

Method for reprogramming cells

Cell therapy, the treatment of patients with severe diseases using cells expanded and potentially modified in the laboratory, is a very promising technology that is delivering
5 excellent results in oncology (both for leukemias and solid tumors) (Ottaviano et al. 2022) and regenerative medicine (for instance in the recovery of burn victims) (Cossu et al. 2018). Cells can be safely derived from patients or donors, manipulated to expand, differentiate and potentially correct genetic defects or enhance their therapeutic capabilities, and then administered back to the patient. Unfortunately, though, the production of these advanced
10 therapy medicinal products (ATMPs) is still very challenging. One of the major limitations is the lack of access to therapeutically useful amounts of cells.

The promise of stem cells for regenerative medicine consists in the unprecedented possibility to derive any type of cell from a human pluripotent stem cell such as an induced
15 pluripotent stem cell (iPSC) or a human embryonic stem cell (hESCs). These pluripotent stem cells have the potential to generate *in vitro* any human tissue, since they are part – or very similar to – the embryonic structures that give rise to all primordial structures in our body. The main challenge, however, is directing these stem cells in robust and scalable ways to develop into specific cell types useful for therapy. The expansion and/or
20 differentiation of clinically useful cell products from donor-derived stem cells requires costly and time-consuming testing and optimisation of cell culture conditions that are tailored to the specific application and cell type of interest. Cell culture conditions must be formulated by carefully considering the developmental cues in the human embryo, many of which are imperfectly understood. In any case, these may not be effective in the absence of the
25 physiological niche which provides additional stimuli that are difficult to model in cell culture, e.g. specific cell-cell interactions, extracellular matrix composition and tissue stiffness. There is no *a priori* method to determine which of these conditions will yield a highly homogeneous, correctly differentiated cell population, with high reproducibility and in a process that is scalable and has the potential to meet regulatory requirements.

30 Cell reprogramming, the process of converting one cell type to another cell type through manipulation of gene regulatory networks that define cell identity, may have great promise for regenerative medicine but has yet to be applied broadly. Although it is now well known that it is possible to switch the phenotype of one cell type to another by manipulating genetic circuits, the elements required for cell fate conversion are difficult to identify and in most
35 instances unknown. The identification of factors to directly reprogram the identity of cell

types is currently limited by, amongst other things, the cost of exhaustive experimental testing of plausible sets of factors, an approach that is inefficient and unscalable.

One of the most important examples of cell reprogramming is the conversion of a somatic cell to a pluripotent state using exogenous transcription factors (TFs). For example,

- 5 Yamanaka (Yamanaka S. Cell 2006; 126:663-76; PMID:16904174) demonstrated that a small set of TFs, namely OCT4, KLF4, SOX2 and MYC (OKSM), when introduced into the cell can reprogram a human fibroblast to an induced pluripotent stem cell (iPSC). Subsequently, other groups have demonstrated that it is possible to convert fibroblasts to hepatocytes, to cardiomyocytes and to several other cell types (Huang P, et al.. Nature
- 10 2011; 475:386-9; Sekiya S, Suzuki A., Nature 2011; 475:390-3; Kogiso T, et al., Hepatol Int 2013; 7:937-44; Ieda M, et al., Cell 2010; 142:375-86; Song K, et al., Nature 2012; 485:599-604; Qian L, et al., Nature 2012; 485:593-8; Takahashi K, Yamanaka S., Nat Rev Mol Cell Biol 2016; 17:183- 93; Sadahiro T, et al., Circ. Res. 2015; 116:1378-91. Tsunemoto RK, et al., EMBO J 2015; 34:1445-55; Heinaniemi M, et al., Nat Methods 2013; 10:577-83; Lang
- 15 AH, et al., PLoS Comput Biol 2014; 10:e1003734; Del Sol ICA, Stem Cells 2013; 31:2127-35; Davis FP, Eddy SR. PLoS One 2013; 8:1-8; D'Alessio et al., Stem Cell Reports 2015; 5:763-75).

- In addition to reprogramming cells to obtain a pluripotent cell (iPSC), a number of related reprogramming strategies have been demonstrated, such as: (i) 'forward programming'
- 20 where a less differentiated cell such as an iPSC is reprogrammed into a specific cell type (Dalby et al, Stem Cell Reports 11:1462-1478) (ii) 'direct conversion' where one differentiated cell type is switched into another without going through a pluripotent or progenitor state (Zhou et al, Nature 455:627-632); as well as combinations of the foregoing such as (iii) 'indirect (or primed) reprogramming' where ectopic OKSM is expressed
- 25 transiently to effect partial reprogramming to a less differentiated (but not fully pluripotent) intermediate, followed by forward programming to obtain a more differentiated cell type that differs from the starting cell type (Efe et al, Nat Cell Biol 13:215-222).

- Note that '*cell reprogramming*' differs qualitatively from traditional '*cell differentiation*', both in biological mechanism and by the practical method by which it is achieved. Cell differentiation
- 30 occurs by the progressive restriction of developmental potential until a terminal state is reached - famously illustrated by Waddington who imagined a ball rolling from the top of a mountain towards a lower energy resting place in the valley below (*The Strategy of the Genes*, pub. Allen & Unwin, London, 1957) - whereas in cell reprogramming this is not necessarily true (and in the case of iPSCs it is precisely the opposite). In cell differentiation

methods, (stem) cells in culture are exposed to developmental cues such as morphogens, growth factors, hormones, small molecules etc. which are extracellular effector molecules added to cell culture media and typically function by binding protein elements of the cell signalling apparatus such as cell surface receptors, and cytoplasmic signalling mechanisms such as those mediated by kinase cascades etc. In the case of cell reprogramming, the cell fate switch is directed by ectopic expression of transcription factors (TFs) which translocate to the cell nucleus and directly control gene expression, typically by binding gene regulatory elements in enhancer and promoter DNA and stimulating (or inhibiting) the transcription of mRNA. These two mechanistically different strategies for manipulating cell fate -- the first of which is mediated by controlling the cellular environment (i.e. directed differentiation) and the second of which is mediated by rewiring genetic circuits that determine cell fate (cell reprogramming) -- have also been termed the "outside-in" and "inside-out" approaches, respectively (Shakiba et al, Cell Syst. 12:561-592).

The TFs that orchestrate cell reprogramming may themselves regulate one or more TFs amongst other genes, inducing and/or stabilising a Gene Regulatory Network (GRN) that specifies cell fate. One of the best understood GRNs is the pluripotency gene regulatory network (PGRN) that is induced and stabilised by adding the Yamanaka factors OKSM to a (differentiated) cell. At the core of the PGRN is a triumvirate of pluripotency-inducing TFs, namely OCT4, SOX2 and NANOG. These TFs regulate a larger network of interconnected secondary genes. The core TFs do so in a cooperative manner, and form autoregulatory and feed-forward gene circuits that result in network stability. It is known that NANOG and LIN28 can replace KLF4 and MYC in the Yamanaka cocktail, indicating a specific GRN can be induced by alternative means. Indeed, it is known that reprogramming can also be affected by non-coding RNAs that regulate gene expression, such as LincRoR and Let7 that affect key nodes in the PGRN, therefore in the present application non-coding RNAs that regulate gene expression are included in the definition of TFs.

GRNs are essentially composed of genomic components such as genes and their cis-regulatory modules at the nodes, and of regulatory state components, i.e. TFs (including non-coding RNAs) which provide regulatory input into these modules. The connections between different regulatory gene products (i.e. TFs) and the cognate cis-regulatory elements controlling other network nodes, specify a network circuitry that can explain developmental processes such as the maintenance of pluripotency in the blastocyst, or e.g. the differentiation of a pluripotent stem cell into a pancreatic beta cell.

Few if any GRNs are as well defined as the PGRN, and also the TFs that effect specific cell reprogramming transitions are largely unknown. Recently, a number of computational methods for predicting TFs useful in promoting reprogramming from one cell type to another have been developed (reviewed by Kamaraj et al (Cell Cycle 15: 3343–3354). These methods typically consider differential gene expression between starting and target cell types and use knowledge of GRN architecture to rank TFs in order of influence in establishing the core GRN of the target cell. The availability of such methods to evaluate TFs *a priori* for their ability to influence cell fate is an important advantage of the “inside-out” approach, in contrast to the “outside-in” approach where currently there is no broadly applicable strategy to predict the instructive components of cell culture media that direct cell differentiation.

Although computational methods such as those referenced above have been used to predict OKSM can induce the PGRN, and also TFs capable of inducing a relatively small number of other cell fate transitions, the algorithms (and the input data they require) are far from perfect and frequently result in TF shortlists with little or no overlap. Consequently, the literature shows cell reprogramming is difficult to achieve, generally of low efficiency and/or reproducibility when at all successful. Experimental determination of useful transcription factor cocktails by methods used in the prior art is laborious and suffers from two related, major limitations.

Firstly, the conventional method of determining reprogramming cocktails is to screen candidate TFs in a large pool (or a set of smaller pools) and the minimal core TFs then identified through a process of eliminating each TF in turn, as is exemplified in Takahashi and Yamanaka’s (Cell 126:663-676) seminal experiment in which pools of 24 TFs were screened in a cell-based assay to identify OKSM in reprogramming fibroblasts to iPSCs. In a further demonstration of this stratagem, Zhou et al (Nature 455:627-633) introduced a pool of 9 TFs into mouse pancreas *in vivo*, and observed pancreatic exocrine cells were reprogrammed into insulin-expressing beta cells. In this case, Zhou et al were able to deduce, again by a process of elimination, that the effective reprogramming cocktail comprises PDX1, NGN3 and MAFA. In both examples, these pooled screening experiments were unable to discover functional combinations of TFs beyond the minimal core cocktail, since the presence of the core TFs in a pool will mask the effect of eliminating other productive TF cocktails from the pool. This limitation could be overcome if a set of TFs could be screened systematically to evaluate all combinations in separate experiments. However, the number of experiments needed is relatively large since there are 10,626 different combinations of four TFs in Yamanaka’s set of 24 candidate TFs. In addition, the number of

cells required to perform such a large screen would also be large, since the experiment should be performed in replicates and the efficiency of reprogramming e.g. human cells with OKSM was only 0.01% - 0.02% (Takahashi et al, Cell 131:861–872). There is therefore a need for scalable screening methods capable of systematically evaluating individual combinations of TFs, since “large scale screening of combinatorial TF cocktails remains
5 challenging” (Li and Hon, Frontiers in Bioengineering and Biotechnology (2021) vol 9 article 748942).

Secondly, the pooled screening approach does not allow the discovery of optimal protocols in which the sequence of TF addition is important for certain types of reprogramming. Recent
10 advances in single cell -omics technologies (e.g. RNA-Seq, ATAC-Seq) have allowed us to study cells during reprogramming at unprecedented resolution. These techniques have revealed cell reprogramming is a dynamic process, in which cells transition through a continuum of cell states characterised by different patterns of TF expression that accompany GRN remodelling. For example, optimal reprogramming of pancreatic progenitor cells to
15 insulin expressing beta cells using TFs PDX1, NGN3 and MAFA is only achieved by differential expression dynamics (Saxena et al, Nature communications 7 Article number 11247). There is therefore a need for screening methods capable of interrogating combinations of TFs expressed in sequences, in order to discover optimal cell reprogramming methods.

20 In our 2004 patent application WO2004/031369, we described a method for determining the factors required to differentiate pluripotent cells into more committed cell types utilising a splitting and pooling culture approach, together with labelling of multicell units in the form of beads, wherein the labels identify the conditions to which the beads have been exposed.

One of the major drawbacks of this system is that the culture conditions to which the cells
25 are exposed have to be determined by careful reading of the literature to formulate hypotheses that are tested by screening according to WO2004/031369. Moreover, the system is only applied to differentiation of cells as in 2004 only a handful of examples of reprogramming existed and these were mediated by a single TF or pairs of TF.

In the present invention, we provide an improvement of our 2004 process, in which cells are
30 optimally reprogrammed from one cell type to another. We combine the computational approach to prediction of transcription factors with an experimental approach which determines the effect of the administration and timing of administration of TFs to cells in culture.

Summary of the Invention

In a first aspect, the present invention provides a method for determining the conditions required for reprogramming of a first cell type to a second cell type, by screening TF combinations in which TFs are added to the cells sequentially.

- 5 Thus, the invention provides a method for determining the conditions required for reprogramming of a first cell type to a second cell type, comprising the steps of:

exposing a cell of a first cell type to a first TF;

removing the first TF and exposing the cell to a different TF;

optionally, repeating the foregoing step;

- 10 identifying one or more cells which displays one or more markers of a second cell type and determining the identity and sequence of TFs to which the cell has been exposed.

The split/pool techniques described in WO2004/031369 are adaptable to sequential screening of the effects of the exposure of cells to multiple individual TFs and combinations of TFs. Cell units in WO2004/031369 are envisaged therein as beads comprising multiple

- 15 cells which can be sorted together; in the present invention, it is envisaged that cell units such as those employed in WO2004/031369 may be used, or single cells may be used and individually sorted. Thus, a test unit may comprise multiple cells or a single cell.

Therefore, in a further aspect the invention comprises the steps of:

- a. separately exposing a first plurality of test units each comprising a cell to different
20 TFs or combinations of TFs and labelling the cells to indicate exposure to said TFs or combinations of TFs;
- b. pooling the plurality of test units, and subdividing the pool of test units to form a second plurality of test units;
- c. separately exposing the second plurality of test units to different TFs or combinations
25 of TFs;
- d. optionally iteratively repeating steps (a) to (c) as required;

- e. identifying the cell type of the cells in the test units and deconvoluting the labels to identify the identity and sequence of TFs to which the cells have been exposed;
- f. optionally, using the data from step (e) to inform the choice of TFs used in step (a), and repeating steps (a) to (e) as required;
- 5 g. identifying a cell of the second cell type, and deconvoluting the labels to thereby identify the identity and sequence of TFs required to reprogram a cell of the first cell type into a cell of the second cell type.

The present invention employs the approach described in WO2004031369 (see Fig. 1). In that approach, known as CombiCult®, a serial splitting and pooling of cells is used to expose
10 multiple cells to many different reagent combinations.

In one embodiment, the TFs used in step (a) and/or step (c) can be chosen on the basis of computational analysis of known cell reprogramming events. The data used for analysis can be based on previous performance of the method of the invention, or derived from the prior art. The computational selection is combined with the experimental selection provided by
15 the split/pool selection methodology to permit unprecedented precision in determining both the identity of the factors required for cell reprogramming, and the ideal timing of their administration.

