分子对接研究

## 研究思路

网络药理学的研究，要遵循一定的研究的思路与方法。本实验的研究步骤是:

首先，通过分子对接技术确定复方成分与AS 靶点的作用关系，构建成分-靶点网络; 然后，计算复方成分分子指纹及Tanimoto系数，得出各成分间相似度，构建成分-成分网络; 再以成分-靶点网络为中心，成分网络和靶点作用网络为子网络，构建二阶异质网络; 最后，应用模块分析技术分解二阶异质网络，并对其中的主要功能模块进行检测，通过GO 注释和kyoto encyclopedia of genes and genomes( KEGG) 通路分析研究复方丹参滴丸治疗AS 的物质基础和作用机制，为同类研究提供方法学参考。

## 分子对接分类

分子对接方法根据不同的简化程度可以划分为三类:

刚性对接：研究体系的构象不发生变化。刚性对接指在对接过程中，受体和配体的构象不发生变化，适合研究比较大的体系如蛋白-蛋白之间以及蛋白-核酸之间，计算简单，主要考虑对象之间的契合程度。

半柔性对接：研究体系尤其是配体的构象允许在一定的范围内变化。半柔性对接常用于小分子和大分子的对接，在对接过程中，小分子的构象可以在一定范围内变化，但大分子是刚性的。这样既可以在一定程度上考察柔性的影响，又能保持较高的计算效率。在药物设计和虚拟筛选过程中一般采用半柔性的分子对接方法。

柔性对接：研究体系的构象基本上是可以发生变化的。柔性对接方法一般用于精确研究分子之间的识别情况，由于允许对接体系的构象变化，可以提高对接准确性但耗时较长。

本实验使用的方法是半柔性对接。

## 分子对接原理

进行分子对接, 也就是从已知结构的受体( 靶蛋白或活性位点) 和配体出发, 通过化学计量学方法模拟分子的几何结构和分子间作用力来进行分子间相互作用识别并预测受体􀀁配体复合物结构的方法称为分子对接。

## 研究工具

现在研究的工具是PyMOL和AutoDock,ChenOffice等相关软件,前两者是用来进行分子对接的软件,最后一个是用来处理分子的结构相关的软件。现在可以对下面进行对接的软件进行分类。可视化软件：PyMOL、

### PyMOL

这是一个分子可视化的软件。

### AutoDock

这个是一个分子对接的软件，但是这个是对接软件出现的问题是比较麻烦，合适的方式是使用基于这个软件开发的SystemDock的在线对接软件，可以实现这个软件的相关功能。

### ChenOffice

这个软件主要是用来进行绘制2D和3D的化合物的结构。使用这个工具可以把对应的小分子化合物的。

### HTdocking

网络药理学预测化合物靶点信息的方法所用到的预测靶点的在线软件，该软件可以可以直接进行分子的对接。并且这个软件使用的是反向对接的技术。

### PharmMapper

网络药理学预测化合物靶点信息的方法所用到的预测靶点的在线软件。但是这个软件主要是使用反向找靶就是，实际上是根据药品的成分反向找到靶点的

### SystemDock

这个软件是在线的分子对接软件，该软件要是想对接，首先要做的就是进行登录账号，这个是一个关键的操作，必须登录否则会导致step3和step4步骤报错，所以登录到账号上，最后做的就是按照相关的教程进行下一步操作，进行分子对接操作。

下面是对接的效果：

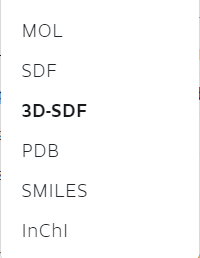


## 研究数据库

进行分子对接研究需要一些原始信息的准备，原始信息的准备有两部分，一个是药品成分信息的准备，另一个是靶点信息的准备。药品成分信息的准备有TCMSP，TCMID，上海有机所的成分数据库等相关数据库。靶点数据库包括NCBI等相关的数据库。

### DrugBank

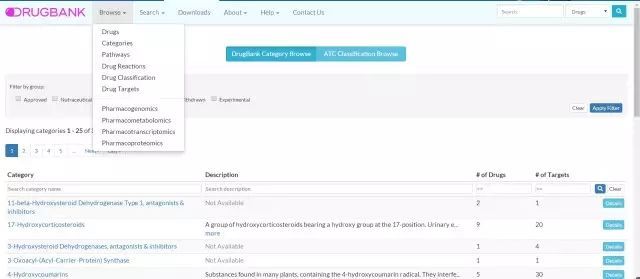
可以搜索小分子的结构等相关的信息。分子的格式有下面几种:



可以从这个数据库中查看小分子的结构等信息。

DrugBank数据库是唯一将详细的药品数据（即化学、药理学和制药）与综合药物靶点信息（即序列、结构和作用通路）相结合的“生物信息学和化学信息学”资源。DrugBank由加拿大卫生研究院，亚伯达省创新-健康解决方案和代谢组学创新中心（TMIC）提供支持，该中心是国家资助的研究以及支持广泛的尖端技术代谢组学研究的核心 。DrugBank数据库查询包含以下信息：药品类型、药品简介、化学结构、药品成分、临床试验、药物靶点、酶、转运体、载体、药品图片、批准情况、批准的处方药、国外上市商品名、药物相互作用、制造商、包装商等。

DrugBank4.0中已增加由世界卫生组织药物统计方法整合中心（WHO Collaborating Centre for Drug Statistics Methodology）定义的解剖学治疗学及化学（ATC）分类系统提供的信息。DrugBank列出了分类系统的前四个级别（解剖学组，治疗亚组，药理学亚组和化学亚组），并且每种药物都包含“分类卡（Category Card）”的链接。每个解剖学组的Category Card包含治疗亚组的列表，可以展开显示药理亚组，进一步展开显示化学亚组，化学亚组可展开最终显示第五级——化学物质或药物名称。分类浏览（Category browse）见下图。



