

Exons as units of phenotypic impact for truncating mutations in autism

Andrew H. Chiang^{1,2,3}, Jonathan Chang^{1,2,3}, Jiayao Wang^{1,2}, and Dennis Vitkup^{1,2*}

¹ Department of Biomedical Informatics, Columbia University, New York, New York, USA

² Department of Systems Biology, Center for Computational Biology and Bioinformatics, Columbia University, New York, New York, USA

³ These authors contributed equally to this work

Correspondence to DV at dv2121@columbia.edu

Abstract

Autism spectrum disorders (ASD) are a group of related neurodevelopmental diseases displaying significant genetic and phenotypic heterogeneity¹⁻⁴. Despite recent progress in understanding ASD genetics, the nature of phenotypic heterogeneity across probands remains unclear^{5, 6}. Notably, likely gene-disrupting (LGD) *de novo* mutations affecting the same gene often result in substantially different ASD phenotypes. Nevertheless, we find that truncating mutations that affect the same exon frequently lead to strikingly similar intellectual phenotypes in unrelated ASD probands. Analogous patterns are observed for two independent proband cohorts and several other important ASD-associated phenotypes. We find that exons biased towards prenatal and postnatal expression preferentially contribute to ASD cases with lower and higher IQ phenotypes, respectively. These results suggest that exons, rather than genes, often represent a unit of effective phenotypic impact for truncating mutations in autism. The observed phenotypic effects are likely mediated by nonsense-mediated decay (NMD) of splicing isoforms, with autism phenotypes usually triggered by relatively mild (15-30%) decreases in overall gene dosage. We find that each gene with recurrent ASD mutations can be described by a parameter, phenotype dosage sensitivity (PDS), which characterizes the quantitative relationship between changes in a gene's dosage and changes in a given disease phenotype. We further demonstrate analogous relationships between LGD mutations and changes in gene expression across human tissues. Therefore, similar phenotypic patterns may be also observed in multiple other systems and genetic disorders.

35 Introduction

36

37 Recent advances in neuropsychiatric genetics⁷⁻¹⁰ and, specifically, in the study of autism spectrum
38 disorders (ASD)¹¹⁻¹⁴ have led to the identification of multiple genes and specific cellular processes that are
39 affected in these diseases^{11, 12, 14-16}. However, phenotypes usually associated with ASD vary considerably
40 across autism probands¹⁻⁴, and the nature of this phenotypic heterogeneity is not well understood^{5, 6}.
41 Despite the complex genetic architecture of ASD¹⁷⁻²², a subset of cases from simplex families, i.e. families
42 with only a single affected child among siblings, are known to be strongly affected by *de novo* mutations
43 with severe deleterious effects^{14, 23, 24}. Interestingly, despite having relatively simpler genetic architecture,
44 simplex autism cohorts often display as much phenotypic heterogeneity as more general ASD cohorts²⁵⁻
45 ²⁷. This provides an opportunity for an in-depth exploration of the etiology of the autism phenotypic
46 heterogeneity, at least for these cohorts, using accumulated phenotypic and genetic data. In this study
47 we performed such an analysis, focusing on severely damaging, so-called likely gene-disrupting (LGD)
48 mutations, which include nonsense, splice site, and frameshift variants. We explored genetic and
49 phenotypic data collected in the Simons Simplex Collection (SSC)²⁸ and then validated our results using an
50 independent ASD cohort from the Simons Variation in Individuals Project (VIP)²⁹.

51 In this paper we investigated the effects of LGD mutations on cognitive and other important ASD-
52 related phenotypes, including adaptive behavior, motor skills, communication, and coordination. These
53 analyses allowed us to understand how the exon-intron structure of human genes contributes to observed
54 phenotypic heterogeneity. We then explored the quantitative relationships between changes in gene
55 dosage induced by nonsense-mediated decay (NMD) and the phenotypic effects of LGD mutations. To
56 that end, we introduced a new genetic parameter quantifying how changes in a gene's dosage affect
57 specific autism phenotypes. Finally, we described how simple linear models of gene dosage can explain a
58 substantial fraction of the phenotypic heterogeneity in the analyzed simplex ASD cohorts.

59

60 Results

61

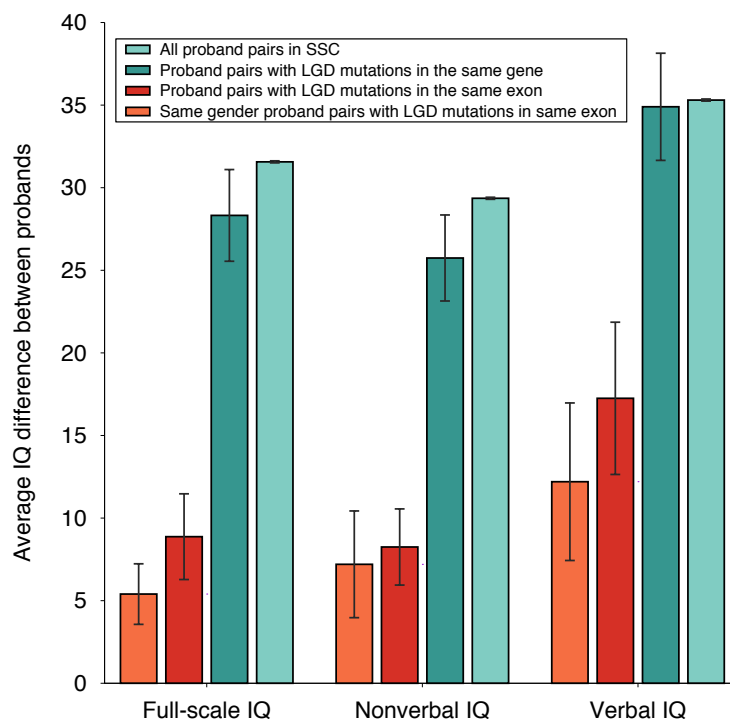
62 We first considered the impact of *de novo* LGD mutations on several well-studied cognitive
63 phenotypes: full-scale (FSIQ), nonverbal (NVIQ), and verbal (VIQ) intelligence quotients^{11, 14, 16}; these
64 scores are normalized by age and standardized across a broad range of phenotypes²⁸. We analyzed *de*
65 *nov*o mutations and the corresponding phenotypes of ASD probands for more than 2,500 families from
66 SSC²⁸. Notably, we found that the average IQ differences between probands with LGD mutations in the
67 same gene were only slightly smaller than the IQ differences between all pairs of probands; the mean
68 pairwise IQ difference for probands with mutations in the same gene was 25.7 NVIQ points, while the
69 mean difference for all pairs of probands was 29.4 NVIQ points (~12% difference, Mann-Whitney U one-
70 tail test $P = 0.14$; Supplementary Table 1).

71 We next asked whether probands with LGD mutations at similar locations within the same gene
72 resulted, on average, in more similar phenotypes (Supplementary Fig. 1). Indeed, IQ differences between
73 probands with LGD mutations closer than 1000 base pairs apart were significantly smaller than the IQ
74 differences between probands with more distant mutations; NVIQ average difference of 10.4 points for \leq
75 1000 bp, NVIQ average difference of 28.6 points for > 1000 bp (MWU one-tail test $P = 0.005$). However,
76 across the entire range of nucleotide distances between LGD mutations in the same genes, we did not
77 observe either a significant correlation or a monotonic relationship between IQ differences and mutation
78 proximity (NVIQ Spearman's $\rho = 0.1$, $P = 0.4$; Mann-Kendall one-tail trend test $P = 0.5$).

