

人乳头瘤病毒芯片寡核苷酸探针的研究*

危 敏¹ 马文丽^{1**} 李 凌¹ 张 宝¹ 郑文岭²

(南方医科大学分子生物学研究所 广州 510515)¹ (华南基因组研究中心 广州 510800)²

摘要: 高危型人乳头瘤病毒 (Human papillomavirus, HPV) 是宫颈癌的主要致病因子。利用 Arraydesigner2.0 和 BLAST 等生物学软件对 10 种型别人乳头瘤病毒全基因组序列进行分析, 设计高特异性、熔解温度 (T_m) 和 GC 含量相近的 60mer HPV 型特异性寡核苷酸探针, 用于 HPV 检测芯片的制备, 并对其中四型最常见 HPV 病毒 (HPV 6, 11, 16, 18) 探针的有效性进行初步验证, 结果表明设计所得的探针型特异性好, 可以应用于 HPV 的检测与分型。

关键词: 人乳头瘤病毒, 寡核苷酸, 芯片, 检测, 分型

中图分类号: Q78 文献标识码: A 文章编号: 0253-2654 (2006) 02-0100-06

Study of Type-specific Oligonucleotide Probes for Microarray Detection of Human Papillomavirus*

WEI Min¹ MA Wen-Li^{1**} LI Ling¹ ZHANG Bao¹ ZHENG Wen-Ling²

(Institute of Molecular Biology, Southern Medical University, Guangzhou 510515)¹

(Southern China Genomics Research Center, Guangzhou 510800)²

Abstract: High-risk human papillomavirus (HPV) types are major causes of cervical cancer. By using bioinformatics softwares such as BLAST, Arraydesigner2.0 to analyze the whole genome sequence of 10 prevalent HPV types and design type-specific oligonucleotide microarray probes of high specificity, identical length and similar melting temperature (T_m). These type-specific oligonucleotide probes were used to prepare microarray for HPV detection. And the validity of four HPV types (HPV 6, 11, 16, 18) probes were validated preliminarily, showing that most of the type specific oligonucleotide probes can be effectively applied to the HPV detection and typing microarray.

Key words: Human papillomavirus, Oligonucleotide, Microarray, Detection, Typing

HPV 感染是目前世界范围内最常见的性传播疾病的原因之一^[1], 近年来感染率持续上升。研究发现某些高危型 HPV 是宫颈癌的主要致病因子, 据美国和欧洲统计发现大约一半的宫颈癌细胞中都有 HPV-16 型的存在, HPV-18, -31 和 -45 型大约占 25 ~ 30%^[2]。全世界 99.7% 的宫颈鳞状细胞癌中都发现有 HPV 的存在^[3]。因此, 为早期发现和治疗宫颈癌, 提高患者生存率, 对 HPV 感染进行分型检测具有重要意义。

新兴的基因芯片技术与传统的分子生物学技术相比, 具有明显的优势, 可以将多种型特异性的基因探针显微集成于极小面积的固体介质表面, 具有大规模、高通量的特点, 不但可以用于分型, 而且可以对多个型别的混合感染情况同时进行监控, 具有

* 广州市重大科技攻关项目 (No. 01-Z-005-01)

** 通讯作者 Tel: 020-61648210, E-mail: wenli@fimmu.com

收稿日期: 2005-07-19, 修回日期: 2005-08-23

重要的临床应用价值。

本研究根据 GenBank 数据库中的序列信息, 利用生物学软件 Array designer 2.0 及美国国家生物技术信息中心 (NCBI) 的 BLAST 检索软件, 设计了 10 种常见 HPV 型别的 60mer 特异性 Oligo 探针共 80 条, 为下一步制备人乳头瘤病毒检测分型芯片奠定基础。

