



# Εθνικό Μετσόβειο Πολυτεχνείο

## Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

### Διπλωματική εργασία

#### Τίτλος διπλωματικής εργασίας:

**«Ανίχνευση γονιδίων-δεικτών για την απόκριση  
του ανθρώπινου οργανισμού στην αέρια  
βιομηχανική ρύπανση με τη χρήση  
μικροσυστοιχιών»**

Όνομα φοιτητή: Έκτωρ-Ξαβιέ Δελαστίκ

Όνομα επιβλέποντος καθηγητή: Α. Γεωργακίλας

Αθήνα, Οκτώβριος 2016

## Ευχαριστίες

Ευχαριστίες προς:

Τα μέλη της τριμελούς επιτροπής για τη συμμετοχή τους σε αυτήν, αλλά και ειδικότερα,

στον κ. Γεωργακίλα, για τη στήριξη και την εμπιστοσύνη που έδειξε κατά τη διάρκεια αυτής της εργασίας

στον κ. Χατζηγιάννου για τις κομβικές συζητήσεις και προτάσεις που επέτρεψαν σε αυτή την έρευνα να φτάσει τόσο μακριά

στον κ. Παπαγιάννη για τις παρατηρήσεις του και την ανάδειξη της δυνατότητας επέκτασης αυτής της εργασίας

Ευχαριστίες στην Ειρήνη Λιάμπα από την ομάδα Βιοπληροφορικής του ΕΙΕ, για τις ώρες συζητήσεων και προτάσεων που επέτρεψαν την αποπεράτωση αυτής της δουλειάς σε αυτό το επίπεδο.

Ιδιαίτερες ευχαριστίες επίσης στα μέλη της ερευνητικής ομάδας Βιοπληροφορικής του ΕΙΕ, γιατί όλες οι συζητήσεις, μικρές και μεγάλες, συνέβαλαν ουσιαστικά :

Ο. Παπαδήμα

Γ. Βαλαβάνη

Ε. Πιλάλη

Κ. Βουτετάκης

Μ. Λογοθέτη

Γ. Κοντογιάννη

Ε. Λιάμπα

Σ. Ι. Βλαχάβας

I. Binembbaum

Κ. Γελές

Ε. Γώσα

Α. Σοφρωνά

Σ. Αλεξίου

# Περιεχόμενα

Περίληψη – 5

## **Κεφάλαιο πρώτο: Θεωρητικά ζητήματα**

- 1.1 : Μικροσυστοιχίες για δεδομένα έκφρασης γονιδίων – σ. 9
- 1.2 : Θεωρητικά επί των εργαλείων της ανάλυσης – σ. 11
- 1.3 : Ανάλυση – σ. 12
- 1.4 : Λειτουργική Ανάλυση – σ. 14
- 1.5 : Περιβαλλοντική ρύπανση και μικροσωματιδιακοί ρύποι (particulate matter) – σ. 16

## **Κεφάλαιο δεύτερο: Υλικά και μέθοδοι**

- 2.1 : Μέθοδος που ακολουθήθηκε κατά τις δειγματοληψίες – σ. 23
- 2.2 : Υλικοτεχνική υποδομή της ανάλυσης – σ. 24

## **Κεφάλαιο τρίτο: Μέθοδος της ανάλυσης**

- Περιορισμοί – σ. 26
- Προεπεξεργασία – σ. 26
- Επιλογή προσέγγισης – σ. 34
- Γραμμικά μοντέλα και επιλογή συγκρίσεων – σ. 34
- Εξαγωγή λιστών – σ. 38
- Έλεγχος των gene symbols – σ.39
- Σύγκριση με οντολογίες – σ. 39
- Διαχωρισμός με τη χρήση γονιδίων-κόμβων – σ. 39

## **Κεφάλαιο τέταρτο: Αποτελέσματα**

- 4.1 : Λίστες στατιστικώς σημαντικά διαφορικώς εκφρασμένων γονιδίων – σ. 41
- 4.2 : Αποτελέσματα σύγκρισης των λιστών με τη Gene Ontology–σ. 62
- 4.3 : Αποτελέσματα σύγκρισης των λιστών με τη Human Phenotype Ontology – σ. 77
- 4.4 : Αποτελέσματα σύγκρισης των λιστών με τη Mammalian Phenotype – σ. 82
- 4.5 : Αποτελέσματα της σύγκρισης των λιστών με τη Reactome Pathways – σ. 87
- 4.6 : Αναζήτηση υποσυνόλου των γονιδίων-κόμβων που να αναδεικνύει το διαχωρισμό μεταξύ των πόλεων – σ. 91
- 4.7 : Σύγκριση με υπάρχουσα βιβλιογραφία – σ. 102

**Κεφάλαιο πέμπτο: Συζήτηση – σ. 104**

**Βιβλιογραφία – σ. 108**

**Παράρτημα: Ο κώδικας της ανάλυσης – σ. 111**

## Συντομογραφίες και όροι

Θα περιοριστούμε όπου αυτό είναι απαραίτητο στις ειδικές σημασίες όρων στα πλαίσια της Βιοπληροφορικής, αντί των γενικότερων σημασιών τους.

**Γονίδια-κόμβοι (hub genes):** Ελαστικός όρος που υποδεικνύει γονίδια τα οποία εντοπίζονται σε αρκετούς οντολογικούς όρους, γεγονός που συχνά υποδεικνύει και αναβαθμισμένο βιολογικό ρόλο.

**Λίστα γονιδίων (gene list):** Πίνακας στατιστικά σημαντικώς διαφορετικά εκπεφρασμένων γονιδίων.

**Οντολογία:** Εντός Βιοπληροφορικής, η κωδικοποίηση μέσω ιεραρχημένων όρων της γνώσης περί γονιδίων, των βιολογικών αποτελεσμάτων των προϊόντων τους και των σχέσεων μεταξύ τους.

**Υβριδισμός:** Η επιλεκτική και ισχυρή σύνδεση δύο συμπληρωματικών αλυσίδων DNA ή RNA.

**Array ή μικροσυστοιχία:** Σύνολο probes, το οποίο αποτελεί πλήρες σύστημα ελέγχου για ένα δείγμα.

**Batch effect correction ή batch correction:** Διαδικασία κατά την οποία υπολογίζεται και αντιμετωπίζεται η συνεισφορά τεχνικών μεταβλητών οι οποίες δεν έχουν σχέση με το υπό μελέτη φαινόμενο.

**Beadchip:** Πλακέτα με σύνολο arrays τα οποία υβριδίζονται ταυτοχρόνως.

**Benzo(a)ylene ή B(a)P:** Ρυπογόνος παράγοντας (πολυκυκλικός αρωματικός υδρογονάνθρακας) συνδεδεμένος με καρκινογενετικές διαδικασίες.

**Data set:** Σύνολο δεδομένων που περιλαμβάνει τις εκφράσεις ανά probe για κάθε δείγμα, καθώς και τις απαραίτητες πληροφορίες για τη δειγματοληπτική διαδικασία που προηγήθηκε.

**Gene symbol:** Συντομογραφία η οποία υποδεικνύει ένα καλώς καθορισμένο γονίδιο

**Heatmap:** Οπτική αναπαράσταση δεδομένων (εντάσεων επιμέρους γονιδίων) κατά την οποία οι σχετικές τιμές τους αποδίδονται σε χρωματική κλίμακα, ενώ οι κατηγορίες των δεδομένων κατανέμονται ιεραρχικά στον κάθε άξονα. Επιτρέπει την αποτελεσματική εποπτεία της συμπεριφοράς των γονιδίων μεταξύ των δειγμάτων.

**Log fold change:** Ο λογαριθμισμένος λόγος των μέσων εκφράσεων ενός γονιδίου μεταξύ δύο κατηγοριών δειγμάτων.

**Principal component:** Βοηθητική μεταβλητή η οποία απεικονίζει στατιστικά σημαντικό ποσοστό της ετερογένειας ενός συνόλου δεδομένων.

**Probe:** Περιοχή της μικροσυστοιχίας η οποία ελέγχει την ύπαρξη μέρους καλώς καθορισμένης αλληλουχίας νουκλεοτιδίων.

**Λέξεις κλειδιά:** Βιοπληροφορική, ατμοσφαιρική ρύπανση, έκφραση γονιδίων, μικροσυστοιχίες, genomics, τοξικογενετική.

## Περίληψη

Ο σκοπός της εργασίας αυτής είναι πρωτίστως η διερεύνηση της πιθανότητας ύπαρξης γονιδίων που εκφράζονται διαφορετικώς παρουσία χρόνιας περιβαλλοντικής ρύπανσης βιομηχανικού τύπου στην ατμόσφαιρα, στο πόσιμο νερό, στην τροφή κλπ, με τη βοήθεια δεδομένων από μικροσυστοιχίες.

Για τη μελέτη αυτή επελέχθη το, δημοσιευμένο στη δημόσια βάση δεδομένων GEO, data set με κωδικό GSE60767. Πρόκειται για μια μελέτη που διενεργήθη στα πλαίσια του Τσέχικου Ινστιτούτου Institute of Experimental Medicine AS CR, στα πλαίσια σειράς μελετών επί των επιδράσεων της ατμοσφαιρικής ρύπανσης. Το συγκεκριμένο data set περιέχει δεδομένα έκφρασης λευκοκυττάρων από μικροσυστοιχίες της Illumina. Οι λόγοι επιλογής αυτού του data set:

- Αποτελεί μελέτη σύγκρισης μεταξύ των πόλεων Πράγα και Οστράβα, με την τελευταία να θεωρείται μία από τις πιο βαριά ρυπασμένες περιοχές εντός της Ευρωπαϊκής Ένωσης.
- Αποτελείται από αξισημείωτα υψηλό αριθμό δειγμάτων (468 δείγματα σε 6 δειγματοληπτικές περιόδους), πράγμα που θα ενισχύσει τη στατιστική σημαντικότητα των αποτελεσμάτων.
- Πρόκειται για έναν αρκετά ομοιογενή πληθυσμό από πλευράς διατροφικών συνηθειών και συνθηκών ζωής, περιορίζοντας τον αριθμό μεταβλητών που θα μπορούσαν να μειώσουν την καθαρότητα των αποτελεσμάτων.
- Το εργασιακό περιβάλλον των δοτών εξασφαλίζει πως υφίστανται το γενικό προφίλ ρύπανσης που υφίσταται το μεγαλύτερο μέρος του πληθυσμού και η ομοιογένεια του συνόλου των δοτών σε φύλο, ηλικία, εργασιακό περιβάλλον μειώνει περαιτέρω τον αριθμό των μεταβλητών που θα περιέπλεκαν την ανάλυση (περισσότερες λεπτομέρειες στην περιγραφή της διαδικασίας δειγματοληψιών).

Η ανάλυση των δεδομένων και η εξαγωγή διαφορετικώς εκφρασμένων γονιδίων έγινε από την αρχή στο περιβάλλον της γλώσσας προγραμματισμού R v3.3.1 και πακέτα ανοικτού κώδικα του Bioconductor, πράγμα που διευκολύνει την εποπτεία και τον έλεγχο της ανάλυσης.

Τα δεδομένα του GSE60767 κατ' αρχήν ανακτήθηκαν από το δημόσιο εναποθετήριο GEO με τη χρήση του πακέτου της R «GEOquery» και ύστερα αναλύθηκαν διεξοδικά με χρήση των πακέτων «limma» και «SVA». Μέσω αυτής της διαδικασίας αναγνωρίστηκαν γονίδια που εκφράζονται διαφορετικά μεταξύ των δύο πόλεων, με το βιολογικό σήμα που μετράται όμως να είναι αναγκαστικώς εξασθενημένο από τη διαδικασία διόρθωσης για batch effects, λόγω λάθους που είχε σημειωθεί κατά την πειραματική διαδικασία, όπως θα αναλυθεί και στο τρίτο κεφάλαιο.

Μετά την εξαγωγή λιστών στατιστικώς σημαντικά διαφορετικώς εκφρασμένων γονιδίων, χρησιμοποιήθηκε η πλατφόρμα BioInfoMiner της e-NiOS ώστε να διερευνηθούν οι ρόλοι τους στις οντολογίες Gene Ontology, Human Phenotype Ontology, MGI Mammalian Phenotype Ontology, Reactome Pathways Ontology, ώστε με τη σύγκριση αυτών να αναγνωριστούν, αφ' ενός οι συνδέσεις των αποτελεσμάτων της ανάλυσης με επιδημιολογικά δεδομένα και αφ' ετέρου να αναγνωριστούν πιθανά γονίδια-κόμβοι (hub genes), με κεντρικό ρόλο σε αρκετές βιολογικές διεργασίες, τα οποία

έχουν συχνά αναβαθμισμένη βιολογική σημασία. Τα αποτελέσματα παρουσιάζονται διεξοδικά στο τέταρτο κεφάλαιο.

Ύστερα ελέγχθη αν κάποιο υποσύνολο των hub genes που εντοπίστηκαν θα μπορούσε να χρησιμεύσει ως δείκτης συστημικής απόκρισης του οργανισμού σε ρυπασμένο περιβάλλον και την αποτελεσματικότητα αυτής της απόκρισης ή προσαρμογής. Λόγω των λαθών στην πειραματική μέθοδο και το αδύναμο βιολογικό σήμα που αυτές επάγουν, η επιτυχία σε αυτό το μέρος της έρευνας ήταν περιορισμένη και ήταν κυρίως στατιστικού χαρακτήρα. Παρ' όλα αυτά όμως, μπόρεσε να αναγνωρισθεί ένα βέλτιστο δίκτυο hub genes το οποίο αναδεικνύει τη διαφορά μεταξύ των προφίλ των δύο πόλεων σε βαθμό χαμηλότερο μεν, της ίδιας τάξης μεγέθους δε, με τη διακριτική ικανότητα που μπορούμε να επιτύχουμε χρησιμοποιώντας τα διαφορικά εκφρασμένα γονίδια μεταξύ των δύο πόλεων.

