5.2 The nature and development of cancer Cancer mu

5.2 The nature and development of cancer: Cancer mutations and their implications 445

ESSENTIALS Cancer is a genetic disease in which progressive accumulation of mutations in the genome of somatic cells induces abnormal bio- logical capabilities. Cancer-inducing mutations may originate from single base substitutions or large chromosomal rearrangements; but ultimately they disrupt normal cellular processes by altering protein function or disturbing the regulation of gene expression. Loss-of-function mutations in tumour suppressor genes inactivate physiological control of cell processes, whereas gain-of-function mutations directly affect physiological networks and, for example, induce pathological activation of signalling pathways. In oncology, so-called driver mutations alter key cellular processes and confer proliferative advantages to the cancer cell which thus expands preferentially as a clonal population; in contrast, mutations that also occur in cancer cells but do not confer selective advantage are termed ‘passenger’ mutations. Ten behavioural hallmarks of cancerous cells result from disruption of physiological processes and ultimately induce clonal expansion of the cancer cell, leading to clinical presentation of a frank tumour mass. These behaviours define abnormalities within the cancer cell as well as the changes in the tumour-cell environment and the host immune system. The principal aim of translational research in oncology is to utilize insights into the molecular basis of cancer rationally to develop treatments—a contemporary effort, frequently described as ‘preci- sion’ or ‘personalized’ medicine. As sequencing technologies are ap- plied increasingly to diagnostic assays carried out on tissue biopsies and body

fluids, the impact of these initiatives is declaring itself in the improved selection of treatments that are specifically targeted to driver mutations in particular tumours. For many common cancers, we are now close to defining unique sets of somatic alterations which confer a specific signa- ture of the cancer type and are also highly specific to the indi- vidual patient. The nature of DNA mutations DNA mutations are stable aberrations which give rise to cancer by altering protein function or dysregulating the control of cellular gene transcription. The classes of DNA mutation are: base substi- tution in which a single nucleotide is changed; small insertions and deletions (indels) and larger structural variants (SV), which include chromosomal rearrangements (in which there is breakage of DNA and abnormal re-joining); or changes in the copy number of DNA segments (Table 5.2.1). In addition, heritable epigenetic changes, particularly methylation of cytosine nucleotides at CpG sequences, also function as stable mutations by directly repressing gene expres- sion or binding of transcription factors to DNA. Other terms are frequently used to describe the functional effects of mutations: nonsynonymous mutations cause an alteration in the amino acid sequence of a protein, whereas silent or synonymous mutations do not change protein coding. Loss-of-function muta- tions are predicted to completely abrogate normal protein function, either by creating abnormal translational stop signals (typically sub- stitutions) or by introducing frameshift and splicing alterations. Loss-of-function mutations are found in tumour suppressor genes which have recessive effects that require mutation of both copies in a cancer cell. Gain-of-function mutations augment protein activity, for instance, by increasing enzymatic activity or changing the af- finity for the substrate of an enzyme. These are typical mutations in oncogenes which have dominant effects as only one altered allele is required. The rate of acquisition of all mutations is proportional to ex- ogenous and endogenous exposures that cause DNA damage. These exposures may include normal or physiological exposure, for ex- ample, to reactive oxygen species produced during chronic inflam- mation, as well as harmful exposures such as alcohol and cigarette smoke in the aerodigestive tract. Other important mutagenic pro- cesses include infrequent errors of DNA editing, replication, and DNA maintenance. 5.2 The nature and development of cancer: Cancer mutations and their implications James D. Brenton and Tim Eisen

 The hallmarks of cancer In a seminal paper in 2000, Hanahan and Weinberg classified the many abnormalities noted in cancer biology into six groups or hall- marks (Fig. 5.2.1). These original six hallmarks are: (1) sustaining proliferative signalling; (2) evading growth suppressors; (3) resisting cell death; (4) enabling replicative immortality; (5) inducing angio- genesis; and (6) activating invasion and metastasis. Reflecting a decade of progress in the understanding of cancer biology, Hanahan and Weinberg updated their classification in 2011 and added four further hallmarks: (7) genome instability and mutation; (8) tumour- promoting inflammation; (9) avoiding immune destruction; and (10) deregulating cellular energetics. Mutations may be described as driver events or driver mutations when they alter key cellular processes and confer clonal advantage. Many driver mutations alter genes involved in pathways implicated in the hallmarks of cancer (Fig. 5.2.1). It is important to realize that driver mutations do not neatly fit into one hallmark and may cause multiple hallmarks of cancer. The close regulation of the cell cycle is necessary to maintain normal growth and maintenance of tissues. Sustained prolifera- tive signalling may result from driver mutations in one or more of the constituents of growth and proliferation signalling pathways. These driver mutations may be found in cell surface growth re- ceptors, their ligands, or in the downstream cytosolic components of the signalling cascade. Mutations in the RAS-RAF-MEK-ERK pathway are some of the commonest in cancer and are present in approximately a third of malignancies. Some of these