The test units in the invention can take the form of beads, capsules, aggregates or the like comprising one or more cells. In some embodiments, distribution of cells to achieve an
20 average of one cell per unit may result in some units being devoid of cells. In certain embodiments, the test unit is a single cell.

Labelling of test units can be performed in a variety of ways, for example as described in WO2004/031369, including nucleic acid labels, radiofrequency encoded tags, fluorescent or optical tags, and spatial encoding of test units on a surface or matrix. Preferably, however,
25 labelling is performed by modifying the nucleic acid of each cell, which allows its exposure to TFs to be tracked precisely. Nucleic acid can be modified, for example, by retrovirus integration or genome editing techniques.

Pooling and splitting of test units, comprising multiple or single cells, permits the cells to be exposed to an array of different TFs and combinations of TFs at different times, and for both
30 the nature of the TFs and the timing of exposure to be monitored.

The split/pool protocol comprises the following steps:

- (a) providing a first set of groups of test units each comprising one or more cells, and exposing said groups to desired TFs or combinations of TFs;
- (b) pooling two or more of said groups to form at least one pool;
- (c) subdividing the pool to create a further set of groups of test units;
- 5 (d) exposing said further groups to desired TFs or combinations of TFs;
- (e) optionally, repeating steps (b) - (d) iteratively as required; and
- (f) assessing the effect on a given test unit of the TFs to which it has been exposed.

Re-use of TFs at different times during the split-pool sequences allows the effect of a TF at different times during the development of a cell to be assessed. For example, a pool may
10 be exposed to a combination of four TFs in a first culture, and then split. Different splits of the pool may be exposed to different combinations of TFs, which may or may not comprise any of the TFs used in the first culture. After pooling and resplitting, further different splits may be exposed to yet further combinations of TFs which may comprise TFs from the first culture or the second culture, as well as novel TFs. Thus, cells in different splits are
15 exposed to different TFs, as well as to the same TFs at different times.

In some embodiments, the TFs are added individually to cell units in sequential method steps. For example, sequential method steps in a procedure may comprise either the addition of single TFs or the addition of combinations of TFs; alternatively sequential method steps in a procedure only comprise the addition of individual TFs. Thus, a procedure may
20 combine method steps which add groups of TFs with steps which add individual TFs, or may only add individual TFs.

A "procedure", as used in this context, represents the performance of an entire method as represented by steps (a) to (f) above.

Although repetitive cycles of splitting and pooling may be used highly efficiently, in a similar
25 manner to combinatorial chemistry protocols, given the necessary processing power protocols involving at least two sequential splitting steps (without re-pooling) may be used. The disadvantage of such protocols is that they can quickly generate a very large number of separate samples, which have been handled differently. The advantage, however, is that each sample does not require laborious deconvolution, since the cell units therein have only

been exposed to one set of conditions. Accordingly, given suitable sample handling facilities, a splitting approach can yield rapid results.

Another advantage of the present invention is that TFs may be added sequentially and individually, or in defined groups, in a large number of combinations, as explained above.

- 5 This prevents the masking of the effect of a TF by dominant TFs which may be present in pooled combinations of TFs used in conventional experiments.

In a preferred embodiment, the TFs are individually added sequentially to cells, such that the effect of each TF can be assessed in isolation.

- 10 The method of the invention allows thousands or millions of TFs and combinations of TFs to be tested, in a multiplexed high-throughput assay, to determine the conditions necessary to achieve the desired result with respect to any cellular process.

- 15 In contrast to methods in which multiple TFs are added together, and then individually removed to identify the effective TFs, the method of the invention allows testing of sequential addition of TFs and therefore the determination of the action of each TF, when added in a plurality of sequential combinations, as well as the testing of multiple combinations amongst groups TFs. This permits the elucidation of the separate steps in a pathway of conversion from one cell type to a second cell type, rather than merely the overall conversion.

- 20 In some embodiments, the invention provides a method for identifying a gene which influences a cellular process, comprising the steps of: a) determining the effect of one or more TFs or combinations of TFs on a test unit, in accordance with the foregoing aspect of the invention; b) analysing gene expression in said cell units when exposed to said TFs; and c) identifying genes which are differentially expressed under desired culture conditions.

- 25 Advantageously, the TFs used cause a change in the cellular process; these TFs are selected for the production of cells in which gene expression is analysed. Gene expression may conveniently be analysed using any comparative expression monitoring technology including PCR-based techniques such as RT-PCR, Serial Analysis of Gene Expression (SAGE), RNA sequencing (RNA seq) including single-cell RNA seq, in-situ hybridisation including single-cell hybridisation, or array technology such as is widely available from suppliers such as Affymetrix. In another aspect, the invention provides a method for
30 producing a nucleic acid which encodes a gene product which influences a cellular process, comprising identifying a gene as above, and producing at least the coding region of said gene by nucleic acid synthesis or biological replication.

In a further aspect, there is provided a method for inducing a cellular process in a cell, comprising the steps of: a) identifying one or more genes which are differentially expressed in association with a cellular process in accordance with the invention; and b) modulating the expression of said one or more genes in the cell. The expression of the genes in the cell can be modulated by, for example, transfecting or otherwise transferring the gene into the cell such that it is overexpressed in a transient or permanent manner. Alternatively, the expression of the endogenous gene may be altered, such as by targeted enhancer insertion or the administration of exogenous agents which cause an increase (e.g. gratuitous inducers) or decrease (e.g. antisense, RNAi, transcription factors) in the expression of the gene. Moreover, the product of the gene may itself be administered to or introduced into the cell to achieve an increase in its activity. Furthermore, agents which increase or decrease the activity of the gene product (e.g. competitive and non-competitive inhibitors, drugs, pharmaceuticals) can be administered to the cell. In a still further aspect, the invention provides a method for identifying the state of a cellular process in a cell, comprising the steps of: a) identifying one or more genes which are differentially expressed in association with a cellular process as set forth above; and b) detecting the modulation of expression of said one or more genes in a cell, thereby determining the state of the cellular process in said cell. Advantageously, the genes employed in this analysis encode cellular markers, which may be detected for instance by immunoassay. Alternatively, the gene products may be enzymes that can be assayed for activity with fluorometric, colorimetric, radiometric, or other methodologies.

The invention further provides a method for regulating a cellular process, comprising the steps of: a) determining the effect of one or more TFs on a cell, in accordance with the foregoing aspect of the invention; b) exposing a cell to TFs which effect a change in the cellular process; and c) isolating the desired cell.

Accordingly, the invention provides for a method for producing a differentiated cell from a second, different differentiated cell type, in accordance with the foregoing aspect of the invention.

There is also provided a method for identifying a TF which is capable of inducing a cellular process, comprising the steps of: a) determining the effect of one or more TFs on a cell, in accordance with the foregoing aspect of the invention; and b) identifying those TF(s) which induce the desired cellular process in the cells. TFs identified in accordance with the invention may be expressed or synthesised by conventional chemical, biochemical or other

techniques, and used in methods for regulating particular cellular processes in cells for example as described herein.

In a further aspect, the invention relates to methods for developing and using a database of cell gene expression data correlated to exposure to TFs, which can provide improved TF

- 5 selection in step (a) of the method of first aspect of the invention. Using machine learning and inputting experimental data from the method of the present invention, the database can be improved, further optimising TF selection and therefore the performance of the method of the invention.

Brief Description of the Figures

- 10 Figure 1 is a schematic of the operation of the CombiCult® split/pool selection method.

Figure 2 is a schematic of the workflow for the CombiCult® method of the present invention.

Figure 3 is a schematic of genetic barcode labelling of cells in the method of the present invention.

- 15 Figure 4 illustrates the screening matrix for 10 transcription factors used in the generation of insulin-producing cells by the method of the invention.

Figure 5 is a photomicrograph showing insulin expression and cd49a expression in differentiated cells obtained by the method of the invention.

- Figure 6. Overview of negative and positive controls. In both the negative and positive control setups, cPPs are placed onto the beads using the same media. Media A consists of
20 DMEM high glucose supplemented with Asc (50µg), 1% B27, EGF (50ng), FGF-7 (50ng), and RA (50nM). The purpose here is to guide the differentiation process towards pancreatic progenitors, particularly considering the early stage of cPPs. Following this, Media B is a combination of 5% KOSR, T3 (1µM), and RA (25 nM), added to DMEM high glucose to foster endocrine progenitors. Finally, Media C comprises DMEM low glucose enriched with
25 10% FBS, T3 (1µM), and Alk5i (10µM). In the negative control, no transfection occurs, while the positive control involves sequential transfection starting with Pdx1, followed by Ngn3 and MafA.

Figure 7A is a graphical illustration of the combinations of transcription factors computationally analysed to determine the frequency of transcription factor addition to hits identified in Figure 5.

Figure 7B shows the results illustrated in Figure 7A in a bar graph.

- 5 Figure 8 illustrates selection of optimal transcription factors by the method of the invention, and also illustrates the deriving of timing information for transcription factor addition.

Detailed Description of the Invention

Unless defined otherwise, all technical and scientific terms used herein have the same
10 meaning as commonly understood by those of ordinary skill in the art, such as in the arts of peptide chemistry, cell culture and phage display, nucleic acid chemistry and biochemistry. Standard techniques are used for molecular biology, genetic and biochemical methods (see Sambrook et al., Molecular Cloning: A Laboratory Manual, 3rd ed., 2001, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; Ausubel et al., Short Protocols in Molecular
15 Biology (1999) 4th ed., John Wiley & Sons, Inc.; Patil and Sivaram, A Complete Guide to Gene Cloning: From Basic to Advanced, Springer: ISBN978-3-030-96853-3, 28 April 2023; Jaroszewicz et al., Phage display and other peptide display technologies, FEMS Microbiology Reviews, fuab052, 46, 2022, 1–25, as well as resources on Addgene.org), which are incorporated herein by reference.

20 **Culture Conditions** As used herein, the term "culture conditions" refers to the environment which cells are placed in or are exposed to in order to promote growth or differentiation of said cells. Thus, the term refers to the medium, temperature, atmospheric conditions, substrate, stirring conditions and the like which may affect the growth and/or differentiation of cells. More particularly, the term refers to specific reagents which may be incorporated into
25 culture media and which may influence the growth and/or differentiation of cells. In a preferred embodiment, the culture conditions are the reagent or combination of reagents to which a cell is exposed.

Cell A cell, as referred to herein, is defined as the smallest structural unit of an organism that is capable of independent functioning, or a single-celled organism, consisting of one or more
30 nuclei, cytoplasm, and various organelles, all surrounded by a semipermeable cell membrane or cell wall. The cell may be prokaryotic, eukaryotic or archaebacterial. For

example, the cell may be a eukaryotic cell. Mammalian cells are preferred, especially human cells. Cells may be natural or modified, such as by genetic manipulation or passaging in culture, to achieve desired properties. A **stem cell** is defined in more detail below, and is a totipotent, pluripotent or multipotent cell capable of giving rise to more than one differentiated cell type. Stem cells may be differentiated *in vitro* to give rise to differentiated cells, which may themselves be multipotent, or may be terminally differentiated. Cells differentiated *in vitro* are cells which have been created artificially by exposing stem cells to one or more agents which promote cell differentiation.

Cellular process A cellular process is any characteristic, function, process, event, cause or effect, intracellular or extracellular, which occurs or is observed or which can be attributed to a cell. Examples of cellular processes include, but are not limited to, viability, senescence, death, pluripotency, morphology, signalling, binding, recognition, molecule production or destruction (degradation), mutation, protein folding, transcription, translation, catalysis, synaptic transmission, vesicular transport, organelle function, cell cycle, metabolism, proliferation, division, differentiation, phenotype, genotype, gene expression, or the control of these processes.

Cell unit A group of cells, which may be a group of one. Pools of cell units may be sorted, subdivided and handled without substantially dissociating the cell units themselves, such that the cell unit behaves as a colony of cells and each cell in the cell unit is exposed to the same culture conditions. For example, a cell unit may comprise a bead to which is adhered a group of cells.

Totipotent A totipotent cell is a cell with the potential to differentiate into any type of somatic or germ cell found in the organism. Thus, any desired cell may be derived, by some means, from a totipotent cell.

Pluripotent A pluripotent cell is a cell which may differentiate into more than one, but not all, cell types.

Label A label or **tag**, as used herein, is a means to identify a cell unit and/or determine a culture condition, or a sequence of culture conditions, to which the cell unit has been exposed. Thus, a label may be a group of labels, each added at a specific culturing step; or a label added at the beginning or the experiment which is modified according to, or tracked during, the culturing steps to which the cell unit is exposed; or simply a positional reference, which allows the culturing steps used to be deduced. A label or tag may also be a device that reports or records the location or the identity of a cell unit at any one time, or assigns a

unique identifier to the cell unit. Examples of labels or tags are molecules of unique sequence, structure or mass; or fluorescent molecules or objects such as beads; or radiofrequency and other transponders; or objects with unique markings or shapes. Preferably, a label is a genetic barcode, as described further herein.

- 5 **Exposure to culture conditions or reagents** A cell is exposed to culture conditions or reagents when it is placed in contact with a medium, or grown under conditions which affect one or more cellular process(es) such as the growth, differentiation, or metabolic state of the cell. Thus, if the culture conditions comprise culturing the cell in a medium in the presence of a reagent, the cell is placed in the medium with a reagent present for a sufficient period of
10 time for it to have an effect. Likewise, if the conditions are temperature conditions, the cells are cultured at the desired temperature.

- Pooling** The pooling of one or more groups of cell units involves the admixture of the groups to create a single group or pool which comprises cell units of more than one background, that is, that have been exposed to more than one different sets of culture conditions. A pool
15 may be subdivided further into groups, either randomly or non- randomly; such groups are not themselves "pools" for the present purposes, but may themselves be pooled by combination, for example after exposure to different sets of culture conditions.

- Proliferation** Cell growth and cell proliferation are used interchangeably herein to denote multiplication of cell numbers without differentiation into different cell types or lineages. In
20 other words, the terms denote increase of viable cell numbers. Preferably proliferation is not accompanied by appreciable changes in phenotype or genotype.

- Differentiation** Cell differentiation is the development, from a cell type, of a different cell type. For example, a bipotent, pluripotent or totipotent cell may differentiate into a neural cell. Differentiation may be accompanied by proliferation, or may be independent thereof. The
25 term 'differentiation' generally refers to the acquisition of a phenotype of a mature cell type from a less developmentally defined cell type, e.g. a neuron, or a lymphocyte, but does not preclude transdifferentiation, whereby one mature cell type may convert to another mature cell type e.g. a neuron to a lymphocyte.

- Differentiation state** The differentiation state of a cell is the level to which a cell has
30 differentiated along a particular pathway or lineage.

