The Application of Clinical Genetics

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Genetic Polymorphisms and QF-PCR Performance Evaluation of 20 Autosomal STR Loci on Chromosomes 13, 18, and 21 in Prenatal Diagnosis Among East Chinese Han Population

Yingwen Liu[1,2], Jiangyang Xue[1,2], Lulu Yan[1,2], Changshui Chen[2], Shumin Zhao[3], Haibo Li[1,2]

[1]The Central Laboratory of Birth Defects Prevention and Control, The Affiliated Women and Children's Hospital of Ningbo University, Ningbo, Zhejiang, 315000, People's Republic of China; [2]Ningbo Key Laboratory for the Prevention and Treatment of Embryogenic Diseases, The Affiliated Women and Children's Hospital of Ningbo University, Ningbo, Zhejiang, 315000, People's Republic of China; [3]School of Life Science, Fudan University, Shanghai, 200000, People's Republic of China

Correspondence: Haibo Li, The Central Laboratory of Birth Defects Prevention and Control, Women and Children's Hospital of Ningbo University, No. 339, Liuting Road, Haishu District, Ningbo, Zhejiang, People's Republic of China, Email lihaibo-775@163.com; Shumin Zhao, School of Life Science, Fudan University, Shanghai, No. 2005, Songhu Road, Shanghai, People's Republic of China, Email shumin.zhao@cubicise.com

Objective: To evaluate the system performance of the 20 autosomal short tandem repeats (STR) polymorphic genetic loci in quality control for prenatal diagnosis.

Methods: A genotyping system consisting of 6 STRs on chromosome 13 (chr13), 6 STRs on chromosome 18 (chr18), 8 STRs on chromosome 21 (chr21), and 10 genetic markers on sex chromosomes were used to analyze the genetic profiles of 2333 unrelated adult females from the Han population in East China. The population allele frequencies of the 20 autosomal STRs were obtained using the genotype dataset of the cohort. The established method in forensic genetic fields was used to calculate allele frequencies of the 20 autosomal STRs, observed heterozygosity (H_{obs}), expected heterozygosity (H), random match probability (PM), power of non-parental exclusion in duos (PE_{duos}), cumulative random match probability (CPM), and cumulative non-parental exclusion rate in duos (CPE_{duos}). The possible influence of STR markers on the detection rates of heterozygotes was assessed by employing the binomial distribution approach for every autosomal chromosome.

Results: The average expected heterozygosity for the STRs on chr13, chr18, and chr21 is 0.8097, 0.7478, and 0.7760, respectively. The CPM for the STRs on the three chromosomes is 4.52E-08, 7.34E-07, and 9.30E-10, respectively. The probability of detecting at least one heterozygosity on each chromosome is 0.999952, 0.999742, and 0.999994, respectively. The CPE_{duos} of the 20 STRs is 0.999982. And the potential linkage effects among the STRs on the autosomal have a negligible impact on the observed heterozygosity.

Conclusion: The 20 autosomal STRs in the 30 plex genotyping system exhibit high polymorphism in the Han population from East China, effective meeting the quality control criteria and providing valuable guidelines in assessing the molecular karyotypes of the chromosomes for prenatal diagnosis.

Keywords: expected heterozygosity, polymorphic genetic marker, QF-PCR, random match probability, prenatal diagnosis

Introduction

Short tandem repeats (STRs) are widely distributed throughout the human genome, highly polymorphic, and easy to genotype using the combined fluorescence multiplex PCR and capillary electrophoresis technique (also known as QF-PCR). Therefore, STRs are widely used in forensic individual identification, paternity testing, and population genetics research fields.[1–5] The STRs genotyping method based on QF-PCR is also used for quality control tasks such as sample identification, determination of the origin of fetal DNA, and detection of maternal cell contamination in prenatal diagnostics.[6,7] These quality control processes are necessary prerequisites to ensure the accuracy of prenatal diagnosis

Received: 12 March 2025
Accepted: 4 June 2025
Published: 12 June 2025

results. Additionally, based on the genotype and peak pattern of STRs, the molecular karyotype of the chromosome can be conveniently indicated.[8,9] However, when STRs are used for testing the molecular karyotype of target chromosomes or quality control of prenatal samples, the test performance is closely related to key system performance parameters, such as population allele frequencies and random match probability. In our routine sample quality control process for prenatal diagnosis, we utilized a multiplex genotyping system including 20 STRs located on chr13, chr18, and chr21 to test 2333 mother-fetus sample pairs. Based on the genotype dataset, we comprehensively evaluated the system performance of the 20 autosomal STRs in sample discrimination and kinship determination of the mother-fetus pair, which are critical aspects of sample quality control in prenatal diagnosis.

This is the first study to evaluate the application values of a genotyping system in prenatal diagnosis from a population genetics perspective among East Chinese Han population. The large-scale sample data significantly improved reliability in detecting maternal cell contamination and increased diagnostic confidence in rapid aneuploidy detection. Although the genotyping system currently covers only three autosomes (chr13, chr18, and chr21), it has demonstrated complete suitability for detecting common aneuploidies in prenatal diagnosis among East Chinese Han population.

Materials and Methods

Sample and Informed Consent

The study cohort comprised 2333 Chinese Han pregnant women who underwent prenatal amniocentesis or chorionic villus sampling at Ningbo Women and Children's Hospital between November 2021 and April 2023. All subjects met the clinical indications required for prenatal diagnosis, peripheral blood samples were collected from the pregnant women using EDTA as an anticoagulant. All samples underwent genotyping analysis for quality control, followed by aneuploidy diagnosis for the prenatal samples. All participants provided informed consent for genetic analysis and future data use. The protocol was approved by the Institutional Review Board of The Affiliated Women and Children's Hospital of Ningbo University (Approval No. EC2020-048). The study complies with the Declaration of Helsinki.

