

1 **Title**

2 Somatic instability of the *FGF14-SCA27B* GAA•TTC repeat reveals a marked expansion bias in the cerebellum

3

4 **Authors**

5 David Pellerin, MD^{1,2,3,*}; Jean-Loup Méreaux, MD^{4,*}; Susana Boluda, MD, PhD⁴; Matt C. Danzi, PhD¹; Marie-
6 Josée Dicaire, BSc²; Claire-Sophie Davoine, BSc⁴; David Genis, MD⁵; Guinevere Spurdens, BSc¹; Catherine
7 Ashton, MBBS^{2,6}; Jillian M. Hammond, BSc^{7,8}; Brandon J. Gerhart, MS⁹; Viorica Chelban, MD, PhD^{3,10}; Phuong U.
8 Le, PhD²; Maryam Safisamghabadi, BSc²; Christopher Yanick, BSc¹¹; Hamin Lee, MSc³; Sathiji K. Nageshwaran,
9 MBBS, PhD^{12,13}; Gabriel Matos-Rodrigues, PhD¹⁴; Zane Jaunmuktane, MD^{15,16}; Kevin Petrecca, MD, PhD²;
10 Schahram Akbarian, MD, PhD¹²; André Nussenzweig, PhD¹⁴; Karen Usdin, PhD¹⁷; Mathilde Renaud, MD,
11 PhD^{18,19,20}; Céline Bonnet, MD, PhD^{18,21}; Gianina Ravenscroft, PhD²²; Mario A. Saporta, MD, PhD¹¹; Jill S.
12 Napierala, PhD⁹; Henry Houlden, MD, PhD³; Ira W. Deveson, PhD^{7,8,23}; Marek Napierala, PhD⁹; Alexis Brice,
13 MD⁴; Laura Molina Porcel, MD, PhD^{24,25}; Danielle Seilhean, MD, PhD⁴; Stephan Zuchner, MD, PhD¹; Alexandra
14 Durr, MD, PhD^{4,**}; Bernard Brais, MDCM, PhD^{2,26,**}

15

16 *Contributed equally to this work.

17 **Jointly supervised this work.

18

19 **Affiliations**

20 ¹Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human
21 Genomics, University of Miami Miller School of Medicine, Miami, FL, USA

22 ²Department of Neurology and Neurosurgery, Montreal Neurological Hospital and Institute, McGill University,
23 Montreal, QC, Canada

24 ³Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology and The National Hospital for
25 Neurology and Neurosurgery, University College London, London, United Kingdom

26 ⁴Sorbonne Université, Institut du Cerveau - Paris Brain Institute- ICM, Inserm, CNRS, APHP, University Hospital
27 Pitié-Salpêtrière, Paris, France

28 ⁵Ataxia and Hereditary Spastic Paraplegia Unit, Service of Neurology, Hospital Universitari de Girona Dr. Josep
29 Trueta (ICS) & Hospital Santa Caterina IAS, Institut d'Investigació Biomèdica de Girona (IDIBGI), Girona, Spain

30 ⁶Department of Neurology, Royal Perth Hospital, Perth, WA, Australia

31 ⁷Genomics and Inherited Disease Program, Garvan Institute of Medical Research, Sydney, NSW, Australia

32 ⁸Centre for Population Genomics, Garvan Institute of Medical Research and Murdoch Children's Research Institute,
33 Australia

34 ⁹Department of Neurology, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center,
35 Dallas, TX, USA

36 ¹⁰Neurobiology and Medical Genetics Laboratory, "Nicolae Testemitanu" State University of Medicine and
37 Pharmacy, Chisinau, Republic of Moldova

38 ¹¹Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA

39 ¹²Department of Psychiatry, Department of Neuroscience and Department of Genetics and Genomic Sciences,
40 Friedman Brain Institute Icahn School of Medicine at Mount Sinai, New York, NY, USA

41 ¹³Neurogenetics Program, Department of Neurology, David Geffen School of Medicine, University of California
42 Los Angeles, Los Angeles, CA, USA.

43 ¹⁴Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA

44 ¹⁵Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, University College London
45 NHS Foundation Trust, London, United Kingdom

46 ¹⁶Department of Clinical and Movement Neurosciences and Queen Square Brain Bank for Neurological Disorders,
47 UCL Queen Square Institute of Neurology, University College London, London, United Kingdom

48 ¹⁷Laboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases,
49 National Institutes of Health, Bethesda, MD, USA

50 ¹⁸INSERM-U1256 NGERE, Université de Lorraine, Nancy, France

51 ¹⁹Service de Neurologie, CHRU de Nancy, Nancy, France

52 ²⁰Service de Génétique Clinique, CHRU de Nancy, Nancy, France

53 ²¹Laboratoire de Génétique, CHRU de Nancy, Nancy, France

54 ²²Centre for Medical Research University of Western Australia and Harry Perkins Institute of Medical Research,
55 Perth, Western Australia, Australia

56 ²³Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia

57 ²⁴Alzheimer's Disease and other Cognitive Disorders Unit, Service of Neurology, Hospital Clínic, Fundació de
58 Recerca Clínic Barcelona-Institut d'Investigacions Biomediques August Pi i Sunyer (FRCB-IDIBAPS), University
59 of Barcelona, Barcelona, Spain

60 ²⁵Neurological Tissue Brain Bank, Biobanc-Hospital Clínic-FRCB-IDIBAPS, Barcelona, Spain

61 ²⁶Department of Human Genetics, McGill University, Montreal, QC, Canada

62

63 **Corresponding author**

64 Dr. Bernard Brais, Montreal Neurological Hospital and Institute, McGill University, 3801 University Street,
65 Montreal, QC, H3A 2B4, Canada.

66 Telephone: +1-514-398-3334

67 Email: bernard.brais@mcgill.ca

68

69 **Short title:** Somatic Instability in SCA27B

70

71 **Character count – Title:** 90 characters

72 **Word count – Abstract:** 229 words

73 **Word count – Manuscript:** 4,454 words

74 **Number of references:** 47

75 **Number of figures:** 6

76 **Number of tables:** 1

77 **Abstract**

78 Spinocerebellar ataxia 27B (SCA27B) is a common autosomal dominant ataxia caused by an intronic GAA•TTC
79 repeat expansion in *FGF14*. Neuropathological studies have shown that neuronal loss is largely restricted to the
80 cerebellum. Although the repeat locus is highly unstable during intergenerational transmission, it remains unknown
81 whether it exhibits cerebral mosaicism and progressive instability throughout life. We conducted an analysis of the
82 *FGF14* GAA•TTC repeat somatic instability across 156 serial blood samples from 69 individuals, fibroblasts,
83 induced pluripotent stem cells, and post-mortem brain tissues from six controls and six patients with SCA27B,
84 alongside methylation profiling using targeted long-read sequencing. Peripheral tissues exhibited minimal somatic
85 instability, which did not significantly change over periods of more than 20 years. In post-mortem brains, the
86 GAA•TTC repeat was remarkably stable across all regions, except in the cerebellar hemispheres and vermis. The
87 levels of somatic expansion in the cerebellar hemispheres and vermis were, on average, 3.15 and 2.72 times greater
88 relative to other examined brain regions, respectively. Additionally, levels of somatic expansion in the brain
89 increased with repeat length and tissue expression of *FGF14*. We found no significant difference in methylation of
90 wild-type and expanded *FGF14* alleles in post-mortem cerebellar hemispheres between patients and controls. In
91 conclusion, our study revealed that the *FGF14* GAA•TTC repeat exhibits a cerebellar-specific expansion bias,
92 which may explain the pure and late-onset cerebellar involvement in SCA27B.

93

94 **Keywords:** SCA27B, spinocerebellar ataxia 27B, GAA-FGF14 ataxia, repeat expansion disorder, mosaicism,
95 expansion

96

97 **Introduction**

98 Spinocerebellar ataxia 27B (SCA27B; GAA-*FGF14*-related ataxia) is a recently described autosomal dominant
99 ataxia caused by the expansion of a polymorphic GAA•TTC repeat located within intron 1 of the fibroblast growth
100 factor 14 (*FGF14*) gene.^{1, 2} SCA27B is phenotypically characterized by a late-onset, slowly progressive pan-
101 cerebellar syndrome that is frequently associated with episodic symptoms and downbeat nystagmus.^{1, 3-5} The size of
102 the GAA•TTC repeat expansion only shows a weak negative correlation with the age at onset^{1, 6} and has not been
103 associated with disease severity or progression.^{3, 4} Furthermore, there exists significant phenotypic variability among
104 patients carrying expansions of similar sizes, suggesting that additional factors influence disease expressivity.^{3, 4, 6}
105 Neuropathological studies in SCA27B have shown that neuronal loss is largely restricted to the cerebellar cortex,
106 involving the vermis more than the hemispheres.^{1, 3} However, it is noteworthy that significant pathological
107 abnormalities are not found in other regions despite the widespread expression of *FGF14* in the brain, albeit at lower
108 levels than in the cerebellum.⁷ This observation suggests that tissue-specific factors may increase the vulnerability of
109 the cerebellum in SCA27B.

110
111 Although the expanded *FGF14* GAA•TTC repeat is highly unstable upon intergenerational transmission,^{4, 8} whether
112 it exhibits somatic instability in the brain has not yet been investigated. Somatic mosaicism is a phenomenon
113 observed in many repeat expansions, which have the ability to form alternative (non-B) DNA structures that
114 predispose them to errors during DNA repair.⁹⁻¹¹ The degree of somatic instability generally increases with the
115 length of the repeat tract and age, and also shows a repeat-specific pattern of variation across different tissues and
116 cell types.¹² Remarkably, previous studies on coding CAG•CTG repeat expansions in SCA1, SCA2, SCA3, SCA7,
117 and dentatorubral pallidoluyian atrophy (DRPLA) have shown that these diseases consistently have smaller
118 expansion sizes and lower degrees of mosaicism in the cerebellum compared to other brain regions.¹³⁻¹⁷ In
119 comparison, the intronic ATTTC pentanucleotide repeat insertion in *DABI* exhibits longer sizes and an expansion
120 bias in the cerebellum compared to blood samples and fibroblasts in patients with SCA37.¹⁸ The intronic *RFC1*
121 AAGGG•CCCTT repeat expansion, however, shows little instability in the cerebellum of patients with cerebellar
122 ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS).¹⁹ Furthermore, in Friedreich ataxia, although the
123 expanded intronic *FXN* GAA•TTC repeat is longer in the cerebellum than in other brain regions, large contractions
124 occur significantly more frequently than large expansions.²⁰⁻²³

125

126 Here, to gain further insight into SCA27B biology, we performed a comprehensive analysis of the *FGF14*
127 GAA•TTC repeat somatic instability in serial blood samples, fibroblasts, induced pluripotent stem cells (iPSCs), and
128 post-mortem brains from individuals with SCA27B.

129

130 **Methods**

131 Detailed methods are provided in the Supporting Information.

132

133 *Human sample collection*

134 Human post-mortem brain tissue was collected from six non-ataxic control individuals and six persons with
135 SCA27B to study the somatic stability of the *FGF14* GAA•TTC repeat tract (**Table 1**). Fresh frozen samples were
136 obtained from various regions of the central nervous system (CNS), as detailed in **Supplementary Table 1**.
137 Additionally, DNA extracted from blood and cerebellar hemispheres of another eight control individuals was
138 obtained to compare allele sizes between both tissues. We also collected fresh frozen tissue of the cerebellar
139 hemispheres from an additional four controls and one SCA27B patient to perform methylation profiling of *FGF14*
140 with Oxford Nanopore Technologies (ONT), as detailed in the Supplementary Methods.

141

142 *Analysis of somatic instability*

143 Somatic stability was studied by capillary electrophoresis or by agarose gel electrophoresis for samples carrying
144 alleles >400-450 triplets. For samples analyzed by capillary electrophoresis, the GeneMapper software v6.0
145 (Applied Biosystems) or Peak Scanner software v1.0 (Applied Biosystems) was used to establish the ‘modal peak’
146 of each of the two alleles in each sample, which was defined as the peak with the highest height. Peaks to the right
147 of the modal allele represent somatically expanded GAA•TTC repeats, while peaks to the left may include both
148 contracted GAA•TTC repeats and PCR stutter products. Expansion indices (EI) were calculated by only taking into
149 account peaks to the right of the ‘modal allele’ and using a 1% relative peak threshold (**Figure 1**), as described
150 previously.^{13, 24} The EI was obtained by first dividing the height of each individual peak by the sum of the heights of
151 all peaks above the threshold. The resulting normalized peak heights were then multiplied by the position of the
152 peak and these normalized peak values were summed to generate the EI. Analysis of blood and CNS tissues revealed

153 that the modal length was remarkably stable across all regions examined, except in the cerebellar hemispheres and
154 vermis where it showed a tendency to expand. We therefore considered the modal allele length as measured in the
155 blood and most CNS regions as the constitutional allele length, which we set as the main allele for calculation of the
156 EI in tissues with an expansion shift of the modal repeat,¹³ tolerating a discrepancy of one triplet due to the
157 variability of our genotyping assay.²⁵ Finally, we correlated the somatic stability indices with the expression of
158 *FGF14* in normal tissue data from the Human Protein Atlas (HPA v.23.0; proteintatlas.org accessed 27/03/2024).²⁶

159

160 *Ethics*

161 The institutional review boards of the Montreal Neurological Hospital, Montreal (MPE-CUSM-15-915), the Centre
162 Hospitalier de l'Université de Montréal, Montreal (ND02.045), the Dr. Josep Trueta University Hospital, Girona
163 (2022.120), the Hôpital Pitié-Salpêtrière, Paris (SPATAX RBM 01-29 and RBM 03-48, BIOMOV NCT05034172),
164 and the University College London Hospitals, London (04/N034) approved this study. All brain donors signed an
165 informed consent for brain tissue usage in medical research. The study complied with all relevant ethical
166 regulations.

167

168 *Data availability*

169 The data supporting this study may be shared at the request of any qualified investigator on reasonable request.

170

171 **Results**

172 *Longitudinal analysis of the FGF14 GAA•TTC repeat somatic instability in blood samples*

173 We performed a longitudinal analysis to determine whether the length of the *FGF14* repeat tracts changes over time
174 in 156 serial blood samples from 69 individuals (see Supplementary Results). Alleles longer than \approx 400-450 triplets
175 could not be measured by capillary electrophoresis and were not included in this analysis. The median interval
176 between the first and last blood sample was 8.9 years (interquartile range [IQR]: 2.9 to 13.9). We observed no
177 significant change of the modal *FGF14* repeat size between the first and last blood samples (**Figure 2a**; median
178 difference [size of last sample – size of first sample], 0 triplet; IQR: 0 to 0; Sign test, $p=0.22$), and found that elapsed
179 time had no significant effect on the modal repeat size as assessed by a linear mixed-effects model (estimate = -
180 0.441, standard error [SE] = 0.678, $p=0.52$).

