Supporting information: Figures and Tables

Figure S1. Repressed genes versus alternative use of 3' UTR. By screening on the probeset level, data can indicate that a gene is repressed, whereas alternative termination may be the real cause. For example, we initially found *Fscn1* repressed in bone marrow based on the expression data of probeset 1416514_a_at. However, when taking other probesets into account, it is clear that *Fscn1* transcript is present in bone marrow (even slightly higher than the average in other tissues). This indicates that in bone marrow, transcription of *Fscn1* terminates before the binding region of probeset 1416514_a_at (alternative termination site indicated by green vertical bar). In subsequent analysis, we took the information of redundant probesets into account in order to filter out this alternative termination effect. Binding region of the probesets is indicated on the gene structure scheme in a corresponding color.

Figure S2. Expression of 14 repressed genes in pancreatic islets of Langerhans, compared to average expression of these genes in all other tissues. Expression data are based on 69 MOE 430 2.0 arrays (dark) and 40 Gene Arrays 1.0ST (light). Bars represent expression (average±SEM) in islets (blue) versus all other tissues (red).

Figure S3. Standard errors of signals for all probe sets. Frequency of genes with an average standard error (calculated over tissues) in a certain range of values for the Novartis set versus our own data. These are used in the test statistics as an offset for the difference between tissue averages and give an impression of the technical variation

between replications corrected for sample size. Clearly, for most genes the standard error is lower in our data than in the Novartis data with the result of larger t-statistics and thus more significant results for our data.

Figure S4. Cell population purity impacts detection of repressed genes. Expression of *Mct1* was measured in whole islets, FACS purified beta cells and acini by microarray (A) and QPCR (B). QPCR expression levels are shown relative to pancreatic acini. The highest expression is observed in acini. Expression in pure beta cells is even lower than in islets since inherent to the isolation procedure islets are slightly (1-5% of cells) contaminated with acini. Two tailed t-tests were used to determine differences between expression in beta cells versus islets or acini. To demonstrate contamination with acini, expression of the exocrine marker *Ptf1a* is shown (C). While expression of this marker in purified beta cells is negligible, a signal at about 2% of expression in acini can be observed for the whole islet preparations.

Figure S5 A: Timecourse of *c-Maf* mRNA expression in pancreatic islets from neonatal rats (1-28 days old) measured by real-time Q-PCR; data were normalized for signal in islets of adult animals (10 weeks; dashed line). *C-Maf* expression displays a linear decreasing trend ($P<0.001$). Statistical significance of the difference between postnatal and adult expression signal was calculated by one-way ANOVA with Dunnett post-hoc test (***: $P<0.001$). **B**: Histone H3 lysine9-acetylation (green bars) and lysine27-trimethylation (red bars) of the promoter region of *c-Maf* in mouse liver and pancreatic islets. A high level of H3 lysine27-trimethylation is seen in islets but not in liver, whereas

H3 lysine9-acetylation is more extensive in the liver. Black bars represent IgG immunoprecipitation control. Data are means ± SEM of 4 QPCRs.

Figure S6. Histone modification during islet development compared to other tissues. Promoter regions of 3 genes deeply repressed in pancreatic islets (*c-Maf, Ldha* and *Mct1*) are shown with for several tissues or cell lines (indicated at the right) enrichment profiles of H3K4me3 (activation marker) and H3K9me3 (silencing marker) (van Arensbergen et al. 2010). H3K9me3 is largely absent in adult adipose tissue, liver and pancreatic progenitor cells and appears in islets, beta-cells and MIN6 cells.

Figure S7. miR-122 is a liver specific microRNA. A: Expression profile of miR-122 in different tissues shows its liver-specificity. **B:** miR-122 is the most abundant liver microRNA. Expression values are relative to the most abundant microRNA. Data obtained from microRNA.org.

Figure S8. MicroRNAs abundant in islets target islet-repressed genes. A: Expression of the most abundant microRNAs in islets, relative to miR-124, data obtained from microRNA.org. high miR124 is due to expression of this miRNA in the MIN6 cell line. **B:** microRNA abundance in adult islets versus MIN6 cells as measured by Exiqon miR microarrays. The abundance of miR-124 is compared to the expression of mir-375 which is islet-specific.

Figure S9. Disallowance of *Oxct1* in pig and zebrafish. GEO datasets GSE10898 (pig) and GSE11107 (zebrafish) were analysed and *Oxct1* expression in the different samples is shown. The pig dataset contains 16 tissue samples, each in 4 replicates. The zebrafish dataset contains only 2 tissues: brain and liver, each in control or starved condition, with for each sample type 4-5 replicates. *Oxct1* is very low in liver as compared to brain (t-test: $p<0.001$).

Table S1. Full list of tissue-specifically repressed genes. This table contains all 1074 genes which were found to be repressed in one of the 21 tissues studied.

