

基于 PCR 技术的人脑基因可变剪接差异性研究 *

张晓宇^{1,2} 熊洁仪² 梁洪雨^{1,2} 邵宁一² Philipp Khaitovich² 杨洪一^{1,Δ}

(1 东北林业大学生命科学学院 黑龙江 哈尔滨 150040 ;

2 中国科学院 - 马普学会计算生物学伙伴研究所, 中国科学院计算生物学重点实验室 上海 200031)

摘要 目的: 对人脑的基因可变剪接位点差异性进行分析。方法: 通过 RNA-Seq 得到的转录组数据, 分析推测出在不同年龄的 7 对样本的大脑前额叶皮层(prefrontal cortex, PFC)表达的某些基因具有可变剪接位点的差异性; 为验证此推测, 设计 7 对引物, 提取不同年龄样本的 RNA, 通过反转录合成 cDNA, 采用 PCR 技术扩增目的序列。结果: 在已剪接和(或)未剪接产物预期片段处有明显的条带。结论: 通过 Quantity One 软件分析后, 确定在不同年龄间存在可变剪接差异。基于 PCR 技术的基因可变剪接分析对 RNA-seq 数据集的预测结果进行验证是一种有效新颖的 PCR 新应用。

关键词: 可变剪接; 聚合酶链式反应; mRNA 前体; RNA-seq; 大脑前额叶皮层

中图分类号: Q75, Q78, R323.1, R338.2 **文献标识码:** A **文章编号:** 1673-6273(2012)06-1013-04

Investigation of Difference of Alternative Splicing In Human Brain Based On PCR*

ZHANG Xiao-yu^{1,2}, XIONG Jie-yi², LIANG Hong-yu^{1,2}, SHAO Ning-yi², Philipp Khaitovich², YANG Hong-yi^{1,Δ}

(1 College of Life Science, Northeast Forestry University, 26 Hexing Road, 150040, Harbin, China;

2 Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences;

320 Yueyang Road, 200031, Shanghai, China)

ABSTRACT Objective: To investigate the aging related alternative splicing by PCR. **Methods:** Differences in splicing sites of the prefrontal cortex (PFC) in brain samples of 2 different ages were detected by transcriptome RNA-Seq data analysis. Seven pairs of primers were designed separately. The mRNA extracted from samples were mixed in equal amount, and then cDNA was synthesized by reverse transcription. **Results:** The target sequences were amplified. **Conclusions:** There were clear bands for different splicing pattern, established the existence of alternative splicing differences between ages.

Key words: Alternative splicing; PCR; Pre-mRNA; RNA-seq; Prefrontal cortex

Chinese Library Classification(CLC): Q75, Q78, R323.1, R338.2 **Document code:** A

Article ID: 1673-6273(2012)06-1013-04

前言

选择性剪接(也叫可变剪接)是指从一个 mRNA 前体中通过不同的剪接方式(选择不同的剪接位点组合)产生不同的 mRNA 剪接异构体的过程,而最终的蛋白产物会表现出不同甚至是相互拮抗的功能和结构特性,或者在相同的细胞中由于表达水平的不同而导致不同的表型。真核生物基因转录产生的 mRNA 前体可按不同方式剪接产生两种或多种 mRNA,这种选择性剪接有些只产生一种蛋白质,有些可产生两种或两种以上的蛋白质。大部分多细胞生物的蛋白编码基因有选择性剪接^[1]。mRNA 这种选择性剪接与细胞的分化、发育及人类的许多疾病密切相关^[2]。

人类的许多恶性肿瘤疾病,都与选择性剪接存在密不可分的关系,如胃癌、肝癌、人乳腺癌,以及丙型肝炎等疾病^[3]。肿瘤癌基因、抑癌基因、肿瘤转移抑制基因可发生选择性剪接^[4],与肿瘤发展关系密切。研究表明,认知能力的下降已经成为老年

