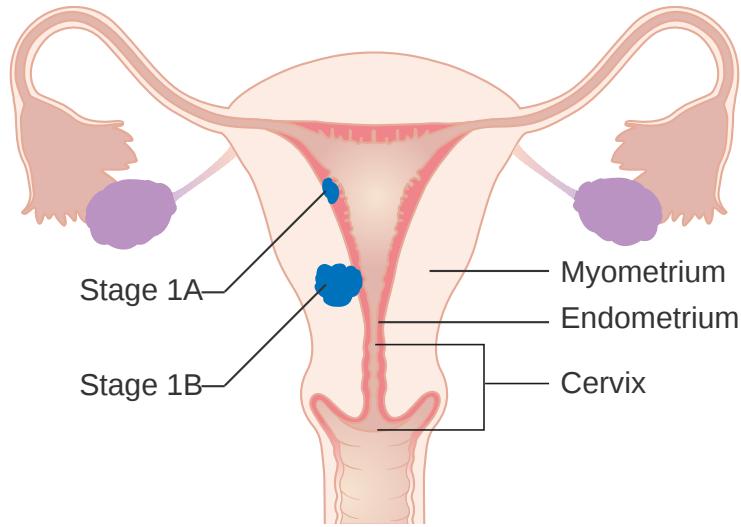


Classification of Endometrial cancer using Machine Learning

Robert Różański

Endometrial cancer



Type I (less aggressive):

- Endometrioid adenocarcinoma

Type II (more aggressive):

- Serous cystadenocarcinoma
- Papillary serous adenocarcinoma
- Clear cell adenocarcinoma

Data

QIAseq Targeted DNA Panels

 Print



Digital DNA sequencing to confidently detect low-frequency variants

- Digital sequencing enabled by molecular barcodes to remove PCR duplicates
- Complete Sample to Insight solution streamlines the workflow
- Compatibility with low-quality DNA enables efficient sequencing of FFPE and cfDNA samples
- Minimal DNA Input to preserve precious samples
- Optimized buffers and conditions to achieve high coverage of GC-rich regions

The QIAseq targeted DNA Panels have been developed as a complete Sample to Insight solution to enable digital DNA sequencing by utilizing molecular barcodes. Digital DNA sequencing is a unique approach to detect low-frequency variants with high confidence by overcoming the issues of PCR duplicates, false positives and library bias.

Data

Adenoma

Tubular Adenoma: **BRAF, FBXW7, KRAS.**

Tubulovillous Adenoma: **BRAF, CTNNB1, KRAS, NRAS, PIK3CA (p110-alpha).**

Villous Adenoma: **BRAF, KRAS.**

Other Adenoma-Related Genes: **APC, DMD, SMAD4 (MADH4), STK11 (LKB1), TCF7L2.**

Carcinoid-Endocrine Tumor

APC, CTNNB1, TP53 (p53).

Carcinoma

Adenocarcinoma: **ACVR1B, AKT1, APC, ATM, ATP6V0D2, AXIN2, BAX, BLM, BMPR1A (ALK3), BRAF, BRCA1, BRCA2, BUB1B, CASP8 (FLICE), CDC27, CDH1 (E-Cadherin), CDK4, CDKN2A (p16INK4a), CHEK2 (RAD53), CTNNA1, CTNNB1, DCC, DMD, EGFR (ERBB1), ENG (EVI-1), EP300, EPCAM, ERBB2 (HER-2, NEU), FBXW7, FGFR3, FLCN, FZD3, GALNT12, GPC6, GREM1, KIT (CD117), KRAS, MAP2K4 (MKK4, JNK1), MAP7, MET, MIER3, MLH1, MLH3, MSH2, MSH3, MSH6, MUTYH, MYO1B, NRAS, PALB2, PIK3CA (p110-alpha), PIK3R1 (p85-ALPHA), PMS1, PMS2, POLD1, POLE, PTEN, PTPN12, RET, RPS20, SLC9A9, SMAD2 (MADH2), SMAD4 (MADH4), SRC, STK11 (LKB1), TCERG1, TCF7L2, TGFBTR2, TP53 (p53), WBSCR17.**

Neuroendocrine: **BRAF, KRAS.**

Serrated Carcinoma: **BRAF, PIK3CA (p110-alpha).**

Squamous Cell Carcinoma: **KRAS, TP53 (p53).**

Other Carcinomas: **BRCA2, CDKN2A (p16INK4a).**

Training Data

NATIONAL CANCER INSTITUTE
GDC Data Portal

Home Projects Exploration Analysis Repository Quick Search Manage Sets Login Cart 0 GDC Apps

Cases Genes Mutations Add a Case Filter

Case input set e.g. TCGA-A5-A0G2, 432fe4a9-2... Upload Case Set

Primary Site Corpus uteri (538) Uterus, NOS (182)

Program TCGA (538) CPTAC (101) FM (81)

Project TCGA-UCEC (538)

Clear Case IN input set

View Files in Repository

Cases (720) Genes (22,160) Mutations (913,229) OncoGrid

Primary Site Project Disease Type Gender Vital Status

Showing 1 - 20 of 720 cases

Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category							# Mutations	# Genes	Slides
					Seq	Exp	SNV	CNV	Meth	Clinical	Bio			
TCGA-A5-A0G2	TCGA-UCEC	Corpus uteri	Female	58 4 5 16 5 1 10 17	42,051	14,357	(3)							
TCGA-EO-A22U	TCGA-UCEC	Corpus uteri	Female	57 4 5 16 5 1 10 16	26,998	12,629	(2)							
TCGA-FI-A2D5	TCGA-UCEC	Corpus uteri	Female	58 4 5 16 5 1 11 16	26,139	12,482	(2)							

