

二代测序技术的技术评估

——以大规模平行测序技术为例

卫生部卫生技术评估重点实验室（复旦大学）

黄葭燕 教授

卫生技术评估报告指南（初稿）

- ◆ 政策问题
- ◆ 卫生技术评估报告的方法学
- ◆ 背景信息
- ◆ 研究问题
- ◆ 技术评估结果
- ◆ 讨论
- ◆ 研究结论
- ◆ 政策建议

第一部分 政策问题

- **大规模平行测序技术（massively parallel sequencing, MPS）**

- 利用对母体外周血中的游离DNA进行深度测序，以获取胎儿染色体信息，来检测胎儿染色体非整倍体的二代测序技术

- 截至目前，该技术在中国尚属临床试验阶段

- **2014年2月14日——《关于加强临床使用基因测序相关产品和技术管理的通知》**

- 要求在相关的准入标准、管理规范出台以前，任何医疗机构不得开展基因测序临床应用，已经开展的要立即停止

- **2014年12月31日——国家卫计委出便函（国卫医护便函[2014]407号）及《高通量基因测序技术临床应用试点工作方案》**

- 宣布在国内开展高通量基因检测技术临床应用试点工作

第一部分 政策问题

- MPS技术的安全性与有效性到底如何？
- 政府部门应如何正确引导和规范该技术的临床应用？
- 在这类新技术的发展过程中，应注意些什么？

第二部分：卫生技术评估报告的方法学

----卫生技术评估计划

- 系统综述

- 通过系统综述及Meta分析，总结目前我国二代测序技术应用于唐氏综合征产前检测的实施情况，由两位作者独立地筛查所纳入文献研究以确定最终满足所有纳入标准的文献，评价其安全性和有效性

- 关键知情人访谈

- 运用质的研究方法中的扎根理论设计访谈提纲，与北京协和医院、深圳市妇幼保健院、华大基因公司、广东省妇幼保健院的相关专家、技术人员对大规模平行测序技术的应用情况开展了知情人访谈

第二部分：卫生技术评估报告的方法学

-----系统综述与Meta分析

• 数据来源

- 纳入文献发表于1997年1月1日至2014年11月7日
- 各大医学数据库包括Cochrane Library、Medline、EMBASE、ISI Web of Science和Biosis previews，中文数据库包括万方数据库、中国知网数据库和中国生物数据库和医学文献数据库
- 搜索未设定语言、国别等任何限制
- 追查所有纳入文献的参考文献

纳入标准

- ✓ 原始研究（例如：随机对照试验，非随机对照试验，队列研究，病例对照研究）
- ✓ 研究对象为怀孕妇女（不限制怀孕年龄，或是否使用辅助生殖技术）
- ✓ 研究检测技术为大规模平行测序技术（**massively parallel sequencing, MPS**）
- ✓ 研究技术用于检测胎儿21-三体综合征（唐氏综合征）风险
- ✓ 参考标准为基于以下一种或多种诊断技术所得的诊断结果，诊断技术包括：绒毛膜活检或羊膜穿刺术所得染色体核型分析结果，胎儿尸检结果或新生儿随访结果
- ✓ 研究汇报了精确数据或计算度量值以完成2 x 2四格表，或汇报检测技术所存在的风险及有害效应的数据值

