

人类新基因 *cegl* cDNA 的克隆及表达研究

付佳琪 王 峦 陈 微 慈宏亮 李亦平*

(¹北京师范大学生命科学学院生物医学研究所, 北京 100875;

²浙江赛尔生物医学研究院, 杭州 310006;

³美国哈佛大学牙医学院 Forsyth 研究所, 波士顿 02115)

摘要 CLECT and EGF-like domain contained Gene 1 (*cegl*) 基因是用电子克隆的方法获得的人类新基因。该基因定位在人类的第14号染色体上,是一个单一外显子的基因。*cegl* 基因的cDNA长度为2050bp,通过生物信息学方法预测它包含一个1340bp的完整阅读框架,编码一个490个氨基酸的蛋白,含有CLECT、EGF-like结构域各一个。以*cegl* 基因全长编码区为探针的整体原位杂交结果显示该基因的小鼠和鸡的同源基因在各自早期胚胎头部中特异表达,并且在不同时期胚胎神经系统增殖迅速的部位中有大量的表达。RT-PCR结果显示该同源基因在小鼠成体各组织中广泛分布。这提示*cegl* 基因可能与头部生长发育有密切关系,并且对维持成体各组织的正常功能起到重要的作用。对*cegl* 基因在胚胎发育的时间和空间表达模式的研究将有助于进一步深入地揭示它在人脑的正常生长发育中的作用。

关键词: *cegl* 基因 生物信息学 EGF-like domain 脑

表皮生长因子(EGF)是一种对上皮及间质细胞的增生具有刺激作用的细胞因子。近年来的研究发现,妊娠妇女血清、羊水及胎儿脐血中都存在高浓度的EGF。因此,EGF可能与妊娠期胚胎及胎儿的生长发育有关^[1]。表皮生长因子(EGF)超家族成员包括EGF、转化生长因子 α (TGF- α)、肝素结合EGF(HB-EGF)和双调蛋白(Ar)等,所有这些细胞因子都是EGF受体(EGFR)的配基,通过与EGFR结合,并激活EGFR后发挥生物学作用^[2]。已知EGFR广泛存在于人体的许多器官和组织中,近年来有报道,在女性生殖系统如子宫平滑肌、子宫内膜、蜕膜、胎膜和胚泡滋养层细胞中有表达,可促进子宫内膜的生长和分化,在滋养细胞的形成和发育中发挥重要作用,并与分娩发动有关^[3,4]。还有研究发现,在人类和啮齿类动物植入前的胚胎也存在EGFR^[5-8],EGF家族中的HB-EGF促进胚胎细胞表面EGFR的表达,从而影响人和啮齿类动物植入前胚胎的发育,介导胚泡的着床^[9,10]。因此含EGF-like结构域的基因在胚胎的生长发育中可能具有重要的作用,具有研究价值。

以计算机和互联网为工具的新基因的搜寻和蛋白功能分析是人类基因组图谱公布后生物医学研究的热点领域。本实验室在日本的RIKEN研究小

组公布的小鼠全长cDNA库的基础上,结合人类基因组数据库,利用基因结构和相似性同源分析方法成功地筛选到了一个全长2050bp的单一外显子人类新基因*cegl*。由于鼠和人的cDNA同源性高达85%,同时在鸡的基因组中进行比对也发现了同源性很高的同源基因,所以本实验室以小鼠和鸡的胚胎为模型对该基因进行了初步的表达谱研究,以揭示它在发育过程中所具有的功能。

1 材料和方法

1.1 实验材料

健康孕鼠购自北京大学医学动物中心;M-MLV逆转录酶购于Invitrogen;高保真的Pfu DNA聚合酶购于宝生物工程大连有限公司;地高辛标记试剂盒购于Roche;所有的引物均由北京博润公司合成;测序工作由上海生工生物工程技术有限公司完成;QIAquick Gel Extraction Kit购于QIAGEN;Trizol购于北京鼎国生物技术有限责任公司。pCMV-SPORT6质粒载体购于Stratagene;大肠杆菌DH5 α 为本实验室储存;其他试剂均为国产分析纯

本文2004年3月29日收到。2004年6月30日接受。

*联系人, Tel: 010-58809071, E-mail: YPLi@forsyth.org

试剂。

1.2 方法

1.2.1 用电子克隆的方法获得 *cegl* 基因

日本的 RIKEN 公布了 21,076 条小鼠全长 cDNAs 的序列,根据与已知基因的同源性,他们把这些基因分为 9 类,其中“motif-containing protein”这一类代表着编码具有一定结构域蛋白的 573 条小鼠新基因。用同源分析的方法搜寻与这些小鼠基因同源的人类新基因获得了 13 个人类新基因,其中包括新基因 *cegl*。具体方法如下:

(1) 下载类别为“motif-containing protein”的小鼠全长新基因;

(2) BLAST 的方法在人类基因组中搜寻与小鼠全长新基因具有同源序列的片段(<http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs>);

(3) 下载人类基因组中与小鼠全长新基因具有同源序列的片段,并用 GenScan 分析这些片段中可能包含的人类新基因的候选基因(<http://genes.mit.edu/GENSCAN.html>);

(4) 用人类新基因候选基因与人类 EST 数据库比较,确定这个基因是真基因还是假基因:如果这条人类新基因候选基因与人类 EST 具有相同的序列,说明它是一个真基因;如果它不与任何人类 EST 同源,则可能是一个假基因,被暂时淘汰(如果过一段时间后,该基因与 NCBI 中新加入的 EST 同源,则认为它是真基因);

(5) 对于真的人类基因,通过把它与非冗余序列库进行比较,淘汰已知功能的人类基因,则得到人类新基因;根据该基因序列设计引物,用 RT-PCR 的方法验证。将 RT-PCR 产物插入到载体 pCMV-SPORT6 的多克隆位点中,酶切鉴定,测序,并用生物学软件对该基因的性质进行分析。

1.2.2 *cegl* 基因的生物信息学分析

利用生物信息学的软件和网上数据库等资源(如:NCBI,Expasy,VECTOR NTI 8.0,DNASTAR 等等)对这个基因的特性进行预测,其中包括:对 mRNA 的 5' 端即转录起始位点区、开放式阅读框架及其编码产物大小的预测(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>); 外显子和内含子的分布及在人类基因组上的定位。在确认该基因具有完整的蛋白编码区后,分析它所编码的蛋白的结构和功能。其中包括:编码的蛋白质在细胞内的定位(ExPASy-

