

DOI:10.12113/j.issn.1672-5565.201802003

# 基于全外显子组数据的 ABO 血型的基因判定

雷建强,张民霞,唐元华\*

(首度生物科技(苏州)有限公司,江苏 苏州 215000)

**摘要:**全外显子组测序研究已经应用在疾病、药物等方面,是临床研究中一种辅助分子诊断方法。该方法也逐步应用在临床 ABO 血型的判定中,由于目前临床血型判定的主要方法是血清学,疑难样本判定等问题无法解决,因此分子水平的基因测序方法提高了血型判定的准确性,比如 PCR 方法和基因芯片方法。为进一步提高 ABO 血型精细分型的准确性,通过分析全外显子组测序数据,开发了相关的 VBA 程序,能够快速自动化判定 ABO 精细分型,并且初步判定结果与临床判定结果一致,可以作为临床血型精确判定的辅助手段。

**关键词:**全外显子组;基因测序;ABO;血型判定

**中图分类号:**Q343.1 **文献标志码:**A **文章编号:**1672-5565(2018)03-184-06

## Gene determination of ABO blood type based on whole exome data

LEI Jianqiang, ZHANG Minxia, TANG Yuanhua\*

(First Dimension Genetic Science and Technology Limited Company, Suzhou 215000, Jiangsu, China)

**Abstract:** Whole Exome Sequencing (WES) has been applied in diseases and drugs. It is a supplementary molecular diagnostic method in clinical research. This method is also gradually applied to the determination of the clinical ABO blood group. The main method of clinical blood group determination is serology. It is difficult to determine the difficult samples and other problems, so the molecular level gene sequencing method such as the PCR method and the gene chip method improves the accuracy of blood type determination. In order to further improve the accuracy of the subtype of ABO blood group, a related VBA program was developed by analyzing the data of the WES, which could quickly and automatically determine the subtype of ABO blood group. The preliminary results were consistent with the clinical judgment results, which could be used as an auxiliary means for accurate diagnosis of clinical ABO blood group.

**Keywords:** Whole exome; Gene sequencing; ABO; Blood type determination

人类基因中大约有 180 000 个外显子,占人类基因组的 1%,约 30 Mb<sup>[1]</sup>。全外显子组测序(Whole Exome Sequencing, WES)是指利用序列捕获或者靶向技术将全基因组外显子区域 DNA 富集后再进行高通量测序的基因组分析方法<sup>[2]</sup>。目前外显子组测序主要用于识别和研究疾病、种群进化相关的编码区及调控区域内的结构变异,结合大量公共数据库提供的全外显子组数据,有利于更好地解释变异之间的关联和疾病的致病机理,因此全外显子组测序具有广阔的应用前景<sup>[3-5]</sup>。

全外显子组测序技术的应用同时促进了分子诊断水平的发展。分子诊断主要指与疾病相关的各种结构蛋白、酶、抗原抗体、免疫活性分子基因的检测。大多数的研究主要集中在疾病相关的基因检测,但 ABO 血型的分子诊断也尤为重要<sup>[6-7]</sup>。ABO 血型判定在临床上有着重大意义,主要用于临床输血、器官移植选择血型相符的供体、不孕症和新生儿溶血症病因分析、亲子鉴定等方面。目前相对较完善和标准化的技术是血清学判定方法,但只能判定 ABO 血型的表现型,而在某些疑难样本、产前胎儿血型判

收稿日期:2018-02-09;修回日期:2018-04-02.

作者简介:雷建强,男,博士,研究方向:E-mail: Leiqj\_fd@163.com.

\*通信作者:唐元华,男,首度基因 CEO, E-mail: ttang@firstdimension.net.

