

Charting the cross-functional map between transcription factors and cancer metabolism

Karin Ortmayr¹, Sébastien Dubuis¹ and Mattia Zampieri^{1,*}

Affiliations:

¹ Institute of Molecular Systems Biology, ETH Zurich, Auguste-Piccard-Hof 1, CH-8093 Zurich, Switzerland.

*To whom correspondence should be addressed: zampieri@imsb.biol.ethz.ch

Abstract

Transcriptional reprogramming of cellular metabolism is a hallmark feature of cancer. However, a systematic approach to study the role of transcription factors (TFs) in mediating cancer metabolic rewiring is missing. Here, we chart a genome-scale map of TF-metabolite associations in human using a new combined computational-experimental framework for large-scale metabolic profiling of adherent cell lines, and the integration of newly generated intracellular metabolic profiles of 54 cancer cell lines with transcriptomic and proteomic data. We unravel a large space of dependencies between TFs and central metabolic pathways, suggesting that the regulation of carbon metabolism in tumors may be more diverse and flexible than previously appreciated. This map provides an unprecedented resource to predict TFs responsible for metabolic transformation in patient-derived tumor samples, opening new opportunities in designing modulators of oncogenic TFs and in understanding disease etiology.

Introduction

Transcription factors (TFs) are at the interface between the cell's ability to sense and respond to external stimuli or changes in internal cell-state¹. In cancer² as well as other human diseases³, alterations in the activity of TFs can remodel the cellular signaling landscape and trigger metabolic reprogramming⁴ to meet the requirements for fast cell proliferation and cell transformation^{5,6}. However, evidence linking alterations of cancer metabolism to TF dysfunction is often based on TF-binding sites upstream of metabolic enzymes⁴, without reporting on functional consequences of detected interactions. Mass-spectrometry based metabolomics approaches are powerful tools for the direct profiling of cell metabolism and to uncover mechanisms of transcriptional (in)activation of metabolic pathways⁷⁻¹⁰. Because of the limitations imposed by commonly used workflows, such as coverage, scalability and comparability, simultaneously quantifying the activity of TFs and metabolic

pathways at genome- and large-scale remains a major challenge. Here, we develop a unique experimental workflow for the parallel profiling of the relative abundance of more than 2000 putatively annotated metabolites in morphologically diverse adherent mammalian cells. This approach overcomes several of the major limitations in generating large-scale comparative metabolic profiles across cell lines from different tissue types or in different conditions, and was applied here to profile 54 adherent cell lines from the NCI-60 panel¹¹. To understand the origin of variance in the metabolome among different cancer cell types, we implemented a robust and scalable computational framework that integrates metabolomics profiles with previously published transcriptomic and proteomic datasets to resolve the flow of signaling information across multiple regulatory layers in the cell. This computational framework enables (i) systematically exploring regulation of metabolic pathways by TFs, (ii) reverse-engineering TF activity from *in vivo* metabolome profiles and (iii) predicting post-translational regulatory interactions between metabolites and TFs. Beyond contributing to the understanding of tumor-induced metabolic changes, this new systematic platform for charting genome-scale functional associations of metabolites with TFs in humans can introduce a new paradigm in the analysis of patient-derived metabolic profiles and the development of alternative strategies to counteract upstream reprogramming of cellular metabolism.

Results

Large-scale metabolic profiling of cancer cells reveals the interplay between transcriptional and metabolic heterogeneity.

Tumor cells, in spite of similar genetic background or tissue of origin, can exhibit profoundly diverse metabolic phenotypes¹²⁻¹⁴. To systematically investigate the origin of heterogeneity in tumor metabolism we exploit the naturally occurring variability among a diverse set of cancer cell lines from the NCI-60 panel of tumor-derived cell lines¹¹, representative of eight different tissue types. In spite of significant advancements in the rapid generation of high-resolution spectral profiles of cellular samples^{15,16}, the accurate comparative profiling of intracellular metabolite abundances across large panels of different cell types is still a major challenge. Here, we present an innovative and robust workflow to enable large-scale metabolic profiling in adherent mammalian cells alongside with a scalable computational framework to compare molecular signatures across cell types with large differences in morphology and size. In contrast to classical metabolomics techniques^{17,18} employing small-scale cultivation formats, laborious and time-consuming extraction procedures, and additional quantification of cell numbers/volume, we use a 96-well plate cultivation format, rapid *in situ* metabolite extraction, automated time-lapse microscopy and flow-injection time-of-flight mass spectrometry¹⁵ (FIA-TOFMS) for high-throughput profiling of cell extract samples (Figure 1a).

We optimized each step, from cultivation and extraction to MS analysis to be compatible with parallel 96-well processing. Different cell lines are seeded in triplicates at low cell density in 96-well microtiter plates, and are grown to confluence within 5 days (37°C, 5% CO₂). Growth is continuously monitored by automated acquisition of bright-field microscopy images in 1.5-hour intervals. Using an in-house developed software for large-scale image analysis, for each cell line we can estimate average adherent cell size (i.e. the area covered on average by a single cell during adherent growth) and cell numbers at sampling time (see Supplementary Figure S1 and Supplementary Note). Replicate 96-well plates are sampled for metabolome analysis in 24-hour intervals. To facilitate sampling, increase the throughput and reduce the risk of sample processing artifacts, we collect metabolomics samples directly in the 96-well cultivation plate without any prior cell detachment (Supplementary Figure S1). Cell extracts were analyzed by FIA-TOFMS¹⁵, allowing rapid full-spectral acquisition within less than one minute per sample.

Normalization of large-scale non-targeted metabolomics profiles is a fundamental aspect of data analysis¹⁹, which becomes particularly challenging when comparing mammalian cell lines with large differences in cell size. We propose a two-fold approach: (i) measure metabolome profiles from the same cell line at different cell densities, enabling a multiple regression approach to decouple cell line-specific metabolic signatures from differences in extracted cell amount, plate-to-plate variance and background noise (Supplementary Figure S2), (ii) image analysis of bright-field microscopy images to correct for differences in cell volume between cell lines (Supplementary Figures S1-2). Notably, by integrating MS readouts at multiple cell densities and time points during growth, the resulting estimates of relative metabolite abundances are invariant to incubation time and cell densities, enabling the direct comparison between cells of different types and/or in different conditions. A detailed description of the workflow and our systematic multivariate approach to normalize metabolome data is provided in the Online methods.

Here, we compared the intracellular metabolomes of 54 adherent cell lines from eight different tissue types in the NCI-60 cancer cell line panel. Consistent with large differences in physiology^{14,20} and growth rates (Supplementary Data Table S1), we observed highly heterogeneous metabolome profiles across cell lines from the same tissue type (Figure 1d). Differences in growth rate correlated with the intracellular abundance of several intermediate metabolites in amino acid biosynthesis (Spearman $|R| > 0.35$, Supplementary Figure S3), reflecting essential biochemical requirements for rapid cancer cell proliferation¹⁴. Differently from metabolome profiles, analysis of previously published transcriptomics²¹ and proteomics²² data sets on the 54 cell lines revealed a stronger molecular signature associating with the tissue of origin (Figure 1b-d, Supplementary Data Table S2). The two major components of metabolic variance across the 54 cell lines (38% explained variance) were

enriched for metabolites in nucleotide- and fatty acid metabolism as well as several signaling pathways (q-value < 0.05, Supplementary Figure S3), and correlated (Spearman $|R| > 0.37$) with the abundance of transcripts in signal transduction pathways regulating cell proliferation, adaptation, cell adhesion and migration (e.g. MAPK, HIF-1, PI3K-Akt, AMPK, Supplementary Figure S3). Altogether, these observations suggest metabolite abundances as closer readouts of cell physiology than mRNA or protein profiles, whilst pointing to a marked interplay between transcriptional regulation and the cell's metabolic state.

Mapping TF activity to metabolic phenotypes

To study the flow of signaling information between transcriptome and metabolome, we sought to quantify the functional interplay between different transcriptional programs and metabolic phenotypes. TFs can directly regulate metabolic fluxes by modulating enzyme abundance, thereby changing maximum flux capacity, or can indirectly affect substrate availability of proximal metabolic reactions, which can in turn result in local changes of fluxes and/or distal changes of reaction kinetics via post-translational regulatory mechanisms^{23,24}. To gain insight into the functional regulatory role of TFs in metabolism, we aimed to establish correlations between TF activity and metabolite abundances across 53 cell lines, using metabolite abundances as reporters of the impact of TF activity on metabolic reactions^{13,25} (see Supplementary Figure S4 and Supplementary Discussion).

