

Étude transcriptomique d'épithéliums olfactifs de souris ayant été exposées à différents composés odorantes

Université Paris 7 Diderot

Master 2 Biologie - Informatique - Bioinformatique

Année 2016 - 2017

Compte Rendu de Projet Long

Dubrunfaut Lucie

Sommaire

I. Introduction

II. Matériel et Méthodes

1. Échantillons de muqueuse olfactive de souris
2. Extraction et séquençage des transcrits
3. Préparation des données
4. Recherche des récepteurs olfactifs
5. Analyse des données

III. Résultats et Discussion

1. Présentation des données
2. Sélection des récepteurs olfactifs
3. Analyse exploratoire non supervisée des données
4. Estimation de la dispersion des données
5. Analyse des gènes ayant une expression différentielle

IV. Conclusions

Références

Introduction

L'odorat est un sens qui permet à l'individu d'obtenir des informations de son environnement essentielles, nécessaires parfois pour des interactions sociales ou la reproduction, mais également des informations vitales comme un danger (prédateur, incendie, etc) ou encore si de la nourriture est proche et si elle est comestible. Tous les composés chimiques associés présents dans l'air sont appelés odorants et sont détectés par une catégorie spécifique de récepteurs qui sont les récepteurs olfactifs (OR). Ils sont notamment reconnus d'après leurs propriétés chimiques (poids moléculaire, groupement chimique, etc) [1]. De plus ils peuvent parfois donner une odeur précise lorsqu'ils sont plusieurs combinés [2]. Ainsi le système olfactif doit être capable de donner une réponse à ces odorants, qui puisse refléter la composition du mélange d'odorants et de la relayer au cerveau. Le nombre de gènes codant pour des ORs d'une espèce à l'autre est très variable, autant en terme de nombre qu'en terme de diversité [3], chez la souris les gènes codant pour les ORs représente presque 5% du génome (1130 ORs) tandis que chez l'Homme environ 400 ORs ont été mis en évidence (auxquels s'ajoute plus de 400 pseudogènes) [4]. Cependant ce nombre d'ORs est bien trop insuffisant devant les millions de composés volatiles possibles pour qu'un OR soit spécifique d'un composé. Ainsi un composé peut être reconnu par de multiples ORs et inversement un OR peut reconnaître de multiples odorants. Ceci conduit à un code combiné de l'information olfactive [5] comme l'illustre la Figure 1.

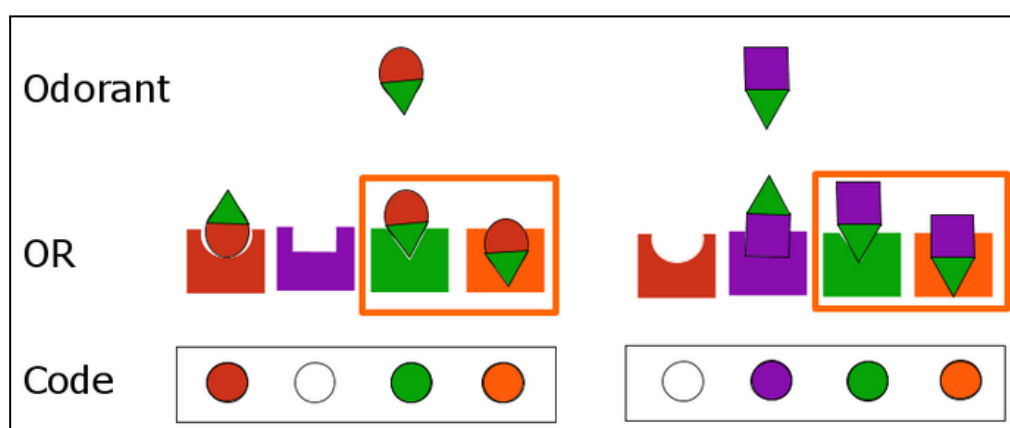


Figure 1 : Représentation schématique du système de codage combiné permettant la reconnaissance des odorants

Actuellement notre connaissance des ORs n'est généralement pas associée au type d'odorant qui les activent, ce qui rend notre connaissance du codage de l'olfaction assez fragmentaire. Ce même si de

nombreuses méthodes (utilisation de souris transgéniques marquées par fluorescence[6], expression *in vitro* des ORs [7], etc) ont été mise en oeuvre pour associer des ORs avec leurs ligands.

C'est dans cette problématique que s'inscrit notre projet qui vise à mettre en évidence chez l'Homme, les gènes codant pour des ORs présentant une expression différentielle suite à une exposition à différents odorants ou combinaison d'odorants. Notre étude visait dans un premier temps à restreindre nos données aux ORs connus chez l'Homme, puis à réaliser une analyse exploratoire non supervisée des données avant de faire l'analyse des gènes exprimés différenciellement selon l'odorant auquel l'individu a été exposé.