定和遗传学研究等方面该技术受到限制。随着血型系统中多种亚型和基因多态性的深入研究,血型分子诊断技术克服了血清学技术中的一些限制,能够更加准确判定血型亚型<sup>[8-9]</sup>。

目前血型分子诊断技术主要基于 PCR 方法比如聚合酶链反应-序列特异性引物(PCR-SSP)和基因芯片技术。PCR-SSP 方法基于已知序列分别设计特异性引物,直接扩增出目的基因片段,通过凝胶电泳分析基因型结果。该方法简单快捷、易于操作,但也有不足之处如扩增引物设计缺陷、技术人员操作水平、电泳图谱的质量低、低丰度变异检测率低等问题都会影响最终结果的判定<sup>[10-13]</sup>。基因芯片技术是指经过标记的待测样本通过与芯片上特定位置的探针杂交,可根据碱基互补配对原则确定靶序列,经激光共聚焦显微镜扫描后以计算机对荧光信号进行比较和检测迅速获取信息。该技术可大量检测和分析 DNA 变异及多态性,自动化程度高,快速分析差异表达基因等,但其缺点是重复多次使用后芯片或微阵列的敏感性会降低,样品的制备和标记较复杂等<sup>[14-15]</sup>。两种分子检测技术的共有的局限性是只能检测出已知的基因序列信息,不能发现新的基因变异。而二代测序中的全外显子组测序技术,相比前两种技术避免了这一不足。该方法是一种序列比对方法,通过检测样本序列与已知序列进行比对,从而找出所有变异位点信息,并对变异信息进行生

物信息大数据统计和分析,最终得到可信度和准确度高的变异信息。全外显子组测序方法位点覆盖度高,除已知位点变异外还可以发现新的变异位点信息,能够更加精确检测 ABO 的基因型。目前国内外还没有关于应用全外显子组测序技术检测 ABO 血型的相关报道。因此,基于以上研究与分析,全外显子组测序用于 ABO 血型的基因型判定更加精确,其遗传规律也更加清晰,是临床血型判定方法较好的辅助手段。

### 1 全外显子组测序中 ABO 血型判定方法

基于全外显子组测序的基因检测结果,利用 Excel VBA 程序对 ABO 进行基因分型。该方法的核心思想为将基因突变转换为数字:在单条染色体下,0 定义为野生型,1 为杂合突变;在两条染色体下,0 定义为野生型,1 为杂合突变,2 为纯合突变。将两条染色体的纯杂合变异转换为两条染色体的数字相加,预先生成所有的二倍体化的 ABO 基因型的组合情况,最后测序结果数据与这个二倍体化的表格进行比对来判别(ABO 血型判定的核心方法见图 1)。该方法方便快捷,节约了时间和人力成本,有利于分析查看样本每个位点信息,有助于积累中国人群血型亚型数据,具有潜在的生物信息分析价值。

