Autosomal Allelic Inactivation: Variable Replication and Dosage Sensitivity

Michael B. Heskett[1], Athanasios E. Vouzas[2], Brian Johnstone[3], Krister P. Freese[4], Phillip A. Yates[5], Philip F. Copenhaver[6], Paul T. Spellman[1], David M. Gilbert[2], Mathew J. Thayer[5,*].

[1]Division of Hematology Oncology
Departments of Medicine and Human Genetics
David Geffen School of Medicine
University of California, Los Angeles

[2]San Diego Biomedical Research Institute
3525 John Hopkins Court
San Diego, CA 92121

[3]Shriners Hospitals for Children—Portland
3101 SW Sam Jackson Park Rd,
Portland, Oregon, 97239

[4] Department of Orthopedics and Rehabilitation
[5]Department of Chemical Physiology and Biochemistry
[6]Department of Cell, Developmental and Cancer Biology
Oregon Health & Science University
3181 S. W. Sam Jackson Park Road
Portland, Oregon 97239

***Corresponding author.**

ORCID: 0000-0001-6483-1661

TEL: (503) 494-2447

FAX: (503) 494-7368

Email:thayerm@ohsu.edu

Abstract

Autosomal monoallelic gene expression and asynchronous replication between alleles are well-established features of imprinted genes and genes regulated by allelic exclusion. Inactivation/Stability Centers (I/SCs) are recently described autosomal loci that exhibit epigenetic regulation of allelic expression and replication timing, with differences that can be comparable to those observed between the active and inactive X chromosomes [1]. Here we characterize hundreds of autosomal loci with allele-specific epigenetic regulation of replication timing and gene expression, defining them as I/SCs. I/SCs are approximately 1 megabase in size and can contain both protein-coding and noncoding genes. In different single cell derived clones, these genes may be expressed from a single allele, the opposite allele, both alleles, or not expressed at all. This stochastic, yet mitotically stable, pattern indicates that the choice of which allele is expressed is independent of parent of origin and independent of the expression status of the other allele. Similarly, alleles within I/SCs show varying replication timing, either earlier or later, that is also independent of the other allele. Additionally, we find corresponding I/SCs in the mouse genome that display conserved synteny with human I/SCs. This allele-restricted regulation creates extensive cellular mosaicism through a stable epigenetic mechanism. This mosaicism impacts numerous dosage-sensitive genes associated with human diseases such as Parkinson, epilepsy, deafness, and impaired intellectual development.

60 **Introduction:**

61 The relationship between monoallelic expression and asynchronous replication between
62 alleles of autosomal genes, exemplified by genes subject to genomic imprinting or allelic
63 exclusion, has been the focus of extensive research for over three decades [2-4]. In recent
64 work, we employed haplotype-phased RNA expression and replication timing assays on
65 multiple single-cell-derived Lymphoblastoid Cell Line (LCL) clones from two unrelated
66 individuals to identify loci exhibiting allele-restricted expression and replication timing [1].
67 This clonal, allele-specific analysis across autosomes revealed an unexpected genomic
68 feature at approximately 200 loci, exhibiting allele-restricted gene expression (including
69 random monoallelic expression); and variable replication timing (including asynchronous
70 replication between alleles). The magnitude of allelic differences in expression and
71 replication timing observed in some of the clones was comparable to the large differences
72 detected between the active and inactive X chromosomes. We have named these
73 autosomal loci Inactivation/Stability Centers (I/SCs) [1,5-7].

74 I/SCs can encompass both coding and noncoding genes that may be expressed
75 from a single allele in some clones, expressed from the opposite allele in other clones,
76 biallelically expressed in other clones, or silent on both alleles in yet other clones. This
77 pattern of allele-restricted expression indicates that each allele independently adopts
78 either an expressed or silent state. Importantly, because these expression states are
79 mitotically stable, allele-autonomous, and independent of parental origin, we refer to the
80 choice of the expressed allele as stochastic [1]. To describe this phenomenon, we and
81 others have used the term Allelic Expression Imbalance (AEI) to refer to the stochastic
82 nature of autosomal random monoallelic expression [1,8-11]. Additionally, the replication
83 timing profiles on autosomes are highly variable and clone-specific, exhibiting both

84 synchronous and asynchronous patterns in different clones. We refer to this as Variable
85 Epigenetic Replication Timing (VERT) [1].

86 Random monoallelic expression in the olfactory, immune, and central nervous
87 systems is well established and plays a critical role in generating cell-specific identities [12-
88 14]. However, increasing evidence indicates that autosomal genes outside these
89 systems also exhibit allele-specific expression, suggesting that AEI plays a broader role
90 across diverse cell types [4,13]. The functional diversity of these genes implies that AEI
91 contributes to a broad array of cellular processes and is not limited to specialized
92 systems.

93 In the present study, we aimed to determine whether I/SCs could be detected in
94 the genomes of clonal populations of human primary cells derived from a tissue distinct
95 from those previously examined. To this end, we employed articular cartilage progenitor
96 (ACP) cells as a second model system [15]. ACP cells are non-transformed yet are
97 particularly well-suited for clonal analysis due to several advantages: 1) they can be
98 readily isolated from amputated tissue of patients with polydactyly, 2) they exhibit high
99 proliferative capacity, and 3) they can be efficiently expanded from single cells [16]. We
100 identify an additional 163 VERT regions in the human genome, increasing the total
101 number of potential I/SCs to 363, which collectively cover approximately 12% of the
102 human genome. In addition, we found that many previously reported "random
103 monoallelic" genes, including those encoding olfactory and vomeronasal
104 receptors[3,17], antigen receptors[18], and homophilic cell adhesion proteins[12], are located
105 within VERT regions. Our findings confirm that these well-established "random
106 monoallelic" genes reside within genomic loci that exhibit the epigenetic hallmarks,
107 namely allele-restricted expression and variable replication timing, that define I/SCs.

The relationship between genotype and phenotype is central to understanding human genetic disease. While most loss-of-function mutations are recessive, indicating that a single normal allele is sufficient for typical function, dominantly inherited disorders often result from gain-of-function mutations. However, a significant fraction of dominant diseases arises through haploinsufficiency (i.e. heterozygous for a loss of function mutation), where a single functional allele fails to maintain normal physiology. Although haploinsufficiency may affect up to 40% of human genes [19-21], the underlying molecular mechanisms remain poorly understood. Current models suggest that insufficient gene dosage may arise from a lack of redundancy, critical expression thresholds, stoichiometric imbalances, or toxicity associated with a lack of overexpression buffering [22-25]. In contrast, insights from X-linked disease inheritance illustrate how cellular mosaicism and cellular selection can shape phenotypic outcomes: in XX individuals, differences in the survival or expansion of cells expressing either the wild-type or mutant alleles can influence whether a disease manifests as recessive or dominant [21,26,27].

The magnitude of AEI is often quantified using allelic expression ratios, which represent the proportion of transcripts derived from each allele. Importantly, recent findings have demonstrated that even modest allelic expression biases can have functional consequences, particularly in disease contexts. One recent study directly linked small but stable AEI to variable clinical outcomes among individuals carrying the same pathogenic mutation, underscoring AEI as a potential modulator of disease expressivity [28]. These observations highlight the importance of considering AEI as a disease modifier that may shape phenotypic diversity and influence the penetrance of genetic disorders.

132 Here, we demonstrate that numerous genes implicated in autosomal human
133 genetic diseases, such as Parkinson (DNAJC6, LRRK2 and SNCA), epilepsy (SCN1A,
134 GABR1, and SAMD12), deafness (ATP11A, EPS8, and MYO6), and neurodevelopmental
135 disorders (KCNA1, RETREG1, ROBO1, and TIAM1), are located within VERT regions
136 and display AEI. Furthermore, mapping I/SCs across all autosomes reveals 87 known
137 dominantly inherited single-gene disorders that map within VERT regions, where loss of
138 function mutations in most of these genes result in haploinsufficiency [29]. We propose
139 that epigenetically regulated stochastic AEI can contribute to pathogenic phenotypes in
140 heterozygous individuals by either reducing the number of functional cells below a critical
141 disease threshold, akin to X linked dominant disorders, or by disrupting the normal
142 mosaicism of gene expression within disease-relevant tissues.

143 **Results**

144 Here we employed articular cartilage progenitor (ACP) cells[15], as a distinct cell type from
145 those from our previous studies, as a second model system to identify I/SCs. For this
146 analysis, we isolated ACP cells from amputated digits of patients with polydactyly. In
147 addition, we sequenced genomic DNA from the parents of the patients with polydactyly
148 to generate phased genomes of the ACP cells. We analyzed two sets of ACP clones
149 isolated from two unrelated individuals, one female (ACP7) and one male (ACP6), for
150 haplotype-phased Repliseq and RNAseq (Fig. 1b). Analysis of the female ACP7 clones
151 allowed us to establish the statistical parameters for the assignment of VERT and AEI
152 present in the ACP clones by directly comparing the magnitude of autosomal allele-
153 specific replication timing and RNA expression to that of the active and inactive X
154 chromosomes present in each clone (Fig. 1c; and see Supplementary Tables 1, 2 and 3).
155 Taken together the combined VERT and AEI analysis allowed us to characterize

156 additional I/SCs in the human genome. An illustration of a hypothetical I/SC, with the
157 region of VERT shaded in gray and the expression of coding and noncoding genes
158 indicated as stippled or smooth rectangles, respectively, is shown in Figure 1d.

159 To identify genomic regions with epigenetic variation between alleles, we
160 calculated the standard deviation (SD) of replication timing across alleles for each 250 kb
161 genomic window. Windows with an SD value ≥ 2.5 standard deviations above the
162 genome-wide median were classified as outliers. This analysis identified 163 additional
163 VERT regions (Supplementary Table 1). Figure 1e illustrates this analysis for
164 both ACP (ACP6 and ACP7) and LCL (EB3-2 and GM12878) clones on Chromosome 1,
165 where windows are indicated as gray circles and outliers indicted in different colors.

166 To detect coding genes with AEI in ACP clones, we defined AEI as differential
167 expression between alleles using recently established criteria [1,28]. Genes were
168 considered AEI-positive if they exhibited ≥80% allelic imbalance (AEI ≥ 0.80 or ≤ 0.20)
169 with FDR ≤ 0.05 in at least one clone. Additional analyses using more stringent thresholds
170 (AEI ≥ 0.90 or ≥ 0.95) are provided in Supplementary Table 2. To exclude imprinted
171 genes and effects due to genetic polymorphisms (i.e. expression Quantitative Trait Loci,
172 eQTL[30]) we required that each AEI gene also be biallelically expressed (AEI < 0.70) in at
173 least one clone from the same individual with ≥ 20 informative reads. Using these criteria,
174 we identified 256 (2.4% of expressed genes) autosomal coding genes exhibiting AEI
175 across the two ACP clone sets (Supplementary Table 2).

176 In addition, we also sought to identify AEI of extremely long noncoding RNAs (i.e.
177 ASAR lincRNAs [1,5,7,31]) expressed in the ACP clones. For this analysis we refer to regions
178 of the genome with >50 kb of contiguous transcription of noncoding DNA as Transcribed

179 Loci (TL) [1]. Using the same criteria used for the coding genes described above, we
180 identified 23 TLs exhibiting AEI across the two ACP clone sets (Supplementary Table 3).

181 **Allelic Expression Imbalance and Variable Epigenetic Replication Timing define**
182 **Inactivation/Stability Centers.**

183 By integrating the location of genes exhibiting AEI with genomic regions displaying VERT,
184 we identified overlapping loci that define I/SCs. Figures 1f and 1g illustrate a
185 representative I/SC located on chromosome 1 at approximately 240 Mb (marked by an
186 arrow in Fig. 1e). Notably, VERT was detected at this locus across both sets of ACP
187 clones (Fig. 1f and 1g). Figure 1h presents the corresponding UCSC Genome Browser
188 view of this locus, with genes showing AEI marked by asterisks (see Supplemental Table
189 2). Additional examples of I/SCs located on chromosomes 18, 6 and 9 are shown in Figure
190 2.

191 One of the primary objectives of this study was to determine whether different
192 tissues utilize distinct sets of I/SCs, as might be expected given that different cell types
193 express unique gene repertoires that could be subject to AEI. To explore this, we
194 performed hierarchical clustering analysis of the VERT regions identified in the two sets
195 of ACP clones and the two sets of LCL clones. Figure 1i shows this analysis and reveals
196 tissue-specific patterns with a relatively small degree of overlap in VERT regions across
197 the two tissue types: ACP6 clones share more VERT regions with ACP7 clones than with
198 either LCL clone sets, and the reverse is true for the LCL clones. These findings indicate
199 that while some I/SCs are shared across tissues, there is a large component of tissue-
200 specific epigenetic regulation of allelic replication timing.

201 At the molecular level, stochastic monoallelic expression contributes to cellular
202 individuality and introduces functional variability among individual cells within complex

203 biological systems [12-14]. This mechanism is particularly well characterized in the olfactory,
204 immune, and central nervous systems, where the expression of diverse receptor-type
205 molecules through random monoallelic expression underlies cell-specific functional
206 identities. To further investigate the functional characteristics of protein-coding
207 genes located within I/SCs, we annotated the genes residing within the VERT
208 regions identified across all four clonal datasets (ACP clones: ACP6 and ACP7; and LCL
209 clones: EB3-2, and GM12878; see Supplementary Table 4). To assess whether these
210 genes are functionally related, we performed a Gene Ontology (GO) enrichment analysis.
211 This analysis revealed a significant enrichment for genes involved in plasma membrane
212 localization, cell-cell adhesion and neurodevelopmental processes, the most significant
213 hits for each set of clones is presented in Figure 1k (see full details in Supplementary
214 Table 5). This analysis suggests that I/SC-associated epigenetic regulation may play a
215 broader role in the establishment of cellular identity and tissue organization particularly in
216 the nervous system.

