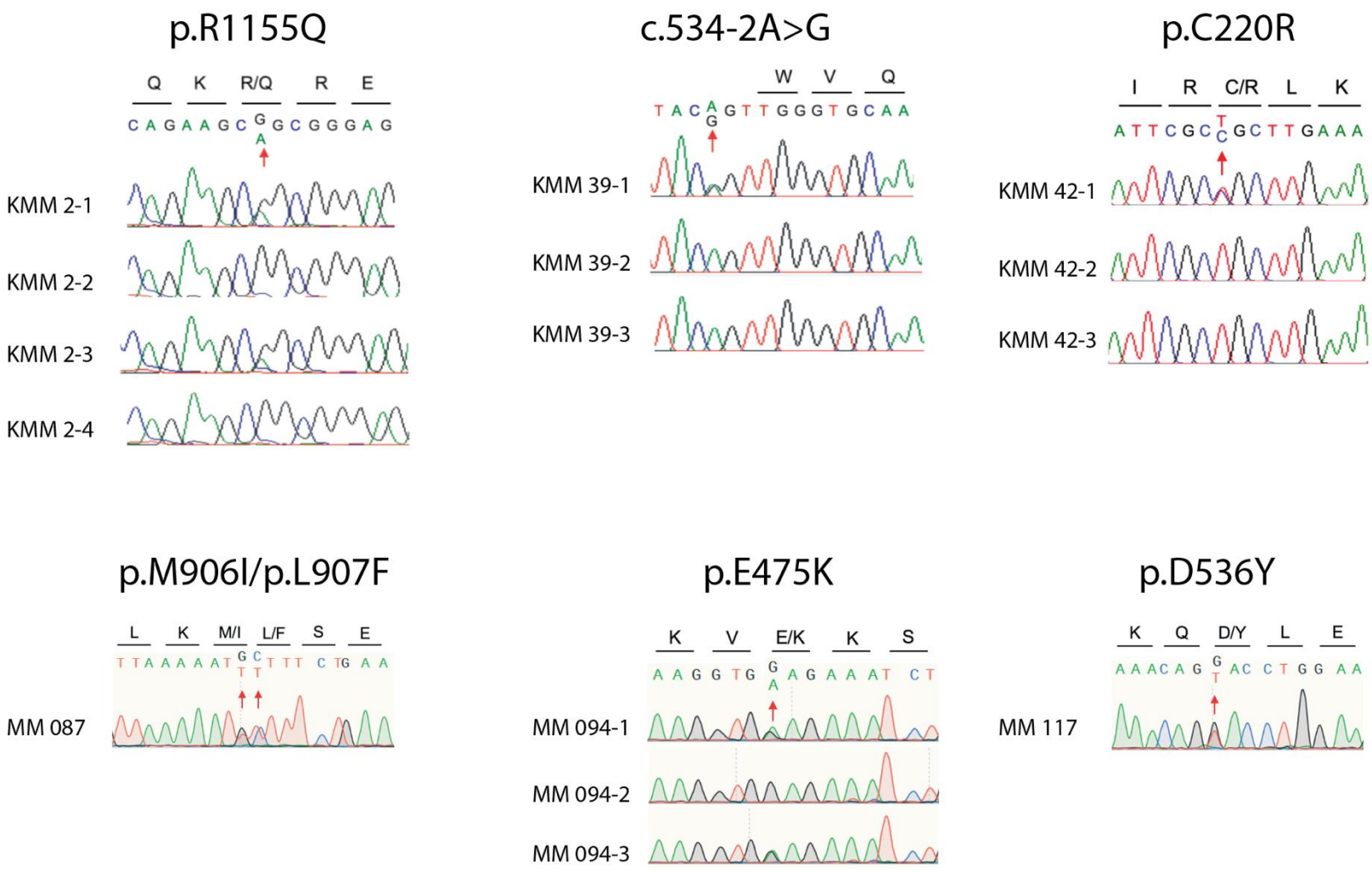
Supplementary Online Content

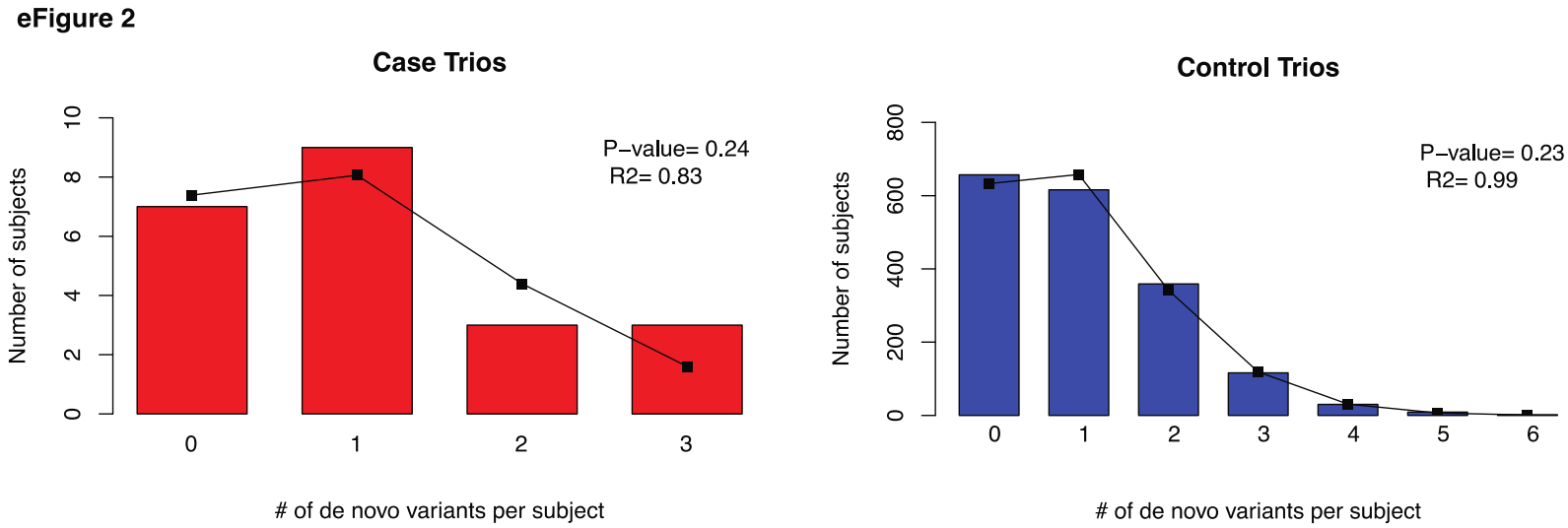
Kundishora AJ, Peters ST, Pinard A, et al. *DIAPH1* variants in non–East Asian patients with sporadic moyamoya disease. *JAMA Neurol*. Published online June 14, 2021. doi:10.1001/jamaneurol.2021.1681

This supplementary material has been provided by the authors to give readers additional information about their work.

