nature research

https://doi.org/10.1038/s41588-020-0703-5

Supplementary information

Discovering functional evolutionary dependencies in human cancers

In the format provided by the authors and unedited

SUPPLEMENTARY NOTE

DATA AVAILABILITY AND PROCESSING

The Cancer Genome Atlas (TCGA) Pan-Cancer Atlas dataset

In this study, we primarily used the latest publicly available data freeze from the TCGA consortium, also referred to as TCGA Pan-Cancer Atlas cohort, encompassing 9125 samples from 32 different tumor types for which all molecular profiling and sample annotation data were available[1]. Data were downloaded from:

- https://gdc.cancer.gov/about-data/publications/pancanatlas
- https://xenabrowser.net/datapages/?cohort=TCGA%20Pan-Cancer%20(PANCAN)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443.

Each type of molecular and clinical data is described separately in this section.

Sample annotation and clinical data

We used previously defined tumor sample annotations[1]. Tumor type (organ of origin) and subtype was available for 9125 samples; tumor type classes and acronyms are defined as in OncoTree (http://oncotree.mskcc.org/#/home). In this work, we manually validated tumor annotations and found some minor inconsistencies in the Microsatellite Instability (MSI) and POLE subtype annotations. We found 4 UCEC samples annotated as MSI but showing low mutation burden (data not shown). Additionally, we found 29 samples annotated as POLE, but without any POLE hotspot mutation (P286X and V411). These samples were ignored in the downstream analyses. The final sample annotation is available in **Supplementary Table 1**. Harmonized clinical data[2] was used in all survival analyses.

TCGA somatic mutation data

Somatic point mutation data curated in the MC3 initiative[3] were used in this work (maf file release 2.8, with LAML patch). We preferred the MC3 version over the original mutation calls separately curated by the TCGA within each single tumor type study, as it guarantees the same standard of data processing, mutation calling, and annotation for each tumor type.

TCGA Copy Number Alterations (CNA) data

We analyzed CNA with GISTIC[4] using the latest array CGH data (segmentation file) generated by TCGA and available at https://gdc.cancer.gov/about-data/publications/pancanatlas. GISTIC was run separately on each single tumor type and once on the pan-cancer cohort as previously

described[1]. The output of these analyses was integrated as described in the section "Copy Number Alterations Catalogue". GISTIC was run with a confidence threshold parameter to 0.95.

TCGA gene fusion and gene expression data

We downloaded detected gene fusions and RNA-seq gene expression data used in the TCGA Pan-Cancer Atlas project at https://gdc.cancer.gov/about-data/publications/pancanatlas.

TCGA inferred tumor phylogenies

Phylogenies inferred from single TCGA samples were retrieved[5]. Reconstructed phylogenies (500 per sample) of tumor clones were available for ~6000 TCGA samples. Each phylogeny was annotated with a confidence score.

TCGA mutational signatures

Mutational signatures for each TCGA sample were derived using the R package 'deconstructSigs' version 1.8[6], using as input the MC3 maf file and the signature set from[7]. The tool was used with default parameters and 'exome2genome' normalization.

MSK-IMPACT dataset

Tumor molecular and clinical data collected and generated by Memorial Sloan Kettering Cancer Center using the MSK-IMPACT platform[8] was downloaded from the cBioPortal[9] http://www.cbioportal.org/. Molecular data were processed as done for the TCGA cohort, retaining only functional events defined in our study.

TRACERx and REVOLVER datasets

We used tumor phylogenies inferred by REVOLVER[10] on multiregional sequencing data of breast and lung cancers generated by the TRACERx initiative. For each lung or breast cancer patient, REVOLVER infers a phylogeny of tumor clones and the set of mutations found within each clone. Trajectories from the TRACERx/REVOLVER dataset were extracted from the Supplementary Data of the corresponding manuscript[10] and processed as explained below.

Cancer Cell Line Encyclopedia (CCLE) dataset

The Broad CCLE[11,12] currently encompasses 1457 cell lines profiled for exonic point mutations, copy number variants, DNA methylation, and gene expression. We used the CCLE molecular profiles in all drug and gene dependence enrichments performed in this work, unless otherwise specified. Data were downloaded from the CCLE portal (https://portals.broadinstitute.org/ccle/about) [January 2019 release]. A detailed and description of the available data is available in[12].

Cancer Cell Line Project (CCLP) dataset

CCLP profiled more than 1000 human cancer cell lines for point mutation, copy number alterations and gene expression[13]. Data were downloaded from the Cosmic Data Portal (http://cancer.sanger.ac.uk/cell_lines/, data freeze v81).

Cell Model Passports (CMP) dataset

Similarly to CCLP, Cell Model Passports (https://cellmodelpassports.sanger.ac.uk/downloads) is an initiative to profile hundreds of human cancer cell lines. We downloaded the release of September 2018, encompassing mutation, copy number and gene expression profiles for 1270 cancer cell lines.

Genomics of Drug Sensitivity in Cancer (GDSC) dataset

The GDSC effort screened a panel of 1074 human cancer cell lines for resistance to 265 compounds[13]. Predicted IC50 values are used as sensitivity measure. We downloaded the drug resistance data from the GDSC portal (http://www.cancerrxgene.org/ , data release version 6, drug resistance release version 17).

