

·论 著·

二代测序试剂盒 SNP 位点遗传学参数和对比

周 密¹, 张 科², 汪 军³

(1. 芜湖市公安局, 安徽 芜湖 241000; 2. 安徽省公安厅物证鉴定管理处, 安徽 合肥 230061; 3. 安徽工程大学计算机与信息学院, 安徽 芜湖 241000)

摘要: 目的 计算二代测序试剂盒 SNP 位点的遗传学参数, 与 STR 基因座进行对比, 以建立 SNP 和 STR 的系统效能换算比例。方法 对二代测序试剂盒 (ForenSeq™ DNA Signature Prep 试剂盒和 Precision ID Identity Panel 试剂盒) 共 101 个 SNP 位点进行 Hardy-Weinberg 平衡检验, 计算 SNP 位点在个体识别、标准三联体鉴定、二联体鉴定及双亲皆疑鉴定中的各项系统效能参数, 并与 STR 基因座进行对比。结果 除无基因型频率数据的 2 个位点外, 99 个 SNP 位点符合 Hardy-Weinberg 平衡检验 ($P>0.05$)。ForenSeq™ DNA Signature Prep 试剂盒 94 个 SNP 位点的累积个体识别率 (cumulative discrimination power, CDP) 为 $1-1.1521 \times 10^{-34}$, 累积三联体非父排除率 (cumulative probability of exclusion in trios, CPE_{trio}) 为 $1-4.4169 \times 10^{-8}$, 累积二联体非父排除率 (cumulative probability of exclusion in duos, CPE_{duo}) 为 $1-8.4837 \times 10^{-5}$, 双亲皆疑累积排除率 (cumulative probability of exclusion in alleged parents cases, CPE_{AP}) 为 $1-1.2227 \times 10^{-12}$ 。Precision ID Identity Panel 试剂盒 90 个 SNP 位点的 CDP 为 $1-2.0524 \times 10^{-33}$, CPE_{trio} 为 $1-8.7093 \times 10^{-8}$, CPE_{duo} 为 $1-1.1638 \times 10^{-4}$, CPE_{AP} 为 $1-3.7257 \times 10^{-12}$ 。在个体识别、标准三联体鉴定、二联体鉴定、双亲皆疑鉴定中, 分别平均有 2.85、4.51、4.88、4.55 个 SNP 位点等于 1 个 STR 基因座的系统效能。结论 二代测序试剂盒 SNP 位点的系统效能较高, 多个 SNP 位点联合可应用于法庭科学中的个体识别和亲子鉴定。

关键词: 法医遗传学; 多态性, 单核苷酸; 亲子关系; 二代测序; 个体识别

中图分类号: DF795.2 文献标志码: A doi: 10.12116/j.issn.1004-5619.2018.03.005

文章编号: 1004-5619(2018)03-0242-06

Genetic Parameters of SNP Loci in Next Generation Sequencing Kits and Their Comparison

ZHOU Mi¹, ZHANG Ke², WANG Jun³

(1. Wuhu Public Security Bureau, Wuhu 241000, China; 2. Management Office of Physical Evidence, Department of Public Security of Anhui Province, Hefei 230061, China; 3. School of Computer & Information Science, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: Objective To calculate genetic parameters of SNP loci in next generation sequencing kits, and to compare them with STR loci for establishing the conversion ratio between SNP and STR system effectiveness. **Methods** Hardy-Weinberg equilibrium tests were performed in 101 SNP loci of next generation sequencing kits (ForenSeq™ DNA Signature Prep kit and Precision ID Identity Panel kit). The parameters of system effectiveness of SNP loci in the cases of personal identification, trios, duos, and alleged parents were calculated, which were compared with the genetic parameters of STR loci. **Results** Except 2 loci without the data of genotype frequency, other 99 SNP loci conformed to Hardy-Weinberg equilibrium tests ($P>0.05$). In ForenSeq™ DNA Signature Prep kit, the CDP of 94 SNP loci was $1-1.1521 \times 10^{-34}$, CPE_{trio} was $1-4.4169 \times 10^{-8}$, CPE_{duo} was $1-8.4837 \times 10^{-5}$, and CPE_{AP} was $1-1.2227 \times 10^{-12}$. In Precision ID Identity kit, the CDP was $1-2.0524 \times 10^{-33}$, CPE_{trio} was $1-8.7093 \times 10^{-8}$, CPE_{duo} was $1-1.1638 \times 10^{-4}$, and CPE_{AP} was $1-3.7257 \times 10^{-12}$. In the cases of personal identification, trios, duos and alleged parents, the system effectiveness of 2.85, 4.51, 4.88 and 4.55 SNP loci was equal to that of 1 STR locus, respectively. **Conclusion** With high system effectiveness of SNP loci, the next generation sequencing kits is suitable for personal identification and paternity testing in forensic science.

