TITLE

A multi-model approach defines function altering MECP2 missense variants identified in individuals with autism spectrum disorder.

AUTHOR LIST

Eric Chen[1]*, Jessica Schmitt[1]*, Graeme McIntosh[1]*, Ajay Singh[2]*, Barry Young[1], Tianshun Lian[1], Jie Liu[1], Vedanta Khan[2], Ke Chen[3], J Beatrice Liston[1], Lily MacDonald[1], Bill Wang[1], Sonia Medina Giro[1], Benjamin Boehme[1], Mriga Das[1], Seevasant Indran[1], Sanja Rogic[3], Paul Pavlidis[3], Jesse T Chao[2], Douglas W Allan[1]#, and Christopher JR Loewen[1]#

[1]Department of Cellular and Physiological Sciences, The Life Sciences Institute, The University of British Columbia, Vancouver, BC, Canada

[2]Department of Medical Biophysics, Sunnybrook Research Institute, The University of Toronto, Toronto, ON, Canada

[3]Department of Psychiatry, The Michael Smith Institute, The University of British Columbia, Vancouver, BC, Canada

(co-first *, co-corresponding #)

ABSTRACT

MECP2 is commonly mutated in Rett syndrome, where MECP2's function as a DNA cytosine methylation reader is believed critical. MECP2 variants are also catalogued in individuals with autism spectrum disorder (ASD), including nine missense variants with no known clinical significance. To assess these nine as risk alleles for ASD, we developed MECP2 variant function assays using yeast, *Drosophila* and human cell lines. We calibrated these assays with known reference pathogenic and benign variants. Our data predict that four ASD variants are loss of function (LoF) and five are functional. Protein destabilization or nuclear delocalization offers insight into the altered function of a number of these variants. Notably, yeast and *Drosophila* lack DNA methylation, yet all Rett reference pathogenic and ASD variants in the methyl DNA binding domain that we analyzed proved to be LoF, suggesting a clinically-relevant role for non-methyl DNA-binding by MECP2.

34 **INTRODUCTION**

35 The X-linked protein MECP2 was first identified in screens for proteins recruited to methylated
36 DNA [1], and LoF alleles of MECP2 were subsequently found as causative in the
37 neurodevelopmental disorder, Rett syndrome [2]. Since then, variants in MECP2 have been found
38 in most cases of Rett, but also in other neurodevelopmental disorders including autism spectrum
39 disorder (ASD) [3–6]. MECP2 has largely been examined as a DNA cytosine methylation reader
40 via its methyl cytosine binding domain (MBD) that recruits co-repressor complexes to DNA to
41 repress transcription via its transcriptional repressor domain (TRD) [7–10]. The MBD, and to a
42 lesser extent the TRD, harbors most Rett LoF variants [11], resulting in variants which are believed
43 to reduce binding to methylated DNA or reduce recruitment of co-regulator complexes,
44 respectively [12,13]. While numerous MECP2 variants have been found in patients that exhibit Rett
45 with ASD, recent reports have annotated six MECP2 missense variants in ASD without an
46 associated Rett diagnosis (R91W, P152L, G195S, A202G, E282G, S411R) [14–18]. It is unclear
47 whether any of these variants are function altering and clinically significant. Two of them, R91W
48 and A202G, are classified by ClinVar as variants of uncertain significance (VUS) in cases of
49 severe neonatal onset encephalopathy with microcephaly, and E282G is classified as likely
50 pathogenic. The remaining four variants have no current ClinVar classification.

51

52 While MECP2 is understood to primarily be a methyl-cytosine DNA (DNAm) binding protein, it
53 has become clear that it also binds unmethylated DNA with a preference for GC-rich sequences
54 [19–22]. This has been proposed as an alternate hypothesis for why MECP2 appears to preferentially
55 bind methylated GC-rich regions of the genome, as opposed to binding DNAm methylation *per*
56 *se*. It can be challenging to segregate DNAm-dependent from DNAm-independent functions of

57 MeCP2 *in vivo*, as DNAm is both abundant and essential in mammals. Consequently, it is still
58 unclear what fraction of MeCP2's functions are dependent on its interaction with methylated
59 and/or unmethylated DNA. Such ambiguity leaves unresolved the question of whether disease-
60 relevant variants in MECP2 may selectively disrupt binding to methylated versus unmethylated
61 DNA.
62
63 There are extremely low levels of methylated cytosine in the *Drosophila* genome, and none in
64 *Saccharomyces cerevisiae* [23,24]. Thus, *Drosophila* and yeast offer avenues to explore the
65 interactions of MECP2 with unmethylated DNA, and to test if Rett and ASD variants alter
66 MECP2 binding to unmethylated DNA. Moreover, while neither organism has an ortholog of
67 MECP2, both organisms are robust genetic models that have proven well-suited to comparing the
68 function of large numbers of gene variants [25–27]. In *Drosophila*, human MECP2 associates with
69 chromatin and mediates transcriptional repression via interactions with orthologs of known
70 MECP2 physical interactors in humans, including Sin3A, N-Cor, REST as well as other
71 chromatin modifiers [28]. In fly neurons, MECP2 impacts dendritic branching and behavioural
72 output [29,30] and C-terminal truncation mutants (R294X specifically) promotes neuronal apoptosis
73 [31]. In yeast, an accompanying article shows that MECP2 binds chromatin and mediates
74 widespread transcriptional repression (Brown et al.).

75 Our goal was to use the *Drosophila* and budding yeast models to test the functional impact of
76 these ASD MECP2 variants to determine whether they may confer autism risk, by virtue of their
77 altered function. Further, we complemented these functional studies with molecular
78 characterization of MECP2 variant stability and nuclear localization in human cell lines. We
79 calibrated our assays with MECP2 variants classified by ClinVar as pathogenic or benign, in

80 accordance with ACMG guidelines for experimental assays [32]; first to validate our assays for
81 variant assessment, and second to provide a rubric for determining if test variants are likely
82 pathogenic or likely benign. All reference variants functioned as predicted in the *Drosophila*
83 assay, indicating that we could functionally assess any test variants. In yeast, reference benign
84 variants functioned as predicted, but only reference pathogenic variants in the MBD functioned
85 as predicted; thus we consider the yeast assay to be exclusively predictive for test variants in the
86 MBD. Our analyses indicate that four of the nine test ASD variants are LoF, making these
87 variants candidate risk alleles for ASD. Interestingly, these variants mostly proved to be less
88 severe LoF than Rett pathogenics, providing a possible explanation for how MECP2 mutations
89 can lead to ASD but not Rett.

90

91 **METHODS**

92 **MECP2 isoform selection**

93 MECP2 is known to have two major isoforms, E1 and E2. We selected to work with MECP2
94 isoform E2 (ENST00000303391, NM_004992) based on the recommendation by the ClinGen
95 Rett/Angelman-like Expert Panel [33]. This transcript is annotated as the APPRIS principal isoform
96 in Ensembl genome browser, based on the annotations for protein structure, function and cross-
97 species conservation, and is the highest expressed transcript in the brain (GTEx portal).

98

99 **Variant aggregation, annotation and prioritization**

100 We utilized our in-house computational pipeline to aggregate MECP2-E2 missense variants from
101 ClinVar and VariCarta and perform comprehensive annotation. We collected ClinVar variants
102 (downloaded in January 2022) that have at least one gold review star, i.e. the review status

103 (CLNREVSTAT) has to have one of the following designations: “criteria provided, single
104 submitter”, “criteria provided, conflicting interpretations”, “criteria provided, multiple
105 submitters, no conflicts” or “reviewed by expert panel”. For each of these variants, we also
106 obtain its interpretation (clinical significance; CLNSIG) and reported condition (disease name;
107 CLNDN). The ASD missense variants were obtained from VariCarta, a database of variants
108 found in individuals diagnosed with ASD and reported in the scientific literature, that we created
109 [34]. We annotated the aggregated variants using the Ensembl VEP tool [35]. We also added protein
110 domain annotation relevant in the context of Rett Syndrome [36].
111

112 **MECP2-E2 cloning**

113 A *MECP2-E2* cDNA clone was obtained (Horizon, Clone Id: 3956518, # MHS6278-
114 202830559), and MECP2 coding sequences were amplified with iProof DNA polymerase (Bio-
115 Rad, California USA, #1725301) using the following primers:
116 GA-attL1-MECP2-522-F (tgtacaaaaaagcaggctccggcaccatggtagctgggatgttag)
117 GA-attL1-MECP2-522-R (ttgtacaagaaagctgggtctatcagctaactctctcggtc)
118 Amplicons were cloned in PCR8-GW-TOPO (Invitrogen, ThermoFisher Scientific, Waltham,
119 USA, catalog #K250020) using NEBuilder HiFi DNA Assembly Master Mix (New England
120 Biolabs, USA, # M5520). The resulting construct was named *MECP2-pENTRY*. A sequence of
121 CCGGCACC was introduced in primer GA-attL1-MECP2-522-F (underlined) preceding the
122 start codon for *MECP2-E2*. After LR cloning of MECP2-pEntry into pDEST vectors, this
123 CCGGCACC sequence can function as a Kozak sequence to initiate translation, or as an in-
124 frame linker for N-terminal peptide sequences. Missense variants were generated in *MECP2-*
125 *pENTRY* either using the Q5 mutagenesis kit (New England Biolabs, USA, # E0554) in house, or

126 were made by GenScript (New Jersey, USA). For MECP2 expression clone construction,
127 *MECP2-pENTRY* plasmids were subcloned in *pAG426GAL-ccdB* (a kind gift from a gift from
128 Susan Lindquist; Addgene plasmid # 14155), *pGW-HA.attB* [37] or *pDEST-CMV-N-EGFP*
129 (Addgene Plasmid #122842) respectively using the Gateway LR Clonase reaction (Thermo
130 Fisher Scientific, Waltham, USA, catalog # 11791020). All transgenic *UAS-MECP2* variants (on
131 *pGW-HA.attB*) *Drosophila* strains used were verified by Sanger sequencing (UBC NAPS,
132 Vancouver, Canada) or whole amplicon sequencing (Plasmidsaurus, CA, USA) (Supplemental
133 Data 1). All HEK variants (on *pDEST-CMV-N-EGFP*) were verified by Sanger (UBC NAPS,
134 Vancouver, Canada) or whole plasmid sequencing (Plasmidsaurus, CA, USA) (Supplemental
135 Data 1).
136
137 **Synthetic dosage lethality screening in yeast**
138 Synthetic dosage lethality screening of MECP2 was performed as described previously [27], with
139 the following modifications. Yeast strain Y7092 was transformed with plasmid pSI7338-
140 MECP2. Following mating of this strain with the yeast deletion collection, diploid selection,
141 sporulation and *MAT**a*** haploid selection, MECP2 expression was induced by plating on media
142 containing 10 nM beta-estradiol. After incubation for one day at 30 °C, these plates were
143 replicated on identical media. Plates were imaged and analyzed the following day. Control plates
144 were generated by replicating haploid cells on the same media in the absence of beta-estradiol.
145 Plate images were analyzed using *Balony* software as previously described [38].
146
147 **Yeast quantitative liquid growth assays**

148 Yeast strain JHY716 was transformed with various plasmids containing MECP2 variants placed
149 downstream of the *GAL1* promoter. Four independent clones were picked from transformation
150 plates and transferred to SD-URA media containing 2% raffinose, 2% galactose to induce
151 expression of MECP2. For each clone, a 2 mL culture in SD-URA with 2% raffinose, 2%
152 galactose was grown overnight at 30 °C in a shaking incubator. The following morning, cultures
153 were diluted to an OD600 of ~0.25 in the same media. After 6 hours of incubation at 30 °C, log
154 phase cultures were diluted to an OD600 of 0.125. Three aliquots of 200 µl were taken from each
155 culture and transferred to a well of a flat-bottomed 96-well dish. Growth of cultures was
156 measured overnight at 30 °C with shaking at 800 rpm in an Agilent Logphase 600 plate reader.
157 All plates contained strains expressing wild type MECP2 and the corresponding empty vector as
158 controls. The exponential growth constant k was determined for each well by fitting to an
159 exponential growth equation during the exponential phase (typically between 5 and 10 hours post
160 dilution).
161
162 **Protein abundance assays**
163 To determine the abundance of MECP2 variants in yeast, strains were grown in SD-URA media
164 containing 2% raffinose to early log phase. MECP2 expression was induced by the addition of
165 2% galactose to the growth medium. Cells were harvested three hours later by centrifugation and
166 stored at -80 °C for subsequent processing. Cell pellets were resuspended in Laemmli sample
167 buffer, and heated to 95 C for 5 minutes. Cell lysis was achieved by the addition of 50 µl acid-
168 washed glass beads followed by two rounds of disruption in a MP Biomedical FastPrep-24
169 homogenizer (maximum speed for 60 s with 5 minutes cooling between cycles). Samples (0.2
170 OD600 equivalents) were separated by SDS-PAGE on Bio-Rad 4-15% pre-cast gels and

