Supplemental Methods

Exome Enrichment and Sequencing

Genomic libraries were prepared using the Illumina Paired End Sample Prep Kit following the manufacturer's instructions. Enrichment was performed as described previously (1) using the Agilent SureSelect Human All Exon 50Mb kit following the manufacturer's recommended protocol but excluding pre-enrichment PCR amplification. Each exome was sequenced using a 76bp paired-end protocol, and two lanes of an Illumina GAII DNA Analyser. Sequencing reads were aligned to the human genome (NCBI build 37) using the BWA algorithm on default settings (2). Reads which were unmapped, PCR derived duplicates, or outside the targeted region of the genome were excluded from the analysis. The remaining uniquely mapping reads provided approximately 70% coverage over the targeted exons at a minimum depth of 30X.

Variant Detection

CaVEMan (Cancer Variants through Expectation Maximisation) was used to call single nucleotide substitutions (1). This uses a naïve Bayesian classifier to estimate the posterior probability of each possible genotype (wild-type, germline, somatic mutation) at each base. We applied several post processing filters to the set of initial CaVEMan mutation calls in order to remove variants reported in poor quality sequence and increase the specificity of the output.

To call insertions and deletions, we used split-read mapping implemented as a modification of the Pindel algorithm (3). Pindel searches for reads where one end is anchored on the genome, and the other end can be mapped with high confidence in two (split) portions, spanning a putative indel. Post processing filters were applied to the output to improve specificity.

Mutations were annotated to Ensembl version 58.

Variant Validation

Validation of all 312 somatic variants was attempted by capillary re-sequencing of the tumour and normal pair. Where capillary sequencing failed, variants were reported as somatic if manual inspection of the aligned sequence reads provided strong evidence to support their validity.

Screening of SPEN and FGFR2 in additional cases by capillary sequencing
The coding exons of SPEN and FGFR2 were sequenced via PCR-based capillary sequencing as described previously. The sequence traces were analysed using semi-automated mutation detection followed by visual inspection of sequencing traces as previously described(4).

Identification of likely driver base substitutions and indels

A subset of the 312 substitution and indel somatic mutations identified in the exome screen were classified as “likely driver mutations” using conservative criteria. To do this we identified the established cancer genes from the Cancer Gene Census (www.sanger.ac.uk/genetics/CGP/Census/) which are known to be mutated by base substitutions and indels in order to contribute to cancer development. We then classified as likely driver mutations those which conformed to the known patterns of cancer-causing mutation for each cancer gene. Thus, for recessive cancer genes truncating mutations and essential splice site mutations were included. Missense mutations were also included where they had been seen previously or conformed to the known pattern of missense mutation in each gene (www.sanger.ac.uk/genetics/CGP/cosmic/ ). For established dominantly acting cancer genes we included mutations that had been previously registered in COSMIC.

Expression array analysis

RNA samples were available for twenty one of the twenty four samples in the original exome screen. RNA samples were hybridised to Illumina HumanHT-12 V4 BeadChips as per the manufacturer’s protocol. Un-normalised probe profiles were exported from Illumina GenomeStudio version 1.5.4. Variance stabilisation, normalisation, and differential gene expression analysis were performed using the lumi package of (5). Batch effect removal was performed using Combat(6).

SPEN expression by quantitative RT-PCR

Quantitative RT-PCR was performed by the Applied Biosystems 7900HT Real-time PCR Systems (Applied Biosystems) with KAPA SYBR FAST kit (KAPA Biosystems). The following primers were used for the expression of SPEN:

- SPEN-e2 (forward in exon 2): 5_-TCACTTCATGCACGAGAAGG-3_,
- SPEN-e3 (reverse in exon 3): 5_-TCGACTCCGAAGGCCTTAAT-3_,
- SPEN-e13 (forward in exon 13): 5_-CACTTGTCTCTGGCAACAA-3_,
- SPEN-e14 (reverse in exon 14): 5_-ATCTGTCTCCACCGTCATCC-3_.

The ACTB gene was used as internal control using primers 5_-TCACCGAGCGCGCTTCTTCTTCACC-3_ and 5_-TAATGTCACGCACGATTTCC-3_. Duplicate samples for each samples were analyzed. The expression of SPEN was determined by the _CT method (Average CT-SPEN
- Average CT-\textit{ACTB}), and relative \textit{SPEN} expression in each sample was calculated relative to \textit{SPEN} expression in normal salivary gland tissue (Clontech).

\textbf{References}


Supplemental Figure 2. *SPEN* expression by quantitative RT-PCR.

*SPEN* expression in each sample was calculated relative to *SPEN* expression in normal salivary gland tissue (Clontech). Expression was measured using primers designed to exons 2-3 (top histogram) and exons 13-14 (bottom histogram). In red are samples in which a *SPEN* mutation was identified; in blue are samples without a *SPEN* mutation. A253 is a salivary gland epidermoid carcinoma cell line.