Perturbation-response genes reveal signaling footprints in cancer gene expression

Michael Schubert[1], Bertram Klinger[2,3], Martina Klünemann[2,3], Mathew J Garnett[4], Nils Blüthgen[2,3], Julio Saez-Rodriguez[1,5,*]

[1] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge CB10 1SD, UK
[2] Institute of Pathology, Charité Universitätsmedizin Berlin
[3] IRI Life Sciences and Institute for Theoretical Biology, Humboldt University Berlin, Germany
[4] Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK
[5] RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine, Aachen 52057, Germany
[*] To whom correspondence should be addressed: saezrodriguez@gmail.com

Abstract

Numerous pathway methods have been developed to quantify the signaling state of a cell from gene expression data, usually from the abundance of transcripts of pathway members, and are hence unable to take into account post-translational control of signal transduction. Gene expression signatures of pathway perturbations can capture this, but they are closely tied to the experimental conditions that they were derived from. We overcome both limitations by leveraging a large compendium of publicly available perturbation experiments to define consensus signatures for pathway activity. We find that although individual expression signatures are heterogeneous, there is a common core of responsive genes that describe pathway activation in a wide range of conditions. These signaling footprints better recover pathway activity than existing methods and provide more meaningful associations with (i) known driver mutations in primary tumors, (ii) drug response in cell lines, and (iii) survival in cancer patients, making them more suitable to assess the activity status of signaling pathways.

Introduction

With omics technologies becoming more and more commonplace, a wealth of molecular data has become available to describe a cell's state in different diseases.The remaining challenge is how to derive predictive and reliable biomarkers for disease status, treatment opportunities, or patient outcome in a way that is both relevant and interpretable. Of particular interest are methods which infer and quantify deregulation of signaling pathways, as those are key for many processes underpinning different diseases. A well-known example of this is cancer, which is largely caused by cell signaling aberrations created by driver mutations and copy number variations [1].

As direct measurements of signaling activity are not widely available, significant effort has been devoted to extract specific, functionally relevant readouts from gene expression. Pathway-level activity has mostly been inferred using the expression level of pre-defined gene sets derived from Gene Set Enrichment Analysis [2] and more sophisticated methods that attempt to quantify the signal flow, such as SPIA [3], PARADIGM [4], or Pathifier [5]. All of these methods [6], however, are based on mapping the mRNA expression on the corresponding signaling proteins, and hence do not take into account the effect of post-translational modifications that are known to govern mammalian signal transduction (Fig. 1a). It is therefore unclear if the pathway scores obtained by those methods correspond to signaling activity.

Another approach is to look at the downstream effect of pathway activity on gene expression. Expression levels of genes regulated by transcription factor or kinases can be used to estimate the activity status of proteins [7,8]. Similarly, the transcripts altered when perturbing a specific pathway can be used to infer pathway activation from gene expression of other samples, i.e. provide a signature of its activity. Such signatures have been derived for breast cancer [9,10], but they are known to be heterogeneous and not replicate well under different experimental conditions [11].

To address both the disregard of post-translational control by pathway methods and the context-specificity of gene expression signatures, we collected a large body of publicly available perturbation experiments to quantify gene expression responses to a specified set of stimuli. We then used those to infer the upstream signal mediating downstream expression changes based on consensus signatures as we have previously suggested [12]. Using these signatures of pathway-responsive genes (PRGs), we performed a systematic comparison of this approach and other commonly used pathway methods in terms of how well they are able to recover constitutive activity mediated by driver mutations in The Cancer Genome Atlas (TCGA) [13]. We also quantified how well they can explain drug sensitivity to 265 drugs in 1,001 cancer cell lines in the Genomics of Drug Sensitivity in Cancer (GDSC) panel [14,15] and patient survival in 7254 primary tumors spanning 34 tumor types using TCGA data. We found that our perturbation-based signatures outperform existing methods for all these tasks.

Results

Consensus gene signatures for pathway activity

We curated (workflow in Fig. 1b) a total of 208 different submissions to ArrayExpress/GEO, spanning perturbations of a total of 11 pathways, 580 experiments and 2652 microarrays (Supplementary Fig. 1). We obtained z-scores of gene expression changes for each experiment, for which we performed a multiple linear regression using the perturbed pathway as input and gene expression as a response variable, taking into account known pathway crosstalk (Fig. 1c). This provided us consensus pathway activity signatures across a range of different conditions irrespective of the expression state of pathway members. For each pathway, we find responsive genes that are highly significant (Supplementary Fig. 2) and do barely overlap between pathways (Supplementary Fig. 3) or known pathway members in Reactome [16] or Gene Ontology [17] (Supplementary Fig. 4).

For the 100 most significant response genes per pathway, we use the coefficients in the above regression to infer pathway activity from gene expression (Supplementary Table 9). We find that this model is able to cluster input experiments by their activation pattern, which individual signatures are not (Fig. 2a). Within experiments, the inferred pathway activation is strongly ($p<10^{-10}$) associated with the pathway that was experimentally perturbed. The associations with other pathways are weaker ($p>10^{-5}$) except for EGFR with MAPK/PI3K and TNFa with NFkB/MAPK, where there is biologically known cross-activation (Fig. 2b)[18]. Relative activation patterns are consistent across input experiments and not driven by outliers (Fig. 2c), suggesting that those genes represent a common core of pathway responses across a wide range of conditions. This is in contrast to methods based on pathway expression that are not able to recover experimental perturbations by means of their inferred pathway score (Supplementary Fig. 5), unless the pathway has been specifically curated to capture the response to a stimulus. Across experiments, PRGs are able to better rank experiments according to the pathway perturbed for 10 out of 11 pathways (Fig. 2d, Supplementary Table 5). For the majority of pathways, mapping methods do not perform significantly better than random, with the exception of NFkB and JAK-STAT pathways. VEGF is not recapitulated well by any method, possibly because of overlap with other pathways. Taken together, these results suggest that PRGs more closely correspond to pathway activation upon perturbation than any method that maps mRNA expression to pathway members.

Knowing how pathway-responsive genes behave when a stimulus is present, we can take the idea one step further and hypothesize that if expression changes characterize pathway perturbation, the existence of a different basal expression level of the responsive genes may in turn correspond to intrinsic signaling activity. In contrast to the other methods (Supplementary Fig. 6), we find that for our method the correlation between different pathway scores in basal expression (Fig. 2e) corresponds to the previously observed cross-activation upon perturbation (Fig. 2b). This indicates that footprints of signaling activity

are present and detectable in basal gene expression. Furthermore, they are robust to changes in the experiments the model was derived from (Fig. 2f).

