# Prize: an R package for prioritization estimation based on analytic hierarchy process

Daryanaz Dargahi

May 2, 2019

daryanazdargahi@gmail.com

## Contents

# 1  Licensing

Under the Artistic License, you are free to use and redistribute this software.

# 2  Overview

The high throughput studies often produce large amounts of numerous genes and proteins of interest. While it is difficult to study and validate all of them. In order to narrow down such lists, one approach is to use a series of criteria to rank and prioritize the potential candidates based on how well they meet the research goal. Analytic hierarchy process (AHP) [1] is one of the most popular group decision-making techniques for ranking and prioritizing alternatives when multiple criteria must be considered. It provides a comprehensive and rational framework to address complicated decisions by modeling the problem in a hierarchical structure, showing the relationships of the goal, objectives (criteria and subcriteria), and alternatives. AHP has unique advantages when communication among team members is impeded by their different specializations or perspectives. It also enables decision makers to evaluate decision alternatives when important elements of the decision are difficult to quantify or compare.

The AHP technique uses pairwise comparisons to measure the impact of items on one level of the hierarchy on the next higher level. It has two models for arriving at a ranking of alternatives. (A) The relative model, where alternatives are compared in a pairwise manner regarding their ability to achieve each of the criteria. (B) The rating model is often used when the number of alternatives is large, or if the possibility of adding or deleting alternatives exists [2]. This model requires establishing a series of rating scales (categories) for each criterion. These scales must be pairwise compared to determine the relative importance of each rating category, and then alternatives are evaluated one at a time by selecting the appropriate rating category for each criterion.

Here, we introduce an R package for AHP, "Prize". Prize offers the implementation of both relative and rating AHP models. In order to rank and prioritize a set of alternatives with AHP, decision makers must take four steps:

1. Define the problem and determine the criteria, subcriteria, and alternatives

2. Structure the decision hierarchy

3. Construct pairwise comparison matrices

4. Estimate and visualize priorities

In the following, we describe a brief example use case for Prize in translational oncology.

# 3 Relative AHP

## 3.1 Defining the problem and determining the criteria, subcriteria, and alternatives

Assume a scenario that a group of scientists identified 10 genes that are being differentially expressed (DE) in tumor tissues in comparison to healthy tissues. They are interested in ranking and prioritizing these genes based on their potential role as a tumor marker or therapeutic target. They decide to consider the (1) gene expression profile in tumor tissue, (2) gene expression profile in healthy tissue, (3) frequency of being DE, and (4) epitopes as the criteria for making their decision. They also subdivide the epitope criterion into the size and number of extracellular regions.

## 3.2 Structuring the decision hierarchy
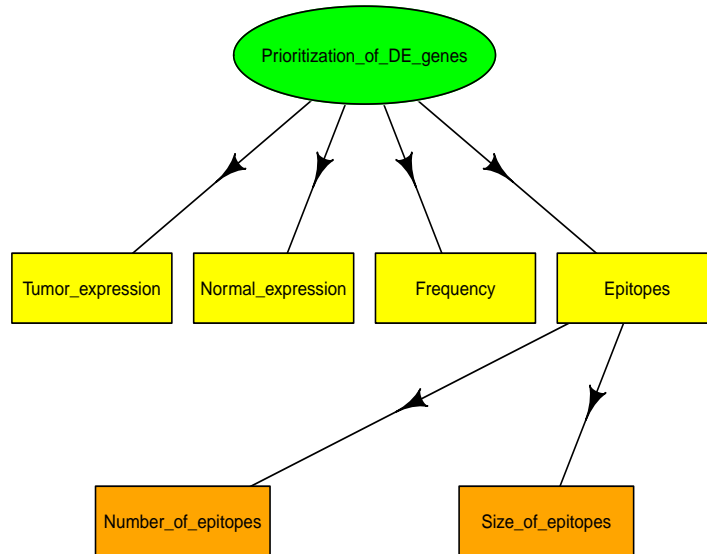
The scientists form their decision hierarchy as follows;

```
> require(Prize)
> require(diagram)

> mat <- matrix(nrow = 7, ncol = 2, data = NA)
> mat[,1] <- c('0', '1','2','3','4','4.1','4.2')
> mat[,2] <- c('Prioritization_of_DE_genes','Tumor_expression','Normal_expression',
+             'Frequency', 'Epitopes', 'Number_of_epitopes', 'Size_of_epitopes')
> mat

      [,1]  [,2]
[1,] "0"   "Prioritization_of_DE_genes"
[2,] "1"   "Tumor_expression"
[3,] "2"   "Normal_expression"
[4,] "3"   "Frequency"
[5,] "4"   "Epitopes"
[6,] "4.1" "Number_of_epitopes"
[7,] "4.2" "Size_of_epitopes"

> ahplot(mat, fontsize = 0.7, cradx = 0.11 ,sradx = 0.12, cirx= 0.18, ciry = 0.07)
```