181
182 As for repeat sizes, the EI did not significantly change over time in repeated blood samples (**Figure 2b-c**, median of
183 differences [EI of last sample – EI of first sample], 0 [IQR: -0.02 to 0.01]; Sign test, $p=0.22$), nor in the subset of
184 alleles longer than 200 triplets (median of differences, 0 [IQR: -0.28 to 0.43]), Sign test, $p=1$). In a linear mixed-
185 effects model, we found that allele size (estimate = $5.530e-03$, SE = $3.253e-04$, $p<0.0001$) had a significant effect on
186 the EI, while elapsed time (estimate = $-7.634e-03$, SE = $6.945e-03$, $p=0.27$), the interaction between allele size and
187 elapsed time (estimate = $9.008e-05$, SE = $4.839e-05$, $p=0.064$), and sex (estimate = $-5.148e-02$, SE = $9.654e-02$,
188 $p=0.60$) did not. Regression analysis conducted on 173 alleles from 100 individuals (the EI of alleles longer than
189 ≈ 400 -450 triplets could not be adequately determined by capillary electrophoresis and were not included in this
190 analysis) confirmed a linear increase in somatic expansion with *FGF14* GAA•TTC repeat size ($r=0.85$ [95%
191 confidence interval: 0.80 to 0.88], slope = 0.007, $p<0.0001$) (**Figure 2d**). Furthermore, we compared the size and EI
192 of the *FGF14* repeats in two pairs of monozygotic twins with SCA27B who had blood samples drawn at the age of
193 65-70 and 60-65 years, respectively. We observed a difference of a single triplet between them (15/225 vs 15/226
194 and 9/427 vs 9/428 triplets, respectively), which falls within the expected range of variability of our genotyping
195 assay.²⁵ The short alleles of both pairs of twins did not exhibit somatic expansion (EI=0), while the EI of the long
196 allele of the first pair of twins differed by only 0.5 (EI = 1.76 vs 1.26; the EI of the long allele of the second pair of
197 twins could not be calculated due to their large size). Overall, our results demonstrate that *FGF14* GAA•TTC repeat
198 size and somatic expansion in blood samples remain largely stable over periods of more than 20 years.

199 200 *Instability of the FGF14 GAA•TTC repeat tract in fibroblasts and induced pluripotent stem cells*

201 To further assess somatic instability of the *FGF14* GAA•TTC repeat in peripheral tissues, we analyzed allele sizes
202 in paired blood and fibroblasts obtained from three patients with SCA27B. These patients respectively carried alleles
203 of 292/304 (patient P1), 9/508 (patient P8), and 16/389 (patient P9) repeat units, as measured in blood samples. We
204 found minimal difference in the number of GAA•TTC repeats and EI between paired blood and fibroblasts,
205 providing further evidence for the lack of significant instability of the *FGF14* repeat locus in peripheral tissues
206 (**Supplementary Figures 4-5 and Supplementary Table 2**). Moreover, the sizes of both wild type and expanded
207 alleles remained unchanged across ten passages in fibroblasts from two SCA27B patients (patient P8: 9/518
208 GAA•TTC repeats; patient P9: 16/409 GAA•TTC repeats).

209
210 Reprogramming of fibroblasts into iPSCs resulted in variable increase of the modal peak for most expanded alleles,
211 but not wild type alleles (**Supplementary Figures 4-5 and Supplementary Table 2**). While two iPSC clones from
212 patient P1 exhibited minimal changes in modal alleles compared to fibroblasts, one clone harbored an expansion of
213 11 triplets of its longer allele, corresponding to an increase in length of 3.6% (**Supplementary Figure 6**). During
214 reprogramming, the longer allele in all four iPSC clones from patient P9 remained relatively stable (modal
215 difference ranging from -19 to +2 triplets; **Supplementary Figure 4**), while it expanded by 7 to 89 triplets in iPSC
216 clones from patient P8 (**Supplementary Figure 5**). Finally, similar to fibroblasts, the sizes of alleles in control ($n=3$;
217 alleles of 8/8, 9/58, and 12/138 repeat) and patient ($n=3$; alleles of 291/315, 9/597, 16/402) iPSCs remained
218 unchanged across serial passages (**Supplementary Figure 7**).

219
220 *Analysis of the FGF14 GAA•TTC repeat somatic instability in post-mortem brain tissues*
221 Because transcription through GAA•TTC repeat tracts enhances somatic instability,²⁷ the very low expression levels
222 of *FGF14* in blood and fibroblasts¹ may account for the minimal instability of the repeat in these tissues. In
223 comparison, *FGF14* is widely expressed in the CNS, with the highest expression levels observed in the cerebellum.⁷
224 Specifically, *FGF14* isoform 1b (*FGF14*-201, transcript 2; ENST00000376131.9; NM_175929.3), within which the
225 GAA•TTC repeat is located in its first intron, has notably higher expression levels in the cerebellum compared to
226 isoform 1a (*FGF14*-202, transcript 1; ENST00000376143.5; NM_004115.4) according to the Genotype-Tissue
227 Expression (GTEx) Project database. We therefore hypothesized that higher degrees of somatic instability may be
228 present in the CNS. To address this question, we conducted a comprehensive analysis of the somatic instability of
229 *FGF14* GAA•TTC repeat tracts in multiple post-mortem CNS regions from six non-ataxic controls and six patients
230 with SCA27B (**Table 1**).

231
232 In controls C1-C6 and patients P1-P3, all of whom carried alleles smaller than 400 GAA•TTC triplets, analysis of
233 tissue-wide somatic instability in multiple brain regions revealed that the modal length and EI of both wild type and
234 expanded alleles were remarkably stable across all analyzed regions, except in the cerebellar hemispheres and
235 vermis, which showed a marked somatic expansion bias (**Figures 3 and 4a**; see also Supplementary Results). In
236 control and patient brains, the modal length of wild type and expanded alleles was identical – within two triplets –

237 across all non-cerebellar regions and blood samples, while, for alleles longer than ≈ 90 repeats, it increased on
238 average by $10 \pm 6\%$ in the vermis and $11 \pm 6\%$ in the cerebellar hemispheres (**Figure 3a** and **Supplementary**
239 **Figures 17-25**). Levels of somatic expansions, as measured by the EI, were markedly increased in the cerebellar
240 hemispheres and vermis compared to other CNS regions (**Figure 4a**). Excluding $(GAA)_9$ alleles, which did not
241 exhibit somatic expansion, the EI in the cerebellar hemispheres was, on average, 3.23 times greater relative to other
242 CNS regions outside the cerebellum. Furthermore, within the cerebellum itself, the EI in the hemisphere was 1.15
243 times greater relative to the vermis and 2.42 times greater relative to the dentate nucleus. We found that increasing
244 allele length was associated with higher degrees of instability of the *FGF14* repeat locus, with instability even
245 observed in alleles as short as 15 triplets, indicating that somatic instability is not exclusive to large, pathogenic
246 *FGF14* alleles (**Supplementary Figure 26**). The pattern of somatic expansion in the CNS was closely correlated
247 with the regional expression of *FGF14*, which is highest in the cerebellar hemispheres and vermis (**Figure 4**). This
248 observation supports the hypothesis that higher levels of transcription through $GAA \cdot TTC$ repeat tracts promote their
249 somatic expansion.²⁷ Furthermore, we found a positive correlation between the EI and the repeat length in many of
250 the examined tissues, with the steepest regression slopes observed in the vermis ($r=0.96$ [95% CI: 0.81 to 0.99],
251 slope=0.130, $p<0.0001$) and the cerebellar hemispheres ($r=0.93$ [95% CI: 0.81 to 0.98], slope=0.133, $p<0.0001$)
252 (**Supplementary Figure 26**). Using a multiple regression model, we found significant relationships between EI and
253 repeat size (estimate = 0.047, SE = 0.005, $p<0.0001$), tissue expression (estimate = 0.133, SE = 0.026, $p<0.0001$),
254 and female sex (estimate = 6.630, SE = 1.744, $p=0.00022$), although the latter association should be interpreted
255 cautiously as it was based on two female cases. In contrast, age at death did not have a significant effect on EI
256 (estimate = 0.040, SE = 0.061, $p=0.51$). These variables together explained approximately 58% of the variance in EI.
257 Overall, our results demonstrate that the *FGF14* repeat locus exhibits a strong somatic expansion bias largely
258 restricted to the cerebellum, that may in part be driven by the high level of *FGF14* expression in this tissue. This
259 observation contrasts with a recent study on SCA1, SCA2, SCA3, and SCA7, which showed that despite high
260 *ATXN1*, *ATXN2*, *ATXN3*, and *ATXN7* gene expression, the cerebellum exhibited the lowest degree of somatic
261 mosaicism.¹⁴

262

263 We also studied the somatic instability of the *FGF14* repeats by agarose gel electrophoresis in the post-mortem
264 brains of three patients with SCA27B (P4-P6), each carrying expansions longer than 400 triplets. This analysis

265 confirmed the cerebellar-specific somatic expansion bias of the *FGF14* repeat locus. Compared to other tissues, the
266 expansion length in the cerebellar hemispheres was 87-159 triplets longer in patient P4, 129-185 triplets longer in
267 patient P5, and 70-168 triplets longer in patient P6 (**Figure 5a, Supplementary Figures 53-55**). Furthermore,
268 compared to blood samples, this corresponds to an increase in expansion length in the cerebellar hemispheres of
269 26% in patient P4 and 41% in patient P6 (**Figure 5b**). Among the three patients, cerebellar vermis was only
270 available for study in patient P4 and exhibited a greater expansion length compared to the cerebellar hemispheres
271 (709 vs. 692 GAA•TTC repeats). Of note, in keeping with the greater levels of *FGF14* expression in cerebellar
272 neurons compared to glia,²⁶ the cerebellar hemispheres and vermis showed longer expansion sizes relative to the
273 cerebellar white matter in all three patients. Finally, although the cervical spinal cord exhibited the largest expansion
274 size of all tissues examined from patient P6, this finding was not replicated in the two other patients (P1 and P5) for
275 whom cervical spinal cord was available for study. This discrepancy could be due to differences in the specific
276 regions of the spinal cord sampled for analysis.

277
278 Comparison of modal lengths in blood samples and cerebellar hemispheres from five patients with SCA27B (P1, P2,
279 P3, P4, P6) and eight control individuals (C7-C14) revealed that sizes were largely stable in both tissues for alleles
280 shorter than ≈ 90 repeat units. However, for longer alleles, the modal length in the cerebellar hemispheres increased
281 with allele size ($r=0.99$ [95% CI: 0.96 to 1.00], slope=1.365, $p<0.0001$) (**Figure 5b**). Remarkably, this threshold of
282 ≈ 90 repeat units for instability in the cerebellar hemispheres is highly similar to the threshold of 75-100 repeat units
283 for instability during intergenerational transmission.^{4, 8} Despite the similar instability thresholds, different
284 mechanisms are likely responsible for instability in the cerebellum and the germline, given the comparatively lower
285 *FGF14* expression in testes and ovaries and the repeat contraction observed with paternal transmission.^{4, 8}

286 287 *Methylation profiling of FGF14 in post-mortem cerebellum*

288 Preliminary investigations of patient-derived post-mortem cerebellum and iPSC-derived neurons have suggested that
289 the intronic GAA•TTC repeat expansion likely leads to loss of function of isoform 1b (transcript 2) by interfering
290 with *FGF14* transcription.¹ Since the *FGF14* expansion shares the same repeat motif as the Friedreich ataxia's *FXN*
291 expansion (GAA•TTC), which is known to cause hypermethylation of the DNA upstream of the expanded repeat
292 locus (differentially methylated region),²⁸ we investigated whether altered methylation was present in post-mortem

293 cerebellum from four patients with SCA27B. Using programmable targeted ONT long-read sequencing,²⁹ we found
294 no evidence of significant difference in 5mC methylation frequencies of *FGF14* (median frequency [IQR] in
295 patients vs controls, 69.75 [68.65 to 69.80] vs 68.95 [67.43 to 69.80]; Mann-Whitney U test, $p=0.63$) and its two
296 putative promoters (first promoter: 2.00 [1.35 to 2.27] vs 2.25 [1.40 to 3.25]; Mann-Whitney U test, $p=0.54$; and
297 second promoter: 0.85 [0.57 to 1.35] vs 0.80 [0.40 to 0.90]; Mann-Whitney U test, $p=0.63$) in controls compared to
298 patients (**Figure 6**). The region surrounding the repeat expansion (10 kb window centred on the repeat locus)
299 showed slightly higher CpG methylation compared to the region surrounding wild type alleles, although the
300 difference was not statistically significant (median frequency [IQR] in expanded alleles vs wild type alleles, 82.37
301 [79.96 to 84.76] vs 78.78 [76.66 to 80.55]; Mann-Whitney U test, $p=0.11$). These results do not exclude the
302 possibility that the expansion is associated with hypermethylation of *FGF14*, which will require further study in a
303 larger number of cases.

304

305 **Discussion**

306 This study comprehensively assessed the somatic instability of the *FGF14* GAA•TTC repeat in peripheral tissues
307 and post-mortem brains of patients with SCA27B and uncovered a marked tendency for the repeat to somatically
308 expand in the cerebellar vermis and hemispheres. Our findings provide new insight into the biology of SCA27B,
309 potentially explaining the selective vulnerability of the cerebellum – specifically the gray matter – and the relative
310 resilience of other regions in this disease.^{1, 3} Although somatic expansion events were detected in all tissues
311 analyzed, the expansion bias was considerably higher in the cerebellar hemispheres and vermis, which are the
312 primary sites of neuronal loss in SCA27B.

