



Exploiting Epigenetics to Systematically Optimise Culture Conditions for Cellular Therapies

One of the main challenges of cell therapies is the maintenance and/or expansion of the required therapeutic phenotype *in vitro*. Moreover, to meet both the traceability and safety requirements for a clinical-grade therapeutic, cells need to be cultured in chemically defined conditions. Over the last two decades, scientists have sequenced the genome and epigenome of all known cell types in the human body, in addition to mapping hundreds of possible protein-protein interactions. The availability of biological data at this resolution and scale has enabled the development of computational tools capable of addressing the challenges associated with the development of scalable cell therapies.

One such platform, recently published in *Cell Systems*¹, systematically predicts the best components for chemically defined culture conditions for cell conversion, maintenance, and expansion. This technology has been used to predict soluble factors that replace Matrigel in the derivation and maintenance of clinically relevant cell types, such as astrocytes and cardiomyocytes¹. By allowing the integration of new datasets, platforms such as this could facilitate the systematic identification of culture conditions, differentiation and transdifferentiation stimuli for future cell therapy applications.

Consortia, Exponential Data Growth and the Rise of Epigenetics
In February 2003, researchers working as part of an international

research consortium, known as the Human Genome Project², completed the 13-year-long hunt for a draft sequence of the human genome^{3,4}. The consortium not only sequenced 94% of the human genome and generated a wealth of data about the architecture, development, and variation of the genomic landscape, but also established new standards for international research collaborations and the accessibility of the results. However, elucidating the exact sequence of the 3.2 billion DNA base pairs encoded within the human genome, left researchers with more questions than answers.

The need to understand which of these base pairs formed functional units, either as templates for transcription or regulators of gene expression, subsequently sparked the foundation of several international research collaborations aiming to pick up where the Human Genome Project left off.

In September 2003, the Encyclopedia of DNA Elements or ENCODE project⁵, and its sub-project, the GENCODE consortium⁶, embarked on a quest to identify all functional elements in the human and mouse genomes across four consecutive phases. The ENCODE project first used microarray-based methods to reveal basic organisational features in 1% of the genome (ENCODE 1), then switched to sequencing-based technologies that interrogated the whole human genome and transcriptome (ENCODE 2 and 3)⁷. ENCODE 3 introduced new types of assays to interrogate chromatin interaction and chromatin conformation, adding to the growing pool of data on epigenetic regulation of the human genome. The results of ENCODE 3, published in





July last year⁸ offered the first version of an encyclopaedia of functional elements in the human genome, containing 20,225 protein-coding and 37,595 non-coding genes, 2,157,387 open chromatin regions, 750,392 regions with histone modifications, 1,224,154 regions bound by transcription factors and chromatin-associated proteins, 845,000 RNA subregions occupied by RNA-binding proteins, and more than 130,000 long-range interactions between chromatin loci.

Now in its fourth phase, the ENCODE project, amongst a multitude of other consortia and independent research, have greatly enriched and enhanced our view of the human genome from its original mapping in 2003, building on and rapidly expanding our understanding of the organisation and function of the human genome and epigenome. New consortia leveraging the latest technologies will continue to expand on this work; for example, the Human Cell Atlas⁹ uses single-cell RNA sequencing technology to create comprehensive reference maps of all human cells, which will add functional elements with high tissue- and/or cell-specificity to our understanding of the human genome.

Considerable growth in large-scale data is compelling scientists and bioinformaticians to develop novel methodologies for efficiently analysing and modelling the data generated. The analysis of such datasets is also not limited to scientists within consortia such as the ENCODE sub-project, the Encode Data Analysis Center, but is also made available to myriad independent researchers and affiliated commercial entities,

increasing the likelihood of their application in the development of cellular therapies. Two complementary examples, developed separately over more than a decade of multi-national research collaborations, utilise consortium and other data from the human genome and transcriptome to direct cellular conversion¹⁰ and the maintenance of cell identity¹.

