

LAB BOOK 1

M101

ZIYI CUI

LAB BOOK 1

M101

ZIYI CUI

→ 03/10/2016

Access to Abysis:

aeron

publications

Basic Unix grammar:

1.1 A single directory tree:

root: '/'

current directory: '.'

the directory above: '..'

full path of directory separated by /

e.g.: /home/bsm2/username

1.3 - To run a program in the background so as to continue to issue commands in the shell,  
put an & after the command name.

firefox &.

- Commands often take options with '-'

e.g. -l.

e.g.: --long-format : long name, can <sup>be</sup> introduced with two '-'

1.4. help:

manual pages: ↗

read the manual pages for command: 'man command'

list all manual pages containing keyword: 'man -k keyword'

1.6 Input & output

By default:

input files read from 'standard input': stdin  
~~output~~ output write to 'standard output': stdout  
(the keyboard)

['cat' : read from stdin and write into stdout. (the screen)]

...  
'CTRL-d' : the end. of the file

'cat /etc/fstab' : read file in '/etc/fstab' instead of stdin  
↳ a way to display this file on screen

- redirect stdin by using '<'  
e.g. cat < /etc/fstab  
redirect /etc/fstab as the stdin
- redirect a stdout by using '>'  
e.g. cat > helloworld.txt.  
redirect to a file rather than the screen
- take the output of a program, rather than ~~send~~ write to the screen  
send it directly into another program using 'pipe': '|'  
e.g. cat /etc/fstab | less

## 2. Simple Commands:

- 'cat filename' : types a file to the screen
- 'cd dir' : change directory
- 'cd' : take back to home directory.
- 'cp file1 file2' : 'CoPy' : a backup of file
- 'emacs' : a configurable editor'
- 'gedit' : a simple graphical text editor. ↗ in Linux
- 'grep' : search for a str
- 'grep string filename' : such for a string in a file.
- 'less filename' : types a file to the screen,  
paused at the end of each screen full of text.  
move to the next page : "space-bar" press  
move back : pres 'b'  
quit : 'q'
- 'lpr filename' : Line PRintter : prints a file to the printer.  
if filename not specific. read from stdin.
- 'ls' : list the files in the current directory.
- 'ls -l' : gives a list of file with sizes and dates

"mkfile"  
"mv file"  
"pwd"  
"rm file"

"sort"  
"stt .."

"ssh m"  
3.  
+ chmod

4. When +

'PATH'

"echo"  
△

"`mkdir dir`" : make directory.

"`mv file1 file2`" : moves or rename file.

"`pwd`" : print working directory — shows the current directory

"`rm filename`" : remove a file.  
delete

"`sort ..`" : can add various options:  
e.g.: `sort xxx -n`

do a numeric sort rather than an alphabetical sort.

"`ssh machine`" : allow to log into another machine

3.

"`chmod options filename`" : change mode;

modify the mode or permission on the file

4.

When typing a command:

shell : the program running all the time

shell exploit "environmental variable" to known where to look for commands.  
↓  
limit the directories.

"`PATH`" : An environmental variable used by shell to store a set of ~~directory~~ directories, where it looks for commands.

typically like: `"/usr/bin:/bin:/usr/local/bin:$HOME/bin"`

↓  
used to separate different directories

"`echo PATH`" : to find what your PATH is set to.



① If it includes \$HOME bin,

- simply create that directory:

`mkdir $HOME/bin`

and move the executable programs into that directory  
"mv"

② If not → type:

`/home/myhome/mydir/myprog myfile.c`

B - Create the \$HOME/bin directory and move the files there and then modify the path so that it have \$HOME/bin in it,

also it can be done by typing:

' export = PATH = "PATH:\$HOME/bin"

↳ take the current value of PATH environment and appends :\$HOME/bin onto the end.

. Shell has hidden file called

"B if not: either '.profile' or '.bashrc'

cd  
echo export  
PATH=\$PATH:  
... ls -a

↑ to get it  
cd  
ls -a

>.bashrc  
(a-s.bashrc, A) If it does exist then run : type:

profile  
cd  
echo 'export PATH="\$PATH:\$HOME/bin"' > .bashrc

5. Emacs

'emacs filename'  
→ the file to be edited by emacs.

'Control Key' sequence: simultaneously press 'Ctrl' and 'S'

↓

'Ctrl-S'

'Meta' key: first press 'Esc' then press a letter.

A) M-f: move forward  
"M-b" by word

then release  
Meta-X

Previous C-p

Backward C-b  
Current cursor C-f

Next line C-n

Examples:

'CTRL-X' 'CTRL-C': Exit emacs

'CTRL-X' 'CTRL-S': save the contents of the buffer

'CTRL-V': move down the page

'M-a' to beginning of line  
'M-e' to end of line  
Whole text

C-a: to beginning of line  
C-e: to end of line  
M-a: to beginning of sentence  
M-e: to end moving the text around the centre of

'CTRL-Meta-V': move up a page

'CTRL-L': Clear screen and redisplay all the text

'CTRL-s' : search for a string.

'CTRL-f' : search backwards for a string.

'CTRL-d' : delete character to the right of the cursor.

'CTRL-k' : kill/remove the line from under the cursor to the end of the line on the right.

'CTRL-y' : Paste back a set of deleted text.

{ 'CTRL-@' set a mark at one end of block and move to where you want to end the block.

'CTRL-w' to wipe the the marked block.

'CTRL-h t' : interactive tutorial

"~~Ctrl-b~~" : move back across the "Newline" character forward  
if the ~~c-f~~ "CTRL-U" : C-u & C-f : move ~~8~~ characters

7. ~~back up~~ <sup>forward</sup> numeric argument.  
1st level: - highest

All users live in: /home/bsm or /home/blm 2.

e.g. /home/blm/martin/

2nd level: space under:

/acrm/home

3rd level: /acrm/data

No additional backup.

Data can be replaced easily, typically mirrored for Internet

& Server.

Main server: acrm5

Web server: acrmwww

of line  
file  
of sentence  
in the  
around  
the centre of screen

Git:

→ 5/10/20

add a

1. add new file,  
in git bash : type :  
"git add filename."

2. change something.

git commit -a -m "A brief description of the change"  
git push

[list]

△ 1. url of the frequency. ✓

△ 2. git . master:

e.g. wget -O H23\_k.txt:

http://www.bioinf.org.uk/abysis2.7/ws/refreq.cgi?quiet=1

&residue=H23&&aa=k

\* quiet=1. :

URL:  
1. #  
2. ? : connecting.  
3 &  
4 -.

single Heavy }  
pick up all  
the single,  
single light  
[artificial]

10/10/2011  
1. P -  
P -  
? h

2. warning

3 choice

→ 15/10/2016 use dictionary to add sequences,  
variable = value

add a variable to parameter: →  
residue + variable + "



residue = value

[ list ]

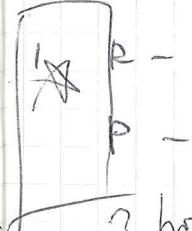
→ 19/10/2016 Integrate data

1. Read.

2. check the time of appearance

{ ① just one time → standout  
② appear in both select [ R. data ]

10/10/2016 Integrate data.



P - only Heavy.

→ don't match?

P - Heavy + light

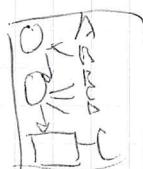
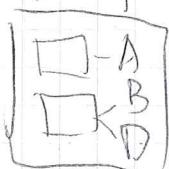
? how different? ; (R) reject

2. warning message.

L if different from the two sources

3

choose sequence from the same source,



~~total~~ (R) —

dict: { }.

248+3

R. /

P.

131

only with sequence.

$$102 + 16 - 1 = 117$$

$$\begin{array}{r} 51 \\ \times 3 - 10 \\ \hline 153 - 10 \\ \hline 142 \end{array}$$

$$51 \times 3 - 10 = 142.$$

$$153 - 11$$

$$116 \rightarrow 116 + 2 - 1 = 117.$$

$$142 \rightarrow 142 + 2 - 1 = 143.$$

$$248 \rightarrow 248 + 4 - 1 = 251.$$

$$106 \in [143:251] :$$

$$[0-49]$$

$$50+49$$

task: n

B' other

Heavy

light

\* Normally,

two chains for one antibody will not appear separately.

[+1 -1 +1.1]

~~11/10/2016~~

① A

② B

③ C

④ D

⑤ E

⑥ F

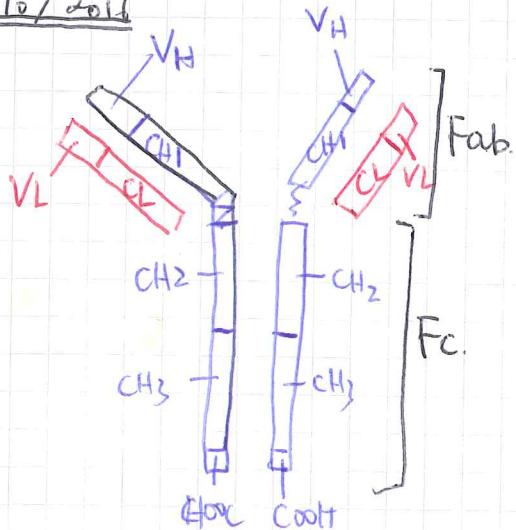
13/10/16

VL X

Sort an

13/10/2011

35 + 30



Light chain  
heavy chain

$$F_v = V_L + V_H$$

$$\frac{25-1}{143} \\ \frac{142}{108}$$

106

450

task: need whole  $\cdot$   $V_H / V_L$  : 115 aa.

B' Orlertuzumab:  $F_v + \text{CH}_2 + \text{CH}_3$   
 ~~$+ \text{CH}_2 + \text{CH}_3$~~

$$450 + 3 = 483$$

Heavy =  $\begin{cases} 550 \text{ aa} \\ 450 \text{ aa.} \end{cases}$

light : 211 ~ 217.

$V_H + V_L + \text{CH}_2 + \text{CH}_3$   
split.

115.

$V_H + \text{Linker} + V_L$   
+ KAPPA + hinge +  
 $\text{CH}_2 + \text{CH}_3$

$V_H: 1-116 \Rightarrow [0:115]$   
 $V_L: 142-248 \Rightarrow [141-247]$   
Tol 6  
hinge +  $\text{CH}_3: 249-483$   
+ 2 "n" [0:117]

[143:251]

106

Sort and format the sequence in the database

- ① A|H: ; A|L: ; ⑥ F|H; F|L|F
- ② B|H ; B|H<sub>2</sub> ; B|L: ; ⑦ G|H; G|H<sub>2</sub>|F; G|L
- ③ C|H ; -
- ④ D|L ; ⑧ F
- ⑤ E|H ; E|H<sub>2</sub> ; E|L ; E|L<sub>2</sub>: ;
- ⑥ F|H ; F|L|F ⑦ G|H; G|H<sub>2</sub>|F

→ 14/10

final

ignore: Placitumab | Heavy.  
~~Othersubmabs | Heavy~~  
 change to { H  
 L.  
 Cetuximab amungtenkin | Heavy 2 | Fusion.

with Fusion: only

Cer - - - amu - - - | H2F.

Citatumab bogatox | H | LF |

- score

check

only

- check

ignore:

{ Placitumab | Heavy  
 } Cetuximab amungtenkin | Heavy 2 | Fusion

Othersubmabs | Heavy => change to { Heavy  
 Light.

→ 17/10

→ Task

1. check

2. check

[na]

fixed

get data in:

ls/acm/data

3. only

→ RL

No

(1) ✓ rec

\*

RL all  
instead

PL an

→ 14/10/2016 converted <sup>all</sup> sequences in images into text.  
find all image in RL and PL115.

- Search "mabum".

check all RL:  
1-75.

only RL75: 1 image

check PL.

PL115: 12 images

- check the "no-sequence".

