










Question 1.1:

Output from GO search:

GO POSITIVE REGULATION OF MOLECULAR FUNCTION [1823]	Any process that activates or increases the rate or extent of a molecular function, an elemental biological activity occurring at the molecular level, such as catalysis or binding. [GO:jl]	31		2.2 e ⁻⁹	2.26 e ⁻⁵
GO REGULATION OF HYDROLASE ACTIVITY [1302]	Any process that modulates the frequency, rate or extent of hydrolase activity, the catalysis of the hydrolysis of various bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3. [EC:3.-.-, GOC:ai]	24		3.87 e ⁻⁸	1.99 e ⁻⁴
GO GOLGI APPARATUS [1598]	A membrane-bound cytoplasmic organelle of the endomembrane system that further processes the core oligosaccharides (e.g. N-glycans) added to proteins in the endoplasmic reticulum and packages them into membrane-bound vesicles. The Golgi apparatus operates at the intersection of the secretory, lysosomal, and endocytic pathways. [ISBN:0198506732]	26		1.18 e ⁻⁷	3.5 e ⁻⁴
GO WHOLE MEMBRANE [1721]	Any lipid bilayer that completely encloses some structure, and all the proteins embedded in it or attached to it. Examples include the plasma membrane and most organelle membranes. [GOC:dos]	27		1.36 e ⁻⁷	3.5 e ⁻⁴
GO PROTEIN HOMODIMERIZATION ACTIVITY [655]	Interacting selectively and non-covalently with an identical protein to form a homodimer. [GOC:jl]	16		2.13 e ⁻⁷	3.87 e ⁻⁴
GO PROTEIN DIMERIZATION ACTIVITY [1023]	The formation of a protein dimer, a macromolecular structure	20		2.26 e ⁻⁷	3.87 e ⁻⁴

	consists of two noncovalently associated identical or nonidentical subunits. [ISBN:0198506732]				
GO NEUROGENESIS [1674]	Generation of cells within the nervous system. [GO_REF:0000021, GOC:cls, GOC:curators, GOC:dgh, GOC:dph, GOC:jid]	26		2.88 e ⁻⁷	3.95 e ⁻⁴
GO SMALL GTPASE MEDIATED SIGNAL TRANSDUCTION [509]	Any series of molecular signals in which a small monomeric GTPase relays one or more of the signals. [GOC:mah]	14		3.08 e ⁻⁷	3.95 e ⁻⁴
GO NEURON PROJECTION [1366]	A prolongation or process extending from a nerve cell, e.g. an axon or dendrite. [GOC:jl, http://www.cogsci.princeton.edu/~wn/]	23		3.73 e ⁻⁷	4.14 e ⁻⁴
GO CELL MOTILITY [1819]	Any process involved in the controlled self-propelled movement of a cell that results in translocation of the cell from one place to another. [GOC:dgh, GOC:dph, GOC:isa_complete, GOC:mlg]				

Relations to ovarian serous cystadenocarcinoma (OV):

GO_POSITIVE_REGULATION_OF_MOLECULAR_FUNCTION:

This output indicates any process that increases the rate of catalysis or binding throughout the cell. It appears in relation to Yap1, a gene whose mutations are known to onset Ovarian Serous Cystadenocarcinoma. Gene cards indicates increases in protein binding with YAP 1 and several other genes are indicative of ovarian serous cystadenocarcinoma.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5428545/>

https://www.malacards.org/card/ovarian_serous_cystadenocarcinoma

GO_REGULATION_OF_HYDROLASE_ACTIVITY:

Anything that changes the rate of hydrolase activity. This is also indicated in relation to the oncogene Yap1 as a risk for ovarian serous cystadenocarcinoma.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5428545/>

GO_GOLGI_APPARATUS:

This indicates increased activity in processing the core oligosaccharides that were added to proteins in the ER and packages them into membrane-bound vesicles. The SYNE1 gene is shown to be involved in nuclear organization and structural integrity, function of the Golgi apparatus, and cytokinesis. It is known to be downregulated in ovarian and other cancers, indicating Golgi apparatus process alterations may also be altered.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7330690/>

GO_WHOLE_MEMBRANE:

This indicates any lipid bilayer that completely encloses some structure and all proteins imbedded or attached to it. Irregular nuclear membranes and larger nuclei are tied to ovarian carcinomas. Atypical cell and organelle formation are also common in ovarian cancers. Given that the proteins that shape the cell necessarily must attach to the membrane, irregularities here could also be considered a link between the GO term and ovarian cancer

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745605/>

GO_PROTEIN_HOMODIMERIZATION_ACTIVITY:

