Supplemental Figure 1. Related to STAR Methods

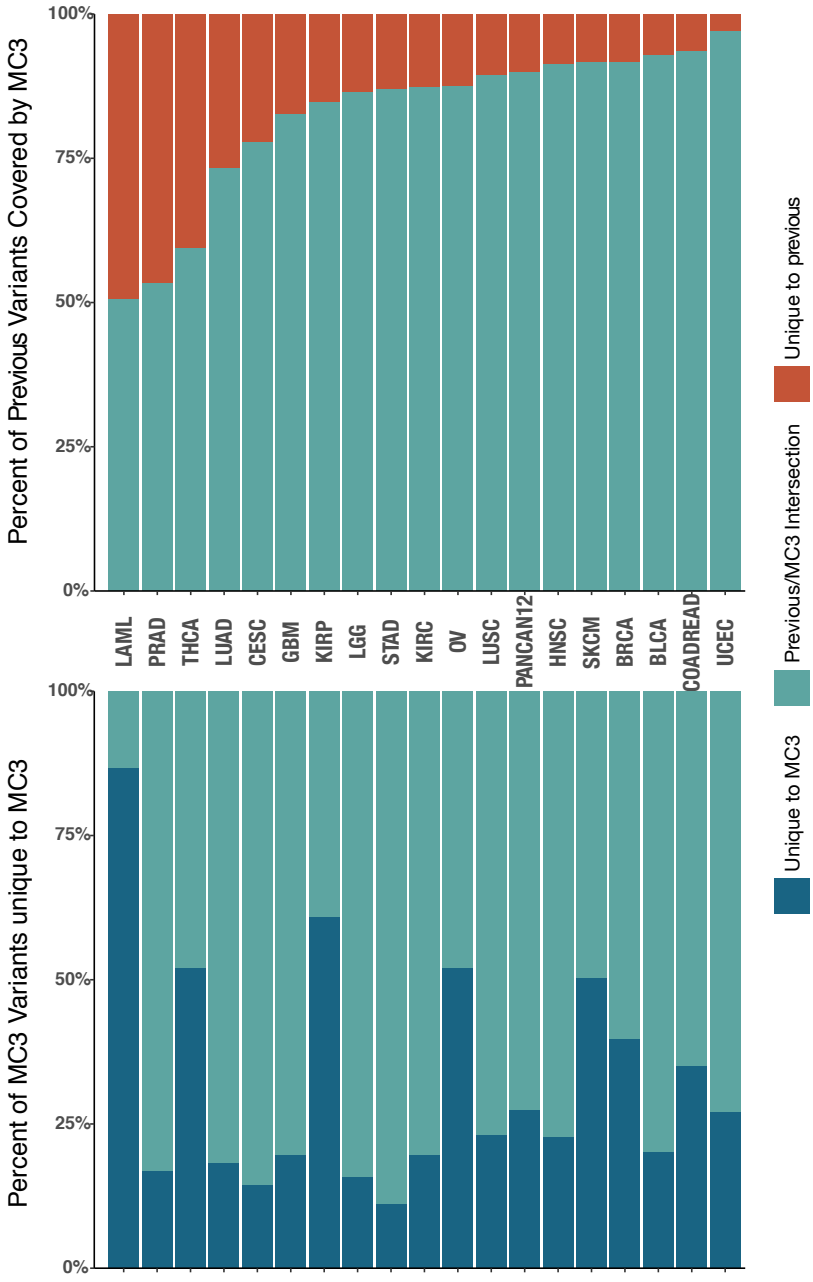


Figure S1. Related to STAR Methods: Overlap of mutation calls with analysis working group MAFs at the time of Pancan12, 2013. Related to STAR Methods. Top) Proportion bar plots indicate the percentage intersection of previous efforts with MC3 variants (teal) and the percentage of variants that are unique to previous variant calling efforts (red). Bottom) Proportion bar plot indicates the percentage of MC3 variants that overlap with variants produced by previous efforts (teal) and the percentage of variants that are unique to the MC3 MAFs (blue).

