CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

Part 1 Introduction to modular organization of the networks of gene functions and cancer

Chapter 1

Systems biology of cancer progression

Sam Thiagalingam

Introduction

Heterogeneity in the genetic and epigenetic alterations of cancers that exhibit similar functional properties during the various stages of cancer progression, including the terminal metastatic stage, has remained as the major challenge to effective diagnosis, prognosis and therapeutic efforts. While there has been significant progress in cataloguing the various genetic and epigenetic alterations with the advent of expanding new high-throughput technologies, streamlining the available and emerging data into a coherent scheme of events depicting drivers, the connectors and the conductors that form multi-modular molecular networks (MMMN) of cancer progression culminating in tumors, requires novel strategies. The ultimate goal of cancer research should be to take advantage of the parallel progress made through both experimental and computational approaches and integrate the data from these fronts using systems biology to generate MMMN cancer progression models. Such models can be cancer specific and can be functionally definable in terms of disease stage to help design biomarker screening tests for effective diagnosis/prognosis and the development of personalized cancer therapies.

Background

Cancer is a genetic and epigenetic disease, which manifests functional properties of target tumor cells at different stages due to the accumulation of specific combinations of alterations. The number of alterations required to assume a given stage of cancer may vary within and between certain types of cancer. While modelling cancer progression has been attempted at various times, the first breakthrough came with the study of the genetics of colon cancer progression, which depicted multiple stages of the disease [1, 2]. Since then, despite an increase in the wealth of knowledge that has emerged on the types of alterations associated with specific cancers as a result of comprehensive profiling and next-generation sequencing (NGS) strategies to decipher genetic and epigenetic alterations, seemingly insurmountable complexity has prevented the streamlining of the various changes into coherent and definable stages, which still awaits the development of novel strategies to make progress.

Multi-modular molecular networks of cancer progression depict heterogeneity in genetic and epigenetic alterations

The lack of consistent and defined genetic and epigenetic alterations affecting a specific set of gene(s) in the majority of sporadic cancers with similar histologic subtypes and stages poses a challenge in understanding the molecular basis for the heterogeneity of molecular aberrations. The inconsistency in these profiles of molecular targets not only imposes a dilemma to gaining a clear understanding of the disease but also complicates efficient early diagnosis, prognosis and strategies for treatment modalities for cancers. To address this challenge faced by the cancer research community, I proposed a strategy for the formulation of a detailed framework known as an MMMN cancer progression model as a road map to dissect the complexity inherent to cancer (Figure 1.1) [3]. This model predicts that cancer initiation and progression are mediated by dysregulation/inactivation of a series of interconnected functional sub-network modules.

Systems Biology of Cancer, ed. S. Thiagalingam. Published by Cambridge University Press. © Cambridge University Press 2015.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