Because TF activity is a complex function of post-transcriptional and post-translational mechanisms, monitoring TF gene expression or protein levels is an inadequate proxy of TF activity^{26,27} (Supplementary Figure S4). Instead, we derived a relative measure of TF activities by integrating previously published transcript abundance data²¹ for 53 cell lines with a genome-scale network of TF-target gene interactions in human²⁸ using Network Component Analysis (NCA)²⁹ (Figure 2a, Supplementary Figure S4, Supplementary Data Table S3). By systematically correlating the activity of 728 TFs to relative levels of individual metabolites across cell lines (Figure 2a), we generated a network of TF-metabolite associations (Supplementary Data Table S4, Supplementary Figure S4). While in human cells most of the known TF binding sites in enzyme-encoding genes map to disease-related signaling pathways, our TF-metabolite network unraveled a complementary large space of associations involving intermediates in central metabolic pathways (Figure 2b). To functionally characterize the role of individual TFs in regulating distinct metabolic pathways, we tested TF-metabolite associations for an overrepresentation of intermediates in KEGG metabolic pathways (Figure 2c, Supplementary Figure S5, Supplementary Data Table S4). We discovered a significant enrichment (q-value < 0.05) for 677 TFs (mean: 8, median: 7 pathways per TF). Because of the characteristic difference in carbon metabolism (e.g. Warburg effect) between cancer- and normal cells^{30,31}, oncogenic TFs that associate with central

metabolic pathways are of particular interest (Figure 2c). A set of 38 oncogenic TFs showed significant associations with glycolysis, pentose phosphate pathway, tricarboxylic acid (TCA) cycle and/or oxidative phosphorylation (q-value < 0.05, Supplementary Figure S5), potentially reflecting a large space of regulatory interactions that can mediate the adaptation to diverse micro-environmental conditions and nutrient availability. For example, several TFs in the NF- κ B signaling pathway (RelA, NF- κ B2 and Bcl-3) exhibit a significant association with the pentose phosphate pathway (q-value < 0.05, Figure 2c), possibly reflecting the role of NF- κ B complex as a mediator of redox homeostasis³² (see Supplementary Discussion).

The inferred associations between metabolic pathways and TFs are complementary to TF-gene regulatory networks. While the latter report on the regulatory capability of TFs to change enzyme abundance, coordinated changes between multiple metabolic intermediates and TFs can reveal the direct functional impact of TFs on pathway activity, and a TF's potential in mediating the reprogramming of specific metabolic pathways in cancer cells. An emblematic example is tumor protein 53 (p53), one of the most well-studied tumor suppressor proteins^{33,34}, for which significantly correlated metabolic intermediates in purine, arginine, proline- and nucleotide metabolism (q-value < 0.05) directly point to the functional role of the TF in sensing and allocating metabolic resources for cell division and DNA replication³³⁻³⁵. Even in cases where the role and target genes of TFs have been extensively characterized, the herein-proposed TF-metabolite associations can refine the condition-specific functional role of TFs in metabolism. For example, hypoxia-inducible factor 1 (HIF-1)³⁶ is reported to act on regulatory elements upstream of nearly hundred enzymes in central metabolic pathways (Figure 2c). Here, we find a significant association with TCA cycle intermediates (q-value < 0.05), suggesting that changes in HIF-1 activity can directly transmit to TCA cycle. The predicted direct regulatory role is consistent with experimental evidence for a HIF-1-mediated activation of pyruvate dehydrogenase kinase (PDK) that actively represses TCA cycle activity^{37,38}. In the following, we test the inferred TF-metabolite associations in an *in vivo* context, exploring the role of HIF-1 as a key oncogene in clear-cell renal cell carcinoma.

Inferring TFs mediating *in vivo* cancer metabolome rearrangements

Here, we ask whether the map of TF-metabolite associations found *in vitro* recapitulates metabolic rearrangements in an *in vivo* setting. To this end, we developed a computational framework that searches for TFs potentially responsible for metabolic differences between healthy and tumor tissue. The approach is based on a scoring function that evaluates the dot product between the TF-metabolite correlation matrix derived *in vitro*, and *in vivo* metabolite fold-changes between normal and cancer tissues (Supplementary Figure S6). Here, we applied this approach to analyze differences in metabolite

abundances between clear-cell renal cell carcinoma (ccRCC) and proximal normal tissue samples in a cohort of 138 patients⁷. For each patient, we estimated the significance of a TF in explaining the observed metabolic changes, and ranked the 728 TFs according to the median of q-values across patients (Figure 3).

Loss-of-function mutations in van Hippel Lindau (VHL) TF gene are the most frequent and specific genetic event observed in ccRCC³⁹, entailing a hyper-activation of hypoxia-inducible factors (HIF-1, HIF-2 and HIF-3)⁴⁰. In agreement with the genetic basis of ccRCC, we identified VHL and HIF-1 α among the top 1% of TFs that potentially mediate metabolic rearrangement in ccRCC (Figure 3). Other top-ranking TFs include YY1 that can further stabilize HIF levels in VHL-null renal carcinoma⁴¹, and TFAP2A, whose gene is frequently hyper-methylated in ccRCC⁴² (see also Supplementary Discussion and Supplementary Figure S6). These results support the relevance of the previously inferred *in vitro* map of TF-metabolite associations and its potential clinical value in the interpretation of *in vivo* metabolome profiles of tumor tissue samples.

To further illustrate the potential of TF-metabolite associations in aiding the interpretation of tumor specific metabolic changes, we collected data from two additional studies monitoring tumor metabolic reprogramming in cohorts of 10 and 21 patients with colon¹³ and lung cancer⁴³, respectively. In contrast to ccRCC, the most recurrent genetic events in colon and lung cancers are mutations in p53^{44,45}, shared across many different tumor types. In such cases, our approach can dissect specific TFs, possibly effectors up- or downstream of p53, that can more directly explain tumor-specific metabolic changes. In colon cancer, we predicted an involvement of Mllt10, a downstream effector of Wnt signaling that is activated in many colorectal cancers⁴⁶, as well as RelB, an NF- κ B subunit which contributes to the rewiring of cancer metabolism through crosstalk with p53 and is involved in colitis-associated cancers⁴⁷ (Figure 3 and Supplementary Discussion). Similarly, the metabolic changes observed in the lung cancer cohort highlight AATF, an upstream regulator of p53⁴⁸, as well as NF-Y, a putative regulator of cancer metabolism that can act in complex with gain-of-function mutant p53 to increase DNA synthesis⁴⁹ (Figure 3 and Supplementary Discussion). Taken together, *in vitro* associations between TFs and metabolites not only provide a means to functionally characterize the role of TFs across diverse cancer types, but could also complement genomic information and guide the analysis, molecular classification and interpretation of metabolome profiles from large patient cohorts.

Systematic mapping of TF activity modulators

While we have shown that metabolic rearrangements in cancer can be described by changes in TF activity, the origin of such changes often remains elusive. Mutations in genes encoding TFs can be directly responsible for altered TF functionality, and in some cases even explain disease etiology.

However, the activation of new transcriptional programs is often an indirect response to changes in the abundance of internal effectors of cell signaling^{50,51}. *In silico* models have proven extremely powerful in finding new allosteric interactions that can regulate enzyme activity⁵² and in testing their *in vivo* functionality⁵³, but little progress has been made in the systematic mapping of effectors of TF activity. Here, we integrated three layers of biological information to obtain first insights into how TFs could be embedded in a regulatory network allowing the cell to activate distinct transcriptional programs in response to changes in the abundance of specific intracellular metabolites.

Besides mutations in TF-encoding genes, interactions with other proteins (co-factors or kinases/phosphates) and metabolites (e.g. allosteric binding) within the cell can repress or enhance TF activity. Despite the increased interest in metabolites as signaling molecules and their potential role in driving cellular transformation^{50,54-56}, resolving the influence of metabolites on TF activity has remained a daunting task⁵⁷. Here, we established an *in silico* framework for generating hypotheses on regulatory interactions between TFs, metabolites and kinases (Figure 4a). To that end, we used model-based fitting analysis to integrate TF activity and metabolome profiles with proteome data²² measuring the abundance of 100 TFs and 64 kinases/phosphatases across 53 cell lines. For each TF, we applied nonlinear regression analysis to determine whether variation in TF activity across the 53 cell lines could be modeled as a function of TF protein abundance and the activating or inhibiting action of individual metabolites and/or kinases. In total we tested 6,753,600 models, and found 1,888 interactions which significantly (FDR \leq 0.1%) improved the explained variance in the activity of 96 TFs. Most of the inferred regulatory interactions (93%) involved the combined action of a kinase and a metabolite, while only few metabolites or kinases could alone significantly improve model-fitting (Figure 4b, Supplementary Data Table S5). This observation suggests that multiple coordinated regulatory mechanisms underlie the post-transcriptional regulation of TF activity. On average, each metabolite engaged in 8 interactions (median: 3), while kinases on the other hand were predicted to be more ubiquitous effectors (13-125 interactions per kinase) (Supplementary Figure S7).