Supplemental Figure 2: Related to STAR Methods

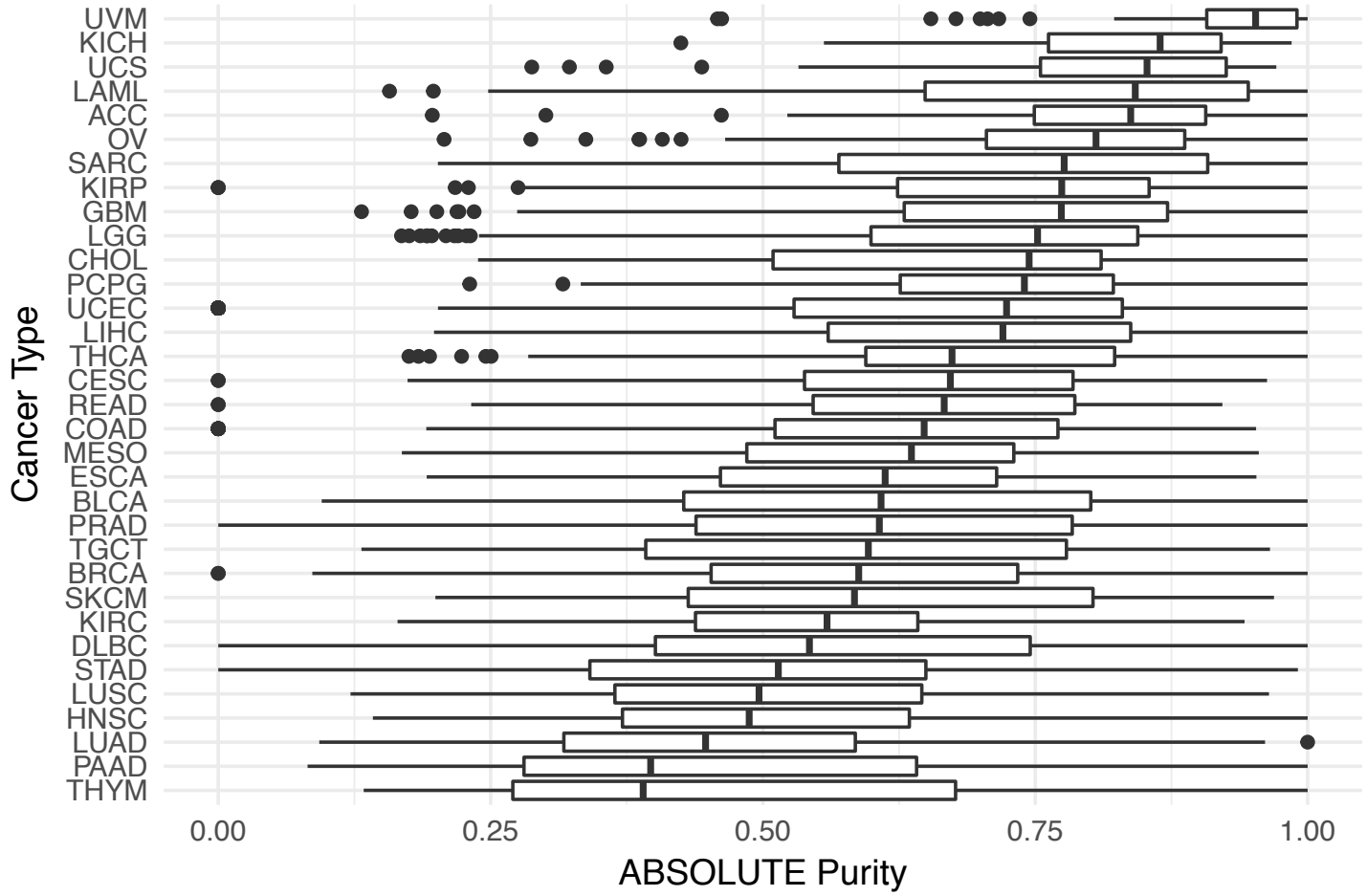


Figure S2. Related to STAR Methods: Tumor purity scores calculated by ABSOLUTE, are shown in box-plot format. Related to STAR Methods. Cancer types on the y-axis are sorted by increasing median purity estimates.

Supplemental Figure 3. Related to Figure 4.