Can Epigenetics be Used to Predict Culture Conditions?

Understanding the development of cells and their conversion from one type to another is one of the great challenges in biology. Cellular conversion happens throughout development, from the first cell divisions of an embryo to the activation of cells in response to an immune stimulus. How can modern medicine draw on this knowledge to develop cell types for therapeutic applications? While the promise of cell therapies is enormous, most protocols do not make it to the clinic, fail in trials, or are not commercially viable, due to issues with efficacy, safety, and scalability. New solutions are required to improve these outcomes and address the challenges associated with the efficacy and manufacturability of cellular therapies.

Next-generation sequencing of the genome and epigenome, alongside high-throughput data approaches, are playing a key role in the identification of gene regulators (e.g. transcription factors and epigenetic modifiers), and soluble factors (e.g. cytokines and matrix proteins). These factors, once identified, can increase the generation, maturation, and maintenance of *in vitro*-derived therapeutic cellular products. Recent progress has seen novel technologies, such as Rackham *et al.* (2016)¹⁰

Structural and Physicochemical Characterization Services



COMPARE



CHARACTERIZE



IDENTIFY



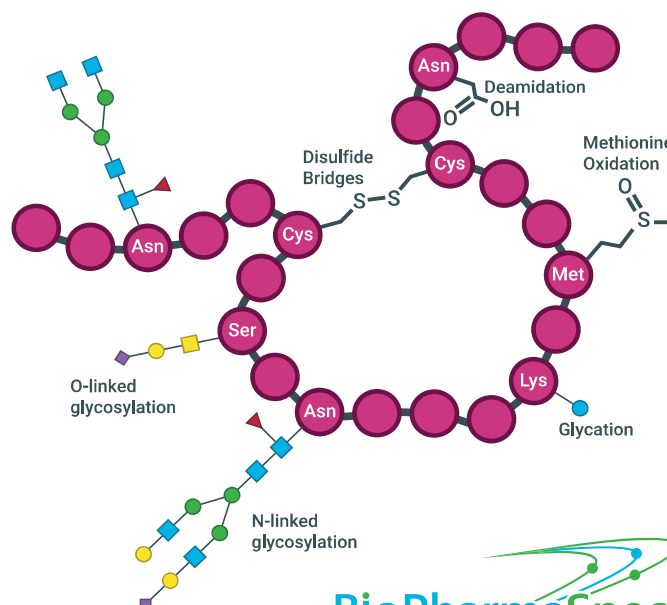
QUANTIFY

Structural

- Primary *de novo* sequencing
- N and C Terminal Sequencing
- Amino Acid Analysis
- Peptide mapping
- Disulfide bridge analysis
- Intact molecular weight analysis
- Monosaccharide analysis
- Oligosaccharide analysis
- Glycosylation site analysis

Physicochemical

- icIEF
- CE-SDS
- Circular Dichroism
- FTIR
- Intrinsic Fluorescence
- μ -DSC
- Dynamic and Static Light Scattering
- SV-AUC and SEC-MALS
- 1D and 2D NMR





and Kamaraj *et al.* (2020)¹ deploy next-generation sequencing, gene regulatory and epigenetic network data to make unbiased predictions: using transcriptomics data to predict the key regulatory switches, such as an optimal combination of transcription factors, required for cell conversion¹⁰, and epigenetic data to rank the most important genes for each cell type, including soluble factors required for enhancing directed cell differentiation and maintenance of cell identity in chemically defined media¹ (Figure 1).

type, called cell identity genes. The broader the breadth of the H3K4me3 peak, the more specific that gene is to a particular cell type (Figure 2). Such correlation is not replicated in gene expression levels; although cell identity genes are usually highly expressed in a particular cell type, as are housekeeping genes, whose functions are generally universal in all cell types. This striking correlation between H3K4me3 breadth and cell identity enables the prioritisation of proteins and signalling pathways that are critical for the survival of a given cell type.

