Author/s:
Lubin, EE;Gonzalez, EM;Sangree, AK;Durham, EL;Klinkhammer, H;Li, JM;Smith, SM;Layo-Carris, DE;Clark, KJ;Melendez-Perez, AJ;Wang, XM;Angireddy, R;Weiss, EE;Barakat, TS;Mercier, S;Cogné, B;Koene, S;Hilhorst-Hofstee, Y;Rydzanicz, M;Ploski, R;de los Ángeles Gómez Cano, M;Palomares-Bralo, M;Arévalo, TB;Tan, TY;Gallacher, L;MacFarland, SP;Ahrens-Nicklas, RC;Nomakuchi, TT;Bhoj, EJK

Title:
Coupling deep phenotypic quantification with next-generation phenotyping for 192 individuals with germline histonopathies

Date:
2025-07-10

Citation:
Lubin, E. E., Gonzalez, E. M., Sangree, A. K., Durham, E. L., Klinkhammer, H., Li, J. M., Smith, S. M., Layo-Carris, D. E., Clark, K. J., Melendez-Perez, A. J., Wang, X. M., Angireddy, R., Weiss, E. E., Barakat, T. S., Mercier, S., Cogné, B., Koene, S., Hilhorst-Hofstee, Y., Rydzanicz, M. ,... Bhoj, E. J. K. (2025). Coupling deep phenotypic quantification with next-generation phenotyping for 192 individuals with germline histonopathies. Human Genetics and Genomics Advances, 6 (3), pp.100440-. https://doi.org/10.1016/j.xhgg.2025.100440.

Persistent Link:
https://hdl.handle.net/11343/360509

HGG Advances

ARTICLE

Coupling deep phenotypic quantification with next-generation phenotyping for 192 individuals with germline histonopathies

Emily E. Lubin,[1,2] Elizabeth M. Gonzalez,[1,2,17] Annabel K. Sangree,[1,2,17] Emily L. Durham,[2,3] Hannah Klinkhammer,[4,5] Jing-Mei Li,[4] Sarina M. Smith,[1,2] Dana E. Layo-Carris,[2] Kelly J. Clark,[1,2] Ashley J. Melendez-Perez,[1,2] Xiao Min Wang,[2] Rajesh Angireddy,[2] Erin E. Weiss,[2] Tahsin Stefan Barakat,[6] Sandra Mercier,[7,8] Benjamin Cogné,[7,8] Saskia Koene,[9] Yvonne Hilhorst-Hofstee,[9] Malgorzata Rydzanicz,[10] Rafal Ploski,[10] María de los Ángeles Gómez Cano,[11] María Palomares-Bralo,[11,12] Tania Barragán Arévalo,[13] Tiong Yang Tan,[14,15] Lyndon Gallacher,[14,15] Suzanne P. MacFarland,[1,16] Rebecca C. Ahrens-Nicklas,[1,2] Tomoki T. Nomakuchi,[2] and Elizabeth J.K. Bhoj[1,2,18,*]

Summary

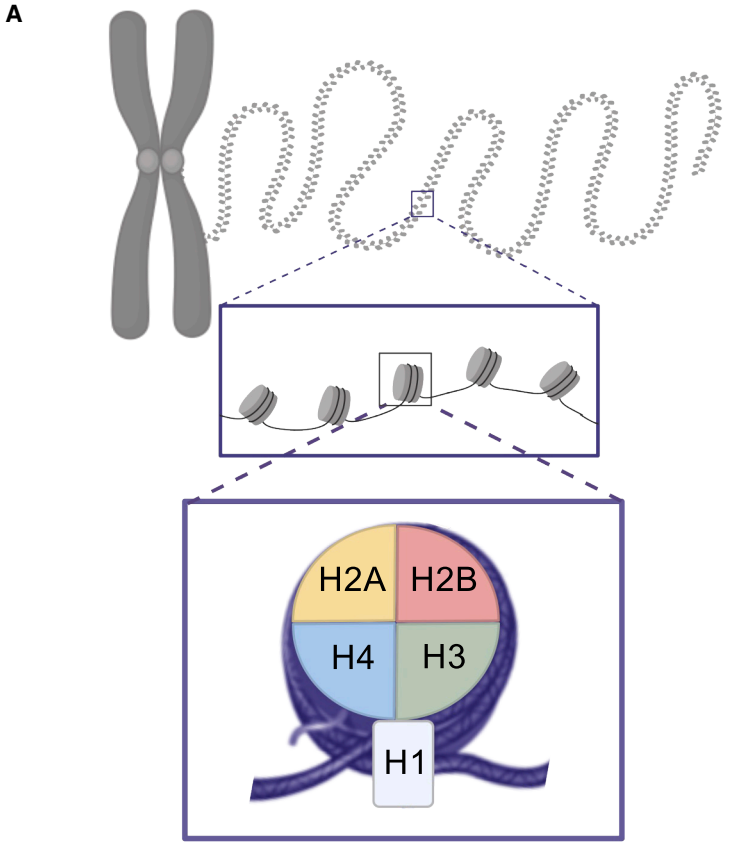
Mendelian histonopathies are rare neurodevelopmental disorders (NDDs) caused by germline variants in histone-encoding genes. Here, we perform a more expansive pan-histonopathy interrogation than previously possible. We analyze data from 192 individuals affected by histonopathies. This analysis includes representation of the 185 published individuals with HIST1H1E syndrome, Bryant-Li-Bhoj syndrome, and Tessadori-Bicknell-van Haaften NDD; as well as from seven unpublished individuals, five of whom harbor variants in genes not previously associated with disease (*HIST1H2AL/H2AC16, H2AFZ/H2AZ1, HIST1H3D/H3C4,* and *HIST3H3/H3-4*). By intersecting clinician-reported phenotypic data with next-generation phenotyping of published 2D facial photographs ($n = 98$), we sought to address the lack of established craniofacial gestalts or characteristic phenotypic patterns for this community. While these analyses may suggest a histone core versus linker protein basis of delineation, they more strikingly highlight data gaps that confound the identification of phenotypic patterns at this time. Based on this, we developed an updated standardized clinical survey, which allowed us to identify the second known individual with a germline histonopathy and a cancer diagnosis. Notably, the community-wide cancer incidence is currently 1%, which falls below the recommended 5% cut off for routine surveillance. Ultimately, this work highlights the ways in which histonopathy-associated phenotypes change throughout the lifespan, necessitating longitudinal re-evaluation; that every identified individual shapes our understanding of these syndromes in a way that improves care for this community; and the value of ongoing translational work to address the outstanding question of cancer predisposition for individuals living with germline histonopathies.

Introduction

Histones are core genome organizers that bind and wrap nuclear DNA, regulating gene expression and chromatin structure. Eight histone proteins come together to form the core nucleosome, made up of two copies each of H2A, H2B, H3, and H4 class histones (Figure 1A). A central tetramer, made up of two H3 proteins and two H4 proteins, sits at the core of the nucleosome octamer flanked by two H2A/H2B heterodimers.[1] The H1 linker protein adds an additional layer of regulation between nucleosomes.[2]

In addition to these five protein delineations, histones can be further divided into two classes: replication-coupled (RC) and replication-independent (RI) (Figure 1B). The expression of RC histones is linked to S-phase of the cell cycle while RI histones demonstrate temporal and tissue specificity. Further, while a single RC histone protein is encoded by many (10–20) genes that are clustered in the genome, each RI histone is encoded by a few (≤ 3) genes that are often found on different chromosomes. Structurally, RC histone-encoding genes lack introns and encode cellular mRNAs that are not polyadenylated, which stands in

[1]University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA; [2]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [3]Director of Research Engagement, The TBCK Foundation, Pittsburgh, PA, USA; [4]Institute for Genome Statistics and Bioinformatics, University Hospital Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany; [5]Institute for Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany; [6]Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands; [7]CHU de Nantes, Service de Génétique Médicale, Nantes, France; [8]L'institut Du Thorax, INSERM, CNRS, UNIV Nantes, Nantes, France; [9]Leiden University Medical Center, Leiden, the Netherlands; [10]Department of Medical Genetics, Medical University of Warsaw, Warsaw, Poland; [11]INGEMM-IdiPaz, Institute of Medical and Molecular Genetics, Madrid, Spain; [12]CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, Madrid, Spain; [13]Medical Genetics Department, Hospital Infantil de México Federico Gómez, Mexico City, Mexico; [14]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, VIC, Australia; [15]Department of Pediatrics, University of Melbourne, Parkville, VIC, Australia; [16]Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[17]These authors contributed equally
[18]Lead contact
*Correspondence: bhoje@chop.edu
https://doi.org/10.1016/j.xhgg.2025.100440.



B

RC H4 (non-polyadenylated)	RC H2A (non-polyadenylated)	RC H2B (non-polyadenylated)	RC H3 (non-polyadenylated)	RC H1 (non-polyadenylated)
H4C1	*H2AC1*	*H2BC1*	*H3C1*	*H1-1*
H4C2	*H2AC4*	*H2BC3*	*H3C2*	*H1-2*
H4C3	*H2AC6*	*H2BC4*	*H3C3*	*H1-3*
H4C4	*H2AC7*	*H2BC5*	***H3C4***	*H1-4*
H4C5	*H2AC8*	*H2BC6*	*H3C6*	*H1-5*
H4C6	*H2AC11*	*H2BC7*	*H3C7*	
H4C7	*H2AC12*	*H2BC8*	*H3C8*	
H4C8	*H2AC13*	*H2BC9*	*H3C10*	
H4C9	*H2AC14*	*H2BC10*	*H3C11*	
H4C11	*H2AC15*	*H2BC11*	*H3C12*	
H4C12	***H2AC16***	*H2BC12*	*H3C13*	
H4C12	*H2AC17*	*H2BC13*	*H3C14*	
H4C13	*H2AC18*	*H2BC14*	*H3C15*	
H4C14	*H2AC19*	*H2BC15*		
H4C15	*H2AC20*	*H2BC17*		
H4C16	*H2AC21*	*H2BC18*		
	H2AC25	*H2BC21*		
		H2BC26		

RI H2A (polyadenylated)	RI H2B (polyadenylated)	RI H1 (polyadenylated)	RI H3 (polyadenylated)
H2AZ1	*H2BW1*	*H1-0*	*H3-3A*
H2AZ2	*H2BW2*	*H1-6*	*H3-3B*
H2AX	*H2BK1*	*H1-7*	***H3-4***
MACROH2A1	*H2BN1*	*H1-8*	*H3-5*
MACROH2A2		*H1-10*	*H3-7*
H2AB1			*CENPA*
H2AB2			*H3Y1*
H2AB3			*H3Y2*
H2AJ			
H2AL1Q			
H2AL3			
H2AP			

C

	Gene	gnomAD constraint metrics								Monarch Initiative – orthologs with phenotypes	Available transcriptomic data in GTEx or BrainSpan (aka polyA)
		gnomAD synonymous z score		gnomAD missense z score		gnomAD pLI		gnomAD LOEUF score			
		v2.1.1	v4.0.0	v2.1.1	v4.0.0	v2.1.1	v4.0.0	v2.1.1	v4.0.0		
Previously published	*H1-4*	-16.76	-8.54	-6.72	-7.37	0.19	0	1.59	1.85	Sc, Dm, Mm	
	H3-3A	-0.33		3.16		0.12		1.16		Ce, Dm, Mm, Dr	Yes
	H3-3B	-7.71	-3.38	2.88	4	0.81	0.96	0.55	0.49		Yes
	H4C3	-8.71	-5.09	-0.02	-0.95	0	0	1.95	1.96		
	H4C4		-4.88		-1.83		0		1.95		
	H4C5		-6.41		0.07		0		1.98		
	H4C6		-4.4		-2.18		0		1.91		
	H4C9		-1.19		-0.06		0.08		1.64		
	H4C11	-8.19	-3	0.68	1.99	0.39	0.45	1.74	1.26		
Previously unreported	***H2AC16***	-4.6	-2.83	-0.17	-0.66	0	0	1.92	1.94		
	H2AZ1	-1.47	-0.49	2.92	3.12	0.88	0.95	0.43	0.4	Sc, Dm, Ce, Dr, Mm	Yes
	H3C4	-4.81	-3.48	0.38	-0.22	0	0	1.88	1.77		
	H3-4	-4.13	-2.66	-2	-2.04	0	0	1.94	1.9	Dm, Sc, Ce	Yes

Figure 1. Germline histonopathy-implicated genes

(A) Schematic overview of the different classes of histone proteins.

(B) Genes encoding the different classes of histone proteins (color-coded based on A), subdivided based on whether a gene encodes an RC histone (top) or RI histone (bottom). Purple highlight indicates genes reported in this paper as linked to germline histonopathies. Bolded are previously unreported.

(C) Integration of gnomAD constraint metrics and Monarch Initiative data on model systems with phenotypes when the human gene orthologue is perturbed. Unfilled fields indicate that data are not available. Gray = HIST1H1E syndrome-associated gene;

(legend continued on next page)

contrast to RI histone-encoding genes that contain introns and encode polyadenylated mRNA transcripts.[3]

Beyond the fundamental roles that histones play in cellular physiology, a role for mutated histones in neuropathology has emerged in the last decade. In 2012, somatic oncohistone mutations were reported to be recurrent in individuals with pediatric high-grade gliomas (HP:0009733).[4,5] Subsequently, driver and passenger mutations in histones were found in a range of tumor types.[3,6–10] A recent pancancer histone mutation atlas identified somatic histone mutations in 11% of analyzed cases, with the highest rates of somatic histone mutations in chondrosarcoma (HP:0006765), pediatric high-grade gliomas (HP:0009733), and lymphoma (HP:0002665).[11] This work also extended the spectrum of histone genes implicated in cancer to include all five protein delineations (H1, H2A, H2B, H3, and H4).

Germline histonopathy-causing variants were reported in 2014, establishing a class of Mendelian neurodevelopmental disorders (NDDs, HP:0012759), which complements the previously delineated Mendelian disorders of epigenetic machinery (MDEMs) that are associated with variants in proteins that read, write, erase, or remodel epigenetic marks on histones.[12] The germline variant in a histone-encoding gene, *HIST1H1E/H1-4* (OMIM: 142220), was reported to cause overgrowth (HP:0001548) and a Sotos-like phenotype (OMIM:117550) in a single individual.[13] In 2017, variants in *HIST1H1E/H1-4* were again reported to cause overgrowth (HP:0001548) and intellectual disability (HP:0001249) in five unrelated individuals, providing the foundation for the classification of the rare Mendelian NDD HIST1H1E syndrome (OMIM: 617537).[14] At the time of this report, a total of 185 individuals with germline variants in histone-encoding genes have now been published across 24 reports, including 55 individuals with HIST1H1E syndrome, affecting the RC H1 linker protein[13–26]; 96 individuals with Bryant-Li-Bhoj syndrome (BLBS [OMIM: 619720 and 619721]), affecting the core nucleosomal RI H3.3 histone protein[27–32]; and 34 individuals with Tessadori-Bicknell-van Haaften NDD (TBvH [OMIM: 619758, 619759, 619950, 619951]), affecting the core nucleosomal RC H4 histone protein.[33–36] With a community of this size, the addition of information from even one individual shapes the clinical understanding of the syndrome and informs translational mechanistic research. Here, we build on a prior pan-histonopathy analysis of 130 individuals with HIST1H1E syndrome, BLBS, and TBvH NDD by including 55 subsequently published individuals with OMIM-identified histonopathies as well as seven previously unpublished individuals with germline histonopathies identified through GeneMatcher and DECIPHER collaborations.[37–39]

Of the previously unreported individuals, two harbor BLBS-causative variants in *H3F3B/H3-3B* while five harbor variants in previously unreported genes (*HIST1H2AL/H2AC16, H2AFZ/H2AZ1, HIST1H3D/H3C4,* and *HIST3H3/H3-4)*, otherwise known as genes of uncertain significance (GUS).[38] These seven unpublished individuals are included in this pan-histonopathy analysis for numerous reasons. In this era of genome sequencing, the pace of discovery of gene-disease associations is estimated to be roughly one per day.[39] Approximately 87% of these disorders are discovered using next-generation sequencing approaches. However, 60% of individuals who opt for clinical exome or genome sequencing receive a non-diagnostic report. This is due in part to the gaps in current understanding of the function of all the genes in the human genome, as well as the fact that establishing a GUS as a gene associated with Mendelian disease is extremely challenging without a prior link between perturbation of that gene and human disease. The field of translational genetics has yet to reach a clear consensus to define the best path to gene discovery for rare Mendelian disorders, so we turn to current evidence-based strategies.[40] Inclusion of all 192 individuals in this analysis advances our understanding of the role of histones in human health, and facilitates community and support for affected individuals and their families, for whom currently only nonspecific palliative medical care and reactive comorbidity management are currently available.

