Supplemental Methods

Next-generation sequencing

Nucleic acid extraction and sample preparation: DNA and RNA from bulk B16F10 cells and DNA fromC57BL/6 tail tissue were extracted in triplicate using Qiagen DNeasy Blood and Tissue kit (for DNA) and Qiagen RNeasy Micro kit (for RNA).

DNA exome sequencing: Exome capture for DNA resequencing was performed in triplicate using the Agilent Sure-Select whole-exome mouse solution-based capture assay (1) following the manufacturer's instructions (Supplemental information). 3 μg purified genomic DNA (gDNA) was fragmented to 150-200 bp using a Covaris S2 ultrasound device. Fragments were end repaired and 5' phosphorylated and 3' adenylated according to the maufacturer's instructions. Illumina paired end adapters were ligated to the gDNA fragments using a 10:1 molar ratio of adapter to gDNA. Enriched pre capture and flow cell specific sequences were added using Illumina PE PCR primers 1.0 and 2.0 for 4 PCR cycles. 500 ng of adapter ligated, PCR enriched gDNA fragments were hybridized to Agilent's SureSelect biotinylated mouse whole exome RNA library baits for 24 hrs at 65 °C. Hybridized gDNA/RNA bait complexes where removed using streptavidin coated magnetic beads, washed and the RNA baits cleaved off during elution in SureSelect elution buffer. These eluted gDNA fragments were PCR amplified post capture 10 cycles. Exome enriched gDNA libraries were clustered on the cBot using Truseq SR cluster kit v2.5 using 7 pM and 50 bps were sequenced on the Illumina HiSeq2000 using Truseq SBS kit-HS 50 bp.

RNA gene expression profiling (RNA-Seq): Barcoded mRNA-seq cDNA libraries were prepared in triplicate, from 5 μg of total RNA (modified Illumina mRNA-seq protocol). mRNA was isolated using Seramag Oligo(dT) magnetic beads (Thermo Scientific) and fragmented using

divalent cations and heat. Resulting fragments (160-220 bp) were converted to cDNA using random primers and SuperScriptII (Invitrogen) followed by second strand synthesis using DNA polymerase I and RNaseH. cDNA was end repaired , 5' phosphorylated and 3' adenylated according to the manufacturer's instructions. 3' single T-overhang Illumina multiplex specific adapters were ligated with T4 DNA ligase using a 10:1 molar ratio of adapter to cDNA insert. cDNA libraries were purified and size selected at 200-220 bp (E-Gel 2% SizeSelect gel, Invitrogen). Enrichment, adding of Illumina six base index and flow cell specific sequences was done by PCR using Phusion DNA polymerase (Finnzymes). All cleanups up to this step were done with 1,8x volume of AgencourtAMPure XP magnetic beads. All quality controls were done using Invitrogen's Qubit HS assay and fragment size was determined using Agilent's 2100 Bioanalyzer HS DNA assay. Barcoded RNA-Seq libraries were clustered and sequenced as described above.

NGS data analysis, gene expression: The output sequence reads from RNA samples were preprocessed according to the Illumina standard protocol, including filtering for low quality reads. Sequence reads were aligned to the mm9 reference genomic sequence (2) with bowtie (version 0.12.5) (3). For genome alignments, two mismatches were allowed and only the best alignment ("-v2 –best") was recorded; for transcriptome alignments the default parameters were used. Reads not alignable to the genomic sequence were aligned to a database of all possible exon-exon junction sequences of RefSeq transcripts (4). Expression values were determined by intersecting read coordinates with those of RefSeq transcripts, counting overlapping exon and junction reads, and normalizing to RPKM expression units (Reads which map per Kilobase of transcript per Million mapped reads) (5).

NGS data analysis, somatic mutation discovery: Somatic mutations were identified as previously described (cite Löwer et al, submitted). 50 nucleotide (nt), single-end reads were aligned to the mm9 reference mouse genome using bwa (default options, version 0.5.8c) (6). Ambiguous reads mapping to multiple locations of the genome were removed. Mutations were identified using three software programs: samtools (version 0.1.8) (7), GATK (version 1.0.4418) (8), and SomaticSniper (9). Potential variations identified in all B16F10 triplicates were assigned a "false discovery rate" (FDR) confidence value (Löwer *et al*., submitted).

Mutation selection, validation, and function

Selection: mutation selection criteria were: (i) present in all B16F10 and absent in all C57BL/6 triplicates, (ii) FDR ≤0.05, (iii) homogeneous in C57BL/6, (iv) occur in a RefSeq transcript, and (v) cause non-synonymous changes. Furthermore, mutations were selected for validation and immunogenicity testing that occur in B16F10 expressed genes (median RPKM across replicates >10) and in an MHC-binding peptide based on the Immune Epitope Database (IEDB) (10).

Validation: DNA-derived mutations were classified as validated if confirmed by either Sanger sequencing or the B16F10 RNA-Seq reads. All selected variants were amplified from 50 ng of DNA from B16F10 cells and C57BL/6 tail tissue, products visualized (QIAxcel system, Qiagen) and purified (QIAquick PCR Purification Kit, Qiagen). The amplicon of the expected size was excised from the gel, purified (QIAquick Gel Extraction Kit, Qiagen) and subjected to Sanger sequencing (Eurofins MWG Operon, Ebersberg, Germany) with the forward primer used for PCR amplification.

Functional impact: The programs SIFT (11) and POLYPHEN-2 (12) predict the functional significance of an amino acid on protein function based on the location of protein domains and

cross-species sequence conservation and were employed to assess the impact of selected mutations. Ingenuity IPA was used to infer gene function

Supplemental References

(1) Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009;27:182-9.

(2) Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520-62.

(3) Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.

(4) Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 2007;35:D61-D65.

(5) Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621-8.

(6) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.

(7) Li H. Improving SNP discovery by base alignment quality. Bioinformatics 2011;27:1157-8.

(8) McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

(9) Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: Identification of Somatic Point Mutations in Whole Genome Sequencing Data. Bioinformatics 2011.

(10) Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 2008;36:W509-W512.

(11) Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.

(12) Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.