Cellular Physiology and Biochemistry

Cell Physiol Biochem 2018;49:295-305

DOI: 10.1159/000492879

Published online: 23 August, 2018

Accepted: 14 August, 2018

Original Paper

Genetic Evaluation of 114 Chinese Short Stature Children in the Next Generation Era: a Single Center Study

Zhuo Huang[a,b] Yu Sun[a,b] Yanjie Fan[a,b] Lili Wang[a,b] Huili Liu[a,b] Zhuwen Gong[a,b] Jianguo Wang[a,b] Hui Yan[a,b] Yu Wang[a,b] Guorui Hu[a,b] Ruifang Wang[a,b] Jun Ye[a,b] Lianshu Han[a,b] Wenjuan Qiu[a,b] Huiwen Zhang[a,b] Lili Liang[a,b] Yu Yang[c] Andrew Dauber[d] Yongguo Yu[a,b] Xuefan Gu[a,b]

[a]Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, [b]Molecular Genetics Group, Shanghai Institute for Pediatric Research, Shanghai, [c]Jiangxi Provincial Children's Hospital, Nanchang, China, [d]Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH USA

Key Words

Short stature • Genetic etiology • Next generation sequencing • Chromosomal microarray analysis

Abstract

Background/Aims: The genetics of human height is a frequently studied and complex issue. However, there is limited genetic research of short stature. To uncover the subgroup of patients to have higher yield and to propose a simplified diagnostic algorithm in the next generation era. ***Methods:*** This study included 114 Chinese children with height SDS ≤ -2.5 and unknown etiology from 2014 to 2015. Target/whole exome sequencing (referred as NGS) and chromosomal microarray analysis (CMA) were performed on the enrolled patients sequentially to identify potential genetic etiologies. The samples solved by NGS and CMA were retrospectively studied to evaluate the clinical pathway of the patients following a standard diagnostic algorithm. ***Results:*** In total, a potential genetic etiology was identified in 41 (36%) patients: 38 by NGS (33.3%), two by CMA (1.8%), and an additional one by both (0.9%). There were 46 different variants in 29 genes and 2 pathogenic CNVs identified. The diagnostic yield was significantly higher in patients with facial dysmorphism or skeletal abnormalities than those without the corresponding phenotype (P=0.006 and P=0.009, respectively, Pearson's χ^2 test). Retrospectively study the cohort indicate 83.3% patients eventually would be evaluated by NGS/CMA. ***Conclusion:*** This study confirms the utility of high-throughput molecular detection techniques for the etiological diagnosis of undiagnosed short stature and suggests that NGS could be used as a primary diagnostic strategy. Patients with facial dysmorphism

Xuefan Gu, MD, PhD
and Yongguo Yu, MD, PhD

Dept. of Pediatric Endocrinology/Genetics, Xinhua Hospital, School of Med., Shanghai Jiao Tong Univ.
Room 801 Sci&Edu Bldg, 1665 Kongjiang Rd, Yangpu district, Shanghai 200092 (China)
Tel. +86-21-25076453, E-Mail guxuefan@xinhuamed.com.cn; yuyongguo@shsmu.edu.cn

and/or skeletal abnormalities are more likely to have a known genetic etiology. Moving NGS forward would simplified the diagnostic algorithm.

Introduction

Short stature is defined as a height that is more than 2 standard deviations (SDS) below the corresponding mean for a given age, sex and population [1]. Children with severe short stature have been found to be vulnerable to diverse developmental, social and educational problems [2]. Over one third of the children seen in our pediatric endocrinology clinic are referred for short stature. It is estimated that approximately 80% of height variation within a given population is under genetic control [3]. Genome-wide association studies have identified 697 common variants clustered in 423 loci that together explained 20% of the variation in adult height [4]. In contrast to contributions from common variants in multiple genes on normal height variation, extremes in height are more likely to be caused by rare pathogenic variants in monogenic genes critical for growth control [5]. Wang et al. [6] investigated 1077 genes in 192 children with short stature with no known genetic etiology revealed some cases carrying pathogenic variants in monogenic genes known to cause short stature. However, variant frequency in that cohort were inferred by pooled targeted sequencing. The overall diagnostic yield is difficult to calculate [6]. Other recent studies have shown that 2.5%-10% patients with idiopathic short stature carry disease-causing rare copy number variants (CNVs), indicating that chromosomal microarray analysis (CMA) can also be used as a technique for unexplained short stature [7, 8]. In recent years, endocrinologists realize the importance of genetic evaluation of short stature, especially the pathogenic short stature, as knowing the molecular diagnosis of the disease would help to predict the development of diseases, guide the potential treatment of short stature and future family planning. Studies have proposed several factors that might increase the likelihood for a monogenic cause of short stature based on clinical experiences [9, 10], such as severe short stature, multiple pituitary hormone deficiency, etc. However, the experiment evidence is lacking.

To examine the genetic causes of short stature, experts have suggested algorithms to diagnose in a stepwise manner. The approach includes various branches, mixed with evaluation of patients' growth-related factors, endocrine system, skeletal anomaly and application of different choices of genetic techniques at different stages of the diagnose and depends on clinicians' experience greatly. Even if the distinct disorders were recognized correctly by clinician, the detect efficiency is low. One reason is the genetic heterogeneity, multiple genes might be involved in one entity, such as Noonan syndrome. Another reason is some of the genes are large in size. It would be both time and labor consuming to investigate by conventional Sanger sequencing. High throughput techniques, like next generation sequencing (NGS) and CMA could detect the variants in genome wide in a single test. And the pricing keep decrease. Thus, the application of such methods in clinical is becoming practical. A guidance of the genetic diagnosis of short stature in this next generation era is in urgent demands.

In this study, we performed NGS and CMA in a cohort of Chinese children with short stature to explore their potential genetic causes. We compared the yields among patients categorized into different phenotypic subgroups and tried to provide meaningful guidance for clinicians to better utilize these high-throughput molecular detection techniques in children with undiagnosed short stature. Furthermore, we retrospectively examined the 114 cases with a standard diagnostic algorithm, and calculated the steps and labors would cost. Finally, we suggested a new algorithm to diagnose short stature in the next generation era.