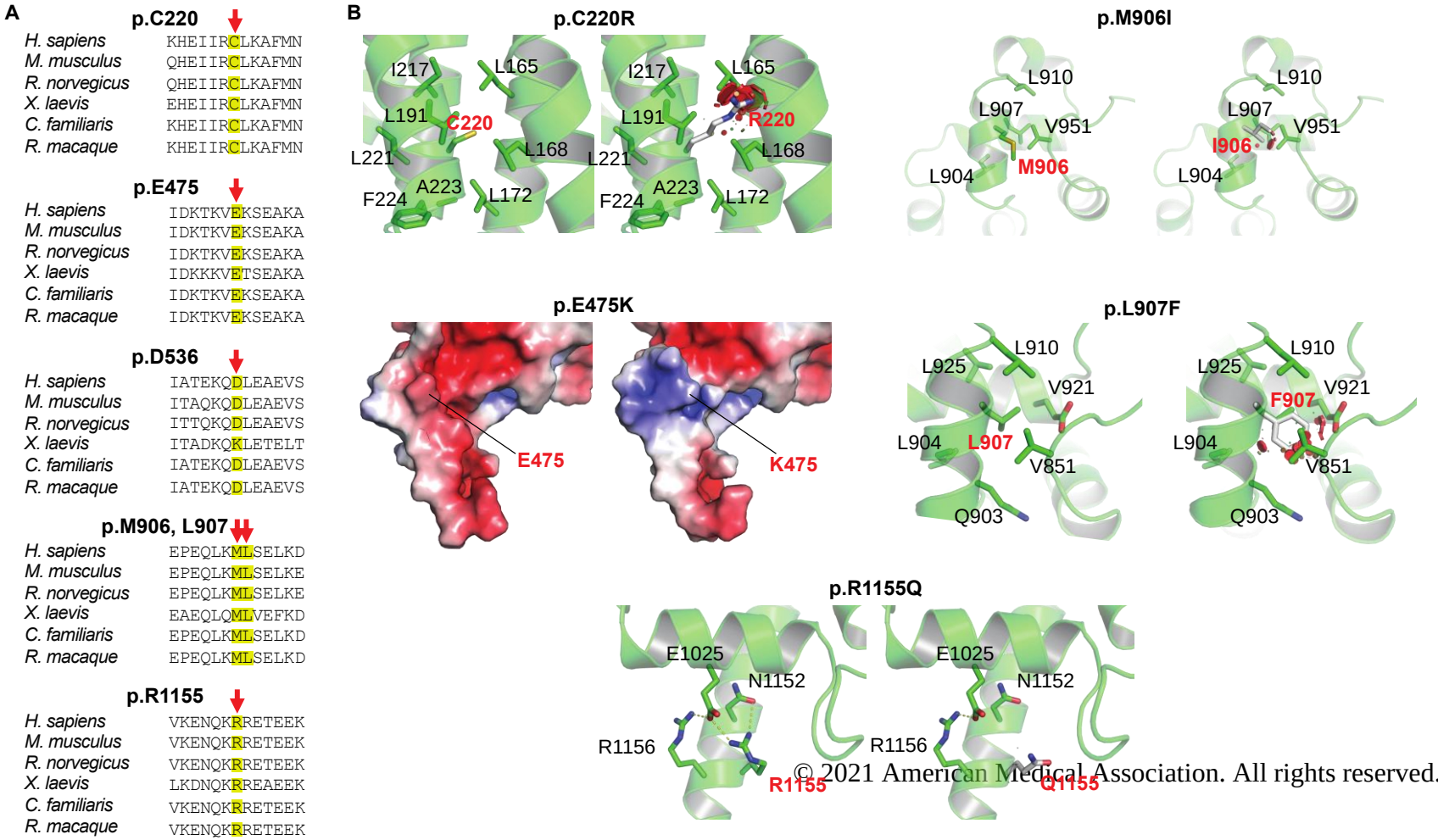
eFigure 1. Sanger Validation Of Identified *DIAPH1* Variants



eFigure 2. *De novo* variant rate closely approximates Poisson distribution in primary MMD cases and controls.
Observed number of *de novo* mutations per subject (bars) compared to the numbers expected (line) from the Poisson distribution in the case (red) and control (blue) cohorts. 'p' denotes chi-squared p-value.



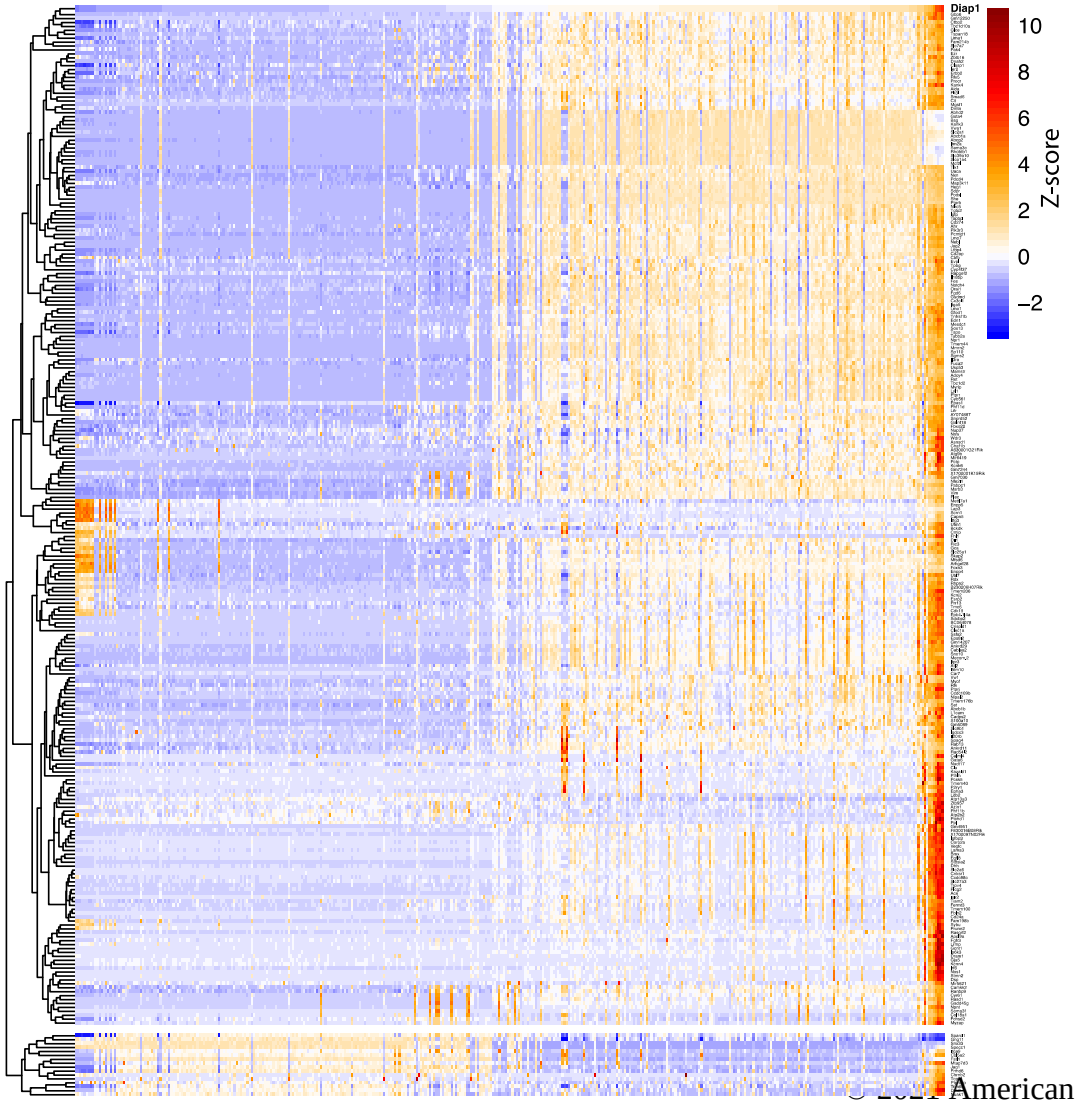
eFigure 3. Phylogenetic conservation and structural implication of De novo and rare, damaging transmitted DIAPH1 variants in moyamoya disease. **(A)** Phylogenetic conservation of amino acid residues in mDia1 mutated in moyamoya disease. The wild-type amino acid is shown, with the mutated amino acid highlighted in yellow. **(B)** Impact of moyamoya disease-associated missense mutations on mDia1 protein structure. The red discs represent the extent of the structural clashes. C220 is positioned in the DID domain. The Cys side chain is surrounded by hydrophobic residues (L165, L168, L172, L191, I217, L221, A223 and F224), which are sequestered within a tightly packed four-helical bundle. A large, bulky, polar, guanidinium side chain of Arg at this position will not fit within the tightly packed hydrophobic region and is predicted to disrupt the tight helical packing. The ΔΔG value for C220R mutation was predicted to be 2.68. E475 is located in a coiled coil region between FH3-Dimerization domain and FH1 domains. The side chain in the structure is solvent exposed and contributes to a negatively charged region. A mutation to a positive charged lysyl side chain alters the charge environment around the region. The ΔΔG value for E475K was predicted to be 0.78. M906 is present in the FH2 domain and is surrounded by hydrophobic residues. A mutation to the branched side chain of Ile results in side chain steric clashes with the other side chains of hydrophobic residues like L907 and V951. The ΔΔG value for M906I mutation was predicted to be 0.17. The side chain of L907 is surrounded in a tight hydrophobic environment by short side chains of L925, L901, and V851. A mutation to a bulky phenylalanine, will make steric clashes with V851 and also the side chain of V921. The ΔΔG value for L907F mutation was predicted to be 2.47. R1155 is a part of the R1155-E1025-R1156 triad, present in the FH2 domain. The guanidinium side chain of R1155 forms a salt bridge with E1025. In addition, R1125 makes hydrogen bond interactions with N1152. These interactions tether α-helices in the FH2 domains together. A mutation to a shorter Gln side chain disrupts the hydrogen bond network and also destabilizing the secondary structure in the region. The ΔΔG value for R1155Q mutation was predicted to be 1.97. See eMethods for further details on predicting protein structure.