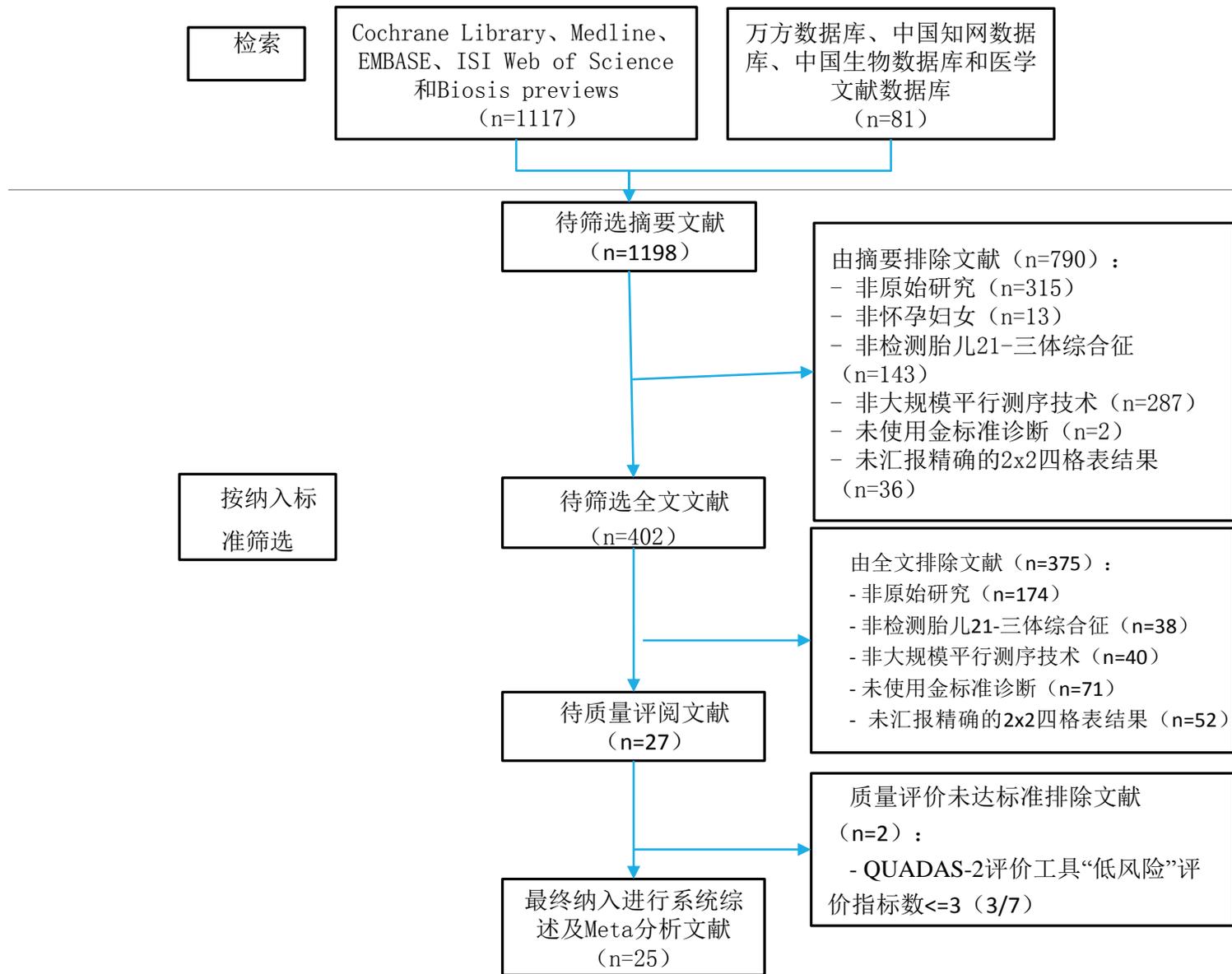
- 排除标准

- ✓ 综述，模拟或模型研究，编辑评论，文献评阅，信件或会议摘要
- ✓ 研究未汇报检测结果的具体数据结果
- ✓ 仅以评价“非侵入性检测安全性”为目的的原始研究
- ✓ 对多种筛查方案进行两两对比，但没有设定统一的参照标准（如羊膜穿刺、绒毛膜取样等）

- 参照标准（金标准）

- ✓ 理想的参照标准是胚胎组织细胞核型分析；分娩后的临床检查也是本研究所认可的参照标准

文献纳入排除流程图



• 资料提取方法

- 研究由两位作者独立地筛查所纳入文献摘要、全文以确定最终满足所有纳入标准的文献
- 研究设计了资料提取表对文献信息进行提取
- 提取的资料主要包括: 文献的基本信息、研究对象的纳入标准、大规模平行测序技术的开展时间、“金标准”诊断方法、真阳性例数、假阳性例数、假阴性例数、真阴性例数、敏感度和特异度等

• 研究证据的质量评价

- 由两名系统评价员使用QUADAS-2工具对所纳入文献进行质量评价
- 共有四个质量评估域：病例的选择，待评价测试，金标准，病例流程和进展情况
- 评分等级分为“高偏倚风险”、“低偏倚风险”、“偏倚风险不确定”、“高临床适用性”、“低临床适用性”以及“临床适用性不确定”
- 制定了针对二代测序技术应用于唐氏综合征检测的理想研究特征列表

