



# A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis

Tiong Yang Tan <sup>1,2,3</sup> · Sebastian Lunke <sup>1,2,4</sup> · Belinda Chong<sup>1,2</sup> · Dean Phelan<sup>1,2</sup> · Miriam Fanjul-Fernandez<sup>1,2,3</sup> · Justine E. Marum<sup>1,2</sup> · Vanessa Siva Kumar<sup>1,2</sup> · Zornitza Stark<sup>1,2,3</sup> · Alison Yeung<sup>1,2,3</sup> · Natasha J. Brown<sup>1,2,3,5</sup> · Chloe Stutterd<sup>1,2,3,5</sup> · Martin B. Delatycki<sup>1,2,3,5</sup> · Simon Sadedin<sup>1,2</sup> · Melissa Martyn<sup>2,3,6</sup> · Ilias Goranitis<sup>2,3</sup> · Natalie Thorne<sup>2,3,6</sup> · Clara L. Gaff<sup>2,3,6</sup> · Susan M. White <sup>1,2,3</sup>

Received: 5 January 2019 / Revised: 24 April 2019 / Accepted: 2 July 2019 / Published online: 18 July 2019  
© The Author(s), under exclusive licence to European Society of Human Genetics 2019

## Abstract

Diagnostic exome sequencing (ES) can be performed on the proband only (singleton; sES) or with additional samples, often including both biological parents with the proband (trio; tES). In this study we sought to compare the efficiencies of exome sequencing (ES) by trio (tES) versus singleton (sES) approach, determine costs, and identify factors to consider when deciding on optimal implementation strategies for the diagnosis of monogenic disorders. We undertook ES in 30 trios and analysed each proband's sES and tES data in parallel. Two teams were randomly allocated to either sES or tES analysis for each case and blinded to each other's work. Each task was timed and cost analyses were based on time taken and diagnostic yield. We modelled three scenarios to determine the factors to consider in the implementation of tES. sES diagnosed 11/30 (36.7%) cases and tES identified one additional diagnosis (12/30 (40.0%)). tES obviated the need for Sanger segregation, reduced the number of variants for curation, and had lower cost-per-diagnosis when considering analysis alone. When sequencing costs were included, tES nearly doubled the cost of sES. Reflexing to tES in those who remain undiagnosed after sES was cost-saving over tES in all as first-line. This approach requires a large differential in diagnostic yield between sES and tES for maximal benefit given current sequencing costs. tES may be preferable when scaling up laboratory throughput due to efficiency gains and opportunity cost considerations. Our findings are relevant to clinicians, laboratories and health services considering tES over sES.

Manuscript editing: all authors contributed to manuscript review and editing

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41431-019-0471-9>) contains supplementary material, which is available to authorized users.

✉ Tiong Yang Tan  
tiong.tan@vcgs.org.au

✉ Susan M. White  
sue.white@vcgs.org.au

- 1 Victorian Clinical Genetics Services, Melbourne, Australia
- 2 Murdoch Children's Research Institute, Melbourne, Australia
- 3 Department of Paediatrics, University of Melbourne, Melbourne, Australia
- 4 Department of Pathology, University of Melbourne, Melbourne, Australia
- 5 Austin Health Department of Clinical Genetics, Melbourne, Australia
- 6 Melbourne Genomics Health Alliance, Parkville, Australia

## Introduction

Genomic sequencing is now firmly established in clinical practice for the diagnosis of individuals with monogenic disorders. The choice of genomic test, whether panel, exome or genome sequencing, is usually determined by specificity of clinical presentation, funding, and local availability. Exome sequencing (ES) aims to capture and determine the sequence of all protein-coding exons representing approximately 2% of the genome, and can be performed on the proband only (singleton; sES) or with additional samples, often including both biological parents with the proband (trio; tES).

The diagnostic yield of sES ranges from 20 to over 50% [1–18], with reanalysis adding an additional 10% over a period of 18–36 months [19–24]. Candidate causal variants identified by sES usually require parental segregation with targeted Sanger sequencing to confirm de novo status or biallelic inheritance, information that is inherently obtained

from tES. Studies comparing diagnostic yield of sES against tES have generally found that tES provides an incremental gain of 10–15%, mostly due to the ease of identifying de novo variants (Table 1) [3, 6, 7, 15, 17, 18, 25]. (and F. Alkuraya, personal communication). However, these studies have compared different individuals sequenced either by sES or tES within a particular setting with the phenotype likely influencing the ordering physician's test choice, which potentially introduces selection bias.

