






Epigenetic control of chromosome-associated lncRNA genes essential for replication and stability

Received: 2 May 2022

Accepted: 13 October 2022

Published online: 22 October 2022

 Check for updates

Michael B. Heskett ^{1,2}, Athanasios E. Vouzas³, Leslie G. Smith⁴, Phillip A. Yates⁴, Christopher Boniface ⁵, Eric E. Bouhassira⁶, Paul T. Spellman ^{2,5}, David M. Gilbert ⁷ & Mathew J. Thayer ⁴ ✉

ASARs are long noncoding RNA genes that control replication timing of entire human chromosomes in cis. The three known ASAR genes are located on human chromosomes 6 and 15, and are essential for chromosome integrity. To identify ASARs on all human chromosomes we utilize a set of distinctive ASAR characteristics that allow for the identification of hundreds of autosomal loci with epigenetically controlled, allele-restricted behavior in expression and replication timing of coding and noncoding genes, and is distinct from genomic imprinting. Disruption of noncoding RNA genes at five of five tested loci result in chromosome-wide delayed replication and chromosomal instability, validating their ASAR activity. In addition to the three known essential cis-acting chromosomal loci, origins, centromeres, and telomeres, we propose that all mammalian chromosomes also contain “Inactivation/Stability Centers” that display allele-restricted epigenetic regulation of protein coding and noncoding ASAR genes that are essential for replication and stability of each chromosome.

For the vast majority of mammalian DNA, homologous regions on chromosome pairs replicate in a highly synchronized manner^{1–3}. However, genetic disruption of non-protein coding ASAR (“ASynchronous replication and Autosomal RNA”) genes causes a delay in replication timing on individual human chromosomes in cis, resulting in highly asynchronous replication between pairs of autosomes^{4–6}. The first ASAR genes were identified from a genetic screen designed to identify loci on human chromosomes that when disrupted resulted in chromosome-wide delayed replication^{4,6,7}. This screen identified five balanced translocations, affecting eight different autosomes, all displaying delayed replication along the length of the chromosomes⁷. Characterization of two of the translocation breakpoints identified discrete cis-acting loci where translocations or deletions resulted in

delayed replication^{4,6}. Molecular examination of the disrupted loci identified two lncRNA genes located on chromosomes 6 and 15, and were named ASAR6 and ASAR15, respectively^{4,6}. The ASAR6 and ASAR15 lncRNAs are extremely long (>200 kb), and remain associated with the chromosome territories where they are transcribed^{4,6,8–10}. These studies defined the first cis-acting loci that control replication timing and structural stability of individual human autosomes.

One unusual characteristic of ASAR genes is that they are expressed from only one allele^{4–6}. In contrast, the majority of genes on mammalian autosomes are expressed from both alleles, i.e., they display bi-allelic or balance expression. Allelic expression imbalance (AEI) of protein coding and noncoding genes is well established, and can arise from several distinct mechanisms. For example, AEI can arise due

¹Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. ²Department of Molecular and Medical Genetics Oregon Health & Science University, Portland, OR 97239, USA. ³Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA. ⁴Department of Chemical Physiology and Biochemistry Oregon Health & Science University, Portland, OR 97239, USA. ⁵Cancer Early Detection Advanced Research Center, Knight Cancer Institute Oregon Health & Science University, Portland, OR 97239, USA. ⁶Department of Cell Biology and Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461, USA. ⁷San Diego Biomedical Research Institute, San Diego, CA 92121, USA. ✉e-mail: thayerm@ohsu.edu

to heterozygosity at DNA sequence polymorphisms within cis-acting elements that influence the efficiency with which a gene will be transcribed (i.e., expression quantitative trait loci or eQTL; reviewed in refs. 11, 12). AEI may also occur through epigenetic programming to regulate gene dosage or to provide “exquisite specificity”, where the most extreme form of AEI is referred to as mono-allelic expression (reviewed in refs. 13–17). Genomic imprinting is a well-established form of epigenetically programmed AEI occurring in a parent of origin specific manner (reviewed in refs. 18, 19). Alternatively, epigenetically programmed AEI occurring in a random manner with respect to parent of origin has been observed for as many as 10% of autosomal genes (e.g., olfactory receptors, immunoglobulins, and T cell receptors; refs. 20–23). The three known ASAR genes display random epigenetically programmed AEI^{4–6,9}.

Another unusual characteristic of ASAR genes is that they display asynchronous replication timing (ASRT) between alleles^{4–6,9,24}. In contrast, the timing of DNA replication on autosome pairs occurs in a highly synchronous manner^{1,2}. ASRT can arise by different mechanisms. For example, ASRT can be caused by heterozygosity at DNA sequence polymorphisms that dictate the time during S phase that a locus will be replicated (i.e., replication timing quantitative trait loci or rtQTL; refs. 25–27). In addition, ASRT occurs at regions of the genome that display epigenetically programmed AEI. Thus, both imprinted and random AEI genes located on autosomes display ASRT^{1,13,28–32}. The three known ASAR genes display random epigenetically programmed ASRT between alleles^{4–6,9,24}.

ASAR RNAs share similarities with the vlinc (very long intergenic non-coding) class of RNAs. The vlincRNAs were characterized as RNA Polymerase II products that are nuclear, non-spliced, non-polyadenylated transcripts of >50 kb of contiguously expressed sequence that are not associated with protein coding genes³³. There are currently >2700 annotated human vlincRNAs, which are expressed in a highly cell type-restricted manner^{33–36}. The annotated vlincRNAs have been referred to as genomic “Dark Matter” because the vast majority of them have unknown function. However, they occupy >10% of the human genome and represent >50% of the non-ribosomal RNA within the cell^{33–36}. Previously, we found that the genomic region annotated as expressing vlinc273 (>185 kb in length) has all of the physical characteristics that are shared between ASAR6 and ASAR15, and CRISPR/Cas9-mediated deletion of the vlinc273 genomic locus resulted in delayed replication of chromosome 6, indicating that vlinc273 is an ASAR (which we named ASAR6-141⁵). ASAR6-141 resides within a cluster of 6 vlincRNA genes, with all 6 being expressed in a partially overlapping set of human tissues^{5,36}. Furthermore, the original ASAR6 gene resides within a larger ~1 mb genomic region of ASRT that also contains two other vlincRNA genes, which are also expressed in different human tissues^{5,9}. Taken together, these observations raise the intriguing possibility that other vlincRNAs are also ASARs, and that multiple ASARs can be clustered at the same locus, expressed in different cell types, and associated with the same ASRT region.

Deletion and ectopic integration analyses demonstrated that the chromosome-wide effects on replication timing of ASAR6 and ASAR15 map to the antisense strand of LINE1 (L1) retrotransposons located within the ASAR6 and ASAR15 RNAs¹⁰. Targeting the L1 sequences, using oligonucleotide-directed RNA degradation, revealed a functional role for the L1 sequences within ASAR6 RNA in controlling chromosome-wide replication timing¹⁰. Previous support for a role for L1s in epigenetically programmed AEI came from the observation that L1s are present at a relatively high local concentration (>18%) near both imprinted and random AEI genes located on autosomes³⁷. The three known ASAR genes contain >30% L1 sequences within the transcribed regions^{4–6}.

Altogether, our findings have suggested that ASARs may be ubiquitous essential cis-acting elements of chromosome replication and integrity. Indeed, we previously found that ~2.5% of chromosome

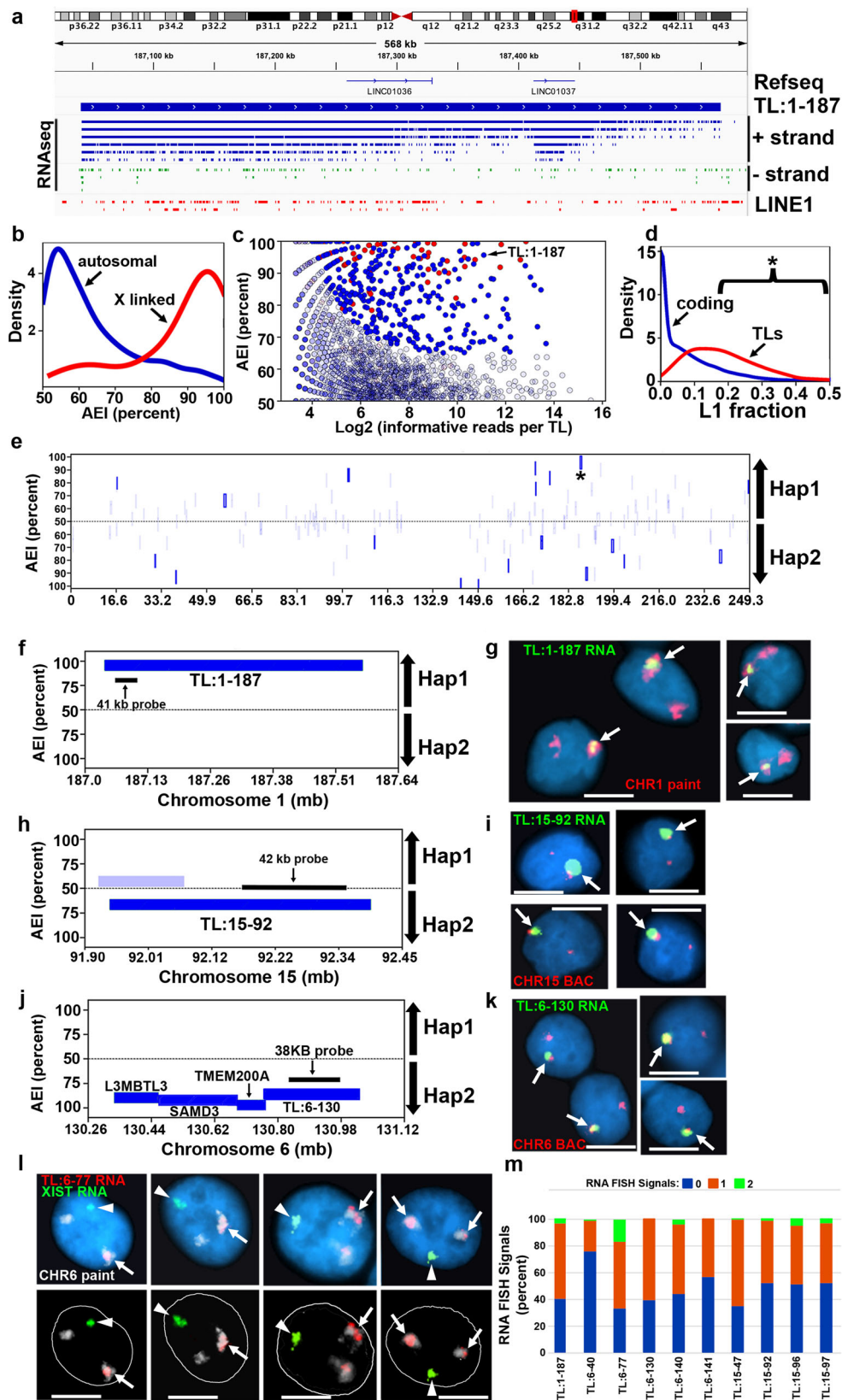
translocation products, induced by two different mechanisms (ionizing radiation or Cre/loxP), result in delayed replication timing of entire human chromosomes^{7,38}. Two of the Cre/LoxP-induced translocations with delayed replication were further characterized and found to have disrupted ASAR genes^{4,6}, suggesting that ~2.5% of the human genome encodes ASARs. Here, we directly tested this hypothesis by taking advantage of several notable characteristics of the three known ASARs to identify additional ASAR genes on human autosomes: (1) long contiguously transcribed regions of >180 kb, (2) epigenetically regulated AEI; (3) epigenetically regulated ASRT; (4) high density of L1 sequences; and (5) spatial retention of the RNA on the parent chromosome. We used RNA-seq, Repli-seq, and RNA-DNA FISH assays on single cell-derived clones from lymphoblastoid cell lines (LCLs) isolated from two unrelated individuals, both with haplotype phased genomes, to identify additional ASAR candidates on human autosomes. We chose female cells for this analysis to take advantage of X chromosome inactivation as an internal control for epigenetically controlled AEI and ASRT that occurs between the active and inactive X chromosomes. We identified hundreds of autosomal loci that display epigenetically controlled AEI and ASRT that is comparable to that observed on the X chromosome. Genetic deletion assays validated the biological activity of five out of five tested ASAR candidates. Our work identifies an unanticipated, widespread epigenetic program that occurs at hundreds of autosomal loci, effects AEI and ASRT of both noncoding and protein coding genes, has profound roles in chromosome structure stability, is predominant over the cis-effects of rtQTLs, and is distinct from genomic imprinting.

Results

Transcribed loci with allelic expression imbalance are widespread on autosomes

One complicating factor associated with both vlinc and ASAR RNAs is their extreme length. Thus, determining if these long RNA species represent single >50 kb contiguous transcripts or represent multiple overlapping transcripts with different start and stop sites is difficult using existing technologies. The vlincRNAs were annotated using tiling of contiguous RNA-seq reads across a given locus allowing gaps in the contigs of ≤7 kb to accommodate repetitive elements, e.g. ALUs and LINEs, that don't allow for unique mapping of the RNA-seq reads to the genome^{34,36}. This complication is also evident for all of the known ASAR RNAs, where the long length (>180 kb) and high L1 content (>30%) make annotation of the transcripts difficult. Given these limitations, we refer to regions of the genome with >50 kb of contiguous transcription of non-coding DNA as transcribed loci (TL) throughout this manuscript.