This blueprint interaction map can serve as a new resource for generating hypotheses and designing target-oriented experimental approaches for in-depth mechanistic studies identifying metabolic regulators of transcriptional reprogramming. The most highly connected metabolite is choline (118 interactions with 36 TFs), substrate of the first step in the biosynthesis of phosphatidylcholine, a key component of eukaryotic cellular membranes. Recent studies have revealed a deep interplay between oncogenic signaling and choline metabolism, such that altered choline levels have become a metabolic hallmark of malignant transformation⁵⁸. Our results support this key role, suggesting that choline availability could be sensed and can trigger a wide range of transcriptional adaptive processes. Further examples of predicted regulatory interactions are provided in the Supplementary Discussion.

Overall, our approach opens the door for a systematic investigation of a previously largely unexplored⁵⁹ interaction space between transcriptional regulators and signaling effectors in human cells. The herein-presented experimental workflow for large-scale comparison of metabolomes in different cancer types, and *in silico* modeling of regulatory actions of metabolites and/or kinases can become an invaluable methodology in finding unique effectors that can trigger changes in the cell's transcriptional program.

Discussion

In recent years, increasing efforts have been made to understand the signals driving metabolic changes in cancer¹³. Transcriptional regulation is at the basis of the decision-making process of a cell and its ability to allocate resources necessary for cell transformation and proliferation. Genome sequencing and transcriptome technologies have revealed an intricate network of transcription factor–target gene interactions in which TF mutations often associate to disease states and aberrant metabolic phenotypes^{7,60,61}. However, it is important to emphasize that gene regulatory interactions between enzymes and TFs per se are not sufficient to functionally regulate the activity of metabolic pathways. Key to unambiguously resolving regulatory circuits at the intersection with metabolism are methods searching for coordinated behavior between the different levels. Here, we integrate biological information across three layers: the transcriptome, the proteome, and the metabolome of 53 phenotypically diverse cell lines from the NCI-60 panel of human tumor cell lines. Key to this approach was the development of a combined experimental and computational framework (Figure 1a) that overcomes important limitations in large-scale metabolome screenings, including (i) the limited throughput and laborious sample preparation of classical metabolomics approaches in mammalian cell cultures^{17,18}, (ii) the lack of scalable methods to adequately normalize metabolomics data across morphologically diverse cell types, and (iii) the need for systematic data integration strategies.

Here, by integrating cross-sectional omics data from diverse tumor cell lines, we constructed a global network model across three layers of biological information, exploiting the naturally occurring diversity in an *in vitro* cell line system. By analyzing the coordinated changes in baseline transcriptome, proteome and metabolome with the aid of a gene-regulatory network and model-based fitting analysis, we investigated the information flow from TFs to metabolic pathways and vice versa. The herein-constructed network of TF-metabolite associations provides an unbiased metabolic characterization of TF functions in human, and a more systematic understanding of functional regulatory interactions that can mediate metabolic adaptation in cancer cells. For many TFs, we found new regulatory associations with central metabolic pathways, suggesting a large space of transcriptional solutions by which cells can fulfill the anabolic and catabolic requirements for rapid

proliferation⁶² and adaptation to nutrient limitations^{63,64}. Secondly, we constructed a map of functionally relevant modulatory interactions between transcription factors, cell metabolism and cell signaling (kinases). Most interactions included the combined action of a metabolite and a kinase, acting as allosteric effectors or post-translational modifiers of TFs. An extensive network of these two types of physical interactions has emerged in model organisms such as *E. coli* or Yeast^{24,65,66}, and based on our findings we expect a similar picture to hold true in humans⁵⁹. However, the identity and condition-dependent relevance of such interactions is much harder to determine experimentally in human cells than in unicellular model organisms. As a result, discoveries of TF-metabolite interactions in human cells have been sporadic⁶⁷. The analysis and predictions proposed in this study offer new opportunities for a more systematic and exploratory discovery of TF post-translational regulatory interactions. The experimental and computational components of our approach are scalable and should in the future be applied to include a direct readout of kinase activity (e.g. phosphoproteomics data), broader protein coverage, larger cohorts of cell lines and more diverse environmental conditions. The relevance of TF-activity modulators extends beyond cancer metabolism, to the general understanding of how fluctuating micro-environments trigger coordinated responses of signaling pathways and culminate in the adaptive rewiring of the cell's functional landscape.

In light of the central regulatory role of TFs in cellular organization, targeting TFs is an extremely attractive way to counteract global gene expression changes that underlie cancer development⁶⁸⁻⁷⁰. Endogenous metabolites capable of modulating TF activity could become invaluable chemical scaffolds to design new therapeutic molecules targeting oncogenic TFs, with the potential to overcome shortcomings related to targeting kinase-mediated signaling cascades⁷⁰. The herein proposed workflow for large-scale metabolome profiling is directly applicable to the study of dynamic metabolic responses to external stimuli, and can scale to larger cohorts that are now within reach of other molecular profiling platforms⁷¹. Altogether, our work also suggests that, while clearly far from typical *in vivo* conditions, *in vitro* cell line systems represent an invaluable discovery tool to investigate metabolic regulatory mechanisms that can still generalize to *in vivo* conditions and clinical settings. The experimental and computational framework proposed in this study is applicable to other systems or diseases, providing us with an unprecedented tool to investigate the origin of metabolic dysregulation in human diseases.

Acknowledgments

We thank the National Cancer Institute (NCI) for providing the cancer cell lines. This work was supported by a Worldwide Cancer Research (WCR-15-1058) project funding to M.Z., K.O. was funded by the Austrian Science Fund (FWF): FWF P26603 and FWF W1224 Doctoral Program BioToP – Biomolecular Technology of Proteins. We thank Uwe Sauer and Nicola Zamboni for supporting this

work and providing laboratory facilities, Dimitris Christodoulou, Maren Diether, Nicola Zamboni and Victor Chubukov for helpful feedback and discussions.

Contributions

M.Z. designed the project. S.D. and K.O. performed the metabolome experiments. M.Z. and K.O. analysed the data. S.D. and M.Z. designed and implemented the image analysis framework. M.Z. and K.O. wrote the manuscript. All authors contributed to preparing the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Materials and Methods

Cell cultivation. The NCI-60 cancer cell lines were obtained from the National Cancer Institute (NCI, Bethesda, MD, USA). After thawing, the adherent cell lines were expanded in cell culture flasks (Nunc T75, Thermo Scientific) at 37°C and 5% CO₂ in RPMI-1640 (Biological Industries, cat.no. 01-101-1A) supplemented with 5% fetal bovine serum (FBS, Sigma Aldrich, F6178), 2 mM L-glutamine (Gibco, cat.no. 25030024), 2 g/L D-glucose (Sigma Aldrich, cat.no. G8644), and 100 U/mL penicillin/streptomycin (P/S, Gibco, cat.no. 15140122). After two passages, the cells were transferred to fresh medium where FBS was replaced by dialyzed FBS (dFBS, Sigma Aldrich, cat.no. F0392) with a reduced content of low molecular weight compounds, to improve accuracy of metabolite quantification. Cells were maintained in medium with dFBS throughout all experiments. The starting cell density for metabolomics experiments was determined for each cell line. To this end, cells were plated in triplicates at eight different starting cell densities and incubated at 37°C and 5% CO₂ for 3 days. On the third day, the medium was changed in all wells by aspirating the spent medium using a multi-channel aspirator, washing once with phosphate-buffered saline solution (PBS, pH 7.4, Gibco, cat.no. 10010015 at 37°C) using a multi-channel dispensing pipet, and finally filling each well again with 150 µL of fresh medium. The plate was imaged to determine the confluence (see below) immediately before and after media change, and after 72 hours. The starting cell density for metabolomics experiments was then chosen to guarantee a minimum of 20-30% cell confluence before media change, and approx. 80% confluence after 72 hours.

Cell imaging and image analysis. We monitor cell growth by measuring cell confluence (i.e. area of the well covered by cells in percentage) directly from 96-well plates using automated time-lapse microscopy imaging. Every 1.5 hours, bright-field microscopy images of each well were acquired using a TECAN Spark 10M plate reader. In addition, we developed an image analysis framework to segment cells and determine the characteristic cell size area (i.e. average surface area of single adherent cells) for each cell line (Supplementary Figure S1). To quantify the number of extracted cells, confluence was divided by the characteristic cell size (Supplementary Figure S1). A detailed description and validation of the algorithm used for estimating cell numbers from bright-field microscopy images is provided in the Supplementary Note, alongside with a Matlab code. Of note, this approach has several important advantages, in that it is non-destructive, and allows quantifying cell growth and cell numbers without any manual sample manipulation.