Materials and Methods

Cohort

This study was approved by Ethics Committee of Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University (XHEC-C-2017-066). All recruited subjects or their legal guardians provided written informed consent. The cohort was selected from children (age < 18 years) characterized by short stature (height SDS (HtSDS) < -2, compared to age- and sex- matched Chinese population [11]) who visited or were referred to the pediatric endocrinology clinic from January 2014 to December 2015. Some recognizable causes of short stature had been excluded through prescreening tests: (1) postoperative pituitary tumor; (2) congenital hypothyroidism without treatment; (3) congenital adrenal hyperplasia; (4) hypophosphatemic rickets; (5) Turner syndrome; (6) Down syndrome; (7) achondroplasia caused by *FGFR3* G380R variant; (8) Prader-Willi syndrome; (9) mucopolysaccharidosis (MPS), GM1-gangliosidosis, and mucolipidosis type II/III; (10) glycogen storage disease type I; (11) inherited metabolic disorders associated with short stature; (12) pathogenic CNVs had been identified in some patients with short stature and intellectual disability, and/or multiple congenital anomalies before this study. In total, there were 656 children included in the initial cohort. We further selected individuals with severe short stature (HtSDS ≤ -2.5) with available DNA samples from nuclear family members. Consequently, 114 patients were enrolled in the final cohort (Fig. 1).

The clinical records of the 114 patients were collected and categorized as follows: age, HtSDS and weight SDS at the first visit/referral or before rhGH treatment. Family history of short stature (HtSDS < -2) or similar phenotypes seen within three generations was considered as positive. Small for gestational age (SGA) was defined as birth weight or/and birth length < -1.88 SDS (i.e. 3rd percentile) relative to gestational age- and sex- matched controls in China. Patients with average DQ/IQ < 70 were diagnosed with developmental delay/intellectual disability (DD/ID). Other congenital anomalies included manifestations presented at birth but not listed separately such as cryptorchidism, hearing loss, cleft lip and palate, and so on. GH provocation tests used arginine and clonidine respectively without sex steroid priming, and the results showed peak serum GH levels < 10 ng/ml were termed as GH deficiency (GHD) and < 5 ng/ml as complete GHD. Insulin like growth factor 1 (IGF1) level lower than age-matched reference was defined as low IGF1 [12]. Patients with GHD and other pituitary hormone deficiency were diagnosed with multiple pituitary hormone deficiency (MPHD). Bone age was assessed from a standard left hand and wrist radiograph and read according to the method of Greulich and Pyle [13], and bone age was defined as normal if the difference between the bone age and chronological age was within +/- one year. Patients without additional phenotype but having positive family history, abnormal GH and/or IGF1 level, or slight abnormalities on imaging were classified as isolated short stature; patients with only one additional congenital anomaly or abnormal biochemical finding were classified as short stature with one additional phenotype; the rest were classified as short stature with more than one additional phenotype.

DNA extraction

Genomic DNA was extracted from peripheral blood of the enrolled patients, their parents and other available family members using blood genomic DNA extraction kits (Zeesan Biotech, China).

NGS and Validation

For target sequencing and whole exome sequencing (WES), capture library was prepared using ClearSeq Inheried Disease panel and SureSelect All Exon V5 (Agilent) respectively. Among the 114 enrolled patients, 102 underwent target sequencing, 12 underwent WES. The library was sequenced by Illumina HiSeq 4000 to generate 150 bp paired end reads. Data analysis was performed as previously described [14]. In general, the raw data was aligned to the human reference hg19 by BWA. Variants were called following GATK best practice (version 3) from pass filtered reads. The output vcf files were annotated by SNPEff. High frequency variants (with a frequency >1% in 1000 Genomes Project, Exome Aggregation Consortium (ExAC), Exome Variant Server (EVS), or >5% in local exome database with 200 exomes) were removed from the candidate variant list. Variants were subsequently filtered through autosomal recessive and autosomal dominant/de novo inherited pattern. Candidate variants were classified according to recent standards and guidelines from the American College of Medical Genetics and Genomics (ACMG) [15]. Sanger sequencing was applied to confirm and to determine the co-segregation of the underlying causative variants. All primers would be available upon request.

CMA and Validation

For undiagnosed patient by NGS in this study, CMA was performed using Affymetrix CytoScan HD array as previously described [16]. The pathogenicity of detected CNVs and loss of heterozygocity regions (LOH) were classified based on the ACMG recommendations [17]. In combination of NGS data, if one patient carries a heterozygous pathogentic/likely pathogenic variant in a certain recessive short stature gene, signal of this gene was checked specifically to detect exonic CNVs that might be missed by standard procedure. Quantitative polymerase chain reaction (qPCR) was applied to confirm and to determine the co-segregation of the candidate CNVs. All primers would be available on request.

Statistical analysis

All statistical analyses were performed using SPSS software version13.0. Pearson's χ^2 test (both the total number of valid cases ≥ 40 and all cells had expected count ≥ 5) / Fisher's exact test (the total number of valid cases < 40, more than 1/5 cells had expected count < 5 or 1 cell had expected count < 1) was used to show the differences of diagnostic categories (solved or unsolved) between patients with and those without a certain phenotype, or between patients with different phenotypes. The odds ratio (OR) and 95% confidence interval (95% CI) of different diagnostic categories for the selected factors (family history, SGA, microcephaly, facial dimorphism, skeletal abnormalities, DD/ID, cardiac anomaly, other congenital anomalites, GHD, low IGF1, other biochemical anomalies, abnormal BA and abnormal brain MRI imaging) were also calculated, and forest plots were generated by Stata 12.0 statistical package.

Results

Cohort description

The clinical characteristics of this cohort were summarized in Table 1 (detailed in Supplementary Table 1).

For all supplemental material see www.karger.com/10.1159/000492879.