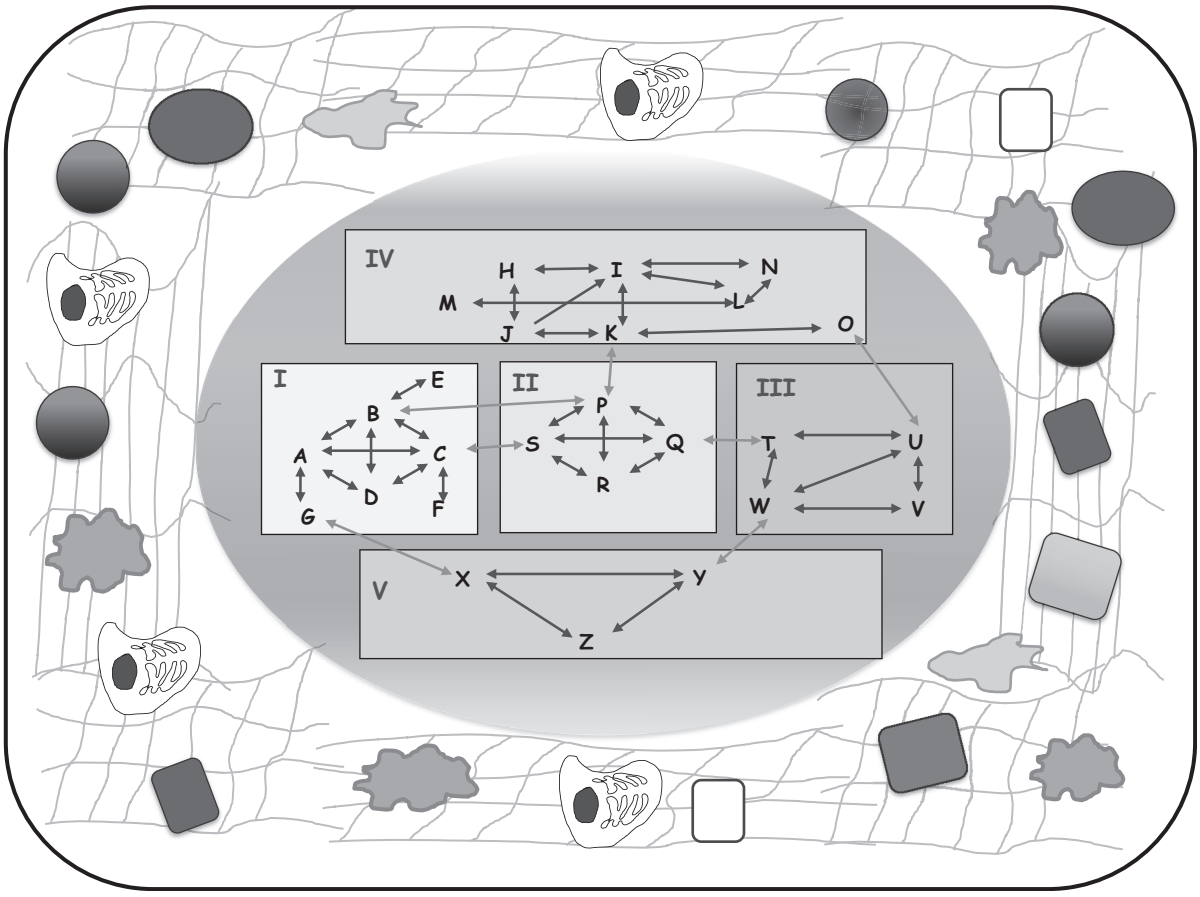

Figure 1.1 A cascade of aberrant network modules defines the multi-modular molecular network (MMMN) model for cancer progression. An MMMN cancer progression model predicts that aberrant activations/inactivations of functional modules of networks in a series of steps would be necessary to elicit properties of metastatic cancer. In this model, cancer initiation is mediated by inactivation of the gatekeeper network module (e.g., module I). We predict that the gatekeeper function is mediated by an interconnecting network of pathways (axes). Dysregulation/inactivation of the gatekeeper module predisposes the cells to become more receptive and susceptible to acquiring additional neoplastic alterations, which occur in a series of modular (modules II, III, IV, etc.) inactivations or hyper-activations leading to intermediate and late carcinoma and finally to the metastatic stage. Modules II and III in this model represent the intermediate stages of tumor progression. The terminal module may represent the metastatic stage (module IV). The fact that there could be alternate target genes in any one of the modules of the network could explain why there is often genetic/epigenetic heterogeneity in multi-step cancer progression resulting in similar histologic subtypes of cancer. In this model, the double-headed light and dark arrows represent intra- and extra-modular connections, respectively. The alphabetical letters represent specific genes or functional protein–protein and protein–DNA interactions that are nodal points/driver alterations in each network. While the modular organization depicted inside the inner oval represents the alterations within the target tumor cells, the web structures that represent the extracellular matrix and the surrounding stromal cells constitute the tumor microenvironment. (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

The provision in the MMMN cancer progression model, which defines a cascade of events encompassing multiple targets within each module, is that one or more alternate target gene(s) could alter the functionality of each of the specific modules. This provides a molecular basis for the genetic and epigenetic heterogeneity that is observed during the progression of tumors that exhibit similar pathological characteristics (Figure 1.1). Furthermore, the absence of consistent alterations in specific gene(s) in sporadic cancers, and in cancers that are primarily induced by environmental effects to generate neoplastic precursor cells, could be predicted to occur via inactivation/overactivation of multiple alternate gatekeeper gene(s) that act in one or more interconnected axes of events, within a defined sub-network in a module of the global network (Figure 1.1). Thus the first network module that becomes inactivated leading to

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

the initiation of cancer is the gatekeeper functional unit [4]. The cancer precursor cells harboring an inactivated gatekeeper module become receptive to additional genetic and epigenetic alterations that occur in interconnected but defined modules of sub-networks representing multiple stages, leading to the development of advanced and terminal stages. Therefore the functional inactivation or aberrant hyper-activation of network modules occurs in a series of events that advance the tumor from the early to late stages of cancer. It is also noteworthy that overlaps in the functional contributions of the specific gene alterations may be responsible for simultaneous dysregulation of different modules of cancer progression. While any alteration capable of inactivating/dysregulating a specific sub-network module could occur at any time, its effect will be fully realized to manifest the corresponding cancer stage only when the preceding module(s) have also become inactivated/dysregulated. Thus the rates at which tumor evolution occurs and the time required for the transition from an early to a later stage of cancer will be dependent upon the preexisting genetic and epigenetic alterations (familial or sporadically acquired) and the tumor microenvironment. This notion is also consistent with an accelerated cancer progression when there is a preexisting inherited alteration that corresponds to a specific module as it has been observed with familial cancers. Despite the possibility that the overall phenotypic effects elicited by the target tumor or tumor precursor cells could be influenced by the surrounding cells and/or extracellular matrix (ECM) components, the epigenetic and genetic alterations in the resident target cells are a prerequisite for the effects caused by the microenvironment and surrounding stromal cells [5].

While interdependent interactions of genes and proteins may consist of physical interactions among proteins, representing inter- and intracellular communications and their binding to DNA elements (e.g., transcription factors, histones harboring specific modifications, etc.) and mRNAs/regulatory RNAs (e.g., miRNA, lincRNA, etc.), there could also be metabolic networks of biochemical reactions that involve distinct substrates and products. The modular organization of the various stages of cancer progression consisting of interconnected networks of events also suggests that changes in alternate targets that render similar functional status can lead to the acquisition of drug resistance. Thus developing drugs that target combinations of distinct landscapes of alterations would be necessary for clinical decision making and to select therapies that increase therapeutic efficacy [6].