eFigure 4. kNN-DREMI Diaph1-associated genes and KEGG terms. **(A)** Heatmap of kNN-DREMI Diaph1-associated genes from Betsholtz scRNAseq data of mouse brain vascular cells. kNN-DREMI was used to find Diaph1-associated genes as described in the Methods. The top 98th percentile of genes associated with Diaph1 expression is shown in a random subset of 400 single-cells. **(B)** Top 10 KEGG terms after genes positively associated with Diaph1 DREMI were enriched, using MsigDB.

eFigure 4

A



eTable 1a. High confidence, monogenic causes of primary MMD

Gene*	OMIM #	Type	Variant(s)*	Variant type	Associated features	Population of discovery	Reference	PMID
RNF213	613378	Common variant risk factor	c.14576G>A (p.Arg4859Lys)	snv	isolated MMD	East Asian	Kamada F *et al.*	21048783
			c.14429G>A (p.Arg4810Lys)	snv			Liu W *et al.*	22688066
			c.12037G>A (p.Asp4013Asn)	snv		East Asian/ Western European	Liu W *et al.*	21799892
			c.12553A>G (p.Lys4185Glu)	snv		Western European	Gagunashvili AN *et al.*	31645973
			c.12343_12345del (p.Lys4115del)	deletion			Cecchi AC *et al.*	25278557
			c.12059G>T (p.Cys4020Phe)	snv			OMIM Evidence	n/a
			c.12360C>G (p.Phe4120Leu)	snv		Mixed population	Pinard A *et al.*	31474762
			c.12040C>A (p.His4014Asn)	snv		Mixed population	Guey S *et al.*	28635853
GUCY1A1	615750	rare large effect	c.1086+1G>A	snv	achalasia	Algerian	Herve D *et al.*	24581742
			c.1045C>T (p.Arg349Ter)	snv		Western European		
			c.1170del (p.Glu391fs)	snv				
			c.1258C>T (p.Arg420Ter)	snv				
			c.1954G>T (p.Gly652Ter)	snv			Wallace S *et al.*	26777256
			c.1550G>A (p.Cys517Tyr)	snv	isolated MMD			
			c.334_335del (p.Glu112fs)	snv				
BRCC3⁺	300617	rare large effect	deletion on Xq28	deletion	Short stature, hypergonadotropic hypogonadism, hypertension, facial dysmorphism, dilated cardiomyopathy, premature grey hair, early-onset cataracts	North African	Miskinyte S *et al.*	21596366
MTCP1⁺	300116							

eTable 1b. High confidence, monogenic causes of syndromes with associated MM pathology

Gene	OMIM #	Type	Variant(s)*	Variant type	Syndrome	Population of discovery	Reference	PMID
RNF213	613378	Common variant risk factor	c.6169G>A (p.Asp2057Asn)	snv	Midaortic Syndrome	Mixed population	Warejko JK *et al.*	29483232
ACTA2	614042	rare large effect	c.773G>A (p.Arg258His)	snv	Smooth muscle dysfunction syndrome	Western European	Guo D *et al.*	17994018
			c.772C>T (p.Arg258Cys)	snv				
			c.536G>A (p.Arg179His)	snv			Milewicz *et al.*	20734336
PCNT	210720	rare large effect	c.844dup (p.Arg281fs)	duplication	microcephalic osteodysplastic primordial dwarfism, type II	Eastern European	Rauch A *et al.*	18174396
			PCNT, 486-BP DEL, NT84	deletion		Middle Eastern	Rauch A *et al.*	18174396
			c.3109G>T (p.Glu1037Ter)	snv		Eastern European	Willems M *et al.*	19643772
YY1AP1	607860	rare large effect	c.724C>T (p.Gln242Ter)	snv	Grange syndrome	Western European	Guo D, *et al.*	27939641

* reported genes and variants include those with high quality evidence reported in OMIM and determined likely pathogenic in ClinVar

[+] BRCC3 and MTCP1 are within the critical overlap region of deletion on Xq28 which is responsible for MMD 4

eTable 2. Number of studied cases from Yale, UT Health, and Simons Simplex Collection (SSC)

Family Type	Total Number from Yale (# Trios)	Total Number from UT Health (# Trios)	Total Number from SSC (# Trios)
Singleton	2	52	0
Trio (affected proband and parents)	18	28	1,798
Quad (affected proband, unaffected sibling, parents)	1 (1)	0 (0)	0 (0)
Other family structures	3 (3)	4 (1)	0 (0)
Total	**24 (22)**	**84 (29)**	**1,798 (1,798)**

eTable 3. Phenotypes for Primary Probands

a. Phenotypes for 24 Primary Moyamoya probands enrolled at Yale

Proband ID	Sex	Ethnicity	Family Type
KMM-4-1	F	European	Trio
KMM_50-1	F	Mexican	Trio
KMM_5-1	M	European	Trio
KMM_45-1	F	European	Trio

KMM_42-1	F	European	Trio
KMM_41-1	M	European	Trio
KMM_40-1	M	European	Trio
KMM_39-1	M	European	Trio
KMM_36-1	F	European	Trio
KMM_35-1	M	Mexican	Trio
KMM_3-1	M	Mexican	Trio
KMM_28-1	F	African American/African	Trio
KMM_27-1	F	European	Trio
KMM_23-1	M	Mexican	Trio
KMM_21-1	F	European	Trio
KMM_2-1	M	European	Trio
KMM_19-1	M	European	Trio
KMM_16-1	F	European	Trio
KMM_14-1	M	European	Trio
KMM_13-3	M	East Asian	Trio
KMM_11-1	M	European	Trio
CHO_0001-01	M	European	Trio
KMM_22-1	F	European	Singleton
KMM_29-1	M	European	Singleton

b. Phenotypes for 84 Primary moyamoya probands enrolled at UTHealth

Proband ID	Sex	Ethnicity	Family Type
MM005	F	European	Trio
MM010	F	Other	Trio
MM048	F	European	Trio
MM050	F	European	Trio
MM052	M	Mexican	Trio
MM055	F	European	Trio
MM071	F	European	Trio
MM072	F	European	Trio
MM075	F	European	Trio
MM092	F	European	Trio
MM094	M	European	Trio
MM123	M	European	Trio
MM131	M	European	Trio
MM150	M	European	Trio
MM151	F	Other	Trio
MM152	M	European	Trio
MM153	F	European	Trio
MM154	M	European	Trio
MM156	M	European	Trio
MM158	F	European	Trio
MM159	F	European	Trio
MM162	M	European	Trio
MM165	F	European	Trio
MM166	F	European	Trio
MM169	M	European	Trio
MM170	M	European	Trio
MM173	F	European	Trio

MM185	M	European	Trio
MM013	F	Other	Trio
MM015	F	European	Singleton
MM035	M	European	Singleton
MM112	F	European	Singleton
MM125	M	European	Singleton
MM139	F	Mexican	Singleton
MM163	F	European	Singleton
MM001	F	European	Singleton
MM130	M	European	Singleton
MM137	F	European	Singleton
MM003	F	East Asian	Singleton
MM006	M	European	Singleton
MM007	F	Other	Singleton
MM009	F	Other	Singleton
MM011	F	European	Singleton
MM012	F	European	Singleton
MM023	F	European	Singleton
MM026	F	European	Singleton
MM028	F	European	Singleton
MM029	M	European	Singleton
MM030	F	European	Singleton
MM039	F	European	Singleton
MM041	M	European	Singleton
MM042	F	European	Singleton
MM051	F	European	Singleton
MM060	F	European	Singleton
MM061	M	Mexican	Singleton
MM078	M	European	Singleton

