Personalized drug repositioning using gene expression

Koudijs, K.K.M.

CHAPTER 8

General discussion

Introduction

Transcriptome Signature Reversion (TSR) has been used as a drug repositioning method to find new anti-cancer treatments, both before[1-4] and after[5-7] we started to work on this thesis. Indeed, the hypothesis that the potential of drugs to normalize the gene expression profile of a cancer type could be predictive of its therapeutic potential in treating that specific cancer type is both plausible and appealing. If TSR indeed works as intended, it would significantly improve our ability to reposition drugs for cancer treatment, saving a lot of time, money and patient life years.

To perform TSR, one first needs to create a gene expression signature of the disease for which the new drug treatments are to be found. In the context of finding new anti-cancer drugs, this is called the tumor gene expression signature. Typically, this list is created by contrasting the average gene expression of all tumor tissue samples belonging to specific tumor type (e.g., breast carcinoma) with the average gene expression of adjacent normal tissue samples, the so-called differential expression analysis and genes which differ in expression level between the two groups based on statistical significance are called differentially expressed genes (DEG). The tumor gene expression signature is then created by either taking all the DEG or a subset of the DEG after applying cutoff, for example an additional filter for the magnitude of the change in expression such as at least a 2-fold increase or decrease in expression. The second step is to create the drug gene expression signatures. These are created by performing a differential expression analysis of the gene expression of cell lines after drug exposure compared to control cell lines without drug exposure. The third and final step then consists of calculating how much each drug gene expression signature is expected to reverse the expression of the genes in the tumor gene expression signature. This is summarized using the connectivity score, which is normalized from -1 (complete expected reversion by the drug gene expression signature of the tumor gene expression signature) to +1 (genes of the drug gene expression signature are differentially expressed in the same direction as the tumor gene expression signature).

From clinical practice, it is well known that there exists a high extent of interindividual variability in drug response among similar tumor types. We therefore hypothesized that instead of using all tumor samples to create a single tumor gene expression signature, it may be a better to create a unique tumor gene expression signature for each individual tumor sample ('individualized drug repurposing'). In **chapter 3** we tested this approach using 534 tumor samples and 72 matched normal samples from 530 clear cell renal cell carcinoma (ccRCC) patients from The Cancer Genome Atlas (TCGA). These 3 different types of tumor gene expression signatures (average, subtype and individual) were used to calculate connectivity scores with the over 20,000 drug gene expression signatures created from the Connectivity Map (CMAP) database and its successor the Library of Integrated Network-based Cellular

Signatures (LINCS) database. We found that more drug gene expression signatures were able to show expected reversion of the individual tumor gene expression signatures than those of the 'ccRCC average' and 'ccRCC subtypes' tumor gene expression signatures, based on statistical significance. In addition, some drugs which are used for ccRCC treatment (sirolimus and temsirolimus) produced statistically significant negative enrichment scores for many of the individual tumor gene expression signatures (i.e., more than expected by chance as demonstrated by simulations), but produced no statistically significant negative enrichment scores for the 'ccRCC average' and 'ccRCC subtypes' tumor gene expression signatures. These drugs were therefore rediscovered by the individual tumor gene expression signature TSR approach but not using the average and subtype TSR approach.

