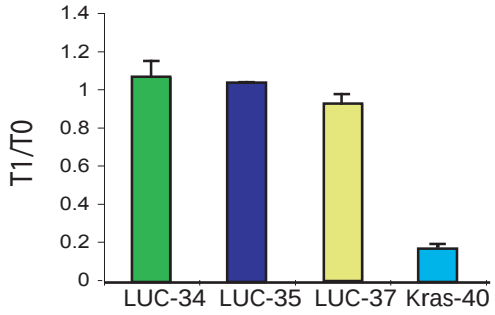
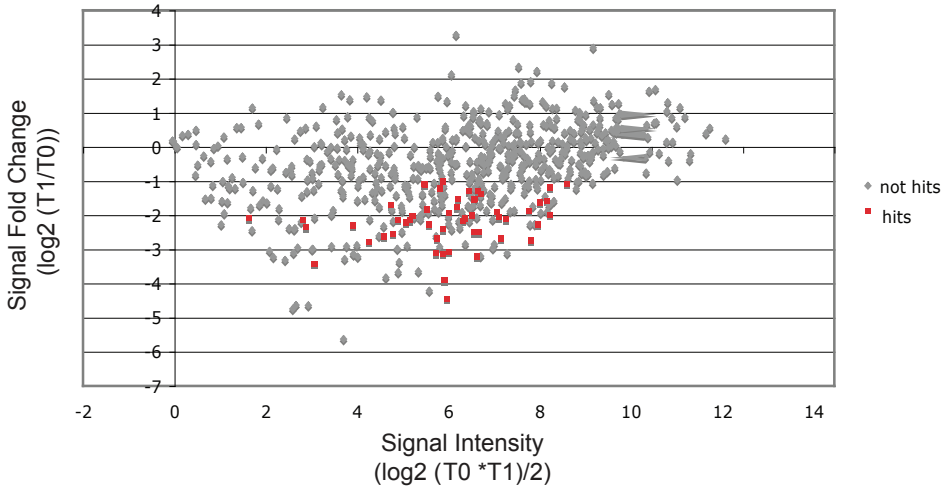
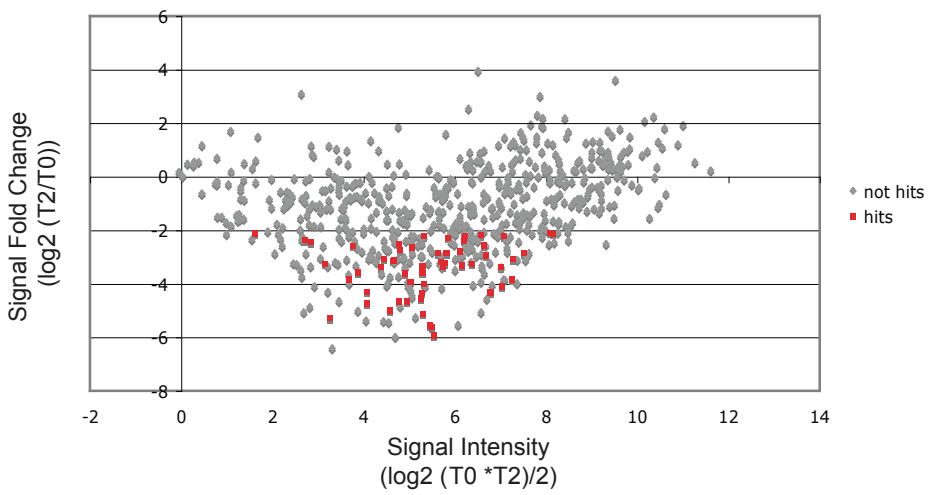
A