Only 18 patients were examined with isolated short stature. The rest (96/114) demonstrated at least one additional abnormality. HtSDS distribution of the enrolled patients was shown in Fig. 2A. Most of them are above -4 HtSDS (63.2%).

Genetic findings

An overview of the NGS data is summarized in Supplementary Table 2. In total, we identified 46 different variants in 29 genes including 19 variants not reported previously, and three pathogenic/likely pathogenic CNVs in two other patients (Table 2), which supports the high genetic heterogeneity

Table 1. Clinical characteristics and corresponding diagnostic yields of the 114 enrolled patients. Abbreviations: HtSDS, standard deviation score of height; NA, not available; GH, growth hormone; CGHD, complete growth hormone deficiency; MPHD, multiple pituitary hormone deficiency; IGF1, insulin like growth factor 1; MRI, magnetic resonance imaging; SGA, small for gestational age. #P value was calculated for the differences of diagnostic categories (solved or unsolved) between patients with and those without a certain phenotype

	n (%) (Total=114)	Diagnostic yield	P value#
Age at the first visit (years): median (range)	4.0 (0.3 to 17.3)	/	
HtSDS at the first visit: median (range)	-3.7 (-9.7 to -2.5)	/	
Sex			
Male	70 (61.4%)	/	
Female	44 (38.6%)	/	
Family history			0.150
Yes	22 (19.3%)	22.7%	
No	92 (80.7%)	39.1%	
GH provocation			0.322 (0.300 for CGHD)
Deficiency	35 (30.7%)	22.9%	
CGHD	13 (11.4%)	15.4%	
Sufficiency	16 (14.0%)	37.5%	
NA	63 (55.3%)	/	
IGF1			0.212
Deficiency	37 (32.5%)	24.3%	
Sufficiency	40 (35.1%)	37.5%	
NA	37 (32.5%)	/	
Bone age			0.739
Delayed	38 (33.3%)	26.3%	
Advanced	3 (2.6%)	0.0%	
Normal	16 (14.0%)	31.3%	
NA	57 (50.0%)	/	
Brain MRI			0.183
Abnormal	32 (28.1%)	25.0%	
Normal	32 (28.1%)	40.6%	
NA	50 (43.9%)	/	
Isolated short stature	18 (15.8%)	0.0%	
Short stature with one additional phenotype	43 (37.7%)	37.2%	
Short stature with more than one additional phenotype	53 (46.5%)	47.2%	
Main additional phenotype			
SGA	21 (18.4%)	38.1%	0.822
Microcephaly	21 (18.4%)	52.4%	0.083
Facial dysmorphism	30 (26.3%)	56.7%	0.006
Skeletal abnormalities	51 (44.7%)	49.0%	0.009
Developmental delay/intellectual disability	31 (27.2%)	45.2%	0.211
Cardiac anomaly	16 (14.0%)	50.0%	0.207
Other congenital anomalies	21 (18.4%)	38.1%	0.822
Other biochemical anomalies	17 (14.9%)	41.2%	0.627

Table 2. NGS and CMA findings associated with a molecular diagnosis. Note: [a] Novel variants (i.e. variants absent in population and disease databases) are labeled underlined. [b] References of variants previously reported in patients were listed in Supplementary Table 3. [c] According to the recently published ACMG standards and guidelines. [d] Clinical information of patients were shown in Supplementary Table 1. [e] This large deletion was called by CMA (15q26.1(91297020-91308577)×1), and confirmed by qPCR (primers were designed ranging from chr15:91292776 to chr15:91304121). [f] This large deletion was validated by qPCR (primers were designed ranging from chr5:60213106 to chr5:60214502). [g] Numbers of overlapping CNVs reported as pathogenic/likely pathogenic in the DECIPHER database. Abbreviations: P, pathogenic; LP, likely pathogenic; VUS, variant of uncertain significance.

