HipSci: a large collection of highly characterized induced pluripotent stem cells (iPSCs)

Laura Gatcombe, Rosa Loffredo, Jessica Pharoah, Bhavna Sidhu and Sharon Bahia

European Collection of Authenticated Cell Cultures (ECACC), Culture Collections, National Infection Service, Public Health England, Porton Down, SP4 0JG

Public Health England

Protecting and improving the nation's health

INTRODUCTION

Stem cells have the ability to differentiate into any type of cell in the human body – they are pluripotent. Induced pluripotent stem cells (iPSCs) are mature adult cells that have been reprogrammed back into a pluripotent state. Since their discovery over a decade ago, iPSCs have become optimised models for the study of human diseases. iPSCs can be produced from a patient's blood samples and skin biopsies, allowing patient-specific iPSCs to be generated. These can be differentiated into disease relevant tissues and has the potential to lead to the development of personalised treatments and medicines [1].

The Human Induced Pluripotent Stem Cell Initiative (HipSci) has produced a large, high-quality reference panel of iPSCs, using a standardised and well-defined experimental pipeline (Figure 1) [2]. The collection includes hundreds of cell lines from patients with inherited genetic diseases e.g. Bardet-Biedl syndrome, Macular Dystrophy and Kabuki Syndrome (Table 1). As well as collecting samples from patients with diseases, HipSci is the first collection to also include material from healthy volunteers. Whilst a lot of iPSC research focuses on the effect of genotype on a diseased phenotype, it is just as important to study common genetic variation in healthy individuals and it's effect on cell behaviour, for example [3]. All of the iPSCs in the HipSci panel are highly characterised and each have a large amount of data available on the HipSci website (http://www.hipsci.org/), including RNA-seq data and cellular phenotyping data.

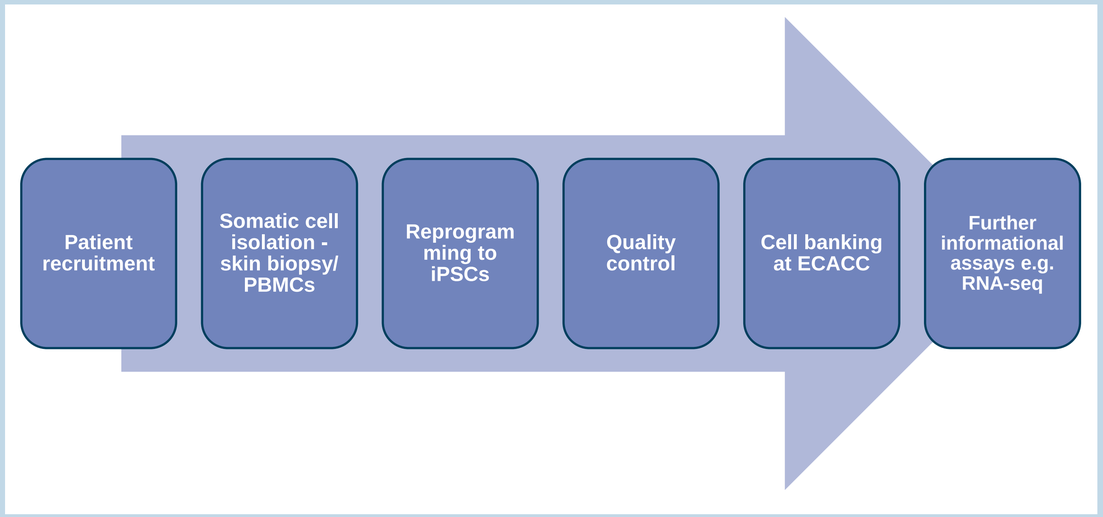

FIGURE 1 Diagram of the HipSci experimental pipeline.

Disease Status	Number of iPSCs available
Normal	496
Bardet-Biedl syndrome	61
Monogenic diabetes	44
Usher syndrome/ congenital eye defects	33
Hypertrophic cardiomyopathy	27
Heredity spastic paraplegia	26
Batten disease	25
Retinitis pigmentosa	24
Hereditary cerebellar ataxia	22
Alport syndrome	16
Primary immune deficiency	14
Bleeding and platelet disorders	13
Genetic macular dystrophy	13
Rare genetic neurological disorder	8
Congenital hyperinsulism	7
Kabuki syndrome	6

TABLE 1 Summary of diseases represented in the HipSci collection (only lines that have been selected for banking)

HIPSCI DATA

The HipSci data portal provides the research community with a collated view of the HipSci collection, allowing them to search for and discover lines that they may want to purchase. Users can filter their search according to disease state, assay type, banking status and source material etc. Users can also view, download and analyse multiple data files for each cell line. The HipSci data is divided into two main groups:

1) Quality control data is collected on all successfully reprogrammed iPSCs from each donor, in addition to the originating somatic cells. The QC assays performed assess the pluripotency and the genomic stability of the lines. The outcome of these tests determines which iPSCs are suitable for banking at ECACC.