171 transferred to nitrocellulose membranes. MECP2 was immunoblotted using Sigma antibody
172 (M9317), with PGK1 (Abcam AB113687) used as a loading control. For visualization,
173 secondary antibodies conjugated to Cy3 (MECP2) and AlexaFluor-488 (PGK1) were used and
174 imaged using a Bio-Rad ChemiDoc MP Imaging System. MECP2 band volume was normalized
175 to PGK1 band volume for each variant.
176
177 ***Drosophila* strains**
178 All flies were maintained on standard cornmeal food at 25C and 70% humidity unless otherwise
179 specified. *P{GAL4}A9* (BL8761) (referred to as *A9-GAL4*) and *Asx[XF23]*(BL6041) were
180 obtained from *Bloomington* Drosophila Stock Centre (Indiana, USA). *UAS-MECP2* reference
181 (*UAS-MECP2-ref*) and variant sequences (collectively termed *UAS-MECP2.#*) were integrated
182 into embryos of genotype *y[1] w[] P{y[+7.7]=nos-phiC31\int.NLS}X;*
183 *P{y[+t7.7]=CaryP}attP40 (referred to as attP40) by Genome ProLab (Quebec, Canada), and*
184 *maintained as w[1118]; P{y[+t7.7] w[+mC]=UAS-MECP2.#.GW.**}attP40**,** where # represents
185 the variant and * represents integrant line number (Supplemental Data 1).
186
187 **Crossing scheme**
188 Experimenters were double blinded to genotype throughout. Homozygous *UAS-MECP2* males
189 and a control *attP40* line were crossed to *A9-GAL4* virgins (ratio 4 males : 12 females), and
190 maintained at room temperature (~21C) for 72 hours. Crosses were then transferred into fresh
191 'assay' vials, and transferred each 24 hrs each for three days, generating three 'assay' vials per
192 cross. After removal of adults from each vial, vials were placed into a 29C incubator (70%

193 humidity). Three days after eclosion, adult flies were sorted by sex and placed in pure
194 isopropanol at -20C. All transgenic flies for comparison were tested in a single batch.
195
196 **Wing imaging**
197 Experimenters were double blinded to genotype throughout. Flies were removed from
198 isopropanol, dried on paper towels, and 10-25 left wings were removed per genotype per assay
199 vial (one wing per fly). Wings were mounted on microscope slides in Canadian Balsam Fir
200 (1.01691.0025, CAS-No: 8007-47-4, Sigma Aldrich, MA, USA), coverslipped and sealed with
201 nail polish. Slides were compressed while drying for 24 hrs. Mounted wings were scanned using
202 a Panoramic Midi II digital slide scanner (3DHISTECH, Budapest, Hungary). Scanned fly wing
203 images were cropped at 15x zoom using CaseViewer (3DHISTECH, Budapest,Hungary). Images
204 were imported into Imaris v10.0 software (Bitplane, MA, USA) for measurement of wing vein
205 area per wing area, using Labkit machine learning and Fiji software (University of Wisconsin-
206 Madison, USA). In Labkit, we developed "BIGSMOOTHCLASSYBOI", a machine learning
207 algorithm to automate measurement of wing vein area for calculation of wing vein area/wing
208 area (details provided in Supplemental Methods 1).
209
210 **Statistical analysis of variant effects**
211 Yeast growth constants (k) and fly relative wing vein area data were analyzed using methods
212 similar to Post et al. [26]. Briefly, linear mixed-effects models were fit to each data set using the
213 *lmerTest* R package [39] to estimate the phenotypic effects of each variant compared to the
214 *MECP2-ref* constructs, treating experimental batches as random effects. For visualization, values
215 were standardized to the range [0-1] where 0 represents the mean phenotype measured for an

216 empty vector, and 1 represents the mean phenotype with the *MECP2-ref* construct. Four
217 biological replicates, each with three technical replicates were performed for yeast variant
218 growth assays, while for *Drosophila* wing assays 10-25 wings each from a different animal were
219 imaged per genotype.
220
221 **HEK cell culture and variant transfection**
222 HEK 293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM)
223 supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin in a
224 humidified incubator at 37°C with 5% CO2. To transfect variants, cells were seeded in 24-well,
225 glass bottom plates (CellVis) and grown to ~80% confluency. Each MECP2E2 variant plasmid
226 was transfected using Lipofectamine 2000 (Thermo Fisher Scientific) according to the
227 manufacturer's protocol. 24 hours post-transfection, cells were fixed with 4% paraformaldehyde
228 and stained with 4',6-diamidino-2-phenylindole (DAPI) nuclear stain for 5 minutes. Prolong
229 Diamond Antifade mountant (Thermo Fisher Scientific) was added to each well and coverslips
230 were placed on top. The plates were then given 24 hours to set. Each variant was tested in three
231 separate transfections.
232
233 **HEK cell microscopy**
234 The plates were imaged using the Olympus scanR high-content screening microscope. For image
235 acquisition, the objective (20x) was focused on the centre plane of the DAPI channel, and 5 Z-
236 stack images were captured at a step width of 0.6 µm. Deconvolution was performed for each
237 image field and the maximum projection was used for quantification. Image analysis, including
238 nuclear and cytoplasmic segmentations, was carried out using the scanR software. We developed

239 a “cyto-ring assay” which uses a U-Net convolutional neural network to segment the nuclear
240 compartment based on DAPI staining, and calculate the cytoplasmic compartment based on the
241 nuclear segmentation. Two features, including protein abundance and protein translocation, were
242 extracted after segmentation from GFP-expressing cells. Protein abundance was calculated by
243 adding the total GFP intensity in the nuclear and cytoplasmic compartment. Protein translocation
244 was calculated as the ratio of GFP intensities within the nucleus divided by that in the cytoplasm.
245 Measurement for each variant was normalized to that of WT on the corresponding plate.
246
247 **HEK cell data analysis**
248 Statistical analysis was performed using GraphPad Prism 10. Multiple unpaired T-tests followed
249 by Holm-Sidak Correction were used to compare the mean values of protein abundance of each
250 variant to that of WT, as well as the protein translocation ratio of each variant to WT before
251 normalization.
252
253 **RESULTS**
254 ***Selection of variants***
255 We selected ‘reference’ benign and pathogenic variants with clear clinical classifications to
256 calibrate our functional assays, and nine test (ASD) variants (Figure 1, Supplemental Data 1).
257 We focused on ClinVar variants with “pathogenic”, “likely_pathogenic”, “benign” and
258 “likely_benign” clinical interpretation as reference variants. Priority was given to those with
259 either “reviewed by expert panel” or “criteria provided, multiple submitters” review status. A
260 few exceptions were made. D121E, a single-submitter benign variant, and T148A, originally
261 classified as benign, but now listed as conflicting were selected because they were the only

262 benign variants found in the MBD. When selecting reference pathogenic variants, in addition to
263 using review status as a criterion, we also prioritized the ones with "Rett syndrome" condition
264 annotation as well as the ones that have been frequently observed in Rett patients [36]. In total, we
265 selected 37 reference variants, 10 benign and 27 pathogenic. We selected all VariCarta MECP2
266 missense variants as ASD variants as long as they didn't have a clear clinical interpretation in
267 ClinVar, for the total of 9 test ASD variants, with the goal to determine the functional
268 consequences of these variants. Since the time of our initial selection, four test ASD variants
269 have been given updated classifications in ClinVar, P217L and E282G becoming likely
270 pathogenic, and P376S and P388S becoming benign.
271
272 ***MECP2 produces a slow growth phenotype in yeast***
273 MECP2-ref was cloned into a plasmid designed to give controlled expression in response to
274 addition of estradiol. Using an anti-MECP2 antibody we monitored MECP2 protein levels by
275 Western blot in response to increasing estradiol (Figure 2A). In the no-estradiol condition we
276 detected low levels of MECP2 protein, which increased with increasing estradiol and then
277 saturated at ~10nm. At the highest expression level, MECP2 levels were comparable to a more
278 conventional vector employing the TDH3 promoter in a multicopy 2 micron vector (Figure 2A).
279 Next we measured the growth rate of WT yeast containing either the MECP2-ref plasmid or an
280 empty vector control in response to estradiol addition (Figure 2B). Compared to empty vector
281 which had little effect in response to increasing estradiol, yeast containing MECP2 showed a
282 dose-dependent decrease in exponential growth rate. We observed a similar result with a more
283 conventional vector employing a galactose-inducible promoter (Supplemental Figure 1). Thus,
284 expression of MECP2 reduced the fitness of WT yeast.

285

286 ***MECP2 impacts DNA-related functions in yeast***

287 To understand how MECP2 might be functioning in yeast and if this is related to the normal
288 function of MECP2 we employed a genetic interaction mapping approach we previously
289 developed to study the function of human genes in yeast, called Sentinel Interaction Mapping
290 (SIM) [27]. Using SIM we express the human gene of interest in a collection of nearly 4,800 yeast
291 deletion mutants and systematically identify aggravating and alleviating genetic interactions
292 between the human and yeast genes. The known functions of the interacting yeast genes inform
293 us on the function of the human gene in yeast and if it is relevant to its known function in
294 humans. SIM also has the capacity to uncover novel functions for human genes. We performed a
295 SIM screen with MECP2-ref in which MECP2 under control of an estradiol responsive promoter
296 was expressed in ~ 4,800 different yeast deletion mutant strains in high-density arrays. We
297 identified 132 interactions in 3/3 biological replicates (Supplemental Data 2). These interactions
298 showed enrichment for functions falling under the gene ontology categories DNA Damage
299 Response, DNA Binding and Cell Cycle Process, consistent with the known function of MECP2
300 as a DNA binding protein with roles in transcriptional repression (Figure 2C). Hence, these
301 interactions suggested that the decrease in fitness observed upon MECP2 expression was a result
302 of disruption of these or related processes in yeast.

303

304 ***Functionalization of MECP2 variants in yeast***

305 We decided to simply exploit the robust slow growth phenotype observed in WT yeast to
306 measure the effects of MECP2 variants, anticipating that variants that interfered with MECP2
307 function would no longer slow down growth. We measured yeast growth in liquid culture for all

308 46 variants using a galactose-inducible vector, calculated exponential growth rates, and using a
309 linear mixed-effects model estimated the effects of the variants on MECP2 function. Raw data is
310 provided in Supplemental Data 3. Shown in Figure 3A are the model's estimates of MECP2
311 function arranged by variant amino acid position. Reference benign variants had little effect on
312 function as expected across all domains. The one exception was the D121E variant which was
313 nearly complete LoF, for which we have additional supporting evidence in fly leading to
314 reclassification of this variant as pathogenic (see below). For the MBD domain, all seven
315 reference pathogenic variants located in the MBD were complete or near-complete LoF. This
316 supported that our assay was disease-relevant for variants in the MBD. L100V, R106W, R133C,
317 and T158M have been previously reported to have reduced DNA binding *in vitro* and *in vivo* [40],
318 suggesting that our assay reported on this function of the MBD. R133 is part of the DNA
319 interface and interacts directly with guanine of the mCpG island, and T158 is part of the Asx-ST
320 motif that interacts with the DNA backbone [40]. The two test ASD variants in this domain, R91W
321 and P152L, were found to have decreased function, although not as severe as the reference
322 pathogenics. In contrast, reference pathogenics outside the MBD generally did not show
323 decreased function. Given the general consistency of the reference annotation with the assay
324 results for the MDB variants, our interpretation is that our assay was not sensitive to variants
325 outside the MBD.
326
327 ***Protein stability of variants in yeast***
328 We measured protein stability for a subset of variants with a focus on the MBD to determine if
329 stability was a factor in our assessment of function. We also included additional non-MBD
330 reference benigns from the TRD (T228S) and the CTD (S356G) for comparison, and P388S

331 because this variant showed decreased function suggesting it was destabilized. MECP-2 protein
332 levels from yeast extracts were measured on Western blots and normalized to MECP2-ref
333 (Figure 3B and Supplemental Figure 1). In the MBD, 8 out of 13 MBD variants were
334 destabilized relative to MECP2-ref, while 5 were not significantly different from Ref. Stability
335 ranged from 75% (T148A) to 37% (R106W). The two reference benigns outside the MBD,
336 T228S and S356G, were stable. Now, comparing stability to function of MBD variants there
337 appeared to be little correlation. Some variants were stable yet non-functional (L100V, P101S,
338 D121E, R133C), while the T148A variant was destabilized yet functional. Some variants showed
339 both destabilization and decreased function (R106W, T158M, G163W). This suggested that in
340 some cases the variant affected MECP2 DNA binding directly (e.g. R133C), whereas in other
341 cases decreased function was also a result of destabilization. It also suggested the functional
342 assay was fairly insensitive to destabilization.
343
344 To determine the sensitivity of our assay to destabilization we looked to variants outside the
345 MBD, because these should not directly affect DNA binding. We identified one variant out of a
346 total of 29, P388S, that showed moderate LoF (approx. 50%), suggesting it might be
347 destabilized. We measured the stability of P388S and found it was highly destabilized (approx.
348 25% stable). A functional variant in the same domain, S356G, was not destabilized. Given the
349 modest decrease in function we observed for P388S relative to its high degree of destabilization,
350 we conclude our functional assay was relatively insensitive to destabilization. This was also
351 supported by T148A, which was 67% stable, but was completely functional. Overall, these data
352 suggested that the yeast functional assay primarily reported on the intrinsic DNA binding ability
353 of the MBD.

