The American Journal of Human Genetics, Volume *100*

Supplemental Data

CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions

Darci T. Butcher, Cheryl Cytrynbaum, Andrei L. Turinsky, Michelle T. Siu, Michal Inbar-Feigenberg, Roberto Mendoza-Londono, David Chitayat, Susan Walker, Jerry Machado, Oana Caluseriu, Lucie Dupuis, Daria Grafodatskaya, William Reardon, Brigitte Gilbert-Dussardier, Alain Verloes, Frederic Bilan, Jeff M. Milunsky, Raveen Basran, Blake Papsin, Tracy L. Stockley, Stephen W. Scherer, Sanaa Choufani, Michael Brudno, and Rosanna Weksberg

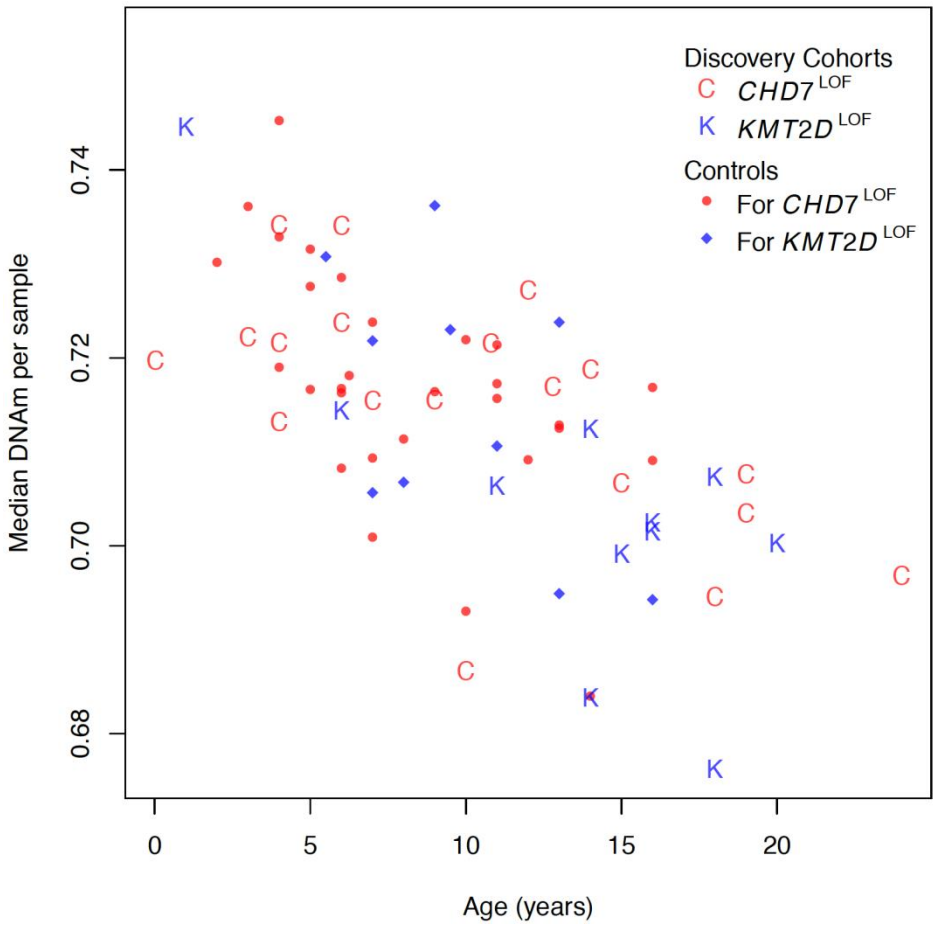

Figure S1. Age distribution of the *CHD7*^{LOF} and *KMT2D*^{LOF} samples in the Discovery cohorts. The age in years of the *CHD7*^{LOF} (red C), *CHD7*^{LOF} matching controls, (red circles), *KMT2D*^{LOF} (blue K) and *KMT2D* ^{LOF} matching controls (blue diamonds) are plotted (X-axis) against the median DNAm per sample (Y-axis).

