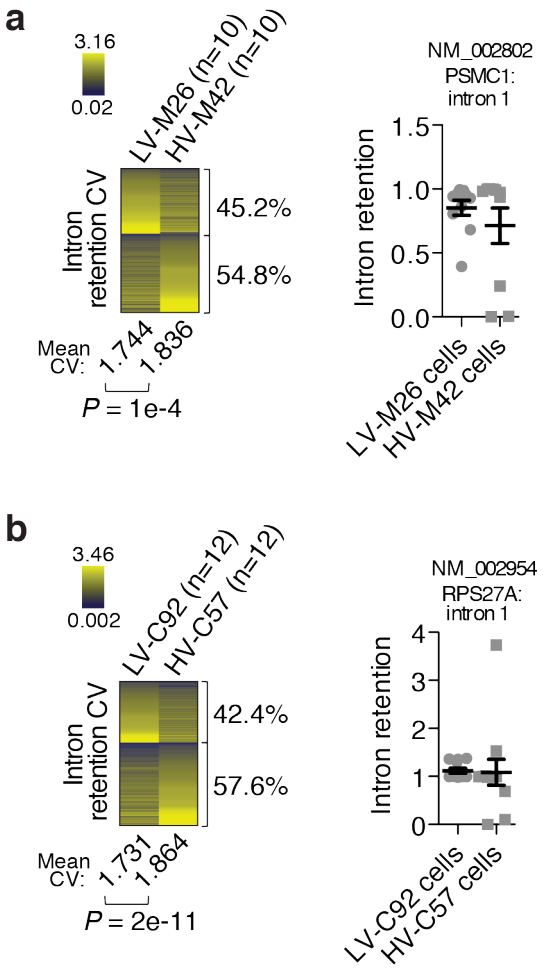

Figure 3.33: Intron retention variability is observed in high variability subpopulations

Intron retention (IR) variability analysis was performed by calculating expression coefficient of variation for each IR measurement between cells from MDA-derived (**a**) and CN-derived (**b**) subpopulations. Mean IR coefficient of variation was calculated. P-value was derived from one-sided paired t-test. Percentages of retained introns with higher CV in each population are shown. Representative high variability retained intron expression in single cells is shown (right); n=10 single cells for MDA-derived populations and n=12 single cells for CN-derived populations.

Exon-exon junction variability

Additionally, if this intron retention variation is propagated, variation should be apparent in further processed forms of pre-mRNA. Indeed, exon-exon junction expression variability measured from single cell sequencing experiments was significantly increased in high variability subpopulations from MDA-derived cells (54.7% exon-exon junctions with higher CV in HV-M42, P = 2e-4 by one-sided paired *t*-test; Figure 3.34) and CN-derived cells (65.3% exon-exon junctions with higher CV in HV-C57, P = 8e-10 by one-sided paired *t*-test; Figure 3.34). This molecular variability could be caused by variation in splicing as well as varied decay of improperly spliced transcripts. Consistent with our findings, intron retention has been previously observed to be the most significant splicing alteration in breast cancer patient samples (Eswaran et al., 2013). These findings reveal enhanced splicing and intron-retention variability as one feature of high variability subpopulations.

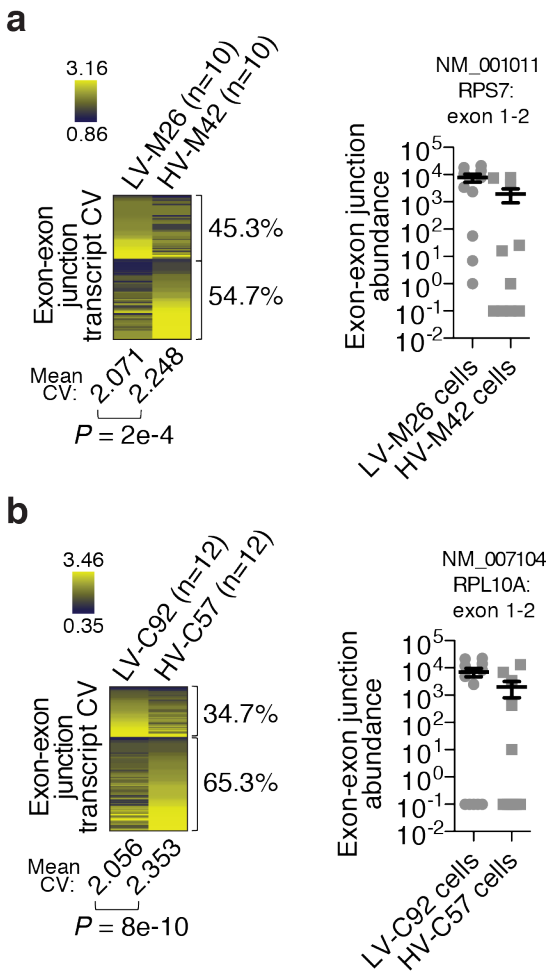

Figure 3.34: Exon-exon junction variability is observed in high variability subpopulations

Exon-exon junction transcript variability analysis was performed by calculating expression coefficient of variation for each exon-exon junction transcript between cells from MDA-derived (**a**) and CN-derived (**b**) subpopulations. Mean exon-exon junction transcript coefficient of variation was calculated. P-value was derived from one-sided paired t-test. Percentages of exon-exon junction transcripts with higher CV in each population are shown. Representative high variability exon-exon junction transcript expression in single cells is shown (right); n=10 single cells for MDA-derived populations and n=12 single cells for CN-derived populations.

3.6 Variable cell-to-cell expression of SNRNP40 increases metastatic colonization

Engineered variation of SNRNP40 increases metastatic colonization

To determine if spliceosomal gene variability contributed to increased metastatic colonization, the next goal was to assess whether enforced modulation of a high variability spliceosomal gene could recapitulate the metastatic capacity seen in high variability subpopulations. While high variability subpopulations display variability in many spliceosomal genes, focused study on a single spliceosomal gene could serve as a model by which other spliceosomal genes function, but may recapitulate only a fraction of the effects seen in high variability subpopulations. Focused study was performed on SNRNP40, a component of the U5 snRNP complex, because this gene exhibited high transcript variability in high variability subpopulations from single-cell sequencing experiments (Figure 3.35), was among the top 3 most variable spliceosomal gene transcripts from single-cell sequencing experiments (Table 3.2), and whose protein expression could be readily quantified by immunofluorescence-based imaging on confocal microscopy. Indeed, high variability subpopulations displayed increased protein-level variability of SNRNP40 (Figure 3.36).

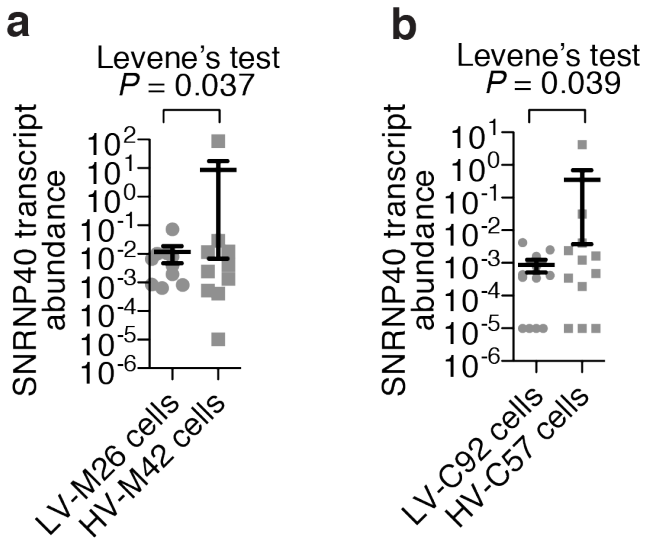

Figure 3.35: SNRNP40 transcript (NM_004814) is variable in high variability subpopulations

Single cell expression of SNRNP40 transcript (NM_004814) from MDA-derived (**a**) and CN-derived (**b**) subpopulations. P-value to test variation was performed using Levene's test.

Table 3.2: Top 5 most variable spliceosomal genes in common to MDA-derived and CN-derived high variability subpopulations

RefSeq	Gene	MDA HV/LV log2-ratio	CN HV/LV log2-ratio
NM_015484	SYF2	1.212	1.456
NM_003092	SNRPB2	1.252	1.100
NM_004814	SNRNP40	0.745	1.250
NM_003016	SRSF2	1.067	0.378
NM_004941	DHX8	0.459	0.807

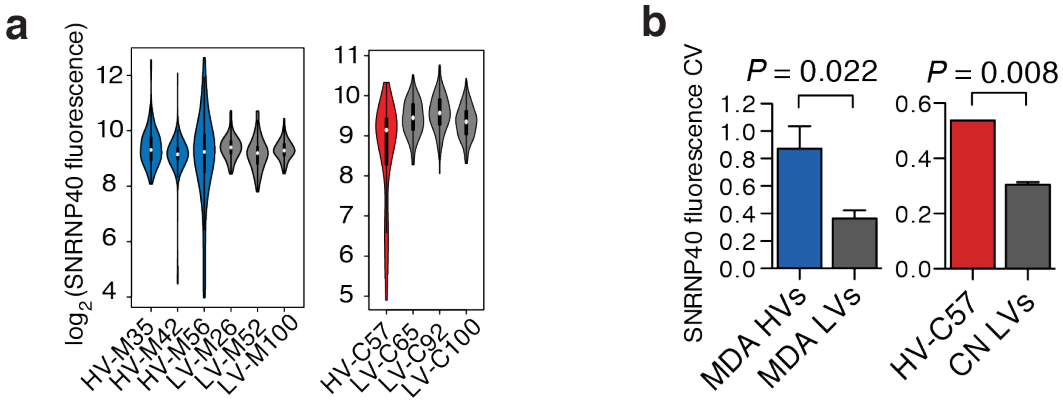

Figure 3.36: Variable cell-to-cell expression of SNRNP40 protein in high variability subpopulations

Subpopulations were quantified for SNRNP40 protein levels by fluorescence confocal microscopy (**a**) and analyzed for protein expression coefficient of variation (**b**). P-values were derived from two-sample (MDA) or one-sample (CN) one-sided *t*-test.