We first sought to identify TL that are expressed in a human lymphoblastoid cell line (LCL). For this analysis we used publicly available data from a nuclear-enriched, ribosomal RNA depleted, strand-specific, RNA-seq dataset from GM12878 (see ENCODE³⁹). Using a strategy of merging contiguous reads^{34,35}, a strand-specific call set of 1570 TL was defined. An example of a TL is shown in Fig. 1a, and shows the genomic location and the strand-specific, contiguous coverage of the RNA-seq reads for TL:1-187. This TL covers ~500 kb of genomic DNA, contains two RefSeq lincRNAs (LINC01036 and LINC01037), is not associated with any known protein coding gene, and contains 562 heterozygous SNPs within the transcribed region. Next, utilizing a fully haplotype-phased reference genotype (i.e., one haplotype block spanning each chromosome) for GM12878 to enumerate informative allele-specific reads⁴⁰ we assessed AEI of all 1570 TL. AEI was defined as biased allelic expression outlying the binomial distribution (FDR q -value <0.01) and greater than 2.5 standard deviations above the mean bias controlled for expression level (see methods). GM12878 cells are female and represent a poly-clonal pool of EBV-transformed lymphocytes. Random X chromosome inactivation among different cells in the population would be indicated by an equal number of RNA-seq reads



from the maternal (haplotype 1) and paternal (haplotype 2) X chromosomes. However, deviation from equal inactivation of each parental allele is common in the general female population, and is known as skewing^{41,42}. We found dramatic AEI of both protein-coding and non-coding RNAs expressed from the X chromosome, indicating that the LCL pool of cells in GM12878 has skewed X inactivation

(see Supplementary Data 1). We identified 78 TL that were expressed from the X chromosome, and 60 of these TL show AEI (Fig. 1b, c). The distribution of AEI of the X-linked TL also revealed the presence of 18 TL with equivalent expression from both X chromosomes. These bi-allelic TL are located in regions of the X chromosome that are known to escape X inactivation (Supplementary Data 1; and see refs. 42–44).

Fig. 1 | Genome-wide allelic expression imbalance and spatial retention patterns of TL in GM12878. **a** Genome browser view of a representative example of an intergenic non-coding contiguously transcribed locus on the plus strand of chromosome 1. The location of TL:1-187 with 436,832 independent RNA-seq reads from the plus strand (+ strand; blue) and 592 RNA-seq reads from the minus strand (- strand; green) are shown. The location of 221 LINE1 sequences, totaling 153,348 bp, representing ~28% of the transcribed sequence are shown in red. **b** Genome-wide distribution of AEI of TLs on all autosomes (blue) and the X chromosome (red). We define AEI as an allelic-bias that is outlying parametric (FDR-Benjamini-Hochberg corrected binomial test q -value ≤ 0.01) and non-parametric (allelic-bias ≥ 2.5 S.D. above autosome-wide mean, controlled for expression level) estimation of the null-distribution of bi-allelic expression. **c** Scatter plot of AEI of autosomal TL (blue dots) and X-linked TL (red dots) as a function of the number of informative reads. Opaque dots are outliers on the genome-wide distribution of AEI. **d** Distribution of

the fraction of LI derived sequence in TL and coding genes. The bracket and * highlight TLs with $>18\%$ LI sequence. **e** Chromosome 1 view of AEI of TL (opaque: statistically significant). The position of TL:1-187 is shown (*). **f, h, j** Zoom-in views of AEI of representative TL (dark blue = statistically significant; light blue not significant). The location and size of Fosmid probes (see Supplementary Data 2) used for RNA FISH are shown. **g, i, k** RNA-DNA FISH images of TL:1-187, TL:15-92, and TL:6-130 expression in single cells. **l** RNA-DNA FISH image of TL:6-77 RNA (red), CHR6 DNA (chromosome paint, white), and XIST RNA (green) visualized within individual cells, top and bottom panels represent the same four cells with the nuclear outline drawn in white. **m** Percent of XIST positive nuclei exhibiting strong AEI of nine representative TLs, as measured by DNA/RNA FISH visualization (Percent of nuclei with 0, 1 or 2 RNA hybridization signals). Source data are provided as a Source data file.

Table 1 | Summary of AEI and VERT loci

Sample	TLs	AEI TLs	AEI coding	Random AEI TL	Random AEI coding	VERT loci	AEI TL within VERT	AEI coding within VERT
EB3_2 (6 clones)	2758	644	941	68	63	105	82	80
GM12878 (parent)	1570	300	380	N/A	N/A	N/A	N/A	N/A
GM12878 (2 clones)	1903	378	613	6	4	144	44	57

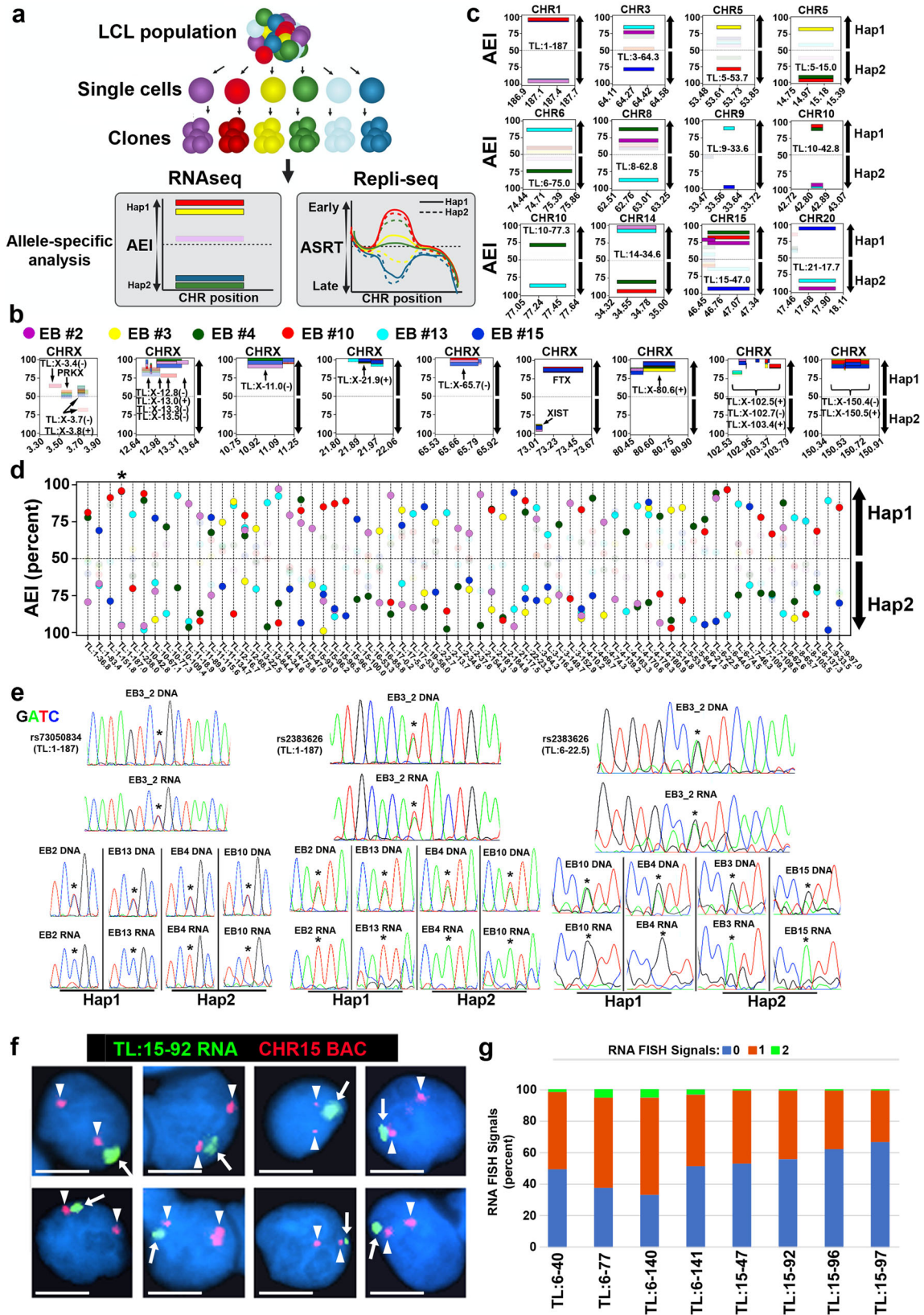
Given the dramatic skew in X inactivation, and anticipating a similar skew in random AEI on autosomes, we next assessed the autosomal TL for allele-specific expression. In contrast to the X chromosome TL, the autosome-wide distribution of AEI is consistent with bi-allelic expression for the majority of TL, but 300 unique TL (encoded by ~90 mb or ~3% of the genome) have AEI that significantly deviated from the null distribution of equal expression from both alleles (Fig. 1b, c; and Table 1). These TL contain a high LI content compared to introns within coding genes (Fig. 1d), which is one of the characteristics of the known ASARs. The TL with AEI are distributed along the length of every autosome pair and are expressed in a reciprocal pattern (see Supplementary Data 1). Figure 1e shows allelic expression analysis of all of the TL on chromosome 1, and highlights the 19 TL that display AEI, including TL:1-187 (see Fig. 1a, c).

To determine if TL RNAs with AEI remain associated with their parent chromosomes, we used RNA-DNA FISH to assay for chromosome territory localized retention of TL:1-187 plus 8 additional TL in GM12878 cells. For this analysis we used Fosmid probes to detect TL RNA (Fig. 1f, h, and j; and see Supplementary Data 2) and a chromosome-specific paint, BAC (Bacterial Artificial Chromosome) or centromeric probe to detect DNA for the chromosome of interest. We detected RNA hybridization signals that remain associated with their respective chromosomes for all 9 TL RNAs (for examples, see Fig. 1g, i, k, l), which is consistent with the spatial-expression pattern observed for the known ASARs⁴⁻⁶. For an independent assessment of mono- and bi-allelic expression, we quantified the frequency of single and double sites of RNA hybridization. For this analysis we included an XIST RNA FISH probe as positive control (GM12878 have a single large site of RNA hybridization signal for XIST in $>90\%$ of cells). Scoring only cells with a single XIST RNA FISH signal indicated a wide variability in the number of TL RNA hybridization signals, with zero, one or two sites of RNA FISH hybridization per nucleus (Fig. 1m). Single sites of RNA hybridization were detected for all TL ranging from 25 to 75% of cells, and two sites of RNA hybridization were detected for 7 out of 9 TL, ranging from 2 to 15% of cells (see Fig. 1l for examples). We note that the size of the RNA hybridization signals detected by all TL probes was variable, ranging in size from large clouds that occupy the entire chromosome territory to relatively small sites of hybridization that remain tightly associated with the expressed allele.

The observations described above indicate that ~300 autosomal TL are expressed in an allele-restricted manner in GM12878. We next sought to determine if the AEI of TL was present in LCLs from a second

unrelated individual, and if the AEI could come from either allele in multiple single-cell derived clones, which would distinguish between random epigenetically programmed AEI from imprinted AEI and AEI caused by genetic polymorphisms (i.e., eQTL). For this analysis, we used EB3_2 LCLs, which were isolated from a female with a haplotype-phased genome⁴⁵. By leveraging the B-cell origin of LCLs, we isolated six single-cell derived clones, as defined by the presence of unique immunoglobulin gene rearrangements (see Supplementary Data 4; and refs. 6, 21, 46). The EB3_2 clones (EB2, EB3, EB4, EB10, EB13, EB15) are expected to be isogenic except at the regions of immune gene-related somatic rearrangements, and were expanded for >25 population doublings prior to the generation of nuclear enriched RNA-seq and Repli-seq libraries (see Fig. 2a). Because EB3_2 cells are female, we first queried AEI on the X chromosome. The six EB clones display AEI of 66 TL and 127 protein coding genes located on the X chromosome (Supplementary Data 1). The orientation of the expressed alleles indicated that all six EB clones have the same active and inactive X chromosomes, haplotype 1 (maternal) and haplotype 2 (paternal), respectively. Figure 2b shows examples of the AEI of X chromosome TL and protein coding genes, and indicates strong expression from haplotype 1 in all 6 clones.

By analyzing expression from autosomes across all six clones, 2692 TL were defined in the EB3_2 clone set, with 644 TL, or ~24%, showed statistically significant AEI (Table 1). Strikingly, 68 TL were identified that display AEI from opposite homologs in two or more clones (Fig. 2c, d), which is consistent with random epigenetically programmed AEI. In addition, some clones display equivalent levels of expression between alleles (i.e., bi-allelic expression) at the same TL that display AEI of either allele in other clones, while in other clones there is undetectable expression. The other 576 autosomal TL with AEI showed expression from single alleles, both alleles or neither allele in one or more clones (Supplementary Data 1). This AEI pattern is not consistent with genomic imprinting, nor with the presence of eQTL, but may be due to random epigenetically programmed AEI where we have not analyzed enough clones to detect expression from opposite alleles. Regardless, validation of the clone-specific AEI from autosomal TL was then performed by Sanger sequencing of DNA and cDNA from the parent multiclonal population and from individual clones (see Fig. 2e for examples). These data indicate that expression of 68 autosomal TL can originate from either haplotype 1, haplotype 2, both or neither in different clones isolated from the same individual. These observations indicate that each allele had acquired the expressed or



silent state independently from the other allele, and because all possible combinations were detected in different clones from the same individual these differential expression states must be under epigenetic control.

Next, to determine if the TL RNAs that show AEI in LCLs are associated with their parent chromosome in human primary cells, we

carried out RNA-DNA FISH using probes to 8 different TL on primary blood lymphocytes (PBLs). We detected chromosome-associated RNA hybridization signals for all 8 TL. An example of this analysis is shown for TL:15-92 in Fig. 2e. Quantification of the number of RNA FISH signals in >100 cells indicated that all 8 TL were expressed from single chromosomes in >80% of cells (Fig. 2g). We note that the size of the

Fig. 2 | Haplotype resolved analysis of allelic expression imbalance of TL RNAs. **a** Sub-cloning and allele-specific genomic analysis scheme for 6 clones from the EB3_2 lymphoblastoid cell line. The left panel shows phased RNA-seq data at one hypothetical TL showing AEI in different clones. The right panel shows the phased Repli-seq data at one hypothetical replication domain showing synchronous or asynchronous replication. **b, c** Zoom-in views of examples of TL RNAs that display AEI (Hap1: haplotype 1, Hap2: haplotype 2) on the Y axis and the genomic positions (megabases) for the X chromosome (CHRX; panel **b**) or autosomes (panel **c**) positions are shown on the X axis. TL and protein coding genes are labeled and marked by arrows. The panel at -3.5 mb of the X chromosome shows bi-allelic expression of PRKX, a known escape gene⁴⁴, and three bi-allelic TL. **c** Each panel represents the AEI and location of a prominent TL from 9 different autosomes

(CHRI, etc). **d** All TL RNAs that display AEI from different alleles within the 6 EB3_2 clones (X-axis: chromosome start-position; Y-axis: AEI as percent). The asterisk marks TL:1-187. **e** Sanger sequencing traces of PCR products generated from genomic DNA and cDNA illustrating heterozygosity in genomic DNA and mono-allelic expression of either allele from TL RNAs (EB3_2: parent cell population, EB2, 3, 4, 10, 13, and 15: clones). **f** RNA-DNA FISH in PBLs using a Fosmid probe to TL:15-92 to detect RNA and a chromosome 15 BAC to detect DNA. **g** Expression of TL in female PBLs. Quantification of the number of RNA FISH signals (shown 0, 1, or 2) for 8 different TL in nuclei that contained a single XIST RNA FISH signal. In the nuclei with TL hybridization signals, we detected >80% of nuclei with single sites of TL RNA hybridization for all 8 TLs. Source data are provided as a Source data file.

