

PAX3 基因及其蛋白的生物信息学分析 *

张 敏 权 力 张 霆[△]

(北京大学首都儿科研究所教学医院 北京 100020)

摘要 目的 对 PAX3 基因和 PAX3 蛋白进行生物信息学分析,更多的了解该基因的相关信息,为进一步研究 PAX3 与神经管畸形的相关性研究提供基础。方法 运用生物信息学方法对 PAX3 基因的基因结构、单核苷酸多态性位点(SNP)、PAX3 基因与其他基因的相互作用网络、PAX3 蛋白结构域、蛋白二级结构、蛋白间相互作用网络、以及 PAX3 蛋白所调控和影响的靶基因进行分析。结果 PAX3 基因有 9 个可变剪切形式,编码区存在 14 个 SNP 位点,其中错意突变 13 个,移码突变 1 个。PAX3 蛋白由 479 个氨基酸组成,分子量 52968Da, PAX3 蛋白可能调控和影响 151 个靶基因的转录和表达,与 PAX3 基因存在相互作用的基因和与 PAX3 蛋白存在相互作用的蛋白多数与发育相关。结论 通过对 PAX3 基因和 PAX3 蛋白的生物信息学分析获得了其相应的分子生物学特征,为进一步研究提供基础。

关键词 PAX3 基因;生物信息学;转录因子;神经管畸形

中图分类号:Q75,Q78 文献标识码:A 文章编号:1673-6273(2011)01-36-05

Bioinformatic analysis of PAX3 gene and PAX3 protein*

ZHANG Min, QUAN Li, ZHANG Ting[△]

(Peking University Teaching Hospital-Capital Institute of Pediatrics, Beijing, 100020, China)

ABSTRACT Objective: To obtain more information about PAX3 gene and PAX3 protein. **Methods:** Bioinformatics approach were applied to analyze PAX3 gene structure, single nucleotide polymorphisms, target genes regulated by PAX3 protein, gene interaction network, protein secondary structure, protein interaction network, multiple sequence alignment of PAX3 gene family, multiple sequence alignment of PAX3 protein and its homologous sequences. **Results:** There are nine transcripts in PAX3 gene and 14 SNPs were found in coding region, including 13 missense and 1 frame shift. 151 target genes of PAX3 regulation factor were found and the interaction network between PAX3 and other genes were get. PAX3 protein has two domains, Homeobox domain and Paired box domain. During the 9 members of PAX3 gene family, PAX3 and PAX7 have the closest relationship. **Conclusion:** We obtained more biological information about PAX3 gene and PAX3 protein by bioinformatic analysis, which is very useful for further research.

Key words: PAX3 gene; Bioinformatic Analysis; Transcription Factor; Neural Tube Defects

Chinese Library Classification: Q75, Q78 **Document code:** A

Article ID: 1673-6273(2011)01-36-05

前言

Pax 基因是一个进化上保守的基因家族。普遍存在于从果蝇到人类的各种生物体中。迄今为止发现的人和小鼠的 Pax 基因各有 9 个(Pax1-Pax9)。人的 PAX3 基因位于 2 号染色体(2q36.1),分布于细胞核,是一类重要的转录调控因子,在胚胎发育过程中对组织和器官的分化起着重要的调控作用^[1,2]。PAX3 蛋白能够于体外结合某些 DNA 靶序列并激活报告基因的转录^[3],这一转录调控作用是 PAX3 基因功能的分子生物学基础。

PAX3 基因表达发生在神经管细胞开始分化之前,在整个神经管轴的侧翼板、神经嵴细胞及顶板均有表达^[4],PAX3 突变或表达失调会引起某些神经系统相关的出生缺陷和遗传综合征。例如人的神经管畸形(NTDs)^[5,6]、Waardenburg 综合征 1 型、2 型和 3 型^[7,8]。小鼠 pax3 突变会引起斑点型突变^[9]。关于 NTDs 的发生机制和预防措施是目前研究的热点,Pax3 表达水平降

低引起 NTDs 的具体的分子生物学机制国外研究较多,但尚未有统一的认识,目前的研究推测其可能的机制包括:PAX3 作为一种转录因子,其表达下调后所调控和影响的基因转录和表达的改变以及细胞凋亡、氧自由基的参与^[10]。

