PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2015, 42(4): 375~389
www.pibb.ac.cn

基于节点-模块置信度及局部模块度双重约束挖掘前列腺癌候选疾病模块 *

王一斌　程咏梅　张绍武 **
(西北工业大学自动化学院，信息融合技术教育部重点实验室，西安 710072)

摘要　前列腺癌病因及发病机理研究有助于前列腺癌预防和治疗．目前，前列腺癌生化试验研究方法成本高、耗时，而基于网络计算方法容易受基因表达谱数据不完整、噪声高及实验样本数量少等约束．为此，本文提出一种基于节点 - 模块置信度及局部模块度的双重约束算法(命名为 NMCOM)，挖掘前列腺癌候选疾病模块．NMCOM 算法不依赖基因表达谱数据，采用候选基因与致病表型之间一致性得分，候选基因与致病基因之间语义相似性得分融合排序策略，选取起始节点，并基于节点 - 模块置信度及局部模块度双重约束挖掘前列腺癌候选疾病模块．通过对挖掘出的模块进行富集分析，最终得到 18 个有显著意义的候选疾病基因模块．与单一打分排序方法及随机游走重开始方法相比，NMCOM 融合排序策略的平均排名比小、AUC 值大，且挖掘出结果明显优于其他模块挖掘算法，模块生物学意义显著．NMCOM 算法不仅能准确有效地挖掘前列腺癌候选疾病模块，且可扩展挖掘其他疾病候选模块．

关键词　前列腺癌，疾病模块挖掘，候选基因排序，节点 - 模块置信度，局部模块度
学科分类号　R318.04，Q78　　**DOI**: 10.16476/j.pibb.2014.0091

前列腺癌在许多西方国家是男性最常见的恶性肿瘤，占男性癌症死因的第二位[1]，仅 2012 年美国就有超过 250 万的男性被诊断出患有前列腺癌[2]．在我国，前列腺癌在男性泌尿、生殖系统恶性肿瘤中发病率已跃居第三位，成为一个不容忽视的问题[3]．近年来，前列腺癌的医疗诊断和治疗技术有了新的发展和成就，例如超声诊断、医学影像诊断和较新的冷冻疗法、荷尔蒙疗法等，但关于前列腺癌的病因和发病机制至今尚未明确．Beillard 等[4]于 2013 年通过试验发现蛋白质 GNA13 对前列腺癌细胞的入侵和转移起着至关重要的作用．随后 Baena 等[5]利用临床小鼠试验和综合全基因组方法，分析了蛋白质 ERG 和 ETV1 在前列腺癌中的功能特异性，发现它们除调控一些常见的基因外，还涉及到类固醇的合成和某些新陈代谢，这些都对前列腺癌的发展起着关键作用．此外，Ding 等[6]利用荟萃分析研究了 CYP1A1 MspI 多态性与前列腺癌的风险关联性，结果表明其多态性对前列腺癌发展有较大风险，尤其是对亚洲男性．这些研究成果虽然极有价值，但通常以花费大量的时间和高昂的人力物力为代价．

当前，不仅可利用的大量生物网络数据不断涌现，而且很多相关的研究工作也已展开，如生物网络分析、模块挖掘、网络比对等．因而，借助生物网络，通过计算方法来预测致病基因或挖掘疾病模块已迅速成为从分子生物学网络层面研究人类复杂疾病的热点．这些计算方法根据所采用的计算模型和机理，可以将其划分为基于传统图理论的检测方法和基于非传统图理论的检测方法两大类．基于传统图理论分析的方法依据检测结果所体现出的特征又可细分为如下三类方法：a．基于密度的聚类方法，其基本假设是网络中的子图通常对应着蛋白质的功能模块，因此搜索浓密连接的子图来发现功能

* 国家自然科学基金资助项目(91430111, 61473232, 61170134)，西南财经大学“中央高校基本科研业务费专项资金 - 青年教师成长项目”(JBK150134).
** 通讯联系人.
Tel: 029-88431308, E-mail: zhangsw@nwpu.edu.cn
收稿日期：2014-12-14，接受日期：2015-03-17

模块．代表方法有极大团方法(maximize cliques)[7]、基于蛋白质的连接值的计算方法(molecular complex detection，MCODE)[8]等．b．基于层次的聚类方法，其基本思想是由于PPI(蛋白质相互作用，protein-protein interaction)网络在本质上具有模块化的层次特性，所以基于层次聚类的方法很自然地适用于蛋白质功能模块的检测．代表方法有中心结构法[9]、蛋白质间距离法[10]等．c．基于划分的聚类方法，其目的是寻找一个最优的网络划分，使得划分到同一簇的对象之间尽可能“接近”或相似，而不同簇中的对象间尽可能“远离”或不同，并且能够对所有稀疏连接的节点做出合理的解释．代表方法有限制邻近搜索聚类方法(restricted neighborhood search clustering，RNSC)[11]、边 - 中介数方法(edge-betweenness)[12]、亲缘传播方法(affinity propagation，AP)[13]等．除利用传统图理论的检测方法外，近年来也涌现出许多将其他机理融合于图聚类过程的检测方法，主要有以下四类：a．基于流模拟的聚类方法，它是通过分析PPI网络中每个蛋白质对其他蛋白质施加的生物学或拓扑影响的程度来实现功能模块的检测．马尔科夫聚类(markov clustering，MCL)就是PPI网络中流的自由行走来检测蛋白质复合物的经典算法[14]．b．基于谱分析的聚类方法，它以谱图为理论基础，采用矩阵分析技术将待求解问题转化为带约束的二次型优化问题．代表方法有谱密度算法[15]、扩散模型方法(adjustable diffusion matrix-based spectral clustering，ADMSC)[16]等．c．基于核心 - 附属关系的聚类方法，该方法认为位于“核心”内的蛋白质具有高度的共表达特性及高度的功能相似性，“附属”蛋白质位于功能模块的“核心”外围，辅助其实现相关的生物功能．代表方法有Core算法[17]、最大基团算法[18]．d．基于群集智能的聚类方法，这一类基于简单个体相互作用时涌现的整体智能行为而提出的元启发搜索方法，并在PPI网络的功能模块检测中开始了一些新的探索．例如基于蚁群最优化的NACO-FDM方法[19]、基于连接强度的蚁群优化聚类算法[20]等．此外，由于人类疾病数目众多，且不同疾病的需求和研究进展参差不齐，因此还有众多以具体疾病为研究对象，利用不同的数据集，从不同角度构建疾病数据库、预测致病基因、挖掘疾病模块或路径的方法．如针对癌症中最常见、综合死亡率最高的肺癌，文献[21−23]构建了不同功能和对象的肺癌数据库，以方便相关研究人员使用．文献[24]将微阵列基因表达谱和PPI网络相结合，利用线性回归模型，构建出生物网络标记，并配合其所提出的致癌关联数值，找出40个重要的与肺癌相关的蛋白质．文献[25]则将人 - 鼠同源相互作用与肺癌相关的哺乳动物表型相结合，并用自然语言的处理方法挖掘出70余个非小细胞肺癌相关的基因．文献[26]对非小细胞肺癌运用前馈环路，并整合miRNA-TF (transcription factor)共调控网络，挖掘其疾病模块，阐明相关的再生调控路径．文献[27]应用rank-based的方法构建基因共表达网络，进而基于肺癌差异表达基因及基因模块功能一致性的联合测度，最终筛选出疾病相关功能模块．目前大多数前列腺癌疾病研究大都利用筛选出差异表达基因或生物标记，将它们映射到PPI网络中构建相关子网络的方法，或者利用基因表达谱数据对PPI网络中的相互作用赋予权重，按照一定的条件挖掘子网络[28−31]．这些方法虽然对该疾病病理和临床诊断提供了一些信息和帮助，但主要依赖基因表达谱数据，而目前基因表达谱数据存在样本数量有限、数据有缺失且包含噪声等问题，同时没有整合诸如基因本体(gene ontology，GO)注释等信息，导致预测结果不完整，预测精度较低．此外，这些方法将辨识出的结果直接利用PPI网络中的相互作用关系来组建子网络，没有进行进一步的筛选和挖掘，导致结果的统计显著性和生物学意义不明显．

鉴于此，我们提出一种新的基于节点 - 模块置信度和局部模块度的算法(node-module confidence and local modularity，NMCOM)，挖掘前列腺癌风险致病基因模块．NMCOM算法首先选用PPI网络、疾病 - 表型网络和GO数据库，借助其拓扑性质和语义相似性对所选的候选基因进行融合打分，并按得分高低排序，避免了基因表达谱数据的局限性；然后依次从得分最高的候选基因开始，构建初始模块，并按节点 - 初始模块的置信度以及局部模块度的变化率判断其他节点是否加入初始模块；最后在模块扩张结束后由模块富集分析来决定该模块最终是否为候选致病基因模块．模块挖掘中对加入的节点进行了双重筛选并以富集分析为依据对模块进行取舍，因此最终能得到具有较好生物学意义和显著性的模块挖掘结果．

1 材料与方法

1.1 数据集

本文采用前列腺癌致病基因、人类蛋白质相互

作用、表型相似性及 GO(http://www.geneontology.org)四个数据集．致病基因数据集由两部分组成，一部分来源于 PGB(human prostate gene dataBase)数据库，另一部分从 OMIM Gene Map 数据库筛选得到．人类蛋白质相互作用数据集来源于 HPRD 数据库(版本 9)，去除蛋白质自身相互作用，最终得到一个包含 9 502 个蛋白质，37 520 个相互作用的网络．表型相似性数据来自 Van Driel 等研究结果[32]，以矩阵形式给出，一共包含了 5 080 种 OMIM 表型以及表型之间的相似性得分．表型以 MIM 编号形式表示，每个致病基因都有对应的表型 MIM 编号.

利用已知致病基因集和表型相似性得分数据，开始选取候选基因集．首先将一个致病基因相对应的表型 MIM 编号映射到表型相似性得分矩阵中，并将该表型与其他所有表型的相似性得分平均值作为一个阈值．凡相似性得分大于此阈值的其他表型都将作为候选表型，然后在 OMIM 中找到候选表型所对应的基因．对每一个候选表型而言，它所对应的基因可能是一个也可能是多个．最后把候选表型所涉及的每一个基因进行连锁间隔查询以构建候选基因，其中连锁间隔区域的选取长度为 10 个连锁基因．对每一个前列腺癌致病基因进行同样的处理过程，构成候选基因集.

1.2　NMCOM 算法

由于局部信息模块挖掘算法具有计算复杂度较低、空间和时间消耗较小、扩展性较高等优点，我们根据局部信息搜索挖掘和 “节点 - 扩充”的思想，提出 NMCOM 模块挖掘算法．NMCOM 方法分别将候选基因与致病表型之间的一致性和候选基因与致病基因之间的语义相似性进行打分，融合两个得分；然后从得分最高节点开始，根据一定条件选择其他节点构建初始模块，按照节点 - 模块置信度的变化率和局部模块度变化率这两个双重约束条件，判断是否允许新节点加入初始模块，进而挖掘出满足条件的模块；最后对模块的显著性进行富集分析，以最终决定该模块是否属于疾病候选模块．如模块确定为疾病候选模块，则将该模块所有节点从 PPI 网络中剔除，反之则在 PPI 网络中保留．选择下一个候选起始节点继续进行同样的挖掘工作，当连续 3 个模块都判定为不属于疾病候选模块时，整个模块挖掘工作结束.

NMCOM 算法流程图如图 1 所示：

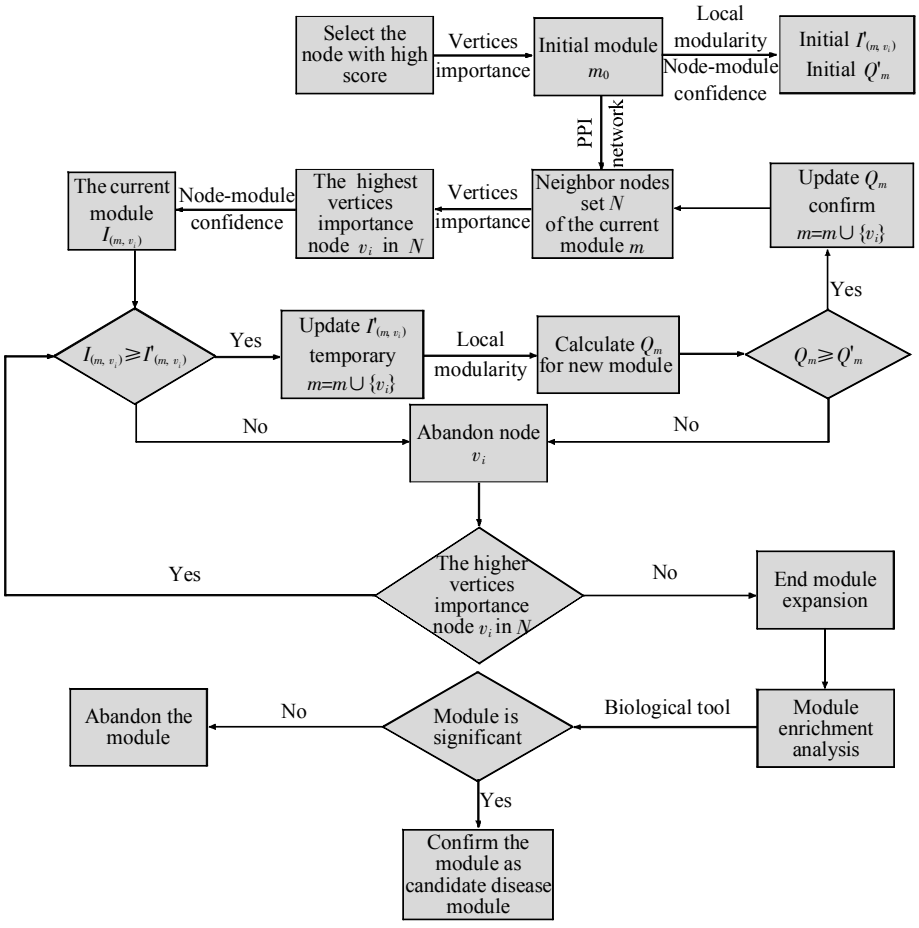

Fig. 1　The flow diagram of NMCOM algorithm

1.2.1 候选基因打分排序

基于“节点 - 扩充”的局部信息搜索挖掘方法需要从一个初始解(某个节点)开始，每一步在当前的邻域内找到一个最优解，使目标函数满足一定的条件或使其逐步优化，直到不能进一步扩充为止. 其中初始解选取对产生模块的质量具有较大影响，因此 NMCOM 算法初始解选取不使用基因表达谱数据，而采用其他数据集，同时为避免使用单一数据集而导致的片面性问题，采取了多种数据集相融合的策略：对某个候选基因 g，它与前列腺癌表型之间的一致性得分为 CS_{pg}，与前列腺癌致病基因集之间的语义相似性得分为 $P_{\text{semantic}}(g)$，融合 CS_{pg} 和 $P_{\text{semantic}}(g)$，其综合得分为 $S(g)$：

$$S(g)=r\times CS_{pg}+(1-r)\times P_{\text{semantic}}(g) \tag{1}$$

式中：r 为权重系数，$r\in(0, 1)$.

根据综合得分 $S(g)$值大小对候选基因排序，依次选取排名靠前的候选基因为模块挖掘起始节点.

a. 基因与表型间一致性计算

依据疾病表型所对应的基因在 PPI 网络中的连接紧密程度，Wu 等[33]提出如下疾病表型相似性计算模型：

$$S_{pp'}=C_p+\sum_{g\in G(p)}\sum_{g'\in G(p)}\beta_{pg}\mathrm{e}^{-L_{gg'}^2} \tag{2}$$

式中：$S_{pp'}$表示两个表型 p 和 p' 之间的相似性得分；$L_{gg'}$表示基因 g 和 g' 在 PPI 网络中的最短路径；$G(p)$表示表型 p 对应的所有致病基因；C_p 为常数；β_{pg} 为权重系数，代表基因 g 对表型 p 的影响程度，从实际角度考虑，设该系数独立于表型 p' 和 g' 基因.

