The urinary protein biomarker database

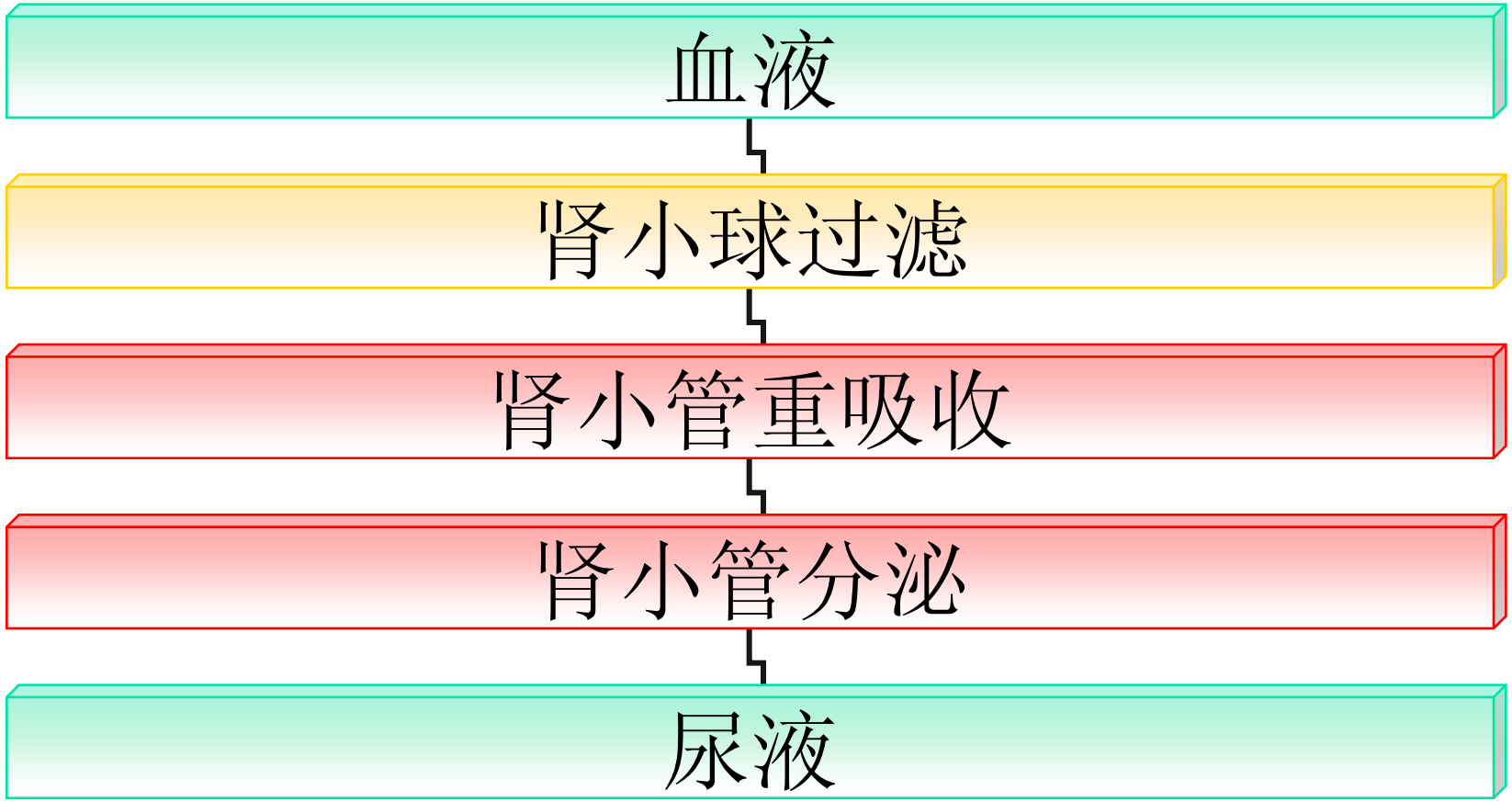
中国医学科学院基础医学研究所

邵晨

尿蛋白作为疾病标志物来源

- 可无创、大量地获得
- 可直接反映肾脏、膀胱等泌尿系统的变化
- 复杂度相对较低，更容易观察其中的低丰度蛋白的变化
- 能够在一定程度上反映血液的变化，因此也可以用来研究全身系统性的疾病

尿液的产生



标志物的疾病特异性问题

- 仅比较疾病和正常样品，不能说明找到的标志物是否是这个疾病特有的。
- 没有实验室为了找到特异标志物可以全面地把所有相关的疾病都包含在研究中。

标志物的可信度问题

- 尿蛋白质组较大的动态变化范围和个体差异的存在、以及当前仪器水平的限制，对同一疾病，不同的实验室的得到的标志物差异很大

我们的工作

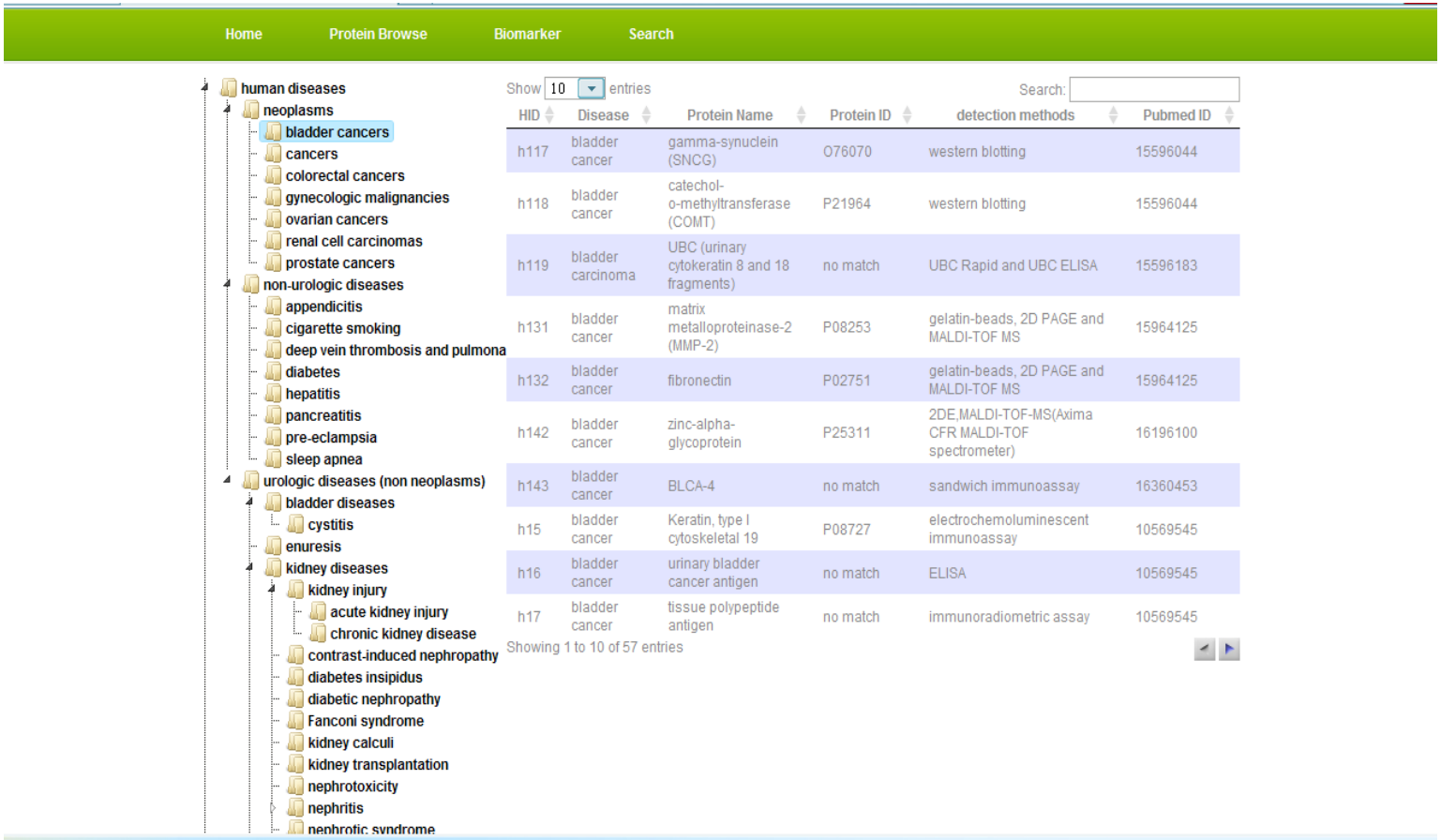
- 建立一个数据库，收录了目前文献报道的所有尿液中疾病标志物的研究成果（不限于蛋白质组研究）
- 内容包括：尿液中鉴定到的所有疾病的候选标志物（人和动物模型）；不包括：CE-MS，SELDI-MS等未鉴定蛋白序列的研究
- 通过数据分析，回答一些生物学问题