mutations have been successfully targeted with mutation-specific inhibitors, for example, BRAF V600E in melanoma. However, the commonly mutated protein RAS has proved a very difficult therapeutic target to date. Successful antagonism of sustained proliferative signalling has the potential to revolutionize therapy. For example, epidermal Table 5.2.1 Types of mutations Small-scale mutations Silent mutation Silent mutation Missense mutation Nonsense mutation Change in a single nucleotide Amino acid sequence is not changed Amino acid sequence is changed Amino acid sequence is changed to a stop
codon, therefore truncating the protein Insertion Addition of one or more extra nucleotides Deletion Removal of one or more nucleotides Large-scale mutations Amplification Multiple copies of chromosomal region Deletion Loss of chromosomal region Translocation Interchange of regions from different chromosomes Inversion Reversal of the orientation of a chromosomal region Loss of heterozygosity Loss of one allele Reprinted from Ma CX and Ellis MJ (2013). The Cancer Genome Atlas: Clinical applications for breast cancer. Oncology, 27(12). EGFR inhibitors Sustaining proliferative signalling Deregulating cellular energetics Resisting cell death Genome instability & mutation Inducing angiogenesis Activating invasion & metastasis Tumor- promoting inflammation Enabling replicative immortality Avoiding immune destruction Evading growth suppressors Aerobic glycolysis inhibitors Proapoptotic BH3 mimetics PARP inhibitors Inhibitors of VEGF signalling Inhibitors of HGF/c-Met Selective anti- inflammatory drugs Telomerase Inhibitors Immune activating anti-CTLA4 mAb Cyclin-dependent kinase inhibitors Fig. 5.2.1 Hallmarks of cancer and potential therapeutic approaches. Reprinted from Cell, 144(5), Hanahan D and Weinberg RA, Hallmarks of Cancer: The Next Generation, 646–74, Copyright © 2011, with permission from Elsevier.

5.2 The nature and development of cancer 447 growth factor receptor (EGFR) inhibitors have greatly extended the life expectancy of patients with non-small cell lung cancer expressing EGFR with activating and sensitizing mutations. This example also illustrates the importance of understanding mech- anisms of resistance to therapy. After initial success with treat- ment with an EGFR inhibitor, a resistant clone arises and the disease progresses. The commonest cause of this resistance is a further mutation in the EGFR, T790M, which renders the pro- tein resistant to first-line EGFR inhibitors. It has proved possible to develop highly effective inhibitors of the T790M mutated EGFR which offer patients a substantial further period of disease control. Healthy cells possess several layers of regulatory control of pro- liferation, including control of the cell cycle at checkpoints. Evasion of growth suppressors allows cancer cells to divide but also leaves cancer cells vulnerable. Further disruption of cellular control by cyclin-dependent kinase inhibitors is effective in cancer cells by leading to mitotic catastrophe and cell death in sensitive cells. Normal cell death may be due to activation of extrinsic cellular death receptors or activation of an intrinsic pathway sensing cellular stress, such as irreparable DNA damage or lack of essential nutri- ents. These stresses are sensed by the regulatory protein p53 which alters the balance between pro-apoptotic and antiapoptotic proteins and results in cell death. Many cancer cells can resist cell death, often due to inactivating mutations in p53, which are present in up to half of all cancers. Other mechanisms of inducing apoptosis, for ex- ample, by mimicking the pro-apoptotic protein BH3, are finding an increasing role in oncology. If a cell survives a severe stress, it may irreversibly exit the cell cycle and become senescent. A characteristic feature of senescent cells, as well as of premalignant cells, is shortening of the telomeres. Cancer cells circumvent this telomere shortening by upregulating the enzyme telomerase which normally maintains the length of the telomere. This property allows cancer cells to continue proliferating and develop replicative immortality. Inhibitors of telomerase have shown activity in several tumour types. Without

developing new blood vessels, tumours are limited to a volume of 3 mm3 because they are dependent on diffusion of nu- trients and waste products. Tumour blood supply may form by a number of mechanisms including co-option of normal vascula- ture and induction of angiogenesis. Targeting of angiogenesis has proved particularly effective in tumours characterized by patho- logical activation of hypoxia inducible factor (HIF) and conse- quent upregulation of vascular endothelial growth factor (VEGF). Inhibitors of VEGF receptors have greatly improved the outlook for patients with clear cell renal carcinomas in which upregulation of VEGF is ubiquitous. Most deaths from cancer are due to metastatic spread of the tu- mour from its original or primary site to other distant sites. In order to invade and metastasize, a cancer cell must successfully complete a sequence of tasks: it must move through the extracellular matrix and basement membrane to reach a blood or lymph vessel; it must invade the vessel and then survive in the circulation in the absence of contact with other cells; it must lodge at a distant site, extravasate and invade the distant organ; it must survive and grow in its new environment and then develop a blood supply as described earlier. Key to the early steps in this sequence in carcinomas is the cell's transition from a polarized epithelial phenotype to a mesenchymal phenotype able to invade surrounding tissues. This is called the epithelial-mesenchymal transition (EMT). Upregulation of cell signals such as hepatocyte growth factor or c-Met is implicated in EMT. C-Met inhibitors show anticancer activity in several tumour types. Genome instability and mutation is a double-edged sword which serves the ability of the cancer cell to evolve, survive, and metas- tasize but also makes the cancer cell vulnerable to therapies which further limit the cancer cells ability to repair DNA damage. The gene breast cancer 2 (BRCA2) encodes one of several proteins in- volved in repair of DNA damage. When BRCA2 is nonfunctional, as is the case in some breast, prostate, and ovarian cancers, the cancer cell is particularly dependent on other DNA repair path- ways. Inhibition of such a rescue pathway would result in mas- sive DNA damage accumulating in the cancer cell and its eventual death. This is the concept of synthetic lethality. The use of poly ADP ribose polymerase (PARP) inhibitors has been highly ef- fective in BRCA2 tumours and is now being developed in a wider group of malignancies. Cells are particularly susceptible to DNA damage and mutation when dividing. States of chronic inflammation lead to increased cell turnover and proliferative stimulation. Environmental toxins such as tobacco smoke may lead to tumour-promoting inflammation and increased cellular turnover including in the stem cell compartment. Coupled with direct DNA damage, this substantially increases the risk of lung malignancy. Withdrawal of the environmental toxin by stopping smoking has a pronounced benefit even after many years of exposure. In other disease set- tings, trials of anti-inflammatory drugs have produced some encouraging results. As for other hallmarks of cancer, inflammation may both be a risk factor for cancer and present a therapeutic opportunity. It has long been recognized that a patient's immune system is able, in certain circumstances, to attack and destroy cancers. It is not uncommon for immune cells to account for up to half of tumour volume. However, cancer cells have become adept at avoiding immune destruction, for example, by downregulating cell surface antigens or by expressing molecules which inactivate or kill cytotoxic T cells, such as pro- grammed death-ligand 1 (PD-L1). Tumours differ in the extent to which they exhibit immune cell infiltration, from complete lack of immune cells (immune desert) to heavily infiltrated or inflamed. In the most exciting development in the systemic treatment of can- cers to date, drugs which inhibit the normal T-cell checkpoints have shown activity in a wide variety of tumours, including the ability to eliminate them completely. Examples include antibodies which in- hibit the T-cell checkpoints' programmed death-1 (PD-1) and cyto- toxic T lymphocyte antigen 4 (CTLA4). The likelihood of success varies between tumours and appears to be related to the degree of mutational load, the degree of inflammatory infiltrate, and