1 材料与方法

1.1 HPV 序列及质粒

本研究中特定型别的 HPV 全基因组序列信息来自 GenBank (括号内为 GenBank 登陆号): HPV6 (NC_ 000904), HPV11 (NC_ 001525), HPV16 (NC_ 001526), HPV18 (NC_ 001357), HPV31 (NC_ 001527), HPV33 (NC_ 001527), HPV35 (NC_ 001529), HPV39 (NC_ 001535), HPV45 (NC_ 001590), HPV56 (NC_ 001594)。

HPV 质粒: 含 HPV 全长基因的 4 种 HPV 质粒 (包括 HPV6、11、16 和 18 型) 由德国海德堡癌症研究中心的 De Villier E M 博士惠赠。

1.2 探针设计靶区域的确定

HPV 基因组是一双链环状 DNA 分子, 大约 7,900bp。基因组从功能上可以分为 3 个区域: (i) 早期区 (early region, E) 含有 ORF E1, E2, E4, E5, E6 和 E7, 是参与病毒复制、编码病毒蛋白、维持细胞内病毒高拷贝数的基因。(ii) 晚期区 (late region, L) 含 L1 和 L2 基因, 分别编码主要衣壳蛋白和次要衣壳蛋白。(iii) 长控制区 (long control region, LCR), 含有多个结合位点, 通过与转录因子的相互作用, 调控 DNA 的复制。这一区域在病毒基因组的变异度也最高^[4]。结构示意图如图 1。

我们首先将 HPV 病毒基因组非编码序列 (LCR)、结构蛋白编码序列 (L1、L2) 和非结构蛋白编码序列 (E1, E1, E2, E4, E5, E6, E7) 都列为 60mer 探针设计的备选区域。

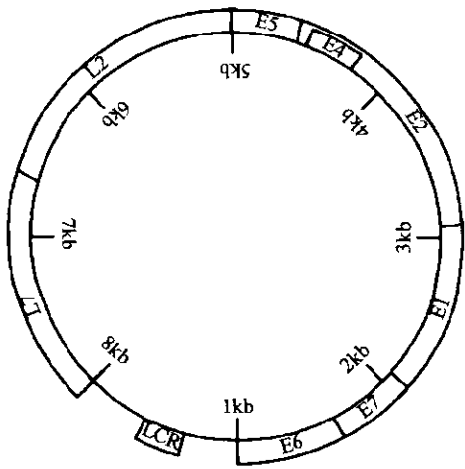


图 1 人乳头瘤病毒基因结构示意图

1.3 型特异性探针的设计

将各探针备选区域的序列输入软件 Array designer 2.0 的程序中, 设置寡核苷酸探

针设计参数, 包括探针长度, 溶解温度范围, 以及二级结构的稳定性 (包括探针发夹结构和二聚体的数目) 等。同时, Array designer 2.0 可自动连接美国国家生物技术信息中心 (NCBI) 的 BLAST 检索软件, 排除备选区域中与 GenBank 中其他基因组交叉同源的序列 (包括人全基因组和其他微生物的基因组)。

本研究设置探针设计参数的主要原则为: (1) 所有探针的 T_m 值应接近整个基因组的平均 T_m 值, 上下波动范围不超过 5°C ; (2) GC 含量的范围在 40% ~ 60%; (3) 寡核苷酸探针内重复的单一碱基连续不超过 6 个; (4) 探针分子最稳定二级结构配对碱基长度少于 6bp; (5) 经 BLAST 比较与其他序列的相似性小于 70%; (6) 经 BLAST 比较与其他基因的同源片段连续不超过 20 个碱基。

以 11 号探针 HPV16-L1 为例, 生物软件 Arraydesigner 2.0 可对其二级结构和探针内部稳定性 (ΔG) 进行分析, 分析界面见图 2, 图 3。探针内部的 hairpin 和自身二聚体长度要求不超过 6bp, ΔG 趋于 0, 这样可保证探针内部不会形成稳定的二级结构而影响杂交效率。

