Sensitivity Analysis I – additive variance analysis of double randomization set

- Created two randomized sets for six cancer cohorts
- Calculated additive variance for SNV-level and Gene-level models

Breast	CNS	Kidney	Ovary	Pancreas	Prostate
1e-6%	2e-6%	2e-6%	2e-6%	1e-6%	1e-6%

Sensitivity Analysis II – evaluating the influence of window size

	50kb model	100kb model
Breast	0.5105	0.5147
CNS	0.1991	0.2014
Kidney	0.5072	0.6409
Ovary	0.6485	0.6426
Prostate	0.3296	0.3326
Pan-cancer	**0.4390**	**0.4664**

Custom MOAT-sim to generate co-variate corrected randomized dataset

- Following conditions were applied to generate new randomized dataset

1. Tri-nucleotide context of permuted location was same to original mutation

2. Permuted mutation lie on the same chromosome as original mutation

3. genome were divided into 10kn non-overlapping bin. Average of multiple co-variates were computed for each bin.

4. Clustering approach was applied to identify relevant bins, where a given mutation can be permuted satisfying condition1 and condition2.

5. Following co-variate corrections were considered for generating the background mutations.
   a) Replication timing
   b) Chromatin accessibility
   c) GC content
   d) Penta-nucleotide context for Liver and Melanoma cohorts

MOAT-sim based additive variance analysis

(SNV-level model)

	Drivers	Coding	Promoters	Non-coding	Total
Breast	0.5132	0.0023	0.0107	0.0528	0.579
CNS	0.1738	0.0065	0.0108	0	0.1911
Kidney	0.426	0.024	0.0046	0.0323	0.4869
Ovary	0.5622	0.0076	0.0513	0.1659	0.787
Pancreas	0.9312	0.0512	0	0	0.9824
Prostate	0.248	0	0	0.1572	0.4051
Pan-cancer	**0.4757**	**0.0153**	**0.0129**	**0.068**	**0.5719**

MOAT-sim based weak drivers (Overall)

(Lower estimates)

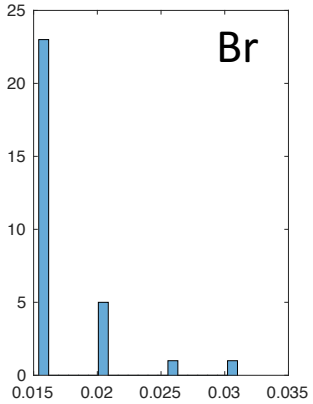
	Breast	CNS	Kidney	Ovary	Pancreas	Prostate	**Pan-cancer**
WD	8.2577	0.2695	4.7817	3.0818	8.2035	2.6616	**4.5426**
DP removed	9.4381	1.3475	4.2465	13.2182	24.7403	5.8535	**9.8074**
DP retained	3.8077	0.1511	1.4227	4.8395	9.7791	1.8056	**3.6343**

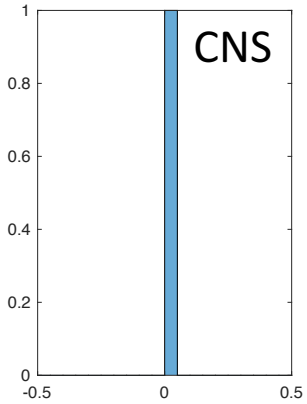
Normalized additive variance

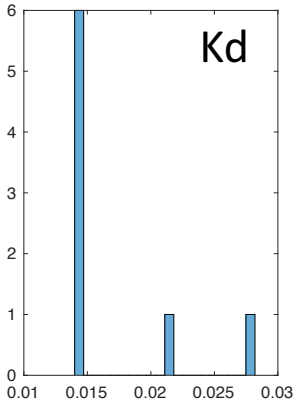
	Cod	Prom	Noncod
Br	0.1915	0.0032	0.003
CNS	1.0548	0.0476	0
Kd	0.7387	0.0042	0.0063
Ov	0	0	0.0191
Pa	0.2071	0	0
Pr	0	0	0.0113
Pan	0.3654	0.0092	0.0066