RNA hybridization signals detected by all of the TL probes was variable, ranging in size from large clouds to relatively small sites of hybridization. We also detected two sites of hybridization for all TL probes in 2 to 20% of cells (Fig. 2g), which is consistent with the LCL RNA-seq data (see Fig. 2b, c), and indicates that expression and chromosome territory localization of this set of TL RNAs can be bi-allelic in at least some cells in the PBL population. Therefore, we conclude that mono- and bi-allelic expression of chromosome-associated RNAs can be detected for all 8 TL in primary cells.

Variable epigenetic replication timing occurs at hundreds of autosomal loci

ASRT occurs at regions of autosomes that display epigenetically programmed AEI^{13,28–32}. To assess the relationship between AEI and ASRT in LCLs, we next performed allele-specific Repli-seq on the six EB3_2 clones (see Fig. 2a), and on two single-cell-derived clones from GM12878. Because EB3_2 and GM12878 cells are female-derived LCLs, we used the differences in replication timing between the active and inactive X chromosomes as internal controls. First, allele-specific replication timing profiles were generated for each haplotype from each clone from both sets of clones, totaling 12 and 4 allele-specific RT profiles per chromosome for the EB3_2 and GM12878 clone sets, respectively. Second, differences in RT among each clone set was assessed by measuring the standard deviation (SD) of three subdivisions of the RT profiles. For each clone set, we generated a “combined RT profile” that included haplotype 1 plus haplotype 2 (e.g. all 12 alleles in the six EB3_2 clones); and two “allele-specific RT profiles” that included only haplotype 1, or only haplotype 2. Third, to identify classical ASRT regions within individual clones, we analyzed the difference between haplotype 1 and haplotype 2 RT profiles within each clone separately by comparing the absolute difference in allele-specific read counts in the Repli-seq data, which can detect ASRT resulting from either genetic or epigenetic mechanisms³. Figure 3a illustrates hypothetical RT profiles for a region of the genome that shows synchronous or asynchronous RT, as well as epigenetic variability that could occur in the RT profiles in different clones from the same individual.

The genome-wide distribution of the SD for the “combined RT profile” on autosomes for each 250 kb genomic region revealed a single large peak ($SD < 1$) of synchronous replication with a shallow but long right tail of outliers ($SD \geq 2.5 \times SD + \text{mean}$: >0.92 for the EB3_2 clones and >0.90 for the GM12878 clones; Fig. 3b, c). In contrast, the distribution of the SD of the “combined RT profile” along the X chromosome revealed two peaks. One peak is superimposed with the synchronous autosomal peak ($SD < 1$), and these X chromosome regions represent the late replicating regions on both the active and inactive X chromosomes (Fig. 3b, c, f). The X chromosome regions within the second broader peak ($SD > 1$) map to the regions of the X chromosome that has early replication (Fig. 3f), and the early replicating DNA is from haplotype 1, which represents the active X chromosome in all six EB3_2 clones (see Fig. 2b and Supplementary Data 1). The “combined RT profile” analysis of the two

GM12878 clones revealed a similar profile for the autosomes and X chromosomes (Fig. 3c), with haplotype 2 representing the early replicating active X chromosome in both clones (see Supplementary Data 1).

For the “allele-restricted RT profiles”, we compared the RT profiles of each haplotype in all 6 EB3_2 and both GM12878 clones independently from the RT profiles of the other haplotype in each clone set (Fig. 3d, e). This analysis allowed us to detect outliers in the RT profiles on each allele in the two clone sets separately ($SD \geq 2.5 \times SD + \text{mean}$: >0.96 for Hap 1 and >0.98 for Hap 2 in the EB3_2 clones; and >0.92 for Hap 1 and >0.94 for Hap 2 in the GM12878 clones). In contrast to the “combine RT profile” analysis described above (Fig. 3b, c), the “allele-restricted RT profile” analysis identified only four relatively small outlier regions on the X chromosome, two on the active and two on the inactive X chromosomes (Fig. 3f–j). All four of these regions contain TL with strong AEI (Fig. 3g–j), and one of these regions contains two loci, XACT and DXZ4, that are known to be important for X inactivation in humans (see Fig. 3j; and refs. 47–50). Because this Repli-seq analysis was performed on clones isolated from the same individual and the differences in replication timing between the active and inactive X chromosomes are known to be under epigenetic control, we conclude that this analysis allows us to detect epigenetically controlled allele-specific replication timing.

The “combined RT profile” analysis indicated that the vast majority of autosomal DNA replicated synchronously ($SD < 1$), with a relatively small number of loci (99 loci in the EB3_2 and 139 loci in the GM12878 clone sets) with significant differences in allele-specific RT (Supplementary Data 3). In addition, analyzing the “allele-restricted RT profiles”, we identified 108 haplotype 1 and 103 haplotype 2 outlier loci in the EB3_2 clones (Fig. 3d, e). In total, we identified outliers in the autosomal RT profiles for 248 loci in the EB3_2 plus GM12879 clones, with 37 loci shared between the two clone sets (211 loci representing -190 mb or -6% of the human genome; Table 1). Figure 4a shows the RT profiles on chromosome 5 in the six EB3_2 clones, and indicates that there are 9 loci in the “combined RT profile” with significant differences in RT, and the “allele-restricted RT profiles” indicated that these differences are associated with either haplotype 1, haplotype 2 or both. Figure 4b–i shows examples of 8 loci identified as outliers in both EB3_2 and GM12878 RT profiles. The differences in the “allele-restricted RT” profiles detected at outlier loci can affect either haplotype 1, haplotype 2 or both. Thus, some clones display large differences in RT between alleles at the same locus that displays synchronous replication in other clones, and the RT profiles in the synchronous clones can either have the earlier or later replication timing profile (Fig. 4b–i; also see Fig. 5b–j). Taken together, these observations indicate that each allele had acquired the earlier or later replication independently from the other allele. Because the differences in RT at these loci are quite variable and affect either allele independently, we refer to these loci as having Variable Epigenetic Replication Timing (VERT). Within the 99 VERT loci in the EB3_2 clone set we identified 82 TL that display AEI (Table 1, Supplementary Data 1 and 3; also see Figs. 4b–i and 5 below).

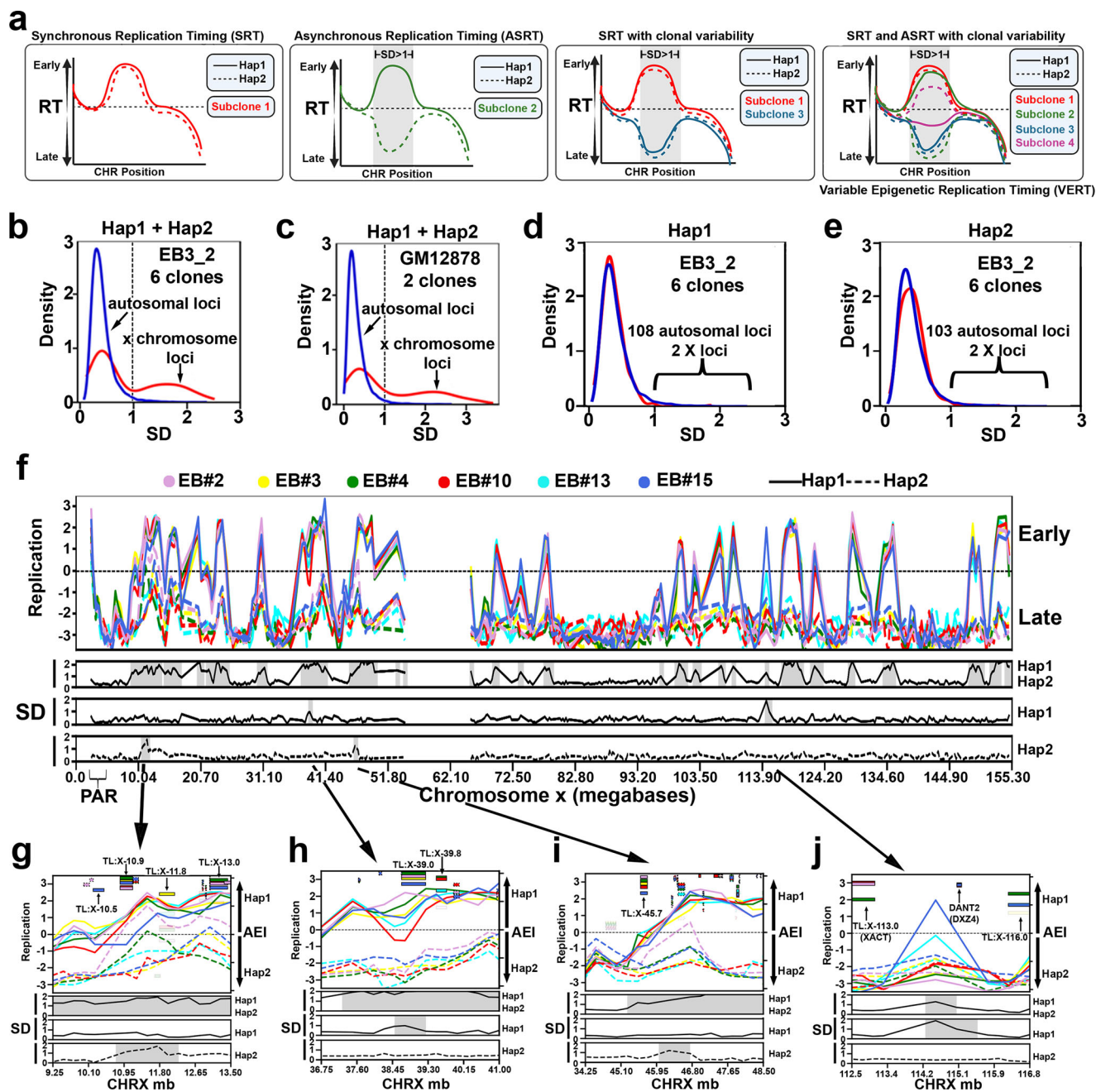


Fig. 3 | Haplotype resolved analysis of asynchronous replication timing.

a Illustration of a hypothetical genomic region with synchronous replication timing (SRT), asynchronous replication timing (ASRT), and possible clonal variability in both SRT and ASRT. Outliers in the standard deviation (SD) are highlighted in gray. **b, c** Genome-wide distribution of the SD of the combined RT of individual alleles among clones derived from EB3_2 (12 alleles) and GM12878 (4 alleles). **d, e** Genome-wide distribution of the SD of the allele restricted replication timing of each individual allele among clones derived from EB3_2 (**d**: 6 Hap1 alleles; **e**: 6 Hap2 alleles). Outliers were identified as: $SD \geq 2.5 \times SD + \text{mean}$: >0.96 for Hap 1 and >0.98 for Hap 2 in the EB3_2 clones and >0.92 for Hap 1 and >0.94 for Hap 2 in the GM12878 clones. **f** Chromosome X Early/Late RT profile with SD of haplotype-resolved replication timing from the 6 EB3_2 clones highlighting (gray shading) outlier regions from the “combined RT profile” (Hap1 plus Hap2) and from the “allele-

restricted RT profile” separately (Hap1 or Hap2). Each clone was color coded as shown, with haplotype 1 shown as a solid line, and haplotype 2 shown as a dotted line for both sets of clones. The left axis shows the RT profiles, with positive numbers representing early replication and negative numbers representing late replication. **g–j** Zoom in views of the 4 locations that contain outliers in SD of the “allele-restricted RT profiles”. Each clone was color coded as shown, with haplotype 1 shown as a solid line, and haplotype 2 shown as a dotted line for both sets of clones. The left Y axis shows the Early/Late RT profiles. The SD of the replication timing across each locus is shown below each panel. Areas highlighted in gray represent outliers in the SD. The right Y axis shows the AEI, with smooth rectangles representing TL and the stippled rectangles representing protein coding genes. The opaque rectangles show AEI, while the transparent rectangles show bi-allelic expression of TL and protein coding regions.

In addition, at any given TL with AEI the expression could come from either the earlier or later replicating allele, indicating that each allele had acquired the expressed or silent state independent of the replication timing within these VERT loci.