217 **Gene clusters with AEI map to VERT regions.**

218 In the olfactory, immune, and central nervous systems, random monoallelic expression of
219 gene family members, often organized in genomic clusters, plays a key role in
220 establishing cell-specific identities [12-14]. Consistent with these observations, we found
221 that VERT regions frequently coincide with gene clusters in the human genome (Table
222 1). For example, Figures 3a and 3b highlight a cluster of 11 olfactory receptor (OR)
223 genes located on human chromosome 19 at approximately 15 Mb. This region also
224 harbors a cluster of Cytochrome P450 (CYP4F family, n = 6) genes. The interspersed
225 genomic arrangement of the CYP4F genes among the OR genes, indicates that both
226 gene families reside within the same I/SC and therefore both may be subject to AEI.

227 A second example of a gene cluster is shown in Figures 3c and 3d, where
228 the major histocompatibility complex (MHC), comprising 16 HLA genes (which are known
229 to be subject to AEI [28,32]), lies within a VERT region. Notably, this locus also contains
230 a cluster of Lymphocyte Antigen 6 genes (LY6G family, n = 6), which are interspersed
231 among the HLA genes. This genome organization indicates that the LY6G genes, like
232 their neighboring HLA genes, are within the same I/SC and suggests that they too may
233 be subject to stochastic allelic expression.
234
235 **Mapping mouse VERTs identifies known AEI gene clusters and syntenic I/SCs.**

236 Much of our current understanding of stochastic autosomal allelic expression is derived
237 from studies conducted in the mouse [4,9-11,17,18,33-40]. To assess whether VERT
238 regions could be identified at expected loci in the mouse genome, specifically at sites of
239 known AEI, we analyzed previously published Repli-seq data from mouse pre-B cell
240 clones reported by Blumenfeld et al [41]. This analysis revealed 117 loci in the mouse
241 genome that display VERT (Supplemental Table 6).

242 Consistent with our findings in the human genome, the VERT regions detected in
243 the mouse genome frequently overlap with gene clusters, including 5 olfactory receptor
244 gene clusters located on four different autosomes (Table 1). Figure 4 presents three
245 examples: two olfactory receptor (Or) gene clusters on mouse chromosome 2 (Figs. 4a–
246 4c) and the protocadherin (Pcdh) gene cluster, which is known to generate cellular
247 individuality in the central nervous system through stochastic expression of different Pcdh
248 family members [12,42], located on mouse chromosome 18 (Figs. 4d–4e). In addition to the
249 gene clusters with VERT, we also detected immunoglobulin and T cell receptor gene loci
250 with VERT. As detailed in Supplemental Figure 1, we detected VERT at the T cell receptor

251 alpha (Tra) locus on mouse chromosome 14 at approximately 54 Mb. One striking feature
252 of this locus is that it also contains an olfactory receptor cluster, containing 37 Or genes,
253 and a vomeronasal receptor gene (Vmn2r88). All of these multigene families are
254 documented sites of AEI [3,12,17,42,43], highlighting the genomic organization and
255 conservation of I/SC-associated regulation at known AEI genes in the mouse. Gene
256 Ontology analysis of the mouse genes located within these regions showed significant
257 enrichment for biological processes related to the plasma membrane and cell-cell
258 adhesion (Fig. 1k), which is consistent with the human VERT gene dataset and reinforces
259 the conservation of this regulatory mechanism across species.

260 Previous studies analyzed a relatively small number of mouse AEI genes to model
261 their human counterparts with respect to autosomal-dominant disorders [38,44]. To further
262 investigate the conservation of I/SC regulation between the human and mouse genomes,
263 we identified VERT regions that map to syntenic loci across the two species. This analysis
264 revealed 21 loci where VERT is detected in both genomes, as detailed in Table 2. Figure
265 5 presents three representative examples of syntenic VERT regions: two loci on the short
266 arm of human chromosome 5 with corresponding regions on mouse chromosome 15
267 (Figs. 5a–5g), and a third locus on human chromosome 16 with its syntenic region on
268 mouse chromosome 16 (Figs. 5h–5k). Additional examples of syntenic loci showing AEI
269 and/or VERT, including mouse App (see Gimelbrant et al. [45] for AEI of human APP),
270 Mocos and Fhod3 (See Fig. 2A for VERT and AEI of human MOCOS and FHOD3), and
271 a cluster of 5 sodium channel genes, Scn1a, Scn2a, Scn3a, Scn7a and Scn9a (see Fig.
272 6a-c below for VERT and AEI of the syntenic human SCN gene cluster) are shown in
273 Supplemental Figure 2.

274 We point out that many of the mouse genes that are located within VERT regions
275 are associated with various human genetic diseases (Supplemental Table 7). One of the
276 loci highlighted in Figure 5a-5d labeled "hi", exhibits both VERT and AEI (Fig. 5c and [1]),
277 affecting the protein-coding gene SEMA5A, a gene implicated in autism and
278 neurodevelopmental disorders [46-48]. Notably, this locus also contains a syntenic pair of
279 taste receptor genes: TAS2R1 in humans and Tas2r119 in mice. These genes are part
280 of the taste receptor family, which are known to harbor chromatin marks characteristic of
281 AEI-regulated loci [49]. Consistent with this observation, we also detected VERT at a taste
282 receptor gene cluster (Tas2r) on mouse chromosome 6 (Table 1).
283 A second example of synteny is shown in Figure 5f-5g, where human RETREG1
284 and mouse Retreg1 are located within VERT regions. Human RETREG1 is a known AEI
285 gene [45], and mutations in RETREG1 cause sensory and autonomic neuropathy [29]. A third
286 example is shown in Figure 5h-5k, where both human GRIN2A and mouse Grin2a are
287 located within VERT regions, and loss of function mutations in GRIN2A are known to
288 cause autosomal dominant (i.e. haploinsufficiency) epilepsy [29,50]. These findings indicate
289 that I/SC-associated epigenetic regulation is conserved across species, affects
290 functionally related gene families, and may influence human genetic disease inheritance
291 patterns (see below).
292
293 **Neurodevelopmental disease genes map to I/SCs.**

294 To further investigate the relationship between I/SC regulation and human disease genes,
295 we annotated the human protein-coding genes located within VERT regions (see
296 Supplemental Table 4). This analysis identified 979 genes overlapping VERT regions, of
297 which 272 are associated with single-gene disorders according to the OMIM database [29].

298 Among these genes, 155 are linked to autosomal recessive diseases, 87 to autosomal
299 dominant diseases, and 30 exhibit both dominant and recessive inheritance
300 patterns (Supplemental Table 8). Consistent with our GO analysis, which revealed an
301 enrichment for neurodevelopmental processes (see Fig. 1k), we found that various
302 neurodevelopmental disease genes were prominently represented among the disease-
303 associated genes (Table 3). Moreover, it was not unexpected to observe gene clusters
304 associated with neurodevelopmental disorders residing within VERT regions. For
305 example, Figures 6a–6c depict the location, I/SC profile, and genomic organization of a
306 cluster of sodium channel genes (SCN family, n = 5). Mutations in SCN1A, SCN2A,
307 SCN3A, and SCN9A genes are known to cause autosomal dominant developmental and
308 epileptic encephalopathies, with haploinsufficiency as a primary mechanism (Table 3 and
309 [29,51,52]). We also note that the syntenic locus in the mouse displays VERT, with the 5
310 syntenic mouse Scn genes residing within a VERT region on mouse chromosome 2
311 (Supplementary Fig. 2e-2f).

312 A second example of a cluster of epilepsy disease genes is shown in Figures 6d–
313 6e, where four members of the gamma-aminobutyric acid type A receptor gene family
314 (GABR family, n = 4) reside adjacent to a VERT region; notably, GABRA1 displays AEI
315 (Fig. 6e, and Supplementary Table 2), and mutations in GABRA1, GABRB2, and
316 GABRG2 cause autosomal dominant developmental and epileptic encephalopathies
317 [29,53,54]. We also note that a second GABR gene cluster (n = 4), where GABRA2 and
318 GABRB1 are also associated with autosomal dominant developmental and epileptic
319 encephalopathies, is adjacent to a VERT region on chromosome 4 (see Table 1). Two
320 additional examples are shown in Figures 6g–6j, where OXR1 and SAMD12 are located

321 in different VERT regions on chromosome 8, and mutations in each result in autosomal
322 dominant epilepsy [29].

323 Another prominent group of human disease genes located within I/SCs are those
324 associated with Parkinson disease (Table 3). As shown in Figure 7, four well-
325 characterized Parkinson disease genes, SNCA, LRRK2, DNAJC6, and PARKN, are
326 situated within or adjacent to VERT regions. Importantly, both SNCA and LRRK2 were
327 among the first genes identified as displaying AEI in the study by Gimelbrant et al. in 2007
328 [45]. In our own dataset, we found that DNAJC6 exhibits AEI in LCL clones [1]. We also
329 highlight that JAK1, a gene involved in immune signaling, lies within the same I/SC as
330 DNAJC6 (Fig. 7f). JAK1 is a known AEI gene [55,56], and mutations in
331 JAK1 cause autosomal dominant autoinflammation, immune dysregulation, and
332 eosinophilia (AIIDE). A recent study indicated that AEI of wild-type and mutant JAK1
333 alleles correlates with variable disease penetrance [28].

334 Taken together, these findings underscore the role of I/SCs in regulating disease-
335 relevant gene clusters and emphasize their conservation, synteny, and functional
336 relevance in both neurological and immune-related disorders. These examples highlight
337 how stochastic epigenetic regulation within I/SCs contributes to phenotypic variability and
338 disease risk.

339

340 **Discussion:**

341 The co-occurrence of random monoallelic expression and asynchronous replication
342 between alleles of autosomal genes is well documented [2-4], yet its developmental
343 potential and contribution to human disease have not been systematically investigated.

344 In this study, we expand on the discovery and characterization of Inactivation/Stability
345 Centers (I/SCs), autosomal loci marked by AEI and VERT, by identifying an
346 additional 163 VERT regions using clonal lines of primary ACP cells. This brings the total
347 number of I/SCs to 363, covering approximately 12% of the human genome.

348 To assess the evolutionary conservation of I/SC regulation, we analyzed
349 replication timing data from mouse pre-B cell clones, identifying 117 mouse VERT loci.
350 Many of these regions overlap with known mouse AEI genes and gene clusters
351 including olfactory, vomeronasal, and antigen receptors as well as protocadherin gene
352 families. Notably, we identified 21 syntenic VERT regions shared between human and
353 mouse genomes. These findings reinforce and extend previous observations indicating a
354 high degree of conservation of AEI genes [4] and suggest that I/SCs are a fundamental and
355 widespread feature of autosomal epigenetic regulation in mammals.

356 **I/SC regulation imparts cell-type specific and tissue-relevant variation**

357 Our clonal analysis of ACP and LCL cells, combined with haplotype-phased replication
358 timing and expression profiling, enabled us to identify tissue-specific I/SCs, while also
359 confirming a subset of I/SCs that are shared across tissues and conserved between
360 species**.** These findings support a dual regulatory model in which certain I/SCs may
361 serve ubiquitous functions and are utilized broadly across cell types, whereas others
362 are selectively engaged based on the gene expression program and functional
363 requirements of a given tissue. This model is supported by hierarchical clustering analysis
364 of VERT regions, which showed greater overlap within tissue type (ACP6 vs. ACP7; EB3-
365 2 vs. GM12878) than between tissues, suggesting that I/SC activity is tightly linked to
366 cellular identity.

367 Importantly, we identified 68 VERT regions shared across two or more clonal sets,
368 representing high-confidence I/SCs that likely reflect core regulatory regions active across
369 individuals and cell types (Fig. 1i). As detailed in Supplemental Table 1, we also
370 identified VERT regions detected exclusively in single clone sets. These private VERT
371 regions may reflect inter-individual variability in epigenetic regulation, context-dependent
372 I/SC usage, but more likely reflect a limitation of sampling—i.e., regions that would be
373 detected in additional clones if a larger number were analyzed. Thus, while our findings
374 provide strong support for both shared and tissue specific I/SCs, they also underscore the
375 importance of clone number and diversity in defining the full landscape of autosomal
376 epigenetic regulation.

377 In the context of development, AEI introduces stochastic variability in gene
378 dosage across clonal populations, contributing to cellular heterogeneity during lineage
379 specification and tissue formation. Such heterogeneity could serve as a substrate
380 for cellular selection, enabling the developing tissue to fine-tune function or respond
381 adaptively to environmental cues. This model aligns with observations that AEI often
382 affects genes involved in cell adhesion, signal transduction, and neurodevelopment,
383 where even small variations in expression levels could significantly impact cell fate and
384 connectivity.

385 At the molecular level, I/SCs contribute to cellular individuality by
386 enabling stochastic monoallelic expression of specific genes. This mechanism is
387 particularly critical in biological systems that require diversity at the single-cell level, such
388 as the olfactory, vomeronasal, immune, and nervous systems. Consistent with this, many
389 I/SCs coincide with gene clusters, including olfactory, vomeronasal, and antigen receptor
390 genes, as well as HLA and protocadherin genes, all of which are known to undergo

391 stochastic monoallelic expression. We identified numerous additional gene clusters
392 residing within I/SCs, suggesting that these other gene families may also participate in
393 tissue mosaicism and/or cellular individuality. Our current model proposes that I/SC
394 regulation is integral to tissue development, where stochastic, epigenetically driven allelic
395 regulation introduces phenotypic variability among cells. This variability may
396 promote functional diversity and selective advantage during development. Under this
397 model, each tissue would engage a distinct but partially overlapping set of I/SC-regulated
398 genes, aligned with its unique transcriptional and functional requirements. This
399 framework positions I/SCs as key players in promoting cellular individuality, tissue
400 mosaicism**,** and developmental plasticity.