In this expanded pan-histonopathy analysis, we seek to address outstanding questions that impact the diagnostic odyssey and proactive clinical management of individuals living with germline histonopathies. Between 30% and 40% of Mendelian syndromes are known to be associated with a characteristic pattern of craniofacial features, known as a craniofacial gestalt.[41] Identifying and defining a pattern of syndrome-specific craniofacial features can aid in shortening the diagnostic odyssey for families awaiting a rare or ultra-rare disease diagnosis.[42] Thus, we first interrogated whether existing phenotypic data could elucidate a craniofacial gestalt for individuals with germline histonopathies. To accomplish this, we combined the power of clinician-reported craniofacial phenotypic information with next-generation phenotyping (NGP). Clinical geneticists are highly trained in identifying facial gestalts. However, recognition of rare or ultra-rare syndromes, such as germline histonopathies, can be very challenging if a clinician has never cared for one of these individuals. Additionally, a robust body of literature demonstrates that the ethnicities of both the clinician and the affected individual significantly impact what features are reported.[43–45] This issue is exacerbated by the fact that most clinical training atlases are heavily biased toward

green = BLBS-associated gene; blue = TBvH NDD-associated gene. Deep orange indicates constraint metric values flagged in gnomAD to be statistically significant. Pale orange indicates constraint metric values with a Z score ≥ 2 away from the mean. Abbreviations: RC = replication-coupled; RI = replication-independent; Sc = *Saccharomyces cerevisiae*; Ce = *Caenorhabditis elegans*; Dr = *Danio rerio*; Dm = *Drosophila melanogaster*; Mm = *Mus musculus*.

individuals of European ancestry with very little representation of other ethnicities. NGP employs computational facial analysis technology on 2D facial photographs to identify syndrome-specific patterns in facial features and has been shown in prior reports to outperform clinicians when identifying patterns in geographically and ethnically diverse populations.[42,45] NGP platforms, such as the GestaltMatcher Database (GMDB), employ image encoders that convert all features of a facial image into vectors, which enable characteristic patterns of facial features to be recognized not just for syndromes included in its training set, but also on ultra-rare disorders, rendering this machine learning algorithm the ideal platform to employ for individuals with germline histonopathies.[42,46,47]

Phenotypic patterns arising from germline variants found in every cell of the body are not restricted to the craniofacial complex and impact the development of all organ systems. Further, recent studies have reported that children with multiple congenital anomalies arising from genetic etiologies are more likely to be diagnosed with cancer.[48,49] The interplay between pathogenic germline variants and the somatic genome is an emerging focus of cancer predisposition research with a growing list of rare genetic syndromes known to be associated with an increased cancer risk.[50] Thus, we also interrogated systemic pan-histonopathy phenotypic patterns, including neurological features, non-neurologic systemic features, growth features, and cancer prevalence. This work builds on preceding analyses that have sought to identify patterns within specific OMIM-classifications,[14–19,23,26,28,30–35] as well as inter-histonopathy patterns,[37] which have shown that there is phenotypic heterogeneity across individuals. By taking a more expansive approach than has been previously possible, we aim to interrogate how these heterogeneous clinical phenotypes highlight ways in which proximate biology converges on common genes, pathways, or functional signatures that may provide insight into the disease pathogenesis, as has been shown for other neurologic syndromes.[51] Thus, with the inclusion of the 192 known individuals with germline histonopathies in this analysis, we work to leverage all existing data to gain translational insights so that more proactive and directed comorbidity management and/or screening recommendations can be developed for this community; to facilitate the robust capture of phenotypic features in lab-based models of these syndromes; and to enable the patient-guided development of therapeutics to slow or prevent progression of disease.

Results

Cohort overview

A total of 185 individuals have been reported in the literature to harbor germline variants in histone-encoding genes.[13–36] HIST1H1E syndrome is caused by heterozygous germline frameshift variants in *HIST1H1E/H1-4*, which encodes the RC H1 linker protein.[13–26] BLBS is caused by heterozygous germline variants in *H3F3A/H3-3A* or *H3F3B/H3-3B*, the genes that encode the RI nucleosomal H3.3 protein.[27–32] TBvH NDD is caused by heterozygous germline variants in *H4C3, H4C4, H4C5, H4C6, H4C9,* or *H4C11*, which encode the RC nucleosomal H4 protein (Figure 1B).[33–36]

In addition to these previously published individuals, we include seven unreported individuals in this analysis (Figure 1B, bold; Table S1). One individual, HIST-A, harbors homozygous stop-gain missense variants in *HIST1H2AL/H2AC16*. Individual HIST-B harbors a *de novo* heterozygous missense variant in *H2AFZ/H2AZ1*. Individual HIST-C harbors a *de novo* heterozygous missense variant in *HIST1H3D/H3C4*. Two individuals harbor variants in *HIST3H3/H3-4* (HIST-D – compound heterozygous variants; HIST-E – paternally inherited heterozygous variant). Finally, two individuals harbor previously unreported BLBS-causative variants in *H3F3B/H3-3B*.

The syndromic phenotypes of affected individuals with germline histonopathies support the pathogenicity of these variants. However, relying on commonly utilized strategies beyond genotype- and phenotype-driven approaches to uplift gene-disease relationships has proven challenging for histone-based disorders.[40] Constraint metrics are not currently available for all histonopathy-causing genes in either the Genome Aggregation Database (gnomAD) v2 release or the gnomAD v4 release, so we performed a combined analysis using data from both versions (Figure 1C). gnomAD missense and loss-of-function constraint metrics support the causality of variants in a subset of previously established and previously unidentified histonopathy-associated genes (Figure 1C, orange). More specifically, constraint metrics support causality exclusively for histonopathy-implicated RI histone genes, such as those that underly BLBS, but not histonopathy-implicated RC histone genes, including those that underly HIST1H1E syndrome and TBvH NDD.

Model organism databases provide orthogonal evidence to support the nomination of a GUS as a disease-associated candidate.[40] The Monarch Initiative integrates information about the phenotype that results from perturbation of human gene orthologues in multiple model organisms, including *Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (worm), *Danio rerio* (zebrafish), *Drosophila melanogaster* (fly), and *Mus musculus* (mouse).[52,53] While Monarch Initiative data support the pathogenicity of variants in *H1-4, H3-3A, H2AZ1,* and *H3-4*, information about model organisms that interrogate all histonopathy-implicated genes is not available, so we once again lack complete data for both previously established and more recently reported histonopathy genes (Figure 1C).

Another recommended strategy to improve Mendelian disease discovery is to utilize multi-omic data, such as RNA-seq transcriptomic expression data.[40] While this can successfully be done for polyadenylated RI histones, nearly two-thirds of histones are RC and these genes

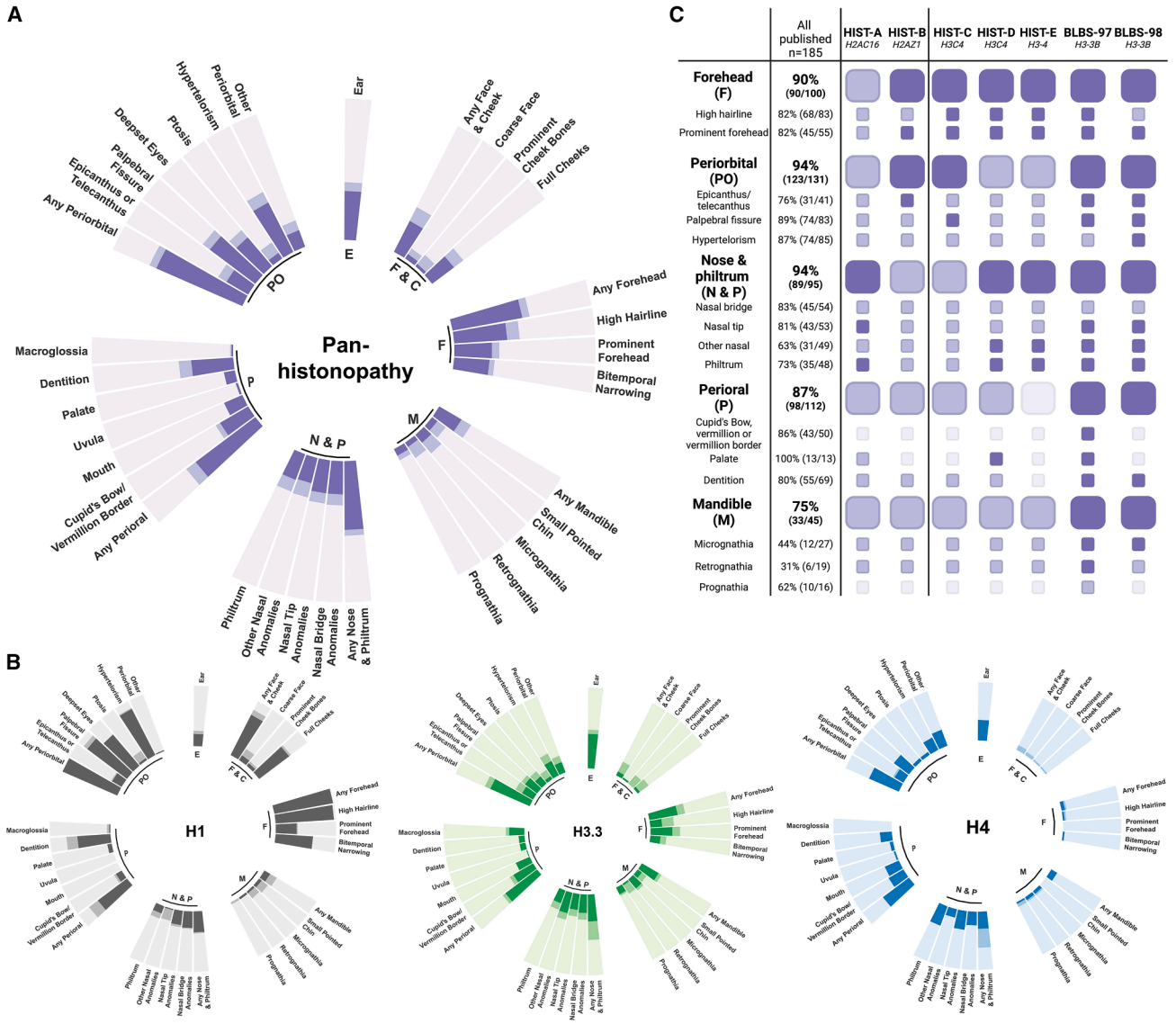


Figure 2. Quantification of clinician-reported craniofacial data

Color-coding for (A)–(C): darkest color indicates a clinician-reported response of yes, intermediate color indicates a clinician-reported response of no, and palest color indicates that a feature was either not reported or not assessed.

(A) Phenotypic analysis of all 191 affected individuals, subdivided by neurologic (N) features, non-neurologic systemic (S) features, growth (G) features and craniofacial (CF) features. Purple color indicates analysis unbiased by OMIM classification (pan-histonopathy).

(B) Analysis of published phenotypes for affected individuals diagnosed with HIST1H1E syndrome (H1, gray), BLBS (H3.3, green), or TBvH NDD (H4, blue) (color-coded based on Figure 1A). Interrogation subdivided by neurologic (N) features, non-neurologic systemic (S) features, growth (G) features, and craniofacial (CF) features.

(C) Reporting of phenotypes for previously unreported affected individuals. Top row includes the identifier for the affected individuals as well as their affected gene. First column lists prevalence of a feature across individuals previously reported.

encode transcripts that are not polyadenylated (Figure 1B).[3,54] Thus, powerful and widely utilized transcriptomic datasets that rely on polyA-positive RNA-selection for library preparation, such as the Genotype-Tissue Expression (GTEx) Project and the BrainSpan Atlas of the Developing Human Brain, do not capture the expression of this entire class of histones (Figure 1C).[55,56] There are unique histone-specific gaps in available tools that pose barriers to the nomination of histone-encoding genes as Mendelian disease associated. Thus, deeply mining the currently available data may help build a framework for more expediently identifying affected individuals based on a conserved constellation of clinical features.

Interrogation of identifiable facial gestalt

We interrogated whether germline histonopathies could be unified or substratified by patterns of facial features. After subdividing the regions of the face based on the *American Journal of Medical Genetics* Special Issue "Elements of Morphology: Standard Terminology," we visualized the clinician-reported craniofacial data for all 192 individuals (Figure 2A).[57–62] We then substratified the data. For the

185 previously reported individuals, we stratified based on OMIM classification (Figure 2B). For the seven unpublished individuals, we stratified based on protein class (H2A v H3) (Figure 2C).

Broadly, we see that 89.7% of individuals present with forehead features (HP:0009890, HP:0011220, HP:0000341, response rate = 55.7%), 92.0% of individuals present with periorbital features (HP:0000316, HP:0000508, HP:0000490, HP:0008050, HP:0000286, HP:0000506, response rate = 71.4%), 92.5% of individuals present with nose and philtrum features (HP:0011825, HP:0011829, HP:0011832, HP:0011831, HP:0000455, HP:0000422, response rate = 53.1%), 84.7% present with perioral features (HP:0000158, HP:0000164, HP:0000174, HP:0000172, HP:0011538, HP:0000153, HP:0011337, HP:0002263, HP:0011339, response rate = 61.4%), 67.0% present with mandibular features (HP:0000307, HP:0000347, HP:0000278, HP:0000303, response rate = 27.0%), 84.8% present with ear anomalies (HP:0000598, response rate = 41.1%), and 68.1% present with face and cheek features (HP:0000280, HP:3000019, HP:0000293, response rate = 35.9%). The most salient feature of these data was the low/inconsistent response rate (Figures 2A and 2B; Table S1). Homing in on the individual features that comprise these more comprehensive groupings, there was not a single queried feature for which a >51% response rate was attained in the pan-histonopathy craniofacial analysis. The HIST1H1E syndrome community had the highest clinician response rate related to facial phenotype queries, achieving a >51% response rate for 7 of 27 interrogated facial features (Figure 2B, left; Table S1). Data for this community suggest a pattern of facial features that include full cheeks (HP:0000293), high hairline (HP:0009890), bitemporal narrowing (HP:0000341), dental anomalies (HP:0000164), palpebral fissure differences (HP:0008050), deepset eyes (HP:0000490), and hypertelorism (HP:0000316). A >51% response rate was not achieved for any queried feature for either the BLBS or TBvH NDD community. This inconsistent availability of information about the qualitative features across published individuals hinders the discernment of unifying or delineating craniofacial gestalts at this time.

Clinician response rates were improved for the previously unpublished individuals, for whom a revised clinician survey was used to elicit data (Figure 2C; Table S2). There is not a clear pattern of salient features associated with H2A-based germline histonopathies based on the data for individuals HIST-A and HIST-B. Conversely, a few notable characteristics may be linked with H3 protein-associated germline histonopathies. High hairline (HP:0009890), prominent forehead (HP:0011220), and philtrum anomalies (HP:0011825) seem to be features conserved across H3-based germline histonopathies, regardless of which gene is perturbed (*H3C4, H3-4, H3-3B)*. There may also be features that are specific to perturbed gene/protein. For instance, of the previously unpublished individuals with variants in H3 class histones, just the two individuals with BLBS exhibit epicanthus/telecanthus (HP:0000286, HP:0000506), palpebral fissure anomalies (HP:0008050), nasal tip anomalies (HP:0011832, HP:0011831, HP:0000455), dental anomalies (HP:0000164), and micrognathia (HP:0000347) (Figure 2C). Future gene-level analyses will provide informative, granular information. Unfortunately, at this time, response rates are not high enough to allow for this level of delineation (Figure S1).

Recognizing the confounding effect of data gaps in available clinician-reported data, we employed a complementary approach to assess the facial gestalt of individuals with germline histonopathies based on 2D facial photographs through the use of the GMDB NGP platform. Ninety-eight images of individuals with germline histonopathies have been published to date.[14–18,22–25,27,28,30,33–36] These images include 51 of individuals with HIST1H1E syndrome, 25 of individuals with BLBS, and 22 of individuals with TBvH NDD. Male and female individuals are equally represented (48M/50F) (Table S3). Unfortunately, no images of previously unreported individuals were available for this analysis.

Based on the developmental stages of the craniofacial complex, we stratified by age into a 5 years and younger cohort (Figures 3A and 4A); a 6- to 12-year-old cohort (Figures 3B and 4B); and a 13 years and older cohort (Figures 3C and 4C). This analysis enables us to understand the similarities among individuals with germline histonopathies compared with the other ∼10,000 images that comprise GMDB.[63–66] In the inter-cohort analysis, the distribution of individuals with variants in HIST1H1E, H3.3, and H4 form distinct clusters, though with slight overlaps between them (Figures 3A–3C and S2). Within all three age groups, the pairwise comparisons between distinct histonopathy communities show that 100% of the distributions exceed the threshold, indicating distinct facial phenotypes among the patient groups. However, the left tails of some comparisons approach the threshold, particularly for HIST1H1E/H3.3 in the under-5 age group and H3.3/H4 in both the under-5 and 6 to 12 age groups (Figures 3A and 3B). To further investigate these overlaps, we conducted an individual-level analysis. The 5 years and younger cohort is composed of 40 total images of individuals, 23 of whom have a diagnosis of HIST1H1E syndrome; 13 of whom have a diagnosis of BLBS; and four of whom have a diagnosis of TBvH NDD (Table S3). Male and female individuals are equally represented (19M/21F). The largest cluster, cluster 1, is composed almost entirely of images of individuals with HIST1H1E syndrome (95%, 21 of 22) (Figure 4A, black/gray).