公式(2)建立的模型中，还可用公式(3)来计算某个基因与某个表型所对应的所有基因之间的紧密性之和：

$$\Phi_{gp'}=\sum_{g'\in G(p')}\mathrm{e}^{-L_{gg'}^2} \tag{3}$$

因此公式(2)又可变为：

$$S_{pp'}=C_p+\sum_{g\in G(p)}\beta_{pg}\Phi_{gp'} \tag{4}$$

根据公式(3)和(4)，前列腺癌疾病表型 p 和其他 n 个疾病表型之间的相似性，候选基因 g 和其他 n 个疾病表型之间的紧密性可分别用向量表示，即 $S_p=(S_{pp_1}, S_{pp_2}, \cdots, S_{pp_n})$和 $\Phi_g=(\Phi_{gp_1}, \Phi_{gp_2}, \cdots, \Phi_{gp_n})$. 最后利用皮尔逊相关系数，计算表型 p 和候选基因 g 之间一致性得分.

$$CS_{pg}=\frac{\mathrm{cov}(S_p, \Phi_g)}{\sigma(S_p)\sigma(\Phi_g)} \tag{5}$$

式中：cov 代表协方差；σ 代表标准偏差. 对于前列腺癌疾病表型以及每一个候选基因，我们可利用公式(5)计算其一致性分数.

在构建前列腺癌候选基因过程中，我们发现致病基因集中的一部分基因，它们所对应的 OMIM 表型编号并没有包含在 van Driel 等构建的表型相似性数据中，因此无法得到与这些表型相关的候选基因. 但这些表型以及所涉及的候选基因对该疾病的发展也起着十分重要的作用，不能忽略. 因而需要估计出这些表型与其他表型之间的相似性.

对于某个疾病表型 p_i 而言，它与多个前列腺癌疾病表型之间已知的相似性值和所涉及基因之间的最短路径可作为先验信息，利用公式(2)，列出一个线性方程组，如公式(6)所示：

$$\begin{cases} S_{p_ip_1}=C_{p_i}+\sum\limits_{g_i\in G(p_i)}\beta_{p_ig_i}\Phi_{g_ip_1} \\ S_{p_ip_2}=C_{p_i}+\sum\limits_{g_i\in G(p_i)}\beta_{p_ig}\Phi_{g_ip_2} \\ \cdots\cdots \\ S_{p_ip_n}=C_{p_i}+\sum\limits_{g_i\in G(p_i)}\beta_{p_ig_i}\Phi_{g_ip_n} \end{cases} \tag{6}$$

式中：$\{p_1, p_2, \cdots, p_n\}$表示表型相似性矩阵中已有的前列腺癌疾病表型.

之后运用最小二乘拟合法估计出权重系数向量 $\beta_{p_ig}=\{\beta_{p_ig_1}, \beta_{p_ig_2}, \cdots, \beta_{p_ig_m}\}$和常数 C_{p_i}，因此对于某个表型相似性矩阵中未包含的前列腺癌疾病表型 p_x，则可利用估计出的系数和公式(2)，计算出它与表型 p_i 之间的相似性大小.

b. 基因之间语义相似性计算

GO 数据库可以呈现为一个有向无环图(directed acyclic graphs，DAGs)，图中各个注释为节点，两种语义关系“is-a”和“part-of”为边. 节点之间能以文本描述进行语义相似性分析.

一个基因有一个或多个 GO 注释对其进行描述，因此我们首先考虑单个 GO 注释间的语义相似性. 并认为如果一个 GO 注释用来描述某个基因或其产物，则它的所有父节点可视为参与描述该基因及其产物[34]. 以一个 GO 注释 A 为例，它可以表示为一个有向无环图 $DAG_A=(A, T_A, E_A)$，起始点为注释 A，终点为某个根术语. T_A 表示 DAG_A 中所有相关的 GO 注释的集合，包括注释 A 以及在它所有

的父注释；E_A 表示 DAG_A 中连接 GO 术语的所有边．显然，在 DAG_A 中，离 A 越近的其他注释，对 A 语义贡献也就越大．我们采用了 Wang 等[35]方法计算两个基因之间的语义相似性．采用 S 值衡量注释 t 对注释 A 的语义贡献大小．对 $DAG_A=(A,T_A,E_A)$ 中任一个注释 t，其 S 值为：

$$\begin{cases} SA(A)=1 \\ SA(t)=\max\{w_e*S_A(t')|t'\in t \text{ 的子注释}\} \text{如 } t\neq A \end{cases} \tag{7}$$

式中：w_e 是连接注释 t 和子注释 t' 的边语义贡献因子，用来表示“is-a”或“part-of”两种语义关系相似程度．一般将“is-a”关系值设为 0.8，“part-of”关系值设为 0.6．进而定义注释 A 的语义值 $SV(A)$：

$$SV(A)=\sum_{t\in T_A} S_A(t) \tag{8}$$

对于任意给定的两个注释 A 和 B，$DAG_A=(A, T_A, E_A)$和 $DAG_B=(A, T_B, E_B)$分别为它们各自的 DAG，则它们之间的语义相似性定义为：

$$S_{GO}(A, B)=\frac{\sum_{t\in T_A\cap T_B}(S_A(t)+S_B(t))}{SV(A)+SV(B)} \tag{9}$$

式中：$S_A(t)$是注释 t 关于注释 A 在 DAG_A 中的 S 值；$S_B(t)$是注释 t 关于注释 B 在 DAG_B 中 S 值．

定义一个 GO 注释 go 和一组 GO 注释集合 $GO=(go_1, go_2, \cdots, go_k)$之间的语义相似性如下：

$$Sim(go, GO)=\max_{1\leqslant i\leqslant k}(S_{GO}(go, go_i)) \tag{10}$$

假设给定两个基因 G_1 和 G_2，$GO_1=(go_{11}, go_{12}, \cdots, go_{1m})$和 $GO_2=(go_{21}, go_{22}, \cdots, go_{2n})$分别为其所对应的注释集，那么基因 G_1 和 G_2 的语义相似性定义为：

$$Sim(G_1,G_2)=\frac{\sum_{1\leqslant i\leqslant m} Sim(go_{1i},GO_2)+\sum_{1\leqslant j\leqslant n} Sim(go_{2j},GO_1)}{m+n} \tag{11}$$

根据上述方法，我们可以计算前列腺癌致病基因和候选基因之间的各个语义相似性得分，并以矩阵 $M_{m\times n}$ 表示，m 和 n 分别代表候选基因和已知致病基因个数，矩阵元素值表示两个基因的语义相似性得分．向量 P_{semantic} 表示各候选基因与致病基因集之间的语义相似性，且定义如下：

$$P_{\text{semantic}}=\frac{M_{m\times n}*P^0_{\text{semantic}}}{n} \tag{12}$$

式中：P^0_{semantic}是元素为 1 的 $n\times 1$ 初始向量．对矩阵 $M_{m\times n}$ 中某一个候选基因 g，它与各致病基因的语义相似性分别为 $S_{g1}, S_{g2}, \cdots, S_{gn}$，那么它与致病基因集的语义相似性为 $P_{\text{semantic}}(g)=(S_{g1}+S_{g2}+\cdots+S_{gn})/n$．

1.2.2　参数选取、节点 - 模块置信度及局部模块度

在对候选基因打分过程中，参数 r 的选择对于候选基因的最终综合得分和排名起着至关重要作用．本文中，我们采用计算平均排名比(mean rank ratio，MRR)来确定参数 r[36].

对于某个特定预测模型，如果根据某种方法得到各元素的排名为($rank_1$, $rank_2$, …, $rank_n$)，则元素 a_i 的排名比为 $rank_1/n$，排名越靠前的元素与该问题的关联程度越紧密．数据集平均排名比 MRR 定义如下：

$$MRR=\frac{rank_1+rank_2+\cdots+rank_n}{n^2} \tag{13}$$

对于已知包含 n 个致病基因的前列腺癌致病基因集，我们预设一组不同 γ 参数并采用留一法验证，在得到不同数值下的平均排名比后，利用 Matlab 多项式曲线拟合工具拟合出参数 γ 和 MRR 的关系曲线，如图 2 所示．最后，选择 MRR 值最小时所对应的参数为 γ 最优值，即：

$$\gamma=\arg\min(MRR) \tag{14}$$

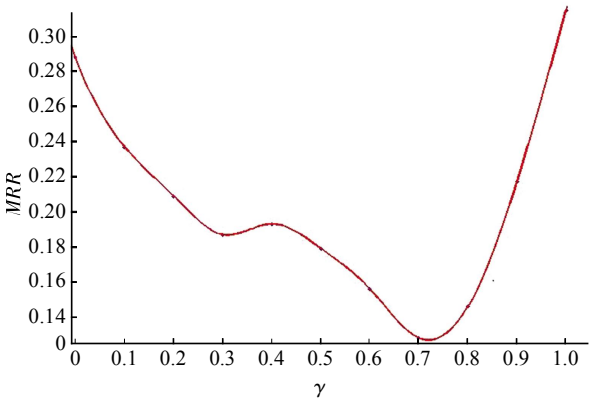

Fig. 2　The relationship fitting curve between parameter γ and MRR

从图 2 拟合曲线可以看出，γ=0.7～0.8，平均排名比值较小，γ=0.72 平均排名比值最小．通过实验验证，平均排名比为 11.73%，故 γ=0.72.

对于无权无向网络 $G=(V, E)$，V 为顶点集合，E 为边集合，$u, v\in V$，$<u, v>\in E$．M 为邻接矩阵，M^2 包含网络中两个节点能否通过一个中转节点相连信息，M^l 包含网络中两个节点能否通过$(l-1)$个中转节点相连信息．考虑到网络复杂度和算法效率，本文取 l=3，定义边 $<u, v>$ 重要度 $S(u, v)$：

$$S(u, v)=M(u, v)+M^2(u, v)+M^3(u, v) \tag{15}$$

对网络中任一节点，可计算出其节点重要度，即将网络中与此节点相连的边重要度相加，即：

$$S(u)=\sum_{v\in V}S(u,v) \quad (16)$$

假设 m 为网络中某一个模块，v 为 m 外面的一个节点，则节点 - 模块置信度定义为：

$$I_{(m,v)}=\frac{|E_{vm}|}{|D_v|}\Big/\frac{2\times|E_m|}{|V_m|\times|V_m-1|} \quad (17)$$

式中，$|E_{vm}|$为节点 v 与模块 m 的连接边数，$|D_v|$为节点 v 的度，$|E_m|$为模块 m 的边数，$|V_m|$为模块 m 的节点数.

模块 m 的局部模块度定义为[37]：

$$Q_m=\frac{L_{in}}{L_{in}+L_{out}} \quad (18)$$

式中：L_{in} 为模块内部节点间边的数量；L_{out} 为模块内部节点与模块外部节点间边的数量.

1.2.3 NMCOM 算法计算步骤

Step 1：对候选基因进行打分，并按值从大到小顺序排序，从排名最高的候选基因开始依次选取一个基因作为起始节点.

Step 2：对于某个起始节点 v_a，在 PPI 网络中找出它的邻居节点，计算这些邻居节点的节点重要度. 选取节点重要度最高节点 v'，节点重要度次高节点 v''，形成初始模块 m_0.

Step 3：计算初始模块的局部模块度 Q'_m 和初始节点 - 模块置信度. 初始节点 - 模块置信度 $I'_{(m,v_i)}=I_{(m',v'')}$，定义为节点 v'' 与模块 m' 间的置信度，模块 m' 由起始节点 v_a 和节点 v' 构成.

Step 4：寻找当前模块 m 的邻居节点，并对这些节点按节点重要度从高到低进行排序，得到节点集合 $N=\{v_1, v_2, \cdots, v_n\}$.

Step 5：在集合 N 中从节点重要度最大的节点为开始，计算该节点 v_i 与当前模块 m 之间的置信度 $I_{(m,v_i)}$，判断 $\Delta I=I_{(m,v_i)}-I'_{(m,v_i)}$，如果 $\Delta I\geqslant 0$，则将 $I'_{(m,v_i)}$更新为 $I_{(m,v_i)}$并跳转到 Step 6；如果 $\Delta I<0$，则舍弃该节点，并从集合 N中选择下一个节点重复 Step 5 的计算工作，以此类推，直至遍历集合 N 中的所有节点.

Step 6：该节点暂时加入当前模块 m 形成新模块，计算新模块的局部模块度 Q_m，判断 $\Delta Q=Q_m-Q'_m$，如果 $\Delta Q\geqslant 0$，则确认该节点加入并形成新模块，同时将更新 Q'_m 为 Q_m；如果 $\Delta Q<0$，则舍弃该节点，从集合 N 中选择下一个节点并跳转到 Step 5 进行计算.

Step 7：判断 Step 6 中是否有新的节点加入模块中. 如是，则转到 Step 4，如否，该模块的挖掘工作结束.

Step 8：将挖掘出的模块进行富集分析. 如果满足显著性阈值条件，则保留该模块为疾病候选模块，如果不满足，则舍弃.

Step 9：将疾病候选模块中的节点从 PPI 网络中删除，并从下一个候选节点起始，重复进行模块挖掘. 如果连续三个模块都被认定为是非疾病候选模块，则整个模块挖掘工作结束.

2 结果与讨论

2.1 模块挖掘结果及候选基因功能描述

在致病基因集中，我们共收集了 150 个前列腺癌致病基因，其中 115 个来自于 PGD 数据库，35 个来自于 OMIM 数据库. 通过表型间相似性估计构建了完整的候选基因集，得到 3 632 个候选基因. 按照 NMCOM 算法进行候选疾病模块挖掘，并应用 Gavin Sherlock 和 Shuai Weng 等开发的在线工具 Go-TermFinder(http://go.princeton.edu/cgi-bin/GOTermFinder?load=5085021)来进行富集性分析和确认. 分析时显著性阈值 *P*-value 默认为 0.01. 最终得到 18 个阈值小于 0.01，且具有一定生物学意义的前列腺癌候选疾病模块.

每一个候选疾病模块都对前列腺癌疾病的发生或发展起着关键的作用. 但由于篇幅所限，这里仅给出排名第一和第二的基因 FGF3、KLK3 的候选疾病模块，分别如图 3、4 所示. 对于其他候选疾病模块，我们在附件表 S1～S32 中列出了较为详细的显著富集分析，并分析预测了它们与前列腺癌疾病的关系以及在该疾病中所起的作用.