To determine whether population-level variation in SNRNP40 observed with high variability cells was sufficient to enforce metastatic fitness, a low variability subpopulation was used to enforce modulation of SNRNP40 expression. Populations of LV-M100 cells were individually transduced with SNRNP40 over-expression constructs and shRNAs targeting SNRNP40 at various titers to generate populations with varying levels of SNRNP40 (Figure 3.37a). These populations were pooled to generate a mixed population with increased cell-to-cell SNRNP40 expression variability without significantly altering mean SNRNP40 expression (Figure 3.37b; mean: Control = 9.0, hiCV=8.9, P = 0.19 by two-sided *t*-test; CV: Control = 0.068, hiCV= 0.104, P = 7e-14 by Levene's test). Functional testing of this population in metastasis revealed that increased

cell-to-cell variability in SNRNP40 expression enhanced the ability of these cells to metastasize more efficiently and colonize more sites systemically (Figure 3.37), suggesting that variation in SNRNP40 between cells could, in part, contribute to enhanced metastatic capacity observed in high variability subpopulations. Next, SNRNP40 over-expression and depletion cell lines were tested individually to determine whether the cells with increased or decreased SNRNP40 expression imparted the enhanced fitness to the pooled population. While increased SNRNP40 expression did not affect metastatic capacity, SNRNP40 depletion promoted systemic metastatic colonization (Figure 3.38).

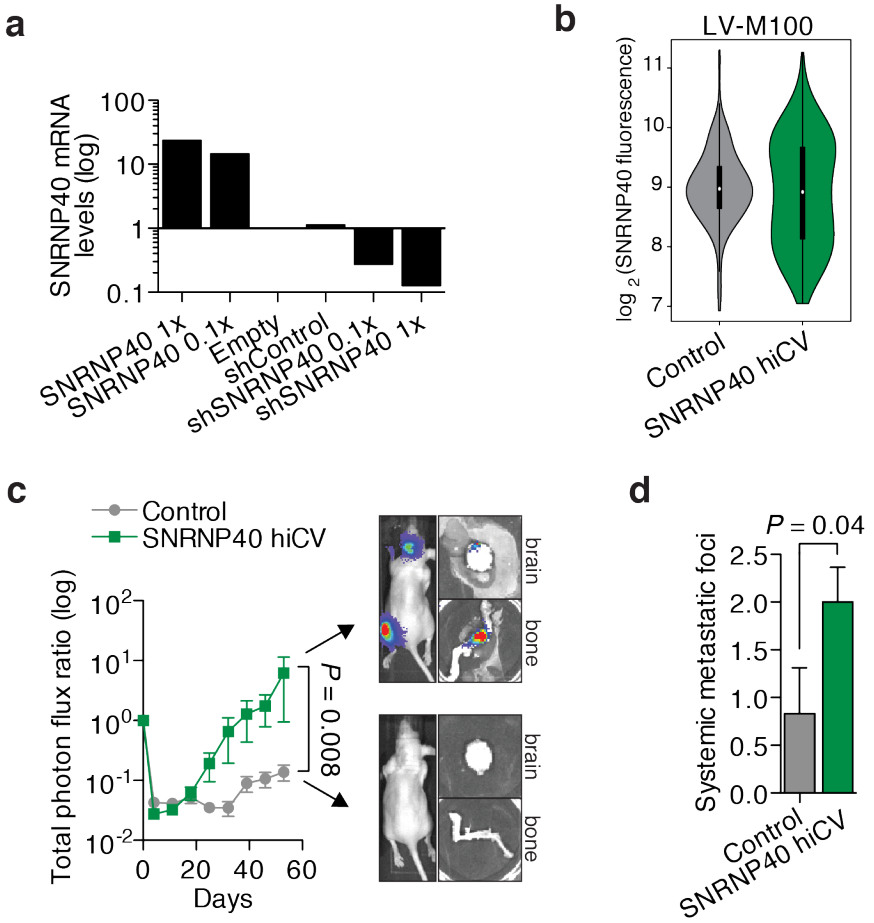

Figure 3.37: Engineered variation of SNRNP40 increases metastatic colonization

(**a**) Populations of LV-M100 cells were generated with over-expression of SNRNP40, shRNA-mediated knockdown of SNRNP40, empty vector, or control shRNA vector at varying titers to generate various population-level expression of SNRNP40 as measured by qRT-PCR. (**b**) Populations generated in (a) were pooled to generate a high SNRNP40 expression variability LV-M100 population and were assayed for protein expression by fluorescence microscopy. (**c**) 10^5 cells from pooled LV-M100 populations were inoculated via intracardiac injection and monitored by bioluminescence flux. Representative mice are shown. P-value was derived from one-sided *t*-test; n = 6. (**d**) Systemic metastatic foci per mice were counted as distinct, minimal bioluminescence signals. P-value was derived using one-sided *t*-test.

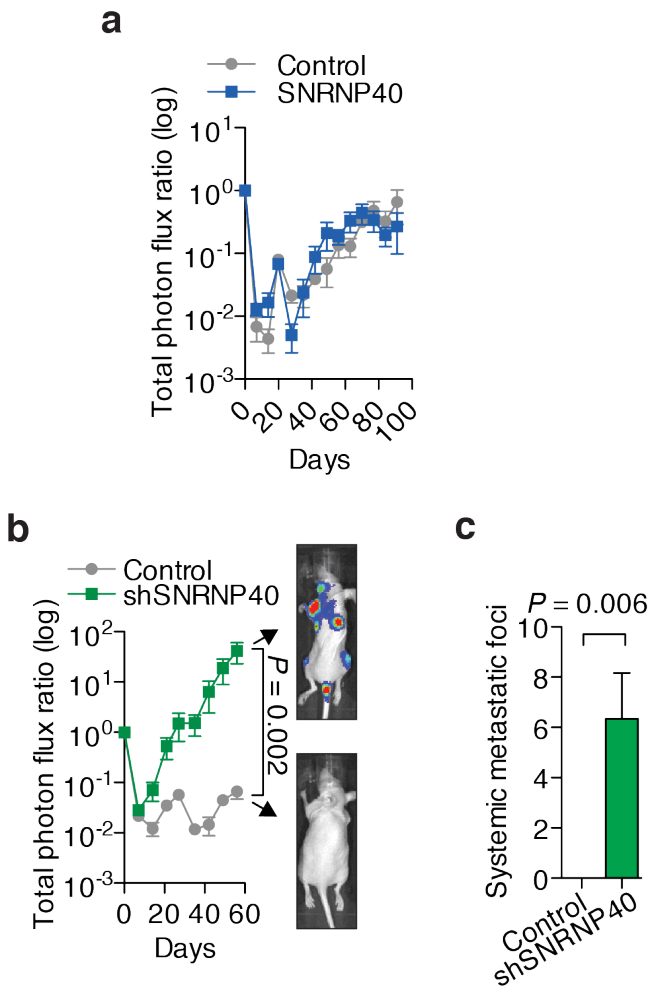

Figure 3.38: SNRNP40 depletion, not over-expression, promotes metastatic colonization

(**a**) 10^5 LV-M100 cells with SNRNP40 over-expression were inoculated by intracardiac injection and monitored by bioluminescence; n=5. (**b**) 10^5 LV-M100 cells with SNRNP40 depletion were inoculated by intracardiac injection and monitored by bioluminescence. Representative mice are shown; n=5. (**c**) Systemic metastatic foci per mice were counted. P-values were derived from one-sided t-test.

SNRNP40 regulates genes found to be variable in high variability subpopulations and is required for proper intron exclusion

The next goal was to determine whether SNRNP40 contributed to the gene expression variability seen in high variability subpopulations. The working model was that variable levels of SNRNP40 allows for transmission of variability to many transcripts dependent on SNRNP40 expression. Thus, siRNA-mediated depletion of SNRNP40 followed by RNA-sequencing was performed to identify these dependent transcripts with high magnitude of expression change indicative of strong transmission. If variable expression of SNRNP40 contributes to gene expression variability seen in high variability subpopulations, SNRNP40-dependent transcripts would be enriched among variable transcripts in high variability subpopulations. Indeed, transcripts with high magnitude expression change in SNRNP40 knockdown cells were enriched among variable transcripts between high variability single cells (Figure 3.39a). Additionally, absolute log-fold expression change significantly correlated with single-cell transcript variability ratios (Figure 3.39b). These findings suggest that SNRNP40 contributes in part to the expression of a set of variable transcripts in breast cancer cells. Additionally, SNRNP40 depletion led to an increase in intron retention (Figure 3.40), consistent with intron retention variability observed in high variability subpopulations, although the modest effect may be attributed to perturbation of only a single spliceosomal gene.

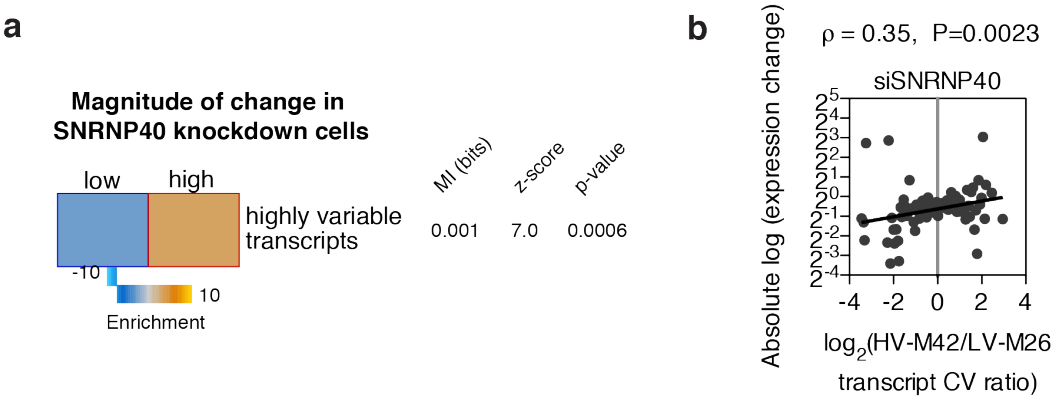

Figure 3.39: SNRNP40 regulates gene expression consistent with gene expression variability seen in highly variable subpopulations.

