



FunkyCells ToolBox

A hands-on practical course

Martin LARSEN

www.FunkyCells.com

INSERM U1135, CHU Pitié-Salpêtrière, Paris, France

TCR analysis

1. Sequence Quality

1. Trimming of bad sequence data

```
>H7U1JGS01AA0ZD LEN=157 QL=15 QR=171
ncctatcccctgtgtgccttggcagtcctcagAGCTCTGAGCTGAATGTGAACGCCTTGTTGCTGGGGGACTCGGCCCTCTATCT
CTGTGCCAGCAGCTTGGCTCCGAGGGCCTACAATGAGCAGTCTTGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01ABGGZ LEN=162 QL=15 QR=176
cctatcccctgtgtgccttggcagtcctcagAGCTCTGAGCTGAATGTGAACGCCTTGTTGCTGGGGGACTCGGCCCTCTATCTC
TGTGCCAGCAGCCGCAGCTGGGCTAGCGGAAGCAATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01AZQV0 LEN=160 QL=15 QR=174
cctatcccctgtgtgccttggcagtcctcagAGCTCTGAGCTGAATGTGAACGCCTTGGAGCTGGAGGACTCGGCCCTGTATCTC
TGTGCCAGCAGCTTGGAACTAGCGGGGCCGTACGAGCAGTACTTCGGGCCCGGGCACCAGGCTCACGGTCACA
>H7U1JGS01AB892 LEN=162 QL=15 QR=176
cctatcccctgtgtgccttggcagtcctcagAGCTCTGAGCTGAATGTGAACGCCTTGGAGCTGGACGACTCGGCCCTGTATCTC
TGTGCCAGCAGCAGGGTAGCGGGAAGAGGGGCCAATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01AFBGF LEN=153 QL=15 QR=167
cctatcccctgtgtgccttggcagtcctcagAGCTCTGAGCTGAATGTGAACGCCTTGGAGCTGGACGACTCGGCCCTGTATCTC
TGTGCCAGCAGTAAGGACCAGAGTAATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01AAUP3 LEN=288 QL=15 QR=302
CCTATCCCCTCCGTTAGTTCCCGTCCGTTTGGGTCGACCGTTTCGGTCCAAGGTACGTCCTACGTAGGACGTCCTTGGAAAGT
CGTTGGAAAACGTCCGTTTGGTTAACGGCCCTTTGGGGGTTACGTCCGGTCCCGGGGGTACGTTACGGTCCCCTTCCTTAGTTC
GTCCCTAGGTCGACCGACGTTCCGGTAGGCTACGGGACGGAACGGTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCA
CA
```

Sequence Quality

Funky Cells Sequence Import

Data Selection | **Sequence Quality** | Trim ends

Select Fasta file: FASTA file

No file selected

Select Quality file: QUALITY file

No file selected

Select category field including sequence names corresponding to imported files:

Partial match Match selected sequences with Fasta sequences ▼

Select field including sequence to identify in corresponding imported files:

NA ▼

HELP ADVANCED CANCEL OK

Sequence Quality

Funky Cells Sequence Import

Data Selection | **Sequence Quality** | Trim ends

Output quality of full Fasta sequence (based on Phred Quality Score)

Phred-threshold: >20 (Base call accuracy 99%) ▼

Output quality of selected sequence (based on Phred Quality Score)

HELP ADVANCED CANCEL OK

Sequence Quality

Funky Cells Sequence Import

Data Selection | Sequence Quality | Trim ends

Trim procedure searches the first sequence match from the specified end.

5' End Trim:

3' End Trim:

Max sequence length::

Reverse complement Fasta sequences before trimming

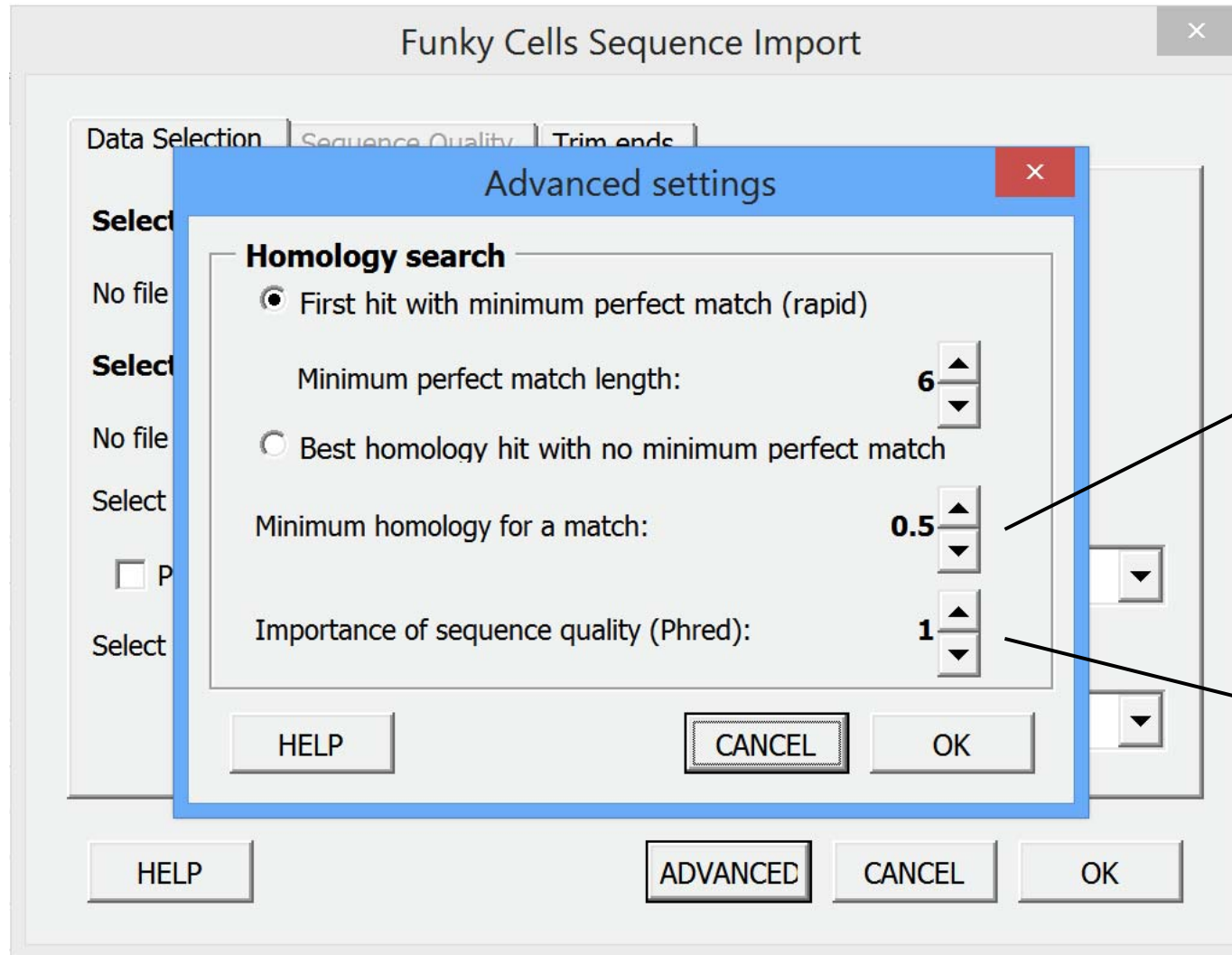
Export trimmed sequences and Phred-quality data as Fasta files

Export filename:

No file selected

HELP ADVANCED CANCEL OK

Sequence Quality



Minimum frequency of correct bases

Range: [0:10]
Corrects homology match based on the quality of the sequence. Badly sequenced nucleotides count least when value is high.

