ESHG

ARTICLE

Check for updates

High molecular diagnostic yields and novel phenotypic expansions involving syndromic anorectal malformations

Raymond Belanger Deloge[1], Xiaonan Zhao[1,2], Pamela N. Luna[1], Chad A. Shaw[1], Jill A. Rosenfeld[1] and Daryl A. Scott[1,3]✉

Evidence suggests that genetic factors contribute to the development of anorectal malformations (ARMs). However, the etiology of the majority of ARMs cases remains unclear. Exome sequencing (ES) may be underutilized in the diagnostic workup of ARMs due to uncertainty regarding its diagnostic yield. In a clinical database of ~17,000 individuals referred for ES, we identified 130 individuals with syndromic ARMs. A definitive or probable diagnosis was made in 45 of these individuals for a diagnostic yield of 34.6% (45/130). The molecular diagnostic yield of individuals who initially met criteria for VACTERL association was lower than those who did not (26.8% vs 44.1%; $p = 0.0437$), suggesting that non-genetic factors may play an important role in this subset of syndromic ARM cases. Within this cohort, we identified two individuals who carried de novo pathogenic frameshift variants in *ADNP*, two individuals who were homozygous for pathogenic variants in *BBS1*, and single individuals who carried pathogenic or likely pathogenic variants in *CREBBP, EP300, FANCC, KDM6A, SETD2*, and *SMARCA4*. The association of these genes with ARMs was supported by previously published cases, and their similarity to known ARM genes as demonstrated using a machine learning algorithm. These data suggest that ES should be considered for all individuals with syndromic ARMs in whom a molecular diagnosis has not been made, and that ARMs represent a low penetrance phenotype associated with Helsmoortel-van der Aa syndrome, Bardet-Biedl syndrome 1, Rubinstein-Taybi syndromes 1 and 2, Fanconi anemia group C, Kabuki syndrome 2, *SETD2*-related disorders, and Coffin-Siris syndrome 4.

European Journal of Human Genetics (2023) 31:296–303; https://doi.org/10.1038/s41431-022-01255-y

INTRODUCTION

Anorectal malformations (ARMs) are common birth defects occurring in approximately 1 in 5000 births [1]. The phenotypic presentation of ARMs ranges from mild (anterior placed anus) to severe (anal atresia). ARMs can occur as isolated defects, but often present with other structural anomalies (syndromic ARMs) such as those that make up VACTERL association (Vertebral, Anorectal, Cardiac, Tracheo-Esophageal, Renal, and Limb anomalies; MIM# 192350) [2]. Surgery is typically required within the first two years of life, and complete restoration of function is not always possible. Hence, individuals with ARMs may have long-term complications including bowel, urological, gynecological, and sexual difficulties.

The relatively high recurrence risk associated with ARMs suggests that genetic factors are likely to contribute to their development. Specifically, within a cohort of 327 individuals with ARMs, Dworschak et al. calculated a 1500-fold increase in recurrence risk for offspring of an affected parent and a 32-fold increase if a sibling was affected [3]. The existence of genetic syndromes in which ARMs are a common phenotype provides additional evidence of the importance of genetic factors in the development of these disorders. For example, ARMs are associated with chromosomal abnormalities and genomic disorders such as Down syndrome (MIM# 190685) and cat-eye syndrome (MIM# 115470), and with single gene disorders such as Townes-Brocks syndrome 1 (MIM# 107480) which is caused by pathogenic variants in *SALL1*. In a recent review, Khanna et al. also suggested the SHH, WNT, and FGF signaling pathways play a major role in the development of ARMs [4].

Despite progress in understanding the genetic etiology of ARMs, the underlying molecular cause of most cases cannot be identified. Exome sequencing (ES) is widely used to identify genetic changes in individuals with multiple congenital anomalies, especially in cases where a clinical diagnosis is not clear. ES has recently been used to identify a molecular diagnosis in individuals with syndromic ARMs [5–7]. However, ES is not always ordered on individuals with syndromic ARMs for whom other genetic tests have failed to identify a cause. This may be due, in part, to uncertainty regarding the efficacy of ES in individuals with syndromic ARMs.

In this study, we analyzed a clinical database of approximately 17,000 ES results to determine the diagnostic yield of ES in individuals with syndromic ARMs. We then used these data to identify eight phenotypic expansions involving ARMs.

SUBJECTS AND METHODS

Database analysis and clinical review

We searched for individuals whose test indications included "anal stenosis", "anal atresia", "imperforate anus", "anterior anus" or similar descriptive entries in a database of ~17,000 individuals referred ES to

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [2]Baylor Genetics, Houston, TX, USA. [3]Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX 77030, USA. ✉email: dscott@bcm.edu

Received: 18 May 2022 Revised: 4 November 2022 Accepted: 22 November 2022
Published online: 6 December 2022

Baylor Genetics. Individuals with an indication of "anal fissure" and those who received a diagnosis on a molecular test other than ES were not included in this study.

Variants reported by Baylor Genetics to be related to clinical phenotypes listed in the indication for ES testing were reanalyzed and classified as pathogenic, likely pathogenic, or variants of uncertain significance (VUS) based on American College of Medical Genetics and Genomics (ACMG) standards for the variant interpretation using the most current data available [8]. Each potential diagnosis was then designated as definitive, probable, or provisional based on previously published criteria set forth by Scott et al. [9]. These criteria take into account the ACMG classification of the variant(s), their inheritance pattern, variant configuration (cis vs. trans), the sex of the proband, and the overlap between the phenotypes listed in the indication and phenotypes previously shown to be associated with disorders caused by the affected gene.

