Empirical Analysis of ONMF-MP Results Based on GO-term Propagation

After stratification, we analyze resulting subtypes based on the GO signatures of each subtype to identify significant GO terms. The GO-MP that we used is based on the most specific terms. However, for interpretation of subtype signatures, looking at more general terms may be more convenient. For this purpose, we propagate the leaf node weight to the ancestors until significant GO terms are found. Details of the process are shown in Alg. 1.

Iteration proceeds until the scores of updated nodes converge (Line 1). For each parent node of the updated i-th node, the propagation is repeated until the stopping criterion is met (Line 3). If the number of parent nodes is 1, it indicates that the targeted parent node will have same score as that of the child node with more general meaning (Line 6). In this case, the propagation stops. $x[i].r$ is the ratio of child nodes having non-zero scores over all of its child nodes, and ϵ is the minimum fraction of child nodes that have nonzero scores to be a valid node. If

$\epsilon < x[p].r$, it indicates that the targeted parent covers a sufficient number of child nodes in terms of the

semantics of the ontology (Line 10). In this case, the propagation continues to accumulate the scores to the ancestor node that represents the child nodes. Finally, based on the accumulated scores of GO terms, the nodes that have larger score than the elbow point are selected as significant nodes called "valid nodes" (Line 18) and the GO terms of the nodes are collected and used. Here significant GO terms for UCEC and GBM data and their clinical interpretations are provided and more examples on other data are provided in Table S1.

Algorithm 1: Identifying significant GO terms.

Data: Initial score vector $\mathbf{w_0}$, GO terms $\mathbf{x}$
Result: A set of significant GO terms, $\mathbf{x}^*$

```
1  w* = w0
2  repeat
3    foreach node i that is updated at the previous step do
4      P = x[i].Parents();
           % An index set of i-th GO term' ancestors
5      scr = w*[i]=|P|
6      if |P| == 1 then
7        continue;
8      end
9      foreach p ∈ P do
10       if ε < x[p].r then
11         w*[p]+ = scr;
12       else
13         w*[p] = scr;
14       end
15     end
16   end
17 until w* does not change;
18 return x* = GetSignificantGOterms(w*)
```

Example on UCEC Data.

In this section, we provide the analysis results of GO terms based on the cancer subtypes. Even with the initial filtering, the cancer subtypes identified by ONMF-MP (basis vectors H) still included over 3,000 GO terms (Fig. S5A), so the significant GO terms in each cancer subtype are difficult to recognize and looking at general terms that are significant can be useful especially if many GO leaf node terms have similar scores. Based on Alg. 1, significant GO terms can be extracted based on the scores accumulated from the leaf nodes (Fig. S5C). In the basis vectors of one UCEC subtype (Fig. S5B) the flow of propagation to "regulation of signal transduction (GO:0009966)" gains most of the scores from its descendants "apoptotic signaling pathway (GO:0097190)" and "positive regulation of signal transduction (GO:0009967)".

Example on GBM Data.

We show stratification results of GBM by representing subtypes by their basis vectors, which indicate the contribution of the GO terms to cancer. In this experiment, data were grouped to five subtypes according to the best χ^2 statistics (Fig.S2C). Fig. S6A shows the distributions of GO term scores from initial stratification results, and Fig. S6B shows the distributions of GO term scores after executing Alg.1. Compared to the initial number of GO terms in each subtype, the number of significant GO terms is much smaller and easier to interpret for empirical examination. Among the five cancer subtypes extracted from GBM data (Fig. S6A), in the first subtype (the first column in the heatmap) had "positive regulation of phosphatidylinositol 3-kinase (PI3K) signaling" and "homophilic cell adhesion" as the top-2 GO terms. PI3K cascade is an important pathway that is involved in proliferation, invasion, and migration in cancer (Samuels et al. (2005); Zeng et al. (2006); Jimenez et al. (2000)). Also genes related to PI3K pathway influence GBM patients' survival (Ruano et al. (2008)). After term propagation (Fig. S6B), the top-2 GO terms for the first subtype were "phosphatidylinositol 3-kinase" and "regulation of signal transduction", which are more general terms than the top-2 GO terms in the initial term list. In the fourth subtype, the top-1 GO term was "positive regulation of pancreatic stellate cell proliferation" in both initial and filtered subtypes. A protein interaction network analysis shows that glioblastoma cancer and pancreatic cancer share network patterns that contain most of the candidate causative mutations (Wu et al. (2010)). In addition, pancreatic stellate cells (PSCs) are responsible for creating a tumor facilitatory environment that stimulates local tumor growth and distant metastasis (Yoshida et al. (2004)).

