

Figure S1 - Mutability adjusted proportion of singletons (MAPS) across region upstream of the acceptor site (including polypyrimidine tract region) split by changes from a pyrimidine to a purine (PyPu) vs all other changes in DDD unaffected parents (A) and ExAC data (B). Deficit of variants in genes with high pLI in unaffected parents recruited as part of DDD study across region upstream of the acceptor site (including polypyrimidine tract region) split by changes from a pyrimidine to a purine (PyPu) vs all other changes (C).

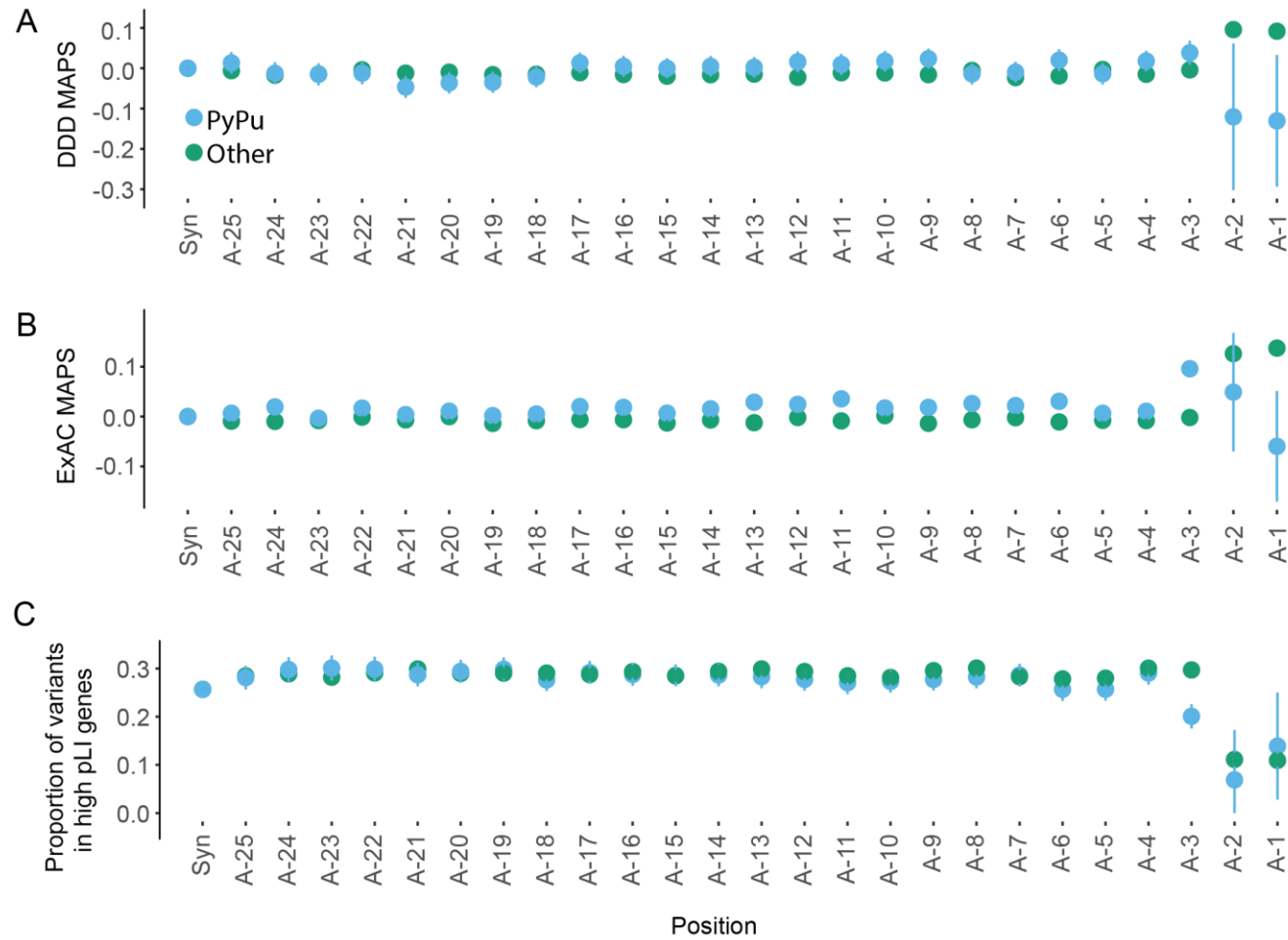


Figure S2 – Positive predictive values (PPVs) for classes of splice and non-splice mutations with MAPS (DDD unaffected parents and ExAC) and proportion of unaffected DDD parental variants in high pLI genes (pLI > 0.9)