Recovering mechanisms of known driver mutations

If our reasoning is correct and pathway-response signatures in basal gene expression correspond to intrinsic signaling activity, we should be able to see a higher pathway score in cancer patients with an activating driver mutation in that pathway and a lower score for pathway suppression compared to patients where no such alteration is present. Depending on how those aberrations are spread across cancer types, we should be able to detect them within or across cancer types.

We selected all cancer types in the TCGA for which there were tissue-matched normals available, in order to make full use of the pathway methods that require them. We calculated pathway scores for those using PRGs, Reactome and Gene Ontology enrichment, SPIA, Pathifier, and PARADIGM. We used an ANOVA to calculate significant associations between the presence and absence of mutations and copy number alterations and the inferred pathway scores, both with and without regressing out cancer types (Fig. 3a and Supplementary Table 6).

In terms of proliferative signaling, we find that *EGFR* amplifications are correlated both with EGFR- and MAPK-responsive genes (FDR<10^{-20}), and to a lesser extent PI3K, VEGF, and Hypoxia (FDR<10^{-9}). *ERBB2* amplifications show an increase in EGFR and PI3K-responsive genes, but also a reduction in the Trail signature (FDR<0.05), suggesting a stronger relative impact on cell survival. *KRAS* mutations show an increase in inferred EGFR activity, and amplifications additionally for MAPK and PI3K (FDR<10^{-5}). *BRAF* mutations have a positive effect on EGFR and MAPK (FDR<10^{-9}) but not PI3K (FDR>0.4).

For *TP53* mutations we find a significant reduction in p53/DNA damage response activity (FDR<10^{-18}) and activation of the pathways for MAPK, PI3K, and Hypoxia (FDR<10^{-4}). This is in contrast to loss of *TP53*, where we only find a reduction in p53/DDR (FDR<10^{-3}) but no modification of any other pathway (FDR>0.15). The dual nature of *TP53* mutations and loss are in line with the recent discovery that *TP53* mutations can act in an oncogenic manner in addition to disrupting its tumor suppressor activity, which has been shown for individual cancer types [19–22]. In addition, this analysis suggests a link between *TP53* mutations and genes that are induced by activation of canonical oncogenic signaling such as MAPK and PI3K. The associations of p53/DDR activity with *MYC*, *RAD21*, *NDRG1*, and *PABPC1* (Fig. 3a) can also be explained by co-occurring *TP53* modifications (p>0.04 if conditioned).

We find that *VHL* mutations (which have a high overlap with Kidney Renal Carcinoma, KIRC) are associated with an expected stronger induction of hypoxic genes [23] compared to other cancer types. More surprisingly, we find that presence of *PIK3CA/B* amplifications and *PTEN* deletions is also more connected to increasing the hypoxic response (FDR<10^{-6}) compared to an effect on the PI3K-responsive genes (FDR between 10^{-2} and 10^{-5}). A role of PI3K signaling in hypoxia has been shown before [24–26].

None of the above processes are sufficiently recapitulated by other pathway methods (Fig. 3b and Supplementary Fig. 7-8). For *TP53* mutations (Fig 3b.; top left) only PRGs are able to recover the expected negative association between the mutation and p53/DDR activity. GO and Reactome showed a much weaker effect in the same direction, while Pathifier and SPIA showed an incorrect positive effect. For *KRAS* mutations (Fig 3b.; top right) only our pathway signature can detect a strong activation of the MAPK/EGFR pathways where the other methods either show no significant effect or an effect in the wrong direction. The same applies to *EGFR* amplifications (Fig 3b.; bottom left). Across tissues, our method is the only one to recover hypoxia as the strongest link with *VHL* mutations (Fig 3b.; bottom right).

Associations with drug response

The next question we tried to answer is how well pathway-responsive gene signatures are able to explain drug sensitivity in cancer cell lines. We took as a measure of efficacy the IC_{50}, i.e. the drug concentration that reduces viability of cancer cells by 50%, for 265 drugs and 805 cell lines from the GDSC project [15] and performed an ANOVA between those and inferred pathway scores of PRGs, Reactome, Gene Ontology, SPIA, Pathifier, and PARADIGM.

We found that there were 199 significant associations (10% FDR in Fig. 4a) for PRGs, dominated by sensitivity associations between MAPK/EGFR activity and drugs targeting MAPK pathway (Fig. 4b) that are consistent with oncogene addiction. The strongest hit we obtained was the association between Nutlin-3a and p53-responsive genes. Nutlin-3a is an MDM2-inhibitor that in turn stabilizes p53, and it has also previously been shown that a mutation in *TP53* is strongly associated with increased resistance to Nutlin-3a [14]. This is thus a well-understood mechanism of sensitivity (presence) or resistance (absence of p53 activity) to this drug that our method recovers but none of the pathway expression-based methods do. We also find strong associations of MAPK/EGFR activation with different MEK inhibitors (Trametinib, RDEA119, CI-1040, etc.), a Raf inhibitor (AZ628) and a TAK1 inhibitor (7-Oxozeaenol).

The other pathway methods (Supplementary Fig. 9) showed a much lower number of associations across the range of significance (Fig. 4c). Using the same significance threshold (10% FDR) mutated driver genes only yield 136 associations. Other methods only provide stronger associations for *TP53*, where the signature is a compound of p53 signaling and DNA damage response, and PLX4720/Dabrafenib, drugs that were specifically designed to target mutated *BRAF*. For 170 out of 265 drugs covered by significant associations with either PRGs or mutations, the PRGs provided stronger associations for 85. They also show significant enrichment in cytotoxic drugs compared to targeted drugs for mutations (Fisher's exact test, p<0.002).

However, stratification using PRGs and mutated driver genes is not mutually exclusive. Our pathway scores are able to further stratify the mutated and wild-type sub-populations into more and less sensitive cell lines (Fig. 4d and Supplementary Table 9). This includes, but is

not limited to, *BRAF*, *NRAS* or *KRAS* mutations using MAPK pathway activity and the MEK inhibitor Trametinib (Fig. 4d; top left) or Raf inhibitor AZ628 (Fig. 4d; bottom left), *BRAF* mutations with Dabrafenib (Fig. 4d; top right), and *TP53* mutations with p53/DDR and Nutlin-3a (Fig. 4d; bottom left). For MAPK- and *BRAF*-mutated cell lines, we find that cell lines with an active MAPK pathway according to the PRGs are 175 (AZ628), 7596 (Trametinib), or 10^5 fold (Dabrafenib) more sensitive than those where it is inactive. For Trametinib, cell lines with active MAPK but no mutation in *BRAF*, *KRAS*, or *NRAS* are six times more sensitive than cell lines that harbor a mutation in any of them but MAPK is inactive.