DNA Extraction

We collected amniotic fluid or chorionic villi from each pregnant women using the standard amniocentesis technique for QF-PCR. Following centrifugation, fetal cells were isolated. Genomic DNA was extracted from peripheral blood samples collected with EDTA anticoagulant using the QIAGEN QIAamp DNA Mini Kit (QIAGEN, Germany). The extracted genomic DNA samples underwent quality testing using the NanoDrop 2000 (Thermo Fisher Scientific, USA). DNA samples that passed the quality assessment were stored at −20°C for subsequent genotyping analysis.

Multiplex PCR Amplification

Multiplex PCR amplification was performed using the AneuFiler^{TM} kit (Approved by National Medical Products Administration of China with No.20233401830; Shanghai Cubicise Medical Co., Ltd., China), which was a clinical diagnosis assay for common aneuploidies of chr13, chr18, chr21 and sex chromosomes based QF-PCR. The PCR amplification system was prepared according to the product manual, and amplification was carried out using the Veriti^{TM} PCR System (Thermo Fisher Scientific, USA) under the provided PCR conditions. The multiplex system is a single-tube 30 plex PCR amplification system, including 6 STRs on chr13(*D13S1822, D13S317, D13S892, D13S788, D13S1817, D13S325*), 6 STRs on chr18(*D18S862, D18S878, D18S391, D18S867, D18S865, D18S1371*), 8 STRs on chr21 (*D21S1437, D21S2052, D21S1446, D21S1444, D21S1432, D21S1414, D21S1411, PentD*), 1 sex locus (*ZFXY*) on chrX and chrY, 5 loci (*IDXp, IDXq, DXS7132, DXS9902, GATA165B12*) on chrX, 1 locus on chrY(*SRY*), 1 quality control marker (*TAF9b*) on chr3 and chrX, 1 quality control marker (*CM11X*) on chr11 and chrX, and 1 quality control marker (*CM01Y*) on chr1 and chrY. The complete marker specifications for the AneuFiler^{TM} kit are provided in the Supplementary Table 1.

Capillary Electrophoresis

Capillary electrophoresis was performed using the 3500Dx Genetic Analyzer with a 50 cm capillary and POP-7™ polymer, all of which are products of Thermo Fisher Scientific (USA). According to the instructions of the AneuFiler™ kit, 1.0 μL of PCR amplification products was taken to prepare the capillary electrophoresis system. CM-LIZ500 was used as the internal DNA fragment size standard in the capillary electrophoresis (Shanghai Cubicise Medical Co., Ltd., China). The capillary electrophoresis parameters were configured according to the instructions of the AneuFiler™ kit.

Capillary Electrophoresis Data Analysis

Capillary electrophoresis raw data were analyzed using the GeneMapper v3.0 software package (Thermo Fisher Scientific, USA). The Panel file and Bin file of the AneuFiler™ kit were provided by the supplier to establish the analysis methods. Following the instructions of the AneuFiler™ kit, capillary electrophoresis profiles of each sample were obtained. The genotype data of the 20 autosomal STRs for each sample were exported into a Microsoft Excel file based on the instructions of the GeneMapper v3.0 software for subsequent statistical analysis. Peak height ratios between 0.80 and 1.50 were considered normal, whereas ratios between 0.45 and 0.65 or between 1.80 and 2.40 were interpreted as trisomy. Three alleles of equal peak area were also considered trisomy. The presence of a single peak was usually considered uninformative. The DNA profiling of normal diploid female sample, normal diploid male sample, and male trisomy21 are provided in the Supplementary Figures 1–3, respectively.

Population Genetic Analysis

The population allele frequencies (*f*) of the 20 STRs in the study group were obtained using the STRAF v2.1.5 online analysis package.[10,11] The Hardy-Weinberg equilibrium test was conducted, and the linkage disequilibrium between pairs of STRs on the same chromosome was also assessed using the STRAF v2.1.5. Additionally, a principal component analysis (PCA) was performed to investigate the presence of substructures within the study population.[11]

System Performance Parameters

According to the guidelines for QF-PCR application in aneuploidies testing in prenatal diagnosis, the number of heterozygous loci (x) among m examined STRs ($m \geq 4$) on the same chromosome is crucial for determining the ploidy of the autosomal chromosome.[8,9] If x equals zero, indicating that all of the m STRs are homozygous, it becomes difficult to determine the ploidy of the chromosome. The observed heterozygosity (H_{obs}) of the 20 autosomal STRs was calculated using a counting method with Microsoft Excel. The expected heterozygosity (H), random match probability (PM), individual discrimination power (DP), power of non-maternal or non-paternal exclusion in duos (PE_{duos}), as well as the cumulative random match probability (CPM), total discrimination power (TDP), and cumulative power of non-parental exclusion in duos (CPE_{duos}) were computed based on the population allele frequencies of STR using formulas from references.[10,12–14]

Impact of Linkage Effects on the Detection Rate of Heterozygosity

The potential linkage effects of pairwise STRs on the same autosomal chromosome were evaluated using the Binomial distribution function. The actual sample proportion, $P(x \leq k)$ (k = 0, 1, 2,., m), in the study population was obtained with direct counting method. When the potential linkage effects of pairwise STRs on the same autosomal chromosome could be ignored, meaning that the genotyped STRs on the same autosome chromosome could be treated independently, the expected sample proportion, $P(x \leq k)$ (k = 0, 1, 2,., m), could be obtained using the Binomial distribution function with H as the population rate.[15] Here, H represented the average expected heterozygosity of m STRs genotyped on the same autosomal chromosome. Nonlinear fitting for the actual sample proportion distribution of $P(x \leq k)$, denoted as *P-true*, and the expected sample proportion distribution of $P(x \leq k)$, denoted as *P-Emulation*, was performed using GraphPad Prism v8.0 software.

