Classification of gene signatures for their information value and functional redundancy

Laura Cantini[1*], Laurence Calzone[1], Loredana Martignetti[1], Mattias Rydenfelt[2,3], Nils Blüthgen[2,3], Emmanuel Barillot[1], Andrei Zinovyev[1*]

1. Institut Curie, INSERM U900, PSL Research University, Mines ParisTech, 26, rue d'Ulm, F-75248 Paris, France.

2. Institute of Pathology, Charite Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany.

3. IRI Life Sciences and Institute for Theoretical Biology, Humboldt University, Philippstr. 13, Haus 18, 10115 Berlin, Germany

Correspondence : Laura Cantini (laura.cantini@curie.fr) and Andrei Zinovyev (andrei.zinovyev@curie.fr)

ABSTRACT

Large collections of gene signatures play a pivotal role in interpreting results of omics data analysis but suffer from compositional (large overlap) and functional (redundant read-outs) redundancy, and many gene signatures rarely pop-up in statistical tests. Based on pan-cancer data analysis, here we define a restricted set of 962 so called informative signatures and demonstrate that they have more chances to appear highly enriched in cancer biology studies. We show that the majority of informative signatures conserve their weights for the composing genes (eigengenes) from one cancer type to another. We construct InfoSigMap, an interactive online map showing the structure of compositional and functional redundancies between informative signatures and charting the territories of biological functions accessible through transcriptomic studies. InfoSigMap can be used to visualize in one insightful picture the results of comparative omics data analyses and suggests reconsidering existing annotations of certain reference gene set groups.

KEYWORDS: gene signature, transcriptome, PCA, network

INTRODUCTION

The majority of the studies exploring gene expression data result in one or more gene signatures, i.e. list of genes sharing a common pattern of expression that can be employed to classify an independent dataset. Together with such "data-derived" signatures, "*a priori* knowledge-based" gene signatures are produced from the available gene ontologies or pathway databases. In recent years, data-derived and *a priori*-knowledge-based reference gene signatures have been widely employed to interpret the results of gene expression data analyses (e.g. differential expression, clustering). The number of available signatures is getting larger allowing users to benefit for a more exhaustive coverage of the existing biological processes. However, not all the signatures contained in these compendia are equally informative and the number of gene sets representing the same biological process is not equilibrated. Intrinsic redundancy and the presence of numerous signatures which have small chances to be selected in any analysis affects the results by heavy p-value corrections producing a higher number of false negative results. Conceptually, the aforementioned gene set redundancy can be of two types: compositional or functional (see Figure 1A). Compositionally redundant signatures are characterized by a large intersection in terms of the composing genes. On the opposite, two signatures will be here called functionally redundant when they represent two different (sometimes, with zero overlap) transcriptional read-outs of the same biological process. The presence of multiple functionally redundant signatures affects the enrichment analysis by highly scoring multiple gene sets belonging to analogous or related biological processes hiding other potentially relevant hits. Of note, any estimation of the functional redundancy is conditioned on the context (e.g., certain cancer type) and, therefore, depends on the data corpus which defines functional (e.g., correlation-based) relations between the individual genes.

To our knowledge, few methods have been proposed to address the problem of gene signature redundancy [1–6]. Currently, the best attempt to define a robust and not-redundant collection of signatures is represented by the MSigDB Hallmarks (H) collection [7]. H was obtained by merging and re-organizing compositionally redundant signatures and then refining the genes of the resulting signatures based on their ability to discriminate the associated phenotype. The Hallmarks collection

methodology thus involves a manual curation and redefinition of signatures steps, which might create certain bias *vis a vis* an expert's opinion leading to the loss of certain signature properties. More importantly, H, as all the other currently proposed procedures, takes into account only compositional redundancy without exploiting the problem of the functional one.

In this paper, a new approach to prioritize and classify gene signatures is proposed. Our method is based on the concept of "an informative signature", which is capable of defining a ranking of samples independently on their labeling. Considering the simplest case of a gene signature composed of only two genes X and Y, their co-variance can define three possible scenarios of samples distribution (ranking), as reported in Figure 1B. Whereas the ranking defined by informative signatures presents a distinguishable axis (anisotropy), labels are required to define a samples ranking in the non-informative ones. As a consequence, when the dysregulation of an informative signature is tested on a two-conditions transcriptomics dataset, a significant enrichment score will be observed whenever the samples membership matches the direction of the largest variance and this is significantly greater than randomly expected (the gene set is "overdispersed"). In all other cases, a very specific distribution of sample labels is needed to obtain a significant enrichment score. Therefore, informative signatures, systematically defining robust and "objective" sample ranking in many datasets, are more valuable for data analysis.

Starting from a vast collection of signature compendia, composed of 12096 *a priori*-knowledge and data-derived signatures, we defined a restricted collection of 962 informative signatures, which is made available to the users for further applications. The collection was defined by exploiting a large pan-cancer TCGA collection of transcriptomic profiles (32 cancer types with totally 8991 transcriptomic profiles) and it is thus cancer biology-oriented. Of note is the fact that among the databases under investigation, a relatively small SPEED collection [8] proved to be the most informative with 55% of its signatures being highly informative. The reliability of our signature collection was then validated by comparing their performances with those of the starting complete collection in some typical data analysis scenarios. In all the considered examples, the informative gene sets were found much more frequently significant than the others, confirming the rationale behind the selection procedure here proposed.

Each signature defines a set of weights for the composing genes (which we term eigengene [9]) in each analyzed dataset. We introduce the notion of "conserved signatures", i.e. those signatures whose

eigengenes are highly correlated across different cancer types. 73% of our informative signatures resulted to be also highly conserved, pointing out that they map well some universal functional blocks of the cellular machinery. The collection of informative gene signatures was classified by computing the average correlation between the sample activity profiles of all signatures across 32 cancers types. This metrics was used as a measure of functional redundancy in our analysis. We found that functional redundancy is an abundant phenomenon that does not result from a significant intersection size of two gene sets. Therefore the Hallmarks and the other previous works, employing only the overlap as a measure of redundancy, are capturing only a small portion of this phenomenon. In order to visually and interactively represent the structure of functional redundancies between informative gene signatures, we developed InfoSigMap (http://navicell.curie.fr/pages/maps_avcorrmodulenet.html), a user-friendly interactive Google Maps-based tool whose nodes are constituted by our set of informative signatures and whose links represent the two types of redundancies (compositional and functional). InfoSigMap can be used as a data visualization tool to provide a quick navigation into any set of scores associated to the informative signatures (e.g., enrichment scores). The use of InfoSigMap is demonstrated in some typical data analysis scenarios showing that it is able to provide a concise and biologically meaningful holistic view on the pattern of differential regulation of various cellular functions.

RESULTS

Informative signatures represent a small fraction of the widely employed gene sets

Defined the concept of “an informative signature” (see Methods), a large pan-cancer TCGA compendium of gene expression data derived from 32 solid cancer types was employed to restrict the input collection of gene signatures (12096) to 962 (see Supplementary Table S2) informative for cancer data analysis and corresponding to our compendium (for the selection procedure and inputs see Methods). Of the 962 identified informative signatures the majority were data-derived (231 knowledge-based, 706 data-derived, 15 MSigDB Hallmark and 10 MSigDB C1), showing that for cancer-oriented applications data-derived signatures tend to be more informative than knowledge-based ones. In order to assess which of the input compendia was more informative, the ratio between the number of informative gene sets and the total number of contained signatures was evaluated

(Table 1). As shown in Table 1, the most informative compendium resulted to be SPEED (55% of informative signatures). The reliability of this database is thus corroborated by our results, suggesting that when dealing with cancer transcriptomics data analysis it should be probably preferred to alternative ones. Overall, good performances were also obtained by CIT (28%), the MSigDB C4 (31%) and the Hallmarks (H) (30%). While the best performing knowledge-based database resulted to be ACSN, with 13% of informative signatures. The distribution of the number of tumors in which the informative signatures were found to have the ROMA L1 and L1/L2 p-values significant were then investigated. As shown in Figure 2A and Supplementary Table S2, the majority of the informative signatures is cancer-specific (significant in only 2 cancer types), another peak is present around 15-20 tumor types and only 8 signatures are pan-cancer significant (significant in more than 25 cancer types). Furthermore, data-driven signatures tend to be more frequently pan-cancer-wise informative than the knowledge-based ones.

Informative gene sets tend to be much more frequently significant in some typical cancer data analysis

Our hypothesis that an informative gene set has much higher chances to be enriched in a typical transcriptomic data analysis, is here tested according to the procedure detailed in the Methods section in three typical scenarios: (i) KRAS mutated vs. wild type colorectal cancer [10]; (ii) metastatic vs. primary colon cancer [11] and (iii) tumor vs. normal in four tissues (Lung [12], Gastric [13], Colon [14], Cervix [15]). As shown in Table 2, in all three cases the number of informative signatures in the GSEA output was strongly enriched (average P-value 10^-79). Note that, while the selection of the informative signature was performed using an unsupervised approach, the validations presented in this section are realized using a supervised one (GSEA). Nevertheless the amount of informative signatures obtained in the output of the GSEA analysis is significantly higher than what could be expected at random.

Informative signatures perform better than the MSigDB Hallmarks in some typical cancer data analysis

Given that the only other attempt to prioritize the most reliable non-redundant signatures is represented by the MSigDB Hallmarks, its performances were compared with those of our

compendium in the three above test cases in terms of Fisher's exact test P-values and P-value of the Kolmogorov-Smirnov (KS) test for the NES distributions, as described in the Methods section. The results of both Fisher's exact test and the KS test are summarized in Table 2. As shown in the table, the Hallmarks signatures obtained significant Fisher P-values in all cases (average P-value 10^-05), confirming the reliability of the procedure employed by Liberzon A. and coauthors. However, the P-values obtained by the Hallmark collection resulted to be always less significant than those of our informative signatures (10^-05 vs 10^-79). Concerning the NES distribution, as shown in Supplementary Figure S1, the informative signatures tend to be always associated with absolute NES higher than those of the Hallmarks. Indeed, the KS P-values are always lower than 0.05, except for the KRAS mutated vs. wild type colorectal cancer example in which the P-value is 0.098. Therefore, not only the informative signatures have higher chances to be significant in a GSEA analysis, but also in the GSEA output they tend to be among those with the highest NES score. This result indicates that our compendium is capturing the strongest sources of expression variation in all three transcriptomic datasets. As a further check, given that 15 over 50 Hallmarks signatures are also contained in our informative collection, the fraction of Hallmarks signatures present in the output of the GSEA analysis and also contained in our informative compendium is evaluated: (i) KRAS mutated vs. wild type colorectal cancer 67%; (ii) metastatic vs. primary colon cancer 67% and (iii) normal tissue vs. tumor in 4 tissues (Lung 48%, Gastric 47%, Colon 44%, Cervix 60%). Such a result shows that among the 50 signatures constituting the MSigDB Hallmarks, those that are found significant in the GSEA analysis are in the majority of the cases also informative. This last observation thus further confirms the reliability of our selection procedure taking into account that the subset of the Hallmarks that we selected tends to be enriched more frequently in the analyzed cancer-specific applications.

The majority of the informative signatures eigengenes are conserved across cancer types

To further investigate the reliability of our informative collection, we verified if the informative signatures were quantitatively reproduced across different cancer types, where the term quantitatively refers to the eigengenes (set of gene weights) resulting from the PCA decomposition. Computing the conservation score as described in the Methods section and employing a threshold of 10^{-6}, 1459 over the 12096 starting signatures (12%) resulted to be conserved across-cancer. On the opposite 703

over the 962 informative signatures (73%) were found to be conserved, showing that the signatures selected with our approach have higher chances to maintain the same quantitative definition across different cancer types and thus they tend to be more robust than the starting ones. Testing then the much stringent threshold of 10^{-10}, while the total number of conserved signatures substantially decreased (from 1459 to 408) the percentage of conserved signatures that were found to be also informative substantially increased (from 48% to 83%). This last result thus proves that the previously obtained results are not affected by the value of the chosen conservation score threshold and, more importantly, that the informative signatures are always among the most across-cancer conserved gene sets.

Functional redundancy of gene sets is poorly explained by their intersection size

Two gene sets with empty intersection can represent incomplete transcriptional read-outs of the same biological process (i.e., cell cycle), being thus functionally redundant. For example, when data-derived expression signatures are constructed, the number of genes whose expression is associated with the phenotype of interest is generally minimized and only the most representative are maintained in the signature. This procedure may lead to the reconstruction of two data-derived signatures associated to the same phenotype but having a poor/null intersection. This leads to a well-known problem of signature reproducibility [16]. In order to quantify the scale of this phenomenon, the two redundancy measures: functional redundancy, computed as described in the Methods section, and gene content intersection in terms of Jaccard-index (JI) are compared in Figure 2B. As expected, high JI value usually results in high functional redundancy (i.e. high average correlation between meta-samples over all cancer types). However, as shown in the figure, high functional redundancy is distributed over a large range of JI values with a surprisingly higher points density in the area corresponding to poorly overlapping gene sets. Therefore, in order to reduce the functional redundancy between gene sets, it is not sufficient to simply take into account their overlap, but also their correlation of activity needs to be considered. A first consequence of this result is that the Hallmarks collection, based on the use of the JI as a measure of redundancy, is not able to completely capture analogous signatures. To better quantify redundancy, also the meta-samples correlation need to be considered. The intrinsic limitation of such approach is that it requires the wide employment of expression data and thus its output is data-dependent.

InfoSigMap a user-friendly interactive representation of the functional redundancy structure between informative signatures for insightful gene set score visualization

Given the predefined collection of informative signatures, GSEA or alternative approaches can be employed in order to score them based on a transcriptomics dataset and a set of sample labels. The output of such analysis consists of a table of gene sets enrichment statistics. This tabular organization of the output, containing functionally connected signatures scattered throughout, frequently does not help the interpretation of the results and the formulation of consistent biological hypothesis. In order to improve this aspect, we developed InfoSigMap (http://navicell.curie.fr/pages/maps_avcorrmodulenet.html), according to the procedure described in Methods. As shown in Figure 3, the obtained network contains six main strongly connected components of informative signatures. The first is associated to core cellular functions (i.e. all those basic functions that are fundamental for the life of the cell) and it contains cell cycle, mRNA translation, splicing, MYC targets, protein degradation and oxidative phosphorylation. Of particular interest is the fact that in this component the network is able to clearly separate the signatures associated to the different cell cycle phases. The second connected component is instead related to the tumor microenvironment and it comprises: immune system, inflammation, TNF-α pathway, interferon and extracellular matrix/EMT. The third and fourth connected components, smaller than the others, correspond to transcription and neuronal system. Finally, the fifth and sixth components contain signatures associated to analyses performed on the same expression compendia (GNF2 [17] and GCM [18]) and simply represent genes with large neighbourhood overlap. The first and the second large components are connected through an area associated to Experimental perturbations of Immune cells (EI) signatures. Indeed the informative signatures derived from EI (dark red nodes in Figure 3) are separated into two main areas: one, as could be expected, belongs to the tumor microenvironment component and it is strongly linked to the immune system/inflammation signatures; the second, instead, is part of the proliferation component, strongly linked to the cell cycle area. We considered that such unexpected configuration could be caused by the presence of gene sets, belonging to the EI category, but obtained from differentiation induction experiments or on immortalized cell lines, and thus characterized by differences in the proliferation rate. Indeed an alteration in the cell cycle process could justify the high correlation of activity (reflected in an high

density of links) present between these two sets of signatures. In order to test this hypothesis, a dataset obtained from the expression profiling of human CD4+ T cell during differentiation induction was employed [19]. The informative pathways altered during the experiment (differentiated vs undifferentiated) were detected employing ROMA according to the procedure described in Methods section. As shown in Supplementary Figure S2, the two areas of experimental perturbations of immune cells signatures have an opposite behavior concordant with that of the signatures around them. Indeed the area near to the cell cycle results to be donwregulated in the differentiated cells, while those that are part of the immune island are upregulated. This result confirms our starting hypothesis that an alteration of the cell cycle process was at the origin of the observed subdivision and it suggests that a reshaping of the EI category would be recommended for their future use in data analysis. Another non-intuitive observation is that signatures coming from the same collection tend to co-localize in the map and data-derived signatures tend to be clearly separated from knowledge-based ones. The discrepancy between data-derived and knowledge-based signatures can be explained by the fact that the transcriptional readouts of a biological process might be very different from the genes involved in the process itself. Yet another unexpected observation is that higher functional redundancy exists between signatures of the same collection rather than between signatures describing the same biological function, suggesting that all the analyzed signature collections can be prone to a common bias. The only two exceptions to this trend to some extent are the MSigDB Hallmark collection and the SPEED signatures (although several non-informative SPEED signatures are clustered together). These compendia indeed resulted to be well spread around the map, confirming that they are able to well capture the main biological signals encoded in the transcriptomic data. Nevertheless, some areas such as mRNA translation, transcription, splicing and protein degradation were not covered by any of the Hallmark and SPEED signatures, indicating that there is still the need of other signatures in order to have a complete portrait of the transcriptomic landscape. As introduced above, InfoSigMap was developed to simplify the navigation and interpretation of the gene set score distributions. In the next section, some examples of typical analysis scenarios where InfoSigMap can be employed to formulate consistent biological hypothesis are presented.

