ARTICLE

doi:10.1038/nature13200

Domains of genome-wide gene expression dysregulation in Down's syndrome

Audrey Letourneau[1]*, Federico A. Santoni[1]*, Ximena Bonilla[1], M. Reza Sailani[1], David Gonzalez[2], Jop Kind[3], Claire Chevalier[4], Robert Thurman[5], Richard S. Sandstrom[5], Youssef Hibaoui[6], Marco Garieri[1], Konstantin Popadin[1], Emilie Falconnet[1], Maryline Gagnebin[1], Corinne Gehrig[1], Anne Vannier[1], Michel Guipponi[1], Laurent Farinelli[7], Daniel Robyr[1], Eugenia Migliavacca[1,8], Christelle Borel[1], Samuel Deutsch[9], Anis Feki[6], John A. Stamatoyannopoulos[5], Yann Herault[4], Bas van Steensel[3], Roderic Guigo[2] & Stylianos E. Antonarakis[1,10]

Trisomy 21 is the most frequent genetic cause of cognitive impairment. To assess the perturbations of gene expression in trisomy 21, and to eliminate the noise of genomic variability, we studied the transcriptome of fetal fibroblasts from a pair of monozygotic twins discordant for trisomy 21. Here we show that the differential expression between the twins is organized in domains along all chromosomes that are either upregulated or downregulated. These gene expression dysregulation domains (GEDDs) can be defined by the expression level of their gene content, and are well conserved in induced pluripotent stem cells derived from the twins' fibroblasts. Comparison of the transcriptome of the Ts65Dn mouse model of Down's syndrome and normal littermate mouse fibroblasts also showed GEDDs along the mouse chromosomes that were syntenic in human. The GEDDs correlate with the lamina-associated (LADs) and replication domains of mammalian cells. The overall position of LADs was not altered in trisomic cells; however, the H3K4me3 profile of the trisomic fibroblasts was modified and accurately followed the GEDD pattern. These results indicate that the nuclear compartments of trisomic cells undergo modifications of the chromatin environment influencing the overall transcriptome, and that GEDDs may therefore contribute to some trisomy 21 phenotypes.

Down's syndrome results from total or partial trisomy of chromosome 21. It is the most frequent live-born aneuploidy affecting 1 in 750 infants. Down's syndrome patients are characterized by a cognitive impairment as well as muscle hypotonia, dysmorphic features, Alzheimer's disease neuropathology or congenital heart defects[1]. The severity and the incidence of those phenotypes are variable within the Down's syndrome population[1]. Among the possible causes, the genetic (or epigenetic) background of each individual may contribute to this phenotypic variability.

It is likely that most of the Down's syndrome phenotypes are related to alteration of gene expression due to the supernumerary copy of chromosome 21 (HSA21). According to the 'gene dosage effect' hypothesis, some Down's syndrome features could be directly explained by the dosage imbalance of genes on HSA21 (refs 2–4). Additionally, the phenotypes may also be due to the presence of extra DNA material[3,5]. Understanding the genomic determinants that contribute to the different phenotypes is a major objective in Down's syndrome research[6].

Several Down's syndrome mouse models have been created to mimic the gene expression changes observed in humans. The models are based either on translocation or on duplication of syntenic regions between HSA21 and segments of mouse chromosomes (MMUs) 10, 16 and 17. Among them, the Ts65Dn mouse model has been extensively used to study the molecular mechanisms of the features of Down's syndrome[7]. Ts65Dn mice harbour a translocation of MMU16 and exhibit some of the Down's syndrome phenotypes[8]. Other mouse models are based on the triplication of single candidate genes such as *SIM2* or *DYRK1A*. These models have emphasized the role of individual HSA21 genes in specific phenotypes, in particular the cognitive impairment[9,10].

Because Down's syndrome is probably due to gene expression disturbances, the investigation of the molecular mechanisms that underlie the phenotypic consequences requires an understanding of the transcriptome differences in trisomic cells and tissues. Several studies have explored the changes of gene expression between trisomic individuals and controls[11–14]. However, the extensive natural gene expression variation occurring in both normal and Down's syndrome individuals[15,16] complicates the identification of changes related to trisomy 21 per se. To assess the perturbations of gene expression in Down's syndrome without genetic variation among the samples, we studied a pair of monozygotic twins discordant for trisomy 21 (ref. 17). The use of these samples eliminates the bias of genome variability and thus most of the transcriptome differences observed are probably related to the supernumerary HSA21. Notably, we have found that the differential gene expression between the trisomy 21 discordant twins is organized in domains along all the chromosomes. We show that those domains are conserved in the Ts65Dn mouse model and correlate with the previously described lamina-associated and replication time domains. The study of histone mark profiles confirmed the role of chromatin modifications in the gene expression changes

[1]Department of Genetic Medicine and Development, University of Geneva Medical School, University Hospitals of Geneva, 1211 Geneva, Switzerland. [2]Center for Genomic Regulation, University Pompeu Fabra, 08003 Barcelona, Spain. [3]Division of Gene Regulation, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. [4]AneuPath 21, Institut de Génétique Biologie Moléculaire et Cellulaire, Translational medicine and Neuroscience program, IGBMC, ICS, PHENOMIN, CNRS, INSERM, Université de Strasbourg, UMR7104, UMR964, 1 rue Laurent Fries, 67404 Illkirch, France. [5]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [6]Stem Cell Research Laboratory, Department of Obstetrics and Gynecology, Geneva University Hospitals, 1211 Geneva, Switzerland. [7]FASTERIS SA, 1228 Plan-les-Ouates, Switzerland. [8]Swiss Institute of Bioinfomatics, 1211 Geneva, Switzerland. [9]DOE Joint Genome Institute, Walnut Creek, California 94598, USA. [10]iGE3 Institute of Genetics and Genomics of Geneva, 1211 Geneva, Switzerland.

*These authors contributed equally to this work.

observed between the twins. Altogether, our results suggest a new molecular mechanism in the regulation of gene expression in trisomic cells.

Differential expression is organized in domains

We used messenger RNA sequencing to study the transcriptome of fetal skin primary fibroblasts derived from both the trisomic (T1DS) and the normal (T2N) twin, in four replicates for each. In total, 63–157 million 100-base-pair (bp) paired-end reads were generated from each sample and mapped with GEM[18]. The normalized gene expression (RPKM, reads per kilobase per million) was compared between the twins. We used EdgeR[19] to evaluate the differences of gene expression and found 182 genes (including 42 long non-coding RNAs (lncRNAs)) significantly differentially expressed between the twins (false discovery rate (FDR) <0.05) (Supplementary Table 1). The gene ontology analysis[20] revealed a reduced expression of secreted proteins involved in signalling (adjusted $P = 1.2 \times 10^{-8}$) and in particular those involved in cytokine–cytokine receptor interaction pathways (adjusted $P = 8 \times 10^{-4}$) and inflammatory response (adjusted $P = 3.8 \times 10^{-4}$). This confirms that trisomic cells may have a defective cell signalling system contributing to the impairment of the immune system[21].

We then focused on general transcriptomic changes occurring in trisomic cells. We assessed the genome-wide differential expression between the discordant twins by looking at the distribution of gene expression fold changes along the chromosomes. This comparison revealed well-defined chromosomal domains composed of neighbouring genes sharing differential expression profiles. In most of the chromosomes, large regions of upregulated genes alternate with large downregulated domains (Figs 1a, b). This observation suggests that the differential expression between T1DS and T2N is not randomly organized but follows a specific pattern along the chromosomes. We used a smoothing function to define the domain borders and identified a total of 337 GEDDs in the trisomy 21 discordant twins (Supplementary Table 2). Those domains vary in size (from 9 kilobases (kb) to 114 megabases (Mb), median size of 3.2 Mb) and contain up to 507 genes with a median number of 20.

