Profiling embryonic stem cell differentiation by MALDI TOF mass spectrometry: development of a reproducible and robust sample preparation workflow

Rachel E. Heap[1], Anna Segarra-Fas[2], Alasdair P. Blain[3], Greg M. Findlay[2] and Matthias Trost[1*]

[1]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle-upon-Tyne, UK

[2]MRC Protein Phosphorylation & Ubiquitylation Unit, University of Dundee, Dundee, Scotland, UK;

[3]Wellcome Trust Centre for Mitochondrial Research, Newcastle University, Newcastle-upon-Tyne, UK;

Corresponding Author

*Email: matthias.trost@newcastle.ac.uk; phone: +44 191 2087009

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Supplementary methods

Chemicals

All reagents used for this study were of HPLC grade unless stated otherwise. Acetonitrile (ACN), water (H_2O), Methanol (MeOH) were purchased from Merck (Darmstadt, GER). Cell lines U2OS, MCF7, THP1 and HEK293 were purchased from ATCC (Manassas, USA). Dulbecco's Modified Eagle Medium (DMEM), Roswell Park Memorial Institute medium (RPMI), phosphate buffered saline solution (PBS), fetal bovine serum (FBS), β-mercaptoethanol (cell culture grade) were all purchased from Gibco, Life Technologies (Darmstadt, Germany). L-glutamine and penicillin-streptomycin were purchased from Lonza (Basel, Switzerland). Trypsin-EDTA solution, 37% paraformaldehyde solution (PFA) and trifluoroacetic acid (TFA) were purchased from Sigma Aldrich. Sinapinic Acid (SA), 2,5-dihydroxybenzoic Acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA) were kindly provided by Bruker Daltonics (Bremen, Germany).

MALDI TOF target preparation. AnchorChip (1536) and Ground steel (384) MALDI targets (Bruker, Bremen, Germany) were first cleaned thoroughly by submersion and sonication in isopropanol for five minutes before subsequent sonication with 30% acetonitrile, 0.1% TFA solution for a further five minutes. Targets were then washed briefly with HPLC grade methanol before drying by nitrogen gas.

MALDI MS Parameters. A random walk pattern (complete sample) on spot laser ablation pattern was used with a M5 Smart beam Parameter at a 45-μm × 45-μm scan range. Spot diameter was limited to 2000 μm and a random walk pattern movement enabled at 1000 shots per raster position. Ionization was achieved using a variable laser power between 50-85% (laser attenuator offset 14%, range 30%) with a detector gain of ×6.8 in the 2000 to 20,000 m/z mass range with a mass suppression up to 1600 m/z. Samples were analysed in a linear geometry with optimized voltages for ion sources (ion source 1, 20 kV, PIE 1.3 kV), lens (8.6 kV), and a pulsed ion extraction of 180 ns. A novel 10-bit digitizer was used at a sampling rate of 1.25 GS/s. Raw data were processed first by a TopHat baseline subtraction followed by smoothing with a SavitzkyGolay algorithm.

The laser parameters were set up as follows: Global Attenuator Offset 0%, Attenuator Offset 42%, Attenuator Range 30%, Focus Offset 0%, Focus Range 100%, Focus Position 19%, Smartbeam index 1.

MALDI-TOF-MS Calibration

Calibration of the instrument was performed prior to each experiment. Protein calibration standard (i) was prepared and used in accordance with the manufacturer's procedure (Bruker Daltonics). Protein calibration standard (i) contains the following calibrations: Insulin [M+H]$^+$ (5734.51 m/z), Ubiquitin I [M+H]$^+$ (8565.76 m/z), Cytochrome C [M+H]$^+$ (12360.97 m/z), Myoglobin [M+H]$^+$ (16952.30 m/z), Cytochrome C [M+H]$^{2+}$ (6180.99 m/z) and Myoglobin [M+H]$^{2+}$ (8476.65 m/z).

Perseus analysis of mESC peak intensities. Raw data were first separated into their corresponding biological and technical replicates with reported relative intensity (Supp. Table 1). A complete matrix was imported into Perseus (v.1.6.0.7) and data was filtered so that only features that were identified in 10 spectra in total were selected. This filtered mass list was then used to generate PCA plots using Perseus. Z-score averaging was performed on the filtered data set and included unique peaks identified to each of the two conditions. Hierarchical clustering was performed with Euclidian distancing and 20 and 10 number of clusters for row and column trees respectively

Table S-1: Primers and their corresponding sequences used for qPCR analysis.

Gene	Forward (5' to 3')	Reverse (5' to 3')
Nanog	CTCATCAATGCCTGCAGTTTTTCA	CTCCTCAGGGCCCTTGTCAGC
Klf4	ACACTTGTGACTATGCAGGCTGTG	TCCCAGTCACAGTGGTAAGGTTTC
Oct4	AGCTGCTGAAGCAGAAGAGG	AGATGGTGGTCTGGCTGAAC
Fgf5	GCTGTGTCTCAGGGGATTGT	CACTCTCGGCCTGTCTTTTC
Dnmt3b	CTGGCACCCTCTTCTTCATT	ATCCATAGTGCCTTGGGACC
Gapdh	CTCGTCCCGTAGACAAAA	TGAATTTGCCGTGAGTGG

Table S-2 Measured diameter of each cell line with standard deviation as well as optimal cell number on spot derived from the titration

