Thursday, 26 July

- Convert (remapped), sorted, filtered, (phased) .vcf to OneMap input file (.raw format)
- Using R
- Linkage mapping with OneMap

Cygwin: a LINUX-like environment for Windows

https://www.cygwin.com/

https://cygwin.com/faq.html

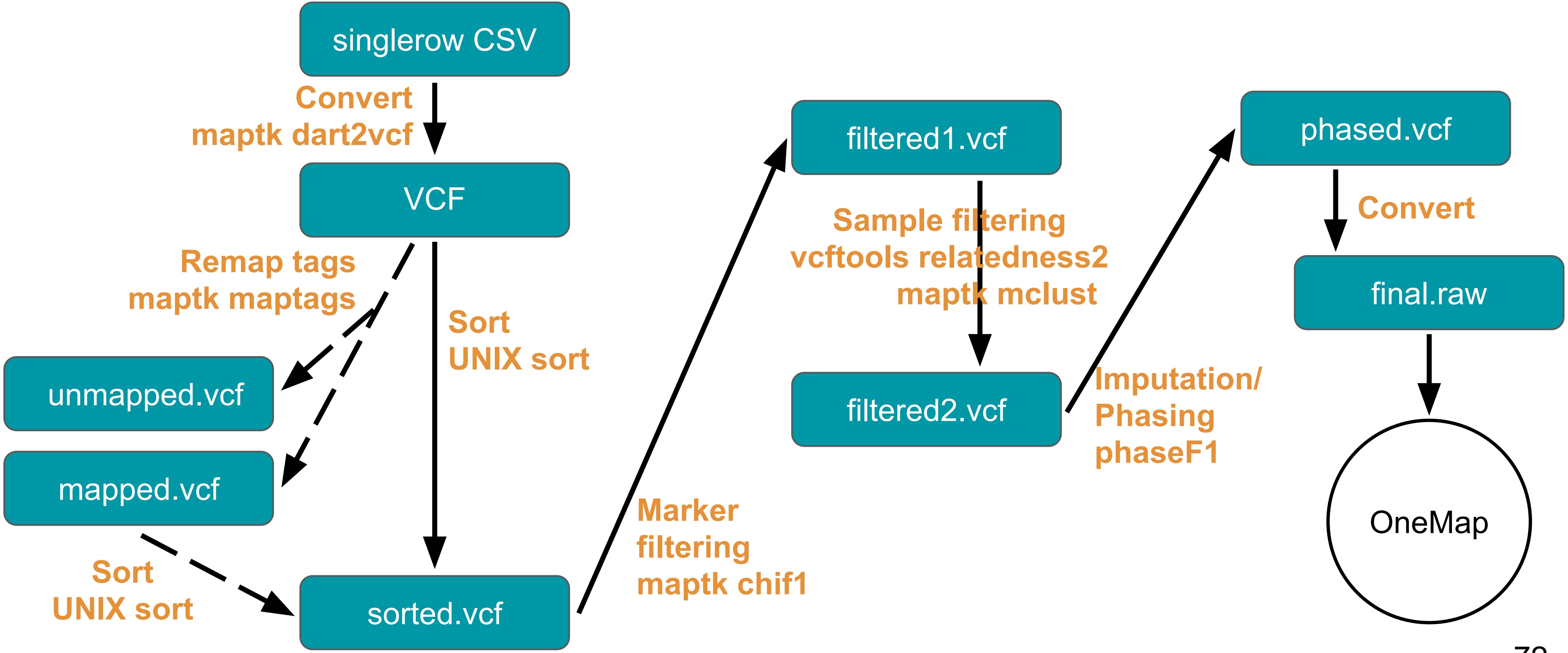
https://cygwin.com/cygwin-ug-net/cygwin-ug-net.html

