Appendix

Datasets and choice of the GO layers

a)

Tissue type	abdomen	adrenal	blood	bone	brain	breast	colon	kidney	liver
#samples	142	83	4283	3525	869	2171	1239	657	730

Tissue type	lung	lymph node	ovary	pancreas	prostate	skin	stomach	uterus	Total
#samples	1415	567	573	243	415	835	154	572	18473

b)

Class	cancer	noncancer	Total
#train	11799	6048	17847
#validation	2950	1512	4462
#test	3688	1890	5578
Total	18437	9450	27887
Class frequency (%)	66.11	33.89	100

Table S1: Description of the microarray dataset. (a) List of the tissue types observed in the entire dataset. (b) Size of each {train,validation,test} set. # indicates the number of samples.

Class	BRCA	HNSC	KIRC	LGG	LIHC	LUAD	LUSC	OV	PRAD	THCA	UCEC	NT	Total
#train	705	320	344	327	238	341	321	239	318	321	353	309	4136
#validation	176	80	86	82	59	85	81	60	80	81	88	77	1035
#test	221	100	108	102	74	107	100	75	100	100	110	96	1293
Total	1102	500	538	511	371	533	502	374	498	502	551	482	6464
Class frequency (%)	17.05	7.74	8.32	7.91	5.74	8.25	7.77	5.79	7.71	7.77	8.53	7.46	100

Table S2: Description of the TCGA dataset where # indicates the number of samples in each set {train,validation,test}. Meaning of the abbreviations: BRCA (Breast invasive carcinoma), HNSC (Head and Neck squamous cell carcinoma), KIRC (Kidney renal clear cell carcinoma), LGG (Brain Lower Grade Glioma), LIHC (Liver hepatocellular carcinoma), LUAD (Lung adenocarcinoma), LUSC (Lung squamous cell carcinoma), OV (Ovarian serous cystadenocarcinoma), PRAD (Prostate adenocarcinoma), THCA (Thyroid carcinoma), UCEC (Uterine Corpus Endometrial Carcinoma), NT (Normal samples). The data are pre-normalized with FPKM (fragments per kilobase per million mapped reads) and transformed using $\log_2$.

The Gene Ontology version used dates from 01-06-2020 and contains originally 29,112 GO-BP terms. The DAG is organized into levels, where the level of a GO term is determined according to its longest path with the root. For each dataset, a GO graph is constituted of the GO terms annotated with the set of input genes. 67.37% of the genes in microarray (resp. 35.51% in TCGA) are linked with at least one GO term from GO-BP. The genes that are not associated with any GO term are removed. Some GO terms do not have any connections with the genes. If a gene is annotated with a GO term and the ancestors of this GO term, we do not consider the annotations with the ancestors. On the basis of the propagation rule, the information will be spread to the ancestors. Only the link between the gene and the smallest descendants is kept. As the entire GO graphs could not fit in memory, we cut the first level of leaves in each graph. Based on the transitivity principle, a parent GO term inherits the set of genes from its children, so we connect the parents of the deleted leaves with the genes with whom they have annotations. It results in a 19-levels GO graph with $K = 10,663$ for microarray, and $K = 10,636$ for TCGA.

