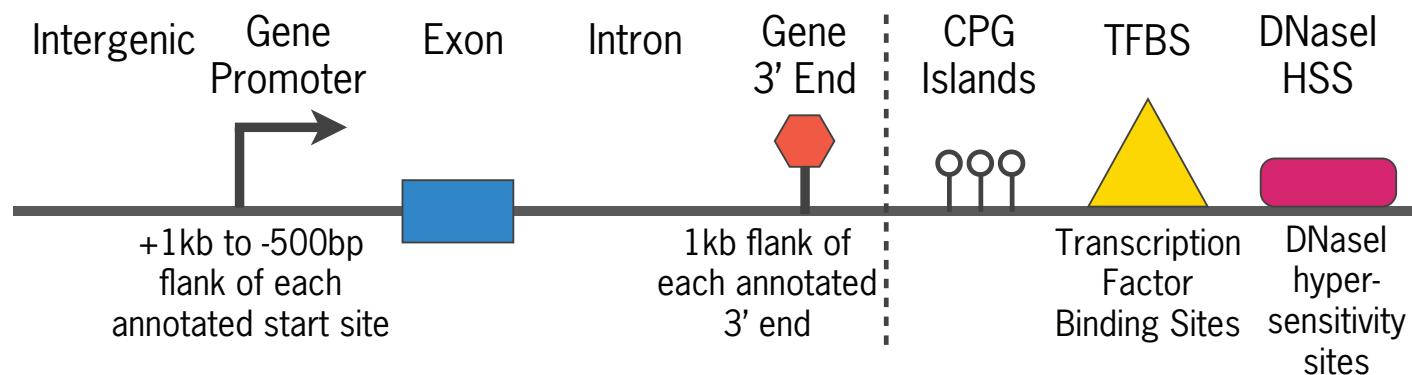


# goldmine

R package to explore relationships between genome regions from experiments versus known genes and features

