

Supplementary material

1 Non-small cell lung cancer data

1.1 GRF results

Table 1: Top GO biological processes using the top 100 most important features selected by GRF.

GOBPID	P-value	Adj-P	Term
GO:0051052	0.018	0.015	regulation of DNA metabolic process
GO:0000723	0.018	0.015	telomere maintenance
GO:0032069	0.018	0.015	regulation of nuclelease activity
GO:0032200	0.018	0.015	telomere organization
GO:0032211	0.018	0.015	negative regulation of telomere maintenance via telomerase
GO:0051098	0.018	0.015	regulation of binding
GO:0048598	0.034	0.028	embryonic morphogenesis
GO:1904357	0.034	0.028	negative regulation of telomere maintenance via telomere lengthening
GO:0042098	0.034	0.028	T cell proliferation
GO:0006303	0.034	0.028	double-strand break repair via nonhomologous end joining
GO:0002562	0.034	0.028	somatic diversification of immune receptors via germline recombination within a single locus
GO:0016444	0.034	0.028	somatic cell DNA recombination
GO:0051053	0.034	0.028	negative regulation of DNA metabolic process
GO:0032205	0.034	0.028	negative regulation of telomere maintenance
GO:2000279	0.034	0.028	negative regulation of DNA biosynthetic process
GO:0090026	0.034	0.028	positive regulation of monocyte chemotaxis
GO:0050778	0.034	0.028	positive regulation of immune response
GO:0007492	0.034	0.028	endoderm development
GO:0042110	0.034	0.028	T cell activation
GO:0006266	0.034	0.028	DNA ligation
GO:2000278	0.034	0.028	regulation of DNA biosynthetic process
GO:0002200	0.034	0.028	somatic diversification of immune receptors
GO:0042129	0.034	0.028	regulation of T cell proliferation
GO:0045088	0.034	0.028	regulation of innate immune response
GO:0008286	0.035	0.028	insulin receptor signaling pathway
GO:0002253	0.036	0.030	activation of immune response
GO:0071417	0.037	0.031	cellular response to organonitrogen compound
GO:0006310	0.038	0.032	DNA recombination
GO:0033143	0.040	0.033	regulation of intracellular steroid hormone receptor signaling pathway
GO:0090025	0.040	0.033	regulation of monocyte chemotaxis
GO:0006282	0.040	0.033	regulation of DNA repair
GO:0051348	0.042	0.035	negative regulation of transferase activity
GO:0007004	0.043	0.035	telomere maintenance via telomerase
GO:0050731	0.043	0.035	positive regulation of peptidyl-tyrosine phosphorylation
GO:0006278	0.043	0.035	RNA-dependent DNA biosynthetic process

GO:1903829	0.044	0.037	positive regulation of cellular protein localization
GO:0050921	0.046	0.038	positive regulation of chemotaxis
GO:0002690	0.046	0.038	positive regulation of leukocyte chemotaxis
GO:0044089	0.046	0.038	positive regulation of cellular component biogenesis
GO:0002252	0.046	0.038	immune effector process
GO:0030335	0.046	0.038	positive regulation of cell migration
GO:0006302	0.046	0.038	double-strand break repair
GO:0050730	0.046	0.038	regulation of peptidyl-tyrosine phosphorylation
GO:1901699	0.046	0.038	cellular response to nitrogen compound
GO:0002366	0.046	0.038	leukocyte activation involved in immune response
GO:0001819	0.046	0.038	positive regulation of cytokine production
GO:0050863	0.046	0.038	regulation of T cell activation
GO:0051972	0.046	0.038	regulation of telomerase activity
GO:0010833	0.046	0.038	telomere maintenance via telomere lengthening
GO:0032869	0.046	0.038	cellular response to insulin stimulus
GO:0031400	0.046	0.038	negative regulation of protein modification process
GO:0002263	0.046	0.038	cell activation involved in immune response
GO:1903037	0.046	0.038	regulation of leukocyte cell-cell adhesion
GO:2000147	0.046	0.038	positive regulation of cell motility

1.2 Random Forest result

Table 2: Top GO biological processes using the top 100 most important features selected by RF.

