Translating the Cancer Genome: going beyond p values

Lynda Chin [1] and Joe W Gray [2]

1. Dana-Farber Cancer Institute and Harvard Medical School, Boston MA, USA
2. Lawrence Berkeley National Laboratory, Berkeley CA, USA

Correspondence should be addressed to LC at lynda_chin@dfci.harvard.edu

Preface

Cancer cells are endowed with diverse biological capabilities driven by myriad inherited and somatic genetic and epigenetic aberrations that commandeer key cancer-relevant pathways. Efforts to elucidate these aberrations began with Boveri's hypothesis of aberrant mitoses causing cancer and continue today with a suite of powerful high-resolution technologies that enable detailed catalogues of genomic aberrations and epigenomic modifications. Tomorrow will likely bring the complete atlas of reversible and irreversible alteration in individual cancers. The challenge now is to discern causal molecular abnormalities from genomic and epigenomic "noise", to understand how the ensemble of these aberrations collaborate to drive cancer pathophysiology. Here, we highlight lessons learned from now classical examples of successful translation of genomic discoveries into clinical practice, lessons that may be used to guide and accelerate translation of emerging genomic insights into practical clinical endpoints that can impact on practice of cancer medicine.

Introduction

The human cancer genome harbors wholesale alterations on the chromatin, chromosome and nucleotide levels, resulting in irreversible numerical and structural aberrations as well as reversible epigenetic modifications (Figure 1) that affect hundreds to thousands of genes or regulatory transcripts. Collectively, these molecular abnormalities serve to activate or neutralize biological events that define diverse aspects of cancer pathophysiology including altered growth, death, metabolism, angiogenesis, immune sequestration and metastasis [1]. Mining the complex cancer (epi)genome for aberrations governing these processes has become a major activity in cancer research, as it is widely appreciated that embedded within the oncogenomic landscape are mechanistic clues to disease pathogenesis that can inform the broader efforts of identifying molecular events for therapeutic intervention and molecular biomarkers for early detection and prognosis, improved diagnosis and response prediction. Recognizing this, multiple national and international efforts, including The Cancer Genome Atlas (TCGA) pilot project by the NCI and NHGRI [2], have been initiated to accelerate the compilation of the comprehensive atlas of cancer genomic alterations.

In recent years, cancer genomics – defined here as the study of the ensemble of DNA-associated abnormalities that enable and accompany cancer development - has exploded as a field, enabled by genome-wide high-resolution high-throughput platforms (see textbox). These technologies now yield informative but dauntingly complex multi-dimensional genomic datasets that describe in detail the myriad changes in epigenomic modifications and DNA copy number and structural aberrations as well as sequence mutations within individual tumors and how these differ between individual tumors. There is real potential that these datasets will transform the practice of cancer medicine as evidenced by therapies that target distinctive molecular events that result from genome aberrations features such as EGFR mutations (gefitinib or erlotinib)[3-5], the BCR-ABL translocation (imatinib mesylate)[6] and ERBB2 amplification (trastuzumab, lapatinib) [7] and assays for these aberrations that are now used to stratify patients for treatment. In parallel, assays for germline mutations identify individuals at high risk of cancer development – for example, p53 mutations are associated the Li-Fraumeni cancer syndrome[8], BRCA1/2 mutations signal increased risk for breast and ovarian cancer risk[9-11], mutations or epigenomic modification of DNA mismatch repair (MMR) genes such as MLH1 or MSH2 or MSH6 associates with hereditary non-polyposis colorectal cancer (HNPCC)[12] and CDKN2A mutations indicate increased risk for familial atypical multiple mole melanoma-pancreatic cancer [13].

These examples have demonstrated the promise of cancer genomics, stimulated rapid advances in genome technologies and computational science, and galvanized an entire generation of multi-disciplinary scientists on the quest to identify the next set of key cancer targets and disease biomarkers. While there has been tremendous success in the rapid accumulation of genomic data, vast majority of these enormous datasets have not yet translated into meaningful clinical endpoints. Historically, translation of each genome aberration discoveries into improved patient management has taken at least a decade and sometimes billions of dollars. This pace and expense will not permit the range of genomic discoveries to be effectively exploited unless more efficient, less costly strategies are developed. What are the barriers to rapid conversion of genomic information into useful diagnostics and effective therapeutics? Is statistical significance in the absence of mechanistic insights sufficient to harness the full translational potential of these complex genomic datasets in a cost-efficient and effective way? Or, is some degree of molecular biological function required for efficient translation? In this regard, the ABL, ERBB2 and EGFR paradigms appear to support the view that the coupling of genomic insights with pathobiology holds the greatest promise for clinical impact. In this article, we will review some examples of successful translation of genomic discoveries to the clinic and lessons learned from these first experiences. Against this backdrop, we will discuss the challenges and potential paths forward to translate the promise of a complete cancer genome atlas.

Historical lessons

There have been several pioneering examples of successful translation of genomic aberration discoveries in cancers into diagnostics and therapeutics with dramatic impact on practice of cancer medicine. Although many of these successes predated present-day high-throughput genome-wide technologies – indeed, some resulted from decades of painstaking work - they nevertheless presage the translation of cancer genomic discoveries into useful clinical tests and effective treatments. We review several here as lessons for modern genome researchers that may guide and accelerate translation of the genome aberrations now being discovered.

Translocations. The first recurrent genome aberration discovered to be associated with a human malignancy was the “Philadelphia chromosome” discovered by Nowell and Hungerford in 1960[14]. In the ensuring decades, cytogenetic and molecular studies showed this to be a translocation between chromosomes 9 and 22, resulting in a fusion product, BCR-ABL. This fusion gene deregulates tyrosine kinase activity in patients with chronic myeloid leukemia (CML)

and some forms of acute lymphoblastic leukemia (ALL). More than 30 years after the discovery of the Philadelphia chromosome, a kinase inhibitor, imatinib mesylate, was developed as an effective therapeutic agent against BCR-ABL in patients with CML[6]. Unfortunately, despite initial dramatic responses, this targeted therapy does not lead to durable cure since resistant tumors emerge that abrogate the inhibitory effect of imatinib mesylate[15]. Genomic analyses of the resistant tumors showed the acquisition of point mutations (sometimes amplified). This insight guided development of new kinase inhibitors designed to counter this resistance mechanism, leading to recent approval of nilotinib and dasatinib[16]. This suggests a recursive paradigm of therapy/biomarker developments in which genomic analysis guides the development of targeted therapies and associated predictive markers followed by genomic studies of resistant tumors to aid development of second and third generation inhibitors to counter resistance mechanisms. Banking of tumor tissues from drug resistant patients will be essential to support these studies. Another lesson to be learned from the imatinib mesylate story is that genomic analyses can effectively guide the use of small molecule inhibitors that show multi-target specificity. Imatinib mesylate, for example, also inhibits the c-Kit receptor tyrosine kinase. Guided by genomic analyses of gastrointestinal stromal tumors (GIST sarcomas)[17] and mucosal melanomas[18] showing that both harbor c-Kit mutations, imatinib has been successfully used to treat patients with GIST (sarcoma) and mucosal melanomas[17-19].

Numerous causal recurrent translocations have been discovered in human leukemias and lymphomas via molecular cytogenetic analyses since the pioneering discovery of the Philadelphia chromosome[20]. However, discovery of causal translocations in solid tumors has been difficult, possibly reflecting the complex genomic profiles and heterogeneous nature of these malignancies. With current day genomic analyses coupled with sophisticated analytical approaches and expanding genomic information, recurrent structural aberrations are being discovered in solid tumors and may be more prevalent that previously thought. A notable discovery is the high-frequency of TMPRSS2:ETS-family translocations in human prostate cancer. Using a novel integrative analytical methodology called COPA (Cancer Outlier Profile Analysis) that identifies associations between genomic and transcriptional abnormalities, Chinnaiyan and colleagues identified this family of common translocations that brings an oncogenic ETS transcription factor under control of the androgen response element TMPRSS2, effectively placing expression of the ETS oncogene under androgen regulation[21]. Molecular assays for the fusion events are now being developed and evaluated as early detection markers for prostate cancers[22]. Similar computational approaches on emerging multi-dimensional datasets will hopefully yield other causal structural aberrations in solid tumors. And this is only

the beginning. Next generation sequencing technologies that enable DNA sequence analysis of entire tumor genomes will be particularly valuable in discovering fusion genes and other structural rearrangements. The promise of this approach is illustrated by the revelation of remarkable structural complexity in the cancer genomes by end sequence profiling[23] or genomic region sequencing[24].

