Noninvasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing

——通过大规模测序技术开展胎儿全染色体异倍体的无创检查

# 研究背景

由于和前两篇文献(Chiu RW, Chan KC, Gao Y, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci USA 2008;105:20458–63. / Fan HC, Blumenfeld YJ, Chitkara U, et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci USA 2008;105:16266–71. )的研究课题相似，所以涉及的相关研究背景相似，重复内容不再赘述。额外需提及的内容有：

由于本课题研究胎儿全染色体异倍体的检查，所以除了常见的常染色体异倍体（如21号、18号、13号）之外，还有出现许多性染色体异倍体的病例，如Turner综合征(45,X)、Klinefelter综合征(47,XXY)、XYY综合征(47,XYY)、X三体综合征(47,XXX)；还会出现不常见的常染色体异倍体现象(本文中的病例是T9)。

另外，本研究主要使用了上述两文献涉及的方法，同时对他们存在的部分问题进行了改进与补充。

# 研究方法

## 研究样本

本研究中使用的测试样本有435例，其中涉及多种染色体异倍体的类型，同一染色体异常也尽可能地使用到不同的类型；与此同时，使用了50例怀女胎的孕妇血浆样本作为参考组。

## 游离血浆DNA测序

本研究使用Illumina HiSeq2000平台和Chiu et al.中提及的改进版ChIP Seq测序流程对样本进行测序，同时在每次测序时增加已知核型(D21、T21)的参考样本。

## 数据分析

### NCR值的计算

标准化染色体表达量(Normalized chromosome representation value, NCR value)=该染色体上的唯一比对序列/样本中所有常染色体的唯一对比序列，同时计算相关的GC含量。

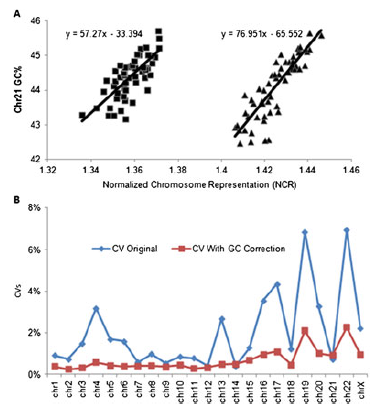
### NCR值的修正

**NCR值与GC含量的线性关系：**通过Chiu et al.和Fan et al.的研究可知NCR值和相应的GC含量呈很强的相关，且每次实验所得到的实际NCR值(GC含量)波动情况和所在染色体固有的GC含量、人为实验因素等有关。同时，NCR值的波动情况(可由变动参数CV = SD/mean表示，mean和SD分别代表所有样本在该染色体上NCR值的平均数和标准偏差)与检测染色体异倍体的难易程度有关。

**NCR值的修正降低变动系数：**在本实验中，通过对两组已知核型的参考样本作NCR值和GC含量的线性回归分析（见下图A），所得结果除14号染色体外均符合上述的结论。由于原始的NCR值对应的变动参数过大，所以使用由50例怀女胎的孕妇血浆参考样本对测试样本中的除Y染色体外的所有染色体进行GC含量偏差修正，修正后的变动参数确实有明显改善（见下图B）。修正公式如下：

NCRgc=NCR-(GC-GCaverage\_ref)/Sloperef

其中，NCR和NCRgc分别表示修正前后的NCR值，GC表示测试样本的GC含量，GCaverage\_ref和Sloperef分别是参考样本的GC含量平均值和相关线性回归分析的斜率。



### Z-cutoff值的计算

**Z值计算：**Zij=(Xij-meanref)/SDref 其中Zij为样本j的第i号染色体的Z值，Xij为样本j的第i号染色体的NCRgc值，meanref和SDref分别为50例怀女胎的孕妇血浆参考样本的第i号染色体的NCRgc值的平均值和标准差。

Chiu et al.中通过对各样本中各染色体的Z值判断筛选出T21的病例，且得到21号染色体的Z-cutoff值为3，本研究也采用该判定标准。另外，由于本研究需分析全染色体的异倍体情况，故除21号染色体外的其他染色体的Z-cutoff值以21号染色体的Z-cutoff值为参考计算可得。*文中反复强调Z-cutoff值是相对于胎儿DNA比重为5.52%而言的，但是该数据的计算过程和原始数据并未给出，且对相关Z-cutoff值的计算并没有影响。*