State of a cellular process The state of a cellular process refers to whether a cellular process is occurring or not and in complex cellular processes can denote a particular step or

stage in that cellular process. For example, a cellular differentiation pathway in a cell may be inactive or may have been induced and may comprise a number of discrete steps or components such as signalling events characterised by the presence of a characteristic set of enzymes or intermediates.

- 5 **Gene** A gene is a nucleic acid which encodes a gene product, be it a polypeptide or an RNA gene product. As used herein, a gene includes at least the coding sequence which encodes the gene product; it may, optionally, include one or more regulatory regions necessary for the transcription and/or translation of the coding sequence.

- 10 **Gene Product** A gene product is typically a protein encoded by a gene in the conventional manner. However, the term also encompasses non-polypeptide gene products, such as ribonucleic acids, which are encoded by the gene.

- 15 **Nucleic acid synthesis** Nucleic acids may be synthesised according to any available technique. Preferably, nucleic acid synthesis is automated. Moreover, nucleic acids may be produced by biological replication, such as by cloning and replication in bacterial or eukaryotic cells, according to procedures known in the art.

- 20 **Differential Expression** Genes which are expressed at different levels in response to cell culture conditions can be identified by gene expression analysis, such as on a gene array, by sequencing such as RNA-seq, or by any methods known in the art. Genes which are differentially expressed display a greater or lesser quantity of mRNA or gene product in the cell under the test conditions than under alternative conditions, relative to overall gene expression levels.

- 25 **Transfection** Genes may be transfected into cells by any appropriate means. The term is used herein to signify conventional transfection, for example using calcium phosphate, but also to include other techniques for transferring nucleic acids into a cell, including transformation, viral transduction, electroporation and the like.

Modulation The term modulation is used to signify an increase and/or decrease in the parameter being modulated. Thus, modulation of gene expression includes both increasing gene expression and decreasing gene expression.

- 30 **Stem Cells** Stem cells are described in detail in *Stem Cells: Scientific Progress and Future Research Directions*. Department of Health and Human Services. June 2001.

<http://www.nih.gov/news/stemcell/scireport.htm>. The contents of the report are herein incorporated by reference.

- Stem cells are cells that are capable of differentiating to form at least one and sometimes many specialised or differentiated cell types. The repertoire of the different cells that can be
- 5 formed from stem cells is thought to be exhaustive; that is to say it includes all the different cell types that make up the organism. Stem cells are present throughout the lifetime of an organism, from the early embryo where they are relatively abundant, to the adult where they are relatively rare. Stem cells present in many tissues of adult animals are important in normal tissue repair and homeostasis. The existence of these cells has raised the possibility
- 10 that they could provide a means of generating specialised functional cells that can be transplanted into humans and replace dead or non-functioning cells in diseased tissues. The list of diseases for which this may provide therapies includes Parkinson's disease, diabetes, spinal cord injury, stroke, chronic heart disease, end-stage kidney disease, liver failure and cancer.
- 15 Different stem cells have differing potential to form various cell types: spermatogonial stem cells are unipotent as they naturally produce only spermatozoa, whereas haematopoietic stem cells are multipotent, and embryonic stem cells are thought to be able to give rise to all cell types and are said to be totipotent or pluripotent. To date three types of mammalian pluripotent stem cell have been isolated. These cells can give rise to cell types that are
- 20 normally derived from all three germ layers of the embryo (endoderm, mesoderm and ectoderm). The three types of stem cell are: embryonal carcinoma (EC) cells, derived from testicular tumours; embryonic stem (ES) cells, derived from the pre-implantation embryo (normally the blastocyst); and embryonic germ (EG) cells derived from the post-implantation embryo (normally cells of the foetus destined to become part of the gonads).
- 25 Differentiated cell types are capable of changing phenotype. This phenomenon, termed transdifferentiation, is the conversion of one differentiated cell type to another, with or without an intervening cell division. Differentiation can sometimes be reversed or altered. In vitro protocols are now available in which cell lines can be induced to transdifferentiate. Moreover, specialised cell types can de- differentiate to yield stem-like cells with the
- 30 potential to differentiate into further cell types.

Many systems that have been devised for the differentiation of stem cells in vitro are complex multi-stage procedures, in which the precise nature of the various steps, as well as the chronology of the various steps, are important. For instance, Lee et al (2000, Nature Biotechnology, vol. 18, p 675-679) used a five stage protocol to derive dopaminergic

neurons from mouse ES cells: 1) undifferentiated ES cells were expanded on gelatin-coated tissue culture surface in ES cell medium in the presence of LIF ; 2) embryoid bodies were generated in suspension cultures for 4 days in ES cell medium; 3) nestin-positive cells were selected from embryoid bodies in ITSFn medium for 8 days after plating on tissue culture surface; 4) nestin-positive cells were expanded for 6 days in N2 medium containing bFGF/laminin; 5) finally the expanded neuronal precursor cells were induced to differentiate by withdrawing bFGF from N2 medium containing laminin. In a second example of serial cell culture, Bonner-Weir et al. [Proc. Natl. Acad. Sci. (2000) 97: 7999-8004] derived insulin producing cells from human pancreatic ductal cells by: 1) selecting ductal cells over islet cells by selective adhesion on a solid surface in the presence of serum for 2-4 days; 2) subsequently withdrawing serum and adding keratinocyte growth factor to select for ductal epithelial cells over fibroblasts for 5-10 days; and 3) overlaying the cells with the extracellular matrix preparation 'Matrigel' for 3- 6 weeks. In a further example of serial cell culture, Lumelsky et al. [Science (2001) 292: 1389- 1394] derived insulin secreting cells by directed differentiation of mouse embryonic stem (ES) cells by: 1) expanding ES cells in the presence of LIF for 2-3 days; 2) generating embryoid bodies in the absence of LIF over 4 days; 3) selecting nestin-positive cells using ITSFn medium for 6-7 days; 4) expanding pancreatic endocrine precursors in N2 medium containing 827 media supplement and bFGF for 6 days; and 5) inducing differentiation to insulin secreting cells by withdrawing bFGF and adding nicotinamide.

However it is not only the sequence and duration of the various steps or the series of addition of different factors that is important in the determination of cell differentiation. As embryonic development is regulated by the action of gradients of signalling factors that impart positional information, it is to be expected that the concentration of a single signalling factor, and also the relative concentration of two or more factors, will be important in specifying the fate of a cell population in vitro and in vivo. Factor concentrations vary during development and stem cells respond differently to different concentrations of the same molecule. For instance, stem cells isolated from the CNS of late stage embryos respond differently to different concentrations of EGF: low concentrations of EGF result in a signal to proliferate, while higher concentrations of EGF result in proliferation and differentiation to astrocytes. Many of the factors that have been found to influence self-renewal and differentiation of stem cells in vitro are naturally-occurring molecules. This is to be expected, as differentiation is induced and controlled by signalling molecules and receptors that act along signal transduction pathways. However, by the same token, it is likely that many synthetic compounds will have an effect on stem cell differentiation. Such synthetic compounds that have high probability of interacting with cellular targets within signalling and

signal transduction pathways (so called drugable targets) are routinely synthesised, for instance for drug screening by pharmaceutical companies. Once known, these compounds can be used to direct the differentiation of stem cells ex vivo, or can be administered in vivo in which case they would act on resident stem cells in the target organ of a patient.

- 5 **Common variables in tissue culture** In developing conditions for the successful culture of a particular cell type, or in order to achieve or modulate a cellular process, it is often important to consider a variety of factors. One important factor is the decision of whether to propagate the cells in suspension or as a monolayer attached to a substrate. Most cells prefer to adhere to a substrate although some, including transformed cells, haematopoietic
10 cells, and cells from ascites, can be propagated in suspension.

Assuming the culture is of adherent cells, an important factor is the choice of adhesion substrate. Most laboratories use disposable plastics as substrates for tissue culture. The plastics that have been used include polystyrene (the most common type), polyethylene, polycarbonate, Perspex, PVC, Teflon, cellophane and cellulose acetate. It is likely that any
15 plastic can be used, but many of these will need to be treated to make them wettable and suitable for cell attachment. Furthermore it is very likely that any suitably prepared solid substrate can be used to provide a support for cells, and the substrates that have been used to date include glass (e.g. alum-borosilicate and soda-lime glasses), rubber, synthetic fibres, polymerised dextrans, metal (e.g. stainless steel and titanium) and others. Some cell types,
20 such as bronchial epithelium, vascular endothelium, skeletal muscle and neurons require the growth substrate to be coated with biological products, usually extracellular matrix materials such as fibronectin, collagen, laminin, polylysine or others. The growth substrate and the method of application (wet or dry coating, or gelling) can have an effect on cellular processes such as the growth and differentiation characteristics of cells, and these must be
25 determined empirically as discussed above. Probably the most obviously important of the variables in cell culture is the choice of culture medium and supplements such as serum. These provide an aqueous compartment for cell growth, complete with nutrients and various factors, some of which have been listed above, others of which are poorly defined. Some of these factors are essential for adhesion, others for conveying information (e.g. hormones, mitogens, cytokines) and others as detoxificants. Commonly used media include RPMI
30 1640, MEM/Hank's salts, MEM/Earle's salts, F12, DMEM/F12, L15, MCDB 153, and others. The various media can differ widely in their constituents - some of the common differences include sodium bicarbonate concentration, concentration of divalent ions such as Ca and Mg, buffer composition, antibiotics, trace elements, nucleosides, polypeptides, synthetic
35 compounds, drugs, etc. It is well known that different media are selective, meaning they

promote the growth of only some cell types. Media supplements- such as serum, pituitary brain and other extracts, are often essential for the growth of cells in culture, and in addition are frequently responsible for determining the phenotype of cells in culture, i.e. they are capable of determining cell survival or directing differentiation. The role of supplements in cell processes such as differentiation is complex and depends on their concentration, the time point at which they are added to the culture, the cell type and medium used. The undefined nature of these supplements, and their potential to affect the cell phenotype, have motivated the development of serum-free media. As with all media, their development has come about largely by trial and error, as has been discussed above. The gas phase of the tissue culture is also important and its composition and volume which should be used can depend on the type of medium used, the amount of buffering required, whether the culture vessel is open or sealed, and whether a particular cellular process needs to be modulated. Common variables include concentration of carbon dioxide and oxygen. Other conditions important to tissue culture include the choice of culture vessel, amount of headspace, inoculation density, temperature, frequency of media changes, treatment with enzymes, rate and mode of agitation or stirring. Varying the cell culture conditions is therefore a method of achieving a desired cellular process. One aspect of the invention recognises that variation of the cell culture conditions in a serial manner can be a highly effective method for achieving a cellular effect. In various applications, for instance in studies of cell differentiation, it will often be the case that a specific series of different tissue culture conditions are required to effect a cellular process. The different conditions may include additions or withdrawals to/from the media or the change of media at specific time points. Such a set of conditions, examples of which are given below, are commonly developed by trial and error as has been discussed above.

Formation of test units An aspect of the present invention is that, in some embodiments, groups of cells (cell colonies) can be grown in cell culture under various conditions and that the colony can largely maintain its integrity under various conditions, when disturbed, and when mixed with other colonies. Such groups or colonies are referred to herein as test units. In some instances, test units are a single cell, in which case it is not necessary to form cell colonies.

Formation of test units may be achieved, by way of illustration, by growing cells as adherent cultures on solid substrates such as carriers. If cell proliferation occurs after seeding on the carriers, the daughter cells will attach on the same carrier and form part of the same colony. In general, live adherent cells do not readily dissociate from their growth substrate, and so

the integrity of the cell colony persists despite any mechanical manipulation of the carrier, agitation of the culture medium, or transfer into another tissue culture system.

Similarly, if at any time multiple carriers are placed in the same vessel (e.g. beads are pooled) then there will be no substantial transfer of cells from one bead to another. One
5 advantage of growing cells in units or colonies is that if a unit is placed serially in a set of different tissue culture media, then all the cells comprising the colony will have been exposed to the same series of culture conditions, in the same order and for the same period of time. Growing cells in units that are not necessarily themselves adherent to the tissue culture vessel has the advantage that individual colonies can be removed at will and
10 transferred to a different culture vessel. One of the advantages of this method is that tissue culture can be miniaturised: relatively few cells are required to colonise a microcarrier bead (see below) compared to even the smallest tissue culture flask. A further advantage of growing cell units formed on carriers is that cell culture can be scaled up. Growth of stem cells on carriers offers a way of scaling up production to provide enough material for stem
15 cell therapy. Equally, differentiation of stem cells on carriers offers a way of scaling up production as differentiation proceeds, eventually providing enough material for cell replacement therapies. Scale up of such cell cultures requires at least 50g (dry weight) of microcarrier, preferably, 100g, 500g, 1kg, 10kg or more. Another important advantage of forming cell units on solid substrates is that the substrate - and therefore the attached cells
20 by reason of association - can be labelled by various means. Glass spheres of 3mm and 5mm have been widely used as cell adhesion substrates, particularly in glass bead bioreactors (e.g. such as manufactured by Meredos GmbH) used for the scale-up of cell cultures. These beads are typically used in packed beds rather than batch culture, to avoid mechanical damage to the adherent cells. In contrast, when cells are grown on smaller
25 carriers they can be treated as a suspension culture. A common method of growing cells on small carriers is referred to as microcarrier cell culture (see 'Microcarrier cell culture, Principles and Methods', Edition AA, available from Amersham Biosciences (18-1140-62); herein incorporated in its entirety by reference). Microcarrier cultures are used commercially for antibody and interferon production in fermenters of up to 4000 litres. A variety of
30 microcarriers is available, ranging in shape and size and made of different materials. Microcarrier beads made of polystyrene (Biosilon, Nunc), glass (Bioglass, Solohill Eng), collagen (Biospheres, Solohill Eng), DEAE sephadex (Cytodex-1, Pharmacia), dextran (Dormacell, Pfeifer & Langen), cellulose (DE-53, Whatman), gelatin (Gelibead, Hazelton Lab), and DEAE dextran (Microdex, Dextran Prod.) amongst others are commercially
35 available. These carriers are well characterised in terms of the specific gravity of the beads, the diameter and the surface area available for cell growth. In addition a number of porous

(micro) carriers are available with greatly increased surface area for cell growth. A further characteristic of these porous carriers is that they are suitable for growth of both anchorage dependent cells, as well as for suspension cells which are carried by entrapment in the network of open, interconnecting pores. Porous carriers are available in materials such as
5 gelatin (Cultispher G, HyClone), cellulose (Cytocell, Pharmacia), polyethylene (Cytoline 1 and 2, Pharmacia), silicone rubber (Immobilisil, Ashby Scientific), collagen (Microsphere, Cellex Biosciences), and glass (Siran, Schott Glassware). These carriers are variously suited to stirred, fluidised or fixed bed culture systems.