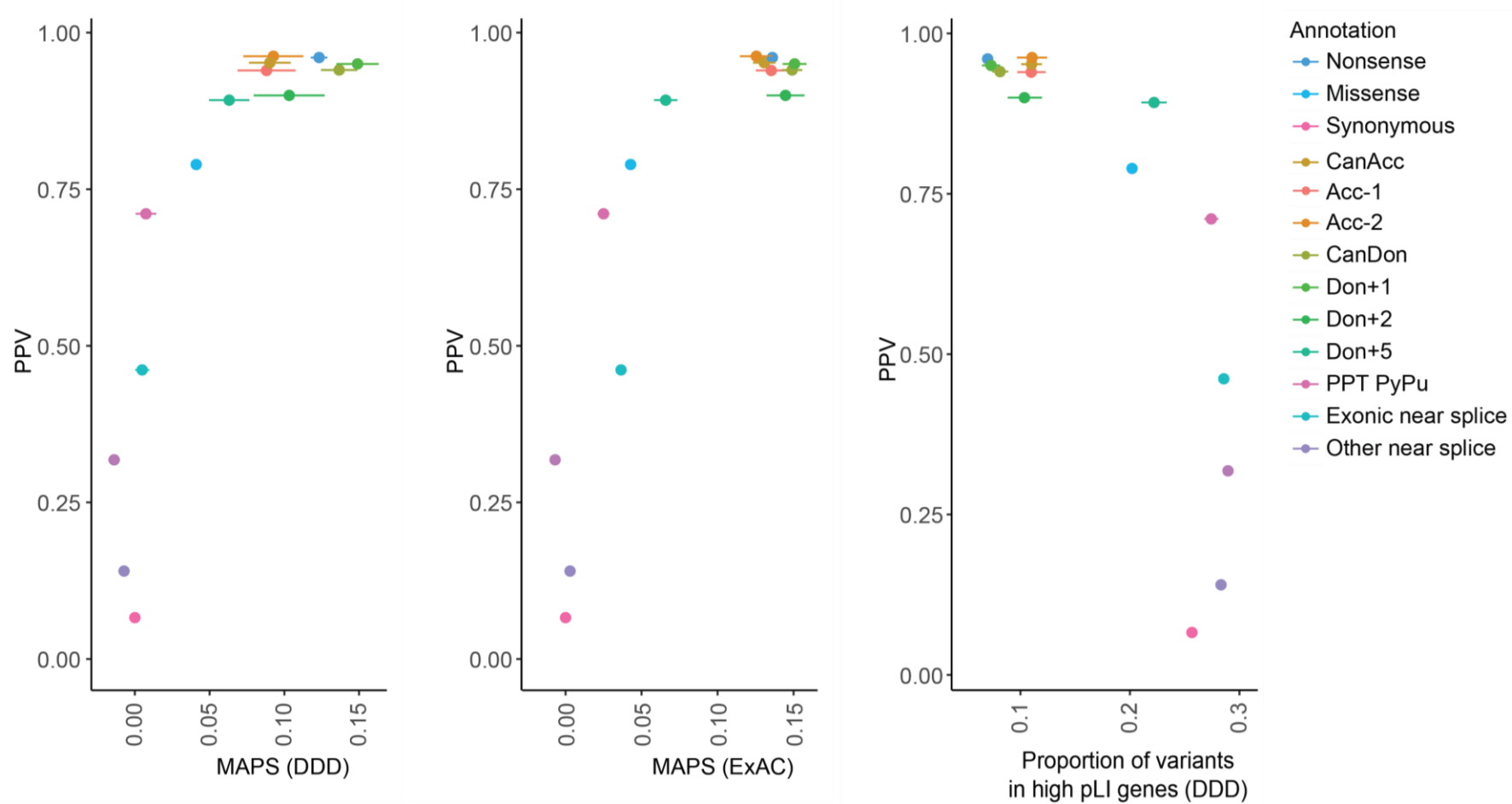
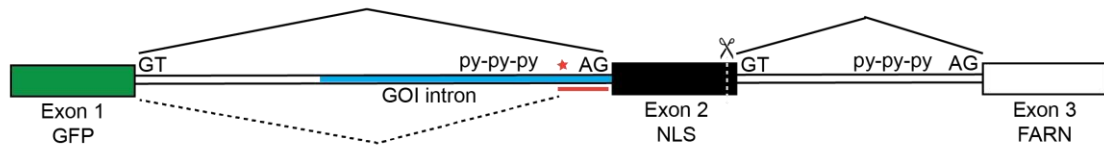