354

355 ***Known MECP2 pathogenic and benign variants function predictably in Drosophila***

356 Expression of *MECP2* in *Drosophila* tissues has a wide range of phenotypes [28–31]. From these
357 reports, we chose to assess MECP2 variants in developing wing tissues; the reported ectopic
358 wing vein phenotype gives a quantitative readout of MECP2 protein function, and MECP2
359 genetically interacts with known co-repressor complexes relevant to MECP2 function in this
360 assay [28].

361

362 We integrated *UAS-MECP2* variants (on the *pGW.HA.attB* transgenic vector backbone) into the
363 same locus, *attP40*, to control for genome position effect and copy number[37]. After pilot
364 experiments testing a range of wing imaginal disc *GAL4* drivers and incubation temperatures, we
365 established assay conditions in which *A9-GAL4* virgin females were crossed to homozygous
366 *UAS-MECP2* males, as well as a control *attP40* line, and raised animals at 29C. We transferred
367 mating adults through three consecutive 'assay' vials. Adult progeny were then collected 3 days
368 post-eclosion, and a single wing from a minimum of 10 adult male progeny per assay vial was
369 removed and slide-mounted, for a total of 30 wings per variant. Mounted wings were scanned
370 and total wing vein area / wing area per wing was analyzed using a Labkit machine learning
371 algorithm that semi-automated the quantitation of wing vein area (see Methods).

372

373 First, we wished to validate our assay by repeating previous observations, by examining the
374 effect of MECP2-ref expression in developing wing tissue [28]. In agreement, we observed ectopic
375 wing vein tissue in wings upon expression of *MECP2-ref* (Figure 4). Further, this phenotype was

376 strongly suppressed in a heterozygotic background for *Asx[XF23],* consistent with known roles
377 for MECP2 in chromatin regulation and gene expression [20].
378
379 Next, we examined wing vein phenotypes induced by expression of all MECP2 variants and
380 controls in parallel. We tested eight reference benign and eight reference pathogenic variants.
381 Representative images for selected variants as well as *attP40* negative controls and MECP2-ref
382 are shown in Figure 5A (see Supplemental Figure 2 for representative images of MECP2 variants
383 and also the masked area that was considered to contain vein tissue). Benign variants all showed
384 increased ectopic wing vein whereas reference pathogenics showed very little, and test (ASD)
385 variants had a range of outcomes. To quantify this phenotype, we trained two machine learning
386 algorithms to quantify (1) the total area of the wing vein region and (2) the total area of the wing
387 (see Supplemental Data 3 for details). Using a linear mixed effects model these two values were
388 used to calculate the estimated effect of variants relative to MECP2-ref and were normalized to
389 the vein area/wing area for MECP2-ref (at 1) and attp40 (at 0) (Figure 5B).
390
391 All reference benign variants functioned similar to MECP2-ref, except for D121E. This variant
392 was classified in ClinVar as benign by a single contributor, but we find both yeast and fly assays
393 characterize it as a strong LoF, hence we predict that D121E is pathogenic. All reference
394 pathogenic variants showed varying degrees of a LoF or reduced function phenotype, indicating
395 the wing vein assay reported on disease relevant functions for MECP2, especially for the MBD
396 and TRD domains for which we have multiple calibrating pathogenics. Most were strong LoF
397 variants, except for P225R, which had reduced function but not a complete LoF. This successful

398 calibration of the assay indicated that clinically pathogenic and benign variants can be correctly
399 classified in the fly model.
400
401 In the same experimental batch as the reference pathogenic and benign variants described above,
402 we determined the wing vein phenotype for MECP2 test (ASD) variants identified in ASD that
403 have no clear classification. We found that both MBD variants R91W and P152L were LoF. The
404 TRD variants P217L and E282G exhibited a partial LoF with data distributed similarly to the
405 TRD pathogenic variant P225R. Finally, the Interdomain (G195S, A202G) and CTD variants
406 (P376S, P388S and S411R) all exhibited normal function that compared to the reference benign
407 variants. Thus, we conclude that only ASD variants in the MBD and TRD domains were LoF.
408
409 Finally, we compared yeast and fly assay results for MBD variants (because the yeast assay
410 calibrated for MBD variants only). We observed remarkable similarity in results between the two
411 assays (compare Figures 3A and 5B). Of particular interest was that both test (ASD) variants
412 (R91W and P152L) appeared more functional in both yeast and fly assays than reference
413 pathogenic variants. Given that T158M is DNA-binding defective and is complete LoF in our
414 assays, this suggested that ASD variants retain more DNA binding capacity than Rett
415 pathogenics.
416
417 ***MECP2 localization and stability in human cells***
418 To investigate whether MECP2 variants had altered nuclear localization we expressed GFP-
419 MECP2 in HEK cells (Raw data available in Supplemental Data 4). GFP-MECP2 localized
420 predominantly to the nucleus (Figure 7A) as expected. Using high-content imaging we assessed

421 the nuclear localization of all 46 variants using our imaging pipeline (Figure 7B). We identified
422 17 variants that showed altered nuclear localization. We were particularly interested in variants
423 that had decreased nuclear localization. 14 of the 17 showed decreased nuclear localization and
424 of these only 2 were reference benigns, whereas the remaining 12 were either reference
425 pathogenics or test (ASD). 9 of these were located in the TRD. The variant showing the greatest
426 delocalization was the pathogenic variant R255G, and consistent with this R255 is known to be
427 part of the nuclear localization signal [41]. Three test (ASD) variants showed substantial nuclear
428 delocalization, A202G, P217L and P388S, suggesting that nuclear delocalization of MECP2 is a
429 mechanism in ASD (see Figure 7A for representative images). Using this same image data we
430 also estimated protein stability of the variants (Figure 7C). 11 variants showed decreased
431 stability, of which 8 were reference pathogenics and 3 were test (ASD) (R91W, P217L, P376S).
432 Interestingly, 8 of the 11 were variants in the TRD domain, suggesting this region was important
433 for MECP2 stability. Twelve out of thirteen MBD variants were found to be stable, which
434 disagreed with the yeast stability measurements that found many MBD variants were
435 destabilized. This also disagreed with published *in vitro* and *in vivo* stability studies (see
436 Discussion), suggesting the HEK assay was less sensitive to destabilization in the MBD. Possible
437 explanations are that the N terminal GFP tag stabilized MBD variants, perhaps by altering
438 proteasome-mediated degradation [42], or that high expression driven by the CMV promoter
439 masked MBD destabilization. Because variants in the TRD are much more distal to the GFP tag
440 it may have had a lesser effect on stabilizing TRD variants.
441

442 **DISCUSSION**

443 ***Clinical classification of variants***

444 Using our assays calibrated with known reference pathogenic and benign variants, we are able to
445 provide functional assessments for nine test ASD variants, six of which have been found in cases
446 of ASD without Rett (underlined): R91W, P152L, G195S, A202G, P217L, E282G, P376S,
447 P388S, S411R (Figure 7). We found that R91W and P152L were LoF variants (yeast and fly)
448 and we classify these as likely pathogenic. S411R functioned normally (fly) and G195S had a
449 modest gain of function (fly), hence we classify these as likely benign. This provides the first
450 classifications for P152L, G195S and S411R, and re-classifies R91W from a variant of uncertain
451 significance (VoUS) in ClinVar to a likely pathogenic variant. E282G showed moderately
452 reduced function (fly) and had reduced nuclear localization (HEK), consistent with its updated
453 likely pathogenic ClinVar classification (originally had no classification). A202G was found to
454 be functional (fly), but showed decreased nuclear localization (HEK). The ClinVar classification
455 for this variant is now conflicting (one benign and one VoUS report; originally it was uncertain)
456 and we recommend this variant be further tested before a classification can be made. For the
457 additional three test (ASD) variants that are ASD and Rett, our results largely support their
458 current ClinVar classifications. We find that R133C and P152R were LoF (yeast and fly),
459 matching their likely pathogenic ClinVar classifications. P217L showed moderately reduced
460 function (fly) and had reduced nuclear localization and decreased stability consistent with its
461 updated likely pathogenic ClinVar classification. P376S was found to be functionally normal
462 (fly), but had reduced stability (HEK) suggesting its updated benign classification in ClinVar
463 should be re-examined. Finally, P388S was found to be functionally normal (fly), which
464 corresponds with its updated benign classification in ClinVar; however our HEK results
465 indicated P388S had reduced nuclear localization. Because we did not test pathogenic variants in
466 the CTD in the *Drosophila* assay we would suggest that P388S is worthy of further examination.

467

468 D121E, which is classified as benign in ClinVar, was LoF in both yeast and *Drosophila* assays.
469 Interestingly, our data was consistent with *in vitro* studies showing a ~7-fold reduction in the
470 DNA binding capacity of the D121E variant [43]. This variant is not in VariCarta or RettBASE. It
471 has one report in ClinVar, but this report lacks a description for how a benign classification was
472 made. Interestingly, the computational predictors SIFT, PolyPhen and CADD all predict this
473 variant is likely to cause LoF. We propose this variant should be re-classified in ClinVar as
474 likely pathogenic.

475

476 ***Calibrating functional assays by domain***

477 MEPC2 is a multi-domain protein, yet it is largely unstructured outside the MBD [44]. We initially
478 focused on identifying numerous reference variants in the MBD and TRD, as these domains
479 harbour many of the clinically relevant variants. The yeast assay only calibrated for the MBD
480 and failed to calibrate for all other domains indicating we can only assess functional effects of
481 variants in the MBD. The likely explanation for this is that the MECP2 MBD binds DNA in
482 yeast without the need of accessory factors. Indeed, in an accompanying paper, Brown et al.
483 show this to be likely the case. The TRD domain mediates interactions with co-repressor
484 complexes, including Sin3A, N-Cor and REST, and Rett-causing mutations disrupt these
485 interactions [45]. It is likely that the assay failed to calibrate for TRD variants because these
486 complexes either are not present in yeast (N-Cor and REST) or the TRD did not interact with the
487 yeast orthologs (Sin3A/HDAC). This may also be the case for the other domains for which
488 pathogenic reference variants failed to exhibit altered function.

489

490 *Drosophila* was successful in stratifying reference pathogenic and benign variants in both the
491 MBD and TRD domains, indicating we can reliably make predictions about test ASD variants
492 located in both these domains. For variants in the MBD, it is likely these disrupt DNA binding,
493 although this has yet to be shown. For the TRD, our data suggest MECP2 is able to interact with
494 the *Drosophila* orthologs of the co-repressor complexes. First, R306C prevents interaction with
495 co-repressors and was complete LoF in our assay [45]. Second, co-repressor complex mutants
496 suppress phenotypes resulting from MECP2 over-expression in the *Drosophila* wing [28]. For
497 example, the Polycomb group co-repressor ortholog of Asxl1, Asx, shows genetic interactions
498 with MECP2 expressed in *Drosophila* [28]. We further confirmed mutation of Asx suppressed
499 wing vein overgrowth resulting from MECP2 overexpression, supporting MECP2 interacts with
500 this co-repressor. Thus, when developing functional assays in surrogate model systems with
501 some evolutionary distance from humans, it is important to consider calibrating assays for each
502 domain before making conclusions about the effects of test variants.
503
504 ***MECP2 protein stability***
505 The stability of a handful of MECP2 variants has been studied both *in vitro* and in *in vivo* mouse
506 models [40,46–48]. We could compare seven variants located in the MBD between our yeast results
507 and these studies (Supplemental Table 1). R106W, P152R and T158M were all destabilized in
508 yeast, mouse and *in vitro* systems. R133C was found stable in yeast and *in vitro* systems, but
509 destabilized in mouse. In contrast, L100V and P101S were stable in yeast but destabilized *in*
510 *vitro*, and R133H was destabilized in yeast but stable *in vitro*. These results highlight the value of
511 comparing variant function across multiple species and model systems. For example, MECP2
512 turnover in cells is regulated by ubiquitylation by the E3 ligase RNF4 [42], indicating an *in vivo*-

