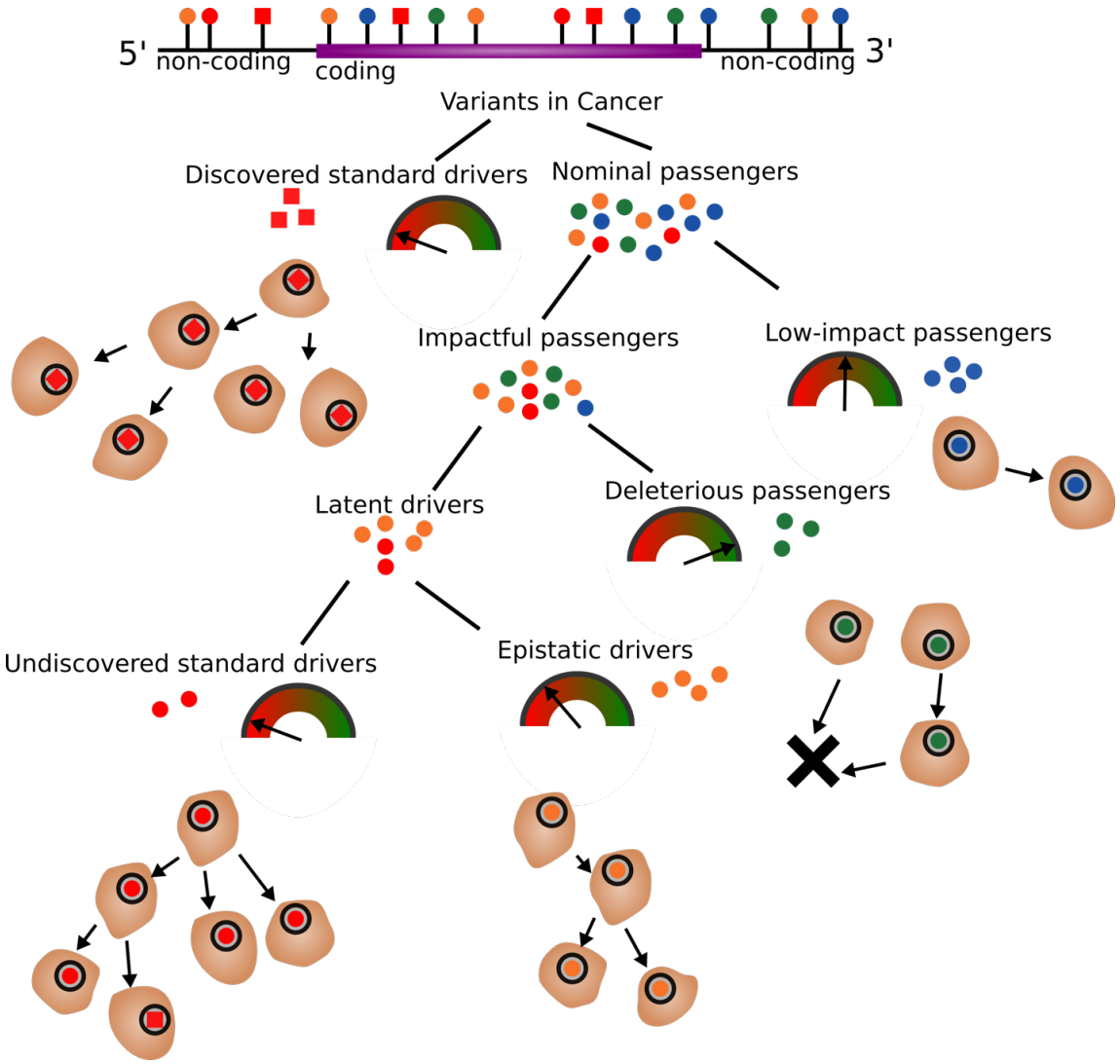


Key points:

Thousands of genomic variants mostly in noncoding region. very few of these (<5/tumor[1]) are thought to drive tumor growth.

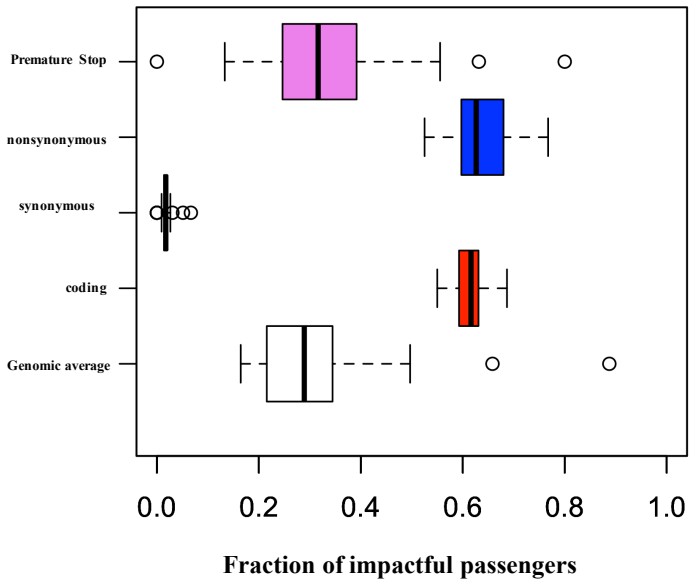
Functional consequence of passenger variants poorly understood.

