

Pan-cancer mutation analysis

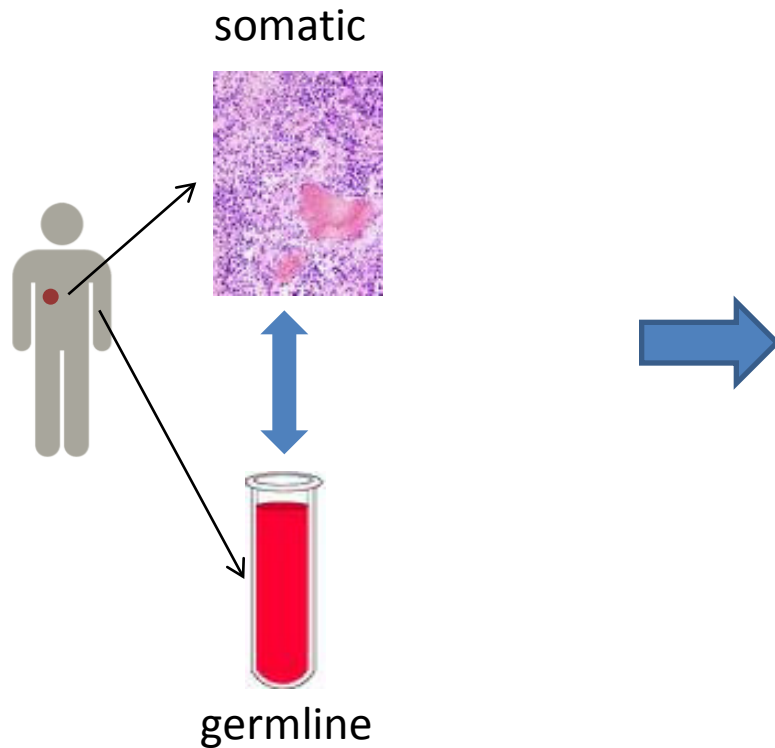
Mike Lawrence
Petar Stojanov
Broad Institute

November 27, 2012

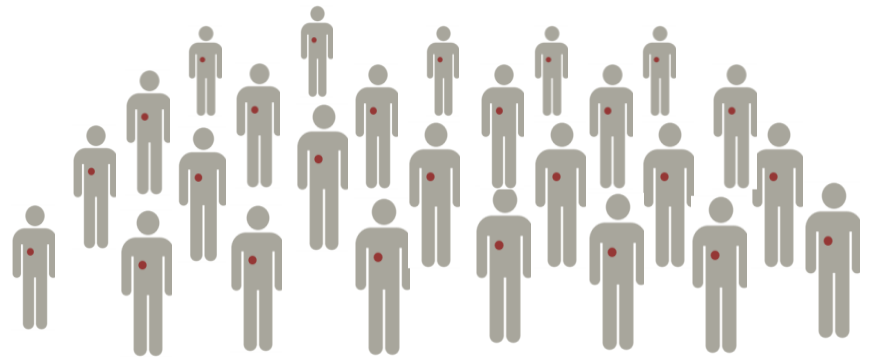
Cancer therapeutics: Discovery

Find cancer genes and pathways → cancer genome projects

Characterization (Individual)



Interpretation (Population)



What is the **full set of genome alterations** within the cancer (and germline)—mutations, copy number, translocations, epigenetic, etc?

- (1) Which genome alterations are **statistically significant** in the population?
- (2) In which **genes** and **pathways** do these alterations occur?

Cancer therapeutics: Discovery

Map of cancer genome projects

20-25 Tumor types

cancergenome.nih.gov

THE CANCER G



Ir
C
C



3 (April 2010)

S,

Neuroblastoma, Mesothelioma ~10s WES

Flood of data – Analysis pipeline

Picard pipeline
Sequencing platform

Tim Fennell

Align to genome (BWA)

Mark duplicate reads

Base quality recalibration

BAM file

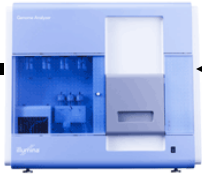
dbGaP

Visualization (IGV)

*Jim Robinson
Helga Thorvaldsdottir*

Purity ploidy

Scott Carter



Cancer Genome Analysis



*Doug Voet
Mike Noble
Dan DiCara*



dbGaP

*Carrie Sougnez
Rui Jing*

QC

Mutations

Indels

Purity/
ploidy

Copy number

Rearrangements

Pathogens ...

*Kristian
Cibulskis*

*Andrey
Sivachenko*

*Scott
Carter*

*Gordon Saksena
Marc-Danie Nazaire*

*Mike Lawrence
Yotam Drier*

Alex Kostic

Characterization

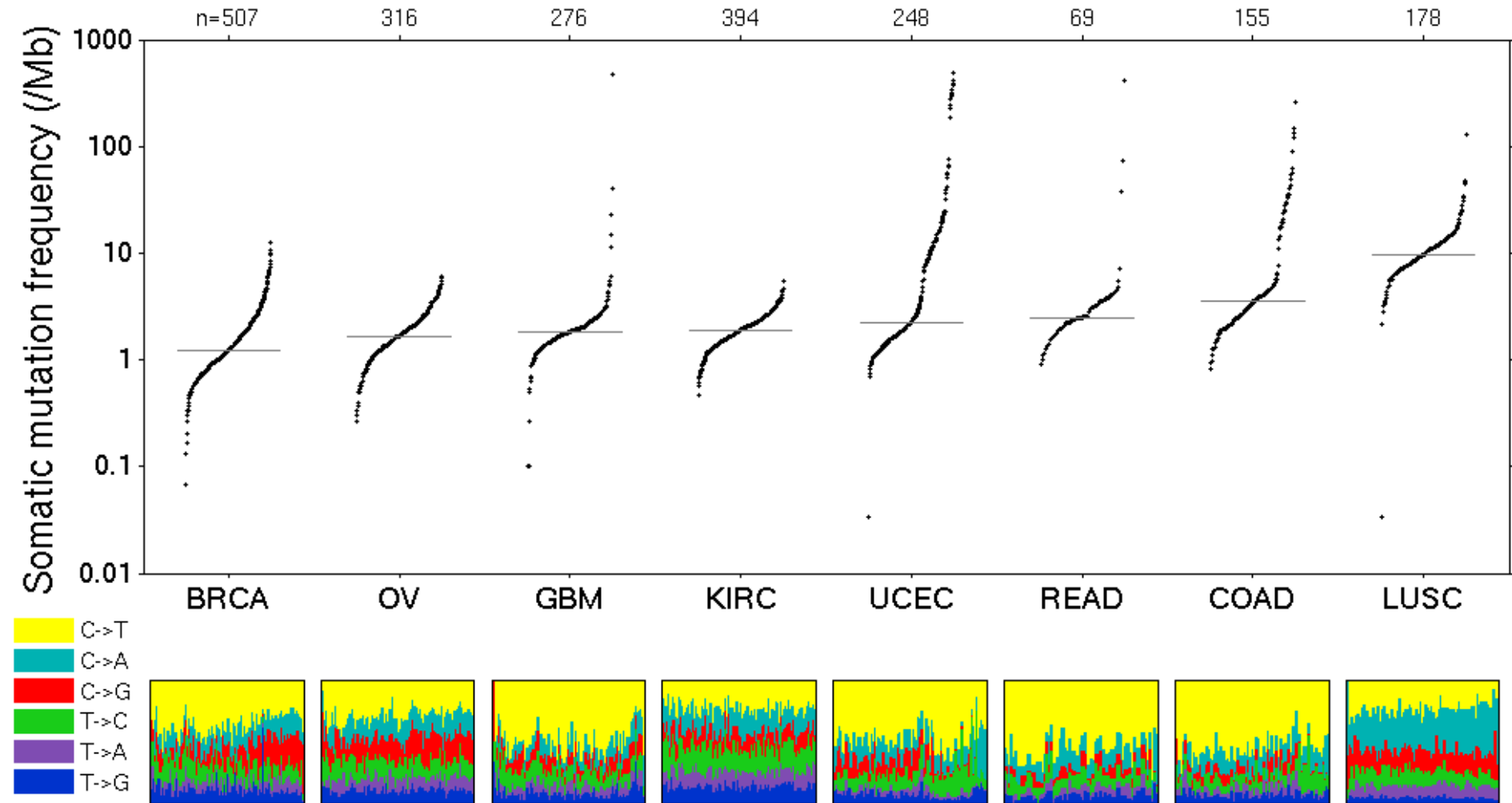
Annotation (*Oncotator*) + Reports

*Alex
Ramos*

PanCan8 data set

tumor_type		#patients		#mutations (coding)	
BRCA	breast	507	24%	28888	7%
COAD	colon	155	7%	59494	14%
GBM	glioblastoma	276	13%	34358	8%
KIRC	kidney	394	18%	26409	6%
LUSC	lung squamous	178	8%	64296	15%
OV	ovarian	316	15%	19156	4%
READ	rectal	69	3%	21591	5%
UCEC	endometrial	248	12%	182563	42%
TOTAL		2143	100%	436755	100%

PanCan8 data set



Mutation Significance: MutSig

Takes into account:

1. Sample-specific, gene-specific, and context-specific background mutation rates
2. Base-level evolutionary conservation
3. Positional configuration
4. Truncating mutations

Number of significant genes ($q < 0.1$)