It has been asserted that the advantages of tES over sES include a lower analysis cost, sensitivity to de novo and compound heterozygous variants and newly published genes, whilst avoiding the need for parental segregation thus improving turnaround time [3, 6]. These benefits may be especially manifest when the presenting phenotype is complex, rather than single-system as the analytical benefit of tES is less informative when the phenotype is well-defined [6]. What remains unclear is the optimal circumstances in which to select tES over sES, and what factors are important in that decision-making process to maximise yield and optimise system costs. In this study we sought to examine some of these factors.

A head-to-head comparison of sES and tES in the same patient is the least biased method of assessing test performance. No prior studies have used this methodology. In this study we used a novel approach of analysing each case as a singleton and trio to quantitate the efficiencies and determine costs of tES compared to sES. To develop a better understanding of the factors that may inform optimum test selection, we model hypothetical scenarios varying the a priori likelihood of a monogenic diagnosis, incremental diagnostic yield from tES, and sequencing models.

## Methods

Individuals were prospectively recruited by clinical geneticists after referral from general and subspecialist physicians or paediatricians from February 2016 to September 2017 as part of the Complex Care Flagship of the Melbourne Genomics Health Alliance Demonstration Project [26]. The Complex Care flagship aimed to recruit individuals suspected of having a monogenic disorder with complex medical needs from three tertiary centres in Melbourne, Australia (Royal Children's Hospital, Austin Hospital and Monash Health). Inclusion and exclusion criteria are detailed in Table 2. Cases were presented by the referring clinical geneticist at a weekly meeting in which a panel of at least two other clinical geneticists decided if they satisfied the inclusion criteria. We decided to include patients at the start of their diagnostic trajectory so excluded those who had been investigated for more than 2 years. Recruited individuals had microarrays (without a causative diagnosis) and no prior single-gene or panel sequencing. Study data were collected and managed using REDCap [27] electronic data capture tools hosted at Murdoch Children's Research Institute. Phenotypic data of all recruited individuals were recorded in PhenoTips [28].

## Sequencing

### Exome sequencing, variant detection and filtering

We performed exome sequencing, variant detection and filtering as described previously [9, 29]. The mean coverage

**Table 1** Studies reporting diagnostic rates of tES across different clinical indications, some providing within-cohort comparisons to sES yield

Ref	Setting	N	Clinical indication	Total dx rate	Singleton dx rate	Trio dx rate
Lee et al. [6]	UCLA	814	Undiagnosed, suspected genetic disorder	213/814 (26%)	74/338 (22%)	127/410 (31%)
Soden et al. (2014)	Children's Mercy Kansas (±rapid WGS)	100	Neurodevelopmental	45/100 (45%)	N/A	N/A
Farwell et al. [3]	Ambry diagnostic lab; molecular geneticist + genetic counsellor curation	500	Multiple; mostly paediatric neurological, multiple congenital anomalies	152/500 (30.4%)	14/68 (20.6%)	82/220 (37.3%)
Reterrer et al. (2015)	GeneDx diagnostic lab	3040	Multiple; paediatric and adult	875/3040 (28.8%)	128/542 (23.6%)	647/2088 (31%)
Zhu et al. (2015)	Duke University MC	119	'Severe' genetic disorders	29/119 (24.4%)	N/A	N/A
Monroe et al. (2015)	Utrecht MC	17	Intellectual disability	5/17 (29.4%)	N/A	N/A
Monies et al. [15]	Saudi Arabia NGS reference lab	1013	Broad range; panels, exome sequencing	149/347 exomes (43%)	138/321 (43%)	9/17 <sup>a</sup> (53%)