SUPPLEMENTARY TABLES

Table S1 (a): Siginificant GO terms for BRCA data

Subtype	GO ID	Description	Score
	GO:0009966	regulation of signal transduction	41.95
	GO:0006366	transcription from RNA polymerase II promoter	35.05
	GO:0007156	homophilic cell adhesion	28.41
	GO:0014065	phosphatidylinositol 3-kinase signaling	19.95
	GO:0035556	intracellular signal transduction	16.11
	GO:0006468	protein phosphorylation	14.48
	GO:0007165	signal transduction	14.07
	GO:0045765	regulation of angiogenesis	13.47
	GO:0097190	apoptotic signaling pathway	13.39
1	GO:0009154	purine ribonucleotide catabolic process	13.31
	GO:0048011	neurotrophin TRK receptor signaling pathway	13.23
	GO:0007254	JNK cascade	13.06
	GO:0001817	regulation of cytokine production	12.97
	GO:0006200	ATP catabolic process	12.85
	GO:0050865	regulation of cell activation	11.92
	GO:0042127	regulation of cell proliferation	11.61
	GO:0061041	regulation of wound healing	11.33
	GO:0001503	ossification	11.01
	GO:0043627	response to estrogen	11.00
	GO:0071158	positive regulation of cell cycle arrest	30.16
	GO:0090399	replicative senescence	18.91
	GO:0097190	apoptotic signaling pathway	17.97
	GO:1901525	negative regulation of macromitophagy	14.18
	GO:0002931	response to ischemia	13.46
	GO:0090400	stress-induced premature senescence	13.36
	GO:0010666	positive regulation of cardiac muscle cell apoptotic process	11.86
	GO:0007005	mitochondrion organization	9.64
	GO:0035583	sequestering of TGFbeta in extracellular matrix	9.55
	GO:2001022	positive regulation of response to DNA damage stimulus	9.48
2	GO:0035556	intracellular signal transduction	9.34
	GO:0042772	DNA damage response, signal transduction resulting in transcription	8.88
	GO:0035264	multicellular organism growth	8.52
	GO:0006284	base-excision repair	7.43
	GO:0070231	T cell apoptotic process	7.40
	GO:2000772	regulation of cellular senescence	7.39
	GO:0010332	response to gamma radiation	7.22
	GO:0007165	signal transduction	6.96
	GO:0090343	positive regulation of cell aging	6.49
	GO:0042554	superoxide anion generation	6.49
3	GO:0038028	insulin receptor signaling pathway via phosphatidylinositol 3-kinase	26.24
	GO:0036092	phosphatidylinositol-3-phosphate biosynthetic process	25.04
	GO:0043491	protein kinase B signaling	18.25
	GO:2000653	regulation of genetic imprinting	15.48
	GO:0014065	phosphatidylinositol 3-kinase signaling	15.04
	GO:0044029	hypomethylation of CpG island	13.69
	GO:0038094	Fc-gamma receptor signaling pathway	10.86
	GO:0002431	Fc receptor mediated stimulatory signaling pathway	8.36
	GO:0035264	multicellular organism growth	8.11
	GO:0097049	motor neuron apoptotic process	8.01
	GO:0030258	lipid modification	7.44
	GO:0050862	positive regulation of T cell receptor signaling pathway	7.15

	GO:0009966	regulation of signal transduction	7.15
	GO:2000667	positive regulation of interleukin-13 secretion	19.34
	GO:0001817	regulation of cytokine production	14.75
	GO:0043049	otic placode formation	13.66
	GO:0072107	positive regulation of ureteric bud formation	12.86
	GO:2000703	negative regulation of fibroblast growth factor receptor signaling pathway	10.19
	GO:2000611	involved in ureteric bud	9.00
	GO:2000734	positive regulation of thyroid hormone generation	8.96
	GO:0061549	negative regulation of glial cell-derived neurotrophic factor receptor signaling	8.72
	GO:2000664	pathway	8.50
	GO:0060037	sympathetic ganglion development	8.36
	GO:0071773	positive regulation of interleukin-5 secretion	7.94
	GO:0072178	pharyngeal system development	7.92
	GO:0035799	cellular response to BMP stimulus	7.33
	GO:0032736	nephric duct morphogenesis	7.33
	GO:0043491	ureter maturation	7.06
	GO:0035455	positive regulation of interleukin-13 production	6.99
	GO:0032633	protein kinase B signaling	6.63
	GO:0043627	response to interferon-alpha	6.41
	GO:0001816	interleukin-4 production	6.31
	GO:0051099	response to estrogen	6.17
	GO:0032609	cytokine production	6.00
4	GO:0060065	positive regulation of binding	5.16
	GO:0031929	interferon-gamma production	4.69
	GO:0042415	uterus development	4.40
	GO:0045444	TOR signaling	4.08
	GO:0032674	norepinephrine metabolic process	3.75
	GO:0051568	fat cell differentiation	3.54
	GO:0060374	regulation of interleukin-5 production	3.49
	GO:0030217	histone H3-K4 methylation	3.34
	GO:0050673	mast cell differentiation	3.28
	GO:0001806	T cell differentiation	3.17
	GO:0030099	epithelial cell proliferation	3.16
	GO:0003207	type IV hypersensitivity	3.11
	GO:0009966	myeloid cell differentiation	3.10
	GO:0039023	cardiac chamber formation	3.03
	GO:2000176	regulation of signal transduction	2.96
	GO:0072577	pronephric duct morphogenesis	2.95
	GO:0051569	positive regulation of pro-T cell differentiation	2.95
	GO:0050678	endothelial cell apoptotic process	2.87
	GO:0045622	regulation of histone H3-K4 methylation	2.82
	GO:0043383	regulation of epithelial cell proliferation	2.73
	GO:0042423	regulation of T-helper cell differentiation	2.73
		negative T cell selection	
		catecholamine biosynthetic process	