GOBPID	P-value	Adj-P	Term
GO:0008544	0.00003	0.00002	epidermis development
GO:0030216	0.003	0.003	keratinocyte differentiation
GO:0061436	0.003	0.003	establishment of skin barrier
GO:0030855	0.003	0.003	epithelial cell differentiation
GO:0033561	0.004	0.004	regulation of water loss via skin
GO:0043588	0.004	0.004	skin development
GO:0002009	0.012	0.011	morphogenesis of an epithelium
GO:0009913	0.015	0.014	epidermal cell differentiation
GO:0048729	0.015	0.014	tissue morphogenesis
GO:0031424	0.025	0.023	keratinization
GO:0043616	0.025	0.023	keratinocyte proliferation
GO:0014032	0.028	0.025	neural crest cell development
GO:0048608	0.028	0.025	reproductive structure development
GO:0060707	0.028	0.025	trophoblast giant cell differentiation
GO:0008593	0.028	0.025	regulation of Notch signaling pathway
GO:0061458	0.028	0.025	reproductive system development
GO:0014031	0.028	0.025	mesenchymal cell development
GO:0014033	0.028	0.025	neural crest cell differentiation
GO:0048864	0.028	0.025	stem cell development
GO:0050891	0.028	0.025	multicellular organismal water homeostasis
GO:0045103	0.028	0.026	intermediate filament-based process
GO:0045104	0.028	0.026	intermediate filament cytoskeleton organization
GO:0030104	0.036	0.033	water homeostasis
GO:0140115	0.036	0.033	export across plasma membrane
GO:0045109	0.036	0.033	intermediate filament organization
GO:0001738	0.045	0.042	morphogenesis of a polarized epithelium

2 Human embryonic stem cell data

2.1 GRF results

Table 3: Top GO biological processes using the top 100 most important features selected by GRF.

GOBPID	P-value	Adj-P	Term
GO:0010718	6.08e-08	0.000153	positive regulation of epithelial to mesenchymal transition
GO:0050808	1.06e-07	0.000153	synapse organization
GO:0010717	2.7e-07	0.000194	regulation of epithelial to mesenchymal transition
GO:0034109	2.7e-07	0.000194	homotypic cell-cell adhesion
GO:0007178	3.38e-07	0.000194	transmembrane receptor protein serine/threonine kinase signaling pathway
GO:0070527	6.82e-07	0.000291	platelet aggregation
GO:0048667	7.23e-07	0.000291	cell morphogenesis involved in neuron differentiation
GO:0048812	8.08e-07	0.000291	neuron projection morphogenesis
GO:1903706	1.12e-06	0.000332	regulation of hemopoiesis
GO:0001837	1.15e-06	0.000332	epithelial to mesenchymal transition
GO:0061564	2.44e-06	0.000639	axon development
GO:0071417	3.41e-06	0.000748	cellular response to organonitrogen compound
GO:1902414	3.55e-06	0.000748	protein localization to cell junction
GO:0090092	3.64e-06	0.000748	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
GO:0030324	4.24e-06	0.000813	lung development
GO:0009611	4.52e-06	0.000813	response to wounding
GO:0030323	5.02e-06	0.00085	respiratory tube development
GO:1901653	7.07e-06	0.00113	cellular response to peptide
GO:0050679	8.13e-06	0.00123	positive regulation of epithelial cell proliferation
GO:0060541	1.05e-05	0.00151	respiratory system development
GO:0030855	1.37e-05	0.00187	epithelial cell differentiation
GO:0033077	1.48e-05	0.00194	T cell differentiation in thymus
GO:0090287	1.65e-05	0.00206	regulation of cellular response to growth factor stimulus
GO:0048514	1.83e-05	0.00214	blood vessel morphogenesis
GO:0060485	1.86e-05	0.00214	mesenchyme development
GO:0071364	1.98e-05	0.00219	cellular response to epidermal growth factor stimulus
GO:0007409	2.15e-05	0.00221	axonogenesis
GO:0050678	2.17e-05	0.00221	regulation of epithelial cell proliferation
GO:0048762	2.23e-05	0.00221	mesenchymal cell differentiation
GO:0030168	2.91e-05	0.00264	platelet activation
GO:0000226	2.93e-05	0.00264	microtubule cytoskeleton organization
GO:0031589	3.03e-05	0.00264	cell-substrate adhesion
GO:0010001	3.03e-05	0.00264	glial cell differentiation
GO:0070849	3.22e-05	0.00272	response to epidermal growth factor
GO:0007507	3.47e-05	0.0028	heart development
GO:2000177	3.51e-05	0.0028	regulation of neural precursor cell proliferation
GO:1901652	4.4e-05	0.00342	response to peptide
GO:0001525	4.52e-05	0.00342	angiogenesis
GO:0070371	5.38e-05	0.00397	ERK1 and ERK2 cascade
GO:0048709	5.53e-05	0.00398	oligodendrocyte differentiation
GO:0034341	5.91e-05	0.00415	response to interferon-gamma
GO:0045637	6.45e-05	0.00442	regulation of myeloid cell differentiation
GO:0045596	6.63e-05	0.00444	negative regulation of cell differentiation
GO:0048589	7.09e-05	0.00449	developmental growth
GO:0044849	7.17e-05	0.00449	estrous cycle
GO:0098974	7.17e-05	0.00449	postsynaptic actin cytoskeleton organization