As the physical properties of carriers are well known it is easy to calculate the number of
10 carriers used in an experiment. Some of the carriers described and many besides are available as dried products which can be accurately weighed, and subsequently prepared by swelling in liquid medium. In addition the number of cells used to inoculate a microcarrier culture can be worked out and varied. For instance, a culture of Cytodex 3 (2 g/litre) inoculated at 6 cells per bead will give a culture containing 8 million 42 microcarriers, on
15 which 48 million cells/litre are grown at a density of 5×10^5 cells/cm.

Harvesting of cells grown on microcarriers, or liberation of labels from microcarriers (see below), can be achieved by enzymatic detachment of cells, and/or by digestion of the carrier where applicable: gelatin carriers can be solubilised by trypsin and/or EDTA, collagen carriers using collagenase and dextran carriers using dextranase.

20 In addition to solid or porous microcarriers, cells may be grouped by immurement, i.e. confined within a medium permeable barrier. Membrane culture systems have been developed where a permeable dialysis membrane retains a group of cells, but allows the culture medium and its constituents to exchange freely with the inner and outer compartments. Cell culture in hollow fibre cartridges has also been developed, and a
25 multitude of fibres and even turn-key systems are commercially available (e.g. from Amicon, Cellex Biosciences). Cell encapsulation in semi-solid matrices has also been developed, where cells are immobilised by adsorption, covalent bonding, crosslinking or entrapment in a polymeric matrix. Materials that have been used include gelatin, polylysine, alginate and agarose. A typical protocol, is to mix 5% agarose at 40°C with a suspension of cells in their
30 normal growth medium, to emulsify the mixture using an equal volume of paraffin oil and to cool in an ice bath, producing spheres of 80-200µm diameter. These spheres can be separated from the oil and transferred to medium in a tissue culture vessel.

Cell entrapment is a simple method for the immobilisation of groups of cells, akin to the use of microcarriers or porous substrates. A simple technique is to enmesh cells in cellulose

fibres such as DEAE, TLC, QAE, TEAE (all available from Sigma). Other more sophisticated devices are ceramic cartridges which are suitable for suspension cells, as in the Opticel culture system (Cellex Biosciences).

One skilled in the art will envisage, in addition to the above methods of creating cell units,
5 other methods of creating groupings of cells including forming 3D cultures of cells such as neural spheres or embryoid bodies, or using tissues and indeed whole organisms such as *Drosophila* or *C. elegans*.

Test units, or the substrates of which they are comprised, can be associated with a particular factor including, but without limitation, proteins, nucleic acids or other chemicals such as
10 drugs. Pre-conditioning of substrates can be achieved in many ways, for instance simply by incubating the substrate with the factor of interest, or by attaching the factor covalently or non-covalently to the substrate. Soluble factors can be incorporated into dry materials by impregnation. This technique relies on the rapid ingress of liquid, carrying soluble factors, into dry porous material that concomitantly becomes swollen and ready for use. Solid factors
15 can be incorporated for example by mixing the factor in fibrinogen with thrombin solution, at which point a fibrin clot containing the factor is formed. Multiple other ways can be envisaged of associating factor(s) with a cell group, in addition to impregnating, entrapping or encapsulating the factor together with cells.

A method for associating a cell group with a number of different factors is to pre-form
20 cocktails of factors which are subsequently associated with a particular cell group. A second method would be to serially condition cell groups in a number of factors. Using dry formulations of cell group growth substrates as an example, this method would involve firstly partially swelling the substrate in a solution containing a first factor and subsequently further swelling the same in a solution containing a second factor, resulting in a substrate to which
25 both factors have become associated. By devising a systematic protocol of associating cell groups with different combinations of different factors it will be possible to sample the effect on the cell group of any combination of a set of factors. Regardless of the method used to condition cell units with factors, the factors are taken up by cells that comprise that cell unit. Factors leaking into the growth medium are diluted to such an extent that their concentration
30 falls below physiologically relevant limits and they have no effect on any additional cell group to which they are exposed. The diffusion of the factor out of e.g. the substrate forming part of the cell unit is governed by parameters such as the nature and dimensions of the material, the mean pore diameter, and the molecular weight and concentration of the factor. To calibrate the process if necessary, factor release can be measured by physical assays such

as HPLC analysis or release of labelled factor into the medium, or by biological assays such as the dorsal root ganglion outgrowth bioassay for neurotrophic factors.

Combinatorial serial culture of cells Split-pool cell culture Forming test units

(particularly microscopic test units) is furthermore useful for sampling multiple tissue culture

5 conditions as each cell unit constitutes an easily handled unit that can be exposed to a variety of cell culture conditions. For simplicity, in this discussion we will assume that cell groupings are produced by growing cells in microcarrier culture, and the terms test unit, cell unit, cell group, single cell, colony and bead are used interchangeably. However, the methods described are equally applicable to any test unit, for instance those described
10 above.

A particularly efficient method for sampling a large number of cell culture conditions is referred to as Combinatorial Cell Culture or split-pool cell culture (Fig 1) and in one embodiment involves the serial subdividing and combining of groups of cell units in order to sample multiple combinations of cell culture conditions. In one aspect of the invention the
15 method operates by taking an initial starter culture (or different starter cultures) of cell units divided into X number of aliquots each containing multiple beads

(groups/colonies/carriers/cells) which are grown separately under different culture conditions. Following cell culture for a given time, the cell units can be pooled by combining and mixing the beads from the different aliquots. This pool can be split again into X2 number
20 of aliquots, each of which is cultured under different conditions for a period of time, and subsequently also pooled. This iterative procedure of splitting, culturing and pooling (or pooling, splitting and culturing; depending on where one enters the cycle) cell units allows systematic sampling of many different combinations of cell culture conditions.

The complexity of the experiment, or in other words the number of different combinations of
25 cell culture conditions tested, is equal to the product of the number of different conditions ($X_1 \times X_2 \times \dots \times X_n$) sampled at each round. Note that the step of pooling all the cell units prior to a subsequent split can be optional - a step in which a limited number of cell units are pooled can have the same effect.

The invention therefore embodies a number of related methods of systematically sampling
30 multiple combinations of cell culture conditions where groups of test units are handled in bulk. Regardless of the precise manner in which a diversity of cell culture conditions is sampled by this means the procedure is efficient because multiple cell units can share a single vessel, where they are cultured under identical conditions, and it can be carried out

using only a few culture vessels at any one time (the number of culture vessels in use is equal to the number of split samples).

- In many respects the principle of this procedure resembles that of split synthesis of large chemical libraries (known as combinatorial chemistry), which samples all possible combinations of linkage between chemical building block groups (see for example: Combinatorial Chemistry, Oxford University Press (2000), Hicham Fenniri (Editor)). Split-pool cell culture can be repeated over any number of rounds, and any number of conditions can be sampled at each round. So long as the number of test units (cells or colonised beads in this example) is greater than or equal to the number of different conditions sampled over all rounds, and assuming that the splitting of cell units occurs totally randomly, it is expected that there will be at least one cell unit that has been cultured according to each of the various combinations of culture conditions sampled by the experiment. This procedure can be used to sample growth or differentiation conditions for any cell type, or the efficiency of biomolecule production (e.g. production of erythropoietin or interferon) by any cell type. Because the procedure is iterative, it is ideally suited to testing multistep tissue culture protocols - for instance those described above in connection with stem cell differentiation. The variables which can be sampled using this technique include cell type, cell grouping (e.g. microcarrier culture, cell encapsulation, whole organism), growth substrate (e.g. fibronectin on microcarrier), duration of cell culture round, temperature, different culture media (including different concentrations of constituents), growth factors, conditioned media, co-culture with various cell types (e.g. feeder cells), animal or plant extracts, drugs, other synthetic chemicals, infection with viruses (incl. transgenic viruses), addition of transgenes, addition of antisense or anti-gene molecules (e.g. RNAi, triple helix), sensory inputs (in the case of organisms), electrical, light, or red-ox stimuli and others.
- Split-split cell culture** The purpose of performing split-pool processes on cell units is to systematically expose these to a pre-defined combination of conditions. The person skilled in the art will conceive of many different means of achieving this outcome. In addition to split-pool processes and variations thereof, it is worthwhile briefly discussing split-split processes. A split-split process involves subdividing a group of cell units at least twice, without intervening pooling of cell units. If split-split processes are used over a large number of rounds, the number of separate samples that are generated increases exponentially. In this case it is important to employ some level of automation, for example the use of a robotic platform and sophisticated sample tracking systems. The advantage of split-split steps is that (since cell units are not combined) it is possible to segregate lineages of the various cell units based on their cell culture history. Consequently split-split steps can be used to deduce

if a particular cell culture condition is responsible for any given cellular process and therefore used to deduce the culture history of cell units.

Predetermined protocols The splitting and/or pooling of cell units may be accomplished totally randomly or may follow a predetermined protocol. Where cell units are split and/or

5 pooled randomly, the segregation of a given cell unit into any group is not predetermined or prejudiced in any way. In order to result in a high probability that at least one cell unit has been exposed to each of the possible combinations of cell culture conditions, it is advantageous to employ a larger number of cell units than the total number of combinations of cell culture conditions that are being tested. Under certain circumstances it is therefore

10 advantageous to split and/or pool cell units according to a predetermined protocol, the overall effect being that adventitious duplications or omissions of combinations are prevented. Predetermined handling of cell units can be optionally planned in advance and logged on a spreadsheet or computer programme, and splitting and/or pooling operations executed using automated protocols, for instance robotics. Labelling of cell units (see below)

15 can be by any of a number of means, for instance labelling by RFID, optical tagging or spatial encoding. Robotic devices capable of determining the identity of a sample, and therefore partitioning the samples according to a predetermined protocol, have been described (see 'Combinatorial Chemistry, A practical Approach', Oxford University Press (2000), Ed H. Fenniri). Alternatively, standard laboratory liquid handling and/or tissue culture

20 robotics (for example such as manufactured by: Beckman Coulter Inc, Fullerton, CA; The Automation Partnership, Royston, UK) is capable of spatially encoding the identity of multiple samples and of adding, removing or translocating these according to pre-programmed protocols.

Analysis and/or separation of cell units Following each round of cell culture, or after a

25 defined number of rounds, the cell units can be studied to observe any given cellular process that may have been affected by the tissue culture conditions. The examples below are illustrative and not intended to limit the scope of the invention.

Following each round of cell culture, or after a defined number of rounds, the cell units can be assayed to determine whether there are members displaying increased cell proliferation.

30 This can be achieved by a variety of techniques, for instance by visual inspection of the cell units under a microscope, or by quantitating a marker product characteristic of the cell. This may be an endogenous marker such as a particular DNA sequence, or a cell protein which can be detected by a ligand or antibody. Alternatively an exogenous marker, such as green fluorescent protein (GFP), can be introduced into the cell units being assayed to provide a

specific readout of (living) cells. Live cells can be visualised using a variety of vital stains, or conversely dead cells can be labelled using a variety of methods, for instance using propidium iodide. Furthermore the labelled cell units can be separated from unlabelled ones by a variety of techniques, both manual and automated, including affinity purification

5 ('panning'), or by fluorescence activated cell sorting (FACS) or broadly similar techniques (Fig. 17). Depending on the application it may be possible to use standard laboratory equipment, or it may be advantageous to use specialised instrumentation. For instance, certain analysis and sorting instruments (e.g. see Union Biometrica Inc., Somerville MA, USA) have flow cell diameters of up to one millimeter, which allows flow sorting of beads

10 with diameters up to 500 microns. These instruments provide a reading of bead size and optical density as well as two fluorescent emission wavelengths from tags such as GFP, YFP or OS-red. Sorting speeds of 180,000 beads per hour and dispensing into multi-well plates or into a bulk receptor are possible. Following each round of cell culture, or after a defined number of rounds, the cell units can be assayed to determine whether there are

15 members displaying a particular genotype or phenotype. Genotype determination can be carried out using well known techniques such as the polymerase chain reaction (PCR), fluorescence in situ hybridisation (FISH), DNA sequencing, and others. Phenotype determination can be carried out by a variety of techniques, for instance by visual inspection of the cell units under a microscope, or by detecting a marker product characteristic of the

20 cell. This may be an endogenous marker such as a particular DNA or RNA sequence, or a cell protein which can be detected by a ligand, conversion of an enzyme substrate, or antibody that recognises a particular phenotypic marker (For instance see Appendix E of *Stem Cells: Scientific Progress and Future Research Directions*. Department of Health and Human Services. June 2001; appendices incorporated herein by reference). A genetic

25 marker may also be exogenous, i.e. one that has been introduced into the cell population, for example by transfection or viral transduction. Examples of exogenous markers are the fluorescent proteins (e.g. GFP) or cell surface antigens which are not normally expressed in a particular cell lineage or which are epitope-modified, or from a different species. A transgene or exogenous marker gene with associated transcriptional control elements can

30 be expressed in a manner that reflects a pattern representative of an endogenous gene(s). This can be achieved by associating the gene with a minimal cell type-specific promoter, or by integrating the transgene into a particular locus (e.g. see European patent No. EP 0695351). The labelled cell units can be separated from unlabelled ones by a variety of techniques, both manual and automated, including affinity purification ('panning'), or by

35 fluorescence activated cell sorting (FACS). Nishikawa et al (1998, Development vol 125, p1747-1757) used cell surface markers recognised by antibodies to follow the differentiation

of totipotent murine ES cells. Using FACS they were able to identify and purify cells of the haematopoietic lineage at various stages in their differentiation.

An alternative or complementary technique for enriching cell units of a particular genotype or phenotype is to genetically select the desired groups. This can be achieved for instance by introducing a selectable marker into the cell units, and to assay for viability under selective conditions, for instance see Soria et al (2000, Diabetes vol 49, p1-6) who used such a system to select insulin secreting cells from differentiated ES cells. Li et al (1998, Curr Biol vol 8, p 971-974) identified neural progenitors by integrating the bifunctional selection marker/reporter f3geo (which provides for f3- galactosidase activity and G418 resistance) into the Sox2 locus by homologous recombination in murine ES cells. Since one of the characteristics of neural progenitors is expression of Sox2, and therefore the integrated marker genes, these cells could be selected from non-neuronal lineages by addition of G418 after inducing differentiation using retinoic acid. Cell viability could be determined by inspection under a microscope, or by monitoring f3-gal activity. Unlike phenotype-based selection approaches, which can be limited by the availability of an appropriate ligand or antibody, genetic selection can be applied to any differentially expressed gene.