- **Expression array** - the pluripotency of the candidate iPSCs and the originating somatic cells are tested using an expression assay or RNA-seq. The results are inputted into Pluritest which compares the expression data collected to a reference set of 450 lines to assess the level of expression of pluripotency markers. It uses a 'Pluripotency Score' and 'Novelty Score' to distinguish iPSCs from other cell types (Figure 2) [4].
- **Genotyping array** – used to compare the genetic stability of the iPSCs and somatic cells. The genotype calls are compared to each other to assess genomic integrity (Figure 3).

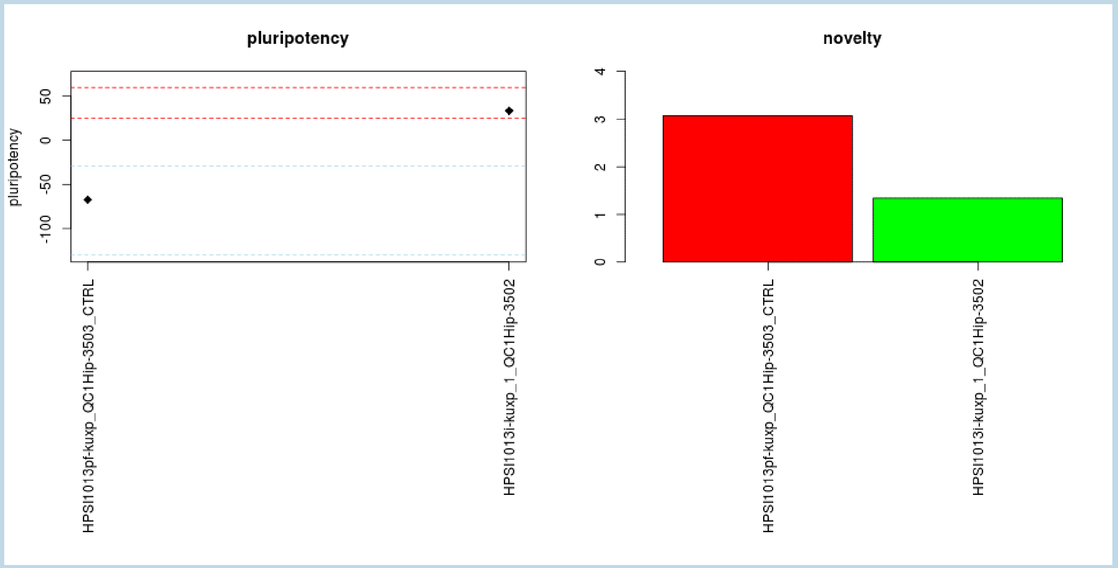

FIGURE 2 PluriTest plots for HPSI1013i-kuxp_1 and originating somatic cell (CTRL) from the HipSci data browser. Pluripotency score = 33.4, Novelty score = 1.35. The 'Pluripotency Score' gauges how pluripotent a sample is, by comparing the gene expression profile to the reference panel of human pluripotent stem cells (high score for iPSC). The 'Novelty Score' is calculated based on the detection of gene expression patterns associated with differentiated cells (low score for iPSC).

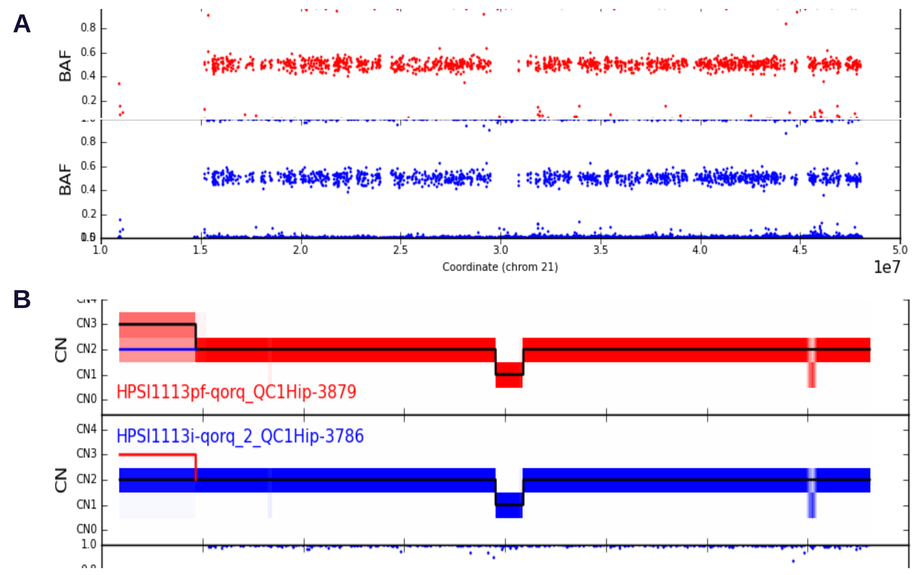

FIGURE 3 Copy number variation check for HPSI1113i-qorq_2 (chr 21) (A) In the plots, each dot corresponds to one marker in the fibroblast (red) or iPSC (blue) line. Homozygous reference genotypes BAF=0, heterozygous genotypes BAF=0.5 and homozygous alternate genotypes BAF=1. Markers which significantly deviate are indicative of non-diploid copy number state. **(B)** Shows transitions between copy number states in the black line, with the thick blue/red background of the black line representing the confidence of the prediction.