for e.g. vcftools, you will have to compile it on your computer

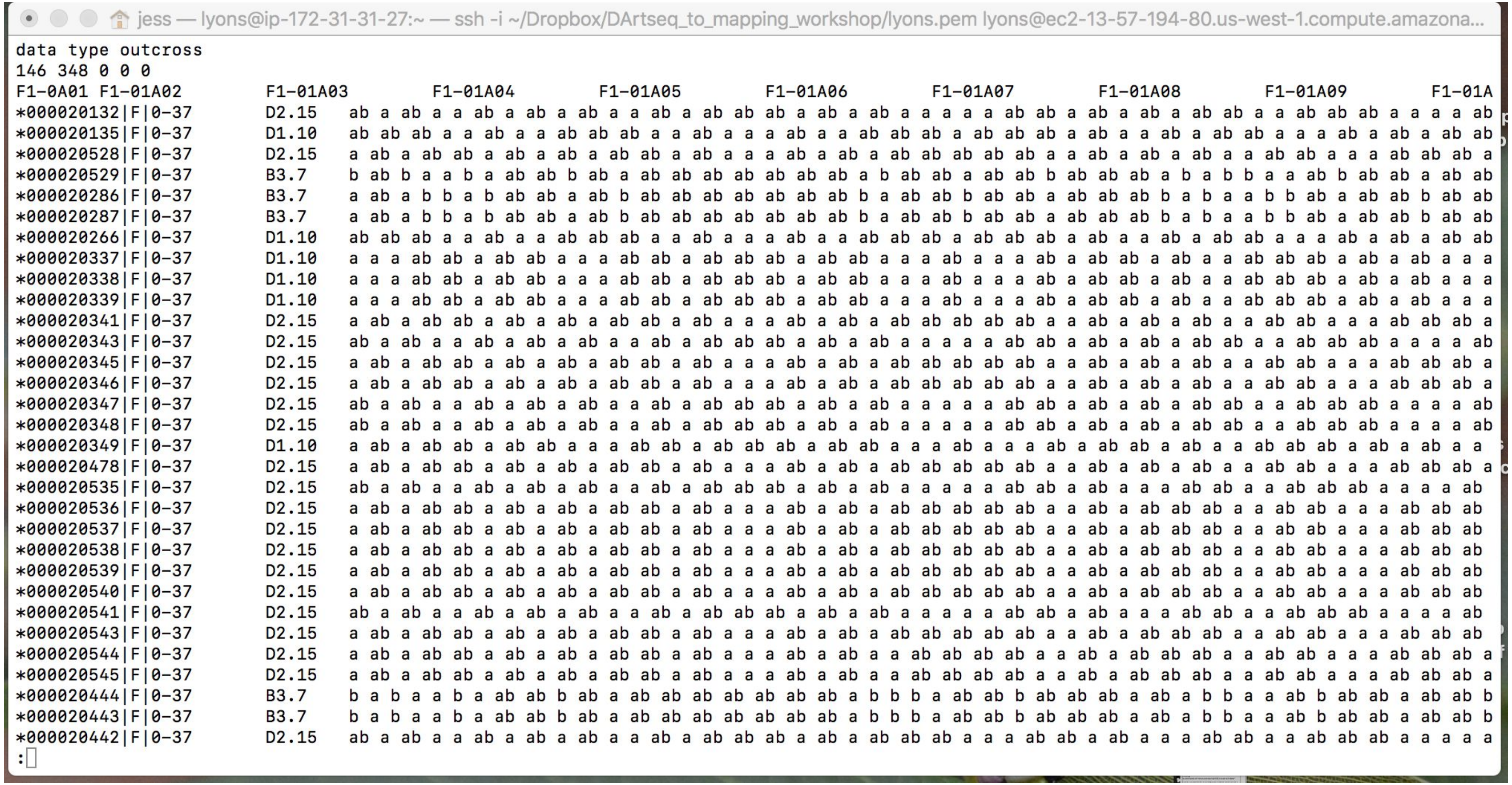
any arguments that reference the filesystem must be in Windows (or DOS) format or translated

➔ `cygpath` utility (https://cygwin.com/cygwin-ug-net/using-effectively.html)