T1/T0
1.4
1.2
1
0.8
0.6
0.4
0.2
0
LUC-34
LUC-35
LUC-37
Kras-40

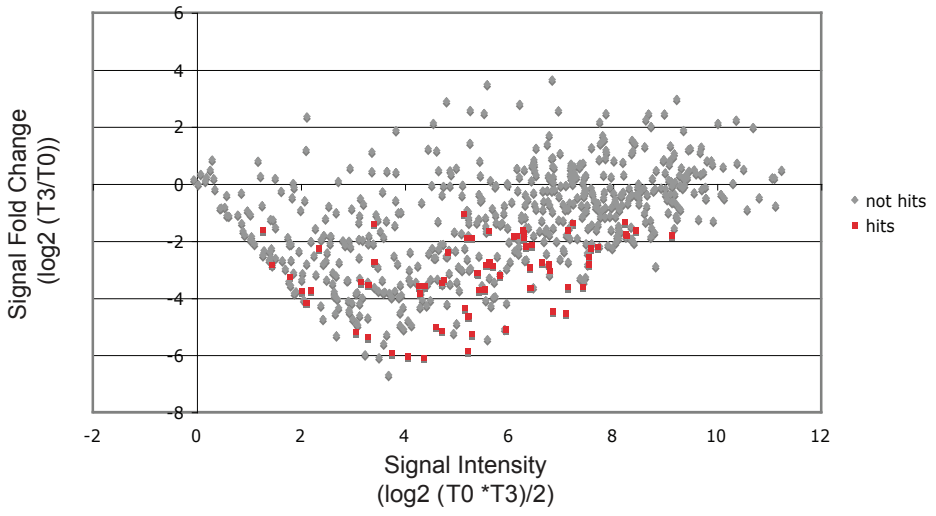
B

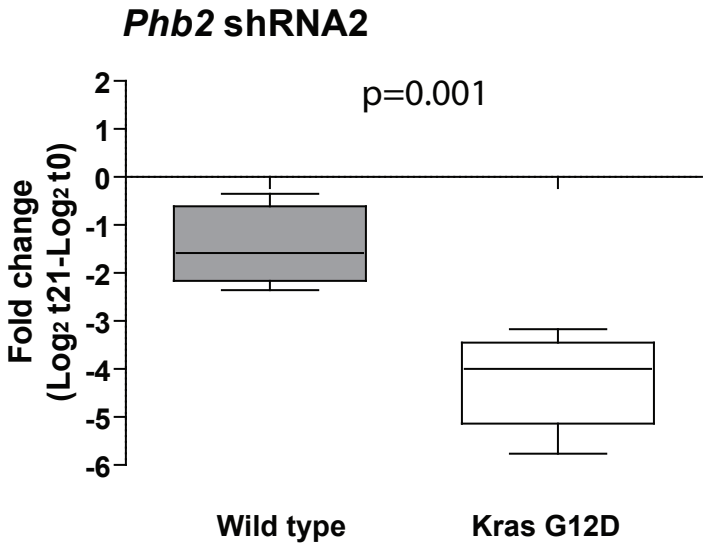

Signal Fold Change
(log2 (T1/T0))
Signal Intensity
(log2 (T0 *T1)/2)
not hits
hits

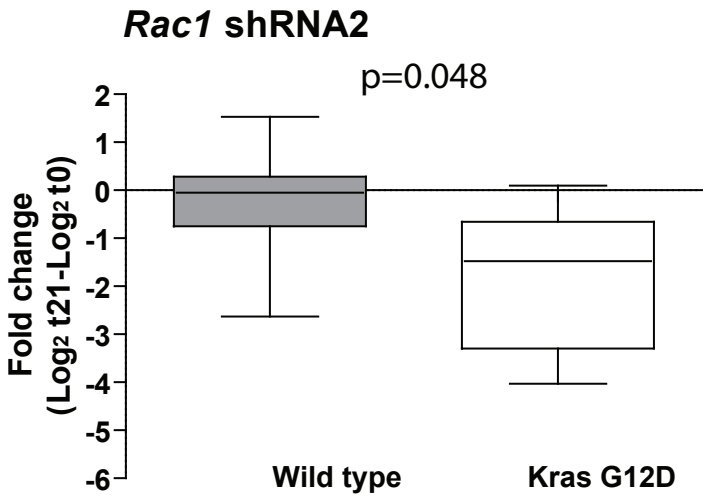
C

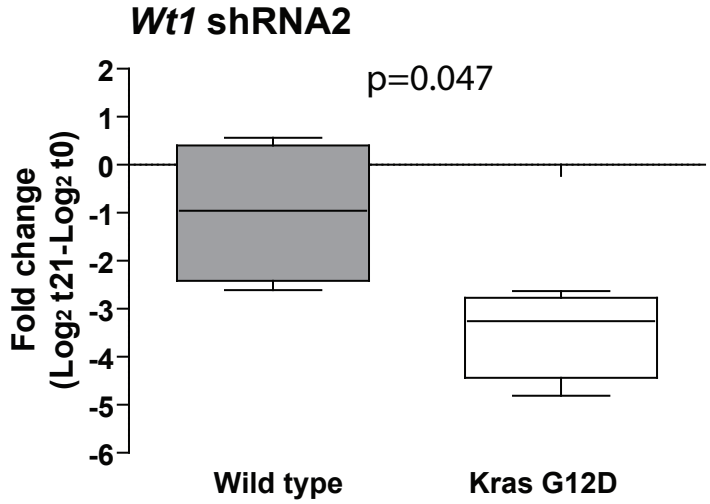

Signal Fold Change
(log2 (T2/T0))
Signal Intensity
(log2 (T0 *T2)/2)
not hits
hits

D

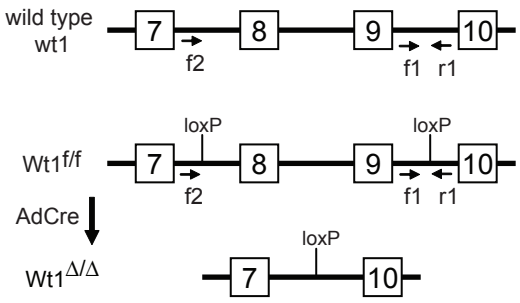

Signal Fold Change
(log2 (T3/T0))
Signal Intensity
(log2 (T0 *T3)/2)
not hits
hits


Phb2 shRNA2
p=0.001
Fold change (Log2 t21-Log2 t0)
2
1
0
-1
-2
-3
-4
-5
-6
Wild type
Kras G12D

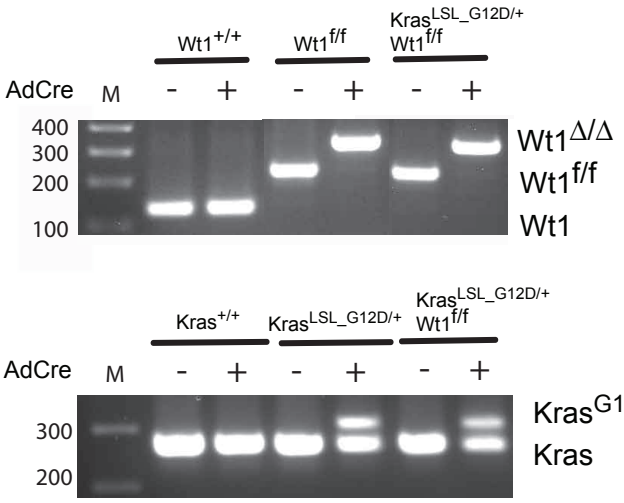

Rac1 shRNA2
p=0.048
Fold change (Log2 t21-Log2 t0)
2
1
0
-1
-2
-3
-4
-5
-6
Wild type
Kras G12D