Keywords: forensic genetics; polymorphism, single nucleotide; parent-child relations; next generation sequencing; individual identification

近年来,二代测序技术蓬勃发展,逐步开始运用于法医学领域。2014年,美国 Illumina 公司推出了

MiSeq FGx™ 平台和 ForenSeq™ DNA Signature Prep 试剂盒,美国 Thermo Fisher Scientific 公司推出了 Ion Torrent PGM™ 平台和 Precision ID Identity Panel 试剂盒。运用二代测序技术,基于一份检材就可以同时获得短串联重复序列 (short tandem repeat, STR)、单核苷酸多态性 (single nucleotide polymorphism, SNP)、

基金项目:安徽省科技强警资助项目(1604d0802002)

作者简介:周密(1982—),男,硕士,主检法医师,主要从事法医物证学研究;E-mail:523603361@qq.com

插入/缺失(insertion/deletion, InDel)、线粒体 DNA(mitochondrial DNA, mtDNA)、信使 RNA(messenger RNA, mRNA)等各种类型的大量遗传标记信息。其中 SNP 被认为是第三代遗传标记,在法医学个体识别、表型预测和始祖研究等方面都有巨大潜力^[1]。本研究将对二代测序试剂盒中 SNP 位点的遗传学参数进行对比分析,探讨其在个体识别和亲子鉴定中的法医学应用价值。

1 材料与方法

1.1 基础数据

ForenSeq™ DNA Signature Prep 试剂盒包含 94 个常染色体 SNP 位点, Precision ID Identity Panel 试剂盒包含 90 个常染色体 SNP 位点, 两个试剂盒共有的 SNP 位点为 83 个, 合计有 101 个 SNP 位点。ForenSeq™ DNA Signature Prep 试剂盒独有位点 11 个: rs763869、rs8037429、rs8078417、rs2399332、rs279844、rs2920816、rs1294331、rs13182883、rs1336071、rs2111980、rs4606077。Precision ID Identity Panel 试剂盒独有位点 7 个: rs1872575、rs2016276、rs2292972、rs4288409、rs560681、rs7520386、rs7704770。

在国际人类基因组单体型图计划网站(<http://www.hapmap.org>)和美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)网站(<https://www.ncbi.nlm.nih.gov>)上查询 101 个 SNP 位点的等位基因频率、基因型频率、群体调查样本量、人群来源等遗传学参数。其中 95 个位点的数据来源于北京汉族人群(CHB)调查结果, 4 个位点的数据来源于中国人群(CHN)调查结果, 2 个位点的数据来源于北京汉族人群和日本人群(CHB+JPT)调查结果, 有 2 个位点仅有等位基因频率数据。

1.2 Hardy-Weinberg 平衡检验

Hardy-Weinberg 平衡检验通常采用 χ^2 检验来度量常用基因型的预期值和观察值之间的吻合程度。公式为:

$$\chi^2 = \sum (\text{观察值} - \text{预期值})^2 / \text{预期值}, \quad (1)$$

其中 χ^2 检验的自由度为:

$$v = \text{基因型数} - 1。$$

将每一种基因型的 χ^2 值相加获得总的 χ^2 值, 然后查 χ^2 值表得 P 值。以 $P > 0.05$ 作为差异无统计学意义的界限^[2]。根据 SNP 位点的等位基因频率和基因型频率进行 Hardy-Weinberg 平衡检验。

1.3 个体识别系统效能

个体识别的系统效能通常用个体识别率(discrimination power, DP)和累积个体识别率(cumulative discrimination power, CDP)来定量评价, 公式^[3]为:

$$DP = 1 - \sum_{i=1}^n P_i^2, \quad (2)$$

$$CDP = 1 - \prod_{i=1}^n (1 - DP_i), \quad (3)$$

式中, P_i 代表第 i 个基因型的频率, DP_i 为第 i 个基因座的 DP。

因为 CDP 为 $1 - DP$ 的连乘乘积, 所以本研究以 $1 - DP$ 对比遗传标记系统的系统效能。在北京汉族人群中, 根据公式(2)和(3), 计算 101 个 SNP 位点的 DP 和平均 DP(DP_{SNP}), 以及两个试剂盒的 CDP。根据中国汉族人群的遗传学参数^[4], 统计 41 个 STR 基因座的平均 DP(DP_{STR})。