最后, 将软件设计的符合以上要求的 60mer 探针与 GenBank 中的所有序列再次进行 BLAST 比对, 排除与非人乳头瘤病毒基因序列 Score 分值高于 40 的探针, 以确保探针与靶基因的特异性杂交。由于 HPV 病毒各型别之间存在着交叉同源性, 设计型特异性探针必须同时排除与其他型别 HPV 核酸序列 Score 分值高于 40 的探针。

#	Sense Hairpin	ΔG (kcal/mol)
1.	<div>CTCCACCCACCCACATCGAAAAGCAAAAGGATTACA 5'</div> <div>A </div> <div>LTCTACCTCTACAACCTGCTAAACGC 3'</div>	... -1.7
2.	<div>ACATCGAAAAGCAAAAGGATTACA 5'</div> <div> </div> <div>LCCACCCACCTCATCTACCTCTACAACCTGCTAAACGC 3'</div>	... -1.2
3.	<div>GAAAAGCAAAAGGATTACA 5'</div> <div>C </div> <div>LTACACCCACCCACCTCATCTACCTCTACAACCTGCTAAACGC 3'</div>	... -1.2
4.	<div>CACCCACATCGAAAAGCAAAAGGATTACA 5'</div> <div>C </div> <div>LACCTCATCTACCTCTACAACCTGCTAAACGC 3'</div>	... -1.1
5.	<div>CACATCGAAAAGCAAAAGGATTACA 5'</div> <div> </div> <div>LCCACCCACCTCATCTACCTCTACAACCTGCTAAACGC 3'</div>	... -0.0
6.	<div>ATCGAAAAGCAAAAGGATTACA 5'</div> <div> </div> <div>LACCCACCCACCTCATCTACCTCTACAACCTGCTAAACGC 3'</div>	... -0.0
7.	<div>CAAAAGGATTACA 5'</div> <div> </div> <div>GAAAAGCTACACCCACCCACCTCATCTACCTCTACAACCTGCTAAACGC 3'</div>	... -0.0

图 2 13 号探针发夹结构

1.4 型特异性探针的初步验证

按照以上方法设计并合成 HPV 6, 11, 16, 18 4 个型别的探针共 20 条, 平均 1 条探针代表一个病毒编码区域。合成后的寡核苷酸探针加 DMSO 调整浓度为 $1\mu\text{g}/\mu\text{L}$, 以一定的点阵打印于玻片上, 每个探针重复打印 6 个点。采用含 HPV 全长基因的 4 型 HPV 质粒为阳性质控样品, 应用限制性荧光标记技术分别进行 Cy3 荧光标记^[5], 标记后的样品与芯片进行杂交检测, 验证寡核苷酸探针的有效性。

#	Sense Self Dimer	ΔG (kcal/mol)
1.	5' ACATTAGGAAAACGAAAAGCTAC... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-2.9
2.	5' ACATTAGGAAAACGAAAAGCTACACCACCCATCTACTCT... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCAC... 	-1.7
3.	5' ACATTAGGAAAACGAAAAGCTACACCACCCAC... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-1.2
4.	5' ACATTAGGAAAACGAAAAGCTACA... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-1.2
5.	5' ACATTAGGAAAACGAAAAGCTACACCACCCATCTACTCT... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-1.1
6.	5' ACATTAGGAAAACGAAAAGCTACACCACCCACCTC... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-0.0
7.	5' ACATTAGGAAAACGAAAAGCTACACCAC... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-0.0
8.	5' ACATTAGGAAA... 3' CGCAAATCGTCAACATCTCCATCTACTCCACCACCCACATCGAAA... 	-0.0

图 3 13 号探针自身二聚体结构

2 结果

2.1 HPV 病毒全基因组序列分析

遵循以上探针设计原则可获得 HPV-6、-11、-16、-18、-31、-35、-39、-45、-51、-56 共 10 个型别的 60mer Oligo 探针共 80 条，每个编码区域设计 1~3 条探针，表 1 中列出了部分 HPV-6、-11、-16、-18 型特异性探针的碱基序列及其重要参数值。

