

初探 E2F1 对基因组内跨膜信号转导基因的调控作用 *

孙红梅 马 晔 陈 方 徐良德[△]

(哈尔滨医科大学生物信息科学与技术学院 黑龙江 哈尔滨 150081)

摘要 目的:从基因组全局性角度研究 E2F1 对包括离子通道和 G 蛋白偶联受体在内的重要的跨膜信号转导基因的调控作用。方法:对从 TRED 和 IUPHAR 获取的 E2F1 靶基因数据和基因组离子通道及 G 蛋白偶联受体基因数据进行数据联配,获取 E2F1 调控的离子通道基因(ICG)和 G 蛋白偶联受体基因,并对调控基因进行家族富集性分析和组织特异性分析。结果:发现 E2F1 对 7 个 ICG 具有调控作用,且具有钾离子通道富集性,调控的离子通道基因具有心脏、脑、消化系统组织表达特异性。获得的 11 个受 E2F1 调控的 G 蛋白偶联受体基因家族富集性不明显,组织特异性表达不一致。结论: E2F1 可能通过对钾离子通道基因的表达调控,实现对相应组织的作用机理影响,相应调控作用的紊乱也将导致心脏、脑或消化系统疾病,但难以确定 E2F1 对 GPCR 的调控作用效果。

关键词 转录调控;离子通道;G 蛋白偶联受体;组织特异性;富集分析

中图分类号 Q75,Q811.4,R318.04,R319 **文献标识码** A **文章编号** 1673-6273(2012)19-3614-04

Study on the E2F1 on Key Trans-membrane Signal Transduction Genes in Genome Regulation*

SUN Hong-mei, MA Ye, CHEN Fang, XU Liang-de[△]

(Bioinformatics, Harbin Medical University, Harbin, Heilongjiang, 150081, China)

ABSTRACT Objective: Our study told about E2F1 regulation on the important trans-membrane transduction genes from the point of genome. **Methods:** We use TRED web service and IUPHAR database to realize E2F1 target genes, ion channel genes and G-protein-coupled receptor genes extraction. We achieve E2F1 regulation signal transduction genes by data alignment. Then, we introduce tissue-specific analysis and gene family enrichment analysis into the regulation data. **Results:** We find 7 ICG genes regulated by E2F1 significantly enrich in potassium ion channels, while these genes have a high heart, brain and digestive system tissue expression. E2F1 regulation of 11 GPCR genes did not show significant enrichment of family, and the tissue-specific expression is not clear. **Conclusions:** E2F1 maybe affect the cell physics such as cardiac muscle cell by the regulation of potassium channels, and lead to the heart, brain or digestive system diseases.

Key words: Transcription regulation; Ion channels; G-protein coupled receptors; Tissue specificity; Enrichment analysis

Chinese Library Classification: Q75, Q811.4, R318.04, R319 **Document code:** A

Article ID: 1673-6273(2012)19-3614-04

前言

E2F1 是目前已知的最重要的转录因子之一,它通过调控靶基因表达水平,在细胞分化、增殖、发育、凋亡过程中发挥重要作用^[1]。离子通道(Ion channel)是细胞膜上的一类特殊亲水性蛋白质微通道,广泛参与细胞信息转导过程,并能够对细胞生理调控产生重要的影响^[2,3]。现有研究表明 E2F1 的调控异常与心血管病的发生存在密切联系,这一过程可能是由 E2F1 调控对多种信号转导基因特别是对离子通道基因的转录调控作用造成的^[4,5]。在此基础上,我们期望通过 E2F1 靶基因预测,从基因组层面研究 E2F1 对重要的跨膜信号转导基因,特别是对离子通道基因的调控作用,从而为进一步研究 E2F1 在疾病发生中可能存在的作用机理提供有益的提示。

1 数据和方法

1.1 数据获取

1.1.1 E2F1 靶基因的收集与预测 TRED 是用于研究基因调控和功能的数据库^[6]。它收集实验证实的哺乳动物顺式调控元件,像启动子和结合基序^[7,8],和反式调控元件,像转录因子,并开发了转录调控预测平台。我们借助 TRED 实现 E2F1 已知靶基因的提取,并进行靶基因的预测。

1.1.2 离子通道基因(ICG)的提取和 G 蛋白偶联受体基因(GPCR)的提取 IUPHAR 数据库包含关于 G 蛋白偶联受体、电压门控离子通道、配体门控离子通道的药理学的、功能的和病理生理的信息^[9]。我们通过 IUPHAR 数据库实现了基因组已知离子通道基因和 G 蛋白偶联受体基因的提取,获得 195 个