## 3.3 Constructing pairwise comparison matrices

Each scientist (decision maker) investigates the values of the decision elements in the hierarchy, and incorporates their judgments by performing a pairwise comparison of these elements. Each decision element in the upper level is used to compare the elements of an immediate inferior level of the hierarchy with respect to the former. That is, the alternatives are compared with respect to the subcriteria, the subcriteria are compared with respect to the criteria and the criteria are compared with respect to the goal. Therefore, each decision maker constructs a set of pairwise comparison matrices reflecting how important decision elements are to them with respect to the goal. Pairwise comparison matrices are built from the comparison between elements, based on the Saaty fundamental scale [3].

Table 1: Saaty fundamental scale for pairwise comparison [3]

| Intensity of importance | Definition | Explanation |
|---|---|---|
| 1 | Equal importance | Two elements contribute equally to the objective |
| 3 | Moderate importance | Experience and judgement slightly favor one element over an other |
| 5 | Strong importance | Experience and judgement strongly favor one element over an other |
| 7 | Very strong importance | One element is favored very strongly over an other, its dominance is demonstrated in practice |
| 9 | Extreme importance | The evidence favoring one element over another is of the highest possible order of affirmation |
| Intensities of 2,4,6, and 8 can be used to express intermediate values. Intensitise 1.1, 1.2, 1.3, etc. can be used for elements that are very close in importance | | |

For instance, to pairwise compare the criteria a total of six comparisons must be done, including Tumor expression/Normal expression, Tumor expression/Frequency, Tumor expression/Epitope, Normal expression/Frequency, Normal expression/Epitope, and Frequency/Epitope. The criteria pairwise comparison matrix (PCM) is shown below.

```
> pcm <- read.table(system.file('extdata','ind1.tsv',package = 'Prize'),
+                   sep = '\t', header = TRUE, row.names = 1)
> pcm

                  Tumor_expression Normal_expression Frequency Epitopes
Tumor_expression                 1                 2         4        5
Normal_expression               NA                 1         3        4
Frequency                       NA                NA         1        2
Epitopes                        NA                NA        NA        1
```

The *ahmatrix* function completes a pairwise comparison matrix by converting the triangular matrix into a square matrix, where diagonal values are equal 1 and pcm[j,i] = 1/pcm[i,j].

```
> pcm <- ahmatrix(pcm)
> ahp_matrix(pcm)

                  Tumor_expression Normal_expression Frequency Epitopes
Tumor_expression              1.00         2.0000000       4.0        5
Normal_expression             0.50         1.0000000       3.0        4
Frequency                     0.25         0.3333333       1.0        2
Epitopes                      0.20         0.2500000       0.5        1
```