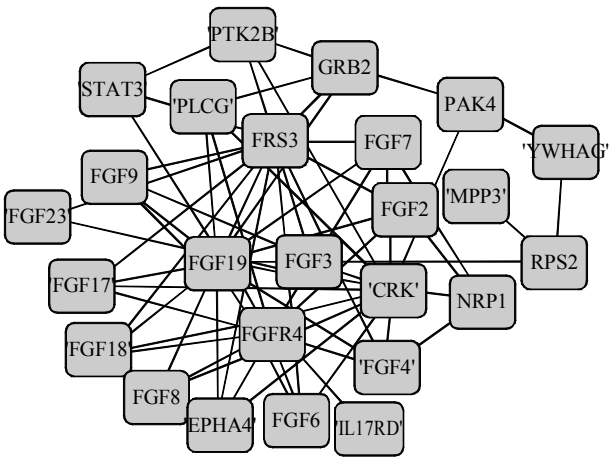

Fig. 3 The candidate disease module of FGF3 gene

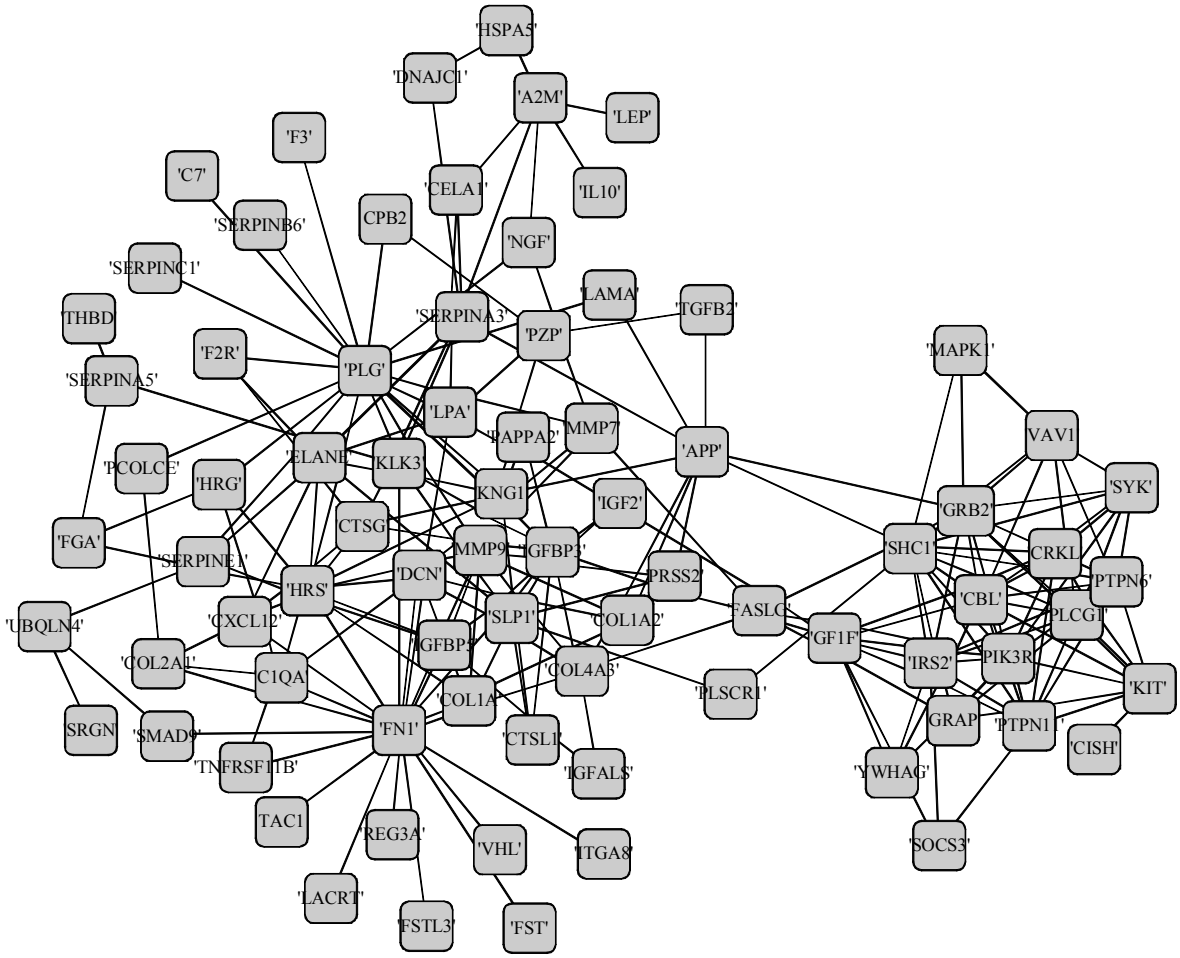

Fig. 4　The candidate disease module of KLK3 gene

此外，我们还在附录中列出了部分排名靠前候选种子基因以及相应的基因描述，每一个候选基因都对应一个候选疾病模块并且该模块都与前列腺癌有密切联系，如附件表 S33 所示．这些基因与前列腺癌的关系是借助于在线数据库查询和文献检索的方法进行描述的．

2.2　候选疾病模块富集分析

我们以富集性分析后筛选出的 FGF3 模块和 KLK3 模块为例，并利用生物信息数据平台 DAVID[38](The database for annotation，visualization and integrated discovery，http://david.abcc.ncifcrf.gov/home.jsp)，分别对这两个候选疾病模块在生物过程(biology process，BP)、细胞成分(cell component，CC)和分子功能(molecular function，MF)以及 KEGG 通路中部分高富集条目($P < 0.01$)进行了展示，如表 1～4 所示．其中，模块频率表示在疾病模块中被注释到 GO 数据库中特定条目的基因数占所有被注释到 GO 数据库中的基因数目的比例，基因组频率表示注释到特定条目的基因数目占人类全基因组基因数目的比例，基因数目表示对应通路下所包含的基因数目．

从图 3、表 1 和表 2 中可以看到，FGF3 候选疾病模块包含 26 个基因，它们之间的相互关系十分密切，其中最多的一条 GO 注释中包含有 19 个模块中的基因，约占模块中所有基因数目的 73%．该模块的富集条目主要集中在细胞生长分裂、增殖和代谢的过程中，例如细胞分裂、细胞增殖、受体信号传导等等．其中在 GO 富集分析中：GO:0008543 涉及到细胞生长及肿瘤的生长和入侵；GO:0044344 涉及到睾丸间质细胞的活动；GO:0044255 对睾丸间质细胞、肾上腺皮质有较强的正向性，文献[39]的研究结果也认为该注释在前列腺癌的发生和发展过程中有显著性的作用．在 KEGG 通路富集分析中：hsa04810 与性腺功能减退疾病

Table 1 Partial results of GO enrichment analysis for FGF3 candidate disease module

	GO ID	GO term	Cluster frequency	Genome frequency	*P*-value
BP	GO:0008543	Fibroblast growth factor receptor signaling pathway	19/26	272/45576	8.83×10^{-29}
	GO:0044344	Cellular response to fibroblast growth factor stimulus	15/26	256/45567	5.59×10^{-27}
	GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	15/26	258/45576	6.10×10^{-26}
	GO:0071774	Response to fibroblast growth factor	16/26	263/45576	3.72×10^{-24}
	GO:0007167	Enzyme linked receptor protein signaling pathway	16/26	268/45576	4.32×10^{-21}
	GO:0008284	Positive regulation of cell proliferation	15/26	259/45576	2.46×10^{-19}
	GO:0048011	Neurotrophin TRK receptor signaling pathway	14/26	285/45576	1.84×10^{-15}
	GO:0044255	Cellular lipid metabolic process	14/26	288/45576	2.07×10^{-15}
	GO:0038093	Fc receptor signaling pathway	13/26	350/45576	1.80×10^{-14}
	GO:0008286	Insulin receptor signaling pathway	14/26	239/45576	3.54×10^{-14}
	GO:0007173	Epidermal growth factor receptor signaling pathway	13/26	252/45576	6.05×10^{-14}
	GO:0038127	ERBB signaling pathway	12/26	259/45576	7.98×10^{-14}
CC	GO:0005615	Extracellular space	22/26	286/45576	1.55×10^{-27}
	GO:0005576	Extracellular region	14/26	224/45576	6.63×10^{-25}
	GO:0044445	Cytosolic part	14/26	237/45576	7.45×10^{-25}
MF	GO:0008083	Growth factor activity	14/26	246/45576	9.56×10^{-26}
	GO:0060090	Binding	13/26	229/45576	3.97×10^{-25}

Table 2 Partial results of KEGG pathway enrichment analysis for FGF3 candidate disease module

KEGG ID	KEGG term	Count	*P*-value
hsa04810	Regulation of actin cytoskeleton	24	3.07×10^{-26}
hsa05200	Pathways in cancer	19	1.21×10^{-25}
hsa04010	MAPK signaling pathway	17	6.63×10^{-19}
hsa05215	Prostate cancer	16	2.28×10^{-17}

Table 3 Partial results of GO enrichment analysis for KLK3 candidate disease module

	GO ID	GO term	Cluster frequency	Genome frequency	*P*-value
BP	GO:0009611	Response to wounding	67/86	7414/45576	1.42×10^{-37}
	GO:0032101	Regulation of response to external stimulus	58/86	6875/45576	2.57×10^{-36}
	GO:0042060	Wound healing	57/86	9258/45576	7.95×10^{-33}
	GO:0030193	Regulation of blood coagulation	53/86	5729/45576	4.09×10^{-28}
	GO:0007599	Hemostasis	51/86	4381/45576	8.77×10^{-28}
	GO:0007596	Blood coagulation	47/86	5729/45576	3.26×10^{-27}
	GO:0050817	Coagulation	46/86	4985/45576	3.39×10^{-27}
	GO:0050878	Regulation of body fluid levels	46/86	8914/45576	1.59×10^{-26}
	GO:0050818	Regulation of coagulation	44/86	4501/45576	1.88×10^{-26}
	GO:0030574	Collagen catabolic process	45/86	7705/45576	1.62×10^{-24}
	GO:0042221	Response to chemical	43/86	4016/45576	1.89×10^{-23}
	GO:0044243	Multicellular organismal catabolic process	44/86	4558/45576	6.32×10^{-23}
	GO:0032963	Collagen metabolic process	42/86	6382/45576	2.64×10^{-22}
	GO:0006950	Response to stress	41/86	5886/45576	4.30×10^{-21}
	GO:0032501	Multicellular organismal process	39/86	6758/45576	7.67×10^{-20}
	GO:0044259	Multicellular organismal macromolecule metabolic process	39/86	5821/45576	8.09×10^{-19}
	GO:0044236	Multicellular organismal metabolic process	38/86	5387/45576	5.43×10^{-18}
CC	GO:0005576	Extracellular region	62/86	8643/45576	3.87×10^{-46}
	GO:0044421	Extracellular region part	57/86	8586/45576	5.06×10^{-45}
	GO:0005615	Extracellular space	56/86	8958/45576	2.17×10^{-44}
	GO:0031012	Extracellular matrix	56/86	8673/45576	1.92×10^{-43}
	GO:0060205	Cytoplasmic membrane-bounded vesicle lumen	53/86	9152/45576	7.25×10^{-42}
	GO:0030141	Secretory granule	53/86	7998/45576	2.97×10^{-41}
MF	GO:0004175	Endopeptidase activity	61/86	9637/45576	9.17×10^{-38}
	GO:0008233	Peptidase activity	57/86	8249/45576	3.23×10^{-35}
	GO:0004252	Serine-type endopeptidase activity	57/86	8647/45576	1.82×10^{-33}
	GO:0070011	Peptidase activity, acting on L-amino acid peptides	55/86	7985/45576	8.37×10^{-33}
	GO:0008236	Serine-type peptidase activity	53/86	8355/45576	1.76×10^{-31}
	GO:0005509	Calcium ion binding	52/86	7983/45576	1.68×10^{-25}
	GO:0019838	Growth factor binding	53/86	7758/45576	1.08×10^{-23}

Table 4　Partial results of KEGG pathway enrichment analysis for KLK3 candidate disease module

KEGG ID	KEGG term	Count	*P*-value
hsa04610	Complement and coagulation cascades	58	1.18×10^{-28}
hsa05200	Pathways in cancer	47	4.22×10^{-23}
hsa04012	ErbB signaling pathway	44	2.57×10^{-21}
hsa05211	Renal cell carcinoma	42	9.23×10^{-19}
hsa05215	Prostate cancer	42	9.37×10^{-19}
hsa05219	Bladder cancer	34	7.11×10^{-18}
hsa04512	ECM-receptor interaction	34	2.68×10^{-16}
hsa04910	Insulin signaling pathway	32	2.95×10^{-15}

有关；hsa04810 与人类多种癌症有关，如家族性腺瘤性息肉病、李法美尼症候群、先天缺牙 - 结肠癌综合征等；hsa04010 涉及到细胞增殖、变异和迁移；hsa05215 则直接涉及到前列腺癌的发展机制．细胞分裂共同调节和维持着组织器官细胞数目的平衡，恶性肿瘤的产生正是这种平衡被破坏的结果．根据该模块在细胞生长因子生长、细胞增殖调控、受体信号、细胞外间隙和区域等 GO 条目表现出的高富集性，推测该模块可能在前列腺癌细胞失控增生及疾病信号传导方面起一定的作用．

从图 4、表 3 和表 4 中可以看到，KLK3 候选疾病模块包含 86 个基因，基因间相互关系也很密切，其中最多的一条 GO 注释中包含有 67 个模块中的基因，约占模块中所有基因数目的 78%．该模块的富集条目主要集中在细胞的血液凝固机制和代谢的过程中，例如凝血调节、止血作用、分解代谢等．其中 GO:0007596 是血液多种凝血因子相互作用的连续过程，GO:0050878 对人类体液有直接或间接的调节作用，包括前列腺液，GO:0032963 涉及到胶原的化学反应和途径，而文献[40]认为Ⅳ型胶原在前列腺癌的表达中有显著意义．在 KEGG 通路富集分析中：hsa04610 与遗传性血管性水肿和多种凝血因子缺乏有关；hsa04012 涉及到多种肿瘤综合征 (如 PTEN 错构瘤、垂体腺瘤等)；hsa05215、hsa05219 和 hsa05211 涉及到前列腺癌和相关的泌尿系癌症(如肾细胞癌、膀胱癌等)；hsa04512 与多种血液相关的疾病有关，hsa04910 所涉及的胰岛素样生长因子则在前列腺癌的发生发展过程中发挥着重要的作用．血液凝固在癌细胞的生长、浸润和转移形成的各个阶段中有重要作用，尤其是癌细胞的血行转移中形成的微血栓，是癌细胞得以着床和增生的条件．根据该模块在细胞血液凝固、体液调控、肽酶活性等 GO 条目表现出的高富集性，推测该模块可能在血液促进或抑制方面，对前列腺癌细胞的增生和转移有一定的影响．

文献[30]也将识别出的前列腺癌功能模块进行了富集分析，列出了模块中具有显著意义的 GO 注释及相应的 *P* 值．其中某些 GO 注释与本文进行模块富集分析时得到的 GO 注释相同，我们将这些相同的 GO 注释和各自的 *P* 值进行比较，结果如表 5 所示．

Table 5　The compare of *P*-value of the same GO annotation between literature [30] and NMCOM

GO ID	GO term	*P*-value from literature [30]	*P*-value from NMCOM
GO:0007229	Integrin-mediated signaling pathway	3.64×10^{-6}	1.77×10^{-10}
GO:0008286	Insulin receptor signaling pathway	6.86×10^{-6}	2.32×10^{-9}
GO:0006916	Negative regulation of apoptotic process	9.88×10^{-6}	8.50×10^{-9}
GO:0001525	Angiogenesis	3.83×10^{-5}	1.61×10^{-7}
GO:0007409	Axonogenesis	4.04×10^{-5}	5.27×10^{-3}
GO:0006261	DNA-dependent DNA replication	1.01×10^{-4}	9.63×10^{-7}
GO:0008285	Negative regulation of cell proliferation	1.12×10^{-4}	1.37×10^{-6}
GO:0000187	Activation of MAPK activity	3.46×10^{-4}	1.01×10^{-6}
GO:0007249	I-kappaB kinase/NF-kappaB signaling	8.48×10^{-4}	7.83×10^{-5}
GO:0007265	Ras protein signal transduction	1.66×10^{-2}	9.61×10^{-5}
GO:0051052	Regulation of DNA metabolic process	1.45×10^{-1}	9.92×10^{-4}

从表 5 中可以看出，对于相同的 GO 注释，NMCOM 算法在进行富集分析时具有更小的 *P* 值，即具有更显著的统计学意义，从而说明 NMCOM 算法更加有效可靠.