* 基金项目 黑龙江省教育厅科学技术研究项目(12511273)

作者简介 孙红梅(1982),女,硕士研究生,助教,主要研究方向:生物信息学研究方法,分子遗传学,

电话 0451-86669617, E-mail: sunhongmei@hrbmu.edu.cn

[△]通讯作者 徐良德, E-mail: xuliangde@hrbmu.edu.cn

(收稿日期 2012-03-23 接受日期 2012-04-18)

离子通道基因 (ICG) 和 368 个 G 蛋白偶联受体基因 (GPCRG)。

1.2 方法

1.2.1 获取 E2F1 调控的 ICG 和 GPCRG 将 TRED 网络服务器预测的 E2F1 靶基因和 IUPHAR 数据库下载的基因组离子通道基因和 G 蛋白偶联受体基因数据进行联配,识别 E2F1 调控的 ICG 和 GPCRG。

1.2.2 E2F1 调控基因的家族富集分析和组织特异性分析 我们使用超几何分布概率统计量对 E2F1 调控跨膜信号转导基因进行家族富集分析^[10],公式如下:

$$P=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

此处 N 为已知的人类基因总数,M 为预测的 E2F1 调控基因数,n 为基因组中已知的某类离子通道数目,即 K⁺、Ca²⁺、正电荷离子等各种离子通道的数目,以及 G 蛋白偶联受体各类受体数目,m 为 E2F1 调控的某种离子通道数目或调控的某个家族的 G 蛋白偶联受体数目。

我们进一步通过 Uniprot 数据库进行文本检索,查询数据

库中基因产物的组织定位情况,进行 E2F1 调控信号转导基因的组织特异性分析^[11]。

2 结果与分析

2.1 数据的分类整理

基于 TRED 网络服务器,我们获得 586 个高可靠性的 E2F1 调控靶基因,表 1 显示了 E2F1 靶基因预测数量及其对应的启动子和结合质量。

我们从 IUPHAR 数据库获得的 195 个离子通道基因中,有 142 个为电压门控离子通道,涉及 4 大类离子,9 个亚家族(见表 2) 有 53 个配体门控离子通道,涉及 6 类配体,6 个亚家族。我们从 IUPHAR 数据库获得的 368 个 G 蛋白偶联受体分为 4 大类,涉及到 62 个亚家族(见表 3)。

2.2 E2F1 调控的离子通道基因分析

通过进行离子通道数据集和 E2F1 靶基因数据集联配,发现 E2F1 调控基因中有 7 个是离子通道基因。E2F1 调控的离子通道基因均为电压门控离子通道,其中有 5 个是钾离子通道,2 个是正电荷离子通道(见表 4)。

表 1 E2F1 靶基因预测

Table 1 Prediction of E2F1 target genes

Promoter	Number of genes	Combined	Number of genes
Known, curated	126	Know	139
Known	321	Likely	18
Refseq, predicted	105	Maybe	429
Refseq	34		

表 2 离子通道基因的家族分类

Table 2 The classification of ion channel gene family

Voltage gated ion channel			Ligand gated ion channel		
Ion type	Number of genes	Number of subfamilies	Ligand type	Number of genes	Number of subfamilies
K ⁺	78	4	5 hydroxytryptamine	5	1
Ca ²⁺	17	2	GABA	19	1
Na ⁺	9	1	Glycine	5	1
+	38	2	Acetylcholine	16	1
			ATP	7	1
			Zn ²⁺	1	1

Note: "+", that is a positively charged ion, the same meaning as the following appears.

表 3 G 蛋白偶联受体基因的家族分类

Table 3 The classification of G-protein coupled receptors gene family

GPCR type	Number of genes	Number of subfamilies
Class A	284	50
Class B	50	6
Class C	23	5
Frizzled	11	1

表 4 E2F1 调控的离子通道

Table 4 The ion channels of the E2F1 regulation

Gene Symbol	Ion type	Type	Channel family
KCNJ2	K ⁺	Voltage	Inwardly Rectifying Potassium Channels
KCNJ16	K ⁺	Voltage	Inwardly Rectifying Potassium Channels
KCNK10	K ⁺	Voltage	Two-P Potassium Channels
KCNC4	K ⁺	Voltage	Voltage-Gated Potassium Channels
KCNG4	K ⁺	Voltage	Voltage-Gated Potassium Channels
TRPC7	+	Voltage	Transient Receptor Potential Channels
TRPM2	+	Voltage	Transient Receptor Potential Channels

