Unbiased human genomic characterization of polyglutamine disorder genes to guide biological understanding and therapeutic strategies

Kevin Lucy Namuli,[1,2] Britt I. Drögemöller,[3] Galen E.B. Wright,[1,2,3]

Author affiliations:

[1]Department of Pharmacology and Therapeutics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada.

[2]PrairieNeuro Research Centre, Kleysen Institute for Advanced Medicine, Health Sciences Centre and Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada

[3]Department of Biochemistry and Medical Genetics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada.

Correspondence to: Galen E.B. Wright
Department of Pharmacology and Therapeutics
University of Manitoba
Winnipeg, MB, Canada
Phone: +1 (204) 272-3130
galen.wright@umanitoba.ca

Running title: Genomic Insights Guide PolyQ Therapeutics

Financial Disclosure/Conflict of Interests: The authors report no competing interests.

Funding Sources: Natural Sciences and Engineering Research Council of Canada (Canada Research Chairs Initiative CRC-2019-145 and Discovery Grant Program RGPIN-2022-04509), Health Sciences Centre Foundation Winnipeg, Canadian Institutes of Health Research (Canada Research Chairs Initiative CRC-2023-00351), Research Manitoba and Rady Faculty of Health Sciences, University of Manitoba.

Abstract

Background: Polyglutamine (polyQ) disorders, such as Huntington disease (HD) and several spinocerebellar ataxias, are severe neurological disorders caused by glutamine codon repeat expansions. These conditions lack effective treatments, with therapeutic research focused on pathogenic gene knockdown.

Objectives: We aimed to profile these genes using diverse human genomic data to inform therapeutic strategies by identifying new biology and assessing the potential on-target effects of knocking down these genes.

Methods: We conducted an unbiased phenome-wide study to identify human traits and diseases linked to polyQ disorder genes (Open Targets L2G>0.5). Network analyses explored shared trait associations and overlapping biological processes among these genes. Lastly, we assessed the theoretical druggability of polyQ disorder genes using recently identified features predictive of clinical trial success and compared them to repeat expansion (HD) modifier genes.

Results: We identified 215 human phenotype/polyQ disorder gene associations from 3,095 studies, indicating potential adverse effects from gene knockdown. Shared trait associations among genes suggested overlapping biological processes despite distinct functions. Drug target profile analysis revealed unfavorable risk profiles for polyQ disorder genes, particularly *ATN1*, *ATXN1*, *ATXN7*, and *HTT*, due to genomic features such as constraint, molecular interactions, and tissue specificity. PolyQ disorder genes also showed significantly more safety-related risks than HD genetic modifier genes (P=7.03x10^{-3}).

Conclusion: Our analyses emphasize the pleiotropic nature of polyQ disorder genes, highlighting their potential risks as drug targets. These findings reinforce the importance of exploring alternative therapeutic strategies, such as targeting genetic modifier genes, as well as allele-selective approaches, to mitigate these challenges.

Introduction

Polyglutamine (polyQ) disorders are a group of ten heritable neurodegenerative conditions characterized by repeat expansions of the glutamine codon in distinct genes.[1] These include Huntington disease (HD), spinal and bulbar muscular atrophy (SBMA), dentatorubral-pallidoluysian atrophy (DRPLA), and several spinocerebellar ataxias (i.e., SCAs: 1-3, 6, 7 and 17).[1] Collectively, they affect approximately 1-10 individuals per 100,000 people, contributing significantly to the increasing burden of neurological disorders globally.[2]

To date, some progress has been made in elucidating the pathogenic mechanisms of polyQ disorders. Unfortunately, these disorders remain without a cure, with current pharmacological treatments focused on alleviating the symptoms.[3,4] While gene knockdown strategies, such as antisense oligonucleotides (ASOs), have been the focus of current research, and some of these ASOs have advanced to clinical trial stages, efficacy and safety issues remain substantial barriers to their clinical implementation.[3,5–7] For example, the ASO tominersen demonstrated adverse events and failed to achieve efficacy. As tominersen suppresses the production of wild-type *HTT*,[8,9] this disrupts normal gene function, highlighting the challenges associated with the therapeutic targeting of polyQ genes. This is particularly important given that all polyQ disorders present in an autosomal dominant manner, except for SBMA.

The role of adverse events in clinical trial failures underscores the importance of understanding the biological consequences of targeting specific genes when designing therapeutics.[10] In line with this, Plenge *et al.*[11] highlighted the significance of evaluating potential adverse clinical consequences *a priori* by identifying human phenotypes associated with genetic variation related to drug target genes.[11] The increasing availability of extensive genome-wide association studies (GWAS) and related repositories, assisted by fine-mapping techniques that link genetic associations to suspected causal genes,[12,13] allows for systematic and unbiased identification of gene-trait links. For example, we previously showed that common *HTT* genetic variation is associated with cognitive function in large human cohorts.[14,15]

Associations between polyQ disorder genes and human traits can also provide insights into the biological processes that these genes are involved in, aiding in the identification of additional

potential therapeutic targets. Several studies have demonstrated that drugs with prior genetic evidence supporting the role that the targeted gene plays in the clinical phenotype being treated are more likely to succeed in clinical trials.[13,16–20] Notably, a recent study also identified additional genetic factors associated with clinical trial outcomes,[19] showing that genes displaying human genetic constraint, several interacting partners and broad gene expression were associated with clinical trial stoppage. These studies illustrate the importance of incorporating human genomic information into drug development pipelines.

The challenges highlighted above underscore the need to comprehensively evaluate polyQ disorder genes to identify potential on-target effects and investigate genomic features associated with therapeutic failures. Here, we aimed to characterize polyQ disorder genes using extensive human genomic data to inform therapeutic studies and disease biology. First, we analyzed and prioritized human trait and disease associations with common genetic variation in polyQ disorder genes using large repositories. We unbiasedly profiled important human genomic information through this approach to identify potential on-target effects. We also uncovered unique and shared biological features by examining phenotypes associated with these genes outside of the polyQ disorders themselves. Finally, we evaluated the theoretical druggability of the polyQ disorder genes, ranking them according to features related to clinical trial success and comparing them to the repeat expansion modifier genes that influence disease onset. Our comprehensive analysis enhances our understanding of the underlying biology of polyQ disorders and offers potential insights into the consequences of targeting the associated genes for polyQ disorder therapeutics. The findings from our study have broad relevance to therapeutic strategies for other repeat expansion disorders, of which more than 50 have been identified.

Material and Methods

Ethical approval and gene selection

Ethical approval was obtained from the University of Manitoba Bannatyne Research Ethics Board (H2022:354 HS25743). We profiled all known polyQ disorder genes (*n*=10, i.e., *AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *CACNA1A*, *HTT*, *TBP* and *THAP11*). For GWAS-based analyses, we excluded *AR* because it is located on the X chromosome, and most current GWAS do not include this information.

64 **Identification of human traits and diseases associated with polyQ disorder genes**

65 To identify human traits and diseases associated with the polyQ disorder genes beyond their
66 primary disorders, an unbiased, comprehensive phenome-wide study was conducted by analyzing
67 data from the Open Targets Genetics database (OTG) (version 8).[12] This repository contains
68 approximately 132,893 curated GWAS association signals from various studies and genomic
69 biobanks (e.g., FinnGen and UK Biobank), including 8,317 human GWAS summary statistics
70 (extracted January 2024). To curate OTG data, we used the related application programming
71 interface, implemented using GraphQL v2.0[12] and executed in R 4.3.3.

72

73 We accessed polyQ disorder gene data using Ensembl identifiers.[21] We identified human traits
74 where significant associations could be attributed to polyQ disorder genes using the locus-to-gene
75 (L2G) fine-mapping tool.[22] L2G is designed to refine genome-wide significant GWAS signals
76 ($P<5x10^{-8}$), identifying likely causal genes attributed to these signals.[12] This tool uses a scoring
77 metric that quantifies the strength of evidence for gene-trait associations, ranging from 0 to 1.[12]
78 Higher scores signify increased support for a particular gene-trait association. Traits overlapping
79 with the polyQ disorder genes with L2G>0.5 were kept for further analysis as recommended by
80 the developers of this tool.[22]

81

82 Finally, to examine other severe Mendelian disorders and other phenotypes associated with polyQ
83 disorder genes outside of the canonical polyQ disorder, we profiled the Online Mendelian
84 Inheritance in Man (OMIM) database.[23] We analyzed the OMIM genemap2 dataset (accessed June
85 2022) containing 5,779 phenotype entries.

86

87 **Assessing shared traits among polyQ disorder genes**

88 To investigate whether polyQ disorder genes are linked to shared traits and the degree of this
89 overlap, we performed a network analysis using *ggraph* (version 2.1.0) and *arcdiagram* (version
90 0.1.12) in R. All gene-trait associations with an L2G>0.5 were included in the network analysis.
91 In the resulting network, each node represented a polyQ disorder gene, and each edge (link)
92 represented a bidirectional relationship shared between two genes. This network analysis aimed to

93 identify networks of genes linked to particular traits, thereby providing insights into underlying
94 mechanisms and shared biology.

95 **Prioritization of informative L2G-significant human trait associations for polyQ disorder**
96 **genes**

97 To prioritize individual traits and disease signals for the polyQ disorder genes, count-based traits
98 (e.g., blood cell types) were removed. This was performed to minimize non-specific signals, as
99 such traits are associated with numerous generic associations across the genome. Further, we
100 restricted our analyses to include only gene-trait pairs derived from peer-reviewed PubMed studies
101 to ensure the robustness of study findings.

102 **Variant annotation and unique signal identification**

103 We annotated the predicted functional effect of index variants related to GWAS trait signals to
104 understand their potential mechanistic contribution to the observed gene-trait associations.
105 Ensembl Variant Effect Predictor was used to identify the *most severe* consequence flag related to
106 each polyQ disorder gene. SpliceAI was used to predict whether the variants alter splicing.[24]
107 Genotype-Tissue Expression (GTEx) data were assessed to determine whether variants were
108 linked to gene expression changes of the related polyQ disorder genes. Further, we identified
109 independent signals at each gene by considering linkage disequilibrium (LD) between index
110 variants. Independent signals, defined as haplotypes, were determined by pruning index variants
111 with r^2>0.5 using the 1,000 Genomes European super-population with LDLink SNPClip.[25] Each
112 haplotype was described as a set of variants in LD (i.e., r^2>0.5) with each other. A single tag variant
113 was considered a proxy for all the variants in high LD and was used to capture these haplotypes.
114 The Finnish population was excluded from this analysis to minimize potential confounding factors
115 from extended haplotype blocks.[26] Resulting haplotypes were assigned to prioritized polyQ
116 disorder gene-trait pairs to determine unique signals at each locus.