the level of PD-L1 expression in the tumour. Deregulation of cellular energetics was first described in 1927 by Warburg and colleagues. The ‘Warburg effect’ is the overuse of glu- cose by cancer cells as an energy source. The most important cur- rent application of this knowledge is the detection of even very small volumes of metabolically active tumour by [18F] fluorodeoxyglucose positron emission tomography (PET) scanners. Therapeutic ex- ploitation of deregulated cellular energetics is a field of very active research.

 Detection of cancer mutations Cytogenetic analysis Driver mutations are typically present in the earliest invasive le- sions but may also arise during subsequent divergent evolution of cells within the tumour mass (see later in this chapter). Passenger mutations do not confer a cellular advantage but also accumulate over time. The task of distinguishing new driver mutations from the background noise of passenger mutations is complex and combines bioinformatic, statistical, and functional analyses. The current tally of proven cancer genes is estimated to be approximately 400, which comprises less than 2% of all human protein coding genes. The earliest evidence for mutations in cancer cells was obtained from light microscopy studies on cancer chromosomes car- ried out at the end of the end of the nineteenth century by von Hansemann and followed by similar studies by Boveri. The Philadelphia chromosome was discovered by karyotyping studies on chronic myeloid leukaemia in 1960 and subsequent improve- ments in cytogenetic analysis led to the identification of recurrent translocations between chromosomes 9 and 22 in 1973. These dis- coveries moved cytogenetic analysis out of research laboratories into clinical practice, and over the next 20 years karyotyping was used to describe differences in chromosome number and chromosomal rearrangements (translocations) resulting in new molecular classifications, particularly in haematological and paediatric malignancies. Despite low spatial resolution and lack of molecular informa- tion, karyotyping arguably provided the first evidence for key differences in the type and number of mutations between mor- phologically similar neoplasms. These methods provided the first demonstration of numerical chromosomal instability, for instance, some colorectal cancers were shown to have normal chromosome number, in contrast to cases with prominent aneu- ploidy. Subsequent discoveries showed that this distinction was due to microsatellite instability in colorectal cancer which does not cause marked changes in chromosome number, but pro- duces a hypermutator phenotype with very large numbers of indel mutations. The development of fluorescent in situ hybridization (FISH) in the 1990s provided the first molecular evidence that specific chromosomal loci were rearranged or amplified in cancer, and was used to detect recurrent translocation between the BCR and ABL genes in chronic myeloid leukaemia (CML) and the later detec- tion of the BCR-ABL fusion oncogene. This discovery was followed by the identification of other recurrent translocations in haem- atological cancers and in some solid tumours, notably sarcomas and paediatric cancers. However, the application of karyotyping and FISH to solid tumour specimens is difficult and over the past 10 years, alternative methods using DNA microarrays or sensitive immunohistochemistry have been developed to detect important structural variants including MYCN amplification in neuro- blastoma and ERRB2 amplification in breast cancer (HER2 testing). Although karyotyping and FISH are still important laboratory assays, they will be progressively replaced by tests based on next- generation sequencing as this becomes the major technology for clinical genomics. Next-generation sequencing The main catalyst for the development of sequencing technology over the past two decades has been the international effort to cata- logue human genetic variation. DNA sequencing became feasible in 1977 with Sanger’s development of chain terminating inhibitors, which allowed the resolution of individual nucleotide

positions using radioactive gel electrophoresis of synthesized DNA prod- ucts. Automated sequencing machines first appeared in 1986 using fluorescent terminators, and were further improved a decade later by the development of capillary electrophoresis which could accom- modate up to 96 individual sequencing reactions (Fig. 5.2.2). These DNA fragmentation In vivo cloning and amplification Cycle sequencing Electrophoresis (1 read/capillary) 3'-... GACTAGATACGAGCGTGA...-5' ...CTGATC ...CTGATCT ...CTGATCTA ...CTGATCTAT ...CTGATCTATG ...CTGATCTATGC ...CTGATCTATGCT ...CTGATCTATGCTC ...CTGATCTATGCTCG G C T C G T A T C 5'-... CTGAT Polymerase dNTPs Labelled ddNTPs (template) (primer) Fig. 5.2.2 Sanger sequencing. Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology (Shendure J and Hanlee J, 2008, Next-generation DNA sequencing, Nature Biotechnology, 26, 1135–1145), copyright © 2008.