2) Characterisation data is collected from iPSCs that have been selected for banking. These lines are subjected to several more genomic, proteomic and phenotyping assays:

- **RNA-seq:** uses next-generation sequencing to analyse the expression of the whole transcriptome allowing the detection of gene expression levels, gene fusion events and single nucleotide variants.
- **Whole exome sequencing:** analyses the protein coding regions of the genome. This analysis can identify coding variants that may have the potential to cause disease
- **DNA methylation**: methylation profiling by array is used to probe the epigenetic state of the iPSCs at single nucleotide resolution.
- **Whole genome sequencing:** a small selection (~150 donors) of iPSCs and their primary somatic cells have been deep whole genome sequenced at 30x coverage.
- **Proteomic mass spec:** used to assay the proteome of the iPSC lines. This data has been analysed on the Encyclopaedia of Protein Dynamics.
- **Cellular phenotyping:** studying how iPSCs respond to chemical, physical and biological stimuli. Novel assays and artificial stimuli are being used to analyse iPSC behaviour in different microenvironments. This data contributes to HipSci's research into the relationship between (epi)genetic and phenotypic variance.

ECACC's ROLE IN THE HIPSCI INITIATIVE

HipSci iPSCs are banked, QC tested and made available to order through the European Collection of Authenticated Cell Cultures (ECACC) - https://www.phe-culturecollections.org.uk/collections/ecacc.aspx. (Figure 4). ECACC is one of the largest cell line depositories in the world and supplies the research community with authenticated and quality-controlled cell lines, nucleic acids and iPSCs. After the iPSCs have been banked at ECACC, they undergo standard QC testing. The standard QC tests include mycoplasma detection by culture isolation, Hoechst DNA staining and PCR, along with culture testing for contaminant bacteria, yeast and fungi.

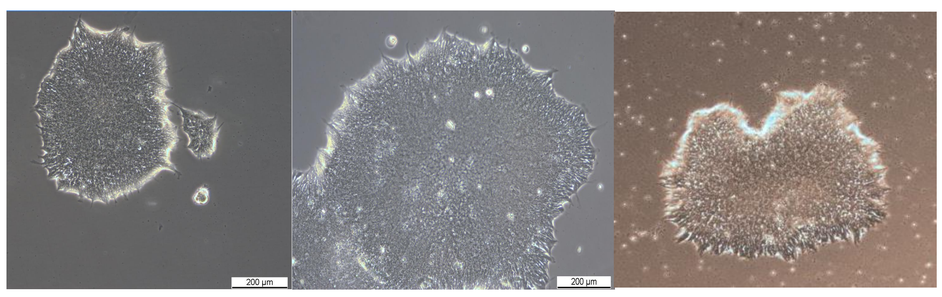

FIGURE 4 Morphology of feeder-free HPSI0514-fiaj_1 and HPSI0913i-eika_2 in culture – cell morphology is graded in an A-D system. 'A' defining well rounded, compacted colonies with smooth and defined edges and 'D' defining irregular colonies with no obvious areas of compaction and high differentiation levels.

CONCLUSION

~~As shown in this poster,~~ What makes the HipSci collection stand out is the extensive catalogue of assay data available for each cell line. In addition to viability, morphology, sterility and authentication testing, a variety of additional assays have also been performed on these iPSC lines. These assays include genotyping and expression arrays, RNA-seq, DNA methylation analysis and proteomic mass spectrometry. All of this characterisation information is available in the data browser on the HipSci website, making it an easily accessible global resource for the research community.

ACKNOWLEDGEMENTS

The HipSci Initiative is funded by the Wellcome Trust and the UK Medical Research Council.

REFERENCES

[1] Ebert AD, Liang P, Wu JC. Induced pluripotent stem cells as a disease modeling and drug screening platform. *J Cardiovasc Pharmacol*. 2012;60(4):408–416. doi:10.1097/FJC.0b013e318247f642

[2] Kilpinen H, Goncalves A, Leha A, et al. Common genetic variation drives molecular heterogeneity in human iPSCs [published correction appears in Nature. 2017 Jun 29;546(7660):686]. *Nature*. 2017;546(7658):370–375. doi:10.1038/nature22403

[3] Streeter I, Harrison P.W, Faulconbridge A, et al. The human-induced pluripotent stem cell initiative-data resources for cellular genetics. *Nucleic Acids Res*. 2017;45(D1):D691–D697. doi:10.1093/nar/gkw928

[4] Müller FJ, Schuldt BM, Williams R, et al. A bioinformatic assay for pluripotency in human cells. *Nat Methods*. 2011;8(4):315–317. doi:10.1038/nmeth.1580