筛选 PAX3 转录因子的靶基因是了解 PAX3 蛋白功能的前提,目前只有少数靶基因被鉴定,本研究运用生物信息学方法,对 PAX3 基因和 PAX3 蛋白进行生物信息学分析,并对 PAX3 蛋白所调控和影响的基因做一筛选。从而为 PAX3 与 NTDs 的机制研究提供基础和线索。

1 材料与方法

从 NCBI 主页 GeneBank 数据库中查找人类 PAX3 基因,获得其基因组定位和基因结构信息,应用 GeneMania 数据库构建 PAX3 基因与其他相关基因的相互作用网络,从 NCBI SNP 数据库获得 PAX3 基因 SNPs 位点信息。以 GeneCard 数据库

* 基金项目 首都医学发展科研基金资助项目(2005-2048),国家自然科学基金资助项目(20905051)

作者简介 张敏(1983-),女,硕士,研究方向:代谢组学 E-mail zmin.zm@gmail.com

[△] 通讯作者 张霆 E-mail zhangtingcv@yahoo.com.cn

(收稿日期 2010-10-23 接受日期 2010-11-18)

获得的 PAX3 蛋白的氨基酸序列信息为基础, 分别应用 Prot-param 软件分析 PAX3 蛋白的理化性质 ,SOPMA 软件预测 PAX3 蛋白的二级结构, 应用 pfam 蛋白家族数据(<http://pfam.sanger.ac.uk/>)和 EBI 数据库(<http://www.ebi.ac.uk/>)进行 PAX3 蛋白的结构域分析, 然后应用 STRING 交互式数据库(<http://string.embl.de>)构建相关蛋白相互作用网络。PAX3 转录因子调控和影响的靶基因的检索通过 TRANSFACT 7.0 数据库获得, 最后应用 Vector 软件进行氨基酸序列多重比对和分子进化树绘制。

2 结果

2.1 基因结构、SNP 分析

通过对 NCBI 的 Gene 数据库、UCSC GenomeBrowser 和 GeneCards 的搜索, 获得以下基因基本信息 :PAX3 基因位于 2 号染色体(2q36.1) ,全长 99110bp,含有 8 个外显子 ,mRNA 全长 3112bp,CDS 在 mRNA 的 382bp-1593bp 区域。编码一个 479 个氨基酸且分子量为 52968Da 的蛋白质。该基因有 9 中可变剪切形式(表 1) 基因结构图见图 1。通过对 NCBI SNP 数据库的搜索 ,发现 PAX3 基因编码区共有 SNP 位点 14 个 ,其中错意突变 13 个 移码突变 1 个(表 2)。

表 1 PAX3 基因的 9 种可变剪切形式
Table 1 nine transcripts in PAX3 gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	CCDS
PAX3-001	ENST00000344493	3109	ENSP00000342092	403	CCDS2450
PAX3-002	ENST00000336840	1715	ENSP00000338767	407	CCDS2449
PAX3-003	ENST00000258387	959	ENSP00000258387	206	CCDS2451
PAX3-004	ENST00000392070	2258	ENSP00000375922	484	CCDS42825
PAX3-005	ENST00000409828	1254	ENSP00000386817	215	CCDS46523
PAX3-006	ENST00000350526	3610	ENSP00000343052	479	CCDS42826
PAX3-008	ENST00000392069	3170	ENSP00000375921	505	CCDS2448
PAX3-009	ENST00000409551	1782	ENSP00000386750	483	CCDS46522
PAX3-007	ENST00000464706	822	No protein product		

