IEDB
Immune Epitope Database & Tools

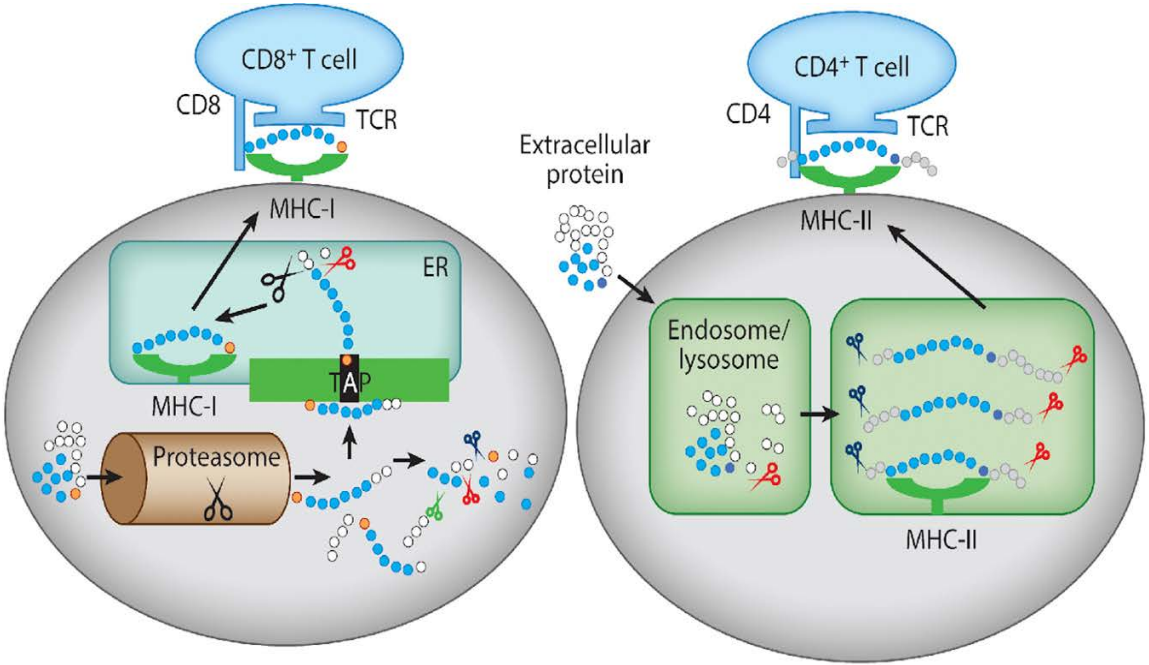
TCRMatch

Predicting T-cell receptor specificity based on sequence similarity to previously characterized receptors

Presented by
Raphael Trevizani

TCRMatch

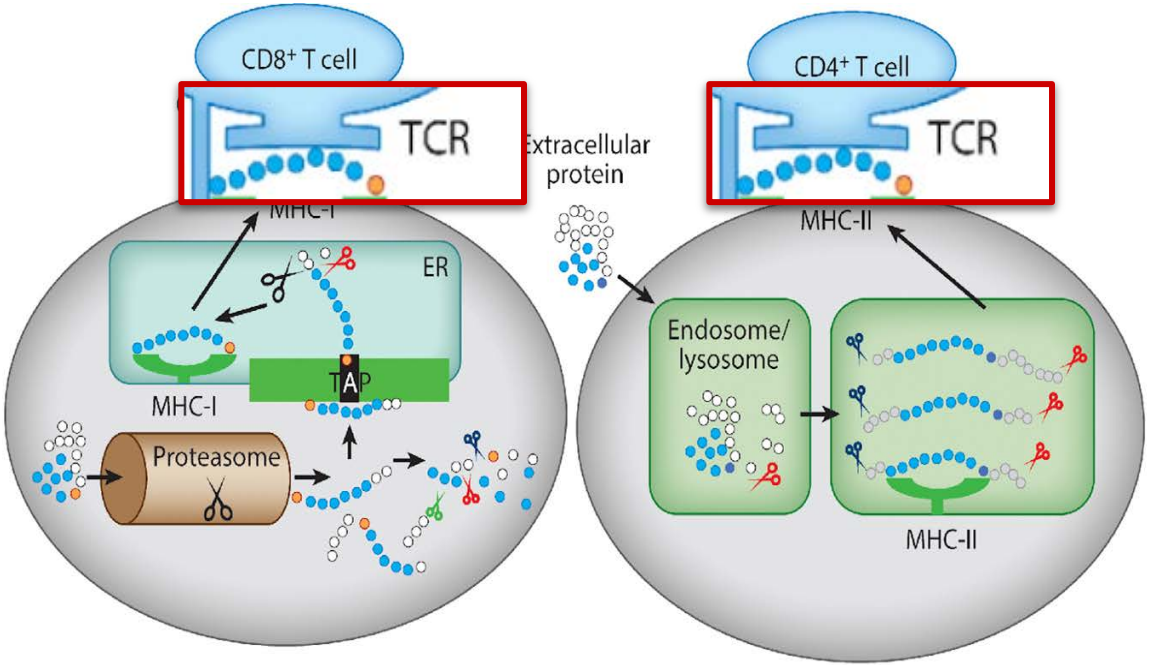
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



Peters et al, Annual Review of Immunology, 2020

TCRMatch

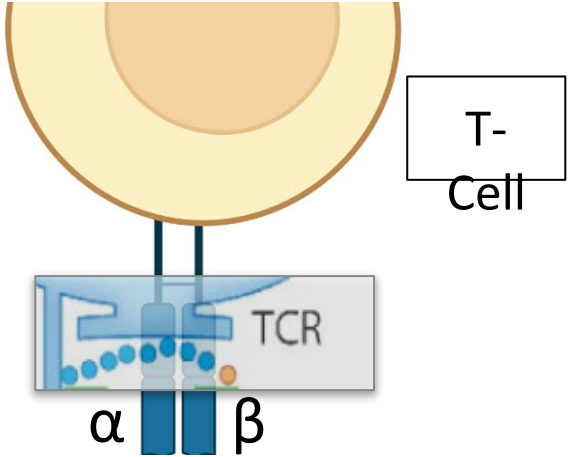
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



Peters et al, Annual Review of Immunology, 2020

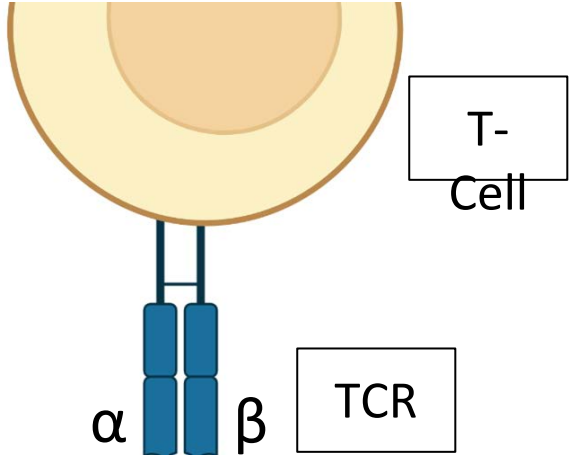
TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



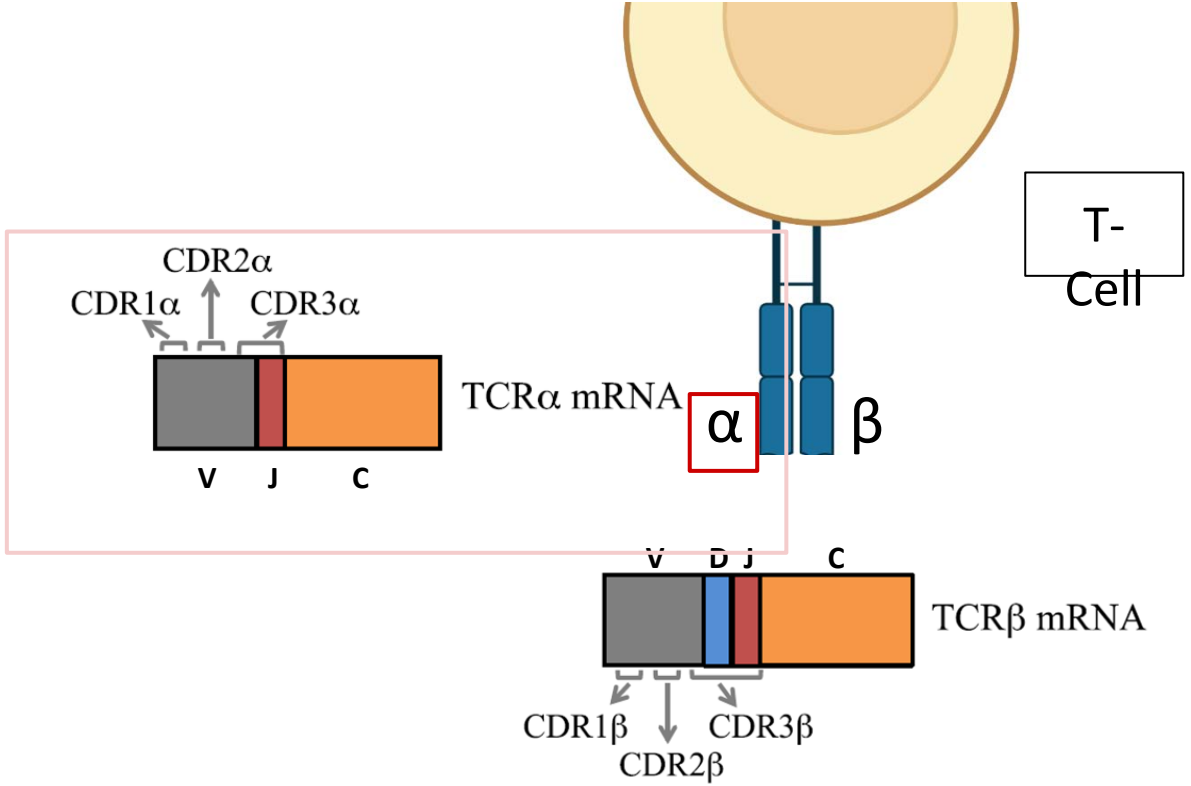
TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



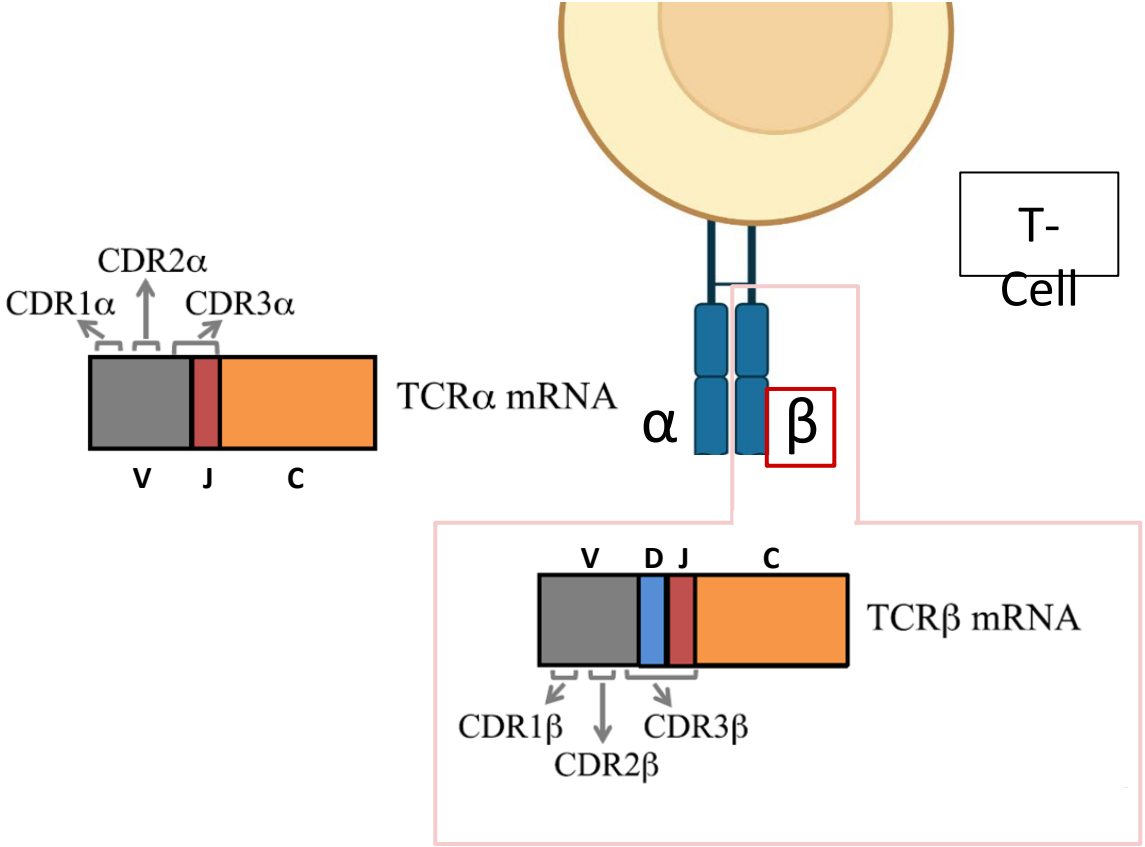
TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



TCRMatch

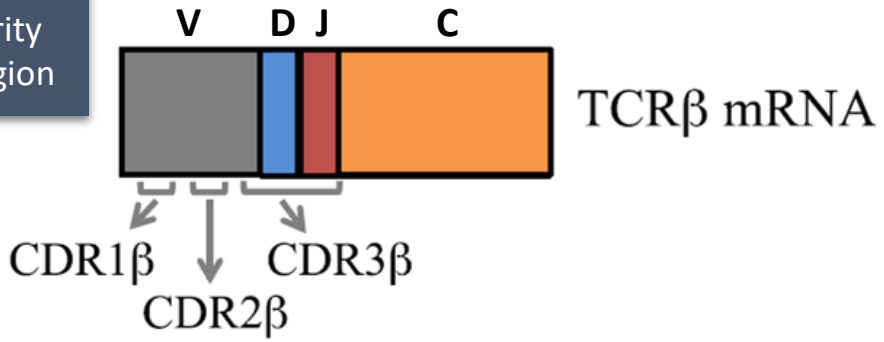
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

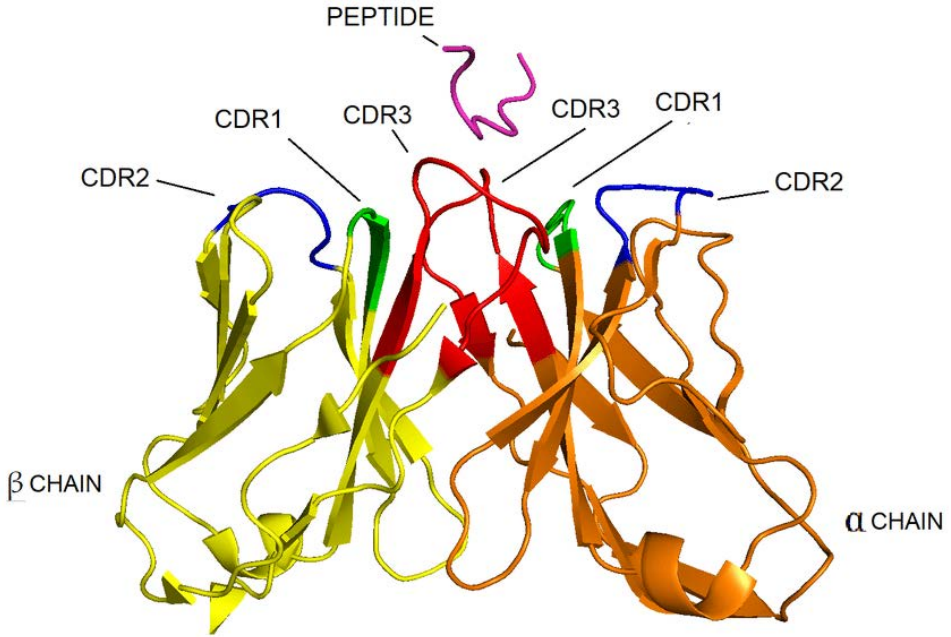


TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

Complementarity determining region

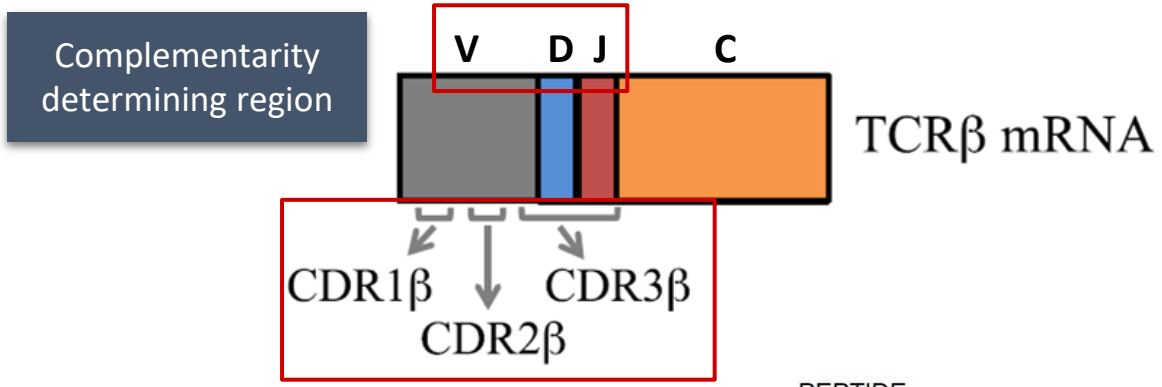


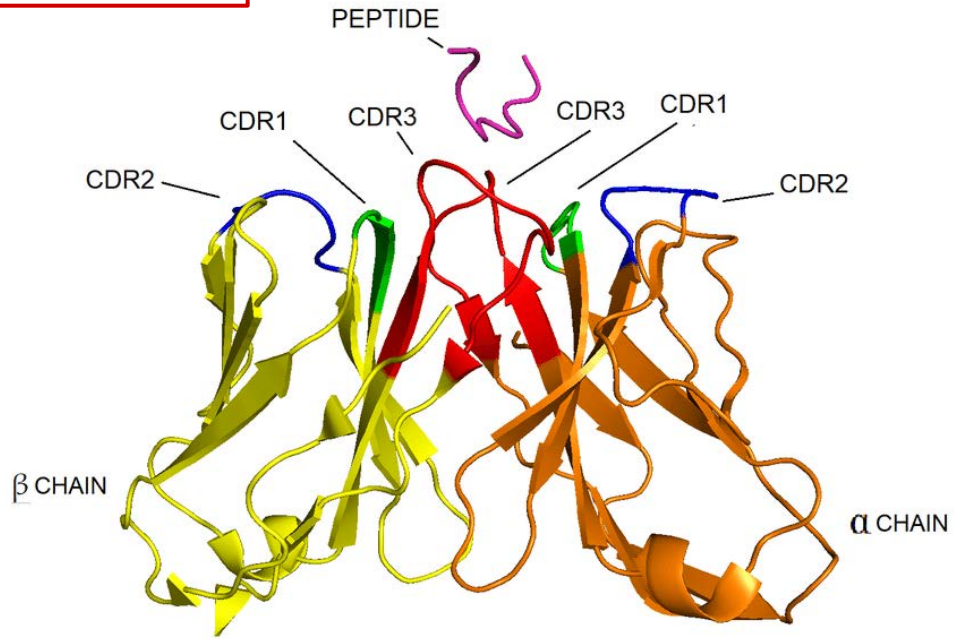


Cinelli, Mattia. (2018). Analysis of murine CDR3β repertoires using machine learning techniques.

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

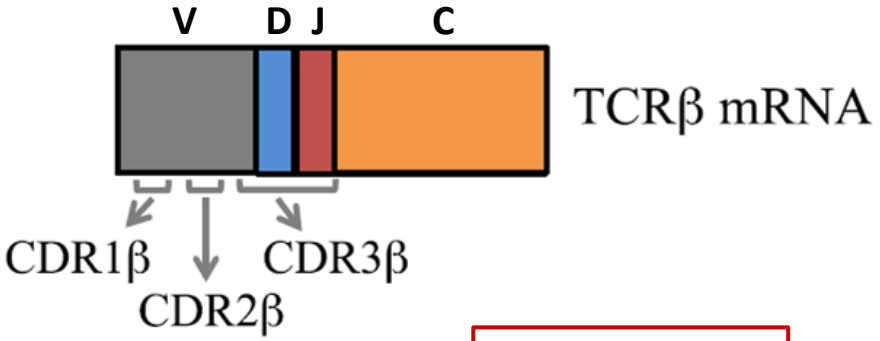


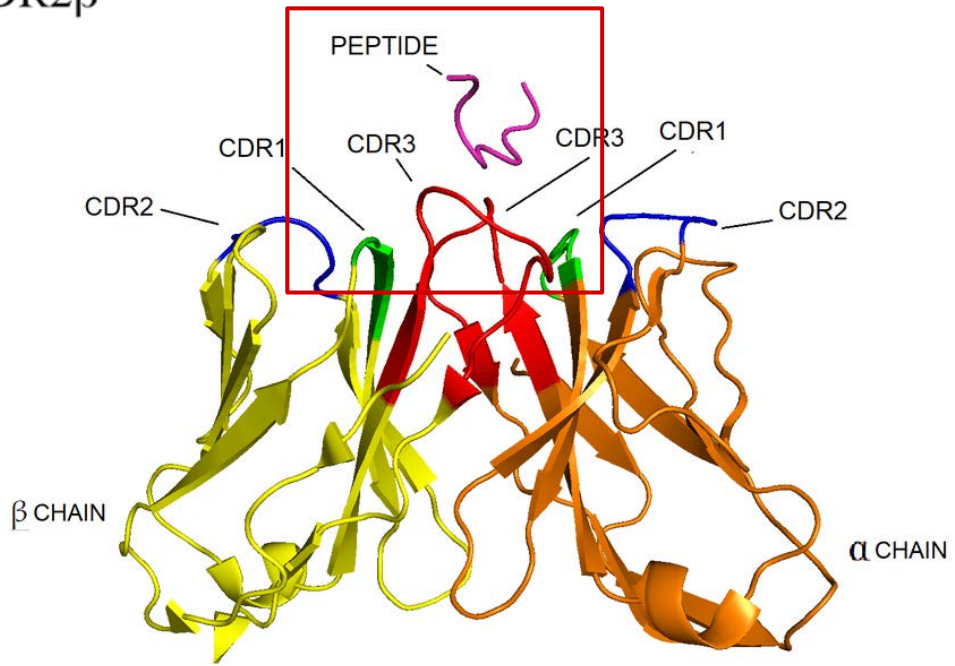


Cinelli, Mattia. (2018). Analysis of murine CDR3β repertoires using machine learning techniques.

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

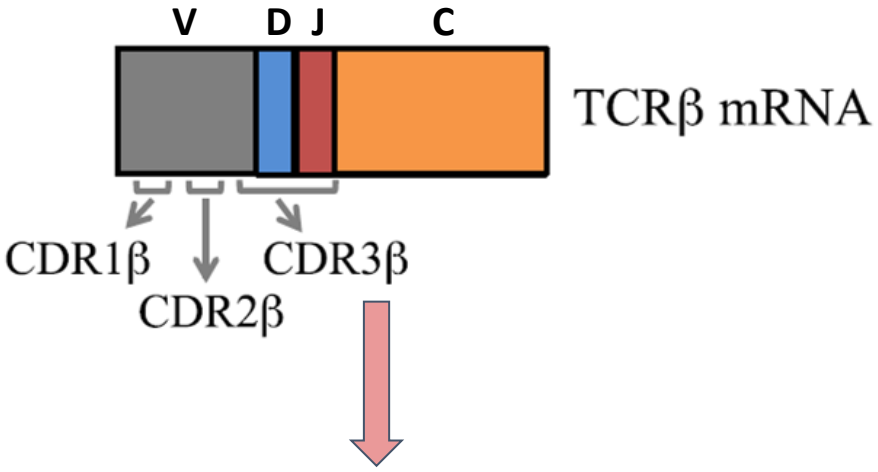




Cinelli, Mattia. (2018). Analysis of murine CDR3β repertoires using machine learning techniques.

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



Repertoire sequencing

...CASSIRSSYEQYF...
...CASSLEGYTEAFF...
...CASSSANYGYTF...
...CASSIRAAETQYF...
...CASSIRSSTEAFF...
...CASSSRSSYEQYF...

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

TCR sequencing does not directly reveal the epitope recognized by a given CDR3β

Repertoire sequencing

...CASSIRSSYEQYF...
...CASSLEGYTEAFF...
...CASSSANYGYTF...
...CASSIRAAETQYF...
...CASSIRSSTEAFF...
...CASSSRSSYEQYF...

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

TCR sequencing does not directly reveal the epitope recognized by a

Problem: how do we know the epitope?

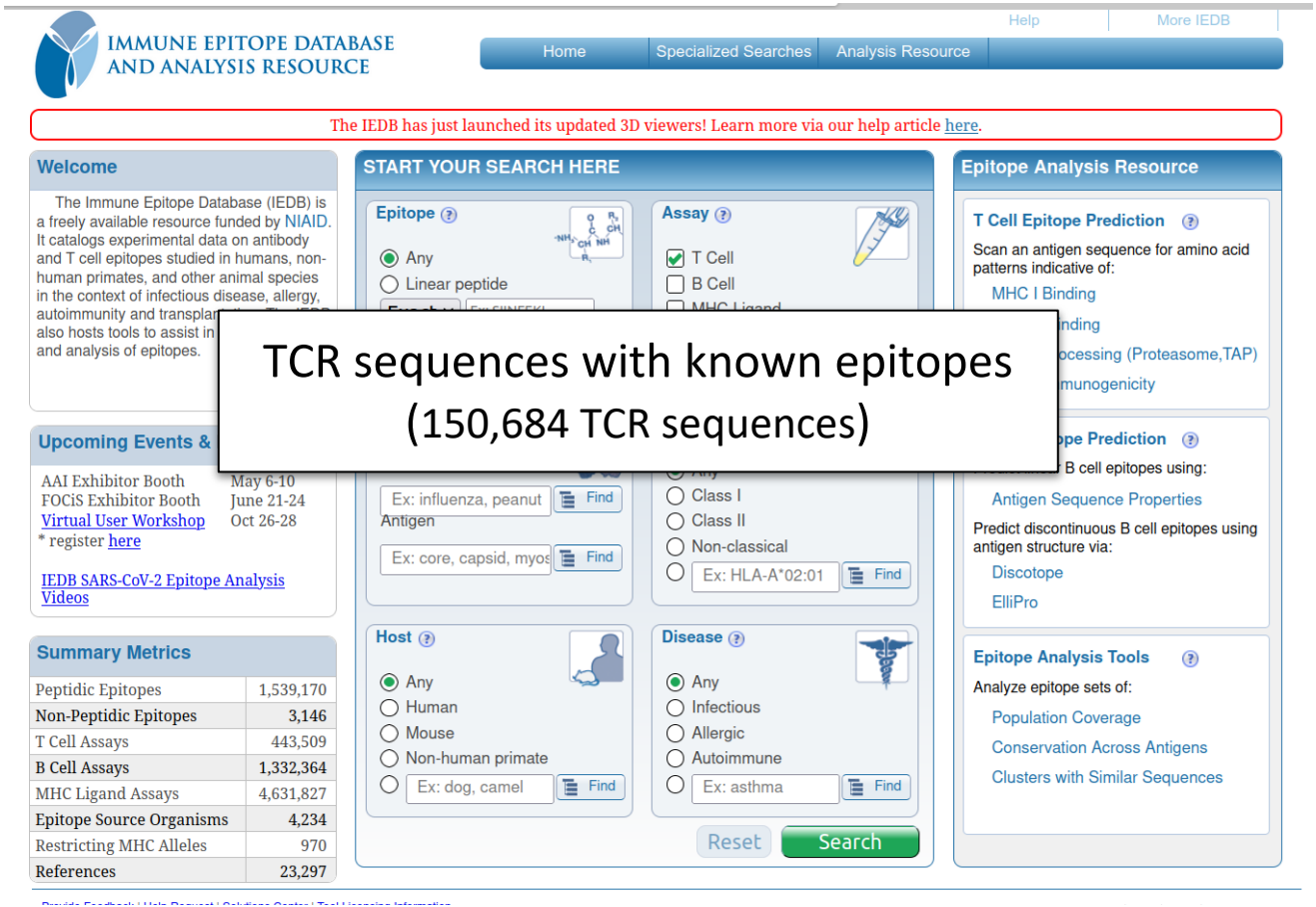
Repertoire sequencing

...CASSIRSSYEQYF...
...CASSLEGYTEAFF...
...CASSSANYGYTF...
...CASSIRAAETQYF...
...CASSIRSSTEAFF...
...CASSSRSSYEQYF...

TCRMatch

http://www.iedb.org/

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

TCR sequencing does not directly reveal the epitope recognized by a

Problem: how do we know the epitope?

Repertoire sequencing

...CASSIRSSYEQYF...
...CASSLEGYTEAFF...
...CASSSANYGYTF...
...CASSIRAAETQYF...
...CASSIRSSTEAFF...
...CASSSRSSYEQYF...