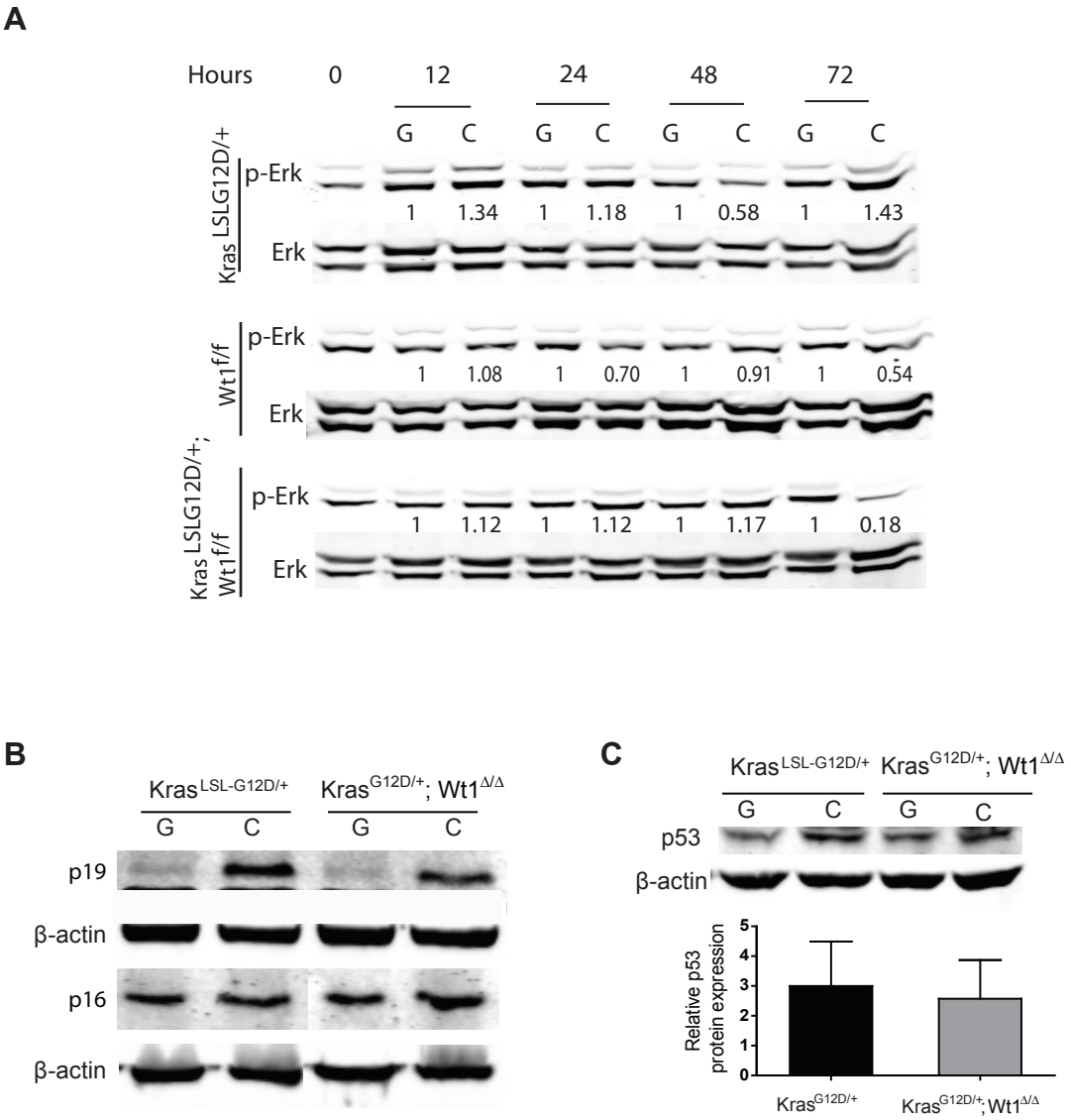

Wt1 shRNA2
p=0.047
Fold change (Log2 t21-Log2 t0)
2
1
0
-1
-2
-3
-4
-5
-6
Wild type
Kras G12D

A

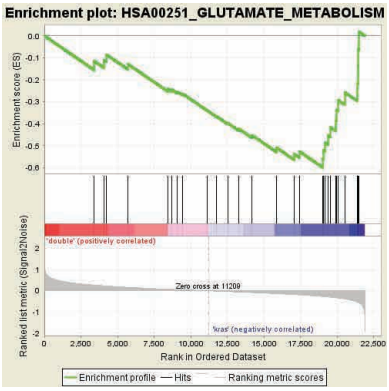

wild type
wt1
7
8
9
10
f2
f1
r1
loxP
loxP
Wt1f/f
AdCre
Wt1Δ/Δ

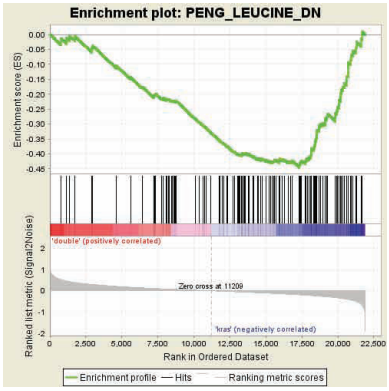
B


Wt1+/+
Wt1f/f
KrasLSL_G12D/+ Wt1f/f
AdCre
M
400
300
200
100
Wt1Δ/Δ
Wt1f/f
Wt1
Kras+/+
KrasLSL_G12D/+
KrasLSL_G12D/+ Wt1f/f
300
200
KrasG12D
Kras