DArTseq to mapping: analysis flow chart



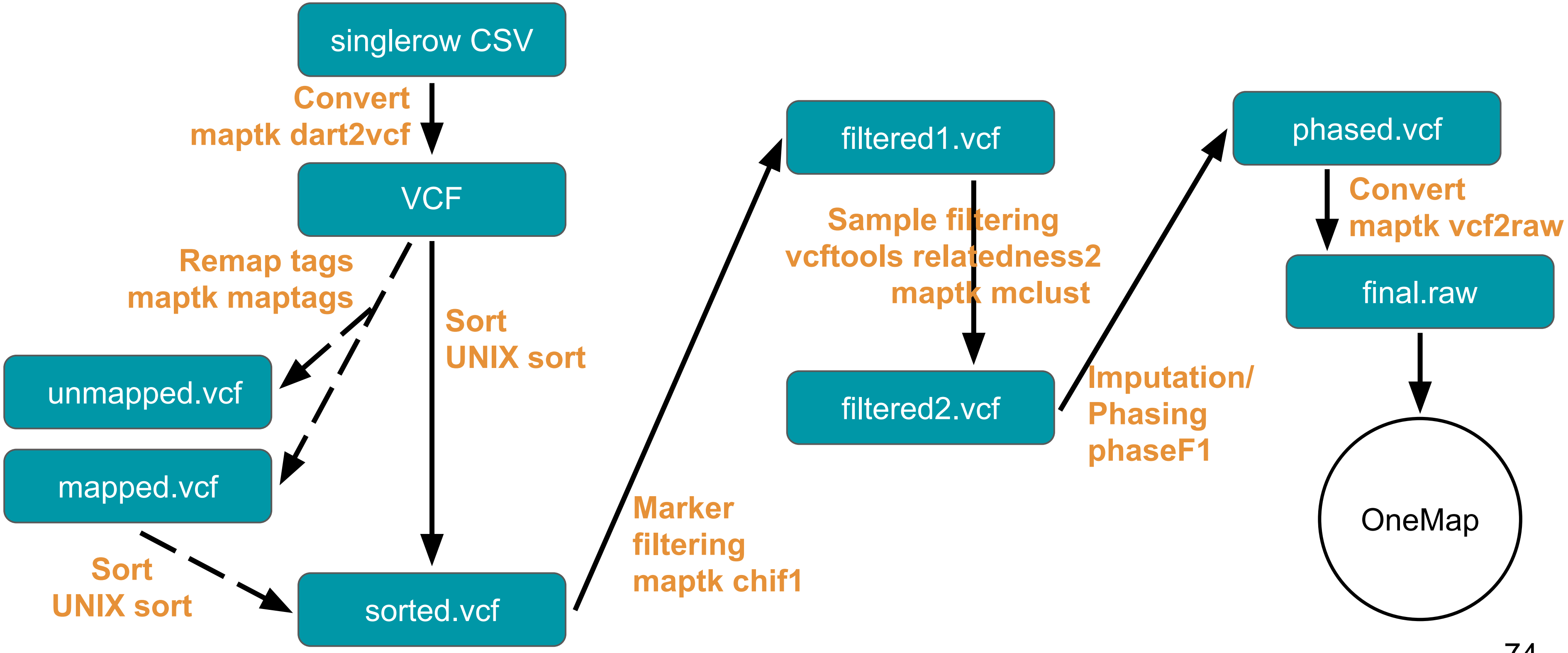
maptk vcf2raw



jess — lyons@ip-172-31-31-27:~ — ssh -i ~/Dropbox/DArtseq_to_mapping_workshop/lyons.pem lyons@ec2-13-57-194-80.us-west-1.compute.amazona...

```
data type outcross
146 348 0 0 0
F1-0A01 F1-01A02        F1-01A03        F1-01A04        F1-01A05        F1-01A06        F1-01A07        F1-01A08        F1-01A09        F1-01A
*000020132|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab a a a a a ab ab a ab a ab a ab ab a a ab ab ab a a a a ab
*000020135|F|0-37       D1.10   ab ab ab a a ab a a ab ab ab a a ab a a a ab a a ab ab ab a ab ab ab a ab a a ab a ab ab a a a ab a ab a ab ab
*000020528|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab a ab a a ab ab a a a ab ab ab a
*000020529|F|0-37       B3.7    b ab b a a b a ab ab b ab a ab ab ab ab ab ab ab a b ab ab a ab ab b ab ab ab a b a b b a a ab b ab ab a ab ab
*000020286|F|0-37       B3.7    a ab a b b a b ab ab a ab b ab ab ab ab ab ab ab b a ab ab b ab ab a ab ab ab b a b a a b b ab a ab ab b ab ab
*000020287|F|0-37       B3.7    a ab a b b a b ab ab a ab b ab ab ab ab ab ab ab b a ab ab b ab ab a ab ab ab b a b a a b b ab a ab ab b ab ab
*000020266|F|0-37       D1.10   ab ab ab a a ab a a ab ab ab a a ab a a a ab a a ab ab ab a ab ab ab a ab a a ab a ab ab a a a ab a ab a ab ab
*000020337|F|0-37       D1.10   a a a ab ab a ab ab a a a ab ab a ab ab ab a ab ab a a a ab a a a ab a ab ab a ab a a ab ab ab a ab a ab a a a
*000020338|F|0-37       D1.10   a a a ab ab a ab ab a a a ab ab a ab ab ab a ab ab a a a ab a a a ab a ab ab a ab a a ab ab ab a ab a ab a a a
*000020339|F|0-37       D1.10   a a a ab ab a ab ab a a a ab ab a ab ab ab a ab ab a a a ab a a a ab a ab ab a ab a a ab ab ab a ab a ab a a a
*000020341|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab a ab a a ab ab a a a ab ab ab a
*000020343|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab a a a a a ab ab a ab a ab a ab ab a a ab ab ab a a a a ab
*000020345|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab a ab a a ab ab a a a ab ab ab a
*000020346|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab a ab a a ab ab a a a ab ab ab a
*000020347|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab a a a a a ab ab a ab a ab a ab ab a a ab ab ab a a a a ab
*000020348|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab a a a a a ab ab a ab a ab a ab ab a a ab ab ab a a a a ab
*000020349|F|0-37       D1.10   a ab a ab ab a ab ab a a a ab ab a ab ab ab a ab ab a a a ab a a a ab a ab ab a ab a a ab ab ab a ab a ab a a
*000020478|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab a ab a a ab ab a a a ab ab ab a
*000020535|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab a a a a a ab ab a ab a a a ab ab a a ab ab ab a a a a ab
*000020536|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab
*000020537|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab
*000020538|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab
*000020539|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab
*000020540|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab
*000020541|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab a a a a a ab ab a ab a a a ab ab a a ab ab ab a a a a ab
*000020543|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a ab ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab
*000020544|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a a ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab a
*000020545|F|0-37       D2.15   a ab a ab ab a ab a ab a ab ab a ab a a a ab a ab a a ab ab ab ab a a ab a ab ab ab a a ab ab a a a ab ab ab a
*000020444|F|0-37       B3.7    b a b a a b a ab ab b ab a ab ab ab ab ab ab ab a b b b a ab ab b ab ab ab a ab a b b a a ab b ab ab a ab ab b
*000020443|F|0-37       B3.7    b a b a a b a ab ab b ab a ab ab ab ab ab ab ab a b b b a ab ab b ab ab ab a ab a b b a a ab b ab ab a ab ab b
*000020442|F|0-37       D2.15   ab a ab a a ab a ab a ab a a ab a ab ab ab a ab a ab ab ab a a a ab ab a ab a a a ab ab a a ab ab ab a a a a a
:
```