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

TCR sequencing does not directly reveal the epitope recognized by a

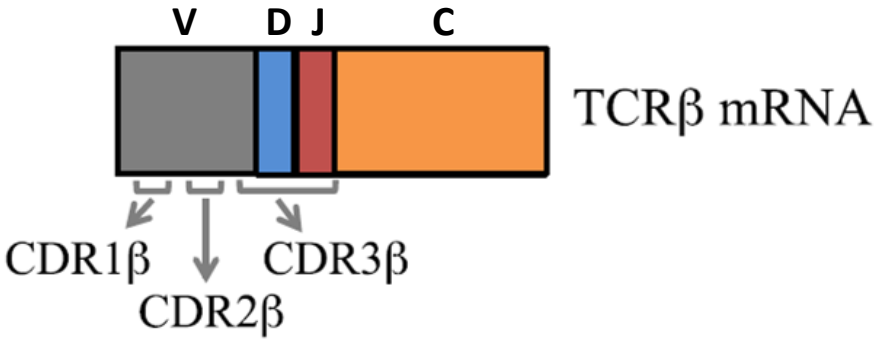
Solution: use the sequence of known TCRs to match the query,

Repertoire sequencing

...CASSIRSSYEQYF...
...CASSLEGYTEAFF...
...CASSSANYGYTF...
...CASSIRAAETQYF...
...CASSIRSSTEAFF...
...CASSSRSSYEQYF...

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

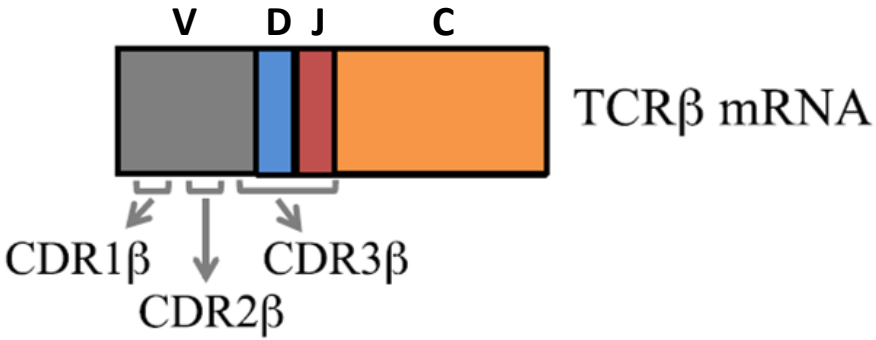


Repertoire sequencing

…CASSIRSSYEQYF…
…CASSLEGYTEAFF…
…CASSSANYGYTF…
…CASSIRAAETQYF…
…CASSIRSSTEAFF…
…CASSSRSSYEQYF…

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

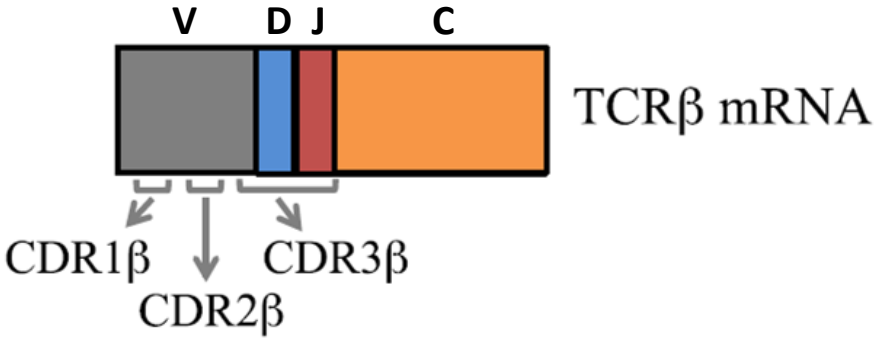


Repertoire sequencing

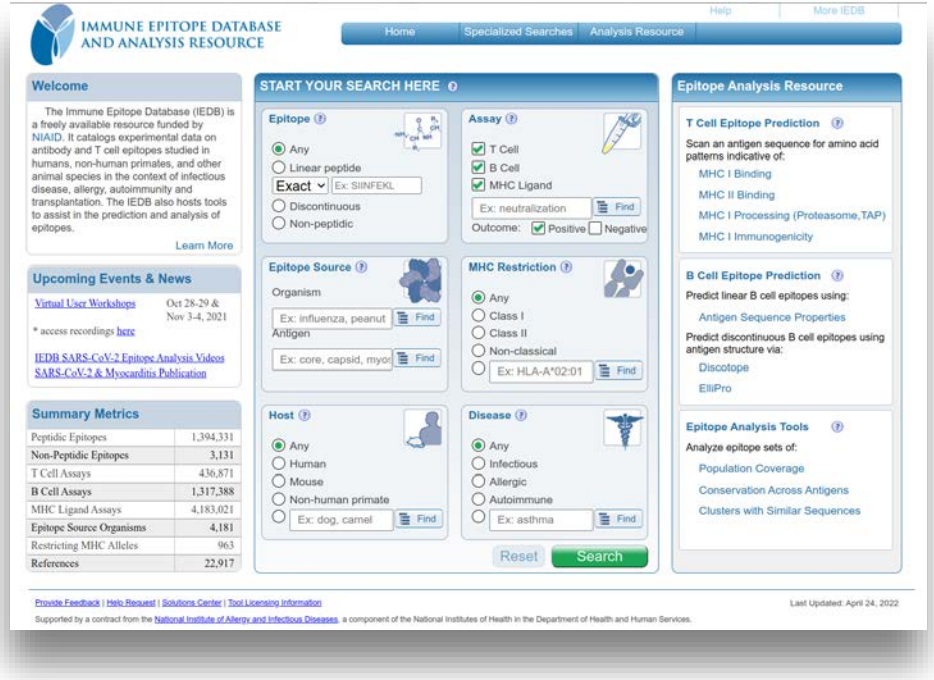
...CASSIRSSYEQYF...
...CASSLEGYTEAFF...
...CASSSANYGYTF...
...CASSIRAAETQYF...
...CASSIRSSTEAFF...
...CASSSRSSYEQYF...

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

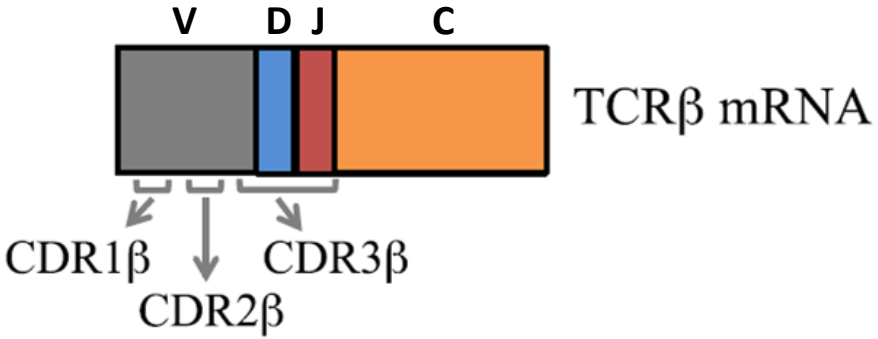


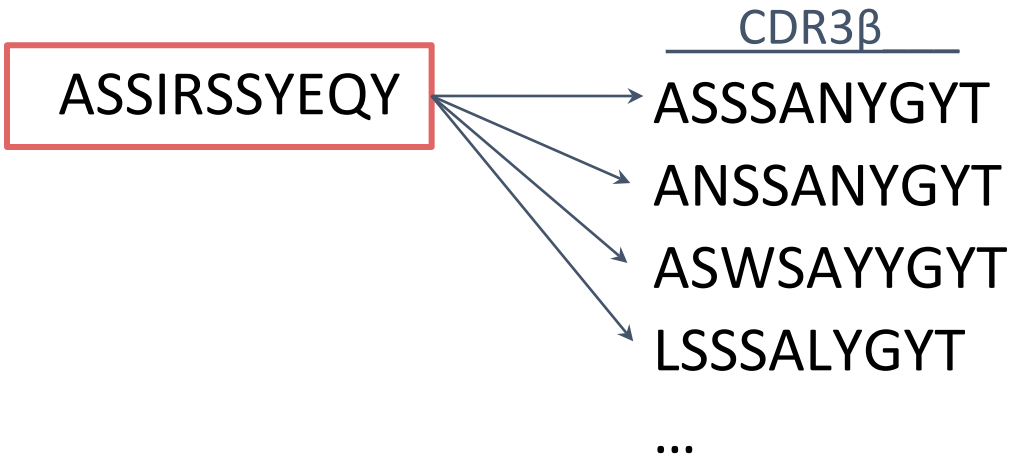
ASSIRSSYEQY



TCRMatch

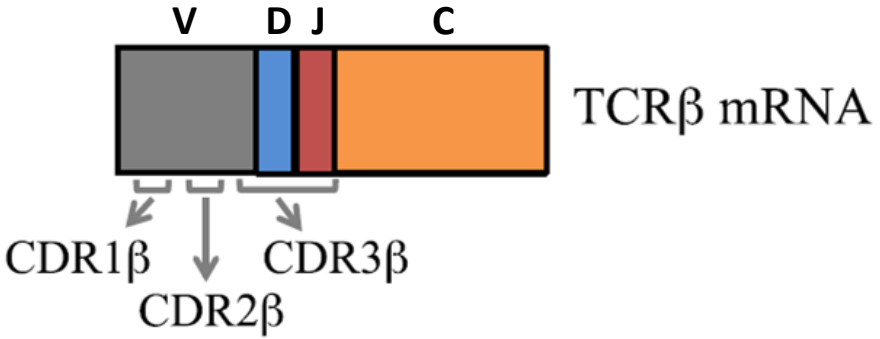
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

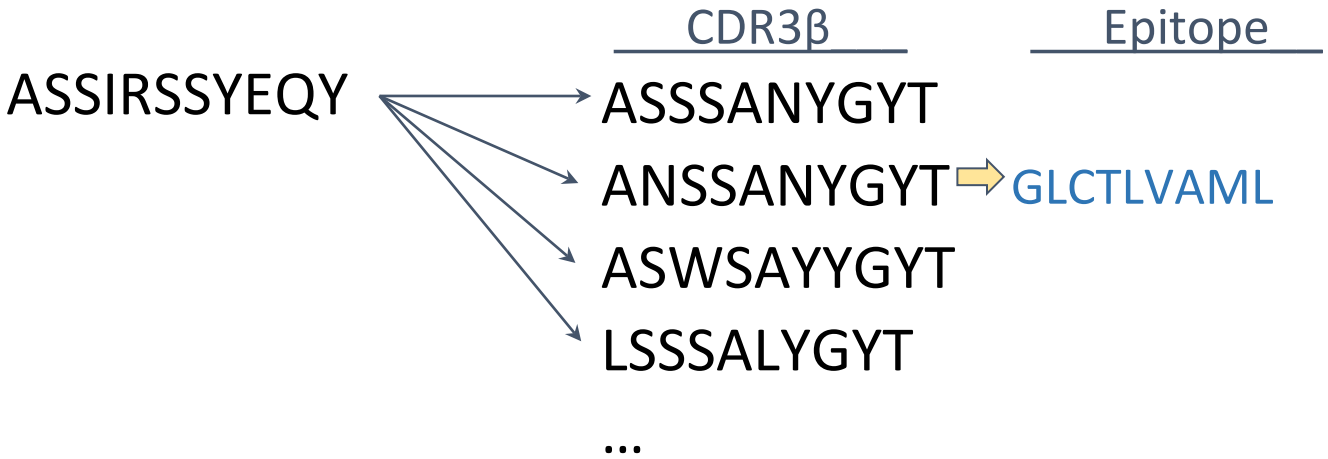




TCRMatch

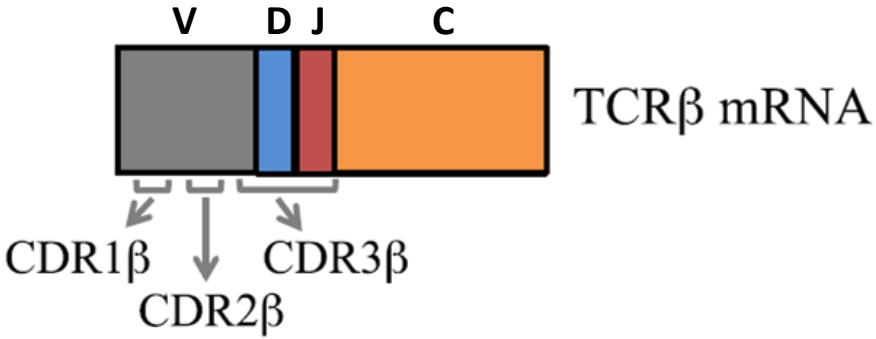
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

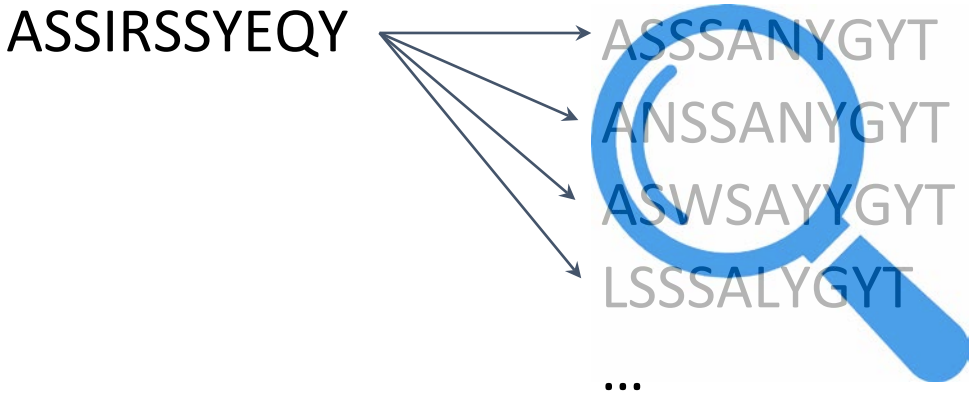




TCRMatch

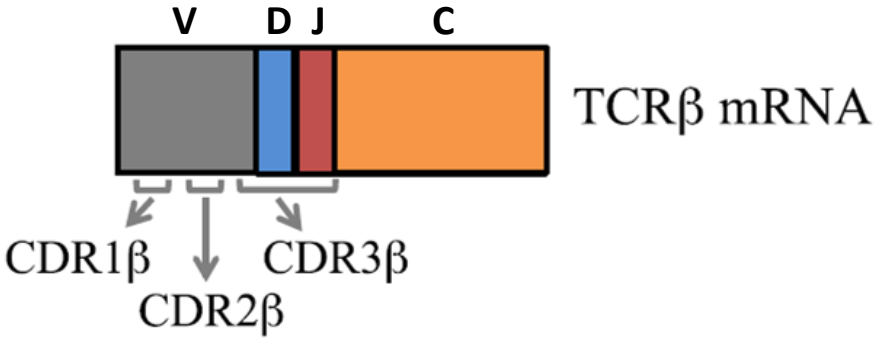
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

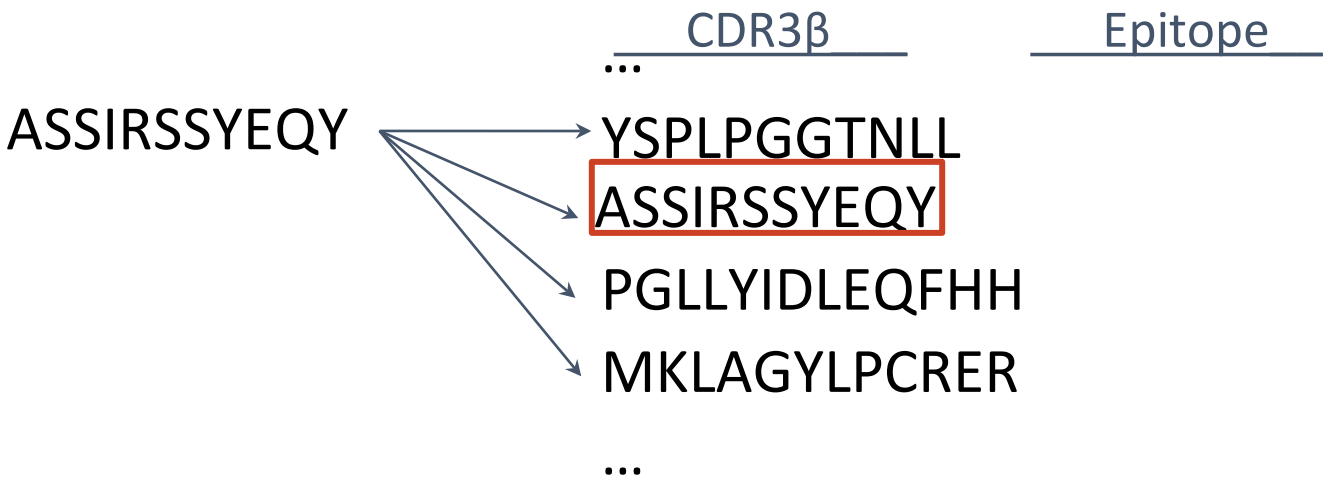




TCRMatch

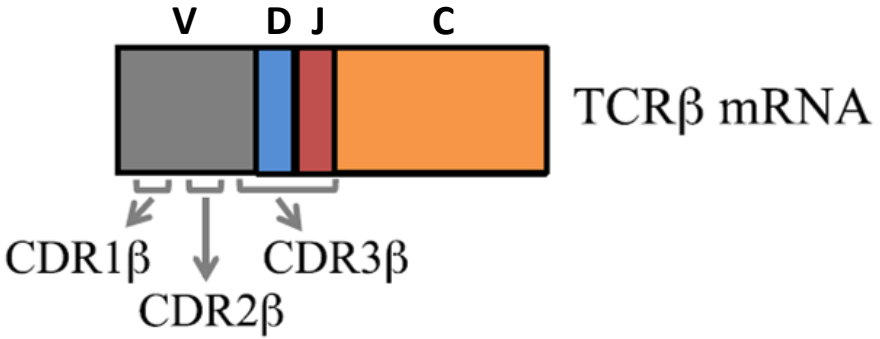
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

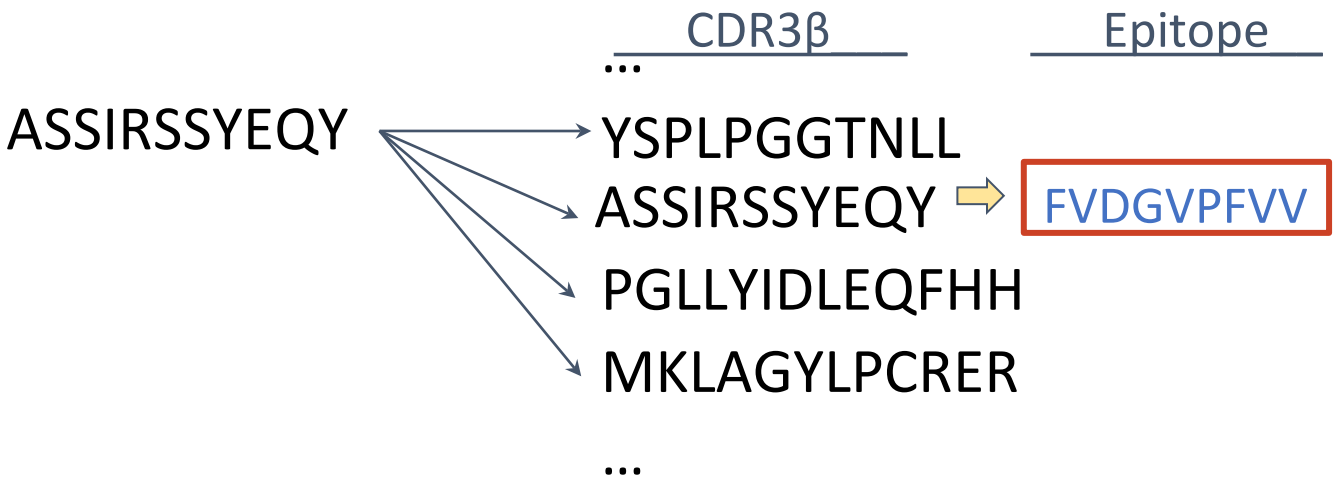




TCRMatch

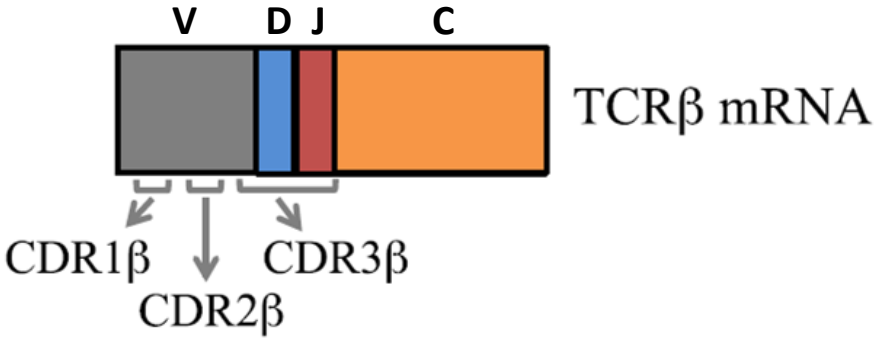
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?





TCRMatch

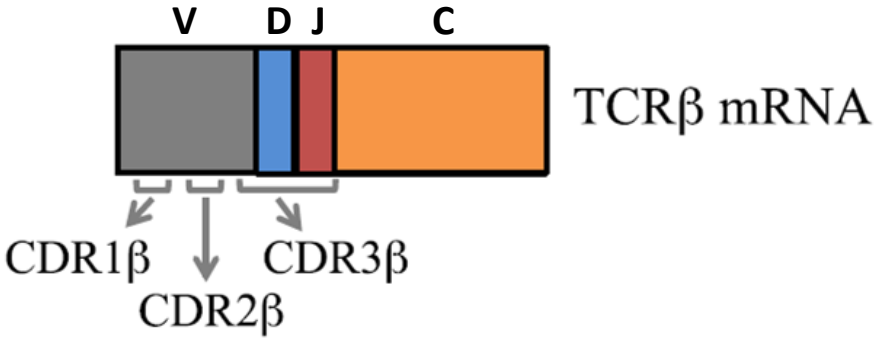
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



CDR3β	Epitope
ASSIRSSYEQY	

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



CDR3β → Epitope

CSSIRSSYEQY → not found → ?

TCRMatch

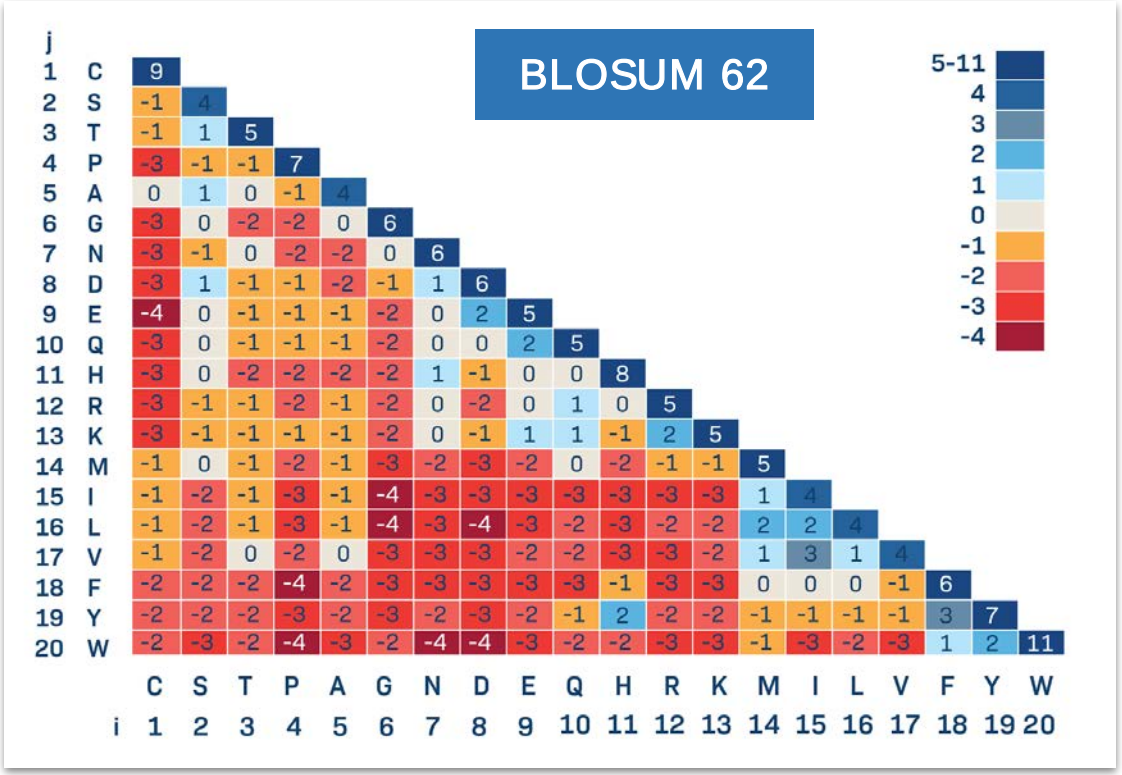
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

TCRMatch

CDR3β Epitope

CSSIRSSYEQY → not found → ?

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

CDR3β → Epitope

CSSIRSSYEQY → not found → ?

Metric	Description
Alignment Score	Alignment score divided by length of alignment
Identity Alignment	Percent identity within length of alignment
Identity Long	Percent identity within length of longer sequence
Identity Short	Percent identity within length of shorter sequence
Levenshtein distance	Minimum number of edits (substitutions, insertions, and deletions) necessary to transform one sequence into another
TCRdist	Similarity-weighted mismatch distance between two sequences
TCRMatch (MAIT Match)	Comprehensive comparison of all possible k-mers using BLOSUM62 observed frequency matrix

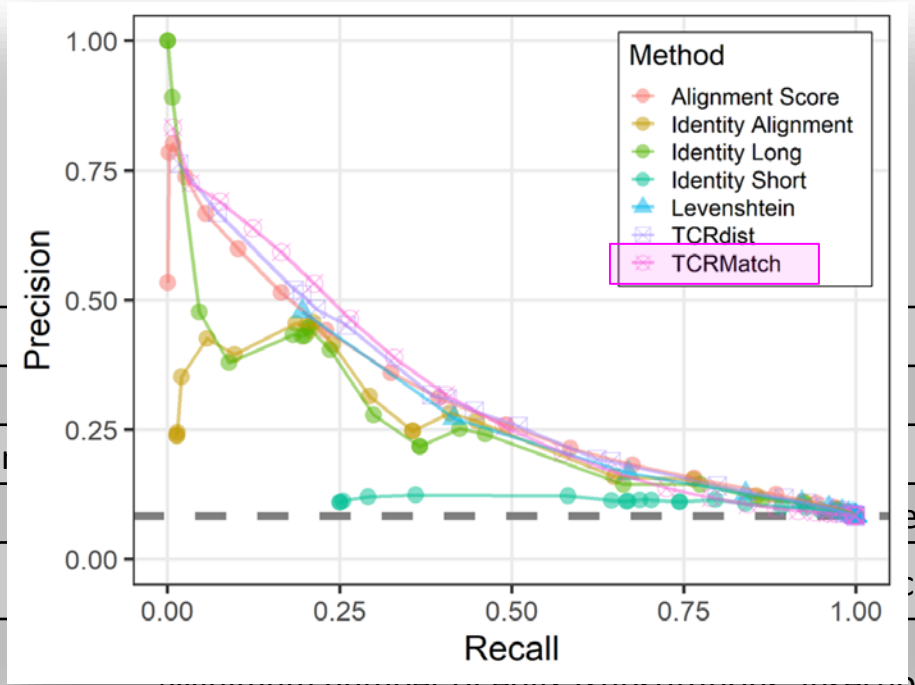
TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?