Cancer Therapeutics Response Portal (CTRP) dataset

CTRP is a resource quantifying the sensitivity of 860 human cancer cell lines to 481 compounds[14]. Predicted EC50 values and AUC are used as sensitivity measure. We used the data release version 2.1 downloaded from
https://portals.broadinstitute.org/ctrp/?page=#ctd2BodyHome.

Dependency Map (DepMap) gene essentiality screenings

The cancer DepMap portal (https://depmap.org/portal/) integrates results from genome-wide gene essentiality screenings performed on hundreds of human cancer cell lines. Two major studies are available in the portal: CERES and DEMETER2.

CERES dataset

CERES[15] is a computational methodology to estimate gene dependency levels from CRISPR-Cas9 essentiality screenings. CERES assigns the dependency score on gene *y* by taking into account the effects of copy number changes in the genomic region targeted by the sgRNAs used in the CRISPR-Cas9 to knock-out gene *y*. The dependency score, called AVANA, is a real value distributed around 0, with negative values indicating increased dependency. In this work, we used the version 18Q3_v3 of the CERES dataset. AVANA scores were available for more than 17'000 genes in 485 cancer cell lines (with annotation and genomic data available).

DEMETER2 dataset

DEMETER2[16] is a computational framework for analyzing RNAi screens and quantify gene dependency across multiple screenings and samples. Similarly, to AVANA, negative values indicate increased dependency. We downloaded the DEMETER2 dataset version v2_6025238. Scores were available for more than 17'000 genes in 712 cancer cell lines.

DRIVE dataset

DRIVE is a gene essentiality screening performed by Novartis using RNAi techniques[17]. In their study, McDonald et al. tested the vulnerability of 398 cancer cell lines to 7837 genes, using on average 20 shRNAs per gene. Dependencies are quantified by means of two related measures, called RSA and Ataris in the original publication. Processed data was obtained upon private request to the contact author (version 4, May 2019).

SCORE dataset

SCORE is a CRISPR-Cas9 based screening[18] encompassing 324 human cancer cell lines and 17'995 genes. The initial release (version 1) was downloaded from https://score.depmap.sanger.ac.uk. Similarly, to the other dependency scores, a negative 'loss of fitness score' represents an increased dependency.

Matching cell line molecular data with gene essentiality and drug screening experiments

In this study, we match molecular profiles of cell line dataset with screening datasets based on maximum overlap of studied cell lines. Precisely, we used CCLE molecular profiles in conjunction with the AVANA, DEMETER2, DRIVE, and CTRP datasets; CCLP molecular profiles with the GDSC dataset; and CMP molecular profiles with the SCORE dataset.

SELECT INPUT PRE-PROCESSING AND PARAMETERS

SELECT requires the following primary input data:

- A binary Genomic Alteration Matrix (GAM), encoding the occurrence of each alteration in each sample (1: altered, 0: wild type);
- A vector of sample annotations, sample categorical covariates are expressed through this annotation vector (e.g. tumor types and subtypes);
- (optional) A vector of alteration annotations, alteration categorical covariates are expressed through this annotation vector (e.g. copy number alterations vs. somatic mutations);

The pre- and post-processing steps used in this work to generate SELECT input data and filter its output are described in the following paragraphs.

Preprocessing of alterations and sample annotations

To generate the final GAM, we first merged copy number alteration events that were occurring within the same chromosome and sharing more than 80% (overlap > 0.8) of alteration occurrences. Precisely, the overlap between the occurrences of two CNAs was calculated as the number of samples with both alterations divided by the maximum number of possible overlapping samples, i.e. the minimum occurrence count between the two events. This step is important to reduce the number of dependencies involving adjacent copy number events. This process reduced the number of deletion and amplification events from 147 and 123 to 124 and 104, respectively.

Sample tumor type annotations were modified to merge colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) into a single colorectal cancer class (CRC), as done in the previous works. The subtype classification for the GBM samples was ignored as several GBM samples were missing the annotation. In the end, only 9083 tumor samples belonging to tumor classes with at least 10 samples were retained. Similarly, only alterations with at least 10 occurrences were considered. Finally, all the functional alteration occurrences were condensed into a binary genomic alteration matrix (GAM) spanning 9083 samples and 659 functional events (data available at http://ciriellolab.org/select/select.html).

Results described in **Fig. 1** were obtained upon running SELECT on 3 different GAMs:

1. GAM of putative Functional mutations (F – **Fig. 1g**). This matrix is a subset of GAM described in the previous sections which include as events (rows) exclusive recurrently mutated genes (n = 399).
2. GAM of all mutations (F+N – **Fig. 1g**). This GAM included exclusively recurrently mutated genes (n = 399), but alteration occurrences were called from all non-synonymous mutations observed in these genes. Alteration occurrences in each gene were subsampled to match its alteration frequency in the F GAM. Alteration subsampling was repeated multiple times using different seeds.
3. GAM of synonymous mutations (S – **Fig. 1g**). This GAM included exclusively recurrently mutated genes (n = 399), but alteration occurrences were called from all synonymous mutations observed in these genes ('synonymous_variant' or 'Silent').

Vetoed interactions

As done in our previous work[19], we did not test for mutual exclusivity and co-occurrence copy number alterations within the same chromosome. Additionally, we ignored fusions and amplifications within the same chromosome, as fusion events tend to be reflected also in copy number changes of the altered locus.