The 6- to 12-year-old cohort is composed of 24 individuals, 13 of whom have a diagnosis of HIST1H1E syndrome, six of whom have a diagnosis of BLBS, and five of whom have a diagnosis of TBvH NDD (Table S3). Male and female individuals are equally represented (13M/11F). Again, the largest cluster, cluster 1, is composed mostly of images of individuals with HIST1H1E syndrome (92%, 13 of 14) (Figure 4B, black/gray). Cluster

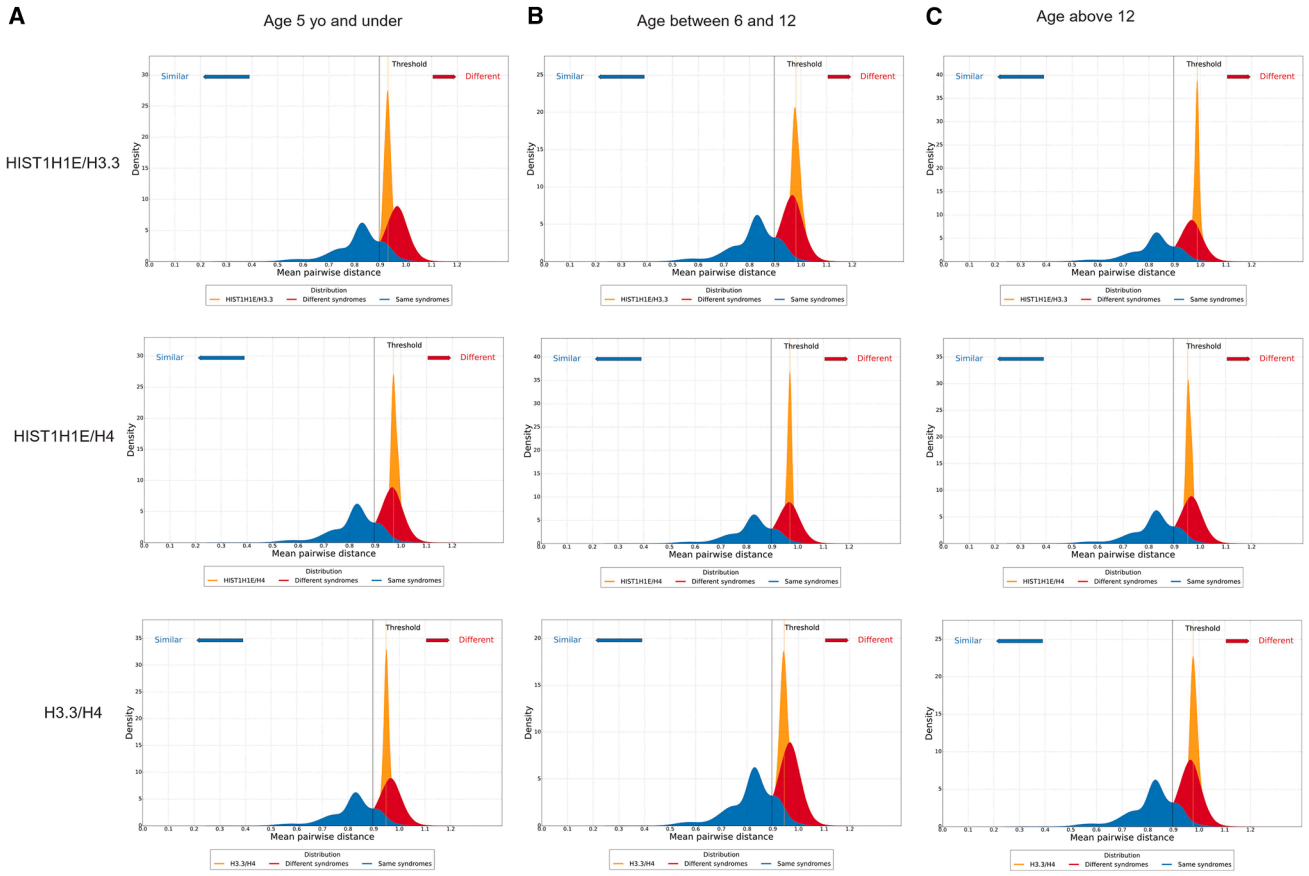


Figure 3. GMDB inter-cohort analysis

Inter-cohort analysis using mean pairwise distance distribution comparing images from individuals with the same syndrome (blue), different syndrome (red), and respective histonopathies (orange). Each row shows a pair of comparison, such as HIT1H1E/H3.3, HIST1H1E/H4, and H3.3/H4. Each column shows the results at different age groups such as age younger than 5 and 5 years old (A), age between 6 and 12 years old (B), and age older than 12 years old (C). The *x* axis is the distance. Therefore, small distance indicates high similarity. The black vertical line is the threshold that classifies whether it is the same disorder or different disorders.

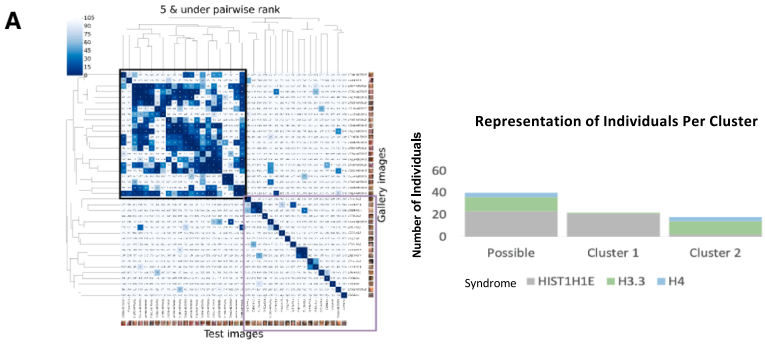
2 is composed of a mixed cohort of individuals with BLBS (60%, 6 of 10) and with TBvH NDD (40%, 4 of 10) (Figure 4B, purple). Thus, these data do not support the definitive subclustering of individuals with BLBS or TBvH NDD in this age group.

Finally, the 13 years and older cohort is composed of 28 individuals, 13 of whom have diagnosis of HIST1H1E syndrome, six of whom have a diagnosis of BLBS, and nine of whom have a diagnosis of TBvH NDD. Male and female individuals are equally represented (15M/13F). Cluster 1 is again composed mostly of images of individuals with HIST1H1E syndrome (86%, 13 of 15) (Figure 4C, black/gray). This cohort, with the largest age range, demonstrates the greatest intracluster variability. If more images of individuals with histonopathies become available in this age group, further substratification could elucidate whether the age range itself confounds this analysis or whether the facial gestalt becomes less consistent over time.

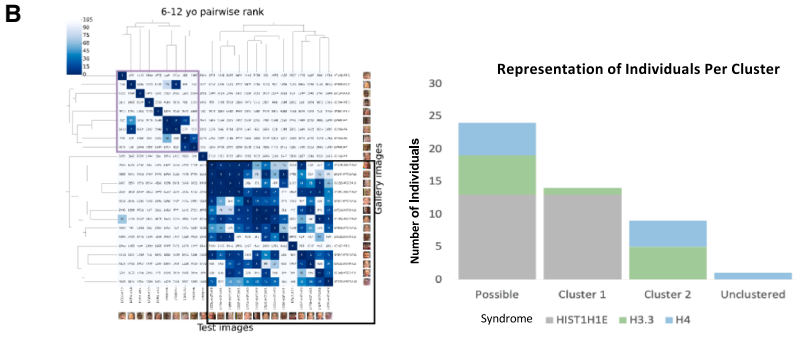
Taken together, this analysis shows that images of individuals with HIST1H1E syndrome consistently form distinct clusters across all age groups (Figures 4A–4C). Nonetheless, one or two images of individuals with BLBS or TBvH NDD occasionally cluster with these HIST1H1E syndrome predominant clusters. For instance, in the under-5 age group, image 17318 (BLBS) clusters with the HIST1H1E group. Further, although the inter-cohort analysis shows that BLBS and TBvH NDD present with different phenotypes (Figures 3A–3C), we did see some potential for shared presentations between these two groups. In the 6- to 12-year age group, image 17330 (BLBS) ranks fourth in similarity to image 17390 (TBvH NDD) (Figure 4B). These findings suggest that while these three groups may exhibit distinct facial phenotypes at a population level, some individuals present with craniofacial features shared by multiple OMIM-classified subgroups.

Interrogating clinician-reported phenotypes to identify systemic constellations of features

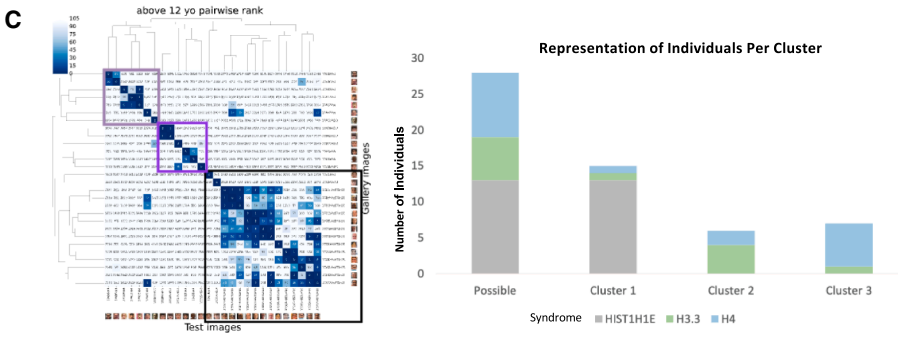
We next performed an unbiased pan-histonopathy quantification of all clinician-reported data of non-craniofacial phenotypic features for this cohort of 192 individuals. We delineated different sub-phenotypes of affected individuals into (1) neurologic features, (2) non-neurologic



5 years old & under	All	Cluster 1	Cluster 2
Images with at least 1 individual in top 10-rank	27/33	21/22	4/8
Images with at least 1 individual in top 20-rank	29/33	22/22	5/8
Images with at least 1 individual in top 30-rank	30/33	22/22	6/8
Images with at least 1 individual in top 40-rank	33/33	22/22	8/8



6– 12 years old	All	Cluster 1	Cluster 2	Un-clustered
Images with at least 1 individual in top 10-rank	18/21	13/14	5/6	0/1
Images with at least 1 individual in top 20-rank	18/21	13/14	5/6	0/1
Images with at least 1 individual in top 30-rank	18/21	13/14	5/6	0/1
Images with at least 1 individual in top 40-rank	18/21	13/14	5/6	0/1



13 years old & over	All	Cluster 1	Cluster 2	Cluster 3
Images with at least 1 individual in top 10-rank	21/28	13/15	2/6	5/7
Images with at least 1 individual in top 20-rank	25/28	13/15	6/6	5/7
Images with at least 1 individual in top 30-rank	25/28	13/15	6/6	5/7
Images with at least 1 individual in top 40-rank	25/28	13/15	6/6	5/7

Figure 4. GMDB craniofacial analysis of previously published images

GestaltMatcher analysis of published photos of individuals with germline histonopathies stratified by age. Pairwise rank dendrograms for 5 years and younger (A); 6–12 years old (B); 13 years and older (C). Pairwise rank dendrograms are annotated with cluster 1 (black/gray) and cluster 2 (purple), cluster 3 (bright purple). The pairwise rank matrix represents the similarity rankings of images, where each

(legend continued on next page)

systemic features, and (3) growth features (Figure 5A). Broadly, we see that 100% of individuals present with a neurologic phenotype (HP:0012759, HP:0001249, HP:0001270, HP:0000750, HP:0030890, HP:0007103, HP:0001252, HP:0001276, HP:0011145, HP:0000708, HP:0002360, HP:0000365, HP:0000657, HP:0000496, HP:0000504, response rate = 99.5%), 85% of individuals present with a non-neurologic systemic phenotype (HP:0001945, HP:0002719, HP:0001005, HP:0005616, HP:0000119, HP:0011968, HP:0011024, HP:0012718, HP:0500015, HP0011028, response rate = 97.9%), and 94.2% of individuals present with a growth phenotype (HP:0001507, HP:0000234, HP:0000252, HP:0000256, HP:0000924, HP:0000002, HP:0004323, HP:0000818, response rate = 98.9%) (Figure S3). However, when we subdivide each phenotypic category into more specific features, we observe varied response rates. A response rate >51% was achieved for 9 of 11 (81.2%) of queried neurologic features and for five of six (83.3%) of queried growth features but just two of eight (25.0%) of queried non-neurologic systemic features (Table 1).

Based on this threshold of response rate >51%, these data suggest that germline histonopathies are associated with neurologic features that include DD/ID (HP:0012759, HP:0001249), motor delay (HP:0001270), speech delay (HP:0000750), anomalies on brain MRI (HP:0030890, HP:0007103), hypotonia (HP:0001252), and vision anomalies (HP:0000504), with head sizes outside the expected range for an individual's age and sex (HP:0000252, HP:0000256), and with skeletal changes (HP:0000924). They also suggest that, at a population level, germline histonopathies are not associated with seizures (HP:0011145), comorbid NDDs (including autism spectrum disorder and attention-deficit hyperactivity disorder) (HP:0000708, HP:0000717, HP:0007018), hearing anomalies (HP:0000365), cardiac/circulatory anomalies (HP:0500015, HP:0011028), genitourinary anomalies (HP:0000119), height outside the expected range for an individual's age and sex (HP:0000002), weight outside the expected range for an individual's age and sex (HP:0004323), or endocrine anomalies (HP:0000818). Response rates at this time are insufficient to interrogate the association between germline histonopathies and any other features.

We then analyzed the phenotypic features of the 185 previously published individuals subdivided by their OMIM-classified histonopathies: HIST1H1E syndrome (Figure 5B, top), BLBS (Figure 5B, middle), and TBvH NDD (Figure 5B, bottom). For HIST1H1E syndrome, a response rate >51% was achieved for 4 of 11 (36.4%) of queried neurologic features, three of eight (37.5%) of queried non-neurologic systemic features, and four of six (66.7%) of queried growth features (Table 2). The data suggest that HIST1H1E syndrome is associated with neurologic features that include DD/ID (HP:0012759, HP:0001249), brain MRI anomalies (HP:0030890, HP:0007103), hypotonia (HP:0001276), and comorbid NDDs (HP:0000708, HP:0000717, HP:0007018), with features of advanced aging (HP:0005616), and with skeletal changes (HP:0000924). An association between HIST1H1E syndrome and cardiac/circulatory anomalies (HP:0500015, HP:0011028), genitourinary anomalies (HP:0000119), or head size/weight/height outside the expected range for an individual's age and sex (HP:0000252, HP:0000256, HP:0000002, HP:0004323) is not supported by these data. Response rates at this time are insufficient to interrogate the association between HIST1H1E syndrome and other features.

For BLBS, a response rate >51% was achieved for 8 of 11 (72.3%) of queried neurologic features, two of eight (25.0%) of queried non-neurologic systemic features, and five of six (83.3%) of queried growth features (Table 2). The data suggest that BLBS is associated with neurologic features that include DD/ID (HP:0012759, HP:0001249), motor delay (HP:0001270), speech delay (HP:0000750), brain MRI anomalies (HP:0030890, HP:0007103), hypotonia (HP:0001252), and vision anomalies (HP:0000657 HP:0000504, HP; 0000496), and with growth features that include head size outside the expected range for age and sex (HP:0000252, HP:0000256), as well as skeletal changes (HP:0000924). These data do not support an association between BLBS and seizures (HP:0011145), comorbid NDDs (HP:0000708, HP:0000717, HP:0007018), hearing anomalies (HP:0000365), cardiac/circulatory anomalies (HP:0500015, HP:0011028), genitourinary anomalies (HP:0000119), height/weight outside the expected range for age and sex (HP:0000002, HP:0004323), or endocrine anomalies (HP:0000818). Again, for BLBS, response rates at this time are insufficient to interrogate the association between BLBS and other clinically relevant features.

For TBvH NDD, a response rate >51% was achieved for 6 of 11 (54.5%) of queried neurologic features, two of eight (25.0%) of queried non-neurologic systemic features, and four of six (66.7%) queried growth features (Table 2). The data suggest that TBvH NDD is associated with neurologic features that include DD/ID (HP:0012759, HP:0001249), speech delay (HP:0000750), and vision anomalies (HP:0000657 HP:0000504, HP; 0000496); feeding difficulties (HP:0011968); and with growth features that include head size outside the expected range for age and sex (HP:0000252, HP:0000256) as well as skeletal changes (HP:0000924). Despite the relatively robust response rate, these data do not support an association between TBvH

column corresponds to a specific test image and each row shows the rank of similarity for the remaining images. Labels indicate the image ID from GMDB and the group classification (e.g., HIST1H1E, H3.3, or H4). The hierarchical clustering highlights phenotypic patterns and relationships among the patient groups. Bar charts report distribution of OMIM-characterized syndromes in each age group (total possible), in each cluster, and in the unclustered images.