Metric	Description
Alignment Score	Alignment score divided by length of alignment
Identity Alignment	Percent identity within length of alignment
Identity Long	Percent identity within length of longer sequence
Identity Short	Percent identity within length of shorter sequence
Levenshtein distance	Minimum number of edits (substitutions, insertions, and deletions) necessary to transform one sequence into another
TCRdist	Similarity-weighted mismatch distance between two sequences
TCRMatch (MAIT Match)	Comprehensive comparison of all possible k-mers using BLOSUM62 observed frequency matrix

TCRMatch

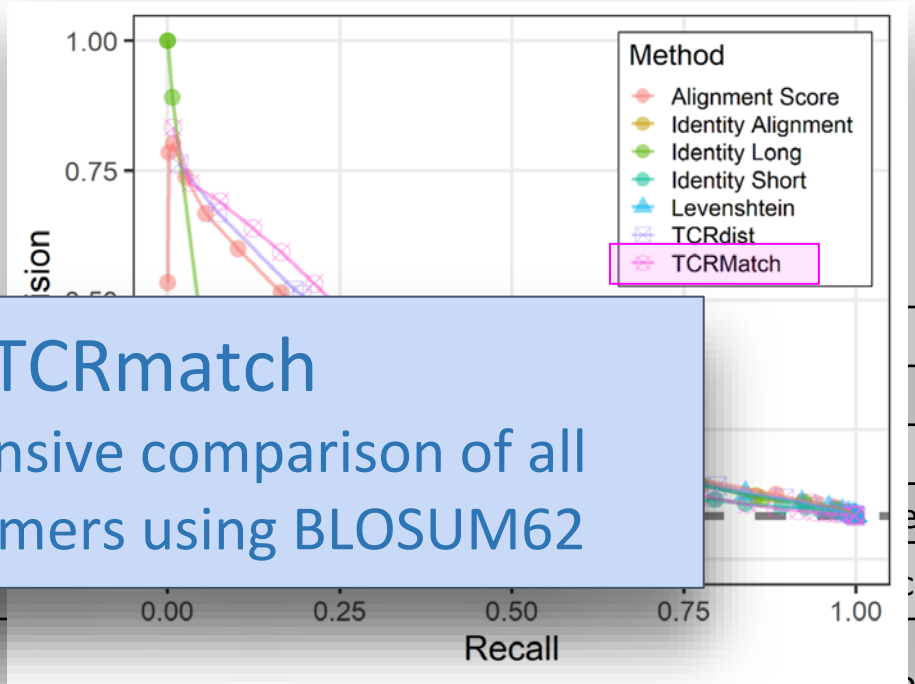
- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombination
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interacts with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



Metric	
Alignment Score	
Identity Alignmer	
Identity Long	
Identity Short	
Levenshtein distance	Minimum number of edits (substitutions, insertions, and deletions) necessary to transform one sequence into another
TCRdist	Similarity-weighted mismatch distance between two sequences
TCRMatch (MAIT Match)	Comprehensive comparison of all possible k-mers using BLOSUM62 observed frequency matrix

TCRMatch

- MHC presenting an antigen to a T-Cell
- T-cell receptor: α/β units
- Somatic recombinat
- Each chain: 3 CDRs
- CDR3β
  - Most variable
  - Directly interact with epitope
- Repertoire sequencing
- Which epitope interacts with *this specific* TCR?
- TCRmatch: searches IEDB for a similar CDR3β
- How to find the most similar sequences?



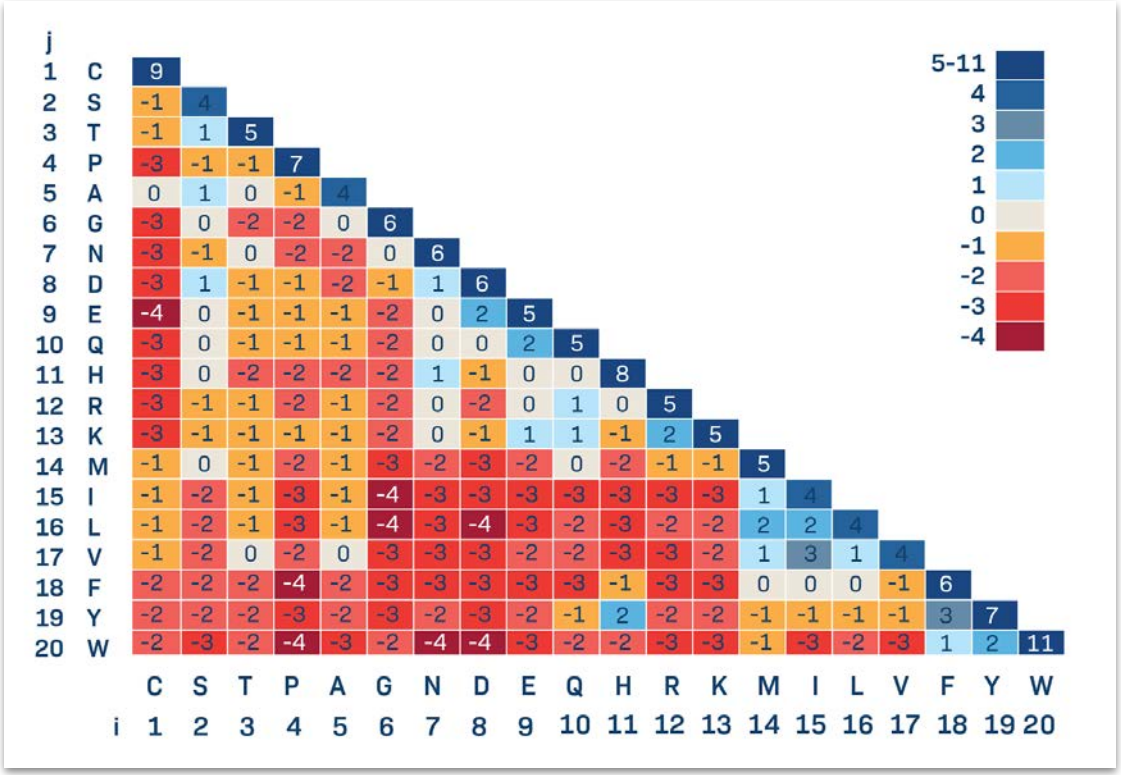
TCRmatch

Comprehensive comparison of all possible k-mers using BLOSUM62

Levenshtein distance	Minimum number of edits (substitutions, insertions, and deletions) necessary to transform one sequence into another
TCRdist	Similarity-weighted mismatch distance between two sequences
TCRMatch (MAIT Match)	Comprehensive comparison of all possible k-mers using BLOSUM62 observed frequency matrix