文献检索和数据收集

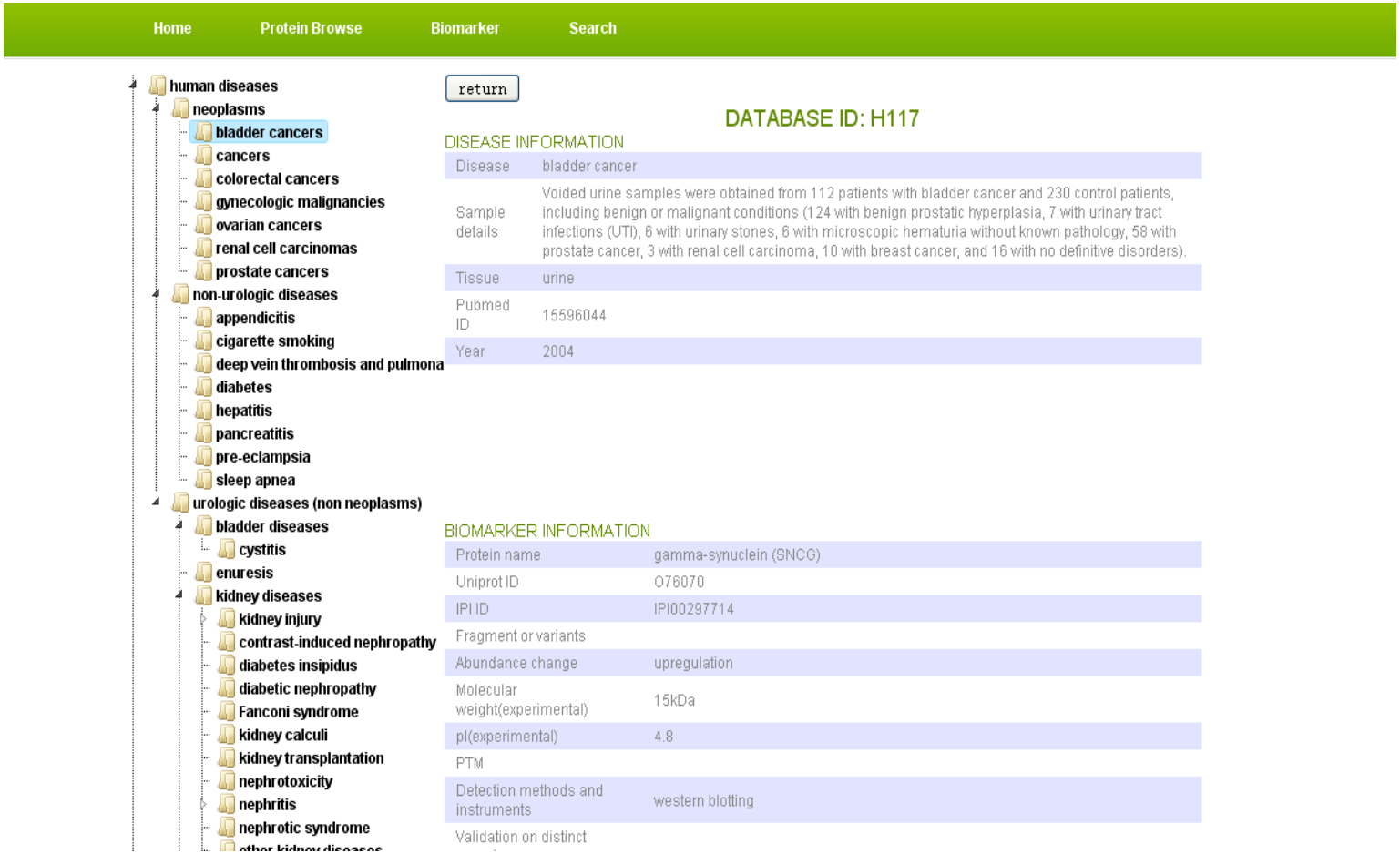
- 2005年以后综述引用的文献
- 直接Pubmed检索文献

- 采用人工阅读文献的方法收集biomarker的信息，以确保其准确性

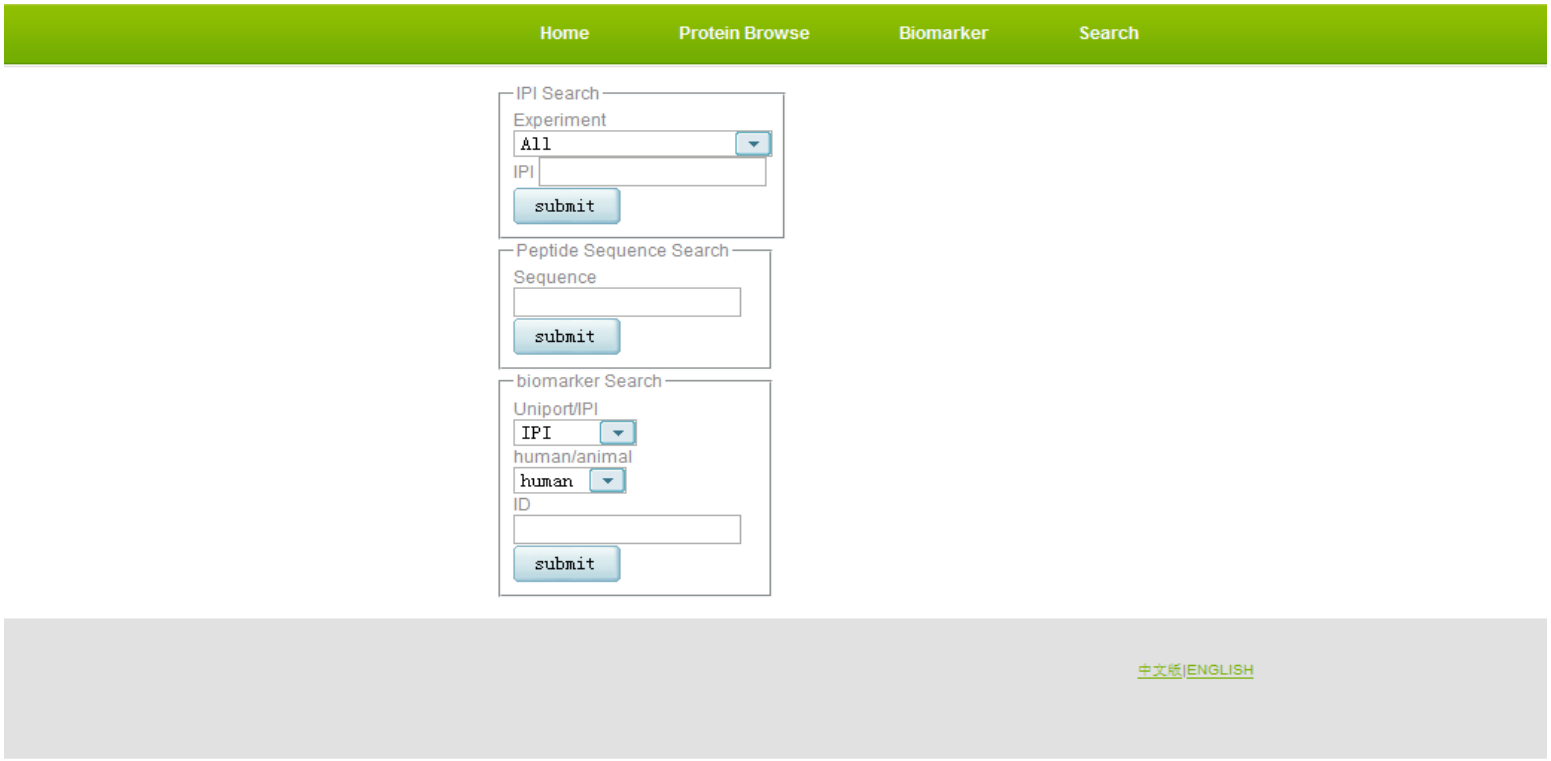
根据疾病分类显示


Home
Protein Browse
Biomarker
Search
human diseases
neoplasms
bladder cancers
cancers
colorectal cancers
gynecologic malignancies
ovarian cancers
renal cell carcinomas
prostate cancers
non-urologic diseases
appendicitis
cigarette smoking
deep vein thrombosis and pulmona
diabetes
hepatitis
pancreatitis
pre-eclampsia
sleep apnea
urologic diseases (non neoplasms)
bladder diseases
cystitis
enuresis
kidney diseases
kidney injury
acute kidney injury
chronic kidney disease
contrast-induced nephropathy
diabetes insipidus
diabetic nephropathy
Fanconi syndrome
kidney calculi
kidney transplantation
nephrotoxicity
nephritis
nephrotic syndrome
Show 10 entries
Search:
HID | Disease | Protein Name | Protein ID | detection methods | Pubmed ID
h117 | bladder cancer | gamma-synuclein (SNCG) | O76070 | western blotting | 15596044
h118 | bladder cancer | catechol-o-methyltransferase (COMT) | P21964 | western blotting | 15596044
h119 | bladder carcinoma | UBC (urinary cytokeratin 8 and 18 fragments) | no match | UBC Rapid and UBC ELISA | 15596183
h131 | bladder cancer | matrix metalloproteinase-2 (MMP-2) | P08253 | gelatin-beads, 2D PAGE and MALDI-TOF MS | 15964125
h132 | bladder cancer | fibronectin | P02751 | gelatin-beads, 2D PAGE and MALDI-TOF MS | 15964125
h142 | bladder cancer | zinc-alpha-glycoprotein | P25311 | 2DE,MALDI-TOF-MS(Axima CFR MALDI-TOF spectrometer) | 16196100
h143 | bladder cancer | BLCA-4 | no match | sandwich immunoassay | 16360453
h15 | bladder cancer | Keratin, type I cytoskeletal 19 | P08727 | electrochemoluminescent immunoassay | 10569545