InfoSigMap can be used to visualize the results of transcriptomics data analysis

InfoSigMap is tested to investigate the alterations affecting the transcriptome of the three aforementioned typical cancer problems: (i) KRAS mutated vs. wild type colorectal cancer; (ii) metastatic vs. primary colon cancer and (iii) tumor vs. normal tissue in lung, gastric, colon and cervix. In all three cases, the procedure employing ROMA and described in Methods is applied resulting in the output shown in Figure 4. Below the obtained results are discussed in detail:

(i) KRAS mutated vs. wild type colorectal cancer

The impact of *KRAS* mutation on CRC transcriptome is investigated with InfoSigMap (Figure 4A). *KRAS* mutated CRC patients are known to be resistant to standard Epidermal Growth Factor Receptor (*EGFR*) inhibitory treatments [20,21]. The output of our analysis can thus give some indications concerning possible new processes to be targeted in KRAS mutated patients. The strongest effect reported in Figure 4A (bright red area) is the upregulation of a subset of the metastatic signatures. This result fits with previous evidences that KRAS mutation is associated to metastasis in patients with CRC [22,23]. Moreover an alteration of the metabolism is detectable from an upregulation of the mitochondria and oxidative phosphorylation area. This result fits with previous experimental evidences. Indeed *KRAS* mutation has already been shown to induce mitochondrial oxidative stress, inducing a phenotype consistent with the so-called Warburg effect, a metabolic alteration fundamental for cancer cell proliferation [24–26]. In CRC, *KRAS* mutation also causes an alteration of the transcriptional response and amino acid metabolism machineries, two fundamental processes for cancer cell proliferation and maintenance [27,28]. This effect is captured in our analysis by the upregulation of the mRNA translation/protein metabolism area on the InfoSigMap network.

(ii) Metastatic vs. primary colon cancer

The differential module activity between metastatic and primary Colon Cancer (CC) is studied (Figure 4B). As expected, an up-regulation of the collagen/EMT area of the network clearly appears on the map Figure 4B. Among the signatures found in this area, we can observe an upregulation of the miR-21 targets whose role in EMT is well-known [29,30]. Moreover, the network areas of mRNA and proteins metabolism and splicing resulted to be significantly upregulated. This is not surprising given that the aberration of the RNA processing machinery (stability, metabolism, splicing and polyadenylation) is

known to be associated with cancer initiation and progression. In CC, beta-catenin (*CTNNB1*), involved in the Wnt pathway, is generally the cause of the RNA processing alterations [31–33]. This is confirmed in InfoSigMap; indeed *CTNNB1* is found active as shown by the upregulation of its targets (node FEVR_CTNNB1_TARGETS). The study of the cancer-specific RNA metabolism is a relatively unexplored area of research, with potentially significant implications for the prevention and treatment of CC. The above results confirm the experimentally observed *CTNNB1*-mediated alteration of the RNA processing machinery. On the other side, a strong downregulation of the cells proliferative activity can be observed. This phenomenon has already been documented and found associated with poor-prognosis in colorectal cancer (CRC) [34,35]. The observed slow-proliferation in metastatic CC may be caused by a high proportion of cancer stem-like cells. Indeed stem cells are in a quiescent state, a phenomenon that could explain the cell cycle downregulation detected in our analysis. The hypothesis of a high stem cell concentration is also confirmed by the significant downregulation of the immune area. Indeed the stem-like phenotype of metastasis-initiating cells is generally associated with immune evasive quiescence, even if this point is not well documented in CC [36].

(iii) Tumor vs. normal in four tissues

When then compare tumor vs normal tissue in cervix, colon, gastric and lung cancer (Figure 4C-F). A global feature present in all four tissue types is the upregulation of the connected component associated to the core cellular functions. This is not a surprising result, given that cancer cells generally inactivate tumor suppressors and hyperactivate oncogenes to promote sustained proliferation, alter autophagy and the various steps of the RNA transcription and translation processing machinery, develop metabolic imbalances and enhance resistance to mitochondrial apoptosis [37]. The microenvironment-associated connected component instead shows a dual behavior. It is indeed significantly upregulated in cervical and gastric cancer and downregulated in colon and lung. The results thus suggest a different role of the immune system in these four tumors. A possible explanation for this result is that the tumors are associated to different levels of antigenicity, i.e. the extent to which tumor cells display HLA-restricted antigens that can be selectively or specifically recognized by T cells [38]. Tumors with low antigenicity hide against cytotoxic attack leading to a passive escape from anti-cancer immune defense. This hypothesis is supported by the observation that the HLA signature present in our network (GNF2_HLA_C) is concordantly downregulated in lung

and colon and upregulated in cervix and gastric. Moreover, concerning lung cancer, its association with low antigenicity had already been observed [39]. The tumor antigenicity is one of the aspects that seem to determine if a patient will respond to a given immunotherapy. In this sense, a comprehensive pan-cancer classification of the immune component behavior could give indications regarding those individuals who are most likely to respond to immune-based therapies.

DISCUSSION

Data-driven and a priori-knowledge-based gene signatures are largely employed in cancer studies to score clinical samples according to distinct tumor subtypes, identify important cellular responses to stimuli, predict clinical outcomes and quantify the activation of signaling pathways. Nowadays signature collections are getting larger providing the benefit of a more complete coverage of the existing biological processes. However, the growth of these compendia is posing two main challenges due to the reliability and redundancy of the collected gene sets.

Here, we developed a new methodology for assessing the value of a gene set which is based on the notion of informative signature, i.e. a gene set able to systematically robustly rank tumor samples in many independent datasets. A restricted collection of 962 informative gene sets is suggested for transcriptomic data analysis in cancer biology. The robustness of the information content enclosed in our compendium is then proved showing that an informative gene set has much higher chances to be selected (enriched) in a typical scenario of transcriptomic data analyses, even in the ones using supervised methods, and that the eigengenes of the majority of the informative signatures tend to be conserved across cancer types. The redundancy of the informative collection is then investigated, showing that functional redundancy is a frequent phenomenon not captured by the approaches previously proposed for gene signatures summarization. To integrate all the obtained results we developed InfoSigMap, a user-friendly interface designed for insightful data visualization. InfoSigMap was applied in some typical scenarios of cancer data analysis. The obtained results proved that a global view of the concordant behavior of functionally redundant signatures leads to an insightful interpretation of the results in respect to what can be deduced from the lists of significant signatures output of gene expression data analysis tools. In all four analyzed cases, the obtained results were found to fit with the previous experimental knowledge, confirming the reliability of our approach. However, also some indications concerning new candidate mechanisms to be experimentally

investigated were extracted, showing how InfoSigMap can help the formulation of new biological hypothesis.

METHODS

Definition of "an informative signature"

To explain the rationale behind the definition of informative signature, let us consider to apply Principal Component Analysis (PCA) to a gene expression data table whose columns correspond to the genes from a selected gene set and whose rows correspond to samples. If we observe that the variance explained by the first principal component computed for such a table is significantly larger than for a random set of genes of the same size then the considered gene set is called overdispersed [40,41]. Intuitively, an overdispersed gene set has a stronger contribution to the data variance than expected by chance. Similarly, if the ratio between the variances explained by the first and second principal components computed for the aforementioned table is larger than for a random set of genes then the given gene set is called coordinated. Intuitively, the existence of a statistically significant gap between the first and the second eigenvalue of the covariance matrix corresponds to an overall increase in the pairwise correlations between the genes of the signature compared to what can be observed at random. The advantage of having a coordinated gene set is that it defines an axis of principal variance in the multi-dimensional distribution of samples and thus it robustly ranks samples independently on their labeling, as discussed in the Introduction (see Figure 1). In the context of cancer biology, we define informative a gene set that is simultaneously overdispersed and coordinated in more than two cancer types.

Transcriptomics data input of the analysis

To systematically search for informative signatures, a large pan-cancer TCGA compendium of gene expression data derived from 32 solid cancer types (ACC, BLCA, BRCA, CESC, CHOL, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS, UVM) was employed. The data were downloaded from TCGA and normalized. An overview of the samples available for the different tumor types is reported in Supplementary Table S1.

Signatures collection input of our analysis

A vast collection composed of both data-derived and *a priori*-knowledge-based signatures was considered as input for our analysis. The signature collections: Molecular Signature Database (MsigDB v5.2) [42], Atlas of Cancer Signaling Network (ACSN) [43], the top-contributing genes of the components identified by Biton et al. (here denoted as CIT) [44] and the Signaling Pathway Enrichment using Experimental Data sets (SPEED) [8] have been downloaded and organized, obtaining a starting collection of 12096 signatures. In the following, we will consider as data-derived signatures: CIT, SPEED and some MSigDB categories (clusters of genes co-expressed in microarray compendia (C4), signatures of oncogenic pathway activation (C6), the large collection of immunological conditions (C7) and chemical and genetic perturbations (CGP) part of the MSigDB collection canonical pathways and experimental signatures curated from publications (C2)). On the other side, ACSN and the MSigDB categories: genes sharing cis-regulatory motifs up- or downstream of their coding sequence (C3), genes grouped according to Gene Ontology (GO) categories (C5) and Canonical Pathways (CP), part of C2 and including the well-known BIOCARTA, KEGG and REACTOME databases, will be denoted as knowledge-based. Finally, the MSigDB collections genes grouped by their location in the human genome (C1) and the Hallmarks (H) will not be associated to any of the two previous classifications.

Procedure for the prioritization of those signatures that are informative in cancer biology

To detect which of the starting 12096 signatures, detailed above, were informative, we employed the Representation and quantification Of Module Activity (ROMA) tool, designed for the robust detection of overdispersed and coordinated modules [41]. The activity of each signature was thus evaluated in all the 32, previously described, expression datasets separately. Only those signatures having the p-values associated to the variance explained by the first principal component (ROMA L1 score) and to the ratio between the variances explained by the first and second principal components (ROMA L1/L2 score) lower than 0.05 in at least two tumor datasets were prioritized.

Test if informative gene sets are more frequently enriched in typical cancer analysis scenarios

Our hypothesis that an informative gene set has much higher chances to be enriched in a typical transcriptomic data analysis, is tested by using Gene Set Enrichement Analysis (GSEA), a well-known

and widely adopted supervised approach [42]. GSEA was applied to three typical cancer-related problems using the entire collection of 12096 starting signatures. The set of significant signatures was determined selecting those with a GSEA FDR q-value lower than 0.05. Then, the enrichment of our set of informative signatures in the output of the GSEA analysis was evaluated through a Fisher's exact test.

Comparison Informative signatures vs Hallmarks in typical cancer analysis scenarios

Given that only one other attempt to prioritize the most reliable non-redundant signatures exists and it is represented by the MSigDB Hallmarks, the procedure described in the previous section was repeated also for this database. The performances of our informative collection were then compared with those of the Hallmarks in terms of Fisher's exact test P-values. Moreover, the distributions of the absolute GSEA Normalized Enrichment Score (NES) for the two collections were studied and the significance of the difference between the two distributions was evaluated through a Kolmogorov-Smirnov (KS) test.

Evaluation of the signatures eigengenes conservation across-cancers

To further investigate the reliability of our informative collection, we verified if the informative signatures were quantitatively reproduced across different cancer types, where the term quantitatively refers to the eigengenes (set of gene weights) resulting from the PCA decomposition. For each informative signature, the pair-wise correlation between the eigengenes obtained in the 32 cancer types were computed and a conservation score was obtained as the geometric mean of the Pearson correlation p-values. We define a conserved gene set by having the conservation score lower than 10^{-6}. To then evaluate how much the results of this test were affected by the threshold used to define a conserved gene set, also a much stringent threshold of 10^{-10} was tested.

Comparison between functional redundancy and intersection size of gene sets

Two gene sets with empty intersection can represent incomplete transcriptional read-outs of the same biological process, being thus functionally redundant. In order to quantify the scale of this phenomenon, we have compared the normalized size of gene set intersection with their functional redundancy measure. For each couple of informative signatures, the average correlation between

their meta-samples was thus computed over the 32 cancer types. This approach captures all those couples of gene sets having a similar pan-cancer behavior and thus representing the same transcriptional read-out, independently on whether they have significant number of common genes or not. To evaluate if the functional redundancy between a couple of signatures is well explained by their overlap size, the intersection in terms of Jaccard-index (JI) between all couples of informative gene sets was computed and the two measures were compared.

InfoSigMap construction procedure

To help the quick navigation of the set of informative signatures and to improve the interpretation of their concordant behavior, we developed InfoSigMap (http://navicell.curie.fr/pages/maps_avcorrmodulenet.html), a user-friendly Google-Maps based visualization method. The construction of InfoSigMap involved three main steps: (i) creation of the signature redundancy graph; (ii) definition of its layout and (iii) representation of the graph as an interactive online map. In line with what has been already done in Enrichment Map, GOlorize and ClueGO [2,4,6], the first step is performed by organizing the 990 signatures (corresponding to the 962 informative collection plus all Hallmarks and SPEED signatures even if they were not shown to be informative) into a weighted network, where each signature is a node and links represent redundancy between couples of gene sets. Differently from the previously mentioned Cytoscape plug-ins, the links of our network are weighted averaging over two measures of signatures redundancy: overlap (JI) and functional redundancy. The functional redundancy was computed as the average correlation coefficients between the metasamples defined by ROMA, for each pair of informative signatures in each cancer type. The signatures having average correlation above 0.7 are connected in the graph and the final weights of the links were obtained as the mean between the average correlation and the JI. The threshold 0.7 is justified by appearance of distinguishable but still connected functional components in the graph. For the second step of graph structure representation, a different shape is used to denote the gene sets that are only informative (diamond) and those that are also conserved (circle), while the node size denotes the number of genes in the signature. Links are also classified into two classes, dark gray is used for those edges that connect signatures being both functionally redundant and having a significant JI, while light gray denotes links only associated to functional redundancy. Finally, the thickness of the links is proportional to their weights, the standard Cytoscape

organic layout is used to spatially organize the largest connected component of the network and smaller components or unconnected nodes were positioned by using the structure of weaker correlations. The areas of the network containing signatures associated to same biological functions are then identified and manually annotated on the top of the map to help the navigation of the users. In addition, also a purely data-driven layout was computed by applying tSNE dimension reduction method to the matrix of average pairwise correlations between the meta-samples defined by our signatures in all cancer types. This view of the InfoSigMap is available at http://navicell.curie.fr/pages/maps_avcorrmodulenet.html (View/tSNE selection in the right-hand panel). Finally, the representation of the network as an interactive online map is achieved by using NaviCell [45], powered by Google Maps API.

Using InfoSigMap to have a global view of the signatures behavior

Gene sets can be tested for differential activity across different experimental conditions using a tool of choice (e.g. GSEA, ROMA). If ROMA is chosen for this test, first the activity of the informative signatures is evaluated by applying ROMA, then the differential module activity is evaluated by Student's t-test and fold-change applied to the ROMA activity scores. Finally, the fold-changes associated to a significant Student's t-test P-value (lower than 0.05) are mapped to the nodes of InfoSigMap as a color gradient, from red (up-regulation) to white (no significant change) to green (down-regulation), using the map staining approach described in [45]. The map is thus colored in the territories around each node creating a continuous colored pattern that helps a qualitative appreciation of the concordant/discordant behavior of large map regions.

AUTHOR CONTRIBUTIONS

LCan, AZ and LCal designed the study. LCan made numerical computations. LCan, AZ wrote the manuscript. NB, MR provided materials and analyzed the results. LCan, LM participated in case studies. All authors have read and edited the manuscript.

ACKNOWLEDGMENTS

This study was financially supported by COLOSYS EU EracoSysMed project.

REFERENCES

1. Bauer, S., Grossmann, S., Vingron, M. & Robinson, P. N. Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24,** 1650–1651 (2008).
2. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25,** 1091–1093 (2009).
3. Doig, T. N. *et al.* Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. *BMC Genomics* **14,** 469 (2013).
4. Garcia, O. *et al.* GOlorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. *Bioinformatics* **23,** 394–396 (2007).
5. Lewin, A. & Grieve, I. C. Grouping Gene Ontology terms to improve the assessment of gene set enrichment in microarray data. *BMC Bioinformatics* **7,** 426 (2006).
6. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS ONE* **5,** e13984 (2010).
7. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1,** 417–425 (2015).
8. Parikh, J. R., Klinger, B., Xia, Y., Marto, J. A. & Bluthgen, N. Discovering causal signaling pathways through gene-expression patterns. *Nucleic Acids Res.* **38,** W109–W117 (2010).
9. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1,** 54 (2007).
10. Marisa, L. *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* **10,** e1001453 (2013).
11. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
12. Landi, M. T. *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PloS One* **3,** e1651 (2008).

13. D'Errico, M. *et al.* Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. *Eur. J. Cancer Oxf. Engl. 1990* **45,** 461–469 (2009).
14. Cordero, D. *et al.* Large differences in global transcriptional regulatory programs of normal and tumor colon cells. *BMC Cancer* **14,** 708 (2014).
15. Scotto, L. *et al.* Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. *Genes. Chromosomes Cancer* **47,** 755–765 (2008).
16. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5923–5928 (2006).
17. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6062–6067 (2004).
18. Ramaswamy, S. *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 15149–15154 (2001).
19. Lee, M. S., Hanspers, K., Barker, C. S., Korn, A. P. & McCune, J. M. Gene expression profiles during human CD4+ T cell differentiation. *Int. Immunol.* **16,** 1109–1124 (2004).
20. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **3,** 11–22 (2003).
21. Malumbres, M. & Barbacid, M. Timeline: RAS oncogenes: the first 30 years. *Nat. Rev. Cancer* **3,** 459–465 (2003).
22. Renaud, S. *et al.* KRAS and BRAF mutations are prognostic biomarkers in patients undergoing lung metastasectomy of colorectal cancer. *Br. J. Cancer* **112,** 720–728 (2015).
23. Tie, J. *et al.* KRAS mutation is associated with lung metastasis in patients with curatively resected colorectal cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **17,** 1122–1130 (2011).
24. Neuzil, J., Rohlena, J. & Dong, L.-F. K-Ras and mitochondria: Dangerous liaisons. *Cell Res.* **22,** 285–287 (2012).

