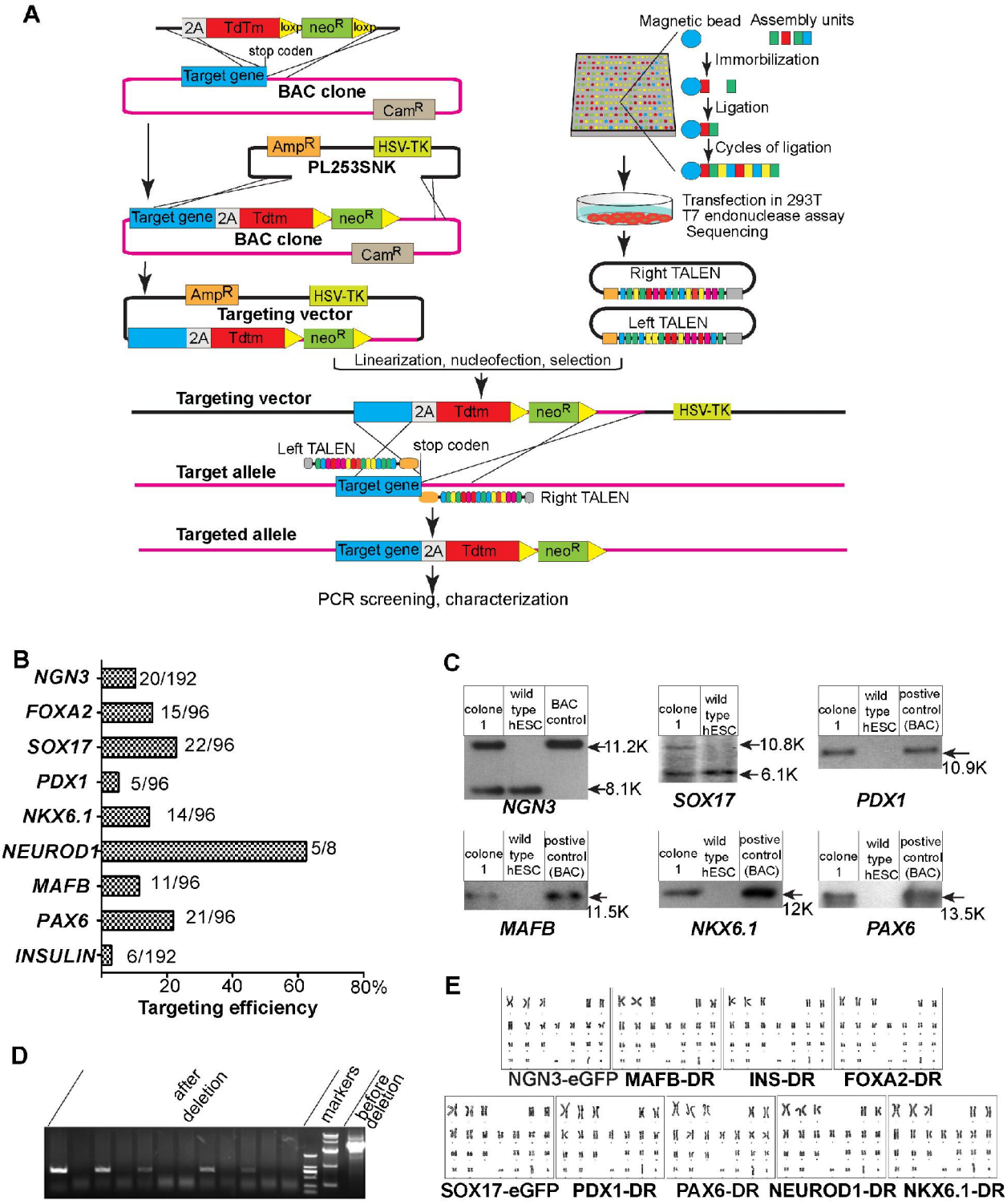
A
2A
TdTm
loxp
neoR
loxp
stop coden
Target gene
BAC clone
CamR
AmpR
HSV-TK
PL253SNK
Target gene
2A
Tdtm
neoR
BAC clone
CamR
AmpR
HSV-TK
Targeting vector
2A
Tdtm
neoR
Magnetic bead
Assembly units
Immorbilization
Ligation
Cycles of ligation
Transfection in 293T
T7 endonuclease assay
Sequencing
Right TALEN
Left TALEN
Linearization, nucleofection, selection
Targeting vector
2A
Tdtm
neoR
HSV-TK
Left TALEN
stop coden
Target allele
Target gene
Right TALEN
Targeted allele
Target gene
2A
Tdtm
neoR
PCR screening, characterization
B
NGN3
20/192
FOXA2
15/96
SOX17
22/96
PDX1
5/96
NKX6.1
14/96
NEUROD1
5/8
MAFB
11/96
PAX6
21/96
INSULIN
6/192
20
40
60
80%
Targeting efficiency
C
colone 1
wild type hESC
BAC control
11.2K
8.1K
NGN3
colone 1
wild type hESC
10.8K
6.1K
SOX17
colone 1
wild type hESC
postive control (BAC)
10.9K
PDX1
colone 1
wild type hESC
postive control (BAC)
11.5K
MAFB
colone 1
wild type hESC
postive control (BAC)
12K
NKX6.1
colone 1
wild type hESC
postive control (BAC)
13.5K
PAX6
D
after deletion
markers
before deletion
E
NGN3-eGFP
MAFB-DR
INS-DR
FOXA2-DR
SOX17-eGFP
PDX1-DR
PAX6-DR
NEUROD1-DR
NKX6.1-DR

Supplementary information, Figure S1. Targeting of *eGFP* or *Tdtm* into a series of pancreatic gene loci in hESCs. (A) Schematic representation of the hESC targeting strategy with the aid of the high-throughput TALEN synthesis technique. Targeting vectors with flanking homology arms were generated based on two rounds of BAC-based recombination (A, upper left). The DNA sequences of TALENs specific to selected gene loci were synthesized by a high-throughput integrated chip method, which allowed DNA fragments encoding different TALE repeats to be assembled on magnetic beads in batch fashion (A, upper right). Gene targeting was performed by nucleofecting the linearized targeting vector and