Current repo: https://github.com/laurenmyersxd/Explainable_AI_Genomics_Study

Breast cancer (Breast-AdenoCa):

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
snv_ms_driver_CBFB_Intron	0.3368811095521300	0.3368811095521300	Yes
snv_ms_driver_RUNX1_Intron	0.3319197663810530	0.3319197663810530	Yes
snv_ms_driver_ESR1_Intron	0.2939985522241130	0.2939985522241130	Yes
snv_ms_driver_BRIP1_Intron	0.2921886714121990	0.2921886714121990	Yes
snv_ms_driver_GATA3_Coding	0.2865230742880090	0.2865230742880090	Yes
snv_ms_driver_NCOR1_Intron	0.2848285920457750	0.2848285920457750	Yes
snv_ms_driver_MAP3K1_Intron	0.2780036798641150	0.278003679864115	Proxy
cna_total_len_Mb	-0.2405017889384050	0.24050178893840500	Yes
snv_ms_driver_HERC2_Intron	0.2401149562071560	0.2401149562071560	Yes
sv_TRA	0.2397203464555420	0.23972034645554200	Proxy

1. **snv_ms_driver_CBFB_Intron**: https://pubmed.ncbi.nlm.nih.gov/32711101/ As a critical transcription factor, CBFB (core binding factor subunit β) is frequently mutated in breast cancer and considered to be of significance in the pathogenesis of cancer
2. **snv_ms_driver_RUNX1_Intron**: https://pmc.ncbi.nlm.nih.gov/articles/PMC5514882/ RUNX1 is the predominant RUNX protein expressed in the normal human breast epithelium and it is the only RUNX family member for which somatic mutations have been identified in human breast cancer. These mutations are primarily loss of function mutations that occur through nonsense, frameshift or missense mutations within the Runt DNA-binding domain
3. **snv_ms_driver_ESR1_Intron**: https://www.nature.com/articles/s41467-022-29498-9 Estrogen receptor alpha (ER/ESR1) is frequently mutated in endocrine resistant ER-positive (ER+) breast cancer and linked to ligand-independent growth and metastasis.

4. **snv_ms_driver_BRIP1_Intron BRIP1** https://pmc.ncbi.nlm.nih.gov/articles/PMC8619737/ (Breast Cancer 1 Interacting Helicase 1) is a tumor suppressor gene that has vital function in preserving the genetic stability by repairing DNA damage though have significant associations with the onset of breast cancer (BC) if mutated or overexpressed.
5. **snv_ms_driver_GATA3_Coding** https://pmc.ncbi.nlm.nih.gov/articles/PMC4758516/ Recent genomic analysis of human breast cancers has revealed high-frequency mutation in GATA3 in luminal tumors, suggesting “driver” function(s).
6. **snv_ms_driver_NCOR1_Intron** https://pubmed.ncbi.nlm.nih.gov/16019133/ Univariate and multivariate prognostic analysis demonstrated that NCOR1 mRNA is an independent prognostic factor for breast cancer.
7. **snv_ms_driver_MAP3K1_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC8756433/ Mutations in MAP3K1 were predominant in patients with the luminal A and B breast cancer subtypes in METABRIC datasets ($P<0.001$), although no significant differences were observed in the GDPH cohort ($P=0.227$)
8. **cna_total_len_Mb** https://pmc.ncbi.nlm.nih.gov/articles/PMC4996377/ Identification of CNAs associated with each subtype was performed by analyzing each subtype separately from the others and by taking the rest of the subtypes as the control. Our results found a new amplification in 11q at the location of the progesterone receptor in the Luminal A subtype.
9. **snv_ms_driver_HERC2_Intron** https://pubmed.ncbi.nlm.nih.gov/24508693/ (human epidermal growth factor receptor 2) receptor is a membrane tyrosine kinase and when activated affects cell proliferation and survival. The HER2 oncogene is located on chromosome 17q12. HER2 amplification is the primary pathway of HER2 receptor overexpression and is a major driver of tumor development and progression in a subset of breast cancers
10. **sv_TRA** https://www.nature.com/articles/s41586-023-06057-w Recurrent amplification boundaries and rearrangement hotspots correlate with oestrogen receptor binding in breast cancer cell

