



**HAL**  
open science

## From Oncogenetic Pedigrees to Family Profiles: A Necessary Step to Enable Statistics

M Arbre, F. Kwiatkowski, L Serlet, Yj Bignon

► **To cite this version:**

M Arbre, F. Kwiatkowski, L Serlet, Yj Bignon. From Oncogenetic Pedigrees to Family Profiles: A Necessary Step to Enable Statistics. *Journal of Proteomics and Bioinformatics*, 2016. hal-01675318

**HAL Id: hal-01675318**

**<https://uca.hal.science/hal-01675318>**

Submitted on 4 Jan 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Research Article

### From Oncogenetic Pedigrees to Family Profiles: A Necessary Step to Enable Statistics

Arbre M<sup>1</sup>, Kwiatkowski F<sup>1,2\*</sup>, Serlet L<sup>2</sup>, Bignon YJ<sup>1\*</sup>

<sup>1</sup>Centre Jean Perrin, Laboratoire d'Oncologie Moléculaire, 58, rue Montalembert, 63011 Clermont-Ferrand, France

<sup>2</sup>Université Blaise Pascal, 24, avenue des Landais, BP 80026, 63171 Aubière Cedex, France

\*Corresponding author: Dr. Kwiatkowski F, Centre Jean Perrin, Laboratoire d'Oncologie Moléculaire, 58, rue Montalembert, 63011 Clermont-Ferrand, France, Email: Fabrice.Kwiatkowski@CJPF.fr

Received: 12-15-2015

Accepted: 05-08-2016

Published:

Copyright: © 2016 Kwiatkowski

**Background:** Cancer has always been a major domain requiring progress in statistics, methodology and bio-informatics. Oncogenetic, focusing on the relationship between genetics and cancer, is particularly concerned with “big data” issues, which includes genealogical pedigrees: their special structure – made of relations between members and possible clinical annotations - is too complex to be directly used for statistical purpose. This article describes a way to condense pedigrees so that they can be handled more easily and compared together.

**Method:** our approach aggregates the genealogical and clinical information of pedigrees containing many generations. Condensed pedigrees, called “subtrees”, are composed of basic 2 or 3-generation pedigrees: for one whole pedigree, a subtree is calculated by the mean of all basic pedigrees it contains. These subtrees can then be grouped together for different subsets of families (for example breast/ovarian cancer families with or without BRCA mutation carrier). Such a grouping named “profile”, besides its reduced structure, is particularly interesting because for each studied characteristic, means and standard deviations are available. Moreover, distances between each subtree and various profiles can be calculated and used as a discriminant index.

**Results:** Subtrees and profiles were validated using a subset of 454 families (22.348 members) with a Lynch syndrome: in 84, at least one member carried an MMR deleterious mutation. Two profiles were computed depending on the presence or the absence of MMR mutation in the families. An ROC analysis showed that distances between each family subtree and both profiles were significant predictors for MMR mutations.

**Conclusion:** Subtrees and profiles show interesting discriminant properties to study pedigree data. This method seems suitable to search for population differences between monogenic cancer risk models and multigenic ones.

**Keywords:** Pedigree; Oncogenetic; Genealogy; Modeling; Subtree

## Introduction

Currently, with the progress of genetic research, more and more predispositions to hereditary diseases are discovered. As pangenomic analysis (genome-wide screening) cannot be realized routinely - partly for ethical reasons - it is necessary to predict which genes are the most likely to be mutated and then perform targeted genomic analysis.

In the oncogenetic routine, pedigrees are frequently used to diagnose hereditary predispositions. These contain two kinds of data: first the genealogy, i.e. the relations between members from which for example fertility and mortality parameters can be calculated and second, possible clinical information that may characterize the phenotype of a hereditary predisposition to a disease. Overall, both types of information are necessary

for the discovery of new deleterious mutations. Indeed, they enable to isolate pedigrees with special characteristics like the occurrence of typical cancer locations. Once this step is achieved, gene analysis on available DNA samples can be performed with increased chances to point out one or more mutations possibly responsible for the phenotype.

Unfortunately, oncogenetic pedigrees are usually too complex to be analyzed other than visually. To solve this issue, some authors limit their inquiry to smaller pedigrees (only 3 to 4 generations) [1]: although easier, this solution appears in fine deleterious because of the loss of phenotypic information that is spread all over generations. Another way to evaluate cancer or mutational risks of one family is to calculate scores. Concerning breast/ovarian cancer risk, different authors have developed indexes upon pedigrees such as Manchester [2], Eisinger [3] or BRCAPro [4]. This kind of index combines clinical parameters included in pedigrees that, after a logistic regression, have kept a sufficient significance (i.e. for Manchester score, only breast, ovarian, prostatic, pancreatic cancers reported in the family are used). But these methods only concern a reduced set of familial predispositions and limit the analysis to only a small part of pedigree data, mainly the occurrence of cancers within the family. We have extended such a research to uncommon information. It enabled us to conclude that fertility parameters could also help predict these risks [5]. However, much work still remains because first, the indexes available to calculate the familial risk of mutation are only adapted to breast/ovarian cancers and second, because known mutations account for only a minority of cancer predispositions: improving the way to select specific sub-groups of families is thus a necessary step for narrowing the research of new deleterious mutations to a reduced set of genes.