Cell line	U2OS	MCF7	THP1	HEK293
Cell diameter (µm)	21.6 ± 1.4	18.1 ± 0.75	8.0 ± 0.25	16.9 ± 0.63
Optimal cell number	250	500	1000	500
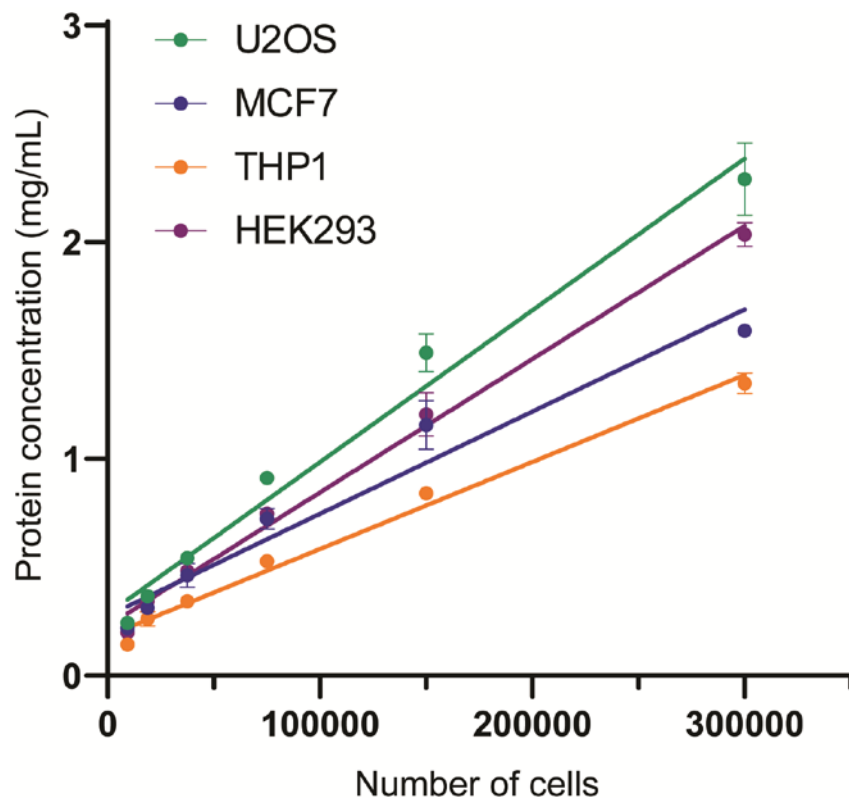
Figure S1. Protein concentration of the four cell lines (U2OS, MCF7, THP1 and HEK293) by BCA quantitation from 300,000 – 9,000 cells. Fitting linear regressions gave an $R^2 > 0.9$, thus showing a correlation between number of cells and protein concentration. These data also show how protein concentration varies between the different cell lines.
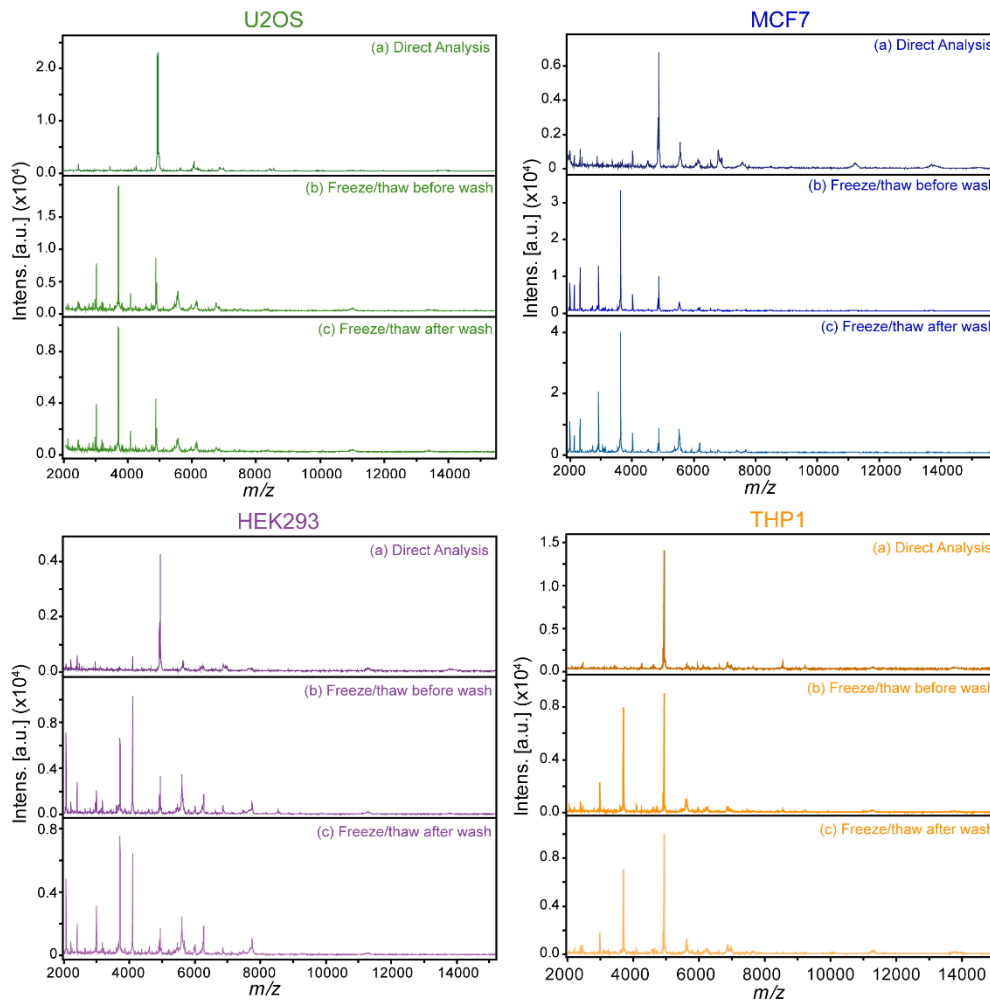

