An *HNRNPK*-specific DNA methylation signature makes sense of missense variants and expands the phenotypic spectrum of Au-Kline syndrome

Authors

Sanaa Choufani, Vanda McNiven, Cheryl Cytrynbaum, ..., Antonie D. Kline, P.Y. Billie Au, Rosanna Weksberg

Correspondence

billie.au@albertahealthservices.ca (P.Y.B.A.), rweksb@sickkids.ca (R.W.)

This study reports the identification of a DNA methylation signature for *HNRNPK* and demonstrates an expanded role for DNA methylation signatures. We show that the DNA methylation signature can be utilized not only to assist with variant interpretation but also to further delineate the phenotypic spectrum of Au-Kline syndrome.

Choufani et al., 2022, The American Journal of Human Genetics *109*, 1867–1884
October 6, 2022

https://doi.org/10.1016/j.ajhg.2022.08.014

CellPress

An *HNRNPK*-specific DNA methylation signature makes sense of missense variants and expands the phenotypic spectrum of Au-Kline syndrome

Sanaa Choufani,[1,48] Vanda McNiven,[2,3,48] Cheryl Cytrynbaum,[1,2] Maryam Jangjoo,[1] Margaret P. Adam,[4] Hans T. Bjornsson,[5,6] Jacqueline Harris,[7] David A. Dyment,[8] Gail E. Graham,[9] Marjan M. Nezarati,[10] Ritu B. Aul,[11] Claudia Castiglioni,[12] Jeroen Breckpot,[13] Koen Devriendt,[13] Helen Stewart,[14] Benito Banos-Pinero,[15] Sarju Mehta,[16] Richard Sandford,[16] Carolyn Dunn,[16] Remi Mathevet,[17] Lionel van Maldergem,[17] Juliette Piard,[17] Elise Brischoux-Boucher,[17] Antonio Vitobello,[18,19] Laurence Faivre,[18,20] Marie Bournez,[18,20] Frederic Tran-Mau,[18,19] Isabelle Maystadt,[21]

(Author list continued on next page)

Summary

Au-Kline syndrome (AKS) is a neurodevelopmental disorder associated with multiple malformations and a characteristic facial gestalt. The first individuals ascertained carried *de novo* loss-of-function (LoF) variants in *HNRNPK*. Here, we report 32 individuals with AKS (26 previously unpublished), including 13 with *de novo* missense variants. We propose new clinical diagnostic criteria for AKS that differentiate it from the clinically overlapping Kabuki syndrome and describe a significant phenotypic expansion to include individuals with missense variants who present with subtle facial features and few or no malformations. Many gene-specific DNA methylation (DNAm) signatures have been identified for neurodevelopmental syndromes. Because *HNRNPK* has roles in chromatin and epigenetic regulation, we hypothesized that pathogenic variants in *HNRNPK* may be associated with a specific DNAm signature. Here, we report a unique DNAm signature for AKS due to LoF *HNRNPK* variants, distinct from controls and Kabuki syndrome. This DNAm signature is also identified in some individuals with *de novo HNRNPK* missense variants, confirming their pathogenicity and the phenotypic expansion of AKS to include more subtle phenotypes. Furthermore, we report that some individuals with missense variants have an "intermediate" DNAm signature that parallels their milder clinical presentation, suggesting the presence of an epi-genotype phenotype correlation. In summary, the AKS DNAm signature may help elucidate the underlying pathophysiology of AKS. This DNAm signature also effectively supported clinical syndrome delineation and is a valuable aid for variant interpretation in individuals where a clinical diagnosis of AKS is unclear, particularly for mild presentations.

Introduction

Recent advances in gene sequencing technologies have led to the molecular characterization of many rare neurodevelopmental disorders (NDDs). In turn, this has led to routine use of targeted and genome-wide sequencing for clinical diagnostics for NDDs. Significant knowledge gaps still hamper diagnosis in many individuals with NDDs because of (1) the significant number of variants of uncertain significance (VUSs) reported by molecular diagnostic

[1]Genetics and Genome Biology Program, Research Institute, the Hospital for Sick Children, Toronto, ON M5G 1X8, Canada; [2]Division of Clinical and Metabolic Genetics, Department of Pediatrics, the Hospital for Sick Children, University of Toronto, Toronto, ON M5G 1X8, Canada; [3]Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, ON M5T 3L9, Canada; [4]Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98105, USA; [5]McKusick-Nathans Institute of Genetic Medicine, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [6]Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland; [7]Kennedy Krieger Institute, Baltimore, MD 21205, USA; [8]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON K1H 5B2, Canada; [9]Division of Genetics, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, ON K1H 8L1, Canada; [10]Genetics Program, North York General Hospital, Toronto, ON M2K 1E1, Canada; [11]Mackenzie Health, Richmond Hill, ON L4C 4Z3, Canada; [12]Departmento de Neurologica Pediatrica, Clinica Las Condes, Santiago 7591046, Chile; [13]Center for Human Genetics, University Hospitals Leuven, Leuven 3000, Belgium; [14]Oxford Centre for Genomic Medicine, Nuffield Orthopaedic Centre, Oxford University Hospitals, NHS Foundation Trust, Headington, Oxford OX3 7HE, UK; [15]Oxford Regional Genetics Laboratories, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 7LE, UK; [16]Department of Clinical Genetics, Addenbrookes Hospital, Cambridge CB2 0QQ, UK; [17]Centre de génétique humaine, CHU Besançon, Université de Bourgogne Franche-Comté, Besançon 25000, France; [18]Inserm UMR1231 GAD, Génétique des Anomalies du Développement, Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement, CHU Dijon Bourgogne - Université de Bourgogne, Dijon 21079, France; [19]UF diagnostic génomiques et maladies rares et FHU TRANSLAD, Centre Hospitalier Universitaire de Dijon-Bourgogne, Dijon 21079, France; [20]Centre de référence Anomalies du Développement et syndromes malformatifs et FHU TRANSLAD, Centre Hospitalier Universitaire de Dijon-Bourgogne, Dijon 21079, France; [21]Département de Génétique Clinique, Institut de Pathologie et de Génétique, Gosselies 6041, Belgium; [22]Department of. Pediatrics and Neurology, Hospital Universitario Quirónsalud, School of Medicine, Universidad Europea de Madrid, Madrid 28224, Spain; [23]Genonics and Medicine, NIMGenetics, Madrid 28108, Spain; [24]Department of Translational Genomics, Center for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia; [25]Center of Excellence for Biomedicine, King Abdulaziz City for Science and Technology, Riyadh 12354, Saudi Arabia; [26]Department of Adult and Pediatric Medical Genetics, Hamad Medical Corporation, Doha 3050, Qatar; [27]Division of Genetic and Genomic Medicine, Sidra Medicine, Doha 26999, Qatar; [28]Department of Pediatrics-Medical Genetics, Duke University, Durham, NC 27710, USA; [29]Department of Pediatrics, Medical Genetics, Cedars-Sinai Medical

(Affiliations continued on next page)

Alberto Fernández-Jaén,[22] Sara Alvarez,[23] Irene Díez García-Prieto,[23] Fowzan S. Alkuraya,[24] Hessa S. Alsaif,[24,25] Zuhair Rahbeeni,[24] Karen El-Akouri,[26,27] Mariam Al-Mureikhi,[26,27] Rebecca C. Spillmann,[28] Vandana Shashi,[28] Pedro A. Sanchez-Lara,[29] John M. Graham, Jr.,[29] Amy Roberts,[30] Odelia Chorin,[31] Gilad D. Evrony,[31] Minna Kraatari-Tiri,[32] Tracy Dudding-Byth,[33] Anamaria Richardson,[34] David Hunt,[35,36] Laura Hamilton,[35] Sarah Dyack,[37] Bryce A. Mendelsohn,[38] Nicolás Rodríguez,[39] Rosario Sánchez-Martínez,[40] Jair Tenorio-Castaño,[41,42,43] Julián Nevado,[41,42,43] Pablo Lapunzina,[41,42,43] Pilar Tirado,[44] Maria-Teresa Carminho Amaro Rodrigues,[45] Lina Quteineh,[45] A. Micheil Innes,[46] Antonie D. Kline,[47] P.Y. Billie Au,[46,*] and Rosanna Weksberg[1,2,*]

laboratories, (2) the variability of clinical phenotypes within a disorder, and (3) overlapping non-specific features across different NDD syndromes. This diagnostic ambiguity has generated renewed interest in developing better clinical criteria to assist with reverse-phenotyping and developing accessible functional tools to classify VUSs.

Au-Kline syndrome (AKS [MIM: 616580]) is a rare neurodevelopmental and multiple congenital malformation syndrome. The condition was reported by Au et al. in 2015[1] as being caused by heterozygous loss-of-function (LoF) variants affecting heterogeneous nuclear ribonucleoprotein K (*HNRNPK* [MIM: 600712]). The AKS phenotype has also been previously described clinically as Okamoto syndrome (MIM: 604916) as well as in individuals with 9q21.3 microdeletions overlapping *HNRNPK*.[2–5] Since the identification of a human phenotype associated with variants in *HNRNPK*, two individuals with an initial clinical diagnosis of Okamoto syndrome were subsequently found to have pathogenic variants in *HNRNPK*, confirming that Okamoto and AKS are the same condition.[2,4] AKS is characterized by hypotonia, global developmental delay, characteristic facies (long palpebral fissures, shallow orbits, ptosis, a broad nasal bridge, hypoplastic alae nasi, downturned corners of the mouth, and a long face), congenital heart defects, genitourinary abnormalities, skeletal abnormalities, and variable other congenital malformations.[1,6–9]

AKS has a recognizable facial gestalt; however, as more individuals are ascertained, the phenotypic spectrum is widening to include more subtle presentations. Clinical diagnosis of AKS may therefore be challenging and clinical diagnosis is further complicated by clinical overlap with other conditions, such as Kabuki syndrome (KS [MIM: 147920 and 300867]).[1,6,7] Expert evaluation may be needed to identify characteristic features in mildly affected individuals. The molecular diagnosis of AKS can also be challenging. Many individuals have now been identified with rare *de novo* missense or intronic variants in *HNRNPK*, which are frequently reported as VUSs. While some of these individuals have the recognizable AKS facial gestalt that has been reported with *HNRNPK* LoF variants, it is challenging to confirm or rule out a diagnosis of AKS if there is a less characteristic facial appearance and/or if inheritance status of a variant is unknown.

We developed clinical diagnostic criteria following phenotypic review of a core group of individuals with a confirmed diagnosis of AKS (based on the presence of a LoF variant in *HNRNPK*). The purpose of these criteria is to guide the clinician's level of clinical suspicion of AKS and to enhance reverse phenotyping in individuals with VUSs in *HNRNPK*, many of whom have mild presentations. Further, we derived a score to grade the severity of disease in individuals with AKS to explore whether variant type correlated with disease burden.