MM083	F	European	Singleton
MM087	F	African American/African	Singleton
MM093	F	African American/African	Singleton
MM095	F	European	Singleton
MM097	F	European	Singleton
MM098	F	European	Singleton
MM100	F	European	Singleton
MM101	F	European	Singleton
MM102	F	European	Singleton
MM106	F	European	Singleton
MM108	F	European	Singleton
MM109	F	African American/African	Singleton
MM110	F	African American/African	Singleton
MM117	F	African American/African	Singleton
MM120	F	European	Singleton
MM124	F	European	Singleton
MM128	M	South Asian	Singleton
MM133	F	East Asian	Singleton
MM135	F	South Asian	Singleton
MM138	F	European	Singleton
MM140	F	European	Singleton
MM160	F	European	Singleton
MM161	F	East Asian	Singleton
MM172	F	European	Singleton
MM175	F	Other	Singleton
MM184	F	European	Singleton
MM190	M	European	Singleton
MM191	F	European	Singleton

eTable 4. Summary sequencing statistics for the primary MMD and control cohorts

Category	Cases (NimbleGen V2; N=3)	Cases (MedExome; N=9)	Cases (xGen IDT; N=56)[&]	Controls (Roche V2; N=5,394)
Read length (bp)	76	76-101	101	50-99
# of reads per sample (M)	73.7	60.8	78.1	111.5
Median coverage at each targeted base (X)	61.7	37.3	85.1	68
Mean coverage at each targeted base (X)	72.7	43.7	90.4	80.8
% of all reads that map to target	62.30%	45.90%	58.80%	46.50%
% of all bases that map to target	46.70%	37.10%	44.10%	35.70%
% of targeted bases read at least 8x	93.70%	96.10%	98.30%	92.70%
% of targeted bases read at least 10x	92.40%	94.80%	98.10%	91.40%
% of targeted bases read at least 15x	89.70%	89.70%	97.40%	87.80%
% Mean error rate	0.40%	0.30%	0.20%	0.50%

eTable 5. *De novo* variants in 22 primary MMD trios

ID	CHROM	POS	REF	ALT	Gene	AA change	Variant_Class	MetaSVM_ pred	CADD v1.6	ExAC_ALL	Bravo	gnomAD _exome_ ALL	gnomAD _genome _ALL	pLI score (gnomAD v2.1.1)	mis_z (gnomAD v2.1.1)
KMM_23-1	16	86612574	G	T	*FOXL1*	p.R82L	D-Mis	D	27.7	.	.	.	.	0.04	-0.28
KMM_21-1	8	87616341	C	G	*CNGB3*	p.G587G	synonymous	.	4.8	.	.	.	.	0.00	-1.18
KMM_11-1	19	15281611	T	G	*NOTCH3*	p.N1588H	D-Mis	T	26.3	.	.	.	.	0.41	3.53
KMM_13-3	18	25589724	G	A	*CDH2*	p.S220L	D-Mis	T	24.8	6.60E-05	.	6.10E-05	3.23E-05	0.99	2.09
KMM_13-3	19	59073731	G	A	*MZF1*	p.S638L	D-Mis	T	25.8	.	.	.	.	0.00	-0.26
KMM_13-3	19	757502	G	C	*MISP*	p.V186L	T-Mis	T	11.9	1.39E-05	.	8.86E-06	.	0.00	-1.17
KMM_14-1	12	119937952	G	A	*CCDC60*	p.E209E	synonymous	.	8.2	.	.	.	.	0.00	0.14
KMM_16-1	2	170163857	T	A	*LRP2*	p.I121F	D-Mis	D	21.2	.	.	.	.	1.00	2.07
KMM_27-1	1	68960137	A	C	*DEPDC1*	p.V97G	D-Mis	T	23.3	.	.	.	.	0.00	0.44
KMM_27-1	12	88586534	T	C	*TMTC3*	p.F620F	synonymous	.	11.3	.	7.96E-06	.	.	0.00	1.62
KMM_2-1	1	186330829	C	T	*TPR*	p.D295N	D-Mis	T	32.0	.	.	.	.	1.00	2.99
KMM_2-1	19	1052268	G	A	*ABCA7*	p.S1068N	T-Mis	T	11.7	.	.	.	.	0.00	-1.47
KMM_2-1	4	57180668	CG	C	*KIAA1211*	p.R334fs	frameshift_deletion	.	.	.	.	.	.	0.00	0.36
KMM_35-1	17	12028609	A	C	*MAP2K4*	c.847-2A>C	splicing	.	33.0	.	.	.	.	1.00	3.23
KMM_36-1	12	72030698	C	T	*ZFC3H1*	p.E623K	D-Mis	T	25.9	.	.	.	.	1.00	2.14
KMM_39-1	5	140962861	T	C	*DIAPH1*	c.534-2A>G	splicing	.	33.0	.	.	.	.	0.92	1.65
KMM_40-1	12	18552718	A	C	*PIK3C2G*	p.H529P	D-Mis	T	24.2	.	.	.	.	0.00	-0.97
KMM_40-1	8	104240298	C	G	*BAALC*	p.R172G	D-Mis	T	20.2	.	.	.	.	0.00	-0.26
KMM_42-1	19	33609999	C	T	*GPATCH1*	p.R833W	D-Mis	T	23.4	3.50E-05	7.96E-06	1.64E-05	.	0.00	0.68
KMM_42-1	5	140961905	A	G	*DIAPH1*	p.C220R	D-Mis	D	28.2	.	.	.	.	0.92	1.65
KMM_42-1	5	6742701	T	C	*TENT4A*	p.N369N	synonymous	.	11.0	.	.	.	.	1.00	2.31
KMM_50-1	10	33200833	G	A	*ITGB1*	p.S563S	synonymous	.	11.6	.	.	.	.	0.98	3.46
KMM_5-1	17	71434185	G	A	*SDK2*	p.D278D	synonymous	.	5.3	.	3.19E-05	4.00E-05	3.23E-05	0.00	2.68
KMM_5-1	3	151163065	A	C	*IGSF10*	p.S1568S	synonymous	.	0.8	.	.	.	.	0.00	-1.63

AA Change - Amino Acid Change; Bravo WGS Freq - Bravo Whole Genome Sequencing Frequency; gnomAD WES Freq - Gnome Aggregation Database Whole Exome Sequence Frequency; MetaSVM - Meta-analytic support vector machine is an ensemble score that predicts the tolerability of a mutation. CADD1.6 - Combined Annotated Dependent Depletion is a validated tool for scoring the deleteriousness of single nucleotide deletions, insertions, or variations. pLI is a gene-wide constraint metric that estimates the probability of being intolerant to loss-of-function mutations. mis_Z is a gene-wide constraint metric that estimates the probability of being intolerant to missense mutations.

eTable 6. Demographic characteristics of primary MMD cases and controls

.

	Cases	UTHealth cases	Autism Sibling Controls
Sample size	24	84	1,798
Gender			
Male	14 (58.3%)	22 (26.2%)	842 (46.8%)
Female	10 (41.7%)	62 (73.8%)	956 (53.2%)
Ethnicity			
European	18 (75.0%)	65 (77.4%)	1,418 (78.9%)
African American/African*	1 (4.2%)	5 (5.9%)	77 (4.3%)
East Asian	1 (4.2%)	3 (3.6%)	40 (2.2%)
South Asian	0 (0.0%)	2 (2.4%)	88 (4.9%)
Mexican	4 (16.6%)	3 (3.6%)	129 (7.2%)
Other	0 (0.0%)	6 (7.1%)	46 (2.6%)

The number of samples is shown in each category with the corresponding percentage in parentheses. Ethnicity is determined by principal component analysis compared to HapMap samples using EIGENSTRAT. Plink --check-sex was used to determine sex.