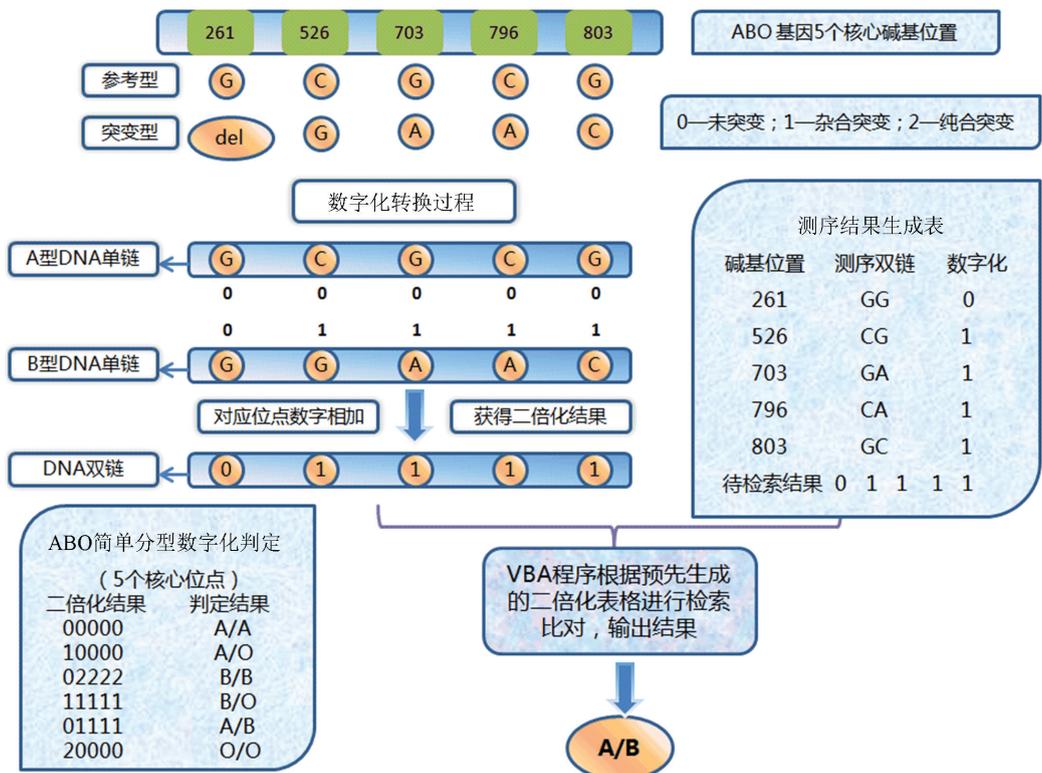


图1 ABO 血型判定的核心方法

Fig.1 The core method of ABO blood type determination

## 2 全外显子组测序中 ABO 血型判定基本流程

本文 ABO 血型判定基本流程主要分为三个部分：ABO 基因型判定数字化过程、获取染色体绝对位置和 WES 数据处理。其中 ABO 基因型判定数字化过程主要通过 4 个 VBA 程序最终获得《ABO 数字化表—二倍体化》工作表；获取染色体绝对位置

部分通过 1 个 VBA 程序和人工修正获得《染色体绝对位置表—数字化》和《染色体绝对位置表—数字化修正》工作表；WES 数据处理部分通过生物信息数据分析后获得《ABO 基因测序结果列表》，然后在程序 6.1 的运行下获得《ABO 基因测序结果列表—数字化》工作表；最后运行程序 6.2，将《ABO 数字化表—二倍体化》和《ABO 基因测序结果列表—数字化》两个工作表进行关联，经过程序自动检索对比，最终输出 ABO 基因型结果，基本判定流程见图 2。