Determination of the identity or cell culture history of a cell unit When handling large numbers of cell units, their identity and/or cell culture history (for example the chronology and the exact nature of a series of culture conditions that any one group or unit may have been exposed to) can become confused. For instance, the split-pool protocol of cell culture necessarily involves mixing cell units in each round, making it difficult to follow individual units. Determining the cell culture history of a cell unit in a mixture of cell units which have been subjected to multiple culture conditions is sometimes referred to as 'deconvolution' of the cell culture history. One method of doing this is to label cell units and it is therefore advantageous to label the cell units. Labelling may be performed at the beginning of an experiment, or during each round of an experiment and may involve a unique label (which may or may not be modified in the course of an experiment) or a series of labels which comprise a unique aggregate. Similarly, reading of the label(s) may take place during each round or simply at the end of the experiment. Preferably, unique labels such as RFID labels are read during each round, whereas labels added serially at each round are read at the end of an experiment. Labelling of cell units may be achieved by a variety of means, for instance labelling either the cells themselves, or any material to which the cells are attached or otherwise associated with. Any of the chemical and non-chemical methods used to encode synthetic combinatorial libraries can be adapted for this purpose and some of these are described in Methods in Enzymology Vol 267 (1996), 'Combinatorial Chemistry', John N.

Abelson (Editor); Combinatorial Chemistry, Oxford University Press (2000), Hicham Fenniri (Editor); K. Braeckmans et al., 'Scanning the code', Modern Drug Discovery (Feb. 2003); K. Braeckmans et al, 'Encoding microcarriers: Present and Future Technologies'; Nature Reviews Drug Discovery, vol. 1, p. 447-456 (2002) all of which are herein incorporated by
5 reference. Some examples of labelling methods follow. One method of labelling cell units involves associating cell units with a tag that becomes sequentially modified as it is placed in different culture conditions. This may involve for instance the addition or subtraction of further units to the tag such that its stereochemistry, sequence or mass is altered; or the alteration of electronic memory as in read-write RF transponders (see below). Another
10 method of labelling cell units involves sequentially associating unique tags with the cell units whenever they are cultured under different conditions, such that subsequent detection and identification of the tags provides for an unambiguous record of the chronology and identity of the cell culture conditions to which the cell unit has been exposed. Tags can be taken up by cells, or attached to the cell surface by adsorption, or a suitable ligand or antibody, or
15 conjugated to a cell-associated matrix such as a carrier by adsorption, colloidal forces or a variety of linkages such as covalent linkage or non-covalent linkage, e.g. biotin-streptavidin linkage. For instance, one simple tag that can be introduced to cells or attached to a matrix associated with cells is an oligonucleotide of defined length and/or sequence. Oligonucleotides may comprise any class of nucleic acid (e.g. RNA, DNA, PNA, linear,
20 circular or viral) and may contain specific sequences for amplification (e.g. primer sequences for PCR) or labels for detection (e.g. fluorophores or quenchers, or isotopic tags). The detection of these may be direct, for instance by sequencing the oligos or by hybridising them to complementary sequences (e.g. on an array or chip), or indirect as by monitoring an oligonucleotide-encoded gene product, or the interference of the nucleotide with a cellular
25 activity (e.g. antisense inhibition of a particular gene). An advantageous method of amplifying nucleic acids is by rolling circle amplification (RCA; 2002, V. Demidov, *Expert Rev. Mol. Diagn.* 2(6), p.89- 95) where nucleic acid tags can comprise RCA templates, elongation primers, or struts that aid the circularization of minicircle templates). Any molecular or macromolecular tag can be used so long as it can be detected, including
30 peptide tags, coloured or fluorescent compounds, secondary amines, halocarbons, mixtures of stable isotopes etc. Tags may attach to cell units directly or via an intermediary, for instance an antibody raised against a component of the cell unit, or via an interacting pair such as biotin-streptavidin. In addition tags can be protected against degradation by the components of the cell culture, for example by chemical or other modification or by
35 encapsulation. Encapsulation of tags can take place in many different media, for example in beads many types of which are available from suppliers such as Bangs Laboratories Inc. (Fishers IN, USA), and encapsulation may be used to standardise tag dosage in addition to

providing components for tag amplification and/or detection (for example by providing PCR primers for use with a DNA tag). A preferred method of labelling cell units employs fluorescent beads such as those manufactured by Luminex Corporation (Austin, TX, USA). The Luminex system comprises polystyrene beads which may or may not be externally derivatised (e.g. with avidin or antibody) and are internally dyed with differing ratios of two spectrally distinct fluorophores, and a reader which is capable of characterising the spectral signature of each bead. A further preferred method employs beads such as those manufactured by Bangs Laboratories Inc. (Fishers IN, USA). The Bangs system comprises bead sets which can be distinguished based on differing sizes (e.g. bead sets of 4.4µm and 5.5µm diameter). Beads within each set can be furthermore distinguished from each other based on differing fluorescence intensity owing to differential loading with a single fluorescent dye. It is possible to use many different dyes with different absorption or emission characteristics, which can be internally loaded or attached externally to carriers by a multiplicity of means. It is furthermore possible to use 'quantum dots' to obtain a very high number of different fluorescent labels which can be read conveniently.

A preferred labelling technique, advantageously used when the test units are single cells, comprises nucleic acid modification of the cell by insertion of a nucleic acid "barcode", by viral infection (such as by a lentivirus) or gene editing, for instance by CRISPR.

Cell growth substrates such as those described in connection with forming cell units can be derivatised or coated with substances that facilitate tagging and do not interfere with cell growth. A preferred method of derivatising carriers is to modify them covalently or non-covalently with biotin, to which a tag can be attached via streptavidin or avidin. In general it will be important to use a tag that will not itself induce a cellular effect (i.e. an inert tag), and that can be distinguished from molecules present in cell units or the culture media, and that can be attached to its target and subsequently detected in the background of such molecules. To facilitate detection, it may be advantageous to selectively elute tags from cell units or to strip off the cells from cell units using selective conditions. More complicated molecular tagging strategies can also be envisaged, including the strategy of 'binary encoding' where information is recorded by a set of binary codes assigned to a set of molecular tags and their mixtures.

Detection of tags can be accomplished by a variety of methods familiar to those skilled in the art. Methods include mass spectrometry, nuclear magnetic resonance, sequencing, hybridisation, antigen detection, electrophoresis, spectroscopy, microscopy, image analysis, fluorescence detection, etc.

Of particular interest are labelling or encoding strategies in which labelling is carried out only once or where labelling and/or detection are non-physical and therefore non-invasive. Radiofrequency Identification (RFID) is an example of a system exhibiting these properties. RFID employs transponders (RF tags), antennae and readers. An RF tag is a small

5 electronic circuit, usually encased in glass or plastic, which in its simplest form provides access to a unique identification code that may be 'read', without contact or line of sight, by suitable electronics. Tags may also store information generated by the user, again without contact or line of sight. A 'reader' is an electronic unit that transfers information to and from one or more tags (it should be noted that the term reader is used interchangeably to mean

10 both a read only and read/write unit). The size and features of a reader may vary considerably, and it may operate in isolation, or be connected to a remote computer system. An antenna is used to transmit information from a reader to a tag, and to receive information sent by an RF tag. The size and format of an antenna will reflect the specific application, and may range from a small circular coil to large planar structures. An RFID system may operate

15 in isolation, or be connected to a remote computer for more comprehensive interpretation and manipulation of identification and associated data derived from a tag. One RFID strategy used in combinatorial chemistry is described in Nicolaou et al (1995, Angew Chem Intl Ed Engl, vol. 34, p. 2289) and comprises: (i) a porous enclosure containing a synthesis substrate and the semiconductor tag; (ii) the solid phase synthesis resin; (iii) a glass-

20 encased Single or Multiple Addressable Radiofrequency Tag semiconductor unit capable of receiving, storing and emitting radiofrequency signals. A similar device could be adapted to growing and following cell units simply by replacing the solid phase synthesis resin with tissue culture microcarriers or suitable cell units. More variations of this can be envisaged including but not limited to (coated or uncoated) RF tags on which cells are grown directly, or

25 RF tags implanted into cell units or organisms.

Thus tags do not necessarily have to be distinguished by their chemical or molecular structure in the first instance. Multiple variations of the non-chemical tagging strategy can be devised to determine the identity of a given cell unit in a mixture or of deducing the identity of the different cell units that comprise a mixture. For instance optical or visual methods of

30 tagging have been described where different shaped objects, graphically encoded objects or different colours denote the identity of a sample (for example see 1998, Guiles et al, Angew. Chem. Intl Ed Engl, vol. 37, p926; Luminex Corp, Austin TX, USA; BD Biosciences; Memobead Technologies, Ghent, Belgium), or where a pattern or bar code is etched onto a substrate such as a ceramic bar and recognised using pattern recognition technology (for

35 example see 1997, Xiao et al, Angew. Chem. Intl Ed Engl, vol 36, p780; SmartBead Technologies, Babraham, UK). A further method of tracking or labelling cell units is to

encode their identity spatially, i.e. by their position in space. In this method different cell units are segregated in defined relative positions, and these positions denote or encode the identity of the units. For instance, cell units may be cultured in an array, whereby the identity and/or culture history of each unit is known and is associated to a particular position in the array. In their simplest forms such arrays can comprise collections of tissue culture flasks, wells of a multi-well plate, or locations on a glass slide or other surface. Examples of positional encoding strategies can be found in Geysen et al. (1984, Proc Natl Acad Sci USA vol. 81, p. 3998-4002), Fodor et al. (1991, Science vol. 251, p. 767-773), Ziauddin and Sabatini (2001, Nature, Vol. 411, p. 107-110), and Wu et al. (2002, Trends Cell Biol. Vol. 12(10), p.485-8). The invention has many facets, each of which may have many forms that may be combined to form numerous permutations of the invention.

It will be apparent that it is not necessary to label all of the cell units in order to be able to deduce information about the outcomes resultant from a combination of cell culture protocols. Thus without labelling of cell units it would still be possible to assay large combinations of cell culture conditions according to the invention, and to determine whether one or more of these was capable of resulting in a particular cellular effect. However, cell units are preferably labelled. Labelling of a cell unit allows the derivation of useful information from the experiment regarding the outcome of the particular conditions sampled by the labelled cell unit, as opposed to all the cell units.

Alternatively it is sometimes advantageous to label one or a few group(s) of cell units which have all been exposed to a certain culture protocol, for instance a group of cell units which have been segregated into the same medium during a particular split or pool step. It will also be apparent that labelling certain cell units allows one to infer the identity of other (perhaps unlabelled) cell units.

Similarly, it will be clear that performing cell culture experiments in which various conditions are omitted can give information regarding the utility of those conditions with respect to a particular experimental outcome. It would therefore be possible to evaluate each of the conditions sampled in a manner according to the invention by repeating the experiment a number of times, each time omitting a different set of conditions. Split-split cell culture steps can also be used to determine the effect of a particular set of conditions on experimental outcome. In effect split-split steps result in the formation of particular lineages of cell units which have each been exposed to a unique cell culture conditions at the time of branching. By studying the different lineages it is possible to determine the utility of the tissue culture conditions studied at the branching point, with respect to a particular experimental outcome.

Cell reprogramming involves changing the identity or function of a cell. This can involve transforming a cell from one type to another or reverting mature cells back to a more pluripotent or stem cell-like state, before re-differentiating into a more committed state.

In the context of the present invention, reprogramming is the conversion of a differentiated,
5 committed cell type into another differentiated, committed cell type. Reprogramming excludes mere conversion of a pluripotent cell into a differentiated cell type.

Several techniques for reprogramming are known to the skilled person; in one embodiment, reprogramming as referred to herein may encompass any one or more of these techniques.

Induced Pluripotent Stem Cells (iPSCs):

10 Discovered by Shinya Yamanaka and his colleagues in 2006, this method involves introducing a specific set of transcription factors (often referred to as Yamanaka factors: Oct4, Sox2, Klf4, and c-Myc) into mature cells.

These factors induce the mature cells to revert to a pluripotent state, similar to embryonic stem cells. The pluripotent cells can then be further differentiated into a desired cell type.

15 Yamanaka used retroviral vectors to deliver TFs, but many other methods are available, including non-integrating viral vectors, mRNA delivery, plasmids and direct addition of TF proteins, as well as the use of non-coding RNA.

Various types of ncRNAs, including microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and others, can be involved in regulating transcription factors. These ncRNAs
20 can be used to replace TFs in the assays of the invention.

microRNAs (miRNAs) are short (~22 nucleotides) ncRNAs that generally bind to the 3' untranslated region (UTR) of target mRNAs, leading to their degradation or translation repression. Through this mechanism, miRNAs can negatively regulate the levels of specific transcription factors (Bartel DP. (2009) MicroRNAs: target recognition and regulatory
25 functions. *Cell*. 136(2):215-33).

Long non-coding RNAs (lncRNAs) are longer than 200 nucleotides and have diverse roles. Some lncRNAs can interact with transcription factors directly, modulating their activity, and can thus serve to modulate TF activity in an assay according to the invention. See Rinn JL, Chang HY. (2012) Genome regulation by long noncoding RNAs. *Annual Review of*
30 *Biochemistry*. 81:145-66.

Non-coding RNAs have been shown to play roles in conjunction or as an alternative to the classic Yamanaka factors:

1. miR-302/367 Cluster: This cluster of microRNAs has been shown to enhance reprogramming efficiency when used in combination with the Yamanaka factors (Oct4, Sox2, Klf4, and c-Myc). See Anokye-Danso F, et al. (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*. 8(4):376-88.
2. ****lncRNA-RoR****: This lncRNA has been implicated in the regulation of reprogramming and maintenance of pluripotency. It was shown that lncRNA-RoR can modulate the core transcriptional network of pluripotency, including influencing the levels of the Yamanaka factors. See Loewer S, et al. (2013) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genetics*. 45(12):1504-9.

Direct Lineage Conversion (Transdifferentiation):

This method bypasses the pluripotent state. Instead, cells are directly converted from one mature cell type to another using a combination of lineage-specific transcription factors.

For example, fibroblasts can be directly converted to neurons, cardiomyocytes, or other cell types, depending on the factors used. Many techniques are currently available for transdifferentiation:

1. Viral Vector-mediated Transcription Factor Introduction:
Using viral vectors, lineage-specific transcription factors are delivered to cells. For example:
Converting fibroblasts to neurons using factors like Ascl1, Brn2, and Myt1l (Vierbuchen, et al., (2010). *Nature*, 463(7284), 1035-1041).
Converting fibroblasts to cardiomyocytes using factors like Gata4, Mef2c, and Tbx5 (Ieda, et al., (2010) *Cell*, 142(3), 375-386).
2. RNA-based Transcription Factor Delivery:

This method employs synthetic modified mRNA to transiently express the necessary transcription factors. RNA-based methods can alleviate concerns associated with potential DNA integrations into the host genome. Warren, et al., (2010) *Cell Stem Cell*, 7(5), 618-630.