79 To explain the observed patterns of phenotypic similarity, we next considered the exon-intron
80 structure of target genes. Specifically, we investigated phenotypes resulting from truncating mutations
81 affecting the same exon in unrelated ASD probands; in this analysis, we took into account LGD mutations
82 in the exon's coding sequence as well as disruptions of the exon's flanking canonical splice sites, since

83 such splice site mutations should affect the same transcript isoforms (Supplementary Fig. 2). Interestingly,
84 the analysis of 16 unrelated ASD probands (8 pairs with LGD mutations in the same exons) showed that
85 they have strikingly more similar phenotypes (Fig. 1, red bars) compared to probands with LGD mutations
86 in the same gene (Fig. 1, dark green bars); same exon FSIQ/NVIQ/VIQ average IQ difference 8.9, 8.3, 17.3
87 points, same gene average difference 28.3, 25.7, 34.9 points (Mann-Whitney U one-tail test $P = 0.003$,
88 0.005 , 0.016). Because of well-known gender differences in autism susceptibility^{11, 30, 31}, we also compared
89 IQ differences between probands of the same gender harboring truncating mutations in the same exon
90 (Fig. 1, orange bars) to IQ differences between probands of different genders; same gender
91 FSIQ/NVIQ/VIQ average difference 5.4, 7.2, 12.2; different gender average difference 14.7, 10, 25.7 (MWU
92 one-tail test $P = 0.04$, 0.29 , 0.07). Thus, stratification by gender further decreases the phenotypic
93 differences between probands with LGD mutations in the same exon. Notably, the patterns of phenotypic
94 similarity between different probands only extended to mutations affecting the same exon. The average
95 IQ differences between probands with LGD mutations in neighboring exons were not significantly
96 different compared to mutations in non-neighboring exons (MWU one-tail test $P = 0.6$, 0.18 , 0.8 ;
97 Supplementary Fig. 3). The observed effects were also specific to LGD mutations; probands with either
98 synonymous ($P = 0.93$, 0.97 , 0.95 ; Supplementary Fig. 4) or missense ($P = 0.8$, 0.5 , 0.8 ; Supplementary Fig.
99 5) mutations in the same exon were as phenotypically diverse as random pairs of ASD probands.

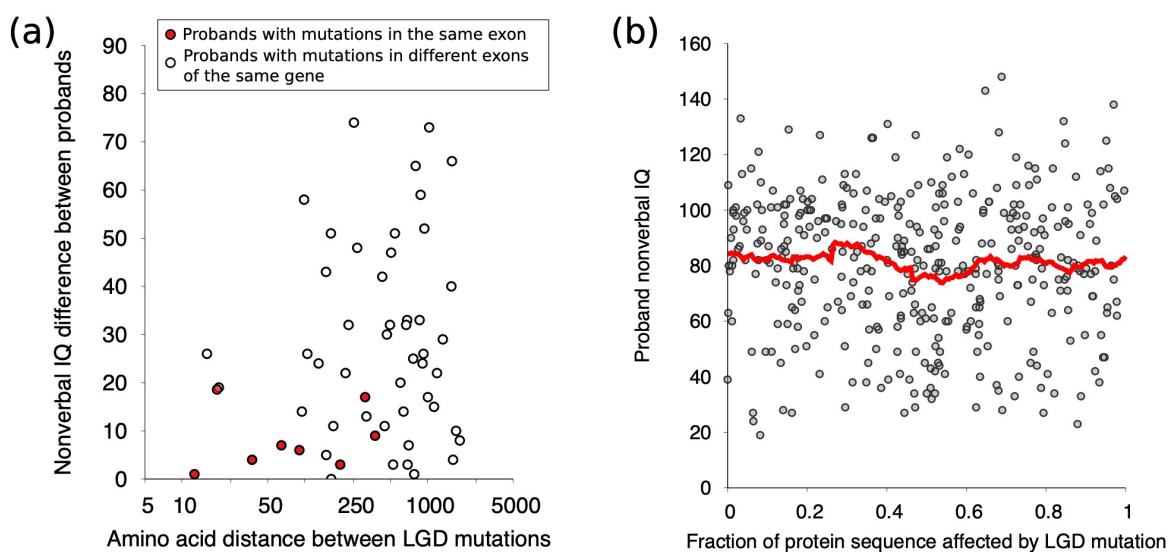
100



101
102 **Figure 1:** Average difference in IQs between SSC probands. From left to right, the sets of bars represent
103 differences between full-scale, nonverbal, verbal IQs. Within each bar set, from right to left, the bars
104 represent the average IQ difference between pairs of probands in the entire SSC cohort (light green),
105 between probands with *de novo* LGD mutations in the same gene (dark green), between probands with
106 *de novo* LGD mutations in the same exon (red), and between probands of the same gender and with *de*
107 *nov* LGD mutations in the same exon (orange). Error bars represent the SEM.

108
109 We next explored the relationship between phenotypic similarity and the proximity of truncating
110 mutations in the corresponding protein sequences. This analysis revealed that probands with LGD

111 mutations in the same exon often had similar IQs, despite being affected by truncating mutations
112 separated by scores to hundreds of amino acids in protein sequence (Fig. 2a; Supplementary Fig. 6).
113 Furthermore, we found probands with LGD mutations in the same exon to be more phenotypically similar
114 than probands with LGD mutations separated by comparable amino acid distances in the same protein
115 sequence but not necessarily in the same exon (NVIQ distance-matched permutation test $P = 0.002$;
116 Supplementary Fig. 7). We also investigated whether *de novo* mutations truncating a larger fraction of
117 protein sequences resulted, on average, in more severe intellectual phenotypes. Surprisingly, this analysis
118 showed no significant correlations between the fraction of truncated protein and the severity of
119 intellectual phenotypes (Fig. 2b); NVIQ Pearson's $R = 0.05$ ($P = 0.35$; Supplementary Fig. 8). We also did
120 not find any significant biases in the distribution of truncating *de novo* mutations across protein sequences
121 compared with the distribution of synonymous *de novo* mutations (Kolmogorov-Smirnov two-tail test $P =$
122 0.9 ; Supplementary Fig. 9). It is possible that the lack of the correlation between phenotypic impact and
123 the fraction of truncated sequence is due to the averaging of various effects across different proteins.
124 Therefore, for genes with recurrent mutations, we used a paired test to investigate whether truncating a
125 larger fraction of the same protein sequence led to more severe phenotypes. This analysis also showed
126 no substantial phenotypic difference due to LGD mutations truncating different fractions of the same
127 protein (average NVIQ difference 0.24 points; Wilcoxon signed-ranked one-tail test $P = 0.44$). We also
128 investigated, using the Pfam database³², whether mutations that truncate the same protein domain lead
129 to more similar phenotypic differences. The results demonstrated that mutations in different exons, even
130 when truncating the same protein domain, resulted, on average, in phenotypes as different as due to LGD
131 mutations in the same protein (average NVIQ differences = 28.1; Supplementary Fig. 10).
132



133
134 **Figure 2:** The relationship between the position of *de novo* LGD mutations in protein sequence and
135 probands' IQs. (a) Amino acid distance between LGD mutations in protein sequence versus differences in
136 nonverbal IQ. Each point corresponds to a pair of probands with LGD mutations in the same gene. The x-
137 axis represents the amino acid distance between LGD mutations, and the y-axis represents the difference
138 between the corresponding probands' nonverbal IQs (NVIQ). Red points represent pairs of probands with
139 LGD mutations in the same exon, and white points represent pairs of probands with mutation in the same
140 gene but different exons. (b) Relative fraction of protein sequence truncated by LGD mutations versus
141 probands' NVIQs. Each point corresponds to a single individual affected by an LGD mutation. The x-axis