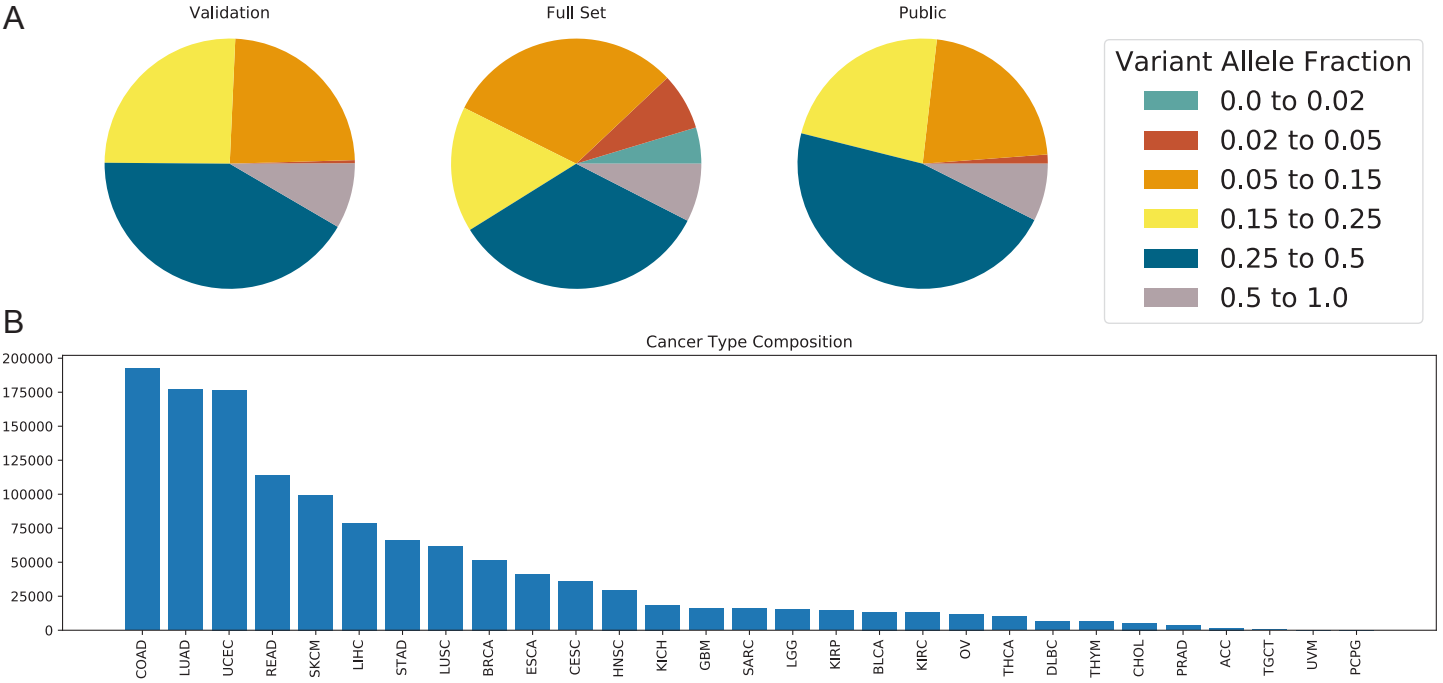
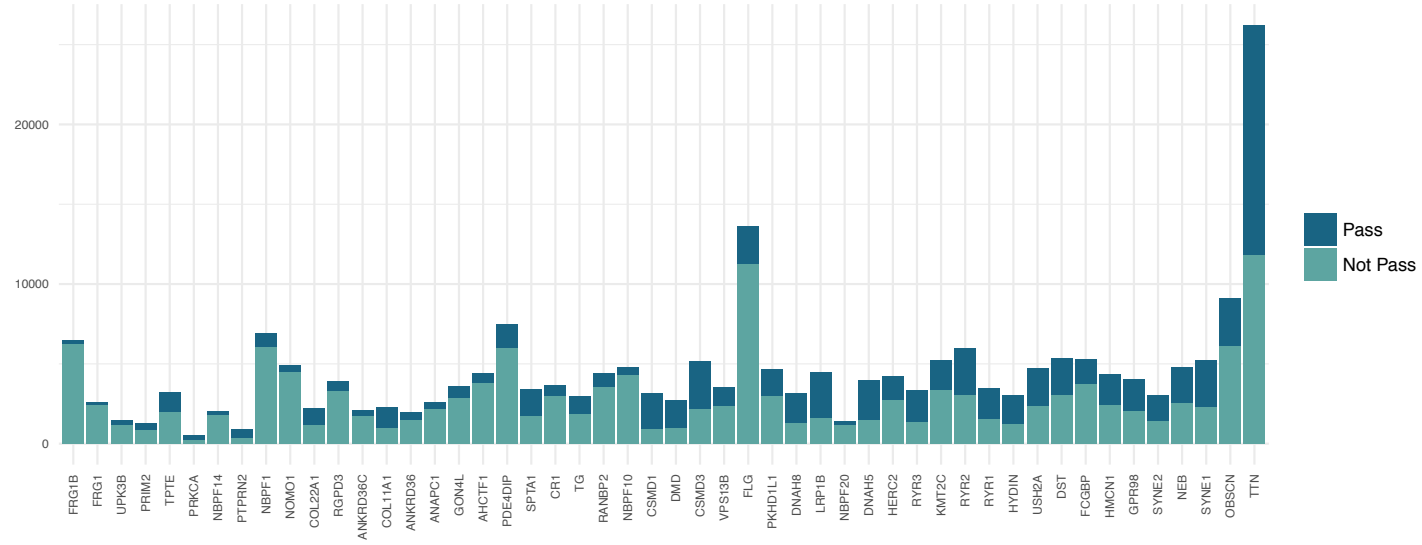