LV-M100 cells were transfected with two siRNAs targeting SNRNP40 and processed for gene expression analysis. (**a**) Absolute log-fold change was calculated to determine the magnitude of change. High magnitude change was defined as absolute log-fold change > 1. Highly variable transcripts were defined using single-cell RNA-sequencing experiments as $\log_2$(HV-M42/LV-M26 transcript CV ratio) > 1.5. The calculated mutual information value (in bits) and z-score are provided. P-value was derived using two-sided Fisher's exact test. Also shown are the enrichment scores, presented as logP (positive for enrichments and negative for depletions), where P is calculated from hypergeometric distribution (shown as a heatmap with blue and gold showing depletion and enrichment, respectively). (**b**) Gene expression variability ratios of 6,906 transcripts derived from single-cell RNA-sequencing were binned into 100 bins. For transcripts in each bin, the absolute log-fold expression change of SNRNP40 siRNA treatment relative to control siRNAs was calculated, averaged, and plotted. Robust linear regression is shown in black. Two-sided spearman's correlation and corresponding p-value are shown.

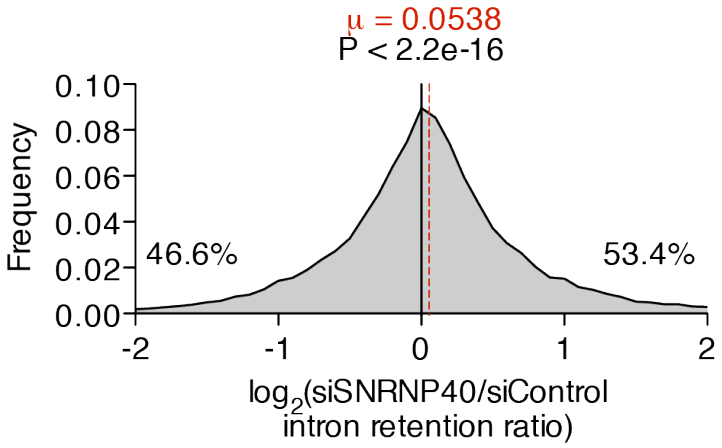

Figure 3.40: SNRNP40 promotes proper intron exclusion
Log-fold intron retention ratio was calculated for each transcript and plotted as a histogram. Dashed red line indicates mean. Percentages of intron retention ratios above and below zero are shown. P-value was derived from two-sided t-test of log-ratio as compared to zero.

Low SNRNP40 is associated with metastasis in human breast cancer

Given the experimental role of SNRNP40 in metastasis, human breast cancer samples were assessed for SNRNP40 expression to determine if this gene was associated with clinical outcomes. Indeed, decreased SNRNP40 transcript-level expression in bulk primary breast cancer sampled was associated with increased metastatic relapse outcomes in multiple independent datasets (Figure 3.41). Furthermore, low SNRNP40 copy number correlated with worse overall survival (Figure 3.41). These findings provide clinical association support for SNRNP40 expression in human breast cancer progression. While tumors can achieve reduced SNRNP40 expression through mean expression or

copy number changes, our experimental model suggests that cancer subpopulations may achieve SNRNP40 silencing through deregulation in a subset of cells that is not easily observed using averaged measurements.

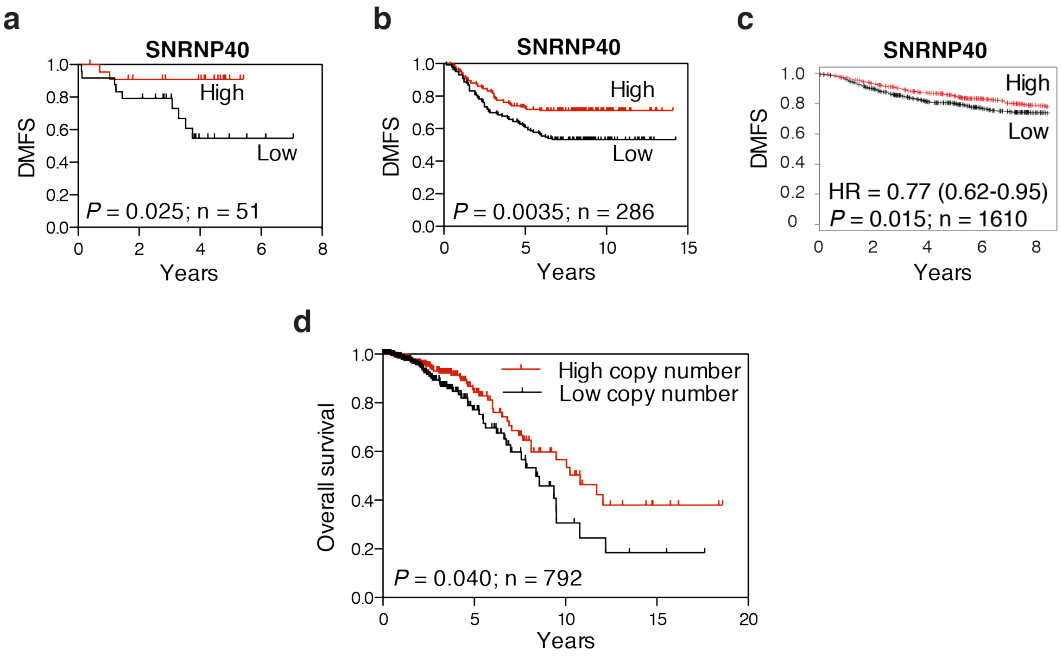

Figure 3.41: Low SNRNP40 expression is associated with metastasic outcomes in human breast cancer

Kaplan-Meier curves of distant metastasis-free survival (DMFS) from GSE33926 (Kuo et al., 2012) (a), GSE2034 (Wang et al., 2005) (b) and kmplotter meta-analysis (Gyorffy et al., 2010) (c) as a function of primary tumor SNRNP40 expression in breast cancer. (d) Kaplan-Meier curve of overall survival from TCGA samples based on SNRNP40 copy number (Cancer Genome Atlas, 2012). P-values were derived from log-rank test.

3.7 Discussion

Cell-to-cell gene expression variability

Gene expression variability in cell populations can be driven by a combination of stochastic and deterministic processes (Hanna et al., 2010). While stochastic changes can impact population fitness, deterministic processes can more reliably benefit evolution over multiple generations. Single-cell transcriptomic analyses have revealed deterministic molecular programs that regulate gene expression heterogeneity in induced pluripotent stem cell differentiation and immune cells responses (Buganim et al., 2012; Shalek et al., 2013). In cancer, analysis of gene expression cell-to-cell variability has been limited though has mainly pertained to stem-like characteristics of individual cells within glioblastoma tumors (Patel et al., 2014). The work presented here illustrates another regulatory program involved in cell-to-cell gene expression variability, namely variable splicesomal gene expression between cells as a catalyst for gene expression variability.

Splicing as regulator of global gene expression networks

While constitutive splicing serves a necessary processing function to generate protein-coding transcripts from genomic sequences, minimal deregulation of splicing factors has the potential for amplified alterations of gene regulatory networks and gene expression states. Key splicing factors are understood to be required for maintenance of robust transcriptomes as feedback loops exist to maintain steady-state levels and

direct molecular programs (Jangi and Sharp, 2014). Furthermore, bone-marrow-derived dendritic cell populations in response to lipopolysaccharide display variability in splice isoforms between cells (Shalek et al., 2013). Consistent with the findings presented here, heterogeneous transcriptional signatures and splice isoform variation have been observed in glioblastoma (Patel et al., 2014) and likely contribute to cancer progression. Additionally, intron retention has been observed to regulate the expression of genes involved in nuclear shape as well as splicing factor genes (Wong et al., 2013), indicating a complex regulatory network as well as phenotypic consequences associated with perturbed splicing.

Upstream source of molecular variability

While spliceosome gene expression variability has the potential to contribute in part to the observed transcriptomic and phenotypic variability in high variability subpopulations, the upstream cause of this transcript-level variability remains unknown. One approach to characterize the source is to determine whether the observed variability is due to dynamic variability within a single cell, namely high fluctuations in gene expression over time, or stable variation between cells, namely certain cells always express low levels while other cells always express high levels. Dynamic variability within cells would suggest that molecular instability at the level of transcription or stability mediate transcript-level variability, whereas stable variation between cells would suggest unequal

cell divisions might be responsible. This question can be addressed by using a fluorescent reporter of endogenous transcripts that display high variability and performing live cell imaging to visualize protein expression levels between cells and over time.

Given the maintained genomic integrity in these populations, epigenetic alterations are a likely candidate for the source of variability. Histone modifications and methylation of CpG islands are known to contribute to cancer progression and have known roles in affecting robustness of transcriptional activity (Rodriguez-Paredes and Esteller, 2011). Future experiments include analysis of transcriptional activity of variable genes as well as analysis of DNA methylation in single cells within highly variable subpopulations. Furthermore, if epigenetic alterations are contributing to phenotypic and molecular variability, the use of epigenetic drugs has the potential to suppress variability and slow cancer evolution.