https://portal.gdc.cancer.gov/exploration?filters=%7B%22content%22%3A%5B%7B%22content%22%3A%7B%22field%22%3A%22cases.case_id%22%2C%22value%22%3A%5B%22set_id%3AAWxYARM-1FJwHPndQLUN%22%5D%7D%2C%22op%22%3A%22IN%22%7D%5D%2C%22op%22%3A%22AND%22%7D

```
97     }
98   ],
99   "genomic_dna_change": "chr12:g.71567009_71567010insTATATAAGTATTACTGGTAGAAGATTCTAGACAGGGT",
100  "ssm_id": "1f850024-4e34-512b-880d-734e133bd006",
101  "mutation_subtype": "Small insertion"
102 }, {
103   "consequence": [
104     {
105       "transcript": {
106         "is_canonical": false,
107         "gene": {
108           "symbol": "PDE12",
109           "gene_id": "ENSG00000174840"
110         },
111         "aa_change": null,
112         "annotation": {
113           "vep_impact": "MODIFIER",
114           "polyphen_impact": "",
115           "sift_impact": ""
116         },
117         "consequence_type": "3_prime_UTR_variant"
118       }
119     },
120   ],
121   {
122     "transcript": {
123       "is_canonical": true,
124       "gene": {
125         "symbol": "PDE12",
126         "gene_id": "ENSG00000174840"
127       },
128       "aa_change": null,
129       "annotation": {
130         "vep_impact": "MODIFIER",
131         "polyphen_impact": "",
132         "sift_impact": ""
133       },
134       "consequence_type": "3_prime_UTR_variant"
135     }
136   }
137 ],
138  "genomic_dna_change": "chr3:g.57560088A>C",
139  "ssm_id": "c088a181-5490-5028-8d0f-01912288c38b",
140  "mutation_subtype": "Single base substitution"
141 }, {
142   "consequence": [
143     {
```

```
97     }
98   },
99 ],
100 "genomic_dna_change": "chr12:g.71567009_71567010insTATATAAGTATTACTGGTAGAAGATTCTAGACAGGGT",
101 "ssm_id": "1f850024-4e34-512b-880d-734e133bd006",
102 "mutation_subtype": "Small insertion"
103 },{
104   "consequence": [
105     {
106       "transcript": {
107         "is_canonical": false,
108         "gene": {
109           "symbol": "PDE12",
110           "gene_id": "ENSG00000174840"
111         },
112         "aa_change": null,
113         "annotation": {
114           "vep_impact": "MODIFIER",
115           "polyphen_impact": "",
116           "sift_impact": ""
117         },
118         "consequence_type": "3_prime_UTR_variant"
119       }
120     },
121     {
122       "transcript": {
123         "is_canonical": true,
124         "gene": {
125           "symbol": "PDE12",
126           "gene_id": "ENSG00000174840"
127         },
128         "aa_change": null,
129         "annotation": {
130           "vep_impact": "MODIFIER",
131           "polyphen_impact": "",
132           "sift_impact": ""
133         },
134         "consequence_type": "3_prime_UTR_variant"
135       }
136     }
137   ],
138   "genomic_dna_change": "chr3:g.57560088A>C",
139   "ssm_id": "c088a181-5490-5028-8d0f-01912288c38b",
140   "mutation_subtype": "Single base substitution"
141 },{
142   "consequence": [
143     {
```

```
97     }
98   },
99 ],
100 "genomic_dna_change": "chr12:g.71567009_71567010insTATATAAGTATTACTGGTAGAAGATTCTAGACAGGGT",
101 "ssm_id": "1f850024-4e34-512b-880d-734e133bd006",
102 "mutation_subtype": "Small insertion"
103 },{
104   "consequence": [
105     {
106       "transcript": {
107         "is_canonical": false,
108         "gene": {
109           "symbol": "PDE12",
110           "gene_id": "ENSG00000174840"
111         },
112         "aa_change": null,
113         "annotation": {
114           "vep_impact": "MODIFIER",
115           "polyphen_impact": "",
116           "sift_impact": ""
117         },
118         "consequence_type": "3_prime_UTR_variant"
119       }
120     },
121     {
122       "transcript": {
123         "is_canonical": true,
124         "gene": {
125           "symbol": "PDE12",
126           "gene_id": "ENSG00000174840"
127         },
128         "aa_change": null,
129         "annotation": {
130           "vep_impact": "MODIFIER",
131           "polyphen_impact": "",
132           "sift_impact": ""
133         },
134         "consequence_type": "3_prime_UTR_variant"
135       }
136     }
137   ],
138   "genomic_dna_change": "chr3:g.57560088A>C",
139   "ssm_id": "c088a181-5490-5028-8d0f-01912288c38b",
140   "mutation_subtype": "Single base substitution"
141 },{
142   "consequence": [
143     {
```

```
97     }
98   },
99 ],
100 "genomic_dna_change": "chr12:g.71567009_71567010insTATATAAGTATTACTGGTAGAAGATTCTAGACAGGGT",
101 "ssm_id": "1f850024-4e34-512b-880d-734e133bd006",
102 "mutation_subtype": "Small insertion"
103 },{
104   "consequence": [
105     {
106       "transcript": {
107         "is_canonical": false,
108         "gene": {
109           "symbol": "PDE12",
110           "gene_id": "ENSG00000174840"
111         },
112         "aa_change": null,
113         "annotation": {
114           "vep_impact": "MODIFIER",
115           "polyphen_impact": "",
116           "sift_impact": ""
117         },
118         "consequence_type": "3_prime_UTR_variant"
119       }
120     },
121     {
122       "transcript": {
123         "is_canonical": true,
124         "gene": {
125           "symbol": "PDE12",
126           "gene_id": "ENSG00000174840"
127         },
128         "aa_change": null,
129         "annotation": {
130           "vep_impact": "MODIFIER",
131           "polyphen_impact": "",
132           "sift_impact": ""
133         },
134         "consequence_type": "3_prime_UTR_variant"
135       }
136     }
137   ],
138   "genomic_dna_change": "chr3:g.57560088A>C",
139   "ssm_id": "c088a181-5490-5028-8d0f-01912288c38b",
140   "mutation_subtype": "Single base substitution"
141 },{
142   "consequence": [
143     {
```

Feature engineering: genes and VEPs

genes

case_id	type	diagnosis	ACVR1B	AKT1	ALK3	APC	ATM	ATP6V0D2
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	1	0	0	0	1	1
119b1761-a	1	Endometrioi	0	1	0	0	1	1
435b57e7-5	1	Endometrioi	1	1	0	1	1	1