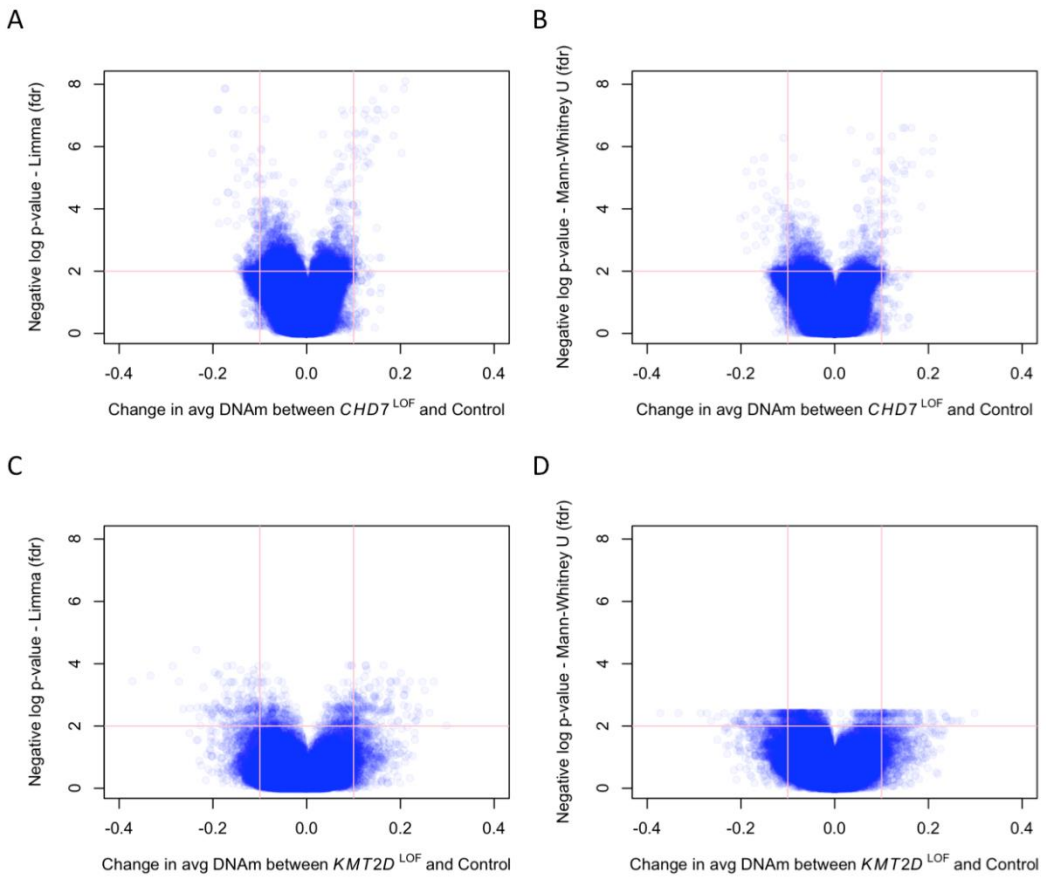

Figure S2. Volcano plots shows the average gain/loss of DNAm using limma regression and Mann-Whitney U tests. Average gain/loss of DNAm (X-axis) at all CpG sites are plotted against the statistical significance of such change after FDR correction for multiple testing (Y-axis, log-scale). Individual CpGs are indicated as semi-transparent circles. A) Average DNAm change in 19 *CHD7*^{LOF} samples with respect to 29 matching controls, with the statistical significance derived from *limma* regression test. B) Average DNAm change in 19 *CHD7*^{LOF} samples with respect to 29 matching controls, with the statistical significance derived from non-parametric Mann-Whitney U test. C) Average DNAm change in 11 *KMT2D*^{LOF} samples with respect to 11 matching controls, with the statistical significance derived from limma regression test. D) Average DNAm change in 11 *KMT2D*^{LOF} samples with respect to 11 matching controls, with the statistical significance derived from non-parametric Mann-Whitney U test. The FDR-adjusted significance level α=0.01 is shown as pink horizontal lines. The effect-size threshold of 10% DNAm difference is shown as pink vertical lines. Relatively few CpG sites have DNAm change over 10% by magnitude while also exhibiting the statistical significance level α=0.01.

