

学校编码: 10384  
学号: 21720061152241

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_  
UDC \_\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

$\beta$  核转运蛋白家族的进化、表达和调控模式  
分析

Evolutionary and Transcriptional Exploration of  
Karyopherin  $\beta$  superfamily proteins

全 宇

指导教师姓名: 纪志梁 教授

陶 涛 教授

专 业 名 称: 生物化学与分子生物学

论文提交日期: 2009 年 4 月

论文答辩时间: 2009 年 5 月

学位授予日期: 2009 年 月

答辩委员会主席: \_\_\_\_\_

评阅人: \_\_\_\_\_

2009 年 5 月

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为(纪志梁教授)课题(组)的研究成果,获得(纪志梁教授)课题(组)经费或实验室的资助,在(厦门大学生物信息学与药物开发)实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，  
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

# 目 录

1 前言 .....	1
1.1 $\beta$ 核转运蛋白的结构 .....	1
1.2 $\beta$ 核转运蛋白的功能 .....	2
1.3 本研究的目的是和意义 .....	4
2 实验步骤 .....	6
2.1 识别同源蛋白 .....	7
2.2 系统发育树的构建 .....	7
2.3 转录模式研究 .....	7
2.4 调控元件研究 .....	9
3 结果与讨论 .....	11
3.1 $\beta$ 核转运蛋白超家族的进化 .....	11
3.1.1 序列比对 .....	11
3.1.2 $\beta$ 核转运蛋白的祖先 .....	14
3.1.3 基因重复 .....	18
3.1.4 逆转录转座 .....	错误! 未定义书签。
3.1.5 小结 .....	22
3.2 $\beta$ 核转运蛋白超家族在细胞和组织中的表达 .....	22
3.3 $\beta$ 核转运蛋白超家族基因的调控机制 .....	26
4 结论 .....	33
参考文献 .....	35
致 谢 .....	39
附录一：攻读硕士研究生期间发表的文章 .....	40

# CONTENTS

<b>1 INTRODUCTION</b> .....	错误! 未定义书签。
<b>1.1 Structure of <math>\beta</math>-karyopherins</b> .....	1
<b>1.2 Function of <math>\beta</math>-karyopherins</b> .....	2
<b>1.3 Object and significance of this work</b> .....	4
<b>2 EXPERIMENTAL PROCEDURES</b> .....	6
<b>2.1 Identification of orthologous proteins</b> .....	错误! 未定义书签。
<b>2.2 Construction of phylogenetic trees</b> .....	错误! 未定义书签。
<b>2.3 Investigation of transcription patterns</b> .....	错误! 未定义书签。
<b>2.4 Investigation of regulatory elements</b> .....	错误! 未定义书签。
<b>3 RESULTS AND DISCUSSION</b> .....	错误! 未定义书签。
<b>3.1 Evolution of the karyopherin <math>\beta</math> superfamily</b> .....	错误! 未定义书签。
3.1.1 The sequence alignments.....	错误! 未定义书签。
3.1.2 The ancestors of $\beta$ -karyopherins .....	错误! 未定义书签。
3.1.3 Gene duplications.....	错误! 未定义书签。
3.1.4 Retroposition .....	错误! 未定义书签。
3.1.5 Summary.....	错误! 未定义书签。
<b>3.2 Expression of karyopherin <math>\beta</math> genes in cells and tissues</b> .....	错误! 未定义书签。
<b>3.3 Mechanism of regulation of karyopherin <math>\beta</math> superfamily genes</b> .....	错误! 未定义书签。
<b>4 CONCLUSIONS</b> .....	错误! 未定义书签。
REFERENCES .....	错误! 未定义书签。
ACKNOWLEDGEMENTS .....	39
Appendix 1: Publication during graduate study.....	错误! 未定义书签。