GO:1902692	7.88e-05	0.00483	regulation of neuroblast proliferation
GO:0030098	8.49e-05	0.00499	lymphocyte differentiation
GO:0050673	8.49e-05	0.00499	epithelial cell proliferation
GO:1903827	9.99e-05	0.00575	regulation of cellular protein localization

2.2 Random Forest results

Table 4: Top GO biological processes using the top 100 most important features selected by RF.

GOBPID	P-value	Adj P	Term
GO:0048667	5.16e-10	1.25e-06	cell morphogenesis involved in neuron differentiation
GO:0048812	4.81e-09	5.83e-06	neuron projection morphogenesis
GO:0007409	1.54e-08	1.24e-05	axonogenesis
GO:0010001	5.91e-08	3.38e-05	glial cell differentiation
GO:0061564	6.96e-08	3.38e-05	axon development
GO:0042063	1.7e-07	6.87e-05	gliogenesis
GO:0009611	8.96e-07	0.00031	response to wounding
GO:0007411	1.77e-06	0.000502	axon guidance
GO:0097485	1.86e-06	0.000502	neuron projection guidance
GO:0022406	2.95e-06	0.000716	membrane docking
GO:0048709	4.62e-06	0.00102	oligodendrocyte differentiation
GO:0006935	7.38e-06	0.00149	chemotaxis
GO:0042330	7.99e-06	0.00149	taxis
GO:0031629	2.28e-05	0.00369	synaptic vesicle fusion to presynaptic active zone membrane
GO:0099500	2.28e-05	0.00369	vesicle fusion to plasma membrane
GO:0000226	2.93e-05	0.00426	microtubule cytoskeleton organization
GO:0048708	2.99e-05	0.00426	astrocyte differentiation
GO:0050804	3.67e-05	0.00468	modulation of chemical synaptic transmission
GO:0099177	3.67e-05	0.00468	regulation of trans-synaptic signaling
GO:0050769	4.22e-05	0.00512	positive regulation of neurogenesis
GO:0060840	5.53e-05	0.00635	artery development
GO:0048499	5.76e-05	0.00635	synaptic vesicle membrane organization
GO:0017156	6.12e-05	0.00645	calcium-ion regulated exocytosis
GO:0061005	6.75e-05	0.00669	cell differentiation involved in kidney development
GO:0048589	7.09e-05	0.00669	developmental growth
GO:1903651	7.17e-05	0.00669	positive regulation of cytoplasmic transport
GO:0031099	8.38e-05	0.00731	regeneration
GO:0120035	8.44e-05	0.00731	regulation of plasma membrane bounded cell projection organization
GO:0051962	9.39e-05	0.0076	positive regulation of nervous system development
GO:0002328	9.8e-05	0.0076	pro-B cell differentiation
GO:0014015	9.81e-05	0.0076	positive regulation of gliogenesis
GO:0072009	0.000102	0.0076	nephron epithelium development
GO:0001505	0.000103	0.0076	regulation of neurotransmitter levels
GO:0060411	0.000117	0.00834	cardiac septum morphogenesis
GO:0021782	0.000123	0.00841	glial cell development
GO:0051960	0.000125	0.00841	regulation of nervous system development
GO:0001889	0.000128	0.00841	liver development
GO:0048713	0.000138	0.0086	regulation of oligodendrocyte differentiation
GO:0048844	0.000138	0.0086	artery morphogenesis
GO:0061008	0.000148	0.00897	hepatobiliary system development
GO:0007423	0.000171	0.0101	sensory organ development
GO:0060560	0.000186	0.0108	developmental growth involved in morphogenesis
GO:0050767	0.000196	0.0111	regulation of neurogenesis