Jeff Bhasin <[jeffb@case.edu](mailto:jeffb@case.edu)>



# All genome-wide assays produce lists of genomic regions that are hard to interpret on their own

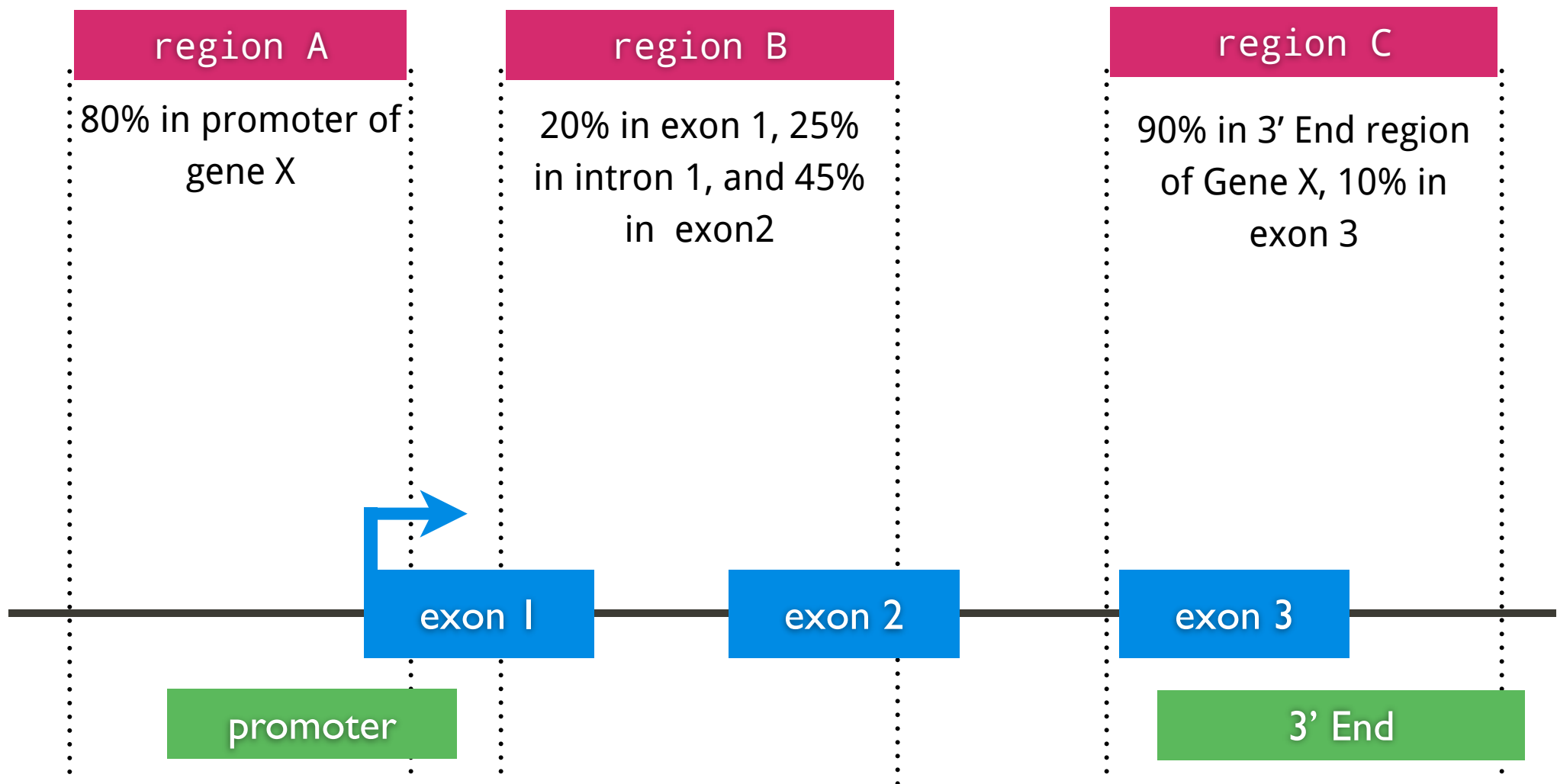
	A	B	C	D	E	F	G	H	I	J
1	<u>chr</u>	<u>start</u>	<u>end</u>	<u>width</u>	<u>strand</u>	<u>anodev.padj</u>	<u>pattern</u>	<u>benign</u>	<u>low.p</u>	<u>high.p</u>
2	chr1	672201	672350	150	*	0.015729669	10	0.7143	1.778	0.963
3	chr1	724251	724400	150	*	0.0159674841	10	44.714	73.5	56.85
4	chr1	980201	980400	200	*	0.00736754	10	13.036	23.21	21.97
5	chr1	1154451	1154700	250	*	0.0006365276	10	1.5429	3.467	2.578
6	chr1	1158251	1158450	200	*	0.0053899155	10	81.25	146	101.4
7	chr1	1194651	1194800	150	*	0.002876592	10	22.19	39.5	35.44
8	chr1	1251251	1251450	200	*	0.0016715211	10	2.25	5.542	4.083
9	chr1	1276551	1276750	200	*	0.0001168894	10	40.214	68.17	50.97
10	chr1	1297151	1297300	150	*	0.0099407967	10	4.2381	8.389	6.407
11	chr1	1409551	1409700	150	*	0.0002362142	10	6.3333	11.83	9.889
12	chr1	1435851	1436000	150	*	0.0013747539	10	1.619	3.611	2.148
13	chr1	1474301	1474500	200	*	0.0091654605	10	18.964	35.54	28.58
14	chr1	1490901	1491050	150	*	0.0221025243	10	18	29.67	25.74
15	chr1	1605701	1605850	150	*	0.0058019203	10	6.2381	11.39	8.889
16	chr1	1674051	1674250	200	*	0.0019693833	10	7.5357	16.83	13.5
17	chr1	1730401	1730550	150	*	0.0072279782	10	3.1905	6.389	5.185
18	chr1	1871051	1871200	150	*	0.0008207126	10	7.1429	11.89	8.852

?

