A schematic illustration of UMI use in CRISPR–Cas9 screens.

(a) Data analysis in CRISPR screens is conventionally based on several sgRNAs targeting the same gene. Introduction of random barcodes at complexities well above total analyzed cell number will tag each individual cell with a unique molecular identifier (UMI). This generates a third layer of information at single cell level, generating biological replicas. (b) Schematic illustration of the generation of single cell derived clones using a limiting dilution followed by clonal expansion.
Supplementary Figure 2

The bioinformatic pipeline of sgRNA prediction.

Doench-scores as measure of predicted sgRNA activity were calculated for all exonic sgRNAs compatible with our cloning strategy. Doench scores were penalized based on a ruleset for biological effects. Those rules combine evaluation of exon length, prediction of protein domains, alternative splicing and ATG start codons, Pol-III terminator sequences, position of the sgRNA within the CDS. The penalties also spread selected sgRNAs over different exons and include off-target prediction penalties.
Supplementary Figure 3
A scheme illustrating the generation of CRISPR-UMI library complexity.

The CRISPR-UMI library is generated by 2 subsequent complex cloning steps. Initially, a random barcode consisting of 10 nucleotides is integrated into the vector backbone. Subsequently, the sgRNA pool of 26,514 sgRNAs is ligated to the barcode library with over 1000 ligation events per sgRNA. Thereby, each of the over 1000 ligation events per sgRNA combines the sgRNA with another random barcode. The combination of sgRNA and random barcodes generates a complexity of >1000 times the number of sgRNAs. We refer to this highly complex combination of sgRNA and barcode as UMI (unique molecular identifier). Our library reached a complexity of 83 million.
Supplementary Figure 4

CRISPR-UMI sgRNA library and cassette amplification.

(a) Vector design for library generation. Upon pooled parallel cloning of a barcode of 10 random nucleotides into retroviral backbones at complexities of 10^6, chip-synthesized sgRNA pools (at a complexity of 2.6514) were cloned into UMI containing backbone at a coverage >1000 clones/guide. Cassette-flanking PacI sites allowed for liberation of small sgRNA containing fragments from mammalian genomic DNA; (b) Library subpools and cloning complexity resulting in overall complexity of 83 million. (c) Ethidium bromide stained agarose gel,
200ng DNA/lane; Digest of genomic DNA after screens and plasmid DNA as control with the octamer recognition site enzyme PacI results in mostly large genomic fragments, while sgRNA fragments are 589 bp long (arrowhead). Long and short fragments can be fractionated using magnetic beads. (d) q-PCR on genomic DNA (gDNA), PacI digest gDNA and size separated fractions of digested gDNA. Error bars are s.d., shown are two biologically independent experiments in technical triplicate, equivalent to Figure 2b, all data points are shown
A pilot screen to identify optimal conditions for UMI-based CRISPR screen analysis.

(a) Setup of screen: Upon editing, various clonal outgrowth regiments, followed by clonal expansion and dropout screening, were run in parallel. Cas9 expression was induced by Dox, selection for cells harboring guide RNAs was performed by neomycin (G418) selection. Limiting dilution and expansion is variable in the experiment. Cells are treated with or without 3.3nM etoposide a LD₃₀ for 8 days. (b) Scheme illustrating variation in clone number and size (c) Average clone numbers and size determined from NGS data (d) Distribution of single cell derived clones in each regimen illustrated with guide_1 against Nhej1. P-value for each clone correlates with read depth but results in less data points. (e) Plot illustrating median dropout for each condition as well as p-value determined by combining multiple clones using MAGeCK. Signal to noise ratios (SNR) are highest in 148 clones of 35 reads, and the percentage of guides expected to have less than 5 clones due to variability in representation is with 0.06% lower than for 52 or 21 clone datasets.
Supplementary Figure 6

The screen layout for a sensitizer screen against etoposide.

Graphical illustration for large scale screen setup used to identify sensitizing mutations for etoposide. Cas9 expression was induced by doxycycline, selection for cells harboring guide RNAs was performed by neomycin (G418) selection. After washout of doxycycline and neomycin single cell derived clones are generated by a limiting dilution and clonal expansion. Cells are treated with 3.3nM Etoposide or mock treatment for 8 days.
a. Common hits

b. Conventional analysis specific hits

c. CRISPR-UMI specific hits
Supplementary Figure 7

Read distributions of single-cell-derived clones.

(a) No strong outlier clones are detected in hits identified by conventional analysis as well as CRISPR-UMI (b) Strong outlier clones with very high read counts as well as depletion are seen in putative false positive hits called by conventional analysis. (c) Genes identified only in CRISPR-UMI show modest but reproducible depletion in multiple independent clones but are often dominated by clones with high read counts that to not deplete.

Nature Methods: doi:10.1038/nmeth.4466
Supplementary Figure 8

A pooled dropout screen without clonal outgrowth showing outliers resulting in false positive calls.

(a) Comparison of CRISPR-UMI with conventional screen analysis on guide level in absence of clonal dilution and outgrowth shows highly correlative results (yellow) as well as discrepancy between both regimen (mixed colors, Pearson correlation: 0.729) (b) While correlating sgRNAs do not contain strong outlier clones based on total read count, guides only called in conventional analysis show outlier clones responsible for overall dropout. (c) Ranking of sgRNAs improves upon removal of outlier clones (top 3 clones by read count) from the dataset illustrating their confounding effects.
Supplementary Figure 9

A comparison of positive control guide detection in conventional analysis versus CRISPR-UMI.

Venn Diagrams illustrating the number of sgRNAs targeting the positive controls of the NHEJ complex (Lig4, Xrcc4-6, Nhej1) called within the top 50/100 hits.
Supplementary Figure 10

Validation of the reprogramming efficiency and predicted size distribution of colonies by UMI analysis from NGS data.

(a) Median size distribution of read counts per UMI for each sgRNA. Reads for each UMI were filtered for sequencing errors and median colony size is plotted relative to median size illustrating a marked size increase per iPS colony in many but not all identified roadblocks of reprogramming. Center line, median; hinges, 25th and 75th quartiles; whiskers, median ± 1.58 Å the interquartile range (IQR)/2. Individual data points represent outliers. (b) Alkaline phosphatase staining in 6 well dishes 10 days after Dox induction in the transgenic system illustrating enhanced iPS colony formation for guides targeting *Men1* or *Pias1*. (c) Representative colonies for comparison with Figure 5e stained with alkaline phosphatase in validation experiment on day 10 after Dox.