VEP 1

case_id	type	diagnosis	('ACVR1B', 'HIGH')	('ACVR1B', 'LOW')	('ACVR1B', 'MODERATE')	('ACVR1B', 'MODIFIER')	('AKT1', 'HIGH')	('AKT1', 'LOW')
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	0	0	1	1	0	0
119b1761-a	1	Endometrioi	0	0	0	0	0	0
435b57e7-5	1	Endometrioi	0	0	0	1	0	0

VEP 2

case_id	type	diagnosis	('ACVR1B', 'MODIFIER')	('ACVR1B', 'VEP_SCORE')	('AKT1', 'MODIFIER')	('AKT1', 'VEP_SCORE')	('ALK3', 'MODIFIER')	('ALK3', 'VEP_SCORE')
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	1	2	0	0	0	0
119b1761-a	1	Endometrioi	0	0	1	2	0	0
435b57e7-5	1	Endometrioi	1	0	1	2	0	0

VEP 3

case_id	type	diagnosis	ACVR1B	AKT1	ALK3	APC	ATM	ATP6V0D2
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	2	0	0	0	3	3
119b1761-a	1	Endometrioi	0	2	0	0	2	3
435b57e7-5	1	Endometrioi	0	2	0	2	2	2

Feature engineering: genes and VEPs

Genes: 0.867

case_id	type	diagnosis	ACVR1B	AKT1	ALK3	APC	ATM	ATP6V0D2
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	1	0	0	0	1	1
119b1761-a	1	Endometrioi	0	1	0	0	1	1
435b57e7-5	1	Endometrioi	1	1	0	0	1	1

VEP 1: 0.882

case_id	type	diagnosis	('ACVR1B', 'HIGH')	('ACVR1B', 'LOW')	('ACVR1B', 'MODERATE')	('ACVR1B', 'MODIFIER')	('AKT1', 'HIGH')	('AKT1', 'LOW')
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	0	0	1	1	0	0
119b1761-a	1	Endometrioi	0	0	0	0	0	0
435b57e7-5	1	Endometrioi	0	0	0	0	1	0

VEP 2: 0.881

case_id	type	diagnosis	('ACVR1B', 'MODIFIER')	('ACVR1B', 'VEP_SCORE')	('AKT1', 'MODIFIER')	('AKT1', 'VEP_SCORE')	('ALK3', 'MODIFIER')	('ALK3', 'VEP_SCORE')
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	1	2	0	0	0	0
119b1761-a	1	Endometrioi	0	0	1	2	0	0
435b57e7-5	1	Endometrioi	1	0	1	2	0	0

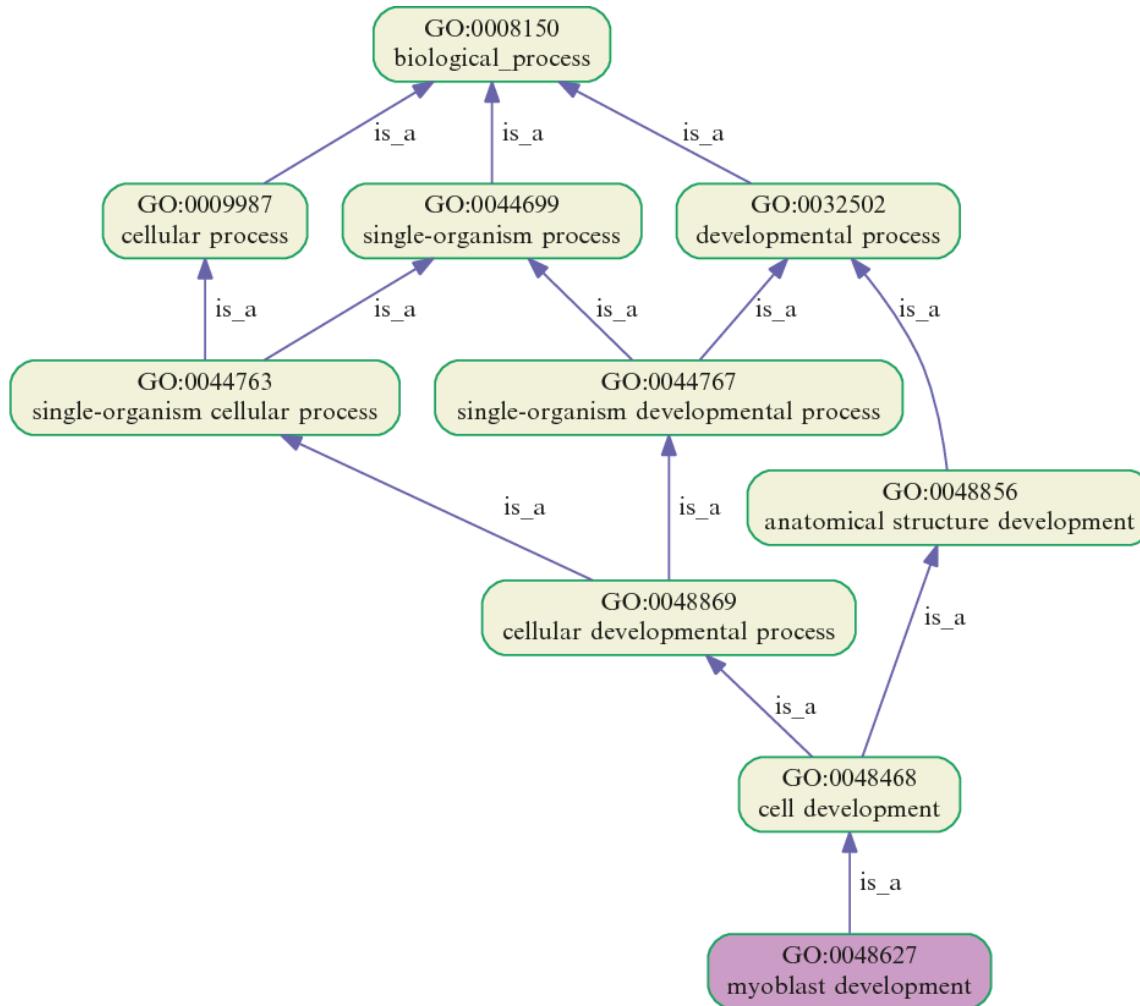
VEP 3: 0.885

case_id	type	diagnosis	ACVR1B	AKT1	ALK3	APC	ATM	ATP6V0D2
8a0a58a0-5	2	Serous cyst	0	0	0	0	0	0
4db38349-2	1	Endometrioi	2	0	0	0	3	3
119b1761-a	1	Endometrioi	0	2	0	0	2	3
435b57e7-5	1	Endometrioi	0	2	0	0	2	2