5.2 The nature and development of cancer 449 first-generation sequencing machines provided most of the data for the Human Genome Project, the first large international sequencing project. The goal of determining the entire nucleotide sequence of human DNA was a highly expensive effort as large banks of sequen- cers and significant infrastructure were needed to support semi-automated parallel processing of the huge number of DNA reads required. By contrast, next-generation (NGS) sequencing machines operate as massively parallel processing devices (Fig. 5.2.3). The Illumina sequencing method is able to array an entire size-selected library of DNA templates onto a glass slide called a flow cell. Each single DNA molecule in the library is amplified more than one million times, forming a discrete cluster, which is large enough to be imaged on the glass slide. Sequencing is performed by sequential synthesis, in which fluorescent bases are incorporated onto the DNA template in each cluster. Base calling is performed between each base in- corporation step by imaging the entire flow cell and analysing the nucleotide-specific colour at each cluster location. Next-generation sequencing provides shorter sequencing reads than the Sanger method with higher error rates and longer run times—but dramatically cheaper costs and more rapid throughput. The key determinant of cost and quality of this sequencing is the theoretical or expected coverage (also called depth of sequencing), defined as the average number of times that each nucleotide is ex- pected to be sequenced given a certain number of reads of a given length and random distribution. For comparison, the Human Genome Project achieved 6–8× average depth across the entire hap- loid human genome of 3.2 billion base pairs (3.2 giga-base pairs [Gb]). With increased error rates and shorter reads, to achieve similar whole genome sequencing fidelity with next-generation methods requires 30× depth. Cataloguing and classifying mutations The task of comprehensive sequencing of cancer genomes was finally made possible in 2000 when the Human Genome Project delivered a near-complete sequence for all human chromosomes. This identi- fied approximately 22 000 protein coding genes which covered only 1–2% of the entire genome. These data provided the essential map to enable targeted sequencing approaches to discover new candi- date cancer genes, for example, by sequencing genes from the same protein families or signalling pathways. Early pioneering studies to identify new cancer genes carried out Sanger sequencing of all pro- tein coding genes using large cancer cell line collections and small cohorts of breast and colorectal cancer specimens. This was fol- lowed by rapid development of next-generation sequencing surveys of individual cancers, including acute myeloid leukaemia and breast cancer, which have been followed by studies including essentially all tumour types and large patient cohorts. These studies use whole genome sequencing (WGS), in which all noncoding and coding regions of the genome are sequenced, or whole exome sequencing (WES) in which the protein coding gene sequences are first puri- fied away from the huge excess of

noncoding sequencing by various Fragments Add adaptors Attach to flowcell Dissociation PCR extension Bind to primer Cluster formation Sequencing Signal scanning C G A A T C G. . . T
Fig. 5.2.3 Next-generation sequencing using sequencing by synthesis method. Reprinted from Lu Y, et al. (2016) Next Generation Sequencing in Aquatic Models, Next Generation Sequencing—Advances, Applications and Challenges, Dr. Jerzy Kulski (Ed.), InTech, DOI: 10.5772/61657. Available from: https:// www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/ next-generation-sequencing-in-aquatic-models.

450 SECTION 5 Principles of clinical oncology purification methods. The advantage of WGS is that it provides a comprehensive analysis of all possible mutations (including SV and changes in intergenic regions) and is technically easier to perform. A key aim of the Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) projects has been to identify all cancer mutations while recognizing that these may be present at low prevalence (c.5%). To provide enough statistical power re- quires more than 500 cases using WES and WGS, and the current analysis suggests that in the majority of cancer genomes there are typically less than 10 mutated driver genes. The main achievement of the international sequencing projects has been to create detailed catalogues for over 50 cancer types and to make them publicly avail- able, which enables clinical translation of mutations as diagnostic and predictive biomarkers. Clinical sequencing The goal of providing a whole genome sequence for less than US$1000 was achieved in 2014 and this has provoked significant debate as to whether this technology should now replace less com- prehensive sequencing methods in the clinic. However, the principal barrier to clinical use is the complexity of the computational ana- lysis required for whole genome data and this is allied to the need to handle the large volume of data produced by the sequencing runs. The possibility of using whole genome sequencing at scale in health services is currently being addressed by many countries. In the United Kingdom, the Genomics England 100 000 genome project in patients with genetic disorders and cancer has been initiated and uses a centralized model for sequencing and analysis. Worldwide, the most frequent application of next-generation sequencing for cancer management is the use of targeted panels covering several hundred cancer genes. A clinically accredited benchtop sequencing device can generate over 5 gigabases of sequence data in nine hours. This sequencing coverage can be used simultaneously to analyse up to 200 patient DNA samples using multiplexed analysis with barcoded DNA primers, making it feasible to sequence 250 or more cancer genes at a price point of £100–200 per patient in 2016. Mutational analysis and its implications Mutational signatures reveal cancer aetiology The availability of numerous cancer genomes has facilitated com- parative analyses: these have revealed new integrative molecular classifications based on detection of mutations. The number and type of mutations show wide variation between and within cancer types (Fig. 5.2.4) and these differences correlate with grade and aetiology of the cancer. For example, the number of DNA substitu- tions and indels is very low in low-grade cancers such as carcinoid tumours and paediatric medulloblastoma, whereas in melanoma and lung cancer very high numbers of mutations are observed, reflecting the powerful mutagenic exposures of UV light and to- bacco carcinogens. Some cancer types show hypermutation with extremely large numbers of mutations, exemplified by colorectal cancers with microsatellite instability and uterine cancers with mutations in DNA polymerase epsilon exonuclease (POLE). The number and type of structural variants also differs between cancer types. A comparative analysis of over 3000 cancer genomes shows a hyperbolic relationship between the frequency of nucleotide substi- tution and structural variants across common cancers. High-grade serous ovarian and triple negative breast cancers have the greatest number of SVs, have very few

