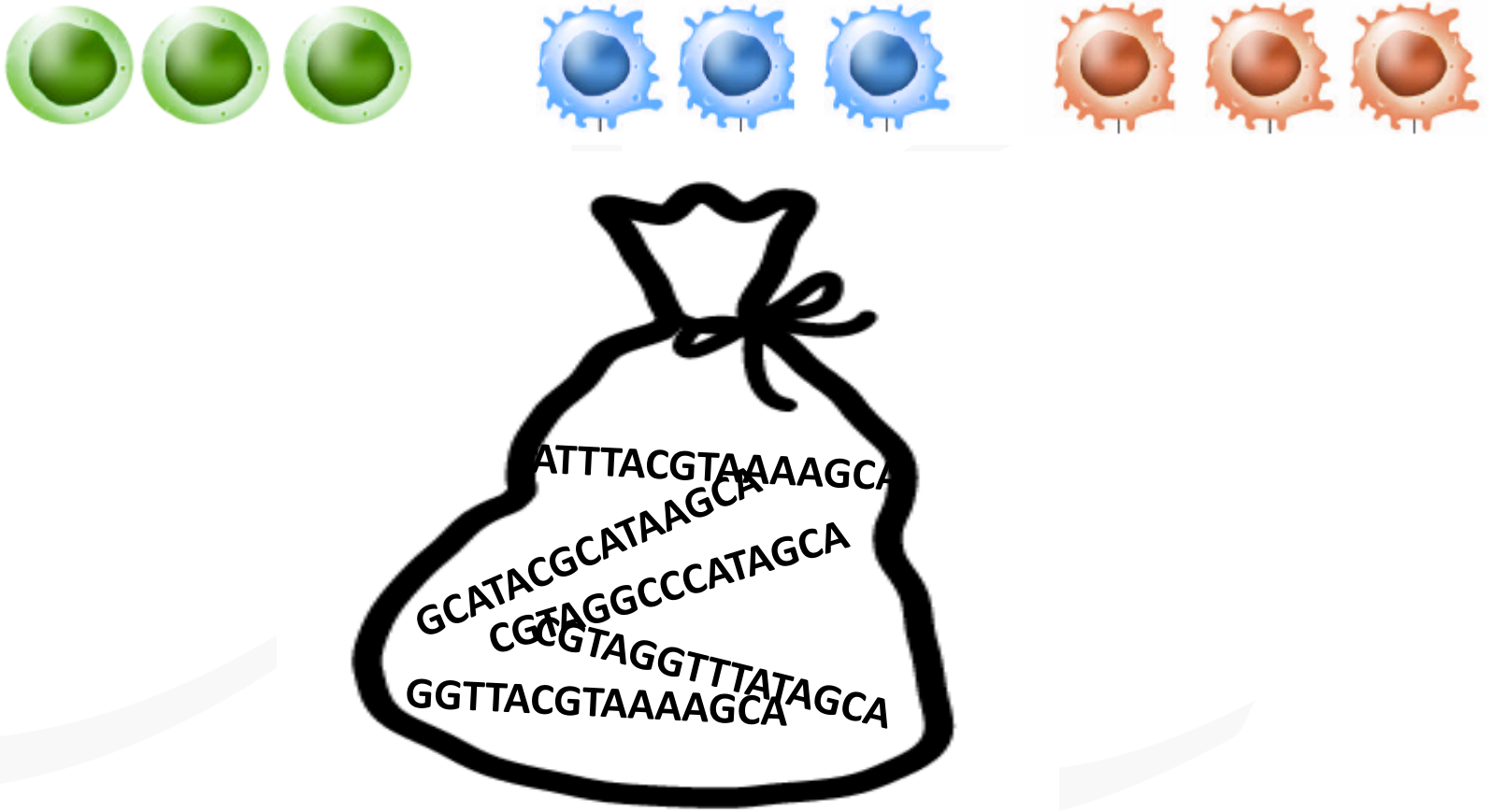