In the literature, if we except indexes, no efficient methodology enables to a group or compare pedigrees. Specialized pedigree software exists but they concern animals [6,7] and they focus on the inbreeding level.

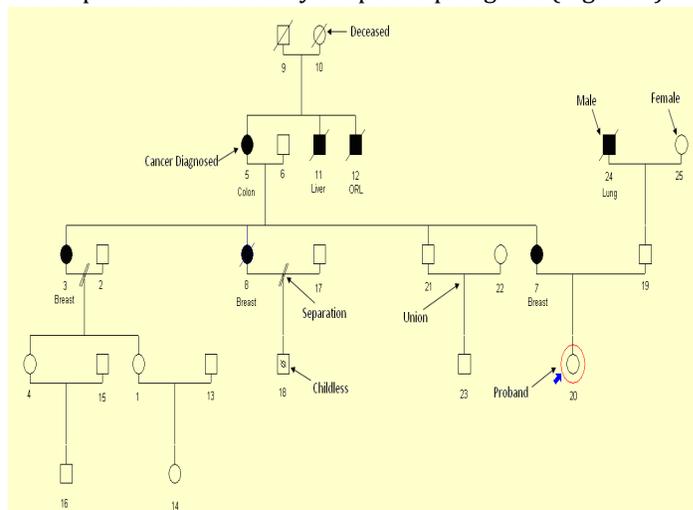
The approach proposed in the next chapter is a modeling of pedigrees into “subtrees”, i.e. family representations condensed into two or three generations by aggregating the information of all family members from all generations. How to create profiles (global subtree for several families together presenting similar characteristics) is the matter of another paragraph. Finally, the use of profiles and subtrees will be demonstrated within a sample of 454 families at colon cancer risk extracted from the database of the Oncogenetic Department of the comprehensive anticancer Center Jean Perrin.

## Methods

### Description of pedigrees

A specialized function of the SEM software [8] has been

developed to automatically shape the pedigrees (Figure 1).



**Figure 1.** Example of a simple pedigree (males are represented by a square and females by a circle. Every striped symbol represents a deceased individual and ones filled with black indicate a cancer).

Most symbols used to draw pedigree are common and have been recommended in the pedigree standardized nomenclature edited by Bennett et al. [9,10]. The «proband» is the person who requests the creation of the medical file. The proband, in the example (Figure 1), is a woman: she is represented by a circle and pointed by a blue arrow. The pedigree is shaped with all the individuals who are related to her. Throughout this article, we will use the family represented in Figure 1 as an example. The Jean Perrin Center database contains families that include sometimes more than 600 members, consequently visual analysis becomes difficult and new representation types are necessary.

### Modeling subtree method

The underlying structure of a pedigree is a reduced 2 or 3-generation pedigree that cumulates the information from all family members. These structures must be distinguished from pedigree branches (for example the paternal and maternal branches of a proband): branches try to isolate members of a pedigree carrying a particular genotype while subtrees gather reduced patterns that occur several times within a pedigree or a branch. Three models are considered: a 2-generation subtree, a detailed 2-generation subtree, and a 3-generation subtree.

#### 2-generation subtree

The way to constitute this 2-generation subtree is to find each pair of [mother or father]/ [son and daughter]. Male and female headers (parents) are separated because men and women are not exposed to the same cancer risk. With these pairs, all the information needed for the construction of the 2-generation subtrees is available in the database and can be collected and aggregated as many times as pairs are available.

From the proband, we can find his/her parents, then the parents of proband's parents, and so forth. Once members at the top of the pedigree have been identified (numbered 9, 10, 24 and 25 in Figure 1), we can browse down the pedigree to keep only genetically related members: children of top members can be selected, then children's children and so on until the most recent generation.