oncogenic mutations, and show fre- quent loss of tumour suppressor genes including TP53, RB1, NF1, and BRCA1/2. 1000 n = 22 20 52 134 26 23 81 227 91 57 121 13 63 214 11 394 219 20 49 181 231 76 88 35 335 179 121 100 10 Somatic mutation frequency (/Mb) 1 0.1 0.01 Rhabdoid tumour Ewing sarcoma Thyroid AML Medulloblastoma Carcinoid Neuroblastoma Prostate CLL Low-grade glioma Breast Pancreas Multiple myeloma Kidney clear cell Kidney papillary cell Ovarian Glioblastoma multiforme Cervical DLBCL Head and neck Colorectal Oesophageal adenocarcinoma Stomach Bladder Lung adeno- carcinoma Lung squamous cell carcinoma Melanoma C T C A C G T C T A T G Fig. 5.2.4 The frequency of mutation in common cancers. Dots show individual cancer cases and tumour types are ordered by median frequency shown on a log scale. Lower panel shows relative proportions of the six possible nucleotide substitutions. Reprinted by permission from Macmillan Publishers Ltd: Nature (Lawrence MS, et al., 2013, Mutational heterogeneity in cancer and the search for new cancer-associated genes, Nature, 499, 214–18), copyright © 2013.

5.2 The nature and development of cancer 451 These differences suggest that the mutational signature (also called the mutational spectrum) of a cancer can reveal the im- print of mutagenic processes accumulated over the lifetime of a cancer cell. This notion was first developed after the demon- stration that the mutagenic effect of UV light induced highly characteristic C to T mutations (C>T) in DNA, particularly at dipyrimidine sites when two cytosines are adjacent (CC>TT). This specific base-change was confirmed by the demonstration of a huge predominance of C>T mutations in squamous cell carcinomas of the skin. Other DNA mutational signatures were discovered by aggregating patterns of transition and transversion mutations (Fig. 5.2.5) discovered in the TP53 gene across many cancer types. Importantly, these signatures could also be shown to be induced in experimental systems using relevant chemical exposures. For example, the TP53 mutational spectrum in lung cancer is dominated by C>A substitutions, which were shown to be the result of misrepair of bulky DNA adducts formed by carcinogenic components of tobacco smoke. In hepatocellular carcinoma, C>A transversions were shown to be induced by en- vironmental aflatoxin exposure. 20% 10% 10% 10% 10% 10% 10% 10% 10% 0% 0% 0% 0% 0% 0% 0% 0% 20% 20% 20% 20% 20% 20% 20% 10% 0% 20% 10% 0% 0% 20% C > A C > G C > T T > A T > C T > G C > A C > G C > T T > A T > C T > G C > A C > G C > T T > A T > C T > G C > A Signature 1 Signature 4 Signature 7 Signature 10 Signature 13 Signature 16 Signature 19 Signature 22 Signature 25 Signature 28 Signature 2 Signature 5 Signature 8 Signature 11 Signature 14 Signature 17 Signature 20 Signature 23 Signature 26 Signature 29 Signature 3 Signature 6 Signature 9 Signature 12 Signature 15 Signature 18 Signature 21 Signature 24 Signature 27 Signature 30 C > G C > T T > A T > C T > G C > A C > G C > T T > A T > C T > G C > A C > G C > T T > A T > C T > G Fig. 5.2.5 Patterns of mutational signatures. Bar plots show proportions of each substitution in each trinucleotide context (labelling not shown). Courtesy of the Catalogue of Somatic Mutations in Cancer (COSMIC).

452 SECTION 5 Principles of clinical oncology These early discoveries of the links between exposures and somatic mutation were extremely important but had several limitations. First, they were only able to elucidate the dominant mutational sig- nature in a cancer and could not distinguish between multiple sig- nals from different mutagenic processes. Secondly, the use of a single gene (usually TP53) to catalogue mutational signatures can bias the statistical analysis, owing to strong selection for particular driver mutations in cancer development. For example, the commonest mu- tations in BRAF in melanoma are T to A transversions, which are not characteristic of UV crosslinking, but nevertheless are likely to have arisen from UV induced mutation. The