117 **Assessment of the theoretical druggability and risk profiles of the polyQ disorder genes**

118 To thoroughly evaluate polyQ disorder genes for theoretical druggability and drug target profiles
119 related to the risk of clinical trial failure, we assessed recently identified features associated with
120 clinical trial success.[19] Genes were categorized as favorable/unfavorable under each category
121 based on previously determined thresholds.[19] The following features were included:

122 (i) Genomic constraint: We examined genes' tolerance to loss-of-function mutations using
123 sequencing data from the Genome Aggregation (gnomAD) Database, which contains
124 125,748 exomes and 15,708 genomes (gnomAD v2.1.1; accessed July 2022).[27] We
125 used the probability of loss-of-function intolerance (pLI) metric, which ranges from 0
126 to 1, with more constrained genes having higher scores. Genes with a high degree of
127 constraint were categorized as unfavourable, indicated by a rank greater than 0.9
128 (corresponding to *"extremely loss of function intolerant"* genes).[28]
129 (ii) Tissue expression: We accessed RNA tissue specificity data from the Human Protein
130 Atlas (HPA) database (accessed February 2023),[29] which uses data from the HPA and
131 GTEx to classify genes into five categories (i.e., *"tissue enriched"*, *"group enriched"*,
132 *"tissue enhanced"*, *"low tissue specificity"* and *"not detected"*) based on fold change
133 in mRNA expression across 37 organ systems.[30] Genes whose RNA tissue specificity
134 was classified as *"low tissue specificity"* were ranked unfavorable as they are broadly
135 expressed.
136 (iii) Molecular interactions (MI): We used the Open Targets Platform to access data from
137 the IntAct database[31] related to MI evidence for the genes (accessed October 2024).
138 We used the MI score, which ranges from 0 to 1 (scores closer to 1 are higher
139 confidence) and is built using diverse data (i.e., publications, experimental detection
140 methods, and interaction types).[32] Genes with more than ten interacting partners with
141 MI scores exceeding 0.42 (considered high confidence[19]) were unfavorable.
142 (iv) Druggability evidence: We assessed the Drug Gene Interaction Database (accessed July
143 2024)[33] to identify any known or potentially druggable genes among the polyQ disorder
144 genes. We assigned a favorable "*druggable any*" metric to polyQ disorder genes that
145 had any known or predicted evidence of interaction with drugs from this database.
146
147 Finally, we compared the number of unfavorable genomic features associated with increased
148 safety-related risk (i.e., constraint, expression and interactions) to these same metrics reported in
149 our recent analysis of HD modifier genes[34] using the Wilcoxon test accompanied by the calculation
150 of the effect size (r) to quantify the magnitude of the observed difference. This analysis aimed to
151 assess whether targeting one of these gene sets presents a theoretically greater safety risk than the
152 other. $P<0.05$ was considered significant in these analyses.

153 **Data sharing**

154 The data used in the analyses described in this study are publicly accessible to researchers
155 through the consortia that generated them. Analysis scripts for data accession, analysis and
156 visualization are accessible at https://github.com/Wright-Lab-Neurogenomics-
157 Research/polyQ_genes_characterization_analyses

158

159 ## Results

160 **Profiling polyQ disorder genes for associations with common human traits and diseases**

161 We identified 3,095 association signals (1,467 unique traits) surrounding polyQ disorder genes
162 from the OTG database. These traits were distributed across 20 categories based on health
163 conditions or biological processes, with the broad *measurement* category containing the most
164 associations (*n*=2,521). Applying a filtering threshold of L2G>0.5 retained seven polyQ disorder
165 genes (i.e., *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *CACNA1A*, *HTT* and *TBP*) and identified 215 gene-
166 trait pairs (149 unique traits), signifying associations with strong evidence for polyQ disorder gene
167 involvement (**Figure 1**). *ATXN1* had the highest number of L2G>0.5 significant traits reported
168 (*n*=57), predominantly related to hematological (measurement) phenotypes. *ATXN3* and *TBP* were
169 exclusively linked to hematological phenotypes, while *ATXN7* was associated solely with disease-
170 related phenotypes. Network analysis of the 215 polyQ disorder gene-trait pairs with L2G>0.5
171 revealed that five polyQ disorder genes shared phenotypic traits. *TBP* and *CACNA1A* did not share
172 traits with any polyQ disorder genes. *ATXN1* had the greatest traits shared with the other polyQ
173 disorder genes. *HTT* and *ATXN1* displayed the most shared traits, including educational attainment
174 and cognitive ability, in addition to neutrophil and white blood cell counts.

175

176 We applied further prioritization steps to obtain high confidence and informative traits linked to
177 polyQ disorder genes, including removing count-based terms and studies not found in PubMed.
178 These analyses retained 46 significant traits (L2G>0.5) and five polyQ disorder genes: *HTT,*
179 *ATXN1, ATXN2, ATXN7*, and *CACNA1A* (**Figure 2**). Furthermore, filtering out duplicate traits for
180 each gene retained 37 unique gene-trait associations. These informative trait associations represent
181 diverse disease phenotypes, including neuropsychiatric traits (e.g., *HTT*-depression, *CACNA1A*-
182 positive affect, *ATXN7*-schizophrenia), metabolic disorders (e.g., *ATXN2*-type 1 diabetes,

183 *HTT/ATXN7*-type 2 diabetes) and autoimmune disorders (e.g., *ATXN1*-multiple sclerosis, *ATXN1*-
184 systemic lupus erythematosus). Additionally, they include several non-neurological traits or
185 diseases, such as *ATXN1*-atrial fibrillation, *ATXN1*-/*ATXN7*-breast cancer, and *ATXN7*-cataracts.
186 PolyQ disorder genes, such as *ATXN1* and *HTT,* were associated with non-pathogenic neurogenetic
187 traits, including educational attainment, cognitive ability, and aging. This collective observation
188 highlights the pleiotropic features related to these genes outside of polyQ pathobiology.

189 **Investigation and annotation of the variants linked to unique polyQ disorder gene-common**
190 **trait associations**

191 The index variants linked with the high-confidence unique traits associated with the polyQ disorder
192 genes fall into three categories: intronic (*n*=26), splice (*n*=1), and variants located in untranslated
193 regions (*n*=1) (**Figure 2**). Notably, all traits associated with the splice and untranslated region
194 variants were annotated to *HTT*. In contrast, traits related to the remaining polyQ disorder genes
195 (i.e., *CACNA1A, ATXN1, ATXN2* and *ATXN7*) were exclusively associated with intronic variants.
196 The *HTT* splice variant showed associations with educational attainment and its noncognitive
197 aspects. *ATXN7* was associated with several pathogenic traits, including type 2 diabetes,
198 schizophrenia, breast cancer and cataracts, each associated with a unique variant (total *n*=4).
199

200 The 37 unique gene-trait associations arose from 30 distinct signals. A variant located in the
201 untranslated region in *HTT* was associated with the highest number of distinct traits (*n*=5),
202 including type 2 diabetes and walking pace (**Table 1**). The remaining unique variants resulted in
203 21 independent signals (i.e., *ATXN1, n*=8; *HTT, n*=6; *ATXN7, n*=3; *ATXN2, n*=2; *CACNA1A, n*=2)
204 referred to as haplotypes. The *HTT* haplotype two (index variant: intronic rs61348208), which we
205 have previously reported,[15] had the largest number of traits (*n*=6) associated with it, including both
206 pathogenic (e.g., depression) and non-pathogenic traits (e.g., aging traits and parental lifespan).
207 *ATXN1* exhibited the highest number of independent signals (*n*=8 haplotypes), with *ATXN1*
208 haplotype three (index variant: intronic rs909788) (*n*=5) demonstrating the most trait associations.
209 However, these were all linked to highly correlated cognitive and socio-economic traits.

210 **Profiling polyQ disorder genes for severe Mendelian diseases beyond canonical polyQ**
211 **disorders**

212 A subset of the polyQ disorder genes (i.e., *AR, ATN1, ATXN2, CACNA1A, HTT, TBP*) had
213 phenotype entries in the OMIM database beyond the primary polyQ disorders (**Table 2**). Rare

214 Mendelian disorder associations included androgen insensitivity and hypospadias related to *AR*,
215 and neurological conditions linked to *HTT* (i.e., neurodevelopmental disorder: Lopes-Maciel-
216 Rodan syndrome), *CACNA1A* (i.e., developmental and epileptic encephalopathy) and *ATN1* (i.e.,
217 congenital hypotonia, epilepsy, developmental delay, and digital anomalies). *ATXN2* and *TBP*
218 were also associated with susceptibility to Parkinson's disease, a common neurological disorder,
219 while *ATXN2* was also associated with susceptibility to amyotrophic lateral sclerosis (ALS), a
220 neurodegenerative disease. Finally, *AR* was also associated with prostate cancer susceptibility in
221 OMIM. These findings illustrate that therapeutic polyQ disorder gene knockdown may impair
222 critical developmental or neurological pathways.