Gene	Associated disease	Inheritance	Sequencing variants (hg19) [ab]	Classification [c]	Patient [d]
Autosomal dominant					
BRAF	Cardio-facio-cutaneous Syndrome	De novo	NM_004333.4: c.1785T>G(p.F595L)	P	P34
COL1A1	Ehlers-Danlos syndrome	De novo	NM_000088.3: c.159G>C(p.W53C)	LP	P87
COL2A1	Spondyloepimetaphyseal dysplasia, Strudwick	De novo	NM_001844.4: c.3121G>A(p.G1041S)	P	P30
		De novo	NM_001844.4: c.2290A>G(p.K764E)	LP	P40
COMP	Pseudoachondroplasia	Maternal	NM_000095.2: c.875G>T(p.C292F)	LP	P3
		De novo	NM_000095.2: c.1417_1419del(p.D473del)	P	P65
CREBBP	Rubinstein-Taybi syndrome	De novo	NM_004380.2: c.4507T>C(p.Y1503H)	LP	P2
FBN1	Acromicric dysplasia	De novo	NM_000138.4: c.5096A>G(p.Y1699C)	P	P21
		De novo	NM_000138.4: c.5206T>C(p.C1736R)	LP	P25
FGFR3	Achondroplasia	De novo	NM_000142.4: c.1138G>A(p.G380R)	P	P8/P48/P74/P105
		De novo	NM_000142.4: c.1138G>C(p.G380R)	P	P108
	Hypochondroplasia	De novo	NM_000142.4: c.1620C>A(p.N540K)	P	P77
GH1	Growth hormone deficiency, isolated, type II	Maternal	NM_000515.4: c.291+1G>C	P	P61
HRAS	Costello syndrome	De novo	NM_005343.2: c.34G>A(p.G12S)	P	P60
KIF22	Spondyloepimetaphyseal dysplasia with joint laxity, type 2	De novo	NM_007317.2: c.443C>T(p.P148L)	P	P110
KMT2A	Wiedemann-Steiner syndrome	De novo	NM_001197104.1: c.8407C>T(p.Q2803*)	P	P80
KRAS	Noonan syndrome 3	De novo	NM_004985.4: c.458A>T(p.D153V)	P	P35
MATN3	Multiple epiphyseal dysplasia 5	Maternal	NM_002381.4: c.361C>T(p.R121W)	P	P31
RAF1	Noonan syndrome 5	De novo	NM_002880.3: c.770C>T(p.S257L)	P	P10/P113
RUNX2	Cleidocranial dysplasia	De novo	NM_001024630.3: c.574G>A(p.G192R)	P	P37
SRCAP	Floating-Harbor syndrome	De novo	NM_006662.2: c.7219C>T(p.Q2407*)	P	P16
X-linked dominant					
HDAC8	Cornelia de Lange syndrome 5	De novo	NM_018486.2: c.806T>G(p.I269R)	LP	P33
NAA10	Ogden syndrome	De novo	NM_003491.3: c.247C>T(p.R83C)	P	P98
Autosomal recessive					
ALPL	Hypophosphatasia, infantile	Paternal	NM_000478.4: c.1120G>A(p.V374M)	LP	P76
		Maternal	NM_000478.4: c.228G>T(p.Q76H)	LP	
BLM	Bloom syndrome	Paternal	NM_000057.3: c.3564del(p.F1189Lfs*10)	P	P75
		Maternal	NM_000057.3: Large deletion involving E3-E8 [e]	P	
DCHS1	Van Maldergem syndrome 1	Paternal	NM_003737.3: c.1753G>A(p.A585T)	VUS	P27
		Maternal	NM_003737.3: c.502C>T(p.R168C)	VUS	
ERCC8	Cockayne syndrome, type A	Paternal	NM_000082.3: c.394_398del(p.L132Nfs*6)	P	P83
		Maternal	NM_000082.3: Large deletion involving E4 [f]	P	
GALNS	Mucopolysaccharidosis IVA	Paternal	NM_000512.4: c.1334G>A(p.G445E)	VUS	P46
		Maternal	NM_000512.4: c.1451C>T(p.P484L)	LP	
GLB1	Mucopolysaccharidosis type IVB	Paternal	NM_000404.2: c.145C>T(p.R49C)	P	P57
		Maternal	NM_000404.2: c.248A>G(p.Y83C)	P	
OBSL1	3-M syndrome 2	Paternal	NM_015311.2: c.128_129insA(p.V44Gfs*210)	P	P53
		Maternal	NM_015311.2: c.664del(p.A222Rfs*36)	P	
PYCR1	Cutis laxa, type IIB	Paternal	NM_006907.3: c.751C>T(p.R251C)	LP	P5
		Maternal	NM_006907.3: c.397A>C (p.T133P)	LP	
SLC12A3	Gitelman syndrome	Paternal	NM_000339.2: c.1181G>A (p.G394D)	LP	P32
		Maternal	NM_000339.2: c.2877_2878del (p.R959Sfs*11)	P	
SLC12A3	Gitelman syndrome	Paternal	NM_000339.2: c.2039del(p.G680Dfs*21)	P	P50
		Maternal	NM_000339.2: c.947G>C (p.G316A)	LP	
SPINK5	Netherton syndrome	Paternal and Maternal	NM_006846.3: c.2474_2475del (p.E825Gfs*2)	P	P101
VPS13B	Cohen syndrome	Paternal	NM_017890.4: c.6940+1G>T	P	P41
		Maternal	NM_017890.4: c.9337A>T(p.N3113Y)	VUS	

CNV (hg19)	Size (kb)	OMIM gene (n)	DECIPHER [g] (gain/loss) (n)	Syndrome	Classification [c]	Patient [d]
7q31.1q31.31(111551633-118284664)×3	6733	20	2/7	/	P	P109
9p24.3(203861-1138636)×3	935	5	18/12	/	LP	P42
17p13.3(525-2117982)×1	2117	30	21/10	Miller-Dieker syndrome	P	P42

of the short stature cohort. Most of the affected genes and variants are not recurrent (Supplementary Fig. 1A, 1D).

18 mutant genes are related to autosomal or X-linked dominant disorders. The identified variants all arose *de novo*, except for the variants carried by P3, P31 and P61 which were maternally inherited. The other affected genes follow an autosomal recessive pattern of inheritance (Supplementary Fig. 1B). There were two large deletions detected in patient P83 and P75, respectively (Table 2, Supplementary Fig. 1E). A "homozygous" deletion was detected in *ERCC8* gene by NGS in P83. However, Sanger sequencing could only find the paternal origin of the variant. We therefore performed qPCR, and confirmed the maternal exon 4 was deleted (Table 2, Supplementary Fig. 2A). Additionally, P75 carries a paternal frameshift variation detected by NGS and a maternal large deletion detected by CMA in exon 3-8 of *BLM* gene (Table 2, Supplementary Fig. 2B). Four variants were classified with uncertain significance according to recently published ACMG recommendations (Table 2, Supplementary Fig. 1F). They were recognized to likely cause the diseases, as the patients' phenotype and inheritance patterns were consistent with the corresponding genetic disorders.

Furthermore, based on the recent relevant references [6, 18-20] and OMIM (https://www.omim.org/), we found that at least 24 of the detected 29 genes are affected in the recognized growth plate regulatory systems and skeletal development (Supplementary Fig. 1C). To our surprise only two cases (P61, P101) harbour mutations in the GH-IGF1 axis. This suggests the clinicians to shift the focus from classical GH-IGF1 axis to an extended pathway that involve the growth plate and skeletal development.