In the last 7 years, a new functional tool that assesses genome-wide DNA methylation (DNAm) has emerged, which is particularly useful for classifying VUSs for a large class of genes that encode epigenetic regulators implicated in NDDs.[10–12] Although the primary disruption may impact histone modification or chromatin remodeling, some of these disorders still show specific genome-wide DNAm signatures, reflecting a layered dysregulation of the epigenetic machinery. Examples of disorders with identifiable DNAm signatures include Kabuki syndrome[10] (*KMT2D* [MIM: 602113]) and Sotos syndrome (MIM: 117550) (*NSD1* [MIM: 606681]),[11] which are associated with genes that impact histone methyltransferases. DNAm signatures

Center, Los Angeles, CA 90048, USA; [30]Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; [31]Center for Human Genetics & Genomics and, Department of Pediatrics, NYU Grossman School of Medicine, New York City, NY 10016, USA; [32]PEDEGO Research Unit, Medical Research Centre and Department of Clinical Genetics, University of Oulu and Oulu University Hospital, Oulu 90220, Finland; [33]The University of Newcastle, Newcastle 2308, Australia; [34]Faculty of Medicine, Department of Pediatrics, BC Children's Hospital and the University of British Columbia, Vancouver, BC V6H 3N1, Canada; [35]Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, UK; [36]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK; [37]Division of Medical Genetics, Department of Pediatrics, Dalhousie University, Halifax, NS B3K 6R8, Canada; [38]Department of Medical Genetics, Kaiser Permanente Oakland, 3505 Broadway, Oakland, CA 94611, USA; [39]Pediatrician Department, Hospital General Universitario de Alicante Dr. Balmis, Pintor Baeza, 11, 03010 Alicante, Spain; [40]Internal Medicine Department, Hospital General Universitario de Alicante Dr. Balmis, Institute for Health and Biomedical Research of Alicante, Pintor Baeza, 11, 03010 Alicante, Spain; [41]CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, Madrid 28042, Spain; [42]INGEMM-Idipaz, Institute of Medical and Molecular Genetics, University Hospital La Paz, Madrid 28046 Spain; [43]ITHACA, European Reference Network, Brussels 1000, Belgium; [44]Neuropediatrics Service, La Paz University Hospital, Madrid 28046, Spain; [45]Department of Genetic Medicine, University Hospitals of Geneva, 1205 Geneva, Switzerland; [46]Alberta Children's Hospital Research Institute for Child and Maternal Health, Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada; [47]Harvey Institute for Human Genetics, Greater Baltimore Medical Center, Baltimore, MD 21204, USA

[48]These authors contributed equally

*Correspondence: billie.au@albertahealthservices.ca (P.Y.B.A.), rweksb@sickkids.ca (R.W.)

https://doi.org/10.1016/j.ajhg.2022.08.014.

have also been identified in CHARGE syndrome (MIM: 214800) (*CHD7* [MIM: 608892]),[10] Nicolaides-Baraitser syndrome (MIM: 601358) (*SMARCA2* [MIM: 600014]),[13] and Floating Harbor syndrome (MIM:136140) (*SRCAP* [MIM: 611421]),[14] which implicate chromatin remodelers. Initially, DNAm signatures were reported for genes with primary roles in epigenetic regulation; however, more recently, evidence has emerged that genes with pleiotropic functions that include various interactions with DNA can also demonstrate DNAm signatures, such as *DYRK1A* (MIM: 600855),[15] *CDK13* (MIM: 603309),[16] and *ADNP* (MIM: 611386).[17] There are now >40 disorder/gene-specific DNAm signatures currently available that can be used as a second-tier test to functionally classify VUSs.[10–12] This approach is particularly helpful in rare disorders where there have been only a limited number of reported individuals and when there are missense variants, splice variants, or non-coding variants, which are more difficult to interpret.

HNRNPK encodes the heterogeneous nuclear ribonucleoprotein K (hnRNP K), a conserved and ubiquitously expressed nucleic-acid-binding protein involved in many gene expression processes, including chromatin remodeling, transcription, RNA stability, splicing, translation, post-translational modification, and signal transduction (reviewed in Barboro et al.[18]). It has been implicated in the regulation of both tumor-suppressive and oncogenic pathways. Given its role in chromatin regulation, we hypothesized that pathogenic variants in *HNRNPK* associated with AKS may be associated with a specific DNAm signature. We analyzed genome-wide DNAm in individuals with AKS confirmed by a LoF variant in *HNRNPK* to investigate the pathophysiology of AKS and its overlap with Kabuki syndrome. We identified a unique and robust DNAm signature for AKS that is distinct from that of the *KMT2D* Kabuki syndrome DNAm signature, showing that the DNAm signature of AKS is specific and that these two conditions have a distinct pathophysiology despite their clinical similarities.

The unique DNAm signature associated with AKS allowed for the functional characterization of VUSs in *HNRNPK* as either consistent with AKS (pathogenic) or not consistent with AKS (benign), which was particularly valuable when diagnostic scoring based on clinical features was equivocal for AKS. An integrated approach using the AKS DNAm signature together with the AKS clinical criteria allows us to confidently report on a group of individuals with missense and intronic variants in *HNRNPK* as having AKS, resulting in a significant phenotypic expansion for this syndrome to include individuals with only subtle clinical features.

Subjects and methods

Research subjects

Informed consent was obtained for all 39 research subjects for phenotype and natural history studies through protocols approved by the Calgary Health Research Ethics Board (REB #16–2419) and the GBMC Institutional Review Board (IRB #1220098). 31 of these individuals were recruited for DNAm studies according to the protocol approved by the Research Ethics Board of the Hospital for Sick Children (REB # 1000038847) or the local IRB of the respective recruiting institution. Photography consent for publication was obtained as needed, either as part of natural history studies or through respective institutions. Individuals with variants in *HNRNPK* and/or a clinically suspected diagnosis of AKS were identified through GeneMatcher[19] and through direct contact with collaborators. Study subjects were recruited through the institutions of the co-authors. One individual was diagnosed through the Deciphering Developmental Disorders (DDD) project.[20] One individual was diagnosed via the 100,000 Genomes Project.[21] One individual was diagnosed through an IRB-approved protocol (KFSHRC RAC#2080 006) at King Faisal Specialist Hospital. Two individuals were diagnosed through the Undiagnosed Disease Network (UDN).

A group of 31 individuals with clinical features suggestive of Au-Kline syndrome and/or *HNRNPK* variants was included in this study for DNA methylation analysis (Table 1, Figure 1). All but one of these individuals had a known pathogenic or candidate variant in *HNRNPK* identified via exome sequencing (ES).

The molecular findings in these individuals include the following: 8/31 with LoF variants in *HNRNPK*, 1/31 with a heterozygous microdeletion encompassing *HNRNPK*, 17/31 with missense variants in *HNRNPK*, 3/31 with intronic variants in *HNRNPK*, 1/31 with an in-frame deletion, and 1/31 (P20) with no detectable coding *HNRNPK* variant (this individual had non-diagnostic trio ES with no candidate variants and was included in this study because of phenotypic overlap with AKS). An additional eight individuals (N1–8) were included for phenotype studies alone (1/8 intronic, 2/8 missense, and 5/8 LoF), and three of these individuals aided in establishing the clinical scoring system.

Of the individuals reported here, six individuals have been previously published: AKS1 (patient 1 in Au et al., 2018[8] and originally published in Au et al., 2015[1]), AKS2 (patient 2 in Au et al., 2018[8]), AKS3 (patient 3 in Au et al., 2018[8] and originally published in Lange et al., 2016[17]), AKS7 (patient 7 in Au et al., 2018[8]), AKS8 (patient 8 in Au et al., 2018[1]), and AKS10 (patient 10 in Au et al., 2018[2]). All other individuals have not been previously reported.

Detailed clinical information was collected on all participating subjects via a phenotype questionnaire (see Table S1).

Development of clinical diagnostic criteria for Au-Kline syndrome

Two clinicians (P.Y.B.A. and A.D.K.) previously evaluated seven individuals with known AKS (defined as having a confirmed *de novo* LoF variant in *HNRNPK*) to assess the frequency of clinical findings, malformations, and medical involvement (previously published AKS1, 2, 7, and 11 as well as three previously unpublished individuals, N2, 3, and 4). Three clinicians (P.Y.B.A., A.D.K., and V.M.) then reviewed the clinical features of this group and compared them to five additional individuals. This set of five individuals had varying levels of clinical suspicion for AKS and included an individual with a predicted LoF variant (AKS8, which served as a positive internal control), two individuals with missense variants affecting recurrent residues (P6 and P11), one individual with a *de novo* missense variant of uncertain significance with a milder presentation (P17), and one individual with a

Table 1. Overview of all study subjects

	Sample ID	Sex (M/F)	Age (years) —sample collection (and last assessment)	Variant based on *HNRNPK* transcript GenBank: NM_002140.4	Mutation type	Coding effect (LoF = loss of function)	Met AKS diagnostic criteria? (Y/N/possible)	Face score	Severity score	DNAm signature (positive/negative/ intermediate)
Discovery cohort	AKS1[a,b]	M	13 (21)	c.953+1dupG	duplication	abnormal splicing (LoF)	Y	6	5	positive
	AKS3[a,b,c]	M	10 (11)	c.931_932insTT (p.Pro311Leufs*40)	insertion	frameshift (LoF)	Y	6	7	positive
	AKS7[a]	M	4 (9)	c.859C>T (p.Arg287*)	substitution	nonsense (LoF)	Y	5	6	positive
	AKS8[a]	M	11 (9)	c.779dupG (p.Asp262*)	duplication	nonsense (LoF)	Y	5	8	positive
	AKS10[a]	F	8 (8)	9q21.32 (86,328,837_ 86,592,487) ×1	deletion (264 kb)	–	Y	5	10	positive
	AKS 11	F	6 (10)	c.257G>A	substitution	abnormal splicing (LoF)	Y	6	11	positive
Validation cohort	AKS2[a,b]	M	7 (14)	c.257G>A	substitution	abnormal splicing (LoF)	Y	6	9	positive
	AKS12	F	13 (13)	c.1304_1322del (p.Ile435Argfs*15)	deletion	frameshift (LoF)	Y	6	3	positive
	AKS13	M	11 (8)	c.1090C>T (p.Gln364*)	substitution	nonsense (LoF)	Y	5	7	positive
Testing cohort	P1	M	10 (15)	c.214−77G>A	substitution	– (intronic variant)	Y	5	4	positive
	P2	F	25 (32)	c.137G>T (p.Arg46Leu)	substitution	missense	Y	5	3	positive
	P3	F	1 month (3)	c.140_143delinsATCA (p.Ile47_Leu48delinsAsnGln)	indel	in-frame	Y	5	7	positive
	P4	F	10 (10)	c.257 + 5G>A	substitution	– (intronic variant)	Y	5	8	positive
	P5	M	4 (3)	c.673T>C (p.Tyr225His)	substitution	missense	Y	6	4	intermediate
	P6	M	7 (7)	c.253G>A (p.Glu85Lys)	substitution	missense	Y	5	3	intermediate
	P7	F	4 (4)	c.137G>T (p.Arg46Leu)	substitution	missense	Y	5	3	intermediate
	P8	F	4 (4)	c.253G>A (p.Glu85Lys)	substitution	missense	possible	4	4	intermediate
	P9	M	4 (7)	c.213+5G>A	substitution	– (intronic variant)	Y	6	4	intermediate
	P10	M	8 (8)	c.248G>A (p.Gly83Asp)	substitution	missense	Y	5	4	intermediate
	P11	F	3 (4)	c.674A>G (p.Tyr225Cys)	substitution	missense	possible	4	4	intermediate
	P12	M	17 (16)	c.203T>G (p.Leu68Arg)	substitution	missense	Y	5	3	intermediate
	P13	M	28 (29)	c.253G>A (p.Glu85Lys)	substitution	missense	possible	4 (unable to score tongue)	2	intermediate
	P14	F	6 (8)	c.176G>A (p.Gly59Glu)	substitution	missense	possible	4	2	negative
	P15	F	6 (5)	c.185G>A (p.Gly62Asp)	substitution	missense	N (unable to determine)	2+ (unable to score fully)	4	negative

(*Continued on next page*)

