Personalized drug repositioning using gene expression

Koudijs, K.K.M.

CHAPTER 2

Transcriptome signature reversion as a method to reposition drugs against cancer for precision oncology

KKM Koudijs
AGT Terwisscha van Scheltinga
S Böhringer
KJM Schimmel
HJ Guchelaar

Cancer Journal, 2019 Mar/Apr;25(2):116-120

ABSTRACT

Transcriptome signature reversion (TSR) has been hypothesized as a promising method for discovery and use of existing non-cancer drugs as potential drugs in the treatment of cancer (i.e. drug repositioning, drug repurposing). TSR assumes that drugs with the ability to revert the gene expression associated with a diseased state back to its healthy state are potentially therapeutic candidates for that disease. This paper review methodology of TSR and critically discusses key TSR studies. In addition, potential conceptual and computational improvements of this novel methodology are discussed as well as its current and possible future application in precision oncology trials.

Introduction

In 2018 an estimated 9.6 million patients died from cancer worldwide and this number is projected to reach over 13 million in 2030.[1,2] This has led to a huge investment into the development of methods to characterize individual cancers at the molecular level (i.e. precision oncology), which has revealed that cancer is a very heterogeneous disease. As a consequence, the treatment of cancer has moved from standard treatments regimens based upon the organ affected to precision medicine based upon the genetic makeup (i.e. somatic mutations) of the tumor.[3] Indeed, most of the novel anti-cancer drugs were developed for treatment of patients with a tumor with a specific somatic mutation such as ALK or BRAF mutations. However, the development of new drugs is not only very costly (> 2 billion[4]) but also takes many years before it reaches the market (> 10 years[5]) and thus is available for patients. In addition, since precision medicine by definition leads to segregation of the drug market, it may prove economically unviable to develop and market new drugs for the increasingly rarer subtypes of cancer.

A potentially cheaper and faster method for discovery of novel drugs in precision medicine of cancer therefore might be to explore whether drugs already on the market for other diseases are effective against specific tumors (i.e. drug repositioning/repurposing). Obviously, the approach of drug repositioning provides the benefit of an established safety and pharmacokinetic drug profile.

The most direct method to discover and pre-clinically validate the anti-cancer activity of drugs would be high-throughput screening to identify associations between characteristics of in vitro/vivo models (e.g. a certain mutation) and the sensitivity to certain drugs. However, the ultimate stand-alone effectiveness of this approach depends both on the internal validity (i.e. technical reproducibility) and external validity (i.e. how well the *in vitro* model represents the drug sensitivity of tumors in patients), both of which have been called into question.[6,7]

Computational drug repositioning approaches may therefore offer a complementary approach to discover new indications of existing drugs. Various computational 'guilt-by-association' approaches exist, which predicts new drug repositioning candidates based on its similarity to an established anti-cancer drug, e.g. through shared side-effects, similarity in chemical structure, similarity in induced gene expression changes upon incubation with the drugs.[8] The benefit of this approach is that the dysregulated pathway being targeted by the original drug has already demonstrated clinical significance. However, drugs identified through a 'guilt-by-association' approach would target the same dysregulated molecular pathway as the original drug and would therefore expected to be of limited additional therapeutic use, unless the new drug is somehow not affected by the same resistance mechanism as the originator.

Transcriptome signature reversion (TSR) is instead ideally suited to discover drugs which are expected to act against dysregulated cancer pathways which are not yet being targeted by established anti-cancer drugs (Figure 2.1).[9,10]

The first step in TSR is to establish which genes are differently expressed (i.e. up- or downregulated) in the cancer cells as compared to healthy cells of the same tissue. The aim is then to find drugs that are able to reverse this expression profile to the expression profile of the healthy control cells. From public databases including over 1,000 different registered drugs, information is available on how these drugs affect the gene expression of cell lines after incubation. The extent of reversal of the tumor expression profile towards the expression of control cells is typically expressed in a score, referred to hereafter as the Reverse Gene Expression Score (RGES). Drugs with the highest RGES are expected to normalize the expression of a large number of genes and are considered potential drug candidates for treating the malignancy.