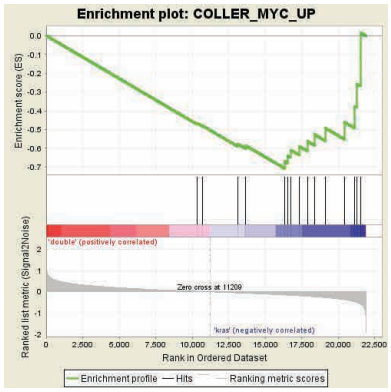

A
Hours
0
12
24
48
72
G
C
Kras LSLG12D/+
p-Erk
1
1.34
1
1.18
1
0.58
1
1.43
Erk
Wt1f/f
p-Erk
1
1.08
1
0.70
1
0.91
1
0.54
Erk
Kras LSLG12D/+; Wt1f/f
p-Erk
1
1.12
1
1.12
1
1.17
1
0.18
Erk
B
Kras LSL-G12D/+
Kras G12D/+; Wt1 Δ/Δ
p19
β-actin
p16
β-actin
C
p53
β-actin
Relative p53 protein expression
0
1
2
3
4
5
Kras G12D/+
Kras G12D/+; Wt1 Δ/Δ

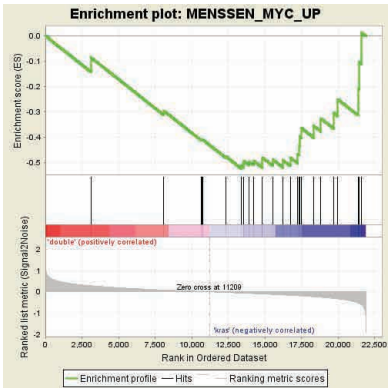
A


Enrichment plot: HSA00251_GLUTAMATE_METABOLISM
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11209
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

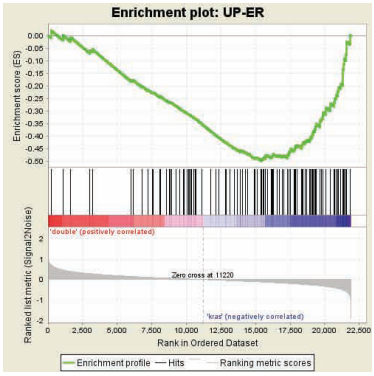

Enrichment plot: PENG_LEUCINE_DN
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11209
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

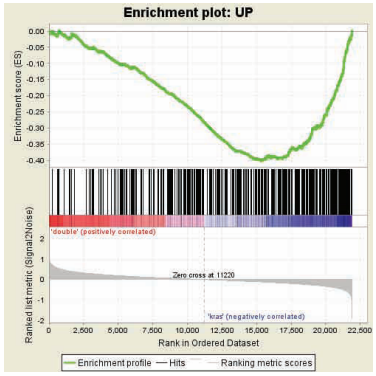
B

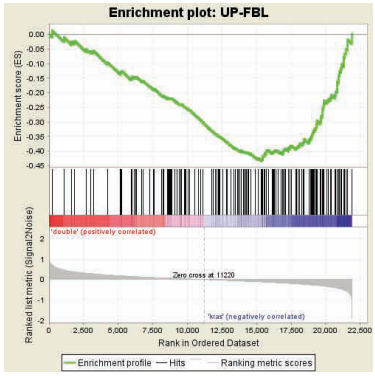

Enrichment plot: COLLER_MYC_UP
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11209
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

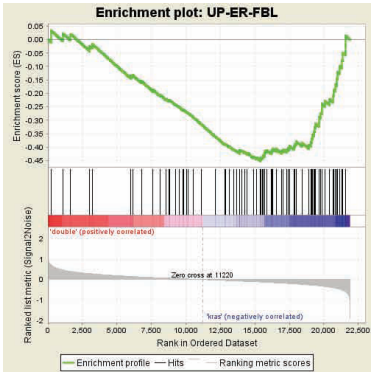

Enrichment plot: MENSSEN_MYC_UP
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11209
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

C

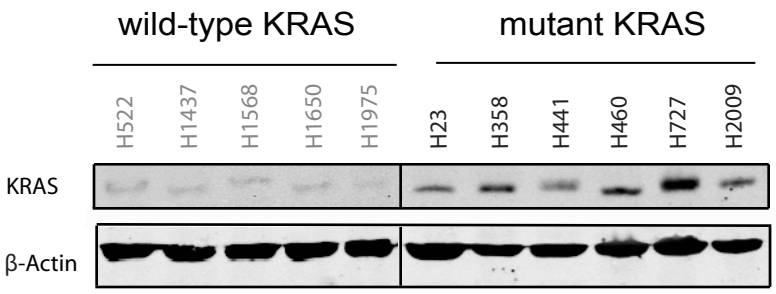

Enrichment plot: UP-ER
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11220
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores


Enrichment plot: UP
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11220
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

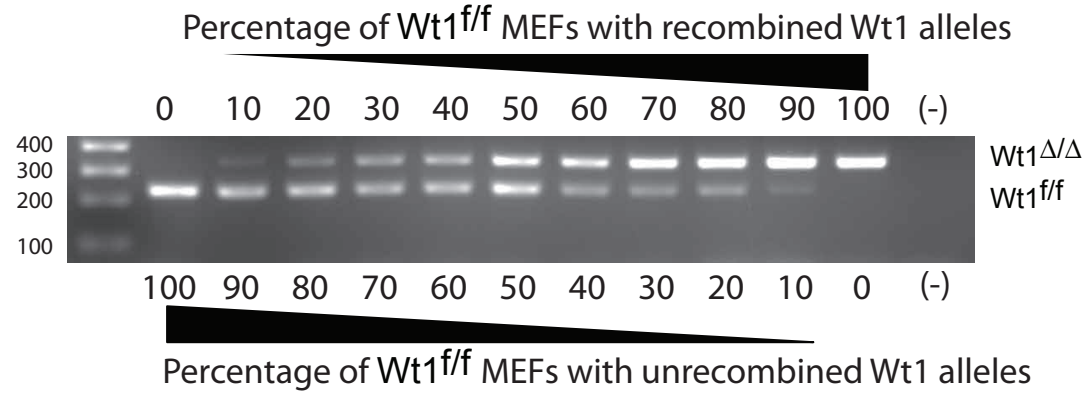

Enrichment plot: UP-FBL
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11220
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores


Enrichment plot: UP-ER-FBL
Enrichment score (ES)
'double' (positively correlated)
Zero cross at 11220
'kras' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

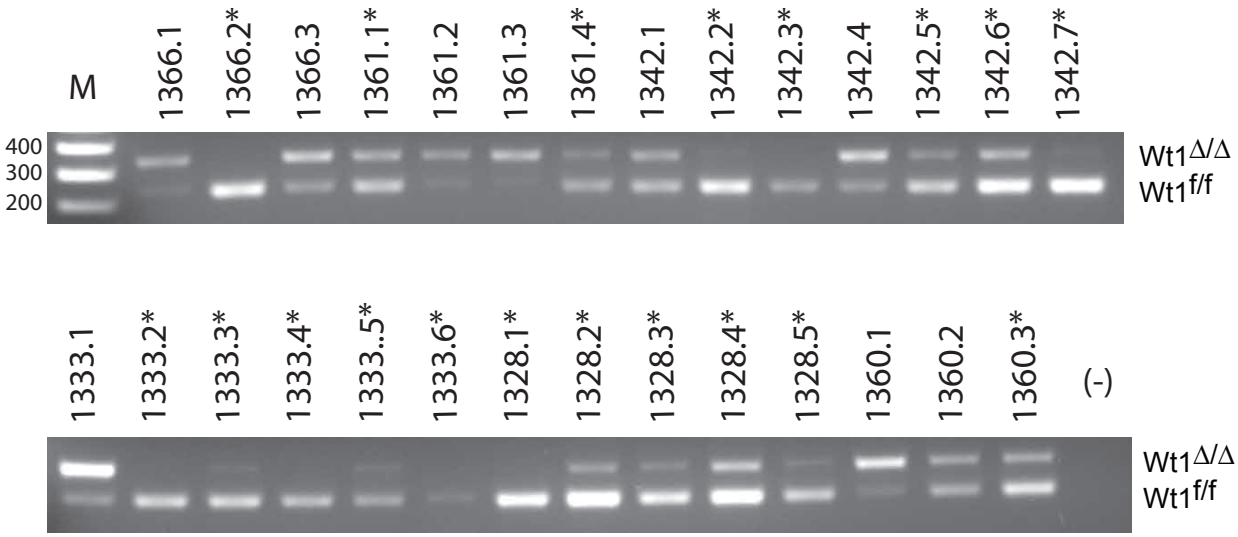
A


wild-type KRAS
mutant KRAS
H522
H1437
H1568
H1650
H1975
H23
H358
H441
H460
H727
H2009
KRAS
β-Actin

A


Percentage of Wt1f/f MEFs with recombined Wt1 alleles
0 10 20 30 40 50 60 70 80 90 100 (-)
400
300
200
100
Wt1Δ/Δ
Wt1f/f
100 90 80 70 60 50 40 30 20 10 0 (-)
Percentage of Wt1f/f MEFs with unrecombined Wt1 alleles

B


M
1366.1
1366.2*
1366.3
1361.1*
1361.2
1361.3
1361.4*
1342.1
1342.2*
1342.3*
1342.4
1342.5*
1342.6*
1342.7*
400
300
200
Wt1Δ/Δ
Wt1f/f
1333.1
1333.2*
1333.3*
1333.4*
1333..5*
1333.6*
1328.1*
1328.2*
1328.3*
1328.4*
1328.5*
1360.1
1360.2
1360.3*
(-)
Wt1Δ/Δ
Wt1f/f

SUPPLEMENTARY INFORMATION

WILMS TUMOR-1 (*WT1*) REGULATES *KRAS*-DRIVEN ONCOGENESIS AND SENESCENCE IN MOUSE AND HUMAN

Silvestre Vicent, Ron Chen, Leanne C. Sayles, Chenwei Lin, Randal G. Walker, Ana K. Gillespie, David E. Root, Aravind Subramanian, Greg Hinkle, Xiaoping Yang, Vicki Huff, William C. Hahn and E. Alejandro Sweet-Cordero

Figure Legends

Supplementary Figure 1. (**A**) A *Kras*-specific shRNA is negatively selected in LKR13 cells. 3 different control (luciferase) lentiviral vectors, each carrying a different shRNAs with a unique barcode, were mixed at equimolar concentrations with a *Kras* specific shRNA linked to a different barcode. Virus was used to infect LKR13 cells and cells were allowed to proliferate (splitting every 3 days) for 3 weeks. Results show negative selection of the K-ras specific shRNA over time. Error bars show ± s.d. (**B-D**) Results of Luminex bead pooled screen for each time point. shRNAs shown in Table S1A were used to infect LKR10 and LKR13 cell lines. Results shown are average across both cell lines (full results for all 3 time points are in Table S1B). T0= 3 days after infection with lentivirus. T1=3 weeks of passage in vitro. T2= 6 weeks of passage in vitro. T3= 3 weeks of subcutaneous tumor growth. (**B**) LKR10 and LKR13 T1 *vs.* T0. (**C**) LKR10 and LKR13 T2 *vs.* T0. (**D**) LKR10 and LKR13 T3 *vs*. T0.

Supplementary Figure 2. Negative selection of *Wt1*, *Phb2* and *Rac1* in a MEFs screen. Graphs show the results of a second shRNA used in the MEFs screen. Box plots indicate the mean and standard deviation of MFI fold change of the respective gene when comparing T1 *vs.* T0. P values were obtained using a two-tailed t test.