3. Small Molecules and Chemicals:

While transcription factors are most commonly used for direct lineage conversion, some studies have successfully employed combinations of small molecules to either enhance the efficiency of the process or, in rare cases, achieve transdifferentiation without the need for genetic factors.

For example:

Hou, et al., (2013). Science, 341(6146), 651-654, successfully reprogrammed mouse somatic cells into pluripotent stem cells using only small molecules, highlighting the potential to generate iPSCs without introducing exogenous genes.

Shi, et al., (2008). Cell stem cell, 2(6), 525-528 showed how a combination of small molecules can significantly enhance the efficiency of iPSC generation from mouse fibroblasts.

Fu, et al., (2015). Cell Research, 25(9), 1013-1024 used a set of small molecules to directly convert mouse fibroblasts into cardiomyocyte-like cells.

4. CRISPR/Cas9-based Activation of Endogenous Genes:

The CRISPR/Cas9 system, originally known for its gene-editing capabilities, can be adapted for gene activation by using a deactivated Cas9 (dCas9) fused to transcriptional activators. This can be used to activate endogenous lineage-specific genes, promoting transdifferentiation. Gilbert, et al., (2013) Cell, 154(2), 442-451.

5. MicroRNAs (miRNAs):

Certain miRNAs have been found to play roles in cell fate decisions and can be used alongside transcription factors or alone to drive direct lineage conversion. Jayawardena, et al., (2012) Circulation research, 110(11), 1465-1473.

6. Epigenetic Modulators:

Direct lineage conversion involves significant changes in the epigenetic landscape of a cell. Incorporating molecules that modulate epigenetic marks can increase the efficiency of transdifferentiation. Polo, Jet al., (2012) Cell, 151(7), 1617-1632

7. Extracellular Cues and Microenvironment:

While transcription factor-mediated conversion is the primary driver, the efficiency and success of the conversion process can be influenced by the microenvironment, including extracellular matrix components, neighboring cells, and culture conditions. Engler, A. J., et al., (2006). *Cell*, 126(4), 677-689.

- 5 In the context of the present invention, reprogramming is preferably transdifferentiation, by which cells are converted directly from one differentiated state to another, without passing through a pluripotent state.

A “cell type”, as referred to herein, is a differentiated cell belonging to a specified lineage. A cell type is not a totipotent cell. In embodiments, a cell type is not a pluripotent cell capable of differentiating into more than one lineage. A cell lineage may be any of a variety of cells
10 types which can develop during normal embryogenesis. Examples include:

Hematopoietic Lineage: This lineage gives rise to all the different blood cells. Hematopoietic stem cells (HSCs) in the bone marrow differentiate into various types of blood cells, such as erythrocytes (red blood cells), various leukocytes (neutrophils, lymphocytes, and
15 monocytes), and platelets.

Neuronal Lineage: Neural stem cells differentiate into various types of neuronal cells and glial cells. This includes motor neurons, sensory neurons, astrocytes, oligodendrocytes, and microglia.

Epidermal Lineage: The epidermal stem cells in the skin differentiate into various skin cells,
20 including keratinocytes, melanocytes, and cells of the hair follicles.

Muscle Lineage: Myogenic stem cells (or satellite cells) can differentiate into muscle fibres or myocytes.

Mesenchymal Stem Cell (MSC) Lineage: MSCs are multipotent and can give rise to a variety of cell types, including osteoblasts, chondrocytes, fibroblasts and adipocytes.

25 Intestinal Lineage: Stem cells in the crypts of the intestines differentiate into various types of intestinal cells, such as absorptive enterocytes, goblet cells, Paneth cells, and enteroendocrine cells.

Hepatic Lineage: Liver stem cells can differentiate into hepatocytes, which are the main functional cells of the liver, and cholangiocytes, which line the bile ducts.

Cardiac Lineage: Cardiac progenitor cells can differentiate into various cell types of the heart, such as cardiomyocytes, endothelial cells, and smooth muscle cells.

Germ Cell Lineage: This lineage leads to the formation of ova in females and sperm in males.

- 5 Retinal Lineage: Retinal progenitor cells can differentiate into various types of retinal cells, such as photoreceptors, bipolar cells, and ganglion cells.

Reprogramming can involve the conversion of a cell from one lineage to another, or from one cell type within a lineage to another cell type within the same lineage, for example from a fibroblast to an osteoblast.

- 10 A “test unit” as referred to herein is a cell or group of cells which can be handled as a unit. It is synonymous with cell unit. For example, a test unit can be a cell unit as referred to in WO2004031369, comprising cells arranged on a microcarrier. In a preferred embodiment, however, a test unit is a single cell.

- 15 Where the test units comprise a carrier which in turn comprises one or more cells, the carrier may be populated with cells in a stochastic manner. Where the intended cell number is low, such as one cell or two cells per unit, the probability of cells populating the test units may be reduced such that some units comprise no cells. Thus, in embodiments not all test units comprise cells. Preferably, however, the test units are single cells and do not comprise any carrier; in such an embodiment, all test units comprise a cell.

- 20 “Reagents” as referred to herein may be any chemical, small molecule, transcription factor, nucleic acid, protein, compound or ion which can interact with a cell. In preferred embodiments, reagents are transcription factors, small molecule or protein compounds, or nucleic acids. In the most preferred embodiment, reagents are transcription factors (TFs) or agents which influence transcription factor activity, such as ncRNAs.

- 25 Certain reagents are known in the art to promote reprogramming, as described in references cited herein and in more detail below. Transcription factors are often used to influence reprogramming.

- “Labelling” can include any technique which is capable of individually labelling a cell. For example, reference is made to the techniques described above. In a preferred embodiment,
30 cells are labelled by genetic modification, such as by the introduction of retroviral markers or modification using genome editing techniques, as described below.

Deconvolution of labels refers to the identification of labels on a cell, and optionally the order in which the labels were acquired by the cell, and thereby determining which reagents or combinations of reagents the cell has been exposed to. Since the cell is labelled whenever it is exposed to a given reagent, the reagent leaves a "mark" on the cell which can be used to map the cell's history in exposure to reagents as well as the timing of such exposure.

Predicting the most effective transdifferentiation-inducing reagents requires a combination of experimental data, computational modelling, and optimization techniques. An algorithm can be used to interrogate the results of exposure of cells to given reagents, and thereby predict which reagents can be employed to induce a desired transdifferentiation in a given cell type.

The results of the reagent exposure should be presented as a consistent metric to the algorithm; for example, the results can be in the form of gene expression data, cell morphology scores, specific gene activation events, and the like. Output from the algorithm may be used to select initial reagents for the performance of the method of the invention. Where the method is performed multiple times, the output data can be used to inform further instances of the performance of the method.

Machine learning (ML, AI) can be employed to better predict input reagents for desired transdifferentiation outcomes. A database may be provided of gene expression data related to reagent, and gene regulatory networks constructed further comprising the machine-learning model configured to output a gene regulatory graph.

Pooling and splitting or subdividing of cell populations is carried out as described in WO2004013369. Iterative repetition of the steps of the process of the present invention can also be performed in a manner analogous to that described in WO2004013369 and above.

Identification of cell types can be performed by any suitable means, including genetic, morphological, immunochemical or other features.

Gene expression analysis includes the analysis of one or more reporter genes, as well as analysis based on a wide variety of cellular genes including using arrays which analyse expression of the entire genome. In one embodiment, gene expression analysis focusses on one or more genes which are markers for the desired cell type. The method of the invention can be used to determine such genes by analysing gene expression changes in cells which are exposed to reagents which are shown to lead to desired transdifferentiation outcomes by analysis of alternative markers of cell differentiation, such as immunochemical markers or cell morphology.

Genes which influence cell transdifferentiation can in turn be purposefully modulated in order to induce transdifferentiation; for example, information about a gene being modulated in response to exposure to a reagent can be used to select a gene modulation technique, such as genome editing or the like, to further modulate the gene in place of administration of the reagent.

CombiCult®

A procedure for splitting and pooling cellular populations to determine the effect of reagents on cell fate has been described in WO2004031369. This combinatorial screening platform allows testing thousands of time-resolved combinations of culture conditions at a time. It is based on combinatorial science, which had already been successfully employed in chemical synthesis of new compounds. Briefly, stem cells are grown in or on beads that can be tagged with condition-specific fluorescent labels. For every step in the differentiation course, beads are pooled and split, and subjected to different conditions, where they are concurrently labelled in a condition-specific fashion.

This process is repeated for several steps, ensuring that every bead is exposed to one specific combination of conditions, with multiple beads going through the same combination. Cells in beads are then screened for markers, using antibodies, and selected through a large particle sorter. Positive beads are isolated and dissociated, and their tags are read by a FACS, thus detecting all the combinations of tags enriched in the positive beads. A computational deconvolution strategy allows to identify and quantify the most frequent tag combination, which will correspond to a specific set and order of addition of molecules to the cell culture (also termed "protocol"). Highly represented protocols are then validated independently in vitro and compared to gold standard differentiation or expansion procedures. CombiCult® has been successfully deployed in several cases, allowing the discovery of for example novel protocols for expansion of hematopoietic stem cells, differentiation of macrophages, neutrophils, natural killer cells, megakaryocytes, smooth muscle cells, and oligodendrocyte progenitors.

The decision on which molecules to include in a CombiCult® screening comes from careful reading of literature, meaning that its power to discover new combinations is somehow limited by already suggested hypotheses. By harnessing the power of modern machine learning models combined with the large amount of high-resolution single cell transcriptomics datasets, the invention permits the generation of new data-driven hypotheses with an unprecedented breadth.

Moreover, the multiplexing potential of CombiCult® is limited by the use of beads and the selection of one positive cell fate per screen, discarding the rest of the material.

The present invention provides:

1. Bespoke computational analyses to define the input of molecules to be tested,
5 including a machine learning algorithm to prioritize molecular drivers of cell fate acquisition and prevent contradictory or opposing pathways to be tested together;
2. A single cell tagging approach to entirely bypass the need for beads and large
particle sorting, while simultaneously acquiring single cell definition and gene expression-
based monitoring of differentiation results through low cost single cell sequencing (i.e.
10 Oxford Nanopore benchtop sequencing);
3. A second computational pipeline to analyse the results of the combinatorial screen
by identifying viral barcodes and their representation in the acquisition of cell fates of
interest.

These 3 elements create a virtuous cycle in which the information from one screen can be
15 used to better refine the initial predictions, as more and more experimental evidence linking
molecules and their transcriptional effects is generated at every round. Eventually, once a
critical mass of evidence is generated from several experiments, a new AI- based approach
can be implemented in which, for a given profile, an ordered set of molecular drivers can be
efficiently predicted, effectively reducing the need for combinatorial testing.

20 **Computational analysis**

The results of one CombiCult screen can be used to select input TFs for a further screen, or
alternatively to determine optimal TFs for a desired transdifferentiation protocol.

The wealth of publicly available single-cell RNA-sequencing (scRNA-seq) and single cell
chromatin accessibility data (scATAC-seq) from cells actively differentiating shows a
25 continuum of gene expression changes.

In general, an algorithm for selection of TFs can process the results of screens as follows:

Input:

1. List of cells with their identities: cell_list

2. List of reagents: reagent_list

3. Data on the effect of each reagent on each cell: effect_data (this could be gene expression profiles, cell phenotype changes, or any measurable metric post-exposure)

4. Desired cell type profile: desired_profile

5 Output:

List of selected reagents: selected_reagents

Steps:

1. Initialize an empty list: selected_reagents = []

2. For each cell in cell_list:

10 a. For each reagent in reagent_list:

i. Get the effect of this reagent on this cell from effect_data.

ii. Compute the similarity score between the effect and the desired_profile. (This could be a correlation score, frequency metric, or any relevant measure of similarity.)

b. Rank the reagents for the cell based on the similarity scores.

15 c. Store the top-ranked reagent for this cell.

3. Aggregate the top-ranked reagents across all cells.

4. For each reagent in the aggregated list:

a. Compute the frequency of its selection across all cells.

b. If the frequency is above a certain threshold (indicating its consistent efficacy):

20 i. Add the reagent to selected_reagents.

5. Return selected_reagents.

By further sorting cells into cell types, it is possible for the algorithm to return selected_reagents associated with specific cell type transdifferentiation events.

Machine learning may be applied to the algorithm, to further refine the selection of reagents. For example, a machine learning algorithm may be applied in the following manner:

Input:

1. Training dataset: Contains cell identities, timing and identity of reagent exposure, and observed effects post-exposure.
2. Desired cell type profile: desired_profile

Output:

Optimal reagent(s) for transdifferentiation: optimal_reagents

Steps:

- 10 1. Data Preprocessing:
 - a. Normalize and standardize the data (e.g., z-score normalization for gene expression levels).
 - b. Split the dataset into training and validation sets.
2. Feature Selection:
 - 15 a. Select relevant features (e.g., specific gene expression levels, cell markers).
 - b. If dataset is high-dimensional, consider dimensionality reduction methods like PCA.
3. Model Selection:
 - a. Choose a suitable ML model. Regression models like Random Forest or Gradient Boosting Machines (GBM) are examples.
 - 20 b. Train the model on the training set.
 - c. Validate the model on the validation set.
4. Hyperparameter Tuning:

a. Use techniques like grid search or random search to find the optimal hyperparameters for the chosen model.

b. Retrain the model using the optimal hyperparameters

5. Prediction:

5 a. Use the trained model to predict the effect of each reagent on transdifferentiation towards the desired cell type profile.

b. Rank reagents based on their predicted effectiveness.

6. Model Interpretation (Optional but recommended):

10 a. Use techniques like SHAP (SHapley Additive exPlanations) or feature importance scores to interpret which features (e.g., specific timings, cell markers) are most influential in the predictions.

7. Return:

a. Return the top-ranked reagent(s) as optimal_reagents.

15 The algorithm of the invention applies gold standard data pre-processing steps and machine learning methods to identify temporal (or pseudo-temporal) trajectories and gene expression changes associated with them. These trajectories will connect stem cells to terminally differentiated cells through a continuum of gene expression changes. This continuum can be treated as a high-resolution sequence of molecular events that are specific to the cell fate acquisition process (**Fig. 2**). Within this continuum, it will be possible to detect a set of
20 transcription factors (TFs) that drive gene expression changes. The algorithm of the invention prioritizes and orders these TFs based on 4 main features:

1. Their trend along the differentiation trajectories (differential expression)
2. Their gene regulatory network, inferred through multiple regression and externally validated TF-target networks and protein-protein interaction networks (PPINs)
- 25 3. Their specificity, i.e. preferential activity within a trajectory compared to all the other ones in a similar physiological process (e.g. selecting a T-cell specific set of drivers compared to other lymphoid lineages in haematopoiesis)
4. Their dynamics in terms of switching time and/or transient expression.