Taken together, these results indicate that our pathway scores can be used to complement mutation-derived biomarkers by either refining them or providing an alternative where no such marker exists.

Implications for patient survival

The implications of inferred pathway activity compared to pathway expression is expected to be less clear for patient survival than for cell line drug response due to more heterogeneity and many more confounding factors involved that affect the phenotype observed. Nonetheless, we were interested in how our inferred pathway activity compared to the pathway expression methods in terms of overall patient survival. We would expect activity of canonical oncogenic pathways to be negatively correlated with patient survival and pro-apoptotic pathways to be positively correlated.

The pathway activity inferred by PRGs showed a strong association with decreased survival for EGFR, MAPK, PI3K, and Hypoxia (Fig. 5a). Gene Ontology found much weaker associations for those pathways, and the other methods missed them almost entirely. In terms of Trail activity, PRGs find an increase in survival while the other methods show either a decrease or no effect. For JAK-STAT, NFkB, p53, and VEGF there are no significant associations that are picked up by more than one method (FDR<0.05).
Compared to pathway scores, driver mutations only showed a significant decrease in survival for *TP53* (FDR<0.03 vs. FDR>0.2) with a weaker effect size.

For individual cancer types, we find a similar separation between oncogenic and tumor-suppressor pathways for the associations using PRGs (Fig. 5b) that other methods fail to provide, and our associations are significant for more cancer types and less correlated between pathways (Supplementary Fig. 10). In addition, we find cancer-specific associations of pathways with no effect in the pan-cancer setting. Adrenocortical Carcinoma (ACC) shows a significant survival increase with p53 activity (FDR<10^{-3}) but none of the reported gain-of-function *TP53* variants [22]. Kidney Renal Clear Cell Carcinoma (KIRC) and Low-Grade Glioma (LGG) show decreased survival with TNFa and the latter with JAK-STAT, pathways where activating mutations are much less well established than for EGFR/MAPK. For these associations, we found a difference in one-year survival of over 25% between the top and bottom quartiles of the assigned pathway scores (Fig. 5c). Compared to mutations,

PRG associations were stronger (FDR 10^{-7} vs. 10^{-3}) and more consistent (strongest associations with small number of mutated genes).

Discussion

The explanation of phenotypes in cancer, such as cell line drug response or patient survival, has largely been focussed on genomic alterations (mutations, copy number alterations, and structural variations). While this approach has generated many important insights into cancer biology, it does not directly make statements about the impact of those aberrations on cellular processes and signal transduction in particular. Pathway methods, mostly used on gene expression, have so far largely fallen short on delivering actionable evidence. This can be due in part to lack of robustness, as suggested by the heterogeneity in responses of individual signatures (Fig. 2a), but arguably also by the fact that extracting features that reflect pathway activity from gene expression is not trivial. Measuring the proteome and in particular its post-translational modifications could in part resolve this, but the amount of available data will lag behind sequencing technologies for the foreseeable future. We thus have a need to address both the need of accurate inference of signaling activity from gene expression and the issue of irreproducible gene signatures.

Here we show progress to overcome these limitations by leveraging a large compendium of pathway-responsive gene signatures derived from a wide range of different conditions. We found that despite the heterogeneity in gene expression experiments, consensus signatures are more strongly correlated with pathway perturbations than other methods. These pathway perturbation signatures translate into generally applicable signaling footprints in basal gene expression, where they can recover the impact of known driver mutations, but also identify cases where a pathway is active without the corresponding mutation. Pathway mapping only recovers known associations where this effect is mediated by expression changes, such as copy number aberrations.

In terms of drug associations, we have shown that (i) signaling footprints outperform commonly used pathway methods, (ii) footprint scores can be used to refine mutation-derived biomarkers, and (iii) they can act as biomarkers themselves when there is no known associated mutation for a given drug. We expect that this method of deriving signatures for pathway activity will also in other contexts be able to provide insight into disease mechanisms as well as treatment opportunities that pathway expression-based methods are not able to. In addition, pathway expression is further removed and thus more likely to be a consequence rather than a cause of the drug sensitivity induced by a signaling aberration. The fact that competing methods do not recover oncogene addiction patterns supports this claim.

For survival associations, only signaling footprints find the pathways that we would most expect to decrease patient survival by accelerating tumour growth (EGFR and MAPK) and promoting survival by apoptosis (Trail) to be associated with the respective outcome in both the pan-cancer and the tissue-specific cohorts. Other methods fail to separate those, only

obtain significant associations for a very limited number of cancer types, and show high correlation between pathways.

Overall, our results suggest that consensus pathway response signatures provide a better measure of pathway activity than pathway expression, irrespective of whether the latter was derived from gene sets or directed paths. We have shown that they are able to refine our understanding of the impact of mutations, as well as their utility for cell line drug response and patient survival. The examples we outlined show that a downstream readout should be used as a proxy for pathway activity instead of mapping mRNA expression levels to signaling molecules.

Methods

Data from The Cancer Genome Atlas (TCGA)

To obtain TCGA data, we used the Firehose tool from the BROAD institute (http://gdac.broadinstitute.org/), release 2016_01_28.

For gene expression, we used all data labelled level 3 RNA-seq v2. We extracted the raw counts from the text files for each gene, discarded those that did not have a valid HGNC symbol, and averaged expression levels where more than one row corresponded to a given gene. We then performed a voom transformation (limma package, BioConductor) for each TCGA study separately, to be able to use linear modeling techniques with the count-based RNA-seq data. The data used corresponds to 34 cancer types and a total of 9737 tumor and 641 matched normal samples.

From clinical data, we extracted the vital status and used known survival time or known time of last follow-up as the survival time for the downstream analyses. We converted the time in days to months by dividing by 30.4. The overlap of TCGA data where we could obtain both mRNA expression levels as well as survival times is 10544 distributed across 33 cancer types. For comparing different pathway methods, we only used cancer types with tissue-matched controls, leaving 5927 samples in 13 cancer types.

Data from the Genomics of Drug Sensitivity in Cancer (GDSC) project

We used version 17a of the GDSC data [15], comprised of molecular data for 1,001 cell lines and 265 anticancer drugs, specifically microarray gene expression data (ArrayExpress accession E-MTAB-3610) and the IC_{50} values for each drug-cell line combination. For computing pan-cancer associations, we used the subset TCGA-like cancer type label, leaving 768 cell lines.