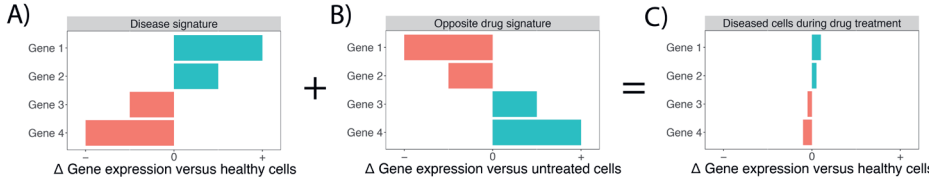

Figure 2.1: Graphical illustration of transcriptome signature reversion (TSR).
A) In the first step, a disease signature is created by determining which genes are upregulated (green) and downregulated (red) in diseased cells compared to the healthy cells. B) In the second step, a database is virtually screened to find drugs with the opposite effect on gene expression. This is determined by comparing the gene expression of drug-treated cell lines with untreated cell lines. C) Drugs with an opposite gene expression signature are considered to be potentially effective drugs to treat the disease because they are expected to revert the gene expression in the diseased cells back to the gene expression levels observed in the healthy cells.

Studies utilizing TSR to find potentially effective anti-cancer drugs

TSR has been used for repositioning drugs against a variety of conditions, such as epilepsy,[11] skeletal muscle atrophy,[12] inflammatory bowel disease[13] and cancer.[14-22] Here, we discuss a selection of studies applying TSR to find new potentially effective anti-cancer drugs.

One of the first and currently most cited (> 450) study to apply was published by Sirota et al. in 2011.[15] First, they integrated the gene expression measurements from 100 diseases and 164 drug compounds. Disease and healthy control tissue gene expression measurements for each disease were downloaded

from the Gene Expression Omnibus (GEO), which is a freely accessible repository for experimentally generated high-throughput gene expression data used by researchers to share their data.[23] The gene expression profiles of drug treated and non-drug treated cancer cell lines were extracted from the 2006 release of the Connectivity Map (CMAP).[24] An integrated analysis of these datasets showed that TSR was able to recover known drug and disease relationships, by clustering drugs and diseases by their predicted therapeutic scores.

Furthermore, to prove that TSR has the ability to predict previously unknown therapeutic relationships, they experimentally validated topiramate as a candidate therapy for Crohn's disease/ulcerative colitis in a rat model of inflammatory bowel disease (IBD) and cimetidine as a candidate therapy for lung adenocarcinoma using cell lines and tumor xenografts implanted in immunodeficient mice. To our knowledge, topiramate and cimetidine have yet to be clinically investigated for these indications. However, cimetidine has been extensively pre-clinically and clinically investigated in other cancer types (even before publication of this study) with encouraging but as of yet inconclusive results.[25] Topiramate as a treatment of IBD may suffer from the critical drawback that is known to induce diarrhea at clinically used doses, which would worsen one of the prominent symptoms of IBD.[26] This case study therefore emphasizes that efficacy at any dose is not enough for clinical application and other factors should always be considered.

In the study by Sirota et al.,[15] both disease and drug signatures were derived from experiments using RNA microarray technology. However, microarrays have recently been surpassed by RNA-sequencing (RNA-seq) as the most preferred method of measuring gene expression.[27] In addition, a new database called the LINCS L1000 (the successor of CMAP) has since become publicly available.[28] The LINCS L1000 contains the gene expression profiles of many more small molecule compounds (currently 19,811) which are tested on a larger panel of cell lines (up to 71) compared to CMAP ($N = 3$). The only downside of the LINCS L1000 database is that a special microarray platform was used which only measures the gene expression of 978 key genes. Although only the expression of a small subset of $\pm$ 20,000 protein coding genes are measured, these genes were preselected based on how well it is possible to computationally infer 81% of the remaining of non-measured transcripts.