h16 | bladder cancer | urinary bladder cancer antigen | no match | ELISA | 10569545
h17 | bladder cancer | tissue polypeptide antigen | no match | immunoradiometric assay | 10569545
Showing 1 to 10 of 57 entries

标志物的具体信息


Home
Protein Browse
Biomarker
Search
human diseases
neoplasms
bladder cancers
cancers
colorectal cancers
gynecologic malignancies
ovarian cancers
renal cell carcinomas
prostate cancers
non-urologic diseases
appendicitis
cigarette smoking
deep vein thrombosis and pulmona
diabetes
hepatitis
pancreatitis
pre-eclampsia
sleep apnea
urologic diseases (non neoplasms)
bladder diseases
cystitis
enuresis
kidney diseases
kidney injury
contrast-induced nephropathy
diabetes insipidus
diabetic nephropathy
Fanconi syndrome
kidney calculi
kidney transplantation
nephrotoxicity
nephritis
nephrotic syndrome
return
DATABASE ID: H117
DISEASE INFORMATION
Disease
bladder cancer
Sample details
Voided urine samples were obtained from 112 patients with bladder cancer and 230 control patients, including benign or malignant conditions (124 with benign prostatic hyperplasia, 7 with urinary tract infections (UTI), 6 with urinary stones, 6 with microscopic hematuria without known pathology, 58 with prostate cancer, 3 with renal cell carcinoma, 10 with breast cancer, and 16 with no definitive disorders).
Tissue
urine
Pubmed ID
15596044
Year
2004
BIOMARKER INFORMATION
Protein name
gamma-synuclein (SNCG)
Uniprot ID
O76070
IPI ID
IPI00297714
Fragment or variants
Abundance change
upregulation
Molecular weight(experimental)
15kDa
pI(experimental)
4.8
PTM
Detection methods and instruments
western blotting
Validation on distinct