人最大的健康威胁之一,超过 50%的老年人承受着阿尔茨海默氏症的折磨^[5]。选择性剪接的变化与种族和性别特异性的变化紧密联系^[6,7]。因此,对 mRNA 前体的可变性剪接的研究具有现实意义。

根据现有 ESTs 数据分析显示,在人类 23000~24000 个基因中大约有 40%~60%的基因具有选择性剪接的形式,这说明选择性剪接是人基因组功能的复杂性的重要组成部分之一^[8,9]。通过 mRNA 前体的选择性剪接功能,至少 90%的 mRNA 前体被当做内含子剪切掉,而只有平均大概 10%的 mRNA 前体成为外显子序列^[10],所以 mRNA 剪接位点的选择也是广泛存在的。在过去的十年里,信使 RNA(mRNA)的表达被广泛的运用到探索随年龄变化的人类和哺乳动物大脑的基因表达水平^[11]。近年来兴起的转录组测序技术(RNA-seq)就是把 mRNA, non-coding RNA 等或者其中一些用高通量测序技术,测出他们的序列,而序列读数反映他们的表达水平。RNA-seq 实验的目的主要以比较两种或多种样本中基因表达或整个转录组的差异为

* 基金项目:中国科学院重要方向项目(KSCX2-EW-R-02-02)

作者简介:张晓宇(1986-),女,研究生,主要研究方向:人类进化研究,电话:15021503065, E-mail: zhangxiaoyu262425@hotmail.com

Δ通讯作者:杨洪一, E-mail: yhy1@sohu.com

(收稿日期:2011-09-01 接受日期:2011-09-23)

主,如比较癌症组织和正常组织的转录组差异等。这些差异既包括通常意义下的差异表达基因,也主要包括选择性剪接模式的差异、剪接异构体表达的差异、非编码转录本的差异等。由于RNA-seq为高通量测序,技术有了明显提高,但仍然存在系统噪声和随机错误的可能,对样本的准备要求较高,导致RNA-seq结果存在可能的在测序结果上的误差。在内含子和基因间区域,一半以上的转录组来源于序列的重复,占全部转录组的42%。对于一些简单并复杂性较低的重复,他们来源于功能性ncRNA,如snRNA,snoRNA和7SK RNA等,RNA的来源影响转录检测,所有这些差异都将影响得到的结论,使其存在假阳性结果的可能。而基于PCR试验验证可变剪接结果更为准确直观。本研究用Illumina公司的Solexa技术对人类大脑14个不同年龄的样本测序,找到一些大脑中随年龄变化的可变剪接基因。根据RNA-seq测序结果分析,推测出可能存在的可变剪接的差异性。为进一步验证RNA-seq特异性和敏感性,本研究基于PCR技术,对7对不同年龄样本的大脑前额叶皮层进行可变剪接位点差异性进行分析^[12],以期对RNA-seq数据集的预测结果进行确实验证。

1 材料与方法

1.1 材料

1.1.1 组织来源 人脑组织大脑前额叶皮层(prefrontal cortex, PFC)组织样本由中国武汉的中国脑库中心获得,所有样本实验的实施都获得伦理委员会的批准,并由组织库病理学家确认为正常健康大脑样本。

1.1.2 主要试剂和酶类 Trizol Reagent与SuperScript Reverse Transcriptase 购自Invitrogen公司,PCR反应试剂dNTP Mix-

ture(各2.5 mmol/L),Taq DNA聚合酶(5U/ μ l)附10× PCR Buffer [Tris-HCl (PH 8.3),KCl MgCl₂] ,6× Loading Buffer ,D2000 DNA Marker 均为TaKaRa公司产品。

自配试剂 100 mmol/L Tris-HCl(PH 8.0) 20 mmol/L EDTA (PH 8.0) ,1.4 mol/L NaCl ,1× TAE 琼脂糖凝胶电泳缓冲液(40 mmol/L Tris·CH₃COOH,1 mmol/L EDTA)。

