**Sox5 regulates adult beta-cell differentiation and is reduced in type 2 diabetes**

Axelsson AS1,\*, Mahdi T1,\*, Nenonen H1,\*, Hänzelmann S2, Bagge A1, Wendt A1, Reinbothe TR1, Mecham B3, Millstein J3, Yang X3, Zhang B3, Tang Y1,4, Wang J1, Andersson LE1, Eliasson L1, Artner I1, Wollheim CB1,5, Derry J3, Friend SH3, Spégel P1, Mulder H1, Costa IG2, Zhang E1, Rosengren AH1,3

Inventory of Supplemental Information

1. Supplemental figures (page 2-5)
2. Supplemental Tables (page 6-20)
3. Supplemental Experimental Procedures (page xx-xx)
4. Supplemental References (page xx-xx)

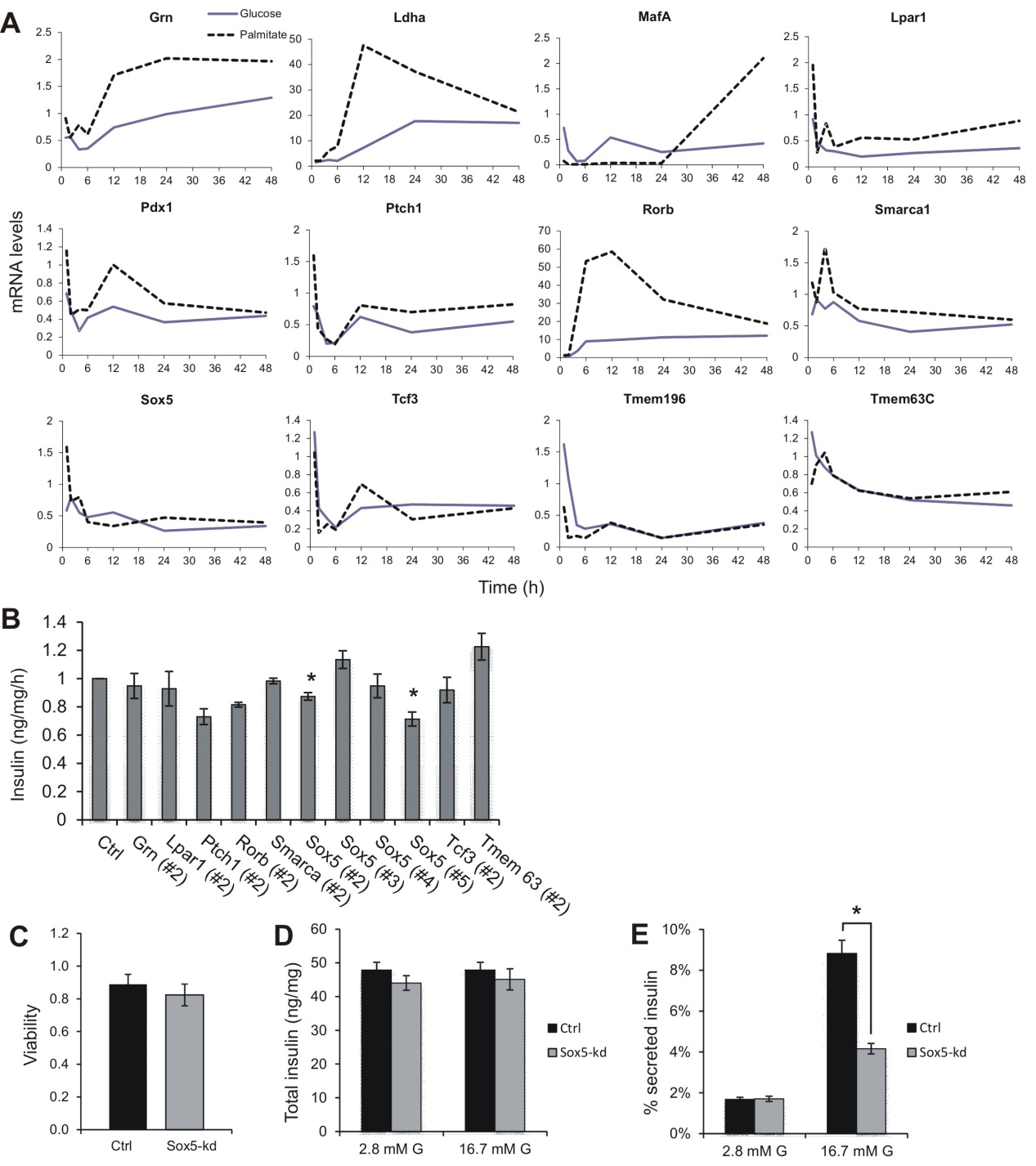


Figure S1, Related to Figure 2

(A) Graphs showing the expression of putative regulators of Sox5 expression in rat pancreatic islets after incubation at 20 mM glucose or 1 mM palmitate over 48 h. Y-axis values are fold-changes compared to expression at time 0. Dashed line: palmitate; solid line: glucose.

(B) Effect of siRNAs against transcription factors on insulin secretion. Bars show average fold-change in insulin secretion compared to control at 16.7 mM (n=3-7). All oligonucleotides were from Life Technologies except for Sox5 #2 and #3 (Sigma-Aldrich) and Sox5 #4 and #5 (Thermo Scientific).

Knock-down efficiency of siRNAs: Grn 75%, Lpar1 80%, Ptch1 73%, Rorb 82%, Smarca 83%, Sox5 #2 43%, Sox5 #3 35%, Sox5 #4 61%, Sox5 #5 72%, Tcf3 71%, Tmem63c 74%, Zbtb10 77%.

(C) Cell viability based on NADH and NADPH levels. Data are based on 5-6 replicates from one experiment.

(D) Total insulin content normalized to total protein content after secretion at 2.8 and 16.7 mM glucose. (n=6)

(E) Secreted insulin normalized to remaining intracellular insulin content at 2.8 and 16.7 mM glucose for the same samples as in (D). (n=6)

Data in bar graphs represent average with error bars showing +- SEM. \* p<0.05, \*\*p<0.01

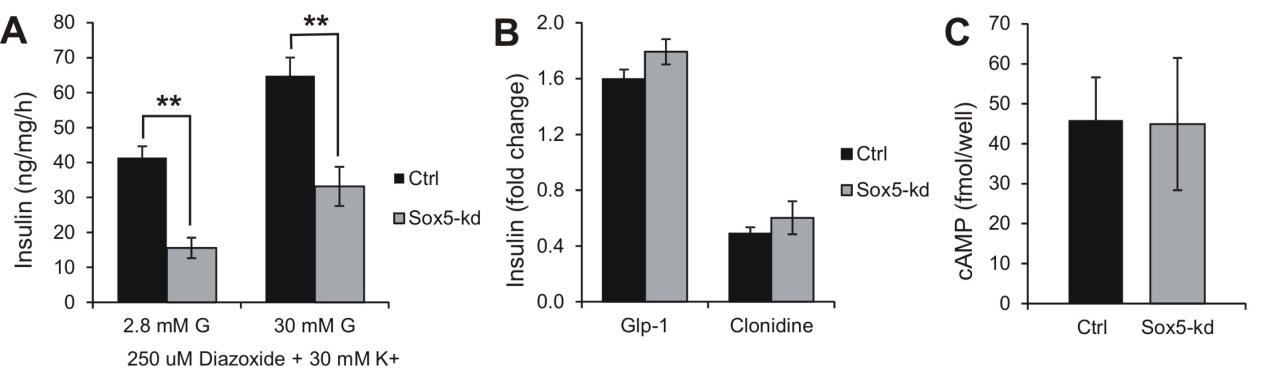


Figure S3, related to Figure 2

(A) Secreted insulin in response to 2.8 mM and 30 mM glucose in the presence of a combination of 250 uM diazoxide and 30 mM K+. (n=3)

(B) Fold change in insulin secretion with 16.7 mM glucose + Glp-1 or clonidine compared to 16.7 mM glucose alone. (n=3) p>0.05 in fold-change for all comparisons of Ctrl vs Sox5-kd.

