Check for
updates

Experience of the first adult-focussed undiagnosed disease program in Australia (AHA-UDP): solving rare and puzzling genetic disorders is ageless

Mathew Wallis[1,9,10], Simon D. Bodek[1,16*], Jacob Munro[2,3], Haloom Rafehi[2,3], Mark F. Bennett[2,3,4], Zimeng Ye[4], Amy Schneider[4], Fiona Gardiner[4], Giulia Valente[1], Emma Murdoch[1], Eloise Uebergang[1,7], Jacquie Hunter[1], Chloe Stutterd[1,6,7,8], Aamira Huq[1,18], Lucinda Salmon[1,11], Ingrid Scheffer[1,4,5], Dhamidhu Eratne[1,4,12], Stephen Meyn[20], Chun Y. Fong[1], Tom John[1,16,19], Saul Mullen[1,4], Susan M. White[6,7,8], Natasha J. Brown[6,7,8], George McGillivray[6,13], Jesse Chen[14], Chris Richmond[15], Andrew Hughes[1,16], Emma Krzesinski[17], Andrew Fennell[1,17], Brian Chambers[1,16], Renee Santoreneos[1,6], Anna Le Fevre[1,2,6], Michael S. Hildebrand[4], Melanie Bahlo[2,3], John Christodoulou[6,7,8], Martin Delatycki[1,6,7,8] and Samuel F. Berkovic[1,4]

Abstract

Background Significant recent efforts have facilitated increased access to clinical genetics assessment and genomic sequencing for children with rare diseases in many centres, but there remains a service gap for adults. The Austin Health Adult Undiagnosed Disease Program (AHA-UDP) was designed to complement existing UDP programs that focus on paediatric rare diseases and address an area of unmet diagnostic need for adults with undiagnosed rare conditions in Victoria, Australia. It was conducted at a large Victorian hospital to demonstrate the benefits of bringing genomic techniques currently used predominantly in a research setting into hospital clinical practice, and identify the benefits of enrolling adults with undiagnosed rare diseases into a UDP program. The main objectives were to identify the causal mutation for a variety of diseases of individuals and families enrolled, and to discover novel disease genes.

Methods Unsolved patients in whom standard genomic diagnostic techniques such as targeted gene panel, exome-wide next generation sequencing, and/or chromosomal microarray, had already been performed were recruited. Genome sequencing and enhanced genomic analysis from the research setting were applied to aid novel gene discovery.

Joint first authors: Mathew Wallis, Simon D. Bodek, Jacob Munro, Haloom Rafehi.

Joint last authors: Michael S. Hildebrand, Melanie Bahlo, John Christodoulou, Martin Delatycki, Samuel F. Berkovic.

*Correspondence:
Simon D. Bodek
simon.bodek@austin.org.au
Full list of author information is available at the end of the article

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 2 of 14

Results In total, 16/50 (32%) families/cases were solved. One or more candidate variants of uncertain significance were detected in 18/50 (36%) families. No candidate variants were identified in 16/50 (32%) families. Two novel disease genes (*TOP3B*, *PRKACB*) and two novel genotype–phenotype correlations (*NARS,* and *KMT2C* genes) were identified. Three out of eight patients with suspected mosaic tuberous sclerosis complex had their diagnosis confirmed which provided reproductive options for two patients. The utility of confirming diagnoses for patients with mosaic conditions (using high read depth sequencing and ddPCR) was not specifically envisaged at the onset of the project, but the flexibility to offer recruitment and analyses on an as-needed basis proved to be a strength of the AHA-UDP.

Conclusion AHA-UDP demonstrates the utility of a UDP approach applying genome sequencing approaches in diagnosing adults with rare diseases who have had uninformative conventional genetic analysis, informing clinical management, recurrence risk, and recommendations for relatives.

Keywords Rare disease, Genome sequencing, Mosaicism, Genotype, Phenotype, Undiagnosed disease, Tuberous sclerosis, *TOP3B*, *PRKACB*, *NARS*

Introduction

Rare disorders affect 8% of Australians, including almost 500,000 people living in Victoria, an Australian State with a population of over 6.5 million [1]. Most have diseases that begin in childhood, persist through adulthood and are inherently complex. Over a third of adult Australians living with a rare disease endure a diagnostic odyssey lasting years [2], resulting in unnecessary or delayed treatment, poorer health outcomes and substantial burden on our health system [3].

It is estimated that over 80% [4] of rare diseases have a genetic origin. Rapid evolution of high throughput sequencing approaches has dramatically increased the speed and reduced the cost of finding a genetic diagnosis. The most notable of these is exome sequencing (ES), which has become the backbone of clinical genetic testing and for some services, the first-line investigation of choice [5–7]. However, clinical ES is limited in the detection of non-coding and large copy number variants and is constrained to established gene-disease associations and well-characterised phenotypes, yielding a diagnostic rate of only 25–50% [5, 8–11]. While periodic ES-reanalysis (12–36 months) increases yield by a further 10% [12–16], for many adult patients, a diagnosis remains elusive.

Individuals with rare diseases benefit from diagnostic clarity through personalised intervention and surveillance, which improves morbidity and mortality and allows clarification of inheritance patterns and recurrence risks, enabling access to predictive and reproductive genetic testing [17]. The Undiagnosed Diseases Program (UDP) was first established by the US National Institute of Health in 2008 [18] to provide answers for individuals with rare diseases who remained undiagnosed despite exhaustive workup and to provide insight into novel disease mechanisms [19]. This evolved into the US Undiagnosed Diseases Network and subsequently into the Undiagnosed Diseases Network International [20]. This network has led to the diagnosis of thousands of individuals and advanced our knowledge of gene-disease mechanisms and genotype–phenotype correlations through extensive clinician-scientist collaboration and multi-omics integration. The International Rare Disease Research Consortium, set an ambitious goal of diagnostic genomic testing being available for almost all patients with rare diseases by the year 2020. This has not yet been achieved and further work must be done to increase implementation of genomic sequencing and analysis to achieve diagnoses [21].

Historically, UDPs have prioritised increased access to clinical genetics assessment and genomic sequencing for children with rare diseases [22], because they are deemed more likely to benefit from rapid diagnosis and treatment implementation than adults. This lack of awareness and perceived low clinical benefit for adults has limited their access to these projects, for whom a large clinical service testing gap also remains. Notably, despite evidence for utility and a comparable diagnostic yield in targetable phenotypes such as syndromic intellectual disability and multi-system disorders [23–27], very few adult-focussed UDPs exist (see Table 1).

The Austin Health Adult Undiagnosed Diseases Program (AHA-UDP) was a two-year project that aimed to address an area of unmet need for adults in Victoria with presumed orphan or undiagnosed Mendelian disorders and was designed to complement the existing UDP-Victoria, which was targeted at paediatric rare disease [22]. We used a combination of ES re-analysis, genome sequencing (GS), ddPCR, advanced bioinformatic algorithms [41, 42] and variant analysis software [43] and collaborated with local, national, and international researchers to increase diagnostic yield, expand the phenotype of known diseases, and contribute to novel gene discovery.