Curation of Perturbation-Response Experiments

Our method is dependent on a sufficiently large number publicly available perturbation experiments that activate or inhibit one of the pathways we were looking at. The following conditions needed to be met in order for us to consider an experiment: (1) the compound or factor used for perturbation was one of our curated list of pathway-perturbing agents (for a list, see table S0); (2) the perturbation lasted for less than 24 hours to capture genes that belong to the primary response; (3) there was raw data available for at least two control arrays and one perturbed array; (4) it was a single-channel array; (5) we could process the arrays using available BioConductor packages; (6) the array was not custom-made so we could use standard annotations.

We curated a list of known pathway activators and inhibitors for 11 pathways, where the interaction between each compound and pathway is well established in literature. We then used those as query terms for public perturbation experiments in the ArrayExpress database [27] and included a total of 223 submissions and 573 experiments in our data set, where each experiment is a distinct comparison between basal and perturbed arrays. If there were multiple time points, different cells, different concentrations, or different perturbing agents within a single database submission, they were considered as different experiments.

Microarray Processing

Started from the curated list of perturbation-induced gene expression experiments, we included all single-channel microarrays with at least duplicates in the basal condition with raw data available that could be processed by either the limma [28], oligo [29], or affy [30] BioConductor packages and for which there was a respective annotation package available. Multiple concentrations or time points in a series of arrays were considered as individual experiments.

We first calculated a probe-level for 573 full series of arrays, where we performed quality control of the raw data using RLE and NUSE cutoffs under 0.1 and kept all arrays below that threshold. If after filtering less than two basal condition arrays remained, the whole experiment was discarded. For the remaining 568 series we normalized using the RMA algorithm and mapped the probe identifiers to HGNC symbols.

Building a Linear Model of Pathway-Response Genes

For each HGNC symbol, we calculated a model based on mean and standard deviation of the gene expression level, and computed the z-score as average number of standard deviations that the expression level in the perturbed array was shifted from the basal arrays. We then performed LOESS smoothing for all z-scores in a given experiment using our null model, as described previously [12].

From the z-scores of all experiments and all pathways, we performed a multiple linear regression with the pathway as input and the z-scores as response variable for each gene separately:

$$Z_g \sim M \ ... \ \forall g$$

Where Z_g is the z-score for a given gene g across all input experiments (as a column vector of experiments). M is a coefficients matrix (rows are experiments, columns pathways, Fig. 1b) that has the coefficient 1 if the the experiment had a pathway activated, -1 if inhibited, and 0 otherwise. For instance, the Hypoxia pathway had experiments with low oxygen conditions set as 1, HIF1A knockdown as -1, and all other experiments as 0. The same is true for EGFR and EGF treatment vs. EGFR inhibitors respectively, with the additional coefficients of MAPK and PI3K pathways set to 1 because of known cross-talk (for a full structure of the cross-talk modeled, see Fig. 1c). As these are fold changes, we do not allow an intercept.

From the result of the linear model, we selected the top 100 genes per pathway according to their p-value and took their estimate (the fitted z-scores) as coefficient. We set all other gene coefficients to 0, so this yielded a matrix with HGNC symbols in rows and pathways in columns, where each pathway had 100 non-zero gene coefficients (Supplementary Table 9).

Signaling Footprint scores

Each column in the matrix of perturbation-response genes corresponds to a plane in gene expression space, in which each cell line or tumor sample is located. If you follow its normal vector from the origin, the distance it spans corresponds to the pathway score P each sample is assigned (matrix of samples in rows, pathways in columns). In practice, this is achieved by a simple matrix multiplication between the gene expression matrix (samples in rows, genes in columns, values are expression levels) and the model matrix (genes in rows, pathways in columns, values are our top 100 coefficients):

$$P = E * G$$

We then scaled each pathway or gene set score to have a mean of zero and standard deviation of one, in order to factor out the difference in strength of gene expression signatures and thus be able to compare the relative scores across pathways and samples at the same time.

Pathway and Gene Ontology scores

We matched our defined set of pathways to the publicly available pathway databases Reactome [16] and KEGG [31], and Gene Ontology (GO) [17] categories (Supplementary Tables 1-2), to obtain a uniform set across pathway resources that makes them comparable. We calculated pathway scores as Gene Set Variation Analysis (GSVA) scores that are able to assign a score to each individual sample (unlike GSEA that compares groups).

SPIA scores

Signaling Pathway Impact Analysis (SPIA) [3] is a method that utilizes the directionality and signs in a KEGG pathway graph to determine if in a given pathway structure the available species are more or less available to transduce a signal. As the species considered for a pathway are usually mRNAs of genes, this method infers signaling activity by the proxy of gene expression. In order to do this, SPIA scores require the comparison with a normal condition in order to compute both their scores and their significance.

We used the SPIA Bioconductor package [3] in our analyses, focussing on a subset of pathways (Supplementary Table 3). We calculated our scores either for each cell line compared to the rest of a given tissue where no normals are available (i.e. for the GDSC and drug response data) or compared to the tissue-matched normals (for the TCGA data used in driver and survival associations).

Pathifier scores

As Pathifier [5] requires the comparison with a baseline condition in order to compute scores, we computed the GDSC/TCGA scores as with SPIA. As gene sets, we selected Reactome pathways that corresponded to our set of pathways (Supplementary Table 2), where Pathifier calculated the "signal flow" from the baseline and compared it to each sample.

PARADIGM scores

We used the PARADIGM software from the public software repository (https://github.com/sbenz/Paradigm) and a model of the cell signaling network [32] from the corresponding TCGA publication (https://tcga-data.nci.nih.gov/docs/publications/coadread_2012/). We normalized our gene expression data from both GDSC and TCGA using ranks to assign equally spaced values between 0 and 1 for each sample within a given tissue. We then ran PARADIGM inference using the same options as in the above publication for each sample separately. We used nodes in the network representing pathway activity to our set of pathways (Supplementary Table 4) to obtain pathway scores that are comparable to the other methods, averaging scores where there were more than one for a given sample and node.

Associations with known driver mutations and CNAs

For comparing the impact of mutations across different pathway methods, we used TCGA cohorts where tissue-matched controls were available, leaving 6549 samples across 13 cancer types. For mutated genes, we considered all genes that had a change of coding sequence (SNP, small indels in MAF files) as mutated and all others as not mutated. For copy number alterations (CNAs), we used the thresholded GISTIC [33] scores, where we considered homozygous deletions (-2) and strong amplifications (2) as altered, no change (0) as basal and discarded intermediate values (-1, 1) from our analysis. We focussed our analysis of the mutations and copy number alterations on the subset of 464 driver genes that were also used in the GDSC. We used the sets of mutations and CNAs to compute the

linear associations between samples for all different methods we looked at. We did not regress out the cancer type in order to keep associations where mutations/CNAs are highly correlated with it, but highlighted all associations that passed the significance threshold of FDR<5% (for each pathway method individually) after such a correction.