Driver versus passenger alterations

Cancer phenotypes are driven by gain-of-function alterations as seen with oncogenes such as the *AKT1*, *ALK*, *BRAF, CTNNB1*, *DDR2*, *EGFR*, *ERBB2*, *FGFR1*, *IDH1*, *IDH2*, *KRAS, MDM2*, *MITF*, *MYC*, *MYCN*, *MYCL1, NKX2.1*, *PIK3CA, REL* and *SOX2* and/or loss-of-function alterations as frequently observed with specific tumor suppressor genes such as the *APC*, *BMPR1A*, *CDH1*, *CDKN2A*, *NF1*, *NF2*, *MAP2K4*, *MLH1*, *MSH2*, *PIK3R1*, *PTEN*, *RB1*, *SMAD4*, *SMARCB1* and *TP53*, mediated by either genetic or epigenetic changes [3, 6]. A "20/20 rule," which requires at minimum >20% of the observed missense mutations at recurrent positions in an oncogene, and >20% of inactivating mutations for a tumor suppressor gene has been proposed [7]. The gene alterations that provide a selective advantage during the evolution of a tumor are regarded as the "drivers" while the alterations that are coincidental in their appearance and do not play a role in the cancer progression are termed the "passengers" [6, 8]. While, traditionally, genetic changes are regarded as the drivers and epigenetic alterations as the passengers, there is accumulating evidence for either type of alteration to be passengers or drivers [9]. It is also noteworthy that not all mutations in the same gene are drivers as exemplified by *APC* mutations in colorectal cancer [2, 7]. Furthermore, some driver genes are more frequently mutated and referred to as the "mountains," while others, despite their importance, are less frequently mutated and are known as the "hills," thus shaping the landscape of genetic alterations during cancer progression [10].

In the MMMN model for cancer progression, the driver genes represent the nodal points and activation of a single module could be effected by one, or possibly a few nodal gene alterations [3]. For example, of the more than one hundred pathways aberrantly regulated in breast cancer, several involved phosphatidylinositol 3-kinase (PI3K) signaling, with *PIK3CA* as the most frequent target and others such as *GAB1*, *IKBKB*, *IRS4, NFKB1*, *NFKBIA*, *NFKBIE*, *PIK3CB*, *PIK3CG*, *PIK3R1*, *PIK3R4*, and *RPS6KA3* as other potential targets [10, 11]. These observations are consistent with the molecular heterogeneity involving

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

aberrations in alternate target genes in modules of the MMMN model for cancer progression [3].

Emerging MMMN models

Being an autonomous complex genomic disease, cancer presents its characteristics in any group of representative cells and subtypes, based on genetic and epigenetic signatures that are often under the influence of microenvironmental effects. There has been significant progress made in visualizing these signatures through the application of genomic technologies to decipher their functional effects at the level of individual genes, the genome, and the pathways and networks of signaling events.

For example, breast cancer has a well-established genetic component exhibiting a greater than ten-fold risk in individuals harboring familial rare mutations in *BRCA1*, *BRCA2*, *TP53* and *PTEN* but elicit at least 18 morphologically distinct tumor types according to the World Health Organization. Recently, it has been classified into six different intrinsic subtypes, which harbor characteristic gene alterations: luminal A (*CCND1*, *ESR1*, *FOXA1*, *GATA3*, *KRT8*, *KRT18*, *LIV1*, *MAP3K1*, *PIK3CA*, *TFF3* and *XBP1*), luminal B (*ESR1*, *FOXA1*, *GATA3*, *KRT8*, *KRT18*, *LAPTM4B, SQLE, TFF3* and *XBP1*), HER2-enriched [*ERRB2* (*HER2* or *Neu*) and *GRB7*], basal-like (*CDH3*, *FABP7*, *ID4*, *KRT5*, *KRT17*, *LAMC2* and *TRIM29*), normal breast-like (*AQP7*, *CD36*, *FABP4*, *ITGA7* and *PTN*) and claudin-low (*ALDH1*, *CD29*, *CD44* and *SNAI3*) based on genomic studies [12–14]. Additionally, multiple technology platforms such as mRNA expression profiling, DNA copy number arrays, massively parallel sequencing as well as the high information content assays to probe DNA methylation, miRNA expression and protein expression, have been used to assess the various abnormalities in the cancer state. These efforts identified mutations previously implicated, in breast cancer: *AKT1*, *BRCA1*, *CDH1*, *GATA3*, *PIK3CA*, *PTEN*, *RB1* and *TP53*; and in other cancers: *APC*, *ARID1A*, *ARID2*, *ASXL1*, *BAP1*, *KRAS*, *MAP2K4*, *MLL2*, *MLL3*, *NF1*, *SETD2*, *SF3B1*, *SMAD4* and *STK11*. Interestingly, new lesions were also identified for the first time in breast cancer: *AFF2*, *AKT2*, *ARID1B*, *CASP8*, *CBFB*, *CCND3*, *CDKN1B*, *MAP3K1*, *MAP3K13*, *NCOR1*, *NF1*, *PIK3R1*, *PTPN22*, *PTPRD, RUNX1, SF3B1, SMARCD1* and *TBX3* [15, 16]. Furthermore, while there were cancer subtype specific mutations, only three genes (*GATA3*, *PIK3CA* and *TP53*) exhibited recurrent mutations in >10% of the breast cancers confirming the complexity and heterogeneity in the profiles of alterations that contribute to the formation of each tumor [3, 15, 16].

