Chapter 26

Pancreatic cancer

Sergii Ivakhno, Kristopher Frese, Simon Tavare, Christine Iacobuzio-Donahue and David Tuveson

Introduction

Pancreatic cancer is highly lethal, with a five-year survival rate of less than 5% [1]. The most common form of pancreatic cancer is pancreatic ductal adenocarcinoma (PDA), accounting for over 250,000 deaths worldwide annually [2] and ranked as the fourth leading cause of cancer mortality in the USA [3]. The poor survival of PDA patients has been attributed to the advanced stage of disease presentation and ineffective therapeutic options. Approximately 20% of PDA patients are diagnosed with localized disease and are candidates for surgical resection with curative intent; however, the majority of these patients unfortunately relapse within several years and carry a five-year post-operative survival of only 20% despite additional treatment with adjuvant chemotherapy [4, 5]. The current worldwide standard treatment for advanced and surgically unresectable pancreatic cancer is the chemotherapeutic agent gemcitabine; however, this intervention only modestly improves patients' symptoms and has little measureable effect on overall survival [6]. Despite tremendous efforts, only the EGFR inhibitor erlotinib has been approved as an additional agent for PDA patients, albeit for an incremental increase in median overall survival of only two weeks [7]. Therefore the development of new therapeutic approaches for PDA patients that are based upon the underlying biological complexity of this disease is of the utmost importance. The formulation of new treatment strategies for PDA patients requires a detailed understanding of the genetic and epigenetic alterations in PDA tumors and the biochemical signaling networks that are active in this disease. By collectively integrating such information, a multimodal molecular network (MMMN) can be established for PDA and used to probe the biological vulnerabilities of this nefarious malignancy.

The proposed progression model of pancreatic cancer

The first integrative model of colon cancer progression proposed by Fearon and Vogelstein correlated the anatomical and histopathological progression from a pre-invasive intestinal polyp to invasive colon cancer with the accumulation of genetic and epigenetic mutations thought to be causal events in these processes [8]. Subsequently, similar models have been postulated for other cancers, including pancreatic cancer [9]. PDA histologically arises in the ductal exocrine compartment of the pancreas due to the presence of pre-invasive neoplasms in small ducts. However, the precise nature of the originating cell type for PDA is unknown and may include either pancreatic progenitor cells or mature epithelial cells that dedifferentiate [10]. The cell of origin for PDA is hypothesized to progress through a series of pre-invasive histopathological states – termed pancreatic intraepithelial neoplasms (PanINs) – prior to the acquisition of invasive and metastatic characteristics. Additionally, pre-invasive cystic pancreatic neoplasms distinct from PanINs develop in either the large pancreatic ducts and are termed intraductal papillary mucinous neoplasms (IPMNs), or peripheral small ducts and are classified as mucinous cystic neoplams (MCNs). Both IMPNs and MCNs can be found in close proximity to invasive PDA in resected specimens, suggesting but not proving that PDA can also develop from these precursor neoplasms. Since PanINs appear to be the predominant precursor, we will focus on these pre-neoplasms for the remainder of the review. Although PanINs of different grades are often conceptualized as discrete steps in PDA progression, it is much more plausible that they represent static pictures of a continuously evolving process.

Systems Biology of Cancer, ed. S. Thiagalingam. Published by Cambridge University Press.

While normal pancreatic ductal epithelial cells are cuboidal with uniformly small round nuclei, the epithelial cells of grade 1 PanINs are characterized by a transition to a columnar appearance and concomitant accumulation of apical cytoplasmic mucin. More severe grade 2 PanINs harbor cells with less prominent mucin, abnormally shaped nuclei and loss of polarity. Finally, grade 3 PanIN exhibit all characteristics of grade 2 and in addition have cells budding into the ductal lumen that occasionally form cribriform structures that bridge across the lumen of the duct. Grade 3 PanINs are generally considered to be carcinoma *in situ*, which makes them an attractive therapeutic target.

Molecular genetics of pancreatic cancer

Despite its poor prognosis, the molecular genetics of pancreatic cancer progression are remarkably well defined (Figure 26.1). The most common mutation found in pancreatic cancer is point mutation of the *KRAS2* gene that produces a constitutively active form of the Kras protein. Active GTP-bound Kras can be self-inactivated through its intrinsic weak GTPase activity resulting in a GDP-bound state. Kras is physiologically activated by a variety of growth factor receptors, most often receptor tyrosine kinases, which promote the membrane recruitment and activation of guanine nucleotide exchange factors (GEFs) that catalyze the exchange of GDP for GTP. Conversely, Kras inactivation is attained through the recruitment of GTPase-activating proteins (GAPs) that dramatically accelerate intrinsic Kras GTPase activity. Oncogenic mutations in Kras invariably occur at amino acids 12, 13, or 61 [11] and produce a constitutively active GTP-bound Kras molecule that is unresponsive to GAP-mediated inactivation. Kras mutations are present in approximately 40% of PanIN-1 and over 90% of invasive pancreatic cancers, suggesting that oncogenic Kras is an initiating factor for pancreatic cancer [12, 13]. Furthermore, *KRAS2* is amplified in approximately 25% and overexpressed in the majority of end-stage cancers. Intriguingly, expression of an oncogenic allele of Kras in the developing murine pancreas promotes the full spectrum of pancreatic cancer, including all grades of PanIN, with complete penetrance [14].

In addition to Kras, mutations in three key tumor suppressor genes (TSG) are frequently found in sporadic pancreatic cancer. Over 90% of all pancreatic cancers harbour $p16^{INK4A}$ mutations including intragenic mutations followed by LOH (40%), homozygous deletions in the 9p21 locus (40%), and promoter hypermethylation (15%) [15–17]. $p16^{INK4A}$ inhibits cdk4/6/cyclin D-mediated phosphorylation of Rb, and thus controls the G1-S transition. Because these mutations are first observed in grade 2 PanINs, this suggests that $p16^{INK4A}$ mutations promote progression rather than initiation of pancreatic cancer [18]. This hypothesis is further supported by the observation that familial atypical multiple mole melanoma (FAMMM) syndrome kindreds with germline mutations in $p16^{INK4A}$ have an increased risk for pancreatic cancer but are not predisposed to early development of the disease [19]. Furthermore, biallelic deletion of the CDKN2A locus in the developing murine pancreas fails to elicit any phenotype; however, in conjunction with a mutant Kras allele results in the development of rapidly invasive PDA [20].