Next Generation Sequence Files

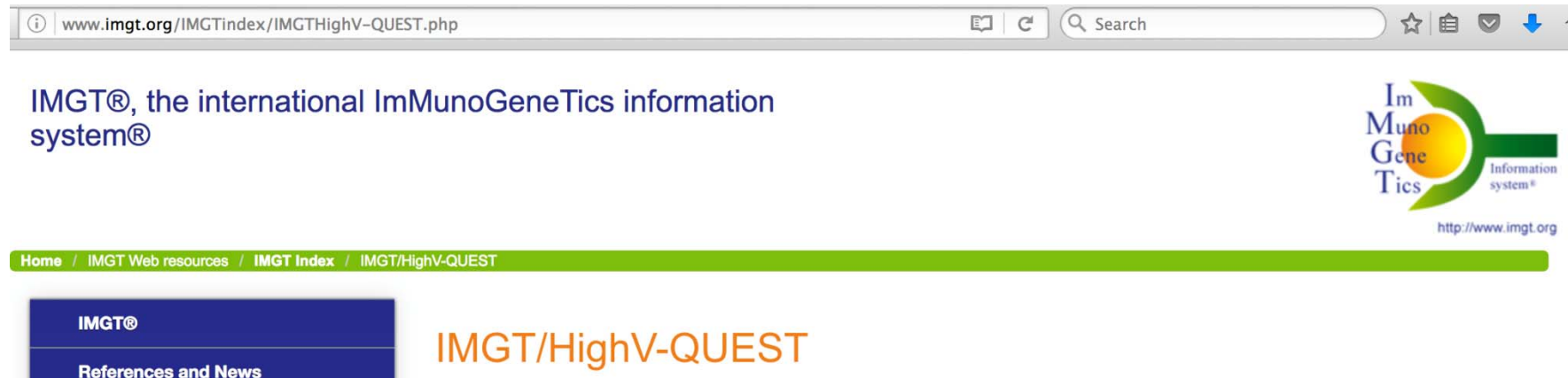
```
>H7U1JGS01AA0ZD LEN=157 QL=15 QR=171
ncctatcccctgtgtgccttggcagctctcagAGCTCTGAGCTGAATGTGAACGCCTTGTGCTGGGGGACTCGGCCCTCTATCT
CTGTGCCAGCAGCTTGGCTCCGAGGGCCACAATGAGCAGTCTTGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01ABGGZ LEN=162 QL=15 QR=176
cctatcccctgtgtgccttggcagctctcagAGCTCTGAGCTGAATGTGAACGCCTTGTGCTGGGGGACTCGGCCCTCTATCTC
TGTGCCAGCAGCCGAGCTGGCTAGCGGAAGCAATGAGCAGTCTTTCGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01AZQV0 LEN=160 QL=15 QR=174
cctatcccctgtgtgccttggcagctctcagAGCTCTGAGCTGAATGTGAACGCCTTGGAGCTGGAGGACTCGGCCCTGTATCTC
TGTGCCAGCAGCTTGAAGTACGCGGGCCGTACGAGCAGTACTTCGGGCCCGGACACAGGCTCACCGTGCTA
>H7U1JGS01AB892 LEN=162 QL=15 QR=176
cctatcccctgtgtgccttggcagctctcagAGCTCTGAGCTGAATGTGAACGCCTTGGAGCTGGAGGACTCGGCCCTGTATCTC
TGTGCCAGCAGCAGGGTAGCGGAAGAGGGGCCAATGAGCAGTCTTTCGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01AFBGF LEN=153 QL=15 QR=167
cctatcccctgtgtgccttggcagctctcagAGCTCTGAGCTGAATGTGAACGCCTTGGAGCTGGAGGACTCGGCCCTGTATCTC
TGTGCCAGCAGTAAGGACAGAGTAATGAGCAGTCTTTCGGGCCAGGGACACGGCTCACCGTGCTA
>H7U1JGS01AAUP3 LEN=288 QL=15 QR=302
CCTATCCCTCCGTTAGTTCCCGTCCGTTTGGGTTCGACCGTTTTCGGTCAAGGTAAGTACGCTCTAGTAGGACGCTCCTGAAAGT
CGTTGAAAACGTCGTTTGGTTAAGCGCCCTTGGGGGTACGTCGGTCCCGGGGTACGTTACGGTCCCTTCCCTTAGTTC
GTCCCTAGGTCGACCGACGTTTCGTAGGCTACGGGACGGAACGGTACGAGCAGTACTTCGGGCCCGGACACAGGCTCACGGTCA
CA
```

```
>H7U1JGS01AA0ZD LEN=157 QL=15 QR=171
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 35 33 33 30 37 37 24 16 16 15
15 22 17 24 24 26 26 30 28 37 37 37 37 33 33 32 32 28 28 26 29 18 26 26 29 30 24 24
24 24 24 28 37 37 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
33 21 24 12 12 12 14 14 21 21 21 21 21 21 23 23 25 35 35 37 36 38 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 37 37
35 35 34 34 26 22 22 19 20 26 25 25 23 0
>H7U1JGS01ABGGZ LEN=162 QL=15 QR=176
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
39 39 36 36 33 40 40 40 29 29 29 29 27 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 39 35 35 40 40 36 36 27 21 16 16 16 16 21 25 30 29 29
>H7U1JGS01AZQV0 LEN=160 QL=15 QR=174
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 31 31 31 40 40 28 28 28 36
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 39 39 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 35 35 36 36 40 38 36 21 21
>H7U1JGS01AB892 LEN=162 QL=15 QR=176
39 40 40 40 40 35 36 30 35 26 35 31 38 38 34 37 27 24 17 17 17 23 30 33 33 30 30 29
38 38 37 40 40 40 40 40 40 40 40 40 37 37 37 38 33 31 27 25 25 21 21 13 19 19 16 16 21
18 30 26 27 27 28 32 36 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 33 21
21 21 35 35 35 39 39 40 37 37 37 37 40 38 38 38 28 28 26 26 26 32 29 32 32 32
37 37 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39
39 39 40 40 40 40 40 40 40 40 37 37 35 35 39 37 37 37 40
>H7U1JGS01AFBGF LEN=153 QL=15 QR=167
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 31 31 31 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 35 37 35 35 39 40 40 40
>H7U1JGS01AAUP3 LEN=288 QL=15 QR=302
36 40 36 40 40 40 40 40 40 40 40 40 40 36 36 40 36 36 36 36 34 34 34 36 33 31 31 31 33
29 28 24 23 23 23 29 29 19 14 15 14 15 16 11 11 11 11 11 11 16 16 17 17 17 16 11
11 11 11 15 18 15 14 14 18 14 15 15 21 25 21 21 21 25 26 23 21 18 18 18 21 21 21 21
25 26 20 23 23 24 26 21 21 21 21 15 15 15 16 25 25 26 29 30 22 31 27 27 29 27 21
13 13 13 13 13 18 18 23 18 13 13 21 23 31 31 32 37 34 30 26 28 28 20 20 19 19 19 22
22 28 26 26 26 30 30 30 30 22 22 14 14 13 18 23 25 23 19 19 11 11 12 15 15 21 21
17 19 19 21 21 21 21 21 27 27 27 27 27 26 26 24 24 24 27 31 31 31 31 29 29 28 28
29 29 29 31 31 30 30 30 29 29 29 31 29 24 18 18 18 29 30 31 31 31 29 26 21 21 21 21
21 25 18 18 16 15 12 12 12 12 15 13 20 20 20 21 21 21 16 18 12 12 12 12 18 17 21
19 19
```

TCR analysis

1. Sequence Quality

1. Trimming of bad sequence data
2. Blast exported data (bioinformatics) – external analysis
(<http://www.imgt.org/IMGIndex/IMGTHighV-QUEST.php>)



The screenshot shows a web browser window with the address bar containing www.imgt.org/IMGIndex/IMGTHighV-QUEST.php. The page header includes the text "IMGT®, the international ImMunoGeneTics information system®" and the IMGT logo. A green navigation bar contains the links "Home / IMGT Web resources / IMGT Index / IMGT/HighV-QUEST". Below this, a blue button labeled "IMGT®" is positioned above a button labeled "References and News". The main heading of the page is "IMGT/HighV-QUEST" in orange text.

Or other external software!