BRCA	22
COAD	14
GBM	12
KIRC	12
LUSC	16
OV	3
READ	3
UCEC	33
PanCan8	150

GBM glioblastoma

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
TP53	tumor protein p53	259875	89	80	63	2.55E-15	3.03E-07	3.03E-07	7.66E-15	9.73E-11
PTEN	phosphatase and tensin homolog (mutated)	267850	80	79	65	5.11E-15	9.60E-03	0.36	1.53E-14	9.73E-11
EGFR	epidermal growth factor receptor (erythr	847000	58	54	26	5.66E-15	2.56E-07	2.56E-07	1.70E-14	9.73E-11
RB1	retinoblastoma 1 (including osteosarcoma	783200	21	21	19	7.11E-15	0.049	0.24	2.13E-14	9.73E-11
PIK3R1	phosphoinositide-3-kinase, regulatory su	512600	17	17	14	1.99E-11	1.33E-06	0.0011	5.96E-11	2.18E-07
NF1	neurofibromin 1 (neurofibromatosis, von	2588300	31	25	30	2.17E-10	4.30E-01	6.70E-01	6.51E-10	1.98E-06
IDH1	isocitrate dehydrogenase 1 (NADP+), solu	276925	15	15	2	1.46E-07	2.5E-07	0.021	4.37E-07	1.00E-03
PIK3CA	phosphoinositide-3-kinase, catalytic, al	714725	19	19	17	1.51E-07	0.015	0.0061	4.53E-07	0.001
ZNF74	zinc finger protein 74	389400	3	1	3	9.90E-01	2.49E-07	0.023	7.48E-07	0.0015
ABCA3	ATP-binding cassette, sub-family A (ABC1	971575	3	1	3	9.90E-01	2.95E-07	0.00063	8.85E-07	0.0016
ITGA7	integrin, alpha 7	770825	3	3	3	8.20E-01	0.0027	2.44E-06	7.32E-06	0.012
TAPBPL	TAP binding protein-like	298650	2	2	2	3.90E-01	1.05E-05	0.017	3.15E-05	0.048
APRT	adenine phosphoribosyltransferase	69850	3	1	3	4.20E-01	2.97E-05	0.27	8.92E-05	0.13
CDKN2C	cyclin-dependent kinase inhibitor 2C (p1	107525	4	4	4	5.79E-05	1.40E-01	9.50E-01	1.70E-04	0.21
SHANK1	SH3 and multiple ankyrin repeat domains	620675	2	2	2	1.00E+00	5.81E-05	0.012	1.70E-04	0.21
C19orf10	chromosome 19 open reading frame 10	76725	2	2	2	2.30E-01	0.014	6.13E-05	1.80E-04	0.21
GABRA6	gamma-aminobutyric acid (GABA) A recepto	297825	9	9	8	6.54E-05	0.08	0.24	2.00E-04	0.21
ACSM2B	acyl-CoA synthetase medium-chain family	376750	2	2	2	1.00E+00	0.00012	0.39	3.60E-04	0.37
MTMR7	myotubularin related protein 7	436425	4	4	2	2.50E-01	0.0002	0.42	6.00E-04	0.53
QKI	quaking homolog, KH domain RNA binding (244200	6	5	6	0.00021	0.77	1.20E-02	6.30E-04	0.53
DHX29	DEAH (Asp-Glu-Ala-His) box polypeptide 2	931425	2	2	2	0.83	0.00021	6.90E-01	0.00064	0.53
STAG2	stromal antigen 2	859375	11	10	10	2.30E-04	0.23	0.45	0.00068	0.53
ACSL6	acyl-CoA synthetase long-chain family me	485100	2	2	1	3.20E-01	0.00023	0.9	0.0007	0.53
PRF1	perforin 1 (pore forming protein)	334400	2	1	2	0.79	0.0003	0.7	0.0009	0.63
SCMH1	sex comb on midleg homolog 1 (Drosophila	429275	2	2	1	1	0.00031	0.0017	0.00094	0.63
SVOP	SVOP-like	224125	5	4	5	0.2	0.00032	0.045	0.00096	0.63
PSG11	pregnancy specific beta-1-glycoprotein 1	215875	2	2	2	0.56	0.39	0.00034	0.001	0.64
LTK	leukocyte receptor tyrosine kinase	462550	2	2	2	0.65	0.0004	0.66	0.0012	0.73
C19orf26	chromosome 19 open reading frame 26	169675	2	2	1	0.052	0.0005	0.92	0.0015	0.88

OV ovarian

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
TP53	tumor protein p53	298620	279	276	143	1.00E-15	4.74E-07	4.74E-07	3.00E-15	5.48E-11
SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral	324848	4	4	2	1.20E-01	2.92E-07	0.00055	8.75E-07	0.008
RB1	retinoblastoma 1 (including osteosarcoma	899968	9	9	9	5.95E-06	0.18	0.51	1.78E-05	0.11
NF1	neurofibromin 1 (neurofibromatosis, von	2974192	15	14	15	2.36E-05	0.28	0.059	7.08E-05	0.3
THADA	thyroid adenoma associated	1300972	2	1	2	1	0.0000273	0.22	8.18E-05	0.3
BRCA1	breast cancer 1, early onset	1441592	10	10	10	6.37E-05	1	3.50E-01	1.90E-04	0.55
MYO19	myosin XIX	689196	2	1	2	9.40E-01	0.0000704	0.059	2.10E-04	0.55
RNASEN	ribonuclease type III, nuclear	982444	2	2	2	9.10E-01	0.0000836	0.00044	2.50E-04	0.57
POTED	POTE ankyrin domain family, member D	101752	4	4	4	1.10E-04	0.27	0.032	3.30E-04	0.66
FAP	fibroblast activation protein, alpha	591236	4	4	4	1.20E-04	0.71	0.15	3.70E-04	0.68
ARID1B	AT rich interactive domain 1B (SWI1-like	1287384	2	1	2	9.90E-01	0.00017	0.83	5.20E-04	0.87
RYK	RYK receptor-like tyrosine kinase	328956	2	2	2	6.50E-01	0.00021	0.28	6.40E-04	0.98
BRCA2	breast cancer 2, early onset	2558968	9	9	9	1.40E-01	0.27	0.82	4.10E-01	1
GPRIN3	GPRIN family member 3	561532	5	5	5	2.60E-04	0.21	0.52	7.80E-04	1
GPX6	glutathione peroxidase 6 (olfactory)	169692	3	3	3	0.00036	9.20E-01	8.00E-01	1.10E-03	1
FAM200A	family with sequence similarity 200, mem	47716	2	2	2	3.70E-04	3.50E-01	0.8	1.10E-03	1
MAP3K15	mitogen-activated protein kinase kinase	815596	3	3	3	7.50E-01	0.00047	0.3	1.40E-03	1
DUSP19	dual specificity phosphatase 19	164952	4	4	4	4.70E-04	0.64	0.54	1.40E-03	1
PAK3	p21 (CDKN1A)-activated kinase 3	417120	5	5	5	4.90E-04	0.96	0.88	1.50E-03	1
C1orf95	chromosome 1 open reading frame 95	75208	3	3	3	4.90E-04	0.27	0.47	1.50E-03	1
CTNND2	catenin (cadherin-associated protein), d	822232	4	4	4	0.96	0.5	5.30E-04	1.60E-03	1
SYT5	synaptotagmin V	210456	2	1	2	0.41	0.00053	9.20E-01	0.0016	1
C5orf33	chromosome 5 open reading frame 33	253748	2	1	2	3.10E-01	0.00063	0.6	0.0019	1
GTF3C3	general transcription factor IIIC, polyp	679084	2	2	2	0.99	6.30E-04	0.98	0.0019	1
USP22	ubiquitin specific peptidase 22	338752	2	1	2	0.5	7.50E-04	0.4	0.0023	1
CHMP4A	chromatin modifying protein 4A	285348	3	3	3	0.00088	0.93	0.71	0.0026	1
LCOR	ligand dependent nuclear receptor corepr	321372	4	4	4	0.00091	0.47	0.0076	0.0027	1
BDKRB1	bradykinin receptor B1	253748	4	4	4	0.00094	0.49	0.93	0.0028	1
MAP7D1	MAP7 domain containing 1	351708	2	2	2	0.72	0.001	0.41	0.003	1