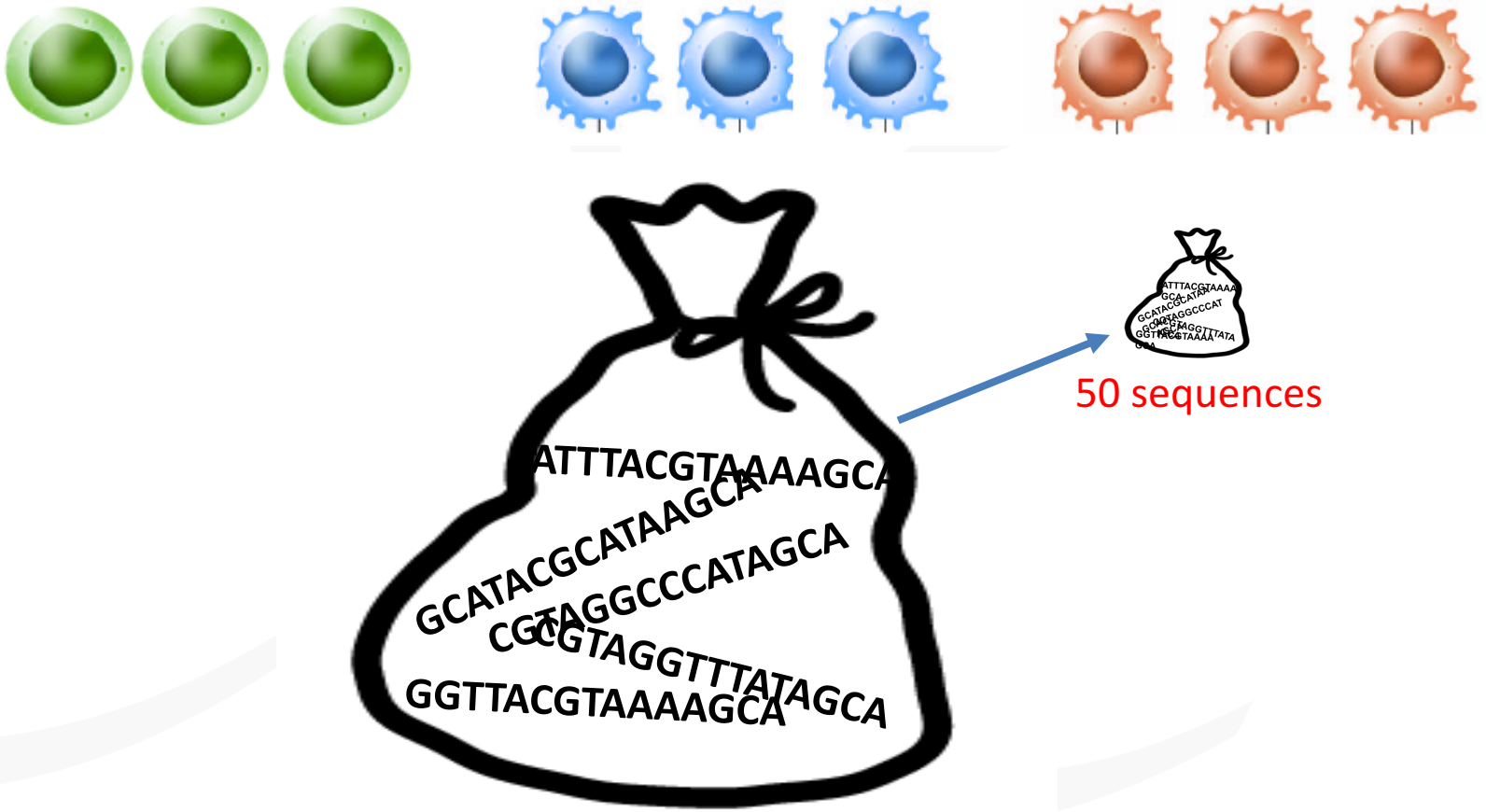
Using AI to classify cancer samples