TCRMatch

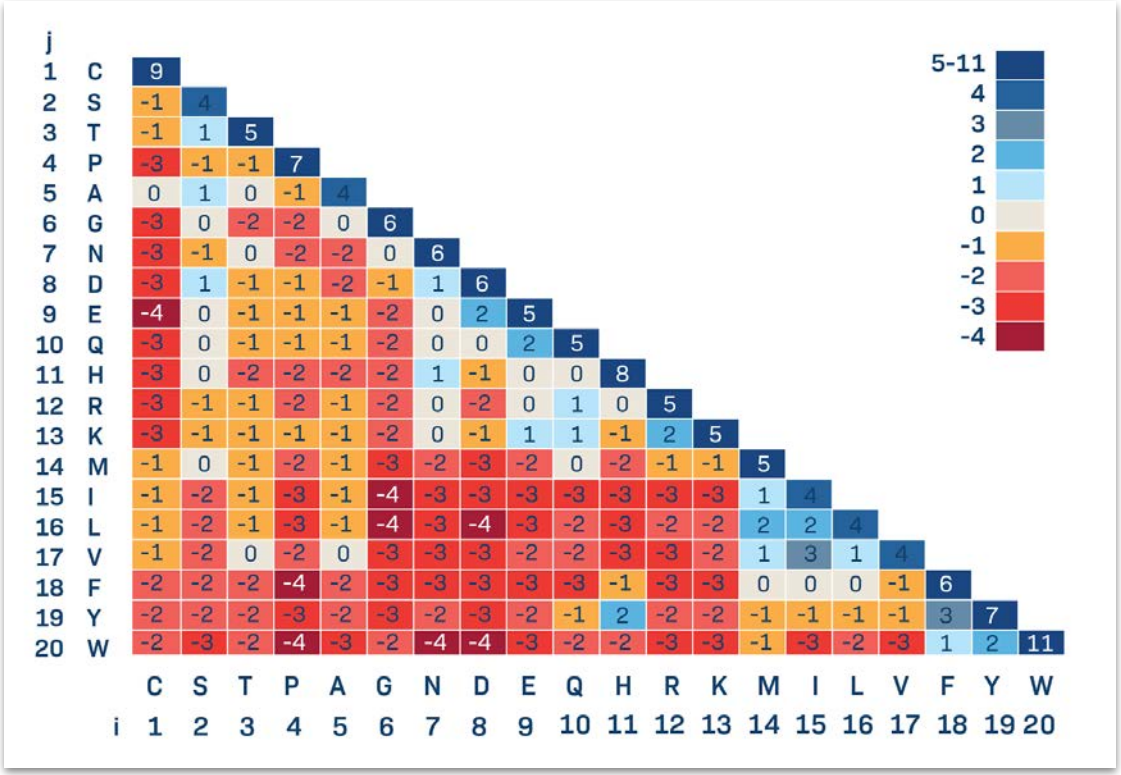
ASSSANYGYT
ASSIRAAETQY



TCRMatch

ASSSANYGYT

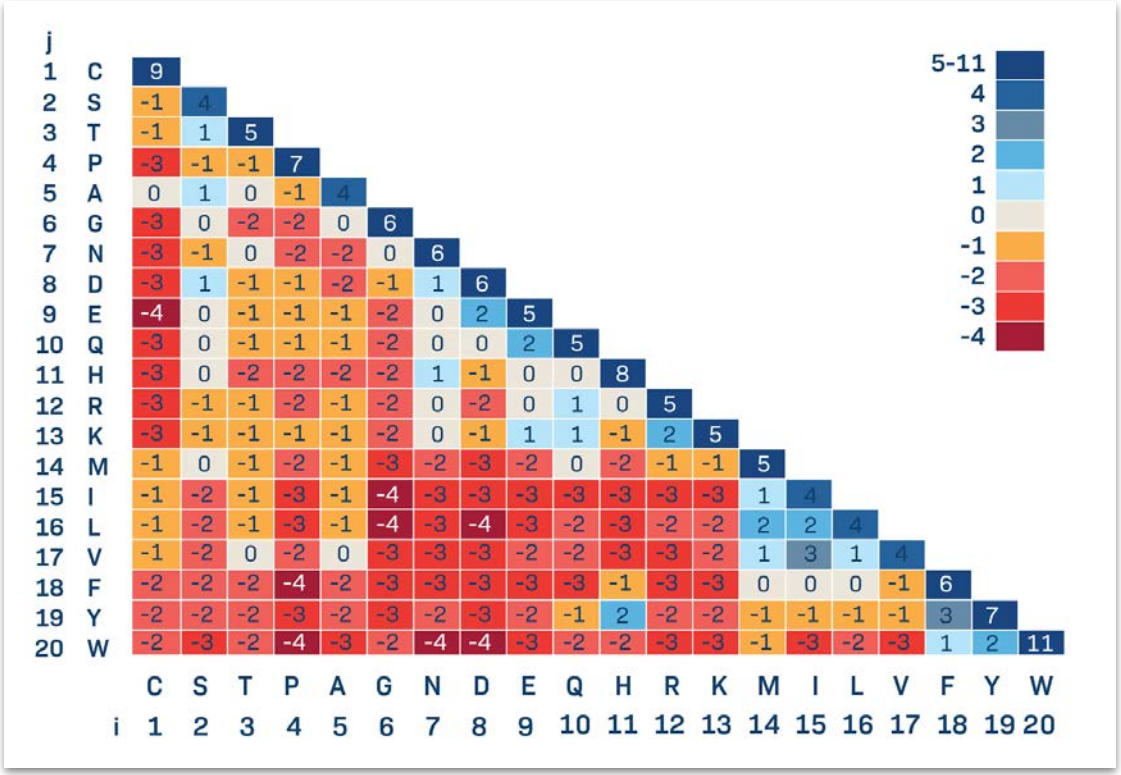
ASSIRAAETQY



TCRMatch

ASSSANYGYT

ASSIRAAETQY

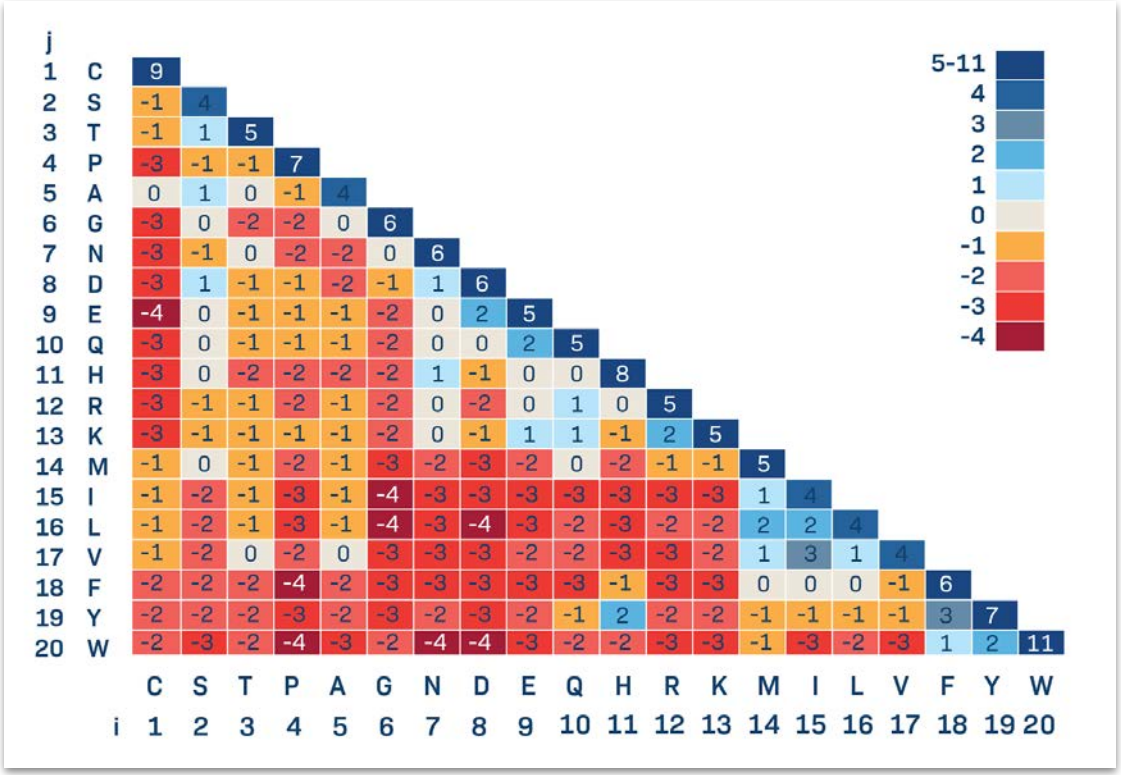


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

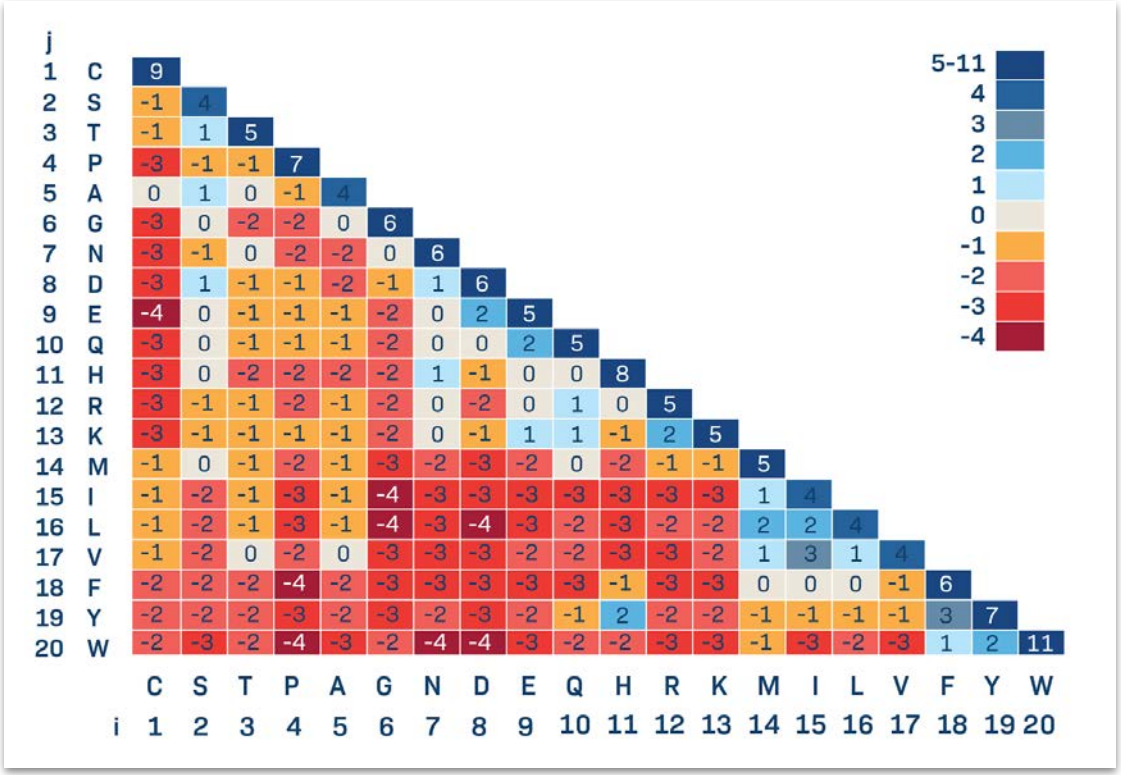


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

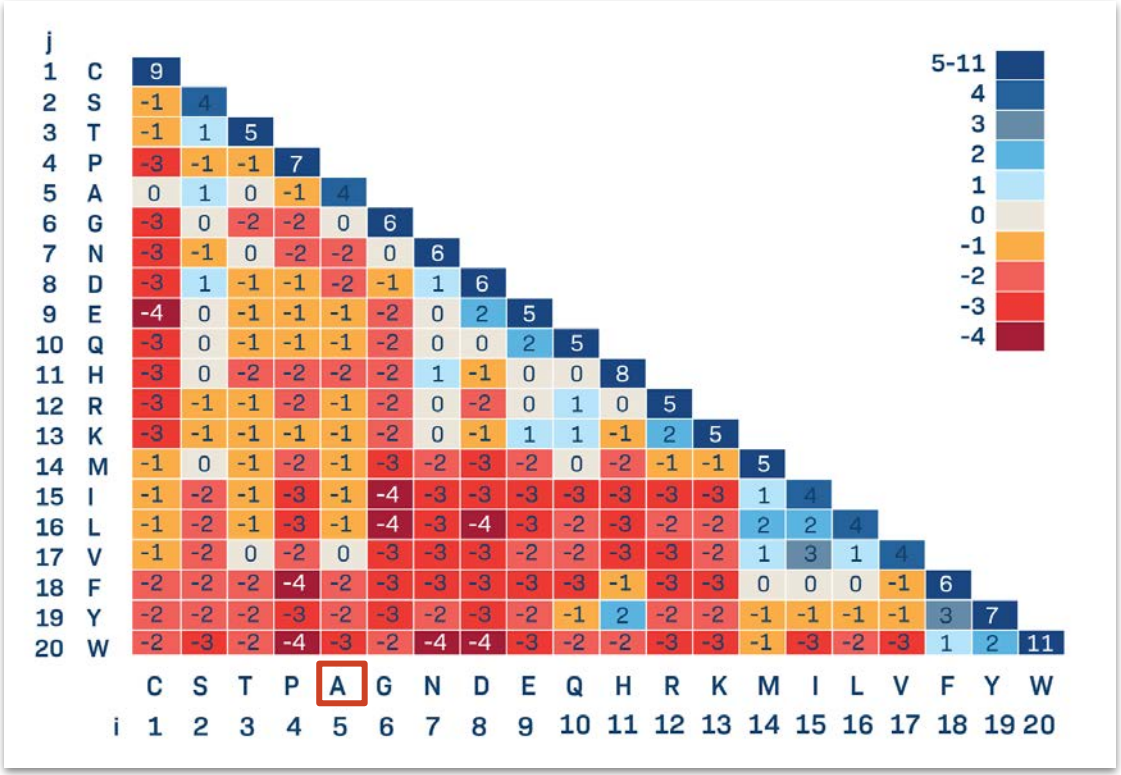


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

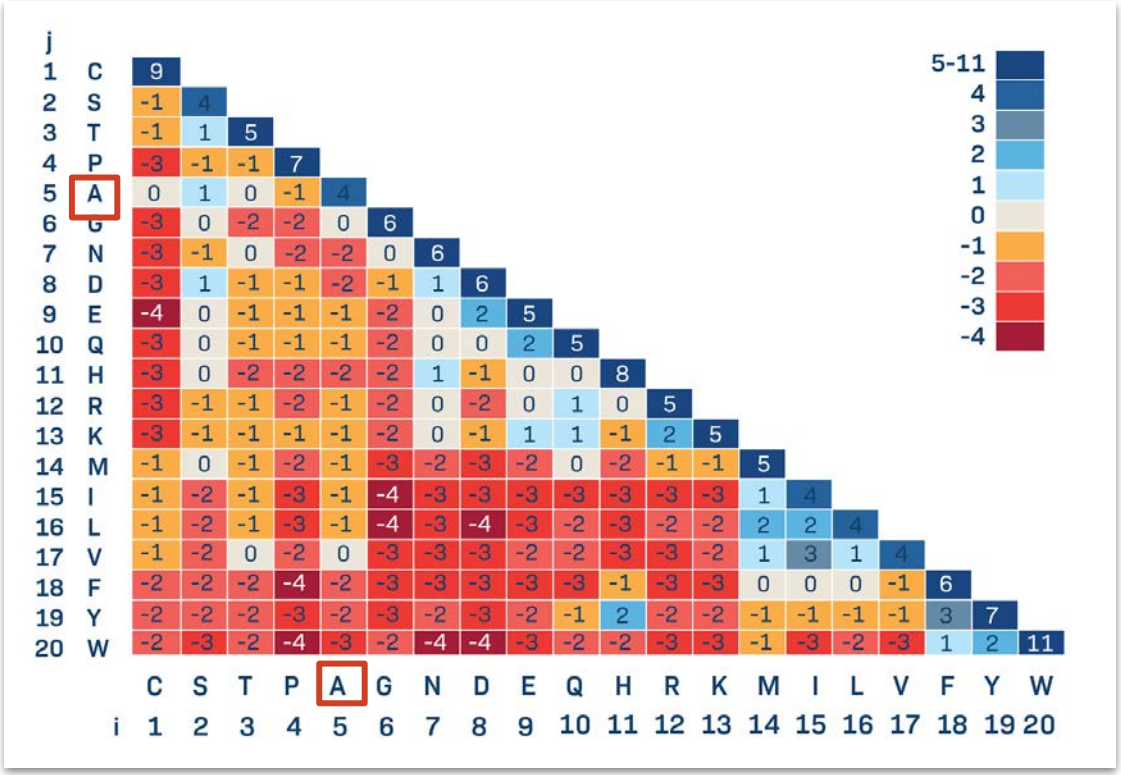


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

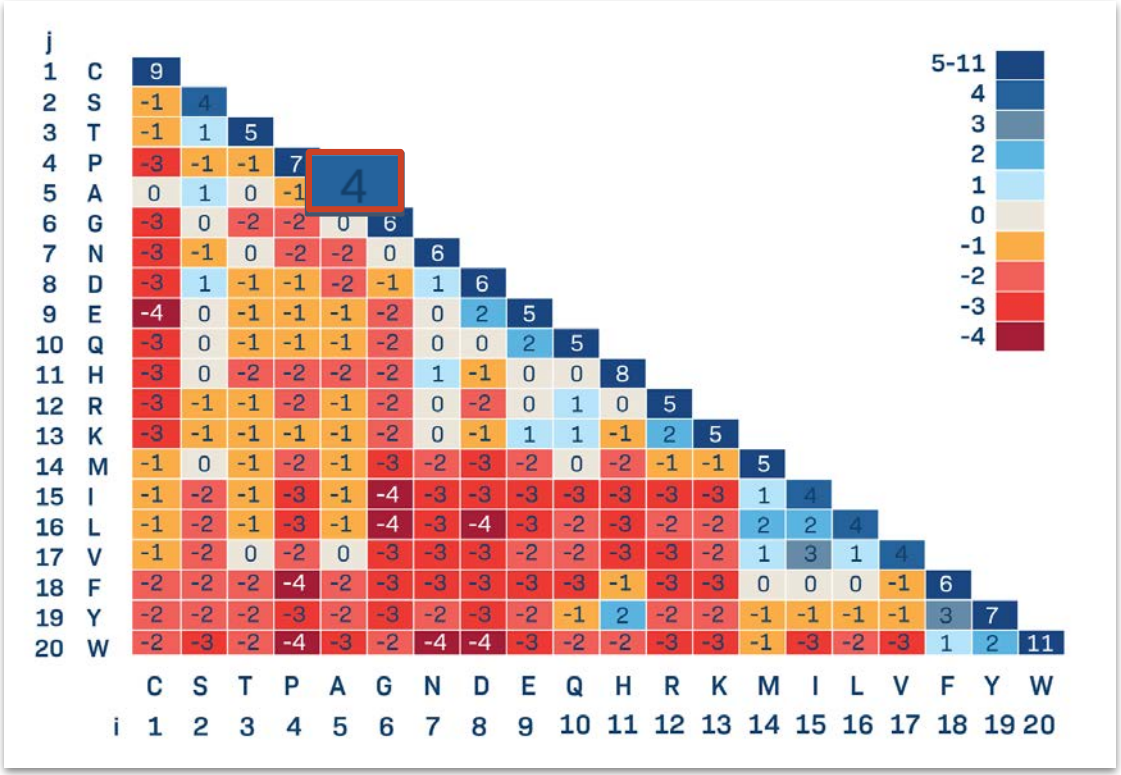


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

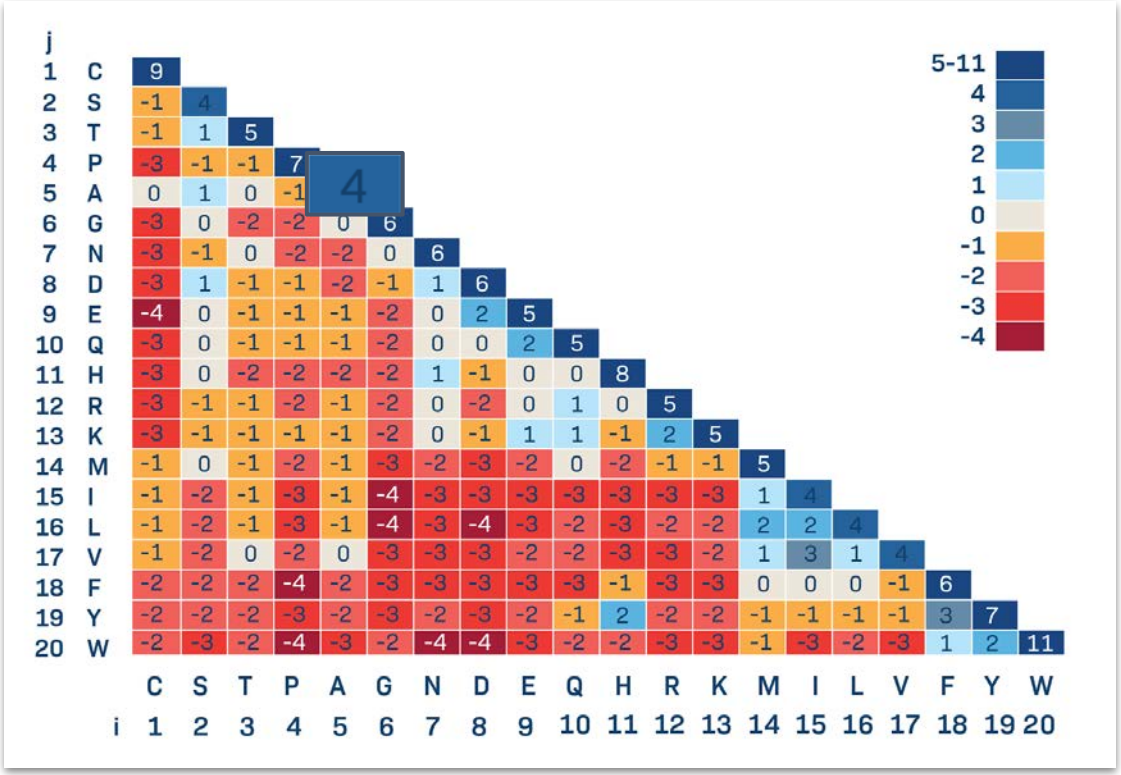


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

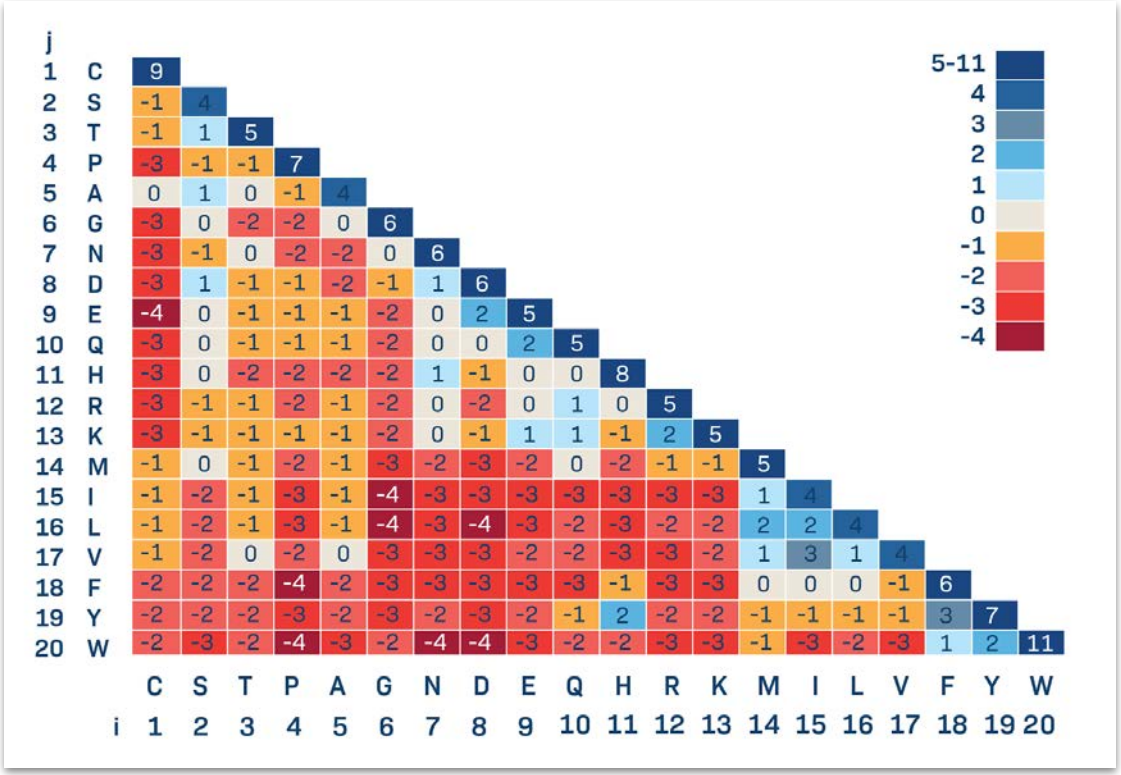


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

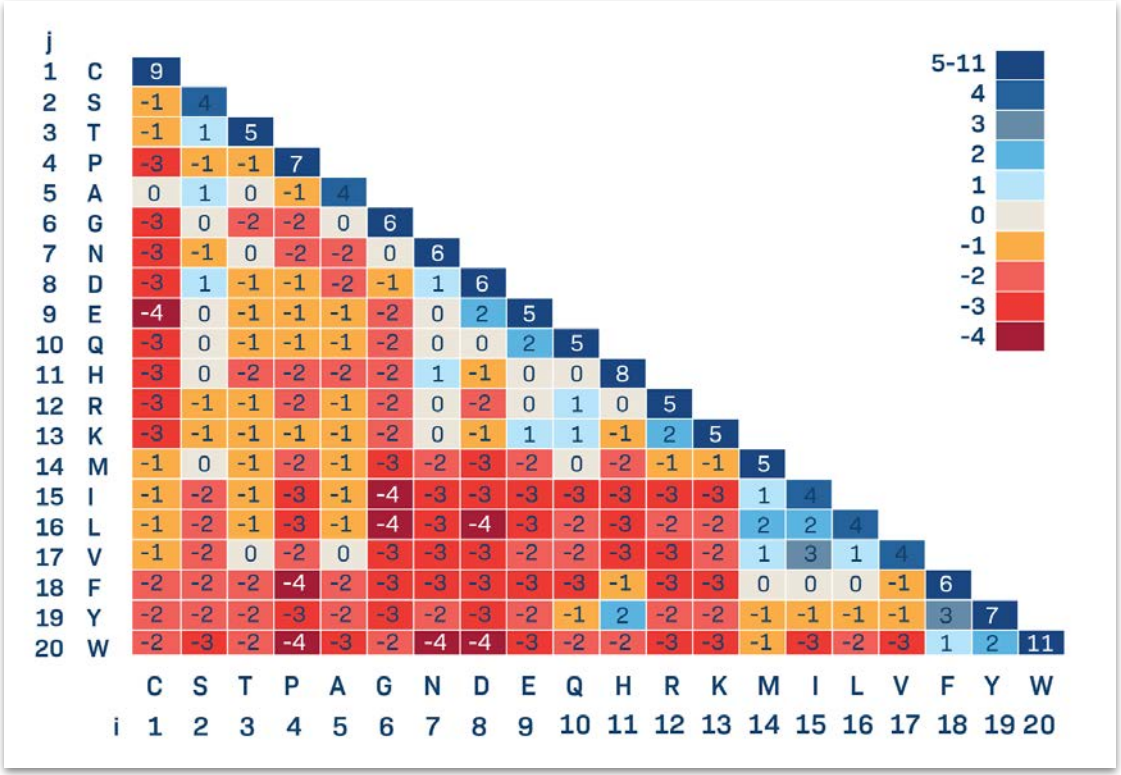


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

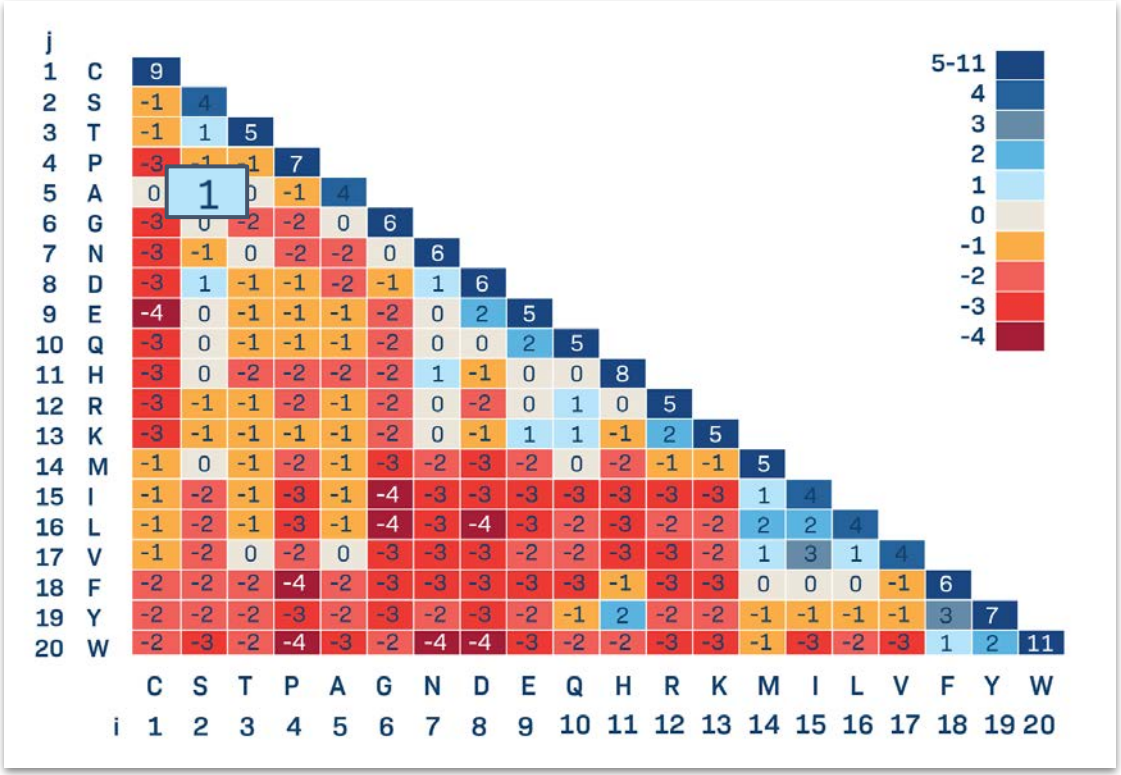


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

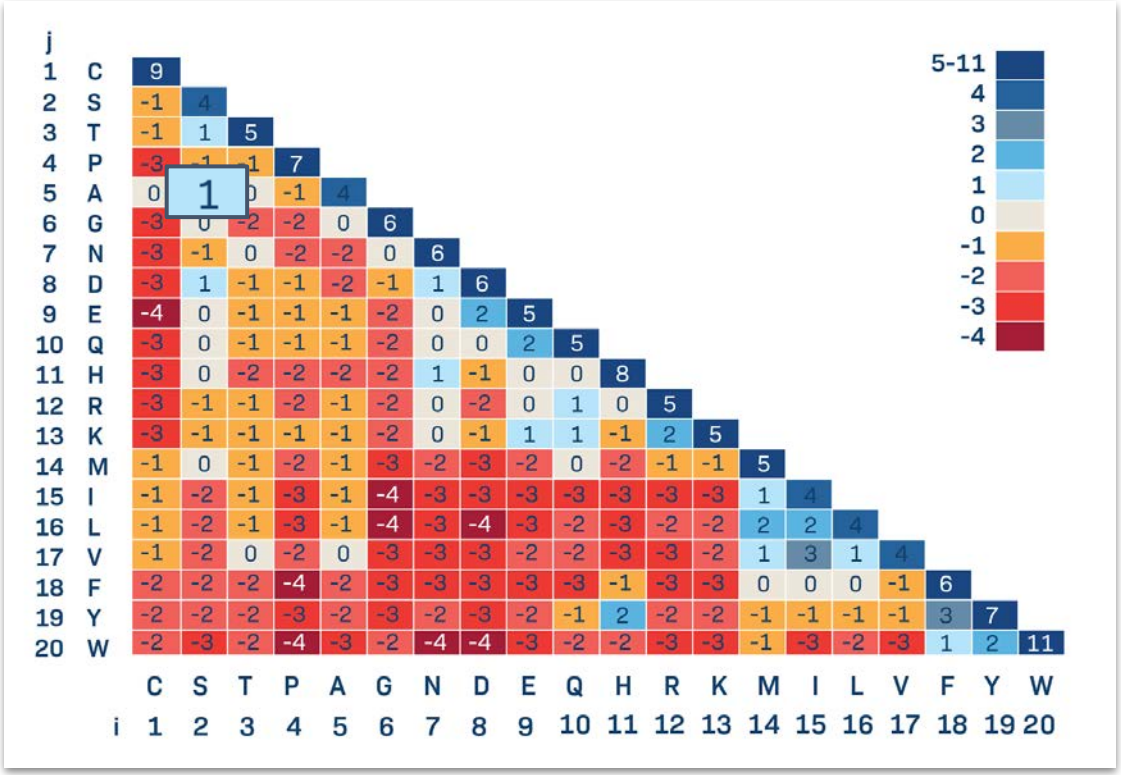


TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

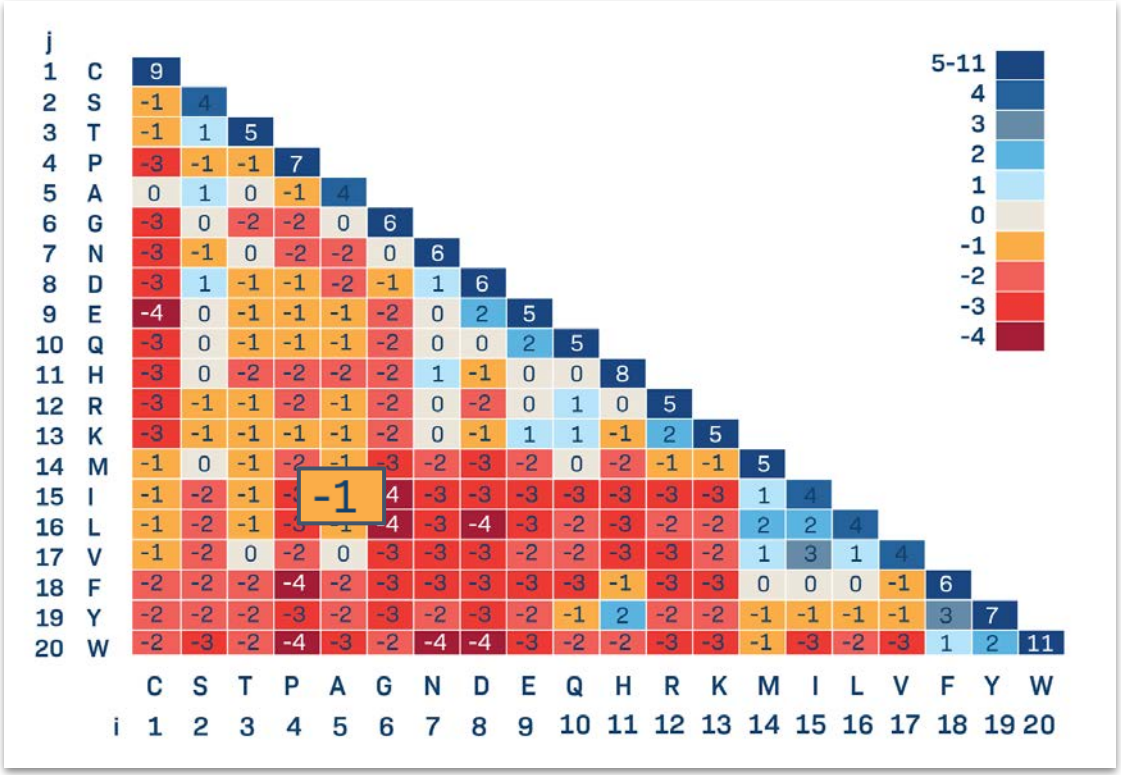


TCRMatch

k = 1

ASSSANYGYT

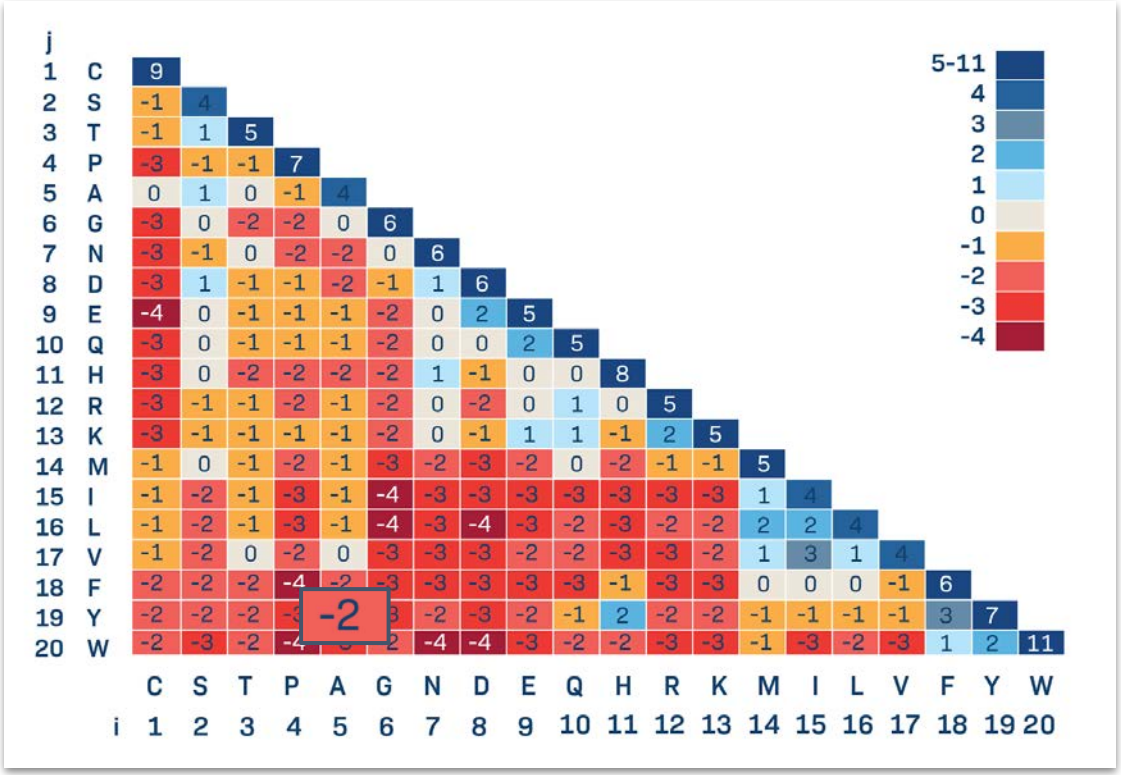
ASSIRAAETQY



TCRMatch

k = 1

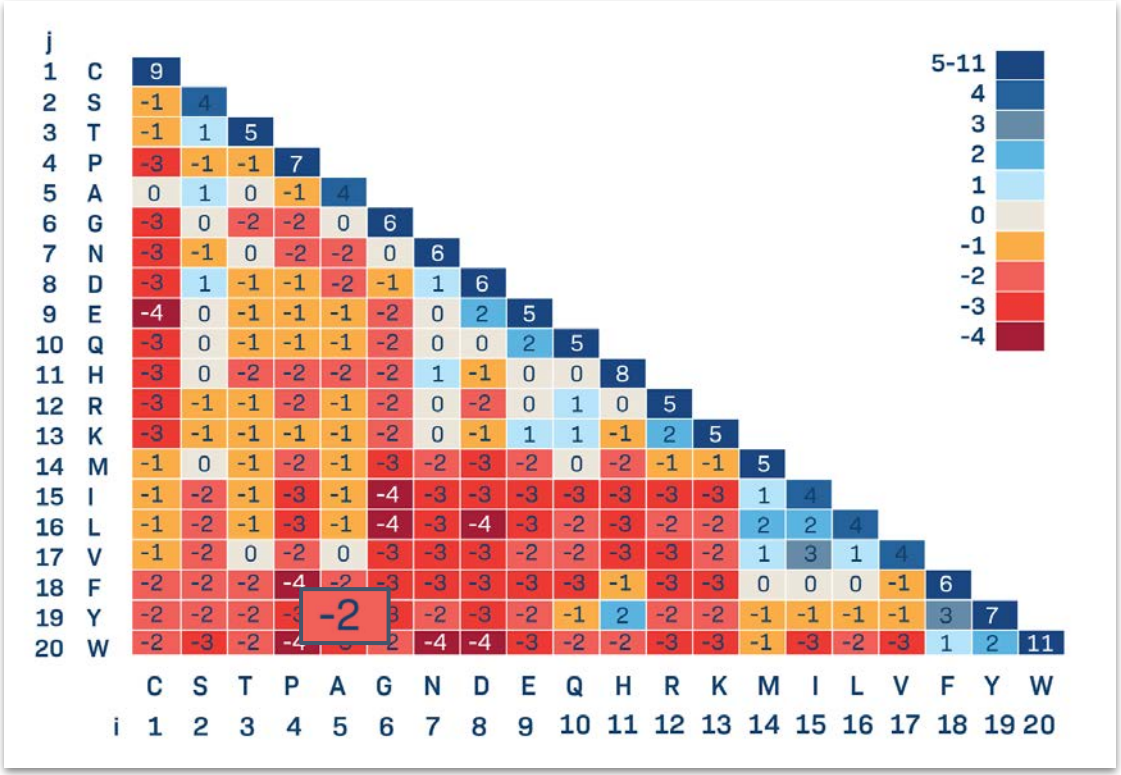
ASSSANYGYT
ASSIRAAETQY
(...)



TCRMatch

k = 1

ASSSANYGYT
ASSIRAAETQY
(...)



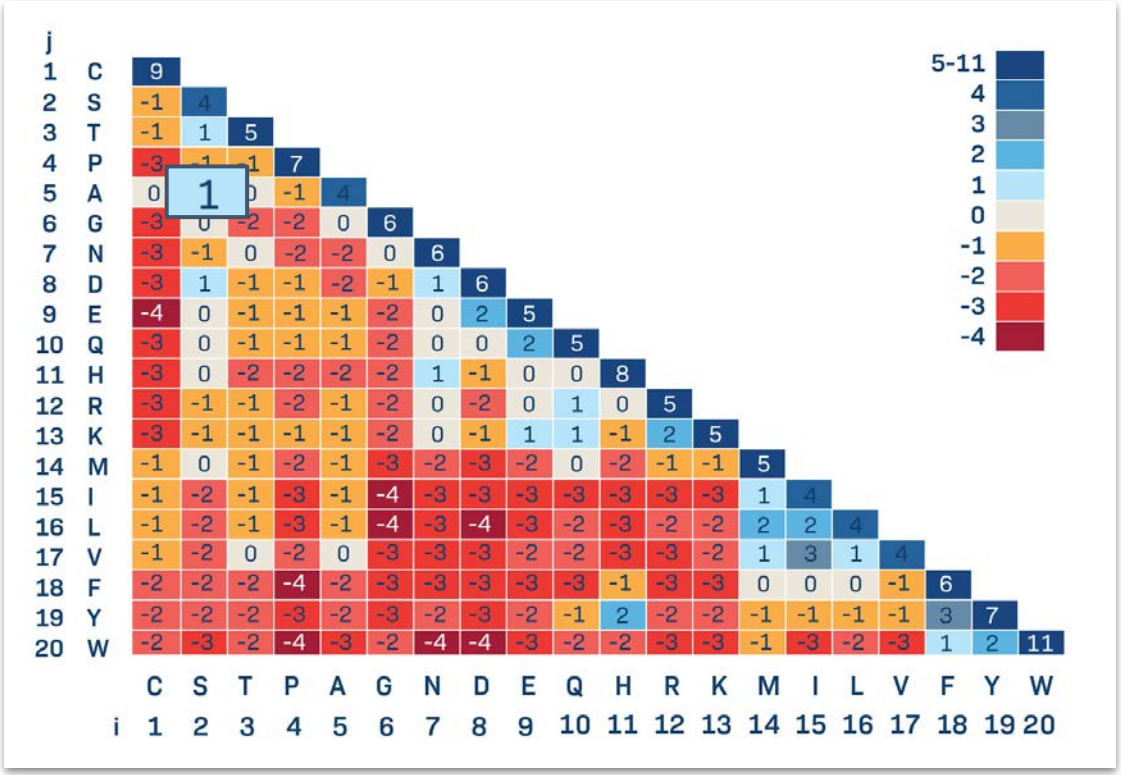
TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

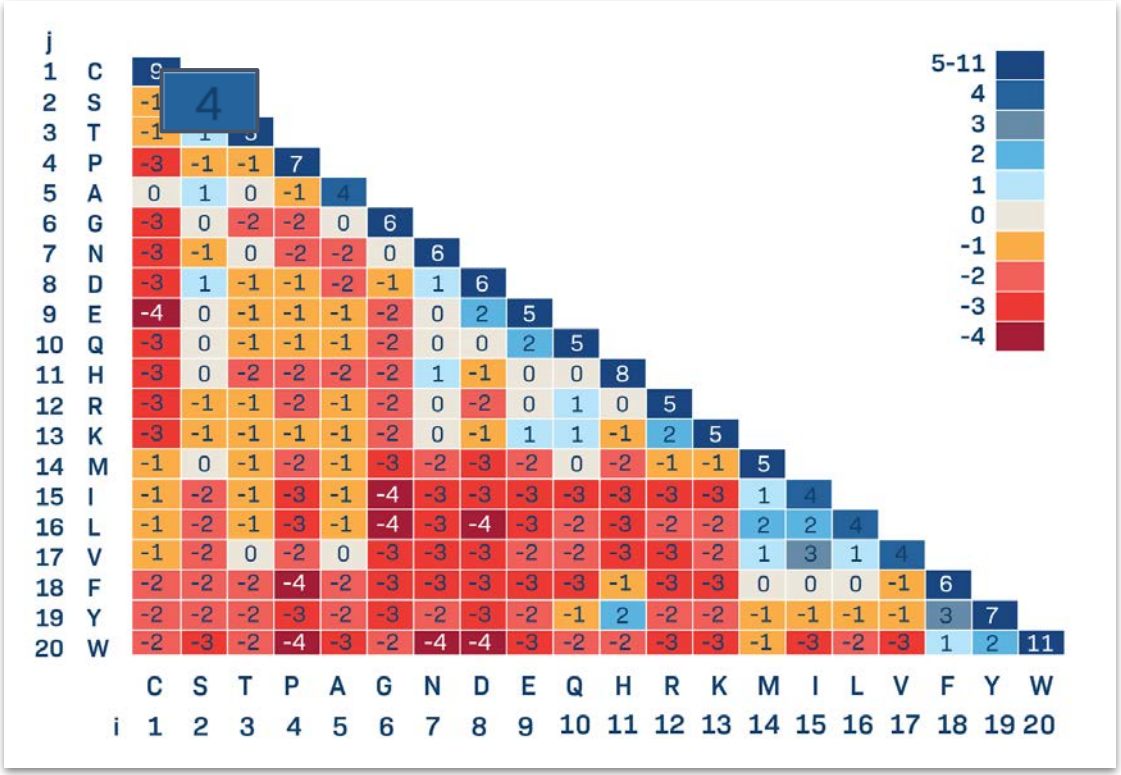
(...)



TCRMatch

k = 1

ASSSANYGYT
ASSIRAAETQY
(...)



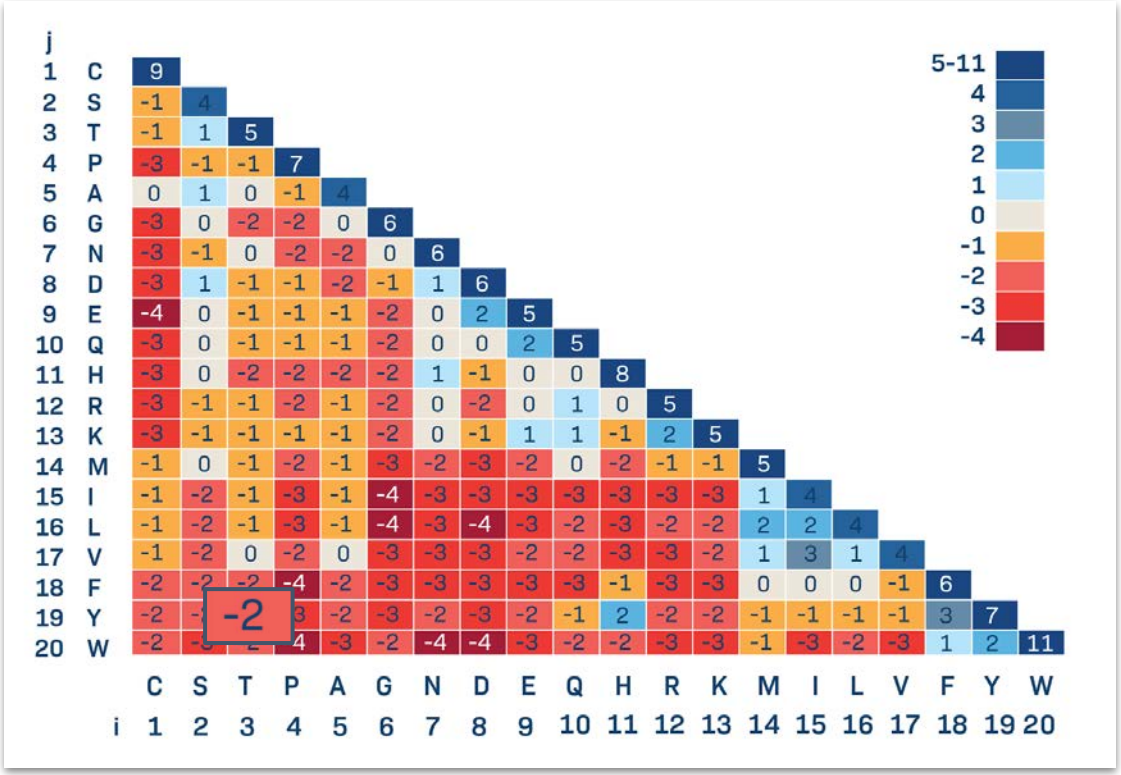
TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

(...)



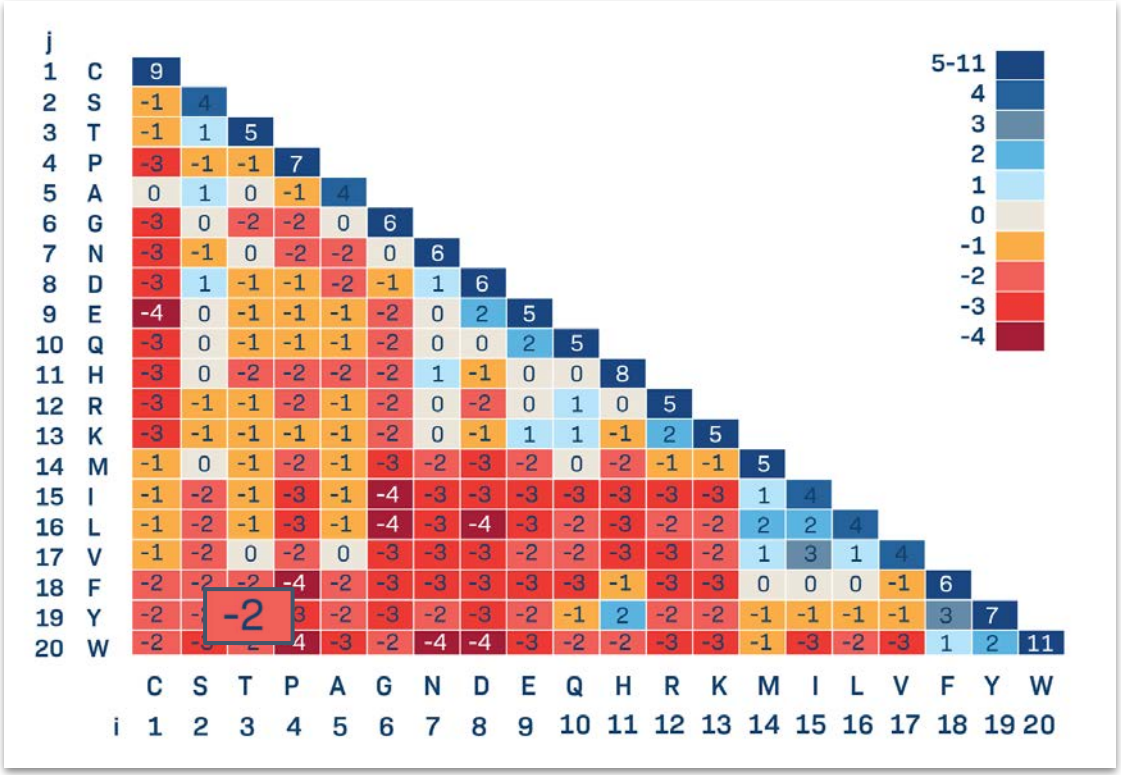
TCRMatch

k = 1

ASSSANYGYT

ASSIRAAETQY

(...)