COAD colon

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
APC	adenomatous polyposis coli	1034160	155	103	100	1.00E-15	3.75E-07	2.90E-02	3.00E-15	5.48E-11
TP53	tumor protein p53	146475	77	75	50	3.00E-15	3.16E-07	3.16E-07	8.99E-15	8.21E-11
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral onco	93775	57	57	8	1.81E-14	2.84E-07	1.39E-05	5.43E-14	3.30E-10
NRAS	neuroblastoma RAS viral (v-ras) oncogene	71455	15	15	7	1.33E-11	0.001	0.009	3.98E-11	1.82E-07
FBXW7	F-box and WD repeat domain containing 7	301010	32	28	20	7.75E-11	2.76E-05	0.048	2.33E-10	8.49E-07
SMAD4	SMAD family member 4	205220	22	18	18	1.04E-07	8.30E-02	9.50E-04	3.11E-07	9.50E-04
FAM123B	family with sequence similarity 123B	383780	19	19	17	1.71E-07	0.019	0.41	5.14E-07	1.30E-03
BRAF	v-raf murine sarcoma viral oncogene homo	267685	20	19	2	5.60E-04	2.72E-07	2.72E-07	8.17E-07	1.70E-03
PCBP1	poly(rC) binding protein 1	124930	4	4	2	2.70E-01	2.82E-07	5.62E-05	8.47E-07	1.70E-03
PIK3CA	phosphoinositide-3-kinase, catalytic, al	402845	33	26	18	2.64E-05	2.55E-05	3.12E-07	9.35E-07	1.70E-03
EYS	eyes shut homolog (Drosophila)	218240	9	7	9	9.00E-01	4.81E-07	0.046	1.44E-06	2.30E-03
FSIP2	fibrous sheath interacting protein 2	588535	6	4	6	1.00E+00	4.98E-07	0.0039	1.49E-06	2.30E-03
SOX9	SRY (sex determining region Y)-box 9 (ca	146010	9	9	9	1.54E-06	0.11	0.43	4.63E-06	0.0065
TCERG1	transcription elongation regulator 1	402845	11	10	11	0.0000124	6.30E-01	7.10E-01	3.73E-05	0.049
TCERG1L	transcription elongation regulator 1-lik	106795	6	4	6	8.10E-01	8.60E-02	0.00005	1.50E-04	0.18
ACTN4	actinin, alpha 4	311705	2	2	2	7.60E-01	0.077	5.79E-05	1.70E-04	0.19
SULT1C4	sulfotransferase family, cytosolic, 1C,	116405	3	3	1	3.60E-01	5.83E-05	0.13	1.70E-04	0.19
NAIP	NLR family, apoptosis inhibitory protein	82460	2	2	2	3.40E-01	6.67E-05	0.52	2.00E-04	0.2
FAT4	FAT tumor suppressor homolog 4 (Drosophi	1780020	56	29	56	1.00E+00	0.000075	0.3	2.20E-04	0.21
RBL1	retinoblastoma-like 1 (p107)	388585	3	3	3	0.98	0.24	7.78E-05	2.30E-04	0.21
HPRT1	hypoxanthine phosphoribosyltransferase 1	78585	2	2	2	0.39	0.023	9.65E-05	0.00029	0.25
C15orf60	chromosome 15 open reading frame 60	73625	4	4	4	1.80E-04	0.69	0.83	0.00055	0.45
SFPQ	splicing factor proline/glutamine-rich (168950	7	6	6	0.38	1.60E-03	0.0002	0.0006	0.48
HTR3B	5-hydroxytryptamine (serotonin) receptor	163525	4	4	3	7.50E-01	0.14	0.00021	0.00064	0.49
ACVR1B	activin A receptor, type IB	176080	13	13	13	0.00024	9.70E-02	0.0019	0.00072	0.52
RP9	retinitis pigmentosa 9 (autosomal domina	67115	3	3	2	0.00027	0.2	0.1	0.00081	0.57
SMAD2	SMAD family member 2	173135	11	10	8	0.00029	0.025	0.062	0.00086	0.58
TGFB2	transforming growth factor, beta recepto	210800	6	6	6	0.31	0.011	0.00032	0.00095	0.6
TEKT5	tektin 5	179025	4	4	3	0.96	0.00032	0.0037	0.00096	0.6

READ rectal

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral onco	41745	37	37	8	1.00E-15	2.96E-07	2.96E-07	3.00E-15	5.48E-11
TP53	tumor protein p53	65205	45	45	30	5.55E-15	2.97E-07	2.97E-07	1.67E-14	1.52E-10
APC	adenomatous polyposis coli	460368	84	57	68	1.13E-14	3.29E-07	0.031	3.40E-14	2.07E-10
ERBB2	v-erb-b2 erythroblastic leukemia viral o	211846	4	4	2	9.30E-01	2.33E-05	0.0085	7.00E-05	0.32
SMAD4	SMAD family member 4	91356	8	8	6	5.36E-05	0.15	0.16	1.60E-04	0.59
NRAS	neuroblastoma RAS viral (v-ras) oncogene	31809	5	5	4	1.50E-04	9.70E-03	2.10E-01	4.60E-04	1
FLG	filaggrin	633972	11	5	11	9.10E-01	0.43	0.00017	5.20E-04	1
RUNX2	runt-related transcription factor 2	87308	2	1	1	4.70E-01	0.00019	0.64	5.70E-04	1
TLL2	tolloid-like 2	167187	3	3	2	7.50E-01	0.00021	0.25	6.40E-04	1
FBXW7	F-box and WD repeat domain containing 7	133998	12	9	10	2.20E-04	0.072	0.018	6.50E-04	1
RBM10	RNA binding motif protein 10	112712	5	5	4	2.40E-04	0.71	0.78	7.10E-04	1
FAM123B	family with sequence similarity 123B	170844	8	6	8	3.60E-04	0.038	0.67	1.10E-03	1
TCF7L2	transcription factor 7-like 2 (T-cell sp	102465	7	7	7	4.10E-04	0.056	0.0077	1.20E-03	1
ELF3	E74-like factor 3 (ets domain transcript	64897	3	3	3	0.00044	4.10E-02	2.20E-01	1.30E-03	1
C11orf16	chromosome 11 open reading frame 16	81249	2	1	1	3.30E-01	4.70E-04	0.98	1.40E-03	1
ZNF395	zinc finger protein 395	80008	2	2	1	1.50E-01	0.0005	0.82	1.50E-03	1
CTLA4	cytotoxic T-lymphocyte-associated protei	39493	2	1	1	2.10E-01	0.0005	0.077	1.50E-03	1
SMC1B	structural maintenance of chromosomes 1B	214912	2	2	2	9.80E-01	0.00056	0.36	1.70E-03	1
FAM169A	family with sequence similarity 169, mem	121618	2	2	2	7.30E-01	0.00075	0.54	2.30E-03	1
KIAA1804		124890	11	9	9	0.00077	0.032	8.90E-03	2.30E-03	1
FCHSD2	FCH and double SH3 domains 2	98118	4	2	4	0.66	0.082	8.00E-04	0.0024	1
CAMKV	CaM kinase-like vesicle-associated	83658	2	1	2	8.40E-02	0.001	0.9	0.003	1
PEX11A	peroxisomal biogenesis factor 11A	38909	2	1	1	0.11	1.40E-03	0.62	0.0042	1
CDKN1B	cyclin-dependent kinase inhibitor 1B (p2	33945	2	2	2	1.40E-03	0.77	0.63	0.0043	1
SMAD2	SMAD family member 2	77073	5	5	5	0.0015	1.30E-02	0.015	0.0045	1
NCOA3	nuclear receptor coactivator 3	236532	5	1	5	0.98	0.15	0.0016	0.0048	1
MMRN1	multimerin 1	199272	5	2	5	0.87	0.9	0.0016	0.0048	1
GRXCR1	glutaredoxin, cysteine rich 1	50224	2	2	2	0.32	0.0028	0.0017	0.005	1
ZFP2	zinc finger protein 2 homolog (mouse)	76797	2	2	1	0.69	0.0018	0.43	0.0054	1

LUSC lung squamous

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
TP53	tumor protein p53	168210	147	141	98	3.44E-15	3.66E-07	3.66E-07	1.03E-14	1.88E-10
CDKN2A	cyclin-dependent kinase inhibitor 2A (me	107156	26	26	23	2.10E-14	6.22E-05	4.2E-07	6.29E-14	5.74E-10
PTEN	phosphatase and tensin homolog (mutated	173372	16	14	15	1.23E-09	0.065	0.31	3.70E-09	2.25E-05
KEAP1	kelch-like ECH-associated protein 1	231400	24	22	21	5.70E-09	0.085	0.19	1.71E-08	7.80E-05
SLC6A5	solute carrier family 6 (neurotransmitte	306516	7	6	7	1.00E+00	2.79E-07	3.63E-06	8.38E-07	0.0023
NFE2L2	nuclear factor (erythroid-derived 2)-lik	250802	28	27	15	3.91E-07	2.81E-07	2.81E-07	8.43E-07	0.0023
EYS	eyes shut homolog (Drosophila)	250624	14	13	14	1.60E-01	2.94E-07	0.028	8.82E-07	0.0023
IFNA10	interferon, alpha 10	79032	3	2	2	3.70E-01	5.95E-07	0.031	1.79E-06	0.0041
PIK3CA	phosphoinositide-3-kinase, catalytic, al	462622	29	27	16	1.40E-04	1.55E-06	0.00018	4.65E-06	0.0091
ZEB2	zinc finger E-box binding homeobox 2	517802	15	15	15	8.10E-01	1.96E-06	0.044	5.88E-06	0.0091
COL6A6	collagen, type VI, alpha 6	894272	20	19	20	1.00E+00	0.000002	0.014	6.00E-06	0.0091
LTBP4	latent transforming growth factor beta b	442330	6	4	6	9.70E-01	0.000002	0.041	6.00E-06	0.0091
ZNF567	zinc finger protein 567	264152	4	3	4	7.00E-01	2.27E-06	3.69E-06	6.82E-06	0.0096
NOTCH1	Notch homolog 1, translocation-associate	710754	16	14	16	0.066	1.70E-01	1.82E-05	5.45E-05	0.071
MLL2	myeloid/lymphoid or mixed-lineage leukem	1885198	41	35	41	2.17E-05	3.00E-01	0.26	6.50E-05	0.078
RB1	retinoblastoma 1 (including osteosarcoma	506944	12	12	12	2.27E-05	0.095	0.57	6.82E-05	0.078
MEF2D	myocyte enhancer factor 2D	187434	3	3	3	5.30E-01	0.0000351	0.0096	1.10E-04	0.11
TNXB	tenascin XB	1512644	19	14	19	9.90E-01	0.0000389	0.43	1.20E-04	0.12
CRLS1	cardiolipin synthase 1	88822	2	1	2	7.80E-01	0.0000501	0.0000531	1.50E-04	0.14
CSRNP3	cysteine-serine-rich nuclear protein 3	246886	6	6	6	0.6	0.02	5.71E-05	1.70E-04	0.16
UBAP2L	ubiquitin associated protein 2-like	477930	4	3	4	1	0.0000667	3.20E-01	0.0002	0.17
CILP2	cartilage intermediate layer protein 2	361162	6	5	6	1.00E+00	0.0000765	0.47	0.00023	0.19
SLC17A8	solute carrier family 17 (sodium-depende	252048	15	13	15	0.000087	6.20E-01	0.51	0.00026	0.21
ASCL4	achaete-scute complex homolog 4 (Drosoph	31328	6	6	6	1.10E-04	0.2	0.53	0.00032	0.24
FBXW7	F-box and WD repeat domain containing 7	345676	11	11	9	0.00013	3.70E-03	0.066	0.00039	0.29
RASGRF2	Ras protein-specific guanine nucleotide-	515666	7	7	7	0.85	0.62	0.00017	0.0005	0.34
SMURF1	SMAD specific E3 ubiquitin protein ligas	301532	2	2	2	0.96	0.00017	0.91	0.00052	0.34
HS6ST1	heparan sulfate 6-O-sulfotransferase 1	106622	5	5	5	0.0037	0.00021	0.005	0.00064	0.39
KCNC2	potassium voltage-gated channel, Shaw-re	235316	6	5	5	1	0.016	0.00023	0.0007	0.4