### 3.3.1 Aggregating individual judgments into a group judgment

Once the individual PCMs are available, *gaggregate* function could be used to combine the opinions of various decision makers into an overall opinion for the group. *gaggregate* offers two aggregation methods including aggregation of individual judgments (AIJ - geometric mean) and aggregation of individual priorities (AIP - using arithmetic mean) [4]. If decision makers have different expertise or perspectives, in order to reflect that in the group judgment, one can use a weighted AIJ or AIP, by simply providing a weight for each decision maker.

```
> mat = matrix(nrow = 4, ncol = 1, data = NA)
> mat[,1] = c(system.file('extdata','ind1.tsv',package = 'Prize'),
+             system.file('extdata','ind2.tsv',package = 'Prize'),
+             system.file('extdata','ind3.tsv',package = 'Prize'),
+             system.file('extdata','ind4.tsv',package = 'Prize'))
> rownames(mat) = c('ind1','ind2','ind3', 'ind4')
> colnames(mat) = c('individual_judgement')
> # non-weighted AIJ
> res = gaggregate(srcfile = mat, method = 'geometric', simulation = 500)

> # aggregated group judgement using non-weighted AIJ
> AIJ(res)

                  Tumor_expression Normal_expression Frequency Epitopes
Tumor_expression         1.0000000         2.0000000 3.4641016 5.143687
Normal_expression        0.5000000         1.0000000 3.6628415 5.383563
Frequency                0.2886751         0.2730121 1.0000000 1.681793
Epitopes                 0.1944131         0.1857506 0.5946036 1.000000

> # consistency ratio of the aggregated group judgement
> GCR(res)

[1] 0.02921679
```
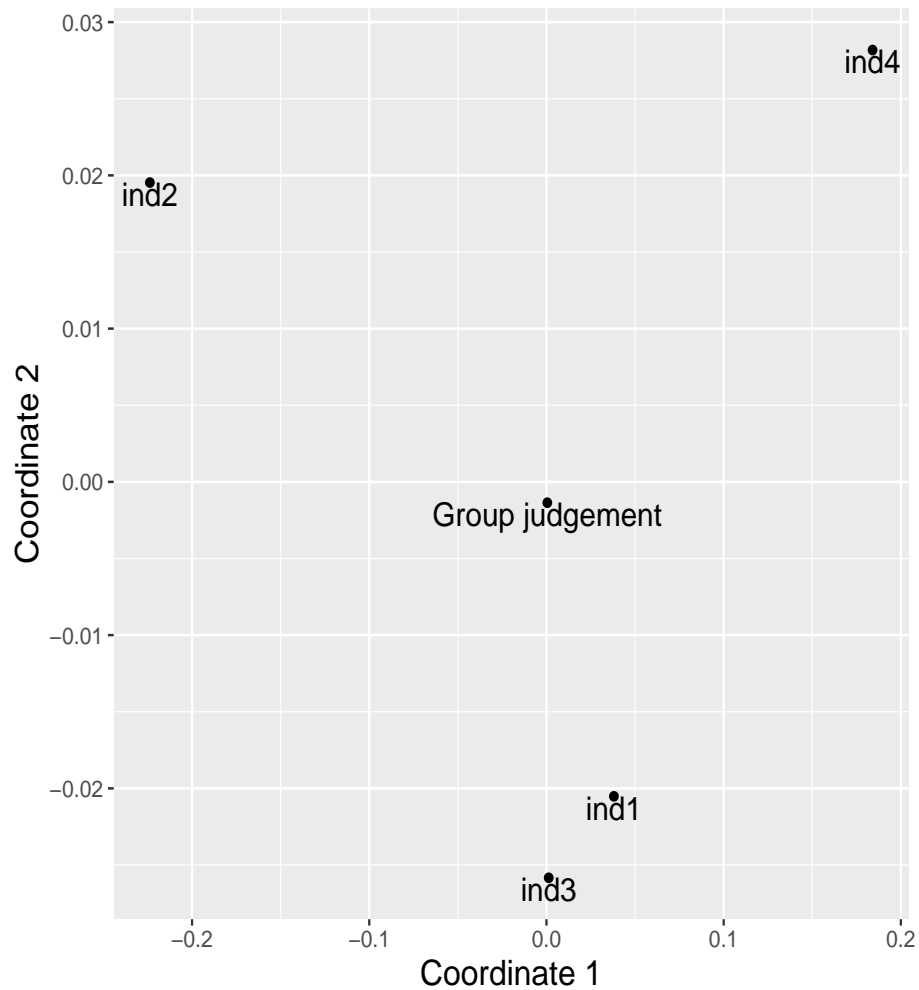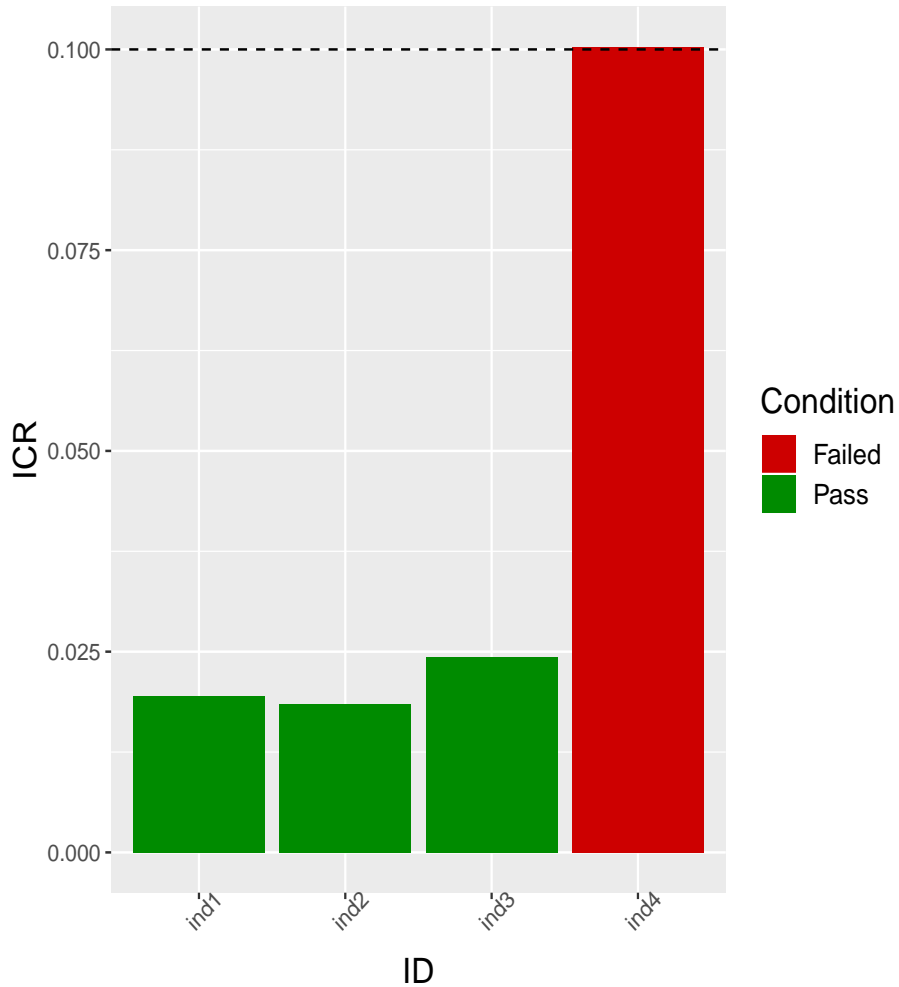
The distance among individual and group judgments can be visualized using the *dplot* function. *dplot* uses a classical multidimensional scaling (MDS) approach [5] to compute the distance among individual and group priorities.

```
> require(ggplot2)

> # Distance between individual opinions and the aggregated group judgement
> dplot(IP(res))
```



