

THE GENETIC EVOLUTION AND NATURAL HISTORY  
OF PANCREATIC ADENOCARCINOMA

by

Alvin P. Makohon-Moore

A dissertation submitted to Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March 20, 2015

## **Abstract**

Pancreatic cancer evolves via the step-wise accumulation of genetic mutations, yet the dynamics of this process are unknown. Multiple cancer tissues from a patient – collected via biopsy, surgery, or autopsy – enable analyses with profound implications for treatment as well as for understanding tumor evolution. My thesis focuses on the mutations acquired during two critical transitions in pancreatic cancer: the advancement of the precursor lesion to the primary tumor, and the evolution of metastatic disease. For the former, we sequenced 21 exomes from tumor precursors and matched cancers. We observed clonality of all concomitant lesions, even when multiple precursors existed within a patient. Yet, most precursors acquired unique mutations – some with numbers comparable to the matched cancer – indicating genetic divergence during carcinogenesis. In addition, known cancer drivers were detected among other somatically acquired alterations. Current efforts aim to determine the order of mutations and the genetic heterogeneity of the lesions – ultimately, a mathematical model will facilitate analysis. For the evolution of metastasis, we analyzed 26 distinct metastatic tumors from four end stage, treatment-naive patients. Using a quantitative measure of genetic relatedness, we found that pancreatic cancers and their metastases demonstrated a level of relatedness that was markedly higher than that expected for any two cells randomly taken from a normal tissue. This minimal amount of genetic divergence among very large, distinct, advanced lesions indicates that genetic heterogeneity, when quantitatively defined, is not a fundamental feature of the natural history of untreated pancreatic cancers. Overall, these analyses reveal the evolutionary history of pancreatic cancer – as recorded by genetic mutations – from initiation to metastasis.

## **Dissertation Readers:**

**Christine Iacobuzio-Donahue**

**Professor of Pathology**

Department of Pathology, The David M. Rubenstein Center for Pancreatic Cancer Research, Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.

**Nickolas Papadopoulos**

**Professor of Oncology**

The Ludwig Center, the Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA

## **Acknowledgements**

The first year of my thesis began with the Pathobiology program. At that time, Wilhelmena Braswell supported the graduate students through the requirements and challenges. Since then, Nancy Nath and Tracie McElroy have perpetuated the program forward – I thank all three for their help. In addition, I thank Dr. Noel Rose – while his passion for scientific approaches to disease may not be in the recruiting brochure, it is a central theme for the program. I also thank my fellow Pathobiology students for creating an encouraging atmosphere, critical for intellectual growth. In particular, I have enjoyed many scientific and personal insights over weekly lunches with Joshua Wang – I am confident there are many more meals to muse over.

I officially joined the Iacobuzio-Donahue laboratory on May 24, 2011. Today, some of the original 11 members remain; some new ones have arrived and departed in the meantime. All of them, however, have taken part in a highly successful and supportive group. Natalie Cole and Karen Blair facilitated many of the key details, including scheduling, funding submissions, and the all-important pizza parties. Richard Morgan and Tyler Saunders, both technicians, maintained a steady flow of pipette tips and good humor. Three other graduate students set a high bar for thesis work: Cate White, Jackie Brosnan-Cashman, and Anne Macgregor-Das – I thank them for unknowable volumes of technical advice and encouragement. I have also enjoyed working with five post-docs: Drs. Yi Zhong, Karen Matsukuma, Salva Naranjo, Vicente Valero, and Justin Poling. I have learned many techniques and perspectives from each of them. My time in the lab did not overlap with Dr. Shinichi Yachida, yet my work directly stems from his tremendous efforts (I was also lucky to inherit his cubicle, highly regarded real estate on our floor).

I have highly benefitted working within the G.I. Pathology department and the larger cancer research community at JHU. Specifically, I thank Drs. Nicholas Roberts and Laura Wood for providing feedback on my projects. I also thank Dr. James Eshleman for support throughout my thesis: first, as interim supervisor, then continuing for grant writing and my thesis committee. I am grateful to Dr. Ralph Hruban for his leadership of our department and for his direct support of my endeavors through project funding and letters of recommendation. I thank Dr. Christopher Wolfgang for many insights the opportunity to observe pancreatic cancer surgery first-hand. I am grateful to Drs. Kathy Burns and Nemanja Rodic for the opportunity to help with a fascinating question in cancer genetics. I also want to thank Dr. Steve Leach for chairing my thesis committee and for insights which challenged (and ultimately improved) this work.

For my thesis, I interacted closely with two research groups in particular. I thank Drs. Martin Nowak and Ivana Bozic for their invaluable insights, and Johannes Reiter for many pages of emails regarding analytical challenges and shared scientific interests. I also had the privilege of working with the Vogelstein group. Fay Wong and Drs. Yuchen Jiao and Ming Zhang provided technical expertise for analyzing cancer genomics data – and for teaching me in the process. I also thank Dr. Nickolas Papadopoulos for many discussions of genetic analysis, for participating in my committee, and for reading this thesis. I thank Dr. Kenneth Kinzler for sharing his expertise of cancer genetics and sequencing as well as for always answering my (often inexpert) questions. I also thank Dr. Bert Vogelstein for countless scientific insights, discussions, and for guiding me through a challenging project while always welcoming my own observations.

I now turn to my mentor, Dr. Christine Iacobuzio-Donahue, by simply stating that she continues to encourage my scientific curiosity and career. Undoubtedly, her intellectual guidance drives the scientific work, but I also want to acknowledge her mentorship directly: Chris has challenged me to consider scientific questions carefully and creatively. She has fostered a consistently supportive environment (with a touch of good humor), inspiring my continued interest in how tumors evolve. I look to her as an example of scientific ingenuity and leadership.

Of course, I thank my family for their continued love and support. I have very much enjoyed discussing science with my younger sister, Sasha (although I suspect I glean more insight from these conversations than she does). She has recently begun her own scientific career – I am already proud of her accomplishments and excited to see what is next. I also want to thank my parents, Norman and Livia, for consistently celebrating my decisions – education was not a choice, it was required; yet, I was free to pursue whatever specific path I chose. My father taught me how to refine arguments and modes of thinking through many discussions and the occasional philosophical dispute. My mother exemplified deep and passionate consideration of people and ideas, while never forgetting the overall humor and absurdity of each. The final thank-you goes to my fiancée, Kara. More than any other, she has encouraged me to push forward, never doubting the outcome or direction we have taken together. I do not have enough space to list all of the late nights, early mornings, and mumbled musings she has dealt with over the past five years – but I do remember each and every one was met with a smile (and perhaps the occasional eye-roll). I can only hope to be as tireless in my own love and support for her.

# Contents

Abstract.....	ii
Dissertation Readers:.....	iii
Acknowledgements.....	iv
Tables.....	viii
Figures.....	ix
Chapter 1 – Pancreatic cancer evolves.....	1
Challenges.....	1
Genetics.....	3
Evolution.....	6
Metastasis.....	10
Chapter 2 – Evolution of pancreatic cancer precursors.....	16
Synopsis.....	16
Introduction.....	18
Results.....	22
Conclusions.....	34
Chapter 3 – Natural history of pancreatic cancer metastasis.....	38
Synopsis.....	38
Introduction.....	40
Results.....	53
Conclusions.....	88
Chapter 4 – Methods and Supplementary Data.....	94
General.....	94
PanINs.....	95
Metastasis.....	101
References.....	112
Curriculum vitae.....	119

## Tables

<b>Table 1-1.</b> Common genetic drivers in pancreatic cancer.....	5
<b>Table 2-1.</b> Cases selected for PanIN and cancer whole exome sequencing.....	20
<b>Table 2-2.</b> Mutation numbers detected via whole exome sequencing.....	23
<b>Table 2-3.</b> Genetic drivers and mutation numbers.....	24
<b>Table 3-1.</b> Clinical characteristics of four pancreatic cancer patients.....	43
<b>Table 3-2.</b> Samples used for Whole genome (WGS) and directed sequencing.....	44
<b>Table 3-3.</b> Samples analyzed.....	45
<b>Table 3-4.</b> Average coverage per base.....	54
<b>Table 3-5.</b> Variants validated by targeted sequencing.....	55
<b>Table 3-6.</b> Major driver gene mutations identified in each patient.....	71
<b>Table 3-7.</b> Relatedness indices of metastases.....	77
<b>Table 3-8.</b> Genetic distances among metastases.....	79



## Figures

<b>Figure 1-1.</b> The genetic progression model of pancreatic cancer.....	9
<b>Figure 1-2.</b> The subclonal evolution and timing of pancreatic cancer.....	13
<b>Figure 2-1.</b> Approach for PanIN evolution.....	17
<b>Figure 2-2.</b> Laser capture microdissection of pancreatic tissue.....	21
<b>Figure 2-3.</b> PIN101 evolutionary relationship.....	32
<b>Figure 2-4.</b> PIN108 evolutionary relationship.....	33
<b>Figure 3-1.</b> Approach for pancreatic cancer metastasis evolution.....	39
<b>Figure 3-2.</b> Metastatic disease of four pancreatic cancer patients.....	51
<b>Figure 3-3.</b> Hierarchical clustering for Pam01.....	73
<b>Figure 3-4.</b> Hierarchical clustering for Pam02.....	74
<b>Figure 3-5.</b> Hierarchical clustering for Pam03.....	75
<b>Figure 3-6.</b> Hierarchical clustering for Pam04.....	76
<b>Figure 3-7.</b> Somatic evolution of normal tissues.....	80
<b>Figure 3-8.</b> Neighbor-joining phylogeny of primary tumor sections and metastases for patient Pam02.....	83
<b>Figure 3-9.</b> Neighbor-joining phylogeny of primary tumor sections and metastases for patient Pam03.....	85
<b>Figure 3-10.</b> Relatedness and three dimensional geography of Pam02 primary tumor sections and metastases.....	89
<b>Figure 3-11.</b> Relatedness and three dimensional geography of Pam03 primary tumor sections and metastases.....	90
<b>Figure 3-12.</b> Relatedness and three dimensional geography of Pam04 primary tumor sections and metastases.....	91
<b>Figure 4-1.</b> PIN103 evolutionary relationship.....	96
<b>Figure 4-2.</b> PIN104 evolutionary relationship.....	97
<b>Figure 4-3.</b> PIN105 evolutionary relationship.....	98
<b>Figure 4-4.</b> PIN106 evolutionary relationship.....	99
<b>Figure 4-5.</b> PIN107 evolutionary relationship.....	100
<b>Figure 4-6.</b> MinDel phylogeny for Pam01.....	106
<b>Figure 4-7.</b> MinDel phylogeny for Pam02.....	107
<b>Figure 4-8.</b> MinDel phylogeny for Pam03.....	109
<b>Figure 4-9.</b> MinDel phylogeny for Pam04.....	110

## Chapter 1 – Pancreatic cancer evolves

---

### Challenges

Solid tumors progress in sequential stages over time – this observation was first supported by epidemiologic data of age-incidence rates for several tumor types<sup>1,2</sup>. Subsequent molecular and genetic studies demonstrated that several distinct mutations accumulate during cancer formation; indeed, these sequential, genetic mutations underpin the multiple stages implied by the epidemiological data<sup>3,4</sup>. These genetic mutations may be inherited or acquired during the lifetime of the individual, the latter caused by environmental mutagens or simply a mistake during cell division in renewing tissues<sup>5-7</sup>. Despite this genetic understanding of cancer origins, the individual tumor progresses by chance, history, and evolution contributing to multiple disease stages over time. Additionally, “cancer” represents a set of tumors which have a staggering variety of particular causes, phenotypes, and clinical behaviors<sup>5,8</sup>. To sum up, a pancreatic tumor differs greatly from a lung tumor; worse, even two pancreatic tumors differ by their unique evolutionary histories.

These differences, however, boil down to a more limited set of general features, specifically perturbed cellular pathways and acquired tumor hallmarks<sup>6,9,10</sup>. Further, evolutionary analysis sets these mutations and phenotypes in a unifying framework<sup>11</sup>: for example, pancreatic cancer fits an evolutionary paradigm very well driven by a series of clonal expansions. Briefly, a pancreatic cell acquires a mutation (or epigenetic alteration) which stably confers a selective advantage allowing the cell to divide and produce a population<sup>6,11,12</sup>. The result of this process, repeated over the life of the patient, is a

clinically detectable pancreatic tumor – these principles may also explain metastasis and resistance to treatment. Thus, understanding how tumors evolve is highly important.

In 2015, pancreatic cancer remains the fourth leading cause of cancer deaths for men and women in the United States<sup>13</sup>. Nearly 50,000 new cases are expected to be diagnosed this year with just over 40,000 deaths – even worse, mortality rates are expected to rise for both genders<sup>13</sup>. The five year survival rate is estimated between 6-7%, yet this rate drops to 2% for metastatic disease<sup>13</sup>.

Pancreatic ductal adenocarcinoma represents the most common and lethal form of pancreatic cancer (the work described here will focus on this type). Two clinical characteristics contribute to the striking lethality of this pancreatic cancer subtype: covert disease and limited treatments. Pancreatic cancer is largely asymptomatic, typically undetected until a very late stage of disease – far past the point of surgical intervention with curative potential<sup>14,15</sup>. Additionally, there are few widely-implemented screening tools for early detection<sup>16</sup>. The most common screening method, the pancreatic tumor marker CA19-9, works less efficiently with smaller tumors: those which could be cured by early intervention<sup>17</sup>. For these reasons, unveiling covert disease is an important endeavor in addition to identifying individuals with elevated risk<sup>14,18</sup>.

For treatment, there are very few chemotherapeutic options available for late-stage pancreatic cancer patients. The standard of care is gemcitabine, a chemotherapeutic with modest effect on survival when administered alone<sup>19,20</sup>. Even in combination with other therapies, survival improves but does not increase more than a few weeks or months (e.g. with erlotinib or FOLFIRINOX)<sup>21,22</sup>. The modest effects of these treatments may be due to biological barriers within the tumor, specifically the stroma<sup>23</sup>. However,

stroma incompletely explains treatment difficulties because metastases often lack stroma, yet demonstrate therapeutic resistance<sup>24</sup>. Another potential approach is the use of synthetic lethal screening to target those cancer cells with particular mutations. However, *KRAS*, the most common driver, has been difficult to target in the past; generally, these measures require greater effectiveness in pancreatic cancer<sup>25–27</sup>.

In summary, pancreatic cancer remains a dire diagnosis – we are in need of improved screening methods and more effective treatments for advanced disease. A better understanding of pancreatic cancer genetics and evolution, especially in the context of metastasis, will be essential for these efforts.

## **Genetics**

As in other solid tumors, such as colorectal cancers<sup>6</sup>, mutations underlie pancreatic cancer formation – genetic sequencing studies continue to reveal which mutations drive the evolution of these tumors. Generally, these studies rely on one of the following levels of analysis: among patient populations or within individual patients. These analyses overlap (both seeking to reveal mutations or patterns), yet each emphasizes different aspects of the mutation data and sample context. The first compares pancreatic cancers across many patients; the second analyzes many tumors within an individual. I will briefly summarize what we have learned from each.

Several large-scale, human sequencing studies have focused on pancreatic cancers among many patients<sup>28–30</sup>. These studies confirm the presence of several genetic drivers which were previously detected by molecular analyses. Drivers are those genetic mutations which occur at a high frequency and confer a selective advantage to the tumor

cell<sup>6</sup>: for pancreatic cancer, these four mutations are in *KRAS*, *CDKN2A*, *TP53*, and *SMAD4* (see Table 1-1). There are other, less-frequently observed mutations which may drive the formation of pancreatic cancer<sup>28-31</sup> as well as those variants which associate with a family history of pancreatic cancer and occur in the germline of these patients<sup>32</sup> – these cases account for ~10% of pancreatic cancer diagnoses. My thesis focuses on the remaining ~90% of cases related to sporadic pancreatic cancer.

In summary, these sequencing studies have confirmed the high prevalence of known drivers, implied novel drivers, and revealed the genomic landscape of pancreatic cancer. One study by Jones *et al.* sequenced the exomes of 17 primary tumors and 7 metastases (for a total of 24 patients), which confirmed the main drivers as well as novel mutations in chromatin-remodeling genes, a previously unrecognized feature of these tumors<sup>28</sup>. More recently, Biankin *et al.* sequenced the exomes of 99 individual pancreatic tumors, confirming previously discovered alterations and a new set of mutated genes: axon guidance, usually expressed during embryogenesis<sup>29</sup>. Additionally, Waddell *et al.* performed whole genome sequencing on 100 tumors to reveal novel drivers and the extent of genomic rearrangements, categorizing pancreatic cancers into four groups based on these changes<sup>30</sup>.

These studies have provided the comprehensive catalogue of genetic changes which occur in pancreatic cancer. Yet, tallying up the mutations detected in tumors will not reveal underlying biological principles or meaningful solutions for patients. This requires evolutionary analyses, the topic of the next section.

**Table 1-1. Common genetic drivers in pancreatic cancer.**

Gene	Full name	Prevalence in PDAC <sup>a</sup>	Mutation	Role in cancer	Cellular function; pathways <sup>b</sup>	Other cancers	References <sup>c</sup>
<i>KRAS</i>	v-Ki- ras2 Kirsten rat sarcoma viral oncogene homolog	90-95%	Point mutation at codons 12, 13, or 61	Activating oncogene	GTPase; Ras signaling, cell division, survival	Bladder, breast, leukemia, lung	33–35
<i>CDKN2A</i>	cyclin-dependent kinase inhibitor 2A gene	90%	Point mutation with allelic loss, homozygous deletion, or hypermethylation	Inactivating tumor suppressor	Kinase inhibitor; cell division	Melanoma	35–38
<i>TP53</i>	tumor protein p53	75%	Point mutation or small intragenic deletion	Inactivating <sup>d</sup> tumor suppressor	DNA expression; cell survival, division, DNA repair	Many	39–42
<i>SMAD4</i>	SMAD family member 4 gene	55%	Point mutation and allelic loss or homozygous deletion	Inactivating tumor suppressor	DNA expression; TGF-β <sup>e</sup> and BMP <sup>f</sup> signaling, cell division	Colorectal	40,43–46

<sup>a</sup>pancreatic ductal adenocarcinoma

<sup>b</sup>only main or well-documented cellular functions and pathways are listed here; these proteins likely participate in many others

<sup>c</sup>for the original discovery of mutations in each driver; large-scale sequencing studies also confirm their presence<sup>28–30</sup>

<sup>d</sup>mutations lead to recessive inactivation of the protein but result in a dominant negative cellular phenotype

<sup>e</sup>transforming growth factor-beta

<sup>f</sup>bone morphogenetic protein

## **Evolution**

Charles Darwin put forward a theory of evolution proposing that natural selection, a creative process, explained how small variations between organisms resulted in new species<sup>47,48</sup>. Decades later, Gregor Mendel's work (at least, its twentieth century rediscovery) catalyzed a synthesis of genetics and evolutionary theory, thereby providing a mechanism which linked individual variation and population evolution<sup>49-52</sup>. Since this foundation in genetics, evolutionary theory has extended beyond species to explain various aspects of medicine, including infectious diseases, antibiotic resistance, and the evolution of cancer. The latter culminated in the 1970s when discussions regarding mutations, cancer, and evolution began to appear in the literature of widely read journals<sup>12,53</sup>. Specifically, Peter Nowell described cancer in an evolutionary sense by hypothesizing that tumor cell variants are selected in a step-wise manner as a tumor develops<sup>12</sup>.

Our contemporary view of cancer is essentially genetic, underpinned by genetic mutations which occur in stages as the cancer evolves<sup>3,6,8</sup>. Further, these mutations are the source of genetic variation – which confer hallmark phenotypes for natural selection to act upon<sup>9,11,54</sup>. This Darwinian process explains how a cell gains a selective advantage and forms a tumor which can become more aggressive<sup>55</sup>. The current challenge is to translate these findings to better understand cancer biology and more effectively treat patients.

As stated previously, pancreatic cancer can be understood in an evolutionary sense, from the initial lesions to advanced stage disease<sup>56</sup>. For example, the precursor lesions in pancreatic cancer are pancreatic intraepithelial neoplasias, abbreviated

PanINs<sup>57,58</sup>. These early lesions develop in sequential stages which may lead to invasion and advanced disease. Histologically, PanINs occur microscopically in the smaller pancreatic ducts, appearing as papillary or flat groups of cells<sup>57,59</sup>. Genetically, the four drivers of pancreatic cancer overlay the different stages of PanIN development (Figure 1-1). Typically, mutations in *KRAS* and *CDKN2A* occur early while *TP53* and *SMAD4* mutations occur in later-stage PanIN lesions<sup>35,40,44</sup>. Other genetic changes include telomere abnormalities and shortening (which may be one of the earliest events in low-grade PanIN lesions<sup>60</sup>) and epigenetic regulation, including aberrant gene and miRNA expression<sup>61-64</sup>. Thus, understanding the timing and order of mutations in pancreatic cancer provides insights into its evolution, especially the drivers which occur early in the cellular precursors of the patient's pancreas. However, the genetics which illustrate the evolution of a late-stage PanIN to a pancreatic cancer, and beyond, remain unpredictable (Figure 1-1).

Beyond the precursor lesions, the previously cited sequencing studies provide key findings for primary tumor evolution. In Jones *et al.*, despite the particular mutations across all 24 tumors, the genetic complexity reduced to a dozen core pathways involved in pancreatic cancer, mostly those affecting cell survival, division, and DNA damage<sup>28</sup>. Yet, the specific proteins which were affected in any given individual varied across a wide spectrum<sup>28</sup>. Clinically, this genetic “heterogeneity” is significant because therapies which target the specific proteins would yield success in only a minority of patients (whose tumors harbor those particular alterations). However, therapies which are designed to instead target pathways will be more broadly applicable – generally, this pathway approach can simplify the complexity of other solid tumors<sup>6,10</sup>.



In addition to the therapeutic implications, genetic mutations provide a record of the natural history – as evolutionary markers, mutations are essential for understanding how a pancreatic cancer developed over time. There are several advantages to use sequencing to ask these evolutionary questions. For instance, next generation sequencing is a more sensitive approach to detecting mutations, useful for tissue with normal and cancer cells mixed or potentially several cancer subclones present. Second, sequencing at the whole genome level can detect large-scale genomic alterations as well as point mutations, allowing for complementary analysis across the genome. The large amount of data which these studies generate is essential for the “global” view of the genome and a long look backward into the tumor’s history<sup>65-67</sup>.

These large sequencing studies reveal much needed information regarding mutations and the evolution of the pancreatic cancer genome. Yet, these studies typically use one sample per patient as an end-point – not ideal for evolutionary questions concerning mutation order, timing, and disease progression. These endeavors require the genetic analysis of multiple tumors within a patient: the topic of the next section.

**Lesion: Normal → PanIN1 → PanIN2 → PanIN3 → Cancer → Metastasis**

*Mutation:*                    *KRAS*.....*CDKN2*.....*TP53*...*SMAD4*.....?.....?

Histo:     Ductal        Columnar        Atypia        Pseudopapillary        Tumor

**Figure 1-1. The genetic progression model of pancreatic cancer.** PanINs are thought to develop from the normal ductal epithelium of the pancreas. Telomere shortening, most likely one of the first events in PanIN development, is not included (but would likely precede *KRAS*). As a PanIN lesion matures, the cells become tall and columnar, eventually gaining atypical nuclei, and finally forming pseudopapillary structures. The four driver genes propel this process forward; other genes which may affect PanIN formation are not included here. Additionally, any drivers which specifically lead from PanIN3 to cancer and metastasis have yet to be characterized.

## Metastasis

Metastasis occurs when tumor cells disseminate and colonize a foreign organ to form a new tumor – a key feature of aggressive disease<sup>68,69</sup>. This phase represents the lethal stage of most solid tumors (except for those in the brain). Besides this general biology, the underlying mechanisms which drive this process are relatively unknown. Nevertheless, some aspects relevant to metastasis have been identified, including the epithelial to mesenchymal transition<sup>70</sup>, angiogenesis<sup>71</sup>, and interactions with the stroma and tumor microenvironment<sup>9,72</sup>; each of these may contribute, but are insufficient for metastasis. Additionally, mutations which contribute directly to metastasis are unknown, yet gene expression of the primary tumor can predict metastatic behavior; still, we lack genetic drivers of metastasis<sup>73</sup>. Another theory suggests that the mutations which determine metastatic propensity are acquired early in tumor evolution (possibly the drivers themselves)<sup>74</sup> – or metastases occur simply as the result of circulating tumor cells, requiring no additional genetic alterations to grow, which happen to survive and travel to a hospitable organ<sup>6</sup>. Finally, it is possible that metastatic tumors represent a spectrum of these patterns; likewise, metastasis represents a step-wise process in which chance and biology most likely influence certain stages. Much work, including mathematical modeling<sup>75,76</sup> and tumor phylogenetics<sup>67,77</sup>, is needed to discover the principles which underlie metastasis.

In some ways, pancreatic cancer is ideal for the study of metastasis. As previously described, the genetic and molecular underpinnings of pancreatic cancer formation are well known<sup>14</sup>. These changes overlay onto a progression model for which we know the precursor lesions and their progression<sup>57,59,78</sup>. Additionally, most patients

have metastatic disease during the final stages of illness and death<sup>79</sup>. Many of these patients also harbor the primary tumor – this is key for the evolutionary analysis of matched primary tumor and metastases<sup>15,80</sup>.

Previous work with the JHMI Rapid Autopsy Program revealed that not all pancreatic cancer patients die with metastatic disease: even those which were metastatic represented a spectrum of number of tumors in one patient, displaying differing patterns of spread<sup>81</sup>. Interestingly, these differences were not correlated with clinical parameters, but rather the status of *SMAD4*. Specifically, the loss of protein expression correlated with the more widespread pattern of metastases<sup>82</sup>. This demonstrates that early-acquired drivers may (directly or indirectly) contribute to late-stage metastasis.

Other previous work began with the whole exome sequencing data from seven metastases from Jones *et al.*<sup>28</sup>. In Yachida *et al.*, the whole exome data were used to screen the additional samples of each patient for the presence or absence of each previously discovered mutation, ultimately revealing mutation order, evolutionary relationships, and timing of various stages of disease progression<sup>83</sup>. In this study, there were several key findings from which my thesis stems.

First, the targeted sequencing revealed that mutations could be categorized into two groups: Founders, present in all tumors in a person (metastases and primary tumor sections), and Progressors present in a subset or only one tumor<sup>83</sup>. Founder mutations were likely to be gained early in the disease, most during PanIN progression. Evolutionarily, Founders represented the genotype of the most recent common ancestor of all tumors within a patient – Progressors were gained later and coincided with tumor divergence. Interestingly, a co-published study analyzing the genomic rearrangements in

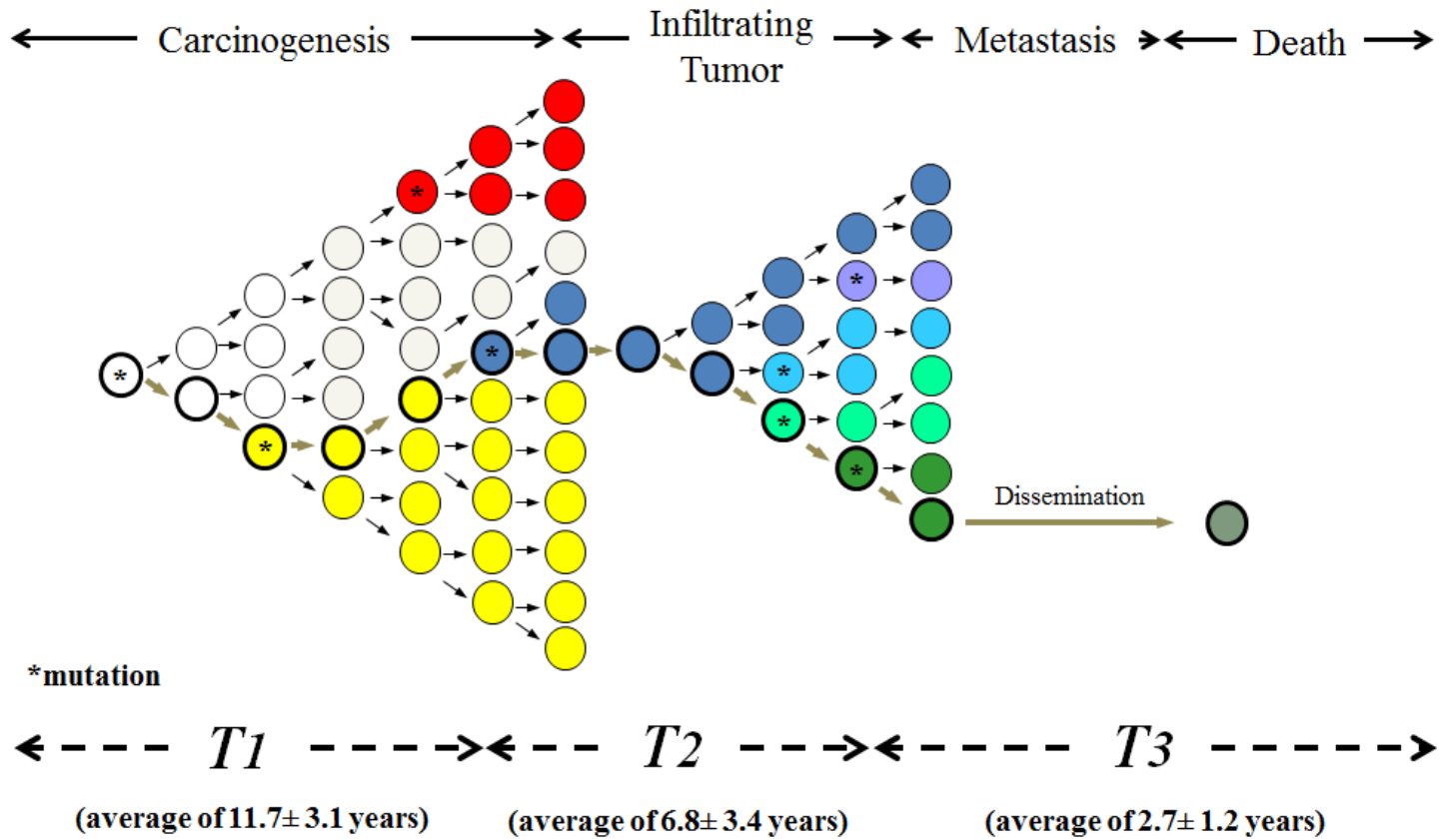
13 pancreatic cancers found the same categories of mutations: the genomic rearrangements were Founders or Progressors<sup>84</sup>.

These studies suggested that point mutations and genomic rearrangements occur during the formation of the parental clone and divergence of the subclonal lineages<sup>83,84</sup>. Interestingly, genetic subclones were also shown to occur in geographically distinct parts throughout the original primary tumor<sup>83</sup>. Additionally, Yachida *et al.* demonstrated through modeling that pancreatic cancer progression occurs over a long period overall (~20 years), with metastasis occurring late and over a short time period of ~2 years (Figure 1-2)<sup>83</sup>.

