Novel Driver Strength Index highlights important cancer genes in TCGA PanCanAtlas patients

Aleksey V. Belikov*, Danila V. Otnyukov, Alexey D. Vyatkin and Sergey V. Leonov

Laboratory of Innovative Medicine, School of Biological and Medical Physics, Moscow Institute of Physics and Technology, 141701 Dolgoprudny, Moscow Region, Russia

*Corresponding author: belikov.research@gmail.com

Abstract

Elucidating crucial driver genes is paramount for understanding the cancer origins and mechanisms of progression, as well as selecting targets for molecular therapy. Cancer genes are usually ranked by the frequency of mutation, which, however, does not necessarily reflect their driver strength. Here we hypothesize that driver strength is higher for genes that are preferentially mutated in patients with few driver mutations overall, because these few mutations should be strong enough to initiate cancer. We propose a formula to calculate the corresponding Driver Strength Index (DSI), as well as the Normalized Driver Strength Index (NDSI), the latter completely independent of the overall gene mutation frequency. We validate these indices using the largest database of human cancer mutations – TCGA PanCanAtlas, multiple established algorithms for cancer driver prediction (2020plus, CHASMplus, CompositeDriver, dNdScv, DriverNet, HotMAPS, IntOGen Plus, OncodriveCLUSTL, OncodriveFML) and four custom computational pipelines that integrate driver contributions from SNA, CNA and aneuploidy at the patient-level resolution. We demonstrate that NDSI provides substantially different rankings of genes as compared to DSI and frequency approach. For example, NDSI highlighted the importance of guanine nucleotide-binding protein subunits *GNAQ*, *GNA11*, *GNAI1, GNAZ* and *GNB3,* General Transcription Factor II family members *GTF2I* and *GTF2F2*, as well as fibroblast growth factor receptors *FGFR2* and *FGFR3*. Intriguingly, NDSI prioritized *CIC*, *FUBP1*, *IDH1* and *IDH2* mutations, as well as 19q and 1p chromosome arm losses, that comprise characteristic molecular alterations of gliomas. KEGG analysis shows that top NDSI-ranked genes comprise PDGFRA-GRB2-SOS2-HRAS/NRAS-BRAF pathway, GNAQ/GNA11-HRAS/NRAS-BRAF pathway, GNB3-AKT1-IKBKG/GSK3B/CDKN1B pathway and TCEB1-VHL pathway. NDSI does not seem to correlate with the number of protein-protein interactions. We share our software to enable calculation of DSI and NDSI for outputs of any third-party driver prediction algorithms or their combinations.

Introduction

Most cancer driver prediction algorithms answer one question – what is a probability of a given gene being a driver. This is definitely a crucial question and the answers are very valuable. However, high confidence that a gene is a driver does not translate to the statement that this gene is a strong driver. We can imagine a gene that is mutated in the majority of cancer patients (e.g. because it has multiple suitable sites for a driver mutation) but has a very weak contribution to cancer progression in each of these patients (e.g. because this gene is redundant). We can also imagine a gene that is mutated rarely (e.g. because it has only one suitable site for a driver mutation) but if the mutation does occur it immediately leads to cancer (e.g. because this gene is in a key position to control cell growth). The former would be an example of high confidence but weak driver, whereas the latter would be low confidence but strong driver. Overall, algorithms based on mutation recurrence cannot determine driver strength.

Some algorithms try to predict driver strength based on data from protein interaction networks (1)(2)(3). The idea is that a gene having multiple connections with other genes, i.e. playing the role of a network hub, will have more dramatic influence on the cell in case of mutation (4). This seems like a great idea at first sight, but a more detailed look shows that this is not the case. Yes, mutations in network hubs are likely to cause more disturbance in the cell, but what are the reasons to believe that all (or any of) these perturbations would be beneficial for cancer progression? In fact, mutations in network hubs are more likely to lead to cell death than to oncogenic transformation.

Here, we propose another approach. We reason that a few strong drivers are sufficient to initiate cancer, and there would be no need to accumulate additional drivers. On the other hand, weak drivers would need to accumulate in much higher quantity, until their combined strength would become sufficient to initiate cancer. Therefore, it should be statistically more likely to find strong drivers in patients that have only few driver mutations in their tumors, and less likely to find them in patients with multiple drivers per tumor. Likewise, it should be statistically less likely to find weak drivers in patients that have only few driver mutations in their tumors, and more likely to find them in patients with multiple drivers per tumor. Hence, we propose the Driver Strength Index (DSI) that takes into account the frequencies of mutation of a given driver gene in groups of patients with different total number of driver mutations, and gives priority weights to groups with fewer mutations. We also propose a modification of this index that is completely independent of the overall frequency of mutation of a given driver gene – the Normalized Driver Strength Index (NDSI).

Calculating these indices requires data on the number of driver mutations in each individual patient. The majority of existing driver prediction algorithms work at the cohort level, i.e. they predict driver genes for large groups of patients, usually having a particular cancer type. This does not allow to look at the composition of driver mutations in individual patients. We wrote specific scripts to convert cohort-level predictions into patient-level events, which also allowed seamless integration of the results from various third-party algorithms, including 2020plus (5), CHASMplus (6), CompositeDriver (7), dNdScv (8), DriverNet (9), HotMAPS (10), IntOGen Plus (11), OncodriveCLUSTL (12), and OncodriveFML (13). This is useful, as each individual driver prediction algorithm has its own strengths and shortcomings, and combining results from multiple algorithms allows to obtain more complete and balanced picture, ensuring that less driver mutations have been missed.

In addition to these existing driver prediction algorithms, we decided to create our own, using clear and simple rules to have an internal reference standard. We called this algorithm SNADRIF – Single Nucleotide Alteration DRIver Finder. It predicts cancer driver genes from the TCGA PanCanAtlas SNA data and classifies them into oncogenes and tumor suppressors. Driver prediction is based on calculating the ratio of nonsynonymous SNAs to silent SNAs (8), whereas driver classification is based on calculating the ratio of hyperactivating SNAs to inactivating SNAs (14). Bootstrapping is used to calculate statistical significance and Benjamini–Hochberg procedure is used to keep false discovery rate under 5%.

Copy-number alterations (CNA) usually involve large chunks of DNA containing tens or hundreds of genes, which makes CNA data not very useful for uncovering individual driver genes. Nevertheless, it is an important source of information about amplifications and deletions of driver genes predicted from SNA data. However, due to CNA data coarseness, we wanted to clarify the actual copy number status of individual genes using mRNA and miRNA expression data available at TCGA PanCanAtlas. For this purpose, we created another pipeline called GECNAV - Gene Expression-based CNA Validator. CNA validation is based on comparing the CNA status of a given gene in a given patient to expression of this gene in this patient relative to the median expression of this gene across all patients.

Aneuploidy – chromosome arm and full chromosome gains and losses – makes a substantial contribution to the number of driver alterations per tumor, and thus we needed to take it into account when calculating our indices. However, there are no existing algorithms to differentiate driver aneuploidies from passenger ones. Therefore, we built our own pipeline called ANDRIF - ANeuploidy DRIver Finder. Driver prediction is based on calculating the average alteration

status for each arm or chromosome in each cancer type. Bootstrapping is then used to obtain the realistic distribution of the average alteration statuses under the null hypothesis and Benjamini–Hochberg procedure is performed to keep the false discovery rate under 5%.

Finally, we needed an algorithm to integrate all data on driver mutations from different algorithms - our own and third-party. We called this algorithm PALDRIC - PAtient-Level DRIver Classifier. It translates cohort-level lists of driver genes or mutations to the patient level, classifies driver events according to the molecular causes and functional consequences, and presents comprehensive statistics on various kinds of driver events in various demographic and clinical groups of patients. Moreover, we developed a modification of PALDRIC that allows analysis and ranking of individual genes, chromosome arms and full chromosomes according to their frequency of occurrence, DSI, and NDSI.

Our overall workflow can be seen in **Fig 1**.

Results

To get three different perspectives on the number and composition of driver events, we performed three different analyses. In the first one we used results of our own SNADRIF algorithm applied to the whole TCGA PanCanAtlas dataset. We will refer to this analysis as "Pancan SNADRIF" (for data and graphs see **Supplementary Data 1**). In the second analysis we used the combination of results from several third-party algorithms - 2020plus, CHASMplus, CompositeDriver, OncodriveFML and consensus results from (7) – also applied to the whole TCGA PanCanAtlas dataset. We will refer to this analysis as "Pancan combined" (for data and graphs see **Supplementary Data 2**). In the third analysis we used the combination of results from 2020plus, CHASMplus, CompositeDriver, dNdScv, DriverNet, HotMAPS, IntOGen, OncodriveCLUSTL, OncodriveFML and consensus results from (7), applied separately to each cancer cohort of TCGA PanCanAtlas. We will refer to this analysis as "Cohorts combined". Applying algorithms to individual cohorts allows to discover cancer type-specific drivers and avoid "contamination" by false positives, i.e. driver genes discovered during Pancan analysis that do not in reality play any role in a given cancer type. On the other hand, much fewer patients are available for cohort-specific analysis, and this decreases statistical power to discover new driver genes. Of note, our SNADRIF algorithm works best for Pancan analysis and struggles with small cohorts, due to scarcity of point mutations. However, when a combination of driver prediction algorithms is used, there are lower chances of missing an important driver gene even in a cohort-specific analysis, as algorithms based on differing principles complement each other. The results of "Cohorts combined" analysis would be presented in the following paragraphs.

We calculated the number of various types of driver events in individual genes, chromosome arms or full chromosomes for each cancer type, tumor stage, age group, as well as for patients with each total number of driver events from 1 to 50. We performed the analyses for total population and for males and females separately, and, for each group, plotted the histograms of top 10 driver events in each class and overall (for data and graphs see **Supplementary Data 3**). In **Fig 2** we present the overall ranking of genes for all TCGA PanCanAtlas cohorts combined. It can be seen that *PIK3CA* is the oncogene with the highest number of SNAs, as well as the highest number of simultaneous occurrences of SNAs and gene amplifications. *MYC* is the oncogene with the highest number of amplifications. *TP53* is the tumor suppressor with the highest number of SNAs, as well as the highest number of instances of simultaneous occurrences of an SNA in one allele and a deletion in the other allele. It is also the top mutated gene when driver events of all classes are counted. *CDKN2A* is the tumor suppressor with the highest number of deletions. Losses of chromosomes 13 and 22 are the most frequent cancer-promoting chromosome losses, whereas gains of chromosomes 7 and 20 are the most frequent cancer-promoting chromosome gains. Losses of 8p and 17p arms are the most frequent cancer-promoting chromosome arm losses, whereas gains of 1q and 8q arms are the most frequent cancer-promoting chromosome arm gains. Overall, these results are expected and confirm that our analytic pipelines work as they should.

Next, we calculated the Driver Strength Index (DSI)

$$DSI_A = \sum_{i=1}^{100} \frac{p_{A\,i}}{i\,p_i}$$

where $p_{A\,i}$ is a number of patients with a driver event in the gene/chromosome *A* amongst patients with *i* driver events in total; p_i is a number of patients with *i* driver events in total. Surprisingly, we do not see much change compared to the simple frequency-of-mutation

approach (**Fig 3**). The only dramatic difference is that *BRAF* became the top SNA-based (and overall) oncogene according to DSI, whereas *PIK3CA* dropped to the second place, lagging behind by a wide margin. Also, *PIK3CA* overtook *MYC* as the top CNA-based oncogene, and *PTEN* displaced *CDKN2A* from the top CNA-based tumor suppressor spot. Moreover, members of several gene families appeared in the top 10 lists, such as *KRAS*, *NRAS* and *HRAS* in the SNA-based oncogenic events list, or lysine methyltransferases *KMT2C* and *KMT2D* in the SNA-based tumor suppressor events list. This indicates that our approach is indeed meaningfully selecting for some biological attributes, which are not selected by simple frequency sorting. Finally, multiple small changes of ranking positions happened, nevertheless not affecting the overall picture. We think the reason for the limited effect of changes is that DSI is still very much affected by the overall frequency of gene mutation, due to $\frac{p_{A\,i}}{p_i}$ component. Hence, to uncover the true driver strength unrelated to the mutation frequency, further normalization is required. Therefore, we propose Normalized Driver Strength Index (NDSI)

$$NDSI_A = \frac{\sum_{i=1}^{100} \frac{p_{A\,i}}{i\,p_i}}{\sum_{i=1}^{100} \frac{p_{A\,i}}{p_i}}$$

that corrects for the effects of mutation frequencies. As can be seen in **Fig 4**, this time the rankings are completely different from both DSI and frequency-based approaches. *GTF2I* conquers the top spot amongst SNA-based oncogenes and overall, *EPRS* becomes number one CNA-based oncogene, and *PDGFRA* occupies the first line of mixed oncogene rating. *NUP214* and *CHEK2* become the top SNA- and CNA-based tumor suppressors, respectively. *FUBP1* and *CIC* occupy the first and the second places in the mixed tumor suppressors list, with minimal difference from each other but big difference from the third place. NDSI reveals the loss of chromosome 1 as the strongest cancer-promoting chromosome loss, whereas gains of chromosomes 17 and 11 as the strongest cancer-promoting chromosome gains. NDSI shows

that the loss of 19q arm is the strongest cancer-promoting chromosome arm loss, whereas gains of 19p and 17p arms are the strongest cancer-promoting chromosome arm gains.

Like DSI, NDSI is able to select for specific gene families. Three members of the guanine nucleotide-binding protein family, *GNAQ*, *GNB3* and *GNAI1*, appeared on the top 10 SNA- and CNA-based oncogenic events lists (**Fig 4**). Additionally, two more G proteins, *GNA11* and *GNAZ*, appeared on the top 100 NDSI-ranked driver list (**Table 1**). Of note, only *GNAS* member of this family is present on the top 100 DSI-ranked driver list (**Table 1**). Two members of isocitrate dehydrogenase family, *IDH1* and *IDH2*, appeared on the top 10 SNA-based oncogenic events list, whereas fibroblast growth factor receptors *FGFR2* and *FGFR3* appeared on the top 10 mixed oncogenic events list (**Fig 4**). Moreover, the strongest SNA-based oncogene, *GTF2I*, and the second strongest CNA-based tumor suppressor, *GTF2F2*, belong to the General Transcription Factor II family. The ability of NDSI to prioritize members of specific protein families suggests that this index has actual biological meaning.

Next, we wanted to analyze top DSI- and NDSI-ranked genes using several common gene list analysis tools. To this aim, we combined the lists of drivers from various classes. If the same gene was affected by more than one kind of alteration, we chose the alteration type with the highest (N)DSI. Also, we removed the data on chromosome arms and full chromosomes, as external pathway and network analysis tools can work only with genes. Then, we selected top 100 DSI- and NDSI-ranked genes. The resulting lists can be seen in **Table 1**.

Table 1. Top 100 DSI- and NDSI-ranked genes.

Rank	Entrez ID	Symbol	DSI	Entrez ID	Symbol	NDSI
1	7157	TP53	1.4145	100093631	GTF2I	0.52439
2	673	BRAF	1.34253	673	BRAF	0.30477
3	5290	PIK3CA	0.96597	11200	CHEK2	0.27057
4	5728	PTEN	0.91662	2776	GNAQ	0.2188
5	1029	CDKN2A	0.63645	80142	GBF1	0.181
6	8289	ARID1A	0.63555	2963	GTF2F2	0.16995
7	3845	KRAS	0.6198	7812	CSDE1	0.16207
8	5925	RB1	0.54114	10010	TRAF2	0.15917
9	1956	EGFR	0.51318	2058	EPRS	0.15088
10	4609	MYC	0.51083	8021	NUP214	0.14232
11	5747	PTK2	0.48049	3417	IDH1	0.13883
12	546	ATRX	0.42089	84376	HK3	0.13671
13	55294	FBXW7	0.40372	8880	FUBP1	0.13633
14	4089	SMAD4	0.39892	3265	HRAS	0.13632
15	5624	APC	0.38758	7249	TSC2	0.13619
16	3417	IDH1	0.38126	3707	ITPKB	0.1345
17	5437	POLR2H	0.37672	3066	HDAC2	0.13328
18	6389	SDHA	0.3739	23152	CIC	0.13091
19	23236	PLCB1	0.35215	9203	ZMYM3	0.13029
20	7037	TFRC	0.33698	1788	DNMT3A	0.12791
21	1387	CREBBP	0.33558	10905	MAN1A2	0.12728
22	472	ATM	0.33481	9997	TYMP	0.12528
23	22916	NCBP2	0.33332	1031	CDKN2C	0.12014
24	7276	TTN	0.33319	3418	IDH2	0.11748
25	108	ADCY2	0.33029	2588	GALNS	0.11309
26	595	CCND1	0.32945	6938	TCF12	0.11206
27	1962	EHHADH	0.32557	10983	CYC1	0.1117
28	2064	ERBB2	0.31659	2770	GNAI1	0.1113
29	6464	SHC1	0.31471	8803	SUCLA2	0.1085
30	5589	PLD1	0.30528	4048	LTA4H	0.10636
31	8972	MGAM	0.29863	715	C1R	0.10503
32	51606	ATP6V1H	0.29832	862	RUNX1T1	0.10456
33	6262	RYR2	0.29226	2534	FYN	0.10427
34	6035	RAC1	0.29067	2784	GNB3	0.1024
35	5295	PIK3R1	0.29043	2194	FAS	0.10225
36	2195	FAT1	0.28314	2885	GRB2	0.09907
37	2033	EP300	0.27801	25836	NIPBL	0.09905
38	4780	NFE2L2	0.27606	6610	SMPD2	0.09835
39	4893	NRAS	0.26471	2182	ACSL4	0.09583
40	777	CACNA1E	0.25906	8517	IKBKG	0.0942
41	10401	PIAS3	0.25776	114788	CSMD3	0.09379

42	8831	RASA1	0.24882	55626	AMBRA1	0.09318
43	3310	HSPA6	0.24837	23533	PIK3R5	0.09061
44	58508	KMT2C	0.24833	7124	TNF	0.09041
45	1499	CTNNB1	0.24444	3312	HSPA8	0.0902
46	4763	NF1	0.24206	3158	HMGCS2	0.08886
47	5105	PCK1	0.23991	51366	UBR5	0.08767
48	55193	PBRM1	0.23983	26137	ZBTB20	0.08748
49	2065	ERBB3	0.23766	25913	POT1	0.0871
50	9631	NUP155	0.22488	29072	SETD2	0.0854
51	5586	PAK2	0.20337	9869	SETDB1	0.08479
52	1894	ECT2	0.20199	8731	MET	0.08452
53	8731	MET	0.20067	7428	VHL	0.08432
54	2157	F8	0.19857	1027	CDKN1B	0.08373
55	3265	HRAS	0.19731	91851	CHRDL1	0.08354
56	9997	TYMP	0.19479	9968	MED12	0.08243
57	1857	DVL3	0.19285	170261	ZCCHC12	0.08231
58	114	ADCY8	0.19272	4893	NRAS	0.08131
59	1131	CHRM3	0.18925	6655	SOS2	0.08067
60	5287	PIK3C2B	0.18911	5580	PRKCD	0.08014
61	1213	CLTC	0.18867	4771	NF2	0.0797
62	9223	BAP1	0.1883	2767	GNA11	0.07937
63	8085	KMT2D	0.18468	55958	KLHL9	0.07922
64	58508	MLL3	0.18319	8242	KDM5C	0.07729
65	196528	ARID2	0.18285	8662	EIF3B	0.07701
66	2778	GNAS	0.17942	8731	RNMT	0.07683
67	6490	SI	0.17815	7410	VAV2	0.07666
68	54965	PIGX	0.17814	567	B2M	0.07487
69	11059	WWP1	0.17722	5093	PCBP1	0.07432
70	197257	DLD	0.17325	4629	MYH11	0.07414
71	1978	EIF4EBP1	0.1725	154	ADRB2	0.07413
72	2963	GTF2F2	0.17073	5156	PDGFRA	0.074
73	6000	RGS7	0.16933	1964	EIF1AX	0.07252
74	8733	GPAA1	0.16928	5050	PAFAH1B3	0.0723
75	6326	SCN2A	0.16847	290	ANPEP	0.07069
76	31	ACACA	0.16824	3683	ITGAL	0.06978
77	29072	SETD2	0.16764	4615	MYD88	0.06945
78	9817	KEAP1	0.16709	6173	RPL36A	0.06936
79	5294	PIK3CG	0.16659	537	ATP6AP1	0.06874
80	8394	PIP5K1A	0.16633	51187	RPL24	0.06844
81	5594	MAPK1	0.1628	2781	GNAZ	0.06793
82	1589	CPS1	0.16218	5430	POLR2A	0.0675
83	9091	PIGQ	0.16007	3792	KEL	0.06746
84	5624	PROC	0.15766	6092	ROBO2	0.06695
85	11200	CHEK2	0.15505	8309	ACOX2	0.0669

86	54880	BCOR	0.15462	6598	SMARCB1	0.06668
87	2194	FASN	0.15432	286530	P2RY8	0.06656
88	7095	SEC62	0.15428	8818	DPM2	0.06628
89	2909	ARHGAP35	0.15427	207	AKT1	0.06578
90	9757	MLL2	0.15194	2932	GSK3B	0.06497
91	57492	ARID1B	0.15193	7114	TMSB4X	0.06444
92	5335	PLCG1	0.1504	7248	TSC1	0.06421
93	9939	RBM8A	0.15019	1594	CYP27B1	0.06415
94	3320	HSP90AA1	0.14937	5727	PTCH1	0.06408
95	5332	PLCB4	0.14864	6921	TCEB1	0.06384
96	3551	IKBKB	0.14798	26047	CNTNAP2	0.0638
97	5313	PKLR	0.14727	23291	FBXW11	0.06362
98	5885	RAD21	0.14565	51755	CDK12	0.06356
99	730249	CAD	0.14039	6778	STAT6	0.06351
100	6531	SLC6A3	0.14035	9223	BAP1	0.06346

First, we uploaded the resulting lists to “Reactome v76 Analyse gene list” tool and studied affected Reactome pathways on Voronoi visualizations (Reacfoam). It can be seen in **Fig 5** that top 100 DSI-ranked genes are significantly overrepresented in such categories as Signal transduction, Diseases of signal transduction by growth factor receptors and second messengers, Chromatin organization, RNA polymerase II transcription, Cell Cycle, Diseases of mitotic cell cycle, Cellular responses to stress, Diseases of cellular response to stress, Programmed cell death, Developmental biology, and even Immune system and Hemostasis. Interestingly, several large categories often deemed important for cancer – such as Metabolism, Autophagy, DNA replication and DNA repair – are not affected. Top 100 NDSI-ranked genes are significantly overrepresented in even fewer categories (**Fig 6**) – Signal transduction, Diseases of signal transduction by growth factor receptors and second messengers, Chromatin organization, RNA polymerase II transcription, Cell Cycle, Metabolism of proteins and Metabolism of RNA, as well as Immune system, Infectious disease, Hemostasis, and Axon guidance.

