

# Méthodes multiblocs pour l'identification de gènes associés au vieillissement cutané chez 502 femmes Caucasiennes adultes

Anne Bernard<sup>\*,\*\*</sup>, Arthur Tenenhaus<sup>\*\*\*</sup>, Jean-Francois Zagury<sup>\*\*\*\*</sup>,  
Gilbert Saporta<sup>\*\*</sup>, Christiane Guinot<sup>\*,‡</sup>

\*CERIES, 20 rue Victor Noir, Neuilly sur Seine, France  
(anne.bernard;christiane.guinot@ceries-lab.com)

\*\*CNAM, laboratoire CEDRIC, 292 rue Saint-Martin, Paris, France  
gilbert.saporta@cnam.fr

\*\*\*SUPELEC, 3, rue Joliot-Curie, Gif-sur-Yvette, France  
arthur.tenenhaus@supelec.fr

\*\*\*\*CNAM, laboratoire GBA, 292 rue Saint-Martin, Paris, France  
zagury@cnam.fr

‡Université Francois Rabelais, département d'informatique, 64 avenue Jean Portalis, Tours, France

## 1 Introduction

Un projet de recherche (GWAS) a été mis en place en 2010 pour étudier l'impact du patrimoine génétique sur l'expression du vieillissement cutané (phénotypes), le but final étant d'identifier des cibles thérapeutiques pour le développement de nouvelles stratégies de diagnostic de la peau. Cette étude nécessite le développement de méthodes statistiques pour le traitement de données à grande dimension. On rapporte ici les premières analyses réalisées à l'aide de la méthode multiblocs Regularized Generalized Canonical Correlation Analysis (Tenenhaus et Tenenhaus, 2011).

## 2 Matériel

L'étude a été conduite sur 502 femmes caucasiennes de la cohorte SU.VI.MAX (Hercberg et al, 1998), vivant en Ile de France et âgées de 44 à 70 ans. A partir de photographies du visage des femmes, un dermatologue a évalué la sévérité du vieillissement cutané, à l'aide d'échelles ordinales spécifiques (Larnier et al, 1994 ; Morizot et al, 2002). De plus, des analyses génétiques ont été réalisées à l'aide de biopuces Illumina à 1140000 marqueurs génétiques (Single Nucleotide Polymorphism). Après contrôle qualité, 795063 SNPs ont été retenus dont 362223 étaient situés dans un total de 15198 gènes (après regroupement des SNPs par gènes, 15198 blocs de SNPs). Dans les premières analyses, 13 gènes déjà connus ont été étudiés.

### 3 Méthode

Les liens entre les génotypes et le vieillissement cutané ont été modélisés grâce à la méthode RGCCA. Son objectif est de trouver des combinaisons linéaires entre blocs de variables, telles que les composantes expliquent correctement leur propre bloc et que les composantes des blocs connectés soient fortement corrélées. Les blocs explicatifs sont les blocs de SNPs  $X_1$ ,  $X_{13}$  et  $X_{14}$  (vieillessement cutané) est le bloc à expliquer. Le problème d'optimisation est le suivant :

$$\text{Max}_{\{a_1, \dots, a_j\}} \sum_{j,k=1, j \neq k}^J c_{jk} \mid \text{cov}(X_j a_j, X_k a_k) \mid$$

sous la contrainte :  $\|a_j\|^2 = 1, j = 1, \dots, J$  avec  $y_j = X_j a_j$  les composantes à optimiser et  $C = \{c_{jk}\}$  tq  $c_{jk} = 1$  si  $X_j$  et  $X_k$  sont liés, 0 sinon.

### 4 Résultats - Conclusion

Les composantes expliquent correctement leur propre bloc (ex :  $\text{AVE}(\text{gene1}) = 0,8344$ ), cependant la qualité du modèle interne est basse ( $\text{AVE} = 0,0198$ ). L'aspect unidimensionnel de la RGCCA ne suffit pas, il faut donc étendre au multidimensionnel. La solution est donc de chercher à construire pour chaque bloc plus d'une composante externe (déflation). Le nombre de composantes nécessaires a été calculé par validation croisée (régression PLS). Par ailleurs, les interactions éventuelles entre gènes n'ont pas été prises en compte, il faudrait donc par la suite ajouter des liens entre blocs (gènes) et généraliser la méthode aux 15198 gènes.

### Références

- Hercberg, S., P. Preziosi, S. Briancon, P. Galan, I. Triol, D. Malvy, et al (1998). A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the su.vi.max study—design, methods, and participant characteristics. supplementation en vitamines et minéraux antioxydants. *Control Clinical Trial* 19, 336–351.
- Larnier, C., J. Ortonne, A. Venot, B. Faivre, J. Béani, P. Thomas, et al (1994). Evaluation of cutaneous photodamage using a photographic scale. *British Journal of Dermatology* 130, 167–173.
- Morizot, F., S. Lopez, C. Guinot, M. Binder, et E. Tschachler (2002). Development of photographic scales documenting features of skin ageing based on digital images. *Annales de dermatologie et de venerologie* 129, 1s402.
- Tenenhaus, A. et M. Tenenhaus (2011). Regularized generalized canonical correlation analysis. *Psychometrika* 76, 257–284.

### Summary

A study was conducted on 502 women to identify genes affecting skin aging. The severity of 4 phenotypes was evaluated and genetic analysis were done with microarrays. The relationships between genes and the phenotypes were tested using a multiblock method.