TCR analysis

1. Sequence Quality

1. Trimming of bad sequence data
2. Blast exported data (bioinformatics) – external analysis
(<http://www.imgt.org/IMGTindex/IMGTHighV-QUEST.php>)
3. Pivot tables to aggregate data
 1. Clonotypes (nucleotide sequence)
 2. Clonotype distribution (singletons versus polytons)
4. Identify quality of a specific sequence

2. Peptide analysis

1. Pivot table to extract clonotypes (AA)
2. Physico-chemical properties

3. Similarity and Diversity

1. Join data (MID2 and MID2)
2. How to compare data from multiple samples

Similarity and Diversity Analyzer

Similarity and Diversity Analyzer

Similarity Analysis

- Overlap Count
- Morisita's overlap index
- Simpson's index
- Horn's modification

Diversity Analysis

- Shannon Entropy
- Normalized
- Bit (base = 2)
- Natural (base = e)
- Digit (base = 10)

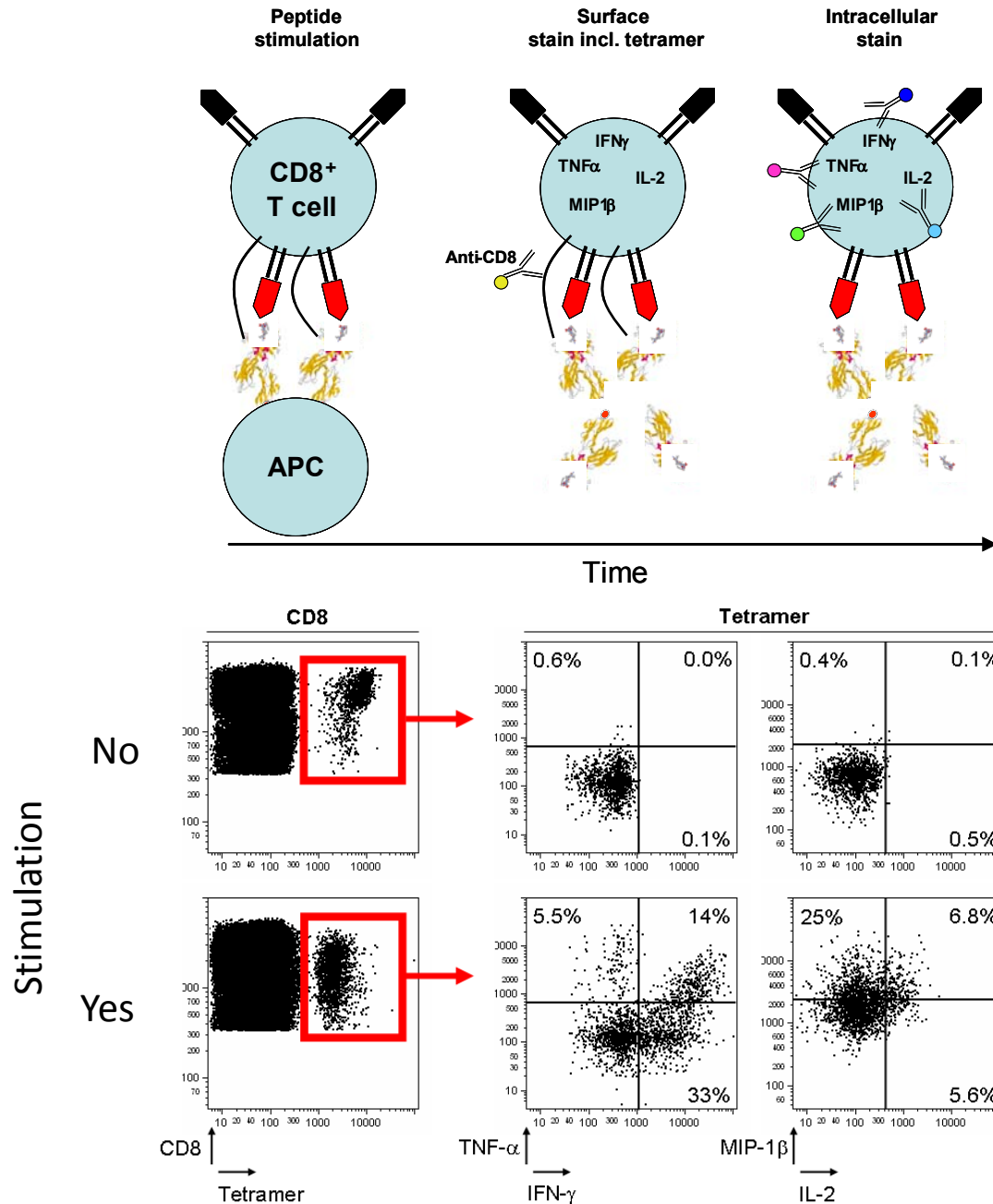
HELP CANCEL OK

Classic ecology
measures
(Google IT)