检索功能


Home
Protein Browse
Biomarker
Search
IPI Search
Experiment
All
IPI
submit
Peptide Sequence Search
Sequence
submit
biomarker Search
Uniport/IPI
IPI
human/animal
human
ID
submit
中文版|ENGLISH

Data statistics-human

- 131 papers （40 proteomics studies）
- 273 proteins
- 40 diseases
- 443 records （including 5 negative results）
- 387 biomarkers

Data statistics-animal model

- 21 papers （14 on rat,7 on mouse）
- 169 proteins
- 19 diseases
- 230 records

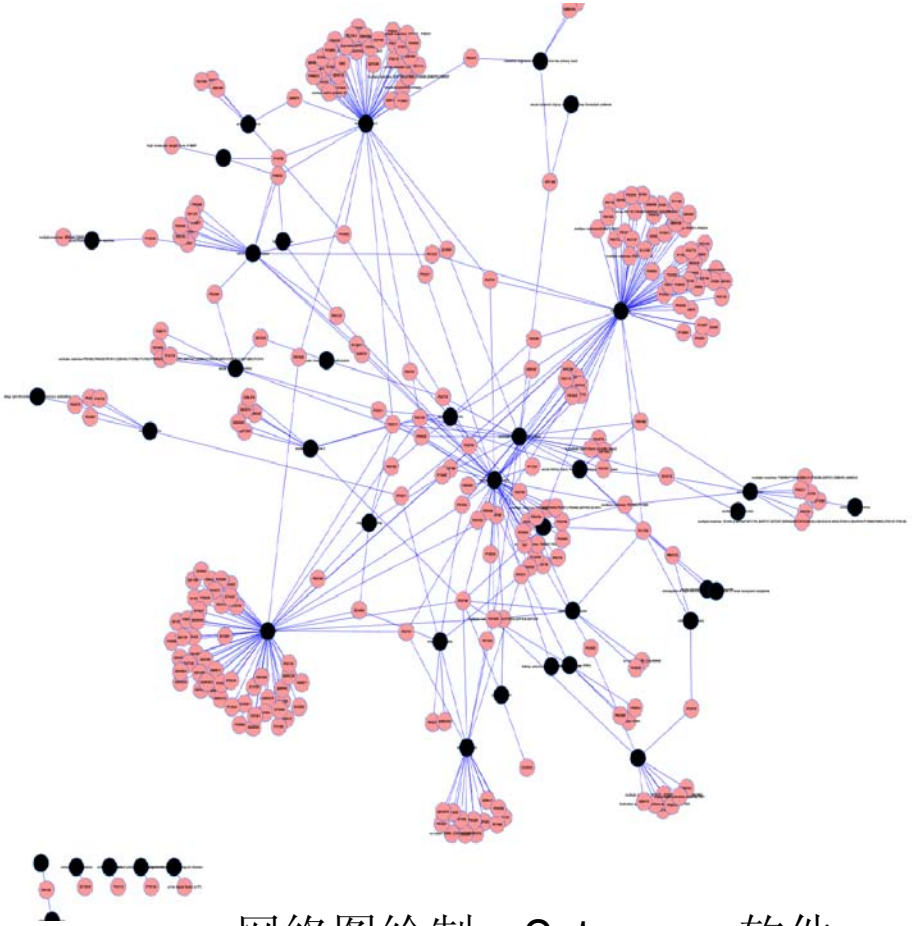
- 约1/3的动物蛋白具有人的直系同源蛋白

network analysis of human data

Human disease-protein network

40 disease nodes
272 protein nodes
387 edges

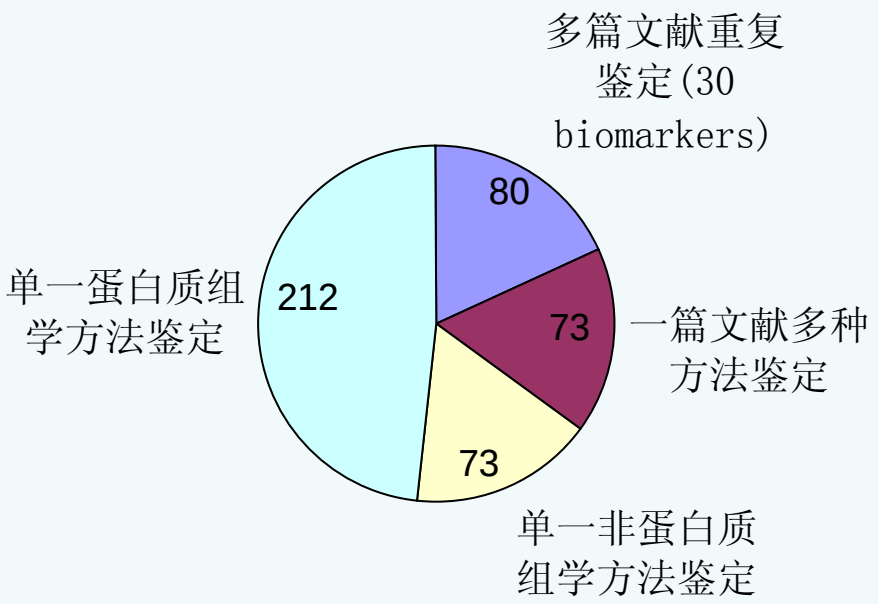
Ave.edges per disease node
9.7
Ave edges per biomarker node
1.4
Cluster coefficient
0.025
Number of compartments
6