313 Although brain mosaicism has only been studied in a few spinocerebellar ataxias, previous studies of SCA1, SCA2,
314 SCA3, SCA7, and DRPLA have consistently shown that their causative coding CAG•CTG repeat expansions exhibit
315 smaller sizes and lower degrees of mosaicism in the cerebellum compared to other brain regions.¹³⁻¹⁷ Studies of
316 Huntington disease, myotonic dystrophy type 1, and spinal and bulbar muscular atrophy, all caused by CAG•CTG
317 repeat expansions, have also shown that the smallest expansions and lowest somatic instability occur in the
318 cerebellum.^{13, 30-32} In comparison, a recent study found that the pathogenic *DABI* intronic ATTTC repeat insertion
319 causing SCA37 exhibits longer sizes and greater somatic expansion in the cerebellum compared to blood samples

320 and fibroblasts of two patients,¹⁸ suggesting that the instability pattern of CAG•CTG repeat expansions in the
321 cerebellum may differ from that observed with other types of repeats. The factors underlying the distinct behavior of
322 the *FGF14* GAA•TTC repeat expansion in the cerebellum compared to CAG•CTG repeat expansions remain
323 unknown, although they may involve *cis*-acting elements, differential effects of mismatch repair pathways on
324 CAG•CTG repeats compared to GAA•TTC repeats, or an increased ability of the long and uninterrupted *FGF14*
325 GAA•TTC repeat to form alternative DNA structures that can become substrates for DNA repair machineries.³³

326 In Friedreich ataxia, although the expanded *FXN* GAA•TTC repeat is longer in the cerebellum than in other CNS
327 regions, large contractions occur significantly more frequently than large expansions in the cerebellum.²⁰⁻²³
328 Progressive expansion of the *FXN* repeat over time has been described in human cells, cerebellum, and other
329 terminally differentiated tissues in both human and mouse models.^{20-23, 34, 35} DNA repair has been established as a
330 major factor modulating GAA•TTC repeat stability.^{9, 33} Studies in a nondividing quiescent *S. cerevisiae* model have
331 shown that large contractions of the GAA•TTC repeat expansion may be driven by non-homologous end joining
332 repair of DNA breaks of the repeat mediated by mismatch repair complexes.³⁶ Furthermore, activity of the mismatch
333 repair complex MutSβ has been shown to promote *FXN* GAA•TTC expansion in Friedreich ataxia patient-derived
334 cells.^{37, 38} Similarly, DNA repair mechanisms may also be involved in somatic expansion of the *FGF14* GAA•TTC
335 repeat, especially considering the high expression levels of mismatch repair genes in the cerebellum compared to
336 other brain regions.¹⁴ Moreover, 5' GAA-strand and 5' TTC-strand nicks (single-strand breaks) have recently been
337 shown to induce expansion of GAA•TTC repeats in non-dividing yeast cells,³⁹ raising the hypothesis that such
338 phenomenon may contribute to somatic instability in post-mitotic neurons, which are known to accumulate single-
339 strand breaks over time.⁴⁰

340 Our findings suggest that tissue-specific transcriptional activity through the GAA•TTC repeat promotes its
341 instability, which align with a previous study on GAA•TTC repeat stability in human cell lines.²⁷ The GAA•TTC
342 repeat is located within intron 1 of *FGF14* isoform 1b, the isoform predominantly expressed in the CNS, particularly
343 in the cerebellum.⁷ Analysis of tissue-wide somatic instability revealed that the pattern of somatic expansion in the
344 CNS closely correlates with the regional expression of *FGF14*. The cerebellar vermis and hemispheres, which have
345 the highest and second highest levels of *FGF14* expression in the CNS, respectively, exhibited the greatest somatic
346 expansion across all tissues analyzed. Moreover, the association between *FGF14* repeat instability and expression

347 levels is further corroborated by our results showing longer expansions in the cerebellar hemispheres and vermis
348 compared to the cerebellar white matter, in keeping with the substantially greater expression of *FGF14* in cerebellar
349 neurons relative to glia.²⁶ Alternative DNA structures called DNA triplexes are formed in expanded GAA•TTC
350 repeats in Friedreich ataxia patient-derived cells,⁴¹ and transcription through these repeats has been shown to induce
351 the formation of triplexes, which may fuel repeat instability.^{9, 42} This observation contrasts with a recent study of the
352 somatic instability of CAG•CTG repeats, which are prone to form DNA hairpins,⁴³ but not DNA triplexes, in the
353 polyglutamine spinocerebellar ataxias SCA1, SCA2, SCA3, and SCA7. These pathogenic CAG•CTG repeats
354 showed lower levels of expansion in the cerebellum despite high expression levels of the *ATXN1*, *ATXN2*, *ATXN3*,
355 and *ATXN7* genes.¹⁴ The absence of significant somatic instability of the CAG•CTG repeat in the cerebellum, which
356 is one of the most pathologically affected brain regions in these spinocerebellar ataxias,⁴⁴ further supports the
357 possibility that different alternative DNA structures and various repair pathways may differentially impact the
358 stability of CAG•CTG repeats compared to GAA•TTC repeats in the cerebellum.

359 Although we observed no change in repeat size and EI in blood samples over time, somatic expansions of the
360 *FGF14* mutant alleles in the cerebellar cortical neurons are likely to progress with age, with the rate of expansion
361 increasing with the number of repeats. Indeed, we found that increasing allele length was associated with higher
362 degrees of instability of the *FGF14* GAA•TTC repeat locus. While our study did not look at mosaicism in fetal
363 brains, somatic instability is likely to progress throughout life rather than arise during CNS development (as
364 previously shown in Friedreich ataxia²⁰ and Huntington disease,⁴⁵ for example) as regions with the same
365 embryological origin, such as the cerebellar cortex, dentate nuclei, and pons, all deriving from the metencephalon,⁴⁶
366 present dramatically different instability patterns. This hypothesis would also be consistent with the late age at onset
367 of SCA27B, suggesting that the disease may only manifest when a critical proportion of cerebellar Purkinje cells
368 accumulate expansions above a cell-specific pathogenic threshold, which is likely to be higher than the currently
369 used 'blood-based threshold'.⁴⁷ It also raises the possibility that loss of function of *FGF14* isoform 1b may only
370 occur when repeat sizes extend beyond this specific threshold in cerebellar Purkinje cells. This hypothesis also
371 provides a potential explanation for the incomplete penetrance of certain *FGF14* expansions, suggesting that
372 asymptomatic carriers may accumulate somatic expansions in the cerebellum at lower rates compared to
373 symptomatic individuals, preventing them from accumulating a sufficiently large burden of somatic expansions to

374 manifest the disease. Furthermore, variable somatic expansion, which may be influenced by variations in DNA
375 repair genes, may underlie the significant phenotypic variability among patients carrying expansions of similar sizes.

376 Our study has several strengths, including the inclusion of a large number of post-mortem controls and SCA27B
377 cases with a variety of allele sizes, the availability of serial blood samples for many patients over long periods of
378 time, and the ability to accurately size and calculate expansion indices in samples carrying expansions shorter than
379 ≈ 400 -450 triplets. However, our results need to be interpreted in light of some limitations. The inability of capillary
380 electrophoresis to resolve low-abundance alleles and alleles longer than ≈ 400 -450 triplets prevented a
381 comprehensive analysis of the full spectrum of allele length variability in the cerebellum. Analysis of bulk PCR
382 products by capillary electrophoresis will preferentially detect shorter GAA•TTC tracts and cannot capture the full
383 extent of GAA•TTC repeat expansions in any given tissue. Furthermore, our analyses did not permit the detection of
384 rare, very large expansion events that may be occurring in Purkinje cells, which are primarily affected in SCA27B.
385 Purkinje cells make up only a small fraction of the total cerebellar cell population and, as such, do not contribute
386 significantly to the instability readout measured in bulk tissue. Additional techniques, such as single-cell sequencing,
387 will be necessary to dissect cell-specific somatic instability in SCA27B. Moreover, in post-mortem tissues from
388 patients with neurodegenerative ataxia, it is unknown whether the genotype of degenerated cerebellar neurons is
389 similar to that of the remaining cells. However, it is possible that neurons carrying longer expansions degenerate
390 faster, leading to an underrepresentation of longer allele sizes in the instability readout.

391 In conclusion, our analysis of the *FGF14* GAA•TTC repeat somatic instability in post-mortem brain samples from
392 individuals with SCA27B revealed a marked expansion bias in the cerebellar hemispheres and vermis. Our study
393 offers new insights into the biology of SCA27B by providing a potential explanation for the significant phenotypic
394 variability and the incomplete penetrance of certain expansions. It also opens up important new lines of inquiry for
395 understanding the mechanisms driving somatic instability of tandem repeats, the cell-specific patterns of instability,
396 and their relationship to neurodegeneration.

397 **Acknowledgments**

398 We thank the patients and their families for participating in this study. Part of this work was carried out in the DNA
399 and cell bank of the Paris Brain Institute - ICM. We gratefully acknowledge Ludmila Jornea and Philippe Martin
400 Hardy for their help as well as Sabrina Leclère-Turbant from Neuro-CEB. We are indebted to the Biobanc-Hospital
401 Clinic-FRCB-IDIBAPS for samples and data procurement.

402

403 **Study sponsorship and funding**

404 This work was supported by the Canadian Institutes of Health Research (grant 189963 to BB), the Fondation Groupe
405 Monaco (to BB), and the NIH National Institutes of Neurological Disorders and Stroke (grant 2R01NS072248-11A1
406 to SZ). VC was supported by the Association of British Neurologists' Academic Clinical Training Research
407 Fellowship (grant ABN 540868) and the Guarantors of Brain (grant 565908). VC and HH received grants from the
408 MSA Trust, MSA Coalition, MANX MSA, King Baudouin Foundation, the National Institute for Health Research
409 (NIHR) University College London Hospitals Biomedical Research Centre, Michael J Fox Foundation (MJFF),
410 Fidelity Trust, Rosetrees Trust, Ataxia UK, Alzheimer's Research UK (ARUK), NIH NeuroBioBank, and MRC
411 Brainbank Network. HH is supported by the Wellcome Trust, the UK Medical Research Council (MRC), and by the
412 UCLH/UCL Biomedical Research Centre. SA is supported by the NIDA (grant DP1DA056018). AN is supported by
413 the Intramural Research Program of the NIH funded in part with federal funds from the NCI under contract
414 HHSN261201500003. GR is supported by an EL2 Investigator Grant (APP2007769) from the Australian National
415 Health and Medical Research Council (NHMRC). IWD holds a fellowship from the Australian Medical Research
416 Future Fund (MRFF, 1173594). MN and JSN are supported by Friedreich's Ataxia Research Alliance and NIH
417 National Institutes of Neurological Disorders and Stroke (grant NS081366). DP holds a Fellowship award from the
418 Canadian Institutes of Health Research and JLM holds a fellowship from the Fondation pour la Recherche Médicale
419 (grant 13338). The funders had no role in the conduct of this study.

420

421 **Declaration of interests**

422 DP reports no disclosures.

423 JLM reports no disclosures.

424 SB reports no disclosures.
425 MCD reports no disclosures.
426 MJD reports no disclosures.
427 CSD reports no disclosures.
428 DG reports no disclosures.
429 GS reports no disclosures.
430 CA reports no disclosures.
431 JMH reports no disclosures.
432 BJG reports no disclosures.
433 VC reports no disclosures.
434 PUL reports no disclosures.
435 MS reports no disclosures.
436 CY reports no disclosures.
437 HL reports no disclosures.
438 SKN reports no disclosures.
439 GMR reports no disclosures.
440 ZJ reports no disclosures.
441 KP reports no disclosures.
442 SA reports no disclosures.
443 AN reports no disclosures.
444 KU reports no disclosures.
445 MR reports no disclosures.
446 CB reports no disclosures.
447 GR reports no disclosures.
448 MAS reports no disclosures.
449 JSN reports no disclosures.
450 HH reports no disclosures.

451 IWD manages a fee-for-service sequencing facility at the Garvan Institute of Medical Research that is a customer of
452 Oxford Nanopore Technologies but has no further financial relationship. IWD has previously received travel and
453 accommodation expenses to speak at Oxford Nanopore Technologies conferences.

454 MN received consultancy honoraria from Reata Pharmaceuticals unrelated to this work.

455 AB reports no disclosures.

456 LMP received consultancy honoraria from Biogen unrelated to this work.

457 DS reports no disclosures.

458 SZ has received consultancy honoraria from Neurogene, Aeglea BioTherapeutics, Applied Therapeutics, and is an
459 unpaid officer of the TGP foundation, all unrelated to the present manuscript.

460 AD serves as an advisor at Critical Path Ataxia Therapeutics Consortium and her Institution (Paris Brain Institute)
461 receives consulting fees on her behalf from Biogen, Huntix, UCB, as well as research grants from the NIH,
462 ANR and holds partly a Patent B 06291873.5 on “Anaplerotic therapy of Huntington disease and other
463 polyglutamine diseases”.

464 BB reports no disclosures.

465

466 **Author contributions**

467 Design or conceptualization of the study: DP, JLM, MCD, AB, SZ, AD, BB.

468 Acquisition of data: DP, JLM, SB, MCD, MJD, CSD, DG, GS, CA, JMH, BJG, VC, PUL, MS, CY, HL, JSN, HH,
469 IWD, MN, AB, LMP, DS, SZ, AD, BB.

470 Analysis or interpretation of the data: DP, JLM, SB, MCD, MJD, CSD, DG, GS, CA, JMH, BJG, VC, PUL, MS,
471 CY, HL, SKN, GMR, ZJ, KP, SA, AN, KU, MR, CB, GR, MAS, JSN, HH, IWD, MN, AB, LMP, DS, SZ, AD, BB.

472 Drafting or revising the manuscript for intellectual content: DP, JLM, SB, MCD, MJD, CSD, DG, GS, CA, JMH,
473 BJG, VC, PUL, MS, CY, HL, SKN, GMR, ZJ, KP, SA, AN, KU, MR, CB, GR, MAS, JSN, HH, IWD, MN, AB,
474 LMP, DS, SZ, AD, BB.

475 All authors read and approved the final version of the manuscript.

476

477 **References**

- 478 1. Pellerin D, Danzi MC, Wilke C, et al. Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset
479 Cerebellar Ataxia. *N Engl J Med* 2023;388:128-141.
- 480 2. Rafehi H, Read J, Szmulewicz DJ, et al. An intronic GAA repeat expansion in FGF14 causes the
481 autosomal-dominant adult-onset ataxia SCA27B/ATX-FGF14. *American journal of human genetics* 2023;110:1018.
- 482 3. Wilke C, Pellerin D, Mengel D, et al. GAA-FGF14 ataxia (SCA27B): phenotypic profile, natural history
483 progression and 4-aminopyridine treatment response. *Brain* 2023;146:4144-4157.
- 484 4. Méreaux JL, Davoine CS, Pellerin D, et al. Clinical and genetic keys to cerebellar ataxia due to FGF14
485 GAA expansions. *EBioMedicine* 2024;99:104931.
- 486 5. Pellerin D, Heindl F, Wilke C, et al. GAA-FGF14 disease: defining its frequency, molecular basis, and 4-
487 aminopyridine response in a large downbeat nystagmus cohort. *EBioMedicine* 2024;102:105076.
- 488 6. Pellerin D, Danzi MC, Renaud M, Houlden H, Synofzik M, Zuchner S, Brais B. Spinocerebellar ataxia
489 27B: A novel, frequent and potentially treatable ataxia. *Clin Transl Med* 2024;14:e1504.
- 490 7. Wang Q, McEwen DG, Ornitz DM. Subcellular and developmental expression of alternatively spliced
491 forms of fibroblast growth factor 14. *Mech Dev* 2000;90:283-287.
- 492 8. Pellerin D, Del Gobbo GF, Couse M, et al. A common flanking variant is associated with enhanced
493 stability of the FGF14-SCA27B repeat locus. *Nature genetics* 2024.
- 494 9. Khristich AN, Mirkin SM. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome
495 instability. *J Biol Chem* 2020;295:4134-4170.
- 496 10. Jones L, Houlden H, Tabrizi SJ. DNA repair in the trinucleotide repeat disorders. *Lancet Neurol*
497 2017;16:88-96.
- 498 11. Matos-Rodrigues G, Hisey JA, Nussenzweig A, Mirkin SM. Detection of alternative DNA structures and
499 its implications for human disease. *Molecular cell* 2023;83:3622-3641.
- 500 12. Depienne C, Mandel JL. 30 years of repeat expansion disorders: What have we learned and what are the
501 remaining challenges? *American journal of human genetics* 2021;108:764-785.
- 502 13. Mouro Pinto R, Arning L, Giordano JV, et al. Patterns of CAG repeat instability in the central nervous
503 system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Human molecular genetics*
504 2020;29:2551-2567.