Gene amplification. Another prominent success story involves the now well-established oncogene, ERBB2. This gene with homology to erb-B and the tumour antigen p185 was initially identified as a transforming oncogene in NIH-3T3 cells[25] that was also amplified in human breast cancer cell lines[26-28]. Shortly thereafter, ERBB2 amplification was found in ~30% of primary tumors wherein amplification was associated with short survival duration and time to relapse[29]. Based on these observations, diagnostic assays for Her2 amplification[30] or overexpression were developed and used in early clinical trials of trastuzumab (a monoclonal antibody directed against the extracellular domain of HER2) to demonstrate survival benefit in Her2 positive metastatic breast cancer patients[7]. More recently, molecular assays for ERBB2 have guided the clinical deployment of the EGFR/ERBB2 targeted small molecule inhibitor, lapatinib [31].

Mutations. Since completion of the human genome sequence, several high-impact discoveries in genome science have come from systematic re-sequencing of cancer genes or pathways or gene family. One of the first and perhaps the most celebrated success from such large-scale sequencing projects is the discovery of frequent activating somatic mutations in *BRAF*, a serine threonine kinase in 60% of malignant melanoma, 10% of colorectal cancer as well as in lower frequencies of other cancers[32]. This discovery has driven multiple BRAF inhibitor development programs with several drugs already entered clinical trials. Other notable discoveries from large-scale sequencing efforts included frequent PI3KCA[33] and AKT1[34] mutations in many cancer types, ERRB2 and EGFR mutations in NSCLC[35,36], among others. The discovery that, in addition to gender, ethnicity, smoking history and histopathological subtypes; EGFR-activating mutations predict responses to anti-EGFR targeted therapies in chemotherapy-refractory advanced NSCLC patient [3-5] has immediately changed the standard of care for patients with NSCLC. EGFR mutation testing is becoming routine prior to treatment decision with EGFR inhibitors[36]. Here, the ability to assay EGFR genotype retrospectively using banked tumor tissues with matched germline DNA was crucial in enabling the stratification of responders and demonstration of efficacy, leading ultimately to FDA approval of erlotinib for NSCLC[37]. A lesson to learn from this is the importance and values of uniform collection of pre- and post-treatment tumor specimens with matched normal controls from clinical trials to enable

future prospectively designed retrospective analyses of clinical responses, perhaps even at the expense of slower patients accrual.

Germline susceptibility. In addition to somatic genetics, genomic science is also revolutionizing our searches for germline susceptibility genes or polymorphisms in inherited disease predisposition including cancers. One of the early successes in this area was the discovery of BRCA1 mutation association with familial breast cancer[9,10]. Genetic screening for germline mutations in BRCA1 and now a second cancer susceptibility gene, BRCA2[11] are now being deployed world-wide to identify patients at high risk of developing early onset breast and ovarian cancer. Moreover, the knowledge that BRCA1 is necessary for error-free double-strand break repair led the way to development of PARP inhibitors, a new treatment paradigm[38]. These and subsequent studies established the concept that efforts to discover inactivating germline mutations associated with increased susceptibility to cancer can be guided by analyses of LOH or reduced genome copy number and/or methylation in the tumors that eventually develop. Application of current and future day genomic technologies in coordinated germline and tumor studies are likely to significantly accelerate that identification of susceptibility genes of this class thereby enhancing our ability to stratify high-risk individuals for aggressive surveillance, prevention and management. However, this will require coordinated collection of tumor specimens along with germline DNA in large cohort genetic susceptibility studies.

Making sense of the oncogenome

Empowered by our improved capability to survey the cancer genome with increasing accuracy and resolution, hundreds of cancer genomic studies have been conducted or initiated with the hope of discovering the next EGFR, HER2 or BRAF. Instead, these analyses are uncovering hundreds of recurrent genomic or genetic alterations impacting thousands of "genetic elements of interest (GEOI)" – including annotated genes, non-coding micoRNA, or other conserved elements – that *might* contribute to the pathophysiology of human cancers. The nature and "strength" of each GEOI, our certainty of its contribution and therefore its translational importance varies substantially. Some will be strong, causal "drivers" of important cancer hallmarks[1], others will be weaker but important "contributors" to the development of cancer pathophysiology while many will be genomic "noise" or "passengers" that are biologically neutral and have been accumulated by chance during the cancer's life history. Distinguishing the drivers and contributors from the passengers is a central challenge in genomic research

today. This is made more challenging by the diversity of GEOI function and the likelihood that GEOI function may be tumor type (or subtype) as well as microenvironment dependent.

In cases of high-frequency events such as amplifications of regions encoding EGFR in GBM (45%) or ERBB2 in breast cancer (20%); deletions of regions encoding CDKN2A or PTEN in upwards of 80% in solid tumors or mutations of p53, RAS, BRAF and PIK3CA in a wide range of solid tumors (see http://www.sanger.ac.uk/genetics/CGP/Census/), assignment of GEOIs as "drivers" is compelling and rests on weight of functional evidence built up over decades, a "luxury" not afforded by novel GEOIs born to present day high-throughput cancer genomics. Furthermore, these prominent "gene-mountains" appear to be few and far between relative to the numerous "hills and valleys" stretching broadly over large regions of the oncogenome [39,40]. Which of these GEIOs are on the critical path to malignancy? What are their relative contributions? These are challenging questions without simple answers, but they do converge on the theme of integration and triangulation (Figure 2). Below we highlight several examples of approaches that have been utilized successfully to find the "needles" – drivers and contributors – in the haystack of cancer genome data.

Integrative analyses of multi-dimensional data

The cancer genome is dysregulated by multiple mechanisms, including DNA and chromatin modifications and changes in DNA structure, copy number and mutations of coding and non-coding sequences that alter RNA transcription, translation, gene function and/or post-translational modification. Technological advances that allow examination of the cancer genome in multiple "omic" dimensions are helping to focus driver and contributor discovery since these GEOIs tend to be deregulated by several different mechanisms. A classic example is the CDKN2A ($p16^{INK4A}$/$p14^{ARF}$) tumor suppressor, which can be inactivated by homozygous deletion of the 9p21 locus, epigenetic silencing of gene expression via promoter methylation, or point mutations crippling $p16^{INK4A}$ functions[41]. Similarly, the oncogene, PIK3CA, can be activated by amplification and over-expression[42] and/or activating mutations[33]. Such multi-mechanism deregulation is clearly illustrated when examining well-known *bona fide* oncogenes in a typical signaling pathway (Figure 3). In other words, if it is important, cancer will find a way to deregulate a genetic element by any mechanism possible. By this reasoning, targeted re-sequencing of resident genes within regions of amplification has yielded clinical fruits, such as KIT in mucosal and acral melanomas[18]. Thus, demonstration of complementary modes of deregulation through integration of multiple dimensions of genomic information is a piece of strong evidence in support of a likely pathogenetic GEOI. The current large-scale cancer

genome projects with coordinated comprehensive genome-wide characterization will be most powerful in leveraging such multi-dimensional data for integrative analyses. Additionally, triangulation across tumor types can be highly informative as well, as it is clear that the mechanisms of deregulation of many *bona fide* cancer genes, such as MYC, EGFR, AKT, RAS or p53, PTEN and CDKN2A vary according to tumor type – for example, genes like MYC that are activated by translocation in leukemias may be activated by amplification in solid tumors. Convergence among different tumor types can rapidly prioritize GEOIs that are likely to be important broadly. As a byproduct, it is likely that the predictive or prognostic power of genome biomarkers will increase substantially if assays are developed that assess the accumulated effect of all mechanisms of deregulation, such as changes in protein level or structure.

