人

----------------------------

Gene：从Human\_G\_G中获取所有用到的Gene的EntrezID，导入Human\_G\_G.txt到matlab矩阵中，使用HumanGene1=unique(HumanGG(:,3:4))获得不同的GeneId，并导出到HumanGene1.txt为G-G使用到的Gene(19491个)

从ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE\_INFO/Mammalia/下载的Homo\_sapiens.gene\_info，执行Human\_Gene.java，输出Human\_Gene.txt，共20956个，注意好多基因有别名

执行A=intersect(EntrezGeneId,VarName1);找到两个基因集的交集GeneBoth.txt(14985),各自的补集GeneOnlyPPI.txt,GeneOnlyPC.txt

--------------------------

Phenotype：

-------------------------

G-G：从http://thebiogrid.org/download.php下载BIOGRID-ORGANISM-3.4.127.tab2压缩包，BIOGRID-ORGANISM-Homo\_sapiens-3.4.127.tab2.txt为人的G-G，执行Human\_G\_G.java,输出Human\_G\_G.txt

---------------------------

P-P:

-------------------------

G-P:从<ftp://ftp.omim.org/OMIM/>下载morbidmap，执行Human\_G\_P.java输出Human\_G\_P.txt（已经去除重复的了）。

现在要把Human\_G\_P.txt中的symbol列对应到entrezID，执行SymbolToEntrezID.java,输出Human\_G\_P\_EntrezID.txt,会发现少了一些条，经研究发现这些基因的类型不是protein-coding一般是unkown。

---------------------------

鼠

----------------------------

Gene：

从ftp://ftp.inforrmatics.jax.org/pub/reports/index.html下载MIG\_Gene.rpt,由Mouse\_Gene.java处理输出Mouse\_Gene.txt

从ftp://ftp.inforrmatics.jax.org/pub/reports/index.html下载MGI\_EntrezGene.rpt,由Mouse\_Gene.java处理输出Mouse\_Gene.txt

从Mouse\_G\_G中获取所有用到的Gene的EntrezID，导入Mouse\_G\_G.txt到matlab矩阵中，使用MouseGene1=unique(MouseGG(:,3:4))获得不同的GeneId，并导出到MouseGene1.txt为G-G使用到的G

--------------------------

Phenotype：

-------------------------

G-G：从http://thebiogrid.org/download.php下载BIOGRID-ORGANISM-3.4.127.tab2压缩包，BIOGRID-ORGANISM-Mus\_musculus-3.4.127.tab2.txt为老鼠的G-G，执行Mouse\_G\_G.java输出Mouse\_G\_G.txt

---------------------------

P-P:

-------------------------

G-P:从MGI下载文件MGI\_Geno\_Disease.rpt，执行Mouse\_G\_P.java输出Mouse\_G\_P.txt,并同时输出MouseP\_HumanP.txt

---------------------------

跨物种

---------------------

G人-G鼠：从ftp://ftp.inforrmatics.jax.org/pub/reports/index.html下载HOM\_MouseHumanSequence.rpt，由Ortholog.java处理输出ortholog.txt

--------------------

P人-P鼠：见老鼠G\_P获取

-----------------------

将矩阵导出到txt

[nrows,ncols]= size(HumanGene1);

filename = 'HumanGene1.txt';

fid = fopen(filename, 'w');

for row=1:nrows

fprintf(fid, '%s\n', num2str(HumanGene1(row,:)));

end

fclose(fid);

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*