- 505 14. Kacher R, Lejeune FX, David I, et al. CAG repeat mosaicism is gene specific in spinocerebellar ataxias.
506 American journal of human genetics 2024;111(5):913-926.
- 507 15. Matsuura T, Sasaki H, Yabe I, Hamada K, Hamada T, Shitara M, Tashiro K. Mosaicism of unstable CAG
508 repeats in the brain of spinocerebellar ataxia type 2. J Neurol 1999;246:835-839.
- 509 16. Cancel G, Gourfinkel-An I, Stevanin G, et al. Somatic mosaicism of the CAG repeat expansion in
510 spinocerebellar ataxia type 3/Machado-Joseph disease. Human mutation 1998;11:23-27.
- 511 17. Takano H, Onodera O, Takahashi H, et al. Somatic mosaicism of expanded CAG repeats in brains of
512 patients with dentatorubral-pallidoluysian atrophy: cellular population-dependent dynamics of mitotic instability.
513 American journal of human genetics 1996;58:1212-1222.
- 514 18. Sanchez-Flores M, Corral-Juan M, Gasch-Navalón E, Cirillo D, Sanchez I, Matilla-Dueñas A. Novel
515 genotype-phenotype correlations, differential cerebellar allele-specific methylation, and a common origin of the
516 (ATTTC)(n) insertion in spinocerebellar ataxia type 37. Human genetics 2024;143(3):211-232.
- 517 19. Currò R, Dominik N, Facchini S, et al. Role of the repeat expansion size in predicting age of onset and
518 severity in RFC1 disease. Brain 2024;147(5):1887-1898.
- 519 20. De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Coccozza S, Bidichandani SI. Somatic
520 instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. Genomics
521 2007;90:1-5.
- 522 21. De Biase I, Rasmussen A, Endres D, et al. Progressive GAA expansions in dorsal root ganglia of
523 Friedreich's ataxia patients. Ann Neurol 2007;61:55-60.
- 524 22. Long A, Napierala JS, Polak U, Hauser L, Koeppen AH, Lynch DR, Napierala M. Somatic instability of
525 the expanded GAA repeats in Friedreich's ataxia. PLoS One 2017;12:e0189990.
- 526 23. Montermini L, Kish SJ, Jiralerspong S, Lamarche JB, Pandolfo M. Somatic mosaicism for Friedreich's
527 ataxia GAA triplet repeat expansions in the central nervous system. Neurology 1997;49:606-610.
- 528 24. Lee JM, Zhang J, Su AI, et al. A novel approach to investigate tissue-specific trinucleotide repeat
529 instability. BMC Syst Biol 2010;4:29.
- 530 25. Bonnet C, Pellerin D, Roth V, et al. Optimized testing strategy for the diagnosis of GAA-FGF14
531 ataxia/spinocerebellar ataxia 27B. Scientific reports 2023;13:9737.

- 532 26. Sjöstedt E, Zhong W, Fagerberg L, et al. An atlas of the protein-coding genes in the human, pig, and mouse
533 brain. *Science* 2020;367.
- 534 27. Ditch S, Sammarco MC, Banerjee A, Grabczyk E. Progressive GAA.TTC repeat expansion in human cell
535 lines. *PLoS genetics* 2009;5:e1000704.
- 536 28. Rodden LN, Chutake YK, Gilliam K, et al. Methylated and unmethylated epialleles support variegated
537 epigenetic silencing in Friedreich ataxia. *Human molecular genetics* 2021;29:3818-3829.
- 538 29. Stevanovski I, Chintalaphani SR, Gamaarachchi H, et al. Comprehensive genetic diagnosis of tandem
539 repeat expansion disorders with programmable targeted nanopore sequencing. *Sci Adv* 2022;8:eabm5386.
- 540 30. Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, Ennis M, Shelbourne PF. Dramatic tissue-specific
541 mutation length increases are an early molecular event in Huntington disease pathogenesis. *Human molecular*
542 *genetics* 2003;12:3359-3367.
- 543 31. Jinnai K, Mitani M, Futamura N, Kawamoto K, Funakawa I, Itoh K. Somatic instability of CTG repeats in
544 the cerebellum of myotonic dystrophy type 1. *Muscle & nerve* 2013;48:105-108.
- 545 32. Tanaka F, Reeves MF, Ito Y, et al. Tissue-specific somatic mosaicism in spinal and bulbar muscular
546 atrophy is dependent on CAG-repeat length and androgen receptor--gene expression level. *American journal of*
547 *human genetics* 1999;65:966-973.
- 548 33. Masnovo C, Lobo AF, Mirkin SM. Replication dependent and independent mechanisms of GAA repeat
549 instability. *DNA Repair (Amst)* 2022;118:103385.
- 550 34. Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M, Bidichandani SI. The GAA triplet-repeat is
551 unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a
552 transgenic mouse model. *Human genetics* 2007;120:633-640.
- 553 35. Al-Mahdawi S, Pinto RM, Ruddle P, Carroll C, Webster Z, Pook M. GAA repeat instability in Friedreich
554 ataxia YAC transgenic mice. *Genomics* 2004;84:301-310.
- 555 36. Neil AJ, Hisey JA, Quasem I, McGinty RJ, Hitczenko M, Khristich AN, Mirkin SM. Replication-
556 independent instability of Friedreich's ataxia GAA repeats during chronological aging. *Proc Natl Acad Sci U S A*
557 2021;118.

- 558 37. Du J, Campau E, Soragni E, Ku S, Puckett JW, Dervan PB, Gottesfeld JM. Role of mismatch repair
559 enzymes in GAA·TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. *J Biol Chem*
560 2012;287:29861-29872.
- 561 38. Halabi A, Ditch S, Wang J, Grabczyk E. DNA mismatch repair complex MutS β promotes GAA·TTC
562 repeat expansion in human cells. *J Biol Chem* 2012;287:29958-29967.
- 563 39. Li L, Scott WS, Khristich AN, Armenia JF, Mirkin SM. DNA Nicks Drive Massive Expansions of (GAA)_n
564 Repeats. *bioRxiv* 2024:2024.2006.2012.598717.
- 565 40. Wu W, Hill SE, Nathan WJ, et al. Neuronal enhancers are hotspots for DNA single-strand break repair.
566 *Nature* 2021;593:440-444.
- 567 41. Matos-Rodrigues G, van Wietmarschen N, Wu W, et al. S1-END-seq reveals DNA secondary structures in
568 human cells. *Molecular cell* 2022;82:3538-3552.e3535.
- 569 42. Neil AJ, Liang MU, Khristich AN, Shah KA, Mirkin SM. RNA-DNA hybrids promote the expansion of
570 Friedreich's ataxia (GAA)_n repeats via break-induced replication. *Nucleic acids research* 2018;46:3487-3497.
- 571 43. Pearson CE, Tam M, Wang YH, Montgomery SE, Dar AC, Cleary JD, Nichol K. Slipped-strand DNAs
572 formed by long (CAG)_n(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic acids research*
573 2002;30:4534-4547.
- 574 44. Rüb U, Schöls L, Paulson H, et al. Clinical features, neurogenetics and neuropathology of the
575 polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. *Prog Neurobiol* 2013;104:38-66.
- 576 45. Kacher R, Lejeune FX, Noël S, Cazeneuve C, Brice A, Humbert S, Durr A. Propensity for somatic
577 expansion increases over the course of life in Huntington disease. *Elife* 2021;10.
- 578 46. Later Development of Embryonic Central Nervous System. Reference Module in Biomedical Sciences:
579 Elsevier, 2014.
- 580 47. Kaplan S, Itzkovitz S, Shapiro E. A universal mechanism ties genotype to phenotype in trinucleotide
581 diseases. *PLoS Comput Biol* 2007;3:e235.

582

583 **Figure legends**

584 **Figure 1: Determination of the *FGF14* GAA•TTC repeat expansion index**

585 Method for calculating the expansion index of the *FGF14* GAA•TTC repeat. The expansion index is calculated by
586 only taking into account peaks to the right of the modal allele, which represent somatically expanded GAA•TTC
587 repeats, and using a 1% relative peak threshold. The modal allele corresponds to the peak with the highest intensity,
588 as measured in relative fluorescence units in GeneMapper or Peak Scanner. The expansion index is calculated by
589 first dividing the height of each individual peak by the sum of the heights of all peaks to the right of the modal allele
590 that are above the set 1% threshold. The resulting normalized peak heights are then multiplied by the position of the
591 peak (change from the modal allele) and these values are summed to generate the index.

592

593 **Figure 2: Longitudinal analysis of the *FGF14* GAA•TTC repeat somatic stability in blood samples**

594 Longitudinal analysis of (a) the *FGF14* GAA•TTC repeat tract size, expressed in triplet repeat counts, and (b)
595 expansion index across 69 individuals (120 alleles analyzed) who underwent serial blood collections over a median
596 period of 8.9 years (interquartile range [IQR]: 2.9 to 13.9). Observations from the same person are connected by a
597 line. In panel b, the color gradient shows the GAA•TTC allele size of each data point. (c) Positive linear relationship
598 between the *FGF14* GAA•TTC repeat tract size, expressed in triplet repeat counts, and expansion index in each of
599 the first (orange) and last (purple) longitudinal blood samples across 69 individuals ($n=120$ alleles). The Pearson's
600 correlation coefficient is $r = 0.73$ (95% confidence interval [CI]: 0.64 to 0.81) for the first blood samples and $r =$
601 0.84 (95% CI: 0.78 to 0.89) for the last blood samples. Both regression lines did not differ significantly (mixed-
602 effect analysis, $p=0.72$), indicating relative stability of the expansion index over time. (d) Positive linear relationship
603 between the *FGF14* GAA•TTC repeat tract size and expansion index across 100 individuals ($n=173$ alleles),
604 including the results from the last blood samples of the 69 individuals featured in panels b and c. The Pearson's
605 correlation coefficient is $r = 0.85$ (95% CI: 0.80 to 0.88). In panels c and d, the shaded areas display the 95%
606 confidence intervals.

607

608 **Figure 3: Somatic instability profile of the *FGF14* GAA•TTC repeat in brain regions**

609 (a) Somatic instability profiles of the *FGF14* GAA•TTC repeat in different brain regions, derived from post-mortem
610 samples of six non-ataxic control and three patients with SCA27B. Each plot shows the average instability profile

611 for each of the two alleles within a given brain region, calculated from replicate PCR reactions. For regions where
612 multiple tissue samples were analyzed, results for each sample are shown individually. Profiles were plotted by
613 normalizing individual peak height data to the height of the modal allele within each brain region, except for
614 SCA27B cases P1 and P3, where data were normalized to the height of the shorter modal allele due to the lack of
615 significant size difference with the longer allele. Peaks left of the modal allele above a 10% threshold and those right
616 of the modal allele above a 1% threshold were plotted. Vertical dashed black lines indicate the size of the modal
617 alleles measured in the blood and/or non-cerebellar regions, while vertical dashed red and orange lines indicate the
618 size of the modal alleles measured in the vermis and cerebellar hemispheres, respectively. Individual instability
619 profiles in each region are shown in **Supplementary Figures 8-16**. (b) Schematic representation of the various
620 brain regions analyzed for somatic instability in this study. The same color scheme is used to represent the brain
621 regions and their corresponding instability profiles shown in panel a. Panel created with BioRender.com.

622

623 **Figure 4: Expansion index of the *FGF14* GAA•TTC repeat in brain regions**

624 Schematic representation of (a) the relative expansion index and (b) the regional expression of *FGF14* (normal
625 tissue expression data in transcripts per million [TPM] from the Human Protein Atlas) across the analyzed brain
626 regions shown in **Figure 3**. Expansion indices for each brain region were normalized relative to those in the
627 cerebellar hemispheres. The relative expansion indices were calculated by averaging the relative expansion index of
628 each allele, excluding (GAA)₉ alleles, which did not exhibit somatic expansion. A corresponding heatmap of the
629 results for each region is shown on the right side of both panels. Expression data were not available for the
630 frontopolar cortex and the pituitary gland. Figure created with BioRender.com.

631

632 **Figure 5: *FGF14* GAA•TTC repeat lengths in brain regions of SCA27B patients**

633 (a) Distribution of *FGF14* GAA•TTC repeat lengths, expressed in triplet repeat counts, across post-mortem brain
634 regions and blood samples in the six SCA27B patients included in this study. For brain regions where multiple
635 tissue samples were analyzed, only the results of the sample with the largest allele size are shown. Observations for
636 each of the two alleles from the same patient are connected by a dashed line. (b) Comparison of *FGF14* GAA•TTC
637 repeat lengths in blood samples (*x*-axis) versus cerebellar hemispheres (*y*-axis) for five SCA27B patients and eight
638 controls with available tissue samples. The dashed black line represents the identity line. In panels a and b, a single

639 modal allele size is shown for patient P1 due to the complete blending of the instability profiles for the short and
640 long alleles in the cerebellar hemispheres and vermis (**Supplementary Figure 33**), preventing the identification of
641 the two distinct modal peaks in these tissues.

642

643 **Figure 6: Methylation profiling of the *FGF14* gene in post-mortem cerebellum**

644 Methylation analysis of the *FGF14* gene in post-mortem cerebellar hemispheres from four controls (C7-C10) and
645 four patients with SCA27B (P1, P2, P3, P7) using programmable targeted long-read sequencing with Oxford
646 Nanopore Technologies. (a) Methylation profiles, expressed as 5'-methylcytosine (5mC) methylation frequencies at
647 all CpG sites within the *FGF14* locus (T2T-CHM13, chr13:100923763-101619864), in the four control and four
648 patient post-mortem cerebellar hemispheres. A diagram of the *FGF14* gene with its two putative promoters (P1:
649 promoter 1 and P2: promoter 2) is shown at the bottom of the panel. (b) Plots shows the distribution of 5mC
650 methylation frequencies at all CpG sites within the *FGF14* locus (top plot), the first promoter (middle plot), and
651 second promoter (bottom plot). The horizontal black bars show the median 5mC methylation frequency.