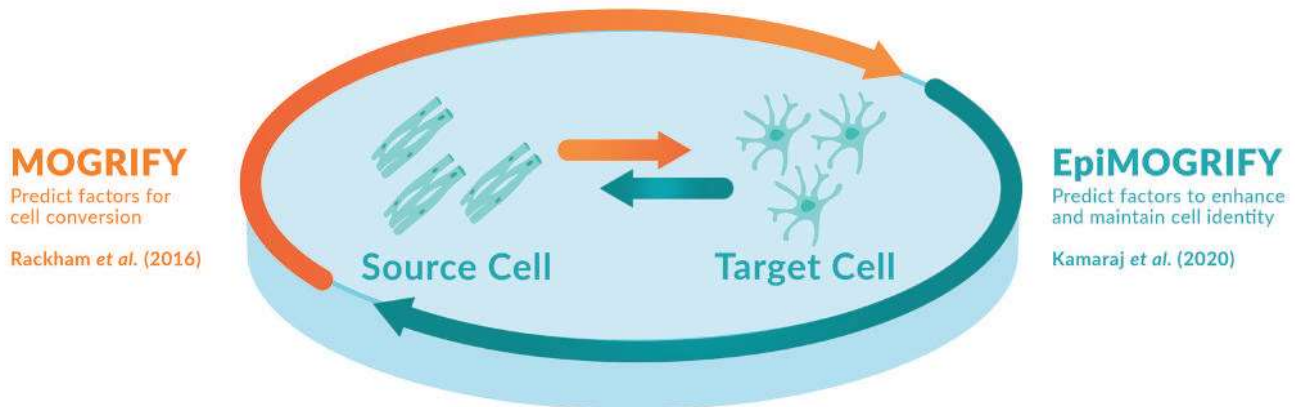


Figure 1. Systematically predicting ligands, receptors and other factors to enhance cell identity, both for cell maintenance and conversion

Uncovering the ideal culture conditions for the derivation of cells *in vitro*, both from pluripotent stem cell differentiation and for direct cell transdifferentiation, is a demanding task. Typically, achieving culture conditions *in vitro* that mimic the *in vivo* microenvironment relies on domain knowledge, and trial-and-error of several combinations of undefined culture medium and matrices containing different components (e.g. Matrigel) and the addition of different sets of cell-type-specific factors, such as growth factors and extracellular matrices. Leveraging epigenetic data, the computational tool dubbed EpiMOGRIFY, published by Kamaraj *et al.* (2020)¹, has been developed as a systematic means of identifying culture conditions that allow for both maintenance and conversion of any human cell type.

Kamaraj *et al.* describe an approach that incorporates data from over 100 human cell types available from the ENCODE and Roadmap Epigenomics consortia. Data from both projects were integrated into the same analysis pipeline to evaluate how different epigenetic marks correlate with gene expression.

Interestingly, the research observed that the breadth of the H3K4me3 peak at the gene promoter region correlates with the expression of genes that are specific to a particular cell

Following the discovery, the authors hypothesised that this could be leveraged to identify signalling molecules such as receptors and ligands that are essential for cell maintenance conditions. A third database was integrated: a receptor-ligand database previously described and recently updated^{11,12}. With this integration, the authors were able to systematically rank receptor-ligand pairs, enabling prediction of the microenvironment required to maintain and support the conversion of target cell types *in vitro*.