## 摘要

$\beta$  核转运蛋白家族是真核细胞中广泛存在的一类蛋白质, 在生物大分子核质转运这一细胞过程中扮演着关键角色。由于  $\beta$  核转运蛋白家族各成员间序列组成及结构上的差异, 其所携带的底物及工作机制也有所不同。目前国际上对该家族的研究尚浅, 对其家族的形成及功能演化也尚未明晰。在本研究中, 我们利用系统发育分析, 表达谱芯片分析, 转录因子调控位点预测等方法, 对酿酒酵母, 秀丽隐杆线虫, 黑腹果蝇, 斑马鱼, 爪蟾, 原鸡, 小鼠和人类基因组中  $\beta$  核转运蛋白家族成员开展了进化、表达和基因调控模式分析。在酵母的 14 个家族成员中, 8 个成员 (KAP95, KAP114, KAP120, KAP123, CSE1, MSN5, CRM1 和 LOS1) 直接与它们在人类中的同源蛋白对应, 2 个成员 (KAP122 和 SXM1) 在进化过程中丢失, 剩下的 4 个成员 (PSE1, NMD5, MTR10 和 KAP104) 在不同的机制下发生了基因重复。 $\beta$  核转运蛋白虽普遍存在于各物种中, 但在不同类型的组织和细胞中却不是均匀表达的。在酵母和小鼠中某些  $\beta$  核转运蛋白家族成员的基因表达水平同时受到细胞周期和发育水平的调控。进一步对该家族基因启动子结合元件的虚拟分析说明了 SP1, NRF-2, HEN-1, RREB-1 和 NF-Y 等转录因子调控着大部分  $\beta$  核转运蛋白成员基因的表达。这些发现揭示了  $\beta$  核转运蛋白家族进化过程中的功能分化是与该蛋白家族多样的基因表达和调控模式密切相关的。

**关键词:** 核转运蛋白; 基因重复; 逆转录转座

## Abstract

Karyopherin  $\beta$  superfamily is a group of proteins which widely spread in eukaryotic cells. They play a key role in mediating nucleocytoplasmic transport of macromolecules. Because of their diversity in sequence compositions and structure, they bind varied cargos and function in different ways. Currently karyopherin  $\beta$  superfamily is not well studied worldwide, and it's still unclear about the origin and function evolution of this family. In this study, we investigated the evolutionary and transcriptional patterns of these proteins in 8 genomes (*Saccharomyces Cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Xenopus tropicalis*, *Gallus gallus*, and *Mus musculus* and *Homo sapiens*) with the help of phylogenetic analysis, microarray expression profile analysis and prediction of transcription factor binding sites. Among 14  $\beta$ -karyopherins in *Saccharomyces cerevisiae*, 8 corresponded to their human orthologs directly without diversification (KAP95, KAP114, KAP120, KAP123, CSE1, MSN5, CRM1 and LOS1), 2 were lost (KAP122 and SXM1), and the remaining 4 proteins (PSE1, NMD5, MTR10 and KAP104) exhibited gene duplications by different mechanisms.  $\beta$ -Karyopherins were ubiquitously but nonuniformly expressed in distinct cells and tissues. In yeast and mice, the titer of some  $\beta$ -karyopherin transcripts appeared to be regulated both during the cell cycle and during development. Further virtual analysis of promoter binding elements suggested that transcription factors SP1, NRF-2, HEN-1, RREB-1 and NF-Y regulate expression of most  $\beta$ -karyopherin genes. These findings illuminated close relationship between functional evolution of  $\beta$ -karyopherins and their diversity in gene expression and regulation patterns.

**Key words:** Karyopherin; Gene Duplication; Retroposition

## 1 前言

在真核细胞中，蛋白质，RNA，核糖体的核质转运是至关重要的生理过程，参与细胞核质转运过程的主要载体蛋白是  $\beta$  核转运蛋白 ( $\beta$ -karyopherins) 家族，该蛋白家族能将大分子底物通过核孔复合体 (nuclear pore complex, NPC) 运输到核内或核外<sup>[1]</sup>。该蛋白家族成员的分子量大约在 95 到 145kDa 间。迄今为止从酵母到人类等真核生物中都有  $\beta$  核转运蛋白被人们发现。在低等真核生物酿酒酵母 (*Saccharomyces Cerevisiae*) 中 14 个  $\beta$  核转运蛋白成员起着运输相应货物进核以及出核的作用<sup>[2, 3]</sup>。然而在哺乳动物细胞中，有大约 20 种  $\beta$  核转运蛋白参与了这些过程<sup>[4]</sup>，各成员蛋白能转运一组特定类型的底物<sup>[5]</sup>。这个蛋白家族同时包括了负责输出和输入的受体，其中负责底物输入的受体蛋白比负责输出的数量更多。比如在酿酒酵母中，10 个成员蛋白负责将底物运输到细胞核内，3 个成员蛋白 (CRM1, LOS1 和 CSE1) 负责输出，1 个成员蛋白 (MSN5) 既能参与底物输入也参与输出过程<sup>[3]</sup>。