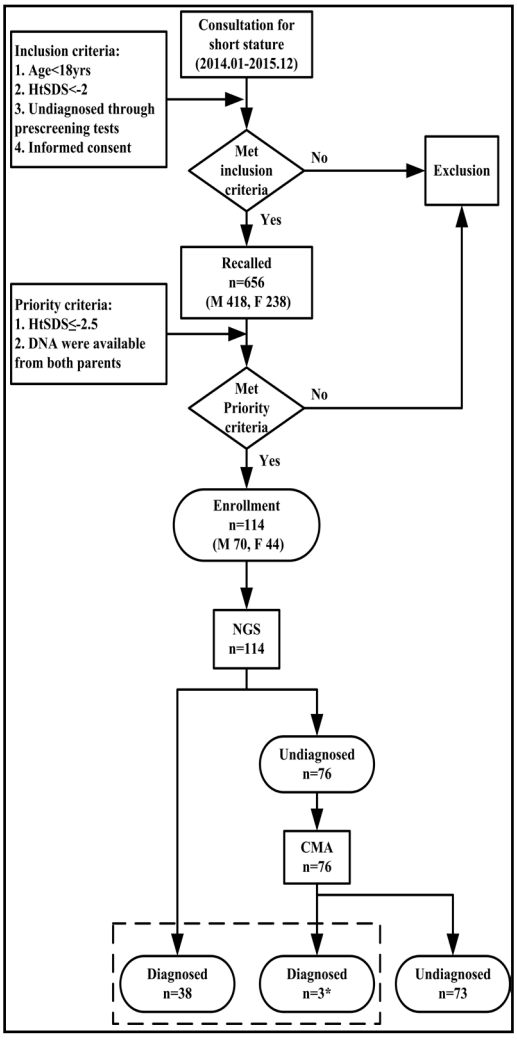

Fig. 1. Flowchart of study subjects. Note: * One patient had a combined diagnosis by NGS and CMA.Abbreviation: M, male; F, female; n, number; HtSDS, standard deviation score of height; NGS, next generation sequencing; CMA, chromosomal microarray analysis; rhGH, recombinant human growth hormone.

In addition, two patients (P42 and P109) were diagnosed by CMA. P42 carries one pathogenic loss known to cause Miller-Dieker syndrome and one likely pathogenic gain in 9p24.3 (Table 2, Supplementary Fig. 2C). P109 was validated to have a *de novo* gain of large size by qPCR (Table 2, Supplementary Fig. 2D). This 6.7 Mb duplication in 7q31.1q31.31 covers 20 OMIM genes and has not been related to a continuous gene duplication syndrome. However, the mechanism remains obscure.

Diagnostic yield of NGS and CMA

38 of the 114 patients (33.3%) was diagnosed by NGS (Fig. 1). The genetic findings are listed in Table 2. Subsequently, three out of 76 unsolved patients were tackled by CMA and NGS (Fig. 1). The overall diagnostic yield is 36.0% (41/114).

By calculating the diagnostic rate of different HtSDS ranges, patients with HtSDS between -4 and -5 and below -6 had a higher diagnostic rate than the overall yield (50.0% and 44.4%, respectively versus 36.0%), which suggests that individuals with more severe short stature tend to have a higher likelihood of finding a genetic cause. However, there had been no difference between each HtSDS group (Fig. 2A, χ^2=4.112, *P*=0.249).

We found that two phenotypes had statistical significance, namely facial dysmorphism (56.7% versus 28.6%, χ^2=7.576, *P*=0.006) and skeletal abnormalities (49.0% versus 25.4%, χ^2=6.829, *P*=0.009) by comparing the diagnostic rates between patients with and those without a given additional phenotype (Fig. 2B, Supplementary Table 4).

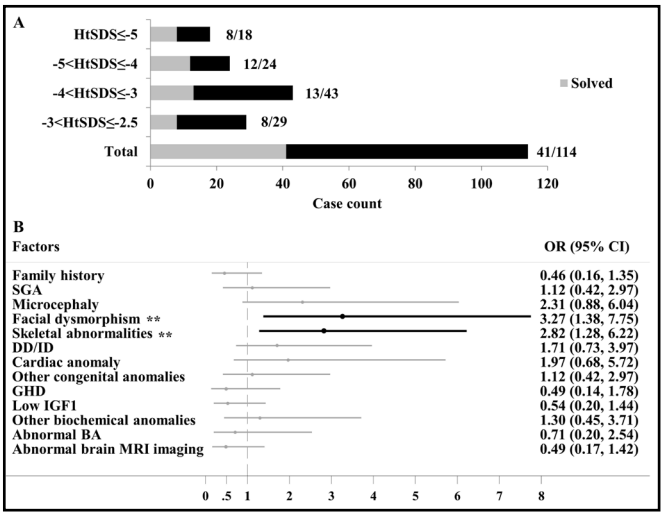

Fig. 2. Diagnostic yields and statistical analyses in patients categorized into different subgroups. (A) HtSDS distribution and corresponding diagnositc yields. Next to each bar, the denominator represents total cases and the numerator represents solved cases (shown in gray rectangles) corresponding to a given HtSDS range. (B) Forest plots with odds ratios (OR) and confidence intervals (CI) for some clinical factors. An OR with a lower 95% CI > 1 (vertical dash line) was considered a positive likehood of molecular diagnosis being associated to a given factor (shown in black bold lines). **P<0.01. Abbreviation: HtSDS, standard deviation score of height; SGA, small for gestational age; DD/ID, developmental delay/intellectual disability; GHD, growth hormone deficiency; IGF1, insulin like growth factor 1; BA, bone age; MRI, magnetic resonance imaging.

18 patients classified as isolated short stature still remained unexplained in this study (Table 1). Of the 96 patients had at least one additional phenotype, the diagnostic yield of patients with more than one additional phenotype was higher than those with only one additional phenotype (25/53, 47.2% versus 16/43, 37.2%, Table 1), but there was no statistical difference between the two subgroups (χ^2=0.963, *P*=0.327). The diagnostic rates vary in different additional phenotype subgroups. Each additional phenotype subgroup could be further divided into two groups: the additional phenotype being only one accompanying phenotype or one of accompanying phenotypes. However, there were no statistical significant differences amongst groups (all *P* > 0.05, Supplementary Table 4).

As the definition of idiopathic short stature is vague, different clinician might catalog patient differs with each other. To avoid such ambiguity, we examined the cohort retrospectively following Dauber's algorithm (Fig. 3, modified from Dauber 2014, JCEM) [9]. 33 patients could be classified as idiopathic short stature (Supplementary Table 1). In this group, genetic background was identified in seven patients. The yield is 21.2% (7/33).