652

653 **Tables**

654 **Table 1: Demographic and genetic characteristics of the post-mortem brain samples analyzed for *FGF14***

655 **GAA•TTC repeat instability**

Study site	Status	ID number	GAA•TTC repeat units	Age at onset (years)	Age at death (years)
Montreal	Control	C1	15 / 92	-	70-80 y
Montreal	Control	C2	9 / 36	-	60-70 y
Paris	Control	C3	16 / 152	-	70-80 y
Paris	Control	C4	23 / 165	-	90-100 y
Paris	Control	C5	47 / 151	-	80-90 y
Paris	Control	C6	17 / 42	-	60-70 y
Montreal	SCA27B	P1	292 / 304	50-60 y	60-70 y
Barcelona	SCA27B	P2	9 / 288	70-80 y	90-100 y
Barcelona	SCA27B	P3	191 / 248	70-80 y	80-90 y
Paris	SCA27B	P4	31 / 548	40-50 y	70-80 y
Paris	SCA27B	P5	70 / 480	NA	80-90 y
Paris	SCA27B	P6	32 / 405	50-60 y	70-80 y

656

657 Data on sex have been removed from the table, and the exact ages at onset and at death are not provided to protect

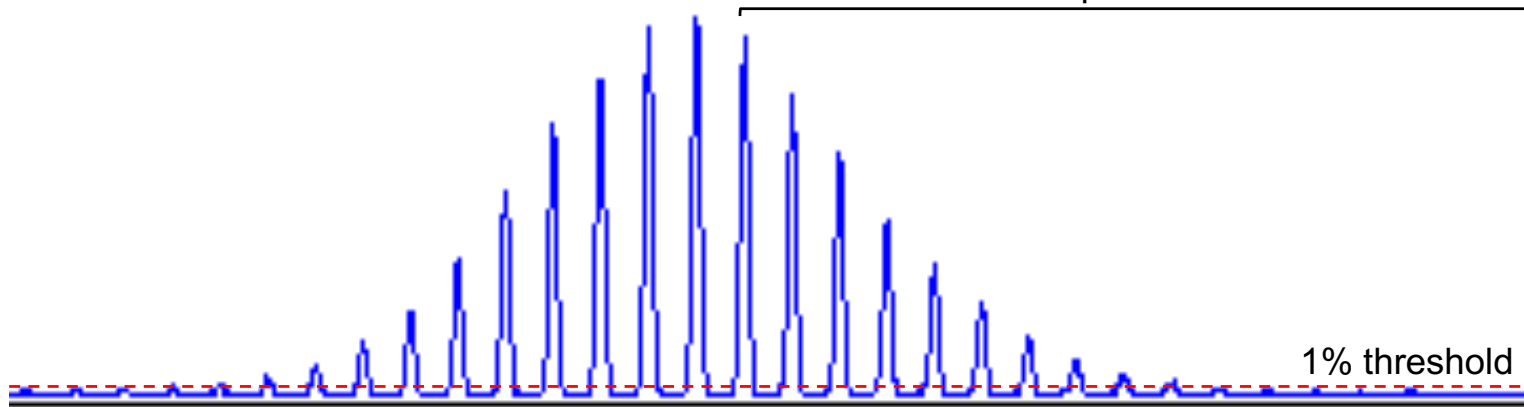
658 the privacy of the brain donors.

659 Abbreviations: NA, not available; SCA27B, spinocerebellar ataxia 27B

Expansion index

Modal
allele

Expansions



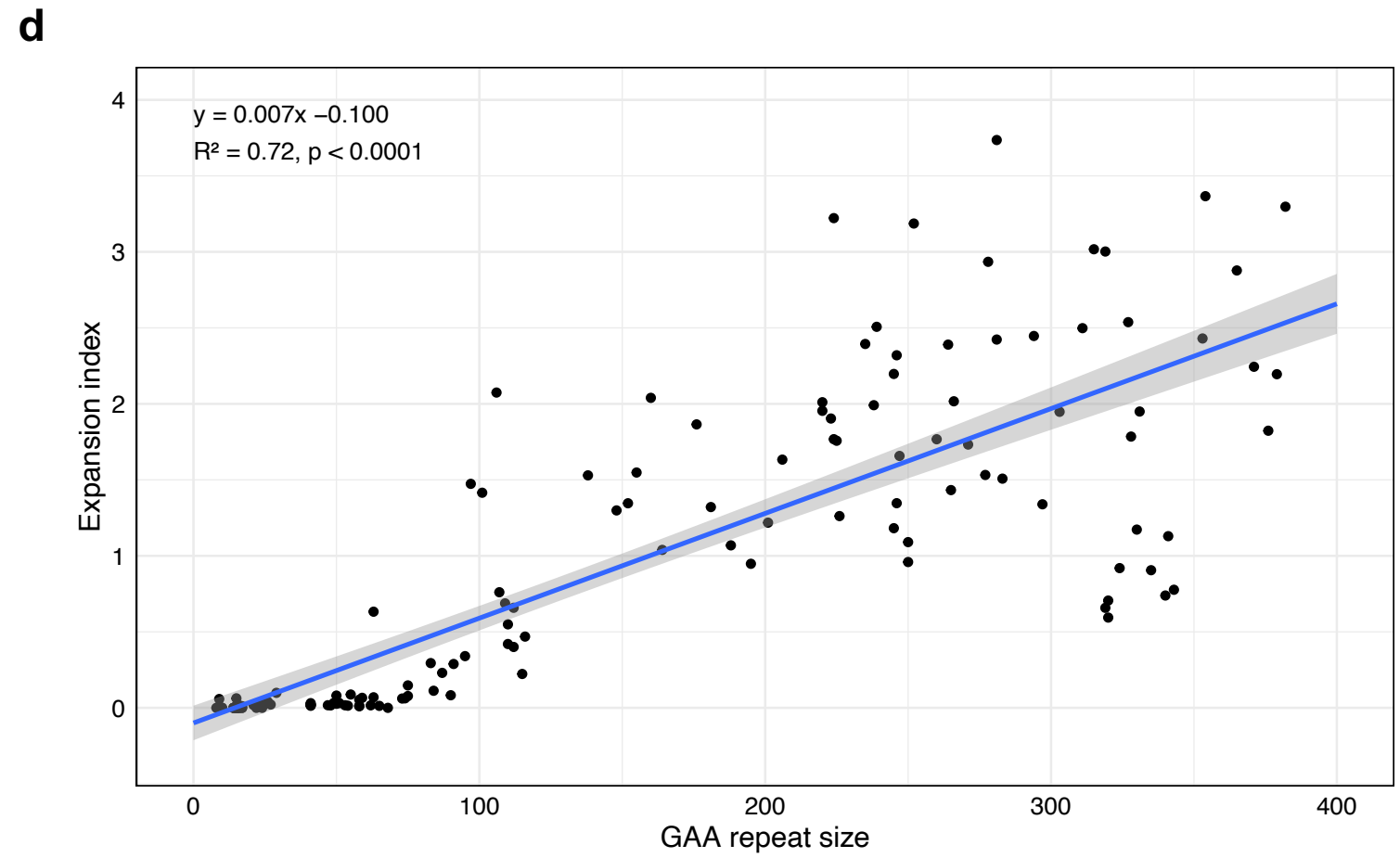
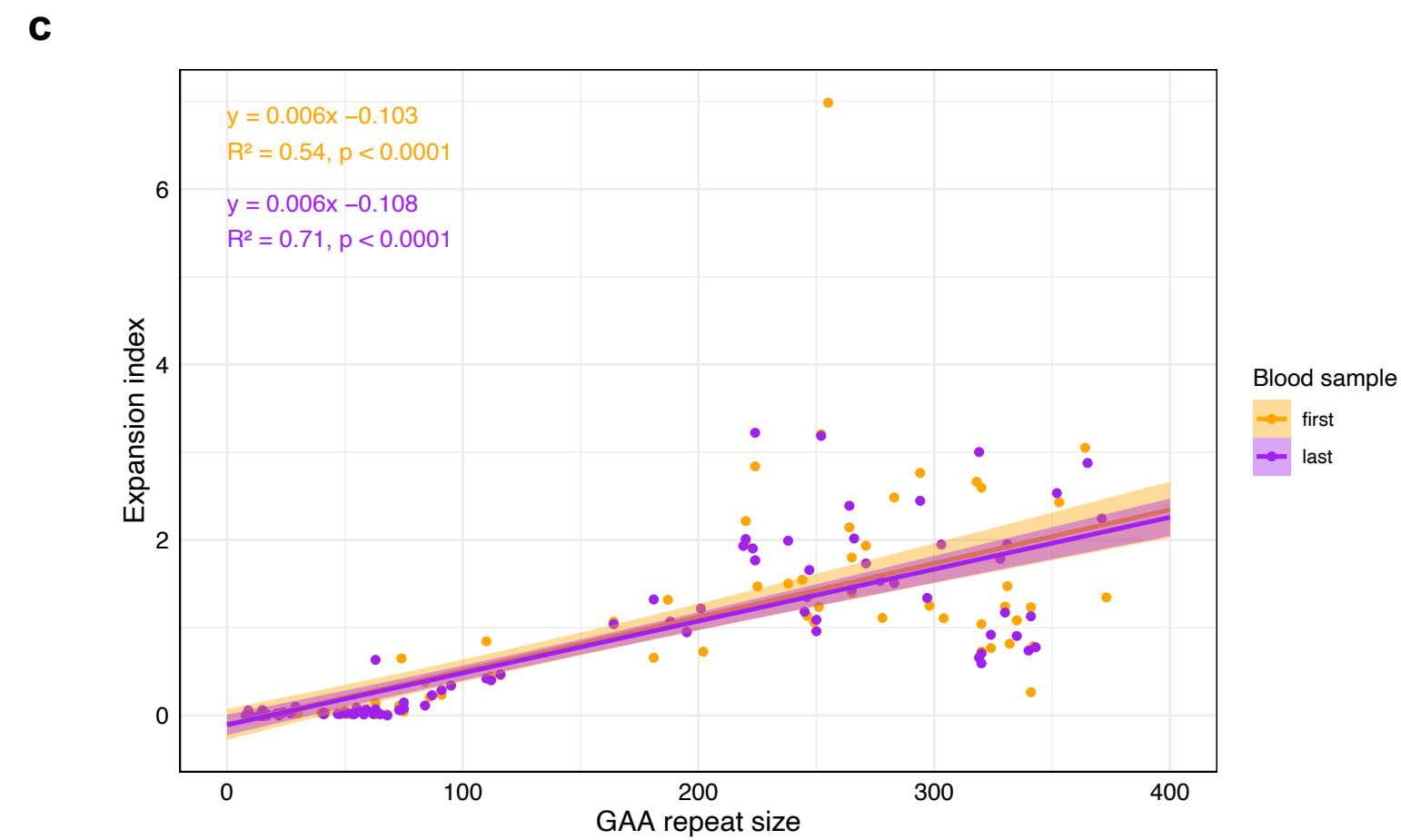
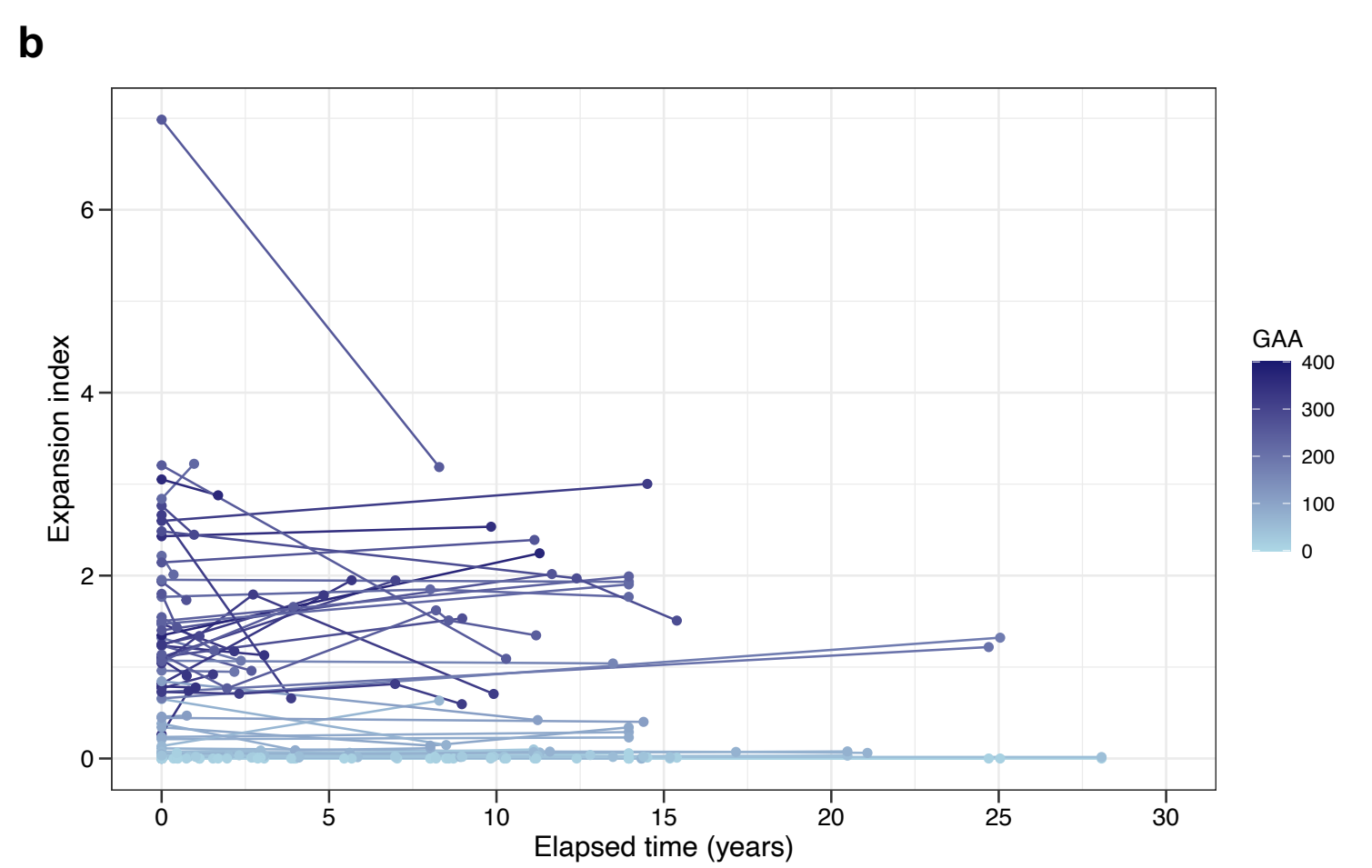
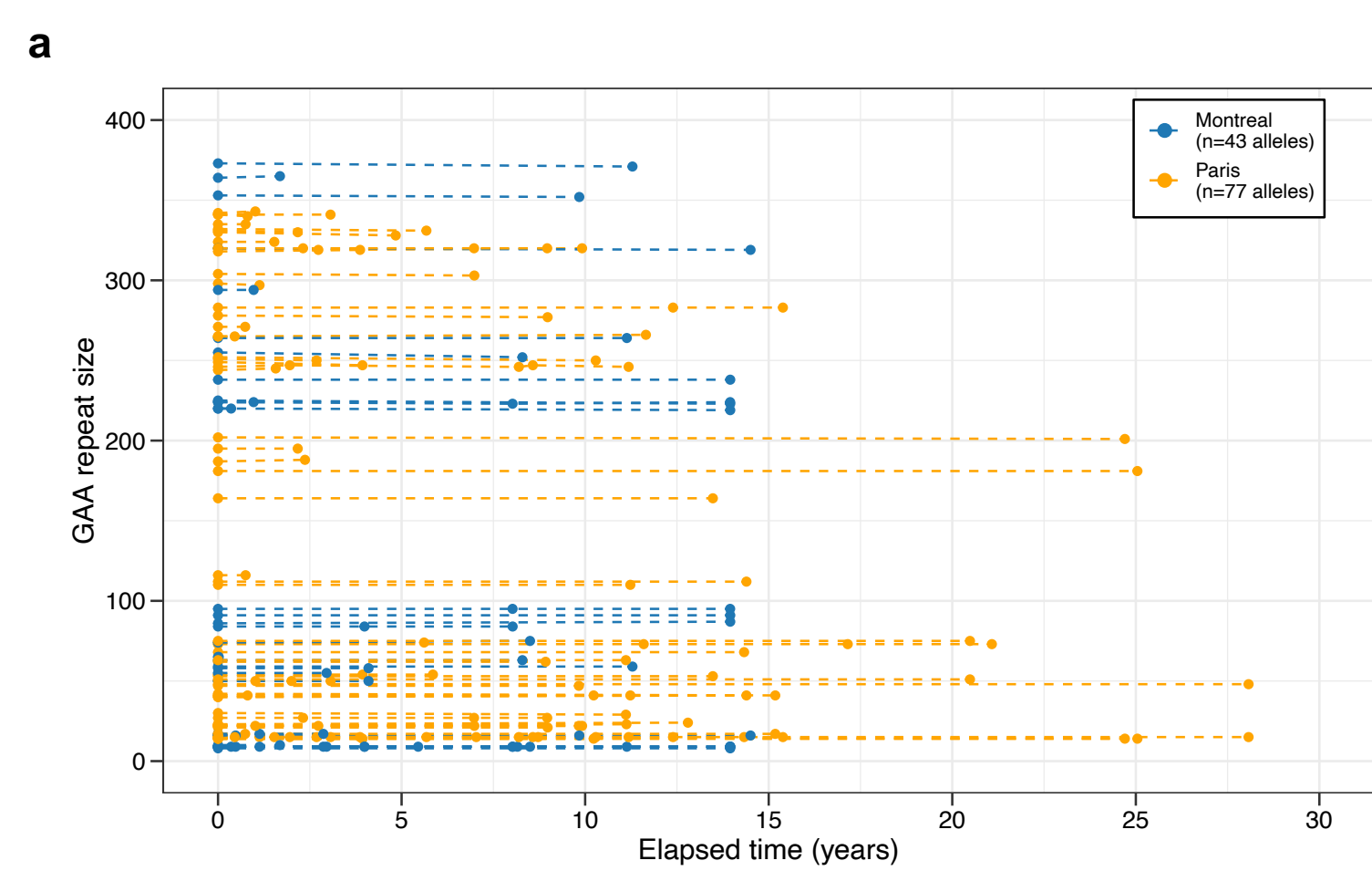
1% threshold