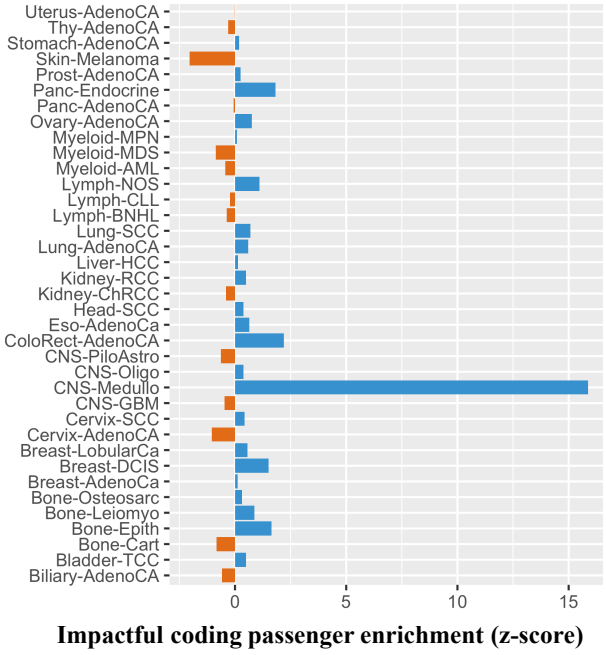
Passengers can be subdivided into neutral and impactful based on functional impact score. Low-impact passengers are thought to be inconsequential for tumor progression.

impactful passengers can alter cancer progression in two distinct mode

deleterious passengers decrease the cancer cell fitness and inhibit cancer growth

impactful passengers can drive cancer progression by interacting epistatically





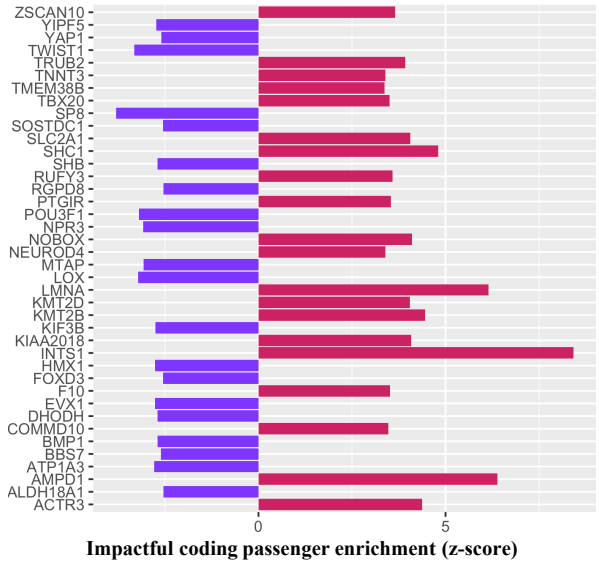
Key points:

Higher fraction of impactful passenger SNVs in the coding region compared to genomic average.

Among coding variants nonsynonymous and synonymous SNVs have the highest and lowest fraction of impactful passengers, respectively.

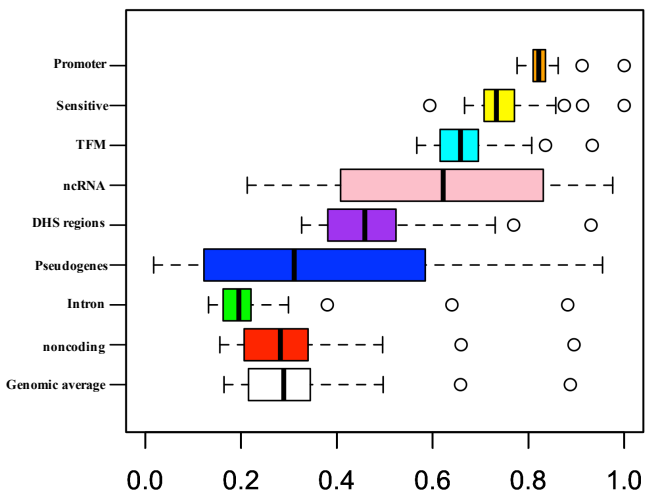
Medulloblastoma cohort is highly enriched in impactful coding passengers SNVs.

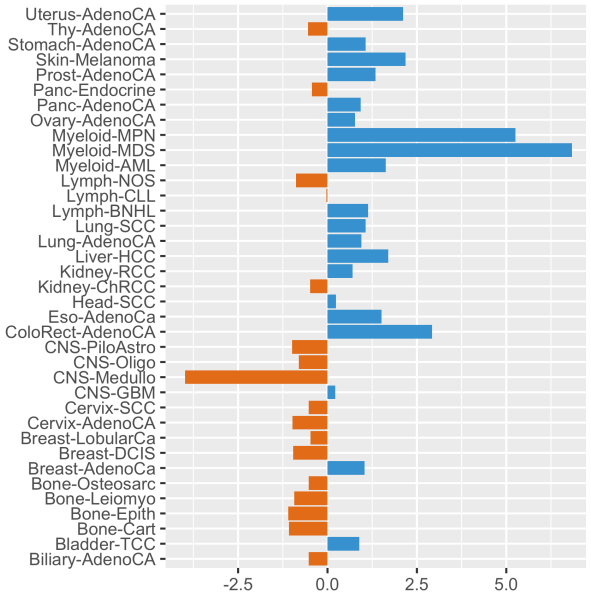
In contrast, Melanoma is depleted in impactful coding passenger variants.