(C) Intracellular cAMP levels after 1 h incubation at 16.7 mM glucose. Data is based on triplicates from one experiment and show mean +- SEM. p>0.05

Data in bar graphs represent average with error bars showing +- SEM. \* p<0.05, \*\*p<0.01

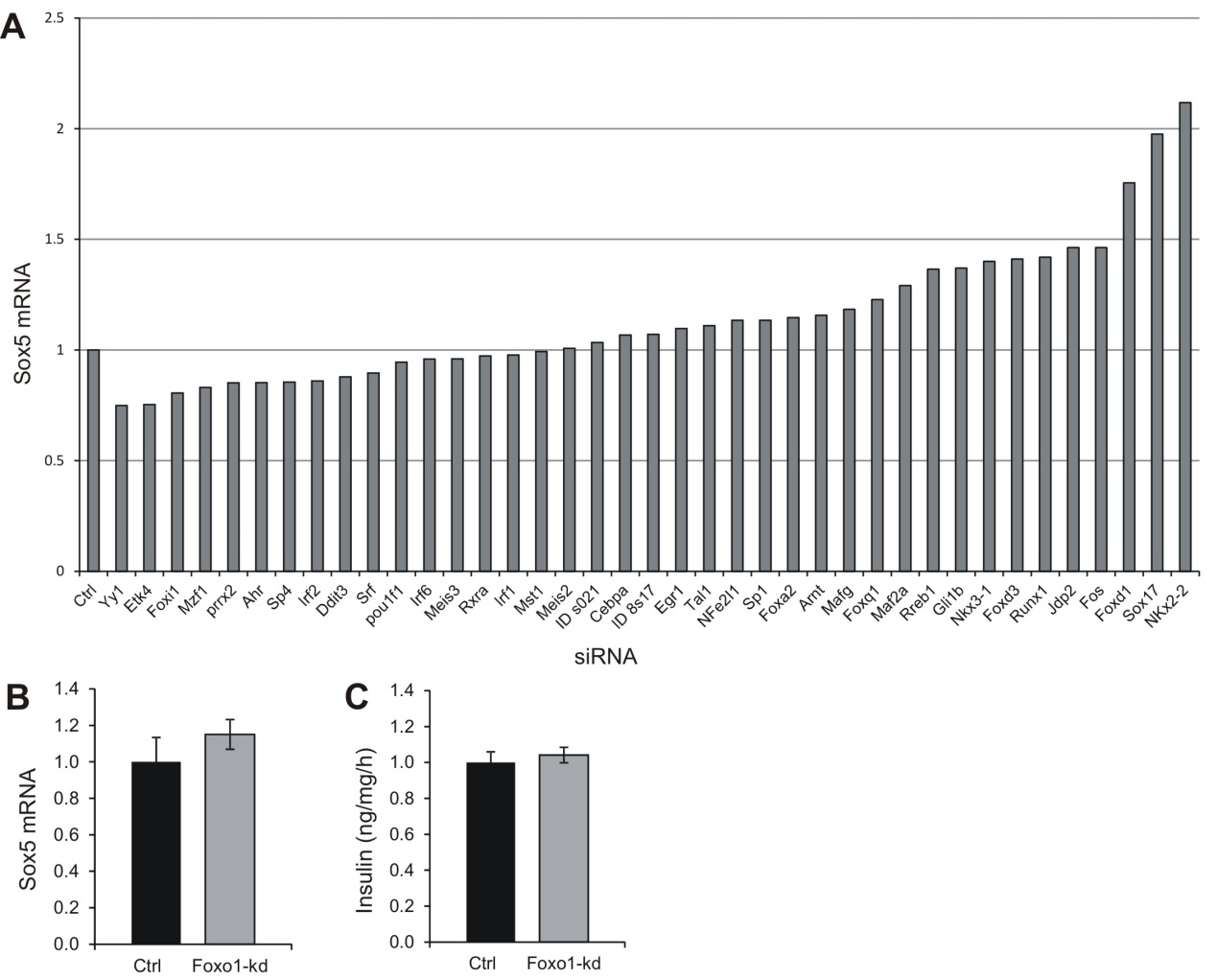


Figure S4, Related to Figure 6

(A) Sox5 mRNA levels after transfection with siRNA against 39 putative regulators of Sox5 expression. Data are from one experiment.

(B) Fold change effect of Fox01-kd on Sox5 mRNA levels (n=7). P>0.05

(C) Fold change effect of Fox01-kd on insulin secretion at 16.7 mM glucose. (n=7). P>0.05

Data in bar graphs represent average with error bars showing +- SEM. \* p<0.05, \*\*p<0.01

**Supplemental Tables**

Table S1Characteristics of human islet donors

|  |  |  |  |
| --- | --- | --- | --- |
|  | All (n=64) | Non-diabetic (n=45) | T2D (n=19) |
| Age (years) | 57 ± 10 | 56 ± 11 | 59 ± 9 |
| Sex (female / male) | 33 / 31 | 23 / 22 | 10 / 9 |
| BMI (kg/m2) | 26.2 ± 3.8 | 25.2 ± 2.9 | 28.7 ± 4.6 |
| Purity (%) | 66±18 | 69±17 | 61±20 |
| Number of donors from which HbA1c was available | 52 | 34 | 18 |
| Number of donors from which insulin secretion measurements were conducted | 48 | 32 | 16 |

Data are presented as mean ± s.d.

Table S2 Analysis of network properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | R2 | mean *k* | max *k* |
| 1 | 0.00904 | 4820 | 7990 |
| 2 | 0.351 | 1290 | 3160 |
| 3 | 0.63 | 448 | 1660 |
| 4 | 0.75 | 185 | 1020 |
| 5 | 0.811 | 86.9 | 685 |
| 6 | 0.834 | 45 | 488 |
| 7 | 0.851 | 25.2 | 363 |
| 8 | 0.876 | 15.1 | 279 |
| 9 | 0.868 | 9.5 | 220 |
| 10 | 0.878 | 6.26 | 177 |
| 11 | 0.887 | 4.28 | 144 |
| 12 | 0.902 | 3.03 | 119 |

Values of  from 1 to 12 were used to generate the adjacency matrix ai,j = |cor(xi,xj)|, where xi represents the expression vector for gene i across the 64 samples. The degree *k* denotes the connectivity of a gene. A linear relation between the logarithm of k and the logarithm of the degree distribution p(k) implies that the network has scale-free topology. R2 denotes the scale-free fit index. To obtain a network with approximately scale-free topology (R2>0.8) and yet high connectivity, =5 was used for subsequent analyses.