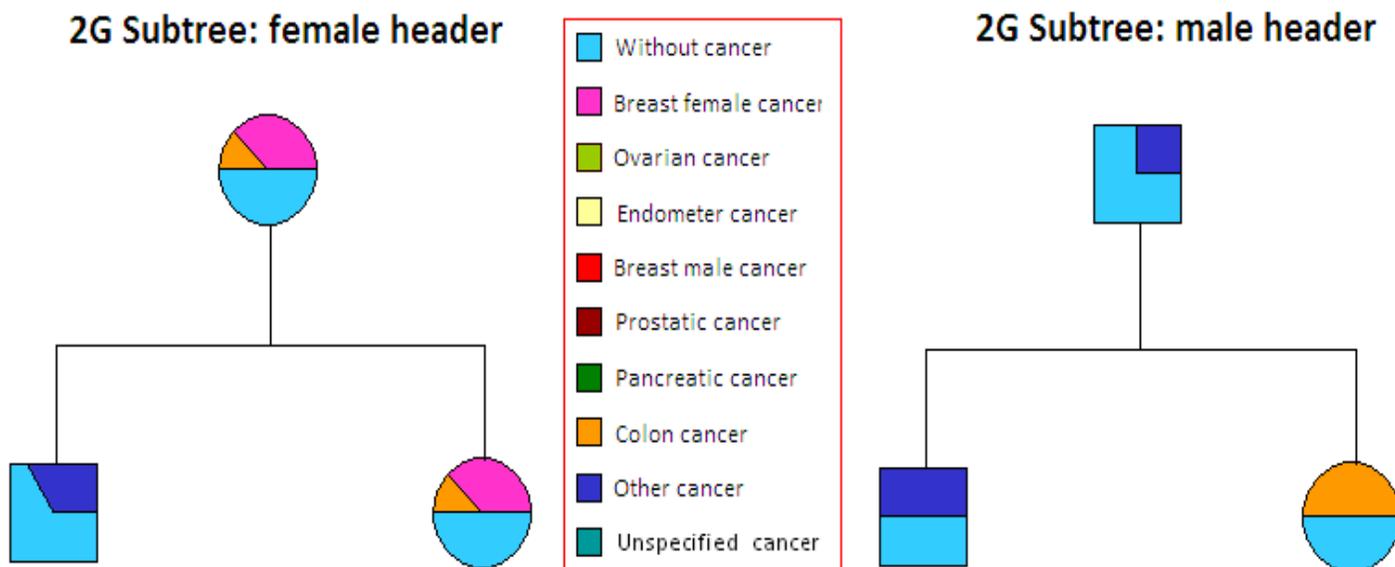
This pruning process excludes a few members who are not supposed to bring genetic information about the cancer risk. Else their presence would "feed the background noise", and increase uselessly the overall variability (i.e. lower the precision of estimates). Pruned members are:

- all spouses who do not provide information about their parents (numbers 2, 6, 13, 15, 17 and 22)
- Childless members (numbers 11, 12 and 18)
- Latest generation members, they are usually too young to have children (numbers 14, 16, 20 and 23)

Finally, each pair "parent/child" can be deduced, keeping in mind that one person can be used as a parent as well as the child.

Once the individuals constituting subtrees are identified, all useful information is collected from clinical data registered in the database. The combination of the information for each item of the 2-generation subtree ends up with 6 composite "family members". They cumulate the following information:

- number of female headers
- number of male headers
- number of males from the female header
- number of females from the female header
- number of males from male header
- number of females from the male header
- number and patient's age at the diagnosis of following cancers:
  - breast male and female
  - ovarian
  - endometer
  - prostatic
  - colon
  - pancreatic
  - other cancers (cumulative)
  - number of members without cancers



**Figure 2.** Basic structure of 2-generation subtree built from the pedigree of Figure 1.

With this selection process, four male headers (numbers 9, 19, 21 and 24) and eight female headers (numbers 1, 3, 4, 5, 7, 8, 10, 25) are identified. Twelve 2-generation subtrees are thus available into the pedigree of figure 1 and a resulting subtree can be built. Basic 2-generation subtree is shaped as in figure 2.

Although this list includes already 104 variables, it can be extended if needed.

With this 2-generation subtree, all the information is condensed whatever the size of the family and it becomes easier to compare 2 or more families. The proportion of cancers by location is represented by a pie chart within circles

and squares at each level (Figure 2):

For the pedigree of figure 1, following characteristics are calculated by the software:

- 8 female headers: 3 breast cancer (occurring in average at 49 years), 1 colon cancer (73 years) diagnosed and 50% without cancer

• 6 male children: 1 liver cancer and 1 ORL cancer, so 2 “other cancer” (70 years) diagnosed and 67% without cancer

• 8 female children: 3 breast cancer (49 years), 1 colon cancer (73 years) diagnosed and 50% without cancer