1.2 方法

1.2.1 总RNA的提取 从不同年龄(新生儿和成年)样本的大脑组织中提取总RNA。每个样本中切取100 mg组织,放入1.5 ml Eppendorf管中,加入1 ml Trizol试剂,将组织完全破碎后,放入Eppendorf低温离心机,12000 g 4℃离心10 min后取出,室温静置5 min;加入0.2 ml 氯仿剧烈震荡15 s使其摇匀,室温静置3 min,12000 g 4℃离心15 min,吸取水相至新的1.5 ml Eppendorf管,加入0.5 ml 异丙醇轻轻摇匀,室温静置10 min后,12000 g 4℃离心10 min,倒掉上清液,加入75%乙醇,7400 g 4℃离心5 min,倒掉上清,室温静置2 min,使乙醇挥发彻底,加入20 μ l水溶解RNA。

1.2.2 RNA-seq 用Oligo(dT)参照Oligotex® mRNA Midi Kit (Qiagen)操作手册两次重复分离提取mRNA,100 ng mRNA为起始量加入5X Fragmentation Buffer,94℃温育2.5 min,完成mRNA的片段化,把片段化的mRNA置于冰上,过柱去片段化离子,用随机引物反转录获取cDNA,送到公司进行测序。

1.2.3 外显子特异性引物的设计 基于上述的大脑组织样本所得到的RNA-seq数据集预测结果,由Primer5软件设计7对引物,结果见表1。表中所有的引物均由上海生工生物技术有限公司合成。

表1 引物序列
Table 1 PCR Primer Sequence

(Gene)	(Up-stream primer sequence)	(Down-stream primer sequence)	Product lengths (long/short) (bp)
HSPA4	CCTACAGGATTAACAGGTAT	CAGATTCCTGGAAACGTAT	1795/76
BCLAF1	TTATCCTGATGGTGGAGATC	GCATCCAGATTCTTGTGTAG	600/81
R3HDM1	CAATGAAGGATTTGGAGG	ACTTATCACTGGCCTTTTC	351/219
FMNL2	CAAAGAGGCAGCAGCAAG	TGGGCAGTTTCATTACAG	396/270
MEGF8	GCCGTGTGCGTGAAGT	TGCTCCACCGAGATGAGG	286/153
MAPT	AGAAGGTGGCAGTGGTCC	ATGGATGTTGCCTAATGAGC	320/227
AKPT1B1	GCAGCCAAGCACAAATAAACT	ATCACCAACCAAGTTCCTCTG	669/71

1.2.4 cDNA的合成 取一Eppendorf管,加入1 μ g总RNA,Random primers 1 μ l (150 ng)为反转录引物,1 μ l dNTP(每种10 mM),加双蒸水至12 μ l,65℃下变性5 min后,立即取出置于冰上,再依次加入5× First-Strand Buffer 4 μ l,0.1M DTT 2 μ l,RNaseOUT 1 μ l,充分混合后25℃下温育2 min;反应结束后加1 μ l SuperScript反转录酶,混合后25℃温育10 min,42℃温育50 min,70℃温育15 min,反应结束后将cDNA保存备用。

1.2.5 聚合酶链式反应 PCR反应总体积为20 μ l,在0.5 ml Eppendorf管中依次加入10× PCR Buffer (Mg²⁺ Plus) 2 μ l,10 mmol/L dNTP Mixture 1.6 μ l,cDNA 0.8 μ l,10 mmol/L上游引

物和下游引物各0.8 μ l,Taq酶0.1 μ l,加双蒸水至20 μ l。反应条件为:94℃预变性2 min后,进行30个循环:95℃ 20 s,72℃ 30 s,72℃ 5 min,循环结束后,72℃自动延伸5 min。

1.2.6 PCR产物的琼脂糖凝胶电泳检测 取PCR产物2.5 μ l,6× Loading Buffer 1 μ l,点样在2.0%(w/v)琼脂糖凝胶,于1× TAE电泳缓冲液中,100 V,90 mA下电泳15 min,溴化乙锭(EB)染色后,在长波紫外光下Bio-RAD凝胶成像仪下呈现目的条带,照相保存。