Additionally, the class of genes known as phenotypic capacitors have the potential to contribute to transcript-level variability. Phenotypic capacitors function to buffer noise and maintain phenotypic robustness; thus, upon failure of a capacitor under external pressures, systems lose robustness and result in greater phenotypic variability. The first capacitor described was Hsp90, whose decreased expression resulted in uncovering of genetically-encoded morphological variation as well as developmental plasticity (Queitsch et al., 2002; Rutherford and Lindquist, 1998). To assess whether capacitors could contribute to variability observed in breast cancer populations, loss-of-function screens could be performed to

identify capacitors that regulate phenotypes and molecular functions identified in highly variable subpopulations. These screens have previously been performed in yeast and have identified chromatin remodeling complexes as capacitors (Levy and Siegal, 2008), supporting the notion of epigenetic alterations as a potential contributor to cell-to-cell variability.

3.7 Materials and Methods

Cell culture

MDA-MB-231, CN34, and derived sub-population cells were propagated as previously described (Tavazoie et al., 2008). MDA-MB-231 cells and their derivatives were maintained in DMEM supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin and Fungizone. CN34 cells and their derivatives were maintained in M199 supplemented with 2.5% FBS, 10 μg ml^{-1} insulin, 0.5 μg ml^{-1} hydrocortisone, 20 ng ml^{-1} EGF, 100 ng ml^{-1} cholera toxin, glutamine, pyruvate, penicillin, streptomycin and Fungizone. Cells in culture were routinely tested for mycoplasma contamination. Clonal populations were generated by seeding cells sparsely, picking individual colonies using cloning discs (Sigma), and expanding the populations to approximately 10^5 cells when cells were seeded to be imaged for size measurements.

Cell size measurements and coefficient of variation analysis

$3x10^4$ cells were seeded on coverslips and stained with HCS CellMask Red (Invitrogen) and DAPI. Numerous fields were imaged on DeltaVision Image Restoration Microscope to capture at least 100 cells per population. Cell size parameters were measured using Cell Profiler 2.0. Debris was filtered out by generating a histogram of (cytoplasmic area – nuclear area) and applying a minimum threshold in R. To calculate coefficient of variation (standard deviation / mean) for each clonal population, a sampling size was determined from the clonal population with the fewest

cells analyzed (96 cells in the MDA-MB-231 population and 206 cells in the CN34 population). Subpopulations with more cells imaged than the sampling size were sampled 100 times, with the average CV used in the final analysis. For each subpopulation, coefficient of variation was calculated for each size parameter, which includes cell area, cytoplasmic area, nucleus, perimeter, major axis length, and minor axis length. Principal component analysis was performed using all subpopulations on the coefficient of variations for all morphological parameters. The first principal component accounted for 93% and 91% of the variance in MDA and CN34 cell lines, respectively, and is used for all size variation analyses.

Three-dimensional size coefficient of variation analysis

Cells were fixed in 8% paraformaldehyde, permeabilized with 0.1% Triton-X, and stained with HCS CellMask Red. Cells were analyzed by ImageStream-X (Amnis) to measure cell size, nuclear size, perimeter, major axis, and minor axis. Principal component analysis was performed, and the first principal component was used for analyses.

Proliferation & colony formation assays

For bulk population proliferation assays, $5x10^3$ cells were seeded in triplicate and assayed WST-1 reagent (Roche) 72 hours after seeding. For colony formation assays, 200 cells were seeded in 10cm plates in triplicate and were allowed to grow for 10 days (MDA-derived cells) or 20 days

(CN-derived cells). Plates were fixed in 6% glutaraldehyde with 0.5% crystal violet and scanned. Colony areas were measured using ImageJ. R was used to remove debris, equalize colony numbers with samplings as above, and calculate coefficient of variation.

Animal studies

Animal experiments were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee at The Rockefeller University. For *in vivo* experiments with MDA-derived cell lines, populations were transduced with a retroviral construct expressing a luciferase reporter (Ponomarev et al., 2004). CN34 parental cell line was previously labeled with luciferase reporter. Athymic female mice were used for intracardiac injection. NOD-SCID female mice were used for tail vein injections with MDA-derived cells, while NOD-SCID gamma female mice were used for tail vein injections with CN-derived cells. For intracardiac injections, cells in PBS were injected into the left ventricle of mice in a volume of 100ul. For tail vein injections, cells in PBS were injected into the tail vein of mice in a volume of 100ul. Portal circulation injections were performed as described in chapter 2.8. For portal circulation injection of mixed population, LV-M100 cells were transduced with pLKO.1 puro (Addgene plasmid #8453), and HV-M42 cells were transduced with pLKO.1 blast (Addgene plasmid #26655). Cells were mixed at a 1:1 ratio immediately before injections. *In vivo* bioluminescence was performed as described in chapter 2.8.

Tissue microarray analysis

NCI CDP Breast Cancer Progression Tissue Microarray slides were deparaffinized, rehydrated, and exposed to Heat Induced Epitope Retrieval at pH 6 (Vector Labs) in a pressure cooker for 4 minutes. Slides were stained with DAPI, and tumor cores were imaged on Leica TCS SP5 system at 40x. Images were analyzed using CellProfiler 2.0 to identify and measure cancer cell nuclei by Otsu Adaptive thresholding. A minimum nuclei size threshold was applied to remove stromal cell nuclei. Image acquisition and analysis were blinded until measurements were completed for all tumor cores. Mitotic index was calculated as follows: # mitotic cells / total cells for each tumor core.

Exome sequencing and analysis

Genomic DNA was extracted using DNEasy kit (Qiagen). Libraries were prepared using Nextera Extended Exome sequencing kit, as per manufacturer's instructions (Illumina), and paired-end sequenced on HiSeq 2500 (Illumina). The analysis pipeline for the exome-seq data was based on the GATK best practices. Reads were aligned to the human genome (build hg19). The paired-end reads were then fixed and filtered using Picard (v. 1.107; http://picard.sourceforge.net/). The duplicates were also removed in the same step. Using GATK (v. 2.5) (McKenna et al., 2010), the reads were realigned and recalibrated. mpileup (samtools (Li et al., 2009)) was used to create an input for VarScan (v2.3.6) (Koboldt et al.,

2012). In VarScan, mpileup2snp and mpileup2indel commands were used to identify variants across all exome-seq samples. To study population genetic divergence, the frequencies of all variants identified (148,234) were used to generate phylogenic tree by Nei's genetic distance using neighbor-joining method in the PHYLIP package.

RNA-sequencing of bulk populations

For bulk population RNA-sequencing, RNA was extracted using total RNA isolation kit (Norgen Biotek) with DNAse I treatment followed by Ribo-Zero Gold rRNA removal (Epicentre). Libraries were generated using ScriptSeq v2 RNA-seq Library Preparation Kit (Epicentre) and run on HiSeq 2500 (Illumina). For RNA-sequencing data analysis, the reads were trimmed to remove matches to linkers and low-quality bases (cutadapt v1.2.1). Tophat (v. 2.0.8)(Kim et al., 2013) was then used to map the reads to the human transcriptome (build hg19). Cufflinks and cuffmerge (v.2.0.2) were then used to calculate reads per kilo base per million (RPKM) and consolidated results across the samples. Finally, cuffdiff (v.2.0.2) was used to calculate log-fold changes and the associated statistics.

Single cell RNA-sequencing

Single cell isolation, cDNA synthesis, amplification, and processing for Illumina sequencing were performed as described in detail (Islam et al., 2012). To describe the protocol briefly, single cells were sorted by fluorescence-activated cell sorting

into 96-well plates containing cell lysis buffer, a biotin-labeled oligo-dT primer for capturing poly-adenylated mRNA, and a template-switching barcoded oligonucleotide to label transcripts from each cell. This template-switching oligonucleotide is incorporated at the 5′ end of complementary DNA generated in the next step. Once cells were captured, Superscript II Reverse Transcriptase (Invitrogen) was used under required buffer conditions to generate complementary DNA that contained the biotin-labeled oligo-dT on the 3′ end and the barcoded oligonucleotide on the 5′ end. Complementary DNA was captured and purified using MyOne carboxylate Dynabeads (Life Technologies), then PCR amplified using Advantage 2 DNA polymerase (Clontech) using a biotinylated oligonucleotide. An aliquot of cDNA was checked for quality by agarose gel electrophoresis. Complementary DNA was next captured and purified using MyOne C1 streptavidin Dynabeads (Life Technologies). The complementary DNA was then processed for Illumina-based sequencing by fragmentation, end repair, dA tailing, adapter ligation, PCR amplification, purification with AMPure XP beads (Beckman Coulter), and gel size selection for amplicons 200-400bp in size. Samples were then sequencing on HighSeq 2500 (Illumina). Ambion ArrayControl RNA spikes (Ambion) were used in each well to assess sequencing quality and fidelity. For data analysis, reads were distributed into separate samples based on their barcodes. Cells were excluded if wells were empty or generated library was low quality as assessed by the number of total reads, number of mapped reads, a high percentage of unmapped reads and a high percentage of spike-in reads (Stegle et al.,

2015). To equalize the number of cells analyzed from each population, cells were randomly selected and removed from analysis. The RNA-sequencing pipeline described above was then used to measure gene expression across RNA-sequencing data from each cell. In parallel, TopHat results were parsed to count the presence of every exon-exon junction across all the samples. Transcripts present in both cell populations, present in more than 25% of all cells and expressed above threshold mean of all cells based on a Gaussian distribution were included for analysis. Non-linear cell clustering by t-Distributed Stochastic Neighbor Embedding was performed using 'tsne' package in R. To measure intron retention, the number of reads mapping to each exon or intron were counted across the transcriptome for all samples. As a measure of intron retention (IR), for each intron, the number of reads mapping to that intron relative to the two spanning exons (r_int_ex) was calculated. For CV measurements, retained introns and exon-exon junctions were included in analysis if present in more than 25% of all cells.