223 **Analysis of human genomic safety profiles and the theoretical druggability of polyQ disorder**
224 **genes highlights potential safety concerns associated with therapeutic targeting**

225 We evaluated genomic features that predict clinical trial success to better understand the theoretical
226 implications of therapeutically perturbing polyQ disorder genes. Most of these genes showed
227 unfavorable profiles based on multiple criteria (**Figure 3**) using thresholds defined by the original
228 study that identified associations with these features and clinical trial stoppage.[35] Notably, four
229 genes (i.e., *ATN1, ATXN1*, *ATXN7* and *HTT*) ranked unfavorably for the three categories associated
230 with clinical trial stoppage due to safety concerns (i.e., genetic constraint, broad expression, and
231 several interactors). Further, all polyQ disorder genes displayed at least one unfavorable signal for
232 one of these three features. All polyQ disorder genes, except for *THAP11*, displayed several
233 molecular interactors (i.e., more than ten protein-protein interactions in the IntAct database). This
234 was especially apparent for *HTT*, *ATXN1*, *ATN1* and *ATXN3*, which showed over 100 high-
235 confidence interactions (i.e., ten times above the threshold used for classification). While five
236 polyQ disorder genes fell within the *'druggable any'* category, this metric indicates whether the
237 gene can be theoretically perturbed, not an indicator of safety. Compared to HD age-of-onset
238 modifier genes, polyQ disorder genes displayed significantly more unfavorable features associated
239 with an increased safety risk (P=7.03x10^{-3}; mean HD modifiers=1.0, n=9; mean polyQ
240 disorders=2.2, n=10; effect size r=0.628, large).

241

242

243 **Discussion**

244 Gene knockdown strategies targeting the primary disease-causing genes in polyQ disorders are
245 being extensively explored as potential therapies in these repeat expansion disorders.[3,5,6,36]
246 However, these strategies have been unsuccessful in polyQ clinical trials, failing to achieve the
247 desired therapeutic value and resulting in adverse safety profiles. To better understand these
248 challenges, we systematically profiled polyQ disorder genes using large-scale genomic databases,
249 highlighting their pleiotropic roles and identifying genomic features associated with increased
250 clinical trial failure. We demonstrated that polyQ disorder genes might be unfavorable targets since
251 they are associated with genomic features linked to an increased failure of drug targets in clinical
252 trials. Notably, *ATN1* (DRPLA), *ATXN1* (SCA1), *ATXN7* (SCA7), and *HTT* (HD) had particularly
253 deleterious risk profiles in our analyses.

254

255 Our analysis of human GWAS revealed that polyQ disorder genes display extensive pleiotropy
256 beyond their canonical roles in polyQ-related neurodegeneration, indicating potential on-target
257 effects of therapeutic knockdown.[37] For example, *HTT, ATXN1, ATXN7* and *CACNA1A* were
258 associated with neuropsychiatric disorders such as depression and schizophrenia. Additionally,
259 metabolic traits like type 1 and type 2 diabetes were linked to *ATXN2* and *ATXN7*, which is in line
260 with the fact that metabolic dysfunction is observed in several polyQ disorders.[38] *ATXN7* was
261 exclusively associated with pathogenic traits, including schizophrenia, diabetes, breast cancer, and
262 cataracts, highlighting the potential risks of targeting this gene. Our OMIM analyses also revealed
263 disease links to polyQ disorder genes outside canonical repeat-mediated disorders. For example,
264 *ATN1* and *HTT* have been linked to neurodevelopmental disorders, while *CACNA1A* has been
265 linked to channelopathies with diverse neurological manifestations. However, susceptibility to
266 Parkinson's disease and ALS are findings correlated with longer CAG repeats,[39] suggesting
267 possible benefits of mutation-specific knockdown related to these phenotypes.

268

269 PolyQ disorder genes were also linked to critical non-pathogenic phenotypes through GWAS. We
270 have previously reported that common *HTT* genetic variation is associated with important
271 neurobiological traits such as educational attainment and longevity,[15] which we replicate here. This
272 study showed that *ATXN1* is also linked to similar traits, including educational attainment and
273 cognitive abilities. When comparing the variant effects between related GWAS and GTEx

274 expression, alleles linked to lower *ATXN1* expression were also associated with lower educational
275 attainment and decreased cognitive abilities. These findings suggest that therapeutically targeting
276 the wild-type versions of these genes might inadvertently affect cognitive function, which should
277 be assessed during future clinical trials. However, a key limitation of this investigation is that we
278 cannot currently infer the direction of effect for all polyQ gene-GWAS trait pairs. Future studies
279 should systematically address this gap using high-throughput approaches (e.g., CRISPR screens
280 and functional assays).
281
282 In addition to individual gene-trait associations, network analyses revealed that specific trait
283 associations were shared between polyQ disorder genes, with *ATXN1*, *ATXN2,* and *HTT* displaying
284 the greatest number of shared interactions. For example, associations with metabolic traits (type 2
285 diabetes and cholesterol levels) were shared by *HTT, ATXN2* and *ATXN7* and susceptibility to
286 breast cancer was shared by *ATXN1* and *ATXN7.* These analyses indicate that specific polyQ
287 disorder genes may be involved in common biological pathways, as previously reported by
288 others.[40,41]
289
290 While half of the polyQ disorder genes were found to be potentially challenging to target
291 therapeutically with traditional approaches, RNA-targeted therapeutics provide a flexible solution
292 by enabling the targeting of any gene product. Nonetheless, our human genomic safety profile
293 analyses revealed that polyQ disorder genes consistently rank unfavorably for predictors of clinical
294 trial success,[19] which could provide some insight into the mechanisms driving gene-lowering trial
295 failures.[6,7,42] Specifically, *ATN1*, *ATXN1*, *ATXN7* and *HTT* ranked poorly across all safety-related
296 categories. These four genes are integral to transcriptional regulation and are, therefore, essential
297 for several critical physiological processes.[38] Therapeutic targeting of these genes, especially using
298 approaches that do not specifically target mutant versions of the related gene products, could
299 therefore disrupt several important processes, leading to adverse effects.
300
301 Most of the polyQ disorder genes (i.e., 80%) displayed broad expression patterns, with this profile
302 of expression associated with increased trial stoppage risks [reported odds ratio (OR): 1.29 95%;
303 confidence interval (CI): 1.19-1.39].[19] Additionally, half of the polyQ disorder genes are highly
304 constrained, with LoF intolerance (pLI>0.9) also associated with an increased risk of trial stoppage

305 (reported OR: 1.37; 95% CI: 1.27-1.48).[19] Further, all polyQ disorder genes, except for *THAP11*,
306 displayed more than ten high-confidence protein-protein interactions, indicating their central
307 physiological roles – an effect that can, in part, be attributed to the polyQ repeats themselves.[43]
308 Specifically, genes with between 11 and 20 partners (e.g., *CACNA1A*) are at an increased risk of
309 clinical trial stoppage (reported OR: 1.31; 95% CI: 1.19-1.44),[19] while those with >20 partners
310 (i.e., 80% of polyQ disorder genes) are at an even greater risk (reported OR: 1.38; 95% CI: 1.26-
311 1.52).[19]
312
313 Finally, when we compared the genomic safety predictions for polyQ disorder genes to those of
314 genetic modifiers of HD onset age, we demonstrated that the modifier genes showed more
315 favourable genomic safety scores. This provides additional support for prioritizing these modifiers
316 in future therapeutic research, especially in HD. Beyond their favourable drug target profiles, a
317 benefit of modifier-based therapeutic targeting is the potential for cross-repeat expansion disorder
318 relevance due to shared pathobiological mechanisms of these modifier genes (e.g., somatic repeat
319 instability).[34,44–47] This may also be useful for rarer repeat expansion disorders, where assembling
320 large cohorts of individuals required for a disorder-specific clinical trial may be difficult.[36] While
321 these modifier genes were identified by GWAS-based approaches, which typically yield
322 associations with modest to small effect sizes with traits being studied,[34] the effect of
323 therapeutically targeting these genes may have a larger impact on disease course. This is illustrated
324 by the fact that several GWAS have confirmed the involvement of variants in known drug targets
325 for numerous diseases.[48] For example, *DRD2*, the target for antipsychotics, has now been robustly
326 associated with schizophrenia risk through GWAS, although with a small effect (rs2514218:
327 GWAS odds ratio = 1.08)[49].
328
329 In conclusion, our findings indicate that the pleiotropic roles and adverse genomic profiles of
330 polyQ disorder genes might diminish their potential as therapeutic targets for their respective
331 conditions. However, it should be noted that these genes do have human genetic evidence for
332 involvement in their primary diseases, a feature associated with increased clinical trial success.[13]
333 Therefore, polyQ disorder genes should not be entirely abandoned as therapeutic targets, especially
334 with some encouraging results from recent early clinical trials.[50] Instead, increased focus should
335 be placed on assessing age-dependent implications of knockdown, as well as therapeutic

approaches aimed at mutant allele-specific perturbation and gene editing. In parallel, building on our findings that genetic modifiers of the age of HD onset have favourable genomic safety scores, future research should intensively investigate therapeutics that target repeat expansion disorder genetic modifiers. In line with this, the promising role of therapeutic targets that prevent somatic expansions has been further emphasized by recent reports that show that the impact of the expanded repeat only becomes evident late in the lifetime of vulnerable cell types.[51] Taken together, this study has revealed the importance of incorporating diverse human genomics data into drug development pipelines to develop safer and more effective treatments.

Declaration of interest

The authors declare no competing interests.

Acknowledgements

We thank the Natural Sciences and Engineering Research Council of Canada through the Canada Research Chairs Initiative and Discovery Grant Program, as well as the Health Sciences Centre Foundation Winnipeg, for the financial support provided to GEBW. BID was also supported through the Canadian Institutes of Health Research Canada Research Chairs Initiative. The Research Manitoba Graduate Studentship and Rady Faculty of Health Sciences Student Top-Up supported NKL.

Author contributions

NKL performed the formal analyses and wrote the manuscript. GEBW and BD designed the study and wrote and edited the manuscript. All authors agreed on the final version of the paper.

REFERENCES

1. Bunting, E.L., Hamilton, J., and Tabrizi, S.J. (2022). Polyglutamine diseases. Curr Opin Neurobiol *72*, 39–47. https://doi.org/10.1016/j.conb.2021.07.001.

2. Fan, H.-C., Ho, L.-I., Chi, C.-S., Chen, S.-J., Peng, G.-S., Chan, T.-M., Lin, S.-Z., and Harn, H.-J. (2014). Polyglutamine (PolyQ) diseases: genetics to treatments. Cell Transplant *23*, 441–458. https://doi.org/10.3727/096368914X678454.