Similar to $p16^{INK4A}$, mutations in *TP53* are found in 55 to 70% of invasive pancreatic cancers [21, 22]. p53 regulates apoptosis and cell cycle arrest through a variety of downstream pathways, and its deletion provides a clear advantage to preinvasive neoplastic cells. Indeed, p53 mutation may allow cell survival in the presence of genomic instability that will otherwise trigger apoptosis. *TP53* mutations occur relatively late in pancreatic carcinogenesis and are not detected until grade 3 PanIN [23, 24]. The importance of *TP53* mutation can be assessed by studying patterns of Li–Fraumeni syndrome, an autosomal-dominant hereditary disorder that greatly increases the susceptibility to cancer and is linked to germline mutations of *TP53*. Although there is little evidence suggesting that Li–Fraumeni patients with germline *TP53* mutations are predisposed to the development of pancreatic cancer, this is likely because these patients often die of other cancers prior to the development of pancreatic malignancy [25]. Mice lacking p53 are not predisposed to PDA and appear to have normal pancreata; however, in conjunction with mutant Kras this leads to invasive PDA [26].

Deletion of *DPC4/SMAD4* at the 18q21 locus occurs in over 50% of all invasive pancreatic cancers [27], and is found in a higher fraction of tumors from patients with advanced disease [28] and portends a poor overall prognosis [29, 28]. Dpc4/Smad4 is a transcriptional regulator that forms heterodimers with other Smads and is required for a subset of TGFβ

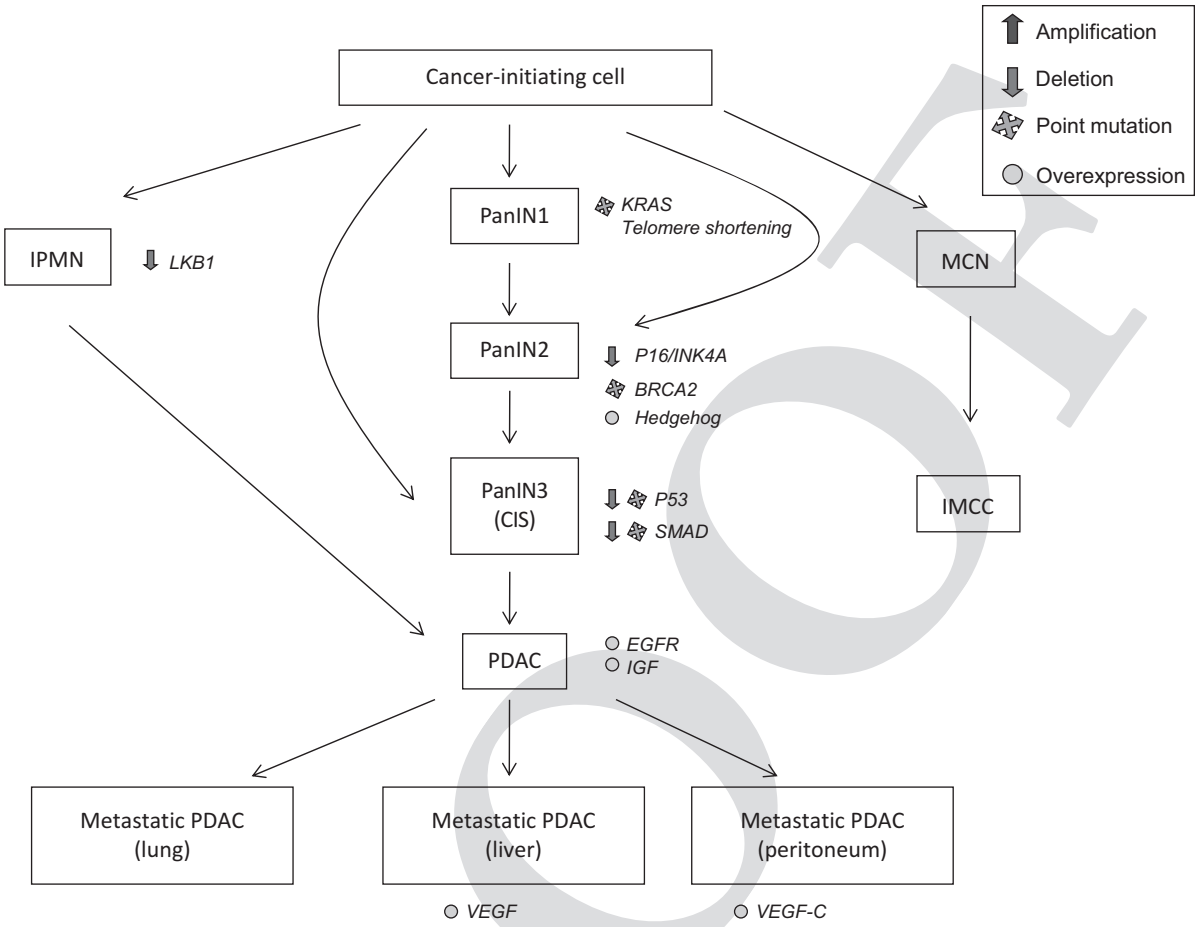

Figure 26.1 Histological and molecular patterns of pancreatic cancer progression. Distinct pathophysiological stages of pancreatic are shown along with the most common genetic alterations prevalent for each stage of disease progression. (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

signaling. Its loss in high-grade PanIN supports its role in progression rather than initiation of pancreatic cancer [24, 30].

The multimodal molecular network (MMMN) and pancreatic cancer

Because the molecular genetics of pancreatic cancer is relatively well understood, the MMMN for pancreatic cancer is perhaps more defined than for other cancers. Indeed, recent global sequencing projects at the level of genomic copy number alterations and exomic sequencing have revealed that the aforementioned "four peaks" of pancreatic cancer represent the only mutations present in over 20% of all PDA cases [31]. Considering these and other results, the mutations and signaling pathways associated with the different steps of pancreatic cancer progression are fairly well understood.