Feature engineering: Gene Ontology

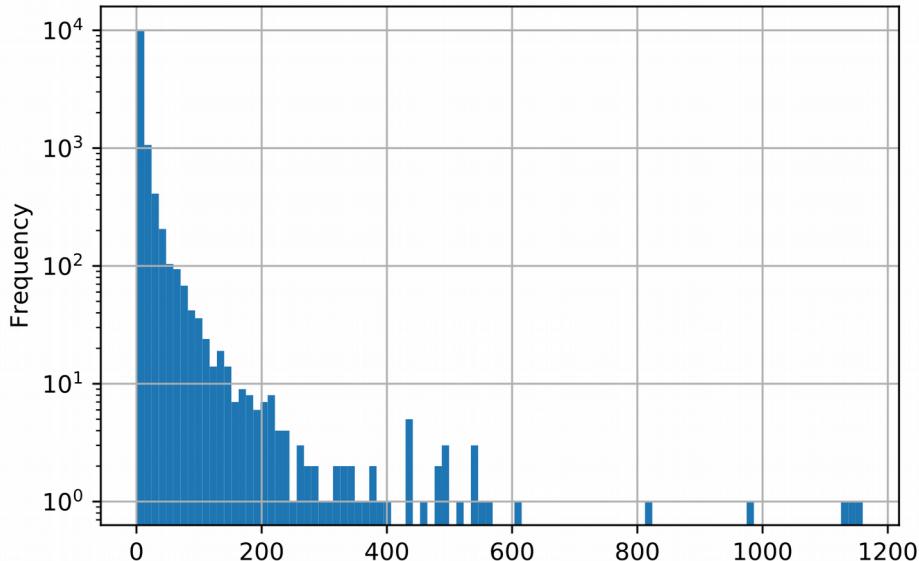
case_id	GO:0006355	GO:0006468	GO:0006508	GO:0006511	GO:0006886	GO:0006915	GO:0006954
8a0a58a0-5	0	2	4	0	0	0	0
4db38349-2	0	10	16	3	0	0	14
119b1761-a	0	4	14	2	0	0	11
435b57e7-5	0	6	17	3	3	0	20