Enrichment of impactful non-synonymous passenger variants in key essential genes such as INTS1, LMNA, AMPD1 and SHC1 genes. Biologically these genes are involved in transcription initiation, signal transduction pathways, cell scaffolding.

Depletion of impactful passengers in genes such as SP8, TWIST1 and POU3F1.





Impactful non-coding passenger enrichment (z-score)

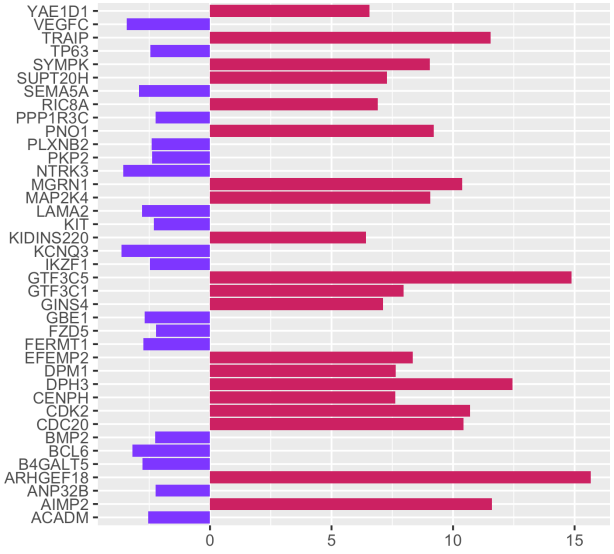
Key points:

Among non-coding variants, promoters tend to harbor larger fraction of impactful passengers compared to other non-coding region.

Whereas, majority of passenger variants falling in intronic regions are neutral.

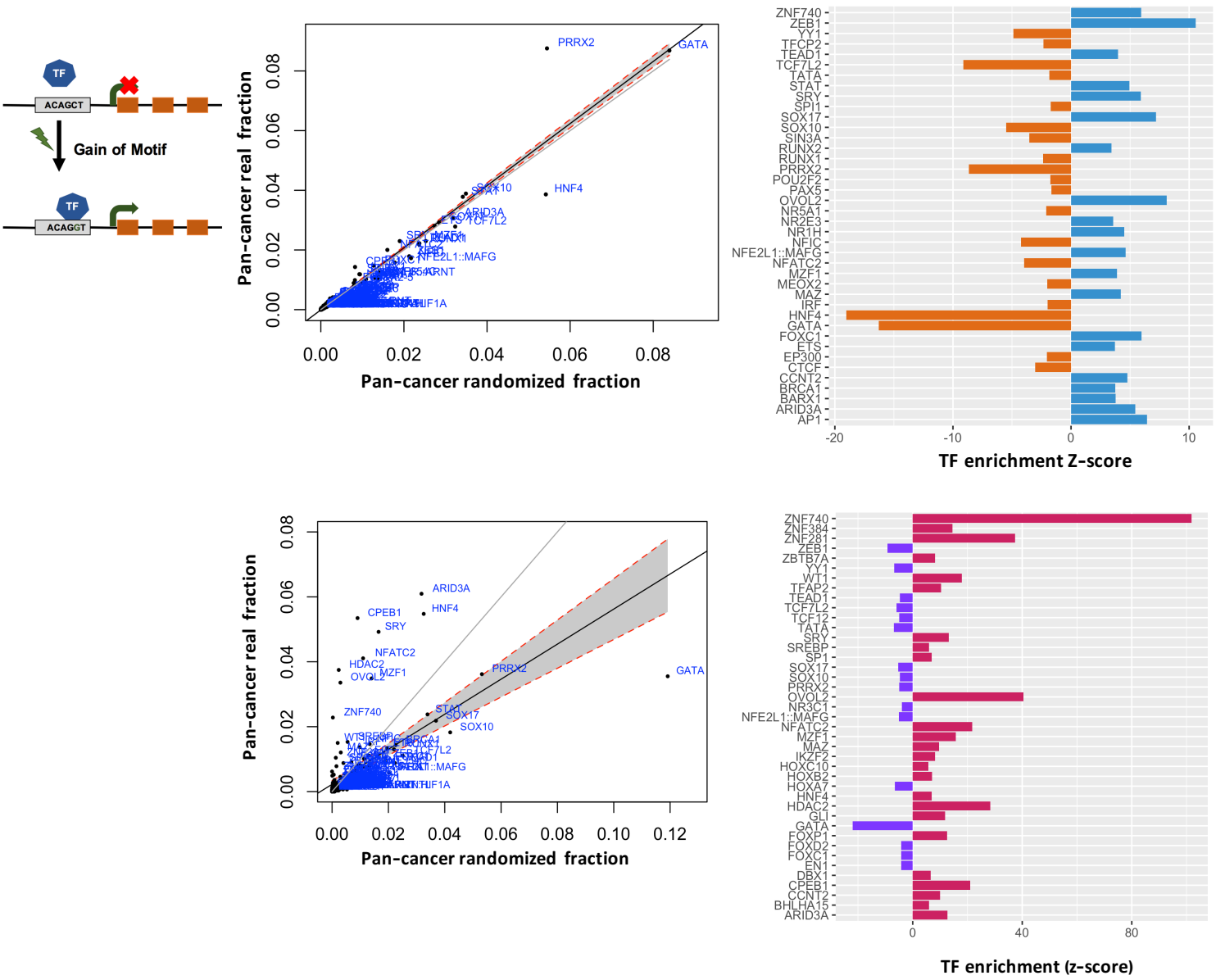
Myeloid-MDS and Myeloid-MPN cohorts are highly enriched with impactful non-coding passenger variants.