应用MPS检测21-三体综合征理想研究特征

低风险研究特征

- 入组孕妇为连续性样本或随机样本
- 不采用病例对照研究
- 不根据金标准结果来排除入组对象
- 在开展数据分析前已制定截断值或阈值
- 所有孕妇均采用金标准检测验证，所采用的金标准包括：羊膜穿刺术后核型分析、绒毛膜绒毛取样后核型分析、出生后核型分析（随访）
- 所有孕妇均纳入数据分析

高临床适用性研究特征

- 孕妇为21-三体综合征高风险人群，且研究目的在于比较大规模平行测序技术与金标准检测之间的准确性
- 在侵入性诊断技术前行大规模平行测序技术
- 研究中仅使用一种截断值，或者两个截断值不连续时，落入不连续区间的病例不超过受试孕妇的5%

● 数据分析

- 效果指标：合并敏感度(**pooled sensitivity**)、特异度(**pooled specificity**)、阳性似然比(**pooled positive LR**)、阴性似然比(**pooled negative LR**)、合并诊断比值比(**pooled diagnostic OR**)
- 使用**Review Manager 5.3.5**软件绘制证据质量评价结果图并使用**Stata12.1**进行Meta分析
- 异质性分析：如果**P < 0.05**，则各纳入研究间存在异质性，使用**Stata**软件中**Midas**模块混合效应模型计算合并值

第三部分：背景信息

- 唐氏综合征——21-三体综合征或先天愚型
 - ICBDSR数据显示，2006年新生儿唐氏综合征发生率中位数为17/万人
 - 我国尚无准确的唐氏综合征出生缺陷网络监测数据，2012人群调查数据显示我国唐氏综合征在新生儿中的发生率约为14.7/万人，每年新增唐氏综合征2.3~2.5万例
- 目前应用最为广泛的中孕期三联血清学筛查技术仍然存在较低的检出率(70%)和较高的假阳性率(5%)
- 急需一种更安全、快速、准确、易标准化的检测方法，以提高现有产前筛查、产前诊断体系的工作效率和质量

第三部分：背景信息

- 胎儿染色体非整倍体的无创产前检测 (non-invasive prenatal testing, NIPT)
 - 产妇血液中存在胎儿游离DNA (cffDNA) 和胎儿游离RNA (cffRNA), 胎儿染色体异常会引起游离DNA含量的微量变化
 - 常用技术
 - 选定区域数字化分析技术 (digital analysis of selected regions, DANSR)
 - 大规模平行测序 (massively parallel sequencing, MPS)
 - 单核苷酸多态性 (single nucleotide polymorphism, SNP)
- 二代测序技术因其本身的检出率高、影响因素少和无创等特点, 被公认为促进了无创产前基因诊断的发展

技术的适用范围

- 适用人群：血清学筛查、影像学检查显示临界风险(即 $1/1000 \leq$ 唐氏风险值 $< 1/270$) 的孕妇; 有介入性产前诊断禁忌证者; 血清学筛查最佳时间, 或错过常规产前诊断时机
- 慎用人群：产前筛查高风险, 预产期年龄 ≥ 35 岁的高龄孕妇, 以及有其他直接产前诊断指征的孕妇; 孕周 < 12 周的孕妇; 高体重(体重 > 100 千克) 孕妇; IVF-ET 方式受孕的孕妇; 双胎妊娠的孕妇; 合并恶性肿瘤的孕妇
- 不适用人群：染色体异常胎儿分娩史; 孕妇 1 年内接受过异体输血等对高通量基因测序产前筛查与诊断结果将造成干扰的; 胎儿影像学检查怀疑胎儿有其他染色体异常可能的; 各种基因病的高风险人群

第三部分：背景信息

• 国内开展现状

- 国内学者在国际期刊上发表研究已多达**16**篇，与国际前沿水平齐头并进
- **2008**年，深圳市妇幼保健院首次与中国华大基因合作；**2013**年至**2014**年，深圳市妇幼保健院在华大基因完成了**5000**多例检测
- 华大基因公司已在临床上完成**40**万份样本检测
- 广州市妇幼保健院，在院内自购了无创检测设备，自**2012**年下半年起共完成**6000**多例高龄或高风险孕妇的唐氏无创产前检测
- 华大基因所做的检测，由于单次检测上机的样本量大且存在一定资金支持，目前所收费用在**500元-1000元**不等

第四部分：研究问题

- MPS技术的检测效能（检出率、假阳性率、阳性预测值、阳性阴性似然比等）
- 现有的证据水平和质量如何？
- 现有证据还存在哪些空白？
 - 对筛查结果（健康效应、健康风险和副作用）的测量是否全面？指标是否恰当？
 - 影响因素（筛查对象特征、服务提供方的培训和技术水平、其他影响因素等）对不同筛查策略的作用