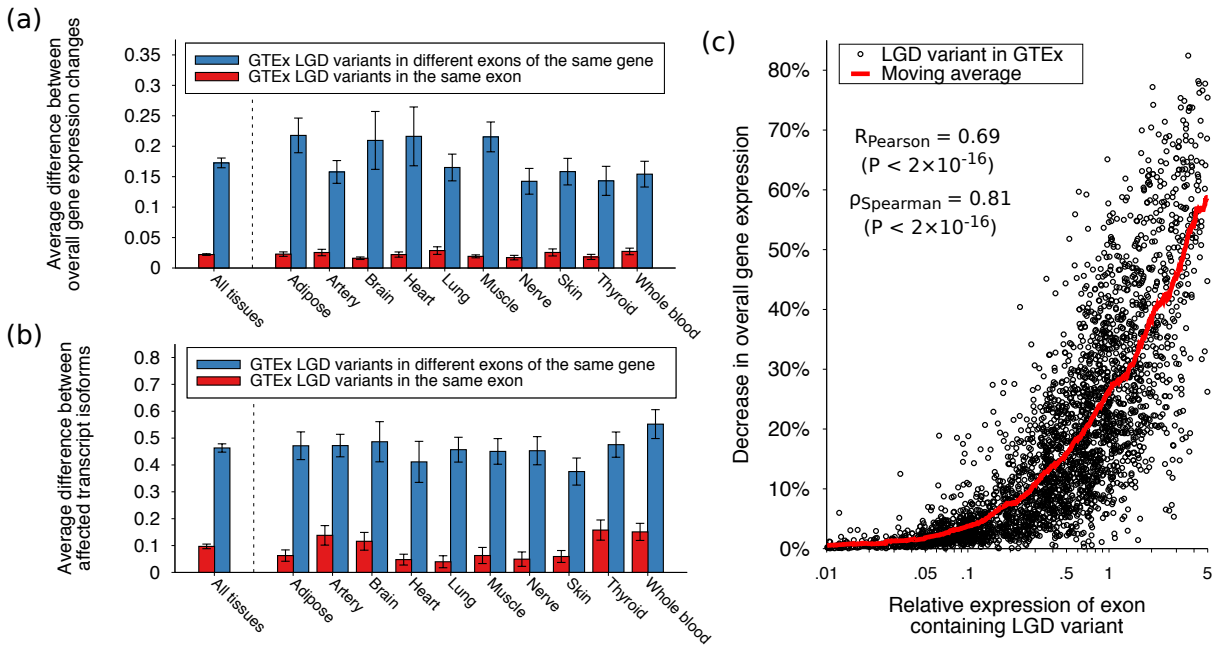
142 represents the fraction of protein sequence (i.e. fraction from the first amino acid) truncated by the
143 mutation, and the y-axis represents the corresponding NVIQ. The red line represents a moving average of
144 the data, calculated on an interval of width 0.05.

145
146 The results presented above suggest that the occurrence of *de novo* LGD mutations in the same
147 exon, rather than simply the proximity of mutation sites in protein or nucleotide sequences, is primarily
148 responsible for similar phenotypic consequences in unrelated probands. To explain this observation, we
149 hypothesized that truncating mutations in the same exon usually affect, due to nonsense-mediated decay
150 (NMD)³³, the expression of exactly the same sets of splicing isoforms. Therefore, such mutations should
151 lead to particularly similar phenotypes, both through similar decreases in overall gene dosage and similar
152 perturbations to the mRNA expression of affected transcriptional isoforms. To evaluate this mechanistic
153 model, we used data from the Genotype and Tissue Expression (GTEx) Consortium^{34, 35}, which collected
154 exome sequencing and corresponding human tissue-specific gene expression data from hundreds of
155 individuals and across multiple tissues. Using ~4,400 LGD variants in coding regions and corresponding
156 RNA-seq data, we compared the expression changes resulting from LGD variants in the same and different
157 exons of the same gene (Fig. 3a,b). Specifically, for each truncating variant, we analyzed allele-specific
158 read counts³⁶ and then used an empirical Bayes approach to infer the effects of NMD on gene expression
159 (see Methods). This analysis confirmed that the average gene dosage changes for individuals with LGD
160 variants in the same exon were ~7 times more similar compared to individuals with LGD variants in
161 different exons of the same gene (Fig 3a); 2.2% versus 17.3% average difference in the decrease of overall
162 gene dosage (Mann-Whitney U one-tail test $P < 2 \times 10^{-16}$). Moreover, by analyzing GTEx data for each tissue
163 separately, we found that across tissues LGD variants in the same exons lead to drastically more similar
164 dosage changes of target genes (Fig. 3a).

165 Distinct splicing isoforms often have different functional properties^{37, 38}. Consequently, LGD
166 variants may affect phenotypes not only through NMD-induced changes in overall gene dosage,
167 considered above, but also by altering the expression levels of different sets of splicing isoforms. To
168 specifically analyze changes in the relative expression of distinct gene isoforms, we next used GTEx
169 variants to quantify the effects of NMD on each isoform of a gene. To compare isoform-specific expression
170 changes in the same gene, we calculated an angular distance metric between vectors representing dosage
171 changes for each isoform (see Methods). This analysis demonstrated that changes in relative isoform
172 expression are also significantly (~5 fold) more similar for LGD variants in the same exon compared to
173 variants in different exons of the same gene (Fig. 3b); 0.1 versus 0.46 for the average angular distance
174 between isoform expression vectors (Mann-Whitney U one-tail test $P < 2 \times 10^{-16}$). These results were also
175 consistent across tissues (Fig. 3b). Overall, the analyses of GTEx data demonstrate that both overall
176 changes in gene dosage and changes in the relative expression levels of different isoforms are
177 substantially more similar for truncating mutations in same exons.

178 Truncating variants in highly expressed exons should lead, on average, to relatively larger NMD-
179 induced decreases in overall gene dosage. To confirm this hypothesis, we used RNA-seq data from GTEx.
180 Specifically, for each exon harboring a truncating variant, we calculated its expression level relative to the
181 expression values of the corresponding gene. We then explored the relationship between the relative
182 expression of exons with the observed NMD-induced decreases in gene expression. The analysis indeed
183 revealed a strong correlation between the relative expression levels of exons harboring LGD variants and
184 the corresponding changes in overall gene dosage (Fig. 3c; Pearson's $R = 0.69$, $P < 2 \times 10^{-16}$; Spearman's ρ
185 = 0.81, $P < 2 \times 10^{-16}$; see Methods). NMD-induced dosage changes may mediate the relationship between
186 the relative expression levels of target exons and the corresponding phenotypic effects of truncating
187 mutations. To investigate this relationship in detail, we used the BrainSpan dataset³⁹, which contains
188 exon-specific expression from human brain tissues. The BrainSpan data allowed us to estimate expression
189 dosage changes resulting from LGD mutations in different exons of ASD-associated genes (see Methods).

190



191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

Figure 3: Gene expression changes across human tissues due to LGD variants in the same exon and in the same gene but different exons. Expression changes (decreases) due to LGD variants were calculated based on data from the Genotype and Tissue Expression (GTEx) Consortium³⁴. (a) Bars represent the average difference across the GTEx cohort in overall gene expression changes induced by distinct LGD variants in the same exon (red) and in the same gene but different exons (blue). Error bars represent the SEM. (b) Bars represent the average difference across the GTEx cohort in isoform-specific expression changes induced by distinct LGD variants in the same exon (red) and in the same gene but different exons (blue). Differences in expression changes across transcriptional isoforms were quantified using the angular distance metric between vectors representing isoform-specific expression changes (see Methods). Error bars represent the SEM. (c) Relationship between the relative expression of exons containing LGD variants and the corresponding NMD-induced decreases in overall gene expression. Each point corresponds to an LGD variant in one of ten human tissues. The x-axis represents the relative expression of an exon harboring an LGD variant in a tissue; the relative expression of an exon was calculated as the ratio between the exon expression and total expression of the corresponding gene (see Methods). The y-axis represents the NMD-induced decrease in overall gene expression (see Methods). Red line represents a moving average of the data, calculated on an interval of width 0.1 (log-scaled).