- 4 male headers: 1 lung cancer (48 years) diagnosed and 75% without cancer

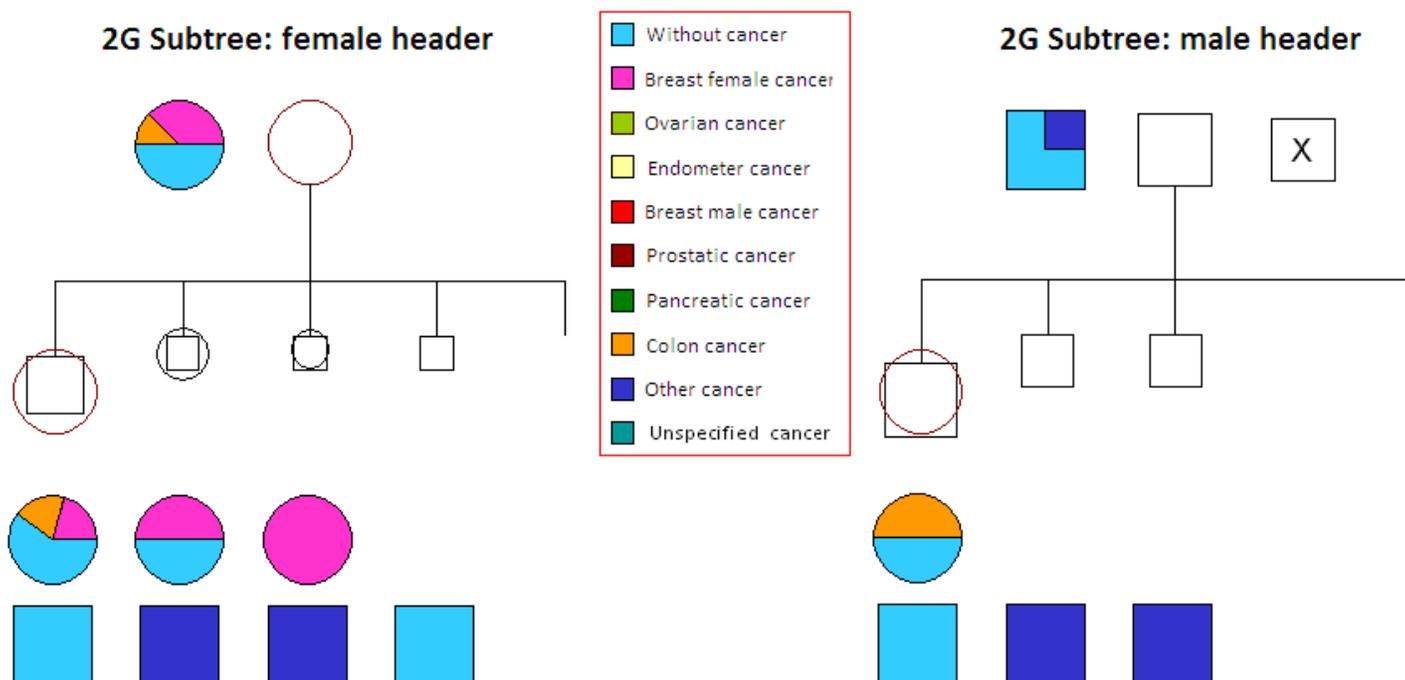
• 4 male children: 1 liver cancer and 1 ORL cancer so 2 “other cancer” (70 years) diagnosed and 50% without cancer

• 2 female children: 1 colon cancer(73 years) diagnosed and 50% without cancer

for most families. Another data is also interesting: childless members and miscarriages, which are also included in this detailed representation.

The member selection process does not differ from the one used for the 2-generation subtree. The same exclusions apply here and headers remain unchanged.

Figure 3 exhibits the basic pattern concerning figure-1 pedigree: circles still represent females and squares males. For children, the 4 first vertical lines correspond to children's rank (born first, born second...) and miscarriages (if any) are positioned at the 5th rank using a lozenge (none in fig. 3). The size of squares and circles is proportional to the number of children for each item. The length of the vertical lines connecting parents to children depends on the parental mean age at their children's birth. Proportions of childless adults have represented aside headers using squares and circles with a cross. Proportions of persons diagnosed with a cancer are represented underneath for the children with the same color-code as in figure 2.



**Figure 3.** Example of a 2-generation subtree with detailed information by children's rank.

## 2-generation subtree with details about children

One might wonder if the birth rank may influence the risk for particular events (example, congenital malformations). This rank is available if dates of birth are known and the number of 1st boys, 1st girls, 2nd boys, etc. can be computed per header. Only four children of each gender are retained, this enables to include a maximum of 8 children which is usually enough