corresponding TALENs into hESCs; this was followed by drug selection, PCR screening and the further characterization of the drug-resistant colonies. The stop codons of the target genes were replaced with a *2A-Tdtm* (or *eGFP*)-*loxp-CAG-neo-loxp* cassette (A, bottom). (B) PCR screening of drug-resistant colonies showed efficient gene targeting for selected pancreatic gene loci. The fractions indicated positive colony numbers (numerators) versus tested colony numbers (denominators). Only the *NGN3* locus was targeted without the aid of TALENs. (C) Southern blotting analyses of randomly selected reporter cell lines confirmed the correct integration of the reporter genes. To analyze the genomic DNA of *SOX17-eGFP* cell lines and the *NGN3-eGFP* cell line (with neo), probes spanning the homology arms were used. A single band was detected in wild-type hESCs, but two bands were observed in the targeted cell lines. To analyze the genomic DNA of the other reporter cell lines, a probe recognizing the Tdtm sequence was used. A single fragment was observed in the targeted cell lines but not in wild type hESCs. (D) PCR analysis showed the successful deletion of the *CAG-neo* cassette in the *NGN3-eGFP* reporter cell line. The PCR products changed from 3.6 kb to 0.7 kb after deletion. On the left side of the markers, the extension time of the PCR reaction was 45 seconds; only the clones with a successful deletion could generate a 0.7-kb band, while the failed clones generated no band. On the right side, there is a control 3.6-kb band, which was amplified with an extension time of 3.5 minutes. (E) The normal karyotypes of wild-type H1 and reporter hESC lines were shown by G-band analysis. Abbreviations: TALE (transcription activator-like effector); TALEN (transcription activator-like effector nuclease); *INS* (*INSULIN*); *Tdtm* (*TdTomato*).