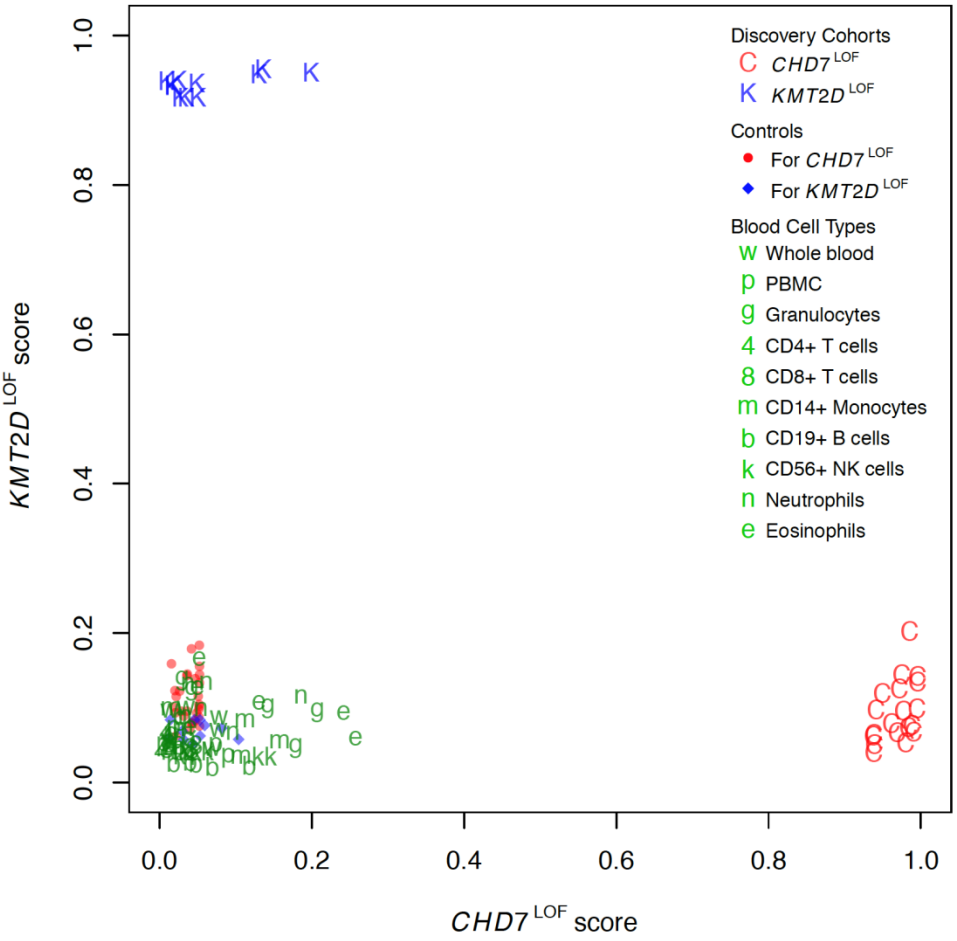

Figure S3. Independence of blood cell type composition from the *CHD7*^{LOF} and *KMT2D*^{LOF} DNAm signatures. We extracted DNAm data from Reinius *et al*., 2012 (GEO: GSE35069) representing 6 samples from each of the following cell types: whole blood, peripheral blood mononuclear cells (PBMC), granulocytes, neutrophils, eosinophils, as well as isolated cell populations (CD4+ T cells, CD8+ T cells, CD56+ NK cells, CD19+ B cells, CD14+ monocytes). All cell-type samples received low scores from the two predictive models built for *CHD7*^{LOF} and *KMT2D*^{LOF} DNAm signatures (X-axis and Y-axis, respectively; compare to Figure 2 in the main text), demonstrating that the predictions are not biased by any of the cell types, and are therefore robust to cell-subtype composition.