The frequency and early occurrence of oncogenic mutations in *KRAS2* makes it the obvious candidate as the gatekeeper of PDA. This is supported by the fact that mutations in Kras occur in practically every pancreatic cancer and that there is no pathological phenotype in the absence of this mutation in mouse models. Therefore a Kras module may be considered a critical event in pancreatic cancer initiation. This module could also encompass upstream receptor tyrosine kinases (RTK) that potentiate Kras signaling or downstream effector molecules such as the Raf/MEK/ERK, PI3K/Akt, or RalGEF/Ral axes. Genomic amplification or overexpression of the RTKs *EGFR*, *ERBB2*, *ERBB3* and their ligands *EGF* and *TGFA* can also occur in low-grade PanIN, suggesting that the

EGF pathway may also be important for PDA initiation [32–34]. Indeed, the small molecule inhibitor erlotinib, which inhibits EGFR kinase activity, is the only targeted therapeutic approved for the treatement of PDA [7]. Although mutations in the Raf/MEK/ERK pathway are exceedingly rare, mutations in *PIK3CA* and *AKT2* are found in 11% and 20% of pancreatic cancers, respectively [35–37]. These observations suggest that although the initiator module is predominantly associated with an activating Kras mutation, other mutations in the other pathways may serve to enhance specific aspects of this signal transduction pathway.

As previously mentioned, mutations in the major tumor suppressor genes appear later in pancreatic cancer progression. The earliest of these events is mutation of p16, which first occurs in PanIN-1B/2 and could therefore be considered a part of an intermediate module leading to the loss of cell cycle checkpoint controls. The predominant role of p16 in regulation of cdk4/6-cyclinD complexes suggests that amplification or overexpression of either kinase may be functionally equivalent to p16 inactivation. Mutations in *RB*, which could convey a similar proliferative advantage, are not found in pancreatic cancer. However, there is some evidence that pancreatic tumors express lower levels of pRb [38]. Although the only known function of p16 is regulation of cdk4/6, it is intriguing to note that FAMMM kindreds harboring mutations in CDK4 that render the kinase resistant to inhibition by p16 are not predisposed to pancreatic cancer [39]. This suggests that p16 may have an additional, as yet undiscovered, role in regulating pancreatic homeostasis.

Deletion of *DPC4/SMAD4* is another intermediate-to-late module that promotes PDA progression. TGFβ signaling, of which Smad4 is a critical regulator, has been implicated in immune evasion, epithelial to mesenchymal transition, and invasion [40]. Deficiencies in Smad4 may be mimicked by amplification or overexpression of inhibitory Smads such as Smad7 [41]. Alternatively, deletion or mutation of TGFβ receptors may elicit similar phenotypes. Indeed, incorporating either of these genes in murine pancreatic cancer models promotes accelerated disease [42, 43].

TP53 mutation, as assessed by the nuclear accumulation of the protein product, does not appear until PanIN-3 and therefore can be considered as a late module in the MMMN framework [24]. Pleiotropic effects of p53 mutation make it unlikely that one or even two mutations in downstream effector genes could recapitulate the multitude of p53-dependent activities, although mutations in genes that regulate p53, such as *HDM2* and *ARF*, have been reported. A polymorphism in the promoter of Hdm2, an E3 ligase that directly regulates p53 protein levels, results in elevated Hdm2 protein levels, reduced p53 function, and an increased risk for tumor progression [44]. Mutations in CDKN2A often affect both p16 as well as $p14^{ARF}$, a negative regulator of Hdm2 (and consequently positive regulator of p53). Interestingly, neither of these mutations result in increased p53 protein levels and therefore would not score positively by immunohistochemistry. Therefore mutations in these genes may significantly increase the number of PDA tumors that are deficient in this critical pathway.

Mutations in the *BRCA2* pathway occur in families predisposed to developing pancreatic cancer [45, 46]. BRCA2 mutations promote genomic instability and may thereby accelerate tumorigenesis through widespread chromosomal aberrations and genomic instability [47–49]. In conjunction with p53 loss of function, BRCA2 mutations abrogate gatekeeper and caretaker functionality of the late MMMN module and subsequently allow cell survival under mutagenic conditions.

Impact of high-throughput genomic approaches on systems biology of PDA

The construction of an MMMN implicated in pancreatic cancer progression was enabled by applying two separate experimental strategies to study the cancer genome. The first strategy involved now standard molecular biological approaches to identify oncogenes and tumor suppressor genes in PDA specimens. The second approach involved recent high-throughput genomic (HTG) techniques that allow the analysis of thousands of genes simultaneously. It is this second set of HTG approaches that holds the most promise for the discovery of new genetic aberrations and epistatic interactions between mutations in PDA. The ultimate aim of oncogenomic research is to identify and catalog both genetic and epigenetic changes in cancer at the genome-wide level in both temporal (e.g., oncogenic progression) and spatial (e.g., metastasis-specific mutations) dimensions. The construction of such a catalog will be pivotal for

selecting patient-tailored treatment programs [50]. From the MMMN network's perspective, the comprehensive oncogenomic information can be utilized to build exploratory and predictive mathematical models for detecting interactions between different genetic and epigenetic alterations and elucidating their impact on temporal and spatial progression of pancreatic cancer. Here we provide some examples of how new genomic approaches gave novel insights into the systems biology of pancreatic cancer.

By application of HTG techniques, current research efforts can be grouped into two categories: those that use only one HTG approach to profile molecular aberrations and those that apply several HTG techniques simultaneously to comprehensively characterize genomic, epigenomic, and physiological changes in cancer. The first category usually includes proof-of-concept applications, new technologies, and studies where an answer to a specific question is the primary concern. It includes early gene sequencing efforts to identify point mutations in either a specific protein family, such as protein kinases, or all protein coding sequences in several different cancers [51, 52]. Since the number of PDA samples in most previous studies was usually limited, most of the conclusions concerned broad mutation patterns in different cancers. For instance, Greenman et al. sequenced coding exons from 518 protein kinase genes in 210 diverse human cancers [51] and later (see below) Jones and colleagues [31] sequenced protein coding genes of 24 pancreatic cancers. When analyzing such datasets the most challenging computational task lies in identifying driver point mutations among a large number of passenger aberrations. Each somatic mutation in a cancer genome, whatever its structural nature, may be classified according to its consequences for cancer development. "Driver" mutations confer a growth advantage on the cells carrying them and have been positively selected during the evolution of the cancer. They reside, by definition, in the subset of genes known as "cancer genes." The remainder of mutations are "passengers" that do not confer growth advantage, but happened to be present in an ancestor of the cancer cell when it acquired one of its drivers. Most algorithms developed to identify relevant mutations utilize the proportion of synonymous to non-synonymous changes for particular nucleotide position to define the baseline for driver mutations.