Similar efforts to catalogue driver genes involved in other cancers are also emerging at this time. The catalogue of genomic alterations in the various cancers are generated using high-throughput technologies at several major institutions such as the Broad Institute and the Johns Hopkins University and through the coordinated efforts of the Cancer Genome Atlas (TCGA) project in the Unites States, the Wellcome Trust Sanger Institute in the United Kingdom and the International Cancer Genome Consortium (ICGC) in Canada. TCGA data can be explored at the gene-based viewing mode using the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu) and the large cohort data also can be analyzed to generate Kaplan–Meier plots [16]. Additionally, pathway-based methods such as the Cytoscape (http://cytoscape.org), Mutual Exclusivity Modules in Cancer (MEMo) (http://cbio.mskcc.org/tools/memo.html), Pathway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM) (https://genome-cancer.ucsc.edu) and cBio Portal (www.cbioportal.org) can be used to elucidate functional connections among the genes of interest [17–19]. While all these efforts are contributing towards building MMMN models for cancer progression of each cancer type, at this time the majority of these alterations are not classifiable to a particular module in the grand scheme of cancer progression. Therefore it will take an improved and organized effort of sampling and profiling strategies of alterations in real time by the shedding of the heavy reliance on snapshots derived from samples corresponding to archived, stationary and/or predetermined randomly fixed time points as is generally the norm at the present time, the use of model systems to infer functional effects, and new bioinformatics tools to achieve what has been predicated by the MMMN hypothesis [3].

Role of the microenvironment in cancer progression

Tumors consist of more than the malignant cells, as the surrounding non-malignant stromal cells such as endothelial cells of blood and lymphatic circulation, fibroblasts, carcinoma-associated fibroblasts (CAFs),

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

myofibroblasts, pericytes, adipocytes, mesenchymal stem cells and immune cells and immunosuppressive cells such as the tumor associated macrophages (TAMs) and myeloid-derived suppressor cells (MDSCs), respectively, embedded in the modified components of the extracellular matrix (ECM) and remodelled vasculature, together form the tumor mass (Figure 1.1) [20, 21]. It is becoming more and more apparent that these diverse components play crucial roles in modulating tumor progression through paracrine/autocrine secretion of cytokines such as transforming growth factor-beta (TGFβ) and interleukin 6 (IL6), growth factors like epidermal growth factor (EGF), fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), insulin-like growth factor-1 (IGF-1) and other factors such as hedgehog (Hh), Notch, periostin (POSTN), vascular endothelial growth factor (VEGF) and Wnts [20–22]. In the context of MMMN models for cancer progression, one can envision that the networks of gene connections and pathways within and between the various modules that constitute the different stages of cancer progression could be influenced by the tumor microenvironment. For example, our previous studies with breast cancer found TGFβ could epigenetically regulate various driver genes involved in epithelial to mesenchymal transition in breast cancer [23]. Thus the functional status of driver genes in the modules of cancer progression could be influenced by TGFβ-like cytokines or other factors and hence impact the functional and phenotypic state of the cancer.

Future perspectives

The success of cancer therapies depends on the fulfillment of two criteria. The first challenge is to offer personalized medicine by treatment with drugs that are tailored to each patient's own tumor(s). The second is the ability to follow up/continue with therapeutic strategies that can prevent therapeutic resistance and the associated relapse to the initial targeted therapy. An optimistic vision for offering the panacea for these major challenges is to develop MMMN models for cancer progression that would provide details of all possible alterations in the tumor and its microenvironment and their contributions, which can be detected in a cancer at the time of diagnosis and used in the future to predict what one could expect to see upon relapse to help with the immediate implementation of effective follow-up therapeutic remedies. While this is not an easy task to achieve at the present time, future research and new technologies may provide the necessary tools to develop combination therapies that achieve the ultimate goal of curing, or at least keeping in check, metastatic disease for the longest term possible.

References

1. Fearon ER and Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. *Cell* 61: 759–767.
2. Kinzler KW and Vogelstein B. 1996. Lessons from hereditary colon cancer. *Cell* 87: 159–170.
3. Thiagalingam S. 2006. A cascade of modules of a network defines cancer progression. *Cancer Res* 66(15): 7379–7385.
4. Kinzler KW and Vogelstein B. 1997. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* 386: 761–763.
5. Jacks T and Weinberg RA. 2002. Taking the study of cancer cell survival to a new dimension. *Cell* 111: 923–925.
6. Leary RJ, Kinde I, Diehl F, et al. 2010. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* 2(20): 20ra14.
7. Vogelstein B, Papadopoulos N, Velculescu VE, et al. 2013. Cancer genome landscapes. *Science* 339(6127): 1546–1558.
8. Haber DA and Settleman J. 2007. Cancer: drivers and passengers. *Nature* 446(7132): 145–146.
9. Sawan C, Vaissière T, Murr R, and Herceg Z. 2008. Epigenetic drivers and genetic passengers on the road to cancer. *Mutat Res* 642(1–2): 1–13.
10. Wood LD, Parsons DW, Jones S, et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853): 1108–1113.
11. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, et al. 2013. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 3: 2650.
12. Prat A and Perou CM. 2011. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* 5(1): 5–23.
13. Alizart M, Saunas J, Cummings M, and Lakhani SR. 2012. Molecular classification of breast carcinoma. *Diagnostic Histopathology* 18(3): 97–103.
14. Eroles P, Bosch A, Pérez-Fidalgo JA, and Lluch A. 2012. Molecular

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev* 38(6): 698–707.