Capitalizing on these new developments, Chen et al. published one of the first studies utilizing both RNA-seq data and the LINCS L1000 database in 2017.[21] RNA-seq data of breast invasive carcinoma (BIC), liver hepatocellular carcinoma (HC), colon adenocarcinoma (COAD) and their respective adjacent normal healthy control tissue from The Cancer Genome Atlas (TCGA) was used to generate 3 tumor signatures. By comparing the tumor signatures with the gene expression profiles induced by drugs on cell lines of the same lineage as the tumor tissue, they showed that RGES depended more on the identity of the cell

line than on variations between replicate conditions (e.g. concentration and incubation duration). Which cell line(s) the drugs are tested on may therefore significantly influence the drug signature. Nonetheless, while less important than cell line identity, higher concentrations and longer incubation duration were predictive of higher RGES which emphasizes that testing conditions are important considerations as well. Another key result is that the RGES calculated using the gene expression changes measured in individual cell lines showed a strong positive correlation with experimental data available in ChEMBL on the half-maximal inhibitory concentration (IC_{50}) measured in the same cell line (Spearman Rho 0.33–0.58, $P < 0.003$ for all 3 cell line collections). The summarized RGES (sRGES) was developed to summarize and standardize (at 24 h and 10 μM) the RGES of a tumor signature-drug signature combination across the average cell line and incubation conditions. In addition, sRGES was validated on drug efficacy data from cell lines of all 3 tumor lineages present in Cancer Therapeutic Response Portal (CTRP v2) and growth rate inhibition metrics measured on BIC cell lines. Lastly, 4 compounds with distinct molecular structures not previously studied for efficacy in LIHC and predicted using sRGES to have a low IC_{50} with LIHC cell lines were experimentally validated in a panel of five LIHC cell lines. The resulting median IC_{50}s ranged between 3.18 μM (CGK-733) to 0.07 μM (pyrvinium pamoate). As pyrvinium pamoate (an FDA-approved drug indicated for the treatment of pinworms) had the lowest IC_{50}, it was subsequently further validated using more specific and advanced in vitro and in vivo assays.

Three important additional findings were reported in the study of Chen et al. The first is that calculating the sRGES without cell lines of the same lineage as the tumor signature decreased the correlation between sRGES and IC_{50}, although it remains positively correlated. This suggests that cell lines from the researcher's tumor type of interest do not necessarily need to be present in LINCS to make valid predictions. The second finding is that targeted drugs ($N = 69$) show more variation in RGES between cell lines than cytotoxic drugs ($N = 9$). This confirms the biological intuition that in order to accurately capture the targeted effect of a drug, the target should be sufficiently present in the cell line. The third finding is that although RGES was predictive of efficacy in aggregate, there are some effective compounds (e.g. microtubule inhibitors) which showed no potency to reverse gene expression. A possible explanation is that the mechanism of action of microtubule inhibitors is not being captured by the 978 genes on the LINCS microarray.

Potential conceptual improvements

One limitation of the two aforementioned key studies is that genes which were differentially expressed between the average tumor sample and the average adjacent normal tissue sample were included in the disease signature. This is of course a simplification, as there exists considerable heterogeneity in genetic

identity and treatment response of tumor cells of the same tissue origin between tumors of different patients, within patients and even within tumors.[29] One of the possible improvements is therefore to reposition drugs based on which genes are differentially expressed in individual tumor samples instead of the average tumor sample of a particular tumor type. We showed by using a systematic approach based on clear cell renal cell carcinoma (ccRCC) RNA-seq data from TCGA, that repositioning drugs based on individual ccRCC samples outperformed TSR based on the average tumor signature or the 4 tumor subtypes previously identified using hierarchical clustering.[22]