Comparison with the standard diagnostic algorithm

Dauber et al. [9] proposed a diagnostic algorithm to search for the genetic background of short stature (Fig. 3, modified from Dauber 2014, JCEM). This diagnostic algorithm counts on clinician's experience greatly, such as the first step "Does the patient have a distinct recognizable genetic syndrome?" Different clinicians might recognize different patients so that the patient might have went through totally different diagnose pathways depending on the specialist he visited. To minimize the effect caused by experience, we examined the 114 cases retrospectively entirely based on the molecular diagnosis result. The 41 diagnosed patients were allocated to the algorithm based on the clinical manifestation. 19 out of the

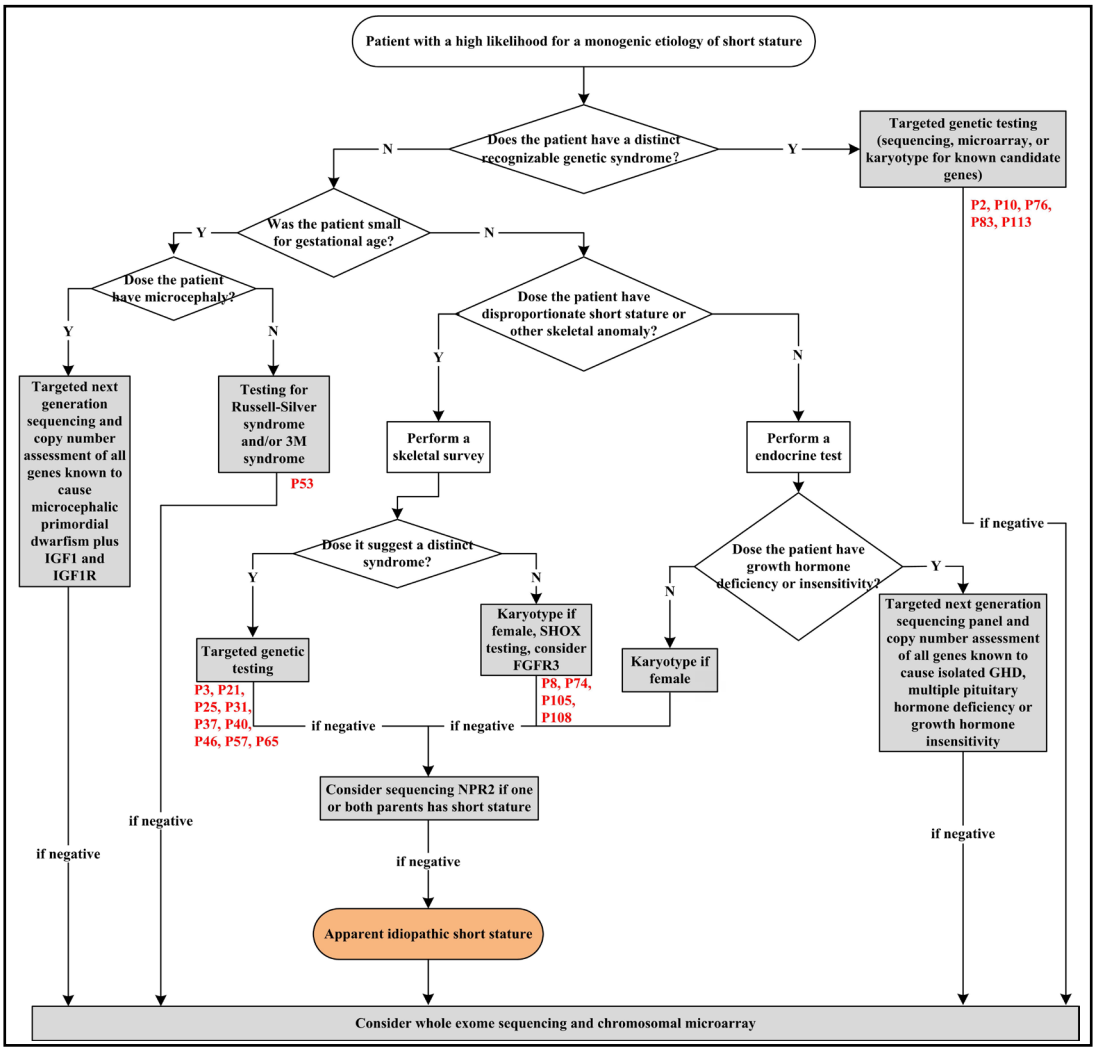

Fig. 3. Current algorithm for the genetic diagnosis of short stature modified from Dauber et al. (9). The diagnosed cases were allocated to the diagram retrospectively based on the etiology.

41 diagnosed patients could be tackled before the last high throughput method (Marked in Fig. 3). Five of the 19 patients would be diagnosed in the first step as they show distinct recognizable features. Majority (13/19) would be diagnosed for the skeletal anomaly, including four patients with FGFR3 defects. 95 (83.3%) patients would be studied by NGS/CMA in the last step eventually. Take into the consideration the heterogeneity of clinical features and possible false negative results of various tests, the number of patients that falls into the last step would be greater.

