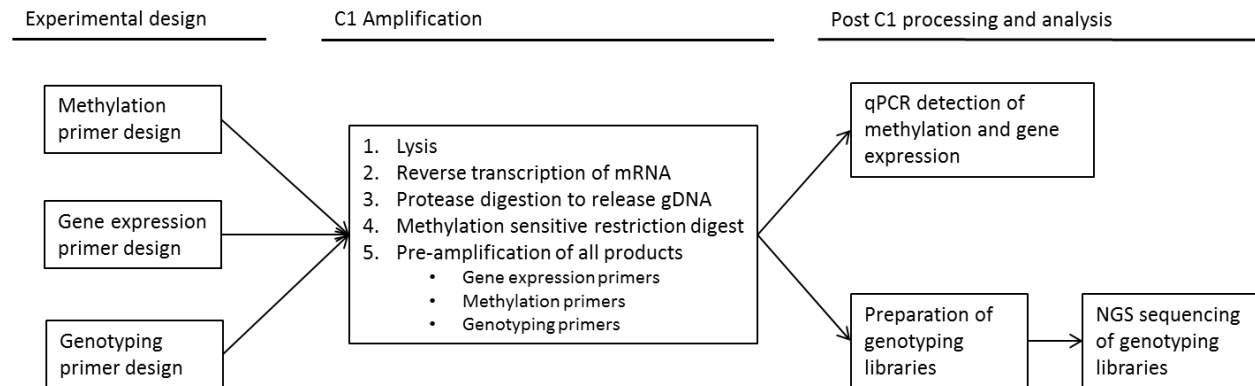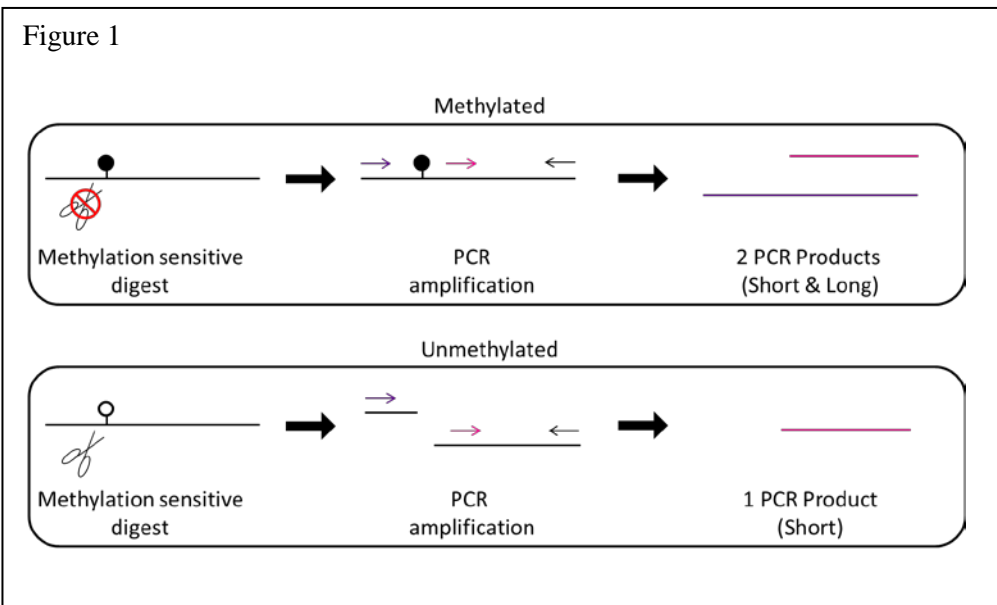## scGEM Workflow

Experimental Design

Single cell DNA methylation primer design

The scGEM DNA methylation assay uses qPCR to measure digestion of target loci by the methylation sensitive restriction endonuclease HpaII. Each locus to be assayed requires two forward primers (Short and Long) positioned on either side of the methylation sensitive endonuclease site of interest, and a single common reverse primer (fig). The short forward primer is downstream of the endonuclease site and will amplify regardless of the methylation status of the locus. The long forward primer is on the opposite side of the endonuclease site and will only amplify if the site is methylated and therefore protected from digestion by the endonuclease. The short amplification product will always be present and is used as a positive control while the presence or absence of the long product indicates the locus is methylated or unmethylated, respectively. Quality primer design is important for success of the single cell DNA methylation assay. The following guidelines will help with the design of primers to assess DNA methylation by this method:

1. Determine potentially methylated site(s) of interest that contain the target sequence for the HpaII endonuclease (5'-CCGG-3').
2. Methylation primers should be designed to non-transcribed regions to avoid amplification of cDNA
3. Design forward (short and long) and reverse primers positioned around the restriction site as shown (fig 1).
4. Primer design is done with Primer3 software, using the default parameters in version 0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm)
5. Amplicon size should be less than 150 bp to prevent amplicon dropout.
6. After primer design, specificity should be checked in silico.

7. It is recommended that primer sets be tested by amplification of undigested bulk genomic DNA to insure single products of the expected size are generated.
8. A nested primer design for the reverse primers can be used to overcome non-specific amplification with the outer pair used in the pre-amplification reaction on the C1 and the inner pair used for qPCR detection.
9. A positive control primers set to a known non-methylated region should be included to assess digestion efficiency.

Figure 1

Methylated

Methylation sensitive digest → PCR amplification → 2 PCR Products (Short & Long)

Unmethylated

Methylation sensitive digest → PCR amplification → 1 PCR Product (Short)

Single cell gene expression primer design

Gene expression primers were designed with Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). To specifically amplify cDNA, primers were designed to span exon-exon junctions or to span large intronic regions (>1000 bp). A nested primer design can be used to overcome non-specific amplification with the outer pair used in the pre-amplification reaction on the C1 and the inner pair used for qPCR detection.

Housekeeping genes can be used for quality control as discussed in the data analysis section.

Single cell genotyping primer design

Genotyping primers were designed to amplify genomic regions of interest. At least one primer from the primer pair was designed to an intronic region to specifically amplify genomic DNA.

Primer sequences used in our experiments. Users please feel free to adapt your own primers for making sequencing libraries:

Forward genotyping primer with universal priming site:
5'-CGACGTAAAACGACGGCCAGT – Gene Specific Sequence-3'

Reverse genotyping primer with universal priming site:
5'-GTGTGTCCTTTGTCGATACTGGTAC–Gene Specific Sequence-3'

Forward universal primer:
5'- AATGATACGGCGACCACCGAGA TCTACAC CGACGTAAAACGACGGCCAGT -3'

Reverse universal primer with sample barcode:
5'-GTTCGTCTTCTGCCGTATGCTCTA xxxxxxxx GTGTGTCCTTTGTCGATACTGGTAC-3'

sc-GEM C1 Post Processing and Analysis:

1. Amplified PCR product is from each C1 capture site was harvested and diluted to a final volume of 28µL in C1 Harvest Reagent.

Single cell gene expression and DNA methylation measurement:

1. Gene expression levels and DNA methylation status was performed in 96.96 Dynamic Array IFCs (Fluidigm) on the Biomark HD System (Fluidigm).
2. 1 µL of preamplification product was diluted with 99 µL C1 DNA Dilution Reagent (Fluidigm) and denatured at 95 °C for 10 minutes.
3. The sample pre-mix was prepared by combining 2.7 µL of denatured preamplification product, 0.3 µL 20X DNA Binding Dye Sample Loading Reagent (Fluidigm PN 100-3738) and 3µL 2X Sso Fast EvaGreen Supermix with Low Rox (Bio-Rad Laboratories, PN 172-5211).
4. The 10X assay mix was prepared by combining 3µL of 2X Assay Loading Reagent (Fluidigm PN85000736), 2.7µL nuclease free water and 0.3µL of a 100 µM each of Forward and Reverse Primer Mix. For DNA methylation qPCR analysis, each site will have two assay mixes, Forward-long/Reverse and Forward-short/Reverse.
5. The Dynamic Array chip was primed and loaded as recommended by the manufacturer.
6. The chip run was performed on the Biomark HD System with the following thermal protocol: 30 cycles of on-chip qPCR (60 s at 95 °C, 30 cycles of 5 s at 96 °C and 20 s at 60 °C)

Single cell genotyping by Sanger sequencing

1. 2 µL of PCR product harvested from the C1 IFC was amplified with the corresponding genotyping primers (0.4 µM each) in 50 µL GoTaq Flexi Buffer (2.5 mM MgCl2, 0.4 mM dNTPs; Promega) (2 min at 95 °C, 35 cycles of 30 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C, followed by 5 min at 72 °C).