There are two clinically-relevant conclusions that can be made from these studies of pancreatic cancer evolution and metastasis<sup>83,84</sup>. First, since pancreatic cancer develops over a long time period, there exists an important window for screening and early detection which could lead to curative interventions<sup>83</sup>. Second, targeted therapies based on alterations which are actually Progressors will inevitably lead to tumor recurrence; ideally, genetic therapies would target the parental clone (i.e. every tumor cell).

These conclusions are especially important given the findings from a computational model of pancreatic cancer showing that very little time exists between clinically detectable disease and metastasis<sup>85</sup>. Additionally, the genetic features of a pancreatic cancer could be used for early detection, as indicated by a follow-up study which showed that the particular mutations of the drivers can predict (and perhaps influence) metastatic behavior of the primary tumor<sup>86</sup>.

**Figure 1-2. The subclonal evolution and timing of pancreatic cancer.** As the pancreatic tumor cells divide, mutations (marked by an asterisk) are acquired in the lineage over time. These genetic mutations identify genetic subclones (represented with different colors) within the primary tumor. The timing of this evolution was modeled using a combination of comparative lesion sequencing and mathematics<sup>83</sup>. As indicated by the figure, pancreatic carcinogenesis and progression take many years to occur, with metastasis occurring toward the final stages of this evolution.



To conclude, I will summarize what we know thus far regarding how pancreatic cancer evolves and metastasizes. First, metastatic cancer cells disseminate from the primary tumor – interestingly, this process is most likely occurring through pancreatic cancer development (perhaps even before tumor formation<sup>87</sup>), yet the genetic data imply that successful colonization and metastasis occurs much later<sup>83</sup>. Second, subclonal evolution occurs in pancreatic cancer: the genetics of the metastases reflect these genetic subclones in the primary tumor<sup>83,84</sup>. Third, the particular driver mutations may indicate the metastatic propensity of the primary tumor<sup>86</sup>. Fourth, metastasis is the final stage of a protracted evolution, yet metastases form relatively quickly<sup>83</sup>.

These conclusions lead to several questions worth exploring. First, when does subclonal evolution arise? Second, how many metastatic subclones exist in the primary tumor, and what are their molecular phenotypes? Third, if subclonal mutations contribute to metastasis, which ones? Fourth, how do multiple metastases within an organ or between organs evolve? Do these relationships imply self-seeding (the exchange of tumor cells between metastases and the primary tumor<sup>88</sup>)?

The following chapters aim to address some of these questions. While these issues (and others) remain provocative<sup>24</sup>, genetic analysis and evolutionary theory will continue to shed much needed light on pancreatic cancer, the deadliest of tumors.



## **Chapter 2 – Evolution of pancreatic cancer precursors**

---

### **Synopsis**

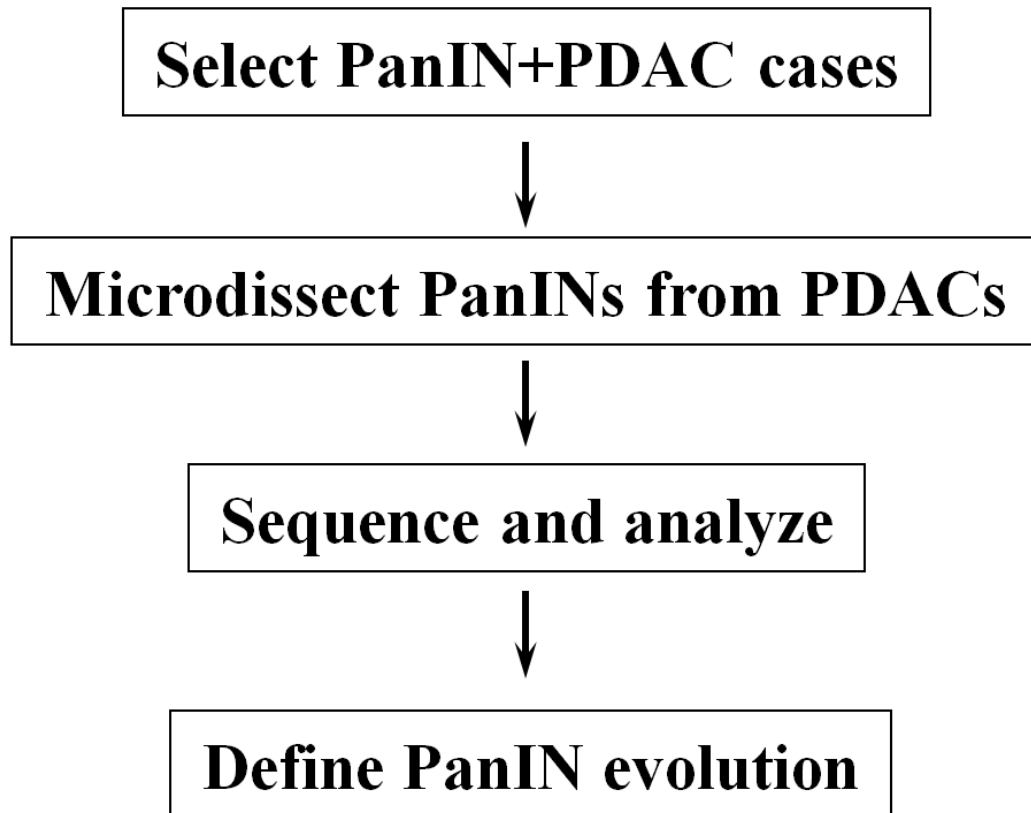
**PROJECT TITLE:** Whole-exome sequencing reveals the genetic evolution of pancreatic intraepithelial neoplasia and pancreatic adenocarcinoma.

**AUTHORS:** Alvin P. Makohon-Moore\*, Karen Matsukuma\*, Yuchen Jiao, Ming Zhang, Ralph H. Hruban, Nickolas Papadopoulos, Kenneth W. Kinzler, Bert Vogelstein, and Christine A. Iacobuzio-Donahue

\*These authors contributed equally to this study

**HYPOTHESIS:** Subclonal evolution occurs early in PDAC and drives the evolution of PanIN to invading carcinoma.

**AIM:** Define the degree of subclonality and phylogenetic relatedness of distinct PanIN lesions and the matching primary carcinoma.



**Figure 2-1. Approach for PanIN evolution.** Each case was selected for the presence of PanIN(s) and matched pancreatic adenocarcinoma, ideally with maximal intervening stroma. The PanINs were microdissected as well as parts of the tumor – patient-matched normal was included for each of the cases. All available PanINs and tumor samples were DNA extracted and underwent whole exome sequencing. The mutation data were filtered for higher quality variants and visualized directly to validate presence. Eventually, all the mutations will be incorporated into an evolutionary analysis.

## Introduction

Pancreatic ductal adenocarcinoma (PDAC) derives from the clonal evolution of pancreatic intraepithelial neoplasia, or PanIN. In the PanIN model of carcinogenesis, a normal pancreatic cell transforms to a low grade PanIN-1 which develops into a higher grade PanIN-2/PanIN-3<sup>59,78</sup>. As a precursor lesion, the PanIN grows and exhibits increasing cellular and ductal atypia over time<sup>14</sup>. This transformation is correlated with the stepwise accumulation of genetic alterations, including *KRAS*, *CDKN2A*, *TP53*, and *SMAD4*<sup>35,41,43</sup>. Eventually, the PanIN lesion evolves into PDAC, and the incipient tumor will grow and possibly metastasize: crossing this threshold leads to the death of the patient<sup>78,79</sup>.

Even though the PanIN model has greatly improved our understanding of PDAC carcinogenesis, many of its predictions require further exploration. For instance, a point mutation in *KRAS* is predicted to occur early in PanIN development before other driver mutations<sup>33,35</sup>. However, this prediction is not a strict prerequisite for a PanIN-1 since other driver events could initiate PanIN formation – in fact, telomere shortening has been observed in a high frequency of early stage PanINs<sup>60</sup>. In the vast majority of colorectal cancers, an alteration in the tumor suppressor *APC* occurs well before a point mutation in *RAS*<sup>3</sup>; in PanIN evolution, whether a tumor suppressor is initially mutated is unknown.

As PanIN development unfolds, it is likely affected by selection pressures and physical growth restrictions, especially within the PanIN itself and the local microenvironment. Yet, little is known regarding the genetic mutations that allow the precursor cells to escape bottlenecks or negative selection. This knowledge is critical for

understanding the transition of a PanIN-3 to PDAC, a progression that has yet to be genetically defined.

Finally, although it is typical for several PanINs to coexist in a single PDAC patient, the significance of this is unrealized<sup>14</sup>. Although a previous study showed that lower grade PanINs harbor around half of the mutations in the matched tumor<sup>89</sup>, how multiple PanINs are evolutionarily related to each other and the concurrent PDAC has not been explored in depth. Specifically, this question requires genetic analysis, combined with evolutionary theory and modeling, to understand PanIN biology.

To address these issues, we applied a whole-exome sequencing (WES) method to DNA extracted from multiple, geographically distinct PanINs and matched tumors from eight patients (Figure 2-1, Table 2-1). The PanINs among the cases ranged from PanIN2 to PanIN3 while the tumors ranged from poor to moderate differentiation as assessed by histology. For five of the cases, we acquired two distinct PanIN lesions – during preparation, each lesion (PanIN or cancer) was collected from either a distinct slide or with maximum intervening stroma.

For detailed techniques, please see the section on Methods. Briefly, each specimen was cut at 5  $\mu$ M slice thickness to visualize for histology and laser capture microdissection (LCM). Figure 2-2 demonstrates the accuracy of LCM for lesions of this size. Once all lesions were microdissected, DNA was extracted and quantified using standard protocols for LCM tissue. DNA libraries were prepared for whole exome sequencing from each sample, including the matched normal.

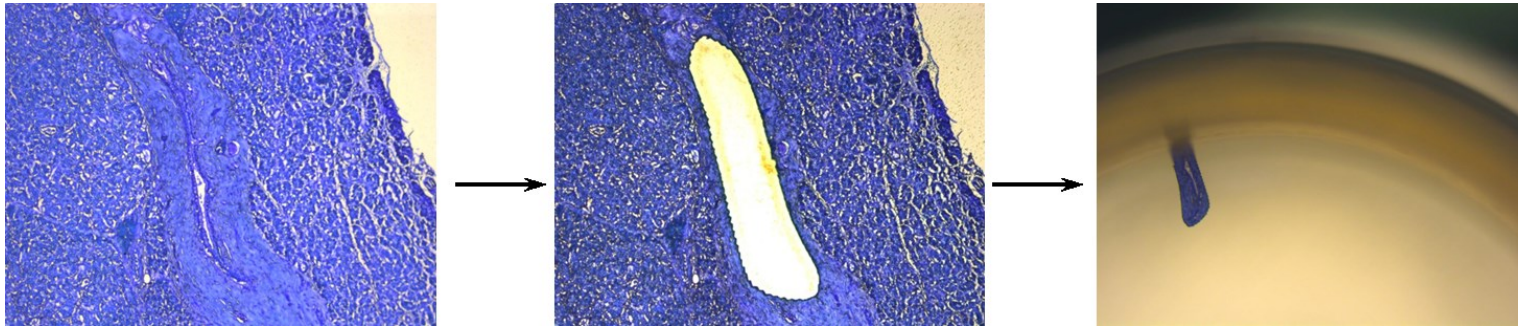
**Table 2-1. Cases selected for PanIN and cancer whole exome sequencing.**

<i>Pathology</i>	<b>PIN101</b>	<b>PIN102</b>	<b>PIN103</b>	<b>PIN104</b>	<b>PIN105</b>	<b>PIN106<sup>a</sup></b>	<b>PIN107</b>	<b>PIN108</b>
<i>PanIN A</i>	PanIN3	PanIN3	PanIN3	PanIN3	PanIN2/3	PanIN3	PanIN3	PanIN3
<i>PanIN B</i>	PanIN2	PanIN2	n/a	PanIN2	n/a	PanIN3	n/a	PanIN2
<i>PT Histo<sup>b</sup></i>	Poor	Mod.	Mod.	Poor	Mod.	Poor	Mod.	Mod.

Mod.: moderate

<sup>a</sup>PIN106 has a third PanIN for analysis

<sup>b</sup>primary tumor histology, differentiation



**Figure 2-2. Laser capture microdissection of pancreatic tissue.** This duct was selected to form a pre-determined field for the laser to cut (not shown). Once cut, the 5  $\mu\text{M}$  thick sample falls into the adhesive cap below – the microdissected duct is pictured in the third panel on the adhesive cap.

## Results

Upon completion of sequencing, the data were analyzed *in silico* and aligned to the hg18 genome. In our data set, >1100 potential somatic mutations were detected at an average high quality coverage of 204X and 90.7% of targeted bases with at least 10 reads. After filtering for high quality mutations which also passed direct visualization (see Methods), we detected 5-52 mutations among the PanINs and 23-53 among the cancers (Table 2-2). One case (PIN102) was of too poor quality for further analysis.

These mutations included point mutations in the common drivers of pancreatic cancer: *CDKN2A*, *KRAS*, and *TP53*. For each case, we detected 1-3 of these driver alterations. Interestingly, *KRAS* mutations were not detected in some PanINs despite the >10X distinct coverage for that site and the presence of other somatic mutations. In addition, TGF $\beta$  receptor genes were mutated in three of these cases – the significance of this (and in the absence of *SMAD4* mutations) is unclear.

Eventually, we will leverage these data using a computational model to understand how divergent the lesions are from each other and to what extent they may have mixed or exchanged cells. However, some evolutionary relationships can be inferred from the presence or absence of the point mutations<sup>65,83</sup>. For example, the cancer and the PanINs in PIN101 share a common ancestor which had already acquired at least 5 mutations before the lesions diverged. The natural history of PIN108 is slightly more complicated: the common ancestor had acquired 22 mutations before the lesions diverged, yet the cancer acquired 13 unique mutations while PanINB acquired 14 unique mutations. This implies that these two subclones, while clonally related, diverged during their evolutionary history – one seeded the cancer, the other remained as a PanIN.

**Table 2-2. Genetic drivers and mutation numbers.**

Case	Number of mutations in PanIN A <sup>a</sup>	Number of mutations in PanIN B <sup>a</sup>	Number of mutations in cancer <sup>a</sup>	Total number of mutations <sup>a</sup>	Median distinct coverage <sup>b</sup>	Drivers <sup>c</sup>
PIN101	5, 14	14	23	25	115X	<i>CDKN2A, KRAS, TP53</i>
PIN103	15	15	29	30	87X	<i>KRAS, TP53</i>
PIN104	26, 29	29	53	53	179X	<i>KRAS, TP53</i>
PIN105	40	n/a	46	79	161X	<i>KRAS</i>
PIN106	49, 52	52	51	74	77X	<i>CDKN2A, KRAS, TP53</i>
PIN107	19	n/a	29	30	171X	<i>CDKN2A, KRAS, TP53</i>
PIN108	23, 39	39	38	53	161X	<i>CDKN2A, KRAS, TP53</i>

<sup>a</sup>after filtering

<sup>b</sup>average distinct coverage of all PanIN and cancer samples from a case, normal included

<sup>c</sup>only included previously known, high-frequency drivers



**Table 2-3. Mutations detected via WES.**

<b>Case</b>	<b>Chr.</b>	<b>Position</b>	<b>BaseFrom</b>	<b>BaseTo</b>	<b>Gene</b>	<b>% Mut</b>
PIN101	chr11	69709387	G	A	ANO1	11.3%
PIN101	chr1	54378055	C	T	CDCP2	22.5%
PIN101	chr9	21964765	G	T	CDKN2A	46.7%
PIN101	chr17	53306031	G	A	CUEDC1	16.9%
PIN101	chr20	34493947	G	A	DLGAP4	19.7%
PIN101	chr6	94124707	C	T	EPHA7	22.6%
PIN101	chr11	113958647	C	T	FAM55D	19.3%
PIN101	chr14	99188468	C	T	HHIPL1	17.1%
PIN101	chr8	83366586	C	A	HNRNPA1P4	32.6%
PIN101	chr14	105392182	G	A	IGHM	21.1%
PIN101	chr12	25289551	C	T	KRAS	31.0%
PIN101	chrX	48317951	G	A	LOC729275	20.8%
PIN101	chr1	160593429	C	T	NOS1AP	20.5%
PIN101	chr11	54818399	A	C	NULL	4.4%
PIN101	chr9	130157870	G	A	SLC27A4	14.5%
PIN101	chrX	142623103	C	T	SPANXN2	26.9%
PIN101	chr12	113288311	G	T	TBX5	15.9%
PIN101	chr6	54299620	C	T	TINAG	20.1%
PIN101	chr19	2940723	T	C	TLE6	32.5%
PIN101	chr17	7517845	C	T	TP53	34.1%
PIN101	chr4	6355046	G	A	WFS1	16.5%
PIN101	chr3	148611193	G	A	ZIC1	23.8%
PIN103	chr11	66087106	C	T	actn3	36.0%
PIN103	chr8	124407378	TC		ATAD2	24.2%
PIN103	chr16	28816868	G	A	ATP2A1	19.2%
PIN103	chr3	10357251	C	T	ATP2B2	57.9%
PIN103	chr9	125173041	G	T	CRB2	38.6%
PIN103	chr16	83464177	G	A	CRISPLD2	22.5%
PIN103	chr11	9148067	G	A	DENND5A	30.2%
PIN103	chr8	1484910	G	T	DLGAP2	11.4%
PIN103	chr17	35133180	C	T	ERBB2	29.6%
PIN103	chr15	46514194	C	A	FBN1	24.0%
PIN103	chr4	79681254	T	A	FRAS1	35.0%
PIN103	chr4	55655860	C	T	KDR	33.3%
PIN103	chr17	20300560	C	A	LGALS9B	30.0%
PIN103	chr6	119711647	C	T	MAN1A1	20.9%

PIN103	chr17	17963432	C	T	MYO15A	75.3%
PIN103	chr20	62309616	A	T	MYT1	11.2%
PIN103	chr8	145185916	C	T	OPLAH	36.0%
PIN103	chr11	5024789	G	A	OR52J3	38.9%
PIN103	chr8	145068443	C	G	PLEC	42.0%
PIN103	chr4	57555718	C	A	POLR2B	22.9%
PIN103	chr10	120344161	C	T	PRLHR	53.2%
PIN103	chr10	104476528	A	G	SFXN2	42.6%
PIN103	chr10	72306332	G	A	SGPL1	38.9%
PIN103	chr12	113596426	G	A	TBX3	34.1%
PIN103	chr17	7518988	G	A	TP53	65.6%
PIN103	chr7	98344463	G	A	TRRAP	32.7%
PIN103	chr2	97707231	G	T	ZAP70	33.3%
PIN103	chr19	11804209	A	T	ZNF440	28.2%
PIN104	chr1	95221336	C	T	ALG14	33.1%
PIN104	chr9	33251116	T	G	BAG1	31.1%
PIN104	chr15	22475177	A	G	C15orf2	10.8%
PIN104	chr3	48241858	A	T	CAMP	34.5%
PIN104	chr10	119956970	T	A	CASC2	34.3%
PIN104	chr4	24447163	T	C	CCDC149	30.7%
PIN104	chr10	96525161	G	C	CYP2C19	26.1%
PIN104	chr4	155506955	G	A	DCHS2	32.3%
PIN104	chr19	14738085	G	A	EMR2	36.6%
PIN104	chr4	126460195	G	A	FAT4	34.5%
PIN104	chr7	5497400	G	A	FBXL18	16.2%
PIN104	chrX	12648702	G	T	FRMPD4	41.6%
PIN104	chr9	129167406	T	G	GARNL3	33.3%
PIN104	chr12	121752804	G	A	GPR109B	14.1%
PIN104	chr12	121766258	G	A	GPR109B	21.1%
PIN104	chr2	196826925	G	A	HECW2	35.6%
PIN104	chr16	1688969	C	G	HN1L	39.4%
PIN104	chr7	27161268	G	A	HOXA7	35.9%
PIN104	chr12	52680738	G	A	HOXC9	34.6%
PIN104	chr3	185237343	G	A	HTR3D	17.8%
PIN104	chr2	89830685	G	A	IGKV1D-13	42.9%
PIN104	chr3	49037217	C	G	IMPDH2	22.3%
PIN104	chr1	144253612	AG		ITGA10	21.8%
PIN104	chr1	1908294	C	T	KIAA1751	25.7%
PIN104	chr17	30345549	G	A	LIG3	23.8%
PIN104	chr7	149665218	C	T	LRRC61	30.8%
PIN104	chr3	49903849	C	T	MST1R	30.2%

PIN104	chr17	10383312	C	T	MYH2	39.3%
PIN104	chr15	18346543	A	G	NULL	15.2%
PIN104	chr5	140731073	G	A	PCDHGB3	25.6%
PIN104	chr2	219209482	C	T	PLCD4	22.3%
PIN104	chr11	45161153	G	A	PRDM11	33.9%
PIN104	chr3	49044653	G	A	QRICH1	32.0%
PIN104	chr1	76032465	G	T	RABGGTB	17.9%
PIN104	chr7	103416847	G	A	RELN	27.0%
PIN104	chr19	44053105	G	C	RINL	43.4%
PIN104	chr2	100988912	C	T	RPL31	36.9%
PIN104	chr11	117542993	G	C	SCN2B	52.5%
PIN104	chr7	54790976	T	C	SEC61G	43.8%
PIN104	chr10	119005138	C	T	SLC18A2	33.8%
PIN104	chr9	130152648	G	C	SLC27A4	20.0%
PIN104	chr19	10997143	C	T	SMARCA4	33.0%
PIN104	chr11	66209961	C	T	SPTBN2	39.1%
PIN104	chr1	114483829	G	A	SYT6	42.7%
PIN104	chr8	133968280	C	G	TG	28.8%
PIN104	chr5	79391351	G	A	THBS4	36.5%
PIN104	chr17	7518947	TC		TP53	59.6%
PIN104	chr11	5582415	G	A	TRIM6	28.9%
PIN104	chr8	35663667	C	T	UNC5D	31.6%
PIN104	chr8	7872844	G	C	USP17L3	15.0%
PIN104	chr4	177950344	G	A	VEGFC	30.8%
PIN104	chr19	23108740	G	C	ZNF730	23.5%
PIN105	chr1	12627281	G	C	AADAACL4	12.2%
PIN105	chr15	87187668	G	A	ACAN	37.9%
PIN105	chr19	45901528	C	T	ADCK4	13.2%
PIN105	chr4	71502935	AGG		AMBN	24.6%
PIN105	chr1	1377328	G	A	ATAD3C	27.8%
PIN105	chr19	47177802	G	A	ATP1A3	37.6%
PIN105	chr12	55279286	G	A	Baz2a	30.5%
PIN105	chr5	944299	C	T	BRD9	33.7%
PIN105	chr12	7133124	C	T	C1r	21.7%
PIN105	chr17	40335587	G	A	CCDC103	25.0%
PIN105	chr13	35910326	G	T	CCNA1	18.1%
PIN105	chr20	59861293	G	A	CDH4	34.0%
PIN105	chr9	122213487	G	A	CDK5RAP2	27.1%
PIN105	chr11	125370123	G	T	CDON	32.0%
PIN105	chr12	6579260	G	C	CHD4	11.8%
PIN105	chr7	146723738	G	A	CNTNAP2	40.9%

PIN105	chr2	237928304	G	A	COL6A3	26.4%
PIN105	chr10	125506748	G	A	CPXM2	28.5%
PIN105	chr4	5675408	C	T	EVC2	29.7%
PIN105	chr14	93465006	G	A	FAM181A	34.8%
PIN105	chr7	32980839	G	A	FKBP9	26.0%
PIN105	chr11	93773760	C	T	GPR83	11.8%
PIN105	chr11	122989492	C	T	Gramd1b	31.2%
PIN105	chr8	102748090	A	G	GRHL2	47.4%
PIN105	chr2	176696448	A	G	HOXD9	43.5%
PIN105	chr2	89830703	T	A	IGKV1D-13	37.5%
PIN105	chr16	28420878		GGG	IL27	44.7%
PIN105	chr12	73887866	C	A	KCNC2	62.1%
PIN105	chr12	21810133	G	A	KCNJ8	27.2%
PIN105	chr8	111054014	C	T	KCNV1	46.4%
PIN105	chr12	51167803	A	T	KRT6A	21.0%
PIN105	chr18	7001349	G	A	LAMA1	32.0%
PIN105	chr19	6329401	C	T	LOC100130856	36.2%
PIN105	chr14	20458108	T	A	LOC643332	36.1%
PIN105	chr11	47254246	G	T	MADD	39.0%
PIN105	chr15	21440336	T	A	MAGEL2	25.2%
PIN105	chr12	12374641	G	T	MANSC1	49.2%
PIN105	chr12	47732294	GG		MLL2	31.0%
PIN105	chr3	196981650	G	A	MUC4	31.0%
PIN105	chr1	146490279	C	G	NBPF14	20.5%
PIN105	chr2	206700818	G	A	NDUFS1	28.7%
PIN105	chr11	20656095	G	A	NELL1	16.1%
PIN105	chr19	60142740	G	A	NLRP7	28.4%
PIN105	chr6	32276729	C	T	NOTCH4	26.0%
PIN105	chr11	65948653	G	C	NPAS4	30.2%
PIN105	chr20	33654374	C	T	NULL	32.8%
PIN105	chr12	50933674	G	A	NULL	67.3%
PIN105	chr11	89243897	C	T	NULL	31.0%
PIN105	chr11	7827187	G	A	OR5E1P	28.2%
PIN105	chr11	56014883	G	C	OR5M8	25.5%
PIN105	chr5	140328418	C	T	PCDHAC1	19.6%
PIN105	chr10	105148201	C	T	PDCD11	20.3%
PIN105	chrX	48916472	C	T	PLP2	68.3%
PIN105	chr10	133608849	A	G	PPP2R2D	28.6%
PIN105	chr20	46705253	C	T	PREX1	24.0%
PIN105	chr1	40864889	C	T	RIMS3	27.1%
PIN105	chr17	55366625	G	A	RPS6KB1	10.6%

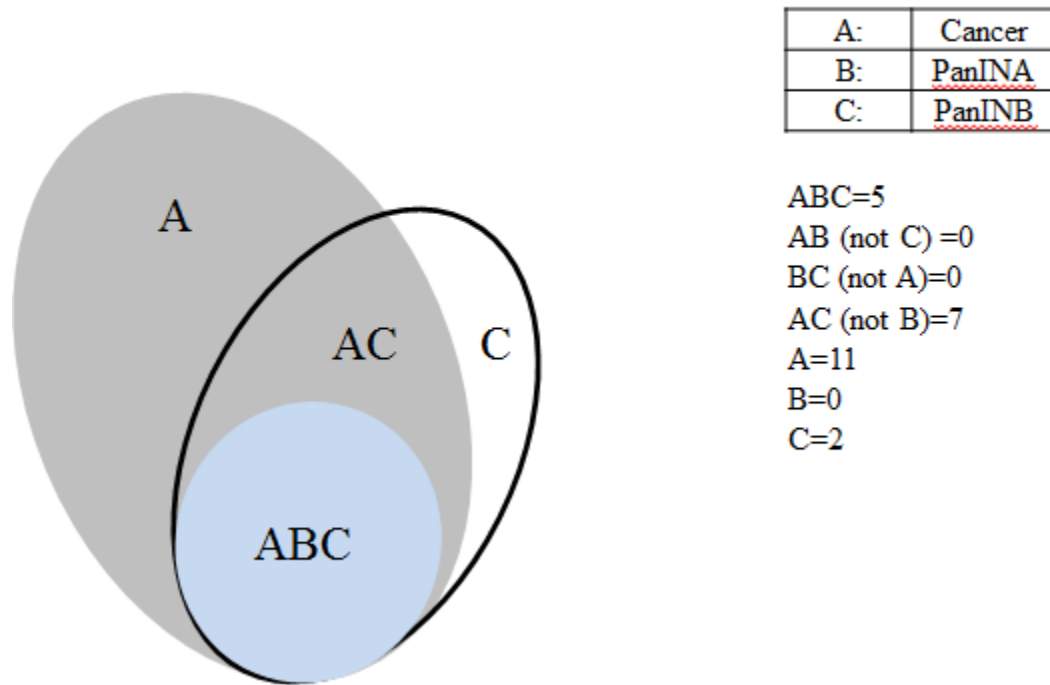
PIN105	chr15	31682692	G	A	RYR3	32.7%
PIN105	chr13	22826833	G	A	SACS	36.8%
PIN105	chr9	129569002	C	T	SH2D3C	18.6%
PIN105	chr2	230870402	A	G	SP140	23.3%
PIN105	chr6	33518736	G	A	SYNGAP1	36.6%
PIN105	chr19	55079379	C	T	TBC1D17	10.0%
PIN105	chr2	137634291	G	A	THSD7B	31.1%
PIN105	chr16	19375559	C	G	TMC5	24.9%
PIN105	chr11	100847318	C	A	TRPC6	29.1%
PIN105	chr5	135720401	G	A	TRPC7	32.2%
PIN105	chr5	110439637	G	A	TSLP	23.3%
PIN105	chr21	43388654	T	C	U2AF1	29.1%
PIN105	chr9	135794147	T	C	VAV2	32.5%
PIN105	chr10	118886091	C	T	VAX1	33.6%
PIN105	chr3	158564186	C	T	VEPH1	29.5%
PIN105	chr1	20542360	G	A	Vwa5b1	27.5%
PIN105	chr3	185440234	C	T	VWA5B2	25.4%
PIN105	chr8	124034343	G	A	ZHX2	32.2%
PIN105	chr5	60870403	T	A	Zswim6	29.3%
PIN106	chr16	16091783	C	G	ABCC1	40.4%
PIN106	chrX	108812640	T	A	ACSL4	53.8%
PIN106	chr5	81584980	G	A	ATG10	24.2%
PIN106	chr19	59193353	G	A	CACNG6	31.7%
PIN106	chr1	156418467	C	T	CD1D	18.0%
PIN106	chr6	75922147	C	T	COL12A1	44.4%
PIN106	chr14	23612046	C	T	CPNE6	13.8%
PIN106	chr7	30688061	G	A	CRHR2	10.6%
PIN106	chr8	2819486	G	A	CSMD1	42.9%
PIN106	chr2	172023133	G	A	DCAF17	21.6%
PIN106	chr1	3778924	C	T	DFFB	12.5%
PIN106	chr7	72735568	C	T	DNAJC30	14.9%
PIN106	chr6	56459870	G	T	DST	13.2%
PIN106	chr12	1007661	C	A	ERC1	12.3%
PIN106	chr5	76048248	C	T	F2R	25.6%
PIN106	chr16	87309697	C	G	FAM38A	50.0%
PIN106	chr11	61320340	T	C	FEN1	20.8%
PIN106	chr13	38322242	G	A	FREM2	16.6%
PIN106	chrX	151570887	G	T	GABRQ	16.9%
PIN106	chr6	55331668	C	T	GFRAL	35.6%
PIN106	chr7	41983705	G	A	GLI3	23.8%
PIN106	chr8	144430294	C	T	gli4	26.9%