ProtParam Tool); 编码蛋白质的亲疏水性质(<http://www.expasy.org/cgi-bin/protscale.pl>) 和可能的跨膜区域(<http://genome.cbs.dtu.dk/services/TMHMM/>); 编码的蛋白与已知功能蛋白的同源性比较;检测编码蛋白可能的功能区域并预测其可能具有的功能(<http://smart.embl-heidelberg.de/>)。

1.2.3 *cegl* 基因的 RT-PCR 扩增及测序检测

根据电子克隆得到的人类新基因的 cDNA 序列设计引物,用 RT-PCR 的方法在 HeLa 细胞株中扩增,检测基因的可靠性。

Sense 5'-GAGCACCAGATGTCTACCCT-3'

Anti-sense 5'-TGTCCCCTGTTCCCTAT-3'

当培养的 HeLa 细胞达到 80% 汇合的时候,使用 Trizol 试剂(Invitrogen),按照其试剂说明书使用一步法提取细胞 RNA,紫外扫描定量后稀释成 0.1g/L,使用逆转录酶 M-MLV(Invitrogen)进行 mRNA 反转录:以 oligo(dT)₁₈ 为起始引物,依次加入无 RNase 的无菌水 10.4μL,5 × RT 缓冲液 4.0μL,dNTP (250μmol/L)1.6μL,mRNA(0.1g/L) 1.0μL,oligo (dT)₁₈ 引物 2.0μL,反应条件为:65℃ 5min,37℃ 60min,75℃ 5min,MMLV 反转录酶 (200MU/L) 在 37℃ 10min 时加入,然后进行 PCR 扩增:取反转录产物 2μL 作为模板,依次加入:10×PCR 缓冲液 3μL,2.5mmol/L dNTPs 3μL,上游引物各 1μL(10pmol),下游引物各 1μL (10pmol),ddH₂O 10μL,混合后加入 2.5U Taq 酶。PCR 循环条件为:94℃ 35s,55℃ 1min,72℃ 35s,32 个循环,最后在 72℃ 延伸 7min。扩增产物进行测序验证。

1.2.4 整体原位杂交

整体原位杂交参照文献^[1]中的方法,以克隆在 pCMV-SPORT6 载体中的 2.1kb 的人类 *cegl* 基因编码区的 cDNA 为模板,使用 DIG 标记试剂盒(Roche)以体外转录的方法制备 DIG 标记的 RNA 探针,电泳检测合成结果后分别与各期的小鼠胚胎和鸡胚进行胚胎的整体原位杂交,实验过程中同时作阴性对照平行组和阳性对照平行组,阴性对照组以 *cegl* mRNA 的 Sence 链为探针,阳性对照组为标记了的、特异性在脑部表达的 CNBP 正义链。由于天数较大的胚胎表皮增厚,探针的穿透能力会受到影响^[11,12],所以对于天数较大的胚胎原位杂交在各期小鼠胚胎的石蜡切片上进行,以确定表达部位。

1.2.5 半定量 RT-PCR

设计 *cegl* 引物为:

Sense 5'-CAGGTGCTGTGACAAACA-3'

Anti-sense 5'-AAGTGTCCCGATCCAAGA-3'

扩增片段长度预计约为 500bp。内参对照为 GAPDH

Sense 5'-ACCACAGTCCATGCCATCAC-3'

Anti-sense 5'-TCCACCACCCTGTTGCTGTA-3'

扩增产物为 452bp。取小鼠成体各组织 100mg, 使用 Trizol 试剂(Invitrogen), 按照其试剂说明以一步法提取组织总 RNA。使用逆转录酶 M-MLV(Invitrogen)进行逆转录(方法同前述)。半定量 PCR 扩增条件为: 94°C 5min, 94°C 45s, 57.2 °C 1min, 72°C 1min, 循环数为 32, 72°C 7min。取 PCR 产物 10ul, 经 2% 琼脂糖凝胶电泳后照相并将扩增产物测序。

2 结 果

2.1 电子克隆的结果

利用前述方法克隆到新基因 *cegl*, 长 2050bp (图 1)。然后将该基因 cDNA 序列在 NCBI 上做 blast 的分析, 与 Genbank 中的已知序列进行比较, 证实了 *cegl* 为一新基因并向 NCBI 提交了该序列, 其 GenBank 登录号为 AY606132。

电子基因定位分析显示该基因定位在人类 14 号染色体上的 14q13 区段, 是一个单一外显子的基因。开放式阅读框架分析结果显示该基因在 28-1497bp 之间有一个完整的开放式阅读框, 编码一个 490 个氨基酸残基, 分子量约为 49KD 的蛋白。

2.2 蛋白质同源性比对

将 *cegl* 和其小鼠同源基因 (Accession No: AK004557) 各自所编码的蛋白进行同源性比对, 如图 2 所示, 显示二者具有 63.1% 的同源性。

2.3 生物信息学分析结果

2.3.1 蛋白质亲疏水性分析

对疏水图谱的分析可以用来预测跨膜螺旋。如果蛋白序列含有跨膜区提示它可能作为膜受体起作用, 也可能是定位于膜的锚定蛋白或者离子通道蛋白等。从而含有跨膜区的蛋白往往和细胞的功能状态密切相关。经分析 *cegl* 基因编码的蛋白在 350-400 氨基酸残基之间有一个明显的疏水区 (data not shown) 提示它可能作为膜受体起作用, 也可能是定位于膜的锚定蛋白或者离子通道蛋白等。

2.3.2 蛋白质跨膜区分析

蛋白的跨膜螺旋特征是通过序列分析直接得到预测并获得较好结果的一个特性。分析显示该

蛋白在 350-400 氨基酸残基之间有一个明显的跨膜区, 蛋白的 N 端位于膜外, C 端位于膜内 (图 3)。这与对蛋白的疏水性分析结果相似。

2.3.3 利用 ExPASy-ProtParam Tool 对 *cegl* 编码蛋白质进行亚细胞定位分析

结果显示该蛋白没有 N 端信号肽 (N-terminal signal peptide), 所以不是定位在胞质内。但是有 73.9 % 的可能是一个跨膜蛋白, 定位在质膜上, 这与前面的分析结果吻合。

2.3.4 用 *cegl* 编码蛋白质与已知功能的蛋白质数据库进行同源性的比对

cegl 编码蛋白与人类 C1qRp (Complement component 1, q subcomponent, receptor 1) 蛋白有 41% 的同源性。

2.3.5 蛋白质的功能预测

如图 4 所示在 31-172 氨基酸残基的位置上有一个 CLECT 结构域 (C-type lectin or carbohydrate-recognition domain)。在 245-287 氨基酸残基的位置有一个 EGF-like 结构域 (Epidermal growth factor-like)。在 399-421 氨基酸残基的位置有一个跨膜结构域。