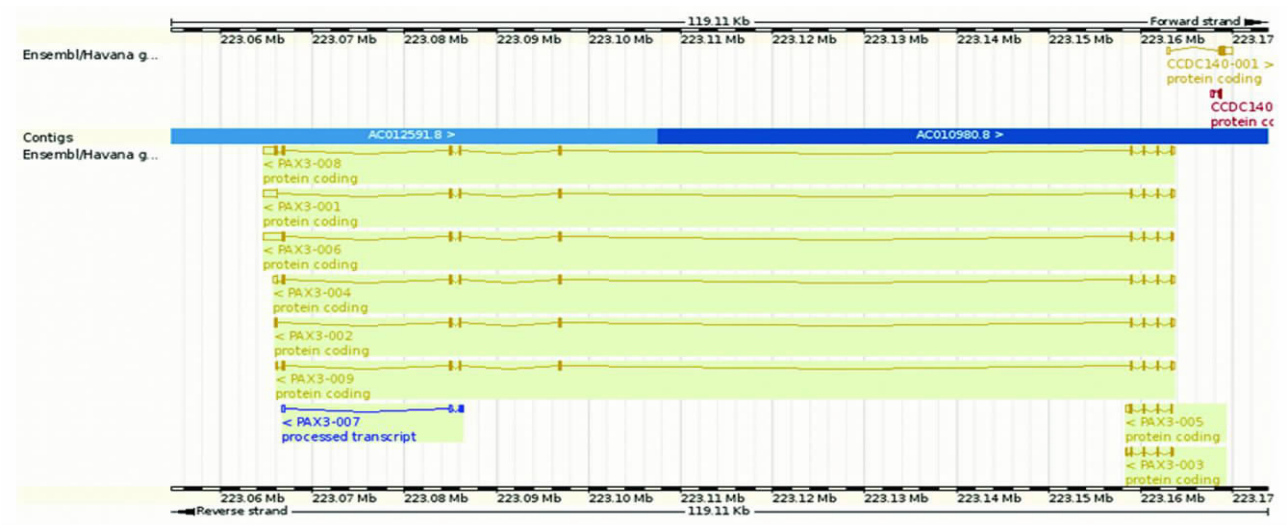


图 1 PAX3 基因结构

Fig. 1 PAX3 gene structure

表 2 PAX3 基因编码区错意突变、移码突变 SNP 位点

Table 2 The missense and frame shift SNPs of PAX3 gene in coding regions

dp SNP ID	Location	Base Changes	Aminoacid Changes
rs1042053	Exon 7	C1600T	L412F
rs1042051	Exon 7	C1598T	A411V
rs45607236	Exon 7	C1487T	S374L
rs2234675	Exon 6	C1310A	T315K
rs72960874	Exon 6	A1294G	T310A

rs80081387	Exon 4	C945A	S193R
rs116473352	Exon 3	G891C	K175N
rs104893654	Exon 2	T634C	Y90H
rs104893651	Exon 2	C617T	S84F
rs104893650	Exon 2	C515T	P50L
rs121908111	Exon 2	G524C	G48A
rs104893652	Exon 2	C507G	N47K
rs104893653	Exon 2	A505C	N47H
rs35487147	Exon 3	CG812A	P149H

2.2 基因相互作用

GeneMANIA 是基因相互作用和基因功能分析的网站 ,应用 GeneMANIA 搜索了同 PAX3 相互作用的其他基因的信息 ,以便更深入地了解 PAX3 以及相关基因的功能 ,所构建的基因相互作用网络如图 2。

2.3 PAX3 蛋白理化性质、二级结构预测

通过 ProtParam 预测 PAX3 蛋白的理论等电点为 8.92 ,半衰期 30 小时 ,分子式为 C₂₃₁H₃₆₄N₆₇O₇₂S₁₈。PredictProtein 是常用的蛋白二级结构预测软件 ,对于 α 螺旋预测精度较好 β 的精度差些 ,对于无规则卷曲的精度很差。因此本研究采用 SOPMA 软件预测 PAX3 蛋白二级结构 ,PAX3 蛋白氨基酸序列中 α 螺旋为 20.25% ,随机卷曲为 69.52% β 转角为 5.43% ,延伸链为 4.80%。

2.4 PAX3 蛋白结构域分析

PAX3 蛋白存在 3 个重要的结构域 ,同源异型类结构域 (Homeobox domain ,HD),成对结构域(Paired box domain ,PD) 还有位于二者之间保守的八肽结构域。HD 和 PD 结构域与 DNA 结合复合物的晶体结构图见图 3。