目前，数据库已更新为5.0版本。根据DrugBank数据库统计，至今数据库中已包含9591条药品记录，其中包括2037个FDA批准的小分子药品，241个FDA批准的生物技术（蛋白质/肽）药品，96个保健品，以及6000多个实验药品。此外，4661非冗余蛋白序列（即药物靶标/酶/转运体/载体）与以上药品条目信息相链接。每个Drug Card条目包含超过200个数据字段，其中一半的信息为药品化学数据，另一半则是药物靶点或蛋白质数据。DrugBank向公众提供免费的资源，使用及重新分配全部或部分数据用于商业用途（包括内部使用）需要许可证。DrugBank要求用户如果下载并使用相关数据在其任何出版物中说明引用DrugBank内容[2]。

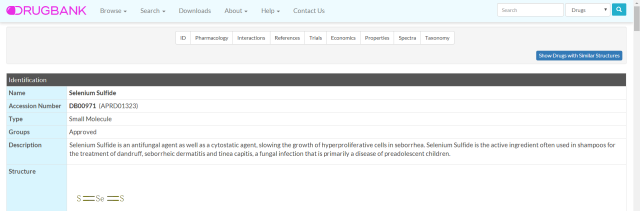
DrugBank数据库每天更新一次。DrugBank下载量每季度发布一次。如果用户需要更频繁的更新，可以通过info@omx.io与DrugBank联系。

2 .DrugBank数据库功能

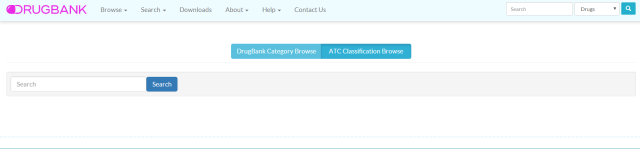
开发DrugBank数据库旨在缩小临床药品资源和化学药品数据库之间“深度与广度”的差距。DrugBank于2006年首次发布，作为全面并完全可搜索的计算机药物资源，将药物分子（包括生物技术药物）的序列、结构和机制数据与其药物靶点的序列、结构和机制数据连接。作为临床导向的药品百科全书，DrugBank能够提供关于药品，药品靶点和药物作用的生物或生理结果的详细、最新、定量分析或分子量的信息。作为化学导向的药品数据库，DrugBank能够提供许多内置的工具，用于查看、排序、搜索和提取文本、图像、序列或结构数据。自数据库首次发布信息起，DrugBank已被广泛应用于计算机检索药物、药物“复原”、计算机检索药物结构数据、药物对接或筛选、药物代谢预测、药物靶点预测和一般制药教育[3]。

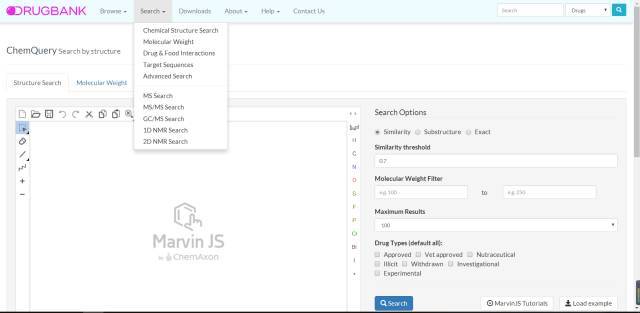
3 .如何检索DrugBank数据库？

用户可以通过多种方式查询DrugBank数据库，最简单方式就是使用普通文本对数据库的所有字段进行检索，通过浏览选项可以以表格的形式浏览该数据库的内容。Drug Card选项允许用户查看特定药物的所有信息。简单文本查询支持数据库整个文本组件的一般文本查询。点击DrugBank导航面板右上方的浏览按钮,接下来则生成DrugBank数据库检索内容的表格概要，例如查询二硫化硒（Selenium Sulfide）：

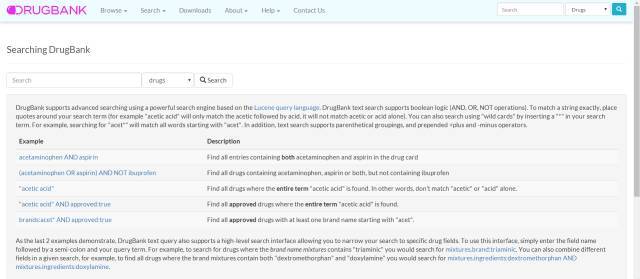
该浏览视图允许用户随意滚动鼠标查询数据或重新排序其查询内容。点击给定的Drug Card按钮（见下图）可显示相应药品的完整数据内容。数据内容标明了所有Drug Card数据范围和来源的完整说明。

http://5b0988e595225.cdn.sohucs.com/images/20170923/90623343b7bf421cba636d0ee45ffb0d.png

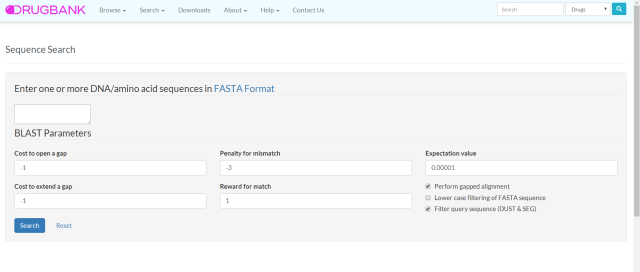
Pharma Browse按钮（见下图）允许用户根据指示分组浏览药物。该按钮不仅有助于药剂师和医生查询药品数据，也方便药物研究人员寻找潜在药物潜力。Chem Query按钮（见下图）允许用户绘制（使用Marvin Sketch小程序或Chem Sketch小程序）或写下（SMILES字符串）化合物，并在DrugBank中检索与查询化合物类似或相同的化学物质。



Text Query按钮支持对DrugBank文本部分进行更复杂的文本搜索（部分词匹配，区分大小写，拼写错误等）。



Seq Search按钮允许用户对DrugBank中包含的18,000个序列进行BLASTP（蛋白）序列搜索。支持单序列和多序列（整个蛋白质组）BLAST查询。



数据提取器是DrugBank最先进的搜索工具。用户可以从DrugBank下载所选择的文本和序列数据，并通过点击下载按钮跟踪最新的DrugBank统计信息[4]。

### STRING

该数据库主要是进行蛋白质之间的关系查找的数据库。STRING 数据是一个可以搜寻已知蛋白质之间和预测蛋白质之间相互作用的系统。蛋白之间的相互作用既包括蛋白质之间直接的物理相互作用，也包括蛋白质之间间接功能的相关性。它包括实验数据、文本挖掘结果、其他数据库的结果以及利用生物信息学的方法预测的结果，这些生物信息学的方法包括染色体临近、基因融合、系统进化谱和基于芯片数据的基因共表达。