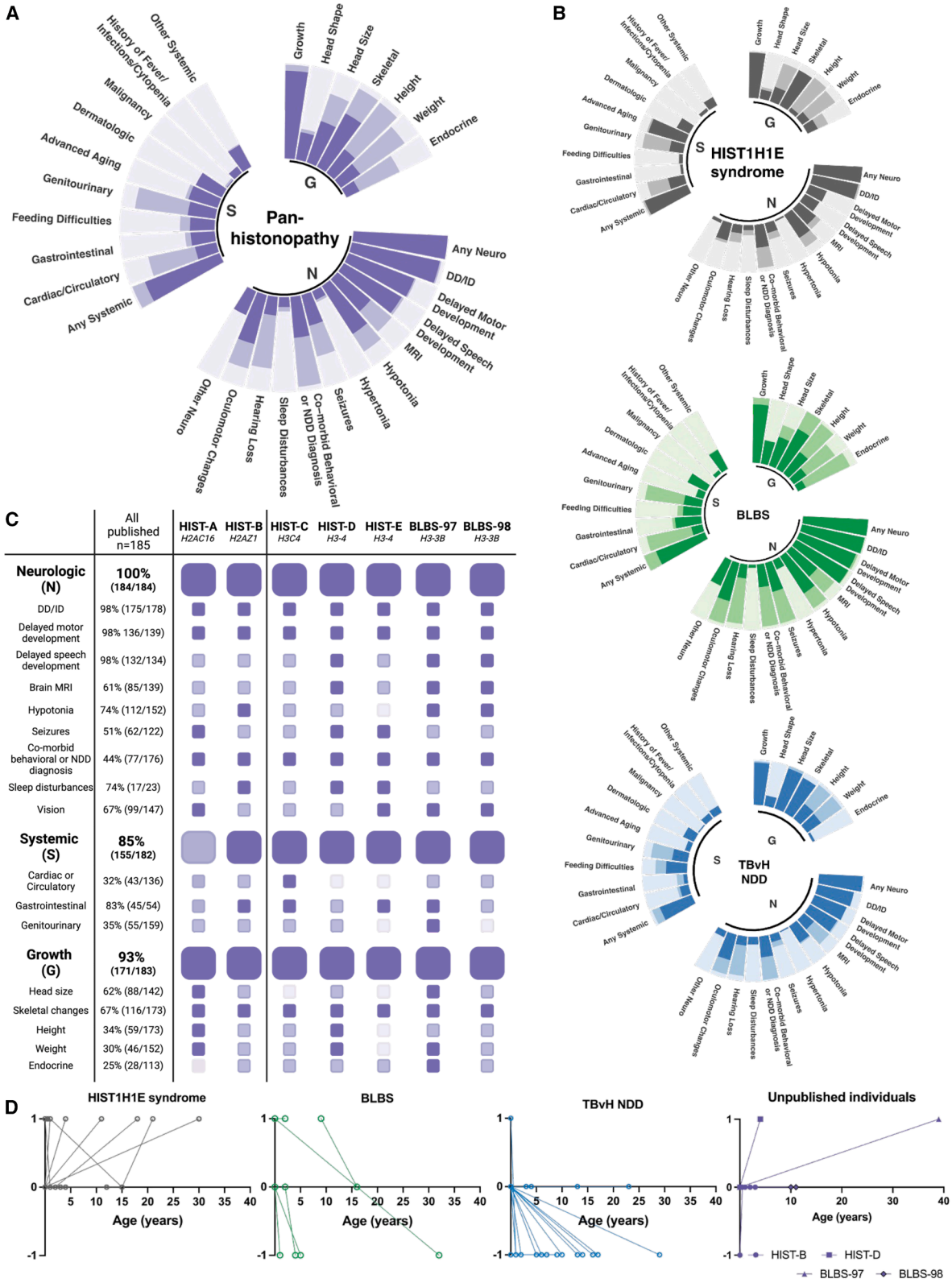


Figure 5. Quantification of non-craniofacial clinician-reported phenotypes

Color-coding for (A)–(C): darkest color indicates a clinician-reported response of yes, intermediate color indicates a clinician-reported response of no, and palest color indicates that a feature was either not reported or not assessed.

(A) Phenotypic analysis of all 191 affected individuals, subdivided by neurologic (N) features, non-neurologic systemic (S) features, growth (G) features, and craniofacial (CF) features. Purple color indicates analysis unbiased by OMIM classification (pan-histonopathy).

(legend continued on next page)

NDD and brain MRI anomalies (HP:0030890, HP:0007103), comorbid NDDs (HP:0000708, HP:0000717, HP:0007018), hearing anomalies (HP:0000365), genitourinary anomalies (HP:0000119), or weight outside the expected range for age and sex (HP:0004323).

With this granular quantification of previously published data, we see that the three OMIM-classified histonopathies share the phenotypic features of DD/ID (HP:0012759, HP:0001249) and skeletal changes (HP:0000924), which is conserved across the previously unreported individuals as well (Table 2; Figure 5C). For the neurologic features, HIST1H1E syndrome and BLBS share brain MRI anomalies (HP:0030890, HP:0007103) and hypotonia (HP:0001252) as salient phenotypes, while BLBS and TBvH NDD share prevalent speech delay (HP:0000750) and vision anomalies (HP:0000657 HP:0000504, HP; 0000496). Interestingly, all seven previously unpublished individuals exhibit delayed motor development (HP:0001270) and comorbid NDDs (HP:0000708, HP:0000717, HP:0007018). Consistent with trends observed in published individuals with BLBS, BLBS-97 and BLBS-98 exhibit delayed speech development (HP:0000750), brain MRI anomalies (HP:0030890, HP:0007103), hypotonia (HP:0001252), and vision anomalies (HP:0000657 HP:0000504, HP; 0000496). Individuals HIST-D and HIST-E, who harbor variants in *H3-4*, have seizures (HP:0011145) and sleep disturbances (HP:0002360). The data do not suggest that seizures (HP:0011145) are a phenotype shared by all individuals with H3-associated germline histonopathies (prevalence <51% for BLBS, not reported for individual HIST-C) (Table 2; Figure 5C).

For growth features, BLBS and TBvH NDD share head size outside the expected range for age and sex (HP:0000252, HP:0000256). However, this head size metric reflects the measurement just at the most recent exam. Thus, we then leveraged available longitudinal mixed qualitative/quantitative head size data to interrogate a possible temporal phenotypic transition. Individuals with HIST1H1E syndrome exhibit a macrocephalic (HP:0000256) trend over time (Figure 5D, gray). Conversely, individuals with BLBS and TBvH NDD exhibit a microcephalic (HP:0000252) trend over time (Figure 5D, green and blue). A putative dichotomy on the basis of linker versus core proteins is not supported by the longitudinal head size data from unpublished individuals, all of whom harbor variants in core nucleosome proteins. Individuals HIST-B, HIST-D, and BLBS-97 exhibit a macrocephalic (HP:0000256) transition while individual BLBS-98 does not exhibit a change in head size over time (Figure 5D, purple).

Data gaps for the non-neurologic systemic features confound the assessment of phenotypic patterns (Figure 5; Tables 1 and 2). However, the data collected from previously unpublished individuals once again highlights how even a single individual can shape the clinical understanding of these syndromes. Here, we report the second individual with a germline histonopathy and malignancy (HP:0002664).[35] Individual BLBS-97, who harbors a p.Lys23Arg variant in *H3-3B* (NM_005324.5: c.71A>G; p.[Lys24Arg]), was diagnosed with an adrenocorticotropic hormone (ACTH)-producing pulmonary carcinoid tumor (HP:0030445) at the age of 33 years. Strikingly, they were not diagnosed with one of the tumor types historically associated with H3.3 somatic oncohistone mutations, which include diffuse midline gliomas, high-grade gliomas of the cerebral cortex (HP:0009733), giant cell tumors of the bone (HP:0011847), and chondroblastomas (HP:0030432).[10,67] Instead, this individual was diagnosed with a rare but resectable lung tumor (HP:0100526). Interestingly, the residue altered in BLBS-97, p.Lys23, is post-translationally modified by KAT6B, a tumor suppressor frequently lost in small cell lung cancer (HP:0030357).[68] Germline variants in *KAT6B* cause a syndromic NDD with a hypothesized cancer predisposition and, notably, an individual with a germline *KAT6B* variant with a neuroendocrine tumor (neuroblastoma, HP:0003006) was recently published.[69,70] For both of these ultra-rare disorders, a single affected individual has been reported with neuroendocrine tumors, albeit of different subtypes.

To better understand what role, if any, *H3-3B* mutations may play in cancer, we performed a more focused analysis of publicly available data. We first pulled all reported cancer-causing somatic mutations in *H3-3B* that are aggregated in cBioPortal for Cancer Genomics, which integrates data from The Cancer Genome Atlas (TCGA) and panCancer studies, pediatric cancer studies and targeted studies, and St. Jude's Pediatric Cancer (PeCan) platform, which integrates data from the Catalog of Somatic Mutations in Cancer (COSMIC) and the Clinical Interpretation of Variants in Cancer (CIViC).[71–73] We plotted known *H3-3B* germline variants (Figure 6A, top) against somatic mutations identified in these datasets (Figure 6B, bottom). Somatic *H3-3B* mutations are found in a range of tumor types, though notably no chondroblastoma (HP:0030432)-associated *H3-3B* mutations were pulled, which is peculiar given prior reports of *H3-3B* (Figure 6B; Table S4). These data show that lung cancers (HP:0100526), the tumor type found in individual BLBS-97, have the third highest *H3-3B* alteration

(B) Analysis of published phenotypes for affected individuals diagnosed with HIST1H1E syndrome (H1, gray), BLBS (H3.3, green), or TBvH NDD (H4, blue) (color-coded based on Figure 1A). Interrogation subdivided by neurologic (N) features, non-neurologic systemic (S) features, growth (G) features, and craniofacial (CF) features.
(C) Reporting of phenotypes for previously unreported affected individuals. Top row includes the identifier for the affected individuals as well as their affected gene. First column lists prevalence of a feature across individuals previously reported.
(D) Mixed qualitative/quantitative longitudinal assessment of head size, limited to affected individuals for whom >1 measurement was available. 1 = clinician-reported macrocephaly (HP:0000256), 0 = clinician-reported head size within expected range for age and sex, −1 = clinician-reported microcephaly (HP:0000252).

Table 1. Quantification of non-craniofacial pan-histonopathy systemic phenotypic features

Feature	Prevalence,[a] %	RR, %
Neurologic features		
DD/ID (HP:0012759, HP:0001249)	98.4	96.3
Motor delay (HP:0001270)	97.9*	76.0*
Speech delay (HP:0000750)	95.7*	73.4*
Brain MRI anomalies (HP:0030890, HP:0007103)	60.8*	74.5*
Hypotonia (HP:0001252)	72.7*	82.3*
Hypertonia (HP:0001276)	76.0	13.0
Seizure (HP:0011145)	50.3	67.2
Comorbid NDD (HP:0000708, HP: 0000717, HP:0007018)	45.9	95.3
Sleep disturbances (HP:0002360)	66.7	15.6
Hearing anomalies (HP:0000365)	26.0	76.0
Vision anomalies (HP:0000657, HP:0000504, HP:0000496)	66.9*	80.2*
Non-neurologic systemic features		
Cardiac/Circulatory anomalies (HP:0500015, HP:0011028)	31.2	73.4
Gastrointestinal anomalies (HP:0011024, HP:0512718)	80.32	31.0
Feeding difficulties (HP:0011968)	70.4	36.9
Genitourinary anomalies (HP:0000119)	34.0	85.4
Features of advanced aging (HP:0005616)	89.0	35.9
Dermatologic anomalies (HP:0001005)	98.3	30.7
Malignancy (HP:0002664)	33.3	3.1
Infection/Cytopenia/Fever (HP:0001945, HP:0002719)	100	8.3
Growth features		
Head shape changes (HP:0000234)	87.5	33.3
Head size[b] (HP:0000252, HP: 0000256)	60.5*	76.5*
Skeletal changes (HP:0000924)	68.3*	93.7*
Height[b] (HP:0000002)	34.1	93.2
Weight[b] (HP:0004323)	31.0	82.3
Endocrine anomalies (HP:0000818)	24.4	61.9

*Prevalence >51% and response rate >51%.
DD/ID = developmental delay/intellectual disability; MRI = magnetic resonance imaging; NDD = neurodevelopmental disorder; RR = response rate.
[a]Based on clinician responses.
[b]Outside range for age and sex at most recent exam.

frequency (Figure 6B, yellow). In fact, somatically mutated residues throughout *H3-3B* are observed in lung cancers (HP:0100526) (Figure 6A, bottom, yellow). It is notable that neuroblastoma (HP:0003006) is the malignancy with the highest *H3-3B* alteration frequency, given that neuroblastomas (HP:0003006) and ACTH-producing pulmonary carcinoids (HP:0030445) are both neuroendocrine tumors (HP:0100634).

Individual BLBS-97's NDD-causative germline variant is just six amino acids away from p.R17Pfs*80, one of three putative driver mutations, NM_005324.5:c.52dup; p.(Arg18ProfsTer80), identified in a blood sample from a 59-year-old man diagnosed with an atypical lung carcinoid (HP:0030445) (identifier P-0019467) (Figure 6A, top; Figure S4). Notably, both of BLBS-97's and P-0019467's variants are located at or near amino acids that are post-transcriptionally modified, which meets criteria for "*a posteriori*" nomination of a candidate histone driver mutation.[67] The delineation between driver and passenger mutations in *H3-3B* is relevant because cancers develop a high mutational burden as they continue to overcome cellular checks on growth and division. While it is not possible to conclusively determine the functional impact of the reported somatic *H3-3B* mutations at this time, additional support comes from the low mutational background and the high variant allele fraction

Table 2. Quantification of non-craniofacial systemic phenotypic features of published individuals by OMIM classification

	HIST1H1E syndrome		BLBS		TBvH NDD	
Feature	Prevalence,[a] %	RR, %	Prevalence,[a] %	RR, %	Prevalence,[a] %	RR, %
Neurologic features						
DD/ID (HP:0012759, HP:0001249)	96.4*	100*	98.9*	98.9*	100*	82.4*
Motor delay (HP: 0001270)	100	38.1	96.7*	95.8*	100	48.1
Speech delay (HP:0000750)	100	34.5	98.9*	93.8*	96.0*	73.5*
Brain MRI anomalies (HP:0030890, HP:0007103)	78.1*	58.2*	60.9*	85.4*	40.0	73.5
Hypotonia (HP:0001252)	71.1*	81.8*	71.4*	94.7*	93.8	47.1
Hypertonia (HP:0001276)	100	3.6	100	15.6	100	2.9
Seizure (HP:0011145)	60.0	36.4	47.8	95.8	60.0	29.4
Comorbid NDD (HP:0000708, HP:0000717, HP:0007018)	52.9*	92.7*	36.9	95.8	48.5	97.1
Sleep disturbances (HP:0002360)	50.0	21.8	100	5.2	100	17.6
Hearing anomalies (HP; 0000365)	47.6	38.1	23.9	91.7	23.3	88.2
Vision anomalies (HP:0000657, HP:0000504, HP:0000496)	87.5	43.6	63.0*	95.8*	64.5*	91.2*
Non-neurologic systemic features						
Cardiac/Circulatory anomalies (HP:0500015, HP:0011028)	44.2	78.2	20.2	87.5	77.8	26.5
Gastrointestinal anomalies (HP:0011024, HP:0012718)	100	5.5	90.7	44.8	37.5	23.5
Feeding difficulties (HP:0011968)	80.0	9.1	77.1	36.5	66.7*	70.6*
Genitourinary anomalies (HP:0000119)	34.0	85.5	32.1	87.5	42.8	82.4
Features of advanced aging (HP:0005616)	95.6*	83.6*	100	12.5	100	11.8
Dermatologic anomalies (HP0001005)	100	29.1	100	32.3	100	26.5
Malignancy (HP:0002664)	0	0	0	0	100	2.9
Infection/Cytopenia/Fever (HP:0001945, HP:0002719)	100	7.3	100	6.3	100	14.7
Growth features						
Head shape changes (HP:0000234)	78.6	25.4	100	36.5	100	23.5
Head size[b] (HP:0000252, HP:0000256)	38.3	85.5	71.4*	65.6*	78.1*	94.1*
Skeletal changes (HP:0000924)	61.7*	85.5*	69.8*	100*	66.7*	88.2*
Height[b] (HP:0000002)	10.2	89.1	43.5	95.8	43.8	94.1
Weight[b] (HP:0004323)	21.7	83.6	31.1	77.1	40.6	94.1
Endocrine anomalies (HP:0000818)	50.0	47.3	15.3	88.5	100	5.9

*Prevalence >51% and Response Rate >51%.
BLBS = Bryant-Li-Bhoj syndrome; DD/ID = developmental delay/intellectual disability; MRI = magnetic resonance imaging; NDD = neurodevelopmental disorder; RR = response rate; TBvH = Tessadori-Bicknell-van Haaften NDD.
[a]Based on clinician responses.
[b]Outside range for age and sex at most recent exam.