401 Many of the genes that reside within I/SCs are regulated by allelic exclusion. Allelic
402 exclusion is a regulatory mechanism by which only one allele of a gene is expressed while
403 the other is actively silenced, ensuring monoallelic expression in each cell. This process
404 is well characterized in the immune system, where it plays a critical role in B-cell and T-
405 cell receptor gene rearrangements, ensuring that each lymphocyte produces a single
406 antigen receptor [18]. Furthermore, functional monoallelic expression is achieved via a
407 secondary mechanism that involves product inhibition. For example, in developing B cells,
408 expression of a productive immunoglobulin heavy chain allele suppresses rearrangement
409 and expression of the second allele, maintaining clonal specificity [18]. Allelic exclusion has
410 also been observed in olfactory receptors, where product feedback ensures expression
411 of a single allele of only one member of a large gene family, ensuring that each olfactory
412 receptor neuron responds to a single odorant [2,33,57-60].

413 **I/SC regulation acts through mechanisms distinct from X inactivation.**

414 Numerous studies in both human and mouse cells support the notion that allele-specific
415 epigenetic regulation of autosomal protein-coding genes is widespread
416 [4,11,13,37,38,40,57,61,62]. Although stochastic monoallelic expression on autosomes may appear
417 similar to X chromosome inactivation (XCI), the mechanisms behind them are distinct:
418 First, XCI is initiated in preimplantation embryos, as early as the 4-8 cell stage in the
419 mouse and 8-16 cell stage in humans [63-67], and later becomes established during
420 differentiation of the inner cell mass of the blastocyst in both species [67,68]. In contrast,
421 autosomal stochastic monoallelic expression appears later in development, with allelic
422 biases that can be erased and reestablished during differentiation [10,38]. For example, a
423 recent study found that allele-specific autosomal inactivation is absent during early blood
424 lineage specification but detectable in hematopoietic stem cells and their differentiated
425 progeny [9]. Second, in XCI, there is strong coordination along the chromosomes: genes
426 on the active X chromosome share both expression and early replication profiles,
427 whereas genes on the inactive X show silencing and late replication. In contrast,
428 autosomal inactivation lacks such coordination: allelic expression and replication timing
429 are unlinked within a chromosome and independent across the genome [1,7,69]. Taken
430 together, these observations indicate that autosomal allelic inactivation operates via a
431 mechanism distinct from XCI [1,9].

432 **An alternative model for haploinsufficiency**

433 The relationship between genotype and phenotype is fundamental to understanding
434 human genetic disease. At the scale of the whole genome, loss of function alleles are
435 generally not deleterious when heterozygous and are therefore haplosufficient [21,24]. Not
436 surprisingly, most of the loss of function mutations in single genes in humans are
437 recessive [21,29]. Therefore, for most human disease genes a single normal allele is

438 sufficient to achieve the typical phenotype. Dominantly inherited disorders comprise a
439 diverse group of conditions that are often severe and poorly understood, except in cases
440 of gain-of-function mutations, where dominant negative or interfering activities induce the
441 abnormal phenotype. In addition, many dominantly inherited diseases result from copy
442 number changes, either gain or loss of wild-type alleles [29,70]. Haploinsufficiency, predicted
443 to apply to as much as 40% of the human genome [19,20], refers to a heterozygous deletion
444 or complete loss-of-function mutation that results in an altered phenotype, indicating that
445 a single normal allele is insufficient. The underlying molecular mechanisms responsible
446 for most haploinsufficient human disease genes remain largely unknown.

447 The existence of two distinct populations of cells in XX individuals, one expressing
448 genes from the maternally derived X chromosome and the other from the paternally
449 derived X, has profound implications for both development and the manifestation of X-
450 linked diseases, where both dominant and recessive inheritance patterns are observed
451 in females [29]. A key factor influencing these inheritance patterns is the presence or
452 absence of cellular selection during development. In certain X-linked disorders, strong
453 selection favors the survival or expansion of cells in which the mutant allele is on the
454 inactive X chromosome, resulting in a predominance of wild-type cells and a clinical
455 presentation consistent with X-linked recessive inheritance [21,26]. In contrast, when cellular
456 selection is absent, heterozygous females typically retain a ~50:50 mosaic of wild-type
457 and mutant-expressing cells. The presence of mutant-expressing cells, essentially null
458 cells, is thought to underlie X-linked dominant phenotypes [27,71]. In such
459 cases, heterozygous loss of function mutations leads to a condition of haploinsufficiency,
460 in which the presence of a single normal allele is insufficient to maintain normal function.
461 Interestingly, although single X-linked alleles are sufficient in males, the mosaic

462 expression pattern in females unmasks phenotypic consequences that are otherwise
463 hidden. This highlights an underappreciated mechanism of haploinsufficiency: the cellular
464 mosaicism inherent to X inactivation can reduce effective gene dosage below functional
465 thresholds in individuals with heterozygous mutations, particularly in tissues where cells
466 expressing the mutant (null) allele persist.

467 One of the most striking implications of I/SC regulation is its relevance to human
468 genetic disease, particularly those involving haploinsufficiency. Among the 979
469 genes mapped within human VERT regions, 272 are associated with monogenic
470 diseases, including 87 with autosomal dominant inheritance, many of which are classified
471 as haploinsufficient [29]. Our findings offer a new mechanistic model for autosomal
472 haploinsufficiency: in the absence of compensatory upregulation and/or cellular selection
473 mechanisms, stochastic allelic inactivation results in or expands a functionally null cell
474 population, even when one allele is wild type. This mirrors the situation observed in X-
475 linked dominant disorders with loss of function mutations, where the absence of cellular
476 selection results in the persistence of null cells. This model expands upon prior theories
477 of haploinsufficiency that assume strict biallelic expression, which attribute dominant
478 phenotypes to factors such as threshold-dependent expression, stoichiometric imbalance
479 in multiprotein complexes, or inability to buffer expression noise [22-25]. In addition, our
480 model fits nicely with a previous study that modeled stochastic gene expression of
481 disease genes and proposed "haploinsufficiency syndromes might result from an
482 increased susceptibility to stochastic delays of gene initiation or interruptions of gene
483 expression" [72]. Our data suggest that epigenetically regulated stochastic AEI can
484 contribute to pathogenic phenotypes in heterozygous individuals by either reducing the
485 number of functional cells below a critical disease threshold or by disrupting the normal

486 mosaicism of gene expression within disease-relevant tissues. This expands the
487 conceptual framework of haploinsufficiency to include epigenetic, rather than purely
488 genetic, sources of dosage reduction.

489 **Limitations of our study.**

490 While our findings provide substantial evidence for the existence and functional relevance
491 of I/SCs, several important limitations remain to be addressed. First, the number of
492 clones analyzed in our studies, while sufficient to identify robust patterns of AEI and
493 VERT, limits our ability to fully capture the full spectrum of stochastic epigenetic states
494 across the genome. For example, we frequently observe AEI of genes located adjacent
495 to VERT regions, indicating that the boundaries of the epigenetic mechanisms functioning
496 at I/SCs remain undefined. This raises important questions about how far the influence of
497 I/SC-associated regulation extends beyond the VERT region. In addition, the specific
498 epigenetic marks, including DNA methylation, histone modifications, and chromatin
499 accessibility, at key regulatory elements such as replication origins, enhancers,
500 and promoters within I/SCs have yet to be mapped. Furthermore, the higher-order
501 chromatin organization of I/SCs, including their relationship to topologically associating
502 domains (TADs) and A/B compartments [49,73], remains largely unexplored and may hold
503 critical insights into their regulatory architecture and functional constraints. Second, our
504 analysis was conducted in a restricted number of cell types, namely LCLs and ACPs, and
505 therefore does not represent the full diversity of I/SC usage across human tissues. Third,
506 our study was limited to a small number of individuals, which may constrain the
507 identification of inter-individual variability in I/SC regulation. Fourth, the developmental
508 timing of I/SC establishment remains unresolved; it is unclear when during lineage
509 commitment or differentiation these regions acquire their allele-specific states. A related

510 and equally important question is whether I/SCs can be reset during reprogramming into
511 induced pluripotent stem cells (iPSCs) and subsequently reestablished upon
512 differentiation. This issue is particularly relevant given the widespread use of iPSCs in
513 human disease modeling and regenerative medicine [74-76]. Understanding how I/SCs
514 behave during cellular reprogramming and lineage specification will be critical for
515 interpreting AEI-related phenotypes in iPSC-derived systems.

516 An additional open question is how the magnitude of stochastic
517 AEI influences disease states, particularly in the context of variable
518 penetrance and expressivity. For example, a recent study employed allelic expression
519 ratios to identify functionally relevant AEI in genes associated with inborn errors of
520 immunity, showing that individuals with the same pathogenic mutation can
521 exhibit divergent clinical outcomes depending on the degree of allelic imbalance [28]. These
522 findings suggest that AEI can modulate disease severity, acting as an epigenetic layer of
523 regulation that determines whether a given mutation manifests clinically. We anticipate
524 that the biological significance of AEI will vary depending on the function of the gene,
525 its dosage sensitivity, and the cellular context in which it is expressed. Addressing these
526 questions will be essential for achieving a comprehensive understanding of the formation,
527 function, and disease relevance of I/SCs in human biology.

528

529 **Conclusion**

530 In conclusion, our work extends the catalog of I/SCs in the human genome and supports
531 a model in which stochastic, allele-specific epigenetic regulation at autosomal loci
532 contributes to cellular diversity and disease susceptibility. Moreover, by
533 identifying syntenic I/SCs in the mouse genome, our findings greatly enhance the
534 potential utility of mouse models for mechanistic studies of disease-causing mutations at

conserved loci. Our findings present a potential mechanism for incomplete penetrance, variable expressivity, and tissue-specific vulnerability, particularly in disorders caused by haploinsufficiency. In such cases, stochastic silencing of the wild-type allele in a subset of cells may lead to functional null states, even in heterozygous individuals, reducing the effective number of functional cells below a critical threshold. This may be especially relevant in the nervous system, where precise gene dosage is often essential for maintaining neuronal function. Moreover, I/SC regulation at disease genes raises the possibility that stochastic epigenetic regulation could act as an epigenetic modifier of disease risk, helping to explain why some individuals carrying pathogenic mutations remain asymptomatic. Understanding the full scope of I/SC regulation in both normal development and disease pathogenesis will be essential for identifying novel regulatory mechanisms and potential therapeutic targets. As tools for single-cell and allele-resolved epigenomics continue to advance, I/SCs may emerge as key elements in both basic biology and precision medicine.

RESOURCE AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Mathew Thayer (thayerm@ohsu.edu)

Materials availability statement

All data generated or analyzed during this study are included in the manuscript and supporting files.

Acknowledgments:

M.J.T. was supported by NIH (R01GM130703, and RF1NS142814).
M.B.H was supported by NIH NCI 4K00CA245677-03
DMG was supported by NIH R01GM083337

Author Contribution Statement

MBH, MJT, and DMG conceived of the experimental design for the RNA expression and replication timing assays. KF performed the surgeries to obtain the amputed digits. BJ and PAY performed the ACP cultures and clone generation. MBH and AEV carried out the RNA-seq assays, including data analysis. MBH and AEV carried out the Repli-seq assays, including data analysis. PFC and PTS provided advice during preparation of the manuscript. MBH, DMG and MJT wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

Literature Cited:

1. Heskett, M.B., Vouzas, A.E., Smith, L.G., Yates, P.A., Boniface, C., Bouhassira, E.E., Spellman, P.T., Gilbert, D.M., and Thayer, M.J. (2022). Epigenetic control of chromosome-associated lncRNA genes essential for replication and stability. Nature communications *13*, 6301. 10.1038/s41467-022-34099-7.
2. Goldmit, M., and Bergman, Y. (2004). Monoallelic gene expression: a repertoire of recurrent themes. Immunol Rev *200*, 197-214.
3. Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834. 10.1016/s0092-8674(94)90562-2.
4. Chess, A. (2016). Monoallelic Gene Expression in Mammals. Annu Rev Genet *50*, 317-327. 10.1146/annurev-genet-120215-035120.
5. Stoffregen, E.P., Donley, N., Stauffer, D., Smith, L., and Thayer, M.J. (2011). An autosomal locus that controls chromosome-wide replication timing and mono-allelic expression. Hum Mol Genet *20*, 2366-2378.
6. Thayer, M.J. (2012). Mammalian chromosomes contain cis-acting elements that control replication timing, mitotic condensation, and stability of entire chromosomes. Bioessays *34*, 760-770.
7. Donley, N., Smith, L., and Thayer, M.J. (2015). ASAR15, A cis-Acting Locus that Controls Chromosome-Wide Replication Timing and Stability of Human Chromosome 15. PLoS Genet *11*, e1004923. 10.1371/journal.pgen.1004923.
8. Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., et al. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nat Genet *47*, 353-360. 10.1038/ng.3222.
9. Kubasova, N., Alves-Pereira, C.F., Gupta, S., Vinogradova, S., Gimelbrant, A., and Barreto, V.M. (2022). In Vivo Clonal Analysis Reveals Random Monoallelic Expression in Lymphocytes That Traces Back to Hematopoietic Stem Cells. Frontiers in cell and developmental biology *10*. 10.3389/fcell.2022.827774.
10. Marion-Poll, L., Forêt, B., Zielinski, D., Massip, F., Attia, M., Carter, A.C., Syx, L., Chang, H.Y., Gendrel, A.V., and Heard, E. (2021). Locus specific epigenetic modalities of random allelic expression imbalance. Nature communications *12*, 5330. 10.1038/s41467-021-25630-3.
11. Zwemer, L.M., Zak, A., Thompson, B.R., Kirby, A., Daly, M.J., Chess, A., and Gimelbrant, A.A. (2012). Autosomal monoallelic expression in the mouse. Genome Biol *13*, R10. 10.1186/gb-2012-13-2-r10.
12. Yagi, T. (2013). Genetic basis of neuronal individuality in the mammalian brain. Journal of neurogenetics *27*, 97-105. 10.3109/01677063.2013.801969.
13. Savova, V., Chun, S., Sohail, M., McCole, R.B., Witwicki, R., Gai, L., Lenz, T.L., Wu, C.T., Sunyaev, S.R., and Gimelbrant, A.A. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat Genet *48*, 231-237. 10.1038/ng.3493.
14. Savova, V., Vigneau, S., and Gimelbrant, A.A. (2013). Autosomal monoallelic expression: genetics of epigenetic diversity? Curr Opin Genet Dev *23*, 642-648. 10.1016/j.gde.2013.09.001.
15. Johnstone, B., Stoddart, M.J., and Im, G.I. (2020). Multi-Disciplinary Approaches for Cell-Based Cartilage Regeneration. J Orthop Res *38*, 463-472. 10.1002/jor.24458.