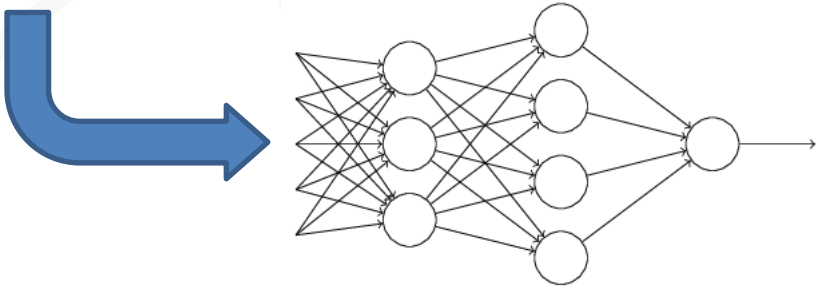
What is the minimum amount of information needed to classify samples?


ATTTACGTAAAAGCA
GCATACGCATAAGCA
CGTAGGCCCATAGCA
GGTTACGTAAAAGCA
CGGTAGGTTTATAGCA

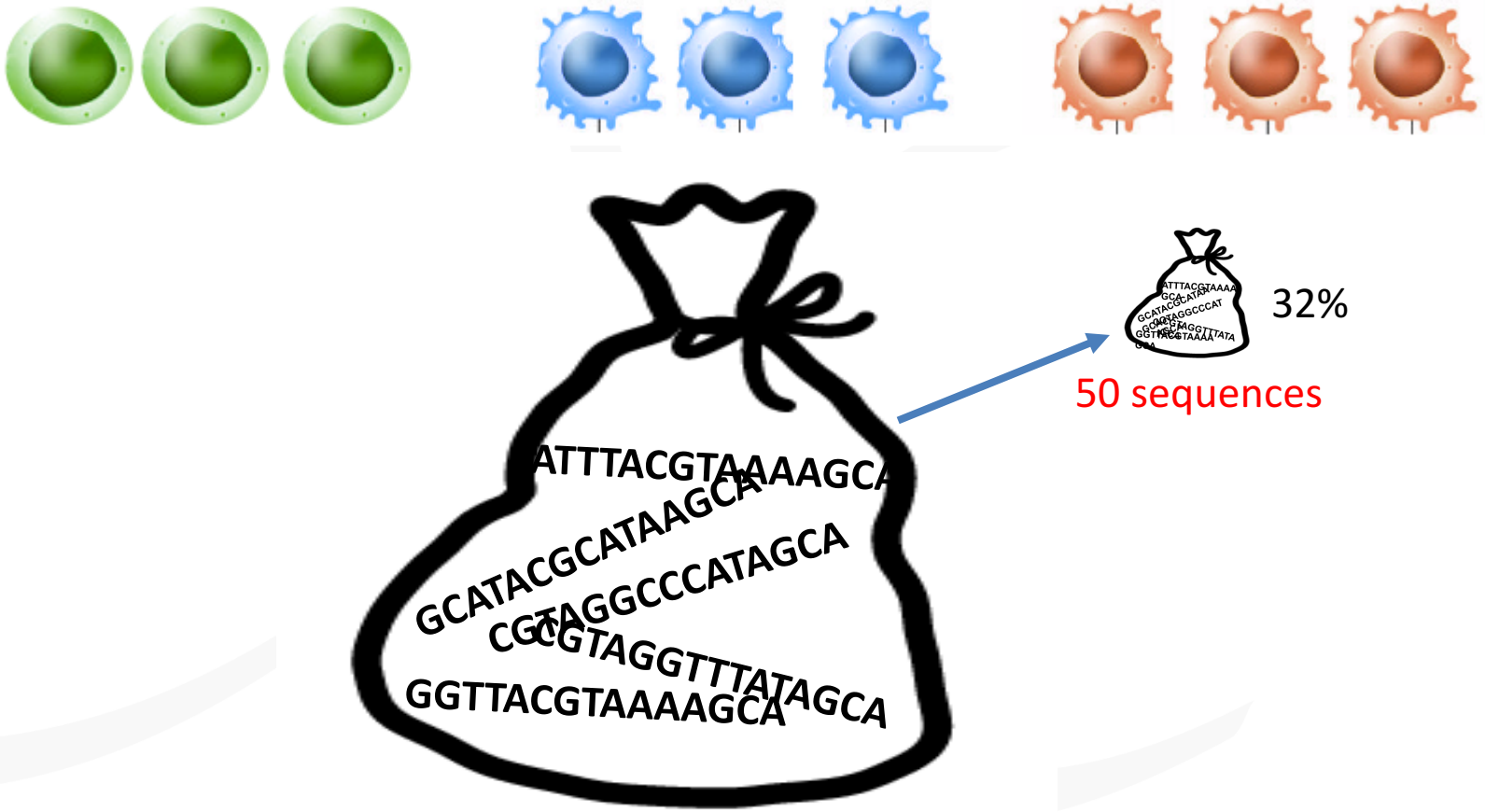

ATTTACGTAAAAGCA
GCATACGCATAAGCA
CGTAGGCCCATAGCA
GGTTACGTAAAAGCA
50 sequences

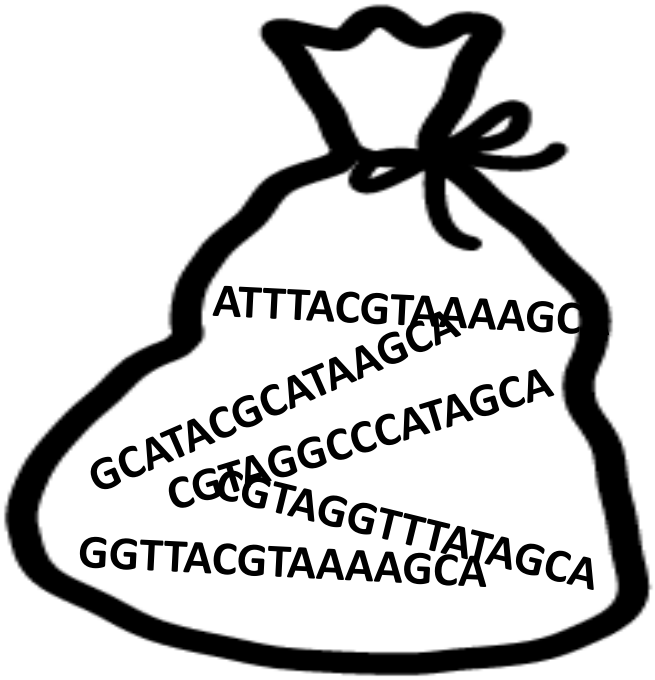
1	CCGATGGACCC	12	13	10	1	0	0	23	12	11
2	ACAATTTTTGA	16	18	11	11	9	17	15	14	12
.										
.										
.										
50	GAGGAGCTAC	121	111	181	189	170	180	123	121	111

1	CCGATGGACCC	12	13	10	1	0	0	23	12	11
2	ACAATTTTTGA	16	18	11	11	9	17	15	14	12
.										
.										
.										
50	GAGGAGCTAC	121	111	181	189	170	180	123	121	111