### UnitProt

这是一个蛋白质数据库。这个数据库类似于DrugBank，该数据库主要是可以获取到靶点蛋白质的结构等信息。

### TTD

这个数据库类似于DrugBank数据库，可以用来查询相关的蛋白质靶点信息。

### TCMSP

该数据库可以用来查询中药材的化合物成分等相关的数据。现在使用TCMSP数据库中获取到数据,现在获取的数据主要有两个,分别是成分数据和靶点数据,无论是成分数据还是靶点基因数据,对应的其实都是3D的数据例如药物成分,其实就是有机物分子。

### OpenBabel

这个数据库可以获取到分子描述符计算和结构特征的统计学分析。

### NCBI

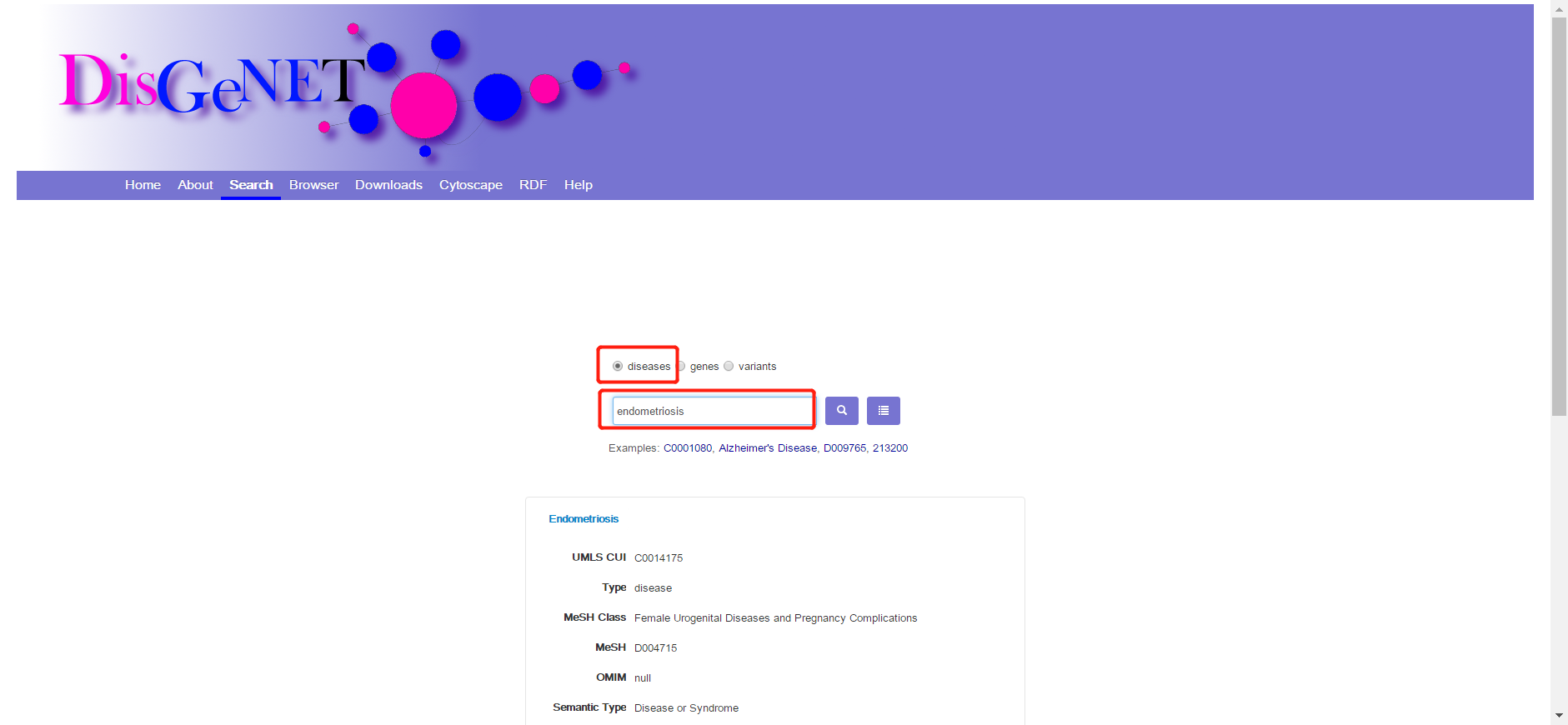
使用这个数据库可以获取到很多的信息，例如靶点。现在为了研究这个分子对接相关的,现在涉及到很多的数据库，上面的都是可能涉及到的数据库信息。现在我们希望使用这个数据库获取到靶点信息