BRCA breast

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
TP53	tumor protein p53	479115	186	183	114	1.00E-15	4.08E-07	4.08E-07	3.00E-15	2.43E-11
GATA3	GATA binding protein 3	430443	58	54	35	1.00E-15	3.13E-07	0.0015	3.00E-15	2.43E-11
RUNX1	runt-related transcription factor 1 (acu	389883	18	18	17	1.33E-15	0.0002	0.26	4.00E-15	2.43E-11
CDH1	cadherin 1, type 1, E-cadherin (epitheli	990678	33	33	30	1.78E-15	0.21	0.33	5.33E-15	2.43E-11
MAP3K1	mitogen-activated protein kinase kinase	1618344	63	39	62	3.00E-15	0.34	0.17	8.99E-15	3.21E-11
PIK3CA	phosphoinositide-3-kinase, catalytic, al	1317693	182	168	32	3.77E-15	3.77E-07	3.77E-07	1.13E-14	3.21E-11
MAP2K4	mitogen-activated protein kinase kinase	449202	21	21	20	4.11E-15	0.016	0.034	1.23E-14	3.21E-11
PTEN	phosphatase and tensin homolog (mutated	493818	18	17	17	8.66E-15	0.13	0.23	2.60E-14	5.93E-11
CBFB	core-binding factor, beta subunit	206856	8	8	8	1.01E-11	0.13	0.073	3.04E-11	6.17E-08
MLL3	myeloid/lymphoid or mixed-lineage leuker	5817825	37	36	37	2.07E-10	0.43	0.18	6.21E-10	1.13E-06
TBX3	T-box 3 (ulnar mammary syndrome)	510042	13	13	13	2.38E-10	0.001	0.63	7.13E-10	1.18E-06
CTCF	CCCTC-binding factor (zinc finger protei	894348	13	13	11	5.65E-09	0.032	0.00047	1.69E-08	2.58E-05
TBL1XR1	transducin (beta)-like 1 X-linked recept	540969	10	8	9	2.29E-07	0.29	0.88	6.86E-07	0.00096
AKT1	v-akt murine thymoma viral oncogene hom	556179	11	11	2	0.19	2.46E-07	4.93E-07	7.39E-07	0.00096
PIK3R1	phosphoinositide-3-kinase, regulatory su	945048	14	14	13	6.81E-07	4.65E-07	0.0055	1.40E-06	0.0017
NCOR1	nuclear receptor co-repressor 1	2889900	18	17	18	2.57E-06	0.52	0.21	7.72E-06	0.0088
RB1	retinoblastoma 1 (including osteosarcoma	1443936	10	9	10	4.49E-06	0.57	0.31	1.35E-05	0.014
ARID1A	AT rich interactive domain 1A (SWI-like)	2264769	11	11	11	5.08E-06	0.23	0.067	1.52E-05	0.015
GPS2	G protein pathway suppressor 2	449709	6	6	6	5.40E-06	0.034	0.074	1.62E-05	0.016
ZNF283	zinc finger protein 283	767598	2	2	2	0.28	0.00053	3.15E-05	9.46E-05	0.086
ZFP36L1	zinc finger protein 36, C3H type-like 1	387348	7	7	7	3.60E-05	0.072	0.91	0.00011	0.09
SF3B1	splicing factor 3b, subunit 1, 155kDa	1583868	9	9	6	0.2	5.71E-05	0.074	0.00017	0.13
NF1	neurofibromin 1 (neurofibromatosis, von	4771884	15	14	15	5.85E-05	0.7	0.86	0.00018	0.13
LMTK3	lemur tyrosine kinase 3	571896	2	2	2	0.44	8.93E-05	0.0036	0.00027	0.2
C17orf71	chromosome 17 open reading frame 71	1162551	6	6	6	0.00012	0.76	0.73	0.00037	0.26
FGFR2	fibroblast growth factor receptor 2 (bac	1059630	4	4	2	0.6	0.0002	0.95	0.0006	0.41
COL11A1	collagen, type XI, alpha 1	2170467	2	2	2	1	0.00032	0.049	0.00096	0.63
CHERP	calcium homeostasis endoplasmic reticulu	785343	2	1	2	0.82	0.00034	0.11	0.001	0.64
KRT38	keratin 38	534378	3	3	1	0.087	0.0004	0.99	0.0012	0.73

KIRC kidney

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
PBRM1	polybromo 1	1602645	146	141	139	1.00E-15	4.80E-01	1.10E-01	3.00E-15	2.74E-11
VHL	von Hippel-Lindau tumor suppressor	120974	145	140	94	1.00E-15	2.76E-07	2.76E-07	3.00E-15	2.74E-11
SETD2	SET domain containing 2	1979377	50	47	49	3.22E-15	0.031	0.0018	9.66E-15	4.41E-11
BAP1	BRCA1 associated protein-1 (ubiquitin ca	637725	41	39	38	7.88E-15	0.06	0.57	2.36E-14	8.63E-11
KDM5C	lysine (K)-specific demethylase 5C	1247506	23	22	23	7.25E-09	8.10E-03	8.20E-02	2.18E-08	6.62E-05
NEFH	neurofilament, heavy polypeptide 200kDa	668244	10	10	9	6.43E-08	2.78E-07	0.0038	1.93E-07	5.00E-04
SCAF1	SR-related CTD-associated factor 1	589895	4	4	3	2.90E-03	2.71E-07	0.027	8.13E-07	0.0019
UQCRCF1	ubiquinol-cytochrome c reductase, Rieske	188800	3	3	1	3.30E-02	4.63E-07	0.046	1.39E-06	0.0028
ARID1A	AT rich interactive domain 1A (SWI-like)	1787148	14	14	14	1.04E-06	0.48	0.29	3.12E-06	0.0057
TP53	tumor protein p53	378000	10	10	10	1.47E-06	0.024	0.21	4.40E-06	0.0069
FOXD4L4	forkhead box D4-like 4	1289980	4	4	3	1.81E-05	1.52E-06	0.15	4.55E-06	0.0069
CY5B	cytochrome b5 type B (outer mitochondria	147600	3	1	3	3.70E-01	3.33E-06	0.012	1.00E-05	0.014
DNAH9	dynein, axonemal, heavy chain 9	4125652	20	19	19	0.0000128	2.60E-01	5.90E-01	3.83E-05	0.047
DHX9	DEAH (Asp-Glu-Ala-His) box polypeptide 9	1220348	2	2	2	8.90E-01	1.29E-05	0.0039	3.87E-05	0.047
FAM200A	family with sequence similarity 200, mem	60835	5	5	5	2.11E-05	0.43	0.034	6.32E-05	0.072
CD7	CD7 molecule	137600	4	3	3	3.00E-04	0.0000294	0.27	8.82E-05	0.095
TPST1	tyrosylprotein sulfotransferase 1	347044	2	2	2	2.40E-01	0.0024	0.0000584	1.80E-04	0.18
TP53I3	tumor protein p53 inducible protein 3	308400	4	2	4	7.50E-02	0.0000647	0.42	1.90E-04	0.19
ZNF800	zinc finger protein 800	632209	7	7	7	0.00013	0.76	8.40E-02	3.90E-04	0.36
TSPAN19	tetraspanin 19	132365	4	4	4	2.00E-04	0.008	0.24	0.00059	0.49
FGGY	FGGY carbohydrate kinase domain containi	542957	2	2	2	0.3	2.10E-04	0.0018	0.00064	0.51
CHSY3	chondroitin sulfate synthase 3	658261	3	3	3	5.40E-01	0.0003	0.52	0.0009	0.58
ROBO1	roundabout, axon guidance receptor, homo	1464595	2	1	2	1	3.00E-04	0.73	0.0009	0.58
RALGAPA1	Ral GTPase activating protein, alpha sub	1936329	9	9	9	0.18	0.00032	0.083	0.00096	0.58
CEBPZ	CCAAT/enhancer binding protein zeta	1094696	2	2	1	0.78	0.00032	0.95	0.00096	0.58
BTNL8	butyrophilin-like 8	492834	5	5	5	0.00033	0.14	0.9	0.001	0.58
SLC4A1AP	solute carrier family 4 (anion exchanger	764226	2	1	2	0.87	0.00038	0.9	0.0011	0.58
CMYA5	cardiomyopathy associated 5	3644765	5	5	5	0.52	0.0004	0.82	0.0012	0.58
TREH	trehalase (brush-border membrane glycopr	299283	2	2	1	0.1	0.0004	0.99	0.0012	0.58