Another limitation inherent in using gene expression measurements from patient tissues is that most cancer and healthy control tissue gene expression data published in the public domain (e.g. from The Cancer Genome Atlas) used for drug repositioning studies have been analyzed using either bulk microarray or RNA-seq methods.[21,22] Bulk methods do not separate the cell types present in the sample before the analysis and therefore observed differences in gene expression might be confounded by differences in cell type compositions:[30] for example, the tumor tissue may contain relatively more blood vessel cells, different or more immune cells, etc. than adjacent normal tissue. When such data is used to generate the disease signature, and a candidate drug is unintentionally selected because it reverses genes expressed primarily in e.g. the immune cells present in the tumor, the patient is unlikely to experience any therapeutic benefit from the drug. Certainly, in some cases (e.g. angiogenesis inhibitors, immunotherapy) the most clinically relevant target of the drug may be in the tumor micro-environment, but an agnostic approach combined with bulk tissue samples would be unlikely to prioritize drugs which target these specific cell types.

Fortunately, it has become technically feasible to analyze the expression of individual cells with single cell RNA-seq.[31] This completely circumvents the potential issue of sample cell type composition. However, even with the use of single cell RNA-seq data two potential issues remain. The first arises from the fact that it has long been established that many driver events in different pathways are needed to gather all the functionality to become a fully functional malignant cell,[32] whereas most candidate drugs probably only target one dysregulated pathway. And given that typically many different pathways are dysregulated, each additional dysregulated pathway increases the size of the haystack (total number of genes to be potentially reversed) while the number of needles (genes targeted by potentially effective drugs) stays the same. Although overall one would still expect a net reversion for a drug, attempting to revert all genes makes it statistically harder to find potentially effective drugs, and lowers the predictive value of the procedure by increasing the ratio of false positive to true positive findings. While this issue may seem most important when comparing tumor cells to normal cells, the same issue can arise with other contrasts. For example, it is likely that even within the same patient and tumor, metastatic/drug resistant tumor clones differ genetically (e.g. different mutations, amplifications, deletions,

etc.) in more than 1 way from drug sensitive/non-metastatic tumor clones. The second potential issue concerns the acquisition of "non-functional" collateral damage sustained by tumor cells along the evolutionary path to become malignant. If this is only caused by random mutations/deletions/amplifications occurring in between acquiring the driver events, it would average out given enough tumor tissue samples are analyzed. However, systematic causes of collateral damage do exist such as a commonly observed mechanism used to deactivate a tumor-suppressor gene: on one chromosome of a pair the allele might be mutated whereas on the other chromosome, the entire chromosomal region might be deleted. Because supposedly the only reason the entire chromosomal region was deleted is to inactivate both copies of a single gene, it is implausible that reversing all the genes on the 1) the deleted chromosomal region; and 2) the genes on other chromosomes affected by the chromosomal deletion back to the normal expression levels will lead to reversion of the undesirable phenotype.

These 2 potential issues might be addressed in the following way. If enough tissue samples are taken from the same tumor and analyzed using single cell RNA-seq, it may be possible to identify pairs of tumor clones which are genetically identical (i.e. same mutations, chromosomal amplifications/deletions, etc.) except for the existence of a key driver event. Alternatively, if such a clean comparison is not possible and the driver event of interest is not completely confounded with other genetic events, it might be reasonable to approximate such a contrast using a regression model containing all genetic events as parameters. Otherwise, the best possible approach might be to take the overlap between 1) all genes dysregulated compared to healthy cells and 2) genes dysregulated after the driver event is edited in using laboratory techniques such as CRISP-CAS.

Potential computational improvements

Batch effect correction

Because of the enormous volume, the data available in CMAP and LINCS was generated and analyzed in a batch wise manner. The basic unit of each batch is the plate, with CMAP using a combination of 6 and 96 well plates and LINCS using 384 well plates. Because of similar handling and environmental conditions, samples within each batch are influenced by the same unwanted technical variation which can affect the statistical power to detect differentially expressed genes and the external validity of generated gene expression signatures.[33] It is therefore recommended to use one of the many different methods available which were developed to reduce this unwanted variation.[34]

More advanced methods to calculate reversal potency

The theoretical advantages and disadvantages of a representative sample of earlier methods have all recently been reviewed by Musa et al.[35] In this section we would therefore like to highlight two newer methods because they claim to represent important recent advancements.