Use of single nucleotide polymorphism (SNP) and oligonucleotide microarrays for copy number analyses is another widely used approach that can detect both copy number changes (CNCs) and copy neutral allelic losses, such as uniparental disomy. In one of the earliest applications to PDA, cDNA microarrays containing 14,160 cDNA clones were used to define copy number alterations in a panel of 24 pancreatic adenocarcinoma cell lines and 13 primary tumor specimens [53]. The high frequency of genomic instability in PDA, in part due to loss of p53, can make it difficult to identify driver CNCs. The authors of the paper developed a prioritization scheme for identifying minimal common regions, which used cross-sample information to find genes most frequently altered in many cancer samples. Sixty-four regions of recurrent copy number changes were identified by this approach that harbored genes known to play important roles in the pathogenesis of pancreatic adenocarcinoma, including the tumor suppressors $p16^{INK4A}$ and *TP53* and the oncogenes *MYC*, *KRAS2*, and *AKT2*.

Since the most informative copy number changes for functional validation of candidate genes are focal high-level amplifications and homozygous deletions, the prominent trend in array comparative genomic hybridization (aCGH) manufacturing was to increase the number and density of features on microarrays. For example, in 2003 Affymetrix, one of the largest SNP array producers, released the Mapping 10 K array, a high-density oligonucleotide array suitable for genotyping and estimation of copy numbers of 10,000 SNPs. Following that, Affymetrix has released platforms that interrogate 100 K and 500 K SNPs. More recently, Affymetrix made available Genome-wide SNP 5.0 and Genome-wide SNP 6.0 arrays, which in addition to SNPs interrogate a large number of non-polymorphic (NP) loci. The 6.0 chip interrogates 900 K SNPs and 900 K NP loci. Following the trend of applying denser arrays to study copy number changes in cancer, Harada et al. used Affymetrix 100 K SNPs arrays to profile 27 microdissected PDA samples [54]. With increased resolution, these arrays allowed more precise delineation for the physical boundaries of chromosomal breakpoints in PDA. For example, homozygous deletions at 9p21.3 (45 kb) and high-level amplifications in three regions of 8q: 8q24.13–q24.21 (2.2 Mb), 8q24.22 (177 kb), and 8q24.23–q24.3 (2.7 Mb) (19% of cases) were identified. *SCAP2* (*SKAP2*, 7p15.2) was the most frequently amplified gene (63% of cases), which has not been described in any type of cancer. Increased copy

number was also identified for *MYC* (8q24.21, 48%), *NCOA3/AIB1* (20q13.12, 44%), *KRAS* (12p12.1, 44%), *ERBB2* (17q12, 41%), and *EGFR* (7p11.2, 33%) genes. On the other hand, two tumor suppressor genes, *CDKN2A* and *CDKN2B*, were included in the locus of 9p21.3 that was deleted at the highest frequency (63% of cases). Genetic losses were also found in genes such as *DCC* (18q21.1, 48%), *SMAD4* (18q21.1, 33%), *MAP2K4* (17p12, 30%), *TP53* (17p13.1, 26%), and *RUNX3* (1p36.11, 22%). As can be seen from this study, high density arrays allowed finding both common (i.e., *MYC*) and PDA-specific (i.e., *SKAP2*) genes implicated in pancreatic cancer progression.

Jones and colleagues [31] undertook a large-scale HTG survey of mutations in 24 pancreatic cancers, providing the largest genome-wide search for new candidate genes reported to date. This study determined the sequences of more than 23 thousand transcripts representing 20,661 protein-coding genes and also searched for homozygous deletions and amplifications in the tumor DNA using Illumina Human1M-Duo SNP arrays. PDA tumors were found to contain on average 63 genetic alterations, the majority of which were point mutations. A core set of 12 cellular signaling pathways and processes were each found to be genetically altered in 67 to 100% of the tumors. Different modes of genetic alterations in these pathways were identified. First were those pathways in which a single, frequently altered gene predominated, such as KRAS signaling, and the regulation of the G1/S cell cycle transition. Some pathways had a few predominant altered genes, such as TGFβ signaling. A large number of pathways had many different altered genes such as integrin signaling, regulation of invasion, homophilic cell adhesion, and small guanine triphosphatase (GTPase) dependent signaling. Unfortunately, it is still unclear how pair-wise interactions within and between different pathways influence cancer progression. Larger sample size will be needed to identify and explain the functional consequences of interactions between different genetic mutations.

The use of DNA microarrays for studying gene expression among different phenotypic categories in PDA was the first oncogenomic technique to be applied to this type of cancer [55]. Buchholz et al. [56] conducted a large-scale expression profiling analysis of microdissected cells from normal pancreatic ducts, PanINs of different grades, and PDA from a total of 51 patients with pancreatic cancer using whole-genome oligonucleotide microarrays representing 21,329 genes in the form of optimized oligonucleotide probes. Differentially expressed genes between different stages were organised into functional categories such as development, structure, and signal transduction according to the Gene Ontology annotations with further divisions into subtypes based on tumor invasiveness. For example, within the "structure" category, clusters of benign/hyperplastic tissue-specific genes and dysplastic/neoplastic tissue-specific genes were readily distinguishable: while the former included the matrix metalloproteinases 3 and 17 (MMP3, MMP17) as well as laminin gamma 3 (LAMC3), the latter encompassed fibronectin 1 (FN1), keratin 16 (KRT16), plastin 3 (PLS3), matrix metalloproteinase 7 (MMP7), and collagen type III alpha-1 (COL3A1).

An alternative to the CNCs detection application of SNP arrays is their use in genome-wide association studies (GWAS) that aim to identify cancer susceptibility alleles [7]. Such studies provide an ample ground for the development of various statistical genetics methods for identifying association of single alleles or a subset/combination of alleles with cancer phenotype. Indeed, a GWAS study has confirmed the linkage of ABO blood types to the risk of developing PDA [57].