Table 1 Comparison of overall diagnostic rate with other published undiagnosed disease programs

Published UDP data	Years recruiting	Site	Participants (N)	Phenotyping and Genotyping completed (N)	% Adult Participants	Phenotypes	Prior genetic testing	Main diagnostic strategy	Overall diagnostic rate (%)	Incidental findings (%)	PMID
Austin Health Adult Undiagnosed Diseases Program (AHA-UDP)	2018–2021	Victoria, Australia; monocentric	104(50 probands)	98	100%	Diverse	100%	WES reanalysis+WGS*	32	0	This paper
Program for undiagnosed rare diseases (UD-PrOZA) [28]	2015–2020	Belgium; monocentric	329	329	93.30%	Diverse	59%	WES + proteomics	18.0	7	35606766
The Italian Undiagnosed Rare Diseases Network (IURDN) [29]	2016–2019	Italian; multicentric	110	13	69.20%	Diverse	N/A	WES	53.90	0	32928283
Undiagnosed Disease Network (UDN) [30]	2015–2017	United states of America; multicentric	601	601	41.80%	Diverse	32%	WES	35.0	N/A	30304647
Spain Undiagnosed Disease Program [31]	2015–2018	Spain; multicentric	147	30	25.90%	Diverse	100%	WES	67.00	N/A	30110963
Singapore Undiagnosed Disease Program [32]	2014–2019	Singapore; multicentric	196	196	10.0%	Neurodevelopmental, congenital malformations	N/A	WES or WGS	37.20	N/A	32819910
The Korean undiagnosed diseases program (KUDP) [33]	2017–2018	Korean; multicentric	97	72	5.20%	Diverse	0%	array>NGS>exome	38.90	N/A	30894207
Initiative on Rare and Undiagnosed Disease in Japan (IRUD) [34]	2015–2020	Japan; multicentric	4205	4205	N/A	Diverse	N/A	WES	42.90	N/A	33997444
Victorian Undiagnosed Diseases Program (UDP-VIC) [22]	2016–2018	Victoria, Australia; monocentric	150	150	4%	Diverse	100%	ES > GS > RNA-seq	42.70	N/A	34740920

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 4 of 14

Table 1 (continued)

Published UDP data	Years recruiting	Site	Participants (N)	Phenotyping and Genotyping completed (N)	% Adult Participants	Phenotypes	Prior genetic testing	Main diagnostic strategy	Overall diagnostic rate (%)	Incidental findings (%)	PMID
Finding Of Rare disease Genes FORGE) Canada Project [35]	2011–2013	Canada; multicentric	375	375	Paediatric	Diverse	N/A	WES	54	N/A	24906018
Deciphering Developmental Disorders study [36–39]	2011	UK; multicentric	13,500	13,500	N/A	Primarily neurodevelopmental	N/A	WES+array CGH	30	N/A	25529582
Care4Rare Canadian Consortium [40]	2013–2018	Canada: multicentric	830	830	Adults are recruitable	Diverse	N/A	N/A	32	N/A	36332610
Care4Rare -SOLVE [40]	2018–2021	Canada: multicentric	601	601	Adults are recruitable	Diverse	N/A	N/A	14	N/A	36332610

A comparison of the cohort characteristics and diagnostic rates between the AHA-UDP and other published UDP studies

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 5 of 14

Methods

Study design and population

The AHA-UDP was a two-year project applying emerging gene discovery technologies in adult participants from a range of settings. It had two major objectives: (a) to identify the causal mutations for a variety of diseases in individuals and families enrolled over two years using the latest advances in high throughput genomic technologies; and (b) discover novel gene/phenotype relationships. It was a prospective cohort study of individuals/families identified with rare, previously undiagnosed genetic conditions.

Probands over the age of 16 years known to Austin Health were referred by their treating clinicians to the AHA-UDP team for consideration of inclusion in the program. Individuals/families were discussed at multidisciplinary case conference with subsequent recruitment offered to eligible families. The inclusion criteria were deliberately broad and included the presence, in a family, of at least one adult member who: (a) was a registered patient of Austin Health; (b) had a rare, undiagnosed condition of presumed genetic aetiology; and (c) in whom previous sequencing or microarray analysis was non diagnostic and (d) was willing to sign consent (individual or appropriate legal guardian), provide access to clinical data and to provide a DNA sample. Proposed individuals were prioritised based on uniqueness of their phenotype, the number of affected individuals within the family, individuals from different families having the same apparently novel phenotype, the presence of consanguinity, ineligibility for an alternate testing pathway or project.

Eligible families were seen through the Austin Health Genetics Service and offered enrolment. Written individual informed consent was obtained from the affected and unaffected participants in each family.

The PhenoTips [44] phenotype/genotype database was utilised. Pedigrees were created for each family and phenotypes were recorded using HPO terms. Candidate variants short listed from the bioinformatic pipeline were recorded on the database and then reviewed at the AHA-UDP Multidisciplinary team meeting. DNA samples were extracted from blood (and other tissues for the tuberous sclerosis complex (TSC) cohort) and stored using de-identified sample IDs at the Melbourne Brain Centre.

Ethical approval for the study was granted by the Austin Health Ethics Committee (HREC/18/Austin/41).

Sequencing

The genomic analysis strategy was case-specific and based on detailed clinical phenotyping which provided diagnostic flexibility. A typical approach was the re-analysis of non-diagnostic massively parallel sequencing data (gene panel or ES) followed by genome sequencing (GS) of the proband and additional family members, as appropriate.

Trio GS (Novogene—Illumina NGS) was performed for an affected proband and their unaffected parents where a de novo mutation or autosomal recessive (AR) pattern of inheritance was suspected. Trio/Quad GS was also performed as appropriate where there were multiple affected individuals in a family (sequencing of more than one affected individual increases the diagnostic yield). In other individuals singleton GS was undertaken (where GS of parents and/or other affected individuals was either not practicable or not thought likely to significantly increase yield).

Variants were reported using the GRCh38/hg38 reference genome. SNP and indel variants were called using GATK v4.1.9.0 [45, 46]. Structural variants were called using an ensemble of methods; CNVnator v0.4.1 [47], Manta v1.6.0 [48] and smoove v0.2.5 [49].

Variant filtering was performed using in house pipeline Cavalier and SVPV v1.00 [43], based on annotation by Ensembl VEP v104.3 and subsequent filtering by variant impact, affected genes and allele frequency in gnomAD. Variants were visualised using IGV snapshots and SVPV pdf generated for candidate variants using Cavalier, an in-house method (available from https://github.com/bahlolab/nf-cavalier). Variants were classified based on review of Cavalier presentation slides.

The following tools were used to examine genomic sequencing data for expansions: exSTRa v0.9.0.0 (expanded STR algorithm: detecting expansion with paired-end Illumina sequencing data [42]), ExpansionHunter v3 (algorithm for expanded known STR detection[50]), and ExpansionHunter Denovo v0.9.0 (genome wide novel STR detection algorithm[41]).

Candidate variants in genes without known disease associations were uploaded to the GeneMatcher database with patient consent.

Microarray analysis

Microarrays were performed on a subset of samples using the Illumina Infinium Omni 2.5 SNP microarray platform.