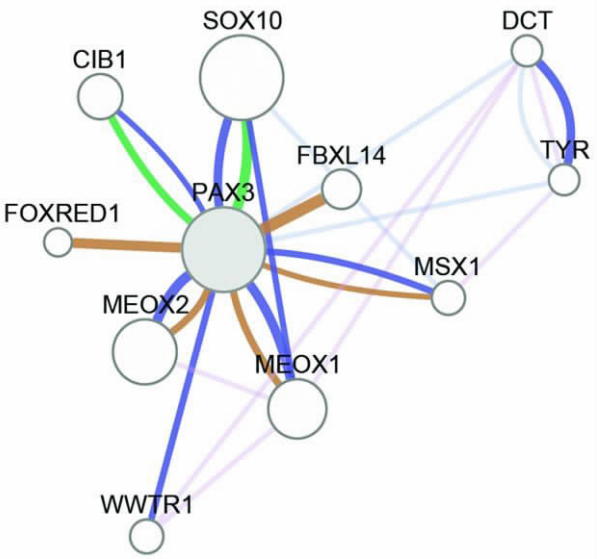


图 2 PAX3 基因与其他基因相互作用网络
Fig. 2 The interaction network of PAX3 gene with other genes

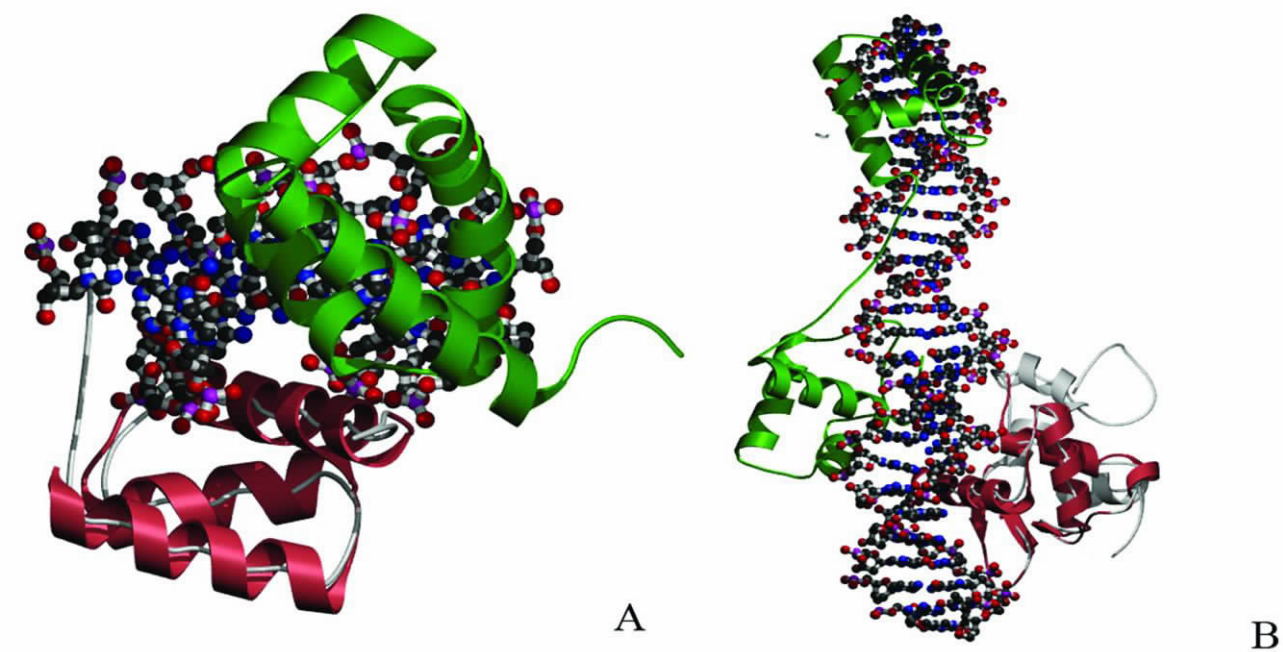


图 3 HD 和 PD 结构域与 DNA 结合复合物的晶体结构图
Fig. 3 Crystal structures for complex consist of Homeobox domain (A) and Paired box domain bounded to DNA (B).