UCEC endometrial

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
PTEN	phosphatase and tensin homolog (mutated)	241552	255	161	149	1.00E-15	4.65E-07	3.60E-02	3.00E-15	1.37E-11
ARID1A	AT rich interactive domain 1A (SWI-like)	1107816	107	83	87	1.00E-15	3.50E-01	0.2	3.00E-15	1.37E-11
TP53	tumor protein p53	234360	74	69	50	1.00E-15	2.63E-07	2.63E-07	3.00E-15	1.37E-11
CTCF	CCCTC-binding factor (zinc finger protei	437472	52	45	41	1.00E-15	0.003	0.14	3.00E-15	1.37E-11
CTNNB1	catenin (cadherin-associated protein), b	456072	80	74	25	1.89E-15	3.05E-07	3.05E-07	5.66E-15	2.07E-11
PIK3R1	phosphoinositide-3-kinase, regulatory su	462272	118	83	89	2.55E-15	3.27E-07	1.70E-02	7.66E-15	2.33E-11
PIK3CA	phosphoinositide-3-kinase, catalytic, al	644552	172	132	76	5.11E-15	2.91E-07	2.91E-07	1.53E-14	3.99E-11
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral onco	150040	53	53	11	6.55E-15	2.84E-07	2.84E-06	1.97E-14	4.48E-11
FBXW7	F-box and WD repeat domain containing 7	481616	47	39	31	8.90E-11	0.0023	0.0067	2.67E-10	5.42E-07
PPP2R1A	protein phosphatase 2 (formerly 2A), reg	328104	30	27	18	9.56E-06	2.53E-07	0.0033	7.58E-07	1.10E-03
MFGE8	milk fat globule-EGF factor 8 protein	215016	4	4	1	1.00E+00	2.74E-07	0.027	8.22E-07	1.10E-03
ABI1	abl-interactor 1	317936	5	4	2	1.00E+00	2.75E-07	0.027	8.24E-07	1.10E-03
PRKAR1B	protein kinase, cAMP-dependent, regulato	192944	4	4	4	9.00E-01	0.00014	2.81E-07	8.43E-07	0.0011
SMTNL2	smoothelin-like 2	159216	9	9	3	0.0001	2.90E-07	1.30E-02	8.70E-07	0.0011
ZRANB3	zinc finger, RAN-binding domain containi	569656	11	8	11	1.00E+00	1.60E-02	3.03E-07	9.09E-07	0.0011
FSIP2	fibrous sheath interacting protein 2	941656	30	15	30	9.90E-01	3.18E-07	0.0077	9.55E-07	0.0011
CABLES1	Cdk5 and Abl enzyme substrate 1	251968	7	6	7	1.00E+00	0.00088	4.35E-07	1.30E-06	0.0014
DST	dystonin	3897072	133	39	129	1.00E+00	4.48E-07	0.043	1.35E-06	0.0014
KIF26B	kinesin family member 26B	806248	25	19	25	1	5.17E-07	2.30E-03	1.55E-06	0.0014
ARID5B	AT rich interactive domain 5B (MRF1-like	692912	35	29	34	2.25E-06	0.036	9.40E-01	0.00000676	0.0059
DCAF4	DDB1 and CUL4 associated factor 4	295120	4	4	3	9.90E-01	0.0033	2.37E-06	0.00000712	0.0059
FGFR2	fibroblast growth factor receptor 2 (bac	518320	35	31	20	2.65E-06	1.80E-02	0.29	0.00000796	0.0063
SPOP	speckle-type POZ protein	226920	23	21	18	3.91E-06	0.0026	0.078	0.0000117	0.0089
NHSL2	NHS-like 2	604376	13	12	13	0.99	4.29E-06	5.21E-06	0.0000129	0.0094
NFE2L2	nuclear factor (erythroid-derived 2)-lik	349432	15	15	12	0.0012	7.32E-06	0.0058	0.000022	0.015
HHIPL1	HHIP-like 1	244528	9	6	9	0.97	0.0048	0.0000152	0.0000457	0.031
ZNF263	zinc finger protein 263	396056	13	8	10	0.69	0.0000169	0.00041	0.0000507	0.033
MORC4	MORC family CW-type zinc finger 4	534936	28	20	26	0.0000182	0.33	0.14	0.0000545	0.034
CCND1	cyclin D1	145328	15	15	13	0.00013	0.0000227	0.14	0.0000682	0.04

PanCan8 (part 1/5)

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
TP53	tumor protein p53	2024190	907	879	291	1.00E-15	9.09E-06	9.09E-06	3.00E-15	5.48E-12
PIK3CA	phosphoinositide-3-kinase, catalytic, al	5594179	452	389	108	1.00E-15	3.70E-07	3.7E-07	3.00E-15	5.48E-12
PTEN	phosphatase and tensin homolog (mutated	2083163	399	295	217	1.00E-15	1.79E-06	0.64	3.00E-15	5.48E-12
APC	adenomatous polyposis coli	14292359	331	216	228	1.00E-15	1.19E-06	0.93	3.00E-15	5.48E-12
PBRM1	polybromo 1	8568833	176	168	167	1.00E-15	0.34	0.32	3.00E-15	5.48E-12
ARID1A	AT rich interactive domain 1A (SWI-like)	9613332	174	146	146	1.00E-15	1.10E-01	2.60E-02	3.00E-15	5.48E-12
PIK3R1	phosphoinositide-3-kinase, regulatory su	3993623	166	128	113	1.00E-15	1.13E-07	0.13	3.00E-15	5.48E-12
NF1	neurofibromin 1 (neurofibromatosis, von	20164924	142	111	132	1.00E-15	0.5	0.74	3.00E-15	5.48E-12
RB1	retinoblastoma 1 (including osteosarcoma	6102796	90	77	79	1.00E-15	0.19	0.16	3.00E-15	5.48E-12
CDH1	cadherin 1, type 1, E-cadherin (epitheli	4205008	62	60	57	1.00E-15	0.14	0.34	3.00E-15	5.48E-12
SETD2	SET domain containing 2	10572521	128	104	121	1.44E-15	0.00013	4.68E-05	4.33E-15	7.18E-12
GATA3	GATA binding protein 3	1817963	74	69	50	2.22E-15	7.5E-08	0.043	6.66E-15	1.01E-11
CTCF	CCCTC-binding factor (zinc finger protei	3778573	78	71	64	7.77E-15	0.0018	0.011	2.33E-14	3.27E-11
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral onco	1296420	157	157	21	1.08E-14	1.10E-07	1.10E-07	3.23E-14	4.21E-11
FBXW7	F-box and WD repeat domain containing 7	4160274	111	96	61	1.21E-14	8.13E-08	5.98E-05	3.63E-14	4.42E-11
BAP1	BRCA1 associated protein-1 (ubiquitin ca	3402381	61	58	56	1.57E-14	0.17	0.56	4.70E-14	5.36E-11
VHL	von Hippel-Lindau tumor suppressor	647054	151	146	98	3.21E-14	1E-07	1E-07	9.63E-14	1.03E-10
MAP3K1	mitogen-activated protein kinase kinase	6870141	115	77	107	9.03E-14	0.0072	0.021	2.71E-13	2.75E-10
CTNNB1	catenin (cadherin-associated protein), b	3958224	99	92	43	1.43E-12	8.15E-08	7.48E-05	4.29E-12	4.12E-09
EGFR	epidermal growth factor receptor (erythr	6596765	105	95	69	1.68E-11	8.79E-05	8.40E-08	5.04E-11	4.6E-08
NRAS	neuroblastoma RAS viral (v-ras) oncogene	987632	33	33	12	1.35E-10	1.61E-07	4.70E-04	4.04E-10	3.51E-07
MAP2K4	mitogen-activated protein kinase kinase	1907020	43	42	37	1.11E-09	0.00018	0.00005	3.34E-09	2.77E-06
RUNX1	runt-related transcription factor 1 (acu	1654975	34	31	32	4.29E-09	7.22E-06	0.15	1.29E-08	0.0000102
MTM1	myotubularin 1	3141746	35	27	33	6.87E-08	0.25	0.83	2.06E-07	0.00015
BRAF	v-raf murine sarcoma viral oncogene homo	3716939	52	48	29	0.00071	6.96E-08	8.84E-06	2.09E-07	0.00015
HCN1	hyperpolarization activated cyclic nucle	4065431	91	74	79	1	7.45E-08	0.1	2.24E-07	0.00016
FSIP2	fibrous sheath interacting protein 2	8189066	44	25	43	0.98	7.86E-08	0.048	2.36E-07	0.00016
NFE2L2	nuclear factor (erythroid-derived 2)-lik	3032342	54	51	30	3.26E-06	8.5E-08	8.5E-08	2.55E-07	0.00017
TNXB	tenascin XB	18341984	72	52	68	1	1.13E-07	0.48	3.4E-07	0.00021