Three independent replicate experiments confirmed these GEDDs in the discordant twins (overall correlation (Spearman correlation) ρ(Rep0, Rep1) = 0.88; ρ(Rep0, Rep2) = 0.76; ρ(Rep0, Rep3) = 0.99, Supplementary Fig. 2). We also compared the gene expression profiles of fibroblasts derived from a healthy pair of monozygotic twins (MZ1 and MZ2) as a control (overall ρ(Rep0, N) = −0.01, Supplementary Fig. 2). The absence of domains in this latter comparison suggests that the domain organization observed in the discordant twins can be mainly attributed to the supernumerary HSA21.

Gene expression levels and GEDDs

Next, we investigated the expression level of the genes within the domains. We classified genes according to low ($0.1 < \text{RPKM} < 10$), medium ($10 < \text{RPKM} < 100$) or high ($\text{RPKM} > 100$) expression level and their position along the chromosomes (Fig. 2a and Supplementary Fig. 3). For each category, we compared the expression fold change distribution with the expected distribution given by the healthy monozygotic twins (Fig. 2b). We found that the highly expressed genes contribute substantially to the domains downregulated in T1DS ($P < 2.2 \times 10^{-16}$ and $P = 1.4 \times 10^{-15}$ for medium and high expression, respectively). In contrast, the low expressed genes are mainly associated with the upregulated regions ($P = 4.4 \times 10^{-4}$). These data show that the expression fold change between trisomic and normal cells is organized in chromosomal domains and that those domains can be partially defined by the expression level of their gene content.

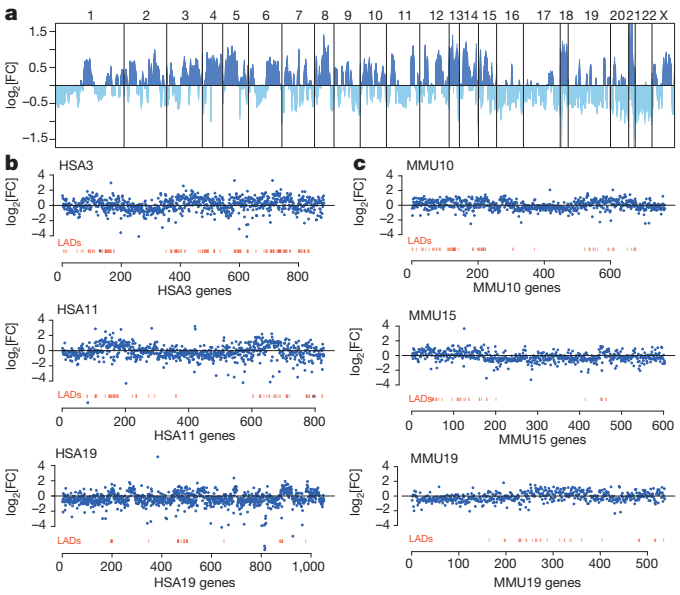

Figure 1 | Gene expression fold change is organized in chromosomal domains. **a**, Lowess smoothed $\log_2$ fold change ($\log_2$[FC] T1DS/T2N Rep0) of gene expression between T1DS and T2N fibroblasts along the human genome. Numbers indicate the chromosomes delimited with vertical lines. Upregulated domains are shown in dark blue and downregulated domains in light blue. **b**, $\log_2$ fold change ($\log_2$[FC] T1DS/T2N Rep0) of gene expression between T1DS and T2N fibroblasts along human chromosomes 3, 11 and 19. **c**, $\log_2$ fold change of gene expression between Ts65Dn and wild-type littermates along mouse chromosomes 10, 15 and 19. Genes (in blue) are sorted according to their position on the chromosome, at equidistance. The human and mouse LADs[27,28] are shown in red.

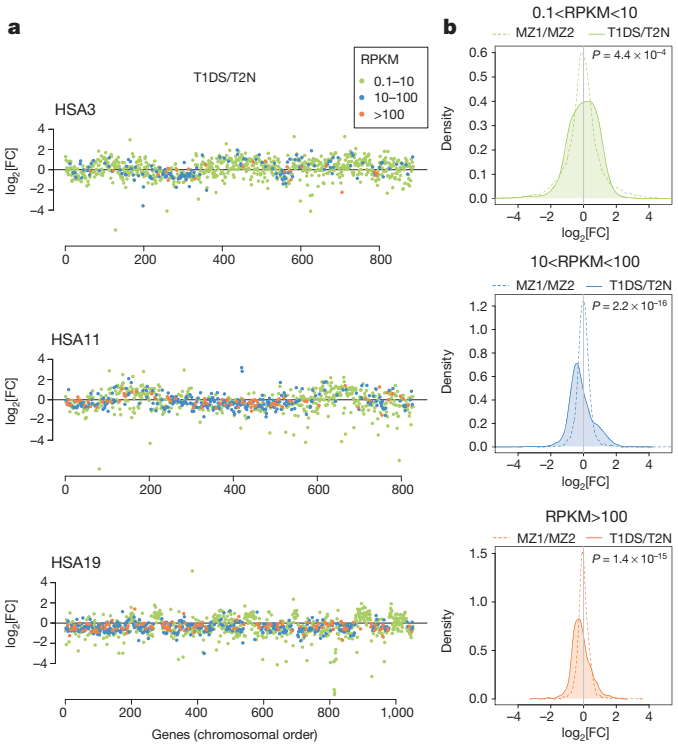

Figure 2 | Weakly and highly expressed genes contribute differently to the domains. **a**, $\log_2$ fold change ($\log_2$[FC] T1DS/T2N Rep0) of gene expression between T1DS and T2N fibroblasts along human chromosomes 3, 11 and 19. Genes are sorted according to their position on the chromosome (at equidistance) and coloured according to their expression level. **b**, Distribution of expression fold changes ($\log_2$) for low ($0.1 < \text{RPKM} < 10$, upper panel), medium ($10 < \text{RPKM} < 100$, middle panel) and high ($\text{RPKM} > 100$, lower panel) expressed genes. Solid lines represent the $\log_2$[FC] between the discordant twins (T1DS/T2N Rep0) and dashed lines represent the $\log_2$[FC] between the healthy twins (MZ1/MZ2). P = Wilcoxon test P value.

Previous studies have reported the clustering of highly expressed genes in domains[22,23]. We examined the gene expression along the chromosomes for each sample to understand the respective contribution of each twin to the GEDD pattern. We confirmed the clustering of genes according to their expression levels in both trisomic and normal fibroblasts. However, we found that the trisomic cells show a decreased dynamics of variation between highly and weakly expressed genes in all chromosomes but HSA13 and HSA18 (Supplementary Fig. 4). This difference of gene expression amplitude between trisomic and normal cells contributes to the observed GEDDs and indicates that gene expression might be less fine-tuned and less dynamic in a trisomic context.

GEDDs are conserved in induced pluripotent stem cells

To verify whether the GEDDs described above could be observed in other cell types, we derived induced pluripotent stem (iPS) cells from each of the fibroblast lines from the discordant twins[24]. We performed mRNA sequencing on both iPS lines and compared the normalized gene expression values along the chromosomes. Notably, for most of the chromosomes, we found that the differential expression patterns in iPS cells are highly similar (overall $\rho = 0.85$) to those observed in the fibroblasts (Fig. 3 and Supplementary Fig. 5). These results indicate that the GEDDs are conserved after dedifferentiation and that the supernumerary HSA21 has similar effects in the genome-wide dysregulation of gene expression in fibroblasts and iPS cells.