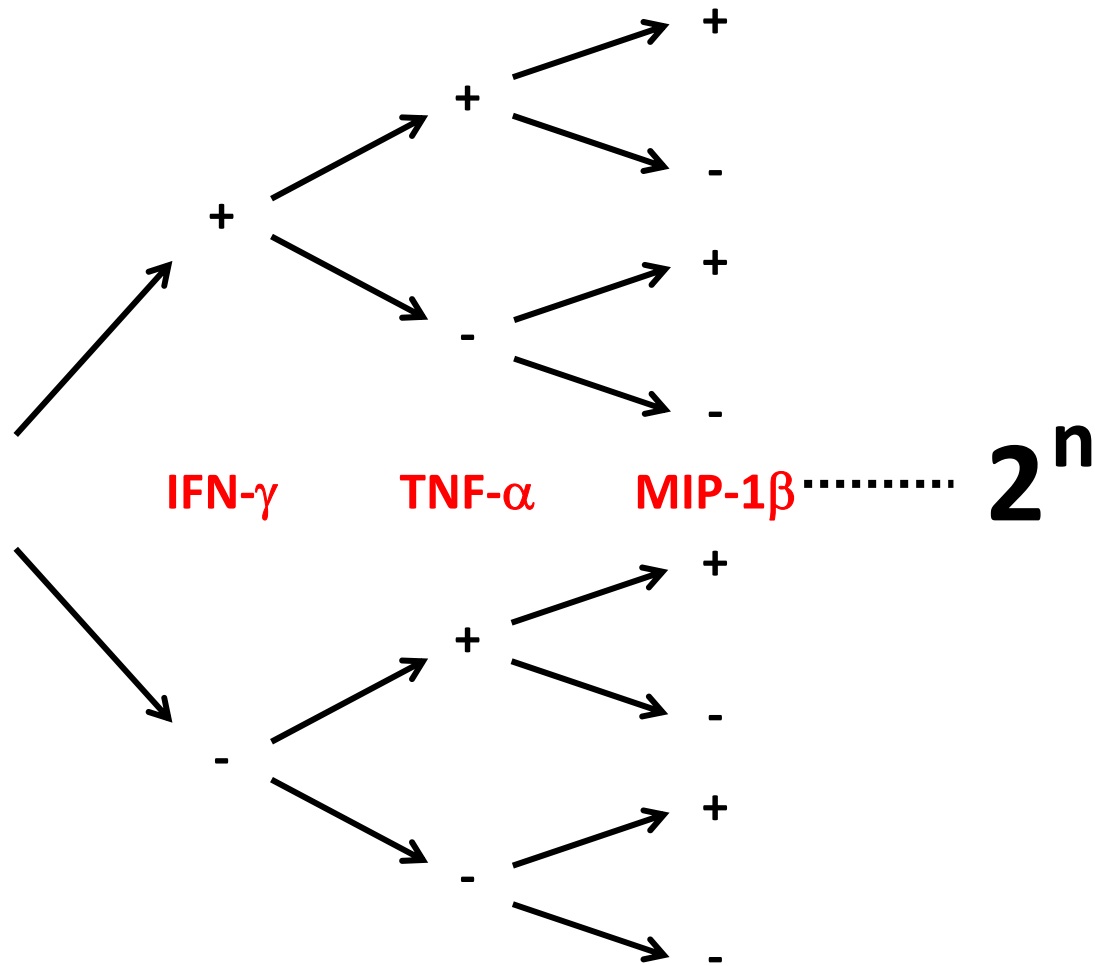
Boolean analysis

1. What is boolean gating?

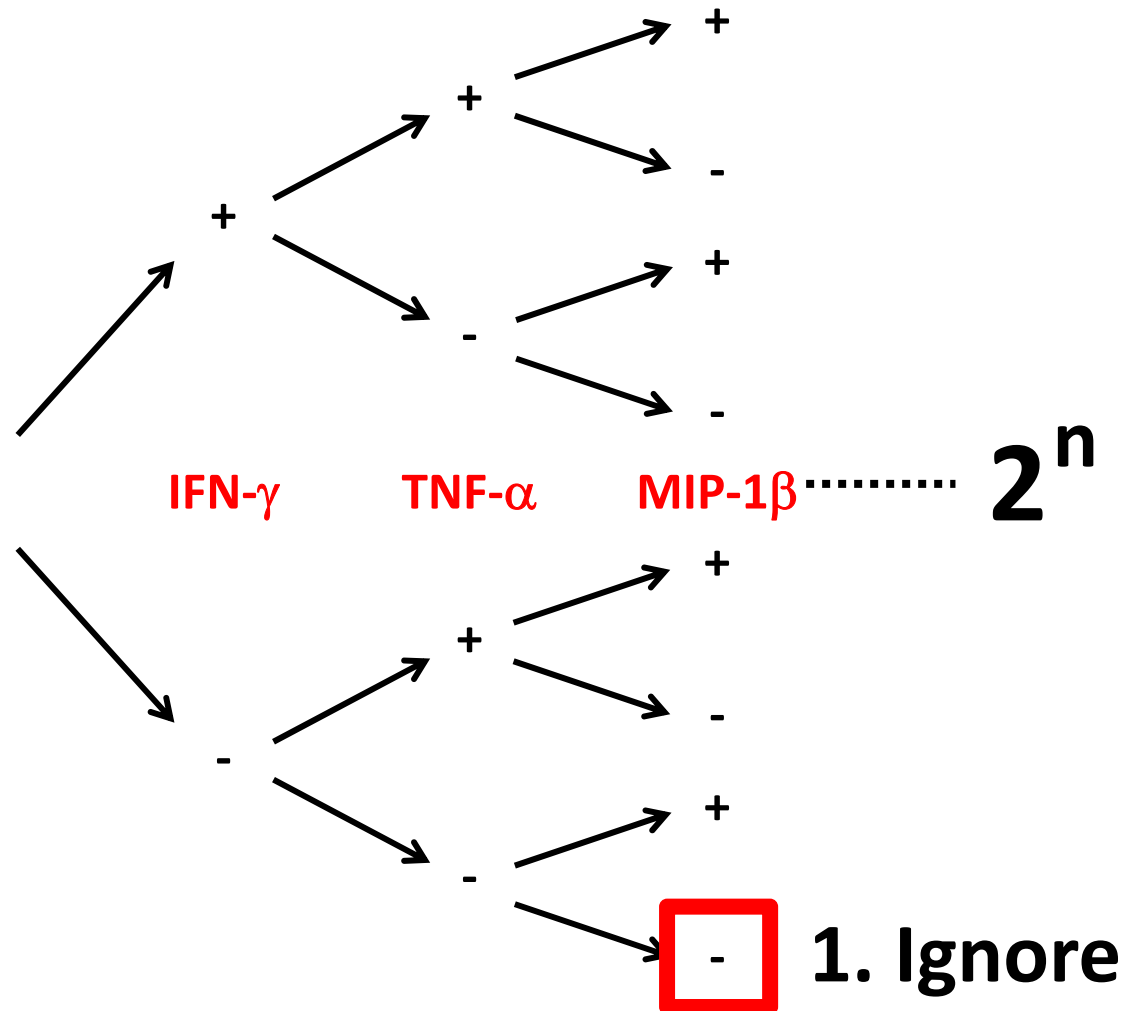
EBV-specific CD8 T cell polyfunctionality



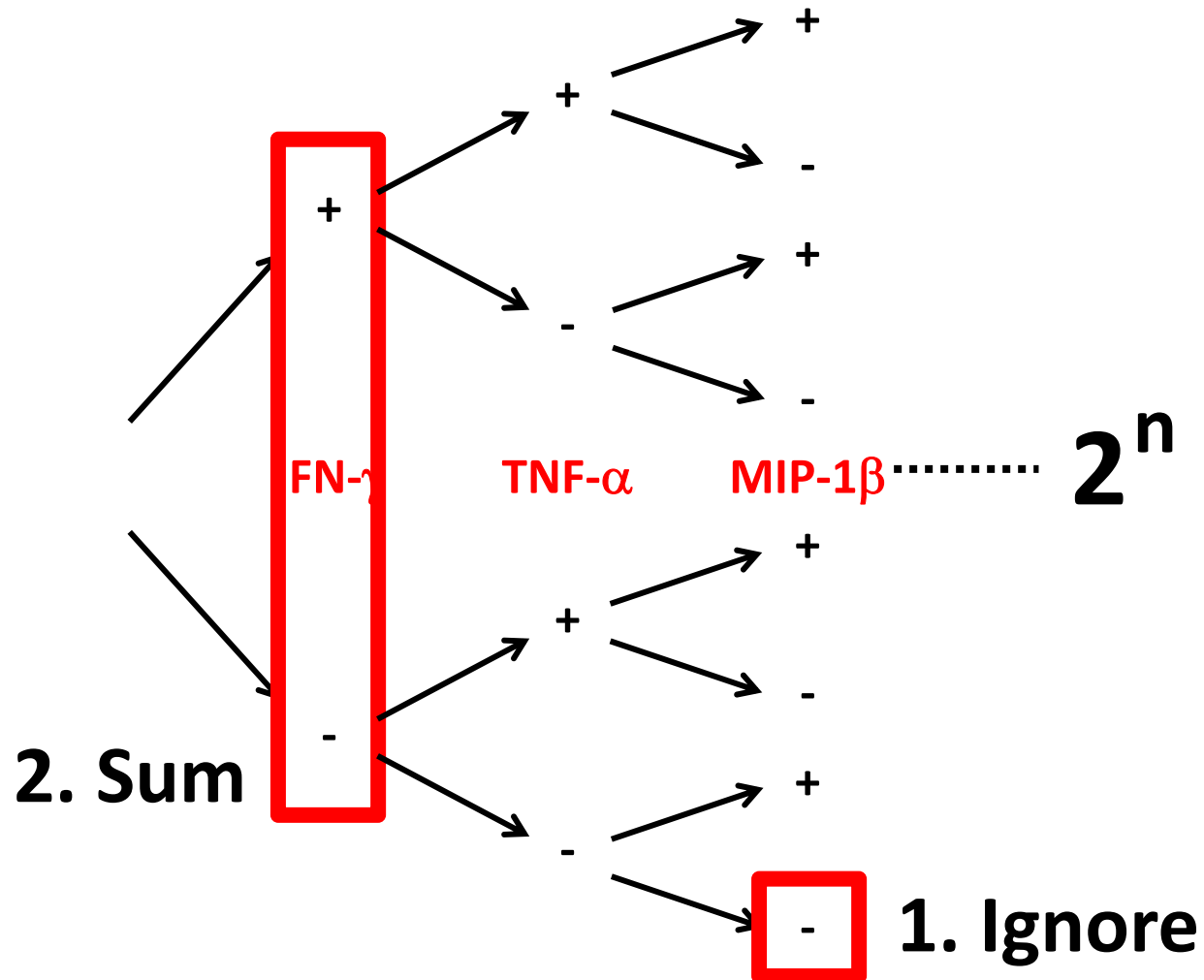
Exhaustive combinatorial boolean analysis



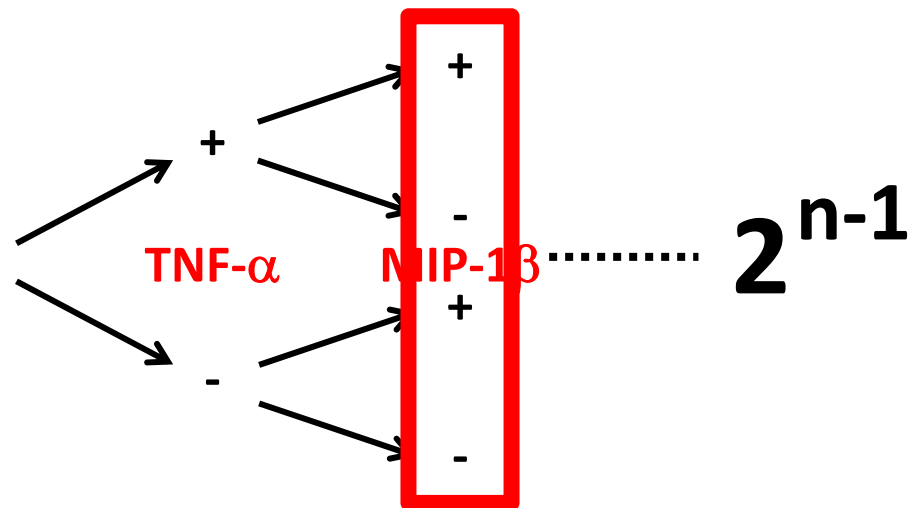
Exhaustive combinatorial boolean analysis