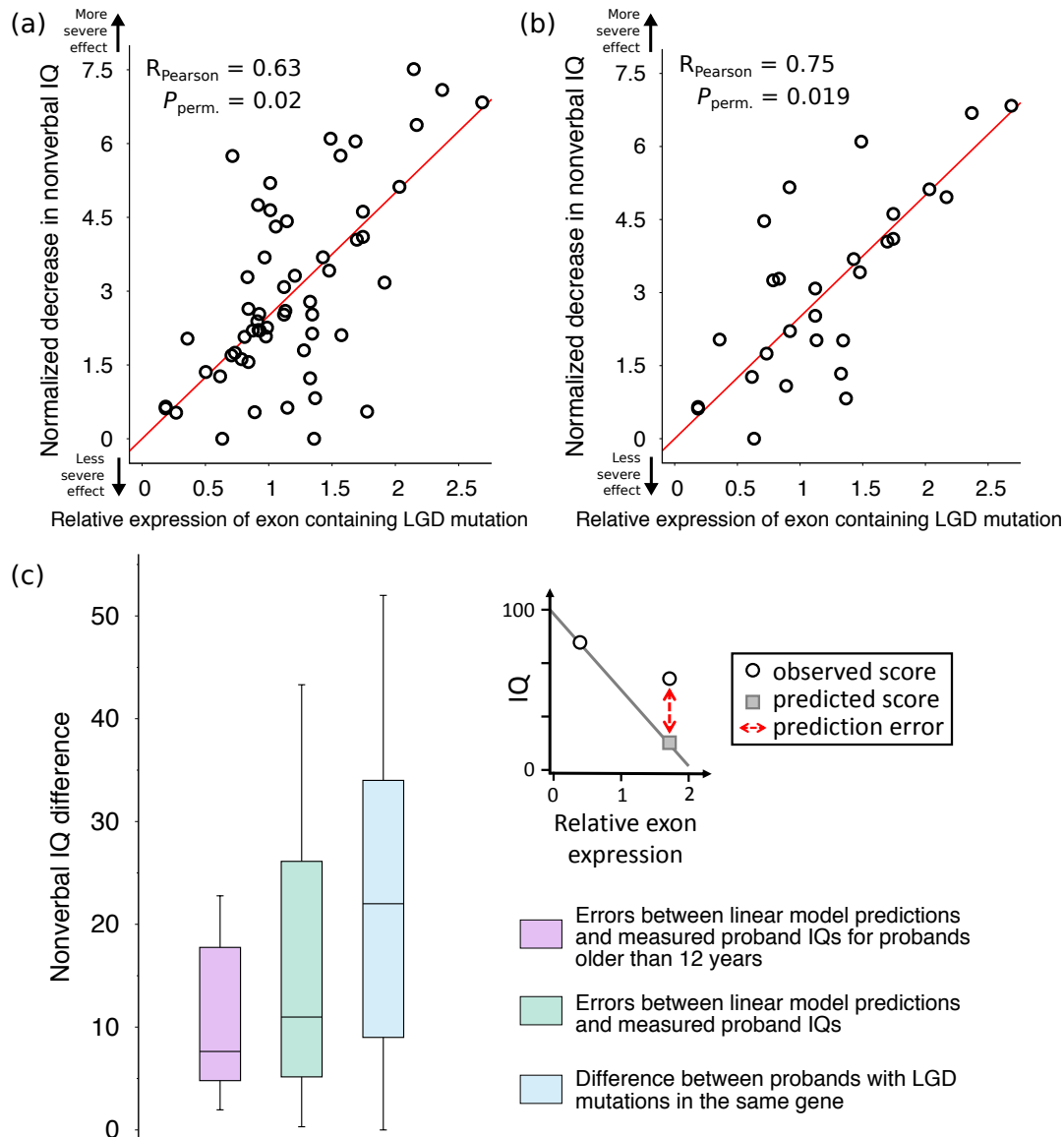
Notably, it is likely that there is substantial variability across human genes in terms of the sensitivity of intellectual and other ASD phenotypes to gene dosage. Therefore, to quantify the sensitivity of IQ to changes in the expression of specific genes, we considered a simple linear dosage model. Specifically, we assumed for genes with recurrent truncating mutations in SSC that changes (decreases) in probands' IQs are linearly proportional to the predicted relative decrease in overall gene dosage due to NMD. We further assumed that each human gene can be characterized by a parameter, which we call its phenotypic dosage sensitivity (PDS), characterizing the linear relationship between changes in gene dosage compared to wild type and the corresponding changes in a given human phenotype. Numerically, we defined IQ-associated PDS to be equal to the average change in IQ resulting from a 10% change in gene dosage. We restricted this analysis to LGD mutations predicted to cause NMD-induced expression

220 changes, i.e. we excluded mutations within 50 bp of the last exon junction complex⁴⁰, and also assumed
221 the average neurotypical IQ (100) for wild type (intact) gene dosage. Using this linear model, for each gene
222 with recurrent truncating ASD mutations, we used predicted changes in gene dosage to estimate gene-
223 specific PDS parameters for intellectual phenotypes (Supplementary Fig. 11; see Methods). Notably, as
224 we expected, PDS values varied substantially across 24 considered human genes ($CV = SD/Mean = 0.57$,
225 for NVIQ).

226 We used the aforementioned linear model to explore the relationship between the relative
227 expression values of exons (i.e. the ratio of exon expression to gene expression) harboring LGD mutations
228 and the corresponding decreases in probands' intellectual phenotypes. To account for differences in
229 phenotypic sensitivity to dosage changes across genes, we normalized the observed changes in IQ by the
230 estimated PDS values of affected genes. Normalized in this way, phenotypic effects represent changes in
231 phenotype relative to the predicted effects for 10% decreases in dosage of affected genes. This analysis
232 revealed that mutation-induced gene dosage changes are indeed strongly correlated with the normalized
233 phenotypic effects; NVIQ Pearson's $R = 0.63$, permutation test $P = 0.02$ (Fig. 4a; Supplementary Fig. 12).
234 Very weak correlations were obtained for randomly permuted data, i.e. when truncating mutations were
235 randomly re-assigned to different exons in the same gene (average NVIQ Pearson's $R = 0.18$; see
236 Methods). Since the heritability of intelligence is known to substantially increase with age⁴¹, we also
237 investigated how the results depend on the age of probands. When we restricted our analysis to the older
238 half of probands in SSC (i.e. older than the median age of 8.35 years), the strength of the correlations
239 between the predicted dosage changes and normalized phenotypic effects increased further; NVIQ
240 Pearson's $R = 0.75$, average permuted $R = 0.2$, permutation test $P = 0.019$ (Fig. 4b; Supplementary Fig. 13).
241 The strong correlations between target exon expression and intellectual ASD phenotypes suggest that,
242 when gene-specific PDS values are taken into account, a significant fraction (30%-45%) of the relative
243 phenotypic effects of *de novo* LGD mutations across genes can be explained by the resulting overall dosage
244 changes of target genes.

245 We next evaluated the ability of our linear dosage model, based on calculated PDS parameters,
246 to explain the effects of LGD mutations on non-normalized IQs. To that end, for each gene with multiple
247 truncating mutations in different probands, we used our linear regression model to perform leave-one-
248 out predictions for IQ scores, i.e. we used PDS values calculated based on all but one probands with
249 mutations in the gene to estimate IQ values for the left out proband (Fig. 4c, inset; see Methods). Despite
250 the minimalism of our model and multiple simplifying assumptions, for LGD mutations that trigger NMD,
251 the model median inference error for NVIQ was 11.1 points (Fig. 4c; Supplementary Fig. 14), which is
252 significantly smaller than median NVIQ difference between probands with LGD mutations in the same
253 gene, 22.0 points (MWU one-tail test $P = 0.014$). The NVIQ inferences based on probands of the same
254 gender had significantly smaller errors compared to inferences based on probands of the opposite gender;
255 same gender NVIQ median error 9.1 points, different gender median error 19.9 points (MWU one-tail test
256 $P = 0.018$). Similar to normalized phenotypic effects (Fig. 4a,b), the inference errors decreased for older
257 probands; for example, for probands older than 12 years, the median NVIQ inference error 7.6 points (Fig.
258 4c, Supplementary Fig. 14 and 15).