513 specific modifier of variant stability. Moreover, the use of GFP-tagged MECP2 in a subset of
514 studies could conceivably affect protein turnover. Interestingly, we identified the yeast orthologs
515 of RNF4, Slx5 and Slx8, as aggravating genetic interactions in our SIM screen, suggesting
516 MECP2 levels may also be regulated similarly in yeast. Thus, intrinsic stability and regulated
517 turnover *in vivo* are important factors contributing to MECP2 variant protein levels and may vary
518 across models.
519
520 For variants outside the MBD, 14 variants showed altered stability using the HEK cell assay.
521 Comparing our results to the literature, 3/3 agreed with animal models (P225R and P322L were
522 destabilized and R306C was stable). This suggested our HEK cell assay was able to detect
523 destabilization for variants outside the MBD. Interestingly, reference pathogenics in the TRD
524 predominately showed decreased stability (7/12 variants). Thus, destabilization could be a
525 mechanism for the pathogenicity of TRD variants. MECP2 is known to be largely unstructured
526 outside the MBD, however when recruited to DNA, secondary structure significantly increases
527 (38-59% for the ID and 22-30% for the TRD) [44]. This suggests disruption of interactions with co-
528 repressors by these variants could be an additional factor leading to destabilization and Rett
529 and/or ASD phenotypes.
530
531 ***MECP2 nuclear localization***
532 Nuclear localization of MECP2 is mediated by its NLS (amino acids 255-271) which binds to the
533 nuclear import factors karyopherins 3 and 4, and by the MBD domain in a karyopherin-
534 independent yet unknown mechanism [41,49]. An engineered mutation in the NLS, R255D, disrupts
535 MECP2 nuclear localization and interaction with karyopherins [41], but there have been no

536 missense mutations identified in the NLS in Rett to date, raising the question of the importance
537 of the NLS to normal brain function. This is of potential interest because of the attractiveness of
538 karyopherin binding as a therapeutic target. Our finding that R255G is strongly nuclear
539 delocalized, together with the fact that this variant is found in an individual with intellectual
540 disability and is predicted to be pathogenic, suggests mutations in the NLS may disrupt MECP2
541 function. Importantly, we found three test (ASD) variants, A202G, P217L and P388S, that
542 showed substantially decreased nuclear localization, suggesting disrupted nuclear targeting is a
543 mechanism in ASD. Because these variants lie outside the NLS suggests nuclear targeting by the
544 MBD may be affected, although the mechanism remains undefined.
545
546 ***MECP2 and DNA methylation***
547 It is noteworthy that all reference pathogenic variants in the MBD were LoF in our functional
548 assays despite there being no, or extremely low levels of DNA cytosine methylation in yeast and
549 *Drosophila* respectively. Our data correlate well with Brown et al (accompanying article)
550 showing that MECP2 binds the unmethylated yeast genome and mediates widespread
551 transcriptional suppression, which is reduced in the pathogenic Rett variant T158M. This raises
552 the possibility that ASD and Rett MECP2 variants may not be specifically deficient in their
553 ability to bind methylated DNA, but may be generally deficient in binding to unmethylated DNA
554 and/or chromatin. We plan on expanding the panel of MECP2 Rett variants tested in our assays,
555 towards testing whether most if not all Rett pathogenic variants are deficient in non-methylated
556 DNA binding. It is challenging to specify differential impacts of specific Rett variants on methyl
557 versus non-methyl DNA binding in models where DNA methylation is present. We propose this
558 may have obscured the impact that pathogenic variants have on non-methylated DNA binding

559 and its importance in healthy and disease states, and propose that the yeast and fly models offer
560 valuable platforms to further examine these mechanisms.
561
562 **REFERENCES**

563 1. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal
564 protein that binds to Methylated DNA. *Cell* **69**, 905–914 (1992).
565 2. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding
566 methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).
567 3. Wan, M. *et al.* Rett Syndrome and Beyond: Recurrent Spontaneous and Familial MECP2
568 Mutations at CpG Hotspots. *Am. J. Hum. Genet.* **65**, 1520–1529 (1999).
569 4. Chahrour, M. & Zoghbi, H. Y. The Story of Rett Syndrome: From Clinic to Neurobiology.
570 *Neuron* **56**, 422–437 (2007).
571 5. Nomura, Y. Early behavior characteristics and sleep disturbance in Rett syndrome. *Brain*
572 *Dev.* **27**, S35–S42 (2005).
573 6. Amir, R. E. *et al.* Influence of mutation type and X chromosome inactivation on Rett
574 syndrome phenotypes. *Ann. Neurol.* **47**, 670–679 (2000).
575 7. Ng, H. H. & Bird, A. DNA methylation and chromatin modification. *Curr. Opin. Genet.*
576 *Dev.* **9**, 158–163 (1999).
577 8. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress
578 transcription. *Nat. Genet.* **19**, 187–191 (1998).
579 9. Harikrishnan, K. N. *et al.* Brahma links the SWI/SNF chromatin-remodeling complex with
580 MeCP2-dependent transcriptional silencing. *Nat. Genet.* **37**, 254–264 (2005).

581 10. Lunyak, V. V. *et al.* Corepressor-Dependent Silencing of Chromosomal Regions Encoding
582 Neuronal Genes. *Science* **298**, 1747–1752 (2002).
583 11. Ehrhart, F. *et al.* A catalogue of 863 Rett-syndrome-causing MECP2 mutations and lessons
584 learned from data integration. *Sci. Data* **8**, 10 (2021).
585 12. Yusufzai, T. M. & Wolffe, A. P. Functional consequences of Rett syndrome mutations on
586 human MeCP2. *Nucleic Acids Res.* **28**, 4172–4179 (2000).
587 13. Ballestar, E., Yusufzai, T. M. & Wolffe, A. P. Effects of Rett Syndrome Mutations of the
588 Methyl-CpG Binding Domain of the Transcriptional Repressor MeCP2 on Selectivity for
589 Association with Methylated DNA. *Biochemistry* **39**, 7100–7106 (2000).
590 14. Leblond, C. S. *et al.* Operative list of genes associated with autism and neurodevelopmental
591 disorders based on database review. *Mol. Cell. Neurosci.* **113**, 103623 (2021).
592 15. Wen, Z. *et al.* Identification of autism-related MECP2 mutations by whole-exome
593 sequencing and functional validation. *Mol. Autism* **8**, 43 (2017).
594 16. Wang, T. *et al.* De novo genic mutations among a Chinese autism spectrum disorder cohort.
595 *Nat. Commun.* **7**, 13316 (2016).
596 17. Guo, H. *et al.* Inherited and multiple de novo mutations in autism/developmental delay risk
597 genes suggest a multifactorial model. *Mol. Autism* **9**, 64 (2018).
598 18. Schaaf, C. P. *et al.* Oligogenic heterozygosity in individuals with high-functioning autism
599 spectrum disorders. *Hum. Mol. Genet.* **20**, 3366–3375 (2011).
600 19. Cheng, T.-L. *et al.* MeCP2 suppresses nuclear microRNA processing and dendritic growth
601 by regulating the DGCR8/Drosha complex. *Dev. Cell* **28**, 547–560 (2014).

602 20. Georgel, P. T. *et al.* Chromatin compaction by human MeCP2. Assembly of novel secondary
603 chromatin structures in the absence of DNA methylation. *J. Biol. Chem.* **278**, 32181–32188
604 (2003).
605 21. Hansen, J. C., Ghosh, R. P. & Woodcock, C. L. Binding of the Rett syndrome protein,
606 MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB Life* **62**, 732–738
607 (2010).
608 22. Skene, P. J. *et al.* Neuronal MeCP2 is expressed at near histone-octamer levels and globally
609 alters the chromatin state. *Mol. Cell* **37**, 457–468 (2010).
610 23. Lyko, F., Ramsahoye, B. H. & Jaenisch, R. DNA methylation in Drosophila melanogaster.
611 *Nature* **408**, 538–540 (2000).
612 24. Buitrago, D. *et al.* Impact of DNA methylation on 3D genome structure. *Nat. Commun.* **12**,
613 3243 (2021).
614 25. Ganguly, P. *et al.* A scalable Drosophila assay for clinical interpretation of human PTEN
615 variants in suppression of PI3K/AKT induced cellular proliferation. *PLoS Genet.* **17**,
616 e1009774 (2021).
617 26. Post, K. L. *et al.* Multi-model functionalization of disease-associated PTEN missense
618 mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nat.*
619 *Commun.* **11**, 2073 (2020).
620 27. Young, B. P. *et al.* Sentinel interaction mapping - a generic approach for the functional
621 analysis of human disease gene variants using yeast. *Dis. Model. Mech.* **13**, dmm044560
622 (2020).
623 28. Cukier, H. N. *et al.* Genetic Modifiers of MeCP2 Function in Drosophila. *PLoS Genet.* **4**,
624 e1000179 (2008).

625 29. Vonhoff, F., Williams, A., Ryglewski, S. & Duch, C. Drosophila as a model for MECP2 gain
626 of function in neurons. *PloS One* **7**, e31835 (2012).
627 30. Williams, A. A. *et al.* MECP2 impairs neuronal structure by regulating KIBRA. *Neurobiol.*
628 *Dis.* **91**, 284–291 (2016).
629 31. Williams, A. A. *et al.* Apoptotic Activity of MeCP2 Is Enhanced by C-Terminal Truncating
630 Mutations. *PLoS ONE* **11**, e0159632 (2016).
631 32. Brnich, S. E. *et al.* Recommendations for application of the functional evidence PS3/BS3
632 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.*
633 **12**, 3 (2019).
634 33. McKnight, D. *et al.* Recommendations by the ClinGen Rett/Angelman-like expert panel for
635 gene-specific variant interpretation methods. *Hum. Mutat.* **43**, 1097–1113 (2022).
636 34. Belmadani, M. *et al.* VariCarta: A Comprehensive Database of Harmonized Genomic
637 Variants Found in Autism Spectrum Disorder Sequencing Studies. *Autism Res. Off. J. Int.*
638 *Soc. Autism Res.* **12**, 1728–1736 (2019).
639 35. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
640 36. Good, K. V., Vincent, J. B. & Ausió, J. MeCP2: The Genetic Driver of Rett Syndrome
641 Epigenetics. *Front. Genet.* **12**, 620859 (2021).
642 37. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis
643 system for Drosophila using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. U.*
644 *S. A.* **104**, 3312–3317 (2007).
645 38. Young, B. P. & Loewen, C. J. R. Balony: a software package for analysis of data generated
646 by synthetic genetic array experiments. *BMC Bioinformatics* **14**, 354 (2013).

647 39. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear
648 Mixed Effects Models. *J. Stat. Softw.* **82**, 1–26 (2017).
649 40. Tillotson, R. & Bird, A. The Molecular Basis of MeCP2 Function in the Brain. *J. Mol. Biol.*
650 **432**, 1602–1623 (2020).
651 41. Baker, S., Lombardi, L. & Zoghbi, H. Karyopherin α 3 and Karyopherin α 4 Proteins
652 Mediate the Nuclear Import of Methyl-CpG Binding Protein 2. *J. Biol. Chem.* **290**, (2015).
653 42. Wang, Y. RING finger protein 4 (RNF4) derepresses gene expression from DNA
654 methylation. *J. Biol. Chem.* **289**, 33808–33813 (2014).
655 43. Free, A. *et al.* DNA Recognition by the Methyl-CpG Binding Domain of MeCP2 *. *J. Biol.*
656 *Chem.* **276**, 3353–3360 (2001).
657 44. Ghosh, R. P. *et al.* Unique physical properties and interactions of the domains of methylated
658 DNA binding protein 2. *Biochemistry* **49**, 4395–4410 (2010).
659 45. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the
660 NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).
661 46. Yang, Y., Kucukkal, T. G., Li, J., Alexov, E. & Cao, W. Binding Analysis of Methyl-CpG
662 Binding Domain of MeCP2 and Rett Syndrome Mutations. *ACS Chem. Biol.* **11**, 2706–2715
663 (2016).
664 47. Kucukkal, T. G., Yang, Y., Uvarov, O., Cao, W. & Alexov, E. Impact of Rett Syndrome
665 Mutations on MeCP2 MBD Stability. *Biochemistry* **54**, 6357–6368 (2015).
666 48. Sperlazza, M. J., Bilinovich, S. M., Sinanan, L. M., Javier, F. R. & Williams, D. C.
667 Structural Basis of MeCP2 Distribution on Non-CpG Methylated and Hydroxymethylated
668 DNA. *J. Mol. Biol.* **429**, 1581–1594 (2017).

669 49. Lyst, M. J. *et al.* Affinity for DNA Contributes to NLS Independent Nuclear Localization of
670 MeCP2. *Cell Rep.* **24**, 2213–2220 (2018).
671

672 **DATA SHARING STATEMENT**

673 Strains and plasmids are available upon request. The authors affirm that all data necessary for
674 confirming the conclusions of the article are present within the article, figures, and tables.
675

676 **ACKNOWLEDGEMENTS**

677 We thank Dr. Janel Kopp, UBC for access to the Panoramic Midi II digital slide scanner. Susan
678 Lindquist for the generous gift of the *pAG426GAL-ccdB* vector. Stocks obtained from the
679 Bloomington *Drosophila* Stock Center (NIH P40OD018537) were used in this study. We thank
680 *Drosophila* Genomics Resource Center, supported by NIH grant 2P40OD010949 for the esc
681 cDNA. We used FlyBase (releases 2020 through 2024) to find information on
682 phenotypes/function/stocks/gene expression (etc). We would like to thank Gino Laberge at
683 Genome ProLab for the kind gift of balancer stocks and also their transgenic services (Genome
684 ProLab Inc, Quebec, Canada). We thanks Libby Natola and Guillaume Poirier-Morency
685 (Pavlidis Lab) for contributions to the variant annotation pipeline. This work was funded by an
686 operating grant from The Simons Foundation Autism Research Initiative (2021 Genomics of
687 ASD: Pathway to Genetic Therapies: AWD-021951. SIMOFOUN 2021).