Interestingly, impactful passengers are highly depleted in the Medulloblastoma cohort in noncoding regions.



Impactful promoter passenger SNV enrichment (z-score)

Gene level analysis indicate enrichment of impactful passenger variants in promoter region of genes such as ARHGEF18, GTF3C5, DPH3 and TRAIP.



Key points:

SNV induced gain of motif events

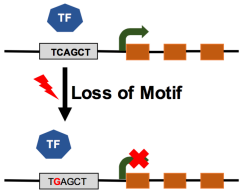
highly enriched among TFs such as ZEB1, ZNF740, OVOL2 and SOX17.

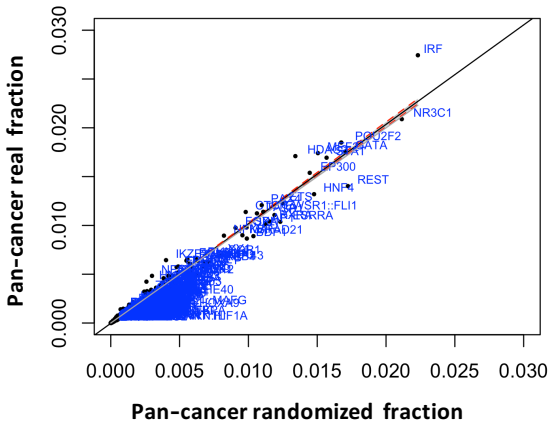
depleted among TFs such as GATA and HNF4.

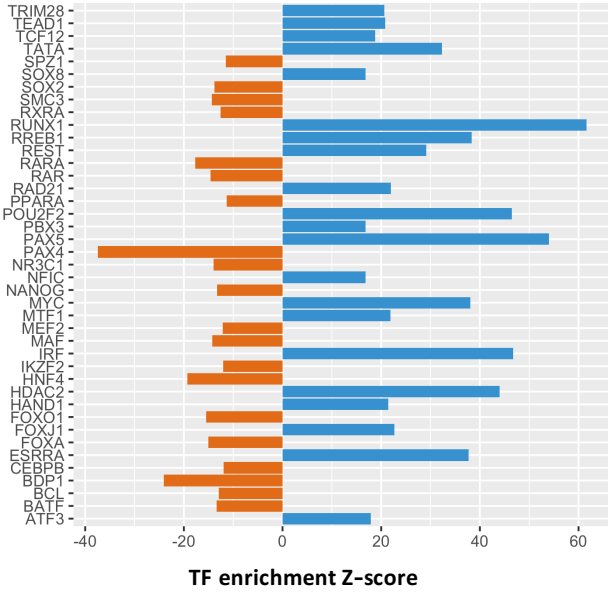
Indel induced gain of motif events

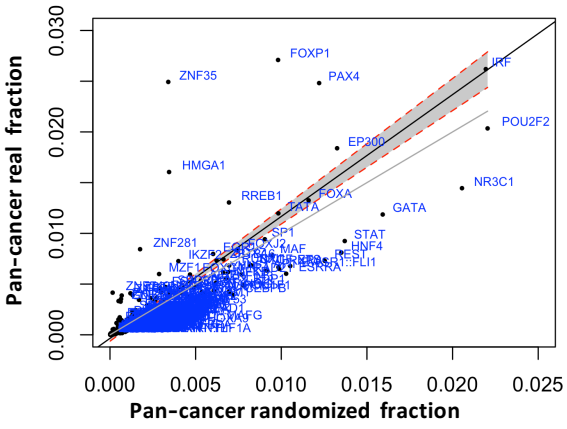
enrichment among of TFs belonging to zinc-finger protein family (ZNF740, ZNF384, ZNF81).

Interestingly, we observe ZNF740, OVOL2 to be highly enriched and GATA to be highly depleted due to both SNV & INDEL induced gain of motif events.









Key points:

SNVs leading to motif breaking events

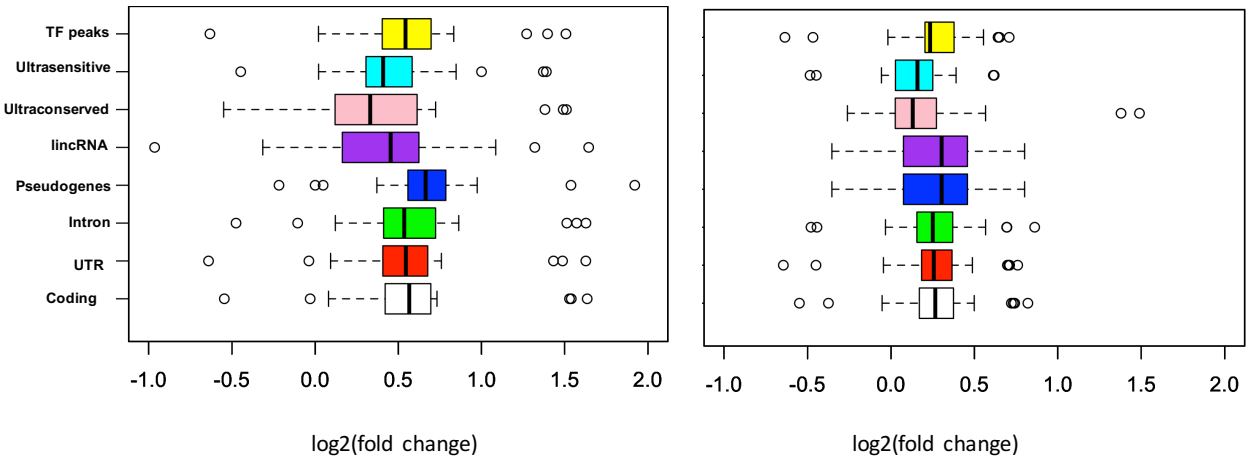
Highly enriched among TFs such as RUNX1, POU2F2, PAX5 and IRF.