2.4 RT-PCR 扩增及测序检测

扩增得到一个 1025bp 左右的单一条带, 经测序后与克隆得到的序列相似性达到 96.6% (data not shown), 可以肯定所得序列的可靠性。

2.5 整体原位杂交结果

第 7.5 天小鼠胚胎染色结果 (图 5-A) 显示胚胎组织被深染, 但是胚外组织没有染色; 第 9.5 天小鼠胚胎的染色结果 (图 5-B) 显示深染主要集中在前脑, 中脑和菱脑, 其中视泡和菱脑的染色最深; 第 10.5 天小鼠胚胎的染色结果 (图 5-C) 显示深染的部位还是集中在前脑、中脑、菱脑和视泡中。其中眼、菱脑还是染色最深, 同时刚刚出现并增殖迅速的鼻凹部分也出现了明显的深染; 第 11 天 (data not shown) 小鼠胚胎染色也基本相似, 集中在脑与神经管, 在神经管部位的表达有所降低, 但在尾神经孔处的表达有些升高; 图 5-D、E 分别是整体原位杂交实验的阴性和阳性对照。

在鸡胚的整体原位杂交中, 第 15 期鸡胚 (按 HH 时期计算, 受精后约 50h) (图 5-F) 染色主要集中在端脑和中脑, 又一次验证了该基因在胚胎早期于脑中表达的特异性; 在第 17 期鸡胚 (受精后约 64h) (图 5-G) 中, 该基因在头部还是有较大量的表

```

                                     M R P A F A L C L L W
1   GCAGCTCTGCGTGCGGGGCTCAGAGAATGAGGCCGGCGTTCGCCCTGTGCCTCCTCTGG
   Q A L W P G P G G G E H P T A D R A G C
61  CAGGCGCTCTGGCCCGGGCCGGCGGGCGAACACCCCACTGCCGACCGTGCTGGCTGC
   S A S G A C Y S L H H A T M K R Q A A E
121 TCGGCCTCGGGGCTGCTACAGCCTGCACCACGCTACCATGAAGCGGCAGGCGGCCGAG
   E A C I L R G G A L S T V R A G A E L R
181 GAGGCCTGCATCCTGCGAGGTGGGGCGCTCAGCACCGTGCGTGCGGGCGCCGAGCTGCGC
   A V L A L L R A G P G P G G G S K D L L
241 GCTGTGCTCGCGCTCCTGCGGGCAGGCCAGGGCCCGGAGGGGGCTCCAAAGACCTGCTG
   F W V A L E R R R S H C T L E N E P L R
301 TTCTGGGTGCGACTGGAGCGCAGGCGTTCCCACTGCACCCTGGAGAACGAGCCTTTGCGG
   G F S W L S S D P G G L E S D T L Q W V
361 GGTTTCTCCTGGCTGTCTCCGACCCCGGCGGTCTCGAAAGCGACACGCTGCAGTGGGTG
   E E P Q R S C T A R R C A V L Q A T G G
421 GAGGAGCCCAACGCTCCTGCACCAGCGGAGATGCGCGGTACTCCAGGCCACCGGTGGG
   V E P A G W K E M R C H L R A N G Y L C
481 GTCGAGCCCGCAGGCTGGAAGGAGATGCGATGCCACCTGCGCGCCAACGGCTACCTGTGC
   K Y Q F E V L C P A P R P G A A S N L S
541 AAGTACCAGTTTGAGGTCTTGTGTCTGCGCCGCGCCCCGGGGCCGCTCTAACTTGAGC
   Y R A P F Q L H S A A L D F S P P G T E
601 TATCGCGCGCCCTTCCAGCTGCACAGCGCCGCTCTGGACTTCAGTCCACCTGGGACCGAG
   V S A L C R G Q L P I S V T C I A D E I
661 GTGAGTGCCTCTGCCGGGACAGCTCCCGATCTCAGTTACTTGCATCGCGGACGAAATC
   G A R W D K L S G D V L C P C P G R Y L
721 GCGCTCGCTGGGACAACTCTCGGGCGATGTGTTGTGTCCCTGCCCGGGAGGTACCTC
   R A G K C A E L P N C L D D L G G F A C
781 CGTGCTGGCAAATGCGCAGAGCTCCCTAACTGCCTAGACGACTTGGGAGGCTTTGCCTGC
   E C A T G F E L G K D G R S C V T S G E
841 GAATGTGCTACGGGCTTCGAGCTGGGGAAGGACGGCCGCTCTTGTGTGACCAGTGGGGAA
   G Q P T L G G T G V P T R R P P A T A T
901 GGACAGCCGACCCTTGGGGGACCGGGGTGCCACCAGGCGCCCGCCGCGCCACTGCAACC
   S P V P Q R T W P I R V D E K L G E T P
961 AGCCCCGTGCCGAGAGAATGGCCAATCAGGGTTCGACGAGAAGCTGGGAGAGACACCA
   L V P E Q D N S V T S I P E I P R W G S
1021 CTTGTCCTGAACAAGACAATTCAGTAACATCTATTCTGAGATTCTCGATGGGGATCA
   Q S T M S T L Q M S L Q A E S K A T I T
1081 CAGAGCAGATGTCTACCTTCAAATGTCCCTTCAAGCCGAGTCAAAGGCCACTATCACC
   P S G S V I S K F N S T T S S A T P Q A
1141 CCATCAGGGAGCGTGATTTCCAAGTTAATTCACGACTTCTCTGCCACTCCTCAGGCT
   F D S S S A V V F I F V S T A V V V L V
1201 TTCGACTCCTCCTCTGCCGTGGTCTTCATATTTGTGAGCACAGCAGTAGTAGTGTGGTG
   I L T M T V L G L V K L C F H E S P S S
1261 ATCTTGACCATGACAGTACTGGGGCTTGTAAGCTCTGCTTTCACGAAAGCCCCTCTTCC
   Q P R K E S M G P P G L E S D P E P A A
1321 CAGCCAAGGAAGGAGTCTATGGGCCCGCGGGCCTGGAGAGTGATCCTGAGCCCCGCTGCT
   L G S S S A H C T N N G V K V G D C D L
1381 TTGGGCTCCAGTTCTGCACATTGCACAAACAATGGGGTGAAGTGGGGACTGTGATCTG
   R D R A E G A L L A E S P L G S S D A
1441 CGGGACAGAGCAGAGGGTGCCTTGCTGGCGGAGTCCCCTCTTGGCTCTAGTGATGCATAG
1501 GGAAACAGGGGACATGGGCACTCCTGTGAACAGTTTTTCACTTTTGATGAAACGGGGAAC
1561 CAAGAGGAACCTACTTGTGTAACGACAATTTCTGCAGAAATCCCCCTTCTAAATTC
1621 CCTTACTCCACTGAGGAGCTAAATCAGAACTGCACACTCCTTCCCTGATGATAGAGGAA
1681 GTGGAAGTGCCTTTAGGATGGTGATACTGGGGGACCGGGTAGTGCTGGGGAGAGATATTT
1741 TCTTATGTTTATTCGGAGAATTTGGAGAAGTGATTGAACTTTTCAAGACATTGGAAACAA
1801 ATAGAACACAATATAATTTACATTAATAATAATTTCTACCAAATGGAAAGGAAATGTT
1861 CTATGTTGTTGAGGCTAGGAGTATATTGGTTTCGAAATCCCAGGAAAAAATAAAAAATAA
1921 AAAATTAAGGATTGTTGATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
1981 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
2041 AAAAAAAAAA