1.2.7 凝胶电泳条带分析 将凝胶电泳图片置于Quantity One软件下,进行定量分析。

2 结果分析

2.1 特异性引物 PCR 扩增的结果

经过优化 PCR 反应体系,7 对引物均扩增得到了预期片段。DNA 条带清楚,未见明显非特异性扩增条带,引物二聚体不明显。Quantity One 软件分析结果如表 2 所示。

表 2 Quantity One 软件分析 DNA 条带结果
Table 2 The results of DNA bands analyzed by Quantity One

基因(Label)	样本(Sample)	短产物信号(LongSig)	长产物信号(ShortSig)
HSPA4	新生儿(New born)	无(None)	105.865
HSPA4	成年(Young)	无(None)	41.876
BCLAF1	新生儿(New born)	无法定量(None quantity)	142.679
BCLAF1	成年(Young)	无法定量(None quantity)	156.843
R3HDM1	新生儿(New born)	195.101	102.425
R3HDM1	成年(Young)	17.872	244.707
FMNL2	新生儿(New born)	41.345	343.516
FMNL2	成年(Young)	222.301	94.411
MEGF8	新生儿(New born)	157.123	155.136
MEGF8	成年(Young)	147.434	160.894
MAPT	新生儿(New born)	457.078	5.618
MAPT	成年(Young)	209.078	134.187
AKPT1B1	新生儿(New born)	126.091	57.861
AKPT1B1	成年(Young)	104.187	12.696

2.2 选择性剪接结果

由 RNA-seq 得到的结果显示 样本中均有可变剪接差异。采用 RT-PCR 技术可以准确的鉴定有无可变剪接及其样本之间的差异性。所得到的结果如图 1。在已发生可变剪接的样本中存在长短产物的两条 DNA 条带,而在没有发生可变剪接的样本中只有长产物 DNA 条带。实验 7 对样本中,有 6 对发生可变剪接,1 对未发生可变剪接。在有可变剪接样本中会有以下几种结果 a 外显子保留产物表达量高于外显子被剪接产物表达量 b 外显子保留产物表达量低于外显子被剪接产物表达量 c 外显子保留产物表达量与外显子被剪接产物表达量几乎

相同。由图 1 可见,在已剪接产物预期片段处有明显条带,而不同年龄样本之间的外显子剪接和未剪接的产物表达量是不同的。基因 HSPA4 中,新生儿大脑(泳道 1)和成年大脑(泳道 2)没有发生可变剪接。基因 R3HDM1 (泳道 5 和 6),基因 FMNL2 (泳道 7 和 8),基因 MAPT (泳道 11 和 12),基因 AKPT1B1(泳道 13 和 14)中,成年大脑与新生儿大脑中可变剪接存在较大差异,基因 MEGF8(泳道 9 和 10)中,成年大脑与新生儿大脑间的可变剪接几乎无差异。基因 BCLAF1 (泳道 3 和 4)中,成年大脑与新生儿大脑中外显子被剪接产物的表达量及其微弱,导致软件无法对其进行定量。