eTable 7 Sequencing metrics for all reported samples

ID	Read Length	Num reads (M)	Num bases (G)	Mean cover-age	Med-ian cover-age	PCR dupli-cates	Multiply mapped	Un-mapped	Reads on-target:	Bases on-target	Mean error rate	1x target base cover-age	2x target base cover-age	4x target base cover-age	8x target base cover-age	10x target base cover-age	15x target base cover-age	20x target base cover-age	30x target base cover-age	40x target base cover-age	50x target base cover-age	100x target base cover-age	Targets
CHO_0001-01	76	65.6	5	65.9	56	0.0476	0.0648	0.0012	0.6333	0.4752	0.0036	0.98	0.973	0.96	0.934	0.92	0.882	0.839	0.746	0.65	0.558	0.212	V2
CHO_0001-02	76	84.7	6.4	83.3	71	0.0592	0.0629	0.0013	0.6205	0.4649	0.0039	0.981	0.974	0.963	0.943	0.932	0.904	0.872	0.802	0.727	0.651	0.33	V2
CHO_0001-03	76	70.7	5.4	69	58	0.0605	0.064	0.0012	0.6155	0.4613	0.0037	0.981	0.973	0.96	0.934	0.92	0.883	0.842	0.753	0.661	0.572	0.236	V2
KMM_11-1	101	57.2	5.8	67.6	65	0.1627	0.0672	0.0022	0.6135	0.4551	0.0019	0.987	0.986	0.985	0.984	0.983	0.978	0.969	0.926	0.842	0.72	0.124	IDT
KMM_11-2	101	74.2	7.5	95.7	92	0.1065	0.0642	0.0012	0.6646	0.4968	0.0018	0.987	0.986	0.985	0.984	0.983	0.982	0.98	0.97	0.947	0.902	0.414	IDT
KMM_11-3	101	59.6	6	77.9	74	0.0957	0.0694	0.0022	0.6759	0.5032	0.0019	0.987	0.986	0.985	0.984	0.983	0.98	0.975	0.951	0.9	0.814	0.215	IDT
KMM_13-1	101	83.9	8.5	102.1	99	0.1264	0.0615	0.0009	0.6263	0.4684	0.0019	0.988	0.988	0.987	0.986	0.985	0.984	0.982	0.973	0.952	0.913	0.493	IDT
KMM_13-2	101	54.9	5.5	62.3	60	0.169	0.0607	0.0021	0.5863	0.4368	0.0019	0.987	0.986	0.985	0.984	0.983	0.979	0.971	0.925	0.826	0.683	0.066	IDT
KMM_13-3	101	64.5	6.5	82.7	79	0.0831	0.0662	0.0008	0.6601	0.4938	0.0019	0.988	0.987	0.986	0.985	0.985	0.982	0.978	0.959	0.914	0.84	0.27	IDT
KMM_14-1	101	47	4.7	60.3	58	0.0925	0.0671	0.0023	0.661	0.4942	0.0018	0.988	0.987	0.986	0.984	0.983	0.977	0.964	0.906	0.794	0.64	0.062	IDT
KMM_14-2	101	63.9	6.5	79.9	77	0.0923	0.0617	0.0008	0.6456	0.4814	0.0018	0.987	0.987	0.986	0.985	0.984	0.982	0.979	0.963	0.921	0.844	0.227	IDT
KMM_14-3	101	79.5	8	100.8	97	0.0997	0.0692	0.0008	0.6533	0.4879	0.0018	0.988	0.987	0.986	0.986	0.985	0.984	0.982	0.973	0.953	0.915	0.47	IDT
KMM_16-1	101	87.9	8.9	110.8	107	0.0981	0.0663	0.0007	0.6494	0.4856	0.0018	0.987	0.986	0.985	0.985	0.984	0.983	0.982	0.977	0.966	0.942	0.573	IDT
KMM_16-2	101	91.7	9.3	115.3	111	0.0872	0.0684	0.0006	0.646	0.4844	0.0017	0.987	0.987	0.986	0.985	0.985	0.984	0.983	0.978	0.967	0.944	0.603	IDT
KMM_16-3	101	108.6	11	133.7	129	0.1019	0.0663	0.0007	0.6335	0.4742	0.0019	0.988	0.988	0.987	0.986	0.986	0.985	0.984	0.981	0.973	0.958	0.721	IDT
KMM_19-1	101	68.8	6.9	82	79	0.1392	0.0584	0.0012	0.6201	0.4589	0.0018	0.988	0.987	0.986	0.985	0.984	0.982	0.979	0.962	0.921	0.85	0.261	IDT
KMM_19-2	101	88.8	9	107.4	100	0.1524	0.064	0.0017	0.6236	0.4658	0.0019	0.987	0.986	0.985	0.984	0.984	0.982	0.98	0.973	0.955	0.919	0.509	IDT
KMM_19-3	101	33.5	3.4	39.4	38	0.1681	0.0657	0.0058	0.6088	0.4523	0.0018	0.987	0.986	0.985	0.98	0.976	0.952	0.902	0.709	0.449	0.232	0.008	IDT
KMM_2-1	101	81.1	8.2	103.4	93	0.1275	0.0671	0.0015	0.6555	0.4912	0.002	0.987	0.987	0.986	0.984	0.983	0.981	0.976	0.957	0.92	0.863	0.447	IDT
KMM_2-2	101	120.7	12.2	146.8	141	0.1208	0.0614	0.0008	0.6251	0.4687	0.0019	0.987	0.987	0.986	0.985	0.985	0.984	0.983	0.981	0.976	0.966	0.78	IDT
KMM_2-3	101	66.9	6.8	41.1	41	0.5255	0.0466	0.0084	0.3296	0.2368	0.0018	0.987	0.986	0.984	0.982	0.979	0.968	0.941	0.802	0.531	0.257	0.002	IDT
KMM_21-1	76	74	5.6	49.5	42	0.0883	0.0769	0.0291	0.4954	0.4105	0.0029	0.984	0.981	0.974	0.957	0.947	0.91	0.858	0.712	0.545	0.391	0.056	MedExome

KMM_21-2	76	90.5	6.9	61.7	53	0.0917	0.0756	0.0056	0.5069	0.4195	0.0025	0.985	0.982	0.978	0.966	0.959	0.936	0.904	0.809	0.681	0.544	0.118	MedExome
KMM_21-3	76	80.5	6.1	54.9	48	0.0874	0.0817	0.0113	0.5037	0.4183	0.003	0.986	0.984	0.979	0.969	0.962	0.938	0.901	0.786	0.63	0.47	0.07	MedExome
KMM_23-1	101	49.2	5	32.6	28	0.1532	0.0678	0.1681	0.3808	0.3057	0.0035	0.985	0.983	0.976	0.95	0.929	0.843	0.72	0.451	0.247	0.13	0.02	MedExome
KMM_23-4	101	61.6	6.2	47.7	41	0.1864	0.0644	0.0045	0.451	0.3567	0.0039	0.986	0.984	0.981	0.97	0.962	0.931	0.879	0.715	0.521	0.35	0.044	MedExome
KMM_23-5	101	55.9	5.6	37.4	32	0.1802	0.067	0.025	0.3885	0.3081	0.0036	0.987	0.985	0.98	0.963	0.949	0.891	0.798	0.556	0.338	0.19	0.025	MedExome
KMM_27-1	101	99	10	123.4	117	0.092	0.0663	0.0007	0.6389	0.4798	0.0018	0.987	0.987	0.986	0.985	0.985	0.984	0.983	0.979	0.969	0.949	0.647	IDT
KMM_27-2	101	101.9	10.3	123.7	117	0.139	0.0674	0.0009	0.6219	0.4674	0.0018	0.987	0.986	0.986	0.985	0.985	0.984	0.982	0.977	0.965	0.941	0.636	IDT
KMM_27-3	101	122.8	12.4	154.5	143	0.1008	0.0668	0.0006	0.6428	0.4849	0.0018	0.988	0.988	0.987	0.987	0.986	0.985	0.984	0.98	0.971	0.955	0.751	IDT
KMM_28-1	101	47.5	4.8	22.4	20	0.644	0.0337	0.0146	0.2511	0.1819	0.0019	0.984	0.982	0.978	0.95	0.916	0.759	0.542	0.211	0.079	0.031	0	IDT
KMM_28-2	101	70.8	7.2	32.7	31	0.6567	0.036	0.0071	0.2466	0.1775	0.0019	0.984	0.983	0.981	0.976	0.972	0.942	0.864	0.554	0.253	0.101	0.002	IDT
KMM_28-3	101	80.1	8.1	89.9	81	0.2406	0.0639	0.0017	0.5785	0.4321	0.002	0.987	0.986	0.985	0.983	0.982	0.98	0.975	0.953	0.907	0.833	0.332	IDT
KMM_3-1	101	80.1	8.1	101.3	94	0.1295	0.0675	0.001	0.6485	0.4869	0.0019	0.987	0.987	0.986	0.985	0.984	0.982	0.979	0.964	0.934	0.884	0.451	IDT
KMM_3-2	101	56.3	5.7	68.2	63	0.151	0.0744	0.0027	0.625	0.466	0.0018	0.986	0.986	0.985	0.983	0.982	0.977	0.967	0.919	0.824	0.692	0.142	IDT
KMM_3-3	101	84.9	8.6	107.9	104	0.0928	0.0644	0.0006	0.6519	0.4892	0.0018	0.988	0.987	0.987	0.986	0.985	0.984	0.982	0.974	0.954	0.919	0.538	IDT
KMM_35-1	101	126.7	12.8	162.2	156	0.1262	0.0688	0.0011	0.6571	0.4934	0.0018	0.988	0.987	0.986	0.985	0.985	0.984	0.983	0.981	0.977	0.969	0.834	IDT
KMM_35-2	101	99.5	10.1	124.3	116	0.1467	0.0683	0.001	0.6405	0.4811	0.002	0.986	0.986	0.985	0.984	0.983	0.982	0.981	0.976	0.965	0.944	0.639	IDT
KMM_35-3	101	73.1	7.4	89.8	85	0.1909	0.0695	0.0022	0.635	0.4728	0.0018	0.987	0.986	0.985	0.982	0.981	0.978	0.971	0.948	0.904	0.837	0.364	IDT
KMM_36-1	101	124.5	12.6	154.2	148	0.1231	0.0642	0.0009	0.6353	0.4773	0.0017	0.987	0.987	0.986	0.985	0.985	0.985	0.984	0.982	0.977	0.968	0.795	IDT
KMM_36-2	101	85.7	8.7	109.6	105	0.1028	0.0674	0.0006	0.6536	0.4924	0.0018	0.987	0.986	0.986	0.985	0.984	0.983	0.981	0.974	0.957	0.924	0.55	IDT
KMM_36-3	101	110.8	11.2	126.5	120	0.2149	0.0588	0.0014	0.5808	0.4398	0.0021	0.988	0.987	0.986	0.986	0.985	0.984	0.982	0.976	0.961	0.936	0.649	IDT
KMM_39-1	101	93.2	9.4	109.1	105	0.1816	0.0614	0.0014	0.5987	0.451	0.0022	0.988	0.987	0.987	0.986	0.985	0.984	0.982	0.973	0.955	0.92	0.55	IDT
KMM_39-2	101	78.8	8	95.7	91	0.1588	0.0647	0.0011	0.6185	0.468	0.0021	0.987	0.986	0.985	0.984	0.984	0.982	0.979	0.966	0.935	0.882	0.418	IDT
KMM_39-3	101	97.6	9.9	115.1	102	0.2085	0.0624	0.002	0.5983	0.454	0.0023	0.987	0.987	0.986	0.984	0.983	0.98	0.976	0.957	0.922	0.87	0.514	IDT
KMM_40-1	101	76	7.7	89.3	86	0.1748	0.0634	0.0014	0.5988	0.4524	0.0022	0.988	0.987	0.986	0.985	0.985	0.982	0.978	0.96	0.918	0.851	0.358	IDT
KMM_40-2	101	58.7	5.9	65.6	63	0.2311	0.0593	0.0014	0.5747	0.43	0.0022	0.986	0.985	0.984	0.983	0.982	0.98	0.975	0.944	0.867	0.737	0.085	IDT