在表 5 所展示的这些共同的 GO 注释中，已有研究文献表明某些注释与前列腺癌有关，例如：“介导整合素信号路径”(GO：0007229)，涉及到前列腺癌细胞的控制增生、存活以及迁移[41]；I-kappaB 致活酶 /NF-kappaB 信号(GO：0007249)，I-kappaB 致活酶能调节转录因子核 kappaB 因子(NF-kappaB)，在哺乳动物细胞中是一个关键的细胞凋亡抗体[42]；MAPK 活性活化作用(GO：0000187)，可实现前列腺癌细胞从雄性激素敏感到雄性激素独立通过不同的多步路径的转换，包括通过 MAPK 的自适应的雄性激素受体路径[43]. 还有一些本文挖掘出的疾病候选模块中的 GO 注释，也涉及到前列腺癌，例如：血管内皮生长因子生成(GO：0010573)，在前列腺癌中异常表达，可造成前列腺中新血管的生成[44]；胆碱氧化途径(GO：0019285)，该路径中的肌氨酸和甘氨酸高基线血清浓度与减少前列腺癌的风险有关[45]；AKT 路径(GO:0051896)，与前列腺癌增生和转移 / 入侵有关[46]；环氧酶路径(GO:0019371)，是一种在不同细胞类型中产生的分裂素、肿瘤激活子、细胞因子和生长因子而诱发的促炎性和诱导酶，它控制着转录和翻译后水平，在前列腺癌的发展过程中通常有过表达[47]等.

2.3　基于全 PPI 网络的模块挖掘

随着生命科学研究的深入和发展，人们逐渐认识到，生物体中的某个蛋白质往往具有不止一个功能，不同的功能模块可能会拥有共同的蛋白质或蛋白质产物，即模块间可能存在交叠[48-50]. 因此，我们尝试将 NMCOM 算法中的 Step 9 做了一些调整：对挖掘出候选疾病模块，其相应的所有节点在 PPI 网络中不予删除，保证每个模块都能从全 PPI 网络中进行挖掘，避免之前删除节点后可能导致下一个模块信息不完整的情况.

根据调整后的方法我们进行了全网络的候选模块挖掘，依旧挖掘出 18 个候选前列腺癌疾病模块. 其中，13 个候选疾病模块没有任何变化，只有 5 个候选疾病模块存在部分模块交叠. 我们将这 5个模块进行了展示(图 5～7). 从图中可以看到，两个模块之间只有少量的蛋白质存在交叠(交叠蛋白质用三角型表示)，同时通过对比富集分析，发现原候选模块补充了重叠的蛋白质后，尽管模块节点信息更为完整，但都并非核心节点，因此模块中显著意义的 GO 和 KEGG 通路富集分析数目及其相应的 *P* 值，几乎没有太大的变化甚至无变化，但是该模块挖掘的效率，却因挖掘范围的扩大而大大降低. 例如 PDS5B 候选模块，在 CPU 为 Pentium E6300，内存大小为 4G，运行环境为 Matlab 2013a 的计算机上，挖掘所用时间之前约为 3～4 min，扩展网络后的运行时间约为 10 min 左右.

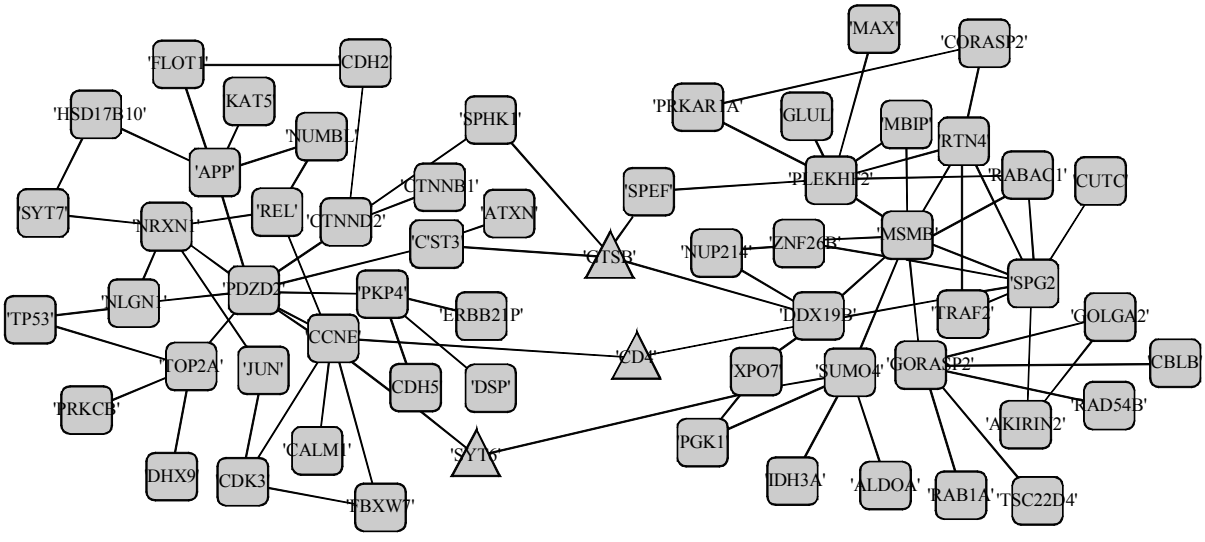

Fig. 5　The overlapping candidate disease modules of PDZD2 and MSMB genes

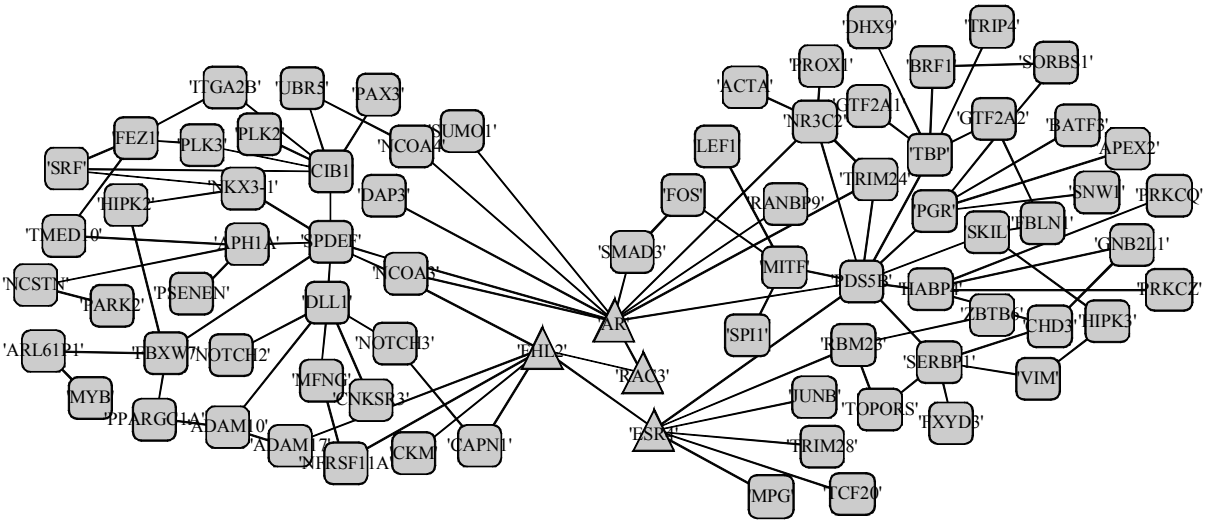

Fig. 6　The overlapping candidate disease modules of SPDEF and PDS5B genes

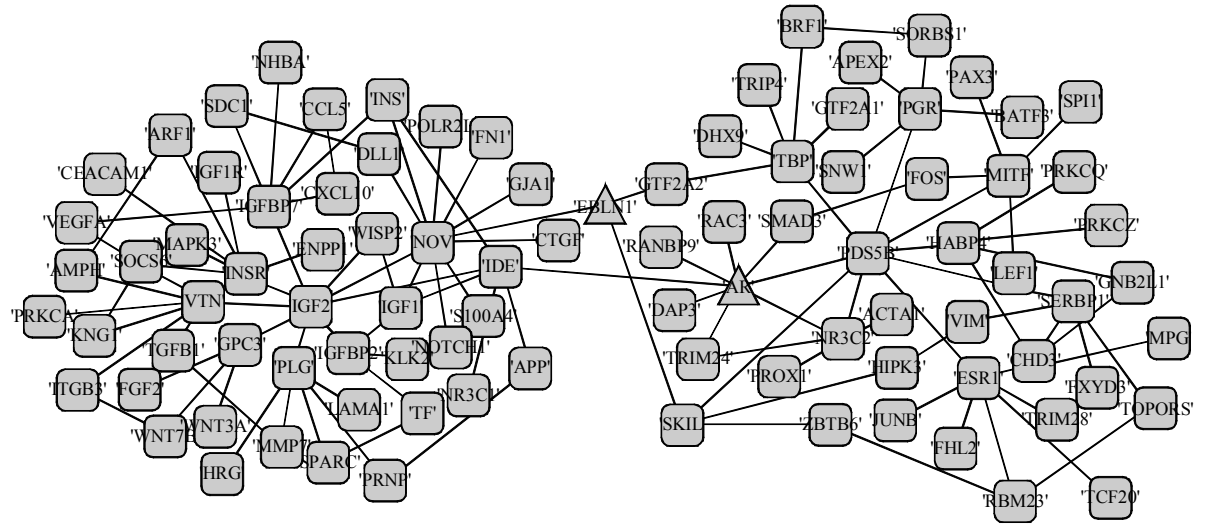

Fig. 7　The overlapping candidate disease modules of IGF2 and PDS5B genes

2.4　NMCOM 与其他算法比较

NMCOM 与重启随机游走(random walk with restart，RWR)算法[5]的致病基因平均排名比对比结果如表 6 所示．图 8 为二种方法的 ROC(receiver operator characteristic)曲线比较结果．

Table 6　The MRR of pathogenic genes using different methods

Method	*MRR*/%
Concordance between phenotypes and genes	24.57
Semantic similarity between genes	19.23
RWR	17.32
Rank method from NMCOM	11.73

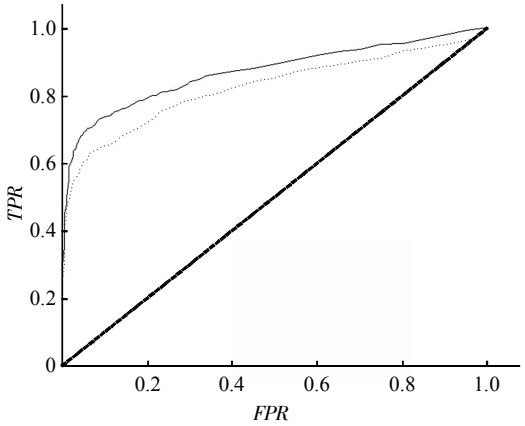

Fig. 8　The ROC curve for rank method of NMCOM and RWR

—— : NMCOMY($AUC \approx 92.27\%$); ·····: RWR($AUC \approx 83.63\%$).

从表 6 中可以看出，NMCOM 方法的平均排名比最小，优于融合前的单一方法和 RWR 方法. 从图 8 中可以看出，NMCOM 方法的 ROC 曲线面积大约比 RWR 方法高 8.64%，说明 NMCOM优于非融合和 RWR 方法.

相对于我们以前提出的基于边、节点信息相融合的网络模块挖掘 CDIEV 算法[52]，NMCOM 算法增加局部模块度约束判定是否形成模块. NMCOM、CDIEV 及 RWR 三种算法的候选疾病模块挖掘结果如表 7 所示.

Table 7 The module mining results for NMCOM, CDIEV and RWR

	NMCOM	CDIEV	RWR
The number of modules	18	5	11
The average *P* value of modules	7.68×10^{-9}	8.56×10^{-3}	1.55×10^{-6}
Top 5/Top 10 average *P* value of modules	2.98×10^{-15}/6.367×10^{-13}	4.57×10^{-9}/7.29×10^{-8}	5.157×10^{-11}/8.53×10^{-10}
The average density of modules	0.792	0.526	0.657

平均模块 *P* 值定义为所得模块中的所有 GO 注释 *P* 值的平均值. 为了避免某个模块中某几项 GO 注释的 *P* 值较大，而其他 GO 注释的 *P* 值都较小，从而导致整个模块的平均 *P* 值变大，而与实际相反的情况，我们定义了模块 Top 5/Top 10 平均 *P* 值，即为不同算法所挖掘的各个模块中 GO 注释的 *P* 值按从小到大的顺序，所有排名前 5 或前 10 的 *P* 值的平均值. 平均模块密度定义为所得各个模块密度的平均值，模块密度定义 D_s 为 $D_s=2|E_c|/|V_c|(|V_c|-1)$，$|E_c|$ 表示模块边数，$|V_c|$ 表示模块节点数，$D_s\in[0, 1]$, 密度越大，说明模块内连接越紧密，则更具有生物学意义.

从表 7 中可以看出，NMCOM 算法与其他两种方法相比，不仅能挖掘出更多数目的模块，而且模块 Top 5/Top 10 平均 *P* 值以及平均模块 *P* 值也小好几个数量级，模块的密度也比较大. 因此 NMCOM 算法对于前列腺癌候选模块挖掘具有更好的显著性和生物意义.

3 结 论

针对目前前列腺癌疾病研究主要利用基因表达谱数据和所构建的子网络没有进行深入筛选和挖掘等问题，本文提出了一种新的基于节点 - 模块置信度和局部模块度双重约束的模块挖掘 NMCOM 算法. 该算法不使用基因表达谱数据，而是整合了疾病表型信息、GO 注释信息、PPI 网络等对候选基因进行打分排序，并使用“种子 - 扩充”的思想进行模块挖掘和富集分析模块确认. 试验结果表明，NMCOM 方法所选取的起始基因大都与该疾病有密切联系、所挖掘出的模块也有较好的富集显著性，在疾病的发生和发展过程中起到了一定的作用和影响. 同时与其他打分方法相比，NMCOM 算法排序策略得到的致病基因平均排名比较低，*AUC* 值大于 RWR 方法；与其他模块挖掘算法相比，NMCOM 算法挖掘的模块不仅有较多数目的显著 GO 注释，而且在模块数目、平均模块 *P* 值、模块 Top 5/Top 10 平均 *P* 值、平均模块密度等方面具有明显的优势.

综上所述，NMCOM 算法对于前列腺癌候选疾病模块挖掘而言，算法简单、结果有效可靠，不仅能挖掘出具有一定意义的疾病候选模块，增强对前列腺癌病理的了解，还能对该疾病的诊断、预防和治疗有很大的帮助.

