

1 **Fragile X Syndrome Carrier Screening Using a Nanopore Sequencing Assay**

2 Running title: FXS Carrier Screening using Nanopore Sequencing

3 Zhongmin Xia^{a,†}, Qiuxiao Deng^{a,b,†}, Ping Hu^{c,†}, Chunliu Gao^d, Yu Jiang^e, Yulin Zhou^{a,b,*}, and

4 Qiwei Guo^{a,b,*}

5 ^aDepartment of Medical Genetics, Women and Children's Hospital, School of Medicine,

6 Xiamen University, Xiamen, Fujian, China.

7 ^bSchool of Medicine, Xiamen University, Xiamen, Fujian, China.

8 ^cDepartment of Prenatal Diagnosis, Women's Hospital of Nanjing Medical University,

9 Nanjing Women and Children's Healthcare Hospital, Nanjing, Jiangsu, China.

10 ^dSchool of Public Health, Xiamen University, Xiamen, Fujian, China.

11 ^eMolecular Diagnostic Laboratory for Precision Medicine, Xiang'an Hospital of Xiamen

12 University, School of Medicine, Xiamen University, Xiamen, Fujian, China.

13 [†]Zhongmin Xia, Qiuxiao Deng, and Ping Hu contributed equally to this work.

14 *Correspondence to:

15 Qiwei Guo, Department of Medical Genetics, Women and Children's Hospital, School of

16 Medicine, Xiamen University, Xiamen, Fujian, China. Tel: +86-592-7805028. Email:

17 guoqiwei@xmu.edu.cn

18 Yulin Zhou, Department of Medical Genetics, Women and Children's Hospital, School of

19 Medicine, Xiamen University, Xiamen, Fujian, China. Tel: +86-592-7805028. Email:

20 zhou_yulin@126.com

21 Keywords: AGG interruption, carrier screening, CGG repeat, fragile X syndrome, nanopore

22 sequencing

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

23 Word count: 2969

24 Figures: 4

25 Tables: 2

26 Nonstandard abbreviations: FXS, fragile X syndrome; TP-PCR, triplet-primed PCR.

27 Human gene: *FMRI*, fragile X messenger ribonucleoprotein 1.

28 **BACKGROUND:** Fragile X syndrome (FXS) is the leading cause of monogenic autism
29 spectrum disorder and inherited intellectual disabilities. Although the value of
30 population-based FXS carrier screening has been acknowledged, appropriate screening
31 methods are urgently required to establish and implement screening programs.

32 **METHODS:** We developed a nanopore sequencing-based assay that includes data analysis
33 software to identify FXS carriers. Reference and clinical samples were used to evaluate the
34 performance of our nanopore sequencing assay. Triplet-primed PCR and PacBio long-read
35 sequencing were used for comparisons.

36 **RESULTS:** Nanopore sequencing identified reference carrier samples with a full range of
37 premutation alleles in single-, 10-, and 100-plex assays, and identified AGG interruptions in
38 an allele-specific manner. Nanopore sequencing revealed no size preference for amplicons
39 containing different length CGG repeat regions. Finally, nanopore sequencing successfully
40 identified three carriers among ten clinical samples for preliminary clinical validation. The
41 observed variation in CGG repeat region size resulted from the base-calling process of
42 nanopore sequencing.

43 **CONCLUSIONS:** Our nanopore sequencing assay is rapid, high-capacity, inexpensive, and
44 easy to perform, thus providing a promising tool and paving the way for population-based
45 FXS carrier screening.

46 **Introduction**

47 Fragile X syndrome (FXS) (#MIM300624) is the leading cause of monogenic autism
48 spectrum disorders and inherited intellectual disability, affecting approximately 1/7,000
49 females and 1/4,000 males worldwide (1, 2). The causative gene of FXS is fragile X
50 messenger ribonucleoprotein 1 (*FMRI*), an X-linked dominant gene with full penetrance in all
51 males and many females that plays a fundamental role in synapse formation and normal
52 dendrite development (3, 4). *FMRI* can be categorized into four allelic forms based on the
53 number of CGG trinucleotide repeats in its 5' untranslated region: (a) normal alleles contain
54 ~5 to ~44 repeats and are stable in meiosis or mitosis; (b) intermediate alleles contain ~45 to
55 ~54 repeats and during intergenerational transmission, their repeat number can change slightly;
56 (c) premutation alleles contain ~55 to ~200 repeats and can expand to full mutation during
57 maternal transmission to offspring; premutation alleles are associated with risk of fragile
58 X-associated primary ovarian insufficiency, fragile X-associated tremor/ataxia syndrome, and
59 fragile X-associated neuropsychiatric disorders; (d) full mutation alleles have more than 200
60 repeats that frequently accompany hypermethylation in adjacent CpG islands and in the repeat
61 region itself, which silences transcription and causes most cases of FXS (4). Asymptomatic
62 women with pre-mutation or full mutation alleles are thus termed FXS carriers because their
63 offspring are at risk of FXS (5). The prevalence of FXS carriers varies among different
64 populations, ranging from approximately 1/149 in Israel to approximately 1/581 in East Asia
65 (6-9). In many cases, CGG repeats are interrupted by one or more AGG trinucleotides (i.e.,
66 AGG interruptions), which can prevent strand slippage during replication, thus functioning as
67 a protective factor that decreases the risk of intergenerational CGG expansion (9). Evaluation

68 of AGG interruptions among CGG repeats is thus essential for genetic counseling, especially
69 for FXS carriers (4).

70 Owing to the severe morbidity of FXS, carrier screening, and subsequent prenatal
71 diagnosis are still warranted for this disease until the development of effective treatments (4,
72 10). Although all major ethnic groups and races appear to be susceptible to FXS, and FXS
73 carrier prevalence is high, whether FXS carrier screening should be offered to the general
74 population has long been a subject of debate (5, 9, 11-13). We previously showed that
75 population-based carrier screening is the dominant strategy for FXS intervention in East Asia
76 (9). Recently, population-based pan-ethnic FXS carrier screening was officially endorsed by
77 the American College of Medical Genetics and Genomics (10). Clinical interest has shifted
78 from “whether to perform” to “how to perform” population-based FXS carrier screening. An
79 assay to capably address this issue should be able to examine a large number of female
80 samples with high accuracy, low cost, short turnaround time, and ease of performance.