Result

Population Allele Frequencies of the 20 Autosomal STRs

The population allele frequencies of the 20 STRs genotyped on chr13, chr18 and chr21 are presented in Tables 1–3, respectively. And 20 autosomal STR genotypes of the 2333 mother-child sample pairs are shown in Supplementary Table 2.

Table 1 Population Allele Frequencies of the 6 STRs on chr13 (2N = 4666)

D13S325		*D13S788*		*D13S1817*		*D13S317*		*D13S892*		*D13S1822*	
allele	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i
13	0.0002	14	0.0002	6	0.0019	7	0.0015	9	0.0967	15	0.0047
14	0.0000	14.2	0.0201	7	0.2312	8	0.2818	10	0.3236	16	0.0129
15	0.0011	15	0.0015	8	0.0189	9	0.1462	10.1	0.0002	17	0.0459
16	0.0011	15.2	0.0009	9	0.3009	10	0.1474	11	0.0956	18	0.1417
17	0.0075	16	0.0006	9.2	0.0060	11	0.2321	12	0.1640	19	0.2426
18	0.0414	16.2	0.0744	10	0.2002	12	0.1451	13	0.2510	20	0.2415
19	0.2390	17	0.0654	10.2	0.0150	13	0.0373	14	0.0667	21	0.2152
20	0.2760	17.2	0.0077	11	0.1764	14	0.0081	15	0.0024	22	0.0808
21	0.2210	18	0.0913	11.2	0.0051	15	0.0004	/	/	23	0.0124
22	0.1414	18.2	0.0223	12	0.0403	/	/	/	/	24	0.0024
23	0.0482	19	0.0084	12.2	0.0002	/	/	/	/	/	/
24	0.0189	19.2	0.0973	13	0.0039	/	/	/	/	/	/
25	0.0021	20	0.1155	/	/	/	/	/	/	/	/
26	0.0017	20.2	0.1989	/	/	/	/	/	/	/	/
27	0.0004	21	0.1031	/	/	/	/	/	/	/	/
/	/	21.2	0.1256	/	/	/	/	/	/	/	/
/	/	22	0.0300	/	/	/	/	/	/	/	/
/	/	22.2	0.0285	/	/	/	/	/	/	/	/
/	/	23	0.0034	/	/	/	/	/	/	/	/
/	/	23.2	0.0049	/	/	/	/	/	/	/	/
/	/	24	0.0002	/	/	/	/	/	/	/	/
total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000

Notes:D13S325, etc. names of corresponding STRs. f_i represents the allele frequencies of the loci;N represents the sample size.

Table 2 Population Allele Frequencies of the 6 STRs on chr18 (2N = 4666)

D18S391		*D18S865*		*D18S867*		*D18S862*		*D18S878*		*D18S1371*	
allele	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i
7	0.0036	6	0.0004	9	0.0026	8	0.0006	9	0.0004	9	0.0291
8	0.0077	7	0.0122	10	0.0904	9	0.0000	10	0.0011	10	0.0064
9	0.2737	8	0.1972	10.2	0.0002	10	0.0109	11	0.0343	10.2	0.0002
10	0.5011	9	0.1513	11	0.3800	11	0.2347	12	0.1813	11	0.0821
11	0.1702	10	0.2675	12	0.3138	12	0.1072	13	0.1987	11.2	0.0030
12	0.0422	10.3	0.0002	12.2	0.0015	13	0.1584	14	0.1987	12	0.3965
13	0.0013	11	0.3210	13	0.1399	14	0.1314	15	0.1903	12.2	0.0036
14	0.0002	12	0.0478	14	0.0615	15	0.1612	16	0.0461	13	0.3875
/	/	13	0.0021	15	0.0094	16	0.1556	17	0.0859	13.2	0.0011
/	/	14	0.0002	16	0.0006	17	0.0362	18	0.0495	14	0.0840
/	/	/	/	/	/	18	0.0034	19	0.0116	14.2	0.0002
/	/	/	/	/	/	19	0.0004	20	0.0019	15	0.0060
/	/	/	/	/	/	/	/	/	/	16	0.0002
total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000

Notes:*D18S391*, etc. names of corresponding STRs. f_i represents the allele frequencies of the loci;N represents the sample size.

https://doi.org/10.2147/TACG.S521043

Table 3 Population Allele Frequencies of the 8 STRs on chr21 (2N = 4666)