3. Ghanekar, S.D., Kuo, S.-H., Staffetti, J.S., and Zesiewicz, T.A. (2022). Current and emerging treatment modalities for spinocerebellar ataxias. Expert Rev Neurother *22*, 101–114. https://doi.org/10.1080/14737175.2022.2029703.

4. Stoker, T.B., Mason, S.L., Greenland, J.C., Holden, S.T., Santini, H., and Barker, R.A. (2022). Huntington's disease: diagnosis and management. Pract Neurol *22*, 32–41. https://doi.org/10.1136/practneurol-2021-003074.

5. Perlman, S.L. (2020). Update on the Treatment of Ataxia: Medication and Emerging Therapies. Neurotherapeutics *17*, 1660–1664. https://doi.org/10.1007/s13311-020-00941-3.

6. Rook, M.E., and Southwell, A.L. (2022). Antisense Oligonucleotide Therapy: From Design to the Huntington Disease Clinic. BioDrugs *36*, 105–119. https://doi.org/10.1007/s40259-022-00519-9.

7. Esteves, S., Duarte-Silva, S., and Maciel, P. (2017). Discovery of Therapeutic Approaches for Polyglutamine Diseases: A Summary of Recent Efforts. Med Res Rev *37*, 860–906. https://doi.org/10.1002/med.21425.

8. Kwon, D. (2021). Failure of genetic therapies for Huntington's devastates community. Nature *593*, 180. https://doi.org/10.1038/d41586-021-01177-7.

9. McColgan, P., Thobhani, A., Boak, L., Schobel, S.A., Nicotra, A., Palermo, G., Trundell, D., Zhou, J., Schlegel, V., Sanwald Ducray, P., et al. (2023). Tominersen in Adults with Manifest Huntington's Disease. N Engl J Med *389*, 2203–2205. https://doi.org/10.1056/NEJMc2300400.

10. Plenge, R.M. (2016). Disciplined approach to drug discovery and early development. Sci Transl Med *8*, 349ps15. https://doi.org/10.1126/scitranslmed.aaf2608.

11. Plenge, R.M., Scolnick, E.M., and Altshuler, D. (2013). Validating therapeutic targets through human genetics. Nat Rev Drug Discov *12*, 581–594. https://doi.org/10.1038/nrd4051.

12. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., et al. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res *49*, D1311–D1320. https://doi.org/10.1093/nar/gkaa840.

13. Minikel, E.V., Painter, J.L., Dong, C.C., and Nelson, M.R. (2024). Refining the impact of genetic evidence on clinical success. Nature *629*, 624–629. https://doi.org/10.1038/s41586-024-07316-0.

14. Schultz, J.L., Saft, C., and Nopoulos, P.C. (2021). Association of CAG Repeat Length in the Huntington Gene With Cognitive Performance in Young Adults. Neurology *96*, e2407–e2413. https://doi.org/10.1212/WNL.0000000000011823.

15. Slike, A.N., and Wright, G.E.B. (2022). Common huntingtin-related genetic variation is associated with neurobiological and aging traits in humans. Cell Death Discov *8*, 311. https://doi.org/10.1038/s41420-022-01114-1.

16. Ghoussaini, M., Nelson, M.R., and Dunham, I. (2023). Future prospects for human genetics and genomics in drug discovery. Curr Opin Struct Biol *80*, 102568. https://doi.org/10.1016/j.sbi.2023.102568.

17. Rusina, P.V., Falaguera, M.J., Romero, J.M.R., McDonagh, E.M., Dunham, I., and Ochoa, D. (2023). Genetic support for FDA-approved drugs over the past decade. Nat Rev Drug Discov *22*, 864. https://doi.org/10.1038/d41573-023-00158-x.

18. Zhang, X., Yu, W., Li, Y., Wang, A., Cao, H., and Fu, Y. (2024). Drug development advances in human genetics-based targets. MedComm (2020) *5*, e481. https://doi.org/10.1002/mco2.481.

19. Razuvayevskaya, O., Lopez, I., Dunham, I., and Ochoa, D. (2024). Genetic factors associated with reasons for clinical trial stoppage. Nat Genet *56*, 1862–1867. https://doi.org/10.1038/s41588-024-01854-z.

20. Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J., et al. (2015). The support of human genetic evidence for approved drug indications. Nat Genet *47*, 856–860. https://doi.org/10.1038/ng.3314.

21. Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., et al. (2023). Ensembl 2023. Nucleic Acids Res *51*, D933–D941. https://doi.org/10.1093/nar/gkac958.

22. Mountjoy, E., Schmidt, E.M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., Fumis, L., Hayhurst, J., Buniello, A., Karim, M.A., et al. (2021). An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet *53*, 1527–1533. https://doi.org/10.1038/s41588-021-00945-5.

23. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res *43*, D789-798. https://doi.org/10.1093/nar/gku1205.

24. de Sainte Agathe, J.-M., Filser, M., Isidor, B., Besnard, T., Gueguen, P., Perrin, A., Van Goethem, C., Verebi, C., Masingue, M., Rendu, J., et al. (2023). SpliceAI-visual: a free

online tool to improve SpliceAI splicing variant interpretation. Hum Genomics *17*, 7. https://doi.org/10.1186/s40246-023-00451-1.

25. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557. https://doi.org/10.1093/bioinformatics/btv402.

26. Huyghe, J.R., Fransen, E., Hannula, S., Van Laer, L., Van Eyken, E., Mäki-Torkko, E., Aikio, P., Sorri, M., Huentelman, M.J., and Camp, G.V. (2011). A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. Eur J Hum Genet *19*, 347–352. https://doi.org/10.1038/ejhg.2010.179.

27. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

28. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.

29. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

30. Digre, A., and Lindskog, C. (2021). The Human Protein Atlas—Spatial localization of the human proteome in health and disease. Protein Science *30*, 218–233. https://doi.org/10.1002/pro.3987.

31. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Baker, J., Malangone, C., Lopez, I., Miranda, A., Cruz-Castillo, C., Fumis, L., et al. (2023). The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. Nucleic Acids Res *51*, D1353–D1359. https://doi.org/10.1093/nar/gkac1046.

32. Villaveces, J.M., Jiménez, R.C., Porras, P., Del-Toro, N., Duesbury, M., Dumousseau, M., Orchard, S., Choi, H., Ping, P., Zong, N.C., et al. (2015). Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study. Database (Oxford) *2015*, bau131. https://doi.org/10.1093/database/bau131.

33. Freshour, S.L., Kiwala, S., Cotto, K.C., Coffman, A.C., McMichael, J.F., Song, J.J., Griffith, M., Griffith, O.L., and Wagner, A.H. (2021). Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res *49*, D1144–D1151. https://doi.org/10.1093/nar/gkaa1084.

109 34. Namuli, K.L., Slike, A.N., Hollebeke, M.A., and Wright, G.E.B. (2025). Genomic
110 characterization of Huntington's disease genetic modifiers informs drug target tractability.
111 Brain Communications *7*, fcae418. https://doi.org/10.1093/braincomms/fcae418.

112 35. Razuvayevskaya, O., Lopez, I., Dunham, I., and Ochoa, D. (2024). Genetic factors
113 associated with reasons for clinical trial stoppage. Nat Genet, 1–6.
114 https://doi.org/10.1038/s41588-024-01854-z.

115 36. Coarelli, G., Coutelier, M., and Durr, A. (2023). Autosomal dominant cerebellar ataxias: new
116 genes and progress towards treatments. Lancet Neurol *22*, 735–749.
117 https://doi.org/10.1016/S1474-4422(23)00068-6.

118 37. Nguyen, P.A., Born, D.A., Deaton, A.M., Nioi, P., and Ward, L.D. (2019). Phenotypes
119 associated with genes encoding drug targets are predictive of clinical trial side effects. Nat
120 Commun *10*, 1579. https://doi.org/10.1038/s41467-019-09407-3.

121 38. Tandon, S., Aggarwal, P., and Sarkar, S. (2024). Polyglutamine disorders: Pathogenesis and
122 potential drug interventions. Life Sciences *344*, 122562.
123 https://doi.org/10.1016/j.lfs.2024.122562.

124 39. Rajan-Babu, I.-S., Dolzhenko, E., Eberle, M.A., and Friedman, J.M. (2024). Sequence
125 composition changes in short tandem repeats: heterogeneity, detection, mechanisms and
126 clinical implications. Nat Rev Genet *25*, 476–499. https://doi.org/10.1038/s41576-024-
127 00696-z.

128 40. Gui, H., Kwan, J.S., Sham, P.C., Cherny, S.S., and Li, M. (2017). Sharing of Genes and
129 Pathways Across Complex Phenotypes: A Multilevel Genome-Wide Analysis. Genetics *206*,
130 1601–1609. https://doi.org/10.1534/genetics.116.198150.

131 41. Peng, J., Zhou, Y., and Wang, K. (2021). Multiplex gene and phenotype network to
132 characterize shared genetic pathways of epilepsy and autism. Sci Rep *11*, 952.
133 https://doi.org/10.1038/s41598-020-78654-y.

134 42. Ward, J.M., Stoyas, C.A., Switonski, P.M., Ichou, F., Fan, W., Collins, B., Wall, C.E.,
135 Adanyeguh, I., Niu, C., Sopher, B.L., et al. (2019). Metabolic and Organelle Morphology
136 Defects in Mice and Human Patients Define Spinocerebellar Ataxia Type 7 as a
137 Mitochondrial Disease. Cell Rep *26*, 1189-1202.e6.
138 https://doi.org/10.1016/j.celrep.2019.01.028.

139 43. Klockgether, T., Mariotti, C., and Paulson, H.L. (2019). Spinocerebellar ataxia. Nat Rev Dis
140 Primers *5*, 24. https://doi.org/10.1038/s41572-019-0074-3.

141 44. Flower, M., Lomeikaite, V., Ciosi, M., Cumming, S., Morales, F., Lo, K., Hensman Moss,
142 D., Jones, L., Holmans, P., TRACK-HD Investigators, et al. (2019). MSH3 modifies somatic
143 instability and disease severity in Huntington's and myotonic dystrophy type 1. Brain *142*,
144 1876–1886. https://doi.org/10.1093/brain/awz115.