SELECT parameters

The number of random matrices to be generated for the estimation of the null model was set to 5000. In order to keep the tumor type heterogeneity into account, SELECT rewires the GAM in a block-wise fashion by splitting the GAM according to sample and alteration annotations provided as annotation vectors. In this work, the tumor subtype and the alteration type (i.e. mutations, amplifications, deletions and fusions) were used as sample and alteration classes, respectively. In general, SELECT uses the edge-flipping technique (implemented in the R BiRewire package) to rewire each block of the GAM preserving the occurrence frequency of each alteration and the number of alterations in each sample. SELECT reverts to a simple random shuffling (instead of using the edge flipping procedure) for the blocks where the number of alterations is less than 20. This is critical to guarantee an effective randomization in setting where a low number of alterations or edges would result in a lot of 'no switch' in the edge flipping process, keeping the randomized matrix too similar to the original one. To improve the stability of the results and their invariance to the seed used in the generation of the null model, we ran SELECT 10 times (using 10 different random seeds) and took the median SELECT score for each pairwise interaction.

SELECT implementation

We improved the original implementation of SELECT in order to reduce execution time and memory requirements. The updated version (SELECT v1.6) is available at http://ciriellolab.org/select/select.html, together with the data generated in this study. The development version of SELECT is available in the Git repository https://bitbucket.org/cso_repo/select/.

EVOLUTIONARY DEPENDENCIES IN THE TCGA COHORT

Recently, it was reported that heterogeneous mutation loads often lead to significant mutual exclusivity and co-occurrence[20]. Moreover, these patterns could be influenced by heterogeneous mutational signatures underlying the emergence (rather than selection) of specific variants[21]. Formally, we tested the existence of statistically significant associations between the number of co-occurrence (CO) and mutual exclusivity (ME) EDs involving a given alteration event *x* and the following properties of *x*: its occurrence frequency ($freq_{SFE}$), its enrichment in the MSI subtype ($MSI_{prevalence}$), the median mutational burden (n_{MUT}) and chromosomal instability (n_{CNA}, i.e. number of CNA segments) of samples in which *x* occurs. $MSI_{prevalence}$ was defined as the fraction of occurrences of *x* in MSI samples over the total number of occurrences of *x* across the entire TCGA pan-cancer cohort. Statistical significance was derived with a multivariate type-II ANOVA analysis on the following linear models:

$$\#CO \sim MSI_{prevalence} + freq_{SFE} + n_{CNA} + n_{MUT}$$
$$\#ME \sim MSI_{prevalence} + freq_{SFE} + n_{CNA} + n_{MUT}$$

with $freq_{SFE}$, n_{MUT} and n_{CNA} in log10 scale. Results are available in **Supplementary Table 5**.

Alteration frequency was a significant predictor in all tests, consistent with the fact that for rare alterations the statistical power is diminished (**Supplementary Table 5**). Interestingly, although n_{MUT} and MSI status were correlated features, only MSI status remained significant when both covariates were tested together (**Supplementary Table 5**). Indeed, MSI tumors not only exhibited high mutation burdens (between 10 to 100 mutations per megabase), but they were also depleted for common cancer mutations (e.g. in *KRAS* or *TP53*) and enriched for mutations rarely found in other cancer types[22,23]. These features led SELECT to occasionally overestimate co-occurrence among events in MSI samples. Hence, in the rest of the analyses, we disregarded CO patterns that were exclusively observed within MSI subtypes. Regarding the influence of heterogeneous mutational signatures on evolutionary dependencies, we found that ME alterations were associated with mutational signatures that were as similar as those associated with CO alterations (**Extended Data Fig. 5b** – solid lines), and their similarity distributions was almost coincident with those obtained upon data randomization (**Extended Data Fig. 5b** – dashed lines). Hence, significant ME and CO patterns identified by SELECT did not emerge as a result of heterogeneous mutational burdens or signatures. Lastly, we wondered what is the likelihood that co-occurring mutations in a given patient are actually harbored in the same cells (*clonal co-occurrence*) or in distinct subclonal populations within a tumor (*subclonal co-occurrence*). Using previously estimated phylogeny distributions for the TCGA cohort[5], we systematically estimated the clonal co-occurrence probability for each patient harboring two evolutionary dependent alterations. We stress that the inference of tumor clonal phylogenies from single-sample whole exome sequencing is underpowered and incorrect predictions are expected. For this reason, we relied on distributions of phylogenies, rather than single predictions, and searched for consistent trends among multiple samples. Our results predicted that clonal co-occurrence was overwhelmingly more likely than subclonal co-occurrence (**Extended Data Fig. 5c**). In particular, although we found instances of subclonal co-occurrence, none of EDs exhibited a consistent trend for this phenomenon (**Extended Data Fig. 5d**).

Estimation of clonal and subclonal co-occurrence

Given an ED (x,y) between two alterations x and y, we checked their co-clonality status in the TCGA samples harboring both alterations. We explored this aspect using tumor phylogenies that we have previously inferred for the TCGA cohort in an independent study[5]. In that study, we inferred for each sample 500 possible phylogenies, each one scored by their likelihood. For each patient i, we checked whether the two alterations x and y were assigned to the same or different clones in the 500 reconstructed phylogenies. For each phylogeny we recorded whether (i) x and y are in the same clone, (ii) x and y are in different clones but along the same lineage, or (iii) x and y are in independent lineages. We then determined the fraction of the 500 phylogenies that predicted x and y exist within the same clone or lineage and computed the probability of clonal co-occurrence as the sum of the likelihoods of these phylogenies divided by the total sum of likelihoods. The complement of this probability is the probability of subclonal co-occurrence. Next, we scored the ED (x,y) by the mean clonal co-occurrence probability across all patients harboring both x and y.