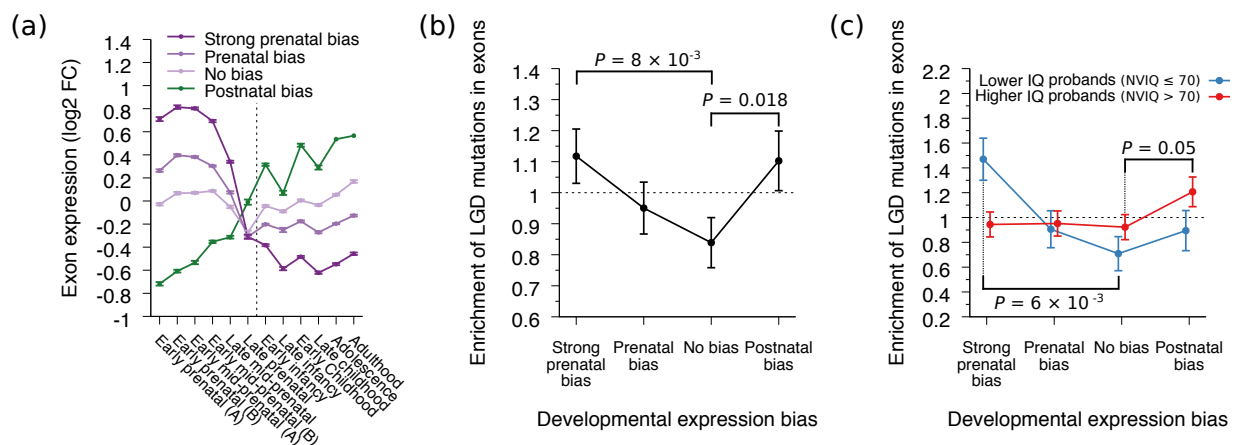
259



260
 261 **Figure 4:** Relationship between the relative expression of exons harboring LGD mutations and the
 262 corresponding decrease in probands' intellectual phenotypes. (a) Each point corresponds to a proband
 263 with an LGD mutation in a gene; only genes with multiple LGD mutations in the SSC cohort were
 264 considered. The x-axis represents the relative exon expression for exons harboring the LGD mutations.
 265 The y-axis represents the normalized decrease in the affected proband's NVIQ, i.e. the NVIQ decrease
 266 divided by the NVIQ phenotypic dosage sensitivity (PDS) of that gene (see Methods). The regression line
 267 across all points is shown in red; P -values were calculated based on randomly shuffled data (see Methods).
 268 (b) Same as (a), but with the analysis restricted to the older half of SSC probands (i.e. older than the
 269 median age 8.35 years). (c) Boxplots represent the distribution of errors in predicting the effects of LGD
 270 mutations on NVIQ (see Methods); NVIQ prediction errors are shown for all probands (green), and for
 271 probands older than 12 years (purple). For comparison, the average differences in NVIQ scores between
 272 probands with LGD mutations in the same gene are also shown (blue). Only genes with multiple LGD
 273 mutations in SSC were considered. The ends of each solid box represent the upper and lower quartiles;
 274 the horizontal lines inside each box represent the medians; and the whiskers represent the 5th and 95th

275 percentiles. The inset panel illustrates the linear regression model used to perform leave-one-out
 276 predictions of probands' NVIQs. Round open points represent observed phenotypic scores for probands
 277 with LGD mutations in the same gene, the grey square point represents the predicted phenotypic score,
 278 and the red dotted line represents the prediction error.

279
 280 Given that relative exon usage varies across neural development^{39, 42}, we investigated the
 281 relationship between developmental expression profiles of exons and ASD phenotypes. To that end, we
 282 sorted exons from genes harboring LGD mutations¹⁴ into four groups (quartiles) based on their
 283 developmental expression bias, which was calculated as the fold-change between prenatal and postnatal
 284 exon expression levels (Fig. 5a). We then analyzed the enrichment of LGD mutations in each exon group
 285 (see Methods). Compared to exons with no substantial developmental bias, we found significant
 286 enrichment of LGD mutations not only in exons with a strong prenatal bias (binomial one-tail test $P =$
 287 8×10^{-3} , Relative Rate = 1.33), but also in exons with postnatal biases ($P = 0.018$, RR = 1.31) (Fig. 5b). To
 288 understand the origin of the observed exon biases, we stratified probands into lower (≤ 70) and higher IQ
 289 (> 70) cohorts (Fig. 5c). Interestingly, while LGD mutations associated with lower IQs were strongly
 290 enriched only in prenatally biased exons (binomial one-tail test $P = 6 \times 10^{-3}$, RR = 1.62), mutations
 291 associated with higher IQs were exclusively enriched in postnatally biased exons ($P = 0.05$, RR = 1.27).
 292 These results demonstrate that mutations in exons with biases towards prenatal and postnatal expression
 293 preferentially contribute to ASD cases with lower and higher IQ phenotypes, respectively. We note that
 294 the observed exon developmental biases for LGD mutations are not simply driven by biases at the gene
 295 level, as mutations associated with both higher and lower IQ phenotypes showed enrichment exclusively
 296 towards genes with prenatally biased expression (Supplementary Fig. 16).
 297



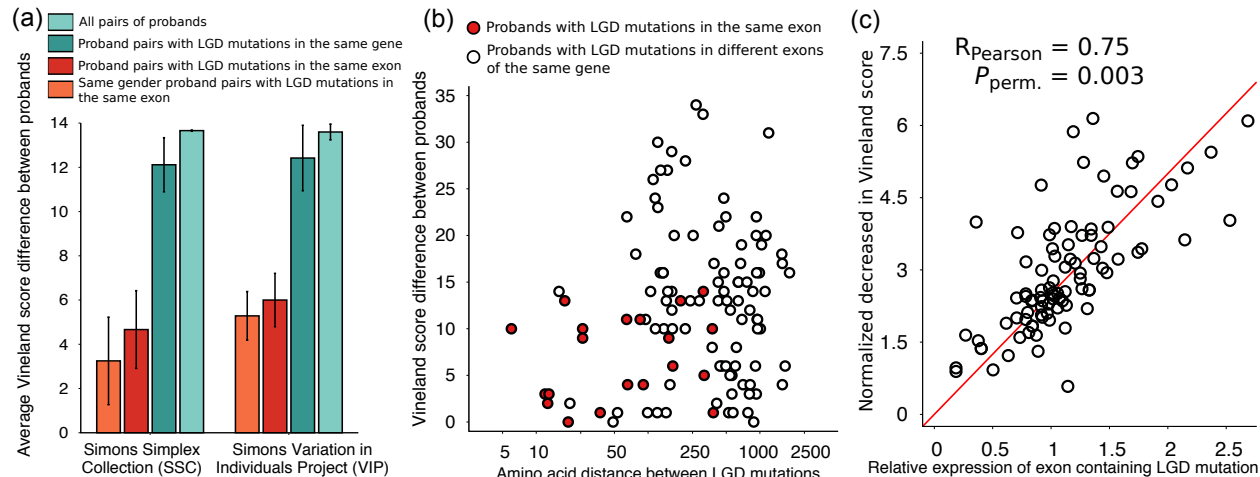
298
 299 **Figure 5:** Relationship between the developmental expression of exons and intellectual ASD phenotypes.
 300 (a) Exon developmental expression profiles for genes with *de novo* LGD mutations in SSC. Exons from all
 301 genes harboring LGD mutations were sorted into four groups (“strong prenatal bias”, “prenatal bias”, “no
 302 bias”, and “postnatal bias”) based on their overall developmental expression bias; the developmental bias
 303 was calculated as the log₂ fold change between the average prenatal and postnatal exon expression levels.
 304 Lines represent the average expression profiles for exons in each group, and the x-axis represents 12
 305 periods of human brain development, based on data from the Allen Institute’s BrainSpan atlas³⁹. The
 306 vertical dotted line delineates prenatal and postnatal developmental periods. Error bars represent the
 307 SEM. (b,c) Enrichment of LGD mutations across the four exon groups with different developmental biases.
 308 The y-axes represent the enrichment (relative rate) of mutations in each exon group; the enrichment was
 309 calculated by randomizing LGD mutations across exons proportionally to the exons’ coding sequence

310 lengths (see Methods). Error bars represent the SEM. **(b)** The overall enrichment of LGD mutations across
311 the four exon groups of exons with different developmental expression biases. **(c)** The enrichment of LGD
312 mutations across the four exon groups calculated separately for ASD probands with higher (>70, red) and
313 lower (≤ 70 , blue) nonverbal IQs.