Table 1. *Continued*

	Sample ID	Sex (M/F)	Age (years) —sample collection (and last assessment)	Variant based on *HNRNPK* transcript GenBank: NM_002140.4	Mutation type	Coding effect (LoF = loss of function)	Met AKS diagnostic criteria? (Y/N/possible)	Face score	Severity score	DNAm signature (positive/negative/ intermediate)
	P16	F	15 (16)	c.17C>G (p.Pro6Arg)	substitution	missense	N	2	5	negative
	P17	M	6 (4)	c.740G>A (p.Arg247His)	substitution	missense	N	0	6	negative
	P18	F	10 (10)	c.184G>T (p.Gly62Cys)	substitution	missense	N	0	3	negative
	P19	F	10 (10)	c.173T>C (p.Ile58Thr)	substitution	missense	N	2	2	negative
	P20	F	6 (5)	nil	–	–	possible	4	7	negative
	P21	F	8 months (19 months)	c.455A>T (p.His152Leu)	substitution	missense	possible	4	1	intermediate
	P22	F	?(11)	c.136C>T (p.Arg46Cys)	substitution	missense	Y	5	2	intermediate
Phenotype cohort	N1	F	(4)	c.214−35A>G	substitution	– (intronic)	Y	6	5	N/A
	N2	M	(5)	c.1250C>A (p.S417*)	substitution	nonsense (LoF)	Y	6	7	N/A
	N3	M	(3)	c.999C>A (p.Tyr333*)	substitution	nonsense (LoF)	Y	6	NR	N/A
	N4	F	(4)	c.998dupA (p.Tyr333*)	duplication	nonsense (LoF)	Y	6	12	N/A
	N5	M	(13)	c.999C>A (p.Tyr333*)	substitution	nonsense (LoF)	Y	5 (unable to score tongue)	8	N/A
	N6	F	(4)	c.253G>A (p.Glu85Lys)	substitution	missense	Y	4 (unable to score mouth	2	N/A
	N7	M	(22)	c.253G>A (p.Glu85Lys)	substitution	missense	Y	5	2	N/A
	N8	F	(7)	c.998dupA (p.Tyr333*)	duplication	nonsense (LoF)	Y	5 (unable to score tongue)	NR	N/A

Total n = 39, n = 31 involved in DNAm signature analysis. Cohorts as used for the DNAm signature are divided into discovery (n = 6), validation (n = 3), and testing (n = 22). Phenotype only cohort (n = 8). N/A, not applicable. Possible AKS = ID + hypotonia + 4 facies OR facies (5+) + one of other two majors. NR = insufficient information. "Nil' refers to no sequence variants in *HNRNPK* was detected by genome sequencing.
[a]Previously published in Au et al., 2018.[8]
[b]Au et al., 2015.[1]
[c]Lange et al., 2016.[6]

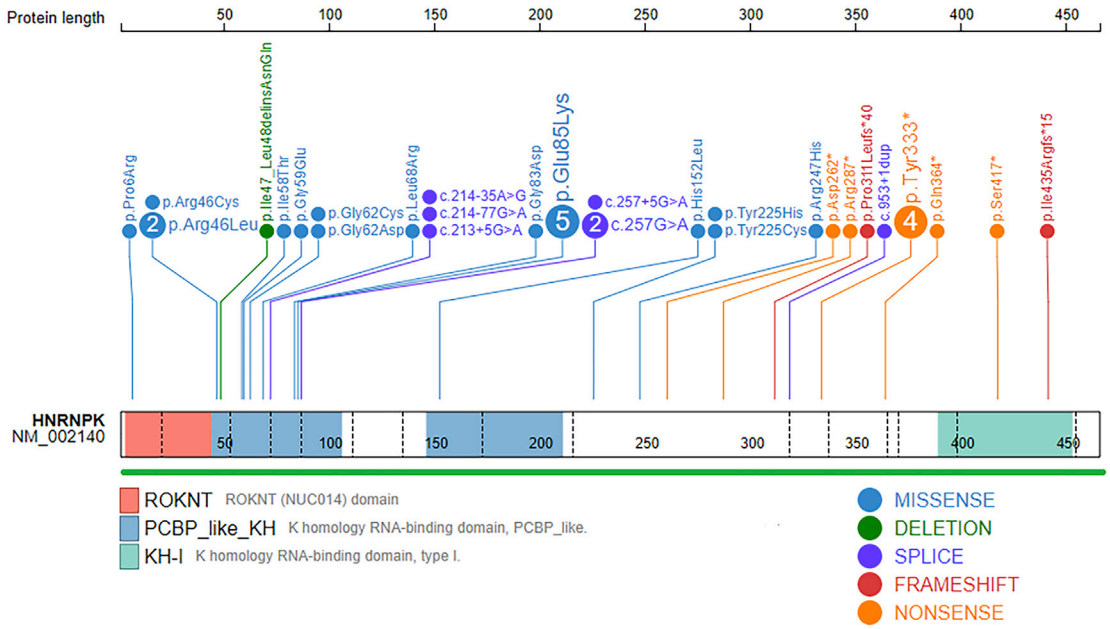

Figure 1. Schematic representation of the *HNRNPK* transcript, its functional domains, and variants used in this study
Each distinct variant in *HNRNPK* is represented by a disc sized in proportion to the number of samples and filled with the color representing its class based on the legend. Sequence variants are positioned by their amino acid coordinates based on *HNRNPK* GenBank: NM_002140.4, hg19. The dotted vertical lines inside the protein delineate the boundaries of coding exons and the filled colors within the protein correspond to known protein domains. ROKNT (NUC014) domain; PCBP_like_KH, K homology RNA-binding domain; KH-I, K homology RNA-binding domain, type I.

clinical phenotype suggestive of AKS but without a detectable *HNRNPK* sequence variant (P20). Review of these individuals helped refine which features would be most specific to include as clinical diagnostic criteria for AKS and helped exclude features that were variably present (see Table S2 and Figure S1).

The diagnostic criteria were then tested on a larger set of individuals with clinical features overlapping AKS and/or presence of an *HNRNPK* variant. This set included a previously published individual with a known LoF variant (AKS10),[8] a published individual with an LoF variant (AKS8), an unpublished individual with an intronic variant confirmed to have a splice effect (N1, unpublished data from UDN), and individuals with missense variants affecting a recurrent residue in multiple probands (P6, 7, and 11). Four clinicians (P.Y.B.A., A.M.I., A.D.K., and V.M.) rated this cohort independently to determine whether the clinical score was reproducible amongst clinicians. One clinician (A.M.I.) was blinded to the *HNRNPK* variant status and the initial process of creating the diagnostic score.

To validate the AKS score's specificity, 22 individuals with a confirmed diagnosis of Kabuki syndrome (KS) due to pathogenic variants in *KDM6A* (MIM: 300128) or *KMTD2* were assessed by three independent clinicians who have expertise in KS (M.A., H.B., and J.H.) by using the AKS clinical diagnostic criteria (see Table S3).

Development of a clinical severity scoring system for Au-Kline syndrome

After clinical diagnostic criteria were established, a clinical severity score was created. Three clinicians (P.Y.B.A., A.D.K., and V.M.) reviewed the phenotypic spectrum of five individuals with AKS to identify clinical features that covered the breadth of the phenotypic spectrum and the main causes of morbidity in AKS. This group of five included individuals with a clinical diagnosis of AKS (based on the newly developed clinical diagnostic criteria), including previously published individuals with LoF variants (AKS1, 7, and 10)[8] and an unpublished individual with a novel LoF variant (N4). This group also included two individuals who had missense variants affecting a recurrent residue (P7 and P11), including one (P11) with a "possible AKS" clinical score, to represent a potential mild presentation of the AKS spectrum. Clinical features included as part of the severity score were selected on the basis of which body systems were frequently affected in AKS (craniofacial, skeletal, cardiac, renal, gastrointestinal, eye, neurodevelopmental, and growth) and included a spectrum of phenotypic variability (Table S4). Manifestations deemed more severe were those that led to increased morbidity, such as more severe disability (e.g., vision deficit), or a need for surgical or invasive intervention (e.g., surgical correction of craniosynostosis or requirement for a feeding tube). Once the severity score was established, the remaining cohort underwent scoring. Of note, the purpose of the clinical severity score is to establish the extent of disease burden; the components of the severity score are not unique to individuals with AKS.

Cohorts for DNA methylation analysis

Disease cohort

For DNAm signature generation and validation, we selected nine individuals with a clinical diagnosis of AKS (based on the AKS diagnostic criteria) and presence of an *HNRNPK* LoF variant (Tables 1 and S1). A random 75% subset of these individuals was then categorized as the discovery cohort (n = 6, [AKS1, 3, 7, 8, 10, and 11]). The remaining 25% (n = 3, [AKS 2, 12, and 13]) was used for validation.

The testing cohort (n = 22) included individuals with *HNRNPK* variants that were either likely pathogenic or of uncertain significance based on ACMG classification criteria.[22] The individual with

a clinical suspicion of AKS but no *HNRNPK* variant (P20) was also included in this cohort.

All gene variant annotations for *HNRNPK* cohorts and *in silico* prediction via PolyPhen-2 and SIFT were generated with Alamut visual 2.11. CADD scores were obtained with https://cadd.gs.washington.edu/snv, v1.4. All DNA was extracted from peripheral blood.

Control cohort

Genomic DNA from peripheral blood was obtained from 16 control individuals selected as age- and sex-matched neurotypical controls to the AKS discovery set. An additional 172 reference blood control DNA methylation profiles were used in the study to determine the specificity of the DNAm signature. These controls were obtained from the POND Network, The Hospital for Sick Children, and The University of Michigan (Dr. Greg Hanna). Neurotypical was defined as healthy and developmentally normal on formal cognitive/behavioral assessments (samples from POND and The University of Michigan) or via physician/parental screening questionnaires (Hospital for Sick Children). For detailed information, see Table S5.

DNAm array processing

Genome-wide DNAm profiling on control and affected individuals matched for age and sex was performed at The Center for Applied Genomics (TCAG), SickKids Research Institute, Toronto, ON, Canada. Genomic DNA from each subject was sodium bisulfite converted with the EpiTect Bisulfite Kit (EpiTect PLUS Bisulfite Kit, QIAGEN, Valencia, CA) according to the manufacturer's protocol. Modified genomic DNA was then processed and analyzed on the Infinium HumanMethylationEPIC BeadChip (Illumina 850K) according to the manufacturer's protocol.[23] The distribution of the samples on the arrays was randomized for both affected individuals and controls. All signature-derivation affected individuals and controls were run in the same batch.

Quality control and normalization

The raw IDAT files were converted into β values, which represent DNAm levels as a percentage (between 0 and 1), with the *minfi* Bioconductor package in *R*. Data preprocessing included filtering out non-specific probes (44,135 probes); probes with detection p value > 0.05 in more than 25% of the samples (771 probes); probes located near single-nucleotide polymorphic sites (SNPs) with minor allele frequencies above 1% (n = 29,958); probes with raw beta = 0 or 1 in >0.25% of samples (n = 21); non-CpG probes (n = 2,932); and X and Y chromosome probes (n = 19,627) for a total of 91,379 probes removed and a total of n = 774,480 probes remaining for differential methylation analysis. Standard quality control (QC) metrics in *minfi* were used, including median intensity QC plots, density plots, and control probe plots: all samples passed quality control and were included in the study.

Differential DNAm analysis

The analysis was performed with our previously published protocol.[12] Differential DNAm analysis between AKS and controls was performed at 774,480 CpG sites with beta scores, representing DNAm levels as a percentage (between 0 and 1). The β value from each sample at the remaining 774,480 CpGs was used for downstream analysis and generation of a DNAm signature. β values were logit transformed to M values with the following equation: $\log2(\beta/(1-\beta))$. We used a linear regression modeling by using *limma* package[24] to identify the differentially methylated probes. We estimated blood cell counts by using Houseman's method implemented in minfi and FlowSorted.Blood.EPIC Bioconductor packages to generate the proportions of $CD8^+$ T cells, $CD4^+$ T, natural killer, B cells, monocytes, and granulocytes (mainly neutrophils [Neu]).[25] The analysis was done on the discovery set of six AKS and 16 controls and was adjusted for age, sex, and blood-cell type. The generated p values were corrected for multiple testing with the Benjamini-Hochberg method. A significant difference in DNAm between AKS and control samples for each CpG site was required to meet the cutoffs of Benjamini-Hochberg-adjusted p values < 0.05 and $|\Delta\beta| \geq 0.10$ (10% methylation differences) as previously reported[12] were considered significant.

