



Supplementary Figure S1: Quantitative germline background profiles.

Quantitative profiles of filtered germline variants potentially favoring tumorigenesis (see Supplementary Material and Methods) in genes involved in nine biological processes classified in the KEGG database. Individual patients have been clustered by cancer type, treatment, age and gender in the four UCSF patients (displayed at top, highlighted) and the 59 primary sarcoma patients from SJCRH. The total numbers of germline variants are shown for each patient (fifth column from the left).

Supplementary Materials and Methods

Bioinformatic Analyses

Read alignment and pre-processing

Alignment of the sequencing reads to the human reference genome (GRCh37 assembly) was performed with BWA [54], followed by duplicates tagging, indel realignment, and base quality score recalibration.

Germline variant calling

Single nucleotide variants (SNVs) and indels were then called across all germline samples using variant quality score recalibration according to GATK Best Practices recommendations [55,56]. Germline variants retrieved from a parallel call using Samtools (minimum base and mapping quality scores ≥ 30) [57] were retained.

Germline variant annotation and filtering

Region- and function-based variant annotations were performed with Annovar [58]. Rare alleles potentially favouring tumorigenesis susceptibility and present in $\leq 1\%$ of the general population have been kept for further analyses (pseudogenes and genes encoding unidentified proteins have been discarded). A proportional scoring model based on predictions of the following algorithms and association study databases has been developed in house in order to assess the pathogenicity level of the selected variants: SIFT [59], PolyPhen HVAR and HDIV [60], LRT [61], MutationTaster [62], MutationAssessor [63], FATHMM and FATHMM-MKL [64], PROVEAN [65], MetaSVM and MetaLR [66] as well as ClinVar database [67].

Dendrogram constructions

Unrooted dendrograms representing the qualitative similarities between the patient's filtered germline mutations were plotted using pairwise Jaccard coefficients (binary method: gene absent or present assigned a score of 0 or 1 respectively), followed by Ward's hierarchical clustering [68].

Gene copy number variation

Analyses of tumors gene copy number alterations, gene breaks, loss of heterozygosity and uniparental disomy have been performed with CNVkit v0.8.3 [69] using the paired germline and tumor samples.