Matériel et Méthodes

1. Échantillons de muqueuse olfactive de souris

Les 24 souris utilisées dans cette étude sont des mâles C57Bl6 issues de l'élevage du laboratoire, elles sont élevées en cages collectives (3 animaux/cage). Elles ont suivi le régime A03 jusqu'à 5 semaines révolues, puis régime A04 à partir de la 6ème semaine (ad libitum). A l'âge de 8 semaines, elles ont été isolées 24h en cages individuelles et exposées au solvant (HM). Ensuite elles ont été exposées au composé (odorant ou solvant) dans une autre salle pendant 5 heures avant de procéder au prélèvement des muqueuses olfactives. Les 20 animaux ont été exposés au composé le même jour, les échantillons prélevés ont été conservés dans de l'azote liquide à -80°C.

2. Extraction et séquençage des transcrits

Les transcrits ont été prélevés en utilisant le kit RNeasy Plus Mini Kit de Qiagen, le contrôle qualité des transcrits a été réalisé par Experion™ et la quantification des transcrits par Nanodrop. Le séquençage nécessitant pour l'analyse 4 µg transcrit/50 µl a été réalisé par la plateforme Genotoul.

3. Description et préparation des données numériques

Les données issues du séquençage ont été préparées (contrôle qualité, cartographiés, etc) par le procédé NG6 [8], les données numériques résultantes ont permis le lancement de ce projet. Elles sont

répertoriées dans une matrice contenant les échantillons en colonne et les gènes au format ENSEMBL en ligne, ainsi pour chaque échantillon il est possible de compter les reads venant de chaque gène. De plus l'utilisation du langage R [9] et plus précisément du package EdgeR [10][11] pour filtrer les transcrits.

4. Recherche des récepteurs olfactifs

La recherche des gènes orthologues chez l'Homme dans nos données au format HGNC a été fait en utilisant l'outil biological DataBase network [12].

5. Analyse des données

L'analyse des données a été réalisée en utilisant le langage R [9] et plus précisément du package EdgeR [10][11] pour les analyses exploratoires non supervisées des données, ainsi que l'analyse des expressions différentielles des gènes. La création des diagrammes de Venn a été assurée par l'outil de la plateforme bioinformatics.psb.ugent.be [13].

Résultats et Discussion

1. Présentation des données

Nous disposons de données transcriptomiques issues de RNA-seq et répertoriées dans une matrice, celles-ci ne sont pas brutes et ont subi au préalable un traitement par le protocole NG6 [8]. Elles ont été générées à partir de transcrits issus de muqueuses olfactives de souris exposées à différents odorants (CA, EM, IE et MEL soit IE + EM). La Table 1 indique la répartition des échantillons dans le modèle expérimental. Il faut noter qu'il n'existe pas de réplicats biologiques c'est à dire que chaque échantillon correspond à un individu.

Groupe de traitement	CA	EM	IE	MEL (IE + EM)
Nombre d'échantillons	4	4	4	4

Table 1 : Répartition des échantillons dans les catégories du plan expérimental.

2. Sélection des récepteurs olfactifs

Notre étude vise à étudier les gènes des ORs connus chez l'Homme ce qui a nécessité l'utilisation des orthologues de ceux-ci connus chez la souris, ce étant donné que cette étude a été menée à partir d'échantillons de muqueuses de souris. Pour ce faire nous avons utilisé l'outil biologique DataBase network [12] qui a permis d'associer les identifiants ENSEMBL des gènes de nos données à des identifiants HGNC correspondant aux orthologues respectifs connus chez l'Homme. Ensuite la base de données HGNC [14][15] nous a permis d'obtenir notre première liste contenant les identifiants HGNC des gènes ORs connus chez l'Homme (contenant 861 gènes). Dans un second temps nous avons choisi d'élargir notre champ de recherche en créant une seconde liste contenant des gènes associés à l'olfaction chez l'Homme mis en évidence dans les travaux de Olender et al. [16] (contenant 101 gènes). Nous avons pour cette dernière également utilisé l'outil biologique DataBase network pour obtenir des identifiants de gène HGNC.

Nous avons alors filtré nos données pour ne conserver que deux matrices contenant chacune seulement les gènes d'une des deux listes obtenues précédemment. Nous disposions alors de deux matrices contenant respectivement les données transcriptomiques associées aux gènes des ORs connus chez l'Homme (491 gènes) et les données transcriptomiques associées à des gènes ayant un lien avec les mécanismes olfactif des muqueuses olfactives (90 gènes). Chacune de ces matrices ont été filtrées en utilisant le package EdgeR [10][11] pour ne conserver que les transcrits qui présentaient au moins plus d'un million de reads dans trois échantillons ou plus.