KMM_40-3	101	50.1	5.1	56.7	54	0.2326	0.0619	0.0032	0.5795	0.4358	0.0022	0.987	0.986	0.985	0.982	0.98	0.971	0.953	0.875	0.742	0.574	0.057	IDT
KMM_41-1	101	74.3	7.5	87.2	83	0.2093	0.0599	0.0014	0.5987	0.4516	0.0022	0.987	0.986	0.985	0.984	0.983	0.98	0.975	0.954	0.908	0.838	0.334	IDT
KMM_41-2	101	104.9	10.6	121.6	114	0.2097	0.0586	0.0012	0.5905	0.4466	0.0022	0.987	0.986	0.985	0.984	0.984	0.983	0.981	0.974	0.958	0.928	0.607	IDT
KMM_41-3	101	53.4	5.4	49.2	45	0.3586	0.0508	0.0037	0.4717	0.3545	0.0021	0.987	0.986	0.985	0.981	0.978	0.962	0.929	0.799	0.611	0.422	0.042	IDT
KMM_42-1	101	64.5	6.5	74.5	72	0.1958	0.0596	0.0015	0.5908	0.4448	0.0022	0.987	0.986	0.985	0.984	0.983	0.981	0.976	0.95	0.89	0.792	0.18	IDT
KMM_42-2	101	54.3	5.5	65.4	63	0.1643	0.061	0.0012	0.6139	0.4641	0.003	0.987	0.986	0.985	0.983	0.982	0.978	0.968	0.92	0.826	0.697	0.101	IDT
KMM_42-3	101	63.9	6.4	76.7	72	0.1877	0.0596	0.0019	0.6107	0.4624	0.0031	0.987	0.986	0.985	0.984	0.983	0.979	0.971	0.934	0.864	0.766	0.23	IDT
KMM_45-1	101	72.6	7.3	66.5	54	0.3904	0.0476	0.0021	0.4705	0.3531	0.0031	0.985	0.984	0.982	0.978	0.975	0.96	0.931	0.827	0.688	0.553	0.179	IDT
KMM_45-2	101	76.6	7.7	53.5	45	0.5107	0.0398	0.004	0.3651	0.2688	0.0024	0.985	0.984	0.982	0.979	0.977	0.963	0.929	0.789	0.597	0.429	0.092	IDT
KMM_45-3	101	61.2	6.2	73.5	61	0.2207	0.0596	0.0018	0.6086	0.4621	0.0031	0.986	0.985	0.983	0.979	0.976	0.96	0.933	0.846	0.733	0.619	0.238	IDT
KMM_5-1	101	91.7	9.3	115.1	102	0.1466	0.0656	0.0017	0.6458	0.4831	0.0019	0.987	0.986	0.986	0.984	0.983	0.981	0.977	0.961	0.93	0.882	0.516	IDT
KMM_5-2	101	66.2	6.7	87	82	0.0996	0.0682	0.0012	0.674	0.5058	0.0018	0.986	0.985	0.985	0.983	0.983	0.981	0.977	0.962	0.925	0.859	0.315	IDT
KMM_5-3	101	87.6	8.9	112.5	104	0.1224	0.0629	0.001	0.6576	0.4944	0.0019	0.987	0.987	0.986	0.985	0.984	0.982	0.979	0.968	0.943	0.901	0.533	IDT
KMM_50-1	101	61.6	6.2	51.9	50	0.4001	0.045	0.0015	0.435	0.324	0.0029	0.986	0.985	0.984	0.982	0.981	0.973	0.953	0.861	0.701	0.512	0.021	IDT
KMM_50-2	101	66.1	6.7	78.4	73	0.1715	0.0599	0.0012	0.6027	0.4571	0.0031	0.987	0.986	0.985	0.984	0.983	0.979	0.971	0.931	0.857	0.759	0.256	IDT
KMM_50-3	101	56.4	5.7	60.9	57	0.232	0.0583	0.0059	0.5525	0.4161	0.0028	0.987	0.987	0.986	0.983	0.981	0.972	0.954	0.877	0.755	0.611	0.097	IDT
KMM-4-1	101	42.8	4.3	36	30	0.1663	0.0695	0.0089	0.4849	0.3881	0.0028	0.985	0.983	0.977	0.957	0.939	0.87	0.766	0.519	0.31	0.174	0.025	MedExome
KMM-4-2	101	47	4.7	37.2	31	0.1725	0.071	0.0236	0.458	0.3653	0.0027	0.985	0.983	0.978	0.958	0.942	0.876	0.777	0.539	0.332	0.193	0.027	MedExome
KMM-4-3	101	45.7	4.6	36.2	31	0.1939	0.0689	0.0089	0.4588	0.3651	0.0027	0.986	0.984	0.978	0.959	0.943	0.88	0.782	0.536	0.319	0.176	0.024	MedExome
KMM_22-1	101	73	7.4	81.3	77	0.1845	0.0654	0.0014	0.572	0.4284	0.0018	0.987	0.986	0.986	0.985	0.984	0.982	0.978	0.958	0.908	0.825	0.26	IDT
KMM_29-1	101	91.9	9.3	73.2	69	0.4197	0.0535	0.0074	0.4169	0.3066	0.002	0.987	0.986	0.985	0.984	0.983	0.979	0.972	0.936	0.863	0.754	0.197	IDT

eTable 8. Rare transmitted/unphased variants in known MMD genes

a. Characteristics of four patients with mutations in *ACTA2* and *RNF213*

Proband ID	Ethnicity	Sex	Type	Class	Gene	Position (GRCh37)	AA Change	Bravo Freq	Meta SVM	CADD 1.6	Cons46aa (cons46diffcnt)
KMM_16-1	European	F	Transmitted from Father	D-Mis	*ACTA2*	10:90707044:T:C	p.I77V	.	D	22.1	IIIIIIIIII/II-II-II/IIIIII-I-IIIIIIIII/I—IIIIIII (0)
KMM_36-1	European	F	Transmitted from Mother	D-Mis	*RNF213*	17:78293080:T:C	p.C998R	$8.0x10^{-6}$	T	23.8	C-CC-CCCCC/-CC-CC-C/-CCCCCC---CCC---C-/C--------- (0)
KMM_14-1	European	M	Transmitted from Father	D-Mis	*RNF213*	17:78357556:T:C	p.I4717T	.	T	21.7	IIIIIIIIII/LIIIVI-I/IIIIIIIII-IIIIIII-/VIIILLLLL-(8)
KMM_40-1	European	M	Transmitted from Father	D-Mis	*RNF213*	17:78359416:C:A	p.T4845K	.	T	22.7	TT-MTTT-TT/TTT-TT--/-TTTTTTTT-TTTTTTTT/TT-M-T---- (2)

b. Phylogenetic conservation of reference amino acid at each variant position described in this work.