Comparative oncogenomics

Evolutionary conservation can be a powerful guide to cancer gene discovery since genes involved in pathways that are deregulated in cancers such as RTK signaling, cell cycle regulation and apoptosis are strongly conserved across species[43,44]. This comparative approach proved to be enormously helpful in refining the draft of the human genome. With respect to cancer, it has been established that cancer genes from one species can effect the malignant transformation of cells derived from different species despite poor primary sequence conservation (e.g., dMyc transformation of rodent cells [45]). Recent large-scale cross-species comparison has established that mouse and human tumors sustain orthologous genomic events which target novel cancer genes in diverse tumor types [46-48], supporting the view that genomic alterations conserved across species are more likely to represent critical events in tumorigenesis, and that evolutionary conservation can provide a potentially powerful solution to the central problem of noise in genomic datasets.

While it began with histopathological diagnoses, cross-species comparison has evolved to include genetic/genomic analyses to demonstrate that genetically engineered mouse models can model genetic aspects of human cancer, as exemplified by cross-species conservation of transcriptional signatures for KRAS activation in lung cancers[49] or somatic mutations of NOTCH1 in mouse and human T cell leukemia [50]. This was followed by proof-of-concept that comparison of genomic profiles of mouse and human tumors enabled discovery of novel oncogenes[46,47]. In the case of the study by Kim *et al*, ability to manipulate stages of tumor evolution, from regression to recurrence to escape *in vivo*, was leveraged to force selection of aberrations conferring metastatic capability. Genome-wide copy number profiles revealed focal amplification in mouse metastastic tumors that were syntenic to human 6p24-25, a region that

sustains copy number gain in 36% of human metastatic, but not primary, melanoma[51]. Although 6p gain is highly recurrent, suggestive of potential pathogenetic and/or prognostic importance in human tumors, its extended nature in human tumors renders identification of drivers/contributors difficult to impossible. Given the focal nature of the event in the mouse, cross-species comparison was able to narrow one region of interest to an 850 KB region encompassing only 8 annotated genes, with *NEDD9* as a putative driver. With that information as a guide, further functional and clinicopathological studies documented NEDD9s metastasis-promoting activities, and elucidated its molecular action via focal adhesion kinase[46]. Likewise, comparisons of recurrent genome copy number aberrations in tumors with ERBB2 amplification in human breast tumors and in a transgenic mouse model in which oncogenic Erbb2 (NeuNT) was expressed under control of the endogenous Erbb2 promoter implicated GRB7 and 14-3-3-σ as contributors in ERBB2 mediated oncogenic process[52].

While syntenic aberrations have been observed between murine and human tumors, it is important to note that genomes of most mouse tumors accumulate far fewer genome aberrations than do human solid tumors. For example, mouse tumors from oncogene-driven mouse models often exhibit few to no copy number aberrations (CNAs), and infrequent (typically simple) CNAs presumably occur only under strong selective pressure. This simplicity facilitates genomic identification driver and contributor aberrations as exemplified by studies by Kim et al and Zender et al [46,47]. On the other hand, the disadvantage is that it does not lend itself to widespread use of cross-species comparison.

Based on the observations that telomere dysfunction-induced DNA breakage events can drive regional amplifications and deletions and that laboratory mouse does not experience telomere-based crisis, DePinho and colleagues knocked out in the mouse germline the RNA component of the telomerase holoenzyme, generating a telomerase-deficient mouse that experienced progressive shortening of telomere length through successive generations, eventually leading to crisis[53]. Tumors from these animals indeed showed high level of instability, harboring large number of non-reciprocal translocations and complex CNAs [54-56]. Unbiased genome-wide comparison of such genome-unstable murine tumors and several human cancers of diverse origins demonstrated compelling non-random overlaps over copy number aberrations, proving that murine and human tumors experience common biological processes driven by the orthologous genetic events [48]. Attesting to the potential of such cross-species comparison for gene discovery, focused re-sequencing of GEOI within syntenic deletions revealed high frequency mutation of FBXW7 in human T-ALL, and PTEN[48], the latter also shown to modify responses to NOTCH1 inhibition in clinic[57]. These studies support the

notion that cross-species synteny serves not only as a measure of validation by virtue of their evolutionary conservation and utilization of different genetic mechanisms (i.e., mutation and copy number), but also guide the discrimination of drivers/contributors from bystanders.

Another aspect where the mouse has proven its value in comparative genomics is in identification of susceptibility loci. Extending the concepts used to identify BRCA1, one might expect to find mutations or polymorphisms that contribute to cancer susceptibility be subjected to positive selection during cancer genome evolution. Thus, these mutations might be found through allele specific analysis of genome copy number and gene expression in defined model systems. As an example, Balmain and colleagues used genomic strategies to identify polymorphic variants of the aurora kinase, AURK or STK15, to be associated with increased risk of developing cancer in multiple anatomic sites in the mouse[58,59]. These studies began with analysis of the genetic localization of quantitative trait loci (QTL) in mice that controlled susceptibility to skin tumor formation in interspecific mouse crosses (Mus musculus x Mus spretus). One of these, Skts13, was orthologous to a region of recurrent copy number increase in human cancers of the breast, colon, ovary and colon at 20q13 that encoded the aurora kinase, AURKA. Analyses of expression of the mouse ortholog, stk6, showed allele specific difference in the mouse intercrosses while genome copy number analyses of AURKA 91A and AURKA 91T showed preferentially amplification of the AURKA 91A allele in human colon tumors. A subsequent meta analysis of the association of AURKA T+91A alleles risk of cancer development of the colon, breast, prostate, skin, lung and esophagus showed an increased risk in both homozygotes and heterozygotes. These results confirmed that the AURKA T+91A variant is a low penetrance cancer susceptibility allele affecting multiple cancer types. Overall, this integrative analysis of quantitative cancer traits in mice, analysis of allele specific copy number change and expression and assessments of susceptibility in large case control studies may be essential to identify the likely large number of low penetrance, high prevalence polymorphisms that influence cancer risk.

Finally, model systems including the mouse are ideally suited for forward genetic screens where one can “listen” and let the cancer cells “tell” us what events they require or prefer on their path toward full malignant transformation. For example, retroviral insertional mutagenesis in the mouse has yielded recurrent and common insertion sites at genomic loci encoding genes such as RAS, Myc, Notch1, Flt3, Kit or p53 ([60] and references therein), attesting to their potential power as cancer gene identification when triangulated with existing and emerging human cancer genomic data.

Cell line model systems

Much of our understanding of cancer cell biology including aspects of gene regulation and signaling has come from studies of cancer cells in culture. The roughly 50,000 publications describing uses of the HeLa cell line and 20,000 publications describing uses of the NIH 3T3 cell line attest to this fact. That said, no cancer biologist or geneticist will argue that established tumor cell lines grown on plastic dishes, in thee dimensional cultures or in immune compromised mice can fully recapitulate all biological aspects of human tumors growing in the complex human microenvironment. Nor can any models fully represent the responses of the range of human tumors to therapy – in part due to differences in biological environment and in part because the models do not capture the range of biological and (epi)genomic diversity found in human tumors. Therefore, it is expected that each model system has pros and cons, strength and weakness. Mouse is one such example; as highlighted above and discussed in greater details elsewhere [61], its value is unequivocal. As long as we are mindful of the limitations of any one model, we can leverage information such system can offer. Integrating across multiple models will bring us closer to a true picture.

So, what CAN we learn about genomic aberrations from cell line models? And why are they important? Simply put, cell lines are essential for functional and biological validation of GEOI. Almost without exception, demonstration of functional activities and molecular bases for their action are necessary for the discovery of novel cancer genetic elements, including genes or microRNAs. Such efforts inevitably begin with various cell model systems, including established cancer cells, for their ease of manipulation and versatility (Figure 4). In such system, one can simulate the cancer-associated events (e.g. enforced expression of a GEOI resident in an amplification or RNAi-knockdown of a GEOI in a deleted region) to interrogate the biological and biochemical consequences of GEOI deregulation and to define its role (e.g. driver vs contributor vs passenger) in cancer development. Clearly, a major obstacle to accurate interpretation of functional data in established cancer cell lines is the lack of clarity on the complements of genetic alterations they carry, as it has become clear that genotypes of the system, be it a cell line, a model or even a patient, can dictate behavior of the cancer cells and alter response to a manipulation such as RNAi knockdown or pharmacological inhibition. As in the case of the tumors from which they were derived, no two cancer cell lines are alike. Moreover, there is the legitimate concern that genomic aberrations will be gained or lost during extended passages in culture. Therefore, it is important that cell line models – on plastic, in 3-diminesional culture or in xenografts - are subjected to same level of comprehensive genomic characterization as human tumor specimens so that interpretation of functional studies can be

guided by knowledge of the similarities and differences between the cell lines and tumors they are intended to model. It is also important that any cell line "system" used for functional oncogenomic studies is comprised of multiple independent cell lines with molecular diversity. If sufficiently diverse, analyses of such cell line collections minimize the risk that the elucidated function of an aberration will be idiosyncratic to a particular cell line.