25. Toda, K. *et al.* Metabolic Alterations Caused by KRAS Mutations in Colorectal Cancer Contribute to Cell Adaptation to Glutamine Depletion by Upregulation of Asparagine Synthetase. *Neoplasia* **18,** 654–665 (2016).
26. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324,** 1029–1033 (2009).
27. DeBerardinis, R. J. & Cheng, T. Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene* **29,** 313–324 (2010).
28. Wise, D. R. & Thompson, C. B. Glutamine addiction: a new therapeutic target in cancer. *Trends Biochem. Sci.* **35,** 427–433 (2010).
29. Ferraro, A. *et al.* Epigenetic regulation of miR-21 in colorectal cancer: ITGB4 as a novel miR-21 target and a three-gene network (miR-21-ITGB4-PDCD4) as predictor of metastatic tumor potential. *Epigenetics* **9,** 129–141 (2014).
30. Huang, K. *et al.* MicroRNA roles in beta-catenin pathway. *Mol. Cancer* **9,** 252 (2010).
31. Korinek, V. *et al.* Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275,** 1784–1787 (1997).
32. Lee, H. K., Choi, Y. S., Park, Y. A. & Jeong, S. Modulation of oncogenic transcription and alternative splicing by beta-catenin and an RNA aptamer in colon cancer cells. *Cancer Res.* **66,** 10560–10566 (2006).
33. Lee, H. K. *et al.* beta-catenin regulates multiple steps of RNA metabolism as revealed by the RNA aptamer in colon cancer cells. *Cancer Res.* **67,** 9315–9321 (2007).
34. Anjomshoaa, A. *et al.* Slow proliferation as a biological feature of colorectal cancer metastasis. *Br. J. Cancer* **101,** 822–828 (2009).
35. Palmqvist, R. *et al.* Low tumour cell proliferation at the invasive margin is associated with a poor prognosis in Dukes' stage B colorectal cancers. *Br. J. Cancer* **79,** 577–581 (1999).
36. Malladi, S. *et al.* Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165,** 45–60 (2016).

37. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).

38. Savage, P. A. Tumor antigenicity revealed. *Trends Immunol.* **35,** 47–48 (2014).

39. Domagala-Kulawik, J. The role of the immune system in non-small cell lung carcinoma and potential for therapeutic intervention. *Transl. Lung Cancer Res.* **4,** 177–190 (2015).

40. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13,** 241–244 (2016).

41. Martignetti, L., Calzone, L., Bonnet, E., Barillot, E. & Zinovyev, A. ROMA: Representation and Quantification of Module Activity from Target Expression Data. *Front. Genet.* **7,** (2016).

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

43. Kuperstein, I. *et al.* Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis* **4,** e160 (2015).

44. Biton, A. *et al.* Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. *Cell Rep.* **9,** 1235–1245 (2014).

45. Bonnet, E. *et al.* NaviCell Web Service for network-based data visualization. *Nucleic Acids Res.* **43,** W560-565 (2015).

FIGURES

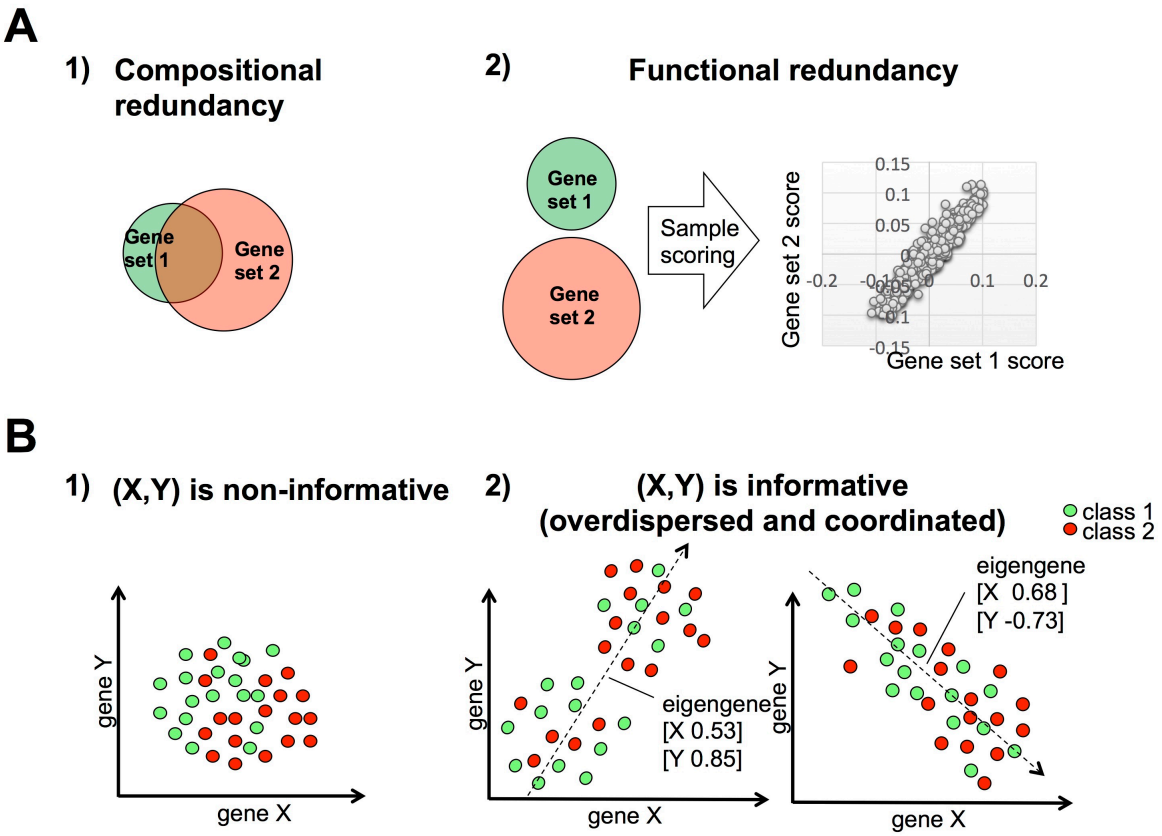

Figure1. Schematic explanation of the basic notions used in this study.

Panel A) schematically summarizes the two possible forms of redundancy between two gene sets. 1) Compositional redundancy corresponds to gene set overlap. 2) Functional redundancy represents instead different transcriptional read-outs of the same biological process and it is possible even for the gene sets with no overlap. Measuring functional redundancy depends on the way a sample is scored based on the expression of its genes and the chosen corpus of data. Panel B) explains the difference between non-informative (1) and informative (2) gene sets. A signature composed of two genes X and Y is here considered. The circles denote biological samples and the two colors correspond to two different labels : class 1 and class 2 (e.g., metastatic vs. primary tumors). Scatter plots are used to represent the expression values of gene X (X-axes) and gene Y (Y-axes) in each sample. Three types of samples distributions are shown. In 1) (isotropic case) no naturally distinguished axis in the points distribution, labeling of samples is needed to define their ranking. In 2) instead, it exists a distinguishable axis in the data distribution that allows a robust ranking of the samples independently on their labeling. Both second and third scenario leads to overdispersion and coordination of the corresponding gene set and are selected in the analysis.

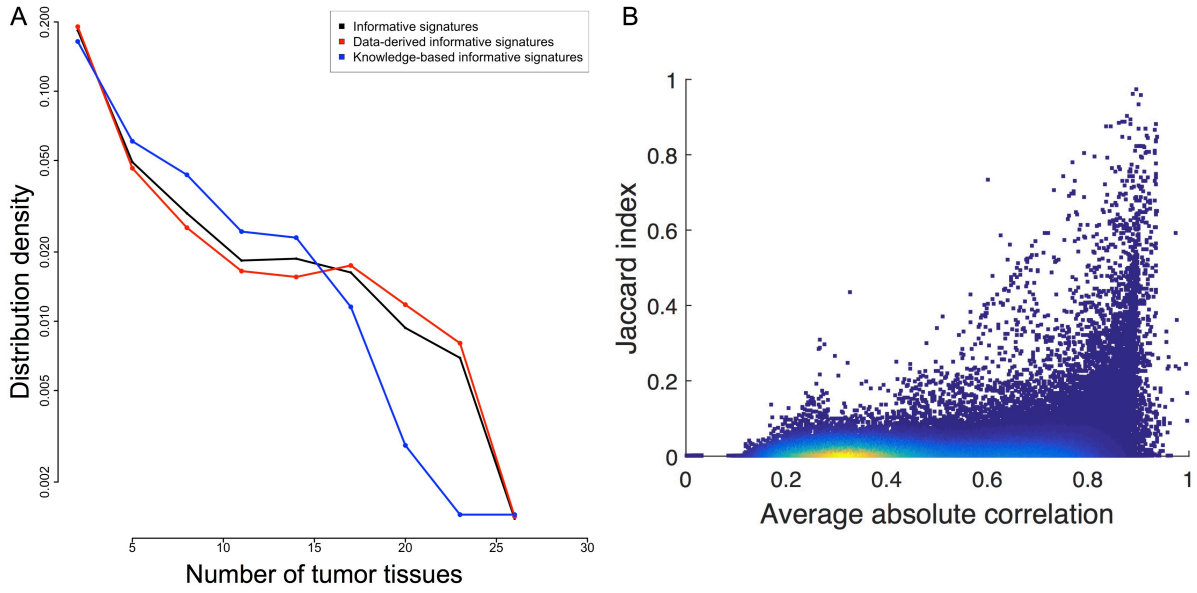

Figure 2. Behavior of the informative signatures across the different cancer transcriptomes. The main results regarding the pan-cancer behavior of the informative signatures are here summarized. In panel A) the distribution of the numbers of cancer types in which the informative signatures have ROMA L1 and L1/L2 p-values significant is reported. The behavior of all informative signatures is represented in black, that of data-driven and knowledge-based informative signatures is denoted in red and blue, respectively. In B) the dependence between informative gene sets overlap (Jaccard-index) and average correlation between the meta-samples defined by the informative gene sets is reported. Each point corresponds to an informative signature and their color is proportional to the points density : from red (high density) to blue (low density).

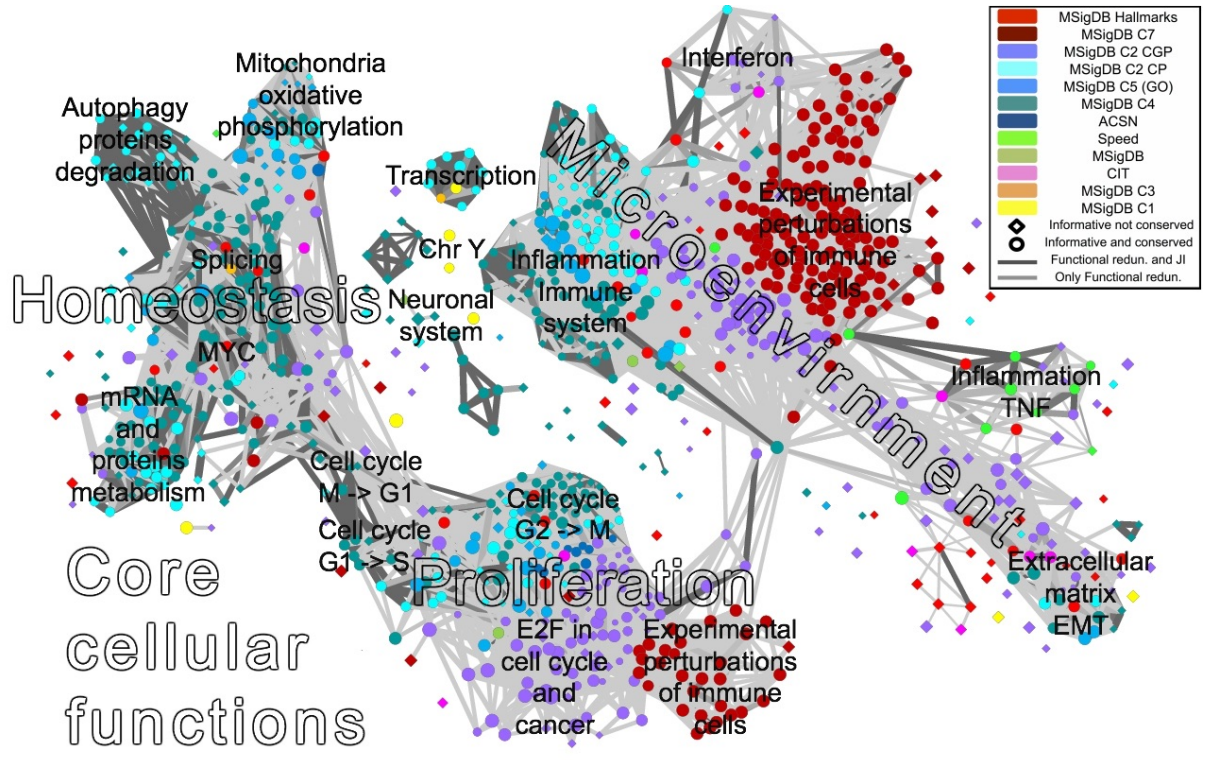

Figure 3. InfoSigMap: user-friendly interactive representation of the informative signatures.

The network map of the 962 informative signatures plus SPEED and Hallmarks is here reported as available on the website (http://navicell.curie.fr/pages/maps_avcorrmodulenet.html). The signatures are organized as nodes of the network. Nodes colors correspond to the different signatures categories, while the shape is a diamond for informative signatures and circular for those that are also conserved. The links correspond to redundancy between couples of gene sets (functional redundancy in light gray, dark gray if also the Jaccard-index intersection is significant). The names annotated on the top of the map denote areas of the network containing signatures associated to same biological function. The interactive on-line version of this map can be browsed as an instance of Google Maps, with the possibility of zooming in and /out, getting description of gene signatures, and visualizing data (various gene set scores, such as GSEA or ROMA scores) on top of the map.

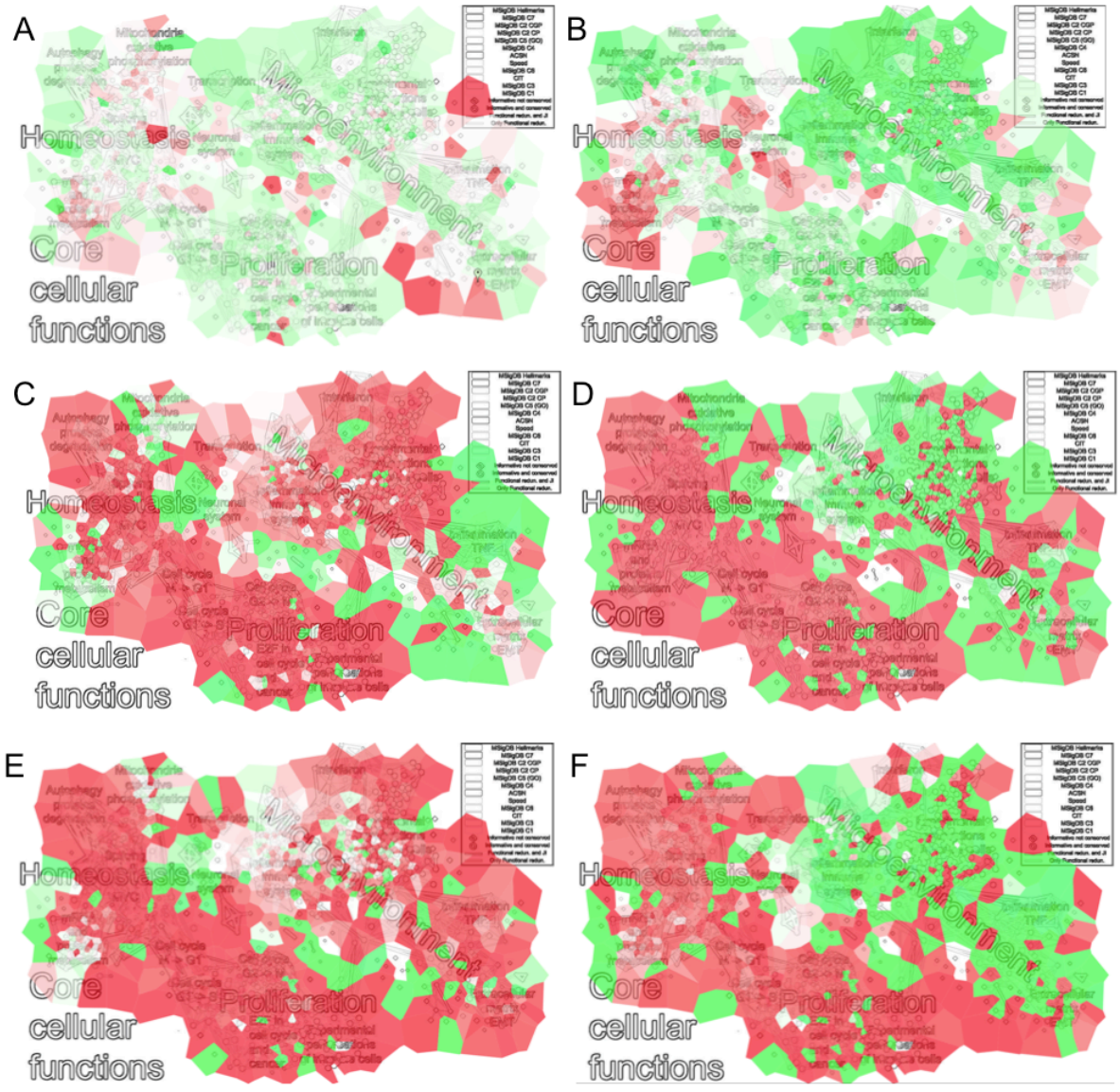

Figure 4. Results of InfoSigMap applied to some typical data analysis scenarios.