Figure S2: Freeze/thaw method increases number of features in spectra. MALDI TOF spectra of each of the four cell lines (U2OS, MCF7, THP1, HEK293) for the freeze/thaw methods: direct analysis (a), freeze/thaw first (b), freeze thaw after wash (c). Freeze/thaw cycle significantly increases number of features identified for each of the cell lines. The point at which this step is performed does not negatively affect spectral sensitivity or information.
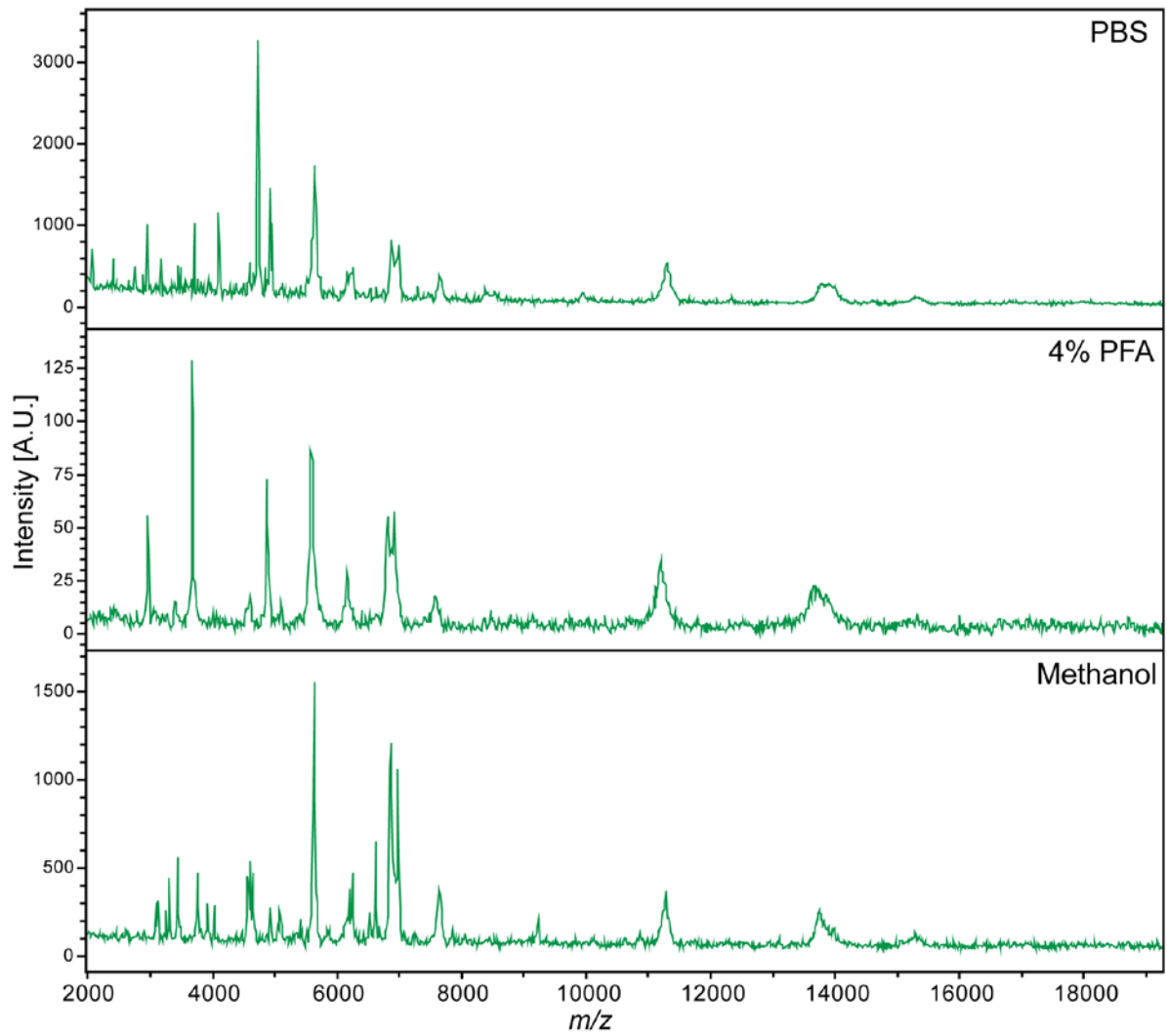