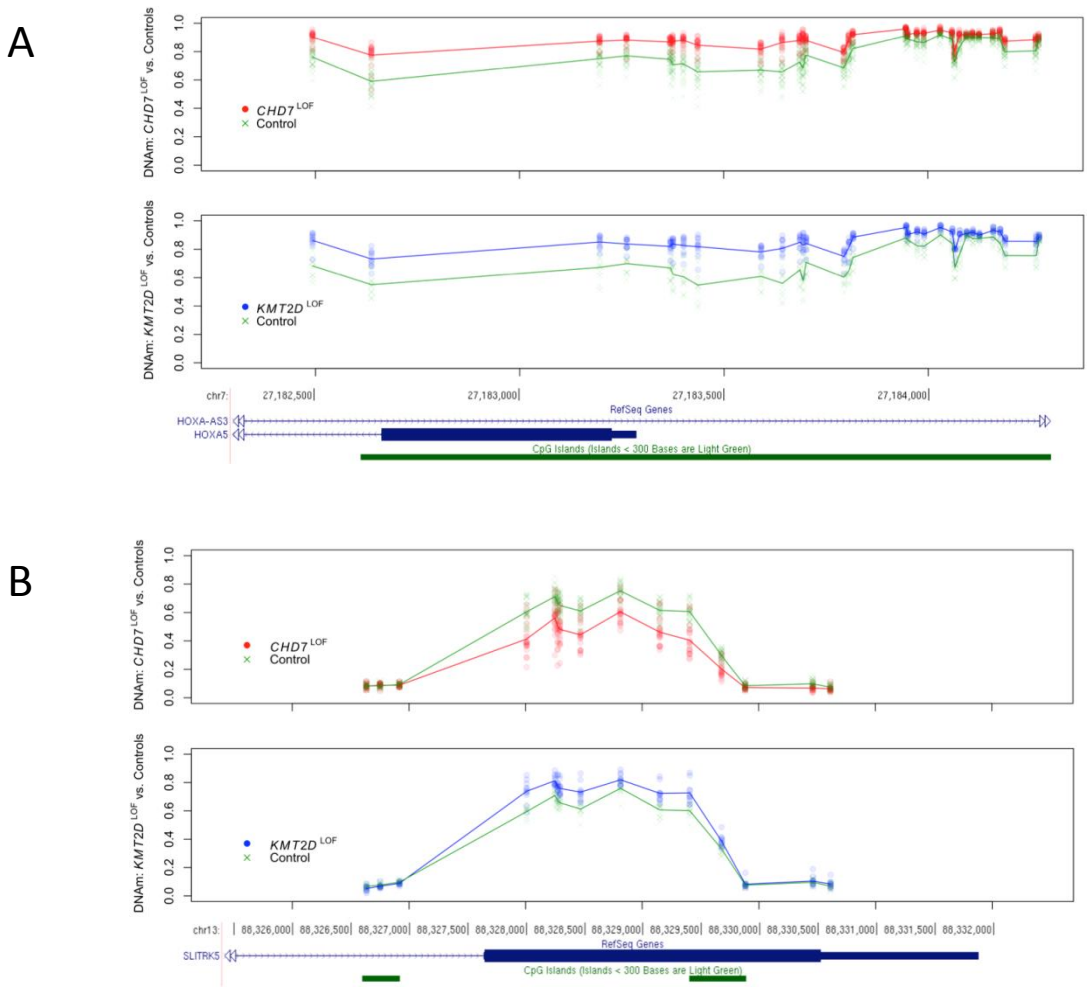

Figure S4. Overlap in differentially methylated regions (DMR) associated with the *CHD7*^{LOF} and *KMT2D*^{LOF} mutations. A) The DMR near the promoter of *HOXA5* shows a gain of methylation in both *CHD7*^{LOF} and *KMT2D*^{LOF} Discovery Cohorts. DNAm values are visualized as semi-transparent circles or crosses, arranged vertically for each CpG in the region spanning the candidate gene promoter. The X-axis shows the position of CpG site along the chromosome. The top panel shows the methylation level (Y-axis) for individual CpGs in each *CHD7*^{LOF} individual (red circles) and each matching control sample (green crosses). Also shown are group-average DNAm levels for the *CHD7*^{LOF} cohort (red line) and controls (green line). Similarly, the bottom panel shows the DNAm level (Y-axis) for individual CpGs in each *KMT2D*^{LOF} individual (blue circles) and each matching controls (green crosses). Also shown are the corresponding group averages (blue and green lines, respectively). The DMRs correspond to the gap between the lines showing the group-average DNAm levels. B) The DMR near the promoter of *SLITRK5* shows a loss of DNAm in *CHD7*^{LOF} but a gain of DNAm in *KMT2D*^{LOF} individuals.