Four examples showing how InfoSigMap provides a more insightful interpretation of the lists of significant signatures output of the classical gene expression analysis tools is here proposed. The significant fold-changes resulting from the differential ROMA analysis are plotted on the top of the

map according to a heatmap coloring highlighting up- (red) and down-regulated (green) gene signatures. The plots are organized as follows : (a) KRAS mutated vs. wild type colorectal cancer; (b) metastatic vs. primary colon cancer and (c-f) tumor vs. normal tissue in cervix, colon, gastric and lung, respectively.

TABLES

Table 1. Contribution of each signature collection to the informative set. Number of informative signature, total dimension and the fraction of the previous two fields are reported for each signature collection.

Signatures collections	Data-Derived						Knowledge-based				H	C1
	SPEED	C4	BITON	CGP	C7	C6	ACSN	CP	C5	C3		
Informative signatures	6	270	11	233	185	1	8	150	64	9	15	10
Total	11	858	40	1698	2436	99	63	1330	1454	615	50	326
Fraction	**55%**	**31%**	**28%**	14%	8%	1%	**13%**	11%	4%	1%	**30%**	3%

Table2. Performances of the informative and Hallmark collection in respect to the full set of starting signatures in three test cases. The columns report for each case: the total number of signatures significant in the test, the Fisher P-value for the informative signatures, the Fisher P-value for the Hallmarks and the Kolmogorov-Smirnov P-value comparing the NES distribution of the informative vs Hallmarks.

Test cases	Total number of significant signatures	Informative signatures Fisher P-value	Hallmarks signatures Fisher P-value	KS P-value informative vs H NES distribution
(i) Colon KRAS mutated vs wild type	896	**1.1E-130**	1E-07	0.098
(ii) Colon metastasis vs primary	396	**2.5E-78**	1.6E-08	**0.02**
(iii) Normal vs	1845	**4.3E-161**	3.9E-06	**0.005**

Tumor cervix				
(iii) Normal vs Tumor lung	3290	**1.2E-214**	9.7E-05	**0.02**
(iii) Normal vs Tumor colon	2159	**1.9E-137**	2.6E-08	**0.001**
(iii) Normal vs Tumor gastric	2644	**4.6E-170**	4.3E-10	**0.005**

Supplementary Information

Pan-cancer classification of gene signatures for their information value and functional redundancy

Laura Cantini[1*], Laurence Calzone[1], Loredana Martignetti[1], Mattias Rydenfelt[2,3], Nils Blüthgen[2,3], Emmanuel Barillot[1], Andrei Zinovyev[1*]

1. Institut Curie, INSERM U900, PSL Research University, Mines ParisTech, 26, rue d'Ulm, F-75248 Paris, France.

2. Institute of Pathology, Charite Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany.

3. IRI Life Sciences and Institute for Theoretical Biology, Humboldt University, Philippstr. 13, Haus 18, 10115 Berlin, Germany

Correspondence : Laura Cantini (laura.cantini@curie.fr) and Andrei Zinovyev (andrei.zinovyev@curie.fr)

Supplementary contents

Table S1. Number of samples available for each transcriptomic dataset used for the analysis

tumor code	tumor name extended	number of samples
ACC	Adrenocortical carcinoma	78
BLCA	Bladder Urothelial Carcinoma	408
BRCA	Breast invasive carcinoma	1085
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma	303
CHOL	Cholangiocarcinoma	36
COAD	Colon adenocarcinoma	276
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	48
ESCA	Esophageal carcinoma	182
GBM	Glioblastoma multiforme	155
HNSC	Head and Neck squamous cell carcinoma	515
KICH	Kidney Chromophobe	65
KIRC	Kidney renal clear cell carcinoma	516
KIRP	Kidney renal papillary cell carcinoma	286
LGG	Brain Lower Grade Glioma	513
LIHC	Liver hepatocellular carcinoma	368
LUAD	Lung adenocarcinoma	510
LUSC	Lung squamous cell carcinoma	488
MESO	Mesothelioma	87
OV	Ovarian serous cystadenocarcinoma	302
PAAD	Pancreatic adenocarcinoma	177
PCPG	Pheochromocytoma and Paraganglioma	181
PRAD	Prostate adenocarcinoma	494
READ	Rectum adenocarcinoma	90
SARC	Sarcoma	255
SKCM	Skin Cutaneous Melanoma	103
STAD	Stomach adenocarcinoma	401
TGCT	Testicular Germ Cell Tumors	138
THCA	Thyroid carcinoma	501
THYM	Thymoma	120
UCEC	Uterine Corpus Endometrial Carcinoma	173
UCS	Uterine Carcinosarcoma	57
UVM	Uveal Melanoma	80

Table S2. List of informative signatures and number of cancer types in which they were found significant

Signature	Compendium of origin	Tumors in which is informative
E2F4_TARGETS	ACSN	15
E2F3_TARGETS	ACSN	12
S_CC_PHASE	ACSN	12
WEE	ACSN	10
E2F6_TARGETS	ACSN	8
E2F2_TARGETS	ACSN	5
E2F1_TARGETS	ACSN	4
MITOCH_METABOLISM	ACSN	4
CIT8_UP	CIT	27
CIT8_DN	CIT	26
CIT7_DN	CIT	25
CIT12_UP	CIT	18
CIT3_UP	CIT	16
CIT14_UP	CIT	7
CIT4_DN	CIT	4
CIT5_DN	CIT	4
CIT11_DN	CIT	3
CIT20_DN	CIT	2
CIT6_DN	CIT	2
chryq11	MSigDB C1	24
chryp11	MSigDB C1	18
chr5q31	MSigDB C1	6
chr16p13	MSigDB C1	3
chr6p22	MSigDB C1	3
chr17q25	MSigDB C1	2
chr19p13	MSigDB C1	2
chr1q21	MSigDB C1	2
chrxp11	MSigDB C1	2
chrxq28	MSigDB C1	2
KOBAYASHI_EGFR_SIGLING_6HR	MSigDB C2 CGP	25
FARMER_BREAST_CANCER_CLUSTER_2	MSigDB C2 CGP	23
CHANG_CYCLING_GENES	MSigDB C2 CGP	20
WALLACE_PROSTATE_CANCER_RACE	MSigDB C2 CGP	20
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_6HR	MSigDB C2 CGP	20
KONG_E2F1_TARGETS	MSigDB C2 CGP	19
MCLACHLAN_DENTAL_CARIES	MSigDB C2 CGP	19
TURASHVILI_BREAST_CARCINOMA_DUCTAL_VS_LOBULAR	MSigDB C2 CGP	19
ZHANG_TLX_TARGETS_60HR	MSigDB C2 CGP	19
CROONQUIST_IL6_DEPRIVATION	MSigDB C2 CGP	18
GUTIERREZ_WALDENSTROEMS_MACROGLOBULINEMIA_2	MSigDB C2 CGP	18
HOFMANN_MYELODYSPLASTIC_SYNDROM_RISK	MSigDB C2 CGP	18
LEE_DIFFERENTIATING_T_LYMPHOCYTE	MSigDB C2 CGP	17
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE	MSigDB C2 CGP	17
TURASHVILI_BREAST_LOBULAR_CARCINOMA_VS_LOBULAR_NORMAL	MSigDB C2 CGP	16
BOSCO_TH1_CYTOTOXIC_MODULE	MSigDB C2 CGP	15
RPS14_DN.V1	MSigDB C2 CGP	15
SHIPP_DLBCL_CURED_VS_FATAL	MSigDB C2 CGP	15
CHUNG_BLISTER_CYTOTOXICITY	MSigDB C2 CGP	14
GRAHAM_NORMAL_QUIESCENT_VS_NORMAL_DIVIDING	MSigDB C2 CGP	14
POMEROY_MEDULLOBLASTOMA_DESMOPLASIC_VS_CLASSIC	MSigDB C2 CGP	14
PUJA_BREAST_CANCER_LIT_INT_NETWORK	MSigDB C2 CGP	14
SMID_BREAST_CANCER_NORMAL_LIKE	MSigDB C2 CGP	14
VILIMAS_NOTCH1_TARGETS	MSigDB C2 CGP	14
JISON_SICKLE_CELL_DISEASE	MSigDB C2 CGP	13
WONG_MITOCHONDRIA_GENE_MODULE	MSigDB C2 CGP	13
JAATINEN_HEMATOPOIETIC_STEM_CELL	MSigDB C2 CGP	12
KOBAYASHI_EGFR_SIGLING_24HR	MSigDB C2 CGP	12
UROSEVIC_RESPONSE_TO_IMIQUIMOD	MSigDB C2 CGP	12
BILANGES_SERUM_AND_RAPAMYCIN_SENSITIVE_GENES	MSigDB C2 CGP	11
CHOI_ATL_CHRONIC_VS_ACUTE	MSigDB C2 CGP	11
FIK_BREAST_CANCER_SDPP_SIGTURE	MSigDB C2 CGP	11
GAURNIER_PSMD4_TARGETS	MSigDB C2 CGP	11
HOFFMAN_CLOCK_TARGETS	MSigDB C2 CGP	11
SA_RESPONSE_TO_IFNG	MSigDB C2 CGP	11
TURASHVILI_BREAST_NORMAL_DUCTAL_VS_LOBULAR	MSigDB C2 CGP	11
ZHAN_MULTIPLE_MYELOMA	MSigDB C2 CGP	11
HOLLEMAN_ASPARAGISE_RESISTANCE_ALL	MSigDB C2 CGP	10
ICHIBA_GRAFT_VERSUS_HOST_DISEASE_D7	MSigDB C2 CGP	10
JIANG_HYPOXIA_VIA_VHL	MSigDB C2 CGP	10
KAUFFMANN_MELANOMA_RELAPSE	MSigDB C2 CGP	10
PIEPOLI_LGI1_TARGETS	MSigDB C2 CGP	10
RUTELLA_RESPONSE_TO_HGF	MSigDB C2 CGP	10
RUTELLA_RESPONSE_TO_HGF_VS_CSF2RB_AND_IL4	MSigDB C2 CGP	10
BENPORATH_PROLIFERATION	MSigDB C2 CGP	9
BOWIE_RESPONSE_TO_EXTRACELLULAR_MATRIX	MSigDB C2 CGP	9
CROONQUIST_NRAS_SIGLING	MSigDB C2 CGP	9
HORTON_SREBF_TARGETS	MSigDB C2 CGP	9
HSIAO_HOUSEKEEPING_GENES	MSigDB C2 CGP	9
MARSON_FOXP3_TARGETS	MSigDB C2 CGP	9
QI_PLASMACYTOMA	MSigDB C2 CGP	9
RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4	MSigDB C2 CGP	9
BROWNE_INTERFERON_RESPONSIVE_GENES	MSigDB C2 CGP	8
GRAHAM_CML_DIVIDING_VS_NORMAL_DIVIDING	MSigDB C2 CGP	8
LI_WILMS_TUMOR_APLASTIC	MSigDB C2 CGP	8
PICCALUGA_ANGIOIMMUNOBLASTIC_LYMPHOMA	MSigDB C2 CGP	8
RODWELL_AGING_KIDNEY	MSigDB C2 CGP	8
RORIE_TARGETS_OF_EWSR1_FLI1_FUSION	MSigDB C2 CGP	8
WEBER_METHYLATED_LCP_IN_SPERM	MSigDB C2 CGP	8
YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_12	MSigDB C2 CGP	8
FUJII_YBX1_TARGETS	MSigDB C2 CGP	7
FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS	MSigDB C2 CGP	7
GABRIELY_MIR21_TARGETS	MSigDB C2 CGP	7
GRADE_METASTASIS	MSigDB C2 CGP	7
MATTIOLI_MULTIPLE_MYELOMA_WITH_14Q32_TRANSLOCATIONS	MSigDB C2 CGP	7
MOLEAR_TARGETS_OF_CCND1_AND_CDK4	MSigDB C2 CGP	7
MORI_IMMATURE_B_LYMPHOCYTE	MSigDB C2 CGP	7
ONO_AML1_TARGETS	MSigDB C2 CGP	7
AMUNDSON_GAMMA_RADIATION_RESISTANCE	MSigDB C2 CGP	6
ASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGTURE	MSigDB C2 CGP	6
BOYLAN_MULTIPLE_MYELOMA_PCA3	MSigDB C2 CGP	6
BROWN_MYELOID_CELL_DEVELOPMENT	MSigDB C2 CGP	6
DIAZ_CHRONIC_MEYLOGENOUS_LEUKEMIA	MSigDB C2 CGP	6
DISTECHE_ESCAPED_FROM_X_ICTIVATION	MSigDB C2 CGP	6
FERREIRA_EWINGS_SARCOMA_UNSTABLE_VS_STABLE	MSigDB C2 CGP	6
GOLDRATH_ANTIGEN_RESPONSE	MSigDB C2 CGP	6
HOFFMANN_PRE_BI_TO_LARGE_PRE_BII_LYMPHOCYTE	MSigDB C2 CGP	6
HONMA_DOCETAXEL_RESISTANCE	MSigDB C2 CGP	6
LINDGREN_BLADDER_CANCER_CLUSTER_2B	MSigDB C2 CGP	6