Table S2. Leave-one-out analysis. We assessed the effect of leaving one tissue out of the (MOE 430 2.0) dataset on the extra number of genes found disallowed in the other tissues and generated a tissue association profile. In rows, the baseline tissues are listed and columns are the tissues left out of the analysis. Each cell of the matrix is the extra number of genes that are found disallowed on top of the number already found disallowed in the analysis with all tissues. Association, reflected by a large number in the matrix (both for a tissue as baseline or when the tissue is left out), is indicated between heart and muscle (red), spleen and bone marrow (orange) and testis vs other tissues (yellow).

Table S3. Disallowed genes in public data. We applied our methodology to the GNF Mouse GeneAtlas V3 CEL files (GEO GSE10246). The GNF set consists of 182 CEL

files which represent 91 tissues and cell lines in duplicate. This resulted in 280 genes that are found repressed in one of 28 tissues (31% of all tissues). Many samples are different parts of the same tissues or even time-series of the same cell type: 9 different brain regions, 25 samples from different immune-related cell types, 8 samples from eye, 5 from bone, 2 from adipose tissue, 2 from intestine and 2 from muscle cells.

Supporting information: Methods

Preparation of tissues and purified cells

Digestion of mouse pancreata was 3 min at 37°C after which islets and acini were separated at 0°C by three rounds of sedimentation (1 x g) and handpicking under a dissection microscope. For rat islet isolation , P1, P7, P15, P21, P28 and adult 10 week old rats were anesthesized and their islets isolated by collagenase digestion with rodent Liberase RI (Roche, Indianapolis IN). Islets of 10 animals were pooled as one sample for both P1 and P7; those from 2-3 animals were pooled for one sample for P15, P21 and P28.

Rat islets from all ages were handpicked under a stereomicroscope to ensure high purity of islet preparations.

Single cells from islets of RIPYY mice were obtained by dispersing the islets with trypsin, and a population containing > 99% pure β-cells was obtained by sorting the cells with a Beckton Dickinson FACSVantage SE (USA) (excit: 488 nm; emission: 530; 50 μm nozzle; 45 psi, 81 KHz).

RNA extraction

Total RNA was extracted using TRIzol Reagent according to the manufacturer's protocol (Invitrogen, Carlsbad, CA), followed by a cleanup procedure with RNeasy columns (Qiagen, Cologne, Germany). Total RNA from pituitary gland, adrenal gland and islets was extracted using the Absolutely RNA microprep from Stratagene (CA). RNA quantity and quality was determined using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DW) and the 2100 Bioanalyzer (Agilent, Waldbronn, Germany), respectively. Total RNA profiles of all tested samples were similar with sharp 18S and 28S rRNA peaks on a flat baseline. Total RNA of rat islets was isolated using the

picoRNA extraction kit from Arcturus or QIAGEN RNeasy Plus Mini Kit depending on the quantity of islets.

Intersection-union test

According to this test, a gene is declared specifically repressed (with type I error at level α) in a baseline tissue if it holds that the null hypothesis of equal or higher expression in the baseline tissue can be rejected at level α for *each* of the 20 baseline tissue – other tissue pairs. The test statistic we used is the t-test as calculated within an ANOVA setting (which implies using the information of all 21 tissues for the calculation of the mean squared error) and we calculated this statistic on the log 2 transformed expression levels obtained after RMA pre-processing. To account for the fact that we perform Berger's test 17344 times (once for each gene), we used Sidak's multiple comparison procedure. For an overall error rate (α_{EW}) of .05 the procedure sets the significance level for each probeset at $\alpha_{CW}=1-(1-(\alpha_{EW})^{1/17344}$, meaning that Berger's procedure was performed for each gene at a significance level of 2.96e-06. The intersection-union test can be adapted to test the hypothesis of repression in two or more specific tissues.

Antagomir treatment

The sequences of antagomir-122 (α-miR-122) and control antagomir (α-miR-122-mut) were: 5'-a*c*aaacaccauugucacacu*c*c*a*-Chol-3' and 5'-a*c*acacaacacugucacauu*c*c*a*-Chol-3', respectively. All the bases were 2' OMe modified; the star indicates phosphorothioate linkage; Chol represents 3'-TEG Cholesteryl.