D21S1432		*D21S1414*		*D21S1437*		*D21S2052*		*D21S1444*		*D21S1411*		*Penta D*		*D21S1446*	
allele	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i	**allele**	f_i
11	0.0006	26	0.0002	7	0.0002	16	0.0004	22	0.0002	34	0.0004	5	0.0002	10	0.0021
12	0.0405	27	0.0028	9	0.0069	17	0.0049	22.2	0.0013	35	0.0019	6	0.0034	11	0.0079
13	0.0396	27.2	0.0002	10	0.1012	18	0.0249	23.2	0.0231	36	0.0034	7	0.0058	12	0.5212
14	0.2859	28	0.0463	11	0.1472	19	0.0894	24	0.0006	37	0.0261	8	0.0482	12.2	0.0002
15	0.3513	28.2	0.0069	12	0.0137	20	0.1252	24.2	0.0148	38	0.1267	9	0.3283	13	0.1635
16	0.2090	29	0.2640	13	0.0384	21	0.2649	25	0.0214	39	0.2265	10	0.1237	13.2	0.0253
17	0.0692	29.2	0.0015	14	0.4895	22	0.2848	25.2	0.0643	40	0.2617	11	0.1374	14	0.0394
18	0.0039	30	0.0240	15	0.1316	23	0.1020	26	0.0697	41	0.1950	12	0.1766	14.2	0.0214
/	/	30.2	0.2568	16	0.0602	24	0.0647	26.1	0.0002	42	0.0917	13	0.1269	15	0.0133
/	/	31	0.0227	17	0.0107	25	0.0279	26.2	0.2608	43	0.0454	14	0.0396	15.2	0.1828
/	/	31.2	0.1027	18	0.0004	26	0.0088	27	0.0381	44	0.0152	15	0.0084	16	0.0009
/	/	32	0.0682	/	/	27	0.0021	27.1	0.0011	45	0.0039	16	0.0015	16.2	0.0204
/	/	32.2	0.0420	/	/	/	/	27.2	0.1513	46	0.0019	/	/	17.2	0.0015
/	/	33	0.1065	/	/	/	/	28	0.0086	/	/	/	/	/	/
/	/	33.2	0.0062	/	/	/	/	28.1	0.0122	/	/	/	/	/	/
/	/	34	0.0429	/	/	/	/	28.2	0.2454	/	/	/	/	/	/
/	/	34.2	0.0013	/	/	/	/	29	0.0009	/	/	/	/	/	/
/	/	35	0.0041	/	/	/	/	29.1	0.0004	/	/	/	/	/	/
/	/	35.2	0.0002	/	/	/	/	29.2	0.0669	/	/	/	/	/	/
/	/	36	0.0006	/	/	/	/	30.2	0.0137	/	/	/	/	/	/
/	/	/	/	/	/	/	/	31.2	0.0015	/	/	/	/	/	/
/	/	/	/	/	/	/	/	32.2	0.0030	/	/	/	/	/	/
/	/	/	/	/	/	/	/	33.2	0.0004	/	/	/	/	/	/
total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000	total	1.0000

Notes:*D21S1432*, etc. names of corresponding STRs. f_i represents the allele frequencies of the loci;N represents the sample size.

A total of 250 alleles were observed from the 20 autosomal STRs, and 8–23 alleles were observed for each locus with the corresponding allele frequencies spanned from 0.0002 to 0.5011. The loci with the highest number of alleles were *D13S325* on chr13 (15 alleles), *D18S1371* on chr18 (13 alleles), and *D21S1444* on chr21 (23 alleles).

Population Genetic Statistic

The observed heterozygosity and expected heterozygosity values spanned from 0.6438(*D18S391*) to 0.8757(*D13S788*) and from 0.6433(*D18S391*) to 0.8904(*D13S788*). Forensic statistical parameters were calculated and reported in Table 4, the *PM, DP*, and PE_{duo} were observed to range from 0.0218(*D13S788*) to 0.1852(*D18S391*), 0.8148(*D18S391*) to 0.9782 (*D13S788*), and 0.2241(*D18S391*) to 0.6380(*D13S788*), respectively. Based on the Hardy-Weinberg genetic equilibrium test, except for *D21S1414* and *D21S1444*, all other 18 STRs reached genetic equilibrium (Table 4, after Bonferroni adjusted, $P > 0.0025$). The linkage disequilibrium test for STRs on the same chromosome suggests that closely linkage was not observed (Table 4). PCA aims to reduce dimensionality by transforming multiple variables into a smaller set of composite indicators, extracting the most important characteristics from data. This analytical method projects the data onto a new coordinate system, where the greatest variance in the data lies on the first coordinate (called the first principal component, PC1), the second greatest variance on the second coordinate (PC2). In Figure 1, the X-axis represents the first principal component (PC1), and the Y-axis represents the second principal component (PC2). PCA analysis showed that the PC1 and PC2 interpreted 0.81% genetic variance and 0.8% genetic variance, respectively. Most of the Han populations were obviously cluster tightly within an elliptical strip in the PCA plot, indicating the study population does not show population stratification in the study samples. The system performance parameters of the multiplex genotyping system composed of the 20 autosomal STRs were presented in Table 5. The probability of at least one heterozygosity among STRs on chr13, chr18, and chr21 is 0.999953, 0.999743, 0.999994, respectively. The probability of all STRs being homozygous on chr13, chr18, and chr21 is 4.09E-05, 0.000195, 4.89E-06, respectively. On chr21, the *CPM* is the

Table 4 Performance Parameters for Each of the 20 Autosomal STRs (N = 2333)

Locus	*HET*	N-allele	*P* value	H_{obs}	*H*	*PM*	*DP*	PE_{duo}
D13S325	1890	14	0.7342	0.8101	0.7935	0.0735	0.9265	0.4174
D13S788	2043	21	0.2143	0.8757	0.8904	0.0218	0.9782	0.6380
D13S1817	1835	12	0.9574	0.7865	0.7826	0.0810	0.9190	0.3965
D13S317	1860	9	0.8034	0.7973	0.8012	0.0685	0.9315	0.4291
D13S892	1843	8	0.7287	0.7900	0.7826	0.0788	0.9212	0.4016
D13S1822	1871	10	0.4651	0.8020	0.8076	0.0647	0.9353	0.4428
D18S391	1502	8	0.9956	0.6438	0.6433	0.1852	0.8148	0.2241
D18S865	1781	10	0.7575	0.7634	0.7613	0.0963	0.9038	0.3593
D18S867	1667	10	0.1452	0.7145	0.7256	0.1197	0.8803	0.3176
D18S862	1937	11	0.3399	0.8303	0.8396	0.0462	0.9538	0.5068
D18S878	1972	12	0.9335	0.8453	0.8388	0.0465	0.9535	0.5078
D18S1371	1575	13	0.0427	0.6751	0.6780	0.1602	0.8398	0.2638
D21S1432	1708	8	0.8177	0.7321	0.7433	0.1081	0.8919	0.3383
D21S1414	1905	20	0.0010	0.8165	0.8310	0.0477	0.9523	0.5051
D21S1437	1634	11	0.0890	0.7004	0.7058	0.1149	0.8851	0.3174
D21S2052	1872	12	0.1288	0.8024	0.8091	0.0611	0.9389	0.4559
D21S1444	2009	23	0.0010	0.8611	0.8324	0.0474	0.9526	0.5069
D21S1411	1906	13	0.4826	0.8170	0.8148	0.0596	0.9404	0.4597
Penta D	1872	12	0.5881	0.8024	0.8069	0.0613	0.9387	0.4517
D21S1446	1542	13	0.8457	0.6614	0.6650	0.1490	0.8510	0.2659