2.2 基因芯片杂交信号扫描结果

基因芯片杂交扫描结果显示，4 型 HPV 探针都有明显的杂交信号（图 4），说明此法设计的探针特异性和杂交灵敏度较好，可用于 HPV 芯片检测。

表 1 型特异性探针的位置、序列及其重要参数

序号	名称	位置	碱基序列	T _m	GC (%)	发夹结构	二聚体	重复序列
1	HPV-6	L2	ACCACCGTGGAGGCTAATGGACATATATTAATTTCTGCACCCACTATAACGTCACACCC	76.4	46.7	4	4	3
2	HPV-6	E2	TTTGATTTAATATCATCAACGTGGCACTGGCCCTCTCAAAGCCACACATAAACATGCC	76.8	45.0	4	4	3
3	HPV-6	E5	CATGTCAATTTAATGATGAGATAGCTGGCTGGCTTTGTGGTTGTATGTGCCCTTATATG	73.7	40.0	3	3	3
4	HPV-6	L1	GGGAACCTGTGCTGATACCTTTATAATTAAGGGTAGTGGAAATCGAACGCTGTGATGGGA	74.9	45.0	3	4	3
5	HPV-6	LCR	CAATTAACATACATATTTGTATATCTTGTTACACCTGTGACTCAGTGGCTGTGCACGGC	74.0	41.7	4	4	3
6	HPV-11	E6	ACCTAAAGGTTCTGTGGCGAGACAACCTTCCCTTTCAGCGCTGTCCCTGTTGCTTAGAAC	79.1	50.0	5	5	3
7	HPV-11	E7	GCACAGACGGACATCATCAGAACTACAAGACCTTTTCTGGGCACACTAAATATTGTGT	76.2	45.0	4	4	4
8	HPV-11	L1	AACAAGTTTGCATTACCTGATTCATCCCTGTTTGACCCCACTACACAGCGTTTACTATCG	75.5	43.3	4	4	4
9	HPV-11	L2	TGACCTATCCCTGACCGCTGCGCAACATCTGTTACACAGTCTTATCTTACCTCCACACC	76.5	48.3	4	4	3
10	HPV-11	E1	AACATAGACAGCGGGAAGCAGTAGACGACAGCACCAGAGAGCATGCAGACACATCAGGAA	80.0	53.3	3	5	3
11	HPV-16	L1	ACATTAGGAAAACGAAAAGCTACACCACCCATCTACTCTTACAACCTGCTAAACCGC	75.8	45.0	3	4	4
12	HPV-16	E6	GCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTACCACAGTTATGCACAGAGCTG	78.8	51.7	3	4	3

续表 1

13	HPV-16	E2	TGCCACACCACTAAGTTGTTGCACAGAGACTCACTGGACAGTGCTCCAATCCTCACTGCA	79.7	51.7	5	5	2
14	HPV-16	LCR	GTACATTGTCATATAAAATAAATCACTATGCGCCAACGCCCTTACATACCGCTGTTAGG	73.3	40.0	3	4	4
15	HPV-16	E7	TGCAAGTGTGACTCTACGCTTCGGTCTGCGTACAAAGCACACAGTAGACATTGCTACT	78.2	48.3	6	6	3
16	HPV-18	L2	AGTCGCTAACCCCTGAGTTTCTTACACGTCCATCCTCTTTAATTACATATGACAACCCGGC	75.7	45.0	3	4	3
17	HPV-18	L1	ATAATAAATTAGATGACACTGAAAGTTCCTATGCCGCCACCTCTAATGTTTCTGAGGACG	74.5	41.7	4	4	3
18	HPV-18	E7	ATCGAGTTAATCATCAACATTTACCAGCCGACGAGCCGAACCACAACGTACACAATGT	77.8	46.7	4	4	3
19	HPV-18	LCR	TAGTCATTITCTGTCCAGGTGCGCTACAACAATTGCTTGCATAACTATATCCACTCGCT	75.7	43.3	3	4	4
20	HPV-18	E1	AAATGCCGAACATCGCGGCAATGTATGTAGTGGCGGCAGTACCGAGCCTATAGACAACGG	80.1	53.3	4	6	3

