1: Install the following packages

“Data.table”, “RCPP”, “ggplot2”, “inline”

2: Set up the following folder

D:\\Rfiles\\Indexsearch

3: Copy the following files to the above folder

1. A MS file in .mgf format, the file is converted from raw MS to mgf format by proteome discoverer. The name of the file is ”sample.mgf”.
2. A psms file in .csv format, the file contain all the psms identified in conventional linear peptide search. At least one column of the file contain the scan of each psm. The name of the file is ”phosX-psms-scan.csv”
3. A type 0 peptides list in .csv format. The name of the file is “Type0pep.csv”.
4. “All peptides” list in .csv format, the name of the file is” Allpep.csv”.

4: Execute the script.

5: The final results is csv format, the name of the result file is “s-mDB.csv”

The format of the example input files.

The psms, “Type0pep” and “Allpep” files should have the following format and columns respectively.

psms

FirstScan
1216
1217
4105
4106
4112
4115
4126
4210
4298
4358

Type0pep

Protein_Name	Sequence
Decoy_sp|P0A9L8|P5CR_ECOLI	HVAIRGGKVGSGGSQFMQAAADASGPHLFWYFVNIVK
P61978	IITITGTQDQIQNAQYLLQNSVKQYSGK
Decoy_sp|P41441|GSPF_ECOLI	FGANLTMKTGVIS
P25786	FVFDRPLPVSR
Decoy_sp|P77718|THII_ECOLI	KSQYGPFVNLRVVTK
Decoy_sp|P77249|SFMC_ECOLI	LKNPNRKPLGPANSR
D6RD63	QDLAQLMNSSGSHKDLAGK
Q9Y266	LQLEIDQKKDAENHEAQLK
Decoy_sp|P0AFG3|ODO1_ECOLI	KQLNDGDTDQAETEHRVAVMSIETENSAK
Q14152	TKNETDEDGWTTVRR
P84098	ILMEHIHKLK
P31948	TVDLKPDWGKGYSRK
A0A0B4J210	VNPWTKNALPPVLTTVNGQSPPEHSAPAK
Decoy_sp|P0ABZ6|SURA_ECOLI	HGTWKSEASDNTATVLQPQDTDASPDFARVTVK
Decoy_sp|P0A8V2|RPOB_ECOLI	SVAHGVHKVQIDPYLKELEIVVELSFR
Decoy_sp|P0AG16|PUR1_ECOLI	GGDRAEDYIQEMVDCDKTYFEAR
P17174	KVEQKIANDNSLNHEYLPILGLAEFR
Decoy_sp|P42641|OBG_ECOLI	MIVHGRKIYR
A0A087WYT3	LTFSCLGGSDNFKHLNEIDLFHCIDPNDSK

Decoy_sp|P0AEJ6|EUTB_ECOLI	GMGANMPAENILLEFSGLKRDFPAAIYVNGTGSVYK
Q5JRR6	QPAENVNQYLTDPKFVER
Q5SQT6	GYIWNYGAIPQTWEDPGHNDKHTGCCGDNDPIDVCEI GSK
P32119	ATAVVDGAFKEVKLSDYK
A0A1W2PNV4	NAYVWSQKDGVWKPTLVILR
Decoy_sp|P0A6S7|GPDA_ECOLI	FLGKNDIARLQAVSVNGTINGPLMAR
Decoy_sp|P52129|RNLA_ECOLI	LTESSKAR
Decoy_sp|P27254|ARGK_ECOLI	ADLYQQTEFNGASPSRAIIEAAGGTEHTSKIR
B4E0P5	PVYHAITKHSPK