Metabolomics experiments. Cells were plated in triplicates in 150 µL of RPMI1640 medium (5% dFBS, 2 g/L glucose, 2 mM glutamine, 1% P/S) in 96-microtiter well plates. After an initial growth phase, the medium in each well was renewed on the third day, and the cultures were subsequently monitored for four more days (96 hours). Ten replicate plates were prepared in each experiment to allow for

generating metabolomics samples at five different time-points (immediately before media change, and at 24, 48, 72 and 96 hours after media change), and one additional plate for continuous growth monitoring (TECAN Spark 10M, 37°C, 5% CO₂).

At each sampling time point, two replicate 96-well plates were processed (plates A and B). Plate A was used to generate cell extracts, by (1) removing the spent medium, (2) washing once with 75 mM ammonium carbonate (pH 7.4), and (3) adding ice-cold extraction solvent (40% methanol, 40% acetonitrile, 20% water, 25 μM phenyl hydrazine⁷²). Finally, the plate is sealed, incubated at -20°C for one hour, and subsequently stored at -80°C until MS analysis. Plate B undergoes the same processing steps, except for the last one, where each well is filled with 37°C PBS (pH 7.4), and the plate is immediately imaged to determine cell confluence for subsequent normalization of MS spectra.

Immediately prior to MS analysis, the plates were thawed on ice, and the extracted cells were scraped off the bottom of each well using a multi-channel pipet with wide-bore tips. Next, the cell extracts were transferred to 96-well plates with conical bottom and centrifuged at 4°C, 4000 rpm for 5 min to separate cell debris. Finally, pooled cell extracts for each experiment (“master mixes”, pooled from five cell lines processed within the same experiment) as well as aliquots of unused extraction solvent were added to each measurement plate as control samples, and the plates were sealed and stored at 4°C until injection.

Metabolome profiling using FIA-TOFMS. Cell extract samples were analyzed by flow-injection analysis time-of-flight mass spectrometry (FIA-TOFMS) on an Agilent 6550 iFunnel Q-TOF LC/MS System (Agilent Technologies, Santa Clara, CA, USA), as described by Fuhrer et al¹⁵. This method allows the generation of high-resolution spectral profiles in less than one minute per sample, allowing for sensitive high-throughput profiling of large sample collections. In brief, a defined sample volume of 5 μL is injected using a Gerstel MPS2 autosampler into a constant flow of isopropanol/water (60:40, v/v) buffered with 5 mM ammonium carbonate (pH 9), containing two compounds for online mass axis correction: 3-Amino-1-propanesulfonic acid, (138.0230374 m/z, Sigma Aldrich, cat. no. A76109) and hexakis(1H,1H,3H-tetrafluoropropoxy)phosphazine (940.0003763 m/z, HP-0921, Agilent Technologies, Santa Clara, CA, USA). The sample plug is delivered directly to the ion source for ionization in negative mode (325°C source temperature, 5 L/min drying gas, 30 psig nebulizer pressure, 175 V fragmentor voltage, 65 V skimmer voltage). Mass spectra were recorded in the mass range 50-1000 m/z in 4 GHz high-resolution mode with an acquisition rate of 1.4 spectra per second. Raw MS profiles were processed to align spectra and pick centroid ion masses using an in-house data processing environment in Matlab R2015b (The Mathworks, Natick).

Metabolite annotation. Measured ions were putatively annotated by matching mass-to-charge ratios to a reference list of calculated masses of metabolites listed in the Human Metabolome Database (HMDB) and the genome-scale reconstruction of human metabolism⁷³ (Recon2) within 0.003 amu mass accuracy. The reference mass list was generated from the respective sum formulae, considering deprotonation as the most prevalent mode of ionization in the chosen acquisition conditions. To allow for the annotation of α-keto acid derivatives formed in presence of phenyl hydrazine in the extraction solvent⁷², sum formulae for the phenylhydrazone derivatives (+C₆H₈N₂ -H₂O) of a total of 30 α-keto acid compounds (selected via KEGG SimComp search <http://www.genome.jp/tools/simcomp/>) were added to the metabolite list for annotation. The final list of putatively annotated metabolites consisted of 689 and 5949 unique compound IDs from Recon2, and HMDB, respectively.

Data normalization. We corrected for systematic errors using a two-step regression model to disentangle the contributions of extracted cell numbers, plate-to-plate variance, instrumental and background noise from the actual variance in metabolite abundances between cell types (Supplementary Figure S2). To this end, raw data were first corrected for instrument drift by

normalizing for possible batch/plate effects. Each plate contains 12 pooled cell extract samples prepared from the 5 different cell lines in each experiment (i.e. batch). Measured intensities for each annotated ion are modeled as follows:

$$I_{i,j,p} = \gamma_p \cdot M_{i,j} \quad (\text{Eq. 1})$$

$$\log(I_{i,j,p}) = \log(\gamma_p) + \log(M_{i,j}) \quad (\text{Eq. 2})$$

Where $I_{i,j,p}$ is the measured intensity for ion i , in pooled sample j and plate p , γ_p is the scaling factor associated to each plate and $M_{i,j}$ represents the actual abundance of metabolite i in sample j . By using a linear regression scheme we can estimate both parameters (γ_p and $M_{i,j}$) within an unknown scaling factor. After correcting for possible instrumental artifacts, we implemented a second step in order to derive comparative measurements of metabolite abundance for each cell line. Here, we follow each cell line along the linear growth phase, sampling every 24 h across 5 days. We typically obtain 25 data points for each cell line at different cell densities. The expectation is that the signal measured for any ion of biological origin (i.e. a genuine metabolite) would increase linearly as the number of cells in the sample increases. The proportionality (i.e. α parameter) between ion intensity and extracted cells depend on the intracellular concentration of the metabolite. Here, by implementing a multiple regression scheme, we estimate the relative abundance of a given metabolite in each of the cell lines, α (Supplementary Figure S2), alongside with its standard error. A linear regression model describes variation in ion intensity as a linear function of cells extracted (α) and a constant parameter (β) that capture MS background noise. For each ion, the α values are specific of each cell line while the constant term in the model is fixed (i.e. expected ion signal at zero confluence that is independent of the cell line).

Because of the large number of measurements for each cell line at different cell densities, we can apply a multiple regression analysis (*fitlm* function in Matlab) to infer all model parameters α s and β at once, by minimizing the Euclidian distance between measured metabolite intensities and model predictions. For each metabolite, we solve the following linear model, including all 54 cell lines:

$$\begin{bmatrix} I_{cell_{1,1}} \\ I_{cell_{1,2}} \\ I_{cell_{1,3}} \\ \dots \\ I_{cell_{2,1}} \\ I_{cell_{2,2}} \\ I_{cell_{2,3}} \\ \dots \\ I_{cell_{c,s}} \end{bmatrix} = \begin{bmatrix} N_{cell_{1,1}} & 0 & \dots & 0 & 1 \\ N_{cell_{1,2}} & 0 & \dots & 0 & 1 \\ N_{cell_{1,3}} & 0 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & N_{cell_{2,1}} & \dots & 0 & 1 \\ 0 & N_{cell_{2,2}} & \dots & 0 & 1 \\ 0 & N_{cell_{2,3}} & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & N_{cell_{c,s}} & 1 \end{bmatrix} \cdot [\alpha_{cell_1} \quad \alpha_{cell_2} \quad \dots \quad \alpha_{cell_c} \quad \beta] \quad (\text{Eq. 3})$$

Where $I_{cell_{c,s}}$ is the measured metabolite intensity in sample s of cell line c , $N_{cell_{c,s}}$ is the number of cells extracted in sample s of cell line c , and α s (for each cell line) and β are the unknown parameters to be fitted. The number of cells per sample is derived from the confluence measurements at sampling, divided by the average cell size area determined using our image segmentation analysis (see Supplementary Figure S2 and Supplementary Note). After this step, we retained 2181 ions with a regression p-value below a threshold value of 3.4×10^{-7} (adjusted by the number of cell lines and ions) in at least one cell line, and that showed a significant dependency with the extracted cell number in more than 80% of cell lines (Supplementary Figure S2). Of note, we found that prior to normalization, the variance across three biological replicates at the same time-point was equally low in cell confluence (median: 7.4% CV) and raw ion intensities (median: 13%, Supplementary Figure S2), reflecting the high quality of MS measurements.

