Affymetrix gene profiling and data analysis

Duplicate independent preparations of total RNA from both of the *taf4*$^{lox/-}$ and *taf4*$^{-/-}$ clones exponentially growing in 10% serum (total of 8) were prepared with the Qiagen column kit and treated with DNAse (5 U/100 µg RNA, Sigma). Biotinylated cRNA was prepared according to the standard Affymetrix protocol (Expression Analysis Technical Manual, 2004; Affymetrix), purified and fragmented according to Affymetrix recommendations. 10 µg of fragmented cRNA were hybridised for 16 hr at 45°C on the Mouse Genome GeneChip 430A 2.0. comprising 22 600 probe sets representing over 14 500 mouse genes. GeneChips were washed and stained in the Affymetrix Fluidics Station 400 and scanned using the Affymetrix GeneChip Scanner 3000. The image data were analysed with Microarray Suite version 5.0. (MAS5) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each chip was arbitrarily set to 100. Absolute analyses generate a signal value for each probe set and a detection call of absent, present or marginal (Hubbell et al., 2002; Liu et al., 2002). Finally, an additional normalization by the Quantil method described by Bolstad *et al.* (Bolstad et al., 2003) was performed.

Pair wise comparisons were made between the 4 values from the *taf4*$^{-/-}$ clones versus the 4 values from the *taf4*$^{lox/-}$ clones. In order to identify differentially expressed genes, we performed a Wilcoxon test where the expression levels were rank-ordered and tied ranks were given ex-aequo expression levels. The ranks for the larger group were summed and that number was compared against a precalculated table to obtain a p-value. All genes with a p-value less than 0,05 were considered as significant. To evaluate the false discovery rate due to multiple comparisons, a bootstrap analysis was carried out (Tsai et al., 2003). A random permutation test of 1000 combinations was done for each gene. From these results, the

greatest percentage of false positive observed was 0,046 at 5% confidence level. Genes whose expression changed significantly between two sample groups were also selected based on the detection call and the signal log ratio. For up-regulation, all genes in the *taf4*$^{-/-}$ samples called "Absent" were eliminated and the tables show those with a signal log 2 ratio above 1,0. For down-regulation, all genes in the *taf4*$^{lox/-}$ samples called "Absent" were eliminated and the tables show those with a signal log 2 ratio below -1,0.

To analyse gene expression in low serum conditions, duplicate samples of cells from the *taf4*$^{-/-}$ clones or the *taf4*$^{lox/-}$ clones were transferred in 0% or 1% serum respectively, for 3 days before preparation of RNA. Profiling and statistical analysis were performed as described above.

For RT-PCR, 1 μg of RNA was reverse transcribed with the SuperscriptII system (Invitrogen) using hexanucleotide primers in a volume of 20μl. 1 μl of a 1/50 dilution was used for PCR. Fragments between 150 and 400 bp were amplified. The sequences of the oligonucleotide primers for each gene are available upon request.

References ;

Bolstad, B.M., Irizarry, R.A., Astrand, M. and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185-193.

Hubbell, E., Liu, W.M. and Mei, R. (2002) Robust estimators for expression analysis. *Bioinformatics*, **18**, 1585-1592.

Liu, W.M., Mei, R., Di, X., Ryder, T.B., Hubbell, E., Dee, S., Webster, T.A., Harrington, C.A., Ho, M.H., Baid, J. and Smeekens, S.P. (2002) Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics*, **18**, 1593-1599.

Tsai, C.A., Hsueh, H.M. and Chen, J.J. (2003) Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics*, **59**, 1071-1081.

Legends to supplemental Figures.

Supplemental Figure 1. Differing morphologies of *taf4*$^{lox/-}$ and *taf4*$^{-/-}$ cells at low density. Phase contrast images (X40) of *taf4*$^{lox/-}$ and *taf4*$^{-/-}$cells grown in serum-rich media. The *taf4*$^{lox/-}$ cells exhibit a flatter and rounder morphology compared to the *taf4*$^{-/-}$ cells that are more elongated. This difference is also evident at higher densities (see text, and Fig. 2A).

Supplemental Figure 2. Rapid and dense growth of *taf4*$^{-/-}$ cells in low serum. Phase contrast images (X40) of *taf4*$^{lox/-}$ and *taf4*$^{-/-}$cells grown in 1% serum. The *taf4*$^{lox/-}$ cells exhibit a flattened morphology typical of quiescent cells. In contrast, the *taf4*$^{-/-}$ cells grow to high density and exhibit an elongated morphology. After prolonged growth periods, the *taf4*$^{-/-}$ cells form superposed mesh-like layers that are never observed with the *taf4*$^{lox/-}$ cells. Compared to figure 3C, the cells shown here were cultured for 10 days.

Supplemental Figure 3. Gene expression changes in *taf4*$^{-/-}$ cells. The results of the Affymetrix gene profiling are presented as Excel files. The genes whose expression is up or down-regulated under the indicated conditions are listed. Only the genes with a signal log 2 value greater than 1 for up-regulated genes, or less than –1 for down-regulated genes are listed. Page 1 lists genes up-regulated upon loss of TAF4 ; page 2, genes down-regulated upon loss of TAF4 ; page 3, genes up-regulated in *taf4*$^{-/-}$ cells in 0% serum ; page 4, genes down-regulated in *taf4*$^{-/-}$ cells in 0% serum ; pages 5 and 6, genes up or down-regulated in *taf4*$^{lox/-}$ cells in 1% serum respectively.

Supplemental Figure. 4. Functions of TAF4-regulated genes. The established lists of regulated genes were analysed using the Onto-express progamme at http://vortex.cs.wayne.edu/. The number of genes (probe sets) in each list are indicated along with their ascribed functions.

Supplemental Figure 5. Excel tables listing genes up-regulated by serum and/or by TAF4 inactivation. These lists were used to generate Fig. 8A. Groups A-C are as shown in Fig. 9A. The genes in each overlapping class are listed (for example A and B, page 1) as well as the genes that are specific for a given group (for example, B minus (A and B) and (B and C) defines the genes specific for group B alone)

Supplemental Figure 6. Excel tables listing genes down-regulated by serum and/or by loss of TAF4. These lists were used to generate Fig. 8B. Groups D-F are as shown in Fig. 8B.