Pathway analysis

iPAGE (Goodarzi et al., 2009) pathway analysis (http://iget.c2b2.columbia.edu) was used to identify gene sets with higher transcript variability in both HV subpopulations. Transcript coefficient of variation log-ratios (HV / LV) were used to categorize transcripts into four bins: bin 1 = negative log-ratio in both MDA-derived and CN-derived comparisons, bins 2&3 = positive log-ratio in either

MDA-derived or CN-derived comparison, and bin 4= positive log-ratio in both MDA-derived and CN-derived comparisons. GO and KEGG annotations were analyzed using a maximum p-value of 0.05 and maximum genes per category of 200. Gene sets were only considered relevant if enrichment was highest in bin 4 and lowest in bin 1.

Flow cytometry

For antibody staining, cells were prepared using Cytofix/Cytoperm (BD). Dead cells were excluded using Live/DEAD Aqua (Invitrogen). Primary antibodies used were anti-XPNPEP3 (Abcam, 1:200), anti-UPF2 (LSBio, 1:4) anti-ALDOA (Abcam, 1:8), anti-PABP (Abcam; 1:200), anti-HNRNPA1 (Cell Signaling, 1:30), anti-CD110 (BD, 1:50), anti-HNRNPA0 (Cell Signaling, 1:50), anti-ESR1 (Thermo, 1:20), anti-MCF2 (1:10, LSBio, 1:10), and anti-CSF2RA (1:30, eBiosciences) conjugated to Alexa555 or Alexa647 Zenon secondary antibodies (Invitrogen). Data aquisition was performed on LSRII (BD). Data analysis was performed on FloJo. CV calculations were performed using greater than $25x10^3$ cells, an equal number of cells per sample within each experiment. Ratios were calculated using the average value of all high variability populations (MDA: HV-M35, HV-M42, HV-M56; CN: HV-C57) and low variability populations (MDA: LV-M26, LV-M52, LV-M100; CN: LV-C65, LV-C92, LV-C100). Cell cycle analysis was performed by employing flow cytometry on fixed cells stained with DAPI and determining cell cycle phases in FloJo.

SNRNP40 protein quantitation

Cells were seeded onto coverslips, fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, and stained with DAPI and anti-SNRNP40 antibody (HPA026527, 1:100, Sigma) followed by fluorescent-conjugated secondary antibodies. Imaging was performed on Leica TCS SP5 system. SNRNP40 relative protein level was determined using Cell Profiler 2.0 by measuring total nuclear SNRNP40 fluorescence intensity as demarcated by DAPI signal.

SNRNP40 cell line generation

To generate high CV SNRNP40 population, cell populations were transduced individually at various titers with either virus for shRNA expression (shRNA#1 TRCN0000074608 and shRNA#2 TRCN0000074610, Sigma) or virus for stable ORF expression (pBabe puro vector, Addgene #1764). Expression in individual populations was confirmed by quantitative RT-PCR. Populations were then pooled at equal ratios to generate mixed populations. Control population was generated similarly using a non-targeting shRNA (SHC016, Sigma) and an empty expression vector.

Quantitative RT-PCR

RNA was extracted using total RNA isolation kit (Norgen Biotek). cDNA was generated using Superscript III (Invitrogen). Fast SYBR Green Master Mix (Life Technologies) was used to analyze samples on Applied

Biosystems 7900HT. Expression was normalized to GAPDH. Primers sequences are as follows: SNRNP40 Forward: CAGTGGAGCAGTGATGGAAT; SNRNP40 Reverse: CCCTCTCACCTGTTTCACTATC; GAPDH Forward: AGCCACATCGCTCAGACAC; GAPDH Reverse: GCCCAATACGACCAAATCC; blasticidin Forward: CCTGGGATCAAAGCCATAGT; blasticidin Reverse: TTAGCCCTCCCACACATAAC; puromycin Forward: GTCACCGAGCTGCAAGAA; puromycin Reverse: CCGATCTCGGCGAACAC

siRNA transfection

The following siRNAs were used (IDT): siSNRNP40 #1: GGAAUAGACAAUGAUAUC; siSNRNP40 #2: GGAUUUGACCGACUGAUA. BLOCK-iT Fluorescent Oligo (Life Technologies) and NC1 (IDT) were used as controls.

10^5 cells were seeded and were transfected the next day with siRNAs via Lipofectaime 2000 (Invitrogen). Cells were extracted 48 hours later for validation of knock-down by quantitative RT-PCR or for RNA-sequencing. For RNA-sequencing, two siRNAs were used for SNRNP40 and control, and independent siRNA replicates were averaged.

Clinical samples

GSE2034 (Wang et al., 2005) and GSE33926 (Kuo et al., 2012) were used to generate Kaplan-Meier curves. Patients were stratified by SNRNP40 expression relative to median. KMplotter was used to assess distant metastasis-free survival with follow up threshold of 8 years (Gyorffy et al., 2010). TCGA sample data was stratified by SNRNP40 copy number relative to median and was used to assess overall survival.(Cancer Genome Atlas, 2012)

CHAPTER 4: Final Summary

The work presented in this thesis addresses two aspects of metastatic colonization, the most significant clinical event in advanced cancer progression. The first part of this thesis explored the cellular and molecular functions of a kinase that promotes liver colonization by colorectal cancer, which is a significant cause of cancer mortality. A systematic, unbiased approach was utilized to identify genes that promote liver colonization in a mouse model of metastasis. PKLR, which encodes the liver and red blood cell isozymes of pyruvate kinase, was identified as a promoter of colonization of the liver in an organ-specific manner. Assessment of PKLR expression in patient samples revealed an association of PKLR levels with metastatic disease, consistent with functional testing results. Characterization of the cellular defect upon PKLR depletion revealed the requirement of PKLR for cell survival in the tumor core, a survival defect that can be recapitulated *in vitro* under conditions of high cell density and hypoxia. While PKLR functions under physiologic conditions to promote glycolysis in hepatocytes and red blood cells, PKLR promotes survival by negatively regulating the glycolytic enzymatic activity of the predominant pyruvate kinase expressed in cancer, PKM2, and allows for the maintenance of the primary intracellular antioxidant, glutathione (Illustration 4.2). Furthermore, inhibition of glutathione synthesis by RNAi as well as small-molecule therapy led to a suppression of metastatic colonization, indicating the potential for clinical utility. Our findings reveal the use of multiple pyruvate kinase isozymes by colon cancer for enhanced regulatory control of glycolysis and antioxidant generation. Furthermore, the microenvironment conditions

that render colon cancer cells susceptible to cell death highlight the metabolically demanding conditions upon liver colonization and reveal potential for therapeutic targeting of these pathways.

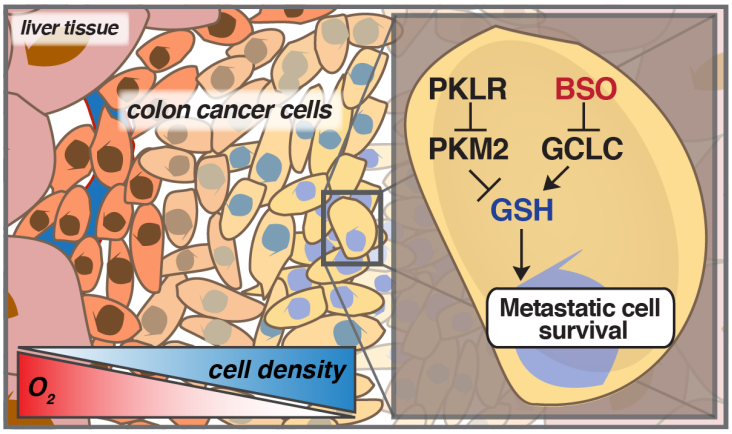

Illustration 4.2: Regulation of glycolytic metabolism for metastasis

The second half of this thesis focused on the study of non-genetic diversity as a contributor to metastatic fitness. First, imaging analysis of clonal subpopulations derived from human breast cancer cells lines was performed to enable the identification of subpopulations that generate phenotypic variability. Characterization of these high variability subpopulations revealed that phenotypic variability was heritable in progeny and was relevant in multiple phenotypes. Consistent with a positive role of diversification in cancer progression, high variability subpopulations displayed increased metastatic fitness. Variability did not appear to be genetic in nature as subpopulation genomic integrity appeared intact. However, single cell transcriptomic analyses reveal

elevated cell-to-cell gene expression variability that was maintained within clonal cancer subpopulations. Furthermore, spliceosome-associated gene expression variability was identified as one potential mechanism by which clonal cancer populations could increase mature transcript expression variability of target genes. One example of a functional role for spliceosome gene expression variability was explored where engineered variable cell-to-cell expression of SNRNP40 enabled enhanced metastatic fitness through existence of a subpopulation with low SNRNP40 expression (Illustration 4.3). Importantly, gene expression profiles upon SNRNP40 deregulation was consistent with this gene's contribution to gene expression variability observed in high variability subpopulations. The consistent molecular variability observed in high variability subpopulations derived from independent human cancer cell populations was measured under different experimental assays and at both the molecular and phenotypic levels, suggesting that the variability is molecularly conserved and deterministic in nature. These experimental observations can be tested in clinical correlates to characterize non-genetic contributions to tumor evolution in patients.

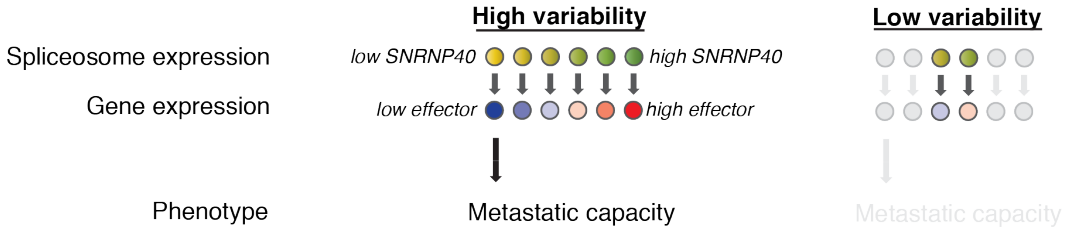

Illustration 4.1: Model of spliceosome expression variability mediating gene expression changes and metastatic capacity.