PIN106	chr17	33739470	G	A	GPR179	15.8%
PIN106	chr7	124191540	G	A	GPR37	16.1%
PIN106	chrX	48558325	G	A	HDAC6	30.0%
PIN106	chr2	176696530	C	T	HOXD9	10.1%
PIN106	chrX	130236939	A	C	IGSF1	12.5%
PIN106	chr4	6158535	G	A	JAKMIP1	20.0%
PIN106	chr16	27537253	G	A	KIAA0556	14.6%
PIN106	chr6	43142223	T	A	KLC4	44.8%
PIN106	chr9	132952963	G	A	LAMC3	33.3%
PIN106	chr3	197872229	C	A	LRRC33	12.3%
PIN106	chr20	5901789	C	G	MCM8	20.7%
PIN106	chr3	99151396	C	T	MINA	24.6%
PIN106	chr19	8926095	A	C	MUC16	12.3%
PIN106	chr10	95231892	A	C	MYOF	10.7%
PIN106	chrM	11064	C	A	ND4L	13.5%
PIN106	chr15	71377840	T	C	NEO1	14.7%
PIN106	chr12	116149628	G	A	NOS1	26.0%
PIN106	chr12	116253142	G	A	NOS1	23.4%
PIN106	chr2	89180760	C	T	NULL	11.0%
PIN106	chr7	128307103	C	G	NULL	32.4%
PIN106	chr11	4892845	C	T	OR51G2	31.7%
PIN106	chr2	240633892	G	A	OR6B3	18.7%
PIN106	chr4	491210	C	T	PIGG	25.0%
PIN106	chr13	27138010	G	A	POLR1D	11.9%
PIN106	chr8	48954771	C	T	PRKDC	100.0%
PIN106	chr6	4002347	T		PRPF4B	10.2%
PIN106	chr10	25266206	A	T	PRTFDC1	21.5%
PIN106	chr6	170686315	A	C	PSMB1	35.3%
PIN106	chr7	5747414	G	A	RNF216	14.2%
PIN106	chr16	1936670	G	C	RPL3L	35.8%
PIN106	chr16	65515054	G	A	RRAD	13.3%
PIN106	chr2	96895829	G	A	SEMA4C	54.5%
PIN106	chr9	38058546	G	A	SHB	16.7%
PIN106	chr9	114692146	C	T	SLC46A2	13.3%
PIN106	chr2	174909259	C	T	SP9	30.0%
PIN106	chr3	9032284	T	A	SRGAP3	53.3%
PIN106	chr6	152818264	C	T	SYNE1	27.3%
PIN106	chr1	150347987	G	A	TCHH	20.2%
PIN106	chr11	120506131	G	A	TECTA	18.3%
PIN106	chr15	79420907	C	T	TMC3	12.6%
PIN106	chr12	124701270	C	T	TMEM132B	42.4%

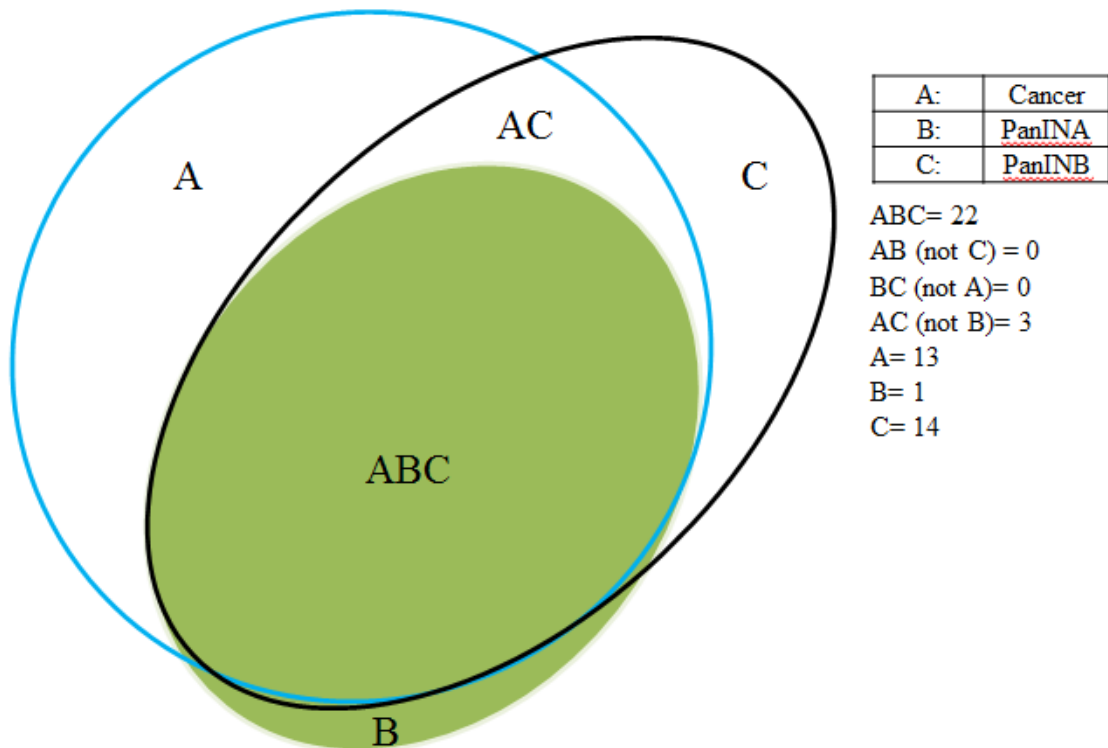
PIN106	chr6	44224565	G	A	TMEM63B	37.8%
PIN106	chr8	59913277	C	T	TOX	34.0%
PIN106	chr17	7517866	C	T	TP53	66.7%
PIN106	chr22	25267343	C	T	TPST2	12.0%
PIN106	chr12	31008031	G	A	TSPAN11	33.3%
PIN106	chr7	149174214	C	T	ZNF862	50.0%
PIN107	chr19	19621569	C	T	ATP13A1	32.2%
PIN107	chr4	81502933	A	G	C4orf22	22.9%
PIN107	chr9	21961120	G	A	CDKN2A	71.0%
PIN107	chr2	9488432	T	A	CPSF3	31.5%
PIN107	chr7	149805145	G	A	GIMAP8	25.1%
PIN107	chr19	50786589	C	T	GPR4	10.5%
PIN107	chr5	54491808	T	C	GPX8	30.7%
PIN107	chr10	105038267	G	T	INA	26.9%
PIN107	chr19	18406746	C	T	ISYNA1	38.1%
PIN107	chr17	21260403	C	A	KCNJ12	11.0%
PIN107	chr13	41042731	C	T	KIAA0564	25.5%
PIN107	chr19	50542564	G	A	KLC3	12.8%
PIN107	chr2	141190036	G	T	LRP1B	11.0%
PIN107	chr11	18151999	G	A	MRGPRX4	35.7%
PIN107	chr12	119268460	C	T	MSI1	35.3%
PIN107	chr5	76643618	C	T	PDE8B	24.3%
PIN107	chr9	8490778	C	T	PTPRD	29.0%
PIN107	chr14	20094999	C	A	RNASE9	25.5%
PIN107	chr17	53795712	C	T	RNF43	32.9%
PIN107	chr3	197188378	C	G	SDHAP1	20.5%
PIN107	chr2	219249141	G	A	STK36	31.4%
PIN107	chr3	30707974	G	A	TGFBR2	46.2%
PIN107	chr17	38720588	T	C	TMEM106A	26.5%
PIN107	chr22	34049063	G	A	TOM1	55.4%
PIN107	chr17	7519251	C	T	TP53	48.9%
PIN108	chr16	46737794	G	A	ABCC12	49.5%
PIN108	chr11	66809336	G	A	ADRBK1	49.2%
PIN108	chr10	24949595		A	ARHGAP21	38.8%
PIN108	chr4	87140756	C	T	ARHGAP24	31.8%
PIN108	chr1	44223250	A	G	B4GALT2	50.0%
PIN108	chr3	47512649	G	A	C3orf75	30.9%
PIN108	chr4	113744269	A	G	C4orf21	12.6%
PIN108	chr7	31701734	C	T	C7orf16	24.9%
PIN108	chr10	105205493	G	A	CALHM1	18.3%
PIN108	chr2	37726799	C	T	CDC42EP3	44.1%

PIN108	chr9	21961111	G	A	CDKN2A	76.2%
PIN108	chr11	99600708	G	C	Cntn5	23.2%
PIN108	chr8	139959592	G	A	COL22A1	19.7%
PIN108	chr8	2955148	C	T	CSMD1	11.7%
PIN108	chr20	23753951	C	A	CST2	28.1%
PIN108	chr2	80654813	G	A	CTNNA2	37.8%
PIN108	chr14	79747411	C	T	DIO2	48.6%
PIN108	chr21	40641515	G	A	DSCAM	33.6%
PIN108	chr8	16979724	G	A	EFHA2	45.9%
PIN108	chr6	159126555	G	A	EZR	19.7%
PIN108	chr8	101215653	G	T	FBXO43	13.7%
PIN108	chr9	34639413	C	A	GALT	51.7%
PIN108	chr16	56166454	G	A	GPR114	29.9%
PIN108	chr6	22678472	C	T	HDGFL1	72.2%
PIN108	chr8	19722251	G	A	INTS10	12.0%
PIN108	chr5	137789137	C	T	KDM3B	37.5%
PIN108	chr19	59836330	C	T	LILRB1	53.0%
PIN108	chr6	40507858	G	T	LRFN2	51.7%
PIN108	chr1	217413856	A	C	LYPLAL1	30.6%
PIN108	chrX	148774326	G	A	MAGEA8	35.9%
PIN108	chr7	72648493	G	A	MLXIPL	12.6%
PIN108	chr11	68234976	C	T	MTL5	27.7%
PIN108	chr17	10360480	C	G	MYH1	43.2%
PIN108	chrM	3878	G	A	ND1	25.4%
PIN108	chr14	36056742	G	A	NKX2-1	25.2%
PIN108	chr1	96917730	G	A	NULL	35.1%
PIN108	chr7	158108180	C	G	NULL	22.1%
PIN108	chr3	13343918	C	T	NUP210	37.0%
PIN108	chr11	131811342	C	A	OPCML	14.7%
PIN108	chr3	197993940	A	T	PAK2	32.7%
PIN108	chr3	78759369	T	A	ROBO1	17.0%
PIN108	chr20	592635	C	T	SCRT2	36.5%
PIN108	chr19	55155641	G	A	SIGLEC11	20.1%
PIN108	chr14	92018836	G	A	SLC24A4	40.1%
PIN108	chr1	19039302	C	T	TAS1R2	21.5%
PIN108	chr8	133980251	A	G	TG	12.6%
PIN108	chr9	100951356	G	A	TGFBR1	55.2%
PIN108	chrX	109303253	T	C	TMEM164	26.7%
PIN108	chr17	7519131	C	T	TP53	63.2%
PIN108	chr11	17499061	T	C	USH1C	49.3%
PIN108	chr7	158387892	C	T	WDR60	12.8%





**Figure 2-3. PIN101 evolutionary relationship.** The grey circle represents the cancer, the blue circle represents PanINA, and the clear circle represents PanINB. Region A indicates mutations unique to the cancer; region AC indicates mutations shared between the cancer and PanINB; region C indicates mutations unique to PanINB; region ABC indicates mutations common to the cancer, PanINA, and PanINB.



**Figure 2-4. PIN108 evolutionary relationship.** The clear circle (blue outline) represents the cancer, the green circle represents PanINA, and the clear circle (black outline) represents PanINB. Region A indicates mutations unique to the cancer; region AC indicates mutations shared between the cancer and PanINB; region B indicates mutations unique to PanINA; region C indicates mutations unique to PanINB; region ABC indicates mutations common to the cancer, PanINA, and PanINB.

## Conclusions

Currently, the analysis of this study is ongoing. In the future, we will analyze these genetic data using a computational approach to quantify the heterogeneity among these lesions, how divergent they are, and the overall phylogenetic relationships. For now, I will summarize some preliminary observations of the data, noting that future conclusions will depend on our planned analyses.

The PanIN progression model explains how a PDAC arises: a PanIN evolves via an increase in cell division and nuclear atypia along with step-wise gains mutations in *KRAS*, *CDKN2A*, *TP53*, and *SMAD4*<sup>14</sup>. Despite the value of this model, it does not fully explain PanIN evolution. Yet, our understanding of the mutations acquired during PDAC formation has greatly improved with recent sequencing-based studies<sup>28-30</sup>, as has our evolutionary understanding of this disease<sup>83,84</sup>. For example, an average PDAC tumor has 3-4 driver mutations and ~45 total somatic alterations in protein-coding genes<sup>6</sup>. Additionally, metastatic PDAC takes ~15-20 years to develop<sup>83</sup>. Thus, sequencing approaches have greatly impacted our understanding of pancreatic cancer – these approaches also hold considerable potential for exploring PanIN evolution.

In this study, we aimed to reveal the genetic factors involved in PanIN evolution, specifically the order of driver events and the relatedness of multiple PanINs and matched PDAC. Given our sequencing approach, the discovered mutations are novel and reliable for two reasons. First, compared to whole genome amplification approaches, whole exome sequencing (WES) of genomic DNA produces mutation lists that are less prone to amplification artifacts or sequencing bias – both common issues in next generation sequencing<sup>65,66</sup>. Second, previous studies of PanINs were restricted to either

comparing many single PanINs from unmatched patients or using non-genetic approaches, neither of which are suitable for evolutionary analysis. Third, although a previous study sequenced multiple PanINs from within a patient, the PanINs were not completely isolated from the matching PDAC<sup>89</sup>. Paradoxically, PanINs near the matched tumor may actually result from ductal cancerization: the PDAC creates PanIN-like lesions as the tumor moves along the duct<sup>14</sup>. Thus, the “PanINs” may not be true precursor lesions because they originate from the PDAC, making it difficult to interpret the mutation data in an evolutionary context. To limit this possibility, we isolated PanINs that were either on separate slides of tissue or were geographically distinct from the matched PDAC tumor with a maximum amount of intervening normal tissue.

Exome sequencing targets the protein-coding regions of the genome, thus driver mutations such as *KRAS*, *CDKN2A*, and *TP53* are readily detectable. In our data, *KRAS* point mutations were the most common, while mutations in *TP53* and *CDKN2A* are second and third, respectively (Table 2-2). Interestingly, *SMAD4* was not mutated in any of our PanINs or PDAC samples: this was due to either normal *SMAD4* status or the inability of WES to detect gene deletions (although WES would have detected point mutations). However, two PDAC tumors were found to have mutations in *TGFβR2*, the receptor known to signal downstream to SMAD proteins. Other potential drivers in these data included point mutations in *MLL2*, *ERBB2*, and *SMARCA4*.

We find that when comparing multiple samples from a patient, we can use the shared mutations to infer the evolutionary relationship of the PanINs and the PDAC tumor. In every patient, this analysis revealed that the PanINs and PDAC derived from a common ancestor prior to their divergence during carcinogenesis. Interestingly, when

comparing the four patients for which we had two PanINs and one PDAC, we observed a wide range in the number of the overall shared mutations (5-24). Thus, there is variability in the genetic relatedness of any two PanINs or PanIN and PDAC. Additionally, the existence of shared mutations between the PanINs (i.e. a clonal relationship) implies that the two PanINs are related and may even represent different components of the same PanIN. This is clinically significant because individual PanINs in a patient may in fact represent a limited total number, which would affect any prognostic assessment using the number of PanINs as a biomarker. This is especially important in the surgical setting, where margins are routinely assessed by the presence and number of PanINs.

Although the unique mutations are not relevant for evolutionary relationships, these alterations can indicate time or rate of evolutionary divergence. Overall, PDAC samples had a higher number of unique mutations than PanINs. However, some PanINs (in separate cases) had equal or more unique mutations compared to the matching PDAC. Further study would need to reveal whether these differences are due to mutation rate, subclonality, or selection pressures.

To summarize, we reveal that multiple PanINs share some mutations with each other and the matched tumor while harboring unique mutations. Interestingly, the data indicate that two PanINs may appear to be histologically distinct but in fact likely derive from a single lesion. We also find that PanIN histology does not always correlate with genetic progression. Collectively, these findings provide new insights into the timing, order, and relationship of genetic mutations underlying PanIN evolution.

### **Author Contributions**

C.I.D. and K.M selected the cases, K.M. performed microdissection, C.I.D., K.M., A.M.M., B.V., K.K., N.P., M.Z., and Y.J. designed experiments, A.M.M, M.Z., and Y.J. performed the experiments, all authors interpreted the data, C.I.D., A.M.M., B.V., and M.Z. wrote the manuscript.

## **Chapter 3 – Natural history of pancreatic cancer metastasis**

---

### **Synopsis**

**PROJECT TITLE:** Metastases in Pancreatic Cancer are Dominated by Genetic Homogeneity

**AUTHORS:** Alvin P. Makohon-Moore\*, Ming Zhang\*, Johannes G. Reiter\*, Ivana Bozic, Benjamin Allen, Deepanjan Kundu, Krishnendu Chatterjee, Fay Wong, Yuchen Jiao, Laura D. Wood, Ralph H. Hruban, Martin A. Nowak, Nickolas Papadopoulos, Kenneth W. Kinzler, Bert Vogelstein, and Christine A. Iacobuzio-Donahue

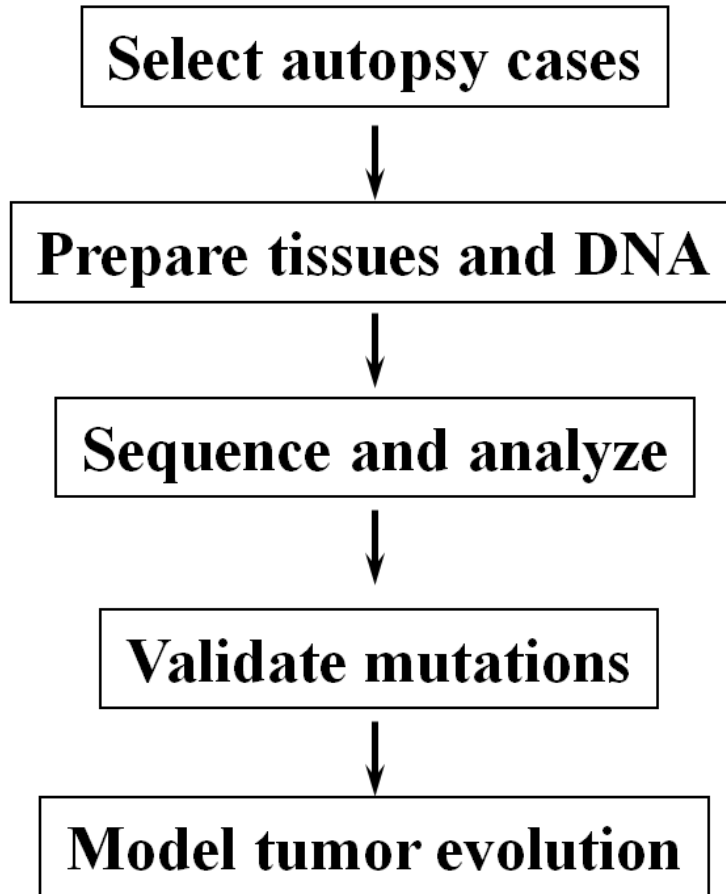
\*These authors contributed equally to this study.

**HYPOTHESIS:** The development of lethal pancreatic cancer metastasis is dependent on subclonal evolution within the primary site.

**AIM:** Determine the extent to which subclonal evolution is underlying PDAC metastasis.

Sub-aim 1: Assess how organ-specific metastases are phylogenetically related to each other.

Sub-aim 2: Determine if metastases in different target organs in the same patient are derived from a single versus multiple metastatic subclones within the matched primary carcinoma.



**Figure 3-1. Approach for pancreatic cancer metastasis evolution.** Autopsies from the JHMI rapid autopsy program were selected using strict criteria; ultimately, four cases with multiple metastases. Metastases and primary tumor sections were prepared and DNA was extracted. Whole genome sequencing for each samples was performed for 60X coverage. Mutations were analyzed and filtered only for those high quality alterations, which were subsequently used for targeted validation and screening. Using these data, we quantified the amount of heterogeneity among the seeding cells of metastases and compared these values to models of normal tissues. In addition, we used phylogenetic analyses to reveal the relationship among samples within each patient.



## Introduction

The primary cause of death from cancers is most often metastasis - dispersal of neoplastic cells from the primary tumor followed by tumor colonization and growth in a distant organ<sup>68,69</sup>. Though primary tumors can often be surgically excised, residual metastatic lesions generally cannot be. The growth of these metastatic lesions following surgery, and the consequent destruction of adjacent normal tissues, also is not well-controlled by systemic therapies<sup>68,69</sup>. This is particularly true of pancreatic ductal adenocarcinomas, the most common form of pancreatic cancers. These cancers are generally not detected until metastases have occurred, resulting in a dismal survival rate<sup>14,56</sup>.

Driver genes frequently mutated in pancreatic ductal adenocarcinomas include the *KRAS* oncogene and the *TP53*, *CDKN2A*, and *SMAD4* tumor suppressor genes<sup>14,28,29</sup>. Somatic mutations of these genes have been observed in various grades of pancreatic intraepithelial neoplasia (PanIN), the non-invasive precursor lesions leading to most invasive pancreatic cancers<sup>14</sup>. The increasing cellular dysplasia that occurs during pancreatic tumorigenesis is underpinned by sequentially-acquired driver gene mutations<sup>6</sup>. In this way, pancreatic cancer fits a conceptual framework in which randomly-acquired genetic variation confers progressive selective advantages to neoplastic cells over time<sup>6,12</sup>. Although well-defined alterations to driver genes occur early in pancreatic cancer evolution, it is unclear whether additional driver gene alterations are acquired later in development. Tumor heterogeneity has been observed in some solid tumor types, yet the extent to which the heterogeneous alterations contribute to the final biologic stage of cancer, i.e., metastasis, is unclear<sup>90-93</sup>. If the single cells (or group of cells) that initiates

metastasis are genetically heterogeneous with respect to driver genes and their pathways, then therapies to target these genes and pathways would not be able to induce complete responses or long-term survival, no matter how potent the therapy<sup>6</sup>. Moreover, the term "heterogeneity" is often used loosely, without quantification or comparison to a reference population of cells.

In the current study, we used whole genome sequencing of DNA from matched metastatic and primary carcinomas of four different patients to identify and quantify the degree of genetic heterogeneity among related lesions from the same individuals (Figure 3-1)<sup>83,84</sup>. We implemented strict clinical and technical criteria to select four patients for whole genome sequencing from autopsied patients in which tissues were available from a Rapid Medical Donation Program (Table 3-1)<sup>81</sup>. For example, patients who received any form of treatment (chemotherapy, radiation or surgery) that might cause genetic bottlenecks in the natural history of the disease<sup>12,94</sup> were excluded. Additionally, patients with a history of a second primary malignancy originating in a different tissue, such as the ovary or lung, or whose pancreatic cancer was of an unusual subtype, were excluded. Patients who were diagnosed with unusual biological variants of infiltrating pancreatic cancer were also excluded because they differ from the more common ductal adenocarcinomas in their pathogenesis, driver gene alterations, and clinical outcome<sup>95</sup>. Finally, we required that all patients have pancreatic adenocarcinoma (PDAC), Stage IV disease, and that the metastases available from each patient accurately represented their disease burden at autopsy (Table 3-1).

Histologic sections were prepared from snap frozen samples of the primary tumor and metastases from these four patients to estimate tumor cellularity and tissue quality.

Samples with low neoplastic cellularity were excluded, as were any tissue samples with confluent necrosis that would yield low quality genomic DNA (gDNA) (Methods). The remaining samples were macrodissected to remove as much normal tissue as possible before purifying gDNA. A similar approach for macrodissection was previously described<sup>83</sup>, but in that study only a single metastasis from each patient was evaluated by genome-wide Sanger sequencing, precluding quantification of heterogeneity and limiting the ability to discern evolutionary relationships. In all, we identified 26 metastatic lesions with relatively high fractions of neoplastic cells, as well as samples of the primary tumor and normal tissues, from these four patients (Tables 3-2 and 3-3).

Genomic DNA from 39 samples (26 metastatic lesions plus distinct parts of primary tumors and normal tissues) were evaluated by 60X whole genome sequencing using an Illumina Hi-Seq 2000 platform<sup>96</sup> (Figure 3-2a, b). Importantly, all metastases were discrete tumors by both gross examination at autopsy and histologic review, ensuring that each metastasis represented an independent neoplasm at that location (Figure 3-2c). The metastases were derived from a diverse set of organ sites including the liver, lung, peritoneum and lymph nodes, all typical secondary sites of pancreatic cancer<sup>79</sup>. DNA from the normal tissues of each patient was used to facilitate the identification of somatic variants.

**Table 3-1. Clinical characteristics of four pancreatic cancer patients.**

<b>Case</b>	<b>Age</b>	<b>Gender</b>	<b>Diagnosis<sup>a</sup></b>	<b>Clinical Stage</b>	<b>Treatment</b>	<b>Survival<sup>b</sup></b>	<b>Metastatic Burden<sup>c</sup></b>	<b>Organ Sites Affected</b>
Pam01	59	Male	PDAC	IV	No	7 mo	5	Liver, Lymph Node
Pam02	69	Female	PDAC	IV	No	0.5 mo	>100	Liver
Pam03	79	Male	PDAC	IV	No	10 mo	50	Liver, Lung
Pam04	74	Male	PDAC	IV	No	3 mo	>100	Peritoneum

<sup>a</sup>PDAC: pancreatic ductal adenocarcinoma

<sup>b</sup>from diagnosis to death

<sup>c</sup>total number of metastases at autopsy based on gross and histologic assessment

**Table 3-2. Samples used for Whole genome (WGS) and directed sequencing.**

<b>Case</b>	<b>Tissue type</b>	<b># of samples</b>	<b># used for WGS</b>	<b># used for Targeted Sequencing</b>
Pam01	Normal tissues	1	1	1
	Distinct parts of primary tumors	2	0	2
	Metastases	4	4	4
Pam02	Normal tissues	1	1	1
	Distinct parts of primary tumor	18	3	18
	Metastases	8	8	8
Pam03	Normal tissues	1	1	1
	Distinct parts of primary tumor	12	3	12
	Metastases	12	8	12
Pam04	Normal tissues	1	1	1
	Distinct parts of primary tumor	27	3	27
	Metastases	6	6	6

**Table 3-3. Samples analyzed.**

Case	MetomeID	Tissue type	Slice #	Slice position (Column)	Slice position (Row)	% Tumor cellularity assessed by histology	Sequencing Method
Pam01	Germline	Normal Lung	N/A	N/A	N/A	N/A	WGS & Targeted
Pam01	LiM1	Left lobe liver met	N/A	N/A	N/A	90	WGS & Targeted
Pam01	LiM2	Right lobe liver met	N/A	N/A	N/A	70	WGS & Targeted
Pam01	NoM1	Pelvic lymph node met	N/A	N/A	N/A	80	WGS & Targeted
Pam01	NoM2	Portal lymph node met	N/A	N/A	N/A	70	WGS & Targeted
Pam01	PT1	Primary tumor	N/A	N/A	N/A	90	Targeted
Pam01	PT2	Primary tumor	N/A	N/A	N/A	80	Targeted
Pam02	Germline	Normal Skin	N/A	N/A	N/A	N/A	WGS & Targeted
Pam02	LiM1	Liver Met	N/A	N/A	N/A	50	WGS & Targeted
Pam02	LiM2	Liver Met	N/A	N/A	N/A	40	WGS & Targeted
Pam02	LiM3	Liver Met	N/A	N/A	N/A	70	WGS & Targeted
Pam02	LiM4	Liver Met	N/A	N/A	N/A	70	WGS & Targeted

Pam02	LiM5	Liver Met	N/A	N/A	N/A	50	WGS & Targeted
Pam02	LiM6	Liver Met	N/A	N/A	N/A	50	WGS & Targeted
Pam02	LiM7	Liver Met	N/A	N/A	N/A	70	WGS & Targeted
Pam02	PT1	Primary tumor	4	1	2	50	Targeted
Pam02	PT10	Primary tumor	4	2	2	80	Targeted
Pam02	PT11	Primary tumor	4	3	4	70	Targeted
Pam02	PT12	Primary tumor	4	5	3	80	Targeted
Pam02	PT13	Primary tumor	4	4	4	80	Targeted
Pam02	PT14	Primary tumor	4	4	2	70	Targeted
Pam02	PT15	Primary tumor	3	3	3	70	Targeted
Pam02	PT16	Primary tumor	4	4	3	80	Targeted
Pam02	PT17	Primary tumor	5	4	3	70	Targeted
Pam02	PT18	Primary tumor	5	5	3	90	WGS & Targeted
Pam02	PT2	Primary tumor	2	1	2	40	Targeted
Pam02	PT3	Primary tumor	6	4	2	60	Targeted
Pam02	PT4	Primary tumor	5	3	2	70	WGS & Targeted

Pam02	PT5	Primary tumor	6	2	2	80	Targeted
Pam02	PT6	Primary tumor	4	3	3	70	Targeted
Pam02	PT7	Primary tumor	4	3	2	60	Targeted
Pam02	PT8	Primary tumor	3	1	3	80	Targeted
Pam02	PT9	Primary tumor	3	2	4	70	WGS & Targeted
Pam03	Germline	Normal Muscle	N/A	N/A	N/A	N/A	WGS & Targeted
Pam03	LiM1	Liver Met	N/A	N/A	N/A	90	WGS & Targeted
Pam03	LiM2	Liver Met	N/A	N/A	N/A	90	WGS & Targeted
Pam03	LiM3	Liver Met	N/A	N/A	N/A	90	WGS & Targeted
Pam03	LiM4	Liver Met	N/A	N/A	N/A	90	WGS & Targeted
Pam03	LiM5	Liver Met	N/A	N/A	N/A	90	WGS & Targeted
Pam03	LiM6	Liver Met	N/A	N/A	N/A	80	Targeted
Pam03	LiM7	Liver Met	N/A	N/A	N/A	70	Targeted
Pam03	LiM8	Liver Met	N/A	N/A	N/A	80	Targeted
Pam03	LiM9	Liver Met	N/A	N/A	N/A	80	Targeted
Pam03	LuM1	Lung Met	N/A	N/A	N/A	10-20	WGS & Targeted