TSC1/2 mosaicism analysis

Targeted gene sequencing of coding regions and splice sites was performed on DNA extracted from blood and other affected tissues (see table [51]). Libraries were prepared and enriched using Sureselect XR target enrichment (Agilent Design ID 0825941). Indexed libraries were pooled and sequenced to a targeted coverage of 700–1000 reads/base (Illumina NextSeq500, 2*75 bp). Seqliner

v0.8 was used to generate aligned reads and call variants against the hg19/GRCh37 human reference genome. PathOS v1.5 was used to annotate and transform variants to standard nomenclature and filter for rare, non-synonymous variants within 20 bp of coding exons. Copy number was analysed using Gaffa 3.0 Targeted. Data was analysed on CLC Genomics Workbench v12.0.2 to screen for low frequency variants present at > 1% VAF. Mosaic variants were then validated with ddPCR as described previously [51].

Report generation and return of results

Following analysis, each case was discussed in a MDT teleconference comprising clinical geneticists, genetic counsellors, clinical scientists and bioinformaticians. Reports were written collaboratively using a cloud office platform before signing off by PI. Participants were notified of their results on completion of the study by letter (referring clinician cc'd) and offer of a follow up appointment, unless a variant of potential clinical significance was found with possible reproductive implications, in which case they were notified sooner. Validation of research results by the referring clinician using a NATA accredited clinical laboratory was recommended.

Mirroring UDP-Vic [22], families were considered solved when a molecular diagnosis in an established disease gene was made, or when a strong candidate for novel gene discovery was identified. For molecular diagnoses in established disease genes, the variant needed to reach criteria for the American College of Medical Genetics and Genomics (ACMG) classification of pathogenic or likely pathogenic (P/LP) [52]. For novel disease-gene variants and new phenotypes in known disease genes, we adopted the same 'strong candidate diagnosis' approach as UDP-Vic [22]. This required the candidate diagnosis meet the following three criteria: (i) phenotypically similar unrelated individuals (matched through data sharing platforms) with a variant in the same gene and population allele frequencies compatible with disease penetrance and inheritance pattern; (ii) in vitro or in vivo functional validation, either planned or underway via local or external research collaborators; and (iii) multidisciplinary agreement that variants in the proposed gene(s) were likely causative for the phenotype(s) and recommended for functional confirmation.

Families were listed as having a variant of uncertain significance (VUS) when they were unsolved, and when: (a) a plausible variant in known gene was found but that variant did not meet Pathogenic/Likely Pathogenic ACMG criteria, or (b) a potential novel disease-gene or known disease gene with a novel phenotype was found, but there was insufficient evidence for the variant to be considered a strong candidate.

Results

Participants

The AHA-UDP commenced in December 2018 and recruited the final family in March 2021. Most cases were referred by a Clinical Geneticist or Neurologist for consideration to the AHA-UDP team. One hundred and fourteen individual participants from 51 families were offered enrolment in the project; one individual withdrew from the study and one individual did not provide a sample, resulting in 104 participants from 50 families who participated in the study.

All probands were adults at time of enrolment, although many had symptom onset in childhood. Four affected relatives under the age of 16 were recruited for potential variant segregation. In the 50 families who participated, the primary phenotype (based on HPO terminology) of the condition under investigation included an abnormality of the nervous system (HP:0000707) (n = 30), a risk of neoplasm (HP:0002664) (n = 12), an abnormality of the genitourinary system (HP:0000119) (n = 4), an abnormality of the cardiovascular system (HP:0001626) (n = 2), an abnormality of the endocrine system (HP:0000818) (n = 1), an abnormality of metabolism/homeostasis (HP:0001939) (n = 1) and an abnormality of the musculoskeletal system (HP:0033127) (n = 1). Eight of the twelve families with a condition associated with a risk of neoplasm were recruited because they were suspected to have mosaic TSC. In those with an abnormality of the nervous system, there were six who were recruited because the proband had intellectual disability believed to be due to an undiagnosed syndrome. Table 1 shows a comparison of the AHA-UDP cohort and diagnostic rates with other published UDP studies. Table 2 shows the characteristics of the AHA-UDP cohort. Table 3 shows a summary of the genetic testing undertaken in the cohort. See supplement 1 for a clinical summary of all families. See supplement 2 for tabulated results and outcomes for all families.

Diagnoses

Ninety-eight individuals from the 50 families were phenotyped and genotyped. All probands had standard clinical genetic testing appropriate to their phenotype prior to recruitment. Singleton GS was completed for 21, trio/quad GS analysis was completed for 11, ES reanalysis was completed for 14, and TSC1/2 deep sequencing (700-1000x) to detect low level mosaicism was undertaken for 8. This was followed by ddPCR validation at Melbourne Brain Centre for those found to have a candidate mosaic variant. Validated mosaic *TSC1* or *TSC2* candidate variants were identified in 4/8 individuals, *PTEN* analysis via the Peter MacCallum Cancer Centre laboratory

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 7 of 14

Table 2 Summary of AHA-UDP cohort

Participants	
Total number of participants*	104 (50 probands)
Sex	55M, 57F
Age at enrolment of participants tested (mean)	48 (SD = 17, 95% CI 44–52)
Total number of clinically affected individuals	73
Total number of families	50
Number of families with a single affected individual	38
Clinical phenotype^	
Abnormality of the nervous system (HP:0000707)	30 (60%)
Intellectual disability believed due to an undiagnosed syndrome	6 (12%)
Risk of neoplasm (HP:0002664)	12 (24%)
Tuberous Sclerosis Complex (TSC)	8 (16%)
Abnormality of the genitourinary system (HP:0000119)	4 (8%)
Abnormality of the endocrine system (HP:0000818)	1 (2%)
Abnormality of the cardiovascular system (HP:0001626)	2 (4%)
Abnormality of the musculoskeletal system (HP:0033127)	1 (2%)
Abnormality of metabolism/homeostasis (HP:0001939)	1 (2%)
Tests prior to enrolment	
Karyotype	3
Chromosome microarray	26
Fragile X	6
Methylation testing	3
Single gene/s testing	17
Gene panel	11
Singleton ES	25
Trio ES	3
ES reanalysis	1

The characteristics of the AHA-UDP cohort. *Some participants were unaffected family members enrolled for trio GS or exclusion mapping. ^The primary phenotype (based on HPO terminology) of the condition under investigation

Table 3 Summary of testing and outcomes

| Testing through UDP (per family)	
High resolution array	5
Single gene/s testing	11
Gene panel re-analysis	1
Singleton whole exome sequencing	1
Exome reanalysis	14
Singleton whole genome sequencing	21
Trio whole exome sequencing	1
Trio whole genome sequencing	8
Quad whole genome sequencing	3
Short tandem repeat analysis	4
ddPCR validation	10
Functional studies	3
Main outcomes	
Solved	16 (32%)
Candidate VUS	18 (36%)
No candidate	16 (32%)

A summary of the tests arranged and main outcomes of the study

was undertaken for one patient, and *PRKACB*, *TOP3B* and *NARS* functional studies were undertaken in three respective families. Chromosomal microarray was undertaken in multiple tissues and individuals in one family. Confirmatory microarray was undertaken in one family. In total, 16/50 families had a diagnosis or strong candidate diagnosis made (32%). One or more candidate variants of uncertain significance were detected in a further 18/50 families (36%). No candidate variants were identified in 16/50 families (32%) (see Fig. 1). Table 3 contains a summary of the tests arranged and main outcomes.