Figure S3. Related to Figure 4: Composition of validation data. Related to Figure 3. (A) Composition of the Variant Allele Fraction (VAF) of mutations in the validation set, the full mutation call set and the filtered open-access data set. Validation data has a clear bias toward lower VAF mutations, selected for validation because they were harder to call. (B) The composition of the validation data by cancer type. Most of the calls come from UCEC, COAD, and LUAD.

Supplemental Figure 4. Related to Figure 3

A



B

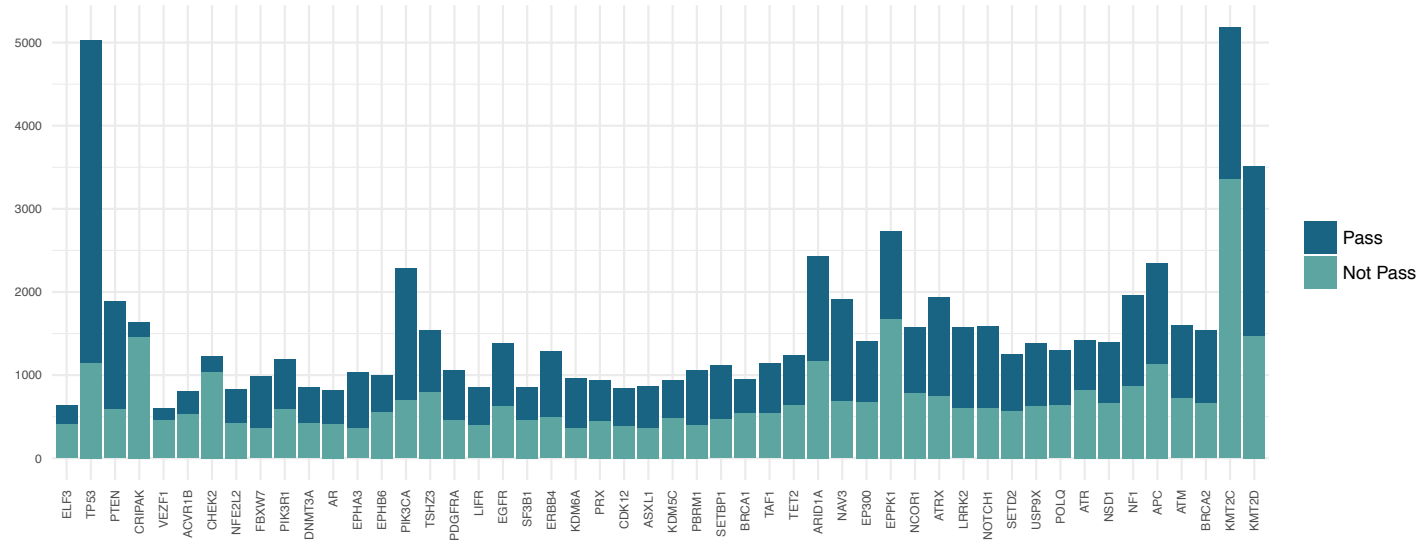


Figure S4. Related to Figure 3: The effects of filtering on mutation counts by gene. Related to Figure 2. Mutation count analysiswas performed for the pre- and post-filtering mutations using the “PASS” filter flag. Variants used for this analysis were restricted to the exonic regions only. (A) The height of each bar represents the total number of called mutations for each gene and is split by “PASS” calls and not “PASS” calls. The top genes 50 genes with the largest difference (‘Not passed’ minus ‘passed’) are plotted in order according to increasing gene length. (B) This panel is identical to panel (A) but is subset to 50 cancer genes identified by Kandoth et. al 2013.

Supplemental Table 1: Related to Figure 1

Program	Flag	Description	Value	Default
pindel				
	-x	The maximum size of structural variations to be detected	1	2
	-w	For saving RAM, divides the reference in bins of X million bases and only analyzes the reads that belong in the current bin	0.1	5
	-m	At the point where the read is split into two, there should at least be this number of perfectly matching bases between read and reference	6	3
	-J		centromere exclusion file	
Somaticsniper				
	-G	Do not report Gain of Reference variants as determined by genotypes		
	-L	Do not report LOH variants as determined by genotypes		
	-q	Filtering reads with mapping quality less than	1	0
	-Q	Filtering somatic snv output with somatic quality less than	40	15
VarScan				