Generation of disease score classification model using correlation analysis

We used a previously described pipeline for generating disease scores by using an established disease-specific DNAm signature.[11,12] At each of the 429 signature CpGs, a median DNAm level was computed across the AKS-affected individuals of the diagnostic cohort (as described above [n = 6]) used to generate the signature, resulting in a reference profile. Similarly, a robust median-DNAm reference profile for the signature controls (n = 16) was created. The classification of each additional gene variant or control DNAm sample was based on extracting a vector $\beta R_{sig}R$ of its DNAm values in the signature CpGs and comparing $\beta R_{sig}R$ to the two reference profiles computed above. *HNRNPK* score was defined as $HNRNPK\ score = r(\beta R_{sig}R, AKS\ profile) - r(\beta R_{sig}R, control\ profile))$, where *r* is the Pearson correlation coefficient. A classification model was developed on the basis of scoring each new DNAm sample with the *HNRNPK* score: a test sample with a positive score is more similar to the AKS reference profile based on the signature CpGs and is therefore classified as "AKS", whereas a sample with a negative score is more similar to the control-blood reference profile and is classified as "not-AKS." To test specificity, we scored and classified EPIC array data from 172 additional neurotypical controls. To test sensitivity, we scored and classified the validation cohort of three additional unrelated AKS individuals with *HNRNPK* pathogenic variants.

Generation of machine learning model for variant classification

Using the R package caret, we removed probes with very similar methylation patterns with correlation greater or equal to 90% (redundant probes) as we previously described.[10,12] Next, we developed a machine learning model, a support vector machine (SVM) model with linear kernel that had been trained on the significant CpG sites from the discovery cohorts after further filtering to remove redundant CpGs. The model was set to the "probability" mode to generate SVM scores ranging between 0 and 1 (or 0% and 100%), thus classifying samples as "AKS" (high scores) or "not-AKS" (low scores). This SVM model was built as a tool for the classification of variants in *HNRNPK* and *KMT2D*.

Gene and genomic regions enrichment analyses

Gene Ontology analysis was performed in Metascape[26] (http://metascape.org) for the 184 human Entrez Gene IDs overlapping the signature CpG sites. Background genes from the Illumina EPIC array were used as the background list. Significant biological process categories were determined by hypergeometric test with Benjamini-Hochberg correction.

Table 2. Au-Kline syndrome clinical diagnostic criteria

Major criteria	Diagnostic criteria for Au-Kline syndrome	Description
1	global developmental delay or intellectual disability	DQ or IQ score below 70
2	congenital hypotonia	clinical diagnosis
3	five of six facial features	facies: long
		eyes: shallow orbits
		eyes: apparently increased length of palpebral fissures
		nose: broad nasal ridge and/or thick alae nasi with narrow nares
		mouth: upper lip with exaggerated cupid's bow (widened M shape)
		mouth: tongue large or bifid and/or with deep midline groove

Three major criteria are required for a clinical diagnosis of AKS. DQ is developmental quotient and IQ is intelligence quotient.

Tissue-specific expression analysis and cell-type-specific expression analysis (TSEA and CSEA, respectively) were performed on the 184 human Entrez Gene IDs with the pSI package.[27] Specificity indices were determined with previously published RNA sequencing data.[28] We used a Chi-square test to determine whether the distribution of cell types in input genes was significantly different. Genomic regions enrichment analysis was performed with the Illumina EPIC array annotation for both "UCSC relation to CpG island" and "DNase hypersensitive sites" to compare CpG sites that overlapped the signature against the background CpG sites from the EPIC array.

Statistical analysis

Continuous variables were presented as medians because of skewed distribution. To determine statistical differences for the clinical severity score between *HNRNPK* positive and intermediate DNAm signature groups, group differences were assessed by nonparametric Mann-Whitney U test/Kruskal-Wallis test.

Results

We collected molecular and clinical information for the 39 individuals in this research study (Table 1; additional details in Table S1). The 31 individuals in the DNA methylation cohort had an age range of 1 month–28 years, with 18 males and 20 females. The eight individuals included only for phenotype studies ranged from age 3 to 22 years, with four females and four males.

Clinical diagnostic criteria for Au-Kline syndrome

Based on review of individuals with *HNRNPK* LoF variants followed by a comparison to a group of more heterogeneous individuals as described above, a core set of six facial features was identified. These features added the most specificity for clinical diagnosis: a long face; shallow orbits; long palpebral fissures; a broad nasal ridge and/or thick alae nasi; an exaggerated Cupid's bow (wide M shape) of the upper lip; and a tongue that is large, bifid, or has a deep midline groove (Table 2 and Figure 2). Other features common in other neurodevelopmental disorders (NDDs) and less specific to AKS were removed. Examples include facial features (such as myopathic face, ptosis, and long ears), craniosynostosis, and congenital anomalies (malformations seen in AKS are common to many NDDs). While not specific to AKS, the presence of global developmental delay (GDD) or intellectual disability (ID) and hypotonia were identified as being required for an AKS diagnosis.

Individuals with facial scores of 5 or 6, in combination with GDD/ID and hypotonia, were classified as "having a clinical diagnosis of AKS" (these individuals will subsequently be referred to as "AKS"). Individuals with GDD/ID and hypotonia with a facial score of 4 were classified as "possible AKS." Individuals who did not have GDD/ID or had a facial score of 3 or below were classified as "unlikely AKS." See Table 2 for the newly developed clinical diagnostic criteria for AKS. When possible, photographs at multiple ages were reviewed.

In addition to the seven individuals with *HNRNPK* LoF variants used to develop the clinical diagnostic criteria (AKS1, 2, 7, 11, N2, 3, and 4), the other individuals (AKS3, 8, 10, and 13) with *HNRNPK* LoF variants also met criteria for having AKS, validating the diagnostic criteria.

For the individuals in the test cohort with missense variants, intronic variants, and the individual with no *HNRNPK* variant, 11/22 met diagnostic criteria for having AKS on the basis of clinical features alone. Facial scores for the individuals with missense variants were extremely variable, ranging from 0 to 6 (see Table 1). 7/17 of the individuals with missense variants met diagnostic criteria for AKS (see Table 1). Of note, facial scores could be variable even for individuals with recurrent variants, e.g., individuals with the recurrent *HNRNPK* variant: c.253G>A (p.Glu85Lys) (GenBank: NM_002140.4) scored either "possible AKS" with 4/5 (P8, 13, and 11) or as "AKS" with >5/6 (P6); however, for some individuals there was insufficient information (photograph or documented exam) to fully score the tongue.

Identification of DNAm signature in Au-Kline syndrome

To identify an *HNRNPK* DNAm signature, we generated genome-wide DNAm profiles by using Infinium Human MethylationEPIC BeadChip arrays with DNA from peripheral blood samples of individuals with pathogenic sequence variants in *HNRNPK* and controls. We identified nine individuals with AKS and with *HNRNPK* LoF variants reported by molecular diagnostic laboratories (Table 1) or through research-based testing. We selected a random 75% subset of these AKS-affected individuals with LoF variants in *HNRNPK* as a discovery cohort (n = 6) for the purpose of feature selection and model training. The remaining 25% (n = 3) was used for validation or assessment of

Figure 2. Facial composite of individuals with AKS diagnosis

(A) Individuals with *HNRNPK* LoF variants in the discovery cohort and (B) in the validation cohort. Individuals with asterisk (*) have been previously published. All individuals in (A) and (B) have LoF variants.

(C) Individuals in the test cohort with VUSs in *HNRNPK* (as indicated) and with a positive AKS DNAm signature. All individuals (A, B, C) with positive AKS DNAm are outlined in red.

(D) Individuals in the test cohort with VUSs in *HNRNPK* with an intermediate AKS DNAm signature. These individuals are outlined in gold.

(E) Individuals in the test cohort who had a negative AKS DNAm signature. These individuals are outlined in blue.

(F) Individuals in phenotype-only group. Photos of N7 show evolution of facial features from infancy to late childhood. For photographs of individuals with AKS (A, B, C, D, and F), an italicized label indicates a facial score of 4/5, whereas regular text indicates a facial score of $\geq$5/6. Not all individuals in the test or phenotype cohorts are shown depending on consent for photograph publication.

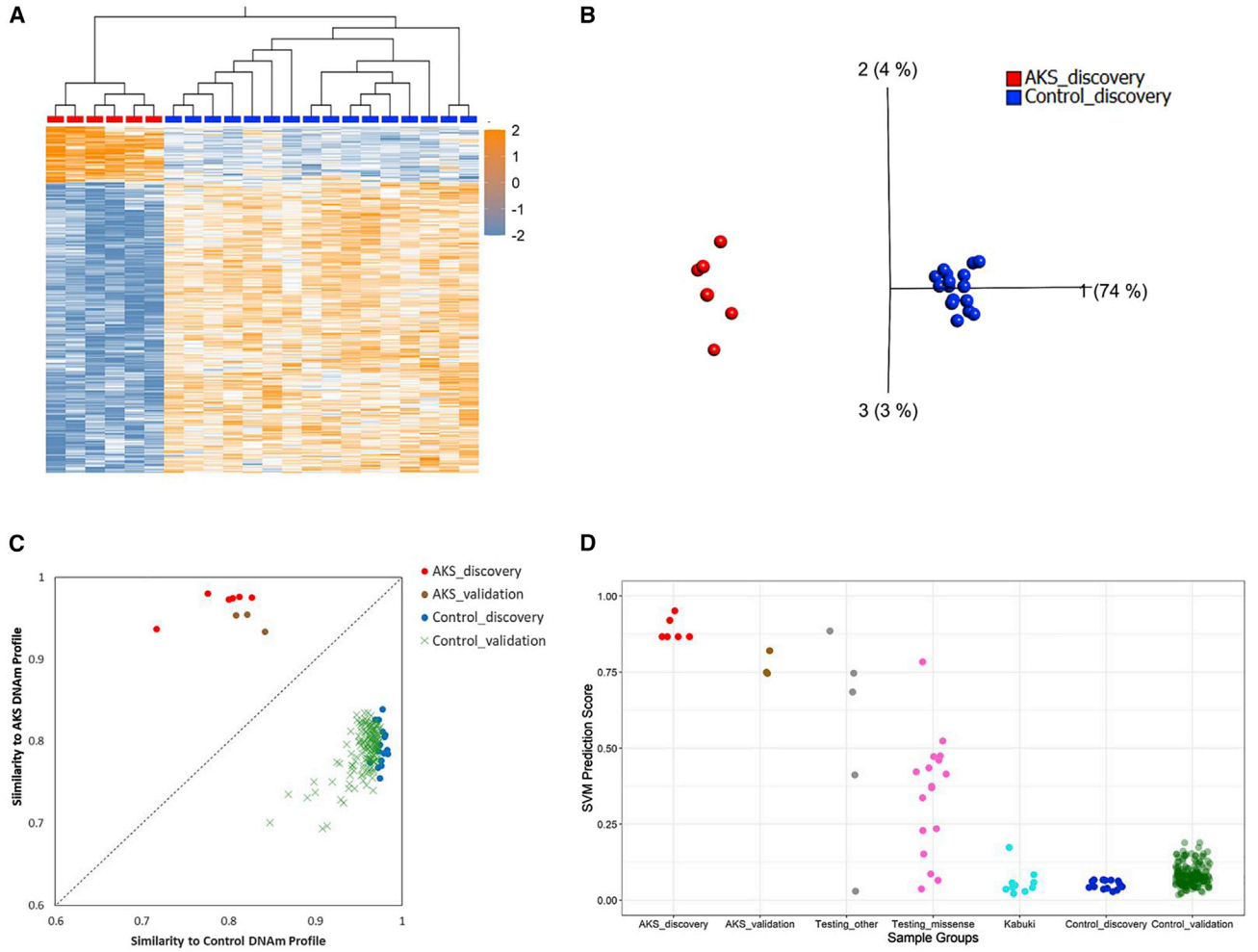

Figure 3. DNA methylation signature of *HNRNPK*

(A) Heatmap showing the hierarchical clustering of discovery *HNRNPK* LoF individuals (n = 6) and age- and sex-matched neurotypical discovery controls (n = 16) used to identify the 429 differentially methylated signature sites shown. The color gradient represents the normalized DNA methylation value from −2.0 (blue) to 2.0 (orange) at each site. DNA methylation at these sites clearly separate discovery individuals (red) from discovery controls (blue). Euclidean distance metric is used for the clustering dendrogram.