Notes: HET: the number of samples with heterozygous genotype of the STR; N-allele: the number of detected alleles of the STR; P value: the significance in Hardy-Weinberg equilibrium test (after Bonferroni adjustment, $\alpha' = 0.05/20 = 0.0025$). The gray background of this column indicates the two STRs did not reach genetic equilibrium in the study population after adjustment.

Abbreviations: DP, individual discrimination power; H, expected heterozygosity; H_{obs}, observed heterozygosity; PE_{duos}, power of non-maternal or non-paternal exclusion in duos; *PM*, random matching probability.

lowest, while the *TDP* is the highest. The CPE_{duos} for 20 STRs is 0.999982. These findings indicated that 20 autosomal STR loci are highly polymorphic and informative in the East Han population and can be used as a powerful tool in personal identification and parentage testing.

Impact of Linkage Effects on the Detection Rate of Heterozygosity

In the study cohort, the actual and expected sample proportion with x heterozygotes, $P(x = k)$ ($k = 0, 1, 2,., m$), on chr13, chr18, and chr21 were presented in Table 6. Based on the results in Table 6, *P-true*, the actual sample proportion with $P(x \leq k)$, and *P-Emulation*, the expected sample proportion with $P(x \leq k)$ could be obtained. And the nonlinear fitting curves of *P-true* and *P-Emulation* were depicted in Figure 2. For the three detected autosomal chromosomes, the two curves completely overlap, indicated that the linkage effect between STR loci on the same chromosome has a negligible impact on the probability of detecting heterozygotes.

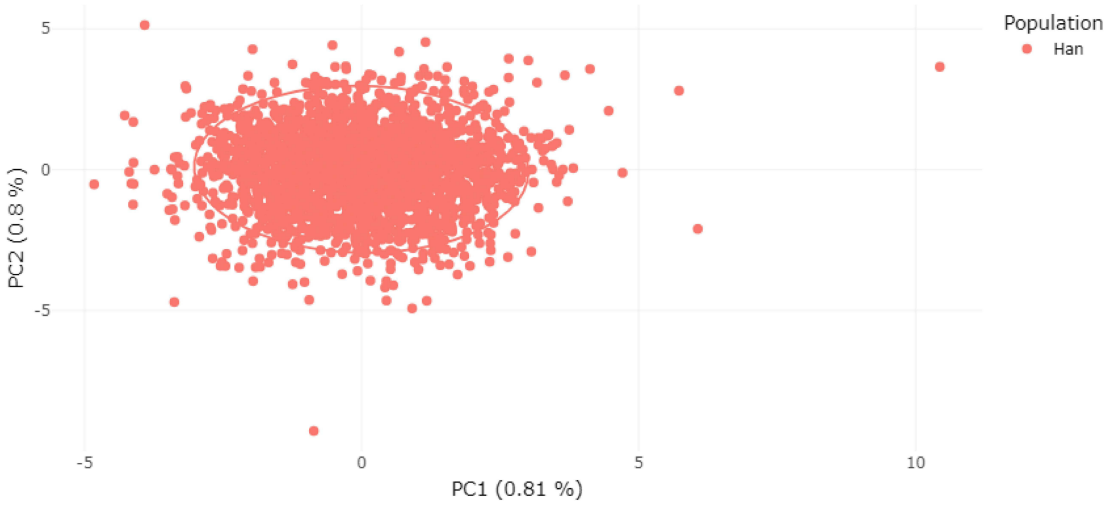

Figure 1 Principal component analysis (PCA) results based on the 20 autosomal STRs genotype dataset. The X-axis represents the first principal component (PC1), and the Y-axis represents the second principal component (PC2). PCA analysis showed that the PC1 and PC2 interpreted 0.81% genetic variance and 0.8% genetic variance, respectively. Most of the Han populations were obviously cluster tightly within an elliptical strip in the PCA plot, indicating the study population does not show population stratification in the study samples.

Discussion

The QF-PCR technology based on STRs has been widely used in quality control of prenatal diagnosis and rapid testing of common chromosomal aneuploidies.[8,9] In 2007, the Association for Clinical Genomic Science (ACGS) published a guidance document on the use of QF-PCR for rapid detection of common aneuploidies, which was updated in 2018 (referred to as international guidelines).[9] In 2016, China also released the "Expert Consensus on the Application of Quantitative Fluorescent PCR Technology in Prenatal Diagnosis" (referred to as the Chinese Expert Consensus).[16] Whether the long-term practice of STRs in forensic genetics,[1,5] or the international guidelines[9] and Chinese expert consensus,[16] all indicate that the number of STRs used and their heterozygosity are critical factors in determining the performance of the corresponding QF-PCR multiplex genotyping system. After obtaining the population allele frequencies of the STRs from a representative population, performance parameters of the STRs, such as the random match probability and power of non-maternal or non-paternal exclusion in duos, etc., can be calculated using the established method in the field of forensic genetics.[5,12,15] These system performance parameters are key indicators for assessing the quality control capability of STRs in prenatal diagnosis. Through these parameters, we can determine whether a multiplex genotyping system is able to differentiate different prenatal samples and accurately determine the kinship between the mother and the fetus. A multiplex genotyping system with a lower *CPM* and higher CPE_{duos} is important to