15. Stephens PJ, Tarpey PS, Davies H, et al. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486(7403): 400–404.
16. Cancer Genome Atlas Network. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418): 61–70.
17. Goldman M, Craft B, Swatloski T, et al. 2013. The UCSC Cancer Genomics Browser: update 2013. *Nucleic Acids Res* 41(Database issue): D949–954.
18. Cline MS, Craft B, Swatloski T, et al. 2013. Exploring TCGA pan-cancer data at the UCSC Cancer Genomics Browser. *Sci Rep* 3: 2652.
19. Eifert C and Powers RS. 2012. From cancer genomes to oncogenic drivers, tumour dependencies and therapeutic targets. *Nat Rev Cancer* 12(8): 572–578.
20. Joyce JA and Pollard JW. 2009. Microenvironmental regulation of metastasis. *Nat Rev Cancer* 9(4): 239–252.
21. Taddei ML, Giannoni E, Comito G, and Chiarugi P. 2013. Microenvironment and tumor cell plasticity: an easy way out. *Cancer Lett* 341: 80–96.
22. Castaño Z, Fillmore CM, Kim CF, and McAllister SS. 2012. The bed and the bugs: interactions between the tumor microenvironment and cancer stem cells. *Semin Cancer Biol* 22(5–6): 462–470.
23. Papageorgis P, Lambert AW, Ozturk S, et al. 2010. Smad signaling is required to maintain epigenetic silencing during breast cancer progression. *Cancer Res* 70(3): 968–978.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

Part 1 Introduction to modular organization of the networks of gene functions and cancer

Chapter 2

Lessons from cancer genome sequencing

Antoine Ho and Jeremy S. Edwards

Introduction

The Human Genome Project (HGP) was one of the greatest achievements of the twentieth century, and the publication of the full human genome sequence in 2001 ushered in the new century by starting the post-genome era in human biology. The great success of the HGP has paved the way to many future discoveries. The human genome sequence represents just the beginning of the payoffs for the biomedical community, and many future benefits are promised and expected in the near future. Specifically, the HGP has enabled the rapid sequencing of more genomes, such as cancer genomes, and this holds the potential to transform cancer research and treatment. Therefore it is more appropriate to look at the completion of the human genome as the end of the beginning, rather than the beginning of the end of the era of human genome sequencing. "Next generation" sequencing (NGS) technologies are providing fast, cheap and high-quality sequencing. As these technologies become less expensive and easier to operate, they will become more widely available. However, the bottleneck in the process will quickly shift to the analysis phases. In other words, making sense of the vast amount of sequence data will be a challenging task, and it will require bioinformatics and systems biology. The analysis of sequencing data will likely have a tremendous impact on many areas of medicine and biomedical research.

Background

The sequencing and publication of the human genome was performed simultaneously by two competing groups, one was publicly funded and the other was privately funded. The publicly funded sequencing project was led by Dr. Francis Collins and was performed in the classical clone-by-clone approach using traditional Sanger sequencing. The private sequencing project was based at Celera and was led by Dr. J. Craig Venter. The Celera group sequenced the human genome using the shotgun sequencing approach, which was made possible for three main reasons: (1) they developed novel assembly algorithms, (2) they utilized data from the public project, and (3) they sequenced a very homogeneous sample, as opposed to a sample representative of a large number of individuals [1].

The HGP's impact on future human genome sequencing has two broad implications. First, the HGP has now established a reference human genome sequence, allowing for relatively rapid sequencing of future genomes while using the reference sequence to align reads. Additionally, a major impact of the HGP has been spin-off technologies and bioinformatics tools, which have led to what is now known as "next-generation" sequencing (NGS) technology [2].

Next-generation sequencing technologies

During the HGP, a number of technologies were developed with the goal of increasing sequencing throughput to allow for cheap and rapid human genome sequencing. The first phases of the improvements were essentially advances in instrumentation and miniaturization of the traditional Sanger sequencing approach. However, a number of true next-generation technologies were also developed and have become widely available.

Sequencing template preparation

The first step of the next-generation sequencing pipeline is the construction of the sequencing library. The library preparation step essentially takes a genomic DNA sample, and converts it into DNA molecules

Systems Biology of Cancer, ed. S. Thiagalingam. Published by Cambridge University Press. © Cambridge University Press 2015.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

that can be sequenced by a given sequencing technology (Figure 2.1). For example, sequencing using the Illumina system fragments the genomic DNA into ~300 bp fragments, amplifies these fragments via PCR and ligates sequencing primer sites to the ends of the fragments [3–5]. These protocols vary in complexity depending on the sequencing platform.