Sensitivity Analysis

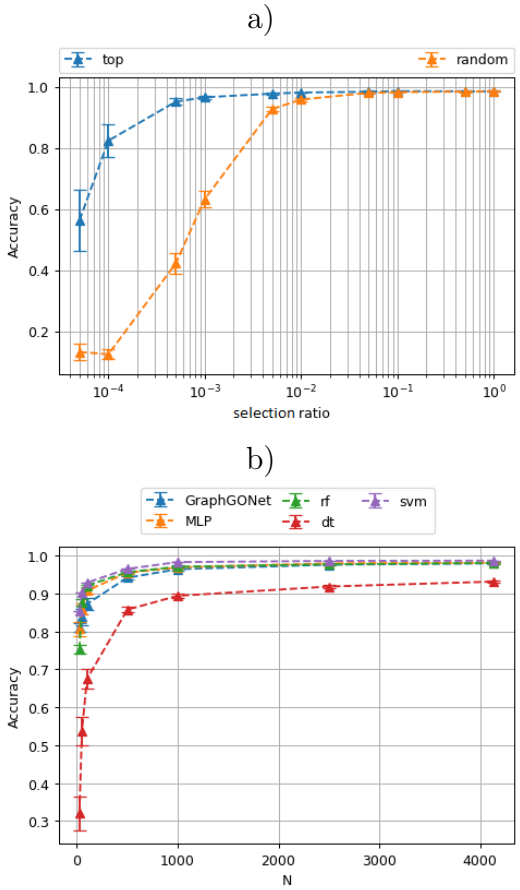


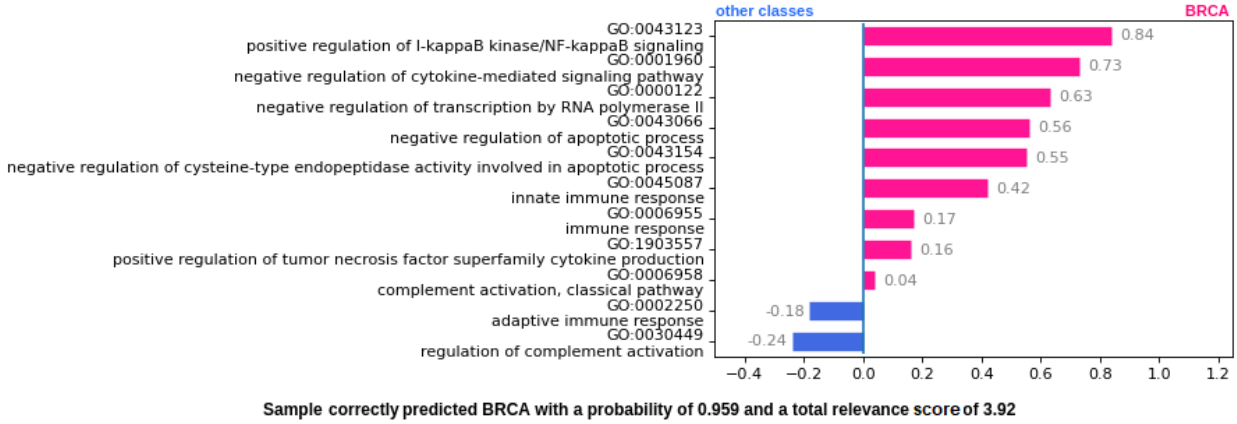
Figure S1: Accuracy of the models according to (a) the number of samples N and (b) the selection ratio r from the TCGA dataset.

Biological Analysis

Interpretation of a patient outcome

Figures S2a/b show a comparison of the explanations provided by GraphGONet on two patients from the TCGA dataset with different diseases (respectively BRCA and LGG). In both cases, the patients are correctly predicted. On these figures, we can see again that some GO terms can be important for different output classes, but with different quantitative contributions. For example, for the GO term GO:0000122, the relevance score is 0.63 for the BRCA patient, contrary to 0.95 for the LGG patient. In contrast, some GO terms can be specific to some tissues. For example, the GO term GO:0007155, which appears in the explanation of the LGG patient, is never selected for any BRCA predicted patients. Besides, the GO term GO:0002250 is not as significant as the other GO terms in the explanation of the LGG patient, but its impact is positive on the final prediction contrary to the explanation of the BRCA patient. Note that the probability of prediction of the BRCA patient is lower than the one of the LGG patient. It can be explained by the existence of negative relevance scores and the fact that the quantitative contributions of the GO terms are lower in the relevance profile of the BRCA patient than in the one of the LGG patient.

a)



b)

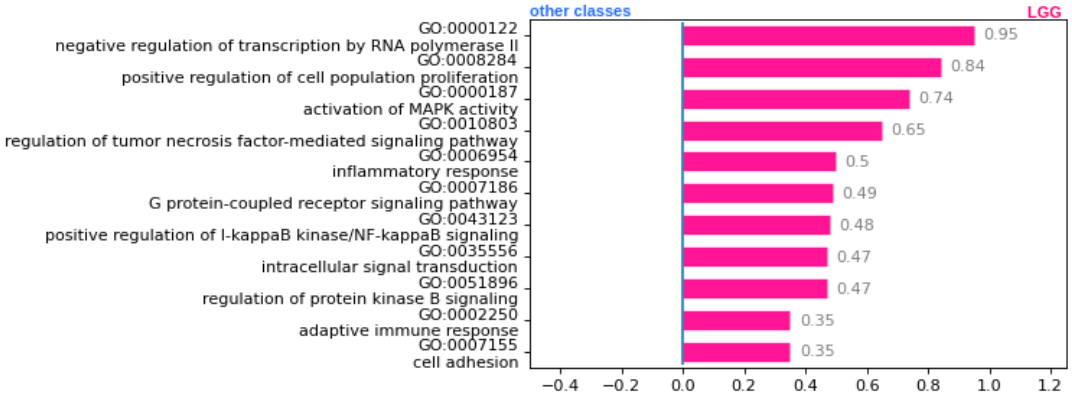


Figure S2: Explanation of (a) a BRCA and (b) a LGG prediction. A subset of eleven GO terms is reported with their relevance score and their description. The color indicates towards which class a GO term influences the signal: blue for other classes and pink for the target class (LGG or BRCA). The total relevance score is the sum of the relevance scores and the bias of the output class.

Interpretation of the model

Tissue type	abdomen	adrenal	blood	bone	brain	breast	colon	kidney	liver
#samples	10	15	454	623	168	366	218	98	120

Tissue type	lung	lymph node	ovary	pancreas	prostate	skin	stomach	uterus	Total
#samples	169	112	95	36	77	79	27	116	2783

Table S3: List of the tissue types observed in the cancer samples from the microarray test set where # indicates the number of samples.