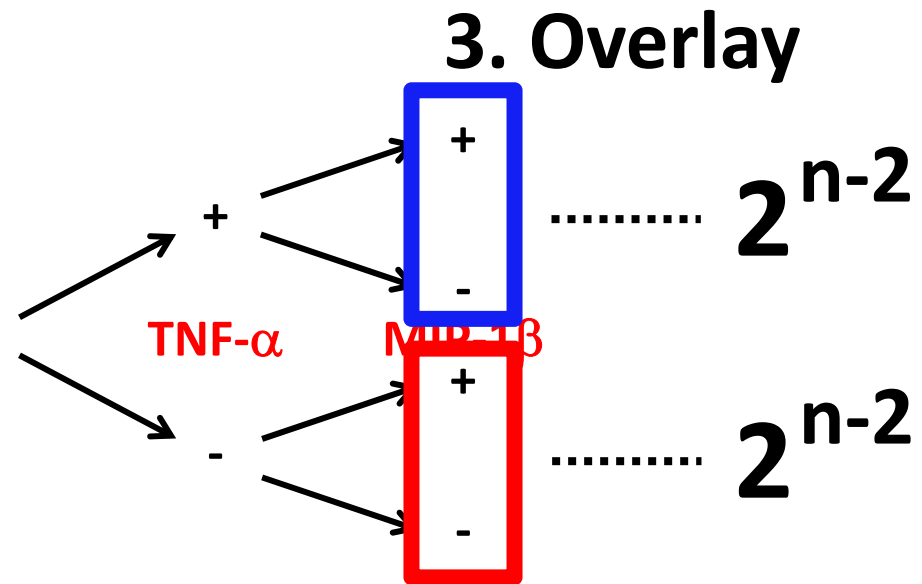
Exhaustive combinatorial boolean analysis



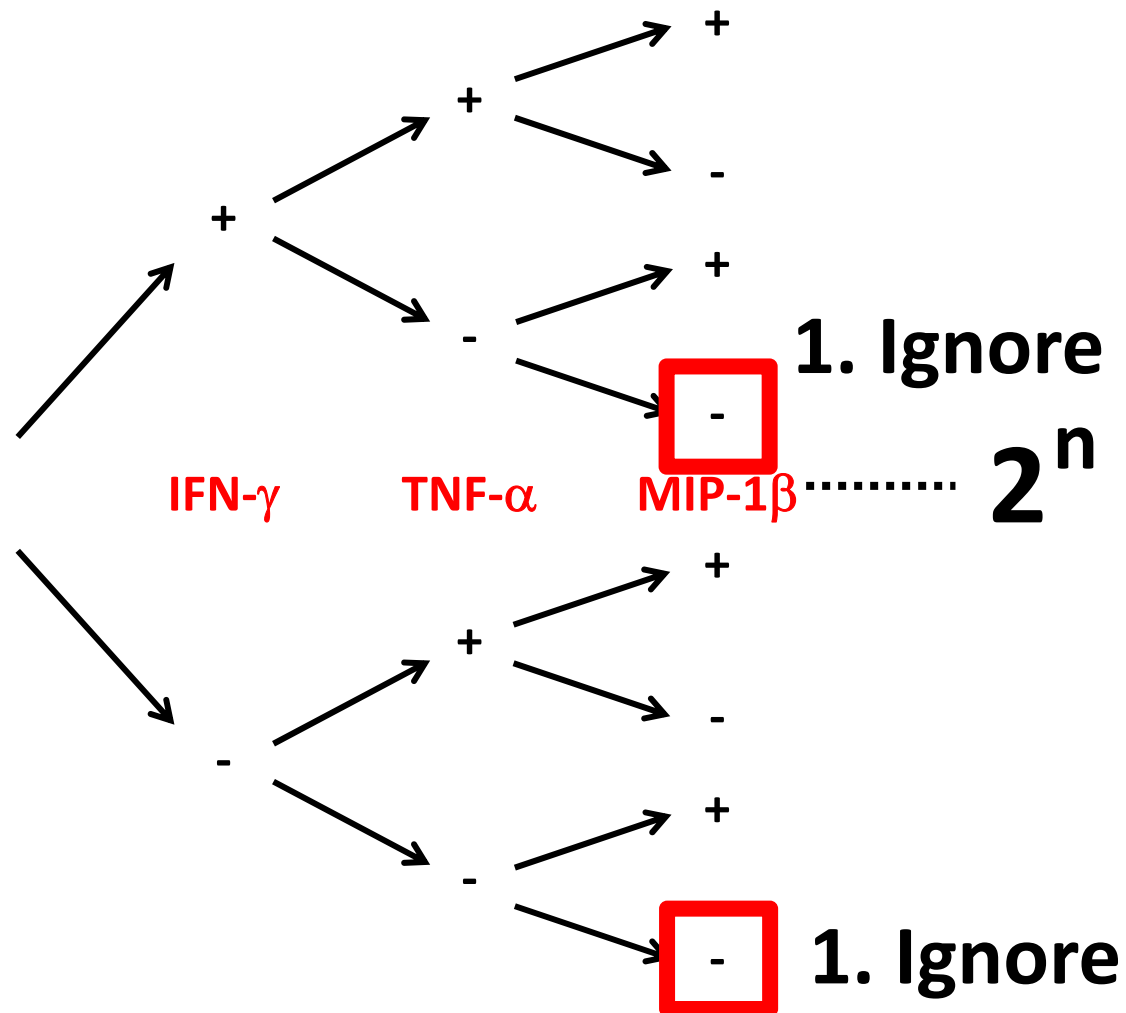
Exhaustive combinatorial boolean analysis



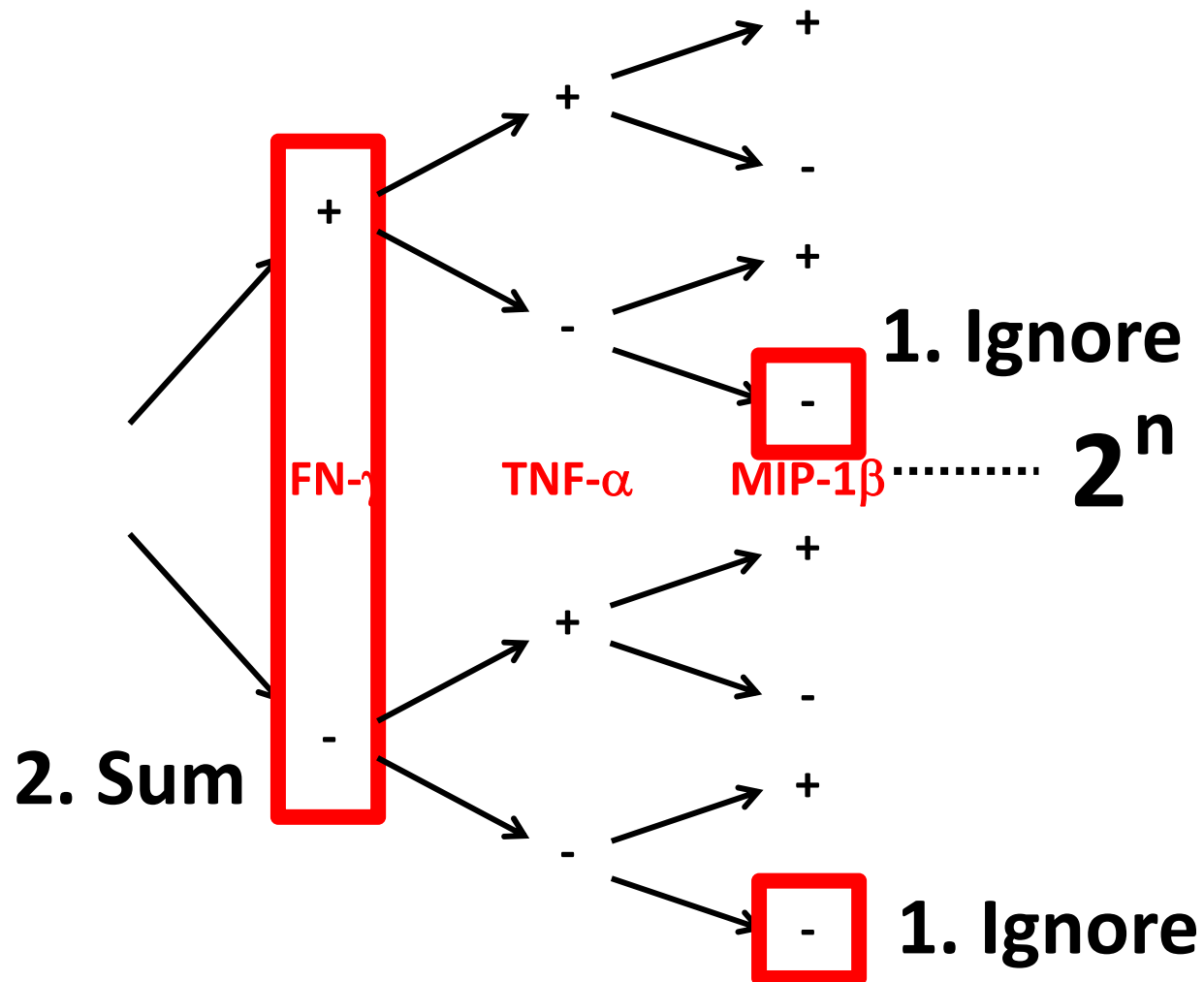
Exhaustive combinatorial boolean analysis