Table S3 Association of module eigengenes with phenotype traits

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | T2D (n=64) | | HbA1c (n=52) | | GSIS (n=48) | | KSIS (n=26) | |
| Module | No of genes | P value | R2 | P value | R2 | P value | R2 | P value | R2 |
| 1 | 4328 | 0.576 | 0.0398 | 0.771 | -0.0492 | 0.215 | -0.167 | 0.325 | -0.121 |
| **2** | **3148** | **0.0133** | **-0.316** | **0.0025** | **-0.39** | **0.00585** | **0.369** | **0.0483** | **0.409** |
| 3 | 1474 | 0.534 | -0.0324 | 0.939 | -0.0352 | 0.727 | -0.126 | 0.267 | -0.307 |
| 4 | 1223 | 0.519 | 0.0602 | 0.926 | 0.0426 | 0.641 | 0.0385 | 0.943 | -0.128 |
| 5 | 1116 | 0.516 | 0.0787 | 0.16 | 0.245 | 0.336 | -0.138 | 0.79 | -0.0188 |
| 6 | 815 | 0.0108 | 0.295 | 0.0314 | 0.315 | 0.101 | -0.24 | 0.316 | -0.402 |
| 7 | 806 | 0.391 | 0.0713 | 0.573 | -0.0364 | 0.122 | -0.198 | 0.247 | -0.221 |
| 8 | 714 | 0.163 | 0.216 | 0.0873 | 0.209 | 0.461 | -0.198 | 0.32 | -0.092 |
| 9 | 702 | 0.0142 | -0.297 | 0.00968 | -0.334 | 0.0087 | 0.33 | 0.105 | 0.384 |
| 10 | 649 | 0.399 | -0.11 | 0.352 | -0.151 | 0.201 | 0.209 | 0.24 | 0.178 |
| 11 | 598 | 0.0343 | -0.245 | 0.00881 | -0.337 | 0.00306 | 0.356 | 0.0892 | 0.382 |
| 12 | 424 | 0.037 | -0.262 | 0.0261 | -0.246 | 0.0175 | 0.332 | 0.0583 | 0.381 |
| 13 | 380 | 0.556 | 0.145 | 0.535 | 0.166 | 0.652 | -0.1 | 0.641 | -0.143 |
| 14 | 358 | 0.000551 | 0.393 | 0.00409 | 0.377 | 0.0328 | -0.253 | 0.216 | -0.425 |
| 15 | 347 | 0.18 | -0.0879 | 0.333 | -0.0634 | 0.606 | 0.0872 | 0.989 | -0.117 |
| 16 | 209 | 0.484 | 0.038 | 0.16 | 0.0998 | 0.687 | -0.0138 | 0.924 | 0.0332 |
| 17 | 203 | 0.0701 | -0.23 | 0.244 | -0.0988 | 0.0632 | 0.215 | 0.104 | 0.398 |
| 18 | 198 | 0.674 | 0.123 | 0.69 | 0.0689 | 0.691 | -0.0316 | 0.831 | -0.0537 |
| 19 | 196 | 0.00641 | 0.316 | 0.002 | 0.408 | 0.00612 | -0.346 | 0.187 | -0.335 |
| 20 | 196 | 0.431 | 0.138 | 0.887 | 0.0872 | 0.273 | 0.187 | 0.0862 | 0.184 |
| 21 | 188 | 0.527 | -0.0879 | 0.6 | -0.0971 | 0.104 | 0.211 | 0.343 | 0.205 |
| 22 | 159 | 0.117 | -0.186 | 0.483 | -0.0186 | 0.297 | 0.155 | 0.348 | 0.275 |
| 23 | 152 | 0.173 | 0.188 | 0.127 | 0.287 | 0.967 | -0.0724 | 0.0732 | 0.0612 |
| 24 | 151 | 0.128 | -0.179 | 0.233 | -0.181 | 0.402 | 0.174 | 0.62 | 0.0496 |
| 25 | 143 | 0.773 | 0.0935 | 0.273 | 0.193 | 0.345 | -0.133 | 0.24 | -0.279 |
| 26 | 140 | 0.0238 | 0.23 | 0.0598 | 0.21 | 0.151 | -0.118 | 0.33 | -0.266 |
| 27 | 125 | 0.0564 | -0.208 | 0.417 | -0.121 | 0.478 | 0.0504 | 0.463 | 0.233 |
| 28 | 120 | 0.0428 | -0.255 | 0.312 | -0.154 | 0.359 | 0.0807 | 0.679 | 0.0872 |
| 29 | 115 | 0.892 | 0.025 | 0.963 | 0.0315 | 0.0748 | 0.196 | 0.186 | 0.233 |
| 30 | 112 | 0.701 | 0.000926 | 0.499 | -0.0786 | 0.438 | 0.189 | 0.0765 | 0.212 |
| 31 | 95 | 0.0396 | 0.268 | 0.125 | 0.264 | 0.817 | -0.1 | 0.965 | -0.0434 |
| 32 | 70 | 0.746 | 0.0713 | 0.261 | 0.234 | 0.375 | -0.0986 | 0.616 | -0.0674 |
| 33 | 65 | 0.305 | 0.0935 | 0.0114 | 0.21 | 0.414 | -0.0505 | 0.825 | 0.00376 |
| 34 | 64 | 0.571 | 0.105 | 0.195 | 0.212 | 0.872 | -0.06 | 0.972 | -0.0708 |
| 35 | 64 | 0.0125 | -0.305 | 0.0238 | -0.299 | 0.0106 | 0.29 | 0.356 | 0.33 |
| 36 | 64 | 0.312 | -0.138 | 0.587 | -0.0809 | 0.444 | 0.0752 | 0.378 | 0.29 |
| 37 | 58 | 0.38 | -0.162 | 0.646 | -0.0482 | 0.403 | 0.133 | 0.388 | 0.272 |
| 38 | 51 | 0.151 | -0.132 | 0.0373 | -0.179 | 0.00425 | 0.312 | 0.0885 | 0.329 |
| 39 | 50 | 0.434 | -0.0787 | 0.553 | -0.203 | 0.335 | -0.171 | 0.112 | -0.343 |
| 40 | 50 | 0.558 | 0.108 | 0.537 | 0.181 | 0.318 | 0.151 | 0.218 | 0.293 |
| 41 | 44 | 0.686 | -0.00833 | 0.746 | 0.0112 | 0.216 | 0.161 | 0.946 | -0.0126 |
| 42 | 43 | 0.102 | 0.219 | 0.655 | 0.137 | 0.174 | 0.0242 | 0.962 | 0.0838 |
| 43 | 40 | 0.298 | -0.116 | 0.466 | -0.186 | 0.197 | -0.242 | 0.077 | -0.441 |
| 44 | 38 | 0.42 | 0.0676 | 0.891 | 0.0367 | 0.751 | 0.0462 | 0.473 | 0.134 |
| 45 | 35 | 0.589 | 0.0435 | 0.862 | -0.0198 | 0.908 | -0.0547 | 0.948 | 0.125 |
| 46 | 32 | 0.521 | -0.0398 | 0.106 | -0.0865 | 0.7 | 0.0135 | 0.993 | -0.169 |
| 47 | 31 | 0.182 | 0.156 | 0.411 | 0.232 | 0.965 | 0.034 | 0.148 | 0.236 |
| 48 | 30 | 0.204 | 0.0953 | 0.152 | 0.0757 | 0.645 | -0.136 | 0.948 | 0.0222 |
| 49 | 26 | 0.527 | -0.0268 | 0.724 | 0.144 | 0.0301 | 0.245 | 0.426 | 0.286 |
| 50 | 24 | 0.801 | 0.012 | 0.802 | -0.0649 | 0.0563 | -0.244 | 0.911 | -0.0338 |
| 51 | 24 | 0.714 | 0.012 | 0.659 | -0.0222 | 0.901 | 0.041 | 0.924 | -0.117 |
| 52 | 21 | 0.54 | -0.0916 | 0.118 | -0.359 | 0.992 | -0.0454 | 0.0556 | -0.348 |
| 53 | 20 | 0.0833 | -0.181 | 0.00933 | -0.301 | 0.0188 | 0.298 | 0.174 | 0.195 |
| 54 | 17 | 0.549 | 0.0639 | 0.482 | 0.167 | 0.161 | 0.0936 | 0.918 | -0.068 |
| 55 | 17 | 0.817 | 0.00648 | 0.552 | 0.199 | 0.076 | 0.208 | 0.243 | 0.203 |
| 56 | 16 | 0.376 | 0.131 | 0.644 | 0.157 | 0.188 | 0.184 | 0.282 | 0.241 |

The module eigengene denotes the 1st principal component of the expression matrix of the module genes across the 48 samples. For each module eigengene the association with T2D status, HbA1c, glucose-stimulated insulin secretion (GSIS) and K+-stimulated insulin secretion (KSIS). The T2D-associated module is indicated in boldface. P values shown in the table are not corrected for multiple testing.