Species	*ACTA2* - p.I77	*RNF213* - p.C998	*RNF213* - p.I4717	*RNF213* - p.T4845
H. sapiens:	YPIEHGIITNWDD	AKTFEKCIIEAVS	SNPVAKIIYGDPV	LRRSLETNGEINL
M. musculus:	YPIEHGIITNWDD	SRSFEKCVIEAVS	SNPVTKIIYGDPA	LRRSLETNGEIKL
R. norvegicus:	YPIEHGIITNWDD	ARSIEKCVIEAVS	SNPVTKIIYGDPA	LRRSLETNGEIKL
X. laevis:	YPIEHGIITNWDD	GKCFEDCAIEAVR	SNPIVKIVYGDPL	LRDSVYANSDIKL
C. familiaris:	YPIEHGIITNWDD	SKCFEKCVVEAVS	SNPVARIVYKDPT	LRRSLQTNGEIKL
R. macaque:	YPIEHGIITNWDD	AKTFEKCIIEAVS	SNPVAKIIYGDPV	LRRSLETNGEINL

AA Change - Amino Acid Change; Bravo Freq - Bravo Whole Genome Sequencing Frequency; MetaSVM - Meta-analytic support vector machine is an ensamble score that predicts the tolerability of a mutation. CADD1.6 - Combined Annotated Dependent Depletion is a validated tool for scoring the deleteriousness of single nucleotide deletions, insertions, or variations. Cond46aa represents “Residues among 46 vertebrate species" as described before in Furey et. al: reference #46

eTable 9. *De novo* variant enrichment analysis for each functional class in 22 primary Moyamoya case trios

	Primary Moyamoya Cases, N = 22					
	Observed		Expected		Enrichment	*p*
	N	Rate	N	Rate		
All genes (N=19,347)						
Total	24	1.09	24.4	1.11	0.99	0.56
Syn	7	0.32	6.9	0.31	1.01	0.54
T-Mis	2	0.09	5.3	0.24	0.38	0.97
D-Mis	12	0.55	10.1	0.46	1.19	0.31
LoF	3	0.14	2.1	0.1	1.42	0.36
Protein altering	17	0.77	17.5	0.8	0.97	0.58
Protein damaging	15	0.68	12.2	0.55	1.23	0.25
High brain-expressed genes (brain expression rank ≥ 75%; N=4,522)						
Total	7	0.32	6.3	0.29	1.11	0.44
Syn	2	0.09	1.8	0.08	1.14	0.52
T-Mis	2	0.09	1	0.05	2	0.26
D-Mis	3	0.14	3	0.14	0.99	0.59
LoF	2	0.09	0.6	0.03	3.52	0.11
Protein altering	5	0.23	4.5	0.2	1.1	0.47
Protein damaging	5	0.23	3.6	0.16	1.38	0.3
Loss-of-function intolerant genes (gnomADv2.1.1 pLI ≥ 0.9; N=3,049)						
Total	9	0.41	5.8	0.26	1.54	0.14
Syn	2	0.09	1.6	0.07	1.21	0.49
T-Mis	0	0	0.8	0.04	0	1
D-Mis	5	0.23	2.9	0.13	1.7	0.18
LoF	2	0.09	0.5	0.02	3.83	0.1
Protein altering	7	0.32	4.2	0.19	1.67	0.13
Protein damaging	7	0.32	3.5	0.16	2.02	0.06
Loss-of-function-intolerant & high brain expressed genes (N=1,495)						

Total	5	0.23	2.8	0.13	1.79	0.15
Syn	2	0.09	0.8	0.04	2.58	0.18
T-Mis	0	0	0.3	0.01	0	1
D-Mis	2	0.09	1.5	0.07	1.38	0.43
LoF	1	0.05	0.3	0.01	3.88	0.23
Protein altering	3	0.14	2	0.09	1.48	0.33
Protein damaging	3	0.14	1.7	0.08	1.75	0.25

N: number of *de novo* variants; Rate: number of *de novo* variants per subject; Enrichment: ratio of observed to expected numbers of *de novo* variants; D-Mis: damaging missense variants as predicted by MetaSVM or CADD v1.6 (≥20); LoF: loss-of-function variants comprised of premature termination, frameshift or splice site mutation; High brain-expressed genes denote genes in the top quartile of expression in the developing mouse brain; Loss-of-function-intolerant genes denote genes with a pLI ≥ 0.9 in the gnomAD database (v2.1.1).

eTable 10. Damaging recessive variants in 26 MMD cases

ID	Variant Type	CHROM:POS:REF:ALT	Gene	AA change	Variant Class	MetaSVM	CADD v1.6	Bravo WGS Freq	gnomAD WES Freq	gnomAD WGS Freq	pLI score (gnomAD v2.1.1)	mis_z (gnomAD v2.1.1)
KMM-4-1	Homozygous	2:218686643:T:A	*TNS1*	p.E1027V	D-Mis	D	29.1	8.28E-04	2.00E-03	3.60E-03	0.06	0.55
CHO_0001-01	Homozygous	3:1369255:G:C	*CNTN6*	p.E400Q	D-Mis	T	25	7.09E-04	1.30E-03	5.00E-04	0.00	-3.00
KMM_13-3	Homozygous	8:121290743:A:G	*COL14A1*	p.K1136R	D-Mis	T	21.4	8.28E-04	1.00E-03	7.00E-04	0.00	1.54
CHO_0001-01	Homozygous	15:52539165:G:A	*MYO5C*	p.T643M	D-Mis	T	23.6	9.16E-04	1.10E-03	6.00E-04	0.00	0.48
KMM_3-1	Homozygous	17:14110391:G:A	*COX10*	p.R398H	D-Mis	D	24.9	1.35E-04	9.83E-05	.	0.08	-0.18
KMM_13-3	compound heterozygous	17:48264049:C:T/17:48276622:G:T	*COL1A1*	p.A1256T/p.P146T	D-Mis/D-Mis	D/D	26/20.1	2.31E-04/ 3.98E-05	2.00E-04/ 9.78E-06	6.47E-05/.	1.00	3.53
KMM_13-3	compound heterozygous	3:89259601:G:A/3:89456470:C:T	*EPHA3*	p.E249K/p.A549V	D-Mis/D-Mis	T/T	24.7/20.5	7.96E-06/ 1.04E-04	1.67E-05/ 2.00E-04	./6.47E-05	0.00	0.65
KMM_13-3	compound heterozygous	2:168097184:T:C/2:168100235:C:T	*XIRP2*	p.L360P/p.P778L	D-Mis/D-Mis	D/T	24.2/26.4	1.59E-05/ 2.39E-05	4.06E-06/0	./3.23E-05	0.00	-2.40
KMM_14-1	compound heterozygous	2:179479288:A:G/2:179647563:C:T	*TTN*	p.I16318T/p.V1024I	D-Mis/D-Mis	D/T	21.8/20.7	9.64E-04/ 2.39E-05	1.20E-03/ 2.48E-05	1.20E-03/ 3.67E-05	0.00	-1.10
KMM_21-1	compound heterozygous	17:10301792:T:C/17:10323469:G:A	*MYH8*	p.I1383V/p.R26W	D-Mis/D-Mis	T/D	22.5/23.9	3.74E-04/ 7.96E-06	6.00E-04/ 8.13E-06	4.00E-04/.	0.00	0.35
KMM_35-1	compound heterozygous	7:122028773:C:T/7:122076416:GT:G	*CADPS2*	p.E856K/p.T671fs	D-Mis/frameshift -deletion	T/.	32/.	7.88E-04/ 1.35E-04	1.50E-03/ 2.00E-04	3.00E-04/ 3.24E-05	0.37	1.09
KMM_36-1	compound heterozygous	1:22155342:C:T/1:22198756:G:A	*HSPG2*	p.G4075S/p.L1382F	D-Mis/D-Mis	D/D	26.2/27	1.59E-05/ 7.96E-06	2.02E-05/ 1.94E-05	3.24E-05/.	0.00	1.14
KMM_40-1	compound heterozygous	16:67985888:C:T/16:68002521:C:T	*SLC12A4*	p.A324T/p.G13R	D-Mis/D-Mis	T/T	20.4/22.9	./4.14E-04	./6.00E-04	./3.00E-04	0.00	1.35
KMM_42-1	compound heterozygous	14:100595081:A:G/14:100604139:G:A	*EVL*	p.Q238R/p.R365K	D-Mis/D-Mis	T/T	21.3/35	7.96E-06/.	./.	./3.23E-05	1.00	1.48
KMM_42-1	compound heterozygous	16:68912084:A:G/16:69008053:C:T	*TANGO6*	p.T399A/p.R942X	D-Mis/stopgain	T/.	22.1/40	4.78E-05/ 3.19E-05	4.89E-05/ 1.63E-05	./3.23E-05	0.00	0.45
KMM_42-1	compound heterozygous	6:31733817:T:C/6:31743113:G:A	*VWA7*	p.N781S/p.Q203X	D-Mis/stopgain	T/.	25.4/36	7.96E-06/ 7.96E-06	./4.18E-06	3.23E-05/.	0.00	1.01
KMM_45-1	compound heterozygous	3:142259830:T:G/3:142261533:T:C	*ATR*	p.H1166P/p.S1142G	D-Mis/D-Mis	T/T	26.1/21.9	1.59E-05/ 7.88E-04	4.47E-05/ 6.00E-04	./4.00E-04	0.00	4.36
KMM_50-1	compound heterozygous	2:26667804:C:T/2:26673482:A:G	*DRC1*	p.L462F/p.D541G	D-Mis/D-Mis	T/T	23/29.6	./7.96E-05	8.25E-06/ 1.00E-04	./.	0.00	-0.05
KMM_50-1	compound heterozygous	2:241658469:G:A/2:241702220:T:C	*KIF1A*	p.A1640V/p.S703G	D-Mis/D-Mis	T/T	24.3/23.1	2.87E-04/.	3.00E-04/.	4.00E-04/.	1.00	5.16
KMM_50-1	compound heterozygous	16:3452194:C:T/16:3454579:C:T	*ZNF174*	p.L64F/p.R186W	D-Mis/D-Mis	T/T	23.7/21.2	./3.19E-05	./1.65E-05	./.	0.01	0.56