availability of detailed catalogues of whole exome- and whole genome sequencing from multiple cancer types has re- sulted in the discovery of many other mutational signatures, by providing thousands of passenger mutations for analysis instead of a limited number of mutations found in driver genes. Whole genome sequencing is superior to whole exome sequencing for analysing signatures because it includes noncoding sequences that comprises 99% of the normal genome, offering many more mutations for analysis. In addition, the application of signal ana- lysis methods to whole genome data can identify many signatures in a single cancer by deconvoluting the imprints of multiple mu- tational processes. Although there are six possible DNA substi- tutions based on Watson–Crick changes in DNA (Fig. 5.2.5), mutational signatures are described as trinucleotide sequences by including the 5′- and 3′-base at each site of mutation. By including the sequence context for the mutation, different mutational pro- cesses can be unambiguously discriminated. For example, a UV light-induced C>T transversion at a dipyrimidine site (NpCpC) can be distinguished from the same C>T change arising from de- amination of 5-methyl cytosine at a NpCpG site (where N indi- cates any nucleotide). Currently, detailed analysis of 7000 genomes from 30 cancer types has identified 30 distinct mutational signatures (Figs. 5.2.5 and 5.2.6). This number is likely to increase as other mutational data, including deletions and large structural variants, are inte- grated into pattern-recognition methods. Several of these signa- tures already provide strong candidate biomarkers for therapy, and it is to be expected that the rapid uptake of whole exome and whole genome methods in clinical genomics will soon lead to pre- cise characterization of tumours as they occur in routine onco- logical practice. The two most common signatures found across all cancer types arise from deamination of 5-methyl-cytosine and are related to ageing and carcinogenesis. Other signatures reveal specific expos- ures and mutational processes. Liver, uterine, and stomach cancer can have up to six distinct signatures, suggesting complex DNA damage and repair phenotypes in their aetiology. Mutational signatures can now provide strong diagnostic and aetiological information for cancer research. Whole genome Signature 1 Adrenocortical carcinoma ALL AML Bladder Breast Cervix Chondrosarcoma CLL Colorectum Glioblastoma Glioma low grade Head and neck Kidney chromophobe Kidney clear cell Kidney papillary Liver Lung adeno Lung small cell Lung squamous Lymphoma B-cell Lymphoma Hodgkin Medulloblastoma Melanoma Myeloma Nasopharyngeal carcinoma Neuroblastoma Oesophagus Oral gingivo-buccal squamous Osteosarcoma Ovary Pancreas Paraganglioma Pilocytic astrocytoma Prostate Stomach Thyroid Urothelial carcinoma Uterine carcinoma Uterine carcinosarcoma Uveal melanoma 40 7+ 4 2 5 13+ 7+ 2 5 4+ 3+ 4 6 3 4 4 10+ 7 4 5+ 6 3 3 5+ 4 5 3 7 6 6 3 6+ 2 3 3 11+ 4 5 8 6 3 22 3 7 40 17 3 2 2 6 2 1 22 2 2 1 6 3 1 2 1 2 1 1 1 4 1 1 1 2 Signature 2 Signature 3 Signature 4 Signature 5 Signature 6 Signature 7 Signature 8 Signature 9 Signature 10 Signature 11 Signature 12 Signature 13 Signature 14 Signature 15 Validated mutational signatures Signature 16 Signature 17 Signature 18 Signature 19 Signature 20 Signature 21 Signature 22 Signature 23 Signature 24 Signature 25 Signature 26 Signature 27 Signature 28 Signature 29 Signature 30 Other signatures Mutational signature present Total validated mutational signatures in a cancer type Total cancer types in which a signature is operative
Fig. 5.2.6 Occurrence of mutational signatures by cancer type. Courtesy of the Catalogue of Somatic Mutations in Cancer (COSMIC).

 sequencing studies have revealed distinct patterns of smoking- related signatures by analysing 5000 tobacco-associated cancers from smokers and nonsmokers. Cancers from smokers had sig- nificantly higher numbers of base substitutions and lung cancers in smokers had higher numbers of copy number aberrations. Signature 4, which

represents exposure to the tobacco carcinogen benzo[a] pyrene, was most prominent in lung and laryngeal cancers, where epithelia are directly exposed to tobacco smoke but was also found in lung cancer from nonsmokers, suggesting contributions from sec- ondary smoke inhalation. Studies on radiation-associated cancers show that the additional induced burden of mutations is relatively low, but in contrast to other mutational signatures, the distribution of mutations is evenly distributed across the genome, and does not show bias with replication timing, sequence complexity, or GC base content. Radiation-induced tumours also show a very rare aberra- tion of balanced inversions which may uniquely identify these can- cers. These data show that the stochastic nature of radiation is not affected by the chromatin context of DNA but induces relatively in- frequent additional mutations, which may explain the low absolute risk of radiation-induced cancer. Cancers with microsatellite instability can be identified by a strong preponderance of signature 6 (Fig. 5.2.6), which reflects very high numbers of substitutions and 1-base pair indels in nucleotide repeats. This signature is most commonly seen in colorectal, uterine, and stomach cancer and is strongly associated with loss of mis- match repair genes. Detection of this signature has important thera- peutic implications as it indicates patients who are likely to benefit from with immune checkpoint inhibitors, owing to increased fre- quency of expressed neoantigens in the tumour microenvironment. Signature 3 is frequent in breast, ovarian, and pancreatic cancers and is strongly correlated with homologous recombination deficiency (HRD) and is commonly caused by mutation in BRCA1 and BRCA2. Cancers with HRD can also be identified by different signatures based on the pattern of structural variants as they characteristic- ally have large indels (≤50 bp) with overlapping microhomology at breakpoint junctions. Patients with HRD show increased sensitivity to platinum-based chemotherapy and poly (adenosine diphosphate [ADP]) ribose polymerase inhibitor (PARPi) therapy. Sequencing for germline and somatic mutations in BRCA1 or BRCA2 (and other less frequent mutations in homologous recombination genes) cannot identify all patients with HRD, whereas the finding of a dominant signature 3 is strongly suggestive of possible benefit from PARPi therapy. The analysis of signature 3 across 32 cancer types has shown that 7–12% of cases of gastric cancer have features of HRD. As gastric cancer is the second commonest cancer worldwide, there may be significant benefits from routine profiling for mutational sig- natures of HRD or MSI. Mutations offer targets for personalized therapies Most targeted therapies currently in the clinic are designed to dis- rupt gain-of-function properties in cancer cells and these com- monly overlap with the hallmarks of cancer. These effects are often mediated by oncogenic mutations that induce constituent activity of the respective protein. They are often driver mutations, and the term ‘oncogene addiction’ is used to describe the dependency of the cancer cell on the induced gain of function, hence targeting of these mutations may have dramatic effects on cancer growth (Table 5.2.2). This is exemplified by the development of therapeutics against BRAF mutations in malignant melanoma. The BRAF gene encodes a serine threonine kinase and is a member of the Raf family of growth Table 5.2.2 Overview of relationships between cancer genotypes and their predicted responses to targeted therapy Cancer type Genotype Therapy Colorectal cancer Mutant KRAS Cetuximab/Panitumimab (no response) Chronic eosinophilic leukaemia (CEL) PDGFR translocations Imatinib Chronic myeloid leukaemia (CML) BCR-ABL translocation Imatinib Resistant CML Mutant BCR-ABL translocation Dasatinib Ponatinib Bosutinib Gastrointestinal stromal tumour (GIST) Mutant KIT Imatinib Breast cancer HER2 amplification Trastuzumaba Pertuzumaba ado-trastuzumab emtansinea Melanoma Mutant BRAF Vemurafenib Dabrafenib Trametinib Dabrafenib/Trametinib Myelofibrosis Mutant JAK2 Ruxolitinib Non-small cell lung cancer Mutant EGFR Erlotinib Gefitinib Afatinib Non-small cell lung cancer ALK translocation Crizotinib Ceritinib Non-small cell lung cancer ROS1 translocation Crizotinib a Eligibility is not