Evolutionary Axes and Trajectories

As discussed in the main text, we built evolutionary axes starting from single EDs inferred by SELECT. Then, we annotated each evolutionary axis with preferential trajectories, that is which alterations tend to occur earlier or later than the other inside a given axis.

Inference of evolutionary axes

Evolutionary axes of a given sample cohort of interest C were inferred by assembling EDs that were either significant in the tumor type-specific analysis for S or that were significant in the pan-cancer analysis and classified as representative of C based on the set of filtering criteria here described.

1. First, we require both X and Y to occur in a fraction of samples *n* greater than F (F depends on the tumor type – see **Supplementary Table 7**) or 10 samples, whichever is the greatest. This filter ensured that both alterations are recurrent in C.

2. Second, we evaluate whether a given ED between X and Y is frequently "satisfied" in C. The frequency of a given ED corresponds to the number of samples that satisfies the ED: if X and Y were found significantly mutually exclusive, the frequency of the ED between X and Y corresponds to the fraction of samples that exhibit either X or Y, but not both; if X and Y were significantly co-occurrent the frequency of the ED corresponds to the fraction of samples that exhibit both X and Y. The observed frequency of a given ED in C (f_C) is then compared to the its expected value (f_{EXP}), which for co-occurrence EDs corresponds to the mean random overlap between X and Y in the randomized GAMs used by SELECT to compute the null model, for mutually exclusive EDs it is 1 minus such value. The ED passes this second filter if:

$$S = \frac{f_C - f_{EXP}}{f_C + f_{EXP}} > 0.$$

This filter ensures that X and Y exhibit the same trend towards mutual exclusivity or co-occurrence that was found significant in the pan-cancer cohort.

3. Lastly, we compare f_C to the corresponding value observed in the pan-cancer cohort f_{PC}. The ED passes this final filter if:

$$\Phi = \frac{f_C - f_{PC}}{f_C + f_{PC}} > 0 \text{ or } f_C > F.$$

This filter ensures that the fraction of samples that satisfies the EDs is either greater than the corresponding fraction in the pan-cancer cohort or greater than a given threshold.

Evolutionary axes in **Extended Data Fig. 9** were manually curated upon intersecting results from individual tumor types.

Inference of evolutionary trajectories

Evolutionary trajectories in **Fig. 4** were extracted from (i) the tumor phylogenies inferred by REVOLVER for the TRACERx lung cancer cohort and available from the corresponding manuscript[10] and (ii) TCGA tumor phylogenies[5]. For each patient, both resources provide a (set of) phylogenies where the nodes represent clones. Each clone is associated to the list of alterations that first emerged in such clone and were not present in any ancestor clone. Given a cancer patient and its reconstructed phylogeny, an alteration x_1 was annotated as 'preceding' x_2 in such patient if x_1 and x_2 were assigned to two clones c_1 and c_2, with c_1 being an ancestor of c_2 in the reconstructed tumor phylogeny. We applied this criterion to all the patients from the two datasets, extracting the trajectory for each pair of co-occurring alterations (x_1, x_2). Each patient was counted as additional evidence for a preferential trajectory x_1 -> x_2 or x_2 -> x_1.

STATISTICAL FRAMEWORK FOR ASSOCIATION STUDIES

We interrogated different functional knockout (KO) and drug sensitivity screenings to assess the impact of combinations of genomic events on gene essentiality and drug resistance. In general, let $X = \{x_1, \dots, x_n\}$ be a set of alterations x_i of interest, and y a phenotype of interest. Let the alteration occurrence profile X_k be a binary vector indicating the occurrence or absence of each alteration x_i in sample k, and let y_k be a real value representing phenotype y in sample k.

We need a way to statistically assess whether cell lines with a specific *X'* occurrence profile have a phenotype different from cell lines with another occurrence profile *X''*. Here we describe the statistical frameworks we designed to study such associations. We explored two distinct statistical approaches, one based on frequentist inference and the other on Bayesian statistics. Below, the two statistical frameworks are first presented for the simpler case of $X = \{x_1\}$, that is when *X* contains only one genomic alteration. Then, the frameworks are generalized to study the impact of multiple genomic alterations, $X = \{x_1, \dots, x_n\}$, with $n > 1$.

The simpler univariate case with $n = 1$ ($X = \{x_1\}$) is used, for instance, when studying the impact of single alterations on cell fitness (**Fig. 2**). The multivariate case ($n > 1$) is suited to study the associations between pairs of evolutionary dependent events and gene essentiality (**Fig. 3**), or between evolutionary axes and drug resistance (**Fig. 4**).

Frequentist Inference Framework

Association between a single alteration and a phenotype

Let X = {x}, where x is a single alteration of interest. X can have two possible statuses in a given sample k: $X_k = [0]$ or $X_k = [1]$. We model the association between X and a given phenotype *y* (real valued) with a linear model considering y as the dependent variable, and the status of alteration *X* (binary value) and the dataset covariates (e.g. tumor type) as independent variables:

$$y \sim mu + cov_1 + \cdots + cov_j + X$$

where *mu* represents the intercept of the model. In this framework, we use a two-way ANOVA to assess the statistical relevance of the association between *X* and *y,* taking into account the effect of the covariates. We used the R implementation of ANOVA (function 'aov'). The resulting P values were used to assess the significance of the association.