81 FXS genetic testing was performed using Southern blotting, a labor-intensive method
82 that allows estimation of CGG expansion size but cannot accurately quantitate CGG repeat
83 number and AGG interruption patterns (4). These drawbacks have been overcome with the
84 triplet-primed PCR (TP-PCR) assay, which has become the mainstay method in clinical use,
85 with several kits, including AmpliDeX™ *FMR1* PCR Kit (Asuragen) and Molecular Fragile X
86 PCR Kit (Biofast) commercially available (4, 14). Typing AGG interruptions at the allelic
87 level in female samples with TP-PCR remains challenging. Long-read sequencing-based
88 assays for FXS genetic testing using Nanopore and PacBio sequencing have recently been
89 developed (15-19). Theoretically, with the ability to directly sequence entire CGG repeats

90 and/or the full-length *FMR1* gene, long-read sequencing-based assays enable quantitative
91 evaluation of CGG repeat number, AGG interruptions, rare intragenic variants, and large
92 deletions in a single allele, dramatically advancing genetic diagnosis of FXS (16). To date, the
93 clinical utility of long-read sequencing-based assays for population-based screening of FXS
94 carriers has not been demonstrated.

95 In this study, we developed a nanopore sequencing assay that includes data analysis
96 software and demonstrated its clinical potential as a promising tool for population-based FXS
97 carrier screening.

98 **Materials and Methods**

99 **SAMPLES**

100 Eight (P2, P3, P4, P5, N1, N2, N3, and N4) and two (NA06968 and NA20239) reference
101 samples with known *FMR1* genotypes were obtained from the National Institutes for Food
102 and Drug Control of China and the Coriell Institute, respectively (20, 21). P3, P4, NA06968,
103 and NA20239 provide genomic DNA from FXS carriers with different extents of CGG
104 expansion, and N1, N2, N3, and N4 provide genomic DNA from women with normal *FMR1*
105 alleles. N1/P2 and N1/P5 mixtures represent genomic DNA from women carrying CGG
106 repeats near the intermediate/premutation boundary (Supplemental Table 1).

107 The genomic DNA of ten women whose *FMR1* genotypes have been examined via
108 clinical FXS genetic testing was obtained from the Department of Prenatal Diagnosis,
109 Women's Hospital of Nanjing Medical University. Signed informed consent was obtained
110 from each participant to authorize the use of their genetic data for research purposes. This
111 study was approved by the Research Ethics Committee of Xiamen University.

112 **AMPLIFICATION OF *FMRI* ALLELES**

113 The CGG repeats and flanking sequences of *FMRI* were amplified using a T100 Thermal
114 Cyclor (Bio-Rad, Hercules, CA, USA). A 50- μ L reaction containing 1 \times Expand Long
115 Template buffer 2 (Roche Diagnostics, Mannheim, Germany), 3.75U Expand Long Template
116 Enzyme mix (Roche Diagnostics, Mannheim, Germany), 0.5 mmol/L dNTPs (Takara, Kyoto,
117 Japan), 2.2 mol/L Betaine (SIGMA-Aldrich, Saint Louis, Missouri, USA), 0.33 mmol/L of
118 each forward and reverse primer (Sangon, Shanghai, China), and 50 ng DNA template. Primer
119 information and amplification conditions are listed in Supplemental Tables 2 and 3,
120 respectively. After amplification, PCR products were purified with a TIANquick Midi
121 Purification Kit (TIANGEN, Beijing, China) and subsequently quantified with a QubitTM
122 dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA) on a QubitTM fluorometer
123 (ThermoFisher Scientific, Waltham, Massachusetts, USA) according to the respective
124 manufacturers' instructions.

125 **LIBRARY PREPARATION AND NANOPORE SEQUENCING**

126 Library preparation and nanopore sequencing methods are described in detail in Supplemental
127 Methods.

128 **DATA ANALYSIS SOFTWARE**

129 Based on the previously described STRique software (15), we developed data analysis
130 software to facilitate the identification of CGG repeat numbers and AGG interruptions for
131 FXS carrier screening in our nanopore sequencing assay. The logic of software is described in
132 detail in Supplemental Methods; the software code was licensed by Xiamen University and is
133 available at <https://github.com/guoqiwei-xmu/FXS-carrier-identifier.git>.

134 **TP-PCR ASSAY**

135 A CE-certified AmplideX™ FMR1 PCR Kit (Asuragen, Minneapolis, Minnesota, USA) and a
136 Chinese Food and Drug Administration-approved Molecular Fragile X PCR Kit (Biofast,
137 Xiamen, Fujian, China) were used as comparative methods. Respective assays were
138 performed on a T100 Thermal Cycler (Bio-Rad, Hercules, CA, USA) and a 3500 DX Genetic
139 Analyzer (Applied Biosystems, Waltham, Massachusetts, USA), according to the
140 manufacturers' instructions.

141 **CAFXS ASSAY**

142 A long-read sequencing-based method termed the CAFXS assay, was used as a comparative
143 method and performed by a commercial service (Berry Genomics, Beijing, China). The
144 principle of CAFXS has been previously described (16).

145 **Results**

146 **DESIGN OF NANOPORE SEQUENCING ASSAY FOR FXS CARRIER SCREENING**

147 As illustrated in Fig. 1, our nanopore sequencing assay comprises three steps: pre-sequencing
148 preparation, nanopore sequencing, and post-sequencing data analysis. During pre-sequencing
149 preparation, CGG repeats and flanking sequences of each sample are amplified with a pair of
150 primers containing a specific identifier sequence at each strand's 5' end. These primers allow
151 the specific identifier sequence to be integrated into the amplicon of each sample to enable the
152 analysis of multiple samples in a single sequencing assay. After purification and quantification,
153 amplicons are pooled to achieve equal masses to prepare the sequencing library. After
154 nanopore sequencing of the library, post-sequencing data analysis is performed using novel
155 data analysis software. Briefly, the nucleotide sequence of each read is translated from a raw

156 ionic current signal (base calling), and qualified reads that contain correct identifier, prefix,
157 and suffix sequences, are retained for further analysis. Next, the CGG repeat numbers of
158 qualified reads are evaluated based on the distance between prefix and suffix sequences. After,
159 the presence and position of AGG interruptions are evaluated for each qualified read. For each
160 sample, the distribution of reads with different CGG repeat numbers and AGG statuses is
161 analyzed. Finally, the CGG repeat number and AGG status of a specific allele are determined
162 based on the peaks generated from the aggregated reads with similar CGG repeat numbers
163 and AGG patterns. When an allele with a CGG repeat number over the threshold (e.g., 54 or
164 55) is identified, an alarm is triggered to flag an FXS carrier.