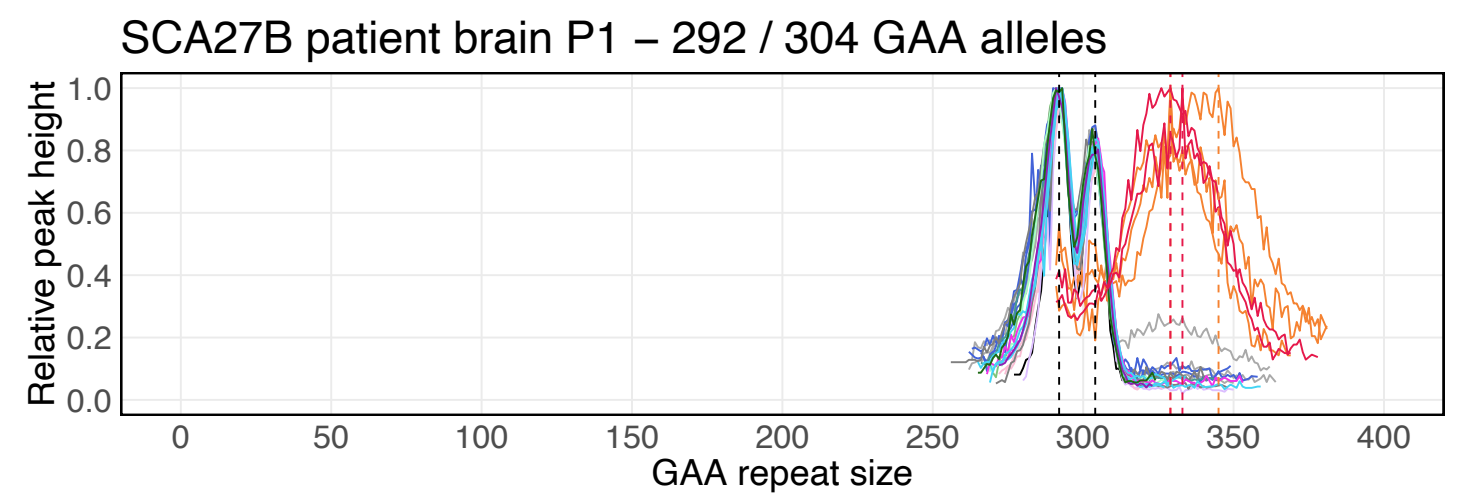
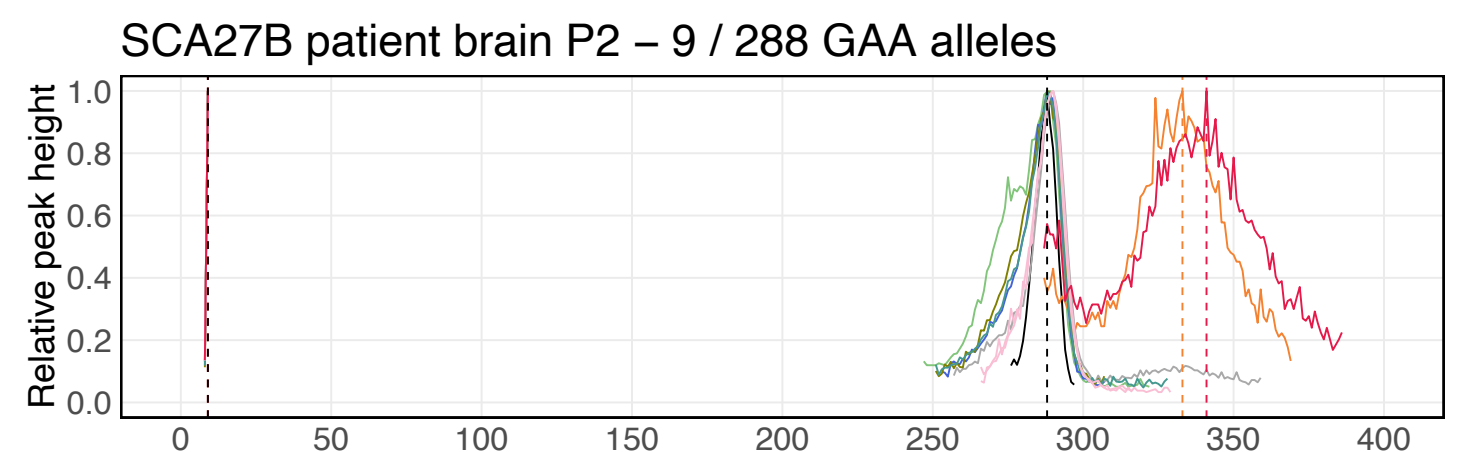
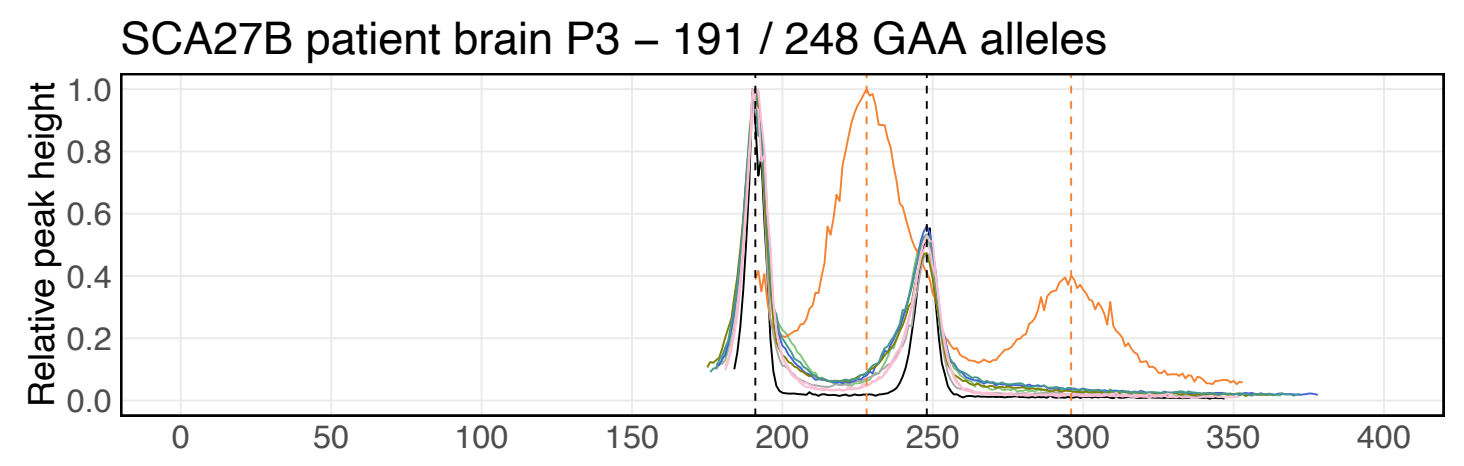
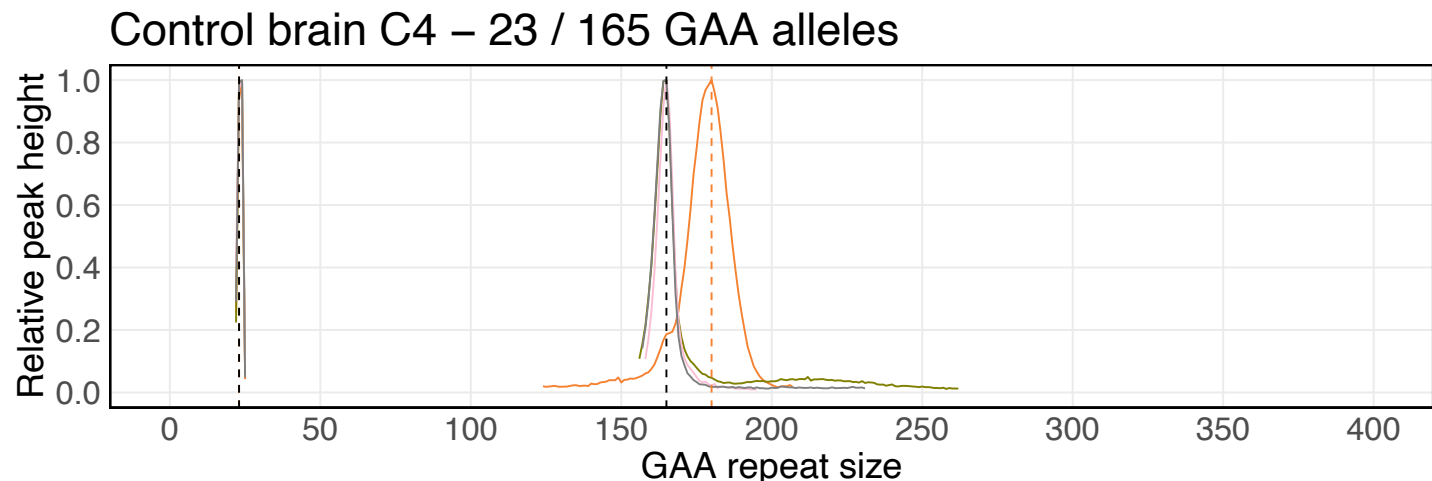
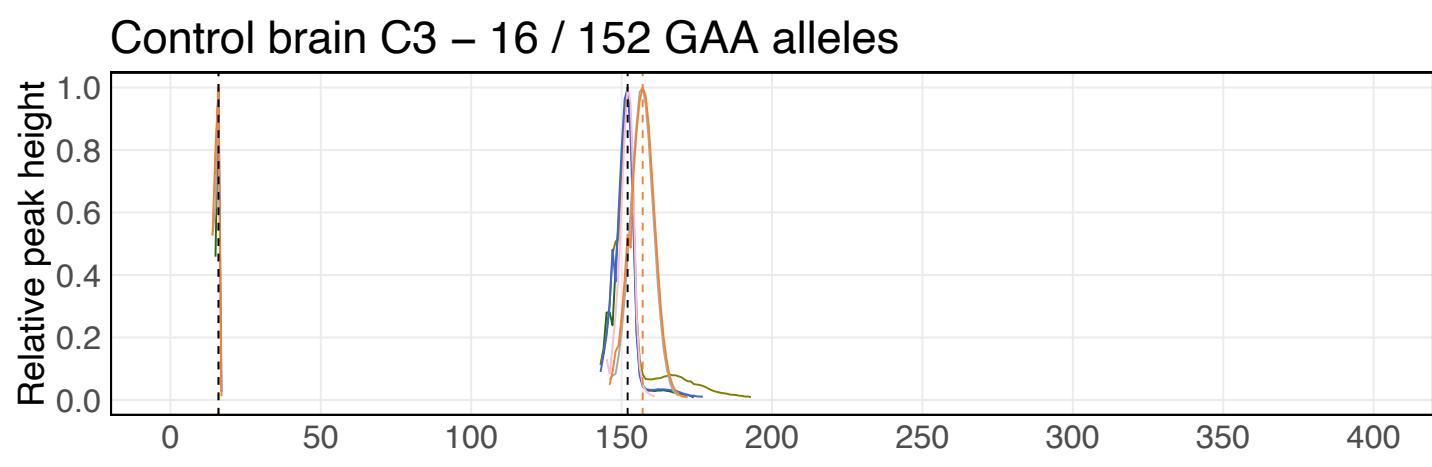
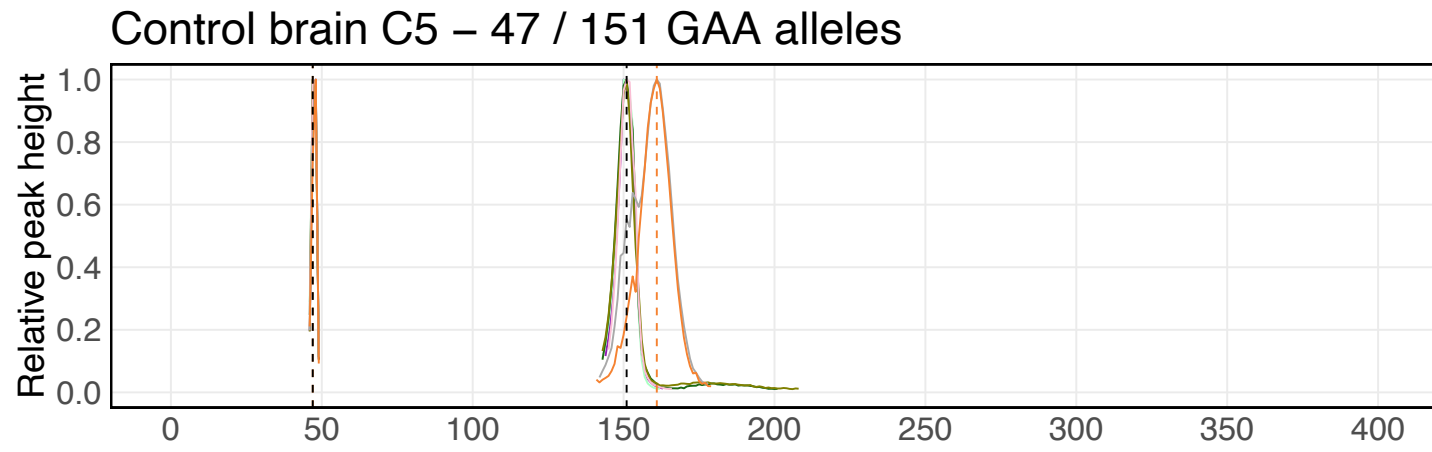
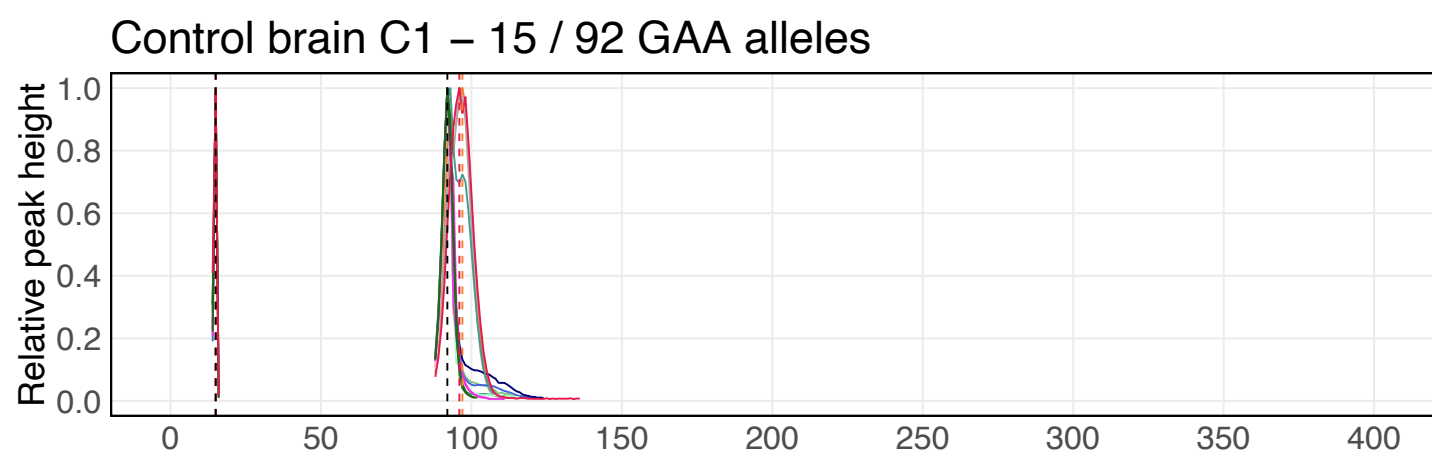
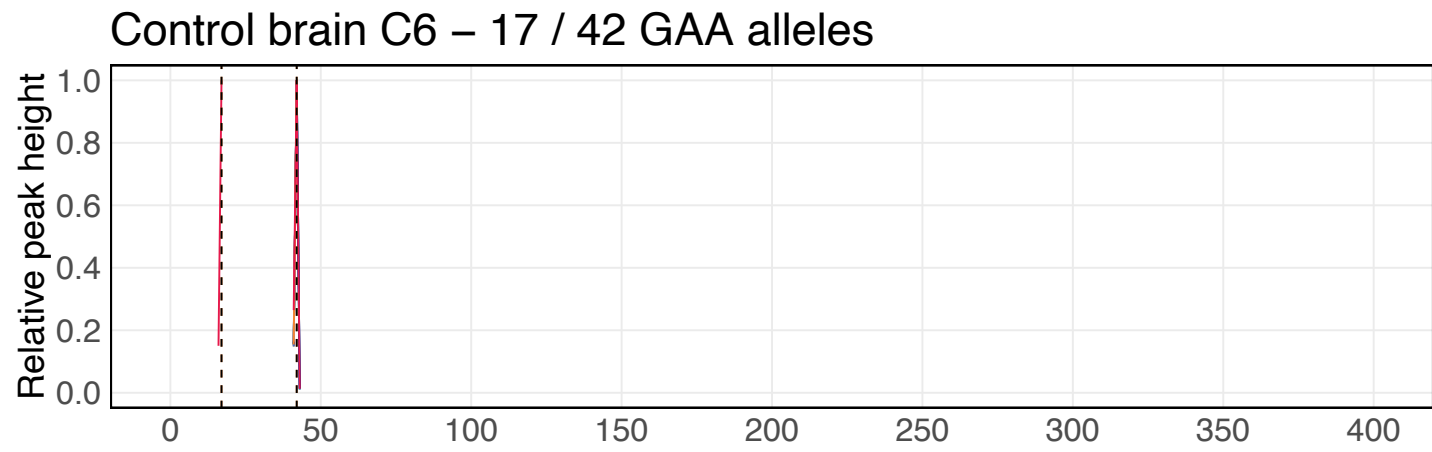
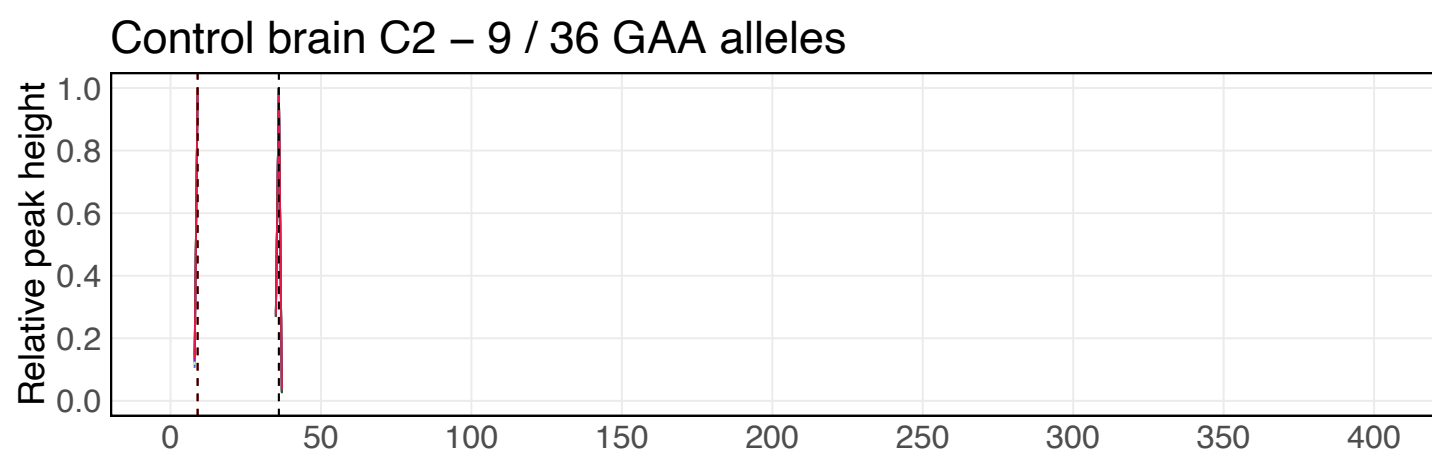
Change from
the modal allele

