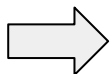
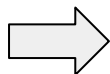


**Transcriptomic  
datasets in GEO**  
(n = 48,893)



**In Gemma**  
(n = 5,997)



**Reference set**  
(relevant to JIA,  
n = 6)

Calculate mean  
embedding



Calculate pairwise  
similarities

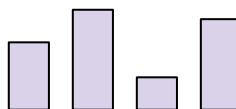
**Comparison  
set A**  
(relevant to JIA,  
n = 6)

**Comparison  
set B**  
(irrelevant to JIA)

Combine,  
sort

Dataset	Similarity
GSE89229	0.983
GSE2492	0.971
GSE9873	0.966
...	...
GSE39269	0.023
GSE93723	0.019

Compare against  
relevant/irrelevant  
labels.



**Metrics**

*Sizes evaluated:*

n = 6

n = 12

n = 30

n = 60

n = 600

n = 1,800

n = 5,985