<sup>a</sup>personal communication, F Alkuraya

**Table 2** Inclusion and exclusion criteria and demographics of recruited individuals

Inclusion criteria	
Individuals of any age considered likely to have a monogenic disorder	
Complexity of medical care (minimum two out of three)	
<ul style="list-style-type: none"> <li>• Multiple organ systems involved and managed by 3 or more specialty units</li> <li>• Severe condition with high morbidity and mortality</li> <li>• Severe limitations on function and activities of daily living</li> </ul>	
Appropriate for trio sequencing	
<ul style="list-style-type: none"> <li>• Both biological parents available</li> <li>• Non-specific phenotype with a broad differential diagnosis</li> <li>• Possibility of blended phenotype</li> </ul>	
Exclusion criteria	
Phenotypes which are typically multi-factorial or environmental-related	
Unrecognised novel phenotypes with no evidence-based gene list (e.g., orphan syndromes) better suited to gene discovery projects	
Prior sequencing of candidate genes (excepting genes in which variant types are poorly detected by exome sequencing, such as copy number variants, nucleotide repeat disorders, methylation testing or mitochondrially encoded variants)	
Individuals who have been evaluated by specialists for diagnostic purposes for more than 2 years	
Demographics	
Proposed for recruitment	37
Excluded	7
Reasons for exclusion	Investigated > 2 years $n = 2$ Non-genetic more likely $n = 1$ Failed complexity criteria $n = 3$ Microarray finding explained phenotype $n = 1$
Males	14 (46.7%)
Females	16 (53.3%)
Median age at recruitment (range)	21.5 months (4–650 months)
Phenotypic category	(see Supplementary Table 1 for details)
Neurodevelopmental without dysmorphism	11 (36.7%)
Neurodevelopmental with dysmorphism	9 (30%)
Multiple congenital anomalies	4 (13.3%)
Failure to thrive with neurodevelopmental issues	4 (13.3%)
Failure to thrive with dysmorphism	2 (6.7%)

obtained was 128.70 (88.8–182.9). Variants in NCBI Reference Sequence Database (RefSeq) listed genes associated with Mendelian disease prior to the end of 2017 (3784 genes; Supplementary Table 2) were available for initial analysis. Variants were assessed using the Melbourne Genomics variant curation database, a modification of the Leiden Open Variation Database [30]. Variants were prioritised based on the phenotype-driven gene lists for each participant and on a bioinformatically computed score reflecting likelihood of functional effect (Variant Prioritisation Index) [29]. We only assessed variants relevant to the participant's phenotype. Criteria for classification were based on the American College of Medical Genetics and Genomics standards for interpretation of sequence variants [31].

### Trio randomisation

All participants underwent trio exome sequencing (proband and both parents), but singleton and trio data were analysed in parallel by two independent teams. Data were analysed in separate instances of the curation database so that teams were blinded to each other's analyses. Each team comprised a clinician and scientist curator who were individually randomised to work on either the tES or sES analysis. The recruiting clinical geneticist, who was inherently familiar with the proband's phenotype, was randomly allocated to sES or tES analysis to avoid potential bias. Random allocation also removed the bias potentially arising from differing levels of curation experience. All tasks of the curation process including the multidisciplinary team (MDT) meetings, which

were conducted separately for sES and tES, were timed using a commercial web-based application MyHours (myhours.com). Clinicians recorded the time taken to prioritise the variants identified within the phenotype-driven gene lists, and scientist curators recorded the time taken to curate each variant. Variant classifications were reviewed in MDT meetings, including clinical geneticists, medical subspecialists, genetic counsellors, molecular geneticists, and bioinformaticians.

Time to prioritise variants (clinician and scientist), variant curation time, MDT discussion time and numbers of variants prioritised for analysis were compared for sES and tES. Mean and standard deviation were determined in order to have a measure of central tendency that best reflects the spread of the data. Differences were investigated using two-sided paired and unpaired t-tests. P values are reported for results where  $p < 0.05$  was considered statistically significant. Analyses were performed using Stata 15.1 (Stata, College Station, TX, USA).

### Cost analysis of hypothetical implementation scenarios

The unit cost of sequencing only (without analysis) was AUD\$1200, while that of segregation Sanger sequencing (including analysis) was AUD\$250 per variant per person. We used the top-tier standard hourly rates for clinician and scientist curator salaries, including 20% on-costs, to calculate the cost of variant analysis. MDT costs were calculated based on a clinician and scientist curator as quorum. We evaluated cost-per-diagnosis for implementation of tES or sES with three approaches. The first approach was to consider only sES (AUD\$1200) in all individuals with Sanger sequencing for parental segregation of identified variants (All sES). The contrasting approach was to undertake only tES (AUD\$3600) in all individuals, without the need for parental segregation Sanger sequencing (“All tES”). The third approach was to consider sES (AUD\$1200) in all individuals upfront, with Sanger sequencing for parental segregation of identified variants and reflexing to post-hoc tES by adding parental ES (additional AUD\$2400) in those who remain undiagnosed by sES (Reflex tES).