附件 表 S1～S33 见本文网络版附录(http://www.pibb.ac.cn)

参 考 文 献

[1] Jathal M K, Chen L, Mudryj M, *et al*. Targeting ErbB3: the new RTK (id) on the prostate cancer block. Immunology. Endocrine & Metabolic Agents in Medicinal Chemistry, 2011, **11**(2): 131

[2] Siegel R, Jemal A. Cancer Facts and Figures 2012. GA: American Cancer Society, 2012, 10

[3] 孙颖浩. 我国前列腺癌的研究现状. 中华泌尿外科杂志，2004, **25**(2): 77-80
Sun Y H. Chinese Journal of Urology, 2004, **25**(2): 77-80

[4] Rasheed S A K, Teo C R, Beillard E J, *et al*. MicroRNA-182 and microRNA-200a control G-protein subunit α-13 (GNA13) expression and cell invasion synergistically in prostate cancer cells. J Biol Chem, 2013, **288**(11): 7986-7995

[5] Baena E, Shao Z, Linn D E, *et al*. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice

and patients. Genes & Development, 2013, **27**(6): 683−698

[6] Ding G, Xu W, Liu H, *et al*. CYP1A1 MspI polymorphism is associated with prostate cancer susceptibility: evidence from a meta-analysis. Molecular Biology Reports, 2013, **40**(5): 3483−3491

[7] Spirin V, Mirny L A. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci USA, 2003, **100** (21): 12123−12128

[8] Bader G D, Hogue C W V. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics, 2003, **4**(1): 2

[9] Ravasz E, Somera A L, Mongru D A, *et al*. Hierarchical organization of modularity in metabolic networks. Science, 2002, **297**(5586): 1551−1555

[10] Arnau V, Mars S, Marín I. Iterative cluster analysis of protein interaction data. Bioinformatics, 2005, **21**(3): 364−378

[11] King A D, Pržulj N, Jurisica I. Protein complex prediction *via* cost-based clustering. Bioinformatics, 2004, **20**(17): 3013−3020

[12] Dunn R, Dudbridge F, Sanderson C M. The use of edge-betweenness clustering to investigate biological function in protein interaction networks. BMC Bioinformatics, 2005, **6**(1): 39

[13] Frey B J, Dueck D. Clustering by passing messages between data points. Science, 2007, **315** (5814): 972−976

[14] Van Dongen S. A cluster algorithm for graphs. Report-Information Systems, 2000, **47**(10): 1−40

[15] Kamp C, Christensen K. Spectral analysis of protein-protein interactions in *Drosophila* melanogaster. Physical Review E, 2005, **71**(4): 041911

[16] Inoue K, Li W, Kurata H. Diffusion model based spectral clustering for protein-protein interaction networks. PloS one, 2010, **5** (9): e12623

[17] Leung H C M, Xiang Q, Yiu S M, *et al*. Predicting protein complexes from PPI data: a core-attachment approach. Journal of Computational Biology, 2009, **16**(2): 133−144

[18] Ma X, Gao L. Predicting protein complexes in protein interaction networks using a core-attachment algorithm based on graph communicability. Information Sciences, 2012, **189**: 233−254

[19] Ji J, Liu Z, Zhang A, *et al*. Improved ant colony optimization for detecting functional modules in protein-protein interaction networks. Springer Berlin Heidelberg, 2012: 404−413

[20] Lei X J, Huang X, Wu S, *et al*. Joint strength based ant colony optimization clustering algorithm for PPI networks. Dianzi Xuebao (Acta Electronica Sinica), 2012, **40**(4): 695−702

[21] Lu J, Getz G, Miska E A, *et al*. MicroRNA expression profiles classify human cancers. Nature, 2005, **435**(7043): 834−838

[22] Massa C, Rusu M, Wang H, *et al*. Structural and functional modeling of pulmonary function in heterogenous lung pathology. Am J Respir Crit Care Med, 2014, **189**: A3572

[23] 赵 丹, 吴宏宇, 韩一平, 等. 在线肺癌病例数据库系统的建立. 解放军医院管理杂志, 2013, **20**(3): 223−225
Zhao D, Wu H Y, Han Y P, *et al*. Hospital Administration Journal of Chinese People's Liberation Army, 2013, (3): 223−225

[24] Wang Y C, Chen B S. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. BMC Medical Genomics, 2011, **4**(1): 2

[25] Li J, Bi L, Sun Y, *et al*. Text mining and network analysis of molecular interaction in non-small cell lung cancer by using natural language processing. Molecular Biology Reports, 2014, **41**(12): 1−9

[26] Mitra R, Edmonds M D, Sun J, *et al*. Reproducible combinatorial regulatory networks elucidate novel oncogenic microRNAs in non-small cell lung cancer. RNA, 2014, **20**(9): 1356−1368

[27] Lv Y, He Y, Miao Z, *et al*. Identification of lung cancer related function modules based on Co-expression network. Biophysics, 2013, **1**(1): 17−24

[28] Wang X Y, Hao J W, Zhou R J, *et al*. Meta-analysis of gene expression data identifies causal genes for prostate cancer. Asian Pacific Journal of Cancer Prevention, 2013, **14**(1): 457−461

[29] Jin G, Zhou X, Cui K, *et al*. Cross-platform method for identifying candidate network biomarkers for prostate cancer. Systems Biology, 2009, **3**(6): 505−512

[30] Guo Z, Li Y, Gong X, *et al*. Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network. Bioinformatics, 2007, **23**(16): 2121−2128

[31] Martinez E, Trevino V. Modelling gene expression profiles related to prostate tumor progression using binary states. Theoretical Biology and Medical Modelling, 2013, **10**(1): 37

[32] van Driel M A, Bruggeman J, Vriend G, *et al*. A text-mining analysis of the human phenome. Eur J Hum Genet 2006, **14** (5): 535−542

[33] Wu X, Jiang R, Zhang M Q, *et al*. Network-based global inference of human disease genes. Molecular Systems Biology, 2008, **4** (1): 189−202

[34] 杨 文, 孙继林. GO 在生物数据整合中的应用. 图书情报工作, 2008, **52**(11): 124−127
Yang W, Sun J L. Library and Information Service, 2008, **52** (11): 124−127

[35] Wang J Z, Du Z, Payattakool R, *et al*. A new method to measure the semantic similarity of GO terms. Bioinformatics, 2007, **23** (10): 1274−1281

[36] Jiang R, Gan M, He P. Constructing a gene semantic similarity network for the inference of disease genes. BMC Systems Biology, 2011, **5**(Suppl 2): S2

[37] Clauset A. Finding local community structure in networks. Physical Review E, 2005, **72**(2): 026132

[38] Alvord G, Roayaei J, Stephens R, *et al*. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol, 2007, **8**(9): R183

[39] Uhlen M, Oksvold P, Fagerberg L, *et al*. Towards a knowledge-based human protein atlas. Nat Biotechnol, 2010, **28** (12): 1248−1250

[40] 杨长俊, 杨艳丽, 吴莉莉. Ⅳ型胶原和层粘连蛋白在前列腺良恶性疾病鉴别诊断中的应用. 中国误诊学杂志, 2008, **8** (33): 8126−8127
Yang C J, Yang Y L, Wu L L. Chinese Journal of Misdiagnostics, 2008, **8**(33): 8126−8127

[41] Fornaro M, Manes T, Languino L R. Integrins and prostate cancer metastases. Cancer and Metastasis Reviews, 2001, **20**(3-4): 321-331

[42] Gasparian A V, Yao Y J, Lu J, *et al*. Selenium compounds inhibit I kappa B kinase (IKK) and nuclear factor-kappa B (NF-kappa B) in prostate cancer cells. Mol Cancer Ther, 2002, **1**(12): 1079-1087

[43] Edwards J, Bartlett J. The androgen receptor and signal-transduction pathways in hormone-refractory prostate cancer. Part 1: modifications to the androgen receptor. BJU International, 2005, **95**(9): 1320-1326

[44] Wozney J L, Antonarakis E S. Growth factor and signaling pathways and their relevance to prostate cancer therapeutics. Cancer and Metastasis Reviews, 2014: 1-14

[45] Vogel S, Ulvik A, Meyer K, *et al*. Sarcosine and other metabolites along the choline oxidation pathway in relation to prostate cancer—a large nested case-control study within the JANUS cohort in norway. International Journal of Cancer, 2014, **134**(1): 197-206

[46] Majid S, Dar A A, Saini S, *et al*. miRNA-34b inhibits prostate cancer through demethylation, active chromatin modifications, and AKT pathways. Clinical Cancer Research, 2013, **19**(1): 73-84

[47] Khan N, Mukhtar H. Modulation of signaling pathways in prostate cancer by green tea polyphenols. Biochemical Pharmacology, 2012, **85**(5): 667-672

[48] Li Y, Liang C, Wong K C, *et al*. Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion. Bioinformatics, 2014: btu373

[49] Pizzuti C, Rombo S E. Algorithms and tools for protein - protein interaction networks clustering, with a special focus on population-based stochastic methods. Bioinformatics, 2014, **30**(10): 1343-1352

[50] Zhang X F, Dai D Q, Ou-Yang L, *et al*. Detecting overlapping protein complexes based on a generative model with functional and topological properties. BMC Bioinformatics, 2014, **15**(1): 186

[51] Köhler S, Bauer S, Horn D, *et al*. Walking the interactome for prioritization of candidate disease genes. The American Journal of Human Genetics, 2008, **82**(4): 949-958

[52] 张绍武, 丁 鹏, 张庭赫. 基于边, 节点信息融合网络社团挖掘算法的海洋微生物作用模式. 科学通报, 2013, **58**(Z2): 2980-2986
Zhang S W, Ding P, Zhang T H. Chinese Science Bulletin, 2013, **58**(Z2): 2980-2986

Uncovering Prostate Cancer Candidate Disease Modules With Dual Constraints Based on Node-Module Confidence and Local Modularity*

WANG Yi-Bin, CHENG Yong-Mei, ZHANG Shao-Wu**

(*School of Automation, Key Laboratory of Information Fusion Technology of Ministry of Education, Northwestern Polytechnical University, Xi'an* 710072, China)

Abstract Researches on the etiology and pathogenesis of prostate cancer are helpful for disease diagnosis and treatment. However, current biochemical experimental methods for prostate cancer are both costly and time-consuming, as well as networks based methods for this disease analysis limited by the nature of gene expression profiles for its incomplete, high noise and small sample size. Therefore, we proposed a dual constraint algorithm based on the confidence of one vertices belonging to the community and local modularity, named as NMCOM, to mine the candidate disease modules of prostate cancer in the present work. The NMCOM algorithm is gene expression independent method. It first integrated the concordance scores between the candidate genes and the causative phenotypes, as well as the semantic similarity scores between the candidate genes and the causative genes for prioritizing the candidate genes together, and then the starting node is selected with a sorting strategy. Finally, the candidate modules of prostate cancer are mined with dual constraint produces constructing on the confidence between node and module, as well as local modularity. 18 significant candidate disease gene modules were detected for the enrichment analysis of the obtained modules. Compared with the single scoring sorting methods and random walk with restart, the NMCOM fusion prioritizing strategy achieved a smaller MRR (Mean Rank Ratio) but bigger AUC value. The results are significantly better than other modules-based mining algorithms, and the biological explanations for these mined modules are more significant. More importantly, the NMCOM algorithm can be easily extended to mine any other diseases candidate modules.

Key words prostate cancer, disease module mining, candidate gene prioritization, node-module confidence, local modularity

DOI: 10.16476/j.pibb.2014.0091

*This work was supported by grants from The National Natural Science Foundation of China (91430111, 61473232, 61170134), and the Fundamental Research Funds for the Central Universities-Yong Scholar Development Project from Southwestern University of Finance and Economics (JBK150134).

**Corresponding author.

Tel: 86-29-88431308, E-mail: zhangsw@nwpu.edu.cn

Received: December 14, 2014 Accepted: March 17, 2015

附 录

Table S1 Partial results of GO enrichment analysis for MSMB candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0042981	Regulation of apoptotic process	4.9×10^{-21}
	GO:0043067	Regulation of programmed cell death	6.5×10^{-20}
	GO:0010941	Regulation of cell death	7.2×10^{-20}
	GO:0051789	Response to protein stimulate	1.6×10^{-18}
	GO:0010033	Response to organic substance	5.4×10^{-17}
	GO:0043066	Negative regulation of apoptotic process	1.2×10^{-16}
	GO:0043069	Negative regulation of programmed cell death	1.4×10^{-16}
	GO:0060548	Negative regulation of cell death	1.5×10^{-16}
	GO:0008219	Cell death	4.6×10^{-16}
	GO:0012501	Programmed cell death	3.6×10^{-15}
CC	GO:0031974	Membrane-enclosed lumen	2.6×10^{-14}
	GO:0070013	Intracellular organelle lumen	3.6×10^{-14}
	GO:0005829	Cytosol	5.3×10^{-13}
	GO:0043233	Organelle lumen	1.2×10^{-12}
	GO:0031981	Nuclear lumen	1.0×10^{-10}
	GO:0043228	Non-membrane-bounded organelle	1.4×10^{-9}
	GO:0043232	Intracellular non-membrane-bounded organelle	1.4×10^{-8}
	GO:0030529	Ribonucleoprotein complex	9.4×10^{-8}
MF	GO:0051082	Unfolded protein binding	4.1×10^{-19}
	GO:0000166	Nucleotide binding	2.5×10^{-19}
	GO:0042802	Identical protein binding	3.9×10^{-19}
	GO:0051920	Peroxiredoxin activity	2.1×10^{-17}
	GO:0003723	RNA binding	5.1×10^{-16}
	GO:0019899	Enzyme binding	7.5×10^{-14}
	GO:0004601	Peroxidase activity	1.6×10^{-13}
	GO:0016836	Hydro-lyase activity	5.3×10^{-13}
	GO:0016209	Antioxidant activity	6.7×10^{-3}

Table S2 Partial results of KEGG pathway enrichment analysis for MSMB candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05223	Non-small cell lung cancer	1.1×10^{-16}
hsa05216	Thyroid cancer	1.1×10^{-14}
hsa05214	Glioma	3.8×10^{-13}
hsa00230	Purine metabolism	2.0×10^{-12}
hsa05200	Pathways in cancer	2.2×10^{-12}
hsa03018	RNA degradation	8.3×10^{-11}
hsa04612	Antigen processing and presentation	10.0×10^{-10}

MSMB 模块的 GO 富集条目分析主要集中在细胞死亡生理活动的调控，尤其是负调控，以及多种化合物的绑定和活性方面．在 KEGG 路径上涉及多种其他癌症疾病以及其他多种生理活动．因此预测该模块对前列腺癌细胞死亡生理活动的负调控起决定性作用，并对多种化合物的活性刺激和绑定活动方面起重要作用．

Table S3　Partial results of GO enrichment analysis for PRAC1 candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0035556	Intracellular signal transduction	8.8×10^{-27}
	GO:0010941	Regulation of cell death	5.2×10^{-26}
	GO:0043067	Regulation of programmed cell death	1.5×10^{-25}
	GO:0042981	Regulation of apoptotic process	1.1×10^{-24}
	GO:0010605	Negative regulation of macromolecule metabolic process	1.3×10^{-24}
	GO:0010604	Positive regulation of macromolecule metabolic process	1.4×10^{-23}
	GO:0019220	Regulation of phosphate metabolic process	2.8×10^{-22}
	GO:0051174	Regulation of phosphorus metabolic process	2.8×10^{-21}
	GO:0042325	Regulation of phosphorylation	4.6×10^{-20}
CC	GO:0005654)	Nucleoplasm	5.2×10^{-20}
	GO:0005829	Cytosol	3.6×10^{-19}
	GO:0031981	Nuclear lumen	1.7×10^{-18}
	GO:0070013	Intracellular organelle lumen	9.8×10^{-15}
	GO:0043233	Organelle lumen	4.8×10^{-14}
	GO:0031974	Membrane-enclosed lumen	7.1×10^{-14}
	GO:0044451	Nucleoplasm part	1.4×10^{-13}
	GO:0043235	Receptor complex	3.2×10^{-12}
MF	GO:0004715	Non-membrane spanning protein tyrosine kinase activity	7.8×10^{-19}
	GO:0004672	Protein kinase activity	1.0×10^{-18}
	GO:0019899	Enzyme binding	4.5×10^{-17}
	GO:0019904	Protein domain specific binding	4.6×10^{-16}
	GO:0060090	Molecular adaptor activity	7.6×10^{-15}
	GO:0032553	Ribonucleotide binding	2.5×10^{-14}
	GO:0032555	Purine ribonucleotide binding	2.5×10^{-13}
	GO:0008134	Transcription factor binding	3.1×10^{-13}

Table S4　Partial results of KEGG pathway enrichment analysis for PRAC1 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa04350	TGF-beta signaling pathway	7.6×10^{-17}
hsa04660	T cell receptor signaling pathwa	5.1×10^{-15}
hsa05220	Chronic myeloid leukemia	1.8×10^{-14}
hsa05200	Pathways in cancer	2.2×10^{-12}
hsa04662	B cell receptor signaling pathway	1.5×10^{-11}
hsa05223	Non-small cell lung cancer	9.0×10^{-10}
hsa05214	Glioma	3.8×10^{-9}

PRAC1 模块的 GO 富集条目分析主要集中在多种生理活动的调控，尤其是细胞死亡、特定化合物的新陈代谢过程以及多种酶和核苷酸的活性和绑定方面，在 KEGG 路径上除涉及多种其他癌症疾病外，还涉及某些细胞的特定信号路径.因此预测该模块对前列腺癌细胞死亡、高分子和含磷化合物的新陈代谢过程调控起着重要的决定性作用，并对多种化合物的活性刺激和绑定活动以及特定物质的生物信号传导有重要的影响.