设个体识别中, 平均 n_1 个 SNP 位点等于 1 个 STR 基因座的系统效能。以 $1 - DP_{SNP}$ 对比遗传标记系统间的系统效能, 则有:

$$(1 - DP_{SNP})^{n_1} = 1 - DP_{STR}, \quad (4)$$

故:

$$n_1 = \lg(1 - DP_{STR}) / \lg(1 - DP_{SNP}). \quad (5)$$

1.4 标准三联体鉴定系统效能

标准三联体鉴定的系统效能通常用标准三联体非父排除率(probability of paternity excluding in trios, PE_{trio})和累积非父排除率(probability of exclusion, CPE)来定量评价, 公式为:

$$PE_{trio} = \sum_{i=1}^n P_i(1 - P_i)^2 + \sum_{i < j} P_i^2 P_j^2 (3P_i + 3P_j - 4)^{[3]}, \quad (6)$$

$$CPE = 1 - \prod_{i=1}^n (1 - PE_i), \quad (7)$$

式中, P_i, P_j 分别代表第 i, j 个等位基因的频率。

因为 CPE 为 $1 - PE$ 的连乘乘积, 所以本研究以 $1 - PE$ 对比遗传标记系统的系统效能。在北京汉族人群中, 根据公式(6)和(7), 计算 101 个 SNP 位点的 PE_{trio} 和平均 PE_{trio} ($PE_{trio-SNP}$), 以及两个试剂盒的 CPE_{trio} 。根据中国汉族人群的遗传学参数^[4], 统计 41 个 STR 基因座的平均 PE_{trio} ($PE_{trio-STR}$)。

设标准三联体鉴定中, 平均 n_2 个 SNP 位点等于 1 个 STR 基因座的系统效能。以平均 PE_{trio} 对比遗传标记系统间的系统效能, 则有:

$$(1 - PE_{trio-SNP})^{n_2} = 1 - PE_{trio-STR}, \quad (8)$$

故:

$$n_2 = \lg(1 - PE_{trio-STR}) / \lg(1 - PE_{trio-SNP}). \quad (9)$$

1.5 二联体鉴定系统效能

二联体鉴定的系统效能通常用二联体非父排除率(probability of paternity excluding in duos, PE_{duo}), 公式^[3]为:

$$PE_{duo} = \sum_{i=1}^n P_i^2(1 - P_i)^2 + \sum_{i < j} 2P_i P_j (1 - P_i - P_j)^2, \quad (10)$$

式中, P_i, P_j 分别代表第 i, j 个等位基因的频率。

与三联体鉴定同理, 在北京汉族人群中, 根据公式(10)和(7), 计算 101 个 SNP 位点的 PE_{duo} 和平均 PE_{duo} , 以及两个试剂盒的 CPE_{duo} 。根据中国汉族人群的遗传学参数^[4], 统计 41 个 STR 基因座的平均 PE_{duo} 。设二联体鉴定中, 平均 n_3 个 SNP 位点等于 1 个 STR 基因座的系统效能。同理有:

$$n_3 = \lg(1 - PE_{duo-STR}) / \lg(1 - PE_{duo-SNP}) \quad (11)$$

1.6 双亲皆疑鉴定系统效能

双亲皆疑鉴定的系统效能通常用双亲皆疑排除率(probability of exclusion in alleged parents cases, PE_{AP})来定量评价, 公式为:

$$PE_{AP} = 1 + \sum_{i=1}^n [P_i^2(2-P_i)]^2 - 2[\sum_{i=1}^n P_i^2(2-P_i)]^2 + 4(\sum_{i=1}^n P_i^3)^2 - 4 \sum_{i=1}^n P_i^{6[5]}, \quad (12)$$