Τέλος, οι λίστες γονιδίων συγκρίθηκαν με την υπάρχουσα βιβλιογραφία μέσω της Συγκριτικής Τοξικογενομικής Βάσης Δεδομένων (Comparative Toxicogenomics Database, CTD, <http://ctdbase.org/>) ώστε να αναδειχθούν αλληλοεπικαλύψεις μεταξύ των αποτελεσμάτων της έρευνάς μας και ήδη στοιχειοθετημένων συνδέσεων μεταξύ της διαφορικής έκφρασης γονιδίων και των αιωρούμενων μικροσωματιδίων, καθώς και του B(a)P.

## Summary

The primary aim of this study is to investigate the possibility of genes which are differentially expressed in the presence of prolonged exposure to environmental pollution of an industrial kind (in the air, water, food etc.), using microarray data.

This study will focus on data set published in the open-access genomics database GEO, bearing the identifier GSE60767. It is a study carried out within the Czech Institute of Experimental Medicine AS CR, as part of a series of studies concerning the effects of environmental pollution. This specific data set contains expression data from leukocytes obtained through the use of Illumina microarrays. The reasons for choosing it were:

- Its being a study between the cities of Prague and Ostrava, the latter of which is considered to be one of the most heavily polluted areas within the European Union.
- It contains an uncommonly high number of samples (468 samples obtained in three discrete sample periods), a feature that will enhance the statistical significance of our findings.
- It refers to a relatively uniform population, in terms of dietary standards and living conditions, limiting the number of probable confounding variables.
- The donors' work environment guarantees their exposure to the pollution profile the general population is exposed to. At the same time, the donors' uniformity in terms of gender, age and line of employment further reduces the confounding variables (a more in-depth description will follow where the sampling procedure is explained).

Analysis and extraction of differentially expressed genes were carried out wholly using the R v3.3.1 programming language and open-source packages from Bioconductor, improving not only the ease, but the verifiability of analysis as well.

The GSE 60767 data were firstly retrieved from the open-access GEO repository using the R package "GEOquery" and were subjected to analysis through the use of the software packages "limma" and "SVA". Through this procedure, genes which were differentially expressed between the two cities were identified, the biological signal of which was inescapably weak due to a mistake in the experimental process and the batch effect correction procedure used, as will be explained in greater detail in chapter 3.

Following the extraction of statistically significant differentially expressed genes, the BiInfoMiner platform created by e-NiOS was used to explore their roles in Gene Ontology, Human Phenotype Ontology, MGI Mammalian Phenotype Ontology and Reactome Pathways Ontology, so that through the comparison of the results, connections between the results and existing epidemiological data could be identified, as well as genes playing a central role in more than one biological processes (hub genes), which often have an elevated biological importance.

Then, we explored the possibility of using a subset of the hub genes as a marker for the systemic response to a polluted environment. Due to the limitations introduced in the study and the ensuing weak biological signal, success in this endeavor was limited and of a statistical nature. Nevertheless, we managed to identify an optimal subset of hub genes that brings forth the difference between the expression profiles of the two cities, at a rate lower than, yet in the same order of magnitude with, the resolution attained by straightforwardly using the genes differentially expressed between these two cities.

Lastly, the gene lists were compared to existing literature through the use of the Comparative Toxicogenomics Database (CTD, <http://ctdbase.org/>) to illustrate overlaps between our results and already-established linkages between differential gene expression and particulate matter and Benzo(a)pyrene (“B(a)P”) pollution.



# Κεφάλαιο Πρώτο: Θεωρητικά Ζητήματα

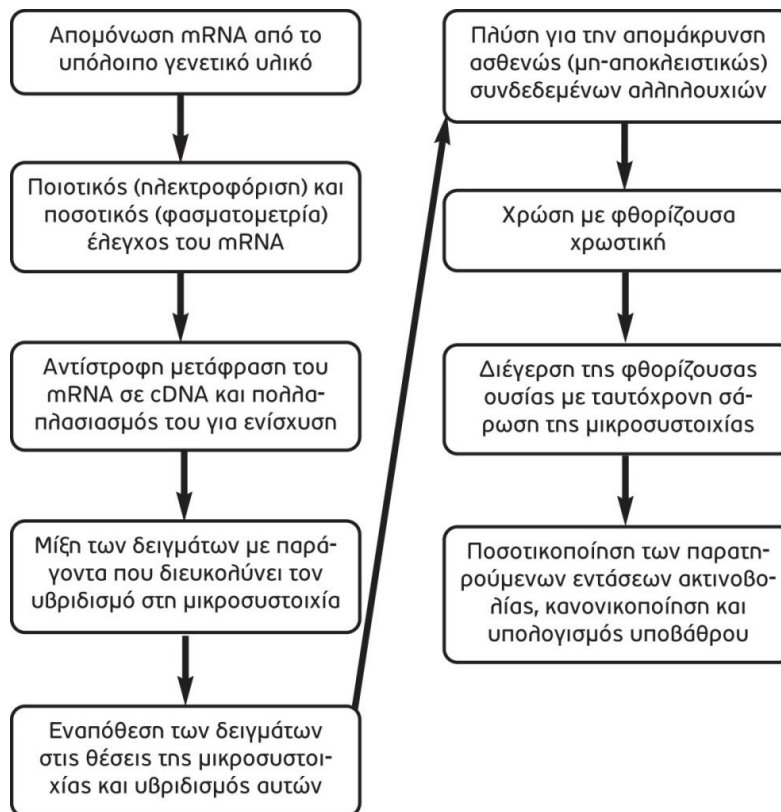
## 1.1 –Μικροσυστοιχίες για δεδομένα έκφρασης γονιδίων

Το βασικό ερώτημα στο οποίο απαντά αυτή η τεχνολογία είναι η παρακολούθηση και ποσοτικοποίηση των σχετικών επιπέδων έκφρασης αρκετών χιλιάδων γονιδίων ταυτοχρόνως από έναν, όσο το δυνατόν πιο συγκεκριμένο κυτταρικό πληθυσμό, ώστε με τη σύγκριση μεταξύ δειγματοληπτικών κατηγοριών να μπορεί να αναγνωριστεί διαφορική έκφραση γονιδίων σχετική με τις πειραματικές μεταβλητές που θέτουμε.

Προς αυτό το σκοπό, η κάθε πλακέτα μικροσυστοιχιών αποτελείται από «ενεργές περιοχές» με μικρές ποσότητες συγκεκριμένων αλληλουχιών DNA, ονομαζόμενες «probes», σε καλώς καταγεγραμμένες θέσεις πάνω σε στερεή επιφάνεια όπως γυαλί ή πλαστικό. Η αρχή λειτουργίας αυτών είναι η ιδιότητα των κλώνων DNA να εντοπίζουν το συμπληρωματικό τους κλώνο και να συνδέονται μαζί του με μεγάλο ποσοστό επιτυχίας. Έτσι, το κάθε probe καθορίζεται ως περιοχή με το συμπληρωματικό κάποιου μέρους cDNA, το οποίο μπορεί να αποτελεί τμήμα γονιδίου που κωδικοποιεί πρωτεΐνη, ή όχι.

Η γενική διαδικασία ξεκινά με τη λήψη των δειγμάτων και την επιβεβαίωση της ομοιομορφίας τους όσον αφορά τη σύνθεση του κυτταρικού πληθυσμού. Ύστερα ακολουθείται ένα πρωτόκολλο που κινείται στις εξής γενικές γραμμές, όσον αφορά την έρευνά μας (εικόνα 1.1.1):

Εικόνα 1.1.1



Κατά το τελευταίο στάδιο της διαδικασίας αυτής, το Illumina Bead Studio Software v3.3.7 με το οποίο ελήφθησαν οι μετρήσεις και αναλύθηκαν οι φωτογραφίες των μικροσυστοιχιών εκμεταλλεύεται τα εξής στοιχεία της δομής των chip της Illumina:

- Την ύπαρξη negative control probes, τα οποία διευκολύνουν την εξαγωγή μιας αξιόπιστης τιμής για το θόρυβο υποβάθρου.
- Την ύπαρξη αρκετών αντιτύπων του κάθε probe σε κάθε μικροσυστοιχία, με τη σύγκριση μεταξύ των τιμών που δίνουν να επιτρέπει την εξαγωγή της αξιοπιστίας της μετρούμενης φωτεινότητας του κάθε probe. Αυτή δίνεται ως πιθανότητα λανθασμένης ανίχνευσης (false discovery rate—FDR) στο διάστημα [0,1], με το μηδέν να υποδεικνύει απόλυτη αξιοπιστία της μέτρησης.

Το αποτέλεσμα αυτής της διαδικασίας είναι ένας πίνακας με τις παρατηρηθείσες σχετικές εντάσεις των probes, τις αξιοπιστίες αυτών και μια σειρά πληροφορίες περί της απόδοσης της μικροσυστοιχίας. Από αυτό το σημείο ξεκινά και η έρευνά μας, η οποία αφορά την ανάλυση των αποτελεσμάτων, την ανάδειξη διαφορικών εκφρασμένων γονιδίων και την ερμηνεία των μοτίβων που τυχόν εμφανίζονται, τόσο σε επίπεδο ανάλυσης και σχέσεων μεταξύ των δειγμάτων του πληθυσμού, όσο και σε επίπεδο λειτουργικής ανάλυσης και σχέσεων μεταξύ των αναδεικνυόμενων γονιδίων.

Από αυτό το σημείο και μετά, η ανάλυση των αποτελεσμάτων πρέπει να προχωρήσει υπό τις εξής υποθέσεις:

- Η ψύξη και διατήρηση των δειγμάτων σε διαφορετικές περιόδους και για διαφορετικά χρονικά διαστήματα δεν αλλοίωσε τα δείγματα.
- Η απομόνωση και ενίσχυση του mRNA δεν αλλοίωσε συστηματικά τις σχετικές αφθονίες των επιμέρους κλώνων που υπήρχαν στο αρχικό μας δείγμα [1].
- Οι σχετικές φωτεινότητες φθορισμού των υβριδισμένων beads είναι ανάλογες των σχετικών αφθονιών των συγκεκριμένων αλληλουχιών στο αρχικό δείγμα

Οι υποθέσεις αυτές είναι καλώς μελετημένες και μπορούν να θεωρηθούν αξιόπιστες στη διαδικασία της ανάλυσης.

## 1.2 – Θεωρητικά επί των εργαλείων της ανάλυσης

### Προ-επεξεργασία

Η προ-επεξεργασία των δεδομένων από μονοκαναλικές μικροσυστοιχίες της Illumina κινείται γύρω από τέσσερα σαφή βήματα: Διόρθωση θορύβου υποβάθρου, κανονικοποίηση ποσοστημορίων, μετασχηματισμό των τιμών των δεδομένων και τέλος, περίληψη του συνόλου των probes.

Στη μελέτη μας, αυτό το στάδιο του pre-processing υλοποιήθηκε με τη μέθοδο `neqc` του πακέτου `limma`, αφ' ότου τα δεδομένα διαβάστηκαν με τη χρήση της μεθόδου `read.ilmn[2]`.

### Θόρυβος υποβάθρου:

Είναι γενικά αποδεκτό πως για το σήμα μικροσυστοιχιών, ο θόρυβος υποβάθρου μπορεί να μοντελοποιηθεί ως μία γραμμική κατανομή, ενώ το ίδιο το σήμα ως μία εκθετική. Ως εκ τούτου, το καθάριστο σήμα, το οποίο αποτελεί το άθροισμα θορύβου και βιολογικού σήματος, μοντελοποιείται ως μία συνέλιξη μεταξύ ενός γραμμικού και ενός εκθετικού όρου, ο υπολογισμός των οποίων απαιτεί τον υπολογισμό τριών παραμέτρων: τη μέση τιμή  $\mu$  και την απόκλιση  $\sigma$  των εντάσεων υποβάθρου, καθώς και τη μέση τιμή  $\alpha$  των εντάσεων του σήματος. Η προσέγγιση που ακολουθεί η `neqc` για τον υπολογισμό αυτών των παραμέτρων είναι η εξής [3]:

$$\begin{aligned}\hat{\mu} &= \bar{b} \\ \hat{\sigma} &= s_b \\ \hat{\alpha} &= \bar{y} - \bar{b}\end{aligned}$$

Όπου

$\bar{b}$ : μέση τιμή υποβάθρου

$s_b$ : κανονική απόκλιση υποβάθρου

$\bar{y}$ : μέση τιμή των εντάσεων των κανονικών probes

### Κανονικοποίηση ποσοστημορίων:

Η διαδικασία αυτή εξασφαλίζει πως οι στατιστικές κατανομές των τιμών αποκτούν τα ίδια χαρακτηριστικά, ώστε να μπορούν τα δείγματα να συγκριθούν μεταξύ τους. [4]

### Μετασχηματισμός των δεδομένων:

Οι σχετικές εντάσεις συχνά κινούνται σε ένα διάστημα  $10^{-10}$  έως  $10^4$ , ενώ έχει παρατηρηθεί πως εμφανίζουν μια εκθετική κατανομή. Ως εκ τούτου, η λογαρίθμηση των τιμών επιτρέπει τη με μεγαλύτερη ευκολία σύγκρισή τους κι εξασφαλίζει μια κοντύτερα σε γραμμική μεταξύ τους σχέση, διευκολύνοντας τις επόμενες φάσεις της ανάλυσης. Η γενική σύμβαση σε αυτό το στάδιο, η οποία ακολουθείται κι εδώ, είναι η λογαρίθμηση στη βάση του 2.