Figure S3: Fixing cells improves spectra (U2OS). Representative MALDI TOF spectra of U2OS cell line that either washed with PBS or fixed with 4% PFA or Methanol.
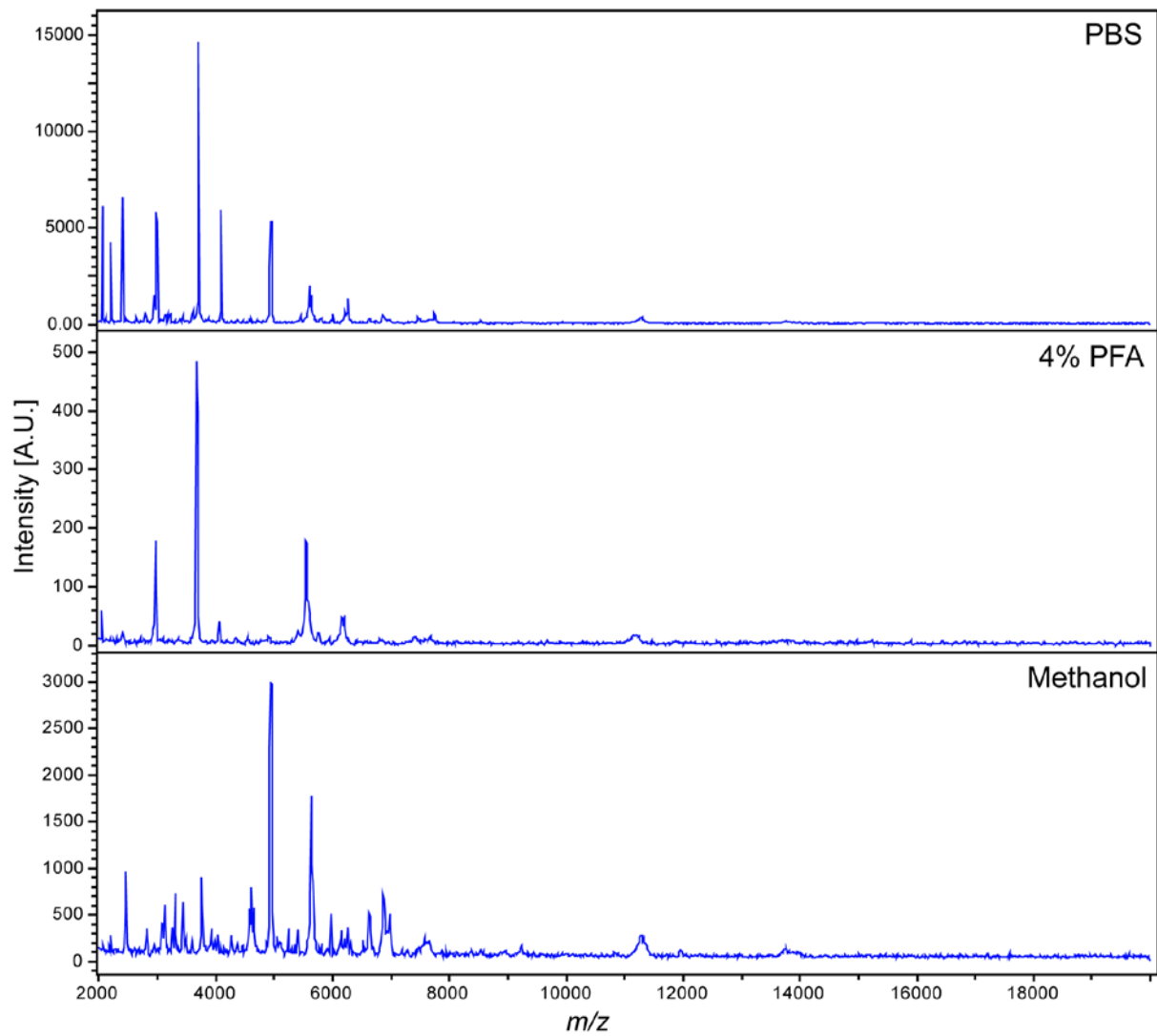
Figure S4: Fixing cells improves spectra (MCSF7). Representative MALDI TOF spectra of MCSF7 cell line that either washed with PBS or fixed with 4% PFA or Methanol.
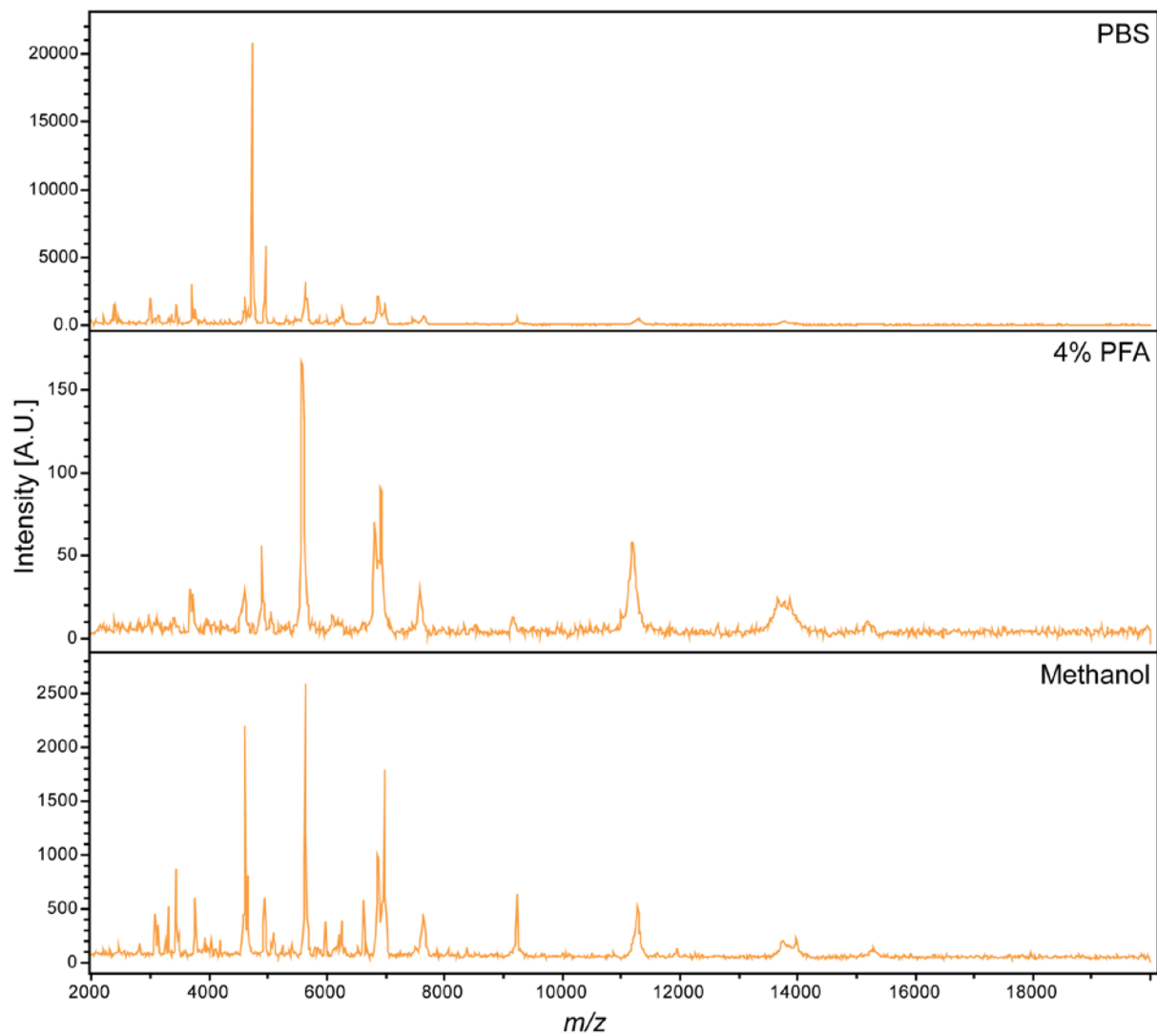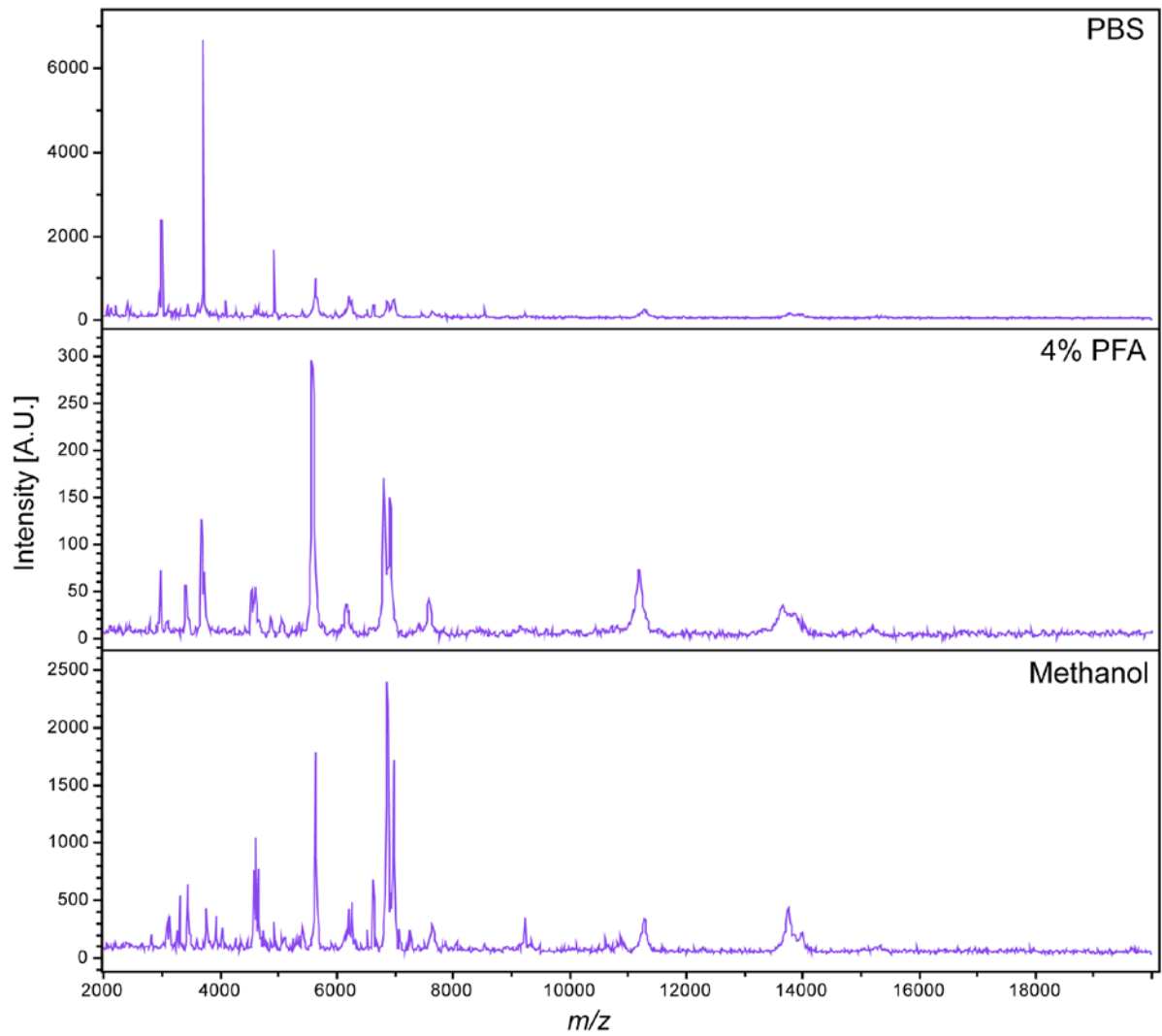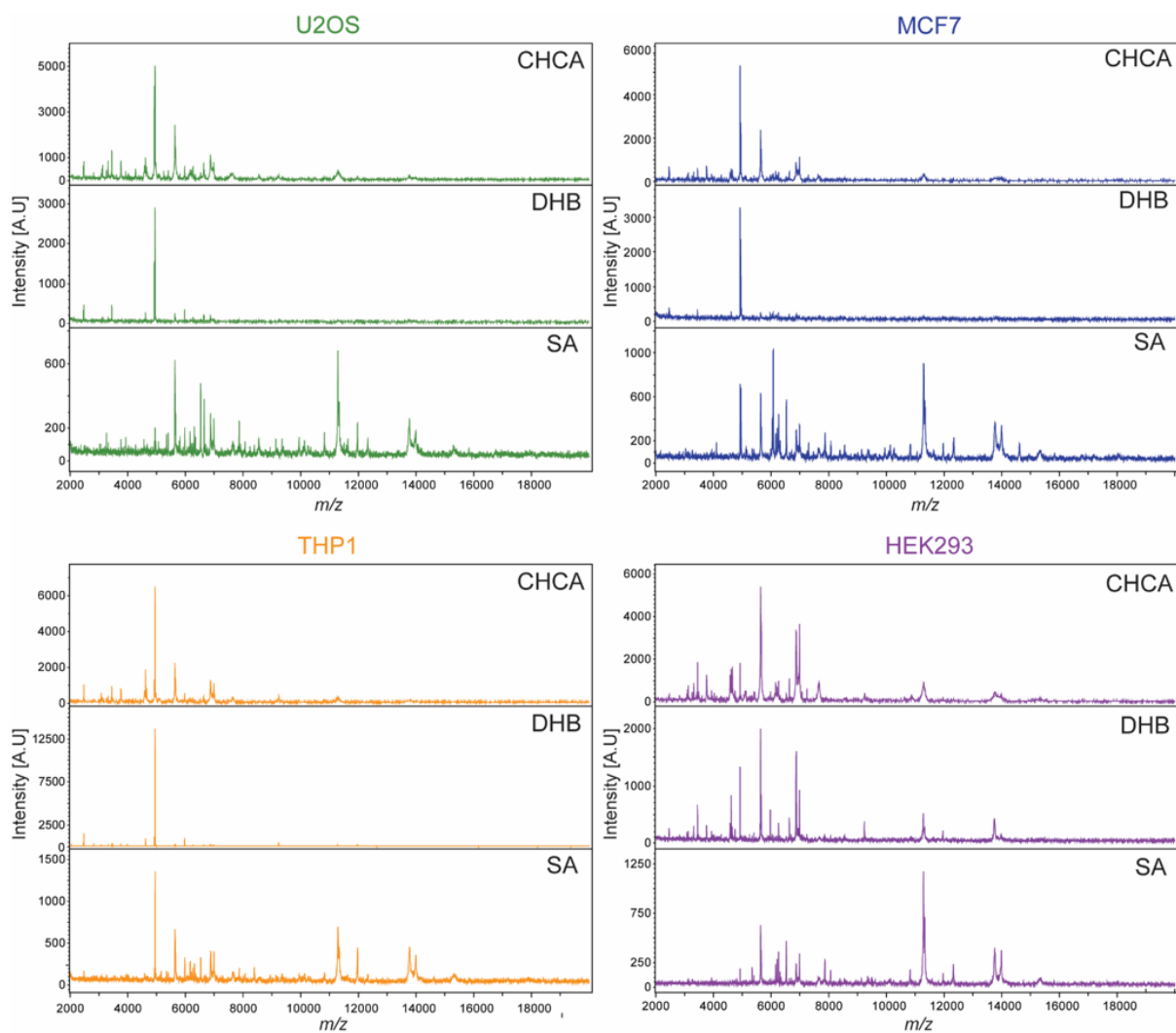
Figure S5: Fixing cells improves spectra (THP1). Representative MALDI TOF spectra of THP1 cell line that either washed with PBS or fixed with 4% PFA or Methanol.