Asynchronous replication can occur at rtQTL, and lacks coordination between loci

By performing allele-specific Repli-seq on single-cell derived clones, we were able to address the consequences of DNA sequence

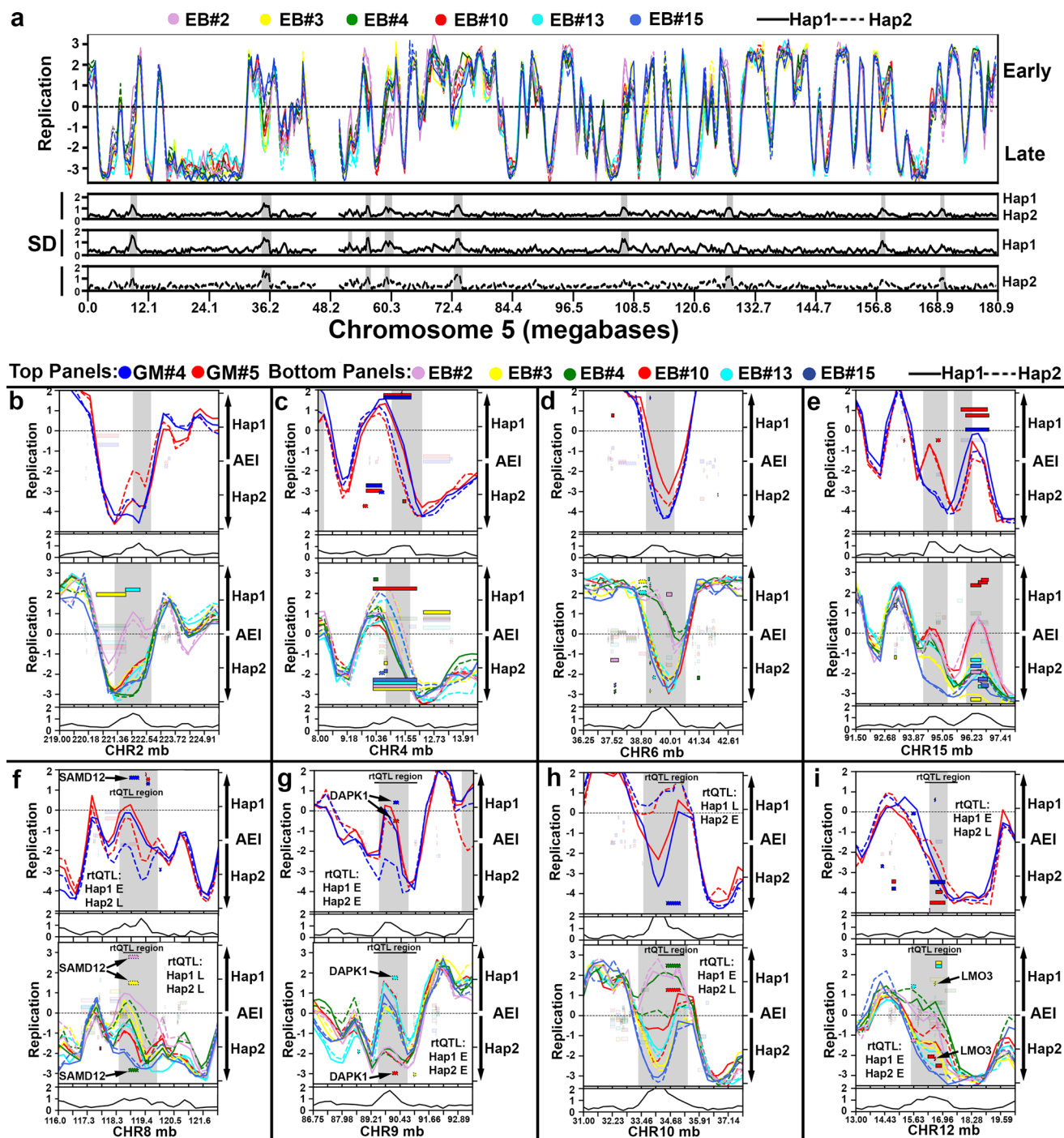


Fig. 4 | Haplotype phased analysis of ASRT on autosomes. **a** Chromosome 5 RT profile from the 6 EB3_2 clones, highlighting regions with standard deviation (SD) > 1 from both the “combined allele RT profile” (Hap1 and Hap2) and from the “allele-restricted RT profile” separately (Hap1 or Hap2). **b–i** Representative examples of regions present in both GM12878 (top panels) and EB3_2 (bottom panels) clones with variable epigenetic replication timing between alleles. Each clone was color coded as shown, with haplotype 1 shown as a solid line, and haplotype 2 shown as a dotted line for both sets of clones. The left Y axis shows the replication timing (Early/Late) profiles. The SD of the replication timing

across each locus is shown below each panel. The right Y axis shows the AEI, with smooth rectangles representing TL and the stippled rectangles representing protein coding genes. The opaque rectangles show AEI, while the transparent rectangles show bi-allelic expression of TL and protein coding regions. Chromosome number and genomic positions (megabases) are shown on the X axis. Areas highlighted in gray represent outliers in the SD. **f–i** The location of rQTL at VERT regions, with the Early (E) and Late (L) alleles²⁶ for both GM12878 and EB3_2 are shown (also see Supplementary Data 5).

polymorphisms on ASRT. A previous study identified 20 polymorphic loci, known as rQTL, by examining RT profiles in human LCLs²⁶. We found that 12 of the 20 previously identified rQTL were within VERT loci detected in either the EB3_2, GM12878, or both Repli-seq data sets. Surprisingly, we found that these 12 loci display strong epigenetic

differences in replication timing regardless of which rQTL allele they contain (Supplementary Data 5). Examples of 4 different VERT loci that were detected in both GM12878 and EB3_2 clones, with the genotype and location of the associated rQTL, are shown in Fig. 4f–i (also see below). We found that differences in RT within these VERT regions

occur at loci that are either homozygous or heterozygous for the rtQTL alleles, demonstrating that the epigenetic effects described here are dominant over any DNA sequence polymorphisms that are present at these loci.

We next sought to address the question of coordination of asynchronous replication timing. We and others previously reported that ASRT of random mono-allelic genes is coordinated with other random mono-allelic genes on the same chromosome, suggesting that there is a chromosome-wide system that coordinates replication asynchrony of random AEI genes^{4,9,28,32,51}. More recently, the asynchrony associated with random AEI genes in mouse pre-B cells was proposed to be coordinated on all autosomes, resulting in only two “mirror-image” patterns of asynchronous replication^{52,53}. For the analysis of coordination at ASRT regions utilizing Repli-seq, we first defined ASRT regions as outliers on the distribution ($SD > 1$) of the difference between haplotype 1 and haplotype 2 RT profiles within each EB3_2 clone. We then classified each allele at each asynchronous region as either early or late within each clone, and compared the Early/Late status across multiple regions on individual chromosomes and between clones. We first analyzed the ASRT on the X chromosome. We detected dramatic coordination in the polarity of the ASRT regions along the length of the X chromosome in all six clones (see Supplementary Fig. 1), which is consistent with previously reported ASRT profiles for the active and inactive X chromosomes in LCLs⁵⁴. While we detected numerous loci on every autosome that display ASRT, we found no evidence for coordination, either on the same autosome or between autosomes (Supplementary Fig. 1b–d). Furthermore, by analyzing the Repli-seq data previously published by Blumenfeld et al for mouse pre-B cell clones⁵² using our “combined RT profile” and ASRT analyses described here, we detected loci in the mouse genome that display VERT, but did not detect coordination in the Early/Late replication timing pattern at loci that show ASRT, and therefore we do not detect a “mirror image” coordination pattern on mouse chromosomes in the Repli-seq data for mouse pre-B cell clones (Supplementary Fig. 2). Our data indicate that the clonal Early/Late pattern at multiple asynchronous loci on human or mouse autosomes is not coordinated on the same chromosome nor between different chromosomes.

Variable epigenetic AEI of protein coding genes occurs at VERT loci

The observations described above indicate that hundreds of TL exhibit AEI, and that this AEI often occurs at genomic loci detected as having VERT. We next determined if protein coding genes also display epigenetically programmed AEI within VERT loci. Utilizing allele-specific analysis of the RNA-seq data at protein coding genes, we identified 941 protein coding genes that display AEI, including 11 known imprinted genes (geneimprint.com; Supplementary Data 1). We also identified 63 protein coding genes that display AEI from opposite homologs in two or more clones (Fig. 5a and Table 1), which is consistent with random epigenetically programmed AEI. Similar to the TL with epigenetically programmed AEI, some clones display equivalent levels of expression between alleles (i.e., bi-allelic expression) while other clones express only one allele, and in yet other clones have undetectable expression. These observations are consistent with previous work that found a similar variability in mono-allelic expression of protein coding genes within LCL clones²¹. We found that 80 of the 941 protein coding genes with AEI are within VERT regions (Table 1; also see Supplementary Data 1 and 3). Figure 5b–j shows examples of the RT profiles and AEI of protein coding genes at genomic regions that also display VERT (also see Fig. 4f, g, i). These observations indicate that the expressed or silent state at these protein coding genes was acquired independently, and that the expressed or silent state is not associated with either earlier or later replication. Utilizing an ensemble of gene enrichment analysis tools⁵⁵, the 63 protein coding genes with random

epigenetically programmed AEI were found to be enriched for several biological processes, with homophilic cell adhesion being the most significant ($P = 1.17 \times 10^{-22}$; Fig. 5k).

Deletion of TL within VERT loci result in delayed replication timing in cis

The previous ASAR genetic disruption studies were carried out in cells that displayed strong AEI at the ASAR loci, and disruptions of the expressed alleles resulted in delayed replication while disruptions on the silent alleles did not^{4–6,10}. However, the variability in the AEI described in this report indicates that TL can be expressed from both alleles in at least some clones, and that the expressed alleles are independent of earlier or later replication at the locus. To determine if TL that reside in VERT regions control replication timing of entire chromosomes, we used CRISPR/Cas9 to delete 5 different TL genes. However, given the low cell cloning efficiencies of LCLs, and the relatively low efficiency of generating large deletions (>200 kb) using CRISPR/Cas9, we used HTD114 cells for this analysis. We chose HTD114 cells for this analysis because the high cloning efficiency allows for large CRISPR/Cas9 deletions, up to ~1.2 mb, as demonstrated during the characterization of ASAR6-141⁵. We also chose HTD114 cells for this analysis because they maintain AEI and ASRT of both imprinted and random mono-allelic genes, they have a stable karyotype that does not change significantly following transfection, drug selection and sub-cloning, and previous knockouts of the three known ASAR genes in HTD114 resulted in delayed replication timing, delayed mitotic condensation and chromosome structure instability in cis^{4–10}. Given the tissue-restricted expression pattern of TL, we first needed to determine which of the TL with AEI and VERT are expressed in HTD114 cells. Therefore, to identify TL that are expressed in HTD114 cells, we first queried RNA-seq data from HTD114 cells (see ref. 5), and subsequently used RNA-DNA FISH and Sanger sequencing of PCR products from reverse transcribed RNA for 16 TL to determine if the RNAs are associated with their parent chromosomes and if they show mono- or bi-allelic expression. For this analysis we used RNA-DNA FISH using Fosmid probes to each TL (see Supplementary Data 2) plus a chromosome centromeric probe to detect DNA. All 16 TL RNA hybridization signals were found to be associated with their parent chromosome centromeric DNA hybridization signals, and displayed single or double RNA FISH signals in the nuclei of HTD114 cells (for examples, see Figs. 6b, c and 7c, d, i, j). Because the epigenetic variability that results in AEI at TL can result in mono- or bi-allelic expression in individual clones, we chose 5 TL that were either mono- or bi-allelic in the HTD114 cells for genetic deletion assays.

TL:1-187 is expressed from chromosome 1 at -187–187.5 mb (see Figs. 1a, e–g and 2c), and Fig. 6a shows the AEI and RT profiles from the EB3_2 and GM12878 clone sets. Figure 6b and c shows results from the RNA-DNA FISH analysis in HTD114 cells, and indicates that single sites of RNA hybridization were detected in ~75% of cells and that two RNA hybridization signals were detected in ~25% of cells, suggesting that HTD114 cells express both alleles. Figure 6d shows sequencing traces from two different SNPs (see Supplementary Data 4) that are heterozygous in genomic DNA, and heterozygous in reversed transcribed RNA isolated from HTD114 cells, confirming that the TL:1-187 transcripts are generated from both alleles in HTD114 cells.

Next, we designed sgRNAs to unique sequences flanking the genomic region that expresses TL:1-187 (Supplementary Data 4). We expressed these sgRNAs in combination with Cas9 and screened clones for deletions using PCR primers that flank the sgRNA binding sites (see Supplementary Data 4). Because TL:1-187 expression is bi-allelic in HTD114 cells, we isolated clones that had heterozygous deletions affecting either homolog. We determined which allele was deleted based on retention of the different base pairs of heterozygous SNPs located within the deleted regions, and arbitrarily assigned the homologs as CHRIA or CHRIB (see Fig. 6d and Supplementary Data 4).