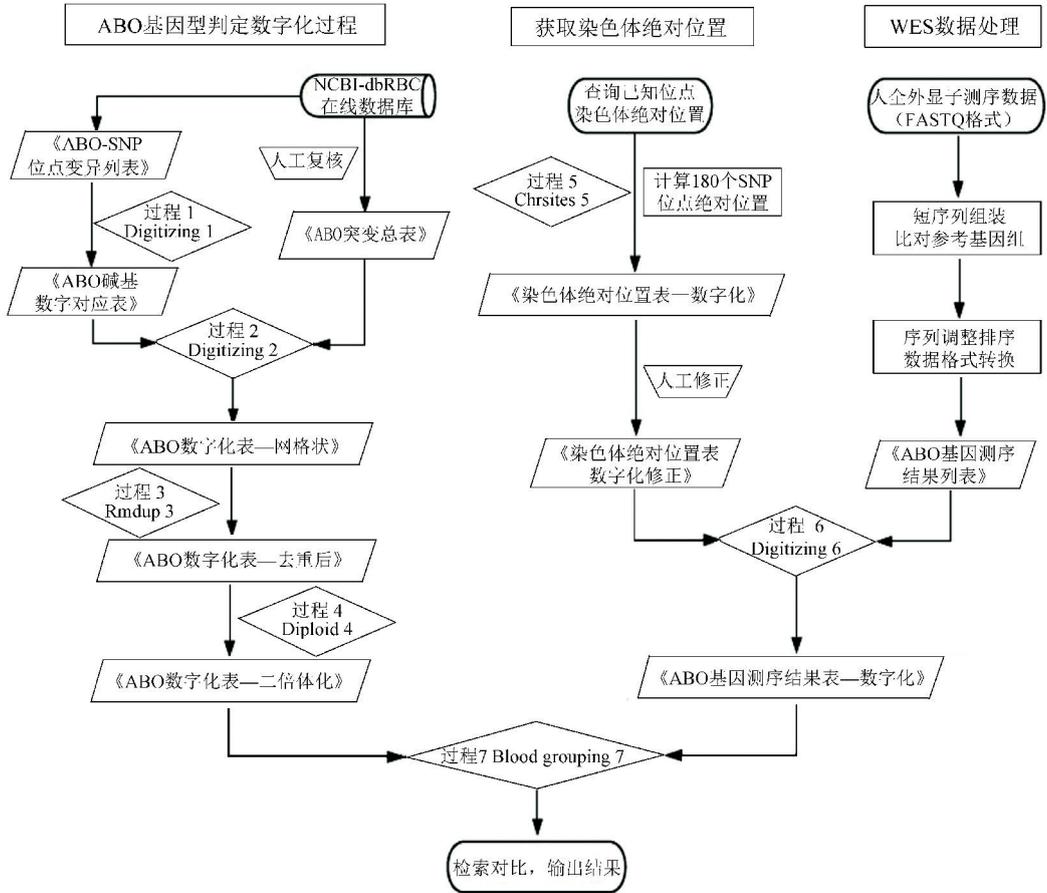


图 2 ABO 血型判定基本流程

Fig.2 Basic flow of ABO blood type determination

## 3 全外显子组测序 ABO 血型判定具体过程

### 3.1 ABO 等位基因汇总表的下载、复核和工作簿建立

(1) 从 NCBI-dbRBC 在线数据库下载 ABO 等位基因汇总表，进行精细划分和处理：去除内含子区域的变异位点（但保留内含子与外显子相接的 50 bp 范围，因某些内含子位点会影响 ABO 的基因型结果），获得《ABO-SNP 位点变异列表》。

(2) 根据 dbRBC 的在线版记录结果进行查漏补

缺，复核每一个位点的碱基突变情况。

以 A101 型为参考型，最终共计 A/B/O/AB 型的 316 种亚型，其中 A 型共有 134 种亚型，包含 A1、A2、A3、Am、Ax、Aw、Ael 等；B 型共有 91 种亚型，包含 B1、B3、Bel、Bm、Bw、Bx 等；AB 杂合亚型共有 17 种，包含 cis-AB、B(A) 等；O 型共有 74 种亚型，包含 O1、O80 等。从碱基位置 1 至 1 116 位，合计 180 个 SNP 位点。该部分结果整理到《ABO 血型的基因判定》工作簿的《ABO 突变总表》工作表中。

(3) 创建《ABO 血型的基因判定》工作簿 (WorkBook)，该工作簿建立以下 7 个工作表 (WorkSheet)：1)《ABO 突变总表》；2)《ABO 碱基数

字对应表》;3)《ABO 数字化表—网格状》;4)《ABO 数字化表—去重后》;5)《ABO 数字化表—二倍体化》;6)《染色体绝对位置表—数字化》;7)《染色体绝对位置表—数字化修正表》。以下 7 个程序运行完毕后,除《临时》工作表,其它工作表都将填写完毕内容。具体程序详见附件。

### 3.2 碱基变异与数字对应

ABO 区分型共计 180 个基因位点,每个位点最多出现 3 种变异型,野生型和变异型分别以数字 0, 1, 3, 7 表示,其中 1 代表出现次数最多的变异型,其次是 3,再次是 7,将每个位点出现的突变的种类统计并以竖排的方式列出来,运行程序 1 后结果输出在《ABO 碱基数字对应表》工作表中。