The second category represents oncogenomic studies that use multiple sources of HTG information. The most prominent example in this category is a recently published paper from the the Cancer Genome Atlas pilot project applied to glioblastoma [58]. The large-scale multidimensional analysis in this study assessed DNA sequence changes, copy number aberrations, chromosomal rearrangements, and DNA methylation. Much smaller studies attempted to use gene expression data from DNA microarrays to pinpoint functional consequences of copy number changes in PDA. For example, Heidenblad et al. carried out copy number analysis of 29 pancreatic carcinoma cell lines using aCGH arrays and compared the results with matching transcriptomic profiling data [59]. They showed that a strong association between DNA copy numbers and mRNA expression levels is present in pancreatic cancer, and demonstrated that as many as 60% of the genes within highly amplified genomic regions display associated overexpression. Another study by Fu and colleagues exemplified the advantage of using several HTG technologies for finding candidate cancer genes. These authors used representational oligonucleotide microarray analysis (ROMA) to identify copy number

changes in pancreatic cancer xenografts, and validated these findings using FISH, quantitative PCR, Western blotting, and immunohistochemical labeling. With this approach, they identified a 0.36-Mb amplification at 18q11.2 containing two known genes, GATA-6 and cTAGE1. However, combined genetic and transcriptional analyses showed consistent overexpression of GATA-6 in all carcinomas with 18q11.2 gain, as well as in the majority of pancreatic cancers examined (17 of 30 cancers, 56.7%) that did not have gain of this region. By contrast, overexpression of cTAGE1 was rare in these same samples suggesting GATA-6 is the true target of this copy number increase.

For ultimate understanding of oncogenic pathways in PDA and for development of targeted therapeutics, not only the use of different genomic technologies, but also utilization of appropriate model systems will be required.

Cancer cell lines in general are invaluable for studying the genetics of PDA, and some recent studies showed that the cancer genomes of a panel of human cancer cell lines reflect the genomic diversity of human cancers [60]. However, cell line models are limited by their inability to recapitulate tumor–stroma interactions. The mouse has become a model system of choice to study PDA and many other types of cancer [61]. Substantial functional genomic evidence corroborates extensive use of this model organism. For example, Maser et al. compared genomes of mouse tumor cells with genetically engineered chromosomal instability to the genomes of various human cancers and showed that there is a significant non-random number of syntenic events [62]. The observations suggest that mouse and human cells can experience common biological processes driven by orthologous genetic events during transformation. In our laboratory, mouse models of pancreatic cancer are used to find syntenic copy number changes with human PDA tissue samples (as assessed by several aCGH/SNP array platforms), followed by functional validation of potential candidate genes. Similar approaches have proven invaluable in prioritizing genes functionally relevant to melanoma, prostate cancer, and hepatocellular carcinoma [63–65].

Informatics support for systems biology of PDA

The vast and diverse information on various pathogenic alterations in the PDA MMMN requires elaborate IT support for data storage, retrieval, preprocessing, and analysis. Both PDA-specific and general cancer databases exist for storage and curation of HTG data. Cancer-specific databases store specialist knowledge of alterations in one particular cancer by combining diverse data types into one database schema, while general cancer databases allow inter-cancer comparisons. Pancreatic Expression database is an example of the earlier approach [66]. It is a data management system based on the BioMart technology, which stores pancreatic gene expression data alongside the human genome, gene and protein annotations, sequence, gene homologs, SNP, and antibody data. Interrogation of the database can be achieved through both a web-based query interface and through web services using combined criteria from pancreatic-cancer specific (disease stages, regulation, differential expression, expression, platform technology, publication) and/or public data (antibodies, genomic region, gene-related accessions, ontology, expression patterns, multi-species comparisons, protein data, SNPs).

The Oncomine database, which stores more than 28 thousand gene expression and aCGH microarrays from 41 cancer types (with more than 200 microarrays representing PDA) is the most comprehensive general cancer microarray database [67]. Oncomine creates a set of differentially regulated genes in a particular set of experiments, commonly referred to as gene signatures, and then compares this set with sets from other experiments using several different strategies including differential expression, correlation, meta-analysis, and COPA score. For instance, the COPA score is calculated by searching for gene expression profiles that display the most profound overexpression in a subset of tumors using rank-ordered statistics. A separate module, Molecular Concept Data, compares gene signatures with independently derived gene sets representing molecular pathways and other biological concepts.

Development of new mathematical tools for comparing and analyzing modules and other entities in the PDA MMMN network is another active area of research spawned by the rapid accumulation of HTG data. There are four different approaches that can be considered for mining MMMN of PDA:

1. Making prognostic and predictive estimates of pancreatic cancer progression based on the HTG data

2. Finding whether genes altered in some way between different pancreatic cancer phenotypes have overrepresentation of a particular set of features, i.e., genes in the same signaling pathway
3. Defining the topological properties of the network, such as identity of hubs or connectivity patterns
4. Integrating different HTG data types to gain comprehensive insight into pancreatic cancer progression.

Again, oncogenomic databases provide both data and programmatic access that make such approaches possible. For instance, data generated by the TCGA consortium can be retrieved and analyzed using extensive sets of database queries [68].

Development of appropriate visualization schemes is also a highly relevant activity for identifying patterns and relationships between genome-wide distribution of genetic aberrations in PDA. Pathway diagrams, networks, and gene signatures are frequently used for identification and visualization of various biological relationships and a number of customizable software tools are available for this purpose [69, 70]. Often it is necessary to create a custom data representation scheme. For example, we use copy number aberration image maps to find regions of recurrent copy number changes in different tumor samples and associate them with cancer phenotype (Figure 26.2).