网络图绘制：Cytoscape软件

Data confidence-human

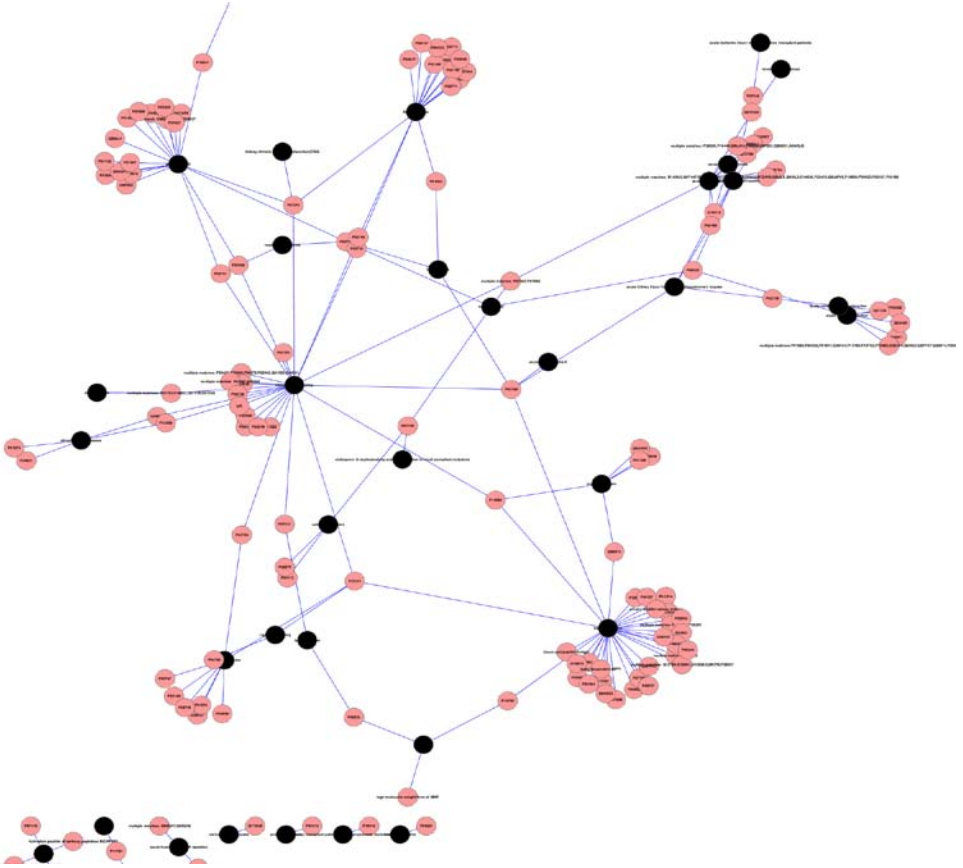
438 human records



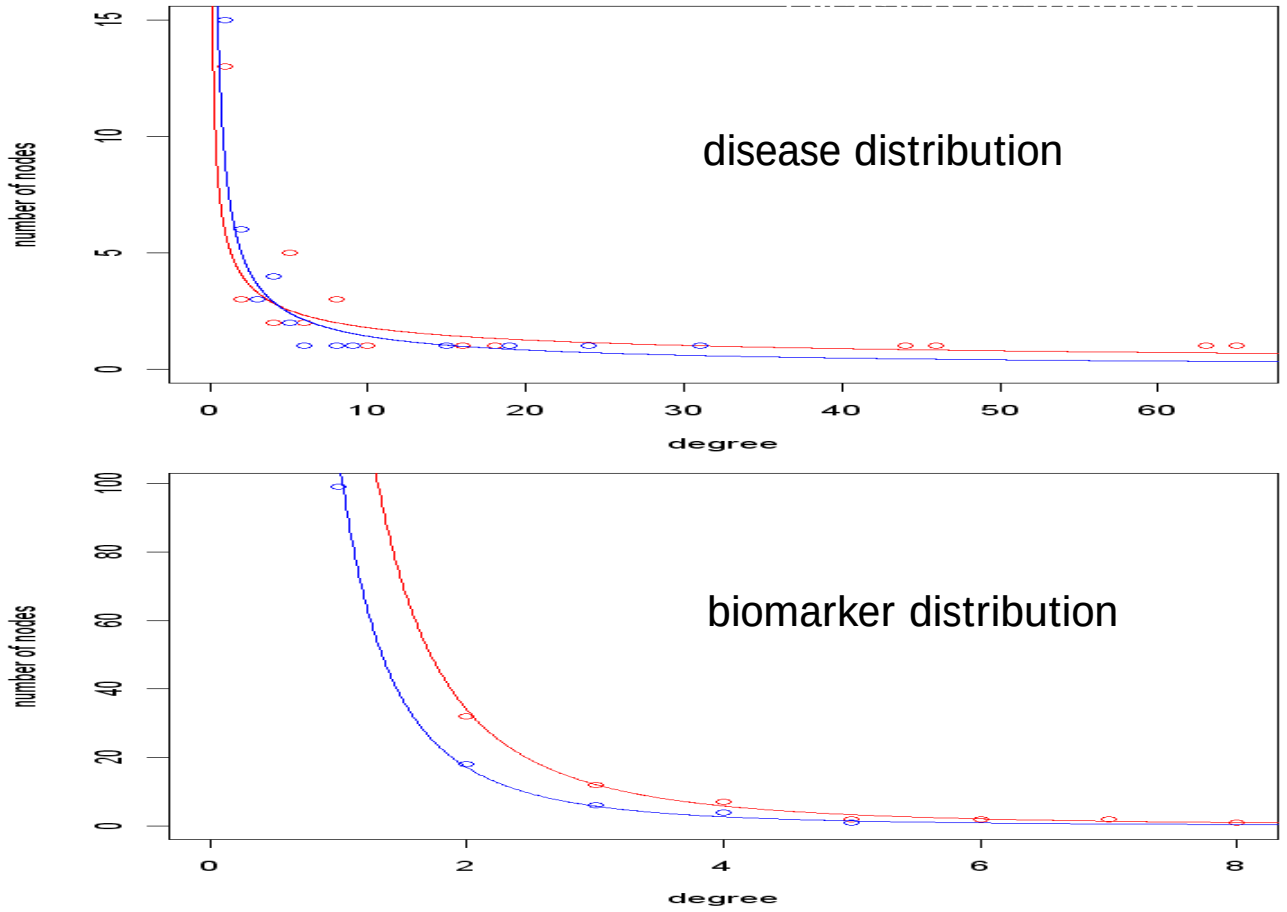
Human_core dataset

37 disease nodes
128 protein nodes
174 edges

Ave.edges per disease node
4.7
Ave edges per biomarker node
1.4
Cluster coefficient
0.027
Number of compartments
10