### Φιλτράρισμα των probes

Ένα δεύτερο ζήτημα που ανακύπτει είναι ποιά από τα probes για τα οποία έχουμε μετρήσεις ( $\sim 48.000$ /δείγμα) θα χρησιμοποιηθούν για την περαιτέρω μελέτη. Δεν πρόκειται για ένα τετριμμένο ερώτημα, καθώς η παρουσία probes χαμηλής αξιοπιστίας, ή που υβριδίζονται συχνά με αποσπάσματα αλυσίδων μπορεί να αλλοιώσει τις σχέσεις μεταξύ των δειγμάτων και να οδηγήσει σε λαθεμένα συμπεράσματα. Χωρίς αμφιβολία, με δεδομένο πως οι μικροσυστοιχίες ως τεχνολογία έχουν υψηλό ποσοστό θορύβου σε διάφορα επίπεδα της ανάλυσης, κάνει σημαντικό το να παρθούν όσο το δυνατόν ορθότερα μέτρα για τον περιορισμό του. Γι αυτό το λόγο υιοθετήθηκε ένα διπλό φιλτράρισμα των probes το οποίο βασίζεται τόσο στην αντιμετώπιση τόσο τεχνικής, όσο και της βιολογικής αναξιοπιστίας. Σε αυτή τη διαδικασία, τα δεδομένα του annotation προέρχονται από το πακέτο της R ονόματι illuminaHumanv3.db [5].

Ένα τρίτο ζήτημα είναι η μέτρηση της απόδοσης του κάθε array όσον αφορά την απόδοσή του. Αυτή η διαδικασία μπορεί να δώσει μία τιμή η οποία δείχνει το βάρος που θα πρέπει να έχει το συγκεκριμένο δείγμα στη μετέπειτα ανάλυση, με το μηδέν να υποδεικνύει τον αποκλεισμό του. Αυτή η διαδικασία επιτρέπει τη βελτίωση της ποιότητας της ανάλυσης και συνηθίζεται ειδικά στα δείγματα αίματος, τα οποία έχουν και πιο μεταβλητή σύνθεση του κυτταρικού πληθυσμού. Αυτόν τον υπολογισμό εκτελεί η arrayWeights της limma [6].

### 1.3 – Ανάλυση

#### Batch effects

Μία από τις σημαντικές εργασίες στο χώρο της βιοπληροφορικής ήταν η αναγνώριση και θεμελίωση των batch effects [7]. Συνοπτικά, πρόκειται για την τάση των δειγμάτων να σχηματίζουν συστάδες (batches) με βάση παράγοντες που δεν έχουν σχέση με τις πειραματικές μεταβλητές που θέτει η έρευνα (και είναι συχνά τεχνικής φύσεως). Η ουσία της μελέτης επί των batch effects βρίσκεται κυρίως στην αναγνώριση του γεγονότος ότι συχνά είναι αρκετά έντονη η επίδρασή τους ώστε να γίνονται σαφώς ισχυρότερα από το βιολογικό σήμα το οποίο προσπαθούμε να μετρήσουμε, καταλήγοντας στο συμπέρασμα ότι αρκετές προγενέστερες έρευνες ανίχνευαν κυρίως τα batch effects της πειραματικής τους διαδικασίας, παρά το όποιο βιολογικό σήμα. Για να δωθεί μία ποιοτική αίσθηση του φαινομένου, έχουν αναγνωριστεί ως παράγοντες batch effect η ημέρα κατά την οποία διενεργείται η ανάλυση, τα επίπεδα ρύπανσης, η ποιότητα της κάθε μικροσυστοιχίας χωριστά, έως και το ποιός τεχνικός τις προετοίμασε. Έτσι, στην ανάλυση που θα ακολουθήσει, μετά την αναγνώριση πιθανών batch effects, χρησιμοποιείται η μέθοδος ComBat [8] του πακέτου sva [7] για την αντιμετώπιση αυτών, καθώς μπορεί να είναι δυνητικά ο σημαντικότερος παράγοντας.

Η διαδικασία αυτή γίνεται σε τρία βήματα:

- Κανονικοποίηση των δεδομένων ώστε οι εκφράσεις των γονιδίων να έχουν αντίστοιχους μέσους όρους και αποκλίσεις, καθώς οι διαφορές αυτές μπορούν να επηρεάσουν την πορεία της εμπειρικής μεθόδου Bayes που ακολουθείται για την εκτίμηση του μεγέθους του batch effect.
- Εμπειρική Bayes εκτίμηση της συνεισφοράς του batch effect ως εξής:  
γ<sub>ig</sub>: η προσθετική συνεισφορά του batch effect

$\delta_{ig}$ : η πολλαπλασιαστική συνεισφορά του batch effect

Με την υπόθεση πως οι κατανομές τους είναι οι ακόλουθες:

$$\gamma_{ig}^* = \frac{n_i \bar{\tau}_i^2 \hat{\gamma}_{ig} + \delta_{ig}^{2*} \bar{\gamma}_i}{n_i \bar{\tau}_i^2 + \delta_{ig}^{2*}} \quad \delta_{ig}^{2*} = \frac{\bar{\theta}_i + \frac{1}{2} \sum_j (Z_{ijg} - \gamma_{ig}^*)^2}{\frac{n_j}{2} + \bar{\lambda}_i - 1}$$

Υπολογίζονται ως εξής:

$$\gamma_{ig} \sim N(Y_i, \tau_i^2) \quad \delta_{ig}^2 \sim \text{Inverse Gamma}(\lambda_i, \theta_i).$$

- Υπολογισμός του καθαρού σήματος:  $\gamma_{ijg}^* = \frac{\hat{\sigma}_g}{\hat{\delta}_{ig}} (Z_{ijg} - \hat{\gamma}_{ig}^*) + \hat{a}_g + X \hat{\beta}_g$ .

### SVA analysis

Η αντιμετώπιση batch effects με τη χρήση της ComBat προϋποθέτει μία σωστή εκτίμηση από πλευράς του ερευνητή για το ποιά είναι αυτά. Δεν αποκλείεται όμως να υπάρχουν batch effects τα οποία να έχουν διαφύγει, ή και «κρυφές μεταβλητές»: μεταβλητές οι οποίες δεν έχουν μετρηθεί στα πλαίσια της πειραματικής διαδικασίας ή της ανίχνευσης των batch effects, αλλά που συνεχίζουν να επηρεάζουν την κατανομή των αποτελεσμάτων με στατιστικά σημαντικό τρόπο. Η surrogate variable analysis εν συντομία αναλύει τί ποσοστό της ετερογένειας του data set εξηγεί κάθε μεταβλητή, και εφαρμοζόμενη μετά τη διόρθωση των batch effects, ανιχνεύει αν υπάρχει ανάγκη εισαγωγής περισσότερων μεταβλητών ώστε να μπορεί να εξηγηθεί η κατανομή των τιμών των δεδομένων. Σε περίπτωση που κριθεί απαραίτητη η εισαγωγή νέων μεταβλητών στο μοντέλο που θα εξηγεί τα δεδομένα, αυτές μπορούν να υπολογιστούν. Οποιαδήποτε πιθανή εξήγηση για τη φύση αυτών των επιπλέον μεταβλητών πρέπει να αναζητηθεί με επιπλέον υποθέσεις και πειράματα, καθώς η μόνη πληροφορία που παρέχεται για αυτές είναι η σχετική τους συνεισφορά στην εξήγηση της ετερογένειας.

Ο έλεγχος για τυχόν επιπλέον μεταβλητές και ο υπολογισμός τους έγιναν με τη χρήση των εντολών num.sv και sva αντιστοίχως, περιεχόμενες στο πακέτο SVA της R [9] [10]. Μπορεί ως διαδικασία να περιγραφεί ως αλληλουχία των εξής τεσσάρων σταδίων:

- Θεώρηση των πειραματικών μεταβλητών ως «θορύβου» και αφαίρεσή τους από το προφίλ έκφρασης, ώστε να απομονωθεί το λανθάνον προφίλ ετερογένειας της γονιδιακής έκφρασης. Ανάλυση του πίνακα των εκφράσεων σε ένα σύνολο ορθομοναδιαίων διανυσμάτων που αναπαράγουν πλήρως τη μορφή του. Εφαρμογή στατιστικού τεστ ώστε να αναγνωριστούν τα μοναδιαία διανύσματα που αναπαράγουν μεγαλύτερο μέρος της ετερογένειας της γονιδιακής έκφρασης απ' όσο μπορεί να δικαιολογηθεί ως «τυχαίο».
- Ανάλυση σημαντικότητας των σχέσεων μεταξύ των γονιδίων και της λανθάνουσας ετερογένειας, με σκοπό την αναγνώριση του υποσυνόλου των γονιδίων που βρίσκονται πίσω από κάθε ορθογώνια συνιστώσα της ετερογένειας έκφρασης.
- Δημιουργία μιας βοηθητικής μεταβλητής με βάση την πλήρη ετερογένεια, για κάθε τέτοιο υποσύνολο γονιδίων.
- Συμπερίληψη των στατιστικά σημαντικών βοηθητικών μεταβλητών ως συμεταβλητών στη συνέχεια της ανάλυσης.

### Γραμμικά μοντέλα

Η υπόθεση που γίνεται σε αυτό τη σημείο της ανάλυσης είναι πως το τελικό προφίλ έκφρασης γονιδίων που παρατηρούμε μπορεί να αναλυθεί γραμμικά σε άθροισμα επιμέρους προφίλ, ένα ανά πειραματική μεταβλητή, τα οποία επιμερίζουν τις συνολικές διαφορικές εκφράσεις που παρατηρούμε μεταξύ αυτών των μεταβλητών. Η διαμόρφωση του πίνακα με βάση τον οποίο θα υπολογιστεί το γραμμικό μοντέλο (design matrix) και κυρίως οι μετέπειτα συγκρίσεις ως πράξεις μεταξύ των επιμέρους προφίλ έκφρασης καθοδηγούνται από τα ερωτήματα που θέτουμε και τους περιορισμούς που έχει θέσει η πρωτύτερη διαδικασία ανάλυσης.

Σε αυτό το σημείο αξίζει να γίνει μία ειδική αναφορά στην αντιμετώπιση των batch effects: μία μεθοδολογία ή οποία χρησιμοποιείται και με την οποία δε συμφωνούμε, είναι η άνευ όρων ένταξη των batch effects στο γραμμικό μοντέλο για τον υπολογισμό τους. Ο λόγος που επιλέξαμε την ξεχωριστή αντιμετώπισή τους με τη χρήση της ComBat είναι πως ο αλγόριθμός της είναι δομημένος ώστε να υπολογίζει τη συνεισφορά των batch effects θεωρώντας ότι η συνεισφορά τους είναι κατ' αρχήν τυχαία, δηλαδή πως δε μπορούμε να γνωρίζουμε εκ των προτέρων αν η επίδραση του batch effect θα είναι θετική η αρνητική. Αντιθέτως, τα γραμμικά μοντέλα λειτουργούν με την παραδοχή πως η επίδραση κάθε μεταβλητής που περιέχουν θα είναι συστηματική και κατ' αρχήν εκ των προτέρων προβλέψιμη. Ως εκ τούτου, ενδέχεται να εξομαλύνουν υπερβολικά οποιεσδήποτε διαφορές θα έπρεπε να παρατηρηθούν, στην προσπάθειά τους να εντάξουν την τυχαία επίδραση στο γραμμικό μοντέλο. Αυτό το στάδιο της ανάλυσης αποπερατώθη με τη χρήση της lmfitt της limma [11].

### Λίστες γονιδίων

Μετά τον υπολογισμό του γραμμικού μοντέλου ακολουθεί ο ορισμός των συγκρίσεων που θα πρέπει να γίνουν μεταξύ των κατηγοριών που ορίζουν οι πειραματικές μεταβλητές, ώστε να εξαχθούν τα προφίλ διαφορικής έκφρασης που θέλουμε να μελετήσουμε περαιτέρω, καθώς και ο βαθμός εμπιστοσύνης σε αυτά. Αυτή η διαδικασία αποπερατώνεται με τη χρήση της μεθόδου eBayes της limma [11], η οποία με τη χρήση μεθόδου empirical Bayes υπολογίζει μία λίστα με τα γονίδια που εμφανίζουν συστηματική διαφορική έκφραση. Ακολουθεί η εξαγωγή από αυτή, με τη χρήση της μεθόδου topTable (του ίδιου πακέτου) [11] της λίστας με τα διαφορικά εκφρασμένα γονίδια του προφίλ διαφορικής έκφρασης που έχουμε δομήσει. Εδώ μπορούν να τεθούν οι απαιτήσεις μας για τη βιολογική σημαντικότητα και τη στατιστική αξιοπιστία των αποτελεσμάτων τα οποία θα δεχτούμε για ανάλυση. Οι τιμές που επελέγησαν για τις δύο αυτές παραμέτρους θα αιτιολογηθούν στο τέταρτο κεφάλαιο.

## **1.4 – Λειτουργική ανάλυση**

### Οντολογίες

Μία οντολογία επιτρέπει τη μοντελοποίηση ενός τομέα με κοινούς όρους αναφοράς, καθορίζοντας τα αντικείμενα που αποτελούν αυτόν τον τομέα, τις ιδιότητες και τις μεταξύ τους σχέσεις. Ο λόγος για τον οποίο είναι ιδιαίτερη η σημασία τους στην ανάλυσή μας είναι η προσπάθεια να χρησιμοποιηθούν στο

χώρο της βιοπληροφορικής ώστε να περιγράψουν σύνθετα βιολογικά συστήματα και δίκτυα και να επιτρέψουν την εξαγωγή δεδομένων και συμπερασμάτων υψηλού επιπέδου από περίπλοκους ή/και μεγάλους όγκους δεδομένων.

Στην προσπάθεια αυτή έχουν δομηθεί μία σειρά οντολογίες, μεταξύ των οποίων θα χρησιμοποιήσουμε τις: Gene Ontology, Human Phenotype Ontology, MGI Mammalian Phenotype Ontology, Reactome Pathways Ontology.