Table S4 Module genes with islet-selective open chromatin

|  |  |  |
| --- | --- | --- |
| Gene symbol | Fold-change in T2D | Connectivity (*k*in) |
| ABCC9 | -1.13 | 225 |
| ABHD10 | -1.19 | 345 |
| ACVR1C | -1.28 | 293 |
| ADAMTS2 | -1.13 | 128 |
| AGBL4 | -1.10 | 92 |
| ANKS1B | -1.04 | 137 |
| ATRNL1 | -1.27 | 330 |
| BARX2 | 1.12 | 150 |
| BMP5 | -1.24 | 172 |
| C1orf158 | 1.01 | 15 |
| C1orf168 | -1.01 | 90 |
| C3orf14 | -1.11 | 222 |
| C9orf150 | 1.03 | 77 |
| CACNA1H | -1.11 | 92 |
| CACNA2D1 | -1.23 | 364 |
| CACNB2 | -1.14 | 278 |
| CADM1 | -1.14 | 380 |
| CADPS | -1.21 | 316 |
| CERKL | -1.13 | 403 |
| CHODL | -1.10 | 61 |
| CNTN1 | -1.19 | 498 |
| CNTN4 | -1.33 | 245 |
| CPE | -1.16 | 474 |
| D4S234E | -1.26 | 314 |
| DACH1 | -1.16 | 459 |
| DCX | -1.13 | 200 |
| DGKB | -1.05 | 51 |
| DIRAS2 | -1.15 | 118 |
| DSCAM | -1.01 | 45 |
| DZIP3 | -1.21 | 364 |
| ELAVL4 | -1.29 | 420 |
| ELP4 | -1.20 | 213 |
| EML5 | -1.17 | 155 |
| ENAM | -1.11 | 174 |
| FAM123C | -1.06 | 85 |
| FAM135A | -1.15 | 50 |
| FAM135B | -1.06 | 247 |
| FAM148A | -1.14 | 20 |
| FAM19A4 | -1.17 | 81 |
| FAM84A | -1.03 | 27 |
| FBXW7 | -1.07 | 105 |
| FGF14 | -1.19 | 475 |
| FLJ25770 | -1.17 | 148 |
| FOXC1 | 1.03 | 69 |
| FOXP2 | -1.06 | 180 |
| G6PC2 | -1.31 | 263 |
| GABRA1 | -1.19 | 19 |
| GABRB3 | -1.13 | 431 |
| GABRG2 | -1.25 | 209 |
| GAD2 | -1.19 | 283 |
| GDAP1 | -1.17 | 338 |
| GLP1R | -1.35 | 262 |
| GNAI1 | -1.10 | 394 |
| GPM6A | -1.30 | 99 |
| GPR27 | -1.05 | 76 |
| GRIA2 | -1.24 | 288 |
| GRIA4 | -1.30 | 152 |
| GRIK2 | -1.10 | 286 |
| HMGCLL1 | -1.22 | 377 |
| HS6ST3 | -1.18 | 330 |
| IAPP | -1.38 | 68 |
| IGSF11 | -1.14 | 246 |
| ISL1 | -1.19 | 223 |
| JAKMIP2 | -1.14 | 382 |
| KBTBD7 | -1.19 | 241 |
| KCNB2 | -1.21 | 327 |
| KCNK16 | -1.23 | 226 |
| KCNMA1 | -1.18 | 415 |
| KIAA0774 | -1.19 | 249 |
| KIAA1257 | -1.04 | 83 |
| KIAA1383 | -1.03 | 37 |
| KIAA1486 | -1.07 | 56 |
| KIAA1804 | 1.08 | 102 |
| KL | -1.13 | 360 |
| KLHL1 | -1.18 | 43 |
| LIG4 | -1.08 | 140 |
| LIMCH1 | -1.11 | 255 |
| LRRTM3 | -1.33 | 249 |
| LSAMP | 1.01 | 62 |
| MAGI2 | -1.16 | 335 |
| MAP9 | -1.04 | 171 |
| MAPK10 | -1.15 | 367 |
| MCTP2 | -1.11 | 41 |
| MEIS2 | -1.22 | 309 |
| MEIS3 | -1.02 | 28 |
| MMP16 | -1.02 | 57 |
| MS4A8B | -1.17 | 89 |
| MYO3A | -1.16 | 269 |
| NALCN | -1.19 | 443 |
| NBEA | -1.19 | 272 |
| NECAB2 | -1.01 | 89 |
| NEGR1 | 1.02 | 156 |
| NEUROD1 | -1.22 | 382 |
| NKX6-1 | -1.27 | 283 |
| NMNAT3 | -1.05 | 118 |
| NOL4 | -1.24 | 482 |
| NR3C2 | -1.11 | 94 |
| NRXN1 | -1.15 | 473 |
| ODZ3 | -1.08 | 256 |
| OXGR1 | -1.10 | 161 |
| PAM | -1.19 | 411 |
| PAPPA2 | -1.10 | 88 |
| PAX6 | -1.20 | 337 |
| PCDH17 | -1.13 | 349 |
| PCDH20 | -1.12 | 74 |
| PCDHAC1 | -1.06 | 269 |
| PCLO | -1.22 | 300 |
| PCP4 | -1.21 | 81 |
| PDK4 | -1.04 | 131 |
| PDX1 | -1.21 | 129 |
| PDZRN3 | -1.09 | 295 |
| PGR | -1.12 | 355 |
| PRKACB | -1.14 | 345 |
| PRUNE2 | -1.08 | 43 |
| PVRL3 | -1.21 | 277 |
| RAB3C | -1.17 | 311 |
| RALYL | -1.08 | 70 |
| RASEF | 1.27 | 373 |
| RGS17 | -1.16 | 108 |
| RGS4 | -1.10 | 202 |
| RIMBP2 | -1.23 | 360 |
| RIMS2 | -1.18 | 342 |
| RNF150 | -1.08 | 245 |
| RNF180 | -1.12 | 250 |
| ROBO2 | -1.35 | 300 |
| ROR1 | -1.07 | 111 |
| RORB | -1.12 | 222 |
| RTN1 | -1.15 | 426 |
| SCG2 | -1.17 | 372 |
| SCGB2A1 | -1.07 | 42 |
| SDK1 | -1.02 | 143 |
| SERPINI2 | 1.08 | 72 |
| SGCB | -1.08 | 216 |
| SGCZ | -1.03 | 58 |
| SIM1 | -1.16 | 279 |
| SLC17A6 | -1.26 | 176 |
| SLC18A2 | -1.21 | 61 |
| SLC30A8 | -1.20 | 259 |
| SLC35F4 | -1.12 | 90 |
| SLC5A1 | -1.07 | 62 |
| SLC5A8 | -1.03 | 39 |
| SLC7A2 | -1.08 | 195 |
| SLCO1A2 | -1.22 | 80 |
| SNAP91 | -1.24 | 455 |
| SOX6 | -1.14 | 125 |
| SPATA7 | -1.12 | 100 |
| SPTB | -1.10 | 260 |
| ST18 | -1.27 | 345 |
| ST6GALNAC5 | -1.23 | 286 |
| ST8SIA3 | -1.12 | 395 |
| STXBP5L | -1.25 | 380 |
| SULT4A1 | -1.14 | 131 |
| SYT14 | -1.27 | 431 |
| SYT4 | -1.24 | 459 |
| TMEM132B | -1.19 | 58 |
| TMEM132D | -1.09 | 242 |
| TMEM196 | -1.21 | 516 |
| TMEM61 | -1.07 | 161 |
| TMEM63C | -1.24 | 490 |
| TRPM3 | -1.24 | 338 |
| TSC22D1 | -1.05 | 106 |
| TSHZ3 | -1.10 | 84 |
| TSPYL5 | -1.11 | 276 |
| TTC8 | -1.16 | 450 |
| UNC5D | -1.17 | 47 |
| ZNF223 | -1.09 | 80 |
| ZNF585B | -1.10 | 217 |
| ZSCAN18 | -1.06 | 99 |

Fold-change in T2D denotes expression in T2D donors vs. non-diabetic donors.

Table S5Characteristics of human islet donors in replication set

|  |  |  |  |
| --- | --- | --- | --- |
|  | All (n=59) | Non-diabetic (n=37) | T2D (n=22) |
| Age (years) | 61 ± 10 | 64 ± 7 | 60 ± 12 |
| Sex (female / male) | 23 / 36 | 17 / 20 | 6 / 16 |
| BMI (kg/m2) | 26.8 ± 3.8 | 26.4 ± 3.7 | 27.4 ± 3.9 |
| Purity (%) | 72 ± 18 | 72 ± 17 | 72 ± 21 |
| Number of donors from which HbA1c was available | 57 | 35 | 22 |

Data are presented as mean ± s.d. No insulin secretion data were available from the replication set.

Table S6 Analysis of network properties in replication set

|  |  |  |  |
| --- | --- | --- | --- |
|  | R2 | mean *k* | max *k* |
| 1 | 0.089 | 2030 | 3233 |
| 2 | 0.192 | 653 | 1510 |
| 3 | 0.459 | 266 | 838 |
| 4 | 0.601 | 126 | 516 |
| 5 | 0.682 | 66 | 340 |
| 6 | 0.730 | 38 | 235 |
| 7 | 0.756 | 23 | 169 |
| 8 | 0.800 | 14 | 126 |
| 9 | 0.803 | 9 | 96 |
| 10 | 0.767 | 6 | 74 |
| 11 | 0.817 | 4 | 58 |
| 12 | 0.834 | 3 | 47 |

Values of  from 1 to 12 were used to generate the adjacency matrix ai,j = |cor(xi,xj)|, where xi represents the expression vector for gene i across the 59 samples in the replication set. The degree *k* denotes the connectivity of a gene. A linear relation between the logarithm of k and the logarithm of the degree distribution p(k) implies that the network has scale-free topology. R2 denotes the scale-free fit index. To obtain a network with approximately scale-free topology (R2>0.8) and yet high connectivity, =8 was used for subsequent analyses.