(b): Siginificant GO terms for LUAD data

Subtype	GO ID	Description	Score
1	GO:0007156	homophilic cell adhesion	104.57
	GO:0009966	regulation of signal transduction	81.09
	GO:0014065	phosphatidylinositol 3-kinase signaling	55.97
	GO:0030154	cell differentiation	48.59
	GO:0001817	regulation of cytokine production	45.06
	GO:0007218	neuropeptide signaling pathway	43.54
	GO:0009154	purine ribonucleotide catabolic process	38.25
	GO:0046034	ATP metabolic process	37.63
	GO:0030326	embryonic limb morphogenesis	37.58
	GO:0035556	intracellular signal transduction	35.94
	GO:0030073	insulin secretion	34.76
	GO:0050796	regulation of insulin secretion	34.76
2	GO:0030154	cell differentiation	42.14
	GO:0009966	regulation of signal transduction	36.46
	GO:0007156	homophilic cell adhesion	33.76
	GO:0019722	calcium-mediated signaling	27.30
	GO:0006941	striated muscle contraction	26.71
	GO:0007005	mitochondrion organization	22.02
	GO:0006366	transcription from RNA polymerase II promoter	21.50
	GO:0006357	regulation of transcription from RNA polymerase II promoter	19.16
	GO:0006937	regulation of muscle contraction	18.65
	GO:0045595	regulation of cell differentiation	18.01
	GO:0097190	apoptotic signaling pathway	17.02
	GO:0042127	regulation of cell proliferation	16.27
	GO:0007165	signal transduction	16.16
	GO:0060070	canonical Wnt signaling pathway	15.91
	GO:0035556	intracellular signal transduction	15.83
	GO:0006468	protein phosphorylation	15.82
3	GO:0009966	regulation of signal transduction	55.43
	GO:0007005	mitochondrion organization	34.19
	GO:0097190	apoptotic signaling pathway	32.32
	GO:0030154	cell differentiation	31.35
	GO:0035556	intracellular signal transduction	28.34
	GO:0006366	transcription from RNA polymerase II promoter	25.75
	GO:0007165	signal transduction	24.45
	GO:0006468	protein phosphorylation	22.96
	GO:0043627	response to estrogen	22.55
	GO:0042127	regulation of cell proliferation	22.07
	GO:0006357	regulation of transcription from RNA polymerase II promoter	21.81
	GO:0060565	inhibition of mitotic anaphase-promoting complex activity	21.13
	GO:0014065	phosphatidylinositol 3-kinase signaling	20.27
	GO:0071158	positive regulation of cell cycle arrest	19.19
	GO:0016310	phosphorylation	17.62
	GO:0043491	protein kinase B signaling	17.42
	GO:0048659	smooth muscle cell proliferation	17.27
	GO:0007254	JNK cascade	16.63
	GO:0000082	G1/S transition of mitotic cell cycle	15.69
	GO:0071310	cellular response to organic substance	15.55
	GO:0006915	apoptotic process	15.42
	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	15.10
	GO:0007166	cell surface receptor signaling pathway	15.05
	GO:2000231	positive regulation of pancreatic stellate cell proliferation	14.83
	GO:0001817	regulation of cytokine production	14.50

GO:0060070	canonical Wnt signaling pathway	14.14
GO:0071260	cellular response to mechanical stimulus	13.96
GO:0045595	regulation of cell differentiation	13.93
GO:0016197	endosomal transport	13.81
GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	13.81