Figure S3 – Outcomes of minigene assays for splicing validations

A. Schematic diagram of splicing construct for polypyrimidine tract variants



Validation outcomes for polypyrimidine tract variants found to affect splicing

DNM1

Construct sequences (intron, exon)

Ref CCTGTCCACCTCTCCCCGGGTGCAGGACGGGTCCGGACTCAGATCTGG

Alt CCTGTCCACCTCTCCCCAGGTGCAGGACGGGTCCGGACTCAGATCTGG

cDNA sequence traces

Ref GGCATGGACGAGCTGTACAAGACTCAGGACGGGTCCGGACTCAGATCTGG

Alt GGCATGGACGAGCTGTACAAGACTCAGGTGCAGGACGGGTCCGGACTCAGATCTGG

Predicted protein outcome

... Exon 9 Exon 10 ...
ATT AGG ACG GGC CTC TTC / 1377 N / TAA
Ile Arg Thr Gly Leu Phe / 459 aa / . (864aa)

... Exon 9 Intron 9/10 Exon 10 ...
ATT AGG TGC AGG ACG GGC CTC TTC / 1377 N / TAA
Ile Arg Cys Arg Thr Gly Leu Phe / 459 aa / . (866aa)

GLI3

Construct sequences (intron, exon)

Ref TCTGTTCCCTGCCCCACCTCTTCTTTAGGAACAGTCCGGACTCAGATC

Alt TCTGTTCCCTGCCCCACCTCTTCTTTAGGAACAGTCCGGACTCAGATC

cDNA sequence traces

Ref GGCATGGACGAGCTGTACAAGACTCAGGAACAGTCCGGACTCAGATCTGG

Alt GGCATGGACGAGCTGTACAAGACTCAGCCTCTTCTTTTCAGGAACAGTCCGGACTCAGATCTGG

Predicted protein outcome

... Exon 9 Exon 10 ...
CAG CAG GAA CAG CCC GAA GGA / 3369 N / TAG
Gln Gln Glu Gln Pro Glu Gly / 1123 aa / . (1580aa)

... Exon 9 Intron 9/10 Exon 10 ...
CAG CAG CCT CTT CTT TCA GGA ACA / 42 N / TGA
Gln Gln Pro Leu Leu Ser Gly Thr / 14 aa / . (472aa)

CHD7

Construct sequences (intron, exon)

Ref TTCTGTGCACGGATGGGCACGGCACAGGCTATGTCCGGACTCAGATCTGGA

Alt TTCTGTGCACGGATGGGCACAGCACAGGCTATGTCCGGACTCAGATCTGGA

cDNA sequence traces

Ref GGCATGGACGAGCTGTACAAGACTCAGGCTATGTCCGGACTCAGATCTGGA

Alt GGCATGGACGAGCTGTACAAGACTCAGCACAGGCTATGTCCGGACTCAGATCTGGA

Predicted protein outcome

... Exon 25 Exon 26 ...
AAA CAT GGC TAT GAG AAG TAC / 3573 N / TAA
Lys His Gly Tyr Glu Lys Tyr / 1191 aa / . (2997aa)

... Exon 25 Intron 25/26 Exon 26 ...
AAA CAT GCA CAG GCT ATG AGA AGT / 69 N / TAG
Lys His Ala Gln Ala Met Arg Ser / 23 aa / . (1830aa)

Key

- █ Exon 1
- █ Intron 1
- █ Retained intron 1
- █ Exon 2
- N Ref nucleotide
- N Alt nucleotide
- Acceptor/donor
- ✂ Self cleaving peptide
- GOI codons

Validation outcomes for polypyrimidine tract variants found to affect splicing (continued)

CREBBP

Construct sequences (intron, exon)

Predicted protein outcome

Ref TTACTGAAGTCAGTGCCTTCGGTTTTTTCACAGCTTTCCCTCCGGACTCAGAT

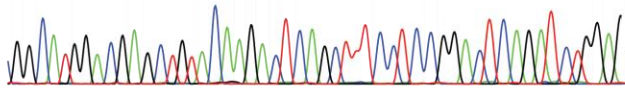
... Exon 14 ACA CCG CTT TCC CAG GCA GCA GCC / 4428N / TAG
 Thr Pro Leu Ser Gln Ala Ala Ala / 1476aa / . (2442aa)