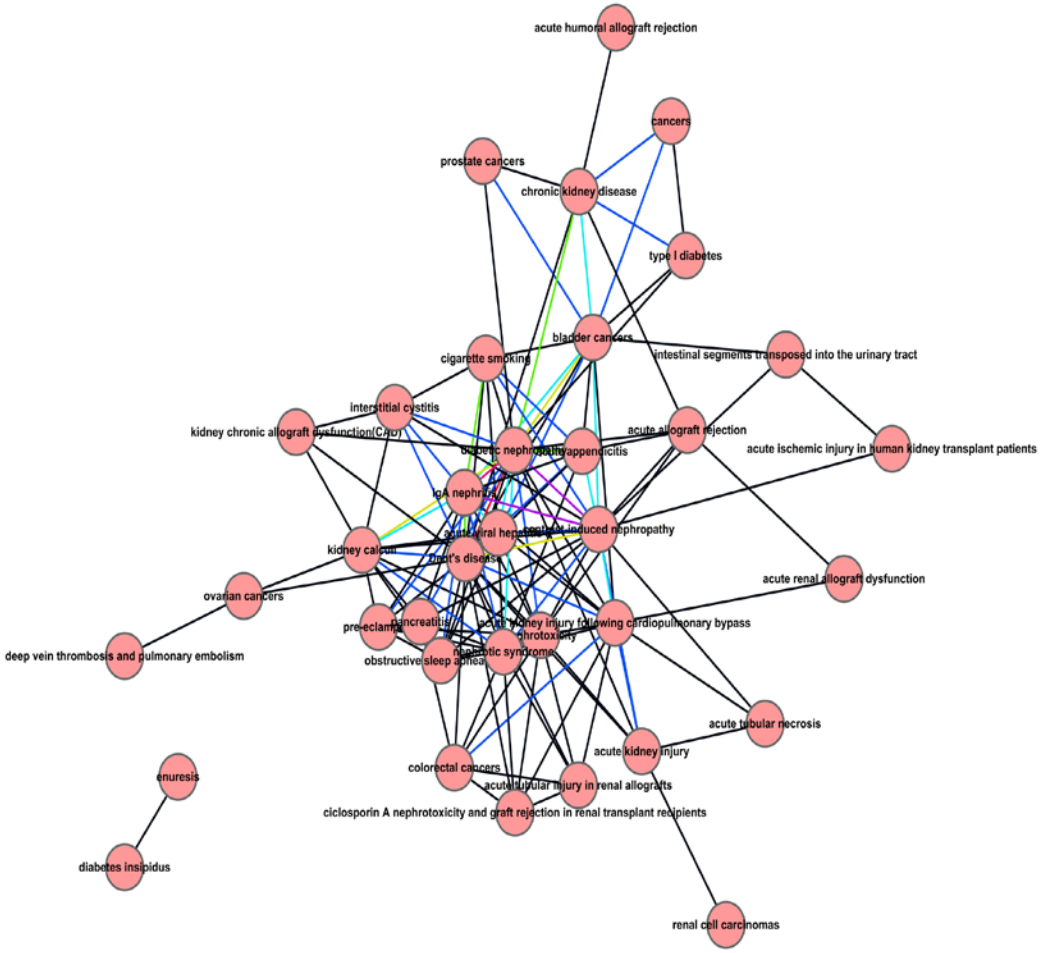


Node distribution



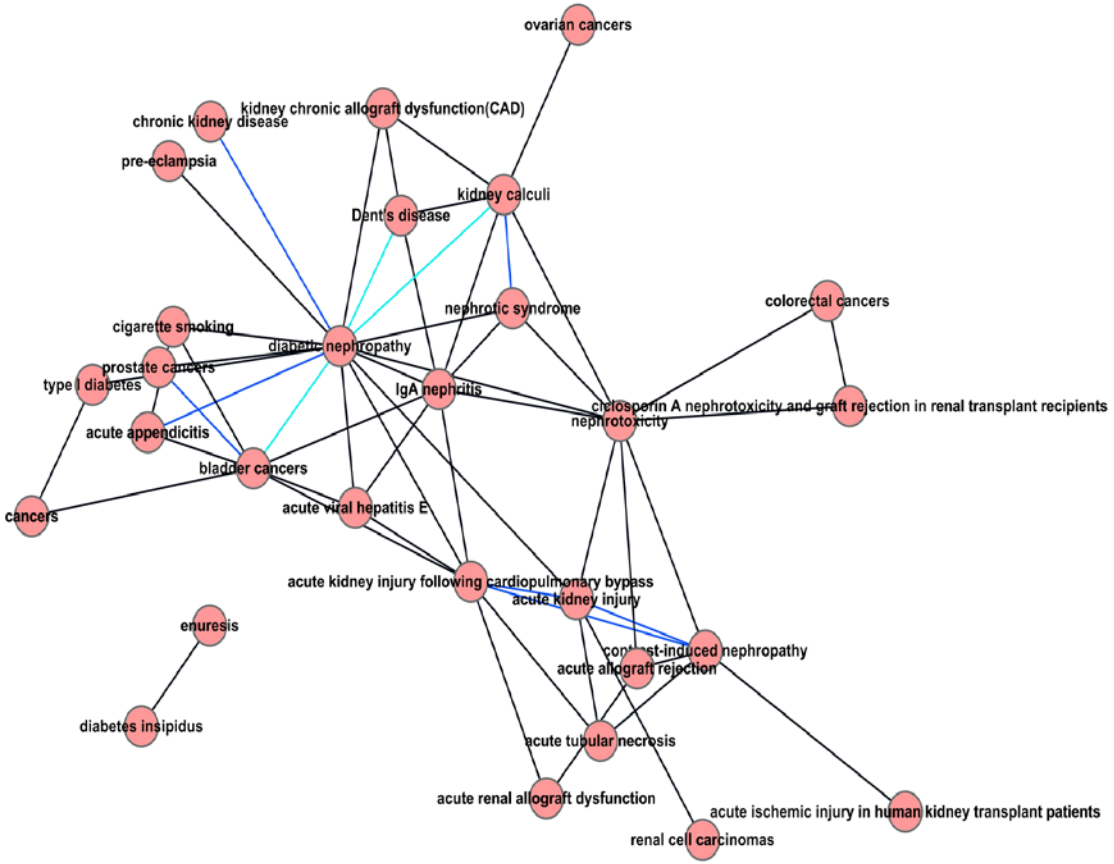
Legend:
Red-human
Blue-core

Human disease-disease network



36 nodes
137 edges
Clustering coefficient
0.59
Ave. degree
7.6
Mean shortest path
2.02