The first is EMUDRA (Ensemble of Multiple Drug Repositioning Approaches).[36] EMUDRA combines the output of a newly developed algorithm (EWCos) with 3 state-of-the-art methods (Cosine, XCor and XSpe). Zhou et al. showed that EMUDRA outperformed all other single and combined approaches on 1) Simulated datasets, 2) Retrieval of drugs from LINCS using drug signatures constructed using CMAP and 3) Predicting which drug pairs share the same fourth level ATC code (i.e. drug pairs with similar chemical, therapeutic and pharmacological properties).

However, one important limitation which all previously published methods (including EMUDRA) have in common is that the researcher still has to choose a cutoff for including genes in the disease signature. Which cutoff is used at what level is not standardized with a result that these cutoffs vary wildly between different studies. To illustrate, in the 14 studies referenced by Musa et al. the following variety of cutoffs were used:[37]

- Only genes which are differentially expressed below a statistical significance cutoff, such as $P < 0.05$, 0.001, 0.0001 or a False Discovery Rate (FDR) below 1%, 5%, 25%;
- Sometimes in combination with a minimum Fold Change above 1.5, 2 or even 4;
- In some studies, no formal statistical significance criterion is used and instead the top X (e.g. 75 or 100) most up- and downregulated genes are selected for reversion.

The number of genes which ended up in the disease signature across these diverse studies subsequently ranged widely from 21 up to 1,000. This has been one of the main motivations behind the development of Dr. Insight (Drug Repurposing: Integration and Systematic Investigation of Genomic High-Throughput data), a "signature free" method, i.e. the researcher does not have to choose a cutoff.[37] This method was benchmarked to 4 existing methods (The KS method implemented in the original CMAP study, ssCMAP, NFFinder and Cogena) and demonstrated superior performance in simulated datasets and three cancer datasets. Additionally, Dr. Insight was shown to be robust to simulated additional experimental noise, even when up to 90% of extra noise was added to the expression profiles.

Unfortunately, to date no independent systematic and quantitative benchmarking of signature reversal methods has been published which includes both EMUDRA and Dr. Insight. Computational validation is further complicated by the fact that there is no general agreement on the best way to validate whether a computational drug repositioning method results in clinically efficacious repositioning hypotheses.[38]

The key issue is that since true positive and true negative hits are unknown, it is impossible to establish the sensitivity and specificity of computational drug repositioning methods in an unbiased manner using real data. Therefore, simulation studies based on realistic assumptions are likely the best alternative to choose the current best TSR algorithm.

Current and possible future application of TSR in precision oncology trials

In all TSR studies reviewed thus far the selected drugs were not given directly to patients. However, in theory TSR can be relatively easily personalized by simply replacing the average tumor signature with the patient's own tumor signature. In August 2018, the first trial was published which utilized RNA sequencing and the underlying principle of TSR to reposition drugs against relapsed multiple myeloma (MM).

From February 2014 to February 2016, 64 patients with relapsed multiple myeloma (MM), a prognosis of 6 months of survival and lack of FDA-approved treatments were included in the trial.[39] DNA and RNA sequencing were performed on CD138+ cells isolated from the bone marrow. In this highly pretreated cohort of patients, an overall response rate of 66% with a median duration of response of 131 days was achieved in the 21 evaluable patients with 5 patients experiencing a durable response.