```

图1 人类新基因 *cegl* 基因 cDNA 的序列

在编码区序列上方同时标出相应的氨基酸序列。起始密码子(ATG)、终止密码子(TGA)和多聚腺苷酸信号(AATAAA)均用下划线标出。

Fig.1 cDNA sequence of the novel human gene *cegl*

The letters up the cDNA open reading frame sequence showed the protein sequence. Initiative code (ATG), stop code (TAG) and polyA signal (AATAA) are underlined.

<i>ceg1</i>	1	-----MKRQAA	50
AK004557	1	MRPALALCLLCPAFWPRPGNGEHTADRAACSASGACYSLHHAPFKRRAA	50
Consensus	1	-----KR AA	50
<i>ceg1</i>	7	EEACILRGGALSTVRGELRAVLALLRAGPGPGGGSKDLLFWVALERRR	100
AK004557	51	EEACSLRGGTLSTVHGFEQAVLLLRAGPGPGGGSKDLLFWVALERSI	100
Consensus	51	EEAC LRGG LSTV AGAE AVL LLRAGPGPGGGSKDLLFWVALER	100
<i>ceg1</i>	57	SHCTLENEPLRGFSWLNSSDPGGLESDTLQWVEEPQRSCTARCAVLQATG	150
AK004557	101	SQCTQEKELRQFSWLNHPDSEDSPLPWVEEPQRSCTVRCALQATR	150
Consensus	101	S CT E EPLRGFSW L D E L WVEEPQRSCT RKCA LQAT	150
<i>ceg1</i>	107	GVEPAGWKEMRCHLRANGYLCKYQFEVLCAPRPGAASNLSRAPFQLHS	200
AK004557	151	GVEPAGWKEMRCHLRDGYLCKYQFEVLCAPRPGAASNLSQAPFRLSS	200
Consensus	151	GVEPAGWKEMRCHLR GYLCKYQFEVLCAPRPGAASNLSF APF L S	200
<i>ceg1</i>	157	ALDFSPPGTEVSAICRGQLPISVTCIAEIGARWDKLSGDVLCPCPGR	250
AK004557	201	ALDFSPPGTEVSAICPGDLSISSTCIQETSANHDGLFPGTVLCPCSGR	250
Consensus	201	AALDFSPPGTEVSALC G L IS TCI DE A WD L G VLCPC GR	250
<i>ceg1</i>	206	YLRAGKCAELPNCLDDLGGFACECATGFELGKDRSCVTSGEGOPTLGGT	300
AK004557	251	YLLAGKVELPDCLDHLGDFTCCEAVGFELGKDRSCETKVEEQLTLEGT	300
Consensus	251	YL AGKC ELP CLD LG F CECA GFELGKDRSC T E Q TL GT	300
<i>ceg1</i>	256	GPTRRPPATATSPVPRRTWPIRVDKLGETPLVPEQDNSITSPETPRW	350
AK004557	301	KPTNRVTATPAGAVTPTWPGQYKPGEPQTEILOW	350
Consensus	301	LPTR AT V NRTWP V DK GE M I EI W	350
<i>ceg1</i>	306	GQSTSTQMLQAESKATITPSGSVSKNSTTSSATPQAFDSSAVV	400
AK004557	341	GQSTPTQKPTKPKVTGTPSGSVLNITSPPVSLTFDCSTVV	400
Consensus	351	GSQSTL TIQ S Q K T TPSGSVI F S S FDS S VV	400
<i>ceg1</i>	356	FIFVSTAVMLVLTITVLGLVKLGFHESPSSQPRKEGPPGIESDPEP	450
AK004557	389	FILVSIAYMLVLTITVLGLFKLGFHKSRSRTGKGDSPGECDAEA	450
Consensus	401	FI VS AVIVLVILTITVLGL KLCFH S SS K AL PGLE D E	450
<i>ceg1</i>	406	ALGSSSAHCTNNGVKVGDCLDRDRAEGALLAESPLGSSDA	491
AK004557	439	TLHSSSTQCTDIGVKSQTVA	491
Consensus	451	AL SS CT GVK G	491