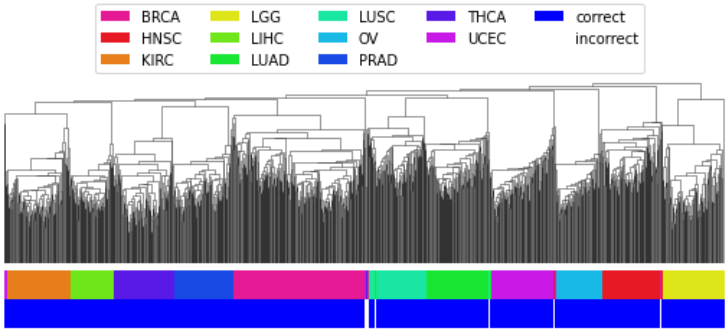


Figure S3: Dendrogram of the relevance matrix on the test cancer samples from the TCGA dataset. The first row displays the type of cancer of each sample, whereas the second row indicates the correctness of the prediction (white: wrong, blue: right).

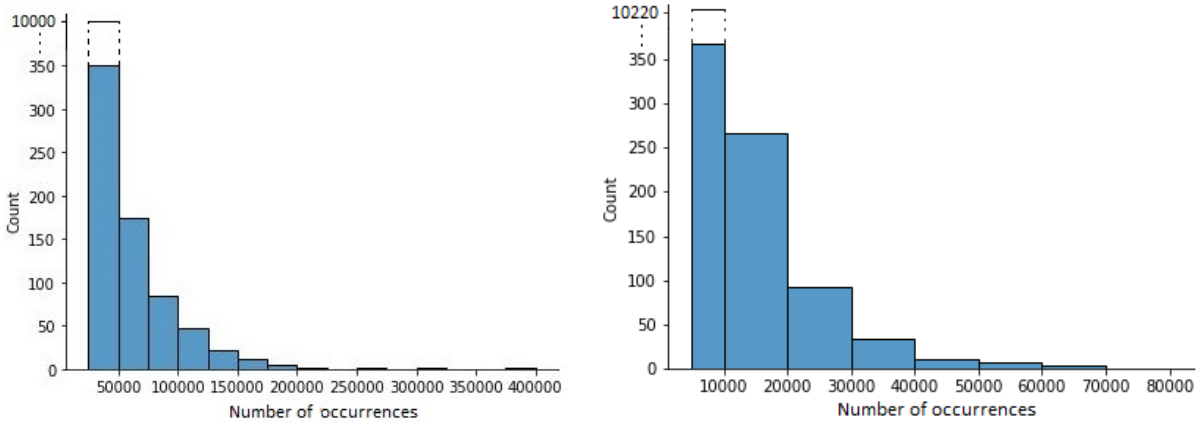


Figure S4: Histogram of number of occurrences based on the occurrence matrix from the (a) microarray and (b) TCGA dataset. The x-axis corresponds to the number of times a GO-BP term is selected across the models and samples, and the y-axis the number of GO-BP terms taking value in the intervals.

Figures S5a and S5b report the top-10 most frequent GO terms according to their occurrence on the cancer samples versus the noncancer samples. We can observe that the most frequent GO terms are similar between the two profiles, but the proportion of positive-negative signs is different. For instance, for the GO term GO:0045944, the most frequent GO term for both outputs, the proportion of positive signs is two-thirds more than the negative ones for the cancer output while, for the noncancer output, it is the opposite. Even if it is not exactly the same ranking, the figures are complementary. It means that a GO term can be important for both types of prediction (same weight), but the sign of the signal (activation) will determine the outcome.

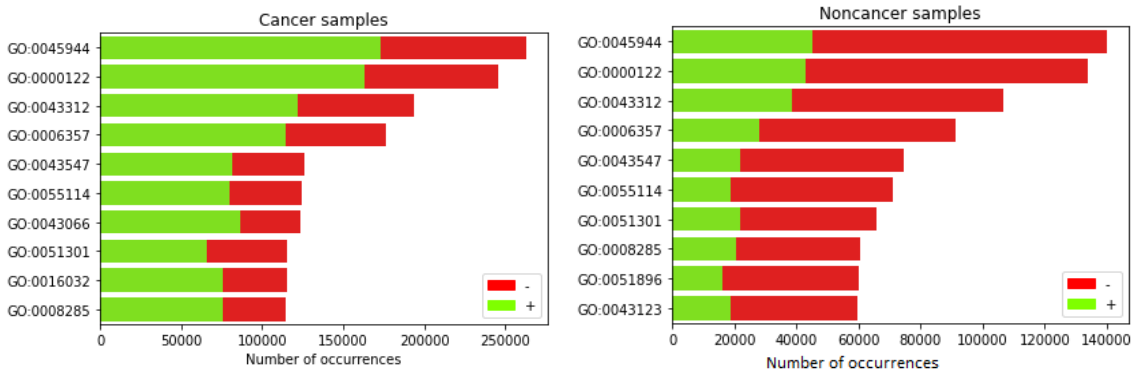


Figure S5: Top-10 most frequent GO terms sorted according to their occurrence for (a) cancer and (b) noncancer output from microarray. The colors indicate the part of occurrences having a negative (red) or positive (green) relevance score. The maximal frequency that can be reached corresponds to the number of cancer (resp. noncancer) samples times the number of models, i.e., 368,800 (resp. 189,000).