As in model organisms, forward functional genetic screen using cancer cell line model, particularly in recent years with advance of RNAi technology, has been touted as one powerful platform to identify cancer relevant genes. Such screens *in vitro* may be limited by the phenotype (life and death, predominantly) amenable to high-throughput biological assays. Nonetheless, recent studies that intersect high-throughput RNAi screen in vitro with human cancer genomic data have led to the identification of REST as a tumor suppressor in colon cancer [62], IKBKE as an oncogene in breast cancers [41] and PIK3CA mutations as important determinants of resistance to trastuzumab [63].

Cell lines also represent an important model system for drug sensitivity and resistance in the quest to identify possible biomarkers to guide early phase clinical trial studies, to identify drugs that may be effective in tumor subtypes that are resistant to the current standard of care and to identify effective drug combinations. Although still in its infancy, a growing literature supports the concept that analyses of responses of collections of molecularly characterized cell lines to chemotherapeutic agents targeting molecular mechanisms intrinsic to the tumor cells will reveal molecular markers that can be used to predict drug response[64 65-68]. As a corollary, these analyses also identify drugs with high specificity to cancer cell subsets defined by specific molecular characteristics. Examples include *in vitro* analyses that predict (a) the known sensitivities of ERBB2 amplified tumors to trastuzumab [69] and lapatinib [66], (b) the sensitivity of tumors carrying EGFR mutations to gefitinib[3-5], and (c) acquired gefitnib resistant mutation in EGFR [70] as well as (d) the resistance of tumors with mutated or amplified BCR-ABL to imatinib mesylate [71].

In short, while not fully recapitulating real tumors in patients, cell model systems with large number of independent established lines of broad molecular and cellular diversity accompanied by comprehensive genomic characterization can be and will be tremendously useful in translation of genomic insights to clinical endpoint. These systems can be further improved by development of (i) co-culture or 3D culture conditions that better model in vivo microenvironment and (ii) strategies to establish primary or short-term cultures that minimize "culture-shock" associated with adaptation to plastics.

Molecular understanding in translation

Through integrative analyses of multi-dimensional data and comparison across multiple model systems or species (Figure 2), the process of identifying driver/contributor GEOIs, especially the relatively weaker or less prevalent ones (e.g. the hills and valleys) can be greatly accelerated. But is this milestone of 'guilt by association' sufficient for translation? We think not. Cancer is a complex and heterogeneous collection of disease entities as defined by clinical, histopathological and genetic parameters. Given this disease heterogeneity, the identification of compelling correlation defined in a test-validation set in a laboratory setting (e.g. a collection of genomic data; behavior in a model system, even responses in a clinical trial), no matter how significant, may not apply generally (i.e., the next new patient entering hospital care or enlisted into a randomized Phase III clinical trial). We believe that, without a definition of the genomic and biological context under which these GEOIs exert their mission-critical roles in cancer (i.e., its neutralization is associated with a robust anti-oncological response), the full therapeutic, diagnostic, and/or prognostic value of these genomic insights will not be realized but rather will be lost in translation.

Consider the example of EGFR mutations in NSCLC and GBM. Mutational activation of EGFR in NSCLC identifies a subpopulation of patients highly responsive to targeted inhibition of EGFR. The percentage of patients with EGFR activation mutation in NSCLC is small (circa 10% in US studies, somewhat higher in Asian populations) [36], and thus the therapeutic response of these patients to gefitnib would not have emerged in the absence of genetic stratification of this clinically distinct population. Conversely, EGFRvIII deletion and amplification are very prevalent in GBM (approximately 45%) [72], yet EGFR TKIs show a strikingly meager clinical impact. A positive, albeit transient, clinical response has been detected in patient subsets with an EGFR event and intact PTEN [73], indicating this key downstream molecule can modify the tumor biological response. However, it is notable that these positive responses are not durable despite documented pharmacological extinction of mutationally activated and amplified target. Here, the proteomic profiling of RTK activation patterns in solid tumors, including GBM and lung cancer, has provided a rational explanation for the patterns of clinical responses. Specifically, Stommel et al showed that established GBM tumor cell lines, xenotransplants and primary patient specimens possess multiple co-activated RTKs, and that inhibition of EGFR alone can lead to replacement by other co-activated RTKs in the PI3K complex, thus maintaining downstream survival signaling[74]. Downstream signaling was extinguished only when multiple RTKs were targeted by RNAi or combination TKI[74]. Thus, the integration of genomic and proteomic insights with molecular dissection of the signaling complex

now provides a more accurate blueprint for the rational deployment of TKIs in GBM, lung and other solid tumors.

While critical to translation, establishing the molecular basis of cancer-relevant action of a GEOI in specific tumor-biological context is perhaps the most difficult step in cancer genomics. Compounding the challenges of lengthy and laborious functional - clinicopathological validation (Figure 4) is the biological phenomenon of false negatives. Such false negatives can arise in many ways, including when (i) cancer-related biological activities of a genetic element are not captured by standard cell-based cancer assays (e.g. interaction with the host stroma); (ii) genetic element plays relevant role but only under specific cellular or genetic context not recreated in the validation assay; or (iii) a genetic element contributes partially to the overall activity conferred by a genomic event, thus single GEOI activity is negligible in the absence of these cooperating partner(s). Therefore, validation must not rely on a single assay by a single manipulation. Gain-of-function and loss-of-function manipulations for multiple tumor-phenotypes using multiple cell lines should be performed to search for the context in which biological activity can be revealed. Here, one can benefit from other tumor biological knowledge of the disease or gene family or pathways, including insights from triangulation data that nominate the specific GEOI. For example, if a GEOI identified by integrative genomic analyses is further prioritized on basis of its known role in neural-stem-cell (NSC) homeostasis, one will specifically assess how its manipulation impacts on NSC renewal/maintenance/differentiation in addition to the more generic assays of anchorage independence or proliferation (Figure 4). Similarly, if a GEOI is identified in a subset of tumors with a particular genotype (e.g. with activated RAS vs EGFR mutation), one needs to assay its biological importance under the appropriate context. This has been demonstrated in two recent studies[46,47]. Kim et al, showed that NEDD9 had gain-of-function pro-invasion activities only in cells with concomitant BRAF or RAS activation, an experimental design informed by the characteristics of the metastatic escapers harboring NEDD9 amplification[46,47]. Zender et al, demonstrated cIAP1 and Yap exhibited oncogenic activities in p53+/- hepatoblasts with Myc activation, but not ones with Akt or RAS activation, consistent with the presence of 9qA1 amplicon (targeting cIAP1 and Yap) in that specific mouse model of HCC[46,47]. In the Zender et al study, not one, but both cIAP1 and Yap were shown to be cooperative targets of 9qA1 amplification, highlighting yet-another level of complexity that contributes to biological false negatives. Here, functional genomics can be an efficient path forward. Not only does such approach allow increased throughput in assaying large number of GEOIs for tumor biological activities, genetic screens with low-complexity libraries representing GEOIs resident within a particular genomic event (especially ones that are

large and gene-rich) will enable identification of cooperating contributors that together confer the biological advantage sought by the cancer cells. This approach will likely be important for sorting out which of the less impressive "hills and valleys" are biologically important.