Table S7 Module genes with islet-selective open chromatin

|  |  |
| --- | --- |
| Gene Symbol | Fold-change in T2D |
| ABCC9 | -1.13 |
| ABHD10 | -1.10 |
| ACVR1C | -1.28 |
| AGBL4 | -1.08 |
| ANKS1B | -1.06 |
| ATRNL1 | -1.10 |
| BARX2 | 1.04 |
| BMP5 | -1.01 |
| C1orf158 | -1.08 |
| C1orf168 | 1.00 |
| C3orf14 | -1.13 |
| CACNA1H | -1.02 |
| CACNA2D1 | -1.09 |
| CACNB2 | -1.03 |
| CADM1 | -1.03 |
| CADPS | -1.15 |
| CERKL | -1.08 |
| CHODL | -1.12 |
| CNTN1 | -1.13 |
| CNTN4 | -1.27 |
| CPE | -1.05 |
| DACH1 | -1.06 |
| DCX | -1.13 |
| DGKB | 1.06 |
| DIRAS2 | -1.14 |
| DSCAM | -1.04 |
| DZIP3 | -1.11 |
| ELAVL4 | -1.07 |
| ELP4 | -1.10 |
| EML5 | -1.07 |
| ENAM | -1.10 |
| FAM123C | -1.01 |
| FAM135A | -1.10 |
| FAM135B | 1.04 |
| FAM19A4 | -1.11 |
| FAM84A | -1.16 |
| FBXW7 | -1.07 |
| FGF14 | -1.06 |
| FOXC1 | 1.08 |
| FOXP2 | -1.06 |
| G6PC2 | -1.16 |
| GABRA1 | -1.31 |
| GABRB3 | -1.14 |
| GABRG2 | -1.38 |
| GAD2 | -1.04 |
| GDAP1 | -1.13 |
| GLP1R | -1.15 |
| GNAI1 | -1.01 |
| GPM6A | -1.11 |
| GRIA2 | -1.11 |
| GRIA4 | -1.05 |
| GRIK2 | -1.21 |
| HMGCLL1 | -1.12 |
| HS6ST3 | -1.15 |
| IAPP | -1.13 |
| IGSF11 | -1.05 |
| ISL1 | -1.24 |
| JAKMIP2 | 1.03 |
| KBTBD7 | -1.22 |
| KCNB2 | -1.15 |
| KCNK16 | -1.11 |
| KCNMA1 | -1.11 |
| KIAA1383 | -1.01 |
| KL | -1.16 |
| KLHL1 | -1.30 |
| LIG4 | -1.05 |
| LIMCH1 | -1.03 |
| LRRTM3 | -1.20 |
| LSAMP | 1.08 |
| MAGI2 | -1.14 |
| MAP9 | -1.04 |
| MAPK10 | -1.07 |
| MCTP2 | -1.09 |
| MEIS2 | -1.12 |
| MEIS3 | -1.02 |
| MMP16 | -1.21 |
| MS4A8B | -1.22 |
| MYO3A | -1.03 |
| NALCN | -1.05 |
| NBEA | -1.11 |
| NECAB2 | 1.04 |
| NEGR1 | 1.00 |
| NEUROD1 | -1.16 |
| NKX6-1 | -1.05 |
| NMNAT3 | -1.14 |
| NOL4 | -1.16 |
| NR3C2 | -1.00 |
| NRXN1 | -1.10 |
| ODZ3 | -1.07 |
| OXGR1 | -1.03 |
| PAM | -1.06 |
| PAPPA2 | 1.07 |
| PAX6 | -1.13 |
| PCDH17 | -1.07 |
| PCDH20 | -1.35 |
| PCDHAC1 | -1.02 |
| PCLO | -1.06 |
| PCP4 | -1.16 |
| PDX1 | -1.22 |
| PDZRN3 | -1.07 |
| PGR | -1.10 |
| PRKACB | -1.10 |
| PRUNE2 | -1.06 |
| PVRL3 | -1.05 |
| RAB3C | -1.22 |
| RALYL | -1.18 |
| RIMBP2 | -1.16 |
| RIMS2 | -1.10 |
| RNF150 | -1.05 |
| RNF180 | -1.02 |
| ROBO2 | -1.16 |
| ROR1 | 1.01 |
| RORB | -1.12 |
| RTN1 | -1.11 |
| SCG2 | -1.04 |
| SCGB2A1 | 1.05 |
| SDK1 | -1.04 |
| SGCB | -1.07 |
| SGCZ | -1.08 |
| SIM1 | -1.12 |
| SLC17A6 | 1.01 |
| SLC18A2 | 1.13 |
| SLC30A8 | -1.14 |
| SLC35F4 | -1.21 |
| SLC5A1 | -1.08 |
| SLC7A2 | -1.13 |
| SLCO1A2 | 1.01 |
| SOX6 | -1.11 |
| SPATA7 | -1.10 |
| SPTB | 1.01 |
| ST18 | -1.19 |
| ST6GALNAC5 | -1.17 |
| ST8SIA3 | -1.02 |
| STXBP5L | -1.16 |
| SULT4A1 | -1.00 |
| SYT14 | -1.15 |
| SYT4 | -1.11 |
| TMEM132B | -1.13 |
| TMEM132D | -1.11 |
| TMEM196 | -1.10 |
| TMEM61 | -1.07 |
| TMEM63C | -1.11 |
| TRPM3 | -1.14 |
| TSC22D1 | -1.05 |
| TSHZ3 | -1.04 |
| TSPYL5 | -1.14 |
| TTC8 | -1.07 |
| UNC5D | -1.15 |
| ZNF585B | -1.06 |
| ZSCAN18 | -1.06 |

Table S8 SNPs associated with module eigengenes

|  |  |  |  |
| --- | --- | --- | --- |
| SNP id | Nearest gene | Associated trait | P value |
| rs4406574 | SMARCA1 | Eigengene T2D module | 1.4200e-04 |
| rs5975111 | SMARCA1 | Eigengene T2D module | 1.9858e-04 |
| rs5932625 | SMARCA1 | Eigengene T2D module | 1.4200e-04 |
| rs11796659 | SMARCA1 | Eigengene T2D module | 1.9858e-04 |
| rs5932624 | SMARCA1 | Eigengene T2D module | 2.1855e-04 |
| rs5977087 | SMARCA1 | Eigengene T2D module | 1.9858e-04 |
| rs1408301 | SMARCA1 | Eigengene T2D module | 1.9858e-04 |
| rs10505917 | SOX5 | Eigengene T2D module | 1.2848e-03 |
| rs7314161 | SOX5 | Eigengene T2D module | 1.1902e-03 |
| rs11047351 | SOX5 | Eigengene 168 genes | 1.6742e-03 |
| rs17498948 | SOX5 | Eigengene 168 genes | 1.6742e-03 |
| rs4569092 | SOX5 | Eigengene 168 genes | 1.1506e-03 |
| rs4963749 | SOX5 | Eigengene 168 genes | 1.6742e-03 |

The islet donors (n=64) were genotyped for single-nucleotide polymorphisms (SNPs) and associations between donor genotype and the eigengene of the T2D-related module and the eigengene representing the 168 open chromatin genes, respectively, was analyzed by linear regression using age, sex and BMI as covariates. The nearest gene and the p values from the linear regression are given. Displayed p values are not corrected for multiple testing.

**Experimental procedures**

**Human islets**

Islets were extracted from cadaveric multi-organ donors after research consent as previously described (3,4). Briefly, the pancreas was perfused with ice-cold collagenase, cut into pieces and placed in a digestion chamber at 37°C. Separation of endocrine and exocrine tissues was achieved by a continuous density gradient. Selected fractions were then centrifuged to enrich for islets. Purity of islets was measured by dithizone staining (3). From this suspension, islets to be used for experiments were hand-picked under a microscope. The islets were cultured at 5.6 mM glucose in CMRL-1066 (INC Biomedicals, Costa Mesa, USA) supplemented with 10 mM HEPES, 2 mM L-glutamine, 50 µg/ml gentamicin, 0.25 µg/ml Fungizone (GIBCO, BRL, Gaitersburg, MD, USA), 20 µg/ml ciprofloxacin (Bayer Healthcare, Leverkusen, Germany), 10 mM nicotinamide and 10% serum at 37°C (95% O2 and 5% CO2) for 1 to 9 days prior to experiments.