Drug associations using GDSC cell lines

We performed drug association using an ANOVA between 265 drug IC_{50}s and 11 inferred pathway scores conditioned on MSI status, doing a total of 2915 comparisons for which we correct the p-values using the false discovery rate. For pan-cancer associations, we used the cancer type as a covariate in order to discard the effect that different tissues have on the observed drug response. While this will also remove genuine differences in pathway activation between different cancer types, we would not be able to distinguish between those and other confounders that impact the sensitivity of a certain cell line from a given tissue to a drug. Our pan-cancer association are thus the same of intra-tissue differences in drug response explained by inferred (our method, GO, or Reactome) pathway scores.

We also selected two of our strongest associations to investigate whether they provide additional information of what is known by mutation data. For two MEK inhibitors, we show the difference between wild-type and mutant MAPK pathway (defined as a mutation in either *MAP2K1*, *NRAS*, *KRAS*, or *BRAF*) with a discretized pathway score (upper and lower quartile vs. the rest), as well as the combination between the upper quartile of tissue-specific pathway scores and presence of a MAPK mutation.

Survival associations using TCGA data

Starting from the pathway scores derived with GO/Reactome GSEA, SPIA, Pathifier, PARADIGM, and our method on the TCGA data as described above, we used Cox Proportional Hazard model (R package survival) to calculate survival associations for pan-cancer and each tissue-specific cohort. For the pan-cancer cohort, we regressed out the effect of the study and age of the patient, and fitted the more for each pathway and method used. For the tissue-specific cohorts, we regressed out the age of the patients. We adjusted the p-values using the FDR method for each method and for each method and study separately. We selected a significance threshold of 5 and 10% for the pan-cancer and cancer-specific associations for which we show a matrix plot and a volcano plot of associations, respectively.

In order to get distinct classes needed for interpretable Kaplan-Meier survival curves (Fig. 4c), we split all obtained pathway scores in upper, the two middle, and lower quartile and respectively to show for the three examples of associations found.

Acknowledgements

Author contributions

MS designed research, performed all analyses, and wrote the manuscript. BK, MK, NB and MJG supported result interpretation. JSR supervised the project and contributed to writing the manuscript. All authors read and approved the final version of the manuscript.

References

1. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100,** 57–70 (2000).

2. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).

3. Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25,** 75–82 (2008).

4. Vaske, C. J. *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26,** i237–i245 (2010).

5. Drier, Y., Sheffer, M. & Domany, E. Pathway-based personalized analysis of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 6388–6393 (2013).

6. Mutation Consequences and Pathway Analysis working group of the International Cancer Genome Consortium. Pathway and network analysis of cancer genomes. *Nat. Methods* **12,** 615–621 (2015).

7. Chen, E. Y. *et al.* Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. *Bioinformatics* **28,** 105–111 (2011).

8. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* (2016). doi:10.1038/ng.3593

9. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439,** 353–357 (2005).

10. Gatza, M. L. *et al.* A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6994–6999 (2010).

11. Chibon, F. Cancer gene expression signatures – The rise and fall? *Eur. J. Cancer* **49,** 2000–2009 (2013).

12. Parikh, J. R., Klinger, B., Xia, Y., Marto, J. A. & Blüthgen, N. Discovering causal signaling pathways through gene-expression patterns. *Nucleic Acids Res.* **38,** W109–17 (2010).

13. The Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45,** 1113–1120 (2013).

14. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483,** 570–575 (2012).

15. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* (2016). doi:10.1016/j.cell.2016.06.017

16. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691–7 (2011).

17. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32,** D258–D261 (2004).

18. Kant, S. *et al.* TNF-stimulated MAP kinase activation mediated by a Rho family GTPase signaling pathway. *Genes Dev.* **25,** 2069–2078 (2011).

19. Olive, K. P. *et al.* Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119,** 847–860 (2004).

20. Zhang, C. *et al.* Tumour-associated mutant p53 drives the Warburg effect. *Nat. Commun.* **4,** 2935 (2013).

21. Weissmueller, S. *et al.* Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor β signaling. *Cell* **157,** 382–394 (2014).

22. Zhu, J. *et al.* Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer

growth. *Nature* **525,** 206–211 (2015).

23. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399,** 271–275 (1999).

24. Zhou, J., Schmid, T., Frank, R. & Brüne, B. PI3K/Akt Is Required for Heat Shock Proteins to Protect Hypoxia-inducible Factor 1α from pVHL-independent Degradation. *J. Biol. Chem.* **279,** 13506–13513 (2004).

25. Yang, X.-M. *et al.* Role of PI3K/Akt and MEK/ERK in mediating hypoxia-induced expression of HIF-1alpha and VEGF in laser-induced rat choroidal neovascularization. *Invest. Ophthalmol. Vis. Sci.* **50,** 1873–1879 (2009).

26. Kilic-Eren, M., Boylu, T. & Tabor, V. Targeting PI3K/Akt represses Hypoxia inducible factor-1α activation and sensitizes Rhabdomyosarcoma and Ewing's sarcoma cells for apoptosis. *Cancer Cell Int.* **13,** 1–8 (2013).

27. Parkinson, H. *et al.* ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* **35,** D747–D750 (2007).

28. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (Springer New York, 2005).

29. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26,** 2363–2367 (2010).

30. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20,** 307–315 (2004).

31. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

32. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).

33. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. U. S. A.* **104,**

20007–20012 (2007).