2.5 蛋白质相互作用分析

利用 STRING 交互式数据库搜索与 PAX3 蛋白相互作用的其他蛋白质信息,以便更加深入的理解 PAX3 蛋白的功能和调控机制,搜索条件限制在 10 个蛋白内,所构建的蛋白相互作用网络如图 4 所示。

2.6 多序列比对分析

运用 Vector NTI Suite 软件分别对 PAX 基因家族基因的碱基序列、人类和其他 8 种物种(猩猩、鸡、小鼠、大鼠、鱼、果蝇、牛、犬)PAX3 蛋白的氨基酸序列进行多重序列比对,并绘制分子进化树(图 5),PAX3 基因家族中 PAX3 与 PAX7 的相似程度最高,有研究表明 PAX3 与 PAX7 可以形成异源二聚体,作用作用于二者共同的靶序列,激活报告基因的转录[3]。不同物种之间猩猩与人的进化距离最小,其次是牛、鸡、犬、大鼠、小鼠、进化距离最远的是鱼类和果蝇。

2.7 PAX3 转录因子调控和影响的基因

TRANSFAC 是收集转录水平上基因表达信息的数据库,通过 TRANSFAC 7.0 数据库 检索 PAX3 基因在基因组启动子

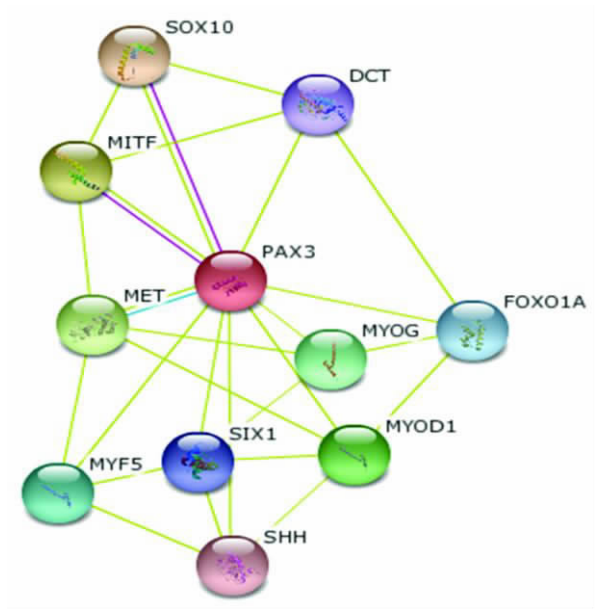


图 4 PAX3 蛋白与其他蛋白的相互作用网络
Fig. 4 The interaction network of PAX3 protien with other proteins