To understand how diagnostic yield might affect costs of each of these implementation models, we considered each of them in scenarios using published minimal and maximal rates of sES (20.6 vs. 43%) and tES (31 vs 53%) from Table 1. Scenario 1 considers a low *a priori* sES diagnostic yield with a small incremental gain with tES (low likelihood of monogenic disorder, low yield from sES and tES). Scenario 2 considers a low *a priori* sES diagnostic yield but a large incremental gain with tES (high likelihood of monogenic disorder, with advantages of tES maximised). Scenario 3 considers a high *a priori* sES diagnostic yield but small incremental gain with tES (high likelihood of monogenic disorder, but most diagnoses made by sES).

To determine confidence intervals on the costs estimated for each of the three implementation approaches, bootstrapping of cost data were performed as previously described [32, 33]. Sensitivity analysis was undertaken with a higher cost of sES (USD\$5000; AUD\$6520.40) and tES (USD \$9000; AUD\$11736.86) to confirm robustness of conclusions.

## Results

### Demographics and complexity of medical care

Thirty families were recruited after seven proposed probands were excluded for various reasons (Table 2). There were 14 (46.7%) males and 16 (53.3%) females with a median age at recruitment of 21.5 months (4–650 months). Neurodevelopmental issues were identified in the majority of probands (24/30, 80%; Table 2). On average, each individual had 4.2 specialty medical units involved in their care, and prior to recruitment had attended a median of 5 (range 1–37) specialist medical appointments for diagnostic purposes. Hospital admissions for diagnostic and management purposes were recorded in 25/30 (83.3%), for a total of 78 admissions (3.1 admissions/individual) with each admission being a median of 6 days (range 1–37). Admissions to an intensive care unit occurred in 5/30 (16.7%) individuals, and the average length of stay per admission was 12.4 days (2–20 days). Each individual had a mean of 3.6 (range 2–6) body systems affected by their condition, and 8.4 (range 3–27) HPO phenotypic descriptor terms designated by their recruiting clinical geneticist. Gene lists prioritised for initial analysis were typically broad, with 18/30 (60%) individuals having >100 genes in each list.

### tES did not significantly increase diagnostic yield compared to sES in this cohort

The diagnostic rate of sES was 11/30 (36.7%) compared to 12/30 (40.0%) by tES (Supplementary Table 1). Of the 12 diagnoses, seven (58.3%) were de novo heterozygous events and five (41.7%) were biallelic variants in autosomal recessive conditions. The single diagnosis missed by sES but identified by tES was a de novo missense variant in *HNRNPU* (MIM 602869). Four individuals were found to have additional findings; two were true incidental variants with medical actionability unrelated to their presenting phenotype while two were variants related to known clinically diagnosed conditions (Neurofibromatosis type 1 and G6PD deficiency). The first true incidental variant (*LZTR1*; susceptibility to Schwannomatosis MIM 615670) was identified in an individual with failure to thrive and global developmental delay and the second (*BAG3*; dilated cardiomyopathy MIM 613881) was identified in an individual

with microcephaly, global developmental delay and stroke. Both variants were maternally inherited and their identification has led to altered medical management of the carrier mothers. The *NFI* and *G6PD* variants were also maternally inherited, but both diagnoses were already known in the family, and neither explained the proband's presenting phenotype. These were not counted as diagnoses for the purposes of the cost analysis.

### **tES led to fewer variants selected for curation, avoided variants of uncertain significance in negative cases and obviated the need for parental segregation sequencing**

At the variant prioritisation stage, the trio sequencing approach almost halved the number of variants selected for curation compared to singleton sequencing without parental alleles (tES 33 variants vs. sES 56 variants). There were 13 variants of uncertain significance in nine cases selected for curation after sES, that were not considered in the tES analysis as they had been filtered out with availability of parental ES. Trio analysis also obviated the need for parental segregation Sanger sequencing of 24 variants.

### **Variant prioritisation and curation times were approximately halved by tES compared to sES**

Trio ES analysis was time-efficient compared to sES, mostly driven by a reduction in the number of variants prioritised and curated. Time taken for clinicians to prioritise variants for curation, scientists to curate the variants, and MDT discussion time were all significantly less for tES compared to sES (Table 3 and Supplementary Table 3). There was no significant difference in the curation time per variant between sES and tES.