MALO_HYPOXIA	MSigDB C2 CGP	6
MOOTHA_HUMAN_MITODB_6_2002	MSigDB C2 CGP	6
STARK_PREFRONTAL_CORTEX_22Q11_DELETION	MSigDB C2 CGP	6
TSUTSUMI_FBXW8_TARGETS	MSigDB C2 CGP	6
CLASPER_LYMPHATIC_VESSELS_DURING_METASTASIS	MSigDB C2 CGP	5
DAVICIONI_MOLECULAR_ARMS_VS_ERMS	MSigDB C2 CGP	5
GRATIAS_RETINOBLASTOMA_16Q24	MSigDB C2 CGP	5
KIM_RESPONSE_TO_TSA_AND_DECITABINE	MSigDB C2 CGP	5
LEOUR_DENDRITIC_CELL_MATURATION	MSigDB C2 CGP	5
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_8D	MSigDB C2 CGP	5
VERHAAK_GLIOBLASTOMA_MESENCHYMAL	MSigDB C2 CGP	5
ASTON_MAJOR_DEPRESSIVE_DISORDER	MSigDB C2 CGP	4
CAIRO_HEPATOBLASTOMA_CLASSES	MSigDB C2 CGP	4
CHEMELLO_SOLEUS_VS_EDL_MYOFIBERS	MSigDB C2 CGP	4
DACOSTA_UV_RESPONSE_VIA_ERCC3_COMMON	MSigDB C2 CGP	4
FAELT_B_CLL_WITH_VH_REARRANGEMENTS	MSigDB C2 CGP	4
GASHIMA_NRG1_SIGLING	MSigDB C2 CGP	4
GRADE_COLON_VS_RECTAL_CANCER	MSigDB C2 CGP	4
HALMOS_CEBPA_TARGETS	MSigDB C2 CGP	4
JACKSONMT1_TARGETS	MSigDB C2 CGP	4
LIU_PROSTATE_CANCER	MSigDB C2 CGP	4
MORI_LARGE_PRE_BII_LYMPHOCYTE	MSigDB C2 CGP	4
NIKOLSKY_BREAST_CANCER_19P13_AMPLICON	MSigDB C2 CGP	4
PENG_RAPAMYCIN_RESPONSE	MSigDB C2 CGP	4
SENGUPTA_SOPHARYNGEAL_CARCINOMA	MSigDB C2 CGP	4
STAMBOLSKY_RESPONSE_TO_VITAMIN_D3	MSigDB C2 CGP	4
VALK_AML_CLUSTER_4	MSigDB C2 CGP	4
WHITFIELD_CELL_CYCLE_S	MSigDB C2 CGP	4
BOYLAN_MULTIPLE_MYELOMA_C_D	MSigDB C2 CGP	3
BURTON_ADIPOGENESIS_7	MSigDB C2 CGP	3
CHNG_MULTIPLE_MYELOMA_HYPERPLOID	MSigDB C2 CGP	3
DING_LUNG_CANCER_MUTATED_FREQUENTLY	MSigDB C2 CGP	3
FEVR_CTNNB1_TARGETS	MSigDB C2 CGP	3
FURUKAWA_DUSP6_TARGETS_PCI35	MSigDB C2 CGP	3
GINESTIER_BREAST_CANCER_ZNF217_AMPLIFIED	MSigDB C2 CGP	3
GRAHAM_CML_DIVIDING_VS_NORMAL_QUIESCENT	MSigDB C2 CGP	3
GUO_HEX_TARGETS	MSigDB C2 CGP	3
HELLER_SILENCED_BY_METHYLATION	MSigDB C2 CGP	3
HISTONE_ACETYLTRANSFERASE_ACTIVITY	MSigDB C2 CGP	3
JAEGER_METASTASIS	MSigDB C2 CGP	3
JECHLINGER_EPITHELIAL_TO_MESENCHYMAL_TRANSITION	MSigDB C2 CGP	3
KAYAMA_SOFT_TISSUE_TUMORS_PCA1	MSigDB C2 CGP	3
KIM_ALL_DISORDERS_OLIGODENDROCYTE_NUMBER_CORR	MSigDB C2 CGP	3
LEE_BMP2_TARGETS	MSigDB C2 CGP	3
LI_INDUCED_T_TO_TURAL_KILLER	MSigDB C2 CGP	3
LUI_THYROID_CANCER_PAX8_PPARG	MSigDB C2 CGP	3
MITSIADES_RESPONSE_TO_APLIDIN	MSigDB C2 CGP	3
MONTERO_THYROID_CANCER_POOR_SURVIVAL	MSigDB C2 CGP	3
MORI_PRE_BI_LYMPHOCYTE	MSigDB C2 CGP	3
OSMAN_BLADDER_CANCER	MSigDB C2 CGP	3
PUJA_BRCA_CENTERED_NETWORK	MSigDB C2 CGP	3
REN_ALVEOLAR_RHABDOMYOSARCOMA	MSigDB C2 CGP	3
RHEIN_ALL_GLUCOCORTICOID_THERAPY	MSigDB C2 CGP	3
RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_POORLY	MSigDB C2 CGP	3
ROETH_TERT_TARGETS	MSigDB C2 CGP	3
ROSS_AML_OF_FAB_M7_TYPE	MSigDB C2 CGP	3
SCHLOSSER_MYC_AND_SERUM_RESPONSE_SYNERGY	MSigDB C2 CGP	3
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D	MSigDB C2 CGP	3
TAYLOR_METHYLATED_IN_ACUTE_LYMPHOBLASTIC_LEUKEMIA	MSigDB C2 CGP	3
TIEN_INTESTINE_PROBIOTICS_24HR	MSigDB C2 CGP	3
TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC	MSigDB C2 CGP	3
VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY	MSigDB C2 CGP	3
VERHAAK_AML_WITH_NPM1_MUTATED	MSigDB C2 CGP	3
WANG_CLIM2_TARGETS	MSigDB C2 CGP	3
WANG_RESPONSE_TO_GSK3_INHIBITOR_SB216763	MSigDB C2 CGP	3
WANG_SMARCE1_TARGETS	MSigDB C2 CGP	3
WUNDER_INFLAMMATORY_RESPONSE_AND_CHOLESTEROL	MSigDB C2 CGP	3
AIGNER_ZEB1_TARGETS	MSigDB C2 CGP	2
AMIT_SERUM_RESPONSE_480_MCF10A	MSigDB C2 CGP	2
BALDWIN_PRKCI_TARGETS	MSigDB C2 CGP	2
BLUM_RESPONSE_TO_SALIRASIB	MSigDB C2 CGP	2
BOQUEST_STEM_CELL	MSigDB C2 CGP	2
BOYAULT_LIVER_CANCER_SUBCLASS_G3	MSigDB C2 CGP	2
CAHOY_NEURONAL	MSigDB C2 CGP	2
CAIRO_HEPATOBLASTOMA	MSigDB C2 CGP	2
CASORELLI_ACUTE_PROMYELOCYTIC_LEUKEMIA	MSigDB C2 CGP	2
CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL	MSigDB C2 CGP	2
CHIANG_LIVER_CANCER_SUBCLASS_CTNNB1	MSigDB C2 CGP	2
CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION	MSigDB C2 CGP	2
CHIANG_LIVER_CANCER_SUBCLASS_UNNOTATED	MSigDB C2 CGP	2
CHIBA_RESPONSE_TO_TSA	MSigDB C2 CGP	2
COLDREN_GEFITINIB_RESISTANCE	MSigDB C2 CGP	2
CREIGHTON_AKT1_SIGLING_VIA_MTOR	MSigDB C2 CGP	2
CUI_TCF21_TARGETS	MSigDB C2 CGP	2
DAIRKEE_CANCER_PRONE_RESPONSE_E2	MSigDB C2 CGP	2
DAUER_STAT3_TARGETS	MSigDB C2 CGP	2
DELASER_MYOD_TARGETS	MSigDB C2 CGP	2
DELYS_THYROID_CANCER	MSigDB C2 CGP	2
DUTERTRE_ESTRADIOL_RESPONSE_24HR	MSigDB C2 CGP	2
EBAUER_MYOGENIC_TARGETS_OF_PAX3_FOXO1_FUSION	MSigDB C2 CGP	2
FAELT_B_CLL_WITH_VH3_21	MSigDB C2 CGP	2
FARMER_BREAST_CANCER_CLUSTER_4	MSigDB C2 CGP	2
FOURNIER_ACIR_DEVELOPMENT_LATE_2	MSigDB C2 CGP	2
FRASOR_RESPONSE_TO_SERM_OR_FULVESTRANT	MSigDB C2 CGP	2
GAUSSMANN_MLL_AF4_FUSION_TARGETS_E	MSigDB C2 CGP	2
GAVIN_FOXP3_TARGETS_CLUSTER_P4	MSigDB C2 CGP	2
GENTILE_UV_HIGH_DOSE	MSigDB C2 CGP	2
GENTILE_UV_RESPONSE_CLUSTER_D4	MSigDB C2 CGP	2
GRAHAM_CML_QUIESCENT_VS_NORMAL_QUIESCENT	MSigDB C2 CGP	2
HADDAD_T_LYMPHOCYTE_AND_NK_PROGENITOR	MSigDB C2 CGP	2
HOEGERKORP_CD44_TARGETS_DIRECT	MSigDB C2 CGP	2
HOEGERKORP_CD44_TARGETS_TEMPORAL	MSigDB C2 CGP	2
HOLLEMAN_VINCRISTINE_RESISTANCE_B_ALL	MSigDB C2 CGP	2
HORIUCHI_WTAP_TARGETS	MSigDB C2 CGP	2
HOXA9_DN.V1	MSigDB C2 CGP	2
ICHIBA_GRAFT_VERSUS_HOST_DISEASE_35D	MSigDB C2 CGP	2
KAUFFMANN_D_REPLICATION_GENES	MSigDB C2 CGP	2
KIM_ALL_DISORDERS_DURATION_CORR	MSigDB C2 CGP	2
KIM_GLIS2_TARGETS	MSigDB C2 CGP	2
LEE_LIVER_CANCER	MSigDB C2 CGP	2
LEE_LIVER_CANCER_SURVIVAL	MSigDB C2 CGP	2
LIEN_BREAST_CARCINOMA_METAPLASTIC	MSigDB C2 CGP	2
LIM_MAMMARY_LUMIL_MATURE	MSigDB C2 CGP	2
LIN_APC_TARGETS	MSigDB C2 CGP	2
LINDGREN_BLADDER_CANCER_CLUSTER_1	MSigDB C2 CGP	2

LINDGREN_BLADDER_CANCER_CLUSTER_3	MSigDB C2 CGP	2
LINDSTEDT_DENDRITIC_CELL_MATURATION_D	MSigDB C2 CGP	2
LIU_VAV3_PROSTATE_CARCINOGENESIS	MSigDB C2 CGP	2
LUI_THYROID_CANCER_CLUSTER_3	MSigDB C2 CGP	2
MA_MYELOID_DIFFERENTIATION	MSigDB C2 CGP	2
MARSON_FOXP3_TARGETS_STIMULATED	MSigDB C2 CGP	2
MORI_MATURE_B_LYMPHOCYTE	MSigDB C2 CGP	2
MULLIGHAN_MLL_SIGTURE_1	MSigDB C2 CGP	2
NIELSEN_MALIGT_FIBROUS_HISTIOCYTOMA	MSigDB C2 CGP	2
NIKOLSKY_BREAST_CANCER_15Q26_AMPLICON	MSigDB C2 CGP	2
ONDER_CDH1_TARGETS_2	MSigDB C2 CGP	2
OXFORD_RALA_OR_RALB_TARGETS	MSigDB C2 CGP	2
PARK_APL_PATHOGENESIS	MSigDB C2 CGP	2
PARK_TRETINOIN_RESPONSE_AND_PML_RARA_FUSION	MSigDB C2 CGP	2
PATIL_LIVER_CANCER	MSigDB C2 CGP	2
PELLICCIOTTA_HDAC_IN_ANTIGEN_PRESENTATION	MSigDB C2 CGP	2
PENG_LEUCINE_DEPRIVATION	MSigDB C2 CGP	2
PHONG_TNF_TARGETS	MSigDB C2 CGP	2
PLASARI_TGFB1_TARGETS_10HR	MSigDB C2 CGP	2
POMEROY_MEDULLOBLASTOMA_PROGNOSIS	MSigDB C2 CGP	2
PYEON_HPV_POSITIVE_TUMORS	MSigDB C2 CGP	2
RADAEVA_RESPONSE_TO_IF1	MSigDB C2 CGP	2
RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_MODERATELY	MSigDB C2 CGP	2
SABATES_COLORECTAL_ADENOMA	MSigDB C2 CGP	2
SANSOM_APC_TARGETS	MSigDB C2 CGP	2
SARRIO_EPITHELIAL_MESENCHYMAL_TRANSITION	MSigDB C2 CGP	2
SCHLINGEMANN_SKIN_CARCINOGENESIS_TPA	MSigDB C2 CGP	2
SCIAN_CELL_CYCLE_TARGETS_OF_TP53_AND_TP73	MSigDB C2 CGP	2
SMID_BREAST_CANCER_LUMIL_B	MSigDB C2 CGP	2
SONG_TARGETS_OF_IE86_CMV_PROTEIN	MSigDB C2 CGP	2
SWEET_KRAS_TARGETS	MSigDB C2 CGP	2
TENEDINI_MEGAKARYOCYTE_MARKERS	MSigDB C2 CGP	2
TIAN_TNF_SIGLING_VIA_NFKB	MSigDB C2 CGP	2
TSENG_IRS1_TARGETS	MSigDB C2 CGP	2
WHITFIELD_CELL_CYCLE_G2_M	MSigDB C2 CGP	2
WILCOX_RESPONSE_TO_PROGESTERONE	MSigDB C2 CGP	2
WU_HBX_TARGETS_1	MSigDB C2 CGP	2
XU_HGF_TARGETS_INDUCED_BY_AKT1_48HR	MSigDB C2 CGP	2
ZWANG_CLASS_2_TRANSIENTLY_INDUCED_BY_EGF	MSigDB C2 CGP	2
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS	MSigDB C2 CP	28
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	MSigDB C2 CP	26
REACTOME_METABOLISM_OF_MRNA	MSigDB C2 CP	22
REACTOME_METABOLISM_OF_RNA	MSigDB C2 CP	22
KEGG_ALLOGRAFT_REJECTION	MSigDB C2 CP	20
KEGG_TYPE_I_DIABETES_MELLITUS	MSigDB C2 CP	20
REACTOME_TRANSLATION	MSigDB C2 CP	20
KEGG_AUTOIMMUNE_THYROID_DISEASE	MSigDB C2 CP	19
KEGG_OXIDATIVE_PHOSPHORYLATION	MSigDB C2 CP	19
REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_REGULATION	MSigDB C2 CP	18
REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION	MSigDB C2 CP	18
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	MSigDB C2 CP	18
BIOCARTA_CTLA4_PATHWAY	MSigDB C2 CP	17
KEGG_GRAFT_VERSUS_HOST_DISEASE	MSigDB C2 CP	17
REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	MSigDB C2 CP	17
REACTOME_INFLUENZA_LIFE_CYCLE	MSigDB C2 CP	17
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	MSigDB C2 CP	17
REACTOME_TCR_SIGNALING	MSigDB C2 CP	17
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	MSigDB C2 CP	16
REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY	MSigDB C2 CP	16
REACTOME_MRNA_SPLICING_MINOR_PATHWAY	MSigDB C2 CP	16
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX	MSigDB C2 CP	16
REACTOME_PEPTIDE_CHAIN_ELONGATION	MSigDB C2 CP	16
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	MSigDB C2 CP	15
REACTOME_G2_M_CHECKPOINTS	MSigDB C2 CP	15
REACTOME_VIF_MEDIATED_DEGRADATION_OF_APOBEC3G	MSigDB C2 CP	15
PID_PLK1_PATHWAY	MSigDB C2 CP	14
PID_TCR_PATHWAY	MSigDB C2 CP	14
KEGG_RIBOSOME	MSigDB C2 CP	13
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	MSigDB C2 CP	13
REACTOME_CELL_CYCLE_MITOTIC	MSigDB C2 CP	13
REACTOME_RNA_POL_I_PROMOTER_OPENING	MSigDB C2 CP	13
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1	MSigDB C2 CP	13
BIOCARTA_NO2IL12_PATHWAY	MSigDB C2 CP	12
PID_IL12_2PATHWAY	MSigDB C2 CP	12
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	MSigDB C2 CP	12
REACTOME_MITOTIC_M_M_G1_PHASES	MSigDB C2 CP	12
KEGG_PARKINSONS_DISEASE	MSigDB C2 CP	11
PID_CD8_TCR_PATHWAY	MSigDB C2 CP	11
REACTOME_CDK_MEDIATED_PHOSPHORYLATION_AND_REMOVAL_OF_CDC6	MSigDB C2 CP	11
REACTOME_DNA_REPLICATION	MSigDB C2 CP	11
REACTOME_FORMATION_OF_ATP_BY_CHEMIOSMOTIC_COUPLING	MSigDB C2 CP	11
REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES	MSigDB C2 CP	11
REACTOME_PD1_SIGNALING	MSigDB C2 CP	11
KEGG_CELL_CYCLE	MSigDB C2 CP	10
KEGG_SPLICEOSOME	MSigDB C2 CP	10
PID_AURORA_B_PATHWAY	MSigDB C2 CP	10
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1	MSigDB C2 CP	10
REACTOME_MITOTIC_PROMETAPHASE	MSigDB C2 CP	10
REACTOME_PHOSPHORYLATION_OF_CD3_AND_TCR_ZETA_CHAINS	MSigDB C2 CP	10
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	MSigDB C2 CP	9
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX	MSigDB C2 CP	9
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0	MSigDB C2 CP	9
REACTOME_DNA_STRAND_ELONGATION	MSigDB C2 CP	9
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	MSigDB C2 CP	9
KEGG_HUNTINGTONS_DISEASE	MSigDB C2 CP	8
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	MSigDB C2 CP	8
KEGG_PROTEASOME	MSigDB C2 CP	8
PID_ATR_PATHWAY	MSigDB C2 CP	8
PID_IL12_STAT4_PATHWAY	MSigDB C2 CP	8
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING	MSigDB C2 CP	8
REACTOME_MITOCHONDRIAL_PROTEIN_IMPORT	MSigDB C2 CP	8
REACTOME_MRNA_SPLICING	MSigDB C2 CP	8
REACTOME_REGULATION_OF_ORNITHINE_DECARBOXYLASE_ODC	MSigDB C2 CP	8
REACTOME_SIGNALING_BY_WNT	MSigDB C2 CP	8
BIOCARTA_IL12_PATHWAY	MSigDB C2 CP	7
BIOCARTA_PROTEASOME_PATHWAY	MSigDB C2 CP	7
KEGG_CELL_ADHESION_MOLECULES_CAMS	MSigDB C2 CP	7
KEGG_LEISHMANIA_INFECTION	MSigDB C2 CP	7
REACTOME_AUTODEGRADATION_OF_CDH1_BY_CDH1_APC_C	MSigDB C2 CP	7
REACTOME_CELL_CYCLE	MSigDB C2 CP	7
REACTOME_G1_S_TRANSITION	MSigDB C2 CP	7
REACTOME_INTERFERON_GAMMA_SIGNALING	MSigDB C2 CP	7
REACTOME_PACKAGING_OF_TELOMERE_ENDS	MSigDB C2 CP	7
BIOCARTA_CTL_PATHWAY	MSigDB C2 CP	6
BIOCARTA_TCYTOTOXIC_PATHWAY	MSigDB C2 CP	6