Highly depleted among PAX4 and BDP1 TFs.

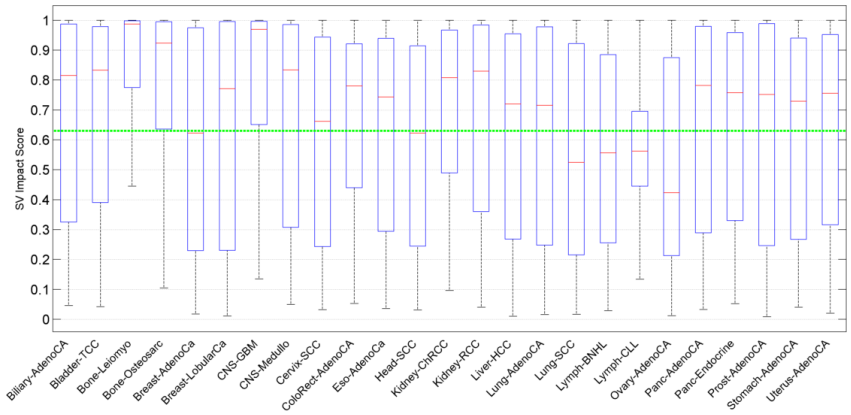
Gain of motif events in INDELs

Enrichment of variants in TFs such as HDAC2, ZNF740, ZNF35 and HMGA1

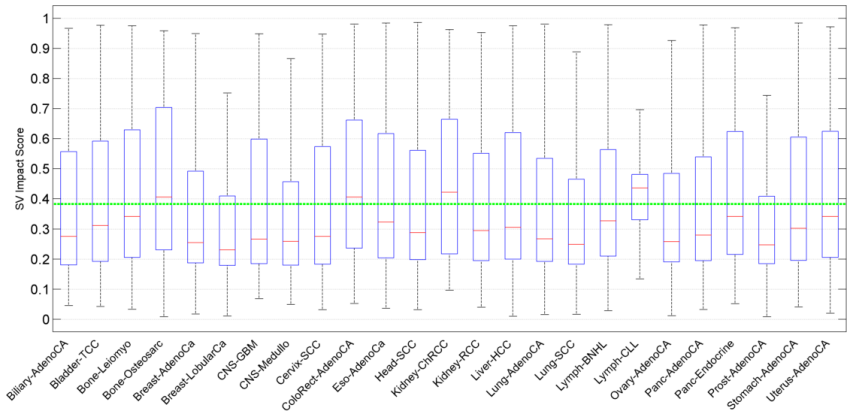
Depleted in REST1 AND GATA



All SVs



SVs<100kb



Key points:

Enrichment of large engulfing somatic deletions is highest among Pseudogenes, coding region, UTRs and intronic region.

Partially overlapping deletions are not highly enriched as engulfing ones.

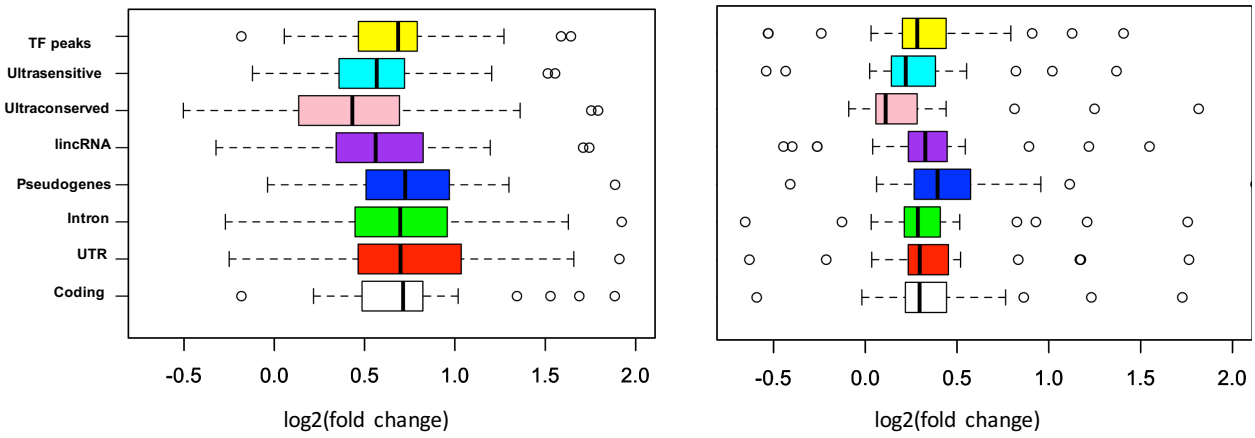
Partially overlapping deletions enriched among lincRNA and pseudogenes to similar extent compare to engulfing scenario.