### 1.1 $\beta$ 核转运蛋白的结构

该家族蛋白的 N 末端都有一个被称为 importin- $\beta$  N-terminal (IBN\_N) 的结构域 (或者叫做 RanGTP 结合结构域)，中部靠近 N 端和 C 端分别有两个核孔蛋白结合区，C 末端还有一个 importin- $\alpha$  结合域 (图 1)。 $\beta$  核转运蛋白家族成员是由 18 个到 20 个连续的 HEAT (Huntingtin, Elongation factor 3, protein phosphatase 2A, TOR1) 重复基序组成，每一个 HEAT 重复基序都是由两段反平行的  $\alpha$  螺旋 (A 螺旋和 B 螺旋) 通过一段环链连接而成。这些 HEAT 重复基序能在不同的功能状态下呈现出不同的构象。柔性的 HEAT 重复基序通过一种诱导-适应机制 (induced-fit mechanism) 促进了其结合物的定位<sup>[6]</sup>。整体上该家族蛋白可以看作一段超螺旋，同时在超螺旋的内侧和外侧具有广阔的相互作用表面<sup>[7]</sup>。这些特征使得  $\beta$  核转运蛋白区别于其他转运底物到非核细胞器的转运蛋白。



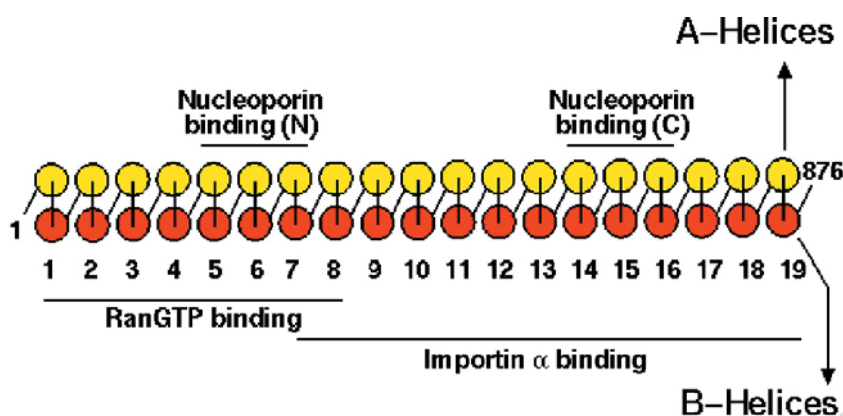


图 1 importin- $\beta$  结构域示意图(引自<sup>[8]</sup>)

Figure 1. Domain organization of importin- $\beta$ .

通过结构分析人们发现  $\beta$  核转运蛋白家族成员具有多个底物结合位点<sup>[9]</sup>。3 项独立研究分别分析了 Importin- $\beta 1$  与  $\alpha$  核转运蛋白片段<sup>[10]</sup>、转录因子 SREBP-2 片段<sup>[11]</sup>和 PTHrP (parathyroid hormone-related protein) <sup>[12]</sup>结合的共结晶结构, 这些研究显示了  $\beta$  核转运蛋白与这些底物间的结合部位是截然不同的。这说明每一类  $\beta$  核转运蛋白都可能含有多个底物结合位点, 这也能解释有限种类的  $\beta$  核转运蛋白家族成员是如何转运各种不同性质的底物。这些不同的结构也显示出  $\beta$  核转运蛋白能够根据不同的底物改变自身的整体构象<sup>[10-12]</sup>。另一项研究通过利用小角度 X 射线散射 (small-angle X-ray scattering, SAXS) 来显示多个  $\beta$  核转运蛋白家族成员的整体形状<sup>[13]</sup>, 这项研究表明在不同种类的核转运蛋白与不同底物的结合中存在显著的构象差异<sup>[13]</sup>。由此看来  $\beta$  核转运蛋白的整体超螺旋结构可能具有很强的柔韧性, 这种特性使得该家族成员能够识别各种类型的底物<sup>[4]</sup>。