第五部分：技术评估结果

安全性与有效性评估结果

资料基本信息

- 共获取**1401**篇文献，其中**Medline 215**篇、**Cochrane Library 233**篇、**Embase 456** 篇，**Web of Science 218**篇，**Biosis previews 198**篇，中文文献库**81**篇，查重后余**1198**篇
- 最终纳入全文**25**篇
- 时间跨度从**2011**年至**2014**年

纳入文献的研究特征

- 文献的发表时间:

- 纳入的25篇文献发表时间为2011~2014年。各年度纳入文献数量显示，二代测序技术作为新兴技术，近年来相关研究数量明显增加，2014年纳入文献数量为11篇



各年度纳入文献数量 (篇)

研究设计

- ✓ 92 % (23/25) 的文献采用了前瞻性队列研究设计， 1项研究采用了病例对照研究设计， 另有一项研究未汇报研究设计类型

- ✓ 7项研究在美国开展， 13项研究在中国开展， 2项研究在中国香港开展， 1项研究在中国台湾开展， 1项研究在德国开展， 1项研究在日本开展
- ✓ 36 % (9/25) 为单中心研究， 52 % (13/25) 为多中心研究， 3篇文献未报告研究中心数量

研究背景

研究对象

- ✓ 13项研究的研究对象是单胎妊娠的孕妇， 1项研究的研究对象是双胎妊娠， 3项研究的研究对象是既有单胎又有双胎妊娠孕妇， 另有8项研究未报告研究对象

表 1 使用MPS检测唐氏综合征纳入文献研究特征

编号	研究	阳性病例数 (例)	阴性病例数 (例)	孕周	孕妇纳入标准	研究类型	敏感度 (95%CI)	特异度 (95%CI)
1	Bianchi, D. W.	5	1898	8-39.4	孕早期、中期 已做唐筛	前瞻性队列	100.00% (47.8%-100%)	99.68% (99%-100%)
2	Bianchi, D. W.	89	404	10-23	计划行侵入性 诊断	前瞻性队列	100.00% (95.9%-100%)	100.00% (99%-100%)
3	Bianchi, D. W.	30	82	10-21	高风险	前瞻性队列	100.00% (88%-100%)	98.80% (93%-100%)
4	Dan, S.	139	2819	9-28	9-28周单胎妊 娠孕妇	前瞻性队列	100.00% (97%-100%)	99.96% (99%-100%)
5	Ehrich, M.	39	409	8-36	高风险	前瞻性队列	100.00% (89%-100%)	99.76% (99%-100%)
6	Jiang, F.	16	887	10-34	?	前瞻性队列	100.00% (79%-100%)	100.00% (100%-100%)
7	Lau, T.K.	11	97	11-28	连续性随机入 组	前瞻性队列	100.00% (72%-100%)	100.00% (96%-100%)
8	Lau, T.K.	23	1959	?	孕周>12周	前瞻性队列	100.00% (85.7%-100%)	100.00% (100%-100%)
9	Palomaki	209	1468	?	高风险	病例对照	98.60% (96%-100%)	99.80% (99%-100%)
10	Bianchi, D. W.	89	443	8-31	高风险； 计划行侵入性 诊断	前瞻性队列	100.00% (96%-100%)	100.00% (99%-100%)

表 1 使用MPS检测唐氏综合征纳入文献研究特征 (续1)