(VAF = 0.24) found in the sample sequenced from individual P-0019467. The somatic *H3-3B* p.R17Pfs*80 putative driver was found to be co-occurring with two other putative drivers in *JAK2* and *DNMT3A*.[74] Due to high homology,[75] variant calls for *H3-3B* can be challenging, which makes it difficult to accurately determine the true VAF and to distinguish between germline and somatic mutations, particularly in blood samples with clonal hematopoiesis. Additional clinical context and validation (e.g., past medical and family history, targeted sequencing of other tissues) would be necessary to fully determine whether the variant is a somatic or germline event, as well as to assess any potential association between the *H3-3B* variant detected in P-0019467 and the

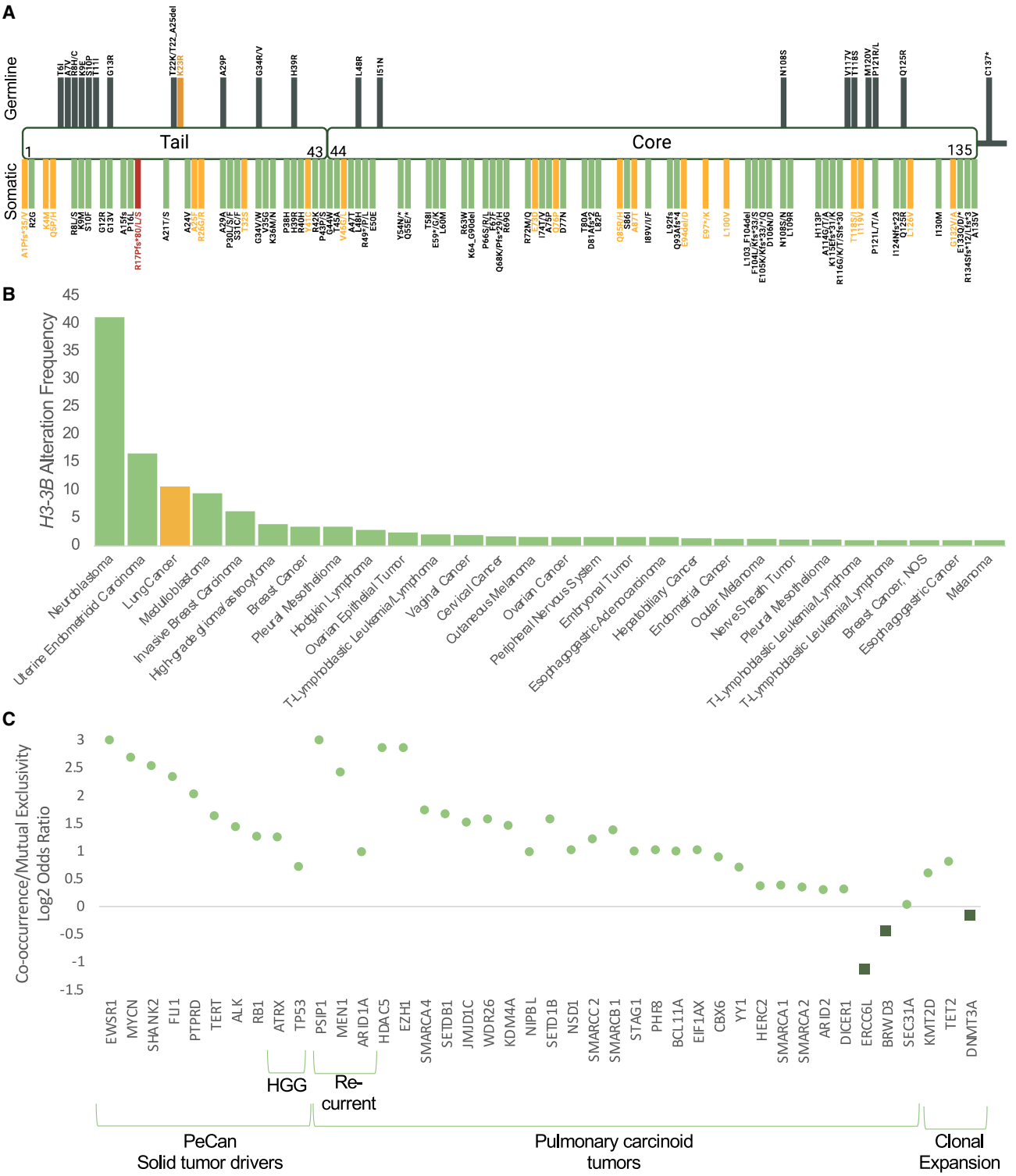


Figure 6. ***H3-3B*** **in cancer**

(A) A map of all the known H3-3B germline variants (top) and H3-3B somatic mutations aggregated in cBioPortal or PeCan (bottom). Yellow germline variant highlights the individual with both a BLBS diagnosis and a history of an ACTH-producing carcinoid tumor (HP:0030445) of the lung. Yellow somatic mutations highlight residues affected in lung tumors (HP:0100526). The red somatic mutation indicates the putative driver mutation in an atypical lung tumor (HP:0100526) reported in cBioPortal.

(B) Tumor types ranked by the percentage of cases with somatically mutated H3-3B (for alteration frequency ≥1) based on data aggregated in cBioPortal. Yellow = lung cancers.

(legend continued on next page)

development of an atypical lung carcinoid (HP:0030445). Nonetheless, the co-occurrence of a *H3-3B* mutation with three other putative mutations supports the further investigation of the role of somatic *H3-3B* mutations.

In furtherance of that objective, we performed a co-occurrence/mutual exclusivity analysis. This was motivated by prior reports that showed co-mutation analyses can both facilitate the identification of pathways implicated in disease pathogenesis, which could be targets for therapeutic intervention, and shed light on combinations of mutations that may be protective against tumorigenesis.[76] In order to interrogate genes with which *H3-3B* mutations are frequently co-occurring or mutually exclusive, we defined a list of genes to query using the cBioPortal for Cancer Genomics beta search function (Figure 6C). Somatic oncohistone mutations in H3.3 in high-grade gliomas often co-occur with mutations in *TP53* or *ATRX*.[9] In addition to mutations in these two genes, non-brain solid tumors, such as lung cancers (HP:0030445), are reported to frequently co-occur with driver mutations in *EWSR1, MYCN, SHANK2, FLI1, PTPRD, TERT, ALK,* and *RB1*.[73] Interestingly, genes encoding histone-modifying proteins, including *PSIP1, MEN1,* and *ARID1A*, are frequently mutated in rare neuroendocrine pulmonary carcinoids (HP:0030445), which are etiologically similar to the one resected from individual BLBS-97.[77] As was the case with germline variants in *EZH2,* germline variants in these chromatin-remodeling proteins are also known to cause MDEMs.[12] Notably, even though somatic *TP53* mutations are common co-occurring drivers in high-grade gliomas (HP:0009733), mutations in *TP53* and *RB1* have been found to rarely co-occur with mutations in histone-modifying genes in pulmonary carcinoids.[77] Finally, *KMT2D, TET2,* and *DNMT3A* are known drivers of clonal expansion in solid tumors.[74] Using this list of genes, we surveyed the co-occurrence/mutual exclusivity log2 odds ratio of *H3-3B* mutations with these reported tumor drivers in cBioPortal (Figure 6C).

Somatic mutations in 69% of queried genes were found to significantly co-occur in solid tumors and/or pulmonary carcinoids (HP:0030445) with somatic mutations in *H3-3B* ($p < 0.05$), including 100% of PeCan-identified solid tumor driver genes, 62% of chromatin-remodeling genes mutated in pulmonary carcinoids (HP:0030445), and 33% of frequently mutated drivers of clonal expansion in solid tumors (Table S4). Somatic mutations in three genes were mutually exclusive with somatic mutations in *H3-3B: ERCC6L, BRWD6,* and *DNMT3A* (Figure 5C, squares). Both *BRWD4* and *DNMT3A* encode epigenetic modifiers. Germline variants in these genes are associated with the MDEMs X-linked intellectual disability disorder 93 (OMIM: 300659) and Tatton-Brown-Rahman syndrome (OMIM: 615879), respectively. Both MDEMs exhibit phenotypic overlap with germline histonopathies. However, other MDEM-associated genes were queried, and support ongoing functional work to delineate the difference between variants that cause NDDs (HP:0012759) and those that cause cancer, as well as potentially protective combinations of mutations.

Discussion

By leveraging existing phenotypic information available for individuals harboring germline histone variants, we were able to begin addressing a few key outstanding questions for this community that have been repeatedly raised by family members of affected individuals as well as clinicians. We first addressed the question of whether there is a unifying phenotype that, once defined, could raise the clinical suspicion for a germline histonopathy diagnosis. As discussed above, there are unique histone-specific gaps in the tools typically used by the translational genetics community to establish the pathogenic role of a gene variant in Mendelian disease (Figure 1C). These barriers likely hinder the ability of individuals to obtain a definitive molecular diagnosis. Thus, we performed a deep quantification of all existing phenotypic information with the objective of delineating a conserved constellation of features that may accelerate clinical identification. As part of that analysis, we had the opportunity to investigate whether a single unifying clinical phenotype could apply to all individuals with germline histonopathies or whether more granular substratification on the basis of affected gene, affected protein, and/or affected histone class (RC vs. RI) was necessary (Figures 1A and 1B). This question of whether to lump or split is being increasingly examined within translational genetics, including by a ClinGen Lumping and Splitting Subgroup of the Curated Disease Entity Working Group, in an ongoing effort to provide the optimal clinical management and eventual therapeutic development for individuals living with rare monogenic syndromes.[78] While significant efforts have been made by our group and others within the international consortium of labs interrogating germline histonopathies to elucidate both the molecular mechanisms and the full spectrum of phenotypic variability of these syndromes, the fact remains that we do not yet have all the necessary pieces. Nonetheless, for the families of affected individuals, there is a responsibility to learn as much as possible from the data that do exist, and to learn from every diagnosed individual.

(C) Co-occurrence or mutual exclusivity log2 odds ratio of somatic H3-3B mutations with somatic mutations in genes frequently reported to harbor driver mutations in solid tumors (sourced from PeCan), including genes with co-occurring mutations in high-grade gliomas (HP:0009733); in chromatin-remodeling genes found to be recurrently mutated in pulmonary carcinoid tumors (HP:0030445) etiologically similar to the one resected from individual BLBS-97; or genes implicated in the clonal expansion of solid tumors. Circles = co-occurring; boxes = mutually exclusive.

Here, we expand the number of candidate histonopathy genes by reporting the phenotypic presentation of individuals with variants in *H2AC16*, *H2AZ1*, *H3C4*, and *H3-4*. As with the majority of previously described individuals with histonopathies, all five of these individuals present with developmental delay and/or intellectual disability, growth phenotypes, and variable systemic differences. Furthermore, we describe two additional individuals with variants in *H3-3B* (BLBS). This includes an affected individual with a cancer diagnosis, highlighting the necessity of investigating cancer predisposition in individuals with histonopathies more deeply while acknowledging there is not yet sufficient information to confirm or dismiss an association between germline histonopathies and cancer.

We performed pan-histonopathy and OMIM-delineated phenotypic quantifications of the 192 individuals with germline histonopathies, which included information from 62 affected individuals not included in a previous pan-histonopathy survey.[37] This report prioritizes granular data about each individual at the center of analyses, which highlights the heterogeneous clinical presentation associated with histonopathies. Affected individuals may exhibit variable phenotypes, which necessitate personalized clinical management, yet population-wide patterns are valuable. Collectively, they may point toward the convergence of proximate biology on common genes, pathways, or functional signatures that provide insight into the disease pathogenesis.[51] Currently, significant data gaps preclude our ability to assess population-wide patterns (Figures 2A–2C, 4A, and 4B), but identifying these gaps helps us prospectively shape the way we survey information. We recommend the implementation of a standardized survey that can be distributed to any clinician who cares for an individual with a germline histonopathy (Table S2). It will be imperative to periodically re-evaluate this form based on longitudinal re-assessment. Affected individuals and their families have so generously shared not just their data but also their stories with us.[79] They have taught both clinicians and researchers a tremendous amount in the past 10 years,[13] including that the clinical presentation associated with these syndromes is neither homogeneous nor static.

Our targeted phenotypic query coupled with our NGP facial analysis provides insights into germline histonopathies. While our analysis of clinician-reported phenotypic data was confounded by prohibitive data gaps, our NGP analysis showed images of individuals with HIST1H1E syndrome segregating away from images of individuals with BLBS and TBvH NDD in all age-stratified groups (Figures 3 and 4). These data may suggest that individuals with variants arising from the H1 linker protein have a distinct craniofacial gestalt from individuals with variants arising from proteins that make up the nucleosomal core. However, it is important to consider that the majority of images used in this analysis are of individuals with HIST1H1E syndrome (52.0% compared with 25.5% from individuals with BLBS and 22.5% from individuals with TBvH NDD). The facial analysis indicates that all three groups exhibit distinct facial phenotypes, with HIST1H1E displaying particularly clear differentiation. The sample sizes for BLBS and TBvH NDD were relatively small, which may contribute to the observed similarities in some cases, and the apparent differences between BLBS and TBvH NDD could be influenced by this limited sample size, as some individuals exhibit notably similar phenotypes. Further reanalysis with a larger cohort will likely be required to validate these findings and better distinguish the phenotypes of these groups. Additionally, these NGP analyses show that with increasing age, at an individual level, images may cluster less distinctly (Figures 3B and 3C). This could indicate that the facial gestalt becomes less consistent or recognizable over time.

The clear and consistent segregation of images of individuals with HIST1H1E syndrome may provide putative support for an unanticipated basis for delineation. Rather than core vs. linker, we often divide histones into RC histones and RI histones based on the biology underlying the unique features and functions of these two classes (Figure 1B). However, this prevailing delineation of RC/RI rather than core/linker likely reflects how little is known about the H1 linker histone, recently referred to as the "forgotten histone," when compared with the core proteins.[80,81] In fact, the crystal structure of an H1/nucleosome complex was not published until 2015, 18 years after the nucleosome was crystalized. Thus, this work highlights the need for both translational and foundational functional work geared toward delineating the roles of core and linker histones.

It is crucial to note this GMDB-based NGP analysis had a number of limitations that impact the strength of this putative linker vs. core dichotomy. From a technical perspective, images of individuals with HIST1H1E syndrome overrepresented, and the number of images per syndrome did not correlate with the number of reported individuals with each germline histonopathy (51 images of individuals with HIST1H1E syndrome, 55 reported individuals with HIST1H1E syndrome; 25 images of individuals with BLBS, 96 reported individuals with BLBS; 22 images of individuals with TBvH NDD, 34 reported individuals with TBvH NDD). Further, no previously unpublished individuals were represented in this analysis. Thus, we did not have representation of individuals with variants in H2A-encoding genes. No individuals have yet been reported with variants in H2B-encoding genes. This is worth noting because we cannot at this time rule out the possibility that the clustering of images of individuals with BLBS and TBvH NDD may specifically reflect the close association of H3 and H4 class proteins, which comprise the tetramer at the core of the nucleosome.[1] Future representation of individuals with H2A/H2B-based germline histonopathies may allow us to observe trends on the basis of core vs. linker or if there is a more specific delineation on the basis of H3/H4 tetramer vs. H2A/H2B heterodimer vs. H1 linker.[82,83]

With this lens, we then also interrogated the non-craniofacial clinician-reported data. Possible discordance in growth phenotypes have been highlighted in prior reports.[37] While initial HIST1H1E syndrome publications suggested that overgrowth (HP:0001548) was a defining phenotypic feature,[13,14] subsequent analyses showed that the linear growth pattern was complex, with some individuals showing decreased height (HP:0000002) percentiles over time.[14,15] An individual with BLBS was reported without a molecular diagnosis to exhibit overgrowth (HP:0001548) of unknown genetic etiology and intellectual disability (HP:0001249) following *in vitro* fertilization and intracytoplasmic sperm injection, with increased linear growth (HP:0000098) and macrocephaly (HP:0000256).[29] At follow-up evaluation a decade later, that same individual exhibited linear undergrowth (HP:0000002) and microcephaly (HP:0000252).[28,32] Finally, individuals with TBvH NDD show a statistically significant decrease in head size (HP:0000234, HP:0040195) percentile from birth to most recent evaluation.[35] Overall, individuals with germline histonopathies show a complex growth pattern, which could provide mechanistic insights if explored more deeply.

We sought to probe this growth phenotype in a way that accounted for confounding factors. Longitudinal assessment of linear growth is dually confounded by the prevalence of skeletal changes (HP:0000924) and endocrine anomalies (HP:0000818), which, in some cases, are treated with growth hormone therapy (Table S1; Figure S5). Additionally, longitudinal assessment of weight (HP:0004323) is confounded by (1) feeding difficulties (HP:0011968), which, in some cases, require intervention with gastrostomy tube; (2) tone and mobility differences (HP:0001270, HP:0001252, HP:0001276); and (3) associated behavioral or neurodevelopmental diagnoses (HP:0000708, HP:0000717, HP:0007018, HP:0012759, HP:0001249), which may be complicated by food aversion (Table S1; Figure S5). Thus, to assess a growth metric less affected by such confounders, we interrogated head size (HP:0000252, HP:0000256) over time in individuals for whom multiple measurements were available (Figure 5D). These data may suggest that, at a population level, individuals with HIST1H1E syndrome exhibit a macrocephalic (HP:0000256) trend over time while individuals with BLBS and TBvH NDD exhibit a microcephalic (HP:0000252) trend over time. This progressive microcephaly (HP:0000252) has recently been recapitulated in a BLBS mouse model.[31] While there is heterogeneity at an individual level, these temporal transitions will be interesting to interrogate mechanistically in lab-based models. Ultimately, in the ultra-rare disease space, the insights we continue to gain through the mining of longitudinal phenotypic data will enable us to optimize our model systems and guide us in the direction of translationally relevant questions.