Table S5 Partial results of GO enrichment analysis for PSMA candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0031398	Intracellular signal transduction	5.3×10^{-16}
	GO:0051437	Positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	3.8×10^{-15}
	GO:0031396	Regulation of protein ubiquitination	4.0×10^{-15}
	GO:0051443	Positive regulation of ubiquitin-protein transferase activity	5.3×10^{-15}
	GO:0051439	Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	6.2×10^{-15}
	GO:0051351	Positive regulation of ligase activity	8.2×10^{-15}
	GO:0051438	Regulation of ubiquitin-protein transferase activity	1.6×10^{-14}
	GO:0051340	Regulation of ligase activity	2.4×10^{-14}
	GO:0051436	Negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	1.9×10^{-13}
	GO:0031145	Anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	1.9×10^{-13}
	GO:0051444	Negative regulation of ubiquitin-protein transferase activity	2.5×10^{-13}
	GO:0043161	Proteasome-mediated ubiquitin-dependent protein catabolic process	2.7×10^{-13}
CC	GO:0005839	Proteasome core complex	9.7×10^{-17}
	GO:0000502	Proteasome complex	2.4×10^{-14}
	GO:0019773	Proteasome core complex, alpha-subunit complex	6.3×10^{-12}
	GO:0005829	Cytosol	1.2×10^{-11}
	GO:0005654	Nucleoplasm	2.9×10^{-9}
	GO:0031981	Nuclear lumen	7.0×10^{-8}
	GO:0070013	Intracellular organelle lumen	3.7×10^{-7}
	GO:0043233	Organelle lumen	4.5×10^{-7}
MF	GO:0004298	Threonine-type endopeptidase activity	4.7×10^{-16}
	GO:0070003	Threonine-type peptidase activity	4.7×10^{-16}
	GO:0004175	Endopeptidase activity	1.6×10^{-14}
	GO:0070011	Peptidase activity, acting on L-amino acid peptides	3.4×10^{-13}
	GO:0008233	Peptidase activity	4.8×10^{-12}
	GO:0042802	Identical protein binding	1.3×10^{-11}
	GO:0051879	Hsp90 protein binding	3.0×10^{-10}

Table S6 Partial results of KEGG pathway enrichment analysis for PSMA candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa03050	Proteasome	9.1×10^{-11}
hsa04110	Cell cycle	2.8×10^{-10}
hsa04810	Regulation of actin cytoskeleton	2.3×10^{-9}
hsa00240	Pyrimidine metabolism	2.3×10^{-8}
hsa00230	Purine metabolism	2.0×10^{-7}

PSMA 模块的 GO 富集条目分析主要集中在泛素蛋白质多个生理活动调控方面，如有丝分裂、连接酶活性以及分解代谢活动等，在 KEGG 路径上也与多种化合物的生理活动有关. 因此推测该模块对前列腺癌疾病细胞中泛素蛋白质的特定生理活动，如连接酶活性、分解代谢等的调控有着重要的决定作用，同时对其他特定化合物，如蛋白酶体、细胞骨架等的生理活动有一定的影响.

Table S7　Partial results of GO enrichment analysis for ACPP candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0007167	Enzyme linked receptor protein signaling pathway	2.1×10^{-20}
	GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	1.0×10^{-18}
	GO:0018108	Peptidyl-tyrosine phosphorylation	1.4×10^{-12}
	GO:0018212	Peptidyl-tyrosine modification	2.2×10^{-12}
	GO:0010647	Positive regulation of cell communication	4.9×10^{-12}
	GO:0006796	Phosphate-containing compound metabolic process	1.7×10^{-11}
	GO:0016310	Phosphorylation	2.9×10^{-11}
	GO:0009967	Positive regulation of signal transduction	7.1×10^{-10}
	GO:0006468	Protein phosphorylation	7.6×10^{-10}
CC	GO:0005829	Cytosol	2.5×10^{-13}
	GO:0005886	Plasma membrane	2.7×10^{-9}
	GO:0044459	Plasma membrane part	1.7×10^{-8}
	GO:0016323	Basolateral plasma membrane	5.1×10^{-7}
	GO:0045121	Membrane raft	1.7×10^{-6}
MF	GO:0004713	Protein tyrosine kinase activity	3.4×10^{-13}
	GO:0004715	Non-membrane spanning protein tyrosine kinase activity	1.2×10^{-9}
	GO:0004714	Transmembrane receptor protein tyrosine kinase activity	6.6×10^{-8}
	GO:0004672	Protein kinase activity	3.9×10^{-7}
	GO:0043560	Insulin receptor substrate binding	1.3×10^{-6}
	GO:0005524	ATP binding	1.5×10^{-6}
	GO:0032559	Adenyl ribonucleotide binding	1.9×10^{-6}

Table S8　Partial results of KEGG pathway enrichment analysis forACPP candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa04012	ErbB signaling pathway	7.7×10^{-14}
hsa04510	Focal adhesion	7.0×10^{-13}
hsa04650	Natural killer cell mediated cytotoxicity	6.6×10^{-12}
hsa05215	Prostate cancer	6.8×10^{-12}
hsa05223	Non-small cell lung cancer	1.1×10^{-11}
hsa05214	Glioma	3.8×10^{-10}
hsa05213	Endometrial cancer	1.1×10^{-9}
hsa04660	T cell receptor signaling pathway	3.0×10^{-9}
hsa05200	Pathways in cancer	9.1×10^{-8}

ACPP 模块的 GO 富集条目分析主要集中在蛋白质络氨酸的某些生理活动方面，如致活酶活性、磷酸化作用、修饰和调控等等，以及蛋白质络氨酸的活性刺激和某些化合物绑定方面，在 KEGG 路径上涉及了多个其他癌症疾病以及特定的信号路径，部分蛋白质还直接涉及到了前列腺癌疾病．因此推测该模块对前列腺癌细胞的蛋白质络氨酸多个生理活动，尤其是磷酸化、致活酶活性、修饰和调控等并起着关键的决定作用，同时对某些特定化合物的生物信号传到有一定影响．

Table S9 Partial results of GO enrichment analysis for SPDEF candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0006357	Regulation of transcription from RNA polymerase Ⅱ promoter	9.6×10^{-43}
	GO:0045941	Positive regulation of transcription, DNA-templated	2.2×10^{-42}
	GO:0010628	Positive regulation of gene expression	1.6×10^{-41}
	GO:0051173	Positive regulation of nitrogen compound metabolic process	3.8×10^{-39}
	GO:0045935	Positive regulation of nucleobase-containing compound metabolic process	1.8×10^{-37}
	GO:0051254	Positive regulation of RNA metabolic process	6.7×10^{-36}
	GO:0045893	Positive regulation of transcription, DNA-templated	1.3×10^{-35}
	GO:0010557	Positive regulation of macromolecule biosynthetic process	8.1×10^{-35}
	GO:0031328	Positive regulation of cellular biosynthetic process	4.9×10^{-34}
	GO:0010604	Positive regulation of macromolecule metabolic process	1.1×10^{-33}
CC	GO:0031981	Nuclear lumen	6.7×10^{-47}
	GO:0005654	Nucleoplasm	2.7×10^{-46}
	GO:0043233	Organelle lumen	2.4×10^{-45}
	GO:0070013	Intracellular organelle lumen	2.7×10^{-45}
	GO:0031974	Membrane-enclosed lumen	6.9×10^{-44}
	GO:0044451	Nucleoplasm part	9.4×10^{-43}
	GO:0005667	Transcription factor complex	2.8×10^{-42}
	GO:0005829	Cytosol	3.3×10^{-41}
MF	GO:0030528	Transcription regulator activity	2.0×10^{-53}
	GO:0008134	Transcription factor binding	3.2×10^{-52}
	GO:0016563	Transcription activator activity	7.3×10^{-51}
	GO:0003700	Sequence-specific DNA binding transcription factor activity	9.6×10^{-50}
	GO:0043565	Sequence-specific DNA binding	5.5×10^{-49}
	GO:0003712	Transcription cofactor activity	2.4×10^{-48}

Table S10 Partial results of KEGG pathway enrichment analysis forSPDEF candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	4.0×10^{-38}
hsa04010	MAPK signaling pathway	8.4×10^{-36}
hsa05220	Chronic myeloid leukemia	5.0×10^{-35}
hsa04722	Neurotrophin signaling pathway	6.4×10^{-35}
hsa05215	Prostate cancer	3.3×10^{-32}
hsa04012	ErbB signaling pathway	8.8×10^{-29}
hsa05221	Acute myeloid leukemia	6.1×10^{-28}
hsa05223	Non-small cell lung cancer	8.6×10^{-25}
hsa05216	Thyroid cancer	1.1×10^{-23}

SPDEF 模块的 GO 富集条目分析主要集中在多种生理活动的调控和活性方面，尤其是新陈代谢和生物合成过程的正调控方面，在 KEGG 路径上涉及多种其他多种癌症疾病，部分蛋白质还直接涉及到了前列腺癌疾病．因此推测该模块对前列腺癌疾病细胞的新陈代谢和生物合成过程的正调控起着决定性作用，并对部分生物功能的转录化合物活性起着重要刺激作用.

Table S11　Partial results of GO enrichment analysis for PAWR candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0043067	Regulation of programmed cell death	2.4×10^{-59}
	GO:0010941	Regulation of cell death	2.8×10^{-59}
	GO:0042981	Regulation of apoptotic process	7.7×10^{-59}
	GO:0010604	Positive regulation of macromolecule metabolic process	1.5×10^{-57}
	GO:0043065	Positive regulation of apoptotic process	4.8×10^{-57}
	GO:0051173	Positive regulation of nitrogen compound metabolic process	5.4×10^{-55}
	GO:0043068	Positive regulation of programmed cell death	6.0×10^{-55}
	GO:0010942	Positive regulation of cell death	7.0×10^{-55}
	GO:0010557	Positive regulation of macromolecule biosynthetic process	9.4×10^{-55}
	GO:0031328	Positive regulation of cellular biosynthetic process	1.05×10^{-54}
	GO:0009891	Positive regulation of biosynthetic process	1.8×10^{-54}
CC	GO:0043233	Organelle lumen	6.9×10^{-59}
	GO:0005654	Nucleoplasm	1.7×10^{-58}
	GO:0070013	Intracellular organelle lumen	2.3×10^{-58}
	GO:0031974	Membrane-enclosed lumen	2.4×10^{-58}
	GO:0031981	Nuclear lumen	2.5×10^{-58}
MF	GO:0019899	Enzyme binding	1.3×10^{-50}
	GO:0004672	Protein kinase activity	4.9×10^{-45}
	GO:0004674	Protein serine/threonine kinase activity	2.3×10^{-43}
	GO:0019901	Protein kinase binding	1.5×10^{-42}
	GO:0019900	Kinase binding	7.8×10^{-42}

Table S12　Partial results of KEGG pathway enrichment analysis forPAWR candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	6.3×10^{-41}
hsa04722	Neurotrophin signaling pathway	9.4×10^{-37}
hsa05212	Pancreatic cancer	5.5×10^{-36}
hsa05220	Chronic myeloid leukemia	1.5×10^{-35}
hsa05215	Prostate cancer	5.4×10^{-35}
hsa05210	Colorectal cancer	1.8×10^{-34}
hsa05223	Non-small cell lung cancer	2.4×10^{-31}

PAWR 模块的 GO 富集条目分析主要集中在多种细胞死亡、某些化学过程的正调控以及多种生物酶的绑定和活性方面，在 KEGG 路径上多涉及多种其他癌症疾病，部分蛋白质还直接涉及到前列腺癌疾病．因此推断该模块在前列腺癌相关细胞死亡、某些化学活动，如新陈代谢和合成代谢等方面的调控，尤其是正调控方面起着重要决定作用．

Table S13 Partial results of GO enrichment analysis forPMEPA1 candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0007242	Intracellular signal transduction	5.2×10^{-25}
	GO:0032446	Protein modification by small protein conjugation	1.0×10^{-23}
	GO:0019941	Modification-dependent protein catabolic process	6.1×10^{-23}
	GO:0043632	Modification-dependent macromolecule catabolic process	6.1×10^{-23}
	GO:0016567	Protein ubiquitination	1.5×10^{-22}
	GO:0030163	Protein catabolic process	2.3×10^{-22}
	GO:0051603	Proteolysis involved in cellular protein catabolic process	3.0×10^{-22}
	GO:0044257	Cellular protein catabolic process	4.3×10^{-22}
	GO:0070647	Protein modification by small protein conjugation or removal	4.8×10^{-22}
CC	GO:0005829	Cytosol	2.3×10^{-28}
	GO:0005654	Nucleoplasm	4.8×10^{-22}
	GO:0031981	Nuclear lumen	7.3×10^{-17}
	GO:0070013	Intracellular organelle lumen	5.9×10^{-16}
	GO:0043233	Organelle lumen	8.0×10^{-16}
	GO:0031974	Membrane-enclosed lumen	1.9×10^{-15}
MF	GO:0046332	SMAD binding	7.7×10^{-21}
	GO:0019787	Ubiquitin-like protein transferase activity	1.1×10^{-20}
	GO:0016881	Acid-amino acid ligase activity	3.5×10^{-18}
	GO:0004842	Ubiquitin-protein transferase activity	2.0×10^{-17}
	GO:0003924	GTPase activity	1.4×10^{-16}
	GO:0019899	Enzyme binding	3.3×10^{-16}
	GO:0016879	Ligase activity, forming carbon-nitrogen bonds	9.7×10^{-17}

Table S14 Partial results of KEGG pathway enrichment analysis forPMEPA1 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa04722	Neurotrophin signaling pathway	1.3×10^{-21}
hsa04120	Uiquitin mediated proteolysis	2.7×10^{-16}
hsa05200	Pathways in cancer	1.9×10^{-14}
hsa05220	Conic myeloid leukemia	5.7×10^{-14}
hsa05210	Colorectal cancer	1.5×10^{-13}
hsa05213	Endometrial cancer	4.3×10^{-12}
hsa05214	Glioma	1.7×10^{-11}
hsa05212	Pancreatic cancer	2.8×10^{-11}
hsa05215	Prostate cancer	4.3×10^{-10}

PMEPA1 模块的 GO 富集条目分析主要集中在特定蛋白质的生物活动以及连接酶的活性方面，如修饰、水解、分解代谢和泛素化等，并在 KEGG 路径上涉及多种其他癌症疾病，部分蛋白质还直接涉及到了前列腺癌疾病．因此推测该模块在前列腺癌细胞的特定蛋白质生物活动中起着重要的作用，并对多种连接酶的活性刺激起着决定性作用.