Supplementary Figure 3. Recombination of conditional alleles for *Kras* and *Wt1* in MEFs. (**A**) Schematic representation of the multiplex PCR used to detect recombination of the *WT1* allele. f1, forward primer 1; f2, forward primer 2; r1, reverse primer 1. (**B**) PCR showing recombined alleles for MEFs with indicated genotypes after Adenoviral Cre infection.

Supplementary Figure 4. *Wt1* loss does not alter Ras protein levels in MEFs yet regulates output of Erk and trp53 signaling. (**A**) Time course of phosphorylation of Erk in MEFs. A second independent MEF line was used to repeat the analysis shown in this figure except that lysates were collected at 12, 24, 48 and 72 hours (data not shown). Western blotting of MEFs with (**B**) p19, p16 and (**C**) p53 antibodies. β-actin was used as loading control. Error bars show ± s.d.

Supplementary Figure 5. Gene Set Enrichment Analysis in $Kras^{G12D/+}$ and $K\text{-}ras^{G12D/+}$; $Wt1^{\Delta/\Delta}$ MEFs. Enrichment plots of (**A**) MSigDB gene sets of glutamine metabolism genes as well as experimentally derived gene-sets down-regulated

by glutamate or leucine starvation, (**B**) MSigDB gene sets for *Myc* pathway genes and (**C**) new curated gene sets of MYC target genes.

Supplementary Figure 6. (**A**) Protein expression levels of endogenous KRAS in wild type and mutant *KRAS* NSCLC cell lines. The lanes were run on the same gel but were noncontiguous. β-actin was used as loading control.

Supplementary Figure 7. (**A**) RT-PCR to detect recombination of $Wt1^{f/f}$ alleles in mixtures of MEFs with specific percentages of recombined alleles. (**B**) RT-PCRs to detect recombination of $Wt1^{f/f}$ alleles in laser capture microdissected lung tumors from $Kras^{G12D/+}$; $Wt1^{\Delta/\Delta}$ mice. Asterisks indicate the tumors that were removed from the analysis.

Supplementary Table Legends

Tables related to **Figure 1**: a negative selection screen to identify *K-ras* effectors.

Supplementary Table 1. List of shRNA included in the screen. Experimental source refers to the reason the gene was included in the screen. A549-genes knocked-down by a *Kras* shRNA in A549 lung cancer cell line[1]. *Kras*sig-genes previously identified as part of a *Kras* signature[1]. Adenocarcinoma-genes upregulated in human NSCLC vs. normal tissue[1]. KSS-gene identified by KSS scanning (see Supplemental Methods and Supplementary Table 3). Annotated-genes found in the literature as potential KRas effectors. Refseq-NCBI reference sequence id. Clone_id-ID # from Broad Institute TRC library. Pool_id-Luminex pool # for screening. Region-region of gene to which shRNA corresponds (cds or 3'UTR). Target Sequence-sequence of shRNA target.

Supplementary Table 2. Luminex LKR screen results for all time points. Results are plotted in Supplementary Fig. 2. Time point 1-in vitro 3 weeks. 2-in vitro 6 weeks. 3-in vivo 3 weeks (subcutaneous tumors). M=log2 (T1, 2 or 3/T0). A-log2 (T1, 2 or 3-T0)/2).

Supplementary Table 3. KSS analysis of transcription factors in the Global Cancer Map. Microarrays in the Global Cancer Map were analyzed using the KSS approach as described in supplemental methods. Enrichment scores are the KSS scores for the *Kras* signature for the rank-ordered list of correlations for each gene (see Supplementarly Methods). Results shown are for all genes (right) as well as only transcription factors (left).

Tables related to **Figure 4**: gene expression analysis of $Kras^{G12D/+}$ and $Kras^{G12D/+}$; $Wt1^{\Delta/\Delta}$ MEFs.

Supplementary Table 4. gct file of MEF microarray data.

Supplementary Table 5. Gene sets enriched in $Kras^{G12D/+}$ vs. $Kras^{G12D/+}$; $Wt1^{\Delta/\Delta}$ MEF using GSEA.

Supplementary Table 6. Gene sets derived from a database of *MYC* target genes (see text).

Supplementary Table 7. PAM analysis of MEF microarray data.

Supplementary Table 8. PAM analysis using only genes for which a single orthologue was found when comparing the mouse and human gene expression data.

Supplementary Tables 9 and 10. *KRAS* signature genes used to separate data in Shedden et al. 2008 into "*KRAS* signature high" and "*KRAS* signature low". The *KRAS* signature genes were defined as described in Luo et al. Cell, 2009.

Supplementary Experimental Procedures

Kolmogorov-Smirnov Scanning

Gene expression data from the Global Cancer Map (GCM; 190 specimens representing 14 tumor types, 16,063 probes, Affymetrix GeneChip Hu6800 and Hu35KsubA)[2] were downloaded from www-genome.wi.mit.edu/cgi-bin/cancer/datasets.cgi. Expression values below a baseline of 20 were set to 20. The dataset was collapsed to gene symbols using the collapse dataset function in the Gene Set Enrichment Analysis (GSEA) software (Broad Institute), resulting in 10,386 genes. Normalization was performed by standardizing each row (gene) to mean 0 and standard deviation 1. A matrix of gene-to-gene correlations was produced using Pearson correlation. To scan this correlation matrix for genes whose expression pattern correlates with the *Kras* signature, we adopted the Kolmogorov-Smirnov Scanning strategy previously described[3]. For each gene, its list of correlations was rank ordered and enrichment score (ES) computed. The score quantifies the degree in which a gene set, in this case the *Kras* signature, is enriched at the top or bottom of a rank-ordered list. Specifically, it is calculated by walking down a list, increasing a running-sum statistic when a gene within the gene set is encountered, or decreasing it when a gene outside of the gene set is encountered[4]. Genes were then ranked by their ES score. Transcription factors were identified by the Molecular Signatures Data Base (MSigDB) (http://www.broad.mit.edu/gsea/msigdb/, total of 921 transcription factors). All transcription factors with a KS score greater than that of *Kras* itself were entered into the screen (Supplementary Table 3).