Published case demonstrating a novel method for rapid short tandem repeat detection
NOP56/SCA36 (AH006)
Spinocerebellar ataxias are often caused by expansions in short tandem repeats (STR). Recent methodological advances have made repeat expansion (RE) detection via GS feasible. We published a family demonstrating the use of the ExpansionHunter and exSTRA algorithms to provide a clear and rapid diagnosis of *NOP56*
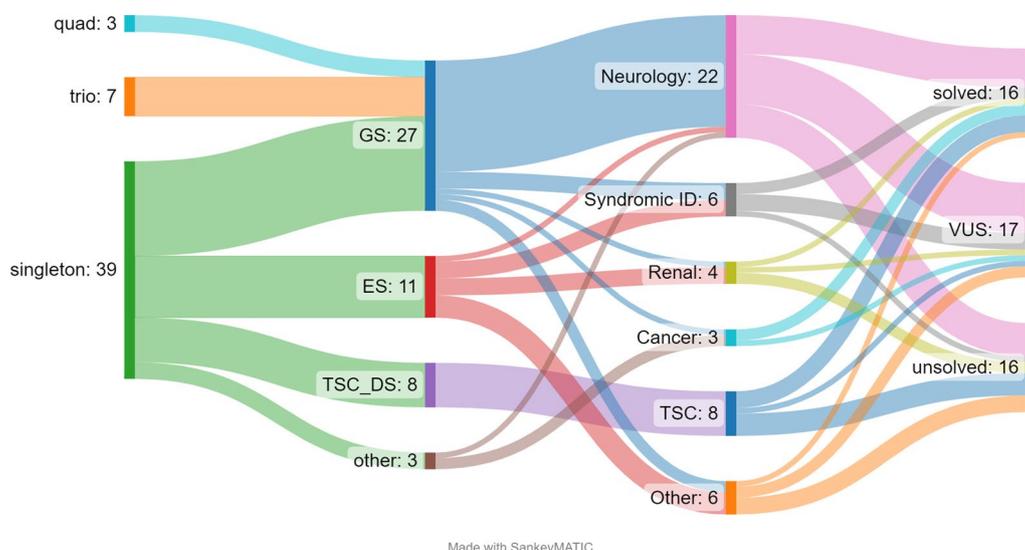
Fig. 1 Sankey diagram of AHA-UDP analysis

SCA36 due to the intronic GGCCTG motif expansion in three affected individuals in a single family with a clinical diagnosis of SCA [53]. A diagnosis was obtained within five days of receiving the sequencing data. A clinical examination confirmed symptoms consistent with SCA36. Molecular diagnostics for SCA36 are not available in Australia- testing was sought via two international providers (Fulgent, Temple City, CA, and MNG Laboratories, Atlanta, GA). Both returned a positive result showing an expansion of a single allele > 70 repeats, however neither diagnostic lab offers exact size estimation methods (such as Southern blot) for this locus.

Novel disease gene discoveries
TOP3B (AH035)
A novel homozygous deletion was found that included *TOP3B* was found on an Infinium Omni2.5 (Illumina) microarray in an individual with bilateral renal cancer [54]. Topoisomerase III beta is one of the least understood members of the topoisomerase family of proteins. Immunoblotting with an antibody against TOP3B showed no detectable protein. Analysis in both the patient and modelled human cells showed that disruption of *TOP3B* causes genome instability with a rise in DNA damage and chromosome bridging (mis-segregation due to formation of Halliday junctions). The molecular defect underlying this pathology is a significant rise in R loop formation (post processing mRNA/DNA dimers). This identified *TOP3B* as a putative cancer gene and supports recent data that R loops are involved in cancer aetiology.

PRKACB (AH033)
We identified a de novo missense variant in *PRKACB* on ES re-analysis in an individual with ID, refractory focal epilepsy, spasticity, periventricular nodular heterotopia, a common atrium / AVSD, polydactyly and several tumours (benign ovarian tumour, liver haemangioma and renal cell carcinoma). *PRKACA* and *PRKACB* encode two catalytic subunits of cAMP dependent protein kinase (PKA). This case was reported along with 3 other *PRKACB* variants and 3 affected individuals with the same *PRKACA* variant (one family) [55]. All affected individuals had the same novel phenotype of atrioventricular septal defects or a common atrium, along with postaxial polydactyly and other features including skeletal anomalies, ectodermal defects of variable severity and cognitive deficits.

Novel disease gene phenotype relationship discoveries
NARS (AH025)
In collaboration with the NARS project at University of Antwerp [56] a missense variant in the NARS protein was identified [c.1025G > A (p.Cys342Tyr)]. Two affected children (LD, DD) had an onset of neuropathy symptoms at age 18 and 19 respectively. Relatively rapid progression of bilateral lower limbs more than upper limbs weakness over 6 months was noted, with subsequent sensory involvement. NCS identified sensori-motor neuropathy with low normal conduction velocities. Their father had small fibre neuropathy symptoms (burning pain in the feet) with onset age 55 whilst the mother was asymptomatic.

KMT2C (AH016)

A 200 kb intragenic heterozygous microdeletion (exons 2–43 of 59) was detected in the *KMT2C* gene on microarray analysis in a proband with mild ID and renal cancer. Loss of function variants in *KMT2C* have been reported in association with a variety of neurodevelopmental phenotypes; autism, a Kleefstra syndrome-like developmental phenotype [57–60], schizophrenia susceptibility, bipolar disorder, and cancer predisposition. Li et al. [59] described a germline mutation in *KMT2C* (MLL3) in four individuals from a multigenerational Chinese family with colorectal cancers and acute myeloid leukemia. Sasaki et al. [60] reported three different variants in *KMT2C* (MLL3) in three closely related family members with familial nasopharyngeal carcinoma using exome sequencing. One variant [p.(Tyr816*)] was predicted to result in a premature stop codon and loss-of-function, the other two variants were missense variants. They hypothesised that inactivating mutations of *MLL3* may be associated with a highly penetrant and previously unknown cancer-predisposition syndrome. No functional studies were performed (PMID: 26014803). Somatic variants in the *KMT2C* gene are often identified in renal cell cancer [61]; however, none of the families reported with germline alterations in this gene have had renal cancer. This variant is classified as pathogenic according to the ACMG criteria[52] (PVS1, PM2, PS2).

There is therefore sufficient evidence to conclude that this microdeletion explains the individual's ID and there is limited evidence to suggest that haploinsufficiency of *KMT2C* is cancer predisposing.

Mosaic TSC

Eight individuals with a definite clinical TSC diagnosis or suspected TSC were enrolled in the AHA-UDP project (see supplement 3). All had prior negative standard of care NGS TSC sequencing. Approximately two thirds of TSC is de novo [62] with mosaicism known to be common in this group and likely accounting for the relatively low sensitivity (~85%) of conventional Sanger or NGS sequencing for individuals with a clinical diagnosis of TSC [62]. Therefore, a strategy of deep sequencing (700–1000* coverage) and ddPCR validation of low level mosaic variants was adopted [51]. In three individuals mosaic low level variant allele fraction (VAF) strong candidate *TSC2* variants were identified and validated. In a further individual a mosaic low level VAF candidate variant was identified but failed ddPCR validation. In another individual a mosaic *TSC1* VUS was identified. The AHA-UDP TSC cohort has been published as part of a collaboration between our groups and a Chinese paediatric TSC cohort [51]. These results are in keeping with previously published cohorts of TSC patients showing a high level of *TSC1* or *TSC2* pathogenic variant mosaicism in individuals with a definite or suspected clinical diagnosis of TSC, in particular for *TSC2* variants, which account for 92% of mosaic patients [63]. This is in contrast to germline TSC where *TSC2* accounts for 69% of cases [64–70]. Of people with a previous NVI (no variant identified) after conventional TSC testing it is estimated that ~50% are mosaic [51, 63].