Similarly challenging is the issue of biological false positives. For instance, RNAi-mediated loss-of-function assay is a powerful mean to determine whether expression of a GEOI is required in a cell for a specific turmorigenic phenotype (e.g. survival, anchorage independence or invasion etc). However, given the innumerable genetic and epigenetic alterations present in established tumor cells (and consequent altered signaling between pathways and networks), it is possible that the phenotype observed may be true-true-unrelated. Here, complementary gain-of-function activity can help to increase the weight of evidence in support of a particular GEOI being a true driver or contributor to cancer. Additionally, type of functional activity also conveys different level of confidence; anchorage independent growth in soft agar is a more stringent assay than enhanced proliferation in fully supplemented cultured media with 10% serum. Biological false positive can also emerge as a direct consequence of the artificial nature of our assays. Consider the possibility that overexpression of a GEOI may confer a strong anchorage independent phenotype, but such may be the result of its supraphysiological level of expression in vitro; conversely, knockdown of a GEOI may lead to cell death for its expression is required for survival of all cells, not just cancerous ones. Here, clinicopathological validation utilizing tissue microarrays can provide added support for cancer-relevance by demonstrating prevalence of dysregulation on DNA (by FISH) and protein (by IHC or IF) levels in independent large cohorts of specific tumor types and of broad tumor spectrum. This can be particularly informative if such TMA cohorts are annotated with clinical outcome as such survey will not only add to the weight of evidence, but also provide invaluable clues and insights into possible clinical context for therapeutic development. In the end, it is the cumulative weight of evidence based on strength of specific functional activities, magnitude of clinicopathological data as well as significance of mechanistic clues that ultimately gives one the confidence of assigning a GEOI as a cancer-relevant "driver" or "contributor" rather than a mere passenger.

Conclusion

Cancer is the phenotypic endpoint of numerous epi/genomic alterations accumulated within the cancer cells and interactions of such alterations with the stromal components in a unique host microenvironment. Some of the major challenges in translation of cancer genomics stem from the fact that many cancer-associated DNA changes represent genomic noise and

there is incomplete understanding of the biological functions of many of the genetic elements in recurrent genomic alterations. Compounding these is the unfortunate reality that cancer is a highly complex, nimble and versatile disease. We have argued here that making sense of this complexity can be greatly facilitated by triangulation and integration with genomic and biological insights from model systems and clinical knowledge of the disease, and that translation can be accelerated by rigorous biological validation and mechanistic exploration in preclinical setting to better define the clinical context(s) in which a specific genetic element (or its linked pathway/network components) represents an effective therapeutic point of intervention. At the same time, we need to be mindful that our current knowledge of what makes a strong driver, a cooperating contributor or, for that matter, a genomic passenger is limited at best and possibly wrong. Therefore, this must be an iterative learning process where results of downstream biological validation and mechanistic studies, even clinical experiences when inhibitors or biomarkers are developed and deployed, can and must inform the integrative analyses and the validation approaches. This effort will be facilitated by development or assembly of model systems that are characterized to the same degree as primary tumors that can be used to quickly test hypotheses suggested by the tumor “omic” analyses. In other words, efficient translation of cancer genomics must go beyond statistical analyses of large genomic datasets and will require the amalgamation of expertise and insights from cancer biology, cancer genetics, cancer modeling and system biology as well as clinical experiences. While each of these components may be pursued in an individual laboratory or research program, effective integration will enable timely exchange and bring synergy not possible otherwise. Therefore, in parallel of multiple national and international large-scale cancer genome projects aimed at systematic and comprehensive characterization of the human oncogenome, it is also important to establish cancer genome translation centers or cooperatives where such diverse expertise can be assembled and efforts coordinated to mine the rich genomic datasets, generate hypotheses for validation, elucidate mechanisms and rationally design clinical development plan for biomarkers and therapeutics.

Textbox: Translating the cancer genome

Comprehensive analyses of genome copy number, transcription, epigenomic modification and DNA sequence in cancers are now underway worldwide. A central challenge in analyzing complex datasets emerging from these efforts is to devise strategies to efficiently prioritize (epi)genomic aberrations for assessment of biological importance and translational potential. Traditionally, such prioritization is based on recurrence of the aberrations, associations with clinical endpoints such as histopathology and outcome, as well as biological activities. A rapidly evolving suite of technological solutions now enable analysis of the oncogenomic landscape with remarkable resolution and accuracy.

Copy number aberrations. Comparative genomic hybridization (CGH) allows changes in genome copy number to be mapped onto a representation of the normal genome thereby allowing ready identification of the genes involved in the aberrations. Modern CGH analysis platforms map these changes onto DNA sequences arranged in microarrays and allow quantitative assessment of changes in genome copy number in cancer genomes - including individual alleles in some platforms – with subgene resolution. Even at this resolution, some aberrations may be missed, especially with platforms that are gene oriented given the growing appreciation of the roles of regulatory, non-coring RNA transcripts in cancer pathogenesis. Next generation technologies that efficiently sequence small genome fragments randomly collected from tumor genomes will complement microarray based copy number analysis strategies by allowing sequencing of sufficiently large numbers of DNA fragments that copy number along the genome can be assessed simply by summing up the number of DNA sequence elements in genome resolution elements distributed along the genome. The resolution of this approached can be made arbitrarily high by sequencing to increasing depth.

Structure aberrations. Structural changes may involve segmental deletions or insertions, translocations or complex rearrangements (e.g. those occurring during gene amplification or copy number change). These may be discovered using sequence based strategies. (a) *End sequence profiling (ESP)* is an adaptation of whole genome shotgun sequencing that allows detection of structural aberrations[23] in which DNA from a tumor is cloned into a large insert vector and the ends of the resulting clones are sequenced and mapped onto the normal human DNA sequence. Paired ends that map farther apart than the maximum size tolerated by the cloning vector indicate the presence of a structural aberration. This approach has the advantage clones containing aberrant DNA sequence fusion can be sequenced to identify the exact DNA sequence at the breakpoint. (b) *Paired end sequencing* combines the

rescue and capture of paired ends of short DNA fragments with high throughput sequencing and a computational approach to map DNA reads onto a reference genome to reveal structural variants [75].

DNA sequence abnormalities. Recent large scale DNA sequence analysis efforts have identified several hundred candidate cancer genes that may play function roles in various human cancers[39,40]. Some occur at relatively high frequency but most are present in only a few percent of tumors. Results from the extensive sequencing and mutation validation efforts now underway will be necessary to establish prevalence and clinicopathological associations for these GEOIs. Both established and next generation sequencing technologies will be brought to bear on this issue. (a) *Sequencing by bybridization (SBH)* [76] is an array based strategy in which mutations are detected based on intensity of hybridization to arrays comprised of comprised of short oligonucleotide probes that are designed to be perfectly complementary to the reference sequence plus oligonucleotide probes that differ by one base at each "substitution position" in the genome to be tested for mutation. This approach is well suited to resequencing. (b) *Dideoxy sequencing* [77] is the current standard mutation detection methodology. Dideoxy sequencing typically is applied to products resulting from PCR amplification using primers that flank regions of interest. Sequence "reads" typically are about 750 bp. Most implementations of mutation detection using dideoxy sequencing will miss mutations that are present in less than about 20% of the cells in the PCR amplified population. Mutations discovered so far are summarized at http://www.sanger.ac.uk/genetics/CGP/cosmic/. (c) *Shotgun sequencing* proceeds by fragmenting a target genome into numerous small segments that are sequenced using dideoxy sequencing. Most regions of the genome will be sequenced several times and computer programs assemble overlapping sequences into a contiguous sequence. (d) *Single Molecule Sequencing Methods* allow genome wide DNA sequencing starting from single molecules rather than a population of molecules [78,79]. Current read lengths range from ~30 to 300 bp and the number of reads per analysis ranges from 300,000 to 30,000,000. These technologies facilitate detection of rare mutations. Recent affinity enrichment techniques allow subsets of the genome to be enriched prior to sequencing (e.g. all known exons) thereby decreasing the cost of targeted sequencing [40,80].