45. Laabs, B.-H., Klein, C., Pozojevic, J., Domingo, A., Brüggemann, N., Grütz, K., Rosales, R.L., Jamora, R.D., Saranza, G., Diesta, C.C.E., et al. (2021). Identifying genetic modifiers of age-associated penetrance in X-linked dystonia-parkinsonism. Nat Commun *12*, 3216. https://doi.org/10.1038/s41467-021-23491-4.

46. Bettencourt, C., Hensman-Moss, D., Flower, M., Wiethoff, S., Brice, A., Goizet, C., Stevanin, G., Koutsis, G., Karadima, G., Panas, M., et al. (2016). DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol *79*, 983–990. https://doi.org/10.1002/ana.24656.

47. Wright, G.E.B., Findlay Black, H., Collins, J.A., Gall-Duncan, T., Caron, N.S., Pearson, C.E., and Hayden, M.R. (2020). Interrupting sequence variants and age of onset in Huntington's disease: clinical implications and emerging therapies. Lancet Neurol *19*, 930–939. https://doi.org/10.1016/S1474-4422(20)30343-4.

48. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet *101*, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.

49. Lam, M., Chen, C.-Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C., et al. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet *51*, 1670–1678. https://doi.org/10.1038/s41588-019-0512-x.

50. Farag, M., Tabrizi, S.J., and Wild, E.J. (2024). Huntington's Disease Clinical Trials Update: September 2024. Journal of Huntington's Disease *13*, 409–418. https://doi.org/10.1177/18796397241293955.

51. Handsaker, R.E., Kashin, S., Reed, N.M., Tan, S., Lee, W.-S., McDonald, T.M., Morris, K., Kamitaki, N., Mullally, C.D., Morakabati, N.R., et al. (2025). Long somatic DNA-repeat expansion drives neurodegeneration in Huntington's disease. Cell *0*. https://doi.org/10.1016/j.cell.2024.11.038.

172 **Figures, titles and legends**

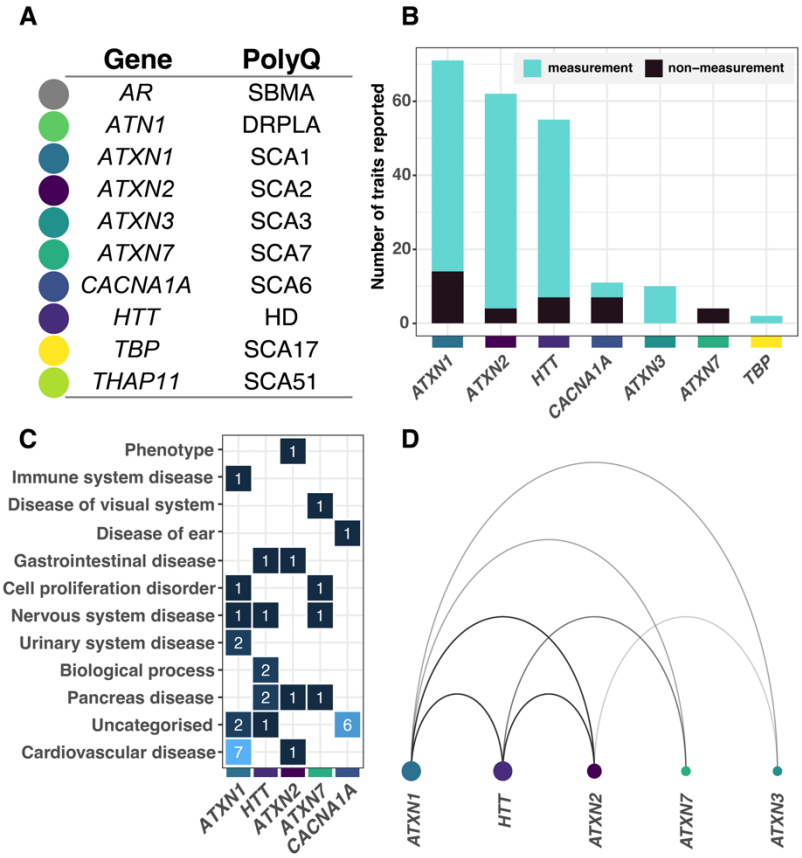


Figure 1. PolyQ disorder genes are pleiotropic and associated with several human traits and diseases in diverse phenotypic categories. (A) polyQ-associated genes and their corresponding polyQ disorders. Colors represent different genes used across panels and figures. **(B)** Human traits and disorders linked to polyQ disorder genes (i.e., L2G>0.5) obtained from an analysis of genome wide association studies in the Open Targets Genetics Database. The highest number of gene-trait pairs were identified for *ATXN1*, *ATXN2*, and *HTT* and most traits were classified as measurements. **(C)** Heatmap of non-measurement-related traits (L2G>0.5) for each polyQ disorder gene, depicting several associations for disease including those related to cardiovascular system (e.g., atrial fibrillation), pancreas (e.g., type 2 and type 2 diabetes) and nervous system. **(D)** Network graph illustrating shared polyQ disorder gene-trait associations. Larger circles represent the number of traits associated with a particular gene, while thicker arcs indicate more shared trait associations between genes. Notable shared traits include *HTT* and *ATXN1* with educational attainment and cognitive ability; *ATXN1* and *ATXN7* with breast cancer; and *ATXN1*, *ATXN2* and *ATXN3* with HDL cholesterol levels. This network highlights the pleiotropic effects of polyQ disorder genes across neurological, metabolic, and oncological traits. **Abbreviations:** L2G; locus-to-gene; *DRPLA*, Dentatorubral-Pallidoluysian Atrophy; *HD*, Huntington Disease; *SBMA*, Spinal and Bulbar Muscular Atrophy; *SCA*, Spinocerebellar Ataxia.

173

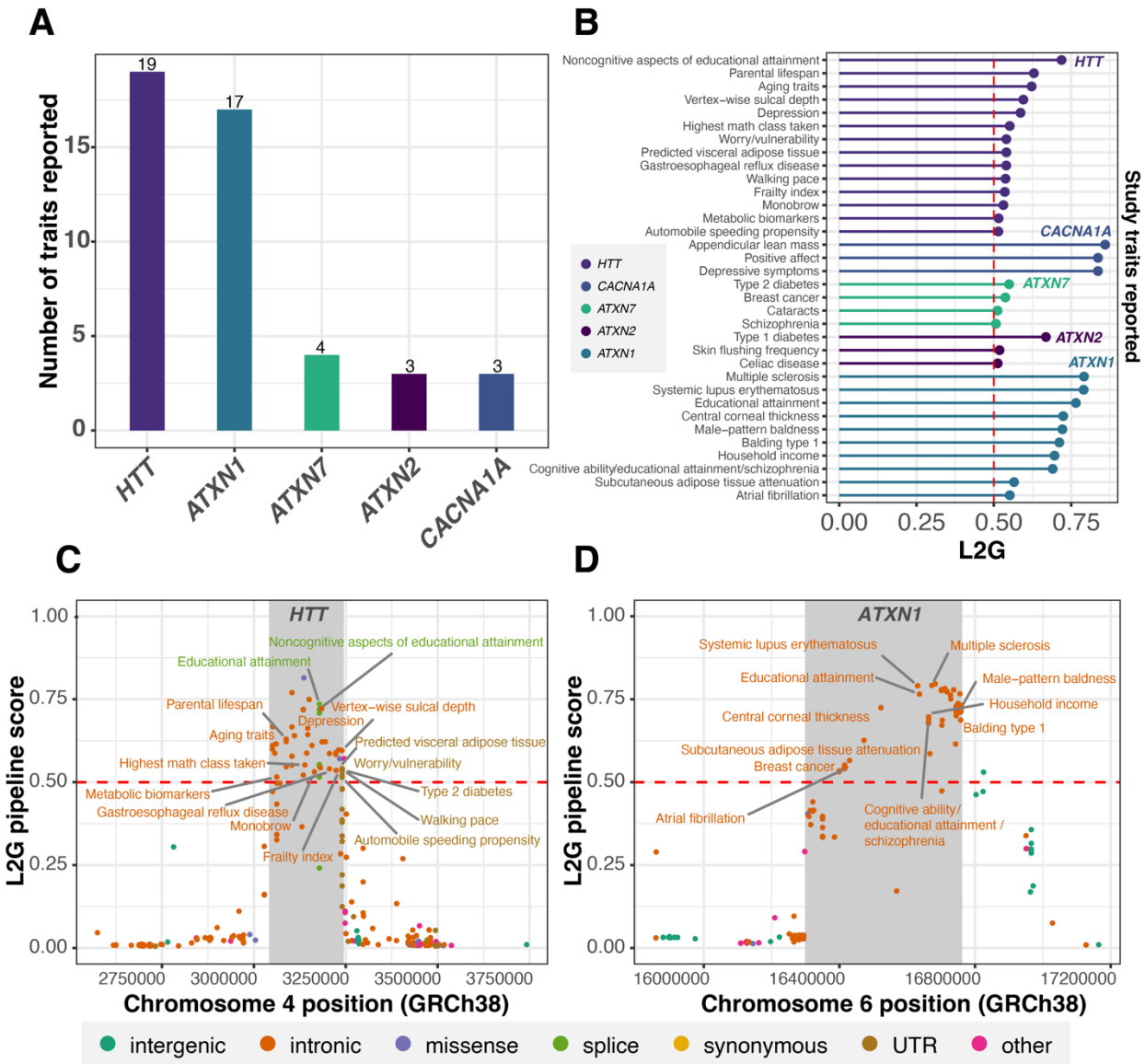


Figure 2. High-confidence and informative human gene-trait associations linked to the polyQ disorder genes. (A) A total of 46 traits linked to five polyQ disorder genes remained after applying a filtering threshold of L2G>0.5, while excluding count-based terms (e.g., blood cell counts) and studies lacking PubMed IDs. (**B**) L2G scores for individual high confidence trait associations (duplicate traits removed). PolyQ disorder genes exhibit associations with both pathogenic and non-pathogenic characteristics, across neurological, metabolic, and immune-related traits. **(C)** *HTT* and **(D)** *ATXN1* displayed the highest number of gene-trait pairs and are linked to neurological disorders and neurobiological phenotypes. In these panels, high confidence gene-trait associations are annotated and plotted by variant position and L2G score. Color indicates predicted variant consequence (*most severe*) of index variant for each trait association. Count based traits were plotted but not annotated. The grey-shaded segment represents polyQ disorder gene locus (GRCh38-genome build coding sequence start-stop position). **Abbreviations:** L2G; locus-to-gene, UTR; Untranslated region, GRCh38; Genome Reference Consortium Human Reference 38.