Feature engineering: Gene Ontology

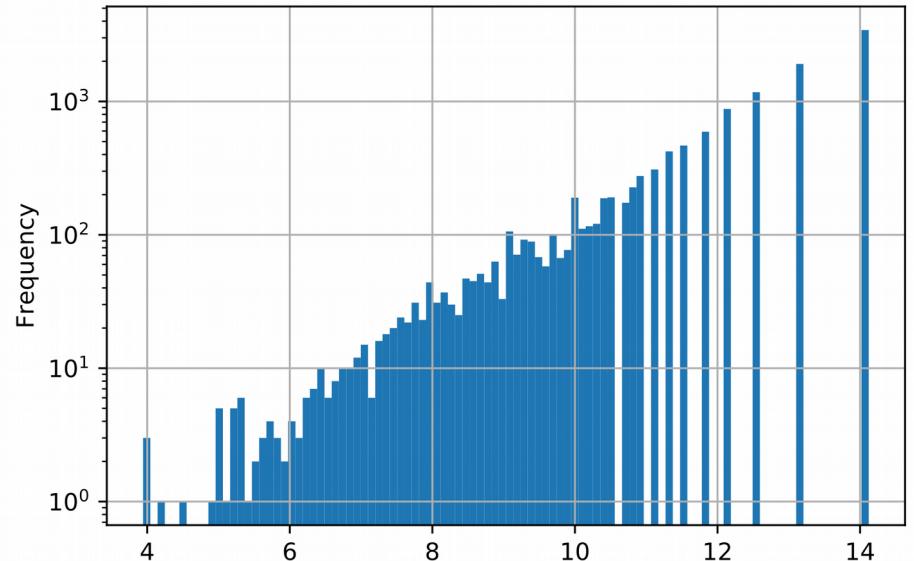


Feature engineering: Gene Ontology

nb. of genes / GO term



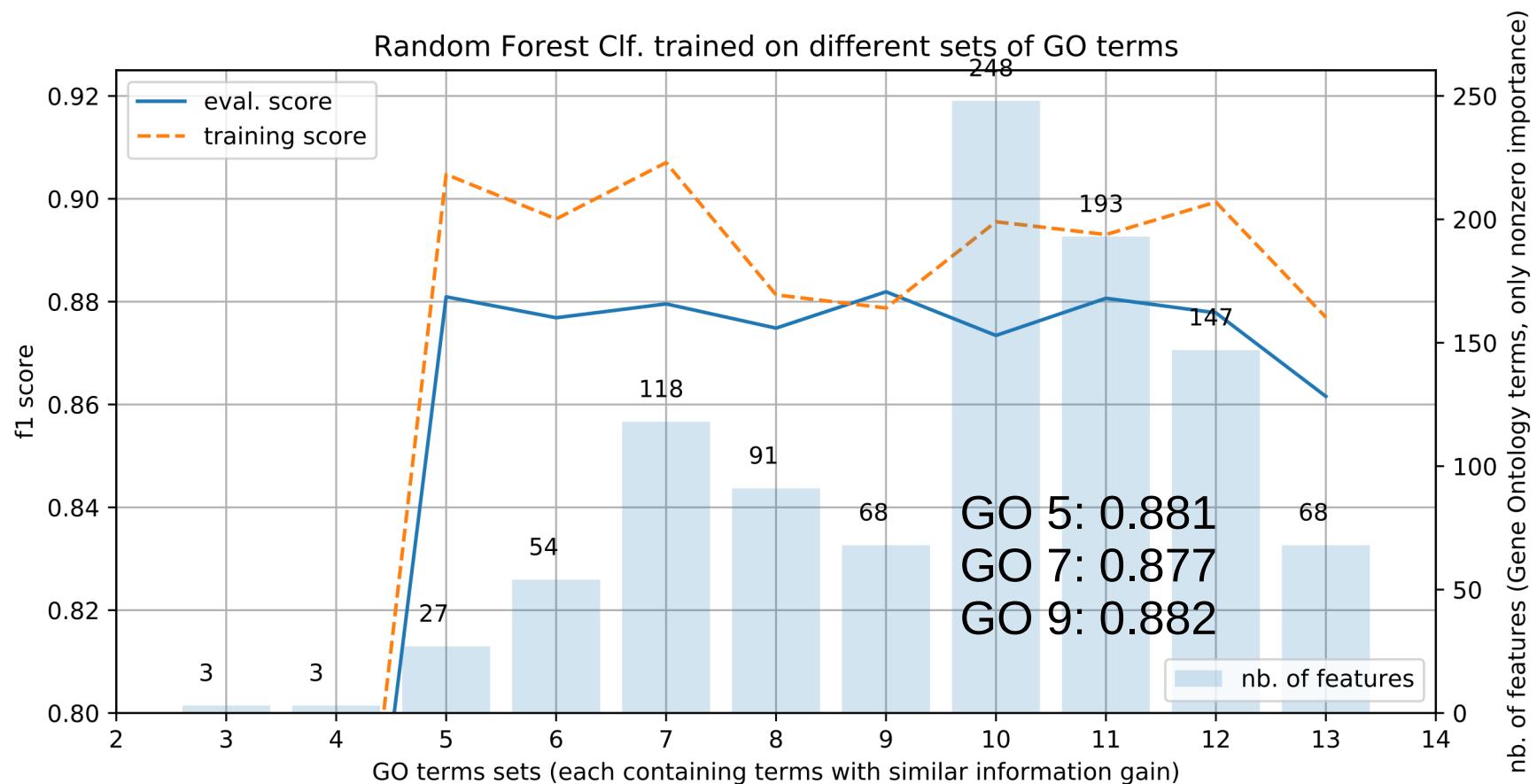
information gain



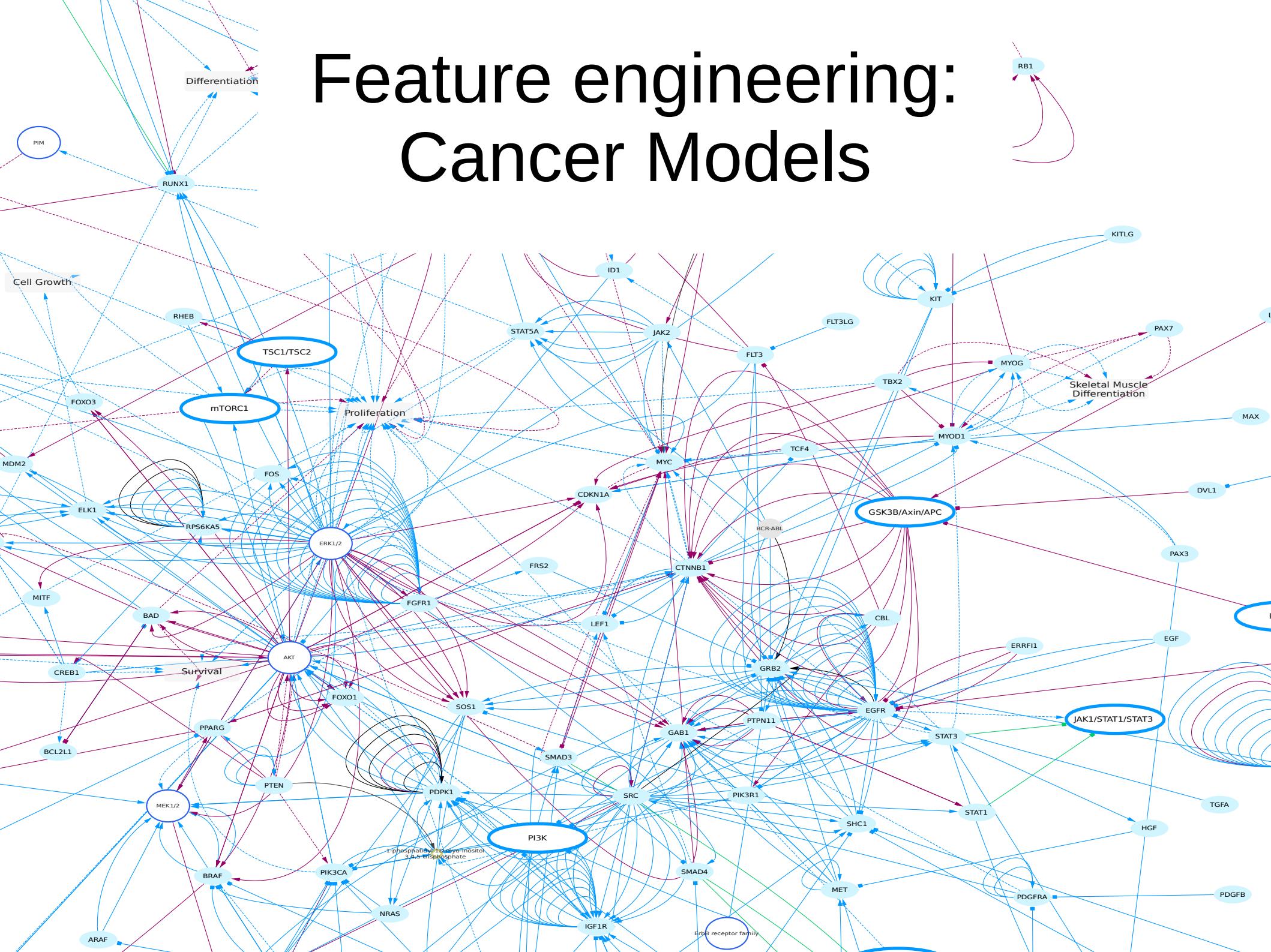
Alterovitz et al., “GO PaD: The Gene Ontology partition database”

Feature engineering: Gene Ontology

case_id	GO:0006355	GO:0006468	GO:0006508	GO:0006511	GO:0006886	GO:0006915	GO:0006954
8a0a58a0-5	0	2	4	0	0	0	0
4db38349-2	0	10	16	3	0	0	14
119b1761-a	0	4	14	2	0	0	11
435b57e7-5	0	6	17	3	3	0	20