→ 17/10/2016

Tasks after discussion

1. check PL76 for imaged sequence

2. check the source of vatezumab

(vareciximab in PL114 ~~in~~)

in RL76 also

corrected: vareciximab → PL105 &  
fixed.

3. only add PL115 seqs.

txt: all txt or pdf  
faa files

→ RL check:

No sequence. use grep.

1) record <sup>these</sup> actually "no sequence" at all.  
\* antibodies with

add.

RL all + 115 PL.faa.

Instead of  
PL and RL

INN.faa.

2) ✓

Check what  
the differences  
is b/w PL.faa.  
& RL.faa.

③ \*

write program to integrate data in RL.faa and PL.faa.  
(take an RL.files + PL.files)

whole p

up to

+

replace the mammal added

git: warn

overwrite replace

leave it

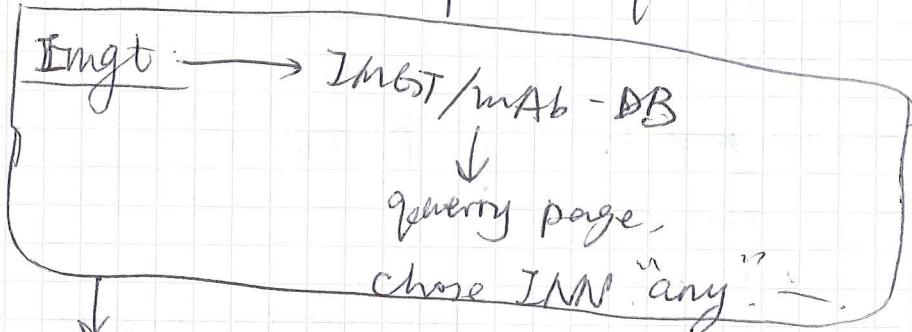
no sequence added in mammal list

if ~~this~~: "mammal"

{ > mab  
> mab

if not: just keep the sequence.

> mab



Can use ~~it~~ to get abs. Instead of convert .png → .txt

> mab/H

{ . cit  
. cit

{ . cergut  
. cergut  
. cergut

[navicixizumab : R7b & P114]

One exception added [in 1986]

whole pdf imaged:

up to R2 45  
PL 85.

git: warning : LF will be replaced by CRLF.  
↓                    ↓  
linux                ↪ windows  
→ gitosis

leave it as what it is. — no split.

{ > mab | Heavy.

{ > mab | Light

> mab | Heavy { · placulumab  
· lexatumumab  
· otlertuzumab

→ wrongly labeled → change to SFv

> mab | light { · emtuzumab  
· naptumomab estafenatox

> mab | H

{ > mab | L

{ > mab | H2

{ > mab | L2.

→ vanucizumab

> mab | H | L | fusion:

{ · citatumumab bogatox | H

{ · citatumumab bogatox | L | F

{ · cengtuzumab amanabekin | H

{ · cengtuzumab amanabekin | H2 | F

{ · cengtuzumab amanabekin | L.

5.

Meeting on 20/10/16 Thur. 12:10

chmod +x integrate.py  
in shell:

automatically in top directory ~> cbioch  
to that shell  
path when open shell  
1. bashrc  
2. export PATH="\$PATH:/acron/user/local/bin"

chmod +x integrateData.py  
./integrate.py

1. write  
allow the Not hard-coded

add the file on the command line. [specify it on command line]  
L pick up from command.

Parameters.

1. RE. 2. PL. 3. manual. optional parameter.

> \$ ./Py Param1 Param2 Param optional

2. correct the manual file. (some get heavy chain)

3. efunctnmb Heavy Light

Heavy chain coding:

o.s CAR.n  
o.g CCAR.n

[WG].

CCCCC-WG.  
5n3oan

the form  
for SCFV

| light  
\* mab/Heavy-Light - .

Light-Heavy - toxin

also put other  
things in

Print  
4. write to the screen, can specify file output

e.g. \$ ./Py > file name

find - or

Pairwise Fusion

{ HIF LIF

{ L H

Multifusion

{ H  
H2F

L

L2F

L

(Light Fusion)  
(Heavy Fusion)

6.

7.

{ Expect  
Actual

① ~~error~~  
\$>mab/

\$>mab/

② mab/

③ >mab/

④ >mab/

⑤ >mab/

⑥ >mab/

all in on

{ citation

{ C O  
C A

C A

WRC

5. better to define function,  $\rightarrow$  as small as possible  
 separate function also separate function



every function should be no longer than 1 <sup>A4</sup> page.

6. also good to write modules.

7. Write "test" before write codes.

for each function

- check if get what you want.

{ Expected -> Print "error"  
 |  
 | actualization | Print "ok\n" }  $\rightarrow$  how to specify the expected  
 \* sample test cases.

- navicixiznumab
- ↑
- ① ~~c~~ mab | H
  - { >mab | Heavy }  $\rightarrow$  mab | H<sub>2</sub>
  - { >mab | Light }  $\rightarrow$  mab | L
  - ② mab | Heavy  $\Rightarrow$  n
  - ③ >mab | Light
  - ④ >mab - no seq
  - ⑤ >mab | P | Fusion.
  - ⑥ >mab | Heavy - light - other  $\Rightarrow$  all in one

Process

dict of chain appearance  
 $d = \{mab : [H, L, F, H2, L2]\}$

sort:  
 $ad = [H, L] = [x_{mab}, y_{mab}]$

↓  
 build b2dict for chain  $\checkmark$  readiness  $\rightarrow$  set key.  
 $D = \{mab : [D2Q, \dots]\}$ , add seq.

citazumab bogatzik heavy

- - - - - - - - - - | Light/Fusion compare with PL, produce warning list  $\checkmark$

{ C a | Heavy  
 { C a | H<sub>2</sub>  
 C a | Light F

$b' = \{mab : [ ]\}, mab : [ ] \dots$   
 ↓  
 only in PL

add manually converted seq  $\checkmark$

$D'' = \{mab : [ ]\}, mab : [ ] \leftarrow, mab : [ ]$

↓  
 classify appearance & sort  
 $mab : [H, F, \dots]$

write into file / print  $\checkmark$

fusion  
 by fusion

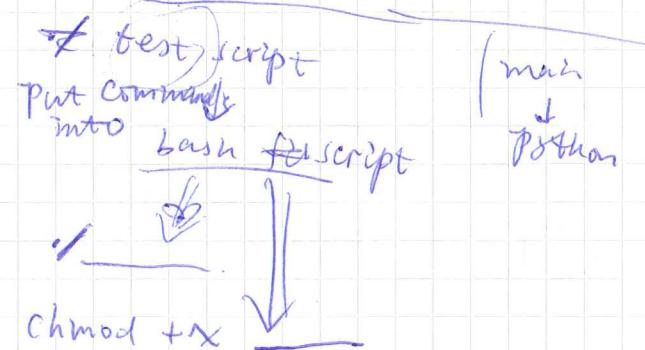
Discussion on 17.00, 20.10.

18 different cases,

creat 9 test files for each cases.  
for main program, 18PL, 18PL.

check manually [output file] → manual created  
before test: [output file created from .ps]

consider all possibilities



If an test part,  
should give no output.

|     |    |
|-----|----|
| 1.  | T  |
| 2.  | T  |
| 3.  | T  |
| 4.  | T  |
| 5.  | T  |
| 6.  | T  |
| 7.  | no |
| 8.  | no |
| 9.  | no |
| 10. | no |
| 11. | no |
| 12. | no |
| 13. | no |
| 14. | no |
| 15. | /  |
| 16. | /  |
| 17. | /  |
| 18. | /  |

If test ✓ →

If test fail → write test for each funct;

PO 24.10

build scenarios :

|     | RL    | PL                |
|-----|-------|-------------------|
| 1.  | txt   | txt \$ ✓          |
| 2.  | txt   | txt or modified ✓ |
| 3.  | txt   | no ✓              |
| 4.  | txt   | image \$          |
| 5.  | txt   | image \$          |
| 6.  | nos   | txt               |
| 7.  | no    | image             |
| 8.  | no    | no                |
| 9.  | image | txt \$            |
| 10. | image | txt \$            |
| 11. | image | image \$          |
| 12. | image | image \$          |
| 13. | image | no                |
| 14. | ...   | ...               |

✓ 13

|     | RL | PL                 | Imag.   |
|-----|----|--------------------|---------|
| 1.  | t. | ts                 | / ✓     |
| 2.  | t. | tol. last line AAA | / ✓     |
| 3.  | t. | ...                | / ✓     |
| 4.  | t. | ...                | P.S. ✓  |
| 5.  | t. | ...                | P.D.    |
| 6.  | /  | /                  | / R     |
| 7.  | /  | /                  | / P     |
| 8.  | /  | /                  | P & R S |
| 9.  | /  | /                  | P & R D |
| 10. | /  | /                  | /       |
| 11. | /  | t                  | R S     |
| 12. | /  | t                  | P D     |
| 13. | /  | t                  |         |

Mod

step ②

step ③

In  
and  
out

model seg

|     | RL  | PL       | IL | IP | IR |
|-----|-----|----------|----|----|----|
| 1.  | (t) | $t =$    |    |    |    |
| 2.  | (t) | $t \neq$ |    |    |    |
| 3.  | (t) | no       |    |    |    |
| 4.  | (t) | no       |    |    |    |
| 5.  | (t) | no       |    |    |    |
| 6.  | (t) | /        |    |    |    |
| 7.  | no  |          |    |    |    |
| 8.  | no  | ✓        |    |    |    |
| 9.  | no  | ✓        |    |    |    |
| 10. | no  |          |    |    |    |
| 11. | no  |          |    |    |    |
| 12. | no  |          |    |    |    |
| 13. | no  |          |    |    |    |
| 14. | no  |          |    |    |    |
| 15. | /   |          |    |    |    |
| 16. | /   |          |    |    |    |
| 17. | /   |          |    |    |    |
| 18. | /   |          |    |    |    |

① add RL.faa →

add PL.faa ②

③ add. ILng.faa

step ① → ~~if true~~ has  
→ ~~if no~~

②

step ②.

PL

$t = \checkmark$

$t \neq \checkmark$

no

$t \checkmark$  mark

no

no

no

no

no

no

PenningList

$P = R$

$R \neq P$

→ could to warning

In integrand  
and unitarily

mark seg

not

$R \checkmark$

$F \checkmark$

26/10/2016

run python in bash on Windo:

wingpy py -3 — .py p1 p2 ...

PL

RL

One type for chain combination

vanuicizumab  
H H  
H L  
H2  
L2

navit haricixizumab  
H H2  
L.

