



# Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA maternal plasma

通过大规模基因组测序技术利用孕妇血浆开展胎儿染色体非整倍型的无创产前诊断

北京交通大学生物医学工程系 蔡正厅

邮箱 : [jety2858@163.com](mailto:jety2858@163.com)

# 目录

## CONTENTS

01

背景介绍

INTRODUCTION

02

研究方法

METHODS

03

相关结论

RESULTS

04

后续讨论

DISCUSSION

# 目录

## CONTENTS

01

背景介绍

INTRODUCTION

02

研究方法

METHODS

03

相关结论

RESULTS




04

后续讨论

DISCUSSION

# 1.1 染色体非整倍体疾病检查



- 胎儿染色体非整倍体疾病检查是孕妇产前检测中的一个重要项目。
- 染色体非整倍体是目前临床上最常见的染色体疾病，其发生具有偶然性和随机性，发病率随孕妇年龄的增高而升高。
- 由于目前没有有效手段，降低生育染色体疾病患儿风险的最好方法就是通过产前遗传咨询、产前筛查及产前诊断，尽早发现并解决问题。

## 1.2 以往的检测方法及其缺陷

### 有创产检（传统）：

#### 1) 绒毛膜绒毛取样、羊膜腔穿刺

- 有一定的流产风险。

### 无创产检（间接）：

#### 2) 超声扫描、孕妇血清生化指标检查

- 仅为染色体病变的副现象；
- 可应用的妊娠期范围较短；
- 需结合多项指标综合考虑以达到临床要求的准确性和敏感性。

## 1.2 以往的检测方法及其缺陷(续)

### 无创产检（直接）：

3) 直接分离检查胎儿有核细胞

- 该方法难度较高。

基于游离胎儿DNA(cell-free fetal DNA, cffDNA)的发现提出如下几种方法：

4) RNA-SNP/表观遗传学等位基因比例法

- 利用等位基因的SNP检测测量胎盘特有的甲基化核酸分子；
- 仅能用于胎儿杂合位点的SNP分析；
- 需多位点的指标提高该方法的普及使用。

5) 多模态独立检测法(polymorphism independent methods)

- 利用高精度的数字PCR技术检测常见的非整倍型的染色体(如：21号染色体)的染色体相对量(relative chromosomal dosage, RCD)，并利用统计学方法分析是否存在异倍体情况；
- 该方法最大问题是需要大量的孕妇血浆样本供实验分析。

## 1.3 本课题的检测方法

### 基于大规模并行测序的无创产前检测法:

【方法优势】 克服上述方法中提及的多种缺陷:

- 无创检查可降低孕妇流产率;
- 直接测量染色体序列能提高诊断准确性;
- 无需测量特定染色体上的特定位点进而不需要大量的孕妇血浆样本。

【实验平台】 Illumina公司“Solexa”技术的GA系列测序仪

【分析软件】 生物信息学分析软件ELAND、统计学分析软件SigmaStat

# 目录

## CONTENTS

01

背景介绍

INTRODUCTION

02

研究方法

METHODS

03

相关结论

RESULTS

04

后续讨论

DISCUSSION



## 2.1 分析流程

### 2. 测序及比对：

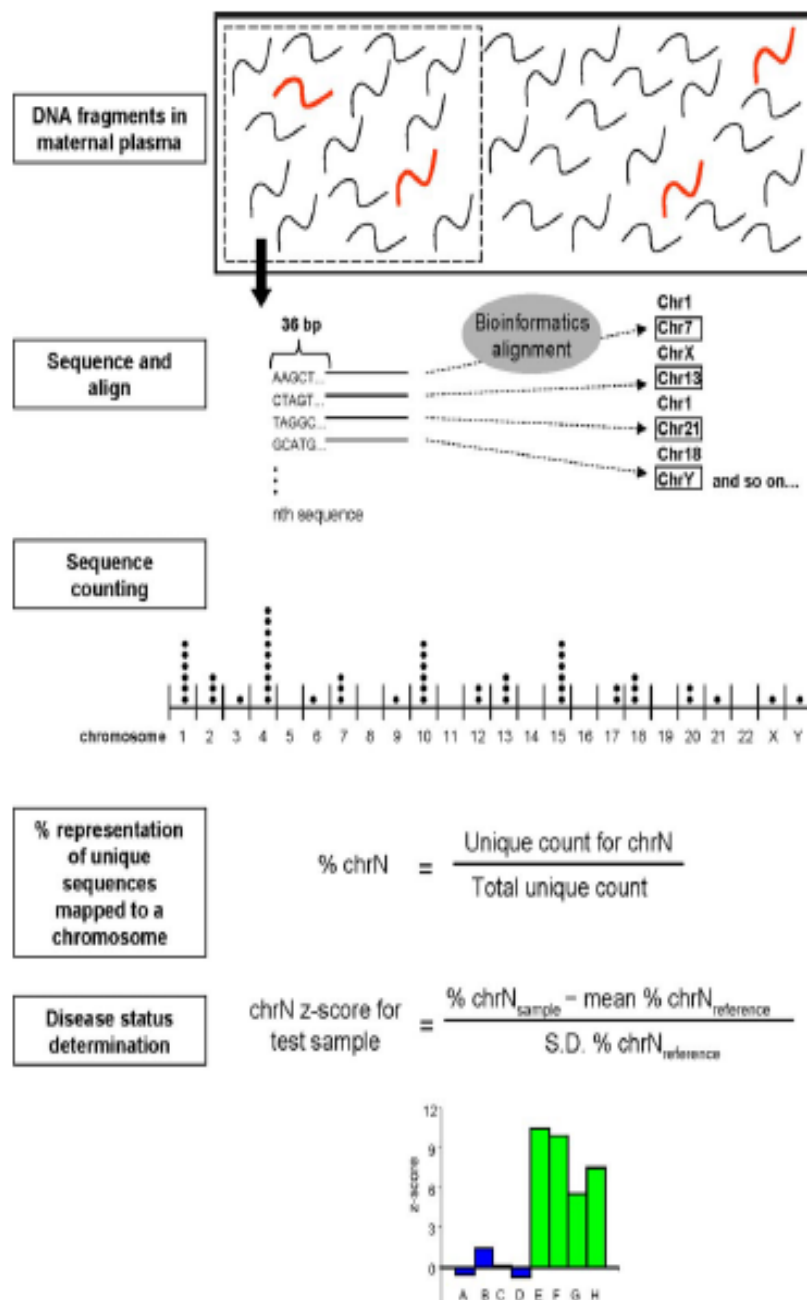
先利用Solexa的边合成边测序技术进行测序，再利用ELAND软件进行分析比对。

### 4. %chrN比值的计算：

分别计算每个染色体匹配上的U0-1-0-0序列数在总的U-0-1-0-0序列数的比值。

### 5. Z值的计算：

计算已确认正常的多组胎儿%chrN的平均数和标准差作为参考，再计算未知特征的染色体的Z值=(该染色体%chrN - 参考平均值)/参考标准差。



### 1. 提取孕妇血浆中的DNA片段：

其中加粗红色片段指胎儿DNA。

### 3. 序列统计：

统计匹配到相关染色体上的唯一并完美匹配的序列数（即该序列只匹配到去重复的人类参考基因组的一个位置且无错配，文中该序列用U0-1-0-0表示）。

### 5. 疾病确认：

通过Z值判断该染色体是否发生异倍体现象，Z值越大，表明越有可能。

## 2.2. 假设验证

上述的流程若能成立，需满足如下的几个假设：

- 1) 在以大量孕妇DNA为背景的孕妇血浆中，大规模并行基因测序(Massively Parallel Genomic Sequencing, MPGS)能够敏锐地捕捉并产生仅有少量的胎儿DNA片段序列(sequence reads)；
- 2) 用于测序的孕妇血浆样本中的各DNA片段能够代表孕妇体内的血浆中各DNA片段所占比例；
- 3) 测序时不存在对某一染色体的偏重，即假设成立情况下各染色体的%chrN比值能够准确代表孕妇与胎儿的染色体综合在血浆中的比重；
- 4) 在上述3)假设成立基础上，还需满足各染色体的%chrN比值与该染色体在人类基因组中所占比重呈正相关。