Extension to the study of multiple alterations

Let $X = \{x_1, \ldots x_n\}$, with $n > 1$ be our set of alterations of interest. The ANOVA framework can easily take into account the effect of multiple alterations, as the X independent variable in the linear model $Y \sim cov_1 + \ldots cov_n + X$ is not constrained to be binary. We simply need to allocate enough symbols to represent all the possible occurrence profiles. For example, the association of an evolutionary dependence between two alterations x_1 and x_2 and a phenotype y can be studied by setting $X = \{x_1, x_2\}$ and representing the 4 possible alteration profiles with four distinct symbols: $X_{dd} = [1, 1]$ representing double-altered cell lines (dd), $X_{dw} = [1, 0]$ for cell lines with only x1 alteration (dw), $X_{wd} = [0, 1]$ for cell lines with only x2 alteration (wd), and $X_{ww} = [0, 0]$ for cell lines without any alteration (ww) (**Extended Data Fig. 6b**). In the worst case, when all the occurrence profiles are present in at least one sample, there would be $s = 2^n$ distinct symbols. Given the exponential growth of the number of symbols, only a limited number n of alterations can be studied at the same time, depending on the number of available samples.

Post-hoc analysis

The multivariate analysis described above informs on whether at least one specific alteration profile X' is significantly different from the others. Sometimes, however, we are interested in assessing whether two specific alteration profiles X' and X'' have significantly different phenotypes. This is typically accomplished with follow-up post-hoc tests. In the case of pairs of evolutionary dependent alterations $X = \{x_1, x_2\}$ we are usually interested in checking whether there is a phenotype difference between dd and dw samples (i.e. between profiles X = [1,1] and X = [1,0]). Note that this corresponds to checking whether double-altered samples have a different phenotype compared to the samples with only one of the two alterations. In the frequentist framework, we use the post-hoc Tukey test to assess significant differences in phenotype between the alteration occurrence patterns in a pairwise fashion. We used the implementation of the Tukey test from the R package 'multcomp'.

Unless otherwise indicated, P values were corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure to estimate the False Discovery Rate (FDR).

<u>Bayesian Inference Framework</u>

We developed a framework based on Bayes Factors[24,25] to test the association between the occurrence of an alteration profile *X* and a phenotype Y. Generally speaking, Bayes Factors are a way of quantifying the strength of evidence (provided by some observed data) in favor of a model over another. A Bayes Factor is a positive real number representing the ratio between the

likelihood that some observed data D are sampled from a model M_1 rather than from another model M_0. More formally, the Bayes factor (BF) is the ratio of the posterior probabilities of observing some data D given the two models (hypotheses) M_1 and M_0:

$$BF(M_1, M_0) = BF_{10} = \frac{P(D \mid M_1)}{P(D \mid M_0)}$$

The Bayes Factor will converge towards 1 when the observed data have similar likelihoods (posterior probability) under the two given models. If one of the two models has a sensibly greater likelihood, the BF will diverge from 1, indicating which model is preferred. Bayes Factors are not simple likelihood-ratio tests. Indeed, the parameters of the models are not defined as exact values, but rather as unknown values sampled from distributions (called priors) over the parameter space. Likelihoods are calculated by integrating over the space of parameters, intrinsically penalizing models including too many parameters or very broad priors and guarding against overfitting.

Association between a single alteration and a phenotype

As in the ANOVA-based approach, we first describe the case X = {*x*}, where *x* is a single alteration of interest. In the Bayesian approach, we wish to contrast the null model M_0 that assumes alteration *x* has no effect on phenotype *y* against the alternative model M_1 that considers the effect of alteration *x*. A Bayes Factor BF_{10} greater than a given threshold would then support the existence of an association between X = {*x*} and *y*.

We define the two alternative models M_0 and M_1, following previously proposed guidelines[26]. Formally, we define M_0 and M_1 as:

$$M_0: y \sim \mu + \varepsilon$$
$$M_1: y \sim \mu + c_{x=0} + c_{x=1} + \varepsilon$$

with:

- μ being the grand mean (intuitively, the intercept)
- $c_{x=0}$ and $c_{x=1}$ are the effect sizes of for the absence (x=0) or occurrence (x=1) of alteration *x*
- ε is a zero-centered noise term.

What distinguishes the two models, in other words, is the selection of the priors for the various c_x, which are fixed to 0 in M_0. It is convenient to reparametrize the models as:

$$M_1: y \sim \mu + \sigma(d_{x=0} + d_{x=1}) + \varepsilon$$

where

- σ is the standard deviation of the model
- $d_{x=0} = {}^{c_{x=0}}/_{\sigma}$
- $d_{x=1} = {}^{c_{x=1}}/_{\sigma}$

d_x corresponds to the standardized effect size of alteration x (the effect size divided by the standard deviation sigma). Within this parameter space, we can now formulate the priors on the effect sizes as recommended[27], using the following combination commonly known as the Jeffreys–Zellner–Siow (JZS) priors:

- Jeffrey Priors (for both M_0 and M_1):
  - μ ~ the grand mean
  - ε ~ Normal(0, σ^2)
  - σ ~ flat non-informative prior on $\log(\sigma^2)$

- Zellner and Siow derived g-priors (only relevant for model M_1):
  - $d_{x=0} \sim Normal(0,\ \sigma^2_{x=0} = g_{x=0})$
  - $d_{x=1} \sim Normal(0,\ \sigma^2_{x=1} = g_{x=1})$
    with the variances of the effect sizes g_x defined as
  - $g_{x=0} \sim Inverse\ Chi\ Squared(1, h^2_{x=0})$
  - $g_{x=1} \sim Inverse\ Chi\ Squared(1, h^2_{x=1})$

The hyper-parameter h^2_x has to be fixed before the analysis and it influences the 'amplitude' of the prior distribution for the effect size parameters $d_{x=0}$ and $d_{x=1}$. A detailed discussion on the nature of h^2_x and how to set it properly is available in [27].