165 **VALIDATION OF NANOPORE SEQUENCING ASSAY BY EXAMINING FXS**

166 **CARRIERS WITH VARIOUS CGG REPEAT NUMBERS AND AGG**

167 **INTERRUPTION PATTERNS**

168 We first validated the nanopore sequencing assay by examining FXS carriers with various
169 numbers of CGG repeats and AGG interruption patterns in a single-plex manner. Reference
170 samples were used as FXS carriers, and TP-PCR and CAFXS were used as methods for
171 comparison. As shown in Fig. 2, nanopore sequencing identified a wide range of expanded
172 alleles. AGG interruption patterns were identified in an allele-specific manner. A slight
173 discordance between nanopore sequencing and comparative method results was observed in
174 the reported CGG repeat numbers of expanded alleles (Table 1). Due to a one CGG repeat
175 difference between methods, the N1/P5 sample was designated as a non-carrier by the
176 nanopore sequencing assay but a carrier by comparative methods.

177 **DETECTION OF FXS CARRIERS FROM WILDTYPE BACKGROUNDS WITH**

178 **THE MULTIPLEX NANOPORE SEQUENCING ASSAY**

179 We next evaluated whether the nanopore sequencing assay could examine multiple samples
180 simultaneously and distinguish FXS carriers from wild-type backgrounds, as shown in Fig. 1.
181 Ten pairs of primers labeled with different identifier sequences were designed, validated, and
182 used to amplify ten respective samples (Fig. 3A and Supplemental Table 2). As illustrated in
183 Fig. 1, amplicons were pooled to achieve equal masses for library preparation and nanopore
184 sequencing. Nanopore sequencing successfully detected a wide range of expanded alleles in
185 the 10-plex assay (Fig. 3B-G).

186 We assessed whether the nanopore sequencing assay could simultaneously examine more
187 samples (e.g., 100 samples). Instead of using 100 identifier sequences to identify 100 different
188 samples, we mimicked the scenario by decreasing the pooling proportion of the six target
189 amplicons to 6% (1% each), while increasing the pooling proportion of the four wild-type
190 amplicons to 94% (23.5% each). Similar to the results of the 10-plex assay, the 100-plex assay
191 successfully detected all expanded alleles within wild-type backgrounds (Fig. 3H-M).

192 **INACCURATE BASE CALLING ATTRIBUTED TO VARIATIONS IN CGG REPEAT** 193 **QUANTIFICATION**

194 Although the nanopore sequencing assay was able to detect various FXS carriers, the reported
195 CGG repeat numbers of expanded alleles differed slightly from those of the comparative
196 methods (Table 1). Moreover, for a specific reference sample, the reported CGG repeat
197 numbers of the expanded alleles were slightly discordant among the different nanopore
198 sequencing assays (Fig. 4A). To investigate the potential causes of these variations, we
199 manually examined the sequences of 20 random reads with fewer CGG repeats than those

200 expected in the nanopore sequencing assay. As shown in Fig. 4B, sequence variants due to
201 inaccurate base calling were distributed in these reads, including prefix, suffix, and CGG
202 repeat sequences, resulting in the miscalculation of CGG repeat numbers in the subsequent
203 data analysis processes. We analyzed the Fastq data (i.e., data after base calling) for N1/P2
204 and N1/P5 reference samples derived from the CAFXS assay using our data analysis software,
205 and the reported CGG repeat numbers of these reference samples were concordant between
206 our data analysis pipeline and the CAFXS assay (Supplemental Fig. 1) confirming the
207 accuracy of post-base-calling processes using our data analysis software. Lastly, we
208 reanalyzed N1/P2 and N1/P5 reference samples based on sequencing data derived from the
209 sense and antisense strands, respectively, and confirmed that there was no strand bias in CGG
210 repeat quantification (Fig. 4C). Collectively, the observed variations in CGG repeat
211 quantification are thus mainly attributed to inaccurate base-calling.

212 **NANOPORE SEQUENCING DEMONSTRATES NO SIZE PREFERENCES FOR** 213 **AMPLICONS WITH DIFFERENT CGG REPEATS**

214 Although the nanopore sequencing assay could detect FXS carriers with a wide range of
215 expanded alleles, we noticed that for a specific carrier, the read ratios of expanded alleles
216 versus their wild-type counterparts tended to decrease with increasing allelic expansion (Fig.
217 4D). We thus evaluated whether the short amplicons of the wild-type alleles more readily
218 passed through the nanopore than the longer amplicons of the expanded alleles. Two libraries
219 derived from reference samples, N1 (29 CGG repeats) and P5 (55 CGG repeats) were pooled
220 in equimolar amounts and sequenced. The read ratio of P5 to N1 was approximately 1.06,
221 suggesting that there was no size preference for amplicons with different CGG repeats

222 (Supplemental Fig. 2). Moreover, we compared the read ratios of the expanded alleles with
223 their wild-type counterparts using single-plex, 10-plex, and 100-plex nanopore sequencing
224 assays. As shown in Fig. 4E, the read ratios of each sample were similar among the three
225 assays, suggesting that in the presence of different wild-type backgrounds, large amounts of
226 short amplicons did not skew the target size-indiscriminative characteristics of nanopore
227 sequencing. Finally, we evaluated the same samples on different nanopore sequencing
228 platforms, namely MinION and PromethION, which had 1190 and 8657 active nanopores on
229 their flow cells at the beginning of sequencing, respectively. As shown in Fig. 4F, the read
230 ratios were similar between the two platforms, suggesting that the target-size indiscriminatory
231 characteristics of nanopore sequencing were not influenced by the number of active
232 nanopores in the flow cells.

233 **PRELIMINARY VALIDATION OF CLINICAL UTILITY OF NANOPORE** 234 **SEQUENCING ASSAY FOR THE FXS CARRIER SCREENING**

235 To validate the clinical utility of the nanopore sequencing assay for FXS carrier screening, we
236 analyzed 10 female clinical samples using a 10-plex nanopore sequencing assay. Considering
237 the variations in CGG repeat quantification, when an allele with ≥ 54 CGG repeats was
238 identified, an alert would be triggered designating a carrier. These samples were analyzed in
239 parallel with CAFXS assay data. As shown in Table 2 and Supplemental Fig. 3, three carriers
240 were identified using the nanopore sequencing assay, concordant with the results of the
241 CAFXS assays.