Epigenome analysis. It is clear that epigenomic modifications are major contributors to tumorigenesis and progression – especially during early stages of development. Several techniques for genome wide assessment of DNA methylation and chromatic structure are now established or emerging that will facilitate further elucidation of the oncogenomic roles

epigenomic aberrations play. (a) *Restriction Length Genomic Scanning (RLGS)* [81] using methylation sensitive enzymes was the first method developed as a genome-wide screen for CpG island methylation. This double-restriction-digest, gel separation technique allows analysis of methylation in up to 4000 loci [82,83]. (b) *Microarray array epigenome analysis* methods [84] proceed via hybridization of tumor and reference DNA samples arrays comprised of oligonucleotides comprised of CpG island sequences after digestion with methylation sensitive restriction enzymes that cut preferentially in CpG islands. Comparison of signal intensities derived from the tumor and reference samples provides a profile of sequences that are methylated in the tumor but not the references (or vise versa). (c) *Reduced representation bisulfite sequencing (RBBS)* [85] is a large-scale genome-wide shotgun sequencing approach in which tumor and reference DNA samples are treated with bisulfate to convert cytosine to uracil while leaving 5-methylcytosine unconverted, digested with a methylation specific enzyme and sequenced. Comparison of CpG sequences in the tumor and reference genomes could then reveal bisulfite induced changes. This method is well suited to next generation single molecule sequencing strategies. (d) *methylation-specific Digital Karyotyping (MSDK)* methodology [86] is a modification of digital karyotyping technique for DNA copy number profiling [87] where sequencing is performed to count accurately tag numbers for comparison between samples, thereby permitting quantitative measurement of methylation events. (e) *Chromatin immunoprecipitation plus microarray analysis* (ChIP on chip)[88] employs immunoprecipitation to enrich DNA sequences associated with chromatin modifications such as histone acetylation and histone H3 methylation for which antibodies are available. Immunoprecipitated DNA sequences are analyzed using microarray technologies or single molecule sequencing strategies.

Acknowledgement

We thank Drs. Ron DePinho, Andy Futreal, Paul Mischel, Alec Kimmelman, Kwok-kin Wong, William Hahn, and Kornelia Polyak for helpful discussion and critical readings of this manuscript. This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Contract DE-AC02-05CH11231) and by the National Institutes of Health, National Cancer Institute P50 Grant CA 58207 and U54 CA 112970 to JWG, and supported by grants from the National Cancer Institute (RO1 CA93947 and CA099041; PO1 CA95616; UO1 CA84313; U24 CA93683) to LC.

Figure Legend

Figure 1. Illustration of various types of genomic and epigenomic aberrations in cancers and type of genomic data that report on such changes. These alterations presumably will lead to altered expression of resident genetic element of interest (GEOI), including coding mRNA and non-coding microRNA. GEOI expression changes can manifest as increased or decreased level of expression or shift in pattern of spliced variant expression or appearance of aberrant transcripts that are cancer specific, such as ones derived from a fusion gene.

Figure 2. Integrative genomic analyses of human cancers of different cell lineages can be triangulated with genomic and biological data from tumors and genetics from model systems and from forward genetic screens in human cell systems. Additionally, association with clinical parameters can prioritize GEOIs. GEOIs identified from such integrative analyses will require biological and clinicopathological validation, a laborious process that can be greatly accelerated by deployment of functional genetic screens. Exploration of mechanistic basis for a GEOI's cancer-relevant activities can provide hints to their uses in the clinics. Results from these downstream activities will feed back to inform and refine analyses and derive improved validation platforms.

Figure 3. Illustrative signaling pathway, highlighted with known examples of *bona fide* cancer genes that are subjected to deregulation by multiple genomic mechanisms.

References

1. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
2. Collins, F.S. & Barker, A.D. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am* **296**, 50-7 (2007).
3. Lynch, T.J. et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129-39 (2004).
4. Paez, J.G. et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497-500 (2004).
5. Pao, W. et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A* **101**, 13306-11 (2004).
6. Druker, B.J. et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* **344**, 1031-7 (2001).
7. Pegram, M. & Slamon, D. Biological rationale for HER2/neu (c-erbB2) as a target for monoclonal antibody therapy. *Semin Oncol* **27**, 13-9 (2000).
8. Malkin, D. et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233-8 (1990).
9. Futreal, P.A. et al. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266**, 120-2 (1994).
10. Miki, Y. et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).
11. Wooster, R. et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789-92 (1995).
12. Marra, G. & Boland, C.R. Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J Natl Cancer Inst* **87**, 1114-25 (1995).
13. Gruis, N.A. et al. Homozygotes for CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds. *Nat Genet* **10**, 351-3 (1995).
14. Nowell, P.C. Discovery of the Philadelphia chromosome: a personal perspective. *J Clin Invest* **117**, 2033-5 (2007).
15. Nardi, V., Azam, M. & Daley, G.Q. Mechanisms and implications of imatinib resistance mutations in BCR-ABL. *Curr Opin Hematol* **11**, 35-43 (2004).
16. Quintas-Cardama, A., Kantarjian, H. & Cortes, J. Flying under the radar: the new wave of BCR-ABL inhibitors. *Nat Rev Drug Discov* **6**, 834-48 (2007).
17. Demetri, G.D. Targeting c-kit mutations in solid tumors: scientific rationale and novel therapeutic options. *Semin Oncol* **28**, 19-26 (2001).
18. Curtin, J.A., Busam, K., Pinkel, D. & Bastian, B.C. Somatic activation of KIT in distinct subtypes of melanoma. *J Clin Oncol* **24**, 4340-6 (2006).
19. Hodi, F. et al. A major response to Imatinib mesylate in KIT mutated melanoma. *JCO* **in press**(2008).
20. Rowley, J.D. The role of chromosome translocations in leukemogenesis. *Semin Hematol* **36**, 59-72 (1999).
21. Tomlins, S.A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).
22. Bradford, T.J., Tomlins, S.A., Wang, X. & Chinnaiyan, A.M. Molecular markers of prostate cancer. *Urol Oncol* **24**, 538-51 (2006).
23. Volik, S. et al. End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci U S A* **100**, 7696-701 (2003).

24. Bignell, G.R. et al. Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Res* **17**, 1296-303 (2007).
25. Schechter, A.L. et al. The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature* **312**, 513-6 (1984).
26. King, C.R., Kraus, M.H. & Aaronson, S.A. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science* **229**, 974-6 (1985).
27. Semba, K., Kamata, N., Toyoshima, K. & Yamamoto, T. A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma. *Proc Natl Acad Sci U S A* **82**, 6497-501 (1985).
28. Coussens, L. et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science* **230**, 1132-9 (1985).
29. Slamon, D.J. et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-82 (1987).
30. Kallioniemi, O.P. et al. ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* **89**, 5321-5 (1992).
31. Cameron, D. et al. A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. *Breast Cancer Res Treat* (2008).
32. Davies, H. et al. Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-54 (2002).
33. Samuels, Y. et al. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
34. Carpten, J.D. et al. A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448**, 439-44 (2007).
35. Stephens, P. et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431**, 525-6 (2004).
36. Sharma, S.V., Bell, D.W., Settleman, J. & Haber, D.A. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* **7**, 169-81 (2007).
37. Blackhall, F., Ranson, M. & Thatcher, N. Where next for gefitinib in patients with lung cancer? *Lancet Oncol* **7**, 499-507 (2006).
38. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-21 (2005).
39. Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8 (2007).
40. Wood, L.D. et al. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-13 (2007).
41. Sharpless, N.E. INK4a/ARF: a multifunctional tumor suppressor locus. *Mutat Res* **576**, 22-38 (2005).
42. Shayesteh, L. et al. PIK3CA is implicated as an oncogene in ovarian cancer. *Nat Genet* **21**, 99-102 (1999).
43. Horvitz, H.R., Shaham, S. & Hengartner, M.O. The genetics of programmed cell death in the nematode Caenorhabditis elegans. *Cold Spring Harb Symp Quant Biol* **59**, 377-85 (1994).
44. Nurse, P., Masui, Y. & Hartwell, L. Understanding the cell cycle. *Nat Med* **4**, 1103-6 (1998).
45. Schreiber-Agus, N. et al. Drosophila Myc is oncogenic in mammalian cells and plays a role in the diminutive phenotype. *Proc Natl Acad Sci U S A* **94**, 1235-40 (1997).
46. Kim, M. et al. Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell* **125**, 1269-81 (2006).