Pilocytic Astrocytoma (CNS-PiloAstro)

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
snv_ms_SBS5	-0.7256925623042500	0.7256925623042500	Yes
snv_ms_driver_BRAF_Intron	0.6182114810599720	0.6182114810599720	Yes
snv_ms_SBS1	-0.6069581796584750	0.6069581796584750	Yes
cna_loss_Mb	-0.5923609831946590	0.5923609831946590	Yes
snv_ms_driver_STAG2_Intron	-0.5845767888747300	0.5845767888747300	No
snv_chr_17	-0.5388182077337360	0.5388182077337360	Yes
snv_ms_SBS8	-0.5148253354895820	0.5148253354895820	Proxy
snv_chr_14	-0.4980643968713260	0.4980643968713260	Yes
cna_frac_altered	-0.4768266482569870 0	0.4768266482569870 0	Yes
snv_chr_12	-0.4670732565487960	0.4670732565487960	No

1. **snv_ms_SBS5** https://pmc.ncbi.nlm.nih.gov/articles/PMC9970869/ The lowest overall (note the negative weight) somatic mutation burden was observed in pilocytic astrocytoma (median 0.034 total mutations per megabase)
2. **snv_ms_driver_BRAF_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC2577184/ Brain tumours are the commonest solid tumours of childhood, and pilocytic astrocytomas (PAs) are the most common central nervous system tumour in 5-19 year-olds. Little is known about the genetic alterations underlying their development. Here we describe a tandem duplication of ~2Mb at 7q34 occurring in 66% of pilocytic astrocytomas. This rearrangement, which was not observed in a series of 244 higher-grade astrocytomas, results in an in-frame fusion gene incorporating the kinase domain of the BRAF oncogene.
3. **snv_ms_SBS1** https://www.nature.com/articles/s43018-022-00509-4 Amongst these, SBS1 and SBS5 were present in 98.2% and 96.6% of samples across the cohort,

respectively (Fig. 2b and Supplementary Tables 6 and 7). As described in adult cancers5,7, and for a small (note small magnitude) fraction of pediatric brain tumors2,

4. **cna_loss_Mb** https://pmc.ncbi.nlm.nih.gov/articles/PMC2761618/ No copy number abnormality was seen in 64% of cases at this resolution. However, whole chromosomal gain (median 5 chromosomes affected) occurred in 32% of tumors.
5. **snv_ms_driver_STAG2_Intron** -
6. **Snv_chr_17** . https://pubmed.ncbi.nlm.nih.gov/8103960/ Allelic loss was observed on chromosome 17 in four cases (three sporadic, one NF1); all lost portions of the long arm in chromosome 17, and one tumor lost the short arm as well. One tumor showed an interstitial deletion on the long arm that included the region of the NF1 gene. These data suggest the presence of a tumor suppressor gene on 17q that is associated with pilocytic astrocytomas.
7. **snv_ms_SBS8** https://www.nature.com/articles/s42003-020-01119-5 While SBS8 is uncommon among mutations in non-malignant tissues, in tumor genomes its proportions increase with replication timing and speed, and checkpoint defects further promote this signature - suggesting that SBS8 probably arises due to uncorrected late replication errors during cancer progression.
8. **Snv_chr_14** https://pmc.ncbi.nlm.nih.gov/articles/PMC5009478/ The most frequent genetic alteration identified in pediatric pilocytic astrocytomas and pilomyxoid variant is the KIAA1549-BRAF fusion, which typically results from a 2.0 Mb tandem duplication in chromosome band 7q34.
9. **Cna_frac_altered** https://pubmed.ncbi.nlm.nih.gov/18398503/ Both the stable silencing of BRAF through shRNA lentiviral transduction and pharmacological inhibition of MEK1/2, the immediate downstream phosphorylation target of BRAF, blocked the proliferation and arrested the growth of cultured tumor cells derived from low-grade gliomas.
10. **Snv_chr_12** Not enough: https://pubmed.ncbi.nlm.nih.gov/17712732/ We screened 25 WHO I and II astrocytomas for mutations of PTPN11, NRAS, KRAS, and HRAS genes and identified the somatic G12A KRAS mutation in one pilocytic astrocytoma

Esophageal adenocarcinoma (Eso-AdenoCa)

*really low weights here!