Next, we uploaded the resulting lists to "KEGG Mapper –Color" tool and mapped them to "Pathways in cancer - Homo sapiens (human)" (hsa05200) KEGG pathway map. **Fig 7** and **Table 1** together suggest that top DSI-ranked genes comprise EGFR/ERBB2/PLCB1/PLCB4/PLCG1-KRAS/NRAS/HRAS-BRAF-MAPK1-MYC-CCND1 pathway, PTK2-PIK3CA/PIK3C2B/PIK3CG-PTEN-IKBKB-CCND1 pathway, GNAS-ADCY2/ADCY8-DVL3-CTNNB1-MYC-CCND1 pathway, KEAP1-NFE2L2 pathway, as well as CDKN2A-TP53-CCND1-RB1 pathway. **Fig 8** and **Table 1** together suggest that top NDSI-ranked genes comprise PDGFRA-GRB2-SOS2-HRAS/NRAS-BRAF pathway, GNAQ/GNA11-HRAS/NRAS-BRAF pathway, GNB3-AKT1-IKBKG/GSK3B/CDKN1B pathway, and TCEB1-VHL pathway.

Finally, we analyzed the data in Cytoscape 3.8.2. We imported BioGRID: Protein-Protein Interactions (H. sapiens) network, appended (N)DSI values from the top 100 (N)DSI-ranked driver list, and mapped node color to (N)DSI values, whereas node size to the degree of connectedness. **Fig 9** shows that although *MYC* and *EGFR* are the biggest hubs of top-DSI-ranked gene network, their DSI values are much lower than those of *BRAF* and *PIK3CA,* which have much less connections. Notably, *TP53* exhibited both high DSI value and high connectedness. Similarly, **Fig 10** shows that although *VHL, AKT1* and *HSPA8* are the biggest, centrally located hubs of top-NDSI-ranked gene network, their NDSI values are much lower than those of *GTF2I*, *BRAF* and *CHEK2*, located on the periphery of the network. This supports our initially proposed idea that network centrality does not equal driver strength. Of note, top-NDSI-ranked gene network appears to have much less edges than top-DSI-ranked gene network.

Discussion

DSI places *BRAF* on the top spot amongst SNA-based oncogenes and drivers of all classes. *BRAF* encodes a protein belonging to the RAF family of serine/threonine protein kinases. *BRAF* plays a role in regulating the MAP kinase/ERK signaling pathway, which affects cell division and differentiation. Mutations in *BRAF*, most commonly the V600E mutation, are the most frequently identified cancer-causing mutations in melanoma, and have been identified in various other cancers as well, including non-Hodgkin lymphoma, colorectal cancer, thyroid carcinoma, non-small cell lung carcinoma, hairy cell leukemia and adenocarcinoma of lung (15). Our analysis shows frequent mutations and amplifications of *BRAF* in TCGA COAD, GBM, KIRP, LGG, LUAD, LUSC, PRAD, SKCM, THCA and UCEC cohorts. From recent studies, a new classification system is emerging for *BRAF* mutations based on biochemical and signaling mechanisms associated with these mutants. Class I *BRAF* mutations affect amino acid V600 and lead to BRAF protein signaling as RAS-independent active monomer, class II mutations make BRAF proteins function as RAS-independent activated dimers, and class III mutations impair BRAF kinase activity but increase signaling through the MAPK pathway due to enhanced RAS binding and subsequent CRAF activation (16). It would be interesting to rate the strength of these *BRAF* mutation classes using NDSI.

DSI prioritizes *KRAS*, *NRAS* and *HRAS* amongst SNA-based oncogenes. These genes belong to the RAS oncogene family, whose members are related to the transforming genes of mammalian sarcoma retroviruses. The products encoded by these genes function in signal transduction pathways. These proteins can bind GTP and GDP, and they have intrinsic GTPase activity. Mutations in the RAS family of proteins have frequently been observed across cancer types, including lung adenocarcinoma, ductal carcinoma of the pancreas, colorectal carcinoma,

follicular thyroid cancer, juvenile myelomonocytic leukemia, bladder cancer, and oral squamous cell carcinoma (17). Our analysis shows frequent mutations and amplifications of *KRAS*, *NRAS* and *HRAS* in TCGA BLCA, BRCA, CESC, CHOL, COAD, ESCA, HNSC, KIRP, LGG, LIHC, LUAD, LUSC, OV, PAAD, PRAD, READ, SKCM, STAD, TGCT, THCA, THYM, UCEC and UCS cohorts. Gain-of-function missense mutations, mostly located at codons 12, 13, and 61, constitutively activate RAS proteins, however, each isoform exhibits distinctive mutation frequency at each codon, supporting the hypothesis that different RAS mutants may lead to distinct biologic manifestations (18).

DSI prioritizes *KMT2C* and *KMT2D* amongst SNA-based tumor suppressors. The proteins encoded by these genes are histone methyltransferases that methylate the Lys-4 position of histone H3 (H3K4me). H3K4me represents a specific tag for epigenetic transcriptional activation. The encoded proteins are part of a large protein complex called ASCOM, which is a transcriptional regulator of the beta-globin and estrogen receptor genes. Whereas *KMT2C* loss disrupts estrogen-driven proliferation, it conversely promotes tumor outgrowth under hormone-depleted conditions (19). In accordance, *KMT2C* is one of the most frequently mutated genes in estrogen receptor-positive breast cancer with *KMT2C* deletion correlating with significantly shorter progression-free survival on anti-estrogen therapy (19). *KMT2D* is among the most highly inactivated epigenetic modifiers in lung cancer (20). Recently, it has been shown that *KMT2D* loss widely impairs epigenomic signals for super-enhancers/enhancers, including the super-enhancer for the circadian rhythm repressor *PER2* (20). Loss of *KMT2D* decreases expression of *PER2*, leading to increase in glycolytic gene expression (20). The role of *KMT2C* and *KMT2D* in cancer has been recently reviewed (21). Our analysis shows frequent mutations and deletions of *KMT2C* and *KMT2D* in TCGA BLCA, BRCA, CESC, COAD, DLBC, ESCA, HNSC, KIRC, KIRP, LIHC, LUSC, PRAD, STAD and UCEC cohorts.

NDSI places *GTF2I* on the top spot both amongst the strongest SNA-based oncogenes and amongst the strongest drivers averaged across all classes. The encoded protein binds to the initiator element (Inr) and E-box element in promoters and functions as a regulator of transcription. *GTF2I* c.74146970 T>A mutation was detected in 82% of type A and 74% of type AB thymomas (22). *GTF2I* β and δ isoforms are expressed in thymomas, and both mutant isoforms are able to stimulate cell proliferation *in vitro* (22). Recently, it has been shown that expression of mutant *GTF2I* alters the transcriptome of normal thymic epithelial cells and upregulates several oncogenic genes (23). *GTF2I* L424H knockin cells exhibit cell transformation, aneuploidy, and increased tumor growth and survival under glucose deprivation or DNA damage (23). Our analysis also shows frequent mutations of *GTF2I* in TCGA THYM (thymoma) cohort. *GTF2I* has been recently named gene of the month and its role in cancer reviewed (24).

Interestingly, the second strongest CNA-based tumor suppressor, *GTF2F2*, belongs to the same General Transcription Factor II family as *GTF2I*. GTF2F2 is a general transcription initiation factor that binds to RNA polymerase II and helps to recruit it to the initiation complex in collaboration with GTF2B. It promotes transcription elongation. GTF2F2 shows ATP-dependent DNA-helicase activity. Our analysis shows frequent deletions of *GTF2F2* in TCGA PRAD (Prostate adenocarcinoma) cohort. Indeed, *GTF2F2* has been shown to be deleted in 20% of prostate cancer patients (25). Interestingly, *GTF2F2* deletions were significantly more frequent in prostate cancers that progressed to metastases than in nonprogressors (26).

EPRS is revealed by NDSI as the strongest CNA-based oncogene and the 8th strongest driver averaged across all classes. The protein encoded by *EPRS* is a multifunctional aminoacyl-tRNA

synthetase that catalyzes the aminoacylation of glutamic acid and proline tRNA species. *EPRS* is upregulated in estrogen receptor positive (ER+) human breast tumors in the TCGA and METABRIC cohorts, with copy number gains in nearly 50% of samples in both datasets, and this overexpression is associated with reduced overall survival of patients (27). Transcriptomic profiling showed that *EPRS* regulates cell cycle and estrogen response genes (27). EPRS is selectively carbonylated in breast tumor tissue compared to matched adjacent healthy tissue (28). *EPRS* is a key upregulated-hypomethylated gene in breast cancer and contributes to significant unfavorable clinical outcome (29). Recently, it has been shown that *EPRS* is frequently overexpressed in gastric cancer tissues compared to the adjacent controls and its overexpression predicts poor survival (30). Mechanistically, EPRS directly binds with SCYL2 to enhance the activation of WNT/GSK-3β/β-catenin signaling pathway and the accumulation of β-catenin in the nucleus, leading to gastric cancer cell proliferation and tumor growth (30). Our analysis shows frequent amplifications of *EPRS* in TCGA CHOL (cholangiocarcinoma) and THCA (thyroid carcinoma) cohorts.

PDGFRA is the strongest mixed (SNA+CNA) oncogene according to NDSI rating. *PDGFRA* encodes a cell surface tyrosine kinase receptor for members of the platelet-derived growth factor family. These growth factors are mitogens for cells of mesenchymal origin. *PDGFRA* plays a role in organ development, wound healing, and tumor progression. Mutations in *PDGFRA* have been associated with somatic and familial gastrointestinal stromal tumors and a variety of other cancers (31) (32). Amplification of the chromosome 4 segment harboring the three receptor tyrosine kinases *KIT*, *PDGFRA*, and *KDR* (4q12amp) is frequent in glioblastomas, angiosarcomas, and osteosarcomas (33). Among 99 pulmonary adenocarcinoma cases harboring 4q12amp, 50 lacked any other known driver (33). Our analysis shows frequent

mutations and amplifications of *PDGFRA* in TCGA GBM (glioblastoma) and LGG (lower grade glioma) cohorts.

NUP214 is ranked by NDSI as the strongest SNA-based tumor suppressor. *NUP214* is a member of the FG-repeat-containing nucleoporins. The protein encoded by *NUP214* is localized to the cytoplasmic face of the nuclear pore complex where it is required for proper cell cycle progression and nucleocytoplasmic transport. Chromosomal translocations involving the *NUP214* locus are recurrent in acute leukemia and frequently fuse the C-terminal region of *NUP214* with *SET* and *DEK*, two chromatin remodeling proteins with roles in transcription regulation (34). *SET-NUP214* and *DEK-NUP214* fusion proteins disrupt protein nuclear export by inhibition of the nuclear export receptor *CRM1*, which results in the aberrant accumulation of *CRM1* protein cargoes in the nucleus (34). *SET-NUP214* is primarily associated with acute lymphoblastic leukemia, whereas *DEK-NUP214* exclusively results in acute myeloid leukemia, indicating different leukemogenic driver mechanisms (34). *NUP214* downregulation elevates mitotic indices, delays degradation of mitotic marker proteins cyclinB1 and cyclinA and dephosphorylation of H3 and enhances chromosomal abnormalities (35). Although classically, majority of studies have shown oncogenic roles of nucleoporins as genetic fusion partners in several types of leukemia, emerging evidence suggests that nucleoporins also modulate many cellular signaling pathways that are associated with several major non-hematological malignancies, such as carcinomas of skin, breast, lung, prostate and colon (36). Our analysis shows frequent mutations and deletions of *NUP214* in TCGA KIRP (kidney renal papillary cell carcinoma), LIHC (liver hepatocellular carcinoma) and THCA (thyroid carcinoma) cohorts.

CHEK2 is ranked by NDSI as the strongest CNA-based tumor suppressor and the second strongest driver averaged across all classes. *CHEK2* is a cell cycle checkpoint regulator and

putative tumor suppressor. CHEK2 contains a forkhead-associated protein interaction domain essential for activation in response to DNA damage and is rapidly phosphorylated in response to replication blocks and DNA damage. When activated, CHEK2 inhibits CDC25C phosphatase, preventing entry into mitosis, and stabilizes the tumor suppressor protein p53, leading to cell cycle arrest in G1. In addition, CHEK2 interacts with and phosphorylates BRCA1, allowing BRCA1 to restore survival after DNA damage. *CHEK2* mutations rank among the most frequent germline alterations revealed by germline genetic testing for various hereditary cancer predispositions (37), including breast (38), prostate (39) and thyroid cancers (40,41). Our analysis shows frequent deletions of *CHEK2* in TCGA THCA (thyroid carcinoma) cohort.

FUBP1 is ranked by NDSI as the strongest mixed tumor suppressor and the 7th strongest CNA-based tumor suppressor. The protein encoded by *FUBP1* is a single stranded DNA-binding protein that binds to multiple DNA elements. FUBP1 is also thought to bind RNA, and contains 3'-5' helicase activity with *in vitro* activity on both DNA-DNA and RNA-RNA duplexes. *FUBP1* cooperates with other tumor suppressor genes to transform mammary epithelial cells by disrupting cellular differentiation and tissue architecture (42). Mechanistically, *FUBP1* participates in regulating N6-methyladenosine (m6A) RNA methylation, and its loss leads to global changes in RNA splicing and widespread expression of aberrant driver isoforms (42). Interestingly, *FUBP1* mutations tend to co-occur with *CIC* mutations and 1p and 19q chromosome arm losses in oligodendrogliomas and oligoastrocytomas (43)(44). 1p is the location of *FUBP1* and 19q is the location of *CIC*. Our analysis also shows frequent instances of simultaneous mutation and hemizygous deletion of *FUBP1* in TCGA LGG (lower grade glioma) cohort.

CIC is ranked by NDSI as the second strongest mixed tumor suppressor and the 6th strongest SNA-based tumor suppressor. *CIC* is an ortholog of the *Drosophila melanogaster* Capicua gene, and is a member of the high mobility group (HMG)-box superfamily of transcriptional repressors. Mutations in *CIC* have been associated with olidogdendrogliomas (43). Our analysis also shows frequent mutations and hemizygous deletions of *CIC* in TCGA LGG (lower grade glioma) cohort. In addition, translocation events resulting in gene fusions of *CIC* with *DUX4* have been associated with round cell sarcomas (45). Inactivation of *CIC* relieves repression of its effector ETV4, driving ETV4-mediated upregulation of MMP24, which is necessary and sufficient for metastasis (46). Loss of *CIC*, or an increase in levels of its effectors ETV4 and MMP24, is a biomarker of tumor progression and worse outcomes in patients with lung and/or gastric cancer (46). Loss of *CIC* leads to overexpression of downstream members of the MAPK signaling cascade (47) via increased histone acetylation (48). Recently, it has been shown that *CIC* deficiency critically enhances cancer stem cell self-renewal without altering their growth rate or invasiveness (49). Loss of *CIC* relieves repression of ETV4 and ETV5 expression, consequently promoting self-renewal capability, EpCAM^{+}/CD44^{+}/CD24$^{low/-}$ expression, and ALDH activity (49). In xenograft models, *CIC* deficiency significantly increases cancer stem cell frequency and drives tumor initiation through derepression of ETV4 (49). Consistent with the experimental data, the CD44^{high}/CD24^{low} cancer stem cell-like feature is inversely correlated with CIC levels in breast cancer patients (49). *SOX2* is a downstream target gene of *CIC* that partly promotes cancer stem cells properties (49). The emerging role of *CIC* in cancer has been recently reviewed (50) (51).

NDSI reveals loss of chromosome 1 as the strongest cancer-promoting chromosome loss. Loss of chromosome 1 is associated specifically with the development of hepatic metastases in patients with sporadic pancreatic endocrine tumors (52). Loss of heterozygosity on

chromosome 1 at one or more loci was detected in 93% of cervical carcinomas (53). Allelic losses were observed in 46% of male germ cell tumor cases on 1p and in 23% of cases on 1q (54). NDSI reveals gains of chromosomes 11 and 17 as the strongest cancer-promoting chromosome gains. Duplication/amplification of chromosome 11 was found by cytogenetic methods in 10 of 119 newly diagnosed patients with acute myeloid leukemia (55). Trisomy of chromosome 17 is characteristic for papillary renal cell tumors (56)(57). Recently, it has been suggested that trisomy/polysomy of chromosome 17 could be a marker of worse prognosis of oral squamous cell carcinoma (58).

NDSI shows that the loss of 19q arm is the strongest cancer-promoting chromosome arm loss. Allelic losses on 19q were found in astrocytomas, glioblastomas, oligodendrogliomas and oligoastrocytomas (59). 1p/19q co-deletion is a classical marker of anaplastic oligodendroglial tumors (60). Interestingly, 1p/19q loss in oligodendrogliomas has been found to co-occur with *CIC* and *IDH1/2* mutations (61). *CIC* is located on 19q. Recently, 1p/19q co-deletion and *CIC* mutation have been discovered as characteristic features of *CDKN2C*-null leiomyosarcoma (62). NDSI shows that gains of 17p and 19p arms are the strongest cancer-promoting chromosome arm gains. 17p gain is a distinguishing feature of type 1 papillary renal cell carcinomas (63). 19p gain is associated with malignant pheochromocytomas (64).

Interestingly, NDSI prioritized five members of guanine nucleotide-binding protein (G protein) family: *GNAQ*, *GNA11, GNAI1, GNAZ* and *GNB3*. Guanine nucleotide-binding proteins function as transducers downstream of G protein-coupled receptors (GPCRs) in numerous signaling cascades. The alpha chain contains the guanine nucleotide binding site and alternates between an active, GTP-bound state and an inactive, GDP-bound state. Signaling by an activated GPCR promotes GDP release and GTP binding. The alpha subunit has a low GTPase activity that

converts bound GTP to GDP, thereby terminating the signal. *GNAQ*-encoded protein, an α subunit in the Gq class, couples a seven-transmembrane domain receptor to activation of PLCβ. Some *GNAQ* cancer mutants display normal basal activity and GPCR-mediated activation, but deactivate slowly due to GTPase activating protein (GAP) insensitivity (65). *GNAQ* mutations occur in about half of uveal melanomas, representing the most common known oncogenic mutation in this cancer (66). The presence of this mutation in tumors at all stages of malignant progression suggests that it is an early event in uveal melanoma (66). Mutations affecting Q209 in *GNAQ* were present in 45% of primary uveal melanomas and 22% of uveal melanoma metastases (67). Our analysis also shows frequent mutations of *GNAQ* in TCGA UVM (uveal melanoma) cohort. Recently, of the 11111 patients screened, 117 patients have been found to harbor *GNAQ/GNA11* mutations, in melanoma, colorectal, liver, glioma, lung, bile duct and gastric cancers (68). *GNA11* encodes subunit α-11 and acts as an activator of PLC. Mutations affecting Q209 in *GNA11* were present in 32% of primary uveal melanomas and 57% of uveal melanoma metastases (67). Our analysis also shows frequent mutations of *GNA11* in TCGA UVM (uveal melanoma) cohort. *GNAI1*-encoded protein represents the α subunit of an inhibitory complex that responds to β-adrenergic signals by inhibiting adenylate cyclase, leading to decreased intracellular cAMP levels. *GNAI1 is* required for normal cytokinesis during mitosis (69). High expression and low DNA hypermethylation of *GNAI1* are significantly associated with poor prognosis for gastric cancer patients (70). Our analysis shows frequent amplifications of *GNAI1* in TCGA PRAD (prostate adenocarcinoma) cohort. *GNAZ* encodes G protein subunit αZ. *GNAZ* was mutated in 5% of melanoma patients (71). Our analysis shows frequent amplifications of *GNAZ* in TCGA SARC (sarcoma) cohort. Interestingly, *GNAZ* expression was required for proper classification of leiomyosarcoma subtypes (72). *GNB3* encodes a β subunit which belongs to the WD repeat G protein β family. β subunits are important regulators of α subunits, as well as of certain signal transduction receptors and

effectors. The β and γ chains are required for the GTPase activity, for replacement of GDP by GTP, and for G protein-effector interaction. Polymorphisms in *GNB3* have been implicated in several cancer types, including extrahepatic cholangiocarcinoma (73), transitional cell carcinoma of the bladder (74) and prostate cancer (75). Our analysis shows frequent amplifications of *GNB3* in TCGA LGG (lower grade glioma) cohort. The current knowledge on cancer-associated alterations of GPCRs and G proteins has been recently reviewed (76). Strikingly, approximately 36% of all drugs approved by the US Food and Drug Administration during the past three decades target GPCRs (77). Overall, G proteins were ranked by NDSI as some of the strongest oncogenes likely because they lead to unconditional activation of two most important oncogenic signaling pathways: RAS-ERK and PI3K-AKT.