314
315 Although we primarily analyzed the impact of autism mutations on intellectual phenotypes,
316 similar dosage and isoform expression changes of affected genes may also lead to analogous results for
317 other quantitative ASD phenotypes^{24, 43}. Indeed, for LGD mutations predicted to lead to NMD, we
318 observed similar patterns for several other key autism phenotypes. Specifically, SSC probands with
319 truncating mutations in the same exon exhibited more similar adaptive behavior abilities compared to
320 probands with mutations in the same gene (Fig. 6a, left set of bars, Supplementary Fig. 17); Vineland
321 Adaptive Behavior Scales (VABS)⁴⁴ composite standard score difference of 4.7 versus 12.1 points (Mann-
322 Whitney U one-tail test $P = 0.017$). In contrast, VABS differences between probands with truncating
323 mutations in the same gene were not significantly different than for randomly paired probands (Fig. 6a,
324 Supplementary Fig. 17); 12.1 versus 13.7 points (MWU one-tail test $P = 0.23$). Furthermore, probands with
325 truncating mutations in the same exon also displayed more similar fine motor skills; in the Purdue
326 Pegboard Test, 1.2 versus 3.0 for the average difference in normalized tasks completed with both hands
327 (MWU one-tail test $P = 0.02$; Supplementary Fig. 18; see Methods). Coordination scores in the Social
328 Responsiveness Scale questionnaire were also more similar in probands with LGD mutations in the same
329 exon compared to probands with mutations in the same gene; 0.6 versus 1.1 for the average difference
330 in normalized response (MWU one-tail test $P = 0.05$; Supplementary Fig. 19).

331 Finally, we sought to validate the observed phenotypic patterns using an independent cohort of
332 ASD probands. To that end, we analyzed an independently collected dataset from the ongoing Simons
333 Variation in Individuals Project (VIP)²⁹. The analyzed VIP dataset contained genetic information and VABS
334 phenotypic scores for 41 individuals with *de novo* LGD mutations in 12 genes. Reassuringly, and consistent
335 with our findings in SSC, probands from the VIP cohort with truncating *de novo* mutations in the same
336 exon also exhibited strikingly more similar VABS phenotypic scores compared to probands with mutations
337 in the same gene (Fig. 6a, right set of bars, Supplementary Fig. 20); VABS composite standard score
338 difference 6.0 versus 12.4 (Mann-Whitney U one-tail test $P = 0.014$). Similar to the SSC cohort, LGD
339 mutations in neighboring exons did not result in more similar behavior phenotypes; VABS composite
340 standard score average difference 13.6 points (MWU one-tail test $P = 0.6$). The fraction of truncated
341 proteins also did not show significant correlation with the VABS scores of affected probands (Pearson's R
342 = -0.08, $P = 0.7$). Using VABS scores from both SSC and VIP, we next investigated whether, analogous to
343 the IQ phenotypes (Fig. 3a), the similarity of VABS scores is primarily due to the presence of mutations in
344 the same exon, rather than proximity of truncating mutations within the corresponding protein sequence.
345 Indeed, LGD mutations in the same exon often resulted in similar adaptive behavior abilities even when
346 the corresponding mutations were separated by hundreds of amino acids (Fig. 6b, red points,
347 Supplementary Fig. 21). By comparing mutations in the same exon to mutations separated by similar
348 amino acid distances in the same protein but not necessarily the same exon, we confirmed that probands
349 with mutations in the same exon were significantly more phenotypically similar (permutation test $P =$
350 3×10^{-4} ; Supplementary Fig. 22; see Methods). When we applied the linear dosage model while accounting
351 for the VABS sensitivity to changes in the dosage of different genes (i.e. gene-specific PDS values), we
352 found substantial correlations between the relative expression of exons harboring LGD mutations and the
353 normalized VABS phenotypes of the affected probands (Pearson $R = 0.75$, permutation test $P = 0.003$; Fig.
354 6c; Supplementary Fig. 23). Overall, these results confirm the generality of the phenotypic patterns
355 observed in the SSC cohort.

356



357
358 **Figure 6:** Validation of the observed IQ patterns using Vineland Adaptive Behavior Scales (VABS) scores.
359 (a) Average VABS score differences between probands using data from SSC (left set of bars) and VIP (right
360 set of bars). Each bar shows the average difference in VABS scores between pairs of probands in different
361 groups. In the SSC and VIP bar sets, from right to left, bars represent differences between all pairs of
362 probands in each cohort (light green), between probands with LGD mutations in the same gene (dark
363 green), between probands with LGD mutations in the same exon (red), and between probands of the
364 same gender with LGD mutations in the same exon (orange). Error bars represent the SEM. (b) Amino acid
365 distance between LGD mutations in the same protein versus differences in VABS score. Each point
366 corresponds to a pair of probands, with individual from either SSC or VIP, with LGD mutations in the same
367 gene. The x-axis represents the amino acid distance between LGD mutations, and the y-axis represents
368 the difference between the affected probands' VABS scores. Red points represent proband pairs with LGD
369 mutations in the same exon, and white points represent proband pairs with LGD mutations in different
370 exons of the same gene. (c) Relationship between the relative expression of exons harboring LGD
371 mutations and the corresponding decrease in probands' normalized VABS scores. Each point corresponds
372 to a proband with an LGD mutation in a gene; only genes with multiple LGD mutations were considered.
373 The x-axis represents the relative expression (exon expression divided by total gene expression) of exons
374 harboring LGD mutations. The y-axis represents the affected probands' normalized decrease in VABS
375 scores, i.e. the VABS decrease divided by the VABS phenotypic dosage sensitivity (PDS) of that gene. The
376 regression line across all points is shown in red; P -values were calculated based on randomly shuffled data
377 (see Methods). The analysis was restricted to *de novo* LGD mutations predicted to trigger NMD, i.e. more
378 than 50 bp upstream from the last exon junction.

379
380

381 Discussion

382

383 Previous studies explored phenotypic similarity in syndromic forms of ASD due to mutations in
384 specific genes⁴⁵⁻⁴⁹. Nevertheless, across a large collection of contributing genes, the nature of the
385 substantial phenotypic heterogeneity in ASD is not well understood. Interestingly, the diversity of
386 intellectual and other important ASD phenotypes resulting from *de novo* LGD mutations in the same genes
387 is usually only slightly (~10%) smaller than the phenotypic diversity across the entire ASD cohort (Fig. 1,
388 Fig. 6a). The presented results suggest that truncating mutations usually result in a range of relatively mild
389 NMD-induced gene dosage changes, on average decreasing gene expression by ~15-30% (Supplementary
390 Fig. 24; see Methods). Our study further suggests a hierarchy of biological mechanisms contributing to

391 phenotypic heterogeneity in simplex ASD cases triggered by LGD mutations in different genes and within
392 the same gene.

393 Across LGD mutations, there is a significant but small correlation between a target gene's brain
394 expression level and the resulting intellectual phenotype ($R^2 = 0.02$, $P = 0.03$). This correlation is small, at
395 least in part, due to the significant variability of expression levels across different exons in a gene. Indeed,
396 intellectual phenotypes correlate significantly better with the relative expression level of exons harboring
397 LGD mutations ($R^2 = 0.10$, $P = 0.011$). In addition to effects associated with different expression levels of
398 exons, there is also substantial variability in the sensitivity of each specific phenotype to dosage changes
399 of different genes. When we account for varying dosage sensitivities using gene-specific PDS values, the
400 correlation between predicted dosage changes and normalized phenotypic effects becomes substantial
401 ($R^2=0.4$, $P = 0.02$, Fig. 4a; Fig. 6c). As the heritability of IQ phenotypes usually increases with age, we
402 observe even stronger dosage-phenotype correlations for older probands ($R^2 = 0.56$, $P = 0.019$, Fig. 4b).
403 Furthermore, even perturbations leading to similar dosage changes in the same gene may result in diverse
404 phenotypes in cases where different, functionally distinct splicing isoforms are truncated. However, when
405 exactly the same sets of isoforms are perturbed, as for LGD mutations in the same exon, the resulting
406 phenotypes, even in unrelated ASD probands, become especially similar (Fig. 1, Fig. 6a). For LGD mutations
407 affecting intellectual phenotypes, we found that same exon membership accounts for a larger fraction of
408 phenotypic variance than multiple other genomic features, including expression, evolutionary
409 conservation, pathway membership, and domain truncation (see Methods). There are likely deviations
410 from the aforementioned patterns for specific genes and specific truncating mutations. For example,
411 truncated proteins that escape NMD may lead to partial buffering, due to remaining activity, or to further
412 damaging effects, due to dominant negative interactions. Nevertheless, our results demonstrate that for
413 *de novo* LGD mutations in ASD, exons, rather than genes, usually represent a unit of effective phenotypic
414 impact.