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
snv_ms_driver_SRGAP3_Intron	0.16840343465161300	0.16840343465161300	No
snv_ms_driver_WWOX_Intron	0.1593187634727600	0.1593187634727600	Yes
snv_ms_driver_ATR_Intron	0.14430483582886700	0.14430483582886700	No
snv_ms_driver_MGA_Intron	0.13850037753774500	0.13850037753774500	Proxy
snv_ms_driver_MSR1_Intron	0.13624429021686300	0.13624429021686300	Yes
sv_DEL	0.12569009266022000	0.12569009266022000	Yes
snv_ms_INS	0.12564898080392700	0.12564898080392700	Yes
snv_ms_driver_ADAM10_Intron	0.12220683905449900	0.12220683905449900	Proxy
sv_total	0.12007744586593600	0.12007744586593600	Proxy
snv_ms_driver_DLG1_Intron	0.11800424507542400	0.11800424507542400	Yes

1. **snv_ms_driver_SRGAP3_Intron** -
2. **snv_ms_driver_WWOX_Intron** https://www.nature.com/articles/6602023 WWOX alterations are common in squamous cell carcinoma of the lung and oesophageal cancer, both of which are related to tobacco smoke and alcohol consumption
3. **snv_ms_driver_ATR_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC2663384/ The ATR and ATM protein kinases are master regulators of the DNA damage response, signaling to control cell cycle transitions, DNA replication, DNA repair, and apoptosis. Recent studies have provided insights into the mechanisms controlling ATR activation, helped to explain the overlapping but non-redundant activities of ATR and ATM in DNA damage signaling, and clarified the critical functions of ATR in maintaining genome integrity.
   * Evidence of DNA damage but not specific to any 1 cancer.
4. **snv_ms_driver_MGA_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC4048962/ Other examples are MGA, whose product competes with Myc for binding to Max and which resides in small focal deletions (containing ≤4 genes) in ovarian and various epithelial

cancers; the interferon regulatory factor IRF6, which is known to have tumor suppressive roles in keratinocytes and is mutated in head and neck squamous cancer; and the delta/notch-like EGF-repeat gene DNER.

5. **snv_ms_driver_MSR1_Intron** https://pubmed.ncbi.nlm.nih.gov/21791690/ MSR1 was significantly associated with the presence of BE/EAC in derivation and validation samples, although it was only present in a small percentage of the cases.
6. **sv_DEL** https://www.nature.com/articles/s42003-022-03238-7 After excluding fragile sites, we identified 51 candidate new drivers in genomic regions disrupted by SVs, including ETV5, KAT6B and CLTC. RUNX1 was the most recurrently altered gene (24%), with many deletions inactivating the RUNT domain but preserved the reading frame, suggesting an altered protein product.
7. **snv_ms_INS** https://www.nature.com/articles/s42003-022-03238-7 Oesophageal cancer, especially the subtype oesophageal adenocarcinoma (OAC), emerged from the PCAWG analysis (n = 100 OACs) as a cancer type with one of the highest burdens of SVs with complex rearrangements1,3. These include breakage-fusion-bridge (BFB) cycles; catastrophic chromothripsis events with oscillating copy number patterns8, deletions in the fragile-sites and the highest rate of somatic mobile element (ME) inserts of any cancer type
8. **snv_ms_driver_ADAM10_Intron** https://pubmed.ncbi.nlm.nih.gov/26986985/ Emerging evidence suggests that ADAM10 is overexpressed in a variety of cancers and participates in tumor progression (18–30). Thus, IHC was initially performed to determine the protein expression level of ADAM10 in ESCC tissues and corresponding normal esophageal mucosal tissues.
9. **Sv_total** https://pmc.ncbi.nlm.nih.gov/articles/PMC10512880/ Further research is required to elucidate the functional implications of these genetic variations to develop targeted therapies that can improve the prognosis of patients with esophageal cancer.
10. **snv_ms_driver_DLG1_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC6186307/ Our study indicated that upregulation of circ-DLG1 promoted esophageal cell proliferation ability and might serve as a novel biomarker for ESCC.