式中, P_i 代表第 i 个等位基因的频率。与三联体鉴定同理, 在北京汉族人群中, 根据公式(12)和(7), 计算 101 个 SNP 位点的 PE_{AP} 和平均 $PE_{AP}(PE_{AP-SNP})$, 以及两个试剂盒的 CPE_{AP} 。根据中国汉族人群的遗传学参数^[4], 统计 41 个 STR 基因座的平均 $PE_{AP}(PE_{AP-STR})$ 。设双亲皆疑鉴定中, 平均 n_4 个 SNP 位点等于 1 个 STR 基因座的系统效能。同理有:

$$n_4 = \lg(1 - PE_{AP-STR}) / \lg(1 - PE_{AP-SNP}) \quad (13)$$

2 结果

2.1 Hardy-Weinberg 平衡检验

101 个 SNP 位点的等位基因频率、样本数、人群来源、Hardy-Weinberg 平衡检验的 P 值见表 1。由表 1 可见, 除无基因型频率数据的 rs722098 和 rs2016276 位点外, 其余 99 个 SNP 位点均符合 Hardy-Weinberg 平衡定律 ($P > 0.05$)。

表 1 101 个 SNP 位点的遗传学参数

序号	SNP 位点	等位基因频率				样本数	人群	DP	PE_{tuo}	PE_{duo}	PE_{AP}	P 值 ⁽¹⁾
		A	C	G	T							
1	rs1005533	0.3780	-	0.6220	-	41	CHB	0.6270	0.1798	0.1106	0.2718	0.9930
2	rs10092491	-	0.5976	-	0.4024	41	CHB	0.6377	0.1827	0.1157	0.2753	0.9905
3	rs1015250	-	0.4333	0.5667	-	45	CHB	0.6104	0.1853	0.1206	0.2785	0.9992
4	rs1024116	0.1111	-	0.8889	-	45	CHB	0.3457	0.0890	0.0195	0.1545	0.9922
5	rs1028528	0.6667	-	0.3333	-	45	CHB	0.5511	0.1728	0.0988	0.2634	0.9802
6	rs1031825	0.5111	0.4889	-	-	45	CHB	0.5412	0.1874	0.1249	0.2812	0.9704
7	rs10488710	-	0.6477	0.3523	-	44	CHB	0.5052	0.1761	0.1041	0.2673	0.9423
8	rs10495407	0.4333	-	0.5667	-	45	CHB	0.6548	0.1853	0.1206	0.2785	0.9737
9	rs1058083	0.3444	-	0.6556	-	45	CHB	0.6064	0.1748	0.1020	0.2657	0.9979
10	rs10773760	0.6463	-	0.3537	-	41	CHB	0.5758	0.1763	0.1045	0.2676	0.9928
11	rs10776839	-	-	0.5000	0.5000	40	CHB	0.6463	0.1875	0.1250	0.2813	0.9950
12	rs1109037	0.4889	-	0.5111	-	45	CHB	0.6598	0.1874	0.1249	0.2812	0.9803
13	rs1294331	0.3500	-	0.6500	-	40	CHB	0.6150	0.1757	0.1035	0.2669	0.9927
14	rs12997453	0.3953	-	0.6047	-	43	CHB	0.5343	0.1819	0.1143	0.2743	0.9656
15	rs13182883	0.5233	-	0.4767	-	43	CHB	0.6047	0.1872	0.1245	0.2809	0.9974
16	rs13218440	0.4222	-	0.5778	-	45	CHB	0.6183	0.1844	0.1190	0.2774	1.0000
17	rs1335873	0.2667	-	-	0.7333	45	CHB	0.5511	0.1573	0.0765	0.2448	0.9506
18	rs1336071	0.3889	-	0.6111	-	45	CHB	0.6301	0.1812	0.1130	0.2734	0.9938
19	rs1355366	0.7556	-	0.2444	-	45	CHB	0.5294	0.1506	0.0682	0.2366	0.9966
20	rs1357617	0.8444	-	-	0.1556	45	CHB	0.4227	0.1141	0.0345	0.1900	0.9999
21	rs1360288	-	0.6556	-	0.3444	45	CHB	0.5709	0.1748	0.1020	0.2657	0.9914
22	rs1382387	-	-	0.3556	0.6444	45	CHB	0.5353	0.1766	0.1050	0.2679	0.9666
23	rs1413212	0.6026	-	0.3974	-	39	CHB	0.5470	0.1821	0.1147	0.2746	0.9736
24	rs1454361	0.5111	-	-	0.4889	45	CHB	0.6479	0.1874	0.1249	0.2812	0.9939
25	rs1463729	0.4667	-	0.5333	-	45	CHB	0.6459	0.1869	0.1239	0.2806	0.9943
26	rs1490413	0.5778	-	0.4222	-	45	CHB	0.6538	0.1844	0.1190	0.2774	0.9639
27	rs1493232	0.3295	0.6705	-	-	44	CHB	0.6043	0.1721	0.0976	0.2625	0.9921
28	rs1498553	-	0.5114	-	0.4886	44	CHB	0.6622	0.1874	0.1249	0.2812	0.9746
29	rs1523537	-	0.4444	-	0.5556	45	CHB	0.6242	0.1859	0.1219	0.2793	1.0000
30	rs1528460	-	0.4556	-	0.5444	45	CHB	0.5561	0.1865	0.1230	0.2800	0.9783
31	rs159606	0.3333	-	0.6667	-	45	CHB	0.5926	0.1728	0.0988	0.2634	1.0000