Patients received drugs based on either DNA sequencing (N = 8), RNA-seq (N = 11) or both (N = 2). This trial therefore illustrates that selecting drugs based on RNA-seq can have a complementary value to selecting drugs based on DNA-sequencing alone. However, recommended drugs were prioritized based on prescriber discretion and on whether there were any specific associations with MM in the CIViC (Clinical Interpretations of Variants in Cancer) database.[40] Furthermore, for most drugs given based on RNA-seq (trametinib, venetoclax, panobinostat), the target gene or pathway was used instead of the corresponding drug signature from the LINCS L1000 database. Seemingly only etoposide was directly chosen because of an opposite relationship with the patient's gene expression profile using their TSR-based drug repositioning algorithm, although observed activation of the HDAC pathway frequently coincided with the selection of vorinostat (a HDAC-inhibitor) using the TSR-based drug repositioning algorithm.

This clinical trial thus raises the question of how TSR fits into the current clinical drug repositioning landscape. If drugs are always prioritized based on existing clinical evidence of anti-cancer activity, it will be hard to clinically test promising non-cancer drug repositioning candidates in precision oncology trials. One way out of this dilemma would be to establish clinically predictive *ex vivo* models of tumor types and screen drug repositioning candidates on these models.

Future perspective and concluding remarks

TSR has already proven itself to be a cost and time efficient method to rapidly screen for potentially effective non-cancer drugs which frequently are proven effective, at least under laboratory conditions. However, up to now attempted validation in clinical trials has been severely lacking.

Because of the heterogenous nature of cancer between and even within patients, it makes scientific and clinical sense to reposition drugs on an individual patient basis. However, a single tumor and healthy tissue sample makes it impossible to distinguish which of the differentially expressed genes compared to healthy cells are actually driving tumor growth and metastasis, and which genes are differentially expressed because of passenger events and/or non-functional collateral damage. Single cell RNA-seq and the analysis of multiple tumor samples from the same patient may make the identification of key genes to revert more tractable in the future.

Another key issue is the external validity of the drug signatures for the individual patient's tumor cells. From the reviewed analysis of cell lines present in LINCS we already know that the drug signature at a standardized concentration and incubation duration varies significantly between cell lines, and thus the predicted expected reversion of genes is expected to vary significantly between tumor clones of individual patients as well. Although drug signatures calculated over the average cell line at standardized conditions proved predictive in aggregate, individual *in silico* prediction of efficacy may have too much uncertainty to directly clinically test predicted drug repositioning candidates on individual patients. This uncertainty is further compounded by the availability of the multitude of computational algorithms and lack of standardization of TSR studies.

Ex vivo validation of individual patient drug repositioning candidates may therefore be the solution to reduce the uncertainty associated with individual drug repositioning predictions.[41] This will require the existence of clinically predictive *ex vivo* models but it can provide an independent check of predicted drug repositioning candidates, and a relative benchmark against drugs hypothesized to be effective through other methods (e.g. actionable mutation-based drug repositioning).

References

1. Worldwide cancer statistics. Cancer Research UK, 2018. (Accessed 2018- 01-23 at https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer.)
2. Key cancer statistics. World Health Organisation. (Accessed 2018-01-23, at https://www.who.int/cancer/resources/keyfacts/en/)

3. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014;158:929-44.
4. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33.
5. Van Norman GA. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. JACC Basic Transl Sci. 2016;1:170-9.
6. Haibe-Kains B, El-Hachem N, Birkbak NJ, et al. Inconsistency in large pharmacogenomic studies. Nature. 2013;504:389-93.
7. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat Med. 2011;17:320-9.
8. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. Brief Bioinform. 2016;17:2-12.
9. Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J. Transcriptional data: a new gateway to drug repositioning? Drug Discov Today. 2013;18:350-7.
10. Nevins JR, Potti A. Mining gene expression profiles: expression signatures as cancer phenotypes. Nat Rev Genet. 2007;8:601-9.
11. Brueggeman L, Sturgeon ML, Martin RM, et al. Drug repositioning in epilepsy reveals novel antiseizure candidates. Ann Clin Transl Neurol. 2018; acn3.703.
12. Dyle MC, Ebert SM, Cook DP, et al. Systems-based discovery of tomatidine as a natural small molecule inhibitor of skeletal muscle atrophy. J Biol Chem. 2014;289:14913-24.
13. Dudley JT, Sirota M, Shenoy M, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med. 2011;3:96ra76.
14. Wang SE, Xiang B, Guix M, et al. Transforming growth factor beta engages TACE and ErbB3 to activate phosphatidylinositol-3 kinase/Akt in ErbB2-overexpressing breast cancer and desensitizes cells to trastuzumab. Mol Cell Biol. 2008;28:5605-20.
15. Sirota M, Dudley JT, Kim J, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011;3:96ra77.
16. Wang G, Ye Y, Yang X, Liao H, Zhao C, Liang S. Expression-based in silico screening of candidate therapeutic compounds for lung adenocarcinoma. PLoS One. 2011;6:e14573.
17. Muthuswami M, Ramesh V, Banerjee S, et al. Breast tumors with elevated expression of 1q candidate genes confer poor clinical outcome and sensitivity to Ras/PI3K inhibition. PLoS One. 2013;8:e77553.
18. Schnell SA, Ambesi-Impiombato A, Sanchez-Martin M, et al. Therapeutic targeting of HES1 transcriptional programs in T-ALL. Blood. 2015;125:2806-14.
19. Fortney K, Griesman J, Kotlyar M, et al. Prioritizing therapeutics for lung cancer: an integrative meta-analysis of cancer gene signatures and chemogenomic data. PLoS Comput Biol. 2015;11:e1004068.
20. Churchman ML, Low J, Qu C, et al. Efficacy of Retinoids in IKZF1-Mutated BCR-ABL1 Acute Lymphoblastic Leukemia. Cancer Cell. 2015;28:343-56.
21. Chen B, Ma L, Paik H, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat Commun. 2017;8:16022.
22. Koudijs KKM, Terwisscha van Scheltinga AGT, Böhringer S, et al. Personalised drug repositioning for Clear Cell Renal Cell Carcinoma using gene expression. Sci Rep. 2018;8:5250.
23. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol. 2016;1418:93-110.

24. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313:1929-35.
25. Pantziarka P, Bouche G, Meheus L, et al. Repurposing drugs in oncology (ReDO)-cimetidine as an anti-cancer agent. Ecancermedicalscience. 2014;8:485.
26. eLetter in response to "Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease by Dudley et al." at http://stm.sciencemag.org/content/3/96/96ra76/tab-e-letters.)
27. Lowe R, Shirley N, Bleackley M, et al. Transcriptomics technologies. PLoS Comput Biol. 2017;13:e1005457.
28. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171:1437-52.e17.
29. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. Br J Cancer. 2013;108:479-85.
30. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015;6:8971.
31. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018;50:96.
32. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.
33. Goh WWB, Wang W, Wong L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. Trends Biotechnol. 2017;35:498-507.
34. Lazar C, Meganck S, Taminau J, et al. Batch effect removal methods for microarray gene expression data integration: a survey. Brief Bioinform. 2013;14:469-90.
35. Musa A, Ghoraie LS, Zhang SD, et al. A review of connectivity map and computational approaches in pharmacogenomics. Brief Bioinform. 2017;18:903.
36. Zhou X, Wang M, Katsyv I, et al. EMUDRA: Ensemble of Multiple Drug Repositioning Approaches to improve prediction accuracy. Bioinformatics. 2018;34:3151-9.
37. Chan J, Wang X, Turner J, et al. Breaking the paradigm: Dr. Insight empowers signature-free, enhanced drug repurposing. Bioinformatics. 2019.
38. Brown AS, Patel CJ. A review of validation strategies for computational drug repositioning. Brief Bioinform. 2018;19:174-7.
39. Laganà A, Beno I, Melnekoff D, et al. Precision Medicine for Relapsed Multiple Myeloma on the Basis of an Integrative Multiomics Approach. JCO Precision Oncology. 2018:1-17.
40. Griffith M, Spies NC, Krysiak K, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017;49:170-4.
41. Lee JK, Liu Z, Sa JK, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. Nat Genet. 2018;50:1399-411.