3. Analyse exploratoire non supervisée des données

Nous souhaitons avoir une vision globale de la structure des données que nous allons analyser. Il s'agit essentiellement d'une approche qualitative et visuelle qui vise à repérer si sans à priori sur les données nous repérons des groupes d'échantillons qui pourraient correspondre à nos groupes de traitement, mais également repérer si certains échantillons sont aberrants (outliers) et doivent être retirés des données ou non. Nous avons utilisé le MDS (Multidimensional Scaling) qui a permis de passer de nos échantillons avec de nombreux gènes (considérés comme des coordonnées) à ces mêmes échantillons dans un espace à deux

dimensions permettant de facilement les visualiser. La Figure 2 illustre les résultats obtenus par MDS pour les deux matrices.

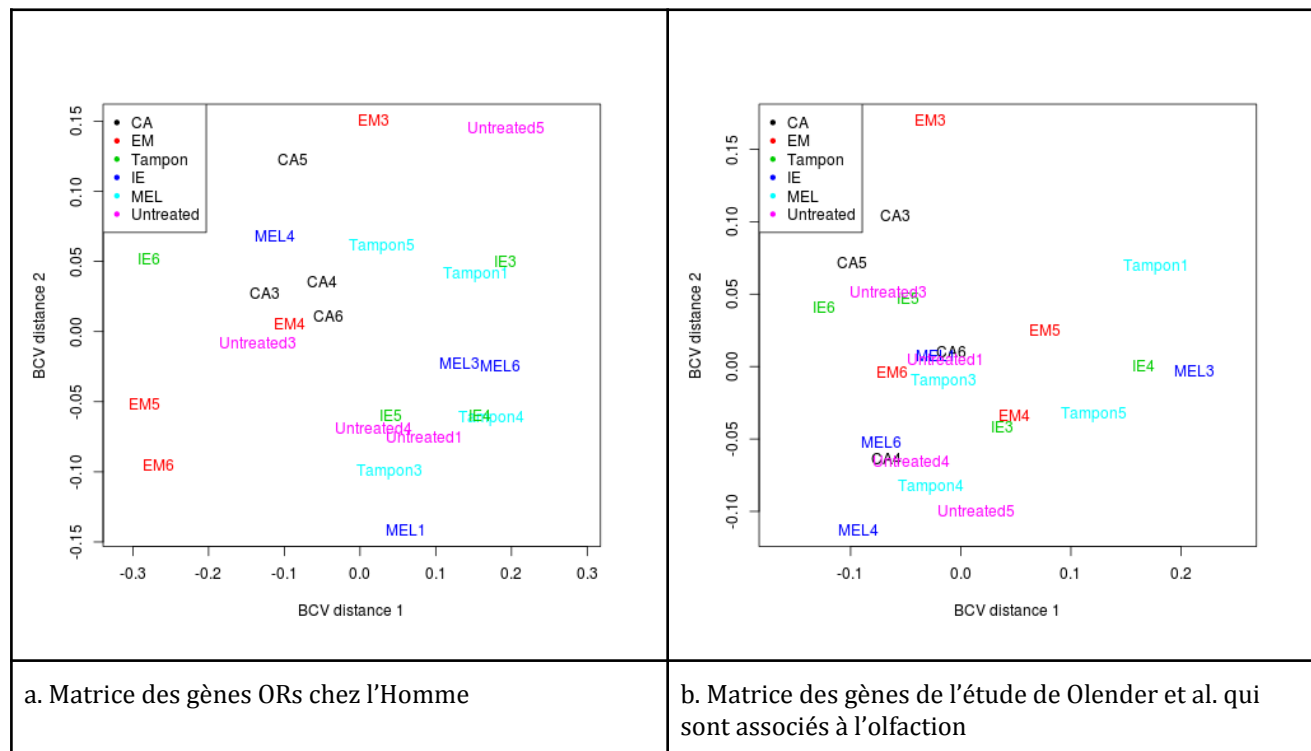


Figure 2 : Distribution des échantillons produite par la méthode MDS pour les deux matrices

En analysant la Figure 2, nous n'observons pas vraiment de groupes d'échantillons clairement distincts à première vue qui puissent correspondre à nos groupes de traitement, ce pour chacune des matrices. Ceci n'indique pas forcément que les groupes de traitements ne diffèrent pas entre eux au niveau de l'expression des gènes olfactifs que nous analysons. En effet, étant donné que nous analysons de nombreux gènes, il se peut que les gènes qui présentent une expression différentielle propre à un groupe de traitement soient peu nombreux et donc que leur signal soit noyé dans la masse. De plus pour les deux matrices nous n'observons pas d'échantillon qui semble à priori aberrant.