IR

IP

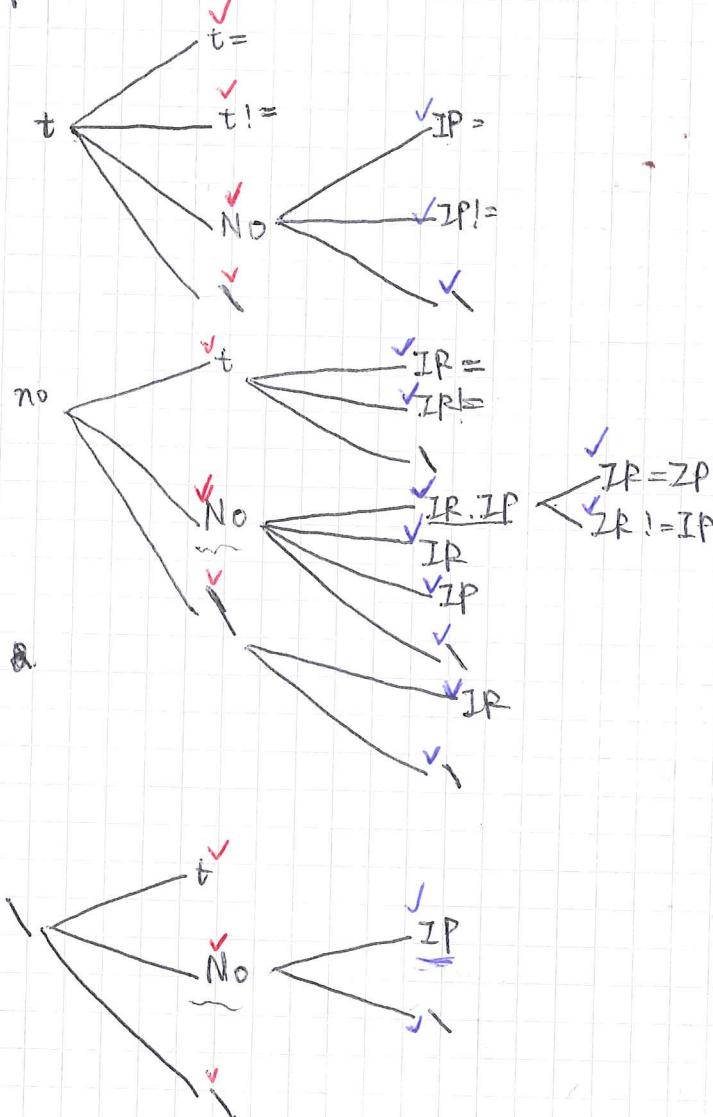
RL

+

IR

IP

All possibilities:



{ mab  
{ no s }

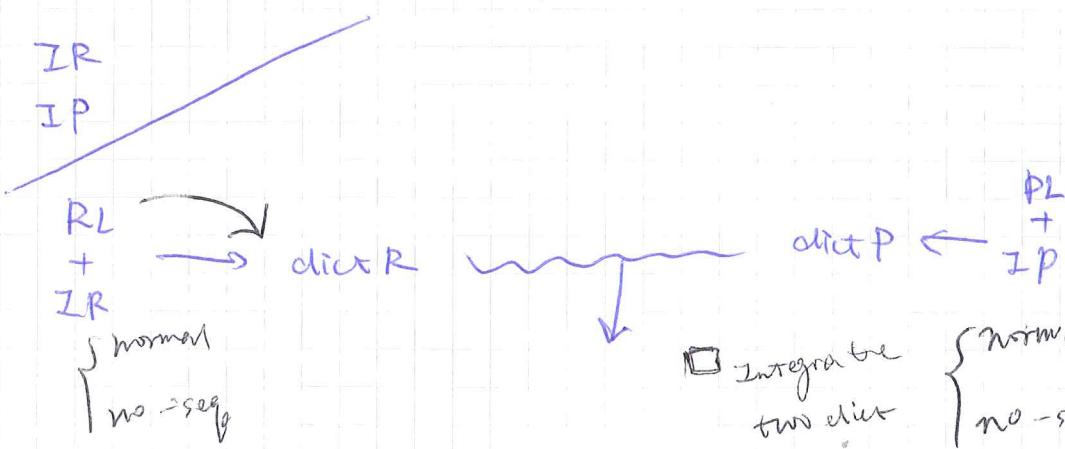
E ]

P

m

no

PL →   
 RL →   
 { "I" IntegratedInfo f y → warninglist  
 { "N" AntibodyNoSeq f y



dictP ← dictP  
 ① same ✓  
 ② diff → warning.  
 ③ If no { no seq. — pop ch seq. } add. — → list only in AP

{ mark!  
 { no seq

✓ mark! — mark! { S d. }  
 ✓ no seq. — no seq. | keep.

: no seq. — mark! : add  
 : pop no seq.

E) ✓ mark! — no seq

→ p  
 p ↓ mark! no seq

|        | mark!    | no seq |  |
|--------|----------|--------|--|
| mark!  | S   d. ✓ | ✓      |  |
| no seq | ✓        | ✓      |  |
| .      | ✓        | ✓      |  |

✓ { mab | H  
| mab | L

mab | H  
mab | L.

✓ mab | H - L - ...

mab | Heavy (1 Fusion)  
mab | Heavy (2 Fusion)  
mab | Light (1 Fusion)  
mab | Light (2 Fusion)

✓ mab - no seq

Note : Pl imagined Doctor.

> ascrinvacumab | Light.  
line 4. the last



> bevacizumab, beta  
(replaced by short in Pl. form.)

1/11/2011  
9:21:00

I don't  
know

2 Fusion

a

3. test

4. test

5. test

6. CP

► 1/11/2016 Discussion with Dr. Andrew Morton  
git: description for each file.

I don't make "global." ✓

format data =

2: Function: testfile ✓

- close after open
- not use global variable
  - use return
- def ...KC

return ...  
 $a \approx k(x, y)$        $y_1 = \dots$   
 $y_2 = \dots$   
 $y_3 = \dots$   
return ( $y_1, y_2, y_3$ )

func



put in,

Pass local variable

from func as argument

$z = f_2(y_1, y_2, y_3)$

global - store externally "trip"

"return" - store in cpu "storage"

3: test.sh ✓

4: testfile: make a expected.out ✓

5: test - chain occurrence. ✓

6: cp ~~.py.out~~ .py.works  
new name. (as a cache) ✓

From ab list:

\* search clinical trial

↳

If anything is obvious for fail.

Compare succeeded/approved.

failed

so new/not tested

Paper:

(1) Analyzing the degree of humanness of

Immuno

\* paper from email:

(2) Immunogenicity of engineered antibodies.

Method 36:3-10 |

then

- apply standard numbering.

program → Chothia

([www.bioinf.org.uk/abs/abnum/](http://www.bioinf.org.uk/abs/abnum/))

- write a python script:

"REST"

add plain=1.

rewrite the numbered sequence  
to a new file.

referee's or form

general

Next

Meeting:

10<sup>th</sup> Nov.

16:00

→ 3/11/20

SET

use

1.

{ m  
} m

2.

mar

3

marb

4.

marb

5.

marb

6.

{ marb  
marb

7

{ mar  
mar

8

{ m  
{ m  
m

9

{ m  
{ m  
m

10

{ m  
m  
m  
m

11

{ m  
{ m  
. m

12

{ mar  
mar  
mar

13

{ mar  
mar  
m  
mar

3/11/2016 testfile 2

SET2 for test

use RL PL IL  
t t! \

PL.

1. { mab / H ✓  
mab / L .

2. mab / H ✓

3 mab / L ✓

4. mab mab - no sep ✓

5. mab / Heavy - Light ✓

6. { mab / H / F ✓  
mab / L .

7. { mab / H ✓  
mab / L / F

8. { mab / H  
{ mab / H2 / F ✓  
mab / L .

9. { mab / H  
{ mab / L  
mab / L2 / F ✓

10. { mab / H  
{ mab / H2 / F ✓  
mab / L  
mab / L2 / F

11. { mab / H  
{ mab / H2 ✓  
. mab / L

12. { mab / H ✓  
mab / L  
mab / L2

13. { mab / H  
{ mab / H2  
mab / L  
. mab / L2

7/11/2016: write python script to use "abnum"

abnum

```
if line[0] = '>'  
    continue.  
if line[0] = "\n":  
    STOP.
```

9/11/2016

ssh acrm8

chmod a+x \*.py

/getabnum.py

File No such file/directory

But:

python > getabnum.py testfile -fae works

10/11/2016

list approved  
and date

↓  
name      approved?      date

Class

dist

\* storm

| Lha

For each  
false

## Clustering mechanism:

1. distance generated (up to 200)

Distance    SD inflexibility

(ori)  
generally  
↑

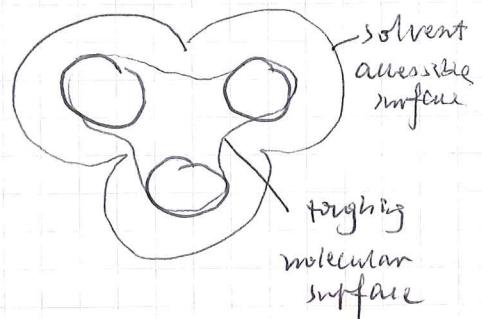
by distance file - propfile - cutoff file

v (Side distance)

or

c-alpha

accessibility  
for each  
residue.  
→ accessible  
area



\* reference \* reference to start

accessibility

(<100%)

some >100%

e.g. end of  
chain

e.g. distorted

can use one  
for all.

precalculated

\* structure info.

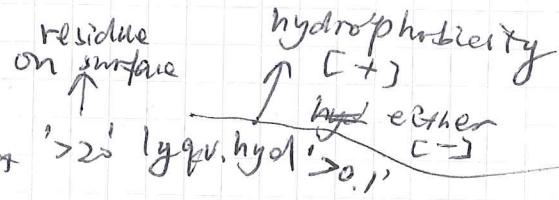
| has programs

For each obs:

take sequence → model → accessibility → distance matrix

# clusterSurfaceHProb.sh

22/11/13



—.pl. "distmat.dat" "access.dat" '>20' lygrv,hyol'>0.1'

Bash:

Perl —.pl distmat.dat

Perl —.pl distmat.dat access.dat '>20' —, num '>20'  
(frequency)

Automate the whole process,  
from getresfreq

work the cluster.pt.  
cluster.pl.

a Bash script.

—.py + — seq + frequency

sh. — s1 s2 s3 s4

then

take each

seq → PDB → accessibility → distance matrix

90% preparing data

10% analysis

Next pt/Need

9:30

ssh acm8 | get access

~~acm8~~

22/11/13

23/11/20

AM save  
write

1. check

2. JS

what  
make  
① b6

\* spec

22/11/2016

~~getresfreq~~ .seq. → frequency file.

process sequence.

not found in added list:

pexelizumab

① ↗ addnosequence.py  
↓ added noseq.txt

for those with sequence:

\* efungumab  
† first am missed

\* chain info unsure:  
radretumab

appear  
no seq  
in INN list

Absent  
(most)

iratumumab  
canakinumab

tanezumab  
denosumab

andecaliximab  
sapilizumab

22/11/2016: change no seq

→ 23/11/2016 meeting at 9:30

AM save the html file,

write script can ~~write~~ find the doc, then gnas

1. check the sequence.

3. sort git.

2. ls ~martin/scripts/clusterResidues/

~~What did in bef before 13.00~~

more specific

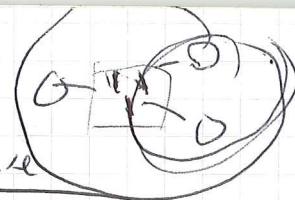
① books read.

\* spelling mistake  
fourth

~~list~~

/acm/bsmhome/martin/

Differ:



dismat.dat : c-alpha distance

↓ change      change.

1. dismat-sc.dat

(~~too~~ need shorter distance cutoff) -al.: [default = 8.5]

look at RasMol.      (3/4)

determine proper thresholds.

/acm/usr/local  
bin/Rasmol

pymol

→ 24/11/2016

Tasks:

1. double check: INN data

2. look at the RasMol to determine proper cutoff

3 testfiles/ added on 9/11/2016 'update test3.sh'

↳ restore

to get access by 'automount':  
In Bash:

ls /acm/usr..

path

In Browser:

/acm/usr

notes:

1. bina

can

For the

2. o-

3. run

① g

②

△ clust

po

S

→ 25/11/2016

1. For th

set 2

✓

✓

2. writ

notes:

1. blinatumomab / L-H-H-L

cannot apply the numbering scheme

For those: all in one, can not apply the numbering scheme:

2. ./\_\_\_\_.pl -m=\_\_\_\_ -d=\_\_\_\_ distmat.dat access.dat '520' -num 'c' frequency.

3. run rasmol

① generate psib files

② SWISS-MODEL.

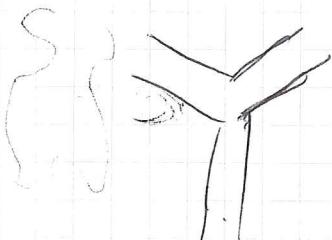
Sample used

1. elotuzumab : approved.

2. solanezumab : failed.

△ cluster:

position: settled. — different ab dif structure



how could clustering results be checked  
on other antibodies' structure?