Preparation of shRNA library and samples for Luminex analysis

To prepare the shRNA library used in our study, one hundred 25-mer "barcodes" were cloned into a lentiviral shRNA vector immediately downstream of the shRNA cloning site[5,6]. In parallel, the anti-sense sequence to each of these barcodes was covalently coupled to each of one-hundred color-coded beads. The quantification of the barcodes using this approach provides a surrogate marker for the number of cells carrying each shRNA in a pooled sample[7].

The preparation of samples for the Luminex run is described hereon. A 5' primer (5'GAGGGCCTATTTCCCATGAT) in the U6 promotor of plko.1s and a 3' (5'GTAATACGACTCACTATAGGGCCTCTTCGGAGATCAGCTTC) primer downstream of the Xma/Nhe insertion site in plko.1s were used to amplify the shRNA region containing the detection oligonucleotide ("barcode"). PCR reactions were set up for each of the templates (denature 94ºC for 45 sec,

anneal 56ºC for 45 sec and extend 72ºC for 45 sec). PCR products were purified using a Zymo DNA Clean & Concentrator 5 (cat # D4004) and 2μl was used as a template for IVT using an Ambion megascript short kit (cat # 1354). 1.5 μl of each biotinylated oligonucleotide (Enzo Life sciences, Bio-16-UTP cat # 42814B and Bio-11-CTP cat # 42818B) was added to the IVT reaction to label the RNA. A Qiagen RNeasy column was used to remove unbound nucleotides and 100ng of RNA was annealed to the Luminex Beads (denature, 94C for 2 min, 45C for 90 min) in the presence of 3M TMAC (sigma T-3411). After incubation a reporter mix containing SAPE (Invitrogen) was added and the sample was read on a Luminex 100 apparatus.

FlexMAP carboxylated microspheres ("beads") were purchased from Luminex Corporation (Austin, Texas, USA). Each bead set was coupled to a unique 24-nucleotide sequence ("capture probe") that contained a 5' amino C12 modification. The microspheres were resuspended by sonication and vortexing and 2.5 million beads in a volume of 200μl was pelleted by centrifugation then resuspended in 25μl of 0.1M MES, pH 4.5 (Sigma M-2933). The capture probes were diluted to 100 μM and 1μl was added to the microspheres after resuspension. 2.5μl of EDC 10mg/mL (cat #22980, Pierce) was added to the samples and incubated for 30 min in the dark. This step was repeated twice to increase density of the probe on the bead surface. The beads were washed with 500 μl 0.02% Tween-20, then with 0.1% SDS and finally TE, pH 8.0. The beads were resuspended in 50 μl TE, pooled together with the other coupled bead sets and stored in the dark at 4C.

Luminex Data Analysis

Background correction and normalization were performed within plate (each plate reflecting processing associated with one or more pools and samples). Specifically, the intensity for a given hairpin/barcode was reduced by the intensity obtained for that barcode using the negative control (water). Background-adjusted intensities were then scaled (using the individual to median ratio of overall sample fluorescence) and converted to the $\log_{(2)}$ scale. To increase discriminatory power, the LKR10 and LKR13 measurements were combined (resulting in six measurements per time point). A student's *t* was subsequently generated for all timepoints relative to baseline.

Statistical significance for each hypothesis test was determined using the local false discovery methodology (locfdr) of Efron et al [8]. In this approach, a mixture model for the test-statistic distribution utilizing an invariant subset is used to refine false discovery rates at the tails. This approach was initially designed for microarray experiments in which a modest proportion of the probes on the array are unlikely to change across the experimental conditions under study. In the current context, each pool contained several barcodes which did not have a corresponding hairpin thereby providing an invariant subpopulation whose test statistic distribution was symmetric and approximately normally distributed. As such, the theoretical null option of locfdr was specified and any hairpin/barcode

with an identified fdr of <0.01 for any time point comparison was retained. The fdr cut off was selected to be highly conservative with an appreciation that the distribution of hairpins which failed to perform could not be empirically differentiated (and therefore used as the invariant subset) and with the objective of rank ordering obtained results.

Remaining hairpin/barcodes which displayed increased negative selection over time and which were concordant between the in vitro and in vivo screens were then identified. This was achieved by imposing a threshold change in the $\log_{(2)}$ difference relative to the initial time point for measurements at 3 weeks (-1 $\log_{(2)}$ or a 2-fold change) and 6 weeks (-2 $\log_{(2)}$ or a 4-fold change). A final filter was applied to ensure that the baseline signal was an order of magnitude (four fold) above the dynamic range of the background.

Genes were included in the final list (Table **1**) if one shRNA met all strict criteria and a second shRNA met the fold change criteria even if the fdr was outside the cut off. In addition, genes with two shRNAs that were the fold change criteria were also included even if the fdr was outside the cut-off.

Laser capture microdissection (LCM)
Lungs embedded in paraffin were cut into 5uM sections onto Leica PEN-Membrane slides (No. 11505158) and baked for 1 hr at 55C. Slides were deparaffinized in xylene (2 times 5 min.) then rehydrated through Ethanol gradient (100%, 96%, then 70%, one minute each) and dipped in distilled water several times. Slides were then stained with Methyl Green for 5 minutes, rinsed in distilled water and allowed to air-dry overnight. Tumors were isolated using a Leica Laser Microdissection System at a 20X magnification and captured into the lid of a 0.2mL PCR tube for genomic DNA isolation. DNA was isolated using the QIAamp DNA Micro Kit following the protocol for the isolation of Genomic DNA from Laser-Microdissected Tissues.