GEDDs are conserved in mouse Ts65Dn fibroblasts

To verify whether the GEDDs described in the twins' fibroblasts are conserved in other organisms, we performed a similar analysis in the partial trisomy 16 Ts65Dn mouse model of Down's syndrome. We analysed the differential expression pattern between primary cultures of Ts65Dn and normal littermate mouse fibroblasts. We first confirmed the expected increased expression of the MMU16 telomeric region that is triplicated in the Ts65Dn mice (Supplementary Fig. 6a). Moreover, those mice present an additional partial trisomy of the centromeric region of MMU17 that is not syntenic to HSA21 (ref. 25); genes in that region are also overexpressed (Supplementary Fig. 6b). Then we investigated the expression fold change between trisomic and wild-type mice and found that it is organized in GEDDs along the Ts65Dn mouse chromosomes (Fig. 1c and Supplementary Fig. 7). The comparison of the largest syntenic blocks (>1 Mb) between mouse and human revealed that the GEDDs are well conserved between the discordant twins' fibroblasts and the mouse fibroblasts and that the direction of dysregulation is maintained between the two species (Fig. 4a). This conservation is independent of the karyotypic context, as the syntenic regions are located in different chromosomes in the mouse. Human and mouse orthologous genes show similar expression fold changes (overall $\rho = 0.44$) (Fig. 4b, top panel). The same comparison using the healthy monozygotic twins did not show a correlation between human and mouse fold changes (overall $\rho = -0.13$) (Fig. 4b, bottom panel). Those results demonstrate that the GEDDs observed between T1DS and T2N are remarkably conserved in the Ts65Dn mouse model for Down's syndrome and indicate that the domain organization is independent of the chromosomal context.

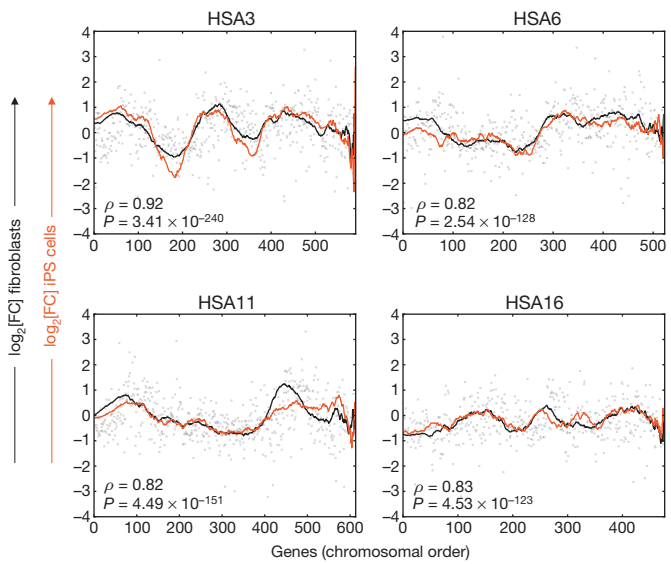

Figure 3 | GEDDs are conserved in iPS cells. Comparison of the $\log_2$ expression fold change profiles between T1DS and T2N in fibroblasts (in black, Rep0) and iPS cells (in red) along human chromosomes 3, 6, 11 and 16. ρ = Spearman correlation; P = associated P value.

Comparison of unrelated individuals

To verify whether GEDDs could be identified when comparing trisomy 21 to normal samples from unrelated individuals, we sequenced the mRNA from fibroblasts of 8 trisomic and 8 sex-matched controls. We reasoned that the natural genetic variability would have masked the domains[26]. Therefore, we averaged the expression values within each group to reduce the influence of individual genetic backgrounds and calculated the expression fold change between the two groups. This comparison did not reveal chromosomal domains (Supplementary Fig. 8), supporting the hypothesis that the natural gene expression variation masks the effects observed with the discordant monozygotic twins. To validate this hypothesis further, we tested whether the gene expression fold change between trisomic and normal unrelated individuals was significantly higher in the upregulated GEDDs than the downregulated GEDDs when selecting the genes with the lowest gene expression variation. We observed a plateau of significant P values (centred around 1×10^{-30} and coefficient of variation <0.45) with approximately 2,500 genes in the upregulated and 3,500 genes in the downregulated GEDDs (Supplementary Fig. 6c, d). These results support the hypothesis that the natural gene expression variation occurring between unrelated individuals is extensive and that only stably expressed genes can unmask the effects observed with the discordant monozygotic twins.

GEDDs and previously described chromosomal domains

The domain organization of the mammalian chromosomes has been previously reported with the identification of the LADs[27]. Therefore we evaluated the correlation between LADs and GEDDs comparing the distribution of gene expression fold changes within and outside the LADs. We observed that the genes located within LADs were, on average, overexpressed in T1DS as opposed to the genes located outside of LADs (in inter-LADs (iLADs), $P = 8.9 \times 10^{-16}$) (Fig. 5a). A similar significant shift was observed in the Ts65Dn fibroblasts when we compared the gene expression fold changes inside and outside the mouse LADs[28] ($P = 4.6 \times 10^{-7}$) (Fig. 5b). The healthy monozygotic twins did not show any expression differences between LADs and iLADs ($P = 0.54$) (Fig. 5c). The LADs are mostly associated with an overall inhibition of gene expression[27,29,30]. In our experiment, the increased expression fold change inside LADs suggests a de-repression of the genes in the LADs of trisomic cells. Therefore, we reasoned that the interaction of the genome with the nuclear lamina might be disturbed in trisomic nuclei. We used the DamID (DNA adenine methyltransferase identification) method[31] to define the position of the LADs in the discordant twins' fibroblasts. The lamin B1 interaction scores along the chromosomes revealed highly correlated profiles of DNA-lamina interactions between T1DS and T2N (Spearman $\rho = 0.94$) (Fig. 5d, e and Supplementary Fig. 6d). We concluded that the overall topology of LADs is not detectably perturbed by the presence of an extra HSA21.

Furthermore, the LADs correlate with replication domains[32]. Early replicating domains contain mostly active genes and tend to localize centrally in the nucleus as opposed to late replicating domains that mainly contain less active genes at the nuclear periphery[33,34]. We assessed

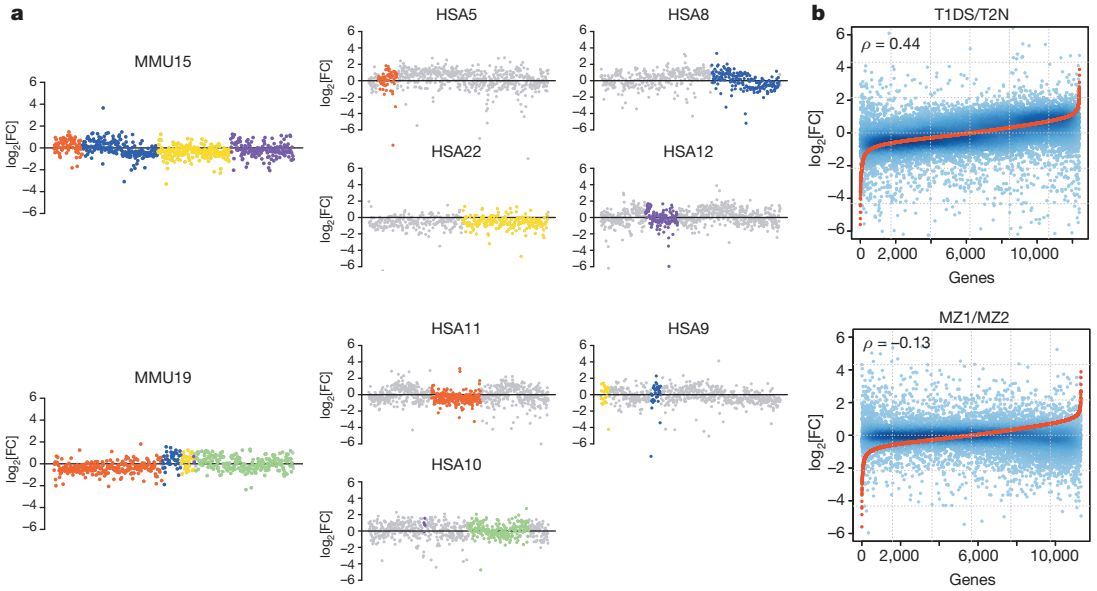

Figure 4 | GEDDs are conserved in Ts65Dn mouse model for Down's syndrome. **a**, Log_2 fold change of gene expression in the Ts65Dn fibroblasts (Ts65Dn/wild type) along mouse chromosomes 15 (MMU15, top panel) and 19 (MMU19, bottom panel). Genes are sorted according to their position on the chromosome (at equidistance). Colours represent the mouse/human syntenic blocks. The corresponding blocks on human chromosomes (T1DS/T2N Rep0) are shown with the same colours. **b**, Spearman correlation plot of log_2 expression fold changes between human (T1DS/T2N Rep0 in top panel and MZ1/MZ2 in bottom panel) and mouse (Ts65Dn/wild type) orthologous genes. Mouse genes (in red) are ranked in the ascending order of log_2[FC]. Human genes (in blue) are plotted in the same order. ρ = Spearman correlation.