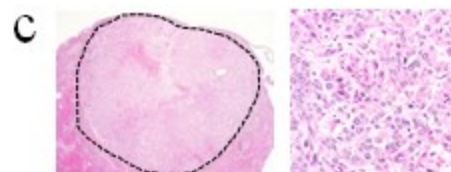
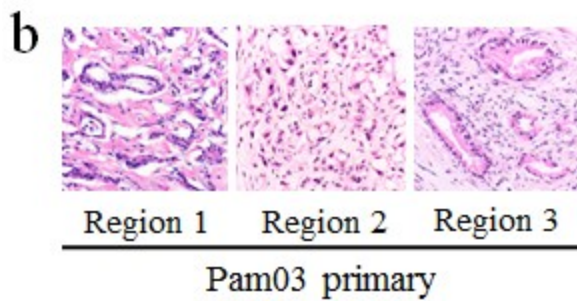
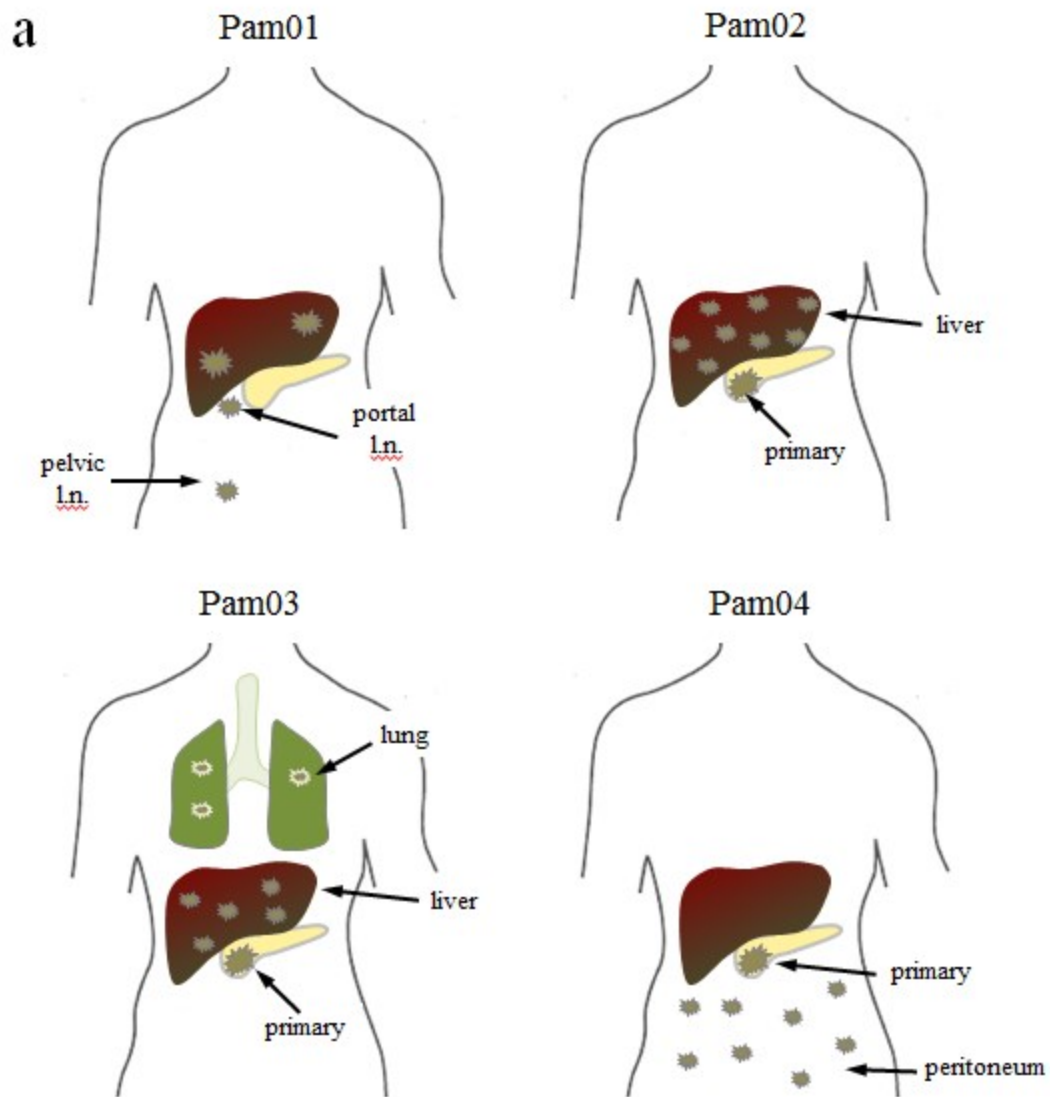


Pam03	LuM2	Lung Met	N/A	N/A	N/A	10	WGS & Targeted
Pam03	LuM3	Lung Met	N/A	N/A	N/A	30-60	WGS & Targeted
Pam03	PT1	Primary tumor	3	2	1	10	Targeted
Pam03	PT10	Primary tumor	4	2	2	70	WGS & Targeted
Pam03	PT11	Primary tumor	2	3	2	70	WGS & Targeted
Pam03	PT12	Primary tumor	6	4	2	70	WGS & Targeted
Pam03	PT2	Primary tumor	2	4	1	70	Targeted
Pam03	PT3	Primary tumor	4	1	3	50	Targeted
Pam03	PT4	Primary tumor	5	1	1	10	Targeted
Pam03	PT5	Primary tumor	7	4	2	60	Targeted
Pam03	PT6	Primary tumor	5	2	2	40	Targeted
Pam03	PT7	Primary tumor	2	3	1	70	Targeted
Pam03	PT8	Primary tumor	3	1	2	60	Targeted
Pam03	PT9	Primary tumor	4	1	1	20	Targeted
Pam04	Germline	Normal Skin	N/A	N/A	N/A	N/A	WGS & Targeted
Pam04	PeM1	Peritoneal Met	N/A	N/A	N/A	70	WGS & Targeted

Pam04	PeM2	Peritoneal Met	N/A	N/A	N/A	70	WGS & Targeted
Pam04	PeM3	Peritoneal Met	N/A	N/A	N/A	70	WGS & Targeted
Pam04	PeM4	Peritoneal Met	N/A	N/A	N/A	60	WGS & Targeted
Pam04	PeM5	Peritoneal Met	N/A	N/A	N/A	60	WGS & Targeted
Pam04	PeM6	Peritoneal Met	N/A	N/A	N/A	70	WGS & Targeted
Pam04	PT1	Primary tumor	5	4	1	20	Targeted
Pam04	PT10	Primary tumor	6	3	3	50	Targeted
Pam04	PT11	Primary tumor	4	1	1	30	Targeted
Pam04	PT12	Primary tumor	8	2	1	50	Targeted
Pam04	PT13	Primary tumor	7	3	1	50	Targeted
Pam04	PT14	Primary tumor	2	4	1	60	Targeted
Pam04	PT15	Primary tumor	8	3	1	40	Targeted
Pam04	PT16	Primary tumor	4	3	1	80	Targeted
Pam04	PT17	Primary tumor	4	2	2	80	Targeted
Pam04	PT18	Primary tumor	7	2	1	50	Targeted
Pam04	PT19	Primary tumor	6	4	3	50	Targeted

Pam04	PT2	Primary tumor	5	2	2	50	WGS & Targeted
Pam04	PT20	Primary tumor	6	5	3	50	Targeted
Pam04	PT21	Primary tumor	7	1	1	60	Targeted
Pam04	PT22	Primary tumor	2	2	1	40	Targeted
Pam04	PT23	Primary tumor	2	3	2	80	Targeted
Pam04	PT24	Primary tumor	2	1	2	60	Targeted
Pam04	PT25	Primary tumor	6	2	3	50	Targeted
Pam04	PT26	Primary tumor	4	2	1	70	WGS & Targeted
Pam04	PT27	Primary tumor	7	2	3	50	WGS & Targeted
Pam04	PT3	Primary tumor	6	4	2	40	Targeted
Pam04	PT4	Primary tumor	3	3	1	80	Targeted
Pam04	PT5	Primary tumor	8	3	2	20	Targeted
Pam04	PT6	Primary tumor	4	4	2	70	Targeted
Pam04	PT7	Primary tumor	7	2	2	50	Targeted
Pam04	PT8	Primary tumor	7	3	3	<10	Targeted
Pam04	PT9	Primary tumor	4	3	2	70	Targeted

**Figure 3-2. Metastatic disease of four pancreatic cancer patients.** a. Anatomic locations of the primary carcinomas (Pam02-Pam04) and discrete metastases (all cases) used for whole genome sequencing. b. Histology of three geographically-independent primary tumor samples from patient Pam03 used for sequencing. c. Low and high power view of a discrete liver metastasis from patient Pam03. The dashed line in the low power view outlines the borders of the metastasis.



## Results

The raw data were filtered for mapping quality and aligned to the hg19 human reference genome, revealing an average coverage of 68X with 97.6% of bases covered at >10x. (Table 3-4). A total of 165,815 potential coding and noncoding somatic mutations, with an average of 4,759 potential mutations per sample, were identified. As *KRAS* is consistently mutated in PDACs, we used the mutant allele fraction (MAF) of *KRAS* as an independent measure of neoplastic cellularity of our samples. This is particularly important given the high non-neoplastic stromal content of many PDACs<sup>23</sup>. The fraction of neoplastic cells in the tissues, as assessed histologically, was highly correlated with the fraction of neoplastic cells judged by *KRAS* MAF (correlation coefficient = 0.44).

It is well known that massively parallel sequencing yields many artificial mutations<sup>66,96</sup>. To enrich for a meaningful and manageable number of genuine somatic alterations within the coding regions of all known genes, we employed criteria that limited potential sequence artifacts and applied them to the sequencing data on these genes (Methods). We then independently assessed every mutation identified via these criteria using a targeted sequencing approach. In brief, capture probes containing ~100 nt surrounding each of 2356 potentially mutant bases were designed. Libraries from each of the 39 samples used for whole genome sequencing were captured with these probes and subjected to massively parallel sequencing. We thereby validated a total of 614 mutations (106 to 233 among each of the four patients [Table 3-5]). Among these, *KRAS* mutations were identified in every sample of all four patients and mutations in other driver genes (e.g., *TP53*, *SMAD4*, *ARID1A*, and *ATM*) were consistent with reports of the genomic landscapes of PDACs<sup>28,29</sup> (Table 3-6).

**Table 3-4. Average coverage per base.**

<b>Case</b>	<b>WGS - Average Coverage</b>	<b>WGS - Average Distinct Coverage</b>	<b>WGS - % bases covered at &gt;10X reads</b>	<b>Targeted Sequencing - Average coverage</b>	<b>Targeted Sequencing - Average Distinct Coverage</b>	<b>Targeted sequencing -% bases covered at &gt;10X reads</b>
Pam01	66.1	60.8	97.5%	649.8	257.1	92.6%
Pam02	66.9	62.2	97.5%	841.6	319.0	89.9%
Pam03	71.2	64.7	97.6%	702.4	207.3	88.7%
Pam04	67.7	64.0	97.7%	543.8	281.2	92.6%

For WGS, averaged across all genomic positions in all 35 cancer samples analyzed (see Supplemental Table 3); for targeted sequencing, averaged over the mutant positions assessed in all 89 samples analyzed.

**Table 3-5. Variants validated by targeted sequencing.**

<b>Case</b>	<b>Chr.</b>	<b>Position</b>	<b>Change<sup>a</sup></b>	<b>Gene</b>	<b>Median %Mut<sup>b</sup></b>
Pam01	chr10p15.3	293378	G>A	ZMYND11	2.00%
Pam01	chr16p13.3	319382	G>A	RGS11	22.10%
Pam01	chr19p13.3	434763	T>G	SHC2	10.10%
Pam01	chr19p13.3	620033	G>A	POLRMT	61.80%
Pam01	chr19p13.3	651661	G>C	RNF126	7.90%
Pam01	chr16p13.3	828671	A>G	MSLNL	7.40%
Pam01	chr11p15.5	1095350	G>A	MUC2	22.30%
Pam01	chr11p15.5	1103892	G>A	MUC2	19.70%
Pam01	chr2p25.3	1437330	G>A	TPO	35.50%
Pam01	chr16p13.3	2835797	G>T	PRSS33	14.70%
Pam01	chr5p15.33	3599646	C>T	IRX1	1.30%
Pam01	chr19p13.3	4652015	C>G	TNFAIP8L1	8.10%
Pam01	chr20p12.3	6057895	C>A	FERMT1	11.10%
Pam01	chr1p36.23	7887205	CGGC>-	PER3	12.10%
Pam01	chr16p13.2	8953059	G>A	CARHSP1	9.20%
Pam01	chr19p13.2	9073689	G>A	MUC16	14.90%
Pam01	chr12p13.31	9220824	A>-	A2M	14.20%
Pam01	chr19p13.2	9869198	A>G	ZNF846	10.40%
Pam01	chr11p15.4	10529739	G>A	MTRNR2L8	10.00%
Pam01	chr12p13.2	11286797	A>G	TAS2R30	7.50%
Pam01	chr12p13.2	11286807	T>C	TAS2R30	7.90%
Pam01	chr19p13.2	11917213	C>T	ZNF491	5.70%
Pam01	chr1p36.22	12353670	G>A	VPS13D	17.10%
Pam01	chr19p13.2	12541716	T>G	ZNF443	8.60%
Pam01	chr20p12.1	14306827	T>A	FLRT3	3.70%
Pam01	chr18p11.21	15325689	A>G	LOC644669	16.20%
Pam01	chr22q11.1	17600899	C>T	CECR6	5.30%
Pam01	chr20p11.23	20610203	G>A	RALGAPA2	11.10%
Pam01	chr14q11.2	20763539	C>G	TTC5	10.30%
Pam01	chr12p12.2	21065268	T>G	SLCO1B3,SLCO1B7	6.30%
Pam01	chr16p12.2	23391867	G>A	SCNN1B	41.70%
Pam01	chr12p12.1	23696264	C>A	SOX5	3.40%
Pam01	chr22q11.23	24126073	G>A	MMP11	15.80%
Pam01	chr16p12.1	24583727	G>A	RBBP6	7.20%
Pam01	chr1p36.11	24995797	CAAA>-	SRRM1	14.40%
Pam01	chr12p12.1	25398284	C>A	KRAS	66.10%
Pam01	chr15q13.1	28885410	A>C	AK307870,AK309255	45.30%
Pam01	chr16p11.2	30288654	G>A	LOC595101	13.40%



Pam01	chr6p21.33	30916645	G>A	DPCR1	13.70%
Pam01	chr21q22.11	31692113	C>T	KRTAP26-1	15.20%
Pam01	chr2p23.1	31751329	G>T	SRD5A2	9.70%
Pam01	chr5p13.3	32233934	AG>-	MTMR12	13.80%
Pam01	chr19q13.11	32845080	T>C	ZNF507	36.20%
Pam01	chr6p21.32	33139540	G>A	COL11A2	15.50%
Pam01	chr2p22.3	33748944	A>G	RASGRP3	23.80%
Pam01	chr1p34.3	36028940	C>T	NCDN	45.70%
Pam01	chr10p11.21	37436367	ATAA>-	ANKRD30A	16.00%
Pam01	chr22q13.1	38051975	C>T	SH3BP1	12.50%
Pam01	chr1p34.3	38274674	->C	C1orf122	2.60%
Pam01	chr15q15.1	41109490	A>C	PPP1R14D	16.80%
Pam01	chrXp11.4	41587159	C>A	GPR82	60.00%
Pam01	chr17q21.31	42855326	A>C	ADAM11	20.40%
Pam01	chr19q13.31	44351168	G>A	ZNF283	6.90%
Pam01	chr19q13.31	44351171	A>G	ZNF283	4.50%
Pam01	chr19q13.31	44351172	T>G	ZNF283	20.10%
Pam01	chr6p21.1	44413604	GG>-	CDC5L	9.80%
Pam01	chr20q13.12	44639816	G>A	MMP9	17.60%
Pam01	chr19q13.31	44890632	A>G	ZNF285	3.30%
Pam01	chr19q13.31	44933564	G>A	ZNF229	43.90%
Pam01	chr7p13	45119334	C>T	NACAD	3.90%
Pam01	chr21q22.3	46032699	T>G	KRTAP10-8	1.50%
Pam01	chr14q21.3	47770744	G>A	MDGA2	6.90%
Pam01	chr12q13.11	48458896	->T	SENP1	15.90%
Pam01	chr19q13.33	48994758	->G	LMTK3	12.20%
Pam01	chr19q13.33	50755952	C>T	MYH14	2.10%
Pam01	chr10q11.23	50822525	AGA>-	CHAT	11.40%
Pam01	chr22q13.33	51065310	G>A	ARSA	19.70%
Pam01	chr1p32.3	51946978	A>G	EPS15	15.80%
Pam01	chr19q13.41	52249338	T>C	FPR1	7.90%
Pam01	chr3p21.1	53853681	G>T	CHDH	17.90%
Pam01	chr12q13.13	54394402	G>T	HOXC9	71.50%
Pam01	chr19q13.42	54966627	C>T	LENG8	12.80%
Pam01	chr13q14.3	55015398	C>T	MIR1297,NONE	18.30%
Pam01	chr19q13.42	55915713	G>A	UBE2S	6.60%
Pam01	chr19q13.43	57176339	G>A	ZNF835	49.80%
Pam01	chr12q13.3	57178696	C>T	HSD17B6	25.60%
Pam01	chr8q12.1	57358423	G>A	PENK	2.10%
Pam01	chr7p11.2	57528688	G>C	ZNF716	15.20%
Pam01	chr12q13.3	58025103	C>-	B4GALNT1	0.60%

Pam01	chr17q23.2	58503559	G>A	C17orf64	16.80%
Pam01	chr11q12.2	60541371	C>T	MS4A15	41.80%
Pam01	chr1p31.3	61553933	G>A	NFIA	24.60%
Pam01	chr17q23.3	62049511	C>T	SCN4A	9.00%
Pam01	chr12q14.2	64724816	->AA	BC042855	25.80%
Pam01	chr16q22.1	67037029	C>T	CES4A	7.10%
Pam01	chr1p31.2	68947256	G>A	DEPDC1	24.30%
Pam01	chr9q21.11	69440046	C>G	ANKRD20A4,DQ581988	11.80%
Pam01	chr16q22.1	70299546	G>A	AARS	23.00%
Pam01	chr8q13.3	70533348	C>T	SULF1	47.40%
Pam01	chr2p13.1	74074606	A>T	STAMBP	16.50%
Pam01	chr4q21.1	77038915	C>A	NUP54	16.60%
Pam01	chr7q11.23	77256331	G>C	PTPN12	21.20%
Pam01	chr10q22.3	78868312	C>T	KCNMA1	13.00%
Pam01	chr17q25.3	79226006	T>C	SLC38A10	22.30%
Pam01	chr10q22.3	79589192	G>C	DLG5	0.30%
Pam01	chr17q25.3	79982951	C>T	LRRC45	23.60%
Pam01	chr15q25.1	80737066	G>C	ARNT2	0.20%
Pam01	chr15q25.1	80737078	G>C	ARNT2	0.30%
Pam01	chr8q21.13	82713832	C>A	SNX16	26.30%
Pam01	chr11q14.1	83173090	C>T	DLG2	22.10%
Pam01	chr16q23.3	84193306	G>A	DNAAF1	20.50%
Pam01	chr7q21.11	84815553	G>A	SEMA3D	26.20%
Pam01	chr9q22.2	91993742	G>A	SEMA4D	16.20%
Pam01	chr3q11.1	93605341	C>A	PROS1	10.70%
Pam01	chr15q26.1	93749856	T>C	RGMA,Mir	16.20%
Pam01	chr14q32.12	93943934	C>A	UNC79	44.40%
Pam01	chr10q23.33	94836364	T>C	CYP26A1	18.20%
Pam01	chr9q22.31	94939672	G>T	LINC00475,AK127087	19.40%
Pam01	chr11q22.1	99690407	G>T	CNTN5	21.20%
Pam01	chr7q22.1	100634402	G>A	MUC12	2.90%
Pam01	chr13q33.1	103387623	C>A	CCDC168	4.60%
Pam01	chr7q22.3	106509677	G>A	PIK3CG	26.40%
Pam01	chr8q23.1	106646486	C>A	ZFPM2	8.40%
Pam01	chr2q12.2	107107550	G>T	RGPD3,ST6GAL2	4.40%
Pam01	chr9q31.1	107298707	G>T	OR13C3	47.70%
Pam01	chr1p13.3	109805033	G>A	CELSR2	16.80%
Pam01	chr13q34	111120553	G>T	COL4A2	12.60%
Pam01	chr12q24.11	111317807	A>G	CCDC63	35.00%
Pam01	chr8q23.3	113563051	G>T	CSMD3	5.30%
Pam01	chr7q31.1	114268728	A>G	FOXP2	20.10%

Pam01	chrXq23	114426605	C>A	RBMXL3	27.50%
Pam01	chr5q22.3	114956347	A>T	TMED7,TMED7-TICAM2	11.70%
Pam01	chr4q26	119145740	C>A	NDST3	16.40%
Pam01	chr1p12	119427951	T>G	TBX15	19.00%
Pam01	chr12q24.31	124398945	G>A	DNAH10	3.90%
Pam01	chr7q31.33	126173205	A>T	GRM8	19.20%
Pam01	chr4q28.1	126241337	G>C	FAT4	23.50%
Pam01	chr3q22.1	130140020	C>G	COL6A5	12.30%
Pam01	chr2q21.1	130869587	C>T	POTEF	19.10%
Pam01	chr8q24.21	131172111	C>T	ASAP1	9.40%
Pam01	chr9q34.11	132623212	A>G	USP20	15.30%
Pam01	chr4q28.3	132779047	T>C	BC131768,BC040219	3.90%
Pam01	chr5q31.1	134263736	C>T	MIR4461	20.50%
Pam01	chr10q26.3	135280763	G>C	LOC619207,CYP2E1	21.30%
Pam01	chr2q21.3	136574957	C>G	LCT	7.50%
Pam01	chr7q33	137172366	C>T	DGKI	13.80%
Pam01	chr9q34.3	138664684	C>T	KCNT1	14.60%
Pam01	chr7q34	143036677	C>A	CLCN1	11.40%
Pam01	chr8q24.3	143857370	C>T	LYNX1	4.30%
Pam01	chr8q24.3	144789429	C>T	CCDC166	12.40%
Pam01	chr7q35	147092825	G>T	CNTNAP2	19.40%
Pam01	chr5q32	147693665	G>A	SPINK7	17.30%
Pam01	chr6q25.1	150464632	G>T	PPP1R14C	14.00%
Pam01	chr7q36.1	151921149	G>A	MLL3	1.60%
Pam01	chr7q36.2	154681009	G>A	DPP6	20.30%
Pam01	chr1q23.2	159505332	C>T	OR10J5	2.70%
Pam01	chr2q24.2	160801442	->T	PLA2R1	0.80%
Pam01	chr1q24.2	169391140	C>G	C1orf114	4.80%
Pam01	chr1q25.3	181727109	C>T	CACNA1E	1.10%
Pam01	chr4q35.1	187026708	C>A	TLR3,FAM149A	31.30%
Pam01	chr4q35.2	189660607	A>G	LOC401164,HSP90AA4P	6.90%
Pam01	chr3q29	195076927	G>C	ACAP2	20.20%
Pam01	chr2q32.3	197183328	C>G	HECW2	8.30%
Pam01	chr1q32.1	201179197	G>A	IGFN1	3.10%
Pam01	chr1q32.1	203316624	C>T	FMOD	3.50%
Pam01	chr2q33.3	207911040	A>G	CPO,KLF7	16.70%
Pam01	chr2q35	215884115	C>T	ABCA12	52.60%
Pam01	chr1q41	220375641	G>C	RAB3GAP2	5.50%
Pam01	chr1q42.13	228468048	G>A	OBSCN	10.60%
Pam01	chr2q37.1	231406681	G>A	SP100	6.20%
Pam01	chr1q42.3	235403665	C>G	ARID4B	9.50%

Pam01	chr1q43	237580371	G>A	RYR2	14.00%
Pam01	chr2q37.3	241710405	C>A	KIF1A	4.10%
Pam01	chr1q44	248487120	A>G	OR2M7	11.00%
Pam02	chrM	14619	A>T	JA040725	4.90%
Pam02	chr19p13.3	1877545	G>A	FAM108A1	5.80%
Pam02	chr1p36.33	2252996	G>A	MORN1	17.50%
Pam02	chr19p13.3	2482823	G>A	GADD45B,GNG7	4.60%
Pam02	chr5p15.33	2755264	G>A	C5orf38	34.70%
Pam02	chr17p13.2	4621284	G>T	ARRB2	29.70%
Pam02	chr11p15.4	5475616	C>T	OR51I2	18.80%
Pam02	chr11p15.4	6150599	G>A	OR56B4,OLFR690	1.40%
Pam02	chr11p15.4	6588737	G>T	DNHD1	24.00%
Pam02	chr17p13.1	7573996	A>G	TP53	27.00%
Pam02	chr1p36.22	9778888	G>A	PIK3CD	27.30%
Pam02	chr1p36.22	10322018	A>C	KIF1B	22.10%
Pam02	chr4p16.1	10445986	G>A	ZNF518B	21.20%
Pam02	chr12p13.2	11214495	T>C	TAS2R46	4.10%
Pam02	chr19p13.2	11598439	G>A	ZNF653	32.80%
Pam02	chr16p13.11	15219553	C>T	FLJ00285	3.80%
Pam02	chr16p13.11	15696493	A>G	KIAA0430	6.90%
Pam02	chr19p13.12	16060195	C>T	OR10H4	1.20%
Pam02	chr9p22.3	16552544	C>T	BNC2	11.40%
Pam02	chr9p21.3	21202040	G>C	IFNA7	3.60%
Pam02	chr9p21.3	21202073	C>G	IFNA7	3.30%
Pam02	chr9p21.3	21202105	G>T	IFNA7	2.90%
Pam02	chr20p11.22	21687122	C>T	PAX1	25.50%
Pam02	chr18q11.2	24496975	G>A	CHST9	16.80%
Pam02	chrXp21.3	25025352	G>A	ARX	23.30%
Pam02	chr12p12.1	25398284	C>A	KRAS	56.80%
Pam02	chr1p36.11	27058029	T>G	ARID1A	28.20%
Pam02	chr12p11.22	28114909	C>T	PTHLH	3.20%
Pam02	chr17q11.2	28545135	C>T	SLC6A4	15.00%
Pam02	chr7p14.3	29519883	G>C	CHN2	0.50%
Pam02	chrXp21.2	31196888	T>G	DMD	19.50%
Pam02	chr17q11.2	31439043	C>T	ACCN1	24.20%
Pam02	chr19q12	31768409	C>G	TSHZ3	34.10%
Pam02	chr13q13.1	32745177	C>T	FRY	1.00%
Pam02	chr6p21.32	33271918	T>C	TAPBP	23.50%
Pam02	chr16p11.2	33730720	C>T	IGH,IGH	11.20%
Pam02	chr17q12	33772558	G>A	SLFN13	20.00%
Pam02	chr9p13.3	33794789	G>A	PRSS3	2.60%

Pam02	chr19q13.12	35850471	T>G	FFAR3	2.80%
Pam02	chrXp11.4	40522273	G>A	MED14	2.40%
Pam02	chr17q21.31	42035354	G>A	PYY	22.10%
Pam02	chr19q13.2	42752861	G>A	ERF	40.70%
Pam02	chr2p21	43768425	G>A	THADA	18.00%
Pam02	chr17q21.31	44248558	G>A	KIAA1267	16.00%
Pam02	chr19q13.31	44351168	G>A	ZNF283	8.20%
Pam02	chr18q21.1	44773444	G>A	SKOR2	68.60%
Pam02	chr12q12	45169896	T>G	NELL2	20.60%
Pam02	chr1p34.1	45295178	C>T	PTCH2	9.30%
Pam02	chr6p12.3	46712103	C>T	PLA2G7,LOC100287718	33.30%
Pam02	chr11p11.2	48347475	C>A	OR4C3	7.80%
Pam02	chr15q21.1	48539175	G>A	SLC12A1	20.30%
Pam02	chrXp11.23	48853689	G>C	GRIPAP1	20.70%
Pam02	chr20q13.13	49493075	C>T	BCAS4	20.70%
Pam02	chr14q22.1	50901249	T>C	MAP4K5	18.40%
Pam02	chr3p21.2	51315131	C>T	DOCK3	30.40%
Pam02	chr3p21.2	51350308	A>G	DOCK3	2.20%
Pam02	chr15q21.2	52553310	G>A	MYO5C	0.00%
Pam02	chr19q13.41	53303182	T>C	ZNF28	5.40%
Pam02	chr19q13.42	54759298	C>T	LILRB5	32.80%
Pam02	chr2p16.1	55201001	G>A	RTN4	12.70%
Pam02	chr11q12.1	55890681	A>T	OR8H3	6.90%
Pam02	chr10q21.1	55944974	C>T	PCDH15	24.70%
Pam02	chr5q11.2	56777811	C>T	ACTBL2	0.50%
Pam02	chr4q12	57777227	C>A	REST	2.40%
Pam02	chr8q12.1	57906117	G>A	IMPAD1	25.80%
Pam02	chr17q24.2	66903933	G>A	ABCA8	27.30%
Pam02	chr9q13	67934780	T>C	ANKRD20A1,ANKRD20A3	2.20%
Pam02	chr1p31.1	70504580	C>T	LRRC7	23.70%
Pam02	chr15q24.1	74426985	G>A	ISLR2	24.10%
Pam02	chr15q25.1	79296424	C>T	RASGRF1	20.10%
Pam02	chr1p31.1	79403500	G>A	ELTD1	14.90%
Pam02	chr16q24.1	84270405	G>T	KCNG4	5.50%
Pam02	chr11q14.3	89502008	->GA	TRIM77P,TRIM49	10.70%
Pam02	chr16q24.3	89764014	G>A	SPATA2L	12.40%
Pam02	chr10q24.1	98588473	G>A	PIK3AP1,LCOR	9.50%
Pam02	chr12q23.1	98989235	C>G	SLC25A3	26.70%
Pam02	chr7q22.1	100346053	C>T	ZAN	27.50%
Pam02	chr14q32.2	100758898	C>T	SLC25A29	23.30%
Pam02	chr5q21.1	102526673	C>T	PIIP5K2	12.00%

Pam02	chr11q22.2	102709871	->T	MMP3	31.30%
Pam02	chr13q33.1	103390124	G>T	CCDC168	23.10%
Pam02	chrXq22.2	103495309	G>A	ESX1	16.30%
Pam02	chrXq22.3	106516773	C>T	CXorf41,KIAA1817	21.60%
Pam02	chrXq23	109247203	C>T	TMEM164	20.30%
Pam02	chr13q34	111532049	C>T	ANKRD10	48.20%
Pam02	chr3q13.2	112991493	G>A	BOC	23.70%
Pam02	chr8q23.3	113563071	A>G	CSMD3	20.00%
Pam02	chr9q32	116132187	->G	BSPRY	23.80%
Pam02	chr5q23.2	122714030	CT>-	CEP120	20.90%
Pam02	chr7q32.1	127250961	G>A	PAX4	8.60%
Pam02	chr7q32.1	128632104	A>G	TNPO3	28.40%
Pam02	chr2q21.1	132288203	G>A	CCDC74A	5.30%
Pam02	chr5q31.1	132535173	C>A	FSTL4	11.60%
Pam02	chrXq26.3	134992667	G>A	SAGE1	22.40%
Pam02	chrXq26.3	136649856	G>A	ZIC3	0.60%
Pam02	chr7q33	136939692	C>T	PTN	13.50%
Pam02	chr8q24.23	139209858	G>A	FAM135B	34.00%
Pam02	chr9q34.3	139996589	G>A	MAN1B1	4.40%
Pam02	chr5q31.3	140530240	C>A	PCDHB6	27.70%
Pam02	chr5q31.3	141325091	C>T	PCDH12	27.30%
Pam02	chrXq28	153172097	G>A	AVPR2	24.10%
Pam02	chr6q25.2	153323704	C>G	MTRF1L	2.90%
Pam02	chr4q31.3	153897517	T>G	FHDC1	13.40%
Pam02	chr4q31.3	155254017	G>A	DCHS2	13.10%
Pam02	chr2q31.1	176973700	C>T	HOXD11	22.70%
Pam02	chr2q31.2	179578779	T>A	TTN	17.90%
Pam02	chr2q31.2	179579932	G>A	TTN	15.40%
Pam02	chr3q27.1	184297520	G>-	EPHB3	19.30%
Pam02	chr3q27.2	185882739	C>T	DGKG	24.90%
Pam02	chr1q42.12	226173221	T>C	C1orf55	15.10%
Pam02	chr2q37.3	239757313	G>A	TWIST2	20.50%
Pam03	chr6p25.3	348133	G>A	DUSP22	6.70%
Pam03	chr11p15.5	1095811	G>A	MUC2	4.80%
Pam03	chr5p15.33	1593282	A>G	SDHAP3	1.70%
Pam03	chr4p16.3	3156068	C>T	HTT	13.30%
Pam03	chr17p13.2	4798417	G>A	MINK1	15.00%
Pam03	chr7p22.1	4901397	C>T	PAPOLB	15.90%
Pam03	chr5p15.32	5232549	G>A	ADAMTS16	16.80%
Pam03	chr4p16.2	5620225	C>T	EVC2	11.80%
Pam03	chrXp22.32	5821413	G>A	NLGN4X,NLGN4Y	24.20%