Κάθε μία από αυτές τις (συνεχώς εξελισσόμενες) οντολογίες αποτελείται από όρους που δημιουργούνται από τις ιδιότητες των βιολογικών προϊόντων των γονιδίων, με τον κάθε όρο να χαρακτηρίζεται από μια σαφή σειρά πληροφοριών όπως το όνομά του, ένα αλφαριθμητικό αναγνωριστικό, αναφορά σε βιβλιογραφία που αιτιολογεί τη σύνδεση του όρου με το γονίδιο, αναφορά στον τομέα στον οποίο ανήκει, αναφορές ίσως σε άλλες βάσεις δεδομένων κ.α.

Στη λειτουργική ανάλυσή μας χρησιμοποιήθηκε η πλατφόρμα BioInfoMiner [12]. Η πλατφόρμα υποβάλλει την κάθε λίστα διαφορετικώς εκφρασμένων γονιδίων σε υπεργεωμετρικό τεστ και χρησιμοποιώντας έναν bootstrapping αλγόριθμο εμφανίζει τη λίστα των όρων της οντολογίας οι οποίοι κρίνονται στατιστικώς σημαντικά υπερεκπροσωπημένοι στην εισαγόμενη λίστα γονιδίων και με ισχυρό βιολογικό περιεχόμενο.

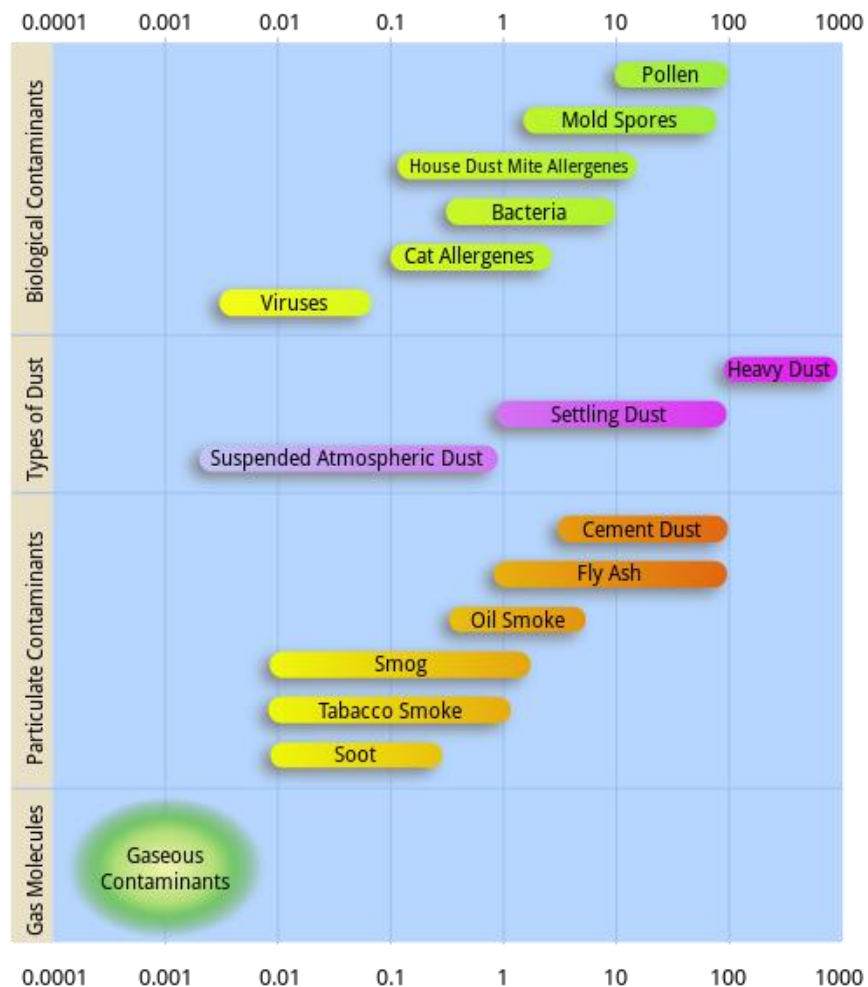
Σε ένα επόμενο επίπεδο, υπάρχει η δυνατότητα να εξαχθούν δενδρογράμματα που δείχνουν τις σχέσεις μεταξύ των οντολογικών όρων που έχουν αναγνωριστεί, οπτικοποιητική διαδικασία που μπορεί να βοηθήσει ιδιαίτερα οποιαδήποτε περαιτέρω μελέτη και ερμηνεία των αποτελεσμάτων.

### 1.5 – Περιβαλλοντική ρύπανση: αιωρούμενα μικροσωματίδια (particulate matter)

Ένας σημαντικός παράγοντας ρύπανσης, στον οποίο γίνεται τα τελευταία χρόνια έντονη έρευνα, είναι τα αιωρούμενα μικροσωματίδια (particulate matter).

Το χαρακτηριστικό τους γνώρισμα είναι το μικρό τους μέγεθος, το οποίο είναι στην τάξη των μm και τους επιτρέπει τη διείσδυση σε ιστούς. Στα μεγέθη της τάξης των 10μm, μπορούν να παρεισφρήσουν στους πνεύμονες [13], στα 2.5 μm έχουν τη δυνατότητα να περάσουν μέσω του ιστού των πνευμόνων στην κυκλοφορία του αίματος [14] και μπορούν να διαπεράσουν τον αιματοεγκεφαλικό φραγμό, ενώ στα 0.1 μm μπορούν να εισέλθουν και εντός μεμονομένων κυττάρων [15].

Μία ενδεικτική εικόνα των τύπων και μεγεθών καθημερινής ρύπανσης δίνεται στο παρακάτω γράφημα (σε κλίμακα μm) :



Εικόνα 1.5.1  
(πηγή: [Wikipedia](#))



Οι βιολογικές επιδράσεις των ρύπων αυτού του είδους δεν περιορίζονται στην παράμετρο του μεγέθους τους και της διεισδυτικής τους ικανότητας. Συχνά δρουν ως «οχήματα» ρύπανσης, μεταφέροντας ουσίες που απορροφούνται στην επιφάνειά τους, όπως πολυκυκλικούς αρωματικούς υδρογονάνθρακες και μέταλλα.

Μέσω της δράσης τους ως φορέων πολυκυκλικών αρωματικών υδρογονανθράκων, μπορούν να συνδεθούν με αυξημένο κίνδυνο καρκίνου, στο μέτρο που μεταφέρουν τέτοιες καρκινογόνες ενώσεις [16] [17].

Η αποβολή τους από τον οργανισμό είναι μια αργή διαδικασία, καθώς υπάρχουν ενδείξεις πως για διάστημα άνω των πέντε μηνών μετά την έκθεση σε μικροσωματιδιακούς ρύπους, το 1/3 της αρχικής συγκέντρωσης παραμένει παρόν και βιολογικά ενεργό [18]. Επίσης θεωρείται δυνατή συχνά και η πρόκληση συστημικής δράσης στον οργανισμό, στοχεύοντας το καρδιαγγειακό και άλλα όργανα.

Παρότι θα θέλαμε να επικεντρωθούμε σε αυτόν τον τύπο ρύπανσης, ο μικρός αριθμός δειγματοληπτικών περιόδων (χειμώνας 2009, καλοκαίρι 2009 και χειμώνας 2010) μας επιτρέπει μεν να κάνουμε εκτιμήσεις για την επίδραση ή μη της ατμοσφαιρικής ρύπανσης, αλλά δεν επιτρέπει να ξεχωρίσουμε την επίδραση κάποιου από τους παράγοντες ρύπανσης από την επίδραση οποιουδήποτε άλλου, παρατηρώντας μακροπρόθεσμα τα αποτελέσματα όταν οι δείκτες ρύπανσης για διάφορους ρύπους αλλάζουν με ανόμοιο μεταξύ τους τρόπο.

Ακόμα, η συνολική βαρύτητα της ατμοσφαιρικής ρύπανσης έχει συνδεθεί με σύνδρομα νευρολογικού χαρακτήρα [19] όπως ενδεικτικά η πέδηση της παιδικής νευρολογικής ανάπτυξης [20] [21], αναπνευστικά, καρδιαγγειακά, αναπαραγωγικά, ανοσολογικά κι αιματολογικά σύνδρομα [22]

Ακόμα και με αυτά τα δεδομένα, αν είχαμε στοιχεία για τις ακριβείς εκθέσεις του κάθε δότη σε διάφορους ρύπους, θα μπορούσε να οργανωθεί ένα μοντέλο το οποίο να επιμερίζει την παρατηρούμενη διαφορική έκφραση μεταξύ τους, επιτρέποντας (τουλάχιστον κατ' αρχήν) να απομονώσουμε τα αποτελέσματα του κάθε ενός. Καθώς τέτοια δεδομένα δεν είναι διαθέσιμα όμως, θα πρέπει να προχωρήσουμε θεωρώντας την πόλη ως μια δυαδική μεταβλητή, με κύρια διαφορά την ύπαρξη βαριάς βιομηχανίας, με τη ρύπανση που αυτή συνεπάγεται και την οποία μπορούμε να ποσοτικοποιήσουμε μέσω των μετρήσεων του Τσέχικου Υδρομετεωρολογικού Ινστιτούτου (<http://portal.chmi.cz/>).

### **Επεκτασιμότητα των αποτελεσμάτων**

Σύμφωνα με τα δεδομένα του Τσέχικου Υδρομετεωρολογικού Ινστιτούτου, οι μέσες συγκεντρώσεις τριών αρκετά ενδιαφέροντων ρύπων τις περιόδους που προηγήθηκαν των δειγματοληψιών για την Οστράβα (χρησιμοποιήθηκαν οι μετρήσεις των σταθμών Bartovice και Zabreh) παρουσιάζονται στον πίνακα 3.1.1.

Pollutant	PM <sub>10</sub> (μg/m <sup>3</sup> )	PM <sub>2.5</sub> (μg/m <sup>3</sup> )	B(a)P (ng/m <sup>3</sup> )
4ο τρίμηνο 2008	52.95	38.33	19.00
1ο τρίμηνο 2009	54.20	43.30	18.53
2ο τρίμηνο 2009	34.30	23.08	2.93
3ο τρίμηνο 2009	34.10	23.87	2.33
4ο τρίμηνο 2009	45.90	41.70	12.73
1ο τρίμηνο 2010	86.00	66.93	14.56

Πίνακας 1.5.1

Μπορεί να τεθεί ερώτημα, όσον αφορά το αν τα πιθανά αποτελέσματα που θα προκύψουν από τη μελέτη μας είναι επεκτάσιμα και στην Ελλάδα, ή και σε άλλες χώρες της Ευρωπαϊκής Ένωσης. Κατ' αρχάς, μια σειρά μελετών [23] [24] [25] πάνω στις συγκεντρώσεις και τη σύνθεση των ρύπων που εντοπίζονται στο λεκανοπέδιο Αττικής δίνουν τιμές συγκεντρώσεων ρύπων σε συγκρίσιμα επίπεδα με της Οστράβας. Συγκεκριμένα, παρατηρούνται τιμές PM<sub>2.5</sub> της τάξης των 25-37 μg/m<sup>3</sup> για το 2006 (προ κρίσης) , έχοντας μάλιστα ημερήσιο μέσο όρο τα 24-29 μg/m<sup>3</sup> εντός του αστικού κέντρου στο επίπεδο του εδάφους, 14-29 μg/m<sup>3</sup> για το 2010. Το 2013, η μέση συγκέντρωση PM<sub>2.5</sub> ανήλθε στα 21 μg/m<sup>3</sup>. Όσον αφορά τα PM<sub>10</sub>, το καλοκαίρι του 2000 ανήλθαν στα 75.5 μg/m<sup>3</sup> [26].

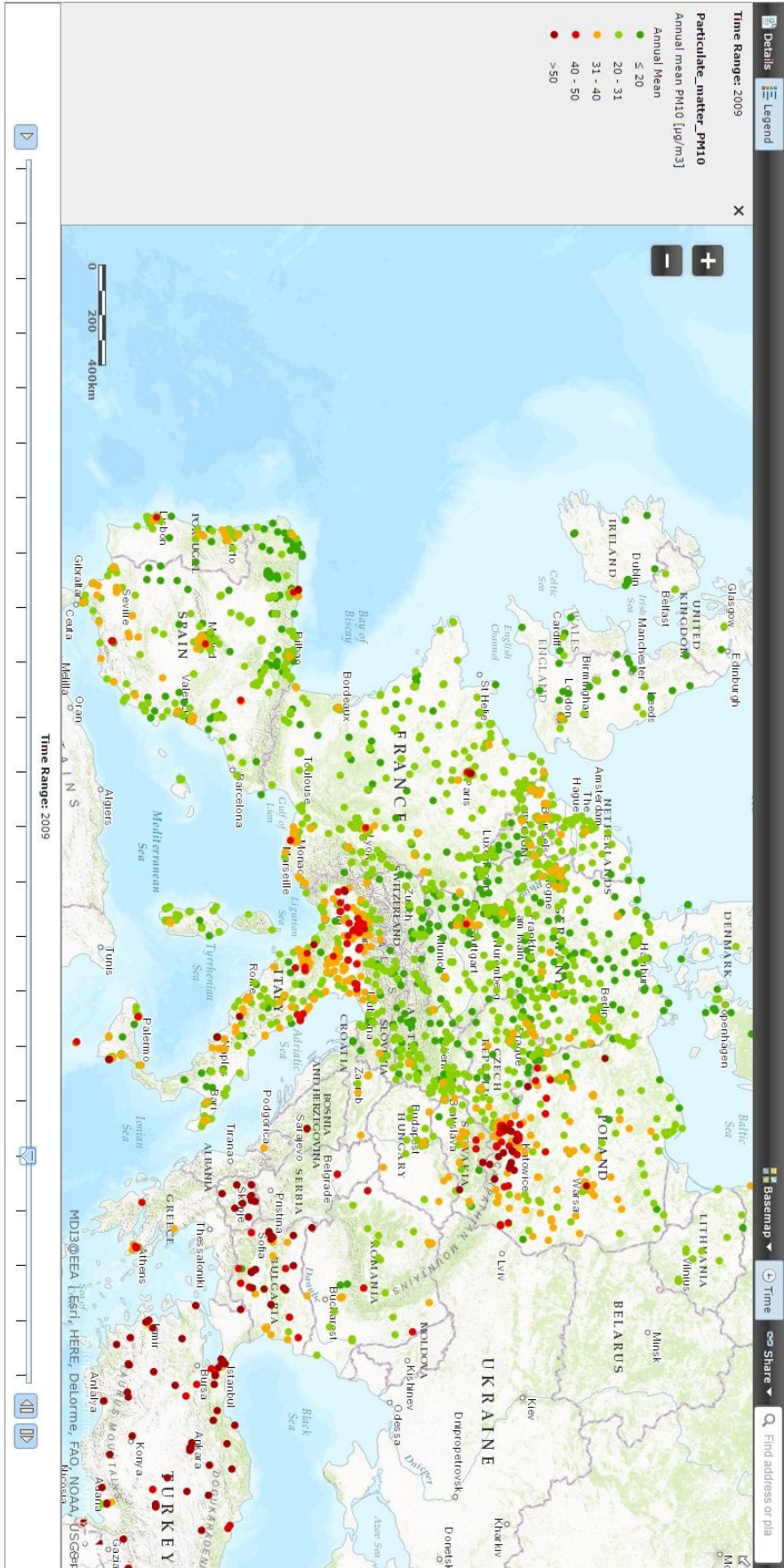
Πέρα από την ένταση του φαινομένου, η οποία ήδη μπορεί να θεωρηθεί σημαντική, η ποιότητα των ρύπων επίσης αλλάζει, καθώς η τιμολογιακή πολιτική στο πετρέλαιο θέρμανσης έχει αυξήσει την καύση ξύλου για οικιακή θέρμανση [27].