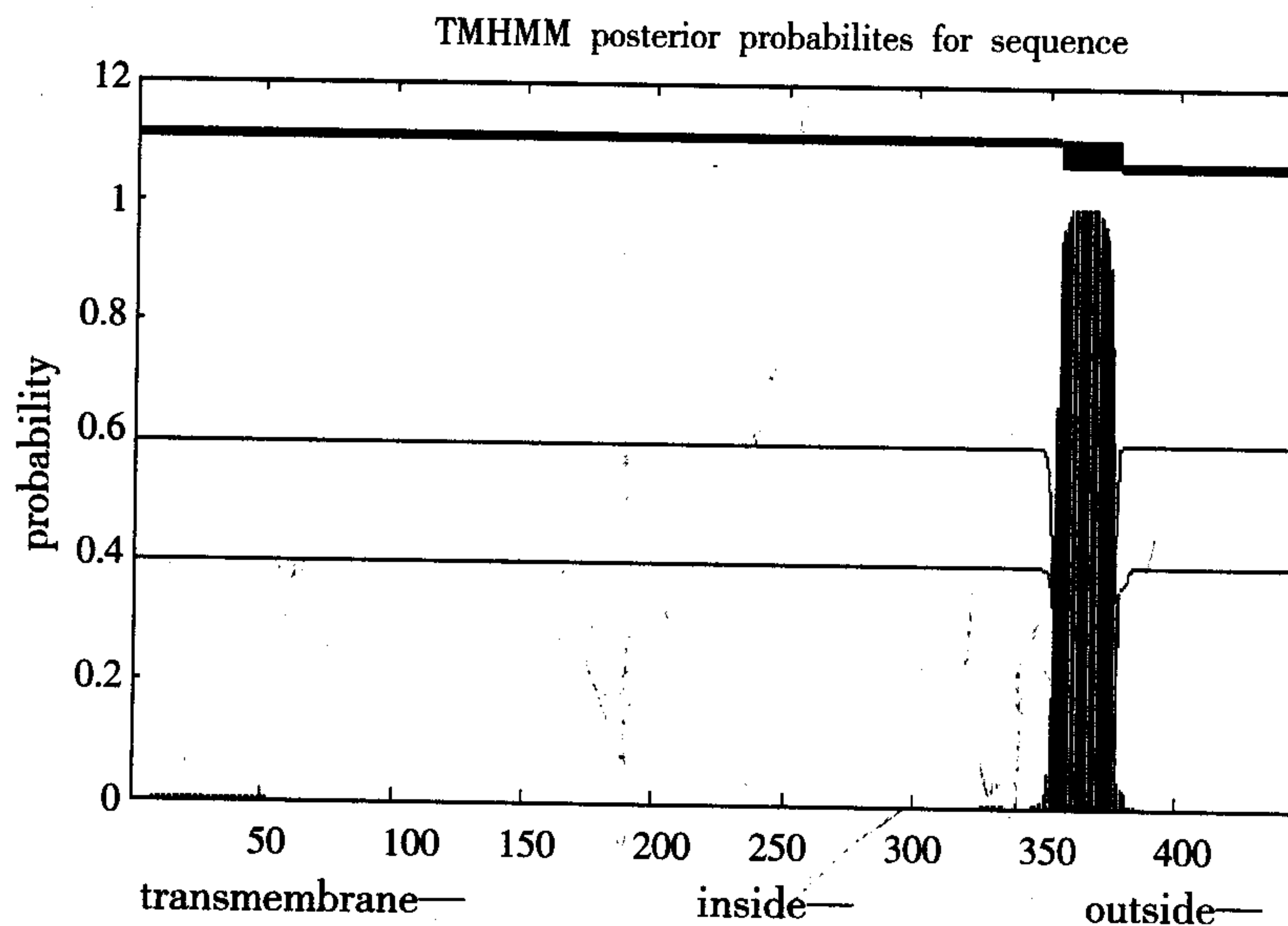
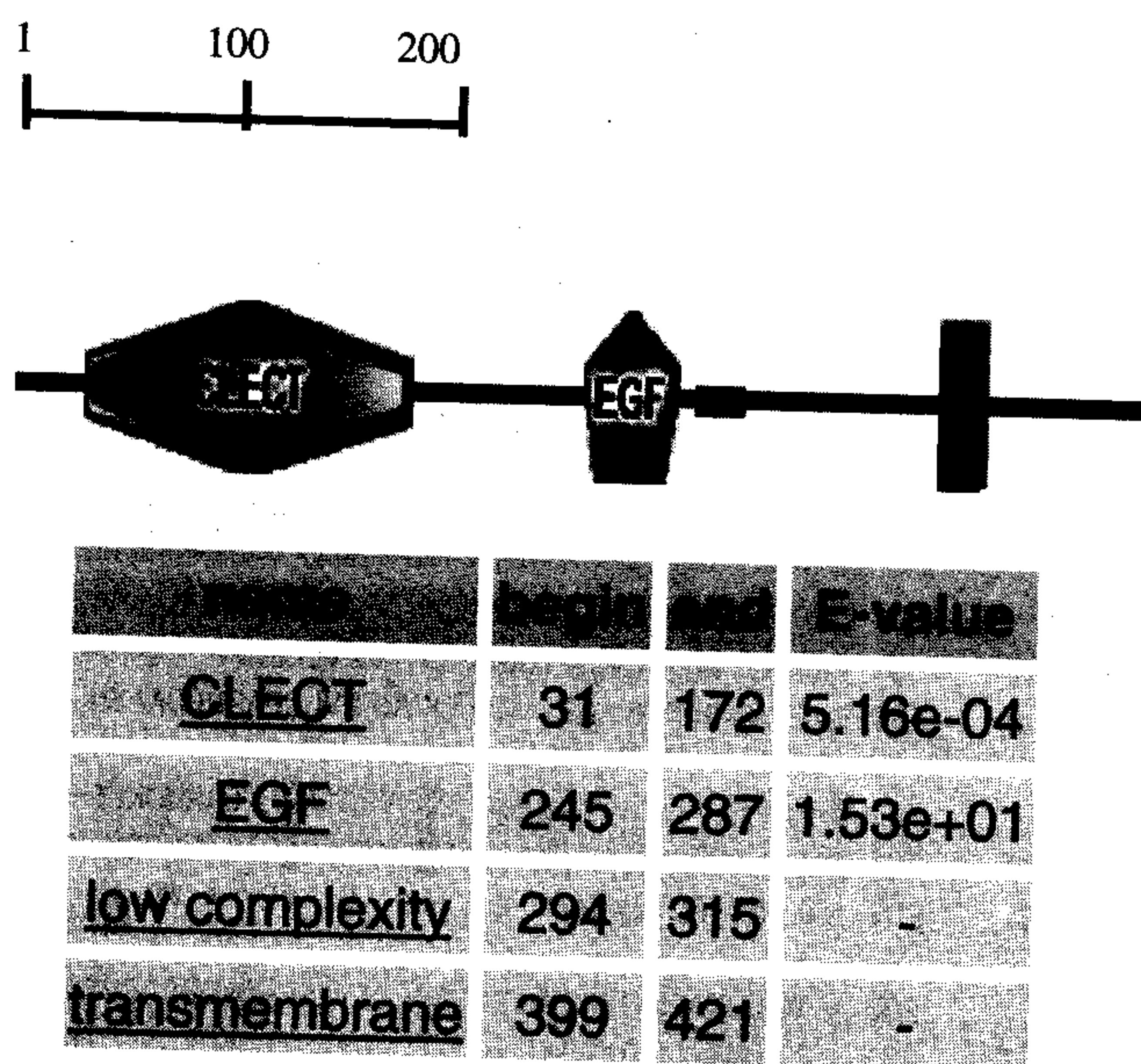
图2 *ceg1* 编码蛋白与小鼠同源基因 AK004557 编码的蛋白的同源性比较Fig.2 Homologous comparison of *ceg1* coding protein and mouse AK004557 coding protein图3 *ceg1* 编码蛋白的跨膜结构域分析

