PDF files:

Supplementary Information

This file contains supplementary information and methods.

Excel files:

Supplementary Table 1

Sample meta data for the HipSci cell lines used in this publication. This is a subset of HipSci's full catalogue of cell lines and data, which can be queried at http://www.hipsci.org/lines.

Supplementary Table 2

CNA results. (a) CNA locations **(b)** Significance of CNA recurrence over 200 kb genome windows **(c)** Properties of the recurrent CNAs, including: peak region, overlap with chromatin fragile sites, *cis* (same chromosome from the CNA) and *trans* (different chromosome) regulated genes (i.e. genes differentially expressed between copy-number 2 and 3 lines), and top candidate genes (identified as described in the main text). **(d)** Genome-wide association of copy numbers at recurrent CNAs with gene expression **(e)** Pathway enrichment analysis of genes regulated in *trans* by the chromosome 17 recurrent CNA region.

Supplementary Table 3

Gene expression variance components analysis. Fraction of variance explained by the factors considered for each expression array probe.

Supplementary Table 4

iPSC eQTL results. (a) eGene level summary of the cis-eQTLs discovered with different sample sets in this study. **(b)** eQTL results for primary and secondary lead eQTL variants of HipSci RNA-seq iPSC eGenes at FDR < 5% (N = 6,631). Primary and secondary eQTLs are defined by the column 'primary_eQTL'. The column 'iPSC_specific' defines whether the eQTL is iPSC-specific. Columns 'N_proxies_used' and 'proxy_positions' give the total number and positions of proxy variants that were tested in the tissue-specific analysis. Additionally, the column 'overlaps_CNA' indicates whether the eQTL lead variant overlaps with a recurrent iPSC CNA.

Supplementary Table 5

Tissue information. (a) Description of the tissue data used in this study to define tissue-specific eQTLs (GTEx V6p, HipSci), including the embryonic origin of each tissue and number of tissue-specific eQTLs identified for each tissue. **(b)** Summary of iPSC eQTL replication tests

in the tissue-specific analysis, showing for each replication tissue how often proxy variants ('ld_buddy', 'best_proxy') were tested instead of the same lead variant ('same_as_lead').

Supplementary Table 6

iPSC eQTL overlap with disease-associated variants. (a) All disease-associated variants in the NHGRI-EBI GWAS catalogue (release 2016-04-10) which are tagged by an iPSC eQTL (lead variant or $r^2 > 0.8$ proxy). For proxy matches, all eQTLs for which the variant is a proxy ($r^2 > 0.8$) are shown. **(b)** Disease-associated variants in the GWAS catalogue that are lead eQTL variants in iPSCs (subset of **(a)**). For each variant, the number of high-LD proxies it has is listed ('N_HIGH_LD_PROXIES'). **(c)** Individual traits in the GWAS catalogue for which iPSC eQTLs show a significant enrichment (BH-adjusted empirical $P < 0.05$, derived from 100 random sets of matched variants; Methods). Shown are traits with minimum five variants tagged by iPSC eQTLs. **(d)** Results of the colocalisation analysis for 14 traits.