Ακολουθούν χάρτες (εικόνες 1.5.2-1.5.5) από την Ευρωπαϊκή Περιβαλλοντική υπηρεσία (<http://www.eea.europa.eu/>), που απεικονίζουν τα επίπεδα PM<sub>2.5</sub> και PM<sub>10</sub> τις χρονιές 2009 και 2010 στην Ευρωπαϊκή Ένωση, ώστε να μπορούν αντιπαραβληθούν τα επίπεδα μόλυνσης μεταξύ των διαφόρων περιοχών.

Με βάση αυτά, μπορούν να προταθούν περιοχές στις οποίες θα είχε ενδιαφέρον η διενέργεια νέων μελετών ώστε να μπορούν να επιβεβαιωθούν και να διαχωριστούν οι επιδράσεις των διαφόρων ρύπων.

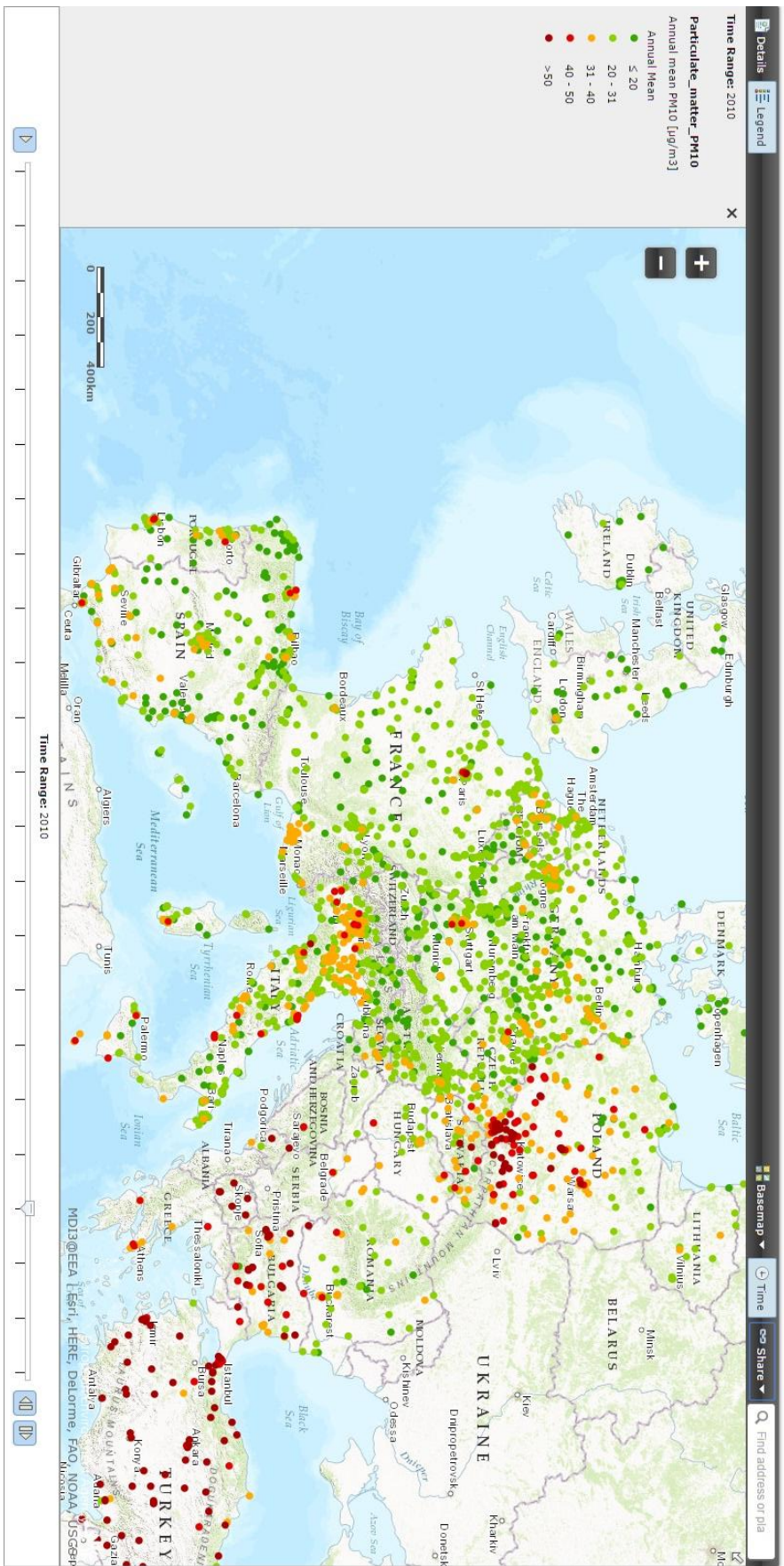
Επίσης, αποτελούν ένα επιπλέον τεκμήριο που στηρίζει τη θέση ότι η Αθήνα είναι μία από τις βαριά ρυπασμένες περιοχές της Ε.Ε. και πως τα αποτελέσματά μας θα είναι κατ' αρχήν επεκτάσιμα και στον ελληνικό πληθυσμό.

# Particulate Matter (PM10) in Europe



Εικόνα 1.5.2

# Particulate Matter (PM10) in Europe

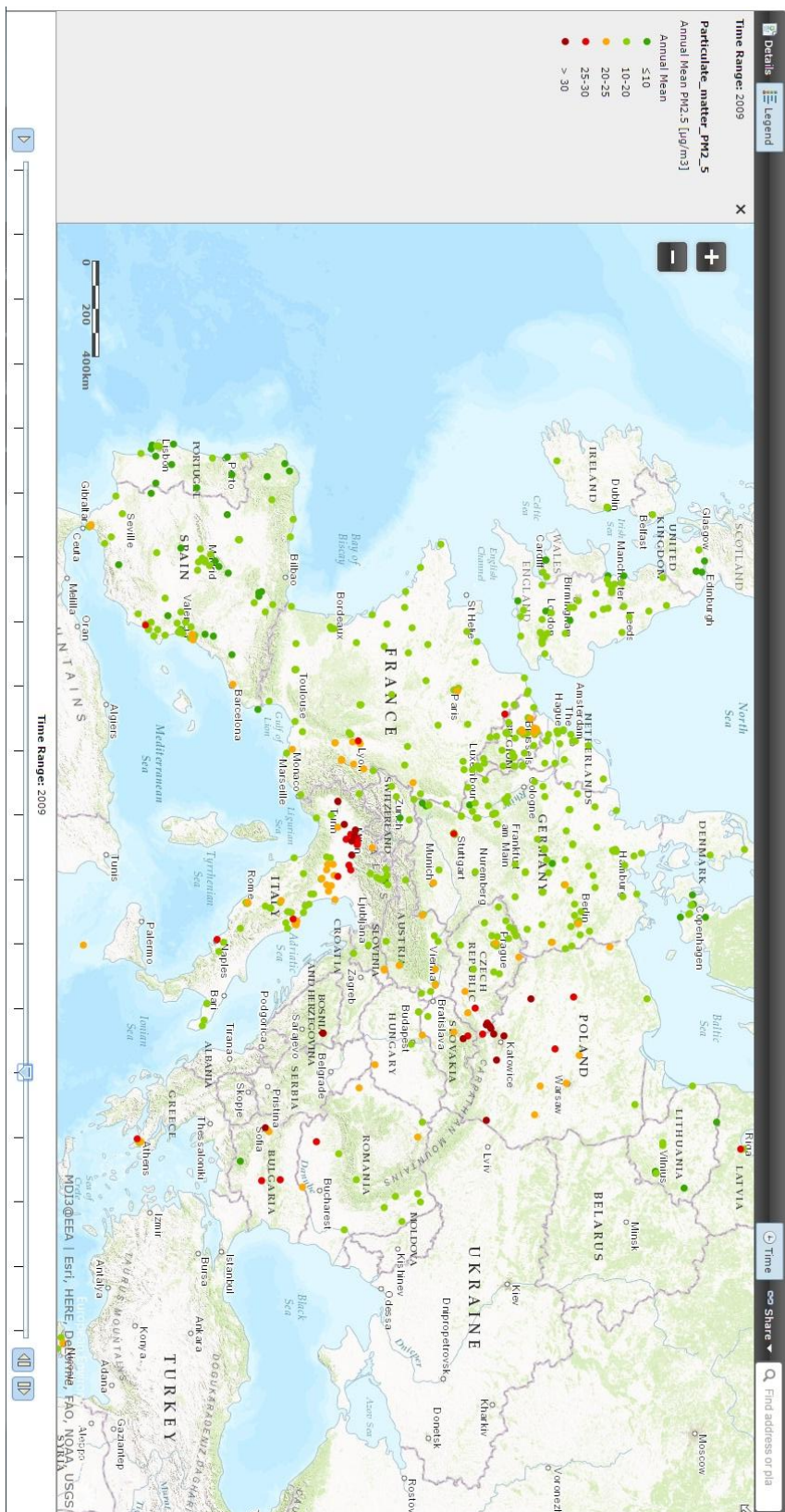


Εικόνα 1.5.3

## Particulate Matter (PM2.5): annual mean concentrations in Europe

GIS Map Application — Published 30 Apr 2013 — Last modified 03 Nov 2015 08:43 AM

The map shows annual mean concentrations of Particulate Matter (PM2.5) in Europe based on daily averages with at least 75% of valid measurements, in  $\mu\text{g}/\text{m}^3$  (source: EEA, AirBase v.8 & AQ e-Reporting)

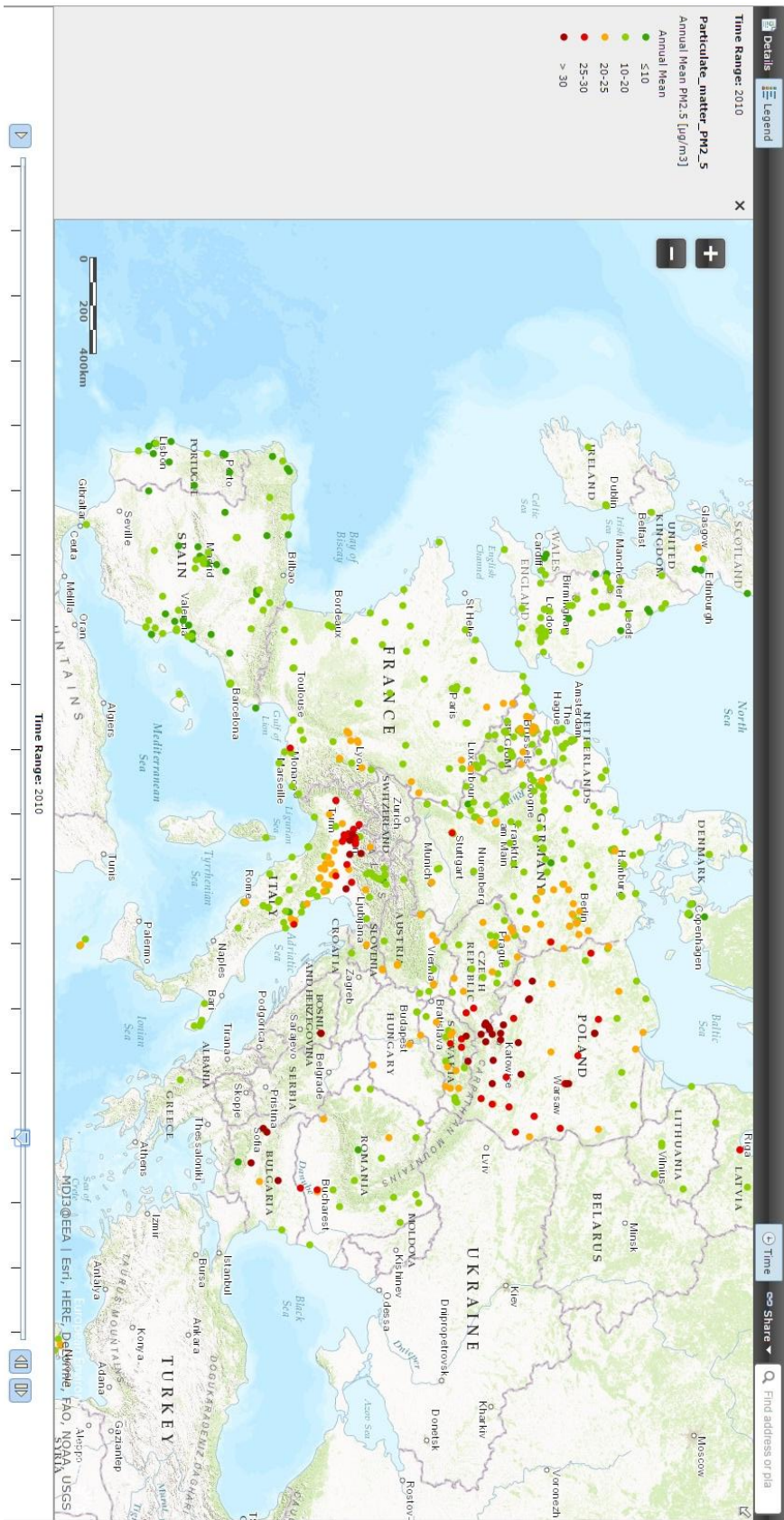


Εικόνα 1.5.4

### Particulate Matter (PM2.5): annual mean concentrations in Europe

GIS Map Application — Published 30 Apr 2013 — Last modified 03 Nov 2015, 08:43 AM