Pam03	chr11p15.4	6567175	C>G	DNHD1	11.50%
Pam03	chr5p15.31	7706895	G>A	ADCY2	10.00%
Pam03	chr17p13.1	8046152	G>A	PER1	7.70%
Pam03	chr19p13.2	8601235	G>A	MYO1F	14.40%
Pam03	chr19p13.2	9874058	A>G	ZNF846	9.70%
Pam03	chr12p13.2	11285909	G>C	TAS2R30	2.30%
Pam03	chr12p13.2	11286172	G>A	TAS2R30	2.20%
Pam03	chr12p13.2	11286188	G>A	TAS2R30	3.70%
Pam03	chr12p13.2	11286239	A>G	TAS2R30	2.90%
Pam03	chr19p13.2	12059902	A>G	ZNF700	25.70%
Pam03	chr1p36.21	12853509	A>C	PRAMEF1	2.80%
Pam03	chr1p36.21	12854510	A>C	PRAMEF1	4.30%
Pam03	chrYq11.1	13368460	G>C	DQ590126,IGL@	1.40%
Pam03	chr17p12	15522550	C>T	CDRT1	3.50%
Pam03	chr19p13.11	16539518	C>T	EPS15L1	14.20%
Pam03	chr8p22	17581311	C>T	MTUS1	2.10%
Pam03	chr6p22.3	21089124	G>A	CDKAL1	17.60%
Pam03	chr19p12	21990994	A>C	ZNF43	20.00%
Pam03	chr3p24.3	22423529	G>C	NONE,UBE2E2	4.90%
Pam03	chr22q11.22	23243367	T>C	abParts	5.60%
Pam03	chr16p12.2	23847616	C>T	PRKCB	12.70%
Pam03	chr14q11.2	24046455	G>A	JPH4	15.60%
Pam03	chr22q11.23	24560377	C>T	CABIN1	12.60%
Pam03	chr22q11.23	25016911	C>T	GGT1	1.80%
Pam03	chr12p12.1	25380275	T>A	KRAS	26.40%
Pam03	chr13q12.13	25671870	G>C	PABPC3	2.20%
Pam03	chr13q12.13	25671892	G>T	PABPC3	2.10%
Pam03	chr6p22.2	26158661	G>A	HIST1H2BD	13.40%
Pam03	chrXp21.3	26212351	G>A	MAGEB6	28.00%
Pam03	chr6p22.2	26409862	C>A	BTN3A1	12.70%
Pam03	chr18q12.1	29992995	G>T	FAM59A	6.20%
Pam03	chr20q11.21	30496504	A>T	TTLL9	13.40%
Pam03	chr10p11.22	33559603	C>G	NRP1	2.40%
Pam03	chr15q14	34331031	G>A	AVEN	20.60%
Pam03	chr10p11.21	37438758	T>A	ANKRD30A	16.00%
Pam03	chr22q13.1	38037409	G>A	SH3BP1	14.20%
Pam03	chr4p14	39219674	C>T	WDR19	15.00%
Pam03	chr19q13.2	39805646	C>T	LRFN1	12.90%
Pam03	chr19q13.2	39908707	C>T	PLEKHG2	50.30%
Pam03	chr20q12	40979337	G>T	PTPRT	1.00%
Pam03	chr8p11.1	43147861	C>T	POTEA	13.60%

Pam03	chr19q13.2	43234183	G>T	PSG3	2.00%
Pam03	chr5p12	43536951	G>C	PAIP1	17.60%
Pam03	chr20q13.12	43727080	G>C	KCNS1	14.90%
Pam03	chr19q13.31	44892228	G>C	ZNF285	2.70%
Pam03	chr21q22.3	45077962	C>T	HSF2BP	11.60%
Pam03	chr19q13.31	45117256	G>A	IGSF23	24.50%
Pam03	chr14q21.2	45579854	G>T	PRPF39	17.00%
Pam03	chr21q22.3	47613623	C>T	LSS	8.20%
Pam03	chr7p12.3	48086062	G>A	C7orf57	12.40%
Pam03	chrXp11.23	48564780	T>G	SUV39H1	18.20%
Pam03	chrXp11.23	49062094	G>A	CACNA1F	28.20%
Pam03	chr8q11.21	49105340	A>G	UBE2V2,EFCAB1	64.30%
Pam03	chr2p16.3	49295429	C>A	FSHR	19.50%
Pam03	chr6p12.3	49456158	GAA>-	CENPQ	10.60%
Pam03	chr1p33	50610767	G>A	ELAVL4	12.70%
Pam03	chr10q11.23	50819316	C>T	SLC18A3	3.30%
Pam03	chr19q13.33	51128445	C>T	SYT3	29.00%
Pam03	chr19q13.33	51161322	G>A	C19orf81	9.90%
Pam03	chr3p21.1	52522145	C>T	NISCH	17.20%
Pam03	chr1p32.3	54605320	->T	CDCP2	0.10%
Pam03	chr11q12.1	56058201	A>C	OR8H1	10.80%
Pam03	chr12q14.1	58131504	G>A	AGAP2	0.10%
Pam03	chr14q23.1	61285437	C>-	MNAT1	20.50%
Pam03	chr11q13.2	66030477	C>T	KLC2	15.40%
Pam03	chr11q13.2	67177043	G>A	TBC1D10C	16.50%
Pam03	chr7q11.22	70597468	G>C	WBSCR17	9.90%
Pam03	chr16q22.1	70604043	G>A	SF3B3	13.80%
Pam03	chr16q22.2	72094691	T>C	HP	2.40%
Pam03	chr14q24.3	75563830	G>A	NEK9	15.30%
Pam03	chr6q14.3	86282022	C>T	SNX14	21.10%
Pam03	chr2p11.2	88124772	C>A	RGPD1,RGPD2	2.40%
Pam03	chr11q21	94310464	->AA	PIWIL4	12.70%
Pam03	chr11q22.3	108192148	G>A	ATM	20.90%
Pam03	chr5q21.3	108387404	G>A	FER	13.50%
Pam03	chr12q24.11	110231805	G>A	TRPV4	18.30%
Pam03	chr7q31.1	110526731	C>T	IMMP2L	17.00%
Pam03	chr11q23.3	116629090	T>C	BUD13	13.50%
Pam03	chr12q24.23	119568591	G>A	SRRM4	16.40%
Pam03	chr1p12	120277295	G>A	PHGDH	3.90%
Pam03	chr8q24.22	131792884	G>T	ADCY8	20.00%
Pam03	chr3q22.1	133119002	G>A	BFSP2	15.60%



Pam03	chr8q24.22	133634926	->T	LRRC6	2.90%
Pam03	chr9q34.3	140652467	A>C	EHMT1	10.80%
Pam03	chr5q31.3	140802733	G>A	PCDHGA11	15.20%
Pam03	chr5q31.3	143853421	G>A	KCTD16	11.10%
Pam03	chr1q21.3	154728481	C>G	KCNN3	18.20%
Pam03	chr5q35.1	169135251	C>T	DOCK2	15.70%
Pam03	chr2q31.1	176972815	C>T	HOXD11	17.80%
Pam03	chr3q26.33	179094854	A>C	MFN1	19.40%
Pam03	chr2q32.1	188293620	A>-	CALCRL	33.30%
Pam03	chr1q31.3	196759282	C>T	CFHR3	9.50%
Pam03	chr1q32.1	201180309	A>G	IGFN1	0.90%
Pam03	chr2q33.2	203927006	C>T	NBEAL1	12.80%
Pam03	chr2q35	219508759	A>T	ZNF142	0.90%
Pam03	chr2q37.1	231115733	A>C	SP140	13.70%
Pam03	chr1q42.2	233497916	C>T	KIAA1804	10.70%
Pam04	chr9p24.3	11071	->G	DDX11L5	20.00%
Pam04	chr5p15.33	236628	C>T	SDHA	3.00%
Pam04	chr4p16.3	779798	C>G	CPLX1	16.30%
Pam04	chr11p15.5	1018091	T>G	MUC6	2.20%
Pam04	chr11p15.5	1018095	G>A	MUC6	1.20%
Pam04	chr11p15.5	1018340	T>C	MUC6	3.30%
Pam04	chr11p15.5	1018341	G>A	MUC6	3.00%
Pam04	chr7p22.3	1050850	A>C	C7orf50	7.00%
Pam04	chr11p15.5	1093549	G>C	MUC2	5.00%
Pam04	chr5p15.33	1096836	A>G	SLC12A7	100.00%
Pam04	chr11p15.5	1284342	G>A	MUC5B	8.00%
Pam04	chr11p15.5	1619670	T>C	LOC338651	0.50%
Pam04	chr18p11.32	2697888	C>T	SMCHD1	7.50%
Pam04	chr4p16.3	3076699	T>G	HTT	6.90%
Pam04	chr4p16.3	3076704	T>C	HTT	4.20%
Pam04	chr16p13.3	3104565	A>G	BC045731	1.90%
Pam04	chr10p15.2	3208559	T>A	PITRM1	3.00%
Pam04	chr5p15.33	3653623	G>A	IRX1,BC034630	27.20%
Pam04	chr11p15.4	5153670	A>C	OR52A5	1.70%
Pam04	chr1p36.32	5306750	G>T	BC037321,AK125078	1.20%
Pam04	chr1p36.32	5306751	T>C	BC037321,AK125078	1.60%
Pam04	chr12p13.31	6128067	G>A	VWF	2.00%
Pam04	chr12p13.31	6128076	A>G	VWF	1.50%
Pam04	chr17p13.2	6329091	C>T	AIPL1	0.10%
Pam04	chr11p15.4	6478026	G>C	TRIM3	8.40%
Pam04	chr11p15.4	6585300	G>C	DNHD1	0.70%

Pam04	chr7p22.1	6906949	C>A	CCZ1B,LOC100131257	0.80%
Pam04	chr17p13.1	7123794	C>A	ACADVL	8.10%
Pam04	chrXp22.31	8434497	A>G	VCX-8r	5.50%
Pam04	chrYp11.2	9383339	G>C	TSPY3,RBMY3AP	14.00%
Pam04	chr18p11.22	10731453	T>-	PIEZO2	0.30%
Pam04	chr8p23.1	11189485	G>T	SLC35G5	3.10%
Pam04	chr12p13.2	11214001	C>G	TAS2R46	2.60%
Pam04	chr12p13.2	11214005	C>T	TAS2R46	2.60%
Pam04	chr12p13.2	11506303	G>T	PRB1	2.40%
Pam04	chr1p36.21	12907701	G>T	HNRNPCL1,LOC649330	2.80%
Pam04	chr1p36.21	12907702	A>G	HNRNPCL1,LOC649330	2.80%
Pam04	chr1p36.21	12907705	T>C	HNRNPCL1,LOC649330	2.90%
Pam04	chr1p36.21	12907708	A>G	HNRNPCL1,LOC649330	3.20%
Pam04	chr3p25.2	12942848	T>-	IQSEC1	5.60%
Pam04	chr12p13.1	13153397	G>C	AK125676,HTR7P1	1.70%
Pam04	chr6p23	13618360	A>G	NOL7	10.50%
Pam04	chr21q11.2	15287733	A>G	DQ579969,ANKRD20A11P	7.90%
Pam04	chr17p11.2	16097809	G>T	NCOR1	0.40%
Pam04	chr6p22.3	17602839	C>T	FAM8A1	1.40%
Pam04	chr15q11.1	20093019	T>C	NONE,DQ576041	4.40%
Pam04	chr15q11.1	20093068	C>T	NONE,DQ576041	3.60%
Pam04	chr15q11.1	20093072	T>C	NONE,DQ576041	3.80%
Pam04	chr15q11.1	20093075	G>C	NONE,DQ576041	3.70%
Pam04	chr22q11.21	20342225	A>T	LOC729444	1.60%
Pam04	chr14q11.2	20404321	C>A	OR4K1	3.60%
Pam04	chr22q11.21	20710577	C>T	USP41	0.40%
Pam04	chr22q11.21	20714957	G>A	USP41	5.70%
Pam04	chr22q11.21	20714962	G>A	USP41	5.30%
Pam04	chr1p36.12	22333374	T>C	CELA3A	1.60%
Pam04	chr19p12	22363736	T>A	ZNF676	1.30%
Pam04	chr3p24.3	22423506	C>T	NONE,UBE2E2	1.90%
Pam04	chr3p24.3	22423518	G>A	NONE,UBE2E2	1.80%
Pam04	chr15q11.2	22473194	G>C	abParts	3.00%
Pam04	chr22q11.22	22988897	G>A	GGTLC2	2.80%
Pam04	chr12p12.1	25398284	C>T	KRAS	9.50%
Pam04	chr13q12.13	25671962	C>T	PABPC3	1.10%
Pam04	chr13q12.13	25671967	G>T	PABPC3	1.10%
Pam04	chr12p11.21	31250830	C>G	DDX11	1.70%
Pam04	chr6p21.33	31324431	->GGG	HLA-B;HLA-B	1.40%
Pam04	chr6p21.32	32557503	A>G	HLA-DRB1,HLA-DRB5	5.30%
Pam04	chr18q12.2	32843207	A>-	ZSCAN30	8.50%

Pam04	chr16p11.2	33020686	C>A	IGH	3.00%
Pam04	chr15q14	34518040	A>T	TMEM85	7.70%
Pam04	chr21q22.12	36042492	A>G	CLIC6	11.10%
Pam04	chr22q12.3	36205978	G>A	RBFOX2	5.40%
Pam04	chr10p11.21	37433975	T>G	ANKRD30A	1.80%
Pam04	chr10p11.21	37433982	C>G	ANKRD30A	1.70%
Pam04	chr7p14.1	38388777	G>A	Y00482	3.70%
Pam04	chr6p21.2	38455405	A>G	BTBD9	6.70%
Pam04	chr6p21.2	38455406	G>A	BTBD9	6.10%
Pam04	chr10p11.1	38740275	T>C	LOC399744	3.30%
Pam04	chr10p11.1	38874587	A>T	LOC399744,ACTR3BP5	3.10%
Pam04	chr9p13.1	39078843	A>G	CNTNAP3	2.10%
Pam04	chr17q21.2	39197549	G>C	KRTAP1-1	1.70%
Pam04	chr22q13.1	40366977	C>T	GRAP2	8.20%
Pam04	chr17q21.2	40764511	C>T	TUBG1	6.80%
Pam04	chr9p13.1	40772275	T>C	ZNF658	1.80%
Pam04	chr10q11.21	42368380	C>T	NONE,LOC441666	14.30%
Pam04	chr19q13.2	43237217	G>T	PSG1,PSG3	0.90%
Pam04	chr19q13.2	43262428	C>T	PSG8	1.90%
Pam04	chr10q11.21	43315730	A>C	BMS1	1.90%
Pam04	chr10q11.21	43315731	G>T	BMS1	1.80%
Pam04	chr20q13.12	44477208	G>A	ACOT8	0.80%
Pam04	chr20q13.12	45014770	G>A	ELMO2	1.40%
Pam04	chr15q21.1	45561643	G>A	SLC28A2	7.50%
Pam04	chr15q21.1	45658258	G>A	GATM	8.80%
Pam04	chr21q22.3	47612603	G>A	LSS	1.40%
Pam04	chr16q12.1	48201472	C>T	ABCC11	9.50%
Pam04	chr18q21.2	48591889	A>G	SMAD4	11.10%
Pam04	chr15q21.1	48736758	A>T	FBN1	9.10%
Pam04	chr4p11	49104722	G>C	CWH43,DQ579288	66.70%
Pam04	chr3p21.31	49722763	C>A	MST1	1.30%
Pam04	chr3p21.31	49723545	G>A	MST1	1.90%
Pam04	chr3p21.31	49725326	C>T	MST1	2.50%
Pam04	chr11p11.12	49878846	T>C	LOC440040,OR4C13	1.40%
Pam04	chr14q21.3	50052749	C>T	RPS29	1.50%
Pam04	chr3p21.31	50253204	G>C	SLC38A3	5.60%
Pam04	chr7p12.1	52204744	A>G	DQ599872,POM121L12	20.00%
Pam04	chr11q12.1	56143198	C>T	OR8U8	3.60%
Pam04	chr7p11.2	56183768	C>T	LOC389493,PSPH	8.20%
Pam04	chr4q12	57777455	C>T	REST	0.60%
Pam04	chr19q13.43	58724101	C>A	ZNF274	1.10%

Pam04	chr11q12.1	59022159	TT>-	MPEG1,OR5AN1	7.70%
Pam04	chr11q13.1	65113214	G>A	DPF2	0.80%
Pam04	chrXq12	67331786	T>C	OPHN1	5.40%
Pam04	chr4q13.2	69681869	A>G	UGT2B10	0.40%
Pam04	chr16q22.1	70048023	T>C	CLEC18A	1.70%
Pam04	chr14q24.2	70924812	C>T	ADAM21	2.00%
Pam04	chr11q13.4	71238711	G>C	KRTAP5-7	1.70%
Pam04	chr2p13.3	71295708	G>A	NAGK	9.00%
Pam04	chr16q22.2	71319430	C>G	FTSJD1	9.60%
Pam04	chr9q21.11	71509492	->T	PIP5K1B	8.50%
Pam04	chr6q13	72892049	G>A	RIMS1	9.30%
Pam04	chr11q13.4	73371897	G>A	PLEKHB1	4.70%
Pam04	chr14q24.3	75199476	C>G	FCF1	7.10%
Pam04	chr3p12.3	75714974	C>A	FRG2C	2.60%
Pam04	chr3p12.3	75790860	C>G	ZNF717	1.90%
Pam04	chr16q23.1	76587326	G>A	CNTNAP4	8.00%
Pam04	chr11q14.1	77813998	AA>-	ALG8	14.00%
Pam04	chr12q21.2	78415548	T>C	NAV3	10.70%
Pam04	chr2p12	79349987	C>G	REG1A	7.90%
Pam04	chr2p12	80085187	C>T	CTNNA2	9.60%
Pam04	chr8q21.13	81471795	A>T	ZBTB10,ZNF704	0.70%
Pam04	chr14q31.1	81943409	C>A	SEL1L	0.10%
Pam04	chr16q23.3	83984147	C>T	OSGIN1	1.60%
Pam04	chr16q24.3	88996549	G>A	CBFA2T3	1.30%
Pam04	chr16q24.3	88996570	A>G	CBFA2T3	1.00%
Pam04	chr16q24.3	89016799	G>A	CBFA2T3	0.50%
Pam04	chr16q24.3	89017404	A>G	CBFA2T3	1.40%
Pam04	chr2p11.2	89326842	T>C	abParts	5.10%
Pam04	chr2p11.2	89417043	T>C	abParts	7.60%
Pam04	chr1p22.2	89448740	G>T	RBMXL1	2.10%
Pam04	chr1p22.2	89448963	G>C	RBMXL1	3.60%
Pam04	chr11q14.3	89595633	G>A	LOC399940	0.20%
Pam04	chr16q24.3	90086148	C>A	GAS8	8.40%
Pam04	chr15q26.1	90904016	C>G	ZNF774	8.70%
Pam04	chr15q26.1	92292700	A>T	TRNA	8.10%
Pam04	chr11q21	94040769	G>A	FOLR4	8.70%
Pam04	chr1p22.1	94476852	C>T	ABCA4	8.40%
Pam04	chr2q11.1	96617111	G>A	DKFZp667P0924,LOC729234	0.90%
Pam04	chr2q11.1	96628157	C>T	DKFZp667P0924,LOC729234	1.40%
Pam04	chr2q11.1	96628162	C>T	DKFZp667P0924,LOC729234	0.80%
Pam04	chr14q32.2	96993782	C>T	PAPOLA	7.60%

Pam04	chr2q11.2	97829023	T>A	ANKRD36	2.20%
Pam04	chrXq22.1	99921839	C>G	SRPX2	15.90%
Pam04	chrXq22.1	100750363	T>G	ARMCX4	16.90%
Pam04	chr3q12.3	101133344	C>A	SENP7	1.50%
Pam04	chr3q12.3	101133383	A>T	SENP7	1.80%
Pam04	chr3q12.3	101133384	C>T	SENP7	1.90%
Pam04	chr3q12.3	101133387	T>C	SENP7	2.00%
Pam04	chr3q12.3	101133420	C>T	SENP7	1.80%
Pam04	chr8q22.3	101717167	G>A	PABPC1	1.00%
Pam04	chr7q22.1	102898150	G>A	DPY19L2P2	1.10%
Pam04	chr12q23.2	103688287	A>G	ASCL1,C12orf42	75.00%
Pam04	chr4q24	103832643	A>C	SLC9B1	7.30%
Pam04	chr4q24	103832656	C>T	SLC9B1	7.50%
Pam04	chr4q24	103866398	A>G	SLC9B1	0.50%
Pam04	chr5q21.2	103978845	A>G	NUDT12,RAB9BP1	4.90%
Pam04	chr14q32.33	106235742	G>C	FLJ00385,abParts	1.30%
Pam04	chr7q22.3	106271741	G>T	AF086203	0.70%
Pam04	chr7q22.3	106272530	T>A	AF086203	3.30%
Pam04	chr2q12.2	107123447	C>T	RGPD3,ST6GAL2	8.20%
Pam04	chr11q22.3	108122664	T>G	ATM	9.40%
Pam04	chr6q21	109767629	C>A	MICAL1	5.60%
Pam04	chr6q21	109996877	A>T	AKD1	1.20%
Pam04	chrXq24	118604922	A>G	SLC25A5	0.90%
Pam04	chrXq24	118604935	T>C	SLC25A5	0.80%
Pam04	chrXq24	118604954	A>G	SLC25A5	0.90%
Pam04	chrXq24	118604969	T>C	SLC25A5	0.80%
Pam04	chrXq24	118604970	G>A	SLC25A5	0.90%
Pam04	chrXq24	118605023	T>G	SLC25A5	0.90%
Pam04	chr7q31.31	119915394	C>T	KCND2	8.20%
Pam04	chr2q14.2	122368261	A>G	CLASP1	5.60%
Pam04	chr11q24.2	123909441	T>C	OR10G7	1.10%
Pam04	chr11q24.2	124121285	T>A	OR8G1	50.00%
Pam04	chr7q32.1	128586018	C>T	IRF5	9.20%
Pam04	chr5q23.3	128797331	A>G	ADAMTS19	2.90%
Pam04	chr5q23.3	128797332	C>G	ADAMTS19	2.90%
Pam04	chr6q22.33	129513873	C>A	LAMA2	9.10%
Pam04	chr2q21.2	133014614	C>T	MIR663B	2.20%
Pam04	chr8q24.22	133088758	G>A	HHLA1	8.20%
Pam04	chr5q31.1	134263939	C>G	PCBD2	3.00%
Pam04	chr9q34.3	137653807	G>A	COL5A1	7.80%
Pam04	chr9q34.3	138586898	G>A	SOHLH1	9.30%

Pam04	chr9q34.3	139565113	T>A	MIR126	7.60%
Pam04	chr9q34.3	139997597	->ACAC	MAN1B1	3.30%
Pam04	chr9q34.3	140040250	C>T	GRIN1	8.00%
Pam04	chr9q34.3	141069495	C>T	TUBBP5	2.10%
Pam04	chr7q34	142149017	G>C	TCRBV5S3A2T	1.10%
Pam04	chr1q21.1	144621649	C>G	NBPF9	2.10%
Pam04	chr8q24.3	145617547	T>A	ADCK5	30.80%
Pam04	chr8q24.3	145617788	G>A	ADCK5	6.80%
Pam04	chr7q35	147075157	T>A	MIR548F4	15.00%
Pam04	chr1q21.2	148594474	C>T	NBPF15	1.40%
Pam04	chr5q33.1	150901466	G>A	FAT2	8.20%
Pam04	chr7q36.1	151904508	C>A	MLL3	8.40%
Pam04	chr1q21.3	152187827	C>G	HRNR	2.00%
Pam04	chr1q21.3	152785170	G>A	LCE1B	8.00%
Pam04	chr5q33.3	158621749	G>A	RNF145	1.10%
Pam04	chr5q33.3	158621792	C>T	RNF145	1.20%
Pam04	chr1q23.1	158687165	G>A	OR6K3	8.40%
Pam04	chr5q33.3	159781949	G>A	C1QTNF2	8.30%
Pam04	chr6q25.3	160132524	T>C	SOD2,WTAP	1.00%
Pam04	chr1q23.3	161596014	A>G	FCGR3B	2.70%
Pam04	chr2q24.3	167085261	C>T	SCN9A	8.10%
Pam04	chr1q24.2	168550338	A>G	XCL1	0.40%
Pam04	chr1q25.2	176668298	C>G	PAPPA2	7.10%
Pam04	chr5q35.3	177378844	C>T	AK126616	1.60%
Pam04	chr5q35.3	177378852	T>C	AK126616	1.00%
Pam04	chr5q35.3	177378857	C>T	AK126616	0.80%
Pam04	chr1q25.3	182354987	A>G	GLUL	8.10%
Pam04	chr1q31.1	186277277	G>A	PRG4	0.80%
Pam04	chr1q31.1	186368339	->T	MIR548F1	8.30%
Pam04	chr4q35.2	187153390	G>A	KLKB1	8.90%
Pam04	chr4q35.2	190426627	A>T	HSP90AA4P,BC087857	1.70%
Pam04	chr4q35.2	190884276	T>A	FRG1	0.80%
Pam04	chr1q32.1	201180054	G>C	IGFN1	3.00%
Pam04	chr1q32.1	201180064	G>A	IGFN1	2.60%
Pam04	chr2q33.3	207653606	C>G	FASTKD2	8.50%
Pam04	chr1q41	216737607	G>C	ESRRG	0.90%
Pam04	chr1q42.11	224216370	->T	AK124970,7SK	1.20%
Pam04	chr1q42.13	230451783	A>T	GALNT2,PGBD5	15.80%
Pam04	chr1q42.13	230451832	A>G	GALNT2,PGBD5	11.10%
Pam04	chr1q42.13	230451834	A>G	GALNT2,PGBD5	13.30%
Pam04	chr1q42.13	230451835	C>T	GALNT2,PGBD5	15.80%

Pam04	chr2q37.3	240982129	G>C	PRR21	5.60%
Pam04	chr2q37.3	241622103	A>C	AQP12B	0.60%

<sup>a</sup>a dash preceding the “>” indicates an insertion of the new base(s), a dash following indicates a deletion

<sup>b</sup>the median mutant allele frequency as calculated by the ratio of mutant reads over total reads

**Table 3-6. Major driver gene mutations identified in each patient.**

Case	Drivers	Chr.	Genomic Position <sup>a</sup>	Base Change	Amino Acid change	Effect <sup>a</sup>	Median MAF <sup>b</sup>
Pam01	KRAS	12	25398284	C>A	G12V	n.s.	65.7%
Pam02	ARID1A	1	27058029	T>G	Y579X	stop	30.2%
	KRAS	12	25398284	C>A	G12V	n.s.	56.8%
	TP53	17	7573996	A>G	L305P	n.s.	30.3%
Pam03	KRAS	12	25380275	T>A	Q61H	n.s.	28.0%
Pam04	ATM	11	108122664	T>G	F570V	n.s.	9.4%
	KRAS	12	25398284	C>T	G12D	n.s.	10.6%
	SMAD4	18	48591889	A>G	D351G	n.s.	11.1%

n.s. → nonsynonymous

<sup>a</sup>based on the UCSC Genome Browser hg19 assembly

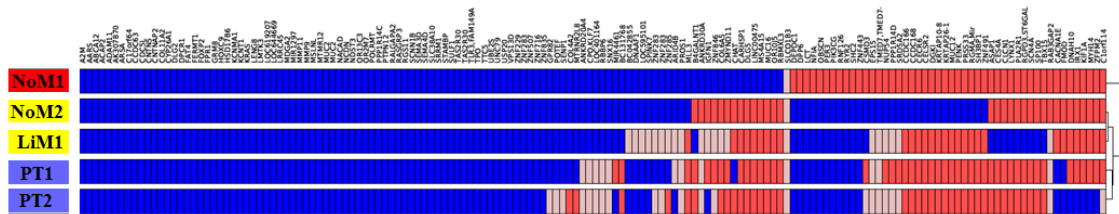
<sup>b</sup>MAF: mutant allele fraction determined from all samples of each indicated patient.



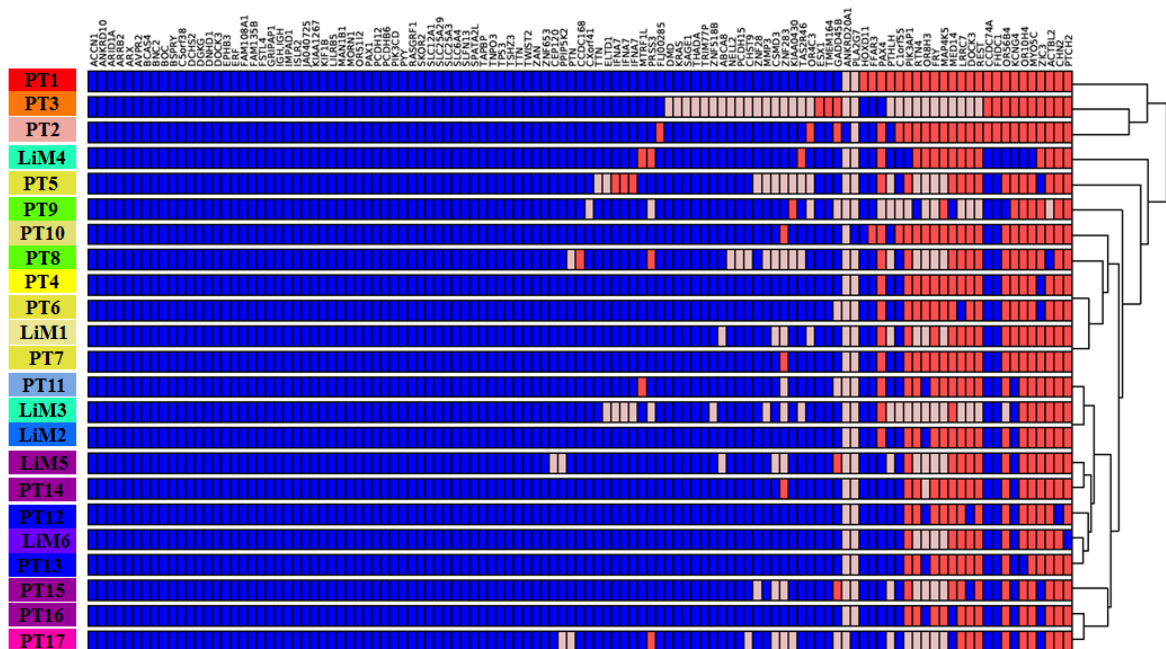
Employing this set of 614 bona fide mutations, we found that a large fraction was shared among all samples from the same patient (Figures 3-3, 3-4, 3-5, and 3-6). Unique somatic mutations ("private") within different metastases were nonetheless identified. To assess whether these private mutations could be identified in the primary tumors, we evaluated the same set of 614 mutations in an additional 46 different regions of the same primary pancreatic tumors (Tables 3-4 and 3-5). The average depth of sequencing of these samples was 684X and the average fraction of bases covered with >10 distinct reads was 91.0% (Table 3-6).