## 2.2.1 验证假设一

**【验证思路】** 如果MPGS能够测序孕妇血浆中的胎儿DNA，那么该方法就应该可以检测到怀有男婴孕妇血浆中的Y染色体。

**【验证方法】** 利用Illumina提供的 beta ChIP-Seq方法(染色体免疫共沉淀测序法，是高通量测序技术发展出的方法可用于检测含量极少的核酸序列，适合于本验证实验)，原流程中涉及“扩增-电泳-再扩增”这一关键步骤。

**【实验样品】** 三个已知怀男婴孕妇血浆样品和一个已知怀女婴孕妇血浆样品。

**【实验结果】** 该方法能够检测到怀男婴孕妇血浆中的Y染色体，具体结果如下：

Sample	Fetal sex	GA (weeks + days)	Input DNA (ng)	Karyotype	Total sequenced counts	Total U0-1-0-0 counts	U0-1-0-0/total sequenced counts (%)	ChrY U0-1-0-0 counts	% chrY <sup>†</sup>
3009	M	17+3	20.9	46XY	10,643,413	1,990,068	18.7	636	0.032
3034	M	17+5	11.3	46XY	9,578,522	1,792,238	18.7	858	0.048
3143	M	17+2	16.8	46XY	9,549,403	1,885,644	19.7	1,054	0.056
3735	F*	41+2	92.6	NA	9,087,258	1,968,465	21.7	177	0.009

M, male; F, female; GA, gestational age; NA, not available.

\*Confirmed to be a healthy female fetus at birth.

<sup>†</sup>Percentage of U0-1-0-0 sequences mapped to chromosome Y among all U0-1-0-0 sequences.

**【实验分析】** 不仅怀男婴孕妇血浆中能检测到Y染色体片段，怀女婴孕妇血浆中也能检测到极少量Y染色体片段，考虑可能在电泳时存在污染。

## 2.2.1 验证假设一(续)

**【误差回顾】** 不仅怀男婴孕妇血浆中能检测到Y染色体片段，怀女婴孕妇血浆中也能检测到极少量Y染色体片段，考虑可能在电泳时存在污染。

**【验证方法】** 将原beta ChIP-Seq方法流程(含“扩增-电泳-再扩增”步骤)设为方法B，去除“电泳-再扩增”这两步仅保留首次“电泳”设为方法A。

**【实验样品】** 等量的怀女婴孕妇血浆DNA(1)、怀男婴孕妇血浆DNA(2)、两成年男性混合血浆DNA(3)各一份，并等分成两份分别用于方法A和B的检测。

**【实验结果】** 怀女婴孕妇血浆中仍能检测到Y染色体片段，且方法A测得的序列数比B的要多，具体结果如下：

Sample	Fetal sex	GA (weeks + days)	Input DNA (ng)	Total sequenced counts	Total U0-1-0-0 counts	U0-1-0-0/total sequenced counts (%)	ChrY U0-1-0-0 counts	% chrY†
1A	F	38 + 4	50	9,194,718	2,096,366	22.8	184	0.009
1B	F	38 + 4	50	10,030,317	2,012,469	20.1	218	0.011
2A	M	38 + 4	50	8,560,897	2,187,094	25.5	1,444	0.066
2B	M	38 + 4	50	9,556,038	2,093,089	21.9	1,615	0.077
3A	NA*	NA*	50	8,573,956	2,010,318	23.4	3,523	0.175
3B	NA*	NA*	50	9,614,592	2,057,491	21.4	3,468	0.169