The map shows annual mean concentrations of Particulate Matter (PM2.5) in Europe based on daily averages with at least 75% of valid measurements, in  $\mu\text{g}/\text{m}^3$  (source: EEA, AirBase v.8 & AQ e-Reporting)



Εικόνα 1.5.5

## Κεφάλαιο Δεύτερο: Υλικά και μέθοδοι

### 2.1 – Μέθοδος που ακολουθήθηκε κατά τις δειγματοληψίες

Καθώς το μέρος αυτό διενεργήθηκε από την ερευνητική ομάδα του Institute of Experimental Medicine AS CR[28] προ της δημοσίευσης του data set, θα πρέπει να περιοριστούμε σε μια πολύ συνοπτική περιγραφή του πληθυσμού και της διαδικασίας με την οποία έγιναν οι μετρήσεις.

Κατ' αρχάς, πρόκειται για μια σειρά μελετών οι οποίες εξετάζουν τις επιδράσεις της ατμοσφαιρικής ρύπανσης με χαρακτήρα βιομηχανικών ρύπων, συγκρίνοντας τις ανταποκρίσεις των ανθρωπίνων οργανισμών μεταξύ των δύο πόλεων με μία σειρά προσεγγίσεων [29] [30] [28]. Στα πλαίσια αυτής της σειράς ερευνών, χρησιμοποιήθηκαν τα δείγματα και για γενομική ανάλυση έκφρασης γονιδίων, ώστε να εξεταστεί αν σε αυτό το επίπεδο μπορεί να αναγνωριστεί διαφορά μεταξύ των προφίλ έκφρασης γονιδίων. Το data set αυτό δημοσιεύτηκε στο ανοικτό εναποθετήριο GEO στις 27 Αυγούστου 2014 με κωδικό αριθμό GSE60767.

#### Σύνθεση του πληθυσμού

Ο πληθυσμός στον οποίο διενεργήθηκαν δειγματοληψίες αποτελείται από άρρενες αστυνομικούς της Πράγας και από αστυνομικούς των πόλεων Χαβίροβ και Καρβίνα, πόλεων στην ευρύτερη περιοχή της Οστράβα (νοτιοανατολικά και βορειοανατολικά αντιστοίχως), οι οποίες παρουσιάζουν πολύ παρόμοιο προφίλ ρύπων.

Κατά την επιλογή των δειγμάτων που θα υφίσταντο τη γενομική ανάλυση, απορρίφθηκαν δείγματα από δότες που ήταν καπνιστές, είχαν υποστεί ιατρική αγωγή, εμβολιασμό ή ακτινογραφία το τελευταίο τρίμηνο πριν τις δειγματοληψίες, ώστε να περιοριστούν οι μεταβλητές που θα μπορούσαν να θολώσουν τα δεδομένα.

Οι συμμετέχοντες στην έρευνα συναίνεσαν μετά πλήρους ενημέρωσης και μπορούσαν να ακυρώσουν ανά πάσα στιγμή τη συμμετοχή τους, ως ορίζει η συνθήκη Ελσίνκι II. Η μελέτη εγκρίθηκε από την Επιτροπή Ηθικής του Ινστιτούτου Πειραματικής Ιατρικής ASCR της Πράγας (Institute of Experimental Medicine AS CR).

Διενεργήθηκαν τρεις δειγματοληψίες και στις δύο περιοχές, όπως παρουσιάζεται στον πίνακα 2.1.1:

	Χειμώνας 2009	Καλοκαίρι 2009	Χειμώνας 2010
Πράγα	47	58	49
Οστράβα	84	95	133

**Πίνακας 2.1.1**

Το 85% των εθελοντών συμμετείχαν σε δύο ή τρεις δειγματοληψίες, ενώ το 15% συμμετείχε μόνο στο καλοκαίρι του 2009 ή το Χειμώνα του 2010. Επίσης, το Χειμώνα του 2010 συμμετείχαν στις

δειγματοληψίες και εθελοντές υπάλληλοι γραφείου από την περιοχή Οστράβα-Μπάρτοβιτς, ανατολικά των μεγαλύτερων βιομηχανιών ατσαλιού της Οστράβα.

### Συλλογή δειγμάτων

Τα δείγματα αίματος συλλέχθηκαν με φλεβοκέντηση, σε φιαλίδια που περιείχαν ethylene diamine tetra acetic acid (EDTA). Τα λευκοκύτταρα απομονώθηκαν και σταθεροποιήθηκαν εντός δωδεκαώρου με τη χρήση του LeukoLOCK™ Total RNA Stabilization and Isolation System (LeukoLOCK; Ambion, Austin, TX, USA). Τα σταθεροποιημένα λευκοκύτταρα αποθηκεύτηκαν στους -80° C μέχρι την ανάλυση.

### Απομονωση RNA

Η εξαγωγή του RNA από τα κατεψυγμένα λευκοκύτταρα έγινε με τη χρήση του LeukoLOCK. Για να μειωθεί η πιθανότητα batch effect, το σύνολο των δειγμάτων υπέστη επεξεργασία την ίδια μέρα, αφού είχε τελειώσει η συλλογή τους. Biotinylated cRNA προετοιμάστηκε από τα 468 δείγματα μεαντίστροφη μετάφραση RNA (200 ng) σε cDNA και το οποίο υπεβλήθη σε in vitro μετάφραση στο Illumina Total Prep RNA Amplification Kit (Ambion, Austin, TX, USA). Το παραχθέν biotinylated cRNA (750 ng) υβριδίστηκε σε Illumina Human-HT12 v3 Expression Bead-Chips (Illumina, San Diego, CA, USA). Η διαδικασία υβριδισμού και η μετέπειτα πλύση, χρώση και στέγνωμα των beadchips έγινε ακολουθώντας το στάνταρ πρωτόκολλο της Illumina. Οι υβριδισμένες πλακέτες διαβάστηκαν στο Illumina BeadStation 500GX. Τα δεδομένα από τις κεφαλές συνοψίστηκαν από το Illumina BeadStudioSoftware v3.3.7. Τέλος, δύο εκ των δειγμάτων εξαιρέθηκαν από τη μετέπειτα ανάλυση ως τεχνικά αναξιόπιστα, δίνοντας τελικό μέγεθος πληθυσμού δειγμάτων ίσο με 466.

## **2.2 – Υλικοτεχνική υποδομή της ανάλυσης**

Η ανάλυση του data set διενεργήθη σε desktop PC με επεξεργαστή Intel Core i5, 4 Gb RAM και λειτουργικό Windows 7 Home Premium SP1.

Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε ήταν η R v3.3.1

Η ανάλυση διενεργήθη στο γραφικό περιβάλλον του RStudio v0.99.902.

Το dataset βρέθηκε από την ανοικτή βάση δεδομένων GEO.

Κατά τη διάρκεια της ανάλυσης, χρησιμοποιήθηκαν τα εξής πακέτα ανοικτού κώδικα της R, από τη σουίτα Bioconductor:

- GEOquery v2.38.4
- limma v3.28.19
- illuminaHumanv3.db v1.26.0
- sva v3.20.0



Στο επίπεδο της λειτουργικής ανάλυσης, χρησιμοποιήθηκε η πλατφόρμα BioInfoMiner για την ανάλυση των διαφορικά εκφραζόμενων γονιδίων με τις οντολογίες Gene Ontology, Human Phenotype Ontology, MGI Mammalian Phenotype Ontology, Reactome Pathways Ontology.

## Κεφάλαιο Τρίτο: Μέθοδος της ανάλυσης

### Περιορισμοί:

- Δεν παρατίθετο, ούτε και έγινε δυνατό να βρεθεί, πληροφορία για το ποιά δείγματα αντιστοιχούν σε κάθε μοναδικό δότη, καθώς το 85% συμμετείχε σε περισσότερες της μίας δειγματοληπτικές περιόδους. Δεδομένου αυτού, η ανάλυση προχώρησε θεωρώντας κάθε δείγμα μοναδικό, παρ' όλο που αρκετά από αυτά προέρχονται από τους ίδιους δότες θα έπρεπε να εισάγουν επιπλέον όρους μεταξύ τους σύνδεσης.
- Επίσης δεν παρατίθετο καμμία πληροφορία για τη δομή του δειγματοληπτικού πληθυσμού όσον αφορά το ποιά δείγματα προέρχονταν από την Χαρβίροβ και ποιά από την Καρβίνα, ποιά από τους επιπλέον εθελοντές όσον αφορά το 2010, τις επιμέρους ηλικίες των δοτών, τα ΒΜΙτους ή οποιοδήποτε άλλο στοιχείο. Τα μόνα δεδομένα που υπάρχουν είναι για το σε ποιά από τις έξι δειγματοληψίες ανήκει το κάθε δείγμα. Έτσι, η ανάλυση θα πρέπει να θεωρήσει τα δείγματα της περιοχής της Οστράβα ως έναν ομοιγενή πληθυσμό χωρίς επιπλέον χαρακτηριστικά.
- Κάθε πλακέτα IlluminaHuman-HT12 v3 Expression Bead-Chip έχει χώρο για 12 δείγματα. Κατά τον υβριδισμό τους, δεν τοποθετήθηκαν με κριτήριο το να αντιπροσωπεύονται σε κάθε πλακέτα όλοι οι δειγματοληπτικοί πληθυσμοί. Ως εκ τούτου, εάν αναγνωριστεί η επίδοση της πλακέτας ως πιθανό batch effect, θα είναι πλέον άρρηκτα συνδεδεμένο με μέρος του βιολογικού σήματος που ψάχνουμε. Ως εκ τούτου, η απαραίτητη σε κάθε περίπτωση διόρθωσή του θα εξομαλύνει και μέρος του βιολογικού σήματος. Βρισκόμαστε σε μία κατάσταση στην οποία η προσπάθεια είναι κυρίως να μειώσουμε το batch effect, διατηρώντας όσο το δυνατόν περισσότερη βιολογική πληροφορία.

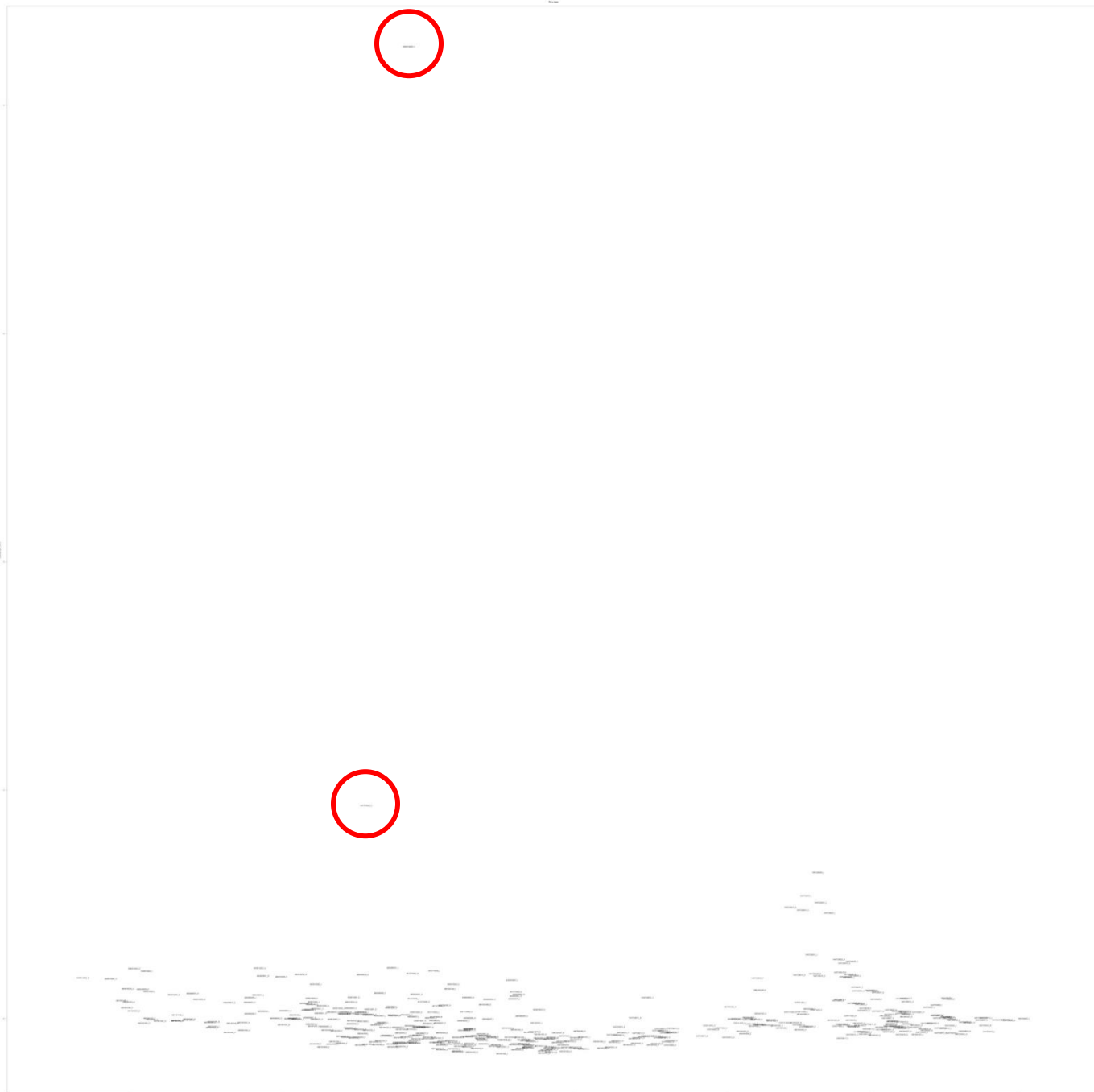
Ο κώδικας θα παρατεθεί στο παράρτημα, καθώς το μεγάλο του μέγεθος κάνει μάλλον μη-βοηθητική την παρουσίασή του εντός του κειμένου. Όπου χρειάζεται, θα δοθούν παραπομπές στις σχετικές σελίδες.