Similarly, there has been a frequent and persistent question of whether individuals with germline histonopathies have an increased cancer risk, since the expansion of the sub-field of histone neurobiology was rooted in the identification of recurrent oncohistone mutations in high-grade pediatric glioma (HP:0009733). This has been further highlighted by previously observed phenotypic trends in individuals with germline histonopathies and raised the concern that cancer is a possible comorbidity of these syndromes, including overgrowth (HP:0001548) and premature aging (HP:0005616) (Figure 7A). Notably, genetic and epigenetic overgrowth disorders, such as Beckwith-Wiedemann Spectrum (BWSp [OMIM: 130650]), Sotos syndrome (OMIM: 11755), and Weaver syndrome (OMIM: 277590), are associated with an increased cancer predisposition.[84–89] Because embryonal tumors, including hepatoblastoma (HP:0002884) and Wilms tumor (HP:0002667), occur in around 10%–15% of children with BWSp, routine cancer surveillance protocols have been developed.[90,91] Conversely, the mildly elevated cancer risk (<5%) associated with Sotos and Weaver syndromes translates to a heightened clinical awareness for physicians without routine surveillance.[92] As with genetic/epigenetic overgrowth (HP:0001548) disorders, genetic disorders of premature aging (HP:0005616), such as ataxia-telangiectasia (OMIM:208900), dyskeratosis congenita (ORPHA:1775), and xeroderma pigmentosum (ORPHA:910), are also associated with an increased cancer risk.[93–98] Thus, individuals with all three OMIM-characterized germline histonopathies exhibit phenotypic features that are potentially associated with an increased cancer risk.

Additionally, existing functional evidence from mechanistic work performed on dermal fibroblasts from affected individuals with germline histonopathies and from zebrafish models support the validity of exploring this potential comorbidity more deeply (Figure 6B). In general, prior reports have asserted that cancer predisposition in individuals with germline variants causing Mendelian syndromes could be related to disruption of gene interaction networks recurrently found in cancer, including telomere maintenance, DNA damage response, and mTOR signaling.[99] We know that deficits in these processes have been observed in models of histonopathies (Figure 7B).[17,28,33–35] Specifically, prior work in dermal fibroblasts donated by individuals with HIST1H1E syndrome show accelerated senescence and an increased sensitivity to DNA damage.[17,37] Dermal fibroblasts donated by individuals with BLBS show increased cell proliferation coupled with an upregulation of genes implicated in mitosis and cell division.[28] Overcoming the constraints on cell proliferation is a key characteristic of cancer.[74] Finally, in addition to dermal fibroblasts donated by individuals with TBvH NDD that show differential expression of genes implicated in cell cycle progression, zebrafish models of this syndrome also show significant increases in double-stranded breaks and cell apoptosis.[33–35]

Until this report, one individual with a germline histonopathy had a reported cancer diagnosis.[35] Individual H4-32 harbored a p.His75Arg variant in *H4C9*,

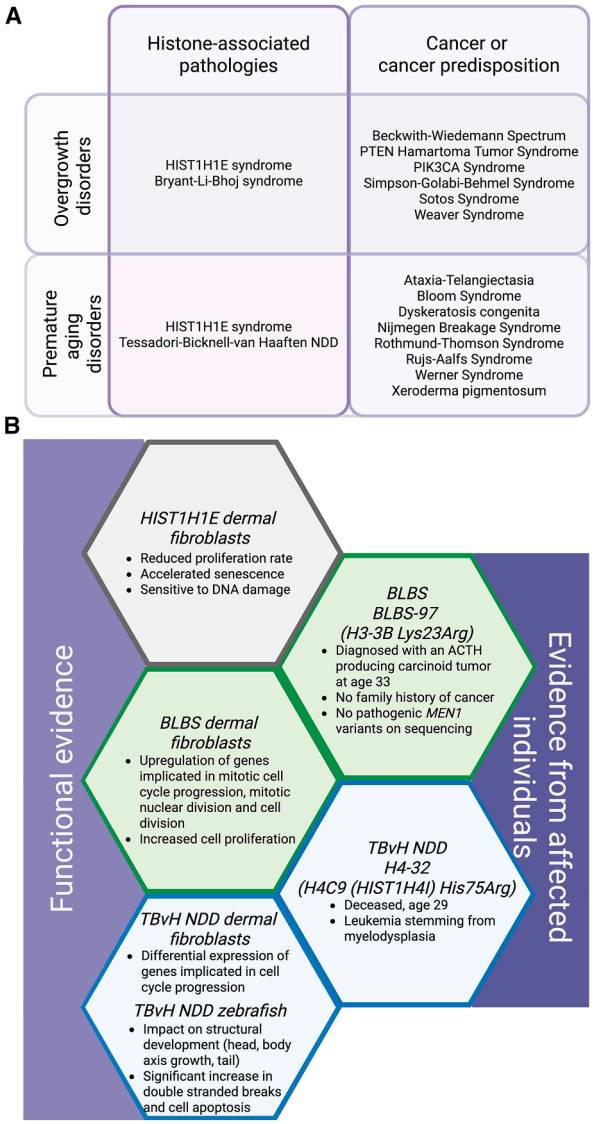


Figure 7. Existing data support the interrogation of cancer risk in individuals with germline histonopathies

(A) Some of the observed phenotypes in individuals with all three OMIM-identified germline histonopathies, including overgrowth and premature aging, overlap with phenotypes that are cardinal features of cancer-predisposition syndromes.

(B) Review of the existing functional evidence performed in dermal fibroblasts from affected individuals and zebrafish models of germline histonopathies motivate exploration of cancer as a comorbidity (left) and are potentially consistent with the two cancer diagnoses reported in the 191 person pan-histonopathy cohort (right).

NM_003495.3:c.227A>G; p.(His76Arg) and developed myelodysplasia (HP:0002863) that progressed to leukemia (HP:0001909), which ultimately caused death at age 29. Based on this report, as well as both the overlap of phenotypic features shared by individuals with germline histonopathies and individuals with known cancer-predisposition syndromes (Figure 7A) and functional work in models of germline histonopathies that showed mechanistic overlap with cancer pathogenesis (Figure 7B), we revised our clinician survey to systematically and explicitly ask about cancer diagnoses. Through this approach, we became aware of another individual with both a germline histonopathy and a cancer diagnosis. This individual was diagnosed with BLBS and cancer. Based on these known cases, the incidence of cancer as a comorbidity in individuals harboring a germline histone variant is 1%, which falls below the recommended 5% cutoff for routine surveillance.[92] However, this statistic is confounded by an inconsistent surveying of cancer frequency among the population, which has been previously noted.[32] This information gap can be remedied by the implementation of a standardized clinical survey Table S2), as well as a heightened awareness among both researchers and clinicians.

There remains an outstanding question of whether the cancers diagnosed in individuals H4-32 and BLBS-97 are independent of the underlying germline histone variants, whether these individuals perhaps harbor a dual molecular diagnosis wherein a non-histone secondary germline variant is associated with an increased cancer susceptibility, or whether germline histone variants themselves may be associated with a potential cancer predisposition.[100,101] However, several crucial limitations impede our ability to answer this question. Unlike the pancancer atlas of somatic histone mutations, we do not have access to paired tumor-germline genomes from these individuals, which hinders our ability to delineate what, if any, role the germline histone variants have in oncogenesis.[11] We also do not have access to the genetic testing results for these individuals, which would allow us to interrogate the possibility of a dual molecular diagnosis or secondary/incidental finding. An estimated 2%–7% of individuals with a syndromic phenotype receive a dual molecular diagnosis, wherein they have two monogenic disorders, and a similar percentage of individuals with a Mendelian syndrome have been reported to harbor a secondary or incidental finding, which is defined as a pathogenic sequence of variants found on genetic testing that are unassociated with a primary genetic diagnosis.[100–103] These incidental findings are particularly noteworthy because they are most commonly associated with cancer susceptibility at a rate of ~1.38%.[102,103] We also acknowledge our limited sample size at this time. Nonetheless, family members of affected individuals and clinicians caring for individuals with germline histonopathies often ask about cancer surveillance for this community. Functional work to understand the different mechanisms leading to NDDs compared with cancer is ongoing within the international consortium of labs interrogating germline histonopathies.[31] We also expect that the co-occurrence/mutual exclusivity analyses presented here will also provide helpful starting points for targeted mechanistic work (Figure 6C). This ongoing mechanistic interrogation coupled with consistent, longitudinal clinical surveying of affected individuals will be imperative to ensure that the gold standards of care are upheld for this community.

In sum, using clinician-reported data, we performed deep phenotypic quantification of individuals affected by germline histonopathies in an attempt to pull out phenotypic patterns (both craniofacial and systemic) shared by individuals. In addition, we describe seven previously unreported individuals affected by histonopathies, which adds to the fields' understanding of this group of disorders and empowers affected individuals and their families. Our analysis highlights a dramatic need for more comprehensive and thoughtful gathering of clinical data as well as potentially shifts focus for treatment and clinical care for individuals with histonopathies.

Subjects and methods

This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia. Informed consent was obtained for all individuals. Analyses and graphs were made in Microsoft Excel, GraphPad Prism v8, or R v4.3.2. Graphics were generated with BioRender.

Interpreting gnomAD constraint metrics

The Genome Aggregation Database (gnomAD) is a publicly available resource that aims to summarize exome and genome sequencing data from 141,456 individuals in gnomAD v2 and 807,162 individuals in gnomAD v4.[104,105] Collating such massive datasets enables the gnomAD investigators to calculate constraint metrics, which suggest how tolerant a particular gene is to missense or loss-of-function variants. The rare observation of variants in a gene suggests that that genes cannot tolerate the accumulation of variants without a deleterious effect.[104]

Cohort assembly—Published individuals

Prior to this report, 185 individuals with molecularly confirmed germline histonopathies had been reported in the literature.[13–36] Through a comprehensive literature review, we aggregated the published phenotypic information for all affected individuals (Table S1).

Cohort assembly—Previously unreported individuals

Previously unreported individuals were referred by their clinicians through GeneMatcher, DECIPHER, or direct communication with E.E.L./E.J.K.B.[106,107] Referring clinicians were sent the standardized survey developed based on the phenotypic assessment of 185 previously reported individuals (Table S4).

Quantification of clinician-reported phenotypes

Once the phenotypic information for all 192 individuals was aggregated, they were subdivided into four main categories: neurologic, non-neurologic systemic, growth, and craniofacial (gross). Neurologic features included developmental delay/intellectual disability (HP:0012759, HP0001249), delayed motor development (HP:0001270), delayed speech development (HP:0000750), anomalies on brain MRI (HP:0030890, HP:0007103), hypotonia (HP:0001252), hypertonia (HP:0001276), seizures (HP:0011145), comorbid behavioral/neurodevelopmental diagnosis (HP:0000708, HP:0000717, HP:0007018), sleep disturbances (HP:0002360), changes in hearing (HP:0000365), changes in vision (HP:0000657, HP:0000504, HP:0000496), as well as "other" to capture any features that did not fit into these categories. Non-neurologic systemic features were subdivided into cardiac/circulatory (HP:0500015, HP:0011028), gastrointestinal (HP:0011024, HP:0512718), feeding difficulties (HP:0011968), genitourinary (HP:0000119), signs of advanced aging (HP:0005616), dermatologic (HP:0001005), history of malignancy (HP:0002664), frequent infections/cytopenia/fever (HP:0001945, HP:0002719), other genetic variants, as well as other to capture any features that did not fit into these categories. Growth encompassed changes to head shape (HP:0000234), head size (HP:0000252, HP:0000256, at birth, last examination, any additional examinations), skeletal changes (HP:0000924), height (HP:0000002, at birth, last examination, any additional examinations), weight (HP:0004323, at birth, last examination, any additional examinations), and endocrine (HP:0000818). These features were grouped because it was determined that changes to head shape, including craniosynostosis (HP:0001363), could confound assessments of head size, that skeletal (HP:0000924) and endocrine (HP:0000818) conditions could confound assessments of linear growth, and that endocrine (HP:0000818) conditions could confound assessments of weight (HP:0004323). Finally, the gross craniofacial regions were defined as forehead, periorbital region, nose and philtrum, perioral region, mandible (distinct from other perioral features), ear, and face and cheeks.

To quantify the prevalence of phenotypic features across the full cohort, as well as in sub-cohorts delineated based on OMIM-characterized syndromes, for all 29 features queried in this analysis, a value of 1 was assigned to all individuals to determine the total number of possible responses (pan-histonopathy = 192; HIST1H1E syndrome = 55; BLBS = 96; TBvH NDD = 34). Then, a value of 1 was assigned for any active responses (yes or no) while a value of 0 was assigned for any queries for which a response was unavailable. Finally, for all queries that received an active response, a value of 1 was assigned if a particular feature was affected in an individual, as reported by the referring clinician, while a value of 0 was assigned if a particular feature was unaffected in an individual, as reported by the referring clinician. Substratified quantifications can be found in Table S1.

GMDB analysis

Only previously published images of affected individuals were included in this analysis. We utilized GestaltMatcher1 to analyze facial phenotypic similarities among 92 patient images in our cohort, encompassing patients with HIST1H1E, H3.3, and H4. To mitigate age-related bias, we categorized the cohort into three age groups: younger than 6 years (40 images), 6–12 years (24 images), and older than 12 years (28 images) and analyzed them separately. GestaltMatcher was trained on 8,547 images spanning 244 disorders from the GMDB2 to learn facial dysmorphic features. To avoid the bias caused during the model training, we ensured that all the HIST1H1E, H3.3, and H4 patients were removed from the training set. We applied test-time augmentation and model ensemble techniques to encode each image into 12 facial phenotype descriptors (FPDs). Facial phenotypic similarity between two images was quantified by averaging the cosine distances across the 12 FPDs, with smaller distances indicating greater similarity in phenotype space. Using these metrics, we assessed both (1) inter-cohort similarities and (2) pairwise patient-level similarities, providing insight into both cohort and patient-level similarities.

We first analyzed the inter-cohort similarity among HIST1H1E, H3.3, and H4 patients. To visualize the distribution of patients, we utilized t-SNE3 to map their facial phenotypic features into a two-dimensional space (Figure S2). Next, we performed random sampling to calculate the distance distributions between two patient groups. Specifically, we sampled two types of distributions from the GMDB: (1) distances between cohorts sampled from the same disorder, and (2) distances between cohorts sampled from two different disorders. Using these distributions, we conducted a receiver operating characteristic (ROC) analysis and determined a threshold value ($c = 0.914$) for distinguishing between groups representing different disorders. This threshold achieved a sensitivity of 0.842 and a specificity of 0.913. Finally, we calculated the percentage of distance distributions between each pair of the three groups (HIST1H1E, H3.3, and H4) above the threshold, providing insights into their inter-cohort phenotypic differences.

For the patient-level analysis, we conducted pairwise comparisons. To simulate a real-world scenario, we populated the phenotype space with 8,547 patient images spanning 244 disorders, which served as controls. We iterated through the cohort for each test image, placing the remaining images into the phenotype space. We then calculated the rank of each image in the phenotype space relative to the test image. All data from the GMDB analysis can be found in Table S2.

Integrated cBioPortal for Cancer Genomics and PeCan analyses

Comprehensive lists of somatic mutations in *H3-3B* and *H4C9* were compiled through intersecting searches of cBioPortal for Cancer Genomics and PeCan.[71–73] Through the cBioPortal web browser (https://www.cbioportal.org), we built a custom case set in the manual query section that included all 255,345 cases that comprise the dataset as of April 3, 2024. Because queries are limited by gene and sample count, we first created a user-defined gene list that included *H3-3B*. These searches enabled us to export tables from the Mutations tab with a list of all reported somatic variants in these two genes across the 34 categories of datasets compiled in cBioPortal (Table S3). Based on these searches, we also exported the tables from the Cancer Types Summary Tab (Figure 6B). These lists of somatic mutations were then amended with non-overlapping mutations aggregated in PeCan (https://pecan.stjude.cloud), which were found in the ProteinPaint tab after performing a search based on gene. These combined lists facilitated the generation of Figure 6A.

Due to limitations on gene/sample count in cBioPortal, we then performed more targeted analyses of co-occurring and mutually exclusive mutations, motivated by the diagnosis of individual BLBS-97. Since individual BLBS-97 was diagnosed with an ACTH-producing pulmonary carcinoid (HP:0030445), a curated list of relevant genes was queried for the co-occurrence or mutual exclusivity with somatic variants in *H3-3B* in a restricted set of studies focused on pancancer interrogations as well as lung neoplasms (HP:0100526, Figure 6C). All data can be found in Table S3.

Data and code availability

The published article includes all datasets/code generated or analyzed during this study within the text, figures, and supplemental files.

Acknowledgments

This work would not be possible without the generosity of affected individuals and their families who entrust us with their stories and their data. We would like to celebrate the members of the HIST1H1E syndrome community, who have built an incredible foundation of support for one another (https://hist1h1e.org/). We recognize that many individuals with unnamed ultra-rare Mendelian disorders, including some with germline histonopathies, may not have access to syndrome-specific online groups or family foundations yet. It is our sincere hope that projects like these will not only lead to more information for families of affected individuals and shorter diagnostic odysseys in the future, but also more resources and community as families navigate their journeys with these syndromes. We would like to thank the members of the Germline Histone Mutations in Human Disease International Consortium for their commitment to data sharing and collaboration to advance our collective research efforts. We would also like to thank attendees of the 43rd and 44th Annual David W. Smith Workshops on Malformations and Morphogenesis for their helpful discussion, with special acknowledgment of Dr. Chaya Murali, who encouraged us to interrogate the comorbidities associated with these syndromes to enhance clinical care. Finally, we would like to acknowledge publicly available resources that were developed with the rare disease community in mind. This study makes use of data generated by the DECIPHER community. A full list of centers that contributed to the generation of the data is available from https://deciphergenomics.org/about/stats and via email from contact@deciphergenomics.org. DECIPHER is hosted by EMBL-EBI and funding for the DECIPHER project was provided by the Wellcome Trust (grant number WT223718/Z/21/Z). This project also uses and displays data and algorithms from the Monarch Initiative. The Monarch Initiative (https://monarchinitiative.org) makes biomedical knowledge exploration more efficient and effective by providing tools for genotype-phenotype analysis, genomic diagnostics, and precision medicine across broad areas of disease. Finally, some of the authors (T.S.B., M.d.l.A.G.C., M.P.-B.) of this publication are members of the European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability ERN-ITHACA (EU Framework Partnership Agreement ID: 3HP-HP-FPA ERN-01-2016/739516).