415 Our results also suggest that changes in ASD phenotypes induced by LGD mutations may be
416 characterized by a simple linear model quantifying the sensitivity of a phenotype to changes in gene
417 dosage. We observe that PDS values for the same phenotype vary substantially across genes
418 (Supplementary Fig. 11), and that PDS differences are a major source of phenotypic variability. Moreover,
419 PDS values for the same gene vary across phenotypes (for example, correlation between PDS values for
420 IQ and VABS across 24 genes, $R^2 = 0.37$, P -values=0.001), which suggests that PDS values are specific to
421 phenotype-gene pairs. Although we evaluated PDS parameters using predicted NMD-induced dosage
422 changes, it may be possible to infer these parameters using other mechanisms of dosage change, such as
423 regulatory mutations. As genetic and phenotypic data accumulate, it will be interesting to estimate gene-
424 specific PDS values for multiple phenotypes and for a substantial number of ASD risk genes. Furthermore,
425 due to consistent patterns of gene expression changes across tissues (Fig. 3), it may be possible to
426 estimate PDS parameters for other genetic disorders and phenotypes. We note in this respect that
427 quantitative gene-dosage relationships have been recently characterized for yeast fitness values in
428 different environmental conditions⁵⁰.

429 In the present study, we focused specifically on simplex cases of ASD, in which *de novo* LGD
430 mutations are highly penetrant and where the contribution of genetic background is minimized. It is likely
431 that differences in genetic background and environment represent other important sources of phenotypic
432 variability^{22, 51, 52}. Therefore, in more diverse cohorts, individuals with LGD mutations in the same exon will
433 likely display greater phenotypic heterogeneity. For example, the Simons Variation in Individuals Project
434 identified broad spectra of phenotypes associated with specific variants in the general population^{29, 53-55}.
435 We also observed significantly larger phenotypic variability for probands from sequenced family trios, i.e.
436 families without unaffected siblings (Supplementary Fig. 25). For these probands, the enrichment of *de*
437 *nov*o LGD mutations is substantially lower and the contribution from genetic background is likely to be
438 larger⁵⁶, resulting in more pronounced phenotypic variability.

439 Our study may have important implications for precision medicine^{51, 57, 58}. The presented results
440 indicate that relatively mild decreases in affected gene dosage may account for a substantial fraction of
441 adverse phenotypic consequences. Thus, from a therapeutic perspective, compensatory expression of
442 intact alleles, as was recently demonstrated in mouse models of ASD⁵⁹⁻⁶¹ and other diseases⁶², may
443 provide an approach for alleviating phenotypic effects for at least a fraction of ASD cases. From a
444 prognostic perspective, our results suggest that by sequencing and phenotyping sufficiently large patient
445 cohorts with truncating mutations in different exons, it may be possible to understand likely phenotypic
446 consequences originating from LGD mutations in specific exons. Furthermore, because we observe
447 consistent patterns of expression changes across multiple human tissues, similar analyses may be also
448 extended to other disorders affected by highly penetrant truncating mutations.

449

450

451 **Acknowledgements**

452

453 We thank Drs. W.K. Chung, I. Pe'er, A. Packer, and all members of the Vitkup lab for helpful
454 scientific discussions. D.V. acknowledges funding from the Simons Foundation (SFARI #308962). This work
455 was supported in part by NIH grant no. T15LM007079 (A.H.C., J.C., J.W.) and Ruth L. Kirschstein National
456 Research Service Award Institutional Research Training grant no. T32GM082797 (A.H.C.).

457

458 References

- 459
- 460 1. American Psychiatric Association (DSM-5 Task Force). *Diagnostic and Statistical*
461 *Manual of Mental Disorders: DSM-5*. 5th edn. American Psychiatric Association:
462 Washington, DC, 2013.
- 463
- 464 2. Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism
465 genetics and molecular neuroscience. *Trends in Neuroscience* 2014; **37**(2): 95-105.
- 466
- 467 3. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics
468 of autism spectrum disorders. *Nature Reviews Genetics* 2014; **15**(2): 133-141.
- 469
- 470 4. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of
471 autism disease mechanisms through genetics. *Nature Medicine* 2016; **22**(4): 345-361.
- 472
- 473 5. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder
474 through genetic findings. *Nature Reviews Neurology* 2014; **10**(2): 74-81.
- 475
- 476 6. Talkowski ME, Minikel EV, Gusella JF. Autism Spectrum Disorder Genetics: Diverse
477 Genes with Diverse Clinical Outcomes. *Harvard Review of Psychiatry* 2014; **22**(2): 65-
478 75.
- 479
- 480 7. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M *et al*. Diverse types of
481 genetic variation converge on functional gene networks involved in schizophrenia.
482 *Nature Neuroscience* 2012; **15**(12): 1723-1728.
- 483
- 484 8. Ayalew M, Le-Niculescu H, Levey DF, Jain N, Changala B, Patel SD *et al*. Convergent
485 functional genomics of schizophrenia: from comprehensive understanding to genetic risk
486 prediction. *Molecular Psychiatry* 2012; **17**(9): 887-905.
- 487
- 488 9. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al*. De
489 novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**(7487):
490 179-184.
- 491
- 492 10. Parikshak NN, Gandal MJ, Geschwind DH. Systems biology and gene networks in
493 neurodevelopmental and neurodegenerative disorders. *Nature Reviews Genetics* 2015;
494 **16**(8): 441-458.
- 495
- 496 11. Chang J, Gilman SR, Chiang AH, Sanders SJ, Vitkup D. Genotype to phenotype
497 relationships in autism spectrum disorders. *Nature Neuroscience* 2015; **18**(2): 191-198.
- 498
- 499 12. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants
500 associated with autism implicate a large functional network of genes involved in
501 formation and function of synapses. *Neuron* 2011; **70**(5): 898-907.
- 502
- 503 13. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D *et al*.
504 Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams
505 syndrome region, are strongly associated with autism. *Neuron* 2011; **70**(5): 863-885.
- 506

- 507 14. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution
508 of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**(7526): 216-
509 221.
- 510
- 511 15. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y *et al.* Large-
512 Scale Exome Sequencing Study Implicates Both Developmental and Functional
513 Changes in the Neurobiology of Autism. *Cell* 2020.
- 514
- 515 16. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism
516 exomes reveal a highly interconnected protein network of de novo mutations. *Nature*
517 2012; **485**(7397): 246-250.
- 518
- 519 17. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB *et al.* Most genetic risk
520 for autism resides with common variation. *Nature Genetics* 2014; **46**(8): 881-885.
- 521
- 522 18. Gratten J, Wray NR, Keller MC, Visscher PM. Large-scale genomics unveils the genetic
523 architecture of psychiatric disorders. *Nature Neuroscience* 2014; **17**(6): 782-790.
- 524
- 525 19. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G *et al.* Individual common
526 variants exert weak effects on the risk for autism spectrum disorders. *Human Molecular*
527 *Genetics* 2012; **21**(21): 4781-4792.
- 528
- 529 20. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K *et al.* Excess of rare,
530 inherited truncating mutations in autism. *Nature Genetics* 2015; **47**(6): 582-588.
- 531
- 532 21. Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC *et al.* Genomic
533 Patterns of De Novo Mutation in Simplex Autism. *Cell* 2017; **171**(3): 710-722.e712.
- 534
- 535 22. Robinson EB, Samocha KE, Kosmicki JA, McGrath L, Neale BM, Perlis RH *et al.* Autism
536 spectrum disorder severity reflects the average contribution of de novo and familial
537 influences. *Proceedings of the National Academy of Sciences* 2014; **111**(42): 15161-
538 15165.
- 539
- 540 23. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J *et al.* Rare de novo and
541 transmitted copy-number variation in autistic spectrum disorders. *Neuron* 2011; **70**(5):
542 886-897.
- 543
- 544 24. Buja A, Volfovsky N, Krieger AM, Lord C, Lash AE, Wigler M *et al.* Damaging de novo
545 mutations diminish motor skills in children on the autism spectrum. *Proceedings of the*
546 *National Academy of Sciences* 2018; **115**(8): E1859-E1866.
- 547
- 548 25. Taylor LJ, Maybery MT, Wray J, Ravine D, Hunt A, Whitehouse AJO. Are there
549 differences in the behavioural phenotypes of Autism Spectrum Disorder probands from
550 simplex and multiplex families? *Research in Autism Spectrum Disorders* 2015; **11**: 56-
551 62.
- 552
- 553 26. Dissanayake C, Searles J, Barbaro J, Sadka N, Lawson LP. Cognitive and behavioral
554 differences in toddlers with autism spectrum disorder from multiplex and simplex
555 families. *Autism Research* 2019; **12**(4): 682-693.
- 556