KEGG_HEMATOPOIETIC_CELL_LINEAGE	MSigDB C2 CP	6
NABA_CORE_MATRISOME	MSigDB C2 CP	6
REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	MSigDB C2 CP	6
REACTOME_INTERFERON_SIGNALING	MSigDB C2 CP	6
REACTOME_MITOTIC_G1_G1_S_PHASES	MSigDB C2 CP	6
REACTOME_OLFACTORY_SIGNALING_PATHWAY	MSigDB C2 CP	6
BIOCARTA_TCAPOPTOSIS_PATHWAY	MSigDB C2 CP	5
BIOCARTA_THELPER_PATHWAY	MSigDB C2 CP	5
KEGG_CHEMOKINE_SIGNALING_PATHWAY	MSigDB C2 CP	5
REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION	MSigDB C2 CP	5
REACTOME_IMMUNE_SYSTEM	MSigDB C2 CP	5
REACTOME_MEIOSIS	MSigDB C2 CP	5
REACTOME_METABOLISM_OF_PROTEINS	MSigDB C2 CP	5
REACTOME_RNA_POL_I_RNA_POL_III_AND_MITOCHONDRIAL_TRANSCRIPTION	MSigDB C2 CP	5
REACTOME_S_PHASE	MSigDB C2 CP	5
BIOCARTA_MCM_PATHWAY	MSigDB C2 CP	4
BIOCARTA_NKT_PATHWAY	MSigDB C2 CP	4
KEGG_ASTHMA	MSigDB C2 CP	4
KEGG_PRIMARY_IMMUNODEFICIENCY	MSigDB C2 CP	4
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	MSigDB C2 CP	4
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	MSigDB C2 CP	4
PID_CD8_TCR_DOWNSTREAM_PATHWAY	MSigDB C2 CP	4
PID_FANCONI_PATHWAY	MSigDB C2 CP	4
PID_FOXM1_PATHWAY	MSigDB C2 CP	4
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS	MSigDB C2 CP	4
REACTOME_CELL_CYCLE_CHECKPOINTS	MSigDB C2 CP	4
REACTOME_MEIOTIC_RECOMBINATION	MSigDB C2 CP	4
REACTOME_SMOOTH_MUSCLE_CONTRACTION	MSigDB C2 CP	4
REACTOME_TRANSCRIPTION	MSigDB C2 CP	4
ST_T_CELL_SIGNAL_TRANSDUCTION	MSigDB C2 CP	4
BIOCARTA_CSK_PATHWAY	MSigDB C2 CP	3
BIOCARTA_TCRA_PATHWAY	MSigDB C2 CP	3
KEGG_DNA_REPLICATION	MSigDB C2 CP	3
REACTOME_ACTIVATION_OF_NF_KAPPAB_IN_B_CELLS	MSigDB C2 CP	3
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S	MSigDB C2 CP	3
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	MSigDB C2 CP	3
REACTOME_CYCLIN_A_B1_ASSOCIATED_EVENTS_DURING_G2_M_TRANSITION	MSigDB C2 CP	3
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM	MSigDB C2 CP	3
REACTOME_DOWNSTREAM_TCR_SIGNALING	MSigDB C2 CP	3
REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUENTLY_THE_43S_COMPLEX	MSigDB C2 CP	3
REACTOME_IL_3_5_AND_GM_CSF_SIGNALING	MSigDB C2 CP	3
REACTOME_P53_DEPENDENT_G1_DNA_DAMAGE_RESPONSE	MSigDB C2 CP	3
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	MSigDB C2 CP	3
REACTOME_REGULATION_OF_APOPTOSIS	MSigDB C2 CP	3
REACTOME_RNA_POL_I_TRANSCRIPTION	MSigDB C2 CP	3
REACTOME_SCFSKP2_MEDIATED_DEGRADATION_OF_P27_P21	MSigDB C2 CP	3
REACTOME_STRIATED_MUSCLE_CONTRACTION	MSigDB C2 CP	3
REACTOME_TRNA_AMINOACYLATION	MSigDB C2 CP	3
BIOCARTA_TH1TH2_PATHWAY	MSigDB C2 CP	2
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	MSigDB C2 CP	2
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	MSigDB C2 CP	2
KEGG_VIRAL_MYOCARDITIS	MSigDB C2 CP	2
NABA_ECM_GLYCOPROTEINS	MSigDB C2 CP	2
PID_CXCR4_PATHWAY	MSigDB C2 CP	2
REACTOME_ABORTIVE_ELONGATION_OF_HIV1_TRANSCRIPT_IN_THE_ABSENCE_OF_TAT	MSigDB C2 CP	2
REACTOME_AMYLOIDS	MSigDB C2 CP	2
REACTOME_BIOLOGICAL_OXIDATIONS	MSigDB C2 CP	2
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	MSigDB C2 CP	2
REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS	MSigDB C2 CP	2
REACTOME_DEFENSINS	MSigDB C2 CP	2
REACTOME_ER_PHAGOSOME_PATHWAY	MSigDB C2 CP	2
REACTOME_GENERIC_TRANSCRIPTION_PATHWAY	MSigDB C2 CP	2
REACTOME_IL_2_SIGNALING	MSigDB C2 CP	2
REACTOME_INNATE_IMMUNE_SYSTEM	MSigDB C2 CP	2
REACTOME_LAGGING_STRAND_SYNTHESIS	MSigDB C2 CP	2
REACTOME_M_G1_TRANSITION	MSigDB C2 CP	2
REACTOME_MRNA_PROCESSING	MSigDB C2 CP	2
REACTOME_MUSCLE_CONTRACTION	MSigDB C2 CP	2
REACTOME_NEURONAL_SYSTEM	MSigDB C2 CP	2
REACTOME_ORC1_REMOVAL_FROM_CHROMATIN	MSigDB C2 CP	2
REACTOME_PHOSPHORYLATION_OF_THE_APC_C	MSigDB C2 CP	2
REACTOME_SIGNALING_BY_ILS	MSigDB C2 CP	2
REACTOME_SYNTHESIS_OF_DNA	MSigDB C2 CP	2
REACTOME_UNWINDING_OF_DNA	MSigDB C2 CP	2
GAZDA_DIAMOND_BLACKFAN_ANEMIA_PROGENITOR	MSigDB C3	11
KRCTCNNNNMANAGC_UNKNOWN	MSigDB C3	5
RCGCANGCGY_V$NRF1_Q6	MSigDB C3	3
V$ELK1_02	MSigDB C3	3
V$NRF2_01	MSigDB C3	3
V$NRSF_01	MSigDB C3	3
ACCAATC,MIR-509	MSigDB C3	2
GGAANCGGAANY_UNKNOWN	MSigDB C3	2
V$GABP_B	MSigDB C3	2
GNF2_EIF3S6	MSigDB C4	27
MORF_TPT1	MSigDB C4	27
GNF2_CCNB2	MSigDB C4	26
MORF_ACTG1	MSigDB C4	26
GNF2_BUB1B	MSigDB C4	25
GNF2_CDC20	MSigDB C4	25
GNF2_HMMR	MSigDB C4	25
GNF2_PCNA	MSigDB C4	25
GNF2_ST13	MSigDB C4	25
MORF_NPM1	MSigDB C4	25
GCM_NPM1	MSigDB C4	24
GNF2_CCNA2	MSigDB C4	24
GNF2_FBL	MSigDB C4	24
GNF2_GLTSCR2	MSigDB C4	24
GNF2_RRM2	MSigDB C4	24
MORF_NME2	MSigDB C4	24
GNF2_CASP1	MSigDB C4	23
GNF2_CD7	MSigDB C4	23
GNF2_CENPE	MSigDB C4	23
GNF2_MCM4	MSigDB C4	23
GNF2_TTK	MSigDB C4	23
MORF_RAN	MSigDB C4	23
GCM_TPT1	MSigDB C4	22
GNF2_CDC2	MSigDB C4	22
GNF2_CKS2	MSigDB C4	22
GNF2_FEN1	MSigDB C4	22
MORF_AP2M1	MSigDB C4	22
MORF_GPX4	MSigDB C4	22
GNF2_BUB1	MSigDB C4	21
GNF2_DAP3	MSigDB C4	21
GNF2_HCK	MSigDB C4	21
GNF2_MKI67	MSigDB C4	21

GNF2_S100A4	MSigDB C4	21
MODULE_22	MSigDB C4	21
MORF_CSNK2B	MSigDB C4	21
MORF_EIF3S6	MSigDB C4	21
MORF_ERH	MSigDB C4	21
MORF_MAP2K2	MSigDB C4	21
MORF_PPP1CA	MSigDB C4	21
MORF_UBE2I	MSigDB C4	21
GNF2_CENPF	MSigDB C4	20
GNF2_RRM1	MSigDB C4	20
GNF2_SMC4L1	MSigDB C4	20
GNF2_TNFRSF1B	MSigDB C4	20
GNF2_TPT1	MSigDB C4	20
GNF2_UBE2I	MSigDB C4	20
MODULE_84	MSigDB C4	20
MORF_SOD1	MSigDB C4	20
GNF2_CKS1B	MSigDB C4	19
GNF2_ESPL1	MSigDB C4	19
GNF2_SMC2L1	MSigDB C4	19
MODULE_54	MSigDB C4	19
MODULE_83	MSigDB C4	19
GCM_ACTG1	MSigDB C4	18
GCM_CSNK2B	MSigDB C4	18
GNF2_CARD15	MSigDB C4	18
GNF2_ITGB2	MSigDB C4	18
GNF2_PECAM1	MSigDB C4	18
GNF2_PTPN4	MSigDB C4	18
GNF2_RFC3	MSigDB C4	18
GNF2_RFC4	MSigDB C4	18
MODULE_152	MSigDB C4	18
MODULE_62	MSigDB C4	18
MORF_EIF3S2	MSigDB C4	18
MORF_FBL	MSigDB C4	18
MORF_G22P1	MSigDB C4	18
GNF2_CD1D	MSigDB C4	17
GNF2_H2AFX	MSigDB C4	17
GNF2_IL2RB	MSigDB C4	17
MODULE_91	MSigDB C4	17
MORF_ANP32B	MSigDB C4	17
MORF_BUB1	MSigDB C4	17
MORF_PPP2R4	MSigDB C4	17
MORF_PSMC1	MSigDB C4	17
GCM_APEX1	MSigDB C4	16
GNF2_CD97	MSigDB C4	16
MODULE_47	MSigDB C4	16
MORF_ATOX1	MSigDB C4	16
MORF_DAP3	MSigDB C4	16
MORF_EIF4A2	MSigDB C4	16
GNF2_ICAM3	MSigDB C4	15
GNF2_MCL1	MSigDB C4	15
GNF2_MYD88	MSigDB C4	15
MODULE_114	MSigDB C4	15
MODULE_151	MSigDB C4	15
MODULE_42	MSigDB C4	15
MORF_HDAC1	MSigDB C4	15
GNF2_CD33	MSigDB C4	14
GNF2_MATK	MSigDB C4	14
GNF2_PTPRC	MSigDB C4	14
GNF2_ZAP70	MSigDB C4	14
MODULE_45	MSigDB C4	14
MORF_BUB1B	MSigDB C4	14
GNF2_ITGAL	MSigDB C4	13
GNF2_MCM5	MSigDB C4	13
GNF2_SELL	MSigDB C4	13
MODULE_292	MSigDB C4	13
MODULE_46	MSigDB C4	13
MODULE_57	MSigDB C4	13
MODULE_75	MSigDB C4	13
MODULE_77	MSigDB C4	13
MORF_JUND	MSigDB C4	13
GNF2_FGR	MSigDB C4	12
GNF2_HLA-C	MSigDB C4	12
MORF_DEK	MSigDB C4	12
MORF_SKP1A	MSigDB C4	12
GNF2_CD53	MSigDB C4	11
MODULE_171	MSigDB C4	11
MORF_BUB3	MSigDB C4	11
MORF_CTBP1	MSigDB C4	11
MORF_PPP6C	MSigDB C4	11
MORF_RAD23A	MSigDB C4	11
GCM_PPP1CC	MSigDB C4	10
GCM_PSME1	MSigDB C4	10
GNF2_TNFSF10	MSigDB C4	10
MODULE_158	MSigDB C4	10
MODULE_208	MSigDB C4	10
MODULE_44	MSigDB C4	10
MORF_EI24	MSigDB C4	10
MORF_RAD21	MSigDB C4	10
MORF_SART1	MSigDB C4	10
GNF2_CDH11	MSigDB C4	9
GNF2_INPP5D	MSigDB C4	9
GNF2_NPM1	MSigDB C4	9
GNF2_RAB7L1	MSigDB C4	9
MORF_AP3D1	MSigDB C4	9
MORF_CCNI	MSigDB C4	9
MORF_RAD23B	MSigDB C4	9
GNF2_APEX1	MSigDB C4	8
GNF2_PA2G4	MSigDB C4	8
GNF2_PTPN6	MSigDB C4	8
GNF2_STAT6	MSigDB C4	8
MODULE_103	MSigDB C4	8
MODULE_125	MSigDB C4	8
MODULE_27	MSigDB C4	8
MODULE_28	MSigDB C4	8
MODULE_32	MSigDB C4	8
MODULE_388	MSigDB C4	8
MODULE_403	MSigDB C4	8
MORF_ESPL1	MSigDB C4	8
MORF_PRDX3	MSigDB C4	8
MORF_PRKAR1A	MSigDB C4	8
MORF_PSMC2	MSigDB C4	8
MORF_RRM1	MSigDB C4	8
MORF_XRCC5	MSigDB C4	8
GNF2_FOS	MSigDB C4	7
GNF2_G22P1	MSigDB C4	7

GNF2_MSH2	MSigDB C4	7
GNF2_RAN	MSigDB C4	7
GNF2_VAV1	MSigDB C4	7
MODULE_128	MSigDB C4	7
MODULE_170	MSigDB C4	7
MODULE_252	MSigDB C4	7
MODULE_315	MSigDB C4	7
MODULE_345	MSigDB C4	7
MORF_FEN1	MSigDB C4	7
MORF_RAB5A	MSigDB C4	7
MORF_RFC4	MSigDB C4	7
MORF_UNG	MSigDB C4	7
GNF2_CD14	MSigDB C4	6
MODULE_1	MSigDB C4	6
MODULE_198	MSigDB C4	6
MODULE_307	MSigDB C4	6
MODULE_436	MSigDB C4	6
MODULE_79	MSigDB C4	6
MORF_AATF	MSigDB C4	6
MORF_ACP1	MSigDB C4	6
MORF_GSPT1	MSigDB C4	6
MORF_HAT1	MSigDB C4	6
MORF_HDAC2	MSigDB C4	6
MORF_UBE2N	MSigDB C4	6
GCM_PFN1	MSigDB C4	5
GNF2_NS	MSigDB C4	5
GNF2_SNRK	MSigDB C4	5
GNF2_XRCC5	MSigDB C4	5
MODULE_124	MSigDB C4	5
MODULE_273	MSigDB C4	5
MORF_DNMT1	MSigDB C4	5
MORF_GNB1	MSigDB C4	5
MORF_PCNA	MSigDB C4	5
MORF_RAC1	MSigDB C4	5
MORF_USP5	MSigDB C4	5
GNF2_ANP32B	MSigDB C4	4
GNF2_CASP8	MSigDB C4	4
GNF2_HAT1	MSigDB C4	4
GNF2_MBD4	MSigDB C4	4
GNF2_PTX3	MSigDB C4	4
MODULE_169	MSigDB C4	4
MODULE_223	MSigDB C4	4
MODULE_5	MSigDB C4	4
MODULE_64	MSigDB C4	4
MORF_DDB1	MSigDB C4	4
MORF_GMPS	MSigDB C4	4
MORF_MBD4	MSigDB C4	4
MORF_MTA1	MSigDB C4	4
MORF_PPP1CC	MSigDB C4	4
MORF_PTPN11	MSigDB C4	4
MORF_RAB1A	MSigDB C4	4
GCM_ANP32B	MSigDB C4	3
GCM_DFFA	MSigDB C4	3
GCM_RAB10	MSigDB C4	3
GCM_UBE2N	MSigDB C4	3
GNF2_MLH1	MSigDB C4	3
GNF2_TTN	MSigDB C4	3
MODULE_119	MSigDB C4	3
MODULE_122	MSigDB C4	3
MODULE_329	MSigDB C4	3
MODULE_387	MSigDB C4	3
MODULE_451	MSigDB C4	3
MODULE_512	MSigDB C4	3
MODULE_98	MSigDB C4	3
MORF_BAG5	MSigDB C4	3
MORF_DAP	MSigDB C4	3
MORF_MSH2	MSigDB C4	3
MORF_PPP2CA	MSigDB C4	3
MORF_PRKAG1	MSigDB C4	3
MORF_PRKDC	MSigDB C4	3
MORF_RAD54L	MSigDB C4	3
CAR_HPX	MSigDB C4	2
CAR_IGFBP1	MSigDB C4	2
GCCATNTTG_V$YY1_Q6	MSigDB C4	2
GCM_BAG5	MSigDB C4	2
GCM_CSNK1A1	MSigDB C4	2
GCM_GSPT1	MSigDB C4	2
GCM_HBP1	MSigDB C4	2
GCM_MLL	MSigDB C4	2
GCM_MYST2	MSigDB C4	2
GCM_NF2	MSigDB C4	2
GNF2_CD48	MSigDB C4	2
GNF2_CDH3	MSigDB C4	2
GNF2_CEBPA	MSigDB C4	2
GNF2_CYP2B6	MSigDB C4	2
GNF2_DEK	MSigDB C4	2
GNF2_DENR	MSigDB C4	2
GNF2_DNM1	MSigDB C4	2
GNF2_GSTM1	MSigDB C4	2
GNF2_HPN	MSigDB C4	2
GNF2_HPX	MSigDB C4	2
GNF2_JAK1	MSigDB C4	2
GNF2_LCAT	MSigDB C4	2
GNF2_LYN	MSigDB C4	2
GNF2_MYL2	MSigDB C4	2
GNF2_MYL3	MSigDB C4	2
GNF2_PPP6C	MSigDB C4	2
GNF2_SERPINI2	MSigDB C4	2
GNF2_SPI1	MSigDB C4	2
GNF2_SPINK1	MSigDB C4	2
GNF2_SPRR1B	MSigDB C4	2
GNF2_TST	MSigDB C4	2
MODULE_11	MSigDB C4	2
MODULE_117	MSigDB C4	2
MODULE_12	MSigDB C4	2
MODULE_149	MSigDB C4	2
MODULE_15	MSigDB C4	2
MODULE_17	MSigDB C4	2
MODULE_177	MSigDB C4	2
MODULE_2	MSigDB C4	2
MODULE_201	MSigDB C4	2
MODULE_202	MSigDB C4	2
MODULE_212	MSigDB C4	2
MODULE_23	MSigDB C4	2
MODULE_297	MSigDB C4	2