DArTseq to mapping: analysis flow chart



singlerow CSV

Convert
maptk dart2vcf

VCF

Remap tags
maptk maptags

Sort
UNIX sort

unmapped.vcf

mapped.vcf

Sort
UNIX sort

sorted.vcf

Marker
filtering
maptk chif1

filtered1.vcf

Sample filtering
vcftools relatedness2
maptk mclust

filtered2.vcf

Imputation/
Phasing
phaseF1

phased.vcf

Convert
maptk vcf2raw

final.raw

OneMap

RRRRR matey

R statistical programming

Very good linear algebraic support (vectors, matrices)

Publication-quality figures

Interactive interpreter and Rscript

Portable on all operating systems

Goals: be able to use R/OneMap and understand what you are doing

Teaching yourself R

Comments:

```
# Write yourself lots of notes
```

Learn more about a function/package:

```
?thing
```

Rstudio, a GUI for R:

https://www.rstudio.com

R: Variables, Math, Logic

Valid variable names may contain A-Z, a-z, 0-9 characters, as well as “_” and “.”
(Cannot begin with a number)

Assignment:

`x = 5`

`x <- 5`

Comparison:

Equivalent to: `x == y`

Not equivalent: `x != y`

Less-than: `x < y`

Less-than-or-equal-to: `x <= y`

Greater-than: `x > y`

Greater-than-or-equal-to: `x >= y`

Logic:

And: x & `y`

Or: `x | y`

Negate: `!x`

Math operations:

Addition: `x + y`

Subtraction: `x - y`

Multiplication: `x * y`

Division: `x / y`

Power: `x^y or x**y`

Modulo: `x %% y`

Log: `log(x)`

Square root: `sqrt(x)`

R: Objects

Variables: are “objects” (containers for data):

```
object = data
```

Belong to a class/have “type” (character, numeric, vector, factor, etc.):

```
class(object)
```

May contain attributes:

```
names(object) # what attributes are there?
object$attribute  # use the attribute as a variable
```

variable types and coercion

NULL: `NULL`

numeric: `-1, 0, 1, -1.0, 0.05, 1e-6`

character: `“string”, “1”, “FALSE”`

vector: `1D sequence`

```
  V = c(1, 5, 10);
  value = V[index];
  V[index] = value
```

matrix: 2D table (typically numeric)

```
  M = matrix(1:9, nrow=3, ncol=3)
  value = M[i,j]
  M[i,j] = value
```

data.frame: Excel-like table

```
   D = data.frame(field1=1:9, field2=11:19)
  value = D[i,j]; D[i,j] = value
  value = D$field1; D$field1 = value
```

logical: `TRUE, FALSE, NA`

list: Multi-D table

```
  L = list(field1=1:9, field2=11:19)
  value = L[i,j]; L[i,j] = value
  value = L$field1; L$field1 = value
  L = list(1:9, 11:19)
  value = L[[i]][j]; L[[i]][j] = value
```