### **tES lowers analysis cost compared to sES, but it approximately doubles the cost of sES when sequencing is also considered. Reflex tES costs less than first line tES on all cases**

When evaluating the cost differences between sES and tES, we considered the cost of analysis first, then the combined

cost of analysis and sequencing. The cost of parental segregation of variants was incorporated into the cost of sES analysis, but was not necessary when calculating tES costs. When considering analysis only, the cost-per-diagnosis of sES was higher ( $\$2011.93 \pm \$428.98$ ) compared to tES ( $\$431.90 \pm \$110.59$ ). However, when the cost of sequencing was included in the analysis, the cost-per-diagnosis of sES ( $\$5284.63 \pm \$217.00$ ) was almost halved compared to tES ( $\$9429.27 \pm \$52.46$ ). The hypothetical intermediate strategy of starting with sES on all cases with reflex to tES only in those who remained undiagnosed resulted in a lower cost-per-diagnosis of  $\$8442.45 \pm \$441.66$  compared to doing tES on all cases up-front in this cohort.

### **An increased differential in diagnostic yield is required to reduce tES costs relative to sES**

When we modelled the cost-per-diagnosis of the three implementation models (All sES vs All tES vs sES followed by reflex tES) and varied the diagnostic yield of sES and tES, we found that the model that gave the lowest cost per diagnosis was reflex tES in a setting where there was the greatest incremental gain in diagnostic yield from tES (Scenario 2, Table 4). This was the only model in which there was a mean cost saving over sES alone in all cases ( $\$3017.55$  per diagnosis). When the incremental gain in diagnostic yield is lower (Scenarios 1 and 3), reflex tES is intermediate in its costs between undertaking sES or tES on all cases. Our sensitivity analysis demonstrated robustness of the data across all scenarios.

## **Discussion**

This study presents a novel approach to understanding the factors to consider when deciding the optimal sequencing strategy of individuals with monogenic disorders, specifically complex phenotypes. Our head-to-head comparison minimises bias that can be potentially introduced by comparing diagnostic rates of sES vs tES between individuals who underwent one approach or the other. It is clear that tES allows for faster recognition of a causative variant, especially when it is de novo. This was illustrated

**Table 3** Mean recorded durations for each curation task for sES compared to tES

Curation task	sES	tES	Difference in means	95% CI around difference in means	<i>p</i> -value
Clinician variant prioritisation per case	54 min 41 s	23 min 13 s	31 min 28 s	12 min 41s–50 min 15 s	0.002
Scientist curation per case	82 min 57 s	51 min 32 s	31 min 25 s	8 min 23 s–54 min 27 min	0.009
Scientist curation per variant	44 min 26 s	46 min 51 s	N.A.	N.A.	N.S.
MDT discussion per case	6 min 4 s	4 min 40 s	1 min 24 s	22 s–2 min 27 s	0.01

CI confidence interval, *min* minutes, *NA* not applicable, *NS* not significant, *s* seconds. Please see Supplementary Table 3 for details



**Table 4** Bootstrapped cost-per-diagnosis of three implementation models in hypothetical diagnostic yield scenarios

Scenario	N	All sES (diagnostic yield %)	All tES (diagnostic yield %)	Reflex tES (Hypothetical)
Observed	30	30	30	30
	No. of diagnoses	11 (36.7%)	12 (40%)	11 + 1
	Mean cost-per-diagnosis	\$5284.63	\$9429.27	\$8442.45
	CI	\$5067.64–\$5501.64	\$9376.81–\$9,481.73	\$8000.79–\$8884.11
Scenario 1: Low likelihood monogenic disorder, low sES yield	No. of diagnoses	6 (20.6%)	9 (31%)	6 + 3
	Mean cost-per-diagnosis	\$9389.21	\$12,166.80	\$10,893.48
	CI	\$9003.66 – \$9774.75	\$12,099.12 – \$12,234.49	\$10,323.60 – \$11,463.37
Scenario 2: High likelihood monogenic disorder, low sES yield	No. of diagnoses	6 (20.6%)	16 (53%)	6 + 10
	Mean cost-per-diagnosis	\$9389.21	\$7116.43	\$6371.66
	CI	\$9003.66 – \$9774.75	\$7076.84 – \$7156.02	\$6038.33 – \$6704.99
Scenario 3: High likelihood monogenic disorder, high sES yield	No. of diagnoses	13 (43%)	16 (53%)	13 + 3
	Mean cost-per-diagnosis	\$4498.08	\$7116.43	\$6371.66
	CI	\$4313.38 – \$4682.79	\$7076.84 – \$7156.02	\$6038.33 – \$6704.99

CI confidence interval, *no.* number

by the individual in our cohort diagnosed by tES whose causative variant was not selected for curation in the sES analysis because it was a missense variant in a gene that was only recently recognised to cause intellectual disability [34, 35]. Parental ES allows immediate prioritisation of de novo or compound heterozygous variants in clinically relevant genes.