### 3.3 数字化总表

以《ABO 突变总表》为结构蓝本,将所有的碱基转换为数字,空白代表与 A101 一致,用 0 表示,其它变异型按《ABO 碱基数字对应表》转换为数字。最终将《ABO 突变总表》中的 316 种亚型和 180 个位点进行数字化标注,得到一张网格状的数字化总表,运行程序 2 后结果输出在《ABO 数字化表—网格状》工作表中。

### 3.4 去重复

《ABO 数字化》工作表中包含的 316 种亚型,其数字化结果会出现所有突变完全相同的结果(也就是突变完全相同的两种亚型),因此需要进行去重复,运行程序 3 去除完全相同的亚型,最终得到去重后的 ABO 数字化表,保留 285 种亚型,结果输出在《ABO 数字化表—去重后》工作表中。

### 3.5 二倍体化

由于等位基因位于成对的染色体上,因此 ABO 血型的基因型由一对等位基因决定。将去重 ABO 数字化总表进行二倍体化,也就是两两组合、对应位置数字相加,得到 285 种亚型全部的二倍体数字化结果,最终得到 40 755 种组合,从 A101/A101、A101/

A102、A101/A103...到 O80/O80。运行程序 4 后结果输出在《ABO 数字化表—二倍体化》工作表中。

### 3.6 染色体绝对位置

(1)参考 NCBI 数据库,ABO 基因共有 7 个外显子区域,每个外显子查询一个已知的位点,获取 GRCh37 版本的染色体绝对位置,运行程序 5,获得每个染色体绝对位置各个碱基测序结果的数字化后的数值,结果记录在《染色体绝对位置表—数字化》工作表的前 4 列中,前 4 列的表头分别为:外显子编号、该外显子最小碱基、碱基位置、染色体绝对位置 (GRCh37);

(2)上表数字化后需要根据下机数据的格式和输出方式进行人工调整,最终结果保存在《染色体绝对位置表—数字化修正》工作表中。

### 3.7 ABO 精细血型判定

WES 测序的下机数据,将 ABO 基因的测序结果单独输出到一个表格《ABO 基因测序结果列表》,结果表格的前 3 列分别为:碱基位置、染色体绝对位置、测序结果。运行程序 6.1,将测序结果按《染色体绝对位置表—数字化修正》工作表的数字化规则转换为数字。运行程序 6.2,将《ABO 基因测序结果列表》与预先生成好的《ABO 数字化表—数字化》工作表进行检索比对,数字差异个数为 0 的即为该样本的 ABO 血型基因判定结果,结果输出在 D2 单元格内。

## 4 VBA 程序运行实例

随机选取 30 个样本的全外显子组测序结果进行测试,其 ABO 基因分型的结果与事先在医院检测的血清学表型一致,这 30 个样本包含 A、B、AB 和 O 型血。本文列举了 8 个样本的血型判定结果如表 1。

表 1 8 个样本的血型判定结果

Table 1 Blood type determination results of 8 samples

样本编号	代号	性别	血型表型	精细分型
DNAMMER9I170303-A02	YYQ	女	A	A102/O01
DNAMMER9I170303-A03	XG	男	A	A101/O01
DNAMMER9I170303-A04	SFL	女	AB	A102/B101
DNAMMER9I170303-A08	GLL	男	O	O01/O01
DNAMMER9I170303-A21	LTT	女	AB	A102/B101
DNAMMER9I170303-A22	FMM	女	A	A101/O02
DNAMMER9I170303-A25	HN	女	A	A102/O01
DNAMMER9I170303-A27	CW	男	B	B101/B101

## 5 讨 论

1) 随着高通量测序技术的普及和发展,本文 ABO 血型的判定方法充分利用了全外显子组测序技术的优势。该技术与现有血型分子诊断方法 PCR 法和基因芯片法相比,简便快捷且能更加精确判定血型亚型,尤其对于临床上疑难样本的鉴定、产前胎儿血型判定和血型遗传学研究等方面有重大作用,是临床上血型精确判定的有效辅助手段。