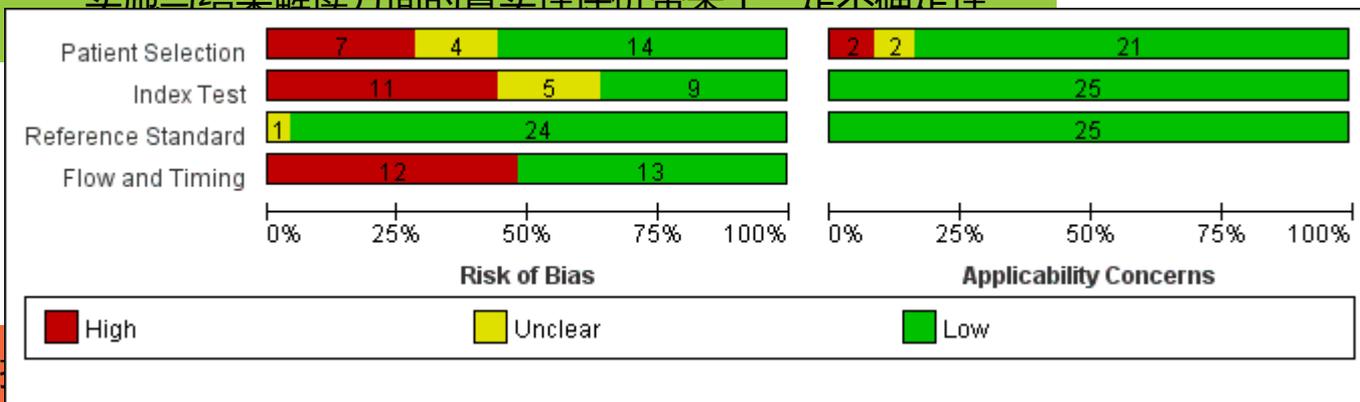
编号	研究	阳性病例数 (例)	阴性病例数 (例)	孕周	孕妇纳入标准	研究类型	敏感度 (95%CI)	特异度 (95%CI)
11	Shaw	11	189	?	孕周>12周	前瞻性队列	100.00% (72%-100%)	100.00% (98%-100%)
12	Song	8	1733	11-21.9	孕妇年龄<35岁	前瞻性队列	100.00% (63%-100%)	100.00% (100%-100%)
13	Stumm	40	430	11-32.1	高风险; 计划行侵入性诊断	前瞻性队列	100.00% (91%-100%)	99.54% (98%-100%)
14	Yu	36	88	?	?	?	100.00% (90%-100%)	100.00% (96%-100%)
15	Huang, X	9	180	11-36	双胎妊娠; 计划行侵入性诊断	前瞻性队列	100.00% (66%-100%)	100.00% (98%-100%)
16	Jensen, T	134	1134	?	高风险	前瞻性队列	100.00% (97%-100%)	99.91% (99%-100%)
17	Liang	40	372	15-39.9	高风险	前瞻性队列	100.00% (91%-100%)	100.00% (99%-100%)
18	邓璐莎	8	89	12-22	12-22周	前瞻性队列	100.00% (63%-100%)	96.74% (91%-99%)

表 1 使用MPS检测唐氏综合征纳入文献研究特征 (续2)

编号	研究	阳性病例数 (例)	阴性病例数 (例)	孕周	孕妇纳入标准	研究类型	敏感度 (95%CI)	特异度 (95%CI)
19	文萍	20	2716	12-26	12-26周	前瞻性队列	100.00% (83%-100%)	99.93% (99%-100%)
20	曹婧	2	101	16-20	高风险	前瞻性队列	100.00% (16%-100%)	100.00% (96%-100%)
21	梅瑾	8	1223	12-28	高风险	前瞻性队列	100.00% (63%-100%)	100.00% (100%-100%)
22	卢守莲	12	987	>13	高风险	前瞻性队列	100.00% (74%-100%)	100.00% (100%-100%)
23	冯暄	20	347	14-20	?	前瞻性队列	100.00% (83%-100%)	100.00% (99%-100%)
24	汪淑娟	3	133	11-13.9	?	前瞻性队列	100.00% (29%-100%)	100.00% (97%-100%)
25	刘红彦	2	151	>14	高风险	前瞻性队列	100.00% (16%-100%)	100.00% (98%-100%)

偏倚风险

- ✓ 入选的文献在研究对象选择、筛检具体实施与结果解读以及病历流程和进展方面有较大可能产生偏倚
- ✓ 绝大多数入选研究文献在金标准的实施与结果解读方面的报告具有较高的真实性
- ✓ 病例选择和待评价试验域中分别有4篇和5篇文献被评估为“偏倚风险未知”，为入选文献在病例选择以及筛检具体实施与结果解读方面的真实性评价带来了一定不确定性



- ✓ 病例选择“高临床不适用性”的文献占比较高，分别占入选文献总数的84%，96%和100%
- ✓ 仅在病例选择域中出现了2篇“高临床不适用性风险”和2篇“临床适用性风险不确定”的文献
- ✓ 在待评价试验域中出现了1篇“高临床不适用性风险”的文献，提示入选研究文献在临床适用性方面报告的总体可信度较高

