CLAIMS

[Claim 1]. A method for determining the transcription factors required for conversion of a source cell to a cell exhibiting at least one characteristic of a target cell type, the method comprising the steps of:

- determining differential expression of genes in the source and target cell types;

- determining a network score for each transcription factor (TF) in each of the source and target cell types based on the differential gene expression over at least one network, wherein the network contains information of interactions that affect gene expression;

- ranking the TFs based on a combination of network scores and differential gene expression information, thereby identifying the set of transcription factors for a conversion from a source cell to a cell exhibiting at least one characteristic of a target cell type.

[Claim 2]. A method according to claim 1, wherein a gene score is determined for each differentially expressed gene in the source and target cell types.

[Claim 3]. A method according to claim 2, wherein the gene score is a combination of the log fold change and adjusted P- value of the differential expression.

[Claim 4]. A method according to any one of claims 1 to 3, wherein the gene score is calculated using a tree-based method, preferably against a background or using Bayesian clustering.

[Claim 5]. A method according to any one of claims 2 to 4, wherein the network score is calculated for each TF by performing a weighted sum of gene scores over at least one subnetwork centered on each TF.

[Claim 6]. A method according to any one of claims 1 to 5, wherein the network contains information of protein-DNA interactions, protein-protein and/or protein-RNA interactions.

[Claim 7]. A method according to claim 6, wherein the network comprises interactions from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database.

[Claim 8]. A method according to any one of claims 1 to 7, wherein the network contains information of the interaction between transcription factors and regulatory regions of a gene.

[Claim 9]. A method according to claim 8, wherein the network comprises interactions from Motif Activity Response Analysis (MARA).

[Claim 10]. A method according to claim 8, wherein the regulatory region is a promoter region of a gene.

[Claim 11]. A method according to any one of claims 1 to 10, wherein the method further comprises the step of collecting expression data for each gene prior to determining a gene score.

[Claim 12]. A method according to any one of claims 1 to 11, wherein the method further comprises the step of removing transcriptionally redundant TFs from the ranked lists from each cell type.

[Claim 13]. A method according to any one of claims 1 to 12, wherein the method further comprises the step of creating a cell conversion landscape by arranging a plurality of cell types on a 2D plane based on their required TFs and adding a height based on the average coverage of the required genes that are directly regulated by the TFs selected.

[Claim 14]. A method according to any one of claims 11 to 13, wherein the expression data identifies the transcriptional start site for each gene.

[Claim 15]. A method according to any one of claims 11 to 14, wherein the expression data comprises expression data obtained by cap analysis gene expression (CAGE).

[Claim 16]. A method according to any one of claims 11 to 15, wherein the expression data comprises expression data taken from the FANTOM5 dataset.

[Claim 17]. A method according to claim 16, wherein the source cell type and the target cell type are each selected from any cell type described in the FANTOM5 dataset.

[Claim 18]. A method according to any one of claims 1 to 17, wherein the source cell is selected from the group consisting of dermal fibroblasts, epidermal keratinocytes, embryonic stem cells (hESCs),induced pluripotent stem cells (iPSCs), mesenchymal cells, monocytes and cardiac fibroblasts.

[Claim 19]. A method according to any one of claims 3 to 18, wherein the gene score combines the log fold change and adjusted P- value using Equation 1:

Equation 1: $$G_x^s = |L_x^s|(-\log_{10} P_x^s)$$

wherein L^s_x is the log-fold change of gene x in sample s and P^s_x is the adjusted p-value of gene x in sample s.

[Claim 20]. A method according to any one of claims 1 to 19, wherein a network score (N^sx,n) for a transcription factor x in cell type s is calculated over a network n using Equation 2:

Equation 2: $$N_{x,n}^s = \sum_{r \epsilon V_x}^s G_r^s \cdot \frac{1}{L_{r,n}} \cdot \frac{1}{O_{r,n}}$$

wherein xϵVx is each gene (r) in the set of nodes (Vx) that make up the local subnetwork of TF x,

Lr,n is the number of steps r is away from x in the network n,

G^sr is a gene score for gene r, and

Or,n is the degree of the parent of r in the network n.

[Claim 21]. A method according to any one of claims 1 to 20, wherein identifying the set of transcription factors for a conversion from a source cell type to a cell exhibiting at least one characteristic of a target cell type comprises comparing the ranked list for the target cell type and a set of genes identified as expressed in the source cell type, wherein if a

TF from the target cell type list is already expressed in the source cell type then it is removed from the ranked list.

[Claim 22]. A method according to any one of claims 1 to 21, wherein the target cell is selected from the group consisting of chondrocytes, hair follicles, CD4+ T cells, CD8+ T cells, NK -cells, haemopoeitic stem cells (HSC), mesenchymal stem cells (MSC) of adipose, mesenchymal stem cells (MSC) of bone marrow, oligodendrocytes, oligodendrocyte precursors, skeletal muscle cells, smooth muscle cells fetal cardiomyocytes; endothelial cells, astrocytes, keratinocytes and epithelial cells.

[Claim 23]. A method according to any one of claims 1 to 22, wherein the method further comprises the step of increasing the amount of the transcription factors, determined as being required for conversion of a source cell type to a target cell type in the source cell type.

[Claim 24]. A method according to any one of claims 1 to 23, wherein determining a network score for each transcription factor (TF) in each of the source and target cell types based on the differential gene expression over at least one network, comprises determining a first network score for each TF over a first network and a second network score for each TF over a second network, wherein the first network comprises protein-DNA interactions between TFs with known binding sites in the promoter region of a gene, and the second network comprises protein-protein, protein-DNA, and protein-RNA interactions.

[Claim 25]. A method according to any one of claims 1 to 24, wherein the at least one characteristic of the target cell is up-regulation of any one or more target cell markers and/or change in cell morphology.

[Claim 26]. A method for determining the transcriptions factors required for conversion of a source cell to a cell exhibiting at least one characteristic of a target cell type the method comprising the steps of:

- collecting expression data for each gene in the source cell type and target cell type;

- calculating the differential expression against a tree-based background for each gene in each sample then combine the log fold change and adjusted P- value to a gene score;

- calculating a network score for each TF by performing a weighted sum of gene scores over at least one subnetwork centered on each TF;

- ranking the TFs based on a combination of gene and network scores;

- calculating the set of transcription factors for a conversion between any two cell types based on comparisons of ranked lists from each cell type; and optionally

- removing transcriptionally redundant TFs from the lists.

thereby determining the transcription factors required for conversion of a source cell type to a target cell type..

[Claim 27]. A method for identifying an agent useful for promoting the conversion of a source cell type to a cell exhibiting at least one characteristic of a target cell type, the method comprising the steps of:

- determining one or more transcription factors required for conversion of a source cell type to a cell exhibiting at least one characteristic of a target cell type by any method according to any preceding claim;

- screening one or more candidate agents for the ability to increase the amount of the one or more transcription factors required for conversion of a source cell type to a cell exhibiting at least one characteristic of a target cell type;

wherein an agent that increases the amount of the one or more transcription factors is an agent useful for promoting the conversion of a source cell type to a cell exhibiting at least one characteristic of a target cell type.