Cartilaginous neoplasm cohort has highest enrichment values for engulfing and partially overlapping deletions compared to other cancer types in the PCAWG. This is consistent across different genomic elements.

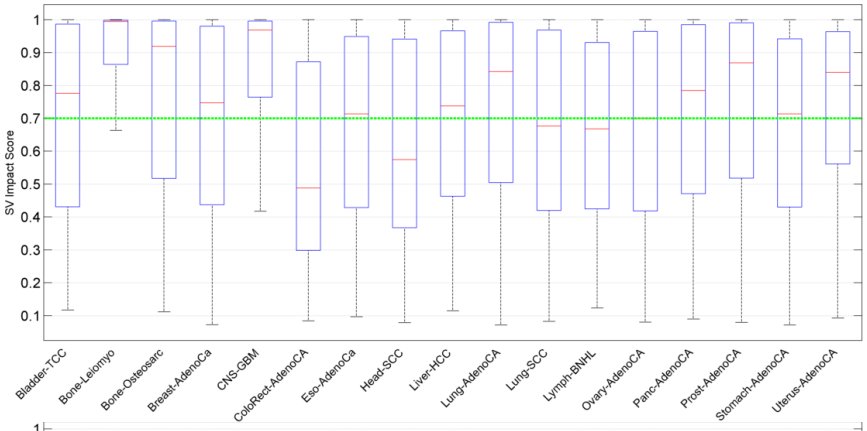
Presence of high impact SVs in Cartilaginous neoplasm, Bone sarcoma and Diffusive glioma (CNS-GBM) cohorts.

Ovary-AdenoCA, Lymph-CLL and Lung-Squmous cohorts have on average lower impact SVs. However, the impact distribution is highly dispersed in these cohorts.

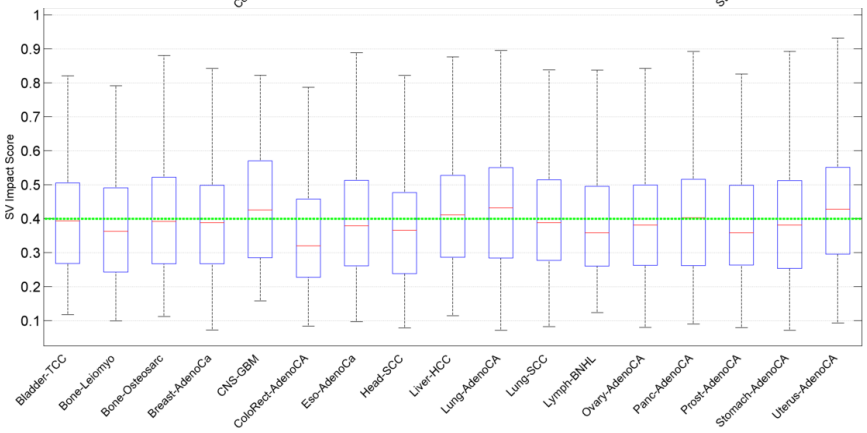
Impact score distribution for deletion below 100Kb indicate distinct profile. Notably, CLL cohort tends to have higher impact deletions on average in this scenario.



All SVs



SVs<100kb



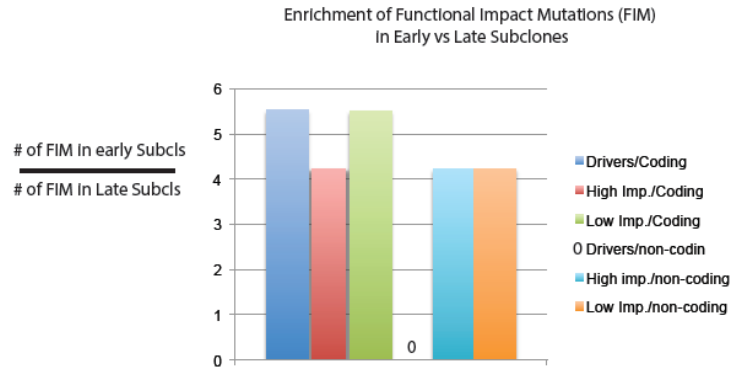
Key points:

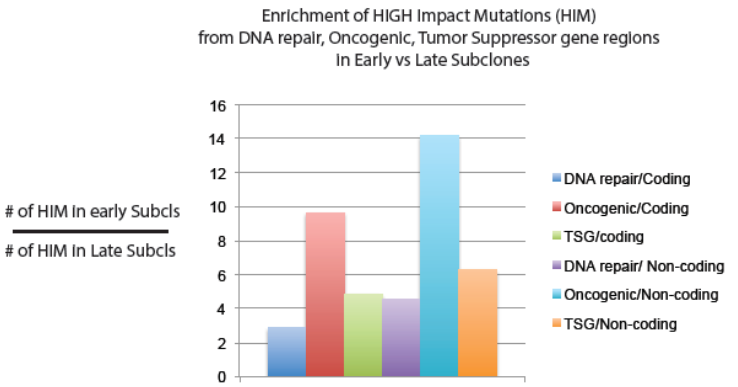
Enrichment pattern remains similar for large duplications as in large deletions.