174

175

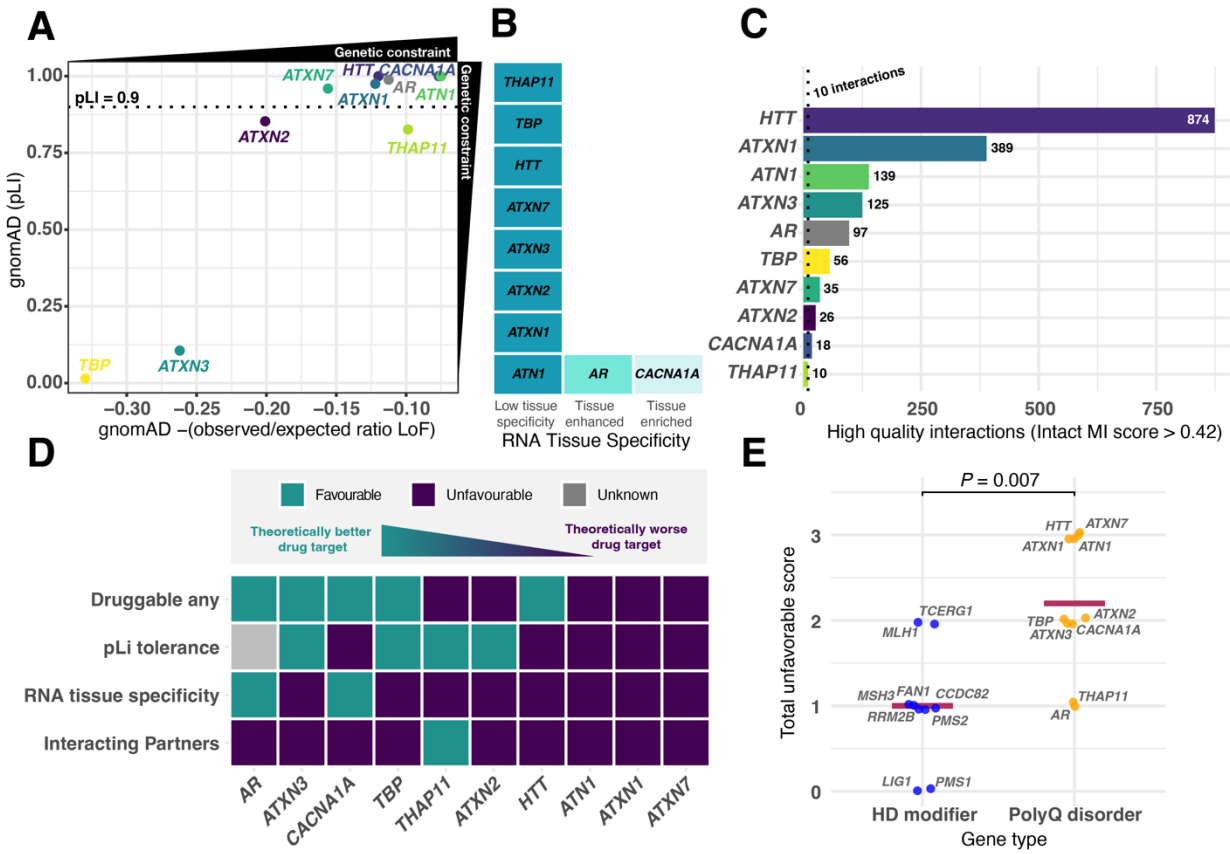


Figure 3. Genetic features associated with clinical trial success suggest that polyQ disorder genes may not be favorable therapeutic targets. **(A)** Genetic constraint features, gnomAD probability of loss-of-function intolerance (pLI) and negative observed vs expected loss-of-function variant ratio, indicate high levels of the constraint of the polyQ disorder genes. pLI allows non-constrained genes (pLI<0.1) to be differentiated from highly constrained ones (pLI>0.9). The observed/expected ratio is a quantitative measure of how tolerant the genes are to certain classes of variations. **(B)** Tissue specificity of the polyQ disorder genes, with *AR* and *CACNA1A* displaying tissue-enhanced and tissue-enriched, respectively, while others show low specificity. **(C)** PolyQ disorder genes interact with several other proteins and only *THAP11* displays less than 11 interacting partners (associated with decreased levels of clinical trial stoppage). **(D)** Aggregated summary of human genomic features used to rank the theoretical druggability and genomic safety of polyQ disorder genes. Genes ranked more unfavorable related to theoretical druggability and genomic safety, such as *HTT*, *ATN1*, *ATXN1* and *ATXN7*, are on the right of the plot. Individual favorable metrics are represented by the teal green tiles, while the deep purple tiles represent unfavorable features. Favorable/unfavorable classification thresholds are based on metrics identified by original study where features associated with clinical trial stoppage were identified (Razuvayevskaya *et al.* 2024 *Nat Gen*). **(E)** PolyQ disorder genes (yellow) display significantly more (*P*=0.007) unfavorable genomic safety scores based on constraint, tissue specificity and interactions (maximum score=3, increased risk of clinical trial stoppage) than Huntington disease modifier genes (blue). Mean scores are indicated by red line (mean polyQ=2.2, mean HD modifier genes=1.0).

176

1 # Tables

2

Table 1 High-confidence GWAS associations linked to polyQ disorder genes are associated with pathogenic and non-pathogenic human traits. L2G prioritization highlights traits linked to unique independent signals identified as haplotypes associated with polyQ disorder genes extending beyond the primary traits directly caused by the pathogenic CAG repeat expansions in these the genes.

Signal (Tag variant)[a]	Reported trait (PMID)	Variant ID	Annotation[b]	Effect AF EUR[c]	GWAS index *P*-value	GWAS index effect (95% CI)	L2G	CADD score
ATXN1: haplotype one (rs719316)	Multiple sclerosis (31604244)	rs719316	intronic	T: 0.529	2.00×10^{-13}	1.07 (1.05-1.09)[d]	0.79	5.92
ATXN1: haplotype two (rs17603856)	Systemic lupus erythematosus (27399966)	rs17603856	intronic	G: 0.346	3.00×10^{-12}	0.88 (0.85-0.91)[d]	0.79	10.94
ATXN1: haplotype three (rs909788)	Cognitive ability, years of educational attainment or schizophrenia (31374203)	rs6459480	intronic	A: 0.509	2.00×10^{-10}	NA	0.69	1.82
	Household income (31844048)	rs6459480	intronic	A: 0.509	5.00×10^{-10}	-0.01 (-0.02--0.01)	0.70	1.82
	Educational attainment (30038396)	rs9297016	intronic	G: 0.576	1.00×10^{-09}	-0.01 (-0.01--0.006)	0.68	2.57
	Educational attainment: years of education (27225129)	rs7772172	intronic	G: 0.576	1.00×10^{-08}	-0.01 (-0.02--0.01)	0.68	5.02

	Educational attainment (years of education) (30038396)	rs909788	intronic	C: 0.545	$4.00x10^{-08}$	-0.01 (-0.011--0.005)	0.77	5.75
ATXN1: haplotype four (rs6459472)	Central corneal thickness (32528159)	rs6459472	intronic	G: 0.573	$3.00x10^{-09}$	NA	0.72	2.97
ATXN1: haplotype five (rs9367926)	Male-pattern baldness (30573740)	rs6915310	intronic	T: 0.182	$5.00x10^{-20}$	0.04 (0.03 - 0.05)	0.72	4.27
	Balding type 1 (30595370)	rs9367926	intronic	G: 0.180	$6.00x10^{-12}$	NA	0.71	6.59
ATXN1: haplotype six (rs2237199)	Subcutaneous adipose tissue attenuation (27918534)	rs2237199	intronic	A: 0.112	$1.00x10^{-08}$	5.67 (3.73 - 7.61)	0.57	3.49
ATXN1: haplotype seven (rs7770062)	Atrial fibrillation (30061737)	rs73366713	intronic	A: 0.131	$1.53x10^{-25}$	0.90 (0.88-0.92)[d]	0.55	2.84
	Atrial fibrillation (29892015)	rs7770062	intronic	A: 0.131	$9.00x10^{-21}$	0.90 (0.88-0.92)[d]	0.54	2.33
ATXN1: haplotype eight (rs3819405)	Breast cancer (29059683)	rs3819405	intronic	T: 0.340	$1.65x10^{-08}$	0.96 (0.95-0.97)[d]	0.53	12.10
ATXN2: haplotype one (rs653178)	Celiac disease (20190752)	rs653178	intronic	T: 0.534	$6.00x10^{-14}$	0.83 (0.78-0.87)[d]	0.51	1.59
ATXN2: haplotype two (rs848130)	Type 1 diabetes (34012112)	rs848130	intronic	A: 0.192	$1.60x10^{-14}$	0.85 (0.82-0,89)[d]	0.67	2.02
ATXN7: haplotype one (rs832190)	Schizophrenia (31740837)	rs832190	intronic	T: 0.603	$4.00x10^{-08}$	NA	0.51	0.58