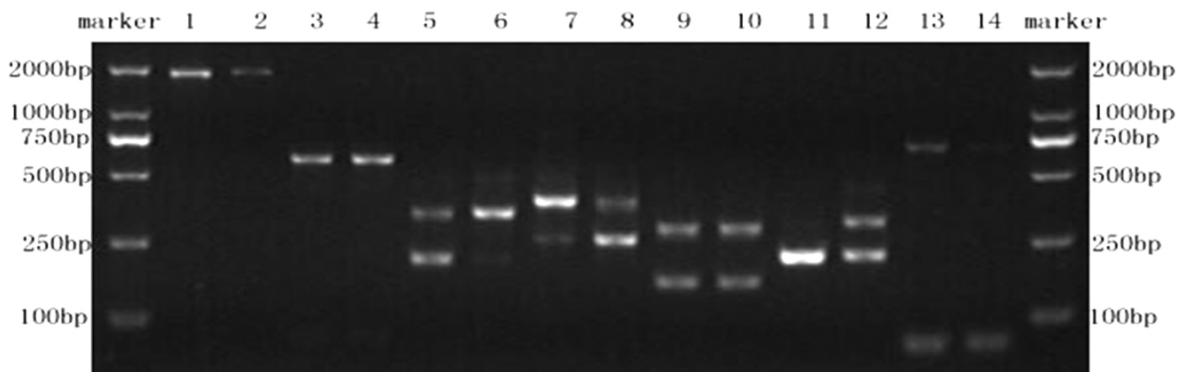


图 1 PCR 检测可变剪接差异的电泳结果 1: 基因 HSPA4(新生儿) 2: 基因 HSPA4(成年) 3: 基因 BCLAF1(新生儿) 4: 基因 BCLAF1(成年) 5: 基因 R3HDM1(新生儿) 6: 基因 R3HDM1(成年) 7: 基因 FMNL2(新生儿) 8: 基因 FMNL2(成年) 9: 基因 MEGF8(新生儿) 10: 基因 MEGF8(成年) 11: 基因 MAPT(新生儿) 12: 基因 MAPT(成年) 13: 基因 AKR1B1(新生儿) 14: 基因 AKR1B1(成年)

Fig.1 Result of alternative splicing by PCR 1: Gene HSPA4 (new born); 2: Gene HSPA4 (young); 3: Gene BCLAF1 (new born); 4: Gene BCLAF1 (young); 5: Gene R3HDM1(new born); 6: Gene R3HDM1(young); 7: Gene FMNL2(new born); 8: Gene FMNL2(young); 9: Gene MEGF8 (new born); 10: Gene MEGF8 (young); 11: Gene MAPT (new born); 12: Gene MAPT (young); 13: Gene AKR1B1(new born); 14: Gene AKR1B1(young)

3 讨论

mRNA 前体的选择性剪接是转录后比较复杂的过程之一,对于选择性剪接的机制和机理的研究成为近年来发展最快的领域之一^[13]。

RNA 编码和可变性剪接是两种不同的转录过程,然而,都可以增加蛋白质组学的复杂性和不同的蛋白质的数量,但却没有增加基因的数目^[14],可变性剪接对于转录本的多样性具有重要的意义,可以说它会影响物种的形成并增加有机体的复杂性,当一种亚型在剪接过程中具有活性,那么它可以有剪接行为或加强剪接的活性,这个取决于目的剪接位点^[15]。可变剪接的错误调节通常导致人类疾病^[16,17]。大多数可变性剪接由可变外显子保留或切除、5'和3'可变剪接位点的转移组成,很少的可变剪接来源于内含子的保留^[18]。RNA-seq 被广泛的应用到转录组结构研究中。在人类中,通过 RNA-seq 技术,已有 31618 个已知的可变剪接被证实(占有已知可变剪接的 11%)并有 379 个新可变剪接被发现^[19]。RNA-seq 读段定位的重要挑战,主要围绕如何处理可变剪接位点,共生同源基因以及拟基因^[20]。目前,基于 RNA-seq 技术的研究发展迅速并日趋成熟,同时也有其不足之处,某些原因可导致 RNA-seq 集的误差,如 1) 有些基因的 mRNA 的表达量太少;2) 某些情况下, RNA-seq 没有重复数据集,所以给误差估计带来困难;3) 不同个体之间有表达差异,不一定是系统差异所导致的;4) 数据分析中的多重检验问题等。作为对 RNA-seq 技术的验证,本实验在得到 RNA-seq 数据基础上,把 PCR 技术运用到 RNA-seq 领域,以排除 snRNA, snoRNA 和 7SK RNA 等 ncRNA 对 RNA-seq 的干扰,对 RNA-seq 的预测结果进行深入验证,改善了 RNA-seq 方法的特异性。通过本实验的研究可证实,在同一基因中不同年龄样本的可变剪接是具有差异性的,而通过 RNA-seq 得到的结果显示所有样本均有不同年龄的可变剪接差异,使实验结果具有更高的可信度。PCR 技术与 RNA-seq 相结合探索 mRNA 的可变性剪接差异,是 PCR 技术在高通量技术在具体生物研究中的具体应用,我们的实验表明这种新的应用领域是可行有效的,为 RNA-seq 在 mRNA 可变剪接研究中的作用提供了有力验证。