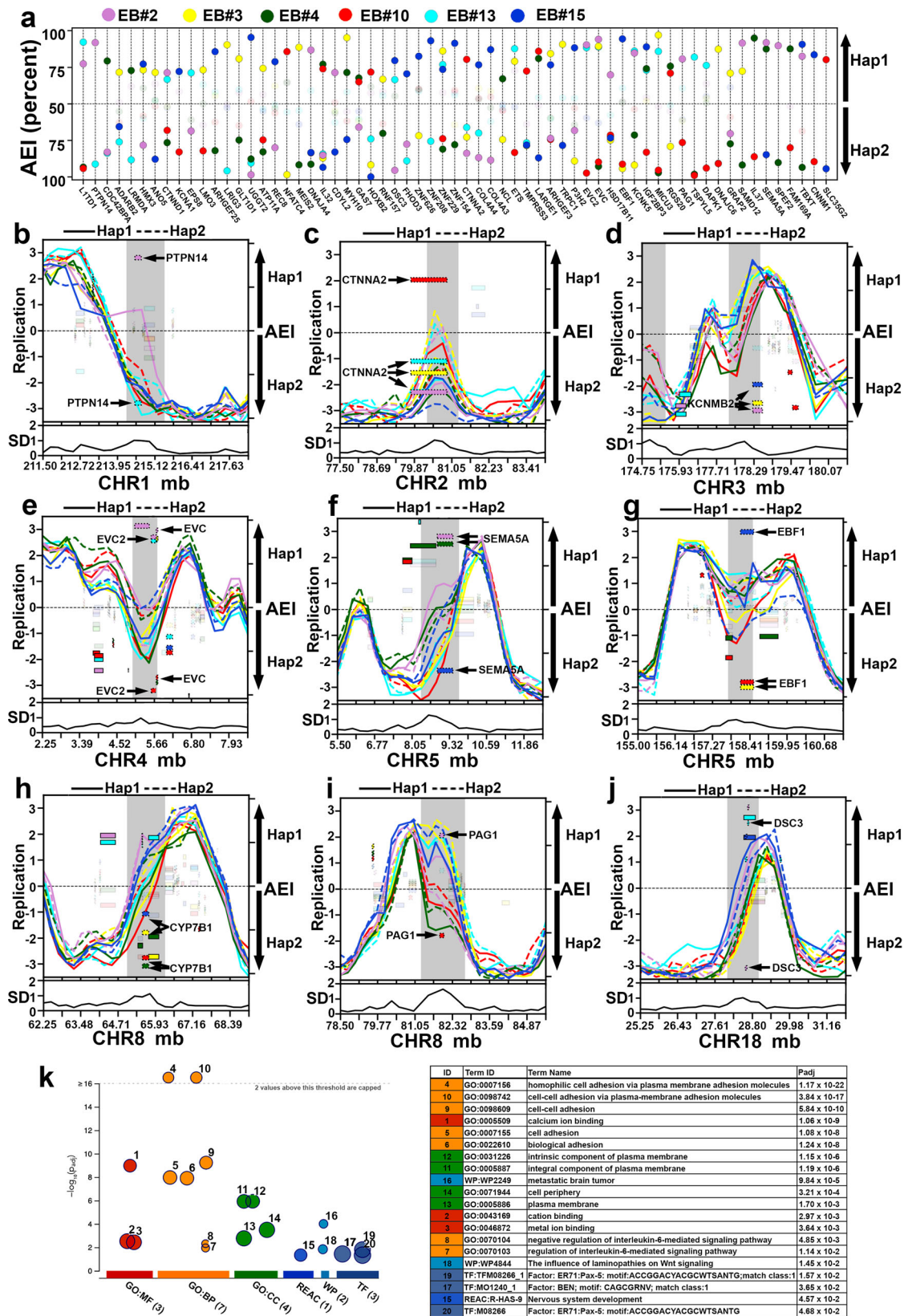
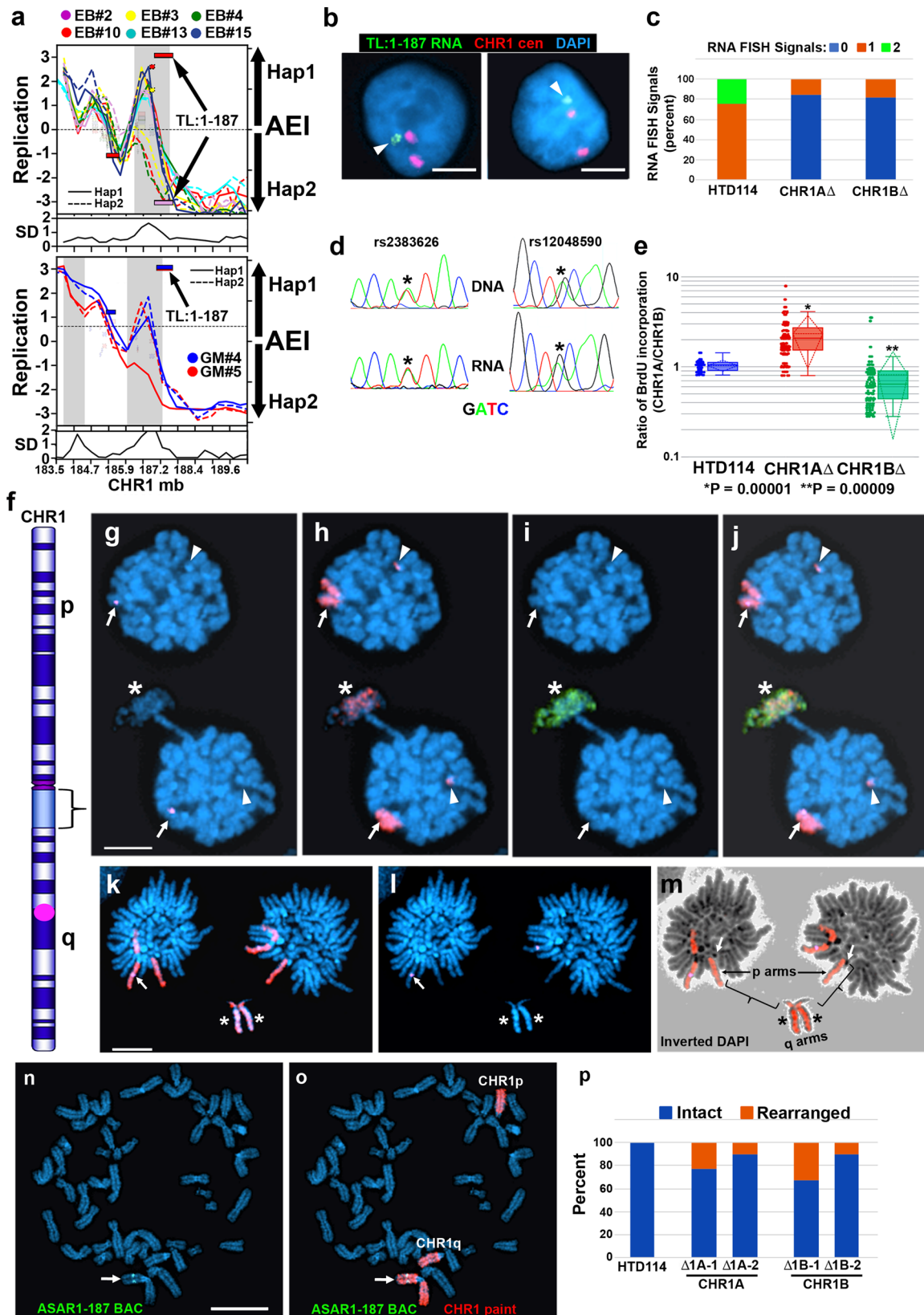


Fig. 5 | Haplotype resolved expression and replication of protein coding genes. **a** 63 protein coding genes display random epigenetic AEI in EB3_2 clones (X-axis: protein coding gene; Y-axis: AEI). **b–j** Examples of genomic regions that contain protein coding genes that display epigenetic AEI and VERT in the six EB3_2 clones. The left axis shows the replication timing (Early/Late) profiles, and the standard deviation (SD) across each locus is shown below each panel. Outliers in the standard deviation (SD) are highlighted in gray. The right Y-axis shows the AEI, with the

stippled rectangles representing protein coding genes and smooth rectangles representing TL. The opaque rectangles show AEI (FDR-BH alpha <= 0.01), while the transparent rectangles show bi-allelic expression of protein coding and non-coding transcripts. Chromosome number and genomic positions (megabases) are shown on the X axis. **k** Functional enrichment analysis of coding genes that display random epigenetic AEI in EB3_2 clones.



In addition, to determine how deletion of each allele affected the number of RNA hybridization signals, we carried out RNA-DNA FISH using the TL:1-187 probe in heterozygous deletion clones. Figure 6c shows the quantification of the RNA FISH signals, and indicates that deletions of TL:1-187 from either CHR1A or CHR1B resulted in cells with single sites of RNA hybridization signals, and no cells with two sites of

hybridization, confirming that both alleles of TL:1-187 are expressed in HTD114 cells.

We next assayed replication timing of the two chromosome 1 homologs using a BrdU incorporation assay⁵⁶. Cultures of cells were pulsed with BrdU for 5.5 h and mitotic cells harvested, processed for BrdU incorporation and subjected to FISH using a chromosome 1 paint

Fig. 6 | Delayed replication following disruption of TL:1-187. **a** AEI and ASRT at TL:1-187 in EB3_2 (top) and GM12878 (bottom) clones. The left Y-axis shows replication profiles and the standard deviation (SD). The right Y-axis shows AEI, haplotype 1 (Hap1), and haplotype 2 (Hap2). Outliers in the standard deviation (SD) are highlighted in gray. Chromosome position (megabases) is shown on the X axis. Smooth rectangles represent TL, stippled rectangles represent coding genes. Opaque rectangles show AEI, and transparent rectangles indicate bi-allelic expression. **b** RNA-DNA FISH using a TL:1-187 probe (green) to detect RNA and a chromosome 1 centromeric (red) probe to detect DNA. Arrowheads mark the RNA hybridization. **c** Quantification of the RNA signals for TL:1-187 in HTD114. An RNA FISH probe to an intron of *KCNQ5* served as positive control⁵. **d** Sequencing traces of PCR products at heterozygous SNPs (asterisks) within TL:1-187. **e** Ratio of BrdU incorporation in chromosome 1A divided by chromosome 1B in cells with heterozygous deletions of TL:1-187. Box plots indicate mean (solid line), standard deviation (dotted line), 25th, 75th percentile (box) and 5th and 95th percentile (whiskers)

and individual cells (single points). *P* values were calculated using the Kruskal–Wallis test⁷⁸. Values for individual cells are shown as dots. **f** Chromosome 1 showing the short (**p**) and long (**q**) arms, G-bands, heterochromatic block (bracket), and TL:1-187 (purple dot). **g–j** Delayed replication and condensation following deletion of TL:1-187. BrdU incorporation (green) and DNA FISH using a chromosome 1 paint (red) plus a TL:1-187 BAC (purple). A chromosome 1 translocation is marked with arrowheads. **k–m** DNA FISH using a chromosome 1 paint (red) plus the TL:1-187 BAC (purple, arrows). The lagging chromosome lacks the TL:1-187 BAC (asterisks). The short (**p**) and long (**q**) arms of the lagging chromosome are indicated, and arrows in panel m mark the centromeres of the lagging chromosome. **n, o** Chromosome following TL:1-187 deletion. DNA FISH using the TL:1-187 BAC (green) and a chromosome 1 paint (red). Scale bars are 10 μ m. **p** Quantification of chromosome 1 rearrangements. **c, e, p** Source data are provided as a Source data file.

probe plus a BAC probe from within the deleted region (Supplementary Data 2). As expected, prior to deletion of the TL:1-187 locus, the two chromosome 1 homologs display synchronous replication (see Fig. 6e). In contrast, cells containing heterozygous deletions of either allele resulted in significantly more BrdU incorporation in the deleted chromosome than the non-deleted chromosome, indicating that deletion of either TL:1-187 allele results in delayed replication timing in HTD114 cells (Fig. 6e). These results are consistent with the observation that both alleles of TL:1-187 are expressed, and that this locus controls replication timing of human chromosome 1. Therefore, we name this lncRNA gene ASARI-187.

One dramatic phenotype associated with chromosomes with disrupted ASARs is instability of the affected chromosomes, which is characterized by an increase in the rate of chromosomal rearrangements⁷, an increase in non-disjunction events resulting in cells with anaphase bridges, lagging chromosomes, and micronuclei^{8,9}, and an increase in endoreduplication resulting in an increase in the ploidy of the affected cells⁸. During the characterization of chromosomes with the ASAR deletions described here, we detected extremely late DNA replication (also called “G2 DNA synthesis”^{57,58}), delayed mitotic chromosome condensation, and lagging chromosomes during anaphase. Figure 6g–j shows an example of an anaphase cell with a lagging chromosome affecting chromosome 1 that is deleted for ASARI-187, and the lagging chromosome also displays delayed mitotic chromosome condensation and extremely late DNA replication with BrdU incorporation that occurred during G2, note that BrdU incorporation was not detected in any other chromosome. Another example of an abnormal anaphase cell with a lagging chromosome 1, also deleted for ASARI-187, is shown in Fig. 6k–m, and shows abnormal mitotic condensation in the large heterochromatic region on chromosome 1. Given the prevalence of these abnormal mitotic figures it is not surprising that we also detected numerous rearrangements involving chromosome 1 in cells with deletions of ASARI-187 in either CHR1A or CHR1B (Fig. 6n–o).

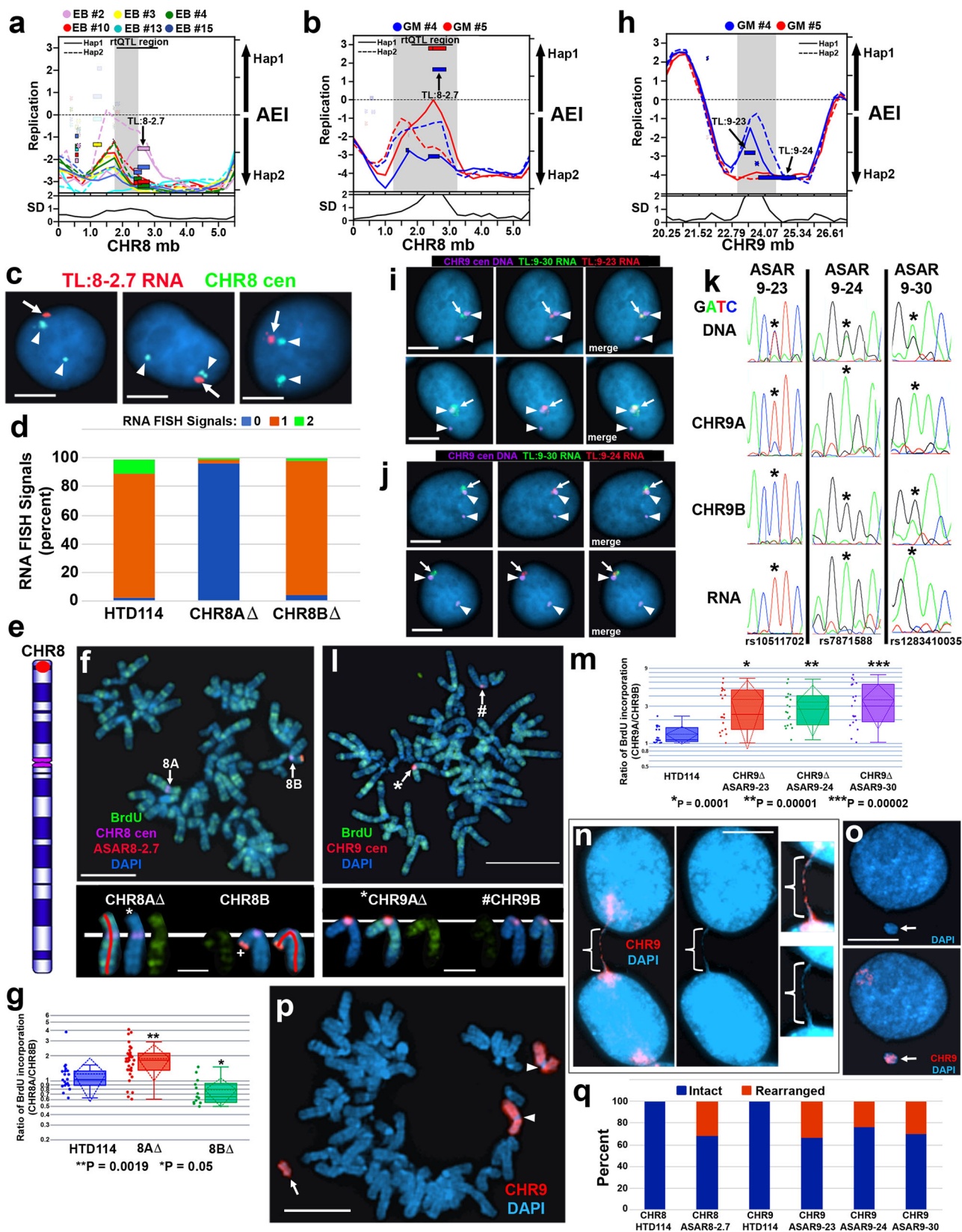
TL:8-2.7 is expressed from chromosome 8 at -2.5–2.8 mb (see Supplementary Data 1), and Fig. 7a and b shows the AEI and RT profiles from the EB3_2 and GM12878 clone sets. Figure 7c shows examples from the RNA-DNA FISH analysis in HTD114 cells, and indicates that single sites of RNA hybridization were detected in ~90% of cells (Fig. 7d). sgRNAs were designed to unique sequences flanking the genomic region that expresses TL:8-2.7 (Supplementary Data 4). The sgRNAs were expressed combination with Cas9, and clones were screened for deletions using PCR primers that flank the sgRNA binding sites (see Supplementary Data 4). Clones were isolated that had heterozygous deletions affecting either chromosome 8 homolog and we determined which allele was deleted based on retention of the different base pairs of heterozygous SNPs located within the deleted regions, and arbitrarily assigned the homologs as CHR8A or CHR8B (see Supplementary Data 4). Next, to determine how deletion of each

allele affected the number of RNA hybridization signals, we carried out RNA-DNA FISH using the TL:8-2.7 probe (Supplementary Data 2) in heterozygous deletion clones. Figure 7d shows the quantification of the RNA FISH signals, and indicates that deletion of TL:8-2.7 from CHR8A resulted in cells with zero hybridization signals in ~95% of cells, and single sites of RNA hybridization signals in ~5% of cells. In contrast, cells containing deletion of TL:8-2.7 from CHR8B contained single sites of hybridization in >95% of cells, and zero sites of hybridization in ~5% of cells. These results indicate that TL:8-2.7 shows AEI in HTD114 cells.

