Read number statistics

Tumor and normal samples were sequenced on Thermo Fisher Personal Genome Machines (PGMs) to an average coverage depth of ~1300X.

Alignment and variant calling

Reads were processed on the Ion Reporter cloud platform using built-in tools:

alignment to the reference human genome build hg19 was done by the Torrent Mapping Alignment Program (TMAP), followed by variant calling and tumor-normal subtraction by the Torrent Variant Caller (TVC) and variant annotation in the Ion Reporter Suite. Raw variant calls and annotation files were then downloaded to a local server for downstream processing.

Pipeline

Starting from the raw variant calls, we processed all samples using an in-house custom pipeline. Source code and documentation is publicly available at https://github.com/aweller/CancerPipeline. First, we set a quality cutoff of at least 400X coverage and a minimum alternative allele frequency of 0.05. Second, we included the annotation sources Annovar and SNPeff (ref1, ref2). in addition to the variant annotation provided by the Ion Reporter suite.

Deamination

Storage of FFPE samples can lead to DNA deamination, a process that is dominated by C:G>T:A transitions. Deaminated samples can be identified by a skewed mutation spectrum and an increased number of SNVs compared to normal samples. Correlating the total number of SNVs with the ratio of C:G>T:A variants among all SNVs per sample reveals a strong correlation (spearmans r: 0.77, p = 1.89e\*24). We excluded all samples with more than 300 SNVs from downstream analysis.

MutSigCV

Significantly mutated genes were identified using MutSigCV (ref3). MutSigCV corrects the mutation frequency for the observed patient-specific background mutation rate and spectrum as well as gene-specific information on gene expression level and replication time, both of which are correlated to mutation frequency.

REFERENCES

1)

Kai Wang, Mingyao Li, and Hakon Hakonarson

ANNOVAR: functional annotation of genetic variants from high-throughput sequencing Nucl. Acids Res. (2010) 38 (16)

2)

"A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.", Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. Fly (Austin). 2012 Apr-Jun;6(2):80-92. PMID: 22728672

3)

Mutational heterogeneity in cancer and the search for new cancer-associated genes  
Nature, 2013/07/11