REFERENCES

(1992). Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. Advanced Colorectal Cancer Meta-Analysis Project. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *10*, 896-903.

Aktipis, C.A., Boddy, A.M., Gatenby, R.A., Brown, J.S., and Maley, C.C. (2013). Life history trade-offs in cancer evolution. Nature reviews Cancer *13*, 883-892.

Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? Cell *141*, 559-563.

Amir el, A.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature biotechnology *31*, 545-552.

Anastasiou, D., Poulogiannis, G., Asara, J.M., Boxer, M.B., Jiang, J.K., Shen, M., Bellinger, G., Sasaki, A.T., Locasale, J.W., Auld, D.S., *et al.* (2011). Inhibition of pyruvate kinase M2 by reactive oxygen species contributes to cellular antioxidant responses. Science *334*, 1278-1283.

Anastasiou, D., Yu, Y., Israelsen, W.J., Jiang, J.K., Boxer, M.B., Hong, B.S., Tempel, W., Dimov, S., Shen, M., Jha, A., *et al.* (2012). Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. Nature chemical biology *8*, 839-847.

Araten, D.J., Golde, D.W., Zhang, R.H., Thaler, H.T., Gargiulo, L., Notaro, R., and Luzzatto, L. (2005). A quantitative measurement of the human somatic mutation rate. Cancer research *65*, 8111-8117.

Badiola, I., Villace, P., Basaldua, I., and Olaso, E. (2011). Downregulation of discoidin domain receptor 2 in A375 human melanoma cells reduces its experimental liver metastasis ability. Oncology reports *26*, 971-978.

Bailey, H.H., Ripple, G., Tutsch, K.D., Arzoomanian, R.Z., Alberti, D., Feierabend, C., Mahvi, D., Schink, J., Pomplun, M., Mulcahy, R.T., *et al.*

(1997). Phase I study of continuous-infusion L-S,R-buthionine sulfoximine with intravenous melphalan. Journal of the National Cancer Institute *89*, 1789-1796.

Bao, S., Ouyang, G., Bai, X., Huang, Z., Ma, C., Liu, M., Shao, R., Anderson, R.M., Rich, J.N., and Wang, X.F. (2004). Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. Cancer cell *5*, 329-339.

Baylin, S.B., and Herman, J.G. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends in genetics : TIG *16*, 168-174.

Beronja, S., Janki, P., Heller, E., Lien, W.H., Keyes, B.E., Oshimori, N., and Fuchs, E. (2013). RNAi screens in mice identify physiological regulators of oncogenic growth. Nature *501*, 185-190.

Bielas, J.H., Loeb, K.R., Rubin, B.P., True, L.D., and Loeb, L.A. (2006). Human cancers express a mutator phenotype. Proceedings of the National Academy of Sciences of the United States of America *103*, 18238-18242.

Bluemlein, K., Gruning, N.M., Feichtinger, R.G., Lehrach, H., Kofler, B., and Ralser, M. (2011). No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis. Oncotarget *2*, 393-400.

Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A., *et al.* (2009). Genes that mediate breast cancer metastasis to the brain. Nature *459*, 1005-1009.

Boutros, M., and Ahringer, J. (2008). The art and design of genetic screens: RNA interference. Nature reviews Genetics *9*, 554-566.

Braun, S., Vogl, F.D., Naume, B., Janni, W., Osborne, M.P., Coombes, R.C., Schlimok, G., Diel, I.J., Gerber, B., Gebauer, G., *et al.* (2005). A pooled analysis of bone marrow micrometastasis in breast cancer. The New England journal of medicine *353*, 793-802.

Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. Nature reviews Genetics *10*, 336-342.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209-1222.

Cabezas, H., Raposo, R.R., and Melendez-Hevia, E. (1999). Activity and metabolic roles of the pentose phosphate cycle in several rat tissues. Molecular and cellular biochemistry *201*, 57-63.

Cailleau, R., Olive, M., and Cruciger, Q.V. (1978). Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. In vitro *14*, 911-915.

Cairns, R.A., Harris, I.S., and Mak, T.W. (2011). Regulation of cancer cell metabolism. Nature reviews Cancer *11*, 85-95.

Cancer Genome Atlas, N. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61-70.

Chaneton, B., Hillmann, P., Zheng, L., Martin, A.C., Maddocks, O.D., Chokkathukalam, A., Coyle, J.E., Jankevics, A., Holding, F.P., Vousden, K.H., *et al.* (2012). Serine is a natural ligand and allosteric activator of pyruvate kinase M2. Nature *491*, 458-462.

Chiang, A.C., and Massague, J. (2008). Molecular basis of metastasis. The New England journal of medicine *359*, 2814-2823.

Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., Fleming, M.D., Schreiber, S.L., and Cantley, L.C. (2008a). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature *452*, 230-233.

Christofk, H.R., Vander Heiden, M.G., Wu, N., Asara, J.M., and Cantley, L.C. (2008b). Pyruvate kinase M2 is a phosphotyrosine-binding protein. Nature *452*, 181-186.

Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z., *et al.* (2008). Dynamic proteomics of individual cancer cells in response to a drug. Science *322*, 1511-1516.

Corley, D.A., Jensen, C.D., Marks, A.R., Zhao, W.K., Lee, J.K., Doubeni, C.A., Zauber, A.G., de Boer, J., Fireman, B.H., Schottinger, J.E., *et al.* (2014). Adenoma detection rate and risk of colorectal cancer and death. The New England journal of medicine *370*, 1298-1306.

Curreri, A.R., Ansfield, F.J., Mc, I.F., Waisman, H.A., and Heidelberger, C. (1958). Clinical studies with 5-fluorouracil. Cancer research *18*, 478-484.

Davies, J.M., and Goldberg, R.M. (2011). Treatment of metastatic colorectal cancer. Seminars in oncology *38*, 552-560.

de Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J., *et al.* (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science *346*, 251-256.

de Gramont, A., Figer, A., Seymour, M., Homerin, M., Hmissi, A., Cassidy, J., Boni, C., Cortes-Funes, H., Cervantes, A., Freyer, G., *et al.* (2000). Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *18*, 2938-2947.

DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G., and Thompson, C.B. (2008). The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. Cell metabolism *7*, 11-20.

Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., *et al.* (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature *464*, 999-1005.

Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D., *et al.* (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature *481*, 506-510.

Domingo, M., Einig, C., Eigenbrodt, E., and Reinacher, M. (1992). Immunohistological demonstration of pyruvate kinase isoenzyme type L in rat with monoclonal antibodies. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society *40*, 665-673.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Eswaran, J., Horvath, A., Godbole, S., Reddy, S.D., Mudvari, P., Ohshiro, K., Cyanam, D., Nair, S., Fuqua, S.A., Polyak, K., *et al.* (2013). RNA sequencing of cancer reveals novel splicing alterations. Scientific reports *3*, 1689.

Ewing, J. (1928). Neoplastic diseases; a treatise on tumors, 3d ed rev. and enl., with 546 illustrations. edn (Philadelphia, London,: W.B. Saunders).

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759-767.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nature reviews Genetics *7*, 21-33.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2014). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer Journal international du cancer.

Fidler, I.J. (1970). Metastasis: guantitative analysis of distribution and fate of tumor embolilabeled with 125 I-5-iodo-2'-deoxyuridine. Journal of the National Cancer Institute *45*, 773-782.

Fidler, I.J. (1986). Rationale and methods for the use of nude mice to study the biology and therapy of human cancer metastasis. Cancer metastasis reviews *5*, 29-49.

Fidler, I.J. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nature reviews Cancer *3*, 453-458.

Fidler, I.J., and Kripke, M.L. (1977). Metastasis results from preexisting variant cells within a malignant tumor. Science *197*, 893-895.

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell *75*, 1027-1038.

Franklin, C.C., Backos, D.S., Mohar, I., White, C.C., Forman, H.J., and Kavanagh, T.J. (2009). Structure, function, and post-translational regulation of the catalytic and modifier subunits of glutamate cysteine ligase. Molecular aspects of medicine *30*, 86-98.

Garcia, C.K., Goldstein, J.L., Pathak, R.K., Anderson, R.G., and Brown, M.S. (1994). Molecular characterization of a membrane transporter for lactate, pyruvate, and other monocarboxylates: implications for the Cori cycle. Cell *76*, 865-873.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., *et al.* (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine *366*, 883-892.

Gerlinger, M., and Swanton, C. (2010). How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. British journal of cancer *103*, 1139-1143.

Glimelius, B., Hoffman, K., Graf, W., Pahlman, L., and Sjoden, P.O. (1994). Quality of life during chemotherapy in patients with symptomatic advanced colorectal cancer. The Nordic Gastrointestinal Tumor Adjuvant Therapy Group. Cancer *73*, 556-562.

Goodarzi, H., Elemento, O., and Tavazoie, S. (2009). Revealing global regulatory perturbations across human cancers. Molecular cell *36*, 900-911.

Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306-313.

Griffith, O.W. (1982). Mechanism of action, metabolism, and toxicity of buthionine sulfoximine and its higher homologs, potent inhibitors of glutathione synthesis. The Journal of biological chemistry *257*, 13704-13712.

Griffith, O.W. (1999). Biologic and pharmacologic regulation of mammalian glutathione synthesis. Free radical biology & medicine *27*, 922-935.

Gupta, G.P., and Massague, J. (2006). Cancer metastasis: building a framework. Cell *127*, 679-695.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633-644.

Gyorffy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast cancer research and treatment *123*, 725-731.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508-525.

Hart, I.R., and Fidler, I.J. (1980). Role of organ selectivity in the determination of metastatic patterns of B16 melanoma. Cancer research *40*, 2281-2287.

Heidelberger, C., Chaudhuri, N.K., Danneberg, P., Mooren, D., Griesbach, L., Duschinsky, R., Schnitzer, R.J., Pleven, E., and Scheiner, J. (1957). Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. Nature *179*, 663-666.