Two members of isocitrate dehydrogenase family, *IDH1* and *IDH2*, appeared on the top 10 SNA-based oncogenic events list as ranked by NDSI. The protein encoded by *IDH1* is the $NADP^+$-dependent isocitrate dehydrogenase found in the cytoplasm and peroxisomes. The cytoplasmic enzyme serves a significant role in cytoplasmic NADPH production. The protein encoded by *IDH2* is the $NADP^+$-dependent isocitrate dehydrogenase found in the mitochondria. It plays a role in intermediary metabolism and energy production. The most frequent mutations R132 (*IDH1*) and R172 (*IDH2*) involve the active site and result in simultaneous loss of normal catalytic activity, the production of α-ketoglutarate, and gain of a new function, the production of 2-hydroxyglutarate (78) (79) (80) (81). 2-hydroxyglutarate is structurally similar to α-ketoglutarate, and acts as an α-ketoglutarate antagonist to competitively inhibit multiple α-ketoglutarate–dependent dioxygenases, including both lysine histone demethylases and the ten-eleven translocation family of DNA hydroxylases (81). Abnormal histone and DNA methylation are emerging as a common feature of tumors with *IDH1* and *IDH2* mutations and may cause altered stem cell differentiation and eventual tumorigenesis (81). In acute myeloid

leukemia, *IDH1* and *IDH2* mutations have been associated with worse outcome, shorter overall survival, and normal karyotype (82). All the 1p19q co-deleted gliomas are mutated on *IDH1* or *IDH2* (83). Our analysis shows frequent mutations of *IDH1* and *IDH2* in TCGA LGG (lower grade glioma) cohort and frequent amplifications of *IDH1* in LIHC (liver hepatocellular carcinoma), as well as less frequent mutations and amplifications of *IDH1* in CHOL, GBM, PRAD and SKCM.

Two fibroblast growth factor receptors *FGFR2* and *FGFR3* appeared on the top 10 mixed oncogenic events list as ranked by NDSI. The extracellular region of these proteins, composed of three immunoglobulin-like domains, interacts with fibroblast growth factors, leading to the activation of a cytoplasmic tyrosine kinase domain that phosphorylates PLCG1, FRS2 and other proteins. This sets in motion a cascade of downstream signals, including RAS-MAPK and PI3K-AKT pathways, ultimately influencing cell proliferation, differentiation, migration and apoptosis. FGFR aberrations were found in 7.1% of cancers, with the majority being gene amplification (66% of the aberrations), followed by mutations (26%) and rearrangements (8%) (84). FGFR1 was affected in 3.5% of 4,853 patients; FGFR2 in 1.5%; FGFR3 in 2.0%; and FGFR4 in 0.5% (84). The cancers most commonly affected were urothelial (32% FGFR-aberrant); breast (18%); endometrial (~13%), squamous lung cancers (~13%), and ovarian cancer (~9%) (84). Our analysis also shows frequent mutations and amplifications of *FGFR2* in TCGA BRCA (breast invasive carcinoma), LUSC (lung squamous cell carcinoma) and UCEC (uterine corpus endometrial carcinoma) cohorts, as well as frequent mutations and amplifications of *FGFR3* in BLCA (bladder urothelial carcinoma) and HNSC (head and neck squamous cell carcinoma) cohorts.

It could be noticed that NDSI prioritized *CIC*, *FUBP1*, *IDH1* and *IDH2* mutations, as well as 19q and 1p chromosome arm losses, that comprise characteristic molecular alterations of gliomas

(43)(44)(59)(60)(61)(83). This could be at least partially explained by our results showing that glioma patients typically have no more than 30 driver mutations per tumor, with the average of 10. This is in contrast with the maximum of 75 and the average of 22 for all cancer types combined (see **Supplementary Data 3**). As NDSI prioritizes genes preferentially occurring in patients with few mutations per tumor, it gives high scores to typical glioma drivers. This, however, does not indicate any pitfall of the index. On the contrary, it validates that the index works as expected. Similar effect can be seen with G proteins. *GNAQ* and *GNA11* mutations are characteristic of uveal melanoma (66)(67), *GNAI1* is amplified in prostate cancer, *GNAZ* in sarcoma and *GNB3* in glioma. All these cancer types have low average number of driver events per patient. The observation that mutations of G proteins are overrepresented in driver-sparse cancer types confirms their high ranking by NDSI.

While both DSI- and NDSI-ranked top 100 genes are significantly overrepresented in such Reactome categories as Signal transduction, Diseases of signal transduction by growth factor receptors and second messengers, Chromatin organization, RNA polymerase II transcription and Cell Cycle, there are also differences. Top 100 DSI-ranked genes are additionally overrepresented in Cellular responses to stress, Diseases of cellular response to stress and Programmed cell death, whereas top NDSI-ranked genes are not. This suggests that although these pathways are frequently mutated in cancer they do not possess strong tumor-promoting activity. It is also peculiar why neither DSI- nor NDSI-ranked top 100 genes are significantly overrepresented in Metabolism, Autophagy, DNA replication and DNA repair categories. This may indicate that the role of these processes in oncogenesis is overestimated.

The major signaling pathway activated by both top DSI- and top NDSI-ranked driver genes is the RAS-RAF pathway. However, these two groups of drivers activate it via different routes. Top

DSI-ranked drivers trigger the pathway via EGFR, ERBB2, PLCB1, PLCB4 or PLCG1, whereas top NDSI-ranked drivers engage it through PDGFRA-GRB2-SOS arm or via G proteins, such as GNAQ or GNA11. This suggests that *EGFR*, *ERBB2*, *PLCB1, PLCB4* and *PLCG1* driver mutations occur more frequently but are weaker than *PDGFRA*, *GRB2*, *SOS*, *GNAQ* and *GNA11* driver mutations. Also, top DSI-ranked drivers appear to engage PI3K-AKT pathway via constitutive PTK2 activation or PTEN inactivation, and lead to IKBKB phosphorylation and CCND1 expression, whereas top NDSI-ranked drivers trigger it via GNB3 and lead to phosphorylation of IKBKG, GSK3B and CDKN1B. Similarly, this suggests that *PTK2, PIK3CA, PIK3C2B, PIK3CG, PTEN, IKBKB* and *CCND1* driver mutations occur more frequently but are weaker than *GNB3, AKT1, IKBKG, GSK3B* and *CDKN1B* driver mutations. Moreover, GNAS-ADCY2/ADCY8-DVL3-CTNNB1-MYC-CCND1 pathway, KEAP1-NFE2L2 pathway, and CDKN2A-TP53-CCND1-RB1 pathway are engaged only by top DSI-ranked drivers, whereas TCEB1-VHL pathway - only by top NDSI-ranked drivers.

Overall, we presented a comprehensive overview on the landscape of cancer driver genes and chromosomes in TCGA PanCanAtlas patients and highlighted particular genes, gene families and pathways deemed strong drivers according to our Normalized Driver Strength Index. A puzzling question that remains in cancer genomics is why mutations in a given driver gene are typically confined to one or a few cancer types, resulting in each cancer type having its own unique set of driver genes (85)? As mutations are supposed to happen randomly as a result of stochastic mutagenesis processes (86), it is logical to suggest that mutations in different tissues can affect the same genes. However, the same mutation can be selected for in some tissues and selected against in others (87). This selection most likely depends on the tissue-specific epigenetic profiles and microenvironments of the cancer-initiating stem or progenitor cells (88,89). Thus, investigating the interplay between stem cell mutations, epigenetic profiles and

microenvironments in various tissues appears to be a promising and exciting avenue for future research.

Methods

Source files and initial filtering

TCGA PanCanAtlas data were used. Files "Analyte level annotations - merged_sample_quality_annotations.tsv", "ABSOLUTE purity/ploidy file - TCGA_mastercalls.abs_tables_JSedit.fixed.txt", "Aneuploidy scores and arm calls file - PANCAN_ArmCallsAndAneuploidyScore_092817.txt", "Public mutation annotation file - mc3.v0.2.8.PUBLIC.maf.gz", "gzipped ISAR-corrected GISTIC2.0 all_thresholded.by_genes file - ISAR_GISTIC.all_thresholded.by_genes.txt", "RNA batch corrected matrix - EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2.geneExp.tsv", "miRNA batch corrected matrix - pancanMiRs_EBadjOnProtocolPlatformWithoutRepsWithUnCorrectMiRs_08_04_16.csv", were downloaded from https://gdc.cancer.gov/about-data/publications/PanCan-CellOfOrigin.

Using TCGA barcodes (see https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA_Barcode/ and https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/sample-type-codes), all samples except primary tumors (barcoded 01, 03, 09) were removed from all files. Based on the information in the column "Do_not_use" in the file "Analyte level annotations - merged_sample_quality_annotations.tsv", all samples with "True" value were removed from all files. All samples with "Cancer DNA fraction" <0.5 or unknown or with "Subclonal genome fraction" >0.5 or unknown in the file "TCGA_mastercalls.abs_tables_JSedit.fixed.txt" were removed from the file "PANCAN_ArmCallsAndAneuploidyScore_092817.txt". Moreover, all samples without "PASS" value in the column "FILTER" were removed from the file

"mc3.v0.2.8.PUBLIC.maf.gz" and zeros in the column "Entrez_Gene_Id" were replaced with actual Entrez gene IDs, determined from the corresponding ENSEMBL gene IDs in the column "Gene" and external database ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia/Homo_sapiens.gene_info.gz. Filtered files were saved as "Primary_whitelisted_arms.tsv", "mc3.v0.2.8.PUBLIC_primary_whitelisted_Entrez.tsv", "ISAR_GISTIC.all_thresholded.by_genes_primary_whitelisted.tsv", "EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2-v2.geneExp_primary_whitelisted.tsv", "pancanMiRs_EBadjOnProtocolPlatformWithoutRepsWithUnCorrectMiRs_08_04_16_primary_whitelisted.tsv".

RNA filtering of CNAs

Using the file "EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2-v2.geneExp_primary_whitelisted.tsv", the median expression level for each gene across patients was determined. If the expression for a given gene in a given patient was below 0.05x median value, it was encoded as "-2", if between 0.05x and 0.75x median value, it was encoded as "-1", if between 1.25x and 1.75x median value, it was encoded as "1", if above 1.75x median value, it was encoded as "2", otherwise it was encoded as "0". The file was saved as "EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2-v2.geneExp_primary_whitelisted_median.tsv." The same operations were performed with the file "pancanMiRs_EBadjOnProtocolPlatformWithoutRepsWithUnCorrectMiRs_08_04_16_primary_whitelisted.tsv", which was saved as "pancanMiRs_EBadjOnProtocolPlatformWithoutRepsWithUnCorrectMiRs_08_04_16_primary_whitelisted_median.tsv"

Next, the file "ISAR_GISTIC.all_thresholded.by_genes_primary_whitelisted.tsv" was processed according to the following rules: if the gene CNA status in a given patient was not zero and had the same sign as the gene expression status in the same patient (file "EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2-v2.geneExp_primary_whitelisted_median.tsv" or "pancanMiRs_EBadjOnProtocolPlatformWithoutRepsWithUnCorrectMiRs_08_04_16_primary_whitelisted_median.tsv" for miRNA genes), then the CNA status value was replaced with the gene expression status value, otherwise it was replaced by zero. If the corresponding expression status for a given gene was not found then its CNA status was not changed. The resulting file was saved as "ISAR_GISTIC.all_thresholded.by_genes_primary_whitelisted_RNAfiltered.tsv"

We named this algorithm GECNAV (Gene Expression-based CNA Validator) and created a Github repository: https://github.com/belikov-av/GECNAV

Aneuploidy driver prediction

Using the file "Primary_whitelisted_arms.tsv", the average alteration status of each chromosomal arm was calculated for each cancer type and saved as a matrix file "Arm_averages.tsv". By drawing statuses randomly with replacement (bootstrapping) from any cell of "Primary_whitelisted_arms.tsv", for each cancer type the number of statuses corresponding to the number of patients in that cancer type were generated and their average was calculated. The procedure was repeated 10000 times, the median for each cancer type was calculated and the results were saved as a matrix file "Bootstrapped_arm_averages.tsv".

P-value for each arm alteration status was calculated for each cancer type. To do this, first the alteration status for a given cancer type and a given arm in "Arm_averages.tsv" was compared to the median bootstrapped arm alteration status for this cancer type in "Bootstrapped_arm_averages.tsv". If the status in "Arm_averages.tsv" was higher than zero and the median in "Bootstrapped_arm_averages.tsv", the number of statuses for this cancer type in "Bootstrapped_arm_averages.tsv" that are higher than the status in "Arm_averages.tsv" was counted and divided by 5000. If the status in "Arm_averages.tsv" was lower than zero and the median in "Bootstrapped_arm_averages.tsv", the number of statuses for this cancer type in "Bootstrapped_arm_averages.tsv" that are lower than the status in "Arm_averages.tsv" was counted and divided by 5000, and marked with minus to indicate arm loss. Other values were ignored (cells left empty). The results were saved as a matrix file "Arm_Pvalues_cohorts.tsv".

For each cancer type, Benjamini–Hochberg procedure with FDR=5% was applied to P-values in "Arm_Pvalues_cohorts.tsv" and passing P-values were encoded as "DAG" (Driver arm gain) or "DAL" (Driver arm loss) if marked with minus. The other cells were made empty and the results were saved as a matrix file "Arm_drivers_FDR5_cohorts.tsv".

Alterations were classified according to the following rules: if the arm status in a given patient (file "Primary_whitelisted_arms.tsv") was "-1" and the average alteration status of a given arm in the same cancer type (file "Arm_drivers_FDR5_cohorts.tsv") was "DAL", then the alteration in the patient was classified as "DAL". If the arm status in a given patient was "1" and the average alteration status of a given arm in the same cancer type was "DAG", then the alteration in the patient was classified as "DAG". In all other cases an empty cell was written. The total

number of DALs and DAGs was calculated, patients with zero drivers were removed, and the results were saved as a matrix file "Arm_drivers_FDR5.tsv".

Using the file "Primary_whitelisted_arms.tsv", the values for the whole chromosomes were calculated using the following rules: if both p- and q-arm statuses were "1" then the chromosome status was written as "1"; if both p- and q-arm statuses were "-1" then the chromosome status was written as "-1"; if at least one arm status was not known (empty cell) then the chromosome status was written as empty cell; in all other cases the chromosome status was written as "0". For one-arm chromosomes (13, 14, 15, 21, 22), their status equals the status of the arm. The resulting file was saved as "Primary_whitelisted_chromosomes.tsv".

The same procedures as described above for chromosomal arms were repeated for the whole chromosomes, with the resulting file "Chromosome_drivers_FDR5.tsv". Chromosome drivers were considered to override arm drivers, so if a chromosome had "DCL" (Driver chromosome loss) or "DCG" (Driver chromosome gain), no alterations were counted on the arm level, to prevent triple counting of the same event.

We named this algorithm ANDRIF (ANeuploidy DRIver Finder) and created a Github repository: https://github.com/belikov-av/ANDRIF

SNA driver prediction

Using the file "mc3.v0.2.8.PUBLIC_primary_whitelisted_Entrez.tsv" all SNAs were classified according to the column "Variant_Classification". "Frame_Shift_Del", "Frame_Shift_Ins", "Nonsense_Mutation", "Nonstop_Mutation" and "Translation_Start_Site" were considered potentially inactivating; "De_novo_Start_InFrame", "In_Frame_Del", "In_Frame_Ins" and

"Missense_Mutation" were considered potentially hyperactivating; "De_novo_Start_OutOfFrame" and "Silent" were considered passengers; the rest were considered unclear. The classification results were saved as the file "SNA_classification_patients.tsv", with columns "Tumor_Sample_Barcode", "Hugo_Symbol", "Entrez_Gene_Id", "Gene", "Number of hyperactivating SNAs", "Number of inactivating SNAs", "Number of SNAs with unclear role", "Number of passenger SNAs".

Using this file, the sum of all alterations in all patients was calculated for each gene. Next, the Nonsynonymous SNA Enrichment Index (NSEI) was calculated for each gene as

$$NSEI = \frac{\textbf{Number of hyperactivating SNAs} + \textbf{Number of inactivating SNAs} + 1}{\textbf{Number of passenger SNAs} + 1}$$

and the Hyperactivating to Inactivating SNA Ratio (HISR) was calculated for each gene as

$$HISR = \frac{\textbf{Number of hyperactivating SNAs} + 1}{\textbf{Number of inactivating SNAs} + 1}$$

saving the results as "SNA_classification_genes_NSEI_HISR.tsv".

Using the file "SNA_classification_patients.tsv", the gene-patient matrix "SNA_matrix.tsv" was constructed, encoding the "Number of hyperactivating SNAs", "Number of inactivating SNAs", "Number of SNAs with unclear role" and "Number of passenger SNAs" as one number separated by dots (e.g. "2.0.1.1"). If data for a given gene were absent in a given patient, it was encoded as "0.0.0.0". By drawing statuses randomly with replacement (bootstrapping) from any cell of "SNA_matrix.tsv" 10000 times for each patient, the matrix file "SNA_matrix_bootstrapped.tsv" was created. The sums of statuses in "SNA_matrix_bootstrapped.tsv" were calculated for each iteration separately, and then the corresponding NSEI and HISR indices were calculated and the results were saved as "SNA_bootstrapped_NSEI_HISR.tsv".

P-value for each gene was calculated as the number of NSEI values in "SNA_bootstrapped_NSEI_HISR.tsv" more extreme than its NSEI value in "SNA_classification_genes_NSEI_HISR.tsv" and divided by 10000. The results were saved as "SNA_classification_genes_NSEI_HISR_Pvalues.tsv". Benjamini–Hochberg procedure with FDR(Q)=5% was applied to P-values in "SNA_classification_genes_NSEI_HISR_Pvalues.tsv", and genes that pass were saved as "SNA_driver_gene_list_FDR5.tsv".

We named this algorithm SNADRIF (SNA DRIver Finder) and created a Github repository: https://github.com/belikov-av/SNADRIF

Driver prediction algorithms and conversion to the patient level

Lists of driver genes and mutations predicted by various algorithms (**Table 2**) applied to PanCanAtlas data were downloaded from https://gdc.cancer.gov/about-data/publications/pancan-driver (2020plus, CompositeDriver, DriverNet, HotMAPS, OncodriveFML), https://karchinlab.github.io/CHASMplus (CHASMplus), as well as received by personal communication from Francisco Martínez-Jiménez, Institute for Research in Biomedicine, Barcelona, francisco.martinez@irbbarcelona.org (dNdScv, IntOGen Plus, OncodriveCLUSTL, OncodriveFML). Additionally, a consensus driver gene list from 26 algorithms applied to PanCanAtlas data (7) was downloaded from https://www.cell.com/cell/fulltext/S0092-8674(18)30237-X. All genes and mutations with q-value > 0.05 were removed. Entrez Gene IDs were identified for each gene using HUGO Symbol and external database ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia/Homo_sapiens.gene_info.gz.

Table 2. Driver prediction algorithms.

Name	Ref.	Repository	Level	Principles
ANDRIF	This paper	https://github.com/belikov-av/ANDRIF	Chromosomal arm, chromosome	Recurrence
20/20plus	(5)	https://github.com/KarchinLab/2020plus	gene	Machine learning, trained on Cancer Genome Landscapes (20/20 rule); Nonsynonymous/Synonymous, clustering, conservation (uses UCSC's 46-way vertebrate alignment and SNVBox), impact (uses VEST), network (uses BioGrid), expression, chromatin, replication (uses MutSigCV)
CHASMplus	(6)	https://github.com/KarchinLab/CHASMplus	mutation	Machine learning, trained on TCGA; clustering (uses HotMAPS 1D), conservation (uses UCSC Multiz-100-way and SNV box), network (uses Interactome Insider)
CompositeDriver	(7)	https://github.com/mi2041/CompositeDriver	gene	Recurrence, impact (uses FunSeq2)
dNdScv	(8)	https://github.com/im3sanger/dndscv	gene	Nonsynonymous/Synonymous
DriverNet	(9)	https://github.com/shahcompbio/drivernet https://bioconductor.org/packages/release/bioc/html/DriverNet.html	gene	Network (uses MGSA and a human functional protein interaction network), impact (uses gene expression outliers)
HotMAPS	(10)	https://github.com/karchinlab/HotMAPS	mutation	3D clustering (uses Protein Data Bank and ModPipe)
IntOGen Plus	(11)	https://www.intogen.org/search https://bitbucket.org/intogen/intogen-plus/src/master/	gene	Combination of dNdScv, CBaSE, OncodriveCLUSTL, HotMAPS, smRegions and OncodriveFML
OncodriveCLUSTL	(12)	http://bbglab.irbbarcelona.org/oncodriveclustl/analysis https://bitbucket.org/bbglab/oncodriveclustl/src/master/	gene	Clustering
OncodriveFML	(13)	http://bbglab.irbbarcelona.org/oncodrivefml/analysis https://bitbucket.org/bbglab/oncodrivefml/src/master/	gene	Recurrence, Impact (uses CADD and RNAsnp)
SNADRIF	This paper	https://github.com/belikov-av/SNADRIF	gene	Nonsynonymous/Synonymous

To convert these population-level data to patient-level data, the following procedures were performed.

For lists of driver *genes*, all entries from the file "mc3.v0.2.8.PUBLIC_primary_whitelisted_Entrez.tsv" were removed except those that satisfied the following conditions simultaneously: "Entrez Gene ID" matches the one in the driver list; cancer type (identified by matching "Tumor_Sample_Barcode" with "bcr_patient_barcode" and "acronym" in "clinical_PANCAN_patient_with_followup.tsv") matches "cohort" in the driver list or the driver list is for pancancer analysis; "Variant_Classification" column contains one of the following values: "De_novo_Start_InFrame", "Frame_Shift_Del", "Frame_Shift_Ins", "In_Frame_Del", "In_Frame_Ins", "Missense_Mutation", "Nonsense_Mutation", "Nonstop_Mutation", "Translation_Start_Site".

For lists of driver *mutations*, the procedures were the same, except that Ensembl Transcript ID and nucleotide/amino acid substitution were used for matching instead of Entrez Gene ID. These data (only columns "TCGA Barcode", "HUGO Symbol", "Entrez Gene ID") were saved as "AlgorithmName_output_SNA.tsv".

Additionally, all entries from the file "ISAR_GISTIC.all_thresholded.by_genes_primary_whitelisted.tsv" were removed except those that satisfied the following conditions simultaneously: "Locus ID" matches "Entrez Gene ID" in the driver list; cancer type (identified by matching Tumor Sample Barcode with "bcr_patient_barcode" and "acronym" in "clinical_PANCAN_patient_with_followup.tsv") matches "cohort" in the driver list or the driver list is for pancancer analysis; CNA values are "2", "1", "-1" or "-2". These data were converted from the matrix to a list format (with columns "TCGA Barcode", "HUGO Symbol", "Entrez Gene ID") and saved as "AlgorithmName_output_CNA.tsv".