25/11/2016

1. For those integrated chain:

set 2 ofca:

✓ 1. split them. and then run the seq

✓ 2. keep the single-chain for database source

Ctrl-D

kill the command

2. write scripts for calculating + the distance matrix

PDB file:

| Column | 1           | 2         | 3            | 4        | 5              | 6 | 7 | 8 | 9         | 10           | 11 | 12 |
|--------|-------------|-----------|--------------|----------|----------------|---|---|---|-----------|--------------|----|----|
| ATOM   | atom number | atom name | residue name | chain id | residue number | x | y | z | occupancy | element name |    |    |

In the unit of Angströms

temperature factor

$$\text{distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

distance matrix:

~~use a binary two-dimensional matrix.~~

average ; SD

→ 28/11/2016 get solvent accessibilities for antibodies.

To run pdbolv,  
require radii.dat,  
set environment variable DATAAD2R

'>10': for access.dat: only surface residues  
unusual:

Rasmol:

'OR': combine two selection

Purple : N    grey : C  
Red : O

distance freq.  
>10  
↓  
>20 +

[How to define a successful cluster]

A) Minimum cluster number: 3?

\* Set DATAAD2R environment:

export DATAAD2R = /home/lsm/martin/data

dist  
[standard] >20

use.

distmat

→ cl=4

- cl=3

- cl=2

→ 30/11/2016

erature  
for

12  
element  
name.

dist  
[standard distance . access.]

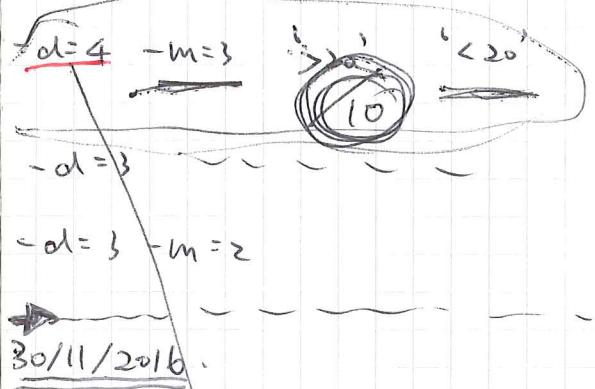
>20

freq

<20

use.

distmat-sc. access.



Basically

one Atom:  $d=1\text{ \AA}$

[no atom btw]

cluster on all atoms

[whether atom in between]  
to say it is clustered

color based on accessibility.

elotuzumab

150 mg/mab

C1:

L28, 31, 33, 96, 97, 98

E<sub>2</sub>:

L31, L32, L50, L55

C1:

L12, 14, 15, 18, 20,  
74, 77

C2:

L29, 31, 50, 91, 93

Not like cluster

C1:

L18 L20 L74

None

None

access file:

.pdb

Discussion 30/11/2016. 4pm.

Q1: how to get distance matrix from accessibility.  
git@github.com:UCL/abysis.git/abysis/src/C/distmat.

Q2: How to define a successful clustering?

- How close b/w residue in continuous patch touching each other  
No. of

✓ - minimum clustering member = 3?

✓ - unusual freq. cutoff: < 20?

2 papers, in the paper provided, 20 of 45  
→ [published]

Q3: probability of unusual score.

Use what techniques?

how to define.

lots of populations } { ① not all get result, some not  
                      } { ② cancer patient has problem in immune.

cross-references

from

a list of

approved group

[up to 5 years]

{ = low

represent.

whole set

{ haven't been approved  
= high  
= low. }

{ 1-side test  
2-side test

The version of T-test that doesn't need Normal Distribution

whether Approved drug is randomly

If normal dist is random: same mean

same variable. if not: not --

~~t-test~~ Mann-Whitney U test, similar to t-test.

doesn't require mean or SD

reflect on the immunogenicity

done with all data

— define randomly

different selected

assume: 'failed'

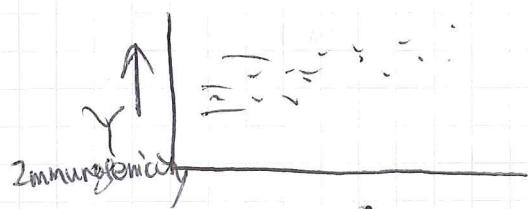
Approved, 'may have high

low immuno'

classify to several class of paper list of immunol.

tolerable

negligible



P { either exact number  
or classed }

r = 1 Pearson's correlation

conservative

multiply testing [Bonferroni]  
correction

If 4 scenarios.

Data: different ways for scoring: P multiplying p-value.

clustering scores: ① sum up all the residue,  $= \sum p_i$

Convert  $\sum$  to a unusual score,

~~J~~  $\rightarrow$  usual score

0.6  $[1 - \frac{1}{\text{unusual}}]$

$\text{or } \frac{1}{x} \text{ or } \frac{1}{n}$

give birth  
Bonferroni  
and  
non-Bonferroni;

one over

② the number of clusters and residues clustering

separate data and programs in github.

- 15/12/2016 Choose the initial criteria for 'clusterResidues' write a seq.
- created olistmat file
  - run clusters on specific examples:

| olst   | freq.                   | elotuzumab   | solanezumab   |
|--|-------------------------|--|---|
| use: CA<br>-m=3,<br>-d=8.5                         | access<br>'>10'<br><20' | None   | None  |
| use: all atoms.<br>-d=3<br>-m=3                    | >10<br><br><20          | None   | None  |
| use: side chain.                                   |                         | None   | None  |
| use: CA -m=3<br><del>freq.</del><br>-d=8.5 >10 <20 |                         | C1: too sparse.<br>C2: not concentrated.<br>C3: C1 ✓: H(33,52,53,56)<br>good cluster<br>C2 ✗: H(93,99) + L(31,50)<br>good cluster<br>C3 ✗: L(31,33,52,53,55) | C1: (52 53 55) H<br>good cluster, touching<br>touching each other |
| use: SC -m=3<br>-d=4 >10 <20                       |                         |  |   |
| task   |                         |  |   |

For each antibody: automatically get data and do clustering

\* re-format the ~~access~~ access file. ✓

\* creat pols for all sequence

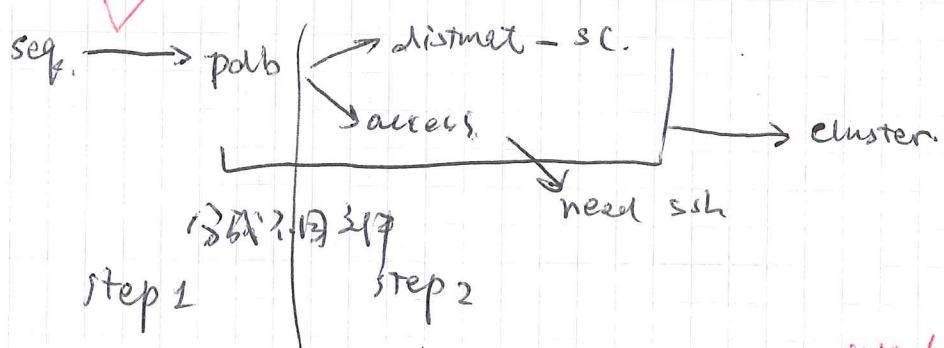
\* how to view in RasMol according to hydrophobicity

use: all . -m=3

-d=4 >10 <20

not all are close enough  
C1: X H(96,98,99,100) but far  
L(31,50,53,55,56, 94)  
C2: ✓ H(33,52,53,56,57)  
have many atoms in between  
C1: ✓ H(31,33,52,53,55)  
C2: ✗ H(50, L(91,93,94))  
C3: ✓ L(30,30A,30C)  
C4: ✓ L(31,50,51) touching  
quite close but not ~~touching~~

blue's write an automatic process to get specific files and run clustering



accessibility:

polbsolv -n -r stdout test.polb | awk '{print \$2,\$3,\$8}' > test.sc

need accessibility  
for sc.

↑  
\$2,\$3,\$8  
test.sc

if sc

↓

for whole residues

distmat:

distmat {  
-c L,H  
-p file.polb  
-a  
-s } > \_\_\_\_\_

1,53,55)  
3,94)  
c]  
] touching  
at ~~the~~

List of antibodies : [CHAMA]

Sequence  
anti-IgM mAb

1. 2H4 and 5D3
2. Anti-Lym1 mAb
3. 10-3D2.
4. <sup>131I</sup>I-labeled anti-HCC mAb/Hepoma-2 I-Hepoma-1 mAb
5. <sup>131I</sup>-T101 no name
6. 14G2a.
7. 16H5
8. 17-1A
9. A7 - NCS
10. Anti-CEA fragment conjugate
11. Anti-CEA

Approved.

X 100

X 38

X 74

X 34

X

89

50

.

.

.

.

0

Appro

✓

12

~~13~~ =anti-CEA  
~~Anti~~citumomab

✓  
BUT No seq

Find seq from Internet

14

X

15

X

16

X

17

X

18

X

19

X

20

X

21.

X

22

X

23

X

24

25. edob

26

27

28 Ibritumoma

29

30

31

32

33

34

35

36

37

38

39

40

41. <sup>131I</sup>-T

44 45  
[HACA]

1. rituxim

2. Abcixim

3 basilixit

|                             |                   |                     |   |   |
|-----------------------------|-------------------|---------------------|---|---|
| 24                          | X                 |                     |   |   |
| 25. <u>edobacombab</u>      | No seq            | possible treatment? | X | 20  |
| 26                          | X                 |                     |   |   |
| 27                          | X                 |                     |   |   |
| 28 Ibrutinib-Tiuxetam       | ② ✓               | ✓                   |   | <2.   |
| 29                          | X                 |                     |   |   |
| 30                          | X                 |                     |   |   |
| 31                          | X                 |                     |   |   |
| 32                          | X                 |                     |   |   |
| 33                          | X                 |                     |   |   |
| 34                          | X                 |                     |   |   |
| 35                          | X                 |                     |   |   |
| 36                          | X                 |                     |   |   |
| 37                          | X                 |                     |   |   |
| 38                          | X                 |                     |   |   |
| 39                          | X                 |                     |   |   |
| 40                          | X                 | ✓                   | ✓ | 8%  |
| 41. <u>I131-tositumomab</u> |                   |                     |   |   |
| 44 \ 45 \ 46 \              | [HACA]            |                     |   |   |
| 1. rituximab                | No seq            | ✓                   | ✓ | 0   |
| 2. Abciximab                |                   | ✓                   | ✓ | 4.8%; after a first readministration, additional 19% became [HACA]<br>$\approx 4.8\% + 95.2\% \times 19\% = 22.9\%$ |
| 3 basiliiximab              | <del>No seq</del> | ✓                   | ✓ | 11.8% with (+OKT3)<br>1.4% (-OKT3)  |

|                 |   |   |              |                 |
|-----------------|---|---|--------------|-----------------|
| 1. Infliximab   | <del>No Seq</del>                           | ✓ | 61%          | → Mabthera      |
| 2. alemtuzumab  | +   | ✓ | 5%           | 1. muromab      |
| 3. rituximab    | ++  | ✓ | 5%           | 2. abciximab    |
| 4. daclizumab   | ++  | ✓ | 1.9%         | 3. rituximab    |
| 5. trastuzumab  | ++  | ✓ | 1.9%         | 4. daclizumab   |
| 6. Palivizumab  | ++  | ✓ | 1.9%         | 5. trastuzumab  |
| 7. basiliximab  | ++  | ✓ | 1.9%         | 6. Palivizumab  |
| 8. infliximab   | ++  | ✓ | 1.9%         | 7. basiliximab  |
| 9. arcitumumab  | ++  | ✓ | 1.9%         | 8. infliximab   |
| 10. canakinumab | ++  | ✓ | 1.9%         | 9. arcitumumab  |
| 11. farnesol    | ++  | ✓ | 1.9%         | 10. canakinumab |
| 12. imciviro    | ++  | ✓ | 1.9%         | 11. farnesol    |
| 13. capromab    | ++  | ✓ | 1.9%         | 12. imciviro    |
| 14. rinfetumab  | ++  | ✓ | 1.9%         | 13. capromab    |
| 15. gemtuzumab  | ++  | ✓ | 1.9%         | 14. rinfetumab  |
| 16. alemtuzumab | ++  | ✓ | 1.9%         | 15. gemtuzumab  |
| 17. ibritumomab | ++  | ✓ | 1.9%         | 16. alemtuzumab |
| 18. adalimumab  | ++  | ✓ | 1.9%         | 17. ibritumomab |
| 19. natalizumab | ++  | ✓ | 1.9%         | 18. adalimumab  |
| 20. omalizumab  | ++  | ✓ | 1.9%         | 19. omalizumab  |
| 21. efalizumab  | ++  | ✓ | 1.9%         | 20. efalizumab  |
| 22. bevacizumab | With Seq H&L                                | ✓ | 6.3%         | 21. tositumomab |
| 23. cantrumab   | With Seq H&L                                | ✓ | 0            | 22. cetuximab   |
| Mertansine      | [Cantrumab has H&L]                         | ✓ | 0            | 23. bevacizumab |
| 24. panitumumab | discontinued from 2005 [phase I for cancer] | ✓ | 0            | [Repeated]      |
|                 |   |   | Not approved | 24. panitumumab |