Renal Cell Carcinoma (Kidney-RCC)

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
snv_ms_driver_VHL_Coding	0.769455762301192	0.769455762301192	Yes
snv_ms_SBS40	0.49293457516447600	0.49293457516447600	Yes
snv_ms_DEL	0.386055645806238	0.386055645806238	Yes
snv_ms_driver_PBRM1_Coding	0.37567001672393000	0.37567001672393000	Yes
snv_ms_DNP	0.29756568291899200	0.29756568291899200	No
snv_chr_22	0.28983880958354400	0.28983880958354400	Yes
snv_ms_driver_TSC2_Intron	0.2801363224884080	0.2801363224884080	Yes
snv_ms_driver_PBRM1_Intron	0.26746412304770400	0.26746412304770400	Yes
snv_ms_driver_PTEN_Intron	0.2657826686206660	0.2657826686206660	Yes
cna_total_len_Mb	-0.2487437669853370	0.2487437669853370	Yes

1. **snv_ms_driver_VHL_Coding** https://pmc.ncbi.nlm.nih.gov/articles/PMC6278085/ VHL mutations were detected in 26/55 (47%) RCC patients.
2. **snv_ms_SBS40** https://www.nature.com/articles/s41586-024-07368-2 Other established measures of kidney function, including cystatin C and creatinine, were correlated with TMAP (P value = 2.5 × 10−30 and 1.7 × 10−69, respectively) and also showed evidence of positive association with SBS40b (P value = 0.023 and 0.058, respectively).
3. **snv_ms_DEL** https://pmc.ncbi.nlm.nih.gov/articles/PMC7186632/ RCC is a tumor with a relatively low TMB, with a median of 1.1 mutations per Mb (3,5). Within the main RCC subtypes, chromophobe RCC carries the lowest TMB (<1 mutation per Mb), while CCRCC and papillary RCC exhibit a similar range of TMB
4. **snv_ms_driver_PBRM1_Coding** https://pmc.ncbi.nlm.nih.gov/articles/PMC8383204/ Renal cell carcinoma is a common solid tumor. PBRM1 is one of the most mutation-prone genes in clear cell renal cell carcinoma (ccRCC) with the occurrence of mutation in 40% of ccRCC patients.

5. **snv_ms_DNP** -
6. **Snv_chr_22** https://pmc.ncbi.nlm.nih.gov/articles/PMC6362797/ Collecting duct carcinoma shows a wide variety of aberrations involving chromosomes 1, X, Y with either translocations or deletions. Furthermore, chromosomes 13 and 22 are affected
7. **snv_ms_driver_TSC2_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC8535193/ TSC1 and TSC2 mutations occur in clear cell RCC, although at a low frequency (2–5% [26,27,28,29,30], Table 2).
8. **snv_ms_driver_PBRM1_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC8383204/ PBRM1 is one of the most mutation-prone genes in clear cell renal cell carcinoma (ccRCC) with the occurrence of mutation in 40% of ccRCC patients
9. **snv_ms_driver_PTEN_Intron** https://pubmed.ncbi.nlm.nih.gov/14598361/ PTEN is a candidate tumor suppressor gene in a variety of malignant tumors, including renal cell carcinoma (RCC).
10. **cna_total_len_Mb** https://pubmed.ncbi.nlm.nih.gov/33668731/#:~:text=Most%20patients%20in%20the%20cohort,cell%20carcinoma;%20copy%20number%20aberrations. Bioinformatics analysis revealed 19 genes mapping to CNA significant regions, including SETD2, BAP1, FLT4, PTEN, FGFR4 and NSD1. Moreover, gain of 5q34-q35.3 (FLT4 and NSD1) and loss of 6q23.2-q23.3 (MYB) and 9p21.3 (MLLT3) had gene expression levels that correlated with TCGA data and was also associated with advanced disease features, such as larger tumors, Fuhrman 3, metastasis at diagnosis and death….