47. Zender, L. et al. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**, 1253-67 (2006).
48. Maser, R.S. et al. Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* **447**, 966-71 (2007).
49. Sweet-Cordero, A. et al. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* **37**, 48-55 (2005).
50. O'Neil, J. et al. Activating Notch1 mutations in mouse models of T-ALL. *Blood* **107**, 781-5 (2006).
51. Chin, L., Garraway, L.A. & Fisher, D.E. Malignant melanoma: genetics and therapeutics in the genomic era. *Genes Dev* **20**, 2149-82 (2006).
52. Hodgson, J.G. et al. Copy number aberrations in mouse breast tumors reveal loci and genes important in tumorigenic receptor tyrosine kinase signaling. *Cancer Res* **65**, 9695-704 (2005).
53. Artandi, S.E. & DePinho, R.A. Mice without telomerase: what can they teach us about human cancer? *Nat Med* **6**, 852-5 (2000).
54. Artandi, S.E. et al. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641-5 (2000).
55. Maser, R.S. et al. DNA-dependent protein kinase catalytic subunit is not required for dysfunctional telomere fusion and checkpoint response in the telomerase-deficient mouse. *Mol Cell Biol* **27**, 2253-65 (2007).
56. O'Hagan, R.C. et al. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2**, 149-55 (2002).
57. Palomero, T. et al. Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nat Med* **13**, 1203-10 (2007).
58. Ewart-Toland, A. et al. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat Genet* **34**, 403-12 (2003).
59. Ewart-Toland, A. et al. Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. *Carcinogenesis* **26**, 1368-73 (2005).
60. Uren, A.G., Kool, J., Berns, A. & van Lohuizen, M. Retroviral insertional mutagenesis: past, present and future. *Oncogene* **24**, 7656-72 (2005).
61. Sharpless, N.E. & Depinho, R.A. The mighty mouse: genetically engineered mouse models in cancer drug development. *Nat Rev Drug Discov* **5**, 741-54 (2006).
62. Westbrook, T.F. et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837-48 (2005).
63. Berns, K. et al. A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* **12**, 395-402 (2007).
64. Staunton, J.E. et al. Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci U S A* **98**, 10787-92 (2001).
65. Bild, A.H. et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353-7 (2006).
66. Konecny, G.E. et al. Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res* **66**, 1630-9 (2006).
67. Hieronymus, H. et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321-30 (2006).
68. Wei, G. et al. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* **10**, 331-42 (2006).
69. Neve, R.M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-27 (2006).
70. Wong, K.K. HKI-272 in non small cell lung cancer. *Clin Cancer Res* **13**, s4593-6 (2007).

71. Scappini, B. et al. Changes associated with the development of resistance to imatinib (STI571) in two leukemia cell lines expressing p210 Bcr/Abl protein. *Cancer* **100**, 1459-71 (2004).
72. Furnari, F.B. et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev* **21**, 2683-710 (2007).
73. Mellinghoff, I.K. et al. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N Engl J Med* **353**, 2012-24 (2005).
74. Stommel, J.M. et al. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**, 287-90 (2007).
75. Korbel, J.O. et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420-6 (2007).
76. Drmanac, R. et al. DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. *Science* **260**, 1649-52 (1993).
77. Sanger, F. & Coulson, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-8 (1975).
78. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80 (2005).
79. Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728-32 (2005).
80. Porreca, G.J. et al. Multiplex amplification of large sets of human exons. *Nat Methods* **4**, 931-6 (2007).
81. Costello, J.F. et al. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**, 132-8 (2000).
82. Dai, Z. et al. An AscI boundary library for the studies of genetic and epigenetic alterations in CpG islands. *Genome Res* **12**, 1591-8 (2002).
83. Plass, C. et al. An arrayed human not I-EcoRV boundary library as a tool for RLGS spot analysis. *DNA Res* **4**, 253-5 (1997).
84. van Steensel, B. & Henikoff, S. Epigenomic profiling using microarrays. *Biotechniques* **35**, 346-50, 352-4, 356-7 (2003).
85. Meissner, A. et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**, 5868-77 (2005).
86. Hu, M. et al. Distinct epigenetic changes in the stromal cells of breast cancers. *Nat Genet* **37**, 899-905 (2005).
87. Leary, R.J., Cummins, J., Wang, T.L. & Velculescu, V.E. Digital karyotyping. *Nat Protoc* **2**, 1973-86 (2007).
88. Collas, P. & Dahl, J.A. Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Front Biosci* **13**, 929-43 (2008).

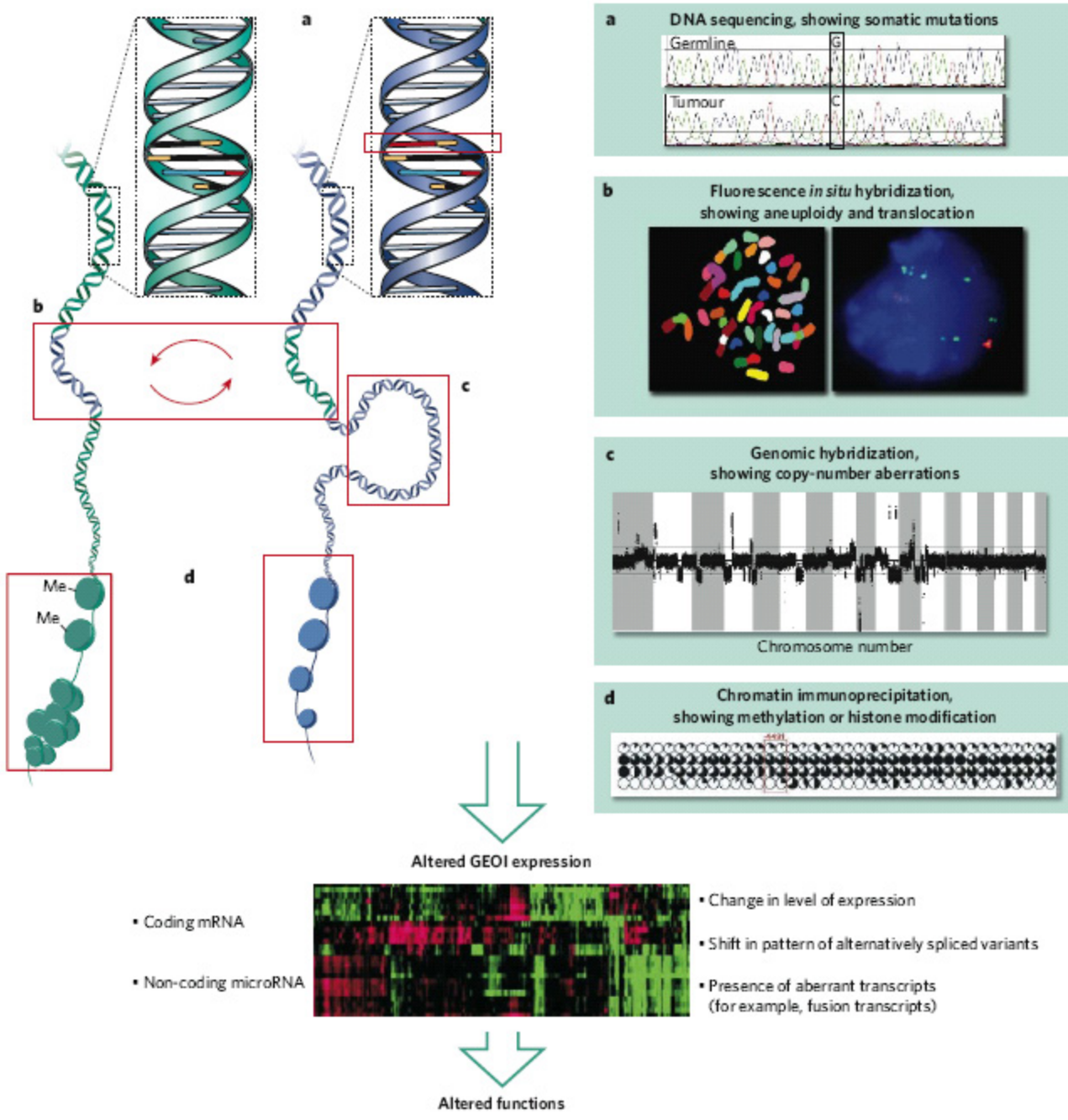

Figure 1 | Various types of genomic and epigenomic aberration in cancers. The main types of genomic and epigenomic aberration are illustrated together with examples of how they can be detected. **a**, Changes in DNA sequence, such as point mutations, can be assessed by DNA-sequencing techniques. **b**, Changes in genomic organization can be assessed by using fluorescence *in situ* hybridization. In the example shown, DNA segments are exchanged between the two (blue and green) DNA molecules. **c**, Changes in DNA copy number, such as those that result from amplification, can be assessed by using comparative genomic hybridization. **d**, Changes in DNA methylation and the resultant changes in chromatin structure can be assessed by using chromatin immunoprecipitation plus microarray analysis of immunoprecipitated DNA. Each of these types of change can alter the expression levels of genes or non-coding microRNAs (referred to here as genetic elements of interest, GEOIs), alter the splicing patterns of transcripts, or change gene function through mutation or through creating chimaeric genes. Many of these events can be as assessed by microarray analysis. These changes ultimately translate into altered functions, leading to a diseased state, such as cancer.