**Microarray analysis.** Total RNA was extracted from human islets with the AllPrep DNA/RNA Mini Kit (Qiagen, Germany). RNA quality and concentration were measured using an Agilent 2100 Bioanalyzer (Bio-Rad, Hercules, CA, USA) and a Nanodrop ND-1000 (NanoDrop Technologies, USA). The Affymetrix GeneChip® Human Gene 1.0 ST microarray chip was used for gene expression analysis in accordance with the standard protocol. Briefly, total RNA was converted into biotin-targeted cRNA following the manufacturer’s specifications, and the biotin-labeled cRNA was fragmented into strands with 35 to 200 nucleotides. This was hybridized onto the chip overnight in a GeneChip® Hybridization 6400 oven using standard procedures. The arrays were washed and stained in a GeneChip® Fluidics Station 450. Scanning was carried out with the GeneChip® Scanner 3000 and image analysis was done with the GeneChip® Operating Software. Data normalization was performed using Robust Multi-array Analysis (RMA).

**Coexpression network analysis.** The co-expression network analysis was performed in R (version 2.15.1) using log2-transformed microarray expression data. Using the weighted gene co-expression network analysis (WGCNA) framework (Zhang and Horvath, 2005) and the corresponding Bioconductor package (Langfelder and Horvath, 2008), we first calculated the pair-wise co-expression for all genes and formed a similarity matrix based on the Pearson correlation coefficients si,j = |cor(xi,xj)|, where xi denotes the expression vector for gene i across the samples.

Next, the similarity matrix was transformed into an adjacency matrix ai,j = |cor(xi,xj)|. The adjacency matrix represents the strength of connection between two genes as a continuous weight in [0,1] and has been shown to be robust with respect to the choice of  as compared with methods using hard thresholds to dichotomize the coexpression matrix (Zhang and Horvath, 2005). The connectivity of a gene in a network (the degree *k*) equals the sum of all connections for that gene.

Biological networks have been suggested to exhibit a scale-free property (Ravasz et al., 2002), which means that the probability that a node is connected with k other nodes (the degree distribution p(k)) decays as a power function p(k) ~ k-. Linear regression analysis of log-transformed k and p(k) was used to estimate how well the co-expression network satisfied the scale-free topology for different values of . We found that for ≥7 R2 for the fit was >0.8 both in the initial analysis and the replication.

Based on the adjacency matrix for =7, the topological overlap, which reflects the relative gene interconnectedness, was calculated for all gene pairs (Ravasz et al., 2002). The non-negative and symmetric topological overlap matrix  = [i,j] was converted to dissimilarity (distance) measures di,j = 1-i,j, which were used for module identification.

The eigengene, defined as the 1st principal component of the gene expression matrix, was determined for each module and was correlated with the phenotype traits. The human islet insulin secretion data showed non-Gaussian distributions. Gaussian distribution was obtained using logarithm transformation, and analyses of human islet insulin secretion data were therefore performed using log-transformed data.

For each gene in the T2D-associated module the connectivity within the module (*k*in) was determined. In addition, the correlation between the gene expression trait and T2D status, was calculated. Next, the correlation between gene connectivity and the trait associations was analyzed across the genes by Spearman’s rank correlation.

Gene enrichment analysis was performed using DAVID version 6.7 (Huang da et al., 2009).

Islet-selective open chromatin genes we identified as having open chromatin in the transcription starting site or gene body (Supplementary Table 2 in Gaulton et al.).

**Comparisons with public microarray data.** Differential gene expression between experimental conditions (e.g. control and gene knockout) was determined for >10,000 public microarray data sets deposited at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo). The overlap between differentially expressed genes in each data set and the 168 open chromatin genes perturbed in human islets from T2D donors was determined and the p value for the enrichment was calculated using Fisher’s exact test.

**In silico analysis of transcription factor binding sites (TFBS).** TFBSwere detected using the Regulatory Genomics Toolbox (regen.googlecode.com).A region 1 kb upstream of the promoter (Ensembl built GRCh37.p13) of genes were used to find binding sites. Next, we performed motif match analysis using the motif matching tool available in Biopython (Cock et al., 2009) with a false positive rate (FPR) of 0.0001 in these regions (Wilczynski et al., 2009). Motifs were obtained in Jaspar and Uniprobe databases (Bryne et al., 2008; Newburger and Bulyk, 2009). The same procedure was repeated 100 times on random genomic regions with the same length as the original regions. We employed a one-tailed Fisher exact test to measure if the proportion of binding sites of a motif inside the gene promoters is higher than the proportion of binding sites in random regions. Final *p*-values were corrected by the Benjamini-Hochberg method (Hochberg and Benjamini, 2009). For detection of candidate regulators of Sox5, we performed motif match in open chromatin regions close to Sox5 (i.e. that overlapped with the gene or promoter region of 1 kb upstream), which are only found in differentiated islet cells.

**Analysis of eSNPs.** The analysis of single nucleotide polymorphisms associated with gene expression traits (eSNPs) the expression data was performed using linear models analyzed by in-house R and Perl code (at Sage Bionetworks) as previously described (Schadt ref) using 10 permutations of the expression file. The permutation was done by swapping column preserve the coexpression structure of the data.

**Transmission electron microscopy**. Human islets were fixated in 2.5 % Glutaraldehyde (GA) added to freshly prepared Millionig buffer and refrigerated for 2h. Following wash in Millionig buffer, the islets were post-fixated in osmium tetroxide (1%) for 1h, and then washed carefully with Millionig solution. Thereafter, the islets were dehydrated and embedded in AGAR 100 (Oxford Instruments Nordiska AB, Sweden) before being cut into 70-90 nm ultrathin sections. The sections were placed on Cu-grids and contrasted with uranyl acetate and lead citrate before examined in a JEM 1230 electron microscope (JEOL-USA. Inc., USA). The number of granules/cell section area divided by granule diameter (Nv) and the number of docked granules/cell section perimeter divided by granule diameter (Ns) were estimated as reported previously (Vikman et al. 2009). Granules were defined as docked when the centre of the granule was located within 150 nm from the plasma membrane. Electron micrographs were analyzed from at least 3 different cells per donor (median 7 cells).

**Rat islets.** Pancreatic islets from rats were prepared by collagenase digestion of the exocrine pancreas. The islets were hand-picked and incubated in a humidified atmosphere at 5 mM glucose in RPMI 1640 tissue culture medium (SVA, Sweden) supplemented with 10% (vol/vol) fetal calf serum, 100 IU/ml penicillin and 100 µg/ml streptomycin.

**INS1-832/13 cells**. Rat insulinoma INS1-832/13 cells (passages 55-70) were cultured at 10 mM glucose in RPMI 1640 (Life Technologies) and supplemented with 10% (vol/vol) fetal calf serum, 100 IU/ml penicillin, 100 μg/ml streptomycin, 10 mM HEPES, 2 mM glutamine, 1 mM sodium pyruvate, and 50 μM β-mercaptoethanol.

**Human -cell/PANC-1 fusion cell lines.** The human 1.1B4 and 1.4E7 -cell/PANC-1 fusion cell lines (ECACC, UK) were cultured in RPMI-1640 (Life Technologies) with 10 mM glucose supplemented with 10% (v/v) fetal calf serum, 100 IU/ml penicillin and 100 µg/ml streptomycin (passages 35-45).

**Insulin secretion and insulin content measurements.** Approximately 350000 - 400000 INS1-832/13 cells were seeded per well in a 24-well plate 72 h before secretion experiments. Cells were washed twice with a secretion assay buffer (SAB) containing 114 mM NaCl, 4.7 mM KCl, 1.2 mM KH2PO4, 1.16 mM MgSO4, 20 mM HEPES pH 7.2, 25.5 mM NaHCO3, 2.5 mM CaCl2, 0.2% BSA with 2.8 mM glucose. This was followed by 2 h preincubation at 37 °C with 5% CO2 in 2 ml SAB. Next, the buffer was removed and cells were incubated at 37 °C for 1 h in SAB supplemented with glucose, tolbutamide (Sigma-Aldrich, Sweden), clonidine, KCl or Glp-1 as indicated. Immediately after incubation, an aliquot of the medium was removed and insulin was analysed directly or after storage at -20 °C for later analysis with the Coat-a-Count® kit (Siemens, Sweden) according to the manufacturer’s protocol. For protein and total insulin analysis the remaining cells were lysed with RIPA buffer (50mM Tris HCl pH 8, 150 mM NaCl, 1% NP-40/Triton X, 0.1% SDS, 0.5% sodium deoxycholate, 2 mM EDTA and 50 mM NaF). Cells were shaken on ice for 30 min followed by collection of the lysate and centrifugation at 10000 g for 5 min (4°C). The supernatant was collected and analysed for total protein and insulin content directly or stored at -20°C for later analysis. Total protein was measured with the Pierce BCA Protein Assay Kit (Thermo Scientific). Total insulin was assessed at a 1:100 dilution in PBS with the Coat-a-Count® kit .