### PDB

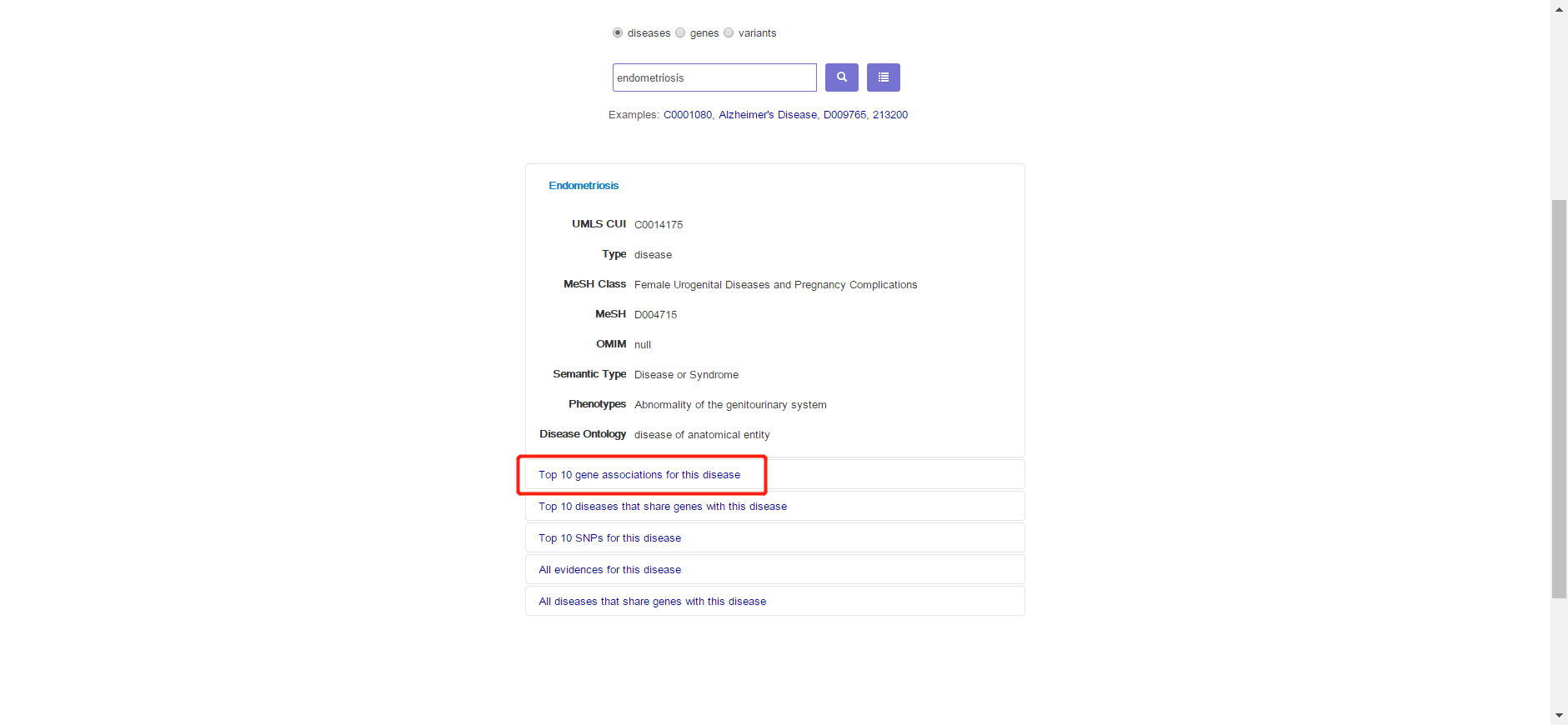
### DisGeNET

地址：<http://www.disgenet.org/web/DisGeNET/menu/search?1>

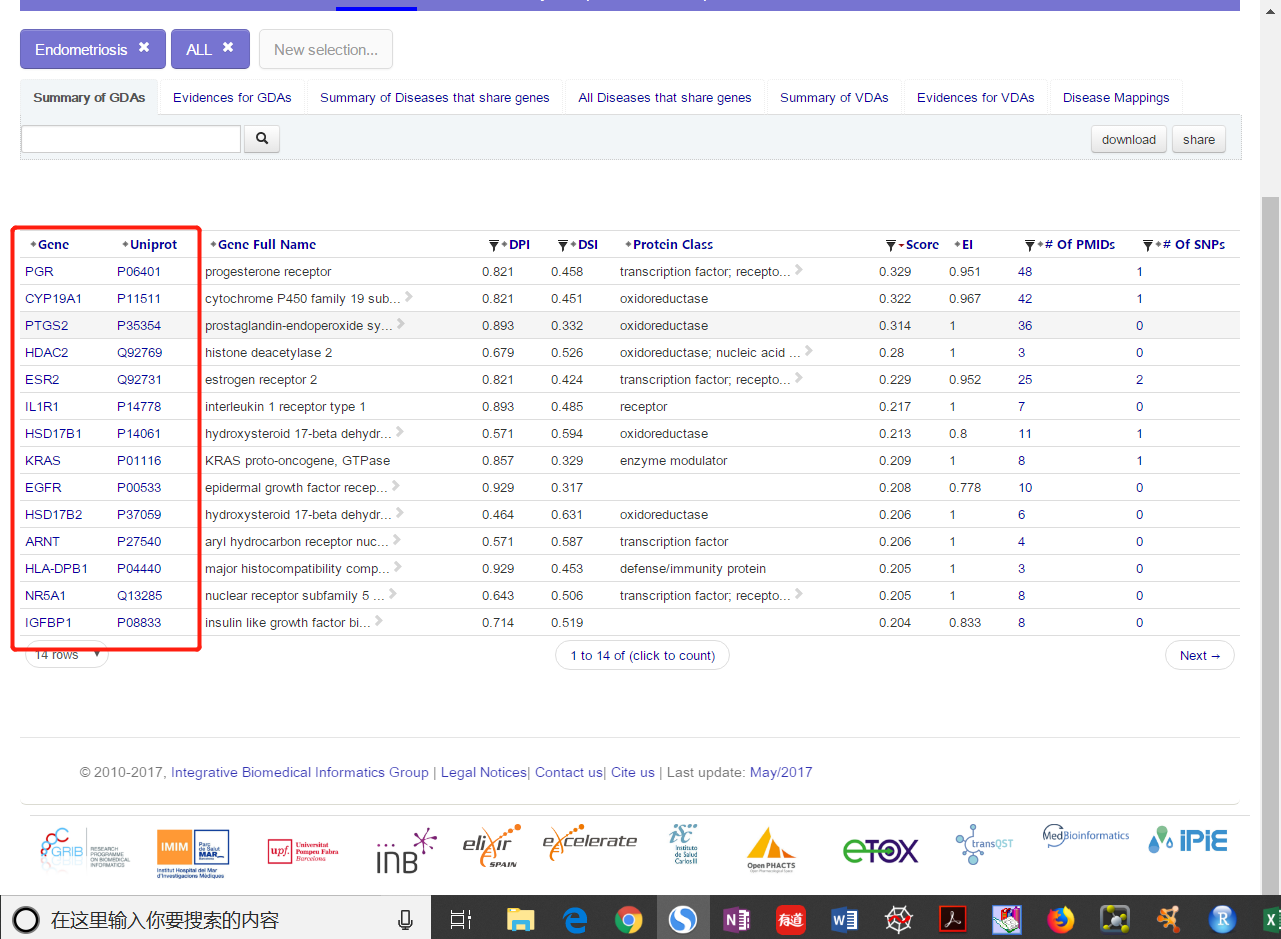
该数据库的使用是，在页面中输入疾病的名称，



输入疾病之后，查询搜索按钮，会得到对应的靶点数据，



这里存在很多的靶点信息，然后需要做的就是查询详细信息，下面是这些把靶点的信息。