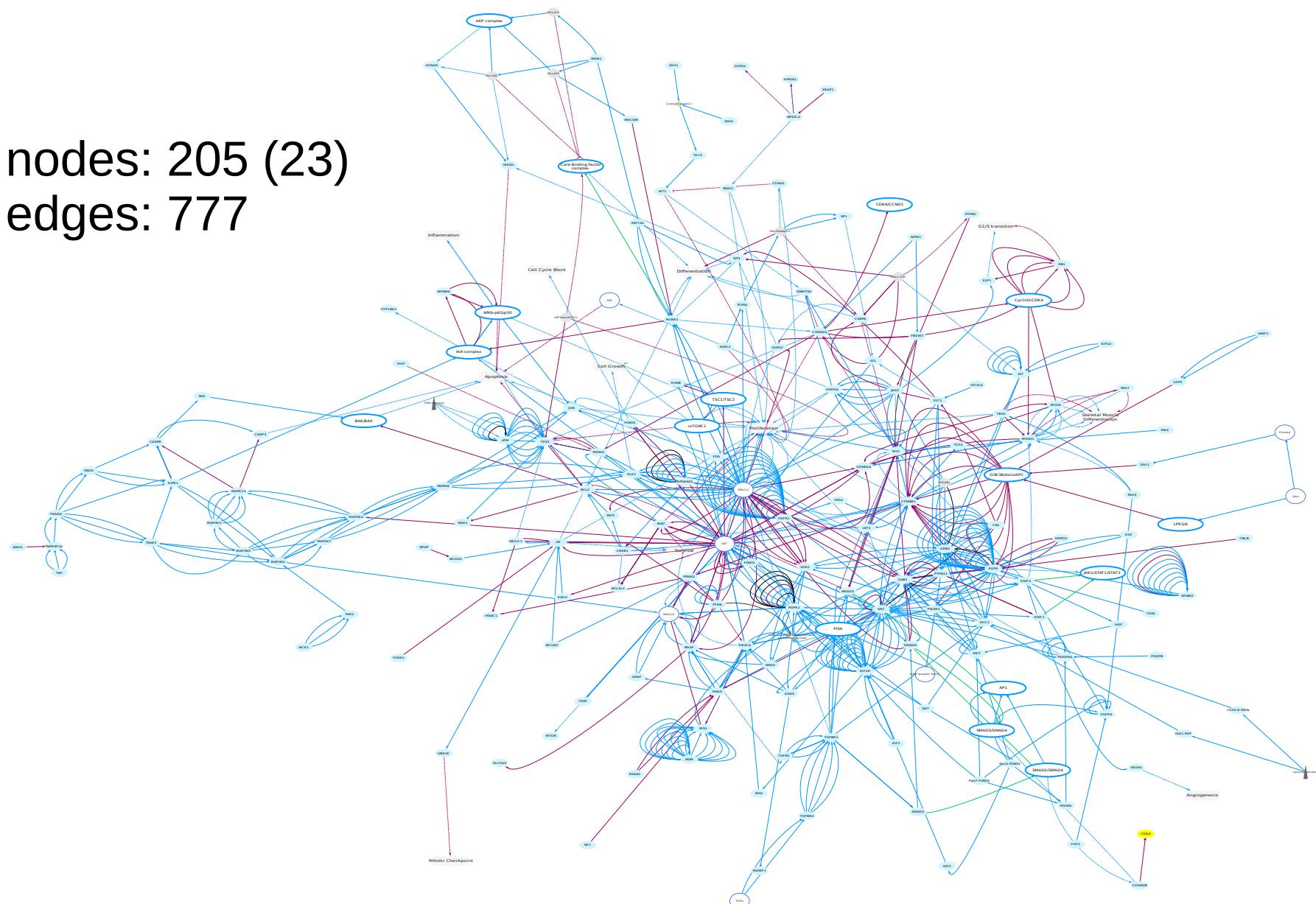


Feature engineering: Cancer Models



Feature engineering: Cancer Models

nodes: 205 (23)
edges: 777



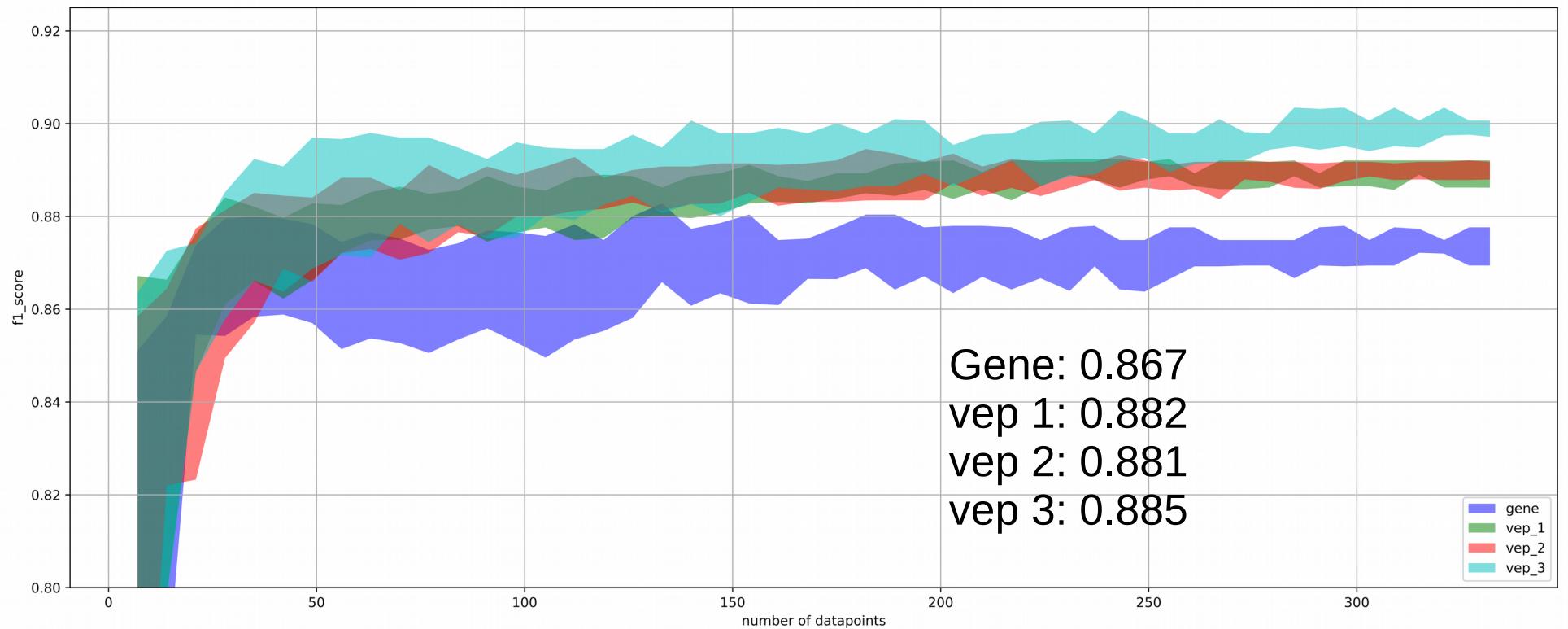
Feature engineering: Cancer Models

case_id	type	diagnosis	Core Binding Factor comp	1-phosphatidyl-1D-myo-ino	2-oxoglutarate(2-)	AEP complex	AKT	AP1