(c): Siginificant GO terms for OV data

Subtype	GO ID	Description	Score
	GO:0071158	positive regulation of cell cycle arrest	28.84
	GO:0097190	apoptotic signaling pathway	27.22
	GO:0007005	mitochondrion organization	26.23
	GO:0090399	replicative senescence	22.04
	GO:0035264	multicellular organism growth	18.53
	GO:1901525	negative regulation of macromitophagy	16.94
	GO:0090400	stress-induced premature senescence	16.34
	GO:0002931	response to ischemia	16.04
	GO:0035556	intracellular signal transduction	15.22
	GO:0010666	positive regulation of cardiac muscle cell apoptotic process	14.87
1	GO:0035583	sequestering of TGFbeta in extracellular matrix	13.92
	GO:0006284	base-excision repair	12.22
	GO:0007165	signal transduction	11.37
	GO:0001824	blastocyst development	9.86
	GO:0010332	response to gamma radiation	9.76
	GO:0070231	T cell apoptotic process	9.64
	GO:0006366	transcription from RNA polymerase II promoter	9.32
	GO:2000772	regulation of cellular senescence	9.21
	GO:0034976	response to endoplasmic reticulum stress	8.91
	GO:0042554	superoxide anion generation	7.95
	GO:0009411	response to UV	7.94
	GO:0006915	apoptotic process	7.93
	GO:0009966	regulation of signal transduction	22.83
	GO:0007156	homophilic cell adhesion	17.62
	GO:0014065	phosphatidylinositol 3-kinase signaling	12.47
	GO:0030154	cell differentiation	10.31
	GO:0048011	neurotrophin TRK receptor signaling pathway	9.52
2	GO:0001817	regulation of cytokine production	9.18
	GO:0050865	regulation of cell activation	8.82
	GO:0006366	transcription from RNA polymerase II promoter	8.23
	GO:0060565	inhibition of mitotic anaphase-promoting complex activity	7.74
	GO:0050796	regulation of insulin secretion	7.68
	GO:0030073	insulin secretion	7.68
	GO:0045765	regulation of angiogenesis	7.20
	GO:0030199	collagen fibril organization	5.93
	GO:0030154	cell differentiation	4.94
	GO:0001935	endothelial cell proliferation	4.80
	GO:0061549	sympathetic ganglion development	4.74
	GO:0008625	extrinsic apoptotic signaling pathway via death domain receptors	4.33
	GO:0021897	forebrain astrocyte development	4.04
	GO:0043534	blood vessel endothelial cell migration	3.86
3	GO:0042063	gliogenesis	3.74
	GO:0050678	regulation of epithelial cell proliferation	3.70
	GO:0097190	apoptotic signaling pathway	3.70
	GO:0042552	myelination	3.61
	GO:0042127	regulation of cell proliferation	3.36
	GO:0007160	cell-matrix adhesion	3.02
	GO:0010001	glial cell differentiation	2.98
	GO:0007166	cell surface receptor signaling pathway	2.73
	GO:0008306	associative learning	2.47
	GO:0008542	visual learning	2.47

(d): Siginificant GO terms for UCEC data

Subtype	GO ID	Description	Score
1	GO:0009966	regulation of signal transduction	197.18
	GO:0007156	homophilic cell adhesion	143.56
	GO:0030154	cell differentiation	110.89
	GO:0006366	transcription from RNA polymerase II promoter	96.93
	GO:0035556	intracellular signal transduction	87.27
	GO:0097190	apoptotic signaling pathway	85.73
	GO:0048011	neurotrophin TRK receptor signaling pathway	80.09
	GO:0014065	phosphatidylinositol 3-kinase signaling	74.75
	GO:0007165	signal transduction	74.71
2	GO:0014065	phosphatidylinositol 3-kinase signaling	76.58
	GO:0030154	cell differentiation	58.66
	GO:0043491	protein kinase B signaling	55.85
	GO:0043627	response to estrogen	54.77
	GO:0060070	canonical Wnt signaling pathway	50.31
	GO:0030326	embryonic limb morphogenesis	47.43
	GO:0050821	protein stabilization	38.51
	GO:0097107	postsynaptic density assembly	37.45
	GO:0009966	regulation of signal transduction	36.95
	GO:0051145	smooth muscle cell differentiation	36.16
	GO:0045792	negative regulation of cell size	34.06
	GO:0045595	regulation of cell differentiation	33.00
	GO:0021542	dentate gyrus development	32.55
	GO:0042127	regulation of cell proliferation	31.53
	GO:0035176	social behavior	27.65
	GO:0060134	prepulse inhibition	27.09
	GO:0042552	myelination	27.00
	GO:0071310	cellular response to organic substance	26.04
	GO:0042711	maternal behavior	25.21
	GO:0036092	phosphatidylinositol-3-phosphate biosynthetic process	24.59
	GO:0034394	protein localization to cell surface	24.51
	GO:0030316	osteoclast differentiation	24.16
	GO:0090071	negative regulation of ribosome biogenesis	24.15
3	GO:0071158	positive regulation of cell cycle arrest	43.40
	GO:0090399	replicative senescence	36.67
	GO:0007005	mitochondrion organization	34.87
	GO:1901525	negative regulation of macromitophagy	33.18
	GO:0002931	response to ischemia	29.62
	GO:0090400	stress-induced premature senescence	25.57
	GO:0010666	positive regulation of cardiac muscle cell apoptotic process	23.92
	GO:0097190	apoptotic signaling pathway	23.32
	GO:0006284	base-excision repair	15.90
	GO:0070231	T cell apoptotic process	15.23
	GO:0035583	sequestering of TGFbeta in extracellular matrix	13.30
	GO:0042554	superoxide anion generation	12.58
	GO:2000772	regulation of cellular senescence	12.54
	GO:0090343	positive regulation of cell aging	12.54