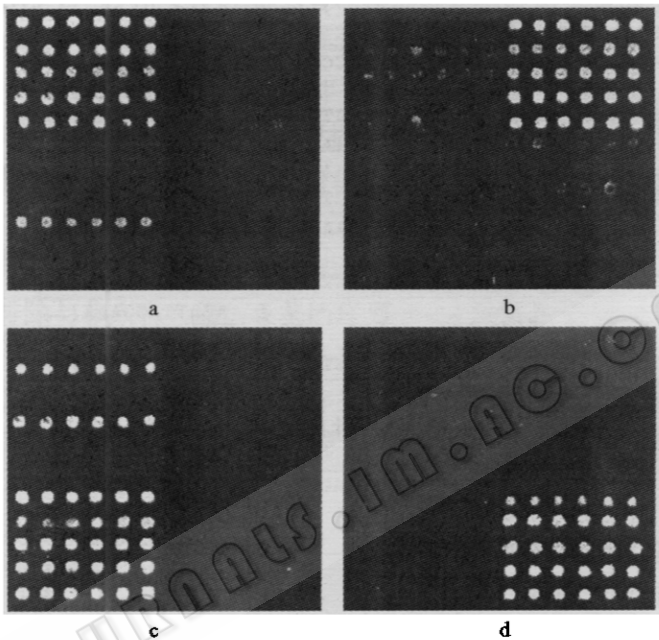


图 4 HPV 芯片杂交信号扫描结果
a HPV6, b HPV16, c HPV11, d HPV18

3 讨论

目前世界上已发现的 HPV 已有 110 余种，共有 20 余种从宫颈癌组织中被分离出来，包括 HPV-16、-18、-31、-35、-39、-45、-51、-56 型等。HPV 感染的临床检测方法主要包括形态学检测和核酸检测。细胞学、阴道镜和组织病理学等形态学检测方法容易因刮片、固定方法、阅片技术等多种因素影响出现漏诊；传统的核酸杂交法如 Southern blot，斑点杂交和原位杂交等操作复杂，一次反应仅能检测单型或双型 HPV 感染，PCR 技术假阳性率高，不利于大规模的临床诊断，因此迫切需要建立一种快速、灵敏、高效的检测方法。

由于基因芯片技术在实现检测标准化和程序化上具有较大优势，利用其多基因多位点同步检测及检测结果的计算机处理，将会成为病毒检测中的有力工具。国内外众多研究者都尝试利用该技术对临床 HPV 感染进行检测和分型^[6~8]。其中包括原位合成的短链寡核苷酸芯片，合成后点样的 cDNA 基因芯片及长链寡核苷酸基因芯片。短链寡核苷酸芯片受到严格的专利限制，cDNA 芯片的制备需要构建 cDNA 文库，克隆的筛选和鉴定操作繁琐、周期很长，而且很难保证探针质量。而设计合成 50 ~ 70mer 长的寡

核苷酸探针更为灵活、简便,在合适的杂交条件下可以获得满意的结果,其中 60 ~ 70mer 探针与全长 cDNA 探针相比检测特异性增强,灵敏度最为一致^[8]。

本研究拟利用长链寡核苷酸芯片对 HPV 病毒检测和分型,其中长链 Oligo 型特异性探针的设计是研究的前提和关键。目前有许多寡核苷酸设计软件包可以利用,设计的原则各有侧重,但这些软件大都利用 BLAST 软件作为基本工具找出探针设计的基因特异区域^[9~11]。本研究通过 BLAST、Arraydesigner 生物信息学软件,针对 10 个型别的 HPV 全基因序列,设计出 80 条 60mer 探针。设计过程中首先就是要确保探针具有高度的特异性,每一条探针都反复经严格的 BLAST 筛选,确保与其他基因序列同源性 score 值小于 40 及低于连续 20 个碱基互补。其次,要考虑到探针二级结构的稳定性,如最稳定二级结构配对碱基长度不超过 6 个,尽量减少发夹结构和二聚体的数目。最后,在同时检测多条探针的杂交反应中,所有探针的 T_m 值和 GC 含量必须在相近范围,以满足杂交动力学要求,获得最理想的杂交效果。