2) 由于全外显子组测序的生物信息数据分析较为复杂,针对 ABO 血型判定需要开发新的算法,因此本文运用 VBA 程序的方法节约了时间和人力成本,方便快捷且自动化程度高,同时也解决了 ABO 亚型分型信息实时更新的问题,保证数据信息的全面可靠性。VBA 程序方法的核心思想为将基因突变转换为数字:0 定义为野生型,1 为杂合突变,2 为纯合突变;单条染色体的二倍体化转换为两条染色体的数字相加;测序结果与已知血型序列之间的同源比对,转换成简单的数字比较。但本方法也存在缺点:无法区分两个相距较远的杂合突变是在一条染色体上还是在分别在两条染色体上,即不能判定各个杂合突变的连锁情况。但是全外显子组测序的读长很短,一般不超过 150bp,各外显子之间很大的内含子区域无测序数据,全外显子组测序本身就比较难以获得一个基因突变的连锁情况,因此本文采用的数字化方法,并没有丢失太多有效的测序信息。该方法通过多个位点的突变来判别一个二倍体的基因型组合,除了 ABO 血型判定外,还可以应用在其它方面的基因型判定,例如药物代谢酶的基因亚型判定等。

3) 利用该方法判定 ABO 血型,经过血型表型的核实,证明该方法的初步判定结果是一致的,进一步表明该方法的准确度很高。目前临床中仍以血清学作为血型判定的主要方法,而该方法的判定结果可以作为临床应用的辅助手段,具有很大的应用潜力。

## 参考文献(References)

[1] BOTSTEIN D, RISCH N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease[J]. *Nature Genetics*, 2003, 33(2): 228-237. DOI: 10.1038/ng1090.

[2] STADLER Z K, VIJAI J, THOM P, et al. Genome-wide association studies of cancer predisposition[J]. *Hematology Clinics of North America*, 2010, 24(7): 629-637. DOI: 10.

1016/j.hoc.2010.06.009.

[3] ZHANG X Y, WEN J, YANG W, et al. Gain-of-Function mutations in SCN11A cause Familial Episodic Pain[J]. *American Journal of Human Genetics*, 2013, 93(5): 957-966. DOI: 10.1016/j.ajhg.2013.09.016.

[4] JONES S, WANG T L, SHIH I M, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma[J]. *Science*, 2010, 330(6001): 228-231. DOI: 10.1126/science.1196333.

[5] ROBERT C, FUENTES P, TROUP K, et al. Design and development of exome capture sequencing for the domestic pig (*Sus scrofa*) [J]. *BMC Genomics*, 2014, 15(1): 1-9. DOI: 10.1186/1471-2164-15-550.

[6] 沈立松, 谢国化. 我国分子诊断的现状 & 问题[J]. *中华检验医学杂志*, 2016, 39: 473-476. DOI: 10.3760/cma.j.issn.1009-9158.2016.07.001.

SHEN Lisong, XIE Guohua. Molecular diagnosis in China: present status and challenges ahead[J]. *Chinese Journal of Laboratory Medicine*, 2016, 39: 473-476. DOI: 10.3760/cma.j.issn.1009-9158.2016.07.001.

[7] 谢兰, 刘冉, 冯娟, 等. 中国分子诊断产业战略研究[J]. *中国工程科学杂志*, 2017, 19(2): 029-036. DOI: 10.15302/J-SSCAE-2017.02.005.

XIE Lan, LIU Ran, FENG Juan, et al. Strategic research on the molecular diagnostics industry in China [J]. *Chinese Journal of Engineering Sciences*, 2017, 19(2): 029-036. DOI: 10.15302/J-SSCAE-2017.02.005.

[8] 孙晓琳, 关晓珍, 于洋, 等. 36 例 ABO 血型亚型检测及血清学分析[J]. *临床输血与检验*, 2012, 14(3): 215-218. DOI: 10.3969/j.issn.1671-2587.2012.03.006.