In the third and last step, we take into account systematic changes in metabolite abundances related to differences in cell size (i.e. cell volume) between the 54 cell lines to derive comparative estimates of intracellular metabolite concentration. Principal component analysis of relative metabolite abundance per cell revealed a strong trend across the 54 cell lines (PC1, 58.9% explained variance, Supplementary Figure S2), which strongly correlates with cell line volumes derived from cell diameters measured in ⁷⁴ as well as with the herein determined adherent cell size area (Supplementary Figure S2, see Supplementary Note). The transitive correlation between adherent cell size and the spherical cell volume in suspension indicates that adherent cell height can be approximated as a constant. To correct MS data for differences in cell line volumes, we selected 987 ions that showed a significant and strong correlation (Pearson's $r > 0.8$, $p < 0.05$, Supplementary Figure S2) with PC1. KEGG pathway enrichment analysis showed that these putatively annotated metabolites were strongly overrepresented in fatty acid metabolism (Supplementary Figure S2), consistent with the expected linear dependency between cell membrane surface (i.e. phospholipid content) and cell volume of adherent cells. For each ion, cell line-specific α -values of the selected metabolites across all 54 cell lines were used to calculate a consensus correction factor for each cell line by taking the mean across the 987 ions. To apply the cell volume correction to the full data set, we divided the cell line-specific α -values for each ion by the consensus correction factor.

Finally, the corrected α -values were normalized using Z-score normalization:

$$Z_{\alpha} = \frac{(\alpha_{Cell} - \bar{\alpha})}{\sqrt{\frac{1}{n} \sum_{Cell} (\alpha_{Cell} - \bar{\alpha})^2}}, \quad (\text{Eq. 4})$$

where n is the number of cell lines (i.e. 54).

The final normalized data set is provided in Supplementary Data Table S1 alongside with p-values and standard errors derived from regression analysis. Missing values ('NaN') correspond to cases where the measured ion abundance for the annotated metabolite was close to the background level in the cell line, and can be considered as zero for further analysis. In cell lines where a significant dependency ($p < 3.4e-7$) of given metabolite's abundance with cell number could be robustly determined and exceeded the background noise, relative standard errors of α calculated during fitting analysis were below 20% (median: 11%, Supplementary Figure S2).

Analysis of tissue signature in transcriptomics, proteomics, metabolomics and drug sensitivity data sets. Principal component analysis (PCA) was used to dissect each data set into major trends, and quantify the extent of variance explained by each of these principal components (Supplementary Figure S3). The PC scores obtained for the different cell lines in each principal component and data set were then analyzed for signatures of individual tissue types based on the pairwise Euclidean distance between cell lines. For each cell line, the enrichment of cell lines from the same tissue type among its nearest neighbors (i.e. smallest distance scores) was tested using a hypergeometric test. The resulting probabilities (p-values) from enrichment analysis were subjected to receiver-operator-characteristic (ROC) curve analysis to evaluate the performance of a given principal component as a predictor of the tissue type (Supplementary Table S2). A principal component was assigned a tissue signature when the area under the ROC curve (AUC) exceeded a value of 0.8 for any tissue type, and the fraction of explained variance was summed up to obtain the overall percentage of variance explained by tissue type in each data set (Figure 1c).

Estimating TF activity using Network Component Analysis (NCA). Originally established by Liao et al.²⁹, NCA provides a mathematical framework for reconstructing TF regulatory signals (TF activity) from gene expression profiles. Here, we adopted sparseNCA implementation by Noor et al.⁷⁵ (Matlab code

downloaded from <https://sites.google.com/site/aminanoor/software>). This methodology adopts a mathematical model to approximate TF-target regulatory interactions and integrates prior network information with the expression of target genes across multiple conditions to regress the activity of the respective TFs, delivering a relative measure of TF activity. We obtained normalized gene expression profiles across the NCI-60 cell lines from Gene Expression Omnibus (accession number GSE32474), containing 54,675 mRNA probes. TRRUST database²⁸ served as the source of TF-target gene interactions relevant in human, including 748 human TFs and 1,975 non-TF gene targets. Intersecting these two resources, we assembled a network of 2,209 unique genes corresponding to 5490 mRNA probes that match target genes of 728 TFs in the TRRUST database (Supplementary Figure S4). We implemented a bootstrapping approach to account for incompleteness of the regulatory network (i.e. missing regulatory interactions), and the fact that there may be multiple optima in the solution space. To this end, for each TF we randomly selected 48 additional TFs and constructed a sub-network containing the 49 TFs and their target genes. Because growth-rate has a pleiotropic effect on gene expression, here reflected in the correlation between first principal component of gene expression data and cell line growth rates (Supplementary Figure S4), we decouple TF activity from the confounding effect of growth-rate by adding an additional TF that targets all genes. This fictitious TF mimics the general effect of proliferation rates on transcription. As a result, each TF is embedded in a sub-network of 50 TFs and their target genes from the full network. Ten such sub-networks were created randomly for each TF to apply NCA. In this bootstrapping scheme, each TF was sampled in on average 490 subnetworks (permutations, min. 423, max. 556 data points per TF). In the final data set, we normalized the calculated TF activity to the maximum across all permutations, and finally calculated the median TF activity and its standard deviation for each TF and cell line (Supplementary Figure S4). It is worth noting that the estimates we obtain with this approach are correct within an unknown scaling factor, and hence we determine a relative measure of activity for each of the 728 TFs across the NCI-60 cell lines.

TF-metabolite association network and inferring TF involvement from *in vivo* metabolic changes. In order to find metabolites whose relative abundances correlate with TF activity, we calculated pairwise Spearman correlations between all 2181 annotated metabolites and 728 TFs across the 54 cell lines. In order to control the false discovery rate (FDR) among network links, we used a bootstrapping approach to calculate 99.5 and 99.9% confidence intervals of correlation coefficients after randomizing the data set. To that end, the cell lines in the metabolome data set were randomized by resampling 100 times with replacement, and Spearman correlation coefficients were calculated for each randomized data set. Correlation coefficients yielding 99.5 and 99.9% confidence intervals (0.5 and 0.1% FDR, respectively) were obtained from the pooled list of absolute correlation coefficients by finding the smallest correlation coefficient that exceeds the maximum value among 99.5 and 99.9% lowest correlation coefficients, respectively).

Based on this association network, the procedure used to estimate the contribution of each of the 728 TFs in mediating metabolic changes observed *in vivo*, consists of 3 main steps: (i) For each patient \log_2 fold-changes of detected metabolites between cancer and adjacent normal tissue are estimated (FC^p), (ii) The dot product between metabolite fold-changes and the TF-metabolite correlation vector (C_{TF} , product of correlation R and $-\log_{10} p$) estimated from *in vitro* cell lines is computed for each TF (S_{TF}^p), (iii) the significance is estimated using a permutation test, where metabolite order is shuffled 10,000 times, the dot product is estimated for each random permutation (\tilde{S}_{TF}^p) and p-values are estimated as follows:

$$S_{TF}^p = \frac{C_{TF} \cdot FC^p}{\|C_{TF}\|^1} \quad (\text{Eq. 5})$$

$$p - value_{TF} = \frac{\sum_k^{10,000} (\delta_{TF}^p \geq S_{TF}^p)}{10,000} \quad (\text{Eq. 6})$$

P-values are corrected for multiple tests by q-value estimation⁷⁶, and the median across patients is calculated. Notably, when analyzing the data published in⁷ we excluded all detected metabolites with more than 10 missing values across patient samples.

KEGG enrichment analysis of TF-metabolite correlation signatures. To assess the over-representation of TF-metabolite associations in KEGG metabolic pathways, the pairwise correlation between TF activity and metabolite relative abundance across cell lines were rank-transformed. A statistical score that models the probability of a KEGG pathway to be significantly associated to a TF is based on the collective activities of multiple metabolites in a pathway following the approach described in⁷⁷. The significance of the rank distribution of all metabolites within the same KEGG pathway is tested by means of an iterative hypergeometric test, indicating the statistical significance of metabolic intermediates of a common metabolic pathway (e.g. TCA cycle) being distributed toward the top ranking ones. P-values were corrected for multiple tests by q-value estimation⁷⁶.