Analysis of a high/low WT1-gene signature in human lung cancer
Among human lung carcinomas[9] a group of "*KRAS* signature high" and "*KRAS* signature low" were identified (Supplementary Table 6), using a gene signature of mutant *KRAS* developed from a published lung[10] and validated in another two public data sets[11-13].

To explore mouse-human relevance, MEFs and human lung carcinoma microarray data[9] were first collapsed to gene symbols, according to maximum of gene probes. Orthologous genes that are present in both human and mouse array were selected, based on the mammalian orthology from the Mouse Genome Informatics (http://www.informatics.jax.org/). The mouse ortholog gene array data was further converted to standard deviations from the mean of each gene. The human lung carcinoma array data were generated at four centers, and so standardization was performed in each center subset separately.

Orthologous gene signals were then subjected to Prediction Analysis for Microarrays (www-stat.stanford.edu/~tibs/PAM/;[14]). After training with MEFs double mutants versus *Kras* mutant data, PAM was able to separate "high WT1-gene signature" and "low WT1-gene signature" samples within human lung carcinoma "*KRAS* signature high" and "*KRAS* signature low" groups. In both "*KRAS* signature high" and "*KRAS* signature low" groups, "high WT1-gene signature" samples were compared with "low WT1-gene signature"" samples, using Kaplan-Meier analysis for patient survival time. Microarray and patient clinical covariates data of Shedden et al. 2008 are available at https://caarraydb.nci.nih.gov/caarray/publicExperimentDetailAction.do?expId=10 15945236141280.

shRNA sequences (mouse shRNAs targeted sequence and clone number in TRC library when available)

GFP shRNA
GCAAGCTGACCCTGAAGTTCAT

Mouse WT1 shRNA 1 (TRCN0000054464)
ACCATACCAGTGTGACTTCAA

Mouse WT1 shRNA 2 (TRCN0000054467)
CAGCCTACCATCCGCAACCAA

Mouse K-ras shRNA 1
TGGTTGGAGCTGATGGCGTAG

Mouse K-ras shRNA 2
TTGGAGCTGATGGCGTAGGCA

Human WT1 shRNA 1 (TRCN0000040066)
GCATCAGAGAAACATGACCAA

Human WT1 shRNA 2 (TRCN0000040067)
GCATCTGAGACCAGTGAGAAA

Primer sequences for rtRT-PCR

GENE	Forward	Reverse
hprt	tgacactggtaaaacaatgca	tgacactggtaaaacaatgca
Apbb2	ccaacgatccatgtcagaag	gccataaacactgccaaggt
Braf	catatagaggccctattggacaaa	tgctggtgtactcttcataggc
Csf2	gcatgtagaggccatcaaaga	cgggtctgcacacatgtta
Eef2	cttcactgacactcgcaagg	gctcgtagaagagggagatgg
Elf3	ccagaaagctgagcaaggaa	ctcggataaactcccacagg
Foxo1	gctgggtgtcaggctaagag	gcatctttggactgctcctc
IL18	catgtacaaagacagtgaagtaagagg	tttcaggtggatccatttcc

Kras	tggatgagtacgaccctacga	tccctcattgcactgtactcc
Mrc1	ggacgagcaggtgcagtt	caacacatcccgcctttc
Nme2	aaacggttcgagcagaagg	gctgcttcaggtgttcttca
Nmt1	cccccaggagaacatcatt	tgtgggttggatggttcat
Pde4c	ccagtttctcatcaacaccaac	agcctaccgcgaggtgat
Pitx1	aatcgtccgacgctgatct	ttcttcttagctgggtcctctg
Phb2	cattgttaatgaggtgctcaaga	cttcggatcaacagggacac
Rac1	gatgcaggccatcaagtgtgt	agcaggcaggttttaccaaca
Rap1ga1	gcgtagacggggatgataca	aggaaggagtcccgcttatg
Rara	tgccatctgcctcatctgt	cagcatgtccaccttgtctg
Rassf1	gagcagcacaaccagcagt	ccgctctacagcctcatcc
Rbm6	tctcagggcaaaatgtcca	gcttgggccagtcctataatc
Slc25a5	gatgccgctgtgtccttc	tatctgccgtgatttgcttg
Tgfbi	gagctgcttatcccagattca	ggcagtggagacgtcagatt
Tkt	caccgtggaggaccactact	tccaggttcacccactacg
Vgll1	ttcaggagaactgaaagacgtg	gaagagagatgcctctgattcg
Wt1	caccaaaggagacacacaggt	gggaaaactttcgctgacaa

Supplementary references

1 Sweet-Cordero, A. *et al.* An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* (2004).

2 Ramaswamy, S. *et al.* Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 98, 15149-15154 (2001).

3 Lamb, J. *et al.* A mechanism of cyclin d1 action encoded in the patterns of gene expression in human cancer. *Cell* 114, 323-334 (2003).

4 Subramanian, A. *et al.* From the Cover: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-15550 (2005).

5 Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124, 1283-1298 (2006).

6 Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *Rna* 9, 493-501 (2003).

7 Peck, D. *et al.* A method for high-throughput gene expression signature analysis. *Genome Biol* 7, R61 (2006).

8 Efron, B. & Tibshirani, R. Empirical bayes methods and false discovery rates for microarrays. *Genet Epidemiol* 23, 70-86 (2002).

9 Shedden, K. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 14, 822-827 (2008).

10 Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 98, 13790-13795 (2001).

11 Beer, D. G. *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8, 816-824 (2002).

12 Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455, 1069-1075 (2008).

13 Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* 137, 835-848 (2009).
14 Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 99, 6567-6572 (2002).