(B) Principal-component analysis (PCA) visualizing the DNAm profiles of the study cohort at the 429 signature sites.

(C) Validation of *HNRNPK* LoF individuals (not used to define the signature sites; red) cluster with discovery individuals, while control validation (n = 172, yellow) cluster with controls.

(D) Support vector machine (SVM) classification model based on the DNA methylation values in the discovery groups. Each sample is plotted on the basis of its scoring by the model. Samples with an SVM score > 0.25 were considered likely disease causing. *HNRNPK* validation individuals from our cohort (n = 3) classified as "AKS" and all control validation individuals (n = 172) classified as "not AKS." Missense variants in the testing cohort showed a spectrum of SVM scores compared to other testing variants. Pathogenic *KMT2D* variants (Kabuki syndrome) also classified as "not AKS".

the performance of the classification model. Both cohorts included previously published individuals from international centers, and all individuals met diagnostic criteria. We used our established pipeline as outlined in the subjects and methods for signature derivation. Of the 774,480 CpG sites tested for differential DNAm between AKS and controls, we identified 429 statistically significant changes in DNAm across the genome at a false discovery rate (FDR)-adjusted p value < 0.05 and $|\Delta\beta| \geq 0.10$ (Table S6). We examined these CpG sites by using hierarchical clustering (Figure 3A) and principal-component analysis (PCA) (Figure 3B) to assess their capacity to separate AKS subjects from controls. As seen in Figure 3B, we could use the significant CpGs to successfully segregate the discovery cohort of AKS from controls. These differentially methylated 429 CpG sites representing the AKS-associated *HNRNPK* DNAm signature will be referred to from here as the AKS signature.

Specificity and sensitivity of the AKS signature

To test the specificity and sensitivity of the AKS signature, we generated median-methylation profiles of controls and AKS subjects from the discovery cohort and classified our independent validation cohort of AKS-affected individuals (n = 3) and controls (n = 172) as either "AKS" (positive disease score) or "not AKS" (negative disease score) on the

basis of their DNAm profiles by using the correlation-based classification model (see subjects and methods). All controls showed DNAm profiles similar to the control profile, had negative disease scores, and were classified as "not AKS" demonstrating 100% specificity (Figure 3C and Table S7). Each individual in the AKS validation cohort clustered with AKS individuals and not with controls and generated positive disease scores. Therefore, the validation cohort demonstrated 100% sensitivity (Figure 3C and Table S7).

Testing the utility of AKS signature for VUS classification

Using the highly specific and sensitive AKS signature, we classified 21 unrelated individuals with *HNRNPK* VUSs (17 missense variants, three intronic variants, and one in-frame deletion) and the one individual (P20) with no detectable variant in *HNRNPK*. These 22 samples were classified with the support vector machine-learning model (SVM). This model generated scores between 0% and 100%, with high scores classified as "AKS" and low scores classified as "not AKS." We found that four variants had high SVM scores $>$ 70% and seven had SVM scores below 25% (i.e., not AKS). Interestingly, 11 variants were classified as intermediate with SVM scores ranging between 34% and 52% (Figure 3D and Table S8). This distinct classification of ten missense variants, which included recurrent missense variants (Table 1), prompted us to look for (epi)-genotype-phenotype correlations associated with AKS and *HNRNPK* missense variants. (see epi-genotype phenotype analysis section below).

Classification of Kabuki syndrome using AKS signature

Since Kabuki syndrome is included in the differential diagnosis of AKS because of clinical overlap, we further tested the specificity of the AKS signature in classifying individuals with Kabuki syndrome and *KMT2D* pathogenic variants. Using an SVM classification model for *HNRNPK*, we classified ten pathogenic variants in *KMT2D* as "not AKS" with an SVM score $<$ 25% (Figure 3D and Table S8). These data provide additional evidence supporting the high specificity of the AKS signature.

Functional enrichment of the *HNRNPK* DNAm signature

We investigated the DNAm signature for its potential to elucidate the molecular pathophysiology of AKS. Gene set enrichment analysis was performed with Metascape. The top ten pathways that reached statistical significance ($p < 0.05$) are shown in Figure 4A. For the GO analysis, the enriched genes function mainly in neuronal projection morphogenesis, synapse organization, and tissue/skeletal development as well as intracellular signaling.

Unbiased tissue-specific enrichment analysis (TSEA) (Figure 4B) and cell-type specific expression analysis (CSEA) (Figure 4C) show AKS-significant genes are most strongly enriched in transcripts expressed in the brain, pituitary, uterus, blood vessels, and specifically in the developing cerebellum.

Analysis of the genomic locations of the CpG sites in the signature showed that CpGs were over-represented in CpG island shores (defined as 0 to 2 kb upstream of CpG island) and DNase hypersensitive sites (Figure 4D).

Epi-genotype phenotype analysis

Next, we compared DNAm classification on the basis of the AKS signature in individuals with *HNRNPK* VUSs and/or unclear clinical diagnosis of AKS for potential epi-genotype phenotype correlations. We identified that four individuals with VUSs in *HNRNPK* with a clinical diagnosis of AKS based on the diagnostic score also had a positive AKS DNAm signature (P1, 2, 3, and 4). This included two individuals with intronic splice variants and the individual with an in-frame deletion-insertion.

Eleven individuals in the test cohort had an intermediate DNAm signature. This included one individual (P9) with an intronic variant who met diagnostic criteria for AKS and six individuals with missense VUSs in *HNRNPK* who also met diagnostic criteria for AKS (P5, 6, 7,10, 12, and 22). The other four individuals with missense variants and an intermediate DNAm signature (P8, 11, 13, and 21) had only facial scores of 4 and were considered as "possible AKS" by diagnostic criteria; however, they were highly suspected of having AKS as they had *de novo* missense variants affecting residues that are recurrently affected in our test cohort, Glu85 or Tyr225. These missense variants may be considered as likely pathogenic by ACMG criteria,[22] as they affect recurrent residues in *HNRNPK* seen in other individuals who meet diagnostic criteria for AKS (see Table 1) (P5 and 6). Given the above observations, we reclassified any variant of uncertain significance associated with either a positive or intermediate AKS DNAm signature to likely pathogenic.

All three individuals (P6, 8, and 13) in the methylation cohort with the recurrent *HNRNPK* variant, GenBank: NM_002140.4 c.253G>A (p.Glu85Lys) (GenBank: NM_002140.4),were intermediate for the AKS DNAm signature (Figure 3 and Table 1). P2 (facial score 5) had a recurrent *HNRNPK* missense variant, c.137G>T(p.Arg46Leu) (GenBank: NM_002140.4), and had a positive DNAm signature; however, P7 (facial score 5), who had the same variant, was intermediate.

Six individuals with VUSs in *HNRNPK* who did not meet the criteria for a clinical diagnosis of AKS (P14, 15, 16, 17, 18, and 19) were all negative for the AKS signature. These individuals all had *de novo* variants, with the exception of P16, for whom the inheritance was unknown as the maternal sample was not available. The individual with no *HNRNPK* variant (P20) who was a "possible AKS" on the basis of a facial score of 4 was also negative for the AKS signature. These negative results support that these *HNRNPK* VUSs are more likely to be benign or potentially associated with a separate condition distinct from AKS (Figure 3 and Table 1).

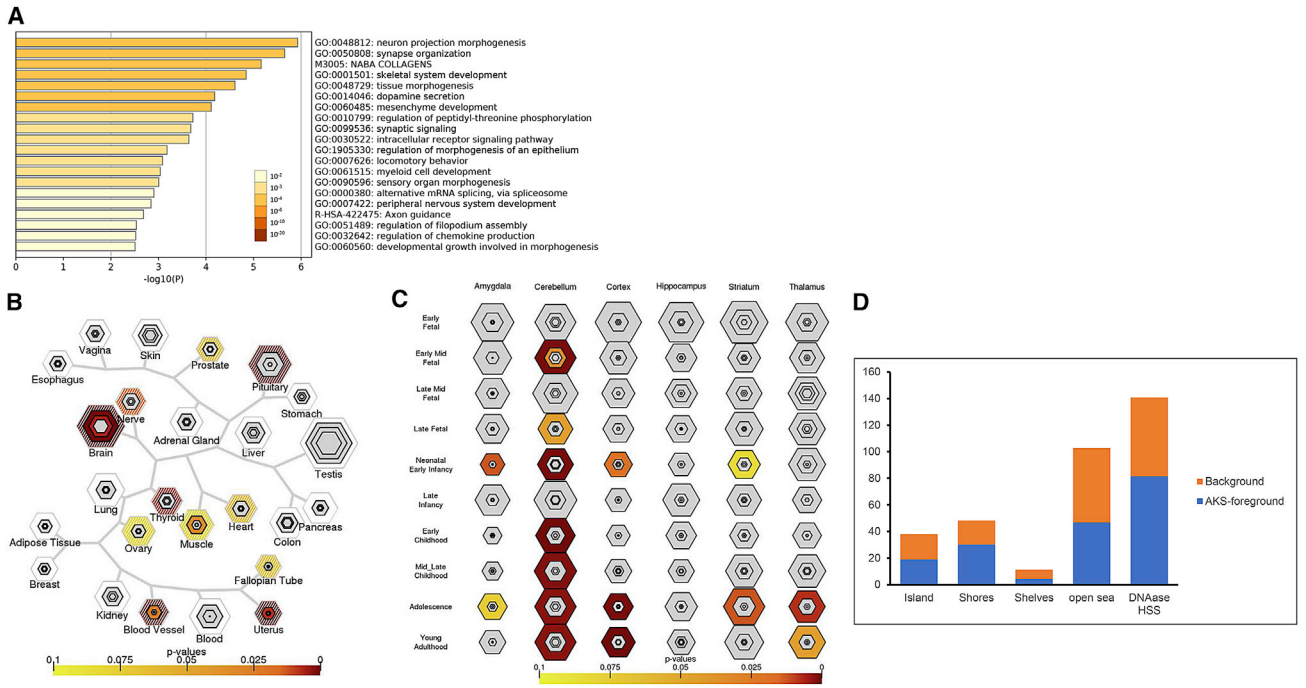

Figure 4. Gene and genomic enrichment analyses

(A) GO enrichment analysis showing the Metascape bar graph for top non-redundant enrichment clusters, one per cluster, using a discrete color scale to represent statistical significance.