PanCan8 (part 2/5)

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
AKT1	v-akt murine thymoma viral oncogene hom	2361092	21	21	10	0.1	1.4E-07	1.4E-07	4.19E-07	0.00025
IDH1	isocitrate dehydrogenase 1 (NADP+), solu	2167238	27	27	13	0.017	1.59E-07	0.82	4.77E-07	0.00028
PPP2R1A	protein phosphatase 2 (formerly 2A), reg	2846835	49	46	32	0.18	1.64E-07	0.054	4.91E-07	0.00028
SGPP2	sphingosine-1-phosphate phosphotase 2	1600754	12	11	10	0.85	0.05	3.13E-07	9.38E-07	0.00052
CLIC6	chloride intracellular channel 6	1258485	11	10	11	0.94	0.00012	4.61E-07	1.38E-06	0.00074
LPA	lipoprotein, Lp(a)	7112749	58	49	56	0.99	6.39E-07	0.64	1.92E-06	0.001
PRKAR1B	protein kinase, cAMP-dependent, regulato	1675039	7	7	7	1	0.15	8.54E-07	2.56E-06	0.0013
MAGEL2	MAGE-like 2	2707825	22	22	22	1	1.25E-06	0.86	3.76E-06	0.0019
HCN4	hyperpolarization activated cyclic nucle	3813564	43	34	40	0.84	1.85E-06	1.08E-05	5.54E-06	0.0027
FOXD4L4	forkhead box D4-like 4	6048412	5	4	3	0.0012	4.08E-06	0.99	0.0000122	0.0056
NEFH	neurofilament, heavy polypeptide 200kDa	3573876	30	29	29	0.0009	4.1E-06	0.43	0.0000123	0.0056
PKD2	polycystic kidney disease 2 (autosomal d	3986262	30	24	29	0.85	2.86E-05	4.17E-06	0.0000125	0.0056
PRSSL1	protease, serine-like 1	622212	3	3	1	0.95	4.94E-06	0.054	0.0000148	0.0059
NHSL2	NHS-like 2	5213509	28	27	26	1	4.96E-06	0.35	0.0000149	0.0059
UGT1A7	UDP glucuronosyltransferase 1 family, po	10594557	15	13	14	0.87	4.97E-06	0.99	0.0000149	0.0059
TRIM7	tripartite motif-containing 7	1662887	14	14	12	0.99	0.000005	0.072	0.000015	0.0059
B2M	beta-2-microglobulin	617798	19	17	15	5.06E-06	0.26	0.11	0.0000152	0.0059
PDE8B	phosphodiesterase 8B	4039540	33	27	32	0.97	0.065	5.07E-06	0.0000152	0.0059
CASKIN1	CASK interacting protein 1	3566194	16	16	15	1	5.34E-06	0.59	0.000016	0.0061
CDKN2A	cyclin-dependent kinase inhibitor 2A (me	1275842	33	33	28	6.01E-06	0.0038	9.76E-05	0.000018	0.0067
STRC	stereocilin	3514738	23	17	22	1	7.46E-06	0.23	0.0000224	0.0082
CACNA1A	calcium channel, voltage-dependent, P/Q	8710531	66	53	65	1	0.037	7.64E-06	0.0000229	0.0082
FAM38B	family with sequence similarity 38, memb	2871336	10	8	10	1	8.64E-06	0.046	0.0000259	0.0091
FAM200A	family with sequence similarity 200, mem	325387	13	13	13	1.25E-05	0.43	0.62	0.0000376	0.013
CABLES1	Cdk5 and Abl enzyme substrate 1	2186258	13	12	13	0.99	0.0012	0.000014	0.000042	0.014
CDKN1B	cyclin-dependent kinase inhibitor 1B (p2	1001202	19	19	17	1.41E-05	0.58	0.47	0.0000424	0.014
CHSY3	chondroitin sulfate synthase 3	3523845	33	30	32	0.97	1.49E-05	0.065	0.0000447	0.015
FAM124B	family with sequence similarity 124B	1217771	16	16	15	0.054	1.65E-05	0.31	0.0000494	0.016
C10orf7	1 chromosome 10 open reading frame 71	3403461	41	33	39	0.89	1.69E-05	0.47	0.0000506	0.016

PanCan8 (part 3/5)

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
LRRC37A	3 leucine rich repeat containing 37, memb	6006559	20	18	20	1	1.69E-05	0.34	0.0000506	0.016
HEATR7A	HEAT repeat containing 7A	1884154	10	9	9	0.96	1.97E-05	0.54	0.0000592	0.018
ING1	inhibitor of growth family, member 1	1980907	22	22	15	0.029	2.06E-05	0.003	0.0000618	0.018
UGT2B10	UDP glucuronosyltransferase 2 family, po	5525247	28	19	24	0.96	2.17E-05	0.83	0.0000652	0.019
PEG10	paternally expressed 10	1727186	14	12	13	0.98	2.54E-05	0.013	0.0000763	0.022
OBSCN	obscurin, cytoskeletal calmodulin and ti	28442270	178	111	174	1	2.69E-05	0.7	0.0000808	0.023
PHLPP1	PH domain and leucine rich repeat protei	5320330	34	30	33	1	0.00038	2.86E-05	0.0000857	0.024
KIF13B	kinesin family member 13B	7164983	40	31	38	1	2.94E-05	0.11	0.0000882	0.024
MYO1B	myosin IB	5933850	48	40	47	0.096	3.26E-05	0.94	0.0000977	0.026
PAN3	PAN3 polyA specific ribonuclease subunit	3792020	28	25	26	0.46	3.26E-05	0.058	0.0000977	0.026
GRIN2D	glutamate receptor, ionotropic, N-methyl	3239868	23	21	22	1	0.00016	3.26E-05	0.0000977	0.026
RHEB	Ras homolog enriched in brain	921491	8	8	5	0.24	3.33E-05	0.056	0.0001	0.026
CBFB	core-binding factor, beta subunit	879843	15	13	14	3.56E-05	0.0089	0.0049	0.00011	0.027
CDK17	cyclin-dependent kinase 17	2700870	29	26	28	3.77E-05	0.22	0.38	0.00011	0.029
FAM122A	family with sequence similarity 122A	1363324	7	7	5	1	0.00004	0.068	0.00012	0.029
CCDC144	A coiled-coil domain containing 144A	4778026	48	32	44	1	0.54	0.00004	0.00012	0.029
SYN1	synapsin I	2885563	23	19	22	1	0.00004	0.029	0.00012	0.029
UGT1A9	UDP glucuronosyltransferase 1 family, po	12032180	16	15	15	0.56	4.05E-05	0.99	0.00012	0.029
PRRX1	paired related homeobox 1	1255724	17	17	16	0.62	4.24E-05	0.11	0.00013	0.03
PLCL1	phospholipase C-like 1	5111815	61	46	57	0.79	0.05	0.00005	0.00015	0.035
CYP4F22	cytochrome P450, family 4, subfamily F,	2563124	29	26	26	1	0.00005	0.26	0.00015	0.035
TAF4	TAF4 RNA polymerase II, TATA box binding	3070753	25	25	25	0.95	0.00022	5.17E-05	0.00016	0.035
UGT2B15	UDP glucuronosyltransferase 2 family, po	5187011	18	18	17	1	5.19E-05	0.077	0.00016	0.035
TWF1	twinfilin, actin-binding protein, homolo	1755233	19	13	16	0.13	5.38E-05	0.83	0.00016	0.036
ST7L	suppression of tumorigenicity 7 like	2945653	17	15	16	5.83E-05	0.022	0.043	0.00017	0.038
IGFBP1	insulin-like growth factor binding prote	901949	7	6	5	0.97	6.09E-05	0.93	0.00018	0.04
ATM	ataxia telangiectasia mutated	15991424	144	93	132	6.31E-05	0.38	0.062	0.00019	0.04
SDK1	sidekick homolog 1, cell adhesion molecu	9983726	92	78	84	1	6.36E-05	0.67	0.00019	0.04
ZEB2	zinc finger E-box binding homeobox 2	6231333	74	64	68	0.99	6.36E-05	0.85	0.00019	0.04

PanCan8 (part 4/5)