32%
accuracy


ATTTACGTAAAAGCA
GCATACGCATAAGCA
CGTAGGCCCATAGCA
GGTTACGTAAAAGCA
GGTAGGTTTATAGCA
32%
50 sequences


ATTTACGTAAAAGC
GCATACGCATAAGCA
CGTAGGCCCATAGCA
GGTTACGTAAAAGCA

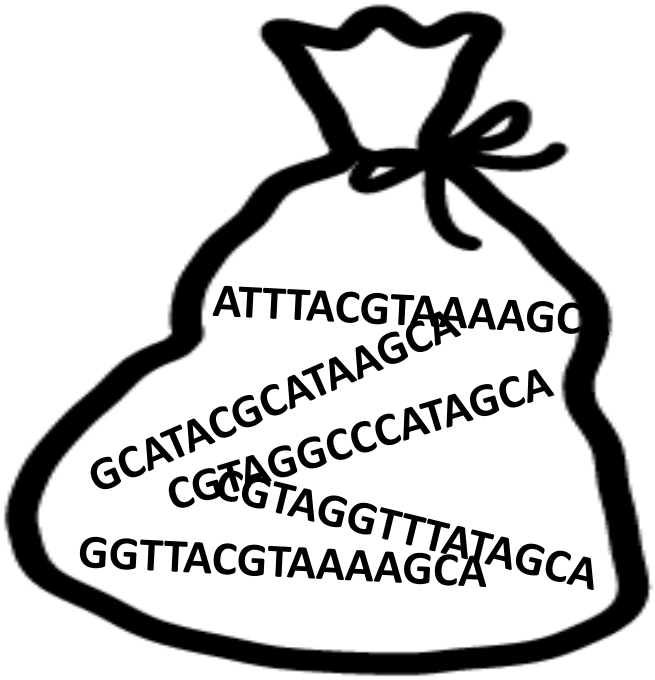
32%

51%

64%

31%

Mutation

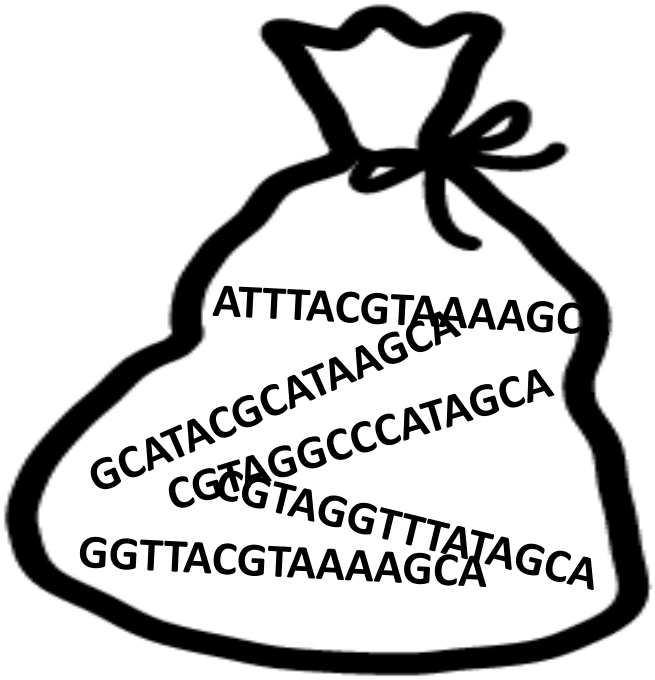


32%

51%

64%

31%



32%

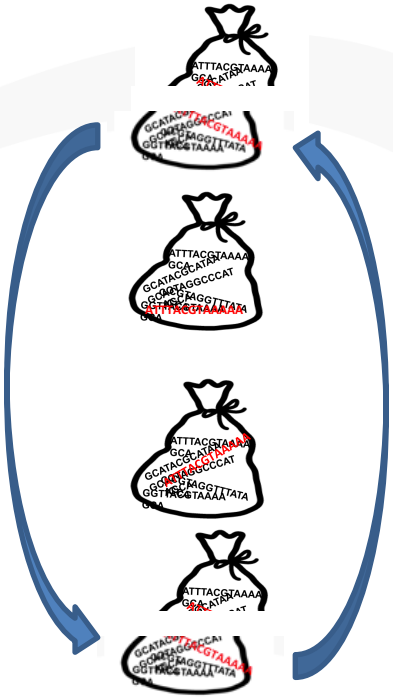
51%

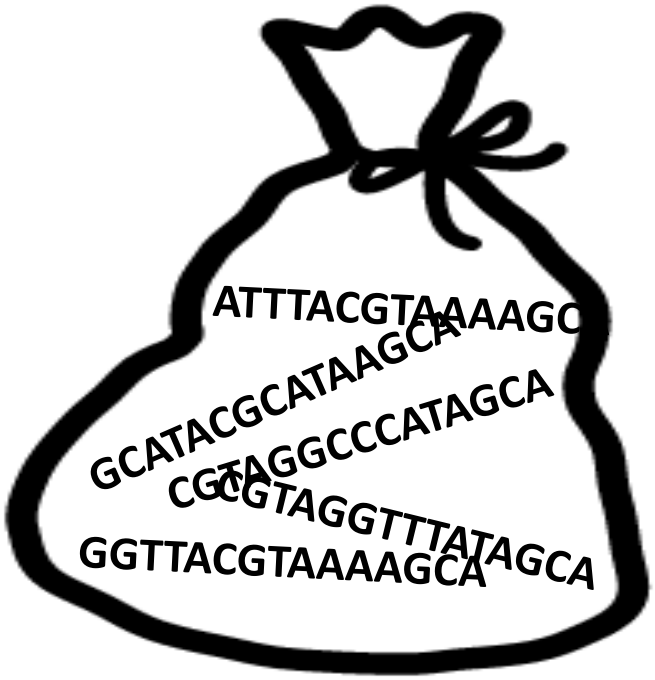
64%

31%