16. Pattappa, G., Reischl, F., Jahns, J., Schewior, R., Lang, S., Zellner, J., Johnstone, B., Docheva, D., and Angele, P. (2021). Fibronectin Adherent Cell Populations Derived From Avascular and Vascular Regions of the Meniscus Have Enhanced Clonogenicity and Differentiation Potential Under Physioxia. Front Bioeng Biotechnol *9*, 789621. 10.3389/fbioe.2021.789621.
17. Rodriguez, I., Feinstein, P., and Mombaerts, P. (1999). Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. Cell *97*, 199-208. 10.1016/s0092-8674(00)80730-8.
18. Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B.E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221-225.
19. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. Nature *599*, 628-634. 10.1038/s41586-021-04103-z.
20. Jurgens, S.J., Wang, X., Choi, S.H., Weng, L.C., Koyama, S., Pirruccello, J.P., Nguyen, T., Smadbeck, P., Jang, D., Chaffin, M., et al. (2024). Rare coding variant analysis for human diseases across biobanks and ancestries. Nat Genet *56*, 1811-1820. 10.1038/s41588-024-01894-5.
21. Zschocke, J., Byers, P.H., and Wilkie, A.O.M. (2023). Mendelian inheritance revisited: dominance and recessiveness in medical genetics. Nat Rev Genet *24*, 442-463. 10.1038/s41576-023-00574-0.
22. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS Genet *6*, e1001154. 10.1371/journal.pgen.1001154.
23. Morrill, S.A., and Amon, A. (2019). Why haploinsufficiency persists. Proc Natl Acad Sci U S A *116*, 11866-11871. 10.1073/pnas.1900437116.
24. Wilkie, A.O. (1994). The molecular basis of genetic dominance. J Med Genet *31*, 89-98. 10.1136/jmg.31.2.89.
25. Veitia, R.A., Zschocke, J., and Birchler, J.A. (2025). Gene Dosage Sensitivity and Human Genetic Diseases. J Inherit Metab Dis *48*, e70058. 10.1002/jimd.70058.
26. Minks, J., Robinson, W.P., and Brown, C.J. (2008). A skewed view of X chromosome inactivation. J Clin Invest *118*, 20-23. 10.1172/jci34470.
27. Boone, P.M., Buenaventura, T., King, J.W.D., and Merkenschlager, M. (2025). X-linked competition - implications for human development and disease. Nat Rev Genet. 10.1038/s41576-025-00840-3.
28. Stewart, O., Gruber, C., Randolph, H.E., Patel, R., Ramba, M., Calzoni, E., Huang, L.H., Levy, J., Buta, S., Lee, A., et al. (2025). Monoallelic expression can govern penetrance of inborn errors of immunity. Nature *637*, 1186-1197. 10.1038/s41586-024-08346-4.
29. McKusick, V. (2025). Online Mendelian Inheritance in Man, OMIM®.
30. Gilad, Y., Rifkin, S.A., and Pritchard, J.K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet *24*, 408-415. 10.1016/j.tig.2008.06.001.
31. Heskett, M.B., Smith, L.G., Spellman, P., and Thayer, M.J. (2020). Reciprocal monoallelic expression of ASAR lncRNA genes controls replication timing of human chromosome 6. RNA (New York, N.Y.) *26*, 724-738. 10.1261/rna.073114.119.

32. Fu, R., Qin, P., Zou, X., Hu, Z., Hong, N., Wang, Y., and Jin, W. (2021). A Comprehensive Characterization of Monoallelic Expression During Hematopoiesis and Leukemogenesis via Single-Cell RNA-Sequencing. Frontiers in cell and developmental biology *9*, 702897. 10.3389/fcell.2021.702897.
33. Zakharova, I.S., Shevchenko, A.I., and Zakian, S.M. (2009). Monoallelic gene expression in mammals. Chromosoma *118*, 279-290.
34. Xu, J., Carter, A.C., Gendrel, A.V., Attia, M., Loftus, J., Greenleaf, W.J., Tibshirani, R., Heard, E., and Chang, H.Y. (2017). Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. Nat Genet *49*, 377-386. 10.1038/ng.3769.
35. Nag, A., Savova, V., Fung, H.L., Miron, A., Yuan, G.C., Zhang, K., and Gimelbrant, A.A. (2013). Chromatin signature of widespread monoallelic expression. eLife *2*, e01256. 10.7554/eLife.01256.
36. Li, S.M., Valo, Z., Wang, J., Gao, H., Bowers, C.W., and Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. PLoS One *7*, e31751.
37. Gendrel, A.V., Marion-Poll, L., Katoh, K., and Heard, E. (2016). Random monoallelic expression of genes on autosomes: Parallels with X-chromosome inactivation. Semin Cell Dev Biol. 10.1016/j.semcdb.2016.04.007.
38. Gendrel, A.V., Attia, M., Chen, C.J., Diabangouaya, P., Servant, N., Barillot, E., and Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. Dev Cell *28*, 366-380. 10.1016/j.devcel.2014.01.016.
39. Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P., and Spector, D.L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. Dev Cell *28*, 351-365. 10.1016/j.devcel.2014.01.017.
40. Barreto, V.M., Kubasova, N., Alves-Pereira, C.F., and Gendrel, A.V. (2021). X-Chromosome Inactivation and Autosomal Random Monoallelic Expression as "Faux Amis". Frontiers in cell and developmental biology *9*, 740937. 10.3389/fcell.2021.740937.
41. Blumenfeld, B., Masika, H., Farago, M., Yehuda, Y., Halaseh, L., Vardi, O., Rapoport, R., Levin-Klein, R., Cedar, H., Bergman, Y., and Simon, I. (2021). Chromosomal coordination and differential structure of asynchronous replicating regions. Nature communications *12*, 1035. 10.1038/s41467-021-21348-4.
42. Mountoufaris, G., Chen, W.V., Hirabayashi, Y., O'Keeffe, S., Chevee, M., Nwakeze, C.L., Polleux, F., and Maniatis, T. (2017). Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. Science *356*, 411-414. 10.1126/science.aai8801.
43. Williams, D.L., Sikora, V.M., Hammer, M.A., Amin, S., Brinjikji, T., Brumley, E.K., Burrows, C.J., Carrillo, P.M., Cromer, K., Edwards, S.J., et al. (2021). May the Odds Be Ever in Your Favor: Non-deterministic Mechanisms Diversifying Cell Surface Molecule Expression. Frontiers in cell and developmental biology *9*, 720798. 10.3389/fcell.2021.720798.
44. Savova, V., Vinogradova, S., Pruss, D., Gimelbrant, A.A., and Weiss, L.A. (2017). Risk alleles of genes with monoallelic expression are enriched in gain-of-function variants and depleted in loss-of-function variants for neurodevelopmental disorders. Mol Psychiatry *22*, 1785-1794. 10.1038/mp.2017.13.

45. Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science *318*, 1136-1140.
46. Ito, H., Morishita, R., and Nagata, K.I. (2017). Autism spectrum disorder-associated genes and the development of dentate granule cells. Med Mol Morphol *50*, 123-129. 10.1007/s00795-017-0161-z.
47. Lin, L., Lesnick, T.G., Maraganore, D.M., and Isacson, O. (2009). Axon guidance and synaptic maintenance: preclinical markers for neurodegenerative disease and therapeutics. Trends Neurosci *32*, 142-149. 10.1016/j.tins.2008.11.006.
48. Mosca-Boidron, A.L., Gueneau, L., Huguet, G., Goldenberg, A., Henry, C., Gigot, N., Pallesi-Pocachard, E., Falace, A., Duplomb, L., Thevenon, J., et al. (2016). A de novo microdeletion of SEMA5A in a boy with autism spectrum disorder and intellectual disability. European journal of human genetics : EJHG *24*, 838-843. 10.1038/ejhg.2015.211.
49. Richer, S., Tian, Y., Schoenfelder, S., Hurst, L., Murrell, A., and Pisignano, G. (2023). Widespread allele-specific topological domains in the human genome are not confined to imprinted gene clusters. Genome Biol *24*, 40. 10.1186/s13059-023-02876-2.
50. Reutlinger, C., Helbig, I., Gawelczyk, B., Subero, J.I., Tönnies, H., Muhle, H., Finsterwalder, K., Vermeer, S., Pfundt, R., Sperner, J., et al. (2010). Deletions in 16p13 including GRIN2A in patients with intellectual disability, various dysmorphic features, and seizure disorders of the rolandic region. Epilepsia *51*, 1870-1873. 10.1111/j.1528-1167.2010.02555.x.
51. Ding, J., Li, X., Tian, H., Wang, L., Guo, B., Wang, Y., Li, W., Wang, F., and Sun, T. (2021). SCN1A Mutation-Beyond Dravet Syndrome: A Systematic Review and Narrative Synthesis. Front Neurol *12*, 743726. 10.3389/fneur.2021.743726.
52. Wolff, M., Johannesen, K.M., Hedrich, U.B.S., Masnada, S., Rubboli, G., Gardella, E., Lesca, G., Ville, D., Milh, M., Villard, L., et al. (2017). Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. Brain *140*, 1316-1336. 10.1093/brain/awx054.
53. Hernandez, C.C., Hu, N., Shen, W., and Macdonald, R.L. (2023). Epileptic Encephalopathy GABRB Structural Variants Share Common Gating and Trafficking Defects. Biomolecules *13*. 10.3390/biom13121790.
54. Hamdan, F.F., Myers, C.T., Cossette, P., Lemay, P., Spiegelman, D., Laporte, A.D., Nassif, C., Diallo, O., Monlong, J., Cadieux-Dion, M., et al. (2017). High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. Am J Hum Genet *101*, 664-685. 10.1016/j.ajhg.2017.09.008.
55. Gruber, C.N., Calis, J.J.A., Buta, S., Evrony, G., Martin, J.C., Uhl, S.A., Caron, R., Jarchin, L., Dunkin, D., Phelps, R., et al. (2020). Complex Autoinflammatory Syndrome Unveils Fundamental Principles of JAK1 Kinase Transcriptional and Biochemical Function. Immunity *53*, 672-684.e611. 10.1016/j.immuni.2020.07.006.
56. Ross, S.H., and Cantrell, D.A. (2020). Of Mosaicism and Mechanisms: How JAK1 Goes Awry. Immunity *53*, 481-484. 10.1016/j.immuni.2020.08.010.
57. Khamlichi, A.A., and Feil, R. (2018). Parallels between Mammalian Mechanisms of Monoallelic Gene Expression. Trends Genet *34*, 954-971. 10.1016/j.tig.2018.08.005.

58. Dalton, R.P., Lyons, D.B., and Lomvardas, S. (2013). Co-opting the unfolded protein response to elicit olfactory receptor feedback. Cell *155*, 321-332. 10.1016/j.cell.2013.09.033.
59. Lyons, D.B., Allen, W.E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An epigenetic trap stabilizes singular olfactory receptor expression. Cell *154*, 325-336. 10.1016/j.cell.2013.06.039.
60. Pourmorady, A.D., Bashkirova, E.V., Chiariello, A.M., Belagzhal, H., Kodra, A., Duffié, R., Kahiapo, J., Monahan, K., Pulupa, J., Schieren, I., et al. (2024). RNA-mediated symmetry breaking enables singular olfactory receptor choice. Nature *625*, 181-188. 10.1038/s41586-023-06845-4.
61. Chess, A. (2012). Mechanisms and consequences of widespread random monoallelic expression. Nat Rev Genet *13*, 421-428.
62. Reinius, B., and Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nat Rev Genet *16*, 653-664. 10.1038/nrg3888.
63. Brockdorff, N., and Turner, B.M. (2015). Dosage compensation in mammals. Cold Spring Harb Perspect Biol *7*, a019406. 10.1101/cshperspect.a019406.
64. Dossin, F., and Heard, E. (2022). The Molecular and Nuclear Dynamics of X-Chromosome Inactivation. Cold Spring Harb Perspect Biol *14*. 10.1101/cshperspect.a040196.
65. Galupa, R., and Heard, E. (2018). X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. Annu Rev Genet *52*, 535-566. 10.1146/annurev-genet-120116-024611.
66. Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet *42*, 733-772. 10.1146/annurev.genet.42.110807.091711 [doi].
67. van den Berg, I.M., Galjaard, R.J., Laven, J.S., and van Doorninck, J.H. (2011). XCI in preimplantation mouse and human embryos: first there is remodelling…. Hum Genet *130*, 203-215. 10.1007/s00439-011-1014-9.
68. van den Berg, I.M., Laven, J.S., Stevens, M., Jonkers, I., Galjaard, R.J., Gribnau, J., and van Doorninck, J.H. (2009). X chromosome inactivation is initiated in human preimplantation embryos. Am J Hum Genet *84*, 771-779. 10.1016/j.ajhg.2009.05.003.
69. Donley, N., Stoffregen, E.P., Smith, L., Montagna, C., and Thayer, M.J. (2013). Asynchronous Replication, Mono-Allelic Expression, and Long Range Cis-Effects of ASAR6. PLoS Genet *9*, e1003423.
70. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen--the Clinical Genome Resource. N Engl J Med *372*, 2235-2242. 10.1056/NEJMsr1406261.
71. Giovenino, C., Trajkova, S., Pavinato, L., Cardaropoli, S., Pullano, V., Ferrero, E., Sukarova-Angelovska, E., Carestiato, S., Salmin, P., Rinninella, A., et al. (2023). Skewed X-chromosome inactivation in unsolved neurodevelopmental disease cases can guide re-evaluation For X-linked genes. European journal of human genetics : EJHG *31*, 1228-1236. 10.1038/s41431-023-01324-w.
72. Cook, D.L., Gerber, A.N., and Tapscott, S.J. (1998). Modeling stochastic gene expression: implications for haploinsufficiency. Proc Natl Acad Sci U S A *95*, 15641-15646. 10.1073/pnas.95.26.15641.

73. Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402-405. 10.1038/nature13986.
74. Li, Y., and Darabi, R. (2022). Role of epigenetics in cellular reprogramming; from iPSCs to disease modeling and cell therapy. J Cell Biochem *123*, 147-154. 10.1002/jcb.30164.
75. Janiszewski, A., Talon, I., Chappell, J., Collombet, S., Song, J., De Geest, N., To, S.K., Bervoets, G., Marin-Bejar, O., Provenzano, C., et al. (2019). Dynamic reversal of random X-Chromosome inactivation during iPSC reprogramming. Genome Res *29*, 1659-1672. 10.1101/gr.249706.119.
76. Tcw, J. (2019). Human iPSC application in Alzheimer's disease and Tau-related neurodegenerative diseases. Neurosci Lett *699*, 31-40. 10.1016/j.neulet.2019.01.043.
77. Pattappa, G., Markway, B.D., Docheva, D., and Johnstone, B. (2023). Physioxic Culture of Chondrogenic Cells. Methods Mol Biol *2598*, 45-63. 10.1007/978-1-0716-2839-3_5.
78. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) *25*, 1754-1760. 10.1093/bioinformatics/btp324.
79. Garrison, E.a.M., G (2024). Haplotype-based variant detection from short-read sequencing. arXiv *1207.3907v2*.
80. Martin, M., Ebert, P., and Marschall, T. (2023). Read-Based Phasing and Analysis of Phased Variants with WhatsHap. Methods Mol Biol *2590*, 127-138. 10.1007/978-1-0716-2819-5_8.
81. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) *25*, 2078-2079. 10.1093/bioinformatics/btp352.
82. Marchal, C., Sasaki, T., Vera, D., Wilson, K., Sima, J., Rivera-Mulia, J.C., Trevilla-García, C., Nogues, C., Nafie, E., and Gilbert, D.M. (2018). Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. Nat Protoc *13*, 819-839. 10.1038/nprot.2017.148.
83. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) *29*, 15-21. 10.1093/bioinformatics/bts635.

854 **Methods**

855 *Cell culture*

856 All human tissue samples were obtained in accordance with protocols approved by the
857 Oregon Health & Science University Institutional Review Board (OHSU IRB# 00019018)
858 and the Western Institutional Review Board (WIRB# 20193303). Informed consent was
859 obtained from all donors or their legal guardians. ACP cells were grown in Dulbecco's
860 Modified Eagle's Medium low glucose (Life Technologies) supplemented with 10% fetal
861 bovine serum (Hyclone). Single cell clones were isolated by plating individual cells in 6
862 well dishes, and were expanded for >20 population doublings [77]. All cells were grown in a
863 humidified incubator at 37°C in a 5% carbon dioxide atmosphere.

864

865 *Haplotype Phasing of ACP cells*

866 Short-read whole genome sequencing was obtained from two parent/child trios with a
867 range of ~7-14x genome coverage. BWA [78] was used for alignment, and FreeBayes [79]
868 used to call putative germline variants jointly for each trio. Variants were filtered to remove
869 sites with Genotype Quality score of less than 30, and sites with missing genotypes for
870 any member of the trio were removed. WhatsHap [80] was then used to phase variants in
871 the child utilizing mendelian inheritance patterns and reads containing multiple variants.
872 1,679,207 high-confidence sites were phased in ACP7, and 837,323 sites were phased
873 in ACP6.

874

875 *Allele-specific Repli-Seq*

876 Short read sequence data was aligned with BWA with standard parameters. BAM files
877 were then deduplicated and sorted with SAMtools [81]. BCFtools mpileup [81] was then used
878 to produce allele-specific read counts at phased heterozygous loci. Library size

879 normalization was performed to produce counts per million informative reads. Read
880 counts were then grouped into 250kb sliding genomic windows with no overlap. E/L Repli-
881 Seq values were generated by taking the Log base 2 of early versus late values [82].
882 Quantile normalization was performed on each group of subclones within each individual
883 sample to allow for comparison among subclones. For each sample, variability of allele-
884 specific RT among subclones was defined as the Standard Deviation (degrees of freedom
885 = 1) of repli-seq at each 250kb window. Genome-wide std. dev. values generated a log-
886 normal distribution, and we set a threshold of Z>=2.25 on the log-normal to identify
887 regions with highly variable allele-specific RT among subclones.
888
889 *Allele-specific RNA-seq*
890 Nuclei were isolated by centrifugation for 0.5 minutes from ACP clones following lysis in
891 0.5% NP40, 140 mM NaCl, 10 mM Tris-HCl (pH 7.4), and 1.5 mM $MgCl_2$. Nuclear RNA
892 was isolated using Trizol reagent using the manufacturer's instructions, followed by
893 DNase treatment to remove possible genomic DNA contamination. Briefly, ribosomal
894 RNAs were removed using the Ribo-Zero kit (Illumina), RNA was fragmented into 250-
895 300bp fragments, and cDNA libraries were prepared using the Directional RNA Library
896 Prep Kit (NEB). Paired end sequencing was done on a NovaSeq 6000 at the OHSU
897 MPSSR core facility. Strand specific short read sequence data was aligned with STAR [83].
898 SAMtools [81] was used to remove duplicates and sort BAM files. Strand-specific, allele-
899 specific reads at heterozygous phased SNPs were counted using BCFtools mpileup [81]
900 and grouped together for all SNPs within each gene. Allelic expression imbalance was
901 calculated as the fraction of reads derived from the maternal vs the paternal allele. For
902 example, completely balanced allelic expression is 50% AEI, and completely imbalanced
903 AEI is 100%. A threshold of AEI>= 70%, and FDR corrected binomial p-value<=0.05 was

904 used to classify AEI in each individual subclone. Standard deviation (degrees of
905 freedom=1) was used for each gene to generate a metric of AEI variability among
906 subclones. A genome-wide distribution of std. dev. was generated across all genes, and
907 the top 2.5% of variable genes were classified as highly variable among subclones.
908

Figure Legends:

Figure 1: Genome wide variable replication-timing and expression. a) Single cell derived ACP clones (color coded) are isolated, expanded and processed for Repliseq and RNAseq. The maternal and paternal haplotypes were used to analyze allele-specific replication timing and expression. b) The top panel illustrates the expected replication timing profile in multiple clones across an autosomal region. The regions that display VERT are highlighted by shading. The bottom panel illustrates the expected expression variability in different clones from the same individual. c) We define AEI as an allelic bias that is ≥80% allelic imbalance (AEI ≥ 0.80 or ≤ 0.20). d) Illustration of an Inactivation/Stability Center showing the clonal and allele-restricted variability in replication timing and expression of both protein coding and noncoding genes. e) The standard deviation (Std. Dev.) in 250 kb windows (circles) across human chromosome 1. Outlier windows from all four sets of clones (ACP6, ACP7, EB3-2, and GM12878) are highlighted in different colors as shown. The top panel illustrates the G-banding (blue shading) pattern for human chromosome 1, with 23 VERT regions highlighted in red. The arrow marks the VERT region shown in panels f-h below. f-h) I/SC located between 239 and 242 mb of chromosome 1 (see panel e). f) VERT (shaded region) and AEI of genes (rectangles) detected in the ACP6 clones, each clone was color coded as shown. For replication, the solid lines represent the maternal (Mat.) allele and the dotted line represents the paternal (Pat.) allele. g) VERT (shaded region) and AEI of genes (rectangles) detected in the ACP7 clones, each clone was color coded as shown. h) UCSC Genome Browser view of the genomic region in panels f and g. The shaded area highlights the VERT region. i) Hierarchical clustering of the VERT regions that are shared between 2 or more clone sets showing both ACP and LCL clones. j) Gene Ontology (GO) enrichment analysis of the genes located within the VERT regions, the most significant hits for each set of clones, organized by cell type, are shown. GO analysis of the mouse genes located within the pre-B cell clone VERT regions are also shown.

Figure 2. Examples of I/SCs. a, d and g) VERT regions on human chromosomes 18, 6 and 9, respectively. The top panels illustrate the G-banding (blue shading) pattern for each human chromosome, with VERT regions highlighted in red. The standard deviation (Std. Dev.) in 250 kb windows (circles) across each human chromosome is shown. Outlier windows from all four sets of clones (ACP6, ACP7, EB3-2, and GM12878) are highlighted in different colors as shown. The arrows and red boxes mark the VERT regions shown in panels b, e, and h. b, e and h) I/SCs highlighted in panels a, d, and g. For replication, the solid lines represent the maternal (Mat.) allele and the dotted line represents the paternal (Pat.) allele. The regions with VERT are shaded and the AEI is shown for each clone in a different color. c, f and i) UCSC Genome Browser views of the genomic regions in panels b, e and h. The shaded areas highlight the VERT regions. Genes with AEI are marked with +, *, or #. b and c) The + marks MOCOS, which shows AEI in ACP cells (Supplemental Table 2) and ++ marks FHOD3, which shows AEI in LCLs [1]. e and f) The * marks from DNAH8, ** marks transcripts from KCNK5, and *** marks non-coding transcripts from ENSG00000293066 (also named: TL:6-39.9_1 [1]). All three of these genes show AEI in LCL clones [1]. h and i) The # marks transcripts from DAPK1, ## marks transcripts from SPIN1. Both genes show AEI in LCL clones [1].

Figure 3. Human gene clusters with AEI map to VERT regions. a and c) VERT regions on human chromosome 9 and 6. The top panels illustrate the G-banding (blue shading) pattern for each human chromosome, with VERT regions highlighted in red. The standard deviation (Std. Dev.) in 250 kb windows (circles) across each human chromosome is shown. Outlier windows from all four sets of clones (ACP6, ACP7, EB3-2, and GM12878) are highlighted in different colors as shown. The arrows and red boxes mark the VERT regions shown in panels b, e, and h. The red box and arrow identify the I/SC in panel b. b) UCSC genome browser view illustrating the genomic location highlighted in panel a. The shaded area highlights the VERT region. The red lines mark 11 olfactory receptor genes, and the green lines mark 6 cytochrome P450 family (CYP4F) genes. d) UCSC Genome Browser view illustrating the genomic location highlighted in panel c. The shaded area highlights the VERT region. The red lines mark 16 HLA genes, and the green lines mark 6 Lymphocyte antigen 6 family (LY6G) genes.

Figure 4. Mouse gene clusters with AEI map to VERT regions. a and d) VERT regions on mouse chromosome 2 and 18. The standard deviation (Std. Dev.) in 50 kb windows (circles) across each mouse chromosome are shown. Outlier windows from pre-B cell clones are highlighted in orange. The red boxes and arrows identify the VERT regions in panels b (i) and c (ii). b and c) UCSC genome browser views illustrating the genomic locations represented by i and ii in panel a. b) Genome Browser view of the VERT region showing the location of 29 olfactory receptor genes (Or genes). The shaded area represents the VERT region. c) Genome Browser view of the VERT region in ii above, showing the location of 242 olfactory receptor genes (Or genes). The shaded area represents the VERT region. e) Genome Browser view of the VERT region in d above, showing the location of clustered Protocaderhin genes (15 Pcdha, 23 Pcdhb, and 22 Pcdhg genes). The shaded area represents the VERT region.

Figure 5. Human and Mouse syntenic regions display VERT. a and b) The short arm of human chromosome 5, with 11 VERT regions (panel a) and the centromeric region of mouse chromosome 15, with 4 VERT regions (panel b) are shown. a) Two human VERT regions are highlighted, hi and hii, and are syntenic with locations on mouse chromosome 15, mi and mii, that also show VERT (panel b). Note that the gene order in both syntenic regions are inverted with respect to the telomeres when comparing the human and mouse loci. c) Illustration of the I/SC on human chromosome 5 between 7.75 and 10.5 mb (hi). The shaded area marks the VERT region, and AEI of the coding gene SEMA5A in different clones is highlighted. For replication, the maternal (Mat.) and paternal (Pat.) alleles are indicated. d) UCSC Genome Browser view of the human VERT region (shaded) in hi, showing the location of SEMA5A and a taste receptor gene (TAS2R1). e) UCSC Genome Browser view of the mouse VERT region (shaded) in mi, showing the location of Sema5a and a taste receptor gene (Tas2r119). f) UCSC Genome Browser view of the human VERT region (shaded) in hii above, highlighting the location of the AEI gene RETREG1 [45]. g) UCSC Genome Browser view of the mouse VERT region (shaded) in mii, showing the synteny between human and mouse for the protein coding genes in panel f, including Retreg1. h) VERT regions on human chromosome 16. The standard deviation (Std. Dev.) in 250 kb windows (circles) across human chromosome 16 is shown. Outlier windows are highlighted in different colors as shown. The red box and arrow mark

the genomic location in panel i. i) UCSC Genome Browser view illustrating the genomic location represented in panel h, showing the location of GRIN2A (red box). j) VERT regions on mouse chromosome 16. The standard deviation in 50 kb windows (circles) is shown. Outlier windows from pre-B cell clones are highlighted in orange. k) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panel j above, with the location of Grin2a highlighted with a red box.