An inherent limitation of the bulk RNA-seq data we used to create the individual tumor gene expression signatures is that the tumor samples consist of varying fractions of tumor cells and non-tumor cell types, which impacts the chance of genes being included in the tumor gene expression signature. For example, if the ccRCC tissue samples contain more immune cells than the adjacent normal tissue samples, a lot of genes associated with immune cell activity may end up in the tumor gene expression signatures which could significantly impact which drug gene expression signatures are expected to reverse the tumor gene expression signature. In **chapter 4** we tested whether the top 8 drug repositioning candidates we identified in **chapter 3** showed more or less negative reversion of the tumor gene expression signatures as the estimated fraction of tumor cells in each sample increased. In our initial analysis, we observed that as the fraction of tumor cells in a sample increased, the connectivity score with these 8 drugs trended from negative (i.e., significant reversion of the tumor gene expression signature) to neutral (i.e., no significant reversion of the tumor gene expression signature). We eventually found a plausible explanation for these surprising results: the dataset not only included ccRCC samples, but was contaminated with samples from 2 different tumor types, namely chromophobe and papillary RCC. Tissue samples from ccRCC contain a lot more endothelial cells as they are highly permeated by blood vessels, and therefore on average contain far lower fractions of tumor cells. After stratification of the samples by how much the tumor sample showed negative reversion to two HIF inhibitors (which target one of the key gene expression pathways which is specifically dysregulated in ccRCC cells), the 307 samples which showed negative reversion to the HIF-inhibitors no longer showed a statistically significant trend of increasing connectivity scores with increasing fraction of tumor cells for 7 out of the 8 drugs. In contrast, for the 214 samples with less apparent reversion with the HIF-inhibitors (thus less likely to be ccRCC tumors), showed a much stronger positive association between the fraction of tumor cells and the connectivity score for all 8 drugs (all P-values < 0.01). From this result we concluded that the drugs we identified were indeed more likely to be targeting ccRCC tumor cells in the samples and not one of the other cell types.

In **chapter 5** we decided to perform a systematic validation of TSR for its intended use as a method to reposition drugs against specific tumor types. Interestingly, this validation study convincingly showed that the reversion of the tumor gene expression signature appears to be only a proxy for the general anti-proliferative effect of a drug but not for the tumor type specific effects. This finding could be substantiated by two crucial analyses. First, when the gene expression signatures of drugs were adjusted to remove the effect of the general anti-proliferative downstream effects of drug-induced decreased cell viability, the amount of reversion of a tumor gene expression signature by the adjusted drug gene expression signature was no longer predictive of the drug's ability to induce cell death to cell lines belonging to the same tumor type. Second, when a much stronger proxy of each drug's anti-proliferative ability, i.e., the mean fraction of cells left after administration of the drug to cell lines not belonging to the same tumor type, was included in the regression analysis alongside with the variable which quantified the amount of gene expression reversion (the connectivity score), the connectivity score variable explains almost none of the variance and is not statistically significant for any of the 18 tumor types after correction for multiple testing.

While indeed the potential to reverse the gene expression signatures of tumor types does have some predictive power for effectiveness, we found a biological explanation for this observation which might limits its usefulness as a means to find new anti-cancer drugs for specific cancer types, especially for individualized treatment. The key to this explanation is that one of the main well-known hallmarks of cancer cells is uncontrolled proliferation.[8,9] Subsequently, when the gene expression of samples consisting mostly of tumor cells is compared to samples consisting mostly of non-tumor cells, genes associated with cell proliferation will appear to be upregulated in the tumor samples (**chapter 5**). Conversely, cell lines exposed to drugs which limit cell proliferation or kill cells, will show downregulation of genes associated with cell proliferation compared to untreated control cell lines.[10] These two phenomena are therefore diametrically opposing, which causes the apparent reversion of the tumor gene expression signature.

In conclusion, while there does appear to be a grain of truth behind the use of TSR as a drug repositioning method, it does not appear to be of any clinical use as it is currently implemented. In our opinion, there are 3 critical issues with the typical current use of TSR as a drug repositioning method to find new anti-cancer drugs which would need to be solved simultaneously for it to become more useful:

1. Bulk tissue gene expression analysis measures the combined expression of all cell types present in the sample;
2. The gene expression impact of driver events and passenger events are treated equally in the analysis;
3. The focus on finding new anti-proliferative drugs.

Each of these 3 issues and their possible solutions are discussed in more detail below.

Issue 1: Bulk tissue gene expression analysis measures the combined expression of all cell types present in the sample

To our knowledge, all TSR-based drug repositioning analyses to date use bulk tissue microarray or RNA-seq analysis of tumor tissue samples to create the tumor gene expression signature. The problem with bulk RNA-seq is that the gene expression of all cell types in the tissue sample (i.e., immune, endothelial, stromal and tumor cells) is measured simultaneously. When measuring the bulk tissue RNA-seq tumor samples, all cell types present in the tumor-microenvironment would therefore contribute to the signal measured. This complicates the interpretation of the results because it is unknowable about how much each cell type contributed to each RNA transcript measured, and therefore the perceived up or downregulation of any transcript may be due either due to 1) expression changes in the tumor cells, 2) expression changes in one or more of the other cell types or 3) changes in the cell type composition of the sample (i.e., higher fractions of some cell types, lower fractions of others).