the correlation between the replication time domains of human fetal fibroblasts[35] and the GEDDs described here. Most chromosomes show a strong correlation between both profiles, indicating that gene expression is increased in late replication domains and decreased in early replication domains of the T1DS cells (overall correlation $\rho = 0.6$, Fig. 6a and Supplementary Fig. 9).

Modification of the chromatin environment in GEDDs

Despite the fact that GEDDs are correlated to structural chromosomal domains, DamID experiments showed that the genome topology is not changed in T1DS fibroblasts. Thus, we reasoned that the alterations of gene expression might be related to epigenetic chromatin modifications within the chromosomal domains of trisomic cells.

We explored the DNA methylation changes in T1DS compared to T2N. We performed reduced representation bisulphite sequencing (RRBS)[36] on DNA from the twins' fibroblasts and evaluated the methylation percentage for each CpG dinucleotide in the genome. We then compared the methylation level of each gene between T1DS and T2N by considering the CpGs either in the gene body or in the promoter region. The profiles of gene methylation differences along the chromosomes were compared to the GEDDs. For some chromosomes, we observed a weak correlation between both patterns (Supplementary Fig. 10a, b). However, the overall degree of concordance between methylation and gene expression (overall $\rho = 0.29$) is not sufficient to correlate methylation status to GEDDs. We concluded that the alteration of gene expression cannot be fully explained by domain-wide changes of cytosine methylation in the trisomic fibroblasts.

The cellular transcription activity can also be influenced by changes of chromatin marks such as H3K4me3 that is known to correlate positively with gene expression[37]. Thus, we used chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) to compare the level of H3K4me3 in the discordant twins. We identified the genomic regions enriched for H3K4me3 in T1DS and T2N, and investigated the differences of signal intensity between all peaks common to both twins. The median of the resulting H3K4me3 log_2 fold change (log_2[FC]) for each gene was estimated to establish chromosomal profiles comparable to GEDDs. The comparison revealed a marked correlation between both patterns (overall $\rho = 0.93$) for all chromosomes except HSA13 and HSA18 (Fig. 6b and Supplementary Fig. 11). Those chromosomes do not show a domain pattern due to the overexpression of almost all of their genes. We conclude that the GEDDs observed in the twins correlate with modification of H3K4me3 in T1DS.

This observation raised the question of the general chromatin accessibility in the presence of an extra copy of HSA21. The DNase I hypersensitivity signal has been widely used to identify the accessible sites in the genome and to recognize the regulatory sequences[38]. We explored the chromatin accessibility in the twins' fibroblasts by performing DNase I hypersensitivity mapping. Interestingly, we detected an overall increased

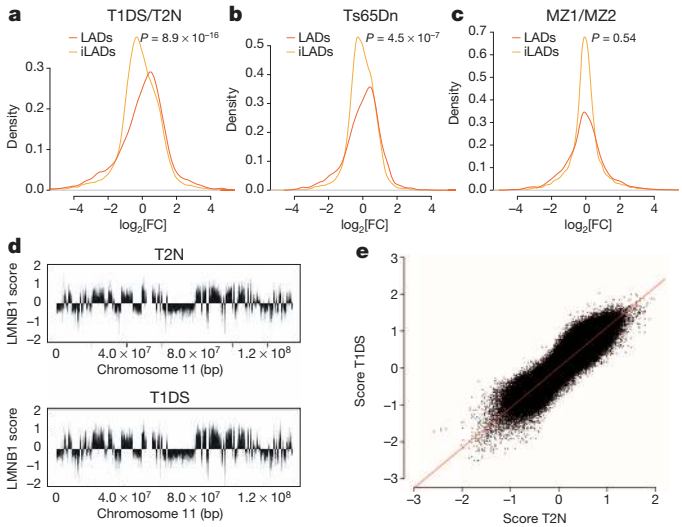

Figure 5 | Correlation with LADs. Distribution of log_2 expression fold changes for genes overlapping a LAD (in red) and genes located outside of a LAD (in yellow, inter-LADs (iLADs)). **a–c**, Profiles are shown for the discordant twins (Rep0) (**a**), the Ts65Dn mice (**b**) and the healthy monozygotic twins (**c**). P = Wilcoxon test P value. **d**, DamID map of lamin B1 (LMNB1) interaction scores for human chromosome 11 in T1DS (top panel) and T2N (bottom panel). **e**, Correlation plot of lamin B1 interaction scores for the entire genome between T1DS and T2N ($\rho = 0.94$). (Samples smoothed with a running median window of 11 probes.)

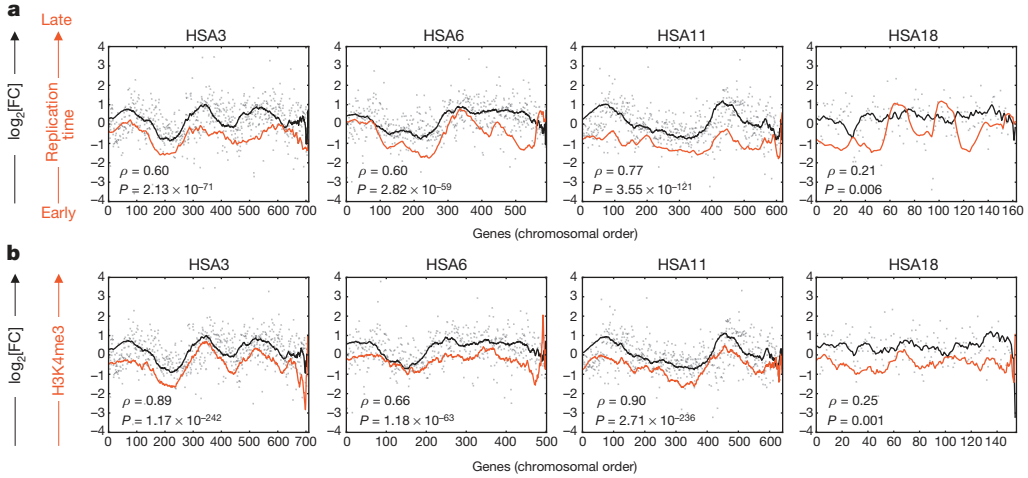

Figure 6 | Correlation with replication time domains and H3K4me3 mark. **a**, **b**, Comparison of the $\log_2$ expression fold change between T1DS and T2N (Rep0, in black) and the replication time (**a**) or the H3K4me3 difference (**b**) profiles (in red) along human chromosomes 3, 6, 11 and 18. ρ = Spearman correlation; P = associated P value.

genomic accessibility in the trisomic twin, particularly marked in the gene bodies, transcription start sites and promoter regions (Supplementary Fig. 6e). Because this modification may influence overall gene expression, we investigated the link between the changes of DNase I hypersensitivity profiles in the twins' fibroblasts and the GEDD pattern. Surprisingly, this comparison revealed that the two patterns were weakly linked and even tend to be inversely correlated (overall $\rho = -0.56$, Supplementary Fig. 12). Given the known correlation between DNase I hypersensitivity signal and marks for open chromatin, especially H3K4me3 (ref. 39), this observation is unexpected and may indicate some compensatory mechanisms for the regulation of gene expression in an open chromatin context.