Finally, the files "AlgorithmName_output_SNA.tsv" and "AlgorithmName_output_CNA.tsv" were combined, duplicate TCGA Barcode-Entrez Gene ID pairs were removed, and the results saved as "AlgorithmName_output.tsv".

Driver event classification and analysis

The file "Clinical with Follow-up - clinical_PANCAN_patient_with_followup.tsv" was downloaded from https://gdc.cancer.gov/node/905/. All patients with "icd_o_3_histology" different from XXXX/3 (primary malignant neoplasm) were removed, as well as all patients not simultaneously present in the following three files:
"mc3.v0.2.8.PUBLIC_primary_whitelisted_Entrez.tsv",
"ISAR_GISTIC.all_thresholded.by_genes_primary_whitelisted.tsv" and
"Primary_whitelisted_arms.tsv". The resulting file was saved as
"clinical_PANCAN_patient_with_followup_primary_whitelisted.tsv".

Several chosen "AlgorithmName_output.tsv" files were combined and duplicate TCGA Barcode-Entrez Gene ID pairs removed. Entries with TCGA Barcodes not present in "clinical_PANCAN_patient_with_followup_primary_whitelisted.tsv" were removed as well. Matching "Number of hyperactivating SNAs" and "Number of inactivating SNAs" for each TCGA Barcode-Entrez Gene ID pair were taken from the "SNA_classification_patients.tsv" file, in case of no match zeros were written. Matching HISR value was taken from "SNA_classification_genes_NSEI_HISR.tsv" for each Entrez Gene ID, in case of no match empty cell was left. Matching CNA status was taken from "ISAR_GISTIC.all_thresholded.by_genes_primary_whitelisted_RNAfiltered.tsv" for each TCGA Barcode-Entrez Gene ID pair, in case of no match zero was written.

Each TCGA Barcode-Entrez Gene ID pair was classified according to the **Table 3**:

Table 3. Driver event classification rules.

Driver type	Number of nonsynonymous SNAs	Number of inactivating SNAs	HISR	CNA status	Count as ... driver event(s)
SNA-based oncogene	≥1	0	>5	0	1
CNA-based oncogene	0	0	>5	1 or 2	1
Mixed oncogene	≥1	0	>5	1 or 2	1
SNA-based tumor suppressor	≥1	≥0	≤5	0	1
CNA-based tumor suppressor	0	0	≤5	-1 or -2	1
Mixed tumor suppressor	≥1	≥0	≤5	-1 or -2	1
Passenger	0	0		0	0
Low-probability driver	All the rest				0

Results of this classification were saved as "AnalysisName_genes_level2.tsv".

Using this file, the number of driver events of each type was counted for each patient. Information on the number of driver chromosome and arm losses and gains for each patient was taken from the files "Chromosome_drivers_FDR5.tsv" and "Arm_drivers_FDR5.tsv". All patients not present in the files "AnalysisName_genes_level2.tsv", "Chromosome_drivers_FDR5.tsv" and "Arm_drivers_FDR5.tsv", but present in the file "clinical_PANCAN_patient_with_followup_primary_whitelisted.tsv", were added with zero values for the numbers of driver events. Information on the cancer type ("acronym"), gender ("gender"), age ("age_at_initial_pathologic_diagnosis") and tumor stage ("pathologic_stage", if no data then "clinical_stage", if no data then "pathologic_T", if no data then "clinical_T") was taken from the file "clinical_PANCAN_patient_with_followup_primary_whitelisted.tsv". The results were saved as "AnalysisName_patients.tsv".

Using the file "AnalysisName_patients.tsv", the number of patients with each integer total number of driver events from 0 to 100 was counted for each cancer type, also for males and females separately, and histograms were plotted for each cancer type-gender combination. Using the same file "AnalysisName_patients.tsv", the average number of various types of driver events was calculated for each cancer type, tumor stage, age group, as well as for patients with each total number of driver events from 1 to 100. Analyses were performed for total population and for males and females separately, and cumulative histograms were plotted for each file.

We named this algorithm PALDRIC (PAtient-Level DRIver Classifier) and created a Github repository: https://github.com/belikov-av/PALDRIC

We later developed a modification of PALDRIC that allows analysis and ranking of individual genes, chromosome arms and full chromosomes – PALDRIC GENE - and created a Github repository: https://github.com/belikov-av/PALDRIC_GENE

Using the files "AnalysisName_genes_level2.tsv", "Chromosome_drivers_FDR5.tsv" and "Arm_drivers_FDR5.tsv", the names of individual genes, chromosome arms or full chromosomes affected by driver events of each type were catalogued for each patient. Information on the cancer type, gender, age and tumor stage was taken from the file "clinical_PANCAN_patient_with_followup_primary_whitelisted.tsv". The results were saved as "AnalysisName_patients_genes.tsv".

Using the file "AnalysisName_patients_genes.tsv", the number of various types of driver events in individual genes, chromosome arms or full chromosomes was calculated for each cancer type, tumor stage, age group, as well as for patients with each total number of driver events from 1 to 100. Analyses were performed for total population and for males and females separately, and histograms of top 10 driver events in each class and overall were plotted for each group.

Driver Strength Index (DSI)

$$DSI_A = \sum_{i=1}^{100} \frac{p_{A\ i}}{i\ p_i}$$

and Normalized Driver Strength Index (NDSI)

$$NDSI_A = \frac{\sum_{i=1}^{100} \frac{p_{A\ i}}{i\ p_i}}{\sum_{i=1}^{100} \frac{p_{A\ i}}{p_i}}$$

were calculated, where $p_{A\ i}$ is a number of patients with a driver event in the gene/chromosome *A* amongst patients with *i* driver events in total; p_i is a number of patients with *i* driver events in total. To avoid contamination of NDSI-ranked driver event lists with very rare driver events and to increase precision of the index calculation, all events that were present in less than 10 patients in each driver event class were removed. To compose the top-(N)DSI-ranked driver list, the lists of drivers from various classes were combined, and drivers with lower (N)DSI in case of duplicates and all drivers with NDSI<0.05 were removed.

Pathway and network analysis of top-(N)DSI-ranked driver genes

First, the chromosome arms and full chromosomes were removed from the top-(N)DSI-ranked driver lists, as external pathway and network analysis services can work only with genes.

Next, top 100 DSI-ranked genes and top 100 NDSI-ranked genes were selected, to facilitate proper comparison.

The resulting lists were uploaded as Entrez Gene IDs to "Reactome v76 Analyse gene list" tool (https://reactome.org/PathwayBrowser/#TOOL=AT). Voronoi visualizations (Reacfoam) were exported as jpg files.

The resulting lists were also uploaded as Entrez Gene IDs to "KEGG Mapper –Color" tool (https://www.genome.jp/kegg/mapper/color.html), "hsa" Search mode was selected, default bgcolor assigned to "yellow", search executed and the top result - "Pathways in cancer - Homo sapiens (human)" (hsa05200) was selected for mapping. The resulting images were exported as png files.

The data were also analyzed in Cytoscape 3.8.2 (https://cytoscape.org). BioGRID: Protein-Protein Interactions (H. sapiens) network was imported and then (N)DSI values appended from the top 100 (N)DSI-ranked driver list. First, Degree Sorted Circle Layout was selected and genes not within the circle were removed. Node Fill Color was mapped to (N)DSI values with Continuous Mapping. Then, Edge-weighted Spring Embedded Layout was selected, and Node Height and Width were mapped to degree.layout parameter (number of connections) with Continuous Mapping. The resulting images were exported as pdf files.

Acknowledgements

AVB acknowledges MIPT 5-100 program support for early career researchers.

References

1. Horn H, Lawrence MS, Chouinard CR, Shrestha Y, Hu JX, Worstell E, et al. NetSig: network-based discovery from cancer genomes. Nature Methods. 2018;15:61–6.

2. Amala A, Emerson IA. Identification of target genes in cancer diseases using protein–protein interaction networks. Network Modeling Analysis in Health Informatics and Bioinformatics. 2019;8:2.

3. Gumpinger AC, Lage K, Horn H, Borgwardt K. Prediction of cancer driver genes through network-based moment propagation of mutation scores. Bioinformatics. 2020;36:i508–15.

4. Bowler EH, Wang Z, Ewing RM. How do oncoprotein mutations rewire protein–protein interaction networks? null. Taylor & Francis; 2015;12:449–55.

5. Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. Evaluating the evaluation of cancer driver genes. Proceedings of the National Academy of Sciences. 2016;113:14330 LP – 14335.

6. Tokheim C, Karchin R. CHASMplus Reveals the Scope of Somatic Missense Mutations Driving Human Cancers. Cell Systems. Elsevier; 2019;9:9-23.e8.

7. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. Elsevier; 2018;173:371-385.e18.

8. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. Elsevier; 2017;171:1029-1041.e21.

9. Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, et al. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome Biology. 2012;13:R124.

10. Tokheim C, Bhattacharya R, Niknafs N, Gygax DM, Kim R, Ryan M, et al. Exome-Scale Discovery of Hotspot Mutation Regions in Human Cancer Using 3D Protein Structure. Cancer

Research. 2016;76:3719 LP – 3731.

11. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nature Reviews Cancer. Springer US; 2020;20:555–72.

12. Arnedo-Pac C, Mularoni L, Muiños F, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers. Bioinformatics. 2019;35:4788–90.

13. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biology. 2016;17:128.

14. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. Science. 2013;339:1546 LP – 1558.

15. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature. 2002;417:949–54.

16. Dankner M, Rose AAN, Rajkumar S, Siegel PM, Watson IR. Classifying BRAF alterations in cancer: new rational therapeutic strategies for actionable mutations. Oncogene. 2018;37:3183–99.

17. Prior IA, Hood FE, Hartley JL. The Frequency of Ras Mutations in Cancer. Cancer Res. 2020;80:2969.

18. Muñoz-Maldonado C, Zimmer Y, Medová M. A Comparative Analysis of Individual RAS Mutations in Cancer Biology. Frontiers in Oncology. 2019;9:1088.

19. Gala K, Li Q, Sinha A, Razavi P, Dorso M, Sanchez-Vega F, et al. KMT2C mediates the estrogen dependence of breast cancer through regulation of ERα enhancer function. Oncogene. 2018;37:4692–710.

20. Alam H, Tang M, Maitituoheti M, Dhar SS, Kumar M, Han CY, et al. KMT2D Deficiency

Impairs Super-Enhancers to Confer a Glycolytic Vulnerability in Lung Cancer. Cancer Cell. Elsevier; 2020;37:599-617.e7.

21. Fagan RJ, Dingwall AK. COMPASS Ascending: Emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. Cancer Letters. 2019;458:56–65.

22. Petrini I, Meltzer PS, Kim I-K, Lucchi M, Park K-S, Fontanini G, et al. A specific missense mutation in GTF2I occurs at high frequency in thymic epithelial tumors. Nature Genetics. 2014;46:844–9.

23. Kim I-K, Rao G, Zhao X, Fan R, Avantaggiati ML, Wang Y, et al. Mutant GTF2I induces cell transformation and metabolic alterations in thymic epithelial cells. Cell Death & Differentiation. 2020;27:2263–79.

24. Nathany S, Tripathi R, Mehta A. Gene of the month: GTF2I. J Clin Pathol. 2021;74:1.

25. Ross-Adams H, Lamb AD, Dunning MJ, Halim S, Lindberg J, Massie CM, et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study. EBioMedicine. 2015;2:1133–44.

26. Paris PL, Hofer MD, Albo G, Kuefer R, Gschwend JE, Hautmann RE, et al. Genomic Profiling of Hormone-Naïve Lymph Node Metastases in Patients with Prostate Cancer. Neoplasia. 2006;8:1083-IN35.

27. Katsyv I, Wang M, Song WM, Zhou X, Zhao Y, Park S, et al. EPRS is a critical regulator of cell proliferation and estrogen signaling in ER + breast cancer. Oncotarget; Vol 7, No 43 [Internet]. 2016 [cited 2016 Jan 1]; Available from: https://www.oncotarget.com/article/11870/text/

28. Aryal B, Rao VA. Specific protein carbonylation in human breast cancer tissue compared to adjacent healthy epithelial tissue. PLOS ONE. Public Library of Science; 2018;13:e0194164.

29. Qi L, Zhou B, Chen J, Hu W, Bai R, Ye C, et al. Significant prognostic values of

differentially expressed-aberrantly methylated hub genes in breast cancer. J Cancer. Ivyspring International Publisher; 2019;10:6618–34.

30. Liu H, Fredimoses M, Niu P, Liu T, Qiao Y, Tian X, et al. EPRS/GluRS promotes gastric cancer development via WNT/GSK-3β/β-catenin signaling pathway. Gastric Cancer [Internet]. 2021; Available from: https://doi.org/10.1007/s10120-021-01180-x

31. Heinrich MC, Corless CL, Duensing A, McGreevey L, Chen C-J, Joseph N, et al. PDGFRA Activating Mutations in Gastrointestinal Stromal Tumors. Science. 2003;299:708.

32. Velghe AI, Van Cauwenberghe S, Polyansky AA, Chand D, Montano-Almendras CP, Charni S, et al. PDGFRA alterations in cancer: characterization of a gain-of-function V536E transmembrane mutant as well as loss-of-function and passenger mutations. Oncogene. 2014;33:2568–76.

33. Disel U, Madison R, Abhishek K, Chung JH, Trabucco SE, Matos AO, et al. The Pan-Cancer Landscape of Coamplification of the Tyrosine Kinases KIT, KDR, and PDGFRA. The Oncologist. John Wiley & Sons, Ltd; 2020;25:e39–47.

34. Mendes A, Fahrenkrog B. NUP214 in Leukemia: It's More than Transport. Cells. 2019;8.

35. Bhattacharjya S, Roy KS, Ganguly A, Sarkar S, Panda CK, Bhattacharyya D, et al. Inhibition of nucleoporin member Nup214 expression by miR-133b perturbs mitotic timing and leads to cell death. Molecular Cancer. 2015;14:42.

36. Roy A, Narayan G. Oncogenic potential of nucleoporins in non-hematological cancers: recent update beyond chromosome translocation and gene fusion. Journal of Cancer Research and Clinical Oncology. 2019;145:2901–10.

37. Stolarova L, Kleiblova P, Janatova M, Soukupova J, Zemankova P, Macurek L, et al. CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate. Cells. 2020;9.

38. Cybulski C, Wokołorczyk D, Jakubowska A, Huzarski T, Byrski T, Gronwald J, et al. Risk of Breast Cancer in Women With a CHEK2 Mutation With and Without a Family History of Breast Cancer. JCO. Wolters Kluwer; 2011;29:3747–52.

39. Seppälä EH, Ikonen T, Mononen N, Autio V, Rökman A, Matikainen MP, et al. CHEK2 variants associate with hereditary prostate cancer. British Journal of Cancer. 2003;89:1966–70.

40. Siołek M, Cybulski C, Gąsior-Perczak D, Kowalik A, Kozak-Klonowska B, Kowalska A, et al. CHEK2 mutations and the risk of papillary thyroid cancer. International Journal of Cancer. John Wiley & Sons, Ltd; 2015;137:548–52.

41. Wójcicka A, Czetwertyńska M, Świerniak M, Długosińska J, Maciąg M, Czajka A, et al. Variants in the ATM-CHEK2-BRCA1 axis determine genetic predisposition and clinical presentation of papillary thyroid carcinoma. Genes, Chromosomes and Cancer. John Wiley & Sons, Ltd; 2014;53:516–23.

42. Elman JS, Ni TK, Mengwasser KE, Jin D, Wronski A, Elledge SJ, et al. Identification of FUBP1 as a Long Tail Cancer Driver and Widespread Regulator of Tumor Suppressor and Oncogene Alternative Splicing. Cell Reports. Elsevier; 2019;28:3435-3449.e5.

43. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, Hruban RH, et al. Mutations in CIC and FUBP1 Contribute to Human Oligodendroglioma. Science. 2011;333:1453.

44. Sahm F, Koelsche C, Meyer J, Pusch S, Lindenberg K, Mueller W, et al. CIC and FUBP1 mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas. Acta Neuropathologica. 2012;123:853–60.

45. Antonescu CR, Owosho AA, Zhang L, Chen S, Deniz K, Huryn JM, et al. Sarcomas With CIC-rearrangements Are a Distinct Pathologic Entity With Aggressive Outcome: A Clinicopathologic and Molecular Study of 115 Cases. The American Journal of Surgical Pathology [Internet]. 2017;41. Available from:

https://journals.lww.com/ajsp/Fulltext/2017/07000/Sarcomas_With_CIC_rearrangements_Are_a_Distinct.9.aspx

46. Okimoto RA, Breitenbuecher F, Olivas VR, Wu W, Gini B, Hofree M, et al. Inactivation of Capicua drives cancer metastasis. Nature Genetics. 2017;49:87–96.

47. LeBlanc VG, Firme M, Song J, Chan SY, Lee MH, Yip S, et al. Comparative transcriptome analysis of isogenic cell line models and primary cancers links capicua (CIC) loss to activation of the MAPK signalling cascade. The Journal of Pathology. John Wiley & Sons, Ltd; 2017;242:206–20.

48. Weissmann S, Cloos PA, Sidoli S, Jensen ON, Pollard S, Helin K. The Tumor Suppressor CIC Directly Regulates MAPK Pathway Genes via Histone Deacetylation. Cancer Res. 2018;78:4114.

49. Yoe J, Kim D, Kim S, Lee Y. Capicua restricts cancer stem cell-like properties in breast cancer cells. Oncogene. 2020;39:3489–506.

50. Wong D, Yip S. Making heads or tails – the emergence of capicua (CIC) as an important multifunctional tumour suppressor. The Journal of Pathology. John Wiley & Sons, Ltd; 2020;250:532–40.

51. Kim JW, Ponce RK, Okimoto RA. Capicua in Human Cancer. Trends in Cancer. 2021;7:77–86.

52. Ebrahimi SA, Wang EH, Wu A, Schreck RR, Passaro E, Sawicki MP. Deletion of Chromosome 1 Predicts Prognosis in Pancreatic Endocrine Tumors. Cancer Res. 1999;59:311.

53. Cheung TH, Hung Chung TK, Poon CS, Hampton GM, Wang VW, Wong YF. Allelic loss on chromosome 1 is associated with tumor progression of cervical carcinoma. Cancer. John Wiley & Sons, Ltd; 1999;86:1294–8.

54. Mathew S, Murty VVVS, Bosl GJ, Chaganti RSK. Loss of Heterozygosity Identifies Multiple Sites of Allelic Deletions on Chromosome 1 in Human Male Germ Cell Tumors. Cancer Res.

1994;54:6265.

55. Šárová I, Březinová J, Zemanová Z, Izáková S, Lizcová L, Malinová E, et al. Cytogenetic manifestation of chromosome 11 duplication/amplification in acute myeloid leukemia. Cancer Genetics and Cytogenetics. Elsevier; 2010;199:121–7.

56. Szponar A, Zubakov D, Pawlak J, Jauch A, Kovacs G. Three genetic developmental stages of papillary renal cell tumors: Duplication of chromosome 1q marks fatal progression. International Journal of Cancer. John Wiley & Sons, Ltd; 2009;124:2071–6.

57. Balint I, Szponar A, Jauch A, Kovacs G. Trisomy 7 and 17 mark papillary renal cell tumours irrespectively of variation of the phenotype. J Clin Pathol. 2009;62:892.

58. Barem Rabenhorst SH, Lima Verde Osterne R, Weege Nonaka CF, Montezuma Sales Rodrigues A, Luiz Maia Nogueira R, Mário Rodriguez Burbano R, et al. Detection of deletions in 1q25, 1p36 and 1pTEL and chromosome 17 aneuploidy in oral epithelial dysplasia and oral squamous cell carcinoma by fluorescence in situ hybridization (FISH). Oral Oncology. 2021;116:105221.

59. von Deimling A, Louis DN, von Ammon K, Petersen I, Wiestler OD, Seizinger BR. Evidence for a Tumor Suppressor Gene on Chromosome 19q Associated with Human Astrocytomas, Oligodendrogliomas, and Mixed Gliomas. Cancer Res. 1992;52:4277.

60. McDonald JM, See SJ, Tremont IW, Colman H, Gilbert MR, Groves M, et al. The prognostic impact of histology and 1p/19q status in anaplastic oligodendroglial tumors. Cancer. John Wiley & Sons, Ltd; 2005;104:1468–77.

61. Yip S, Butterfield YS, Morozova O, Chittaranjan S, Blough MD, An J, et al. Concurrent CIC mutations, IDH mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. The Journal of Pathology. John Wiley & Sons, Ltd; 2012;226:7–16.

62. Williams EA, Sharaf R, Decker B, Werth AJ, Toma H, Montesion M, et al. CDKN2C-Null Leiomyosarcoma: A Novel, Genomically Distinct Class of TP53/RB1–Wild-Type Tumor With

Frequent CIC Genomic Alterations and 1p/19q-Codeletion. JCO Precision Oncology. Wolters Kluwer; 2020;955–71.

63. Jiang F, Richter J, Schraml P, Bubendorf L, Gasser T, Sauter G, et al. Chromosomal Imbalances in Papillary Renal Cell Carcinoma: Genetic Differences between Histological Subtypes. The American Journal of Pathology. 1998;153:1467–73.

64. Sandgren J, Diaz de Ståhl T, Andersson R, Menzel U, Piotrowski A, Nord H, et al. Recurrent genomic alterations in benign and malignant pheochromocytomas and paragangliomas revealed by whole-genome array comparative genomic hybridization analysis. Endocrine-Related Cancer. Bristol, UK: Society for Endocrinology; 2010;17:561–79.

65. Garcia-Marcos M, Maziarz M, Leyme A. A novel BRET biosensor for Gαq-GTP reveals unique properties of cancer-associated GNAQ mutants. The FASEB Journal. John Wiley & Sons, Ltd; 2018;32:557.3-557.3.

66. Onken MD, Worley LA, Long MD, Duan S, Council ML, Bowcock AM, et al. Oncogenic Mutations in GNAQ Occur Early in Uveal Melanoma. Investigative Ophthalmology & Visual Science. 2008;49:5230–4.