688

689

690 **AUTHOR CONTRIBUTIONS**

691 DWA, CJRL, BPY, SMG, JTC, MD, SR, MD, PP contributed to conception of the work

692 DWA, CJRL, GM, EC, BPY, SI, TL, JL, JBL, SMG, BB, MD, JS, AS, JTC SR, PP contributed

693 to design of the work

694 DWA, CJRL, GM, EC, BPY, JBL, LM, JS, JBL, AS, VK, JTC, SR, PP contributed to

695 acquisition, analysis, or interpretation of data

696 DWA, SR, PP, GM, EC, JBL, MD, BB, BW, JTC contributed to creation of new software used

697 DWA, CJRL, JS, BPY, JTC, JBL, GM, EC, LM, JL, JBL, AS, JTC, SR, PP contributed to

698 drafting the paper or substantively revised it

699

700 **FIGURE CAPTIONS**

701 **Figure 1:** Forty-six MECP2 variants by amino acid position and class. Reference pathogenic and

702 benign variants are clinically annotated variants used to calibrate assays and test (ASD) variants

703 are variants with no clinical classification found in individuals with ASD. NTD – N-terminal

704 domain, MBD – Methyl-CpG-binding domain, ID – Interdomain, TRD – Transcription repressor

705 domain, CTD – C-terminal domain.

706

707 **Figure 2:** MECP2 expression and genetic interactions in yeast. (A) Untagged MECP2-ref

708 expressed either from a plasmid utilizing the estradiol-responsive promoter or TDH3 constitutive

709 promoter (CEN low-copy or 2 micron high-copy) in WT yeast. Top band corresponds to full-

710 length MECP2. Western blot labelled with anti-MECP2 antibody. (B) Effect of MECP2-ref

711 expression on growth of WT yeast from a plasmid utilizing the estradiol-responsive promoter.

712 EV – empty vector, MECP2 – MECP2-ref. (C) Results of SIM genetic interaction screen with
713 MECP2-ref. Functional enrichment for genetic interactions is shown where nodes indicate
714 enriched GO terms and edges indicate statistically significant associations between terms. Node
715 size and edge thickness indicate statistical significance (minimum $p < 0.05$). Bolded labels
716 indicate most significant terms within networks.
717
718 **Figure 3:** Analysis of 46 MECP2 variants using the yeast assay. (A) Estimated functional effects
719 of MECP2 variants expressed in WT yeast. Data is normalized to MECP2-ref (=1) and empty
720 vector (=0). Error bars indicate 95% confidence intervals. (B) Protein stability of MECP2
721 variants. Protein levels normalized to MECP2-ref and plotted on a log scale. Error bars indicate
722 S.D. ($n = 3$). * indicates $p < 0.05$, ** indicates $p < 0.01$. Colour coding same as in (A). All
723 variants are located in MBD except T228S (TRD), S356G (CTD) and P388S (CTD).
724
725 **Figure 4:** MECP2 induces ectopic wing-vein formation which is suppressed by Asx loss-of-
726 function. (A) Crossing scheme to introduce MECP2-ref expression into WT and Asx[XF23]
727 background. (B) Representative images of wings of the indicated genotypes at the same scale.
728 Arrows indicate ectopic wing vein.
729
730 **Figure 5:** Analysis of MECP2 variants in the *Drosophila* wing assay. (A) Representative images
731 of wings expressing the indicated variants. *attp40* – no MECP2, *MECP2-ref* – WT MECP. All
732 images same scale. (B) Graph showing normalized activity of MECP2 variants where *MECP2-*
733 *ref* (WT MECP2) = 1 and *attp40* (no MECP2) = 0. Error bars indicate 95% confidence interval.
734

735 **Figure 6:** MECP2 localization and stability in HEK cells. (A) Representative images of MECP2-
736 ref (WT) and test (ASD) variants showing altered nuclear localization. (B) Translocation
737 analysis for MECP2 nuclear:cytoplasmic localization using high-throughput imaging. (C) Protein
738 abundance analysis for MECP2 using high-throughput imaging.
739
740 **Figure 7:** Summary of results for variants in MECP2 tested in yeast, *Drosophila* and HEK-cell
741 assays.
742
743 **SUPPLEMENTAL FIGURE CAPTIONS**
744 **Supplemental Figure 1**: (A) Growth rate of PGAL1-MECP2-ref compared to empty vector
745 (EV). Error bars indicate standard deviation (n = 3). (B) Representative MECP2 Western blots.
746 Arrow indicates the full-length MECP2 protein. Antibody labelling is as follows: red - anti-
747 MECP2, green - anti-PGK1.
748
749 **Supplemental Figure 2**: Representative wing phenotypes for all MECP2 variants. For each
750 variant, raw images are on the left and the masked image used to quantitate total wing vein per
751 wing is on the right.
752
753 **SUPPLEMENTAL METHODS CAPTIONS**
754 **Supplemental Methods 1:** Detailed outline of wing vein quantitation methods.
755
756 **SUPPLEMENTAL DATA CAPTIONS**

757 **Supplemental Data 1:** Detailed information for each variant, primers used in cloning, the

758 molecular genetic reagents made for each organism, and the *Drosophila* stocks generated.

759 **Supplemental Data 2:** Yeast SIM screen raw data

760 **Supplemental Data 3:** Raw data for yeast and *Drosophila* assays

761 **Supplemental Data 4:** Raw data for HEK assays

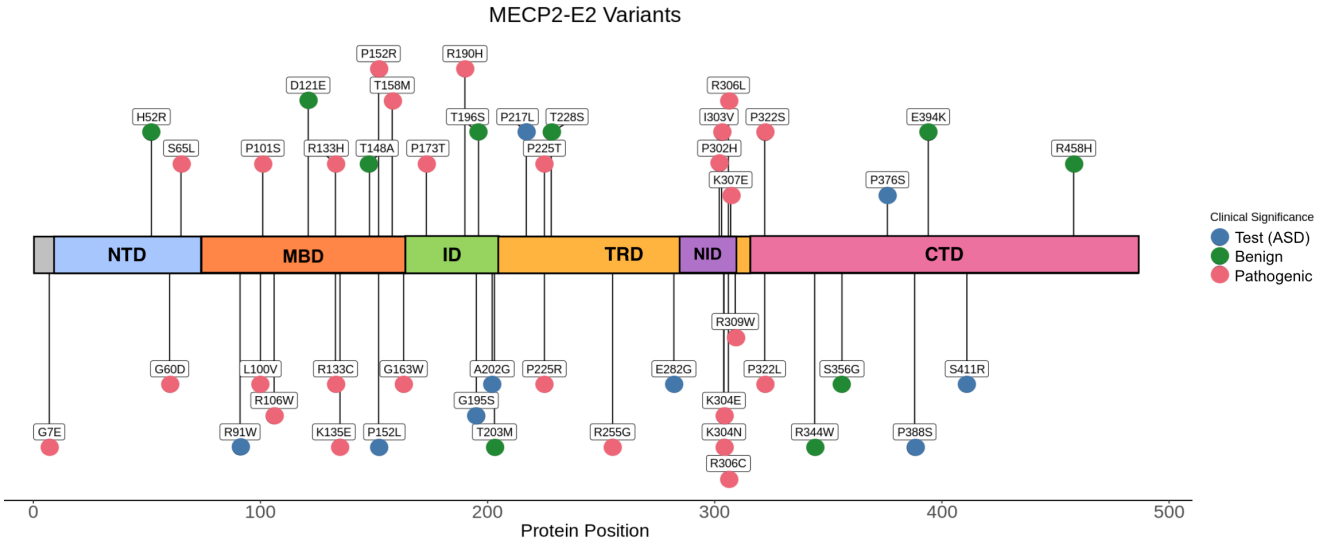
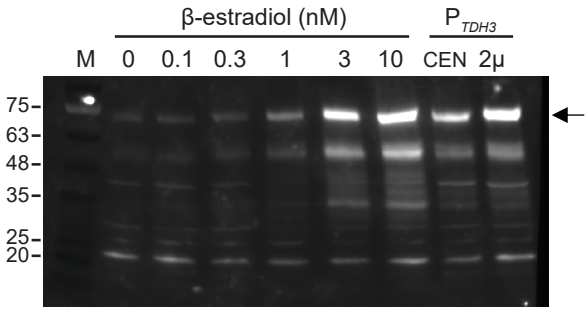
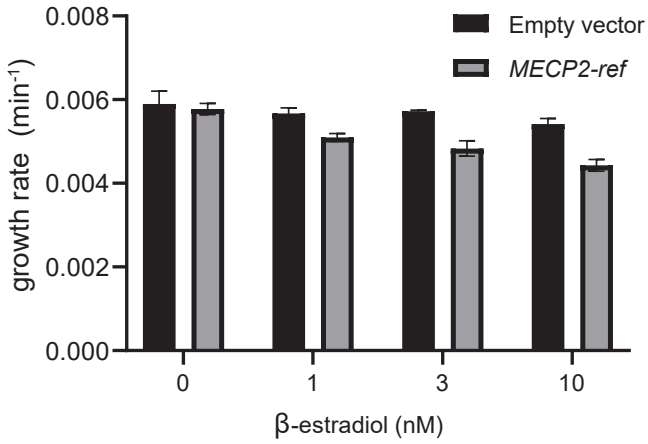

MECP2-E2 Variants
P152R
R190H
D121E
T158M
R306L
H52R
T196S
P217L
T228S
I303V
P322S
E394K
S65L
P101S
R133H
T148A
P173T
P225T
P302H
R458H
K307E
P376S
NTD
MBD
ID
TRD
NID
CTD
Clinical Significance
Test (ASD)
Benign
Pathogenic
R309W
G60D
L100V
R133C
G163W
A202G
P225R
E282G
P322L
S356G
S411R
R106W
G195S
K304E
G7E
R91W
K135E
P152L
T203M
R255G
K304N
R344W
P388S
R306C
0
100
200
300
400
500
Protein Position

Figure 1

A


β-estradiol (nM)
P_{TDH3}
M
0
0.1
0.3
1
3
10
CEN
2μ
75
63
48
35
25
20

B


Empty vector
MECP2-ref
growth rate (min-1)
0.008
0.006
0.004
0.002
0.000
0
1
3
10
β-estradiol (nM)