Average values based on technical replicates (2-6 wells?) from three or more experiments were compared using paired Student’s *t*-test. For experiments correlating insulin secretion to mRNA expression, ctrl siRNA and Sox5-kd secretion values were first normalized to values for non-treated cells.

**Incubation of INS1-832/13 cells with palmitate and glucose.** INS1-832/13 cells were incubated for 48 h at 20 mM glucose or with 0.25, 0.5 and 1 mM palmitate. Palmitate-BSA 5 mM stock solution was prepared as described by Cousin et al. (ref). Control cells were cultured at 10 mM glucose as described above. Cells with addition of 10 mM mannitol were used as osmotic control for the 20 mM glucose sample. After 48 h, RNA was isolated and used for quantification of mRNA levels.

**Incubation of INS1-832/13 cells with valproic acid.** A 250 mM stock of valproic acid (VPA) was prepared fresh at the time of use in distilled water and filter sterilized. VPA at 0.1, 0.3 or 1 mM was added to INS1-832/13 cells at the time of transfection. Medium was changed to fresh VPA-containing medium the day after transfection, and insulin secretion or RNA isolation performed 48 h post-transfection as described in the section for insulin secretion and insulin content measurements and quantitative PCR.

**RNA interference.** The day before siRNA transfection 350000 - 400000 INS1-832/13 cells were seeded per well in a 24-well plate. Cells were transfected with RNA interference oligonucleotides using DharmaFECT 1 (Thermo Scientific, USA) according to the manufacturer’s description. We used Silencer Select® siRNAs from Life Technologies. For Sox5, two additional siRNAs from Sigma and two siRNAs from Thermo Scientific were tested. Final oligonucleotide concentration was 30 nM (45 nM for initial experiments?). The negative control siRNA used was from the same manufacturer as the active siRNA; Silencer Select® Negative Control no. 2 (Life Technologies), Mission® siRNA universal negative control # 1 (Sigma) and On-Target plus non-targeting siRNA # 1 (Thermo Scientific, USA), respectively. Assays were performed 48 h after transfection. For immunohistochemistry cells were re-seeded after 24 h onto a 0.175 mm thick glass and incubated for another 24 h. Transfection efficiency was assessed by qPCR as described below. For capacitance measurements, cells were re-seeded in new plates after 48 h and used for experiments 72 h post-transfection.

**Capacitance measurements.** The electrophysiological measurements were conducted using an EPC-10 patch clamp amplifier with the PULSE software (HEKA, Germany). The plastic Petri dishes were used as the experimental chamber with a plastic insert to reduce the volume to approximately 0.5 ml. The dish was continuously perfused at a rate of ~2 ml/min at 31-33°C. Patch pipettes were pulled from borosilicate glass, coated with Sylgard and fire-polished to an average resistance of 4-6 MΩwhen filled with pipette solution. The zero-current potential of the pipette was adjusted with the pipette in the bath. Exocytosis was elicited by a train of ten depolarizations from -70 to 0 mV, which was applied to simulate glucose-induced electrical activity. Exocytosis was monitored as increases in cell capacitance using the sine + DC mode of the lock-in amplifier included in the PULSE software and the standard whole-cell configuration. The extracellular solution consisted of (mM) 118 NaCl, 20 TEACl, 5.6 KCl, 2.6 CaCl2, 1.2 MgCl2, 5.0 HEPES, and 5.0 glucose (pH 7.4 with NaOH). The pipette solution was composed of (mM) 125 Cs-glutamate, 10 CsCl, 10 NaCl, 1 MgCl2, 5 HEPES, 3 ATP, and 0.1 cAMP and  EGTA, (pH 7.2 with CsOH). All reagents were of analytical grade. Isradipine was added to the extracellular solution at 2 M. The current-voltage relationship was measured by 100 ms depolarizations from the holding potential at -70 mV. The depolarization potential was increased stepwise from -40 to +40 mV.

**Ca2+ Imaging (Enming please check).** Fluo-5F (Kd = 2.3 mM) (Invitrogen, USA) was used for measuring intracellular Ca2+. Cells were loaded using 1 mM Fluo-5F in room temperature for 30 min. Firstly cells were perfused in Krebs-Ringer bicarbonate (KRB) buffer containing 116 mM NaCl, 4.7 mM KCl, 2.6 mM CaCl2, 1.2 mM KH2PO4, 1.2 mM MgSO4, 20 mM NaHCO3, 16 mM HEPES, 2 mg/ml BSA, supplemented with 5 mM glucose. Stimulation was carried out with a 20 mM glucose KRB buffer at room temperature.

Images were acquired by confocal microscopy using a 406water immersion objective. A ratio was calculated by taking the fluorescence intensity in time lapse divided by the average fluorescence intensity under pre-stimulatory conditions. The time

integral of the fluorescence signal (Area Under the Curve, A.U.C) was calculated using the formula: Σ*P*ii~0

0:5:Dt:<eth>RizRiz1?, where Dt,

the time interval; Ri , the ratio at time point i.

**Immunostaining (Enming, Thomas and Isabella please check).** Cells were washed twice in PBS and fixed with 3% PFA in PBS for 30 min at 37 °C followed by washing 3 x 10 min with PBS before permeabilisation for 30 min with Phosphflow II (BD Scientific). Samples were blocked in 5% normal donkey serum in PBS for 30 min at RT. Cells were incubated overnight at 4 °C with primary antibodies against Cav1.2 (1:200, Sigma-Aldrich) and Cav1.3 (1:100; XX) diluted in blocking solution. Immunoreactivity was detected using fluorescently labeled secondary antibodies (*specification* 1:200) and visualized by confocal microscopy (Carl Zeiss, Germany). Co-localization analysis was performed by using a ZEN2009 software based on Pearson’s correlation coefficient analysis which recognizes the colocalized pair by comparison pixel by pixel intensity [ref?31]. The internalization was indicated by a ratio that is defined by the mean intensity of plasma membrane to the mean intensity in cytosol, according to the formula:

Ratio =

((i1\_a1{i2\_a2)

(a1{a2) )

(

(i2\_a2{i3\_a3

(a2{a3) ).

Where i1, i2 and i3 represent the intensities of whole cell, cytosol

and nucleus, a1, a2 and a3 represent the area of whole cell, cytosol

and nucleus respectively.

The immunostaining for MafA, MafB and Sox5 followed a slightly different protocol. Cells were washed twice with PBS prior to fixation in 4 % PFA for 30 min at RT.

After washing 2 x 10 min with PBS, cells were permeabilised with 0.1% TritonX 100 in PBS and thereafter blocked with 5% NDS in PBS. Primary antibodies targeting MafA …. (1:100), MafB …. (1:250) and Sox5 (1:100) (Sigma-Aldrich) were diluted in blocking solution and cells were incubated for 2 h at RT. Samples were washed and fluorescently labeled secondary antibodies diluted in blocking buffer (1:400) were added and incubated at RT for1 h. Cells were then washed 3 times with PBS and post-fixed in 4% PFA for 10 min. Samples were washed with water, mounted and visualized by confocal microscopy (Carl Zeiss, Germany).

Isabella’s methodology

**Quantitative PCR.** Islets were homogenized in Qiazol reagent (Qiagen, USA) followed by vortexing. RNA was extracted with chloroform precipitation using the mRNeasy kit (Qiagen). Reverse transcription was performed using either SuperScript III Reverse transcriptase (Life Technologies) or SuperScript VILO cDNA synthesis kit (Life Technologies). Quantitative PCR was performed on a ViiA 7 Real-Time PCR System (Life Technologies) using TaqMan Gene Expression Assays and TaqMan Universal PCR Master Mix (Life Technologies). Gene expression was measured from triplicate median Cq-values normalized to the geometric mean of Hprt, B2m and Polr2a Cq-values (relative expression).