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
SMAD4	SMAD family member 4	2836348	43	39	35	6.41E-05	0.083	0.0037	0.00019	0.04
APRT	adenine phosphoribosyltransferase	548783	3	1	3	0.97	6.67E-05	0.98	0.0002	0.041
CPEB2	cytoplasmic polyadenylation element bind	2539795	13	13	12	1	0.00042	7.37E-05	0.00022	0.044
MYH3	myosin, heavy chain 3, skeletal muscle,	10112722	67	42	59	1	0.47	7.37E-05	0.00022	0.044
PPM1E	protein phosphatase 1E (PP2C domain cont	3275203	33	25	30	0.62	0.0071	7.37E-05	0.00022	0.044
GLDC	glycine dehydrogenase (decarboxylating)	4865388	28	25	27	1	0.046	0.000075	0.00022	0.044
RASGRF2	Ras protein-specific guanine nucleotide-	6206649	46	38	44	1	0.09	0.000075	0.00022	0.044
LPPR3		944797	7	6	7	0.94	7.65E-05	0.092	0.00023	0.044
MYCL1	v-myc myelocytomatosis viral oncogene ho	1394148	10	10	10	0.7	7.89E-05	0.31	0.00024	0.045
CLIP2	CAP-GLY domain containing linker protein	4239138	23	21	23	1	0.00008	0.96	0.00024	0.045
TMEM132	E transmembrane protein 132E	3957395	25	25	25	1	8.24E-05	0.14	0.00025	0.046
OPRM1	opioid receptor, mu 1	4604149	33	29	27	1	0.00028	8.33E-05	0.00025	0.046
GPS2	G protein pathway suppressor 2	1908911	17	17	17	8.47E-05	0.13	0.1	0.00025	0.046
C3orf63	chromosome 3 open reading frame 63	7190388	22	19	22	1	8.57E-05	0.13	0.00026	0.046
SLC12A2	solute carrier family 12 (sodium/potassi	5324910	40	30	36	1	8.75E-05	0.18	0.00026	0.047
CHD4	chromodomain helicase DNA binding prote	9983737	91	79	79	0.022	9.33E-05	0.073	0.00028	0.049
NRG1	neuregulin 1	4996691	56	44	54	1	9.33E-05	0.5	0.00028	0.049
CNNM3	cyclin M3	2033988	14	13	14	1	9.33E-05	0.57	0.00028	0.049
ERBB2	v-erb-b2 erythroblastic leukemia viral o	6243886	41	36	29	1	0.0038	9.86E-05	0.0003	0.05
SSX4B	synovial sarcoma, X breakpoint 4B	297324	6	6	6	0.044	0.0001	0.94	0.0003	0.05
ZNF498	zinc finger protein 498	2670309	17	14	16	1	0.0001	0.39	0.0003	0.05
NOBOX	NOBOX oogenesis homeobox	1476295	22	20	21	1	0.0001	0.27	0.0003	0.05
GATA6	GATA binding protein 6	1342390	10	10	9	0.95	0.0001	0.0032	0.0003	0.05
LCMT2	leucine carboxyl methyltransferase 2	3267020	19	15	19	1	0.00011	0.94	0.00032	0.053
NRXN2	neurexin 2	6289762	46	43	46	1	0.00011	0.087	0.00032	0.053
SAP30	Sin3A-associated protein, 30kDa	882494	7	7	5	0.9	0.0046	0.00011	0.00033	0.053
ZIM3	zinc finger, imprinted 3	2470151	31	29	27	0.63	0.27	0.00011	0.00034	0.055
AGAP2	ArfGAP with GTPase domain, ankyrin repea	3828451	30	27	28	1	0.00012	0.069	0.00035	0.055
CDAN1	congenital dyserythropoietic anemia, typ	4936166	28	24	26	1	0.55	0.00012	0.00035	0.055

PanCan8 (part 5/5)

gene	description	N	n	npat	nsite	pcv	pcl	pco	p	q
CMAS	cytidine monophosphate N-acetylneuraminic acid	1953249	20	17	17	0.36	0.027	0.00012	0.00035	0.055
NLGN2	neuroligin 2	2932915	13	13	12	1	0.0014	0.00012	0.00036	0.056
C6orf15	chromosome 6 open reading frame 15	1459143	8	8	8	0.98	0.76	0.00013	0.00038	0.058
PFN2	profilin 2	662816	6	6	6	0.29	0.22	0.00013	0.00038	0.058
KAT2B	K(lysine) acetyltransferase 2B	3886523	30	21	25	0.99	0.12	0.00013	0.00038	0.058
MDGA2	MAM domain containing glycosylphosphatidylinositol transferase 2	3706596	60	52	57	0.82	0.00014	0.95	0.00042	0.063
TMEM185	A transmembrane protein 185A	802783	12	10	12	0.32	0.00014	0.0013	0.00042	0.063
GSX2	GS homeobox 2	900122	7	7	6	1	0.00015	0.92	0.00044	0.064
KIF3A	kinesin family member 3A	3607196	24	22	23	0.85	0.00015	0.78	0.00045	0.065
PTGES3	prostaglandin E synthase 3 (cytosolic)	900058	2	2	2	0.97	0.36	0.00015	0.00045	0.065
E2F1	E2F transcription factor 1	1481446	14	14	14	0.83	0.00024	0.00016	0.00048	0.068
CCND1	cyclin D1	1261855	17	17	15	0.13	0.00016	0.49	0.00048	0.068
BTBD11	BTB (POZ) domain containing 11	4101167	28	25	27	1	0.0078	0.00017	0.00052	0.073
MAGI2	membrane associated guanylate kinase, with PDZ domain containing 2	6744036	71	52	66	1	0.00019	0.0012	0.00056	0.078
CCDC105	coiled-coil domain containing 105	1476945	22	20	20	0.93	0.00019	0.17	0.00056	0.078
LONRF1	LON peptidase N-terminal domain and ring 1	2725801	22	21	22	0.62	0.00058	0.00019	0.00057	0.079
FAM69C	family with sequence similarity 69, member C	553847	10	9	9	0.22	0.00019	0.0013	0.00058	0.079
IRF2BP1	interferon regulatory factor 2 binding protein 1	2003123	10	9	9	1	0.00019	0.024	0.00058	0.079
TCF7	transcription factor 7 (T-cell specific)	1639225	14	13	13	0.72	0.0002	0.24	0.0006	0.081
KIAA036	8 KIAA0368	8558540	54	39	54	1	0.15	0.0002	0.0006	0.081
CTSG	cathepsin G	1254268	15	15	15	0.0002	0.93	0.084	0.00061	0.081
TGFB2	transforming growth factor, beta receptor type 2	2927068	27	25	25	0.89	0.0034	0.00021	0.00064	0.083
KCTD1	potassium channel tetramerisation domain containing 1	1326067	14	14	13	1	0.00021	0.051	0.00064	0.083
COL4A5	collagen, type IV, alpha 5 (Alport syndrome)	7834416	69	51	65	0.61	0.44	0.00021	0.00064	0.083
PDGFR1	platelet-derived growth factor receptor-1	1871979	6	6	6	1	0.0061	0.00021	0.00064	0.083
KIAA180	18 KIAA180	3876680	41	34	38	0.67	0.00023	0.026	0.0007	0.088
LTBP4	latent transforming growth factor beta binding protein 4	5333713	39	29	39	1	0.00023	0.13	0.0007	0.088
MAP3K7	mitogen-activated protein kinase kinase 7	3112411	22	19	22	1	0.43	0.00023	0.0007	0.088
STK35	serine/threonine kinase 35	1608414	6	6	6	1	0.065	0.00023	0.0007	0.088

Top genes: sorted by max % patients mutated

1-30

	BRCA	COAD	GBM	KIRC	LUSC	OV	READ	UCEC	PanCan8
TP53	36	48	29	3	79	87	65	28	41
APC	1	66	1	1	4	2	83	12	10
PTEN	3	3	29	4	8	1	4	65	14
KRAS	1	37	0	0	1	1	54	21	7
PIK3CA	33	17	7	2	15	1	10	53	18
PBRM1	0	3	0	36	4	0	0	4	8
VHL	0	0	0	36	1	0	1	1	7
ARID1A	2	10	1	4	7	1	7	33	7
PIK3R1	3	3	6	1	1	0	6	33	6
CTNNB1	0	5	0	0	2	1	6	30	4
MLL2	2	6	2	3	20	0	3	13	5
EGFR	1	6	20	2	3	2	1	3	4
FAT4	1	19	1	2	15	1	14	16	6
MUC5B	2	3	4	1	19	0	3	17	5
CTCF	3	5	0	1	1	0	3	18	3
FBXW7	0	18	1	0	6	1	13	16	4
DMD	3	17	3	2	17	2	9	17	6
HCN1	1	5	4	0	17	0	3	7	3
MLL3	7	6	2	4	16	2	3	10	6
SDK1	1	16	3	2	5	0	3	7	4
DNAH9	2	9	3	5	16	2	6	12	6
DST	2	10	2	4	8	4	4	16	6
OBSCN	2	6	7	1	10	2	6	16	5
NFE2L2	0	1	0	2	15	0	1	6	2
CDKN2A	0	1	1	0	15	0	0	1	2
CHD4	1	8	0	1	3	3	7	14	4
ATM	2	14	1	3	4	1	6	12	4
KIAA1804	0	4	1	0	3	0	13	4	2
FGFR2	1	1	1	0	2	0	3	13	2
RIMS2	1	7	5	1	12	2	1	9	4

31-60

	BRCA	COAD	GBM	KIRC	LUSC	OV	READ	UCEC	PanCan8
KEAP1	0	2	0	1	12	0	0	2	2
FAM123B	1	12	1	1	4	1	9	7	3
BRAF	1	12	2	0	4	1	3	3	2
SETD2	1	6	3	12	3	2	4	9	5
NF1	3	3	9	2	12	4	4	8	5
RELN	3	12	5	1	12	3	4	9	5
FMN2	1	3	0	1	12	1	4	8	3
ARID5B	0	2	1	1	2	1	3	12	2
DCHS2	1	12	1	1	9	1	7	8	4
SMAD4	0	12	0	0	3	0	12	2	2
PPP2R1A	0	3	0	1	4	1	0	11	2
COL6A6	1	9	2	3	11	2	3	8	4
GATA3	11	2	1	0	3	0	0	1	3
EYS	2	5	0	2	7	0	1	10	3
BAP1	0	1	1	10	1	1	0	4	3
NRAS	0	10	0	0	1	1	7	4	2
LPHN3	1	5	0	2	9	1	9	6	3
MDGA2	1	5	1	1	9	1	3	6	2
MAP3K1	8	3	2	1	2	0	1	8	4
PLCL1	0	2	1	0	7	0	4	8	2
SIN3A	1	2	1	0	3	1	1	8	2
SPOP	0	1	0	0	1	0	0	8	1
ZEB2	1	7	0	1	8	1	6	8	3
TRIM58	1	1	1	1	8	1	1	3	2
COL4A5	1	8	1	2	2	1	3	6	2
RB1	2	2	8	0	7	3	3	8	4
MORC4	0	2	1	1	2	1	0	8	2
ADCY2	0	7	1	1	8	2	4	6	2
TNXB	1	6	1	1	8	1	4	5	2
NUP210L	1	3	3	1	8	1	3	4	2