Figure S6: Fixing cells improves spectra (HEK293). Representative MALDI TOF spectra of HEK923 cell line that either washed with PBS or fixed with 4% PFA or Methanol.

Figure S7: Spectra of cell lines with different matrices. Representative MALDI TOF spectra of each of the four cell lines (U2OS, MCF7, THP1, HEK293) for each of the saturated matrix conditions: α-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB) and sinapinic acid (SA). Unique mass spectra were acquired for each cell line when mixed with the different matrices.
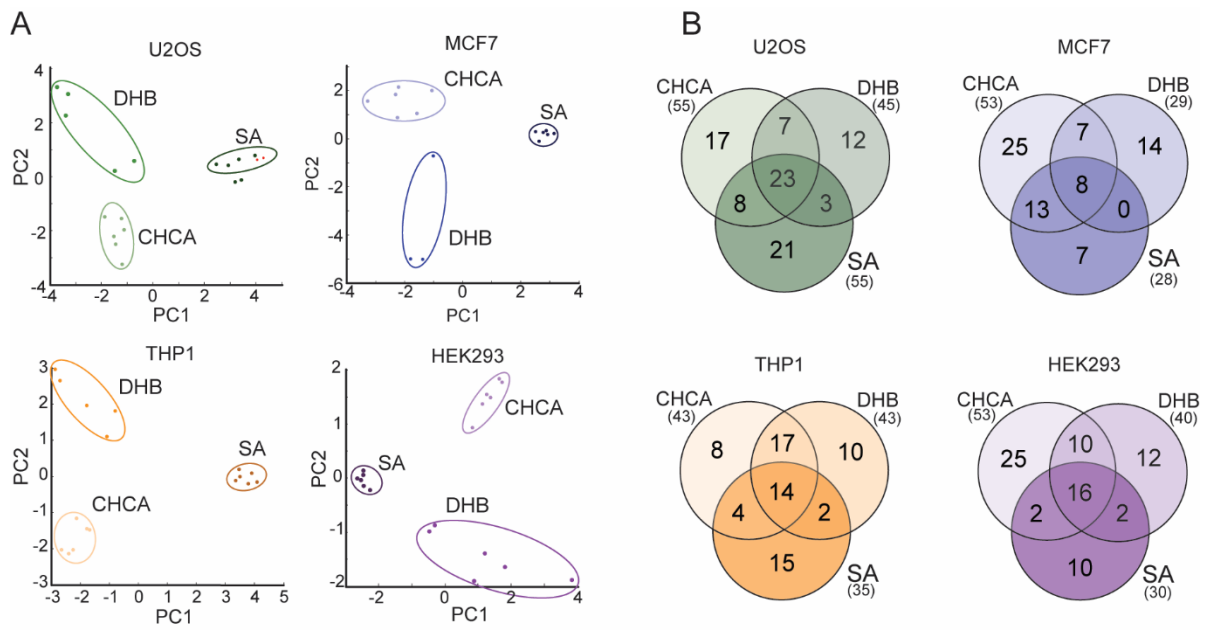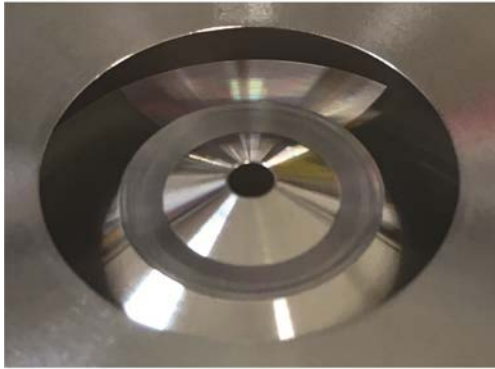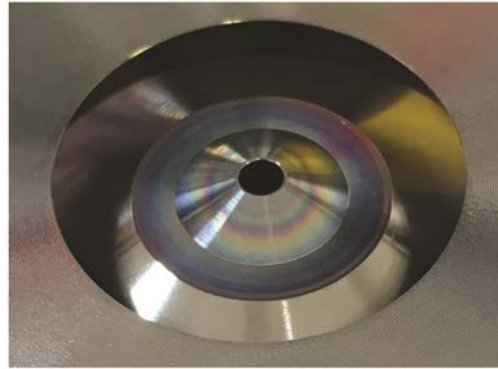
Figure S8: Unique features and reproducibility of each matrix condition. PCA plots showing unique groupings for each of the matrix conditions and Venn diagrams showing unique and common peaks identified for each of the four cell lines (U2OS, MCF7, THP1, HEK293) over six technical replicates.
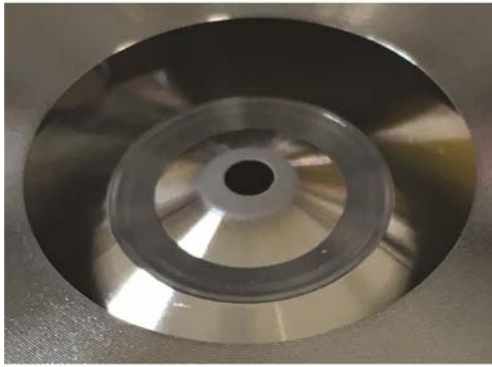