Additionally, if matching public scATAC data is available, it can supplement TF activity by performing footprinting/TF binding site analysis within the open chromatin regions, thus providing orthogonal evidence for the gene regulatory network, and complementing the other dataset(s).

- 5 Once TFs have been identified, target networks from each of them will be screened for annotated interactions. If the interactors of two TFs show a significant degree of overlap but are involved in opposing interactions (e.g., one TF is reported to inhibit these interactors, the other one is reported to activate them), they will be flagged for incompatibility within the same screen, thus providing a more rational way to arrange combinations.
- 10 The algorithm of the invention can take advantage of regression-based gene- gene correlation networks and existing PPINs to identify molecules that act as transducers and receptors in a signalling pathway that is upstream of the TFs and is also changing along the trajectory. Once candidate receptors have been identified, their ligands can be inferred and included in the combinatorial screen.
- 15 Finally, small molecule-protein interaction networks overlaid on the PPINs, and drug-transcriptome correlation databases whose signature can be interrogated along a trajectory, can also be included in Daedalus for their prioritization and inclusion in the screening.

- Along the generation of new hypotheses, the CombiCult platform can be improved by greatly extending its multiplexing and resolution capabilities. The implementation in WO2004031369
- 20 is limited to the identification of one cell differentiation at a time, which means that several other possibilities that may have emerged in culture – such as alternative cell fates – remain unexplored. In this proposal I will design and implement a system to tag (“barcoding”) single cells using viral transduction. To avoid using beads (whose encapsulation represents a time-consuming and resource-intensive step) and to track all single cells during the screening,
 - 25 barcoding is achieved through a unique, short, expressed DNA sequence that functions as a Unique Condition Tag (UCT) delivered through a lentiviral vector, which stably integrates in cells in culture (**Fig. 3**). As cells go through different splits and are treated with different molecules/TFs/growth factors (hereafter referred to as “drivers”), they will be tagged by lentiviral vectors bearing a barcode unique to the condition. By using viral, integrating tags at
 - 30 low doses – so that cells do not become saturated at every transduction cycle – cells can be uniquely barcoded in every condition, and the UCT stays within each cell, tracking the condition in which they were inserted. In the first split, candidate drivers and their corresponding vectors are added to the culture media, which can be refreshed to remove viral vectors after transduction according to the needs of the protocol.

Then, cells are detached, resuspended, pooled together and split in different plates for the second split.

Pooled and split cells undergo a second round of transduction and driver treatment; media is refreshed, and drivers are re-added until the third pool and split process.

- 5 This is repeated for as many splits as necessary, until the final split is reached and cells are detached, optionally sorted via FACS and prepared for single cell RNA sequencing. The fundamental difference between this approach and the approach of WO2004031369 is that in the present invention barcoding happens at the single cell level – and the data consist of the transcriptome of every cell, with its corresponding combination of UCTs.

10 **Development of a new analytical pipeline to select protocols**

- The analysis of single cell RNA coupled with UCT detection resolves the effect of different cell fate choices (both successful and unsuccessful) in a single snapshot. Data from the transcriptome are analysed, annotating cells through their similarity and expression of cell-type specific signatures. By reconstructing UCT "lineage" trees and matching them to
15 signatures, it is possible to chart the sequence of variables that gives rise to cell-fate specific transcriptional profiles. There are three fundamental advantages in using single cell data coupled to UCTs: first, it allows to resolve even small differences in cell fate acquisition; second, if the whole experiment is sequenced and analysed, it allows to identify several cell fates at once; third, it generates data that can be reused and integrated.

- 20 Regarding the first two aspects, scRNA-seq has been already deployed in similar ways to track guide RNA constructs in CRISPR screens (Dixit et al. 2016. Cell 167, 1853–1866) or lentivirus-mediated TF transduction (Joung et al. 2023. Cell 186, 209–229). Furthermore, complex studies using sequential genome editing ("scarring") for lineage tracing at the single cell level have been successfully carried out (Spanjaard et al. Nat Biotechnol . 2018 June ;
25 36(5): 469–473), providing a first analytical framework to connect different events in a single experiment.

- These pioneering approaches have been used to study transcriptional responses to perturbations at a systematic level, including cell fate acquisition, proving the feasibility in principle of a genetic tag-based tracing of experimental and/or physiological conditions at the
30 single cell level. The innovative and ground-breaking aspect of this proposal lies in the application of CombiCult® principles, which allow to keep track of thousands of combinations of sequential addition of molecules or TFs in culture, thus resolving a much

larger experimental space in a single assay. The third advantage allows to greatly enhance predictive power and ties the three elements of this platform together: the generation of datasets where transcriptional changes at the single cell level are matched to specific cell culture conditions provides an excellent opportunity to feed the algorithm again with experimentally validated driver-target interaction data. This virtuous cycle allows the algorithms of the invention to become more and more precise, learning from their own experimental results, and paves the way for a supervised learning framework that, in the long run, can predict with high precision the molecules required for any type of differentiation. In fact, given a large number of ground truth (i.e., data resulting from a known intervention) it is in principle possible to create a Language Model (LM) that is able to predict the steps needed to generate specific cell types. Different types of LMs have been successfully deployed in studies of viral evolution and antibody evolution (Hie et al. 2023. Nat biotechnol doi:10.1038/s41587-023-01763-2) and chemical modifications (Kosonocky et al. arXiv 2023 doi:10.48550/arXiv.2305.16330).

The successful application of the sequencing-based platform of the invention will quickly generate many such data points, thus greatly accelerating ATMP engineering and production.

Transcription Factors

The most commonly recognised reagents for promoting transdifferentiation are transcription factors. Transdifferentiation can be driven by overexpressing specific transcription factors, as has been observed various transdifferentiation experiments:

1. Neuron:

- Ascl1, Brn2, and Myt1l can reprogram fibroblasts into induced neurons (iNs).

- NeuroD1 alone or in combination with other factors can also convert fibroblasts into neurons.

2. Cardiomyocytes:

- Gata4, Mef2c, and Tbx5 can convert fibroblasts into induced cardiomyocyte-like cells (iCMs).

3. Endoderm cells:

- FoxA2 and Hnf1 α can convert fibroblasts into hepatic cell types.

4. Pancreatic beta cells:

- Pdx1, Ngn3, MafA, and sometimes with other factors, can convert various cell types into insulin-producing beta-like cells.

5 5. Endothelial cells:

- Flt1, FoxC2, and Etv2 can convert fibroblasts into endothelial-like cells.

6. Myocytes:

- MyoD is a classic example that can convert fibroblasts into skeletal muscle cells.

Further transcription factors useful in transdifferentiation can be derived from the literature.

10 The foregoing are examples of agents known in the art to promote reprogramming of cells.

Improved reagent database

The method of the invention permits the generation of a database which can improve the starting parameters for an experiment designed to identify TFs suitable for reprogramming a cell. Therefore, the invention provides a method for selecting the starting TFs in a method

15 according to the preceding aspects of the invention, comprising the steps of:

compiling a database of TFs and cell transdifferentiating results; and selecting from the database a set of TFs or combinations of TFs for use in a method for determining the conditions for reprogramming a cell.

The invention also provides a method for improving a database of TFs and cell
20 transdifferentiating results, comprising using the database to select TFs as in the preceding embodiment, and inputting the results of the methods for determining the conditions for reprogramming a cell into the database, thus providing further data points.

The algorithms and machine learning approaches described above may also be used to further improve the database.

25 In another aspect, the method according to the invention provides: a database of TFs (for example transcription factors) useful for reprogramming cells; at least one computer

processor; a memory in operative communication with the processor, the memory comprising instructions configuring the processor to:

- (a) generate gene regulatory networks (GRN) from the database of transcription factors;
- (b) determine a candidate transcription factor for a desired reprogramming event
- 5 (c) analyse an impact of the candidate transcription factor in cell reprogramming;
- (d) optionally, iteratively repeat steps (a)-(c); and
- (e) output a set of optimal transcription factors.

The database preferably comprises gene expression data, such as RNA-seq data.

- 10 Determining a candidate transcription factor comprises interrogating a gene regulatory graph to identify the set of optimal transcription factors to differentiate any given cell type from any given starting cell. A machine-learning model can be used to generate a metric calculation as a function of the gene regulatory graph.

- Examples of a metric calculation include a criticality algorithm, which may be configured for time-series RNA-seq data; see Oh et al., Biomed Res Int. 2013;2013:203681. doi: 10.1155/2013/203681. Epub 2013 Mar 24.
- 15

- The invention further provides a method for determining optimal transcription factors cell reprogramming, comprising: curating, using a computing device, a gene expression database related to use of TFs in cell culture; generating gene regulatory networks from a plurality of gene expression datasets; determining a candidate optimal TF; analysing an impact of the optimal TF in cell differentiation; and outputting a set of optimal TFs.
- 20

Gene regulatory networks can be predicted computationally, as described in Kamaraj, Cell Cycle 2016, vol. 15, no. 24, 3343–3354.

- Methods include CellNet (et al., Cell 2014; 158:903-15), which uses microarray gene expression data and correlation-based network score, the method of D'Alessio et al. (D'Alessio et al., Stem Cell Reports 2015; 5:763–75) whose method uses microarray gene expression data and an information theoretic approach (Jensen- Shannon Divergence) and Mogrify (Rackham et al., Nat Genet 2016; 48(3):331–5; PMID:26711105) which uses cap analysis of gene expression (CAGE) data and a network based score.
- 25

- Regardless of the computational approach selected, the iteration of the experimental steps with computational analysis provides a novel advantage over computational and experimental approaches used alone.
- 30

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made to the Examples, which are not intended to limit the invention in any way.

Examples

5 Materials and Methods

mRNA production and transfection

DNA template preparation and mRNA synthesis were conducted according to HiScribe™ T7 ARCA mRNA Kit (with tailing) (NED, #E2060S) protocol.

10 For transfection with jetMESSENGER® (Axil Scientific, #150-07), 2 µg of mRNA were diluted in 200 µl of mRNA buffer (provided by the kit) followed by thorough mix for 10 seconds. Then 4 µl of jetMESSENGER® reagent was added and gently combined to the mix. After 10 minutes incubation at RT, transfection was performed by addition of the transfection mix to the cells in medium without antibiotics. For serial transfection, the transfection mix was sequentially added in the culture every 24 hours.

15 Immunofluorescence

For Immunostaining, cells were fixed with 4% Paraformaldehyde (PFA) (Sigma Aldrich, #78775) at RT for 30 minutes and washed three times with DPBS for 5 minutes each. For soluble proteins, blocking solution is 10% Bovine Serum Albumin (BSA) (Sigma Aldrich, #05470) or 10% Fetal Bovine Serum (FBS) (Thermo Scientific, #26140079) with 0.5% Triton X-100 (Sigma Aldrich, #T8787) whereas for membrane proteins, Triton X-100 was excluded. Cells were incubated with blocking solution for an hour at RT. Primary antibody diluted in blocking solution, was added to cells and incubated at 4°C overnight. Next day, three washes were performed with DPBS for 5 min each. Secondary Antibody was diluted at 1:500 in DPBS and incubated at RT for 2 hours in dark. Subsequently, three washes with DPBS were performed for 5 min each. For the visualization of cell nuclei, Hoechst 3342 dye (Thermo Fisher Scientific, #62249) was employed, with a dilution ration of 1:10,000 (in DPBS). The dye was then incubated in dark for 5-10 minutes followed by three washes in DPBS. Lastly, the staining was detected using fluorescence microscopy (Inverted Live Cell Confocal Microscope-LSM800 Airy, Clinical Science Building (CSB)) (Inverted Fluorescence Live Cell Microscope-A07, CSB)

Flow cytometry immunostaining

For the quantification of the efficiency of the serial transfection, FACS was performed. cPPs were dissociated from the beads and fixed with 100µl of 4% PFA for 15 min at RT. Two washes were apply with staining buffer (BD, #554657) before permeabilized with blocking buffer for 30 min at RT. After centrifugation, conjugated antibodies were diluted at 1:20 in blocking buffer for 30 min on ice in dark. Following incubation, two washes were performed with DPBS. At last cells were suspended in DPBS and analyzed through FACS.

For the quantification of c-peptide⁺ cells using primary and secondary antibodies, a standardized protocol was implemented. Initially, cells were dissociated and fixed with 100 µl of 4% PFA for 30 minutes on ice. Subsequently, a single wash with 200 µl of DPBS was carried out to remove excess fixative and cellular debris before proceeding to blocking. Blocking was performed on ice for 30 minutes to minimize non-specific binding of antibodies. Following this, incubation with the primary antibody was conducted overnight at low temperature to facilitate specific binding of the primary antibody to its target antigen. The next day, the cells underwent two washes with blocking solution to remove unbound primary antibody, followed by blocking with the secondary antibody on ice for 2 hours to enable detection of the primary antibody. After two additional washes with DPBS to remove excess secondary antibody, the cells were carefully collected in FACS tubes for subsequent flow cytometry analysis to quantify the c-peptide⁺ cells accurately.

Table 1. The table below summarizes antibodies and respectively dilution used.