Alt TTACTGAAGTCAGTGCCTTCAGTTTTTTCACAGCTTTCCCTCCGGACTCAGAT

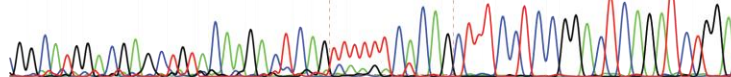
... Exon 14 ACA CCG TIT TIT CAC AG C TTT CCC / 102 N / TGA
 Thr Pro Phe Phe His Ser Phe Pro / 34 aa / . (1000aa)

cDNA sequence traces

Ref GG CATGGACGAGCTGTACAAGACTCAGCTTT CCT CCGGACTCAGAT CTGGA



Alt GG CATGGACGAGCTGTACAAGACTCAGTTTTTTCACAGCTTT CCT CCGGACTCAGAT CTGGA



MEF2C

Construct sequences (intron, exon)

Predicted protein outcome

Ref CAGTAATGCTTTTTATTATTTTAAAGAATCAATCCGGACTCAGAT

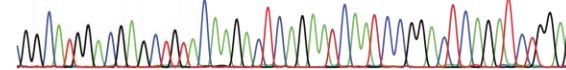
... Exon 9 TCA GTG AAT CAA AGG ATA AAT AAC / 471N / TGA
 Ser Val Asn Gln Arg Ile Asn Asn / 157aa / . (483aa)

Alt CAGTAATGCTTTTTATTAGTTTAAAGAATCAATCCGGACTCAGAT

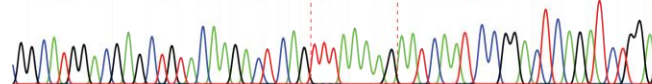
... Exon 9 TCA GTG TTT AAA AG A ATC AAA GGA TAA
 Ser Val Phe Lys Arg Ile Lys Gly . (294aa)

cDNA sequence traces

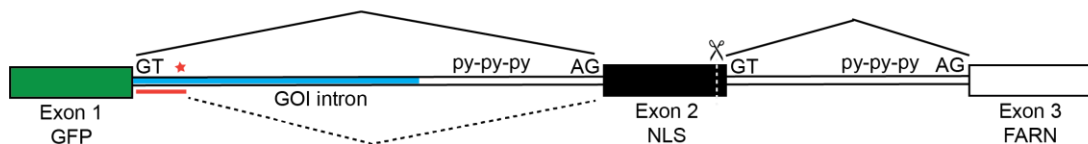
Ref GG CATGGACGAGCTGTACAAGACTCAGAAATCAATCCGGACTCAGAT CTGGA



Alt GG CATGGACGAGCTGTACAAGACTCAGTTTAAA GAATCAATCCGGACTCAGAT CTGGA



B. Schematic diagram of splicing construct for don+5 variant



Validation outcomes for don+5 variant found to affect splicing

MBD5

Construct sequences (exon, intron)

Predicted protein outcome

Ref GCATGGACGAGCTGTACAAGGAACAAGTATGTAATATGGTGAAGGTTTCAG

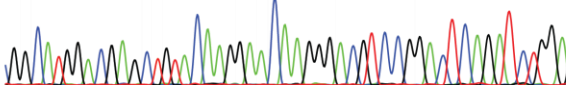
... Exon 8 GGA ACA A AT GCA ACT CCA GTA GTA / 4074 N / TAA
 Gly Thr Asn Ala Thr Pro Val Val / 1358 aa / . (1494aa)

Alt GCATGGACGAGCTGTACAAGGAACAAGTATCTAATATGGTGAAGGTTTCAG

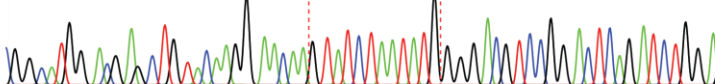
... Exon 8 GGA ACA A GT ATC TAA TAT G AT GCA ...
 Gly Thr Ser Ile . (134aa)

cDNA sequence traces

Ref GG CATGGACGAGCTGTACAAGGAACAAGGACGTC CCGGACTCAGAT CTGGA



Alt GG CATGGACGAGCTGTACAAGGAACAAGGATATCTAATATGGGACGTC CCGGACTCAGATCTGGA



C. Validation outcomes for polypyrimidine tract variants that did not affect splicing

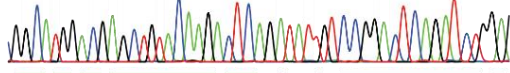
CTNNB1

Construct sequences (intron, exon)