Figures
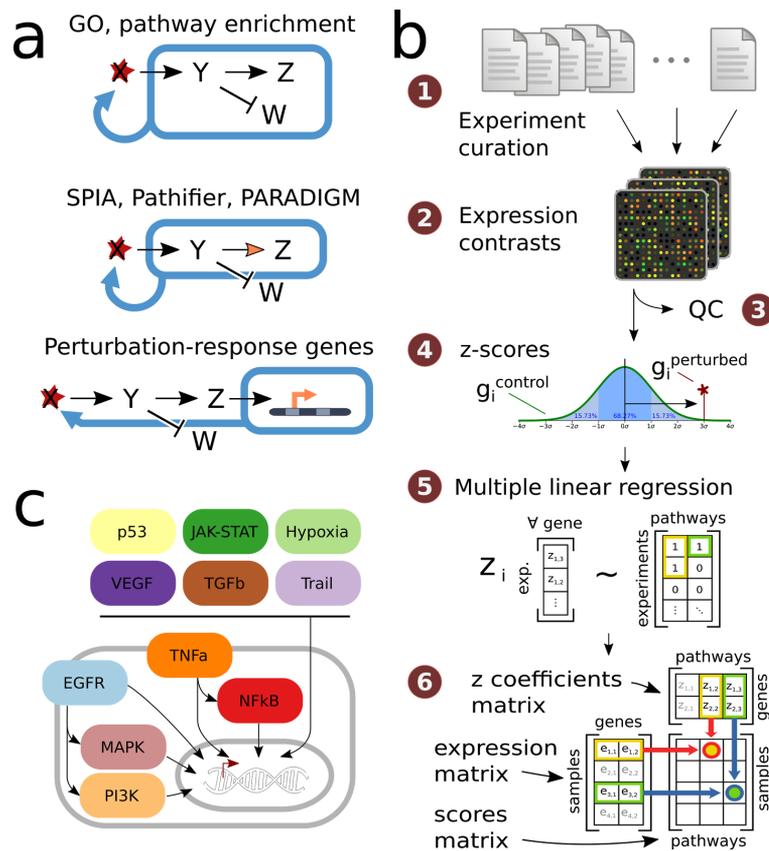
1. Deriving pathway-response signatures for 11 pathways

A. Reasoning about pathway activation. Most pathway approaches make use of either the set (top panel) or network (middle panel) of signaling molecules to make statements about a possible activation, while our approach considered the genes affected by perturbing them.

B. Workflow of data curation and model building. (1) Finding and curation of 208 publicly available experiment series in the ArrayExpress database, (2) Extracting 556 perturbation experiments from series' raw data, (3) Performing QC metrics and discarding failures, (4) Computing z-scores per experiment, (5) Using a multiple linear regression to fit genes responsive to all pathways simultaneously obtaining the z-coefficients matrix, (6) Assigning pathway scores using the coefficients matrix and basal expression data. See methods section for details. Image credit Supplementary Note 1.

C. Structure of the perturbation-response model. For the multiple linear regression, we set the coefficients or perturbed pathways to 1 if a pathway was perturbed, 0 otherwise. In addition, EGFR perturbation also had MAPK and PI3K coefficients set, and TNFa had NFkB set.
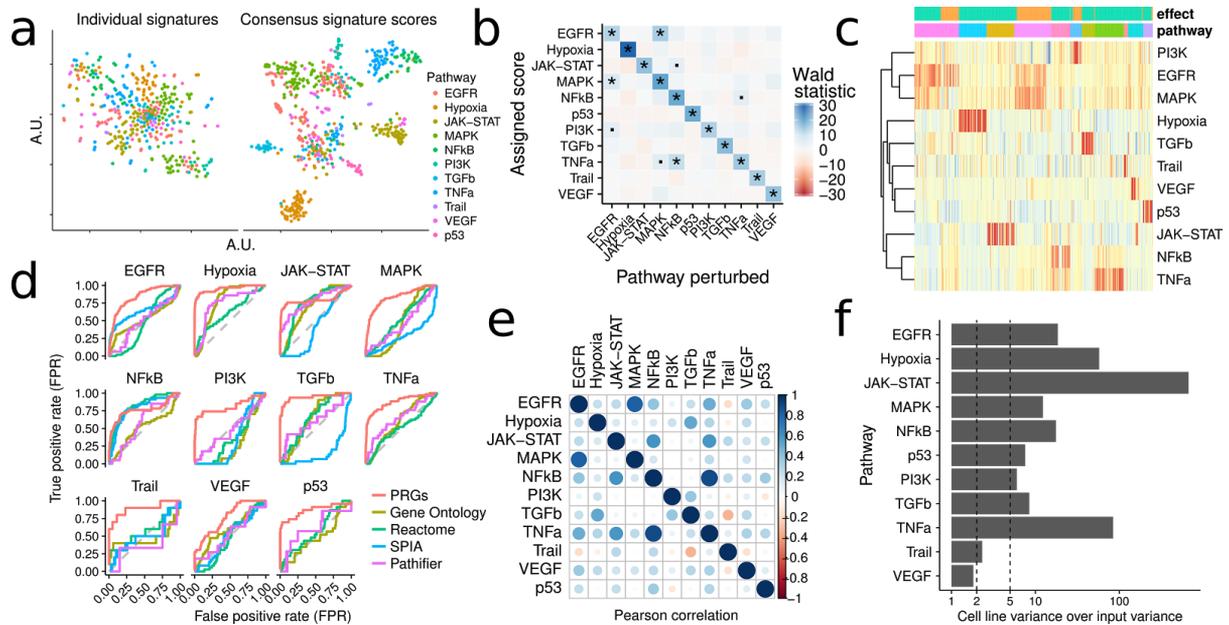
2. Evaluation of pathway-response signatures

A. T-SNE plots for separation of perturbation experiments with different pathway perturbations in different colors. Fold changes of genes in individual perturbation experiments (10% FDR) do not cluster by pathway (left). Using a consensus signature of genes whose z-score is most consistently deregulated for each pathway instead, we can observe distinct clusters of perturbed pathways (right). Details Supplementary Note 2.

B. Associations between perturbed pathways and the scores obtained by the model of pathway-responsive genes (PRGs). Along the diagonal each pathway is strongly ($p<10^{-10}$) associated with its own perturbation. Significant off-diagonal elements are sparse and only occur ($p<10^{-5}$) where there is biologically known cross-activation.

C. Heatmap of relative pathway scores in each perturbation experiment. 523 experiments in columns, annotated with the perturbation effect (green for activation, orange for inhibition) and pathway perturbed (same order as b). Pathway scores in rows cluster between EGFR/MAPK and to a lesser extent PI3K, and TNFa/NFkB. Color indicates activation or inhibition strength.

D. ROC curves for different methods ranking perturbation experiments by their pathway score. PRGs show better performance for all pathways except JAK-STAT and NFkB, where other methods are equal. Gene Ontology and Reactome scores obtained by Gene Set Variation Analysis (GSVA). Pathifier using Reactome gene sets.

E. Correlation of pathway scores in basal gene expression of cell lines in the GDSC panel. Positive correlation in blue, negative in red. Circle size and shade correspond to correlation strength. Pathways that showed cross-activation in point b are more highly correlated in basal expression as well.