8a0a58a0-5	2	Serous cyst	0	2	0	0	2	0
4db38349-2	1	Endometrioid	0	5	0	0	4	0
119b1761-a	1	Endometrioid	0	7	0	0	5	0
435b57e7-5	1	Endometrioid	0	5	0	0	5	0

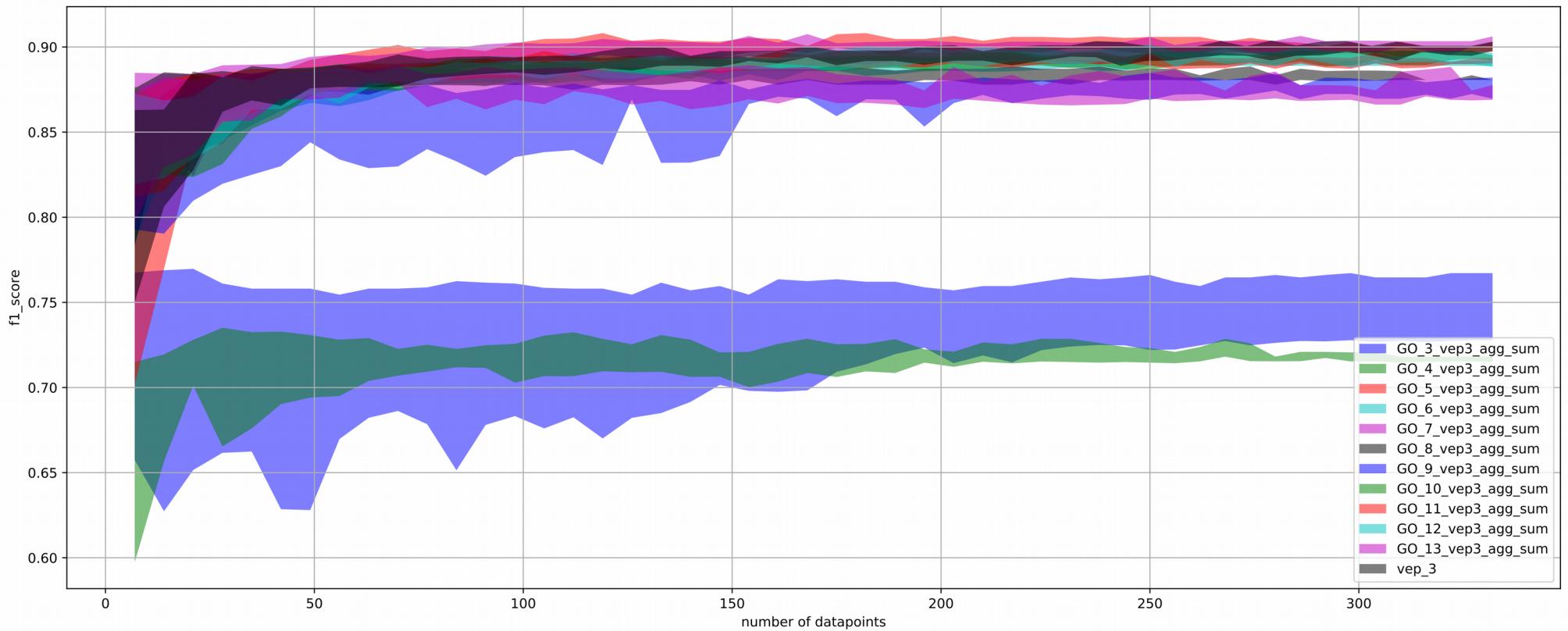
Summary

Gene	VEP 3	GO 5	GO 7	GO 9	model
0.867	0.885	0.881	0.877	0.882	0.896
89*	89*	33	209	68	168