We quantified the genetic differences among metastases using a relatedness index based on the fraction of shared mutations between any two metastatic tumor cells within a patient. Our approach was similar to a Jaccard index, a measure of genetic similarity based on the ratio of discordant variants over all (shared plus discordant) variants between two data sets (Table 3-7 and Table 3-8). This calculation provided a measure of the relatedness of any two tumors that was employed to estimate the overall diversity of the metastases. For example, two metastatic lesions from the same patient sharing zero mutations would result in an index of zero while two lesions with completely identical mutations would lead to an index of 1.

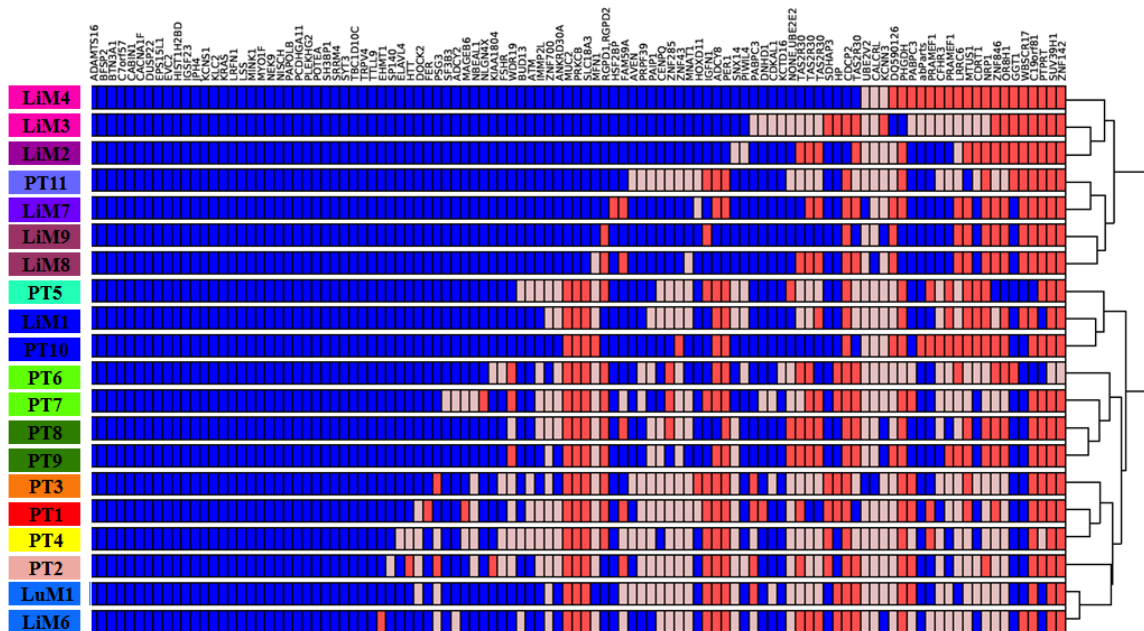
To contextualize the tumor data, we modeled the somatic evolution of three normal tissues by a stochastic evolutionary process with differing renewal patterns (Figure 3-7). In the simplest case, we modeled an organ that grows via a pure birth branching process to  $N_{cell}$  cells with no further cell divisions (modeling neuronal cells, for example). For  $N_{cell} = 10^{10}$ , the expected relatedness index was  $\sim 0.03$ . For a more complex case, we derived the index for an organ with  $N_{crypt}$  cells, where each



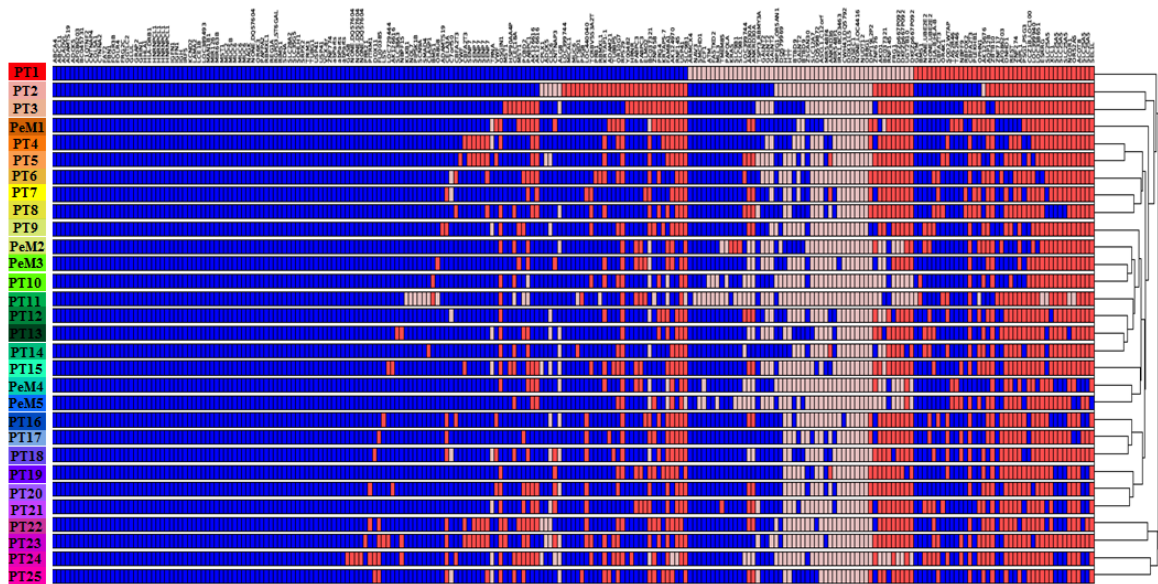
**Figure 3-3. Hierarchical clustering for Pam01.** Samples are indicated along the y-axis and variants are listed along the top of the diagram. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. The variant status in each sample is shown in blue (for present), dark red (for negative), or light red (for unknown). Sample clustering is indicated on the right y-axis. Primary tumors are labeled at “PT” followed by a number, the remaining samples are two lymph node mets (labeled NoM followed by a number) and one liver met labeled as LiM and also followed by a number.



**Figure 3-4. Hierarchical clustering for Pam02.** Samples are indicated along the y-axis and variants are listed along the top of the diagram. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. The variant status in each sample is shown in blue (for present), dark red (for negative), or light red (for unknown). Sample clustering is indicated on the right y-axis. Primary tumors are labeled at “PT” followed by a number, the remaining samples are liver mets labeled as LiM and also followed by a number.



**Figure 3-5. Hierarchical clustering for Pam03.** Samples are indicated along the y-axis and variants are listed along the top of the diagram. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. The variant status in each sample is shown in blue (for present), dark red (for negative), or light red (for unknown). Sample clustering is indicated on the right y-axis. Primary tumors are labeled at “PT” followed by a number, the remaining samples are liver mets labeled as LiM (also followed by a number) and one lung met followed by a number.



**Figure 3-6. Hierarchical clustering for Pam04.** Samples are indicated along the y-axis and variants are listed along the top of the diagram. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. The variant status in each sample is shown in blue (for present), dark red (for negative), or light red (for unknown). Sample clustering is indicated on the right y-axis. Primary tumors are labeled at “PT” followed by a number, the remaining samples are peritoneal mets labeled as PeM and also followed by a number.

**Table 3-7. Relatedness indices of metastases.**

Case		LiM1	LiM2	NoM1	NoM2						Median
<b>Pam01</b>	<i>LiM1</i>	1.00									0.75
	<i>LiM2</i>	0.77	1.00								
	<i>NoM1</i>	0.74	0.66	1.00							
	<i>NoM2</i>	0.80	0.75	0.68	1.00						
<b>Pam02</b>		<b>LiM1</b>	<b>LiM2</b>	<b>LiM3</b>	<b>LiM4</b>	<b>LiM5</b>	<b>LiM6</b>	<b>LiM7</b>	<b>LiM8</b>		0.97
	<i>LiM1</i>	1.00									
	<i>LiM2</i>	0.99	1.00								
	<i>LiM3</i>	0.99	1.00	1.00							
	<i>LiM4</i>	0.91	0.92	0.96	1.00						
	<i>LiM5</i>	0.97	0.98	0.97	0.90	1.00					
	<i>LiM6</i>	0.97	0.98	0.98	0.91	0.98	1.00				
	<i>LiM7</i>	0.95	0.97	0.97	0.94	0.97	0.97	1.00			
<i>LiM8</i>	0.97	0.95	0.99	0.92	0.96	0.96	0.97	1.00			
<b>Pam03</b>		<b>LuM1</b>	<b>LuM2</b>	<b>LuM3</b>	<b>LiM1</b>	<b>LiM2</b>	<b>LiM3</b>	<b>LiM4</b>	<b>LiM5</b>		0.88
	<i>LuM1</i>	1.00									
	<i>LuM2</i>	0.95	1.00								
	<i>LuM3</i>	0.90	0.93	1.00							
	<i>LiM1</i>	0.92	0.91	0.89	1.00						
	<i>LiM2</i>	0.81	0.86	0.80	0.88	1.00					
<i>LiM3</i>	0.83	0.83	0.81	0.85	0.95	1.00					

<i>LiM4</i>	0.78	0.86	0.81	0.84	0.90	0.92	1.00	
<i>LiM5</i>	0.86	0.87	0.84	0.88	0.95	0.97	0.98	1.00

	<b>PeM1</b>	<b>PeM2</b>	<b>PeM3</b>	<b>PeM4</b>	<b>PeM5</b>	<b>PeM6</b>	
<b>Pam04</b>   <i>PeM1</i>	1.00						0.84
<i>PeM2</i>	0.8	1.00					
<i>PeM3</i>	0.84	0.89	1.00				
<i>PeM4</i>	0.82	0.85	0.89	1.00			
<i>PeM5</i>	0.85	0.86	0.87	0.88	1.00		
<i>PeM6</i>	0.76	0.84	0.82	0.83	0.82	1.00	

**Table 3-8. Genetic distances among metastases.**

<b>A. Pam01</b>	LiM1	LiM2	NoM1	NoM2
LiM1	0			
LiM2	25	0		
NoM1	30	39	0	
NoM2	24	29	44	0

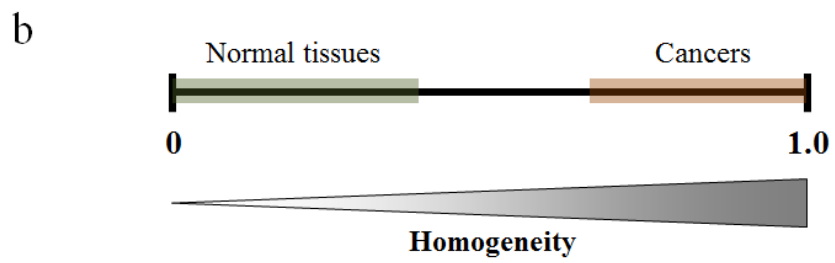
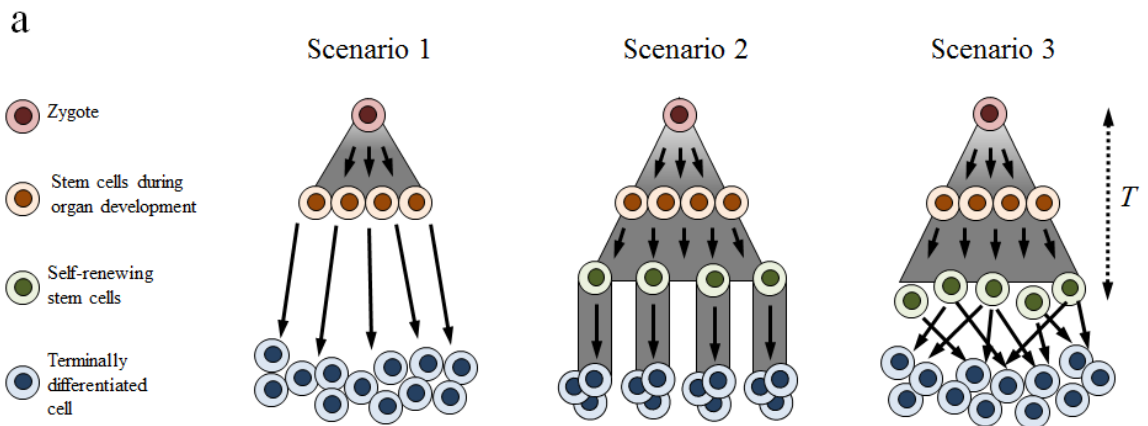
<b>B. Pam02</b>	LiM1	LiM2	LiM3	LiM4	LiM5	LiM6	LiM7	LiM8
LiM1	0							
LiM2	1	0						
LiM3	1	0	0					
LiM4	8	8	3	0				
LiM5	3	2	2	9	0			
LiM6	3	2	2	9	2	0		
LiM7	3	2	2	4	2	2	0	
LiM8	2	4	1	6	3	3	2	0

<b>C. Pam03</b>	LuM1	LuM2	LuM3	LiM1	LiM2	LiM3	LiM4	LiM5
LuM1	0							
LuM2	2	0						
LuM3	5	3	0					
LiM1	5	4	6	0				
LiM2	13	7	12	9	0			
LiM3	11	8	10	10	4	0		
LiM4	16	7	11	12	9	6	0	
LiM5	8	6	8	7	3	2	1	0

<b>D. Pam04</b>	PeM1	PeM2	PeM3	PeM4	PeM5	PeM6
PeM1	0					
PeM2	33	0				
PeM3	28	19	0			
PeM4	31	26	19	0		
PeM5	25	24	22	20	0	
PeM6	26	17	19	18	19	0



**Figure 3-7. Somatic evolution of normal tissues.** a. Three hypothetical scenarios for normal tissue somatic evolution are considered, T (far right) indicates time. In (1) the organ follows a pure birth process for development with no further cell division. In (2), the organ follows a pure birth process for development of stem cells, each founding a single crypt and dividing over time. The organ in (3) follows the same development as (2) but with substantial mixing and replacement of stem cells. b. The expected relatedness index, with 0 for no identical mutations and 1 for complete genetic identity among any two cell lineages. The index ranges for normal tissues and metastases are shown in green and red, respectively. The relatedness index among the stem-cell-like cells (orange cells in (1); green cells in (2) and (3)) was always below 0.39 in all three scenarios. Accounting for possibly additional mutations in short-lived terminally differentiated cells would further increase the heterogeneity within an organ.

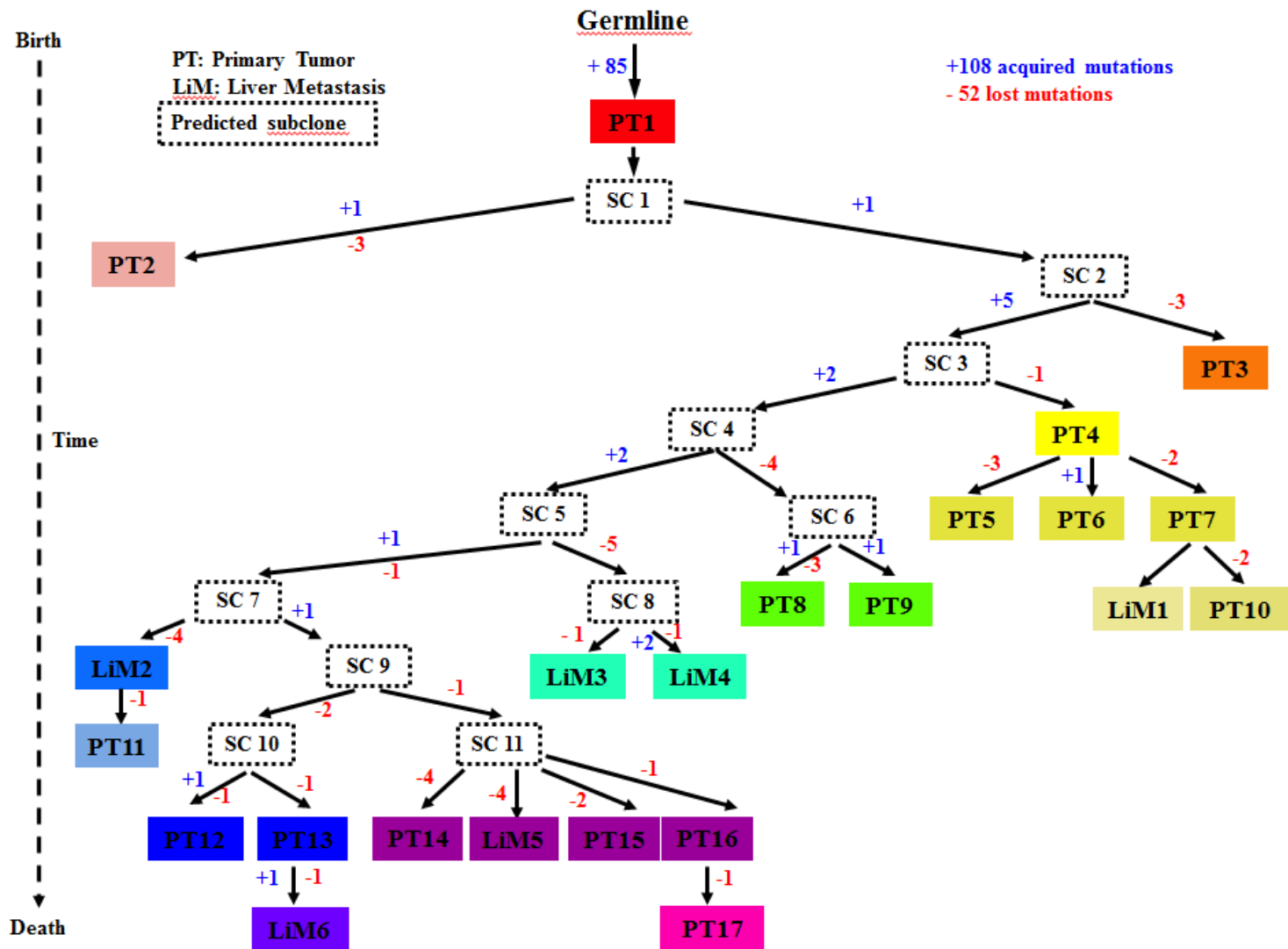


cell founded a single geographically isolated compartment (such as a crypt) that continuously replenishes itself, with no mixing or replacement (modeling intestinal epithelial cells). In this case, the index was less than 0.04 (assuming  $N_{crypt} = 10^7$  cells). Finally, we considered a tissue with  $N_{stem}$  stem cells that not only continuously divides but also exhibits high replacement and mixing (modeling hematopoietic cells). For this final scenario, the index was bounded at 0.39 for relevant time scales and diminished with the population size of an organ. Thus, for healthy tissues of various renewal types, the expected relatedness index ranged from near zero to  $\sim 0.39$  (Figure 3-7b).

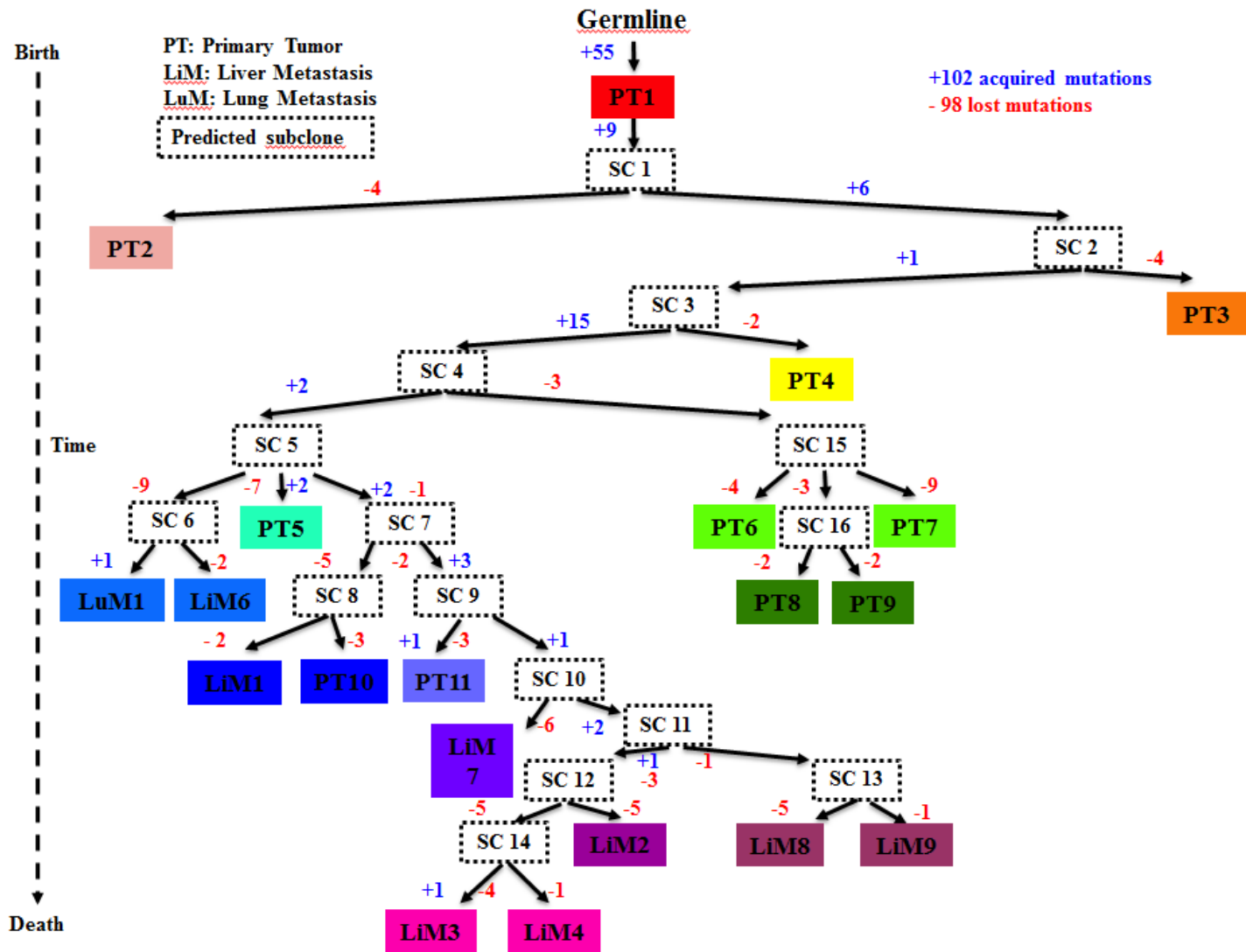
We then applied the relatedness index to the validated mutations detected in the various metastases from each patient. The relatedness indices averaged 0.91 ( $\pm 0.08$ ) (Figure 3-7b and Table 3-7). The relatedness of any metastatic lesion to any other metastatic lesion was far higher (minimum of 0.66) than of any modeled normal tissues (maximum of 0.39). Thus these data demonstrate that metastatic lesions have much higher homogeneity than expected in normal tissues ( $p < 0.002$ , Welch's *t*-test).

To determine whether this high similarity was consistent with the evolution of these tumors, we employed phylogenetic methods. The high similarity complicated the evolutionary reconstruction, so we used three approaches to derive phylogenetic trees. Details of these approaches are provided in the Methods. In general, these trees were consistent with the expected generation of metastases from primary tumors. In Pam02 and Pam03, we found that metastases within the liver were more likely to be derived from different subclones in the primary tumor than from each other (Figures 3-8 and 3-9). This suggested that there are a variety of subclones in the primary tumor that independently gave rise to highly related, but independent metastatic lesions;

**Figure 3-8. Neighbor-joining phylogeny of primary tumor sections and metastases for patient Pam02.** Time is represented on the left axis, divergence is represented on the horizontal axis. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. See Table 3-3 for sample identity. Primary tumors are labeled at “PT” followed by sequential numbers, the remaining samples are liver mets labeled as LiM and also followed by a sequential number. Hypothetical subclones are indicated by “SC” followed by the subclone number and enclosed by a dashed outline. The numbers of acquired mutations are in blue font with a “+” sign while the numbers of discontinuous (i.e. non-persistent) mutations are in red font with a “-“ sign.



**Figure 3-9. Neighbor-joining phylogeny of primary tumor sections and metastases for patient Pam03.** Time is represented on the left axis, divergence is represented on the x axis. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. See Table 3-3 for sample identity. Primary tumors are labeled at “PT” followed by a number, the remaining samples are mets labeled by organ. Hypothetical subclones are indicated by “SC” followed by the subclone number. The numbers of acquired mutations are in blue font with a “+” sign while the numbers of discontinuous (i.e. non-persistent) mutations are in red font with a “-“ sign.



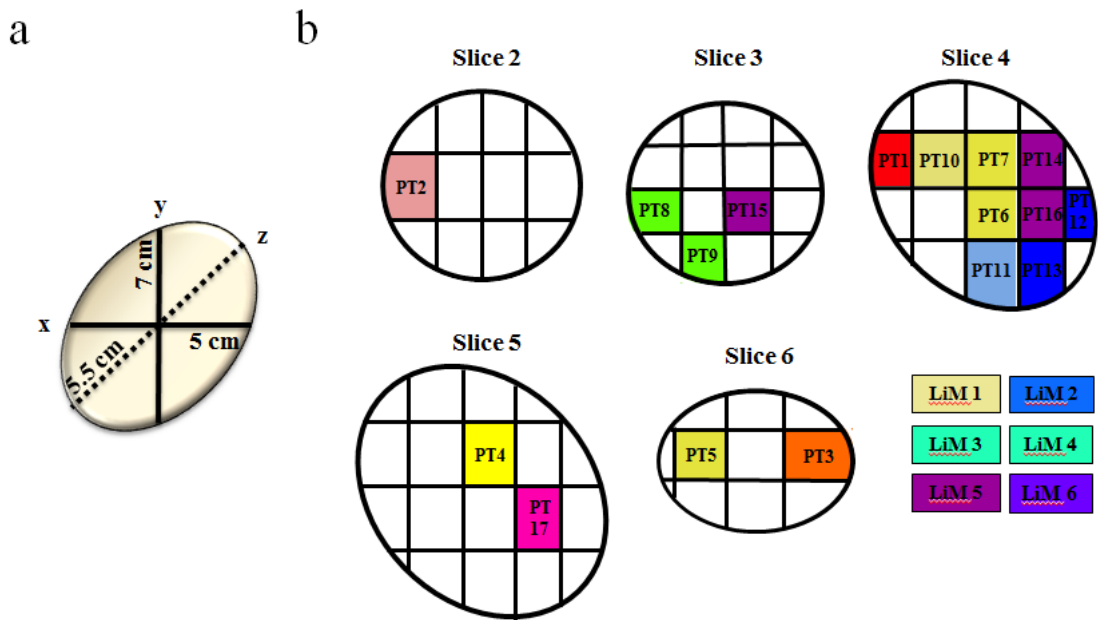
no single subclone within the primary tumors gave rise to all, or even most of, the metastases (Fig. 3-10 and 3-11). The phylogenetic analyses also showed that there were no definable geographic boundaries among the subclones in the primary tumor. Migration of subclones throughout the tumor was apparent (Figure 3-10, 3-11, 3-12). Though we could not exclude that metastases seed the primary tumor in a process reverse to the conventional one, the data from Pam 02 argue against this possibility in at least one patient: LiM1 was genetically identical to sample PT7, even though sample PT7 was located in the very center of the primary tumor.



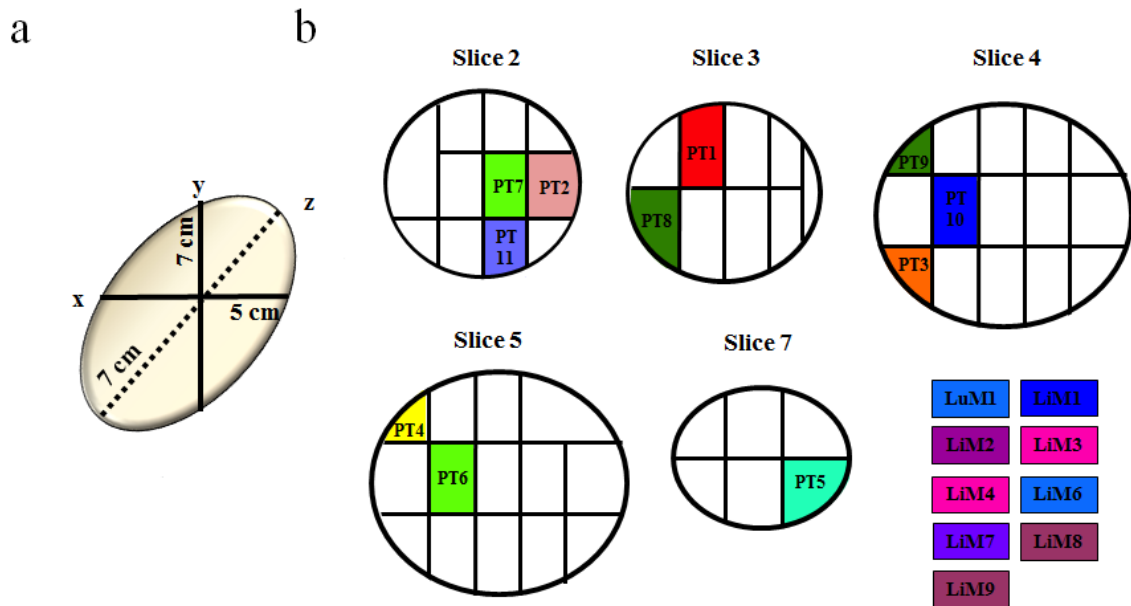
## Conclusions

These data demonstrate a high genetic relatedness of the cells that seeded each distinct metastasis in an individual patient. This relatedness determines whether all, or only a subset, of the metastases in the patient are likely to respond to a therapy targeted to the cancer genome. The high relatedness observed in our study is the direct result of the genetic bottlenecks that occurred throughout the long evolutionary history of the cancers, up to and including the final genetic events required for metastases. We did not attempt to interrogate the relatedness of cells *within* any individual metastatic lesion. This sort of relatedness does not determine the proportion of metastatic lesions that will initially respond to therapy, i.e., whether the patient has the opportunity to respond in a clinically meaningful way. However, if all lesions respond, the intercellular heterogeneity within each metastatic lesion has a major impact on whether a treated lesion will eventually recur<sup>76,97</sup>.