Testing for type:

`is.`***`type`***`(object)`

Coerce to type:

`new.type = as.`***`type`***`(old.type)`

Conditional statements

Template:

```
if (test1) {
    execute this line

} else if (test2) {
    execute this line

} else { # defaults to this
    execute this line
}
```

Template:

```
if (test1) {
    if (test2) {
        execute this line
    }
    execute this line

} else { # defaults to this
    execute this line
}
```

Conditional statements

Example:

```
i = 0
if (i < 10) {
    print(“< 10”)

} else { # defaults to this
    print(“>= 10”)
}
```

Example:

```
i = 5
if ((0 <= i) & (i <= 5)) {
    print(“in range [0, 5]”)

} else if (i < 0){
    print(“negative”)


} else {
    print(“greater than 5”)
}
```

for and while

Template:

```
for (variable in sequence) {
    execute this code
}


while (condition-is-true) {
    execute this code
}
```

Example:

```
for (i in 1:10) {
    print(i)
}

j = 0
while (j < 15) {
    print(j)
    j = j + 1
}
```

functions

Template:

```
f = function(x=default) {
    compute something
    return(result)
}

y = f(x)
```

Example:

```
f = function(x, add=FALSE) {
    if (add) {
        return(x + 1)
    } else {
        return(x - 1)
    }
}

y = f(x)
```

Other useful functions

`plot`	Plot X, Y data	`read.csv`	Read CSV file
`hist`	Plot a histogram	`write.csv`	Write CSV file
`seq`	Create a sequence	`length`	Calc. length of object
`pdf`	Open a PDF device for writing	`names`	Get attribute names
`dev.off`	Close device to save	`append`	Append item to vector
`paste`	Paste together data to make a string	`setwd`	Set working dir
`sort`	Sort a vector, list, etc.	`getwd`	Get working dir
`read.table`	Read in a space-/tab-delimited file	`ls`	Show defined variables
`write.table`	Write a table to file	`summary`	Summarize object data
`install.packages`	Install R packages from CRAN	`lapply`	Iterate and create list

R exercises

1. Use R to calculate the `sum` of 5 and 8 and assign the output to a variable `x`.
2. Use R to determine if `x` is divisible by 2 (hint: see modulo operator).
3. Use `is.integer` type test to determine whether `x` is an integer. If not, what type is it?
4. Use a conditional statement to test whether `x` is equivalent to 12 AND 13
5. Use a conditional statement to test whether `x` is equivalent to 12 OR 13
6. Use the `seq` function to output even numbers between 0 and 50
7. Combine your solution to question 6 with a `for` loop to print odd numbers.
8. Create a function to calculate the difference between two numbers.
9. Use `data(cars)` to load some example data. The `cars` variable becomes defined. What columns does it contain? Plot a histogram of the first column. What is the `mean`? Use `sd` to calculate the standard deviation.

R exercises

10. To save a plot to PDF, you must first call `pdf(“name_of.pdf”)`, then the plotting function of interest, then `dev.off()` to close the file. Save a PDF of your histogram from question 9.

Genetic linkage map estimation

Linkage mapping with OneMap

Why OneMap?

- Free!
- Supports many different cross types (F1, F2, RIL, BC)
- Parallelizable
- Very flexible (not only one analysis)
- Portable to any system

Loading the OneMap package and reading data

To load the OneMap package, use `require()`

```
require(onemap)
```

When you load mapping data, cross type is read from the file

```
cross = read_onemap(“/path/to/dir”, “F1.raw”)
```

Checking your data quality

Plots color genotype matrix

```
plot(cross)
```

Was our chi-squared test stringent enough?

```
plot(test_segregation(cross))
```

OneMap can give you a recommended LOD score

```
lod.suggested = suggest_lod(cross)
```

$$LOD = Z = \log_{10} \frac{\text{probability of birth sequence with a given linkage value}}{\text{probability of birth sequence with no linkage}} = \log_{10} \frac{(1-\theta)^{NR} \times \theta^{R}}{0.5^{(NR+R)}}$$

https://en.wikipedia.org/wiki/Genetic_linkage