Figure 6. Epilepsy genes map to I/SCs. a, d, and g) VERT regions on human chromosomes 2, 5, and 21. The top panels illustrate the G-banding (blue shading) pattern for each human chromosome, with VERT regions highlighted in red. The standard deviation (Std. Dev.) in 250 kb windows (circles) across each human chromosome is shown. Outlier windows from all four sets of clones (ACP6, ACP7, EB3-2, and GM12878) are highlighted in different colors as shown. The arrows and red boxes mark the VERT regions shown in panels b, c, e, f, h, i, j and k. b, e, h, j) Illustrations of I/SCs on chromosomes 2, 5, and 8. The shaded areas mark the VERT regions. Each clone was color coded to illustrate the VERT and AEI in different clones. The maternal (Mat.) and paternal (Pat.) alleles are indicated for replication and expression. c) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panels a and b, with the location of 5 sodium channel genes (SCN3A, SCN2A, SCN1A, SCN9A, and SCN7A) highlighted with a red line. e) Illustration of an I/SC on chromosome 5. The shaded area marks the VERT region. AEI of GABRA1 is shown. f) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panels d and e above, with the location of 4 GABA receptor genes (GABRB2, GABRBA6, GABRA1, and GABRG2) highlighted with a red line. h) Illustration of a I/SCs (A) on chromosome 8. Each ACP6 clone was color coded to illustrate the VERT and AEI in different clones. The maternal (Mat.) and paternal (Pat.) alleles are indicated for replication and expression. The shaded area marks the VERT region. The location and expression of OXR1 is indicated. i) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panels g (A) and h above, with the location of the OXR1 gene highlighted with a red line. We note that ANGPT1 (marked with *), located within this I/SC displays AEI (see Supplementary Table 2). j) Illustration of an I/SC (B) on chromosome 8. AEI of SAMD12 is indicated. k) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panels g (B) and j above, with the location of the SAMD12 highlighted with a red line.

Figure 7. Parkinson disease genes map to I/SCs. a, c, e, and g) VERT regions on human chromosomes 4, 12, 19, and 20, respectively. The top panels illustrate the G-banding (blue shading) pattern for each human chromosome, with VERT regions highlighted in red. The standard deviation (Std. Dev.) in 250 kb windows (circles) across each human chromosome is shown. Outlier windows from all four sets of clones (ACP6, ACP7, EB3-2, and GM12878) are highlighted in different colors as shown. The arrows and red boxes mark the VERT regions shown in panels b, d, f, and h, respectively. b) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panel a, with the location of the SNCA gene highlighted with a red line. d) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) highlighted in panel c, with the location of the LRRK2 gene highlighted with a red line. f) UCSC Genome Browser view illustrating the genomic location of the VERT region (shaded area) represented in panel e, with the location of the DNAJC6 gene highlighted with a red line; also highlighted with a green line is the JAK1 gene. h) UCSC Genome Browser view illustrating the genomic location

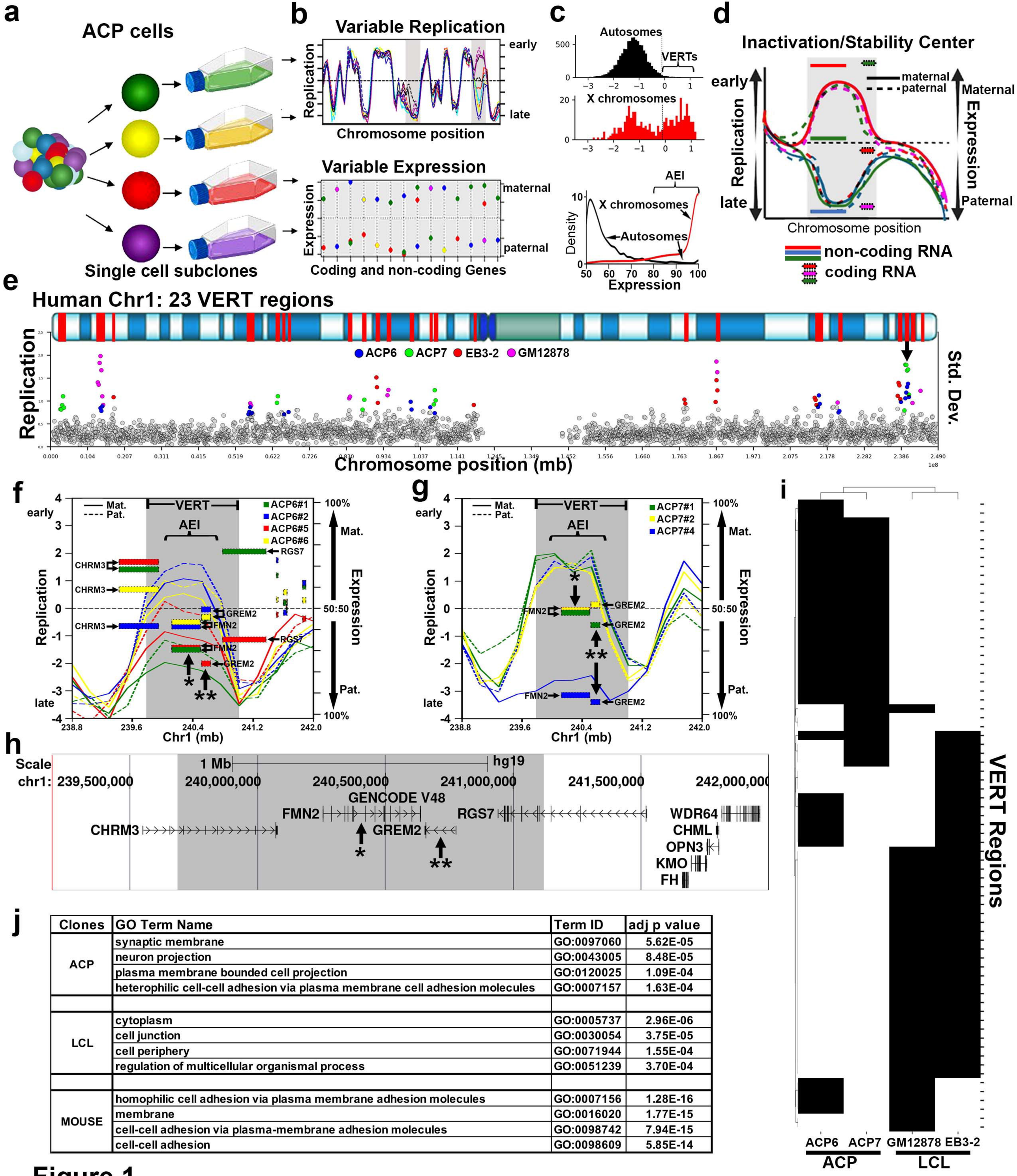
of the VERT region (shaded area) represented in panel g, with the location of the PRKN gene highlighted with a red line.

RESOURCE AVAILABILITY

Source data are provided with this paper. Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Mathew Thayer (thayerm@ohsu.edu).

Materials availability statement

All reagents generated in this study will be made freely available upon request.



j

Clones	GO Term Name	Term ID	adj p value
ACP	synaptic membrane	GO:0097060	5.62E-05
	neuron projection	GO:0043005	8.48E-05
	plasma membrane bounded cell projection	GO:0120025	1.09E-04
	heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	GO:0007157	1.63E-04
LCL	cytoplasm	GO:0005737	2.96E-06
	cell junction	GO:0030054	3.75E-05
	cell periphery	GO:0071944	1.55E-04
	regulation of multicellular organismal process	GO:0051239	3.70E-04
MOUSE	homophilic cell adhesion via plasma membrane adhesion molecules	GO:0007156	1.28E-16
	membrane	GO:0016020	1.77E-15
	cell-cell adhesion via plasma-membrane adhesion molecules	GO:0098742	7.94E-15
	cell-cell adhesion	GO:0098609	5.85E-14

Figure 1

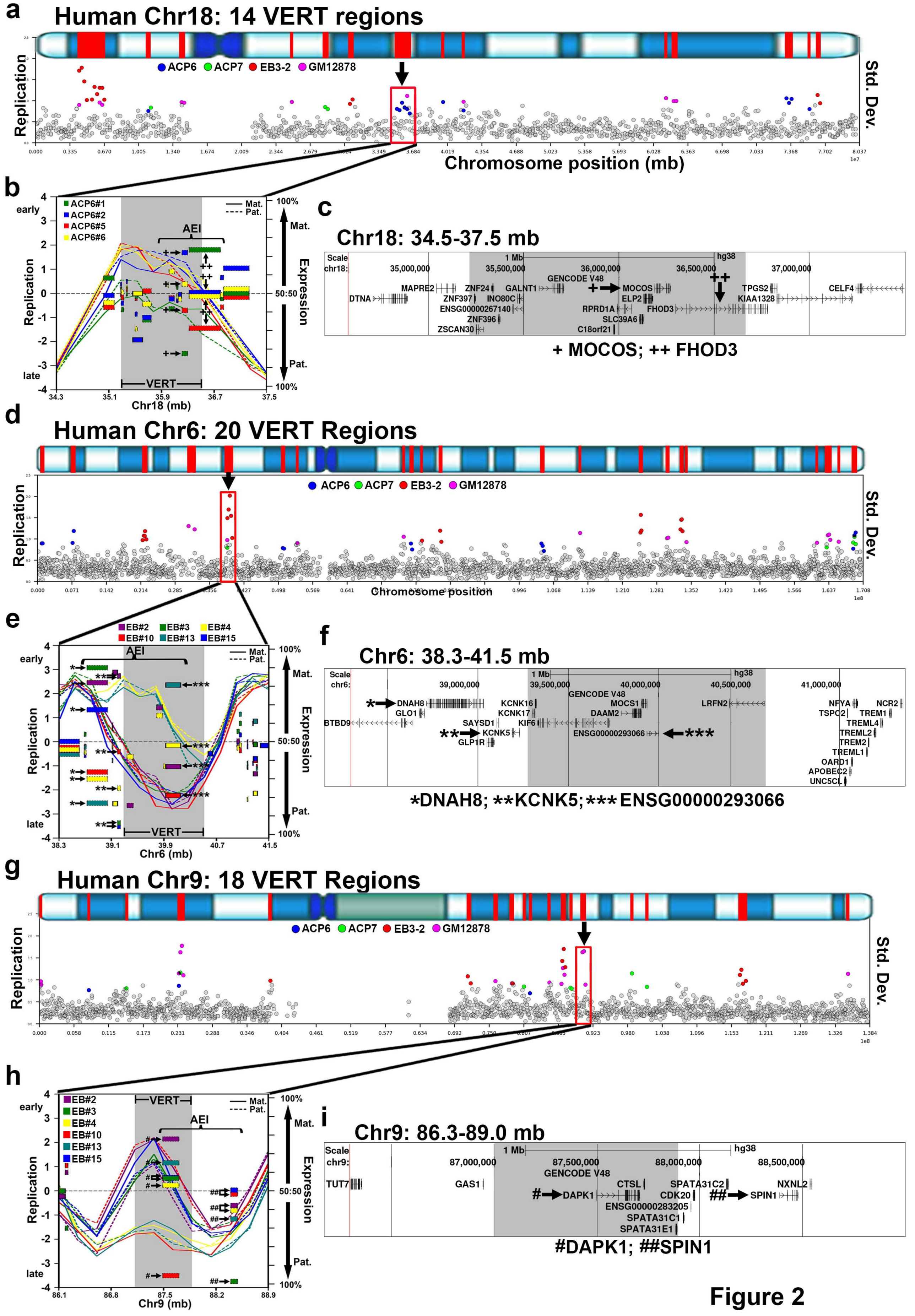

a
Human Chr18: 14 VERT regions
ACP6 ACP7 EB3-2 GM12878
Replication
Std. Dev.
Chromosome position (mb)
b
ACP6#1
ACP6#2
ACP6#5
ACP6#6
Mat.
Pat.
AEI
VERT
early
late
Replication
Expression
Mat.
Pat.
50:50
Chr18 (mb)
c
Chr18: 34.5-37.5 mb
+ MOCOS; ++ FHOD3
d
Human Chr6: 20 VERT Regions
ACP6 ACP7 EB3-2 GM12878
Replication
Std. Dev.
Chromosome position
e
EB#2
EB#3
EB#4
EB#10
EB#13
EB#15
AEI
VERT
Mat.
Pat.
early
late
Replication
Expression
50:50
Chr6 (mb)
f
Chr6: 38.3-41.5 mb
*DNAH8; **KCNK5; *** ENSG00000293066
g
Human Chr9: 18 VERT Regions
ACP6 ACP7 EB3-2 GM12878
Replication
Std. Dev.
h
EB#2
EB#3
EB#4
EB#10
EB#13
EB#15
VERT
AEI
Mat.
Pat.
early
late
Replication
Expression
50:50
Chr9 (mb)
i
Chr9: 86.3-89.0 mb
#DAPK1; ##SPIN1