参考文献(References)

- [1] Ast G. How did alternative splicing evolve [J]. Nat. Rev. Genet, 2004, 5: 773-782
- [2] Ben-Dov C, Hartmann B, Lundgren G, et al. Genome-wide analysis of alternative pre-mRNA splicing [J]. Biol. Chem, 2008, 283: 1229-1233
- [3] Tazi J, Nadia B, Stefan S. Alternative splicing and disease [J]. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2009, 1792(1): 14-26
- [4] Venables J, Klinck R. Cancer-associated regulation of alternative splicing [J]. Nature Structural & Molecular Biology, 2009, 16(6): 670-676
- [5] Bishop NA, Lu T, Yankner BA. Neural mechanisms of ageing and cognitive decline [J]. Nature, 2010, 464(7288): 529-535
- [6] Blekhman R, Marioni JC, Zumbo P, et al. Sex-specific and lineage-specific alternative splicing in primates [J]. Genome Res, 2010, 20(2): 180-189
- [7] Rinn JL, Snyder M. Sexual dimorphism in mammalian gene expression [J]. Trends Genet, 2005, 21: 298-305
- [8] Modrek B, Lee C. A genomic view of alternative splicing [J]. Nature Genetics, 2002, 30(1): 13-19
- [9] Calarco JA, Xing Y, Caceres M, et al. Global analysis of alternative splicing differences between humans and chimpanzees [J]. Gene Dev, 2007, 21(22): 2963-2975
- [10] Stamm S, Ben-Ari S. Function of alternative splicing [J]. Gene, 2005, 344: 1-20
- [11] Somel M, Guo S, Fu N, et al. MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain [J]. Genome Research, 2010, 20(9): 1207
- [12] Yeo G, Holste D, Kreiman G, et al. Variation in alternative splicing across human tissues [J]. Genome Biology, 2004, 5(10): R74:1-15
- [13] Amy E, House and Kristen W. Lynch. Regulation of alternative splicing: more than just the ABCs [J]. Biol. Chem, 2008, 283, 1217-1221
- [14] Koralewski TE, Krutovsky KV. Evolution of exon-intron structure and alternative splicing [J]. PLoS ONE, 2011, 6(3): e18055:1-10
- [15] Tran H, Gourrier N, Lemerrier-Neuillet C, et al. Analysis of exonic regions involved in nuclear localization, splicing activity and dimerization of muscleblind-like-1 isoforms [J]. Journal of Biological Chemistry, 2011, 286: 16435-16446
- [16] Barash Y, Calarco JA, Gao WJ, et al. Deciphering the splicing code [J]. Nature, 2010, 465(7294): 53-59
- [17] Blencowe, B.J. Alternative splicing: New insights from global analyses [J]. Cell, 2006, 126: 37-47
- [18] Lev-Maor G, Ram O, Kim E, et al. Intronic μ s influence alternative splicing [J]. Plos Genet, 2008, 4(9): e1000204:1-12
- [19] Wang Z, Gerstein M. RNA-Seq: a revolutionary tool for transcriptomics [J]. Nature Reviews Genetics, 2009, 10(1): 57-63
- [20] Pepke S, Wold B, Mortazavi A, et al. Computation for ChIP-seq and RNA-seq studies [J]. Nature Methods, 2009, 6: S22-S32