Discussion

Herein, we present genetic findings from 114 Chinese children with previously unexplained short stature. We identified causes in 41 patients through NGS and CMA, attaining a positive rate of 36.0%, which was comparable to that previously reported in a similar study (5/14, 35.7%) [21], and higher than that provided by WES in a large cohort of patients with suspected genetic conditions (504/2000, 25.2%) [22]. Our study confirmed the utility of high-throughput molecular detection techniques in the diagnosis of short stature with unknown etiology. In this study, more than 90% of solved patients were diagnosed by

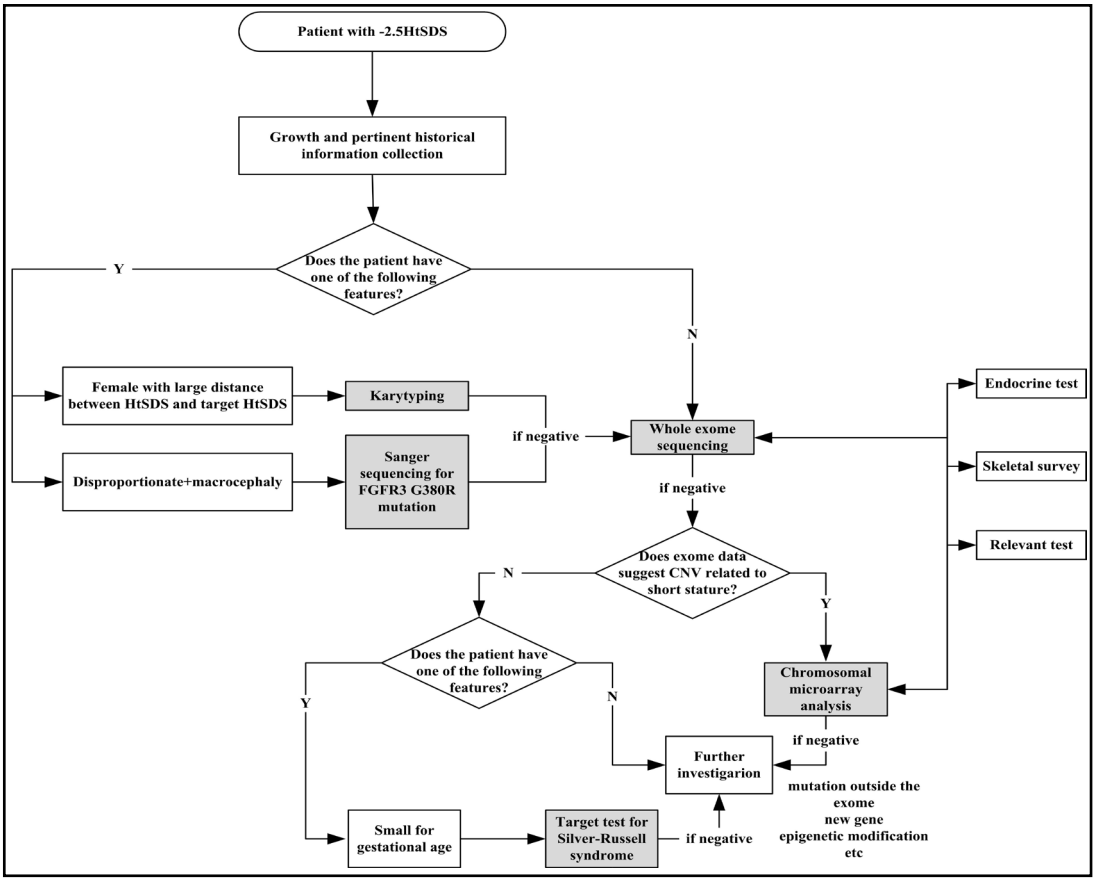

Fig. 4. The proposed diagnostic algorithm to identify the genetic etiology of short stature.

NGS, suggesting that NGS would be a preferred option for genetic evaluation of undiagnosed short stature.

It is proposed based on experience that the severity of short stature, SGA without catch up growth, presence of one or more additional clinical/biochemical abnormalities, presence of sibling or parent with similar features may indicate a genetic cause of short stature [9, 18]. To validate the notion, we adopted statistical analysis to test these factors in our cohort. Although we observed the diagnostic yields varied in different groups of short stature severity, with or without a given additional phenotype, and presence of one or more additional phenotypes, none was statistically different except for facial dysmorphism and skeletal abnormalities (Fig. 2B). That is, patients with short stature who has additional features including facial dysmorphism or skeletal abnormalities might have a potential genetic etiology. Therefore, we provided the first evidence that clinicians would be more likely to identify disease-causing variants in these subgroups. However, evidence in a larger cohort is required to confirm our conclusion. Additionally, no variant was detected in children with isolated short stature in this study (Table 1, Supplementary Table 1). Their pathogenic variants might be located in genes or regions not present in the current panel. Alternatively, they may be constitutional delay of growth and puberty and their final adult height might be normal, thus followup is needed of these individuals.

As compared with the standard diagnostic algorithm, even in the ideal scenario, ~80% patients would be examined by the high throughput methods-NGS/CMA. In this next generation era, NGS become affordable by diagnosis. In addition, new algorithm allows to call CNVs from NGS data. Indeed, we applied XHMM to call CNVs from the 114 datasets. CNVs of P42, P75 and P109 could be detected with the exception of P83 (data now shown). The negative CNV result of P83 is probably due to the size of deletion is too small. As previously

suggested "next generation sequencing demands next generation phenotyping" [23], we propose a new diagnostic algorithm (Fig. 4).

After the growth and pertinent historical information collection, the patients were differentiated of Turner syndrome and FGFR3 related disorders. The rest of the patients will be evaluated by panel/exome sequencing. The candidate mutation would guide the clinical evaluations such as the endocrine test, skeletal survey and other relevant test. The results would help to confirm the NGS data and target the genetic etiology. For the negative cases, if the NGS data suggest CNVs related to short stature, a chromosomal microarray would be ordered. Based on the phenotypes, the rest cases would be tested for Silver-Russell syndrome. This algorithm takes into consideration the techniques that would be used and simplify the standard one by moving the panel/exome sequencing forward. Nevertheless, NGS/CMA would not solve every case. The negative cases need further investigation.

In conclusion, our study confirmed the utility of high-throughput molecular detection techniques in diagnosis of short stature. 41 of 114 Chinese children with undiagnosed short stature were tackled by NGS and CMA. The diagnostic yield reaches 36.0% including 46 different variants in 29 genes and two pathogenic CNVs. Patients with facial dysmorphism and/or skeletal abnormalities had a significantly higher diagnostic rate than those without the corresponding phenotype, which suggested these two phenotypes might be applied as predictors for etiology of short stature by genetic testing, but more samples are required to further validate these findings. By comparison with the standard diagnostic algorithm, 83.3% patients would be evaluated by NGS/CMA. We proposed a simplified algorithm by moving NGS forward. Hopefully in the next generation era, it would help to improve the yield of the molecular diagnosis of short stature.