Table 5 Performance Parameters for the 20 STRs Genotyping System

System Performance Parameters	chr13	chr18	chr21
Number of detected STRs (*m*) / chr	6	6	8
Average of observed heterozygosity	0.8103	0.7454	0.7742
Average of expected heterozygosity	0.8097	0.7478	0.7760
Number of expected heterozygous marker/ person /chr	4.9	4.5	6.2
Probability of *m* STRs being homozygous	4.09E-05	0.000195	4.89E-06
The probability of at least one heterozygosity among *m* STRs	0.999953	0.999743	0.999994
CPM	4.52E-08	7.34E-07	9.32E-10
TDP(1-CPM)	0.999999955	0.999999266	0.999999999
CPE_{duos}	0.999982		

Abbreviations: CPE_{duos}, cumulative non-parental exclusion rate in duos; CPM, cumulative random match probability; TDP, total discrimination power.

https://doi.org/10.2147/TACG.S521043

Table 6 Sample Proportion with $P(x = k)$ (k = 0, 1, 2,., m) on the 3 Chromosomes (N = 2333)

x	chr13 (*m* = 6)		chr18 (*m* = 6)		chr21 (*m* = 8)	
	P-true	*P-Emulation*	*P-true*	*P-true*	*P-Emulation*	*P-true*
0	0.000857	4.76E-05	0.000000	0.000258	0.000000	6.33E-06
1	0.000857	0.001214	0.004715	0.004581	0.000429	0.000175
2	0.014575	0.012907	0.029147	0.033954	0.003858	0.002128
3	0.079297	0.073202	0.144449	0.134205	0.013716	0.014747
4	0.223746	0.233539	0.297900	0.298381	0.059580	0.063875
5	0.387055	0.397370	0.358337	0.353812	0.175311	0.177063
6	0.294042	0.281721	0.165452	0.174809	0.323618	0.306764
7	/	/	/	/	0.300471	0.303699
8	/	/	/	/	0.123018	0.131541
Total	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
≥1	0.999143	0.999953	1.000000	0.999742	1.000000	0.999994

Notes: P-true: sample proportion with P(x = k) (k = 0, 1, 2,., m) obtained by directly counting method. P-Emulation: sample proportion with P(x = k) (k = 0, 1, 2,., m) obtained with the binomial distribution function Bx (H, m). Here, H is the Average of expected heterozygosity in Table 5.

eliminate the possibility of sample confusion in prenatal diagnosis procedure. Additionally, the allele frequency of STRs can generally be used to estimate the capability of a multiplex genotyping system for testing the chromosomal molecular karyotype. When multiple STRs on a chromosome are genotyped, this capability is closely related to the probability of at least one heterozygote being detected.

In this study, allele frequencies of 20 autosomal STRs on ch13, chr18, and chr21 were obtained from 2333 Chinese Han adult female individuals from East China that it is one of the first to systematically evaluate the performance of AneuFiler^{TM} multiplex genotyping system for prenatal diagnosis in a large East Chinese Han cohort. The Hardy-Weinberg equilibrium test showed that only two STRs, *D21S1414* and *D21S1444*, did not meet genetic equilibrium in the study population. Further analysis revealed that this may be due to the presence of multiple rare alleles for the two STRs. After excluding genotype data containing the rare alleles for the two STRs from the dataset, no significant deviation from genetic equilibrium was observed in the distribution of their genotypes. Therefore, the obtained