Support for this work was provided by NICHD F30 1F30HD112125 (E.E.L.); NIGMS T32 5T32GM008638 (E.L.D., T. T.N.); NHGRI T32HG009495 and the Eagles Autism Foundation (D.E.L.-C.); ongoing support by the Netherlands Organisation for Scientific Research (ZonMw Vidi, grant 09150172110002) (T.S.B.); support from the Royal Children's Hospital Foundation, the Murdoch Children's Research Institute, the Harbig Foundation, and the Victorian Government's Operational Infrastructure Support Program (T.Y.T., L.G.); the Chan-Zuckerberg Initiative Neurodegeneration Challenge Network (R.C.A.-N., E.J.K.B.); as well as the Burroughs Wellcome Fund and the Hartwell Foundation (E.J.K.B.).

Author contributions

E.E.L., E.M.G., A.K.S., D.E.L.-C., and E.J.K.B. collected and compiled clinical data for all participants; E.E.L., E.M.G., A.K.S., E.L.D., H.K., J.-M.L., S.M.S. analyzed and interpreted the data; E.E.L. drafted the manuscript; E.M.G., A.K.S., E.L.D., H.K., J.-M. L., S.M.S., D.E.L.-C., K.J.C., A.J.M.-P., X.M.W., R.A., T.S.B., S.M.,

B.C., S.K., Y.H.-H., M.R., R.P., M.d.l.A.G.C., M.P.-B., T.B.A., S.P. M., R.C.A.-N., T.T.N., and E.J.K.B. contributed the critical review of the manuscript; T.S.B., S.M., B.C., S.K., Y.H.-H., M.R., R.P., M.d. l.A.G.C., M.P.-B., and T.B.A. diagnosed affected individuals and provided clinical data; and E.E.L., E.M.G., A.K.S., T.T.N., and E. J.K.B. conceived of the study and review of the manuscript.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2025.100440.

Web resources

cBioPortal: https://www.cbioportal.org
GestaltMatcher Database: https://db.gestaltmatcher.org/publications?locale = en
St. Jude PeCan platform: https://pecan.stjude.cloud

Received: February 26, 2025
Accepted: April 10, 2025

References

1. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260. https://doi.org/10.1038/38444.
2. Fyodorov, D.V., Zhou, B.-R., Skoultchi, A.I., and Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function. Nat. Rev. Mol. Cell Biol. *19*, 192–206. https://doi.org/10.1038/nrm.2017.94.
3. Lubin, E., Bryant, L., Aicher, J., Li, D., and Bhoj, E. (2022). Analysis of histone variant constraint and tissue expression suggests five potential novel human disease genes: H2AFY2, H2AFZ, H2AFY, H2AFV, H1F0. Hum. Genet. *141*, 1409–1421. https://doi.org/10.1007/s00439-022-02432-1.
4. St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat. Genet. *44*, 251–253. https://doi.org/10.1038/ng.1102.
5. Schwartzentruber, J., Korshunov, A., Liu, X.-Y., Jones, D.T. W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.-A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226–231. https://doi.org/10.1038/nature10833.
6. Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M. A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333–339. https://doi.org/10.1038/nature12634.
7. Behjati, S., Tarpey, P.S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D.C., Cooke, S.L., Gundem, G., Davies, H., et al. (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat. Genet. *45*, 1479–1482. https://doi.org/10.1038/ng.2814.
8. Lindroth, A.M., and Plass, C. (2013). Recurrent H3.3 alterations in childhood tumors. Nat. Genet. *45*, 1413–1414. https://doi.org/10.1038/ng.2832.
9. Andrade, A.F., Chen, C.C.L., and Jabado, N. (2023). Oncohistones in brain tumors: the soil and seed. Trends Cancer *9*, 444–455. https://doi.org/10.1016/j.trecan.2023.02.003.
10. Espinoza Pereira, K.N., Shan, J., Licht, J.D., and Bennett, R.L. (2023). Histone mutations in cancer. Biochem. Soc. Trans. *51*, 1749–1763. https://doi.org/10.1042/BST20210567.
11. Bonner, E.R., Dawood, A., Gordish-Dressman, H., Eze, A., Bhattacharya, S., Yadavilli, S., Mueller, S., Waszak, S.M., and Nazarian, J. (2023). Pan-cancer atlas of somatic core and linker histone mutations. npj Genom. Med *8*, 23. https://doi.org/10.1038/s41525-023-00367-8.
12. Bjornsson, H.T. (2015). The Mendelian disorders of the epigenetic machinery. Genome Res. *25*, 1473–1481. https://doi.org/10.1101/gr.190629.115.
13. Helsmoortel, C., Vandeweyer, G., Ordoukhanian, P., Van Nieuwerburgh, F., Van Der Aa, N., and Kooy, R.F. (2015). Challenges and opportunities in the investigation of unexplained intellectual disability using family-based whole-exome sequencing. Clin. Genet. *88*, 140–148. https://doi.org/10.1111/cge.12470.
14. Tatton-Brown, K., Loveday, C., Yost, S., Clarke, M., Ramsay, E., Zachariou, A., Elliott, A., Wylie, H., Ardissone, A., Rittinger, O., et al. (2017). Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. Am. J. Hum. Genet. *100*, 725–736. https://doi.org/10.1016/j.ajhg.2017.03.010.
15. Takenouchi, T., Uehara, T., Kosaki, K., and Mizuno, S. (2018). Growth pattern of Rahman syndrome. Am. J. Med. Genet. *176*, 712–714. https://doi.org/10.1002/ajmg.a.38616.
16. Duffney, L.J., Valdez, P., Tremblay, M.W., Cao, X., Montgomery, S., McConkie-Rosell, A., and Jiang, Y.H. (2018). Epigenetics and autism spectrum disorder: A report of an autism case with mutation in H1 linker histone *HIST1H1E* and literature review. American J of Med Genetics Pt B *177*, 426–433. https://doi.org/10.1002/ajmg.b.32631.
17. Flex, E., Martinelli, S., Van Dijck, A., Ciolfi, A., Cecchetti, S., Coluzzi, E., Pannone, L., Andreoli, C., Radio, F.C., Pizzi, S., et al. (2019). Aberrant Function of the C-Terminal Tail of HIST1H1E Accelerates Cellular Senescence and Causes Premature Aging. Am. J. Hum. Genet. *105*, 493–508. https://doi.org/10.1016/j.ajhg.2019.07.007.
18. Burkardt, D.D., Zachariou, A., Loveday, C., Allen, C.L., Amor, D.J., Ardissone, A., Banka, S., Bourgois, A., Coubes, C., Cytrynbaum, C., et al. (2019). *HIST1H1E* heterozygous protein-truncating variants cause a recognizable syndrome with intellectual disability and distinctive facial gestalt: A study to clarify the HIST1H1E syndrome phenotype in 30 individuals. Am. J. Med. Genet. *179*, 2049–2055. https://doi.org/10.1002/ajmg.a.61321.
19. Ciolfi, A., Aref-Eshghi, E., Pizzi, S., Pedace, L., Miele, E., Kerkhof, J., Flex, E., Martinelli, S., Radio, F.C., Ruivenkamp, C. A.L., et al. (2020). Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. Clin. Epigenet. *12*, 7. https://doi.org/10.1186/s13148-019-0804-0.
20. Li, Y., Luo, Y., Sun, G., Baoerhan, R., Julaiti, D., and Maimaiti, M. (2020). Rahman syndrome caused by HIST1H1E

gene mutation: a case report and literature review. Journal of Clinical Pediatrics *38*, 347–350. https://doi.org/10.3969/j.issn.1000-3606.2020.05.008.

21. Ahmed, M.S.O., Rafey, M., McDonnell, T., and Smith, D. (2021). HIST1H1E syndrome with type 2 diabetes. BMJ Case Rep. *14*, e241907. https://doi.org/10.1136/bcr-2021-241907.
22. Indugula, S.R., Ayala, S.S., Vetrini, F., Belonis, A., and Zhang, W. (2022). Exome sequencing identified a novel HIST1H1E heterozygous protein-truncating variant in a 6-month-old male patient with Rahman syndrome: A case report. Clin. Case Rep. *10*, e05370. https://doi.org/10.1002/ccr3.5370.
23. Zhao, J., Lyu, G., Ding, C., Wang, X., Li, J., Zhang, W., Yang, X., and Zhang, V.W. (2022). Expanding the mutational spectrum of Rahman syndrome: A rare disorder with severe intellectual disability and particular facial features in two Chinese patients. Mol. Genet. Genomic Med. *10*, e1825. https://doi.org/10.1002/mgg3.1825.
24. Tanabe, Y., Nomura, N., Minami, M., Takaya, J., Okamoto, N., Yanagi, K., Kaname, T., Fujii, Y., and Kaneko, K. (2023). HIST1H1E syndrome with deficiency in multiple pituitary hormones. Clin. Pediatr. Endocrinol. *32*, 195–198. https://doi.org/10.1297/cpe.2023-0002.
25. Yüksel Ülker, A., Uludağ Alkaya, D., Çağlayan, A.O., Usluer, E., Aykut, A., Aslanger, A., Vural, M., and Tüysüz, B. (2023). An investigation of the etiology and follow-up findings in 35 children with overgrowth syndromes, including biallelic *SUZ12* variant. Am. J. Med. Genet. *191*, 1530–1545. https://doi.org/10.1002/ajmg.a.63180.
26. Zhao, W., Zhang, Y., Lv, T., He, J., and Zhu, B. (2023). A case report of a novel *HIST1H1E* mutation and a review of the bibliography to evaluate the genotype–phenotype correlations. Mol. Genet. Genomic Med. *11*, e2273. https://doi.org/10.1002/mgg3.2273.
27. Maver, A., Čuturilo, G., Ruml, S.J., and Peterlin, B. (2019). Clinical next generation sequencing reveals an *H3F3A* gene as a new potential gene candidate for microcephaly associated with severe developmental delay, intellectual disability and growth retardation. Balkan J. Med. Genet. *22*, 65–68. https://doi.org/10.2478/bjmg-2019-0028.
28. Bryant, L., Li, D., Cox, S.G., Marchione, D., Joiner, E.F., Wilson, K., Janssen, K., Lee, P., March, M.E., Nair, D., et al. (2020). Histone H3.3 beyond cancer: Germline mutations in *Histone 3 Family 3A and 3B* cause a previously unidentified neurodegenerative disorder in 46 patients. Sci. Adv. *6*, eabc9207. https://doi.org/10.1126/sciadv.abc9207.
29. Shah, P.S., Weksberg, R., and Chitayat, D. (2006). Overgrowth with severe developmental delay following IVF/ICSI: A newly recognized syndrome? Am. J. Med. Genet. *140*, 1312–1315. https://doi.org/10.1002/ajmg.a.31274.
30. Okur, V., Chen, Z., Vossaert, L., Peacock, S., Rosenfeld, J., Zhao, L., Du, H., Calamaro, E., Gerard, A., Zhao, S., et al. (2021). De novo variants in H3-3A and H3-3B are associated with neurodevelopmental delay, dysmorphic features, and structural brain abnormalities. npj Genom. Med. *6*, 104. https://doi.org/10.1038/s41525-021-00268-8.
31. Khazaei, S., Chen, C.C.L., Andrade, A.F., Kabir, N., Azarafshar, P., Morcos, S.M., França, J.A., Lopes, M., Lund, P.J., Danieau, G., et al. (2023). Single substitution in H3.3G34 alters DNMT3A recruitment to cause progressive neurodegeneration. Cell *186*, 1162–1178.e20. https://doi.org/10.1016/j.cell.2023.02.023.
32. Layo-Carris, D.E., Lubin, E.E., Sangree, A.K., Clark, K.J., Durham, E.L., Gonzalez, E.M., Smith, S., Angireddy, R., Wang, X.M., Weiss, E., et al. (2024). Expanded phenotypic spectrum of neurodevelopmental and neurodegenerative disorder Bryant-Li-Bhoj syndrome with 38 additional individuals. Eur. J. Hum. Genet. *32*, 928–937. https://doi.org/10.1038/s41431-024-01610-1.
33. Tessadori, F., Giltay, J.C., Hurst, J.A., Massink, M.P., Duran, K., Vos, H.R., van Es, R.M., Scott, R.H., Van Gassen, K.L.I., et al.; Deciphering Developmental Disorders Study (2017). Germline mutations affecting the histone H4 core cause a developmental syndrome by altering DNA damage response and cell cycle control. Nat. Genet. *49*, 1642–1646. https://doi.org/10.1038/ng.3956.
34. Tessadori, F., Rehman, A.U., Giltay, J.C., Xia, F., Streff, H., Duran, K., Bakkers, J., Lalani, S.R., and Van Haaften, G. (2020). A de novo variant in the human HIST1H4J gene causes a syndrome analogous to the HIST1H4C-associated neurodevelopmental disorder. Eur. J. Hum. Genet. *28*, 674–678. https://doi.org/10.1038/s41431-019-0552-9.
35. Tessadori, F., Duran, K., Knapp, K., Fellner, M., Smithson, S., Beleza Meireles, A., Elting, M.W., O'Donnell-Luria, A., O'Donnell-Luria, A., et al.; Deciphering Developmental Disorders Study (2022). Recurrent de novo missense variants across multiple histone H4 genes underlie a neurodevelopmental syndrome. Am. J. Hum. Genet. *109*, 750–758. https://doi.org/10.1016/j.ajhg.2022.02.003.
36. Borja, N., Borjas-Mendoza, P., Bivona, S., Peart, L., Gonzalez, J., Johnson, B.K., Guo, S., Yusupov, R., Network, U.D., Bademci, G., et al. (2023). *H4C5* missense variant leads to a neurodevelopmental phenotype overlapping with Angelman syndrome. American J of Med Genetics Pt A *191*, 1911–1916. https://doi.org/10.1002/ajmg.a.63193.
37. Knapp, K., Naik, N., Ray, S., Van Haaften, G., and Bicknell, L.S. (2023). Histones: coming of age in Mendelian genetic disorders. J. Med. Genet. *60*, 213–222. https://doi.org/10.1136/jmg-2022-109085.
38. 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. https://doi.org/10.1038/gim.2015.30.
39. Antonarakis, S.E. (2021). History of the methodology of disease gene identification. Am. J. Med. Genet. *185*, 3266–3275. https://doi.org/10.1002/ajmg.a.62400.
40. Seaby, E.G., Rehm, H.L., and O'Donnell-Luria, A. (2021). Strategies to Uplift Novel Mendelian Gene Discovery for Improved Clinical Outcomes. Front. Genet. *12*, 674295. https://doi.org/10.3389/fgene.2021.674295.
41. Hart, T.C., and Hart, P.S. (2009). Genetic studies of craniofacial anomalies: clinical implications and applications. Orthod. Craniofac. Res. *12*, 212–220. https://doi.org/10.1111/j.1601-6343.2009.01455.x.
42. Hsieh, T.C., and Krawitz, P.M. (2023). Computational facial analysis for rare Mendelian disorders. American J of Med Genetics Pt C *193*, e32061. https://doi.org/10.1002/ajmg.c.32061.
43. Lumaka, A., Cosemans, N., Lulebo Mampasi, A., Mubungu, G., Mvuama, N., Lubala, T., Mbuyi-Musanzayi, S., Breckpot, J., Holvoet, M., De Ravel, T., et al. (2017). Facial

dysmorphism is influenced by ethnic background of the patient and of the evaluator. Clin. Genet. *92*, 166–171. https://doi.org/10.1111/cge.12948.