(B) Specific tissue expression and (C) specific brain region enrichment as determined by SEA, showing enrichment of expression of the signature genes and we found that the genes that overlapped the significant CpGs are highly expressed in brain tissue. Hexagons represent list of genes enriched in each tissue or cell type going from pSI threshold to include larger but less stringent gene lists to thresholds for the most stringent subsets (smallest, central hexagons). Fisher's exact testing is used to establish p values for each target and each gene list at each threshold, which are corrected by Benjamini–Hochberg for the number of cell types (color bar).

(D) Bar chart representing the percentage distribution of the CpG sites according to genomic annotations extracted from the Illumina EPIC array annotation file. The CpG sites that overlap the signature were compared to the distribution of the CpGs within the dataset (background, 774,480 CpGs) for the regulatory feature group: "relation to CpG island" and "DNase hypersensitive sites (DHSs)."

Clinical severity score for AKS

We identified fourteen clinical features as contributing to the severity of the AKS phenotype (Table S4). These include three categories: growth abnormalities, major organ system abnormalities (craniofacial, cardiac, genitourinary, musculoskeletal, and nervous system), and developmental abnormalities. Each feature receives a single point, and a higher score reflects a higher disease burden.

All individuals with a positive or intermediate AKS DNAm were scored for severity. When considering all 13 individuals with a positive AKS DNAm signature (including the diagnostic, validation, and test cohorts), the median severity score was 7 (range 3–11). This group primarily represents individuals with LoF variants. For the 11 individuals with an intermediate AKS DNAm signature (P5, 6, 7, 8, 9, 10, 11, 12, 13, 21, and 22), the median severity was 3 (range 1–4) (see Table 1 and Figure S2). This group primarily represents individuals with missense variants. The difference between these groups was statistically significant with $p < 0.001$ via Kruskal-Wallis test.

Although most individuals with AKS-associated missense variants had an intermediate DNAm score, one individual (P2) with a missense variant had a positive DNAm score and had a severity score of 3. There were insufficient numbers of individuals with missense variants in the positive DNAm score group to determine whether there was a significant difference in severity between these individuals and individuals with missense variants in the intermediate DNAm group.

The difference between the severity scores of the positive AKS DNAm group and the intermediate DNAm group may be due to a difference in organ malformation burden. 10/13 positive DNAm signature individuals (nine have LoF variants) had congenital heart disease compared to 5/11 with intermediate DNAm (nine have missense variants). 10/13 positive DNAm signature individuals had renal anomalies compared to 3/10 with intermediate DNAm. 4/13 of the positive DNAm group had frank cleft palate compared to only one in the intermediate DNAm group (P9). 8/13 in the positive DNAm group had scoliosis compared to only one individual in the intermediate DNAm group. Of the components of the severity score, congenital heart disease was found to be the main contributor to the difference between these groups ($F = 6.14$, $p = 0.023$; ANOVA). There is a significant difference in severity score between positive DNAm and intermediate DNAm

groups even if hypotonia and motor and speech delay points are removed and the comparison is made with only systemic involvement (growth and malformations) ($p < 0.001$ via Kruskal-Wallis).

Given the wide age range and the limited number of individuals in the cohort presented here, it was not possible to correlate developmental outcomes with *HNRNPK* variant type or DNAm status. However, there may be a trend toward better development outcomes (i.e., more likely to be ambulatory or have verbal communication at a younger age) for individuals with missense variants.

Overview of clinical features and phenotypic expansion

The DNAm signature confirmed the AKS diagnosis in a broader cohort of individuals, including multiple individuals with missense and intronic variants. An overview of the clinical features seen in individuals with a diagnosis of AKS as confirmed by the DNAm signature can be seen in Table 3.

Review of the entire expanded cohort of AKS individuals demonstrates that malformations are common in AKS; renal (hydronephrosis in 52%) and congenital heart defects (present in 62%) are the most frequent anomalies. However, some individuals with missense variants do not have malformations or have only isolated malformations. Furthermore, some malformations such as vertebral segmentation anomalies are currently not observed in individuals with missense variants. Aortic dilation was present in three individuals (AKS2, P13, and N3); however, given that most individuals are still young, the true incidence and screening requirement remain unclear. Two individuals now have a formal diagnosis of dysautonomia and more than 1/3 of the individuals have a combination of high pain tolerance, GI dysmotility, abnormal sweating, and heat intolerance. CNS malformations are variable, but delayed myelination/hypomyelination and hypoplasia of the corpus callosum are recurrent findings. Skeletal issues, particularly affecting the spine, are common. Scoliosis can occur even in the absence of segmentation anomalies. Craniosynostosis (typically sagittal or metopic sutures) occurs in approximately 20%.

Global developmental delay and/or intellectual disability and hypotonia are universally reported. The distinctive facial features associated with AKS are key in recognizing the syndrome (see Figure 2 for facial composite). The AKS facies appear to evolve and may not be as easily recognizable at a young age, such as in the 1-month-old individual, where the face was rounder (P3; subsequent photos at 3 years of age met the AKS facies criterion), and as demonstrated by N7 (see Figure 2).

Thus, the phenotypic spectrum for AKS is much broader than previously recognized. Individuals with loss-of-function *HNRNPK* variants are more likely than individuals with missense *HNRNPK* variants to have organ malformations.

Discussion

We describe the expanded phenotypic spectrum of Au-Kline syndrome based on 32 individuals with a confirmed diagnosis. All individuals had sequence variants in *HNRNPK* and were assessed with the newly developed AKS clinical diagnostic criteria. Twenty-four individuals with AKS clinical diagnosis were also characterized with DNAm profiling, including 11 individuals with missense variants that are now reclassified as likely pathogenic. Six individuals with *de novo HNRNPK* variants were excluded from having AKS, as these individuals did not have the characteristic AKS DNAm signature.

AKS was initially described in 2015 and 2018[1,8] as a multiple malformation syndrome with a clinically recognizable facial gestalt. Similar to other newly described genetic syndromes, syndrome delineation was facilitated by the identification of multiple individuals with *de novo* LoF variants in a common gene. This ascertainment meant that the initial AKS cohort represented a narrow phenotype of individuals with more severe phenotypic manifestations. However, with the widespread clinical availability of ES, individuals with rare missense or intronic *HNRNPK* variants presenting with only subtle facial features and/or without significant organ involvement are now being identified. This highlights a need for clinical and molecular tools to clarify clinical diagnostic suspicion and variants of uncertain significance.

Although the facies of AKS are considered recognizable to experienced dysmorphologists, there is considerable overlap with other conditions, such as Kabuki syndrome (KS).[6] Individuals with AKS with subtle features may be difficult to identify or differentiate from these other genetic syndromes. Additionally, facial features evolve with aging and may become more "classic" over time. Younger individuals tend to have a rounder face, fuller cheeks, a smaller nose, and a shorter columella. A clinical diagnosis of AKS may be difficult during infancy.

To assist in clinically identifying AKS, we defined six facial features as being most helpful and specific in supporting a clinical diagnosis (see Table 2). KS has been recognized by many as being the condition with the most obvious overlap with AKS,[1,6,7] and in scoring a large cohort of KS-affected individuals, we supported the specificity of the AKS diagnostic criteria (see Table S3).

Non-facial characteristics, such as cleft palate and congenital heart disease, were also considered for the diagnostic criteria; however, these features are often present in other syndromes (e.g., KS) and were less helpful in differentiating AKS. Although not included in the diagnostic criteria, a combination of these less-specific features (e.g., vertebral segmentation anomalies, craniosynostosis, hyporeflexia, and autonomic dysfunction) may help support a diagnosis of AKS. While non-specific, hypotonia and GDD/ID are required for clinical diagnosis, as the absence of these features would be atypical for AKS.

Table 3. Summary of clinical features seen in the cohort of individuals with Au-Kline syndrome

Clinical feature		Total proportion (%)	LoF individuals proportion (%), [proportion in DNAm individuals; proportion in phenotype only individuals]	Missense individuals proportion (%), [proportion in DNAm individuals; proportion in phenotype only individuals]	Intronic/indel individuals (%)
CNS involvement	global developmental delay/ intellectual disability	32/32(100%)	14/14 (100%) [9/9; 5/5]	13/13 (100%) [11/11; 2/2]	5/5 (100%) [4/4, 1/1]
	hypotonia	32/32 (100%)	14/14 (100%) [9/9, 5/5]	13/13 (100%) [11/11; 2/2]	5/5 (100%) [4/4, 1/1]
	CNS malformations[a]	15/30 (50%)	8/13 (62%) [5/9; 3/4]	5/12 (42%) [5/11; 0/1]	2/5 (40%) [1/4, 1/1]
	abnormal myelination	3/31 (10%)	3/13 (23%) [2/9; 1/4]	0/13 (0%) [0/9; 0/2]	0/5 (0%) [0/4, 0/1]
	hypo- or areflexia	12/31 (39%)	7/13 (54%) [4/9; 3/4]	2/13 (15%) [2/11; 0/2]	3/5 (60%) [2/4, 1/1]
	high pain tolerance	11/30 (37%)	8/13 (62%) [5/9; 3/4]	3/13 (23%) [2/11; 1/2]	0/4 (0%) [0/4]
	seizures	1/31 (3%)	1/13 (8%) [1/9; 0/4]	0/13 (0%) [0/11; 0/2]	0/5 (0%) [0/4, 0/1]
Facies[b]		26/31 (84%)	13/13 (100%) [9/9; 4/4]	8/13 (69%) [7/11; $^1/_2$ 1/2]	5/5 (100%) [4/4, 1/1]
Inverted nipples		11/28 (39%)	5/10 (50%) [5/9; 0/1]	1/13 (8%) [1/11; 0/2]	5/5 (100%) [4/4, 1/1]
Congenital heart defect[c]		20/32 (62%)	12/14 (86%) [7/9;5/5]	3/13 (23%) [3/11; 0/2]	5/5 100%) [4/4, 1/1]
Aortic dilation		3/31 (10%)	2/13 (15%) [1/9; 1/4]	1/13 (8%) [1/11; 0/2]	0/5 (0%) [0,4, 0/1]
Palate involvement[d]		16/31 (52%)	8/13 (62%) [6/9; 2/4]	6/13 (46%) [5/11; 0/2]	2/5 (40%) [2/4, 0/1]
Cleft palate		8/31 (26%)	6/13 (46%) [4/9; 2/4]	0/13 (0%) [0/11; 0/2]	2/5 (40%) [2/4, 0/1]
Bifid uvula		5/31 (16%)	2/13 (15%) [2/9; 0/4]	2/13 (15%) [2/11; 0/2]	1/5 (20%) [0/4, 1/1]
High-arched palate		6/31 (19%)	3/13 (23%) [1/9; 2/4]	3/13 (23%) [3/11; 0/2]	0/5 (0%) [0/4, 0/1]
Gastrointestinal (GI) issues[e]		22/31 (71%)	12/13 (92%) [8/9; 4/4]	6/13 (46%) [6/11; 0/2]	4/5 (80%) [3/4, 1/1]
Musculoskeletal	scoliosis	12/31 (39%)	9/13 (69%) [7/9; 2/4]	2/13 (15%) [2/11; 0/2]	1/5 (20%) [0/4, 1/1]
	vertebral segmentation defects	3/31 (10%)	3/13 (23%) [3/9; 0/4]	0/13 (0%) [0/11; 0/2]	0/5 (0%) [0/4, 0/1]
	other skeletal anomalies[f]	22/32 (69%)	11/14 (79%) [7/9; 4/5]	8/13 (62%) [6/11; 2/2]	3/5 (60%) [3/4, 0/1]
	joint hypermobility	19/32 (59%)	8/14 (57%) [3/9; 5/5]	9/13 (69%) [8/11; 1/2]	2/5 (40%) [1/4, 1/1]
	muscle weakness	14/31 (45%)	5/13 (38%) [4/9; 1/4]	7/13 (54%) [6/11; 1/2]	2/5 (40%) [2/4, 0/1]
Metopic ridging		16/32 (50%)	9/14 (64%) [7/9; 2/5]	4/13 (31%) [4/11; 0/2]	3/5 (60%) [2/4, 1/1]
Craniosynostosis		7/31 (23%)	4/13 (31%) [3/9; 1/4]	1/13 (8%) [1/11; 0/2]	2/5 (40%) [1/4, 1/1]
Genitourinary (GU) system	cryptorchidism (males only)	13/17 (76%)	7/9 (78%) [4/6; 3/3]	4/6 (67%) [4/5; 0/1]	2/2 (100%)
	hydronephrosis	16/31 (52%)	9/13 (69%) [6/9; 3/4]	4/13 (31%) [3/11; 1/2]	3/5 (60%) [3/4, 0/1]
Hearing loss (sensorineural and/or conductive)		5/22 (23%)	3/13 (23%) [2/9; 1/4]	1/11 (8%) [1/9; 0/2]	1/5(20) [0/4, 1/1]

Divided into sub-groups of individuals confirmed with AKS DNA methylation signature with (1) a loss-of-function (LoF) or deletion variant in *HNRNPK* (n = 9), (2) a missense variant in *HNRNPK* (n = 11), and (3) non-canonical intronic/indel/splice variants (n = 4). The phenotype-only cohort is also included with LoF n = 5 and missense n = 2 and intronic n = 1.