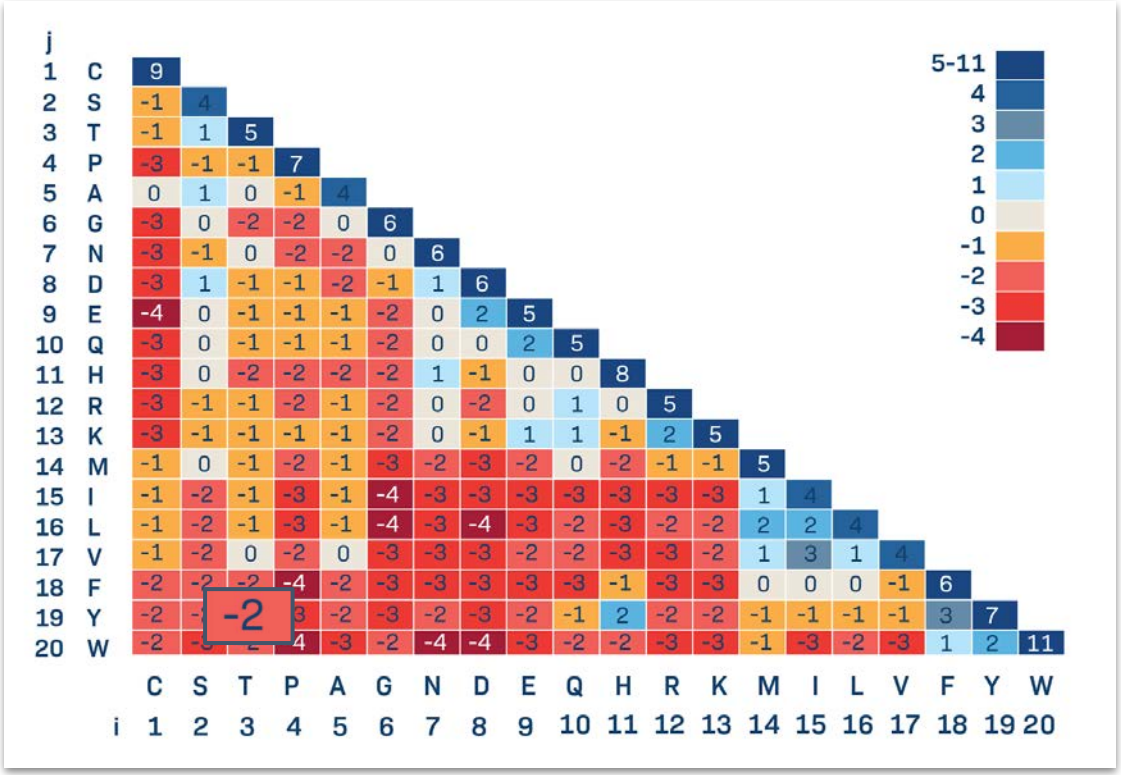
TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

(...)