Implementation details

The R package BayesFactor (https://cran.r-project.org/web/packages/BayesFactor/) supports the calculation of Bayes Factors with the priors described above. We used its function 'lmBF' to compute the Bayes Factors $BF(M_1, M_0)$ between the two models M_0 and M_1. To estimate the likelihood over all the parameter space, the BayesFactor package integrates over the $g_{x=0}$ and $g_{x=1}$ parameters through Monte Carlo sampling. The integration of all the other parameters, instead, has a closed-form solution[28]. We use the same h^2 value for both $h^2_{x=0}$ and $h^2_{x=1}$, running the lmBF function with the following parameters: rscaleFixed = medium, rscaleRandom = nuisance, rscaleCont = medium. The number of iterations used to estimate the Bayes factor was set to 1000.

We opted for relying on the R package bayesFactor as it derives the posterior distributions for some of the model parameters by using a closed-form solution, hence being considerably faster than other solutions (e.g. pyMC3) on our model and specific selection of priors. Given the significant amount of simulations to be performed to generate the random datasets, estimate null distributions and perform the Bayesian inference, all the analyses were parallelized on three workstations equipped with 22 or 44 physical Intel Xeon cores and 256 or 512 GB of RAM each. All the code was implemented in R (bitbucket.org/cso_repo/eda).

Controlling for the effects of covariates

As in the ANOVA analysis, we need to take into account the effect of the dataset covariates (e.g. tumor type). Bayes Factors can be used to compare multiple models[27]. Given two Bayes Factors $BF(M_1, M_0)$ and $BF(M_2, M_0)$ with the same model M_0 at the denominator, we can calculate:

$$BF(M_2, M_1) = \frac{BF(M_2, M_0)}{BF(M_1, M_0)}$$

representing the evidence in favor of M_2 compared to M_1.

We exploit this property to compare the model M_{cov} that takes into account all the covariates (but no *X* alteration status) against the 'full' model M_{x+cov} considering both covariates and *x* alteration status. Formally, we define the models as:

$$M_{cov}: y \sim \mu + c_{covariates} + \varepsilon$$
$$M_{x+cov}: y \sim \mu + c_{covariates} + c_{x=0} + c_{x=1} + \varepsilon$$

It follows that:

$$BF(M_{x+cov}, M_{cov}) = \frac{BF(M_{x+cov}, M_0)}{BF(M_{cov}, M_0)}$$

With M_0 being the same model defined above. The two $BF(M_{x+cov}, M_0)$ and $BF(M_{cov}, M_0)$ are calculated as described in the previous section, using for all the new covariates the same priors used for the effect of x, that is the Zellner and Siow g-priors defined above:

- $d_{cov_i} \sim Normal\left(0, \sigma^2_{cov_i} = g_{cov_i}\right) \forall\ covariate\ i$
- $g_{cov_i} \sim Inverse\ Chi\ Squared(1, h^2)$

The same hyper-parameter h^2 used for g_x is imposed here.

Extension to the study of multiple alterations

As for the multi-class extension of the ANOVA-based approach, the Bayesian approach can handle more complex combinations of *n* alterations X = {x_1, … x_n}. The set of binary strings $\hat{X}$

$$\hat{X} = \{\ \ 0 \dots 0_n\ \ ,\ \ 0 \dots 0_{n-1}1\ \ ,\ \ \dots\ \ ,\ \ 1 \dots 1_n\ \}$$

represents all the 2^n possible alteration occurrence profiles, encoded as a binary string. The string $0 \dots 0_n$, for instance, represents the absence of any alteration in the sample, whereas $0 \dots 1_i \dots 0$ indicates that only alteration *i* occurs in the sample. We can now redefine M_1 as:

$$M_1: y \sim \mu + c_{x=0 \dots 0_n} + c_{x=0 \dots 0_{n-1}1_n} + \dots + c_{x=1 \dots 1_n} + \varepsilon$$

That is, the new model M_1 has a new variable c_… for each possible value (string) of X'. As before, we can assign to each variable the priors:

- $d_x \sim Normal(0, \sigma_x^2 = g_x) \; \forall \; class \; x$
- $g_x \sim Inverse \; Chi \; Squared(1, h^2) \; \forall \; class \; x$

Interpretation of the Bayes Factors and Posterior parameter estimates

As in the ANOVA framework, this extended approach will return a Bayes Factor that informs on whether at least one specific alteration profile differs from any others. If the Bayes Factor is greater than a given threshold, we call the association between the alteration profile *X* and phenotype *y* significant. Additionally, one of the main advantages of the Bayesian approach is that it implicitly derives the posterior distributions of the model parameters. For instance, in **Extended Data Fig. 6b** we show the estimated BF and posterior distributions for the parameters of the model from the case-study with X = {x_1, x_2}. In the context of this study, the Bayesian framework allows us to recollect the distribution of the mean phenotype *y* in the various alteration classes of *X* and can be used as a quantitative indication of the difference between different alteration profiles. In general, these posterior distributions describe the uncertainty around the estimated parameters and of the model itself.