Ως πρώτο βήμα, μετά την εγκατάσταση των σχετικών πακέτων χρησιμοποιήθηκε το GEOQuery για την ανάκτηση των δεδομένων και την προετοιμασία τους. Στο ίδιο σημείο της διαδικασίας, η χρήση της getGEO, του πακέτου της R GEOquery μας επέτρεψε να ανακτήσουμε την πληροφορία για την προέλευση του κάθε δείγματος όσον αφορά την περίοδο και την πόλη δειγματοληψίας, η οποία δεν είχε καταχωρηθεί με εύχρηστο τρόπο στα αρχικά δεδομένα κατά την ανάρτηση του data set στο GEO.

### Pre-processing

Μετά το διάβασμα των δεδομένων (468 δείγματα), έγινε η κανονικοποίησή τους και ελέγχθη η κατανομή τους με multi-dimensional scattering plot (στο εξής, MDSplot). Αναγνωρίστηκε η ύπαρξη δύο δειγμάτων που και όσον αφορά τις αποδόσεις τους και όσον αφορά τη σχέση τους με τα άλλα

δείγματα, η πιθανότητα να πρόκειται για πειραματικά λάθη κατά τον υβριδισμό ήταν ιδιαίτερα υψηλή. Το εύρημα αυτό ελέγχθη και φαίνεται να συμφωνεί με τη δημοσίευση του data set, ενώ ταυτοποιήθηκε πως όντως πρόκειται για τα ίδια δύο δείγματα. Απορρίφθηκαν από την περαιτέρω ανάλυση, με το τελικό μέγεθος του data set να διαμορφώνεται στα 466 δείγματα.

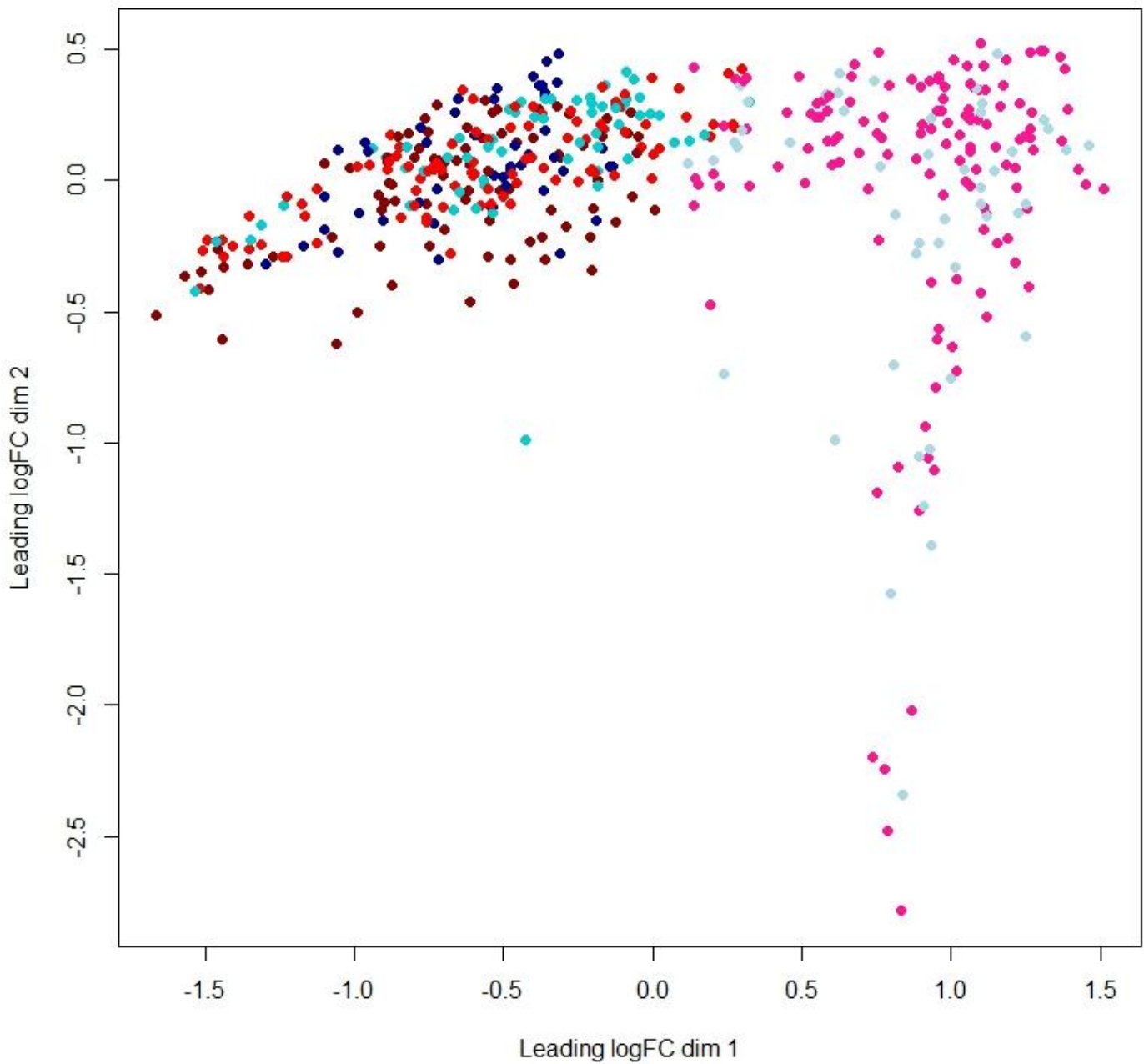


**Εικόνα 3.1**

Εμφανής ύπαρξη outliers, κατά πάσα πιθανότητα τεχνικού χαρακτήρα

- Πράγα, Χειμώνας 2009
- Πράγα, Καλοκαίρι 2009
- Πράγα, Χειμώνας 2010
- Όστραβα, Χειμώνας 2009
- Όστραβα, Καλοκαίρι 2009
- Όστραβα, Χειμώνας 2010

Initial data, color-coded



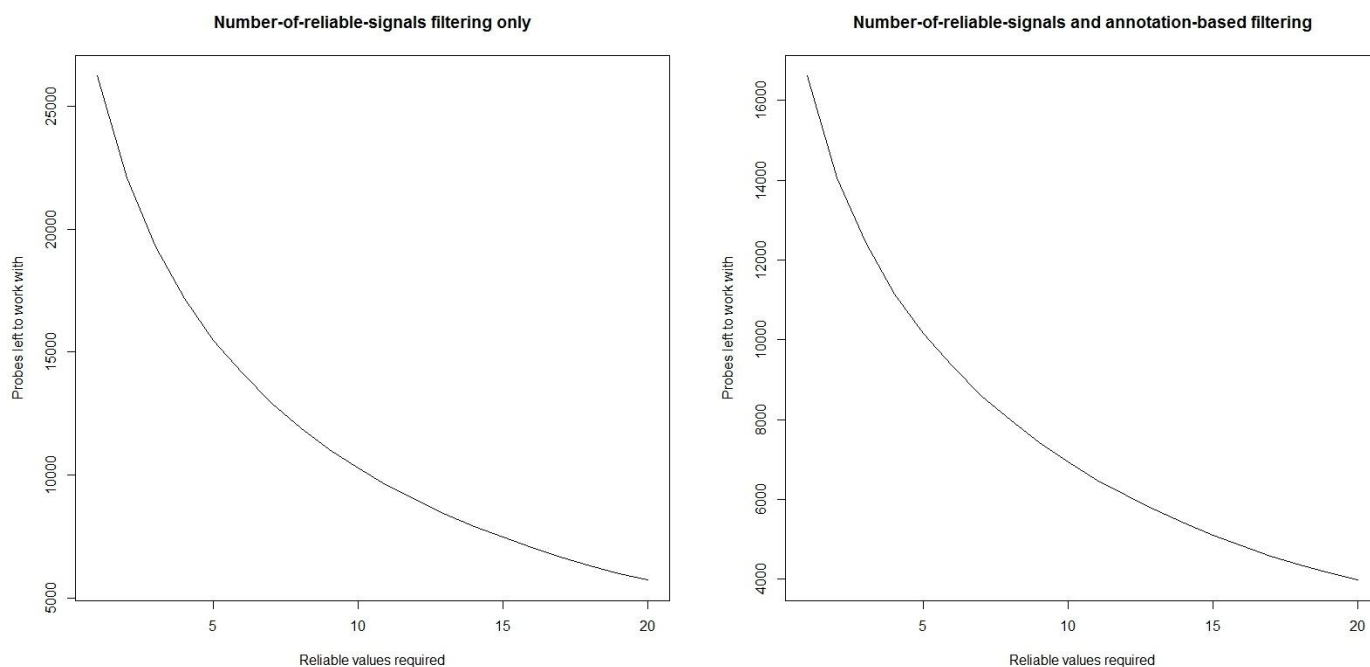
Εικόνα 3.2

Μετά την εκ νέου κανονικοποίηση και την επαλήθευση ότι δεν υπάρχουν άλλες σαφείς υπόνοιες για τεχνικά λάθη, ξεκίνησε η διαδικασία διαλογής των probes που θα χρησιμοποιηθούν στην ανάλυση:

Προτιμήθηκε ένα διπλό φιλτράρισμα:

1. Τεχνικό – κρατήθηκαν τα probes τα οποία έχουν εκφραστεί με αξιοπιστία μεγαλύτερη ή ίση του 95% σε τουλάχιστον ένα συγκεκριμένο αριθμό δειγμάτων, ώστε probes που δε μπορούν να χρησιμοποιηθούν ως βάση σύγκρισης να μην επηρεάσουν την έρευνα.
2. Βιολογικό – αξιοποιώντας πληροφορία για το annotation της συγκεκριμένης έκδοσης των συγκεκριμένων μικροσυστοιχιών (HumanHT-12 v3.0 Gene Expression BeadChip), probes τα οποία έχουν χαρακτηριστεί ως χαμηλής ποιότητας και υβριδίζονται συχνά χωρίς να αντιστοιχούν σε πραγματικό βιολογικό σήμα (επί παραδείγματι επειδή περιέχουν επαναλαμβανόμενα κωδικόνια), αφαιρέθηκαν ώστε να αυξηθεί η αξιοπιστία της βιολογικής πληροφορίας.

Παρατηρήθηκε, λόγω του τεχνικής φύσεως φιλτραρίσματος, μία πολύ έντονη εξάρτηση του αριθμού των εναπομεινάντων probes από κατώφλι το οποίο τίθεται, η οποία ξεκινά να εξωμαλύνεται περίπου στα 10.000 probes, όπως φαίνεται στα παρακάτω γράφημα (Εικόνα 3.3):