- 557 27. Berends D, Dissanayake C, Lawson LP. Differences in Cognition and Behaviour in
558 Multiplex and Simplex Autism: Does Prior Experience Raising a Child with Autism
559 Matter? *Journal of Autism and Developmental Disorders* 2019; **49**(8): 3401-3411.
560
- 561 28. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of
562 autism genetic risk factors. *Neuron* 2010; **68**(2): 192-195.
563
- 564 29. Simons VIP Consortium. Simons Variation in Individuals Project (Simons VIP): a
565 genetics-first approach to studying autism spectrum and related neurodevelopmental
566 disorders. *Neuron* 2012; **73**(6): 1063-1067.
567
- 568 30. Fombonne E. Epidemiology of Pervasive Developmental Disorders. *Pediatric Research*
569 2009; **65**: 591-598.
570
- 571 31. Robinson EB, Lichtenstein P, Anckarsäter H, Happé F, Ronald A. Examining and
572 interpreting the female protective effect against autistic behavior. *Proceedings of the*
573 *National Academy of Sciences* 2013; **110**(13): 5258-5262.
574
- 575 32. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC *et al.* The Pfam protein
576 families database in 2019. *Nucleic Acids Research* 2018; **47**(D1): D427-D432.
577
- 578 33. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance
579 pathway. *Annual Review of Biochemistry* 2007; **76**: 51-74.
580
- 581 34. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot
582 analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235): 648-660.
583
- 584 35. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M *et al.* Human
585 genomics. The human transcriptome across tissues and individuals. *Science* 2015;
586 **348**(6235): 660-665.
587
- 588 36. Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ *et al.* Human
589 genomics. Effect of predicted protein-truncating genetic variants on the human
590 transcriptome. *Science* 2015; **348**(6235): 666-669.
591
- 592 37. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon
593 definition and function. *Nature Reviews Genetics* 2010; **11**(5): 345-355.
594
- 595 38. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A *et al.*
596 Widespread expansion of protein interaction capabilities by alternative splicing. *Cell*
597 2016; **164**(4): 805-817.
598
- 599 39. Kang HJ, Kawasaki YI, Cheng F, Zhu Y, Xu X, Li M *et al.* Spatio-temporal
600 transcriptome of the human brain. *Nature* 2011; **478**(7370): 483-489.
601
- 602 40. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes:
603 when nonsense affects RNA abundance. *Trends in Biochemical Sciences* 1998; **23**(6):
604 198-199.
605

- 606 41. Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM
607 *et al.* The heritability of general cognitive ability increases linearly from childhood to
608 young adulthood. *Molecular Psychiatry* 2009; **15**(11): 1112-1120.
609
- 610 42. Weyn-Vanhentenryck SM, Feng H, Ustianenko D, Duffie R, Yan Q, Jacko M *et al.*
611 Precise temporal regulation of alternative splicing during neural development. *Nature*
612 *Communications* 2018; **9**(1): 2189.
613
- 614 43. Bishop SL, Farmer C, Bal V, Robinson E, Willsey AJ, Werling DM *et al.* Identification of
615 Developmental and Behavioral Markers Associated with Genetic Abnormalities in Autism
616 Spectrum Disorder. *The American Journal of Psychiatry* 2017; **174**(6): 576-585.
617
- 618 44. Zerbino DR, Achuthan P, Akanni W, Amode M R, Barrell D, Bhai J *et al.* Ensembl 2018.
619 *Nucleic Acids Research* 2017; **46**(D1): D754-D761.
620
- 621 45. Sztainberg Y, Zoghbi HY. Lessons learned from studying syndromic autism spectrum
622 disorders. *Nature Neuroscience* 2016; **19**(11): 1408-1417.
623
- 624 46. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O *et al.* Disruptive CHD8
625 mutations define a subtype of autism early in development. *Cell* 2014; **158**(2): 263-276.
626
- 627 47. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende
628 J *et al.* A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP.
629 *Nature Genetics* 2014; **46**(4): 380-384.
630
- 631 48. Van Bon B, Coe B, Bernier R, Green C, Gerds J, Witherspoon K *et al.* Disruptive de
632 novo mutations of DYRK1A lead to a syndromic form of autism and ID. *Molecular*
633 *Psychiatry* 2016; **21**(1): 126-132.
634
- 635 49. Ben-Shalom R, Keeshen CM, Berrios KN, An JY, Sanders SJ, Bender KJ. Opposing
636 Effects on NaV1.2 Function Underlie Differences Between SCN2A Variants Observed in
637 Individuals With Autism Spectrum Disorder or Infantile Seizures. *Biological Psychiatry*
638 2017; **82**(3): 224-232.
639
- 640 50. Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S *et al.*
641 Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness.
642 *Cell* 2016; **166**(5): 1282-1294.e1218.
643
- 644 51. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH. The road to precision
645 psychiatry: translating genetics into disease mechanisms. *Nature Neuroscience* 2016;
646 **19**(11): 1397-1407.
647
- 648 52. Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J *et al.*
649 Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general
650 population. *Nature Genetics* 2016; **48**(5): 552-555.
651
- 652 53. Qureshi AY, Mueller S, Snyder AZ, Mukherjee P, Berman JI, Roberts TP *et al.* Opposing
653 brain differences in 16p11. 2 deletion and duplication carriers. *The Journal of*
654 *Neuroscience* 2014; **34**(34): 11199-11211.
655

- 656 54. Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG *et al.* The
657 cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained
658 population. *Biological Psychiatry* 2015; **77**(9): 785-793.
659
- 660 55. D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L *et al.* Defining
661 the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities.
662 *JAMA Psychiatry* 2016; **73**(1): 20-30.
663
- 664 56. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K *et al.* A unified
665 genetic theory for sporadic and inherited autism. *Proceedings of the National Academy
666 of Sciences* 2007; **104**(31): 12831-12836.
667
- 668 57. Collins FS, Varmus H. A New Initiative on Precision Medicine. *New England Journal of
669 Medicine* 2015; **372**(9): 793-795.
670
- 671 58. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to
672 precision medicine. *The Lancet Neurology* 2015; **14**(11): 1109-1120.
673
- 674 59. Guy J, Gan J, Selfridge J, Cobb S, Bird A. Reversal of Neurological Defects in a Mouse
675 Model of Rett Syndrome. *Science* 2007; **315**(5815): 1143-1147.
676
- 677 60. Mei Y, Monteiro P, Zhou Y, Kim J-A, Gao X, Fu Z *et al.* Adult restoration of Shank3
678 expression rescues selective autistic-like phenotypes. *Nature* 2016; **530**(7591): 481-484.
679
- 680 61. Ehninger D, Han S, Shilyansky C, Zhou Y, Li W, Kwiatkowski DJ *et al.* Reversal of
681 learning deficits in a Tsc2+/- mouse model of tuberous sclerosis. *Nature Medicine* 2008;
682 **14**(8): 843-848.
683
- 684 62. Matharu N, Rattanasopha S, Tamura S, Maliskova L, Wang Y, Bernard A *et al.* CRISPR-
685 mediated activation of a promoter or enhancer rescues obesity caused by
686 haploinsufficiency. *Science* 2019; **363**(6424): eaau0629.
687
- 688
- 689 **Conflict of Interest statement:** The authors declare that there is no conflict of interest.
690