Post-hoc Bayesian Analysis

As for the frequentist framework, we designed a post-hoc analysis based on the Bayesian framework to infer whether two specific alteration profiles *X'* and *X''* have significantly different phenotypes. We designed two different post-hoc tests, henceforth referred to as direct and indirect post-hoc tests (**Extended Data Fig. 6c,d**).

Direct post-hoc Bayes Factor

In the direct approach, similarly to the Tukey post-hoc test, we directly compare the phenotype *y* between the two specific alteration profiles *X'* and *X''* (**Extended Data Fig. 6c**). Formally, we calculate the Bayes Factor $BF(M_{X'-X''}, M_0)$, defining the model $M_{X'-X''}$ as

$$M_{X'-X''}: y \sim \mu + c_{X'} + \; c_{X''} + \varepsilon$$

and evaluating

$$Direct \; BF(M_{X'-X''}, M_0) = \frac{P(D' \mid M_{X'-X''})}{P(D' \mid M_0)}$$

Where D' are observations from the samples in classes *X'* and *X''*.

All covariates are accounted for in the Post-hoc analysis as well, as done in the main analysis, by comparing $M_{X'-X''+cov}$ to the base model M_{cov} that only considers the covariates as dependent variables.

Indirect Post-hoc Bayes Factor

The direct approach described above, albeit intuitive and adequate for moderate sample or effect sizes, is not always able to correctly detect a real difference between classes when the sample size is low at low false discovery rates (see section 'assessment with synthetic data'). In some cases, the estimated direct post-hoc Bayes Factor favors the alternative model $M_{X'-X''}$ over the null model M_0 (BF > 1) but does not attain a desired significance threshold. In this work we developed an 'indirect' post-hoc approach to recover as much as possible these otherwise discarded cases maintaining a low false discovery rate. Contrary to the direct post-hoc approach and the Tukey's test, which are applicable to any pairwise comparison, this indirect approach is suitable to study the variation of a phenotype y when:

- The two classes of alteration profiles X' and X'' to be compared are such that the alterations defining X' are a superset of the alterations defining X''
- We want to assess if the phenotype y is more extreme in class X' than in class X''.

To clarify with an example where X = {x_1, x_2} (**Extended Data Fig. 6b**), the indirect approach could be used to study the difference in phenotype between samples with both alterations x_1 and x_2 (X' = [1, 1]) versus samples with only alteration x_1 (X'' = [1, 0]). Conversely, the indirect approach is not indicated to study the difference between X' = [1, 0] and X'' = [0, 1], that is, when one of the alterations defining X' and X'' is not a subset of the other.

To compute an indirect BF, we fit the same full model M_1 separately on the observed data (D) and on a resampled version of the original data (D_r) where the class X' = X_{dd} = [1, 1] is randomly bootstrapped from the samples in class X'' = X_{dw} = [1, 0], preserving the total number of samples (**Extended Data Fig. 6d**). The indirect post-hoc BF is defined as

$$Indirect\ BF(M_{X'-X''}) = \frac{P(D|\ M_1)}{P(D_r|\ M_1)}$$

that is the ratio between the BF of model M_1 on the observed data D and the BF of model M_1 on the bootstrapped data D_r. If the ratio is ≤ 1, then there is no strong difference between the two classes X and X'', implying that the interaction between alterations x_1 and x_2 is not likely to influence phenotype y. A high ratio, instead, indicates that the full model considering the original data for the double-altered class has more explanatory power than the reduced one, thus the interaction between alterations informs over the phenotype. This procedure is repeated multiple times (10 in this work), and the median $Indirect\ BF(M_{X'-X''})$ is retained as final indirect BF score to attain stability over random fluctuations.

The resampling approach described above is inspired by the previously introduced concept of 'shadow analysis'[29,30]. In line with the 'shadow analysis' developed in these studies, we only calculate the indirect post-hoc BF score if the direct post-hoc BF > 1, that is if there is evidence that a real difference might be present.

Putting all together: Post-hoc Bayes Factors

To summarize, given a phenotype y and an alteration profile X, the Bayesian analysis proceeds as it follows:

1. Calculate the multivariate BF. This assesses whether any class differs from any other. If BF score > $threshold_{BF\_multivariate}$, call the multivariate association significant. The framework calculates the posterior estimates of the effect sizes as byproduct.
2. If multivariate BF > $threshold_{BF\_multivariate}$ and we are interested in the difference between two specific classes X' and X'': calculate the direct post-hoc significance. If direct post-hoc BF score > $threshold_{BF\_direct}$, call the pairwise association significant.
3. If direct post-hoc BF score > 1 and the alterations defining X' are a subset of alterations defining X'': calculate the indirect post-hoc significance. If indirect post-hoc BF score > $threshold_{BF\_indirect}$, call the pairwise association significant. (The minimum direct BF score required can be increased if more conservative results are desired.)