Hepatocellular carcinoma (Liver-HCC)

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
snv_ms_SBS12	0.3615360718149640	0.3615360718149640	Yes
snv_ms_driver_ALB_Intron	0.33666858742897400	0.33666858742897400	Yes
snv_ms_driver_APOB_Intron	0.25194827412044000	0.25194827412044000	Yes
snv_ms_driver_ACVR2A_Intron	0.23295895970951200	0.23295895970951200	Yes
snv_ms_driver_ARID1B_Intron	0.23294337493623600	0.23294337493623600	Proxy
snv_ms_DNP	0.23251074439315400	0.23251074439315400	Yes
snv_ms_driver_HNF4A_Intron	0.21970389780193400	0.21970389780193400	Yes
snv_ms_driver_RTN4_Intron	0.21872017599974700	0.21872017599974700	Yes
snv_ms_driver_DYRK1A_Intron	0.2038314524989900	0.2038314524989900	Yes
snv_ms_driver_MACF1_Intron	0.20121107203836300	0.20121107203836300	Yes

1. **snv_ms_SBS12** https://pmc.ncbi.nlm.nih.gov/articles/PMC9445436/ Meanwhile, SBS12 and SBS40 are also closely related to cancers, although their etiologies are still not clearly identified, SBS12 contributes to a small proportion (<20%) of the mutations of liver cancer and SBS40 is correlated with patients' ages of some cancers.
2. **snv_ms_driver_ALB_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC9299813/ Worldwide, hepatocellular carcinoma (HCC) is one of the most common causes of death in people. Albumin (ALB) is considered as an important indicator for HCC prognosis, and evidence has shown HCC cell growth can be regulated by ALB.
3. **snv_ms_driver_APOB_Intron** https://www.nature.com/articles/s12276-018-0174-2 Recent findings from The Cancer Genome Atlas project have provided a comprehensive map of genomic alterations that occur in hepatocellular carcinoma (HCC), including unexpected mutations in apolipoprotein B (APOB).
4. **snv_ms_driver_ACVR2A_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC12047472/ Although ACVR2A mutations are prevalent in non-viral hepatocellular carcinomas (HCCs), the underlying mechanism remains unelucidated. Our molecular investigation

reveals that ACVR2A impairment induces hyperglycolysis through the inactivation of the SMAD signaling pathway.

5. **snv_ms_driver_ARID1B_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC5458188/ The tumor suppressor role of AT-rich interactive domain containing protein 1B (ARID1B) has drawn much attention in area of cancer etiology. However, it had remained unknown whether or not genetic variants of ARID1B involved in development of hepatocellular carcinoma (HCC).
6. **snv_ms_DNP** https://pmc.ncbi.nlm.nih.gov/articles/PMC9445436/ Additionally, differentially expressed miRNAs are inclined to concentrate in cancer-related signaling pathways. Some of these RNAs also serve as prognostic factors that help predict the survival outcome of HCCs with certain mutational signatures.
7. **snv_ms_driver_HNF4A_Intron** https://www.sciencedirect.com/science/article/pii/S0304383525002988 HNF4A acts as HCC oncogene or suppressor dependent on AMPK pathway activity.
8. **snv_ms_driver_RTN4_Intron** https://pubmed.ncbi.nlm.nih.gov/16201230/ Our data indicated RTN4-C gene was expressed differently in hepatocellular carcinoma and its paracancerous tissues. By tranferring mutant p53 protein from nucleus to cytoplasm and decreasing c-Fos, Hsp70 protein expression, RTN4-C inhibited SMMC7721 cells growth and promoted its apoptosis.
9. **snv_ms_driver_DYRK1A_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC9164892/ The results of the current study demonstrated that dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A) was upregulated in HCC tissues compared with normal liver tissues.
10. **snv_ms_driver_MACF1_Intron** https://www.nature.com/articles/s41698-025-01010-8 In the TCGA-LIHC cohort, patients with high CAIPS scores exhibited a higher frequency of SNVs, specifically in genes TP53, MACF1, ZFHX4, and TSC2. TP53, known as an oncogene, has been associated with poor tumor differentiation and microvascular infiltration in Chinese HCC patients38.