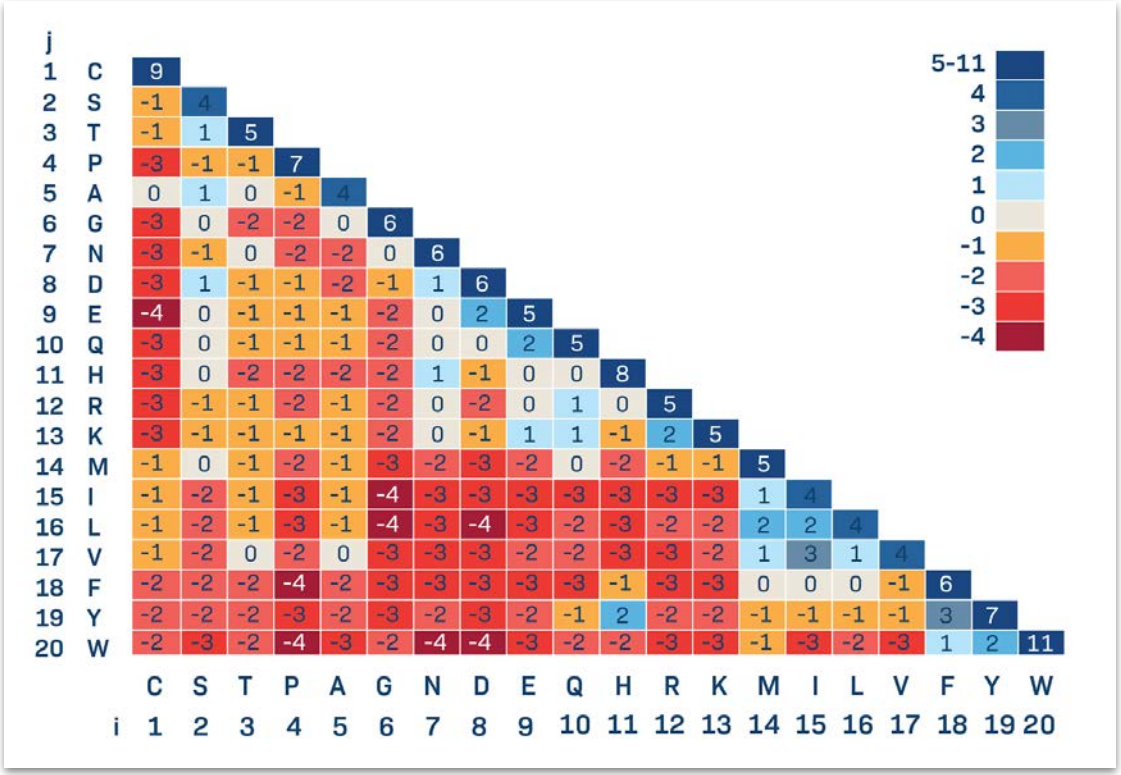


TCRMatch

k = 2

ASSSANYGYT

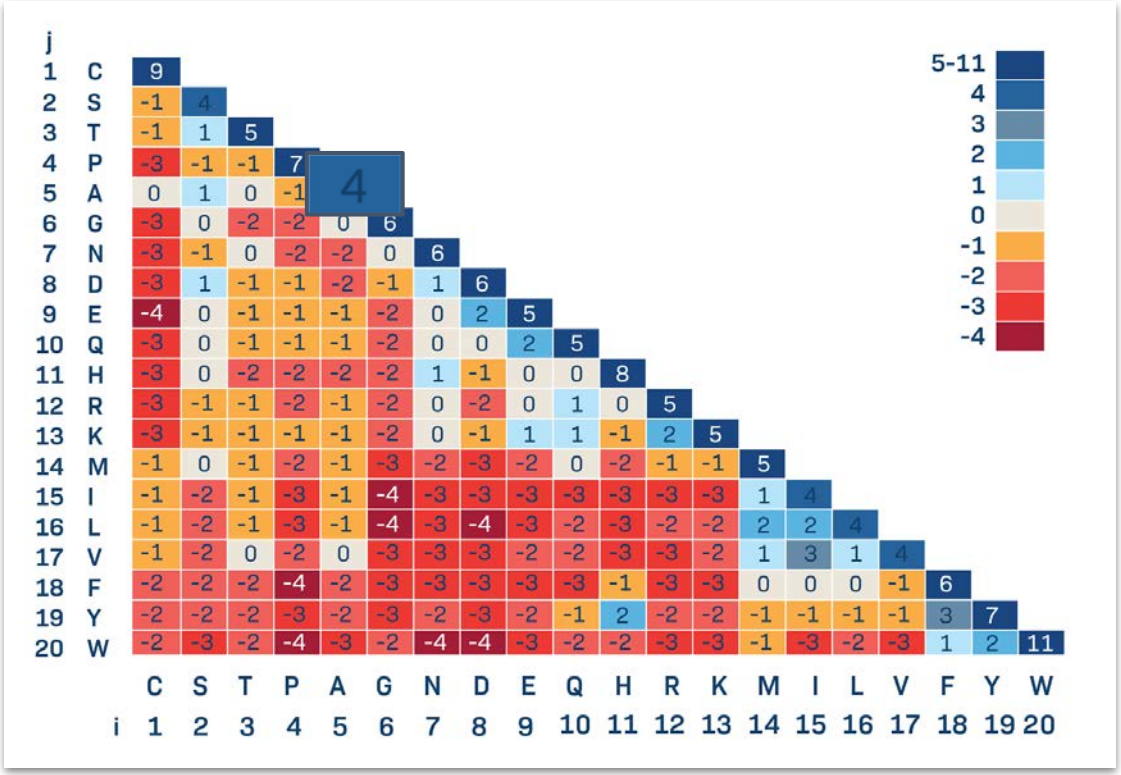
ASSIRAAETQY



TCRMatch

k = 2

ASSSANYGYT
ASSIRAAETQY

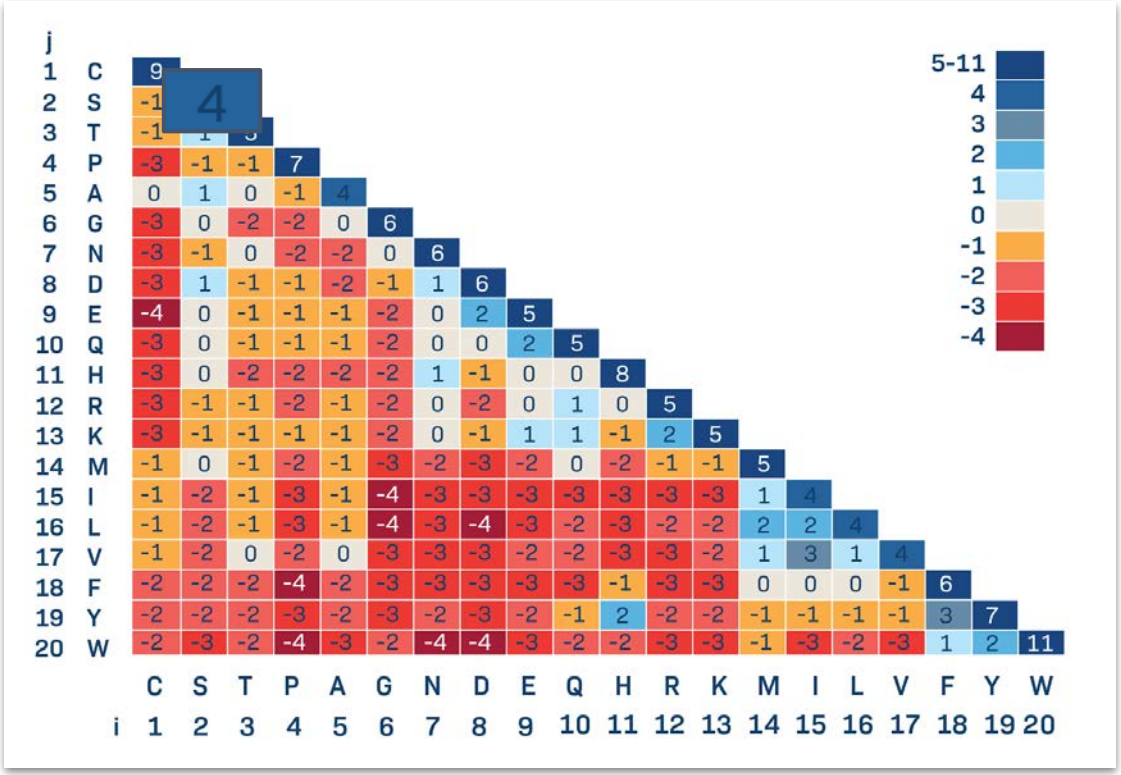


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

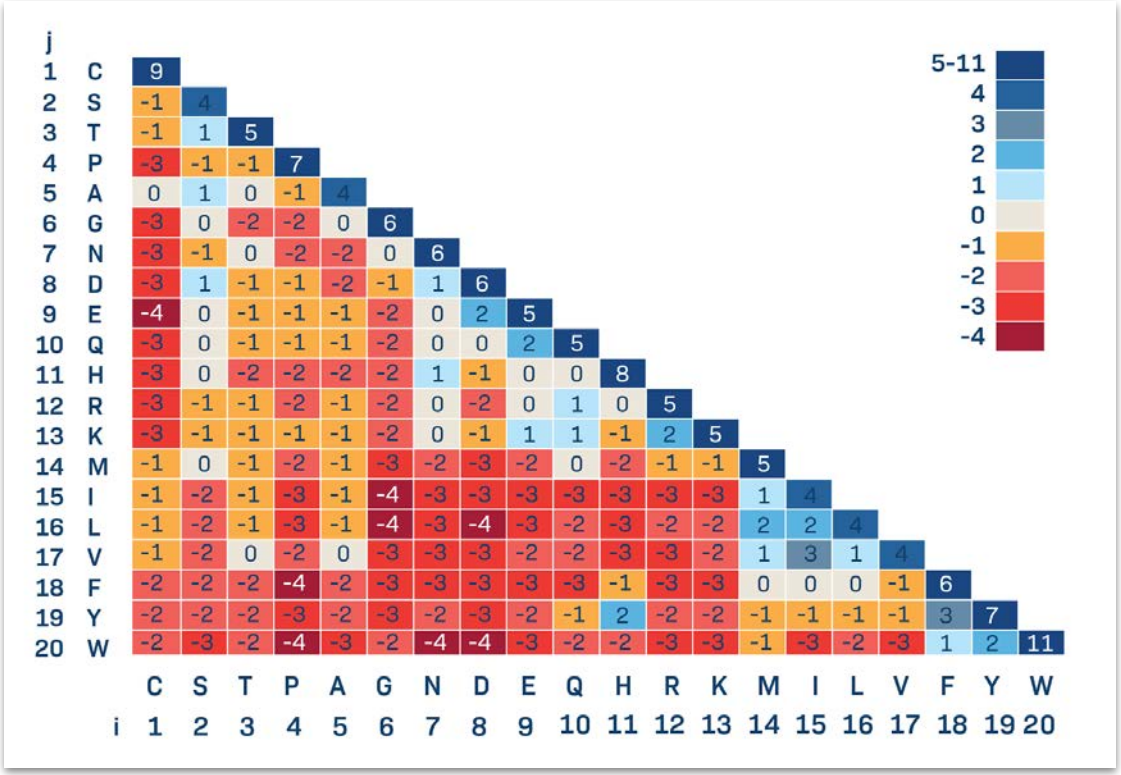


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

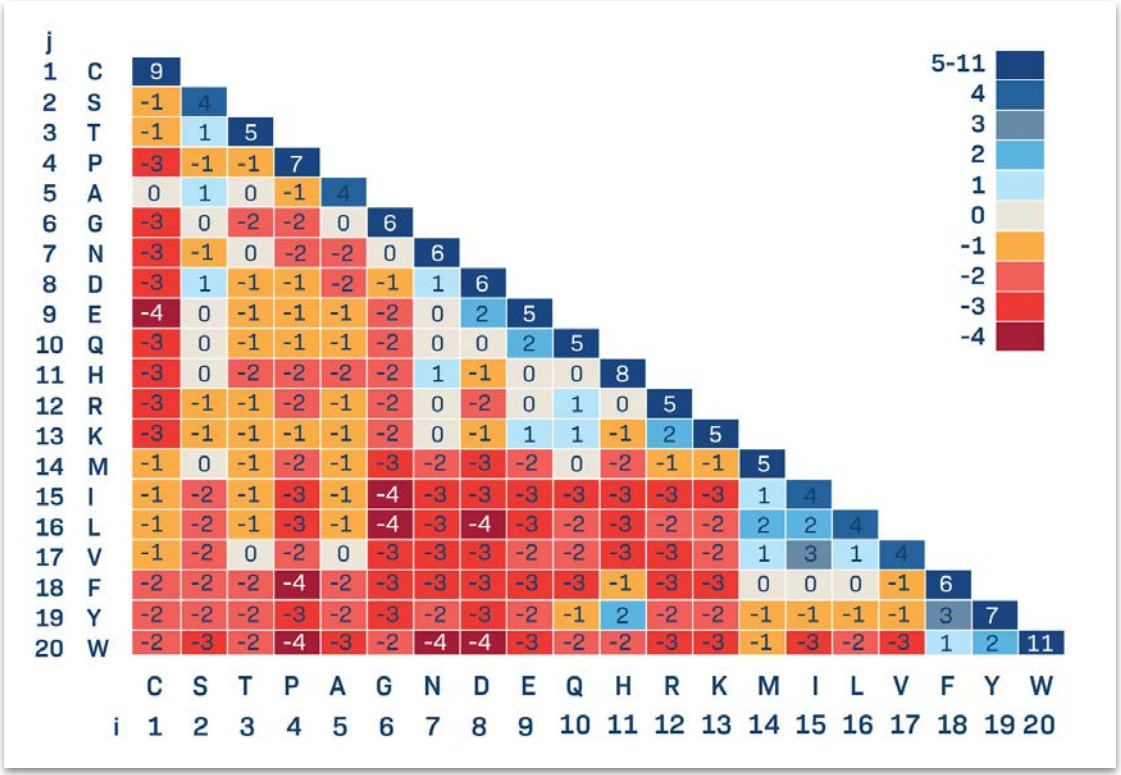


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

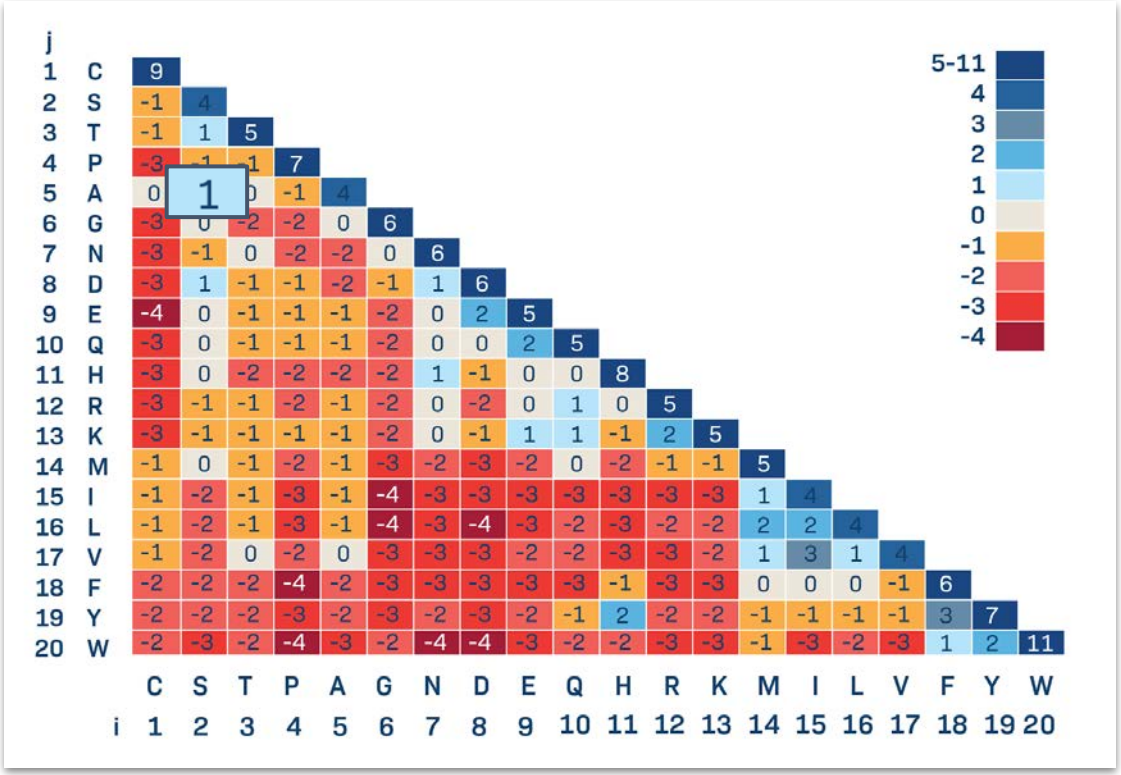


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

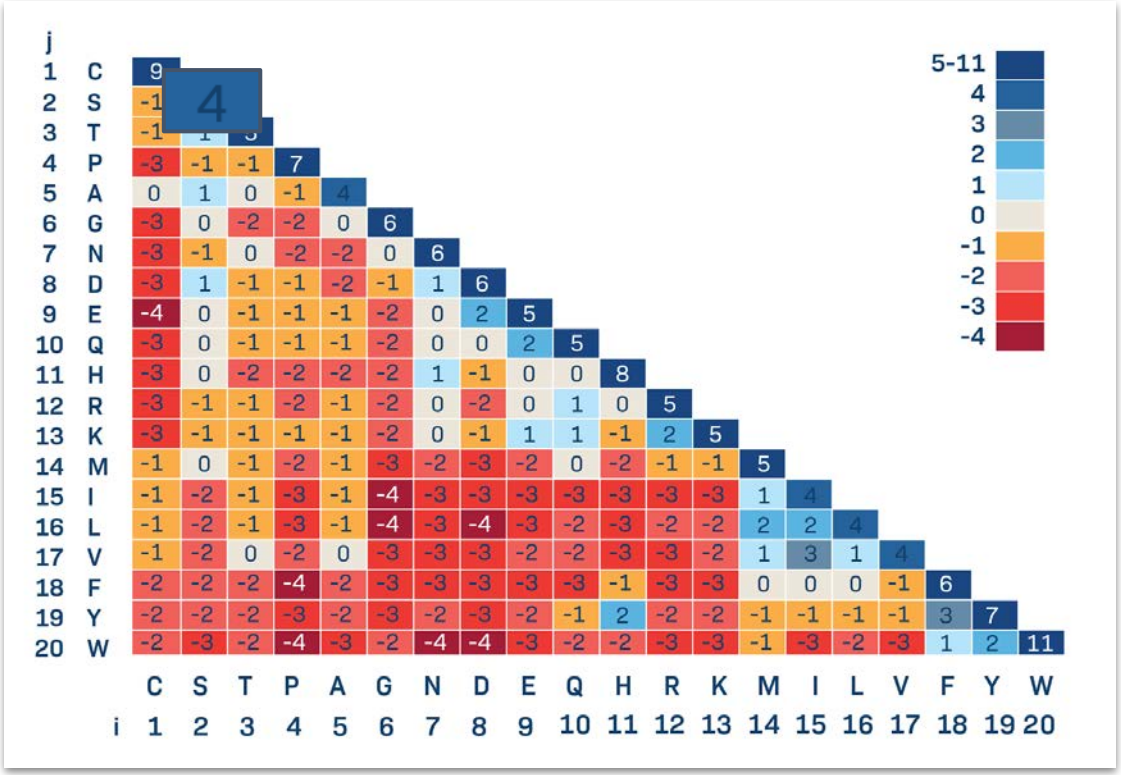


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

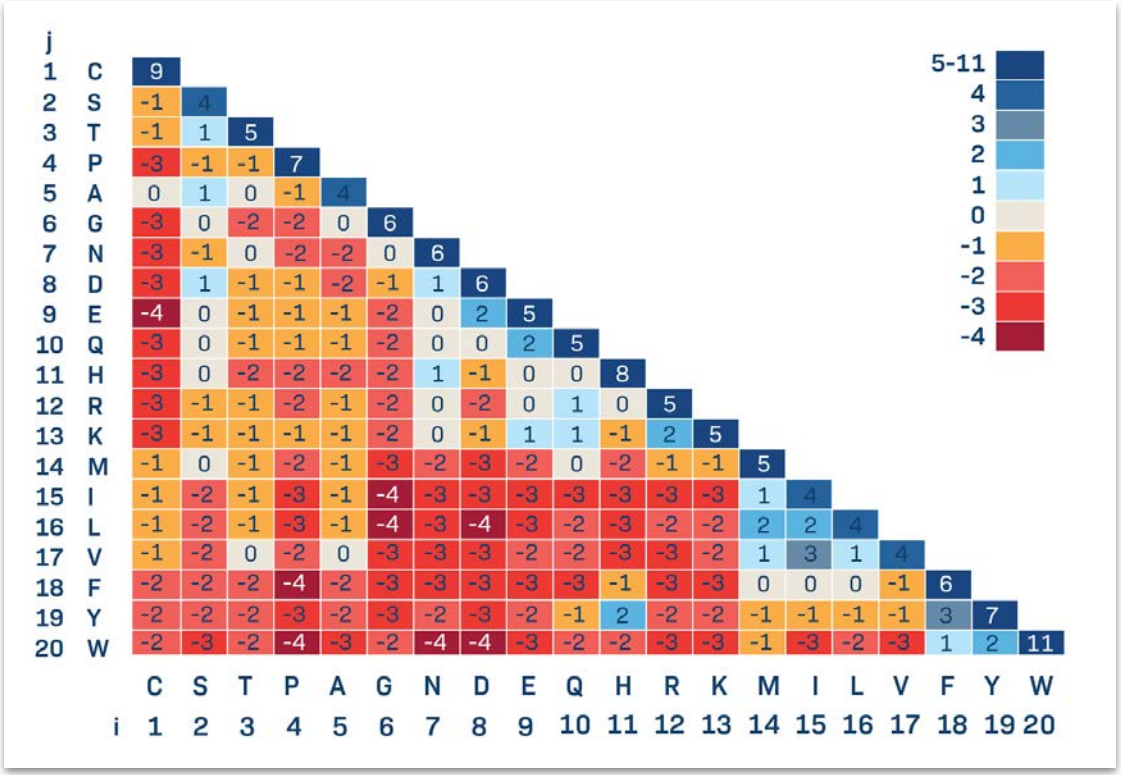


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

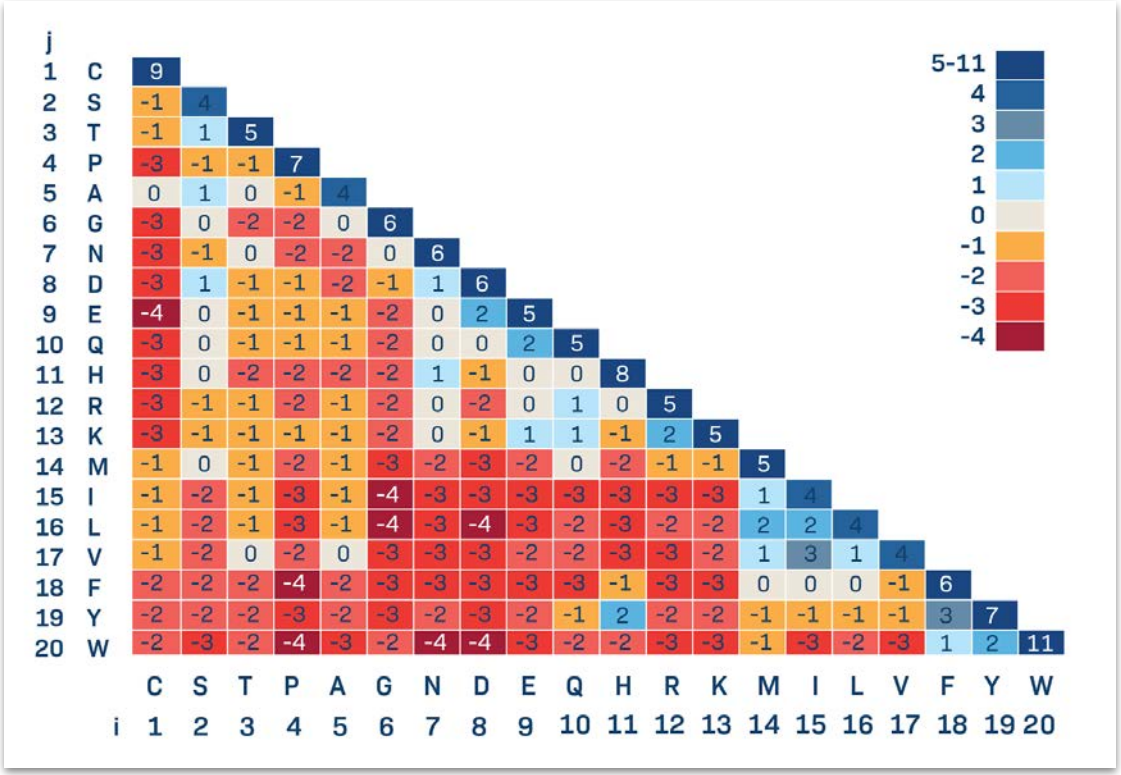


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

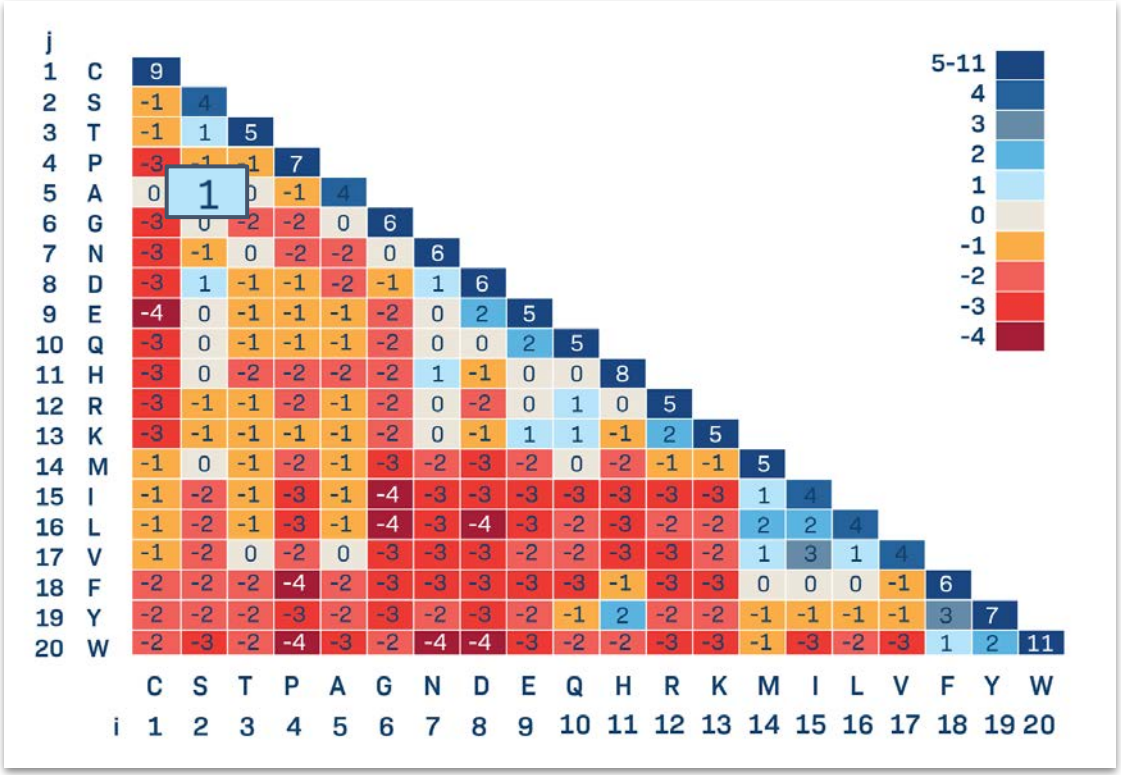


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

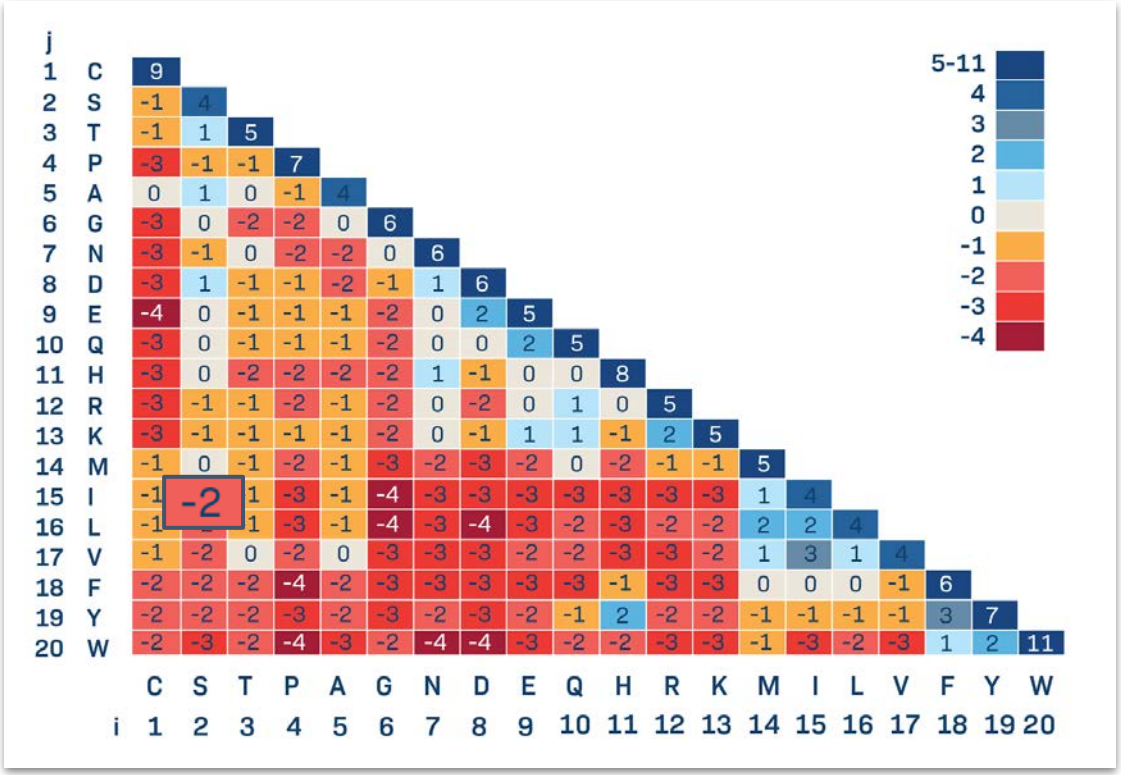


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

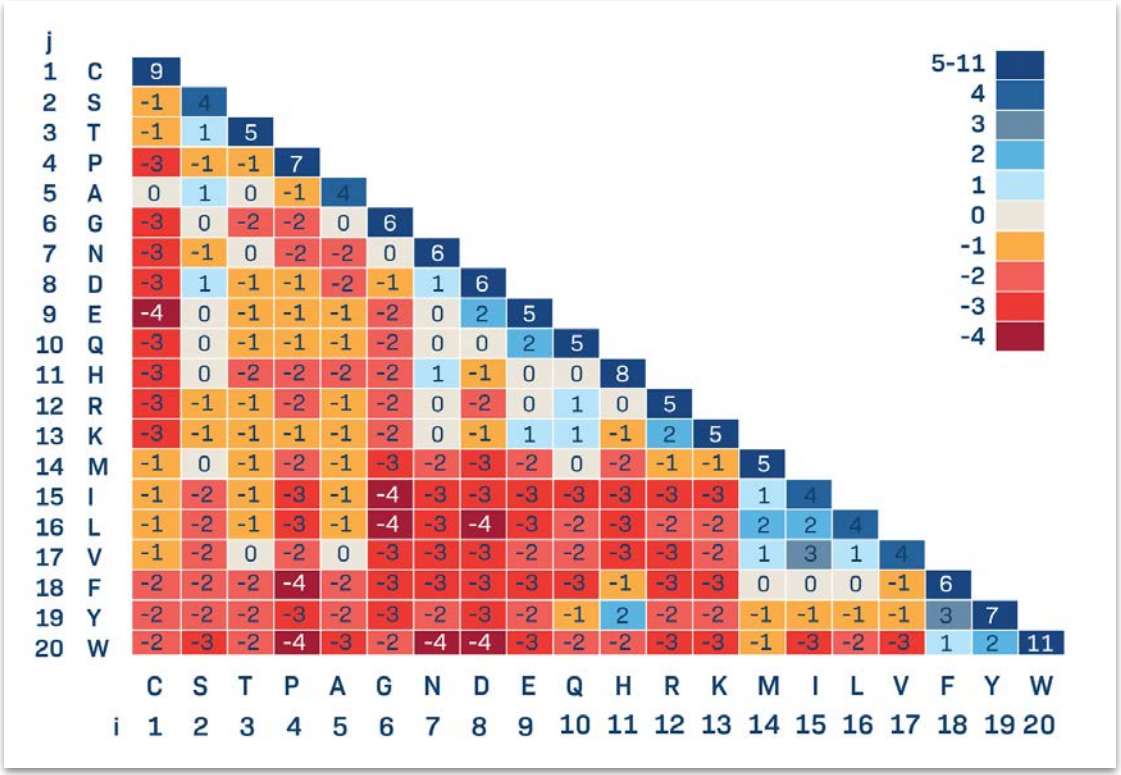


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

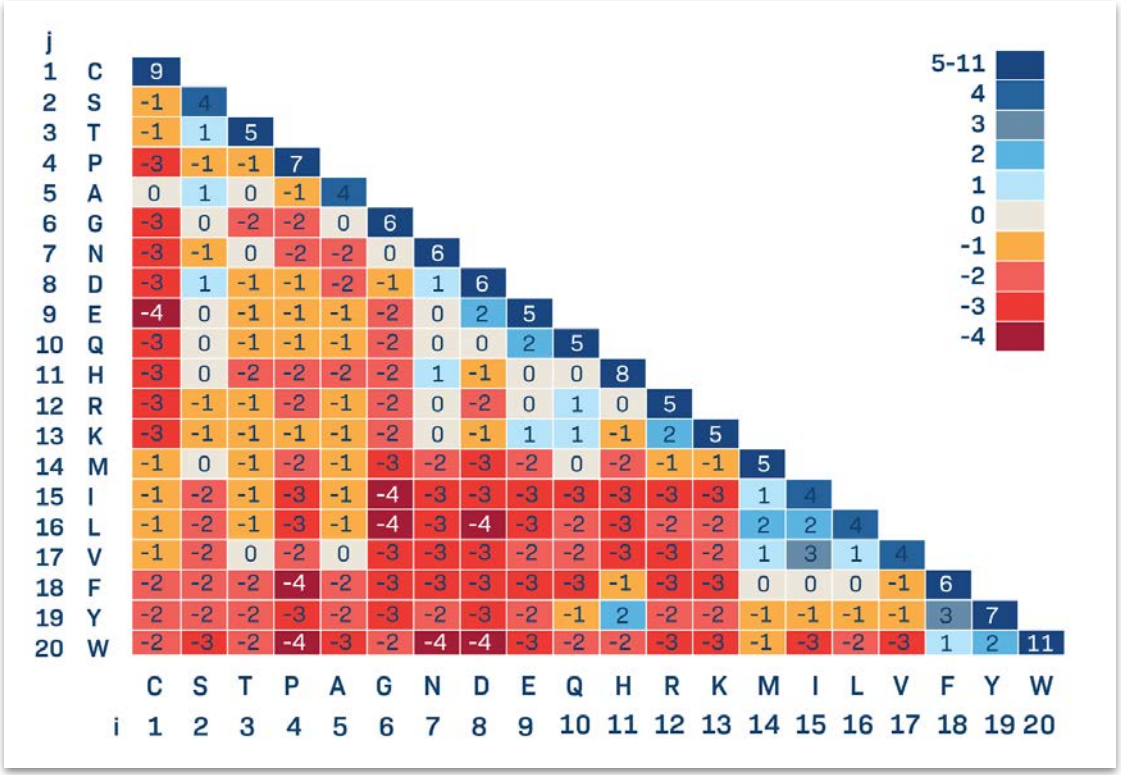


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

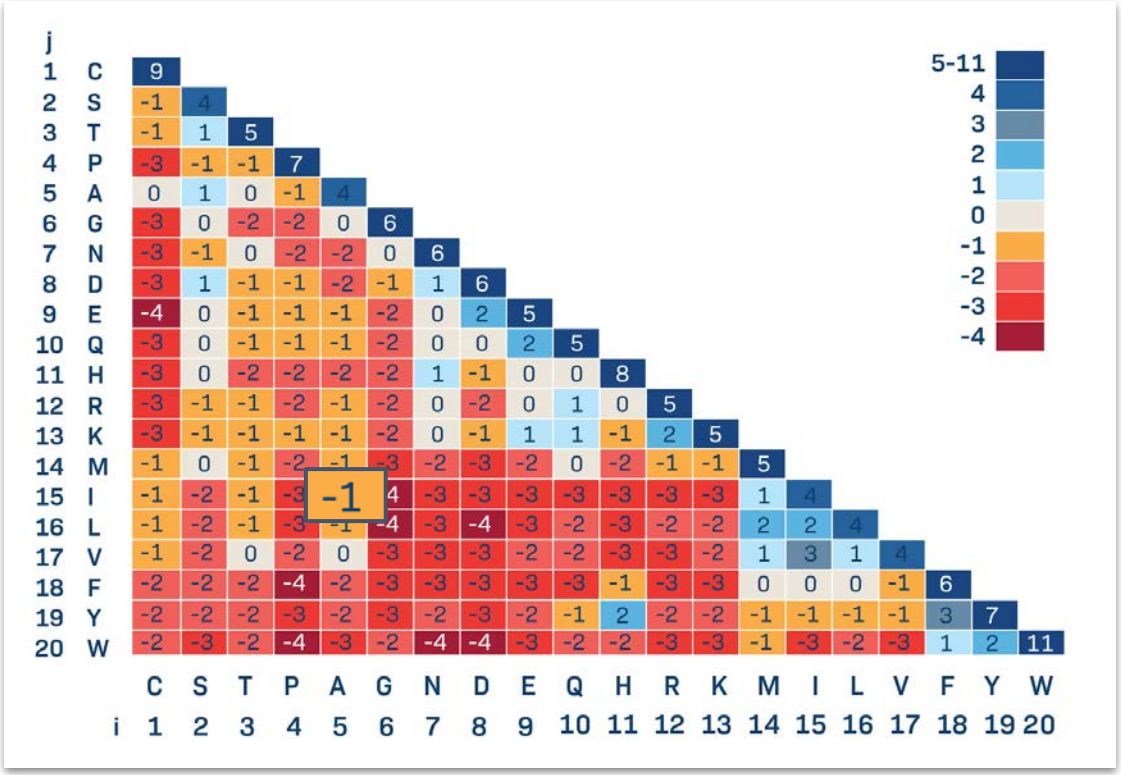


TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

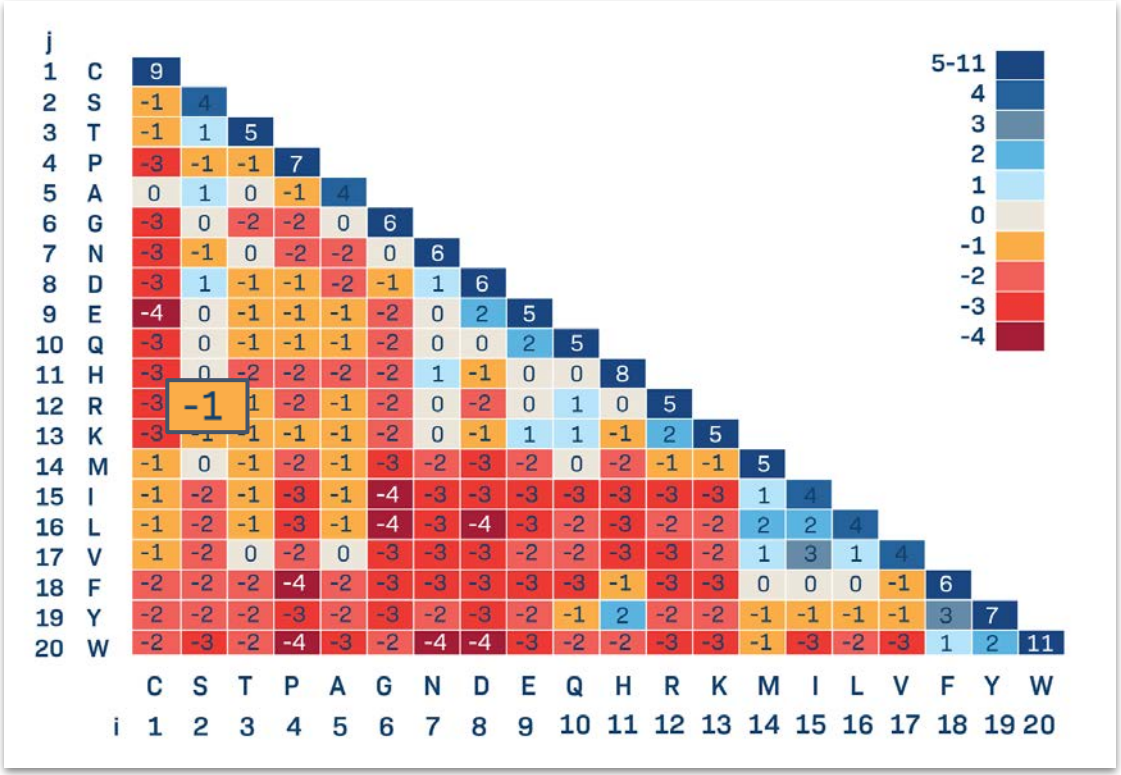


TCRMatch

k = 2

ASSSANYGYT

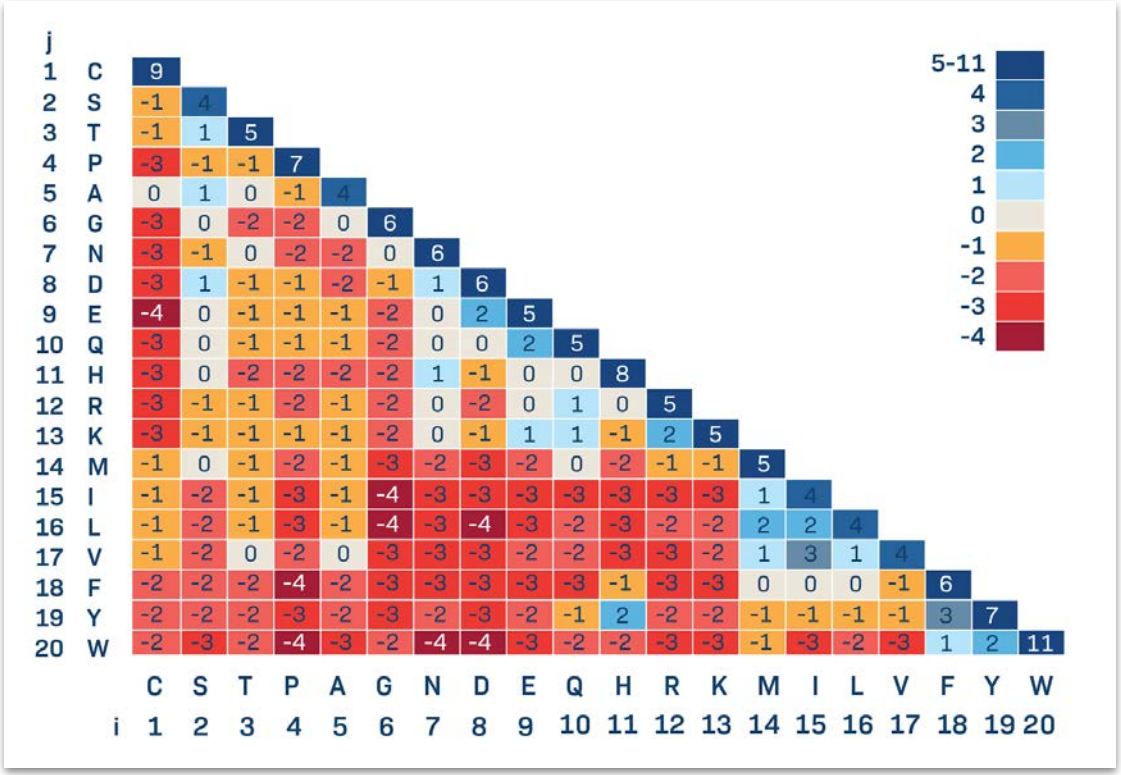
ASSIRAAETQY



TCRMatch

k = 2

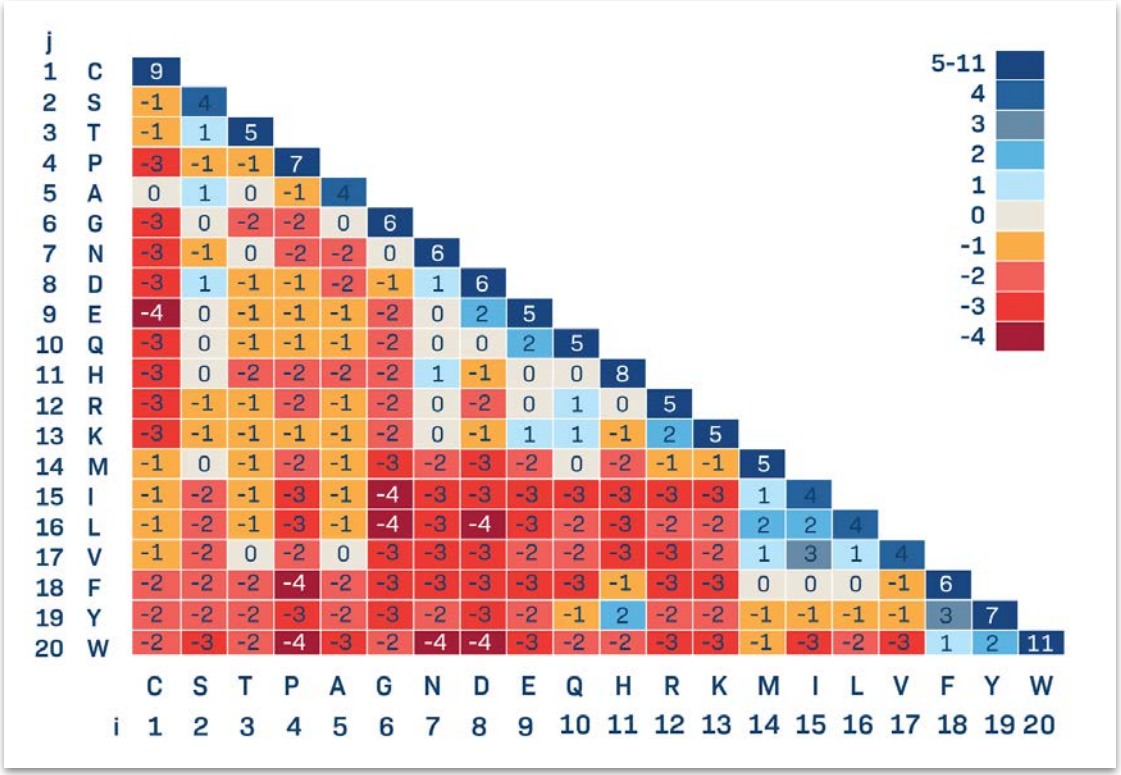
ASSSANYGYT
ASSIRAAETQY
(...)



TCRMatch

k = 2

ASSSANYGYT
ASSIRAAETQY
(...)



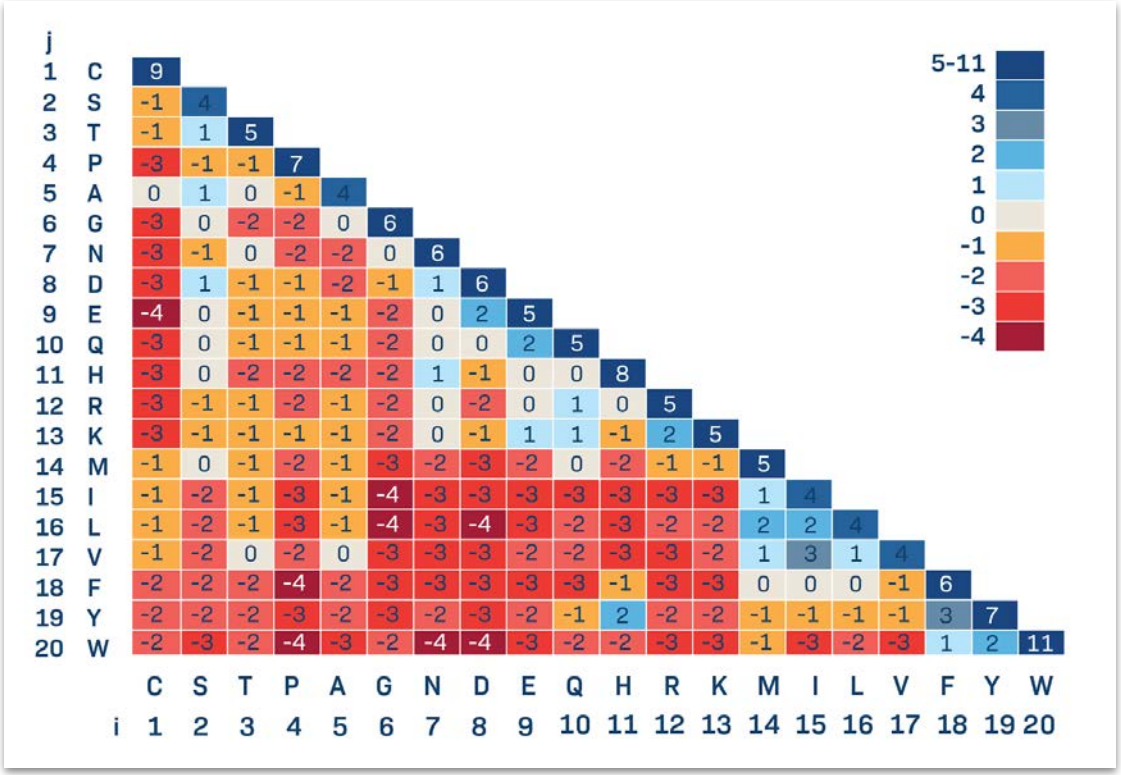
TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

(...)



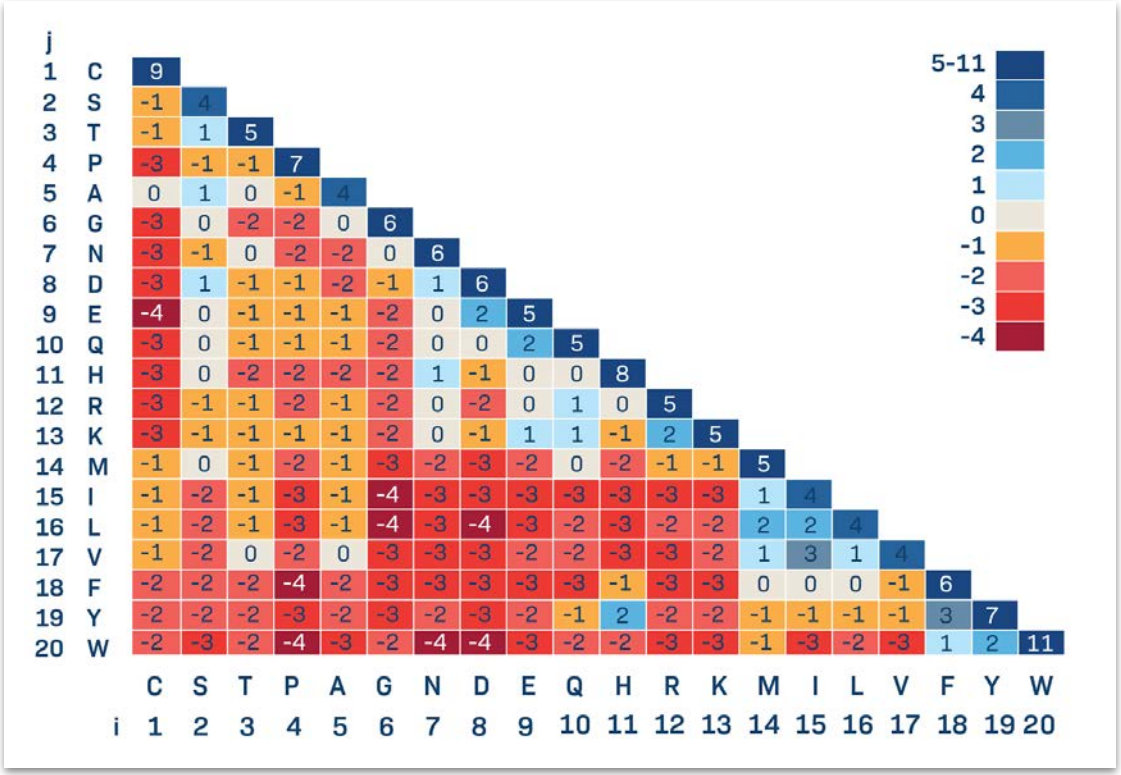
TCRMatch

k = 2

ASSSANYGYT

ASSIRAAETQY

(...)