strictly on genomic amplification of HER2, as strong HER2-positivity by immunohistochemistry is also an eligibility criterion for treatment. Reprinted from Tursz T and Bernards R (2015). Hurdles on the road to personalized medicine. Molecular Oncology, 9(5), 935–9. Published under a Creative Commons Attribution (CC BY) License.

 promoting proteins. Activating mutations in BRAF were first iden- tified by the Cancer Genome Anatomy Project from melanoma cell lines. Sequencing studies on clinical specimens showed that 50–70% of malignant melanoma cases have the V600E BRAF mutation, and this is also found in 8–12% of colorectal cancers and less commonly in other cancers. The V600E mutation induces constitutive activa- tion of the BRAF kinase, while successful pharmacological targeting of the kinase CP-pocket was achieved with the clinical develop- ment of the drug vemurafenib. Use of vemurafenib in patients with metastatic melanoma has shown dramatic responses in over 70% of patients treated, but resistance to single agent therapy develops by activation or mutation of other pathways that circumvent BRAF in- hibition, the implications of which are discussed later. Mutational analysis measures cancer evolution Genomic studies carried out over the past 20 years have revealed considerable intratumoural genetic heterogeneity and strong evi- dence of evolutionary selection of subclonal tumour populations during treatment, as well as during metastasis. These studies were initially carried out in paediatric acute lymphoblastic leukaemia, but similar processes of clonal evolution and selection have now been confirmed in many epithelial tumours. This has led to the clinical realization that profiling a single sample of an individual's cancer may not represent a comprehensive depiction of all driver genes. In addition, reliance on the original diagnostic cancer sample cannot provide accurate information about subsequent somatic changes contributing to therapy resistance. Optimal decision-making for personalized medicine now re- quires contemporaneous genomic information from repeat or se- quential biopsies. The use of image-guided biopsies, particularly using ultrasound with 14 G or 16 G cutting needles, can provide high-quality samples for next-generation sequencing. However, biopsy may be difficult at metastatic sites such as bone or abdom- inal lymph nodes, and the expense and discomfort for the patient may limit repeatability over their treatment. Mutational profiling of cell-free DNA in blood offers an alternative 'liquid biopsy' and is being rapidly developed for clinical use in many cancer types and for different applications, from diagnostic information to detection of emerging resistance mechanisms. Detecting cancer mutations in cell-free DNA Cell-free DNA was first demonstrated in blood in 1948 and shown to be increased in the serum of cancer patients in 1977. The possi- bility that it could be used as a cancer biomarker was established in the early 1990s by the detection of cancer mutations in plasma, stool, and sputum as well as other body fluids. Most cell-free DNA in plasma originates from normal haemopoietic cells and in healthy individuals the concentration of plasma DNA concentrations range between 1 and 10 ng/ml. However, these levels are dynamic and may be altered by disease processes causing tissue injury such as trauma or stroke, as well as physiological changes during pregnancy and physical exercise. Cell-free DNA in blood is fragmented into short fragments of approximately 166 base pairs, which likely reflect the length of DNA strands in close contact with the nucleosome pro- teins in chromatin. Deep sequencing of cell-free DNA can also pre- dict likely gene expression in the parental cells by predicting sites of DNA localization to nucleosomes. Cancer circulating tumour DNA (ctDNA) molecules are typically shorter than normal DNA at around 140 base pairs, although the reasons for this are unknown. The mechanisms determining how DNA enters the bloodstream as ctDNA have not been fully defined, but are most likely caused by apoptosis of tumour cells. The half-life of ctDNA is between 16 min-

utes and 2.5 hours and this, together with the wide dynamic range and relatively low cost of ctDNA assays, make cell-free DNA a highly attractive diagnostic and tumour response indicator. By contrast, the detection of circulating tumour cells is more difficult and has higher expense. Critically, often only low numbers of circulating tumour cells can be detected from a blood sample which limits potential diagnostic and prognostic information. Clinical evidence that ctDNA can indeed be used as a ‘liquid’ biopsy that can replace tissue biopsy has emerged from improvements in the diagnostic pathway for lung cancer (Fig. 5.2.7). Lung Size of clone Time Clone 1 Clone 2 Clone 3 Surgery (or other) Cancer detection: screening or earlier diagnosis Molecular profiling or prognostication Detection of residual disease Monitoring response Monitoring clonal evolution Quantitative analysis • Disease staging • Response monitoring • Prognostication Genomic analysis • Mutation profiling • Treatment selection • Monitoring clonal evolution Serial liquid biopsies Treatment selection Treatment 1 Treatment 2 (a) (b) Fig. 5.2.7 Applications of circulating tumour DNA for cancer diagnosis, prognostication, detection of minimal residual disease and monitoring clonal evolution. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer (Wan JCM, et al., 2017, Liquid biopsies come of age: towards implementation of circulating tumour DNA, Nature Reviews Cancer, 17, 223–38), copyright © 2017.