Hickson, J., Ackler, S., Klaubert, D., Bouska, J., Ellis, P., Foster, K., Oleksijew, A., Rodriguez, L., Schlessinger, S., Wang, B., *et al.* (2010). Noninvasive molecular imaging of apoptosis in vivo using a modified firefly luciferase substrate, Z-DEVD-aminoluciferin. Cell death and differentiation *17*, 1003-1010.

Horie-Inoue, K., and Inoue, S. (2006). Epigenetic and proteolytic inactivation of 14-3-3sigma in breast and prostate cancers. Seminars in cancer biology *16*, 235-239.

Imamura, K., and Tanaka, T. (1972). Multimolecular forms of pyruvate kinase from rat and other mammalian tissues. I. Electrophoretic studies. Journal of biochemistry *71*, 1043-1051.

Imhof, M., and Schlotterer, C. (2006). E. coli microcosms indicate a tight link between predictability of ecosystem dynamics and diversity. PLoS genetics *2*, e103.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nature protocols *7*, 813-828.

Jangi, M., and Sharp, P.A. (2014). Building Robust Transcriptomes with Master Splicing Factors. Cell *159*, 487-498.

Jones, S., Chen, W.D., Parmigiani, G., Diehl, F., Beerenwinkel, N., Antal, T., Traulsen, A., Nowak, M.A., Siegel, C., Velculescu, V.E., *et al.* (2008). Comparative lesion sequencing provides insights into tumor evolution. Proceedings of the National Academy of Sciences of the United States of America *105*, 4283-4288.

Jungermann, K., and Kietzmann, T. (1996). Zonation of parenchymal and nonparenchymal metabolism in liver. Annual review of nutrition *16*, 179-203.

Juric, D., Castel, P., Griffith, M., Griffith, O.L., Won, H.H., Ellis, H., Ebbesen, S.H., Ainscough, B.J., Ramu, A., Iyer, G., *et al.* (2015). Convergent loss of PTEN leads to clinical resistance to a PI(3)Kalpha inhibitor. Nature *518*, 240-244.

Kaelin, W.G., Jr., and McKnight, S.L. (2013). Influence of metabolism on epigenetics and disease. Cell *153*, 56-69.

Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. (2003). A multigenic program mediating breast cancer metastasis to bone. Cancer cell *3*, 537-549.

Ki, D.H., Jeung, H.C., Park, C.H., Kang, S.H., Lee, G.Y., Lee, W.S., Kim, N.K., Chung, H.C., and Rha, S.Y. (2007). Whole genome analysis for liver metastasis gene signatures in colorectal cancer. International journal of cancer Journal international du cancer *121*, 2005-2012.

Kikuchi, A., Ishikawa, T., Mogushi, K., Ishiguro, M., Iida, S., Mizushima, H., Uetake, H., Tanaka, H., and Sugihara, K. (2013). Identification of NUCKS1 as a colorectal cancer prognostic marker through integrated expression and copy number analysis. International journal of cancer Journal international du cancer *132*, 2295-2302.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Kim, S.K., Kim, S.Y., Kim, J.H., Roh, S.A., Cho, D.H., Kim, Y.S., and Kim, J.C. (2014). A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. Molecular oncology.

Klein, C.A. (2006). Random mutations, selected mutations: A PIN opens the door to new genetic landscapes. Proceedings of the National Academy of Sciences of the United States of America *103*, 18033-18034.

Klein, C.A. (2013). Selection and adaptation during metastatic cancer progression. Nature *501*, 365-372.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research *22*, 568-576.

Kreso, A., O'Brien, C.A., van Galen, P., Gan, O.I., Notta, F., Brown, A.M., Ng, K., Ma, J., Wienholds, E., Dunant, C., *et al.* (2013). Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. Science *339*, 543-548.

Kuo, W.H., Chang, Y.Y., Lai, L.C., Tsai, M.H., Hsiao, C.K., Chang, K.J., and Chuang, E.Y. (2012). Molecular characteristics and metastasis predictor genes of triple-negative breast cancer: a clinical study of triple-negative breast carcinomas. PloS one *7*, e45831.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M., *et al.* (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell *75*, 1215-1225.

Levy, S.F., and Siegal, M.L. (2008). Network hubs buffer environmental variation in Saccharomyces cerevisiae. PLoS biology *6*, e264.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Loeb, L.A. (1991). Mutator phenotype may be required for multistage carcinogenesis. Cancer research *51*, 3075-3079.

Loo, J.M., Scherl, A., Nguyen, A., Man, F.Y., Weinberg, E., Zeng, Z., Saltz, L., Paty, P.B., and Tavazoie, S.F. (2015). Extracellular metabolic energetics can promote cancer progression. Cell *160*, 393-406.

Lunt, S.Y., Muralidhar, V., Hosios, A.M., Israelsen, W.J., Gui, D.Y., Newhouse, L., Ogrodzinski, M., Hecht, V., Xu, K., Acevedo, P.N., *et al.* (2014). Pyruvate Kinase Isoform Expression Alters Nucleotide Synthesis to Impact Cell Proliferation. Molecular cell.

Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J.S., Beroukhim, R., Weir, B.A., *et al.* (2008). Highly parallel identification of essential genes in cancer cells. Proceedings of the National Academy of Sciences of the United States of America *105*, 20380-20385.

Luo, W., and Semenza, G.L. (2012). Emerging roles of PKM2 in cell metabolism and cancer progression. Trends in endocrinology and metabolism: TEM *23*, 560-566.

Luzzi, K.J., MacDonald, I.C., Schmidt, E.E., Kerkvliet, N., Morris, V.L., Chambers, A.F., and Groom, A.C. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. The American journal of pathology *153*, 865-873.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912-1934.

Markowitz, S.D., and Bertagnolli, M.M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. The New England journal of medicine *361*, 2449-2460.

Marusyk, A., and Polyak, K. (2010). Tumor heterogeneity: causes and consequences. Biochimica et biophysica acta *1805*, 105-117.

Marusyk, A., Tabassum, D.P., Altrock, P.M., Almendro, V., Michor, F., and Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature *514*, 54-58.

Matsuyama, T., Ishikawa, T., Mogushi, K., Yoshida, T., Iida, S., Uetake, H., Mizushima, H., Tanaka, H., and Sugihara, K. (2010). MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer. International journal of cancer Journal international du cancer *127*, 2292-2299.

Mazurek, S. (2011). Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells. The international journal of biochemistry & cell biology *43*, 969-980.

Mazurek, S., Boschek, C.B., Hugo, F., and Eigenbrodt, E. (2005). Pyruvate kinase type M2 and its role in tumor growth and spreading. Seminars in cancer biology *15*, 300-308.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research *20*, 1297-1303.

Mehlen, P., and Puisieux, A. (2006). Metastasis: a question of life or death. Nature reviews Cancer *6*, 449-458.

Merlo, L.M., Pepper, J.W., Reid, B.J., and Maley, C.C. (2006). Cancer as an evolutionary and ecological process. Nature reviews Cancer *6*, 924-935.

Minn, A.J., Kang, Y., Serganova, I., Gupta, G.P., Giri, D.D., Doubrovin, M., Ponomarev, V., Gerald, W.L., Blasberg, R., and Massague, J. (2005). Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. The Journal of clinical investigation *115*, 44-55.

Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. Science *336*, 183-187.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., *et al.* (2011). Tumour evolution inferred by single-cell sequencing. Nature *472*, 90-94.

Nishihara, R., Wu, K., Lochhead, P., Morikawa, T., Liao, X., Qian, Z.R., Inamura, K., Kim, S.A., Kuchiba, A., Yamauchi, M., *et al.* (2013). Long-term colorectal-cancer incidence and mortality after lower endoscopy. The New England journal of medicine *369*, 1095-1105.

Nordlinger, B., Guiguet, M., Vaillant, J.C., Balladur, P., Boudjema, K., Bachellier, P., and Jaeck, D. (1996). Surgical resection of colorectal carcinoma metastases to the liver. A prognostic scoring system to improve case selection, based on 1568 patients. Association Francaise de Chirurgie. Cancer *77*, 1254-1262.

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

Obrador, E., Benlloch, M., Pellicer, J.A., Asensi, M., and Estrela, J.M. (2011). Intertissue flow of glutathione (GSH) as a tumor growth-promoting mechanism: interleukin 6 induces GSH release from hepatocytes in metastatic B16 melanoma-bearing mice. The Journal of biological chemistry *286*, 15716-15727.

Osterman, J., and Fritz, P.J. (1974). Pyruvate kinase isozymes from rat intestinal mucosa. Characterization and the effect of fasting and refeeding. Biochemistry *13*, 1731-1736.

Paget, S. (1989). The distribution of secondary growths in cancer of the breast. 1889. Cancer metastasis reviews *8*, 98-101.

Park, S.Y., Gonen, M., Kim, H.J., Michor, F., and Polyak, K. (2010). Cellular and genetic diversity in the progression of in situ human breast

carcinomas to an invasive phenotype. The Journal of clinical investigation *120*, 636-644.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., *et al.* (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396-1401.

Pencheva, N., Buss, C.G., Posada, J., Merghoub, T., and Tavazoie, S.F. (2014). Broad-spectrum therapeutic suppression of metastatic melanoma through nuclear hormone receptor activation. Cell *156*, 986-1001.

Pencheva, N., Tran, H., Buss, C., Huh, D., Drobnjak, M., Busam, K., and Tavazoie, S.F. (2012). Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. Cell *151*, 1068-1082.

Png, K.J., Halberg, N., Yoshida, M., and Tavazoie, S.F. (2012). A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. Nature *481*, 190-194.

Ponomarev, V., Doubrovin, M., Serganova, I., Vider, J., Shavrin, A., Beresten, T., Ivanova, A., Ageyeva, L., Tourkova, V., Balatoni, J., *et al.* (2004). A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. European journal of nuclear medicine and molecular imaging *31*, 740-751.