SUN Xiaolin, GUAN Xiaozhen, YU Yang, et al. The Detection of serology and analysis on ABO blood subgroup in 36 cases [J]. *Journal of Clinical Transfusion and Laboratory Medicine*, 2012, 14(3): 215-218. DOI: 10.3969/j.issn.1671-2587.2012.03.006.

[9] 喻琼. ABO 血型基因研究进展[J]. *中国输血杂志*, 2006(19): 75-80. DOI: 1004-549X(2006)01-0075-05.

YU Qiong. Progress in the research of ABO blood group gene [J]. *Chinese Journal of Blood Transfusion*, 2006(19): 75-80. DOI: 10.13303/j.issn.1004-549X(2006)01-0075-05.

[10] 赵桐茂. 血型分子分型技术的进展和应用[J]. *临床输血与检验*, 2017, 19(5): 530-536. DOI: 10.3969/j.issn.1671-2587.2017.05.039.

ZHAO Tongmao. Progress and application of blood type molecular typing [J]. *Journal of Clinical Transfusion and Laboratory Medicine*, 2017, 19(5): 530-536. DOI: 10.3969/j.issn.1671-2587.2017.05.039.

[11] 谭茜茜, 何涛, 邹海曼, 等. 血型血清学和 PCR-SSP 方法在 ABO 疑难血型鉴定中的互补性应用[J]. *中国输血杂志*, 2017, 30(11): 1248-1250. DOI: 10.13303/j.

cjbt.issn.1004-549X.2017.11.010.

TAN Qianqian, HE Tao, ZOU Haiman, et al. Complementary application of PCR-SSP and serology techniques in identifying ambiguous ABO blood groups [J]. Chinese Journal of Blood Transfusion, 2017, 30(11): 1248-1250. DOI: 10.13303/j.cjbt.issn.1004-549X.2017.11.010.

[12] 杨晓玲, 钟青, 许业栋, 等. 疑难 ABO 血型标本的表型与相应等位基因检测在临床中的应用 [J]. 中国优生与遗传杂志, 2017, 25(11): 23-25. DOI: 10.13404/j.cnki.cjbhh.2017.11.008.

YANG Xiaoling, ZHONG Qing, XU Yedong, et al. The phenotype of difficult ABO blood group specimens and the detection of corresponding alleles in clinical use [J]. Chinese Journal of Birth Health & Heredity, 2017, 25(11): 23-25. DOI: 10.13404/j.cnki.cjbhh.2017.11.008.

[13] 张建军, 曾艳. 无创胎儿血型基因型检测的应用进展 [J]. 中国优生与遗传杂志, 2018, 26(1): 1-3. DOI: 10.13404/j.cnki.cjbhh.2018.01.001.

ZHANG Jianjun, ZENG Yan. Improvements of non-invasive methods to genotyped fetal ABO and RhD blood group [J]. Chinese Journal of Birth Health & Heredity, 2018, 26(1): 1-3. DOI: 10.13404/j.cnki.cjbhh.2018.01.001.

[14] 陈岩, 潘龙. 基因芯片技术研究进展 [J]. 齐齐哈尔医学院学报, 2011, 32(17): 2828-2830. DOI: 10.3969/j.issn.1674-6341.2014.02.012.

CHEN Yan, PAN Long. Progress in the research of gene chip technology [J]. Journal of Qiqihar University of Medicine, 2011, 32(17): 2828-2830. DOI: 10.3969/j.issn.1674-6341.2014.02.012.

[15] 苏宁, 杨丽, 刘娟, 等. 基因芯片技术的国内应用研究进展 [J]. 生物技术通讯, 2016, 27(2): 289-292. DOI: 10.3969/j.issn.1009-0002.2016.02.033.

SU Ning, YANG Li, LIU Juan, et al. Application research progress of gene chip technology in China [J]. Letters In Biotechnology, 2016, 27(2): 289-292. DOI: 10.3969/j.issn.1009-0002.2016.02.033.