Figure S1. Comparison of χ^2 statistics of ONMF and k-means subtypes with clinical features.
ONMF is the stratification via Section 2.4.3, k-means (on W) is the stratification by K-means clustering on ONMF-MP, and k-means (on X) is the stratification by K-means clustering on GO-based profile.
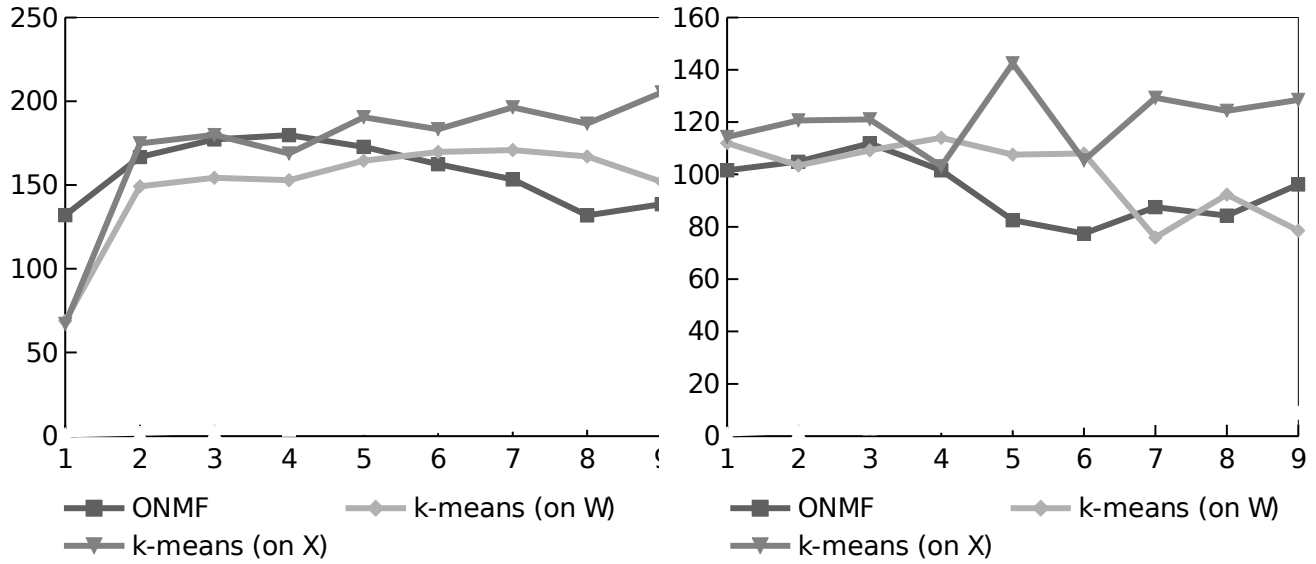
(a): χ^2 statistics of subtypes with histological basis feature on UCEC data

**(b):** χ^2 statistics of subtypes with estrogen receptor status on BRCA data

Figure S2. χ^2 statistics of subtypes with clinical features. X-leaf is a mutation profile represented in GO terms at leaf nodes only, X-raw is a raw mutation profile without Gene-Ontology, and X-all is a mutation profile represented in all GO terms.
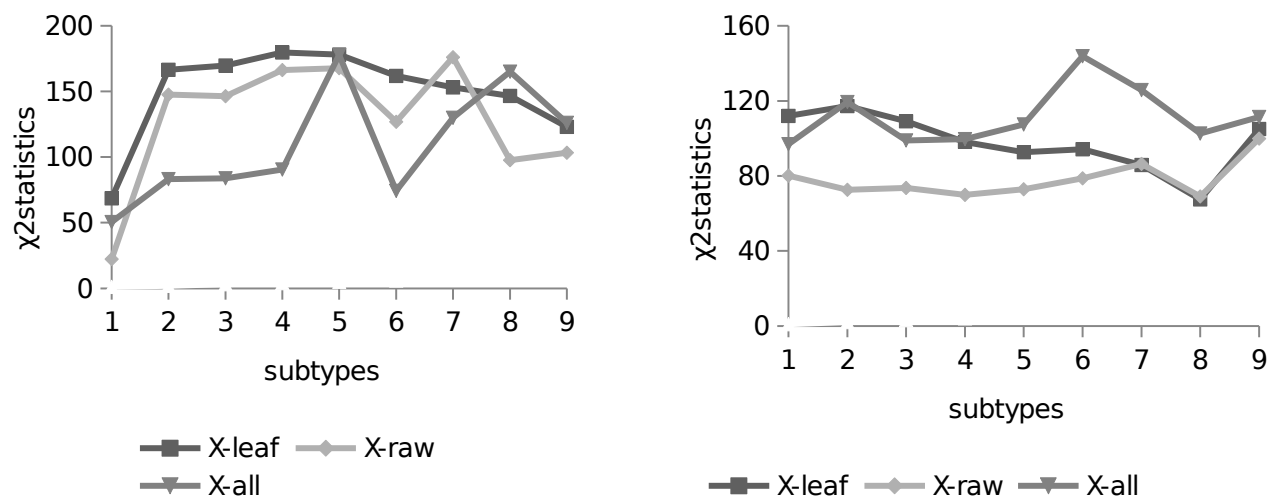
(a): χ^2 statistics of subtypes with histological basis feature on UCEC data