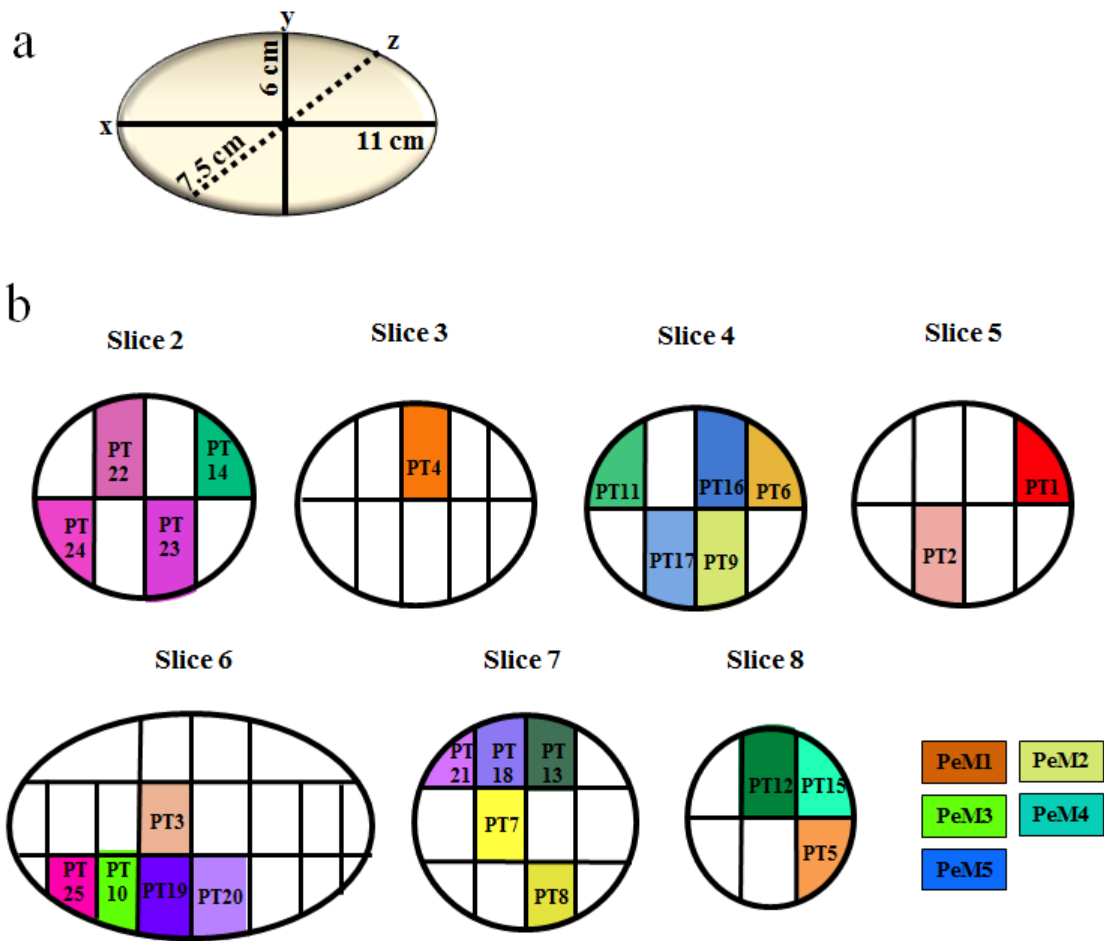
Some sequencing studies of tumors, such as those of renal cell cancers, have suggested that driver gene mutations may occur heterogeneously during tumor evolution<sup>90,91</sup>. Yet the data on pancreatic cancers reported here, as well as data on lung cancers<sup>92</sup>, provide strong evidence against the idea that driver gene heterogeneity is a feature of all common tumor types. In particular, known driver gene alterations, such as those in *KRAS* or *TP53*, were found to be completely homogeneous among the metastatic lesions and distinct anatomic sections of the tumors we studied (Table 3-6). We did not evaluate genetic alterations other than point mutations in this study, given that the major driver gene alterations in PDACs are point mutations and that structural alterations are much less common<sup>83,84</sup>. Copy number alterations can also contribute to heterogeneity, but this



**Figure 3-10. Relatedness and three dimensional geography of Pam02 primary tumor sections and metastases.** a. The three dimensional size of the original primary tumor in centimeters. b. Primary tumor slices are numbered according to the original sectioning and plane order. See Table 3-3 for sample identity Primary tumors are labeled at “PT” followed by a sequential number, the remaining samples are liver mets labeled as LiM and also followed by a sequential number. Colors are matched to the colors used for each sample in Figure 3-8.



**Figure 3-11. Relatedness and three dimensional geography of Pam03 primary tumor sections and metastases.** a. The dimensions of the original primary tumor in centimeters. b. Primary tumor slices are numbered according to the original sectioning and plane order. See Table 3-3 for sample identity. Metastases are labeled by organ followed by a metastasis number. Colors are matched to the colors used for each sample in Figure 3-9.



**Figure 3-12. Relatedness and three dimensional geography of Pam04 primary tumor sections and metastases.** a. The dimensions of the original primary tumor in centimeters. b. Primary tumor slices are numbered according to the original sectioning and plane order. See Table 3-3 for sample identity. Metastases are labeled by organ followed by a metastasis number. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships (as indicated in Supplemental Figure 4).

factor is extremely difficult to reliably measure in tumors with varying non-neoplastic cellular contents.

The term "heterogeneity" is often used in a qualitative rather than quantitative fashion. Every time a normal cell divides, it acquires ~3 mutations<sup>98</sup>, and therefore any two normal cells are not identical. Hence, any statements about the degree of "heterogeneity" in tumors are of uncertain significance in the absence of a comparison to normal tissues. Because heterogeneity cannot yet be easily measured in normal tissues, we used modeling, based on conventional principles and known mutation rates in normal tissues. We were thus able to show that these cancers demonstrate much higher homogeneity than expected in normal tissues. It is important to point out that our analysis was confined to point mutations and did not consider chromosomal changes such as translocations. However, given that point mutations form the vast majority of genetic alterations in cancers<sup>6</sup>, including in PDACs<sup>6,83,84</sup>, our data indicate that bottlenecks that reduce heterogeneity are the predominant force shaping cancer evolution. It is also intriguing that the cellular products of these bottlenecks (subclones) appear to migrate through the very large volumes of the primary tumors (e.g., Figure 3-12). Extensive migrations of subclones have also been observed in primary colorectal cancers<sup>99</sup>. Such extensive migration can be explained by a recent spatial model<sup>100</sup>.

These results have important clinical implications. If driver gene mutations were truly heterogeneous among metastases, then there would be little hope for therapies that target such mutations. Our conclusions about the homogeneity of driver genes among metastatic lesions thereby provide optimism in this regard. These conclusions are also consistent with the results of clinical trials in patients with metastatic cancers<sup>101-103</sup>. All

such trials incorporate measures of the response of several metastatic lesions in each patient. Objective responses are only scored as such if all lesions respond. As so many objective, sometimes striking, responses have been observed in the trials, homogeneity of the targeted driver genes can be inferred.

### **Author Contributions**

C.I.D. and A.M.M. performed the autopsies, C.I.D., A.M.M., B.V., K.K., N.P., M.Z., F.W., and Y.J. designed experiments, A.M.M and M.Z. performed the experiments, J.R., I.B., D.K, B.A., K.C., and M.N. performed modeling, all authors interpreted the data, C.I.D., A.M.M., B.V., J.R., M.Z., I.B., R.H., and K.K. wrote the manuscript.

### **Project Acknowledgements**

This work was supported by National Institutes of Health grants CA179991, CA.I-D.), F31CA180682 (A.M-M.), CA43460 (B.V.), P50 CA62924, The Monastra Foundation, The Virginia and D.K. Ludwig Fund for Cancer Research, Lustgarten Foundation for Pancreatic Cancer Research, and The Sol Goldman Center for Pancreatic Cancer Research, ERC Start grant 279307: Graph Games (J.G.R, D.K., C.K.), Austrian Science Fund (FWF) grant no P23499-N23 (J.G.R, D.K., C.K.), FWF NFN grant no S11407-N23 RiSE (J.G.R, D.K., C.K.).

## Chapter 4 – Methods and Supplementary Data

---

### General

**Tissues.** For dissection, the two projects described here used microdissection from membrane slides and macrodissection from frozen sections cut onto glass slides, respectively. The optimal method for each project was dependent on the cellularity of each sample, the amount of available tumor tissue, and the fact that next-generation sequencing was used for both. Our standard for cutting tissue is typically 5  $\mu$ M thickness. Hematoxylin and eosin staining was executed for all slides as standard protocol dictates. All histology was reviewed under a standard light microscope.

**DNA extraction and quantification.** Genomic DNA (gDNA) was extracted from each tissue piece using a standard phenol and chloroform extraction or a Qiagen QIAmp DNA micro kit (depending on expected yield) followed by precipitation in ethanol. The gDNA was quantified by LINE assay (i.e. counting long interspersed elements (LINE) using real-time PCR), a particularly sensitive method for calculating gDNA concentration for whole genome sequencing. The LINE forward primer was 5'-AAAGCCGCTCAACTACATGG-3' and the reverse primer was 5'-TGCTTTGAATGCGTCCCAGAG-3'. The real-time PCR protocol was 50°C for 2 min, 95°C for 2 min, 40 cycles of 94°C for 10 s, 58°C for 15 s, and 70°C for 30 s, 95°C for 15 s, and 60°C for 30 s. The PCR reactions were carried out using Platinum SYBR Green qPCR mastermix (Invitrogen).

**Sequencing alignment and filtering.** Sequencing was performed on an Illumina Hi-Seq 2000 platform for the target coverages appropriate for each project (see Chapter 2 and 3). Following the completion of sequencing, the data were retrieved and analyzed *in silico* to determine overall coverage and read quality. Reads were aligned to the human reference genome – hg18 for Chapter 2, hg19 for Chapter 3. All low-quality, poorly aligned, or dbSNP containing reads were removed from further analysis. Remaining mutations were analyzed for alterations in known cancer driver genes using information from the COSMIC database. Putative variants were visualized directly using Tag Viewer.

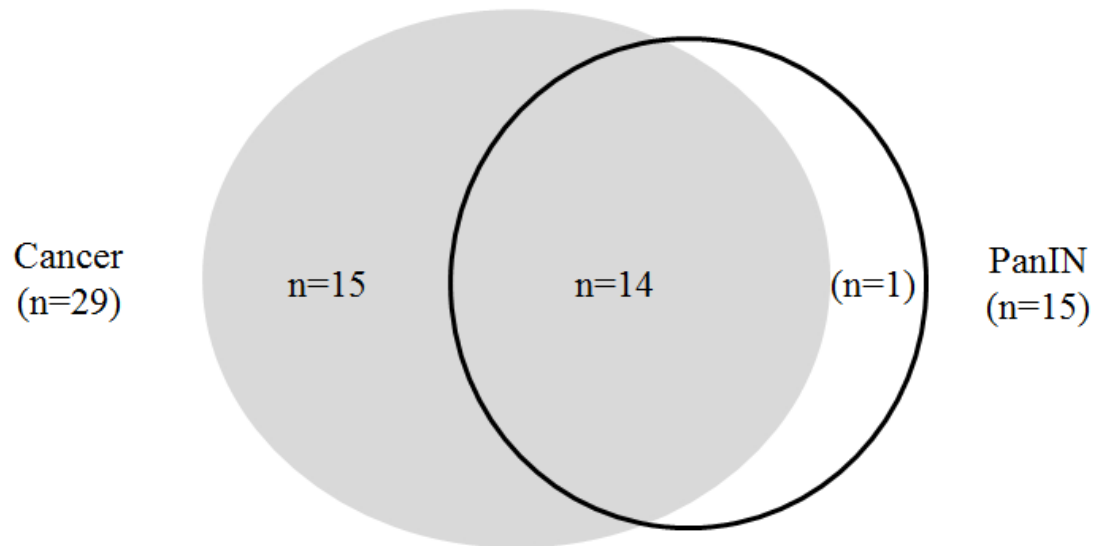
## **PanINs**

**LCM of PanINs and cancer.** Tissue sections were cut at 5  $\mu$ M thickness onto UV-treated membrane slides. Toluidine blue staining was used to visualize PanINs and matched tumors – staining followed manufacturer’s protocols. However, for LCM microdissection, each tissue slide was dried immediately prior to dissection. This facilitates the cutting of the laser as well as the removal of dissected material from the rest of the sample via gravity.

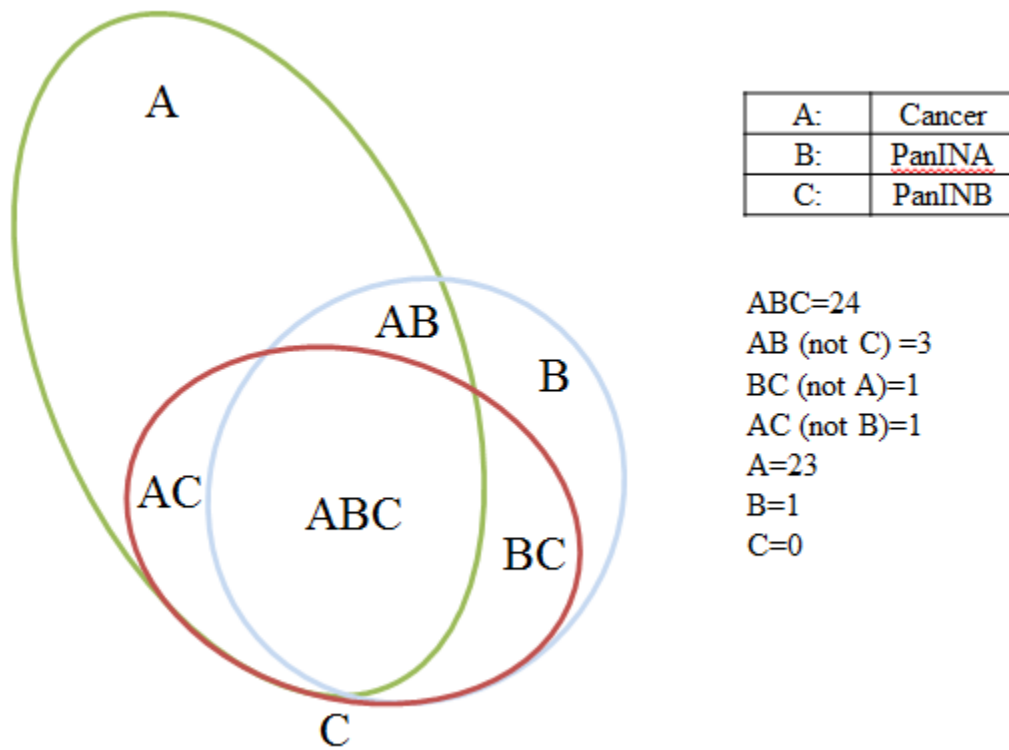
**Filtering of WES data.** Beyond what was generally described for *in silico* analysis, the WES data (containing >500 candidate variants) were filtered for high quality mutations as follows: each variant had to be >10% mutant allele fraction, more than 5 tag pairs, observed in both forward and reverse directions, <2% mutant alleles in the corresponding normal, no SNPs.

**Supplemental Data.** The following five figures (4-1, 4-2, 4-3, 4-4, and 4-5) summarize the evolutionary analysis of each remaining PanIN case thus far.

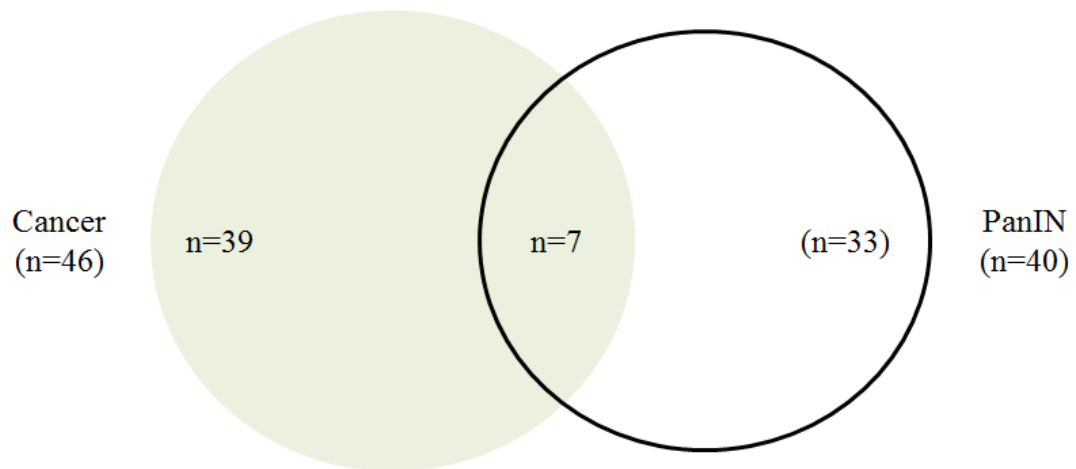




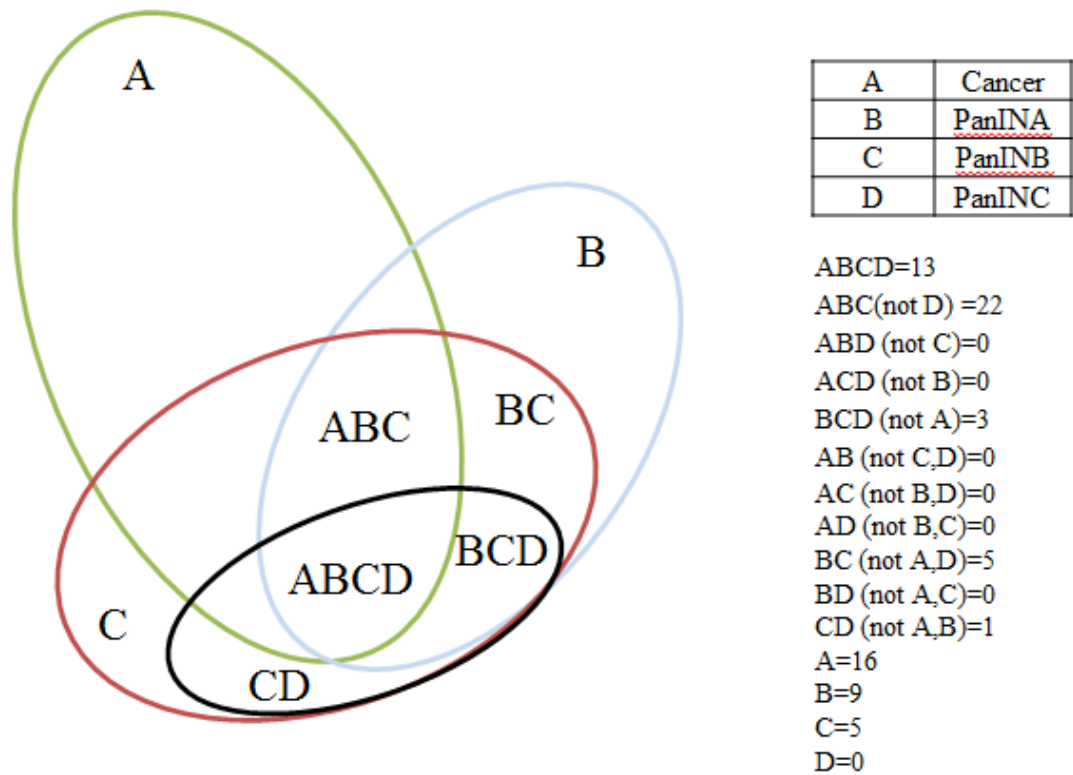
**Figure 4-1. PIN103 evolutionary relationship.** The grey circle represents the cancer, the black circle represents the PanINA. Numbers of mutations are indicated in shared or unique regions.



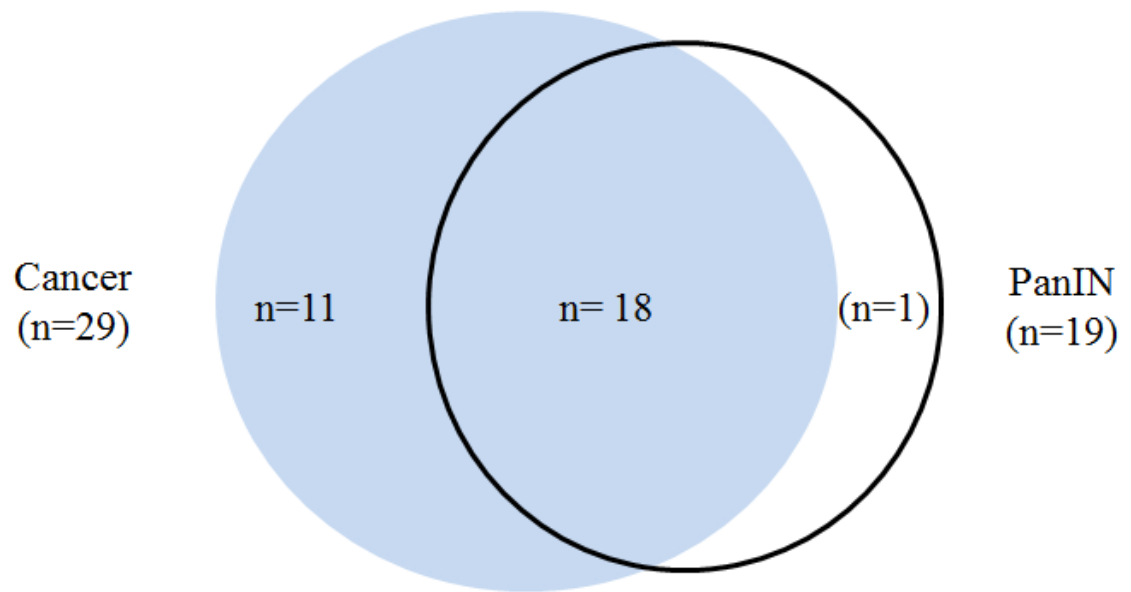
**Figure 4-2. PIN104 evolutionary relationship.** The green circle represents the cancer, the red circle represents the PanINA, and the blue circle represents PanINB. Numbers of mutations are indicated in shared or unique regions.



**Figure 4-3. PIN105 evolutionary relationship.** The green circle represents the cancer, the black circle represents the PanINA. Numbers of mutations are indicated in shared or unique regions.



**Figure 4-4. PIN106 evolutionary relationship.** The green circle represents the cancer, the blue circle represents PanINA, the red circle represents PanINB, and the black circle represents PanINC. Numbers of mutations are indicated in shared or unique regions.



**Figure 4-5. PIN107 evolutionary relationship.** The blue circle represents the cancer, the black circle represents the PanIN. Numbers of mutations are indicated in shared or unique regions.

## **Metastasis**

**Selection of patient autopsies.** The four patients and their respective tissues originated from the Gastrointestinal Cancer Rapid Medical Donation program, a collection of over 150 autopsy cases. This program has been described previously and was deemed in accordance with the Health Insurance Portability and Accountability Act and approved by the Johns Hopkins institutional review board<sup>81</sup>.

**Processing of tissue samples.** Once the body cavity was opened using standard autopsy techniques, the entire pancreas and primary tumor were removed along with each grossly identified metastasis. All tissues were immediately flash-frozen in liquid nitrogen and stored at -80°C. Each primary tumor was serially sliced into 0.5 cm thick slices followed by sectioning of each slice into 1x1 cm tissue samples as described previously. One half of each tissue sample was fixed in 10% buffered-formalin while the remaining tissue was preserved at -80°C. Each metastasis was macrodissected to remove surrounding non-neoplastic tissue. Each frozen primary tumor sample was embedded and frozen in Tissue-Tek OCT for sectioning using a Leica Cryostat. For each tissue sample, a 5 µM thick section was taken to create a hematoxylin and eosin slide to visualize neoplastic cellularity using a microscope. A different set of lesions from patient Pam01 were evaluated in Ref. 83; the three other patients described in this paper had not been evaluated previously. We estimated that the neoplastic cellularity was >50% for Pam01, Pam02, and Pam 03 and >20% for Pam 04.

**Whole genome sequencing and alignment.** For whole genome sequencing (WGS), several metastases were chosen from all four cases along with one or two geographically

distinct sections of the primary tumor from three of the four cases for a total of 35 samples. Only those tissue samples that were confirmed to be of high quality and suitable concentration (>25 ng/ $\mu$ l of amplifiable gDNA) were used for WGS.

**Filtering of whole genome sequencing data and visualization.** Whole genome sequencing generated a large list of potential mutations, even after conventional filtering based on quality scores. We assessed these data with the goal of identifying bona fide mutations and eliminating sequence artifacts. The criteria used to achieve this goal were relaxed compared to what we have previously used<sup>6,104</sup> given the variation in neoplastic cell content among the samples (Table 3-3) and our desire to identify all coding mutations (i.e., high sensitivity, intermediate specificity). Furthermore, we planned on experimentally validating each mutation and could thereby tolerate a higher fraction of false positives in this analysis. The criteria we used were as follows. Each mutant must have been observed in > 3 reads (read defined as the output from one cluster on the Illumina instrument); each mutant must have been observed in at least one read in both directions (i.e., 5' to 3' and 3' to 5' relative to reference genome); each mutant must not have been observed in any reads of the matched normal sample; minimum mutant allele frequency of 10% in samples from patients Pam01, 02, & 03 and 5% in samples from patient Pam04 given the lower neoplastic cell content of the samples from Patient Pam04 (Table 3-3). This analysis, combined with manual inspection of the raw data, yielded a total of 2,356 potential mutations for subsequent validation.

We leveraged the passenger mutations in each case and inferred an overall false positive rate for the mutations of 0.23%, given that two independent cancers were extremely

unlikely to harbor identical passenger mutations. We used a statistical approach, based on mutant read counts and overall coverage, to determine whether a variant was present, absent, or unknown in each matched sample by calculating a p-value for each variant via a binomial distribution. The null hypothesis was that the mutation was absent. We used the step-up method of Benjamini and Hochberg<sup>105</sup> to control for an average false discovery rate (FDR) of 1% in the combined set of p-values from all samples in a patient. Variants with a rejected null hypothesis were labeled as present. The remaining variants (failed to reject the null hypothesis) were labeled as absent if their coverage was  $\geq 100X$  and otherwise labeled as unknown. For phylogenetic analysis, we excluded tumor samples with a median coverage  $< 100X$  or a median all-mutant allele fraction of  $< 5\%$ .

**Targeted sequencing design and validation.** Strict filtering of the WGS data resulted in a highly conservative and confident list of potential mutations that defined the clonal mutation profile for each tumor sample. These lists were used to design a targeted sequencing effort that incorporated the mutation position  $\pm 50$  base pairs in either direction. Additionally, we aimed to increase our sensitivity of mutation detection by increasing our target coverage to  $> 200X$ . To do this, we implemented an Illumina chip-based targeted sequencing approach. Once sequencing was completed, the raw data were aligned and processed as described for the WGS data. To filter the chip data for high quality mutations, we removed those that had greater than 2% mutant allele fraction or no distinct coverage in the corresponding normal sample.



**Evolutionary analysis methods.** We derived phylogenies for each patient. The evolutionary trees were “rooted” at the matched patient’s normal sample and the leaves were formed by the tumor samples (i.e. distinct parts of primary tumors or distinct metastatic lesions). First, we applied hierarchical clustering using UPGMA (unweighted pair group method with arithmetic mean) based on Jaccard distances to find basic genetic similarities among the samples (Tables 3-7 and 3-8). The Jaccard distances were calculated from the partially shared mutations (parsimony-informative mutations).

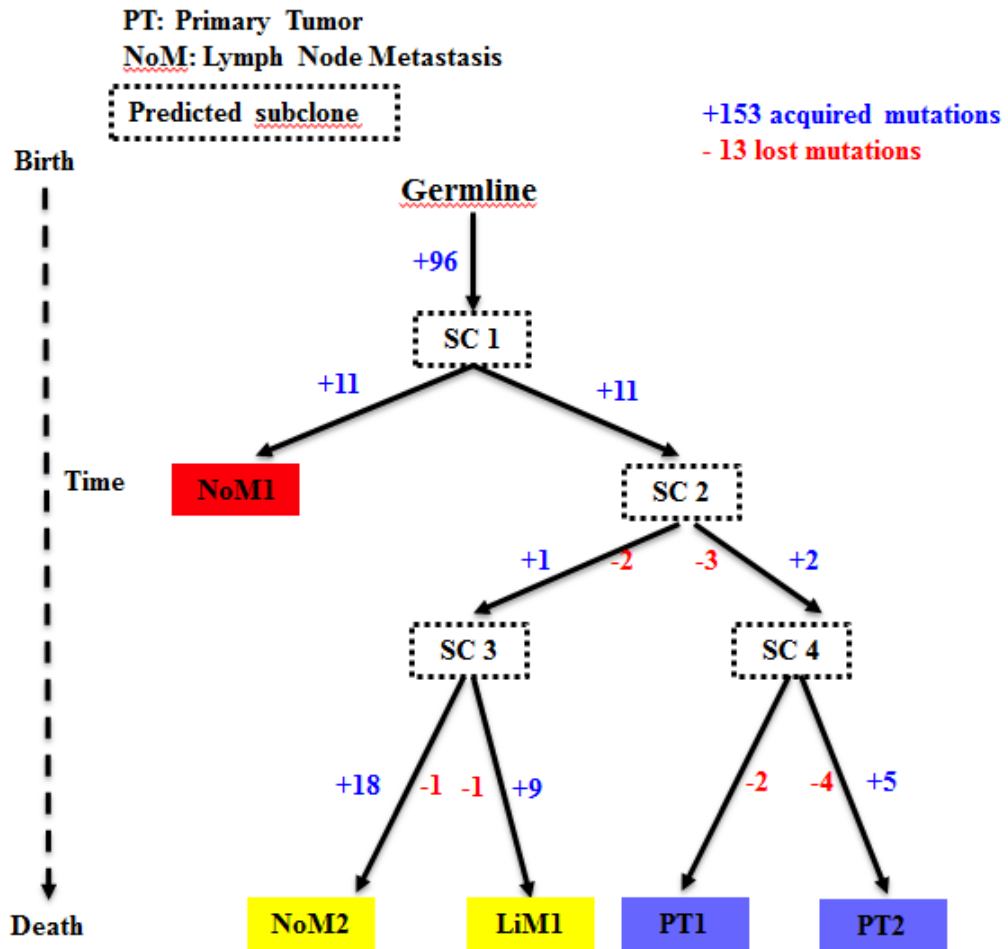
Second, we developed a new phylogenetic approach *MinDel* (extending the clonal ordering introduced by Ref. 90) to infer the most parsimonious tree for each sample set (i.e. the tree minimizing the number of required discontinuities to explain the observed data). Phylogenetic methods, used to analyze germ line rather than somatic mutations, often assume that mutations are both perfect (mutations occur once) and persistent (mutations are not lost)<sup>106</sup>. The perfect phylogeny assumption remains valid in cancer genomics (infinite sites model), however, we removed the persistence assumption in our *MinDel* (minimal deletions) algorithm because some mutations were observed to be frequently discontinuous in our data leading to evolutionarily incompatible explanations. While the apparent discontinuous mutation may be due to sample cellularity and sequencing sensitivity, it may also be seen in association with chromosome-level genetic events that culminate in the loss of heterozygosity at that location<sup>84</sup>. In a *MinDel* phylogeny, each identified mutation is acquired exactly once but can be lost throughout tumor evolution. If multiple tree topologies minimize the number of required “losses”, the number of subclones during tumor progression (parsimony) and the depth of the evolutionary tree (shallowness) is also minimized<sup>107</sup>.