67. Van Raamsdonk CD, Griewank KG, Crosby MB, Garrido MC, Vemula S, Wiesner T, et al. Mutations in GNA11 in Uveal Melanoma. N Engl J Med. Massachusetts Medical Society; 2010;363:2191–9.

68. Yang H, Zhu H, Li H, Wang D, Ma T, Zhang C. 1963P A pan-cancer study of GNAQ/GNA11 mutations in Chinese cancer patients. Annals of Oncology. Elsevier; 2020;31:S1104–5.

69. Cho H, Kehrl JH. Localization of Giα proteins in the centrosomes and at the midbody: implication for their role in cell division. Journal of Cell Biology. 2007;178:245–55.

70. Chen Z, Liu B, Yi M, Qiu H, Yuan X. A Prognostic Nomogram Model Based on mRNA Expression of DNA Methylation-Driven Genes for Gastric Cancer. Frontiers in Oncology. 2020;10:2531.

71. Cárdenas-Navia LI, Cruz P, Lin JC, NISC Comparative Sequencing Program, Rosenberg SA, Samuels Y. Novel somatic mutations in heterotrimeric G proteins in melanoma. null. Taylor & Francis; 2010;10:33–7.

72. Lee Y-F, Roe T, Mangham DC, Fisher C, Grimer RJ, Judson I. Gene expression profiling identifies distinct molecular subgroups of leiomyosarcoma with clinical relevance. British Journal of Cancer. 2016;115:1000–7.

73. Fingas CD, Katsounas A, Kahraman A, Siffert W, Jochum C, Gerken G, et al. Prognostic Assessment of Three Single-Nucleotide Polymorphisms (GNB3 825C>T, BCL2-938C>A, MCL1-386C>G) in Extrahepatic Cholangiocarcinoma. null. Taylor & Francis; 2010;28:472–8.

74. Eisenhardt A, Siffert W, Rosskopf D, Musch M, Mosters M, Roggenbuck U, et al. Association study of the G-protein β3 subunit C825T polymorphism with disease progression in patients with bladder cancer. World Journal of Urology. 2005;23:279–86.

75. Safarinejad Mohammad Reza, Safarinejad Shiva, Shafiei Nayyer, Safarinejad Saba. G Protein β3 Subunit Gene C825T Polymorphism and its Association with the Presence and Clinicopathological Characteristics of Prostate Cancer. Journal of Urology. WoltersKluwer; 2012;188:287–93.

76. Larribère L, Utikal J. Update on GNA Alterations in Cancer: Implications for Uveal Melanoma Treatment. Cancers. 2020;12.

77. Rask-Andersen M, Almén MS, Schiöth HB. Trends in the exploitation of novel drug targets. Nature Reviews Drug Discovery. 2011;10:579–90.

78. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 Mutations in Gliomas. N Engl J Med. Massachusetts Medical Society; 2009;360:765–73.

79. Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA, et al. The Common Feature of Leukemia-Associated IDH1 and IDH2 Mutations Is a Neomorphic Enzyme Activity Converting α-Ketoglutarate to 2-Hydroxyglutarate. Cancer Cell. Elsevier; 2010;17:225–34.

80. Bendahou MA, Arrouchi H, Lakhlili W, Allam L, Aanniz T, Cherradi N, et al. Computational Analysis of IDH1, IDH2, and TP53 Mutations in Low-Grade Gliomas Including Oligodendrogliomas and Astrocytomas. Cancer Inform. SAGE Publications Ltd STM; 2020;19:1176935120915839.

81. Yang H, Ye D, Guan K-L, Xiong Y. IDH1 and IDH2 Mutations in Tumorigenesis: Mechanistic Insights and Clinical Perspectives. Clin Cancer Res. 2012;18:5562.

82. Marcucci G, Maharry K, Wu Y-Z, Radmacher MD, Mrózek K, Margeson D, et al. IDH1 and IDH2 Gene Mutations Identify Novel Molecular Subsets Within De Novo Cytogenetically Normal Acute Myeloid Leukemia: A Cancer and Leukemia Group B Study. JCO. Wolters Kluwer; 2010;28:2348–55.

83. Labussière M, Idbaih A, Wang X-W, Marie Y, Boisselier B, Falet C, et al. All the 1p19q codeleted gliomas are mutated on IDH1 or IDH2. Neurology. 2010;74:1886.

84. Helsten T, Elkin S, Arthur E, Tomson BN, Carter J, Kurzrock R. The FGFR Landscape in Cancer: Analysis of 4,853 Tumors by Next-Generation Sequencing. Clin Cancer Res. 2016;22:259.

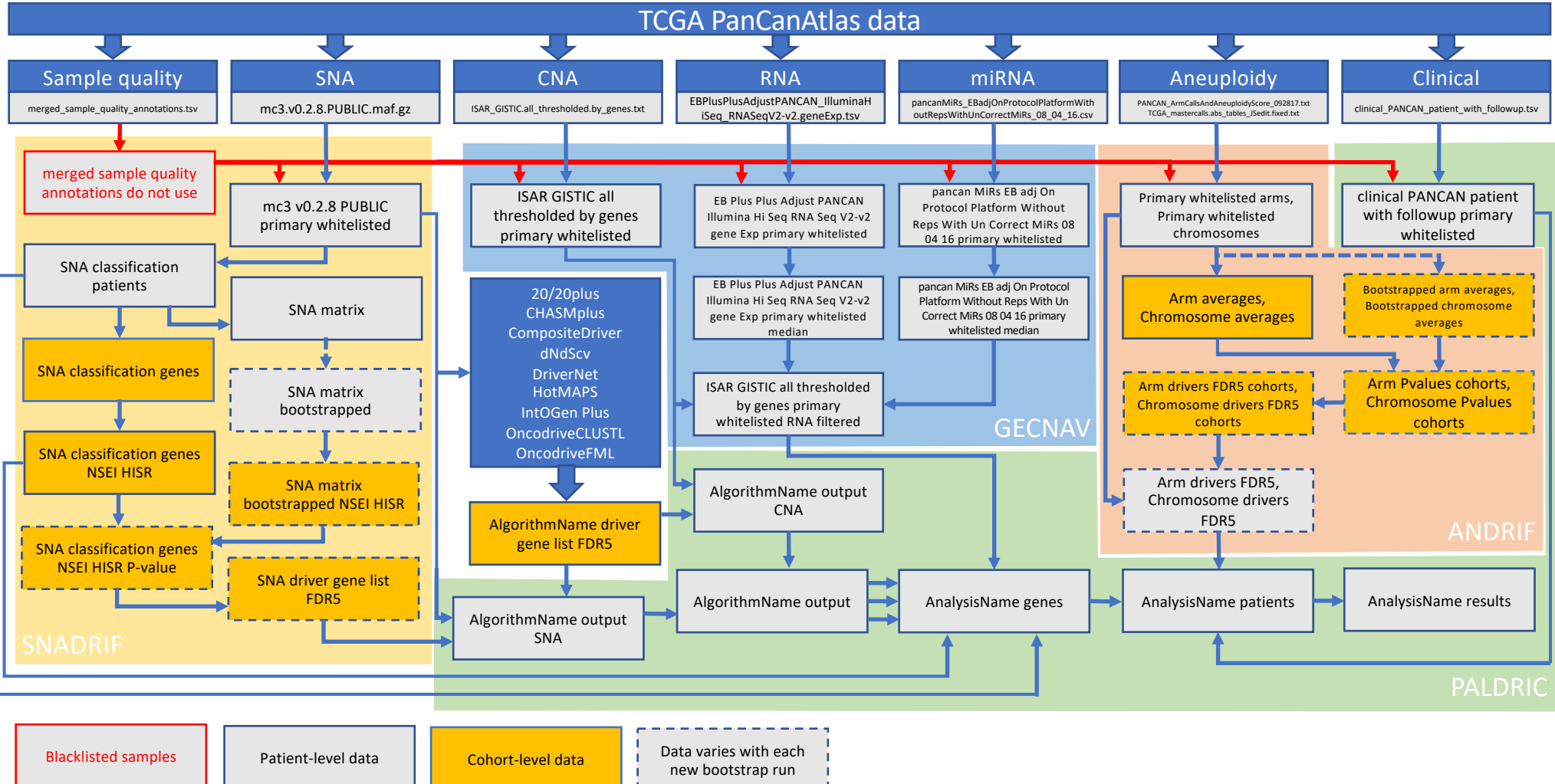
85. Iranzo J, Martincorena I, Koonin EV. Cancer-mutation network and the number and specificity of driver mutations. Proc Natl Acad Sci USA. 2018;115:E6010.

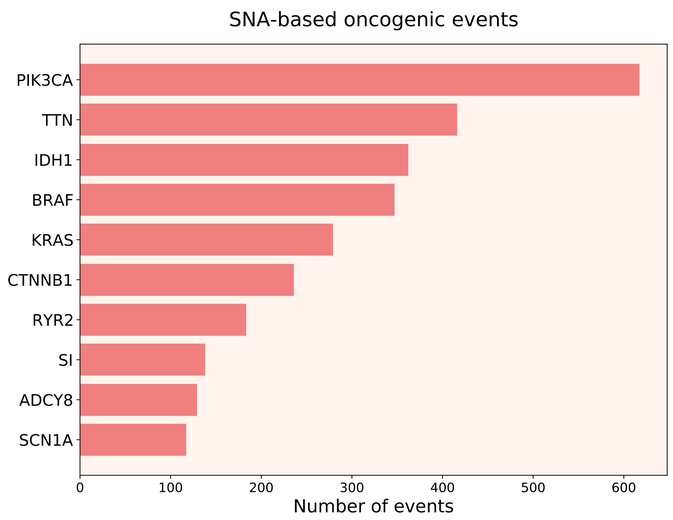
86. Belikov AV. The number of key carcinogenic events can be predicted from cancer incidence. Scientific Reports. 2017;7.

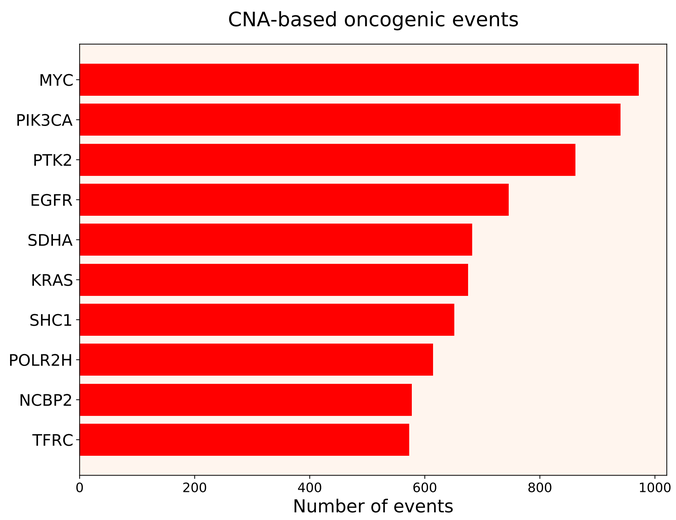
87. Levine AJ, Jenkins NA, Copeland NG. The Roles of Initiating Truncal Mutations in Human Cancers: The Order of Mutations and Tumor Cell Type Matters. Cancer Cell. Elsevier; 2019;35:10–5.

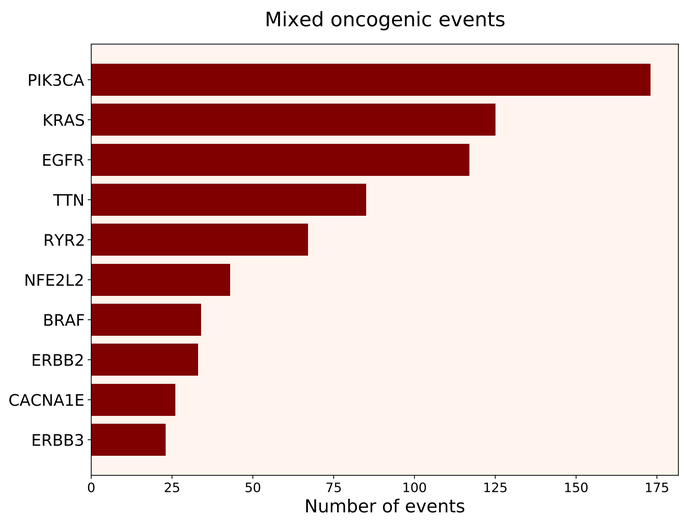
88. Hass R, von der Ohe J, Ungefroren H. Impact of the Tumor Microenvironment on Tumor Heterogeneity and Consequences for Cancer Cell Plasticity and Stemness. Cancers. 2020;12.

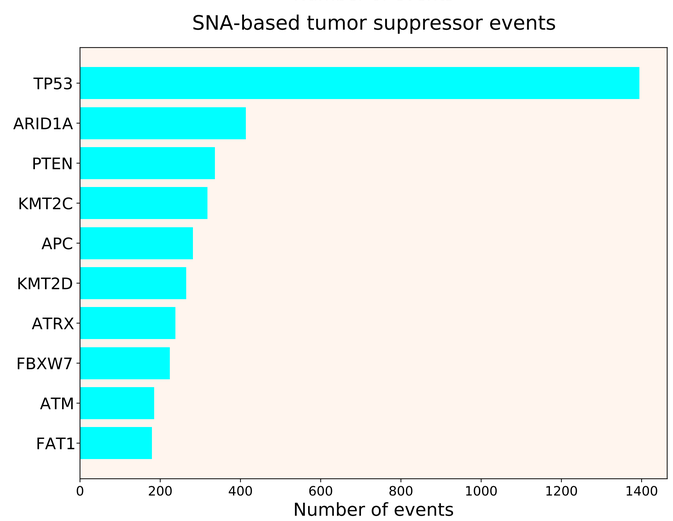
89. Shlyakhtina Y, Moran KL, Portal MM. Genetic and Non-Genetic Mechanisms Underlying Cancer Evolution. Cancers. 2021;13.

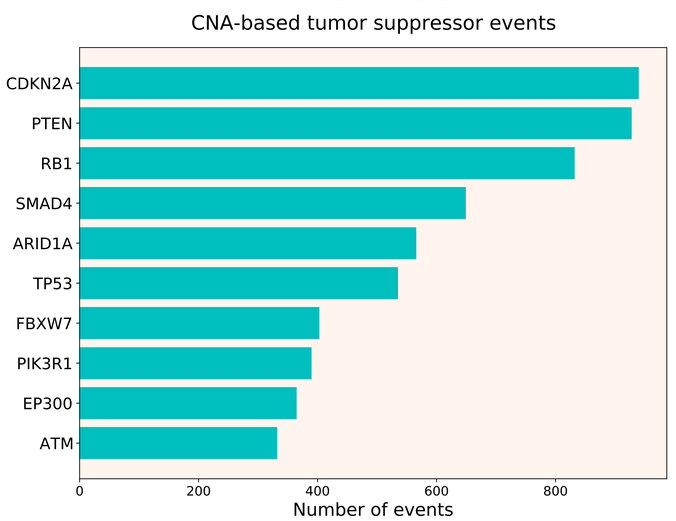

TCGA PanCanAtlas data
Sample quality
merged_sample_quality_annotations.tsv
SNA
mc3.v0.2.8.PUBLIC.maf.gz
CNA
ISAR_GISTIC.all_thresholded.by_genes.txt
RNA
EBPlusPlusAdjustPANCAN_IlluminaHiSeq_RNASeqV2-v2.geneExp.tsv
miRNA
pancanMiRs_EBadjOnProtocolPlatformWithoutRepsWithUnCorrectMiRs_08_04_16.csv
Aneuploidy
PANCAN_ArmCallsAndAneuploidyScore_092817.txt
TCGA_mastercalls.abs_tables_JSedit.fixed.txt
Clinical
clinical_PANCAN_patient_with_followup.tsv
merged sample quality annotations do not use
mc3 v0.2.8 PUBLIC primary whitelisted
ISAR GISTIC all thresholded by genes primary whitelisted
EB Plus Plus Adjust PANCAN Illumina Hi Seq RNA Seq V2-v2 gene Exp primary whitelisted
pancan MiRs EB adj On Protocol Platform Without Reps With Un Correct MiRs 08 04 16 primary whitelisted
Primary whitelisted arms, Primary whitelisted chromosomes
clinical PANCAN patient with followup primary whitelisted
SNA classification patients
SNA matrix
20/20plus
CHASMplus
CompositeDriver
dNdScv
DriverNet
HotMAPS
IntOGen Plus
OncodriveCLUSTL
OncodriveFML
EB Plus Plus Adjust PANCAN Illumina Hi Seq RNA Seq V2-v2 gene Exp primary whitelisted median
pancan MiRs EB adj On Protocol Platform Without Reps With Un Correct MiRs 08 04 16 primary whitelisted median
Arm averages, Chromosome averages
Bootstrapped arm averages, Bootstrapped chromosome averages
SNA classification genes
SNA matrix bootstrapped
ISAR GISTIC all thresholded by genes primary whitelisted RNA filtered
GECNAV
Arm drivers FDR5 cohorts, Chromosome drivers FDR5 cohorts
Arm Pvalues cohorts, Chromosome Pvalues cohorts
SNA classification genes NSEI HISR
SNA matrix bootstrapped NSEI HISR
AlgorithmName output CNA
Arm drivers FDR5, Chromosome drivers FDR5
ANDRIF
SNA classification genes NSEI HISR P-value
SNA driver gene list FDR5
AlgorithmName driver gene list FDR5
AlgorithmName output SNA
AlgorithmName output
AnalysisName genes
AnalysisName patients
AnalysisName results
SNADRIF
PALDRIC
Blacklisted samples
Patient-level data
Cohort-level data
Data varies with each new bootstrap run

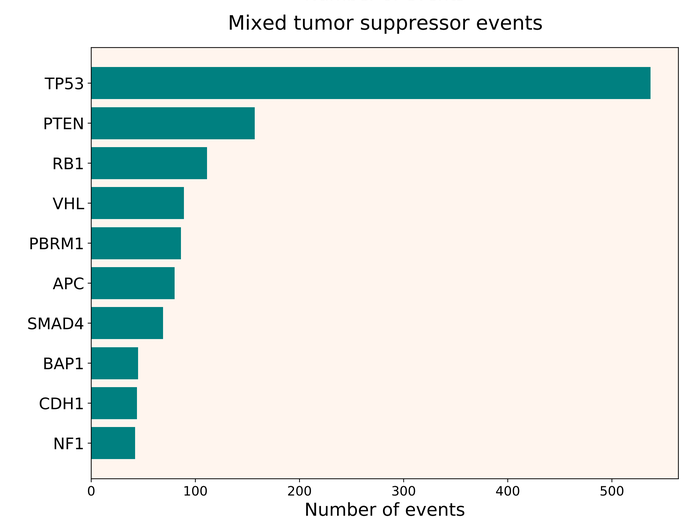

SNA-based oncogenic events
PIK3CA
TTN
IDH1
BRAF
KRAS
CTNNB1
RYR2
SI
ADCY8
SCN1A
0
100
200
300
400
500
600
Number of events

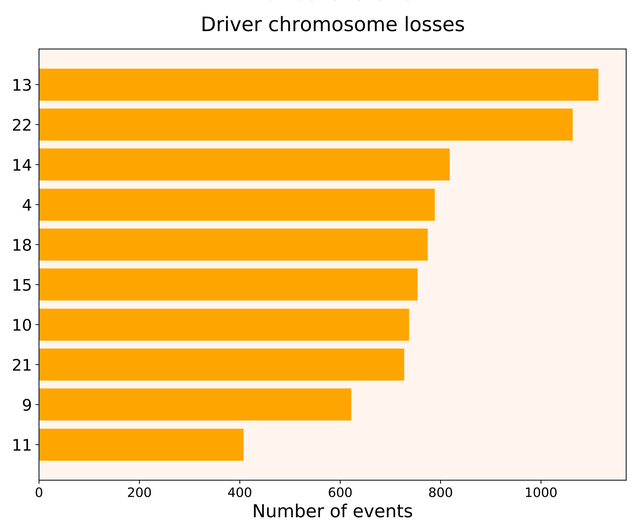

CNA-based oncogenic events
MYC
PIK3CA
PTK2
EGFR
SDHA
KRAS
SHC1
POLR2H
NCBP2
TFRC
0
200
400
600
800
1000
Number of events

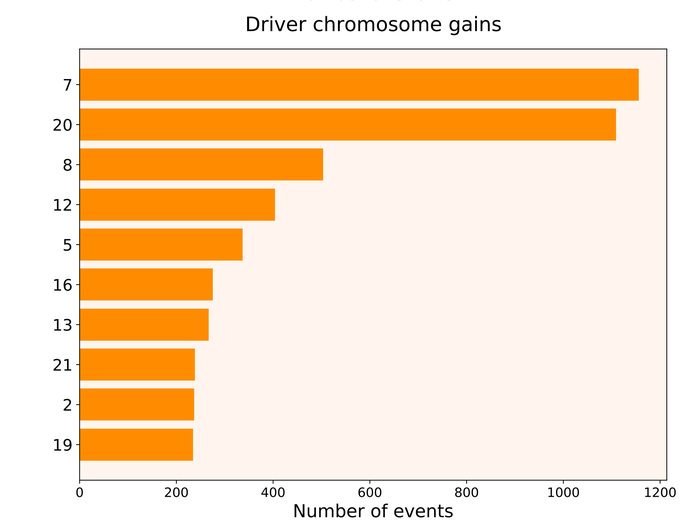

Mixed oncogenic events
PIK3CA
KRAS
EGFR
TTN
RYR2
NFE2L2
BRAF
ERBB2
CACNA1E
ERBB3
0
25
50
75
100
125
150
175
Number of events