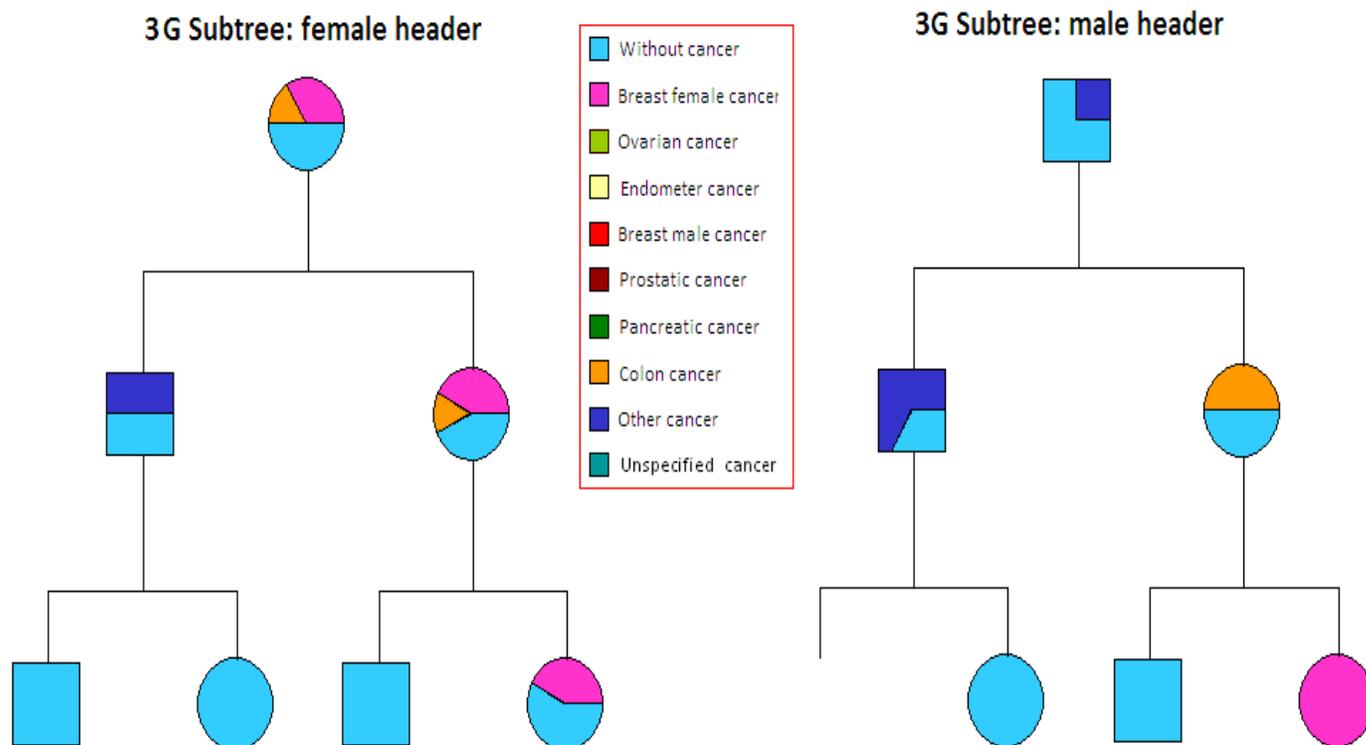
## 3-generation subtree

To highlight possible “variations” of intergenerational cancer transmission, we decided to shape a 3-generation subtree. This 3<sup>rd</sup> synthetic representation includes 3 generations instead of 2: triplets are now identified, with parents/children/grand-children for both genders at each level. The

same process is applied to the members' selection and the same information as previous 2-generation subtrees are collected. Figure 4 shows the basic structure of such a representation drawn for our family example.

### Combining several subtrees to create group profiles

A new concept needs to be introduced if several subtrees are to be grouped together in order to constitute "family profiles"



**Figure 4.** Basic structure of 3-generation subtree from the family of figure 1.

Two recent articles [11,12] have reported that cancers, in mutated families, tended to appear at an earlier age over generations, i.e. daughters had breast cancers sooner than their mothers. Narod [13] suggested this could happen because daughters' exposure time is necessary shorter and late cancers have not enough time to occur. We thus decided to add a correction so that intergenerational comparisons would not be biased by differences of elapsed lifetime (this consideration seems less relevant in 2-generation subtrees because of the number of generations available in our pedigrees).

Correction of the cancer proportions at each level is made according to the average exposure time in "person-year". Following ages are accumulated for each generation:

- the age of living persons without cancers (or age at death)
- the age at cancer diagnosis for other persons

Once the average age is calculated in person-years (P-Y) per generation, the cancer frequency of the Nth generation is multiplied by the ratio:

$$\frac{\text{Average age in P-Y for the N-1 generation}}{\text{Average age in P-Y for the N generation}}$$

or "group profiles". Concerned families are selected when they present particular characteristics. For example, one might want to design a specific profile for BRCA mutated families, another one focusing families with several lung cancers, or in a completely different domain, families where several suicides are reported, and so on.

We have to choose between the three representations of subtrees. The first one, 2-generation subtree, seemed to be the more suitable to create these profiles because both other representations scatter too much the information (i.e. multiply the number of characteristics per subtree and thus diminish the density of parameters).

Three steps are necessary to build such profiles:

- First, families with the chosen characteristics are selected and grouped into a set: this step depends on each software/database that contains the family records.
- At the second step, subtrees are designed for each family of the set, including as many characteristics as needed.
- At the last step, all the information per subtree is combined into a more global object: for this, average and standard de-

viation are calculated for each variable (i.e. personal, familial or clinical characteristic of interest) and registered in a new table of the database. The set of averages and variances per profile corresponds to a multidimensional object which can be represented by a barycenter surrounded by a “cloud of points”.

This group of distribution parameters can then be used to realize statistics, to compare several profiles, to calculate the distance between them and a new subtree and also identify particular families’ subset.

### Statistical considerations

Several statistical tests are used in this study. Distribution parameters (mean and standard deviation) characterize numerical data and numbers / frequencies categorical variables. Best cutoffs optimizing sensitivity and specificity of predictive parameters are calculated using an ROC analysis [14] while the performance of associated ROC curves is evaluated using their area under the curve (AUC) [15]. To build a score predictive for MMR mutation based on standard parameters cited in the literature, a logistic regression model was performed. The corresponding predictive score was calculated using its regression formula.