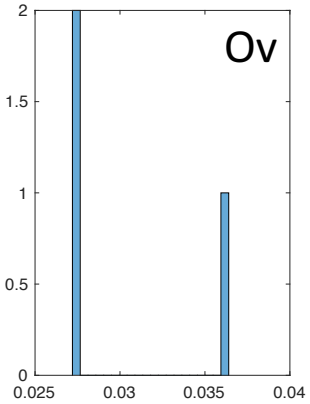
Prevalence analysis, SNV level

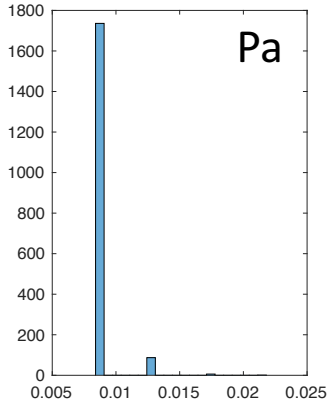
Lower

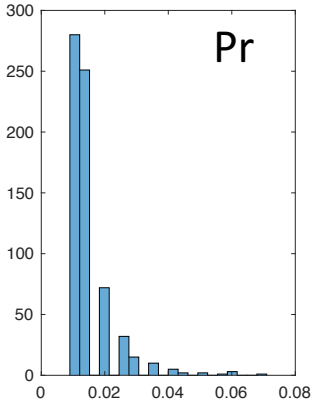












Intersected specific drivers (Gene level)

- Br: 'TP53' 'PLEKHS1' 'TBC1D12' 'BPIFB2'
- CNS: -
- Kd: 'BRINP3' 'UBE2U'
- Ov: 'TP53' 'SH3BGRL3' 'INTS4' 'POLR3E'
- Pa: 'TP53' 'SMAD4' 'KAT8' 'ZFP36L2' 'LANCL2' 'CDKN2A' 'GNAS' 'KRAS' 'ACVR1B'
- Pr: 'PTEN'

Extra Slides

MOAT-sim based additive variance analysis

Nested model:

$$y_j = \mu + z_j^{\mathrm{dr}} u^{\mathrm{dr}} + \sum_i z_{ij}^{\mathrm{cd}} u_i^{\mathrm{cd}} + \sum_i z_{ij}^{\mathrm{prm}} u_i^{\mathrm{prm}} + \sum_i z_{ij}^{\mathrm{ncd}} u_i^{\mathrm{ncd}} + e_j$$

$$u_i^{\mathrm{dr}} \sim N(0, \sigma_{\mathrm{dr}}^2) \qquad u_i^{\mathrm{prm}} \sim N(0, \sigma_{\mathrm{prm}}^2)$$

$$u_i^{\mathrm{cov}} \sim N(0, \sigma_{\mathrm{cov}}^2) \qquad u_i^{\mathrm{ncd}} \sim N(0, \sigma_{\mathrm{ncd}}^2)$$

$$u_i^{\mathrm{cd}} \sim N(0, \sigma_{\mathrm{cd}}^2) \qquad e_j \sim N(0, \sigma_{\mathrm{e}}^2)$$

MOAT-sim based weak drivers (Split)

Coding	Breast	CNS	Kidney	Ovary	Pancreas	Prostate	**Pan-cancer**
WD	0.6082	0.0709	0.4648	0.2636	1.2251	0.2778	**0.4851**
DP removed	1.067	0.1064	0.5352	1.0818	2.7229	0.601	**1.0191**
DP retained	0.2833	0.0124	0.1367	0.2591	0.7175	0.0934	**0.2504**

Promoters	Breast	CNS	Kidney	Ovary	Pancreas	Prostate	**Pan-cancer**
WD	0.8866	0.0284	0.2676	0.2091	0.7532	0.2929	**0.4063**
DP removed	0.4381	0.1844	0.493	1.0091	1.2424	0.5556	**0.6538**
DP retained	0.1772	0.019	0.0667	0.4258	0.5008	0.1847	**0.229**

Other NC	Breast	CNS	Kidney	Ovary	Pancreas	Prostate	**Pan-cancer**
WD	6.9485	0.1702	4.1197	2.6545	6.2857	2.1414	**3.72**
DP removed	8.1598	1.0709	3.2817	11.3727	21.3766	4.7525	**8.3357**
DP retained	3.3408	0.1197	1.1939	4.2005	8.4475	1.5122	**3.1358**