0 1 2 3 4 5 6 7 8 9 10 11

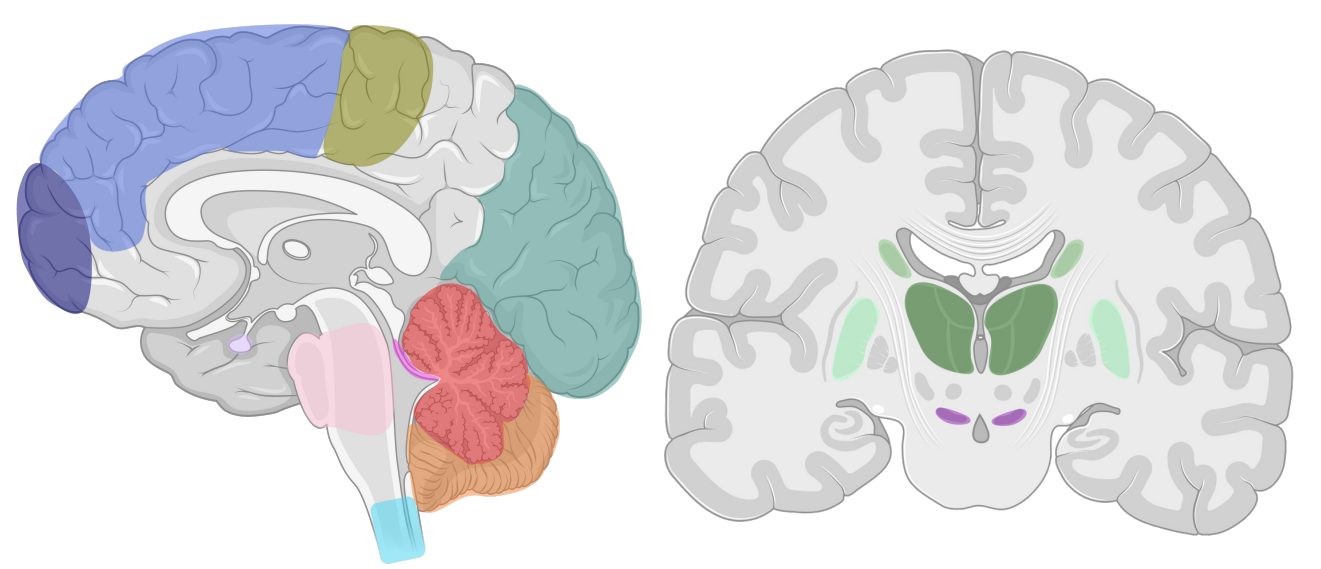
Expansion index = sum of (normalized peak heights x change from the modal allele)

Normalized peak heights = individual peak height / sum of the heights of all 'expansion' peaks above threshold



a**b**

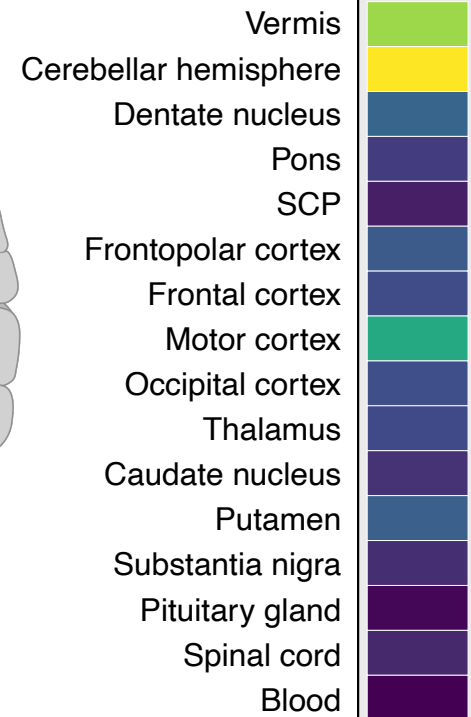
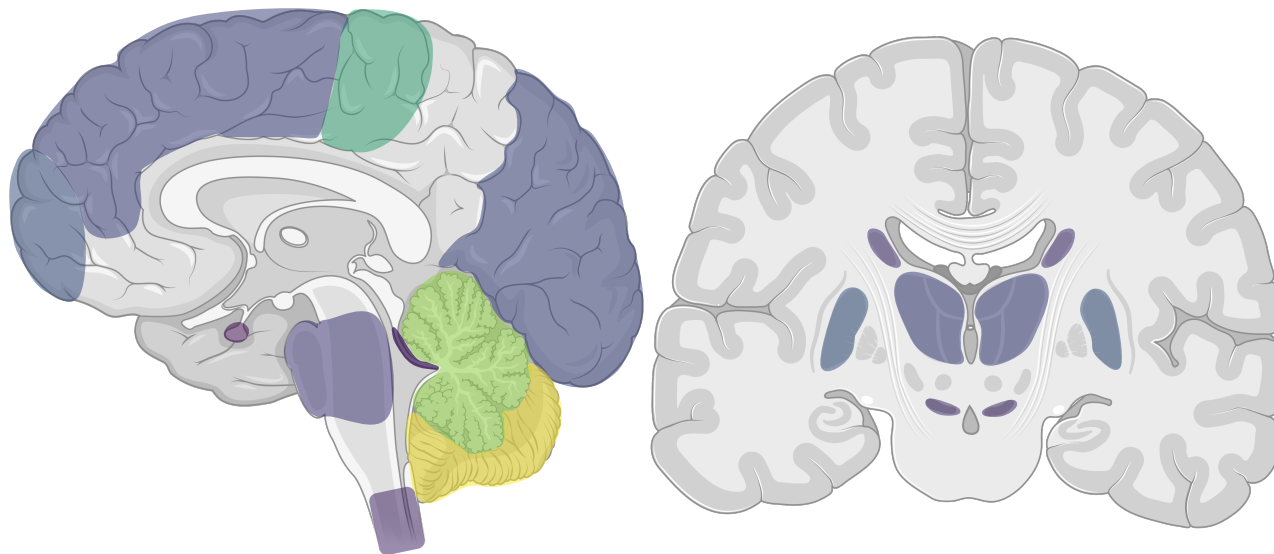
Analyzed brain regions