Chronic lymphocytic leukemia (Lymph-CLL)

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
snv_ms_driver_STAG2_Intron	0.7912893335258450	0.7912893335258450	Yes
snv_ms_driver_CD36_Intron	0.7908222157986760	0.7908222157986760	Yes
cna_gain_Mb	-0.5956321032400950	0.5956321032400950	Yes
snv_ms_SBS8	0.5812378483904030	0.5812378483904030	Yes
sv_TRA	-0.4829807903598870	0.48298079035988700	Yes
snv_ms_SBS5	0.4371014245701050	0.4371014245701050	yes
snv_ms_SBS40	-0.4175509573628980	0.41755095736289800	Proxy
snv_ms_SBS9	0.3903589623672410	0.39035896236724100	yes
snv_ms_driver_GNAI1_Intron	0.3737518316649080	0.37375183166490800	Proxy
snv_ms_driver_TRIO_Intron	0.3566396979200620	0.3566396979200620	No

1. **snv_ms_driver_STAG2_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC7469420/ Mutations in members of the cohesin complex are known early drivers of myelodysplastic syndromes (MDS) and acute myeloid leukaemia (AML), with STAG2 the most frequently mutated complex member.
2. **snv_ms_driver_CD36_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC5940394/ We found that CD36 levels are higher in CLL cells than in normal B cells, and that small interfering RNA, CD36 neutralizing antibodies or sulfosuccinimidyl oleate (SSO) that inhibits CD36 significantly reduced the oxygen consumption of CLL cells incubated with FA
3. **cna_gain_Mb** https://pubmed.ncbi.nlm.nih.gov/31974198/#:~:text=To%20further%20determine%20the%20clinical,tool%20for%20CLL%20risk%20stratification. Lowering the size cutoff to 1 Mb in 647 patients did not significantly improve risk assessment. Genomic arrays detected more chromosomal abnormalities and performed at least as well in terms of risk stratification compared to simultaneous chromosome banding analysis as determined in 122 patients. Our findings highlight genomic array as an accurate tool for CLL risk stratification.

4. **snv_ms_SBS8** https://www.nature.com/articles/s41588-022-01211-y Considering signatures with known or probable etiology, the most prevalent were SBS5 (clock-like), DBS11 (APOBEC activity) and ID2 followed by other clock-like signatures: SBS1 (deamination of 5-methylcytosines), SBS8, DBS2 and the AID signature SBS9.
5. **sv_TRA** https://pubmed.ncbi.nlm.nih.gov/16179374/ This approach revealed that translocations occurred in 33 of 96 (34%) of our patients with CLL.
6. **snv_ms_SBS5** https://www.nature.com/articles/s41588-022-01211-y Considering signatures with known or probable etiology, the most prevalent were SBS5 (clock-like), DBS11 (APOBEC activity) and ID2 followed by other clock-like signatures: SBS1 (deamination of 5-methylcytosines), SBS8, DBS2 and the AID signature SBS9.
7. **snv_ms_SBS40** https://www.sciencedirect.com/science/article/pii/S2666979X2300071X In CLL, SBS1, SBS5, SBS9, and SBS40 are established as the predominant mutational signatures.16,45,46,47SBS1, SBS5, and SBS40, are clock-like signatures—highly ubiquitous signatures of unknown etiology that increase in abundance with age.
8. **snv_ms_SBS9** https://www.nature.com/articles/s41588-022-01211-y Considering signatures with known or probable etiology, the most prevalent were SBS5 (clock-like), DBS11 (APOBEC activity) and ID2 followed by other clock-like signatures: SBS1 (deamination of 5-methylcytosines), SBS8, DBS2 and the AID signature SBS9.
9. **snv_ms_driver_GNAI1_Intron** https://pmc.ncbi.nlm.nih.gov/articles/PMC12170017/ Notably, GNAI1 emerged as a potential biomarker, demonstrating marginal significance with a P value of 0.056. Enrichment analyses elucidated that GNAI1 predominantly participates in key signaling pathways, notably oxidative phosphorylation and ubiquitin-mediated proteolysis.
10. **snv_ms_driver_TRIO_Intron** -

Pancreatic endocrine (Panc-Endocrine)