表 1(续)

序号	SNP 位点	等位基因频率				样本数	人群	DP	PE _{info}	PE _{dbp}	PE _{AP}	P 值 ¹⁾
		A	C	G	T							
32	rs1736442	0.4000	-	0.6000	-	45	CHB	0.6459	0.1824	0.1152	0.2749	0.9400
33	rs1821380	-	0.3111	0.6889	-	45	CHB	0.5768	0.1684	0.0919	0.2580	0.9993
34	rs1886510	-	0.8667	-	0.1333	45	CHB	0.3793	0.1022	0.0267	0.1736	0.9993
35	rs1979255	-	0.6000	0.4000	-	45	CHB	0.5867	0.1824	0.1152	0.2749	0.8452
36	rs2040411	0.1905	-	0.8095	-	42	CHB	0.4751	0.1304	0.0476	0.2115	0.9967
37	rs2046361	0.6556	-	-	0.3444	45	CHB	0.6064	0.1748	0.1020	0.2657	0.9979
38	rs2056277	-	0.8556	-	0.1444	45	CHB	0.4020	0.1083	0.0305	0.1821	0.9999
39	rs2076848	0.4111	-	-	0.5889	45	CHB	0.6479	0.1835	0.1172	0.2763	0.9219
40	rs2107612	0.9111	-	0.0889	-	45	CHB	0.2687	0.0744	0.0131	0.1324	0.9845
41	rs2111980	0.6444	-	0.3556	-	45	CHB	0.6242	0.1766	0.1050	0.2679	0.9497
42	rs214955	0.5444	-	0.4556	-	45	CHB	0.6360	0.1865	0.1230	0.2800	0.9982
43	rs221956	-	0.6136	-	0.3864	44	CHB	0.6322	0.1809	0.1124	0.2731	0.9906
44	rs2269355	-	0.2708	0.7292	-	24	CHN	0.5590	0.1585	0.0780	0.2462	0.9664
45	rs2342747	0.3537	-	0.6463	-	41	CHB	0.6139	0.1763	0.1045	0.2676	0.9957
46	rs2399332	0.3293	0.6707	-	-	41	CHB	0.5604	0.1721	0.0976	0.2625	0.9873
47	rs251934	-	0.1667	-	0.8333	45	CHB	0.4415	0.1196	0.0386	0.1974	0.9992
48	rs279844	0.4556	-	-	0.5444	45	CHB	0.5561	0.1865	0.1230	0.2800	0.9783
49	rs2830795	0.4889	-	0.5111	-	45	CHB	0.5768	0.1874	0.1249	0.2812	0.9879
50	rs2831700	0.4778	-	0.5222	-	45	CHB	0.6183	0.1873	0.1245	0.2809	0.9997
51	rs2920816	-	0.1951	-	0.8049	41	CHB	0.4688	0.1324	0.0493	0.2140	0.9753
52	rs321198	-	0.5667	-	0.4333	45	CHB	0.6104	0.1853	0.1206	0.2785	0.9992
53	rs338882	-	0.3556	-	0.6444	45	CHB	0.5886	0.1766	0.1050	0.2679	0.9978
54	rs354439	0.5333	-	-	0.4667	45	CHB	0.6044	0.1869	0.1239	0.2806	0.9975
55	rs3780962	-	0.5000	-	0.5000	45	CHB	0.6400	0.1875	0.1250	0.2813	0.9978
56	rs430046	-	0.6463	-	0.3537	41	CHB	0.5985	0.1763	0.1045	0.2676	0.9999
57	rs4364205	-	-	0.6222	0.3778	45	CHB	0.5116	0.1798	0.1105	0.2718	0.9490
58	rs445251	-	0.3333	0.6667	-	45	CHB	0.5748	0.1728	0.0988	0.2634	0.9950
59	rs4530059	0.2333	-	0.7667	-	45	CHB	0.5215	0.1469	0.0640	0.2321	0.9984
60	rs560681	0.5778	-	0.4222	-	45	CHB	0.6183	0.1844	0.1190	0.2774	1.0000
61	rs576261	0.6364	0.3636	-	-	44	CHB	0.6074	0.1779	0.1071	0.2694	0.9998
62	rs6444724	-	0.3750	-	0.6250	44	CHB	0.6312	0.1794	0.1099	0.2713	0.9847
63	rs6811238	-	-	0.6222	0.3778	45	CHB	0.5768	0.1798	0.1105	0.2718	0.9910
64	rs6955448	-	0.7667	-	0.2333	45	CHB	0.5215	0.1469	0.0640	0.2321	0.9984
65	rs7041158	-	0.6667	-	0.3333	45	CHB	0.6044	0.1728	0.0988	0.2634	0.9950
66	rs717302	0.8889	-	0.1111	-	45	CHB	0.3457	0.0890	0.0195	0.1545	0.9922
67	rs719366	-	0.2222	-	0.7778	45	CHB	0.5057	0.1430	0.0597	0.2273	0.9877
68	rs722098	0.4667	-	0.5333	-	60	CHB+JPT	0.6239	0.1869	0.1239	0.2806	-
69	rs722290	-	0.4444	0.5556	-	45	CHB	0.6420	0.1859	0.1219	0.2793	0.9950
70	rs727811	0.6667	0.3333	-	-	45	CHB	0.5511	0.1728	0.0988	0.2634	0.9802
71	rs729172	0.1667	0.8333	-	-	45	CHB	0.4326	0.1196	0.0386	0.1974	0.9928
72	rs733164	0.0444	-	0.9556	-	45	CHB	0.1620	0.0407	0.0036	0.0764	0.9989
73	rs735155	0.8444	-	0.1556	-	45	CHB	0.4286	0.1141	0.0345	0.1900	0.9832
74	rs737681	-	0.8778	-	0.1222	45	CHB	0.3546	0.0958	0.0230	0.1644	0.9977
75	rs740598	0.4778	-	0.5222	-	45	CHB	0.5916	0.1873	0.1245	0.2809	0.9936
76	rs740910	0.9222	-	0.0778	-	45	CHB	0.2627	0.0666	0.0103	0.1199	0.9964
77	rs763869	-	0.3111	-	0.6889	45	CHB	0.5768	0.1684	0.0919	0.2580	0.9993
78	rs8037429	-	0.5778	-	0.4222	45	CHB	0.5946	0.1844	0.1190	0.2774	0.9957
79	rs8078417	-	0.6222	-	0.3778	45	CHB	0.5768	0.1798	0.1105	0.2718	0.9910
80	rs826472	-	0.7917	-	0.2083	24	CHN	0.4965	0.1377	0.0544	0.2208	0.9999
81	rs873196	-	0.1333	-	0.8667	45	CHB	0.3793	0.1022	0.0267	0.1736	0.9993