Chromatin immunoprecipitation

ChIP was performed according to a modified protocol of Upstate (Millipore) EZ-ChIP. Tissue was cross-linked in 1% formaldehyde for 10 min at room temperature. The reaction was stopped by addition of glycine to a final concentration of 0.125 M and samples were washed once with cold PBS. Cell membranes were lysed by pestle homogenization and incubation for 15 min on ice in cell lysis buffer (5mM PIPES [pH 8.0], 85 mM KCl, 0.5% NP-40) supplemented with 1 mM PMSF and protease inhibitors (Complete protease inhibitor cocktail; Roche). Nuclei were isolated by centrifugation at 3000 x g for 5 min at 4°C and incubated on ice for 15 min in nuclei lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, [pH 8.1], 1mM PMSF, protease inhibitors). Isolated DNA was sheared by sonication for 20 cycles of 30 sec on/off at 200 W (Bioruptor; Diagenode) to an average length of 200-500 bp and centrifuged at 20 000 x g for 30 min at 4°C. DNA in supernatant was diluted 10-fold in dilution buffer. Chromatin was precleared for 1h at 4°C by addition of protein A-TSK (Affiland) beads blocked with salmon sperm DNA (1 mg/ml) and BSA (1 mg/ml). 1% of total input (10 µl of precleared chromatin) was set aside at 4°C until elution. Precleared chromatin (5 µg DNA per ml) was incubated overnight at 4°C using 2 µg of rabbit anti-acetyl-histone H3 (Lys9), rabbit anti-trimethyl-histone H3 (Lys27) or normal rabbit IgG (all antibodies purchased from Upstate). Antibody-chromatin complexes were collected by adding blocked protein A-TSK beads for 1 h at 4°C and washed sequentially in low-salt buffer, high-salt buffer, LiCl buffer and twice in TE buffer. Immune complexes were eluted 2 times for 15 min in 100 µl 1% SDS, 0.1M $NaHCO_3$ at room temperature and cross-links were reversed by overnight incubation at 65°C with 200 mM NaCl. Samples were treated with 10 U RNase A and 5 U RNase T1 (RNase A/T1 mix; Fermentas) for 30 min at 37°C and 0.6 U

proteinase K (Fermentas) for 1 h 30 min at 45°C. DNA was purified using spin columns (Geneclean Turbo; Qbiogene) and eluted in 60 µl DNase-free water. For assays involving islets samples (*Mct1*, *Ldha*), input was 5 µl of precleared chromatin, 650 ng DNA was incubated in 500 µl and 1 µg antibody was used and elution was in 30 µl DNase-free water.

Q-PCR primers

Rat islets

	Forward primer	Reverse primer
S25	GTGGTCCACACTACTCTCTGAGTTTC	GACTTTCCGGCATCCTTCTTC
Ins2	TCTTCTACACACCCATGTCCC	GGTGCAGCACTGATCCAC
Pc	TTGAAGGATGTGAAGGGCC	ACCTTTCGGATAGTGCCCTC
Ldha	AGACTTGGCCGAGAGCATAATG	ATAGAGACCCTTAATCATGGTGGAA
Mct1	AGTGCAACGACCAGTGAAGTGT	AGGACCTCCGGCATACATGA

Mouse islets

RNA Pol II	FW	GCACCACGTCCAATGATATTGTG
	Rev	GGAGATGACATGGTACAGTTCTCG
	Probe	(6-Fam)CTTCCGCACAGCCTCAATGCCCAGT(Tamra)
Mct1	FW	GCTTGGTGACCATTGTGGAATG
	Rev	CCCAGTACGTGTATTTGTAGTCTCC
	Probe	(6-Fam)CCCTGTCCTCCTAGGGCCACCACT(Tamra)

Mouse liver

	Forward primer	Reverse primer
Oxct1	GAGCCATGCAGGTTTCTAAGTATG	GCTCCTCCCATTCCTTTCAC
Hmgcs2	TGGTGGATGGGAAGCTGTCTA	TTCTTGCGGTAGGCTGCATAG
Actb	TCCTGAGCGCAAGTACTCTGT	CTGATCCACATCTGCTGGAAG
miR-122	ACACTCCAGCTGGGTAACACTGTCT GGTAA	CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGCTACCTGC
18S	GGCGCCCCCTCGATGCTCTTAG	GCTCGGGCCTGCTTTGAACACTCT

ChIP Primers/probes:

Gene	**Amplicon**	**Oligonucleotide sequence (5' to 3')**

	location	
Mct1	-324 to -233	TTTCTCCTCCAGAGCCTGA ACATCCTTATCAGCGCTCTGA (6-FAM)GGACGCCATCGTGGGCCTCA(Tamra)
Ldha	-109 to -39	CCAGCCTACACGTGGGTT TTAAATGGAAGCTCCGCGCT (6-FAM)TCCGCTGGGCTCCCACTCTGA(Tamra)
Actb	+29 to +134	CTTCTTTGCAGCTCCTTCGTTG CCCTGCAGTGAGGTACTAGC (6-FAM)CCACACCCGCCACCAGGTAAGCAG(Tamra)
Oxct1	-121 to -2	TCTCGCTTCTTGACAAGCCA AAGAGAGAAGGCAAGGCGA (6-FAM)ACCGCTCTGAGGTCCCGAGGAC(Tamra)
c-Maf	-133 to -7	GTG TGC ACG TTC GAG CTT TC CAG ATG GGC TGC AGG AGA (6-FAM)CCG CTG GCC ACC CAG CAC AG(Tamra)