Indicates selective interaction with an identical protein, non-covalently to form a homodimer. Yap1 also shows an association here in conjunction with its role on ovarian cancer. Indicating a possible tie. Protein binding and identical protein binding is also listed on gene cards as related to OSC.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5428545/>

https://www.malacards.org/card/ovarian_serous_cystadenocarcinoma#goterms

GO_PROTEIN_DIMERIZATION_ACTIVITY:

Noncovalent binding of two or more identical or not subunits. The above indication applies here as well. Several oncogenes such as PDCD6 also indicate non-homologous dimers as well.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5428545/>

https://www.malacards.org/card/ovarian_serous_cystadenocarcinoma#goterms

<http://www.cancerindex.org/geneweb/PDCD6.htm>

GO_NEUROGENESIS:

Generation of cells within the nervous system. Genecards notes negative regulation of the neuron apoptotic process. The attached paper also shows evidence of neurogenesis pathways being significant for ovarian cancer.

https://www.malacards.org/card/ovarian_serous_cystadenocarcinoma#goterms

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6710300/>

GO_SMALL_GTPASE_MEDIATED_SIGNAL_TRANSDNSDUCTION:

Any signaling pathway including a small monomeric GTPase relaying signal(s). Several GTPase proteins show up on genecards for Ovarian Serous Cystadenocarcinoma, indicating that GTPase pathways are involved in that cancer.

https://www.malacards.org/card/ovarian_serous_cystadenocarcinoma#goterms

GO_NEURON_PROJECTION:

This is the process of prolonging or extending nerve cells. eEF1A Proteins serve as neuron elongation factors and their disfunction is linked to ovarian cancers, indicating a strong relation of ovarian cancer and neuron projection.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4387925/>

GO_CELL_MOTILITY:

Any process involved in the self-propelled movement of the cell from one place to another. This is definitionally tied to all cancers, if it is a cancer it can move from one tissue to another. This is necessary for metastasis. Motility is often used as an indicator of cancer-like activity, with several pathways implicated for this cancer.

[https://www.cell.com/cell-reports/pdf/S2211-1247\(18\)30102-5.pdf](https://www.cell.com/cell-reports/pdf/S2211-1247(18)30102-5.pdf)






<https://pubmed.ncbi.nlm.nih.gov/30979497/>

<https://pubmed.ncbi.nlm.nih.gov/29759990/>

Question 2.2:

List of genes: ACTG1, ACTN4, P4HB, CFL1, COL3A1, PTMA, TPM4, ALDOA, RPS6, RPL5, GNB2L1, ARF1, MIF, RPS14, RPSAP58, FN1, CD81, PSAP, SPON1, HSPA8, TAGLN2, SPARC, RPS4X, RPL10A, COL1A1, RPL15, RPLP2, TUBA1B, TMSL3, EEF1G, GNAS, HSP90AB1, AHNAK, SPINT2, RPL30, RPS3, RPL29, MDK, CLU, TMSB10, EEF2, ACTB, JUP, HSPB1, TUBB, UBC, RPL13, RPS9, UBA52, LGALS3BP, MYH9, COL1A2, SLC34A2, DHCR24, TMBIM6, PKM2, RPL37, RPL32, RPLP0, LY6E, RPS16, EEF1A1, RPL3, HSP90AA1, B2M, CTSB, PLEC, RPS8, RPS2, EIF4G2, C3, CANX, LDHA, RPL28, RPS20, YWHAZ, PTPRF, ATP5B, RPS17, WFDC2, RPL41, RPL37A, GAPDH, CTSD, HDGF, HSP90B1, RPL11, CD24, HLA-DRA, LAPTM4B, RPS24, YBX1, HDLBP, PABPC1, CD74, FTH1, KRT8, APP, RPL27, PFN1, RPS19, RPL4, RPL35A, UBA1, FTL, RPL8, RPLP1, AGRN, MALAT1, RPL19, MYL6, SRRM2, BCAM, RPL18, ITM2B, RPS11, FLNA, CALR, HLA-C, HLA-B, VIM, IGFBP2, TPT1, RPS5, ATP1A1, ENO1, KRT18, IFITM3

Output from GO search:

Gene Set Name [# Genes (K)]	Description	# Gene s in Over lap (k)	k/K	p- valu e ?	FDR q- valu e ?
GO CYTOSOLIC RIBOSOME [110]	A ribosome located in the cytosol. [GOC:mtg_sensu]	38		1.38 e ⁻⁶⁸	1.42 e ⁻⁶⁴
GO COTRANSLATIONAL PROTEIN TARGETING TO MEMBRANE [106]	The targeting of proteins to a membrane that occurs during translation. The transport of most secretory proteins, particularly those with more than 100 amino acids, into the endoplasmic reticulum lumen occurs in this manner, as does the import of some proteins into mitochondria. [ISBN:0716731363, PMID:10512867, PMID:16896215]	37		6 e ⁻⁶⁷	2.51 e ⁻⁶³
GO NUCLEAR TRANSCRIBED MRNA CATABOLIC PROCESS NONSENSE MEDIATED DECAY [120]	The nonsense-mediated decay pathway for nuclear-transcribed mRNAs degrades mRNAs in which an amino-acid codon has changed to a nonsense codon; this prevents the translation of such mRNAs into truncated, and potentially harmful, proteins. [GOC:krc, GOC:ma, PMID:10025395]	38		7.34 e ⁻⁶⁷	2.51 e ⁻⁶³
GO ESTABLISHMENT OF PROTEIN LOCALIZATION TO ENDOPLASMIC RETICULUM [121]	The directed movement of a protein to a specific location in the endoplasmic reticulum. [GOC:mah]	37		2.12 e ⁻⁶⁴	5.44 e ⁻⁶¹
GO SYMBIOTIC PROCESS [980]	A process carried out by gene products in an organism that enable the	61		3.44 e ⁻⁶³	7.07 e ⁻⁶⁰

	<p>organism to engage in a symbiotic relationship, a more or less intimate association, with another organism. The various forms of symbiosis include parasitism, in which the association is disadvantageous or destructive to one of the organisms; mutualism, in which the association is advantageous, or often necessary to one or both and not harmful to either; and commensalism, in which one member of the association benefits while the other is not affected. However, mutualism, parasitism, and commensalism are often not discrete categories of interactions and should rather be perceived as a continuum of interaction ranging from parasitism to mutualism. In fact, the direction of a symbiotic interaction can change during the lifetime of the symbionts due to developmental changes as well as changes in the biotic/abiotic environment in which the interaction occurs. Microscopic symbionts are often referred to as endosymbionts. [GOC:cc, https://study.com/academy/lesson/symbiont-definition-lesson-quiz.html]</p>				
GO_TRANSLATIONAL_INITIATION [192]	The process preceding formation of the peptide bond between the first two	40		3.91×10^{-62}	6.69×10^{-59}

	amino acids of a protein. This includes the formation of a complex of the ribosome, mRNA or circRNA, and an initiation complex that contains the first aminoacyl-tRNA. [ISBN:019879276X]				
GO STRUCTURAL MOLECULE ACTIVITY [709]	The action of a molecule that contributes to the structural integrity of a complex or its assembly within or outside a cell. [GOC:mah, GOC:vw]	55		1.07 e^{-61}	1.58 e^{-58}
GO PROTEIN LOCALIZATION TO ENDOPLASMIC MIC RETICULUM [149]	A process in which a protein is transported to, or maintained in, a location within the endoplasmic reticulum. [GOC:mah]	37		1.57 e^{-60}	2.02 e^{-57}
GO RIBOSOMAL SUBUNIT [187]	Either of the two subunits of a ribosome: the ribosomal large subunit or the ribosomal small subunit. [GOC:jl]	38		1.7 e^{-58}	1.94 e^{-55}
GO CELL SUBSTRATE JUNCTION [423]	A cell junction that forms a connection between a cell and the extracellular matrix. [GOC:aruk, GOC:bc, GOC:hb, GOC:mah, PMID:10419689, PMID:1643657, PMID:16805308, PMID:26923917, PMID:8314002]	46		6.73 e^{-58}	6.92 e^{-55}

GO_CYTOSOLIC_RIBOSOME: Genes that relate to the Cytosolic rhybosomes

GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE: Genes that relate to the targeting of proteins to the membrane during translation.

GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY: The nonsense mediated decay pathway nt-mRNAs that degrades mRNAs in which the codon has changed to the nonsense codon. This prevents translation of the mRNAs into truncated and/or harmful proteins.

GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM: the direct movement of a protein to the ER.

GO_SYMBIOTIC_PROCESS: processes that enable symbiotic relationships with a host organism.

GO_TRANSLATIONAL_INITIATION: The process preceding the first peptide bond in peptide synthesis. This includes formation of the ribosome, mRNA or circRNA, as well as an initiation complex that contains the first aminoacyl-tRNA.

GO_STRUCTURAL_MOLECULE_ACTIVITY: Molecular actions that contribute to the structural integrity of a complex in or outside a cell.

GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM: A process where a protein is transported to or maintained within the ER.

GO_RIBOSOMAL_SUBUNIT: either of the two ribosomal subunits.

GO_CELL_SUBSTRATE_JUNCTION: a junction that forms between a cell and the ECM.

Question 3.2:

Accuracy is better when taking the top 1000 because the genes with lower P-values are removed providing little to no value for sorting. Without removing lower P-value genes, the machine is likely to get a lot of false signals from extraneous data, ultimately reducing its accuracy.