→ Matthew's paper Table 1.

|   | seq? | approved? | Immunogenicity |
|---|------|-----------|----------------|
| 1. muromomab                                  | x    | ✗         | 6% - 12%       |
| 2. abciximab (PAI ready<br>in previous paper) | x    | ✓         | 6% - 44%       |
| 3. rituximab ✓                                | x    | ✓         | 11%            |
| 4. daclizumab ✓                               | x    | ✓         | 14% - 34%      |
| 5. trastuzumab (1)                            | x    | ✓         | <1%            |
| 6. Palivizumab ✓                              | x    | ✓         | 0.7% - 2%      |
| 7. basiliximab ✓                              | x    | ✓         | 1-2%           |
| 8. infliximab ✓                               | x    | ✓         | 10% - 15%      |
| 9. alemtuzumab ✓                              | x    | ? ✓       | <1%            |

~~divotype Ab~~ → Canakinumab

|                                   |                 |   | Approved    |
|-----------------------------------|-----------------|---|-------------|
| 10. canakinumab                   | ✓               | ✓   | 0           |
| 11. fadotrelveomab                | x               | Still remain approved                                       | 0 - 16.6%   |
| 12. imciviroc                     | x               | ! Approved for imaging<br>? Market suspended<br>✗ withdrawn | <1%         |
| 13. capmatamab                    | x               | ✗ ✓   | 8% - 19%    |
| 14. nafetumomab                   | Not in database | ? ✓   | 6%          |
| 15. gemtuzumab<br>repeated        | x ✓             | ✓   | 0           |
| 16. alemtuzumab ✓                 | x ✓             | ✓   | 1.9% - 8.3% |
| 17. ibritumomab ✓                 | Not in database | ✓   | 1.3%        |
| 18. adalimumab ✓                  | ✗ ✓             | ✓   | 2.6% - 26%  |
| 19. omalizumab ✓                  | ✗ ✓             | ✓   | <0.1%       |
| 20. efalizumab ✓                  | x ✓             | ✓   | 6.3%        |
| 21. tositumomab ✓                 | x ✓             | ✓   | 11%         |
| 22. cetuximab ✓                   | x ✓             | ✓   | 5%          |
| 23. bevacizumab (1)<br>[Repeated] | ✓               | ✓   | 0           |
| 24. panitumumab                   | x               | ✓   | 4.6         |

|                                 |       |   |              |
|---------------------------------|-------|---|--------------|
| 25. Ramilizumab                 | ✓ ✗ ✓ | ✓ | 1-6%<br>4-6% |
| 26. eculizumab                  | ✗ ✗   | ✓ | 2%           |
| 27. natalizumab                 | ✗     | ✗ | 9%           |
| 28. golimumab                   | ✗     | ✓ | 4%           |
| 29. certolizumab pegol [Not in] | ✓     | ✓ | 8%           |
| 30. ofatumumab                  | ✗ ✓   | ✓ | 0            |
| 31. ustekinumab                 | ✓     | ✓ | 3-5%         |
| 32. tocilizumab                 | ✗     | ✓ | 2%           |
| 33. alemtuzumab                 | ✓     | ✓ | <1%          |

Ab with seg & % = 6

Ab with -seg | +% = 29.

$$\frac{43 + 15 + 21 + 33}{58 + 54} = \frac{112}{112} = 99$$

→ 8/12/2016 meeting with Andrew Martin

total data: doc 260 (258) check it again!  
 [about 4 missed]  
 4 (256) 2/3

1. Check those ab without name on Google.

Obtain a complete INN list with links → check whether all are included

whether approved one is randomly selected?

hypergeometric distribution Or t-test. Or U-test

24

t-test

e.g.

or

cont

try

general

Common

1. CD

2. RR

they both { s use both

unusual

source

1. diff

2. stan

str

3. do m  
'diction

If show correlation btw {

t-test / U-test.  
e.g. No. of patches  
or  $(1-\eta)$  score.

continuous distribution.

Or discrete into groups  
based on properties' combination

use hypergeometric test.

discrete probability

Try both of them.

generate all of them

Comments on talk:

1. CDR - loops within the region of variable

2. RRC: not relevant

they XML is not ... but plain text.

both (simple eg) script : at the server end.

use both screen scraper "Client" → when have info pairs → get info

(freq)

don't have to pass

the html just take

the request in

the format that to

request website,

just

keywords

unusualness

Source of data not mentioned : — background.

1. difficulties in obtaining the data.

2. standard numbering scheme : it gives a good mapping to structures

str

— structure & sequence usually given.

3. do not need to mention  
'dictionaries' in python

4 check PL and PL  
L detail not included

14/12 Wed.

2pm.

→ 12/12/2016

1. generate all pdb files
2. for bash scripts for pdb, access, distmat  
↓  
[sc] [sc]  
test file ✓

→ 13/12/2016

check pdb files:

abnormal:

1. ✓ actoxumab: unable to generate pdb → input file is empty (1)
2. alacizumab pegyl. pdb → due to space in filename, make it unreadable, use '-'  
fixed.  
[should not in total number]
3. alitocumab. pdb (1)
4. benralizumab. pdb (1)
5. bezlotoxumab. pdb (1)
6. bimekizumab. pdb (1)
7. blonertumab. pdb (1) can't read PDB coordinate [L106A, H1 not in pdb files]
8. brantictuzumab. pdb (1)
9. caplacizumab. pdb (3) → sequence only have H

10. ✓ claz
11. ✓ dupil
12. efung
13. ✓ emic
14. ✓ figit
15. flanro
16. gantu
17. girenti
18. ✓ hexatu
19. lucaru
20. ✓ lulizu
21. mavnil
22. ✓ navici
23. ortic
24. ✓ otler
25. ✓ ozora
26. panot
27. patri
28. ✓ placul
29. pogal
30. ponez
31. prez
32. pritox
33. ✓ rafif

10. clazakizumab. pdb ① can't read pdb coordinate
11. dupilumab. pdb ①
12. efungumab Scattered ② ✓ fixed.
13. emicizumab. pdb → only L. till e. starts from f.
14. fitatumumab. pdb ① can't read pdb coordinate
15. flavotumumab ③ H not numbered correctly. ✓ fixed
16. ganitumab. pdb ② ✓ fixed
17. girentiximab. pdb ② ✓ fixed
18. hexatumumab. pdb ③ only H.
19. lucatumumab. pdb ② ✓ fixed
20. lutzumab-pegl ③ only L. ~~not in this file~~.
21. marrilimumab. pdb ① can't read pdb coordinate
22. navicixizumab 2. seq. ①
23. orticumab. pdb ② ✓ fixed
24. otteruzumab. pdb ③ only H
25. ozoralizumab. pdb ③ only H
26. panobacumab. pdb ② ✓ fixed
27. patritumab. pdb ② ✓ fixed.
28. placitumab. pdb ③ \* Data in pdf: wrongly labeled.  
should be only L.
29. pegalizumab. pdb ✓ fixed
30. ponezumab. pdb ✓ fixed
31. prezalumab. pdb ✓
32. pritoxaximab. pdb ✓
33. rafivirumab. pdb ① can't read pdb coordinate

Dataset

• 3 abs h

④ data

epi

\* bapi

\* nim

\* Ocre

wom

mo

14/12/2

Y.A list

✓. revis

✓ 3. refer

— k :

• add ve

\* in py

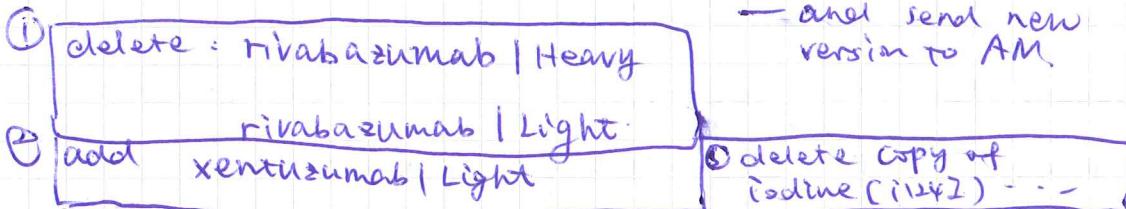
put to

"aerm

②

34 Xentuzumab.pdb ③ H can't number correctly ✓

\* In Dataset: total number : need new calculation



35 Tramipatumab.pdb ① can't read pdb file

36 Tenatumumab.pdb ① can't read PDB file

37 Teprotumumab.pdb ① can't read PDB file

38 Tovetumab.pdb ① can't read --

39 Tremelimumab.pdb ① --

40 Ytrocublumab.pdb ① --

41 Robarilizumab.pdb ③ only H

42 Xentuzumab.pdb ③ L not added correctly. ✓

~~42 = 41~~

~~27 + 15 =~~  
~~[fixed]~~

43 Napntumomab estafenox ③ pdb : only L.

{ trastuzumab - d : same VH & VL, delete one.

and renamed as  
trastuzumab.seq

total

29 - pdb 0 bytes

19 - ①

10 - ③ = only H/L

## Dataset:

- 3 abs has "2"

### ④ dataset

| <del>epitumomab - no sequence</del> | <del>epitumomab</del> | Check the sequence  |
|-------------------------------------|-----------------------|---------------------|
| * bapineuzumab - no sequence        | => ....               | ✓                   |
| * nimotuzumab - no seq              | =>                    | unable to get ≈ pdb |
| * ocrelizumab - no --               | =>                    |                     |

with seq : 261 } 371  
no seq : 110

14/12/2016 [Discussion with AM]

- A list of unable to 'get pdb'

- review topic : problem of immunogenicity
  - reference upload.
- ① different antigenicity
- ② immunogenicity \*
- Geneenel → ③ problem in triimmunogenicity
- in biotherapy antibody drug
- k : keep files from templates [the first part of medology]
- add version description to "#..."
  - in Python → remove `if` Linux (return)  
`r` → remove it Windows (return)
  - put testfiles to TEST/

"aerm4"

②

→ Tom's thesis : predict B-cell epitopes & looking at  
AB stability and tendency

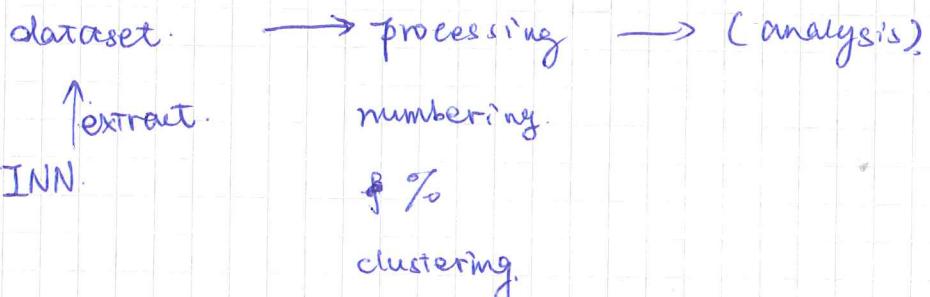
② two papers about immunogenicity

Friday 12pm. takes in meeting room

Interview report:

everything before bioactivity

aim of project: other way to predict immunogenicity



\* explain: patches

[approved or not]

Term 2.