The consistency ratio of individual judgments can be visualized using the *crplot* function. If the consistency ratio is equal or smaller than 0.1, then the decision is considered to be consistent.

```
> # Consistency ratio of individal opinions
> crplot(ICR(res), angle = 45)
```

## 3.4 Estimating and visualizing priorities

In order to obtain the priorities of decision elements to generate the final alternatives priorities, local and global priorities are required to be obtained from the comparison matrices. Local priorities are determined by computing the maximum eigenvalue of the PCMs. The local priorities are then used to ponder the priorities of the immediately lower level for each element. The global priorities are obtained by multiplying the local priorities of the elements by the global priority of their above element. The total priorities of the alternatives are found by the addition of alternatives global prioritiese.

The *pipeline* function computes local and global priorities, as well as final prioritization values. *Pipeline* can simply be called by a matrix including the

8

problem hierarchy and group PCMs. The scientists use the following matrix
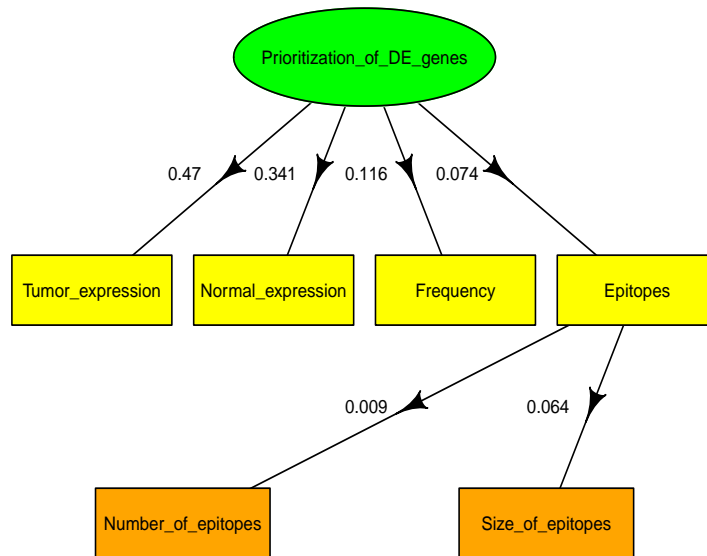(*mat*) to call the *pipeline* function;

```
> require(stringr)

> mat <- matrix(nrow = 7, ncol = 3, data = NA)
> mat[,1] <- c('0', '1','2','3','4','4.1','4.2')
> mat[,2] <- c('Prioritization_of_DE_genes','Tumor_expression','Normal_expression',
+               'Frequency', 'Epitopes', 'Number_of_epitopes', 'Size_of_epitopes')
> mat[,3] <- c(system.file('extdata','aggreg.judgement.tsv',package = 'Prize'),
+               system.file('extdata','tumor.PCM.tsv',package = 'Prize'),
+               system.file('extdata','normal.PCM.tsv',package = 'Prize'),
+               system.file('extdata','freq.PCM.tsv',package = 'Prize'),
+               system.file('extdata','epitope.PCM.tsv',package = 'Prize'),
+               system.file('extdata','epitopeNum.PCM.tsv',package = 'Prize'),
+               system.file('extdata','epitopeLength.PCM.tsv',package = 'Prize'))

> # Computing alternatives priorities
> prioritization <- pipeline(mat, model = 'relative', simulation = 500)
```