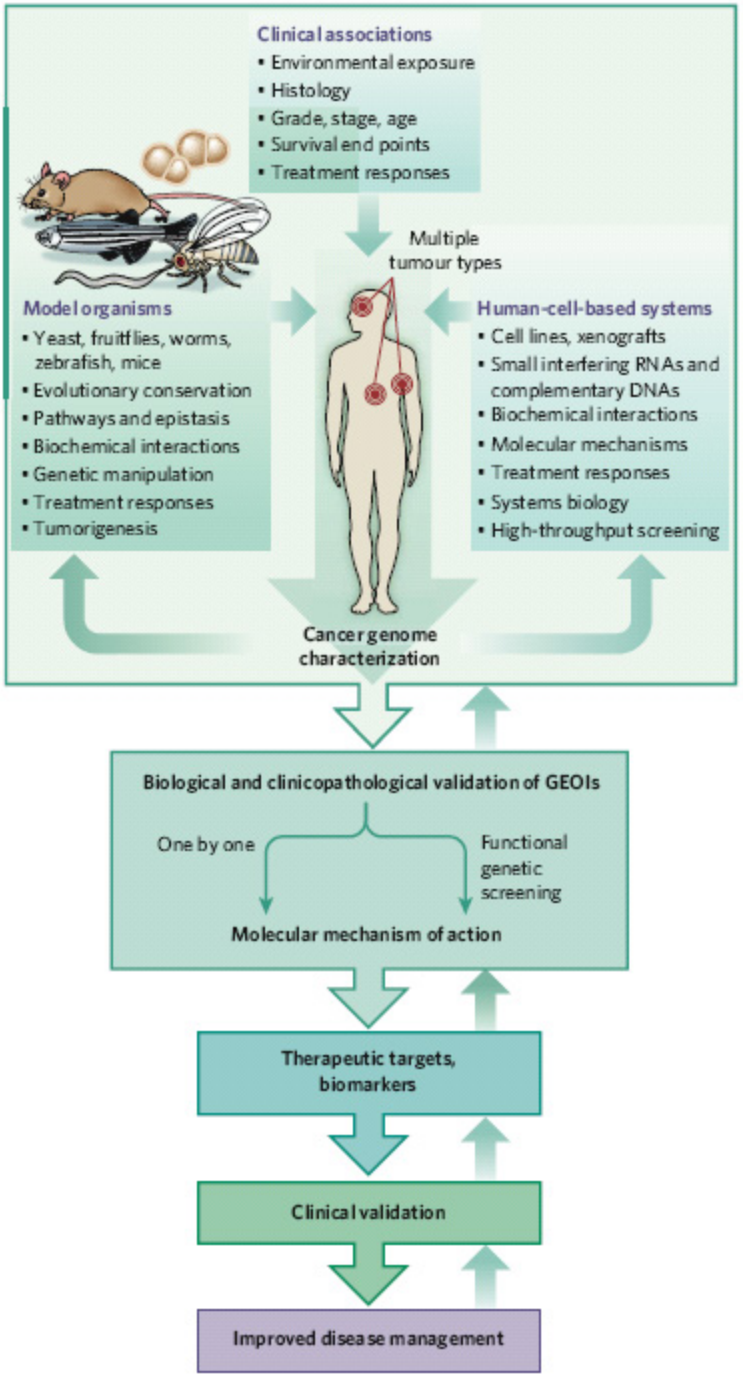

Figure 2 | Integration of complex multidimensional genomic data with insights from other model systems. The identification of cancer drivers or contributors from multidimensional genomic data (such as that shown in Fig. 1) from a particular human tumour type can be facilitated by integration with similar data from other tumour types (for example, by searching for GEOIs that are common to cancers of different lineages). Incorporating clinical information into the analysis of this genomic data helps to narrow the focus to clinically relevant GEOIs. This genomic knowledge can be further filtered by integrating it with insights obtained from studies in model systems. These systems can include non-human model organisms (such as mice, zebrafish, nematodes, fruitflies and yeast), which can be studied to identify evolutionarily conserved GEOIs, to define pathways that GEOIs influence and to elucidate the roles of GEOIs in normal development. Cell-line model systems also can be useful, particularly for functional genetic screening or monitoring responses to drugs. This type of integrative analysis, which extends beyond the cancer genome, is an informative way to identify GEOIs that are likely to be drivers or contributors. After such GEOIs have been identified, they need to undergo stringent biological and clinicopathological validation (Box 2), a labour-intensive process that can be accelerated by carrying out functional screening with a library of GEOIs rather than by assessing one GEOI at a time. For successful translation into the clinic — that is, development of a therapeutic agent that targets the GEOI or a biomarker for the GEOI — a basic understanding of the molecular mechanism of action of the GEOI is helpful, particularly in terms of the specific cellular and genetic context in which it maintains the tumour. Such a biomarker or therapeutic agent will then need to be clinically validated before it can be adopted for routine clinical practice. At each step of this process, the results can be fed back to inform and refine the analyses and to help improve the validation platforms.

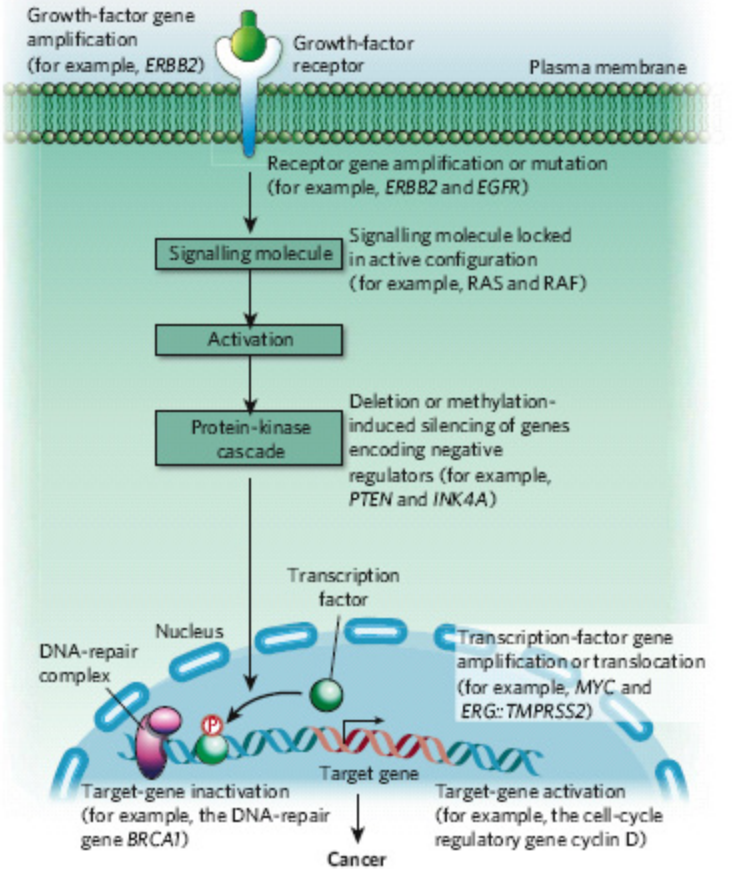

Figure 3 | Disruption of intracellular signalling by alterations in the cancer genome. A simplified signalling pathway is depicted to highlight known examples of *bona fide* oncogenes that are subjected to dysregulation by various mechanisms. It is clear that a signalling pathway can be disrupted at multiple points, and a variety of genomic and epigenomic alterations can contribute to this, ultimately leading to cancer.