(b): χ^2 statistics of subtypes with estrogen receptor status on BRCA data

Figure S3. Survival curves of the identified subtypes; For each experiment, the number of subtypes is determined based on the log-rank statistics test.
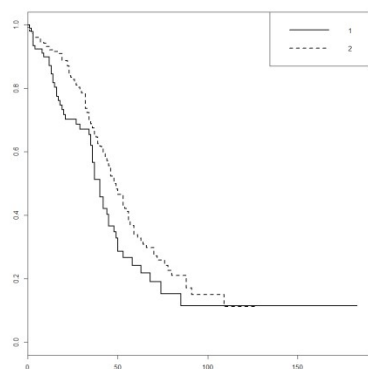
(a): NBS – OV data
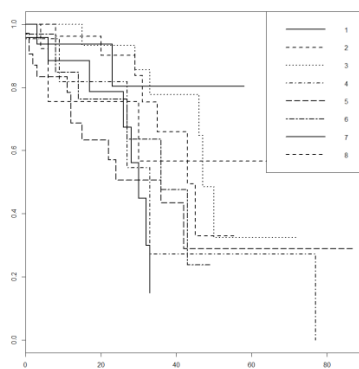
(b): NBS – LUAD data
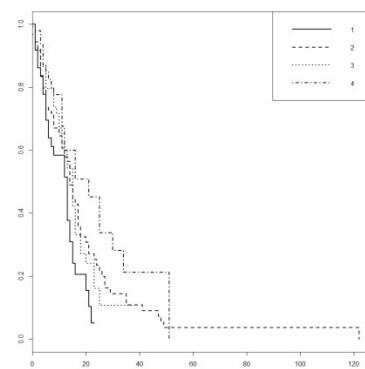
(c): NBS – GBM data
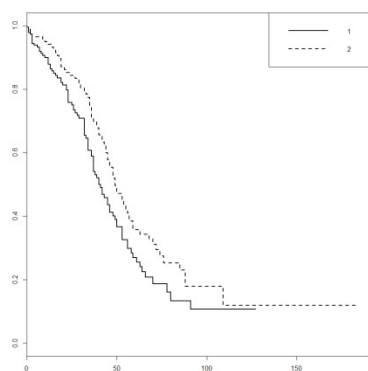
(d): GOS – OV data
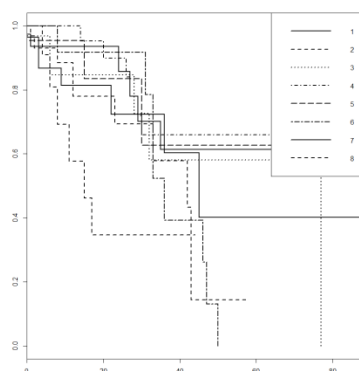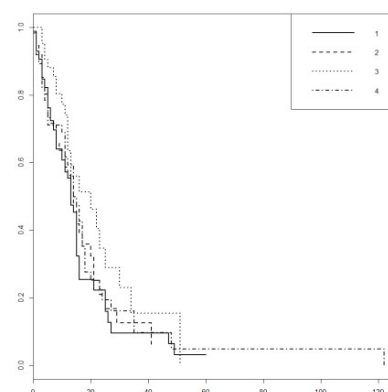
(e): GOS – LUAD data
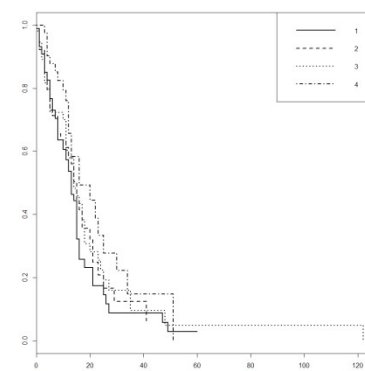
(f): GOS – GBM data
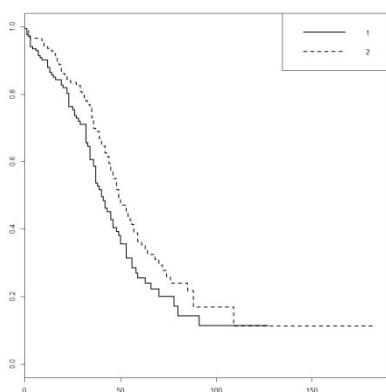
(g): ORGOS – OV data
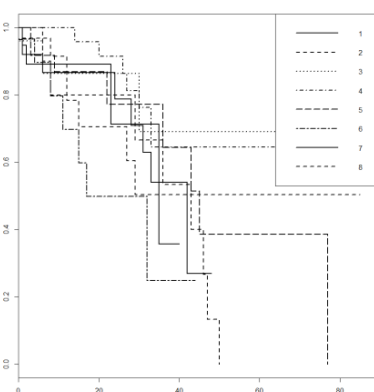
(h): ORGOS – LUAD data