242 **Discussion**

243 Given that the value of population-based FXS carrier screening has been acknowledged (10),

244 appropriate screening methods are urgently needed to implement such screening programs. A
245 favorable testing method should be able to examine large-scale samples with high accuracy,
246 low cost, short turnaround times, and ease of performance. Our nanopore sequencing assay
247 would be one option that can meet these demands. First, we developed reliable and
248 user-friendly software to facilitate sequencing data analysis and output results. Second, using
249 reference and clinical samples, we demonstrated that the nanopore sequencing assay was able
250 to identify the full range of premutation *FMRI* alleles, which account for the majority of FXS
251 carriers, and the nanopore sequencing assay was able to reliably quantify CGG repeats and
252 AGG interruptions, which are two important determinants for assessing the risk of full
253 mutation expansion in carriers, thus facilitating genetic counseling for carrier screening. Third,
254 using identifier sequences, we tested multiple samples in one assay, dramatically increasing
255 test capacity and decreasing cost. In our 100-plex assay, the cost per sample was less than
256 USD 10. Moreover, nanopore sequencing demonstrated no size preferences for amplicons
257 with different CGG repeats, suggesting that the detection of expanded alleles will not be
258 compromised by wild-type allele abundance as long as sufficient data are collected for each
259 sample, while tens of thousands of identifier sequences are available in the form of different
260 short combinations of nucleotides. In this regard, we could further increase sample capacity
261 and decrease cost per sample by collecting more data using an advanced nanopore sequencing
262 platform, such as PromethION. Fourth, the nanopore sequencing assay can be easily
263 accomplished with a short turnaround time (for example, a 100-plex assay requires
264 approximately two days) by a technician in a regular molecular diagnostics laboratory. Lastly,
265 in comparison with TP-PCR and CAFXS, the equipment for nanopore sequencing assays is

266 more available and portable, and thus could be more readily and widely adapted, particularly
267 in underdeveloped regions of the world (22).

268 In comparison with NGS or PacBio-based long-read sequencing, nanopore sequencing
269 has a relatively higher sequencing error rate in single reads owing to inaccuracies in ionic
270 current signal-to-nucleotide sequence translation (i.e., the base calling process) (22). This
271 innate error would cause one to several copy number variations during quantification of CGG
272 repeats in our nanopore assays, which may cause misclassification of subjects whose CGG
273 repeat numbers are near the boundary between intermediate and premutation alleles.
274 Technical limitation-associated variations in CGG repeat sizing are common to most methods
275 and are allowed to a certain extent in technical standards (4). The accuracy of CGG repeat
276 sizing can be improved, rendering it not a major concern in the use of our nanopore
277 sequencing assay. For example, base calling is educable with machine learning models, and
278 its accuracy continuously increases with the accumulation of nanopore sequencing data
279 (23-25). Even at the current stage, a reference sample, such as N1/P5, can be used as a
280 quantitative control to monitor variation in base calling and direct the threshold setting for
281 carrier identification.

282 In comparison with amplification-free methods (15, 18), amplification of *FMRI* alleles
283 drastically increases the sequencing depth of the target region and decreases costs. However,
284 owing to the high GC content of the target region and competitive effects of wild-type alleles,
285 the amplification efficiencies of expanded alleles tend to decrease with increasing allelic
286 expansion (4). In this study, carriers with a full range of premutation alleles were identified by
287 our nanopore sequencing assay; however, carriers with full mutation alleles, which constitute

288 approximately 3% of FXS carriers (9), could be missed by our assay. This issue can be
289 partially resolved by further optimizing the amplification system through the optimization of
290 reagents or reaction conditions. While the sensitivity of mosaicism detection has not been
291 systematically examined because its clinical significance in carrier screening remains unclear,
292 this sensitivity could be compared between nanopore sequencing and TP-PCR (~5-10%) (26)
293 using the mosaicism detection results of reference sample P3.

294 In conclusion, we developed a nanopore sequencing-based assay to identify FXS carriers
295 that includes data analysis software. This assay was demonstrated to be rapid, high-capacity,
296 inexpensive, and easy to perform, thus providing a promising tool for population-based FXS
297 carrier screening.

298 **Conflict of interest**

299 None

300 **Research Funding**

301 This work was supported by the National Natural Science Foundation of China [grant
302 numbers 82071662] and the Natural Science Foundation of Xiamen Municipality [grant
303 numbers 3502Z20227141] .

304 **Acknowledgement**

305 We thank all the clinical participants, and I-Fan Chiu, Maoli Chen, Sihao Wu, Hanwei Wang,
306 and Lingfeng Mao for their kind supports to this work.

307 **References**

308 1. Mila M, Alvarez-Mora MI, Madrigal I, Rodriguez-Revenga L. Fragile x syndrome: An
309 overview and update of the *fmr1* gene. *Clinical Genetics* 2017;93:2:197-205 doi:

- 310 10.1111/cge.13075.
- 311 2. Lozano R, Azarang A, Wilaisakditipakorn T, Hagerman RJ. Fragile x syndrome: A review
312 of clinical management. *Intractable & Rare Diseases Research* 2016;5:3:145-57 doi:
313 10.5582/irdr.2016.01048.
- 314 3. Pieretti M, Zhang F, Fu Y-H, Warren ST, Oostra BA, Caskey CT, et al. Absence of
315 expression of the *fmr-1* gene in fragile x syndrome. *Cell* 1991;66:4:817-22 doi:
316 10.1016/0092-8674(91)90125-i.
- 317 4. Spector E, Behlmann A, Kronquist K, Rose NC, Lyon E, Reddi HV. Laboratory testing for
318 fragile x, 2021 revision: A technical standard of the american college of medical genetics
319 and genomics (acmg). *Genet Med* 2021;23:5:799-812. Epub 20210401 doi:
320 10.1038/s41436-021-01115-y.
- 321 5. Committee opinion no. 691: Carrier screening for genetic conditions. *Obstet Gynecol*
322 2017;129:3:e41-e55. Epub 2017/02/23 doi: 10.1097/AOG.0000000000001952
323 00006250-201703000-00048 [pii].
- 324 6. Zlotogora J, Grotto I, Kaliner E, Gamzu R. The israeli national population program of
325 genetic carrier screening for reproductive purposes. *Genetics in Medicine* 2016;18:2:203-6
326 doi: 10.1038/gim.2015.55.
- 327 7. Westemeyer M, Saucier J, Wallace J, Prins SA, Shetty A, Malhotra M, et al. Clinical
328 experience with carrier screening in a general population: Support for a comprehensive
329 pan-ethnic approach. *Genetics in Medicine* 2020;22:8:1320-8 doi:
330 10.1038/s41436-020-0807-4.
- 331 8. Archibald AD, Smith MJ, Burgess T, Scarff KL, Elliott J, Hunt CE, et al. Reproductive