Mutation

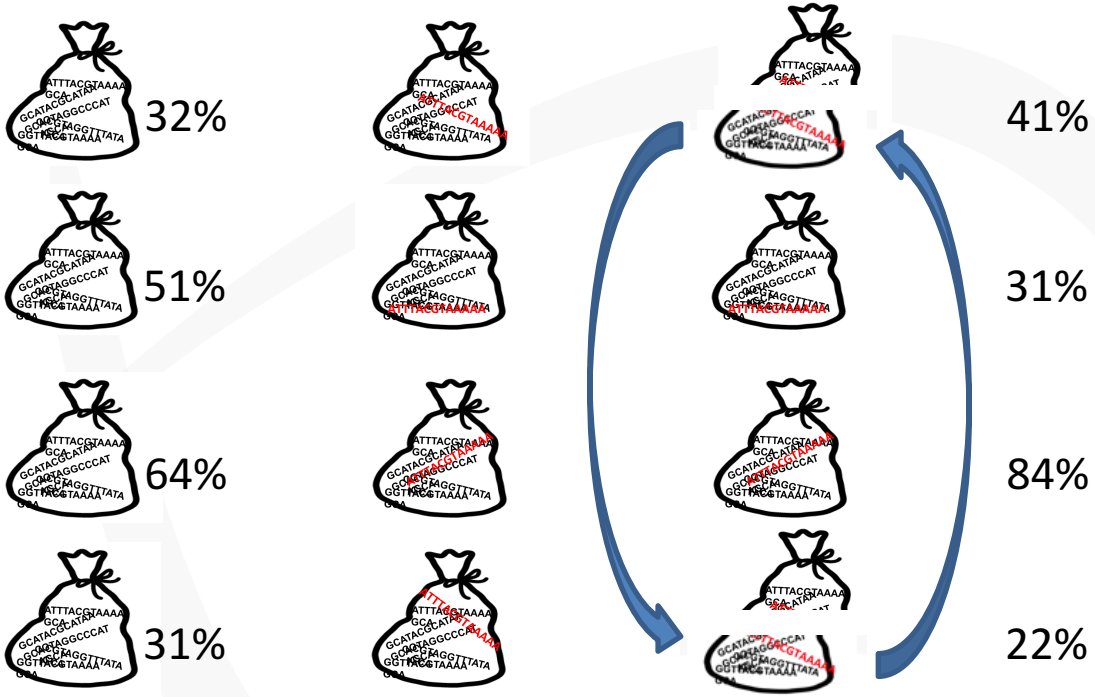
Cross-over

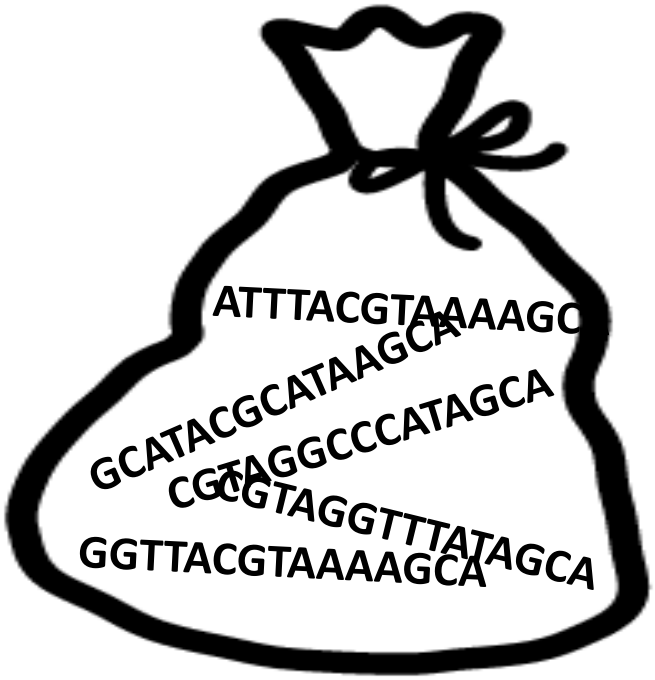




Mutation

Cross-over





Mutation

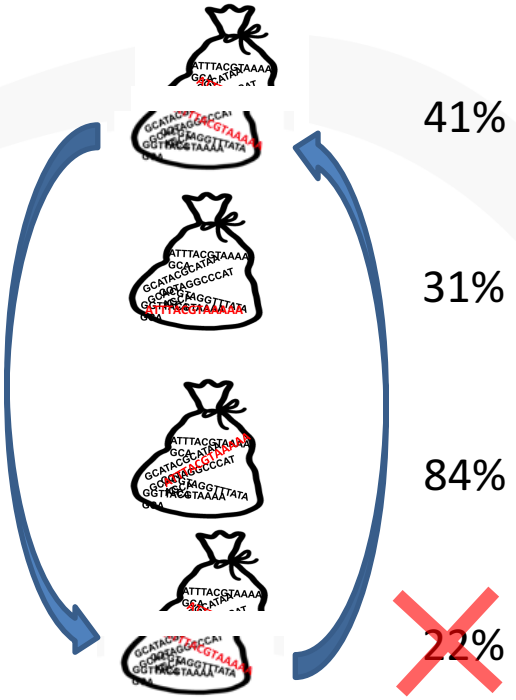
Cross-over

32%

51%

64%

31%



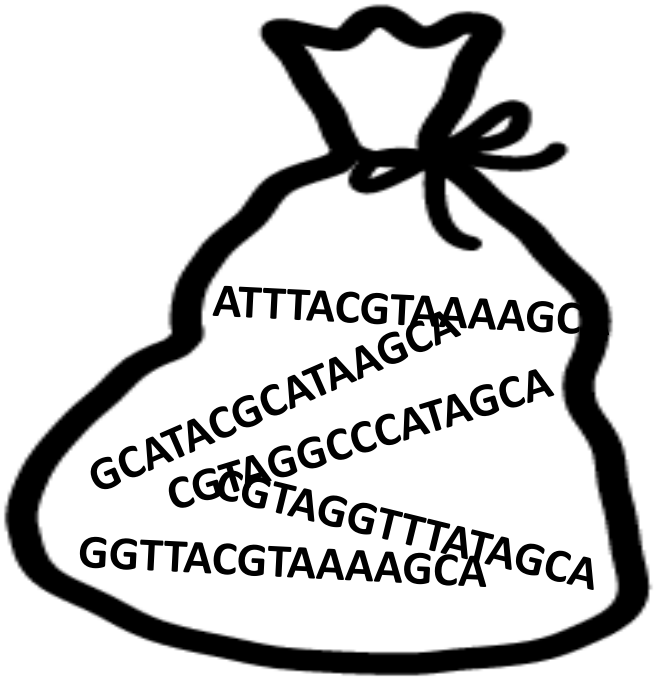
41%

31%

84%

22%

Mutation Cross-over



41%

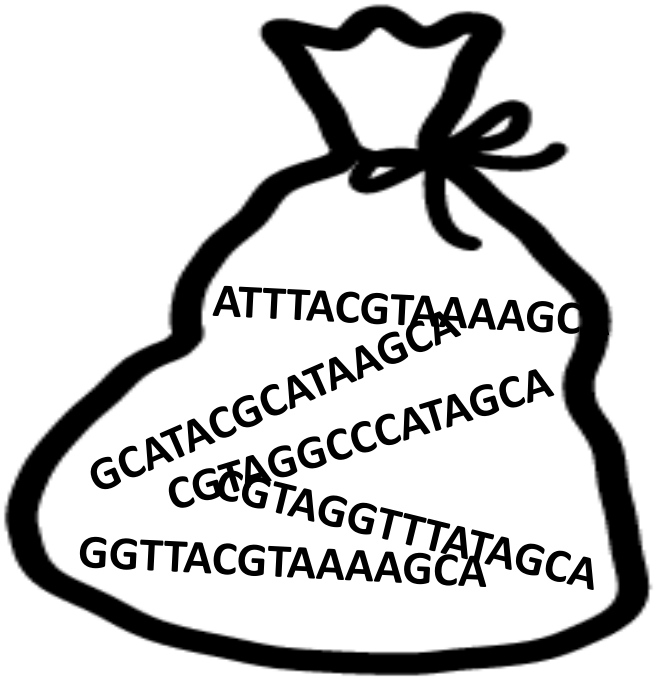
31%

84%

32%

Mutation

Cross-over

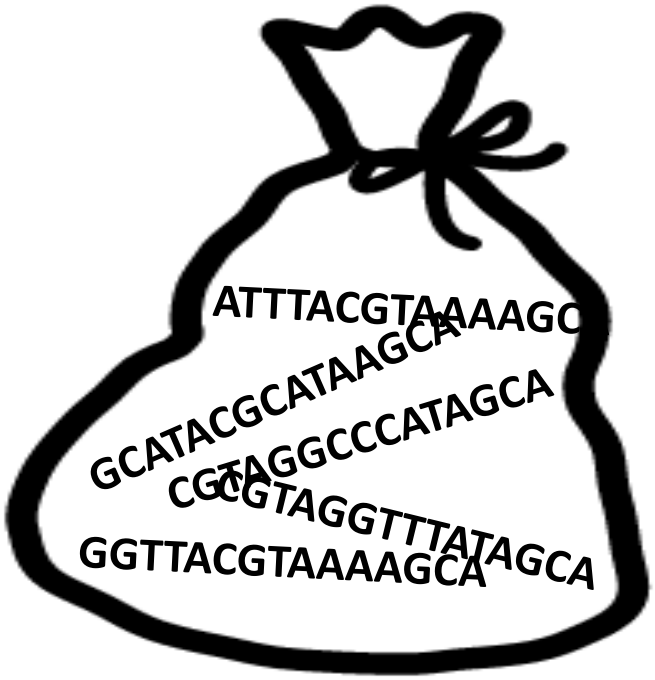