Allpep

Protein_Name	Sequence
Decoy_sp|Q92878|RAD50_HUMAN	NMGEKSIAFIYVEAGMAKNTLIIFAR
Decoy_sp|Q09028|RBBP4_HUMAN	NNCDLHLTHKRSPMDRYFWPDR
Decoy_sp|P53004|BIEA_HUMAN	SKGEVYKLESLEIK
Decoy_sp|Q96J01|THOC3_HUMAN	AFVCSVSAIDPLDLKMTSPR
sp|Q92598|HS105_HUMAN	NKDEK
sp|O75179|ANR17_HUMAN	ILLNAGAEINSRTGSKLGISPLMLAAMNGHTAAVK
Decoy_sp|P35998|PRS7_HUMAN	KPGRGMKPVFSMIGVKILELK
Decoy_sp|Q06265|EXOS9_HUMAN	EKLITAIQIMSPEERIIQNTSQK
Decoy_sp|P42696|RBM34_HUMAN	LSYKSKQIK
sp|P55060|XPO2_HUMAN	LVLDAFALPLTNLFKATIELCSTHANDASALR
Decoy_sp|Q13464|ROCK1_HUMAN	ILSEQQMRYGSNLVMCEFSGLQSDVNAKLFEK
sp|Q13439|GOGA4_HUMAN	AYEEQLAQLQQKLLDLETER
Decoy_sp|P13010|XRCC5_HUMAN	ESDKYFLIEFLEQIK
sp|P52343|RIR1_HHV6U	HLDLSNPLPLIGKCSDGVVMHVK
sp|Q32ZD5|POLG_KOKV	YVSAISQKENVGQEDGAEIEDNWFR
sp|P78417|GSTO1_HUMAN	MSGESARSLGKGSAPPGPVPEGSIR
Decoy_sp|Q7Z7E8|UB2Q1_HUMAN	QIRDATGEAGGSLLGILAPQPEAGFLNGEFKVLTGASLK
sp|P78344|IF4G2_HUMAN	LDHERAKSLMDQYFARMCSLMLSK
Decoy_sp|P12081|SYHC_HUMAN	AESIVLDLFAAEQKVGR
Decoy_sp|Q14008|CKAP5_HUMAN	QLKCGAEKLIAAQSYNYGQEGK
sp|Q7LBC6|KDM3B_HUMAN	FEDLMENLPLPEYTKR
sp|Q9UBT2|SAE2_HUMAN	SRFDIKSMAGNIIPAIATTNAVIAGLIVLEGLK
Decoy_sp|O60841|IF2P_HUMAN	GKERDVEASVGK
Decoy_sp|P36871|PGM1_HUMAN	MLEHSIDGLLKLK
Decoy_sp|Q13838|DX39B_HUMAN	VDKSCDFVRCADR
Decoy_sp|O75369|FLNB_HUMAN	DDGGSPPLKVYWDCPKGNTGEQLFCTIPWDCK
sp|Q8IY81|SPB1_HUMAN	EKVAQLRSLYK
Decoy_sp|Q5VTL8|PR38B_HUMAN	DGYSHGDEQSDSKSMQR

Decoy_sp|P48643|TCPE_HUMAN	QLGKGDLMPHELYDELGSLKLEATMK
Decoy_sp|P50748|KNTC1_HUMAN	KFIKPVLDPLTSLCGSRQSEFER
Decoy_sp|Q9C005|DPY30_HUMAN	YPKNIR
Decoy_sp|P52907|CAZA1_HUMAN	FESKPRPTWEKMGIAFKDQEYEVEIESDVNIRK
Decoy_sp|Q92616|GCN1_HUMAN	EAGGVDLVGTQGSGCVHELAKESWRLKK
Decoy_sp|P48147|PPCE_HUMAN	VVGLPEPYLHEASKLDLCSSFQRNLGSYICAFR
sp|Q13439|GOGA4_HUMAN	TNELINISSSKTNAILSRISHCQHR
sp|Q14683|SMC1A_HUMAN	LPLSKGTMDDISQEEGSSQGEDSVSGSQR
sp|P13639|EF2_HUMAN	KLPRTFCQLILDPIFKVFDAIMNFK
sp|Q5T0N5|FBP1L_HUMAN	KKLQQR