4. Estimation de la dispersion des données

Une des premières étape pour réaliser l'analyse de l'expression différentielle est d'estimer le paramètre de dispersion pour chaque gène, soit donner une mesure du degré de variation inter-échantillon pour ce gène. L'estimation de la dispersion commune donne une idée de la variabilité globale à travers les gènes considérés pour cet ensemble de données. D'après les résultats de Landau et al. [17] on choisit d'utiliser la méthode quantile-adjusted conditional maximum likelihood (qCML) plutôt que la méthode Generalized linear models (GLM) qui

convient mieux à des données avec des facteurs multiples testés. Les résultats obtenus pour les deux matrices sont présentés en Figure 3.

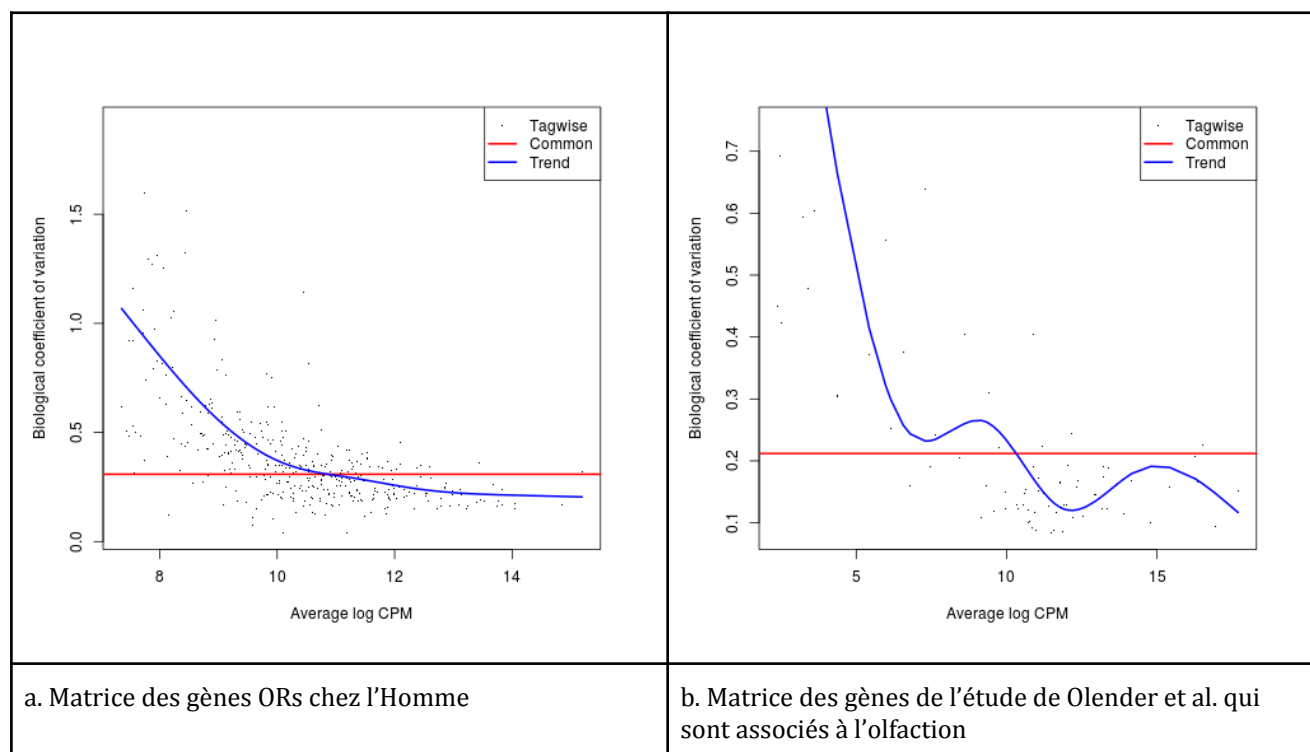


Figure 3 : Estimation du paramètre de dispersion en utilisant la méthode quantile-adjusted conditional maximum likelihood (qCML)

L'analyse de la Figure 3 permet de vérifier que le paramètre de dispersion estimé par notre méthode semble adapté aux données. On constate que la courbe trend suit relativement bien la dispersion des tagwise, ainsi la méthode semble donner une estimation de la dispersion plutôt bonne.

5. Analyse des gènes ayant une expression différentielle

L'analyse exploratoire donnaient une information sur le profil d'expression des gènes de façon globale, nous souhaitons à présent mettre en évidence les gènes ayant une expression significativement différente entre les groupes de nos données. Pour y parvenir nous avons réalisé des test exacts de chaque groupe deux à deux, permettant de générer des valeurs de FDR (False Discovery Rate). Ensuite nous n'avons conservé dans chacune des comparaisons, que les gènes ayant une expression différentielle significative, soit qui présentaient une valeur de FDR ≤ 0.05 . Nous faisons remarquer que "Composé A vs Composé B" signifie que pour chaque gènes l'expression du composé A a été soustraite du composé B.

