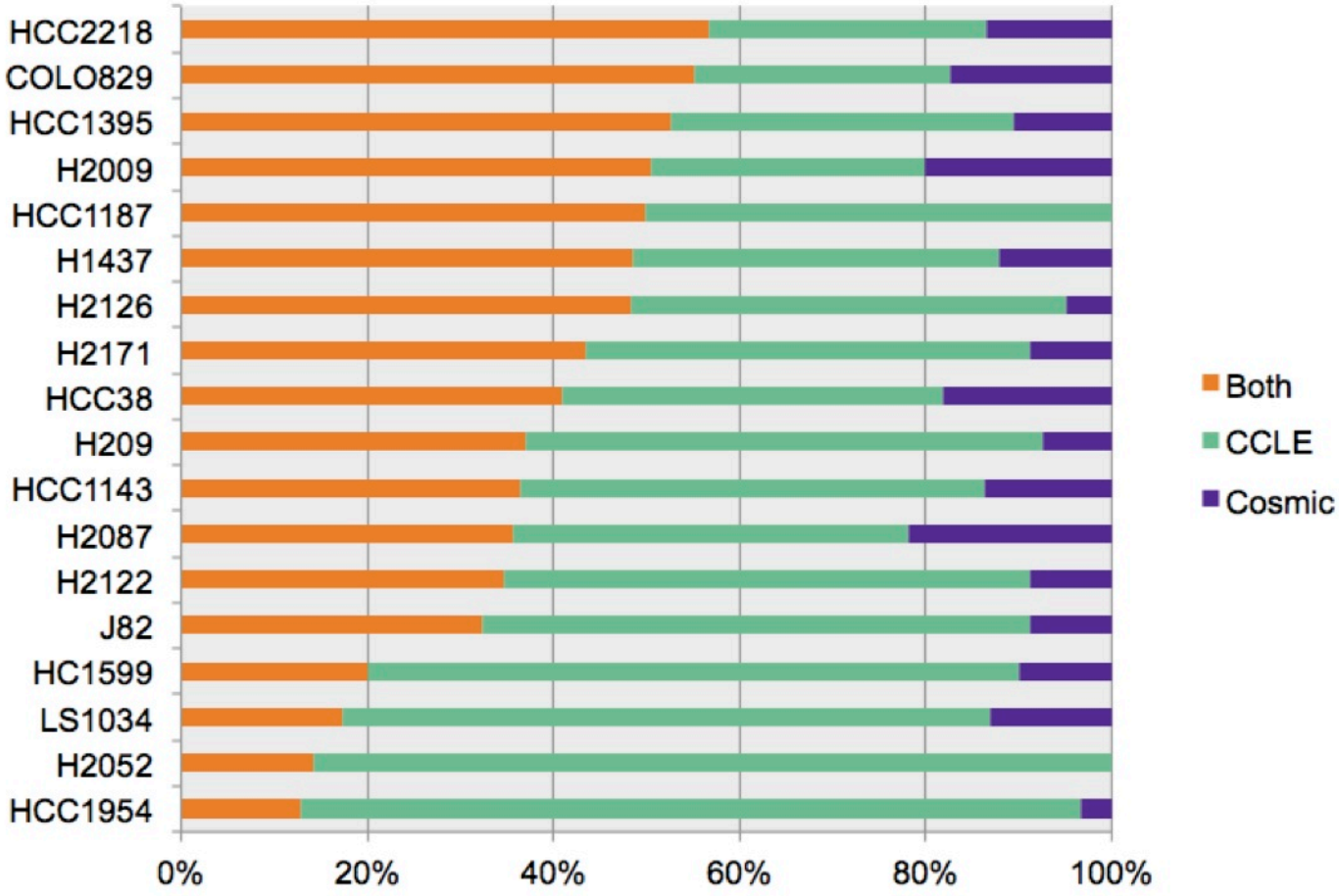
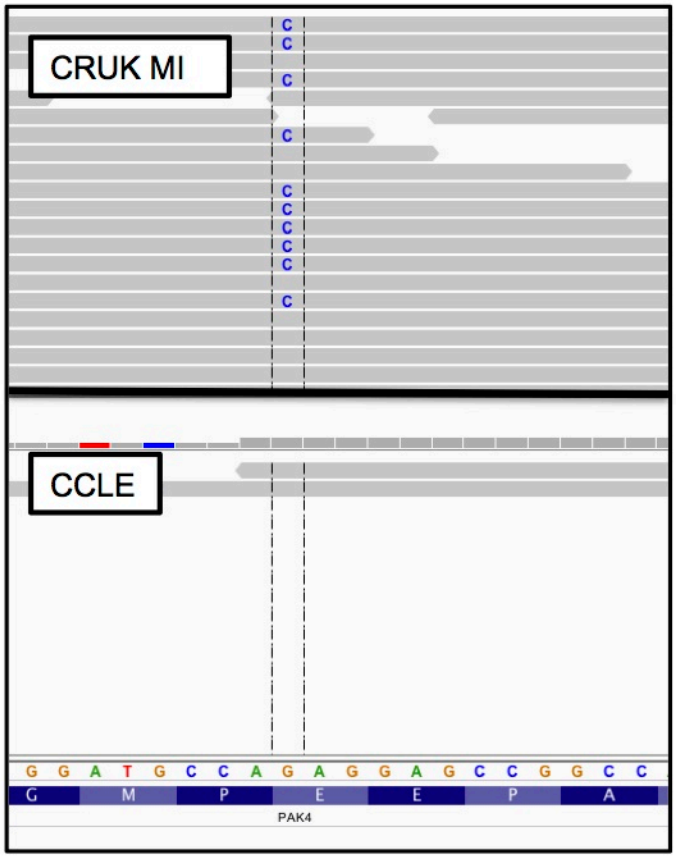
Supp. Figure 1