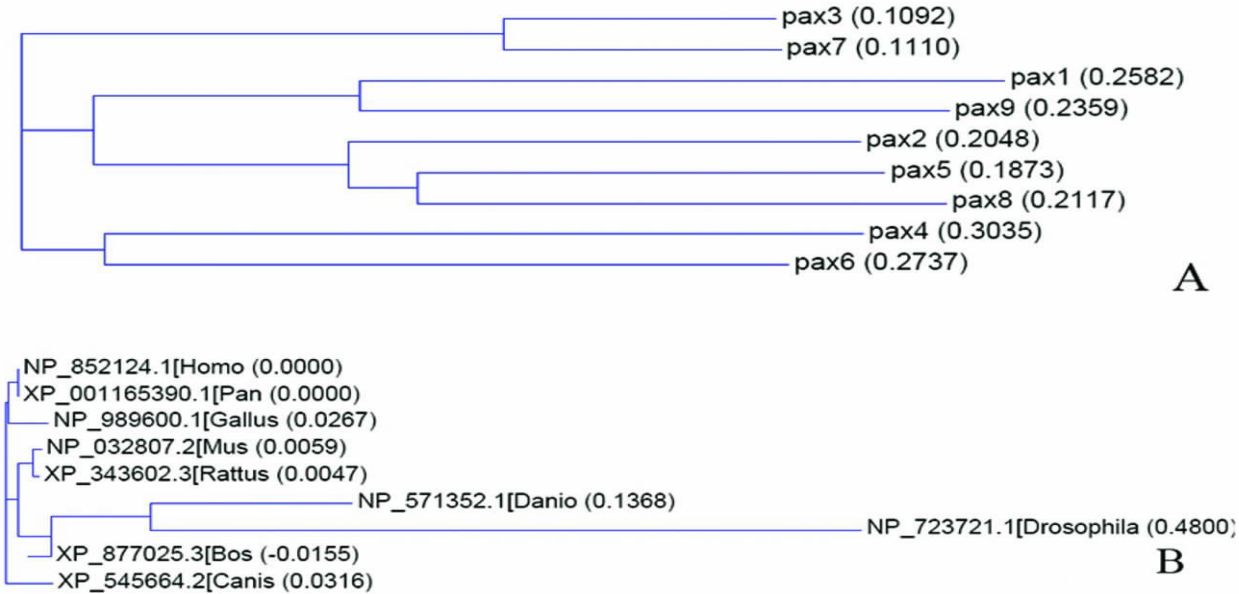


图 5 PAX 基因家族分子进化树(A)和 PAX3 蛋白与其同源序列的分子进化树(B)
Fig. 5 The phylogenetic tree of PAX gene family (A) and the phylogenetic analysis of PAX3 protein and its homologous sequences.

区可能存在的结合位点(结合模式:CGTSACG_V\$PAX3_B),从而找出 PAX3 转录因子结合位点下游潜在的靶基因,这些可能受到 PAX3 转录因子调控和影响的基因包括: HMGN1, RAD51C, RFT1, LTBP1, FLJ45187, PTPN23, CASK, FTHFS-DC1, CBX6, FLJ20308, EPC1, CUL5, ELOVL5, KIAA0082, DHX36, CRY1, RNF146, BHLHB2, PITX2, PDPR, MGC14376, BRAF, ANKS1, C11ORF30, RELB, MGC4399, POLB, RAD9A, ZNF6, GEM, MGC11308, HTATIP, MASK-BP3, FLJ33817, PIAS4, SMARCE1, SMARCA5, ZZZ3, YME1L1, CHGB, BIG1, PB1, KIAA0251, SLC38A2, MGC10471, PPT2, NS, C20ORF177, AF5Q31, NICE-3, FLJ12666, RAB7, ASK, PPP1R3D, SULT4A1, N-PAC, ZFH1B, TNRC15, HOXA9, NKX2-3, PDZK7, PHACTR3, SLC6A12, NR4A2, MGC29643,

NR4A1, PCNP, EN2, MRRF, KIAA0669, CDKN1A, EIF4E, PLK2, PNRC1, SUPT16H, HABP4, GK, TPARL, IKBKB, MAB21L1, HPS5, CDC14B, CTCF, MGC35521, PEX11G, ZZEF1, CANX, PTMA, HMOX2, AP2B1, PAK2, SLC25A3, ZNF288, MASTL, FLJ21019, EGR1, PPP2R1A, HSCARG, E-GR3, NOL4, EGR2, ADAM10, PEPD, EGR4, ZFY,STXBP1, ARP3BETA, CX36, PPM2C, RBBP8, GTF2H1, TAF11, HOXC10, MGC32124, IPO7, NAB2, GNAS, NRK, RASD1, IRX6, USP8, C21ORF63, MCFP, MINK, TSPYL5, EDEM1, NDUFB2, TRIB1, DKFZP434K1421, XBP1, BA108L7.2, BCL11B, BCL11A, FOXD3, KCNE4, EHD4, SOAT1, SBB154, ANKHD1, FLJ12750, WSB1, TEX14, PLEKHE1, GYG, SLC18A2, RBM18, NEUROD2, C14ORF127,WDR23, GGA1,

VPS26。

3 讨论

PAX3 的 PD 结构域是含三个 α 螺旋的 124 个氨基酸的高度保守结构域,最初被发现于 *prd* 基因中,几乎存在于所有 *pax3* 基因中。HD 结构域含有螺旋-折叠-螺旋结构(Helix-turn-helix, HTH motif) PAX3 转录因子可以结合在特异性的 DNA 区段上调节和影响下游基因的转录和表达,这一结构在转录因子与 DNA 结合时发挥作用。HD 和 PD 结构域均在 PAX3 蛋白与 DNA 的结合过程中发挥作用,HD 结构域中的螺旋-折叠-螺旋结构(Helix-turn-helix, HTH motif)中的第二个螺旋通过氢键和疏水作用结合 DNA,第一个螺旋在第二个螺旋与 DNA 结合后协助稳定结合复合体的结构。HD 和 PD 结构域通常具有协同效应,并且会改变 DNA 构象^[11]。