In general, we chose the three significance thresholds ($threshold_{BF\_multivariate}$, $threshold_{BF\_direct}$, $threshold_{BF\_indirect}$) guaranteeing a given FDR. To estimate the FDR, we computed the null distributions of expected Bayes Factors by repeatedly randomizing the phenotype data (preserving the covariates) and running the association analysis on the randomized data. The significance thresholds used in each specific analysis (i.e. gene essentiality association to EDs, **Fig. 3**, and drug sensitivity between evolutionary axes, **Fig. 4**) are reported and discussed in the corresponding sections below. A systematic validation of the statistical frameworks on synthetic data is provided the section "Comparative analysis of the Frequentist and Bayesian inference frameworks".

Comparative analysis of the Frequentist and Bayesian inference frameworks

We extensively assessed on synthetic data the performance of the Bayesian statistical framework and to what extent it surpasses the frequentist approach. We built a synthetic dataset by generating random instances of simulated case-studies similar by design to the real scenarios we studied in this work. As we study the effect of evolutionary dependencies between pairs of alterations X = {x_1, x_2} on a phenotype y, in each synthetic case-study a collection of samples assigned to four classes: X_{dd} = [1, 1], X_{dw} = [1, 0], X_{wd} = [0, 1] and X_{ww} = [0, 0]. The phenotype y of a sample from class X_k is randomly drawn from a normal distribution with mean m_k and variance sd_k (constant for each class). We set the variance of each class k to the same level (sd_k=1), and attain the desired effect size between classes by changing the means m_k. We set m_{ww} = m_{wd} = 0, and D_p = m_{dd} - m_{dw} > 0 to define true positive cases (P), and D_n = m_{dd} - m_{dw} = 0 to define true

negative cases (N) (**Extended Data Fig. 7a**). We generated multiple case-studies at different effect size D and with variable number of samples per class, to study the impact of effect and sample size in the performance of the inference. For each specific combination of number of samples per class and effect size we generated 10'000 case-studies.

We applied our statistical frameworks to estimate post-hoc BF and P values as described in the previous sections. We estimated True Positive Rate (Recall), False Positive Rate and False Discovery Rate (FDR) of the inference methodologies. In general, we observed that the performance of frequentist and Bayesian approaches converge towards the same levels of recall and FDR when the number of samples in the P cases is comparable to the number of samples in N cases (**Extended Data Fig. 7b**). The difference between the methodologies becomes apparent in the 'worst case scenarios' where P cases have a low number of samples (n_p = 5 to 10 samples) and a small effect size (D_p is smaller than the standard deviation), whereas N cases have high number of samples (n_n > 49 samples). True and False Positive rates for frequentist and Bayesian approaches are shown for two of these worst-case scenarios with low sample (n_p) and low effect size (D_p) in **Extended Data Fig. 7c**. The frequentist-based approach (gray line) performs worse than the direct or indirect post hoc approaches based on Bayesian statistics. The indirect BF approach achieves the highest performance, with a recall between 10-15% at a False Positive Rate around 3%. **Extended Data Fig. 7d** shows the performance of the method for a 'less extreme' scenario with D_p = 0.5, n_p = 10 and n_n = 50. Similar trends appear at different combinations of D_p, N_p and N_n, confirming that our solution based on the Bayesian framework surpasses frequentist approaches at low sample and effect sizes and when the number of samples varies greatly among hypotheses.

References

1. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e10 (2018).
2. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11 (2018).
3. Ellrott, K. *et al.* Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *cels* **6**, 271-281.e7 (2018).
4. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
5. Raynaud, F., Mina, M., Tavernari, D. & Ciriello, G. Pan-cancer inference of intra-tumor heterogeneity reveals associations with different forms of genomic instability. *PLOS Genetics* **14**, e1007669 (2018).
6. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biology* **17**, 31 (2016).
7. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
8. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).

9. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404 (2012).
10. Caravagna, G. *et al.* Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature Methods* **15**, 707 (2018).
11. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
12. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
13. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* (2016) doi:10.1016/j.cell.2016.06.017.
14. Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 (2016).
15. Meyers, R. M. *et al.* Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779–1784 (2017).
16. McFarland, J. M. *et al.* Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat Commun* **9**, 1–13 (2018).
17. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577-592.e10 (2017).
18. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
19. Mina, M. *et al.* Conditional Selection of Genomic Alterations Dictates Cancer Evolution and Oncogenic Dependencies. *Cancer Cell* **32**, 155-168.e6 (2017).
20. Haar, J. van de *et al.* Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* **177**, 1375–1383 (2019).
21. Degasperi, A. *et al.* A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. *Nature Cancer* **1**, 249–263 (2020).
22. Network, C. G. A. & others. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
23. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
24. Berger, J. O. & Sellke, T. Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association* **82**, 112–122 (1987).
25. Jeffreys, S. H. *The Theory of Probability*. (Oxford University Press, 1998).
26. Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology* **56**, 356–374 (2012).
27. Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R. & Wagenmakers, E.-J. Bayesian analysis of factorial designs. *Psychol Methods* **22**, 304–321 (2017).
28. Zellner, A. & Siow, A. Posterior odds ratios for selected regression hypotheses. *Trabajos de Estadistica Y de Investigacion Operativa* **31**, 585–603 (1980).
29. Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).
30. Alvarez, M. J. *et al.* Network-based inference of protein activity helps functionalize the genetic landscape of cancer. *Nat Genet* **48**, 838–847 (2016).