Interpreting the differential expression analysis of tumor bulk RNA-seq samples to adjacent 'normal' bulk RNA-seq tissue samples is even more difficult, as the adjacent normal tissue also consists of varying mixtures of different non-cancer cell types, none of which may be of the cell type from which the tumor cell originated. For example, in the case of ccRCC the cell-of-origin is highly suspected to be a cell type found in the proximal convoluted tubule of a nephron of the kidney.[11] It is also known that tumor cells retain much of the expression patterns from their cell-of-origin.[12] Therefore, if a tumor type with on average 60% tumor cells is compared to adjacent normal tissue consisting on average of 10% of the cell type from which the tumor cells originated, one of the strongest expected signals in such a differential expression analysis would actually be in genes associated with the 'healthy' cell-of-origin functioning (i.e., 6-fold upregulation assuming expression stayed constant and the transcript is exclusively expressed in tumor cells and their cell-of-origin). Normalizing these genes back to the 'normal' levels would therefore be unlikely to produce any therapeutic benefits and may be even be detrimental to the normal functioning of the cell-of-origin.

A possible solution to this problem would be to somehow correct for the fraction of tumor cells in each sample, so that you could extrapolate what would be the gene expression at a sample with 100% tumor cells. We did something similar in **chapter 4** plotting the connectivity score against the fraction of tumor cells as estimated from the ABSOLUTE-algorithm. However, we did not see a strong trend for the connectivity score becoming more negative as the estimated fraction of tumor cells increased. We suspect this is the case because ABSOLUTE algorithm is based on the DNA-seq data which is performed on a different fraction of the physical tissue sample as the RNA-seq. Indeed, as other research has shown, there is little concordance between DNA- and mRNA-derived estimates of the fraction of tumor cells in

each sample.[13] From our own experience with comparing the many different mRNA-based algorithms to estimate the fraction of tumor cells, we concluded that these algorithms give conflicting results, both in the fraction of tumor cells present as well as the fractions of other cell types present. However, perhaps in the future one of the mRNA-based algorithms is able to accurately estimate the fraction of tumor cells. With accurate estimates of the fraction of tumor cells in each sample, you could theoretically infer the expression at 100% tumor cells for most genes. However, it would remain difficult to accurately infer the expression changes for genes which are expressed at lower levels inside the tumor cells as compared to other cells of the tumor-microenvironment. Assume for example that the typical tumor sample contains 50% tumor cells and the gene of interest is expressed at twice the rate in the non-tumor cells on average. Thus, 1/3 of all transcripts from this gene come from the tumor cells in this example. If the mutation of interest halves the expression inside of the tumor cells, the total expression would drop only by 1/6 (17%), which is difficult to reliably detect with statistical significance. In fact, because the observed effect size would be 3 times lower, you would need 9 times as many samples (3^2) to have the same statistical power to detect the same difference. This example even leaves out the additional noise the varying fractions of non-tumor cell types have on the transcript level, which increases the standard error and further increases the number of samples needed.

Therefore, the far superior solution to this problem would be to use single cell RNA-seq data instead. Using single cell RNA-seq, it is currently already possible to measure the gene expression of up to 20,000 individual cells at the same time.[14] This would make it possible to compare the expression of the tumor cells directly to the expression of the cell-of-origin, assuming any are present in the tissue samples. Categorizing all the differentially expressed genes into different groups based on the function each gene serves therefore becomes the next crucial step, which leads directly into issue 2.