The global priorities of decision elements can be visualized using the *ahplot*
function.

```
> ahplot(ahp_plot(prioritization), fontsize = 0.7, cradx = 0.11 ,sradx = 0.12,
+         cirx= 0.18, ciry = 0.07, dist = 0.06)
```
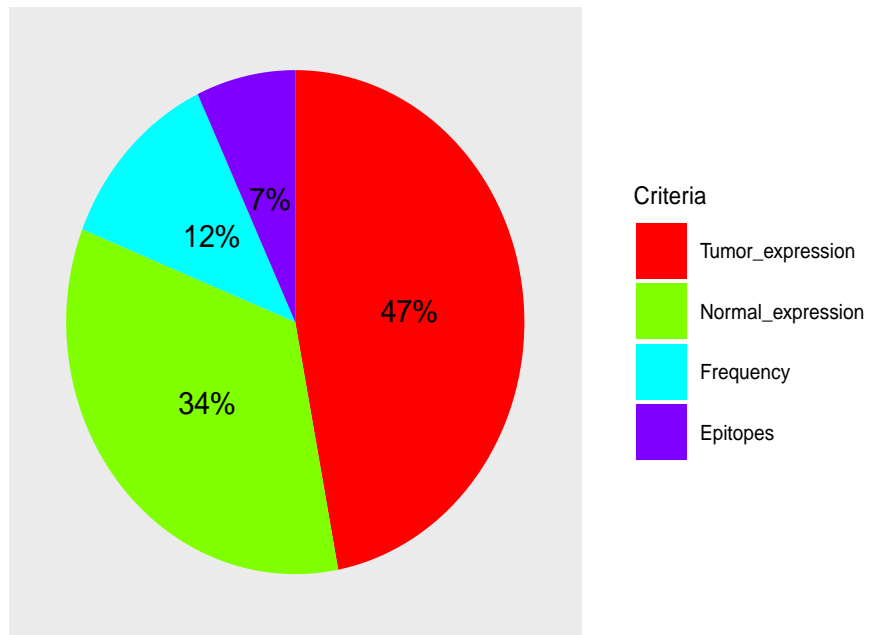
Contribution of decision elements in the final priority estimation could also be visualized using *wplot*.
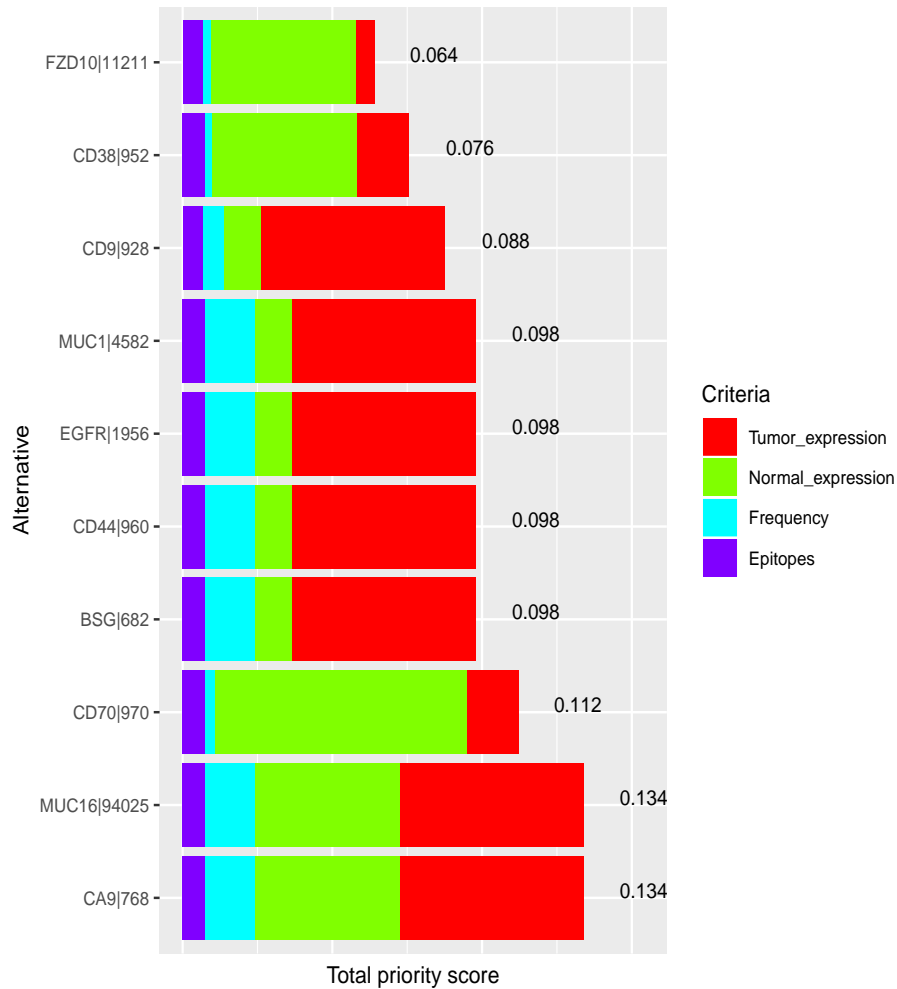
```
> require(reshape2)

> wplot(weight_plot(prioritization)$criteria_wplot, type = 'pie',
+       fontsize = 7, pcex = 3)
```

