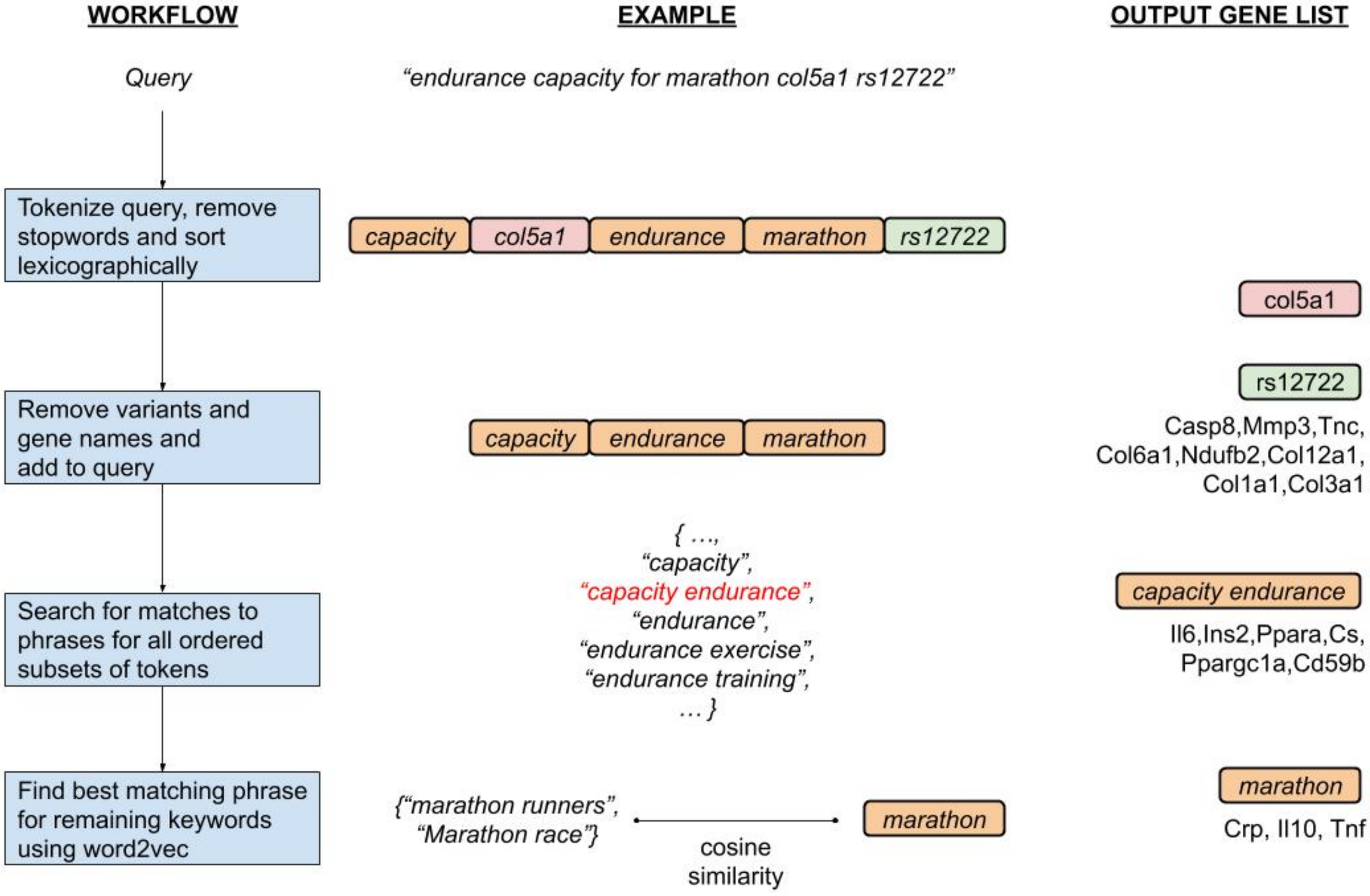
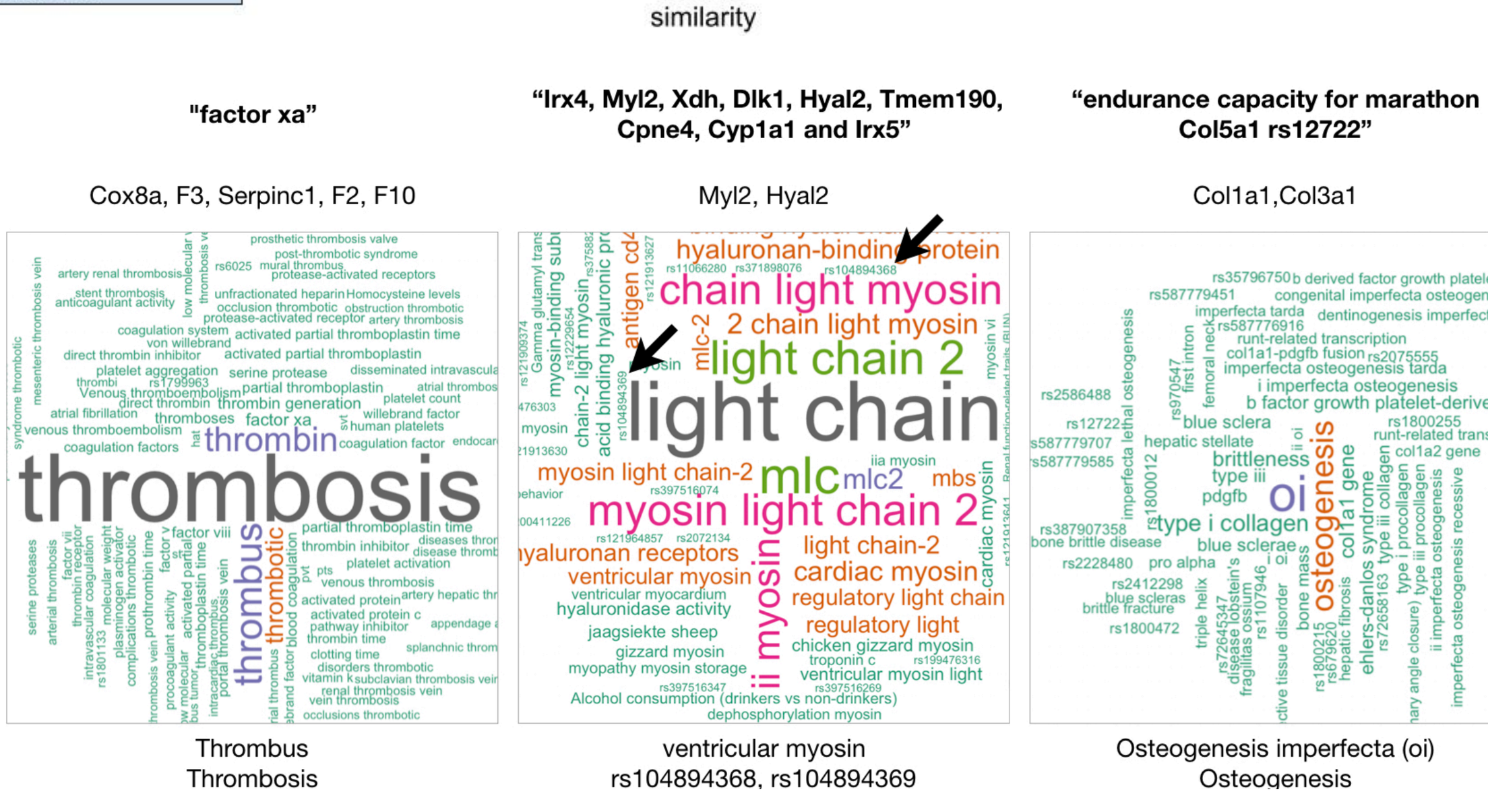
A