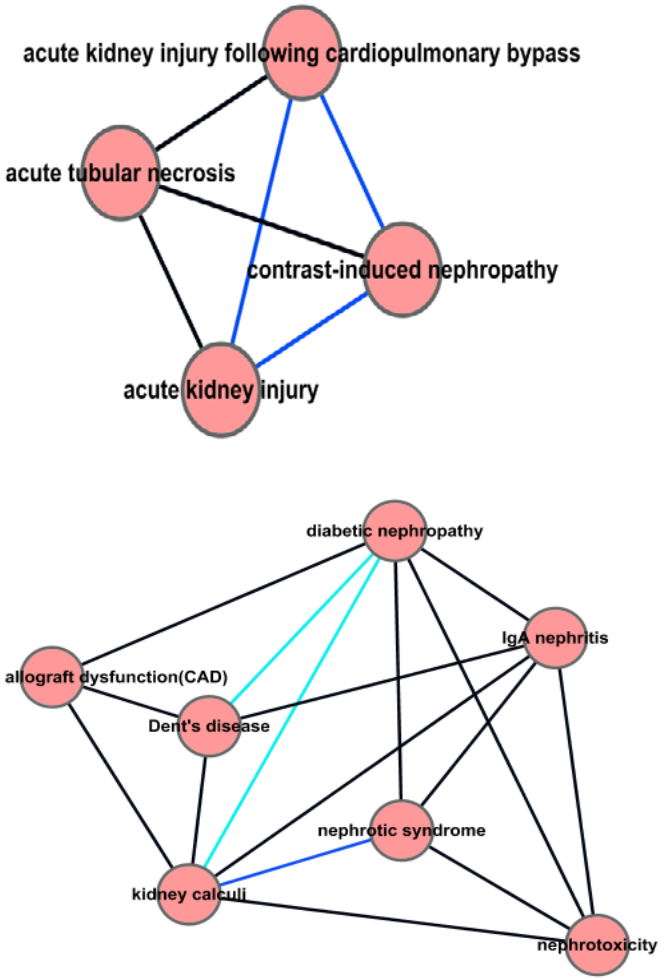
Human_core dataset



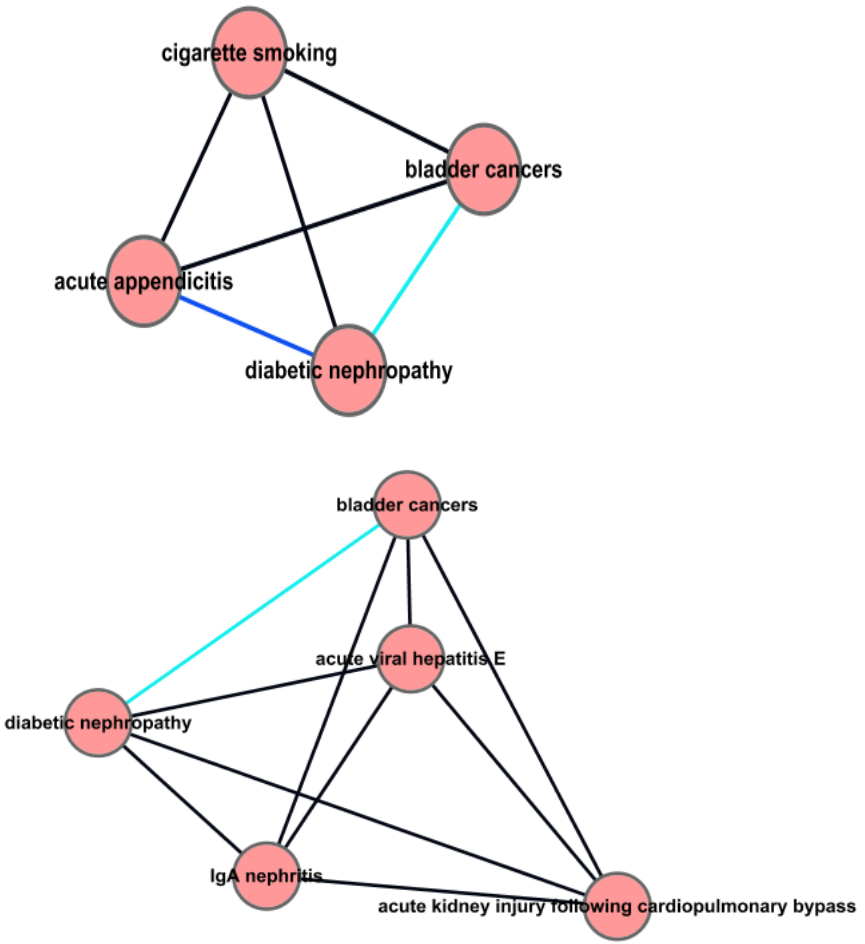
29 nodes
57 edges
Clustering coefficient
0.23
Ave. degree
3.9
Mean shortest path
1.64

Densely connected communities in the human_core network (4-clique clustering by CFinder)

左：相似的疾病聚类在一起



右：不同的疾病聚类在一起



通过建立数据库，回答疾病标志物研究中的一些问题：

问题一：蛋白质组学研究的重复性如何？

	疾病	文献	文献两两组合	共同鉴定的标志物	
human	8	26	64	6对文献，10个标志物	平均两次实验共鉴定到41个候选标志物，才有1个共同鉴定的
animal	2	4	2	1对文献，10个标志物	两疾病的标志物数分别为（11，24）和（16，2）

不同样品处理和蛋白鉴定方法对病人样品的研究的重合率极低

问题二：动物模型 vs 病人样本

- 实验动物间的个体差异较小，动物模型可以方便的监测病程变化，进行各种干预，具有不可替代的优点。

- 从进化角度讲，uniprot数据库中，仅有1/3左右的大鼠蛋白具有人的直系同源蛋白。
- 从蛋白序列相似性来看，也仅有1/3左右的大鼠蛋白与人类蛋白的相似性>90%。

问题二：动物模型 vs 病人样本

	human	animal	Co-identified biomarker
Interstitial cystitis	5	11	1
Diabetic nephrology	41	2	1
Acute kidney injury	12	13	0

考虑到同是人类样品，两次蛋白质组实验的重合率都极低，
没有证据说明动物模型和病人样品之间的差距很大

问题三：标志物的疾病特异性

- 272个人类蛋白中57个蛋白至少与两个以上疾病相关，其中34(45)个是血浆蛋白

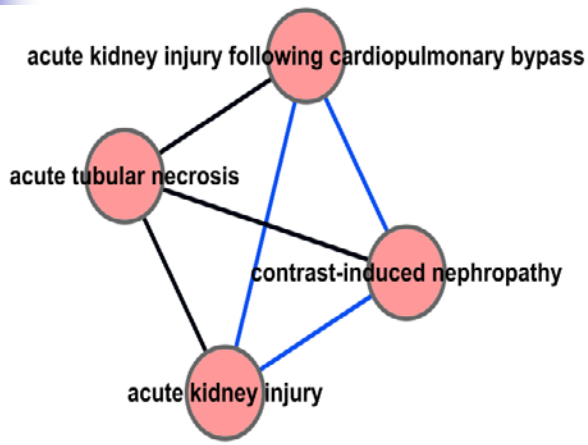
不具备疾病特异性的标志物（前三位）

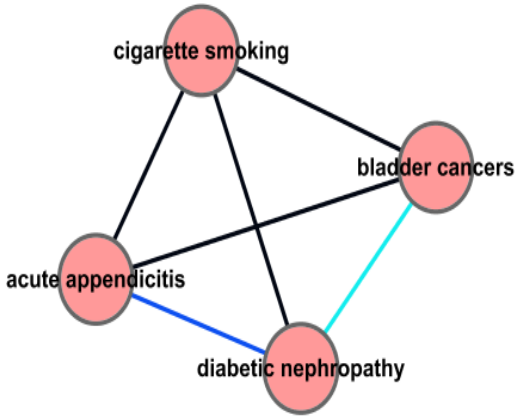
protein	Plasma	Related diseases
albumin	yes	acute viral hepatitis E,obstructive sleep apnea (OSA),pre-eclampsia, diabetic nephropathy,IgA nephritis,kidney stones,nephrotic syndrome,nephrotoxicity
Zinc-alpha-2-glycoprotein	yes	acute appendicitis,acute viral hepatitis E,bladder cancers,cigarette smoking, contrast-induced nephropathy,diabetic nephropathy,IgA nephritis
Alpha-1-microglobulin	yes	acute viral hepatitis E,bladder cancers, contrast-induced nephropathy,diabetic nephropathy,IgA nephritis, acute kidney injury following cardiopulmonary bypass