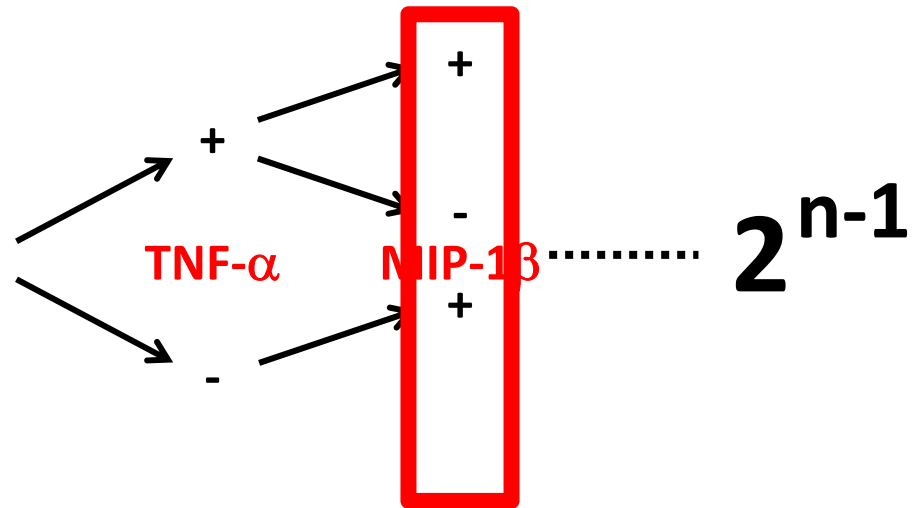
Exhaustive combinatorial boolean analysis



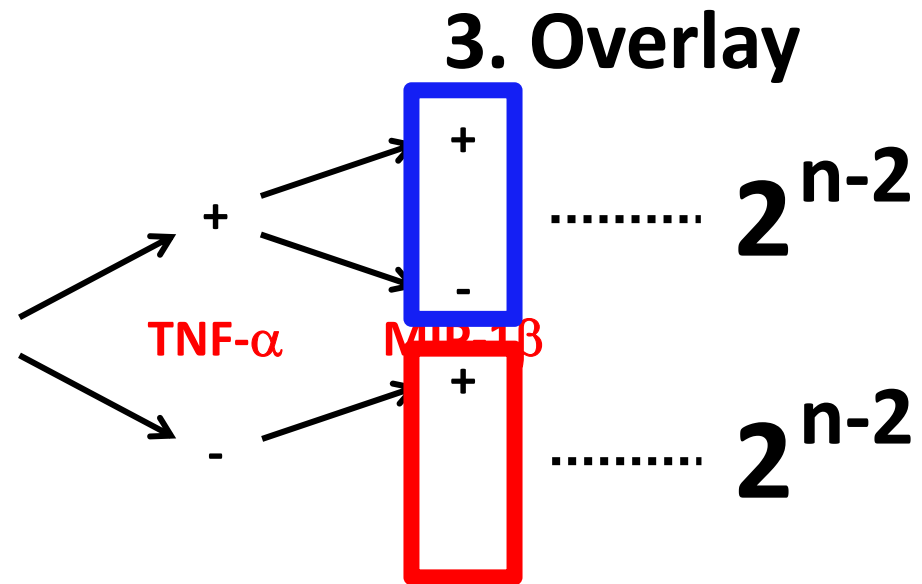
Exhaustive combinatorial boolean analysis



Exhaustive combinatorial boolean analysis



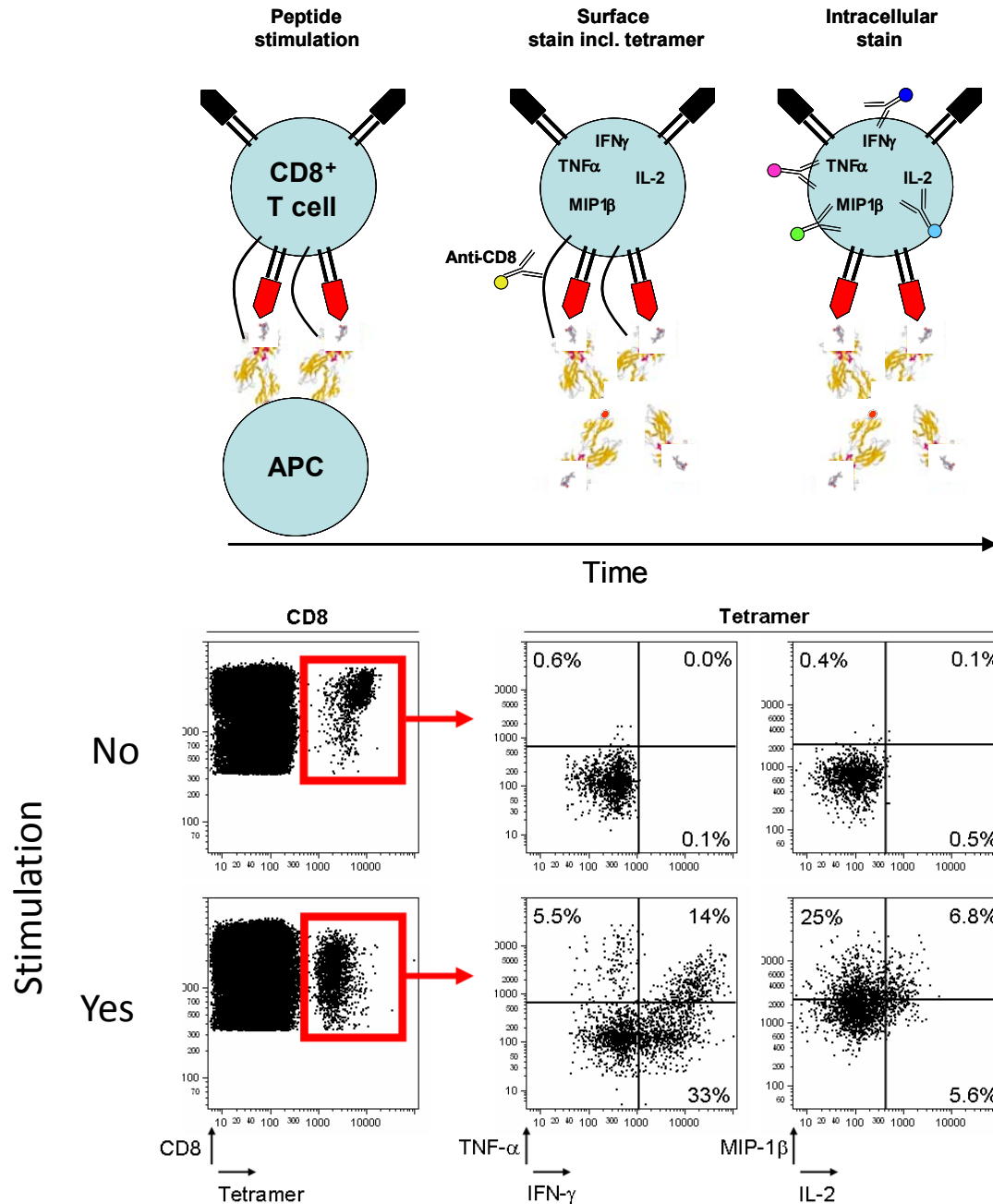
Exhaustive combinatorial boolean analysis