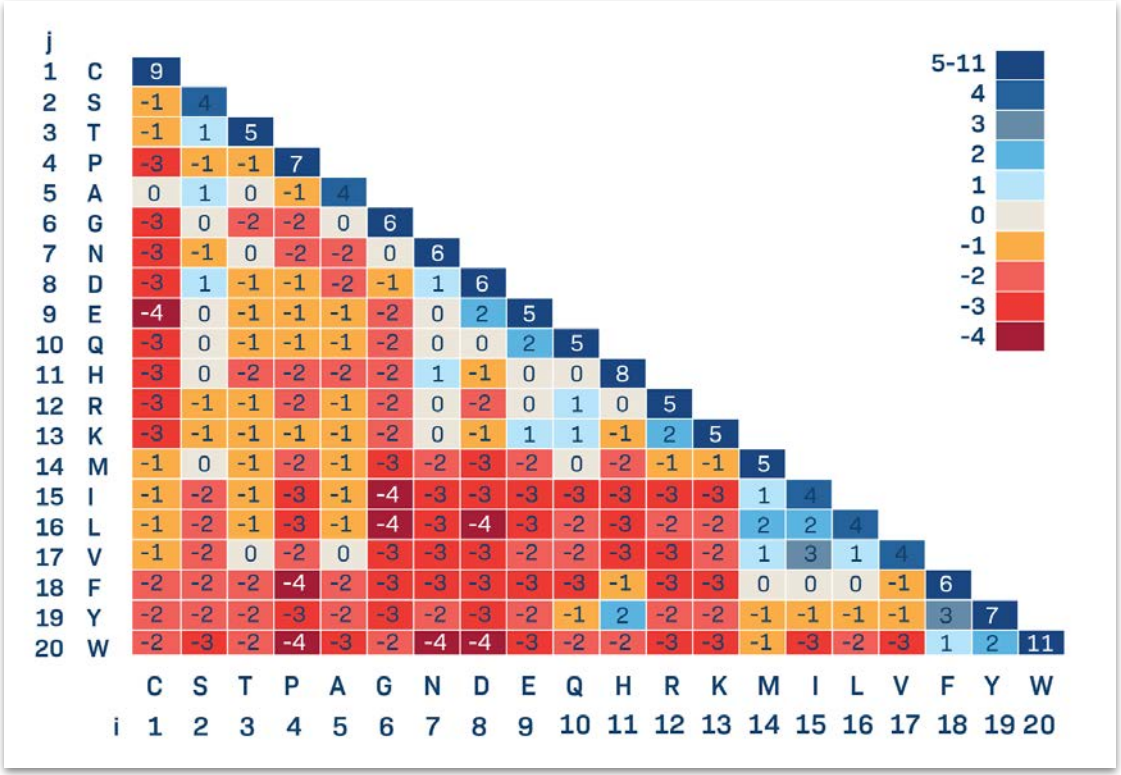


TCRMatch

k = 3

ASSSANYGYT

ASSIRAAETQY

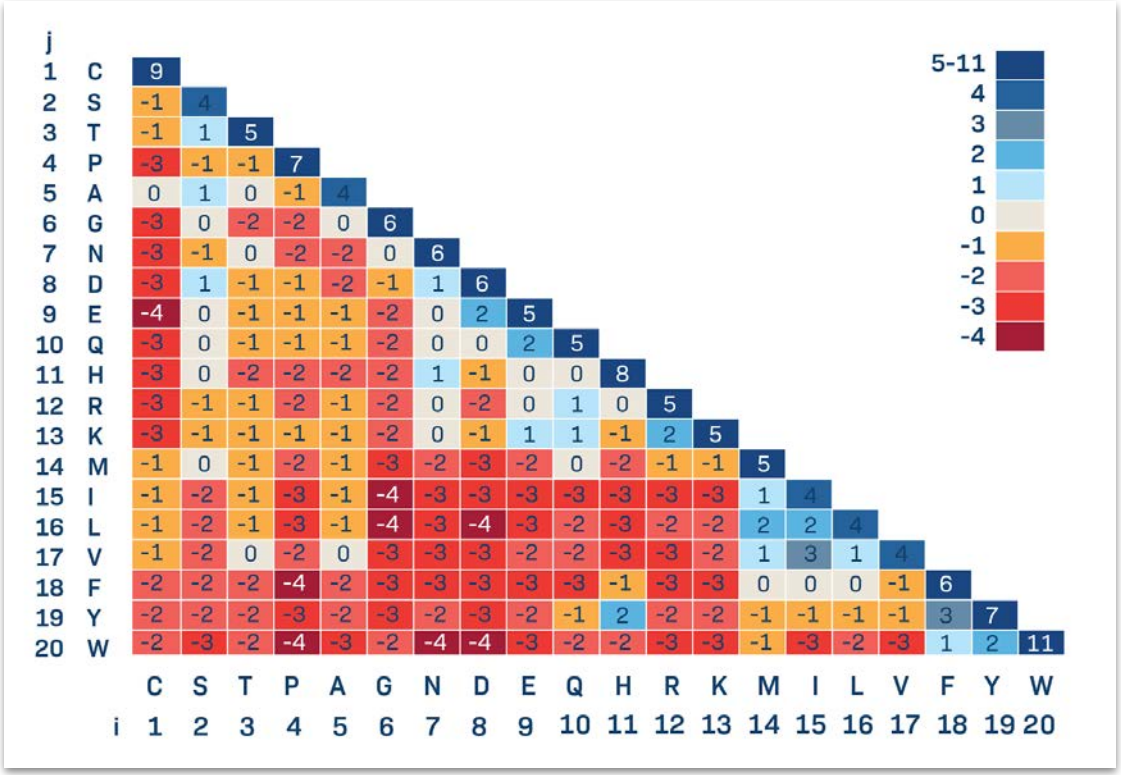


TCRMatch

k = 4

ASSSANYGYT

ASSIRAAETQY

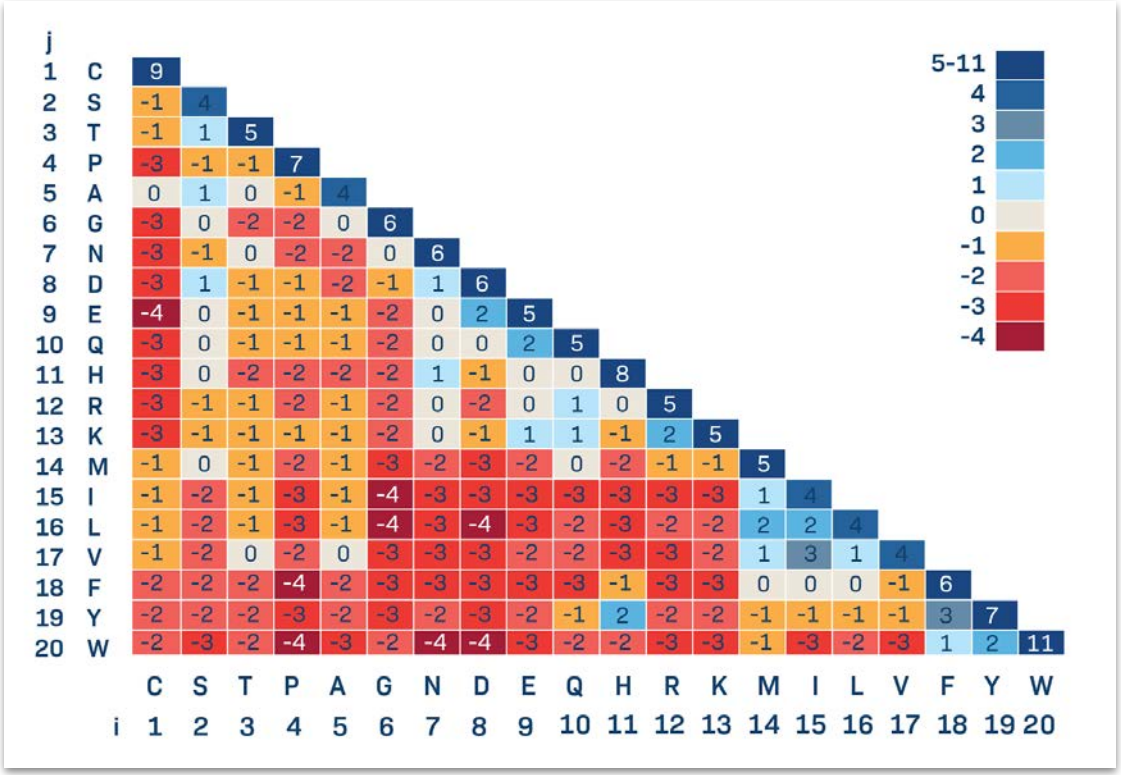


TCRMatch

k = 5

ASSSANYGYT

ASSIRAAETQY

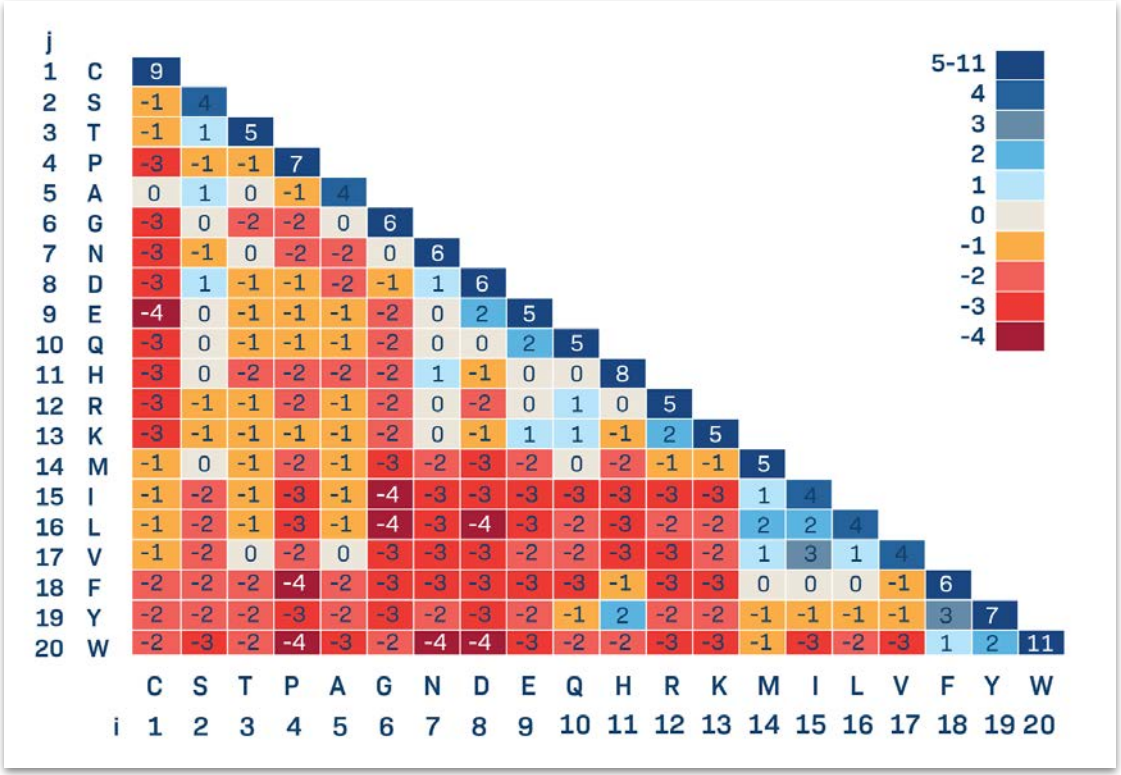


TCRMatch

k = 10

ASSSANYGYT

ASSIRAAETQY

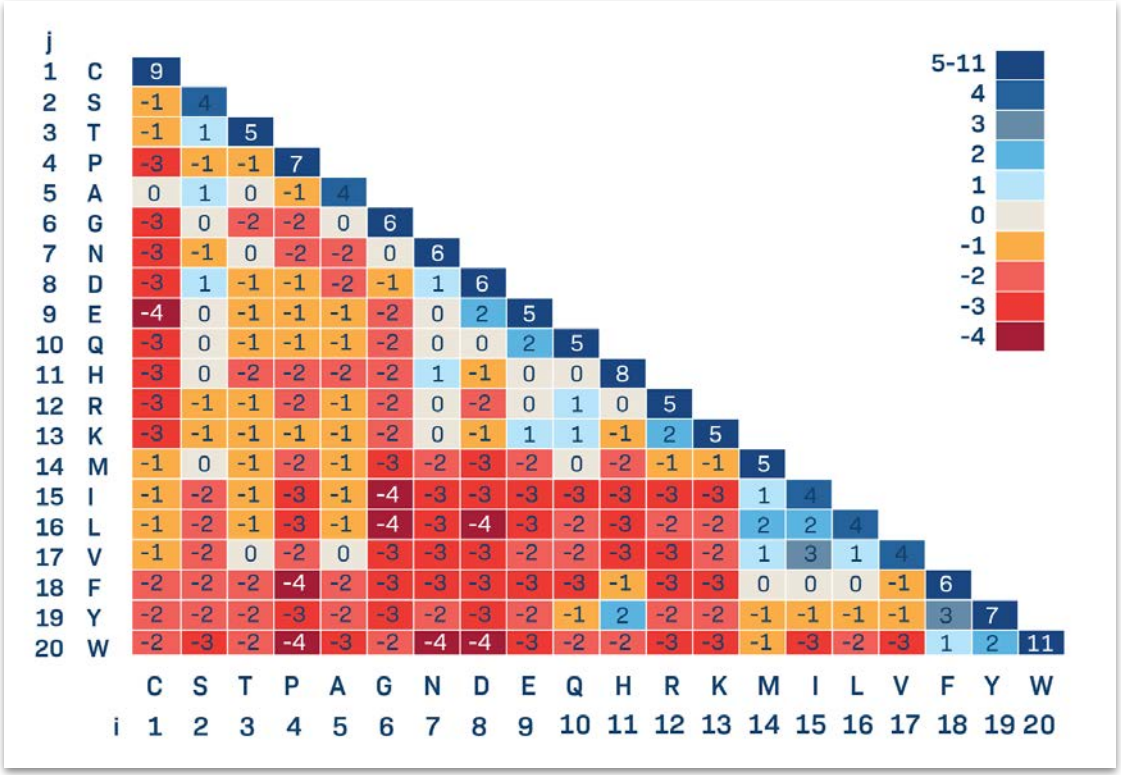


TCRMatch

k = 10

ASSSANYGYT

ASSIRAAETQY

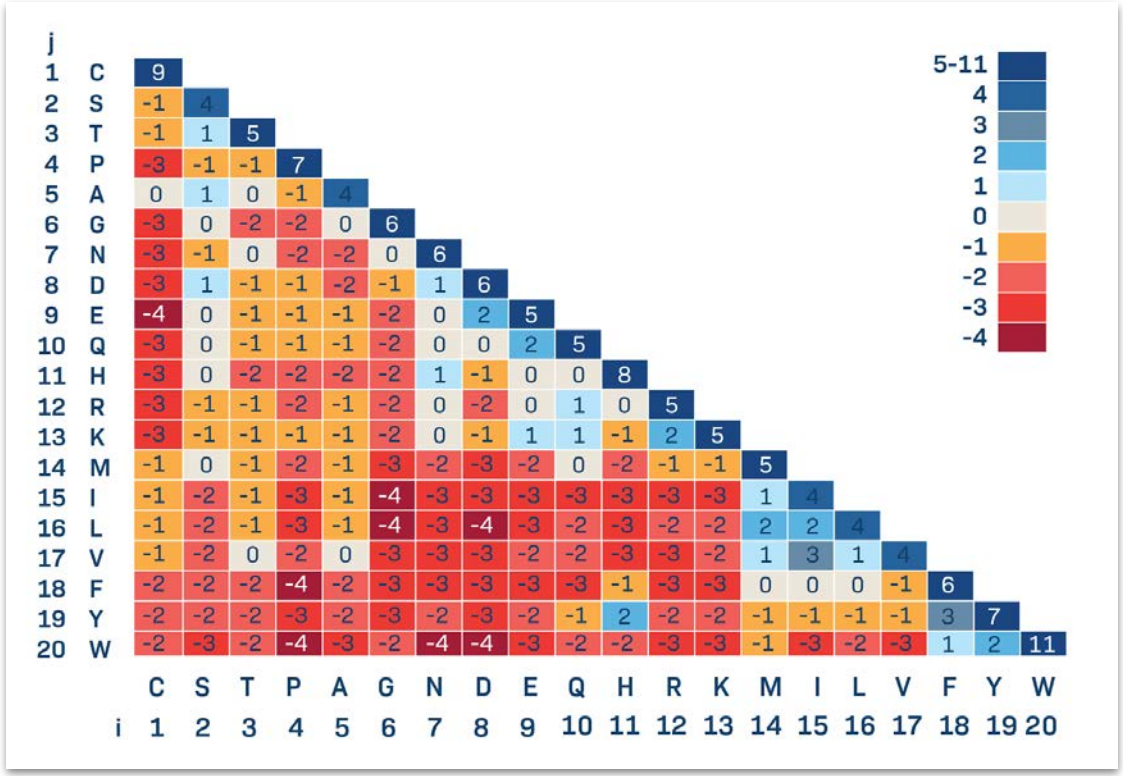


TCRMatch

k = 10

ASSSANYGYT

ASSIRAAETQY

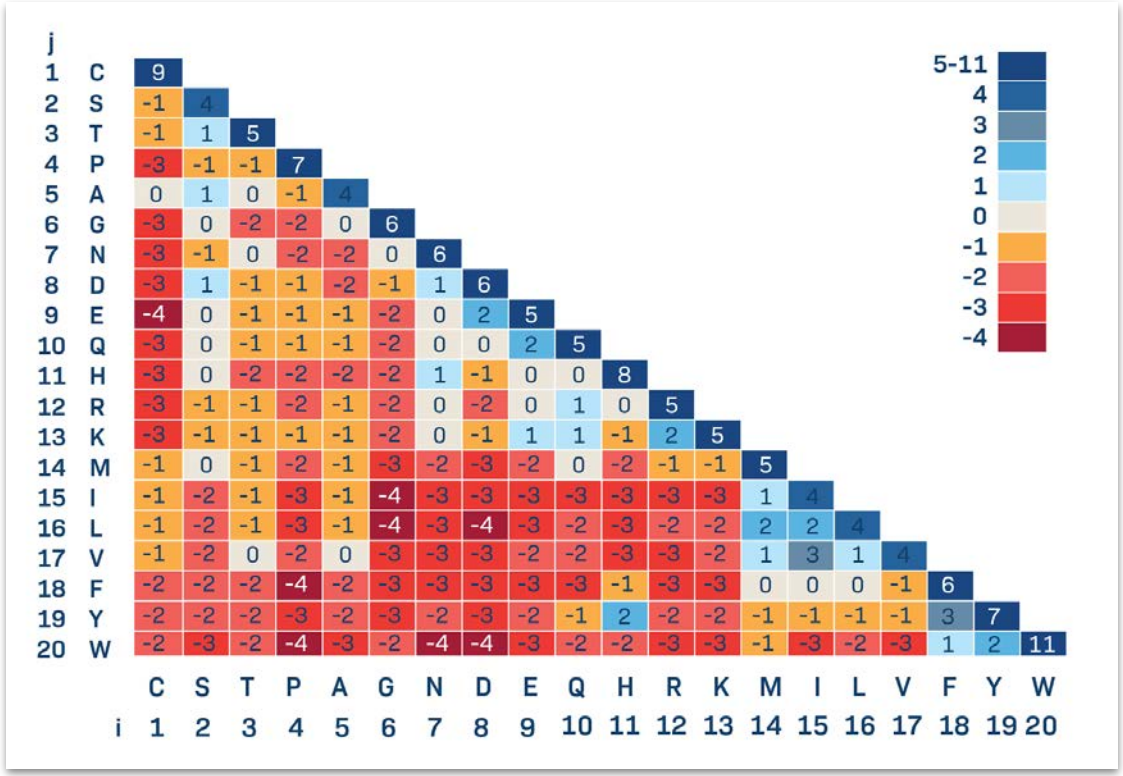


TCRMatch

k = 10

ASSSANYGYT

ASSIRAAETQY



TCRMatch

k = 10

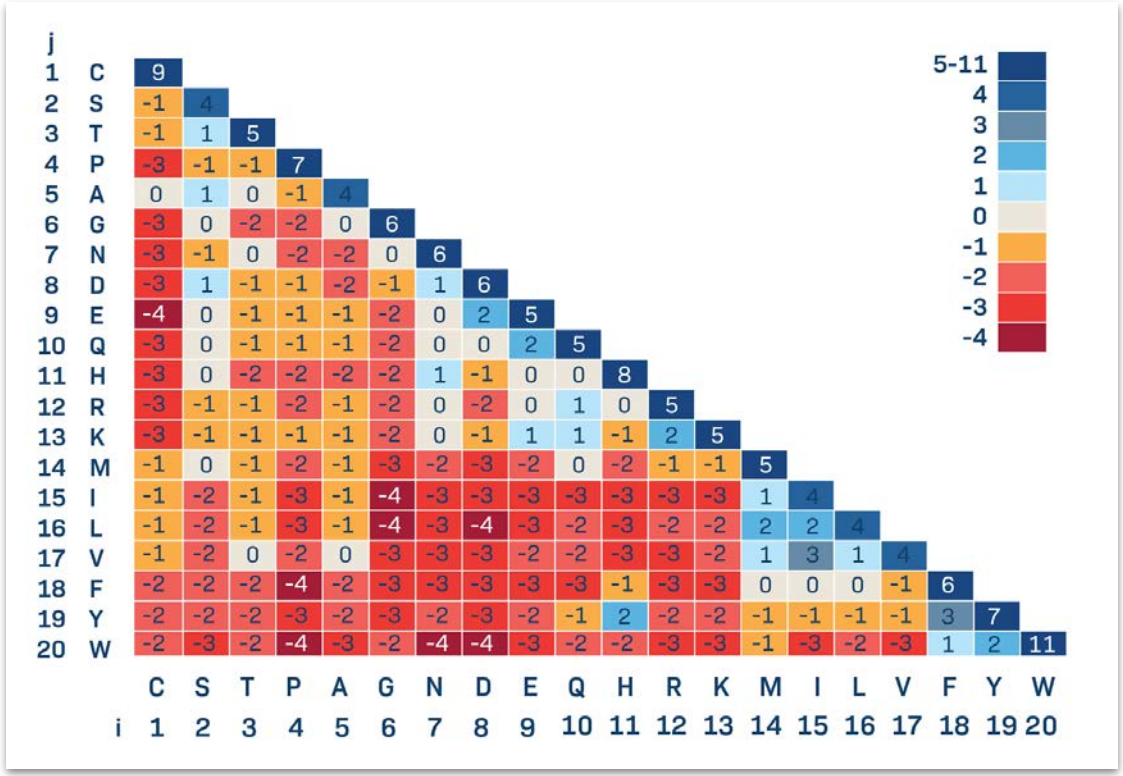
ASSSANYGYT

ASSIRAAETQY

Final score:

0 - 1

(1: perfect match)



TCRMatch

TCRMatch output

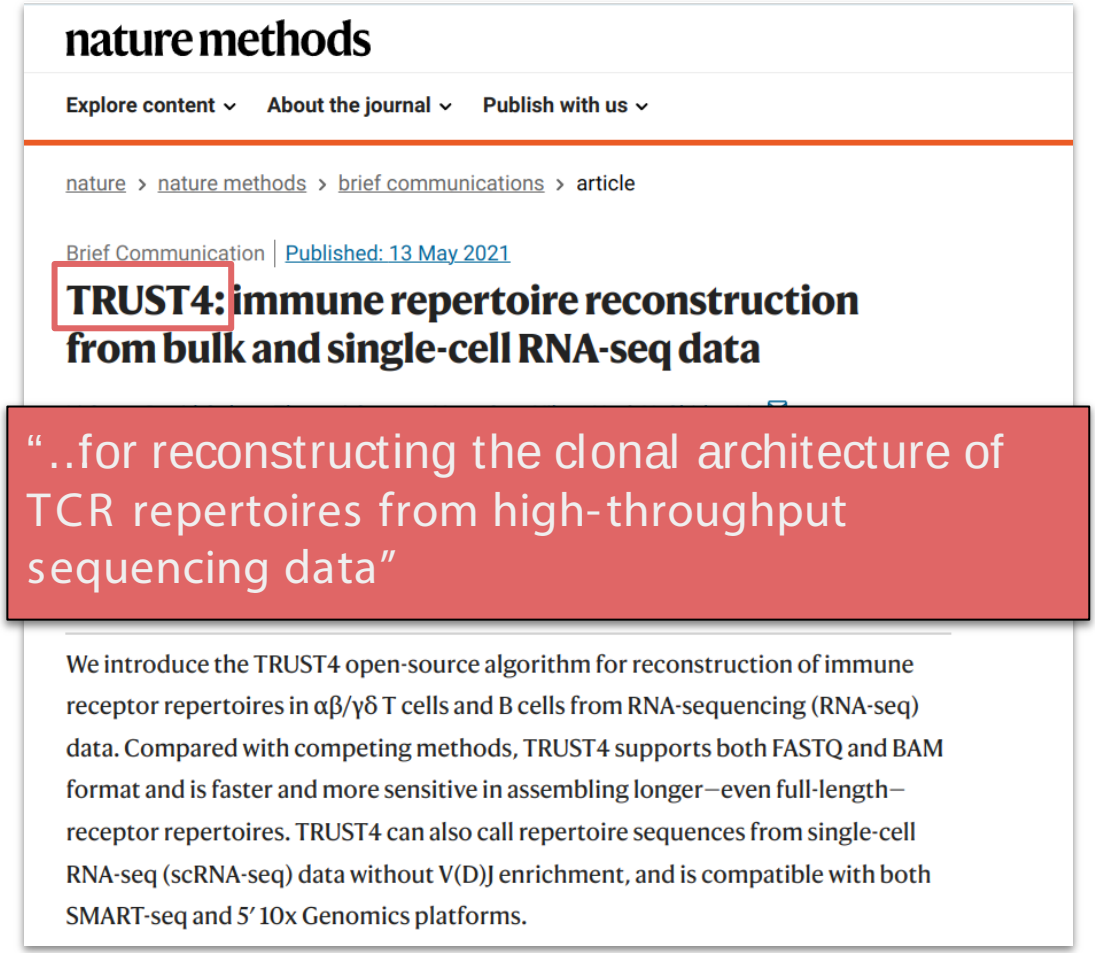
- Original CDR3β - query
- Matching CDR3β - IEDB
- Matching epitope
- Receptor group
- TCRMatch score
- Antigen
- Organism
- TRUST4

input_sequence	match_sequence	score	receptor_group	epitope	antigen	organism
AISEVGVGQPQH	ASSQDRDTQY	0.7716	47	VMAPRTLIL	glycoprotein	Human herpesvirus 5 (Human cytomegalovir
AISEVGVGQPQH	ASGDAGGGYEQY	0.7891	8606	ASQKRPSQRSK	Myelin basic protein	Mus musculus (mouse)
AISEVGVGQPQH	ASGDAGGGYEQY	0.7891	18226	ASQKRPSQR	Myelin basic protein	Mus musculus (mouse)
AISEVGVGQPQH	ASTYHGTGY	0.7686	57	GELIGILNAAKVPAD,GELIGTLNA	triosephosphate isomerase 1	Homo sapiens (human)
AISEVGVGQPQH	ASSYLGNTGELF	0.7986	94	SLLMWITQC	Cancer/testis antigen 1	Homo sapiens (human)
AISEVGVGQPQH	ASGDASGAETLY	0.8028	1243	ASQKRPSQR,ASQKRPSQR,AS	Myelin basic protein,myelin bas	Mus musculus (mouse),Mus musculus (mous
AISEVGVGQPQH	ASGDASGGNTLY	0.7859	103	RGGASQYRPSQ		
AISEVGVGQPQH	ASGDFWGDTLY	0.7353	34695	GSLFLKTPKIVSSKD	L-lactate dehydrogenase A cha	Mus musculus (mouse)
AISEVGVGQPQH	ASGDFWGDTLY	0.7353	18548	WIYVYRPM		
AISEVGVGQPQH	ASRYRDDSYNEQF	0.7206	29969	FLRGRAYGL	nuclear antigen EBNA-3	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	ASRYRDDSYNEQF	0.7206	109	EENLLDFVRF	Epstein-Barr nuclear antigen 6	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	ASRYRDDSYNEQF	0.7206	27756	EENLLDFVRF	Epstein-Barr nuclear antigen 6	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	ASRPGLAGGRPEQY	0.7822	56537	LLFGYPVYV	Protein Tax-1	Human T-cell leukemia virus type I (Human
AISEVGVGQPQH	ASRPGLMSAQPEQY	0.7903	21075	LLFGYPVYV	Protein Tax-1	Human T-cell leukemia virus type I (Human
AISEVGVGQPQH	ASGGGGTLY	0.754	27474	SIYRYYGL		
AISEVGVGQPQH	ASGGGGTLY	0.754	21188	AMKRHGLDNYREYSLGN		
AISEVGVGQPQH	ASPGLAGEYEQY	0.8053	116	LPEPLPQGQLTAY,LPEALPQGQ	BZLF1,Trans-activator protein	Human herpesvirus 4 (Epstein Barr virus),Hu
AISEVGVGQPQH	ASSYVGNTGELF	0.8131	181935	SLLMWITQV		
AISEVGVGQPQH	ASSYVGNTGELF	0.8131	179539	SLLMWITQC	Cancer/testis antigen 1	Homo sapiens (human)
AISEVGVGQPQH	ASSDWVSYEQY	0.774	1926	ALWGFFPVL,FAPGFFPYL	chromosome 15 open reading f	Homo sapiens (human)
AISEVGVGQPQH	SARDLTSGANNEQF	0.7812	26847	ENPVVHFFKNIVTP,SLGNIHFFK	Myelin basic protein,polysacch	Homo sapiens (human),Sulfurovum sp. NBC
AISEVGVGQPQH	SARDLTSGANNEQF	0.7812	181941	ENPVVHFFKNIVTPR,VHFFKNI	Myelin basic protein,MBP,Myel	Homo sapiens (human),Homo sapiens (hum
AISEVGVGQPQH	AWSETGLGTGELF	0.8111	227	EAAGIGILTV,ELAGIGILTV,AAGI	Melanoma antigen recognized	Homo sapiens (human),Homo sapiens (hum
AISEVGVGQPQH	ASSWDRAGNTLY	0.7675	21130	AHHPIWARMDA,AMKRHGLDNY	Chain A, Hen Egg White Lysoz	Gallus gallus (chicken)
AISEVGVGQPQH	ASSWDRAGNTLY	0.7675	237	AMKRHGLDNYRGYSLGN,GAM	Chain A, Hen Egg White Lysoz	Gallus gallus (chicken)
AISEVGVGQPQH	ASSARSGELF	0.8034	25617	HPVGEADYFEY	EBNA-1	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	ASSARSGELF	0.8034	251	HPVGEADYFEY,HPVAEADYFE	EBNA-1,EBNA-1,Epstein-Barr	Human herpesvirus 4 (Epstein Barr virus),Hu
AISEVGVGQPQH	SARDGTGNGYT	0.7825	176648	GLCTLVAML	Transcriptional regulator IE63 h	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	SARDGTGNGYT	0.7825	190444	GLCTLVAML	Transcriptional regulator IE63 h	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	SARGGSYNSPLH	0.782	1953	FSWGAEGQRPGFG	MBP protein	Homo sapiens (human)
AISEVGVGQPQH	ATSALGDTQY	0.8338	1919	ENPVVHFFKNIVTPR	Myelin basic protein	Homo sapiens (human)
AISEVGVGQPQH	ASSLNNANSDYT	0.7588	1251	ADLIAYLKQATK	Cytochrome c	Manduca sexta (Carolina sphinx)
AISEVGVGQPQH	ASSLNWSQDTQY	0.7699	1932	ANERADLIAYLKQATK	Cytochrome c	Manduca sexta (Carolina sphinx)
AISEVGVGQPQH	ASGQGNFDIQY	0.789	323	FLRGRAYGL	nuclear antigen EBNA-3	Human herpesvirus 4 (Epstein Barr virus)
AISEVGVGQPQH	ASSYVSQNNEQF	0.7935	324	CLGGLLTMV,ALGGLLTMV,CLAC	Latent membrane protein 2	Human herpesvirus 4 (Epstein Barr virus)

TCRMatch

TCRMatch output

- Original CDR3β - query
- Matching CDR3β - IEDB
- Matching epitope
- Receptor group
- TCRMatch score
- Antigen
- Organism
- TRUST4



nature methods

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature methods › brief communications › article

Brief Communication | Published: 13 May 2021

TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data

"..for reconstructing the clonal architecture of TCR repertoires from high-throughput sequencing data"

We introduce the TRUST4 open-source algorithm for reconstruction of immune receptor repertoires in αβ/γδ T cells and B cells from RNA-sequencing (RNA-seq) data. Compared with competing methods, TRUST4 supports both FASTQ and BAM format and is faster and more sensitive in assembling longer—even full-length—receptor repertoires. TRUST4 can also call repertoire sequences from single-cell RNA-seq (scRNA-seq) data without V(D)J enrichment, and is compatible with both SMART-seq and 5′ 10x Genomics platforms.