Ref GTGAGTGTGAATTAACCTTTCCAGATATTGTCGGACTCAGATCTGGAGG

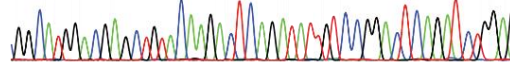
cDNA sequence traces

GGCATGGACGAGCTGTACAAGACTCAGATATTGTCGGACTCAGATCTGGA



Alt GTGAGTGTGAATTAACCTTGTCCAGATATTGTCGGACTCAGATCTGGAGG

GGCATGGACGAGCTGTACAAGACTCAGATATTGTCGGACTCAGATCTGGA



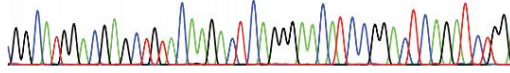
SMARCB1

Construct sequences (intron, exon)

Ref GACTGGGAGGACTTTTCTTGTATCTCCTCAGGAAACATCCGGACTCAGATC

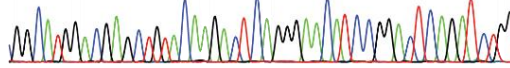
cDNA sequence traces

GGCATGGACGAGCTGTACAAGACTCAGGAAACATCCGGACTCAGATCTGGA



Alt GACTGGGAGGACTTTTCTTGTATCTCCTCAGGAAACATCCGGACTCAGATC

GGCATGGACGAGCTGTACAAGACTCAGGAAACATCCGGACTCAGATCTGGA



D. Validation outcomes for control variants in unaffected parents

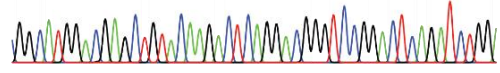
DNM1

Construct sequences (intron, exon)

Ref CCTGTCCCACCTCCTCCCGGGTGCAGGACGGGTCCGGACTCAGATCTGG

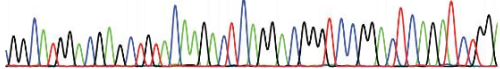
cDNA sequence traces

GGCATGGACGAGCTGTACAAGACTCAGGACGGGTCCGGACTCAGATCTGG



Alt CCTGTCCCACCTCCTCCTCCCGGGTGCAGGACGGGTCCGGACTCAGATCTGG

GGCATGGACGAGCTGTACAAGACTCAGGACGGGTCCGGACTCAGATCTGG



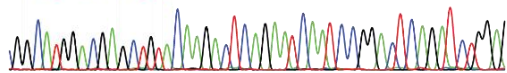
MEF2C

Construct sequences (intron, exon)

Ref CAGTAATGCTTTTATTATTTAAAGAATCAATCCGGACTCAGAT

cDNA sequence traces

GGCATGGACGAGCTGTACAAGACTCAGTAATCAATCCGGACTCAGATCTGGA



Alt CAGTAATGCTTTTATTATTTAAAGAATCAATCCGGACTCAGAT

GGCATGGACGAGCTGTACAAGACTCAGTAATCAATCCGGACTCAGATCTGGA

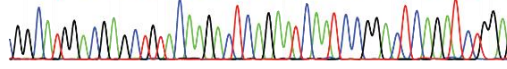


Figure S4 - Proportion of variants in 13,750 unaffected parents of DDD probands which fall within genes with high pLI (>0.9) for pathogenicity score brackets (least to most severe), with Spearman correlation coefficient.