①

Process:

Database →

for a given sequence →

automated → for a <sup>single</sup> antibody

numbered → freq. → clusters

Pdb → mat  
acc

② analysis

database:  
with ab: 26 [some get several seq.],

actual 296 ab's processed

Remain

In

In Fi

29

110

\* aut

"t

Task f

double

✓ Check

In alert

Not

311

294 + 1

294

27

167

29

Remove : ^M [when generates from DOS/Windows to Linux]  
transfer  
~~dos2unix filename~~

In emass:

M-x revert-buffer-with-coding-system:  
~~utf-8 dos~~

M-x replace-string RET C-q C-m RET RET

In Final Data:

294 sequenced. [2 antibodies has double sequences]

110 no seq.

\* automated .sh:

"unable to open file or directory", ?

Task for 13/01/2017:

double-  
✓ Check the list from two paper: whether they have  
① name  
② sequence

In database:

Not approved, with sequence:

311

$$\underline{294 + 17 = 311}$$

$$\begin{array}{r} 294 - 27 = \\ \hline 167 \end{array}$$

$$\begin{array}{r} 296 + 18 = 314 \\ \hline \text{(some get repetition)} \end{array}$$

\* ~~gemtuzumab~~ name

should be

~~gemtuzumab~~

gemtuzumab

- M-x auto-fill-mode :

set margin:

C-x f : ← add argument  
↓  
C-u 20 C-x f

- ~~search~~ C-s : search for search

the 1<sup>st</sup> C-s: give one occurrence

the 2<sup>nd</sup> C-s: move to the next occurrence

{ C-s = search forward (after cursor)  
C-r = search backward (before cursor)

- C-u 0 C-l

- C-x 2 : split screen into 2 windows

- <esc> C-v : scroll the bottom window

- C-w o : move cursor to the bottom window

- M-x make-frame <return>

- M-x delete-frame <return>

- recursive editing level: [ ]

to get out of it: <esc><esc><esc> also: an "all-purpose  
"get-out"

C-h C +  
~~←~~ ~~h~~ ~~+~~ uncharacter for help

or

M-x help <return> ~~←~~

RET : return ↵

{ C-h i + : plot into buffer \*info\*  
C-h C + : a brief description  
C-h K + : documentation  
C-h F + : describe a function  
C-h V : for variable  
C-h A - : type a keyword,  
Emacs finds all command  
with that keyword.

•  $C-X C-B$  : list buffer ← emacs stores each file's text inside it

•  $C-X B$  : switch buffer

$C-X B + \underline{\text{filename}}$

•  $C-X S = C-X S$  : save some buffers

$\left\{ \begin{array}{l} C-X + \rightarrow \text{one character,} \\ \text{or} \end{array} \right.$

$\left\{ \begin{array}{l} M-X + \rightarrow \text{a name,} \\ \text{type: } M-X, \text{ then name of} \end{array} \right.$

•  $C-X C-C$  : end the emacs and save any changes command

•  $C-Z$  : exit emacs "temporarily", and can go back again.  
Or suspend.

— to resume emacs: "fg" or "%emacs"

<tab> : find the name when with a beginning.

•  $M-X replace-string$  : changed  $\downarrow$  free be replaced  
                                  altered the one  $\rightarrow$  to take place

• unsaved file : has an " $\#$ " ~~at the~~ # filename

• \*\* : means have made change

•  $C-h M$  : view documentation on your current mode

•  $M-X \underline{\text{mode name}}$   $\left\{ \begin{array}{l} \text{return} \rightarrow \text{turn on} \\ \text{return again} \rightarrow \text{turn off} \end{array} \right.$

- When trigger a ~~dis~~ disabled command,  
type  $y/c$  (space bar) to ~~g~~ yes  
 $m$  to no
- $C-h k + "C-f"$   
 ↓ display documentation  
 help for  $C-f$
- $C-X 1$  : get rid of extra windows of emacs,  
 and only keep one window  
 ↓  
 the ~~one~~ one contain the cursor  
 - start with  $C-X$ ,

- to do with the window
- {  $C- <Delete>$  : delete the previous character  
 $C-d$  : delete the character just after the cursor.
  - {  $m- <Delete>$  : delete the previous word,  
 $m-d$  : delete the next word after cursor.  
 $C-k$  : kill from the cursor to end of line  
 $m-k$  : kill to the end of current sentence.

$C-@/C-ccpc$  [.....]  $C-w$   
 ↑  
for this part

Killing — can be reinserted →  $C-y$  (yank the recent kill)  
 delete — no  
 $m-y$  (for yank more previous kill)

- $C-X u$  : undo the change

~~=  $C-A$~~   
 $C-_-$

- $C-X C-f$  : find a file  
 $C-X C-f +$  file name, path

- $C-X C-s$  : save the file ; when saved, original one is renamed  
 to " $n$ " + filename

Emacs command: C for "CTRL" M for Meta/Alt.

• C-x C-c : exit emacs

• C-v : move forward

• C-v : move up of the page      M-v : move ~~down~~ <sup>backward ↑</sup> the page

• C-l : Clear screen and redisplay ~~all the~~ all the text, moving the text around the cursor to the centre of screen.

Previous line C-p

↓

↓

backward character --- current cursor --- forward character  
C-b    C-f

Move ~~bit~~

Previous

↑

↑

previous word M-b    forward word M-f

Next line C-n

{ C-a : to beginning of line  
C-e : to end of line

{ M-a : to beginning of sentence  
M-e : to end of sentence

M-< > : to beginning of whole text  
M-< : to end of whole text

• ~~repeat~~

• C-u : numeric argument, a repeat-count.

e.g C-u 8 C-f

move forward 8 characters

C-u 8 C-v : move ↑ by 8 lines  
C-u 8 M-V : move ↓ by 8 lines

(C-g) : - stop a command

- discard a numeric argument or the beginning of a command that do not want to finish.

↓

e.g ~~C-u 8 C-f~~  
C-u 8 C-g C-f  
      ↓  
      C-f

<ESC> C-g

↓ get rid of Esc.

LAB BOOK 2

M101

ZIYI CUI

LAB BOOK2

M101

ZIYI CUI

M martin - col 2

→ 16/01/2017

updated database:  
311 + (294 previous)  
96 - (110 previous)

update:  
abnum/  
abpath/  
refreg/

\* automate process to get cluster info.

{ Version 1: Process many at one time  
V 2: Only process one antibody.

### U-test & T-test

T-test:  
If randomly selected.  
two sets:

same means

\* Whole Set      approved

{ high I  
low I.

using I.

↓

{ from two lists?  
from database

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$\bar{x}$ : mean  
 $s^2$ : variance  
 $N$ : sample size

Welch's T-test:

for testing sample  
with unequal variance  
- to test whether they  
have equal mean

Mann-Whitney U test: regarding the distribution of the underlying variables, a non-parametric test of the null hypothesis that two samples are drawn from the same population.

being compared

\* suitable when the distributions are not normal.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

n: size; R: sum of ranks (pool two sets of data together)

## S0302 - Statistics M

Analysis

Step 1: based on immunogenicity data from two papers:

- using Welch's t-test: do not assume equal population

Sample 1

Sample 2

Approved

~~unapproved~~. All unapproved.

u test

\* When get repeat value:

$$\text{rank} = \frac{\sum \text{order number}}{n}$$

$$z = \frac{U - \text{mean}}{\text{SD}}$$

calculate  $\sigma$ : [SD]

$$\tau = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$\text{Ocorr} = \sqrt{\frac{n_1 n_2 ((n_1 + 1) - \sum_{i=1}^k \frac{t_i^3 - t_i}{n(n-1)})}{12}}$$

$t_i$ : number of subjects sharing rank  $i$

t' test

\* degree of freedom

$$= \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^2 V_1} + \frac{s_2^4}{N_2^2 V_2}}$$

Both two tests confirm that approved immu- is significantly different than unapproved one.

Task for week 23/01/2017:

get all clusters automatically.

The automated process:

Problem: current machine : do modelling  
across : { do request  
          : solve  
          : distanc

BOTW

: do clustering

27 also  
21

27 do m

of 28

scoring

① bootstrap  
usually

= [i-unknown]

② sum up  
in each

Test of

- limited  
only for

ap

27 pols get 0 bytes of solv.  
also  
27

→ side-chain distance.

[ ]

27 do not get proper polfile

of 287:

158 - No.

129 - clusters

scoring of  $X$  clusters

j/ convert

0 instability

① unusual score.

$$= [i - \underline{\text{unusual}}] \quad \left( \frac{1}{\underline{\text{unusual}}} + n \right)$$

Y

homogeneity

① actual data

② group them

② sum up all the residue frequency  
in each clusters?

Test of about other drugs' clusters info.

- limited sequence & clusters  
only for those approved.

should have sequences

approved

47

unapproved [base in database]

$$311 - 5 - 27 = 278$$

\* choose 5 years older at least.

### ④ Method 1.

frequency of residue  $x$

$$U_x = 1 - f_x$$

$$\downarrow \quad \downarrow t.$$

unusualness

may be affected by number

$$\rightarrow U_c = \sum U_x \quad \text{①}$$

$$\bar{U}_c = \frac{\sum U_x}{n} \quad \text{②}$$

If set significant for first approx test, then stop.

### Method 2.

$$i_x = \frac{1}{(f_x + n)} \quad \text{③}$$

$$\textcircled{5} \bar{i}_c = \frac{\sum i_x}{n} \quad \text{④}$$

$$U_c = \sum i_x$$

### ③ Method 3:

$$\textcircled{3} U_c = N$$

$\downarrow$   
number  
of residue

\*  $t=0.2$ , threshold

$$\text{unusualness} = t - f_x$$

$$\text{score} = n \cdot \frac{t - f_x}{t}$$

### Review

#### Introduction:

biological ~~therapeutics~~

use & problems

where to put antibody immunogenicity

X antibody structure Info  
antigenicity & immunogenicity

① development

② factors

③ detection

④ prediction

⑤ eliminating method

#### discussion

at the <sup>a short summary</sup> we can do  
antibody drug ... better on antibody

[mainly focus on]  
the problem

immunogenic test lead to immunogenicity

Model  
base  
modi  
export \$P  
PA

Wed :

→ 30/01/10

①  $U_c$

$t$

$U_t$

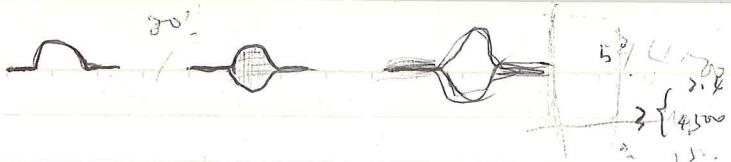
②  $U_c$

$t$

$U_t$

20+15

Modify path:



.. bashrc in home

=  $\Sigma u_i x_i$

modify that.

export \$PATH

$PATH = "\$PATH : /aeron/usr/local/apps/Python-3.6.0/bin:/aeron/usr/local/bin"$

\*△ Permission denied.