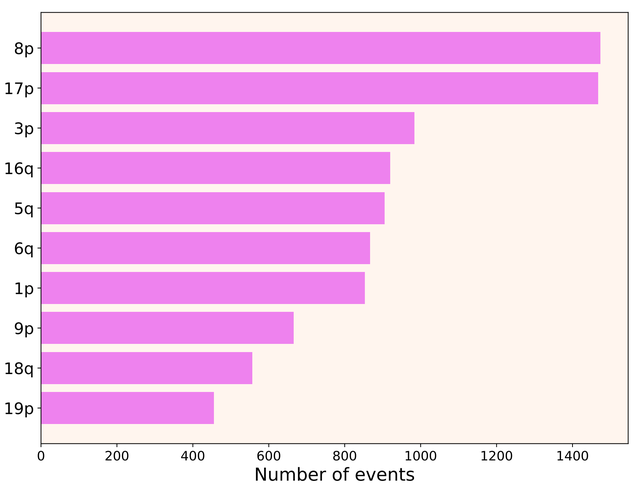

SNA-based tumor suppressor events
TP53
ARID1A
PTEN
KMT2C
APC
KMT2D
ATRX
FBXW7
ATM
FAT1
0
200
400
600
800
1000
1200
1400
Number of events

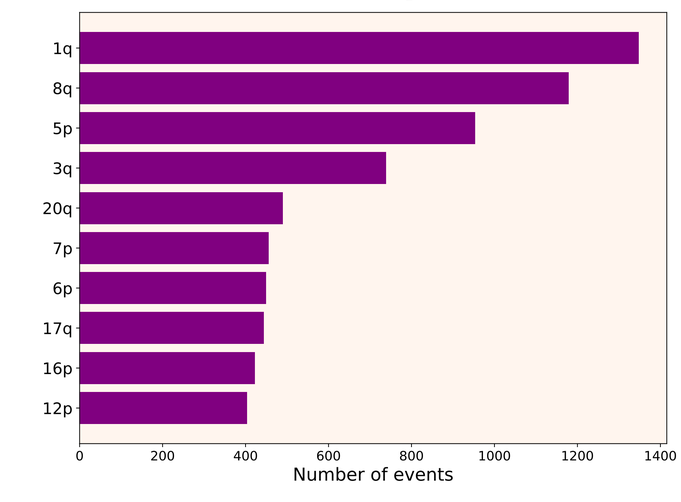

CNA-based tumor suppressor events
CDKN2A
PTEN
RB1
SMAD4
ARID1A
TP53
FBXW7
PIK3R1
EP300
ATM
0
200
400
600
800
Number of events

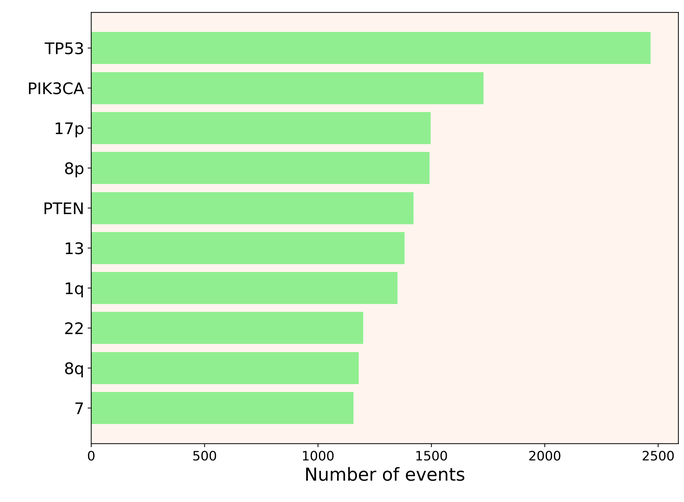

Mixed tumor suppressor events
TP53
PTEN
RB1
VHL
PBRM1
APC
SMAD4
BAP1
CDH1
NF1
0
100
200
300
400
500
Number of events

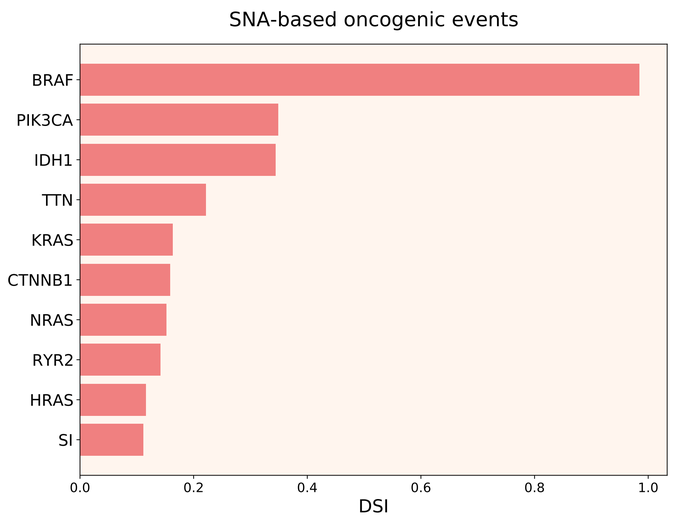

Driver chromosome losses
13
22
14
4
18
15
10
21
9
11
0
200
400
600
800
1000
Number of events

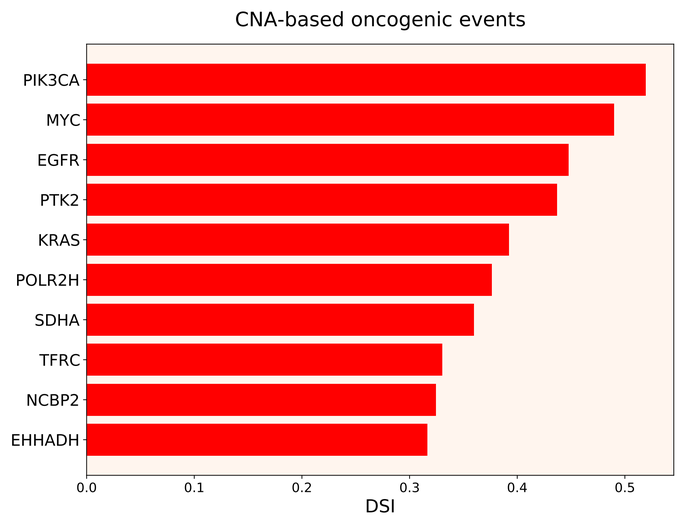

Driver chromosome gains
7
20
8
12
5
16
13
21
2
19
0
200
400
600
800
1000
1200
Number of events

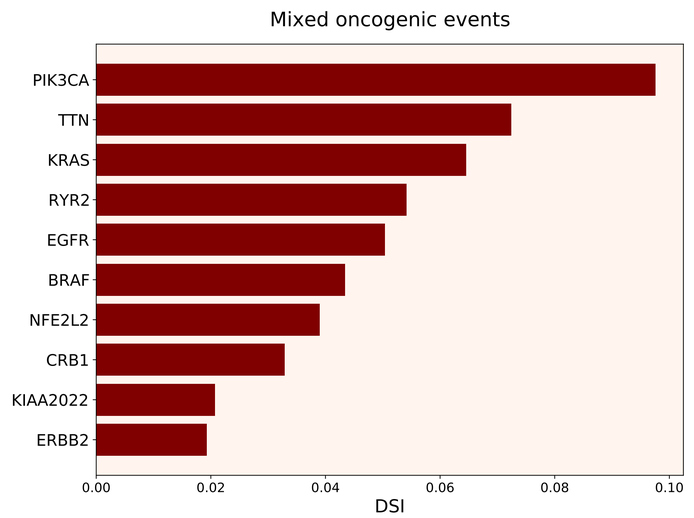

Driver arm losses
8p
17p
3p
16q
5q
6q
1p
9p
18q
19p
0
200
400
600
800
1000
1200
1400
Number of events

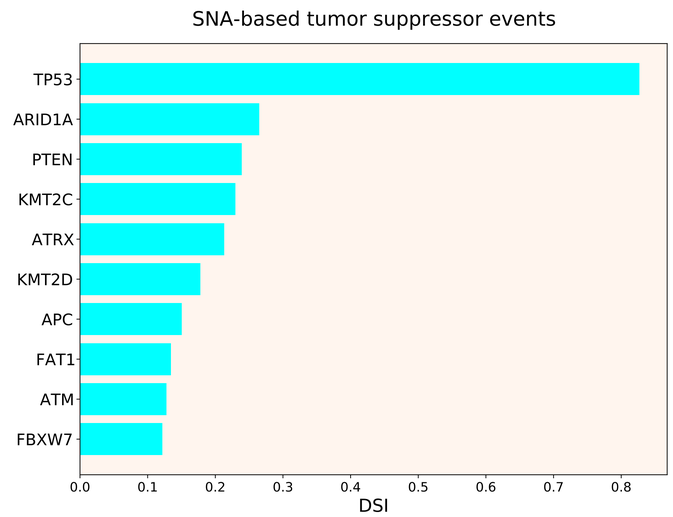

Driver arm gains
1q
8q
5p
3q
20q
7p
6p
17q
16p
12p
0
200
400
600
800
1000
1200
1400
Number of events

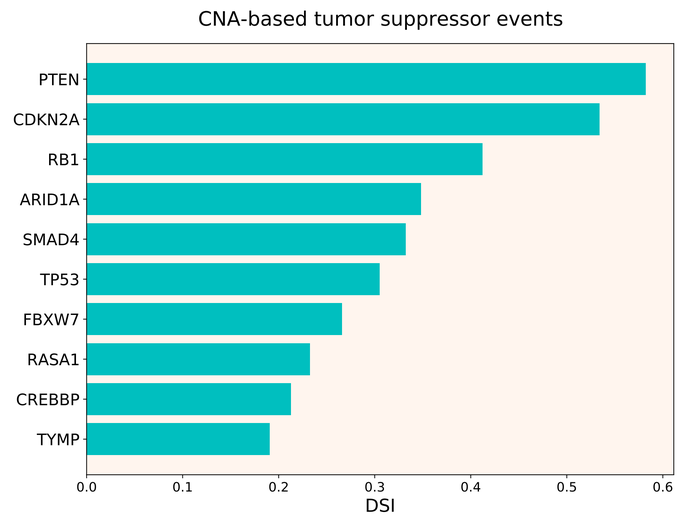

Driver events of all classes
TP53
PIK3CA
17p
8p
PTEN
13
1q
22
8q
7
0
500
1000
1500
2000
2500
Number of events

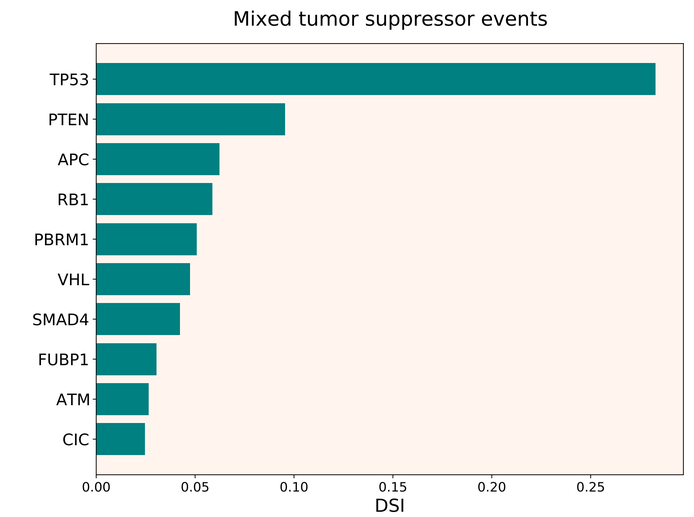

SNA-based oncogenic events
BRAF
PIK3CA
IDH1
TTN
KRAS
CTNNB1
NRAS
RYR2
HRAS
SI
0.0
0.2
0.4
0.6
0.8
1.0
DSI

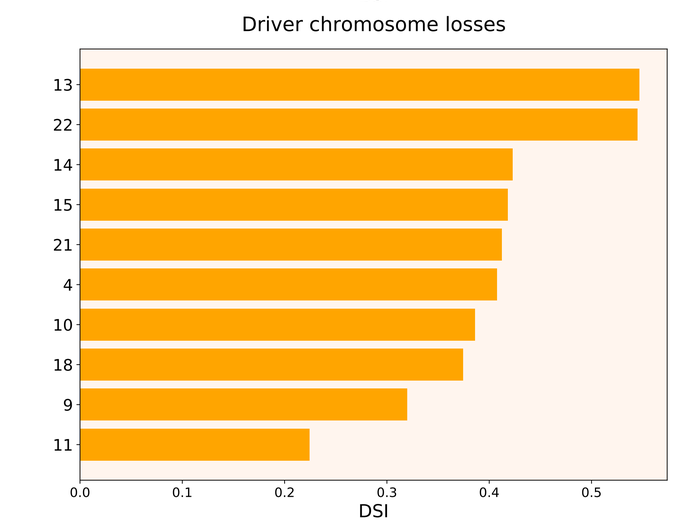

CNA-based oncogenic events
PIK3CA
MYC
EGFR
PTK2
KRAS
POLR2H
SDHA
TFRC
NCBP2
EHHADH
0.0
0.1
0.2
0.3
0.4
0.5
DSI

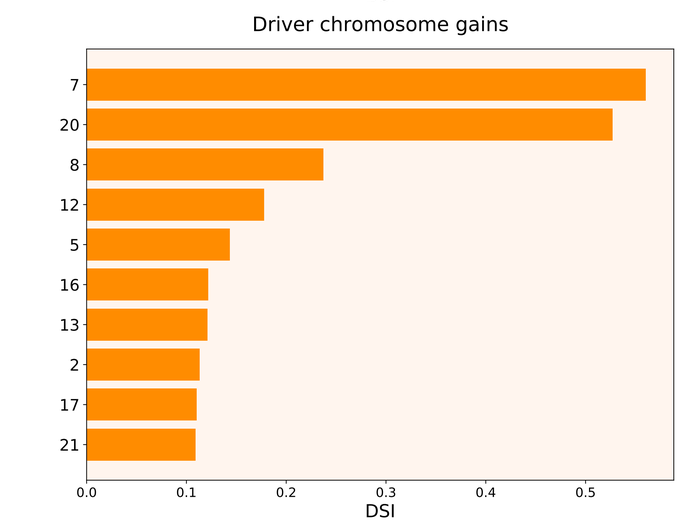

Mixed oncogenic events
PIK3CA
TTN
KRAS
RYR2
EGFR
BRAF
NFE2L2
CRB1
KIAA2022
ERBB2
0.00
0.02
0.04
0.06
0.08
0.10
DSI

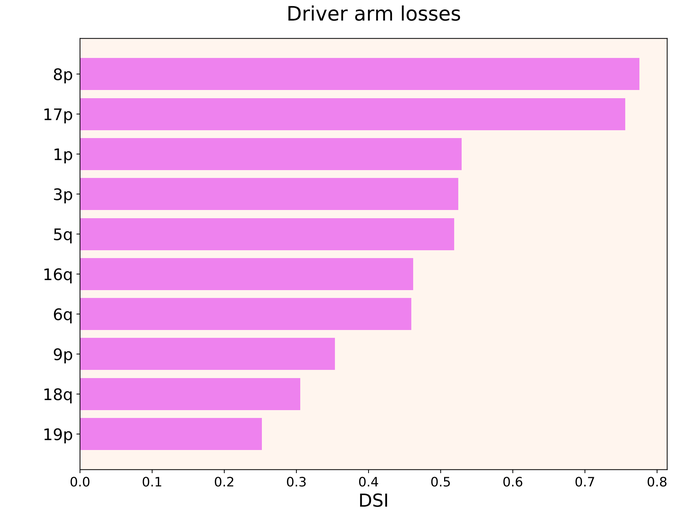

SNA-based tumor suppressor events
TP53
ARID1A
PTEN
KMT2C
ATRX
KMT2D
APC
FAT1
ATM
FBXW7
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
DSI

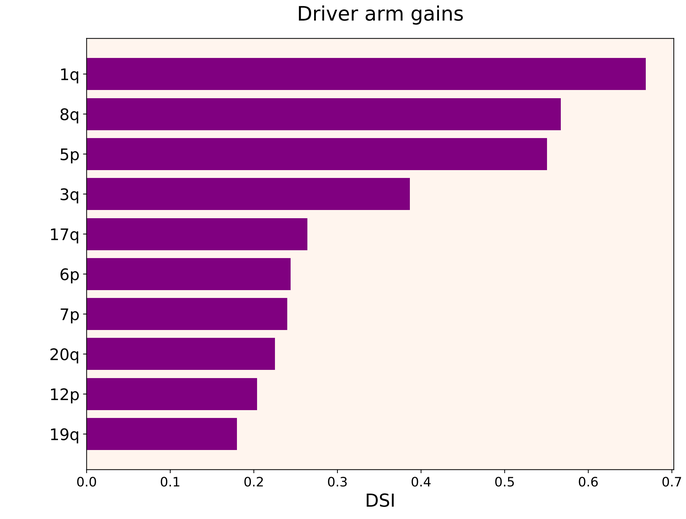

CNA-based tumor suppressor events
PTEN
CDKN2A
RB1
ARID1A
SMAD4
TP53
FBXW7
RASA1
CREBBP
TYMP
0.0
0.1
0.2
0.3
0.4
0.5
0.6
DSI

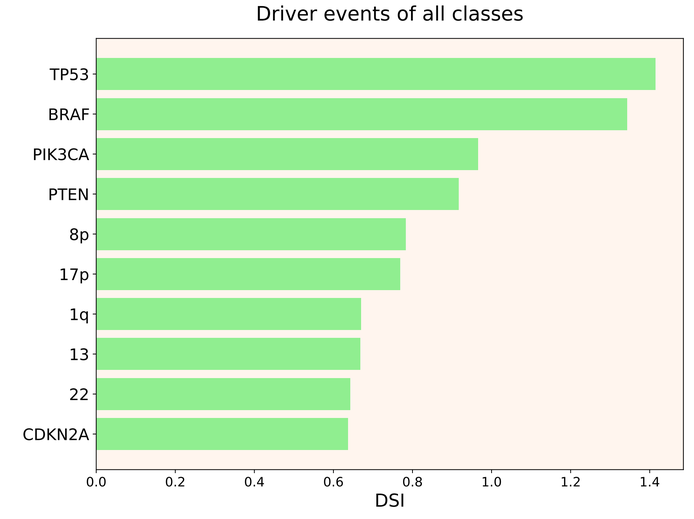

Mixed tumor suppressor events
TP53
PTEN
APC
RB1
PBRM1
VHL
SMAD4
FUBP1
ATM
CIC
0.00
0.05
0.10
0.15
0.20
0.25
DSI

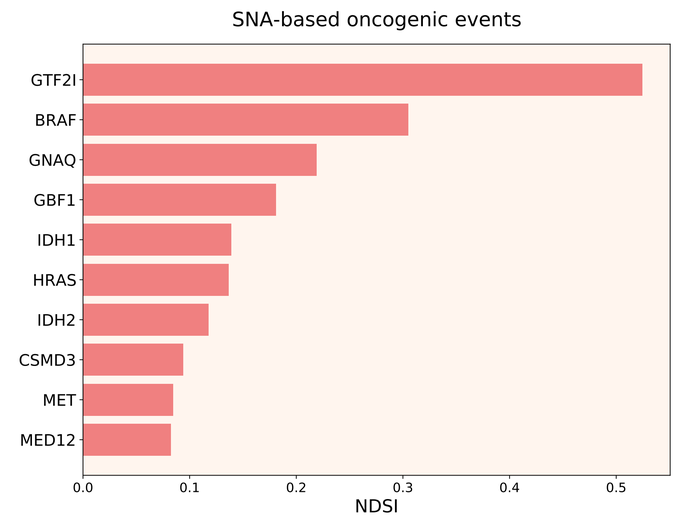

Driver chromosome losses
13
22
14
15
21
4
10
18
9
11
0.0
0.1
0.2
0.3
0.4
0.5
DSI

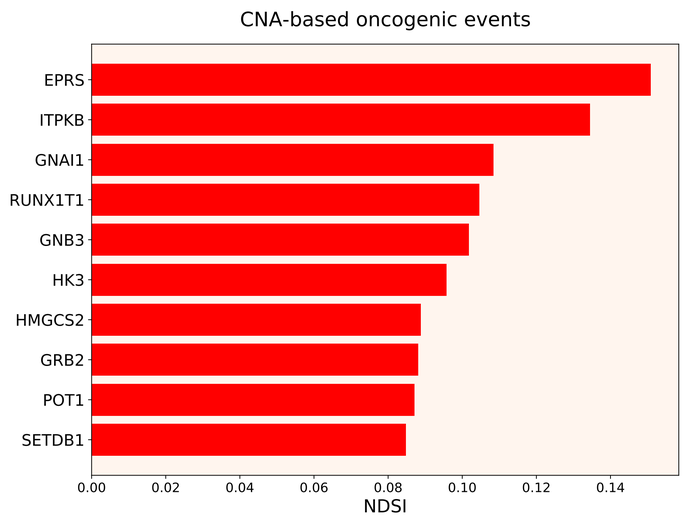

Driver chromosome gains
7
20
8
12
5
16
13
2
17
21
0.0
0.1
0.2
0.3
0.4
0.5
DSI

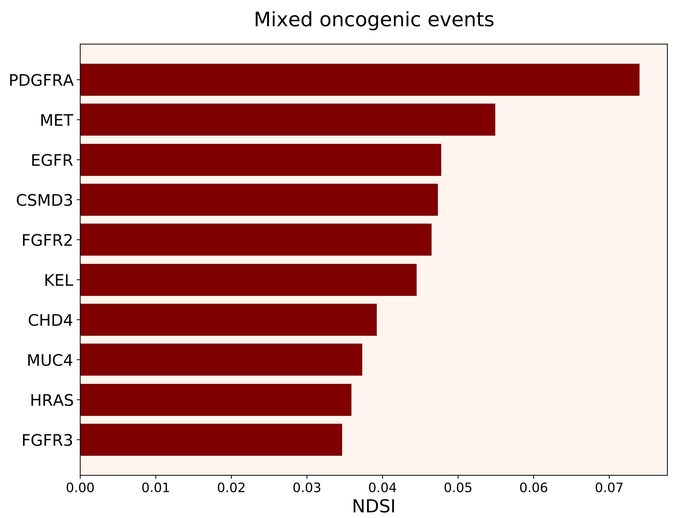

Driver arm losses
8p
17p
1p
3p
5q
16q
6q
9p
18q
19p
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
DSI