这里边的靶点进行筛选即可。

### DAVID

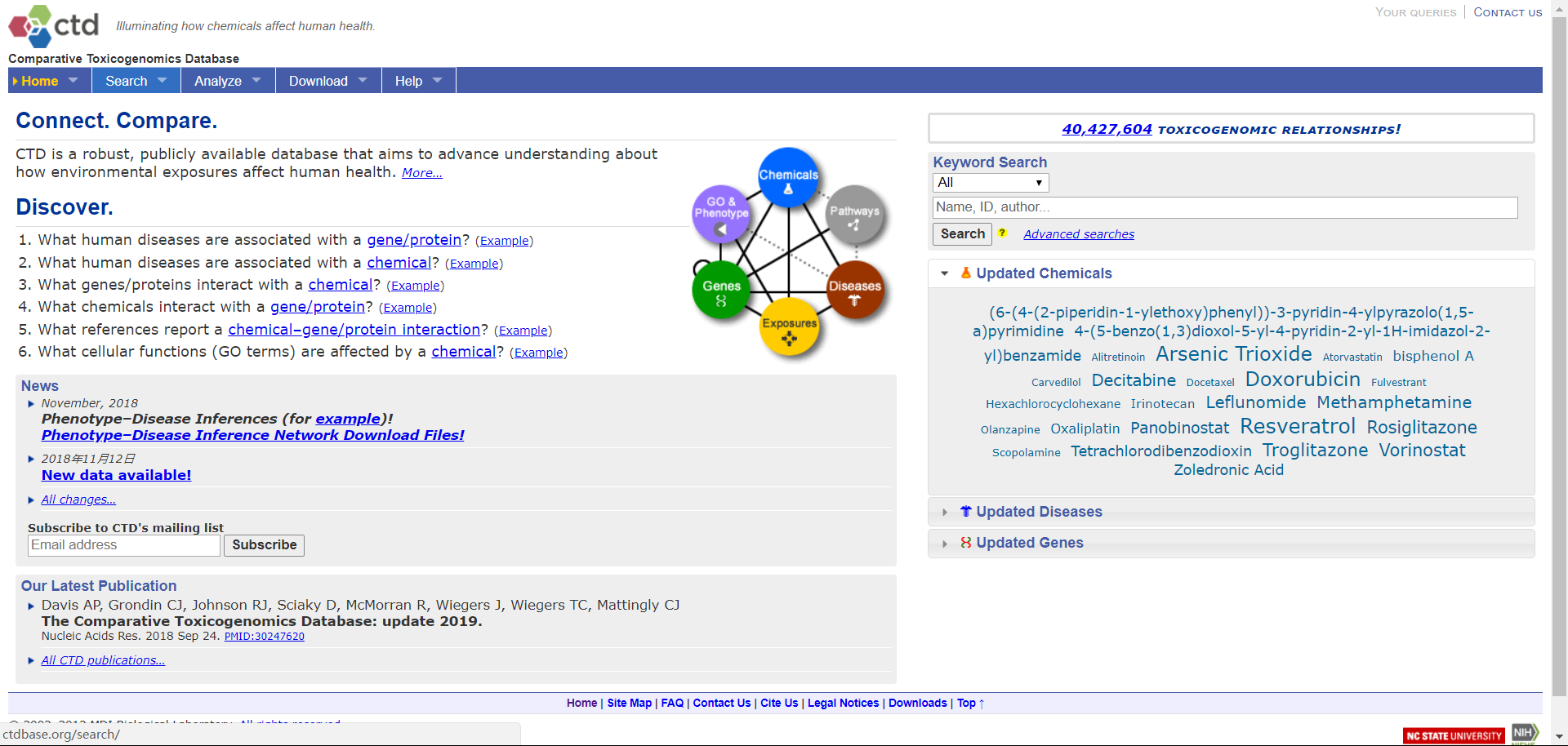
地址：<https://david.ncifcrf.gov/home.jsp>

### MSigDB

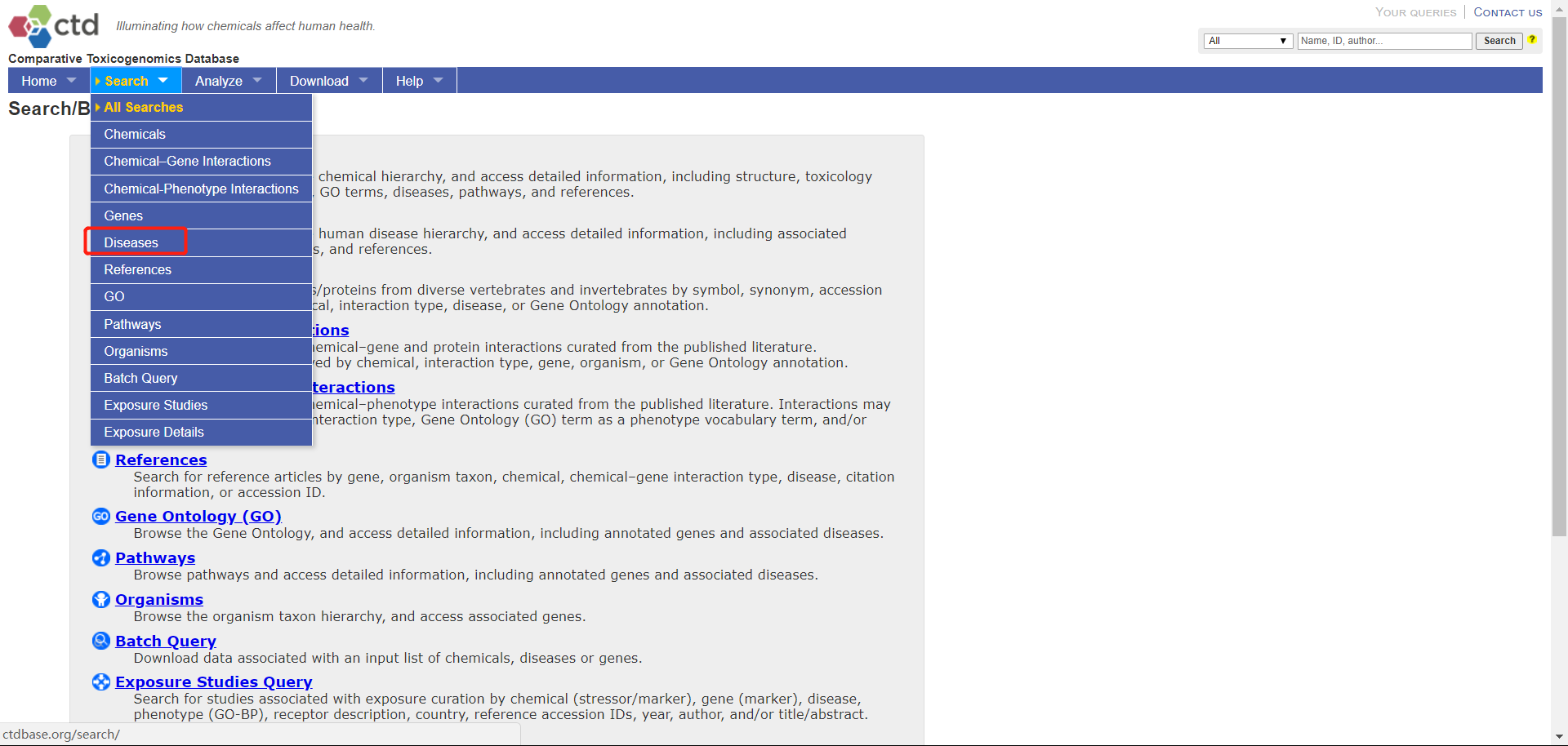
地址：<http://software.broadinstitute.org/gsea/msigdb/index.jsp>