41%

31%

84%

32%



Mutation

Cross-over

41%

31%

84%

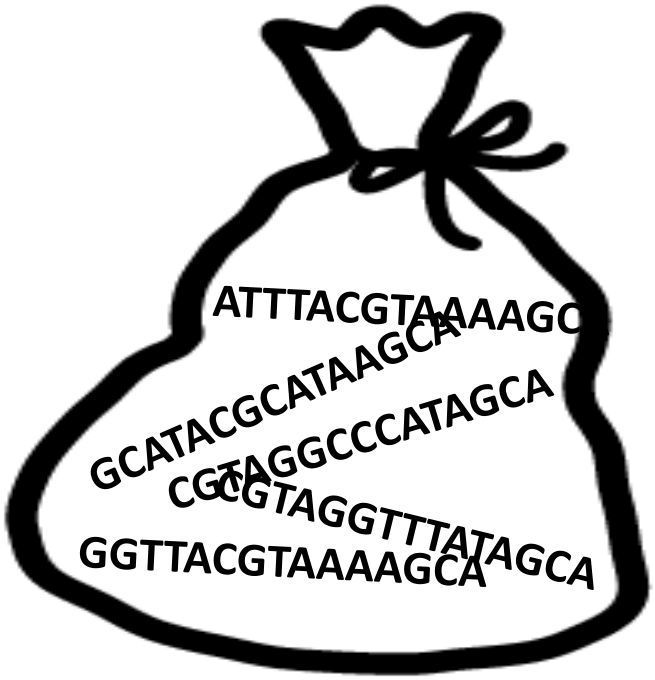
32%

71%

44%

89%

44%



Mutation

Cross-over

41%

62%

84%

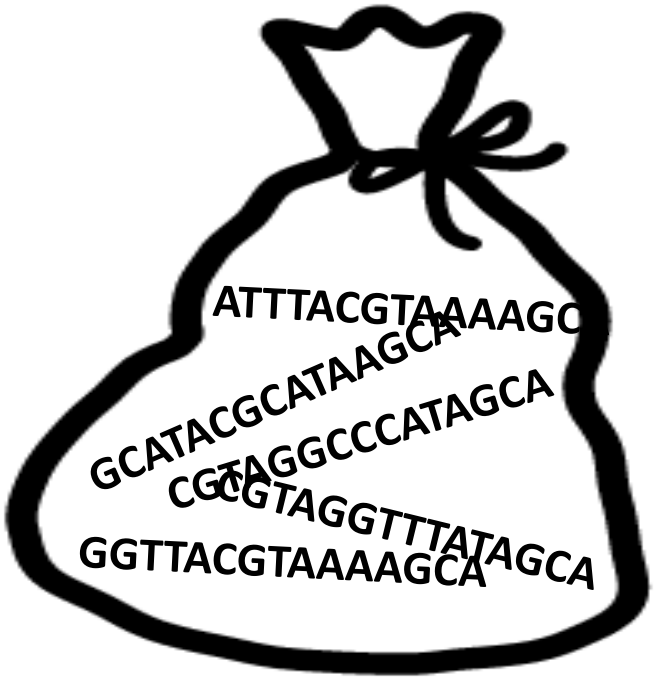
12%

71%

44%

89%

44%


ATTTACGTAAAAGC
GCATACGCATAAGCA
CGTAGGCCCATAGCA
CGTAGGTTTATAGCA
GGTTACGTAAAAGCA

99.8%

nature | ENCODE

32 normal tissues

10 sequences
100%

Title: Detecting the Presence and Progression of Premalignant Lung Lesions via Airway Gene Expression

Authors:
Jennifer Beane[1*], Sarah A. Mazzilli[1], Anna M. Tassinari[1], Gang Liu, [1] Xiaohui Zhang[1], Hanqiao Liu[1], Anne Dy Buncio[4], Samjot S. Dhillon[2], Suso Platero[3], Marc Lenburg[1], Mary E. Reid[2], Stephen Lam[4], Avrum Spira[1]

50 pre-malignant lesions
25 normal lung
280 genes

5 sequences
100%

Circulating Tumor DNA Mutation Profiling by Targeted Next Generation Sequencing Provides Guidance for Personalized Treatments in Multiple Cancer Types

Yongqian Shu, Xue Wu, Xiaoling Tong, Xiaonan Wang, Zhili Chang, Yu Mao, Xiaofeng Chen, Jing Sun, Zhenxin Wang, Zhuan Hong, Liangjun Zhu, Chunrong Zhu, Jun Chen, Ying Liang, Huawu Shao & Yang W. Shao

Lung
Pleural fluid
Stomach
Lymph node
Colon
liver

10 sequences
83%

AI & Gene Regulation