## 1.2 $\beta$ 核转运蛋白的功能

分子量比核孔复合体的扩散限度 (diffusion limit,  $\sim 50$  kDa) 大的入核蛋白主要通过主动运输进入核内, 一般其序列中含有一段核定位信号 (nuclear localization signal, NLS) 被受体蛋白识别。经典的核定位信号是这样一小段碱性氨基酸 PKKKRKV<sup>[14]</sup>, 另一种典型的 NLS 则是一段由两部分组成的碱性氨基酸残基 KRPAATKKKAGQAKKKK, 它首次在核浆素 (nucleoplasmin) 序列中被科

学家发现<sup>[15]</sup>。同时，从核内被运送到核外的出核蛋白一般含有出核信号（nuclear export signal, NES）<sup>[16]</sup>。这些货物的运输是由核转运蛋白与鸟苷三磷酸酶（GTPase）Ran 蛋白协同介导的，其中 Ran 蛋白能够在它的 GDP 结合状态（胞质中）和 GTP 结合状态（核内）间循环变化。RanGTP 通过控制  $\beta$  核转运蛋白的构象和它们结合底物的能力使得货物的运输具有方向性。

如图 2 所示，底物的入核转运过程首先发生在细胞质中。含有 NLS 序列的底物通过接头蛋白 importin- $\alpha$  与 importin- $\beta$  结合，该复合物通过核孔复合体入核后，在 RanGTP 的作用下解离，底物被释放。随后 importin- $\beta$  与 RanGTP 的结合物又通过核孔复合体回到细胞质，而 importin- $\alpha$  则同时与 RanGTP 和它的核输出蛋白 CAS 结合后回到细胞质。最后，在 RanGAP（Ran GTPase-activating protein）的作用下，细胞质中的 RanGTP 发生水解，importin- $\beta$  被释放出来并进入了下一轮蛋白入核循环。核内蛋白的出核是一个类似的细胞过程：在细胞核内，核输出蛋白 Crm1 同时与 RanGTP 和含有 NES 序列的底物结合后通过核孔复合体进入细胞质，随后 RanGTP 的水解使得该复合体解聚，底物随之释放到细胞质，Crm1 也开始进入下一轮入核循环<sup>[6]</sup>。

除了核质转运， $\beta$  核转运蛋白还被人们发现参与了其他细胞过程。 $\beta$  核转运蛋白能够屏蔽碱性分子的电荷，还能调控有丝分裂中纺锤体的装配，中心体的运动，以及细胞核核膜、核孔的形成<sup>[7]</sup>。迄今为止，在哺乳动物中还没有 importin 基因敲除的研究报告，然而在果蝇<sup>[17]</sup>和秀丽隐杆线虫<sup>[18]</sup>中对 importin 基因的突变却导致了胚胎致死现象和异常的卵子发生。这些研究结果说明了  $\beta$  核转运蛋白在个体胚胎发育和生殖过程中起着重要作用。

尽管大部分核转运蛋白在细胞中都有表达，但仍有证据表明至少某些核转运蛋白的表达是受发育水平调控的<sup>[19]</sup>，这可能是由于运输条件的不同。例如，啮齿动物 Kpnb1 和 Ranbp5 的表达受精子发生水平的调控<sup>[20]</sup>。大鼠和人类 IPO13 的表达同时受激素和发育水平调控<sup>[21, 22]</sup>。



许多新近发现的核转运蛋白家族成员。文中作者通过系统发育分析也认为 KAP95、KAP121 和 KAP104 是最早出现的核转运蛋白，原始的 KAP95 基因的重复可能发生在完整的核孔复合体出现之前<sup>[24]</sup>。可是由于该研究涉及的对象面比较广，它并没有详细地解释  $\beta$  核转运蛋白家族起源与进化的过程。