Calculating diagnostic yields

The number of cases with a definitive or probable diagnosis was divided by the total number of syndromic ARM cases to determine the diagnostic yield. We repeated this process for individuals with syndromic ARMs who initially met criteria for VACTERL association by having at least three VACTERL component features [10]. Some have argued that individuals with neurodevelopmental phenotypes should not be classified as having VACTERL association [11]. We have chosen to include these individuals in our VACTERL sub-cohort since ARMs, and other phenotypes associated with VACTERL association, are typically identified in the neonatal period before neurodevelopmental phenotypes such as developmental delay, intellectual disability and autism spectrum disorder become apparent. This is also the time point at which genetic testing is most likely to be initiated.

Literature and database searches

To identify additional cases of syndromic ARMs associated with various genes/genetic disorders, we performed literature searches to identify reports in which a gene's symbol, or the name of its associated genetic disorder(s), was found in association with key words such as "anal", "anus", "anorectal", "anorectal malformations", "anal stenosis", "anal atresia", and/or "imperforate anus".

Machine learning

We have previously developed a machine learning algorithm that integrates knowledge from genome-scale data sources including Gene Ontology (GO), the Mouse Genome Database (MGI), the Protein Interaction Network Analysis (PINA) platform, the GeneAtlas expression distribution, and transcription factor binding and epigenetic histone modifications data from the NIH Roadmap Epigenomics Mapping Consortium to rank genes based on their similarity to a set of training genes known to cause a phenotype of interest [9, 12, 13].

To generate ARM-specific pathogenicity scores for all RefSeq genes, we trained this machine learning algorithm with a set of 26 manually-curated genes that are known to cause ARMs in humans or are the human homologs of genes known to cause ARMs in mice: *CDC45L, CDX1, CRIM1, DACT1, DCHS1, EFNB2, FAM58A* (*CCNQ*), *FREM1, GLI3, INTU, KMT2D, MED12, MID1, MINK1, NOG, PCSK5, PITX2, RECQL4, RIPK4, SALL1, SALL4, SHH, SPECC1L, TBX3, WNT5A, ZIC3* [4, 14].

Cross validation can be used to demonstrate the performance of a machine learning procedure. In these analyses, a subset of the training genes is used to fit the machine learning procedure, which is then used to evaluate the genes that have been excluded or "left-out". This approach enables an objective calculation to characterize the computational procedure using a known input training set while avoiding a circular evaluation that conflates the fitting procedure with performance testing [12, 13].

In our cross-validation analysis, the full set of training genes was randomly broken into two subsets of equal size. The machine learning procedure was trained on each respective set and evaluated on the excluded subset. For each cross validated instance, a genome wide evaluation of all genes was performed including the excluded subset of training genes. The percentiles of the excluded genes were then recorded to assess performance. The procedure was repeated, reciprocally, so that all training genes received cross-validated scores.

These scores were then plotted to characterize the performance of the procedure by tabulating the fraction of training set genes with score percentiles exceeding each cutoff, forming positive receiver operation (ROC) style curves where the effectiveness of the procedure corresponds to the area under the curve and above the diagonal line which represents the result that would be generated by chance alone. These studies generated positive ROC curves based on data from each knowledge source, and the average of the scores across all knowledge sources (Fig. 1A). This demonstrated the ability of our scoring procedure to identify ARMs training genes more efficiently than random chance.

Having validated the algorithm, we generated ARM-specific pathogenicity scores for each gene. This was done by determining the centile rank of each gene as compared with all other RefSeq genes using an omnibus score based on the average fit generated using all knowledge sources. Hence, the ARM-specific pathogenicity score for each RefSeq gene ranges from 0 to 100% with a mean and median of 50%.

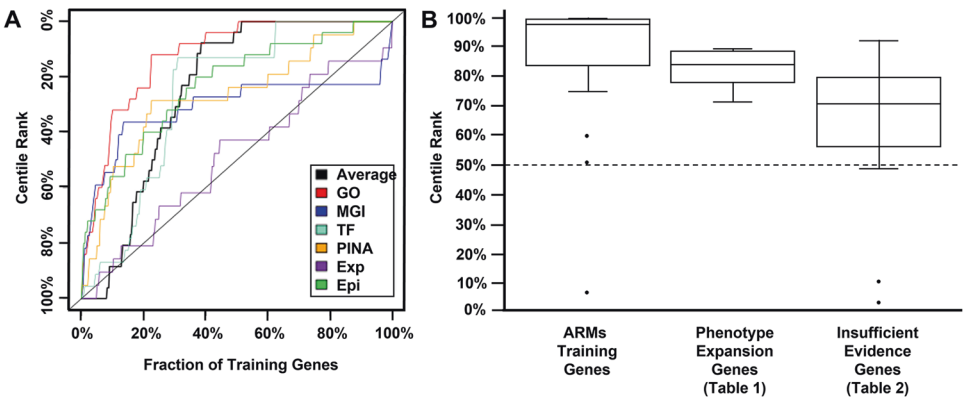

Fig. 1 Machine learning allows all RefSeq genes to be ranked based on their similarity to genes known to cause ARMs. A Receiver operating characteristic (ROC) curves were generated in validation studies of our machine-learning scoring approach. In this figure, colored ROC curves were generated using data from a single knowledge source, and the black ROC curve represents an omnibus score generated using the average score of all knowledge sources. The positive area underneath each curve indicates that our scoring approach identified training set genes known to cause ARMs more efficiently than random chance (diagonal dashed line). **B** After validation, ARMs-specific pathogenicity scores were calculated for all RefSeq genes. Box plots were generated based on the ARM-specific pathogenicity scores of (1) training set genes, (2) genes for which there is sufficient evidence to support a phenotype expansion involving ARMs (Table 1), and (3) genes for which there is currently insufficient evidence to support a phenotype expansion involving ARMs (Table 2). The median pathogenicity scores of the genes listed in Table 1 (83.3%) and Table 2 (70.5%) are lower than median pathogenicity score of the training set (98%) but exceed the median for all RefSeq genes (50%) indicated by the dashed line. This indicates that each of these groups is enriched for genes that are similar to the known ARMs genes in the training set. Epi = epigenetic histone modifications data from the NIH Roadmap Epigenomics Mapping Consortium, Exp = the GeneAtlas expression distribution, GO = Gene Ontology, MGI = the Mouse Genome Database, PINA = the Protein Interaction Network Analysis platform, TF = transcription factor binding data from the NIH Roadmap Epigenomics Mapping Consortium.