AA Change - Amino Acid Change; Bravo WGS Freq - Bravo Whole Genome Sequencing Frequency; gnomAD WES Freq - Gnome Aggregation Database Whole Exome Sequence Frequency; MetaSVM - Meta-analytic support vector machine is an ensemble score that predicts the tolerability of a mutation. CADD1.6 - Combined Annotated Dependent Depletion is a validated tool for scoring the deleteriousness of single nucleotide deletions, insertions, or variations. pLI is a gene-wide constraint metric that estimates the probability of being intolerant to loss-of-function mutations. mis_Z is a gene-wide constraint metric that estimates the probability of being intolerant to missense mutations.

eMethods. Detailed Methods

Kinship analysis

Relationship between proband and parents was estimated using the pairwise identity-by-descent (IBD) calculation in PLINK [60]. The IBD sharing between the proband and parents in all trios is between 45% and 55%.

Ruling out duplicated samples

To identify subjects who could have been recruited multiple times in case cohorts, control cohorts, or both, we calculated the overlap of high-confidence rare variants (MAF = 0% in ExAC, gnomAD, and Bravo) between each pair of individuals. For pairs that share ≥ 80% of rare variants, the sample with greater sequence coverage was kept in the analysis and the other discarded.

Principal component analysis

To determine the genetic ancestry of each subject, we used the EIGENSTRAT software [1] to analyze tag single-nucleotide polymorphisms (SNPs) in cases, controls, and HapMap subjects as described [2].

***De novo* variant expectation model**

Because the MMD trios were captured by three different reagents (i.e., Nimblegen V2, MedExome. and IDT), we took the union of all bases covered by different capture reagents and generated a Browser Extensible Data (BED) file representing a unified capture for all trios. We used bedtools (v2.27.1) to extract sequences from the BED file [3]. A total of 1,798 control trios consisting of unaffected siblings and parents of patients with autism spectrum disorder from the Simons Simplex Collection were analyzed in parallel [4]. We then applied a sequence context-based method to calculate the probability of observing a DNV for each base in the coding region, adjusting for sequencing depth in each gene as described previously [5]. Briefly, for each base in the exome, the probability of observing every tri-nucleotide mutating to other tri-nucleotide was determined. ANNOVAR (v2015Mar22) was used to annotate the consequence of each possible substitution. RefSeq was used to annotate variants (based on the file "hg19_refGene.txt" provided by ANNOVAR). For each gene, the coding consequence of each potential substitution was summed for each functional class (synonymous, missense, canonical splice site, frameshift insertions/deletions, stop-gain, stop-loss, start-lost) to determine the gene-specific variant probabilities [5]. The probability of a frameshift mutation was determined by multiplying the probability of a stop-gain mutation by 1.25, as described previously [5]. In-frame insertions or deletions are not accounted for by the model and were not considered in the downstream statistical analyses. To align with ANNOVAR annotations, analysis was limited to variants that were located in the exonic or canonical splice site regions and were not annotated as "unknown" by ANNOVAR. Following the inclusion criteria, we identified potential coding mutations and generated gene-specific mutation probabilities for 19,347 unique genes. Owing to differences in the exome capture kits, DNA sequencing platforms, and variable sequencing

coverage used to study case and control cohorts, respectively, separate *de novo* probability tables were generated for cases and controls.

De novo mutation enrichment analysis

The burden of DNVs in MMD cases was determined using the denovolyzeR package as previously described [2]. Briefly, the expected number of DNVs in the case cohort across each functional class was calculated by taking the sum of each functional class-specific probability multiplied by the number of probands in the study x 2 (diploid genomes). Then, the expected number of DNVs across functional classes was compared to the observed number in each study using a Poisson test [5]. Gene-set enrichment analyses only considered variants observed or expected in genes within the specified gene set (i.e., high brain-expressed, LoF-intolerant).

Prediction of variant impact on mDia1 structure

The mDia1 protein sequence (UniProt O60610) was used to query the Protein DataBank for structural homologs using MolSoft ICM-Pro ver 1.8.7c [7]. The search identified two structures of mDia1 from mouse (PDB id 2BNX[76] and 1V9D[77]). The homology model was built in ICM-Pro based on a 96.1% sequence identity with the mouse template over the resolved residues of mouse mDia1 separately. The ΔΔG value was calculated using the Molmechanics tool of ICM-Pro and used to predict the impact of the variant on protein structure [8]; this is mathematically expressed by the equation:

$\Delta G = \Delta G_{misfolded} - \Delta G_{unfolded}$

In this equation, $\Delta G_{unfolded}$ refers to the sum of the energies attributed to the individual residues before the mutation, while $\Delta G_{misfolded}$ refers to the sum of the energies attributed to the individual residues after the mutation. A positive $\Delta\Delta G$ indicates a destabilization effect on protein structure caused by the point mutation. More detailed information can be found in **eFigure 2**.

eReferences

1) Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics.* 2006;38(8):904-909.

2) Jin SC, Homsy J, Zaidi S, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nature genetics.* 2017;49(11):1593-1601.

3) Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841-842.

4) Krumm N, Turner TN, Baker C, et al. Excess of rare, inherited truncating mutations in autism. *Nature genetics.* 2015;47(6):582-588.

5) Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nature genetics.* 2014;46(9):944-950.

6) Ware JS, Samocha KE, Homsy J, Daly MJ. Interpreting de novo Variation in Human Disease Using denovolyzeR. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al].* 2015;87:7 25 21-15

7) Abagyan R, Totrov M, Kuznetsov D. Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J Comput Chem.* 1994;15(5):488-506.

8) Otomo T, Otomo C, Tomchick DR, Machius M, Rosen MK. Structural basis of Rho GTPase-mediated activation of the formin mDia1. *Mol Cell.* 2005;18(3):273-281.