在这样的研究背景下，我们还不能确定酵母的 14 个  $\beta$  核转运蛋白成员是否直接导致了人类 20 个该家族成员的产生。 $\beta$  核转运蛋白基因在不同细胞类型中的表达是否反映了相应的功能需求，也还是一个未解之谜。此外，研究  $\beta$  核转运蛋白家族起源和进化的模式，对于我们了解真核生物的起源也有重要意义。为了探索这些问题，我们选取了包括酵母和人类在内的 8 种真核模式生物作为研究对象，使用多种工具和方法搜寻各物种基因组中可能的  $\beta$  核转运蛋白家族基因。通过对这些基因的系统发育分析和对基因结构的比较分析，我们基本确定了  $\beta$  核转运蛋白家族各成员的进化途径。该蛋白家族主要通过基因复制进行扩张，其中 RANBP6 的逆转录转座过程更引起了我们的关注。此外我们还通过对芯片表达图谱和转录因子结合位点的分析系统研究了每个  $\beta$  核转运蛋白家族成员的表达模式和转录调控规律。我们发现该家族基因的表达水平与细胞周期和发育阶段密切相关，SP1, NRF-2, HEN-1, RREB-1 和 Nuclear Factor Y 等转录因子可能对该家族基因的表达起着关键性的调控作用。这些研究结果加强了我们对  $\beta$  核转运蛋白家族进化过程的认识，并为  $\beta$  核转运蛋白家族功能分化和细胞核质转运过程的研究提供了重要的理论依据。

## 2 实验步骤

我们计划从 3 个方面入手：(1)进化分析，(2)表达分析，(3)调控分析。首先，为了理清  $\beta$  核转运蛋白家族各基因在进化中的位置和搞清楚该家族的进化过程，我们将利用基因和蛋白序列进行系统发育分析。由于同源基因往往具有相似的外显子与内含子的排布，所以除了序列信息，搜集并比较  $\beta$  核转运蛋白家族成员的基因结构也是必要的。为了进一步理解  $\beta$  核转运蛋白家族在进化过程中的功能变化，我们需要了解该家族基因在表达方面的特点和规律，所以我们计划通过分析基因表达图谱芯片数据来获得该家族成员的基因表达模式。此外，基因调控模式的分析对于该家族的功能研究也是很有价值的。为此我们利用 MAPPER 搜索引擎预测该家族成员基因可能的转录因子结合位点，希望能发现该家族基因调控的特点和规律。

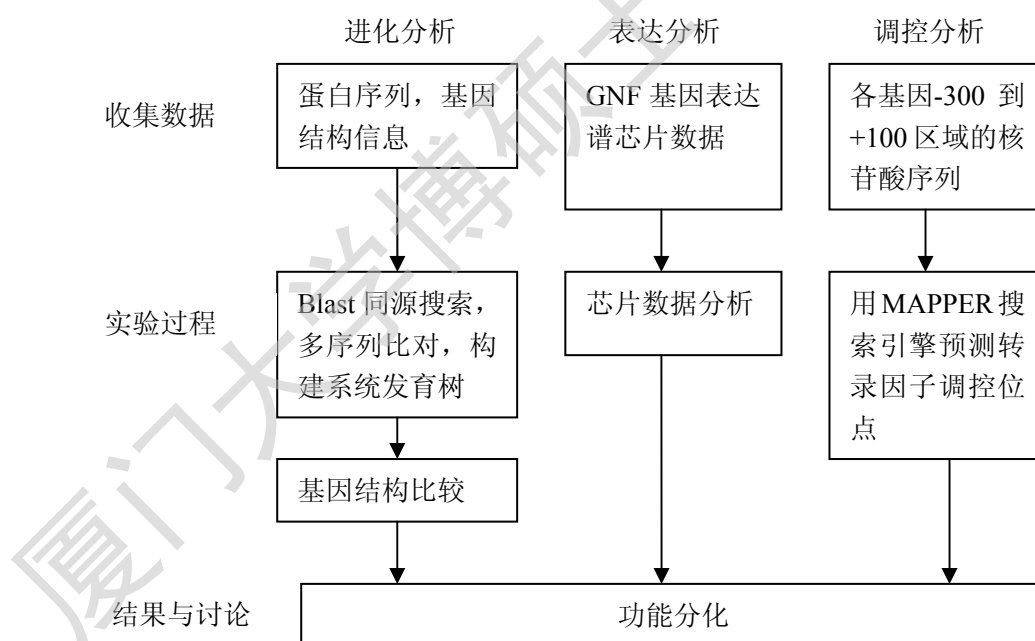


图 3: 实验流程概览

Fig 3. Summary of the research process.