MODULE_357	MSigDB C4	2
MODULE_397	MSigDB C4	2
MODULE_484	MSigDB C4	2
MORF_BECN1	MSigDB C4	2
MORF_PAPSS1	MSigDB C4	2
MORF_REV3L	MSigDB C4	2
MORF_SP3	MSigDB C4	2
MORF_TPR	MSigDB C4	2
MITOCHONDRIAL_PART	MSigDB C5	17
IMMUNE_SYSTEM_PROCESS	MSigDB C5	14
SPINDLE	MSigDB C5	14
M_PHASE	MSigDB C5	12
MITOCHONDRIAL_MEMBRANE_PART	MSigDB C5	12
DEFENSE_RESPONSE	MSigDB C5	11
IMMUNE_RESPONSE	MSigDB C5	11
T_CELL_ACTIVATION	MSigDB C5	11
ORGANELLAR_RIBOSOME	MSigDB C5	10
M_PHASE_OF_MITOTIC_CELL_CYCLE	MSigDB C5	9
MITOCHONDRIAL_ENVELOPE	MSigDB C5	9
MITOCHONDRIAL_MEMBRANE	MSigDB C5	9
MITOCHONDRION	MSigDB C5	9
RIBONUCLEOPROTEIN_COMPLEX	MSigDB C5	9
RIBOSOME	MSigDB C5	9
LYMPHOCYTE_ACTIVATION	MSigDB C5	8
MITOCHONDRIAL_INNER_MEMBRANE	MSigDB C5	8
MITOCHONDRIAL_RESPIRATORY_CHAIN	MSigDB C5	8
CELL_CYCLE_PROCESS	MSigDB C5	7
MITOCHONDRIAL_RIBOSOME	MSigDB C5	7
MITOSIS	MSigDB C5	7
MITOTIC_CELL_CYCLE	MSigDB C5	7
POSITIVE_REGULATION_OF_IMMUNE_SYSTEM_PROCESS	MSigDB C5	7
STRUCTURAL_CONSTITUENT_OF_RIBOSOME	MSigDB C5	7
CELL_CYCLE_CHECKPOINT_GO_0000075	MSigDB C5	6
CELL_CYCLE_PHASE	MSigDB C5	6
CHROMOSOME	MSigDB C5	6
RNA_BINDING	MSigDB C5	6
SPINDLE_MICROTUBULE	MSigDB C5	6
CHROMOSOMEPERICENTRIC_REGION	MSigDB C5	5
MITOCHONDRIAL_LUMEN	MSigDB C5	5
MITOCHONDRIAL_MATRIX	MSigDB C5	5
MITOCHONDRIAL_RESPIRATORY_CHAIN_COMPLEX_I	MSigDB C5	5
RIBOSOMAL_SUBUNIT	MSigDB C5	5
CELLULAR_DEFENSE_RESPONSE	MSigDB C5	4
EXTRACELLULAR_MATRIX	MSigDB C5	4
LEUKOCYTE_ACTIVATION	MSigDB C5	4
NADH_DEHYDROGENASE_COMPLEX	MSigDB C5	4
PROTEINACEOUS_EXTRACELLULAR_MATRIX	MSigDB C5	4
SPINDLE_CHECKPOINT	MSigDB C5	4
SPINDLE_POLE	MSigDB C5	4
ADAPTIVE_IMMUNE_RESPONSE	MSigDB C5	3
ADAPTIVE_IMMUNE_RESPONSE_GO_0002460	MSigDB C5	3
CELL_CYCLE_GO_0007049	MSigDB C5	3
CHROMOSOMAL_PART	MSigDB C5	3
DNA_REPLICATION	MSigDB C5	3
LOCOMOTORY_BEHAVIOR	MSigDB C5	3
MITOTIC_SISTER_CHROMATID_SEGREGATION	MSigDB C5	3
REGULATION_OF_IMMUNE_SYSTEM_PROCESS	MSigDB C5	3
REGULATION_OF_MITOSIS	MSigDB C5	3
RESPIRATORY_CHAIN_COMPLEX_I	MSigDB C5	3
SPLICEOSOME	MSigDB C5	3
STRUCTURAL_CONSTITUENT_OF_MUSCLE	MSigDB C5	3
CELL_ACTIVATION	MSigDB C5	2
CHROMOSOME_SEGREGATION	MSigDB C5	2
DNA_METABOLIC_PROCESS	MSigDB C5	2
INFLAMMATORY_RESPONSE	MSigDB C5	2
MITOTIC_CELL_CYCLE_CHECKPOINT	MSigDB C5	2
MYOSIN_COMPLEX	MSigDB C5	2
ORGANELLE_INNER_MEMBRANE	MSigDB C5	2
RESPONSE_TO_EXTERNAL_STIMULUS	MSigDB C5	2
SARCOMERE	MSigDB C5	2
SISTER_CHROMATID_SEGREGATION	MSigDB C5	2
SMALL_NUCLEAR_RIBONUCLEOPROTEIN_COMPLEX	MSigDB C5	2
GCNP_SHH_UP_LATE.V1_UP	MSigDB C6	2
GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_MYELOID	MSigDB C7	17
GSE22886_NAIVE_CD8_TCELL_VS_MONOCYTE	MSigDB C7	16
GSE10325_CD4_TCELL_VS_MYELOID	MSigDB C7	15
GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_BCELL	MSigDB C7	15
GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE	MSigDB C7	15
GSE10325_CD4_TCELL_VS_BCELL	MSigDB C7	14
GSE10325_BCELL_VS_MYELOID	MSigDB C7	13
GSE15750_DAY6_VS_DAY10_EFF_CD8_TCELL	MSigDB C7	13
GSE22886_NAIVE_TCELL_VS_MONOCYTE	MSigDB C7	13
GSE15750_DAY6_VS_DAY10_TRAF6KO_EFF_CD8_TCELL	MSigDB C7	12
GSE11057_PBMC_VS_MEM_CD4_TCELL	MSigDB C7	11
GSE40685_TREG_VS_FOXP3_KO_TREG_PRECURSOR	MSigDB C7	11
GSE19888_ADENOSINE_A3R_INH_PRETREAT_AND_ACT_BY_A3R_VS_TCELL_MEMBRANES_ACT_MAST_CELL	MSigDB C7	10
GSE10325_LUPUS_BCELL_VS_LUPUS_MYELOID	MSigDB C7	9
GSE14415_TCONV_VS_FOXP3_KO_INDUCED_TREG	MSigDB C7	9
GSE21063_WT_VS_NFATC1_KO_8H_ANTI_IGM_STIM_BCELL	MSigDB C7	9
GSE24634_TREG_VS_TCONV_POST_DAY10_IL4_CONVERSION	MSigDB C7	9
GSE29618_PDC_VS_MDC_DAY7_FLU_VACCINE	MSigDB C7	9
GSE37533_PPARG1_FOXP3_VS_FOXP3_TRANSDUCED_CD4_TCELL	MSigDB C7	9
GSE22886_NAIVE_BCELL_VS_NEUTROPHIL	MSigDB C7	8
GSE29618_BCELL_VS_MONOCYTE	MSigDB C7	8
GSE3982_MEMORY_CD4_TCELL_VS_BCELL	MSigDB C7	8
GSE13485_CTRL_VS_DAY7_YF17D_VACCINE_PBMC	MSigDB C7	7
GSE13547_CTRL_VS_ANTI_IGM_STIM_BCELL_12H	MSigDB C7	7
GSE14415_INDUCED_VS_NATURAL_TREG	MSigDB C7	7
GSE22140_GERMFREE_VS_SPF_MOUSE_CD4_TCELL	MSigDB C7	7
GSE22140_HEALTHY_VS_ARTHRITIC_GERMFREE_MOUSE_CD4_TCELL	MSigDB C7	7
GSE29618_MONOCYTE_VS_PDC	MSigDB C7	7
GSE3039_NKT_CELL_VS_ALPHAALPHA_CD8_TCELL	MSigDB C7	7
GSE39556_CD8A_DC_VS_NK_CELL_MOUSE_3H_POST_POLYIC_INJ_UP	MSigDB C7	7
GSE42724_NAIVE_BCELL_VS_PLASMABLAST	MSigDB C7	7
GSE14000_UNSTIM_VS_4H_LPS_DC	MSigDB C7	6
GSE21360_NAIVE_VS_QUATERNARY_MEMORY_CD8_TCELL	MSigDB C7	6
GSE21360_PRIMARY_VS_TERTIARY_MEMORY_CD8_TCELL	MSigDB C7	6
GSE22140_GERMFREE_VS_SPF_ARTHRITIC_MOUSE_CD4_TCELL	MSigDB C7	6
GSE29618_BCELL_VS_MDC_DAY7_FLU_VACCINE	MSigDB C7	6
GSE29618_MONOCYTE_VS_MDC	MSigDB C7	6
GSE30962_PRIMARY_VS_SECONDARY_ACUTE_LCMV_INF_CD8_TCELL	MSigDB C7	6
GSE37532_WT_VS_PPARG_KO_VISCERAL_ADIPOSE_TISSUE_TREG	MSigDB C7	6
GSE7218_UNSTIM_VS_ANTIGEN_STIM_THROUGH_IGG_BCELL	MSigDB C7	6
GSE7509_UNSTIM_VS_IFNA_STIM_IMMATURE_DC	MSigDB C7	6
GSE13547_2H_VS_12_H_ANTI_IGM_STIM_BCELL	MSigDB C7	5

GSE14415_NATURAL_TREG_VS_TCONV	MSigDB C7	5
GSE22886_CTRL_VS_LPS_24H_DC	MSigDB C7	5
GSE23568_CTRL_TRANSDUCED_VS_WT_CD8_TCELL	MSigDB C7	5
GSE24634_IL4_VS_CTRL_TREATED_NAIVE_CD4_TCELL_DAY3	MSigDB C7	5
GSE24634_TEFF_VS_TCONV_DAY3_IN_CULTURE	MSigDB C7	5
GSE25088_WT_VS_STAT6_KO_MACROPHAGE	MSigDB C7	5
GSE29618_MONOCYTE_VS_MDC_DAY7_FLU_VACCINE	MSigDB C7	5
GSE29618_PDC_VS_MDC	MSigDB C7	5
GSE37533_PPARG1_FOXP3_VS_PPARG2_FOXP3_TRANSDUCED_CD4_TCELL	MSigDB C7	5
GSE41176_UNSTIM_VS_ANTI_IGM_STIM_BCELL_1H	MSigDB C7	5
GSE13484_UNSTIM_VS_YF17D_VACCINE_STIM_PBMC	MSigDB C7	4
GSE14000_TRANSLATED_RNA_VS_MRNA_DC	MSigDB C7	4
GSE1432_1H_VS_24H_IFNG_MICROGLIA	MSigDB C7	4
GSE18791_UNSTIM_VS_NEWCATSLE_VIRUS_DC_6H	MSigDB C7	4
GSE19198_CTRL_VS_IL21_TREATED_TCELL_6H	MSigDB C7	4
GSE21670_UNTREATED_VS_TGFB_IL6_TREATED_CD4_TCELL	MSigDB C7	4
GSE22196_HEALTHY_VS_OBESE_MOUSE_SKIN_GAMMADELTA_TCELL	MSigDB C7	4
GSE23568_ID3_KO_VS_WT_CD8_TCELL	MSigDB C7	4
GSE24142_EARLY_THYMIC_PROGENITOR_VS2_THYMOCYTE_FETAL_UP	MSigDB C7	4
GSE24634_TEFF_VS_TCONV_DAY10_IN_CULTURE	MSigDB C7	4
GSE24634_TEFF_VS_TCONV_DAY7_IN_CULTURE	MSigDB C7	4
GSE25085_FETAL_BM_VS_ADULT_BM_SP4_THYMIC_IMPLANT	MSigDB C7	4
GSE27241_WT_VS_RORGT_KO_TH17_POLARIZED_CD4_TCELL_TREATED_WITH_DIGOXIN	MSigDB C7	4
GSE29618_BCELL_VS_MDC	MSigDB C7	4
GSE29618_BCELL_VS_MONOCYTE_DAY7_FLU_VACCINE	MSigDB C7	4
GSE29618_MONOCYTE_VS_PDC_DAY7_FLU_VACCINE	MSigDB C7	4
GSE36476_CTRL_VS_TSST_ACT_72H_MEMORY_CD4_TCELL_YOUNG	MSigDB C7	4
GSE39110_DAY3_VS_DAY6_POST_IMMUNIZATION_CD8_TCELL	MSigDB C7	4
GSE3920_UNTREATED_VS_IFNA_TREATED_ENDOTHELIAL_CELL	MSigDB C7	4
GSE41867_NAIVE_VS_DAY30_LCMV_ARMSTRONG_MEMORY_CD8_TCELL	MSigDB C7	4
GSE42021_CD24HI_VS_CD24INT_TREG_THYMUS	MSigDB C7	4
GSE42021_TREG_PLN_VS_CD24INT_TREG_THYMUS	MSigDB C7	4
GSE7509_UNSTIM_VS_FCGRIIB_STIM_DC	MSigDB C7	4
GSE9988_LOW_LPS_VS_VEHICLE_TREATED_MONOCYTE	MSigDB C7	4
GSE9988_LPS_VS_VEHICLE_TREATED_MONOCYTE	MSigDB C7	4
GSE11057_CD4_EFF_MEM_VS_PBMC	MSigDB C7	3
GSE13485_CTRL_VS_DAY3_YF17D_VACCINE_PBMC	MSigDB C7	3
GSE1432_CTRL_VS_IFNG_24H_MICROGLIA	MSigDB C7	3
GSE1432_CTRL_VS_IFNG_6H_MICROGLIA	MSigDB C7	3
GSE1460_DP_VS_CD4_THYMOCYTE	MSigDB C7	3
GSE19198_1H_VS_24H_IL21_TREATED_TCELL	MSigDB C7	3
GSE19401_UNSTIM_VS_RETINOIC_ACID_AND_PAM2CSK4_STIM_FOLLICULAR_DC	MSigDB C7	3
GSE21063_WT_VS_NFATC1_KO_3H_ANTI_IGM_STIM_BCELL	MSigDB C7	3
GSE21380_NON_TFH_VS_TFH_CD4_TCELL	MSigDB C7	3
GSE22886_NAIVE_TCELL_VS_DC	MSigDB C7	3
GSE23568_ID3_TRANSDUCED_VS_ID3_KO_CD8_TCELL	MSigDB C7	3
GSE2405_0H_VS_9H_A_PHAGOCYTOPHILUM_STIM_NEUTROPHIL	MSigDB C7	3
GSE24634_IL4_VS_CTRL_TREATED_NAIVE_CD4_TCELL_DAY5	MSigDB C7	3
GSE24634_TREG_VS_TCONV_POST_DAY3_IL4_CONVERSION	MSigDB C7	3
GSE25085_FETAL_LIVER_VS_ADULT_BM_SP4_THYMIC_IMPLANT	MSigDB C7	3
GSE25123_CTRL_VS_ROSIGLITAZONE_STIM_PPARG_KO_MACROPHAGE	MSigDB C7	3
GSE3039_CD4_TCELL_VS_B1_BCELL	MSigDB C7	3
GSE30971_CTRL_VS_LPS_STIM_MACROPHAGE_WBP7_KO_2H	MSigDB C7	3
GSE31082_VS_DP_THYMOCYTE	MSigDB C7	3
GSE34156_NOD2_LIGAND_VS_TLR1_TLR2_LIGAND_6H_TREATED_MONOCYTE	MSigDB C7	3
GSE34156_UNTREATED_VS_6H_TLR1_TLR2_LIGAND_TREATED_MONOCYTE	MSigDB C7	3
GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_TCELL_OLD	MSigDB C7	3
GSE36476_CTRL_VS_TSST_ACT_72H_MEMORY_CD4_TCELL_OLD	MSigDB C7	3
GSE37533_PPARG2_FOXP3_VS_FOXP3_TRANSDUCED_CD4_TCELL	MSigDB C7	3
GSE39556_UNTREATED_VS_3H_POLYIC_INJ_MOUSE_CD8A_DC	MSigDB C7	3
GSE42021_TREG_PLN_VS_TREG_PRECURSORS_THYMUS	MSigDB C7	3
GSE43863_DAY6_EFF_VS_DAY150_MEM_LY6C_INT_CXCR5POS_CD4_TCELL	MSigDB C7	3
GSE44649_NAIVE_VS_ACTIVATED_CD8_TCELL	MSigDB C7	3
GSE45365_NK_CELL_VS_CD11B_DC	MSigDB C7	3
GSE45365_WT_VS_IFNAR_KO_BCELL	MSigDB C7	3
GSE45739_UNSTIM_VS_ACD3_ACD28_STIM_NRAS_KO_CD4_TCELL	MSigDB C7	3
GSE45837_WT_VS_GFI1_KO_PDC	MSigDB C7	3
GSE4984_UNTREATED_VS_GALECTIN1_TREATED_DC	MSigDB C7	3
GSE7219_WT_VS_NIK_NFKB2_KO_DC	MSigDB C7	3
GSE9006_TYPE_1_VS_TYPE_2_DIABETES_PBMC_AT_DX	MSigDB C7	3
GSE9988_ANTI_TREM1_VS_ANTI_TREM1_AND_LPS_MONOCYTE	MSigDB C7	3
GSE9988_ANTI_TREM1_VS_LOW_LPS_MONOCYTE	MSigDB C7	3
GOLDRATH_EFF_VS_MEMORY_CD8_TCELL	MSigDB C7	2
GOLDRATH_NAIVE_VS_MEMORY_CD8_TCELL	MSigDB C7	2
GSE10325_CD4_TCELL_VS_LUPUS_CD4_TCELL	MSigDB C7	2
GSE11057_CD4_CENT_MEM_VS_PBMC	MSigDB C7	2
GSE11386_NAIVE_VS_MEMORY_BCELL	MSigDB C7	2
GSE13485_DAY1_VS_DAY7_YF17D_VACCINE_PBMC	MSigDB C7	2
GSE13485_PRE_VS_POST_YF17D_VACCINATION_PBMC	MSigDB C7	2
GSE13547_2H_VS_12_H_ANTI_IGM_STIM_ZFX_KO_BCELL	MSigDB C7	2
GSE16266_LPS_VS_HEATSHOCK_AND_LPS_STIM_MEF	MSigDB C7	2
GSE17301_ACD3_ACD28_VS_ACD3_ACD28_AND_IFNA5_STIM_CD8_TCELL	MSigDB C7	2
GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_72H_ACT_CD4_TCELL	MSigDB C7	2
GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_10H	MSigDB C7	2
GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_6H	MSigDB C7	2
GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_8H	MSigDB C7	2
GSE18791_UNSTIM_VS_NEWCATSLE_VIRUS_DC_10H	MSigDB C7	2
GSE18893_TCONV_VS_TREG_24H_TNF_STIM	MSigDB C7	2
GSE19401_NAIVE_VS_IMMUNIZED_MOUSE_PLN_FOLLICULAR_DC	MSigDB C7	2
GSE19888_ADENOSINE_A3R_INH_VS_ACT_WITH_INHIBITOR_PRETREATMENT_IN_MAST_CELL	MSigDB C7	2
GSE2128_CTRL_VS_MIMETOPE_NEGATIVE_SELECTION_DP_THYMOCYTE_NOD	MSigDB C7	2
GSE21360_SECONDARY_VS_QUATERNARY_MEMORY_CD8_TCELL	MSigDB C7	2
GSE21546_ELK1_KO_VS_SAP1A_KO_AND_ELK1_KO_DP_THYMOCYTES	MSigDB C7	2
GSE21546_WT_VS_SAP1A_KO_DP_THYMOCYTES	MSigDB C7	2
GSE21670_STAT3_KO_VS_WT_CD4_TCELL_IL6_TREATED_UP	MSigDB C7	2
GSE21670_UNTREATED_VS_IL6_TREATED_STAT3_KO_CD4_TCELL	MSigDB C7	2
GSE22140_HEALTHY_VS_ARTHRITIC_MOUSE_CD4_TCELL	MSigDB C7	2
GSE22601_IMMATURE_CD4_SINGLE_POSITIVE_VS_CD8_SINGLE_POSITIVE_THYMOCYTE	MSigDB C7	2
GSE22886_CD4_TCELL_VS_BCELL_NAIVE	MSigDB C7	2
GSE22886_CD8_TCELL_VS_BCELL_NAIVE	MSigDB C7	2
GSE22886_DAY0_VS_DAY1_MONOCYTE_IN_CULTURE	MSigDB C7	2
GSE22886_DC_VS_MONOCYTE	MSigDB C7	2
GSE22886_NAIVE_BCELL_VS_MONOCYTE	MSigDB C7	2
GSE22886_TCELL_VS_BCELL_NAIVE	MSigDB C7	2
GSE22886_UNSTIM_VS_IL15_STIM_NKCELL	MSigDB C7	2
GSE22886_UNSTIM_VS_IL2_STIM_NKCELL	MSigDB C7	2
GSE22935_WT_VS_MYD88_KO_MACROPHAGE	MSigDB C7	2
GSE23568_CTRL_VS_ID3_TRANSDUCED_CD8_TCELL	MSigDB C7	2
GSE2405_S_AUREUS_VS_UNTREATED_NEUTROPHIL	MSigDB C7	2
GSE24634_TREG_VS_TCONV_POST_DAY5_IL4_CONVERSION	MSigDB C7	2
GSE26030_TH1_VS_TH17_DAY5_POST_POLARIZATION	MSigDB C7	2
GSE26343_UNSTIM_VS_LPS_STIM_NFAT5_KO_MACROPHAGE	MSigDB C7	2
GSE26495_NAIVE_VS_PD1LOW_CD8_TCELL	MSigDB C7	2
GSE29614_CTRL_VS_DAY7_TIV_FLU_VACCINE_PBMC	MSigDB C7	2
GSE29618_BCELL_VS_PDC	MSigDB C7	2