- 332 genetic carrier screening for cystic fibrosis, fragile x syndrome, and spinal muscular
333 atrophy in australia: Outcomes of 12,000 tests. *Genet Med* 2018;20:5:513-23. Epub
334 20171026 doi: 10.1038/gim.2017.134.
- 335 9. Guo Q, Chang YY, Huang CH, Hsiao YS, Hsiao YC, Chiu IF, et al. Population-based
336 carrier screening and prenatal diagnosis of fragile x syndrome in east asian populations. *J*
337 *Genet Genomics* 2021;48:12:1104-10. Epub 20210607 doi: 10.1016/j.jgg.2021.04.012.
- 338 10. Gregg AR, Aarabi M, Klugman S, Leach NT, Bashford MT, Goldwaser T, et al. Screening
339 for autosomal recessive and x-linked conditions during pregnancy and preconception: A
340 practice resource of the american college of medical genetics and genomics (acmg).
341 *Genetics in Medicine* 2021;23:10:1793-806 doi: 10.1038/s41436-021-01203-z.
- 342 11. Dimmock DP. Should we implement population screening for fragile x? *Genetics in*
343 *Medicine* 2017;19:12:1295-9 doi: 10.1038/gim.2017.81.
- 344 12. Metcalfe SA, Delatycki MB, Cohen J, Archibald AD, Emery JD. Fragile x population
345 carrier screening. *Genetics in Medicine* 2018;20:9:1091-2 doi: 10.1038/gim.2017.209.
- 346 13. Metcalfe SA, Martyn M, Ames A, Anderson V, Archibald AD, Carter R, et al. Informed
347 decision making and psychosocial outcomes in pregnant and nonpregnant women offered
348 population fragile x carrier screening. *Genetics in Medicine* 2017;19:12:1346-55 doi:
349 10.1038/gim.2017.67.
- 350 14. Juusola JS, Anderson P, Sabato F, Wilkinson DS, Pandya A, Ferreira-Gonzalez A.
351 Performance evaluation of two methods using commercially available reagents for
352 pcr-based detection of fmr1 mutation. *J Mol Diagn* 2012;14:5:476-86. Epub 20120702 doi:
353 10.1016/j.jmoldx.2012.03.005.

- 354 15. Giesselmann P, Brändl B, Raimondeau E, Bowen R, Rohrandt C, Tandon R, et al.
355 Analysis of short tandem repeat expansions and their methylation state with nanopore
356 sequencing. *Nat Biotechnol* 2019;37:12:1478-81. Epub 20191118 doi:
357 10.1038/s41587-019-0293-x.
- 358 16. Liang Q, Liu Y, Duan R, Meng W, Zhan J, Xia J, et al. Comprehensive analysis of fragile
359 x syndrome: Full characterization of the *fmr1* locus by long-read sequencing. *Clin Chem*
360 2022;68:12:1529-40. Epub 2022/09/29 doi: 10.1093/clinchem/hvac154
361 6726663 [pii].
- 362 17. Ardui S, Race V, Zablotskaya A, Hestand MS, Van Esch H, Devriendt K, et al. Detecting
363 agg interruptions in male and female *fmr1* premutation carriers by single-molecule
364 sequencing. *Human Mutation* 2017;38:3:324-31 doi: 10.1002/humu.23150.
- 365 18. Stevanovski I, Chintalaphani SR, Gamaarachchi H, Ferguson JM, Pineda SS, Scriba CK,
366 et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with
367 programmable targeted nanopore sequencing. *Sci Adv* 2022;8:9:eabm5386. Epub 20220304
368 doi: 10.1126/sciadv.abm5386.
- 369 19. Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, et al. Sequencing the
370 unsequenceable: Expanded cgg-repeat alleles of the fragile x gene. *Genome Research*
371 2013;23:1:121-8 doi: 10.1101/gr.141705.112.
- 372 20. Amos Wilson J, Pratt VM, Phansalkar A, Muralidharan K, Highsmith WE, Beck JC, et al.
373 Consensus characterization of 16 *fmr1* reference materials: A consortium study. *The Journal*
374 *of Molecular Diagnostics* 2008;10:1:2-12 doi: 10.2353/jmoldx.2008.070105.
- 375 21. Gao F, Huang W, You Y, Huang J, Zhao J, Xue J, et al. Development of chinese genetic

- 376 reference panel for fragile x syndrome and its application to the screen of 10,000 chinese
377 pregnant women and women planning pregnancy. *Molecular Genetics & Genomic*
378 *Medicine* 2020;8:6 doi: 10.1002/mgg3.1236.
- 379 22. Ameer A, Kloosterman WP, Hestand MS. Single-molecule sequencing: Towards clinical
380 applications. *Trends in Biotechnology* 2019;37:1:72-85 doi: 10.1016/j.tibtech.2018.07.013.
- 381 23. Singh G, Alser M, Denolf K, Firtina C, Khodamoradi A, Cavlak MB, et al. Rubicon: A
382 framework for designing efficient deep learning-based genomic basecallers. *Genome*
383 *Biology* 2024;25:1 doi: 10.1186/s13059-024-03181-2.
- 384 24. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and
385 challenges in long-read sequencing data analysis. *Genome Biology* 2020;21:1 doi:
386 10.1186/s13059-020-1935-5.
- 387 25. Alser M, Lindegger J, Firtina C, Almadhoun N, Mao H, Singh G, et al. From molecules to
388 genomic variations: Accelerating genome analysis via intelligent algorithms and
389 architectures. *Computational and Structural Biotechnology Journal* 2022;20:4579-99 doi:
390 10.1016/j.csbj.2022.08.019.
- 391 26. Hantash FM, Goos DG, Tsao D, Quan F, Buller-Burckle A, Peng M, et al. Qualitative
392 assessment of fmr1 (cgg)n triplet repeat status in normal, intermediate, premutation, full
393 mutation, and mosaic carriers in both sexes: Implications for fragile x syndrome carrier and
394 newborn screening. *Genetics in Medicine* 2010;12:3:162-73 doi:
395 10.1097/GIM.0b013e3181d0d40e.

396 **Figure 1. Flowchart of screen for fragile X syndrome carriers using our nanopore**
397 **sequencing assay.**

398 **Figure 2. Comparison of nanopore sequencing, TP-PCR, and CAFXS for detection of**
399 **different FXS carrier reference samples.**

400 **Figure 3. Detection of FXS carriers within wildtype backgrounds with multiplex**
401 **nanopore sequencing assay.** (A) Efficacy of ten pairs of primers labeled with different
402 identifier sequences was validated by amplification of wildtype samples. (B-G) CGG repeat
403 numbers associated with reference samples in the 10-plex nanopore sequencing assay.
404 (H-M) CGG repeat numbers associated with reference samples in the 100-plex nanopore
405 sequencing assay.