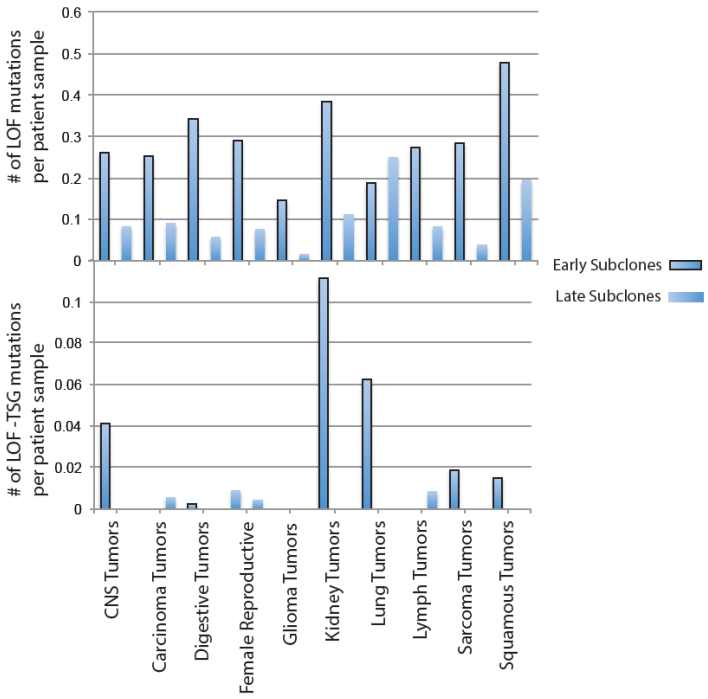
Coding, intronic, UTR and pseudogenes overlap to a larger extent for both engulfing and partial overlap scenarios.

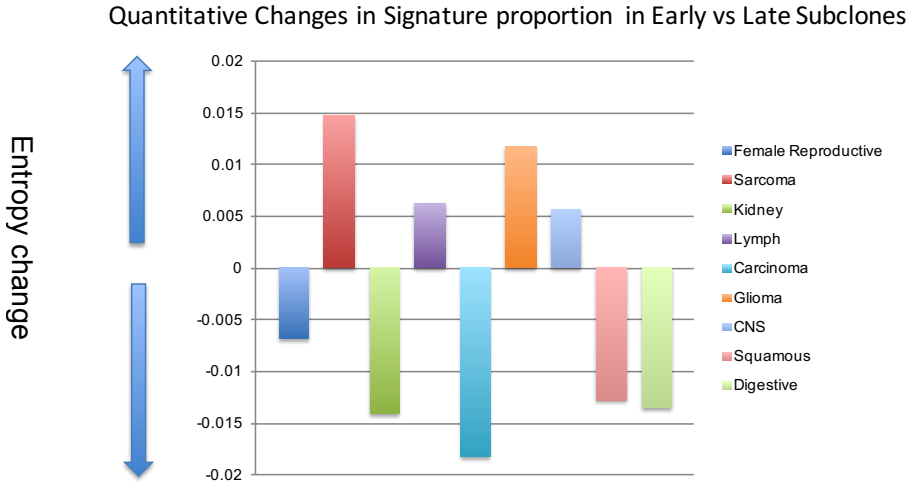
Cartilaginous neoplasm, Insitu Breast Adenocarcinoma, Pancreatic Adenocarcinoma and PiloAstrocytmo cohorts have higher enrichment values for duplication overlap.

Presence of high impact duplications in Cartilaginous neoplasm, Bone sarcoma, Diffusive glioma (CNS-GBM) cohorts. Prostate and Lung Adenocarcinomas also tend to have higher impact duplications.









Key points:

Enrichment of functional Impact mutations in Early vs Late Subclones

- Mutations of positive fitness should appear with higher frequency (cell prevalence) or enriched in early/dominant subclones. Mutations of negative fitness should wash off the population

- Functional Impact mutations are enriched in Early vs Late Subclones and High Freq vs Low Frequency bins (Frequency graph not shown here)

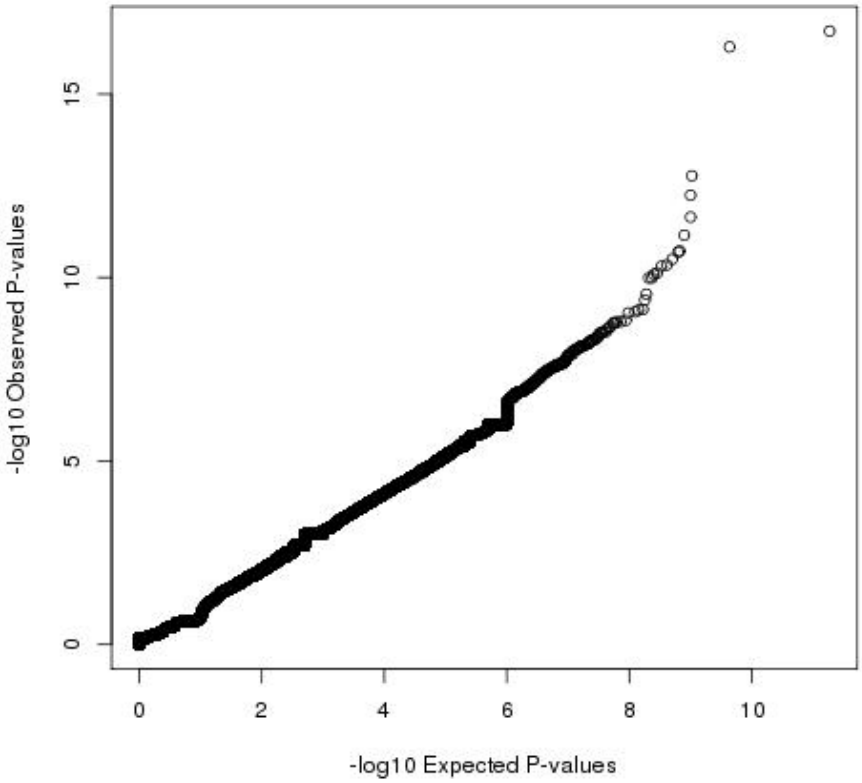
- High Impact mutations in oncogenic regions are particularly enriched in early subclones showing higher fitness