Statistical analysis

To compare the diagnostic yields between sub-cohorts, two-tailed Fisher's exact tests were performed using a 2×2 contingency table calculator available through GraphPad Quick Calcs (https://www.graphpad.com/quickcalcs/contingency1/). To compare the diagnostic yields between individual ARM phenotypes, chi-square tests were performed using a 3×2 contingency table calculator available through Social Science Statistics (https://www.socscistatistics.com/tests/chisquare2/default2.aspx). Box plots were generated using the Alcula.com Statistical Calculator: Box Plot program (http://www.alcula.com/calculators/statistics/box-plot/).

RESULTS

Diagnostic yield of ES

From a cohort of ~17,000 individuals referred for clinical exome sequencing, we identified 130 individuals (including Subjects S1-S61) with imperforate anus/anal atresia, anal stenosis, or anteriorly placed anus, who had at least one additional birth defect or neurodevelopmental phenotype (syndromic ARMs). No cases of non-syndromic ARMs were referred for ES. A definitive ($n = 30$; 23.1%) or probable ($n = 15$; 11.5%) diagnosis was made in 45 individuals for a molecular diagnostic yield of 34.6% (45/130). Additionally, a provisional diagnosis was made in 16 individuals. If these were to be included, ES diagnostic yield would increase to 46.9% (61/130). The clinical and molecular data for all subjects in which a definitive, probable, or provisional diagnosis was made are shown in Supplemental Table S1.

Of the 130 individuals with syndromic ARMs, 71 initially met criteria for VACTERL association, defined as having at least three VACTERL component features, and 59 did not. Considering only individuals with a definitive or probable diagnosis, the ES diagnostic yield for the sub-cohort that initially met criteria for VACTERL was 26.8% (19/71). This was significantly lower than the ES diagnostic yield for the sub-cohort that did not initially meet criteria for VACTERL association (44.1%, 26/59; $p = 0.0437$). If individuals with a provisional diagnosis were also included, the ES diagnostic yield for the sub-cohort that initially met criteria for VACTERL (40.8%, 29/71) was still less than that of the sub-cohort that did not initially meet criteria for VACTERL (54.2%, 32/59), but the difference was no longer statistically significant ($p = 0.1586$).

Recurrently altered genes, and genes associated with ARMs

Putatively deleterious variants in several genes were recurrently identified in our cohort. These genes included *ADNP* (S4, S5), *BBS1* (S11, S12), *FGFR3* (S26, S30), *KMT2D* (S16, S31, S37, S38), *LRP2* (S21, S32, S39), *NIPBL* (S31, S44), and *SALL1* (S16, S30).

A subset of individuals in our cohort carried variants in genes that have previously been associated with an increased risk of developing ARMs. These genes included *AMER1, ARID1A, BRCA2, CDH1, CHD7, DHCR7, FAM58A, FGFR3, GRIP1, JAG1, KAT6B, KIF7, KMT2D, MID1, MNX1, MYCN, NIPBL, POR, PQBP1, RAD51, SALL1, SALL4*, and *SPECC1L* (Table S1). The remaining subjects in our cohort only had changes in genes not clearly associated with ARMs. These were considered ARM candidate genes.

To determine which of these candidate genes were most likely to contribute to the development of ARMs, we performed a literature review to identify previously published cases in which these genes were mutated in an individual with ARMs, or in which an individual with ARMs was diagnosed with one of their corresponding genetic syndromes. As an additional means of determining the likelihood that a candidate gene could contribute to the development of ARMs, we used a previously published machine learning algorithm to generate ARM-specific pathogenicity scores for all of the candidate genes [12, 13]. These scores represent the percentile rank of the similarity of each RefSeq gene to a set of 26 genes known to cause ARMs in humans or the human homologs of genes known to cause ARMs in mice [4, 14].

Among candidate genes that carried variants from which a definitive or probable diagnosis was made in our cohort, there is sufficient evidence to suggest a phenotypic expansion involving ARMs for *ADNP*, *BBS1, CREBBP, EP300, FANCC, KDM6A, SETD2*, and *SMARCA4*. The evidence for these associations is summarized in Table 1. In contrast, there is currently insufficient evidence to suggest that the other genes that carried variants from which a definitive or probable diagnosis was made are associated with the development of ARMs. The evidence for these genes is summarized in Table 2.

We then compared the ARM-specific pathogenicity scores of the training genes, the genes for which there was sufficient evidence to suggest a phenotypic expansion involving ARMs (Table 1) and those for which there is currently insufficient evidence to support a phenotypic expansion involving ARMs (Table 2). As expected, the training set had the highest median score (98%), followed by the median scores of the genes for which there was sufficient evidence to support a phenotypic expansion involving ARMs (83.8%), and the median of the genes for which current data were insufficient to support a phenotype expansion (70.5%) (Fig. 1B). The medians of the genes listed in Tables 1 and 2 exceeded the median for all RefSeq genes (50%), indicating that each of these groups are enriched for genes that are similar to the known ARM genes in the training set.

DISCUSSION

ES is widely used to identify genetic changes in individuals with multiple congenital anomalies, and the clinical utility of ES has been clearly demonstrated. ES has specifically been shown to be effective in identifying the molecular etiology of syndromic ARM cases [15]. However, uncertainty about its diagnostic yield may explain, in part, why ES is not universally ordered in individuals with syndromic ARMs. Here, we used data from 130 individuals to estimate the diagnostic yield of ES in syndromic ARM cases and to identify new phenotypic expansions.

High diagnostic yield of clinical ES in syndromic ARM

In this study we found that the molecular diagnostic yield of ES in individuals with syndromic ARMs was high: 34.6% (45/130) when considering only definitive and probable diagnoses and 46.9% (61/130) when provisional diagnoses were included. To our knowledge, this is the first study to specifically report on the molecular diagnostic yield of ES in this patient population.