Table S1. Molecular data for *CHD7*^{LOF} and *CHD7* sequence variants.

a) *CHD7* Loss of Function Mutations (*CHD7*^{LOF}) Analyzed to Derive the DNAm Signature

Sample ID	mutation DNA	mutation protein	coding effect/splice site mutation
CHD7-1	c.7282C>T	p.Arg2428*	nonsense
CHD7-2	c.3526C>T	p.Gln1176*	nonsense
CHD7-3	c.934C>T	p.Arg312*	nonsense
CHD7-4	c562C>T	p.Gly188*	nonsense
CHD7-5	c.1327delATGGG	p.Met443Asnfs*130	frameshift
CHD7-6	c.2504_2508delATCTT	p.Tyr835Serfs*14	frameshift
CHD7-7	c.1990G>T	p.Glu664*	nonsense
CHD7-8	c.3377dupT	p.Leu1126Phefs*46	frameshift
CHD7-9	c.2585delA	p.Leu862Serfs*26	frameshift
CHD7-10	c.2905_2906del	p.Arg969Glyfs*25	frameshift
CHD7-11	c.7636G>T	p.Glu2546*	nonsense
CHD7-12	c.361delC	p.Gly121Valfs*90	frameshift
CHD7-13	c.2504_2508delATCTT	p.Arg835Serfs*14	frameshift
CHD7-14	c.7717-7720del	p.Gln2537*	nonsense
CHD7-15	c.5458C>T	p.Arg1820*	nonsense
CHD7-16	exon 1 deletion	ND	exon 1 deletion
CHD7-17	c.5405-17G>A	ND	splice site mutation
CHD7-18	c.5405-7G>A	ND	splice site mutation
CHD7-19	c.2097-1G>A	ND	splice site mutation

b) *CHD7* Sequence Variants Classified Using the *CHD7*^{LOF} DNAm Signature

Sample ID	mutation DNA	mutation protein	coding effect/splice site mutation
CHD7-20	c.6322G>T	p.Gly2108Trp	missense
CHD7-21	c.3746G>A	p.Arg1249Gln	missense
CHD7-22	c2751G>A	p.= (p.Thr917Thr)	synonymous
CHD7-23	c.-15G>A	ND	splice site mutation
CHD7-24	c.4225G>A	p.Val1409Met	missense
CHD7-25	c.5436C>G	p.Asp1812Glu	missense
CHD7-26	c.5633A>G	p.Asp1878Gly	missense
CHD7-27	c.5848G>A	p.Ala1950Thr	missense
CHD7-28	c.6304G>T	p.Val2102Phe	missense
CHD7-29	c.3566G>A	p.Arg1189His	missense
CHD7-30	intron4:c2238+1del	ND	splice site mutation
CHD7-31	c.2049_2050insAAAGCA	p.Ala685_Lys686dup	missense
CHD7-32	c.6377A>T	p.Asp2126Val	missense

Abbreviations

ND not determined

Table S2. Molecular data for *KMT2D*^{LOF} and *KMT2D* sequence variants.

a) *KMT2D* Loss of Function Mutations (*KMT2D*^{LOF}) Analyzed to Derive the DNAm Signature

Sample ID	mutation DNA	mutation protein	Coding Effect
KMT2D-1	c.15061C>T	p.Arg5021*	nonsense
KMT2D-2	c.16318delG	p.Glu5440Argfs*16	frameshift
KMT2D-3	c15030dupA	p.Glu5011Argfs*13	frameshift
KMT2D-4	c.8172_8173delC	p.Phe2724Glnfs*5	frameshift
KMT2D-5	c.6595delT	p.Tyr2199Ilefs*65	frameshift
KMT2D-6	c.14055-14056delCA	p.His4685Glnfs*4	frameshift
KMT2D-7	c.6295C>T	p.Arg2099*	nonsense
KMT2D-8	c.4135_4136delA	p.Met1379Valfs*52	frameshift
KMT2D-9	c.12592C>T	p.Arg4198*	nonsense
KMT2D-10	c.4135_4136delA	p.Met1379Valfs*52	frameshift
KMT2D-11	c.11710C>T	p.Gln3904*	nonsense