Last, we used neighbor joining based on genetic distances among the samples to confirm the discovered evolutionary relationships. In general, all three approaches showed that most mutations were acquired prior to genetic divergence in our samples. The phylogenies also indicated that multiple mutations were either lost during progression or undetected during sequencing. The genetic homogeneity among the samples resulted in relatively “low” support (i.e. a limited number of mutations) for branching events, a difficulty also observed in sequencing studies of incipient or recently diverged species<sup>108</sup>. Across the four patients, the number of acquired mutations ranged from 102-204 whilst the number of lost mutations ranged from 13-98. Restricting our analyses to samples that underwent WGS minimized this range to 5-51.

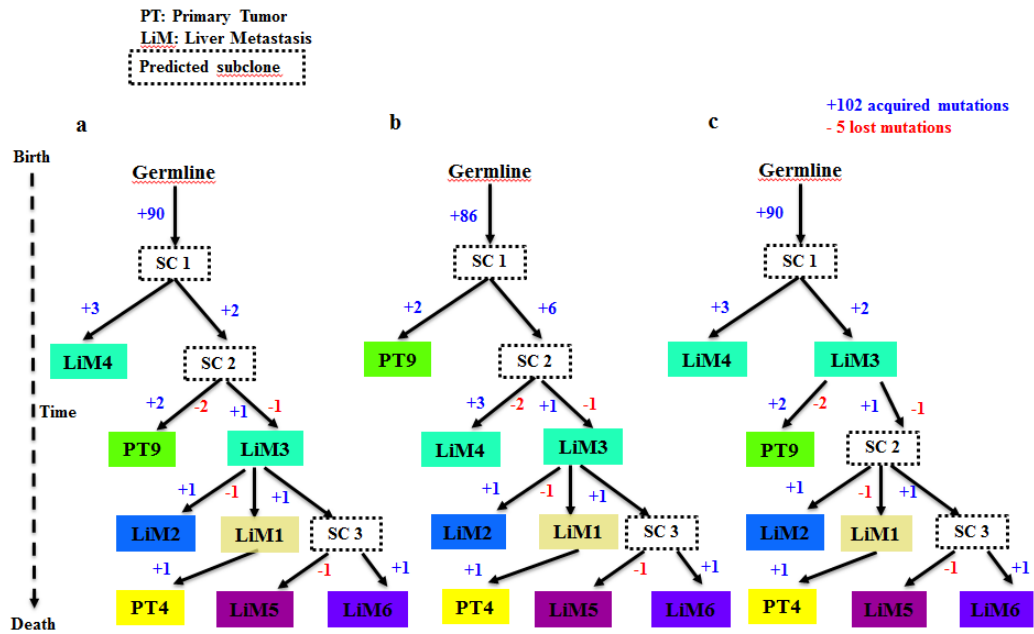
#### **Supplemental Data.**

All three phylogenetic approaches resulted in completely identical trees for Pam01. These trees clearly indicated that the pelvic lymph node metastasis seeded earlier than either the portal lymph node metastasis or the liver metastasis. The portal lymph node metastasis and the liver metastasis were seeded by the identical subclone (Figure 4-6).

For Pam02, we found that many of the liver metastases were more closely related to one of the primary tumor sections than they were to each other (Figure 3-4). This provided strong evidence that coexistent metastases within the liver were more likely to be derived from different subclones (or regions) in the primary tumor than from each other (Figure 3-8 and 4-7). Interestingly, we also noted that these regions were not obviously related to adjacent regions, and were often more related to



**Figure 4-6. MinDel phylogeny for Pam01.** Time is represented on the left axis, divergence is represented on the x axis. See Table 3-3 for sample identity. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. Hypothetical subclones are indicated by “SC” followed by the subclone number. The numbers of acquired mutations are in blue font with a “+” sign while the numbers of discontinuous (i.e. non-persistent) mutations are in red font with a “-“ sign.

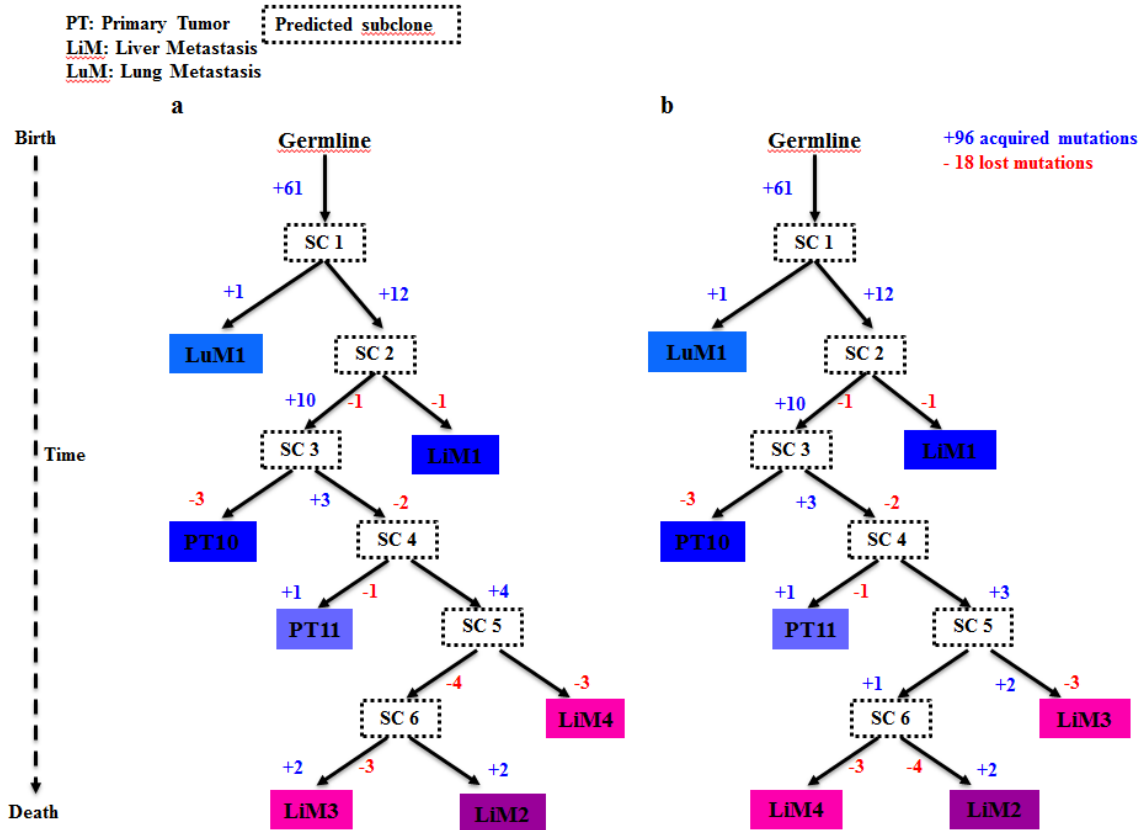


**Figure 4-7. MinDel phylogeny for Pam02 whole genome sequencing samples.** Time is represented on the y axis, divergence is represented on the x axis. See Table 3-3 for sample identity. Hypothetical subclones are indicated by “SC” followed by the subclone number. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. The numbers of acquired mutations is in blue font with a “+” sign while the numbers of discontinuous (i.e. non-persistent) mutations are in red font with a “-” sign. Trees (a), (b), and (c) are equally parsimonious because they have the same number of theoretical subclones and acquired and lost mutations. Due to the genetic homogeneity the temporal ordering is slightly different from the one inferred by neighbor joining (Figure 3-8). The results suggest that LiM1 seeded PT4. However, since not all areas of the primary tumor were evaluated, most likely a related subclone gave rise to these genetically similar samples.

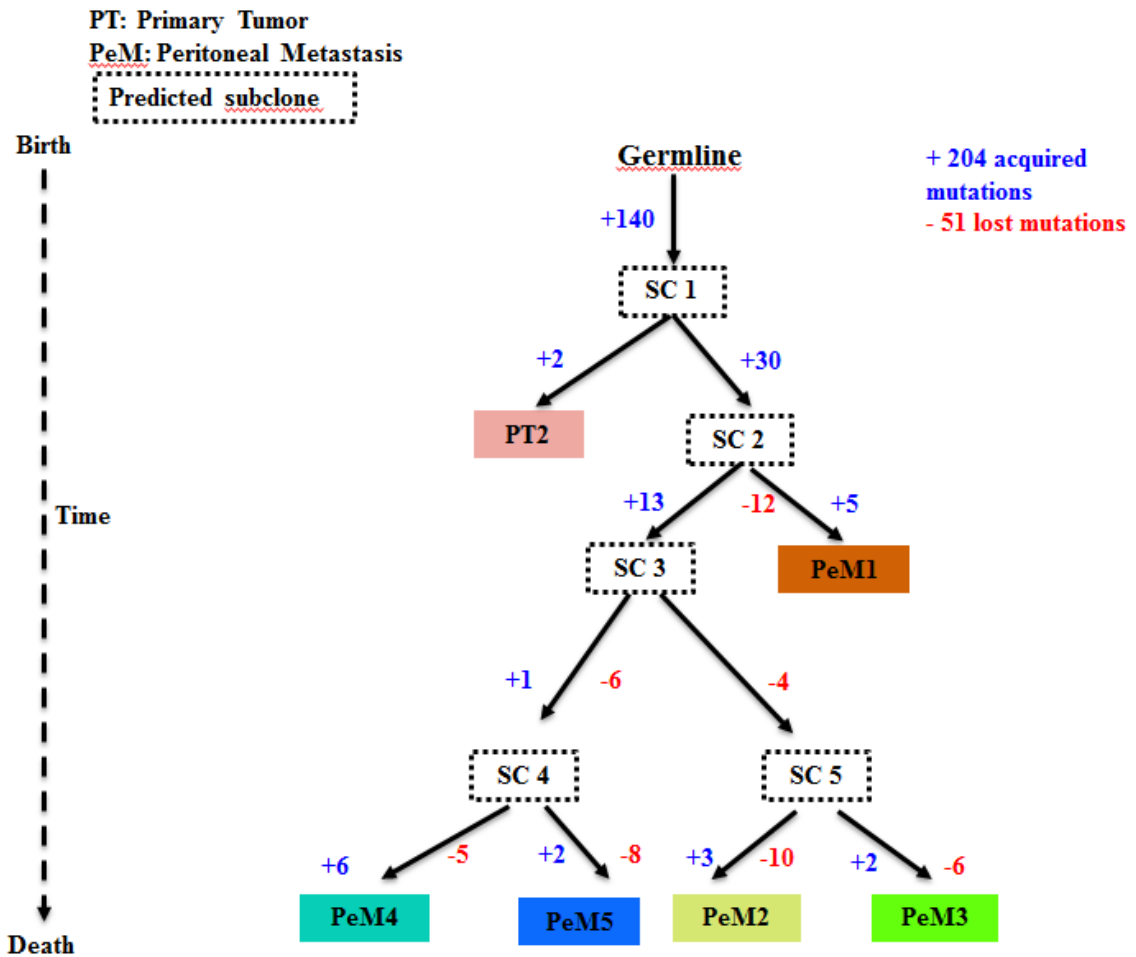
distant subclones within the primary tumor (Figure 3-10). This suggests a considerable degree of migration of the cancer cells throughout the tumor. In one instance, a liver metastasis (LiM2) appeared to be the ancestor to a primary tumor section (PT11) (Figure 3-8). Because we could not possibly evaluate all areas of the primary tumor, the most likely explanation for this observation is that another subclone gave rise to both PT11 and the liver metastasis LiM2. To refine the evolutionary relationships among the metastases, we used the *MinDel* algorithm and focused on the targeted sequencing data of the WGS samples to minimize the necessary deletions which might be caused by differently mixed subclones in some primary samples (Figure 4-7). The support for the branching events remained weak and robust evolutionary relationships among the samples could not be inferred due to genetic homogeneity.

In Pam03, more parsimony-informative mutations were identified. We evaluated four additional metastases (LiM6-9) to assist our phylogenetic analysis. All three phylogenetic approaches showed that the three liver metastases (LiM2, 3, and 4) were closely related, indicating that they were seeded at nearly the same time from an identical subclone (Figures 3-5, 3-9, and 4-7). The analyses also indicate that the lung metastasis diverged earlier in the evolution of this tumor than the liver metastases (Figure 4-7).

For Pam04, no clear evolutionary trees could be derived with confidence except by limiting the analysis to the WGS samples (Figure 4-8). Interestingly, the primary tumor for this case demonstrated a similar mixed pattern without evidence for strictly spatially distinct subclones (Figure 3-12). It was notable that this patient was the only one with a peritoneal pattern of metastasis, a form of metastasis



**Figure 4-7. MinDel phylogeny for Pam03 whole genome sequencing samples.** Time is represented on the left axis, divergence is represented on the x axis. See Table 3-3 for sample identity. Hypothetical subclones are indicated by “SC” followed by the subclone number. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. The numbers of acquired mutations are in blue font with a “+” sign while discontinuous (i.e. non-persistent) mutations are in red font with a “-” sign. Trees (a) and (b) are equally parsimonious because they have the same number of predicted subclones and lost mutations. Both trees suggest very similar evolutionary trajectories and an early seeding of the lung metastases LuM1.



**Figure 4-8. MinDel phylogeny for Pam04 whole genome sequencing samples.** Time is represented on the left axis, divergence is represented on the x axis. See Table 3-3 for sample identity. Hypothetical subclones are indicated by “SC” followed by the subclone number. Colors indicate discrete tumor samples and follow the rainbow spectrum, scaling from ancestral to descendant as indicated by the evolutionary relationships. (as indicated in Figure 3-6). The numbers of acquired mutations are in blue font with a “+” sign while the numbers of discontinuous (i.e. non-persistent) mutations are in red font with a “-” sign.

unrelated to hematogenous spread<sup>69</sup>. Exchange of cells within the peritoneum could in theory explain the difficult tree reconstructions. Though meaningful trees could not be inferred, a large number (65) of mutations were present in all lesions from this patient, similar to the situation in the other cases.



## References

---

1. Frank, S. *Dynamics of Cancer: Incidence, Inheritance, and Evolution*. (2007).
2. Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* **8**, 1–12 (1954).
3. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–67 (1990).
4. Knudson, A. G. Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer* **1**, 157–62 (2001).
5. Weinberg, R. *The Biology of Cancer*. (2006).
6. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–58 (2013).
7. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
8. Vogelstein, B. & Kinzler, K. W. *The Genetic Basis of Human Cancer*. (2002).
9. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–74 (2011).
10. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–99 (2004).
11. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–13 (2012).
12. Nowell, P. C. The Clonal Evolution of Tumor Cell Populations Acquired genetic lability permits stepwise selection. *Science* **194**, 23–28 (1976).
13. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA. Cancer J. Clin.* **65**, 5–29 (2015).
14. Maitra, A. & Hruban, R. H. Pancreatic cancer. *Annu. Rev. Pathol.* **3**, 157–88 (2008).
15. Sohn, T. A. *et al.* Resected adenocarcinoma of the pancreas-616 patients: results, outcomes, and prognostic indicators. *J. Gastrointest. Surg.* **4**, 567–79 (2000).
16. Goggins, M. Markers of pancreatic cancer: working toward early detection. *Clin. Cancer Res.* **17**, 635–7 (2011).
17. Rosty, C. & Goggins, M. Early detection of pancreatic carcinoma. *Hematol. Oncol. Clin. North Am.* **16**, 37–52 (2002).

18. Bhutani, M. S., Thosani, N., Suzuki, R. & Guha, S. Pancreatic cancer screening: what we do and do not know. *Clin. Gastroenterol. Hepatol.* **11**, 731–3 (2013).
19. Burris, H. A. *et al.* Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J. Clin. Oncol.* **15**, 2403–13 (1997).
20. Verslype, C. *et al.* The management of pancreatic cancer. Current expert opinion and recommendations derived from the 8th World Congress on Gastrointestinal Cancer, Barcelona, 2006. *Ann. Oncol.* **18**, vii1–vii10 (2007).
21. Moore, M. J. *et al.* Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J. Clin. Oncol.* **25**, 1960–6 (2007).
22. Conroy, T. *et al.* FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N. Engl. J. Med.* **364**, 1817–25 (2011).
23. Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324**, 1457–61 (2009).
24. McDonald, O. G., Maitra, A. & Hruban, R. H. Human correlates of provocative questions in pancreatic pathology. *Adv. Anat. Pathol.* **19**, 351–62 (2012).
25. Makohon-Moore, A., Brosnan, J. A. & Iacobuzio-donahue, C. A. Pancreatic cancer genomics : insights and opportunities for clinical translation. *Genome Med.* (2013).
26. Gallmeier, E. & Kern, S. E. Targeting Fanconi anemia/BRCA2 pathway defects in cancer: the significance of preclinical pharmacogenomic models. *Clin. Cancer Res.* **13**, 4–10 (2007).
27. Hustinx, S. R. *et al.* Homozygous deletion of the MTAP gene in invasive adenocarcinoma of the pancreas and in periampullary cancer: a potential new target for therapy. *Cancer Biol. Ther.* **4**, 83–6 (2005).
28. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–6 (2008).
29. Biankin, A. V *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405 (2012).
30. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
31. Su, G. H. *et al.* ACVR1B (ALK4, activin receptor type 1B) gene mutations in pancreatic carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3254–7 (2001).
32. Hruban, R. H., Canto, M. I., Goggins, M., Schulick, R. & Klein, A. P. Update on familial pancreatic cancer. *Adv. Surg.* **44**, 293–311 (2010).

33. Almoguera, C. *et al.* Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell* **53**, 549–54 (1988).
34. Jancík, S., Drábek, J., Radzioch, D. & Hajdúch, M. Clinical relevance of KRAS in human cancers. *J. Biomed. Biotechnol.* **2010**, 150960 (2010).
35. Moskaluk, C. A., Hruban, R. H. & Kern, S. E. p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. *Cancer Res.* **57**, 2140–3 (1997).
36. Caldas, C. *et al.* Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat. Genet.* **8**, 27–32 (1994).
37. Liggett, W. H. & Sidransky, D. Role of the p16 tumor suppressor gene in cancer. *J. Clin. Oncol.* **16**, 1197–206 (1998).
38. Schutte, M. *et al.* Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. *Cancer Res.* **57**, 3126–30 (1997).
39. Barton, C. M. *et al.* Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. *Br. J. Cancer* **64**, 1076–82 (1991).
40. Lüttges, J. *et al.* Allelic loss is often the first hit in the biallelic inactivation of the p53 and DPC4 genes during pancreatic carcinogenesis. *Am. J. Pathol.* **158**, 1677–83 (2001).
41. Nigro, J. M. *et al.* Mutations in the p53 gene occur in diverse human tumour types. *Nature* **342**, 705–8 (1989).
42. Redston, M. S. *et al.* p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. *Cancer Res.* **54**, 3025–33 (1994).
43. Goggins, M. *et al.* Genetic alterations of the transforming growth factor beta receptor genes in pancreatic and biliary adenocarcinomas. *Cancer Res.* **58**, 5329–32 (1998).
44. Wilentz, R. E. *et al.* Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. *Cancer Res.* **60**, 2002–6 (2000).
45. Heldin, C.-H. & Moustakas, A. Role of Smads in TGF $\beta$  signaling. *Cell Tissue Res.* **347**, 21–36 (2012).
46. Hahn, S. A. *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–3 (1996).
47. Ayala, F. J. Darwin's greatest discovery: design without designer. *Proc. Natl. Acad. Sci. U. S. A.* **104 Suppl**, 8567–73 (2007).

48. Darwin, C. & Mayr, E. *On the origin of species by means of natural selection*. 502 (Harvard University Press).
49. Orr, H. A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–27 (2005).
50. Bateson, W. & Mendel, G. *Mendel's Principles of Heredity*. 464 (Courier Corporation, 1902).
51. Fisher, R. A. *The Genetical Theory of Natural Selection*. 318 (Oxford University Press, 1930).
52. Gould, S. J. *The Structure of Evolutionary Theory*. (Harvard University Press, 2002).
53. Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255**, 197–200 (1975).
54. Greaves, M. *Cancer: The Evolutionary Legacy*. 276 (Oxford University Press, 2001).
55. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9**, M57–60 (1999).
56. Yachida, S. & Iacobuzio-Donahue, C. A. Evolution and dynamics of pancreatic cancer progression. *Oncogene* 1–8 (2013).
57. Hruban, R. H. *et al.* Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for pancreatic duct lesions. *Am. J. Surg. Pathol.* **25**, 579–86 (2001).
58. Brat, D. J., Lillemoe, K. D., Yeo, C. J., Warfield, P. B. & Hruban, R. H. Progression of pancreatic intraductal neoplasias to infiltrating adenocarcinoma of the pancreas. *Am. J. Surg. Pathol.* **22**, 163–9 (1998).
59. Hruban, R. H. *et al.* An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms. *Am. J. Surg. Pathol.* **28**, 977–87 (2004).
60. Van Heek, N. T. *et al.* Telomere shortening is nearly universal in pancreatic intraepithelial neoplasia. *Am. J. Pathol.* **161**, 1541–7 (2002).
61. Iacobuzio-Donahue, C. A., Ryu, B., Hruban, R. H. & Kern, S. E. Exploring the host desmoplastic response to pancreatic carcinoma: gene expression of stromal and neoplastic cells at the site of primary invasion. *Am. J. Pathol.* **160**, 91–9 (2002).
62. Iacobuzio-Donahue, C. A. *et al.* Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. *Cancer Res.* **63**, 8614–22 (2003).

63. Szafranska, A. E. *et al.* MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. *Oncogene* **26**, 4442–52 (2007).
64. Prasad, N. B. *et al.* Gene expression profiles in pancreatic intraepithelial neoplasia reflect the effects of Hedgehog signaling on pancreatic ductal epithelial cells. *Cancer Res.* **65**, 1619–26 (2005).
65. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4283–8 (2008).
66. Campbell, P. J. *et al.* Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13081–6 (2008).
67. Naxerova, K. & Jain, R. K. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nat. Rev. Clin. Oncol.* **advance on**, (2015).
68. Talmadge, J. E. & Fidler, I. J. The biology of cancer metastasis. *Cancer Res.* **70**, 5649–69 (2010).
69. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9**, 274–84 (2009).
70. Yang, J. & Weinberg, R. a. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. Cell* **14**, 818–29 (2008).
71. Folkman, J. Angiogenesis. *Annu. Rev. Med.* **57**, 1–18 (2006).
72. Barcellos-Hoff, M. H., Lyden, D. & Wang, T. C. The evolution of the cancer niche during multistage carcinogenesis. *Nat. Rev. Cancer* (2013).
73. Nguyen, D. X. & Massagué, J. Genetic determinants of cancer metastasis. *Nat. Rev. Genet.* **8**, 341–52 (2007).
74. Bernards, R. & Weinberg, R. a. A progression puzzle. *Nature* **418**, 823 (2002).
75. Michor, F., Nowak, M. a & Iwasa, Y. Stochastic dynamics of metastasis formation. *J. Theor. Biol.* **240**, 521–30 (2006).
76. Bozic, I. & Nowak, M. A. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15964–8 (2014).
77. Brosnan, J. A. & Jacobuzio-Donahue, C. A. A new branch on the tree: next-generation sequencing in the study of cancer evolution. *Semin. Cell Dev. Biol.* **23**, 237–42 (2012).
78. Hruban, R. H., Maitra, A. & Goggins, M. Update on pancreatic intraepithelial neoplasia. *Int. J. Clin. Exp. Pathol.* **1**, 306–16 (2008).

79. Yachida, S. & Iacobuzio-Donahue, C. A. The Pathology and Genetics of Metastatic Pancreatic Cancer. (2009).
80. Stathis, A. & Moore, M. J. Advanced pancreatic carcinoma: current treatment and future challenges. *Nat. Rev. Clin. Oncol.* **7**, 163–72 (2010).
81. Embuscado, E. *et al.* Immortalizing the Complexity of Cancer Metastasis. *Cancer Biol. Ther.* 548–554 (2005).
82. Iacobuzio-Donahue, C. a *et al.* DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *J. Clin. Oncol.* **27**, 1806–13 (2009).
83. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–7 (2010).
84. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–13 (2010).
85. Haeno, H. *et al.* Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell* **148**, 362–75 (2012).
86. Yachida, S. *et al.* Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. *Clin. Cancer Res.* **18**, 6339–47 (2012).
87. Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–61 (2012).
88. Comen, E., Norton, L. & Massagué, J. Clinical implications of cancer self-seeding. *Nat. Rev. Clin. Oncol.* **8**, 369–77 (2011).
89. Murphy, S. J. *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* **145**, 1098–1109.e1 (2013).
90. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* (2012).
91. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* 1–12 (2014).
92. De Bruin, E. C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251–256 (2014).
93. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014).
94. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368**, 842–51 (2013).

95. Iacobuzio-Donahue, C. a, Velculescu, V. E., Wolfgang, C. L. & Hruban, R. H. Genetic basis of pancreas cancer development and progression: insights from whole-exome and whole-genome sequencing. *Clin. Cancer Res.* **18**, 4257–65 (2012).
96. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9530–5 (2011).
97. Diaz, L. a *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–40 (2012).
98. Tomasetti, C., Vogelstein, B. & Parmigiani, G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc. Natl. Acad. Sci. U. S. A.* 1–6 (2013).
99. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
100. Waclaw, B. *et al.* Spatial model predicts dispersal and cell turnover causes reduced intratumor heterogeneity (submitted). *Nature* (2015).
101. Chapman, P. B. *et al.* Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
102. Kwak, E. L. *et al.* Anaplastic Lymphoma Kinase Inhibition in Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
103. Sharma, S. V, Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7**, 169–81 (2007).
104. Jiao, Y. *et al.* Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat. Genet.* **45**, 1470–3 (2013).
105. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
106. Felsenstein, J. Numerical Methods for Inferring Evolutionary Trees. *Q. Rev. Biol.* (1982).
107. Strino, F., Parisi, F., Micsinai, M. & Kluger, Y. TrAp: a tree approach for fingerprinting subclonal tumor composition. *Nucleic Acids Res.* **41**, e165 (2013).
108. Seehausen, O. *et al.* Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).

## Curriculum vitae

### **Alvin P. Makohon – Moore**

Born in Ann Arbor, Michigan, United States – September 29, 1987  
The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins School of Medicine  
1550 Orleans Street, CRBII Rm 316, Baltimore, Maryland 21231  
(410) 955-3511 • [amakohol@jhmi.edu](mailto:amakohol@jhmi.edu)

#### Education

---

**Ph.D. Candidate, Pathology**

Laboratory of Dr. Christine Iacobuzio-Donahue  
Johns Hopkins University, School of Medicine  
Baltimore, Maryland, USA

**July 2010 - March 2015**  
(expected)

**Bachelor of Science, Zoology**

Concentration: Cell and Developmental Biology  
Michigan State University, College of Natural Science  
East Lansing, Michigan, USA

**August 2005 - May 2010**

#### Honors and Awards

---

**National Cancer Institute Pre-doctoral Fellowship**  
Award ID F31CA180682

**September 2013 - present**

**AACR Minority Fellowship Award**  
Special Conference in Tumor Invasion and Metastasis

**January 2013**

**MPI Research, Inc. Undergraduate Research Award**

**February 2010**

**Michigan State University Dean's List**

**2005-2010**

#### Research Experience

---

**Ph.D. Thesis: Evolution of pancreatic cancer**  
Laboratory of Dr. Christine Iacobuzio-Donahue  
Department of Pathology, Johns Hopkins

**March 2011 – present**

My research focuses on using genetic techniques to understand how pancreatic cancer evolves and metastasizes. I have:

- Harvested tissue samples for the Rapid Autopsy program;
- Processed tissue and genomic DNA for experimentation;
- Designed large-scale sequencing experiments;
- Interpreted diverse types of sequencing data;
- Prepared scientific presentations and manuscripts.



## **Undergraduate: Michigan State University**

West Nile Virus microbiology, Dr. Gabe Hamer

**August 2009 – May 2010**

Protein Crystallization, Dr. Dennis Arvidson

**June 2008 – August 2009**

Aquatic Entomology, Dr. Rich Merritt

**September 2007 – January 2010**

## Teaching Experience

---

### **Student Advisor**

**May – August 2009**

Department of Zoology, Michigan State University

I assisted in summer advising for incoming students:

- Worked independently to advise curriculum;
- Advised students on electives and schedule planning.

### **Undergraduate Teaching Assistant**

**November 2007 – August 2009**

Biological Sciences 110, Michigan State University

I was a teaching assistant for a biology course and participated in an educational research project:

- Assisted in classroom management and grading;
- Analyzed and interpreted classroom data.

## Publications and Presentations

---

**Makohon-Moore, A.**, Brosnan, J. A., & Iacobuzio-Donahue, C. A. (2013). Pancreatic cancer genomics : insights and opportunities for clinical translation. *Genome Medicine*.

**Makohon-Moore, A.**, & Iacobuzio-Donahue, C. A. (2013). Considerations for Sequencing Analyses of Pancreatic Cancer Progression and Metastasis. In G. H. Su (Ed.), *Pancreatic Cancer: Methods and Protocols* (Vol. 980, pp. 121–129). Totowa, NJ: Humana Press.

Hamer, G. L., Anderson, T. K., Berry, G. E., **Makohon-Moore, A. P.**, Crafton, J. C., Brawn, J. D., Dolinski A.C., Krebs B.L., Ruiz M.O., Muzzall P.M., Goldberg T.L., Walker, E. D. (2013). Prevalence of filarioid nematodes and trypanosomes in American robins and house sparrows, Chicago USA. *International Journal for Parasitology. Parasites and Wildlife*, 2, 42–9.

**Makohon-Moore\***, A., Zhang\*, M., Reiter\*, J.G., Bozic, I., Wong, F., Jiao, Y., Kundu, D., Allen, B., Chatterjee, K., Wood, L.D., Hruban, R.H., Nowak, M.A., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., and Iacobuzio-Donahue, C.A. \*these authors contributed equally (2015). The natural history of pancreatic cancer is dominated by genetic homogeneity. Abstract for poster presentation, American Association for Cancer Research Annual Meeting 2015, Philadelphia, PA.

**Makohon-Moore, A.P.**, and Iacobuzio-Donahue, C.A. 2011. The Genetics of *Dicer1* in pancreatic cancer metastasis. Abstract for poster presentation, Annual Johns Hopkins Pathobiology Retreat, Baltimore MD.

**Makohon-Moore, A.P.**, Hamer, G.B., and Walker, E.D. 2010. Avian Filaroid nematodes and the potential for enhancement of West Nile virus. Abstract for oral presentation, MSU Undergraduate Research and Arts Forum, East Lansing MI.