Average number of LOF mutations in early vs late Subclones

- We expect: a) either higher fitness for each cell and therefore in higher prevalence or b) lower fitness and lower prevalence
- Tumor Suppressor gene LOF should -by definition- provide a higher fitness for tumor cells and therefore exist in higher than expected prevalence

Quantitative Changes in Signature proportion in Early vs Late Subclones

- Values of change in entropy closer to 1 signify higher Entropy in early signatures (Fewer of higher prevalence). Negative values signify higher entropy for late signatures. Closer to 0 Signify no change in signature proportions.
    - Female Reproductive, Lymph and CNS tumors show the most conservative nature, remaining mostly unchanged.
    - Sarcoma and Glioma show fewer and more dominant signatures in early/dominant subclones
    - Kidney, Lymph, Squamous and Digestive tumors show fewer signatures in later subclones



	Observed significantly co-mutated gene-gene pairs	Randomized significantly co-mutated gene-gene pairs
Biliary	1	0
Bladder	1	0
GBM	9	1
Medullo	2	0
Head-SCC	28	1
Lung-Adeno	11	1
Lung-SCC	105	17
Lymph-BNHL	3	3
Lymph-CLL	1	0
Uterus-AdenoCA	9	0

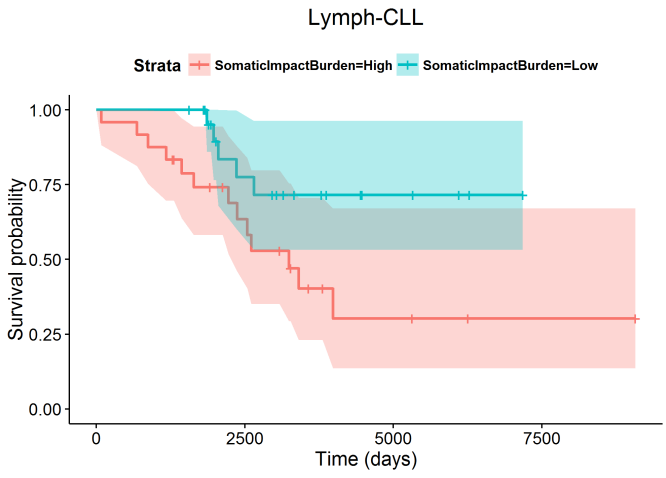
Key points:

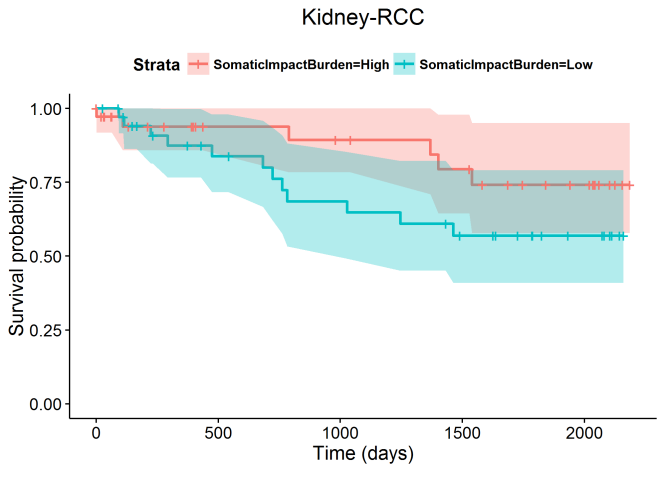
Despite being neutral on their own, two or more variants might confer a selective advantage to tumor cells when mutated.

Investigated epistatic events among the PCAWG variants in the form of gene-pairs that are co-mutated more frequently than expected under additive-effects assumptions.

Nine subtypes are represented among the nine gene-gene-subtype triples, especially squamous cell cancers – both of the lung and of the head.

Comparison with randomized sets suggests a 13% false discovery rate.





Key points:

survival analysis to see if impact burden predicted patient survival within individual cancer subtypes.

somatic mutation burden predicted substantially earlier death in chronic lymphocytic leukemia (CLL) and substantially prolonged survival in renal cell carcinoma (RCC), respectively.

These observations remained after adjusting for patient age at diagnosis and low-impact mutation load and after defining tumor impact burden in relation to the burdening of corresponding randomized sets.

These results lend support to the hypothesis that the aggregate amount of impactful passengers is clinically meaningful.

More specifically, the results suggest that latent drivers are more important than deleterious passengers in CLL, but that the situation is reversed in RCC.

This can be explained by the large share of missing drivers in CLL, which suggests a greater role for latent drivers in CLL.