	-B (samtools)	Disables BAQ computation		
	--min-coverage	Minimum coverage in normal and tumor to call variant	3	8
	--min-var-freq	Minimum variant frequency to call a heterozygote	0.08	0.1
	--p-value	P-value threshold to call a heterozygote	0.1	0.99
muse	-E	Exome mode		

Supplemental Table 2: Related to Figure 4

	Validated	Unvalidated	Germline	True positive	True negative	False positive	False negative
SOMATICSNIPER	230361	3171	81	230361	45935	3252	97493
RADIA	277255	23806	38	277255	25343	23844	50599
MUTECT	309470	7338	7	309470	41842	7345	18384
MUSE	305950	19154	58	305950	29975	19212	21904
VARSCANS	289477	22705	226	289477	26256	22931	38377
PINDEL	14071	780	2	14071	3272	782	10688
INDELOCATOR	18631	2331	4	18631	1719	2335	6128
VARSCANI	21642	1394	24	21642	2636	1418	3117
1+ SNP Callers	327854	48886	301	327854	0	49187	0
2+ SNP Callers	315362	16914	67	315362	32206	16981	12492
3+ SNP Callers	291147	4649	26	291147	44512	4675	36707
4+ SNP Callers	267484	3355	15	267484	45817	3370	60370
5 SNP Callers	210666	2370	1	210666	46816	2371	117188
1+ INDEL Callers	24759	4024	30	24759	0	4054	0
2+ INDEL Callers	17369	349	0	17369	3705	349	7390
3 INDEL Callers	12241	138	0	12241	3916	138	12518

Supplemental Table 3: Related to Figure 4

	MUTECT	ONE or MORE	MUTECT	TWO or MORE	MUTECT	TWO or MORE
Filter	PASS	PASS	PASS	PASS	PASS	PASS
Variant	SNP	SNP	SNP	SNP	SNP	SNP
Validation type	targeted+wg	targeted+wg	wgs	wgs	targeted	targeted
Number of valdated sites with power > 0.95	327854	327854	251210	251210	85769	85769
Number of Calls by CALLER	305056	332137	229650	231732	84139	84650
Validated Calls	299906	312820	227864	230005	80705	81130
Unvalidated calls	5150	19317	1786	1727	3434	3520
Germline validatione evidence	4	21	3	5	1	1
Validated but not called	27948	15034	23346	21205	5064	4639
TRP (G/(G+J))	0.9148	0.9541	0.9071	0.9156	0.9410	0.9459
FDR = (H+I)F	0.0169	0.0582	0.0078	0.0075	0.0408	0.0416
F1	0.9477	0.9480	0.9477	0.9525	0.9500	0.9521
Fp5	0.9686	0.9443	0.9739	0.9761	0.9555	0.9559
Fp1	0.9824	0.9420	0.9913	0.9917	0.9590	0.9583