**【实验分析】** 将污染完全归结于电泳的污染并不成立，需进一步验证。

## 2.2.1 验证假设一(续)

**【误差回顾】** 在使用新的方法A的基础上，仍能在怀女婴孕妇血浆DNA中检测出Y染色体片段，分析是ELAND软件在匹配上存在错误导致的。

**【验证方法】** 将含Y染色体在内的所有染色体匹配上的U0-1-0-0序列，用于NCBI的BLAST进行比对，观察比对匹配到相关染色体上的序列数及唯一匹配的序列数，并计算两者的比值。

**【实验数据】** 上次实验中得到的三组两种方法的六套数据。

**【实验结果】** 在常染色体和X染色体上唯一匹配的序列数和总匹配上的序列数的比值高达97%以上，而在Y染色体上(2)(3)组的比值高达90%以上，怀女婴的(1)组仅30%左右但仍存在相关的唯一匹配的序列。

**【该一系列的验证实验结论】** 上述的多次实验可证明假设一：MPGS能够捕捉并产生孕妇血浆中的胎儿DNA，但是在利用ChIP-Seq方法进行检测时使用去除电泳步骤的方法A可减少污染，同时软件的错配以及其他未知因素(在4.1部分会进一步讨论)会造成微小误差，但是不影响分析结果。

## 2.2.2 验证假设二~四

**【验证思路】** 为验证血浆样品能否代表孕妇体内的血浆成分，与MPGS的测序结果中各染色体比重能否与人类基因组中各染色体比重呈相关性，可通过获得已知的整套人类基因组中各染色体的相对大小值作为参考，与测序得到的样本中各染色体%chrN比值进行比较分析。