Figure S9: Lens contamination depending on matrix. Photo images showing the different lens contamination patterns after analysing 3072 spots with each matrix at the described laser powers in Table 2 and below.
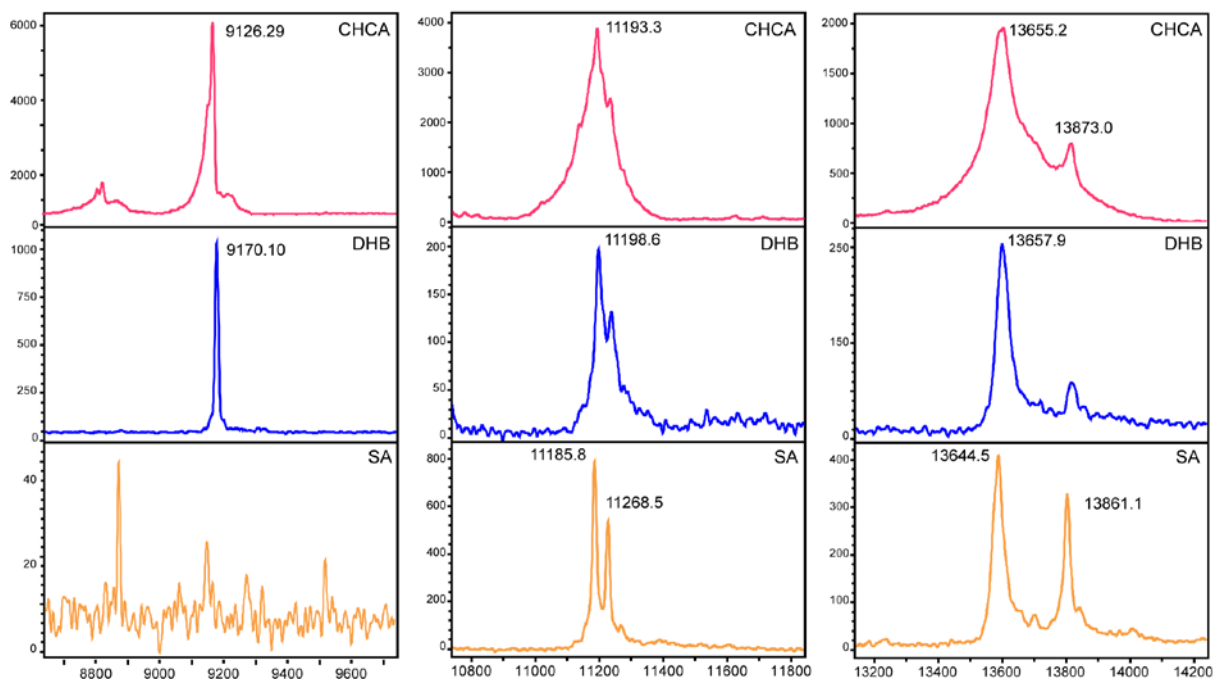
Figure S10: Peak resolution with the three different matrices. Resolution of three peaks 9,150, 11,200 and 13,650 *m/z* showing how SA yields the greatest peak resolution at the higher molecular weight features compared with DHB and SA. However does not detect all the common peaks such as 9,150 *m/z*. DHB exhibits greater resolution than CHCA but much poorer intensity.
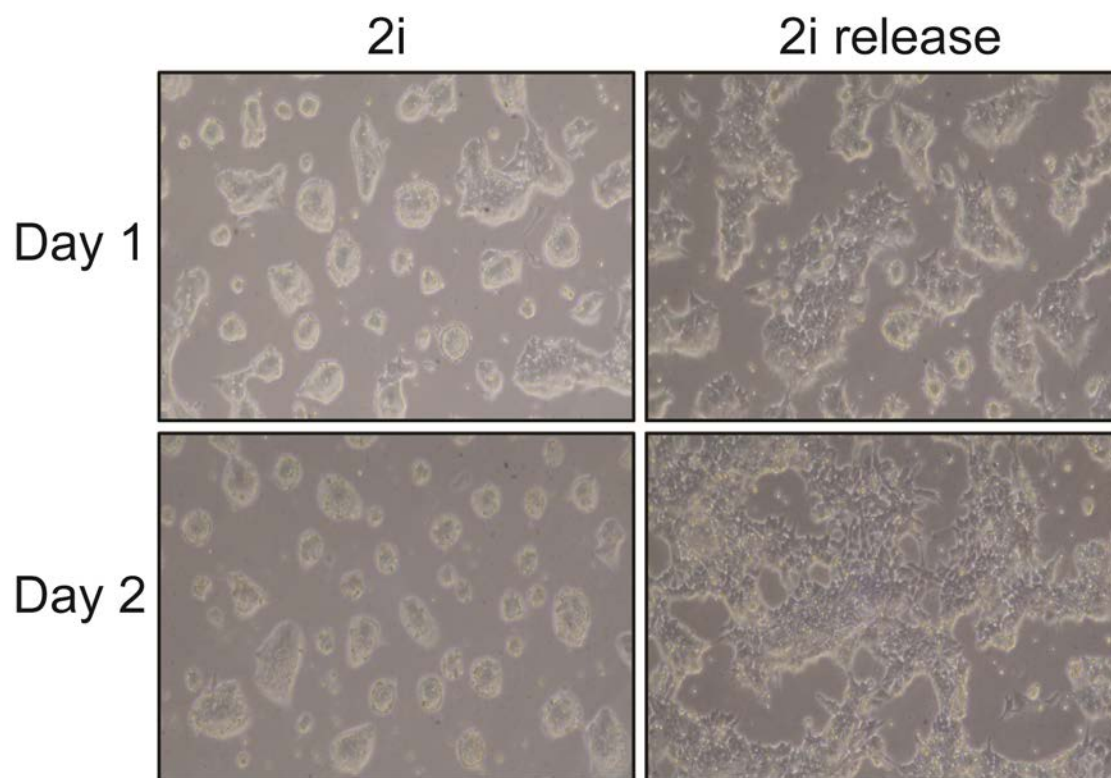
Figure S11: Morphological changes of mESCs light microscopy. Image shows the morphological change in naïve mESCs upon 2i release after 1 and 2 days.
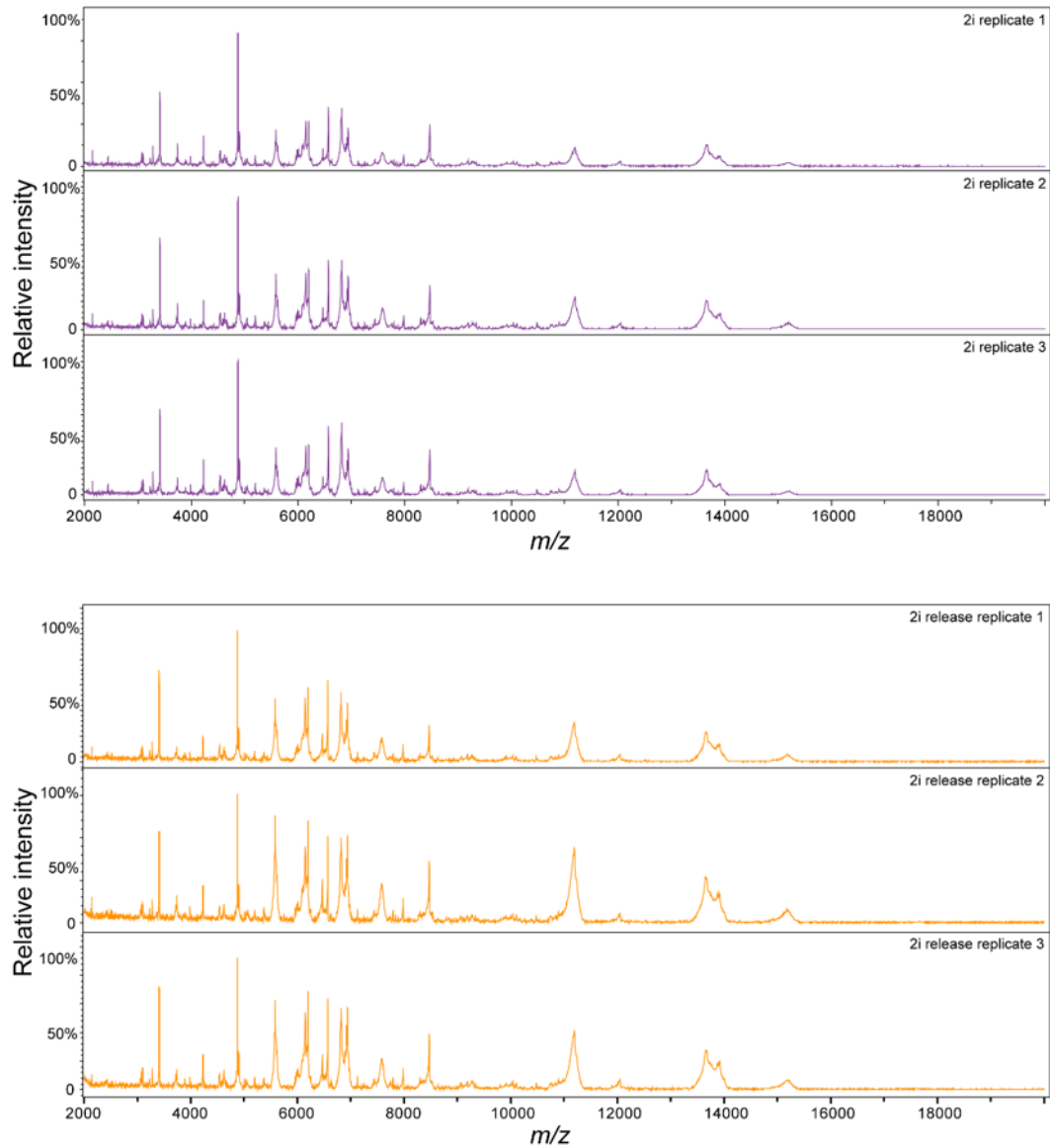
Figure S12: 2i and 2i release spectra. Representative MALDI TOF spectra of each of the three individual biological replicates for 2i and 2i release cell populations.
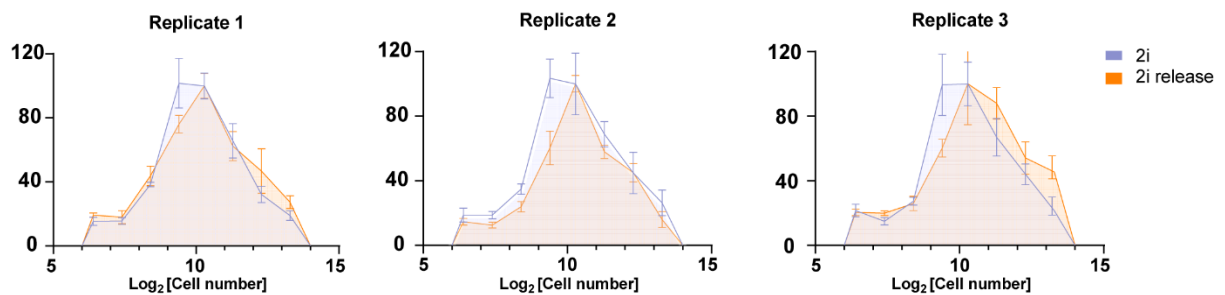