To validate the approach, the predicted ligands required for maintenance of primary astrocytes and cardiomyocytes, and differentiation of these cells from pluripotent stem cells or neural progenitors, were tested. The predicted ligands for each cell type were shown to be sufficient to maintain the primary cells and differentiate pluripotent stem cells in the absence of Matrigel. In the case of astrocytes, one of the predicted ligands, LN1, was able to replace Matrigel alone, creating a simple chemically defined astrocyte differentiation and maintenance media. Of most interest, it was observed that a marker of mature cardiomyocytes, cardiac troponin T, was expressed in 45% of generated cardiomyocytes when the five predicted ligands were present, in comparison to 5% when Matrigel was used. Further validation has revealed a similar trend, showing that the addition of predicted ligands increased the percentage of

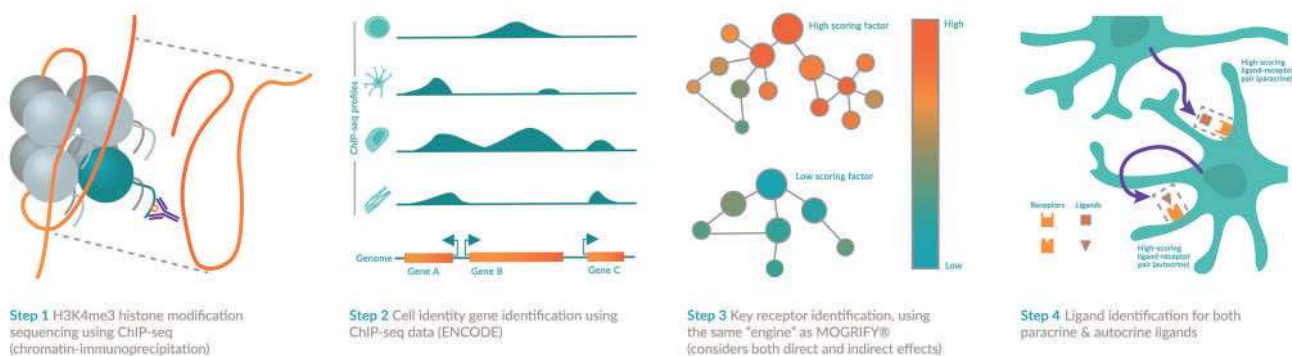


Figure 2. Integrating epigenetic and protein-protein interaction data to predict factors for cell culture



CD13/CD82-expressing cells five-fold by day 12. Even more impressively, the expression of maturation markers increased 80-fold after the addition of systematically predicted ligands by day 21¹³ (Figure 3).

6. www.encodegenes.org, visited on 15 Jan 2021.
7. Abascal, F. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nat.* 583 (7818), 699–710 (2020).
8. Abascal, F. et al. Perspectives on ENCODE. *Nat.* 583 (7818), 693–698 (2020).
9. <https://www.humancellatlas.org/>, visited on 15 Jan 2021.