临床适用性

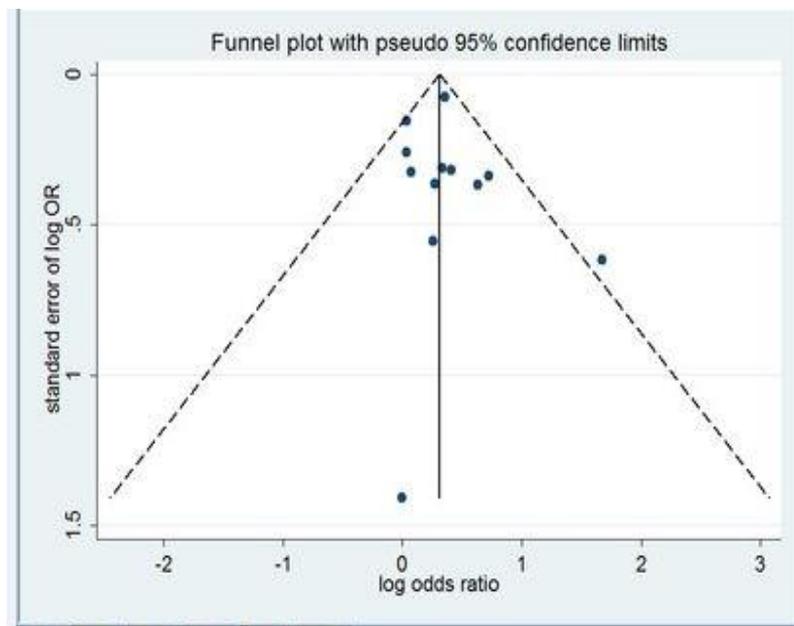
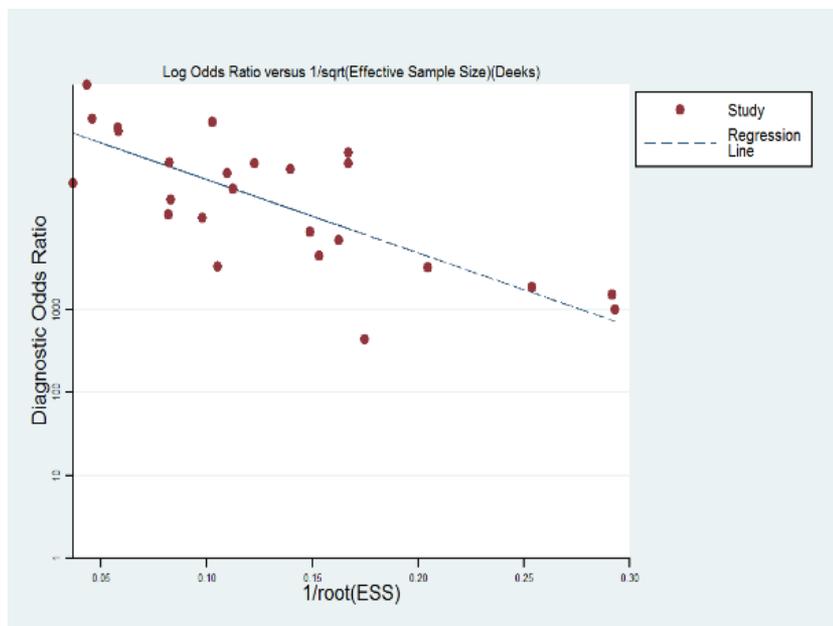
Meta分析结果

指标	效应值	95%CI	异质性检验 (I ² %)	P值
合并敏感度	0.999	0.948-1.000	36.36	0.04
合并特异度	1.000	0.999-1.000	92.32	0.00
合并阳性似然比	2255.465	666.074-7637.469	81.95	0.00
合并阴性似然比	0.001	0.000-0.055	86.31	0.00
合并诊断比值比	3.4*10 ⁶	3.1*10 ⁴ -3.8*10 ⁸	100.00	0.00