Figure 2

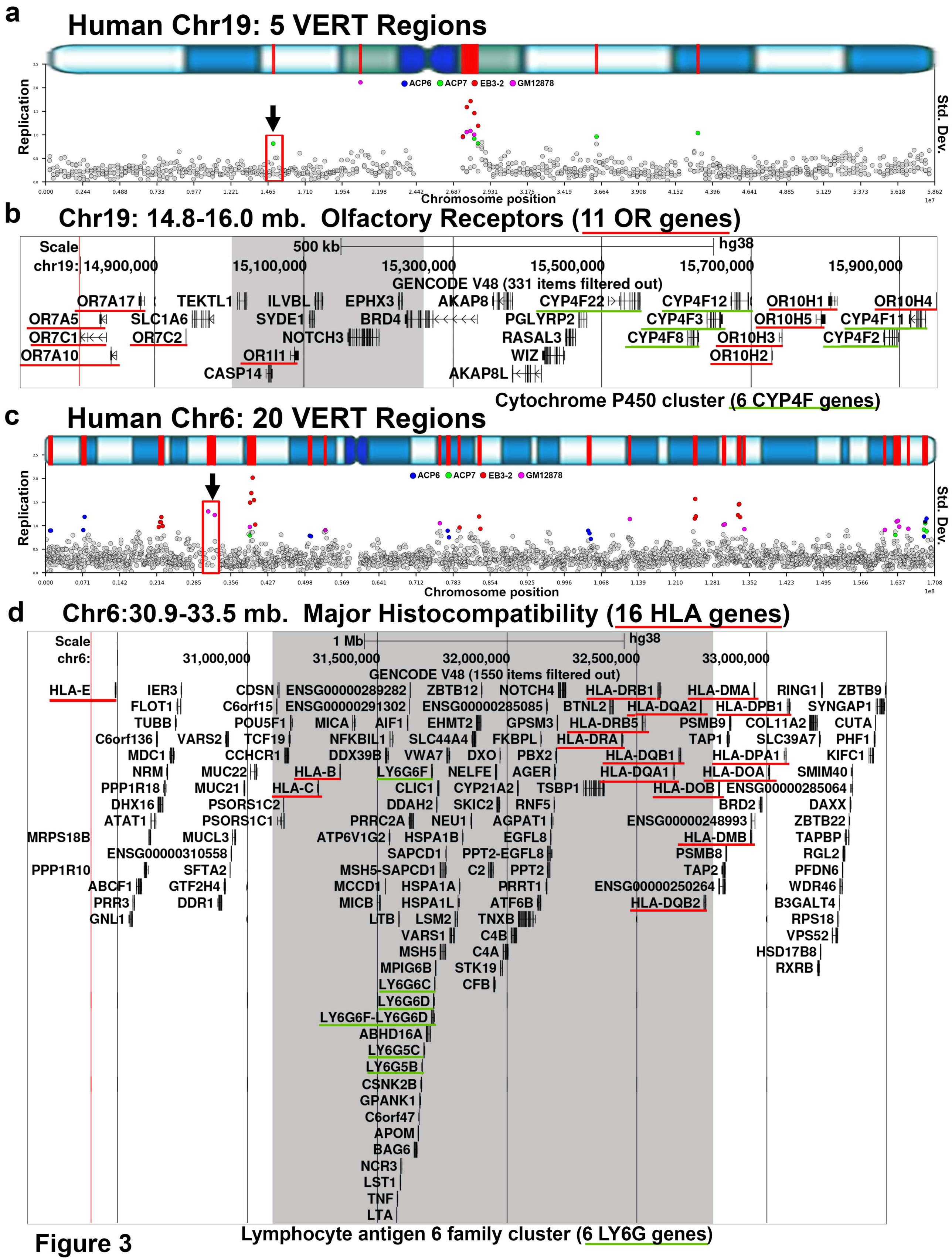

a
Human Chr19: 5 VERT Regions
ACP6 ACP7 EB3-2 GM12878
Replication
Std. Dev.
Chromosome position
b
Chr19: 14.8-16.0 mb. Olfactory Receptors (11 OR genes)
Scale
chr19:
500 kb
hg38
14,900,000
15,100,000
15,300,000
15,500,000
15,700,000
15,900,000
GENCODE V48 (331 items filtered out)
OR7A17
TEKTL1
ILVBL
EPHX3
AKAP8
CYP4F22
CYP4F12
OR10H1
OR10H4
OR7A5
SLC1A6
SYDE1
BRD4
PGLYRP2
CYP4F3
OR10H5
CYP4F11
OR7C1
OR7C2
NOTCH3
RASAL3
CYP4F8
OR10H3
CYP4F2
OR7A10
OR1I1
WIZ
OR10H2
CASP14
AKAP8L
Cytochrome P450 cluster (6 CYP4F genes)
c
Human Chr6: 20 VERT Regions
ACP6 ACP7 EB3-2 GM12878
Replication
Std. Dev.
Chromosome position
d
Chr6:30.9-33.5 mb. Major Histocompatibility (16 HLA genes)
Scale
chr6:
1 Mb
hg38
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
GENCODE V48 (1550 items filtered out)
HLA-E
IER3
CDSN
ENSG00000289282
ZBTB12
NOTCH4
HLA-DRB1
HLA-DMA
RING1
ZBTB9
FLOT1
C6orf15
ENSG00000291302
ENSG00000285085
BTNL2
HLA-DQA2
HLA-DPB1
SYNGAP1
TUBB
POU5F1
MICA
AIF1
EHMT2
GPSM3
HLA-DRB5
PSMB9
COL11A2
CUTA
C6orf136
VARS2
TCF19
NFKBIL1
SLC44A4
FKBPL
HLA-DRA
TAP1
SLC39A7
PHF1
MDC1
CCHCR1
DDX39B
VWA7
DXO
PBX2
HLA-DQB1
HLA-DPA1
KIFC1
NRM
MUC22
HLA-B
LY6G6F
NELFE
AGER
HLA-DQA1
HLA-DOA
SMIM40
PPP1R18
MUC21
HLA-C
CLIC1
CYP21A2
TSBP1
HLA-DOB
ENSG00000285064
DHX16
PSORS1C2
DDAH2
SKIC2
RNF5
BRD2
DAXX
ATAT1
PSORS1C1
PRRC2A
NEU1
AGPAT1
ENSG00000248993
ZBTB22
MRPS18B
MUCL3
ATP6V1G2
HSPA1B
EGFL8
HLA-DMB
TAPBP
ENSG00000310558
SAPCD1
PPT2-EGFL8
PSMB8
RGL2
PPP1R10
SFTA2
MSH5-SAPCD1
C2
PPT2
TAP2
PFDN6
ABCF1
GTF2H4
MCCD1
HSPA1A
PRRT1
ENSG00000250264
WDR46
PRR3
DDR1
MICB
HSPA1L
ATF6B
HLA-DQB2
B3GALT4
GNL1
LTB
LSM2
TNXB
RPS18
VARS1
C4B
VPS52
MSH5
C4A
HSD17B8
MPIG6B
STK19
RXRB
LY6G6C
CFB
LY6G6D
LY6G6F-LY6G6D
ABHD16A
LY6G5C
LY6G5B
CSNK2B
GPANK1
C6orf47
APOM
BAG6
NCR3
LST1
TNF
LTA
Lymphocyte antigen 6 family cluster (6 LY6G genes)

Figure 3

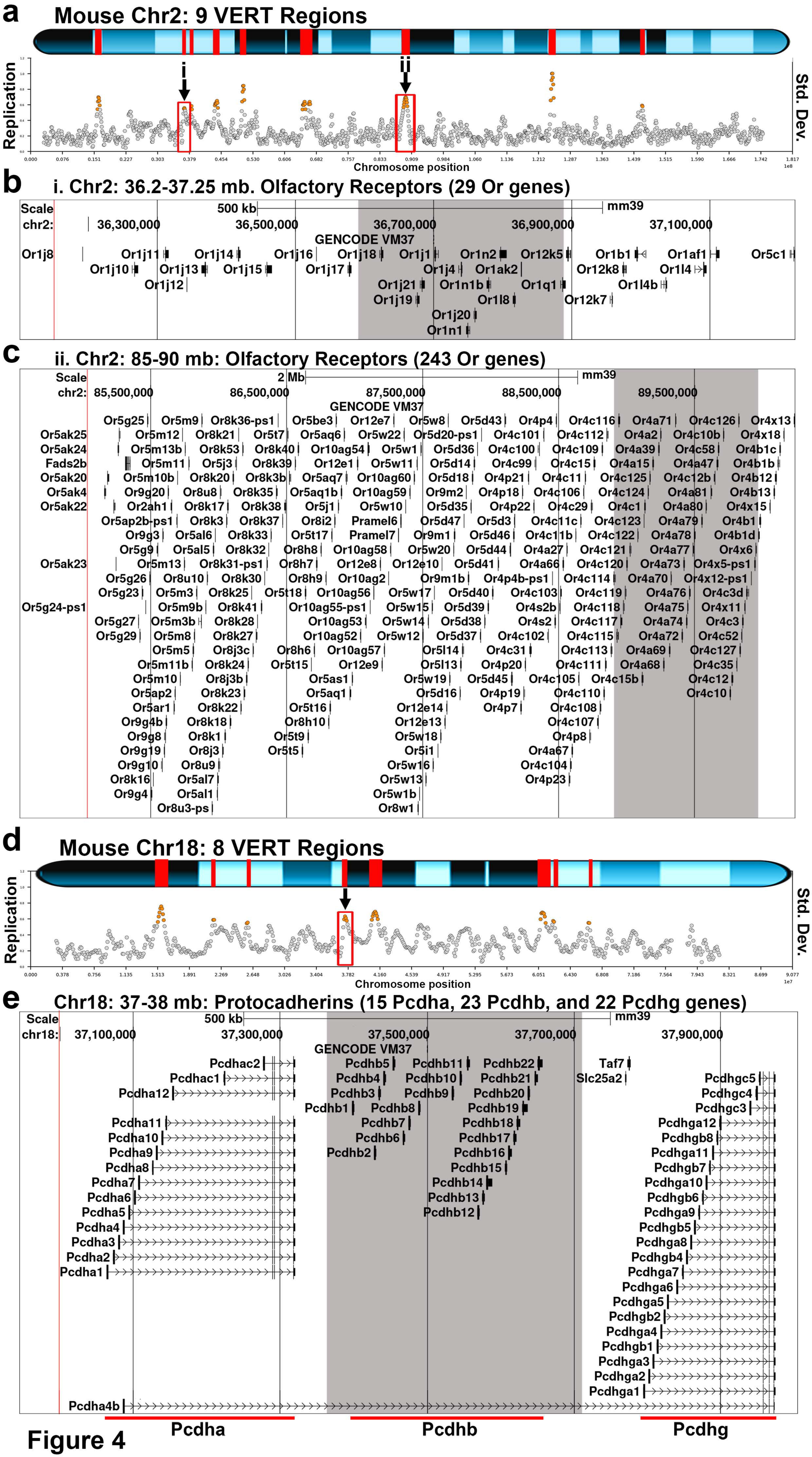

a
Mouse Chr2: 9 VERT Regions
Replication
Std. Dev.
i
ii
Chromosome position
b
i. Chr2: 36.2-37.25 mb. Olfactory Receptors (29 Or genes)
Scale
chr2:
500 kb
mm39
36,300,000
36,500,000
36,700,000
36,900,000
37,100,000
GENCODE VM37
c
ii. Chr2: 85-90 mb: Olfactory Receptors (243 Or genes)
Scale
chr2:
2 Mb
mm39
85,500,000
86,500,000
87,500,000
88,500,000
89,500,000
GENCODE VM37
d
Mouse Chr18: 8 VERT Regions
Replication
Std. Dev.
Chromosome position
e
Chr18: 37-38 mb: Protocadherins (15 Pcdha, 23 Pcdhb, and 22 Pcdhg genes)
Scale
chr18:
500 kb
mm39
37,100,000
37,300,000
37,500,000
37,700,000
37,900,000
GENCODE VM37
Pcdha
Pcdhb
Pcdhg

Figure 4

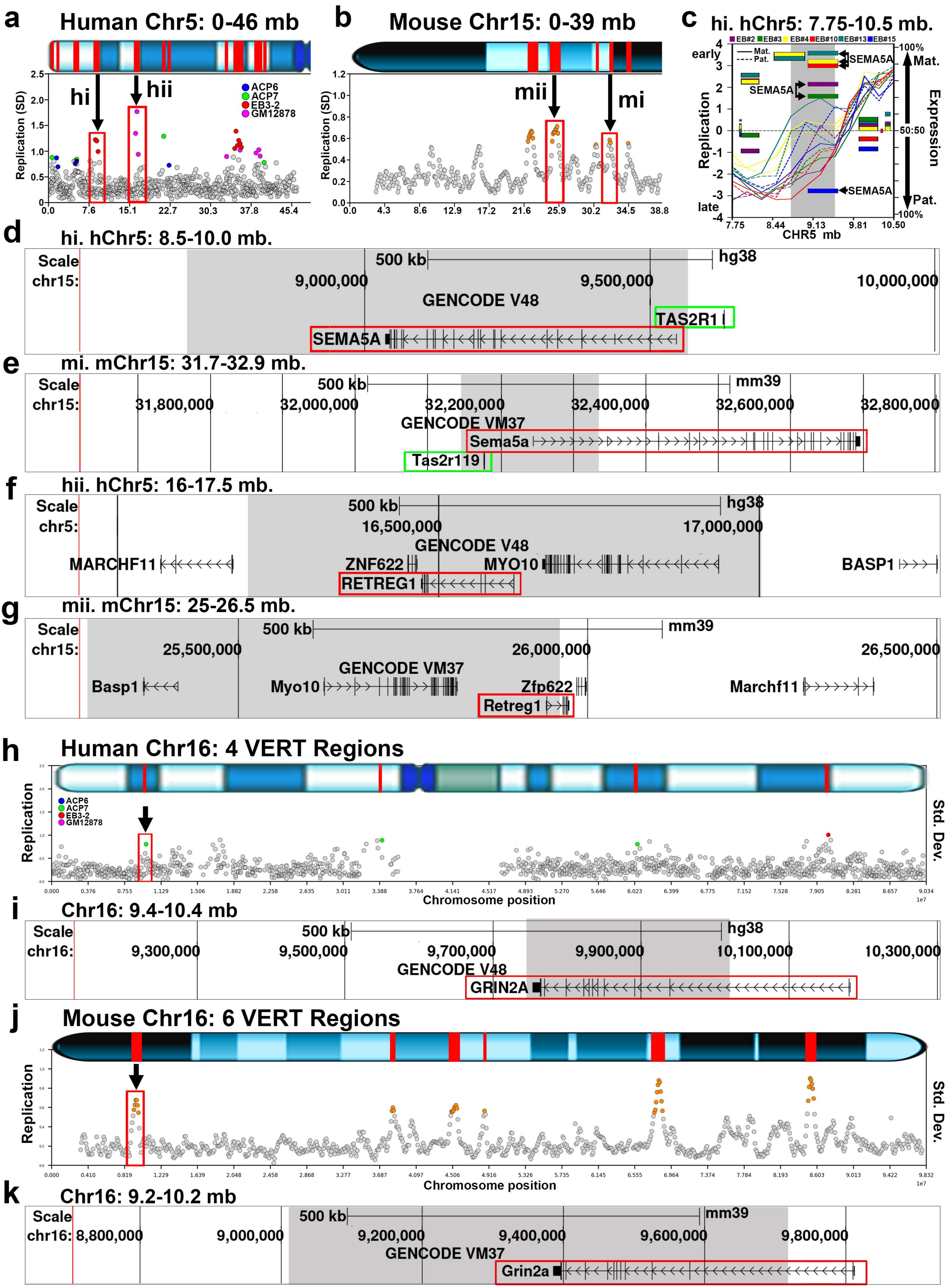

a
Human Chr5: 0-46 mb
hi
hii
ACP6
ACP7
EB3-2
GM12878
Replication (SD)
b
Mouse Chr15: 0-39 mb
mii
mi
Replication (SD)
c
hi. hChr5: 7.75-10.5 mb.
EB#2 EB#3 EB#4 EB#10 EB#13 EB#15
early
late
Mat.
Pat.
SEMA5A
Replication
Expression
50:50
100%
CHR5 mb
d
hi. hChr5: 8.5-10.0 mb.
Scale
chr15:
500 kb
hg38
9,000,000
9,500,000
10,000,000
GENCODE V48
TAS2R1
SEMA5A
e
mi. mChr15: 31.7-32.9 mb.
Scale
chr15:
500 kb
mm39
31,800,000
32,000,000
32,200,000
32,400,000
32,600,000
32,800,000
GENCODE VM37
Sema5a
Tas2r119
f
hii. hChr5: 16-17.5 mb.
Scale
chr5:
500 kb
hg38
16,500,000
17,000,000
GENCODE V48
MARCHF11
ZNF622
MYO10
BASP1
RETREG1
g
mii. mChr15: 25-26.5 mb.
Scale
chr15:
500 kb
mm39
25,500,000
26,000,000
26,500,000
GENCODE VM37
Basp1
Myo10
Zfp622
Marchf11
Retreg1
h
Human Chr16: 4 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
Chromosome position
i
Chr16: 9.4-10.4 mb
Scale
chr16:
500 kb
hg38
9,300,000
9,500,000
9,700,000
9,900,000
10,100,000
10,300,000
GENCODE V48
GRIN2A
j
Mouse Chr16: 6 VERT Regions
Replication
Std. Dev.
Chromosome position
k
Chr16: 9.2-10.2 mb
Scale
chr16:
500 kb
mm39
8,800,000
9,000,000
9,200,000
9,400,000
9,600,000
9,800,000
GENCODE VM37
Grin2a