Fig.3 Analyzing of trans-membrane domain

图4 *cegl* 编码蛋白质的功能域分析

图中菱形表示 CLECT domain, 五边形表示 EGF-like domain, 小长方形表示 low complexity 区域, 大长方形表示跨膜结构。

Fig.4 Analysing of CEG1 function domain

达,但是表达部位已经不再局限在头部;在第 19 期鸡胚(受精后 72h)(图 5-H)中,该基因表达较广泛。

第 13.5 天鼠胚脑部切片(图 5-I)和 14.5 天鼠胚整体切片(图 5-J)的切片原位杂交结果也显示在脑室表面细胞增殖迅速的部位有相对的深染(如箭头所示)。

2.6 半定量 RT-PCR 结果

图 6 结果显示在成体鼠的组织中除眼部以外该基因的同源基因普遍表达,其中在脾脏、肾脏和心脏中表达量比较高。提示该基因可能在胚胎早期的脑部发育中起重要作用,而在成体中的普遍表达说明该基因可能与维持机体的正常功能有关。

3 讨 论

cegl 基因是利用生物信息学的方法进行电子克隆并通过生物学实验的方法进行序列验证得到的人类新基因。利用 Vector NTI 等相关软件预测该基因编码一个 490 个氨基酸残基,分子量约为 49KD 的蛋白。该蛋白有 1 个疏水跨膜结构,在胞外区有一个 EGF-like 结构域和一个 CLECT 结构域。这样的结构与人类及鼠源的 C1qRp 蛋白非常相似^[13]。鼠源 C1qRp 基因作为胚胎干细胞的标志,与造血系统和心脏血管的发育有关^[14];而人源 C1qRp 受激后能够增加单核细胞的噬菌作用^[15],并与凝血系统及其辅助系统有关^[16-18]。具有 CLECT 结构域的蛋白大多是一个膜上的 Ca^{2+} 依赖性的糖蛋白结合受体,介

导糖蛋白的内吞作用;含有 EGF-like 结构域的蛋白作为膜受体的一个重要的功能是参与细胞的生长调控^[19],具有一个或多个 EGF-like 结构域的蛋白如 EGF、TGF- α 等在肿瘤组织中多为高表达^[20-22],结合疏水性分析结果初步推测这个新基因可能作为膜上具有一个跨膜区的 EGF 受体而发挥对细胞增殖的调控作用。

在小鼠胚胎的整体原位杂交中,由于第 9.5 天时小鼠胚胎的神经管还没有闭合,神经嵴开始发生,管壁增厚,此期神经管上皮细胞增殖旺盛,所以图 5-B 中深染主要集中在前脑,中脑和菱脑,其中视泡和菱脑的染色最深。而第 9.75 天时小鼠胚胎的前脑膨大,眼泡末端膨大,眼泡渐与头部外胚层接近,并诱导其局部增厚,变为晶状体板。因此图 5-C 中深染的部位还是集中在前脑、中脑、菱脑和视泡中。其中眼部染色最深,结合 *cegl* 的功能区预测此眼泡深染有可能反映了其外胚层上皮细胞增厚的趋势。因此在各期脑室增殖迅速的部位都有该基因的大量表达,验证了 *cegl* 基因作为膜受体参与细胞的增殖调控的推测。

在鸡胚的整体原位杂交中,第 15 期鸡胚深染主要集中在端脑和中脑,又一次验证了该基因于胚胎早期在脑中表达的特异性。随着鸡胚的发育,在第 17 和 19 期鸡胚中该基因表达渐趋广泛。这与在成体小鼠组织中所做的 RT-PCR 的结果一致。因此该基因可能对于维持各发育成熟组织的正常功能