- **Liste des gènes codants pour des ORs**

L'ensemble des résultats de ces comparaisons pour les gènes codant pour des ORs, est fournie dans les Supplementary Data (fichier disponible dans l'archive : Gene_DE_comparaison.xlsx).

Sur l'ensemble des comparaisons on constate que seulement 5 gènes ORs différents (**OR2A** - HGNC:15001 ; **OR2L13** - HGNC:19578 ; **OR5K3** - HGNC:31290 ; **OR13C7** - HGNC:15102 ; **OR11A1** - HGNC:8176) ressortent comme étant significativement exprimés différemment dont 2 (OR2A et OR2L13) le sont pour chaque comparaison où des gènes exprimés différemment ressortent.

Dans un premier temps on vérifie nos deux comparaisons témoins qui sont présentées en diagramme de Venn dans la Figure 4.

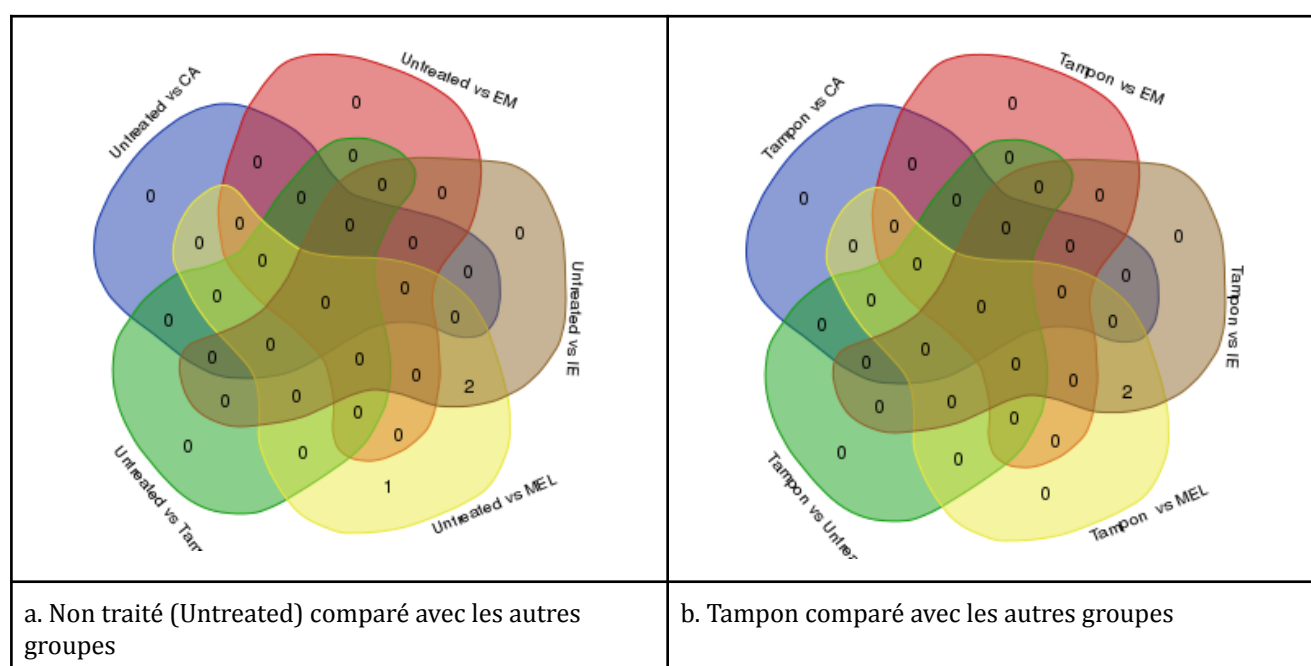


Figure 4 : Diagramme de Venn des listes de gènes exprimés différemment dans les comparaisons témoins

L'analyse de la Figure 4.a montre que seuls les groupes exposés aux odorants MEL (EM + IE) et IE présentent une expression différentielle des gènes (pour MEL et IE : OR2A, OR2L13 et pour MEL : OR5K3) par rapport au groupe non traité (Untreated). Qu'il n'y ait pas de gènes exprimés différemment entre le groupe Tampon et Untreated était attendu puisque le tampon ne devait pas avoir d'effet parasite dans l'étude. En revanche il est étonnant que les groupes CA et EM ne présentent aucune expression différentielle par rapport au groupe Untreated. Une explication possible serait que l'expression différentielle n'a pas pu être mise en évidence en raison d'échantillon(s) aberrants qui auraient brouillé le signal mais que nous n'aurions pas vu lors de l'analyse exploratoire. Il est également possible qu'il n'y ait pas d'expression différentielle dans ces comparaisons car les récepteurs olfactifs ne sont pas sensibles aux odorants des groupes CA et EM. La Figure 4.b confirme que le Tampon n'induit dans aucun groupe des