Discussion

Case selection mirrors paediatric UDP cohorts

This study identified a cause for 32% of individuals/families recruited to an adult UDP program with a further 36% having a genomic finding that did not meet criteria for a definite diagnosis.

We placed an emphasis on recruiting patients with neurological conditions—there are approximately 7211 (OMIM) [71] phenotypes with a known genetic molecular basis; caused by pathogenic variants in approximately 4660 genes. Panel App Australia [72] lists 3935 green genes on their "Mendeliome" gene panel (v1.1894) and 5382 genes in total (including putative disease genes). Of these phenotypes a significant proportion are neurological. For example for intellectual disability syndromes (v0.6063) alone there are 1621 known causative genes (Panel App Australia) [72]. A subset of recruits were judged to have phenotypes likely due to repeat disorders. Molecular diagnosis for dominant ataxia repeat expansions (SCA 1, 2, 3, 6 and 7) are readily available, however no molecular diagnostics are available in Australia for rarer ataxias such as SCA36, or recently discovered ataxias such as CANVAS [73] and SCA27B [74]. The rapid diagnosis of SCA36 highlights the utility for WGS in the diagnosis of rare repeat expansion disorders.

Diagnostic rates are equivalent

Paediatric UDP program diagnostic rates are generally higher than comparable adult cohorts [23] due to differences in case selection (severe neurodevelopmental phenotypes and/or congenital malformations) and low survivability of the most severe disorders into adulthood. Additionally, patients with the more severe genetic conditions may already have been diagnosed in childhood. It is of note that prenatal diagnostic rates are higher again than similar paediatric cohorts [75], likely for similar reasons. Also of note is that rare diseases are a leading cause of death in childhood whereas this is not the case in adults (only ~5% of adults have a rare genetic disease) [4]. An Italian population based study [76] compared the mortality rates across several types of rare genetic diseases in adults and demonstrated improved efficacy of treatment and reduced mortality through genetic

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 10 of 14

diagnosis. The diagnostic rate from the only other adult-focused UDP identified in the literature (UD-PrOZA) was 18%.

In adult UDPs (AHA-UDP and UD-PrOZA) with careful case selection, and controlling for severity (survivability), the diagnostic rates were equivalent to paediatric cohorts; children with undiagnosed rare diseases become adults, and the population prevalence of genetic conditions are unlikely to change significantly over time (although, the proportion with a diagnosis will increase). Lower diagnostic rates could also be explained by a loss of history or medical information over time with deep phenotyping often being key to diagnosis, as well as less capacity for family (trio) sequencing if relatives are deceased or unavailable. Adults are also likely to have other medical phenotypes in addition to those caused by their genetic condition, further complicating diagnosis. Adult-onset multifactorial diseases are often difficult to distinguish from less common monogenic diseases with similar phenotypes (e.g. familial cancer syndromes, maturity onset diabetes of the young (MODY), familial hypercholesterolaemia (FH), cardiomyopathy syndromes, dementia syndromes etc.). This makes case selection more challenging. It is generally considered that a > 10% chance of a monogenic condition is an acceptable threshold for the deployment of genetic testing in adults. This figure comes from cancer genetics [77], is to some extent arbitrary, and is dependent on the cost of genetic testing at a given time. Additionally, trio analysis is less likely to yield additional diagnoses in adults (even when parents are available), because a greater proportion of adult-onset genetic disease follows an autosomal dominant pattern of inheritance and a smaller proportion of adult genetic disease is due to de novo mutations in the proband, rendering the trio de novo approach less powerful.

The AHA-UDP project met its aims of demonstrating the utility of novel analysis techniques (exome re-analysis, genome sequencing, deep sequencing, ddPCR, novel short tandem repeat detection algorithms) in improving diagnostic rates and identifying novel gene disease associations. We identified a diagnosis in 32% of the enrolled families (all of whom had had uninformative previous standard of care clinical diagnostic genetic testing). We identified two novel disease genes and five new genotype–phenotype associations. The detection of a rare repeat expansion in one family highlights the exceptional utility of these novel repeat expansion analyses; particularly in adult disease programs given that many of the diseases caused by repeat expansions have onset in adulthood.

Our diagnostic rates are potentially subject to ascertainment bias as most referrals were from Clinical Geneticists or experienced Neurologists, and cases were selected on the basis of a perceived high prior probability of a monogenic cause. However, genomic testing is still at a premium and so it is likely that in most clinical and research contexts where it is being utilised as a diagnostic test that it would be reserved for cases judged to have a reasonable diagnostic yield.

A study of the illness narratives of participants in the American UDN [78] compared the experiences of adult probands with those of parents of paediatric probands. They found that adult probands hoped a diagnosis would help them resume their former lives. In contrast the parental group understood that their child was unlikely to be cured but hoped a diagnosis would lead to an improvement in medical management. For adult patients, frustration often stemmed from the need to validate symptoms in light of nondiagnostic testing, while for parents it was caused by a concern that evidence crucial to making a diagnosis was being overlooked. That adult patients and parents experience undiagnosed conditions differently, demonstrates the need to further understand and address the psychological impact of undiagnosed conditions in these groups in order to provide optimal support.

Federally funded genomic testing through the Medicare program to identify the cause of childhood syndromes and intellectual disability for children aged ten years and younger became available in Australia on 1 May 2020 [79] due to the evidence that had been gathered from paediatric populations[80]. However, for individuals in Australia with an undiagnosed syndrome who are older than 10 years of age, publicly funded diagnostic testing remains difficult to access because the evidence for utility in adulthood has been underexplored. Yet clear cases for adult-onset genetic testing already exist for example for conditions where disease prevention is possible through screening, risk reducing surgery or chemoprophylaxis and for conditions associated with severe outcomes where predictive testing can enable future planning for at risk individuals.

There can be unique benefit to adults of reproductive age

By way of example, we targeted mosaicism and milder conditions that do not present in childhood. Through targeting mosaicism in patients with suspected TSC and previous uninformative conventional TSC sequencing, we were able to confirm mosaic likely disease causing mutations in 3 out of 8 individuals analysed and a mosaic variant of uncertain significance was identified in one further individual. This is consistent with other recent reports that mosaicism is more common in TSC than had previously been recognised.

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 11 of 14

Adults with undiagnosed conditions are generally underserved

As the utility of UDP programs becomes more established they will leverage novel technologies to increase diagnostic yield. The Austin UDP group will be involved in recruiting patients (initially re-analysing data) to the Undiagnosed Diseases Network-Australia (UDN-Aus) [81]. This is an Australia wide project which began recruiting patients in 2022, leveraging a partnership with the Broad Institute [82] and their *Seqr* platform [83] to enable cloud-based exome/genome analysis by the recruiting clinician (rather than a bioinformatician). The intention is that this model will increase access to the UDP model for clinicians and patients and enable upskilling of clinicians in genomic analysis and a more cost effective/ sustainable model for UDP delivery.