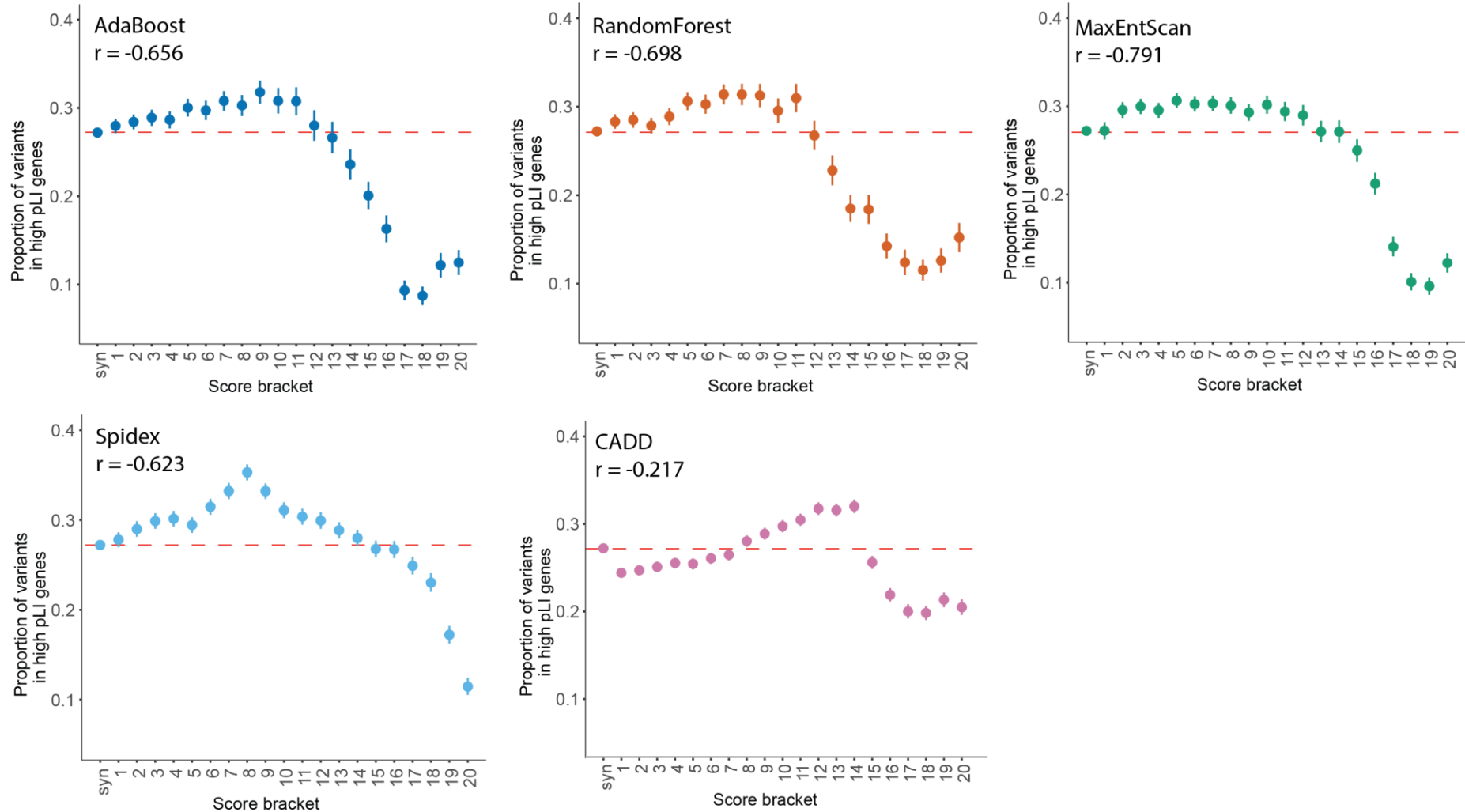


Figure S5 - Mutability adjusted proportion of singletons (MAPS) calculated for pathogenicity score brackets (least to most severe), excluding canonical dinucleotide positions, in 13,750 unaffected parents from the DDD project, with Spearman correlation coefficient.