- | | | | |
|-------------------------|----------------------|-------------------|--------------------|
| ■ Blood | ■ Frontal cortex | ■ Pituitary gland | ■ Spinal cord |
| ■ Caudate nucleus | ■ Frontopolar cortex | ■ Pons | ■ Substantia nigra |
| ■ Cerebellar hemisphere | ■ Motor cortex | ■ Putamen | ■ Thalamus |
| ■ Dentate nucleus | ■ Occipital cortex | ■ SCP | ■ Vermis |

a

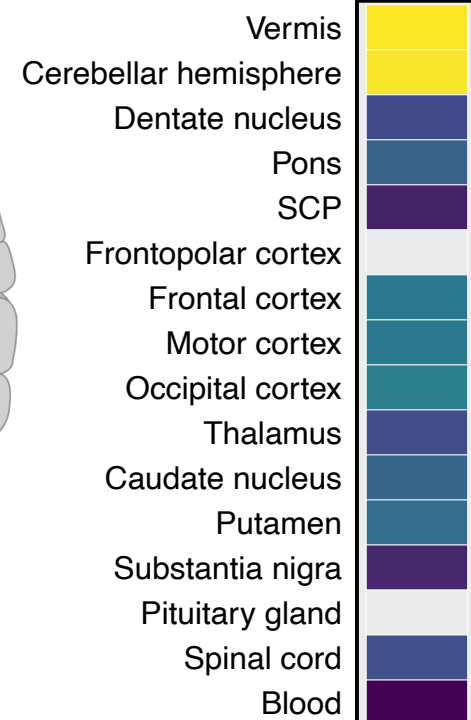
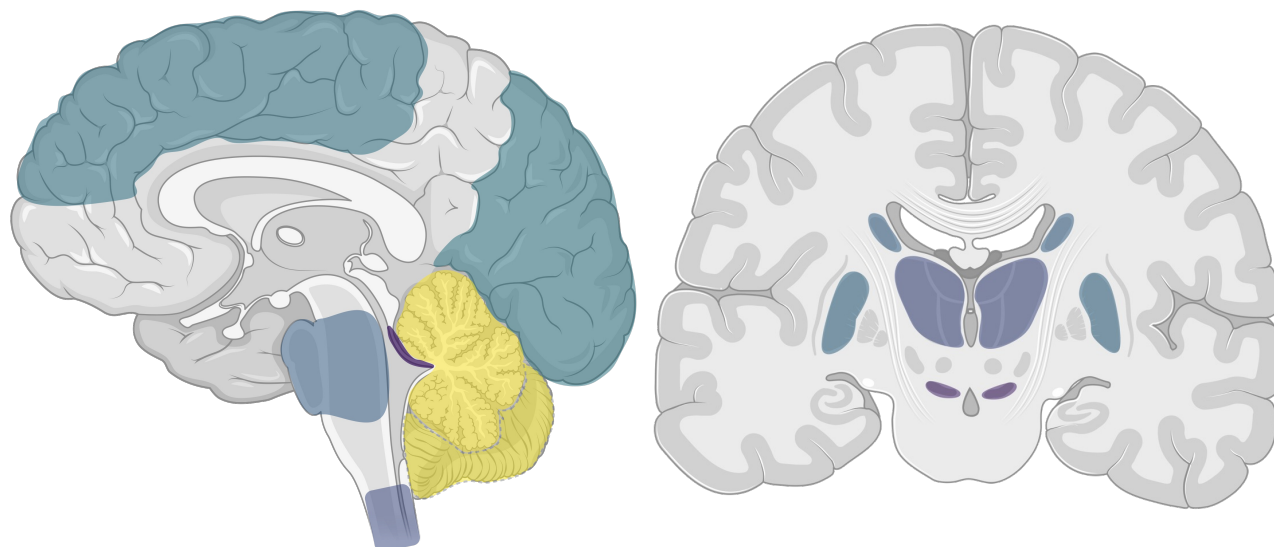
Relative expansion index



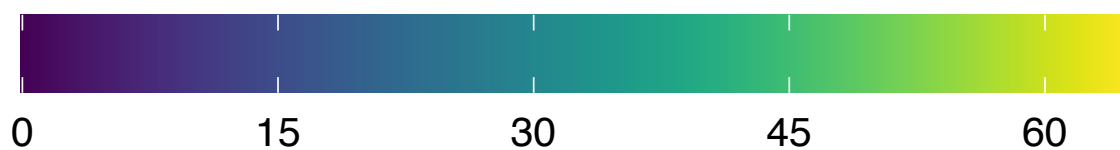
Relative Expansion Index

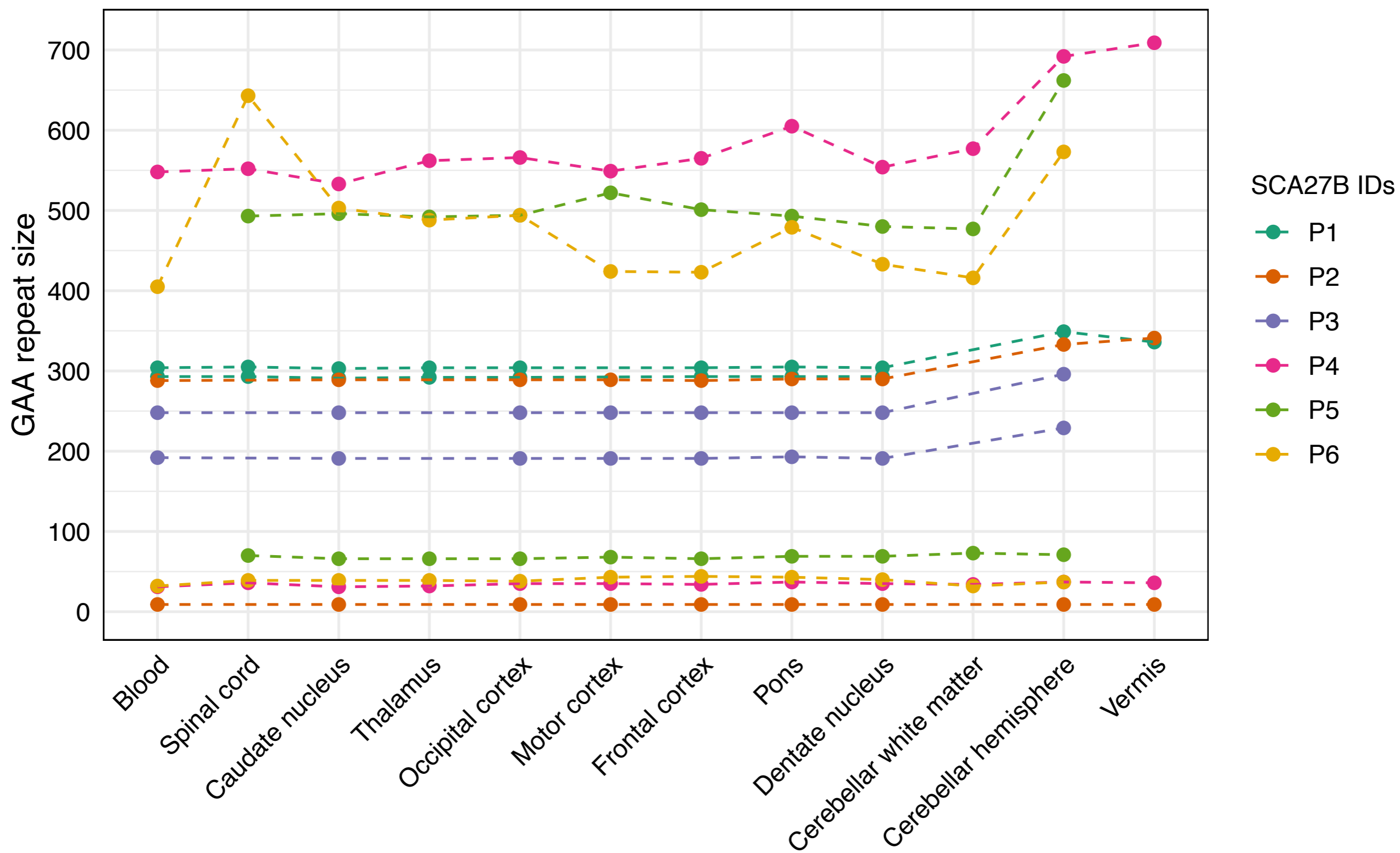
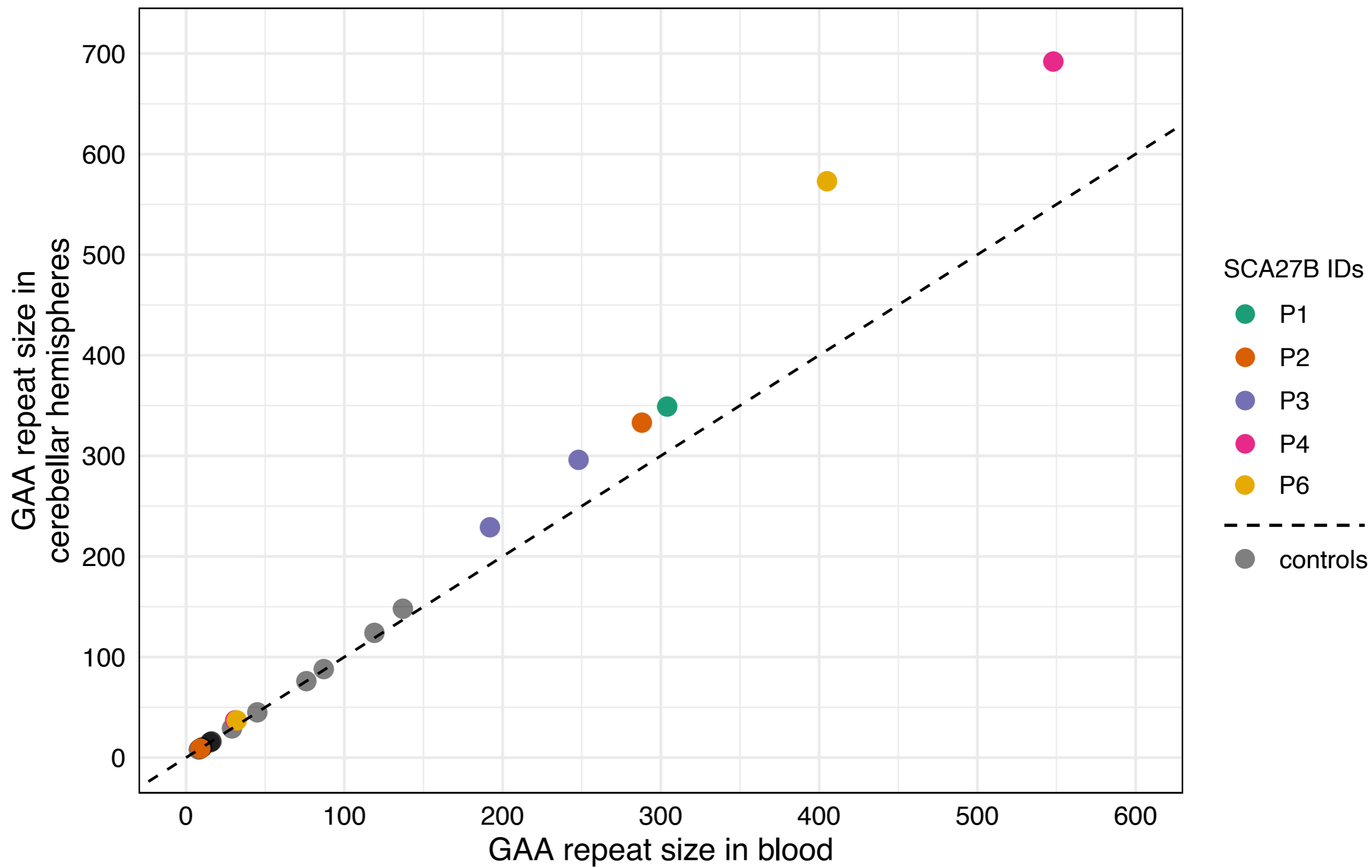
**b**

Regional expression of FGF14 (Human Protein Atlas)



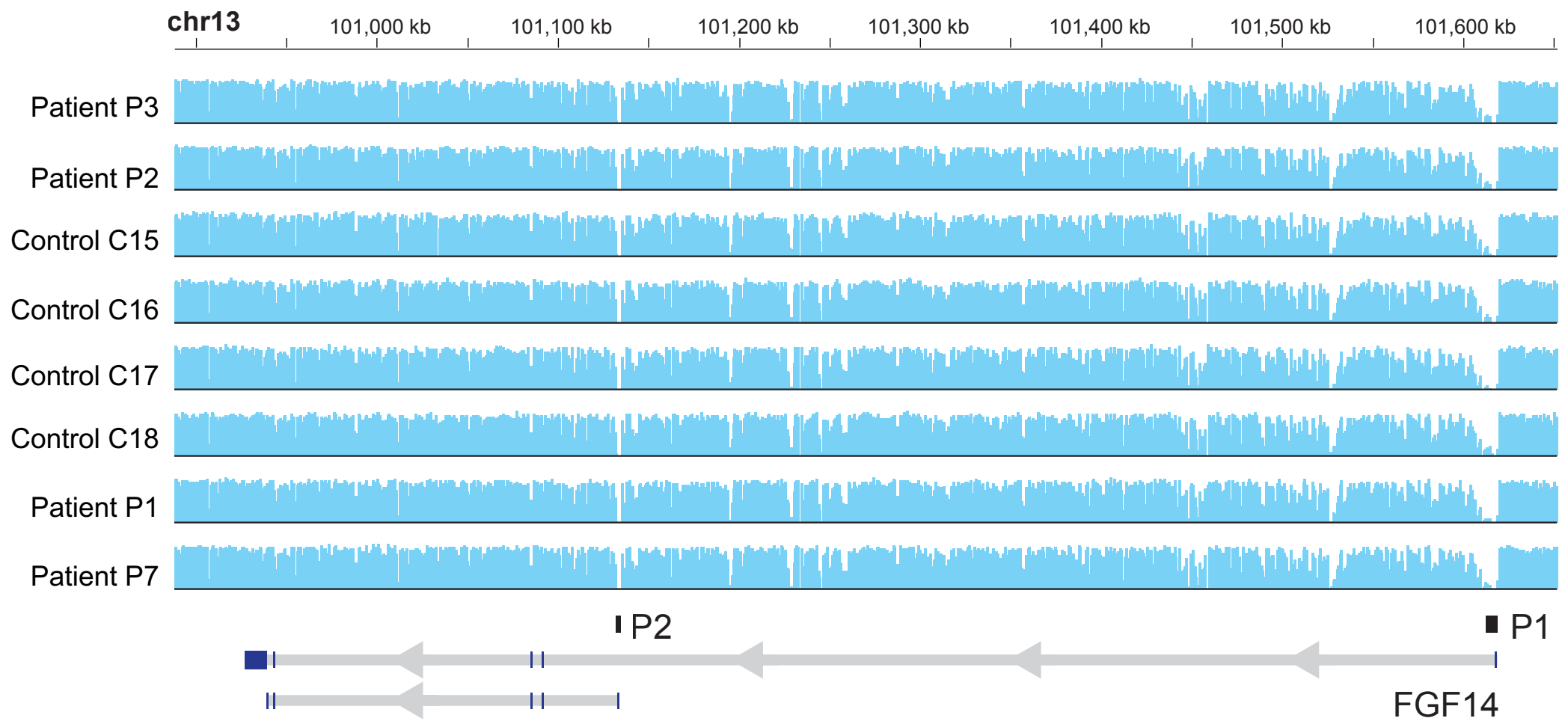
TPM



a**b**

a

5mC methylation frequencies



b

FGF14 promoter methylation levels