The *rainbow* function illustrates prioritized alternatives detailing the contribution of each criterion in the final priority score.

```
> rainbowplot(rainbow_plot(prioritization)$criteria_rainbowplot, xcex = 3)
```

The Carbonic anhydrase 9 (CA9) and Mucin-16 (MUC16) with a global priority of 0.134 are the alternative that contribute the most to the goal of choosing the optimal tumor marker/therapeutic target among the identified DE genes. Drugs targeting CA9 and MUC16 are currently in pre-clinical and clinical studies [6, 7].

```
> rainbow_plot(prioritization)$criteria_rainbowplot
```

|  | Tumor_expression | Normal_expression | Frequency | Epitopes |
|---|---|---|---|---|
| BSG\|682 | 0.061335201 | 0.01248896 | 0.016786400 | 0.007461660 |
| CD44\|960 | 0.061335201 | 0.01248896 | 0.016786400 | 0.007461660 |
| CD38\|952 | 0.017114410 | 0.04853918 | 0.002480381 | 0.007461660 |
| CA9\|768 | 0.061335201 | 0.04853918 | 0.016786400 | 0.007461660 |
| MUC16\|94025 | 0.061335201 | 0.04853918 | 0.016786400 | 0.007461660 |
| CD9\|928 | 0.061335201 | 0.01248896 | 0.006816627 | 0.006950588 |

```
EGFR|1956          0.061335201          0.01248896 0.016786400 0.007461660
CD70|970           0.017114410          0.08407232 0.003506147 0.007461660
MUC1|4582          0.061335201          0.01248896 0.016786400 0.007461660
FZD10|11211        0.006154539          0.04853918 0.002480381 0.006950588
                   total_priorities
BSG|682            0.09807222
CD44|960           0.09807222
CD38|952           0.07559563
CA9|768            0.13412244
MUC16|94025        0.13412244
CD9|928            0.08759138
EGFR|1956          0.09807222
CD70|970           0.11215454
MUC1|4582          0.09807222
FZD10|11211        0.06412469
```

# 4 Rating AHP

As the number of alternatives increase, the amount of pairwise comparison becomes large. Therefore, pairwise comparisons take much time and also the possibility of inconsistency in the comparisons increases. Rating AHP overcomes this problem by categorizing the criteria and/or subcriteria in order to classify alternatives. In another words, rating AHP uses a set of categories that serves as a base to evaluate the performance of the alternatives in terms of each criterion and/or subcriterion. The rating procedure is also suitable when the possibility of adding/removing alternatives exists. The rating AHP reduces the number of judgments that decision makers are required to make.

The rating AHP differs from the relative AHP in the evaluation and obtaining the priority of alternatives. Hence, the decision markers define their decision problem, structure the problem into a hierarchy, and collect PCM matrices for each criteria/subcriteria similar to the relative AHP approach. Then, they use a rating approach to evaluate alternatives.