Interestingly, the molecular diagnostic yield in individuals with syndromic ARMs who initially met criteria for VACTERL association was significantly lower than those who did not meet criteria: 26.8% (19/71) vs. 44.1% (26/59); $p = 0.0437$. In a 2017 study, Meng et al. identified a similar trend, where the ES diagnostic yield for individuals with congenital heart defects (CHD) who initially met criteria for VACTERL association was relatively low compared to other non-VACTERL phenotypes [16]. Recently, Sy et al. also reported that the efficacy rate of ES in individuals with syndromic esophageal atresia/tracheoesophageal fistula (EA/TEF) who initially met criteria for VACTERL association was lower than that of individuals with EA/TEF that did not initially meet criteria (13% versus 18.2%), although this difference did not reach statistical significance [17]. These data suggest that tests designed to identify monogenic etiologies may have lower diagnostic yields in individuals who initially meet the criteria for VACTERL association.

In considering why we see a lower molecular diagnostic yield, we note that epigenetic factors have been described as possibly contributory to VACTERL association, and de novo epivariants have been associated with congenital anomaly syndromes [18, 19]. Non-genetic considerations such as the maternal risk factors of conception via assisted reproductive technologies, pregestational diabetes, and chronic lower obstructive lower pulmonary diseases are also associated with an increased risk of

Table 1. Genes with sufficient evidence to support a phenotypic expansion involving ARMs.

Gene (gnomAD pLI)	Disorder (MIM number)	Subject ID; variant; ACMG interpretation	Number of individuals in our cohort with changes in this gene; level of diagnostic certainty	ARM specific pathogenicity score	Other cases of ARM reported for this gene/ disorder	References
ADNP (pLI = 1)	Helsmoortel-van der Aa syndrome [MIM# 615873]	S4; c.539_542del [NM_015339.5] p.(V180Gfs*17); Pathogenic S5; c.95_96insT [NM_015339.5] p.(K32Nfs*8); Pathogenic	2; Definitive, Definitive	82.9%	Yes/Yes	[24]
BBS1 (pLI = 0)	Bardet-Biedl syndrome 1 [MIM# 209900]	S11; c.[1169T>G];[1169T>G] [NM_024649.5] p.[(M390R)];[(M390R)]; Pathogenic S12; c.[951+1G>T];[951+1G>T] [NM_024649.5] p.[(?)];[(?)]; Pathogenic	2; Definitive, Definitive	84.6%	Yes/Yes	[28, 29]
CREBBP (pLI = 1)	Rubinstein-Taybi syndrome 1 [MIM# 180849]	S18; c.2436_2437insGCTG [NM_004380] p.(P813Afs*20); Pathogenic	1; Definitive	89.2%	Yes/Yes	[34]
EP300 (pLI = 1)	Rubinstein-Taybi syndrome 2 [MIM# 613684]	S22; c.2225del [NM_001429.4] p.(P742Lfs* 34); Pathogenic	1; Definitive	78.1%	Yes/Yes	[33]
FANCC (pLI = 0)	Fanconi anemia complementation group C [MIM# 227645]	S24; c.[1642C>T];[1642C>T] [NM_000136.3] p.[(R548*)];[(R548*)]; Pathogenic/Pathogenic	1; Definitive	89.0%	No/Yes	[35, 36]
KDM6A (pLI = 1)	Kabuki syndrome 2 [MIM# 300867]	S33; c.2441_2442del [NM_021140.4] p.(S814Cfs*20); Pathogenic	1; Definitive	71.1%	Yes/Yes	[42]
SETD2 (pLI = 1)	Luscan-Lumish syndrome [MIM# 616831]	S15; c.5218C>T [NM_014159.7] p.(R1740W); Pathogenic	1; Definitive	77.6%	Yes/Yes	[45]
SMARCA4 (pLI = 1)	Coffin-Siris syndrome 4 [MIM# 614609]	S6; c.2656A>G [NM_001128849.3] p.(M886V); Likely pathogenic	1; Probable	85.7%	No/Yes	[49]

Table 2. Genes with definitive or probable diagnoses that do not currently have sufficient evidence to support a phenotypic expansion involving ARMs.