ATXN7: haplotype two (rs3821902)	Breast cancer (29059683)	rs3821902	intronic	G: 0.150	$2.99x10^{-12}$	1.06 (1.05-1.08)[d]	0.54	0.81
ATXN7: haplotype three (rs13434089)	Type 2 diabetes (32541925)	rs13434089	intronic	C: 0.141	$4.00x10^{-31}$	-0.06 (-0.07--0.05)	0.55	4.59
CACNA1A: haplotype one (rs5021328)	Appendicular lean mass (33097823)	rs5021328	intronic	C: 0.644	$6.60x10^{-09}$	-0.01 (-0.02--0.01)	0.86	0.63
CACNA1A: haplotype two (rs16003)	Positive affect (30643256)	rs16003	intronic	T: 0.463	$3.00x10^{-08}$	0.01 (0.00 - 0.01)	0.84	8.26
	Depressive symptoms (30643256)	rs16003	intronic	T: 0.463	$3.00x10^{-08}$	0.01 (0.00 - 0.01)	0.84	8.26
HTT: haplotype one (rs363096)	Noncognitive aspects of educational attainment (33414549)	rs363096	splice	C: 0.591	$1.00x10^{-12}$	0.05 (0.04 - 0.06)	0.72	6.29
	Educational attainment: years of education (30595370)	rs363096	splice	C: 0.591	$1.00x10^{-10}$	NA	0.74	6.29
HTT: haplotype two (rs61348208)	Depression (30718901)	rs7685686	intronic	G: 0.433	$6.00x10^{-15}$	0.983 (0.979-0.987)[d]	0.59	6.39
	Frailty index (34431594)	rs82334	intronic	C: 0.323	$3.10x10^{-10}$	-0.02 (-0.03--0.02)	0.54	4.13
	Parental lifespan (30642433)	rs61348208	intronic	T: 0.389	$6.00x10^{-09}$	0.23 (0.15 - 0.31)	0.63	5.46

	Gastroesophageal reflux disease (34187846)	rs7685686	intronic	G: 0.433	$1.10x10^{-08}$	0.97 (0.96-0.98)[d]	0.54	6.39
	Aging traits: health span, parental lifespan or longevity (32678081)	rs61348208	intronic	T: 0.389	$3.00x10^{-08}$	NA	0.62	5.46
	Vertex-wise sulcal depth (34910505)	rs2071703	intronic	C: 0.325	$3.00x10^{-08}$	5.57 (3.60 - 7.54)	0.6	2.52
HTT: haplotype three (rs113928896)	Highest math class taken (30038396)	rs113928896	intronic	T: 0.143	$2.00x10^{-09}$	0.02 (0.01 - 0.02)	0.55	0.32
HTT: haplotype four (rs362307)	Automobile speeding propensity (30643258)	rs362307	UTR	T: 0.063	$6.00x10^{-14}$	-0.03 (-0.04--0.02)	0.51	3.23
	Worry/vulnerability: special factor of neuroticism (30867560)	rs362307	UTR	T: 0.063	$1.00x10^{-09}$	0.01 (0.005 - 0.011)	0.54	3.23
	Walking pace (33128006)	rs362307	UTR	T: 0.063	$1.00x10^{-09}$	-0.01 (-0.02--0.01)	0.54	3.23
	Type 2 diabetes (30297969)	rs362307	UTR	T: 0.063	$1.00x10^{-09}$	1.08 (1.05-1.11)[d]	0.54	3.23
	Predicted visceral adipose tissue (31501611)	rs362307	UTR	T: 0.063	$2.00x10^{-09}$	0.03 (0.02 - 0.04)	0.54	3.23
HTT: haplotype five (rs6828882)	Monobrow (27182965)	rs6828882	intronic	G: 0.101	$1.00x10^{-22}$	NA	0.53	8.54

HTT: haplotype six (rs2798297)	Metabolic biomarkers (33980691)	rs2798297	intronic	A: 0.353	$1.00x10^{-24}$	NA	0.52	2.08

Haplotype, Group of variants in high LD with one another. Tag variant, Variant proxy for all variants in high LD at a haplotype. [a]LDlink SNPclip pruned signals using an r2 of 0.5 in the 1000 Genomes Project European superpopulations. [b]Ensembl VEP 'most severe' annotations. [c]1000 Genomes Project European superpopulation effect allele frequency. [d]Odds ratio (all other effects are beta estimates). **Abbreviations:** AF, Allele frequency; CADD, Combined Annotation Dependent Depletion; EUR, 1000 Genomes European super-population; L2G, Open Targets locus-to-gene scoring metric; NA, data Not available; PMID, PubMed Identifier; UTR, Untranslated region.

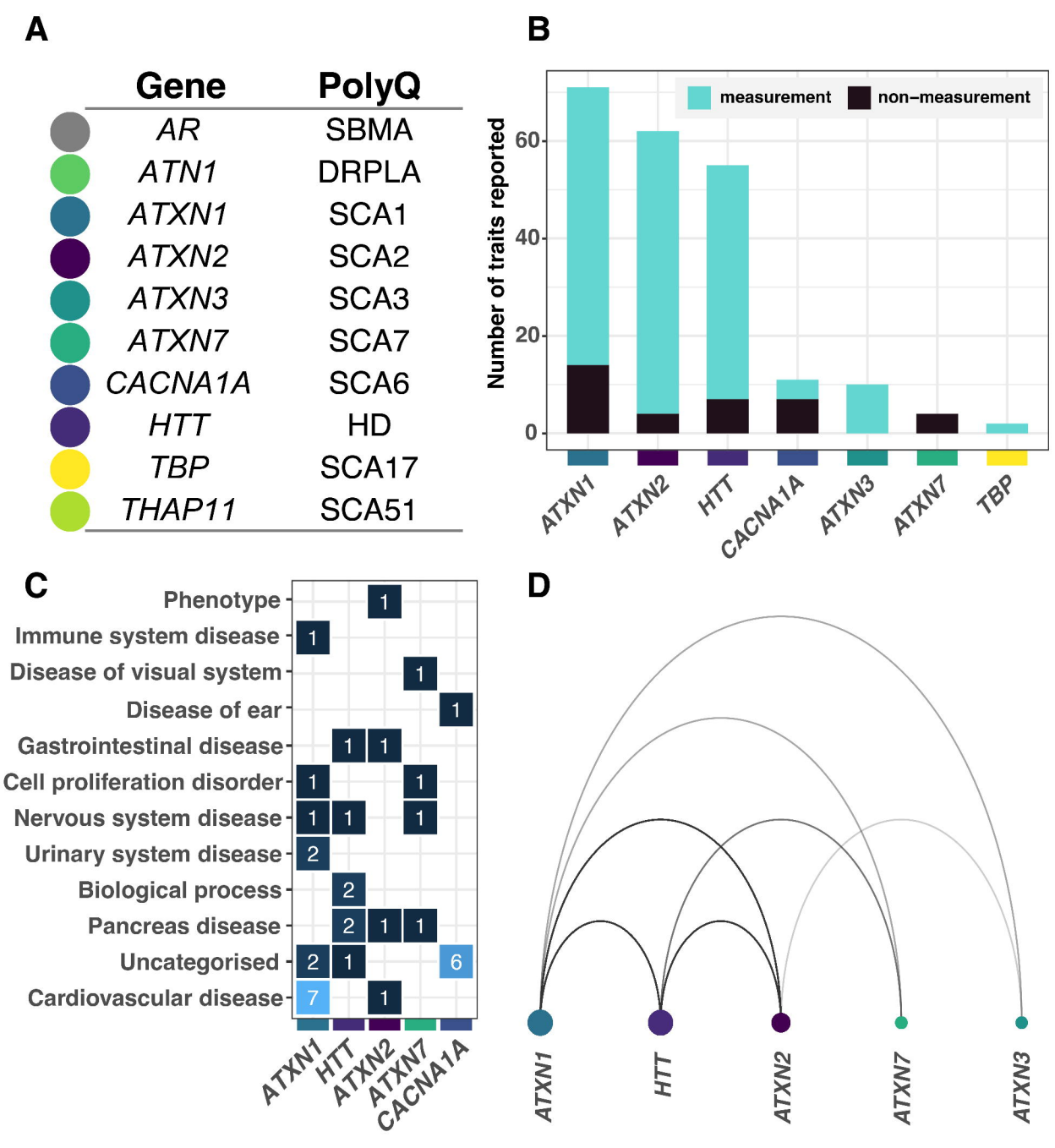
3

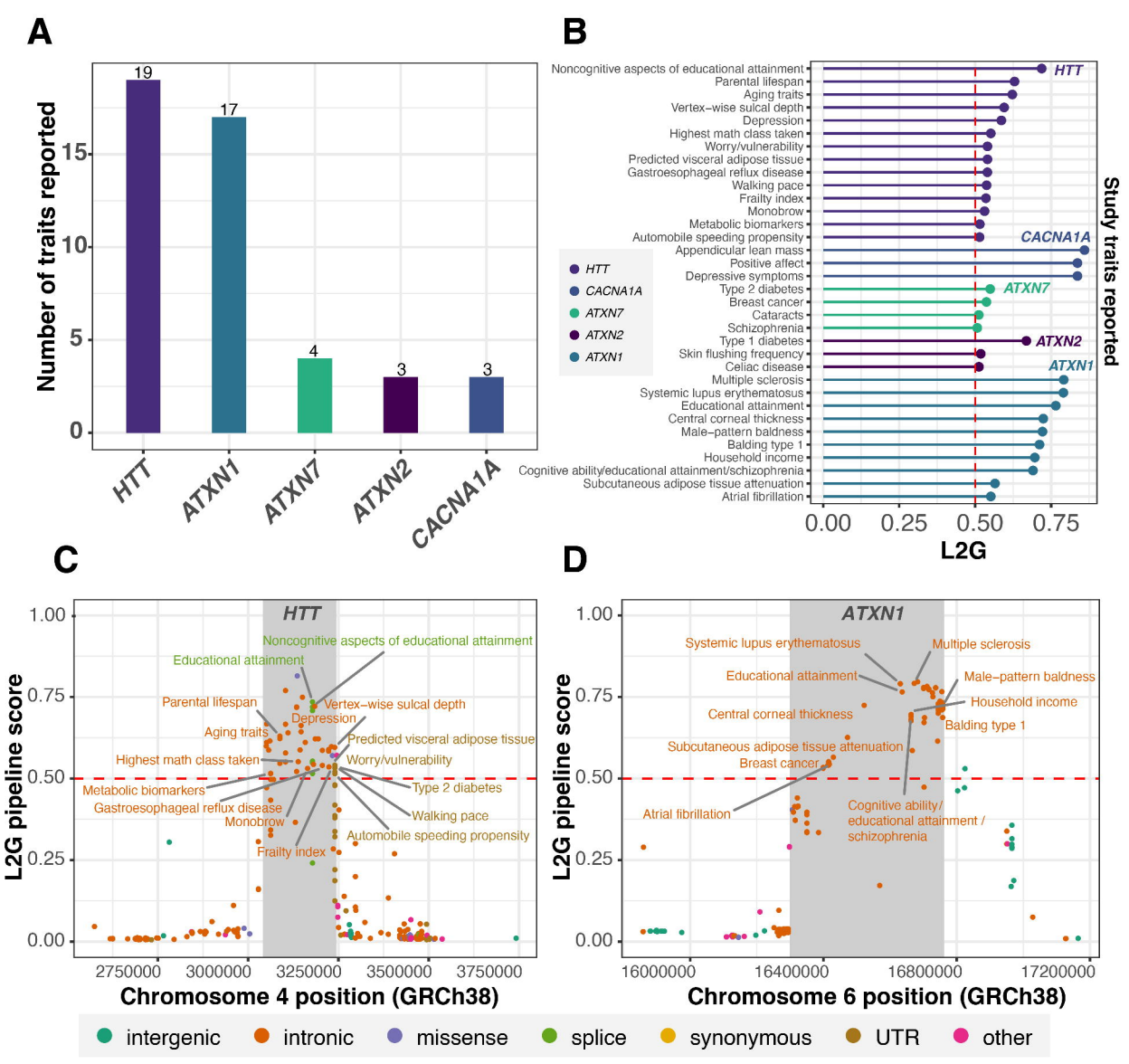
Table 2. Non-canonical disease associations of polyQ disorder genes potentially informs off target effects of therapeutic knockdown of pathogenic disease genes. Six polyQ disorder genes have OMIM phenotype entries outside of their main polyQ disorder. These phenotypes range from endocrine and metabolic conditions (e.g., androgen insensitivity), to neurological and neurodevelopmental disorders (e.g., developmental and epileptic encephalopathy, Lopes-Maciel-Rodan syndrome), and susceptibility to complex diseases.