Total RNA was extracted from INS832/13 using the RNeasy® Plus Mini kit (Qiagen, Germany) and gene expression was assessed as described for islets above.

For experiments with low gene expression (?) preamplification was performed for 10 or 14 cycles prior to quantitative PCR using TaqMan PreAmp Master Mix (Life Technologies) and TaqMan Gene Expression Assays at a dilution of 0.05X.

In experiments where data were available from both non-treated and negative control samples, the relative gene expression in treated samples was normalized to the relative expression in the non-treated sample. Average value and SEM from all experiments were then normalized to the average of the negative controls. Statistical comparisons were made using paired Student’s *t*-test.

**Sox5 overexpression.** INS1-832/13 cells were co-transfected with a plasmid expressing rat Sox5 (BlueHeron, custom made) under the CMV promoter and a plasmid encoding GFP using the Lipofectamine LTX Plus reagent (Life Technologies) according to the manufacturer’s manual. Total plasmid concentration was 0.625 µg/ml and the concentration of Lipofectamine LTX was 4 µl/µg plasmid. The ratio of Sox5 plasmid to GFP plasmid was 1:1 for single-cell capacitance measurements to enable identification of transfected cells. For microarray analysis the Sox5 plasmid was co-transfected with an empty vector instead of the GFP plasmid at a ratio of 1:1000.

The human 1.1B4 and 1.4E7 cells were co-transfected with a human SOX5 plasmid (Origene) and a GFP reporter plasmid (Origene) to enable visualization of transfected cells for single-cell capacitance recordings. Total plasmid concentration was 1.25 µg/ml and the concentration of Lipofectamine LTX 3 µl/ug plasmid.

Control cells were transfected with empty vector together with GFP.

For capacitance measurements, cells were re-seeded in new plates after 48 h and used for experiments 72 h post-transfection. RNA for microarray analysis was extracted 48 h post-transfection.

**cAMP measurements.** Intracellular levels of cyclic AMP (cAMP) were measured with the cAMP Biotrak Enzyme Immunoassay (EIA) System (GE Healthcare Life Sciences, UK) using the non-acetylation protocol as described by the manufacturer. INS1-832/13 cells (300000 cells/well) were seeded in a 24-well plate, transfected with Sox5 siRNA the following day and lysed with 1.5 ml Lysis reagent 1B 48 h post-transfection.

**Proinsulin-insulin ratio.** The ratio of human proinsulin to human insulin in the secretion buffer of INS1-832/13 cells was determined using Mercodia Proinsulin ELISA and Mercodia ELISA (Mercodia, Sweden) according to the manufacturer's instructions. Insulin secretion in Sox5-kd cells was performed 48 h post-transfection as described in the previous **section for insulin secretion and insulin content measurements,** and an aliquot of the secretion buffer was used for proinsulin analysis.

**Ins 832/13 cell viability.** Cell viability was measured using a CellTiter 96 Aqueous One Solution Cell Proliferation Assay Reagent (Promega, Stockholm, Sweden) according to the manufacturer’s instructions. The measurement is based on the spectrophotometric detection of a colored formazan product converted from a (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) (MTS) compound by NADPH or NADH in metabolically active cells. After a culture period of 24 h at 10 mM glucose cells were washed three times with fresh culture medium. Thereafter, the cells were incubated for 2 h in CellTiter 96 Aqueous One Solution reagent before measuring the absorbance at 490 nm with a 96-well plate reader.

**Metabolomics (Peter S please check).** Metabolite profiling in INS1-832/13 was performed as previously described in detail (1,2). In brief, cells were treated as described for insulin secretion, followed by a quick wash in ice-cold PBS and quenching of metabolism by addition of 70 µl ice cold double distilled water. Metabolites were extracted using a one phase extraction, as previously described in detail (1). Metabolites were derivatised and analysed on a gas chromatograph (GC; Agilent 6890N, Agilent Technologies, Atlanta, GA) connected to a time-of-flight mass spectrometer (TOF-MS; Leco Pegasus III TOFMS, Leco Corp., St. Joseph, MI). Data were acquired using Leco ChromaTof (Leco Corp.), exported as NetCDF files, and processed using hierarchal multivariate curve resolution (HMCR) (3) in MATLAB 7.0 (Mathworks, Natick, MA). Metabolites were normalized using the scores from the first component of a principal component analysis (PCA) calculated in Simca P+ 12.0 (Umetrics, Umeå, Sweden) on the uncentered and unit variance scaled areas of internal standards (4).

1) Spégel, P., Malmgren, S., Sharoyko, V. V., Newsholme, P., Koeck, T., and Mulder, H. (2011) Metabolomic analyses reveal profound differences in glycolytic and tricarboxylic acid cycle metabolism in glucose-responsive and -unresponsive clonal beta-cell lines. Biochem. J. 435, 277-284

2) Spégel, P., Sharoyko, V. V., Goehring, I., Danielsson, A. P. H., Malmgren, S., Nagorny, C. L. F., Andersson, L. E., Koeck, T., Sharp, G. W. G., Straub, S. G., Wollheim, C. B., and Mulder, H. (2013) Time resolved metabolomics analysis of beta-cells implicates the pentose phosphate pathway in the control of insulin release. Biochem J 450, 595-605

3) Jonsson, P., Sjövik Johansson, E., Wuolikainen, A., Lindberg, J., Shuppe-Koistinen, I., Kusano, M., Sjöström, M., Trygg, J., Moritz, T., and Antti, H. (2006) Predictive metabolite profiling applying hierchal multivariate curve resolution to GC-MS data - A potential tool for multi-parametric diagnosis. J. Proteome Res. 5, 1407-1414

4) Chorell, E., Moritz, T., Btranth, S., Antti, H., and Svensson, M. B. (2009) J. Proteome Res. 8, 2966-2977

**Mitochondria spare respiratory capacity measurements (Seahorse, Annika B please check details).** 70 000 INS1-832/13 cells were seeded onto L-lysine coated Seahorse microplates 72 h prior to analysis on the Seahorse XF24 extracellular flux analyzer following the manufacturer’s instructions. Before starting the experiments cells were pre-incubated in medium containing 2.8 mM glucose for 2 h. The experiments were performed at 37°C and the media contained X mM glucose and 10? mM pyruvate.

Oxygen consumption rate data consist of mean rates during each measurement cycle consisting of a mixing time of 30s and a wait time of 3 min followed by a data acquisition period of 3 min. Rates displayed are basal respiration (0–3’) and rates following addition of 2 µg/ml oligomycin (3–6’), 6 µm FCCP (6–9’), rotenone (1µM) and myxothiazol (2 µM).

**Incubation of INS1-832/13 cells with palmitate and glucose.** INS1-832/13 cells were incubated for 48 h at 20 mM glucose or with 0.25, 0.5 and 1 mM palmitate. Palmitate-BSA 5 mM stock solution was prepared as described by Cousin et al. (ref). Control cells were cultured at 10 mM glucose as described above. Cells with addition of 10 mM mannitol were used as osmotic control for the 20 mM glucose sample. After 48 h, RNA was isolated and used for quantification of mRNA levels.

**Incubation of INS1-832/13 cells with valproic acid.** A 250 mM stock of valproic acid (VPA) was prepared fresh at the time of use in distilled water and filter sterilized. VPA at 0.1, 0.3 or 1 mM was added to INS1-832/13 cells at the time of transfection. Medium was changed to fresh VPA-containing medium the day after transfection, and insulin secretion or RNA isolation performed 48 h post-transfection as described above.

**Statistical analyses.** Student’s t-test was used for statistical comparisons unless otherwise specified. Fisher’s exact test with Benjamini-Hochberg multiple testing correction was used for enrichment analyses. Linear regression was used for comparisons between SOX5 expression in the microarray and continuous phenotypes and logistic regression was used to analyze the association between SOX5 expression and T2D status. Details on the statistical procedure used to analyze data from specific methods are given in the relevant sections. Statistical analyses were performed using R, python 2.6 (additional packages: scipy 0.7.0, fisher 0.1.0 and statsmodels 0.4.0), IBM SPSS Statistics (ver 20.0) or Excel.