## 2.1 识别同源蛋白

同源蛋白的识别与指定对于  $\beta$  核转运蛋白家族的进化分析是至关重要的。为了这个目标，我们在 8 个物种中搜寻和比较同源蛋白：人类 (*Homo sapiens*)，小鼠 (*Mus musculus*)，原鸡 (*Gallus gallus*)，斑马鱼 (*Danio rerio*)，爪蟾 (*Xenopus tropicalis*)，黑腹果蝇 (*Drosophila melanogaster*)，秀丽隐杆线虫 (*Caenorhabditis elegans*) 和 酿酒酵母 (*Saccharomyces cerevisiae*)。每一组同源蛋白的指定都经过了多方面证据的证实：1. 作为最基础的方法，我使用 BLAST 搜索全基因组数据库来识别并验证可能的同源蛋白，搜索中采用的是美国国立生物技术信息中心 (NCBI) 提供的 blastp 序列比对算法和缺省参数。2. 大部分  $\beta$  核转运蛋白的同源关系能通过搜索 NCBI 蛋白质直系同源簇 (COG) 数据库<sup>[25]</sup> 获得。3. Ensembl 数据库<sup>[26]</sup> 的“同源预测”功能也能帮助我们在全基因组层面识别  $\beta$  核转运蛋白家族可能的同源蛋白。

## 2.2 系统发育树的构建

首先从 Ensembl 基因组数据库 46 版和 GenBank<sup>TM</sup> 数据库获取 8 个物种  $\beta$  核转运蛋白家族的序列和外显子信息。为了说明  $\beta$  核转运蛋白家族从酵母到人类的进化过程，需要构建系统发育树。首先用 MUSCLE 3.6<sup>[27]</sup> 进行蛋白质序列的多序列比对，然后参照蛋白的二级结构手工改进前面的序列比对结果，最后用 ClustalX 1.83<sup>[28]</sup> 将序列比对结果图形化。接下来基于多序列比对的结果，我们利用软件包 MEGA4<sup>[29]</sup> 提供的邻接法 (neighbor joining) 去构建系统发育树。其中氨基酸替代模型我们选择了泊松校正模型 (Poisson Correction Model)，位点进化速率选择均匀进化速率 (uniform rates)，同时我们使用 1000 次重复的自展分析 (bootstrap analysis) 来检验这些系统发育树的可信度。

## 2.3 转录模式研究

为了研究  $\beta$  核转运蛋白基因的转录模式，我们需要了解该家族各成员基因

在不同组织和不同发育阶段的表达水平。传统的基因表达检测技术主要有 Northern Blot 和实时定量 RT-PCR 等，但由于这些方法步骤繁琐，仅适用于少量基因的表达分析。为了高效地获得大量表达数据，我们采用了近年来发展比较迅速的表达谱芯片技术。

表达谱 DNA 芯片是指将大量的 DNA 片段或寡核苷酸固定在以玻璃、硅、塑料等硬质载体上制备成基因芯片。待测样品中的 mRNA 被提取后，通过逆转录获得 cDNA，并在此过程中标记荧光，然后与包含上千个基因的 DNA 芯片进行杂交反应 30min~20h 后，将玻片上未发生结合的片段洗去，再对玻片进行激光共聚焦扫描，测定芯片上各点的荧光强度，从而推算出待测样品中各种基因的表达水平。若要比对不同的两个细胞系或不同组织来源的细胞的基因表达差异，则从不同的两个细胞系或不同来源的组织细胞中提取 mRNA，比较两种荧光在各点阵上的强度，并推算出各基因在不同环境中的相对表达水平。