Gene	Disorder* (MIM number)	Subject ID; variant; ACMG interpretation	Number of individuals in our cohort with changes in this gene; Level of diagnostic certainty	ARM specific pathogenicity score	Other cases of ARM reported for this gene/ disorder in humans
ACTB	Baraitser-Winter syndrome 1 [MIM# 243310]	S1; c.629G>A [NM_001101.5] p.(R210H); Likely pathogenic	1; Probable	76.8%	N/N
ADAT3	Neurodevelopmental disorder with brain abnormalities, poor growth, and dysmorphic facies [MIM# 615286]	S3; c.586del [NM_138422.4] p.(A196Rfs*20); Likely pathogenic	1; Probable	64.2%	N/N
CSNK2A1	Okur-Chung neurodevelopmental syndrome [MIM# 617062]	S19; c.580T>C [NM_001256686.2] p.(S194P); Likely pathogenic	1; Probable	72.8%	N/N
DDX3X	Mental retardation, X-linked 102 [MIM# 300958]	S20; c.1595C>T [NM_001193416.3] p.(T532M); Pathogenic	1; Definitive	48.5%	N /N
EHMT1	Kleefstra syndrome 1 [MIM# 610253]	S21; c.2625del [NM_024757.5] p.(M877*); Pathogenic	1; Definitive	50.0%	N/N
FOXF1	Alveolar capillary dysplasia with misalignment of pulmonary veins [MIM# 265380]	S27; c.668C>A [NM_001451.3] p.(S223*); Pathogenic	1; Definitive	82.6%	N/N
GJB2	Deafness, autosomal recessive, 1A [MIM# 220290]	S28; c.[35del];[35del] [NM_004004.6] p.(G12Vfs*2); Pathogenic/Pathogenic	1; Definitive	86.6%	N/N
KLHL40	Nemaline myopathy 8, autosomal recessive [MIM# 615348]	S34; c.[703G>T]; [703G>T] [NM_152393.4] p.(E235*); Likely pathogenic	1; Probable	82.2%	N/N
MSTO1	Recessive myopathy, mitochondrial, and ataxia [MIM# 617675]	S3; c.[1259del];[1600C>T] [NM_018116.4] p.[(G420Vfs*2)];[(R534C]); Pathogenic/VUS	1; Probable	3.4%	N/N
MYBPC3	Cardiomyopathy, familial hypertrophic 4 [MIM# 115197]	S10; c.1624+2T>C [NM_000256.3] p.(?); Pathogenic	1; Definitive	70.2%	N/N
NEDD4L	Periventricular nodular heterotopia 7 [MIM# 617201]	S43; c.814-6T>A [NM_015277.5] p.(?); Likely pathogenic	1; Probable	88.0%	N /N
PACS2	Epileptic encephalopathy, early infantile, 66 [MIM# 618067]	S46; c.625G>A [NM_001100913.3] p.(E209K); Pathogenic	1; Definitive	10.5%	N/N
PHOX2B	Central hypoventilation syndrome, congenital, with or without Hirschsprung disease [MIM# 209880]	S47; c.385G>A [NM_003924.4] p.(E129K); Likely pathogenic	1; Probable	91.8%	N/N
PIK3CA	[MIM# 171834]	S49; c.1357G>A [NM_006218.4] p.(E453K); Pathogenic	1; Definitive	56.1%	N/N
PTPN11	Noonan syndrome 1 [MIM# 163950]	S51; c.922A>G [NM_002834.5] p.(N308D)	1; Definitive	69.2%	N/N
RET	Hirschsprung disease 1 [MIM# 142623]	S53; c.1947G>A [NM_020975.6] p.(Ser649=); Likely pathogenic	1; Probable	72.4%	N/N
TPO	Thyroid dyshormonogenesis 2A [MIM# 274500]	S59; c.[1472G>A];[1184_1187dup] [NM_000547.6] p.[(R491H)];[(A397Pfs*76)]; VUS/ Pathogenic	1; Probable	74.0%	N/N
TTN	[MIM# 188840]	S21; c.59264del [NM_133378.4] p.(N19755Mfs*2); Pathogenic	1; Definitive	72.7%	N/N
UBE3B	Kaufman oculocerebrofacial syndrome [MIM# 244450]	S14; c.[2624dup];[3014A>C] [NM_130466.4] p.[(N875Kfs*2)];[(Q1005P)]; Pathogenic/VUS	1; Probable	55.6%	N/N
ZEB2	Mowat-Wilson syndrome [MIM# 235730]	S61; c.2886+1G>A [NM_014795.4] p.(?); Pathogenic	1; Definitive	70.5%	N/N
ZNF711	Mental retardation, X-linked 97 [MIM# 300803]	S22; c.930_931del [NM_021998.5] p.(R310Sfs*6); Pathogenic	1; Definitive	58.5%	N/N

N no, *Y* yes.
*If more than one disorder is associated with gene, only the MIM number for the gene is listed.

having a child with VACTERL association [20]. Further research into the genetic, epigenetic, and environmental factors that contribute to the development of VACTERL association is warranted.

Although the data presented here provide clear evidence that ES can be used to identify a molecular diagnosis in a significant percentage of syndromic ARM cases, we recognize the limitation imposed by the retrospective and deidentified nature of this study. A prospective, clinic-based study may provide confirmation of these findings and may also allow comparisons between the yields of ES and other genetic tests—such as chromosome microarray analysis (CMA)—in individuals with syndromic ARMs.

Phenotypic expansions involving ARMs

ADNP. Pathogenic variants in *ADNP* are associated with Helsmoortel-van der Aa syndrome (HVDAS; MIM# 615873). HVDAS is characterized by intellectual disability, motor delay, autism spectrum disorder, hypotonia, dysmorphic facial features, vision complications, congenital heart disease, and gastrointestinal complications such as gastroesophageal reflux and constipation [21–23]. In our cohort, S4 and S5 carried de novo pathogenic frameshift variants in *ADNP*. S4 presented with anal stenosis, while S5 presented with an anteriorly placed anus. One other individual with HVDAS and an ARM has been described [24]. The identification of three individuals with HVDAS and syndromic ARM combined with *ADNP*'s high ARM-specific pathogenicity score (82.9%) lead us to conclude that individuals that carry pathogenic variants in *ADNP* can present with ARMs as part of HVDAS.

BBS1. Bardet-Biedl syndrome (BBS) is a genetically heterogenous disorder [25]. Pathogenic variants in *BBS1* are the cause of Bardet-Biedl syndrome 1 (BBS1; MIM# 209900) and are the most common cause of BBS occurring in 23.4% of all individuals with this disorder [26, 27]. ARMs have been previously described in individuals with BBS. Specifically, Baheci et al described an individual with BBS who had congenital anal atresia, and Hedge et al described a 10-month-old female with BBS and an abnormal site of the anal opening [28, 29]. Unfortunately, molecular diagnoses were not reported for these individuals. Additionally, the Clinical Registry Investigating Bardet-Biedl Syndrome [CRIBBBS] database includes a small percentage of individuals with anomalies of the gastrointestinal tract [30]. In our cohort, we identified 2 individuals—S11 and S12—who carried homozygous pathogenic variants in *BBS1*. These data, combined with the high ARM-specific pathogenicity score of *BBS1* (84.6%) leads us to conclude that ARMs can be a presenting feature of BBS, particularly BBS1 caused by pathogenic variants in *BBS1*.