expression différentielles de nouveaux gènes par rapport aux comparaisons avec le groupe Untreated. En revanche l'expression différentielle du gène OR5K3 dans la comparaison entre MEL et Untreated n'apparaît pas dans la comparaison MEL et Tampon. Il se peut que celle-ci n'ait pas été retenue car elle n'apparaissait pas comme significative ($FDR > 0.05$) dans cette comparaison. En explorant davantage nous avons pu constater que même en augmentant le champ d'acceptation avec un seuil de sélection avec un $FDR \leq 0.20$ le gène n'apparaît tout de même pas dans la comparaison entre MEL et Tampon. En revanche nous avons pu constater que dans la comparaison entre MEL et Untreated ce gène présente un FDR de 0.048 ce qui signifie qu'il était très proche du seuil de rejet de FDR à 0.05. Nous devons donc tenir compte de cette information pour la suite de l'étude.

Dans un second temps nous nous sommes intéressé à regarder les gènes ORs qui étaient exprimés différentiellement dans les comparaisons impliquant les groupes IE et MEL puisqu'ils étaient les seuls à présenter une expression différentielle qu'on suppose liée à l'odorant en question (IE et MEL). La Figure 5 représente des diagrammes de Venn réalisés à partir des listes de gènes différentiellement exprimés dans respectivement chaque comparaison. De plus la Table 2 indique les valeurs de fold change transformées par fonction logarithmique, pour les cinq gènes dans les comparaisons impliquant IE, de même pour MEL dans la Table 3.

Composé IE	HGNC:15001 OR2AJ	HGNC:19578 OR2L13	HGNC:31290 OR5K3	HGNC:15102 OR13C7	HGNC:8176 OR11A1
IE vs CA	1,40 ; 2,19	0,93	1,44	-	-
IE vs EM	2,21 ; 1,12	1,26	1,37 ; 0,82	-	-
IE vs MEL	-	-	-	-	-
IE vs Tampon	2,42 ; 1,30	0,95	-	-	-
IE vs Untreated	2,16 ; 1,37	1,05	-	-	-

Table 2 : Valeurs des expressions différentielles des gènes associés aux comparaisons avec IE (Les valeurs présentées pour chaque gène sont des $\log(\text{Fold-Change})$)

L'analyse de la Figure 5.a et de la Table 2 permet de constater que pour les gènes OR2AJ et OR2L13, les groupes CA et EM se comportent comme les groupes Tampon et Untreated lors de la comparaison avec IE. Une explication pourrait être que du point de vue des récepteurs olfactifs ces groupes ont le même profil d'expression que les groupes témoins (Tampon et Untreated). Cependant pour le gène OR5K3 les groupes CA et EM semblent avoir une expression de ce gène plus grande que dans le groupe IE, avec pour le groupe CA $\log(FC)=1.44$ avec une $FDR \approx 0.0001$ et pour le groupe EM $\log(FC) \approx 1.0$ avec une $FDR \approx 0.001$. Une étude en réduisant le seuil de sélection du FDR à 0.20 nous a permis de remarquer que ce gène OR5K3 apparaît

également dans d'autres comparaisons avec le groupe EM mais également dans la comparaison entre Tampon et Untreated.

Composé MEL (IE + EM)	HGNC:15001 OR2AJ	HGNC:19578 OR2L13	HGNC:31290 OR5K3	HGNC:15102 OR13C7	HGNC:8176 OR11A1
MEL vs CA	2,24 ; 1,80	1,02 ; 0,9	1,58		-0,77
MEL vs EM	2,26 ; 1,53	1,35	1,50 ; 1,05	1,9	-
MEL vs IE	-	-	-	-	-
MEL vs Tampon	2,47 ; 1,70	1,04 ; 0,86	-	-	-
MEL vs Untreated	2,22 ; 1,77	1,14 ; 0,89	1,01	-	-

Table 3. : Valeurs des expressions différentielles des gènes associés aux comparaisons avec MEL (Les valeurs présentées pour chaque gène sont des log(Fold-Change))