[a]CNS malformations include hypoplasia of the corpus callosum, abnormal morphology of the corpus callosum, heterotopia, spinal syrinx, hypomyelination, delayed myelination, cyst in corpus callosum, and spinal multicyst dilatation.

[b]Facies: meets facial criteria if having five or six of the following: long palpebral fissures, shallow orbits, a broad nose with wide nasal bridge, downturned corners of the mouth with upper lip M shape, and a deeply grooved or bifid tongue.

[c]Congenital heart defect: can be of any type; reported cardiac defects include ASD, VSD, AVSD, tetralogy of Fallot, bicuspid aortic valve, coarctation of the aorta, left ventricular hypertrophy, double outlet right ventricle, and dilation of the ascending aorta.

[d]Palate involvement can include cleft palate, high-arched palate, bifid uvula, and short immobile palate.

[e]GI (gastrointestinal) issues include constipation, feeding difficulties, gastroesophageal reflux, and tube feeding.

[f]Other skeletal anomalies include some or all of the following: hip dysplasia, pes planus, talipes, polydactyly, genu recurvatum, coxa valga, a segmented manubrium, and reduced bone mineral density.

In contrast to the clinical diagnostic score, the severity score is not considered specific to AKS and was developed to reflect the breadth and degree of disease burden for AKS. Consequently, the severity score is primarily based on the number of malformations requiring significant intervention (e.g., surgery) and the impact of functional

deficits (e.g., requirement of a feeding tube or mobility or communication impairment). We noted a trend where individuals with LoF variants were more likely to have moderate to severe systemic involvement, whereas individuals with missense variants tended to have milder involvement. The increased severity associated with LoF versus missense appears to be due primarily to the malformation burden; however, whether there is correlation of variant type with developmental outcomes may become clearer as more individuals at different ages are ascertained.

Historically, clinical diagnostic criteria were used to diagnose rare syndromes when genetic testing such as ES was not readily available and to identify individuals for appropriate targeted testing. With increased access to hypothesis-free broad-based genetic testing, clinical criteria are now helpful in reverse phenotyping to assist in determining the likelihood that an individual's phenotype matches the diagnosis suggested by a variant. The availability of a detectable "molecular phenotype," such as a specific DNAm signature, is also invaluable in this context. Previously reported DNAm signatures have already demonstrated this utility in predicting the pathogenicity of VUSs.[10–12]

Using a discovery set of individuals with AKS defined by clinical diagnostic criteria and the presence of *HNRNPK* LoF variants, we identified a highly sensitive and specific AKS DNAm signature. We demonstrate that the AKS signature has 100% specificity and is distinct from the DNAm signature associated with KS. This AKS signature can also differentiate the functional effects of different types of variants in *HNRNPK* with two previously validated classification models, specifically correlation based[11,12] and machine learning.[10,12,14]

Variant interpretation is challenging, particularly when an individual with a rare *de novo* missense *HNRNPK* variant presents with a mild phenotype and/or does not meet diagnostic criteria for AKS. Here, we have overcome a major challenge in clarifying the pathogenic significance of these VUSs in *HNRNPK* by using this functional DNAm assay, which robustly supports variant classification.

The highly specific AKS signature allowed reclassification of six individuals with *de novo* rare missense variants in *HNRNPK* as not AKS. For some individuals, this suggests that the identified variant may be more likely benign (e.g., *HNRNPK* variant, c.740G>A [p.Arg247His] [GenBank: NM_002140.4] for P17), as the individual already appears non-syndromic and *in silico* predictors suggest the variant is not deleterious (Table S1). For other individuals, it is less clear whether the *HNRNPK* variant is likely benign versus potentially associated with a separate condition related to *HNRNPK*.

The AKS DNAm signature, particularly in combination with the syndrome definition provided by the diagnostic criteria, allows us to report on a large cohort of individuals with missense variants in *HNRNPK* (n = 13, 11 with AKS DNAm signature plus two additional individuals in the phenotype study with the recurrent p.Glu85Lys variant). We thus describe with high confidence an expanded phenotypic spectrum of AKS. On the basis of these data, we establish that AKS has a wide range of presentations that may include more subtle facies and the absence of organ malformations (see Table 3).

The DNAm analysis also identified an interesting subset of individuals with an "intermediate" AKS signature (SVM prediction scores of 34%–52%). In this group, the impacted CpG sites remain consistent with an AKS signature but with a lesser degree of hypomethylation or hypermethylation difference in the same direction. Except for P9 (*HNRNPK* variant, c.213+5G>A [GenBank: NM_002140.4]), all the individuals with an intermediate DNAm score harbored an *HNRNPK* missense variant. It is clear, however, that this intermediate DNAm signature is still well within the spectrum of AKS and not associated with a separate *HNRNPK* related condition, as multiple individuals who clearly have AKS by diagnostic criteria classified with an intermediate DNAm score (P5, 6, 7, 9, 10, 12, and 22). Furthermore, one of the recurrent *HNRNPK* variants, c.137G>T (p.Arg46Leu) (GenBank: NM_002140.4),classified as positive for the AKS signature in one individual (P2), whereas it classified as intermediate for the AKS signature in a different individual (P7). There are few published reports of DNA methylation variation for recurrent sequence variants. We have previously reported in Choufani et al., 2020,[12] that the same variant within a family or across many unrelated families can vary by about 20% in their SVM classification. This is comparable to a second recurrent *HNRNPK* variant, c.253G>A (p.Glu85Lys) (GenBank: NM_002140.4), reported in three AKS-affected individuals, which showed about 15% variation within the three unrelated individuals. In contrast, the difference in the SVM score between P2 and P7 with the recurrent *HNRNPK* variants, c.137G>T (p.Arg46Leu) (GenBank: NM_002140.4),is ~40%. Further molecular follow-up will clarify the degree of variation that occurs in DNAm profiles of individuals that carry recurrent sequence variants.

The presence of an intermediate DNAm signature suggests possible DNAm signature-phenotype correlation and may provide insight into the pathophysiological origin of the trends noted earlier for LoF and missense variants, where individuals with missense variants often have milder disease severity in comparison to those with LoF variants. The individuals with missense variants are also likely to have an intermediate DNAm score, although it is notable that some individuals with a positive SVM for AKS signature can also have mild phenotypic severity.

Modifiers, either genetic, environmental, and/or age-related, are most likely influencing both the DNAm signature and clinical phenotype, and these are not yet understood. There can be discordance in the DNAm signature for individuals with the same variant, as previously discussed for P2 (positive) and P7 (intermediate), who both have the same mild severity score of 3. There is also a wide range in severity scores for individuals with loss-of-function variants (ranging from 3 to 9). We highlight

two individuals (AKS2 and AKS12) who have the same known *HNRNPK* splice variant, c.257G>A (GenBank: NM_002140.4), but have severity scores of 9 and 3, respectively. Similar observations have been made for other genetic conditions, where individuals with the same pathogenic variant can have extremely variable expressivity. For example, significant intra-familial variability is seen in neurodevelopmental conditions such as 22q11 deletion syndrome or neurofibromatosis type 1 (MIM: 162200) and in malformation conditions such as branchio-oto-renal (MIM: 113650)[29] or Van der Woude syndromes (MIM: 119300).[30] The degree of variability in the DNAm signature as observed here, in association with the same *HNRNPK* variant (c.137G>T [p.Arg46Leu] [GenBank: NM_002140.4]), has not been reported previously. We expect further multi-omic studies in other tissues, including genome sequencing and RNA and protein analyses, will elucidate these issues. While we speculate with caution that the DNAm signature may become helpful in predicting clinical outcomes for AKS, at this time, it is still challenging to use DNAm as an outcome prediction tool for single unique individuals.

There are known gene domain-specific DNAm signatures, where there may be multiple phenotypically separate conditions with unique signatures for the same gene. For *SMARCA2*, variants clustering outside the helicase domain are associated with a unique DNAm signature and phenotype that is distinct from Nicolaides-Baraitser syndrome (MIM: 601358), which is typically caused by variants clustering within the helicase domains.[13,31] Truncating variants in exons 33 or 34 in *SRCAP* are associated with a DNAm signature specific to Floating Harbor Syndrome (FLHS) (MIM: 136140), whereas proximal variants in *SRCAP* are associated with a separate DNAm signature and neurodevelopmental disorder without the typical facial gestalt of FLHS.[14] For *HNRNPK*, we identified two individuals with dysmorphic facies and ID who had missense variants affecting adjacent residues (P14 *HNRNPK* variant, c.176G>A [p.Gly59Glu] [GenBank: NM_002140.4]; P19 *HNRNPK* variant, c.173T>C [p.Ile58Thr] [GenBank: NM_002140.4]), who did not classify as having the AKS signature, which is in contrast to the missense variants in the AKS intermediate signature group. While it is possible these individuals have a condition that is unrelated to *HNRNPK*, we also wonder whether there could be a separate, non-AKS, but still *HNRNPK*-associated, condition. Additional individuals will be required to clarify this possibility.

The functional roles of the genomic regions within the AKS-associated *HNRNPK* signature are of great importance, as they demonstrate enrichment in CpG island shores where DNA methylation levels are susceptible to change under several conditions, such as tissue differentiation, reprogramming, aging, and disease.[32] CpG island shores are defined as 2-kb-long regions that lie on both sides of a CpG island.[32] Interestingly, AKS signature CpG sites correlated with marks of open chromatin and active transcription, including DNase hypersensitive sites (HSSs), further highlighting the potential functional significance of the underlying signature to inform on the pathophysiological role of *HNRNPK* in AKS. Gene set enrichment highlighted functional categories associated with brain and skeletal development, identifying specific genes that are highly expressed in the relevant tissues as candidate deregulated genes in the pathogenesis of AKS.

While it is clear that there are multiple complementary approaches to interrogate disorder pathophysiology, we posit that insight from DNAm is inherently valuable, as previously described for Sotos syndrome and Nicolaides-Baraitser syndrome.[11,13] These disorders, similar to AKS, impact multiple tissues in the body from early development, leading to broad constellations of clinical features affecting multiple organs. As *HNRNPK* is ubiquitously expressed, with important roles in a wide spectrum of tissues, we suggest that some DNAm patterns are established early in development and may be maintained for a lifetime and provide an opportunistic window to identify candidate disease biomarkers that reflect important developmental molecular changes relevant to disease pathophysiology.