### CTD

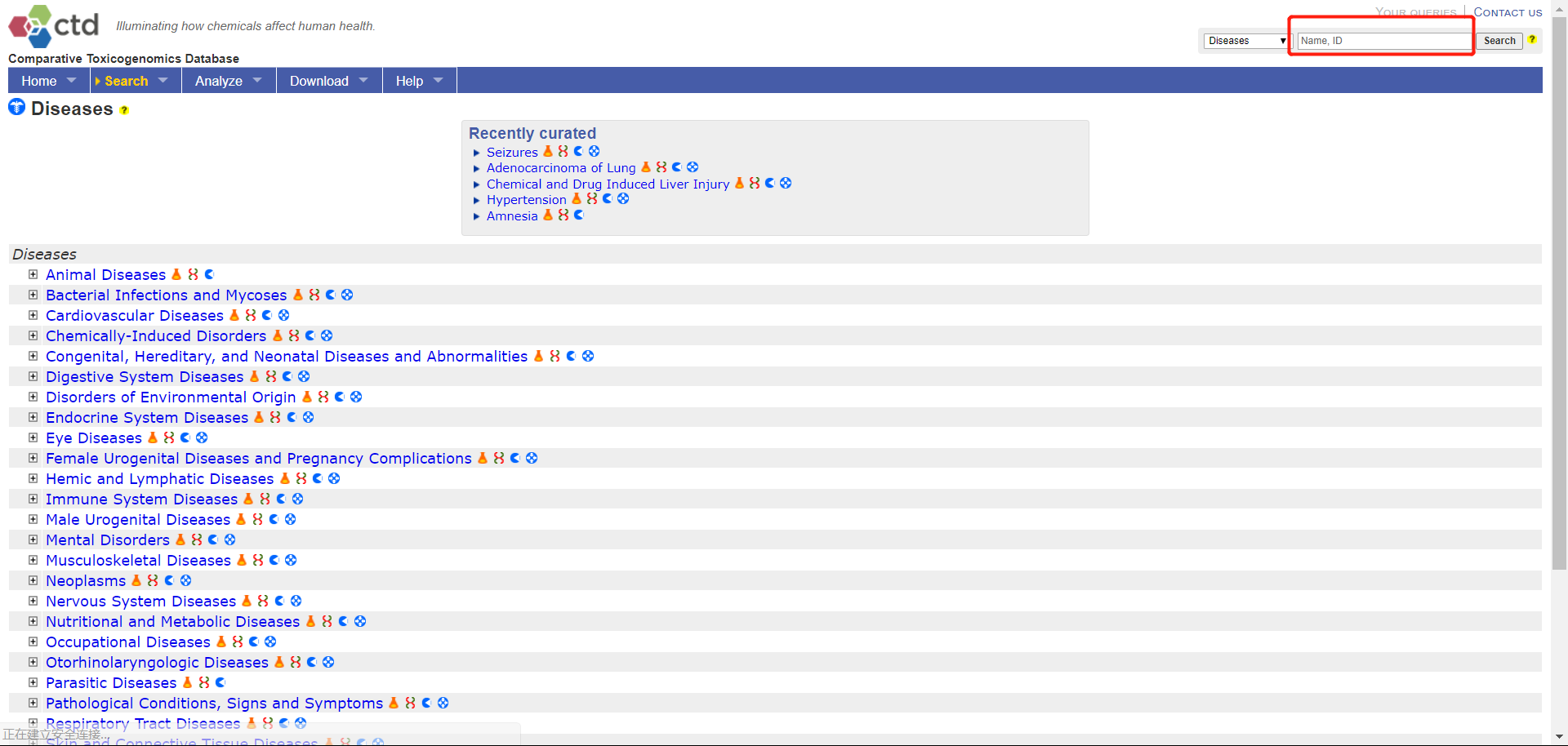
毒性与基因比较数据库CTD( Comparative Toxicogenomics Database, <http://ctdbase.org>)。这个数据库可以进行靶点的筛选，如下面的界面，