WORKFLOW
EXAMPLE
OUTPUT GENE LIST
Query
"endurance capacity for marathon col5a1 rs12722"
Tokenize query, remove stopwords and sort lexicographically
capacity col5a1 endurance marathon rs12722
col5a1
rs12722
Remove variants and gene names and add to query
capacity endurance marathon
Casp8,Mmp3,Tnc, Col6a1,Ndufb2,Col12a1, Col1a1,Col3a1
Search for matches to phrases for all ordered subsets of tokens
{ …, "capacity", "capacity endurance", "endurance", "endurance exercise", "endurance training", … }
capacity endurance
Il6,Ins2,Ppara,Cs, Ppargc1a,Cd59b
Find best matching phrase for remaining keywords using word2vec
{"marathon runners", "Marathon race"}
cosine similarity
marathon
marathon
Crp, Il10, Tnf

B


Query
"factor xa"
"Irx4, Myl2, Xdh, Dlk1, Hyal2, Tmem190, Cpne4, Cyp1a1 and Irx5"
"endurance capacity for marathon Col5a1 rs12722"
Optimised genes
Cox8a, F3, Serpinc1, F2, F10
Myl2, Hyal2
Col1a1,Col3a1
Wordcloud
Highlighted texts
Thrombus
Thrombosis
ventricular myosin
rs104894368, rs104894369
Osteogenesis imperfecta (oi)
Osteogenesis