HNRNPK belongs to a molecularly related gene family of hnRNPs. Many of the genes in this family have been implicated in neurodevelopmental disorders (NDDs). Interestingly, we found that ~85% of pathogenic *HNRNPK* missense variants (n = 11/13) cluster in the K homology RNA-binding domain, which is important for RNA binding and recognition, as previously reported.[33] Clustering of variants in RNA or DNA interaction domains have also been observed for other hnRNPs (e.g., *HNRNPUL1* [MIM: 605800][33]). It will be important to test the specificity of this DNAm signature against NDD phenotypes caused by pathogenic variants in other hnRNPs, such as *HNRNPU* (MIM: 602869), *HNRNPH1* (MIM: 601035), and *HNRNPR* (MIM: 607201).[33] Such comparisons may elucidate our understanding of the clinical overlap and differences in these related conditions.

In summary, we have developed clinical diagnostic criteria, which refine the most specific features of AKS. These criteria have utility in not only identifying individuals likely to have AKS but also support reverse phenotyping when VUSs in *HNRNPK* are found. We have also identified a highly sensitive and specific AKS DNAm signature that is a valuable molecular functional tool that can be used to classify VUSs in *HNRNPK*. We show how the integration of data from the DNAm signature and clinical criteria contribute to a high confidence expansion of the phenotypic spectrum and pathophysiological insights into genotype-phenotype correlations. Although initially described as a multiple malformation syndrome with a recognizable facial gestalt, AKS now includes individuals with only subtle craniofacial differences and relatively few, or only minor, malformations. This study highlights the considerable potential of a combined clinical and molecular diagnostic approach. We anticipate that this

integrated approach will be further facilitated by the use of machine learning techniques that support deep phenotyping with facial recognition and/or electronic medical record data. In addition, emerging technologies, such as long-read sequencing, will in the future detect DNAm alterations and sequence variants concurrently.

Data and code availability

The datasets supporting the current study have not been deposited in a public repository because of institutional ethics restrictions. All software and R packages used in the study are publicly available as described in the subjects and methods section and can also be accessed through EpigenCentral at https://epigen.ccm.sickkids.ca.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.08.014.

Acknowledgments

We are grateful to all the families who participated in this research and to the many clinicians who recruited them into the study. We acknowledge the technical assistance of Youliang Lou and Chunhua Zhao. This work was supported by a Canadian Institutes of Health Research (CIHR) grants to R.W. (PJT-178315) and the Ontario Brain Institute (Province of Ontario Neurodevelopmental Disorders [POND] network [IDS11-02]) grants to R.W. This work was also supported by a grant from SFARI (887172 for R.W.). F.A., H.A., and Z.R. acknowledge the King Salman Center for Disability Research for funding this work through Research Group no. RG-2022-010. J.T.-C., J.N., and P.L. were funded through FEDER funds PI20/01053. The work of G.E. was supported by the Jacob Goldfield Foundation and the RTW Charitable Foundation. P.Y.B.A. acknowledges the Rare Disease Foundation, which provided funding for AKS phenotype studies, and the Care4Rare Consortium, which supported the initial gene discovery. See supplemental information for detailed acknowledgment statements regarding the DDD study, the 100,000 Genomes Project, UDN, and the NIH Common Fund.

Declaration of interests

H.T.B. is a consultant for Mahzi therapeutics.

Received: May 27, 2022
Accepted: August 29, 2022
Published: September 20, 2022

Web resources

CADD, https://cadd.gs.washington.edu/snv.
GenBank, https://www.ncbi.nlm.nih.gov/genbank/.
Metascape, https://metascape.org.
OMIM, http://www.omim.org/.
ProteinPaint, https//proteinpaint.stjude.org.
TSEA and CSEA, http://genetics.wustl.edu/jdlab/tsea/.

References

1. Au, P.Y.B., You, J., Caluseriu, O., Schwartzentruber, J., Majewski, J., Bernier, F.P., Ferguson, M., Care for Rare Canada Consortium, Valle, D., Parboosingh, J.S., et al. (2015). GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue abnormalities caused by de novo variants in HNRNPK. Hum. Mutat. *36*, 1009–1014.
2. Okamoto, N. (2019). Okamoto syndrome has features overlapping with Au-Kline syndrome and is caused by HNRNPK mutation. Am. J. Med. Genet. *179*, 822–826.
3. Hancarova, M., Puchmajerova, A., Drabova, J., Karaskova, E., Vlckova, M., and Sedlacek, Z. (2015). Deletions of 9q21.3 including NTRK2 are associated with severe phenotype. Am. J. Med. Genet. *167A*, 264–267.
4. Maystadt, I., Deprez, M., Moortgat, S., Benoît, V., and Karadurmus, D. (2020). A second case of Okamoto syndrome caused by HNRNPK mutation. Am. J. Med. Genet. *182*, 1537–1539.
5. Pua, H.H., Krishnamurthi, S., Farrell, J., Margeta, M., Ursell, P.C., Powers, M., Slavotinek, A.M., and Jeng, L.J.B. (2014). Novel interstitial 2.6 Mb deletion on 9q21 associated with multiple congenital anomalies. Am. J. Med. Genet. *164A*, 237–242.
6. Lange, L., Pagnamenta, A.T., Lise, S., Clasper, S., Stewart, H., Akha, E.S., Quaghebeur, G., Knight, S.J.L., Keays, D.A., Taylor, J.C., et al. (2016). A de novo frameshift in HNRNPK causing a Kabuki-like syndrome with nodular heterotopia. Clin. Genet. *90*, 258–262.
7. Dentici, M.L., Barresi, S., Niceta, M., Pantaleoni, F., Pizzi, S., Dallapiccola, B., Tartaglia, M., and Digilio, M.C. (2018). Clinical spectrum of Kabuki-like syndrome caused by HNRNPK haploinsufficiency. Clin. Genet. *93*, 401–407.
8. Au, P.Y.B., Goedhart, C., Ferguson, M., Breckpot, J., Devriendt, K., Wierenga, K., Fanning, E., Grange, D.K., Graham, G.E., Galarreta, C., et al. (2018). Phenotypic spectrum of Au-Kline syndrome: a report of six new cases and review of the literature. Eur. J. Hum. Genet. *26*, 1272–1281.
9. Yamada, M., Shiraishi, Y., Uehara, T., Suzuki, H., Takenouchi, T., Abe-Hatano, C., Kurosawa, K., and Kosaki, K. (2020). Diagnostic utility of integrated analysis of exome and transcriptome: Successful diagnosis of Au-Kline syndrome in a patient with submucous cleft palate, scaphocephaly, and intellectual disabilities. Mol. Genet. Genomic Med. *8*, e1364.
10. Butcher, D.T., Cytrynbaum, C., Turinsky, A.L., Siu, M.T., Inbar-Feigenberg, M., Mendoza-Londono, R., Chitayat, D., Walker, S., Machado, J., Caluseriu, O., et al. (2017). CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. Am. J. Hum. Genet. *100*, 773–788.
11. Choufani, S., Cytrynbaum, C., Chung, B.H.Y., Turinsky, A.L., Grafodatskaya, D., Chen, Y.A., Cohen, A.S.A., Dupuis, L., Butcher, D.T., Siu, M.T., et al. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun. *6*, 10207.
12. Choufani, S., Gibson, W.T., Turinsky, A.L., Chung, B.H.Y., Wang, T., Garg, K., Vitriolo, A., Cohen, A.S.A., Cyrus, S., Goodman, S., et al. (2020). DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes. Am. J. Hum. Genet. *106*, 596–610.

13. Chater-Diehl, E., Ejaz, R., Cytrynbaum, C., Siu, M.T., Turinsky, A., Choufani, S., Goodman, S.J., Abdul-Rahman, O., Bedford, M., Dorrani, N., et al. (2019). New insights into DNA methylation signatures: SMARCA2 variants in Nicolaides-Baraitser syndrome. BMC Med. Genomics *12*, 105.
14. Rots, D., Chater-Diehl, E., Dingemans, A.J.M., Goodman, S.J., Siu, M.T., Cytrynbaum, C., Choufani, S., Hoang, N., Walker, S., Awamleh, Z., et al. (2021). Truncating SRCAP variants outside the Floating-Harbor syndrome locus cause a distinct neurodevelopmental disorder with a specific DNA methylation signature. Am. J. Hum. Genet. *108*, 1053–1068.
15. Courraud, J., Chater-Diehl, E., Durand, B., Vincent, M., Del Mar Muniz Moreno, M., Boujelbene, I., Drouot, N., Genschik, L., Schaefer, E., Nizon, M., et al. (2021). Integrative approach to interpret DYRK1A variants, leading to a frequent neurodevelopmental disorder. Genet. Med. *23*, 2150–2159.
16. Rouxel, F., Relator, R., Kerkhof, J., McConkey, H., Levy, M., Dias, P., Barat-Houari, M., Bednarek, N., Boute, O., Chatron, N., et al. (2022). CDK13-related disorder: Report of a series of 18 previously unpublished individuals and description of an epigenetic signature. Genet. Med. *24*, 1096–1107.
17. Bend, E.G., Aref-Eshghi, E., Everman, D.B., Rogers, R.C., Cathey, S.S., Prijoles, E.J., Lyons, M.J., Davis, H., Clarkson, K., Gripp, K.W., et al. (2019). Gene domain-specific DNA methylation episignatures highlight distinct molecular entities of ADNP syndrome. Clin. Epigenetics *11*, 64.
18. Barboro, P., Ferrari, N., and Balbi, C. (2014). Emerging roles of heterogeneous nuclear ribonucleoprotein K (hnRNP K) in cancer progression. Cancer Lett. *352*, 152–159.
19. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.
20. Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzetinova, T., et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet *385*, 1305–1314.
21. Jim, D., Dennys, M., Elbahy, L., Fowler, T., Hill, S., Hubbard, T., Jostins, L., Maltby, N., Mahon-Pearson, J., McVean, G., Nevin-Ridley, K., et al. (2017). The National Genomics Research and Healthcare Knowledgebase (figshare. Dataset).
22. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.
23. Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.
24. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
25. Salas, L.A., Koestler, D.C., Butler, R.A., Hansen, H.M., Wiencke, J.K., Kelsey, K.T., and Christensen, B.C. (2018). An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol. *19*, 64.
26. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523.
27. Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res. *38*, 4218–4230.
28. Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.
29. Smith, R.J.H. (1993). Branchiootorenal spectrum disorder. In GeneReviews((R)), M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K.W. Gripp, G.M. Mirzaa, and andA. Amemiya, eds..
30. Busche, A., Hehr, U., Sieg, P., and Gillessen-Kaesbach, G. (2016). Van der Woude and Popliteal Pterygium Syndromes: Broad intrafamilial variability in a three generation family with mutation in IRF6. Am. J. Med. Genet. *170*, 2404–2407.
31. Cappuccio, G., Sayou, C., Tanno, P.L., Tisserant, E., Bruel, A.L., Kennani, S.E., Sa, J., Low, K.J., Dias, C., Havlovicová, M., et al. (2020). De novo SMARCA2 variants clustered outside the helicase domain cause a new recognizable syndrome with intellectual disability and blepharophimosis distinct from Nicolaides-Baraitser syndrome. Genet. Med. *22*, 1838–1850.
32. Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.
33. Gillentine, M.A., Wang, T., Hoekzema, K., Rosenfeld, J., Liu, P., Guo, H., Kim, C.N., De Vries, B.B.A., Vissers, L.E.L.M., Nordenskjold, M., et al. (2021). Rare deleterious mutations of HNRNP genes result in shared neurodevelopmental disorders. Genome Med. *13*, 63.