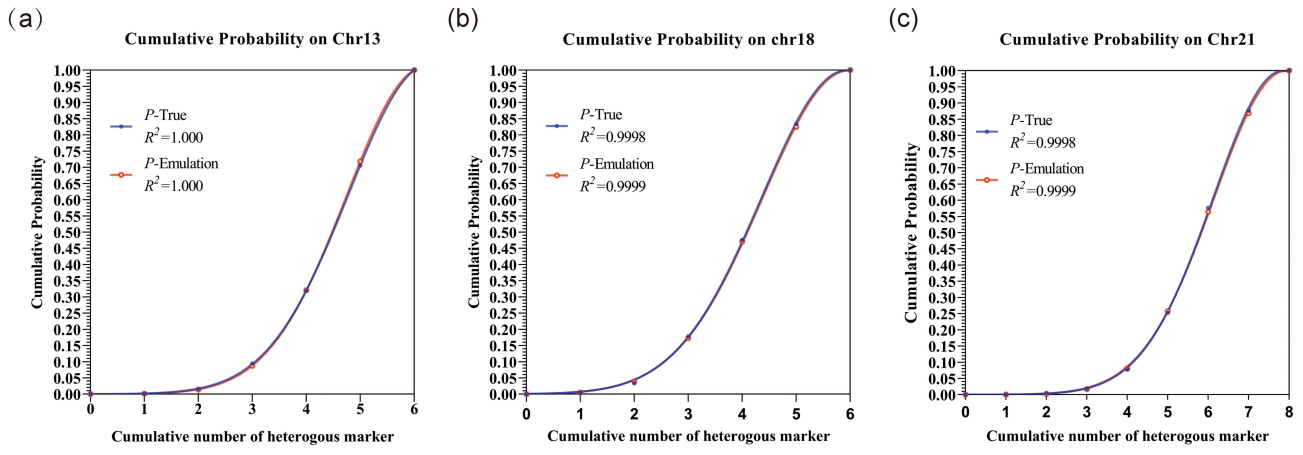

Figure 2 Nonlinear fitting curves of *P-true* and *P-Emulation* for the chromosome 13,18,21. The Y-axis stands for cumulative probability corresponding to the cumulative number of heterozygous markers, whereas the X-axis represents cumulative number of heterozygous markers. The blue curves represent for *P-True* which was the sample proportion with $P(x = k)$ ($k = 0, 1, 2,., m$) obtained by directly counting method, the red curves represents *P-Emulation* which was the sample proportion with $P(x = k)$ ($k = 0, 1, 2,., m$) obtained with the binomial distribution function *Bx* (*H, m*) and R^2 stands for coefficient of determination for non-linear fitting. The two curves completely overlap, indicated that the linkage effect between STR loci on the same chromosome has a negligible impact on the probability of detecting heterozygotes, since the curve of *P-Emulation* was obtained under the hypothesis that all the detected STRs on the same chromosome were completely independent. With the cumulative number of heterozygous markers rising, cumulative probability went up: (**a**) Chromosome 13, (**b**) Chromosome 18, (**c**) Chromosome 21.

https://doi.org/10.2147/TACG.S521043

population allele frequencies can be used for estimating relevant parameters in subsequent analysis.[12] Based on the dataset of 20 autosomal STRs, PCA analysis and linkage disequilibrium tests showed no significant substructure or linkage disequilibrium in the study population, supporting the removal of rare genotypes from the dataset.[12]

Based on the analysis of population genetics, Tables 4 and 5 present the key parameters, such as observed heterozygosity and expected heterozygosity, for the 20 autosomal STRs. The 20 autosomal STRs exhibit high levels of polymorphism within the study population, with average expected heterozygosity values of 0.8097, 0.7478, and 0.7760 for chr13, chr18, and chr21, respectively, meeting the requirements of STR loci with high heterozygosity in international guidelines.[9] The cumulative random match probability for these 20 STRs is 3.10E-23, indicating that their individual identification capability surpasses that of the 13 CODIS STRs and is comparable to that of the NDIS, which is an expanded panel of 20 core forensic STR loci.[17] The CPE_{duos} for the 20 autosomal STRs also meet the level of commonly used parentage testing STR genotyping tools,[15] meeting the requirements specified in the "Technical Specifications for Parentage Testing"[18] These results also indicated that the 20 STRs have reached the same level as commonly used forensic STR genotyping tools in prenatal diagnosis quality control, including differentiating different prenatal samples and kinship determining for the mother-fetus pair. Additionally, prenatal samples such as cord blood, amniotic fluid, and chorionic villus sampling are prone to contamination by maternal cells (maternal cell contamination, MCC) or mixing with other individual samples, which can interfere with prenatal diagnostic results.[16] Previous practices using forensic STR genotyping tools have demonstrated the capability of STRs in MCC testing. Theoretically, the 20 STRs utilized in this study also suitable for MCC testing. Moreover, since this multiplex genotyping system includes six or more STRs on chr13, chr18, and chr21, respectively, it enables rapid testing for these chromosomal aneuploidies while performing quality control in prenatal diagnostic procedures.

The profiles of STR, including genotype and peak-height patterns, are the main basis for determining the corresponding chromosome molecular karyotype.[8,9,16] It is evident that when all the tested STRs on a chromosome are homozygous (ie, with a heterozygote count of $x = 0$), it is difficult to determine the chromosome molecular karyotype. Based on the average expected heterozygosity of the STRs in this multiplex typing system, the probabilities of STRs on chr13, chr18, and chr21 being homozygous [$P(x = 0)$] is as low as 4.76E-05, 0.000258, and 6.33E-06, respectively. This indicates that the 20 autosomal STRs have ideal specificity in determining the corresponding chromosomes as disomy, achieving respective values of 0.999952, 0.999742, and 0.999994. However, the evaluation of this key parameter assumes that these STRs are independent of each other. This premise is the basis for the calculation with the multiplication principle of probabilities.[5,13] In traditional forensic multiplex STR genotyping systems, to ensure the independence of STR loci and facilitate the application of the calculation with the multiplication principle of probabilities, multiple STRs from the same chromosome are typically not included.[1] However, STR multiplex genotyping systems designed for prenatal aneuploidy testing typically require the inclusion of four or more genetic markers on the same chromosome.[16] To evaluate whether the potential linkage effects between different genetic markers on the same chromosome have a negligible impact on critical parameters such as *CPM* and *CPEduos*, this study assessed the nonlinear fitting of cumulative probability distributions of samples with varying heterozygosity counts in the population. The actual cumulative probability distribution of different heterozygote counts on each chromosome shown in Figure 2 does not significantly differ from the simulated cumulative probability distribution based on the average expected heterozygosity, indicating that the potential linkage effects of these STRs on the same chromosome can be ignored, which is associated with uniform distribution of STRs on each chromosome in the AneuFiler™ multiplex genotyping system. This finding is consistent with the results of the linkage disequilibrium test. Additionally, the genetic distances between loci on each chromosome may introduce potential linkage effects. Based on the genetic distances of autosomal chromosomes provided in the UCSC Hg19 database, the average genetic distance of approximately 10.0 cM or more between them (average genetic distance on chr13 is 18.4 cM, on chr18 is 24.1 cM, and on chr21 is 9.9 cM).[3] In the field of forensic genetics, extensive research has been conducted to assess the potential linkage effects between *vWA* and *D12S391*, two STRs located on the short arm of chr12 with a genetic distance of approximately 10.0 cM, on the calculation of the corresponding system performance parameters. Studies involving over 50,000 pedigrees have demonstrated that the linkage disequilibrium between *vWA* and *D12S391* has a negligible impact on the calculation of critical forensic parameters, including *CPM, CPEduos*, and Kinship Index.[3]

One of the limitations of this study could be the lack of the STR allele frequencies in other ethnic populations, which constrains the generalizability of our findings. And the study can also expand on other potential applications such as distinguishing between monozygotic and dizygotic twins, distinguishing surrogacy cases, and risk mitigation in high-throughput prenatal labs. Additionally, our study was primarily based on allele frequencies in typical diploid samples and did not assess the diagnostic performance of this genotyping system in abnormal cases such as aneuploidy or maternal cell contamination, nor did it evaluate locus mutation rates. These aspects remain part of our ongoing research. Nonetheless, our research provided a framework for clinical implementation, and we plan to enlarge our cohort in other populations to confirm the applicability of these findings.