- 该方法的合并敏感度和合并特异度都非常高
- 该方法的诊断价值很高

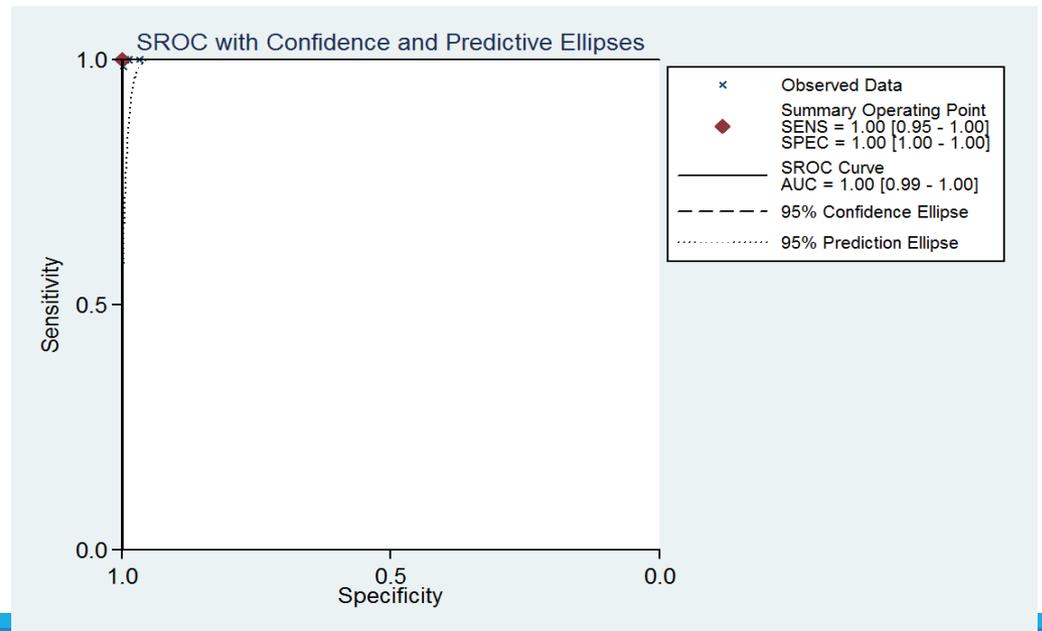
发表偏移问题

- Meta分析结果存在高估MPS技术诊断价值的可能



SROC曲线

- 绘制了25项被纳入研究的合并受试者工作曲线(summary receiver operating characteristic curve, SROC)
- 结果提示曲线下的面积是1.00 (95%CI: 0.969-1.00)



与其他筛查策略/技术的比较

检测技术	合并敏感度 (95%CI)	合并特异度 (95%CI)	合并阳性似然 比(95%CI)	合并阴性似然 比(95%CI)	合并诊断OR 值
孕早期二联血清检测	0.93 (0.93~0.94)	异质性	未报告	未报告	未报告
孕早期联合筛查	0.92 (0.85-0.97)	异质性	未报告	未报告	未报告
孕中期二联血清检测	异质性	异质性	未报告	未报告	未报告
孕中期三联血清检测	0.77 (0.72~0.81) 固定效应模型	0.95 (0.96~0.96) 固定效应模型	17.80 (13.52~23.42) 固定效应模型	0.25 (0.16~0.38) 固定效应模型	未报告
孕中期四联血清检测	未报告	未报告	未报告	未报告	未报告
大规模平行测序技术	0.999 (0.948-1.000) 混合效应模型	1.000 (0.999-1.000) 混合效应模型	2255.465 (666.074-7637.469) 混合效应模型	0.001 (0.000-0.055) 混合效应模型	3.4*10 ⁶ (3.1*10 ⁴ -3.8*10 ⁸) 混合效应模型

*: 此处所有数据皆来自本项目的相关研究

技术评估结果

经济学评价

文献中NIPT价格差别较大

加拿大\$795 (€583, £472) ;

比利时€460 (£373) ;

美国 : \$368 , \$400 , \$400-\$700 , \$795 ;

英国 : £400-£900 ;

澳大利亚 : US\$795 ;

荷兰 : € 311.74-660.94

年份不同 , 国情不同 , 可比性不大

各医疗机构收费标准

具体情况为：

- 浙江2000-2400元
- 湖南2550元
- 山东2000-2850元
- 上海1800-2400元

价格基本上是1800-2850元之间

技术评估结果

社会伦理学影响

● 信息过度(information overload)与敏感信息

- “信息过度” :孕妇夫妇要能真正理解所有信息及其背后的含义，且在有限的时间内做出决策，难度可想而知
- 可能暴露非医学信息，如父系信息(**paternity**)

● 选择性流产vs未来儿童的权益

- 反对者们认为这等同于基因表型的歧视（**expressivist objection**），是对唐氏患者的歧视
- 支持者认为，若孕妇夫妇不希望拥有一个患有严重疾病或残疾的孩子，那么流产就被认为公正