b) *KMT2D* Sequence Variants Classified Using the *KMT2D*^{LOF} DNAm Signature

Sample ID	mutation DNA	mutation protein	Coding Effect/Splice Site Mutation
KMT2D-12	c.15143G>A	p.Arg5048His	missense
KMT2D-13	c.12028T>C	p.Ser4010Pro	missense
KMT2D-14	c.16522-5_16522-4delTT	ND	splice site mutation
KMT2D-15	c.15910A>G	p.Ile5304Val	missense
KMT2D-16	c.15659G>A	p.Arg5220His	missense
KMT2D-17	c.10256A>G	p.Asp3419Gly	missense
KMT2D-18	c.8974G>A	p.Glu2992Lys	missense
KMT2D-19	c.8831A>G	p.Asn2944Ser	missense
KMT2D-20	c.832G>A	p.Ala278Thr	missense
KMT2D-21	c.682C>G	p.Arg228Gly	missense

Abbreviations

ND not determined

Table S5. Validation cohort and classification utilizing *CHD7*^{LOF} and *KMT2D*^{LOF} signatures.

a) Validation Samples with *CHD7* Mutations Classified Using the *CHD7*^{LOF} DNAm Signature

SampleID	Signature (positive/negative)	Gene	Mutation	Protein Change	Coding Effect/Splice Site Mutation
CHD7-33	negative	*CHD7*	c.4851T>G	p.=(p.Gly1617Gly)	synonymous
CHD7-34	positive	*CHD7*	c.5097dupA	p.Ala1700Serfs*37	frameshift
CHD7-35	positive	*CHD7*	c.8791 G>A	p.Val2931Met	missense
CHD7-36	positive	*CHD7*	c.799G>T	p.Glu267*	nonsense
CHD7-37	negative	*CHD7*	c.8802C>G	p.Ser2934Arg	missense
CHD7-38	positive	*CHD7*	c.2516_2518delAGT	p.Gln839_Trp840delinsArg	in-frame deletion
CHD7-39	negative	*CHD7*	c.8759G>C	p.Gly2920Ala	missense
CHD7-40	positive	*CHD7*	c.1312C>T	p.Gln438*	nonsense
CHD7-41	positive	*CHD7*	c.6322G>A	p.Gly2108Arg	missense
CHD7-42	positive	*CHD7*	c.127A>G	p.Ile43Val	missense
CHD7-43	negative	*CHD7*	c.1405A>G	p.Arg469Gly	missense
CHD7-44	positive	*CHD7*	c.7763A>G	p.Asn2588Ser	missense
CHD7-45	positive	*CHD7*	c.3871A>C	p.Lys1291Gln	missense
CHD7-46	positive	*CHD7*	c.5050G>A	p.Gly1684Ser	missense
CHD7-47	positive	*CHD7*	c.5210+3A>G	ND	splice site mutation
CHD7-48	positive	*CHD7*	c.6193C>T	p.Arg2065Cys	missense
CHD7-49	positive	*CHD7*	c.3762T>A	p.His1254Gln	missense
CHD7-50	negative	*CHD7*	c.583C>T	p.Arg195Cys	missense
CHD7-51	positive	*CHD7*	c.4087delC	p.Leu1363Serfs*9	frameshift
CHD7-52	positive	*CHD7*	c.2498+1G>T	ND	splice site mutation
CHD7-53	positive	*CHD7*	c.1918delG	p.Gly640Lysfs*71	frameshift
CHD7-54	negative	*CHD7*	c.1562C>T	p.Pro521Leu	missense
CHD7-55	positive	*CHD7*	c.604 C>T	p.Gln202*	nonsense
CHD7-56	positive	*CHD7*	c.4393C>T	p.Arg1465*	nonsense
CHD7-57	positive	*CHD7*	c.5666-9C>G	ND	splice site mutation
CHD7-58	positive	*CHD7*	c.5029C>T	p.Arg1677*	nonsense
CHD7-59	negative	*CHD7*	c.1797_1799delGAA	p.Lys602del	in-frame deletion
CHD7-60	positive	*CHD7*	c.3177T>G	p.Tyr1059*	nonsense
CHD7-61	positive	*CHD7*	c.2839C>T	p.Arg947*	nonsense
CHD7-62	positive	*CHD7*	c.5429G>C	p.Arg1810Pro	missense
CHD7-63	positive	*CHD7*	c.4361_4362delAG	p.Gln1454Profs*21	frameshift
CHD7-64	positive	*CHD7*	c.2362C>T	p.Gln788*	nonsense
CHD7-65	negative	*CHD7*	c.5827C>T	p.Arg1943Trp	missense
CHD7-66	negative	*CHD7*	c.317A>G	p.His106Arg	missense
CHD7-67	negative	*CHD7*	c.6529G>A	p.Gln2177Lys	missense
CHD7-68	positive	*CHD7*	c.8507delC	p.Pro2836Argfs*53	frameshift
CHD7-69	positive	*CHD7*	c.3655C>T	p.Arg1219*	nonsense
CHD7-70	positive	*CHD7*	c.6356A>G	p.Asp2119Gly	missense
CHD7-71	positive	*CHD7*	c.1141_1142delAT	p.Met381Alafs*23	frameshift
CHD7-72	positive	*CHD7*	c.3082A>G	p.Ile1028Val	missense