### Results

An example of the use of subtrees and profiles is detailed hereafter. We first describe some characteristics of two profiles and then, we explain how these profiles can help predict the mutational risk.

#### Description of the family set (454 Lynch syndrome)

The accrual in our pedigree database started in 1988. Today it contains 6,500 families including over 190,000 individuals with clinical information (family diagnosis, mutated gene if any...). Most of these families correspond to a breast/ovarian cancer risk. Another important group represents the Lynch syndrome (or HNPCC = hereditary nonpolyposis colorectal cancer). Less than 20% of families diagnosed with this syndrome will present a mutation in APC gene or one of main MMR genes (MLH1, MSH2, MSH6, PMS2). Thus, even using NGS analyzers, a systematic sequencing of all 5 genes is not relevant. Up to now in Lynch syndrome, no good algorithm developed using pedigree information can predict with a good accuracy the mutation probability [16]. Two main algorithms exist, but they have weak predictive properties: Amsterdam index [17,18] with sensitivity around 80% and specificity of 46% and 68% across studies, and the revised Bethesda index [19] associated with a 89% sensitivity and 58% specificity; to increase the prediction strength of these indexes, two complementary tests can be performed on blood sample: an Immuno-Histo-Chemical test (IHC) and a microsatellite instability test (MSI). They enable to bring up sensitivity and specificity to values close to 100% but they are not cheap. We thus decid-

ed to check if profiles could help us develop a new strategy and enable to avoid the intermediate use of IHC/MSI test.

Two profiles have been calculated among families presenting with a Lynch syndrome. The first profile corresponded to 84 mutated families, with at least one member diagnosed with a deleterious mutation in APC gene or on a “mismatch repair” gene. The second included 370 families without any member diagnosed with such mutations. All families needed also to contain at least 10 known members to be sufficiently informative. Respectively 4,218 and 18,130 individuals belonged to these two profiles.

Means and SD were computed for about 100 parameters, respectively 23 and 21 per “synthetic” subtree mother or father and 14 or 12 per “synthetic” daughter or son (mothers’ daughters, mothers’ sons, fathers’ daughters, fathers’ sons) and 8 familial fertility scores: the number of features differed by gender because some cancers are gender-specific (prostate, ovaries) and fertility parameters are calculated only for subtree headers. Some of these features are presented in table 1 (mainly of the female header).

Obvious differences can be noticed between both profiles, in particular regarding ages at colon cancer diagnosis for both mothers and fathers (table 1). Cancer frequency is also doubled in fathers if a known deleterious mutation is diagnosed in their family.

#### Distance calculation between a profile and a subtree

Profiles enable statistical computations. A first method is to calculate distances (Fig. 5) between profiles and a new family (i.e. a subtree), in order to find the nearest one. Profiles can be represented as a cloud with a barycenter (average) and a width (using standard deviations).

The spreading of the cloud can be figured by a disk and its radius by a double arrow between the center of the cloud and the edge. A new subtree corresponds to a new cloud which standard deviation is null, thus a point. Two kinds of distance were envisaged: Euclidean and correlation coefficient.

#### Several measures are possible:

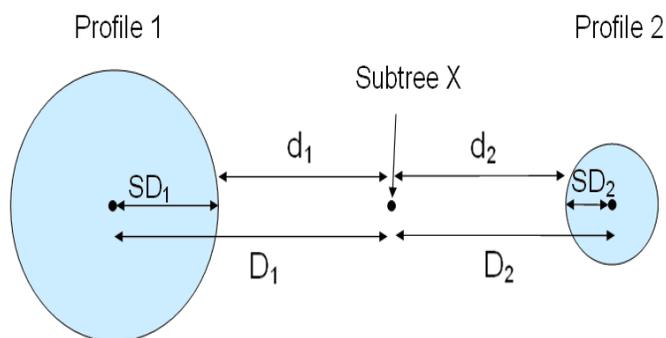
- D = Distance between the center of the cloud and the new family (Figure 5, double arrow between the center of the cloud for profiles 1 and 2 and Subtree X)
- d = distance between the extremity of the cloud and the new family (Figure 5, the double arrow between the extremity of clouds (profiles 1 and 2) and Subtree X)
- R = ratio between the distance D and the associated cloud spreading (Standard Deviation) =  $D / SD$