Figure S13: Cell titration of mESCs. Ionisation profiles using average spectral intensity of both the 2i and 2i release conditions showing how both populations have the same optimal cell number on target.

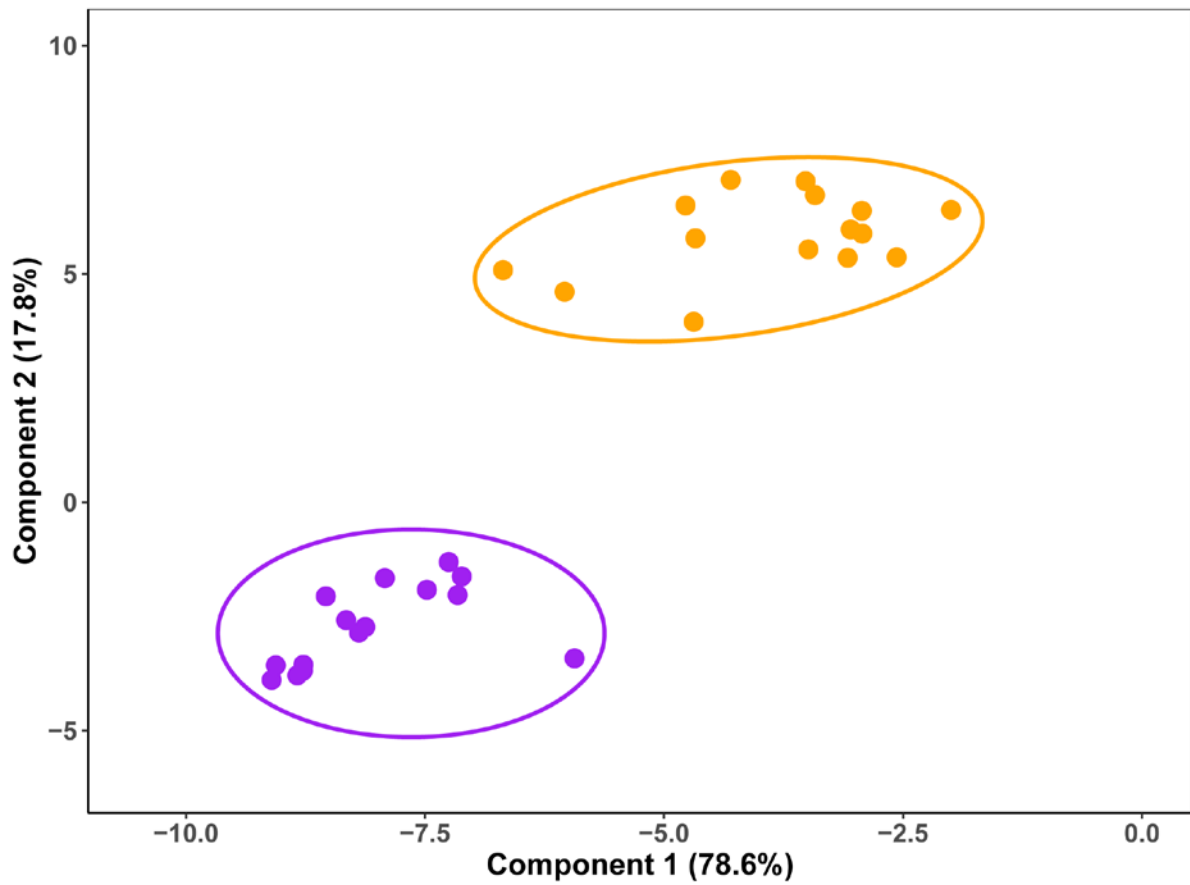**Figure S14:** PCA plot of the 5 technical of three biological 2i vs 2i release samples using Jackknife analysis by which we could differentiate the two conditions. Uncertainty in the obtained loadings estimates was assessed via the jackknife method.
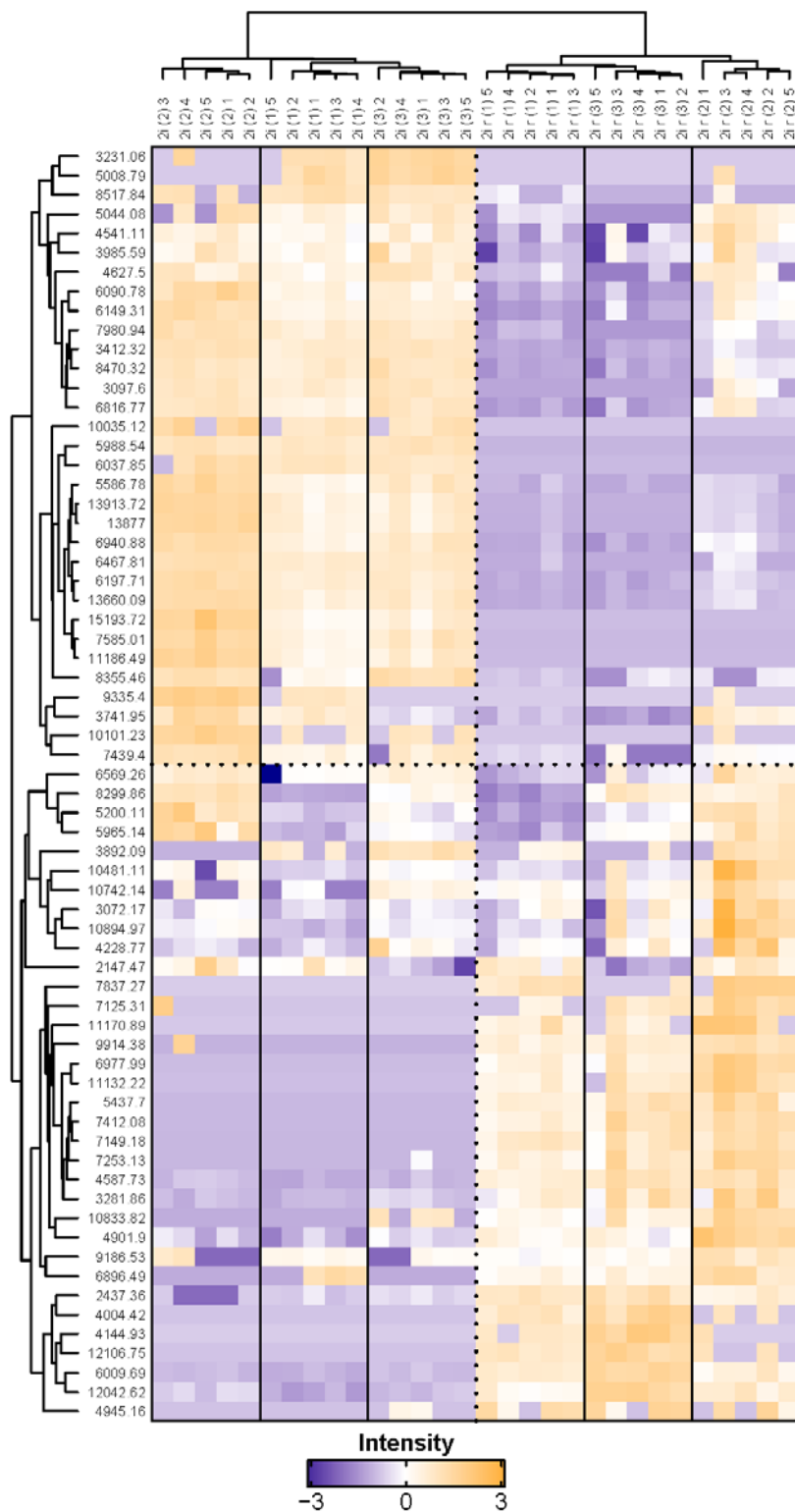
Figure S15: Hierarchical cluster intensity-based heatmap of the five technical replicates of three biological replicates within the 2i and 2i release cell populations using the pvclust R tool, with 10,000 bootstrap iterations.