Discussion

The transcriptome analysis of monozygotic twins discordant for trisomy 21 highlights the existence of chromosomal domains of gene expression dysregulation between trisomic and normal fibroblasts. This observation includes not only the expression of protein-coding genes but also lncRNAs (Supplementary Fig. 6f). We attribute this discovery to the use of samples from monozygotic twins, which eliminates the influence of individual genetic backgrounds. The investigation of gene expression level in each discordant twin independently revealed that the GEDD pattern could be attributed to a reduced dynamics of gene expression in the trisomic fibroblasts. The conservation of this transcriptome pattern in iPS cells suggests that the mechanism responsible for the dysregulation domains is not specific to fibroblasts but is maintained after dedifferentiation. The observed conservation also indicates that trisomy 21 has a consistent influence on the transcriptome of iPS cells.

We have demonstrated that GEDDs are conserved in the Ts65Dn mouse model for Down's syndrome. Remarkably, these domains are present in a genomic chromosomal context that is different in human and mouse. The conservation in mouse also suggests that the domain organization may be the consequence of the overexpression of one or more HSA21 gene(s), the orthologues of which are found on the MMU16 syntenic region. Alternatively, the domains could be related to the physical presence of the extra DNA material in the nucleus.

We have also shown that GEDDs significantly correlate with the lamina-associated and replication domains. Our results indicate that the trisomic cells undergo de-repression of the genes located at the nuclear periphery and repression of the early replicating active genes. The topology of LADs is not disturbed in trisomic cells; however, GEDDs are related to epigenetic modifications of the chromatin environment as shown by the almost perfect correlation with differential enrichment of H3K4me3 observed between T1DS and T2N. These modifications may occur to compensate general changes of chromatin accessibility as suggested by the DNase I hypersensitivity experiment.

Several HSA21 genes emerge as potential candidates for the epigenetic modification of chromosomal domains. *HLCS*, which is triplicated in the Ts65Dn mouse model, catalyses the binding of biotin to histones, and participates in chromatin condensation and gene repression[40,41]. Interestingly, HLCS is associated to the nuclear lamina[42] and can physically interact with chromatin-modifying proteins such as DNMT1 or MeCP2 (ref. 41).

The HSA21 protein HMGN1 also influences the structure and function of the chromatin through histone modifications[43] and is considered to be an important modulator of gene expression[44].

Other HSA21 proteins that are directly or indirectly involved in epigenetic mechanisms such as DYRK1A, BRWD1 and RUNX1 may influence the epigenetic architecture of the nucleus[45–47]. Additional experiments may help to explore their contribution to the GEDD pattern.

Furthermore, the correlation of GEDDs with the replication time domains and a prolonged permissive chromatin state in trisomic cells (as shown by our DHS experiments) raises the hypothesis of an excess of background transcription in trisomic cells. Several studies have shown that the cell cycle is extended in trisomic nuclei[48,49] without affecting the replication time[35]. This perturbation of the cell cycle may therefore result in a prolongation of chromatin access time and in a general increase of stochastic transcription[50].

We propose that the overexpression of one or more HSA21 candidate gene(s) modifies the chromatin environment of the nuclear compartments in trisomic cells. These modifications would lead to a general perturbation of the transcriptome that could explain some of the Down's syndrome phenotypes. Alternatively, the GEDDs could be the result of the additional chromosomal material of trisomy 21. The prediction of this latter hypothesis is that other trisomies may have a similar effect on the dysregulation of gene expression.

METHODS SUMMARY

Details on the following methods (and references within) can be found in the full Methods: cell culture and RNA preparation, mRNA sequencing, mapping and normalization, definition of chromosomal domains, expression level analysis, human/mouse comparison, lamina-associated domain analysis and DamID experiments, reduced representation bisulphite sequence (RRBS) library preparation, methylation computational analysis, H3K4me3 ChIP-sequencing, DNase I hypersensitivity mapping, and correlation analysis between $\log_2$ fold change in fibroblasts versus iPS $\log_2$[FC], IMR90 replication time, DNA methylation, H3K4me3 and DNase I hypersensitivity profiles.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 September 2013; accepted 4 March 2014.

1. Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Reymond, A. & Deutsch, S. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nature Rev. Genet.* **5,** 725–738 (2004).
2. Korenberg, J. R. *et al.* Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl Acad. Sci. USA* **91,** 4997–5001 (1994).
3. Pritchard, M. A. & Kola, I. The "gene dosage effect" hypothesis versus the "amplified developmental instability" hypothesis in Down syndrome. *J. Neural Transm. Suppl.* **57,** 293–303 (1999).
4. Lyle, R. *et al.* Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21. *Eur. J. Hum. Genet.* **17,** 454–466 (2009).
5. Shapiro, B. L. Down syndrome–a disruption of homeostasis. *Am. J. Med. Genet.* **14,** 241–269 (1983).
6. Letourneau, A. & Antonarakis, S. E. Genomic determinants in the phenotypic variability of Down syndrome. *Prog. Brain Res.* **197,** 15–28 (2012).
7. Davisson, M. T. *et al.* Segmental trisomy as a mouse model for Down syndrome. *Prog. Clin. Biol. Res.* **384,** 117–133 (1993).
8. Reeves, R. H. *et al.* A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genet.* **11,** 177–184 (1995).
9. Chrast, R. *et al.* Mice trisomic for a bacterial artificial chromosome with the single-minded 2 gene (Sim2) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of Down syndrome. *Hum. Mol. Genet.* **9,** 1853–1864 (2000).
10. Ahn, K. J. *et al.* DYRK1A BAC transgenic mice show altered synaptic plasticity with learning and memory defects. *Neurobiol. Dis.* **22,** 463–472 (2006).
11. Costa, V. *et al.* Massive-scale RNA-Seq analysis of non ribosomal transcriptome in human trisomy 21. *PLoS ONE* **6,** e18493 (2011).
12. Esposito, G. *et al.* Genomic and functional profiling of human Down syndrome neural progenitors implicates S100B and aquaporin 4 in cell injury. *Hum. Mol. Genet.* **17,** 440–457 (2008).
13. Lockstone, H. E. *et al.* Gene expression profiling in the adult Down syndrome brain. *Genomics* **90,** 647–660 (2007).
14. Sommer, C. A., Pavarino-Bertelli, E. C., Goloni-Bertollo, E. M. & Henrique-Silva, F. Identification of dysregulated genes in lymphocytes from children with Down syndrome. *Genome* **51,** 19–29 (2008).
15. Prandini, P. *et al.* Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. *Am. J. Hum. Genet.* **81,** 252–263 (2007).
16. Deutsch, S. *et al.* Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum. Mol. Genet.* **14,** 3741–3749 (2005).
17. Dahoun, S. *et al.* Monozygotic twins discordant for trisomy 21 and maternal 21q inheritance: a complex series of events. *Am. J. Med. Genet. A.* **146A,** 2086–2093 (2008).
18. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nature Methods* **9,** 1185–1188 (2012).
19. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
20. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
21. Ram, G. & Chinen, J. Infections and immunodeficiency in Down syndrome. *Clin. Exp. Immunol.* **164,** 9–16 (2011).
22. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291,** 1289–1292 (2001).
23. Gierman, H. J. *et al.* Domain-wide regulation of gene expression in the human genome. *Genome Res.* **17,** 1286–1295 (2007).
24. Hibaoui, Y. *et al.* Modelling and rescuing neurodevelopmental defect of Down syndrome using induced pluripotent stem cells from monozygotic twins discordant for trisomy 21. *EMBO Mol. Med.* **6,** 259–277 (2014).
25. Duchon, A. *et al.* Identification of the translocation breakpoints in the Ts65Dn and Ts1Cje mouse lines: relevance for modeling Down syndrome. *Mamm. Genome* **22,** 674–684 (2011).
26. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–511 (2013).
27. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–951 (2008).
28. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).
29. Zullo, J. M. *et al.* DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149,** 1474–1487 (2012).
30. Akhtar, A. & Gasser, S. M. The nuclear envelope and transcriptional control. *Nature Rev. Genet.* **8,** 507–517 (2007).
31. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nature Protocols* **2,** 1467–1478 (2007).
32. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107,** 139–144 (2010).
33. Gilbert, D. M. Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.* **14,** 377–383 (2002).
34. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6,** e245 (2008).
35. Pope, B. D. *et al.* Replication-timing boundaries facilitate cell-type and species-specific regulation of a rearranged human chromosome in mouse. *Hum. Mol. Genet.* **21,** 4162–4170 (2012).
36. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols* **6,** 468–481 (2011).
37. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
38. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
39. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
40. Singh, M. P., Wijeratne, S. S. & Zempleni, J. Biotinylation of lysine 16 in histone H4 contributes toward nucleosome condensation. *Arch. Biochem. Biophys.* **529,** 105–111 (2013).
41. Pestinger, V., Wijeratne, S. S., Rodriguez-Melendez, R. & Zempleni, J. Novel histone biotinylation marks are enriched in repeat regions and participate in repression of transcriptionally competent genes. *J. Nutr. Biochem.* **22,** 328–333 (2011).
42. Narang, M. A., Dumas, R., Ayer, L. M. & Gravel, R. A. Reduced histone biotinylation in multiple carboxylase deficiency patients: a nuclear role for holocarboxylase synthetase. *Hum. Mol. Genet.* **13,** 15–23 (2004).
43. Postnikov, Y. & Bustin, M. Regulation of chromatin structure and function by HMGN proteins. *Biochim. Biophys. Acta* **1799,** 62–68 (2010).
44. Zhu, N. & Hansen, U. Transcriptional regulation by HMGN proteins. *Biochim. Biophys. Acta* **1799,** 74–79 (2010).
45. Canzonetta, C. *et al.* DYRK1A-dosage imbalance perturbs NRSF/REST levels, deregulating pluripotency and embryonic stem cell fate in Down syndrome. *Am. J. Hum. Genet.* **83,** 388–400 (2008).
46. Huang, H., Rambaldi, I., Daniels, E. & Featherstone, M. Expression of the Wdr9 gene and protein products during mouse development. *Dev. Dyn.* **227,** 608–614 (2003).
47. Bakshi, R. *et al.* The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. *J. Cell. Physiol.* **225,** 569–576 (2010).
48. Contestabile, A. *et al.* Cell cycle alteration and decreased cell proliferation in the hippocampal dentate gyrus and in the neocortical germinal matrix of fetuses with Down syndrome and in Ts65Dn mice. *Hippocampus* **17,** 665–678 (2007).
49. Williams, B. R. *et al.* Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322,** 703–709 (2008).
50. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Rev. Genet.* **15,** 69–81 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Swiss National Science Foundation (SNF-144082), the European Research Council (ERC-249968), AnEUploidy and BluePrint EU grants, the Lejeune, and ChildCare foundations for supporting the S.E.A. laboratory. The laboratories of R.G. were supported by Spanish MICINN (BIO2011-26205) and ERC-294653, B.v.S. by NWO-ALW-VICI, Y.He. by CNRS, INSERM, University of Strasbourg and ANR-10-INBS-07, J.A.S. by NIH U54HG007010, and A.F. by Genico and Ernest Boninchi foundation. We thank S. Dahoun and J. L. Blouin for the discordant twins sample collection.

Author Contributions The project was coordinated by S.E.A. A.L. coordinated/undertook the main laboratory work. F.A.S. coordinated/undertook the main bioinformatics/statistical analyses. X.B. performed ChIP-seq experiments. M.R.S. performed DNA methylation, A.L., F.A.S., M.G., R.G. and D.G. processed NGS data. J.K. and B.v.S. performed DamID experiments. C.C. and Y.He. maintained the mouse colony and contributed mouse samples. R.T., R.S.S. and J.A.S. performed DNase experiments; Y.Hi. and A.F. derived the iPS cells; and K.P., D.R. R.G. and E.M. performed additional statistical analyses. E.F., M.G., C.G., A.V., M.G., L.F., C.B. and S.D. assisted with wet lab experiments and contributed to performing NGS experiments. The main findings were interpreted by S.E.A., A.L. and F.A.S., who also wrote the manuscript. All authors made comments on the manuscript.

Author Information All sequencing data have been deposited in the Gene Expression Omnibus (GEO) data repository under accession number GSE55426. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.A. (Stylianos.Antonarakis@unige.ch).

METHODS

Cell culture and RNA preparation. Forearm primary fetal skin fibroblasts were collected post mortem from the T1DS and T2N discordant twins at 16 fetal weeks, after IRB approval from the Ethical Committee of University Hospitals of Geneva, and written informed consent by both parents. Three replicate experiments (Rep0, Rep1 and Rep2) were performed from independent cultures of the same fibroblasts. The fourth replicate (Rep3) was performed using the RNA material of Rep0. Primary fibroblasts from MZ1 and MZ2 healthy monozygotic twins were derived from umbilical cord tissue collected as part of the GenCord project in the University Hospitals of Geneva[51]. Primary skin fibroblasts from trisomy 21 and normal unrelated individuals were taken from ref. 15 (GM08447, GM05756, GM00969, GM00408, GM02036, GM03377, GM03440, PM9726F, GM04616, AG07409, AG06872, AG06922, GM02767, AG08941, AG08942, D-S99124M). All primary fibroblasts were grown in DMEM GlutaMAX (Life Technologies 31966) supplemented with 10% fetal bovine serum (Life Technologies 10270) and 1% penicillin/streptomycin/fungizone mix (Amimed, BioConcept 4-02F00-H) at 37 °C in a 5% CO_2 atmosphere.

Mouse embryonic fibroblasts (MEFs) were obtained from 14.5-day-old mouse embryos. Embryos were collected, eviscerated, decapitated and then further treated with trypsin to obtain the MEFs. MEFs were cultured at least two passages in classical DMEM (Gibco) medium supplemented with antibiotics, glutamine and 10% fetal calf serum. Subconfluent MEFs were harvested in TRIzol lysis buffer (Life Technologies 15596).

Primary fetal skin fibroblasts isolated from T1DS and T2N (10 passages) were used to establish normal (Twin-N-iPSCs) and Down's syndrome (Twin-DS-iPSCs) induced pluripotent stem cells (iPS cells). These iPS cells were generated by transducing the parental fibroblasts with polycistronic lentiviral vectors expressing *OCT4*, *SOX2, KLF4* and *MYC* reprogramming factors[52,53]. The generated iPS cells were cultured on irradiated human foreskin fibroblasts (ATCC, CCD 1112Sk) that were mitotically inactivated by irradiation at 35 Gy. iPS cell colonies were maintained with daily changes in Knockout Dulbecco's Minimal Essential Medium (KO DMEM) supplemented with 20% KO serum replacement, 1 mM L-glutamine, 100 μM non-essential amino acids, 100 μM 2-mercaptoethanol, 50 U ml^{-1} penicillin and 50 mg ml^{-1} streptomycin (all from Gibco, Invitrogen) and 100 ng ml^{-1} human basic fibroblast growth factor (bFGF, Peprotech). These iPS cells were passaged by manual dissection of cell clusters in the presence of 10 μM ROCK-inhibitor Y-27632 (Sigma-Aldrich).