| Label  | Profile 1 (N = 84)<br>MMR mutated families |                    | Profile 2 (N = 370)<br>Families without MMR mutation |                    |
|--|--|--------------------|--|--------------------|
|  | Mean                                       | Standard Deviation | Mean   | Standard Deviation |
| Header Female (number per subtree)               | 7.94                                       | 5.10               | 7.43   | 5.54               |
| Mothers' sons (N)                                | 1.18                                       | 3.76               | 1.14   | 3.90               |
| Mothers' daughters (N)                           | 1.08                                       | 3.50               | 1.17   | 3.98               |
| Childless women (N)                              | 2.45                                       | 3.09               | 2.77   | 3.01               |
| Mean mothers' age at daughters' birth            | 27.55                                      | 5.93               | 28.04  | 6.50               |
| Mean mothers' age at sons' birth                 | 27.36                                      | 6.24               | 28.24  | 6.57               |
| Mothers' age at 1 <sup>st</sup> daughter's birth | 25.11                                      | 5.23               | 25.64  | 5.79               |
| Mothers' age at 1 <sup>st</sup> son's birth      | 25.58                                      | 5.52               | 26.09  | 6.21               |
| Mothers' colon cancer frequency                  | 0.13                                       | 0.53               | 0.09   | 0.39               |
| Mothers' age at colon cancer                     | 53.92                                      | 15.76              | 58.14  | 16.23              |
| Mothers' breast cancer frequency                 | 0.07                                       | 0.43               | 0.06   | 0.35               |
| Mothers' ovarian cancer frequency                | 0.02                                       | 0.14               | 0.01   | 0.12               |
| Mothers' pancreatic cancer frequency             | 0.01                                       | 0.09               | 0.01   | 0.09               |
| Mothers' endometrial cancer frequency            | 0.07                                       | 0.42               | 0.02   | 0.22               |
| Other cancer frequency in mothers                | 0.15                                       | 0.68               | 0.13   | 0.58               |
| Not documented cancers in mothers                | 0.08                                       | 0.43               | 0.07   | 0.37               |
| Fathers' colon cancer frequency                  | 0.21                                       | 0.80               | 0.10   | 0.44               |
| Fathers' age at colon cancer                     | 50.82                                      | 14.30              | 59.56  | 13.21              |

**Table 1.** Example of characteristics (among 104 available) calculated for 2 groups of families at colorectal cancer risk.

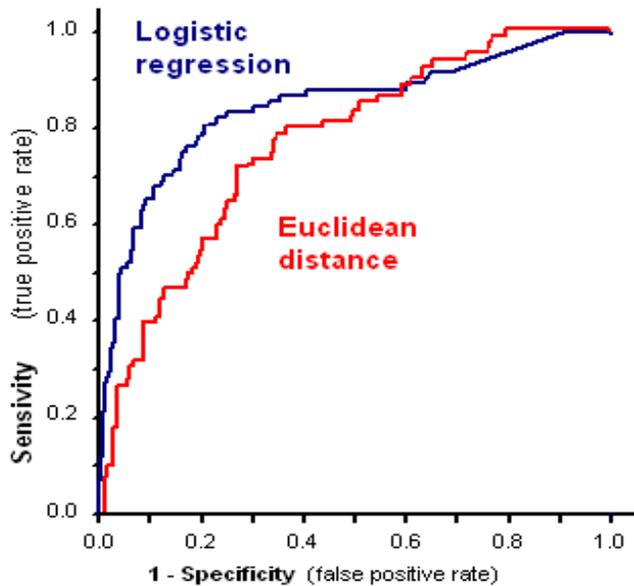
A previous comparison of the predictive values for BRCA mutations of each calculation mode among a very large sample of breast/ovarian cancer-prone families showed that the first Euclidean distance  $D$  performed slightly better than other methods (results not shown). We used the ratio  $D1/D2$  to study its discriminant power for MMR mutations. An ROC analysis was performed to compare this result with a logistic regression calculated on best known significant clinical predictors. Figure 6 presents the two results:

The ROC analysis calculated using the ratio of Euclidean distances between subtrees and both profiles is associated with a good AUC (area under the curve) = 0.76 [0.70; 0.81], a 71% sensitivity and a 72% specificity. The positive predictive value (PPV) is limited = 38% while the negative one (NPV) is rather high = 91%. Overall 72% of families are well classified (70% of mutated families and 72% of not mutated ones). Prediction of mutation by the logistic regression selects only 5 clinical parameters calculated per the whole family (independently from filiation): the number of colon cancers, lower age at colon cancer; the number of endometrial cancers, prostatic cancers, and multiple cancers, this latter parameter diminishing the likelihood of an MMR mutation. The regression formula



**Figure 5.** Distances between profiles and a new family X.

associated with these clinical factors yields a slightly better ROC curve (blue curve in figure 6, difference  $p < 0.01$ ): AUC = 8.5 [0.79; 0.90], sensitivity = 80%, specificity = 80%, PPV = 48% and NPV = 95%. Well, the classified rate is 80% overall and for each subgroup. Despite the superiority of the well-adjusted regression model, profiles that require neither selection nor hypothesis on covariates, appear to possess interesting discriminant properties with a fair ROC AUC ( $> 0.70$ ).



**Figure 6.** ROC curves comparing the predictive value for MMR mutations of best regression model and the ratio of Euclidean distance between profiles (difference between curves:  $p < 0.01$ ).