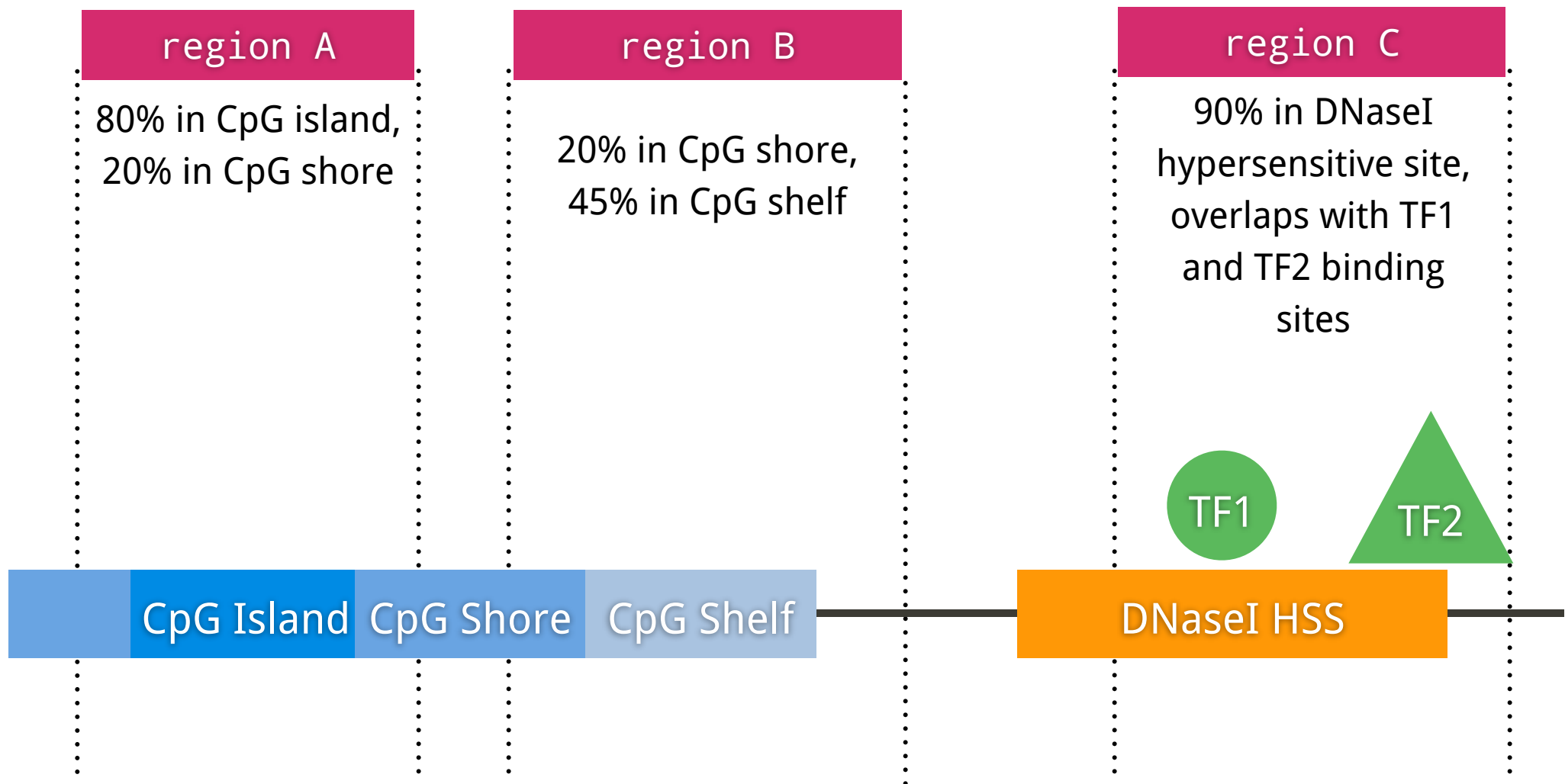
?