Original comparison of 18 cell lines sequenced by COSMIC and CCLE show that the conformity of missense mutation detection ranges from 56.75% in HCC2218 cell line to 12.90% in HCC1954.



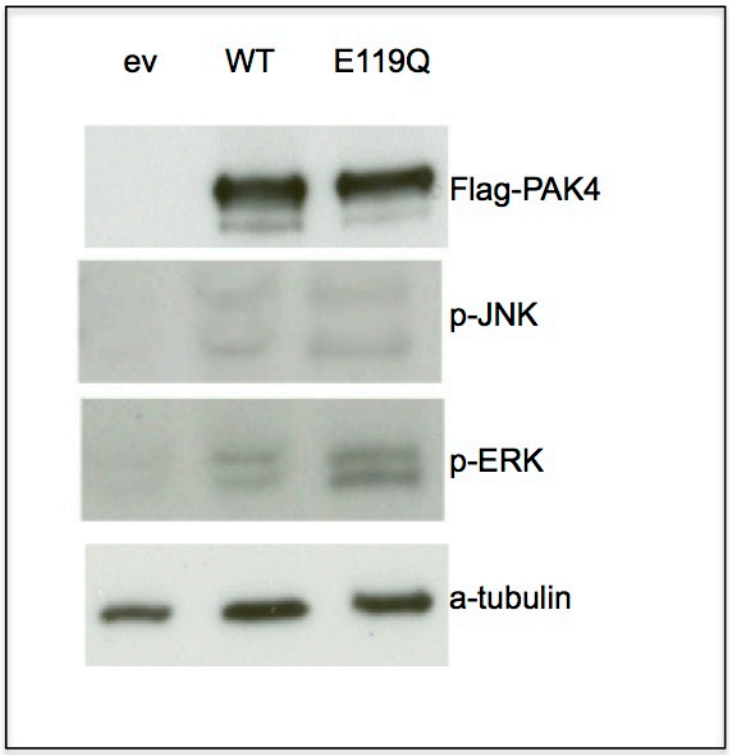
Supp. Figure 2

Images from the Intergrative Genomics Viewer illustrating the read coverage in the CRUK MI sequencing bamfiles and the CCLE hybrid capture bamfiles for the PAK4 p.E119Q mutation in H2009 cell line. The mutation was only reported in the CRUK MI sequencing with mutation seen in 51% of reads in a good coverage area but not reported by CCLE as the area is poorly covered by hybrid capture (only 2 reads and neither showing mutation).