We next assayed replication timing of the two chromosome 8 homologs using a BrdU incorporation assay⁵⁶. Cultures of cells were pulsed with BrdU for 5.5 h and mitotic cells harvested, processed for BrdU incorporation and subjected to FISH using a chromosome 8 centromeric probe plus a Fosmid probe from within the deleted region (Supplementary Data 2). As expected, prior to deletion of the TL:8-2.7 locus, the two chromosome 8 homologs display synchronous replication (see Fig. 7e). In contrast, cells containing heterozygous deletions in CHR8A resulted in significantly more BrdU incorporation in the deleted chromosome than the non-deleted chromosome (Fig. 7e). Therefore, we name this lncRNA gene ASAR8-2.7.

We next wanted to determine if multiple ASARs expressed from the same chromosome homolog all regulate replication timing. For this analysis, we chose three TLs located on chromosome 9. Using RNA-DNA FISH in combination with Sanger sequencing of PCR products containing heterozygous SNPs we found that TL:9-23, TL:9-24, and TL:9-30 RNAs are expressed from the same chromosome 9 homolog (CHR9A) in HTD114 cells (Fig. 7h–k). To fully characterize each chromosome 9 homolog for heterozygosity at SNPs within each TL in HTD114 cells, and to confirm expression of all three TL RNAs occurred in cis, we generated mouse monochromosomal hybrid cells containing either chromosome 9 homolog in different hybrids (see Fig. 7k). This allowed us to fully characterize each heterozygous SNP as being located on either CHR9A or CHR9B (see Fig. 7k for examples). We next designed sgRNAs to unique sequences flanking each genomic region and generated single heterozygous deletions of each of the genomic regions expressing TL:9-23, TL:9-24, or TL:9-30 (Supplementary Data 4). We expressed pairs of sgRNAs in combination with Cas9 and screened clones for deletions using PCR primers that flank the sgRNA binding sites (see Supplementary Data 4). We isolated clones that had heterozygous deletions affecting the expressed allele (CHR9A) for all three TLs (Fig. 7k and Supplementary Data 4).

We next assayed replication timing of the two chromosome 9 homologs using a BrdU incorporation assay⁵⁶. Cultures of cells were pulsed with BrdU for 5.5 h and mitotic cells harvested, processed for BrdU incorporation and subjected to FISH using a chromosome 9 centromeric probe. For this analysis we took advantage of a chromosome 9 centromeric polymorphism to distinguish the two homologs and assigned the larger centromere containing chromosome as CHR9A and the smaller centromere containing chromosome as



CHR9B (see Fig. 7l). As expected, prior to any deletions, the two chromosome 9 homologs display synchronous replication (see Fig. 7m). In contrast, cells containing deletions of TL:9-23, TL:9-24, or TL:9-30 from in CHR9A resulted in significantly more BrdU incorporation in the deleted chromosomes than the non-deleted chromosomes (Fig. 7m). Therefore, we conclude that all three TL control

replication timing of the same chromosome 9 homolog and name these lncRNA genes ASAR9-23, ASAR9-24, and ASAR9-30, respectively.

The chromosome structure instability associated with ASAR disruptions is characterized by numerous chromosomal abnormalities resulting in frequent chromosomal rearrangements⁷⁻⁹. Figure 7n-p shows examples of cells containing a chromosome bridge,

Fig. 7 | Delayed replication and instability. Deletions of TL:8-2.7 (a–g) or TL:9-23, TL:9-24, or TL:9-30 (h–p). **a, b, h** AEI and ASRT in EB3_2 and GM12878 clones. The left Y-axis shows replication, and standard deviation (SD). Outliers in the standard deviation (SD) are highlighted in gray. The right Y-axis shows AEI. The X axis shows chromosome position (megabases). **c** FISH for TL:8-2.7 RNA (red, arrows) and chromosome 8 centromere DNA (green, arrowheads). **d** Quantification of RNA signals for TL:8-2.7 before and after deletion. Percent of nuclei with 0, 1, or 2 signals are shown. **e** Chromosome 8 showing the location of TL:8-2.7 (red dot). **f** Delayed replication following ASAR8-2.7 deletion. BrdU incorporation (green) and DNA FISH (CHR8 cen, purple) plus an ASAR8-2.7 Fosmid (red). **g, m** Quantification of BrdU incorporation following ASAR8-2.7 (g) or ASAR9-23, ASAR9-24, or ASAR9-30 (m) deletion. Box plots indicate mean (solid line), standard deviation (dotted line), 25th, 75th percentile (box), 5th and 95th percentile (whiskers) and individual cells (single points). *P* values were calculated using the Kruskal–Wallis test⁷⁸. **i, j** RNA-

DNA FISH for TL:9-30 RNA (green, arrows) and TL:9-23 RNA (i) or TL:9-24 RNA (j) (red, arrows) and chromosome 9 centromere DNA (purple, arrowheads). **k** Sequencing traces of PCR products from DNA or cDNA (RNA) isolated from HTD114, and DNA from cell hybrids containing either CHR9A or CHR9B. **l** ASRT after deletion of ASAR9-23. Mitotic cell showing BrdU incorporation (green) and DNA FISH (CHR9 centromere; red). **n–p** DNA FISH using chromosome 9 paint (CHR9; red). **m** Deletions of ASAR9-23, ASAR9-24 or ASAR9-30 were from chromosome 9A (*CHR9A). **n** The bracket marks a chromosome bridge, and insets show longer exposures. **o** Micronucleus containing chromosome 9. **p** Mitotic cell containing a chromosome 9 fragment. Arrowheads mark intact chromosome 9 s. Scale bars are 10 μ m (c, f, i–j, l, n–p), and 2.5 μ m in insets in (f) and (l). **q** Percent cells containing rearrangements identified with chromosome paints. Deleted ASARs are indicated, and chromosomes 8 and 9 were scored in parental HTD114 cells. Source data are provided as a Source data file (d, g, m, q).

micronuclei, and rearranged chromosomes affecting chromosome 9 in cells with deletion of ASAR9-23. Given the prevalence of these chromosomal abnormalities it is not surprising that we also detected numerous rearrangements involving chromosome 8 or 9 in cells with deletions of ASAR8-2.7, ASAR9-23, ASAR9-24, or ASAR9-30 (Fig. 7q).

Discussion

Chromosome-associated lncRNAs have become well established as regulators of chromosome scale replication timing, gene expression, and structural stability^{58,59}. The three previously identified ASAR genes display random epigenetically programmed AEI, epigenetically programmed ASRT, contain a high LI content, and express RNAs that remain associated with the chromosome territories where they are transcribed^{4–6,9}. In this report, we identified 68 autosomal TL that share all of these “ASAR characteristics”, and genetic disruption of five ASAR candidates resulted in delayed replication timing of human chromosomes in cis. These results suggest that ASAR genes are abundant throughout the human genome, located in regions of the genome that display epigenetically controlled AEI and ASRT, and function to promote proper replication timing and structural stability of each human chromosome.

In the present study, we used haplotype phased expression and replication timing assays on multiple single-cell derived LCL clones isolated from two unrelated individuals to identify loci that display AEI and ASRT. The clonal and allele-specific analysis of the expression and replication timing profiles for the autosomes revealed an unanticipated genomic behavior at >200 loci with differences in AEI and ASRT that are comparable to the differences observed between the active and inactive X chromosomes. These autosomal loci contain both protein coding and noncoding genes that are expressed from single alleles in some clones, expressed from the opposite allele in other clones, expressed from both alleles in other clones, and are not expressed from either allele in yet other clones. The stochastic nature of the AEI at these loci indicates that each allele had acquired one of two states, expressed or silent, and that each allele chose between these two states independently from the other allele. Because this type of AEI is mitotically stable and is not pre-determined by parent of origin, we refer to the choice of the on-off state as being random.

Asynchronous replication timing between alleles was originally observed for loci containing olfactory receptor⁶⁰ and adaptive immune system³¹ gene arrays, genes that are clearly associated with the exquisite specificity provided by mono-allelic expression. Subsequent studies suggested that all epigenetically programmed mono-allelic genes display asynchronous replication between alleles^{28,31,32,51}. In the present study, we found that each allele can adopt earlier or later replication timing in different clones, which is independent of the other allele. This variability in replication timing results in some clones with dramatic ASRT between alleles, but in other clones the two alleles have synchronous replication, with both alleles adopting either the

earlier or later replication timing. Given this wide variability in replication timing we refer to these loci as having VERT. These observations are also consistent with a previous study that identified “megabase sized” regions of asynchronously replicating domains (ARDs) that contain imprinted genes on human autosomes¹. The relationship between ARDs and VERTs is currently not clear, but several ARDs overlap with regions described here as having VERT. In addition, the stochastic nature of the allelic expression and replication timing present at VERT loci indicates that each allele acquires expression and replication timing via an epigenetic mechanism that is not dependent on parent of origin, not dependent on the expression or replication timing status of the opposite allele, and the choice of which allele is expressed is not dependent on replication timing. Thus, the Early/Late pattern of asynchronous replication detected in this study is not correlated with which allele is expressed, further supporting the conclusion that asynchronous replication is not a consequence of transcription of only one allele^{4,9,28,32}.

The ASRT of random mono-allelic genes has been proposed to be coordinated with other random mono-allelic genes on the same chromosome^{4,9,28,32,61}. The results described here do not support coordinated ASRT either on the same chromosome nor throughout the genome. Our analysis of the RT profiles from the published mouse pre-B cell data⁵² shows evidence of VERT, with numerous examples of clone-specific ASRT (see Supplementary Fig. 2), but the Early/Late RT profiles between the two alleles do not support coordinated ASRT on the same chromosome nor does it support a “mirror image” pattern of ASRT throughout the genome. Instead, we find that each allele can choose between earlier or later replication independently from the other allele, and independently from the other asynchronous loci on the same chromosome.

The variable epigenetic AEI of protein coding genes described here is consistent with an earlier study that also used a clonal analysis and allele-specific expression assay to detect AEI of protein coding genes in human LCLs²¹. This heterogeneity in expression state at autosomal protein coding genes is thought to introduce cellular mosaicism in what would be otherwise similar cell populations, and has also been referred to as ‘clone-specific’ mono-allelic expression⁶². Our results are consistent with this interpretation and indicate that expression of each allele is independent of the other allele resulting in bi-allelic as well as silent clones, implying that the mosaicism that is generated is even more extensive than previously believed. The protein coding genes with random epigenetic AEI described here are consistent with previous observations that indicate that random mono-allelic expression of clustered Pcdh genes, which are homophilic cell adhesion molecules, are critical for the generation of cellular individuality in the nervous system^{63,64}. Taken with the observation that homophilic cell adhesion genes display tissue restricted expression and are located within regions of the genome that show epigenetic AEI, it suggests that the cells of different tissues use the epigenetic program described here to generate cellular heterogeneity

using different homophilic cell adhesion molecules that are distinctive for each cell type.

In this report, we identified 85 TL expressed from the X chromosome, with 66 being expressed exclusively from the active X chromosome and 19 expressed from both the active and inactive X chromosomes. The bi-allelic X chromosome TL reside within regions that are known to escape X inactivation^{42–44}. The hypothesis that the X chromosome contains cis-acting loci important for X chromosome inactivation that are separate from the X inactivation center was initially proposed by Drs. Stan Gartler and Arthur Riggs, which they called the “Way Station” model for X inactivation⁶⁵. Subsequently, Dr. Mary Lyon proposed that LIs represent “Booster Elements” that function during the spreading of Xist RNA in cis along the X chromosome facilitating the process of inactivation^{66,67}. This notion was supported by the observation that the X chromosome contains ~27% and autosomes contain ~13% LI derived sequences⁶⁸. In addition, LIs are present at a lower concentration in regions of the X chromosome that escape inactivation, supporting the hypothesis that LIs serve as signals to propagate inactivation along the X chromosome⁶⁸. Further support for a role of LIs in mono-allelic expression came from the observation that LIs are present at a relatively high local concentration (>18%) near both imprinted and random mono-allelic genes located on autosomes³⁷.

Previous studies have suggested that long nascent transcripts that are detected by COT1 DNA (which is primarily LINE and SINE sequences) play a dynamic structural role that promotes the open architecture of active chromosome territories^{69,70}. The RNA species detected by COT1 DNA in RNA FISH assays are predominantly LI sequences, and these LI containing RNAs remain associated with the chromosome territories where they are transcribed⁷⁰. We also note that COT1 DNA, when used as an RNA FISH probe, can detect ASAR6 RNA that is expressed and localized to an individual chromosome territory⁷¹, indicating that at least some of the RNA FISH signal detected by COT1 DNA represents ASARs. Silencing of COT1 RNA has been used as a convenient assay for the gene silencing function of XIST transgenes when integrated into autosomes^{72,73}. Furthermore, Xist-mediated silencing of COT1 RNA expressed from the future inactive X chromosome precedes protein coding gene silencing during early mouse development⁷⁴. We recently used ectopic integration of transgenes and CRISPR/Cas9-mediated chromosome engineering and found that LI sequences, oriented in the antisense direction, mediate the chromosome-wide effects of ASAR6 and ASAR15¹⁰. In addition, oligonucleotides targeting the antisense strand of the one full-length LI within ASAR6 RNA restored normal replication timing to mouse chromosomes expressing an ASAR6 transgene¹⁰. These results provided the first direct evidence that LI antisense RNA plays a functional role in replication timing of mammalian chromosomes. Taken together, these observations suggest that one role of XIST-mediated X inactivation involves silencing of ASAR counterparts (i.e., LI rich X chromosome TL RNA genes) on the X chromosome, and that regions of the X chromosome that escape inactivation do so by maintaining expression of TL RNAs that promote the maintenance of euchromatin within the escape regions. Under this scenario the LI rich TL do not function as “Booster Elements” on the X chromosome, but as targets of XIST-mediated silencing that precedes inactivation of the protein coding genes on the future inactive X chromosome.