GO:0045687	0.000223	0.0122
GO:0030855	0.000227	0.0122
GO:0022612	0.000233	0.0123
GO:0030324	0.000252	0.013
GO:0016082	0.000264	0.0133
GO:0030323	0.000286	0.0137
GO:0035107	0.000288	0.0137

positive regulation of glial cell differentiation
epithelial cell differentiation
gland morphogenesis
lung development
synaptic vesicle priming
respiratory tube development
appendage morphogenesis

3 Simulation results with ba-power = 1

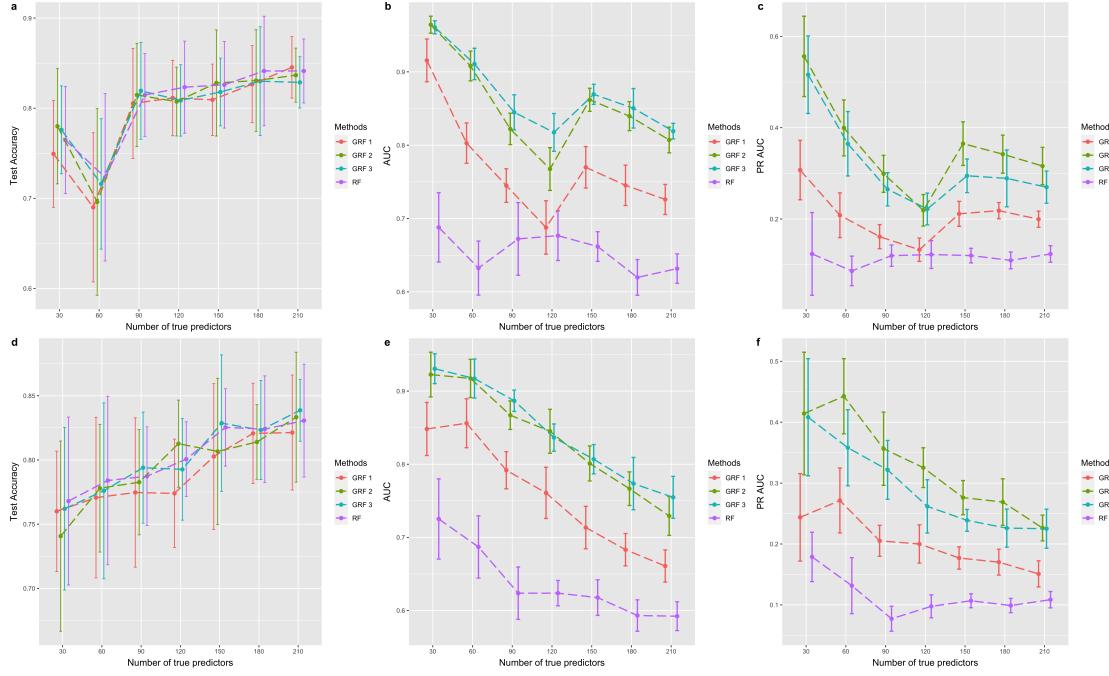


Figure 1: The figures show AUC and PR-AUC of the model with different number of true predictors using logistic link function. The two figures(a,b) in the first line is the AUC with one core and two cores, and the figures(c,d) in the second line show the PR-AUC with one core and two cores respectively. In each figure, we show the metric using random forest and our graph-based random forest with different parameters of selective range, i.e. G_RF 2 means that for each core node, we use the nodes within two hops to set a tree model.