2. PCR product was column purified (Qiagen PCR Purification kit) and Sanger sequenced with both the forward and reverse primers to determine single-cell genotype.

Single cell genotyping by next generation sequencing
1. 1 µL of each product harvested from the C1 IFC was amplified separately, with each corresponding set of genotyping primers containing a universal priming site (40 nM) and a barcoded universal primer containing Illumina-compatible adaptor sequence (0.4 µM) in 20 µL $1\times$ Express Sybr GreenER mix (Life Technologies) (5 min at 95 °C, 35 cycles of 30 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C).
2. Products from all genotyping amplicons generated from up to 96 different samples (with different sample barcodes) were pooled together at equal volume.
3. Pooled amplicons were purified with AMPure XP beads by first adding equal volume of well mixed Ampure XP Beads to the pooled amplicons and mixing thoroughly. The mixture was incubated at room temperature for 5 minutes and placed on a magnetic stand for 5 minutes. The supernatant was removed and discarded without disturbing the beads, 200 µL freshly prepared 80% EtOH was added and incubated for 30 seconds. The 80% EtOH supernatant was removed and discarded without disturbing the beads and 200 µL EtOH was added again and incubated for 30 seconds. The 80% EtOH supernatant was removed and discarded again and the beads were air dried on the magnetic stand for 5 minutes. The beads were resuspended with 32 µL of resuspension buffer by mixing thoroughly with a pipette 10 times. The mixture was incubated at room temperature for 5 minutes and placed on a magnetic stand for 5 minutes. 30 µL of the supernatant (purified library) was transferred to a new tube for subsequent step.
4. Purified library was quantified using the KAPA Library Quantification kit (KK4824), and sequenced on a next-generation-sequencing instrument (Illumina MiSeq) using 150-bp paired-end reads.

Data processing

Single cell gene expression and DNA methylation

The Fluidigm Real-Time PCR Analysis software was used to calculate Ct values from microfluidics qPCR. Raw data was first filtered to remove bad quality cells. Samples in which none of the PCR amplicons from the DNA methylation assay were detected were removed from the analysis as DNA was possibly lost in these samples. We also removed samples in which housekeeping gene transcripts were not amplified (β-actin for the reprogramming experiments, both β-actin and β2-microglobulin for the LUAD experiments), indicating low quality RNA, or in which the unmethylated β-actin promoter region containing an HpaII recognition site was amplified, indicating incomplete digestion with HpaII.

For the DNA methylation assay, the presence or absence of the long restriction site-spanning amplicon indicates whether a particular locus is methylated or unmethylated. To denote the binary nature of DNA methylation state, we converted the Ct values for the long amplicons to 15 if they were amplified (Ct<30) and to 999 if they were not amplified. We observed that the short control amplicons (positive control for presence of DNA) amplified in almost all the tested loci (99.6%) indicating very effective amplification and no loss of single cell DNA. Thus we can assay for DNA methylation status of a locus with one PCR assay (instead of two previously3) and significantly increase the number of loci that can be assayed on a single microfluidics qPCR chip.

The modified Ct values were analyzed with the SINGuLAR Analysis Toolset (Fluidigm) and Custom R Scripts. Scripts are available upon request. The limit of detection was set to be 30 threshold cycles (Ct) representing the last cycle of qPCR. Genes that were not expressed were given a Ct value of 999. Log2 gene expression and DNA methylation values were calculated with the formula $Log2Ex = 30 - Ct$ if $Ct < 30$ and $Log2Ex = 0$ if $Ct > 30$.

Single cell genotyping by next generation sequencing

To analyze genotypes from next-generation-sequencing data, reads from the MiSeq was demultiplexed by the sample barcode and aligned to the human genome (hg18) using Bowtie2. Aligned reads were visualized on IGV. Due to amplification from very small amounts of DNA from single cells, stochastic differences in amplification rates during early cycles of PCR could significantly affect the observed allelic ratios of mutant bases. We called a mutation in a single cell if the allele ratio of the mutant base exceeded 5% and if the same mutation was observed in at least one other cell.