Table S15　Partial results of GO enrichment analysis for PTOV1 candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0006357	Regulation of transcription from RNA polymerase II promoter	3.1×10^{-73}
	GO:0045941	positive regulation of transcription, DNA-templated	1.1×10^{-66}
	GO:0010628	positive regulation of gene expression	2.0×10^{-65}
	GO:0045449	Regulation of transcription, DNA-templated	5.0×10^{-63}
	GO:0045935	positive regulation of nucleobase-containing compound metabolic process	7.1×10^{-63}
	GO:0010557	positive regulation of macromolecule biosynthetic process	5.8×10^{-62}
	GO:0005117	wishful thinking binding	1.8×10^{-61}
	GO:0051252	Regulation of RNA metabolic process	8.3×10^{-61}
	GO:0031328	Regulation of RNA metabolic process	1.7×10^{-59}
CC	GO:0005654	Nucleoplasm	7.4×10^{-84}
	GO:0031981	Nuclear lumen	9.7×10^{-97}
	GO:0070013	Intracellular organelle lumen	8.2×10^{-66}
	GO:0043233	Organelle lumen	7.6×10^{-65}
	GO:0031974	Membrane-enclosed lumen	1.5×10^{-63}
	GO:0044451	Nucleoplasm part	7.0×10^{-59}
	GO:0005829	Cytosol	2.3×10^{-37}
MF	GO:0030528	Transcription regulator activity	1.2×10^{-79}
	GO:0008134	Transcription factor binding	2.7×10^{-65}
	GO:0016563	Transcription activator activity	1.2×10^{-64}
	GO:0003712	Transcription cofactor activity	2.0×10^{-45}
	GO:0003700	Sequence-specific DNA binding transcription factor activity	1.9×10^{-42}
	GO:0003713	Transcription coactivator activity	1.4×10^{-41}

Table S16　Partial results of KEGG pathway enrichment analysis forPTOV1 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	7.3×10^{-26}
hsa04520	Adherens junction	8.0×10^{-15}
hsa04110	Cell cycle	9.0×10^{-15}
hsa05222	Small cell lung cancer	9.3×10^{-13}
hsa04722	Neurotrophin signaling pathway	5.6×10^{-12}
hsa05215	Prostate cancer	3.1×10^{-11}
hsa05220	Chronic myeloid leukemia	7.0×10^{-11}
hsa05212	Pancreatic cancer	5.6×10^{-9}
hsa05210	Colorectal cancer	6.3×10^{-9}
hsa05223	Non-small cell lung cancer	2.1×10^{-8}

PTOV1 模块的 GO 富集条目分析主要集中在特定生理活动的调控和转录化合物的活性方面，如新陈代新过程的调控．在 KEGG 路径上多涉及其他癌症疾病，某些蛋白质更直接涉及到了前列腺癌．因此预测该模块在前列腺癌细胞中对其相应的化合物新陈代谢过程起着重要的调控作用，并对某些转录化合物活性刺激起重要作用．

Table S17 Partial results of GO enrichment analysis for HOXB13 candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0045449	Regulation of transcription	1.2×10^{-98}
	GO:0006355	Regulation of transcription, DNA-templated	3.4×10^{-96}
	GO:0051252	Regulation of RNA metabolic process	4.3×10^{-94}
	GO:0006350	Transcription, DNA-templated	1.3×10^{-84}
	GO:0006357	Regulation of transcription from RNA polymerase Ⅱ promoter	3.1×10^{-83}
	GO:0045941	Positive regulation of transcription	6.5×10^{-83}
	GO:0010628	Positive regulation of gene expression	1.6×10^{-81}
	GO:0051173	Positive regulation of gene expression	8.2×10^{-80}
CC	GO:0031981	Nuclear lumen	3.7×10^{-87}
	GO:0005654	Nucleoplasm	1.8×10^{-83}
	GO:0070013	Intracellular organelle lumen	1.5×10^{-74}
	GO:0043233	Organelle lumen	9.4×10^{-74}
	GO:0031974	Membrane-enclosed lumen	1.5×10^{-72}
	GO:0044451	Nucleoplasm part	8.0×10^{-57}
	GO:0005667	Transcription factor complex	1.5×10^{-46}
MF	GO:0030528	Transcription regulator activity	5.6×10^{-128}
	GO:0003700	Sequence-specific DNA binding transcription factor activity	1.1×10^{-95}
	GO:0003677	DNA binding	6.9×10^{-86}
	GO:0043565	Sequence-specific DNA binding	2.9×10^{-80}
	GO:0016563	Transcription activator activity	6.3×10^{-69}
	GO:0008134	Transcription factor binding	1.2×10^{-59}

Table S18 Partial results of KEGG pathway enrichment analysis forHOXB13 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	7.1×10^{-29}
hsa05215	Prostate cancer	3.9×10^{-17}
hsa05220	Chronic myeloid leukemia	9.5×10^{-17}
hsa05221	Acute myeloid leukemia	5.6×10^{-14}
hsa04110	Cell cycle	5.5×10^{-13}
hsa04330	Notch signaling pathway	4.1×10^{-10}
hsa05210	Colorectal cancer	3.2×10^{-9}
hsa05212	Pancreatic cancer	2.7×10^{-8}
hsa05219	Bladder cancer	2.9×10^{-7}

HOXB13 模块的 GO 富集条目分析主要集中在某些化合物的转录调控以及转录活性和绑定方面，在 KEGG 路径上多涉及其他癌症疾病，部分蛋白质还直接涉及到了前列腺癌疾病．因此预测该模块在前列腺癌细胞中对其相应的化合物的特定转录调控起着重要的决定性作用，并对某些转录化合物的活性和绑定等生物活动起重要作用．

Table S19　Partial results of GO enrichment analysis for STEAP4candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0043067	Regulation of programmed cell death	8.0×10^{-44}
	GO:0006468	Protein phosphorylation	9.3×10^{-44}
	GO:0042981	Regulation of apoptotic process	1.0×10^{-43}
	GO:0010941	Regulation of cell death	1.3×10^{-43}
	GO:0016310	Phosphorylation	2.0×10^{-39}
	GO:0007242	Intracellular signal transduction	3.6×10^{-3}
	GO:0006796	Phosphate-containing compound metabolic process	3.6×10^{-36}
	GO:0006793	Phosphorus metabolic process	3.6×10^{-36}
CC	GO:0005829	Cytosol	2.5×10^{-48}
	GO:0005654	Nucleoplasm	3.2×10^{-15}
	GO:0005856	Cytoskeleton	5.1×10^{-15}
	GO:0015630	Microtubule cytoskeleton	3.8×10^{-13}
	GO:0031981	Nuclear lumen	1.2×10^{-11}
	GO:0044430	Cytoskeletal part	8.0×10^{-11}
MF	GO:0004674	Protein serine/threonine kinase activity	3.2×10^{-40}
	GO:0004672	Protein kinase activity	1.6×10^{-39}
	GO:0005524	ATP binding	3.5×10^{-27}
	GO:0032559	Adenyl ribonucleotide binding	1.9×10^{-26}
	GO:0030554	Adenyl nucleotide binding	3.9×10^{-24}
	GO:0001882	Nucleoside binding	7.9×10^{-24}
	GO:0004674	Protein serine/threonine kinase activity	9.1×10^{-24}

Table S20　Partial results of KEGG pathway enrichment analysis forSTEAP4 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa04722	Neurotrophin signaling pathway	2.3×10^{-22}
hsa04210	Apoptosis	6.4×10^{-19}
hsa05200	Pathways in cancer	4.4×10^{-16}
hsa04620	Toll-like receptor signaling pathway	2.5×10^{-15}
hsa04510	Focal adhesion	2.1×10^{-13}
hsa05215	Prostate cancer	1.9×10^{-12}
hsa05222	Small cell lung cancer	6.9×10^{-11}
hsa05210	Colorectal cancer	2.0×10^{-9}
hsa05212	Pancreatic cancer	7.1×10^{-9}

STEAP4 模块的 GO 富集条目分析主要集中在细胞死亡调控、含磷化合物新陈代谢以及嘌呤核苷酸绑定方面，在 KEGG 路径上多涉及其他癌症疾病，部分蛋白质还直接涉及到了前列腺癌疾病．因此预测该模块对前列腺癌某些细胞的死亡、特定含磷化合物新陈代谢过程起着关键的调控作用，并对某些嘌呤核苷酸的生理活动起决定作用．

Table S21 Partial results of GO enrichment analysis for OR51E2candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0051173	Positive regulation of nitrogen compound metabolic process	7.8×10^{-83}
	GO:0045935	Positive regulation of nucleobase-containing compound metabolic process	3.9×10^{-82}
	GO:0010604	Positive regulation of macromolecule metabolic process	5.9×10^{-82}
	GO:0031328	Positive regulation of cellular biosynthetic process	6.8×10^{-80}
	GO:0045449	Regulation of transcription, DNA-templated	1.1×10^{-79}
	GO:0010628	Positive regulation of gene expression	1.8×10^{-79}
	GO:0045941	Positive regulation of transcription, DNA-templated	2.7×10^{-79}
	GO:0010557	Positive regulation of macromolecule biosynthetic process	6.9×10^{-79}
	GO:0009891	Positive regulation of biosynthetic process	1.3×10^{-78}
CC	GO:0005654	Nucleoplasm	3.6×10^{-91}
	GO:0031981	Nuclear lumen	2.3×10^{-87}
	GO:0043233	Organelle lumen	6.3×10^{-72}
	GO:0070013	Intracellular organelle lumen	3.1×10^{-71}
	GO:0031974	Membrane-enclosed lumen	2.2×10^{-70}
	GO:0044451	Nucleoplasm part	2.8×10^{-64}
	GO:0005667	Transcription factor complex	1.2×10^{-39}
MF	GO:0008134	Transcription factor binding	1.2×10^{-109}
	GO:0030528	Transcription regulator activity	4.6×10^{-94}
	GO:0003712	Transcription cofactor activity	1.6×10^{-74}
	GO:0016563	Transcription activator activity	2.2×10^{-63}
	GO:0003713	Transcription coactivator activity	1.32×10^{-45}
	GO:0035257	Nuclear hormone receptor binding	9.4×10^{-45}
	GO:0016564	Transcription repressor activity	4.8×10^{-44}
	GO:0003700	Sequence-specific DNA binding transcription factor activity	5.2×10^{-41}

Table S22 Partial results of KEGG pathway enrichment analysis for OR51E2 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	3.3×10^{-33}
hsa05215	Prostate cancer	9.5×10^{-21}
hsa04110	Cell cycle	2.4×10^{-18}
hsa04722	Neurotrophin signaling pathway	6.8×10^{-17}
hsa05222	Small cell lung cancer	3.6×10^{-16}
hsa05220	Chronic myeloid leukemia	3.5×10^{-15}
hsa05223	Non-small cell lung cancer	4.1×10^{-13}
hsa05212	Pancreatic cancer	4.3×10^{-13}

OR51E2 模块的 GO 富集条目分析主要集中在某些特定生理活动的调控，尤其是正调控以及各转录化合物活性方面，在 KEGG 路径上多涉及其他癌症疾病，某些蛋白质还直接涉及到了前列腺癌.因此预测该模块主要对前列腺癌细胞的特定生理活动正调控起主要决定性作用，并对某些转录化合物的活性刺激起重要作用.

Table S23 Partial results of GO enrichment analysis for ALKBH3candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	3.6×10^{-49}
	GO:0007167	Enzyme linked receptor protein signaling pathway	9.3×10^{-47}
	GO:0006468	Protein phosphorylation	5.5×10^{-42}
	GO:0006796	Phosphate-containing compound metabolic process	4.1×10^{-41}
	GO:0006793	Phosphorus metabolic process	4.1×10^{-41}
	GO:0016310	Phosphorylation	3.7×10^{-37}
CC	GO:0005829	Cytosol	5.1×10^{-27}
	GO:0044459	Plasma membrane part	1.1×10^{-19}
	GO:0005886	Plasma membrane	5.5×10^{-17}
	GO:0005887	Integral component of plasma membran	2.7×10^{-13}
	GO:0031226	Intrinsic component of plasma membrane	8.0×10^{-13}
MF	GO:0004713	Protein tyrosine kinase activity	9.3×10^{-44}
	GO:0004672	Protein kinase activity	9.5×10^{-42}
	GO:0004714	Transmembrane receptor protein tyrosine kinase activity	5.1×10^{-31}
	GO:0004715	Non-membrane spanning protein tyrosine kinase activity	1.3×10^{-23}

Table S24 Partial results of KEGG pathway enrichment analysis for ALKBH3 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	3.3×10^{-24}
hsa04012	ErbB signaling pathway	6.5×10^{-22}
hsa04722	Neurotrophin signaling pathway	1.4×10^{-21}
hsa04660	T cell receptor signaling pathway	1.0×10^{-20}
hsa05220	Chronic myeloid leukemia	1.8×10^{-19}
hsa04510	Focal adhesion	3.4×10^{-17}

ALKBH3 模块 GO 富集条目分析主要集中在某些含磷化合物的化学反应过程、新陈代谢过程以及多种化合物激活酶活性方面，在 KEGG 路径上除部分涉及其他癌症疾病外，其余大多涉及不同细胞器的信号路径．因此预测该模块对前列腺癌细胞中特定含磷化合物的某些生理活动以及多种激活酶的活性刺激起重要作用，并在多种细胞器的生物信号传导方面起关键作用．

Table S25 Partial results of GO enrichment analysis for PCDH11Ycandidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0006357	Regulation of transcription from RNA polymerase II promoter	2.2×10^{-85}
	GO:0045449	Regulation of transcription, DNA-templated	6.6×10^{-76}
	GO:0010604	Positive regulation of macromolecule metabolic process	1.3×10^{-74}
	GO:0010628	Positive regulation of gene expression	2.2×10^{-72}
	GO:0045941	Positive regulation of transcription, DNA-templated	3.5×10^{-71}
	GO:0051173	Positive regulation of nitrogen compound metabolic process	2.2×10^{-70}
	GO:0045935	Positive regulation of nucleobase-containing compound metabolic process	7.9×10^{-69}
	GO:0031328	Positive regulation of cellular biosynthetic process	1.7×10^{-67}
	GO:0010557	Positive regulation of macromolecule biosynthetic process	1.8×10^{-67}
CC	GO:0031981	Nuclear lumen	2.1×10^{-115}
	GO:0005654	Nucleoplasm	6.6×10^{-113}
	GO:0070013	Intracellular organelle lumen	6.6×10^{-101}
	GO:0043233	Organelle lumen	3.0×10^{-99}
	GO:0031974	Membrane-enclosed lumen	4.5×10^{-98}
	GO:0044451	Nucleoplasm part	4.5×10^{-74}
MF	GO:0030528	Transcription regulator activity	1.6×10^{-97}
	GO:0008134	Transcription factor binding	1.5×10^{-79}
	GO:0003712	Transcription cofactor activity	1.6×10^{-57}
	GO:0016563	Transcription activator activity	1.7×10^{-56}
	GO:0016564	Transcription repressor activity	1.5×10^{-55}
	GO:0003700	Sequence-specific DNA binding transcription factor activity	2.9×10^{-53}

Table S26 Partial results of KEGG pathway enrichment analysis for PCDH11Y candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	3.1×10^{-34}
hsa04110	Cell cycle	3.6×10^{-24}
hsa04010	MAPK signaling pathway	2.2×10^{-22}
hsa04722	Neurotrophin signaling pathway	3.2×10^{-20}
hsa05220	Chronic myeloid leukemia	7.0×10^{-20}
hsa05210	Colorectal cancer	7.6×10^{-15}
hsa05212	Pancreatic cancer	3.3×10^{-14}
hsa05216	Thyroid cancer	1.8×10^{-12}
hsa05215	Prostate cancer	1.4×10^{-11}

PCDH11Y 模块 GO 富集条目分析主要集中在特定生化过程的正调控以及某些转录化合物活性方面，在 KEGG 路径上大多涉及其他癌症疾病，少部分涉及某些化合物的信号路径，部分蛋白质还直接涉及到了前列腺癌疾病．因此预测该模块对前列腺癌细胞的转录、新陈代谢和生物合成等特定的生化过程正调控，以及某些转录化合物活性刺激起着重要作用，并在一定程度上影响某些化合物的信号传导功能.