Issue 2: The gene expression impact of driver events and passenger events are treated equally in the analysis

While using single cell RNA-seq to compare the expression of the tumor cells to those of the cell-of-origin is already conceptually much better than the much noisier bulk tumor vs. adjacent normal tissue contrast, not all gene expression changes between the tumor cells and cell-of-origin cells are likely to be beneficial to be a target for reversion. Many of the original functions performed by the cell-of-origin likely will not have benefitted the survival of the tumor cells and will have been progressively 'turned off' through evolutionary selection to conserve resources for many of the new functions the tumor cells have to perform, such as evading the immune system, suppressing cell death signals, sustaining angiogenesis, etc. These new functions are developed by acquiring driver events. A driver

event is defined as any alteration which causally contributes to tumor evolution and progression (i.e., confers a causal selective advantage to the cancer cell) whereas a passenger event confers little to no selective advantage[15] and can be seen as merely associated with progression. A typical tumor is the result of many years of evolution, in which many different driver and passenger events accumulate in the increasingly unstable genomes of the dividing cancer cells.

To illustrate this point, from the paper describing the same dataset we used for **chapters 3 and 4**,[16] it is reported that inside the average ccRCC tumor cell, 17% of the genome is altered by amplifications/deletions of large chromosomal regions and another 0.4% is altered by more limited, focal alterations. The number of mutations per sample are relatively low compared to other tumor types, with only a median number of detected mutations of around 50. However, a mutation in a single gene (SETD2) was associated with widespread DNA hypomethylation, potentially affecting thousands of genes. Complicating matters further, various possible fusion genes can occur. The most common genetic events in ccRCC are visualized in Figure 8.1.

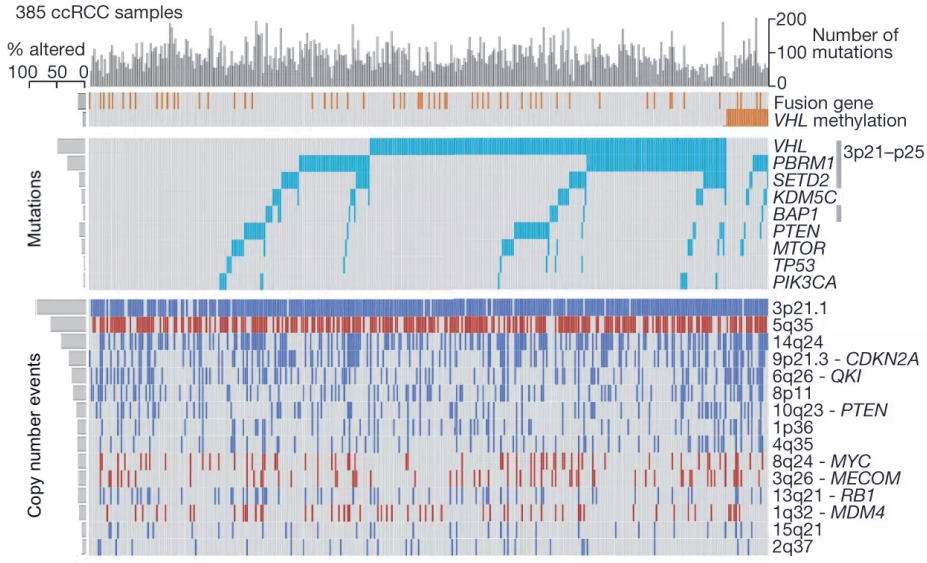

Figure 8.1: Most common genetic events in ccRCC.
Figure adapted from [16] and altered to reduce the complexity.

Some patterns are easy to see in Figure 8.1. For example, to completely deactivate a gene, both working copies of the gene need to be disabled. Therefore, deletion of chromosome 3p, the near-universal initiating event in the origin of ccRCC tumors,[17] results in only a single working copy of genes located on chromosome 3p remaining. Consequently, the genes with the most common ccRCC mutations (i.e.,

VHL, PBRM1, SETD2, BAP1) are all located on chromosome 3p, because a single mutation is enough to deactivate all working copies of these genes. In some cases, the remaining working VHL gene is not mutated but deactivated through methylation instead. Alternatively, in the rare cases of ccRCC in which chromosome 3p is not deleted, VHL is deactivated by 2 mutations.