**【验证方法】** 从Ensembl数据库中下载去重复的人类女性全套基因组数据(由于样本中的DNA以女性为主)，计算各个染色体的碱基数及其与总碱基数的相对值，并将这些相对值与样本中的%chrN做线性回归分析。

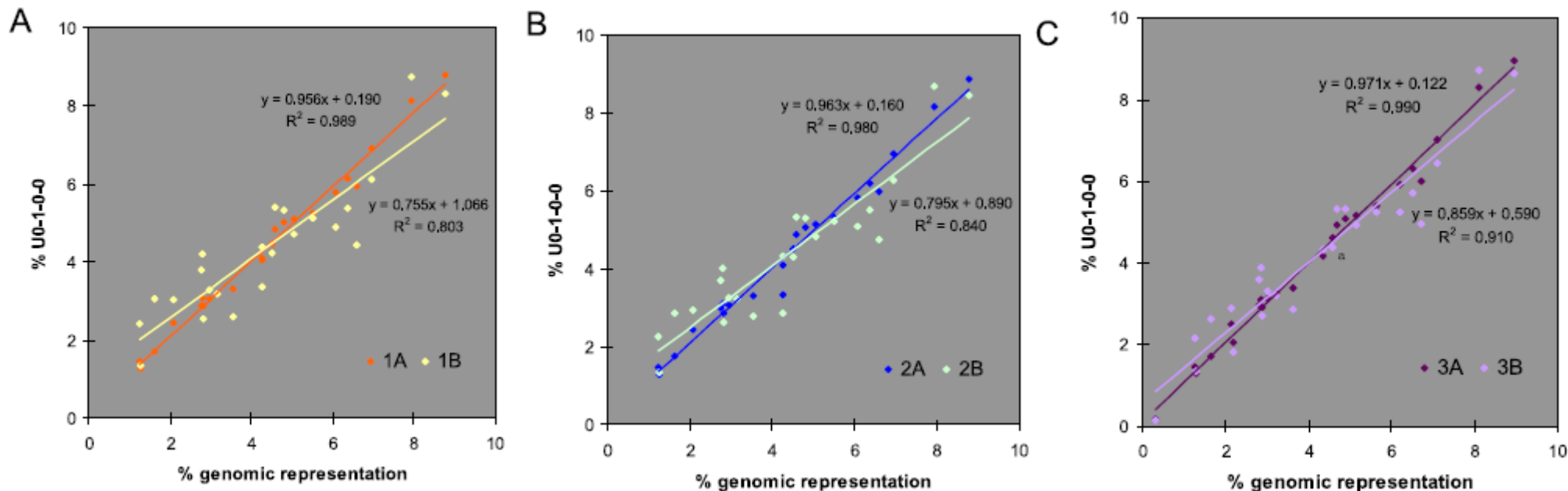
**【实验数据】** 验证假设一中使用的三组两种方法的六套数据。



## 2.2.2 验证假设二~四(续)

**【实验结果】** 用方法A测得的样本中各染色体比重分布比用方法B测得的，更加接近于参考的人类基因组中各染色体的分布情况；从线性回归分析中也能看出：样本1A,2A,3A的斜率均  $> 0.95$ ， $R^2$ 均  $> 0.98$ ，而样本1B,2B,3B中的斜率最大值也仅为0.91，具体结果如下。

**【该验证实验结论】** 该实验再一次证明方法A比方法B更加准确；更重要的是，该实验成功验证假设二~四：孕妇血浆中的各染色体比重能够反映人类基因组中各染色体的相对值；而且孕妇血浆中的DNA成分不仅和女性而且和男性的血浆中DNA成分分布相似。