Future directions in PDA systems biology research

Recent technological improvements in next-generation sequencing provide a great opportunity for rapid advancement in our understanding of genome-wide patterns of somatic aberrations in PDA. The following example illustrates the impact of sequencing technologies on oncogenomic research. Researchers reported approximately 100,000 somatic mutations from cancer genomes in the quarter of a century since the first somatic mutation was found in *HRAS*. With projected advancements in next-generation sequencing technologies, over the next few years several hundred million more will be revealed by large-scale, complete sequencing of cancer genomes [71]. With sufficient genome coverage, sequencing allows comprehensive detection of major mutation types, including copy number changes, point mutations, and genomic rearrangements, which could not have been possible with microarray technologies (Figure 26.3). For example, recent sequencing of an AML genome identified ten non-synonymous somatic mutations, of which only two

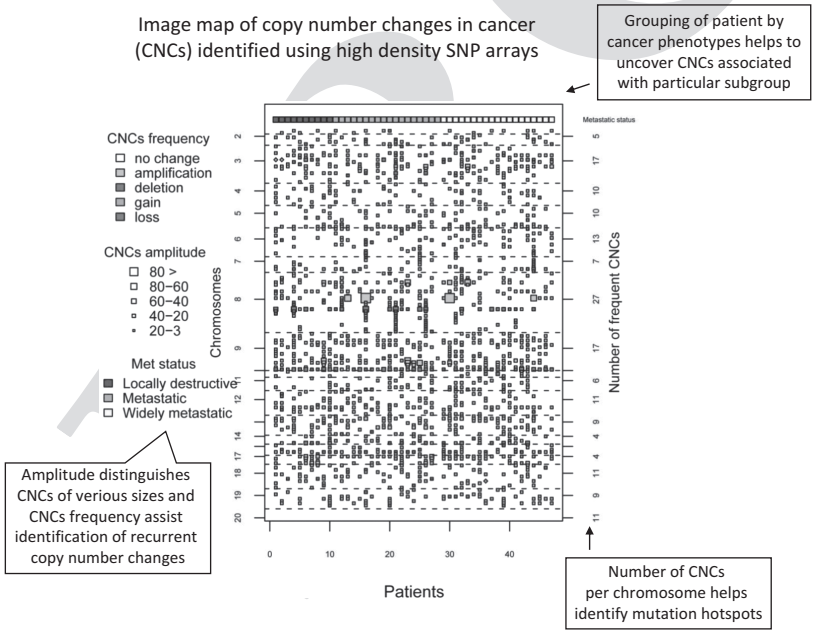

Figure 26.2 Example of an image map of recurrent copy number changes (CNCs) in cancer. Each box represents a CNC event in a particular sample. The size and color of boxes represents amplitude of copy number changes (i.e., gains versus high-level amplifications) and their size. The samples can also be associated with different cancer phenotypes (metastatic status in this example). (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

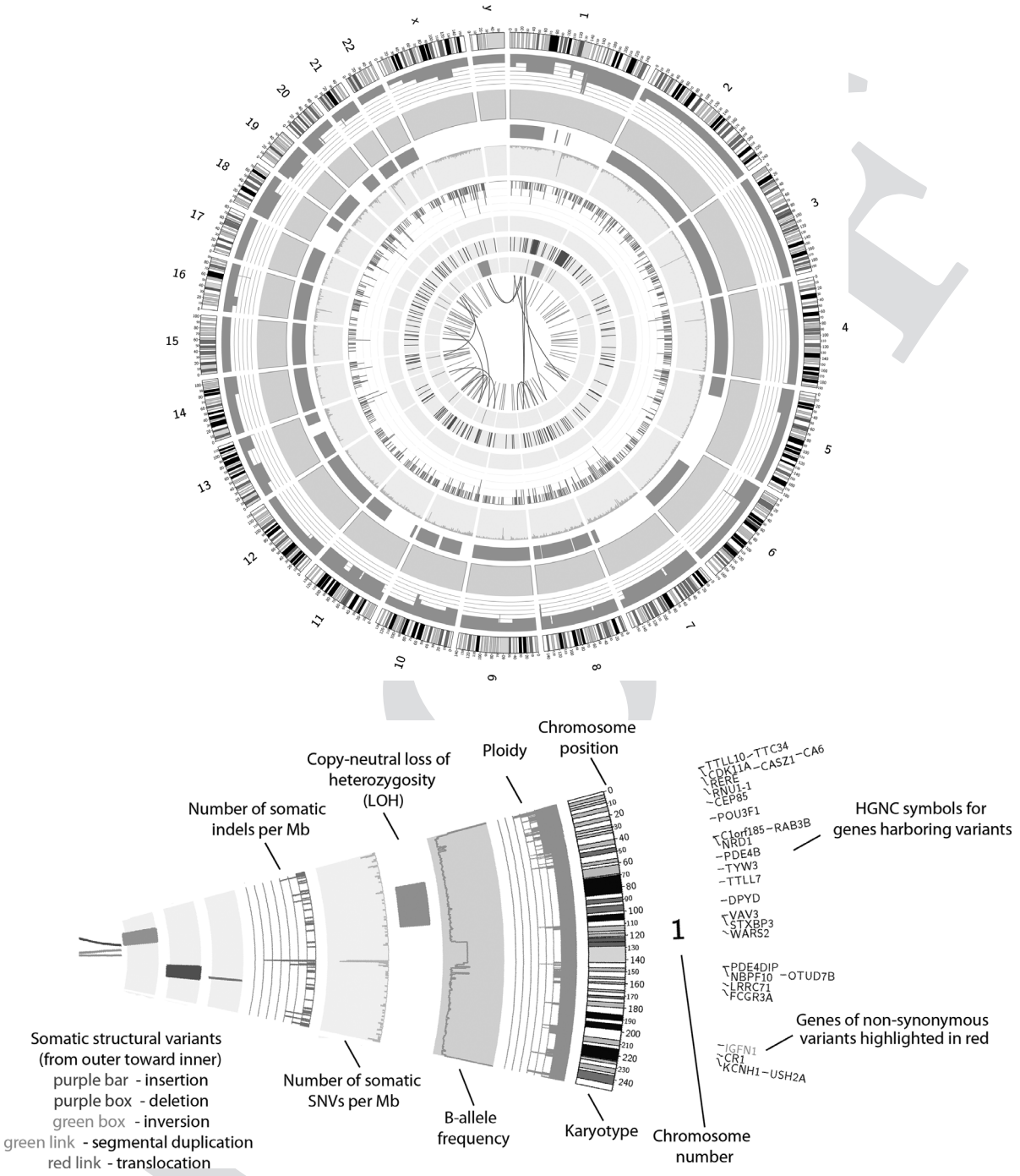

Figure 26.3 Comprehensive categorization of genetic abnormalities in cancers using next-generation sequencing technologies. The example shows part of a catalog of somatic mutations in the small-cell lung cancer cell line NCI-H2171. Individual chromosomes are depicted on the outer circle followed by concentric tracks for point mutation, copy number, and rearrangement data relative to mapping position in the genome. Arrows indicate examples of the various types of somatic mutation present in this cancer genome. (Reprinted from [71] with permission from the publisher.) (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

were previously known AML-associated mutations. The other eight somatic mutations detected were all single base changes, and none had previously been detected in an AML genome. Moreover, four among these eight somatic mutations occurred in genes not previously implicated in cancer pathogenesis, but whose potential functions in metabolic pathways suggest mechanisms by which they could act to promote cancer, thereby providing new avenues for exploration of diagnostic and therapeutic approaches for AML treatment [72].