在 PAX3 基因的相互作用网络图中可见,PAX3 基因与 SOX10 基因之间存在相互作用,二者编码的蛋白也存在相互作用,SOX10 是一种对神经嵴和外周神经系统的发育至关重要的转录因子^[12],除 SOX10 基因外,MSX1(HOX7)、MEOX2、MEOX1 等基因均与发育和肿瘤形成相关^[13]。在与 PAX3 相互作用的蛋白中,MITF 和 FOXO1A 同为转录因子,MET 是肝细胞生长因子受体前体,三者均与细胞的分化和发育密切相关^[14],SHH 则与神经管发育相关^[16]。

与传统的基于实验室生物学研究相比,生物信息学方法可以用较低的成本和较快的时间获得比较可靠的结果。本研究对 PAX3 基因和 PAX3 蛋白进行生物信息学分析,并检索了 PAX3 蛋白所调控和影响的靶基因。从而为 PAX3 基因和 PAX3 蛋白功能研究以及 PAX3 与 NTDs 的机制研究提供重要依据。

参考文献(Reference)

- [1] Mansouri A. The role of Pax3 and Pax7 in development and cancer [J]. Crit Rev Oncog, 1998, 9(2): 141-149
- [2] Buckingham M,Relaix F. The Role of Pax Genes in the Development of Tissues and Organs: Pax3 and Pax7 Regulate Muscle Progenitor Cell Functions [J]. Annual Review of Cell and Developmental Biology, 2007, 23(1): 645-673
- [3] Schäfer B W, Czerny T, Bernasconi M, et al. Molecular cloning and

characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes [J]. Nucleic Acids Research, 1994, 22 (22): 4574-4582

- [4] Stoykova A,Gruss P. Roles of Pax-genes in developing and adult brain as suggested by expression patterns [J]. J. Neurosci., 1994, 14(3): 1395-1412
- [5] Lu W, Zhu H, Wen S, et al. Screening for novel PAX3 polymorphisms and risks of spina bifida [J]. Birth Defects Research Part A: Clinical and Molecular Teratology, 2007, 79(1): 45-49
- [6] Greene N D E, Stanier P, Copp A J. Genetics of human neural tube defects [J]. Human Molecular Genetics, 2009, 18(R2): R113-R129
- [7] Tassabehji M, Read A P, Newton V E, et al. Mutations in the PAX3 gene causing Waardenburg syndrome type 1 and type 2 [J]. Nat Genet, 1993, 3(1): 26-30
- [8] Hoth C F, Milunsky A, Lipsky N, et al. Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I) [J]. Am J Hum Genet, 1993, 52(3): 455-462
- [9] Epstein D J, Vekemans M, Gros P. *spotch* (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3 [J]. Cell, 1991, 67(4): 767-774
- [10] 毛东伟. Cx43、同型半胱氨酸及 Pax3 与神经管缺陷畸形 [J]. 国外医学妇幼保健分册, 2001, 12(1): 13-15
- [11] Schofield P. The riddle of the homeobox [J]. Trends in Neurosciences, 1987, 10(3): 6
- [12] Potterf S B, Mollaaghababa R, Hou L, et al. Analysis of SOX10 Function in Neural Crest-Derived Melanocyte Development: SOX10-Dependent Transcriptional Control of Dopachrome Tautomerase [J]. Developmental Biology, 2001, 237(2): 245-257
- [14] Qian C-N, Guo X, Cao B, et al. Met Protein Expression Level Correlates with Survival in Patients with Late-stage Nasopharyngeal Carcinoma [J]. Cancer Research, 2002, 62(2): 589-596
- [15] Bondurand N, Pingault V, Goerich D E, et al. Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome [J]. Human Molecular Genetics, 2000, 9(13): 1907-1917
- [16] Ye W, Shimamura K, Rubenstein J L R, et al. FGF and Shh Signals Control Dopaminergic and Serotonergic Cell Fate in the Anterior Neural Plate [J]. Cell, 1998, 93(5): 755-766