In one way, using gene expression data is convenient because regardless of the way a gene is inactivated, the expected effect on the gene expression phenotype would likely be similar. However, the downside of gene expression analysis of clinical samples is that the impact of all events is measured at the same time, making it hard to disentangle which events are responsible for which changes in the gene expression. This is absolutely critical, because it is likely only to be therapeutically useful to reverse the gene expression changes caused by driver events (e.g., those resulting from inactivation of VHL, PBRM1, SETD2 and BAP1 genes).

To perfectly disentangle the impact of a single driver event, the single cell RNA-seq gene expression of tumor clone with the driver event should be compared to a tumor clone without the driver event, with both tumor clones having the same other driver and passenger events (i.e., the direct ancestor of the new clone with the driver event). Unfortunately, encountering such a perfect contrast inside a clinical tissue sample would be statistically highly unlikely. The next best solution would be to combine the gene expression results from many different tumor clones with different driver and passenger events using a statistical model. These different tumor clones can come from the same tumor, but also from tumors from many other patients so that the causal impact of the driver event can be estimated more robustly across different tumor micro-environments and the effect of different driver and passenger events is averaged out. It is also potentially possible to identify driver events statistically, if unknown.

As experimental confirmation, the driver event could be introduced to cell lines or in implanted tumors in animals using e.g., CRISPR/CAS to introduce the driver event inside the tumor cells of the experimental group and compare the gene expression to tumor cells without the driver event. The disadvantage of this method would be the reduced external validity for the situation inside patients due to the lack of a representative human tumor micro-environment, which could have taken years to grow in patients. However, there would be some way to check the internal validity of the results by observing whether the tumor cells with the driver event indeed 'outperform' the control tumor cells in the relevant hypothesized metric specific to the driver event (such as e.g., increased tumor cell proliferation, resistance to drug treatment or increased metastasis rate).

Using statistical analysis of the mutation frequency of genes in 33 different tumor types, researchers have already identified around 3,400 driver mutations in 299 driver genes, of which in a limited sample 60–85% were experimentally confirmed.[18] Some of the identified driver genes are considered

drivers in many tumor types (e.g., *TP53* was an identified driver gene in 27 tumor types) but yet other identified driver genes are unique to only a single tumor type, underlining the context dependence of each driver event. How beneficial each driver event is to the evolutionary fitness of the tumor cells most likely depends on the interaction between active gene expression pathways retained from the cell-of-origin, the other cell types present in the early and late tumor micro-environment, preceding driver events, the patient's immune system and past drug treatments.

While it is unlikely that the effect of any specific driver event can be targeted using any of the drugs already on the market, because of the many different driver events, it seems plausible that at least a few driver events will be. However, most drugs on the market are non-cancer drugs with very limited directly measurable anti-proliferative properties when being used at marketed dosages. For this reason, broadening the focus to disrupt other cancer hallmarks might prove to be a more successful strategy. This is discussed further in Issue 3.

Issue 3: The focus on finding new anti-proliferative drugs

Testing the effectiveness of the drugs should be broadened to a larger context outside of *in vitro* cell proliferation assays. While this is the simplest way to measure the effectiveness of a drug, it only covers a limited selection of the functions tumor cells are required to perform to grow and spread *in vivo*. Furthermore, if a drug has anti-proliferative properties, it probably would have been identified already unless it only works that way in a very specific context: as **chapter 5** showed, over 95% of the variation in decreased cell viability after drug exposure in cell lines of a specific tumor type could already be explained by the average decreased cell viability in other cell lines.