The lower incremental diagnostic gain (3.3%) of tES over sES in this study compared to other published rates [3, 6, 15, 18] may be due to our experience with sES analysis and utilisation of a clinically integrated approach where clinical geneticists and curation scientists work in close collaboration on every case, resulting in a relatively high sES detection rate, with a lower latent gain from tES [9, 16]. It may also reflect an unintended selection bias, where cases with non-specific phenotypes were preferentially proposed for tES, some of which may be multifactorial or non-genetic. This selection bias may also apply to the real-world setting where tES might be chosen for non-specific cases where the differential diagnosis might be especially broad, e.g., non-syndromic intellectual disability [36], or there may be limited phenotypic information, e.g., fetal presentation [37]. However, the primary objective of this study was to evaluate the analytic process of tES compared to sES, with diagnostic yield being a secondary outcome.

Much of the efficiency of tES over sES comes from the nearly two-fold reduction in the number of variants selected for curation because of the use of parental ES information. This is predicated on the underlying assumptions that both biological parents have been sequenced, are unaffected by the disorder in the proband and that the disorder has complete penetrance and its causative gene is not subject to imprinting, which may result in erroneous tES filtering [38]. The use of parental alleles with which to annotate each of the variants in the proband can be powerful in excluding those that cause severe conditions expected to be penetrant in the carrier parent, as well as to establish phase when considering two variants in the same gene causing an autosomal recessive condition. In our cohort, 7/30 (23.3%) probands were found to have a de novo variant causative for their phenotype, and 5/30 (16.7%) had biallelic variants, with each parent being a heterozygous carrier. Thirteen heterozygous variants of uncertain significance in nine probands (from a total of 56 variants selected from sES analysis) (34%) were rapidly excluded as causative by tES because they were inherited. In the sES analysis, nine families would have been informed of a potentially causative variant, with that variant subsequently excluded once segregation results were available, potentially causing undue anxiety or distress. An alternative strategy to address this might be to collect parental DNA on all cases prior to sES, thus avoiding the need to contact parents to arrange

segregation testing. This would add costs of extraction and storage of DNA samples that might not be required.

The reduction in analysis time of tES over sES reflects fewer variants to prioritise and curate, given that the curation time per variant was similar between sES and tES. Further efficiencies came from the avoidance of 24 variant segregations because of the availability of tES data, resulting in a faster time-to-report, although this was not specifically quantified in our study. Furthermore, reducing time of both variant prioritisation and variant curation translates to less costly analysis and higher laboratory throughput, including number of cases discussed at a MDT meeting. This opportunity cost is an important consideration in the overall workflow of a health service, especially when scaling up operations. The monetary value of this opportunity cost is difficult to estimate, but is likely to be substantial. Additionally, the staff cost of the MDT meeting is significant and any workflow efficiencies gained by tES will directly impact this. Finally, whilst the curation bottleneck and scalability of service are important reasons for a health service to consider tES upfront, an additional factor is related to missed opportunities for novel gene discovery with sES alone.

Our finding of a nearly 5-fold reduction in analysis cost-per-diagnosis for tES compared to sES ( $\$431.90 \pm \$110.59$  vs.  $\$2011.93 \pm \$428.98$ ) reflects the efficiency gains from tES. This cost difference would likely be much larger if the entire analysis was performed by a curation scientist without integrated clinician input. Clinical input to variant prioritisation and curation is useful due to direct knowledge of the patient's phenotype and ability to correlate with previously published cases, with resultant boosts in diagnostic yield [1, 39]. Direct involvement of the primary clinician in variant prioritisation and tight clinical integration may not always be possible in all settings and may not be scalable, but we believe it is beneficial in exome analysis.

The patients recruited to our study were phenotypically complex and well-suited to tES. The latent gain in diagnostic yield from tES may be lower in patients with less complex phenotypes, such as single-system disease, where the differential diagnosis and gene list are shorter. In simpler cases, there are fewer variants to curate and less reliance on clinical correlation. The difference in analysis cost-per-diagnosis may be less marked and a reasonable strategy might involve sES with a limited analysis of a short gene list. Our study did not model the possible impact of falling costs of data generation on our conclusions. However, data generation costs in a clinical setting include many factors over and above raw reagent costs. The impact of a decline in reagent costs on the overall cost per diagnosis in tES would be significantly diluted and also apply to the cost per diagnosis of sES. A decline in reagent costs is likely to have

only minor impact on the overall conclusions of this study in the foreseeable future.