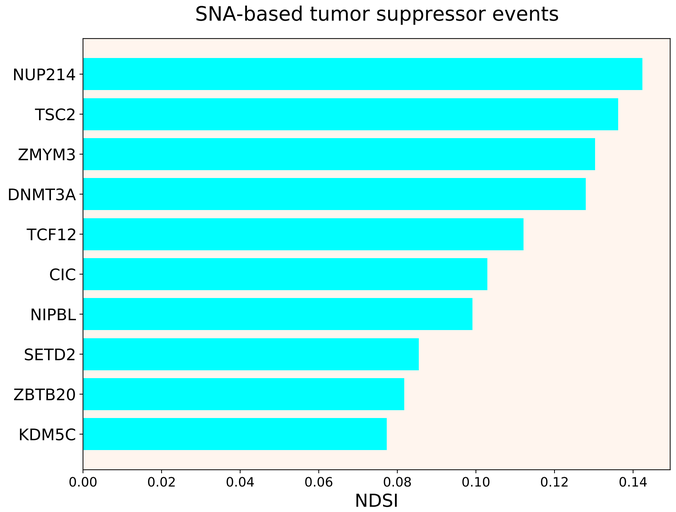

Driver arm gains
1q
8q
5p
3q
17q
6p
7p
20q
12p
19q
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
DSI

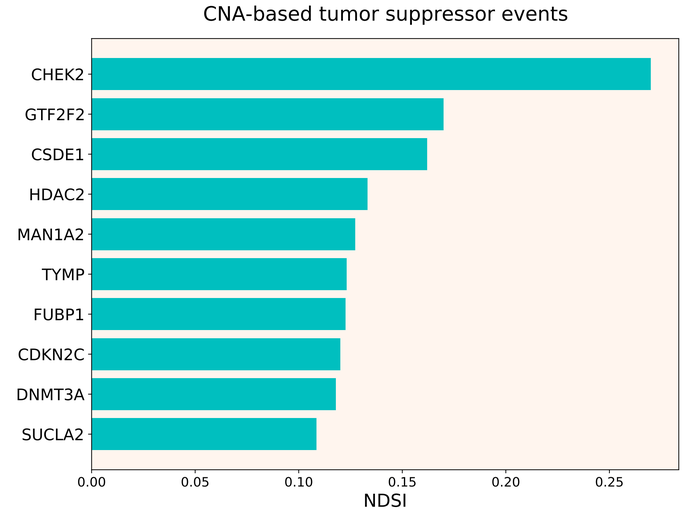

Driver events of all classes
TP53
BRAF
PIK3CA
PTEN
8p
17p
1q
13
22
CDKN2A
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
DSI

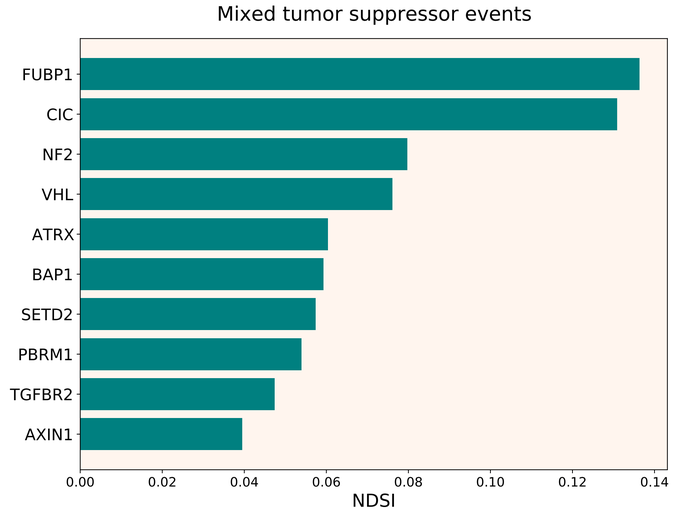

SNA-based oncogenic events
GTF2I
BRAF
GNAQ
GBF1
IDH1
HRAS
IDH2
CSMD3
MET
MED12
0.0
0.1
0.2
0.3
0.4
0.5
NDSI

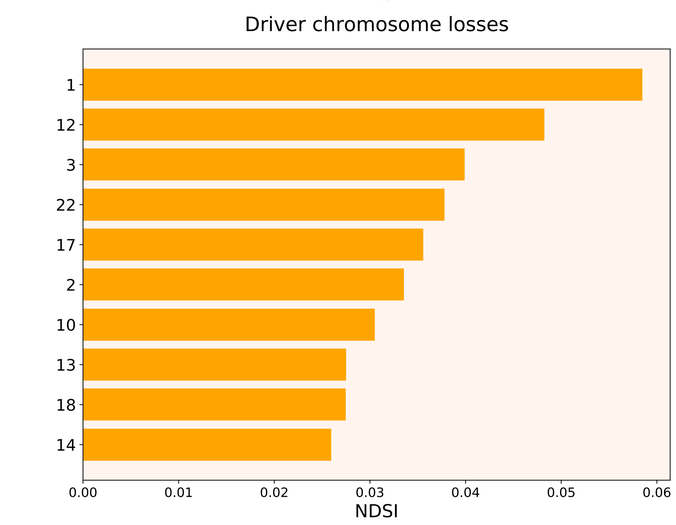

CNA-based oncogenic events
EPRS
ITPKB
GNAI1
RUNX1T1
GNB3
HK3
HMGCS2
GRB2
POT1
SETDB1
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
NDSI

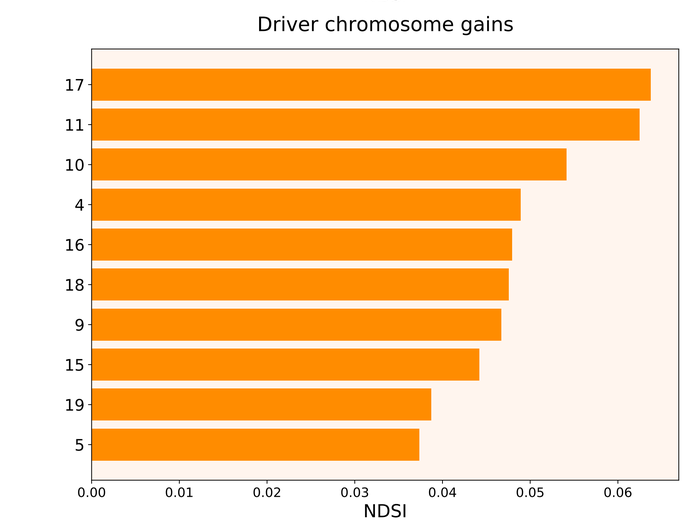

Mixed oncogenic events
PDGFRA
MET
EGFR
CSMD3
FGFR2
KEL
CHD4
MUC4
HRAS
FGFR3
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
NDSI

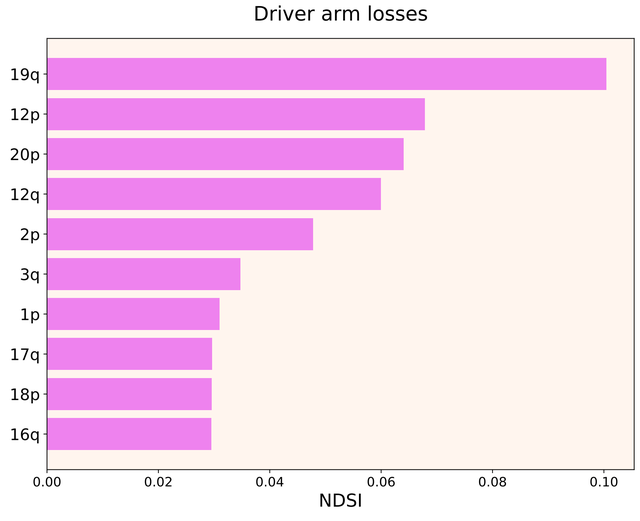

SNA-based tumor suppressor events
NUP214
TSC2
ZMYM3
DNMT3A
TCF12
CIC
NIPBL
SETD2
ZBTB20
KDM5C
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
NDSI

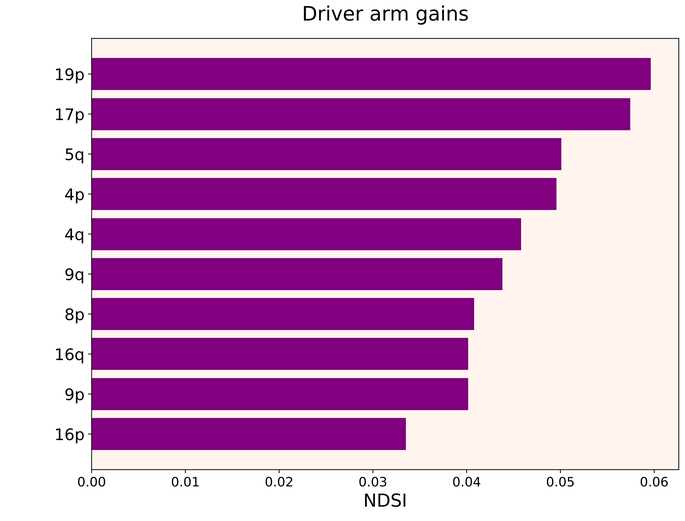

CNA-based tumor suppressor events
CHEK2
GTF2F2
CSDE1
HDAC2
MAN1A2
TYMP
FUBP1
CDKN2C
DNMT3A
SUCLA2
0.00
0.05
0.10
0.15
0.20
0.25
NDSI

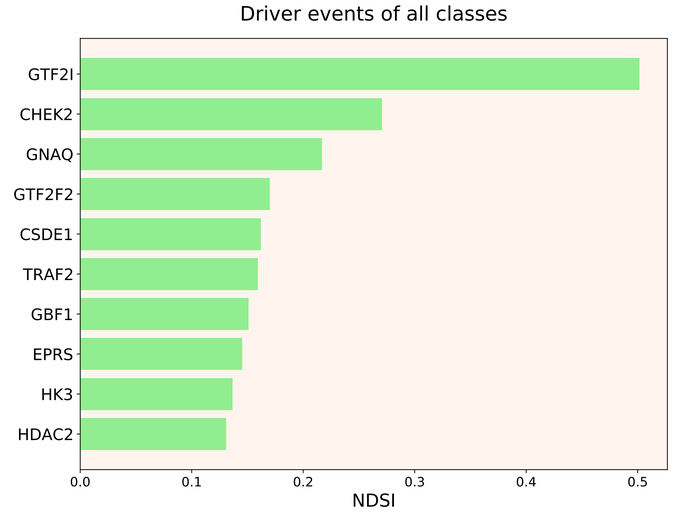

Mixed tumor suppressor events
FUBP1
CIC
NF2
VHL
ATRX
BAP1
SETD2
PBRM1
TGFBR2
AXIN1
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
NDSI

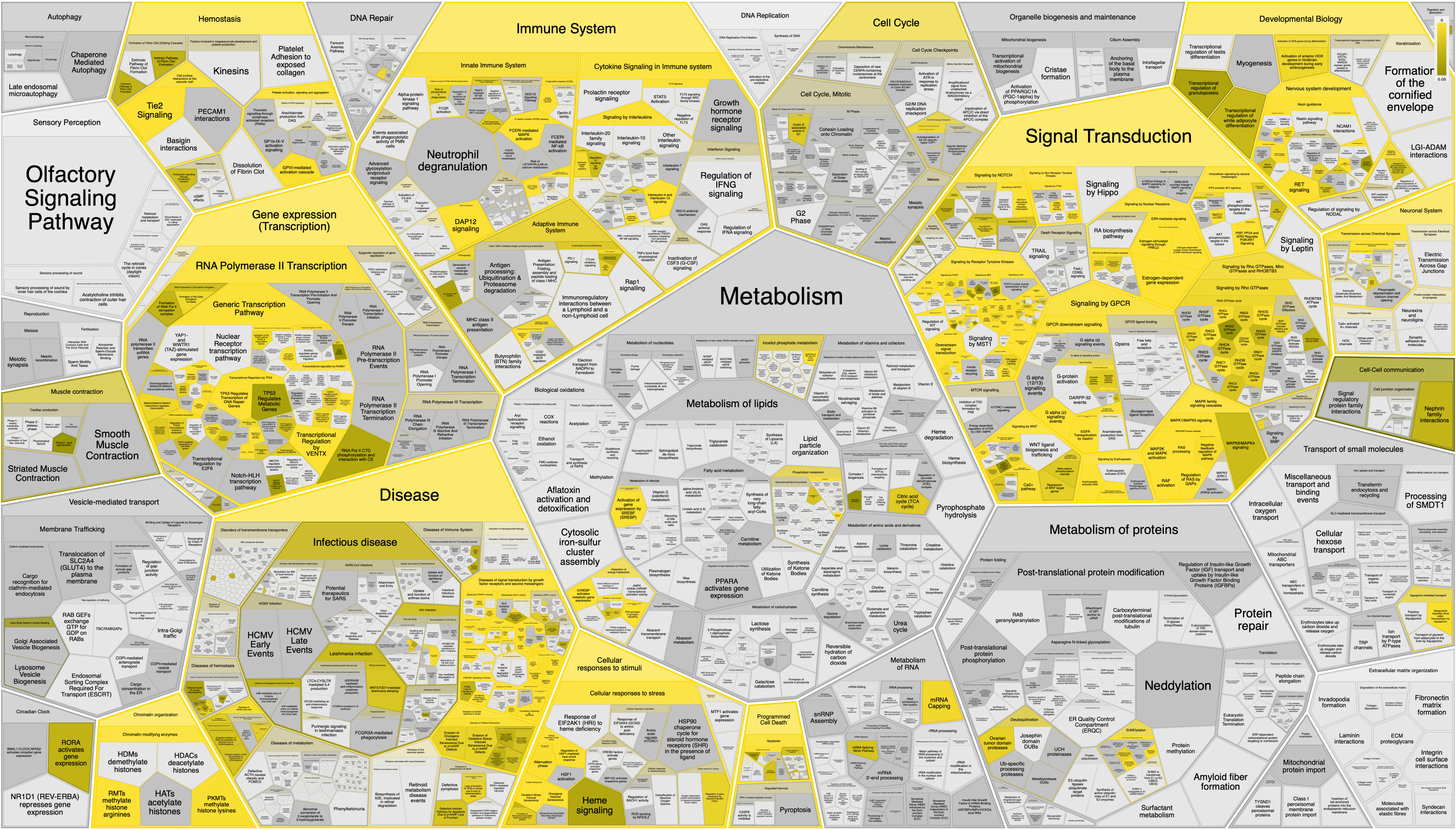

Driver chromosome losses
1
12
3
22
17
2
10
13
18
14
0.00
0.01
0.02
0.03
0.04
0.05
0.06
NDSI


Driver chromosome gains
17
11
10
4
16
18
9
15
19
5
0.00
0.01
0.02
0.03
0.04
0.05
0.06
NDSI

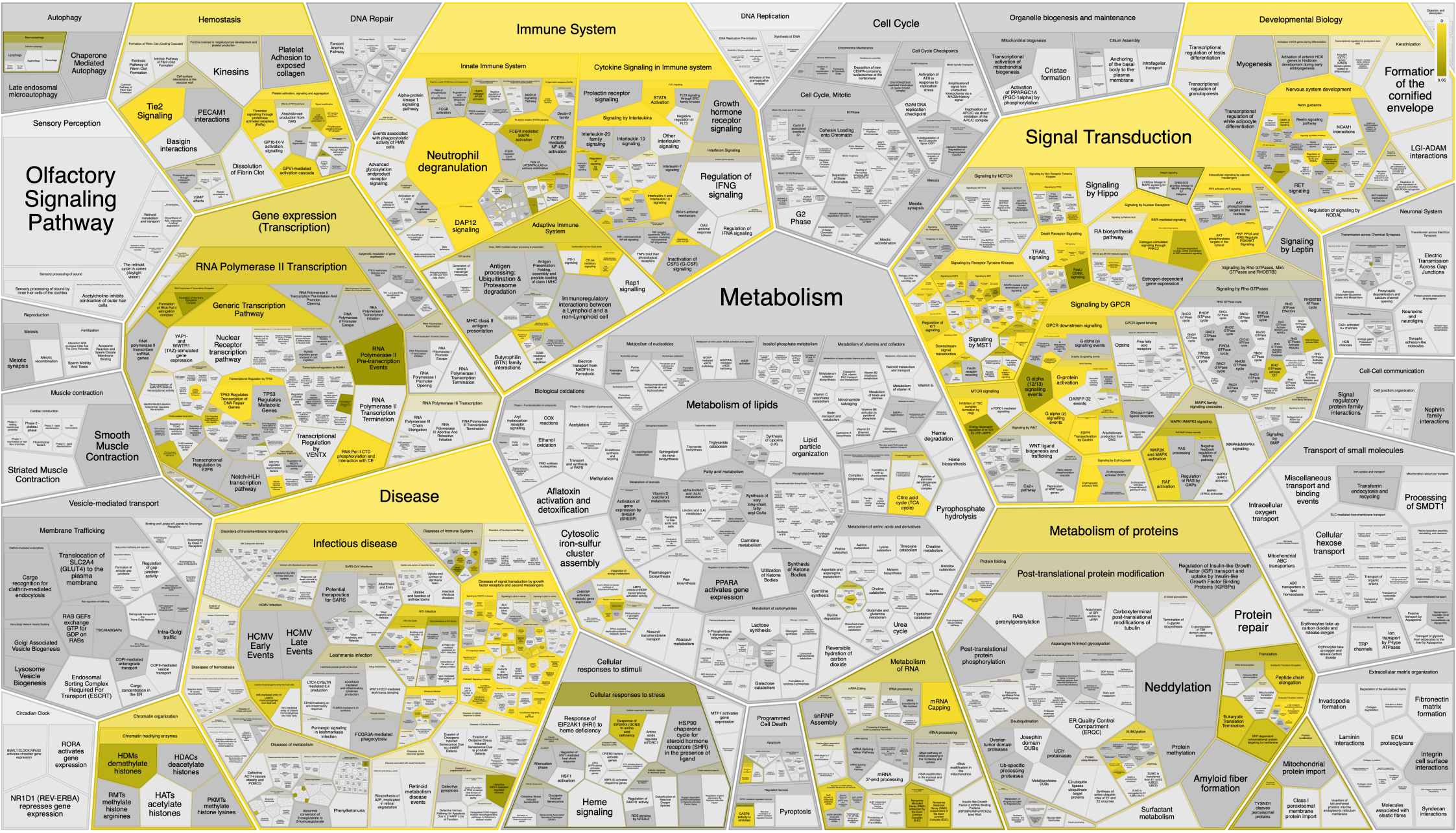

Driver arm losses
19q
12p
20p
12q
2p
3q
1p
17q
18p
16q
0.00
0.02
0.04
0.06
0.08
0.10
NDSI

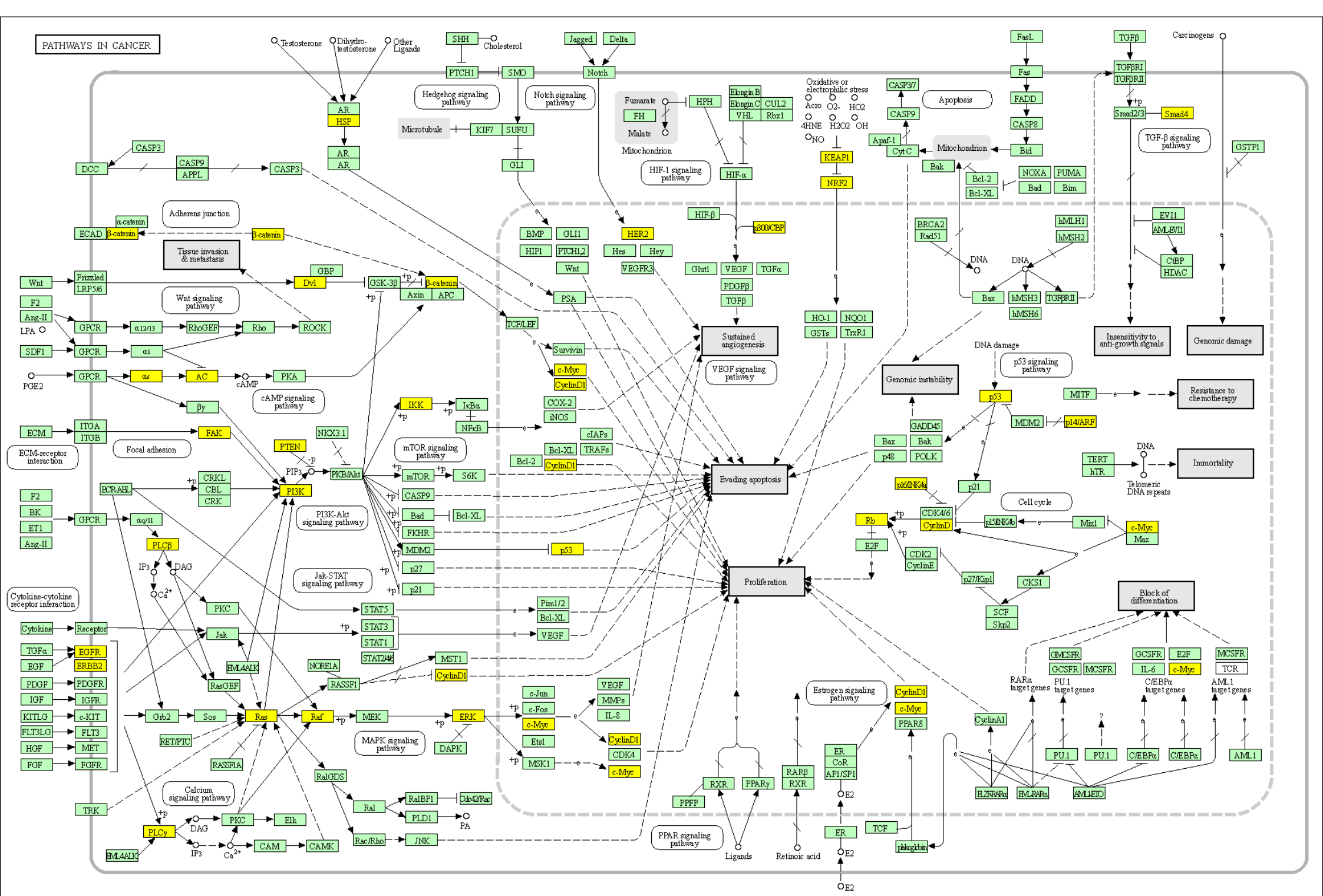

Driver arm gains
19p
17p
5q
4p
4q
9q
8p
16q
9p
16p
0.00
0.01
0.02
0.03
0.04
0.05
0.06
NDSI