L'analyse de la Figure 5.b et de la Table 3 ont également mis en évidence que l'expression différentielle des gènes OR2AJ et OR2L13 dans les comparaisons avec les groupes CA et EM étaient sensiblement du même ordre et dans le même sens que pour les groupes témoins (Tampon et Untreated). Concernant le gène OR2K3 qui apparaît également dans cette comparaison, encore une fois la significativité des valeurs affichées est forte (groupe CA $FDR \approx 0.000009$ et groupe EM $FDR \approx 0.0001$). De plus le gène OR13C7 apparaît dans la comparaison entre le groupe Ca et MEL, avec $\log(FC) = 1.9$ avec une $FDR \approx 0.0012$, en cherchant davantage autour de l'expression différentielle de ce gène nous avons constaté que dans la comparaison entre MEL et Untreated il aurait eu $\log(FC) = 1.34$ avec une $FDR \approx 0.0544$. Ainsi ce gène aurait une expression moins forte dans le groupe MEL que dans le groupe EM, alors qu'entre le groupe MEL et le groupe EI ce gène ne ressort pas. Nous aurions pu penser qu'entre l'odorant IE et MEL qui est le mélange de EI et MEL, il n'y avait pas de différence d'expression des ORs; mais le gène OR13C7 pourrait être le marqueur de la différence induite par la combinaison IE + EM. Pour finir nous avons constaté que la comparaison du groupe CA et MEL mettait en évidence le gène OR11A1 mais avec un $FDR = 0.045$ et donc très proche du seuil d'exclusion que nous avons défini, il est donc possible que ce gène ORs un faux positif.

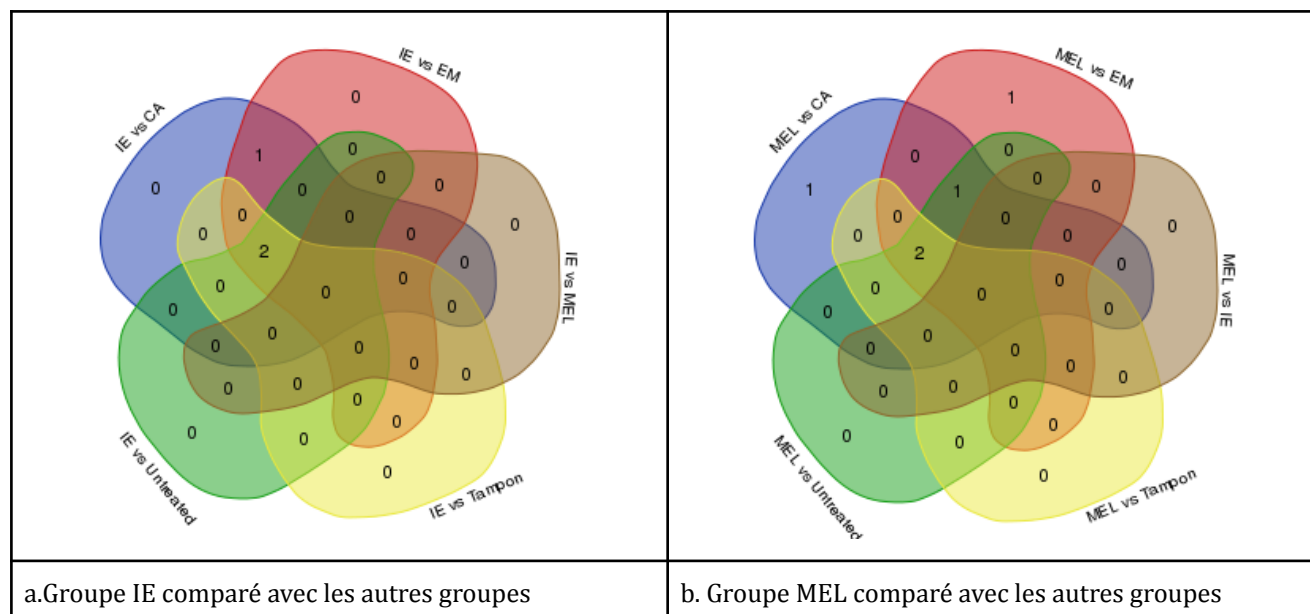


Figure 5: Diagramme de Venn des listes de gènes exprimés différemment dans les comparaisons IE et MEL

Les gènes OR2AJ et OR2L13 sont très probablement impliqués dans la reconnaissance des composés IE et MEL, tandis que le gène OR13C7 permettrait de dissocier le composé MEL du composé IE par une très légère variation d'expression, mais il est difficile de déterminer la fiabilité de l'implication de ce gène ORs dans l'exposition aux composés de cette étude.

- **Liste des gènes de l'étude de Olender et al. qui sont associés à l'olfaction**

Les résultats des gènes associés à l'étude de Olender et al. n'a permis de mettre en évidence sur l'ensemble des comparaisons effectuées que deux gènes exprimés différemment dans une comparaison. Il s'agit de la comparaison entre le groupe non traité et le groupe exposé au composé MEL : NRN1L (HGNC:29811 - Uniprot : Q496H8) avec $\log(\text{FC}) \approx -0.52$ et $\text{FDR} \approx 0.004$; GNG8 (HGNC:19664 - Uniprot : Q9UK08) $\log(\text{FC}) \approx -0,80$ et $\text{FDR} \approx 0.03$. Des recherches nous ont permis de savoir que GNG8 (Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-8) est un composé impliqué dans de très nombreux processus et le FDR associé étant proche de la valeur de rejet il est raisonnable de penser qu'il s'agit d'un faux positif. En ce qui concerne NRN1L (Neuritin-like protein) il est associé à des activités neuronale, étant donné que les gènes ORs sont exprimés à la surface de neurones ce gène exprimé est cohérent. Il est cependant difficile d'expliquer pourquoi ce gène est légèrement surexprimé dans le groupe MEL par rapport au groupe Untreated.