这里选择下面的Diseases疾病来进行选择，



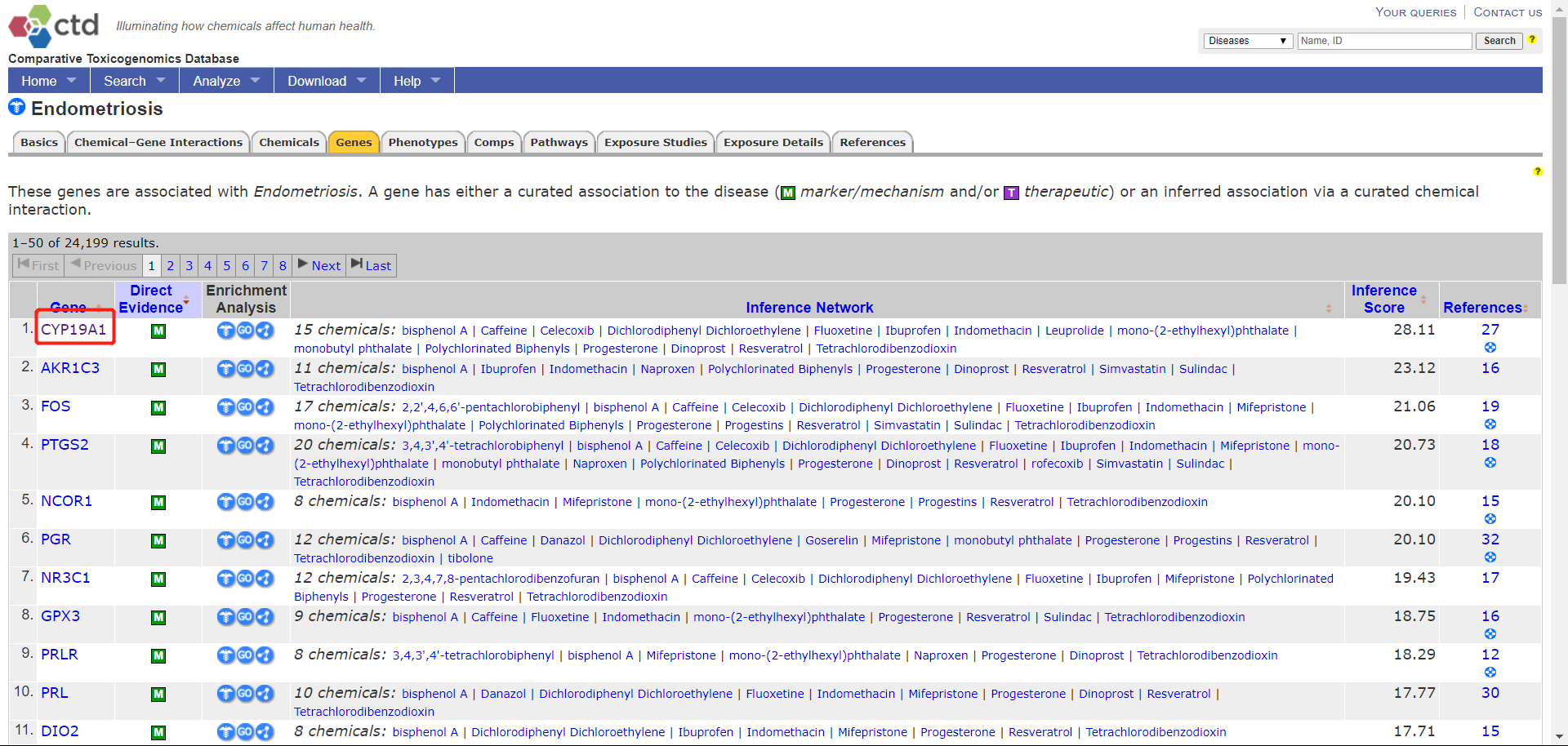
在这样的选择下，输入Diseases名称，输入某种疾病，例如endometriosis，得到查询后的数据，



选择Genes,就可以得到靶点数据，



下面就是得到的结果,



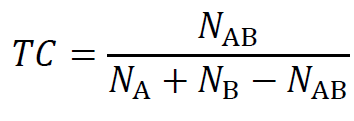
里边的Direct Evidence选择M就是说经过了特定的筛选结果，查看上面的M就是基因信息。得到这些基因信息后，再获取到对应的靶点蛋白信息。

## 分子对接的评价

在进行分子对接的时候,对接的好坏主要是使用“打分函数”来评价受体小分子与配体的匹配程度, 筛选出能高度和受体匹配的小分子。实际上，分子对接主要还是从两个方面判定对接的好坏。

## 相关概念

Tanimoto coefficient: Tanimoto coefficient (*TC*) was calculated with a perl program to compare the gene module profiles of each compound pair.



where *N*A and *N*B are the number of bits set for gene module profiles of compounds A and B, espectively, and *N*AB is the set bits that A and B have in common. If *TC* = 1, the compound pair have the same module profiles; if *TC* = 0, the pair have totally different module profiles.