5.2 The nature and development of cancer 455 cancer is an important exemplar for development of ctDNA as- says owing to the need to identify patients who can benefit from targeted therapy and the relative difficulty of obtaining tissue bi- opsies. Existing CT-guided and bronchoscopic biopsy have a comparatively high failure rate and costs. The first FDA and EMA approvals for the use of ctDNA in lung cancer are for assays based on real-time polymerase chain reaction (rtPCR) to detect common oncogenic mutations in EGFR. These assays depend upon allele- specific primers that generate different fluorescent PCR products that discriminate between mutations (deletion of exon 19, L858R, or T790M) from wild-type sequences and can be used for tumour specimens or ctDNA. These assays are now in clinical use for the selection of gefitinib, erlotinib, and osimertinib therapy and regu- latory approval now allows a therapeutic decision to be made based solely on a ctDNA result if a tumour sample is not available. These advances widen the availability of therapy to lung cancer patients who may be too frail for biopsy or who have a tumour that is not easily accessible. The rapid uptake of ctDNA assays in lung cancer trials has also provided strong evidence for utility in the clinic that will be relevant to other cancer types. Over 4000 patients have had ctDNA results in lung cancer trials, and meta-analysis of these data show an overall sensitivity of 60% and specificity of 94% for detection of EGFR mu- tation. Direct comparison between plasma ctDNA and lung cancer biopsies in over 650 patients showed a sensitivity of 66% and a spe- cificity of 100%. A key question now is how effective ctDNA assays will be for the detection of secondary resistance mutations during therapy and on follow-up. Perhaps not surprisingly, the sensitivity of detecting the classical EGFR resistance mutation T790M is less good, reflecting lower abundance of revertant mutations and the effects of tumour heterogeneity. In colorectal cancer the presence of KRAS mutation is a strong predictor of resistance to the EGFR inhibitor cetuximab. Use of sequential ctDNA assays for KRAS has shown the emer- gence of clonal populations marked by KRAS mutations that are resistant to cetuximab therapy. In addition, these low-frequency populations can be demonstrated at diagnosis in some patients, which are selected to become the predominant population with therapy. There are now intensive efforts to improve the sensitivity of next-generation sequencing methods for ctDNA and to widen the number of genes included in these assays. These advances are likely to overcome current sensitivity limitations by better discriminating between early driver mutations (sometimes called ‘stem mutations’), which have higher abundance in plasma and the rarer revertant mutations. Clinical trials are now needed to

address whether outcomes and response rates are different for patients treated on the basis of ctDNA and tumour monitoring. These impacts may further change therapeutic approaches. A key challenge will be the ability to detect ctDNA in patients with early stage disease as ctDNA strongly correlates with tumour volume and can vary by 100–640-fold between stage I and stage IV disease. See Chapter 3.10 for broader discussion of the development and uses of circulating DNA for molecular diagnostics. Conclusion The development of precision medicine using DNA sequencing has had several dramatic successes in the past decade, notably with the targeting of BRAF and EGFR. There are grounds for considerable opti- mism about the wider applicability of sequencing technology in clinical practice. An absolute requirement for the wider development of clin- ical sequencing will be the need to use fresh or frozen specimens from biopsy and surgical procedures because—although current technology can provide high-quality sequencing data for multiple genes using DNA extracted from formalin-fixed paraffin-embedded tissues—the DNA from these tissues is degraded and not suitable for whole genome sequencing. As whole genome sequencing becomes cheaper and the utility of integrative genomic signatures becomes established, the rou- tine pathological processing of fresh or frozen material will be essential. The second main challenge will require clear recognition that re- sistance to conventional and targeted therapies often reflects evolu- tionary or selective pressure, which induces expansion of new clonal populations of tumour cells. This observation immediately prompts investment in suitable infrastructure for sequential or repeat biopsy of patients undergoing routine standard of care treatment, and the wider use of noninvasive methods such as circulating tumour DNA. Finally, despite the impending arrival of thousands of cancer gen- omes, we are faced with major challenges for data interpretation, sharing of the data with patients, and more widely for increased knowledge and the training of healthcare staff to interpret the clin- ical meaning of genomic change. These challenges will, in part, be addressed by an increasing focus on clinical trials based on the molecular targeting of patients har- bouring particular mutations in genes implicated in cancer develop- ment, and also the rapid dissemination of positive or negative results of such trials. It is difficult, however, to conceive how even dedicated oncology specialists will be able to keep abreast of all potential muta- tions, functional changes, and therapy. One solution to this challenge may be wider data sharing and the use of computerized decision sup- port tools. At the same time, it will be critically important to have high-level clinicians working in molecular pathology, as these indi- viduals with specialized knowledge will be at the nodal point of most clinical decisions based on tumour mutation analysis. FURTHER READING Alexandrov LB, et al. (2013). Signatures of mutational processes in human cancer. Nature, 500, 415–21. Diaz LA, Jr., et al. (2012). The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature, 486, 537–40. Hanahan D, Weinberg RA (2011). Hallmarks of cancer: the next gen- eration. Cell, 144, 646–74. Helleday T, Eshtad S, Nik-Zainal S (2014). Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet, 15, 585–98. Kucab JE, et al. (2019). A compendium of mutational signatures of en- vironmental agents. Cell, 177, 821–36. Misale S, et al. (2014). Resistance to anti-EGFR therapy in colorectal cancer: from heterogeneity to convergent evolution. Cancer Discov, 4, 1269–80.

Revision #1
Created 2026-01-22 16:44:27 UTC by Omar Ayman
Updated 2026-01-22 16:44:27 UTC by Omar Ayman