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
cna_loss_Mb	0.7849873275748980	0.7849873275748980	Yes
cna_frac_altered	0.7284166163273550	0.7284166163273550	Yes
snv_ms_driver_DAXX_Coding	0.6034940452069440	0.6034940452069440	Yes
cna_gain_Mb	0.5487045189936780	0.5487045189936780	Yes
snv_ms_driver_DYNC1I1_Intron	0.5167092205988270	0.5167092205988270	Proxy
snv_ms_driver_SETD2_Intron	0.4331959991915590	0.4331959991915590	Yes
snv_ms_TNP	-0.4292164918373850	0.4292164918373850	No
snv_ms_driver_MEN1_Coding	0.4076034914044600	0.4076034914044600	Yes
cna_total_len_Mb	-0.3440594072532090	0.3440594072532090	Proxy
snv_ms_driver_TP53_Coding	-0.3412145926208350	0.3412145926208350	No

1. **cna_loss_Mb** https://pmc.ncbi.nlm.nih.gov/articles/PMC4265611/ This study suggests that several frequent CNAs in numerous candidate regions are involved in the pathogenesis and metastatic progression of PanNET.
2. **cna_frac_altered** https://pmc.ncbi.nlm.nih.gov/articles/PMC4265611/ This study suggests that several frequent CNAs in numerous candidate regions are involved in the pathogenesis and metastatic progression of PanNET.
3. **snv_ms_driver_DAXX_Coding** https://pmc.ncbi.nlm.nih.gov/articles/PMC6185985/ The commonly mutated genes in pancreatic neuroendocrine tumors (PanNETs) are ATRX, DAXX, and MEN1
4. **cna_gain_Mb** https://pmc.ncbi.nlm.nih.gov/articles/PMC4265611/ This study suggests that several frequent CNAs in numerous candidate regions are involved in the pathogenesis and metastatic progression of PanNET.
5. **snv_ms_driver_DYNC1I1_Intron** https://www.nature.com/articles/s41467-022-31862-8 Also several genes involved in actin and MT cytoskeletal rearrangement were differentially expressed. Tubulins (Tuba1a, Tuba4a, Tubb3), MT organizers (Mtcl1, Mapre3, Kif26a), MT-associated trafficking proteins (Map1a, Map1b, Dynll2, Dync1i1, Kif5b), actin remodeling proteins (Tmsb4x, Dbn1, Scin, Vil1, Mical2) and the cytoskeletal

linker protein dystonin (Dst) were increased in Syt13high compared to Syt13low precursors

6. **snv_ms_driver_SETD2_Intron** https://www.sciencedirect.com/science/article/pii/S2211124724000317 Pancreatic deletion of Setd2 leads to abnormalities in exocrine and endocrine lineages
7. **snv_ms_TNP** (trinucleotide polymorphisms) only SNP (single) evidence.
8. **snv_ms_driver_MEN1_Coding** https://www.ncbi.nlm.nih.gov/books/NBK536980/ Multiple endocrine neoplasia type 1 (MEN1) is a rare endocrine tumor syndrome with high penetrance. It primarily causes neoplasia of the parathyroid glands, the anterior pituitary gland, and the neuroendocrine tissue of gastro-entero-pancreatic organ systems.
9. **cna_total_len_Mb** https://pubmed.ncbi.nlm.nih.gov/39062773/ Our results indicate that copy-number changes at chromosomal loci 4p16.3, 7q31.2, 9p21.3, 17q12, 18q21.2, and 19q12 may be used as diagnostic and prognostic NET biomarkers. This involves a rapid, cost-effective approach to determine the primary tumor site for patients with metastatic liver NETs and to guide risk-stratified therapeutic decisions.
10. **snv_ms_driver_TP53_Coding** https://pmc.ncbi.nlm.nih.gov/articles/PMC7106707/ Although somatic TP53 mutations have been identified in various grades of PNETs as well as gastric, small bowel, colorectal, and appendiceal neuroendocrine tumors (NETs), there are no reported cases of NETs associated with a germline TP53 mutation [2, 5, 6].