文中和参考材料中并没有给出相关Z-cutoff值的计算方法，但是经过研究得到如下：

(M21-mean21)/SD21 = Z21-cutoff → M21=mean21+SD21\*Z21-cutoff

已知参考样本中各染色体NCR值对应的平均值和标准偏差，和21号染色体的Z-cutoff值，可以计算出21号染色体正常含量下最大的NCR值M21。

Mx = M21\*meanx/mean21

通过x号染色体的平均值meanx和21号染色体平均值mean21的比例，求出x号染色体正常含量下最大的NCR值Mx。

Zx-cutoff = (Mx-meanx)/SDx

然后通过Z值求解的公式得到X号染色体的Z-cutoff值。经过验证，本计算Z-cutoff值的方法为本研究所采用的方法。

另外，由于参考样本均为怀女胎的孕妇血浆样本，所以相关的Y染色体Z-cutoff值并不能通过上述方法得到，文中直接给出了Y染色体的Z-cutoff值为3。

## 测试样本的检测实验

### 常染色体的异倍体检测

测试样本中出现异倍体情况的染色体为21号、18号、13号和9号，对应的Z-cutoff值分别是3、5.91、5.72和7.45。

#### 21染色体的异倍体检测

经分析可知，T21的病例的检测灵敏度(sensitivity)和特异性(specificity)为100%。

#### 其他染色体的异倍体检测

经分析可知， T18的病例的检测灵敏度(sensitivity)和特异性(specificity)为100%，T13病例的检测灵敏度为100%、特异性为99.75%(407/408)，假阳性样本为Turner综合征 (45,X)，其Z13值为5.81；只有一例的T9病例也能被成功检测出。

#### 染色体结构异常（罗伯逊易位）检测

T21中有两例罗伯逊易位的病例能够被成功检测出。

#### 嵌合体检测

T21中有一例嵌合体的病例能够被成功检测出。

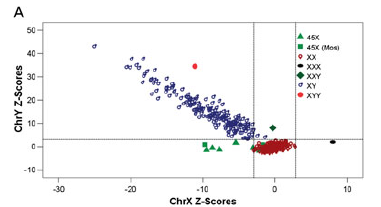
#### 双胞胎检测

T13中存在一组双胞胎的病例能够被成功检测出，不过在T13中Z13值最小，为7.17。

### 性染色体的异倍体检测

Y染色体和X染色体的Z-cutoff值分别是3和2.91。利用ZY-cutoff值对测试样本进行胎儿性别的分类，分类结果的灵敏度和特异性均为100%；利用ZX-cutoff值对测试样本进行胎儿性别的分类，得到的结果是灵敏度为100%,但是将两个男胎错分成女胎，特异性为99.51%。

研究组发现，Y染色体和X染色体的Z值存在相互关系，对不同样本的Y染色体和X染色体作散点图（如下图A），得到的结果基本符合预期：大多数正常样本集中在X轴的-2.91~+2.91之间，Y轴的-3~+3之间；5个Turner综合征被成功检测出（灵敏度为100%），将一个正常样本错分成Turner综合征（特异性为99.75%）；一个XXY病例、一个XYY病例和一个XXX病例均能被成功地检测出。



# 研究结论及延伸

以高通量测序为基础的无创产前诊断方法的确能很好地检测出胎儿的染色体异倍体情况。

本研究同样验证了染色体表达量和相应的GC含量存在线性关系，并且相关的GC含量对于染色体的检测有一定影响。过去的研究虽然能够揭露该现象但是并没有提出很好的改进方法，本研究中采用参考样本对测试样本进行修正，很好的降低了相关的变动参数，也提高了除21号（该染色体的检测率原本就很高）2外的其他染色体的检测灵敏度和特异性。

文中涉及的GC修正算法同样可以应用于不同的测序平台和不同的试剂，只是需要重新进行相关参考样本的测序分析。