这里只涉及到钾离子和正电荷离子,而正电荷离子数不足以达到统计要求,所以只进行 E2F1 调控钾离子通道富集分析^[12-14],即 N=33202,M=78,n=586,m=5,经过计算 $P=0.0098 < 0.01$,即 E2F1 调控基因显著的富集在钾离子通道家族,表明 E2F1 对钾离子基因的转录调控可能是其影响信号转导过程的重要方式。

E2F1 调控离子通道的组织特异性分析,发现除 TRPC7 信息未知外,其他的离子通道基因均具有较高的心脏组织、脑组织、消化系统组织的分布倾向,这提示 E2F1 基因可能通过对这

些离子通道的调控,影响心脏、脑和消化系统的生理和病理机制。

2.3 E2F1 调控 G 蛋白偶联受体基因分析

表 5 列出了 E2F1 调控的 11 个 G 蛋白偶联受体基因。采用超几何分布方法^[15,16]对 I 类 GPCR 进行富集性分析获得 $P=0.065$,具有一定的富集倾向,但由于 GPCR 本身作用形式多样,不足以说明 E2F1 能够通过一类 GPCR 进行信号转导调控,在此基础上,我们对调控基因进行功能注释和组织特异性注释^[17,18],期望从中获得有用的信息。

表 5 E2F1 调控的 GPCR 基因信息

Table 5 The information of GPCR gene with E2F1 regulation

Gene Symbol	GPCR family	Type	Location	Promoter	Combined
HTR1F	5-Hydroxytryptamine receptors	Class A	3p12	3.2:refseq	Maybe
CXCR4	Chemokine receptors	Class A	2q21	2:known	Known
FPR1	Formylpeptide receptors	Class A	19q13.4	2:known	Maybe
P2RY4	P2Y receptors	Class A	Xq13	3.2:refseq	Maybe
PTGER2	Prostanoid receptors	Class A	14q22	2:known	Known
SSTR1	Somatostatin receptors	Class A	14q13	3.1:refseq, predicted	Maybe
OXTR	Vasopressin and oxytocin receptors	Class A	3p25	2:known	Known
OPN3	Class A Orphans	Class A	1q43	2:known	Maybe
EMR1	Class B Orphans	Class B	19p13.3	3.2:refseq	Maybe
GPR114	Class B Orphans	Class B	16q13	3.2:refseq	Maybe
CASR	Calcium-sensing receptors	Class C	3q21-24	2:known	Maybe

利用 Uniprot 数据进行 E2F1 调控 GPCR 基因的组织特异性分析,发现 CXCR4、EMR1、CASR 基因均为广泛表达基因,表明他们是细胞生长、发育过程中普遍发挥作用的重要基因,FPR1 特异的在中性粒细胞中表达,P2RY4 特异的在胰腺组织中表达,PTGER2 在胎盘和肺组织中表达,SSTR1 倾向于在胚胎时期组织中表达,OPN3 倾向于在神经系统细胞中表达。E2F1 调控 GPCR 基因的组织表达信息提示我们没有特定的证据说明 E2F1 调控 GPCR 能够影响心脏组织机能。

3 讨论

E2F1 是细胞生长、发育、增殖、分化过程中行使转录调控作用,产生重要影响的转录因子。由于每个转录因子的结合位点通常都有特定的模式,被称为基序^[7,8],因此找到这些特定的序列片段对研究 E2F1 的转录调控有着重要意义。在我们的研究中对 E2F1 的靶基因进行了收集和预测,并对基因组离子通道基因和 G 蛋白偶联受体基因进行了整理和分类,在此基础上研究了 E2F1 对重要的跨膜信号转导基因的调控作用。

经过我们的研究发现,E2F1 对 7 个离子通道基因具有调控作用。E2F1 调控离子通道基因具有钾离子通道富集性,调控基因同时具有心脏、脑、消化系统组织表达特异性,这提示

E2F1 可能通过对钾离子通道基因的表达调控,实现对相应组织的作用机理影响,相应调控作用的紊乱也将导致心脏、脑或消化系统疾病的发生。

在对 G 蛋白偶联受体与 E2F1 调控研究中,我们也获得了 11 个受 E2F1 调控的 G 蛋白偶联受体基因。但这些基因的表达不具有一致的表达特异性,难以确定 E2F1 对 GPCR 的调控作用效果。

我们的研究从基因组全局性角度出发,在一定程度上表明 E2F1 可以通过对离子通道、G 蛋白偶联受体等重要的跨膜信号转导基因的转录调控,实现对某些组织中相应基因的表达产生影响,从而引起生理和病理机制的改变,对于了解心脏病、脑病,以及消化系统疾病的发生有一定借鉴意义。

参考文献(References)