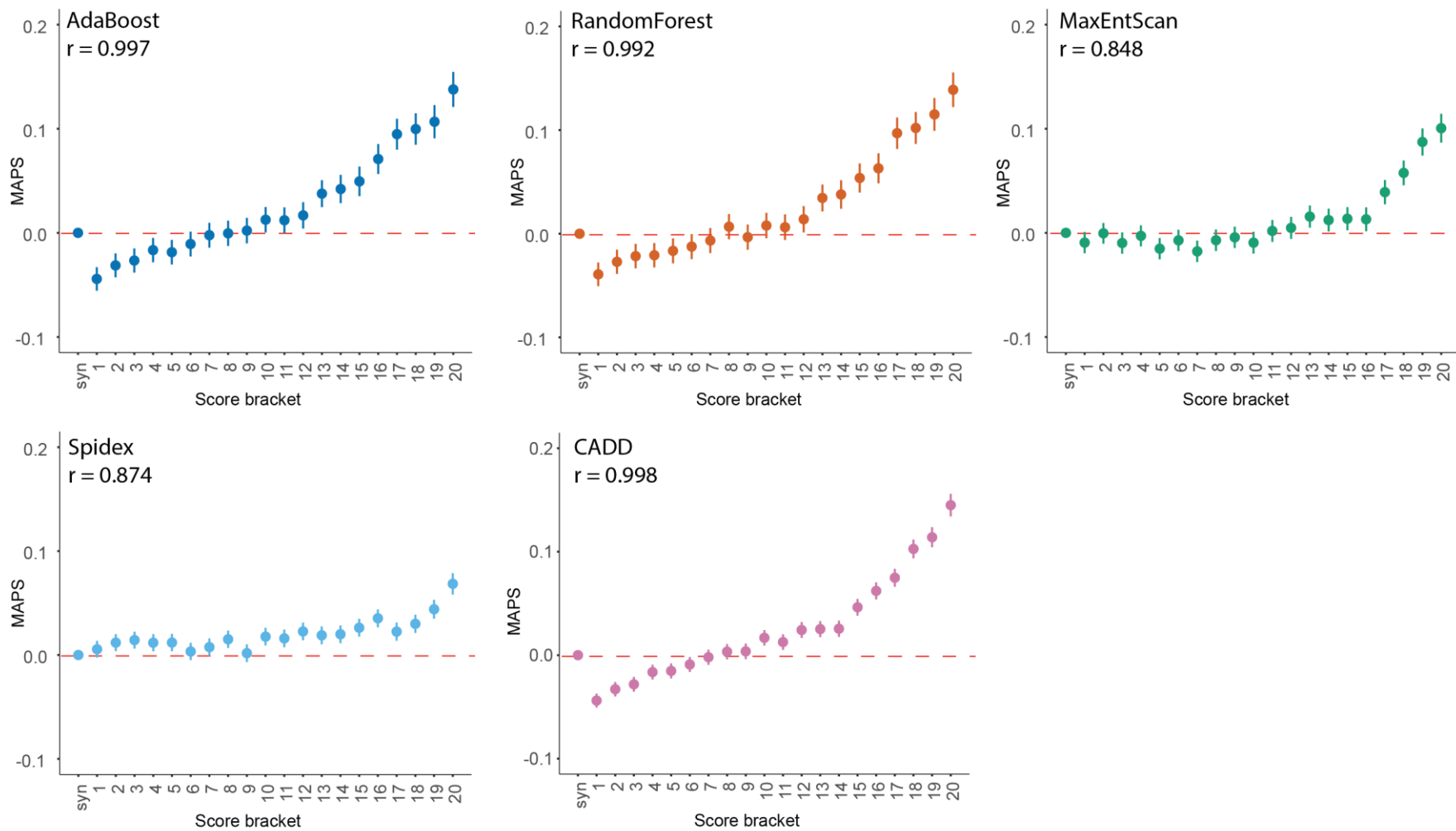


Figure S6 - Proportion of variants in 13,750 unaffected parents of DDD probands which fall within genes with high pLI (>0.9) for pathogenicity score brackets (least to most severe), excluding canonical dinucleotide positions, with Spearman correlation coefficient.