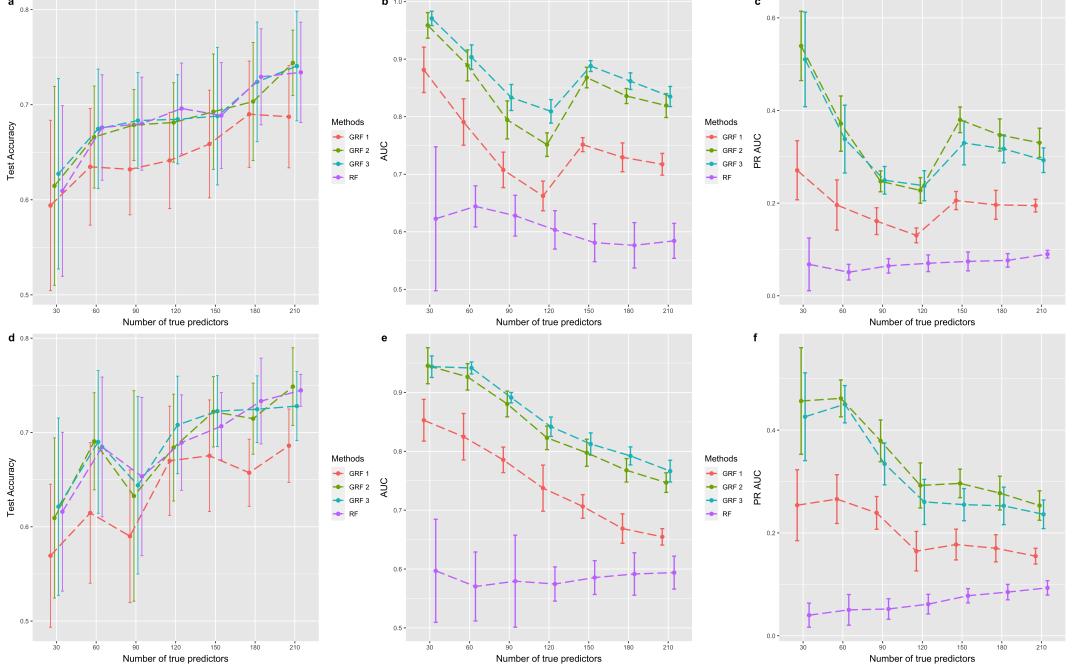


Figure 2: The figures show AUC and PR-AUC of the model with different number of true predictors using absolute link function. The two figures(a,b) in the first line is the AUC with one core and two cores, and the figures(c,d) in the second line show the PR-AUC with one core and two cores respectively. In each figure, we show the metric using random forest and our graph-based random forest with different parameters of selective range, i.e. G.RF 2 means that for each core node, we use the nodes within two hops to set a tree model.