(i): ORGOS – GBM data

Figure S4. χ^2 statistics of cancer subtypes and clinical features of UCEC data using combinations of Gene-Ontology domains (BP: biological process; MF: molecular function; CC: cellular component).
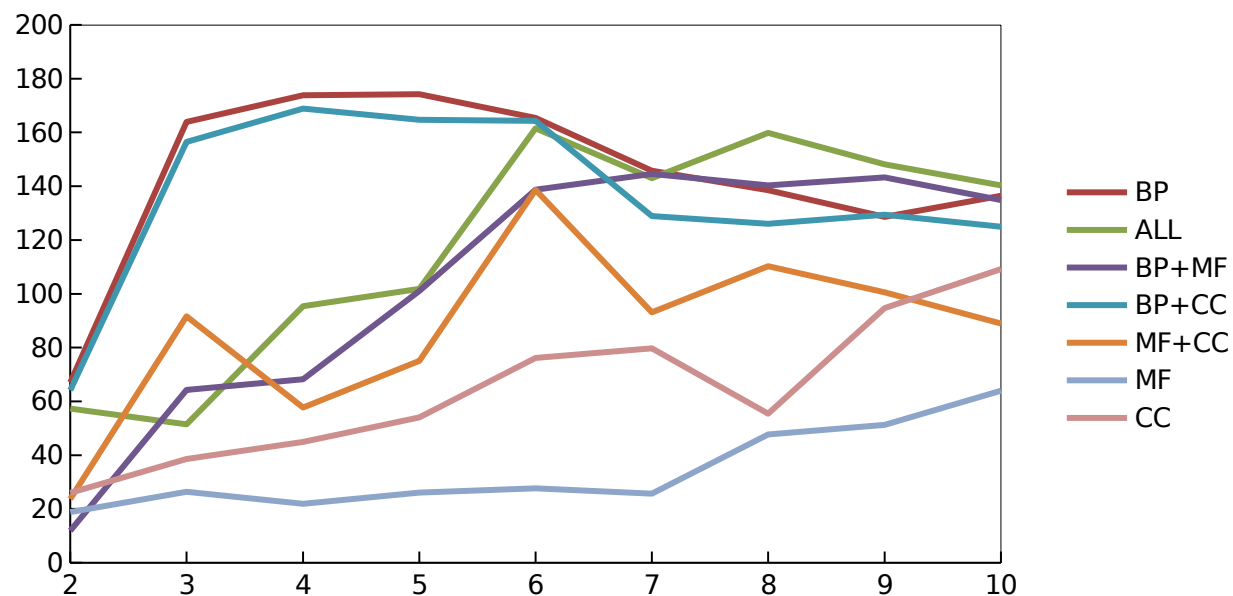
Figure S5. GO-term propagation on UCEC data. The circle sizes are proportional to the accumulated scores and to enhance the visibility only the valid nodes are represented. **A.** Initial leaf node GO terms in ONMF-MP basis vectors. **B.** A subgraph of network propagation where node size indicates the amount of score. The highlighted node is one of the significant GO terms of first subtypes in UCEC data (GO:0009966). **C.** Valid GO terms after score propagation.
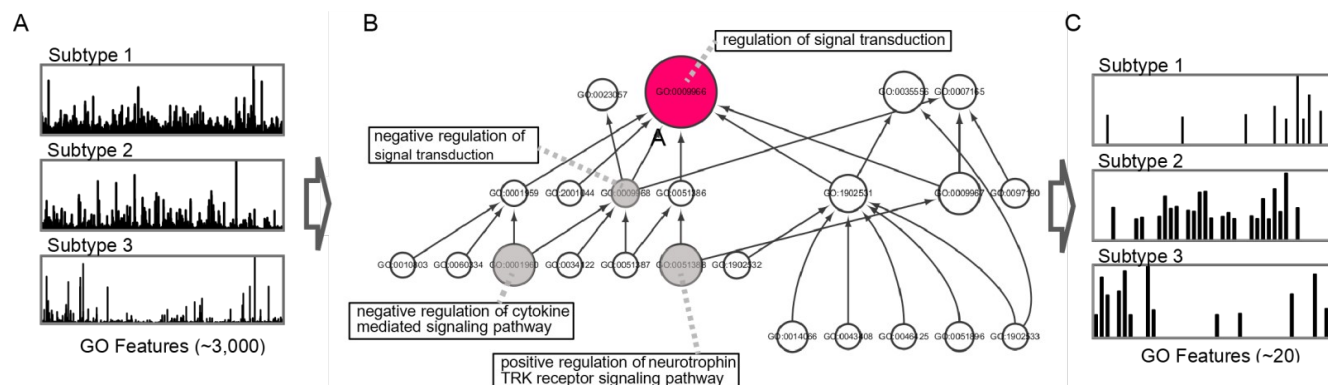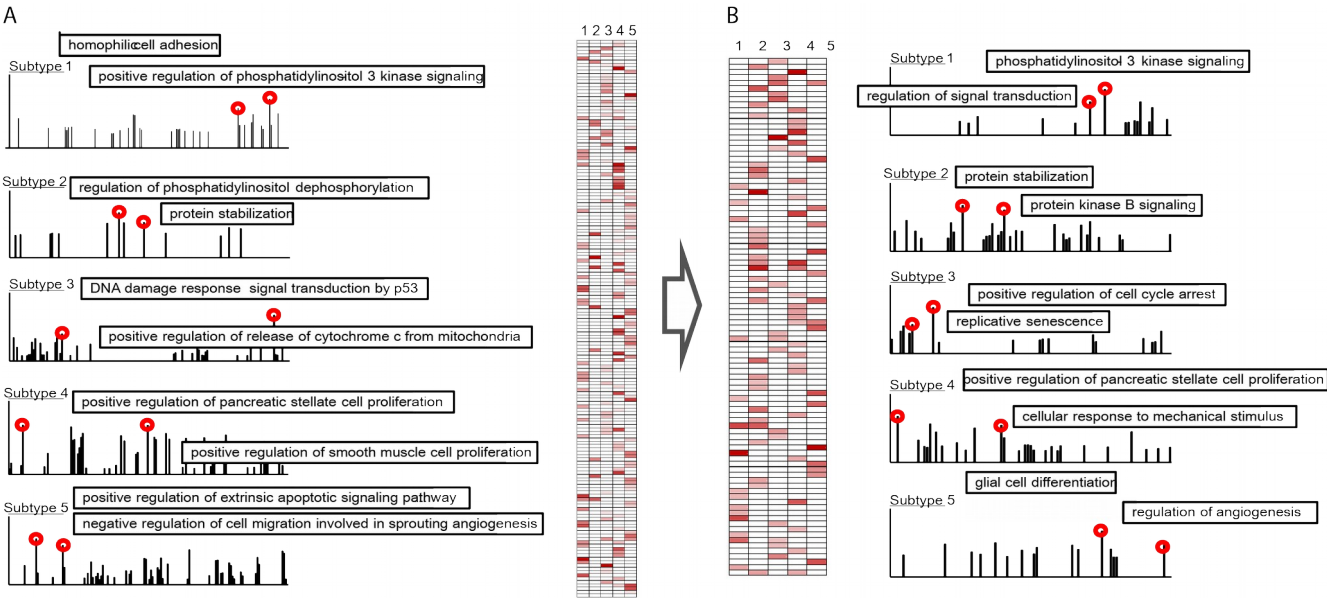
Figure S6. The distributions of GO term scores in cancer subtypes of GBM data. A Initial distribution of terms in basis vectors of ONMF-MP. B Distribution of GO terms after term propagation. The circles in the distribution chart highlight GO terms with the two highest scores.