CREBBP and EP300. Pathogenic variants in *CREBBP* and *EP300* are associated with Rubinstein-Taybi syndrome 1 (RTS1; MIM# 180849) and 2 (RTS2; MIM# 613684), respectively. RTS is characterized by developmental delay, postnatal growth deficiency, microcephaly, broad thumbs and halluces, and dysmorphic facial features [31]. Pathogenic variants in *CREBBP* make up approximately 50–70% of all individuals with RTS, while only 5–8% of individuals with RTS have pathogenic variants in *EP300* [32]. Enomoto et al. reported one individual with a de novo pathogenic deletion in *CREBBP* and anal atresia, and Cohen et al. described two individuals with ARMs who carried *EP300* variants [33, 34]. In our cohort, S18 carried a de novo pathogenic frameshift variant in *CREBBP*, and S22 carried a pathogenic frameshift variant in *EP300*. These data, along with their positive ARM-specific pathogenicity scores (*CREBBP* = 89.2%; *EP300* = 78.1%) suggest that individuals with either RTS1 or RTS2 may present with ARMs.

FANCC. ARMs are a known feature of Fanconi anemia (FA). In our cohort there were five individuals with changes in genes associated with FA (S13, S14 *BRCA2;* S24, *FANCC;* S25, *FANCI;* S52, *RAD51*). Biallelic changes in *BRCA2* and heterozygous variants in *RAD51* have been observed in individuals with ARMs [35, 36]. However, *FANCC* and *FANCI* have not been previously associated with ARMs. The diagnostic certainty for S24 was considered definitive as this individual carried a homozygous pathogenic *FANCC* variant. The ARM-specific pathogenicity score of *FANCC* is 89.0%. Taken together, these data suggest that *FANCC* is associated with ARMs. In contrast, the diagnostic certainty for S25 was considered provisional since this individual only carried only a single pathogenic variant in *FANCI* which is associated with an autosomal recessive form of FA. We also note that the ARM-specific pathogenicity score for *FANCI* was only 41.5%. Hence, there is currently insufficient evidence to support the association between *FANCI* and ARMs.

KDM6A. Variants in *KMT2D*, which are associated with Kabuki syndrome 1 (KABUK1; MIM# 147920), were identified in four Subjects (S16, S31, S37, S38), making it the most commonly affected gene in our cohort. The diagnostic certainty for S37 and S38 was definitive while the certainty for S16 and S31 was provisional. Although ARMs are known to be associated with Kabuki syndrome 1 [37–40], they are not a common feature of Kabuki syndrome 2 (KABUK2; MIM# 300867), which is caused by pathogenic variants in *KDM6A* [41]. S33 carried a de novo pathogenic variant in *KDM6A* and presented with an anteriorly placed anus. One other individual with Kabuki syndrome 2 and an ARM has been reported [42], and *KDM6A* has a positive ARM-specific pathogenicity score (71.1%). Taken together, these data suggest *KDM6A* is associated with the development of ARMs.

LRP2. Variants in *LRP2*, which is associated with Donnai-Barrow syndrome [DBS; MIM# 222448] were identified in three Subjects (S21, S32, S39). However, their diagnoses of Donnai-Barrow were classified as provisional, and all three had variants in at least one other gene included on their ES report. The ARM-specific pathogenicity score for *LRP2* is 64.1%. The presence of three ARMs cases in our cohort suggest the possibility that *LRP2* deficiency contributes to the development of ARMs, however, additional evidence is needed to confirm this association.

SETD2. Pathogenic variants in *SETD2* are associated with Luscan-Lumish syndrome (LLS; MIM# 616831) which is characterized by macrocephaly, intellectual disability, speech delay, low sociability, and behavioral problems. Other more variable features include postnatal overgrowth, obesity, advanced carpal ossification, developmental delay, and seizures [43, 44]. In our cohort, S15 carried a de novo, pathogenic missense variant in *SETD2* (c.5218 C > T [NM_014159.7], p.(R1740W)). This individual presented with an anteriorly placed anus, ventriculomegaly, a cleft palate, congenital heart defects, bilateral 2–3 syndactyly of the hands and feet, a renal cyst, feeding difficulties, respiratory distress, and dysmorphic features. Rabin et al. previously reported two individuals with anteriorly placed anus who carry the same *SETD2* pathogenic variant seen in S15 [45]. They suggested that this variant may have a gain-of-function effect and cause phenotypes that are divergent, and more severe, than those associated with LLS [45]. Specifically, they reported brain anomalies, cleft palate, congenital heart defects, abnormalities of the hands and feet, genitourinary anomalies, respiratory abnormalities, feeding difficulties, and dysmorphic features in individuals with the *SETD2* c.5218 C > T [NM_014159.7], p.(R1740W) variant.

Additionally, Lovrecic et al. reported two individuals with rare 3p21.31 deletions who presented with anal atresia [46]. *SETD2* is located in the genomic region of overlap between the deletions identified in these individuals along with 12 other protein-coding genes, and a portion of *MAP4* (Supplementary Fig. 1) [47]. The features of these genes are summarized in Supplementary Table 2. Among these genes, *SETD2, SMARCC1*, and *DHX30* are loss-of-function intolerant with pLI scores of 1 in gnomAD [48]. However,

only *SETD2* has been observed to be independently associated with ARMs. The ARM-specific pathogenicity score for *SETD2* is 77.6%. Hence, it is possible that pathogenic single nucleotide variants in *SETD2*, including the c.5218 C > T [NM_014159.7], p.(R1740W) variant, and haploinsufficiency of *SETD2* may predispose to the development of ARMs.

SMARCA4. In our cohort, there were 2 individuals who carried variants in genes associated with Coffin-Siris syndrome (S9, *ARID1A*; S6, *SMARCA4*). Pathogenic variants in *ARID1A* have been previously observed in individuals with ARMs, but variants in *SMARCA4* have not [49]. S6 carried a de novo likely pathogenic *SMARCA4* variant, and the ARM-specific pathogenicity score for *SMARCA4* is 85.7%. These data suggest that deleterious variants in *SMARCA4* may lead to the development of ARMs.