In addition to the autosomal loci that display VERT, we detected four regions on the X chromosome, two on the active X and two on the inactive X, that showed allele-specific epigenetic variability in replication timing (see Fig. 3g–j). All four of these regions contain TL with strong AEI, consistent with haplotype 1 representing the expressed and therefore the active X allele. In addition, one of these regions contains two loci that are known to be important for X inactivation in humans. The XACT lncRNA gene is expressed from the active X chromosome and competes with XIST early in human development for the control of

X chromosome activity^{47,48}. We note here that the XACT gene is located within a large TL with an annotated vlincRNA (vlinc483³⁶) that is expressed from the active X chromosome in all six EB3.2 clones (i.e., haplotype 1; Fig. 3j and Supplementary Data 1). The relationship between the spliced XACT lncRNA product and the larger non-spliced TL is unknown at this time. In addition to XACT, this locus contains DXZ4 and expresses the DANT2 lncRNA⁷⁵ from the active X chromosome in all six EB3.2 clones (Fig. 3j). DXZ4 is a macrosatellite repeat that lies at the boundary of a “superdomain” on the inactive X chromosome, and the inactive X allele extensively binds CTCF⁴⁹. Deletion of DXZ4 leads to the disappearance of “superdomains”, changes in compartmentalization patterns, and changes in the distribution of chromatin marks on the inactive X chromosome⁵⁰. The presence of VERT loci on the active and inactive X chromosomes suggests that the epigenetic program described here is not limited to autosomes but also functions on the X chromosomes in female cells. Whether or not VERT is present on the single X chromosome in male cells is currently not known.

In the immune, olfactory, and central nervous systems, cellular individuality is generated by expression of diverse receptor-type molecules, and involves random mono-allelic expression resulting in specific functional properties. To ensure that individual B or T cells express single Ig or TCR genes, respectively, mono-allelic expression functions to limit the number of expressed genes to one per cell³¹. At the molecular level, the mechanism of stochastic mono-allelic expression can help generate cellular individuality and provides potential functional variation among the individual cells of a complex system^{23,62,63}. The protein coding genes that display epigenetically controlled AEI described here are primarily involved in homophilic cell adhesion, suggesting that B cells may also use mono-allelic expression of cell adhesion molecules in addition to the Ig genes during the adaptive immune response.

Genomes are defined by their sequence. However, the linear arrangement of nucleotides along chromosomes is only their most basic feature. Within this linear arrangement of nucleotides exists three different types of cis-acting loci known to be essential for normal function of every eukaryotic chromosome; origins of replication, centromeres, and telomeres are present on all linear chromosomes functioning to promote proper replication, segregation, and stability of each chromosome. The identification of ASARs present on all human chromosomes and essential for timely replication, condensation, and genetic stability supports a fourth type of essential cis-acting chromosomal locus, which we call “Inactivation/Stability Centers (I/SCs)”. I/SCs likely are present in all mammals and may pre-date the evolution of XIST, which is a critical component of the X inactivation center present in eutherian mammals but absent in metatherian mammals⁷⁶. I/SCs are complex loci that contain protein coding genes that are important for cell-cell adhesion and cellular individuality, contain multiple non-coding ASAR genes that are expressed in different tissues and are responsible for proper replication timing and structural stability of each chromosome.

Methods

Cell culture

GMI2878 cells were obtained from ATCC and EB3.2 cells were from ref. 45. LCLs were grown in RPMI 1640 (Life Technologies) supplemented with 15% fetal bovine serum (Hyclone). Single-cell clones were isolated by plating individual cells in 96-well dishes, and were expanded for >25 population doublings. Primary blood lymphocytes were isolated after venipuncture into a Vacutainer CPT (Becton Dickinson, Franklin Lakes, NJ) per the manufacturer’s recommendations and grown in 5 mL RPMI 1640 (Life Technologies) supplemented with 10% fetal bovine serum (Hyclone) and 1% phytohemagglutinin (Life Technologies). HTD114 cells are a human *APRT* deficient cell line derived from HT1080 cells⁷⁷, and were grown

in DMEM (Gibco) supplemented with 10% fetal bovine serum (Hyclone). All cells were grown in a humidified incubator at 37 °C in a 5% carbon dioxide atmosphere.

RNA-seq

Nuclei were isolated by centrifugation for 0.5 min from GM12878 and EB3_2 clones following lysis in 0.5% NP40, 140 mM NaCl, 10 mM Tris-HCl (pH 7.4), and 1.5 mM MgCl₂. Nuclear RNA was isolated using Trizol reagent using the manufacturer's instructions, followed by DNase treatment to remove possible genomic DNA contamination. Briefly, ribosomal RNAs were removed using the Ribo-Zero kit (Illumina), RNA was fragmented into 250–300 bp fragments, and cDNA libraries were prepared using the Directional RNA Library Prep Kit (NEB). Paired-end sequencing was done on a NovaSeq 6000 at the OHSU MPSSR core facility. Sequences were aligned to the human genome (hg19) using the STAR aligner⁷⁸ with default settings. Duplicate reads and reads with map quality below 30 were removed with SAMtools⁷⁹.

Early/Late Repli-seq

Repli-seq libraries were generated using the E/L Repli-seq protocol⁸⁰. In summary, rapidly growing cells are pulse labeled with BrdU for 2 h, trypsinized, and fixed with cold ethanol. Next, cells are FACS sorted by DNA content into 'early S' and 'late S' fractions until both fractions have a minimum of 120,000 cells. DNA is then fragmented and sequencing libraries are prepared with the NEBNext Ultra DNA Library prep kit, followed by BrdU immunoprecipitation to enrich for BrdU incorporated DNA, and standard short-read sequencing.

Identification of transcribed loci (TLs)

Nuclear-enriched, strand-specific, ribo-minus, total-RNA sequencing libraries were first aligned to the hg19 reference genome using default parameters with STAR⁷⁸. Next SAMtools⁷⁹ was used to remove duplicate and low quality (\leq MAPQ20) reads. TLs were defined using a strategy of sequential merging of strand-specific, contiguous intergenic reads^{34–36}. Utilizing strand-specific reads, stranded reads separated by 1000 bp or less were merged to create contigs, and the contigs merged again while allowing for gaps of 7 kb to allow for regions that are not uniquely mappable due to presence of full-length LINE elements. Vlincs were then classified above a minimum cutoff length of 50 kb of strand specific, contiguous expression. For each cell line, TLs were defined after combining aligned reads from all derivative subclones, and considered expressed in each subclone if a minimum of 20 informative reads (i.e., covering a heterozygous SNP locus) were detected.

Allelic expression imbalance analysis of TLs and coding genes in GM12878 and EB3_2

AEI was defined as the fraction of strand-concordant, allele-specific reads to the total number of informative reads. Thus, perfect allelic balance yields a value of 0.5, and perfect allelic bias towards one haplotype yields a value of 1. For convenience, allelic bias from Haplotype 1 is plotted as positive values, allelic bias from Haplotype 2 is plotted as negative values, and AEI is transformed to percentages in the figures.

$$AEI = \frac{\max(\text{Haplotype 1 Reads}, \text{Haplotype 2 Reads})}{\text{Total Informative Reads}} \times 100 \quad (1)$$

A fully haplotype-resolved reference genotype for GM12878 was obtained from the Illumina Platinum Genome project⁴⁰, and a fully haplotype-resolved reference genotype for EB3_2⁴⁵.

The threshold to call AEI for coding genes and TLs was defined as the union of outliers identified by a parametric and non-parametric statistical outlier identification: (1) Two-sided

Binomial-test p -value ≤ 0.001 and FDR Benjamini–Hochberg adjusted q -value ≤ 0.01 ; and (2) To address the increase of variance with higher expressed genes, linear regression was used to model the relationship between AEI and expression level for all annotated expressed genes and TLs. Outliers were defined as expressed TLs or genes displaying AEI greater than the predicted transcriptome-wide average value plus 2.5 times the standard deviation of AEI for a given expression level.

Allele-specific DNA replication-timing analysis

-10x whole genome sequencing was performed on the Early and Late fractions of each Repli-seq sample and reads aligned using BWA MEM⁸¹. Utilizing fully haplotype-resolved reference genotypes, allele-specific Repli-seq reads were enumerated for each allele of the Early and Late sequencing fractions. After quantile normalization, the Log2(Early/Late) ratio was calculated for non-overlapping genomic windows of 250 kb to create a genome-wide replication timing profile for each haplotype. To identify regions with epigenetic variation between alleles, the standard deviation of replication timing among alleles, SD_{alleles} , was calculated for each genomic window. Windows were considered outliers if displaying an SD_{alleles} value 2.5 standard deviations over the median SD_{alleles} value of all genomic windows measured.

PCR and expression analysis

Genomic DNA and total RNA were isolated from tissue culture cells using TRIZOL Reagent (Invitrogen). cDNA was prepared using the SuperScript™ III First-Strand Synthesis System (Invitrogen). Reverse transcriptase reactions were performed in the presence or absence of reverse transcriptase on 5 µg of total RNA. PCR (genomic and RT-PCR) was performed in a 25–50 µL volume using 50–100 ng of genomic DNA or 1–2 µL of cDNA (50–100 ng of input RNA equivalent), 1x Standard Taq Buffer (New England Biolabs, Inc.), 200 µM each deoxynucleotide triphosphates, 0.2 µM of each primer, and 3 units of Taq DNA Polymerase (New England Biolabs, Inc.) under the following reaction conditions: 95 °C for 2 min, followed by 30–40 cycles of 95 °C for 30 s, 55–62 °C for 45 s, and 72 °C for 1 min, with a final extension time of 10 min at 72 °C. PCR products were separated on 1% agarose gels, stained with ethidium bromide, and photographed under ultraviolet light illumination. Sequencing of PCR products was carried out at the Vollum Institute DNA Sequencing Core facility.

DNA FISH

Mitotic cells were harvested by trypsinization in the absence of colcemid. The cells were treated with 75 mM KCl for 15 min at 37 °C, fixed in 3:1 methanol:acetic acid and dropped on wet slides. The chromosomes were denatured in 70% formamide in 2_{SSC} (1_{SSC} is 150mMNaCl, 15 mM Na-citrate) at 70 °C for 3 min⁸². After RNase (100 µg/ml) treatment for 1 h at 37 °C, slides were washed in 2XSSC and dehydrated in an ethanol series and allowed to air dry. Chromosomal DNA on the slides was denatured at 75 °C for 3 min in 70% formamide/2XSSC, followed by dehydration in an ice-cold ethanol series and allowed to air dry. BAC and Fosmid DNAs were labeled using nick translation (Vysis, Abbott Laboratories) with Spectrum Orange-dUTP, Spectrum Aqua-dUTP, or Spectrum Green-dUTP (Vysis). Final probe concentrations varied from 40–60 ng/µL. Centromeric probe cocktails (Vysis) and/or whole chromosome paint probes (Metasystems) plus BAC or Fosmid DNAs were denatured at 75 °C for 10 min and prehybridized at 37 °C for 10 min. Probes were applied to denatured slides and incubated overnight at 37 °C. Post-hybridization washes consisted of one 3-min wash in 50% formamide/2XSSC at 40 °C followed by one 2-min rinse in PN (0.1M Na₂HPO₄, pH 8.0/2.5% Nonidet NP-40) buffer at RT. Coverslips were mounted with Prolong Gold antifade plus DAPI (Invitrogen) and viewed under UV fluorescence (Olympus).

RNA-DNA FISH

Cells were plated on Poly-L-Lysine coated (Millipore Singa) glass microscope slides at ~50% confluence and incubated for 4 h in complete media in a 37 °C humidified CO₂ incubator. Slides were rinsed 1X with sterile RNase-free PBS. Cell Extraction was carried out using ice-cold solutions as follows: Slides were incubated for 30 s in CSK buffer (100 mM NaCl/300 mM sucrose/3 mM MgCl₂/10 mM PIPES, pH 6.8), 10 min in CSK buffer/0.1% Triton X-100, followed by 30 s in CSK buffer. Cells were then fixed in 4% paraformaldehyde in PBS for 10 min and stored in 70% EtOH at -20 °C until use. Just prior to RNA FISH, slides were dehydrated through an EtOH series and allowed to air dry. Denatured probes were prehybridized at 37 °C for 10 min, applied to non-denatured slides and hybridized at 3 °C for 14–16 h. Post-hybridization washes consisted of one 3-min wash in 50% formamide/2XSSC at 40 °C followed by one 2-min rinse in 2XSSC/0.1% TX-100 for 1 min at RT. Slides were then fixed in 4% paraformaldehyde in PBS for 5 min at RT, and briefly rinsed in 2XSSC/0.1% TX-100 at RT. Coverslips were mounted with Prolong Gold antifade plus DAPI (Invitrogen) and slides were viewed under UV fluorescence (Olympus). Z-stack images were generated using a Cytovision workstation. After capturing RNA FISH signals, the coverslips were removed, the slides were dehydrated in an ethanol series, and then processed for DNA FISH, beginning with the RNase treatment step, as described above. We included an RNA FISH probe to an intron of *KCNQ5* as positive control⁵ in HTD114 cells.