Prediction of metabolite-TF effectors. Because of the poor correlation between protein levels and NCA estimated TF activities (Supplementary Figure S4), we systematically investigate whether variation in TF activities across cell lines could be explained by alternative models. Instead of assuming a base model where the TF activity is a simple linear function of protein abundance, we model TF activity as a function of TF protein levels and the post-translational regulatory functions of metabolites and/or kinases, individually or in combination as follows:

$$TF_{activity}^j = \frac{K_1}{1 + \frac{M_i}{K_2}} \cdot P_{TF^j} + K_3 \quad \text{for a metabolite inhibitor,}$$

$$TF_{activity}^j = \frac{K_1 \cdot M_i}{M_i + K_2} \cdot P_{TF^j} + K_3 \quad \text{for a metabolite activator,}$$

$$TF_{activity}^j = \frac{K_1}{1 + \frac{P_x}{K_2}} \cdot P_{TF^j} + K_3 \quad \text{for a kinase inhibitor,}$$

$$TF_{activity}^j = \frac{K_1 \cdot P_x}{P_x + K_2} \cdot P_{TF^j} + K_3 \quad \text{for a kinase activator,}$$

$$TF_{activity}^j = \left(\frac{K_1 \cdot P_x}{P_x + K_2} + \frac{K_1}{1 + \frac{M_i}{K_4}} \right) \cdot P_{TF^j} + K_3 \quad \text{for a kinase activator \& metabolite inhibitor,}$$

$$TF_{activity}^j = \left(\frac{K_1}{1 + \frac{P_x}{K_2}} + \frac{K_1}{1 + \frac{M_i}{K_4}} \right) \cdot P_{TF^j} + K_3 \quad \text{for a kinase inhibitor \& metabolite inhibitor,}$$

$$TF_{activity}^j = \left(\frac{K_1 \cdot P_x}{P_x + K_2} + \frac{K_1 \cdot M_i}{M_i + K_4} \right) \cdot P_{TF^j} + K_3 \quad \text{for a kinase activator \& metabolite activator,}$$

$$TF_{activity}^j = \left(\frac{K_1}{1 + \frac{P_x}{K_2}} + \frac{K_1 \cdot M_i}{M_i + K_4} \right) \cdot P_{TF^j} + K_3 \quad \text{for a kinase inhibitor \& metabolite activator.}$$

Where $K_{1,2,3,4}$ are free parameters in the model, P_{TF}^j is the relative protein abundance associated to TF^j , P_x represents the relative protein abundance of kinase x and M_i the level of metabolite i.

For models in which we assume the non/inhibitory actions of a single metabolite or kinase, we use a non-linear model fitting scheme to find the best set of 3 parameters to describe changes in TF activity across cell lines. To minimize the probability of local minima, we adopted the GlobalSearch function of Matlab and 50,000 starting points. By using the same approach, we tested all possible combinations

between pairs of metabolites and kinases that can act either as activators and/or inhibitors, to find the best set of 4 parameters that describe TF activity. For each pair or triplet of TFs and metabolites and/or kinases we estimate the Mean Squared Error (MSE) associated to each tested model. In addition, we estimated the MSE when fixing the model parameters and randomly permuting TF activity, TF protein, metabolite and kinase levels across cell lines (MSE^e). For each TF we then repeat model fitting, each time by randomly shuffling metabolite and kinases levels. We used this approach to empirically assess the significance of each model in better explaining TF activity. To this end, we calculated MSE values for each model in which kinases and metabolite levels were randomly permuted (MSE_{Random} and MSE^e_{Random}). Finally we retain only those models in which MSE and MSE^e are both above the 0.1% of the respective distributions of MSE_{Random} and MSE^e_{Random} . The proteome dataset (ArrayExpress, project accession: E-PROT-2) report protein levels for 100 annotated TFs and 63 kinases, and to reduce computation time we here consider only differentially abundant metabolites annotated to KEGG identifiers. Fitting analysis was performed in Matlab using the “fitnlm” function, on a cluster with 900 computational nodes.

References

1. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* **514**, 376–379 (2014).
2. Kruiswijk, F., Labuschagne, C. F. & Vousden, K. H. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat. Rev. Mol. Cell Biol.* **16**, 393–405 (2015).
3. Green, D. R., Galluzzi, L. & Kroemer, G. Metabolic control of cell death. *Science* **345**, 1250256 (2014).
4. Yuneva, M. O. *et al.* The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab.* **15**, 157–170 (2012).
5. Boroughs, L. K. & DeBerardinis, R. J. Metabolic pathways promoting cancer cell survival and growth. *Nat. Cell Biol.* **17**, 351–359 (2015).
6. Reznik, E., Wang, Q., La, K., Schultz, N. & Sander, C. Mitochondrial respiratory gene expression is suppressed in many cancers. *eLife* **6**, e21592 (2017).
7. Hakimi, A. A. *et al.* An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **29**, 104–116 (2016).
8. Bartel, J. *et al.* The Human Blood Metabolome-Transcriptome Interface. *PLOS Genet.* **11**, e1005274 (2015).
9. Eckel-Mahan, K. L. *et al.* Coordination of the transcriptome and metabolome by the circadian clock. *Proc. Natl. Acad. Sci.* **109**, 5541–5546 (2012).
10. Kochanowski, K. *et al.* Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol. Syst. Biol.* **13**, 903 (2017).
11. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6**, 813–823 (2006).
12. Hensley, C. T. *et al.* Metabolic Heterogeneity in Human Lung Tumors. *Cell* **164**, 681–694 (2016).
13. Hu, J. *et al.* Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat. Biotechnol.* **31**, 522–529 (2013).
14. Jain, M. *et al.* Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation. *Science* **336**, 1040–1044 (2012).
15. Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal. Chem.* **83**, 7074–7080 (2011).
16. Paglia, G. & Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* **12**, 797–813 (2017).
17. Dietmair, S., Timmins, N. E., Gray, P. P., Nielsen, L. K. & Krömer, J. O. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Anal. Biochem.* **404**, 155–164 (2010).
18. Dettmer, K. *et al.* Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. *Anal. Bioanal. Chem.* **399**, 1127–1139 (2011).
19. Li, B. *et al.* Performance Evaluation and Online Realization of Data-driven Normalization Methods Used in LC/MS based Untargeted Metabolomics Analysis. *Sci. Rep.* **6**, srep38881 (2016).
20. Dubuis, S. *et al.* Metabotypes of breast cancer cell lines revealed by non-targeted metabolomics. *Metab. Eng.* **43**, 173–186 (2017).
21. Pfister, T. D. *et al.* Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. *Mol. Cancer Ther.* **8**, 1878–1884 (2009).

22. Guo, T. *et al.* Rapid proteotyping reveals cancer biology and drug response determinants. in revision. Data publicly available in ArrayExpress, Project accession: E-PROT-2
23. Cascante, M. *et al.* Metabolic control analysis in drug discovery and disease. *Nat. Biotechnol.* **20**, 243–249 (2002).
24. Reznik, E. *et al.* Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Rep.* **20**, 2666–2677 (2017).
25. Pirhaji, L. *et al.* Revealing disease-associated pathways by network integration of untargeted metabolomics. *Nat. Methods* **13**, 770–776 (2016).
26. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).
27. Bellis, A. D. *et al.* Cellular arrays for large-scale analysis of transcription factor activity. *Biotechnol. Bioeng.* **108**, 395–403 (2011).
28. Han, H. *et al.* TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* **5**, srep11432 (2015).
29. Liao, J. C. *et al.* Network component analysis: Reconstruction of regulatory signals in biological systems. *Proc. Natl. Acad. Sci.* **100**, 15522–15527 (2003).
30. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.* **23**, 27–47 (2016).
31. Sciacovelli, M., Gaude, E., Hilvo, M. & Frezza, C. Chapter One - The Metabolic Alterations of Cancer Cells. in *Methods in Enzymology* (eds. Galluzzi, L. & Kroemer, G.) **542**, 1–23 (Academic Press, 2014).
32. Morgan, M. J. & Liu, Z. Crosstalk of reactive oxygen species and NF- κ B signaling. *Cell Res.* **21**, 103–115 (2011).
33. Biegging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* **14**, 359–370 (2014).
34. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer* **9**, 749–758 (2009).
35. Kruiswijk, F., Labuschagne, C. F. & Vousden, K. H. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. *Nat. Rev. Mol. Cell Biol.* **16**, 393 (2015).
36. Semenza, G. L. HIF-1: upstream and downstream of cancer metabolism. *Curr. Opin. Genet. Dev.* **20**, 51–56 (2010).
37. Kim, J., Tchernyshyov, I., Semenza, G. L. & Dang, C. V. HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* **3**, 177–185 (2006).
38. Papandreou, I., Cairns, R. A., Fontana, L., Lim, A. L. & Denko, N. C. HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell Metab.* **3**, 187–197 (2006).
39. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
40. Gossage, L., Eisen, T. & Maher, E. R. VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer* **15**, nrc3844 (2014).
41. Petrella, B. L. & Brinckerhoff, C. E. PTEN suppression of YY1 induces HIF-2 α activity in von Hippel Lindau null renal cell carcinoma. *Cancer Biol. Ther.* **8**, 1389–1401 (2009).
42. Dalgin, G. S. *et al.* Identification of Novel Epigenetic Markers for Clear Cell Renal Cell Carcinoma. *J. Urol.* **180**, 1126–1130 (2008).
43. Chaudhri, V. K. *et al.* Metabolic alterations in lung cancer-associated fibroblasts correlated with increased glycolytic metabolism of the tumor. *Mol. Cancer Res. MCR* **11**, 579–592 (2013).

44. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
45. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
46. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19**, 619–625 (2013).
47. Cooks, T. *et al.* Mutant p53 Prolongs NF- κ B Activation and Promotes Chronic Inflammation and Inflammation-Associated Colorectal Cancer. *Cancer Cell* **23**, 634–646 (2013).
48. Höpker, K. *et al.* AATF/Che-1 acts as a phosphorylation-dependent molecular modulator to repress p53-driven apoptosis. *EMBO J.* **31**, 3961–3975 (2012).
49. Di Agostino, S. *et al.* Gain of function of mutant p53: The mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation. *Cancer Cell* **10**, 191–202 (2006).
50. Sciacovelli, M. *et al.* Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. *Nature* **537**, 544 (2016).
51. Sciacovelli, M. & Frezza, C. Oncometabolites: Unconventional triggers of oncogenic signalling cascades. *Free Radic. Biol. Med.* **100**, 175–181 (2016).
52. Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat. Biotechnol.* **31**, 357–361 (2013).
53. Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, aaf2786 (2016).
54. Sullivan, L. B., Gui, D. Y. & Heiden, M. G. V. Altered metabolite levels in cancer: implications for tumour biology and cancer therapy. *Nat. Rev. Cancer* **16**, nrc.2016.85 (2016).
55. De Smet, F., Christopoulos, A. & Carmeliet, P. Allosteric targeting of receptor tyrosine kinases. *Nat. Biotechnol.* **32**, 1113–1120 (2014).
56. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
57. Koehler, A. N. A complex task? Direct modulation of transcription factors with small molecules. *Curr. Opin. Chem. Biol.* **14**, 331–340 (2010).
58. Glunde, K., Bhujwala, Z. M. & Ronen, S. M. Choline metabolism in malignant transformation. *Nat. Rev. Cancer* **11**, 835 (2011).
59. Alam, M. T. *et al.* The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nat. Commun.* **8**, ncomms16018 (2017).
60. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
61. Chen, J. C. *et al.* Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks. *Cell* **159**, 402–414 (2014).
62. Heiden, M. G. V., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029–1033 (2009).
63. Mayers, J. R. & Vander Heiden, M. G. Famine versus feast: understanding the metabolism of tumors in vivo. *Trends Biochem. Sci.* **40**, 130–140 (2015).
64. DeBerardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism. *Sci. Adv.* **2**, e1600200 (2016).
65. Studer, R. A. *et al.* Evolution of protein phosphorylation across 18 fungal species. *Science* **354**, 229–232 (2016).
66. Gonçalves, E. *et al.* Systematic Analysis of Transcriptional and Post-transcriptional Regulation of Metabolism in Yeast. *PLoS Comput. Biol.* **13**, e1005297–e1005297 (2017).

67. Sodi, V. L. *et al.* Nutrient sensor O-GlcNAc transferase controls cancer lipid metabolism via SREBP-1 regulation. *Oncogene* (2017). doi:10.1038/onc.2017.395
68. Illendula, A. *et al.* A small-molecule inhibitor of the aberrant transcription factor CBF β -SMMHC delays leukemia in mice. *Science* **347**, 779–784 (2015).
69. Moellering, R. E. *et al.* Direct inhibition of the NOTCH transcription factor complex. *Nature* **462**, 182–188 (2009).
70. Gonda, T. J. & Ramsay, R. G. Directly targeting transcriptional dysregulation in cancer. *Nat. Rev. Cancer* **15**, 686–694 (2015).
71. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles. *bioRxiv* 136168 (2017). doi:10.1101/136168
72. Zimmermann, M., Sauer, U. & Zamboni, N. Quantification and Mass Isotopomer Profiling of α -Keto Acids in Central Carbon Metabolism. *Anal. Chem.* **86**, 3232–3237 (2014).
73. Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419–425 (2013).
74. Dolfi, S. C. *et al.* The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. *Cancer Metab.* **1**, 20 (2013).
75. Noor, A., Ahmad, A. & Serpedin, E. SparseNCA: Sparse Network Component Analysis for Recovering Transcription Factor Activities with Incomplete Prior Information. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **PP**, 1–1 (2015).
76. Storey, J. D. A Direct Approach to False Discovery Rates. *J. R. Stat. Soc. Ser. B* **64**, 479–498 (2002).
77. Konig, R. *et al.* A probability-based approach for the analysis of large-scale RNAi screens. *Nat Meth* **4**, 847–849 (2007).
78. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).

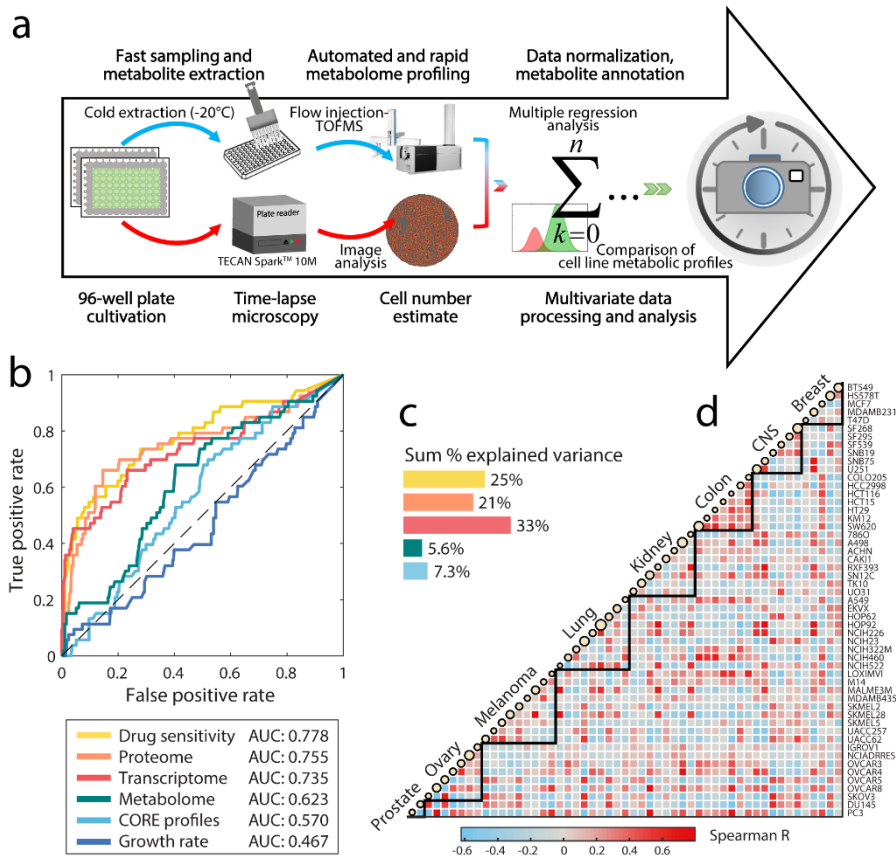


Figure 1. Comparative metabolome profiling in 54 adherent cancer cell lines. (a) Schematic overview of the combined workflow for high-throughput metabolome profiling in adherent cell lines. Multiple cell lines are cultivated in parallel to collect cell extracts for MS-based metabolome profiling (Supplementary Figure S1). A new software tool for the segmentation of bright-field microscopy images (Supplementary Note) is used for automated cell number quantification. Raw MS ion intensities are normalized using a multiple regression scheme (Supplementary Figure S2). (b) Signatures of tissue type in intracellular

metabolome profiles compared with transcriptome, proteome, drug sensitivity, growth rate and the uptake and secretion of 140 metabolites (CORE profiles¹⁴). ROC curve analysis quantifies the performance of each of the data layers in predicting the tissue of origin (Supplementary Figure S3). (c) Major patterns in each data set were extracted by principal component analysis and tested individually for tissue signatures (Supplementary Figure S3). Colored bars indicate the sum of explained variance by tissue-associated principal components (AUC ≥ 0.75). (d) Pairwise similarity (Spearman correlation) of metabolome profiles among 53 cell lines from the NCI-60 panel. The size of circles on the diagonal is proportional to the growth rate (doubling time) of each cell line.