F. Stability of basal pathway scores when bootstrapping input experiments. The variance of pathway scores in cell lines given bootstraps more than five times as high compared to the variance of bootstraps given cell lines for all pathways except two (Trail and VEGF), where it is roughly twice as high.
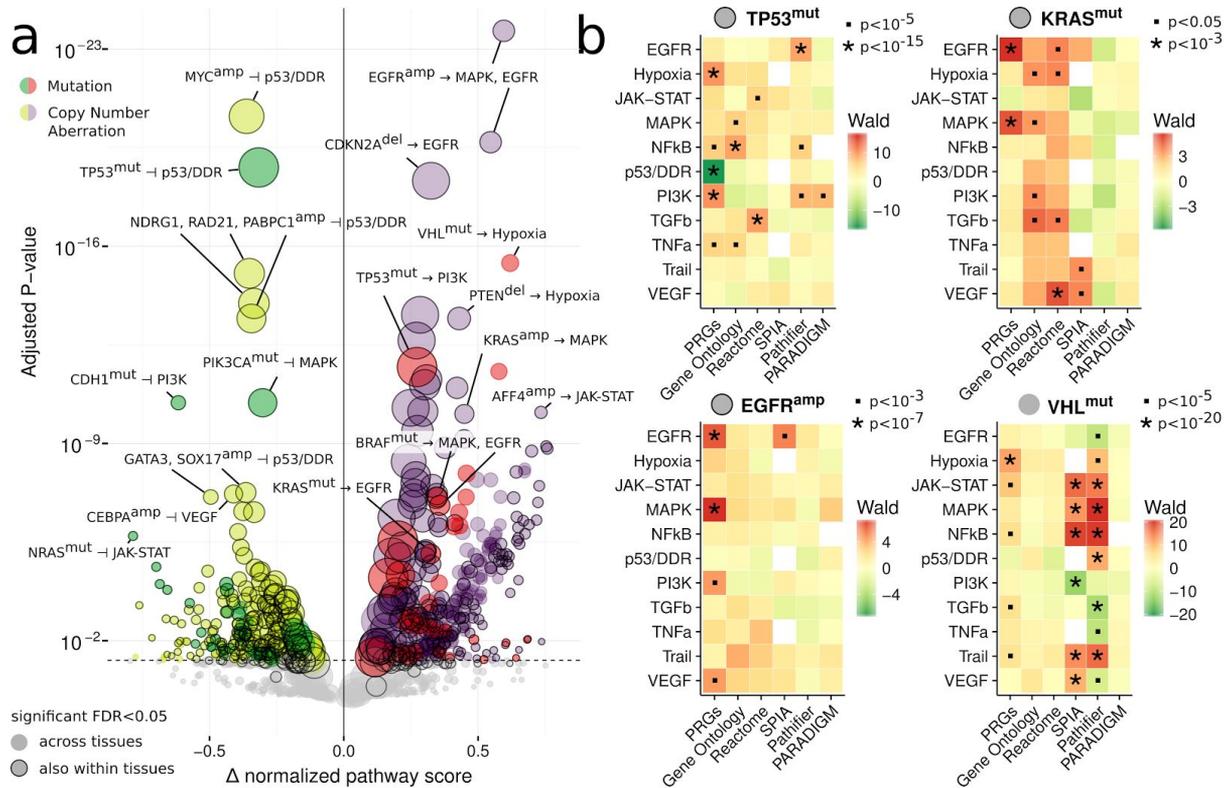
3. Ability of pathway methods to recover well-known mutations

A. Volcano plot of pan-cancer associations between driver mutations and copy number aberrations with differences in pathway score. Pathway scores calculated from basal gene expression in the TCGA for primary tumors. Size of points corresponds to occurrence of aberration. Type of aberration is indicated by superscript "mut" if mutated and "amp"/"del" if amplified or deleted, with colors as indicated. Effect sizes on the horizontal axis larger than zero indicate pathway activation, smaller than zero inferred inhibition. P-values on the vertical axis FDR-adjusted with a significance threshold of 5%. Associations shown without correcting for different cancer types. Associations with a black outer ring are also significant if corrected.

B. Comparison of pathway scores (vertical axes) across different methods (horizontal axes) for *TP53* and *KRAS* mutations, *EGFR* amplifications and *VHL* mutations. Wald statistic shown as shades of green for downregulated and red for upregulated pathways. P-value labels shown as indicated. White squares where a pathway was not available for a method.
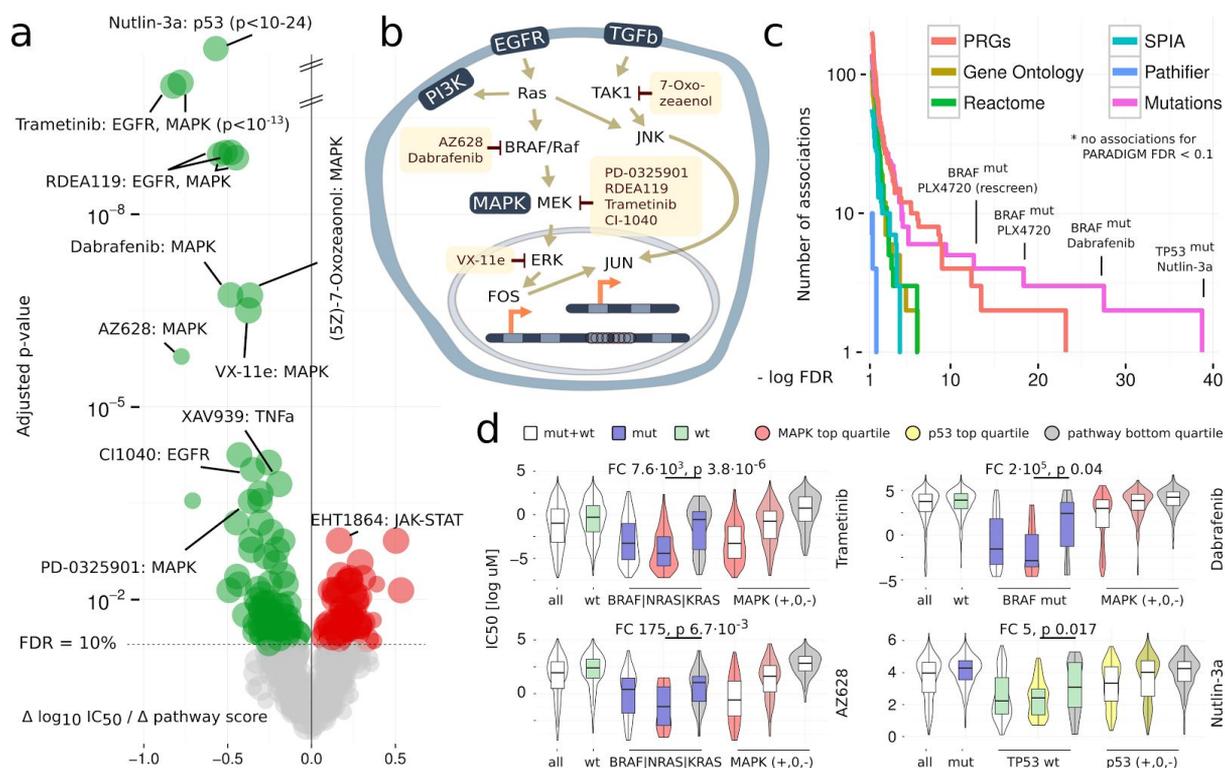
4. MAPK and p53 scores drive drug response across all cancer types