Conclusions

In summary, our study indicates that the 20 autosomal STRs included in the AneuFiler^{TM} multiplex genotyping system exhibit highly polymorphic in East Chinese Han population and are independent on chr13, chr18 and chr21. The comprehensive system facilitated the implementation of quality control measures in clinical prenatal diagnosis, including sample identification, identification of maternal contamination, and kinship verification of prenatal maternal-fetal sample pairs, thereby enabling accurate determination of chromosomal ploidy. Moreover, this system demonstrates extended clinical utility in zygosity testing for twin pregnancies, uniparental disomy (UPD) detection, pre-NGS sample quality control and so on. Nevertheless, further analysis and research involving more diverse populations and aneuploidy cases are still needed and may increase the diversity of applicable populations.

Data Sharing Statement

The original contributions presented in the study are included in the article or Supplementary Materials.

Acknowledgment

We would want to extend our sincere appreciation to the pregnant women and their families for their invaluable participation in this research endeavor. The medical professionals of The Affiliated Women and Children's Hospital of Ningbo University are duly acknowledged for their provision of care to the pregnant women as well as for their diligent efforts in collecting and preserving samples.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by grants from: Social development public welfare Foundation of Ningbo (No.2022S035), Ningbo Science and Technology Project (NO.2023Z178), Ningbo Top Medical and Health Research Program (No.2022020405), Ningbo Sci-Tech Innovation YONGJIANG 2035 (No.2024Z221).

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Butler JM. Advanced topics in forensic DNA typing. *Methodology*. 2011.
2. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet*. 1998;63(3):861–869. doi:10.1086/302011
3. Kong A, Gudbjartsson DF, Sainz J, et al. A high-resolution recombination map of the human genome. *Nat Genet*. 2002;31(3):241–247. doi:10.1038/ng917

4. Dib C, Faure S, Fizames C, et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*. 1996;380 (6570):152–154. doi:10.1038/380152a0
5. Butler JM. Advanced Topics in Forensic DNA Typing: interpretation. 2014.
6. Prenatal Screening And Diagnosis Group Birth Defect Prevention And Control Professional Committee Chinese Preventive Medical Association, Prenatal Diagnosis Group Society Of Medical Genetics Chinese Medical Association JL. Guidelines for the application of chromosomal microarray analysis in prenatal diagnosis (2023). *Chin J Med Genet*. 2023;40(9):1051–1061.
7. Buchovecky CM, Nahum O, Levy B. Assessment of maternal cell contamination in prenatal samples by quantitative fluorescent PCR (QF-PCR). *Methods Mol Biol*. 2019;1885:117–127. doi:10.1007/978-1-4939-8889-1_8
8. Mann K, Ogilvie CM. QF-PCR: application, overview and review of the literature. *Prenat Diagn*. 2012;32(4):309–314. doi:10.1002/pd.2945
9. Mann K, Hamilton S, Evans J, Sibbring J, Dore J. ACGS best practice guidelines for use of Quantitative Fluorescence-PCR for the detection of aneuploidy v4. *Association for Clinical Genomic Science*; 2018. Available from: https://www.acgs.uk.com/media/11402/qf-pcr-bpg-2018.pdf.
10. *Specification of the Calculation Methodology of Forensic Parameters for Autosomal STR. SF/Z JD0105010*. China: Standards Press of China; 2018.
11. Gouy A, Zieger M. STRAF—A convenient online tool for STR data evaluation in forensic genetics. *Forensic Sci Int Genet*. 2017;30:148–151. doi:10.1016/j.fsigen.2017.07.007
12. Hd Z, Sm Z, Chen YX, Li CT. Formula derivation for the probability distribution of ibs score in unrelated individual pairs. *Fa yi xue za zhi*. 2018;34(4):370. doi:10.12116/j.issn.1004-5619.2018.04.005
13. Welch LA, Gill P, Phillips C, et al. European Network of Forensic Science Institutes (ENFSI): evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers. *Forensic Sci Int Genet*. 2012;6(6):819–826. doi:10.1016/j.fsigen.2012.03.005
14. Jamieson A, Taylor CS. Comparisons of three probability formulae for parentage exclusion. *Anim Genet*. 1997;28(6):397–400. doi:10.1111/j.1365-2052.1997.00186.x
15. Sm Z. kinship determination. In: Zhao SM, Li CT, editors. *Forefront of Forensic DNA Typing*. Beijing: Science; 2011:159–247.
16. Prenatal Diagnosis Cooperative Group on Quantitative Fluorescent Polymerase Chain Reaction. Expert concensus on quantitative fluorescent polymerase chain reaction in the prenatal diagnosis of aneuploidies. *Chin J Obstet Gynecol*. 2016;51(5):321–324. doi:10.3760/cma.j.issn.0529-567X.2016.05.001.
17. Moretti TR, Moreno LI, Smerick JB, et al. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. *Forensic Sci Int Genet*. 2016;25:175–181. doi:10.1016/j.fsigen.2016.07.022
18. *Specification of Parentage Testing. GB/T 37223-2018*. China: Standards Press of China; 2018.