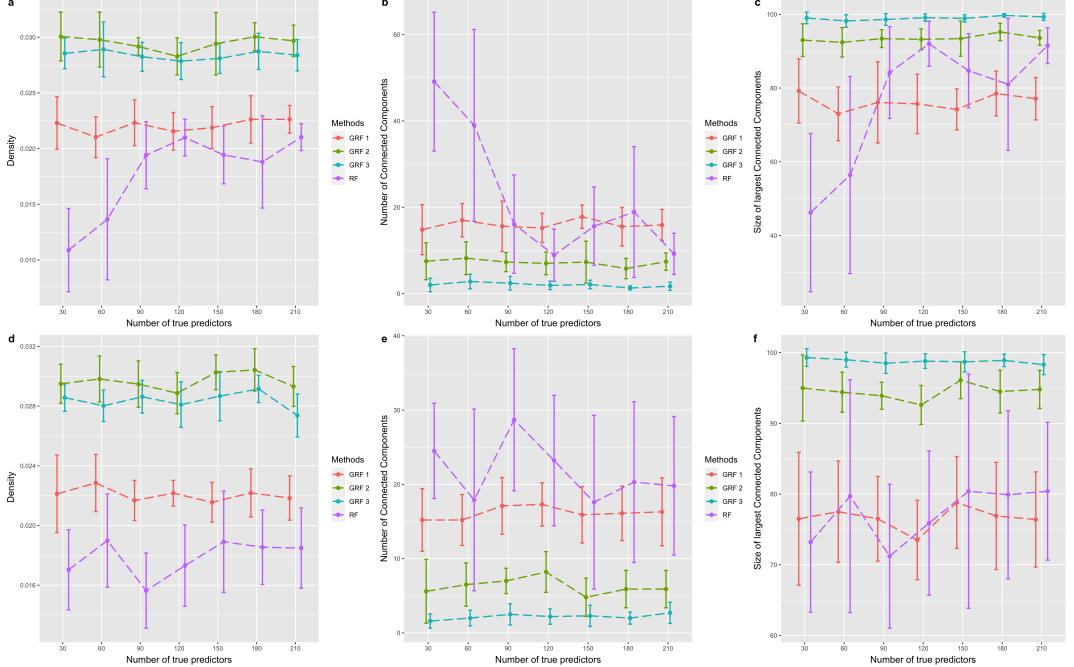


Figure 3: The figures show the selected subgraph density and number of connected components of the model with different number of true predictors using logistics link function. For each methods, we select the top 100 features and build the corresponding subgraph. The two figures(a,b) in the first line is the density with one core and two cores, and the figures(c,d) in the second line show the number of connected components with one core and two cores respectively.

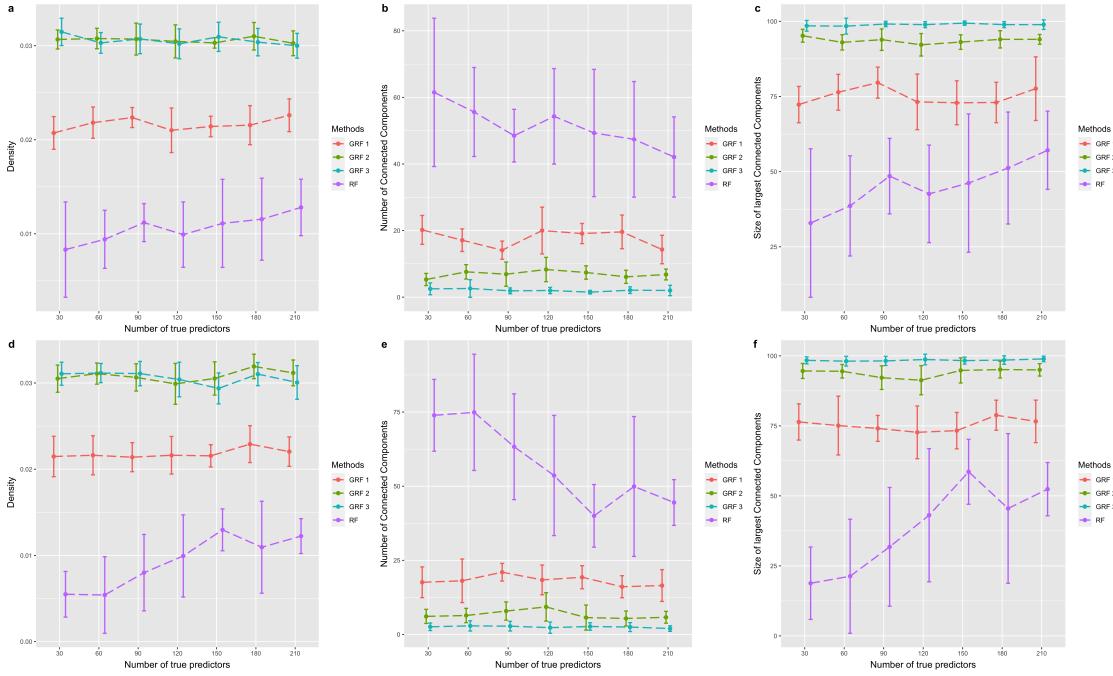


Figure 4: The figures show the selected subgraph density and number of connected components of the model with different number of true predictors using absolute link function. For each methods, we select the top 100 features and build the corresponding subgraph. The two figures(a,b) in the first line is the density with one core and two cores, and the figures(c,d) in the second line show the number of connected components with one core and two cores respectively.