Part of the reason could be that the current *in vitro* high-throughput assays only test a single aspect (i.e., reduced or increased tumor cell proliferation in a limited timeframe) whereas tumor cells *in vivo* have many more functions to fulfill. To illustrate this point, consider the following 10 cancer hallmarks and enabling characteristics which have been described in the literature:[9] #1. 'Evading growth suppressors', #2. 'Enabling replicative immortality', #3. 'Genome instability and mutation', #4. 'Resisting cell death', #5. 'Deregulating cellular energetics', #6. 'Sustaining proliferative signaling', #7. 'Avoiding immune destruction', #8. 'Tumor promoting inflammation', #9. 'Activating invasion and metastasis' and #10. 'Inducing angiogenesis'. The first six hallmarks could arguably be tested in a relatively simple *in vitro* assay, but hallmarks #7-#10 require interaction with a tumor micro-environment to test. The effect of many drugs on suppressing these features could be highly context dependent (i.e., based on cell-of-origin, combination of driver events and tumor micro-environment), and not only work on the tumor cells, but could change the gene expression of other cell types in the tumor micro-environment

as well. This would be an extension of the TSR hypothesis: instead of reversing the genes of pathways inside tumor cells which are directly dysregulated by driver events, the goal would be to reverse the gene expression of non-tumor cells which are indirectly deregulated by the driver events.

For example, a direct downstream consequence of VHL inactivation in ccRCC is secretion of Vascular Endothelial Growth Factor (VEGF) into the tumor micro-environment. VEGF stimulates the growth of blood vessels (angiogenesis) throughout the tumor.[19] During treatment with drugs which target this mechanism (e.g., VEGF receptor inhibitors), the tumor cells likely start to depend more on other mechanisms to stimulate angiogenesis, which could be measured by comparing the changes in the gene expression of vascular endothelial cells using single cell RNA-seq analysis. A drug which could reverse these observed gene expression changes in the endothelial blood vessel cells might prove therapeutically useful as an add-on to the original drug treatment or as a follow-up treatment.

Future research

The potential of TSR for the drug repositioning of new anti-cancer is still alluring. If anyone can successfully refine TSR so that it becomes able to predict with reasonable accuracy which drugs are going to be effective in treating which tumors, it could literally save billions of dollars in R&D. However, future research should take into account the aforementioned 3 main issues with its current use.

The recommendation would be to use single cell RNA-seq data instead of bulk RNA-seq data of clinical samples. This should be combined with single cell DNA-seq so that it is known which driver and passenger events each tumor cell has acquired. This data should become the input into a model which disentangles the effect of driver events from passenger events on the gene expression of tumor cells and other cell types in the tumor micro-environment. Most likely, new statistical and bioinformatics models are required. The results of such models (i.e., which genes in which cell types are affected by each driver event) should be validated in the appropriate experimental model(s) using single cell RNA-seq and the internal validity of the model should be confirmed by proving the tumor cells with the driver event indeed have a selective advantage over the tumor cells without the driver event. Perhaps this final step of validating the findings in an experimental model is not necessarily needed, when it has been validated that the statistical model already provides reliably enough driver event gene expression signatures.

After the successful completion of all steps, the researcher should have a highly reliable gene expression signature of the impact of the driver event under investigation on the tumor cells and other cell types in the tumor micro-environment. If TSR is to be of any use in predicting new drug repositioning

candidates with any reliability, the gene signature of the driver event should strongly inversely match with the drug gene expression signature of existing targeted therapies intended to treat that driver event (e.g., the gene expression signature of an activating EGFR mutation should be strongly reversed by the drug gene expression signatures of clinically used EGFR inhibitors). Ideally, the drug gene expression signatures should be determined by testing the targeted therapies on the same experimental models with and without the driver event in question, to determine how sensitive the determination of the drug gene expression signature is to the existence of the driver event in question. If the drug gene expression signatures of clinically used targeted drugs do not show strong reversal of the driver event gene expression signatures of the driver events they were designed to target, then TSR would not be expected to be of any use for drug repositioning in cancer treatment.