Figure S16: Hierarchical cluster dendogram of the five technical replicates of three biological replicates within the 2i and 2i release cell populations using the 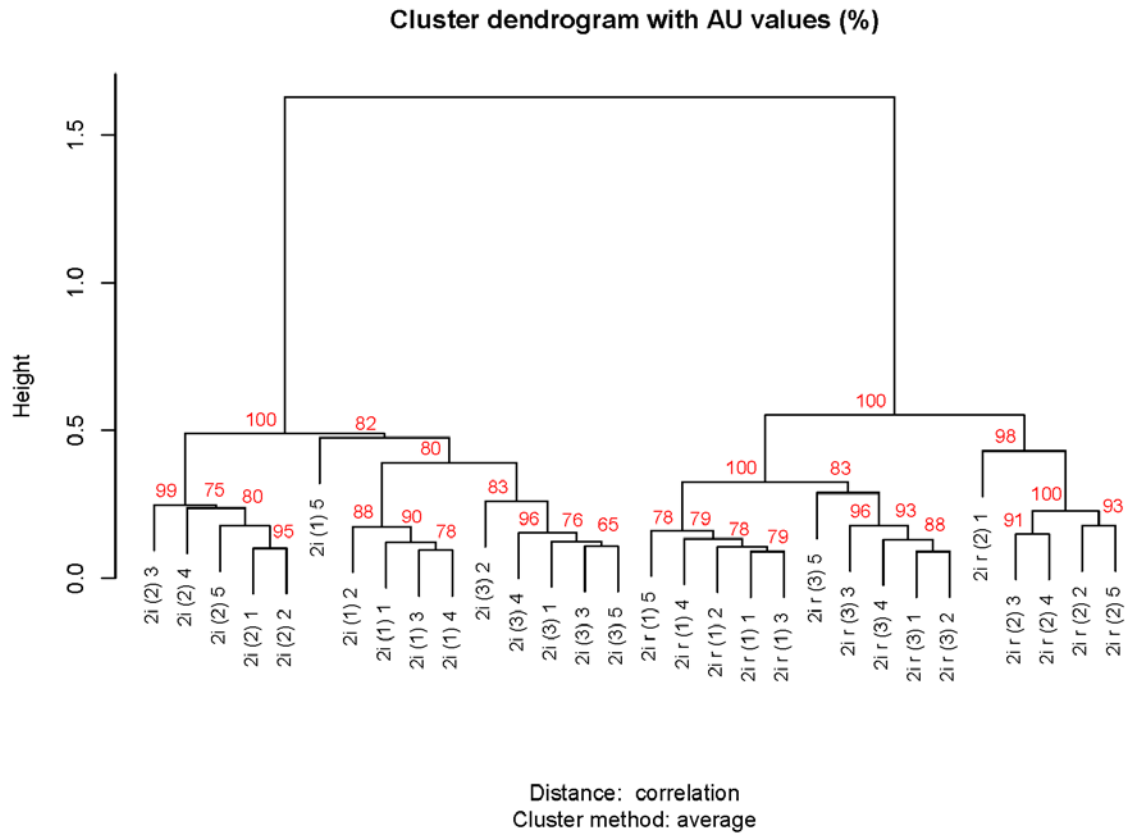pvclust R tool, with 10,000 bootstrap iterations. Scores in red represent the approximately unbiased (AU) statistical significance of each dendogram node on a 1 to 100 scale, scores above 95 are considered significant.

Replicate	PC1 (95% CI)	PC2 (95% CI)
2i.1.T1	-7.25 (-7.92:-6.59)	-1.31 (-2.18:-0.44)
2i.1.T2	-7.16 (-7.44:-6.88)	-2.03 (-2.39:-1.68)
2i.1.T3	-7.48 (-7.95:-7.01)	-1.92 (-2.61:-1.23)
2i.1.T4	-7.12 (-7.79:-6.44)	-1.62 (-2.54:-0.71)
2i.1.T5	-5.94 (-8.44:-3.43)	-3.42 (-6.16:-0.67)
2i.2.T1	-9.06 (-9.45:-8.67)	-3.57 (-4.13:-3.01)
2i.2.T2	-8.78 (-9.27:-8.29)	-3.69 (-4.26:-3.12)
2i.2.T3	-8.77 (-9.25:-8.29)	-3.56 (-4.26:-2.86)
2i.2.T4	-8.84 (-9.34:-8.33)	-3.79 (-4.36:-3.22)
2i.2.T5	-9.1 (-9.77:-8.43)	-3.89 (-4.85:-2.93)
2i.3.T1	-7.92 (-8.67:-7.17)	-1.66 (-2.56:-0.77)
2i.3.T2	-8.54 (-9.11:-7.97)	-2.06 (-2.66:-1.45)
2i.3.T3	-8.12 (-8.49:-7.76)	-2.73 (-3.14:-2.32)
2i.3.T4	-8.32 (-8.72:-7.93)	-2.58 (-2.96:-2.2)
2i.3.T5	-8.19 (-8.58:-7.81)	-2.86 (-3.19:-2.52)
2i.r.1.T1	-3.49 (-3.84:-3.14)	5.54 (5.04:6.04)
2i.r.1.T2	-2.93 (-3.28:-2.58)	5.88 (5.38:6.39)
2i.r.1.T3	-3.05 (-3.48:-2.61)	5.97 (5.54:6.41)
2i.r.1.T4	-3.08 (-3.46:-2.69)	5.35 (4.83:5.88)
2i.r.1.T5	-2.57 (-3.05:-2.08)	5.36 (4.39:6.34)
2i.r.2.T1	-4.69 (-5.28:-4.09)	3.95 (3.07:4.83)
2i.r.2.T2	-4.77 (-5.11:-4.43)	6.5 (5.84:7.16)
2i.r.2.T3	-6.68 (-7.71:-5.66)	5.08 (3.55:6.61)
2i.r.2.T4	-6.04 (-7.08:-5)	4.61 (3.52:5.69)
2i.r.2.T5	-4.67 (-5.28:-4.07)	5.78 (4.58:6.98)
2i.r.3.T1	-3.52 (-3.98:-3.06)	7.03 (6.08:7.98)
2i.r.3.T2	-3.42 (-3.73:-3.11)	6.72 (5.89:7.56)
2i.r.3.T3	-4.3 (-4.82:-3.79)	7.06 (6.47:7.65)
2i.r.3.T4	-2.93 (-3.4:-2.46)	6.38 (5.27:7.5)
2i.r.3.T5	-2 (-2.71:-1.28)	6.4 (4.6:8.2)

Table S3. Principle Component loadings (with 95% confidence intervals) for the first two axes. PCA model fit via the expectation–maximization algorithm, with uncertainty in the obtained loadings estimates assessed via the jackknife method, using the R package MetabolAnalyze.