- 警惕知情同意的偏失

- 产前筛查前的知情同意实际执行不理想

- 给孕妇某种“无声的讯号”——该项检测是一项“不该避免”或“所期望”的检测，使得孕妇没有机会深思熟虑就接受

- B2C模式的隐忧

- 在相关规范制度尚不健全，质量管控缺位，人们对检测结果存在理解偏误，专业规范化的遗传咨询服务与咨询流程缺乏的情况下，商业机构直接面向顾客提供服务会产生伦理与社会问题

第六部分：讨论

- 技术开展存在的问题

- 检测结果的局限性

- 较难检测出微缺失、微重合综合征，无法完全替代产前诊断

- 市场准入随意

- 孕产妇及家属首先接受了市场宣传的言论，对于可能的其他风险便容易选择充耳不闻，不利于孕产妇们做出具有医学依据的决定，也更难实行监督管理
 - 从卫生经济学角度来说，一旦该项技术随意准入市场，个别大的实验室将会很强势，大量的数据库流入市场也难以实行监管

第六部分：讨论

- 技术开展存在的问题
 - 知情同意工作的新挑战
 - 将无创检测视作仅需一次验血的检测步骤，而无需传统检测方法之外的额外考虑，将会造成后续更多的伦理问题
 - 需要对患者进行健康教育，也需要对提供服务的检测人员进行培训

第七部分：研究结论

- 本研究结果显示，MPS技术在筛查唐氏综合征方面有着非常显著的安全性和有效性
- 本研究纳入文献证据尚存在发表偏倚，该技术的诊断价值可能存在被高估的风险

第八部分：政策建议

● 形成并颁布NIPD临床应用指南

- 各国指南一致认为对于在高危人群中开展胎儿染色体异常的筛查，无创产前检测确实是一项安全而有效的技术
- 很多报告中也指出了该技术存在的问题，如检测局限性、反馈时间长
- 各国的临床应用指南对于我国二代测序技术试点与推广具有借鉴与指导意义

● 建立并推广新型的诊断模式

- 根据发达国家已经建立的临床指南，建议可以建立一个新的唐氏综合征的诊断模式
- 高龄孕妇（如35岁以上）、孕11-14周B超检测NT结果异常、孕中期的唐氏生化免疫筛查高危或者临界风险等有产前诊断指征的，可以建议其在12-24周行二代测序检测；如果结果高危，再建议其进行侵入性产前诊断
- 对于单基因病、微小缺失综合征高危的孕妇建议直接行有创性检测

第八部分：政策建议

● 在妇幼保健三级网络下设立NIPT实验室

● 三种模式

- 医院抽取血液样本后，交由商业机构进行检测。知情同意、结果告知、随访等工作均由医院负责，商业机构只负责检测技术；
- 商业机构在医院内设立实验室分支，在医院内完成样本检测，之后与商业机构已有的数据库进行比对，得出结论；
- 医疗机构根据地理覆盖范围建立NIPT实验室及数据库，辐射到下属临检中心，统一进行二代测序检测

● 加强NIPT实验室质量管理

- 检测中心的医疗机构，要严格按照2010年卫生部颁布的行业标准——《胎儿常见染色体异常与开放性神经管缺陷的产前筛查与诊断技术标准》做好各环节质量管理工作，科学监控和评价各个重要环节的流程及质量

第八部分：政策建议

● 加大临床遗传咨询培训的力度

- 临床应重视对NIPT的检测前和检测后的遗传咨询，并给予孕妇正确的指导意见
- 如有超声异常、染色体异常家族史、复发性流产、高龄孕妇、筛查高危等情况，无论NIPT简介结果如何都应进行详细的遗传咨询

● 完善全国出生缺陷监测网络

- 我国现有的出生缺陷监测网络无法提供完整、准确的出生缺陷数据信息
- 覆盖人口有限，监测方式以被动监测为主
- 尽快完善出生缺陷监测网络，以实现产对产前筛查策略的持续评价和修正

同行评议结果

评估内容	平均分	最高分	最低分
研究的重要性	9.86	10	9
研究的必要性	9.86	10	9
研究的创新性	8.43	9	7
对该领域相关研究的熟悉程度	8.86	10	7
资料的真实性与可靠性	9.29	10	8
研究方法的科学性与合理性	9.00	10	8
研究过程的逻辑性	8.86	10	8
研究结果的可靠性	9.00	10	8
结论的恰当性	9.00	10	8
政策建议的针对性	8.86	10	8
综合评价	9.17	10	8.5

谢谢