表 1(续)

序号	SNP 位点	等位基因频率				样本数	人群	DP	PE _{tno}	PE _{duo}	PE _{AP}	P 值 ¹⁾
		A	C	G	T							
82	rs876724	-	0.5667	-	0.4333	45	CHB	0.6311	0.1853	0.1206	0.2785	0.9988
83	rs891700	0.4889	-	0.5111	-	45	CHB	0.6064	0.1874	0.1249	0.2812	0.9977
84	rs901398	-	0.2667	-	0.7333	45	CHB	0.5570	0.1573	0.0765	0.2448	0.9793
85	rs907100	-	0.4432	0.5568	-	44	CHB	0.6291	0.1859	0.1218	0.2792	0.9995
86	rs914165	0.3636	-	0.6364	-	44	CHB	0.5620	0.1779	0.1071	0.2694	0.9842
87	rs917118	-	0.6818	-	0.3182	44	CHB	0.5589	0.1699	0.0941	0.2598	0.9885
88	rs938283	-	0.1250	-	0.8750	24	CHN	0.3750	0.0974	0.0239	0.1668	0.9898
89	rs964681	-	0.3000	-	0.7000	45	CHB	0.5837	0.1659	0.0882	0.2551	0.9950
90	rs987640	0.5444	-	-	0.4556	45	CHB	0.5886	0.1865	0.1230	0.2800	0.9928
91	rs9905977	0.4000	-	0.6000	-	45	CHB	0.6459	0.1824	0.1152	0.2749	0.9670
92	rs993934	-	0.5667	-	0.4333	45	CHB	0.6311	0.1853	0.1206	0.2785	0.9988
93	rs9951171	0.4773	-	0.5227	-	44	CHB	0.5981	0.1872	0.1245	0.2809	0.9957
94	rs4606077	-	0.7262	-	0.2738	42	CHB	0.5431	0.1593	0.0791	0.2472	0.9906
95	rs1872575	-	0.4444	-	0.5556	45	CHB	0.6242	0.1859	0.1219	0.2793	1.0000
96	rs2016276	0.5833	-	0.4167	-	60	CHB+JPT	0.6178	0.1840	0.1182	0.2769	-
97	rs2292972	-	0.3690	-	0.6310	42	CHB	0.6315	0.1786	0.1084	0.2704	0.9732
98	rs4288409	0.3667	0.6333	-	-	45	CHB	0.5570	0.1783	0.1079	0.2699	0.9809
99	rs560681	0.5778	-	0.4222	-	45	CHB	0.6183	0.1844	0.1190	0.2774	1.0000
100	rs7520386	0.4722	-	0.5278	-	18	CHN	0.6235	0.1871	0.1242	0.2808	1.0000
101	rs7704770	0.6111	-	0.3889	-	45	CHB	0.6153	0.1812	0.1130	0.2734	0.9998