406 **Figure 4. Evaluation of length variation and length preference of the nanopore**
407 **sequencing assay.** (A) Slight variations were observed in reported CGG repeat numbers for
408 expanded alleles in reference samples among single-plex, 10-plex, and 100-plex nanopore
409 sequencing assays. (B) Representative reads with sequence variants due to inaccurate base
410 calling. (C) Reported CGG repeat numbers of expanded alleles were identical based on
411 reads derived from forward and reverse strands, respectively. (D) Read ratios of expanded
412 alleles versus their wild-type counterparts tend to decrease with increasing allelic expansion.
413 Error bars indicate standard deviation among triplicate samples. (E) Reads ratios of
414 expanded alleles versus their wildtype counterparts were similar among single-plex,
415 10-plex, and 100-plex nanopore sequencing assays. (F) Reads ratios of expanded alleles
416 versus their wildtype counterparts were similar between MinION and PromethION
417 platforms whose respective flow cells had 1190 and 8657 active nanopores.

418 **Table 1. Comparison of nanopore sequencing, TP-PCR, and CAFXS for the detection of different FXS carrier reference samples.**

Sample	Manufacturer's instruction	CGG repeat number				AGG interruption			
		TP-PCR	TP-PCR	Nanopore	CAFXS	TP-PCR	TP-PCR	Nanopore	CAFXS
		Asuragen	Biofast	sequencing		Asuragen	Biofast	sequencing	
N1/P2	29±1/54	29/54	29/54	29/53	29/54	2AGG	2AGG	9A9A9/none	9A9A9/none
N1/P5	29±1/56±1	29/55	29/55	29/54	29/55	3AGG	3AGG	9A9A9/11A42	9A9A9/11A43
P3	30±1/69±3	30/54/67	30/54/67	30/53/66	30/67	2AGG	2AGG	10A9A9/none/none	10A9A9/none
NA06968	32/107	33/111	33/111	33/110	33/113	1AGG	1AGG	9A23/none	9A23/none
P4	29±1/155±5	30/156	30/156	30/155	30/153	2AGG	2AGG	10A9A9/none	10A9A9/none
NA20239	20/183-193	20/>200	20/>200	20/203	20/202	3AGG	3AGG	10A9/9A9A183	10A9/9A9A182

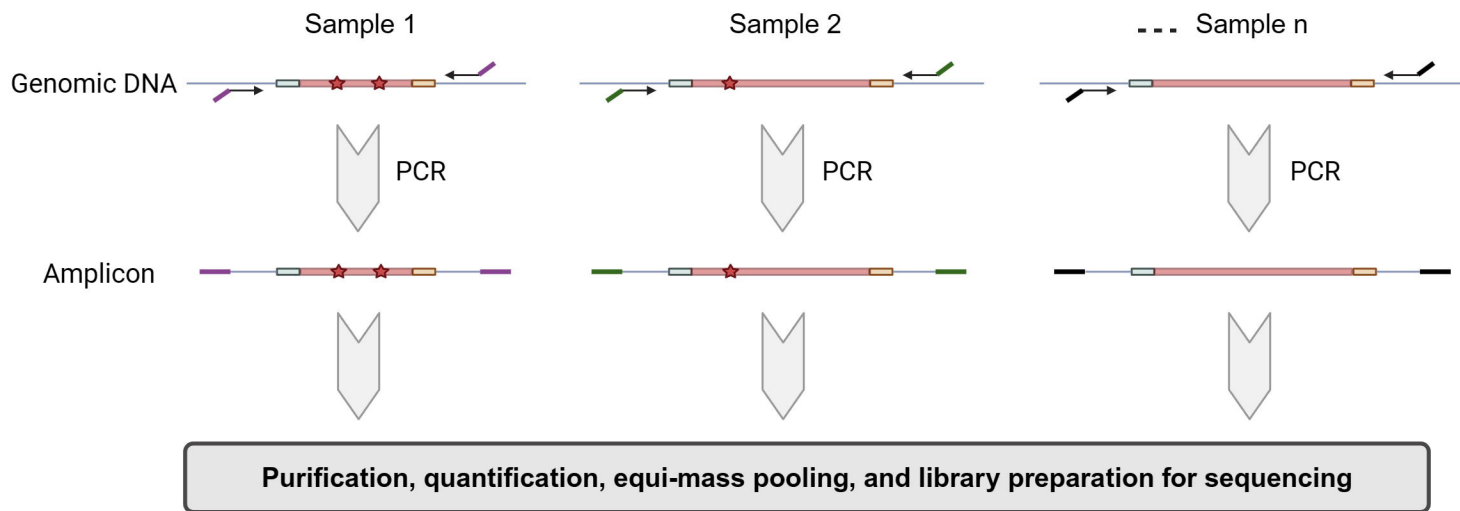
419

420 **Table 2. Evaluation of nanopore sequencing assay with clinical samples**

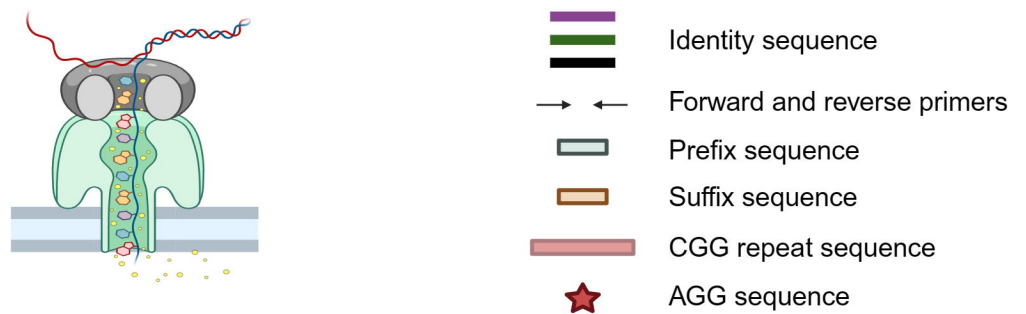
Sample	CGG repeat number		AGG interruption	
	Nanopore sequencing	CAFXS	Nanopore sequencing	CAFXS
1*	36/151	36/153	9A9A6A9/9A141	9A9A6A9/9A143
2	33/50	33/51	9A6A6A9/10A39	9A6A6A9/10A40
3	29/52	29/53	9A9A9/9A9A5A9A6A9	9A9A9/9A9A6A9A6A9
4*	29/54	29/55	9A9A9/9A9A10A6A6A9	9A9A9/9A9A11A6A6A9
5*	36/60	36/61	9A9A6A9/10A49	9A9A6A9/10A50
6	29/29	29/29	9A9A9/9A9A9	9A9A9/9A9A9
7	20/30	20/30	10A9/none	10A9/none
8	29/29	29/29	9A9A9/9A9A9	9A9A9/9A9A9
9	36/36	36/36	9A9A6A9/9A16A9	9A9A6A9/9A16A9
10	26/29	26/29	10A9A5/9A9A9	10A9A5/9A9A9