## 分子对接方法

### （一）小分子化合物结构数据库的构建

基于本文第一部分养心氏片化学物质基础分析的结果，通过 ChEMBL

（www.ebi.ac.uk/chembl） 和 NCBI PubChem Database （ www.ncbi.nlm.nih.gov/

pccompound）等化学成分数据库检索养心氏片化学成分的分子结构，以 mol2 的文件格式保存并构建小分子化合物结构数据库。

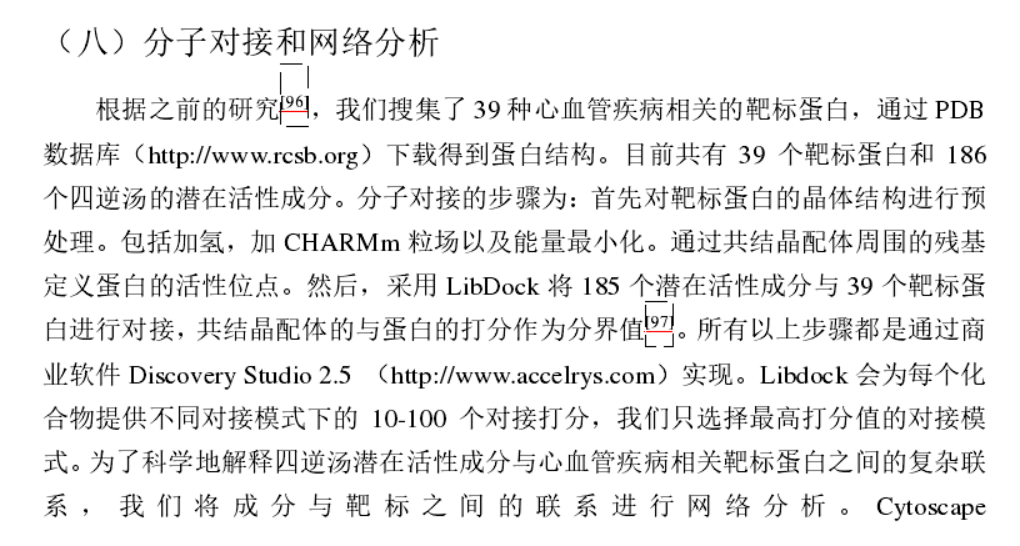
### （二）分子对接

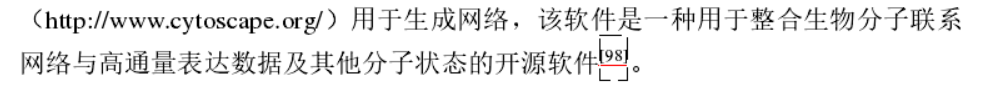
首先，从 PDB bank 网站中下载蛋白结构，导入 maestro 中；其次使用蛋白准备功能对受体蛋白进行处理，删除水分子，对蛋白进行主侧链修补加氢；最后加力场（Forcefield：OPLS\_2005），完成蛋白受体准备。使用受体网格生成功能，选中准备好的受体蛋白中的配体分子，对接位点以配体为中心，其他设置按照默认，运行结束后完成受体网格生成。使用 glide docking 功能，选用上一步生成的受体网格，将化合物导入并选中，选择标准对接模式，其他设置按照默认，运行结束后完成对接。

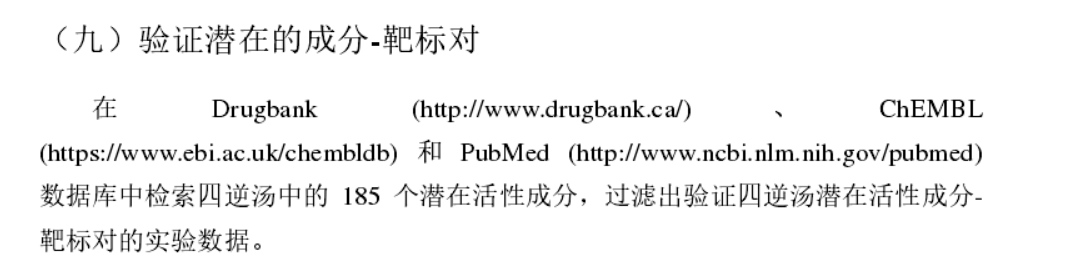
### （三）网络药理学分析

根据分子对接结果，采用 Cytoscape 2.8.3（www.cytoscape.org）构建养心氏片化学成分与作用靶点及相关通路的网络图。网络中节点表示化学成分或者潜在靶点、作用通路，连线表示化学成分对靶点、通路的作用。为了明确节点在网络中的重要性和影响力，网络图的参数由 NetworkAnalysis plugin 进行统计。

### 另一种对接方法







## 分子对接问题

根据分子对接结果，将化合物与靶点的对接得分大于6.5且高于原配体得分的分子与靶蛋白导入Cytoscape 3.1.1 网络分析软件。

上面是进行分子对接的时候提出的一个方案，但是这个方案中提出的化合物与靶点的对接得分是需要理解的，这里提出的对接是使用软件autodock软件来进行的，但是这个对接的得分是怎么的一个数字，另外，原配体是什么，都是需要进行分析的。

另外值得关注的就是对接的基本思路：分子对接这类方法首先要建立大量化合物（例如几十至上百万个化合物）的三维结构数据库，然后将库中的分子逐一与靶标分子进行“对接”（docking），通过不断优化小分子化合物的位置（取向）以及分子内部柔性键的二面角（构象），寻找小分子化合物与靶标大分子作用的最佳构象，计算其相互作用及结合能。在库中所有分子均完成了对接计算之后，即可从中找出与靶标分子结合的最佳分子（前50名或前100名）

从上面的研究中可以看到，无论是使用aotudock软件进行分子对接还是使用discovery studio软件进行分子对接，都需要给一些数据，用来判断是不是合适的，这个数据就是对接的打分。打分这个数据也是在进行研究中的重点。打分的问题解决了，基本的问题就解决了。

## 对接算法的研究

在进行对接研究的时候，对接算法的研究是一个关键和核心的东西。分子对接的原理是研究的关键与核心。另外，对接的算法是对接研究的重点与难点，也是论文书写的比较好的。

## 分子对接工具

根据相关的研究进展，分子对接的工具用的最多的是Autodock和Discovery Studio这两个软件，这两个软件是进行分子对接最常用的软件。涉及到的对接的原理以及相关的算法也是最关键最核心的部分。

一般情况下分子对接的步骤:

(1) 构建受体和配体分子的三维模型;

(2) 充分采集受体-配体各种可能的结合模式;

(3) 初步过滤掉一些明显不可能的结合模式;

(4) 从剩余的对接结构中筛选出与复合物天然结构相类似的近天然结构。

# 蛋白质分子对接的原理

## 分子对接的一般原理

分子对接是将已知三维结构的蛋白质分子从蛋白质数据库中取出（大多数

以PDB文件来组织），把它们逐一放在靶标分子的活性位点处。通过不断优化受体蛋白质的位置、构象、分子内部可以旋转的化学键的二面角和受体小分子的氨基酸残基侧链和骨架，寻找受体小分子化合物与靶标大分子作用的最佳构象，并预测其结合模式、亲和力，通过打分函数挑选出接近天然构象的与受体亲和力最佳的配体-受体复合物的一种理论模拟分子间作用的方法。

# 参考文献

[1]Wishart David S,Knox Craig,Guo An Chi,Shrivastava Savita,Hassanali Murtaza,Stothard Paul,Chang Zhan,Woolsey Jennifer. DrugBank: a comprehensive resource for in silico drug discovery and exploration.[J]. Nucleic Acids Research,2005,34(Database i).

[2] https://www.drugbank.ca/

[3]Wishart David S,Knox Craig,Guo An Chi,Cheng Dean,Shrivastava Savita,Tzur Dan,Gautam Bijaya,Hassanali Murtaza. DrugBank: a knowledgebase for drugs, drug actions and drug targets.[J]. Nucleic Acids Research,2007,36(Database i).

[4] https://www.drugbank.ca/about