我们应用 4 型 HPV 阳性对照样品对寡核苷酸探针的有效性进行了验证,从结果可见,除 HPV6 与 HPV11 之间少数探针出现共同的杂交信号以外,其他探针均表现出较好的型特异性。对型与型之间交叉杂交的探针,如 HPV-6 型 4 号探针,重新对其序列进行分析后发现,该序列第 1 ~ 35 位碱基序列与 HPV11 型基因组第 794 ~ 828 位碱基序列同源性较高,达到 31/35 (88%),因此尽管其 score 值为 38.2,也没有连续 20 个碱基互补,但在杂交反应中仍然出现了型别之间交叉杂交的现象。值得注意的是,尽管寡核苷酸探针在设计上难以达到完美的要求,但由于完全正常的 Watson-Crick 配对双链比具有错配 (Mismatch) 碱基的双链具有更高的热力学稳定性,所以,前者的荧光强度要比后者强出 5% ~ 35%。另外在高严谨性的杂交和洗膜条件下,完全配对的探针-靶序列之间结合效率更高,从而可以提高芯片检测的特异性。

利用合适的生物学软件进行 HPV 寡核苷酸探针的设计,是一种非常简便、可行、有效的方法。但由于目前的生物信息学软件都缺乏经实际证实过的核酸生物物理学特性作为设计寡核苷酸探针的准则,因此所有设计合成的探针必须经过多次杂交筛选,直至确定出特异性强、符合杂交条件的探针。更为关键的是临床样品的情况非常复杂,对芯片检测系统的灵敏度及特异性提出了很高的要求,因此还需要对现有的探针设计方法和杂交条件不断优化。

参 考 文 献

- [1] 徐 芸,王孝华,陈禹保,等. 微生物学通报, 1997, **24** (4): 231 ~ 233.
- [2] Harro C D, Pang Y Y, Roden R B, *et al.* J Natl Cancer Inst, 2001, **93** (4): 284 ~ 292.
- [3] Walboomers J M, Jacobs M V, Manos M M, *et al.* J Pathol, 1999, **189**: 12 ~ 19.
- [4] Adam E, Berkova Z, Daxnerova Z, *et al.* Am J Obstet Gynecol, 2000, **182** (2): 257 ~ 264.
- [5] Li L, Ma W L, Zhu J, *et al.* Journal of Virological Method, 2003, **114**: 71 ~ 75.
- [6] Klaassen C H, Prinsen C F, de Valk H A, *et al.* J Clin Microbiol, 2004, **42** (5): 2152 ~ 2160.
- [7] Delrio-Lafreniere S A, Browning M K, McGlennen R C. Diagn Microbiol Infect Dis, 2004, **48** (1): 23 ~ 31.
- [8] Liu C H, Ma W L, Shi R, *et al.* J Biochem Mol Biol, 2003, **36** (4): 349 ~ 353.
- [9] Wang H Y, Malek R L, Kwitek A E, *et al.* Genome Biol, 2003, **4** (1): R5.
- [10] Chou H H, Hsia A P, Mooney D L, *et al.* Bioinformatics, 2004, **20** (17): 2893 ~ 2902.
- [11] Nielsen H B, Wernersson R, Knudsen S. Nucleic Acids Res, 2003, **31** (13): 3491 ~ 3496.
- [12] Rouillard J M, Zuker M, Gulari E. Nucleic Acids Res, 2003, **31** (12): 3057 ~ 3062.