Figure 5

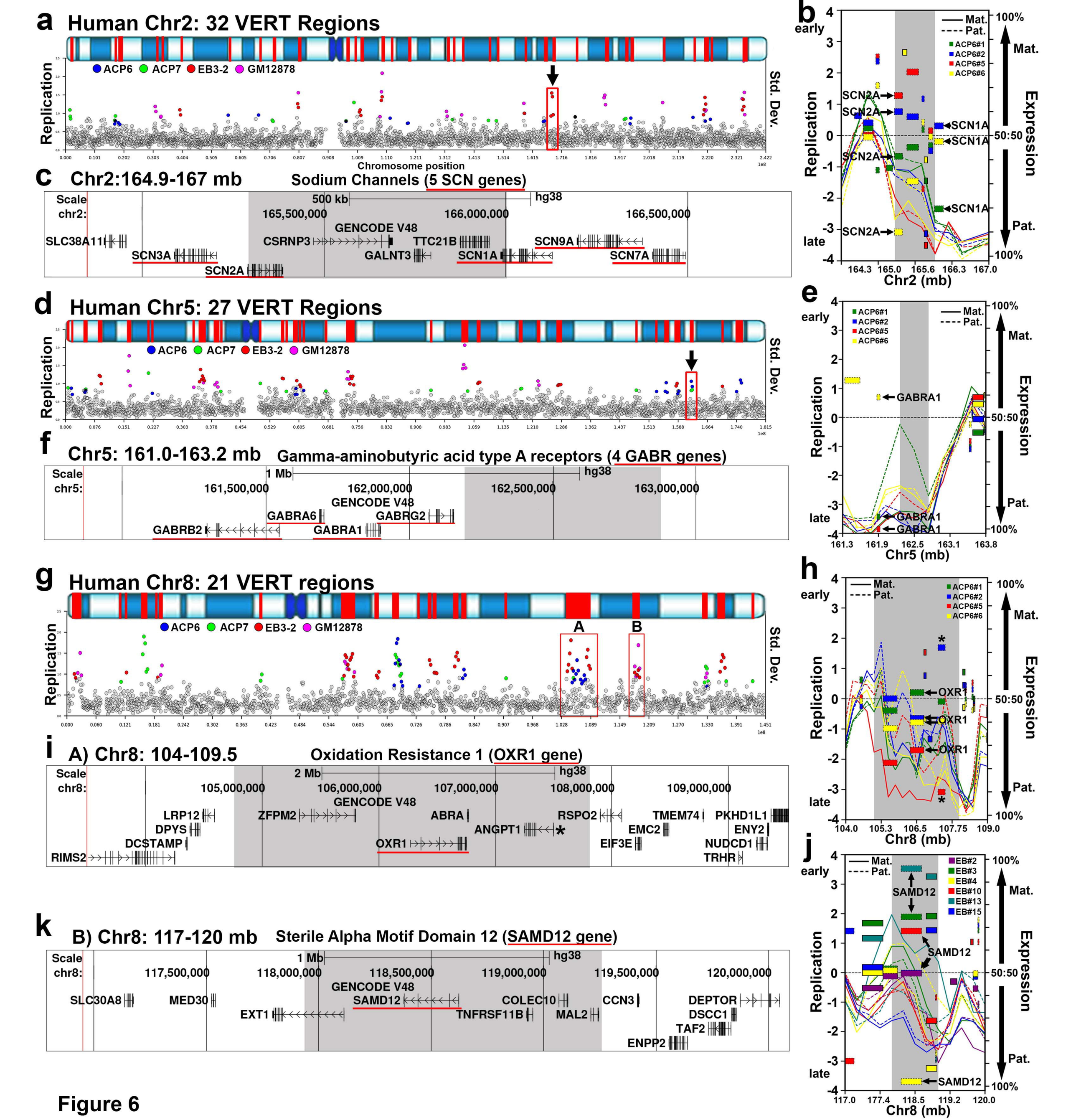

a
Human Chr2: 32 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
Chromosome position
b
early
late
Mat.
Pat.
ACP6#1
ACP6#2
ACP6#5
ACP6#6
SCN2A
SCN1A
100%
50:50
Replication
Expression
Chr2 (mb)
c
Chr2:164.9-167 mb
Sodium Channels (5 SCN genes)
Scale
chr2:
500 kb
hg38
165,500,000
166,000,000
166,500,000
GENCODE V48
SLC38A11
SCN3A
SCN2A
CSRNP3
GALNT3
TTC21B
SCN1A
SCN9A
SCN7A
d
Human Chr5: 27 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
e
early
late
ACP6#1
ACP6#2
ACP6#5
ACP6#6
Mat.
Pat.
GABRA1
100%
50:50
Replication
Expression
Chr5 (mb)
f
Chr5: 161.0-163.2 mb
Gamma-aminobutyric acid type A receptors (4 GABR genes)
Scale
chr5:
1 Mb
hg38
161,500,000
162,000,000
162,500,000
163,000,000
GENCODE V48
GABRA6
GABRG2
GABRB2
GABRA1
g
Human Chr8: 21 VERT regions
ACP6
ACP7
EB3-2
GM12878
A
B
Replication
Std. Dev.
h
early
late
Mat.
Pat.
ACP6#1
ACP6#2
ACP6#5
ACP6#6
OXR1
100%
50:50
Replication
Expression
Chr8 (mb)
i
A) Chr8: 104-109.5
Oxidation Resistance 1 (OXR1 gene)
Scale
chr8:
2 Mb
hg38
105,000,000
106,000,000
107,000,000
108,000,000
109,000,000
GENCODE V48
LRP12
DPYS
DCSTAMP
RIMS2
ZFPM2
OXR1
ABRA
ANGPT1
RSPO2
EIF3E
EMC2
TMEM74
NUDCD1
TRHR
PKHD1L1
ENY2
j
early
late
Mat.
Pat.
EB#2
EB#3
EB#4
EB#10
EB#13
EB#15
SAMD12
100%
50:50
Replication
Expression
Chr8 (mb)
k
B) Chr8: 117-120 mb
Sterile Alpha Motif Domain 12 (SAMD12 gene)
Scale
chr8:
1 Mb
hg38
117,500,000
118,000,000
118,500,000
119,000,000
119,500,000
120,000,000
GENCODE V48
SLC30A8
MED30
EXT1
SAMD12
TNFRSF11B
COLEC10
MAL2
CCN3
ENPP2
DEPTOR
DSCC1
TAF2

Figure 6

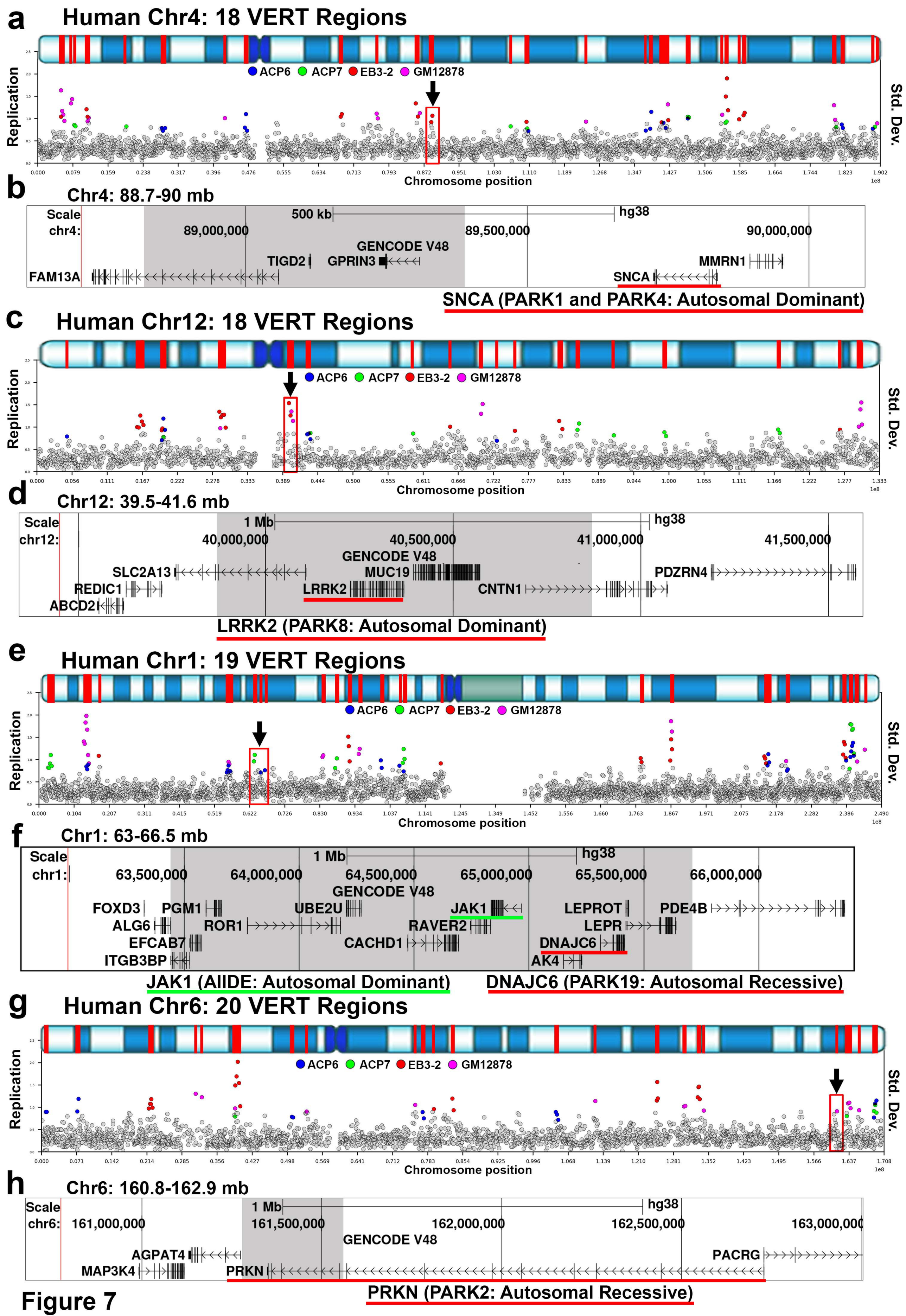

a
Human Chr4: 18 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
Chromosome position
b
Chr4: 88.7-90 mb
Scale
chr4:
500 kb
hg38
89,000,000
89,500,000
90,000,000
GENCODE V48
TIGD2
GPRIN3
MMRN1
FAM13A
SNCA
SNCA (PARK1 and PARK4: Autosomal Dominant)
c
Human Chr12: 18 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
Chromosome position
d
Chr12: 39.5-41.6 mb
Scale
chr12:
1 Mb
hg38
40,000,000
40,500,000
41,000,000
41,500,000
GENCODE V48
SLC2A13
MUC19
PDZRN4
REDIC1
LRRK2
CNTN1
ABCD2
LRRK2 (PARK8: Autosomal Dominant)
e
Human Chr1: 19 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
Chromosome position
f
Chr1: 63-66.5 mb
Scale
chr1:
1 Mb
hg38
63,500,000
64,000,000
64,500,000
65,000,000
65,500,000
66,000,000
GENCODE V48
FOXD3
PGM1
UBE2U
JAK1
LEPROT
PDE4B
ALG6
ROR1
RAVER2
LEPR
EFCAB7
CACHD1
DNAJC6
ITGB3BP
AK4
JAK1 (AIIDE: Autosomal Dominant)
DNAJC6 (PARK19: Autosomal Recessive)
g
Human Chr6: 20 VERT Regions
ACP6
ACP7
EB3-2
GM12878
Replication
Std. Dev.
Chromosome position
h
Chr6: 160.8-162.9 mb
Scale
chr6:
1 Mb
hg38
161,000,000
161,500,000
162,000,000
162,500,000
163,000,000
GENCODE V48
AGPAT4
PACRG
MAP3K4
PRKN
PRKN (PARK2: Autosomal Recessive)

Figure 7