421 *FXS carrier

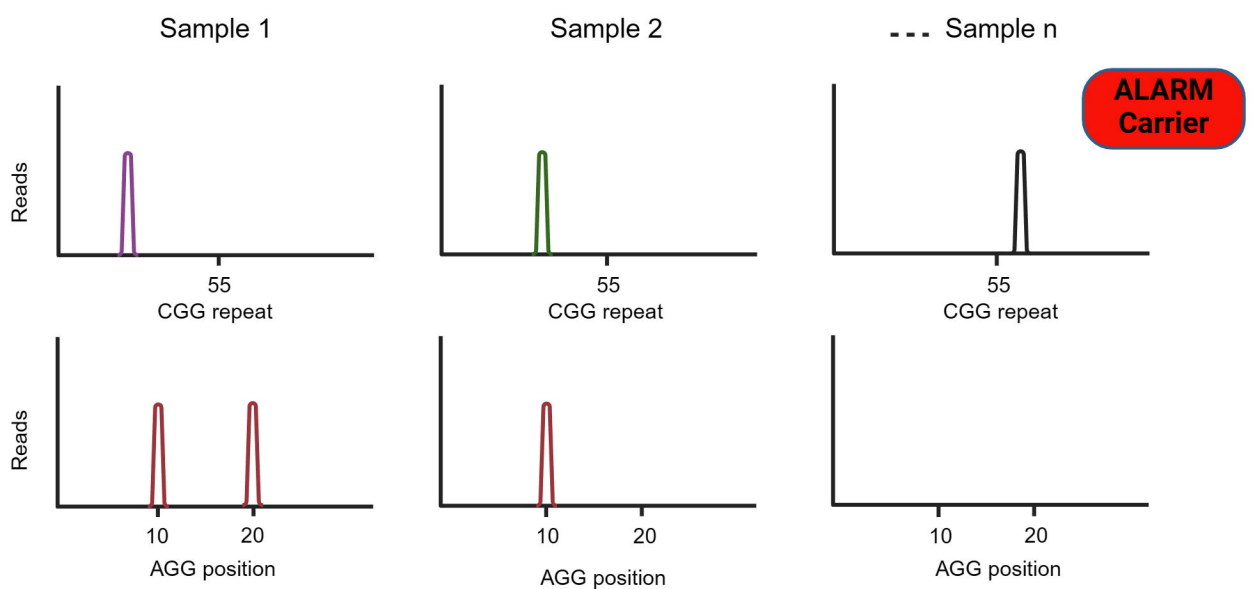
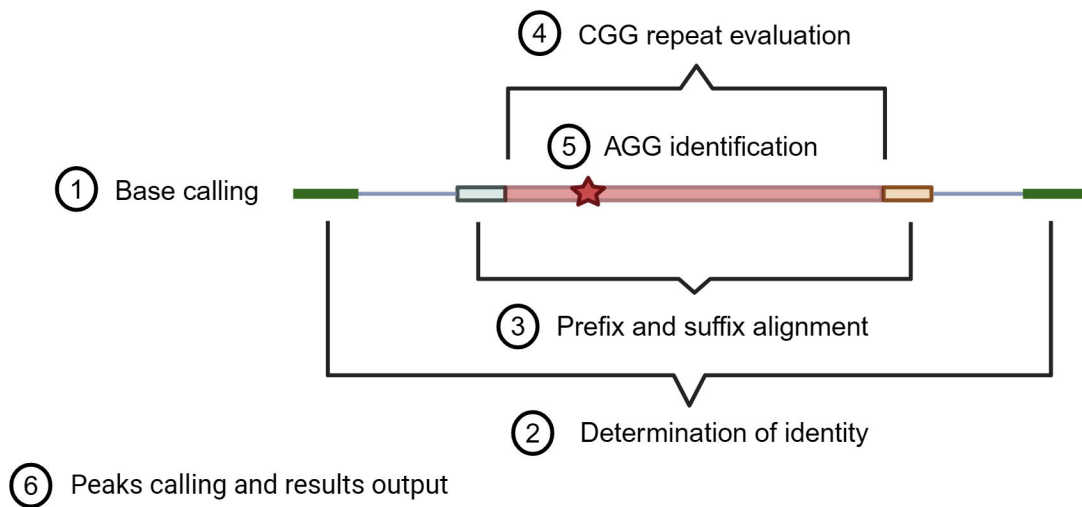
Pre-sequencing preparation