Our analysis demonstrates that the cost of different sequencing strategies is context-dependent. In a setting where individuals who have a low *a priori* chance of a monogenic disorder are being referred for ES (Table 4, Scenario 1), the diagnostic yield will be low with high cost-per-diagnosis regardless of sES or tES approach, and tES costs more than sES. However, tES has lower costs when there is a large incremental gain in diagnostic yield (Table 4, Scenario 2). In a real-world setting, this is akin to a reference laboratory that might have a typical diagnostic yield of 20–25% for sES [25], but rigorous case selection by the ordering clinician increases the likelihood of a positive test result. In this context, tES costs less, with the greatest potential incremental gain in diagnostic yield by adding parental alleles to the analysis. The real-world setting of our study is best reflected in Scenario 3, where the likelihood of a monogenic disorder is high due to rigorous case selection and accurate phenotyping, but the baseline *a priori* sES diagnostic yield is high because of clinical integration to the analysis workflow. In this scenario, sES is less costly than doing all cases as tES up-front, with the compromise being a possible missed diagnosis that would have been diagnosed with tES.

When we consider the implementation model of starting with sES and reflexing to post-hoc tES in those who remain undiagnosed, this has lower costs than tES on all cases upfront in all diagnostic yield scenarios. However, reflex tES costs less than sES on all cases in the setting of a low *a priori* diagnostic yield but high likelihood of a monogenic disorder (Scenario 2, Table 4). This suggests that at point of ordering, if a reference lab is being considered, then it may cost least to start with sES then reflex to tES after a negative result as long as the clinician still has a strong suspicion for a monogenic condition. If the result is time-critical, then ordering tES up-front will cost more than reflex tES, but is more time-efficient without the need to take a two-step approach to do post-hoc parental sequencing. If a laboratory with close clinical integration is being considered then sES up-front on all cases is the least costly approach, but additional diagnoses can be achieved by reflexing to tES with additional cost. Reflex tES requires two additional analyses per negative patient (sES followed by tES) and this has implications for clinical and laboratory workflows and scalability and may ultimately render this option impracticable at scale.

For the clinician, the choice of sES or tES takes into account a complex set of variables including family structure and availability, funding, clinical urgency, clinic structure and preferred laboratory. Three key factors which emerge in our analysis for clinicians to consider are: what is the *a priori* chance of a monogenic condition in this patient? And what is the typical diagnostic yield for their laboratory

of choice for sES compared with tES? Finally, does the laboratory incorporate clinical expertise in their data analysis, which may influence diagnosis rate by either sES or tES? Knowledge of these factors will assist clinicians in choosing the optimal test for their patient.

Our findings not only provide an evidence-base for assertions about the diagnostic efficiencies of tES compared to sES [3], but provides cost data on sES and tES allowing for a head-to-head comparison of the two tests. Our study highlights the importance of test context in terms of clinical-laboratory integration, patient phenotype and a priori chance of a monogenic condition as important factors when considering test selection. Finally, as the demand for genomics escalates, the study draws attention to systemic factors relevant in considering test selection including opportunity costs for clinicians and scientists, and scalability of test options.

**Acknowledgements** The authors thank all doctors who referred patients to the study, and the families who participated.

**Funding** The study was funded by the State Government of Victoria (Department of Health and Human Services) and the 10 member organisations of the Melbourne Genomics Health Alliance (The Royal Melbourne Hospital, The Royal Children's Hospital, The University of Melbourne, The Walter and Eliza Hall Institute, Murdoch Children's Research Institute, CSIRO, Australian Genome Research Facility, Peter MacCallum Cancer Centre, Austin Health and Monash Health).

**Author contributions** Project conception, design, leadership: TYT, SL, CLG, SMW. Data generation: SL, BC, DP, MFF, JEM, VSK, ZS, AY, NJB, CS, MBD, SS, MM, IG. Cost and statistical analyses: MM, IG, NT. Manuscript major draft writing: TYT, SMW.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