Top genes: sorted by total % patients mutated

1-30

	BRCA	COAD	GBM	KIRC	LUSC	OV	READ	UCEC	PanCan8
TP53	36	48	29	3	79	87	65	28	41
PIK3CA	33	17	7	2	15	1	10	53	18
PTEN	3	3	29	4	8	1	4	65	14
APC	1	66	1	1	4	2	83	12	10
PBRM1	0	3	0	36	4	0	0	4	8
KRAS	1	37	0	0	1	1	54	21	7
VHL	0	0	0	36	1	0	1	1	7
ARID1A	2	10	1	4	7	1	7	33	7
DMD	3	17	3	2	17	2	9	17	6
MLL3	7	6	2	4	16	2	3	10	6
PIK3R1	3	3	6	1	1	0	6	33	6
FAT4	1	19	1	2	15	1	14	16	6
DNAH9	2	9	3	5	16	2	6	12	6
DST	2	10	2	4	8	4	4	16	6
OBSCN	2	6	7	1	10	2	6	16	5
NF1	3	3	9	2	12	4	4	8	5
MLL2	2	6	2	3	20	0	3	13	5
MUC5B	2	3	4	1	19	0	3	17	5
SETD2	1	6	3	12	3	2	4	9	5
RELN	3	12	5	1	12	3	4	9	5
FBXW7	0	18	1	0	6	1	13	16	4
EGFR	1	6	20	2	3	2	1	3	4
ATM	2	14	1	3	4	1	6	12	4
CTNNB1	0	5	0	0	2	1	6	30	4
RIMS2	1	7	5	1	12	2	1	9	4
COL6A6	1	9	2	3	11	2	3	8	4
CHD4	1	8	0	1	3	3	7	14	4
SDK1	1	16	3	2	5	0	3	7	4
DCHS2	1	12	1	1	9	1	7	8	4
MAP3K1	8	3	2	1	2	0	1	8	4

31-60

	BRCA	COAD	GBM	KIRC	LUSC	OV	READ	UCEC	PanCan8
RB1	2	2	8	0	7	3	3	8	4
HCN1	1	5	4	0	17	0	3	7	3
CTCF	3	5	0	1	1	0	3	18	3
GATA3	11	2	1	0	3	0	0	1	3
FAM123B	1	12	1	1	4	1	9	7	3
FMN2	1	3	0	1	12	1	4	8	3
EYS	2	5	0	2	7	0	1	10	3
ZEB2	1	7	0	1	8	1	6	8	3
CDH1	7	3	1	1	2	1	1	5	3
KDM5C	1	5	1	6	2	2	0	6	3
LPHN3	1	5	0	2	9	1	9	6	3
BAP1	0	1	1	10	1	1	0	4	3
NCOR1	3	5	1	1	4	0	3	6	3
ADCY2	0	7	1	1	8	2	4	6	2
ATP8A2	0	5	2	2	6	1	3	7	2
CACNA1A	1	5	1	1	6	2	3	6	2
MDGA2	1	5	1	1	9	1	3	6	2
TNXB	1	6	1	1	8	1	4	5	2
MAGI2	1	4	1	1	7	1	4	6	2
NFE2L2	0	1	0	2	15	0	1	6	2
COL4A5	1	8	1	2	2	1	3	6	2
NUP210L	1	3	3	1	8	1	3	4	2
PAPPA	1	7	2	0	4	1	4	7	2
LPA	1	5	1	1	8	1	1	4	2
BRAF	1	12	2	0	4	1	3	3	2
ATG2B	1	4	0	1	8	0	3	6	2
CELSR1	1	4	2	2	5	0	0	6	2
FGFR2	1	1	1	0	2	0	3	13	2
PPP2R1A	0	3	0	1	4	1	0	11	2
PLCL1	0	2	1	0	7	0	4	8	2

Increased power from combining datasets

Number of significant genes ($q < 0.1$)

BRCA	22	LUSC	16	PanCan8	150
COAD	14	OV	3		
GBM	12	READ	3		
KIRC	12	UCEC	33		

Many new genes come up in PanCan8 analysis:

not significant in any tumor type alone (all $p_{\text{individual}} > 0.2$)
but highly significant in the combined analysis.

e.g.	HCN1	($p = 2e-7$ npat = 74)	hyperpolarization activated cyclic nucleotide channel
	HCN4	($p = 1e-6$ npat = 34)	(HCN2 and HCN3 are also near significance)
	B2M	($p = 1e-5$ npat = 17)	beta-2-microglobulin
	UGT2B10	($p = 6e-5$ npat = 13)	UDP glucuronosyltransferase 2 family B #10
	UGT1A7	($p = 1e-5$ npat = 19)	UDP glucuronosyltransferase 1 family A #7
	MTM1	($p = 2e-7$ npat = 27)	myotubularin

(npat_tot = 2143)

Increased power from combining datasets

Number of significant genes ($q < 0.1$)

BRCA	22	LUSC	16	PanCan8	150
COAD	14	OV	3		
GBM	12	READ	3		
KIRC	12	UCEC	33		

Many new genes come up in PanCan8 analysis:

not significant in any tumor type alone (all $p_{\text{individual}} > 0.2$)
but highly significant in the combined analysis.

e.g.	HCN1	($p = 2e-7$ npat = 74)	hyperpolarization activated cyclic nucleotide channel
	HCN4	($p = 1e-6$ npat = 34)	(HCN2 and HCN3 are also near significance)
	B2M	($p = 1e-5$ npat = 17)	beta-2-microglobulin
	UGT2B10	($p = 6e-5$ npat = 13)	UDP glucuronosyltransferase 2 family B #10
	UGT1A7	($p = 1e-5$ npat = 19)	UDP glucuronosyltransferase 1 family A #7
	MTM1	($p = 2e-7$ npat = 27)	myotubularin

(npat_tot = 2143)

Conclusions:

- Combining tumor types gives us significantly more power to detect putative driver genes that we are underpowered to detect in each tumor type separately.

Conclusions:

- Combining tumor types gives us significantly more power to detect putative driver genes that we are underpowered to detect in each tumor type separately.
- It also dilutes the power to detect driver genes that are potentially important in their respective tumor types.

Future steps:

- Incorporate other information for potential functional role apart from conservation (polyphen2, mutation assessor, CHASM, etc.).
- Perform this significance analysis on curated genesets.
- Extend this analysis to look at correlation and mutual exclusivity (MeMo) within and across tumor types, taking into account the variable mutation rate across the genome.
- Extend this analysis to pathways by performing significance analysis of altered gene sub-networks, by working with HotNet.

Acknowledgements

Analysis Team

Paz Polak
Gregory Kryukov
Scott Carter
Adam Kiezun
Peter Hammerman
Chip Stewart
Eran Hodis
Marcin Imielinski
Bryan Hernandez
Kristian Cibulskis
Andrey Sivachenko
Gordon Saksena
Yotam Drier
Alex Ramos
Aaron McKenna
Rui Jing
Lihua Zou
David DeLuca
Elena Helman
Jaegil Kim
Cheng-Zhong Zhang
Sylvan Baca
Nicolas Stransky
Trevor Pugh
Andrew Cherniak
Alex Kostic
Peter Chen

(cont.)

Mike Berger
Mike Chapman
Jens Lohr
Craig Mermel
Roel Verhaak
Derek Chiang
Barbara Weir
Nikhil Wagle
Shantanu Banerji
Firehose
Doug Voet
Mike Noble
Dan DiCara
Lee Lichtenstein
Pei Lin
Peter Carr

NHGRI NCI



Gaddy Getz

Stacey Gabriel

Matthew Meyerson

Levi Garraway

Eric Lander

Lynda Chin

Todd Golub

Broad Institute of Harvard and MIT

Dana Farber Cancer Institute

Rameen Beroukhim

Steve Schumacher
Barbara Tabak

Cathy Wu

Jennifer Brown

Lili Wang
Youzhong Wan
Dan-Avi Landau
Aviv Regev
Nathalie Pochet

Adam Bass

Austin Dulak

Project Management

Carrie Sougnez
Erica Shefler
Elizabeth Nickerson

Daniel Auclair

Marisa Cortes

Kristin Thompson

Jim Robinson

Helga Thorvaldsdottir
Marc-Danie Nazaire

Jill Mesirov

Shamil Sunyaev

Leonid Mirny

Broad Institute Sequencing Pro
and Platform

Tim Fennell

Sara Chauvin
Lauren Ambrogio
Sheila Fisher
Joshua Levin
Xian Adiconis
Andreas Gnirke

David Jaffe

Toby Bloom

Chad Nusbaum

Broad Institute Biological
Genetic Analysis Platform

Robb Onofrio
Brendan Blumenstiel
Huy Nguyen
Mellisa Parkin

Wendy Winckler

Broad Institute Biological Samp
Platform

Kristin Ardlie

Mark DePristo

Eric Banks
Kiran Garimella