Clinical practice recommendations

These data suggest that ES should be considered for all individuals with syndromic ARMs in whom genetic testing has failed to identify a molecular diagnosis. Our data also suggest that additional testing aimed at identifying an independent cause of ARMs may not be warranted in individuals with a diagnosis of Helsmoortel-van der Aa syndrome, Bardet-Biedl syndrome 1, Rubinstein-Taybi syndromes 1 and 2, Fanconi anemia group C, Kabuki syndrome 2, *SETD2*-related disorders, or Coffin-Siris syndrome 4.

DATA AVAILABILITY

The data generated during this study can be found within the published article and its Supplementary Files. All variants reported here have been submitted to the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/; SUB11464682).

REFERENCES

1. Wood RJ, Levitt MA. Anorectal malformations. Clin Colon Rectal Surg. 2018;31:061–70.
2. Wijers CH, van Rooij IA, Marcelis CL, Brunner HG, de Blaauw I, Roeleveld N. Genetic and nongenetic etiology of nonsyndromic anorectal malformations: a systematic review. Birth Defects Res Part C: Embryo Today: Rev. 2014;102:382–400.
3. Dworschak GC, Zwink N, Schmiedeke E, Mortazawi K, Märzheuser S, Reinshagen K, et al. Epidemiologic analysis of families with isolated anorectal malformations suggests high prevalence of autosomal dominant inheritance. Orphanet J Rare Dis. 2017;12:180.
4. Khanna K, Sharma S, Pabalan N, Singh N, Gupta D. A review of genetic factors contributing to the etiopathogenesis of anorectal malformations. Pediatr Surg Int. 2018;34:9–20.
5. Kause F, Zhang R, Ludwig M, Schmiedeke E, Rissmann A, Thiele H, et al. HSPA6: a new autosomal recessive candidate gene for the VATER/VACTERL malformation spectrum. Birth Defects Res. 2019;111:591–7.
6. Kim SY, Ko HS, Kim N, Yim SH, Jung SH, Kim J, et al. A missense mutation in EBF2 was segregated with imperforate anus in a family across three generations. Am J Med Genet Part A. 2018;176:1632–6.
7. Zhu Z, Peng L, Chen G, Jiang W, Shen Z, Du C, et al. Mutations of MYH14 are associated to anorectal malformations with recto-perineal fistulas in a small subset of Chinese population. Clin Genet. 2017;92:503–9.
8. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. ACMG Laboratory Quality Assurance Committee 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. 2015;17:405–24.
9. Scott TM, Campbell IM, Hernandez-Garcia A, Lalani SR, Liu P, Shaw CA, et al. Clinical exome sequencing data reveal high diagnostic yields for congenital diaphragmatic hernia plus (CDH+) and new phenotypic expansions involving CDH. J Med Genet. 2022;59:270–8.
10. Solomon BD. The etiology of VACTERL association: Current knowledge and hypotheses. American journal of medical genetics Part C. Semin Med Genet. 2018;178:440–6.
11. Solomon BD, Bear KA, Kimonis V, de Klein A, Scott DA, Shaw-Smith C, et al. Clinical geneticists' views of VACTERL/VATER association. Am J Med Genet Part A. 2012;158A:3087–100.
12. Campbell IM, Rao M, Arredondo SD, Lalani SR, Xia Z, Kang S-HL, et al. Fusion of large-scale genomic knowledge and frequency data computationally prioritizes variants in epilepsy. PLoS Genet. 2013;9:e1003797.
13. Callaway DA, Campbell IM, Stover SR, Hernandez-Garcia A, Jhangiani SN, Punetha J, et al. Prioritization of candidate genes for congenital diaphragmatic hernia in a critical region on chromosome 4p16 using a machine-learning algorithm. J Pediatr Genet. 2018;7:164–73.
14. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE. Mouse genome database (MGD) 2019. Nucleic Acids Res. 2019;47:D801–6.
15. Van De Putte R, Dworschak GC, Brosens E, Reutter HM, Marcelis CL, Acuna-Hidalgo R, et al. A genetics-first approach revealed monogenic disorders in patients with ARM and VACTERL anomalies. Front Pediatrics. 2020;8:310.
16. Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Vetrini F, et al. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. JAMA Pediatr. 2017;171:e173438.
17. Sy MR, Chauhan J, Prescott K, Imam A, Kraus A, Beleza A, et al. Exome sequencing efficacy and phenotypic expansions involving esophageal atresia/tracheoesophageal fistula plus. Am J Med Genet Part A. 2022;188:3492–504.
18. Lubinsky M. An epigenetic association of malformations, adverse reproductive outcomes, and fetal origins hypothesis related effects. J Assist Reprod Genet. 2018;35:953–64.
19. Barbosa M, Joshi RS, Garg P, Martin-Trujillo A, Patel N, Jadhav B, et al. Identification of rare de novo epigenetic variations in congenital disorders. Nat Commun. 2018;9:2064.
20. van de Putte R, van Rooij I, Haanappel CP, Marcelis CLM, Brunner HG, Addor MC, et al. Maternal risk factors for the VACTERL association: a EUROCAT case-control study. Birth Defects Res. 2020;112:688–98.
21. Breen MS, Garg P, Tang L, Mendonca D, Levy T, Barbosa M, et al. Episignatures Stratifying Helsmoortel-Van Der Aa Syndrome show modest correlation with phenotype. Am J Hum Genet. 2020;107:555–63.
22. Van Dijck A, Vulto-van Silfhout AT, Cappuyns E, van der Werf IM, Mancini GM, Tzschach A, et al. Clinical presentation of a complex neurodevelopmental disorder caused by mutations in ADNP. Biol Psychiatry. 2019;85:287–97.
23. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. Nat Genet. 2014;46:380–4.
24. Gozes I, Patterson MC, Van Dijck A, Kooy RF, Peeden JN, Eichenberger JA, et al. The eight and a half year journey of undiagnosed AD: gene sequencing and funding of advanced genetic testing has led to hope and new beginnings. Front Endocrinol (Lausanne). 2017;8:107.
25. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet. 1999;36:437–46.
26. Niederlova V, Modrak M, Tsyklauri O, Huranova M, Stepanek O. Meta-analysis of genotype-phenotype associations in Bardet-Biedl syndrome uncovers differences among causative genes. Hum Mutat. 2019;40:2068–87.
27. Shamseldin HE, Shaheen R, Ewida N, Bubshait DK, Alkuraya H, Almardawi E, et al. The morbid genome of ciliopathies: an update. Genet Med. 2020;22:1051–60.
28. Bahceci M, Dolek D, Tutuncuoglu P, Gorgel A, Oruk G, Yenen I. A case series of Bardet-Biedl syndrome in a large Turkish family and review of the literature. Eat Weight Disord. 2012;17:e66–9.
29. Hegde HV, Pai RB, Yaliwal VG, Annigeri VM, Halgeri AB, Rao PR. Management of a 10-month-old child with a rare combination of Bardet-Biedl syndrome and ano-rectal malformation undergoing anterior sagittal ano-rectoplasty. J Anesth. 2012;26:132–3.
30. Forsyth RL, Gunay-Aygun M. Bardet-Biedl Syndrome Overview. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al. editors. GeneReviews®. Seattle (WA): University of Washington; 1993.
31. Hennekam RC. Rubinstein-Taybi syndrome. Eur J Hum Genet. 2006;14:981–5.
32. Milani D, Manzoni FM, Pezzani L, Ajmone P, Gervasini C, Menni F, et al. Rubinstein-Taybi syndrome: clinical features, genetic basis, diagnosis, and management. Ital J Pediatr. 2015;41:4.
33. Cohen JL, Schrier Vergano SA, Mazzola S, Strong A, Keena B, McDougall C, et al. EP300-related Rubinstein-Taybi syndrome: Highlighted rare phenotypic findings and a genotype-phenotype meta-analysis of 74 patients. Am J Med Genet A. 2020;182:2926–38.
34. Enomoto Y, Yokoi T, Tsurusaki Y, Murakami H, Tominaga M, Minatogawa M, et al. Divergent variant patterns among 19 patients with Rubinstein-Taybi syndrome uncovered by comprehensive genetic analysis including whole genome sequencing. Clin Genet. 2022;101:335–45.
35. Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. J Med Genet. 2007;44:1–9.
36. Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, et al. A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. Nat Commun. 2015;6:8829.