- [1] Muller H, Bracken AP, Vernell R, et al. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis[J]. *Genes Dev*,2001,15(3):267-285
- [2] Gaborit N, Le Bouter S, Szuts V, et al. Regional and tissue specific transcript signatures of ion channel genes in the non-diseased human heart[J]. *J Physiol*,2007,582(Pt 2):675-693
- [3] Greener ID, Tellez JO, Dobrzynski H, et al. Ion channel transcript expression at the rabbit atrioventricular conduction axis [J]. *Circ Arrhythm Electrophysiol*,2009,2(3):305-315
- [4] Baldini E, Camerini A, Sgambato A et al. Cyclin A and E2F1 overexpression correlate with reduced disease-free survival in node-negative breast cancer patients[J]. *Anticancer Res*,2006,26(6B): 4415-4421
- [5] Movassagh M, Bicknell KA, Brooks G. Characterisation and regulation of E2F-6 and E2F-6b in the rat heart: a potential target for myocardial regeneration[J]. *J Pharm Pharmacol*,2006,58(1):73-78
- [6] Jiang C, Xuan Z, Zhao F, et al. TRED: a transcriptional regulatory element database, new entries and other development [J]. *Nucleic Acids Res*, 2007,35(Database issue):137-140
- [7] Patel PD, Bochar DA, Turner DL, et al. Regulation of tryptophan hydroxylase-2 gene expression by a bipartite RE-1 silencer of transcription/neuron restrictive silencing factor (REST/NRSF) binding motif[J]. *J Biol Chem*,2007,282(37):26717-26724
- [8] Rao JN, Neumann L, Wenzel S, et al. Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription[J]. *Biochem J*,2006,400(3):449-456
- [9] Harmar AJ, Hills RA, Rosser EM, et al. IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels [J]. *Nucleic Acids Res*,2009,37(Database issue):680-685
- [10] Tilford CA, Siemers NO: Gene set enrichment analysis[M]. *Methods Mol Biol*,2009,563:99-121
- [11] Wu CH, Apweiler R, Bairoch A, et al. The Universal Protein Resource (UniProt): an expanding universe of protein information[J]. *Nucleic Acids Res*,2006,34(Database issue):D187-191
- [12] Klein H, Vingron M. Using transcription factor binding site co-occurrence to predict regulatory regions[J]. *Genome Inform*,2007,18:109-118
- [13] Qian Z, Lu L, Liu X, et al. An approach to predict transcription factor DNA binding site specificity based upon gene and transcription factor functional categorization[J]. *Bioinformatics*,2007,23(18):2449-2454
- [14] McCall M, Brown T, Hunter WN, et al. The crystal structure of d (GGATGGGAG): an essential part of the binding site for transcription factor IIIA[J]. *Nature*,1986,322(6080):661-664
- [15] Rani TS, Bapi RS. Analysis of n-gram based promoter recognition methods and application to whole genome promoter prediction [J]. *In Silico Biol*,2009, 9(1-2):S1-16
- [16] Wang X, Xuan Z, Zhao X, et al. High-resolution human core-promoter prediction with CoreBoost_HM[J]. *Genome Res*,2009,19(2):266-275
- [17] Westholm JO, Xu F, Ronne H, et al. Genome-scale study of the importance of binding site context for transcription factor binding and gene regulation[J]. *BMC Bioinformatics*,2008,9:484
- [18] Berman BP, Nibu Y, Pfeiffer BD, et al. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome[J]. *Proc Natl Sci USA*,2002,199(2):757-762
- [9] Kottom TJ, Han J, Zhang Z, et al. Pneumocystis carinii expresses an active Rtt109 histone acetyltransferase[J]. *Am J Respir Cell Mol Biol*, 2011,44(6):768-776
- [10] Masumi A. Histone acetyltransferases as regulators of nonhistone proteins: the role of interferon regulatory factor acetylation on gene transcription[J]. *J Biomed Biotechnol*,2011,42(4)
- [11] Akiba Y, Cave JW, Akiba N, et al. Histone deacetylase inhibitors de-repress tyrosine hydroxylase expression in the olfactory bulb and rostral migratory stream [J]. *Biochem Biophys Res Commun*, 2010, 393(4):673-677
- [12] Carneiro K, Donnet C, Rejtar T, et al. Histone deacetylase activity is necessary for left-right patterning during vertebrate development[J]. *BMC Dev Biol*,2011,25(11):29
- [13] Morita K, Gotohda T, Arimochi H, et al. Histone deacetylase inhibitors promote neurosteroid-mediated cell differentiation and enhance serotonin-stimulated brain-derived neurotrophic factor gene expression in rat C6 glioma cells [J]. *J Neurosci Res*,2009,87(11): 2608-2614

(上接第 3644 页)