C

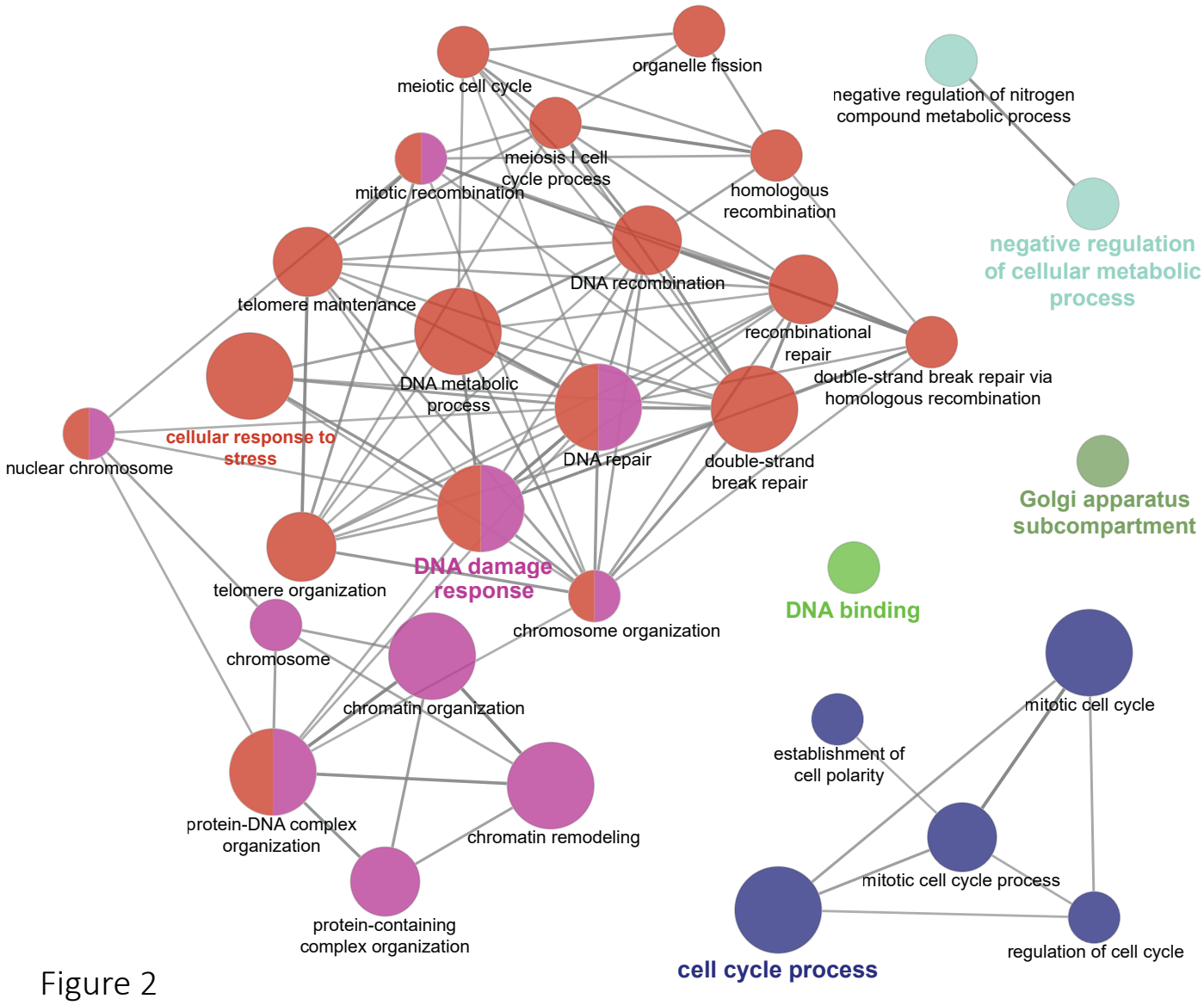
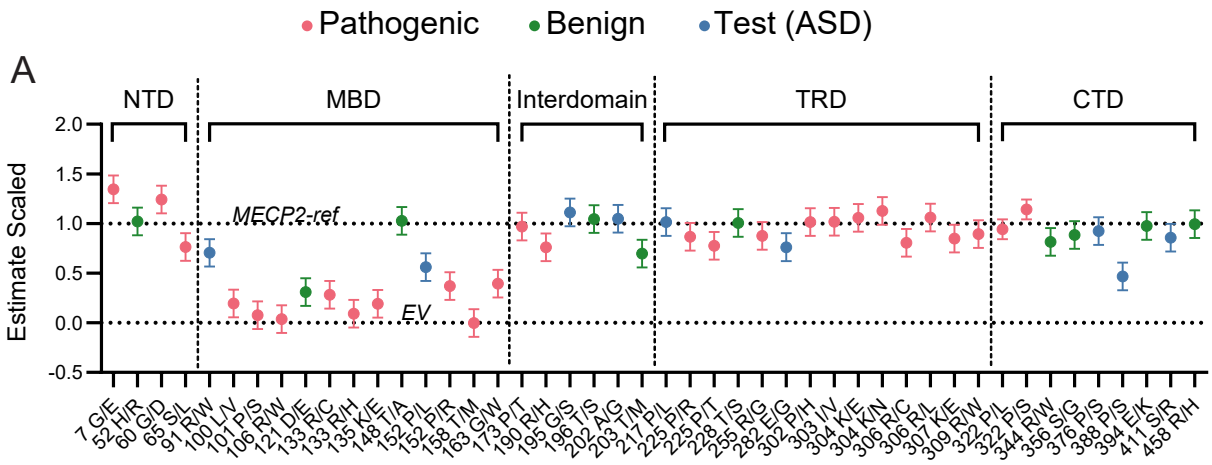

meiotic cell cycle
organelle fission
negative regulation of nitrogen compound metabolic process
meiosis I cell cycle process
mitotic recombination
homologous recombination
negative regulation of cellular metabolic process
DNA recombination
telomere maintenance
recombinational repair
double-strand break repair via homologous recombination
DNA metabolic process
cellular response to stress
nuclear chromosome
DNA repair
double-strand break repair
Golgi apparatus subcompartment
DNA damage response
telomere organization
DNA binding
chromosome organization
chromosome
chromatin organization
mitotic cell cycle
establishment of cell polarity
protein-DNA complex organization
chromatin remodeling
mitotic cell cycle process
protein-containing complex organization
regulation of cell cycle
cell cycle process

Figure 2


Pathogenic
Benign
Test (ASD)
A
NTD
MBD
Interdomain
TRD
CTD
Estimate Scaled
2.0
1.5
1.0
0.5
0.0
-0.5
MECP2-ref
EV
7 G/E
52 H/R
60 G/D
65 S/L
91 R/W
100 L/V
101 P/S
106 R/W
121 D/E
133 R/C
133 R/H
135 K/E
148 T/A
152 P/L
152 P/R
158 T/M
163 G/W
173 P/T
190 R/H
195 G/S
196 T/S
202 A/G
203 T/M
217 P/L
225 P/R
225 P/T
228 T/S
255 R/G
282 E/G
302 P/H
303 I/V
304 K/E
304 K/N
306 R/C
306 R/L
307 K/E
309 R/W
322 P/L
322 P/S
344 R/W
356 S/G
376 P/S
388 P/S
394 E/K
411 S/R
458 R/H

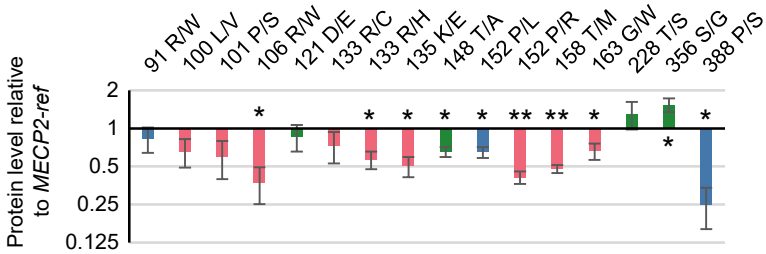

B
Protein level relative to MECP2-ref
2
1
0.5
0.25
0.125
91 R/W
100 L/V
101 P/S
106 R/W
121 D/E
133 R/C
133 R/H
135 K/E
148 T/A
152 P/L
152 P/R
158 T/M
163 G/W
228 T/S
356 S/G
388 P/S

Figure 3

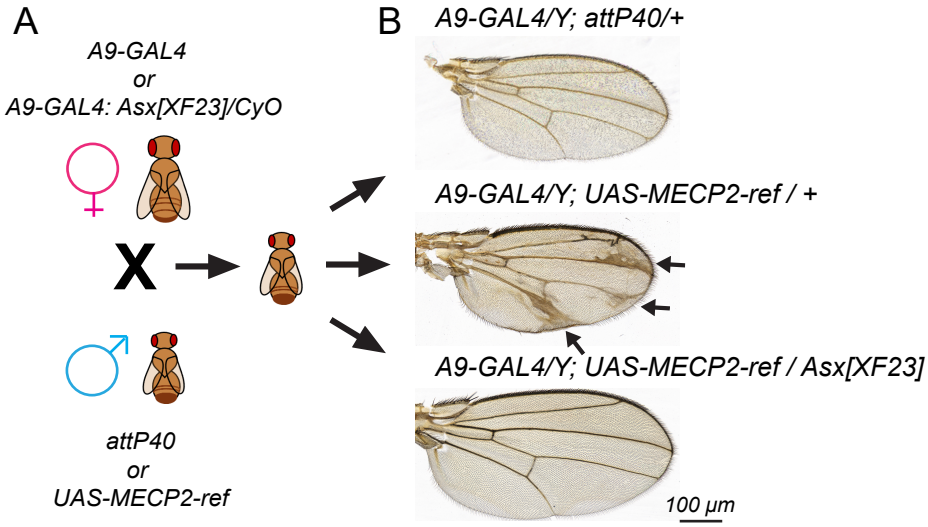
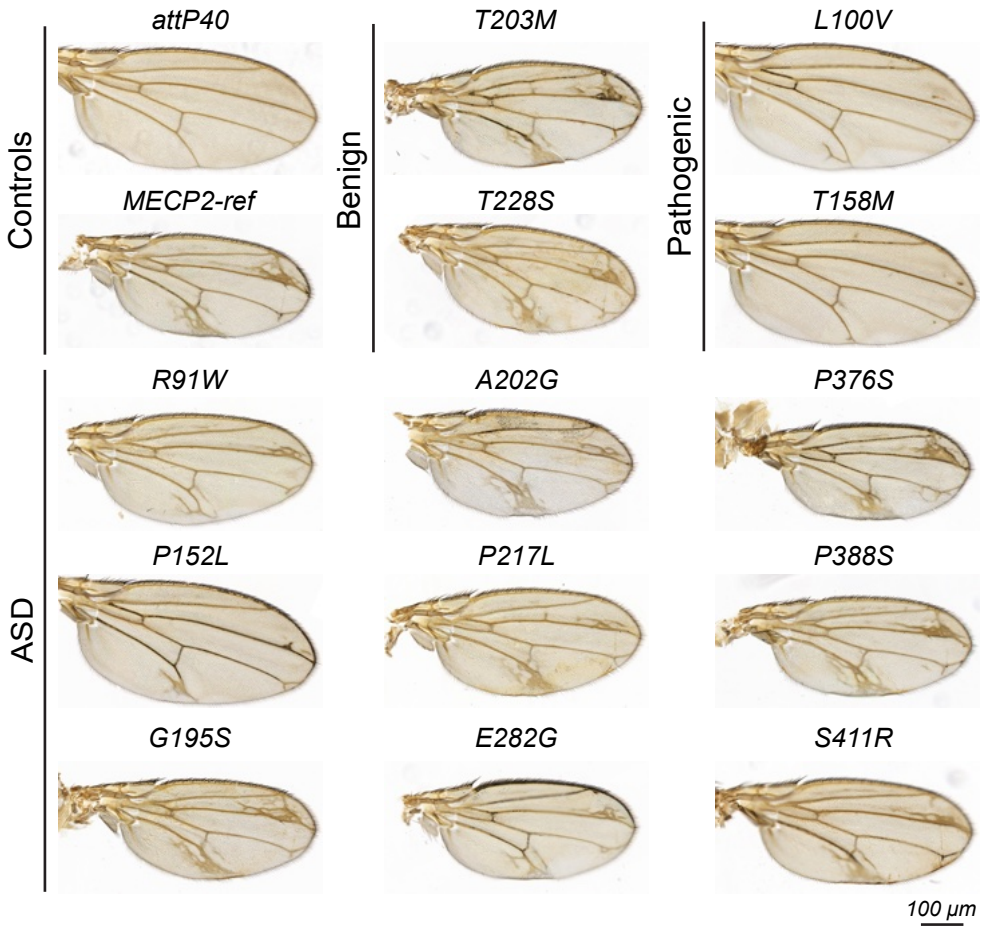

A
A9-GAL4
or
A9-GAL4: Asx[XF23]/CyO
X
attP40
or
UAS-MECP2-ref
B
A9-GAL4/Y; attP40/+
A9-GAL4/Y; UAS-MECP2-ref / +
A9-GAL4/Y; UAS-MECP2-ref / Asx[XF23]
100 µm

Figure 4

A


attP40
T203M
L100V
Controls
Benign
Pathogenic
MECP2-ref
T228S
T158M
R91W
A202G
P376S
P152L
P217L
P388S
ASD
G195S
E282G
S411R
100 µm