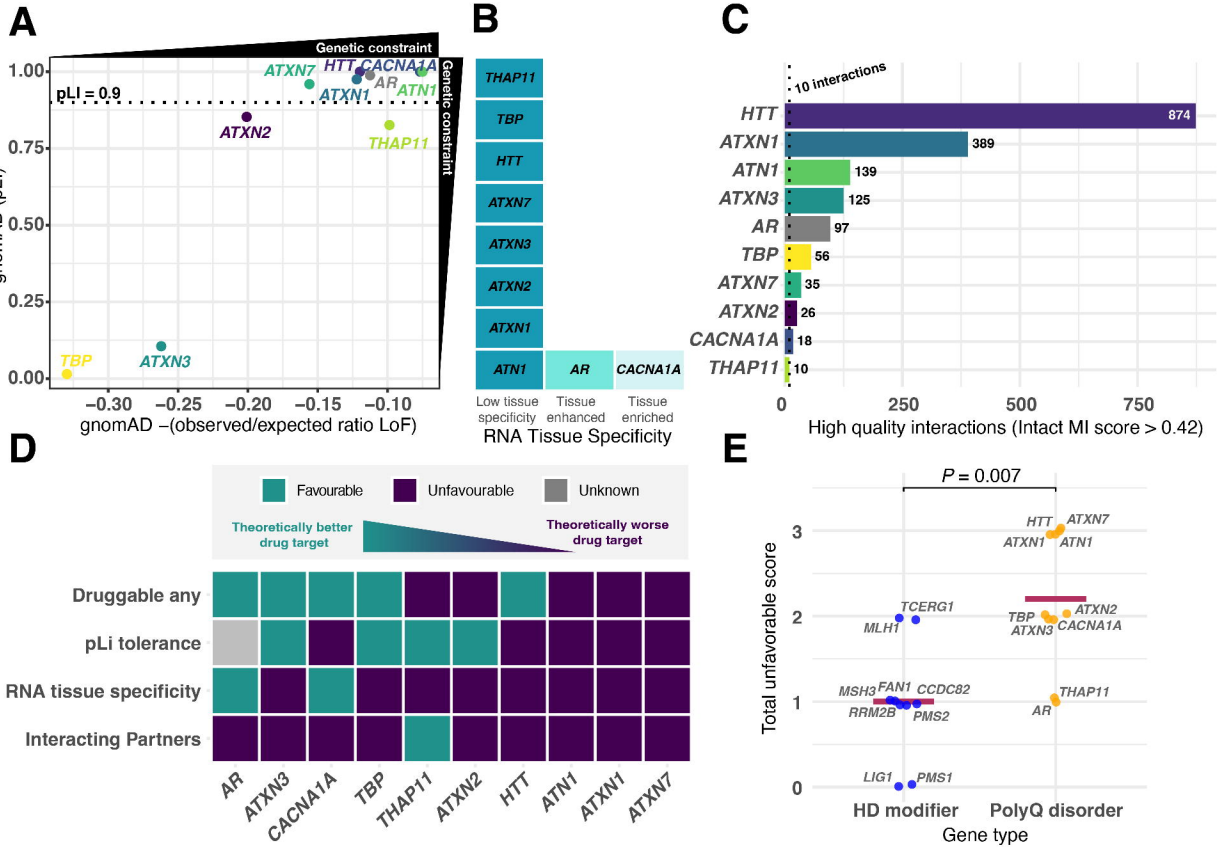
Gene	OMIM phenotypes
AR	• Androgen insensitivity, partial, with or without breast cancer (MIM: 312300, X-linked recessive) • Androgen insensitivity (MIM: 300068, X-linked recessive) • Hypospadias 1, X-linked (MIM:300633, X-linked recessive) • *Susceptibility to: Prostate cancer (MIM: 176807, autosomal dominant, somatic mutation)*
ATN1	• Congenital hypotonia, epilepsy, developmental delay, and digital anomalies (MIM: 618494, autosomal dominant)
ATXN2	• *Susceptibility to: Amyotrophic lateral sclerosis (MIM: 183090, autosomal dominant)** • *Susceptibility to: Parkinson disease, late-onset (MIM: 168600, autosomal dominant, multifactorial)**
CACNA1A	• Developmental and epileptic encephalopathy 42 (MIM: 617106, autosomal dominant) • Episodic ataxia, type 2 (MIM: 108500, autosomal dominant)* • Migraine, familial hemiplegic, 1, with progressive cerebellar ataxia (MIM: 141500, autosomal dominant)
HTT	• Lopes-Maciel-Rodan syndrome (MIM: 617435, autosomal recessive)
TBP	• *Susceptibility to: Parkinson disease (MIM: 168600, autosomal dominant, multifactorial)**

*CAG repeat mediated disorder/susceptibility. **Abbreviations: OMIM**, Online Mendelian Inheritance in Man

4


A
Gene
PolyQ
AR
SBMA
ATN1
DRPLA
ATXN1
SCA1
ATXN2
SCA2
ATXN3
SCA3
ATXN7
SCA7
CACNA1A
SCA6
HTT
HD
TBP
SCA17
THAP11
SCA51
B
measurement
non−measurement
Number of traits reported
0
20
40
60
ATXN1
ATXN2
HTT
CACNA1A
ATXN3
ATXN7
TBP
C
Phenotype
Immune system disease
Disease of visual system
Disease of ear
Gastrointestinal disease
Cell proliferation disorder
Nervous system disease
Urinary system disease
Biological process
Pancreas disease
Uncategorised
Cardiovascular disease
ATXN1
HTT
ATXN2
ATXN7
CACNA1A
D
ATXN1
HTT
ATXN2
ATXN7
ATXN3


A
Number of traits reported
19
17
4
3
3
HTT
ATXN1
ATXN7
ATXN2
CACNA1A
B
Noncognitive aspects of educational attainment
Parental lifespan
Aging traits
Vertex−wise sulcal depth
Depression
Highest math class taken
Worry/vulnerability
Predicted visceral adipose tissue
Gastroesophageal reflux disease
Walking pace
Frailty index
Monobrow
Metabolic biomarkers
Automobile speeding propensity
Appendicular lean mass
Positive affect
Depressive symptoms
Type 2 diabetes
Breast cancer
Cataracts
Schizophrenia
Type 1 diabetes
Skin flushing frequency
Celiac disease
Multiple sclerosis
Systemic lupus erythematosus
Educational attainment
Central corneal thickness
Male−pattern baldness
Balding type 1
Household income
Cognitive ability/educational attainment/schizophrenia
Subcutaneous adipose tissue attenuation
Atrial fibrillation
HTT
CACNA1A
ATXN7
ATXN2
ATXN1
Study traits reported
L2G
C
HTT
L2G pipeline score
Chromosome 4 position (GRCh38)
D
ATXN1
L2G pipeline score
Chromosome 6 position (GRCh38)
intergenic
intronic
missense
splice
synonymous
UTR
other


A
Genetic constraint
pLI = 0.9
ATXN7
HTT
CACNA1A
AR
ATN1
ATXN1
ATXN2
THAP11
TBP
ATXN3
gnomAD (pLI)
1.00
0.75
0.50
0.25
0.00
−0.30
−0.25
−0.20
−0.15
−0.10
gnomAD −(observed/expected ratio LoF)
B
THAP11
TBP
HTT
ATXN7
ATXN3
ATXN2
ATXN1
ATN1
AR
CACNA1A
Low tissue specificity
Tissue enhanced
Tissue enriched
RNA Tissue Specificity
C
10 interactions
HTT 874
ATXN1 389
ATN1 139
ATXN3 125
AR 97
TBP 56
ATXN7 35
ATXN2 26
CACNA1A 18
THAP11 10
0
250
500
750
High quality interactions (Intact MI score > 0.42)
D
Favourable
Unfavourable
Unknown
Theoretically better drug target
Theoretically worse drug target
Druggable any
pLi tolerance
RNA tissue specificity
Interacting Partners
AR
ATXN3
CACNA1A
TBP
THAP11
ATXN2
HTT
ATN1
ATXN1
ATXN7
E
P = 0.007
Total unfavorable score
3
2
1
0
HTT
ATXN7
ATXN1
ATN1
TCERG1
MLH1
ATXN2
TBP
ATXN3
CACNA1A
MSH3
FAN1
CCDC82
RRM2B
PMS2
THAP11
AR
LIG1
PMS1
HD modifier
PolyQ disorder
Gene type