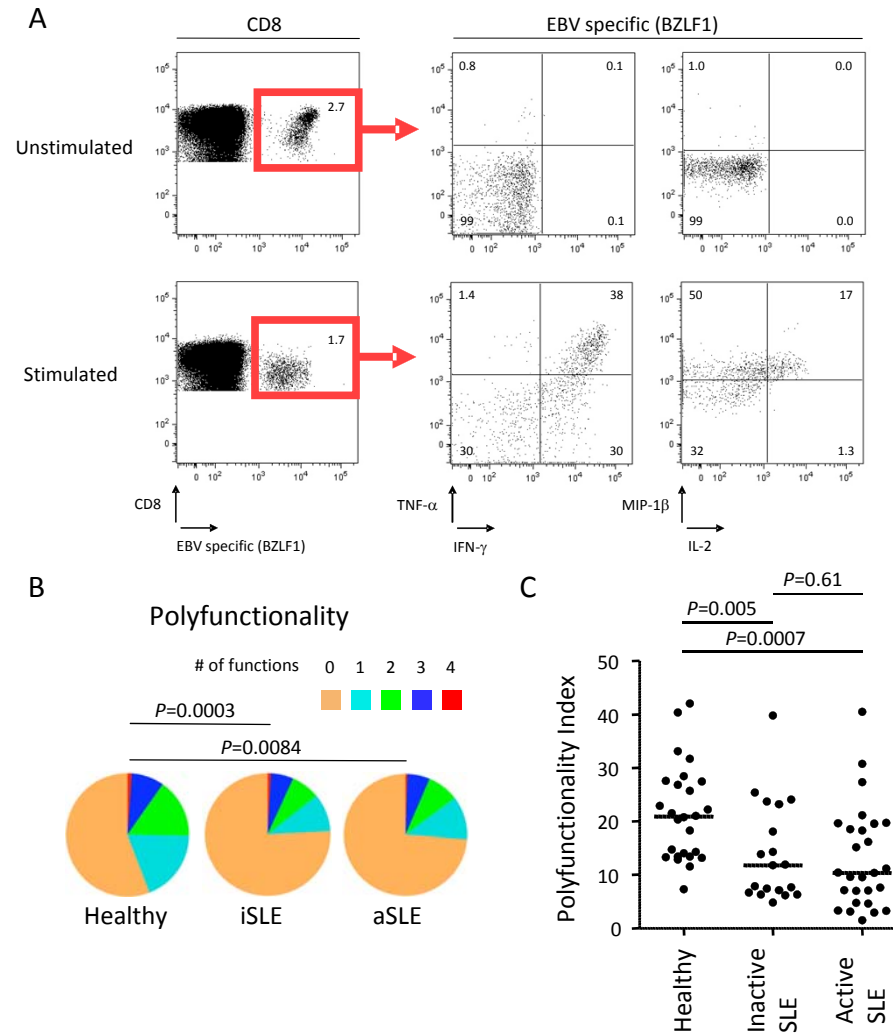
Boolean analysis

1. What is boolean gating?
2. Polyfunctionality analysis
 1. Boolean data miner
 2. Find optimal parameter estimates
 1. Track Function
 2. Bootstrapping

EBV-specific CD8 T cell polyfunctionality



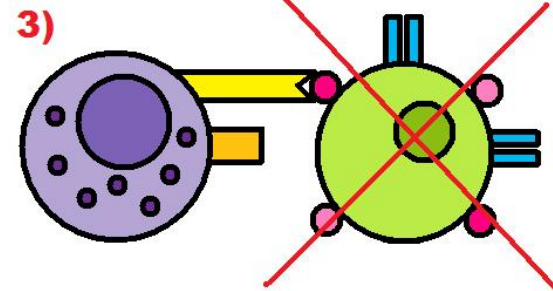
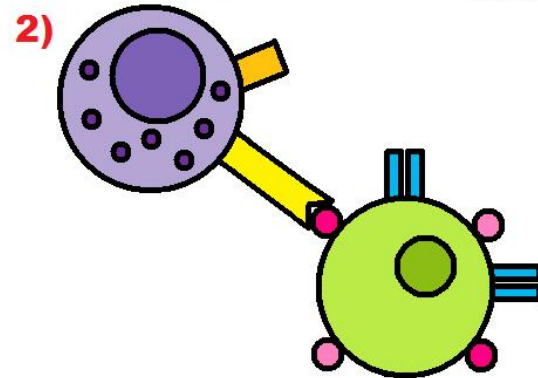
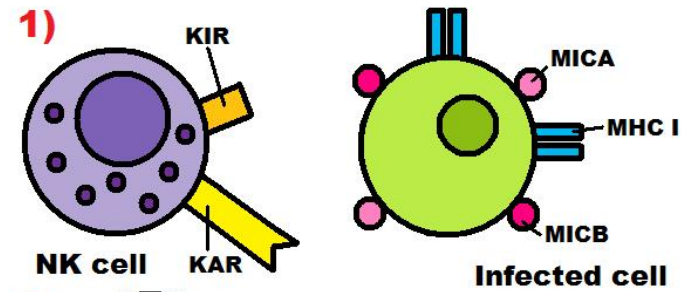
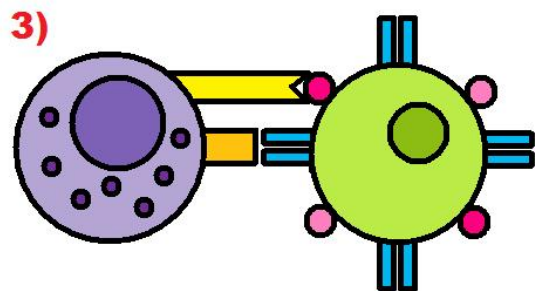
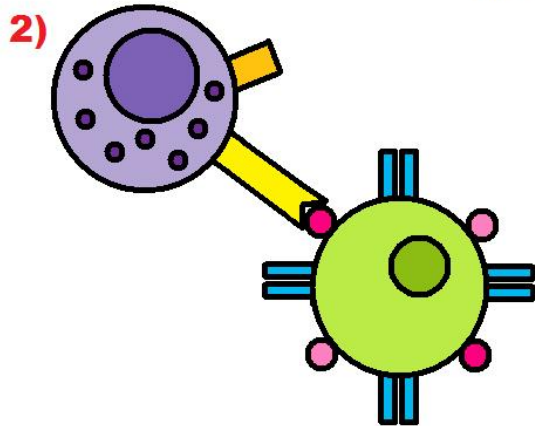
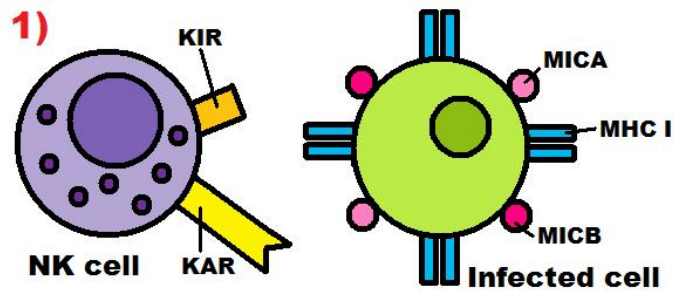
EBV-specific CD8 T cell polyfunctionality



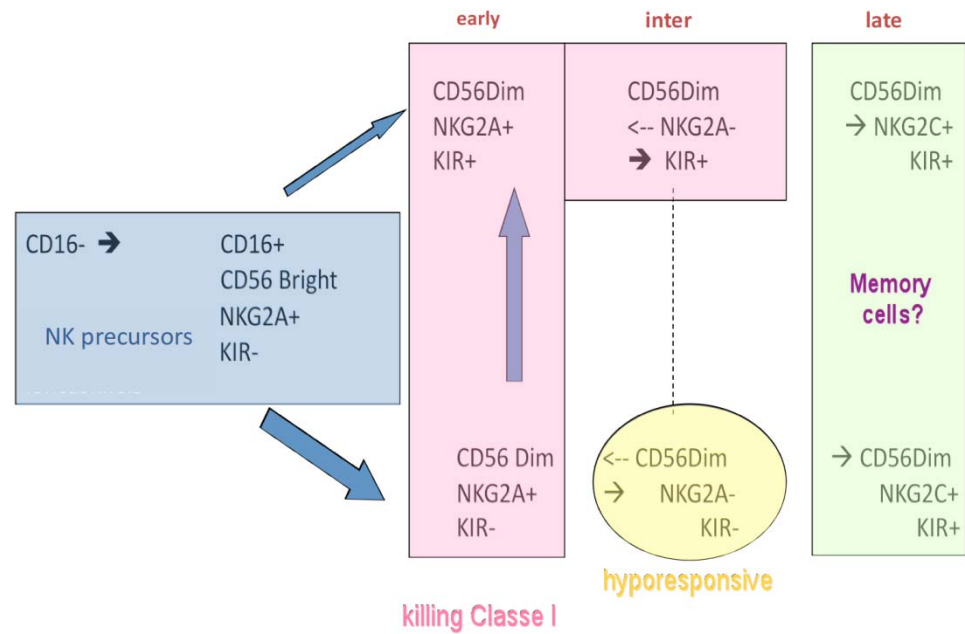
Boolean analysis

1. What is boolean gating?
2. Polyfunctionality analysis
 1. Boolean data miner
 2. Find optimal parameter estimates
 1. Track Function
 2. Bootstrapping
3. Phenotype analysis
 1. NK cell subsets
 2. Scatter plotter