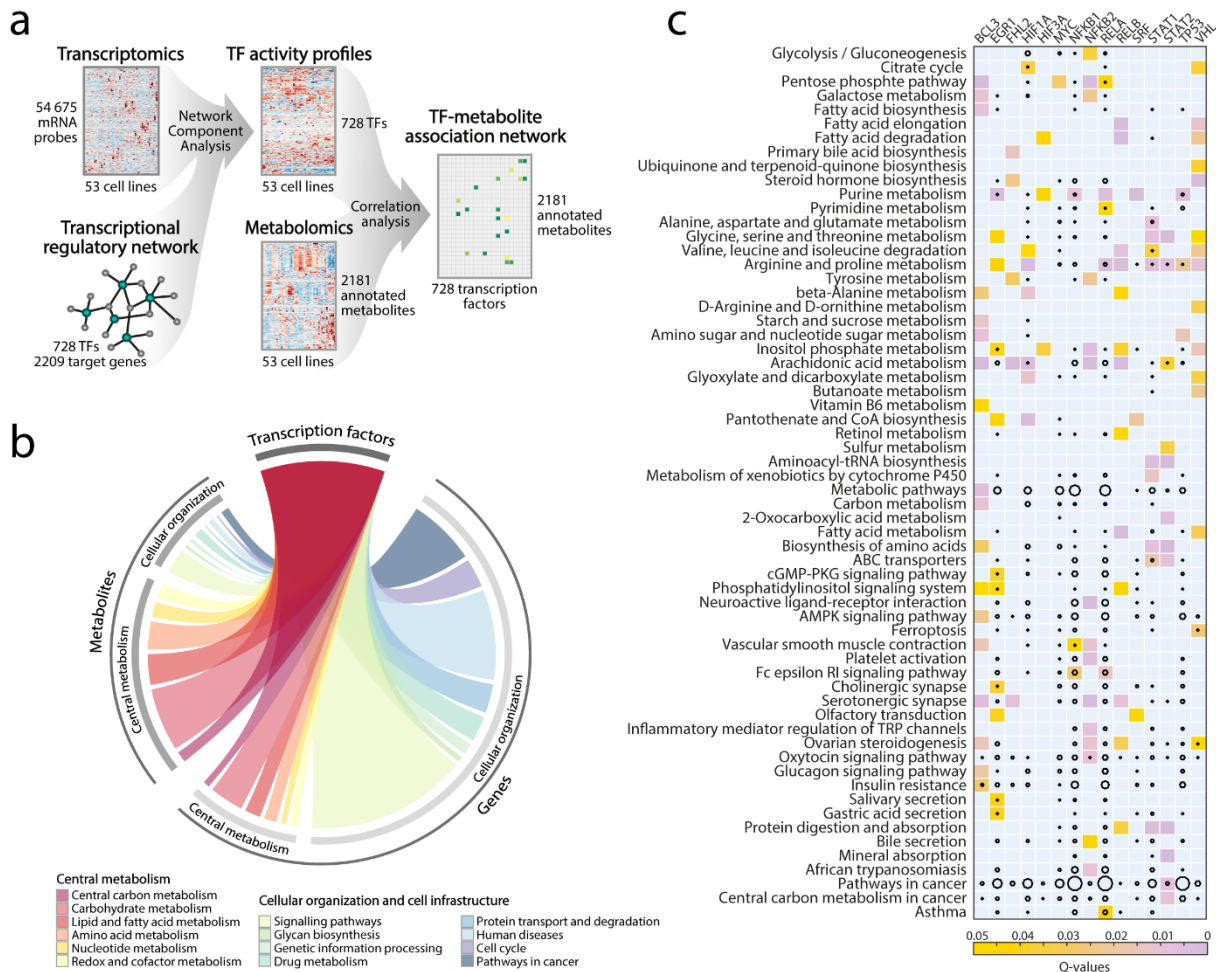


Figure 2. Inferring a TF-metabolite association network by integrating transcriptome and metabolome data. **(a)** Schematic overview of the computational framework for finding TF-metabolite associations. The activity of 728 human TFs was calculated using Network Component Analysis (NCA)²⁹ which estimates TF activity from gene expression profiles of the NCI-60 cell lines and the transcriptional regulatory network between TFs and their target genes. In the second step, correlation analysis was used to find associations between TFs and metabolites (Figure 1). The resulting TF-metabolite association network resembles a scale-free network (Supplementary Figure S4). **(b)** Distribution of links to KEGG pathways in the TF-metabolite association network (left-hand section), and in the TF-gene regulatory network (right-hand section), respectively. KEGG pathways were grouped into super-pathways. Edge-size connecting the TF hub and a super-pathway reflects the number of links found in the network. **(c)** Metabolites associated with 15 known oncogenic TFs were tested for a significant overrepresentation in KEGG metabolic pathways (Supplementary Figure S5). Colored squares indicate the significance of the enrichment analysis (q -value < 0.05), while the size of black circles scales with the number of known TF-gene targets in the metabolic pathway.

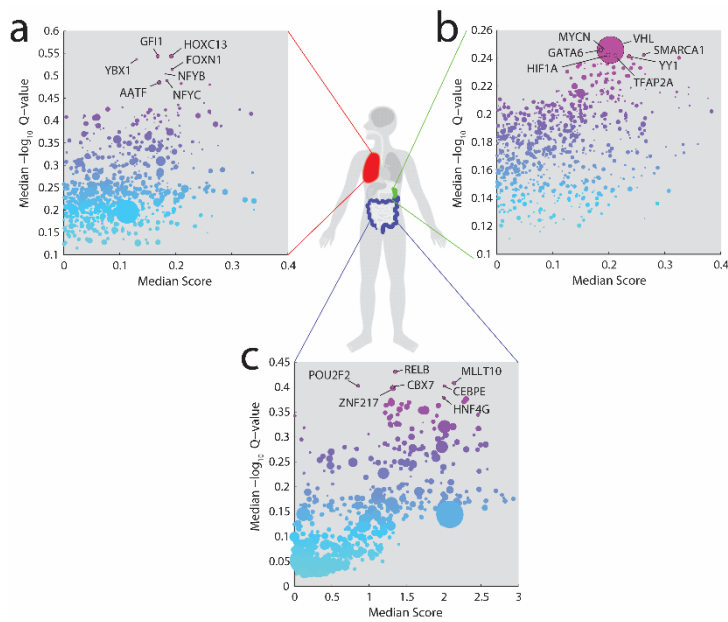


Figure 3. Inferring TF as mediators of *in vivo* metabolic changes. Prediction results in three previously published data sets comprising metabolic profiles of tumor vs. proximal healthy tissue from 21, 138 and 10 patients with lung cancer⁴³ (panel a), clear-cell renal cell carcinoma⁷ (panel b), and colon cancer¹³ (panel c). Each dot represents the strength (x-axis, median score) and significance (y-axis, Q-value) of a TF in mediating *in vivo* metabolic changes. Dot-size indicates the frequency of mutations in the TF-encoding gene for the respective cancer type, derived from the Catalogue Of Somatic Mutations In Cancer (COSMIC)⁷⁸. Gene-name labels are shown for the top 1% most significant hits (see Supplementary Discussion and Supplementary Figure S6).

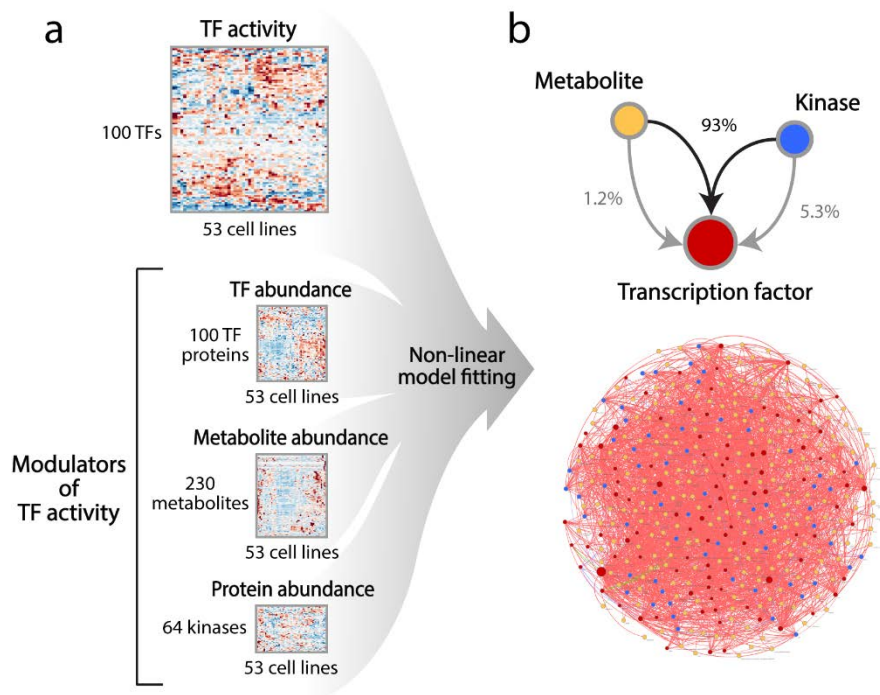


Figure 4. Modeling TF activity using an ensemble of non-linear models. (a) Schematic overview of the computational framework. Three layers of biological information are integrated using a model-based fitting approach: protein abundance of 100 TFs and 64 kinases, TFs activity derived from transcriptome data and relative levels of 230 metabolites across the 53 cell lines (see Online methods). (b) Predicted network of modulatory interactions between TFs (red nodes), metabolites (yellow nodes) and kinases (blue nodes). Most interactions involve the combined action of a metabolite and a kinase (93% of interactions). In 22 cases, changes in metabolite abundance alone were sufficient to improve the description of changes in TF activity (Supplementary Data Table S5).