## 4.1 Defining a rating scale and obtaining alternatives priorities

In the example scenario, the scientists would like to rank and prioritize 10 genes based on their potential role as a tumor marker/therapeutic target. To build a PCM matrix consisting of 10 alternatives 45 pairwise comparisons are required. The large number of pairwise comparisons makes this step time consuming and increase the possibility of inconsistency in the comparisons. Therefore, scientists decide to use rating AHP by defining a series of categories with respect to the criteria and/or subcriteria to evaluate alternatives. They also compute a PCM of these categories. For instance, they define two categories, single and multiple, for the $number of epitopes$ subcriteria, and compute their PCM.

```
> category_pcm = read.table(system.file('extdata','number.tsv', package = 'Prize')
+                          , sep = '\t', header = TRUE, row.names = 1)
> category_pcm

         Single Multiple
Single        1        2
Multiple     NA        1
```

Then, decision makers evaluate the alternatives against the defined categories and build an alternative matrix showing the category that each alternative belongs to.

```
> alt_mat = read.table(system.file('extdata','numEpitope_alternative_category.tsv',
+                      package = 'Prize'), sep = '\t', header = FALSE)
> alt_mat

             V1       V2
1      BSG|682   Single
2     CD44|960   Single
3     CD38|952   Single
4      CA9|768   Single
5   MUC16|94025  Single
6      CD9|928 Multiple
7    EGFR|1956   Single
8     CD70|970   Single
9     MUC1|4582   Single
10 FZD10|11211 Multiple
```

To compute the idealised priorities of alternatives, the *rating* functions can be called by a category PCM and an alternative matrix.

```
> result = rating(category_pcm, alt_mat, simulation = 500)
> # rated alternatives
> RM(result)

              scale_category idealised_priorities
BSG|682       "Single"       "1"
CD44|960      "Single"       "1"
CD38|952      "Single"       "1"
CA9|768       "Single"       "1"
MUC16|94025   "Single"       "1"
CD9|928       "Multiple"     "0.5"
EGFR|1956     "Single"       "1"
CD70|970      "Single"       "1"
MUC1|4582     "Single"       "1"
FZD10|11211   "Multiple"     "0.5"
```

The matrix of idealised priorities (rated alternatives) can be used to call *pipeline* function to estimate final priorities of alternatives.

```
> mat <- matrix(nrow = 7, ncol = 3, data = NA)
> mat[,1] <- c('0', '1','2','3','4','4.1','4.2')
> mat[,2] <- c('Prioritization_of_DE_genes','Tumor_expression','Normal_expression',
+               'Frequency', 'Epitopes', 'Number_of_epitopes', 'Size_of_epitopes')
> mat[,3] <- c(system.file('extdata','aggreg.judgement.tsv',package = 'Prize'),
+               system.file('extdata','tumor_exp_rating.tsv',package = 'Prize'),
+               system.file('extdata','normal_exp_rating.tsv',package = 'Prize'),
+               system.file('extdata','freq_exp_rating.tsv',package = 'Prize'),
+               system.file('extdata','epitope.PCM.tsv',package = 'Prize'),
+               system.file('extdata','epitope_num_rating.tsv',package = 'Prize'),
+               system.file('extdata','epitope_size_rating.tsv',package = 'Prize'))
> # Computing alternatives priorities
> prioritization <- pipeline(mat, model = 'rating', simulation = 500)
```

# References

[1] T.L. Saaty. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234 – 281, 1977.

[2] T.L. Saaty. Rank from comparisons and from ratings in the analytic hierarchy/network processes. *European Journal of Operational Research*, 168(2):557–570, January 2006.

[3] T.L. Saaty. *The Analytic Hierarchy Process, Planning, Piority Setting, Resource Allocation*. McGraw-Hill, New york, 1980.

[4] E. Forman and K. Peniwati. Aggregating individual judgments and priorities with the analytic hierarchy process. *European Journal of Operational Research*, 108(1):165 – 169, 1998.

[5] J.C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3/4):325 – 338, 1966.

[6] M. Felder, A. Kapur, J. Gonzalez-Bosquet, S. Horibata, J. Heintz, R. Albrecht, L. Fass, J. Kaur, K. Hu, H. Shojaei, R. J. Whelan, and M. S. Patankar. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Mol. Cancer*, 13:129, 2014.

[7] P. C. McDonald, J. Y. Winum, C. T. Supuran, and S. Dedhar. Recent developments in targeting carbonic anhydrase IX for cancer therapeutics. *Oncotarget*, 3(1):84–97, Jan 2012.