Total RNA from all cell types was collected using the TRIzol reagent (Life Technologies 15596) following the manufacturer's instructions. RNA isolation was performed after 10 or 13 passages of the twins' primary fibroblasts. iPS cells were grown for 17 passages before the RNA isolation step. RNA quality was verified on the Agilent 2100 Bioanalyzer (RNA 6000 Nano kit, 5067) and quantity was measured on a Qubit instrument (Life Technologies).

mRNA sequencing, mapping and normalization. mRNA-seq libraries were prepared from 4 to 10 μg of total RNA using the Illumina mRNA-seq Sample Preparation kit (RS-100-0801), following the manufacturer's instructions. Libraries were sequenced on the Illumina HiSeq 2000 (Rep0 and Rep1) and on the Illumina Hiseq 2500 (Rep2 and Rep3) using paired-end sequencing 2 × 100 bp. Reads were mapped against the human (hg19) or mouse (mm9) genomes using the default parameters of the GEM mapper through the GRAPE pipeline[18,54]. An independent mapping was performed with TopHat using default parameters. Quantile normalization was performed on the RPKM data (reads per kilobase per million) for all the human samples on one side and for the mouse samples on the other side. Differential expression analysis was performed on the raw read counts using the default parameters of EdgeR on the genes expressed in at least two samples[19].

Definition of chromosomal domains. For each gene, we determined the $\log_2$ expression fold change ($\log_2$[FC]) by calculating the ratio of normalized expression values (after quantile normalization) between the trisomic and the euploid samples (that is, a positive $\log_2$[FC] reflects the overexpression of the gene in the trisomic sample). For the healthy monozygous twins the $\log_2$[FC] was based on the ratio between MZ1 and MZ2. For the unrelated individuals, the $\log_2$[FC] was based on the ratio between the mean value of all trisomy 21 individuals and the mean value of all normal individuals. For all the comparisons, only the genes with at least 0.1 RPKM in the normal individual(s) (T2N, average value of MZ1 and MZ2 or average value of the normal unrelated individuals) were considered. The $\log_2$[FC] values were plotted equidistantly based on the gene positions along the chromosomes. We used the lowess function in R (http://www.r-project.org) to smooth the $\log_2$[FC] data (smoother span 3 to 30%) and identify upregulated and downregulated domains. An upregulated domain was defined as a set of at least two consecutive genes with positive smoothed $\log_2$[FC] values. On the contrary, a downregulated domain was defined as a set of at least two consecutive genes with negative smoothed $\log_2$[FC] values.

Expression level analysis. The genes were classified according to their expression level (RPKM value after quantile normalization) in three categories: low ($0.1 < RPKM < 10$), medium ($10 < RPKM < 100$) or high ($RPKM > 100$) expression. In the discordant twins comparison, we used the expression level of the normal twin (T2N) as a reference. In the healthy monozygous twins comparison, we used the average value between MZ1 and MZ2 to assess the gene expression level. For each category, the distribution of $\log_2$[FC] was plotted (density function in R). The difference of distribution between the discordant twins and the healthy monozygous twins was assessed by the Wilcoxon test.

Gene expression dynamics for each sample has been evaluated by calculating the standard deviation of the lowess smoothed $\log_2$ expression of each gene sorted according to chromosomal coordinates.

Unrelated individuals analysis. We calculated the coefficient of variation (CV) of each gene over all 8 + 8 samples as the standard deviation of its relative expression divided by the mean. The genes with CV above a decreasing threshold were iteratively discarded. At each step, we evaluated the expression fold change of all the remaining genes by considering all possible pairwise comparisons of the 8 trisomic and 8 control samples. We applied a non-parametric Mann–Whitney *U*-test between the fold change values obtained for the genes located in the upregulated GEDDs and the genes located in the downregulated GEDDs.

Human/mouse comparison. The mouse/human syntenic block coordinates were downloaded from the "Comparative Genomics" track of the UCSC genome browser (http://genome.ucsc.edu). Only the largest mouse syntenic blocks (>1 Mb) were conserved for the comparative analysis. $\log_2$[FC] of gene expression between Ts65Dn and wild-type mice were plotted along the mouse chromosomes. Each gene was assigned to a syntenic block (total overlap between the gene coordinates and the block coordinates) and coloured accordingly. The corresponding blocks in human were shown with the same colours in the GEDD plots.

The list of mouse and human orthologous genes was obtained from the Mouse Genome Informatics Database project, The Jackson Laboratory, Bar Harbour, Maine (ftp://ftp.informatics.jax.org/pub/reports/HMD_HumanPhenotype.rpt). We compared the $\log_2$[FC] of expression obtained in the mouse with the values obtained for their human orthologues. The comparison was done first with the discordant twins and then with the healthy monozygous twins. The degree of similitude between the 2 sets was given by the Spearman correlation.

LADs analysis and DamID. The coordinates of human LADs were taken from ref. 27, and the mouse LAD coordinates were taken from ref. 28 (mouse embryonic fibroblasts). A gene was assigned to a LAD if its coordinates overlap from at least 1 bp with the LADs coordinates. The distribution of $\log_2$[FC] for the overlapping genes was compared to the distribution of $\log_2$[FC] for the non-overlapping genes using a Wilcoxon test. DamID was performed as previously described[31].

Reduced representation bisulphite sequencing (RRBS) library preparation. Reduced representation bisulphite sequencing (RRBS) libraries were made according to ref. 36 with some modifications. In short, the Qiagen DNeasy Blood and Tissue kit was used to extract DNA from the twin fibroblast cells. We then digested 2 μg genomic DNA with 2 μl 20 U μl^{-1} MspI restriction enzyme that cuts DNA regardless of cytosine methylation status at CCGG sequence in 5 μl NEB buffer 4 in a total reaction volume of 50 μl. This reaction was incubated for 18 h at 37 °C followed by heat inactivation at 80 °C for 20 min to generate DNA fragments containing CpG dinucleotides at the end. Phenol extraction and ethanol precipitation steps were used to clean up DNA from enzymatic reaction. DNA fragments were subsequently end-repaired and A-tailed by adding 2 μl dNTP mix (10 mM), 2 μl 5 U μl^{-1} Klenow fragment, and 4 μl of NEB2 in a total reaction volume of 40 μl. This reaction was incubated at 30 °C for 20 min, followed by 20 °C for 37 min. Heat inactivation was done at 75 °C for 20 min. The end-repaired and A-tailed DNA was purified by phenol extraction and ethanol precipitation steps and ligated to the methylated version of Illumina adapters: ilAdap Methyl PE1 (ACACTCTTTCCC TACACGACGCTCTTCCGATC*T) and ilAdap Methyl PE2 (GATCGGAAGAG CGGTTCAGCAGGAATGCCGA*G); all Cs are methylated, 5′ phosphate, asterisk indicates phosphorothioate bond. The ligation was mediated by adding 1 μl T4 DNA ligase (2,000 U μl^{-1}), 2 μl T4 ligase buffer (10×), and 1 μl methylated adapters (paired-end adapters) (15 μM) in a total reaction volume of 20 μl. The reaction was incubated at 16 °C overnight. The adaptor-ligated DNA was purified by phenol extraction and ethanol precipitation steps and dissolved in 15 μl EB buffer. Size selection of the adaptor-ligated DNA fragments (170 bp to 350 bp) was done by electrophoresing the 15 μl ligation reaction in a 2.5% NuSieve GTG agarose gel (Lonza). We subsequently purified the DNA using a Qiagen Qiaquick Gel Extraction kit as described in the manufacturer's instructions, except that we eluted the purified DNA from the column using 25 μl buffer EB. We then used 20 ml of this purified DNA in the sodium bisulphite conversion, which was performed using the QIAGEN Epitect Bisulphite kit (Qiagen) per manufacturer's recommended protocol with the following modifications: incubation after the addition of bisulphate conversion reagents was conducted in a thermocycler with the following conditions:

95 °C for 5 min, 60 °C for 25 min, 95 °C for 5 min, 60 °C for 85 min, 95 °C for 5 min, 60 °C for 175 min (95 °C for 5 min and 60 °C for 90 min) × 6. The bisulphate-converted DNA was eluted two times in 20 µl of EB buffer yielding 40 µl at the end. Illumina PCR primers PE1 and PE2 were used for the final library amplification. The sequences of PCR primers are as follow: ilPCR PE1 (AATGATACGGCGACCACC GAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T) and ilPCR PE2 (CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCC TGCTGAACC GCTCTTCCGATC*T), where an asterisk indicates a phosphorothioate bond. The purified bisulphate-treated DNA was then PCR-amplified in a reaction containing 100 µl KAPA2G Robust DNA Polymerase mixture (Kapabiosystems), 40 µl bisulphite-treated DNA, 0.5 mM ilPCR PE1, 0.5 mM ilPCR PE2 DNA oligonucleotide primers in a total reaction volume of 200 µl. The PCR mixtures were divided among eight PCR tubes and were incubated at 95 °C for 2 min, followed by 20 cycles of (95 °C for 20 s and 65 °C for 30 s and 72 °C for 30 s) and 72 °C for 7 min. The PCR products were pooled and purified by adding 360 µl AMPure magnetic beads (Agencourt Bioscience). We confirmed the amplification and correct product size range by running one-fifth of the reaction on a 2% agarose gel. We quantified the purified product using the Qubit fluorometer (Invitrogen). We also checked the template size distribution by running an aliquot of the library on an Agilent Bioanalyzer (Agilent 2100 Bioanalyzer). We then diluted each library to 10 nM and proceeded to sequence each library (pair-ended 100 bp) in a single lane on the Illumina HiSeq 2000 according the manufacturer's instructions. We also ran one lane of a standard (non-RRBS) library (Exome) in the same flow cell as the control lane and used this lane (rather than the RRBS lane) to calculate the dye matrix for base calling. Also, 5% phiX genomic DNA (control) was spiked into each lane. We also limited the cluster number for RRBS libraries to an 80% optimized cluster number per tile on Illumina HiSeq 2000. Image capture, analysis and base calling were performed using Illumina's CASAVA 1.7.

Methylation computational analyses. We obtained 121.7 million reads for T1DS and 104.3 million reads for T2N of which after trimming of reads for base quality and adaptor contamination (trim_galore wrapper http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) 59% and 59.6% of reads were uniquely mapped against human genome hg19 using Bismark aligner (-n 3 -l 20 -I 20)[55]. We used MethylKit R package[56] to calculate methylation percentages for each CpG. Per cent methylation values for CpG dinucleotides are calculated by dividing number of methylated Cs by total coverage on that base. CpGs with at least 20× read coverage and at least a Phred score threshold of 20 were retained for calling CpG methylation. 3,984,938 CpGs in T1DS and 3,690,212 CpGs in T2N met these criteria of which 2,598,760 CpGs are covered in both samples.

H3K4me3 ChIP sequencing. Primary human skin fibroblasts from both twins were grown in 10 cm dishes without reaching confluence. 21 million cells were crosslinked by adding enough formaldehyde for final concentration 0.5% to the growing media at room temperature. Dishes were incubated for 10 min and the reaction was then quenched by adding glycine to a final concentration of 125 mM. The crosslinked cells were collected and stored at −80 °C. For cell lysis and sonication, the pellets were re-suspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8.0) and sonicated in a Covaris S2 instrument (duty cycle 5%, intensity 5, 200 cycles per burst, 6 min).

Chromatin from 7.5 million cells was incubated with 10 µl of H3K4me3 antibody (Abcam ab8580 0.5 mg ml^{-1}) coupled to IgG magnetic beads (Cell Signaling) at room temperature for 1 h and subsequently at 4 °C for another hour. The magnetic beads were washed five times with LiCl wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and one time with TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA). Subsequently, the bound DNA was eluted at 65 °C in elution buffer (1% SDS, 0.1 M $NaHCO_3$) for 1 h and then separated from the beads with a magnetic stand. The immunoprecipitated DNA was incubated overnight at 65 °C with 2 µl Proteinase K to reverse crosslink. The samples were then purified with the QIAquick purification kit (Qiagen).

ChIP-seq libraries were prepared using a ChIP-seq DNA Sample Prep kit from Illumina with the following modifications to the protocol: mRNA adaptor indexes from the TruSeq RNA kit were used instead of the Adaptor oligo mix. The enrichment PCR was carried out with PCR reagents from the TruSeq mRNA kit from Illumina. The enriched library was purified with AMPure magnetic beads (Agencourt), its concentration verified with Qubit (Invitrogen) and the distribution and size of fragments with a Bioanalyzer (Agilent). Four samples were pooled in equimolar proportion and sequenced in a HiSeq2500 (single read, 36 bp). The obtained reads were demultiplexed with Illumina CASAVA 1.8 software and mapped to hg19 with BWA. Peaks were called using unique tags with HOMER (http://sdcsb.ucsd.edu/resources/homer/). Peak comparison between twins was performed with HOMER and in house scripts.

DNase I hypersensitivity mapping. DNase I hypersensitivity mapping and analysis were performed as previously described[38,57]. DNase I hypersensitivity density plots were generated by counting the reads covering each gene body ± 5 kb and normalizing by the respective gene length. Each gene body has been divided in 100 non-overlapping bins +50 bins for ± 5 kb extensions.

Correlation analysis between log_2[FC] fibroblast versus iPS log_2[FC], IMR90 replication time, DNA methylation, H3K4me3 and DNase I hypersensitivity profiles. iPS log_2[FC] chromosomal profiles have been obtained from RNA-seq data as previously described.

Replication time domains of IMR90 fetal fibroblasts were downloaded from the ReplicationDomain database (http://www.replicationdomain.org). Gene replication times have been estimated for each gene considering the median of each overlapping replication time domain and ordered according to chromosomal coordinates.

H3K4me3 profiles have been obtained as follows. For each gene body ±1,000 bp, the overlapping H3K4me3 ChIP-seq peaks were isolated and the log_2[FC] of the size of the matching peaks (as provided by HOMER) of the trisomic and the healthy monozygotic twin were calculated. The median of the resulting H3K4me3 log_2[FC] for each gene was eventually retained to compose the chromosomal profiles according to gene coordinates.

Similarly, DNase I hypersensitivity profiles have been obtained considering the log_2[FC] of the size of the overlapping DNase peaks for each gene body ±1,000 bp. The median of the resulting DNase log_2[FC] for each gene was retained to compose the chromosomal profiles according to gene coordinates.

DNA methylation profiles in gene bodies and promoters have been obtained from averaging per cent methylation values for CpG dinucleotides in gene bodies and in promoter regions encompassing the transcription start sites [TSS−2000 bp, TSS+100 bp], respectively.

Locally weighted scatterplot smoothing (lowess) with 30% bandwidth has been applied before calculating the Spearman correlation between fibroblast log_2[FC] gene expressions and iPS log_2[FC], gene replication times, DNA methylation, DNase I hypersensitivity and H3K4me3 profiles for each chromosome. Significance of each chromosomal correlation has been estimated by 10^6 random permutations. When the obtained empirical P value was zero, we reported the theoretical one.

The above-mentioned chromosomal lowess smoothed profiles have been concatenated from HSA1 to HSAX to calculate the overall Spearman correlation with gene expression. |Overall correlations| >0.2 are all significant ($P < 10^{-300}$).

51. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325,** 1246–1250 (2009).
52. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
53. Grad, I. *et al.* NANOG priming before full reprogramming may generate germ cell tumours. *Eur. Cells Mat.* **22,** 258–274 (2011).
54. Knowles, D. G., Roder, M., Merkel, A. & Guigo, R. Grape RNA-Seq analysis pipeline environment. *Bioinformatics* **29,** 614–621 (2013).
55. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
56. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13,** R87 (2012).
57. John, S. *et al.* Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol. Biol.* Ch. 27, Unit 21 27 (2013).