SUPPLEMENTARY REFERENCES

Jimenez, C., Portela, R. A., Mellado, M., Rodriguez-Frade, J. M., Collard, J., Serrano, A., Martinez-A, C., Avila, J., and Carrera, A. C. (2000). Role of the PI3K regulatory subunit in the control of actin organization and cell migration. J. Cell Biol., 151(2), 249–262.

Ruano, Y., Mollejo, M., Camacho, F. I., Rodriguez de Lope, A., Fiano, C., Ribalta, T., Martinez, P., Hernandez-Moneo, J. L., and Melendez, B. (2008). Identification of survival-related genes of the phosphatidylinositol 3'-kinase signaling pathway in glioblastoma multiforme. Cancer, 112(7), 1575–1584.

Samuels, Y., Schmidt-Kittler, O., Cummins, J., Delong, L., Cheong, I., Rago, C., Huso, D., Lengauer, C., Kinzler, K., and Velculescu, B. V. V. (2005). Mutant pik3ca promotes cell growth and invasion of human cancer cells. Cancer Cell, 7, 561–573.

Wu, G., Feng, X., and Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. Genome Biol., 11(5), R53.

Yoshida, S., Yokota, T., Ujiki, M., Ding, X. Z., Pelham, C., Adrian, T. E., Talamonti, M. S., Bell, R. H., and Denham, W. (2004). Pancreatic cancer stimulates pancreatic stellate cell proliferation and TIMP-1 production through the MAP kinase pathway. Biochem. Biophys. Res. Commun., 323(4), 1241–1245.

Zeng, Z. Z., Jia, Y., Hahn, N. J., Markwart, S. M., Rockwood, K. F., and Livant, D. L. (2006). Role of focal adhesion kinase and phosphatidylinositol 3'-kinase in integrin fibronectin receptor-mediated, matrix metalloproteinase-1-dependent invasion by metastatic prostate cancer cells. Cancer Res., 66(16), 8091–8099.