注:1)Hardy-Weinberg 平衡检验 P 值;“-”表示无数据

2.2 个体识别系统效能和对比

101 个 SNP 位点的 DP 见表 1。ForenSeq™ DNA Signature Prep 试剂盒的 CDP 为 1-1.1521×10⁻³⁴, Precision ID Identity Panel 试剂盒的 CDP 为 1-2.0524×10⁻³³。1-DP_{SNP} 为 0.4404, 1-DP_{STR} 为 0.0969, n₁ 为 2.85。结果表明,在个体识别中,平均 2.85 个 SNP 位点等于 1 个 STR 基因座的系统效能。

2.3 标准三联体鉴定系统效能和对比

101 个 SNP 位点的 PE_{tno} 见表 1。ForenSeq™ DNA Signature Prep 试剂盒的 CPE_{tno} 为 1-4.4169×10⁻⁸, Precision ID Identity Panel 试剂盒的 CPE_{tno} 为 1-8.7093×10⁻⁸。1-PE_{tno-SNP} 为 0.8334, 1-PE_{tno-STR} 为 0.4394, n₂ 为 4.51。结果表明,在标准三联体鉴定中,平均 4.51 个 SNP 位点等于 1 个 STR 基因座的系统效能。

2.4 二联体鉴定系统效能和对比

101 个 SNP 位点的 PE_{duo} 见表 1。ForenSeq™ DNA Signature Prep 试剂盒的 CPE_{duo} 为 1-8.4837×10⁻⁵, Precision ID Identity Panel 试剂盒的 CPE_{duo} 为 1-1.1638×10⁻⁴。1-PE_{duo-SNP} 为 0.9030, 1-PE_{duo-STR} 为 0.6076, n₃ 为 4.88。结果表明,在二联体鉴定中,平均 4.88 个 SNP 位点等于 1 个 STR 基因座的系统效能。

2.5 双亲皆疑鉴定系统效能和对比

101 个 SNP 位点的 PE_{AP} 见表 1。ForenSeq™ DNA Signature Prep 试剂盒的 CPE_{AP} 为 1-1.2227×10⁻¹², Pre-

cision ID Identity Panel 试剂盒的 CPE_{AP} 为 1-3.7257×10⁻¹²。1-PE_{AP-SNP} 为 0.7464, 1-PE_{AP-STR} 为 0.2641, n₄ 为 4.55。结果表明,在双亲皆疑鉴定中,平均 4.55 个 SNP 位点等于 1 个 STR 基因座的系统效能。

