

Human Molecular Genetics

MBB 438



We talk about a diverse range of topics related to human genetics and genetic diseases.

Topics

- Human genetic disease
- Disease gene mapping and characterization
- Gene therapy using Crispr
- Cloning and stem cells
- Genetic counselling
- Cancer genetics
- Animal models of disease

INSTRUCTOR:

Esther Verheyen



sort & uniq

16

sort sorts
its input

\$ sort names.txt

the default sort is
alphabetical.

sort -n
numeric sort

'sort' order	'sort -n' order
12	12
15000	48
48	96
6020	6020
96	15000

sort -h: human sort

'sort -n' order	'sort -h' order
15 G	45 K
30 M	30 M
45 K	15 G
200 G	200 G

useful example:
du -sh * | sort -h

uniq removes duplicates

a
b
b => a
a
c
c

notice there
are still 2
'a's! uniq
only uniquifies
adjacent
matching lines

sort + uniq = ♥

Pipe something to
'sort | uniq' and you'll
get a deduplicated list
of lines! **sort -u** does the
same thing.

b
a
b
a | sort -u => a
b
or sort | uniq

uniq -c
counts each line it saw.

Recipe: get the top 10 most
common lines in a file:

\$ sort foo.txt
| uniq -c
| sort -n
| tail -n 10

I use
this
a lot

Group Data

dplyr::group_by(iris, Species)

Group data into rows with the same value of Species.

dplyr::ungroup(iris)

Remove grouping information from data frame.

iris %>% group_by(Species) %>% summarise(...)

Compute separate summary row for each group.



iris %>% group_by(Species) %>% mutate(...)

Compute new variables by group.