One of the ways TSR could be enhanced is through the development of new computational methods. In **chapter 6** we developed a new method to normalize the gene expression of RNA-seq tumor samples. It could perhaps be used in the single cell RNA-seq analysis of tumor cells to deconvolute the effect of driver events from passenger events. It highlights the need to understand and model the many influencing factors on gene expression levels, including the measurement process. Unfortunately, to date, measuring single cell RNA-seq with all other additional characterizations needed for this process are cost-prohibitive to perform in the sample sizes required to deconvolute the effect of all the driver events. Hopefully in the nearby future datasets similar in size (i.e., 10,000+ tumor samples) to The Cancer Genome Atlas (TCGA) will be recreated using single cell analysis methods and become similarly freely available, so that the driver event gene expression signatures can be determined at no additional cost. Also, improved statistical models can contribute to reduce the required sample sizes. For example, in **chapter 7** we explored whether convolutional neural networks (CNNs) can be applied to gene expression data.

The simplicity of TSR must mean that on some abstraction level, it must be true. However, this does not mean it can be used to make reliable predictions with all the additional preconditions we have described here. For example, it is possible that in tumor cells from clinical samples, the gene expression changes affected by a specific driver event may not be active at all times and therefore not observable in clinical samples. Regardless, with so many different driver genes and the ability to quantify the effects on the tumor micro-environment with single cell analysis methods, it does seem very likely that some of the interactions which can be measured in clinical tumor tissue samples can potentially be disrupted using some of the drugs already on the market. For this reason, we believe that further investigation is warranted, when it will become cost-effective to do so. Validating TSR at this new level of resolution would then become achievable for single research institutes or a small research consortium.

References

1. Bhat-Nakshatri P, Goswami CP, Badve S, Sledge GW, Nakshatri H. Identification of FDA-approved drugs targeting breast cancer stem cells along with biomarkers of sensitivity. Sci Rep. 2013;3:2530.
2. van Noort V, Schölch S, Iskar M, Zeller G, Ostertag K, Schweitzer C, et al. Novel drug candidates for the treatment of metastatic colorectal cancer through global inverse gene-expression profiling. Cancer Res. 2014;74(20):5690-9.
3. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011;3(96):96ra77.
4. Chen HR, Sherr DH, Hu Z, DeLisi C. A network based approach to drug repositioning identifies plausible candidates for breast cancer and prostate cancer. BMC Med Genomics. 2016;9(1):51.
5. Chen B, Wei W, Ma L, Yang B, Gill RM, Chua MS, et al. Computational Discovery of Niclosamide Ethanolamine, a Repurposed Drug Candidate That Reduces Growth of Hepatocellular Carcinoma Cells In Vitro and in Mice by Inhibiting Cell Division Cycle 37 Signaling. Gastroenterology. 2017;152(8):2022-36.
6. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat Commun. 2017;8:16022.
7. Vásquez-Bochm LX, Velázquez-Paniagua M, Castro-Vázquez SS, Guerrero-Rodríguez SL, Mondragon-Peralta A, De La Fuente-Granada M, et al. Transcriptome-based identification of lovastatin as a breast cancer stem cell-targeting drug. Pharmacol Rep. 2019;71(3):535-44.
8. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
10. Szalai B, Subramanian V, Holland CH, Alföldi R, Puskás LG, Saez-Rodriguez J. Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. Nucleic Acids Res. 2019;47(19):10010-26.
11. Frew IJ, Moch H. A clearer view of the molecular complexity of clear cell renal cell carcinoma. Annu Rev Pathol. 2015;10:263-89.
12. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018;173(2):291-304.e6.
13. Haider S, Tyekucheva S, Prandi D, Fox NS, Ahn J, Xu AW, et al. Systematic Assessment of Tumor Purity and Its Clinical Implications. JCO Precis Oncol. 2020;4.
14. Li X, Wang CY. From bulk, single-cell to spatial RNA sequencing. Int J Oral Sci. 2021;13(1):36.
15. Chatterjee A, Rodger EJ, Eccles MR. Epigenetic drivers of tumourigenesis and cancer metastasis. Semin Cancer Biol. 2018;51:149-59.
16. Network CGAR. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43-9.
17. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018;173(3):611-23.e17.
18. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018;173(2):371-85.e18.
19. Djordjevic G, Mozetic V, Mozetic DV, Licul V, Ilijas KM, Mustac E, et al. Prognostic significance of vascular endothelial growth factor expression in clear cell renal cell carcinoma. Pathol Res Pract. 2007;203(2):99-106.