Antibodies used	dilution
Primary: Mouse anti-pro-Insulin c-peptide (Millipore, #05-1109)	1:1000
Secondary: Donkey anti-Mouse, Alexa Fluor 488 (Thermo Fisher Scientific, #A-21202)	1:500

Fluorescence-activated cell sorting (FACS)

FACS (LSRFortessa X-20, CSB) was performed using an LSRFortessa X-20 (CSB) to assess and quantify transfected cells within the population. A dot plot of side scatter (SSC) versus forward scatter (FSC) was utilized to identify and separate live cells from doublets and dead cells. For the live cells, a dot plot of GFP/APC/PerCP versus FSC was created for additional analysis. All FACS data were evaluated using FlowJo software.

a) **CombiCult® screen.**

- i. **Seeding of cells onto the beads.** cPPs were seeded onto CombiCult® beads. For effective seeding, the beads were coated with matrigel (dilution 1:100; BD, #354277) for at least 2 hours at 37°C. Single cells were generated from confluent cPPs with GCDR incubated for 10 min at 37°C. Cells were counted to achieve the desired seeding density, and following centrifugation, cPPs were reconstituted in their basal media plus supplements and Rock inhibitor Y27632 (15). Approximately 4.0×10^7 cPPs (250 cells/bead) were seeded onto 160,000 beads (40,000/condition), and the mixture (cells and beads) were incubated at 37°C overnight.
- ii. **Pool/Split process.** Based on the designed matrix the cells were distributed onto the beads in predetermined conditions (10 conditions expressing specific TF) for four days using a split-pool process, allowing to test 10,000 thousand different combinations of TFs.
- iii. **mRNA transfections and addition of fluorescence tags.** In subsequent days, mRNAs-TFs transfections were performed according to the matrix followed by addition of CombiCult® specific fluorescent tags at each split (Figure 2). Simple medium with few growth factors and known molecules was used to push cPPs into β -cells via endocrine progenitors. First 2 days, DMEM high glucose was supplemented with Asc (50 μ g), 1% B27, EGF (50ng), FGF-7 (50ng) and RA (50nM), while next two days 5% KOSR, T3 (1 μ M) and RA (25 nM) were added to DMEM high glucose. From day 5 to day 11, DMEM low glucose was enriched with 10% FBS, T3 (1 μ M) and Alk5i (10 μ M).
- iv. **Screening assay.** Following the pool/split process, the beads were stained at day 11 with antibody against c-peptide, functional pancreatic β -cells marker.
- Concurrently with the CombiCult® screen, negative and positive controls were implemented to fine-tune the c-peptide staining protocol and establish criteria and parameters for isolating the positive beads. In both the cases, same media of CombiCult® experiment were adopted. However, positive reprogramming sample went through serial mRNA transfections of *Pdx1* (day 1), *Ngn3* (day 2) and *MafA* (day 3), known trio of factors that induce reprogramming to β -cells, while mRNA transfections were not performed in the negative control. The beads from the screen were transferred in Eppendorf tubes. Once the beads were settled, the media was removed and a wash with DPBS was performed to discard remaining media. Next, the beads were fixed with 4% PFA (in DBPS) at RT for 30 minutes and washed three times with DPBS for 5 minutes each. Subsequently, beads were incubated with blocking solution 10% BSA or FBS with 0.5% Triton X-100 for an hour. Primary antibody (Mouse anti-pro-Insulin c-peptide (Millipore, #05-1109)) was diluted in blocking solution (1:1000), added

to beads and incubated at 4°C overnight. Next day, three washes were performed with DBPS for 5 min each. Secondary Antibody (Donkey anti-Mouse, Alexa Fluor 488 (Thermo Fisher Scientific, #A-21202)) was diluted at 1:500 in DPBS and incubated at RT for 2 hours in dark. Lastly, before detection, three washes with DPBS were performed for 5 min each.

5

10

- b) **COPAS™ sorting.** Positive and negative controls (**Figure 3**), both unstained and stained, were utilized to define gating parameters, facilitating the discrimination between negative and positive bead populations. Subsequently, all stained beads from the CombiCult® experiment were subjected to screening using a particle flow sorter known as COPAS (Complex Object Parametric Analyzer and Sorter, Union Biometrica FP 2000 PRO, CSB). By utilizing the gating thresholds established by the positive and negative controls, the COPAS system analyzed the fluorescence intensity of the beads. Through this analysis, the COPAS system identified positive hits containing reprogrammed β -cells, specifically c-peptide⁺ beads.
- 15
- c) **Bead digestion and isolation of tags.** The c-peptide⁺ beads were individually digested according to Plasticell's proprietary protocol to release the fluorescence tags accumulated by the bead during the experiment.
- 20
- d) **Tag analysis.** The released tags were analyzed using FACS and the data produced were submitted and deconvoluted by proprietary CombiCult® bioinformatics program Ariadne® .
- 25
- e) **Ariadne® Bioinformatics.** Ariadne® has the capability to discern specific fluorescence signal intensities as distinct tags, thereby elucidating the pathways or combinations of four TFs responsible for reprogramming cPPs into β -cells. Furthermore, through statistical analysis, it can rank the most effective reprogramming combinations which are subsequently validated.
- 30
- f) **Validation.** 1) The validation of the top-ranked protocols started with bead-based replication of the CombiCult® results. cPPs were seeded onto matrigel-coated beads, followed by sequential delivery of TFs over four consecutive days or

simultaneous delivery of TFs in a single day, using the same media as the screening process. After an incubation period of 11 days, the beads were stained with antibody against c-peptide and screened using COPAS to identify those exhibiting c-peptide signals. The most effective combinations were determined by comparing the distribution of c-peptide fluorescence signals among negative and positive controls, as well as between sequential and simultaneous TF-delivery methods. 2) Subsequently, the ranked protocols underwent validation in spheroids, a common and natural method for obtaining β -cells. In this approach, cPPs were seeded as single cells in 10 cm dishes coated with matrigel (one dish containing 12 million cells, with two dishes per sample). TFs were then delivered sequentially over four days, followed by the formation of spheroids using 24 million cells (200 cells per spheroid). Transfection was performed as described previously; for each 10 cm dish, 10 μ g of mRNA was diluted in 1000 μ l of mRNA buffer, and 20 μ l of jetMESSENGER® was added to the mix. Media was changed regularly until day 11, as for the reprogramming experiments. At this point, some spheroids were stained with antibody against c-peptide for immunofluorescence analysis, while others were dissociated into single cells and stained with c-peptide to quantify through FACS the reprogramming efficiency as a measure of c-peptide expression. Additionally, another subset of spheroids was used to measure the secretion of c-peptide upon stimulation with elevated glucose levels.

Glucose-stimulated insulin secretion

Glucose-stimulated insulin secretion (GSIS) assays are frequently employed in research to evaluate the functional competence of differentiated and reprogrammed β -cells. These assays involve exposing these cells to different glucose concentrations, typically low glucose (2 mM) followed by high glucose (20 mM), to assess their insulin secretion response to glucose stimulation. C-peptide secreted in low and high glucose conditions was collected and quantified using an ELISA assay (Ultrasensitive C-peptide ELISA, Mercodia #10-1141-01).

Example 1: Reprogramming of adult pancreatic cells to beta cells

Zhou et al. (2008) demonstrated the reprogramming of adult pancreatic exocrine cells to beta cells testing a cocktail of 9 transcription factors, including Pdx1, Ngn3, MafA, Nkx6.1, Nkx2.2, Isl1, NeuroD1, Pax4 and Pax6. The combination which proved most effective was Pdx1, Ngn3 and MafA.

In order to test the method of the invention, a series of transcription factors was used in a split/pool experiment, and cells labelled according to the identity, sequence and timing of exposure to different transcription factors (see Fig. 4)

Accordingly, adult pancreatic exocrine cells were exposed to different combinations of the transcription factors Pdx1, Ngn3, MafA, Nkx6.1, Nkx2.2, Isl1, NeuroD1, Rfx6, Pax4 and Pax6, using a split and pool approach. Cells were screened for expression of cd49a and c-peptide, or insulin. CD49a is a marker for c-peptide expressing cells. See Fig. 5.

Transcription factor combinations leading to reprogramming were analysed computationally, and the data tabulated (Fig. 6A, B)

Pdx1, MafA and Ngn3 were confirmed as the most effective transcription factors, replicating the results of Zhou et al (Fig. 7). Rfx6 is also shown to be an important factor.

In addition, it was determined that the optimal order of exposure to the transcription factors is Pdx1, followed by Ngn3 at 4 hours, and finally a combination of Pdx1 and MafA at between 6 and 11 hours (Fig. 7).

Example 2: T lymphopoiesis

The potential of the invention to concurrently explore several cell fate acquisition experiments at once permits us explore novel questions: how do we generate the entire naïve T-lymphocyte repertoire?

T lymphocytes are fundamental actors of cellular adaptive immunity, as they use their T-Cell Receptor (TCR) to recognize foreign/pathological antigens and initiate a powerful response. The maturation of T lymphocytes happens in the thymus, where they upregulate different molecules that correlate with their functions. CD8+ T cells display cytotoxic activity, actively killing cells that positively interact with their TCR through their antigens; CD4+ cells secrete molecules that modulate the activity of other immune cells in the environment of diseased tissue, and regulatory T cells fine-tune the responses from either population in order to avoid a potentially harmful exaggerated T cell-mediated immune response. While CD8+ T cells have been successfully employed in cancer immunotherapy, harnessing their cytotoxic ability via immune-checkpoint inhibitors or by adoptive cell therapy with chimaeric antigen receptor (CAR-T), the cell therapy field at large has not yet managed to consistently differentiate CD4+ and regulatory T lymphocytes for therapeutic uses (Montel- Hagen et al. Cell Stem Cell 2019 – doi:10.1016). These represent a great opportunity for the creation of

ATMPs as they play a critical role not only in the organism's defence against cancer, but also and especially in the regulation (or in the emergence) of autoimmune diseases and immunodeficiencies. In fact, the use of CD4+ and regulatory T lymphocytes in therapy is still far from being a reality (Hippen et al. Front Immunol 2022 – doi:10.3389), owing to the

5 difficulties in expanding them or generating them from stem cell precursors.

Claims

- 5 1. A method for determining the conditions required for reprogramming of a first cell type to a second cell type, by screening transcription factor (TF) combinations in which TFs are added to the cells sequentially.
- 10 2. A method according to claim 1 comprising the steps of:
 - a. exposing a cell of a first cell type to a first TF;
 - b. removing the first TF and exposing the cell to a different TF;
 - c. optionally, repeating the foregoing step (b); and
 - d. identifying one or more cells which displays one or more markers of a second cell type and determining the identity and sequence of TFs to which the cell has been exposed.
- 15 3. A method according to claim 1 or claim 2, which comprises the steps of:
 - a. separately exposing a first plurality of test units each comprising a cell to different TFs or combinations of TFs and labelling the cells to indicate exposure to said TFs or combinations of TFs;
 - 20 b. pooling the plurality of test units, and subdividing the pool of test units to form a second plurality of test units;
 - c. separately exposing the second plurality of test units to different TFs or combinations of TFs;
 - d. optionally iteratively repeating steps (a) to (c) as required;
 - 25 e. identifying the cell type of the cells in the test units and deconvoluting the labels to identify the identity, sequence and timing of TFs to which the cells have been exposed;
 - f. optionally, using the data from step (e) to inform the choice of TFs used in step (a), and repeating steps (a) to (e) as required;
 - 30 g. identifying a cell of the second cell type, and deconvoluting the labels to thereby identify the identity and sequence of TFs required to reprogram a cell of the first cell type into a cell of the second cell type.
- 35 4. A method according to claim 3, wherein the TFs are added individually to cell units in sequential method steps.
5. A method according to claim 3, wherein sequential method steps in a procedure comprise either the addition of single TFs or the addition of combinations of TFs.
- 40 6. A method according to claim 3, wherein sequential method steps in a procedure only comprise the addition of individual TFs.
- 45 7. A method according to claim 1, wherein at least one plurality of test units comprises test units which contain at least one cell and test units which do not contain cells.

8. A method according to any preceding claim, wherein the TFs or combinations of TFs used in the first iteration of step (a) are selected from TFs known in the art to promote cell reprogramming.
- 5 9. A method according to claim 8, wherein the TFs or combinations of TFs used in the first iteration of step (c) are selected from TFs known in the art to promote cell reprogramming.
- 10 10. A method according to any preceding claim, wherein the TFs or combinations of TFs used in second and further iterations of steps (a) and (c) are selected by analysis of the data produced by step (e).
- 15 11. A method according to claim 10, wherein the data produced in step (e) are processed by an algorithm which analyses the effect of the identity and timing of administration of each TF on the reprogramming of the first cell type.
- 20 12. A method according to claim 3, wherein the TFs or combinations of TFs used in the first iteration of step (a) are selected using data produced by step(e) in a previous performance of the method.
- 25 13. A method according to any preceding claim, wherein each test unit is a single cell.
14. A method according to claim 13, wherein the cells are labelled by modifying the nucleic acid of the cell.
- 30 15. A method according to claim 14, wherein the nucleic acid of the cell is modified by genome editing or by viral integration.
- 35 16. A method according to claim 15, wherein the genome editing comprises CRISPR.
- 40 17. A method according to claim 12, wherein the test unit is a single cell, and data derived from the fate of cells exposed to different TFs are combined to correlate transcriptional changes at the single cell level with specific cell culture conditions, to generate a model which predicts the culture conditions required to generate specific cell types.
- 45 18. A method according to claim 17, wherein the model is used to generate initial TF choices for steps (a) to (c), the iterative performance of the steps resulting in an improved model.
19. A method for identifying a gene which influences a cellular process, comprising the steps of:
- a) determining the effect of one or more TFs or combinations of TFs on a cell, in accordance with any one of the preceding claims;
- b) analysing gene expression in said cells when exposed to said TFs or combinations of TFs; and

c) identifying genes which are differentially expressed when exposed to said TFs or combinations of TFs.

5 20. A method for producing a nucleic acid which encodes a gene product which influences a cellular process, comprising identifying a gene in accordance with claim 19, and producing at least the coding region of said gene by nucleic acid synthesis or biological replication.

10 21. A method for inducing a cellular process, comprising the steps of:

- a) identifying one or more genes which are differentially expressed in association with the cellular process in accordance with claim 19; and
- b) modulating the expression of said one or more genes in the cell.

15 22. A method according to claim 21, wherein modulation of gene expression in the cell comprises transfection of said one or more genes into the cell.

23. A method for selecting the starting TFs in a method according to claim 3, comprising the steps of:

20 compiling a database of TFs and cell transdifferentiating results; and

selecting from the database a set of TFs or combinations of TFs for use in a method according to claim 1.

24. A method for improving a database of TFs and cell transdifferentiating results, comprising using the database to select TFs according to claim 23, and inputting the
25 results of the method according to claim 1 into the database, thus providing further data points.

25. A system comprising a database of TFs useful for reprogramming cells, at least one computer processor and a memory in operative communication with the processor, the memory comprising instructions configuring the processor to:

- 30 (a) generate gene regulatory networks (GRN) from the database of transcription factors;
- (b) determine a candidate transcription factor for a desired reprogramming event;
- 35 (c) analyse an impact of the candidate transcription factor in cell reprogramming;
- (d) optionally, iteratively repeat steps (a)-(c); and
- output a set of optimal transcription factors.

40 26. The system according to claim 25, wherein the database comprises gene expression data.

27. A method for determining optimal transcription factors for cell reprogramming, comprising: using a computing device, curating a gene expression database related to use of TFs in cell culture; generating gene regulatory networks from a plurality of gene expression datasets; determining a candidate optimal TF; analysing an impact of the optimal TF in cell differentiation; and outputting a set of optimal TFs.
- 5