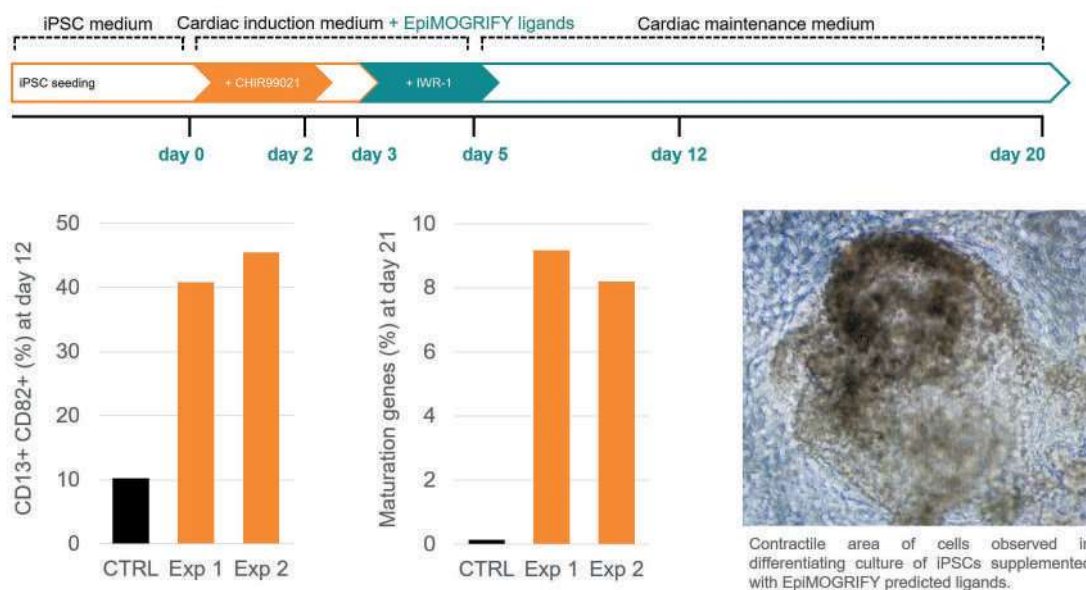


Figure 3. Enhancement of cardiomyocyte differentiation from iPSC

Synergistic Input of Transcription Factors and Signalling Pathways for Cellular Reprogramming

The use of computational power to analyse and model transcriptomic and epigenetic datasets is driving a revolution in how we view developmental biology and the production of *in vitro* cell therapies. The combinatorial use of computational tools predicting key transcriptomic switches, and chemically defined culture conditions, will enable the systematic generation and maintenance of clinically valuable cell types *in vitro*. Moreover, with the adoption of technologies enabling the sequencing of the transcriptome and epigenetic marks at a single-cell level, scientists are starting to gather information about defined sub-sets of cell types that were unknown or hard to capture. Combined, it is possible to imagine a future where the generation and maintenance of cell types is no longer a black-box exercise, being instead a predictable and systematic process powered by computational methodologies.

Acknowledgements

The authors would like to acknowledge Wei Wei, Senior Bio-informatician, and David Anderson, Principal Scientist, from Mogrify Limited, for the help in the manuscript and figure preparation.

REFERENCES

1. Kamaraj, U. et al. EpiMogrify Models H3K4me3 Data to Identify Signaling Molecules that Improve Cell Fate Control and Maintenance, *Cell Syst.* 1–14 (2020).
2. <https://www.genome.gov/human-genome-project>, visited on 15 Jan 2021.
3. Lander, E. S. et al. Initial sequencing and analysis of the human genome: International Human Genome Sequencing Consortium. *Nat.* 412 (6846), 565–566 (2001).
4. Collins, F.S. et al. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nat.* 431 (7011), 931–945 (2004).
5. www.encodeproject.org, visited on 15 Jan 2021.

10. Rackham, O. J. L. et al. A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* 48 (3), 331–335 (2016).
11. Denisenko, E. et al. Predicting cell-to-cell communication networks. *Nat. Commun.* 1–11 (2020).
12. Ramilowski, J. A. et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* 6 (2015).
13. Skelton, R. J. P. et al. SIRPA, VCAM1 and CD34 identify discrete lineages during early human cardiovascular development. *Stem Cell Res.* 13 (1), 172–179 (2014).



Dr. Joachim Luginbühl

Dr. Joachim Luginbühl is a Senior Scientist at Mogrify Limited. After completing his PhD at the children's hospital in Zurich in a laboratory focused on tissue-engineered skin substitutes, he moved to Japan to join the laboratory of Jay W. Shin at RIKEN Yokohama and develop novel methods to study the direct reprogramming of human fibroblasts into different types of neurons.

Email: joachim.luginbuhl@mogrify.co.uk



Dr. Rodrigo Santos

Dr. Rodrigo Santos is the Director of Cell Technologies at Mogrify Limited. He completed his PhD at the Stem Cell Institute, University of Cambridge, focused on the investigation of the biological mechanisms underlying the generation of induced pluripotent stem cells. Prior to Mogrify, Rodrigo was the Head of Technology at Bit Bio, and Principal Scientist at Horizon Discovery.

Email: rodrigo@mogrify.co.uk