Additionally, genome libraries can be constructed to contain mate-pair sequences. This means that the genome tags will be adjacent in the library molecule, but will have a kilobase or more separation in the genome. The mate-pair approach complicates library preparation, but assists in genome assembly/mapping, especially when dealing with very short read lengths, as is typical in most next-generation sequencing technologies (Figure 2.2) [3–5].

There are many ways to sequence DNA, and because of this there are many ways in which to prepare the DNA libraries for sequencing. First, the template can be clonally amplified unless sequencing can be performed on single molecules without the need for amplification. Methods that do not rely on

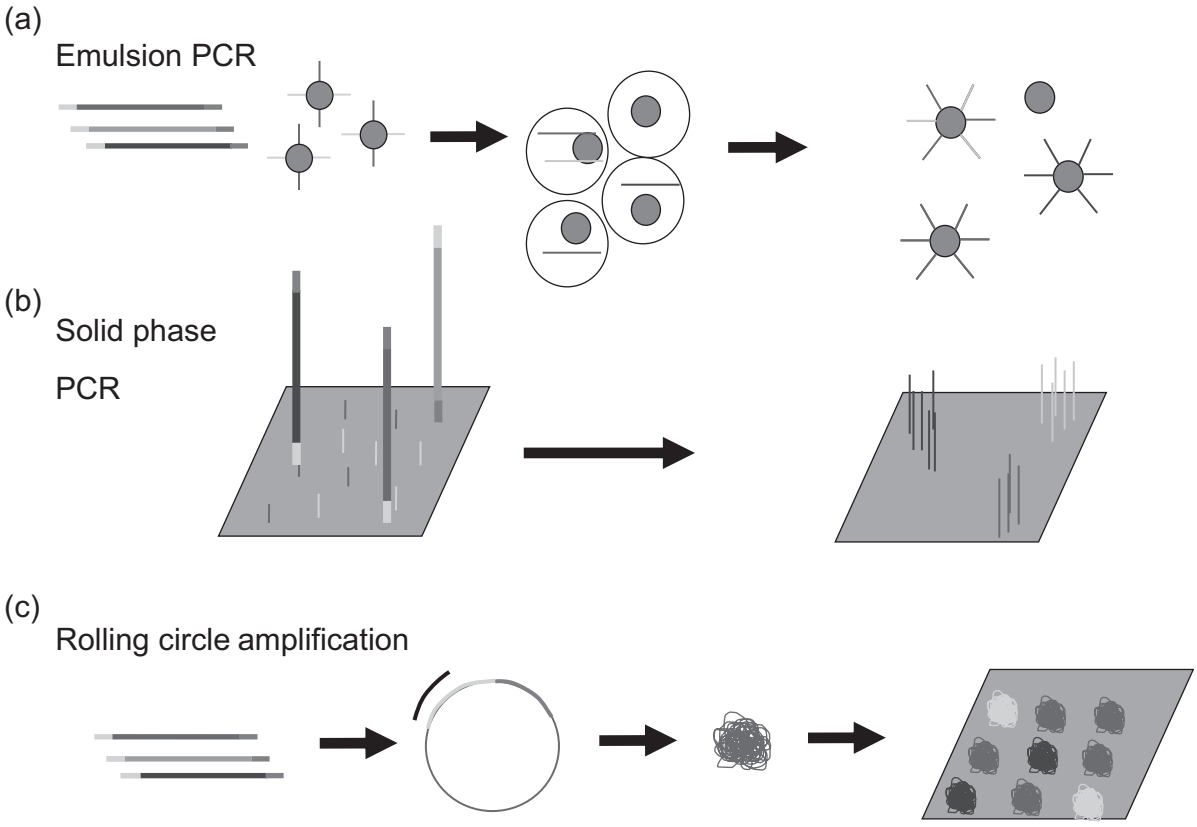

Figure 2.1 (a) Emulsion PCR (ePCR). Template DNA and beads are mixed and then put into an emulsion mixture consisting of an oil phase and an aqueous phase of PCR reagents. These beads have primers complementary to the ends of the template strands coupled to them, allowing the PCR reaction to extend these primers and cover the bead in copies of the template DNA. Template DNA is diluted to maximize the number of emulsions having exactly one template strand and one bead. Proceed with PCR temperature cycling. Sequencing is performed on beads with only clones of a single template DNA, as beads with no DNA and beads with more than one template DNA do not provide usable data. These beads can then be fixed onto an array for sequencing and imaging.
(b) Solid phase PCR. Very similar to ePCR, but without beads. Template DNA is diluted and then added to a slide with primers complementary to end regions of the template DNA coupled to the slide, which allows hybridization and priming. Through a series of PCR temperature cycling, a slide is covered in clonal patches of DNA to be sequenced.
(c) Rolling circle amplification (RCA). A piece of linear DNA is circularized enzymatically. Once circularized, RCA is performed with a polymerase that has displacement activity. This results in a ball of clonal DNA, effectively amplifying the DNA but without the need for emulsions or beads. These balls of DNA are then coupled to an array and sequenced. (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