本项目所用的基因表达芯片数据集 GNF1H\_GCRMA 和 GNF1M\_GCRMA 是从 GNF SymAtlas 数据库<sup>[30]</sup>获得，这两套芯片数据集分别涵盖人类 79 种组织中的 24,277 个基因和小鼠 61 种组织（包括 7 个发育表达状态）中的 32,905 个基因。

我们对  $\beta$  核转运蛋白家族的基因表达模式进行了定量分析。几何比较（相似性测度，similarity measure, SM）是用来说明两个基因的表达谱的相似程度。接近 1 的相似性测度值说明了这 2 个基因表达模式是高度相似的，尽管不同基因的绝对表达水平可能差异很大。这样的基因可能具有相关的生物学作用。如果用向量  $\mathbf{X}$  来表示一个基因的表达图谱， $x_i$  表示该基因在  $i$  组织或时序下的表达水平， $n$  是组织或时序的数量，也就是向量  $\mathbf{X}$  的维数，那么

$$\mathbf{X} = (x_1, x_2, x_3, \dots, x_n), \quad (1)$$

如果  $|\mathbf{X}|$  表示向量  $\mathbf{X}$  的长度， $\theta$  是向量  $\mathbf{X}$  与  $\mathbf{Y}$  在多维空间的夹角，此时基因之间的相似性测度 SM 就可以用下面的公式表示

$$SM(\cos \theta) = \frac{\mathbf{X} \cdot \mathbf{Y}}{|\mathbf{X}||\mathbf{Y}|}, \quad (2)$$

特异性测度（Specificity Measure, SPM）是用来估计某一特定基因在特定组织中的特异性程度与表达丰度，计算公式如下

$$SPM(\cos \alpha) = \frac{x_i}{|\mathbf{X}|}, \quad (3)$$

其中  $\alpha$  是多维空间中该样本向量与坐标轴的夹角<sup>[31]</sup>。特异性测度 SPM 是我们识别组织特异性表达的基因的重要标准，它对于我们认识基因的生理学意义有很大帮助。SPM 取值范围是从 0 到 1，越接近 1 说明该基因表达的组织特异性越强，即该基因集中于在该组织中表达，而在其他组织中不表达或表达水平很低。此外我们还采用了 3 组不同的酿酒酵母芯片数据集（用  $\alpha$  因子处理，cdc15 和 cdc28 突变型）来研究  $\beta$  核转运蛋白家族可能的与细胞周期相关的表达模式。

## 2.4 调控元件研究

转录调控模式研究也是很重要的工作，它对于我们理解生物发育，细胞定型与分化的机制，以及外部和内部的信号是如何转化成为特定的基因表达模式起着关键作用。为了研究  $\beta$  核转运蛋白家族基因的调控模式，我们决定从预测基因上游调控区域的转录因子结合位点（transcription factor binding sites, TFBSs）着手，并选择使用 MAPPER 作为工具分析  $\beta$  核转运蛋白家族基因潜在的调控位点与调控模式。

MAPPER 是多物种转录因子结合位点的计算化识别平台。这个平台应用了创新的技术，它结合了 TRANSFAC 和 JASPAR 数据库的海量数据和剖面隐马尔可夫模型（profile hidden Markov model）的强大搜索能力<sup>[32]</sup>。TRANSFAC 和 JASPAR 数据库的转录因子结合位点序列都是通过实验手段确定的。基于这些序列数据，MAPPER 为大量转录因子建立了对应的剖面隐马尔可夫模型。这些模型被用来开发一种搜索引擎，该引擎将用于检索已知的或全新的序列中可能的转录因子结合位点。MAPPER 搜索引擎能够识别和图像化基因不同区域中（启动子区域或其他区域）可能出现的转录因子结合位点。目前该引擎已经内置了人类，小鼠，黑腹果蝇，秀丽隐杆线虫和酿酒酵母等 5 个物种的基因信息，通过网页可以很方便地搜索到用户感兴趣的基因。此外，它还允许用户查询或建立一个新的 HMM 模型，这需要用户自己提供一个转录因子结合位点的多序列比对。

为了进行转录因子结合位点的分析，我们从 Ensembl 数据库和 GenBank 数



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库