## Discussion

Pedigrees used in oncogenetics contain a large amount of clinical and biological information. Besides, large pedigrees provide complementary information, in particular regarding natality/fertility. This approach by “subtrees” represents a helpful solution to use more widely all available data whatever the size of pedigrees. Standardized and synthetic subtrees allow performing statistics on pedigrees, to build standard profiles according to specific characteristics and give indications about familial mutation risk. With the creation of profiles, the comparison between a new family (a single subtree) and various profiles becomes possible. This approach in our example concerning HNPCC predisposition, although not optimal, enabled to well classify most members carrying an MMR mutation without requiring hypothesis and/or restriction about selected criteria.

In the future, geneticists could gain time trying to “categorize” families with this method: they could be more specific when choosing which gene to sequence. We intend to test how the use of subtrees and profiles may help confirm or contradict, within our breast/ovaries cancer-prone families, a hypothesis regarding a monogenic or multigenic etiology.

A current weakness of our computer program is that it is only compatible with SEM software, used almost exclusively in the Jean Perrin Comprehensive Cancer Center. It should be re-developed for different working environments. The Microsoft Visual Basic source code is available on request to the corresponding author.

The purpose of our work was to contribute to the study of familial risks for any type of cancer, in relation to known or unknown deleterious mutations. Of course, such an approach may also be considered for other purposes than mutational risk prediction.

## Competing interest

We declare no competing interests: this work was supported by FEDER (European Funding of Regional Development) and Conseil Régional Auvergne (France).

## Authors' contributions

Article draft: Marie Arbre, Fabrice Kwiatkowski

Article revising: Pr Yves-Jean Bignon, Pr Laurent Serlet

Project responsible: Pr Yves-Jean Bignon

Subtrees software development: Marie Arbre

Pedigrees software development: Fabrice Kwiatkowski

Mathematical contribution: Pr Laurent Serlet

Statistical analysis: Marie Arbre, Fabrice Kwiatkowski

## Acknowledgments

Claire Laquet, the oncogenetic counselor, Laurence Boulègue, Sandrine Casteker, Mélanie Teurio and Sandra Charbonnier, secretaries of Oncogenetics Department.

## References

1. Glazner C, Thompson EA. Improving Pedigree-based Linkage Analysis by Estimating Coancestry Among Families. *Stat Appl Genet Mol Biol*. 2006, 11(2): pii.
2. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med*. 2004, 23: 1111-1130.
3. Eisinger F, Bressac B, Castaigne D, Cottu PH, Lansac J et al. Identification and management of hereditary predisposition to cancer of the breast and the ovary (update 2004). *Bull Cancer*. 2004, 91(3): 219-237.
4. Evans DG, Eccles DM, Rahman N, Young K, Bulman M et al. A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. *J Med Genet*. 2004, 41: 474-480.
5. Kwiatkowski F, Arbre M, Laquet C, Uhrhammer N, Bignon YJ. BRCA mutations increase fertility in families at hereditary breast/ovarian cancer risk. *Plos One*. 2015, 10(6): e0127363.

6. Mc Parland S, Kearney JF, Rath M, Berry DP. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *J Anim Sci.* 2007, 85(2): 322-331.
7. Cole JB. PyPedal: A computer program for pedigree analysis. *Comp Electro. Agric.* 2007, 57(1): 107-113.
8. Kwiatkowski F, Girard M, Hacene K, Berlie J. Sem: A suitable statistical software adapted for research in oncology. *Bull Cancer.* 2000, 87(10): 715-721.
9. Bennett RL, Steinhaus KA, Uhrich SB, O'Sullivan CK, Resta RG et al. Recommendations for Standardized Human, Pedigree Nomenclature. *Am J Hum Genet.* 1995, 56(3): 745-752.
10. Bennett RL, French KS, Resta RG, Doyle DL. Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. *J Genet Couns.* 2008, 17(5): 424-433.
11. Martinez-Delgado B1, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M et al. Genetic anticipation is associated with telomere shortening in hereditary breast cancer. *PLoS Genet.* 2011, 7(7): e1002182.
12. Litton JK, Ready K, Chen H, Gutierrez-Barrera A, Etzel CJ et al. Earlier age of onset of BRCA mutation-related cancers in subsequent generations. *Cancer.* 2011, 118(2): 321-325.
13. Narod SA. Earlier age of onset in BRCA carriers-anticipation or cohort effect? *Curr Oncol.* 2011, 18(6): 257-258.
14. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol.* 1975, 12(4): 387-395.
15. Hanley JA, McNeil BJ. The meaning and the use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982, 143(1): 29-36.
16. Terdiman JP, Gum JR, Conrad PG, Miller GA, Weinberg V et al. Efficient detection of hereditary nonpolyposis colorectal cancer gene carriers by screening for tumor microsatellite instability before germline genetic testing. *Gastroenterology.* 2001, 120(1): 21-30.
17. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on HNPCC. *Anticancer Res.* 1994, 14(4B): 1661-1664.
18. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology.* 1999, 116(6): 1453-1456.
19. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst.* 2004, 96(4): 261-268.