Table S27 Partial results of GO enrichment analysis for TUSC3candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0043067	Regulation of programmed cell death	1.6×10^{-31}
	GO:0010941	Regulation of cell death	2.3×10^{-31}
	GO:0042981	Regulation of apoptotic process	2.7×10^{-31}
	GO:0042127	Regulation of cell proliferation	6.7×10^{-31}
	GO:0010033	Response to organic substance	4.0×10^{-26}
	GO:0007167	Enzyme linked receptor protein signaling pathway	1.1×10^{-25}
	GO:0008284	Positive regulation of cell proliferation	5.5×10^{-25}
	GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	7.9×10^{-23}
	GO:0008543	Fibroblast growth factor receptor signaling pathway	8.0×10^{-23}
CC	GO:0005829	Cytosol	2.6×10^{-22}
	GO:0043233	Organelle lumen	1.1×10^{-18}
	GO:0031974	Membrane-enclosed lumen	2.9×10^{-18}
	GO:0044421	Extracellular region part	5.6×10^{-18}
	GO:0031981	Nuclear lumen	2.9×10^{-16}
	GO:0005615	Extracellular space	3.5×10^{-15}
	GO:0070013	Intracellular organelle lumen	5.0×10^{-14}
	GO:0005654	Nucleoplasm	1.0×10^{-14}
MF	GO:0019899	Enzyme binding	8.3×10^{-21}
	GO:0005539	Glycosaminoglycan binding	1.4×10^{-16}
	GO:0008201	Heparin binding	3.1×10^{-16}
	GO:0001871	Pattern binding	5.6×10^{-16}
	GO:0030247	Polysaccharide binding	5.6×10^{-16}
	GO:0008083	Growth factor activity	1.4×10^{-14}
	GO:0046982	Protein heterodimerization activity	1.4×10^{-12}

Table S28 Partial results of KEGG pathway enrichment analysis for TUSC3 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	2.6×10^{-30}
hsa04010	MAPK signaling pathway	5.3×10^{-23}
hsa05218	Melanoma	3.1×10^{-17}
hsa04510	Focal adhesion	9.9×10^{-16}
hsa04810	Regulation of actin cytoskeleton	3.7×10^{-15}
hsa05215	Prostate cancer	5.1×10^{-14}
hsa05219	Bladder cancer	6.8×10^{-10}
hsa05212	Pancreatic cancer	7.2×10^{-9}
hsa05210	Colorectal cancer	2.8×10^{-8}

TUSC3 模块的 GO 富集条目分析主要集中在细胞的死亡、增殖等生理活动调控、某些化合物的信号路径以及多种物质的绑定活动方面，在 KEGG 路径上多涉及其他癌症疾病，部分蛋白质还直接参与了前列腺癌的路径．因此推测该模块对前列腺癌细胞的死亡和增殖调控活动以及特定化合物的生物信号传导起重要作用，并对多种物质的绑定活动起决定性作用.

Table S29　Partial results of KEGG pathway enrichment analysis for PDS5B candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0007049	Cell cycle	3.8×10^{-35}
	GO:0022402	Cell cycle process	2.8×10^{-33}
	GO:0022403	Cell cycle phase	2.1×10^{-32}
	GO:0007059	Chromosome segregation	8.2×10^{-32}
	GO:0000279	M phase	8.2×10^{-32}
	GO:0000278	Mitotic cell cycle	2.9×10^{-30}
	GO:0006974	Cellular response to DNA damage stimulus	3.1×10^{-29}
	GO:0006281	DNA repair	4.3×10^{-29}
	GO:0006259	DNA metabolic process	4.9×10^{-28}
	GO:0007067	Mitotic nuclear division	10.0×10^{-28}
CC	GO:0005694	Chromosome	7.3×10^{-38}
	GO:0044427	Chromosomal part	6.2×10^{-33}
	GO:0043228	Non-membrane-bounded organelle	1.9×10^{-30}
	GO:0043232	Intracellular non-membrane-bounded organelle	1.9×10^{-30}
	GO:0005654	Nucleoplasm	1.8×10^{-28}
	GO:0000228	Nuclear chromosome	5.7×10^{-28}
MF	GO:0001882	Nucleoside binding	5.9×10^{-36}
	GO:0043566	Structure-specific DNA binding	1.1×10^{-35}
	GO:0003677	DNA binding	1.2×10^{-35}
	GO:0005524	ATP binding	1.5×10^{-35}
	GO:0032559	Adenylribonucleotide binding	1.7×10^{-35}
	GO:0030554	Adenyl nucleotide binding	3.0×10^{-35}
	GO:0001883	Purine nucleoside binding	3.5×10^{-35}

Table S30　Partial results of KEGG pathway enrichment analysis for PDS5B candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa04110	Cell cycle	1.8×10^{-40}
hsa03030	DNA replication	2.9×10^{-40}
hsa05210	Colorectal cancer	3.9×10^{-39}
hsa04114	Oocyte meiosis	6.3×10^{-38}
hsa03430	Mismatch repair	8.3×10^{-37}

PDS5B 模块的 GO 富集条目分析主要集中在细胞周期及其某些生理活动、DNA 的某些生理功能以及多种化合物，尤其是核苷酸的绑定活动方面，在 KEGG 路径上也多与细胞和 DNA 活动有关，部分蛋白质涉及到癌症疾病．因此推测该模块在前列腺癌细胞周期的多个方面以及 DNA 的损伤响应、修复等功能活动中起着关键作用，并对多种化合物的绑定活动，尤其是嘌呤核苷酸的绑定起决定性作用．

Table S31 Partial results of KEGG pathway enrichment analysis for PAGE4 candidate disease module

	GO ID	GO Term	*P*-value
BP	GO:0006357	Regulation of transcription from RNA polymerase II promoter	6.2×10^{-62}
	GO:0045941	Positive regulation of transcription, DNA-templated	3.6×10^{-61}
	GO:0051173	Positive regulation of nitrogen compound metabolic process	5.4×10^{-61}
	GO:0045935	Positive regulation of nucleobase-containing compound metabolic process	1.7×10^{-60}
	GO:0009891	Positive regulation of biosynthetic process	2.8×10^{-59}
	GO:0031328	Positive regulation of cellular biosynthetic process	5.4×10^{-59}
	GO:0010557	Positive regulation of macromolecule biosynthetic process	2.9×10^{-58}
	GO:0010628	Positive regulation of gene expression	8.1×10^{-61}
CC	GO:0031981	Nuclear lumen	6.8×10^{-52}
	GO:0005654	Nucleoplasm	7.7×10^{-50}
	GO:0070013	Intracellular organelle lumen	1.2×10^{-42}
	GO:0043233	Organelle lumen	1.8×10^{-41}
	GO:0031974	Membrane-enclosed lumen	3.4×10^{-41}
	GO:0044451	Nucleoplasm part	1.3×10^{-39}
	GO:0005667	Transcription factor complex	1.4×10^{-33}
MF	GO:0030528	Transcription regulator activity	2.5×10^{-74}
	GO:0016563	Transcription activator activity	4.5×10^{-59}
	GO:0003700	Sequence-specific DNA binding transcription factor activity	1.6×10^{-51}
	GO:0008134	Transcription factor binding	4.7×10^{-50}
	GO:0043565	Sequence-specific DNA binding	6.8×10^{-41}
	GO:0003712	Transcription cofactor activity	5.6×10^{-33}
	GO:0003713	Transcription coactivator activity	2.2×10^{-29}

Table S32 Partial results of KEGG pathway enrichment analysis for PAGE4 candidate disease module

KEGG ID	KEGG Term	*P*-value
hsa05200	Pathways in cancer	4.3×10^{-20}
hsa05210	Colorectal cancer	9.2×10^{-11}
hsa04310	Wnt signaling pathway	7.0×10^{-10}
hsa05220	Chronic myeloid leukemia	9.1×10^{-10}
hsa05215	Prostate cancer	2.1×10^{-9}
hsa05221	Acute myeloid leukemia	1.2×10^{-8}

PAGE4 模块的 GO 富集条目分析主要集中在多种生理过程的正调控以及多种转录化合物的活性方面，在 KEGG 路径上大多涉及其他癌症疾病，部分蛋白质还直接涉及前列腺癌疾病．因此推测该模块对前列腺癌细胞的某些化学反应过程，如转录、新陈代谢和生物合成等的正调控起主要的决定性作用，并对多个转录化合物的活性刺激起重要作用．

Table S33 Function of top 10 candidate genes

Gene symbol	Gene summary
FGF3	The protein encoded by this gene is a member of the fibroblast growth factor (FGF) family. FGF family membersare involved in a variety of biological processes including embryonic development, cell growth, morphogenesis, tissue repair, tumor growth and invasion. Frequent amplification of this gene has been found in human tumors, which may be important for neoplastic transformation and tumor progression. In addition, there is a close relationship between FGF3 and WT1, which plays an important role in the urogenital system, and literature[1−2]analyze the relationship in the urogenital system about it detailedly.
KLK3	Kallikreins are a subgroup of serine proteases having diverse physiological functions.Growing evidence suggests that many kallikreins are implicated in carcinogenesis and some have potential as novel cancer and other disease biomarkers. Its protein product is a protease present in seminal plasma. Serum level of this protein, called PSA in the clinical setting, is useful in the diagnosis and monitoring of prostatic carcinoma.Literature [3−5] describe its function and analyze the relationship between protate cancer and this gene.
MSMB	The protein encoded by this gene is a member of the immunoglobulin binding factor family. It is synthesized by the epithelial cells of the prostate gland and secreted into the seminal plasma.The expression of the encoded protein is found to be decreased in prostate cancer. Literature[6−7] analyze the relationship between protate cancer and this gene.
PRAC1	This gene is reported to be specifically expressed in prostate, rectum and distal colon. Sequence analysis suggests that it may play a regulatory role in the nucleus. Literature[8−9] analyze its expressionand function in prostate cancer.
PSMA	This gene encodes a type II transmembrane glycoprotein belonging to the M28 peptidase family. The proteinis expressed in a number of tissues such as prostate, central and peripheral nervous system and kidney. In the prostate the protein is up-regulated in cancerous cells and is used as an effective diagnostic and prognostic indicator of prostate cancer. Literature [10−11] describe its expressionand function in prostate cancer.
ACPP	This gene encodes an enzyme that catalyzes the conversion of orthophosphoric monoester to alcohol and orthophosphate. It is synthesized under androgen regulation and is secreted by the epithelial cells of the prostate gland. Literature[12−13] analyze its structure and function in prostate cancer.
SPDEF	The protein encoded by this gene belongs to the ETS family of transcription factors. It is highly expressed in the prostate epithelial cells, and functions as an androgen-independent transactivator of prostate-specific antigen (PSA) promoter. It shows better tumor-association than other cancer-associated molecules, making it a more suitable target for developing specific cancer therapies.Literature [14−15] claim that it has obvious inhibitory effect on the development of prostate cancer.
PAWR	The tumor suppressor represses transcription.This protein is specifically upregulated during apoptosis of prostate cells. Literature[16] analyzes its function detailedly.
PMEPA1	Expression of this gene is induced by androgens and transforming growth factor beta, and the encoded protein suppresses the androgen receptor and transforming growth factor beta signaling pathways though interactions with Smad proteins. Overexpression of this gene may play a role in multiple types of cancer. Literature[17−18] describe its role in prostate cancer.
PTOV1	This gene encodes a protein that was found to be overexpressed in prostateadenocarcinomas.Literature[19] suggests that it is likely to be a biomarker which help to study prostate carcinogenesis and process.

表 S33 中候选基因功能描述支持文献目录

[1] Bruening W, Bardeesy N, Silverman B L, *et al*. Germlineintronic and exonic mutations in the Wilms' tumour gene (WT1) affecting urogenital development. Nature Genetics, 1992, **1**(2): 144-148

[2] Patek C E, Little M H, Fleming S, *et al*. A zinc finger truncation of murine WT1 results in the characteristic urogenital abnormalities of Denys-Drash syndrome. Proceedings of the National Academy of Sciences, 1999, **96**(6): 2931-2936

[3] Du J, Yang Q, Chen X S, *et al*. Changes in fPSA level could discriminate tPSA flare-up from tPSA progression in patients with castration-refractory prostate cancer during the initial phase of docetaxel-based chemotherapy. Cancer Chemotherapy and Pharmacology, 2013, **72**(5): 1055-1061

[4] Larsen S B, Brasso K, Iversen P, *et al*. Baseline prostate-specific antigen measurements and subsequent prostate cancer risk in the Danish Diet, Cancer and Health cohort. European Journal of Cancer, 2013

[5] Park D S, Oh J J, Hong J Y, *et al*. Serum prostate-specific antigen as a predictor of prostate volume and lower urinary tract symptoms in a community-based cohort: a large-scale Korean screening study. Asian Journal of Andrology, 2013

[6] Whitaker H C, Kote-Jarai Z, Ross-Adams H, *et al*. The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. PloS One, 2010, **5**(10): e13363

[7] Waters K M, Stram D O, Le Marchand L, *et al*. A common prostate cancer risk variant 5' of microseminoprotein-beta (MSMB) is a strong predictor of circulating beta-microseminoprotein (MSP) levels in multiple populations. Cancer Epidemiol Biomarkers Prev, 2010, **19**(10): 2639-2646

[8] Liu X F, Olsson P, Wolfgang C D, *et al*. PRAC: A novel small nuclear protein that is specifically expressed in human prostate and colon. The Prostate, 2001, **47**(2): 125-131

[9] Eeles R A, Al Olama A A, Benlloch S, *et al*. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nature Genetics, 2013, **45**(4): 385-391

[10] Lee S J, Lee K R, Yang X, *et al*. NFATc1 with AP-3 site binding specificity mediates gene expression of prostate-specific-membrane-antigen. Journal of Molecular Biology, 2003, **330** (4): 749-760

[11] Leek J, Lench N, Maraj B, *et al*. Prostate-specific membrane antigen: evidence for the existence of a second related human gene. British Journal of Cancer, 1995, **72**(3): 583

[12] Fong L, Brockstedt D, Benike C, *et al*. Dendritic cell-based xenoantigen vaccination for prostate cancer immunotherapy. The Journal of Immunology, 2001, **167**(12): 7150-7156

[13] Sharief F S, Li S S L. Structure of human prostatic acid phosphatase gene. Biochemical and Biophysical Research Communications, 1992, **184**(3): 1468-1476

[14] Sabherwal Y, Mahajan N, Helseth D L, *et al*. PDEF downregulatesstathmin expression in prostate cancer. International Journal of Oncology, 2012, **40**(6): 1889

[15] Turner D P, Findlay V J, Moussa O, *et al*. Mechanisms and functional consequences of PDEF protein expression loss during prostate cancer progression. The Prostate, 2011, **71**(16): 1723-1735

[16] Chaudhry P, Singh M, Parent S, *et al*. Prostate apoptosis response 4 (Par-4), a novel substrate of caspase-3 during apoptosis activation. Molecular and Cellular Biology, 2012, **32**(4): 826-839

[17] Hirokawa Y S, Takagi A, Uchida K, *et al*. High level expression of STAG1/PMEPA1 in an androgen-independent prostate cancer PC3 subclone. Cellular & Molecular Biology Letters, 2007, **12** (3): 370-377

[18] Li H, Xu L L, Masuda K, *et al*. A feedback loop between the androgen receptor and a NEDD4-binding protein, PMEPA1, in prostate cancer cells. Journal of Biological Chemistry, 2008, **283**(43): 28988-28995

[19] Morote J, Fernńndez S, Alaña L, *et al*. PTOV1 expression predicts prostate cancer in men with isolated high-grade prostatic intraepithelial neoplasia in needle biopsy. Clinical Cancer Research, 2008, **14**(9): 2617-2622