Densely connected communities in the human_core network (4-clique clustering by CFinder)

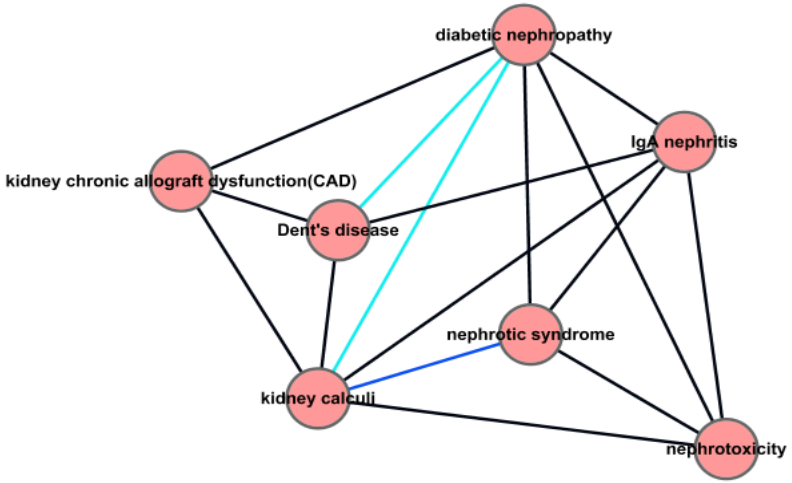
左：相似的疾病聚类在一起

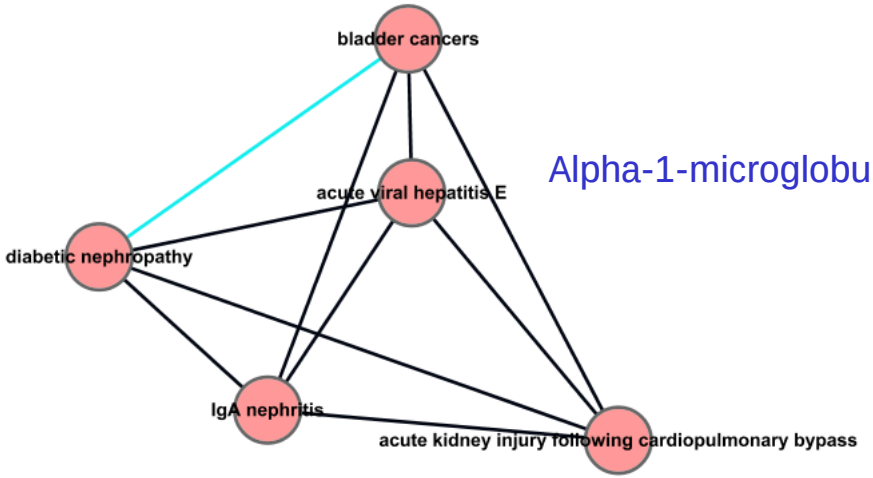
右：不同的疾病聚类在一起





Zinc-alpha-2-glycoprotein





Alpha-1-microglobulin

问题三：标志物的疾病特异性

未鉴定到特异性标志物的疾病

disease	papers	biomarkers
ciclosporin A nephrotoxicity and graft rejection in renal transplant recipients	1	1
kidney chronic allograft dysfunction(CAD)	1	1
nephrotoxicity	2	4
deep vein thrombosis and pulmonary embolism	1	1
diabetes insipidus	1	1
type I diabetes	2	2
cigarette smoking	1	5
renal cell carcinomas	1	1
colorectal cancers	1	3
enuresis	1	1

问题三：标志物的疾病特异性

多篇文献重复鉴定的疾病特异性标志物

protein name	plasma	papers	disease
prostate-specific antigen (PSA)		4	prostate cancers
trypsinogen activation peptide(TAP)		6	pancreatitis
Trypsinogen-2		5	pancreatitis
survivin		2	bladder cancers
nuclear matrix protein 22		6	bladder cancers
UBC		2	bladder cancers
Keratin, type I cytoskeletal 19		5	bladder cancers
BLCA-4		3	bladder cancers
Haptoglobin precursor		2	diabetic nephropathy
CTGF		2	diabetic nephropathy
Ceruloplasmin	yes	2	diabetic nephropathy

问题四：发现疾病/标志物参与的生命过程

- 通过比较标志物关联的多种疾病，发现疾病/标志物参与的共同的生命过程
- 下表中的4个蛋白，相关的疾病都是慢性肾脏病和癌症，可推测它们涉及的是衰老的过程

protein	disease
vascular endothelial growth factor (VEGF)	bladder cancers,chronic kidney disease, diabetic nephropathy,prostate cancers
transforming growth factor beta1	chronic kidney disease,diabetic nephropathy, type I diabetes
matrix metalloproteinase-2 (MMP-2)	bladder cancers,cancers, chronic kidney disease,type I diabetes
matrix metalloproteinase-9 (MMP-9)	bladder cancers,cancers,chronic kidney disease

问题五：肾脏病的鉴别诊断

数据库中只在肾小管疾病中有变化的蛋白

Hemopexin	Apolipoprotein A1
N-acetyl-beta-(D)-glucosaminidase	Plasma retinol-binding protein precursor
Polymeric-immunoglobulin receptor	Apolipoprotein A-IV
TNF	Complement factor B
Vitamin D binding protein (VDBP)	Polymeric-immunoglobulin receptor

另两个普遍承认的肾小管损伤标志物还和其它的疾病相关

beta2-microglobulin	colorectal cancer
uromodulin	interstitial cystitis

问题五：肾脏病的鉴别诊断

多种肾小球疾病的鉴别诊断（animal model）

	total biomarkers	specific biomarkers
focal segmental glomerulosclerosis	58	46
membranoproliferative glomerulonephritis	4	4
mesangioproliferative glomerulonephritis	34	24
nephrotoxic serum nephritis	2	1
membranous nephritis	46	33

总结——尿蛋白标志物研究现状

- 蛋白质组实验的重复性很低
- 已经发现了一些特异的标志物，正在向临床应用转化
- 绝大部分标志物都只被鉴定一次，可信度不高
- 很多高丰度的血浆蛋白都不是特异性的标志物
- 一些疾病研究较少，肾小球疾病缺乏病人样本

数据库的生命力

- 以上的一切分析都是建立在现有数据的结果上的
- 随着质谱技术和蛋白质组学方法的发展，数据会被逐渐更新，分析的结果也会随之产生变化
- 这种利用数据库研究生物标志物特异性的方法，利用数据库研究疾病关联的生物过程的方法将保持着生命力。

对未来研究的展望

- 建立统一的样品处理、鉴定、定量平台，使不同实验室间的数据具有可比性
- 找到更多的高可信度的疾病特异标志物
- 在大样本集上验证候选标志物的特异性和灵敏度，计算表达量随病程变化的曲线，制定诊断的标准

谢谢