Nanopore sequencing



Post-sequencing data analysis

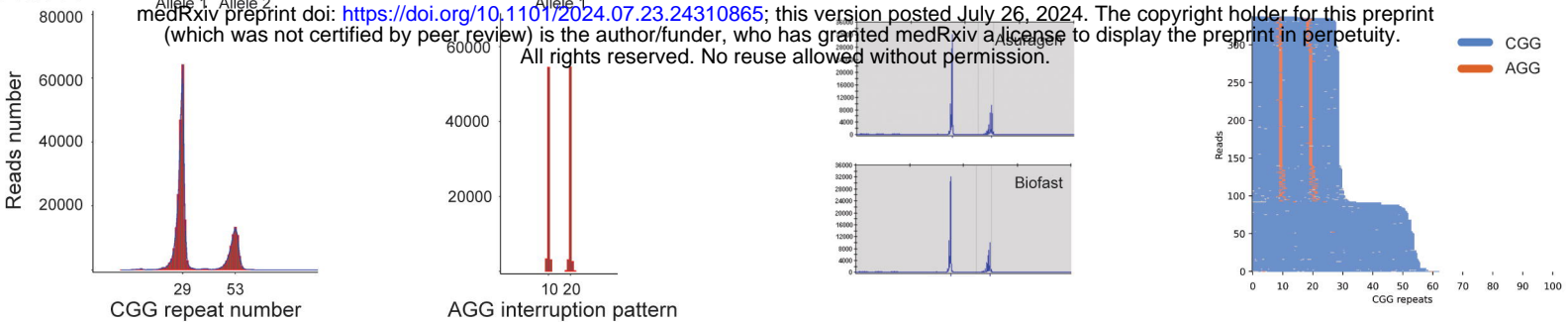


Nanopore sequencing

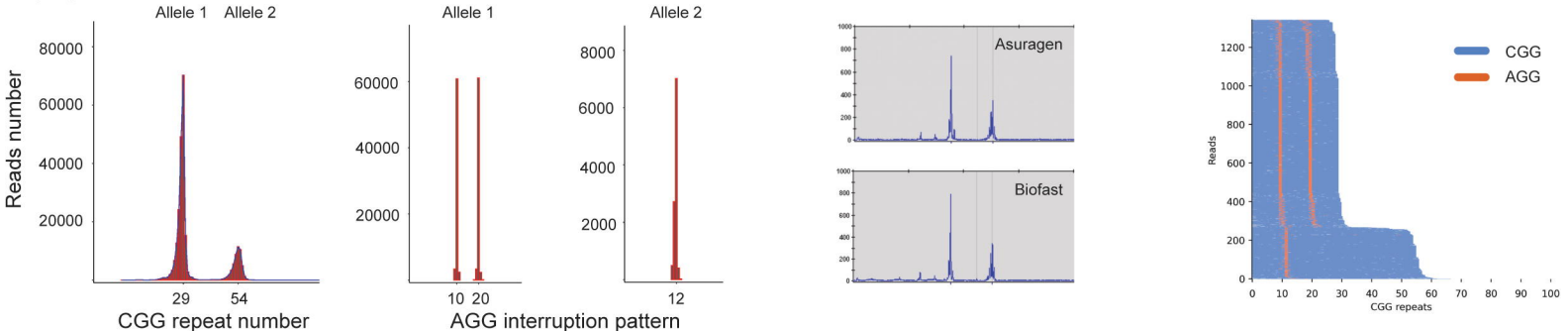
TP-PCR

CAFXS

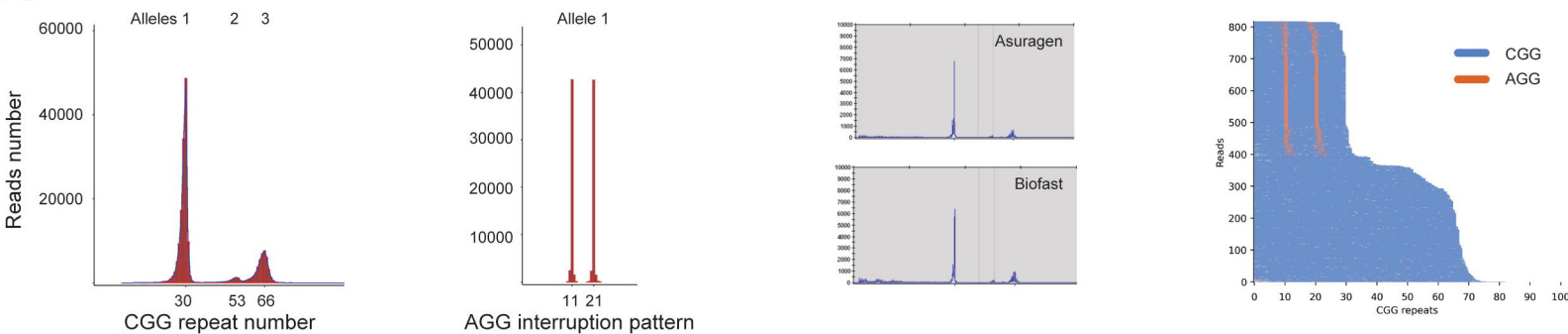
N1/P2



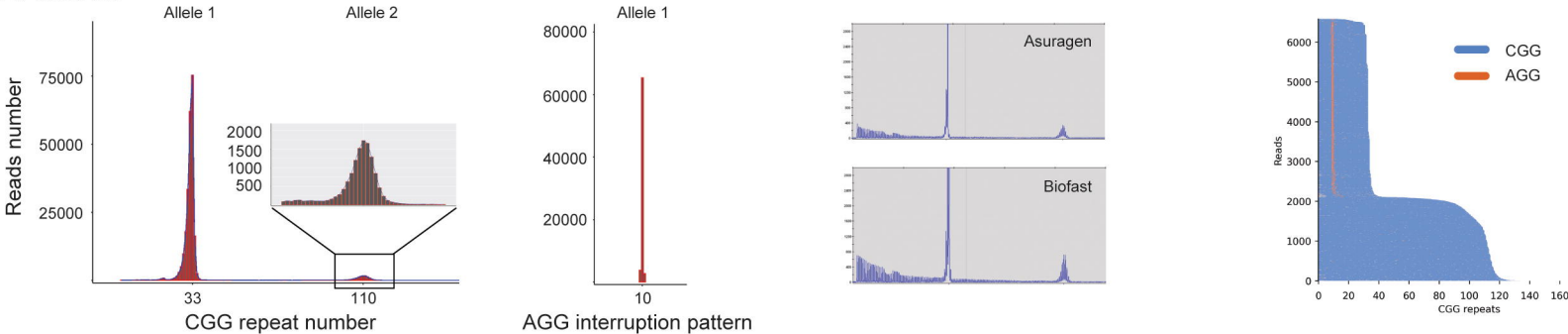
N1/P5



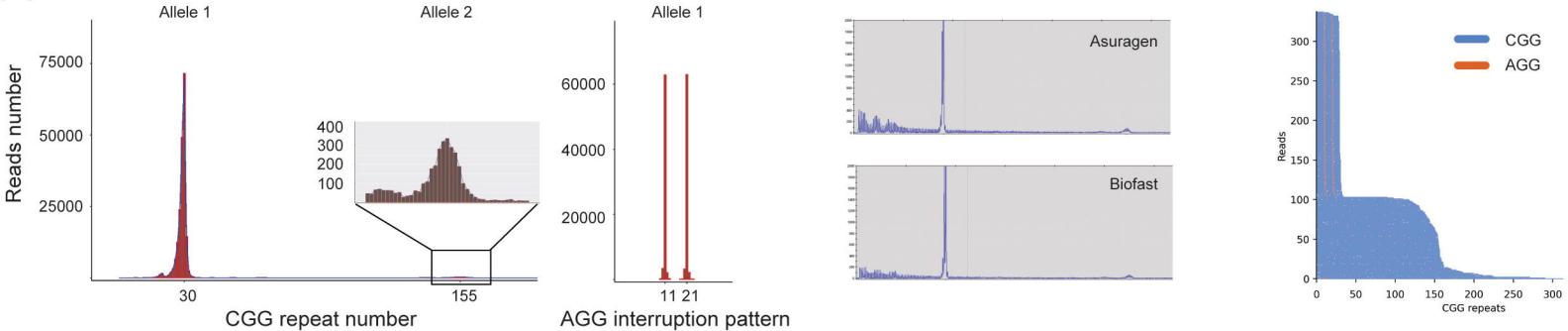
P3



NA06968



P4



NA20239