Poon, M.A., O'Connell, M.J., Moertel, C.G., Wieand, H.S., Cullinan, S.A., Everson, L.K., Krook, J.E., Mailliard, J.A., Laurie, J.A., Tschetter, L.K., *et al.* (1989). Biochemical modulation of fluorouracil: evidence of significant improvement of survival and quality of life in patients with advanced colorectal carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *7*, 1407-1418.

Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. Nature medicine *19*, 1423-1437.

Queitsch, C., Sangster, T.A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. Nature *417*, 618-624.

Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C., and Perucho, M. (1997). Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science *275*, 967-969.

Rees, M., Tekkis, P.P., Welsh, F.K., O'Rourke, T., and John, T.G. (2008). Evaluation of long-term survival after hepatic resection for metastatic colorectal cancer: a multifactorial model of 929 patients. Annals of surgery *247*, 125-135.

Reinmuth, N., Fan, F., Liu, W., Parikh, A.A., Stoeltzing, O., Jung, Y.D., Bucana, C.D., Radinsky, R., Gallick, G.E., and Ellis, L.M. (2002). Impact of insulin-like growth factor receptor-I function on angiogenesis, growth, and metastasis of colon cancer. Laboratory investigation; a journal of technical methods and pathology *82*, 1377-1389.

Renan, M.J. (1993). How many mutations are required for tumorigenesis? Implications from human cancer data. Molecular carcinogenesis *7*, 139-146.

Renault, B., Calistri, D., Buonsanti, G., Nanni, O., Amadori, D., and Ranzani, G.N. (1996). Microsatellite instability and mutations of p53 and TGF-beta RII genes in gastric cancer. Human genetics *98*, 601-607.

Rodriguez-Paredes, M., and Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. Nature medicine *17*, 330-339.

Root, D.E., Hacohen, N., Hahn, W.C., Lander, E.S., and Sabatini, D.M. (2006). Genome-scale loss-of-function screening with a lentiviral RNAi library. Nature methods *3*, 715-719.

Rutherford, S.L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. Nature *396*, 336-342.

Rygaard, J., and Povlsen, C.O. (1969). Heterotransplantation of a human malignant tumour to "Nude" mice. Acta pathologica et microbiologica Scandinavica *77*, 758-760.

Saheki, S., Harada, K., Sanno, Y., and Tanaka, T. (1978). Hybrid isozymes of rat pyruvate kinase. Their subunit structure and developmental changes in the liver. Biochimica et biophysica acta *526*, 116-128.

Saltz, L.B., Cox, J.V., Blanke, C., Rosen, L.S., Fehrenbacher, L., Moore, M.J., Maroun, J.A., Ackland, S.P., Locker, P.K., Pirotta, N., *et al.* (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. The New England journal of medicine *343*, 905-914.

Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., Ang, X.L., *et al.* (2008). Cancer proliferation gene discovery through functional genomics. Science *319*, 620-624.

Schulze, A., and Harris, A.L. (2012). How cancer metabolism is tuned for proliferation and vulnerable to disruption. Nature *491*, 364-373.

Seligson, D.B., Horvath, S., Shi, T., Yu, H., Tze, S., Grunstein, M., and Kurdistani, S.K. (2005). Global histone modification patterns predict risk of prostate cancer recurrence. Nature *435*, 1262-1266.

Semenza, G.L. (2007). Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1. The Biochemical journal *405*, 1-9.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., *et al.* (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236-240.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., *et al.* (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69-80.

Shea, K., and Chesson, P. (2002). Community ecology theory as a framework for biological invasions. Trends Ecol Evol *17*, 170-176.

Sheffer, M., Bacolod, M.D., Zuk, O., Giardina, S.F., Pincas, H., Barany, F., Paty, P.B., Gerald, W.L., Notterman, D.A., and Domany, E. (2009). Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America *106*, 7131-7136.

Siegel, R., Ma, J., Zou, Z., and Jemal, A. (2014). Cancer statistics, 2014. CA: a cancer journal for clinicians *64*, 9-29.

Siegel, R.L., Miller, K.D., and Jemal, A. (2015). Cancer statistics, 2015. CA: a cancer journal for clinicians *65*, 5-29.

Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science *319*, 617-620.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. Molecular cell *55*, 319-331.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428-432.

Spudich, J.L., and Koshland, D.E., Jr. (1976). Non-genetic individuality: chance in the single cell. Nature *262*, 467-471.

Stange, D.E., Engel, F., Longerich, T., Koo, B.K., Koch, M., Delhomme, N., Aigner, M., Toedt, G., Schirmacher, P., Lichter, P.*, et al.* (2010). Expression of an ASCL2 related stem cell signature and IGF2 in colorectal cancer liver metastases with 11p15.5 gain. Gut *59*, 1236-1244.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nature reviews Genetics *16*, 133-145.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Sullivan, W.J., and Christofk, H.R. (2015). The metabolic milieu of metastases. Cell *160*, 363-364.

Tabaries, S., Dong, Z., Annis, M.G., Omeroglu, A., Pepin, F., Ouellet, V., Russo, C., Hassanain, M., Metrakos, P., Diaz, Z., *et al.* (2011). Claudin-2 is selectively enriched in and promotes the formation of breast cancer liver metastases through engagement of integrin complexes. Oncogene *30*, 1318-1328.

Talmadge, J.E., and Fidler, I.J. (2010). AACR centennial series: the biology of cancer metastasis: historical perspective. Cancer research *70*, 5649-5669.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533-538.

Tarin, D., Price, J.E., Kettlewell, M.G., Souter, R.G., Vass, A.C., and Crossley, B. (1984). Mechanisms of human tumor metastasis studied in patients with peritoneovenous shunts. Cancer research *44*, 3584-3592.

Tavazoie, S.F., Alarcon, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massague, J. (2008). Endogenous human microRNAs that suppress breast cancer metastasis. Nature *451*, 147-152.

Thangaraju, M., Carswell, K.N., Prasad, P.D., and Ganapathy, V. (2009). Colon cancer cells maintain low levels of pyruvate to avoid cell death caused by inhibition of HDAC1/HDAC3. The Biochemical journal *417*, 379-389.

Tokutake, N., Hiratake, J., Katoh, M., Irie, T., Kato, H., and Oda, J. (1998). Design, synthesis and evaluation of transition-state analogue inhibitors of Escherichia coli gamma-glutamylcysteine synthetase. Bioorganic & medicinal chemistry *6*, 1935-1953.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., *et al.* (2010). Towards a knowledge-based Human Protein Atlas. Nature biotechnology *28*, 1248-1250.

Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. Cell *147*, 275-292.

Van Cutsem, E., Kohne, C.H., Hitre, E., Zaluski, J., Chang Chien, C.R., Makhson, A., D'Haens, G., Pinter, T., Lim, R., Bodoky, G., *et al.* (2009).

Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. The New England journal of medicine *360*, 1408-1417.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029-1033.

Vermeulen, P.B., Colpaert, C., Salgado, R., Royers, R., Hellemans, H., Van Den Heuvel, E., Goovaerts, G., Dirix, L.Y., and Van Marck, E. (2001). Liver metastases from colorectal adenocarcinomas grow in three patterns with different angiogenesis and desmoplasia. The Journal of pathology *195*, 336-342.

Virchow, R. (1989). Cellular pathology. As based upon physiological and pathological histology. Lecture XVI--Atheromatous affection of arteries. 1858. Nutrition reviews *47*, 23-25.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., and Bos, J.L. (1988). Genetic alterations during colorectal-tumor development. The New England journal of medicine *319*, 525-532.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J.*, et al.* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671-679.

Warburg, O. (1956). On the origin of cancer cells. Science *123*, 309-314.

Watanabe, T., Kobunai, T., Yamamoto, Y., Matsuda, K., Ishihara, S., Nozawa, K., Iinuma, H., Konishi, T., Horie, H., Ikeuchi, H.*, et al.* (2011). Gene expression signature and response to the use of leucovorin, fluorouracil and oxaliplatin in colorectal cancer patients. Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico *13*, 419-425.

Watanabe, T., Sagisaka, H., Arakawa, S., Shibaya, Y., Watanabe, M., Igarashi, I., Tanaka, K., Totsuka, S., Takasaki, W., and Manabe, S. (2003). A novel model of continuous depletion of glutathione in mice treated with L-buthionine (S,R)-sulfoximine. The Journal of toxicological sciences *28*, 455-469.

Weigelt, B., Peterse, J.L., and van 't Veer, L.J. (2005). Breast cancer metastasis: markers and models. Nature reviews Cancer *5*, 591-602.

Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., *et al.* (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nature genetics *38*, 787-793.

Whitehurst, A.W., Bodemann, B.O., Cardenas, J., Ferguson, D., Girard, L., Peyton, M., Minna, J.D., Michnoff, C., Hao, W., Roth, M.G., *et al.* (2007). Synthetic lethal screen identification of chemosensitizer loci in cancer cells. Nature *446*, 815-819.

Winawer, S.J., and Zauber, A.G. (2002). The advanced adenoma as the primary target of screening. Gastrointestinal endoscopy clinics of North America *12*, 1-9, v.

Wong, J.J., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., *et al.* (2013). Orchestrated intron retention regulates normal granulocyte differentiation. Cell *154*, 583-595.

Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A., *et al.* (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature *467*, 1114-1117.

Yamada, K., and Noguchi, T. (1999). Nutrient and hormonal regulation of pyruvate kinase gene expression. The Biochemical journal *337 (Pt 1)*, 1-11.

Zauber, A.G., Winawer, S.J., O'Brien, M.J., Lansdorp-Vogelaar, I., van Ballegooijen, M., Hankey, B.F., Shi, W., Bond, J.H., Schapiro, M., Panish, J.F., *et al.* (2012). Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. The New England journal of medicine *366*, 687-696.

Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., *et al.* (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524-528.