Genetic Data Analysis of Rheumatoid arthritis and Clouston's disease

Module: M1 Méthodologie de traitement de données NGS

Authors

Ngoc VU, Denys BURYI

Introduction

We have carried out linkage analyses (lod-score or affected sib-pairs) and association analyses (case-control or familial) for a monogenic disease, then for a multifactorial disease:

1. Study of a monogenic disease: Clouston's disease

    (a) Genetic linkage analysis (lodscore)

    (b) Familial association analysis

2. Study of a multifactorial disease: rheumatoid arthritis

    (a) Genetic linkage analysis (affected sib-pairs)

    (b) Genome-Wide case-control association analysis

The data and programs were provided by our professor VALÉRIE CHAUDRU.

1 Monogenic disease: Clouston's disease

Clouston syndrome is mainly characterized by abnormal nails (thickened, slow-growing, brittle) and alopecia (hair loss). *Palmoplantar hyperkeratosis* (commonly known as horniness) is also sometimes observed. This is a rare disease, with a prevalence ranging from 1 to 9 in 100,000. It is autosomal dominant. Penetrance of the disease is complete, but its expression is variable, even in patients from the same family. Our aim in this part of the analysis was to:

- Perform linkage analysis using the lodscore method to locate the Clouston disease gene using genetic markers located on chromosome 13.

- Interpret data from a `.vcf` file containing genetic variants identified in a case-control sample and perform association analysis for a candidate variant.

- Search for the causal genetic variant using familial association analysis.

1.1 Genetic linkage analysis - lodscore method

We have analyzed the data contained in the `fam.txt` file using `paramlink` package in R. First we performed some initial data exploration using `linkdat()` and `summary()` functions:

```
1    fam = read.table('fam.txt')
2    x = linkdat(fam)
3    summary(x)
```

We found that the file contains data of a large family of 47 individuals, comprising 22 affected and 25 unaffected members. These individuals are organized into 10 nuclear subfamilies and have been genotyped for 13 genetic markers on chromosome 13. For 11 individuals in the dataset no parent data was available (further reffered to as "founders"). Figure 1 shows the pedigree plot of the family.
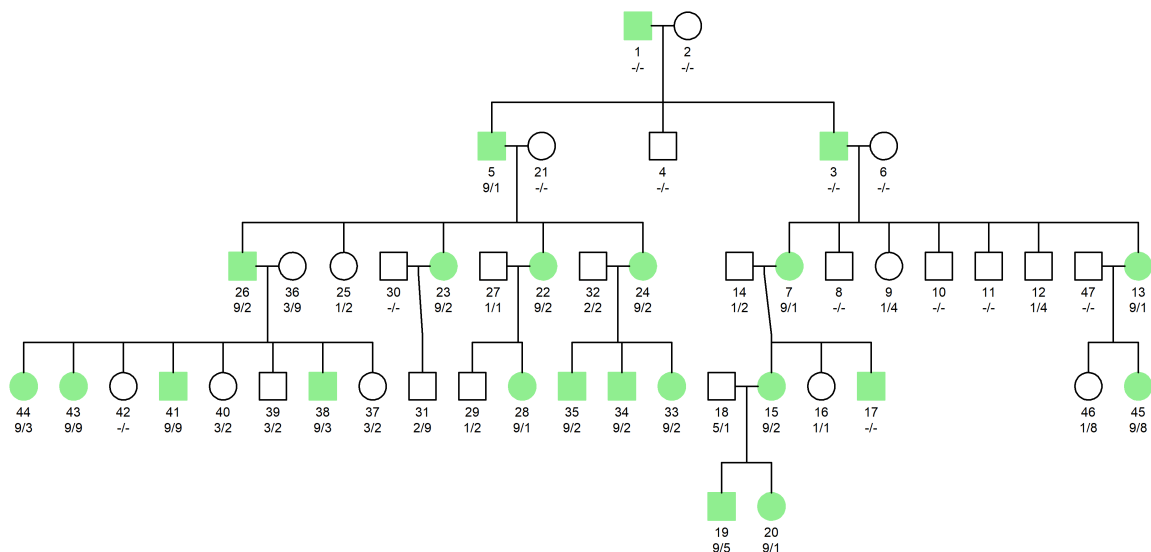

Figure 1: Pedigree plot of the family along with their affection status and genotypes with regards to marker 1. Affected - light green, Unaffected - white. Sex: square/circle - M/F. Genotype: alleles of the marker 1.

In addition to helping summarize the data, the `linkdat()` function also transforms the original data and outputs a `linkdat` object that can be used to perform linkage analysis. `linkdat` object is a list containing among others, the following important components:

- `pedigree` - data.frame with 5 columns (ID, FID, MID, SEX, AFF) describing the pedigree in linkage format.

- `orig.ids` - vector of original IDs of the individuals in the pedigree.

- `subnucs` - list of vectors of IDs of individuals in each nuclear family.

- `markerdata` - a list of `marker` objects describing information about the genetic markers.

- `model` - a list of parameters for the linkage analysis.

2 Multifactorial disease: Rheumatoid Arthritis