[Wed : 14:30 next]

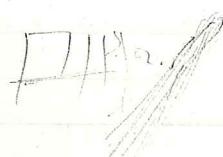
command bash or shell

/emars-nw.bashrc

→ 30/01/2017

for test for approved and unapproved regarding their clustering scores

$$\textcircled{1} U_c = \sum U_{xi} \quad U_x = 1 - f_x \\ (100 - f_x \cdot 100)$$

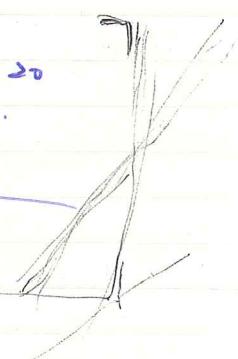


t test:  $|t| = 0.9722 < 1.6652$  at  $p=0.05$

U test:  $Z = -0.6956$  at  $Z = -0.69$ ,  $p=0.2451 > 0.05$ .

$$\textcircled{2} U_c = \sum U_{xi} \quad U_x = \frac{100}{f_x + c} \rightarrow c = 20$$

t test:  $|t| = 0.8396 < 1.6652$  at  $p=0.05$



U test: same result from \textcircled{1}

$$20 + 15 = 35$$

list excluding (unapproved within 5 years)

bavituximab - unapproved.

approved  
46

unapproved  
160

older  
than  
5 years

206

→ Meeting on 01/02/2017

→ ~~Meet for week of 2017~~  
search the two lists of drugs.

↳ try find the sequence.  
hunting back to the literature.

get at PHASE I trial.  
INN name

① for those earlier than 5 years

Marketing - Who produce it  
and papers

try to find the names

[ distribution

Some do failed due to immunogenicity

② drugs

whether

immunogenicity in cluster score.



plot against

③ change cutoff of setting clusters

maybe 10/5 for frequency.

distance matrix

automating the process

from cutoff → several explore set of cutoff.

machine learning.

might learn features distig

Wed. 12:00 Next meeting

using Method ③:

$$U_C = N$$

t test:

$$(t = -0.8875) \text{ at } df = 7, p = 0.25$$



u test:

$$z = -1.0016$$

at  $z = -1.00$ ,  $p = 0.15 > 0.05$ .

$$\textcircled{4} \quad \bar{U}_C = \frac{\sum u_x}{n}$$

④

03/02/2017:

✓ Automate the process to do statistical testing using Python.

## 1 Generate Data

- ① } choose which function to use:
- ② }
- ③ }
- ④ }
- ⑤ }

↓ python3 .py    folder    function number  
[→] [←]

generate Python dict.

{ mabs : score ; ... }

↓  
CSV files. ✓

OR generate CSV from bash script.

The data type that can be used in Python's  
for testing

[scipy.stats.ttest\_ind] module

grouping data into

- Approved /
- Unapproved /

## 2. Testing.

array  
approx  
score

→ Meeting

① Problem

[ Some ]

② drug  
No p

③ immu

not

many

✓ review

\* stem

\* penetr

---  
Solution

under  
[ /aermy ]

[ - user ]

E

[ system ]

array 1  
approved  
score list

array 2  
unapproved  
score list

→ Meeting on 08/02/2017

① Problem in drug list:

[ Some one ~~been~~ approved  $\Rightarrow$  unapproved? ] <sup>pip</sup> under Python 3/  
now withdrawn  
solution: to see the reason.

② drug list in paper:

No paper has name, but also have sequence.

③ immunogenicity vs. scores

not linear correlation

maybe due to too small dataset?

④ review.

- slow down project
- penel for reviewing

— Solution ①. ✓

under

[ ~~acm/nor/bin~~ <sup>local</sup> ~~= /app/Python-3.6.0/bin/Python3.6~~ /pip3 ]

[ --user ]

△ Problem, SSL module ~~too~~ missing

see  
pip3 help

[ ] install --user ↓  
modulte name.

[ ] install ↓ --user  
modulte

[ yum install — ]

→ 21/02

Task

1. obj

2. pyt

3.

1. Segu

[13]

+ 22

- update

cancel

+ update

Cfir

2) stat

→ 22/02

1) among  
@ En

ii

on

11+22=33

2) statist

• ~~stat~~ w

\* Matl

• perform

③ try to do group instead of actual values

\* .c .o various lib .py dependence  
Source → object linked executable  
source -> build

① Configure — where to put ; check if optional present or not

② make — building softwares

③ make test — make sure different files works

④ make install

4 steps to install large modules

\* tutorial for python :

[www.bioinfo.org.uk/teaching/bb1k/tutorials](http://www.bioinfo.org.uk/teaching/bb1k/tutorials)

By Monday noon

22/2 , Wed 11:00

~~→~~ 21/02/2017

[ Task :

1. abry list - sequencing
2. python script for multiple scoring
- 3.

1. Sequence added:

13 antibody (new)  $\Rightarrow$  35 antibody  
 $+ 22$  antibody (old)

• update autoClustering file.sh

~~autoClusterForDatabase~~

• Update autoClusterForDatabase.sh ✓

(first draft, need to format it.)

2) statistics

folder

abnum 2102 /

refreq 2102 /

abpdb 2102 /

& rotv 2102 /

dist 2102 /

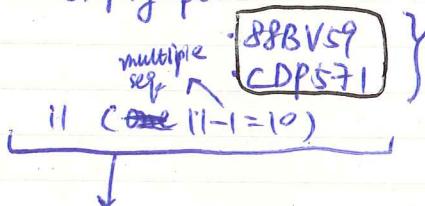
not be

\* integrated into  
the ~~get~~ main folders  
(e.g. abnum/)

~~→~~ 22/02/2017.

1) \* among 13 antibodies:

\* Empty pdb:



ERROR:  
Can't read PDB coordinates

Input file is empty.

only found 5 clusters

$11+21=32 \rightarrow$  linear regression : No trend.

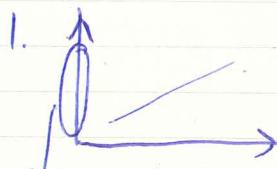
2) statistics

\* ~~write~~ write two csv files:

\* Module itertools Not included.

\* perform test : cannot perform reduce with flexible type.

Meeting at 24/02/2017

1.  do pearson correlation

ignore this.  
[argue that]

\* Email the complete "empty pdb"

Run abymob: repeatedly..

- choose template basing on chemical classes
- assemble  $\downarrow$  together
- refinement : — energy minimization

$\rightarrow$  lev Readme.mol.

( $-v = 3$ )

\* 1. only heavy chain,

$-v$  : verbose . : give detail.  
 $= 5$

$-k=2$  : keep all intermediate files

\* use .seq, run abymob with

$-v = 5$

$\Rightarrow$  see in which stage  
error occurs

$-k=2$ .

$\hookrightarrow$  give the directory where temp files are in

$\downarrow$

~~hh~~

ls -ltr (modelling of groups of side chain)

sorder.dat

\* .seq.tpl in self directory  $\rightarrow$  in .seq directory

$\hookrightarrow$  for test

① each one - each file  
 $\hookrightarrow$  just look for the first template

② +test Compose sequences:

pdbs2pir /acm/L8mhome/acymol/DATA/abpolib/\_.pol

template name

pdbs2pir /acm/L8mhome/acymol/DATA/abpolib/\_.pol



will give the sequence of template

③ +write script:

take .seq to pir file:

>P1; PDBPIR.

[

] command,

(1+) Sequence + "

\\";

Compose:

nw -d \_-pir \_-pir.



Give the alignment.

Look at 'IDshort' should be 100%



means structure already solved:

just use template file as structure:

Category (2+2) empty pols:-

+test:

+test. /acm/L8mhome/martin/scripts/ttestsp1.

nofscript

nofmartin

[Wed 10]

27/02/2017.

task:

- ✓ ① check abymod.pl.
- ✓ ② find corr.coeff.

Empty pdb:

27 mabs + 88BV59  
CPP571.

folder

Finding Errors:

\*1 abymod.pl -V=5 -k=2 .seq → .pdb

\*2 ls -ltr directory for temp

\*3 ✓ find .tpl files. — look for the 1<sup>st</sup> tpl.  
look for template name.

results

on

ofh

→ 28/02/2017

✓ calc

✓ score

• try "j

(to edit .bashrc file:  
emails w/.bashrc )

change from:

export PATH = "/acrm/usr/local/apps/Python-3.6.0/bin:\$PATH:  
/acrm/usr/local/bin"

↓

"- - - - - + \$PATH:/home/bsm/martin/bin" ~~PATH~~

\*4 .seq → .sir.

use Python script "seq2pir.py", .pdb → .sir  
.pdb > pir → .sir

template

.pdb → .sir

① empty

② plot

③ score

Meeting

1. Create

.out

② step c

\*5 Compose

export DATADIR = /home/bsm/martin/cluster

nw -d 40jf0.pir mabs.pir

folder "abpab2to2":

found polb on '27 +2' list. with empty polb

results:

only 1 works in modelling:

others {① only H/L : choose seq. from ("choose complementary")  
② model not match  
③ low identity → need double check.

→ 28/02/2017.

✓ calc r. ✗ checked abnormal value

✓ score + test by python

try "grouped" method.

- for drugs from two papers (not working)

- for our database

\* hypergeometric test

→ 1/03/2017.

① empty polb - solution

② plot: person correlation  $r=0.36 \Rightarrow$  Grouped - not working.

③ score & test (automated)

Meeting

1. Create directory understandg.

.err (from)

2. Create .seq files understandg/scripts/striper perl — do3 file

.seqc

transfer from DOS → Linux format.

② check

→ 09/03

Solution

② DOS file: convert → polbarr/DOSseg Corrected.

adel

test

We

- \* replaced .seg in abnum/ [no DOS files now]
- \* new data [polb + dist + solv + ~~file~~] △
  - | unable H chain only
  - | unable to get distmat
  - | file, but get solv file.
- +
  - | L chain only polb [o]

total

13 antibodies

11            2  
not found    no cl.

| unable H chain only

| unable to get distmat

| file, but get solv file.

|  
|  
|

problem: Command wrong:

should be

✓ distmat -s -c L,H -p

↓

Corrected for files in "dist/"  
"dryDist"

Since using wrong distmat files,  
we need to do the  
t test & u test again

~~~~~

In python:

~~t = p =~~ p =

u = \_\_\_\_ ; p = \_\_\_\_ ;

how to interpret?

• Patch

L

J

try

• The m

don'

\* Ma

O

$$\frac{23+4-9}{2} = \frac{24+27+25+24}{4} = \frac{98}{4} = 24.5$$

→ 09/03/2017.

adef polberr/cluster → mobClusterExtracted  
 13 + 207 => 220

test result:

Welch's t test better than Mann Whitney U

- Method 1 : the best

→ 10/03/2017. Meeting with Andrew

\* patch parameter - cutoff

↳ regenerate the cutoff

↓

try each combination

\* The rest 14 [unmodel]:

don't get sidechabckbone - can't build model

model: CA only

\* Manual fix:

~~~~~ remove the CA-only model.

see whether CA-only

OR:

run modeling:

[ -exclude = XXXX[ ] ]  
 ↓ (only name, no filetype)

structure in abpolb/fb/.

calculate canonical class:



put by the Numbering Scheme  
(based on how defining CDR)

determine the key residue of CDR.

Each highly ativariable:  
overall consistency

clusters of CDR.

e.g. 21/1A

CHOTDA length, clusters

β  
P  
C  
G

Rank of template:

✓ ① sequence alignment

② chemical matches

several templates → Models

splicing inaccuracy [the same canonical class  
ensure some accuracy]

CDR L1  
L1 } the highest for L1.