## 2.3 T21胎儿的检测实验

**【实验前提】** 通过上述的系列实验和数据分析，成功验证了四个前提假设。再根据实验流程，应计算样本的Z值来检测胎儿的染色体非整倍体情况。

**【理论基础】** 胎儿染色体数目异常会使母体血浆中DNA含量产生微量变化，理论上通过区分这一微小的差异，便可以实现胎儿染色体疾病的产前检测。

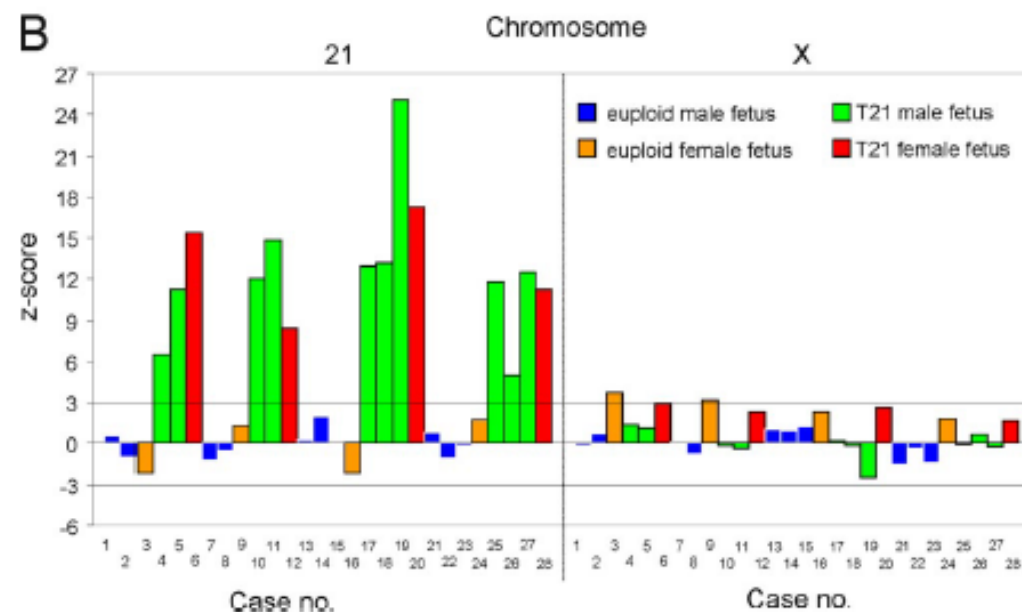
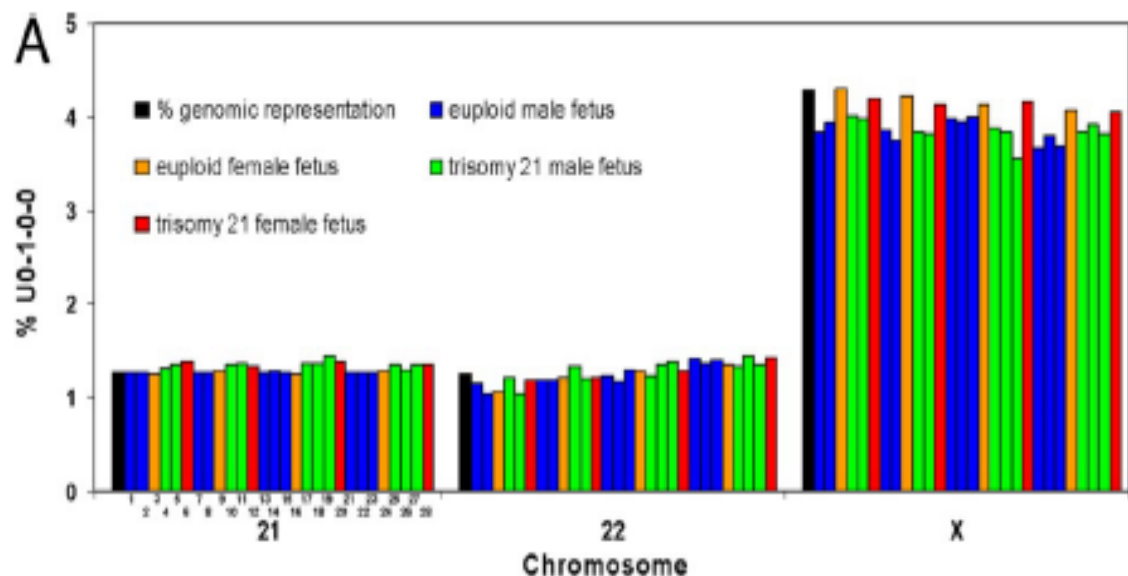
**【实验样品】** 28个已知胎儿表型的妊娠中期孕妇血浆样本，其中14例为正常胎儿，剩余14例为T21(唐氏综合征)胎儿。

**【实验方法】** 利用新的流程A对28个样本进行MPGS测序，计算各个样本的各个染色体的%chrN比值。取其中正常的10个怀男婴孕妇血浆样本数据，求各个染色体的%chrN比值的平均数(mean)和标准差(Standard Deviation, SD)作为参考值。以上述参考值为基础，求28个样本的各个染色体的Z值，加以分析判断(除Y染色体之外)。



## 2.3 T21胎儿的检测实验(续)

【实验结果】 由于参考值选择的是男婴的数据，所以女婴X染色体的%chrN比值相对较高且Z值均  $> 1.67$ ；另外，所有T21胎儿的21号染色体Z值较高均  $> 3$ ，说明该方法能够敏锐地检测到T21孕妇血浆中染色体的变化；而上述未提到的染色体的Z值均  $< 3$ 。具体结果如下：



## 2.4 检测比例的可重复性

**【实验现象】** 从T21的检测实验数据中可发现：T21胎儿的21号染色体%chr21的绝对偏差相对较小但是z值却很大，相反的是%chrX的绝对偏差较大而Z值却很小。

**【实验分析】** 为判断后续的实验可重复上述有趣的现象，作以下的分析：

- 1) 女婴和男婴之间的X染色体之间的绝对偏差大于T21和正常胎儿之间的21号染色体之间的：女性的chrX含量是男性的两倍，而T21的chr21是正常的1.5倍；
- 2) Z值反映的是待测值与参考平均值之间的偏差程度的标准化结果(以标准差为参考)。由于标准差反映的是各染色体多次测序的准确度，故利用参考数据集的10个正常男婴的数据求变动参数(coefficient of variation,  $CV=SD/mean$ ),发现21号染色体的CV值的确很小，而X染色体的相对较大。