Knowing how our regions **relate** to known **genes** and **features** is vital for **making sense** of this data

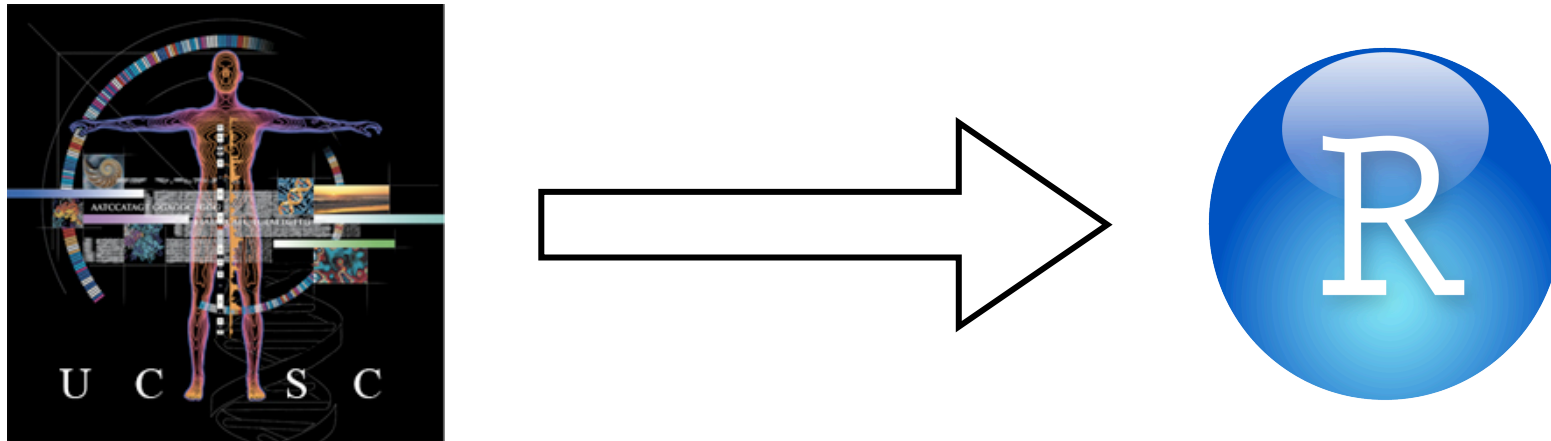
# goldmine takes any list of genomic ranges and reveals all relationships with genes