TCRMatch

```
./run-trust4 -f human_IMGT+C.fa --ref human_IMGT+C.fa -1 example/example_1.fq -2 example/example_2.fq -o TRU
```

TCRMatch output

- Original CDR3β - query
- Matching CDR3β - IEDB
- Matching epitope
- Receptor group
- TCRMatch score
- Antigen
- Organism
- TRUST4

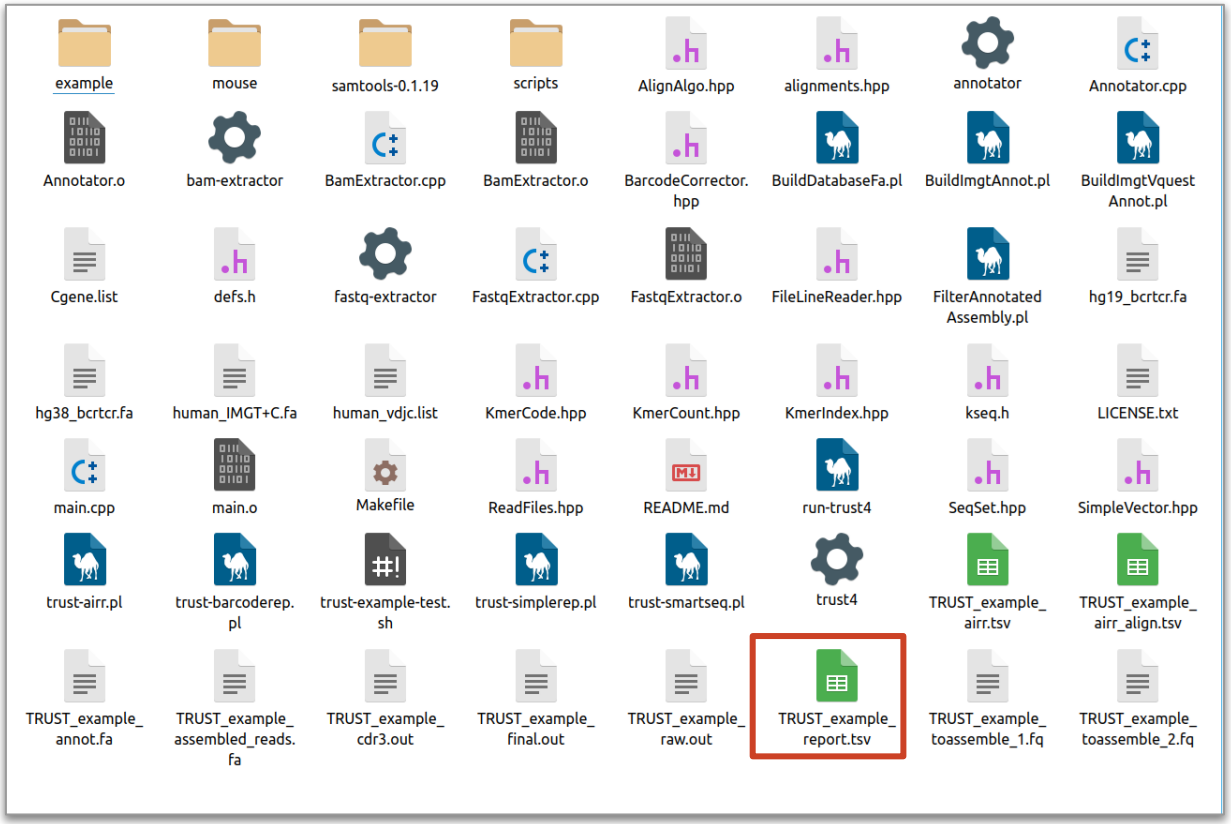
```
epoch ~/prog/TRUST4 master ?27                                                                    ✘ INT
→ ./run-trust4 -f hg38_bcrtcr.fa --ref human_IMGT+C.fa -1 example/example_1.fq -2 example/example_2.fq -o TRUST_example
```

```
→ ls TRUST_example_*
TRUST_example_airr_align.tsv
TRUST_example_airr.tsv
TRUST_example_annot.fa
TRUST_example_assembled_reads.fa
TRUST_example_cdr3.out
TRUST_example_final.out
TRUST_example_raw.out
TRUST_example_report.tsv
TRUST_example_toassemble_1.fq
TRUST_example_toassemble_2.fq
```

TCRMatch

TCRMatch output

- Original CDR3β - query
- Matching CDR3β - IEDB
- Matching epitope
- Receptor group
- TCRMatch score
- Antigen
- Organism
- TRUST4



TCRMatch

TCRMatch output

- Original CDR3β - query
- Matching CDR3β - IEDB
- Matching epitope
- Receptor group
- TCRMatch score
- Antigen
- Organism
- TRUST4

#count	frequency	CDR3nt	CDR3aa
8	0.0833333	TGTGCGAGAGGGCAGGACGGTGACTACGTGGGCGAGCCCGGGGAGTCTACTACTACTACTACATGGACGTCTGG	out_of_frame
6	0.0625	TGTGCGAGGGGGCGAACGATAATATTGTAGTAGTACCAGCTGCTATCTTCCATTTGCTTTTGATATCTGG	out_of_frame
5	0.0520833	TGTGCGAGAGATGGTACCCCGGATGTAGTAGTACCAGCTGCTATGTTTCTTAAACCGGTTGTCTGACTACTGG	out_of_frame
4	0.0416667	GGGGTATTACGATTTTTGGAGTGGTTATAAAAGCGGGGGTACTACTACTACGGTATGGACGTCTGG	GVLRFLEWL_KRGYYYYGMDVW
3	0.03125	TGTGCAAGAGATGTAGGAGGGGTATTTCCATTGTAGTGGTGGTAGCTGCTATGCTGGTTCGACCCCTGG	CARDVGGVFPL_WW_LLCWFDPW
3	0.03125	TGTGCGAGACCCTATAGCAGCAGCTGGTACGAAAGTTTACTACTACTACTACGGTATGGACGTCTGG	out_of_frame
3	0.03125	TGTGCGAAGGAGGTTGGGTGGTGACTGCTAGGGAAACTACTACTACGGTATGGACGTCTGG	out_of_frame
3	0.03125	TGTGCGAGAGACCCTCCCCGGAGCAGTGGCTGGTGGGCGTCCCCTCTTCGTTTTTGACTACTGG	out_of_frame
3	0.03125	TGTGCGAGAAAACCGACCCGGAAACGGATATTGTAGTGGTGGTAGCTGCTACTCGCGGTACCCTCTACACTTGGTTCGACCCCTGG	out_of_frame
3	0.03125	TGTGCGAAAGATCTAGGAAAGGTAGTCGCCGTATAGCAGCTCGTCCGGAGGGGATTTCTTCCTACTACTACTACGGTATGGACGTCTGG	out_of_frame
3	0.03125	TGTGCGAGCGCAGCACGGGATATTGTAGTGGTGGTAGCTGCTACGGTTCATTTATGACTACTGG	out_of_frame
3	0.03125	TGTGCGAGAAATGGGAGTTCGGATTTTGATCCAATCTCCTCGGGGGATTGTAGTAGTACCAGCTGCCTACTACTACTACTACATGGACGTCTGG	out_of_frame
3	0.03125	TGTGCGAGACAAGTGGGAGCGACCCTCATGATGCTTTTGATATCTGG	out_of_frame
2	0.0208333	TGTGCGAGAGATAGGGAGTTGTAGTAGTACCAGCTGCTGTCACGGGACTACTACTACTACTACTACATGGACGTCTGG	CARDREL__YQLLSRDVYYYYYYMDVW
2	0.0208333	TGTTATTGTAGTAGTACCAGCTGCTATTACTACTACTACGGTATGGACGTCTGG	CYCSSTSCYYYYYGMDVW
2	0.0208333	TGTGCGAGAGGCGATACCCGGTATTACGATTTTTGGAGTGGTTATTATCCGCAAAGAGGACCAGGGCAGTCCTACTACTACTACGGTATGGACGTCTGG	CARGDTRYYDFWSGYYPQRGPGQSYYYYGMDVW
2	0.0208333	TGTGGGGTTCGGGGAGTTAACATCGATAAGAAGTAATACTACTACTACTACGGTATGGACGTCTGG	CGVRGVNIDKK_YYYYYGMDVW
2	0.0208333	TGTGCAAGAGACCTAAGACATACCAGCTGCTATACTTTGACTACTGG	out_of_frame
2	0.0208333	TGTGCGAGAGAGGGATAGCAGTGGCTTGGGTACTACTACTACTACATGGACGTCTGG	CAREG_QWLGYYYYYMDVW
2	0.0208333	TGTGCGAGAGATATTGTTGTAGTGGTGGTAGCTGCTACTCCCAGGGCCCAAATTTACTGG	CARDIVVVVVAATPRAQIYW
2	0.0208333	TGTGCGAAAGACGTTGCAAGCACCAGGGCACTGCCGAAACCCCCAGTTGTAGTAGTACCAGCTGCTATCCCCTTTGACTACTGG	CAKDVASTRALPKPPVVVVPAAIPFDYW
2	0.0208333	TGTGCGAGAGTGCAAGTGACCTAATAGCAGTGGCTGGTACGATTCCATACTTTGACTACTGG	out_of_frame
2	0.0208333	TGTAGTGGTGGTAGCTGCTATAGTAAGTACTACTTTGACGACTGG	CSGGSCYSKYYFDDW
2	0.0208333	TGTGCGGCTTTACTATGGTTCAGCCCGGGGTGCTTTTGATATCTGG	out_of_frame
2	0.0208333	TGTGCGAGCAACGCTGGATTGTACTGGTGGTGTATGCTAAACAATGGATTTAGATCGAGGGGAAGAAGTTACTACTACTACTACATGGACGTCTGG	CASNAGLYWWCMLNNGFRSRGRSYYYYYMDVW
2	0.0208333	TGTGCACGGCACCCGGGTCTGTGCCTAGTATTACTATGGTTCAGGGATACTACTACTACATGGACGTCTGG	out_of_frame
2	0.0208333	TGTGTGTCCCTCAGAATTACTATGGTTCAGATTCGAACTCGAAGAACTACTACTACTACTACATGGACGTCTGG	out_of_frame
2	0.0208333	TGTGCGATGGGGTGGGTCGGTCCTCGGGCAGTACAACTGGTTCGACCCCTGG	out_of_frame
2	0.0208333	TGTGCGAGGGTACCCCGCGATGATAGTAGTGGTTAAACAACCCTTTGACTACTGG	out_of_frame
1	0.0104167	TGTGCGAGAGACCTCGGCCGTAGGTATAGCAGTGGCTGGTTCTGTTTCGACCCCTGG	CARDLGRRYSSGWFCFDPW
1	0.0104167	TGTGCGAGAGCTGTAACTTATATTGTAGTAGTACCAGCTGCGGGCTGACAGTTTACTACTACTACTACGGTATGGACGTCTGG	out_of_frame

TCRMatch

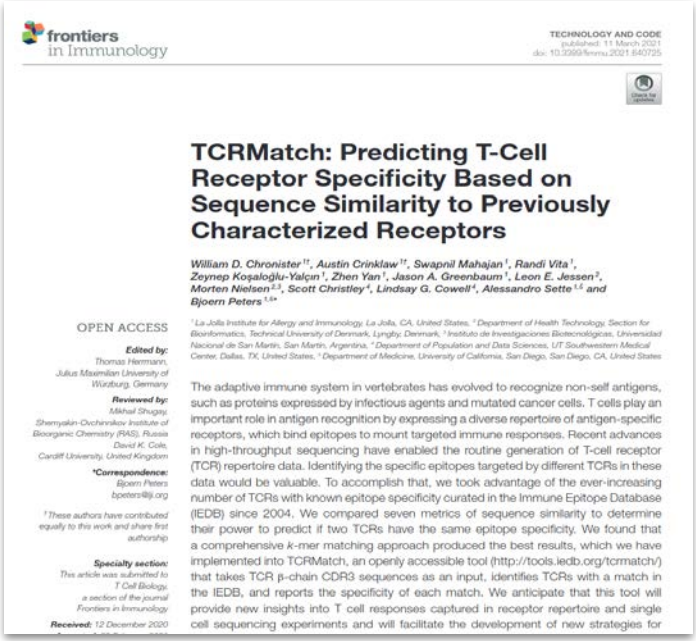
TCRMatch output

- Original CDR3β - query
- Matching CDR3β - IEDB
- Matching epitope
- Receptor group
- TCRMatch score
- Antigen
- Organism
- TRUST4

#count	frequency	CDR3aa	V	D	J	C	cid	cid_full_length
8	0.0833333	out_of_frame	IGHV1-3*01	IGHD1-26*0	IGHJ6*0	.	assemble1	0
6	0.0625	out_of_frame	IGHV4-61*01	IGHD2-2*01	IGHJ3*0	.	assemble2	0
5	0.0520833	out_of_frame	IGHV4-61*01	IGHD2-2*01	IGHJ4*0	IGHM	assemble2	0
4	0.0416667	GVLRFLEWL_KRGYYYYGMDVW	IGHV3-48*04	IGHD3-3*01	IGHJ6*0	IGHM	assemble1	0
3	0.03125	CARDVGGVFPL_WW_LLCWFDPW	IGHV3-74*01	IGHD2-15*0	IGHJ5*0	IGHM	assemble2	0
3	0.03125	out_of_frame	IGHV1-3*01	IGHD6-13*0	IGHJ6*0	IGHM	assemble3	0
3	0.03125	out_of_frame	IGHV3-NL1*0	IGHD2-21*0	IGHJ6*0	.	assemble3	0
3	0.03125	out_of_frame	IGHV3-11*01	IGHD6-19*0	IGHJ4*0	IGHM	assemble5	0
3	0.03125	out_of_frame	IGHV3-69-1*	IGHD2-15*0	IGHJ5*0	.	assemble1	0
3	0.03125	out_of_frame	IGHV3-NL1*0	IGHD6-6*01	IGHJ6*0	.	assemble3	0
3	0.03125	out_of_frame	IGHV3-11*04	IGHD2-15*0	IGHJ4*0	IGHM	assemble2	0
3	0.03125	out_of_frame	IGHV1-18*01	IGHD2-15*0	IGHJ6*0	IGHM	assemble1	0
3	0.03125	out_of_frame	IGHV5-51*01	IGHD1-26*0	IGHJ3*0	IGHM	assemble1	0
2	0.0208333	CARDREL__YQLLSRDYYYYYYMDVW	IGHV3-11*04	IGHD2-2*01	IGHJ6*0	IGHM	assemble3	0
2	0.0208333	CYCSSTSCYYYYYGMDVW	IGHV3-13*01	IGHD2-2*01	IGHJ6*0	.	assemble1	0
2	0.0208333	CARGDTRYYDFWSGYYPQRGPGQSYYYYGMDVW	IGHV3-11*04	IGHD3-3*01	IGHJ6*0	IGHM	assemble4	0
2	0.0208333	CGVRGVNIDKK_YYYYYGMDVW	IGHV3-49*02	IGHD3-10*0	IGHJ6*0	IGHM	assemble6	0
2	0.0208333	out_of_frame	IGHV3-47*01	IGHD2-2*02	IGHJ4*0	IGHM	assemble1	0
2	0.0208333	CAREG_QWLGYYYYYMDVW	IGHV5-51*01	IGHD6-19*0	IGHJ6*0	.	assemble4	0
2	0.0208333	CARDIVVVVVAATPRAQIYW	IGHV3-66*01	IGHD2-15*0	IGHJ4*0	IGHM	assemble3	0
2	0.0208333	CAKDVASTRALPKPPVVVVPAAIPFDYW	IGHV3-23*01	IGHD2-2*01	IGHJ4*0	IGHM	assemble3	0
2	0.0208333	out_of_frame	IGHV1-46*01	IGHD6-19*0	IGHJ4*0	IGHM	assemble6	0
2	0.0208333	CSGGSCYSKYYFDDW	IGHV3-23*01	IGHD2-15*0	IGHJ4*0	IGHM	assemble4	0
2	0.0208333	out_of_frame	IGHV3-41*02	IGHD3-10*0	IGHJ3*0	IGHM	assemble3	0
2	0.0208333	CASNAGLYWWCMLNNGFRSRGRSYYYYYMDVW	IGHV1-69*01	IGHD2-8*02	IGHJ6*0	IGHM	assemble9	0
2	0.0208333	out_of_frame	IGHV2-70*01	IGHD3-10*0	IGHJ6*0	IGHM	assemble4	0
2	0.0208333	out_of_frame	IGHV1-NL1*0	IGHD3-10*0	IGHJ6*0	IGHM	assemble2	0
2	0.0208333	out_of_frame	IGHV1-3*01	IGHD1-1*01	IGHJ5*0	.	assemble3	0
2	0.0208333	out_of_frame	IGHV4-38-2*	IGHD3-22*0	IGHJ4*0	IGHM	assemble2	0
1	0.0104167	CARDLGRRYSSGWFCFDPW	IGHV3-71*01	IGHD6-19*0	IGHJ5*0	.	assemble1	0
1	0.0104167	out_of_frame	IGHV1-3*01	IGHD2-2*01	IGHJ6*0	IGHM	assemble5	0
1	0.0104167	out_of_frame	IGHV1-46*01	IGHD6-19*0	IGHJ4*0	.	assemble7	0
1	0.0104167	out_of_frame	IGHV3-52*01	IGHD4-23*0	IGHJ6*0	IGHM	assemble6	0
1	0.0104167	out_of_frame	IGHV4-61*01	IGHD3-22*0	IGHJ5*0	.	assemble7	0
1	0.0104167	CARDGSGVLRFLEWFFRPYYYYYMDVW	IGHV3-11*04	IGHD3-3*01	IGHJ6*0	IGHM	assemble6	0
1	0.0104167	out_of_frame	IGHV3-NL1*0	IGHD5-12*0	IGHJ4*0	.	assemble5	0
1	0.0104167	CARDEITIF_LVII_PSWF_VP_NPPNLRNYW	IGHV1-3*01	IGHD3-9*01	IGHJ4*0	.	assemble5	0

Acknowledgments

Bjoern Peters
Alessandro Sette
Nina Blazeska
Jason Greenbaum
Austin Crinklaw
Swapnil Mahajan
Randi Vita
Zeynep Kosaloglu-Yalcin
Will Chronister
Zhen Yan
Leon Jessen (DTU)
Morten Nielsen (DTU)
Scott Christley (UTSW)
Lindsay Cowell (UTSW)



frontiers in Immunology

TECHNOLOGY AND CODE
published: 11 March 2021
doi: 10.3389/fimmu.2021.640725

TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors

William D. Chronister[†], Austin Crinklaw[†], Swapnil Mahajan[1], Randi Vita[1], Zeynep Koşaloğlu-Yalçın[1], Zhen Yan[1], Jason A. Greenbaum[1], Leon E. Jessen[2], Morten Nielsen[2,3], Scott Christley[4], Lindsay G. Cowell[4], Alessandro Sette[1,5] and Bjoern Peters[1,5]

OPEN ACCESS

The adaptive immune system in vertebrates has evolved to recognize non-self antigens, such as proteins expressed by infectious agents and mutated cancer cells. T cells play an important role in antigen recognition by expressing a diverse repertoire of antigen-specific receptors, which bind epitopes to mount targeted immune responses. Recent advances in high-throughput sequencing have enabled the routine generation of T-cell receptor (TCR) repertoire data. Identifying the specific epitopes targeted by different TCRs in these data would be valuable. To accomplish that, we took advantage of the ever-increasing number of TCRs with known epitope specificity curated in the Immune Epitope Database (IEDB) since 2004. We compared seven metrics of sequence similarity to determine their power to predict if two TCRs have the same epitope specificity. We found that a comprehensive *k*-mer matching approach produced the best results, which we have implemented into TCRMatch, an openly accessible tool (http://tools.iedb.org/tcrmatch/) that takes TCR β-chain CDR3 sequences as an input, identifies TCRs with a match in the IEDB, and reports the specificity of each match. We anticipate that this tool will provide new insights into T cell responses captured in receptor repertoire and single cell sequencing experiments and will facilitate the development of new strategies for

Chronister 2021, Front. Immunol. DOI: 10.3389/fimmu.2021.640725

…there's one more thing we'd like to show you…

IEDB
Immune Epitope Database & Tools

BCRMatch

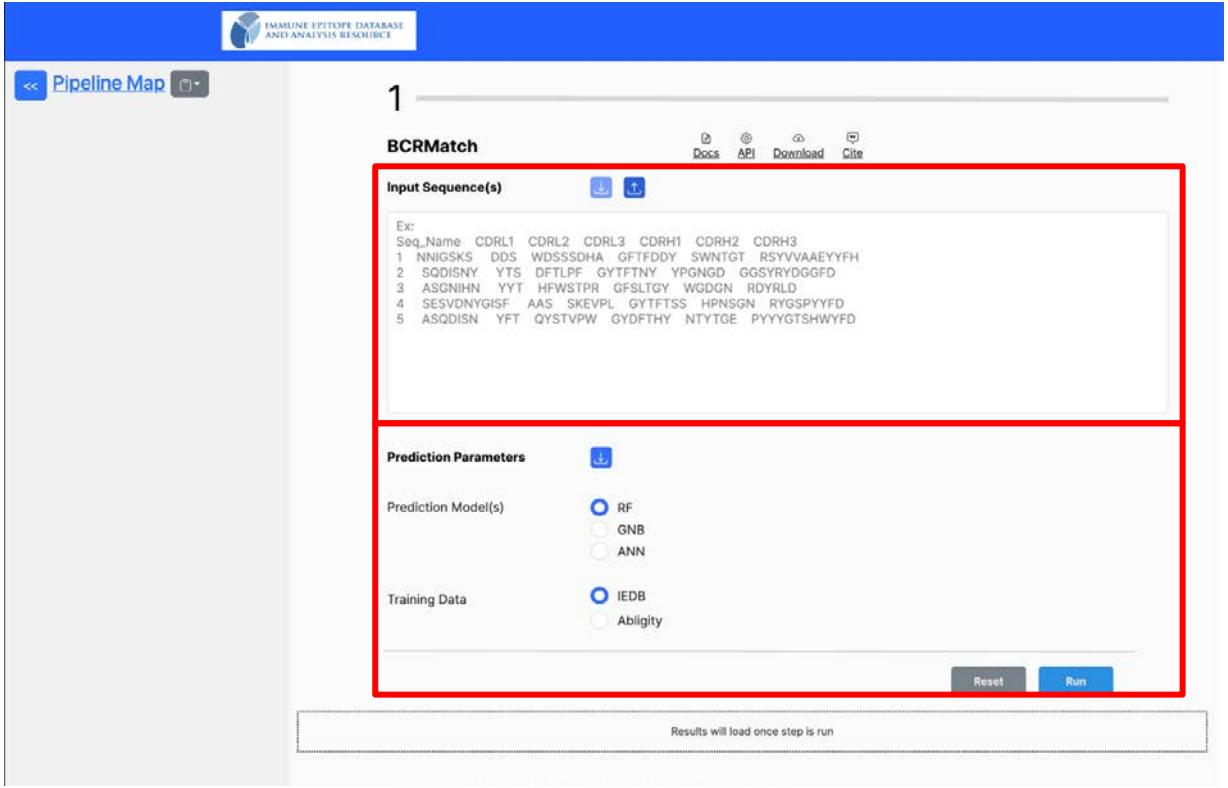
A webserver to predict antibodies binding to a common epitope

Presented by
Mahita Jarjapu

BCRMatch

COMING SOON!

A webserver to predict antibodies binding to a common epitope



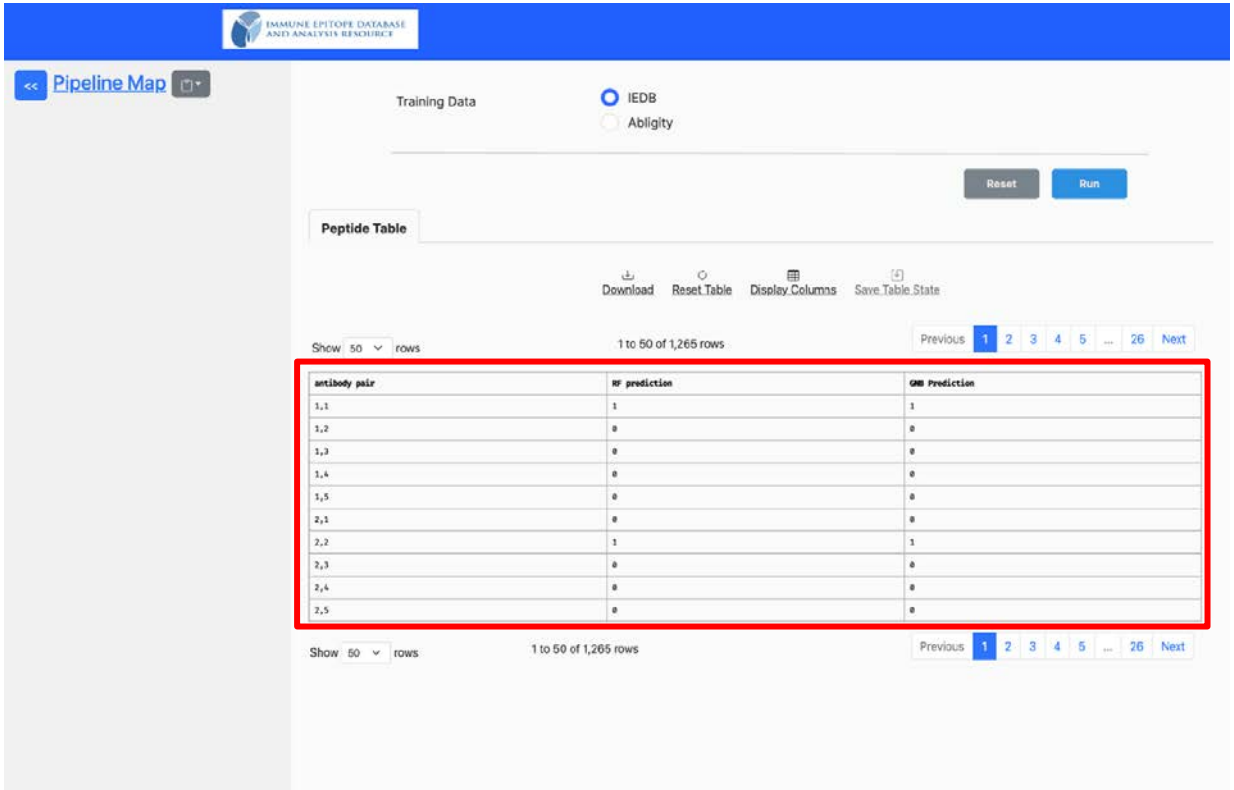
i. User provides sequences of CDR loops of antibodies with and without known epitope (referred to as antibody of interest)

ii. User has the option to select machine learning models (five ML models are provided) and training dataset for making predictions.

BCRMatch

COMING SOON!

A webserver to predict antibodies binding to a common epitope



iii. Webserver uses the selected pre-trained models to predict what antibodies share the same epitope with the antibody of interest

iv. Results of each prediction is listed in a table. We recommend using a consensus of the predictions of all five machine learning models to guide the final result.