进一步地，将CV和各个染色体GC含量进行方差分析(ANOVA)，发现两者之间存在统计学意义上的显著性差异( $P < 0.001$ )。说明染色体的标准差值与该染色体自身特点有关。

# 目录

## CONTENTS

01

背景介绍

INTRODUCTION

02

研究方法

METHODS

03

相关结论

RESULTS

04

后续讨论

DISCUSSION

### 3 结论总结

通过上述所有的实验及分析，可得出如下的结论：

- 1) MPGS可以作为无创产检的诊断方法，用于检测胎儿的染色体非整倍体情况。T21胎儿的孕妇血浆中的染色体异常能够清晰明了地被探测；
- 2) 同时，通过XY染色体含量分析胎儿的性别也具有很强的鲁棒性；
- 3) 将Illumina提供的ChIP-Seq流程中删除“电泳-再扩增”步骤，得到的测序结果无论在数量和准确性都较高；
- 4) 该检测方法有别于其他方法对特定基因位点的探测，仅利用样本血浆中染色体微量的变化就能检测出胎儿染色体数异常，十分重要和喜人。

# 目录

## CONTENTS

01

背景介绍

INTRODUCTION

02

研究方法

METHODS

03

相关结论

RESULTS

04

后续讨论

DISCUSSION

## 4.1 文章延伸

课题组在得出重要结论同时，还提出了如下的探讨内容：

**【样本采样时期】** 实验中的样本主要在妊娠中期进行采样(14-15周左右)；其中11例怀T21胎儿的孕妇在进行有创检测确诊后取样，不过先前的研究表明进行穿刺后数天采样所得的样本中，胎儿DNA含量无显著差异。

**【提高检测准确性】** ① 理论上，通过检测胎儿的DNA含量可提高准确性，但所需样本量较大且原始的数据已经非常准确足够支撑得出检测结论；②通过优化测序后的数据分析方法可提高准确性，比如考虑拷贝数变异、唯一但含一两个错配的序列数等。

**【软件对chrY的错误分析】** 存在少量这种现象与Y染色体中含有大量的重复序列、MPGS的高精度、实验中不可避免的污染和已知人类基因组中存在的未知序列空隙等有关，但这不影响对胎儿性别的判定。

**【测序费用】** 当前用MPGS测序价格有些昂贵，但这只是暂时的现象随着技术进步会有所下降，况且使用其他无创方法也未必更便宜。

**【同行工作】** Fan et.al. 同样用MPGS方法进行无创产检研究，但是他们的研究存在样本数量少、正常胎儿的妊娠周期比患儿的较短、采样时间不佳等缺陷。

## 4.2 个人想法

本课题利用先进的MPGS技术开展无创产前检查，并成功地诊断出相关患病的样例，具有开创性和很好的应用前景。

鉴于能力有限，虽未发现本研究的问题，但就本课题中的部分方法存在疑问：

- 1) 假设二提出要求样本中DNA成分含量能够代表孕妇体内的含量。该假设依据常识虽然显而易见且验证难度较高，但是本课题中只是通过验证测序后的%chrN和参考基因组中各染色体含量分布相当，进而可推断测序不存在偏差，但仍**不能很好证明假设二**。根据个人理解，是否应该比较孕妇不同部位的静脉血和体内血液的情况来验证假设，当然这在实际操作中可行性不大；
- 2) 理论上，在数据分析时测试集和训练集(本课题为参考数据集)应该是分别独立的，但是本课题中的训练集是测试集的一部分，显得不够严谨；
- 3) 不是十分理解**为什么一定要采用唯一完美匹配的序列(U0-1-0-0)**，如果不采样这样的序列结果会如何呢！

**Thank you !**