# goldmine can also relate the query regions to any feature sets of interest



# goldmine caches gene and feature data from the UCSC genome browser (goldenPath)



## Direct access in R to:

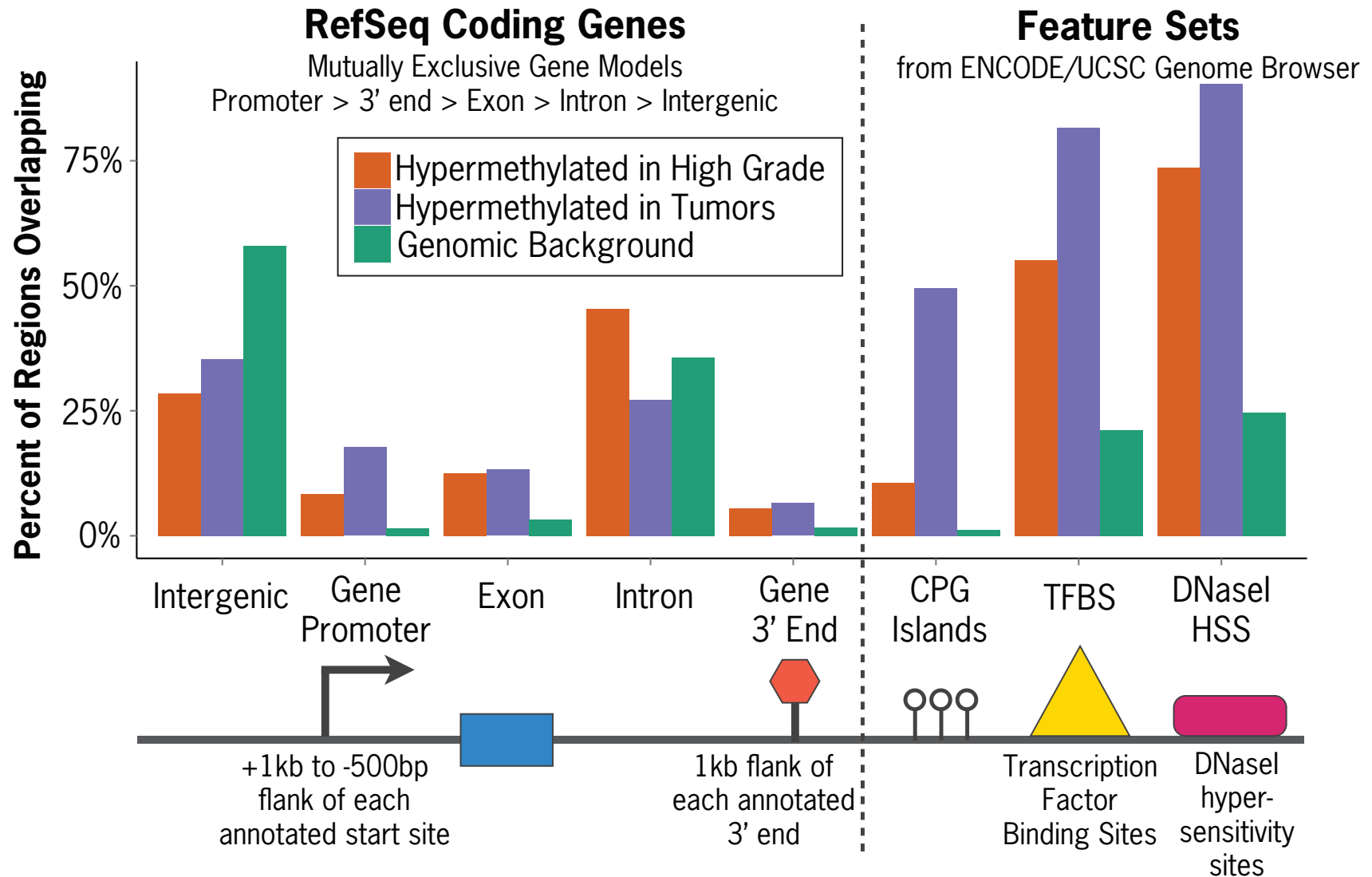
### gene sets

UCSC knownGene  
RefSeq  
ENSEMBL genes

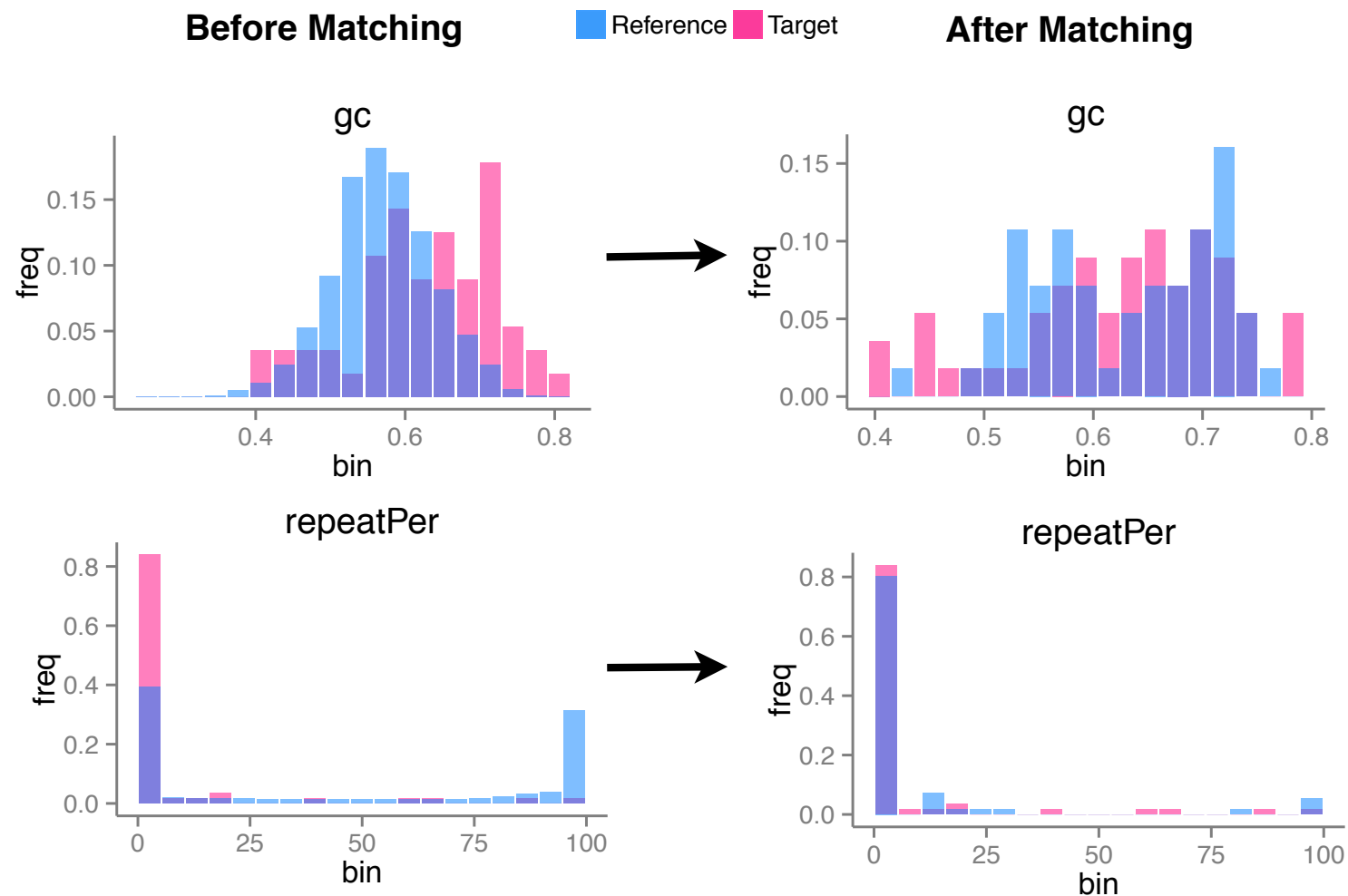
### feature sets

CpG islands  
DNaseI Hypersensitive Sites  
Transcription Factor Binding Sites  
Conserved Motifs and Regions  
Any ENCODE data set

# goldmine can summarize relationships across the query set and perform enrichment testing

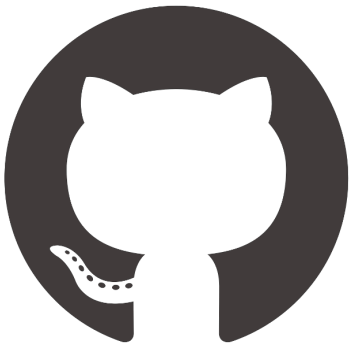


**goldmine can draw random genomic background sets and match for any covariate (length, %GC, repeat %, CpG density, etc)**



# goldmine is available from GitHub

```
library(devtools)  
install_github("jeffbhasin/goldmine")
```



<http://github.com/jeffbhasin/goldmine>