The recruitment and diagnostic framework established by AHA-UDP will continue to be utilised to identify, consent and phenotype adult patients with likely monogenic undiagnosed diseases. Suitable cases will be submitted for consideration of recruitment to the recently activated nationwide Australian Undiagnosed Diseases Network (UDN-Aus).

Conclusion

Here we describe the outcomes from AHA-UDP. We hope our results will encourage other, similarly placed hospitals to also embark on such programs.

The AHA UDP study was conceived to address the absence of a UDP project for adult patients with rare/ monogenic disease in Australia. The study met its aims in terms of demonstrating the utility of an adult UDP framework in achieving comparable diagnostic rates to previously published paediatric UDP studies and in addition a comparable rate of novel genotype phenotype relationship discovery. This has been demonstrated by the results presented in this paper and the already published papers arising out of this study.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13023-024-03297-5.

Additional file 1.

Additional file 2.

Additional file 3.

Author contributions

SDB, MW, JM and HR are joint first co-authors of this research article. The AHA-UDP was designed and initiated by SMe, MW, MBa, SFB, NJB and CYF with assistance and support from TJ, MD and JoC. MW served as initial PI, recruited patients and oversaw sequencing and analyses of half the cases. SDB assumed role of PI halfway through the project, oversaw sequencing and analysis of half the cases, wrote the research variant reports for the families, fed back results and co-wrote this manuscript. MSH, MBa, JoC, MD and SFB all served as co senior authors on the project. MSH oversaw the TSC analysis. MFB, ZY, FG contributed to TSC analysis. MBa oversaw the ES and GS data analysis. MW, SDB, MD, BC, AS, JeC, CS, AH, IS, DE, SMu, SMW, NJB, TJ, LS, GM, CR, RM, SFB, ALF, EK and AF all referred patients who were recruited to the study. EU, JH, LS and GV served as study coordinators. RS contributed to writing up this manuscript. All authors read and approved the final Manuscript.

Data availability

Sequence data is stored at the Walter and Eliza Hall Institute of Medical Research. Supplemental resources are available online. Additional information not subject to ethical restrictions can be obtained from the corresponding author upon request.

Declarations

Conflict of interest

Ingrid E. Scheffer has served on scientific advisory boards for UCB, Eisai,GlaxoSmithKline, BioMarin, Nutricia, Rogcon, Chiesi, Encoded Therapeutics, Knopp Biosciences and Xenon Pharmaceuticals; has received speaker honoraria from GlaxoSmithKline, UCB, BioMarin, Biocodex, Chiesi, Liva Nova and Eisai; has received funding for travel from UCB, Biocodex, GlaxoSmithKline, Biomarin and Eisai; has served as an investigator for Zogenix, Zynerba, Ultragenyx, GW Pharma, UCB, Eisai, Xenon Pharmaceuticals, Anavex Life Sciences, Ovid Therapeutics, Epigenyx, Encoded Therapeutics and Marinus; and has consulted for Zynerba Pharmaceuticals, Atheneum Partners, Ovid Therapeutics, Care Beyond Diagnosis, Epilepsy Consortium and UCB; and is a Nonexecutive Director of Bellberry Ltd. She may accrue future revenue on pending patent WO61/010176 (filed: 2008): Therapeutic Compound; has a patent for SCN1A testing held by Bionomics Inc and licensed to various diagnostic companies; has a patent molecular diagnostic/theranostic target for benign familial PCT/AU2012/001321 (TECH ID:2012-009). JC serves on the Drug Safety Monitoring Board for an Anavex Clinical Trial, and is a member of an endpoint adjudication committee for a Taysha gene therapy trial. The remaining authors declare no competing interests or conflicts of interest.

Ethical approval and consent to participate

Project was reviewed and approved by Austin Health Human Research Ethics Committee. All participants completed an ethics committee approved specific consent form for the study.

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 12 of 14

Consent for publication

The approved consent form for the study signed by all participants included consent to publication for all participants.

Author details

[1]Austin Health Clinical Genetics Service, Austin Health, Melbourne, Australia. [2]The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia. [3]Department of Medical Biology, University of Melbourne, Parkville, Australia. [4]Epilepsy Research Centre, University of Melbourne, Austin Health, Melbourne, Australia. [5]Department of Paediatrics, Austin Health, Melbourne, Australia. [6]Victorian Clinical Genetics Service, Melbourne, Australia. [7]Murdoch Children's Research Institute, Melbourne, Parkville, Australia. [8]Department of Paediatrics, The University of Melbourne, Parkville, Australia. [9]Tasmanian Clinical Genetics Service, Tasmanian Health Service, Hobart, TAS, Australia. [10]School of Medicine and Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia. [11]Genetics Service, Royal Prince Alfred Hospital, Melbourne, Australia. [12]Neuropsychiatry, The Royal Melbourne Hospital, Melbourne, Australia. [13]Genetics Service, Mercy Hospital for Women, Melbourne, Australia. [14]Neurology Service, Austin Health, Melbourne, Australia. [15]Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia. [16]Medicine, Dentistry and Health Science, The University of Melbourne, Parkville, Australia. [17]Monash Health Genetics Clinic, Melbourne, Australia. [18]Genetic Medicine Service, The Royal Melbourne Hospital, Melbourne, Australia. [19]Peter MacCallum Cancer Centre, Melbourne, Australia. [20]Centre for Human Genomics and Precision Medicine, University of Wisconsin-Madison, Madison, WI, USA.