B

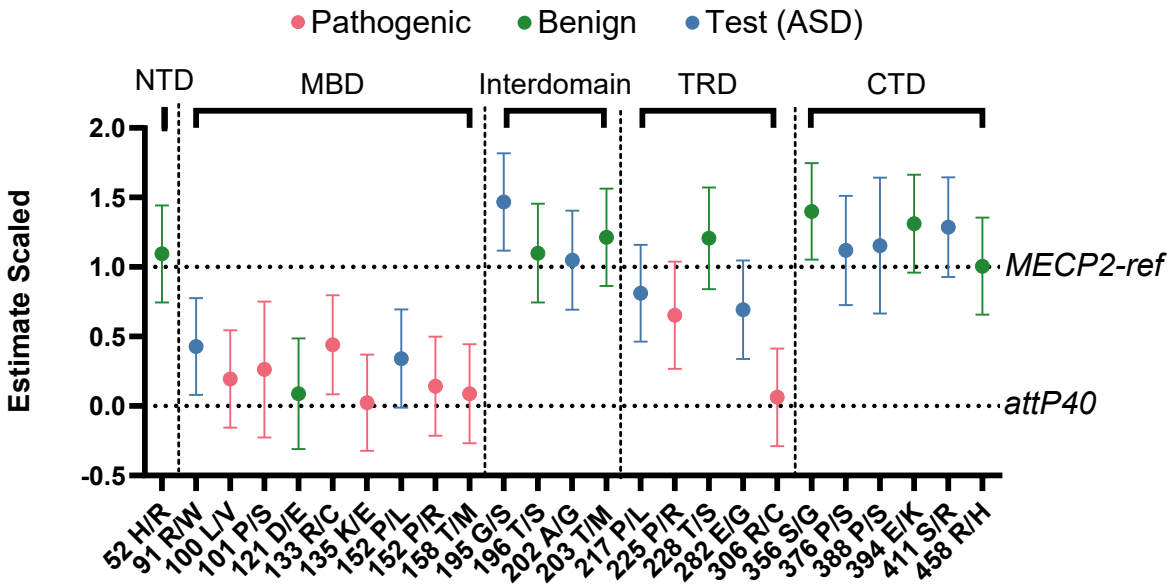
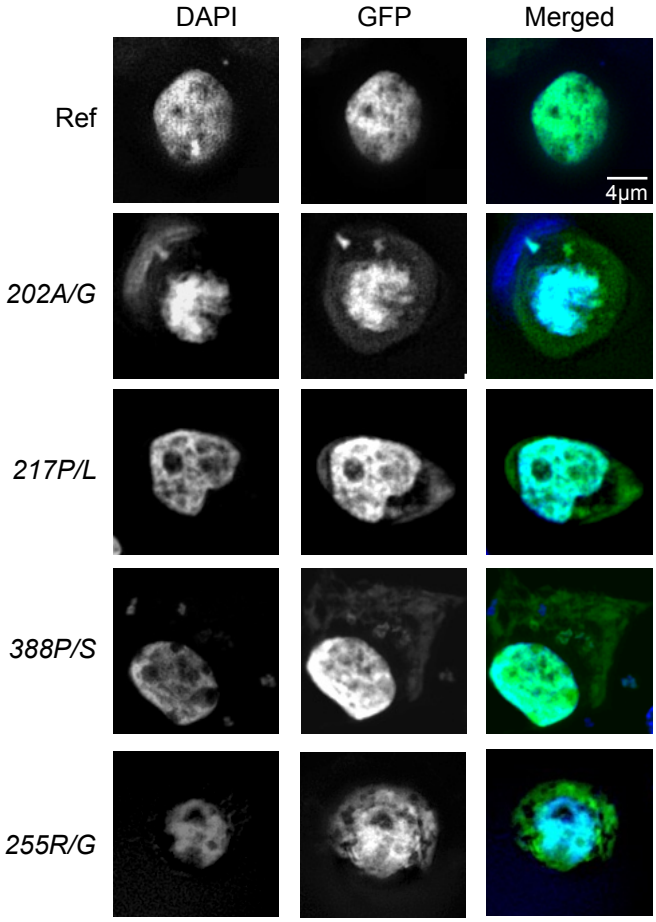
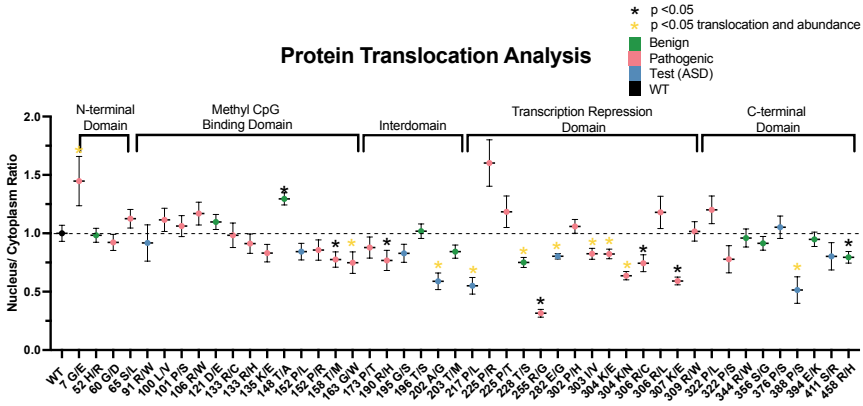

Pathogenic
Benign
Test (ASD)
NTD
MBD
Interdomain
TRD
CTD
Estimate Scaled
2.0
1.5
1.0
0.5
0.0
-0.5
MECP2-ref
attP40
52 H/R
91 R/W
100 L/V
101 P/S
121 D/E
133 R/C
135 K/E
152 P/L
152 P/R
158 T/M
195 G/S
196 T/S
202 A/G
203 T/M
217 P/L
225 P/R
228 T/S
282 E/G
306 R/C
356 S/G
376 P/S
388 P/S
394 E/K
411 S/R
458 R/H

Figure 5

A


DAPI
GFP
Merged
Ref
4µm
202A/G
217P/L
388P/S
255R/G

B


Protein Translocation Analysis
* p <0.05
* p <0.05 translocation and abundance
Benign
Pathogenic
Test (ASD)
WT
N-terminal Domain
Methyl CpG Binding Domain
Interdomain
Transcription Repression Domain
C-terminal Domain
Nucleus/ Cytoplasm Ratio
2.0
1.5
1.0
0.5
0.0
WT
7 G/E
52 H/R
60 G/D
65 S/L
91 R/W
100 L/V
101 P/S
106 R/W
121 D/E
133 R/C
133 R/H
135 K/E
148 T/A
152 P/L
152 P/R
158 T/M
163 G/W
173 P/T
190 R/H
195 G/S
196 T/S
202 A/G
203 T/M
217 P/L
225 P/R
225 P/T
228 T/S
255 R/G
282 E/G
302 P/H
303 I/V
304 K/E
304 K/N
306 R/C
306 R/L
307 K/E
309 R/W
322 P/L
322 P/S
344 R/W
356 S/G
376 P/S
388 P/S
394 E/K
411 S/R
458 R/H

C

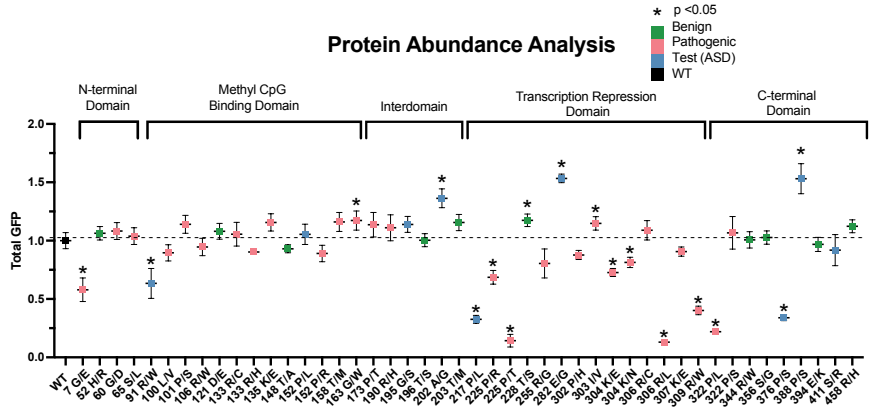

Protein Abundance Analysis
* p <0.05
Benign
Pathogenic
Test (ASD)
WT
N-terminal Domain
Methyl CpG Binding Domain
Interdomain
Transcription Repression Domain
C-terminal Domain
Total GFP
2.0
1.5
1.0
0.5
0.0
WT
7 G/E
52 H/R
60 G/D
65 S/L
91 R/W
100 L/V
101 P/S
106 R/W
121 D/E
133 R/C
133 R/H
135 K/E
148 T/A
152 P/L
152 P/R
158 T/M
163 G/W
173 P/T
190 R/H
195 G/S
196 T/S
202 A/G
203 T/M
217 P/L
225 P/R
225 P/T
228 T/S
255 R/G
282 E/G
302 P/H
303 I/V
304 K/E
304 K/N
306 R/C
306 R/L
307 K/E
309 R/W
322 P/L
322 P/S
344 R/W
356 S/G
376 P/S
388 P/S
394 E/K
411 S/R
458 R/H

Figure 6

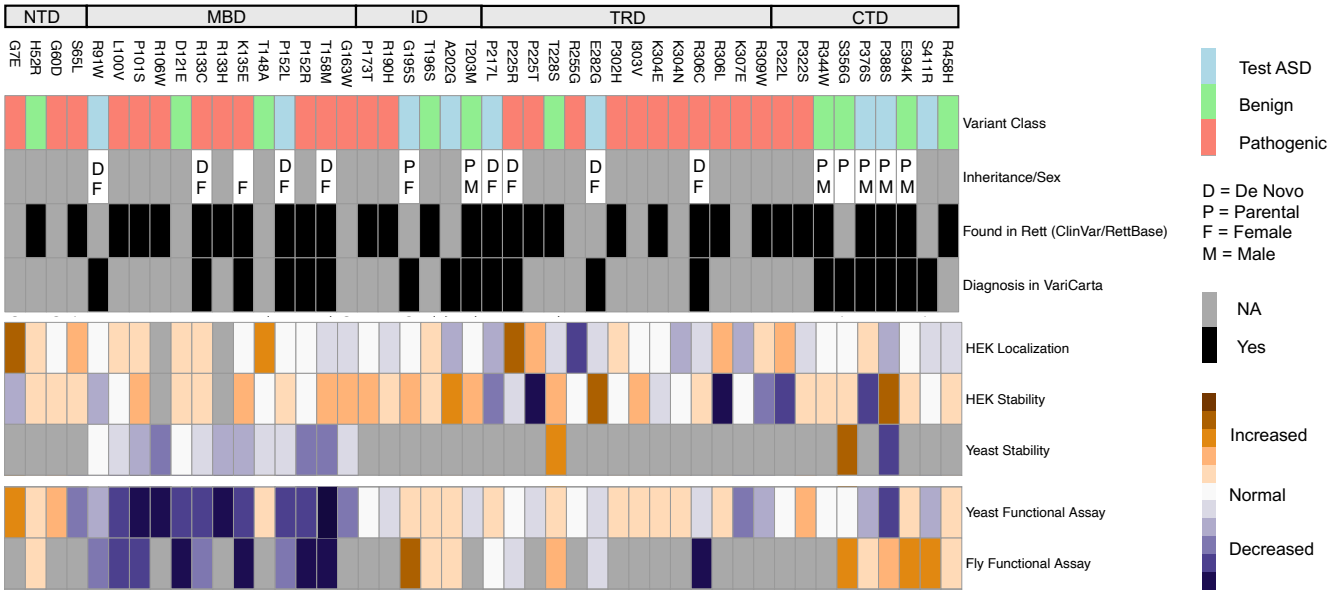
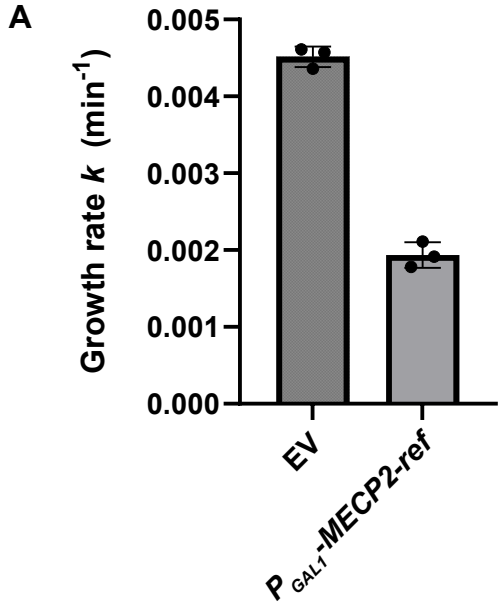

NTD
MBD
ID
TRD
CTD
G7E
H52R
G60D
S65L
R91W
L100V
P101S
R106W
D121E
R133C
R133H
K135E
T148A
P152L
P152R
T158M
G163W
P173T
R190H
G195S
T196S
A202G
T203M
P217L
P225R
P225T
T228S
R255G
E282G
P302H
I303V
K304E
K304N
R306C
R306L
K307E
R309W
P322L
P322S
R344W
S356G
P376S
P388S
E394K
S411R
R458H
Variant Class
Inheritance/Sex
Found in Rett (ClinVar/RettBase)
Diagnosis in VariCarta
HEK Localization
HEK Stability
Yeast Stability
Yeast Functional Assay
Fly Functional Assay
Test ASD
Benign
Pathogenic
D = De Novo
P = Parental
F = Female
M = Male
NA
Yes
Increased
Normal
Decreased