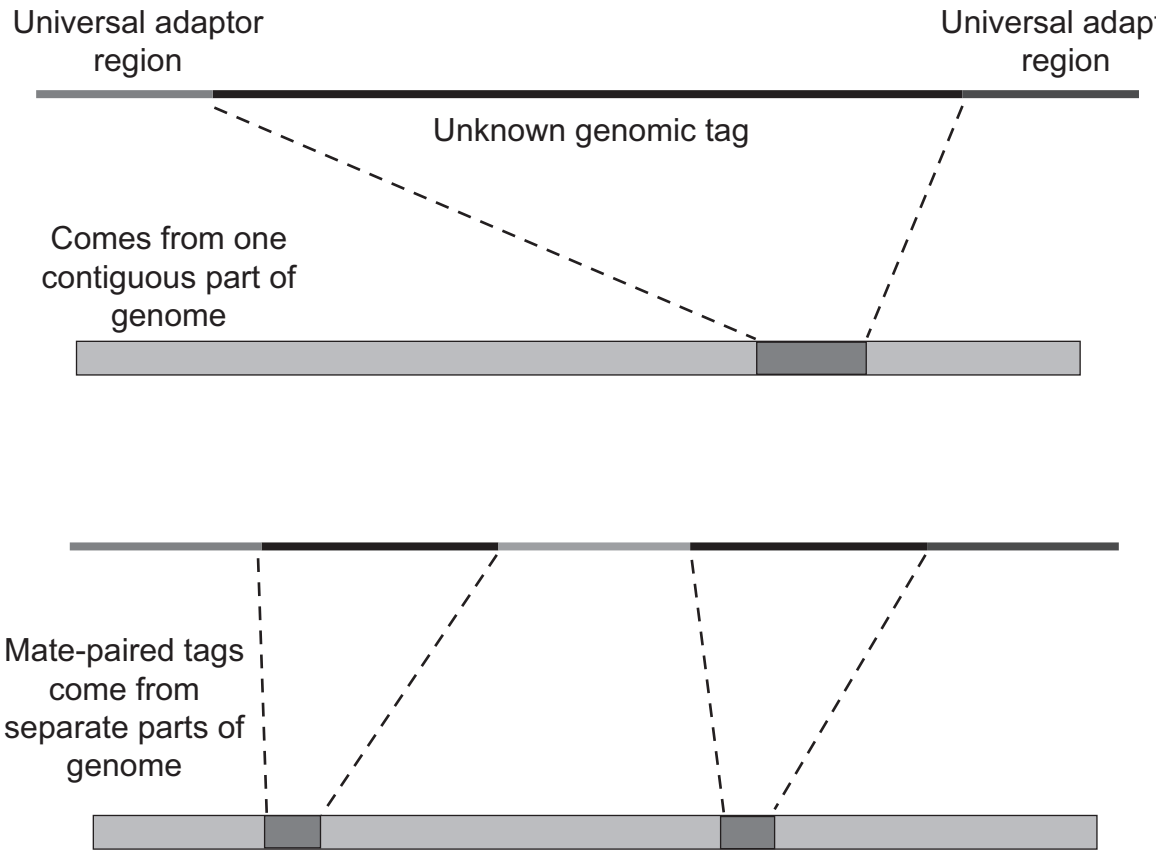

Figure 2.2 Mate-paired libraries. Mate-paired libraries can provide alignment information that is very valuable, especially when trying to sequence large redundant regions with short reads. The most ideal way to sequence a large redundant region is to simply get a single contiguous read of the entire region; however, that may not be technologically possible, which is why this mate-paired strategy is key. Because the mate-paired reads come from two different regions, a set distance apart, it is possible, even with short reads, that one half of the mate-pair will be in a uniquely identifiable region, and even though the other will be in the redundant, difficult to map region, that read will still provide useful alignment data. (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

an amplification step are known as single-molecule sequencing methods. Amplification is necessary for many sequencing approaches because a signal, whether it is light or electrical, must be amplified or would be too weak to identify otherwise. This amplification can occur through an emulsion PCR (ePCR) step [6] or through solid phase PCR as in the Illumina Inc. system. Additionally, rolling circle amplification (RCA) can be utilized to amplify the DNA into a ball, which may itself be coupled to an array (see Figure 2.1) [7]. Clonal amplification may make certain sequencing approaches possible; however, when clonal amplicons are being sequenced, the issue of phasing arises. For example, when a clonal population of DNA molecules is being sequenced, the initial signals for sequencing each base are near identical for all molecules. However, as sequencing progresses, inefficiencies in biochemistry, enzymatic activity, chemical cleavage steps, or incomplete washing cause the signal to become noisy and may contain an earlier (lag phasing) or later (lead phasing) position.

Single-molecule sequencing template preparation is greatly simplified, as there is no need for amplification, and there are no amplification biases that may occur. Some single-molecule sequencing methods also make real-time sequencing possible, though there are obstacles to single-molecule sequencing that methods must take into account, such as being able to recognize the signal of a single molecule, which requires more expensive and larger sequencing equipment [8].