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

## References

- Baldrige D, Heeley J, Vineyard M, Manwaring L, Toler TL, Fassi E et al. The Exome Clinic and the role of medical genetics expertise in the interpretation of exome sequencing results. *Genet Med.* 2017;19:1040–8.
- Bick D, Fraser PC, Gutzeit MF, Harris JM, Hambuch TM, Helbling DC, et al. Successful application of whole genome sequencing in a medical genetics clinic. *J Pediatr Genet.* 2017;6:61–76.
- Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med.* 2015;17:578–86.
- Iglesias A, Anyane-Yeboah K, Wynn J, Wilson A, Truitt Cho M, Guzman E, et al. The usefulness of whole-exome sequencing in routine clinical practice. *Genet Med.* 2014;16:922–31.
- Kuperberg M, Lev D, Blumkin L, Zerem A, Ginsberg M, Linder I, et al. Utility of whole exome sequencing for genetic diagnosis of previously undiagnosed pediatric neurology patients. *J Child Neurol.* 2016;31:1534–9.
- 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.
- 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.
- Srivastava S, Cohen JS, Vernon H, Baranano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. *Ann Neurol.* 2014;76:473–83.
- 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.
- Stavropoulos DJ, Merico D, Jobling R, Bowdin S, Monfared N, Thiruvahindrapuram B et al. Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom Med.* 2016; 1.
- Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadeh-Djebbar S, et al. Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin Genet.* 2016;89:700–7.
- Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med.* 2013;369:1502–11.
- 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.
- Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med.* 2014;312:1870–9.
- Monies D, Abouelhoda M, AlSayed M, Alhassnan Z, Alotaibi M, Kayyali H, et al. The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. *Hum Genet.* 2017;136:921–39.
- 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.
- Sawyer SL, Hartley T, Dymont DA, Beaulieu CL, Schwartzentruber J, Smith A, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet.* 2016;89:275–84.
- 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.* 2015;18:696–704.
- Costain G, Jobling R, Walker S, Reuter MS, Snell M, Bowdin S, et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur J Hum Genet.* 2018;26:740–4.
- 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.
- Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med.* 2017;19:209–14.
- 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.* 2018;21:173–80.
23. Nambot S, Thevenon J, Kuentz P, Duffourd Y, Tisserant E, Bruel AL, 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.
  24. 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.
  25. Clark MM, Stark Z, Farnaes L, Tan TY, White SM, Dimmock D, et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med.* 2018;3:16.
  26. Gaff CL, M. Winship I, M. Forrest S, P. Hansen D, Clark J, M. Waring P, et al. Preparing for genomic medicine: a real world demonstration of health system change. *npj Genom Med.* 2017;2:16.
  27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inf.* 2009;42:377–81.
  28. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chenier S, et al. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat.* 2013;34:1057–65.
  29. Sadedin SP, Dashnow H, James PA, Bahlo M, Bauer DC, Lonie A, et al. Cpipe: a shared variant detection pipeline designed for diagnostic settings. *Genome Med.* 2015;7:68.
  30. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat.* 2011;32:557–63.
  31. Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med.* 2008;10:294–300.
  32. Goranitis I, Coast J, Day E, Copello A, Freemantle N, Frew E. Maximizing Health or Sufficient Capability in Economic Evaluation? A Methodological Experiment of Treatment for Drug Addiction. *Med Decis Mak.* 2017;37:498–511.
  33. Goranitis I, Bellanca L, Daley AJ, Thomas A, Stokes-Lampard H, Roalfe AK, et al. Aerobic exercise for vasomotor menopausal symptoms: a cost-utility analysis based on the Active Women trial. *PLoS ONE.* 2017;12:e0184328.
  34. Bramswig NC, Ludecke HJ, Hamdan FF, Altmüller J, Beleggia F, Elcioglu NH, et al. Heterozygous HNRNPU variants cause early onset epilepsy and severe intellectual disability. *Hum Genet.* 2017;136:821–34.
  35. Epi KC, Epilepsy Phenome/Genome P, Allen AS, Berkovic SF, Cossette P, Delanty N, et al. De novo mutations in epileptic encephalopathies. *Nature.* 2013;501:217–21.
  36. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 2012;367:1921–9.
  37. Best S, Wou K, Vora N, Van der Veyver IB, Wapner R, Chitty LS. Promises, pitfalls and practicalities of prenatal whole exome sequencing. *Prenat Diagn.* 2018;38:10–19.
  38. Aten E, Fountain MD, van Haeringen A, Schaaf CP, Santen GW. Imprinting: the Achilles heel of trio-based exome sequencing. *Genet Med.* 2016;18:1163–4.
  39. Salmon LB, Orenstein N, Markus-Bustani K, Ruhrman-Shahar N, Kilim Y, Magal N et al. Improved diagnostics by exome sequencing following raw data reevaluation by clinical geneticists involved in the medical care of the individuals tested. *Genet Med.* 2018;21:1443–51.