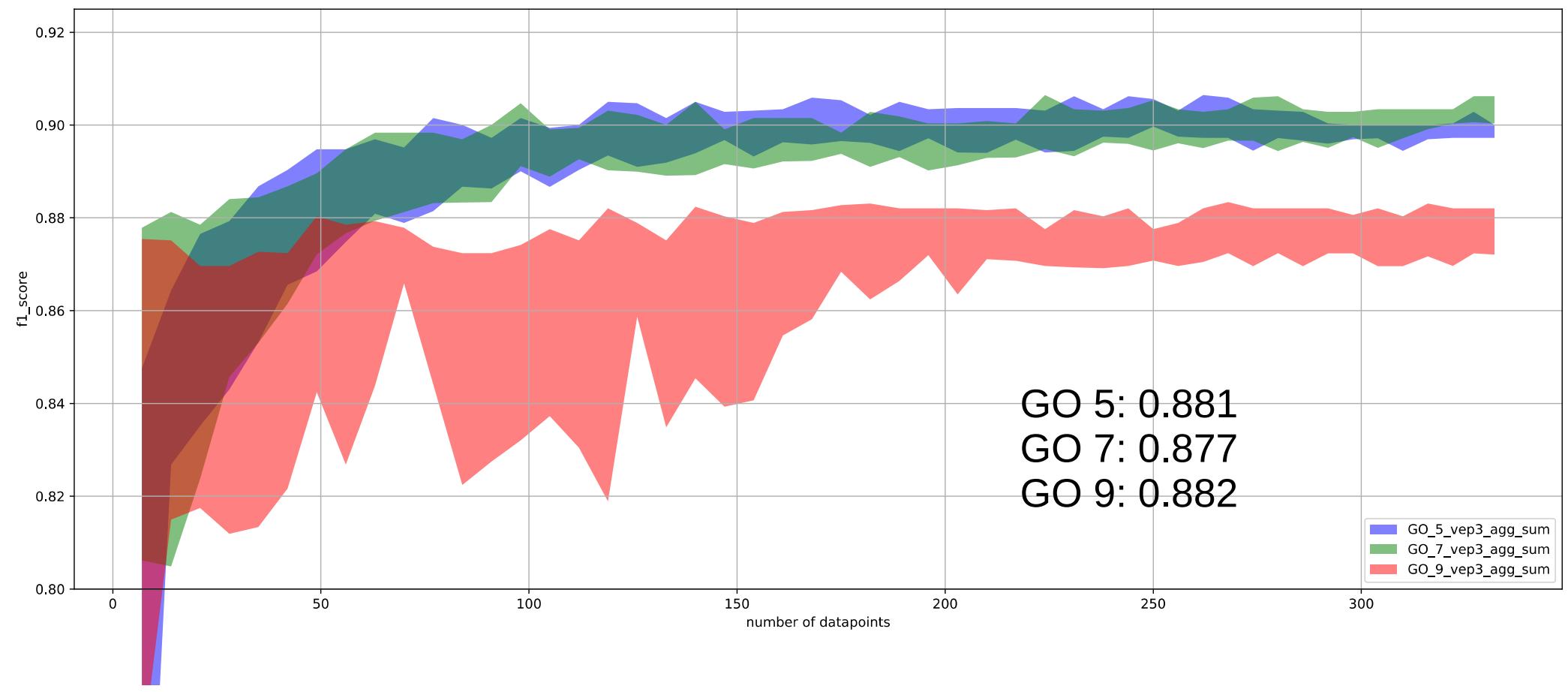
Feature engineering: genes and VEPs



Feature engineering: Gene Ontology

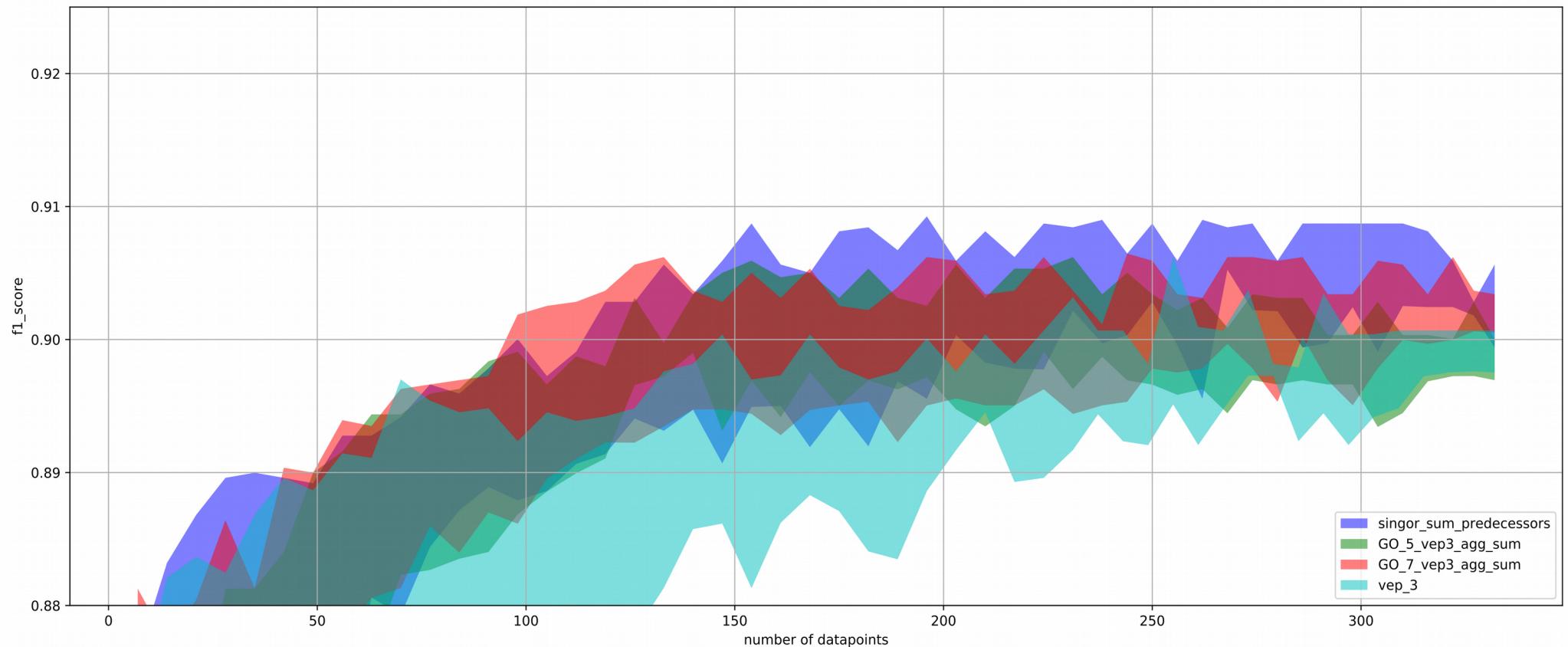


Feature engineering: Gene Ontology



Feature engineering: Cancer Models

case_id	type	diagnosis	Core Binding Factor comp	1-phosphatidyl-1D-myo-ino	2-oxoglutarate(2-)	AEP complex	AKT	AP1
8a0a58a0-5	2	Serous cyst	0	2	0	0	2	0
4db38349-2	1	Endometriotic	0	5	0	0	0	4
119b1761-a	1	Endometriotic	0	7	0	0	0	5
435b57e7-5	1	Endometriotic	0	5	0	0	0	5



Feature engineering: Cancer Models