Sequencing of the lung cancer genome, albeit at low coverage to robustly detect point mutations, in addition to copy number changes identified 103 somatic rearrangements to the base-pair level of resolution. The other advantage of sequencing is that it, on the contrary to microarray technologies, provides an unlimited dynamic range for detecting copy number changes in cancer.

Finally, characterization of the cancer cell transcriptome and active set of transcription factors, combined with genomic aberration profile, will build a comprehensive picture of cancer cell physiology. Further single-cell sequencing will uncover subclones carrying drug-resistance mutations and allow reconstruction of cancer cells' lineage [71].

We therefore expect in the next few years the appearance of a number of reports that categorize PDA genomes using next-generation sequencing. Analyses of genomic information will be only the first step in our understanding of the molecular genetics of PDA progression. The vast amount of new data will create new and exciting opportunities for systems biology research. Mining DNA sequence datasets will require application of statistical methodologies at two distinct levels. First, there will be a growing need for methods to facilitate discovery of recurrent "driver" mutations. Once the driver mutations are identified, the second-level analysis will focus on the interactions between driver mutations present in different individuals. For example, structurally different mutations inactivating protein phosphatases or activating protein kinases could lead to functionally identical cellular responses (i.e., constitutive kinase activity). The use of pathway-based information derived from literature curation and other data sources will be pivotal for reconstructing MMMN PDA networks from sequencing data, and proposing novel therapeutic approaches.

References

1. Jemal, A. et al. 2008. Cancer statistics, 2008. *CA Cancer J Clin*, **58**(2): 71–96.
2. Cancer Research UK 2009. Pancreatic cancer mortality statistics. http://info.cancerresearchuk.org/cancerstats/types/pancreas/mortality/.
3. Jemal, A. et al. 2009. Cancer statistics, 2009. *CA Cancer J Clin*, **59**(4): 225–49.
4. Wang, M.L. and Foo, K.F. 2009. Adjuvant chemoradiotherapy for high-risk pancreatic cancer. *Singapore Med J*, **50**(1): 43–8.
5. Russo, S. et al. 2007. Locally advanced pancreatic cancer: a review. *Semin Oncol*, **34**(4): p. 327–34.
6. Lowy, A.M. 2008. Neoadjuvant therapy for pancreatic cancer. *J Gastrointest Surg*, **12**(9): p. 1600–8.
7. Moore, M.J. et al. 2007. 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**(15): 1960–6.
8. Fearon, E.R. and Vogelstein, B. 1990. A genetic model for colorectal tumorigenesis. *Cell*, **61**(5): 759–67.
9. Hruban, R.H., Wilentz, R.E. and Kern, S.E. 2000. Genetic progression in the pancreatic ducts. *Am J Pathol*, **156**(6): 1821–5.
10. Maitra, A. and Hruban, R.H. 2008. Pancreatic cancer. *Annu Rev Pathol*, **3**: 157–88.
11. Caldas, C. and Kern, S.E. 1995. K-ras mutation and pancreatic adenocarcinoma. *Int J Pancreatol*, **18**(1): 1–6.
12. Lohr, M. et al. 2005. Frequency of K-ras mutations in pancreatic intraductal neoplasias associated with pancreatic ductal adenocarcinoma and chronic pancreatitis: a meta-analysis. *Neoplasia*, **7**(1): 17–23.
13. Smit, V.T. et al. 1988. KRAS codon 12 mutations occur very frequently in pancreatic adenocarcinomas. *Nucleic Acids Res*, **16**(16): 7773–82.
14. Hingorani, S.R. et al. 2003. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell*, **4**(6): 437–50.
15. Caldas, C. et al. 1994. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat Genet*, **8**(1): 27–32.
16. Schutte, M. et al. 1997. Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. *Cancer Res*, **57**(15): 3126–30.

17. Ueki, T. et al. 2000. Hypermethylation of multiple genes in pancreatic adenocarcinoma. *Cancer Res*, **60**(7): 1835–9.

18. Wilentz, R.E. et al. 1998. Inactivation of the p16 (INK4A) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. *Cancer Res*, **58**(20): 4740–4.

19. Landi, S. 2009. Genetic predisposition and environmental risk factors to pancreatic cancer: a review of the literature. *Mutat Res*, **681**(2–3): 299–307.

20. Aguirre, A.J. et al. 2003. Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev*, **17**(24): 3112–26.

21. Redston, M.S. et al. 1994. p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. *Cancer Res*, **54**(11): 3025–33.

22. Rozenblum, E. et al. 1997.Tumor-suppressive pathways in pancreatic carcinoma. *Cancer Res*, **57**(9): 1731–4.

23. Boschman, C.R. et al. 1994. Expression of p53 protein in precursor lesions and adenocarcinoma of human pancreas. *Am J Pathol*, **145**(6): 1291–5.

24. Luttges, J. et al. 2001. Allelic loss is often the first hit in the biallelic inactivation of the p53 and DPC4 genes during pancreatic carcinogenesis. *Am J Pathol*, **158**(5): 1677–83.

25. Birch, J.M. et al. 2001. Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene*, **20**(34): 4621–8.

26. Bardeesy, N. et al. 2002. Obligate roles for p16(Ink4a) and p19 (Arf)-p53 in the suppression of murine pancreatic neoplasia. *Mol Cell Biol*, **22**(2): 635–43.

27. Wilentz, R.E. et al. 2000. Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. *Cancer Res*, **60**(7): 2002–6.

28. Embuscado, E.E. et al. 2005. Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer Biol Ther*, **4**(5): 548–54.

29. Blackford, A. et al. 2009. SMAD4 gene mutations are associated with poor prognosis in pancreatic cancer. *Clin Cancer Res*, **15**(14): 4674–9.