Prostatic adenocarcinoma (Prost-AdenoCA)

feature	weight	abs_weight	Did we find medical evidence that this feature is a factor of cancer?
cna_total_len_Mb	0.6813469847441590	0.6813469847441590	Yes
sv_t2tINV	0.6585677624394790	0.6585677624394790	Proxy
snv_ms_driver_MAD1L1_Intron	0.6350089878805020	0.6350089878805020	Yes
snv_ms_driver_SPOP_Coding	0.5777254987839500	0.5777254987839500	Yes
snv_chr_Y	0.559498534435773	0.559498534435773	Yes
snv_ms_driver_SCAI_Intron	0.5134274838960030	0.5134274838960030	Proxy
snv_ms_driver_ZFHX3_Intron	0.48614262344089000	0.4861426234408900 0	Yes
snv_ms_SBS1	0.44934916931648600	0.4493491693164860 0	Yes
sv_TRA	0.42101370251867500	0.4210137025186750 0	Proxy
cna_loss_Mb	-0.40902466362136000	0.4090246636213600 0	Yes

1. **cna_total_len_Mb** https://www.nature.com/articles/s41598-023-49811-w. Copy number alterations (CNAs) are frequently observed in early-stage prostate cancer and are associated with disease recurrence and tumor aggressiveness
2. **sv_t2tINV** https://www.nature.com/articles/s43018-023-00711-y First, as chromothripsis is a localized phenomenon, it leads to high density breakpoint clustering in a short genomic interval1… Finally, as rejoining of shattered chromosomes is a random process, breakpoints from the four classes of intrachromosomal rearrangements (tandem duplications, deletions, and head-to-head and tail-to-tail inversions) should be represented in approximately equal proportions… in prostate cancer, classical single-chromosome chromothripsis is frequent
3. **snv_ms_driver_MAD1L1_Intron** https://pubmed.ncbi.nlm.nih.gov/11423979/ Frequency of mutations was relatively high in prostate cancer (2/7 cell lines and 2/33 tumor specimens). We placed a mutant truncated MAD1L1, found in a lymphoma sample, into HOS, Ht161 and SJSA cell lines and found that it was less inhibitory than wild type MAD1L1 at decreasing cell proliferation.
4. **snv_ms_driver_SPOP_Coding** https://pubmed.ncbi.nlm.nih.gov/34783071/ Inactivating missense mutations in the SPOP gene, encoding speckle-type poxvirus and zinc-finger protein, are one of the most common genetic alterations in prostate cancer.

5. **snv_chr_Y** https://pmc.ncbi.nlm.nih.gov/articles/PMC6152832/ The Y chromosome is prone to high mutation rates, created exclusively in sperm cells due to the highly oxidative environment of the testis… Y chromosome related genetic abnormalities, likely to be involved in the development and progression of prostate cancer.
6. **snv_ms_driver_SCAI_Intron** https://www.mdpi.com/1422-0067/21/3/1078 On the contrary, genes whose loss/down-regulation was implicated in prostate cancer cell invasion, metastasis formation and worse prognosis (AR, IGF1, IGF2, TGFB3, SEMA3E) were down-regulated in NE-like versus AdenoPCa tumors (SCAI is suppressor of cancer cell invasion)
7. **snv_ms_driver_ZFHX3_Intron** https://www.nature.com/articles/s41420-024-02060-w Zinc-finger homeobox 3 (ZFHX3, also known as ATBF1) suppresses prostatic tumorigenesis. ZFHX3 is frequently found to have numerous deletions in human prostate cancer (PCa).
8. **snv_ms_SBS1** https://www.nature.com/articles/s41698-023-00435-3 The distribution of mutational signatures (predominance of signatures SBS1 and 5 (“clock-like” signatures, associated with aging), as well as SBS40 (unknown etiology)) (Fig. 2c) in primary tumors and metastases was consistent with what has been reported in prostate cancer (TCGA).
9. **sv_TRA** https://pubmed.ncbi.nlm.nih.gov/38947031/ Established that structural variations (SVs) are major contributors to human disease and prostate tumourigenesis, their role is under-appreciated in familial and therapeutic testing.
10. **cna_loss_Mb** https://pmc.ncbi.nlm.nih.gov/articles/PMC8556363/#:~:text=S4A).,are%20shown%20in%20Table%20S5). For SVs at the PTEN locus, we evaluated inter-chromosomal translocations and deletions separately from inversions within the chr10:89 Mbp bin (hg19), which regulate PTEN RNA abundance in localized prostate cancer