References

1. Australian Government. National Strategic Action Plan for Rare Disease. 2020. Available from: https://www.health.gov.au/sites/default/files/documents/2020/03/national-strategic-action-plan-for-rare-diseases.pdf. Accessed 16 Aug 2022.
2. Molster C, Urwin D, Di Pietro L, Fookes M, Petrie D, van der Laan S, et al. Survey of healthcare experiences of Australian adults living with rare diseases. Orphanet J Rare Dis. 2016;11:30.
3. Dye DE, Brameld KJ, Maxwell S, Goldblatt J, Bower C, Leonard H, et al. The impact of single gene and chromosomal disorders on hospital admissions of children and adolescents: a population-based study. Public Health Genomics. 2011;14:153–61.
4. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020;28:165–73.
5. Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, et al. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. Genet Med. 2016;18:1090–6.
6. Tan TY, Dillon OJ, Stark Z, Schofield D, Alam K, Shrestha R, et al. Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. JAMA Pediatr. 2017;171:855–62.
7. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet Med. 2019;21:2413–21.
8. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014;312:1870–9.
9. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA. 2014;312:1880–7.
10. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18:696–704.
11. Burdick KJ, Cogan JD, Rives LC, Robertson AK, Koziura ME, Brokamp E, et al. Limitations of exome sequencing in detecting rare and undiagnosed diseases. Am J Med Genet A. 2020;182:1400–6.
12. Ewans LJ, Schofield D, Shrestha R, Zhu Y, Gayevskiy V, Ying K, et al. Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. Genet Med. 2018;20:1564–74.
13. Wright CF, McRae JF, Clayton S, Gallone G, Aitken S, FitzGerald TW, et al. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. Genet Med. 2018;20:1216–23.
14. Nambot S, Thevenon J, Kuentz P, Duffourd Y, Tisserant E, Bruel A-L, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. Genet Med. 2018;20:645–54.
15. Costain G, Walker S, Marano M, Veenma D, Snell M, Curtis M, et al. Genome sequencing as a diagnostic test in children with unexplained medical complexity. JAMA Netw Open. 2020;3:e2018109.
16. Stark Z, Schofield D, Martyn M, Rynehart L, Shrestha R, Alam K, et al. Does genomic sequencing early in the diagnostic trajectory make a difference? A follow-up study of clinical outcomes and cost-effectiveness. Genet Med. 2019;21:173–80.
17. Splinter K, Adams DR, Bacino CA, Bellen HJ, Bernstein JA, Cheatle-Jarvela AM, et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. N Engl J Med. 2018;379:2131–9.
18. Gahl WA, Tifft CJ. The NIH undiagnosed diseases program: lessons learned. JAMA. 2011;305:1904–5.
19. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, et al. The NIH undiagnosed diseases program: insights into rare diseases. Genet Med. 2012;14:51–9.
20. Taruscio D, Baynam G, Cederroth H, Groft SC, Klee EW, Kosaki K, et al. The undiagnosed diseases network international: five years and more! Mol Genet Metab. 2020;129:243–54.
21. Lochmüller H, Torrent i Farnell J, Le Cam Y, Jonker AH, Lau LP, Baynam G, et al. The international rare diseases research consortium: policies and guidelines to maximize impact. Eur J Hum Genet. 2017;25:1293–302.
22. Cloney T, Gallacher L, Pais LS, Tan NB, Yeung A, Stark Z, et al. Lessons learnt from multifaceted diagnostic approaches to the first 150 families in Victoria's Undiagnosed Diseases Program. J Med Genet. 2022;59:748–58.
23. Benson KA, White M, Allen NM, Byrne S, Carton R, Comerford E, et al. A comparison of genomic diagnostics in adults and children with epilepsy and comorbid intellectual disability. Eur J Hum Genet. 2020;28:1066–77.
24. Zacher P, Mayer T, Brandhoff F, Bartolomaeus T, Le Duc D, Finzel M, et al. The genetic landscape of intellectual disability and epilepsy in adults and the elderly: a systematic genetic work-up of 150 individuals. Genet Med. 2021;23:1492–7.
25. Walsh M, West K, Taylor JA, Thompson BA, Hopkins A, Sexton A, et al. Real world outcomes and implementation pathways of exome sequencing in an adult genetic department. Genet Med. 2022;24:1536–44.
26. Minardi R, Licchetta L, Baroni MC, Pippucci T, Stipa C, Mostacci B, et al. Whole-exome sequencing in adult patients with developmental and epileptic encephalopathy: it is never too late. Clin Genet. 2020;98:477–85.
27. Eratne D, Schneider A, Lynch E, Martyn M, Velakoulis D, Fahey M, et al. The clinical utility of exome sequencing and extended bioinformatic analyses in adolescents and adults with a broad range of neurological phenotypes: an Australian perspective. J Neurol Sci. 2021;420:117260.
28. Schuermans N, Hemelsoet D, Terryn W, Steyaert S, Van Coster R, Coucke PJ, et al. Shortcutting the diagnostic odyssey: the multidisciplinary Program for Undiagnosed Rare Diseases in adults (UD-PrOZA). Orphanet J Rare Dis. 2022;17:210.
29. Salvatore M, Polizzi A, De Stefano MC, Floridia G, Baldovino S, Roccatello D, et al. Improving diagnosis for rare diseases: the experience of the Italian undiagnosed Rare diseases network. Ital J Pediatr. 2020;46:130.
30. Gahl WA, Wise AL, Ashley EA. The undiagnosed diseases network of the national institutes of health: a national extension. JAMA. 2015;314:1797–8.
31. López-Martín E, Martínez-Delgado B, Bermejo-Sánchez E, Alonso J, The SpainUDP Network, Posada M. SpainUDP: the Spanish undiagnosed rare diseases program. Int J Environ Res Public Health. 2018;15:1746.
32. Bhatia NS, Lim JY, Bonnard C, Kuan J-L, Brett M, Wei H, et al. Singapore undiagnosed disease program: genomic analysis aids diagnosis and clinical management. Arch Dis Child. 2021;106:31–7.
33. Kim SY, Lim BC, Lee JS, Kim WJ, Kim H, Ko JM, et al. The Korean undiagnosed diseases program: lessons from a one-year pilot project. Orphanet J Rare Dis. 2019;14:68.

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 13 of 14

34. Adachi T, Kawamura K, Furusawa Y, Nishizaki Y, Imanishi N, Umehara S, et al. Japan's initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey. Eur J Hum Genet. 2017;25:1025–8.

35. Beaulieu CL, Majewski J, Schwartzentruber J, Samuels ME, Fernandez BA, Bernier FP, et al. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am J Hum Genet. 2014;94:809–17.

36. The Deciphering Developmental Disorders (DDD) study—ProQuest. [cited 2022 Nov 29]. Available from: https://www.proquest.com/openview/c74f3e3fc754edc82de4fab719b77813/1?pq-origsite=gscholar&cbl=34809.

37. Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet. 2015;385:1305–14.

38. Kaplanis J, Samocha KE, Wiel L, Zhang Z, Arvai KJ, Eberhardt RY, et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature. 2020;586:757–62.

39. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.

40. Boycott KM, Hartley T, Kernohan KD, Dyment DA, Howley H, Innes AM, et al. Care4Rare Canada: outcomes from a decade of network science for rare disease gene discovery. Am J Hum Genet. 2022;109:1947–59.

41. Dolzhenko E, Bennett MF, Richmond PA, Trost B, Chen S, van Vugt JJFA, et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol. 2020;21:102.

42. Tankard RM, Bennett MF, Degorski P, Delatycki MB, Lockhart PJ, Bahlo M. Detecting expansions of tandem repeats in cohorts sequenced with short-read sequencing data. Am J Hum Genet. 2018;103:858–73.

43. Munro JE, Dunwoodie SL, Giannoulatou E. SVPV: a structural variant prediction viewer for paired-end sequencing datasets. Bioinformatics. 2017;33:2032–3.

44. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat. 2013;34:1057–65.

45. GATK. [cited 2022 Aug 25]. Available from: https://gatk.broadinstitute.org/hc/en-us.

46. Van der Auwera GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. 1st ed. Sebastopol: O'Reilly Media; 2020.

47. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21:974–84.

48. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220–2.

49. Pedersen B. Smoove: structural variant calling and genotyping with existing tools, but, smoothly. [cited 2022 Aug 25]. Available from: https://hpc.nih.gov/apps/smoove.html.

50. Dolzhenko E, Deshpande V, Schlesinger F, Krusche P, Petrovski R, Chen S, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics. 2019;35:4754–6.

51. Ye Z, Lin S, Zhao X, Bennet MF, Brown NJ, Wallis M, et al. Mosaicism in tuberous sclerosis complex—lowering the threshold for clinical reporting. Hum Mutat. 2022;43(12):1956–69.

52. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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.

53. Rafehi H, Szmulewicz DJ, Pope K, Wallis M, Christodoulou J, White SM, et al. Rapid diagnosis of spinocerebellar ataxia 36 in a three-generation family using short-read whole-genome sequencing data. Mov Disord. 2020;35:1675–9.

54. Zhang T, Wallis M, Petrovic V, Challis J, Kalitsis P, Hudson DF. Loss of TOP3B leads to increased R-loop formation and genome instability. Open Biol. 2019;9:190222.

55. Palencia-Campos A, Aoto PC, Machal EMF, Rivera-Barahona A, Soto-Bielicka P, Bertinetti D, et al. Germline and mosaic variants in PRKACA and PRKACB cause a multiple congenital malformation syndrome. Am J Hum Genet. 2020;107:977–88.

56. Beijer D, Marte S, Li JC, De Ridder W, Chen JZ, Tadenev ALD, et al. Dominant NARS1 mutations causing axonal Charcot–Marie–Tooth disease expand NARS1-associated diseases. Brain Commun. 2024;6(2):fcae070. https://doi.org/10.1093/braincomms/fcae070.

57. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LELM, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am J Hum Genet. 2012;91:73–82.

58. Koemans TS, Kleefstra T, Chubak MC, Stone MH, Reijnders MRF, de Munnik S, et al. Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLoS Genet. 2017;13:e1006864.

59. Li W-D, Li Q-R, Xu S-N, Wei F-J, Ye Z-J, Cheng J-K, et al. Exome sequencing identifies an MLL3 gene germ line mutation in a pedigree of colorectal cancer and acute myeloid leukemia. Blood. 2013;121:1478–9.

60. Sasaki MM, Skol AD, Bao R, Rhodes LV, Chambers R, Vokes EE, et al. Integrated genomic analysis suggests MLL3 is a novel candidate susceptibility gene for familial nasopharyngeal carcinoma. Cancer Epidemiol Biomark Prev. 2015;24:1222–8.

61. Yan L, Zhang Y, Ding B, Zhou H, Yao W, Xu H. Genetic alteration of histone lysine methyltransferases and their significance in renal cell carcinoma. PeerJ. 2019;7:e6396.

62. Treichel AM, Hamieh L, Nathan NR, Tyburczy ME, Wang J, Oyerinde O, et al. Phenotypic distinctions between mosaic forms of tuberous sclerosis complex. Genet Med. 2019;21:2594–604.

63. Giannikou K, Lasseter KD, Grevelink JM, Tyburczy ME, Dies KA, Zhu Z, et al. Low-level mosaicism in tuberous sclerosis complex: prevalence, clinical features, and risk of disease transmission. Genet Med. 2019;21:2639–43.

64. Northrup H, Koenig MK, Pearson DA, Au KS. Tuberous sclerosis complex. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LJ, et al., editors. GeneReviews®. Seattle: University of Washington, Seattle; 1993 [cited 2022 Nov 22]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1220/.

65. Jones AC, Shyamsundar MM, Thomas MW, Maynard J, Idziaszczyk S, Tomkins S, et al. Comprehensive mutation analysis of TSC1 and TSC2-and phenotypic correlations in 150 families with tuberous sclerosis. Am J Hum Genet. 1999;64:1305–15.

66. Dabora SL, Jozwiak S, Franz DN, Roberts PS, Nieto A, Chung J, et al. Mutational analysis in a cohort of 224 tuberous sclerosis patients indicates increased severity of TSC2, compared with TSC1, disease in multiple organs. Am J Hum Genet. 2001;68:64–80.

67. Au K-S, Williams AT, Gambello MJ, Northrup H. Molecular genetic basis of tuberous sclerosis complex: from bench to bedside. J Child Neurol. 2004;19:699–709.

68. Sancak O, Nellist M, Goedbloed M, Elfferich P, Wouters C, Maat-Kievit A, et al. Mutational analysis of the TSC1 and TSC2 genes in a diagnostic setting: genotype–phenotype correlations and comparison of diagnostic DNA techniques in Tuberous Sclerosis Complex. Eur J Hum Genet. 2005;13:731–41.

69. Au KS, Williams AT, Roach ES, Batchelor L, Sparagana SP, Delgado MR, et al. Genotype/phenotype correlation in 325 individuals referred for a diagnosis of tuberous sclerosis complex in the United States. Genet Med. 2007;9:88–100.

70. Tyburczy ME, Dies KA, Glass J, Camposano S, Chekaluk Y, Thorner AR, et al. Mosaic and intronic mutations in TSC1/TSC2 explain the majority of TSC patients with no mutation identified by conventional testing. PLoS Genet. 2015;11:e1005637.

71. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43:D789-798.

72. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019;51:1560–5.

73. Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NLM, Pope K, Smith KR, et al. Bioinformatics-based identification of expanded repeats: A nonreference intronic pentamer expansion in RFC1 causes CANVAS. Am J Hum Genet. 2019;105:151–65.

Wallis *et al. Orphanet Journal of Rare Diseases*   (2024) 19:288

Page 14 of 14

74. Rafehi H, Read J, Szmulewicz DJ, Davies KC, Snell P, Fearnley LG, et al. An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14. Am J Hum Genet. 2023;110:105–19.

75. Mellis R, Oprych K, Scotchman E, Hill M, Chitty LS. Diagnostic yield of exome sequencing for prenatal diagnosis of fetal structural anomalies: a systematic review and meta-analysis. Prenat Diagn. 2022;42:662–85.

76. Gorini F, Coi A, Mezzasalma L, Baldacci S, Pierini A, Santoro M. Survival of patients with rare diseases: a population-based study in Tuscany (Italy). Orphanet J Rare Dis. 2021;16:275.

77. Carver T, Hartley S, Lee A, Cunningham AP, Archer S, de Villiers CB, et al. CanRisk Tool—A web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants. Cancer Epidemiol Biomark Prev. 2021;30:469–73.

78. Spillmann RC, McConkie-Rosell A, Pena L, Jiang Y-H, Adams CJ, Adams DR, et al. A window into living with an undiagnosed disease: illness narratives from the Undiagnosed Diseases Network. Orphanet J Rare Dis. 2017;12:71.

79. Sachdev R, Field M, Baynam GS, Beilby J, Berarducci M, Berman Y, et al. Paediatric genomic testing: navigating medicare rebatable genomic testing. J Paediatr Child Health. 2021;57:477–83.

80. Schofield D, Rynehart L, Shresthra R, White SM, Stark Z. Long-term economic impacts of exome sequencing for suspected monogenic disorders: diagnosis, management, and reproductive outcomes. Genet Med. 2019;21:2586–93.

81. The Australian Undiagnosed Diseases Network. Australian Genomics. [cited 2022 Nov 23]. Available from: https://www.australiangenomics.org.au/research/the-australian-undiagnosed-diseases-network/.

82. Broad Institute [Internet]. Broad Institute. [cited 2022 Aug 25]. Available from: https://www.broadinstitute.org/home.

83. Pais LS, Snow H, Weisburd B, Zhang S, Baxter SM, DiTroia S, et al. seqr: a web-based analysis and collaboration tool for rare disease genomics. Hum Mutat. 2022;43:698–707.

Publisher's Note