GSE29618_BCELL_VS_PDC_DAY7_FLU_VACCINE	MSigDB C7	2
GSE30971_CTRL_VS_LPS_STIM_MACROPHAGE_WBP7_KO_4H	MSigDB C7	2
GSE32164_RESTING_DIFFERENTIATED_VS_ALTERNATIVELY_ACT_M2_MACROPHAGE	MSigDB C7	2
GSE32533_MIR17_KO_VS_MIR17_OVEREXPRESS_ACT_CD4_TCELL	MSigDB C7	2
GSE3337_4H_VS_16H_IFNG_IN_CD8POS_DC	MSigDB C7	2
GSE34006_A2AR_KO_VS_A2AR_AGONIST_TREATED_TREG	MSigDB C7	2
GSE34156_TLR1_TLR2_LIGAND_VS_NOD2_AND_TLR1_TLR2_LIGAND_24H_TREATED_MONOCYTE	MSigDB C7	2
GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_TCELL_YOUNG	MSigDB C7	2
GSE36527_CD69_NEG_VS_POS_TREG_CD62L_LOS_KLRG1_NEG	MSigDB C7	2
GSE36826_NORMAL_VS_STAPH_AUREUS_INF_IL1R_KO_SKIN	MSigDB C7	2
GSE36888_UNTREATED_VS_IL2_TREATED_TCELL_6H	MSigDB C7	2
GSE40274_CTRL_VS_FOXP3_TRANSDUCED_ACTIVATED_CD4_TCELL	MSigDB C7	2
GSE41087_WT_VS_FOXP3_MUT_ANTI_CD3_CD28_STIM_CD4_TCELL	MSigDB C7	2
GSE41176_WT_VS_TAK1_KO_ANTI_IGM_STIM_BCELL_24H	MSigDB C7	2
GSE41867_NAIVE_VS_DAY15_LCMV_EFFECTOR_CD8_TCELL	MSigDB C7	2
GSE42021_TREG_VS_TCONV_PLN	MSigDB C7	2
GSE43863_DAY6_EFF_VS_DAY150_MEM_TH1_CD4_TCELL	MSigDB C7	2
GSE45365_HEALTHY_VS_MCMV_INFECTION_CD11B_DC	MSigDB C7	2
GSE5099_CLASSICAL_M1_VS_ALTERNATIVE_M2_MACROPHAGE	MSigDB C7	2
GSE6259_33D1_POS_DC_VS_CD4_TCELL	MSigDB C7	2
GSE7509_DC_VS_MONOCYTE	MSigDB C7	2
GSE7768_OVA_ALONE_VS_OVA_WITH_LPS_IMMUNIZED_MOUSE_WHOLE_SPLEEN_6H	MSigDB C7	2
GSE8835_CD4_VS_CD8_TCELL_CLL_PATIENT	MSigDB C7	2
GSE9988_ANTI_TREM1_VS_CTRL_TREATED_MONOCYTES	MSigDB C7	2
GSE9988_ANTI_TREM1_VS_LPS_MONOCYTE	MSigDB C7	2
GSE9988_ANTI_TREM1_VS_VEHICLE_TREATED_MONOCYTES	MSigDB C7	2
GSE9988_LOW_LPS_VS_CTRL_TREATED_MONOCYTE	MSigDB C7	2
GSE9988_LPS_VS_CTRL_TREATED_MONOCYTE	MSigDB C7	2
HALLMARK_ALLOGRAFT_REJECTION	MSigDB H	24
HALLMARK_OXIDATIVE_PHOSPHORYLATION	MSigDB H	20
HALLMARK_E2F_TARGETS	MSigDB H	17
HALLMARK_G2M_CHECKPOINT	MSigDB H	17
HALLMARK_INTERFERON_GAMMA_RESPONSE	MSigDB H	17
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	MSigDB H	16
HALLMARK_INTERFERON_ALPHA_RESPONSE	MSigDB H	16
HALLMARK_MYC_TARGETS_V1	MSigDB H	13
HALLMARK_INFLAMMATORY_RESPONSE	MSigDB H	10
HALLMARK_TNFA_SIGNALING_VIA_NFKB	MSigDB H	7
HALLMARK_MYC_TARGETS_V2	MSigDB H	5
HALLMARK_IL6_JAK_STAT3_SIGNALING	MSigDB H	4
HALLMARK_COMPLEMENT	MSigDB H	3
HALLMARK_MYOGENESIS	MSigDB H	3
HALLMARK_MTORC1_SIGNALING	MSigDB H	2
IL-1	Speed	6
JAK-STAT	Speed	5
MAPK_PI3K	Speed	5
TLR	Speed	5
TNFa	Speed	5
TGFB	Speed	2

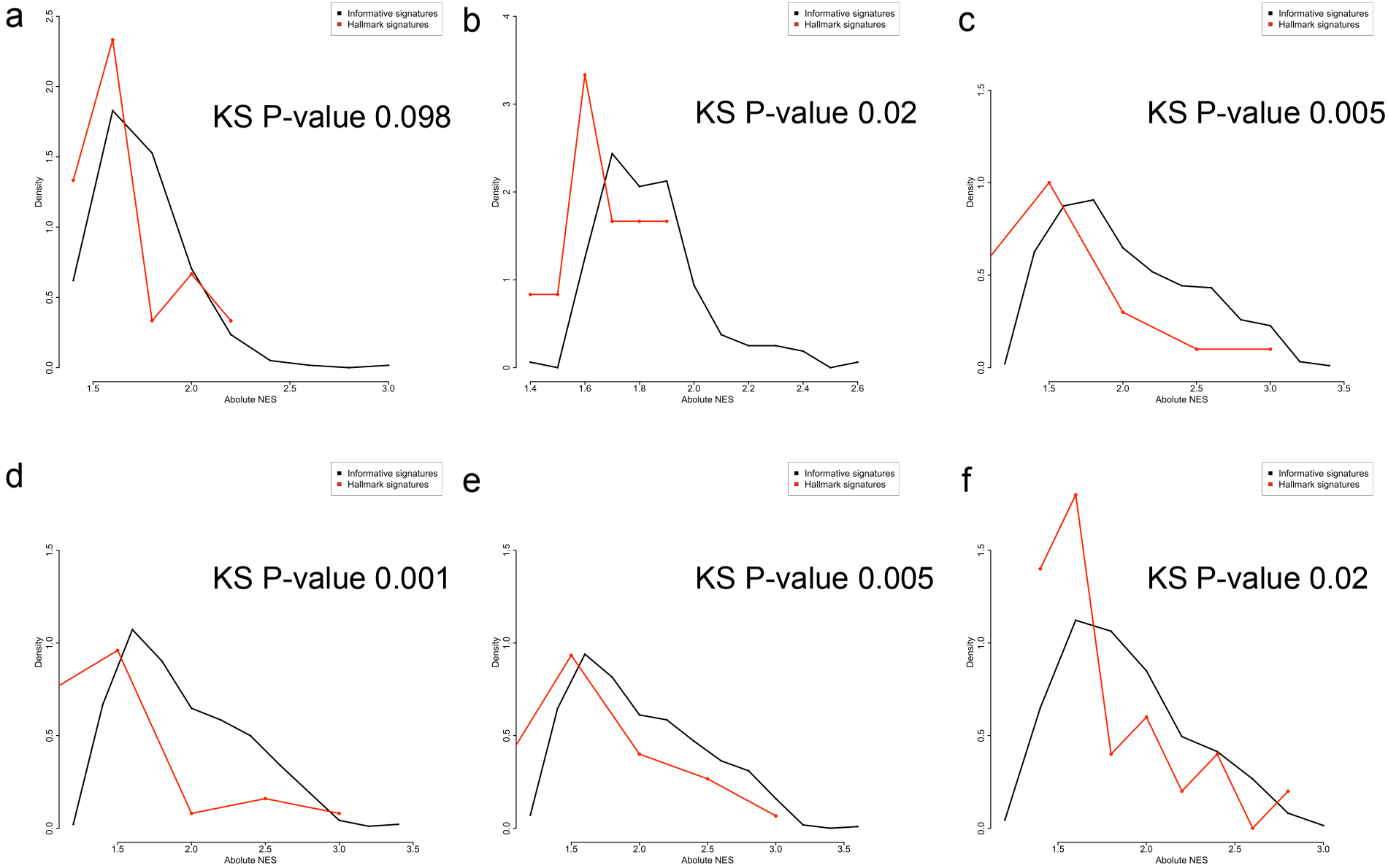

Figure S1. Comparison GSEA absolute NES values for the informative GSEA significant (black) vs. Hallmark GSEA significant (red) in (a) KRAS mutated vs. wild type colorectal cancer; (b) metastatic vs. primary colon cancer and (c-f) tumor vs. normal in cervix, colon, gastric and lung, respectively.

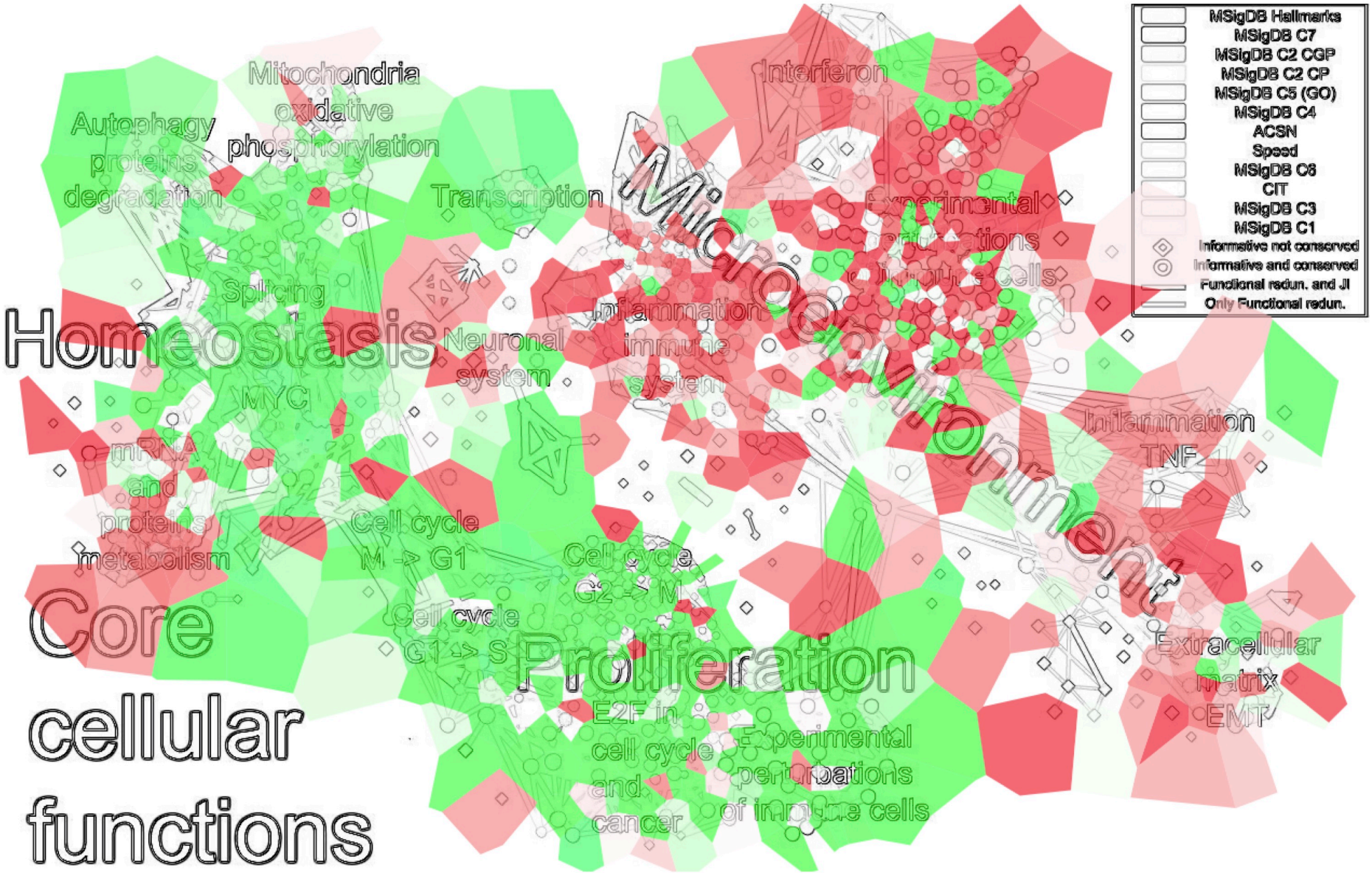

Figure S2. Fold changes of ROMA activity on the expression profiling of human CD4+ T cell during differentiation induction plotted on InfoSigMap. Red, and green indicate upregulation and downregulation, respectively.