Figure 7


A
Growth rate k (min-1)
0.005
0.004
0.003
0.002
0.001
0.000
EV
P GAL1-MECP2-ref

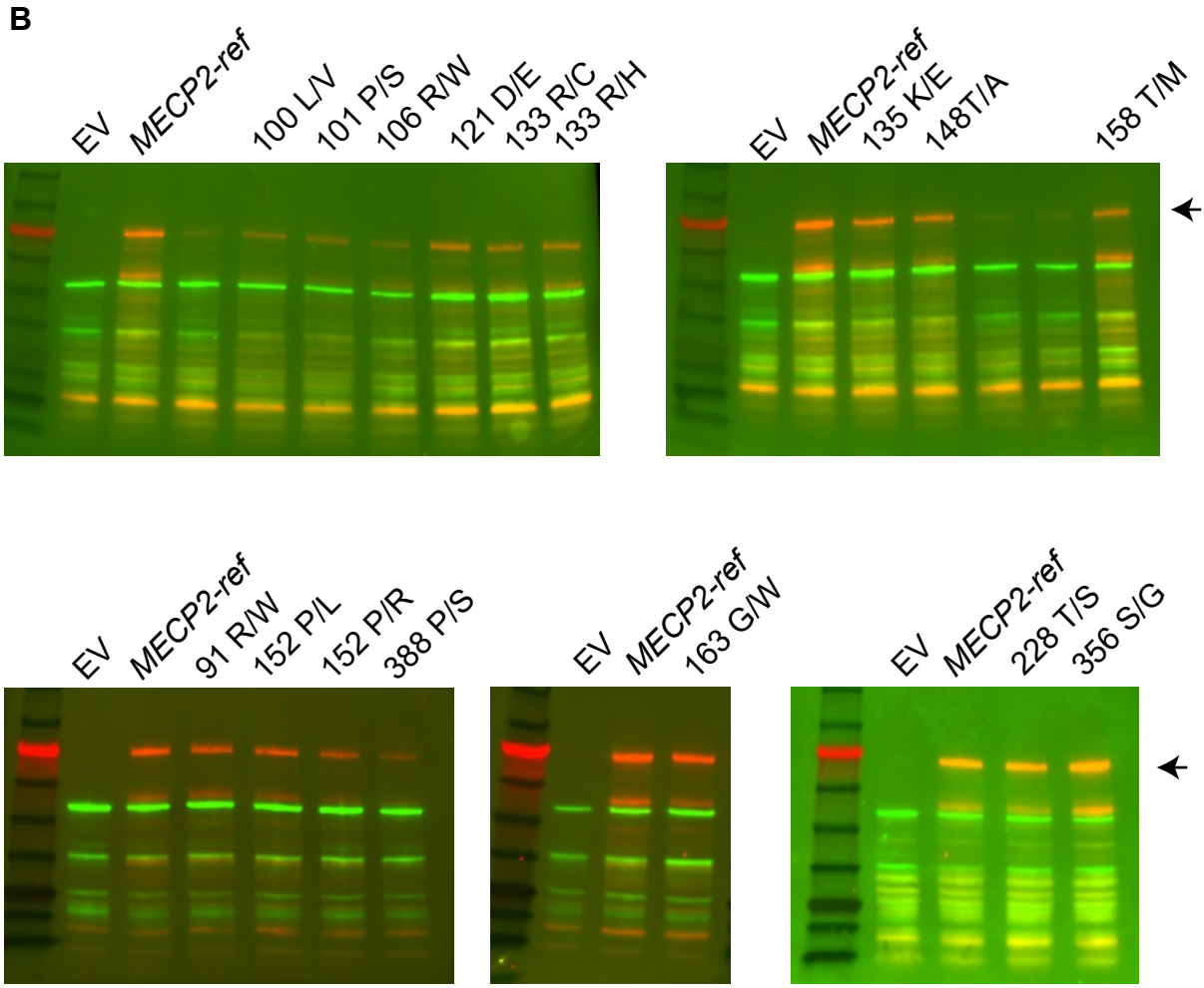

B
EV
MECP2-ref
100 L/V
101 P/S
106 R/W
121 D/E
133 R/C
133 R/H
EV
MECP2-ref
135 K/E
148T/A
158 T/M
EV
MECP2-ref
91 R/W
152 P/L
152 P/R
388 P/S
EV
MECP2-ref
163 G/W
EV
MECP2-ref
228 T/S
356 S/G

Figure S1

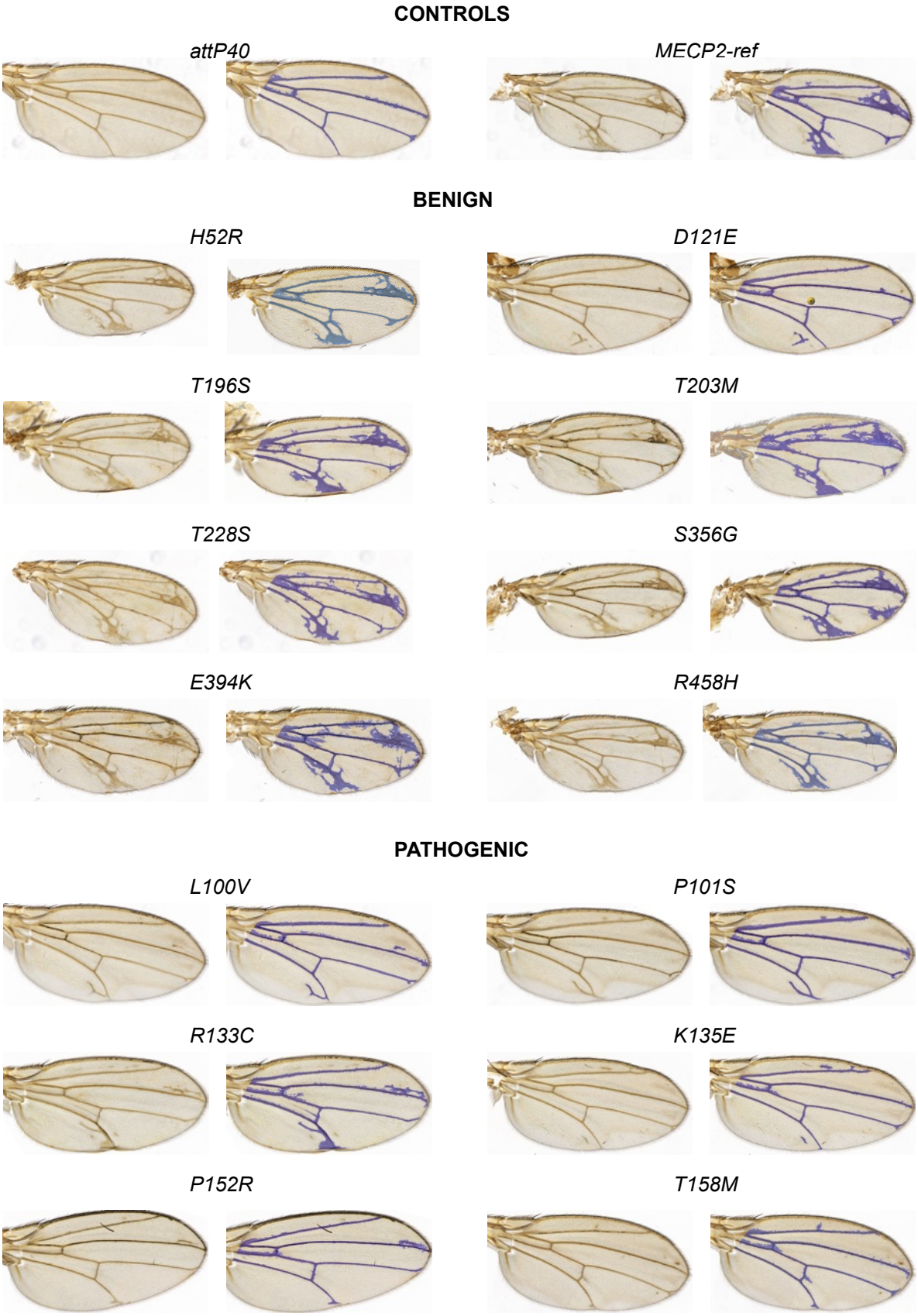

CONTROLS
attP40
MECP2-ref
BENIGN
H52R
D121E
T196S
T203M
T228S
S356G
E394K
R458H
PATHOGENIC
L100V
P101S
R133C
K135E
P152R
T158M

Figure S2-1

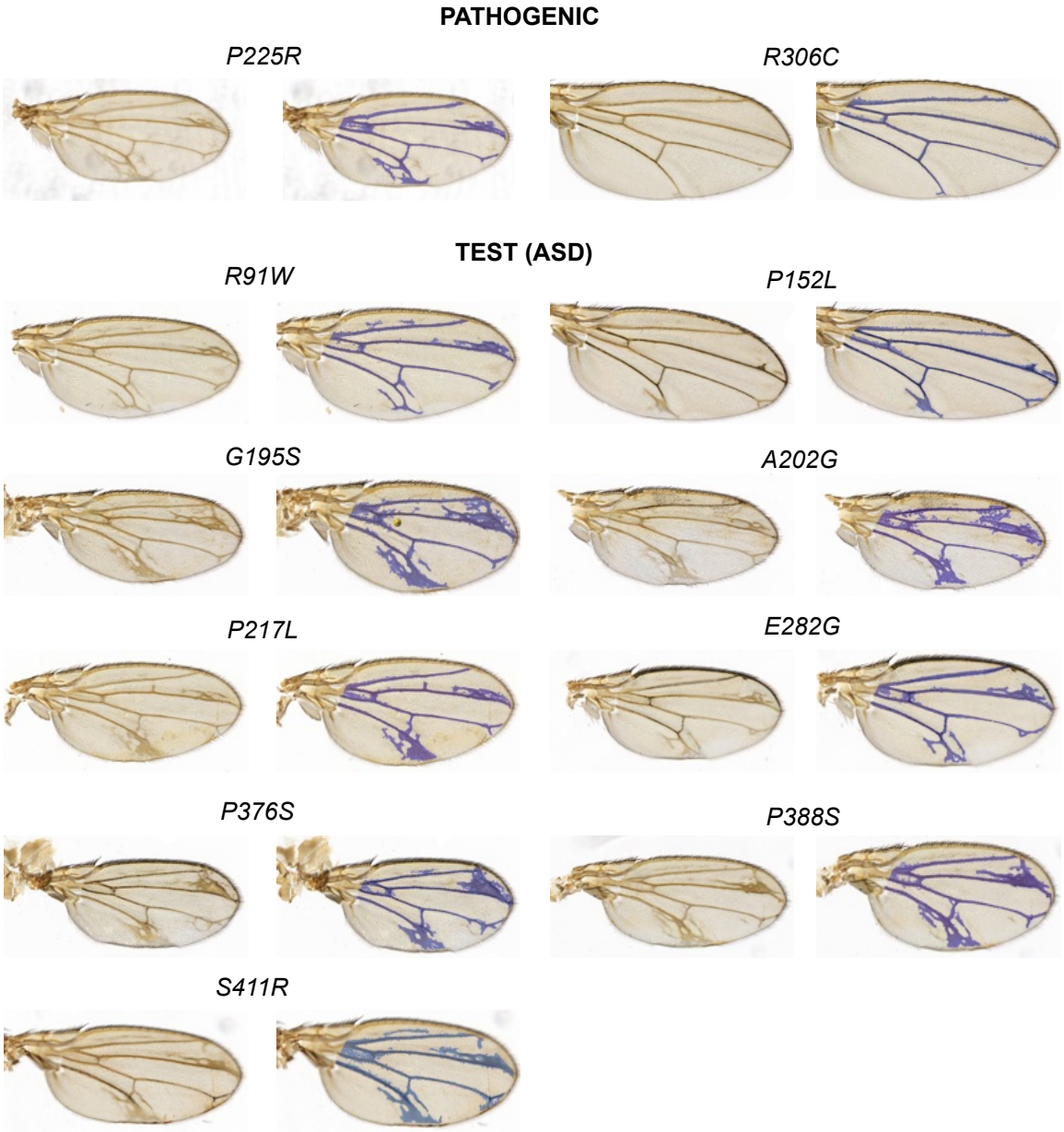

PATHOGENIC
P225R
R306C
TEST (ASD)
R91W
P152L
G195S
A202G
P217L
E282G
P376S
P388S
S411R

Figure S2-2

Supplementary Table 1: Summary of MECP2 protein stability results

Variant	Yeast*	HEK*	Yang [1] (in vitro)	Kucukkal [2] (in vitro) (ΔΔG)	Sperlazza [3] (NMR) (+ slight, ++ more destabilized)	Tillotson [4] (mouse models)
G7E		0.6				
R91W	1	0.65				
L100V	1	1	0.5	1.25		
P101S	1	1			+	
R106W	0.4	1	1	0.1	++	0.1 (TAVI)
D121E	1	1				
R133C	1	1	1	0.4		0.6 (eGFP)
R133H	0.6	1	1	0.25		
K135E	0.5	1				
T148A	0.75	1				
P152L	0.75	1				
P152R	0.4	1	0.25	1.5		
T158M	0.5	1	1	0.25	+	0.3 (eGFP), 0.25 (TAVI)
G163W	0.75	1.25				
A202G		1.4				
P217L		0.35				
P225R		0.7				0.22
P225T		0.25				
T228S	1	1.25				
E282G		1.5				
I303V		1.2				
K304E		0.75				
K304N		0.8				
R306C		1				1.0
R306L		0.2				
R309W		0.5				
P322L		0.3				0.03
S356G	1.5	1				
P376S		0.5				
P388S	0.25	1.6				

* values not significantly different than MECP2-Ref are set to 1, all others are significant

References:

1. Yang, Y., Kucukkal, T. G., Li, J., Alexov, E. & Cao, W. Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. *ACS Chem. Biol.* **11**, 2706–2715 (2016).