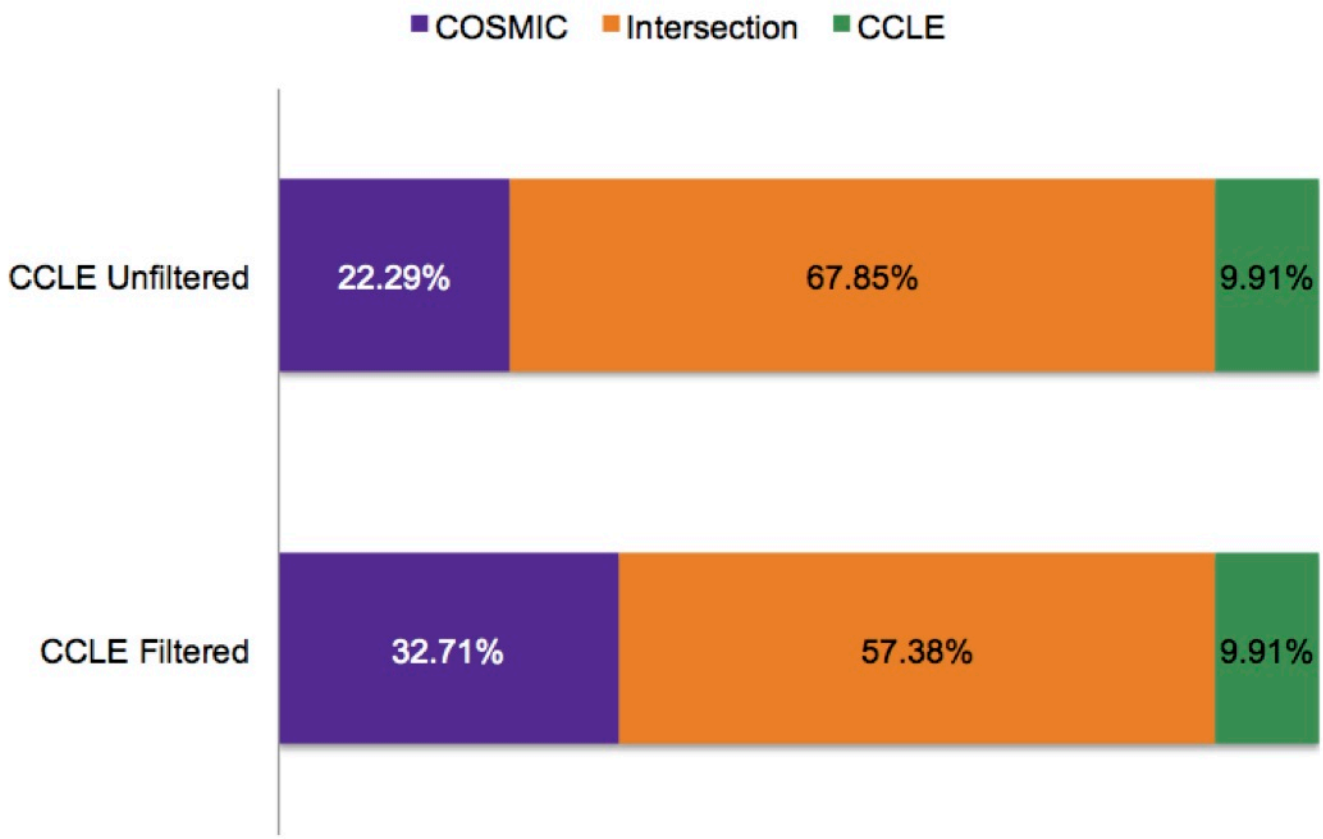
Supp. Figure 3

Western Blot showing PAK4 overexpression in 293T cells. Over-expression of the p.E119Q mutant causes enhanced phosphorylation of ERK compared to over-expression of the wild type plasmid. JNK phosphorylation activity is not affected.



Supp. Figure 4

The comparison of COSMIC and CCLE mutational data for 568 cell lines revealed an overall conformity of 57.38%. Taking the COSMIC only mutations from this analysis and comparing with the dataset from CCLE reported as being unfiltered for germ line SNPs increased the conformity to 67.85% indicating that differences in germ line SNP variant calling is not the sole cause of discrepancy between this data with 10091 unique COSMIC only mutations still remaining. The absence of published germ line SNP unfiltered COSMIC data prevented the opposite analysis being performed with CCLE only mutations.



Supp. Figure 5

Comparison of mutation detection in four cell lines (H2009, H1437, H2122, H2087) between COSMIC and CCLE along with sequencing from our own institute (CRUK MI) reveals 29.41% of mutations missed by CRUK MI. The conformity increases when the comparison is repeated with CRUK MI data that has not been filtered for germ line SNPs with 17.65% now being missed.