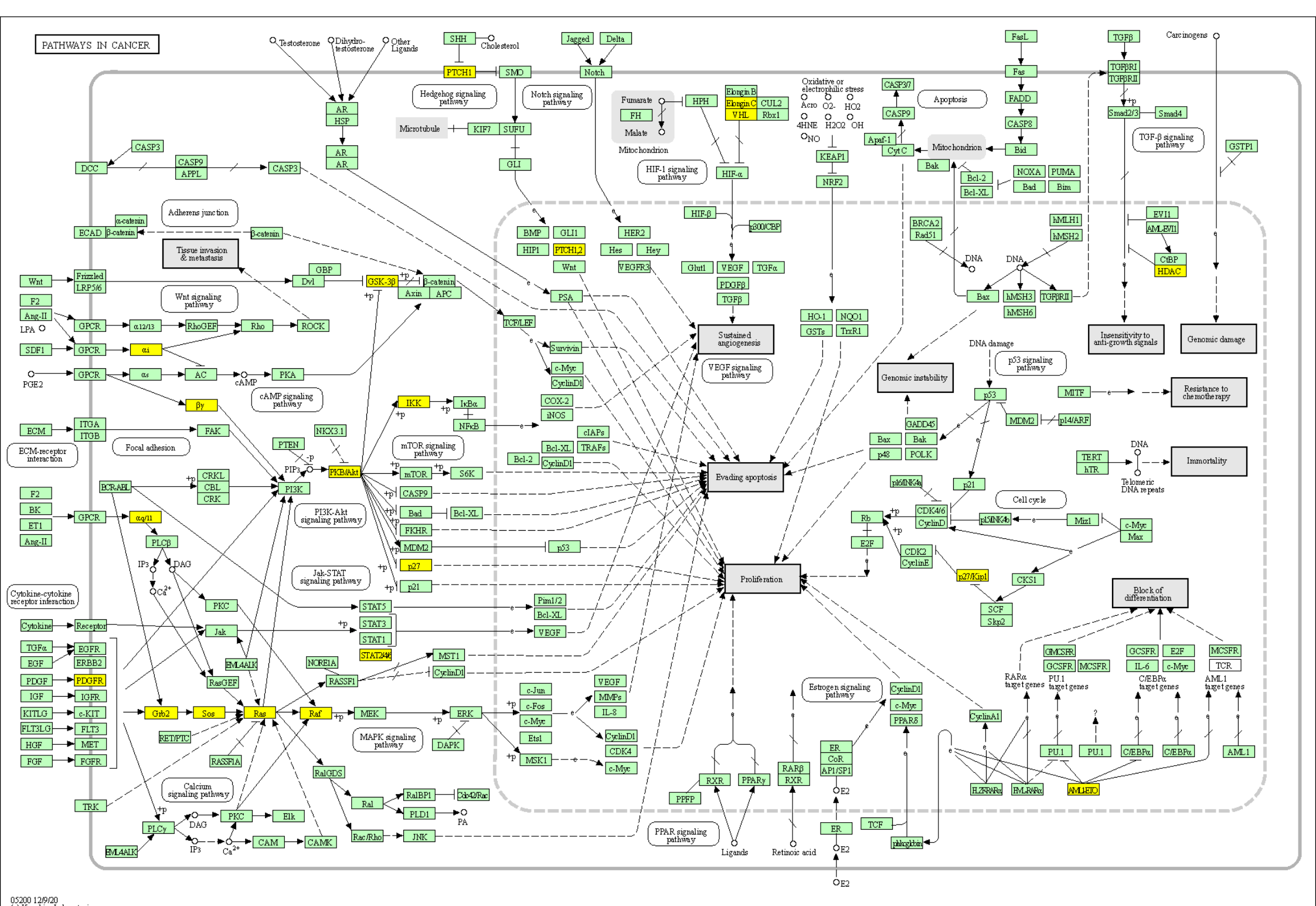

Driver events of all classes
GTF2I
CHEK2
GNAQ
GTF2F2
CSDE1
TRAF2
GBF1
EPRS
HK3
HDAC2
0.0
0.1
0.2
0.3
0.4
0.5
NDSI

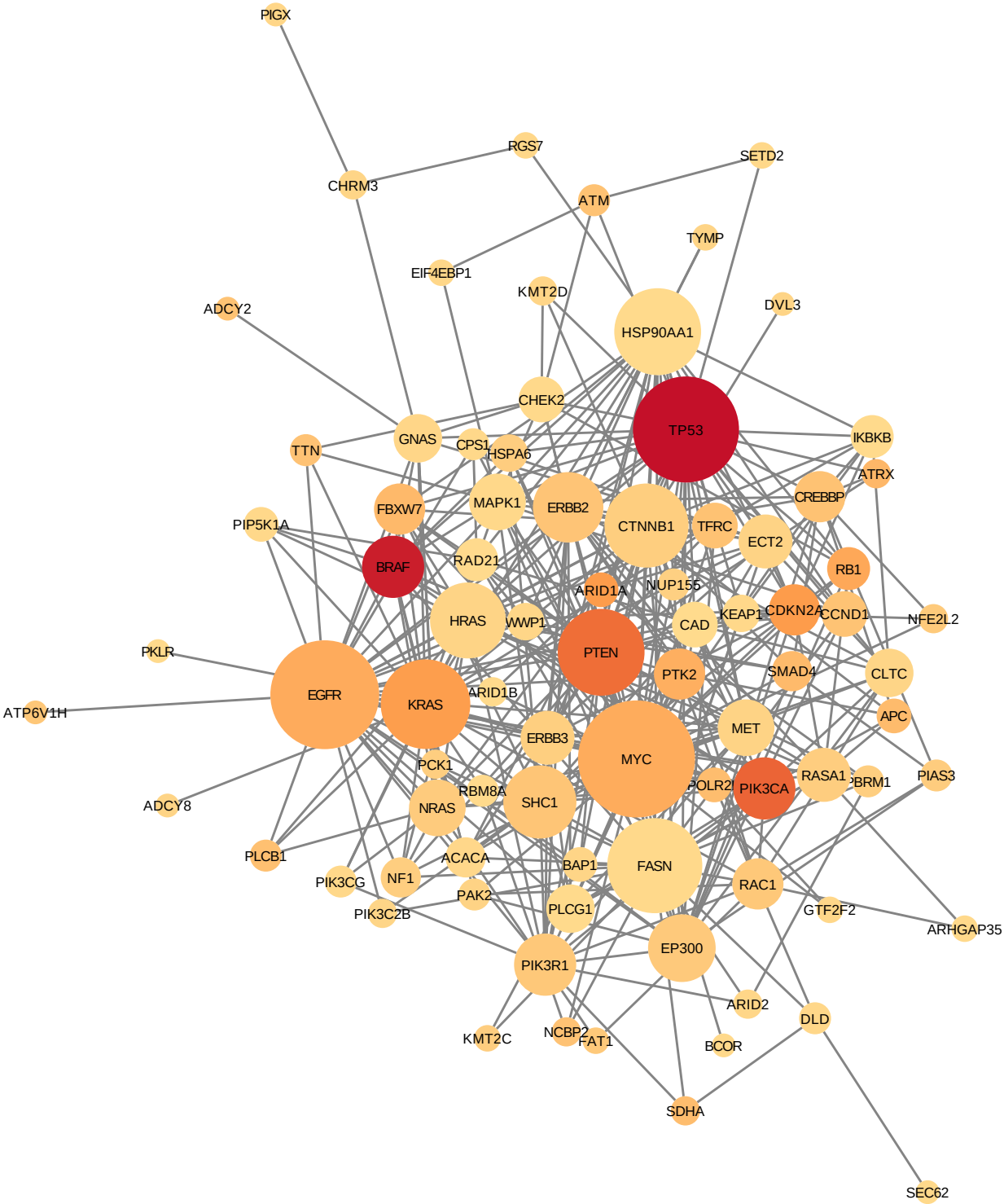


Autophagy
Chaperone Mediated Autophagy
Late endosomal microautophagy
Hemostasis
Platelet Adhesion to exposed collagen
Kinesins
Tie2 Signaling
PECAM1 interactions
Basigin interactions
Dissolution of Fibrin Clot
DNA Repair
Immune System
Innate Immune System
Cytokine Signaling in Immune system
Neutrophil degranulation
Prolactin receptor signaling
Growth hormone receptor signaling
Regulation of IFNG signaling
Signaling by Interleukins
DAP12 signaling
Adaptive Immune System
Antigen processing: Ubiquitination & Proteasome degradation
Rap1 signaling
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell
MHC class II antigen presentation
Butyrophilin (BTN) family interactions
DNA Replication
Cell Cycle
Cell Cycle, Mitotic
Cell Cycle Checkpoints
G2 Phase
Organelle biogenesis and maintenance
Cristae formation
Developmental Biology
Myogenesis
Formation of the cornified envelope
Signal Transduction
Signaling by Hippo
LGI-ADAM interactions
Neuronal System
Signaling by Leptin
Signaling by GPCR
Signaling by MST1
Sensory Perception
Olfactory Signaling Pathway
Gene expression (Transcription)
RNA Polymerase II Transcription
Generic Transcription Pathway
Nuclear Receptor transcription pathway
TP53 Regulates Metabolic Genes
Transcriptional Regulation by VENTX
Notch-HLH transcription pathway
RNA Polymerase II Pre-transcription Events
RNA Polymerase II Transcription Termination
Metabolism
Metabolism of lipids
Lipid particle organization
Biological oxidations
Aflatoxin activation and detoxification
Cytosolic iron-sulfur cluster assembly
PPARA activates gene expression
Pyrophosphate hydrolysis
Heme degradation
Metabolism of RNA
Metabolism of proteins
Post-translational protein modification
Protein repair
Neddylation
Amyloid fiber formation
Surfactant metabolism
Reproduction
Muscle contraction
Smooth Muscle Contraction
Striated Muscle Contraction
Vesicle-mediated transport
Membrane Trafficking
Disease
Infectious disease
HCMV Early Events
HCMV Late Events
Leishmania infection
Cellular responses to stimuli
Cellular responses to stress
Heme signaling
Programmed Cell Death
Pyroptosis
mRNA Capping
snRNP Assembly
Chromatin organization
HDMs demethylate histones
HDACs deacetylate histones
HATs acetylate histones
RMTs methylate histone arginines
PKMTs methylate histone lysines
Circadian Clock
RORA activates gene expression
NR1D1 (REV-ERBA) represses gene expression
Transport of small molecules
Miscellaneous transport and binding events
Processing of SMDT1
Cellular hexose transport
Cell-Cell communication
Nephrin family interactions
Extracellular matrix organization
Fibronectin matrix formation
Mitochondrial protein import
Invadopodia formation
Laminin interactions
ECM proteoglycans
Integrin cell surface interactions


Olfactory Signaling Pathway
Sensory Perception
Autophagy
Chaperone Mediated Autophagy
Late endosomal microautophagy
Hemostasis
Tie2 Signaling
Kinesins
Platelet Adhesion to exposed collagen
PECAM1 interactions
Basigin interactions
Dissolution of Fibrin Clot
GPVI-mediated activation cascade
DNA Repair
Immune System
Innate Immune System
Cytokine Signaling in Immune system
Neutrophil degranulation
DAP12 signaling
Adaptive Immune System
Antigen processing: Ubiquitination & Proteasome degradation
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell
MHC class II antigen presentation
Butyrophilin (BTN) family interactions
Growth hormone receptor signaling
Regulation of IFNG signaling
Regulation of IFNA signaling
Rap1 signalling
DNA Replication
Cell Cycle
Cell Cycle, Mitotic
G2 Phase
Organelle biogenesis and maintenance
Cristae formation
Cilium Assembly
Developmental Biology
Formation of the cornified envelope
Myogenesis
Nervous system development
Axon guidance
NCAM1 interactions
LGI-ADAM interactions
Neuronal System
Signal Transduction
Signaling by Hippo
Signaling by NOTCH
Signaling by Nuclear Receptors
Signaling by GPCR
Signaling by Leptin
Signaling by Rho GTPases
Signaling by MST1
Signaling by Rho GTPases, Miro GTPases and RHOBTB3
Cell-Cell communication
Gene expression (Transcription)
RNA Polymerase II Transcription
Generic Transcription Pathway
Nuclear Receptor transcription pathway
RNA Polymerase II Pre-transcription Events
Transcriptional Regulation by VENTX
Notch-HLH transcription pathway
RNA Polymerase II Transcription Termination
RNA Polymerase III Transcription
Metabolism
Metabolism of lipids
Lipid particle organization
Aflatoxin activation and detoxification
Cytosolic iron-sulfur cluster assembly
PPARA activates gene expression
Pyrophosphate hydrolysis
Heme degradation
Heme biosynthesis
Citric acid cycle (TCA cycle)
Metabolism of proteins
Post-translational protein modification
Protein repair
Neddylation
Amyloid fiber formation
Metabolism of RNA
mRNA Capping
Disease
Infectious disease
HCMV Early Events
HCMV Late Events
Potential therapeutics for SARS
Cellular responses to stimuli
Cellular responses to stress
Heme signaling
Response of EIF2AK1 (HRI) to heme deficiency
HSP90 chaperone cycle for steroid hormone receptors (SHR) in the presence of ligand
Programmed Cell Death
Pyroptosis
Vesicle-mediated transport
Membrane Trafficking
Translocation of SLC2A4 (GLUT4) to the plasma membrane
Golgi Associated Vesicle Biogenesis
Lysosome Vesicle Biogenesis
Endosomal Sorting Complex Required For Transport (ESCRT)
Intra-Golgi traffic
Circadian Clock
Chromatin organization
Chromatin modifying enzymes
HDMs demethylate histones
HDACs deacetylate histones
RMTs methylate histone arginines
HATs acetylate histones
PKMTs methylate histone lysines
NR1D1 (REV-ERBA) represses gene expression
RORA activates gene expression
Reproduction
Meiosis
Meiotic synapsis
Muscle contraction
Smooth Muscle Contraction
Striated Muscle Contraction
Transport of small molecules
Miscellaneous transport and binding events
Processing of SMDT1
Cellular hexose transport
Intracellular oxygen transport
Mitochondrial protein import
Class I peroxisomal membrane protein import
Extracellular matrix organization
Invadopodia formation
Fibronectin matrix formation
Laminin interactions
ECM proteoglycans
Integrin cell surface interactions
Molecules associated with elastic fibres
Syndecan interactions
Nephrin family interactions
Neurexins and neuroligins
Synaptic adhesion-like molecules
Electric Transmission Across Gap Junctions
Transmission across Chemical Synapses
Transmission across Electrical Synapses



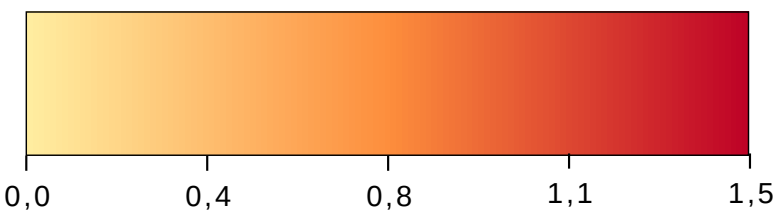

PATHWAYS IN CANCER
Testosterone
Dihydro-testosterone
Other Ligands
AR
HSP
AR
AR
SHH
Cholesterol
PTCH1
SMO
Hedgehog signaling pathway
Microtubule
KIF7
SUFU
GLI
Jagged
Delta
Notch
Notch signaling pathway
Fumarate
FH
Malate
Mitochondrion
HPH
Elongin B
Elongin C
CUL2
VHL
Rbx1
HIF-1 signaling pathway
HIF-α
HIF-β
p300/CBP
Oxidative or electrophilic stress
Acro
O2-
HO2
4HNE
H2O2
OH
NO
KEAP1
NRF2
CASP3/7
CASP9
Apaf-1
Cyt C
Apoptosis
Mitochondrion
FasL
Fas
FADD
CASP8
Bid
Bak
Bcl-2
Bcl-XL
NOXA
PUMA
Bad
Bim
TGFβ
TGFβRI
TGFβRII
Smad2/3
Smad4
TGF-β signaling pathway
Carcinogens
GSTP1
EVI1
AML-EVI1
CtBP
HDAC
BRCA2
Rad51
DNA
hMLH1
hMSH2
hMSH3
hMSH6
TGFβRII
Bax
HO-1
NQO1
GSTs
TrxR1
Insensitivity to anti-growth signals
Genomic damage
DCC
CASP3
CASP9
APPL
CASP3
ECAD
α-catenin
β-catenin
β-catenin
Adherens junction
Tissue invasion & metastasis
Wnt
Frizzled
LRP5/6
GBP
Dvl
GSK-3β
Axin
APC
β-catenin
Wnt signaling pathway
F2
Ang-II
LPA
GPCR
α12/13
RhoGEF
Rho
ROCK
SDF1
GPCR
αi
PGE2
GPCR
αs
AC
cAMP
PKA
cAMP signaling pathway
βγ
TCF/LEF
PSA
BMP
GLI1
HIP1
PTCH1,2
Wnt
HER2
Hes
Hey
VEGFR3
Glut1
VEGF
TGFα
PDGFβ
TGFβ
Sustained angiogenesis
VEGF signaling pathway
Survivin
c-Myc
CyclinD1
IKK
IκBα
NFκB
COX-2
iNOS
cIAPs
TRAFs
Bcl-2
Bcl-XL
CyclinD1
ECM
ITGA
ITGB
FAK
ECM-receptor interaction
Focal adhesion
NKX3.1
PTEN
PIP3
PKB/Akt
mTOR signaling pathway
mTOR
S6K
CASP9
Bad
Bcl-XL
FKHR
MDM2
p53
p27
p21
Evading apoptosis
Genomic instability
DNA damage
p53 signaling pathway
p53
MDM2
p14/ARF
MITF
Resistance to chemotherapy
GADD45
Bax
Bak
p48
POLK
TERT
hTR
DNA
Telomeric DNA repeats
Immortality
p16/INK4a
p21
Cell cycle
Rb
E2F
CDK4/6
CyclinD
p15/INK4b
Miz1
c-Myc
Max
CDK2
CyclinE
p27/Kip1
CKS1
SCF
Skp2
Proliferation
Block of differentiation
BCR-ABL
CRKL
CBL
CRK
PI3K
PI3K-Akt signaling pathway
F2
BK
ET1
Ang-II
GPCR
αq/11
PLCβ
IP3
DAG
Ca2+
PKC
Jak-STAT signaling pathway
Cytokine-cytokine receptor interaction
Cytokine
Receptor
Jak
STAT5
STAT3
STAT1
STAT2/p
Pim1/2
Bcl-XL
VEGF
TGFα
EGF
EGFR
ERBB2
PDGF
PDGFR
IGF
IGFR
KITLG
c-KIT
FLT3LG
FLT3
HGF
MET
FGF
FGFR
EML4ALK
RasGEF
Grb2
Sos
Ras
Raf
MEK
ERK
DAPK
MAPK signaling pathway
RET/PTC
RASSF1A
NORE1A
RASSF1
MST1
CyclinD1
c-Jun
c-Fos
c-Myc
Etsl
MSK1
VEGF
MMPs
IL-8
CyclinD1
CDK4
c-Myc
RalGDS
Ral
RalBP1
Cdc42/Rac
PLD1
PA
Rac/Rho
JNK
Calcium signaling pathway
TRK
PLCγ
DAG
IP3
Ca2+
PKC
Elk
CAM
CAMK
EML4ALK
PPAR signaling pathway
PPFP
RXR
PPARγ
Ligands
RARβ
RXR
Retinoic acid
Estrogen signaling pathway
ER
CoR
AP1/SP1
E2
ER
E2
E2
TCF
ydroxyldoin
CyclinD1
c-Myc
PPARδ
CyclinA1
RARα target genes
GMCSFR
GCSFR
MCSFR
PU.1 target genes
GCSFR
IL-6
E2F
c-Myc
C/EBPα target genes
MCSFR
TCR
AML1 target genes
PU.1
PU.1
C/EBPα
C/EBPα
AML1
PLZF/RARα
PML/RARα
AML1/ETO
05200 12/9/20
(c) Kanehisa Laboratories

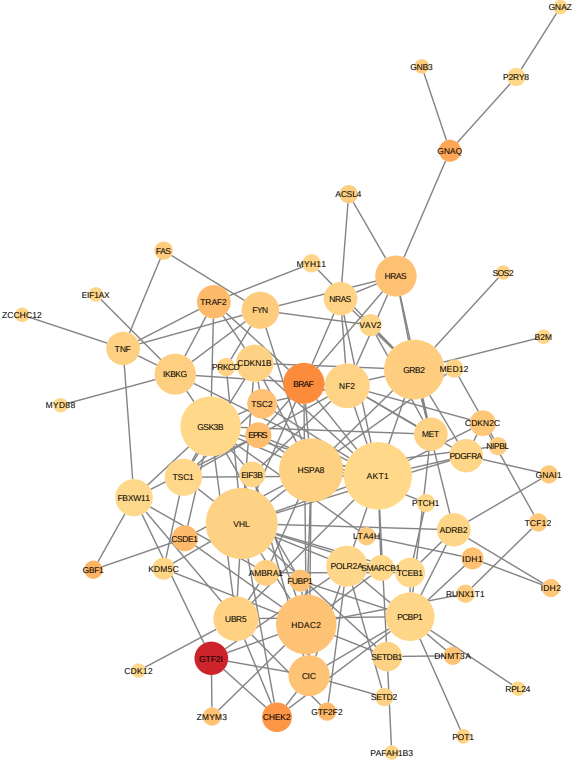


BioGRID: Protein-Protein Interactions (H. sapiens)

yFiles Organic Layout

Node Fill Color: DSI





BioGRID: Protein-Protein Interactions (H. sapiens)

yFiles Organic Layout

Node Fill Color: NDSI