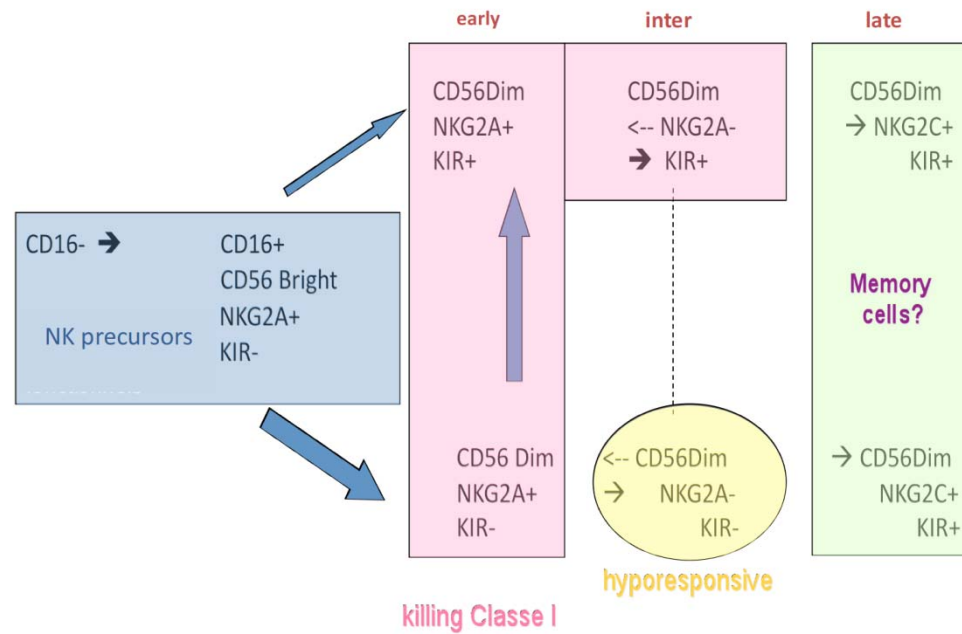
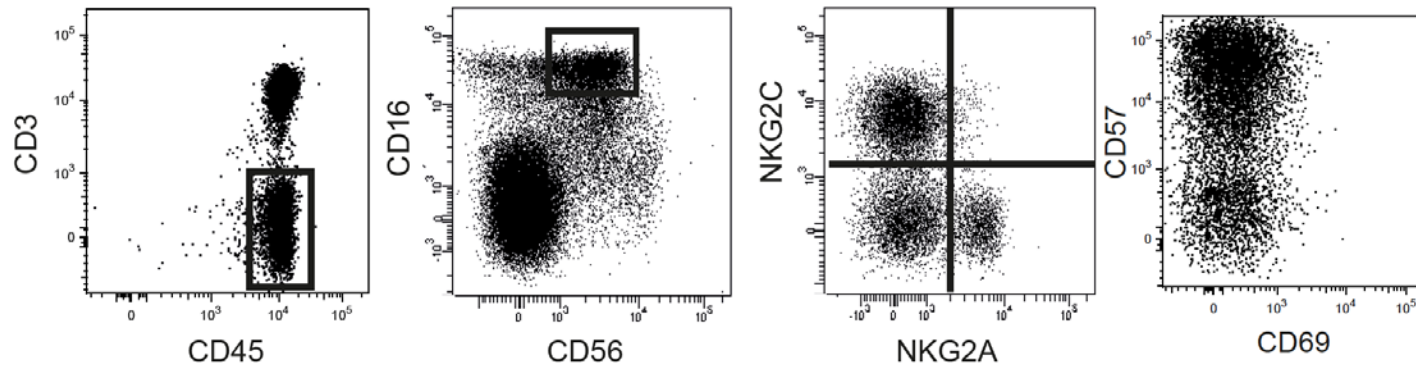
NK function



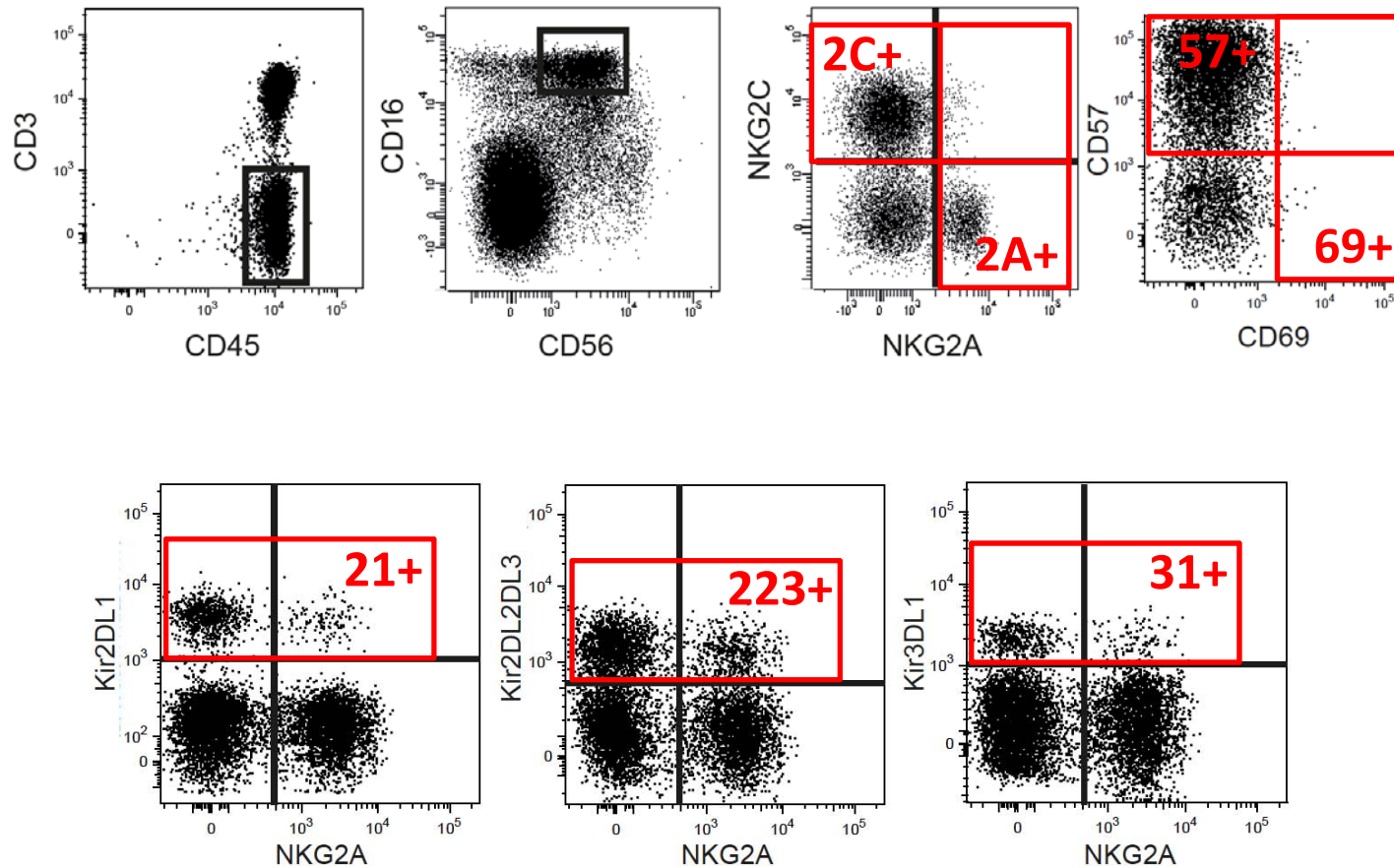
NK phenotype and KIR expression



NK phenotype and KIR expression



Boolean NK phenotype analysis



7 parameters -> $2^7 = 128$ boolean gates

Boolean NK phenotype analysis

Assignments:

1. Identify the frequency of 2A+2C-, 2A-2C- and 2A-2C+ NK cell subsets.
2. Identify the frequency of the three KIRs within the three subsets.
3. Exclude CD69 from the analysis.
4. Make comparative scatter plots of the data.