44. Kruszka, P., Tekendo-Ngongang, C., and Muenke, M. (2019). Diversity and dysmorphology. Curr. Opin. Pediatr. *31*, 702–707. https://doi.org/10.1097/MOP.0000000000000816.
45. Kruszka, P., and Tekendo-Ngongang, C. (2023). Application of facial analysis Technology in Clinical Genetics: Considerations for diverse populations. American J of Med Genetics Pt C *193*, e32059. https://doi.org/10.1002/ajmg.c.32059.
46. Hsieh, T.-C., Bar-Haim, A., Moosa, S., Ehmke, N., Gripp, K. W., Pantel, J.T., Danyel, M., Mensah, M.A., Horn, D., Rosnev, S., et al. (2022). GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat. Genet. *54*, 349–357. https://doi.org/10.1038/s41588-021-01010-x.
47. Lesmann, H., Hustinx, A., Moosa, S., Klinkhammer, H., Marchi, E., Caro, P., Abdelrazek, I.M., Pantel, J.T., Ten Hagen, M., Thong, M.-K., et al. (2023). GestaltMatcher Database - A global reference for facial phenotypic variability in rare human diseases. Preprint at medRxiv. https://doi.org/10.1101/2023.06.06.23290887.
48. Merks, J.H.M., Ozgen, H.M., Koster, J., Zwinderman, A.H., Caron, H.N., and Hennekam, R.C.M. (2008). Prevalence and patterns of morphological abnormalities in patients with childhood cancer. JAMA *299*, 61–69. https://doi.org/10.1001/jama.2007.66.
49. Lupo, P.J., Schraw, J.M., Desrosiers, T.A., Nembhard, W.N., Langlois, P.H., Canfield, M.A., Copeland, G., Meyer, R.E., Brown, A.L., Chambers, T.M., et al. (2019). Association Between Birth Defects and Cancer Risk Among Children and Adolescents in a Population-Based Assessment of 10 Million Live Births. JAMA Oncol. *5*, 1150–1158. https://doi.org/10.1001/jamaoncol.2019.1215.
50. Aaltonen, L.A., Abascal, F., Abeshouse, A., Aburatani, H., Adams, D.J., Agrawal, N., Ahn, K.S., Ahn, S.-M., Aikata, H., et al.; ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93. https://doi.org/10.1038/s41586-020-1969-6.
51. Minaya, M.A., Mahali, S., Iyer, A.K., Eteleeb, A.M., Martinez, R., Huang, G., Budde, J., Temple, S., Nana, A.L., Seeley, W. W., et al. (2023). Conserved gene signatures shared among MAPT mutations reveal defects in calcium signaling. Front. Mol. Biosci. *10*, 1051494. https://doi.org/10.3389/fmolb.2023.1051494.
52. McMurry, J.A., Köhler, S., Washington, N.L., Balhoff, J.P., Borromeo, C., Brush, M., Carbon, S., Conlin, T., Dunn, N., Engelstad, M., et al. (2016). Navigating the Phenotype Frontier: The Monarch Initiative. Genetics *203*, 1491–1495. https://doi.org/10.1534/genetics.116.188870.
53. Shefchek, K.A., Harris, N.L., Gargano, M., Matentzoglu, N., Unni, D., Brush, M., Keith, D., Conlin, T., Vasilevsky, N., Zhang, X.A., et al. (2020). The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. *48*, D704–D715. https://doi.org/10.1093/nar/gkz997.
54. Clark, K.J., Lubin, E.E., Gonzalez, E.M., Sangree, A.K., Layo-Carris, D.E., Durham, E.L., Ahrens-Nicklas, R.C., Nomakuchi, T.T., and Bhoj, E.J. (2024). NeuroTri2-VISDOT: An open-access tool to harness the power of second trimester human single cell data to inform models of Mendelian neurodevelopmental disorders. Preprint at bioRxiv. https://doi.org/10.1101/2024.02.01.578438.
55. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.
56. Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483–489. https://doi.org/10.1038/nature10523.
57. Allanson, J.E., Biesecker, L.G., Carey, J.C., and Hennekam, R.C.M. (2009). Elements of morphology: Introduction. Am. J. Med. Genet. *149A*, 2–5. https://doi.org/10.1002/ajmg.a.32601.
58. Allanson, J.E., Cunniff, C., Hoyme, H.E., McGaughran, J., Muenke, M., and Neri, G. (2009). Elements of morphology: Standard terminology for the head and face. Am. J. Med. Genet. *149A*, 6–28. https://doi.org/10.1002/ajmg.a.32612.
59. Hall, B.D., Graham, J.M., Cassidy, S.B., and Opitz, J.M. (2009). Elements of morphology: Standard terminology for the periorbital region. Am. J. Med. Genet. *149A*, 29–39. https://doi.org/10.1002/ajmg.a.32597.
60. Hunter, A., Frias, J.L., Gillessen-Kaesbach, G., Hughes, H., Jones, K.L., and Wilson, L. (2009). Elements of morphology: Standard terminology for the ear. Am. J. Med. Genet. *149A*, 40–60. https://doi.org/10.1002/ajmg.a.32599.
61. Hennekam, R.C.M., Cormier-Daire, V., Hall, J.G., Méhes, K., Patton, M., and Stevenson, R.E. (2009). Elements of morphology: Standard terminology for the nose and philtrum. Am. J. Med. Genet. *149A*, 61–76. https://doi.org/10.1002/ajmg.a.32600.
62. Carey, J.C., Cohen, M.M., Curry, C.J.R., Devriendt, K., Holmes, L.B., and Verloes, A. (2009). Elements of morphology: Standard terminology for the lips, mouth, and oral region. Am. J. Med. Genet. *149A*, 77–92. https://doi.org/10.1002/ajmg.a.32602.
63. Asif, M., Kaygusuz, E., Shinawi, M., Nickelsen, A., Hsieh, T.-C., Wagle, P., Budde, B.S., Hochscherf, J., Abdullah, U., Höning, S., et al. (2022). De novo variants of CSNK2B cause a new intellectual disability-craniodigital syndrome by disrupting the canonical Wnt signaling pathway. HGG Adv. *3*, 100111. https://doi.org/10.1016/j.xhgg.2022.100111.
64. Aerden, M., Denommé-Pichon, A.-S., Bonneau, D., Bruel, A.-L., Delanne, J., Gérard, B., Mazel, B., Philippe, C., Pinson, L., Prouteau, C., et al. (2023). The neurodevelopmental and facial phenotype in individuals with a TRIP12 variant. Eur. J. Hum. Genet. *31*, 461–468. https://doi.org/10.1038/s41431-023-01307-x.
65. Blackburn, P.R., Ebstein, F., Hsieh, T.-C., Motta, M., Radio, F.C., Herkert, J.C., Rinne, T., Thiffault, I., Rapp, M., Alders, M., et al. (2023). Loss-of-function variants in *CUL3* cause a syndromic neurodevelopmental disorder. Preprint at medRxiv. https://doi.org/10.1101/2023.06.13.23290941.
66. Ebstein, F., Küry, S., Most, V., Rosenfelt, C., Scott-Boyer, M.-P., Van Woerden, G.M., Besnard, T., Papendorf, J.J., Studencka-Turski, M., Wang, T., et al. (2023). PSMC3 proteasome subunit variants are associated with neurodevelopmental delay and type I interferon production. Sci. Transl. Med. *15*, eabo3189. https://doi.org/10.1126/scitranslmed.abo3189.
67. Nacev, B.A., Feng, L., Bagert, J.D., Lemiesz, A.E., Gao, J., Soshnev, A.A., Kundra, R., Schultz, N., Muir, T.W., and Allis,

C.D. (2019). The expanding landscape of 'oncohistone' mutations in human cancers. Nature *567*, 473–478. https://doi.org/10.1038/s41586-019-1038-1.

68. Simó-Riudalbas, L., Pérez-Salvia, M., Setien, F., Villanueva, A., Moutinho, C., Martínez-Cardús, A., Moran, S., Berdasco, M., Gomez, A., Vidal, E., et al. (2015). KAT6B Is a Tumor Suppressor Histone H3 Lysine 23 Acetyltransferase Undergoing Genomic Loss in Small Cell Lung Cancer. Cancer Res. *75*, 3936–3945. https://doi.org/10.1158/0008-5472.CAN-14-3702.
69. Zhang, L.X., Lemire, G., Gonzaga-Jauregui, C., Molidperee, S., Galaz-Montoya, C., Liu, D.S., Verloes, A., Shillington, A.G., Izumi, K., Ritter, A.L., et al. (2020). Further delineation of the clinical spectrum of KAT6B disorders and allelic series of pathogenic variants. Genet. Med. *22*, 1338–1347. https://doi.org/10.1038/s41436-020-0811-8.
70. Knight, S., VanHouwelingen, L., Cervi, D., Clay, M.R., Corkins, M., Hines-Dowell, S., Hamilton, K.V., Mostafavi, R., Ward, J., Furman, W.L., and Murphy, A.J. (2018). Genitopatellar syndrome and neuroblastoma: The multidisciplinary management of a previously unreported association. Pediatr. Blood Cancer *65*, e27373. https://doi.org/10.1002/pbc.27373.
71. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. *6*, pl1. https://doi.org/10.1126/scisignal.2004088.
72. De Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T.N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., et al. (2023). Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Res. *83*, 3861–3867. https://doi.org/10.1158/0008-5472.CAN-23-0816.
73. McLeod, C., Gout, A.M., Zhou, X., Thrasher, A., Rahbarinia, D., Brady, S.W., Macias, M., Birch, K., Finkelstein, D., Sunny, J., et al. (2021). St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. Cancer Discov. *11*, 1082–1099. https://doi.org/10.1158/2159-8290.CD-20-1230.
74. Bolton, K.L., Ptashkin, R.N., Gao, T., Braunstein, L., Devlin, S.M., Kelly, D., Patel, M., Berthon, A., Syed, A., Yabe, M., et al. (2020). Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat. Genet. *52*, 1219–1226. https://doi.org/10.1038/s41588-020-00710-0.
75. Zverinova, S., and Guryev, V. (2022). Variant calling: Considerations, practices, and developments. Hum. Mutat. *43*, 976–985. https://doi.org/10.1002/humu.24311.
76. Jiang, L., Yu, H., Ness, S., Mao, P., Guo, F., Tang, J., and Guo, Y. (2022). Comprehensive Analysis of Co-Mutations Identifies Cooperating Mechanisms of Tumorigenesis. Cancers *14*, 415. https://doi.org/10.3390/cancers14020415.
77. Fernandez-Cuesta, L., Peifer, M., Lu, X., Sun, R., Ozretić, L., Seidal, D., Zander, T., Leenders, F., George, J., Müller, C., et al. (2014). Frequent mutations in chromatin-remodelling genes in pulmonary carcinoids. Nat. Commun. *5*, 3518. https://doi.org/10.1038/ncomms4518.
78. Thaxton, C., Goldstein, J., DiStefano, M., Wallace, K., Witmer, P.D., Haendel, M.A., Hamosh, A., Rehm, H.L., and Berg, J.S. (2022). Lumping versus splitting: How to approach defining a disease to enable accurate genomic curation. Cell Genom. *2*, 100131. https://doi.org/10.1016/j.xgen.2022.100131.
79. Scott, A.L., and Odgers, B. (2021). A journey towards answers: Bonnie Odgers Meets Dr. John Graham. Am. J. Med. Genet. *185*, 2627–2629. https://doi.org/10.1002/ajmg.a.62210.
80. Brockers, K., and Schneider, R. (2019). Histone H1, the forgotten histone. Epigenomics *11*, 363–366. https://doi.org/10.2217/epi-2019-0018.
81. Saha, A., and Dalal, Y. (2021). A glitch in the snitch: the role of linker histone H1 in shaping the epigenome in normal and diseased cells. Open Biol. *11*, 210124. https://doi.org/10.1098/rsob.210124.
82. Feierman, E.R., Louzon, S., Prescott, N.A., Biaco, T., Gao, Q., Qiu, Q., Choi, K., Palozola, K.C., Voss, A.J., Mehta, S.D., et al. (2024). Histone variant H2BE enhances chromatin accessibility in neurons to promote synaptic gene expression and long-term memory. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.29.575103.
83. Aicher, J.K., Jewell, P., Vaquero-Garcia, J., Barash, Y., and Bhoj, E.J. (2020). Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. Genet. Med. *22*, 1181–1190. https://doi.org/10.1038/s41436-020-0780-y.
84. Bharathavikru, R., and Hastie, N.D. (2018). Overgrowth syndromes and pediatric cancers: how many roads lead to *IGF2*. Genes Dev. *32*, 993–995. https://doi.org/10.1101/gad.317792.118.
85. Madsen, R.R., Vanhaesebroeck, B., and Semple, R.K. (2018). Cancer-Associated PIK3CA Mutations in Overgrowth Disorders. Trends Mol. Med. *24*, 856–870. https://doi.org/10.1016/j.molmed.2018.08.003.
86. Burkardt, D.D., Tatton-Brown, K., Dobyns, W., and Graham, J.M. (2019). Approach to overgrowth syndromes in the genome era. American J. of Med. Genetics Pt. C. *181*, 483–490. https://doi.org/10.1002/ajmg.c.31757.
87. Yehia, L., Keel, E., and Eng, C. (2020). The Clinical Spectrum of *PTEN* Mutations. Annu. Rev. Med. *71*, 103–116. https://doi.org/10.1146/annurev-med-052218-125823.
88. Connolly, G.K., Harris, R.D., Shumate, C., Rednam, S.P., Canfield, M.A., Plon, S.E., Nguyen, J., Schraw, J.M., and Lupo, P.J. (2024). Pediatric cancer incidence among individuals with overgrowth syndromes and overgrowth features: A population-based assessment in seven million children. Cancer *130*, 467–475. https://doi.org/10.1002/cncr.35041.
89. Brioude, F., Kalish, J.M., Mussa, A., Foster, A.C., Bliek, J., Ferrero, G.B., Boonen, S.E., Cole, T., Baker, R., Bertoletti, M., et al. (2018). Clinical and molecular diagnosis, screening and management of Beckwith–Wiedemann syndrome: an international consensus statement. Nat. Rev. Endocrinol. *14*, 229–249. https://doi.org/10.1038/nrendo.2017.166.
90. Brioude, F., Toutain, A., Giabicani, E., Cottereau, E., Cormier-Daire, V., and Netchine, I. (2019). Overgrowth syndromes — clinical and molecular aspects and tumour risk. Nat. Rev. Endocrinol. *15*, 299–311. https://doi.org/10.1038/s41574-019-0180-z.
91. Klein, S.D., Nisbet, A., and Kalish, J.M. (2023). Overgrowth syndromes, diagnosis and management. Curr. Opin. Pediatr. *35*, 620–630. https://doi.org/10.1097/MOP.0000000000001298.
92. Villani, A., Greer, M.-L.C., Kalish, J.M., Nakagawara, A., Nathanson, K.L., Pajtler, K.W., Pfister, S.M., Walsh, M.F.,

Wasserman, J.D., Zelley, K., and Kratz, C.P. (2017). Recommendations for Cancer Surveillance in Individuals with RASopathies and Other Rare Genetic Conditions with Increased Cancer Risk. Clin. Cancer Res. *23*, e83–e90. https://doi.org/10.1158/1078-0432.CCR-17-0631.

93. Diderich, K., Alanazi, M., and Hoeijmakers, J.H.J. (2011). Premature aging and cancer in nucleotide excision repair-disorders. DNA Repair *10*, 772–780. https://doi.org/10.1016/j.dnarep.2011.04.025.
94. Coppedè, F. (2013). The epidemiology of premature aging and associated comorbidities. CIA *8*, 1023. https://doi.org/10.2147/CIA.S37213.
95. Ahmed, M.S., Ikram, S., Bibi, N., and Mir, A. (2018). Hutchinson–Gilford Progeria Syndrome: A Premature Aging Disease. Mol. Neurobiol. *55*, 4417–4427. https://doi.org/10.1007/s12035-017-0610-7.
96. Foo, M.X.R., Ong, P.F., and Dreesen, O. (2019). Premature aging syndromes: From patients to mechanism. J. Dermatol. Sci. *96*, 58–65. https://doi.org/10.1016/j.jdermsci.2019.10.003.
97. Schnabel, F., Kornak, U., and Wollnik, B. (2021). Premature aging disorders: A clinical and genetic compendium. Clin. Genet. *99*, 3–28. https://doi.org/10.1111/cge.13837.
98. Aguado, J., Gómez-Inclán, C., Leeson, H.C., Lavin, M.F., Shiloh, Y., and Wolvetang, E.J. (2022). The hallmarks of aging in Ataxia-Telangiectasia. Ageing Res. Rev. *79*, 101653. https://doi.org/10.1016/j.arr.2022.101653.
99. Melamed, R.D., Emmett, K.J., Madubata, C., Rzhetsky, A., and Rabadan, R. (2015). Genetic similarity between cancers and comorbid Mendelian diseases identifies candidate driver genes. Nat. Commun. *6*, 7033. https://doi.org/10.1038/ncomms8033.
100. Smith, E.D., Blanco, K., Sajan, S.A., Hunter, J.M., Shinde, D. N., Wayburn, B., Rossi, M., Huang, J., Stevens, C.A., Muss, C., et al. (2019). A retrospective review of multiple findings in diagnostic exome sequencing: halF.A.re distinct and halF.A.re overlapping diagnoses. Genet. Med. *21*, 2199–2207. https://doi.org/10.1038/s41436-019-0477-2.
101. Spedicati, B., Morgan, A., Pianigiani, G., Musante, L., Rubinato, E., Santin, A., Nardone, G.G., Faletra, F., and Girotto, G. (2022). Challenging Occam's Razor: Dual Molecular Diagnoses Explain Entangled Clinical Pictures. Genes *13*, 2023. https://doi.org/10.3390/genes13112023.
102. Saeidian, A.H., March, M.E., Youssefian, L., Watson, D.J., Bhandari, E., Wang, X., Zhao, X., Owen, N.M., Strong, A., Harr, M.H., et al. (2024). Secondary ACMG and non-ACMG genetic findings in a multiethnic cohort of 16,713 pediatric participants. Genet. Med. *26*, 101225. https://doi.org/10.1016/j.gim.2024.101225.
103. Gordon, A.S., Zouk, H., Venner, E., Eng, C.M., Funke, B.H., Amendola, L.M., Carrell, D.S., Chisholm, R.L., Chung, W. K., Denny, J.C., et al. (2020). Frequency of genomic secondary findings among 21,915 eMERGE network participants. Genet. Med. *22*, 1470–1477. https://doi.org/10.1038/s41436-020-0810-9.
104. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
105. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L. D., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. Nature *625*, 92–100. https://doi.org/10.1038/s41586-023-06045-0.
106. 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. https://doi.org/10.1002/humu.22844.
107. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. *84*, 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010.