b) Validation Samples with *KMT2D* or *KDM6A* Mutations Classified Using the *KMT2D*^{LOF} DNAm Signature

SampleID	Signature (positive/negative)	Gene	Mutation	Protein Change	Coding Effect/Splice Site Mutation
KDM6A-1	positive	*KDM6A*	c.2668_2669dupTA	p.Pro891Thrfs*8	frameshift
KMT2D-22	positive	*KMT2D*	c.11158C>T	p.Gln3720*	nonsense
KMT2D-23	negative	*KMT2D*	c.15587T>A	p.Met5196Lys	missense
KMT2D-24	negative	*KMT2D*	c.11150A>C	p.Gln3717Pro	missense
KMT2D-25	positive	*KMT2D*	c.16521+1G>T	ND	splice site mutation
KMT2D-26	positive	*KMT2D*	c.1940dupC	p.Pro648Thrfs*2	frameshift
KMT2D-27	negative	*KMT2D*	c.11578_11580dupCAG	p.Gln3863dup	in-frame duplication
KMT2D-28	negative	*KMT2D*	c.10909C>A	p.Pro3637Thr	missense
KMT2D-29	positive	*KMT2D*	c.15088C>T	p.Arg5030Cys	missense
KMT2D-30	positive	*KMT2D*	c.5135delA	p.Lys1712Argfs*10	frameshift
KMT2D-31	negative	*KMT2D*	c.12662_12664dupAGC	p.Gln4221dup	in-frame duplication
KMT2D-32	positive	*KMT2D*	c.16052G>A	p.Arg5351Gln	missense
KMT2D-33	positive	*KMT2D*	c.11203C>T	p.Gln3735*	nonsense
KMT2D-34	negative	*KMT2D*	c.2334C>G	p.Cys778Trp	missense
KMT2D-35	negative	*KMT2D*	c.14731_14733delCCT	p.Pro4911del	in-frame deletion
KMT2D-36	positive	*KMT2D*	c.15536G>A	p.Arg5179His	missense
KMT2D-37	positive	*KMT2D*	c.15626G>T	p.Gly5209Val	missense