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-49339-0 - Systems Biology of Cancer
Edited by Sam Thiagalingam
Excerpt
More information

Antoine Ho and Jeremy S. Edwards

Sequencing by synthesis

Fluorescent methods

The most popular next-generation sequencing approach is known as sequencing by synthesis (SBS). In SBS, a DNA polymerase is used to extend a primer on the template strand (Figure 2.3) [3–5]. The DNA template to be sequenced must contain a known region at its 3′ end to hybridize a primer. Once hybridized, synthesis is allowed to occur under controlled conditions with specific reagents. The goal is to allow only the incorporation of a single nucleotide onto this growing strand and to visualize the base that was incorporated. The key is to modify (block) the nucleotides in some fashion that not only allows termination of synthesis once incorporated, but also can be reversible. These can, for example, involve a blocking group on the 3′ OH of the growing DNA strand that can be removed enzymatically or by a chemical cleavage reaction [3–5]. The second element is to attach unique fluorophores onto each of the four different nucleotides to allow visualization. After imaging, and storing this data, the termination must be reversed by removing this blocking group, to allow the addition of another single nucleotide, and then the fluorophores must be cleaved to visualize the signal of the newly incorporated nucleotide. This process is repeated to sequencing up to ~150 bases. Sequencing

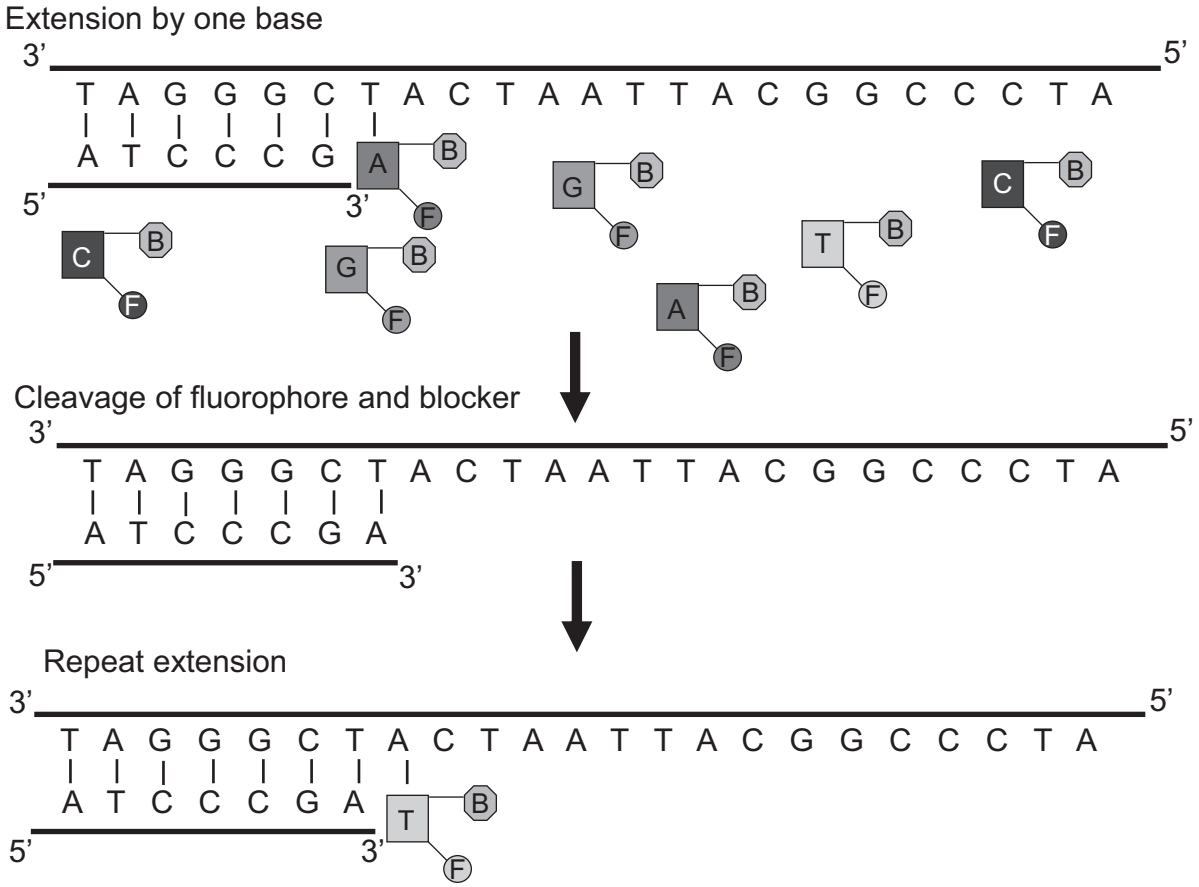

Figure 2.3 Sequencing by synthesis with fluorophores. A primer is hybridized onto the template DNA onto a universal region to allow extension by a polymerase. A single nucleotide will incorporate due to a blocking group on the nucleotides, and the DNA will be able to be visualized by the fluorophores attached to each nucleotide type. If there is a saturation step, as is often the case when dealing with amplified DNA template, it would be performed following the first extension step (not shown). A saturation step is identical to the first step except that there is no fluorophores, though there are still blockers on the nucleotides, and the nucleotides are usually at a very high concentration to saturate. The fluorophores are then cleaved chemically, and then the blocking group is removed so extension can continue to another base. This cycle then repeats. (A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

© in this web service Cambridge University Press www.cambridge.org