30. Maitra, A. et al. 2003. Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. *Mod Pathol*, **16**(9): 902–12.

31. Jones, S. et al. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, **321**(5897): 1801–6.

32. Dancer, J. et al. 2007. Coexpression of EGFR and HER-2 in pancreatic ductal adenocarcinoma: a comparative study using immunohistochemistry correlated with gene amplification by fluorescencent in situ hybridization. *Oncol Rep*, **18**(1): 151–5.

33. Hermanova, M. et al. 2004. Amplification and overexpression of HER-2/neu in invasive ductal carcinomas of the pancreas and pancreatic intraepithelial neoplasms and the relationship to the expression of p21(WAF1/CIP1). *Neoplasma*, **51**(2): 77–83.

34. Yamanaka, Y. et al. 1993. Coexpression of epidermal growth factor receptor and ligands in human pancreatic cancer is associated with enhanced tumor aggressiveness. *Anticancer Res*, **13**(3): 565–9.

35. Miwa, W. et al. 1996. Isolation of DNA sequences amplified at chromosome 19q13.1-q13.2 including the AKT2 locus in human pancreatic cancer. *Biochem Biophys Res Commun*, **225**(3): 968–74.

36. Ruggeri, B.A. et al. 1998. Amplification and overexpression of the AKT2 oncogene in a subset of human pancreatic ductal adenocarcinomas. *Mol Carcinog*, **21**(2): 81–6.

37. Schonleben, F. et al. 2006. PIK3CA mutations in intraductal papillary mucinous neoplasm/carcinoma of the pancreas. *Clin Cancer Res*, **12**(12): 3851–5.

38. Huang, L. et al. 1996. Molecular and immunochemical analyses of RB1 and cyclin D1 in human ductal pancreatic carcinomas and cell lines. *Mol Carcinog*, **15**(2): 85–95.

39. Borg, A. et al. 2000. High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutation-positive melanoma families. *J Natl Cancer Inst*, **92**(15): 1260–6.

40. Bierie, B. and Moses, H.L. 2006. Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. *Nat Rev Cancer*, **6**(7): 506–20.

41. Kleeff, J. et al. 1999. The TGF-beta signaling inhibitor Smad7 enhances tumorigenicity in pancreatic cancer. *Oncogene*, **18**(39): 5363–72.

42. Ijichi, H. et al. 2006. Aggressive pancreatic ductal adenocarcinoma in mice caused by pancreas-specific blockade of transforming growth factor-beta signaling in cooperation with active Kras expression. *Genes Dev*, **20**(22): 3147–60.

43. Kuang, C. et al. 2006. In vivo disruption of TGF-beta signaling by Smad7 leads to premalignant ductal lesions in the pancreas. *Proc Natl Acad Sci U S A*, **103**(6): 1858–63.

44. Asomaning, K. et al. 2008. MDM2 promoter polymorphism and pancreatic cancer risk and prognosis. *Clin Cancer Res*, **14**(12): 4010–15.

45. Maitra, A., Kern, S.E. and Hruban, R.H. 2006. Molecular pathogenesis of pancreatic cancer. *Best Pract Res Clin Gastroenterol*, **20**(2): 211–26.

46. Jones, S. et al. 2009. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science*, **324**(5924): 217.

47. Shi, C., Hruban, R.H. and Klein, A.P. 2009. Familial pancreatic cancer. *Arch Pathol Lab Med*, **133**(3): 365–74.

48. Ozcelik, H. et al. 1997. Germline BRCA2 6174delT mutations in Ashkenazi Jewish pancreatic cancer patients. *Nat Genet*, **16**(1): 17–8.

49. Goggins, M., Hruban, R.H. and Kern, S.E. 2000. BRCA2 is inactivated late in the development of pancreatic intraepithelial neoplasia: evidence and implications. *Am J Pathol*, **156**(5): 1767–71.

50. Chin, L. and Gray, JW. 2008. Translating insights from the cancer genome into clinical practice. *Nature*, **452**(7187): 553–63.

51. Greenman, C. et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature*, **446**: 153–8.

52. Wood, L.D. et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science*, **318**: 1108–13.

53. Aguirre, A.J. et al. 2004. High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc Natl Acad Sci U S A*, **101**(24): 9067–72.

54. Harada, T. et al. 2008. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. *Oncogene*, **27**(13): 1951–60.

55. Han, H. et al. 2002. Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. *Cancer Res*, **62**(10): 2890–6.

56. Buchholz, M. et al. 2005. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. *Oncogene*, **24**(44): 6626–36.

57. Amundadottir, L. et al. 2009. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet*, **41**(9): 986–90.

58. Cancer Genome Atlas Research Network 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216): 1061–8.

59. Heidenblad, M. et al. 2005. Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: implications for the interpretation of genomic amplifications. *Oncogene*, **24**(10): 1794–801.

60. Neve, R.M. et al. 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, **10**(6): 515–27.

61. Tuveson, D.A. and Jacks, T. 2002. Technologically advanced cancer modeling in mice. *Curr Opin Genet Dev*, **12**(1): 105–10.

62. Maser, R.S. et al. 2007. Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature*, **447**(7147): 966–71.

63. Ellwood-Yen, K. et al. 2003. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell*, **4**(3): 223–38.

64. Kim, M. et al. 2006. Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell*, **125**(7): 1269–81.

65. Lee, J.S. et al. 2004. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nat Genet*, **36**(12): 1306–11.

66. Chelala, C. et al. 2007. Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets. *BMC Genomics*, **8**: 16.

67. Rhodes, D.R. et al. 2007. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*, **9**(2): 166–80.

68. *The Cancer Genome Atlas (TCGA) Data Portal Use Case Workshop* 2008, National Cancer Institute and National Human Genome Research Institute National Institutes of Health p. 5.

69. Freeman, T.C. et al. 2007. Construction, Visualisation, and Clustering of Transcription Networks from Microarray Expression Data. *PLoS Comput Biol*, **3**(10): e206.

70. Shannon, P. et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**(11): 2498–504.

71. Stratton, M.R., Campbell, P.J. and Futreal, P.A. 2009. The cancer genome. *Nature*, **458**(7239): 719–24.

72. Ley, T.J. et al. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, **456**(7218): 66–72.