Chromosome replication timing assay

The BrdU replication timing assay was performed as on exponentially dividing cultures and asynchronously growing cells. Mitotic chromosome spreads were prepared and DNA FISH was performed as described above. The incorporated BrdU was detected using a FITC-labeled anti-BrdU antibody (Roche). Coverslips were mounted with Prolong Gold antifade plus DAPI (Invitrogen), and viewed under UV fluorescence. All images were captured with an Olympus BX Fluorescent Microscope using a 100X objective, automatic filter-wheel and Cytovision workstation. Individual chromosomes were identified with either chromosome-specific paints, centromeric probes, BACs or by inverted DAPI staining. Utilizing the Cytovision workstation, each chromosome was isolated from the metaphase spread and a line drawn along the middle of the entire length of the chromosome. The Cytovision software was used to calculate the pixel area and intensity along each chromosome for each fluorochrome occupied by the DAPI and BrdU (FITC) signals. The total amount of fluorescent signal in each chromosome was calculated by multiplying the average pixel intensity by the area occupied by those pixels. The BrdU incorporation into human chromosomes containing CRISPR/Cas9 modifications was calculated by dividing the total incorporation into the chromosome with the deleted chromosome divided by the BrdU incorporation into the non-deleted chromosome within the same cell. Boxplots were generated from data collected from 8 to 12 cells per clone or treatment group. Differences in measurements were tested across categorical groupings by using the Kruskal–Wallis test⁸³ and listed as *P*-values for the corresponding plots.

CRISPR/Cas9 engineering

Using Lipofectamine 2000, according to the manufacturer's recommendations, we co-transfected HTD114 cells with plasmids encoding GFP, sgRNAs, and Cas9 endonuclease (Origene). Each plasmid-encoded sgRNA was designed to bind at the indicated locations (Supplementary Data 4). 48 h after transfection, cells were plated at clonal density and allowed to expand for 2–3 weeks. The presence of deletions was identified by PCR using the primers described in Supplementary Data 4. The single-cell colonies that grew were analyzed for heterozygous deletions by PCR. We used retention of

heterozygous SNPs (see Supplementary Data 4) to identify the disrupted allele, and homozygosity at these SNPs confirmed that cell clones were homogenous.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Repli-seq and RNA-seq sequencing data generated in this study have been deposited in the European Nucleotide Archive database under accession code [PRJEB52794](https://doi.org/10.5281/zenodo.7094766), and are available without restriction. Source data are provided with this paper.

Code availability

Computer code for RNA-seq and Repli-seq analysis is publicly available at (<https://doi.org/10.5281/zenodo.7094766>).

References

- Mukhopadhyay, R. et al. Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. *PLoS Genet.* **10**, e1004319 (2014).
- Dileep, V. & Gilbert, D. M. Single-cell replication profiling to measure stochastic variation in mammalian replication timing. *Nat. Commun.* **9**, 427 (2018).
- Rivera-Mulia, J. C. et al. Allele-specific control of replication timing and genome organization during development. *Genome Res* **28**, 800–811 (2018).
- Donley, N., Smith, L. & Thayer, M. J. ASAR15, a cis-acting locus that controls chromosome-wide replication timing and stability of human chromosome 15. *PLoS Genet.* **11**, e1004923 (2015).
- Heskett, M. B., Smith, L. G., Spellman, P. & Thayer, M. J. Reciprocal monoallelic expression of ASAR lncRNA genes controls replication timing of human chromosome 6. *RNA* **26**, 724–738 (2020).
- Stoffregen, E. P., Donley, N., Stauffer, D., Smith, L. & Thayer, M. J. An autosomal locus that controls chromosome-wide replication timing and mono-allelic expression. *Hum. Mol. Genet.* **20**, 2366–2378 (2011).
- Breger, K. S., Smith, L. & Thayer, M. J. Engineering translocations with delayed replication: evidence for cis control of chromosome replication timing. *Hum. Mol. Genet.* **14**, 2813–2827 (2005).
- Chang, B. H., Smith, L., Huang, J. & Thayer, M. Chromosomes with delayed replication timing lead to checkpoint activation, delayed recruitment of Aurora B and chromosome instability. *Oncogene* **26**, 1852–1861 (2007).
- Donley, N., Stoffregen, E. P., Smith, L., Montagna, C. & Thayer, M. J. Asynchronous replication, mono-allelic expression, and long range cis-effects of ASAR6. *PLoS Genet.* **9**, e1003423 (2013).
- Platt, E. J., Smith, L. & Thayer, M. J. L1 retrotransposon antisense RNA within ASAR lncRNAs controls chromosome-wide replication timing. *J. Cell Biol.* **217**, 541–553 (2018).
- Bryois, J. et al. Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* **10**, e1004461 (2014).
- Petretto, E. et al. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet.* **2**, e172 (2006).
- Alexander, M. K. et al. Differences between homologous alleles of olfactory receptor genes require the Polycomb Group protein Eed. *J. Cell Biol.* **179**, 269–276 (2007).
- Gendrel, A. V. et al. Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* **28**, 366–380 (2014).
- Li, S. M. et al. Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS ONE* **7**, e31751 (2012).

16. Lin, M. et al. Allele-biased expression in differentiating human neurons: implications for neuropsychiatric disorders. *PLoS ONE* **7**, e44017 (2012).
17. Gendrel, A. V., Marion-Poll, L., Katoh, K. & Heard, E. Random monoallelic expression of genes on autosomes: parallels with X-chromosome inactivation. *Semin. Cell Dev. Biol.* **56**, 100–110 (2016).
18. Bartolomei, M. S. Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev.* **23**, 2124–2133 (2009).
19. Tucci, V., Isles, A. R., Kelsey, G. & Ferguson-Smith, A. C. Genomic imprinting and physiological processes in mammals. *Cell* **176**, 952–965 (2019).
20. Chess, A. Mechanisms and consequences of widespread random monoallelic expression. *Nat. Rev. Genet.* **13**, 421–428 (2012).
21. Gimelbrant, A., Hutchinson, J. N., Thompson, B. R. & Chess, A. Widespread monoallelic expression on human autosomes. *Science* **318**, 1136–1140 (2007).
22. Reinius, B. & Sandberg, R. Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat. Rev. Genet.* **16**, 653–664 (2015).
23. Savova, V. et al. Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat. Genet.* **48**, 231–237 (2016).
24. Heskett, M. B., Spellman, P. T. & Thayer, M. J. Differential allelic expression among long non-coding RNAs. *Noncoding RNA* **7**, 66 (2021).
25. Ding, Q. et al. The genetic architecture of DNA replication timing in human pluripotent stem cells. *Nat. Commun.* **12**, 6746 (2021).
26. Koren, A. et al. Genetic variation in human DNA replication timing. *Cell* **159**, 1015–1026 (2014).
27. Zhao, P. A., Sasaki, T. & Gilbert, D. M. High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biol.* **21**, 76 (2020).
28. Ensminger, A. W. & Chess, A. Coordinated replication timing of monoallelically expressed genes along human autosomes. *Hum. Mol. Genet.* **13**, 651–658 (2004).
29. Goldmit, M. & Bergman, Y. Monoallelic gene expression: a repertoire of recurrent themes. *Immunol. Rev.* **200**, 197–214 (2004).
30. Gribnau, J., Hochedlinger, K., Hata, K., Li, E. & Jaenisch, R. Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. *Genes Dev.* **17**, 759–773 (2003).
31. Mostoslavsky, R. et al. Asynchronous replication and allelic exclusion in the immune system. *Nature* **414**, 221–225 (2001).
32. Singh, N. et al. Coordination of the random asynchronous replication of autosomal loci. *Nat. Genet.* **33**, 339–341 (2003).
33. St Laurent, G., Savva, Y. A. & Kapranov, P. Dark matter RNA: an intelligent scaffold for the dynamic regulation of the nuclear information landscape. *Front. Genet.* **3**, 57 (2012).
34. Caron, M. et al. Very long intergenic non-coding RNA transcripts and expression profiles are associated to specific childhood acute lymphoblastic leukemia subtypes. *PLoS ONE* **13**, e0207250 (2018).
35. St Laurent, G. et al. VlnCRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol.* **14**, R73 (2013).
36. St Laurent, G. et al. Functional annotation of the vlnCRNA class of non-coding RNAs using systems biology approach. *Nucleic Acids Res.* **44**, 3233–3252 (2016).
37. Allen, E. et al. High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc. Natl Acad. Sci. USA* **100**, 9940–9945 (2003).
38. Bregar, K. S., Smith, L., Turker, M. S. & Thayer, M. J. Ionizing radiation induces frequent translocations with delayed replication and condensation. *Cancer Res.* **64**, 8231–8238 (2004).
39. Moore, J. E. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
40. Eberle, M. A. et al. A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. *Genome Res.* **27**, 157–164 (2017).
41. Shvetsova, E. et al. Skewed X-inactivation is common in the general female population. *Eur. J. Hum. Genet.* **27**, 455–465 (2019).
42. Tukiainen, T. et al. Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
43. Posytnick, B. J. & Brown, C. J. Escape from X-chromosome inactivation: an evolutionary perspective. *Front. Cell Dev. Biol.* **7**, 241 (2019).
44. Navarro-Cobos, M. J., Balaton, B. P. & Brown, C. J. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. *Am. J. Med. Genet. C. Semin. Med. Genet.* **184**, 226–238 (2020).
45. Lajugie, J. et al. Complete genome phasing of family quartet by combination of genetic, physical and population-based phasing analysis. *PLoS ONE* **8**, e64571 (2013).
46. Wardemann, H. et al. Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).
47. Vallot, C. et al. XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239–241 (2013).
48. Vallot, C. et al. XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell* **20**, 102–111 (2017).
49. Chadwick, B. P. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* **18**, 1259–1269 (2008).
50. Darrow, E. M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci. USA* **113**, E4504–E4512 (2016).
51. Schlesinger, S., Selig, S., Bergman, Y. & Cedar, H. Allelic inactivation of rDNA loci. *Genes Dev.* **23**, 2437–2447 (2009).
52. Blumenfeld, B. et al. Chromosomal coordination and differential structure of asynchronous replicating regions. *Nat. Commun.* **12**, 1035 (2021).
53. Bergman, Y., Simon, I. & Cedar, H. Asynchronous replication timing: a mechanism for monoallelic choice during development. *Front. Cell Dev. Biol.* **9**, 737681 (2021).
54. Koren, A. & McCarroll, S. A. Random replication of the inactive X chromosome. *Genome Res.* **24**, 64–69 (2014).
55. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
56. Smith, L. & Thayer, M. Chromosome replicating timing combined with fluorescent in situ hybridization. *J. Vis. Exp.* **10**, e4400 (2012).
57. zur Hausen, H. Chromosomal changes of similar nature in seven established cell lines derived from the peripheral blood of patients with leukemia. *J. Natl Cancer Inst.* **38**, 683–696 (1967).
58. Thayer, M. J. Mammalian chromosomes contain cis-acting elements that control replication timing, mitotic condensation, and stability of entire chromosomes. *Bioessays* **34**, 760–770 (2012).
59. Galupa, R. & Heard, E. X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. *Annu. Rev. Genet.* **52**, 535–566 (2018).
60. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
61. Masika, H. et al. Programming asynchronous replication in stem cells. *Nat. Struct. Mol. Biol.* **24**, 1132–1138 (2017).
62. Savova, V., Vigneau, S. & Gimelbrant, A. A. Autosomal monoallelic expression: genetics of epigenetic diversity? *Curr. Opin. Genet. Dev.* **23**, 642–648 (2013).

63. Yagi, T. Genetic basis of neuronal individuality in the mammalian brain. *J. Neurogenet.* **27**, 97–105 (2013).
64. Mountoufaris, G. et al. Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. *Science* **356**, 411–414 (2017).
65. Gartler, S. M. & Riggs, A. D. Mammalian X-chromosome inactivation. *Annu. Rev. Genet.* **17**, 155–190 (1983).
66. Lyon, M. F. X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet.* **80**, 133–137 (1998).
67. Lyon, M. F. The Lyon and the LINE hypothesis. *Semin Cell Dev. Biol.* **14**, 313–318 (2003).
68. Bailey, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl Acad. Sci. USA* **97**, 6634–6639 (2000).
69. Creamer, K. M., Kolpa, H. J. & Lawrence, J. B. Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction. *Mol. Cell* **81**, 3509–3525 (2021).
70. Hall, L. L. et al. Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907–919 (2014).
71. Thayer, M. J., Heskett, M. B., Smith, L. G., Spellman, P. T. & Yates, P. A. ASAR RNAs control replication through interactions with dozens of RNA binding proteins. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.06.04.494840> (2022).
72. Hall, L. L. et al. An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc. Natl Acad. Sci. USA* **99**, 8677–8682 (2002).
73. Minks, J. & Brown, C. J. Getting to the center of X-chromosome inactivation: the role of transgenes. *Biochem. Cell Biol.* **87**, 759–766 (2009).
74. Namekawa, S. H., Payer, B., Huynh, K. D., Jaenisch, R. & Lee, J. T. Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. *Mol. Cell Biol.* **30**, 3187–3205 (2010).
75. Figueroa, D. M., Darrow, E. M. & Chadwick, B. P. Two novel DXZ4-associated long noncoding RNAs show developmental changes in expression coincident with heterochromatin formation at the human (*Homo sapiens*) macrosatellite repeat. *Chromosome Res.* **23**, 733–752 (2015).
76. Elisaphenko, E. A. et al. A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS ONE* **3**, e2521 (2008).
77. Zhu, Y., Bye, S., Stambrook, P. J. & Tischfield, J. A. Single-base deletion induced by benzo(a)pyrene diol epoxide at the adenine phosphoribosyltransferase locus in human fibrosarcoma cell lines. *Mutat. Res.* **321**, 73–79 (1994).
78. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
79. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
80. Marchal, C. et al. Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat. Protoc.* **13**, 819–839 (2018).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
82. Smith, L., Plug, A. & Thayer, M. Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc. Natl Acad. Sci. USA* **98**, 13300–13305 (2001).
83. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27 (1964).

Acknowledgements

M.J.T. was supported by NIH NIGMS (R01GM114162 and R01GM130703). M.B.H. was supported by NIH NCI 4K00CA245677-03. D.M.G. was supported by NIH R01GM083337.

Author contributions

M.B.H., M.J.T., P.S., E.E.B., and D.M.G. conceived of the experimental design for the RNA expression and replication timing assays. L.S.G. carried out the RNA-DNA FISH and chromosome replication timing assays. M.B.H. and M.J.T. carried out the RNA-seq assays, including data analysis. M.B.H. and A.E.V. carried out the Repli-seq assays, including data analysis. C.B. provided technical assistance during Repli-seq library preparation. M.J.T. and P.A.Y. generated the CRISPR/Cas9 deletions. M.B.H. and M.J.T. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-34099-7>.

Correspondence and requests for materials should be addressed to Mathew J. Thayer.

Peer review information *Nature Communications* thanks Anne Gendrel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022