3 讨论

SNP 广泛存在于人类基因组中的编码区和非编码区,平均每 721 bp 有 1 个 SNP^[6],是目前分布最广泛、数量最多的一种遗传标记。SNP 的特点有:(1)扩增片段极短;(2)大都表现为二等位基因标记;(3)突变率约为 10⁻⁹,远低于 STR 突变率(约 10⁻³)^[2]。这些特点决定了 SNP 在法医学个体识别与亲子鉴定领域中具有巨大潜力。遗传学参数是二代测序 SNP 位点结果解释的前提,是二代测序技术在个体识别和亲子鉴定中大规模应用的根本基础。因此本研究首先查询并计算二代测序试剂盒检测的 101 个 SNP 位点的常用遗传学参数,为 SNP 分型结果解释奠定基础。其次,了解了在法医学个体识别和亲子鉴定中,需要检测多少个 SNP 位点才能达到目前 STR 基因座的检验水平。本研究进一步计算了个体识别、标准三联体鉴定、二联体鉴定和双亲皆疑鉴定 4 种最常见的鉴定类型中,SNP 位点和 STR 基因座系统效能换算比例,为今后设计二代测序试剂盒的 SNP 位点提供重要参考。

在整理 101 个 SNP 位点群体遗传学数据的过程

中,笔者发现如下问题:(1)群体调查样本量过少,多为41~45人,最少的仅有18人,而STR群体调查样本量往往有数千人;(2)中国汉族人群群体遗传学数据较少,现有数据多为北京汉族人群,有2个位点还混有日本人群;(3)2个位点无基因型频率,既不能进行Hardy-Weinberg平衡检验,也不能直接计算DP值。除无基因型频率数据的rs722098和rs2016276位点外,本研究中其他99个SNP位点均符合Hardy-Weinberg平衡检验。只有处于Hardy-Weinberg平衡的群体中获得的群体数据,才能保证其可靠性和准确性,才能应用于法医学个体识别和亲子鉴定。此外,因为本研究无法获得相关数据,不能检验各SNP位点的独立性,即连锁不平衡检验,而连锁不平衡检验是个体识别的累积似然率(cumulative likelihood rate, CLR)和亲子鉴定的累积亲权指数(cumulative paternity index, CPI)计算的前提。考虑到群体数据的上述不足,笔者认为应进一步对中国汉族人群进行较大规模的SNP位点群体调查,以获取更准确的数据进行适合性检验和独立性检验,统计学数据的计算也将更为精确。

本文两个二代测序试剂盒分别包含90和94个SNP位点,个体识别系统效能换算后相当于32和33个STR基因座,远超一代测序;亲子鉴定系统效能换算后相当于18~21个STR基因座,接近一代测序^[4]。在系统效能相当的情况下,SNP位点具有突变案例数极低的优势。以二联体鉴定为例,检测20个STR基因座,则平均50个案例出现1例突变,必须考虑突变案例;而检测100个SNP位点,则平均1000万个案例出现1例突变,无需考虑突变案例。可见,STR突变案例是SNP的20万倍。此外,如果联合应用二代测序的SNP位点和STR基因座,充分利用二代测序的超

大信息量优势,则祖孙、同胞、半同胞等较远亲缘关系的亲子鉴定还将有进一步发展。

总之,二代测序SNP分型技术在法庭科学中的个体识别和亲子鉴定领域展现出明显的优势,具备极大的应用价值,对于公安实战也有强大吸引力。但该技术的大规模普及仍存在如下挑战:(1)科学的结果解释。如何将二代测序数据结果转化为鉴定结论和法庭证据,有待法医DNA概率统计学理论的深入研究。(2)二代测序和一代测序数据库的接轨。公安机关DNA数据库已超过6000万条数据,而且数据量还在高速增长,二代测序结果势必要与DNA数据库平滑对接。(3)分析软件的支持。二代测序信息量极高,数据结果的分析、比对、计算靠人工几乎不可能,必须依赖计算机软件。(4)测序成本须有效降低。二代测序基于开放性平台,开发国产试剂盒可大幅降低成本。

参考文献:

- [1] 王乐,叶健,白雪,等.二代测序技术及其在法医遗传学中的应用[J].刑事技术,2015,40(5):353-358.
- [2] 郑秀芬.法医DNA分析[M].北京:中国人民公安大学出版社,2002:387.
- [3] 侯一平.法医物证司法鉴定实务[M].北京:法律出版社,2013.
- [4] 吴微微,刘冰,郝宏蕾,等.中国28个省/区汉族人群41个STR基因座多态性数据分析[J].中国法医学杂志,2016,31(1):27-32.
- [5] GUNDEL H, REETZ I. Exclusion probabilities obtainable by biochemical polymorphisms in dogs[J]. Animal Genetics, 1981, 12(2):123-132.
- [6] 张凤珍,郝岗平,张媛英.人类SNP多态性的研究及应用[J].泰山医学院学报,2003,24(2):203-206.

(收稿日期:2016-07-28)

(本文编辑:边英男)