What this presentation will do?

In the presentation, I will tell you about…

1. what did I do.
2. what am I going to do.

What did I do?

1. I modularised the code: `DeenaGendoo_Generate_MVA_DNF.R`
2. I added the `Logging.R` feature.
3. I automated the computation of DNF.
4. I read and tried to understand the following codes:
    1. `DeenaGendoo_Generate_MVA_DNF.R` ✔️
    2. `DeenaGendoo_PermutationTestAndFiltering.R` ☑️

What did I modularise?

```
Preprocessing
├── Dependencies.R
├── FunctionsBank.R
└── Logger.R
```

- `Dependencies.R` do all the package loading/installing.
- `FunctionBank.R` contains all the operations function, including the `dnf` generation process: `get_dnf(...)`
- `Logger.R` is created for better debugging.

Let’s talk about the `Logging.R` feature.

I develop it from `log4r`.

```
log4r-package               package:log4r               R Documentation

A simple logging system for R, based on log4j.

Description:

     logr4 provides an object-oriented logging system that uses an API
     roughly equivalent to log4j and its related variants.
```

Why `logging`?

- Retain execution history.
- Faster problem-shooting ➡️ Happier debugging.
- Generate report.

How does it look like?

see the following `*.log` file snapshot.

```
INFO  [2021-07-04 19:19:57] *** [ read_gmt ] { c2.cp.kegg.v7.4.symbols.gmt } loaded succes
        - number of pathways: { 186 }
INFO  [2021-07-04 19:21:58] *** [ read_gmt ] { c2.cp.reactome.v7.4.symbols.gmt } loaded su
        - number of pathways: { 1604 }
DEBUG [2021-07-04 19:19:59] [ drug_sanity_check ] checking:  ncol(sensData) != ncol(pertDa
DEBUG [2021-07-04 19:19:59] ...passed
INFO  [2021-07-04 19:20:04] gmt:  c2.cp.kegg.v7.4.symbols.gmt [ 22 ]:
KEGG_NON_HOMOLOGOUS_END_JOINING < min_num_common_genes { 2 }
INFO  [2021-07-04 19:20:07] gmt:  c2.cp.kegg.v7.4.symbols.gmt [ 26 ]: KEGG_RENIN_ANGIOTENS
```

DNF automation

DNF automation is done by `get_dnf()`, a custom function:

```r
get_dnf <- function(pathway_name, pathway_genes,
                    pertData, sensData, strcData,
                    min_num_common_genes = 2, logger = get_logger("DNF.log", log_lv = "DEE
```

It returns the following `list`.

```r
dnf <- list(
    "pathway" = pathway_name,
    "common_genes" = common_genes,
    "strc_layer" = strcAffMat,
    "sens_layer" = sensAffMat,
    "pert_layer" = pertAffMat,
    "integreated_network" = integrtStrctSensPert
  )
```

- The function depends on `min_num_common_genes` between `pertData and the pathway_genes`
- `min_num_common_genes` is specified by user

⚠️ `min_num_common_genes` must be >= 2 for computing the correlation matrix, otherwise there is **error**.

Implementation 📌

Essentially, It is 2 `for loop`:

1. create a empty list: `DNFs`
2. get all file path of `*.gmt` under `Data/GMT` directory
3. for each `*.gmt`:
    1. load and read the `*.gmt`
    2. for each `pathway` in the `*.gmt`:
        1. `dnf <- get_dnf(pathway ...)`
            1. if `num_common_gene` < `min_num_common_genes`
                1. skip the pathway and continue
        2. add `dnf` to `DNFs`
4. save `DNFs` to `DNFs.RData`
5. generate `DNFs_report.log`.

DNFs report

Setting: minimum number of common genes = 2

```
INFO  [2021-07-05 17:11:08] == DNFs Report ==
        - number of gmt files processed: 2,
        - gmt files: [c2.cp.kegg.v7.4.symbols.gmt, c2.cp.reactome.v7.4.symbols.gmt],
        - minimum number of common genes: 2
        - number of dnfs generated: 1210,
        - number of unconsidered pathways: 580
```

*Realistically though, I would only keep pathways that have a minimum of **5 genes*** *Deena,2021*

Setting: minimum number of common genes = **5**

```
INFO  [2021-07-05 18:34:18] == DNFs Report ==
        - number of gmt files processed: 2,
        - gmt files: [c2.cp.kegg.v7.4.symbols.gmt, c2.cp.reactome.v7.4.symbols.gmt],
        - minimum number of common genes: 5
        - number of dnfs generated: 685,
        - number of unconsidered pathways: 1105
```

what am I going to do?

I will…

- continue reading `DeenaGendoo_PermutationTestAndFiltering.R`
- try to
    - Re-execute permutation testing & z-score calculation
    - Generate top drug hits against query drugs

Thank you for your attention