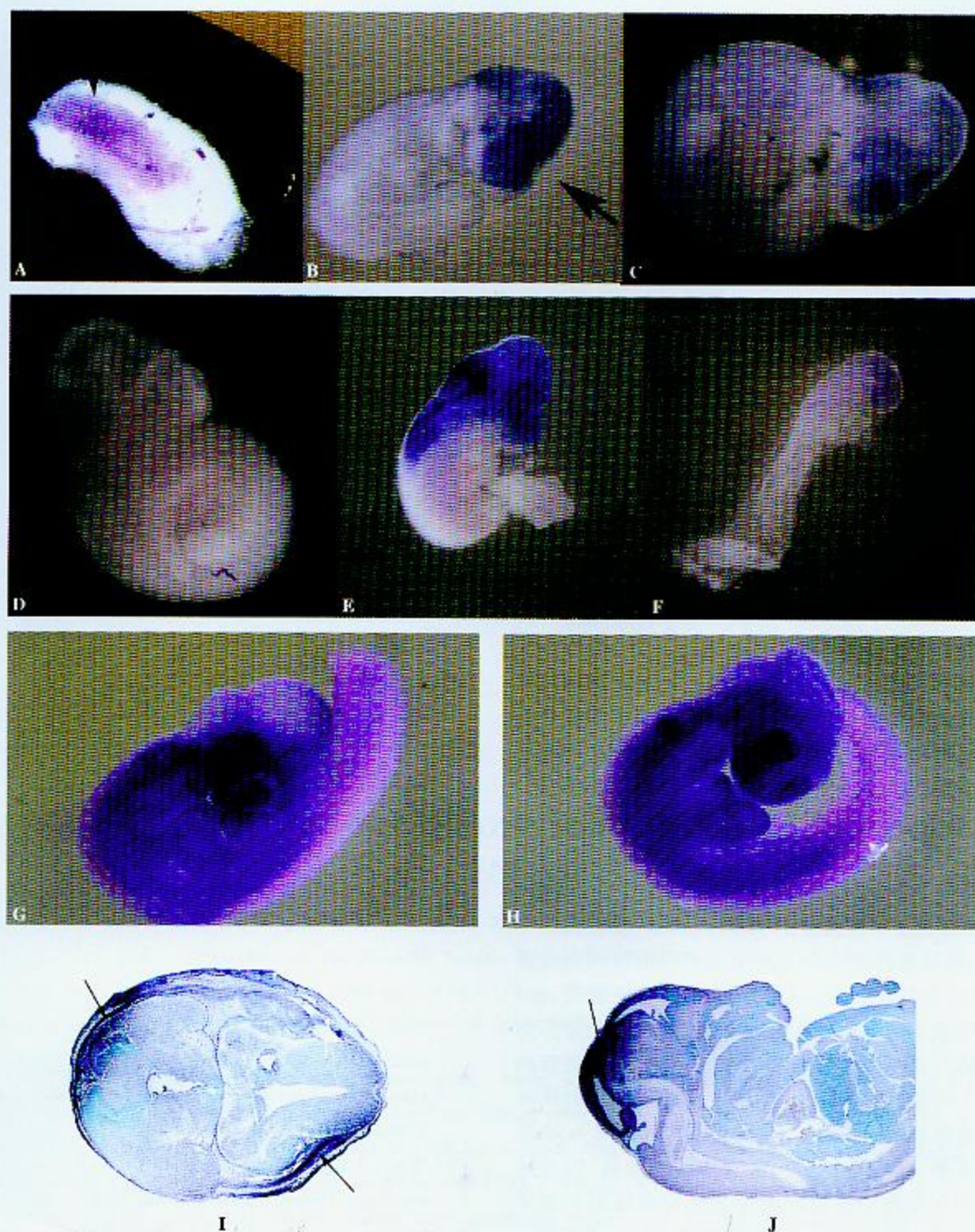


图5 在小鼠胚胎和鸡胚中进行的整体原位杂交

图 A-C 为小鼠各期的杂交结果, 箭头显示深染部位; 图 D 为阴性对照, 图 E 为使用 CNBP 基因作为探针的阳性对照。图 F-H 为各时期鸡胚的杂交结果。小鼠的切片原位杂交图 I 为 13.5 天小鼠胚胎的脑切片, 图 J 为 14.5 天小鼠胚胎的整体切片结果。箭头示深染部位。

Fig.5 Whole-mount *in situ* hybridization of mouse embryo and chicken embryo and Section *in situ* hybridization of 13.5 day mouse embryo brain section and 14.5 day mouse embryo section

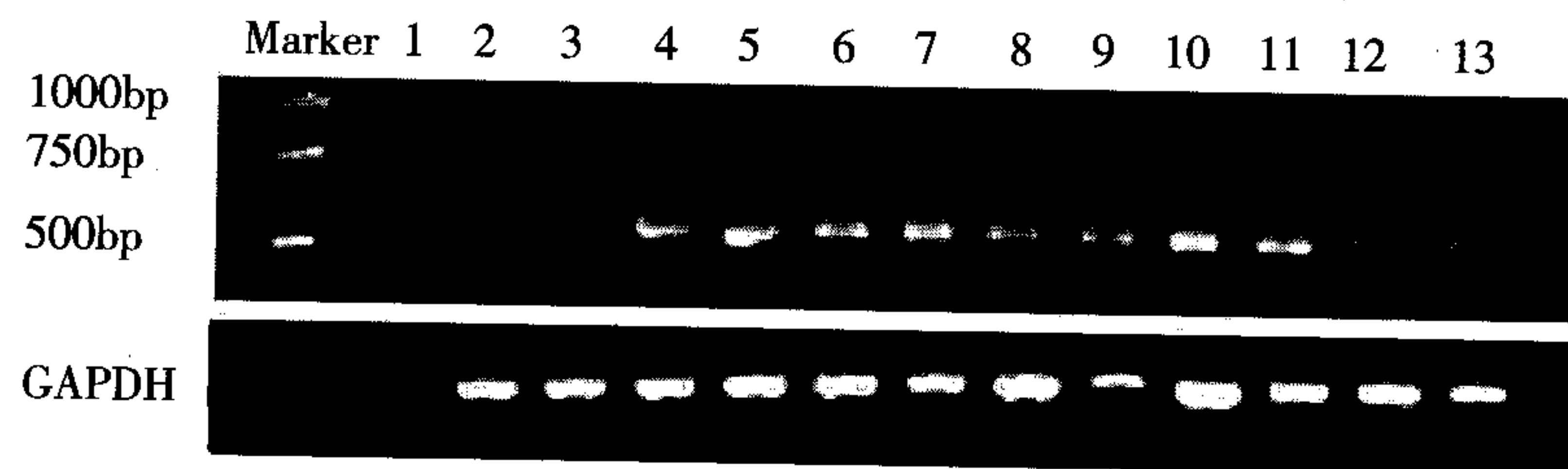


图6 小鼠成体组织的 RT-PCR

上图:从第 1-13 泳道:阴性对照,眼,小脑,子宫,心脏,肺部,肌肉,全脑,睾丸,肾脏,脾脏,肠,肝脏。下图是用作定量的 GAPDH。

Fig.6 The RT-PCR result of Matured mouse organs

Upper pattern: from Lane 1-13: Negative-control, eye, cerebella, uterus, heart, lung, muscle, brain, testis, kidney, spleen, intestines, liver. GAPDH was displayed lower as a control for differences in loading.

起到重要的作用。同时在小鼠胚胎和鸡胚中表达模式的相似性也在一定程度上说明了该基因在高等真核生物中的表达方式相当保守,推测它在人类的发育和正常生长过程中有可能起到类似的作用。

胚胎发育早期正是脑部生长增殖旺盛的时期,从以上的整体原位杂交结果可以看出,作为一个相对保守的基因,*cegl* 基因在胚胎发育时期于头部特异表达,而在成体各个组织中广泛分布,提示 *cegl* 基因可能在胚胎发育中与头部的发育,尤其是与神经细胞的增殖密切相关,同时也对维持成体各个组织的正常功能起到重要的作用。这说明 *cegl* 可能是一个重要的膜结合受体,通过胞外的两个 EGF-like 和 CLECT 结构域与细胞生长因子结合从而调节细胞(包括神经上皮细胞)的增殖,从而参与了细胞的增殖调控和胚胎发育等重要过程。为了进一步探讨 *cegl* 的功能,本实验室正在进行一些后续实验其中包括:在用整体原位杂交确定该基因的时间表达谱后,通过切片原位杂交及荧光染色的方法研究 *cegl* 的空间表达情况;构建原核蛋白表达载体制造多克隆抗体对该基因进行转录后水平的研究;用最新的 siRNA 的实验方法对该基因进行 knockdown 的研究以明确该基因的确切功能。我们希望通过进一步的研究,明确它们和细胞增殖的关系,以及这个基因在细胞中表达的改变与脑发育细胞生物学行为之间的关系。对 *cegl* 基因的研究将有助于进一步揭示脑的发育以及相关疾病的分子机制。

参 考 文 献

- [1] 王渠源、杨秀琴、王丽、王少为、张为远,1998,表皮生长因子及其受体与胎儿出生体重的关系。中华妇产科杂志, 11: 664-666。
[2] Stewart, C.L., S.L., Abbondanzo & E.B.Cullinan, 1995,