37. Boniel S, Szymańska K, Śmigiel R, Szczałuba K. Kabuki syndrome-clinical review with molecular aspects. Genes (Basel). 2021;12:468.
38. Siminas S, Baillie CT, Turnock R. Kabuki syndrome and anorectal malformations: implications for diagnosis and treatment. Eur J Pediatr Surg Rep. 2015;3:54–8.
39. Abdel-Salam GM, Afifi HH, Eid MM, el-Badry TH, Kholoussi NM. Anorectal anomalies, diaphragmatic defect, cleft palate, lower lip pits, hypopigmentation and hypogammaglobulinemia A in Kabuki syndrome: a rare combination. Genet Couns. 2008;19:309–17.
40. Baldridge D, Spillmann RC, Wegner DJ, Wambach JA, White FV, Sisco K, et al. Phenotypic expansion of KMT2D-related disorder: beyond Kabuki syndrome. Am J Med Genet A. 2020;182:1053–65.
41. Lederer D, Grisart B, Digilio MC, Benoit V, Crespin M, Ghariani SC, et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. Am J Hum Genet. 2012;90:119–24.
42. Guo HX, Li BW, Hu M, Si SY, Feng K. Novel KDM6A mutation in a Chinese infant with Kabuki syndrome: a case report. World J Clin Cases. 2021;9:10257–64.
43. Luscan A, Laurendeau I, Malan V, Francannet C, Odent S, Giuliano F, et al. Mutations in SETD2 cause a novel overgrowth condition. J Med Genet. 2014;51:512–7.
44. Lumish HS, Wynn J, Devinsky O, Chung WK. Brief report: SETD2 mutation in a child with autism, intellectual disabilities and epilepsy. J Autism Dev Disord. 2015;45:3764–70.
45. Rabin R, Radmanesh A, Glass IA, Dobyns WB, Aldinger KA, Shieh JT, et al. Genotype-phenotype correlation at codon 1740 of SETD2. Am J Med Genet A. 2020;182:2037–48.
46. Lovrecic L, Bertok S, Žerjav, Tanšek M. A new case of an extremely rare 3p21.31 interstitial deletion. Mol Syndromol. 2016;7:93–8.
47. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet. 2009;84:524–33.
48. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
49. Slavotinek A, Lefebvre M, Brehin AC, Thauvin C, Patrier S, Sparks TN, et al. Prenatal presentation of multiple anomalies associated with haploinsufficiency for ARID1A. Eur J Med Genet. 2022;65:104407.

AUTHOR CONTRIBUTIONS

DAS conceived the study. RBD wrote the first draft of the manuscript. CAS and PL were responsible for the machine learning. XZ was responsible for providing updated variant interpretation based on ACMG criteria. JAR obtained and provided clinical and molecular data. RBD and DAS analyzed clinical and molecular data. All authors reviewed, edited, and approved the final draft.

FUNDING

This work was supported, in part, by National Institutes of Health/Eunice Kennedy Shriver National Institute of Child Health and Human Development grant R01HD098458 to DAS.

COMPETING INTERESTS

The Department of Molecular & Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing completed at Baylor Genetics.

ETHICS APPROVAL

This study was approved by the institutional review board of Baylor College of Medicine (protocol H-47546) and was conducted in accordance with the ethical standards of this institution's committee on human research and international standards.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-022-01255-y.

Correspondence and requests for materials should be addressed to Daryl A. Scott.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.