L2

L2

L2

Then energy minimization       $H\beta = 30^\circ$

\* What if

loop  
β<sub>1</sub>  
β<sub>2</sub>  
β<sub>3</sub>

B-sh  
enough  
table  
full  
of β-sh

replacem

do M

crack

Fixing

truncat

↓

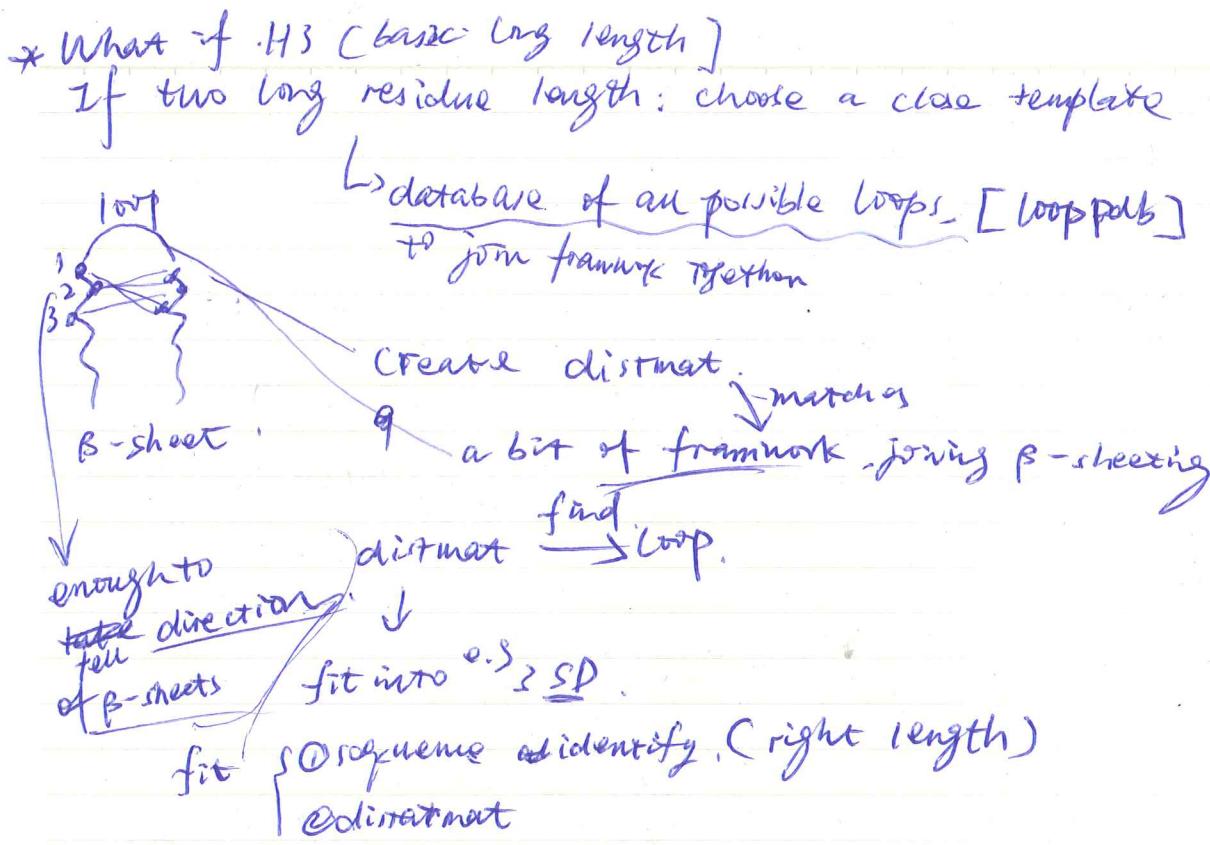
renumber

↓

re-forma

< 2

change  
the gra



do the less important

! !  
 crucial be the constraints

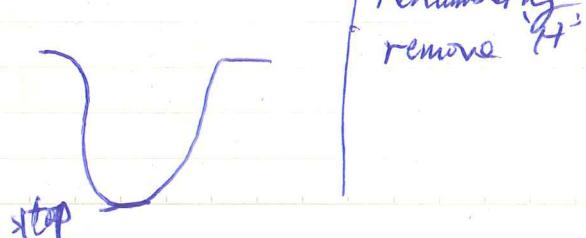
Fixing any missing errors:

truncate additional residue, usually at C-terminus

↓  
 renumber (will be correct in the last step).

↓  
 reformat

$< 2$  step minimization  
 change the gradient

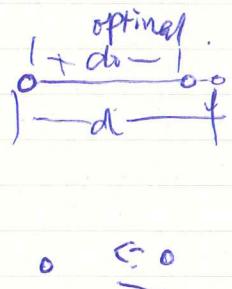


• Rings can get bend during minimization  
 ↓  
 replace by its on sidechain  
 note.

MCC

ROC:

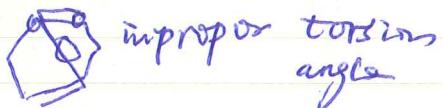
T



$$F = k(d - d_o)^2$$

$$E = \frac{1}{2}k(d - d_o)^2$$

$\chi_2$  torsion angle



Minimization: first calculate the direction for each atom to move

Each step calculate the direction + move

#### ① Structure

② run optimize the cutoff.

(Method 1/2)

baked

③ Reorganize the atypatch

✓ git rm [~] backup:  $rm *{~}$

stop backup:

edit: `~/.emacs`

`(setq make-backup-files nil)`

`(setq auto-save-default nil)`

Tue 2

• edit README.md [ README.md for each ~~fit~~ folder ]

finish

④ score { below threshold. } above threshold.  
 { non-immunogenic }  
 { immunogenic } ↓  
 adjust it.

|           |    | actual |    |
|-----------|----|--------|----|
|           |    | AP     | AN |
| predicted | TP | FP     | TN |
|           | FN |        |    |

TP = True Positive  
 FP = False Positive  
 FN = False Negative  
 TN = True Negative

For each threshold, calculate statistical parameters

S<sub>r</sub>  
 S<sub>p</sub>

MCC [Matthews' Correlation Coefficient]

a table we can compare in with

(MCC) is the most informative value to use

ROC: receiver operating characteristic

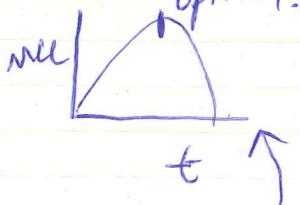
TN rate      FP rate  
TNR vs FPR :  
 $\frac{Sn}{Sp}$        $1 - Sp$

for the perfect predictor  
if it is a perfect predictor,  
only TP and TN,  
no FN FP

a random

does not give whether it is good.

It does tell how much info this predictor gives.  
optimal.



based on the scores, use  $t$  to get optimal

$$6.5 - 6.6$$

nil) ~~stop~~

t nil)

Tue 27<sup>th</sup> 11.am.

finish report around one

concise.

\* Introduction.

variables

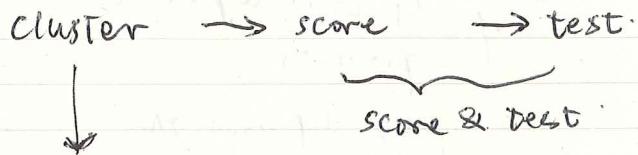
parameters

u'  
in  
nt

→ a table that  
we can calculate  
in wikipedia

$$311 + 96 = 407$$

15/03/2017



1. delete all ".~" files:

find . -name ".~" -type f -delete)

also stop autosave file in gedit  
& emacs

2. rearrange:

DATA TEST.

\* change in scoreCluster.py : ← change in .sh.

Function input:

◦ ./distmat { -S -C L,H -P -pdb } → ~~dist~~ { dist }  
    { -a }

◦ ./clusterResidues.pl [-m=3] [-d=4] { dist } { sa } " > " { freq } " < 50 "

→ generate cluster files. { 10 }

t method → Welch t test

Score

$p < 0.05$

$p < 0.1$

-d  
freq  
-c

ERROR

o Re-run

Result

$p < 0$

Fixing

Zn fo

Meet

1. for

1. no

2. ss

plot

-d: 1-10

freq <10 - <30

-c: diameter

ERROR:

ScoreCluster.py line 251, in rate\_clusters\_4

~~listoffreqs.append(freqDict[field[0]])~~ → Error  
fixed.

• Re-run again

Result of training.

p<0.05 OR p<0.1

w.

Fixing the rest of 14 unmatched seq [Colony]

In folder ipdbCorrectedForCAonly/

Meeting True.

1. ~~false~~ show average: not work

1. not do averaging method

2. so we decide to try those — counting No is the best

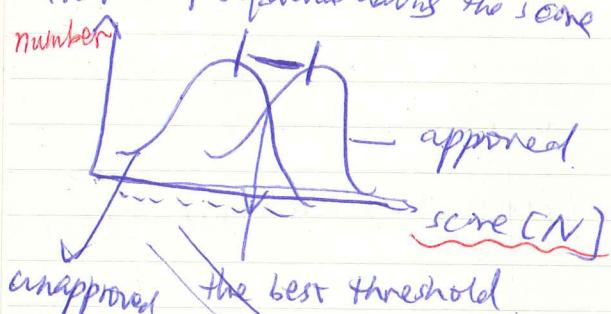
Plot the distribution

→ certainly for the best one

3-14-3

### 3-14-3. predictor.

L plot: to show the means are significant  
the No of sequences having the score number



but try ~~other~~ these threshold, divider to  $n/100$

so, if ~~< above &~~ predict bad  
~~less &~~, predict ok.

Then create

|    |    |
|----|----|
| TP | TN |
| FP | FN |

for each threshold,  
calculate to MCC  
to determine the best one

17/18 10

a. needed

b still can't correct. Top directory: Readme test general

+ green means correct ~~Readme~~ Readme.md for each folder

L send list of

fixed - method

and where to find seq files

Use th

→ Monday

"PUTT"

[Machine]

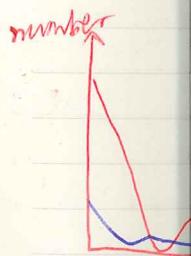
[practical]

par

type

before

3-14-3



thresh

see u

Use those failed nabs as the test for our predictor.

→ Monday, 27<sup>th</sup>. 4pm.

"PUTTY": allow to log in to remote machine

[Machine] = sec1.biochem.ucl.ac.uk — hostname

[protocol] ssh

password

log in as: zcbtzcw  
password: 88!!plap

then: ssh martin-cd04

type: ssh < martin-cd04 ~~-log in as~~

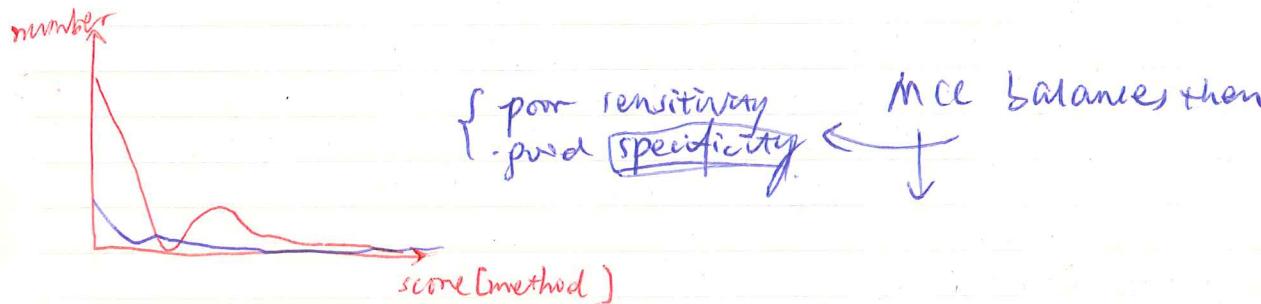
password

before next Monday

3-14-3 : -d=3

freq < 14

Method 3: count the number of residues in clusters



threshold = 1 [score]

see which antibody in approved [score = 3].

★ \* If chimeric, — then reasonable.

$\frac{18}{160} = 11\%$  Although low, but still saves ~~many~~ much money

\* calculate MEL for each combination,  
just like using — for t-test.

No new  
to update

t-test: separation

math

MEL: predictor of performance

early

\* try other methods.

new comp

\* test three new fixed antibodies to see the  
predictor

seen lab

bar chart instead of curves.

min number of cluster member: not large,  
but contain reversal residues

high score → higher probability to be immunogenic.

bavituximab: chimeric, approved  
high score in method 3  
score = 5

Investigational:

Mostly they are antibody conjugate

4

- \* no need to update
  - \* update the approved list:  
[72 now]
- make sure now where in my approval list one am in
- \* early draft send to Am
  - \* re complete draft once -

seen lab book.