Conclusions

Cette étude a permis de mettre en évidence les gènes OR2AJ et OR2L13 comme gène ORs permettant très certainement la reconnaissance du composé IE et donc de celui-ci dans le mélange IE + EM. Le gène OR13C7 serait un gène ORs qui ne reconnaîtrait pas EM mais qui serait capable reconnaître le mélange des composés IE et EM induisant une très légère variation d'expression. Ces variations d'expression des gènes ORs sont à confirmer par des méthodes de biologie moléculaire. De plus une analyse fonctionnelle du gène NRN1L aurait été souhaitable afin d'explorer davantage de nouvelles perspectives dans le domaine de l'olfaction.

Références

1. Haddad, Rafi, Rehan Khan, Yuji K. Takahashi, Kensaku Mori, David Harel, et Noam Sobel. « A Metric for Odorant Comparison ». *Nature Methods* 5, n° 5 (mai 2008): 425-29. doi:10.1038/nmeth.1197.
2. Weiss, Tali, Kobi Snitz, Adi Yablonka, Rehan M. Khan, Danyel Gafso, Elad Schneidman, et Noam Sobel. « Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white ». *Proceedings of the National Academy of Sciences of the United States of America* 109, n° 49 (4 décembre 2012): 19959-64. doi:10.1073/pnas.1208110109.
3. Niimura, Yoshihito, Atsushi Matsui, et Kazushige Touhara. « Extreme Expansion of the Olfactory Receptor Gene Repertoire in African Elephants and Evolutionary Dynamics of Orthologous Gene Groups in 13 Placental Mammals ». *Genome Research* 24, n° 9 (9 janvier 2014): 1485-96. doi:10.1101/gr.169532.113.
4. Niimura, Yoshihito, et Masatoshi Nei. « Evolution of Olfactory Receptor Genes in the Human Genome ». *Proceedings of the National Academy of Sciences* 100, n° 21 (14 octobre 2003): 12235-40. doi:10.1073/pnas.1635157100.
5. Malnic, B., J. Hirono, T. Sato, et L. B. Buck. « Combinatorial Receptor Codes for Odors ». *Cell* 96, n° 5 (5 mars 1999): 713-23.
6. Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* 22, 3033–3043 (2002).
7. Saito, H., Kubota, M., Roberts, R.W., Chi, Q. & Matsunami, H. RTP family members induce functional expression of mammalian odorant receptors. *Cell* 119, 679–691 (2004).
8. Mariette, Jérôme, Frédéric Escudié, Nicolas Allias, Gérald Salin, Céline Noirot, Sylvain Thomas, et Christophe Klopp. « NG6: Integrated next generation sequencing storage and processing environment ». *BMC Genomics* 13 (2012): 462. doi:10.1186/1471-2164-13-462.
9. R Core Team (2016). R: A language and environment for statistical computing . R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.Rproject.org/>
10. Robinson MD, McCarthy DJ and Smyth GK (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics*, **26**, pp. -1.
11. McCarthy, J. D, Chen, Yunshun, Smyth and K. G (2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation." *Nucleic Acids Research*, **40**(10), pp. -9.

12. Mudunuri, Uma, Anney Che, Ming Yi, et Robert M. Stephens. « bioDBnet: the biological database network ». *Bioinformatics* 25, n° 4 (15 février 2009): 555-56. doi:10.1093/bioinformatics/btn654.
13. <http://bioinformatics.psb.ugent.be/webtools/Venn/>
14. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. **genenames.org: the HGNC resources in 2015**. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D1079-85. doi: 10.1093/nar/gku1071
15. HGNC Database, HUGO Gene Nomenclature Committee (HGNC), EMBL Outstation - Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK www.genenames.org.
16. Olender, Tsviya, Ifat Keydar, Jayant M. Pinto, Pavlo Tatarsky, Anna Alkelai, Ming-Shan Chien, Simon Fishilevich, et al. « The human olfactory transcriptome ». *BMC Genomics* 17 (2016): 619. doi:10.1186/s12864-016-2960-3.
17. Landau, William Michael, et Peng Liu. « Dispersion Estimation and Its Effect on Test Performance in RNA-seq Data Analysis: A Simulation-Based Comparison of Methods ». *PLOS ONE* 8, n° 12 (9 décembre 2013): e81415. doi:10.1371/journal.pone.0081415.