Supplemental Material

Supplemental Material 1: [PDF file, Supplement 1 PCA results.pdf] Unsupervised clustering of 2902 most varying genes between the wild type (WT) and *GK*-overexpressing cell line (GK2). The two biological replicates of each cell line are shown separately. **(a)** Three-dimensional biplot. Open circles denote WT, filled circles denote the GK2. **(b)** Average linkage clustering dendrogram.

Supplemental Material 2: [Microsoft Excel file "Supplement 1 Raw microarray data.xls"]. Microarray data for lines WT and GK2, expressed as intensities of 31,099 Affymetrix probesets. CEL files for this data are uploaded to the National Institutes of Health Gene Expression Omnibus.

Supplemental Material 3: [Microsoft Excel file "Supplement 3a NCA connectivity matrix.xls" (Supplemental Material 3a) and PDF file "Supplement 3 NCA connectivity references.pdf" (Supplemental Material 3b)]. Supplemental material 3a contains the initial connectivity matrix linking potential transcription factors to genes differentially expressed in GK2 cells, as compared to line WT. Sheet 1 contains the connectivity matrix and expression levels of genes. An entry of '1' in the connectivity matrix indicates that the corresponding transcription factor is known to regulate the corresponding gene; an entry of '0' indicates no known regulation. All '1' entries are supported by literature references, which are indicated by two-letter codes alongside the entry. Supplemental material 3b contains a listing of these references.

Supplemental Material 4: [Microsoft Excel file "Supplement 4 NCA results.xls"]. Output of NCA. The block B3–J4 contains transcription factor activities (of arbitrary sign) deduced by

NCA for the two biological replicates of line GK2, and the block B10–J61 contains connectivity strengths (also deduced by NCA) between the transcription factors listed in B2–J2 and the genes listed in A10–A61. We multiplied the entire NCA-deduced column for each transcription factor (that contains the activity of and all connectivity strengths corresponding to the transcription factor) by an appropriate sign (either –1 or +1). This multiplication was necessary because the NCA decomposition is unique up to a scaling factor and the decomposition algorithm does not directly determine the signs of the transcription factor activities and connectivity strengths that it calculates (Liao et al., 2003). This sign has to be determined on the basis of experimentally determined connectivity directions (up- or down- regulation) between that transcription factor and one or more of its genes. As an illustration, in case of PPARα there is experimental evidence in Tamura et al. (2006) that PPARα negatively regulates the genes cadherin 17; procollagen, type III, alpha 1; and procollagen, type V, alpha 2 (Tamura et al., 2006). However, the initial NCA output (column G) shows positive signs for the connectivity strengths between PPARα and all these three genes; therefore, the entire PPARα column (transcription factor activities and connectivity strengths; column G) should be multiplied by –1 to be consistent with experimental connectivity evidence in Tamura et al. (2006). The sign-corrected column thus obtained (and reported in the paper) is column Q. This multiplication was performed for all transcription factors on the basis of experimental information available for the genes indicated with thick borders, the experimental connectivity directions being obtained from the references listed in Supplemental Material 3A/3B. The transcription factor activities (and their means and standard deviations) as well as connectivity strengths with correct signs are reported in the block L3–T61.