Supplementary Methods

Somatic variant identification

Data analysis was performed essentially as described previously (1). Briefly, variants called by default Personal Genome Machine (PGM) low stringency somatic variant filtering were further filtered to identify potential driving somatic mutations by removing synonymous or noncoding variants, those with flow corrected read depth (FDP) < 400, flow corrected variant allele containing reads (FAO) < 40, variant allele frequencies (FAO/FDP) < 0.05, flow variant allele calling forward to reverse read ratio < 0.2 or > 5, or indels within homopolymer runs ≥ 4. Variants occurring exclusively in reads containing other variants (single nucleotide variants or indels) or those occurring in the last mapped base of a read were excluded. Variants with 1) allele frequencies > 0.5 % in ESP6500 or 1000 genomes or 2) those reported in ESP6500 or 1000 genomes with observed variant allele frequencies between 0.40 and 0.60 or > 0.9 were considered germline variants unless occurring variants were at known hotspot locations. All variants were visually confirmed in Integrative Genomics Viewer (IGV, Broad Institute, https://www.broadinstitute.org/igv/) and paired samples were inspected in IGV to confirm lack of substantial read support for the called variant.

Reference