Abbreviations

n/a	not applicable
ND	not determined
VUS	Variants of unknown significance

Variant Designation	Align GVGD	SIFT (score)	Mutation Taster (p-value)	PolyPhen-2 (score)	ESEfinder Splicing Predictions
VUS	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Tolerated (0.21)	Disease Causing (0.998)	probably damaging (0.998)	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Tolerated (0.15)	polymorphism (0.998)	benign (0.001)	n/a
VUS	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Deleterious (0)	Disease Causing (1)	probably damaging (0.999)	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	Class C65	Deleterious (0)	Disease Causing (1)	probably damaging (1)	n/a
VUS	Class C0	Tolerated (0.23)	Polymorphism (0.998)	benign (0.001)	n/a
VUS	Class C0	Tolerated (0.09)	Disease Causing (0.99)	probably damaging (0.963)	n/a
VUS	Class C0	Tolerated (0.72)	Disease Causing (0.999)	benign (0.032)	n/a
VUS	Class C45	Deleterious (0)	Disease Causing (1)	probably damaging (1)	n/a
Pathogenic	ClassC55	Deleterious (0)	Disease Causing (1)	probably damaging (0.990)	n/a
Pathogenic	n/a	n/a	n/a	n/a	no ESE prediction disruption
VUS	Class C45	Deleterious (0)	Disease Causing (1)	probably damaging (1)	n/a
VUS	Class C15	Deleterious (0)	Disease Causing (1)	probably damaging (0.988)	n/a
VUS	Class C15	Deleterious (0.03)	Disease Causing (1)	probably damaging (0.997)	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	unclear prediction
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Tolerated (0.36)	Disease Causing (1)	benign (0.072)	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	n/a	n/a	n/a	n/a	disrupts SRp40 site
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C65	Deleterious (0)	Disease Causing (1)	probably damaging (1)	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Deleterious (0)	Disease Causing (1)	probably damaging (0.998)	n/a
VUS	Class C0	Tolerated (0.19)	Disease Causing (1)	possibly damaging (0.634)	n/a
VUS	Class C0	Tolerated (0.9)	Disease Causing (0.963)	benign (0.006)	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Tolerated (0.06)	Disease Causing (1)	probably damaging (1)	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
Pathogenic	Class C25	Deleterious (0)	Disease Causing (1)	possibly damaging (0.752)	n/a

Variant Designation	Align GVGD	SIFT (score)	Mutation Taster(p-value)	PolyPhen-2 (score)	ESEfinder Splicing Predictions
Pathogenic	n/a	n/a	n/a	n/a	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Deleterious (0.01)	Disease Causing (1)	probably damaging (0.986)	n/a
VUS	Class C0	Tolerated (0.22)	Disease Causing (1)	probably damaging (0.999)	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	disrupts a SF2/ASFsite
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Tolerated (0.18)	polymorphism (0.506)	possibly damaging (0.900)	n/a
Pathogenic	Class C0	Deleterious (0)	Disease Causing (1)	possibly damaging (1)	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	n/a	n/a	n/a	n/a	n/a
Pathogenic	Class C0	Deleterious (0.02)	Disease Causing (1)	probably damaging (1)	n/a
Likely Pathogenic	n/a	n/a	n/a	n/a	n/a
VUS	Class C0	Tolerated (0.08)	polymorphism (1)	possibly damaging (0.758)	n/a
VUS	n/a	n/a	n/a	n/a	n/a
Pathogenic	Class C0	Deleterious (0)	disease causing (1)	possible damaging (0.840)	n/a
VUS	Class C0	Deleterious (0)	Disease Causing (1)	probably damaging (1)	n/a

Table S14. Sequences of primers used for sodium bisulfite pyrosequencing validation.

		PCR product size
SLITRK5		170bp
SLITRK5F	TGGGAAATTGTATTTGTTGTAGGTGT	
SLITRK5R	ACTCCACAACTTTATCCATATACTAC	
SLITRK5S	ACTACAAAAACCCCAC	
MYOF1		183bp
MYOF1F	GATTTATTGGAGTTTTTGGGTAGT	
MYOF1R	CCCCCAAACTTTCTTCTCT	
MYOF1S	GAAGTAGAGGGAGAAGGT	
FOXP2		248bp
FOXP2F	AGAAAGATTATGGTAAGTATGTTGGTTTAG	
FOXP2R	CCACCATCAAACAACTATTTACAACAA	
FOXP2S	TGATTAAATGTTGATTTTGTGTA	
HOXA5		160bp
HOXA5-F4	TGAATTATGGAAATGATTGGGATATGTAT	
HOXA5-R4	TCCACCCAACTCCCCTATTA	
HOXA5-S4	AGGTATTTAAATATGGGGT	