A. Volcano plot of pan-cancer associations between PRG pathway scores and drug response ($\log_{10}$ IC$_{50}$). Pathway scores computed using basal gene expression in GDSC cell lines. Associations corrected for cancer type. Size of points corresponds to number of cell lines screened with a particular drug. Effect size corresponds to 10-fold change in IC$_{50}$ per standard deviation of the pathway score. Values smaller than zero indicate sensitivity markers (green) and greater than zero resistance markers (red). P-values FDR-corrected.

B. Pathway context of the strongest associations between EGFR/MAPK pathways and their inhibitors.

C. Comparison of the associations obtained by different pathway methods. Number of associations on the vertical, FDR on the horizontal axis. PRGs yield more and stronger associations than all other pathway methods. Mutation associations are only stronger for TP53/Nutlin-3a and drugs that were specifically designed to bind to a mutated protein. PARADIGM not shown because no associations < 10% FDR.

D. Comparison of stratification by mutations and pathway scores. MAPK pathway (*BRAF*, *NRAS*, or *KRAS*) mutations and Trametinib on top left panel, AZ628 bottom left, *BRAF* mutations and Dabrafenib top right, and p53 pathway/*TP53* mutations/Nutlin-3a bottom right. For each of the four cases, the leftmost violin plot shows the distribution of IC$_{50}$s across all cell lines, followed by a stratification in wild-type (green) and mutant cell lines (blue box). The three rightmost violin plots show stratification of all the cell lines by the top, the two middle, and the bottom quartile of inferred pathway score (indicated by shade of color). The two remaining violin plots in the middle show mutated (*BRAF*, *KRAS*, or *NRAS*; blue color) or wild-type (*TP53*; green color) cell lines stratified by the top- and bottom quartiles of MAPK or p53 pathways scores (Mann-Whitney U test statistics as indicated).
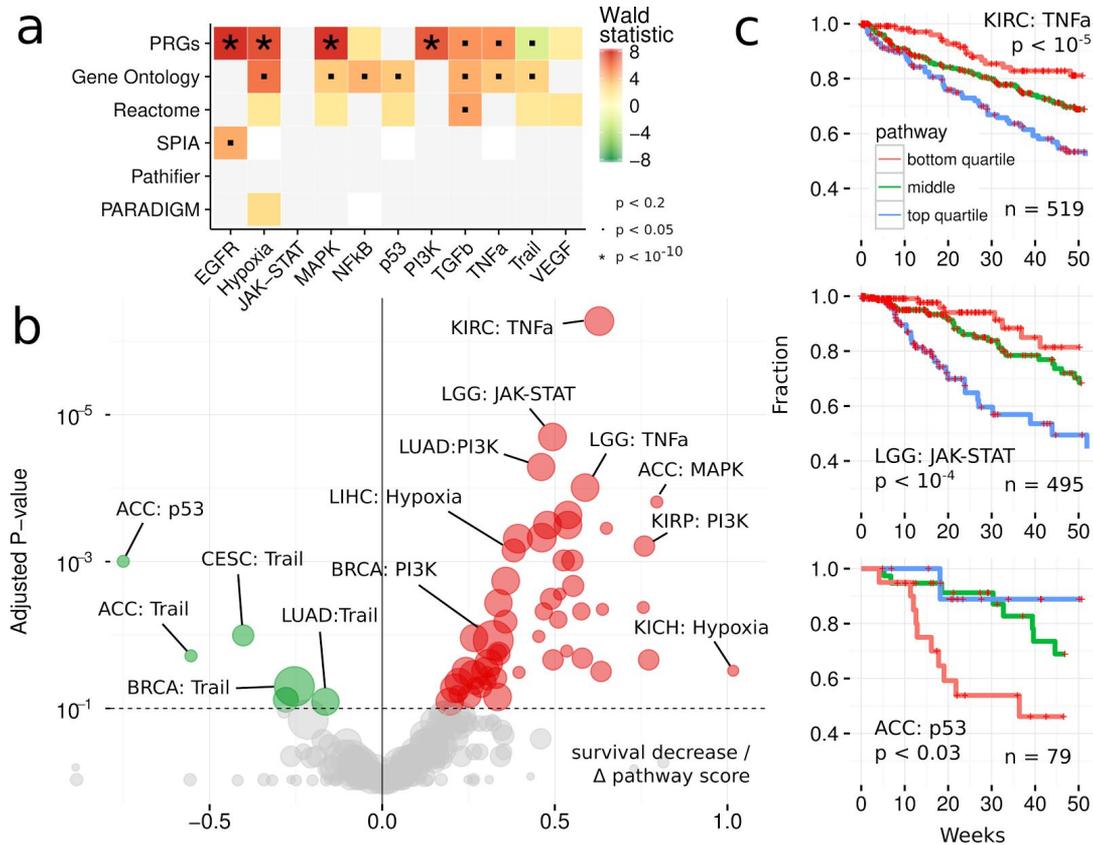
5. Response signatures outperform pathway methods for patient survival

A. Pan-cancer associations between pathway scores and patient survival. Pathways on the horizontal, different methods on the vertical axis. Associations of survival increase (green) and decrease . Significance labels as indicated. Shades correspond to effect size, p-values as indicated.

B. Volcano plot of cancers-specific associations between patient survival and inferred pathway score using PRGs. Effect size on the horizontal axis. Below zero indicates increased survival (green), above decreased survival (red). FDR-adjusted p-values on the vertical axis. Size of the dots corresponds to number of patients in each cohort.

D. Kaplan-Meier curves of individual associations for kidney (KIRC), low-grade glioma (LGG) and adrenocortical carcinoma (ACC). Pathway scores are split in top- and bottom quartiles and center half. Lines show the fraction of patients (vertical axis) that are alive at a given time (horizontal axis) within one year. P-values for discretized scores.