- Regulation by maternally derived cytokines of pre-implantation development and uterine receptiveness. *Contracept Fertil. Sex*, 23: 555-561.
[3] 张丽凤、苏应宽、盖凌、王春霞、刘辉贤,1999,EGF 受体在子宫内膜、早孕蜕膜及滋养细胞中的表达。中华妇产科杂志, 34: 357-359。
[4] 陈健美、曾文洁、吴佳捷,2000,子宫平滑肌和胎膜组织 EGF 受体 mRNA 表达水平与分娩发动的关系。中华妇产科杂志, 35: 29-31。
[5] Chia, C. M., Winston, R. M. & Handyside, A. H., 1995, EGF, TGF- α and EGFR expression in human preimplantation embryos. *Development*, 121: 299-307.
[6] Smotrich, D.B., R.J. Stillman, E.A. Widra, P.R. Gindoff, P. Kaplan, M. Graubert & K. E. Johnson. 1996, Immunocytochemical localization of growth factors and their receptors in human pre-embryos and Fallopian tubes. *Hum Reprod*, 11: 184-190.
[7] Kim, C.H., H.D. Chae, Y.P. Cheon, B.M. Kang, Y.S. Chang & J. E. Mok, 1999, The effect of epidermal growth factor on the preimplantation development, implantation and its receptor expression in mouse embryos. *J Obstet Gynaecol Res*, 25: 87-93.
[8] Wiley, L.M., J.X. Wu, I. Harari & E.D. Adamson, 1992, Epidermal Growth Factor Receptor mRNA and Protein Increase after the Four-Cell Preimplantation Stage in Murine Development. *Dev Biol*, 149: 247-260.
[9] Martin, K.L., D.H. Barlow & I.L. Sargent, 1998, Heparin-binding epidermal growth factor significantly improves human blastocyst development and hatching in serum-free medium. *Hum Reprod*, 13: 1645-1652.
[10] Das, S.K., X. N. Wang, D. Damm, J. A. Abraham, M. Klagsbrun, G.K. Andrews & S. K. Dey, 1994, Heparin-

- binding EGF-like growth factor gene is involved in the mouse uterus temporally by the blastocyst solely at the sites of its apposition: a possible ligand for interaction with blastocyst EGF receptor in implantation. *Development*, **120**: 1071-1083.
- [11] Wilkinson, D. G., 1993, In situ hybridization. In: Stern CD, Holland PWH(eds.) *Essential Development Biology: A Practical Approach*[M]. IRL Press, Oxford, UK,257-274.
- [12] Vladimir Pekarik, Dimitris Bourikas, Nicola Miglino, Pascal Joset, Stephan Preiswerk & Esther T. Stoeckli, 2003, Screening for gene function in chicken embryo using RNAi and electroporation. *Nat Biotechnol*, **21**(1):93-96.
- [13] Tenner, A. J., 1995, Annosebinding protein (MBP) enhances mononuclear phagocyte function via a receptor that contains the 126,000 Mr component of the C1q receptor[J]. *Immunity*, **3**: 485-493.
- [14] Guan, E., S. L.Robinson & E. B. Goodman, 1994, Cell surface protein identified on phagocytic cells modulates the C1q-mediated enhancement of phagocytosis. *J. Immunol.*, **152**: 4005-4016.
- [15] Nepomuceno, R. R., A. H. Henschen-Edman & W. H. Burgess, 1997, cDNA cloning and primary sequence analysis of C1qRp, the human C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro. *Immunity*, **6**: 119-129.
- [16] Tenner, A. J, S. L. Robinson & R. A. B. Ezekowitz, 1999, Membrane receptors for soluble defense collagens. *Curr. Opin. Immunol*, **11**: 34-41.
- [17] Petrenko, O., A.Beavis & M.Klaine, 1999, The molecular characterization of the fetal stem cell marker AA4. *Immunity*, **10**: 691-700.
- [18] Guro, Lovik, Vaage, John T. & Erik Dissen, 2000, Characterization and molecular cloning of rat C1qRp, a receptor on NK cells. *Eur. J. Immunol.* **30**: 3355-3362.
- [19] Johan Stenflo, Yvonne Stenberg & Andreas Muranyi,2000, Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions. *Biochimica et Biophysica Acta* **1477**: 51-63.
- [20] Choi, J.H., H.C. Kim & H. Y. Lim, 1999, Detection of transforming growth factor-alpha in the serum of gastric carcinoma patients. *Oncology*, **57**: 236-241.
- [21] Xia, L., Y.Z. Yuan & C.D. Xu, 2002, Effects of epidermal growth factor on the growth of human gastric cancer cell and the implanted tumor of nude mice[J]. *World J Gastroenterol*, **8**: 455-458.
- [22] Ghanem, M.A., T.H. Van Der KwastDen & J.C. Hollander, 2001, Expression and prognostic value of epidermal growth factor receptor, transforming growth factor-alpha, and c-erb B-2 in nephroblastoma. *Cancer*, **92**: 3120-3129.

CLONING AND CHARACTERIZATION OF A NOVEL HUMAN *cegl* GENE cDNA

FU Jia Qi¹ WANG Luan^{1,2} CHEN Wei³ CI Hong Liang^{1,2} LI Yi Ping^{1,3}*

¹Biomedical Research Institute, College of Life Sciences, Beijing Normal University, Beijing 100875, China;

²Zhejiang Cell Biomedical research College, Hangzhou 310006, China;

³The Forsyth Institute, Harvard School of Dental Medicine, Boston MA 02115, USA)

ABSTRACT *cegl* gene was a novel human gene, which was cloned by bioinformatics research and RT-PCR. It was a single exon gene and located on human chromosome 14. The length of the cDNA was 2050bp. Bioinformatics analysis predicted a 1340 bp complete open reading frame (ORF) which encoded a 446 amino acid protein, containing an EGF-like and CLECT domain. We found that the homologous genes of *cegl* in mouse embryo and chicken embryo were specifically expressed in the brain by in-situ hybridization. The result of RT-PCR of the mature mouse organs showed it was widely expressed in many organs. The result indicates that *cegl* gene may have an essential role in the development of brain and the maintenance of the organs' normal function. The analysis of expression and function profile of *cegl* gene may provide valuable insights into the functions of *cegl* in the development and function of human body.

Key words: *cegl* gene. Bioinformatics. EGF-like domain. Brain

*Corresponding author, Tel: 010-58809071, E-mail: YPLi@forsyth.org