Acknowledgements

We are grateful to all the patients and their families who participated in this study. We also express our gratitude to the nursing staff for obtaining the blood samples.

Author contributions:

XFG and YGY designed and supervised the study. ZH and YS performed the data analysis, wrote the main manuscript and prepared the figures and tables. ZH, LLW, HLL, ZWG, JGW, HY, YW and GRH performed the experiments. YJF, LLW, HLL, ZWG, JGW, HY, YW, GRH and RFW participated in the data analysis. XFG, YGY, JY, LSH, WJQ, HWZ, LLL and YY provided patients' data and recruited the patients. AD modified the article using scientific English. All authors reviewed the manuscript.

This work was funded by National Natural Science Foundation of China (81670812, to YGY), and Shanghai Municipal Commission of Health and Family Planning Foundation (2017YQ020, to YS and 201740192, to YGY), Shanghai Jiao Tong University School of Medicine (2014XJ10044, to YS), Jiaotong University Cross Biomedical Engineering" (No. YG2017MS72, to YGY), Shanghai Shen Kang Hospital Development Center new frontier technology joint project" (No. SHDC12017109, to YGY).

Disclosure Statement

The author reports no conflicts of interest in this work.

References

1 Wit JM, Clayton PE, Rogol AD, Savage MO, Saenger PH, Cohen P: Idiopathic short stature: definition, epidemiology, and diagnostic evaluation. Growth Horm IGF Res 2008;18:89-110.

2 Siegel PT, Clopper R, Stabler B: Psychological impact of significantly short stature. Acta Paediatr Scand Suppl 1991;377:14-18; discussion 19.
3 McEvoy BP, Visscher PM: Genetics of human height. Econ Hum Biol 2009;7:294-306.
4 Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J et al.: Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46:1173-1186.
5 Durand C, Rappold GA: Height matters-from monogenic disorders to normal variation. Nat Rev Endocrinol 2013;9:171-177.
6 Wang SR, Carmichael H, Andrew SF, Miller TC, Moon JE, Derr MA, Hwa V, Hirschhorn JN, Dauber A: Large-scale pooled next-generation sequencing of 1077 genes to identify genetic causes of short stature. J Clin Endocrinol Metab 2013;98:E1428-1437.
7 Zahnleiter D, Uebe S, Ekici AB, Hoyer J, Wiesener A, Wieczorek D, Kunstmann E, Reis A, Doerr HG, Rauch A, Thiel CT: Rare copy number variants are a common cause of short stature. PLoS Genet 2013;9:e1003365.
8 Hu G, Fan Y, Wang L, Yao RE, Huang X, Shen Y, Yu Y, Gu X: Copy number variations in 119 Chinese children with idiopathic short stature identified by the custom genome-wide microarray. Mol Cytogenet 2016;9:16.
9 Dauber A, Rosenfeld RG, Hirschhorn JN: Genetic evaluation of short stature. J Clin Endocrinol Metab 2014;99:3080-3092.
10 Wit JM, Kiess W, Mullis P: Genetic evaluation of short stature. Best Pract Res Clin Endocrinol Metab 2011;25:1-17.
11 Li H, Ji CY, Zong XN, Zhang YQ: [Height and weight standardized growth charts for Chinese children and adolescents aged 0 to 18 years]. Zhonghua Er Ke Za Zhi 2009;47:487-492.
12 Xu S, Gu X, Pan H, Zhu H, Gong F, Li Y, Xing Y: Reference ranges for serum IGF-1 and IGFBP-3 levels in Chinese children during childhood and adolescence. Endocr J 2010;57:221-228.
13 Greulich WW, Pyle SI: Radiographic atlas of skeletal development of the hand and wrist. The American Journal of the Medical Sciences 1959;238:393.
14 Sun Y, Hu G, Liu H, Zhang X, Huang Z, Yan H, Wang L, Fan Y, Gu X, Yu Y: Further delineation of the phenotype of truncating KMT2A mutations: The extended Wiedemann-Steiner syndrome. Am J Med Genet A 2017;173:510-514.
15 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL: 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. 2015;17:405-424.
16 Fan Y, Qiu W, Wang L, Gu X, Yu Y: Exonic deletions of AUTS2 in Chinese patients with developmental delay and intellectual disability. Am J Med Genet A 2016;170a:515-522.
17 Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST: American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 2011;13:680-685.
18 Wit JM, Oostdijk W, Losekoot M, van Duyvenvoorde HA, Ruivenkamp CA, Kant SG: MECHANISMS IN ENDOCRINOLOGY: Novel genetic causes of short stature. Eur J Endocrinol 2016;174:R145-173.
19 Bonafe L, Cormier-Daire V, Hall C, Lachman R, Mortier G, Mundlos S, Nishimura G, Sangiorgi L, Savarirayan R, Sillence D, Spranger J, Superti-Furga A, Warman M, Unger S: Nosology and classification of genetic skeletal disorders: 2015 revision. Am J Med Genet A 2015;167a:2869-2892.
20 Baron J, Savendahl L, De Luca F, Dauber A, Phillip M, Wit JM, Nilsson O: Short and tall stature: a new paradigm emerges. Nat Rev Endocrinol 2015;11:735-746.
21 Guo MH, Shen Y, Walvoord EC, Miller TC, Moon JE, Hirschhorn JN, Dauber A: Whole exome sequencing to identify genetic causes of short stature. Horm Res Paediatr 2014;82:44-52.
22 Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM: Molecular findings among patients referred for clinical whole-exome sequencing. Jama 2014;312:1870-1879.
23 Hennekam RC, Biesecker LG: Next-generation sequencing demands next-generation phenotyping. Hum Mutat 2012;33:884-886.