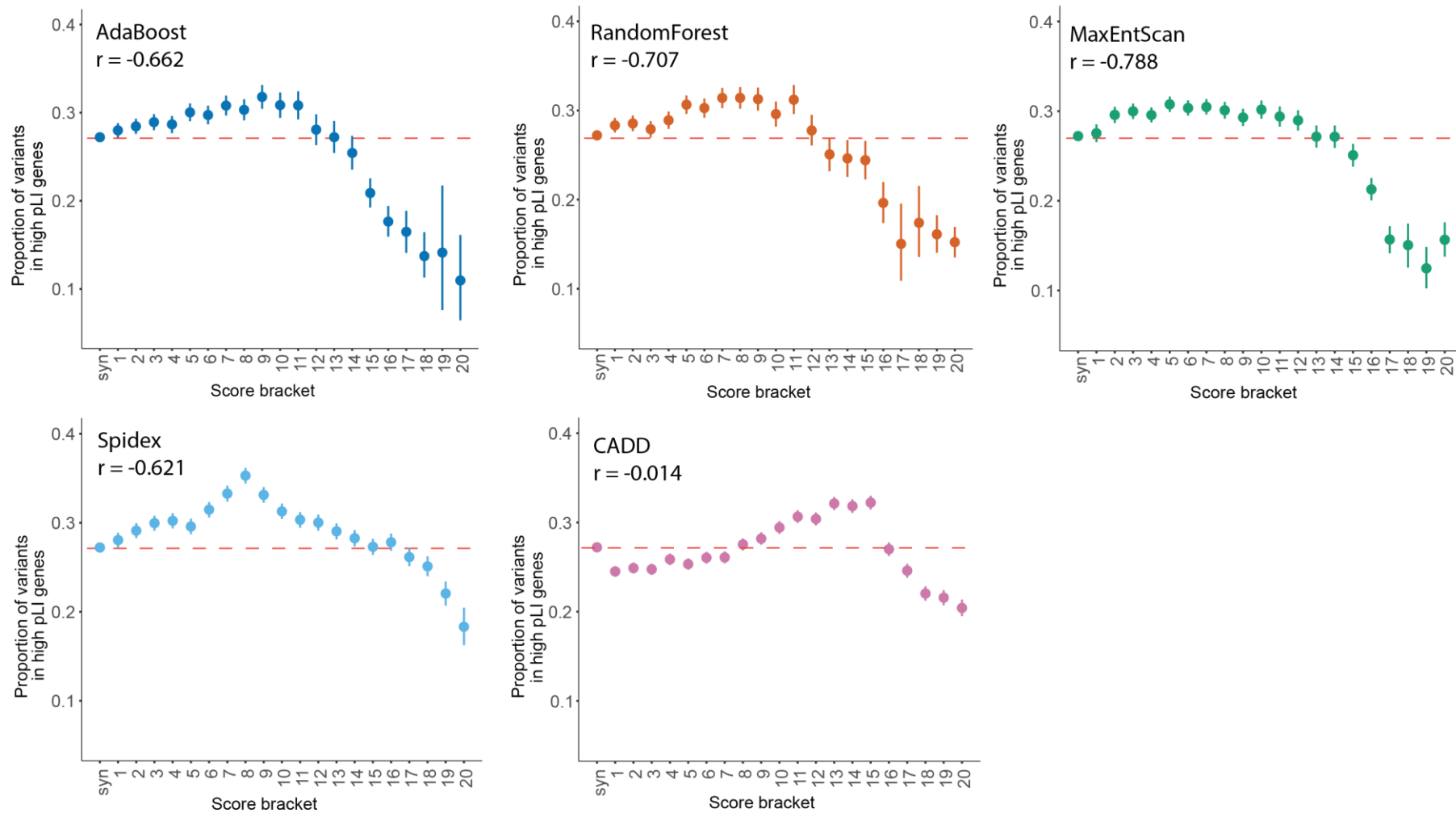


Table S1 – Details of variants selected for validation assay (hg19 coordinates)

Gene of interest	Variant	Intron	Splice annotation	Strand	Reason for validation	Genomic region assayed	Minigene assay outcome
GLI3	7:42063221_G/C	9-10	acc -14	-	PolyPy likely pathogenic	7:42063202-42063393	13bp intron retention
MEF2C	5:88025173_A/C	8-9	acc -9	-	PolyPy likely pathogenic	5:88025159-88025327	8bp intron retention
CHD7	8:61763045_G/A	25-26	acc -7	+	PolyPy likely pathogenic	8:61762849-61763057	5bp intron retention (mixed product)
MBD5	2:149221493_G/C	8-9	don+5	+	Don+5 likely pathogenic	2:149221483-149221651	12bp intron retention (mixed product)
CREBBP	16:3819367_C/T	14-15	acc -13	-	PolyPy likely pathogenic	16:3819349-3819526	11bp intron retention
SMARCB1	22:24143120_T/G	3-4	acc -11	+	PolyPy likely pathogenic	22:24142916-24143136	No effect on splicing
DNM1	9:130988306_G/A	9-10	acc -8	+	PolyPy likely pathogenic	9:130988199-130988319	6bp intron retention
CTNNB1	3:41266439_T/G	3-4	acc -6	+	PolyPy uncertain significance	3:41266294-41266450	No effect on splicing
MEF2C control	5:88025173_A/G (paternal)	8-9	acc -9	-	Negative control	5:88025159-88025327	No effect on splicing
DNM1 control	9:130988302_C/T (maternal)	9-10	acc -12	+	Negative control	9:130988199-130988319	No effect on splicing