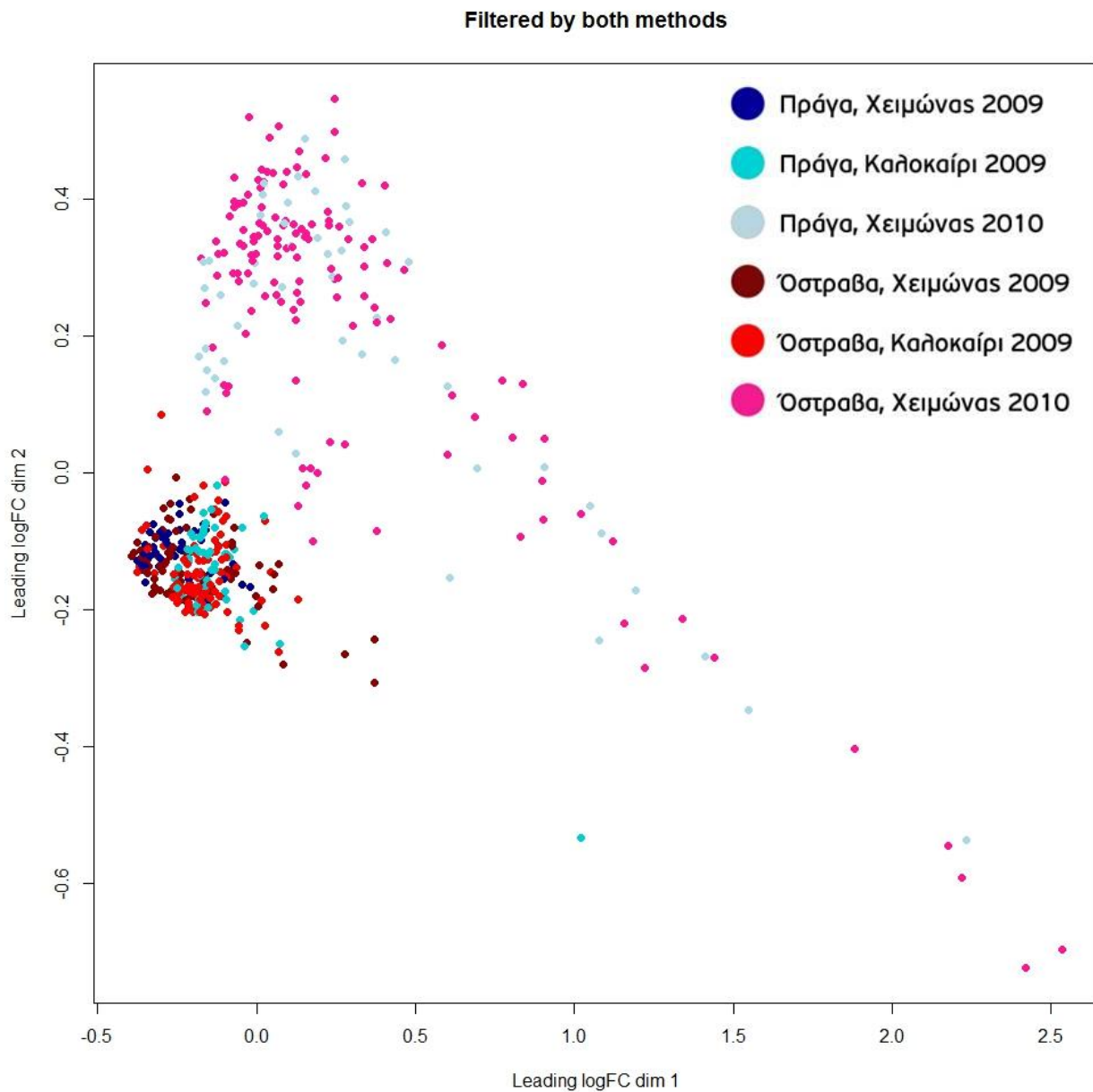
**Εικόνα 3.3**

Ταυτόχρονα, παρατηρήθηκε πως κάτω από μία «κρίσιμη μάζα» 9.500-10.000 probes, το μέγεθος των εξαγόμενων λιστών άρχιζε να έχει έντονη εξάρτηση από τον αριθμό των probes, υποδεικνύοντας πως από αυτό το σημείο και κάτω αρχίζει πλέον να χάνεται αρκετή πληροφορία ώστε να μειώνεται η αξιοπιστία της ανάλυσης που ακολουθεί. Πάνω από αυτά τα όρια, υπάρχει σχετική ανεξαρτησία από

τον αριθμό των probes, οπότε και η τιμή για το τεχνικό φιλτράρισμα επελέγη με κριτήριο το να μην πέσουμε κάτω από αυτό τον αριθμό.

Το MDSplot του φιλτραρισμένου dataset δείχνει μία καλή συμπεριφορά, καθώς υπάρχει μεν μία παραμόρφωση των σχέσεων μεταξύ των δειγμάτων, αλλά:

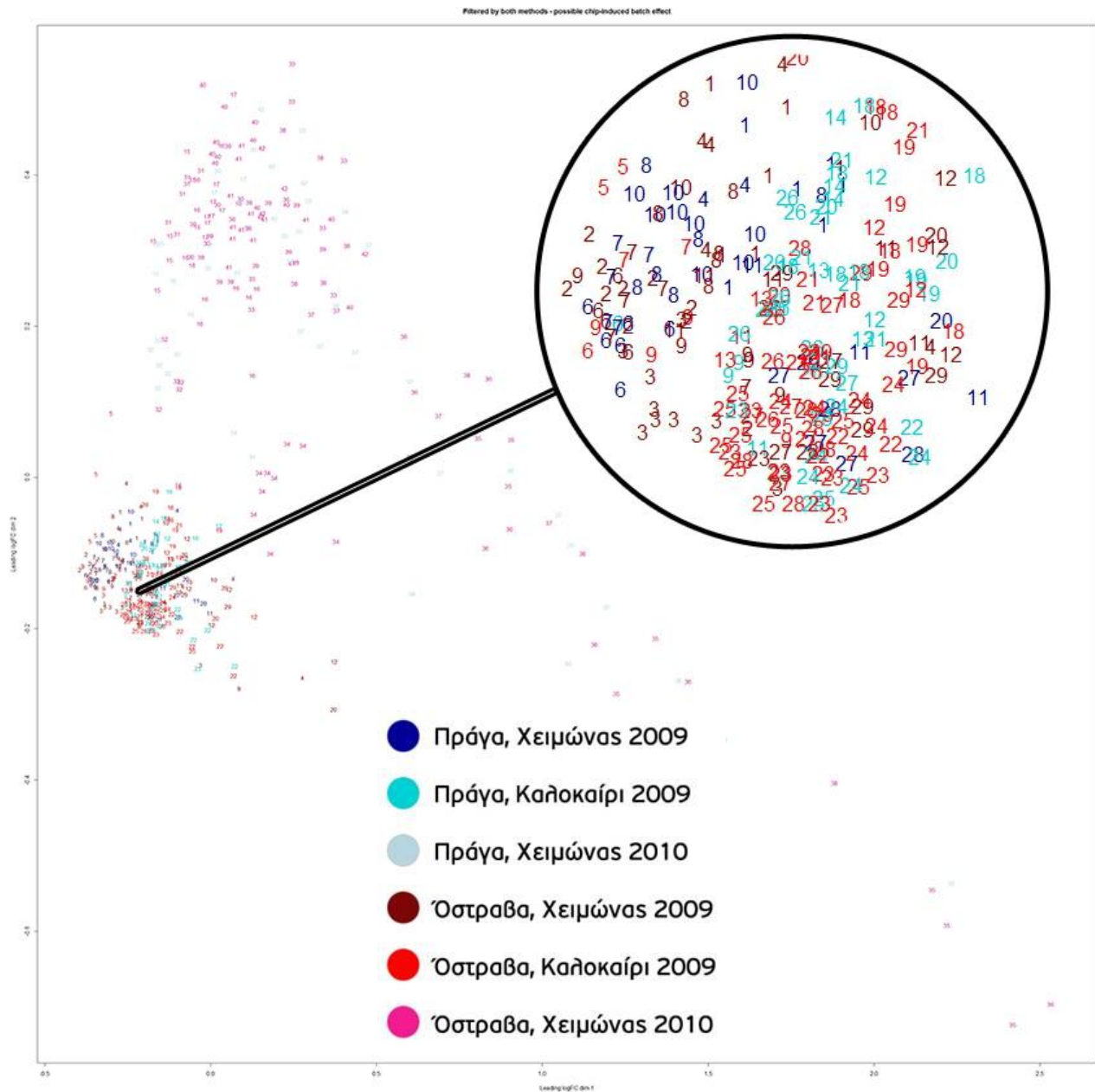
- Το γενικό του σχήμα δεν αλλάζει παρά λίγο, δείχνοντας ότι δεν έχουμε εισάγει σημαντικό ποσό πληροφορίας κατά το φιλτράρισμα, οπότε και διατηρούνται σε σοβαρό ποσοστό οι σχέσεις μεταξύ των δειγμάτων.
- Μειώνεται το μέγεθος της αναπαράστασης, όπως φαίνεται από τους άξονες, πράγμα που υποδεικνύει πως μέρος της ετερογένειας οφειλόταν όντως σε probes χαμηλής αξιοπιστίας. Υποδεικνύεται έτσι πως μάλλον αυξάνεται η αξιοπιστία της έρευνας, καθώς αντιμετωπίζεται μέρος τους θορύβου.



**Εικόνα 3.4**

Παρατηρούμε επίσης πως το MDSplot, τόσο πριν όσο και μετά το φιλτράρισμα, δείχνει έναν ισχυρό διαχωρισμό με βάση τη χρονιά, αναδεικνύοντας τη μάλλον στον πρώτο και ισχυρότερο παράγοντα διαφορικής έκφρασης.

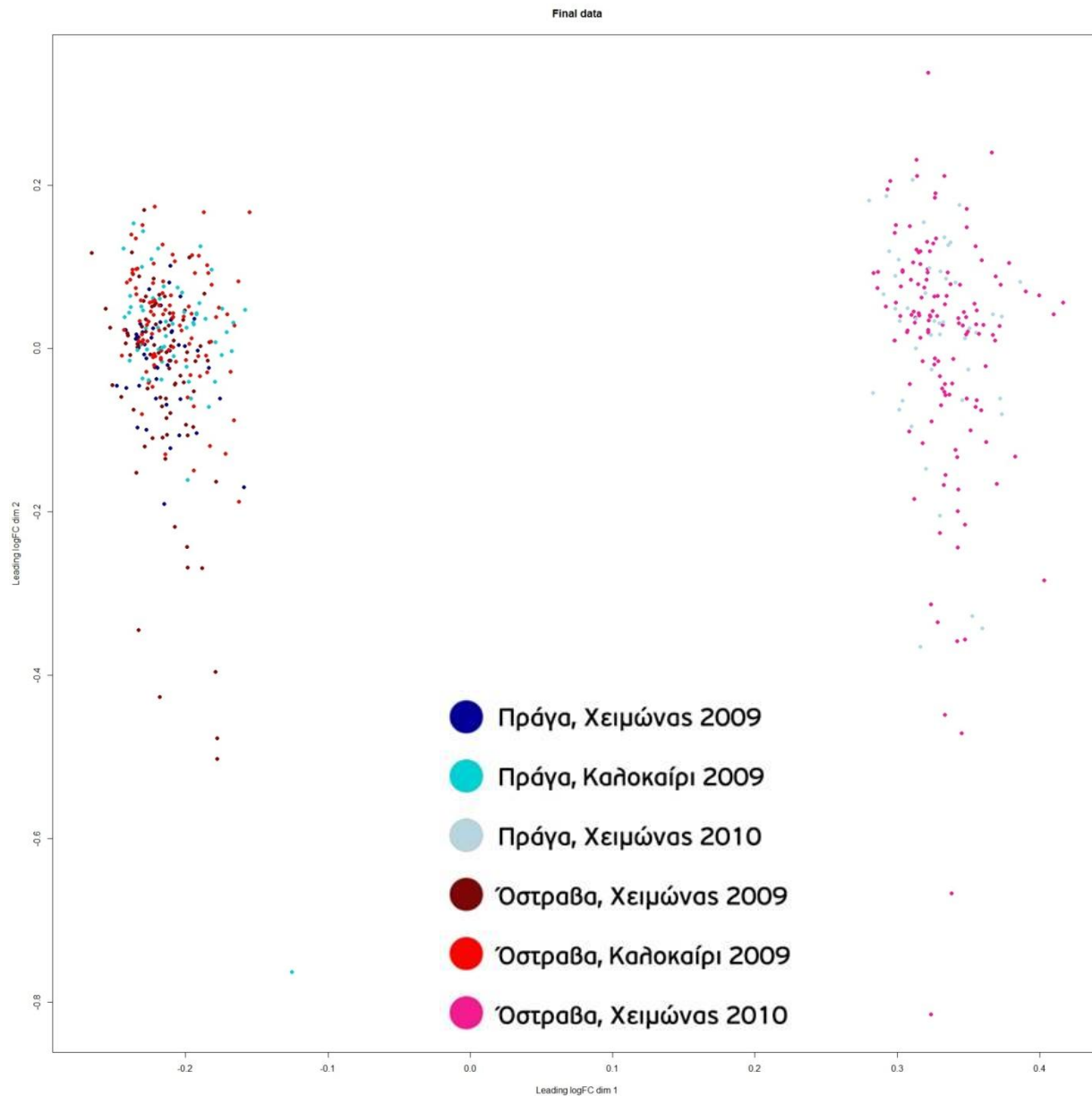
Πριν προχωρήσουμε στην περαιτέρω ανάλυση, δοκιμάσαμε να δούμε πώς κατανέμονται τα δεδομένα με βάση το beadchips στο οποίο έχει υβριδιστεί κάθε δείγμα. Για το σκοπό αυτό, επαναλήφθηκε το MDSplot των φιλτραρισμένων δεδομένων, στο οποίο κάθε δείγμα δεν αναπαρίσταται πλέον από ένα σημείο, αλλά από έναν αριθμό, ο οποίος υποδεικνύει σε ποιο από τα 42 beadchips που χρησιμοποιήθηκαν στην έρευνα υβριδίστηκε (βλ. Εικόνα 3.5):



**Εικόνα 3.5**

Παρατηρούμε μία πολύ ισχυρή συγκέντρωση των δειγμάτων με βάση το chip στο οποίο υβριδίστηκαν, τόσο όσον αφορά την τάση των δειγμάτων που υβριδίστηκαν στο ίδιο beadchip να σχηματίζουν clusters, όσο και από το σημαντικότερο γεγονός ότι αυτά τα clusters έχουν συχνά σαφώς μικρότερη ακτίνα από το clustering με βάση τη χρονιά, την εποχή ή την πόλη. Ως εκ τούτου, πρέπει απαραίτητως να θεωρηθεί η ύπαρξη batch effect με βάση το chip στο οποίο το κάθε δείγμα υβριδίστηκε και να υπάρξει η σχετική διόρθωση. Τα αποτελέσματα αυτής παρουσιάζονται στην εικόνα 3.6:





**Εικόνα 3.6**

Ακόμα περισσότερο από πριν, μετά τη διόρθωση, ενισχύεται η διαφορά μεταξύ των χρονιών, σε σημείο που φαίνεται πως θα υπερκαλύπτει κάθε άλλη συνεισφορά στην ετερογένεια των δεδομένων. Εν ολίγοις, πρόκειται για μια εικόνα που περισσότερο παραπέμπει σε δύο διαφορετικά datasets. Ως εκ τούτου, αποφασίστηκε πως η ορθότερη προσέγγιση θα ήταν τα δεδομένα του 2009 και του 2010 να αναλυθούν από την αρχή ως δύο ξεχωριστά datasets και οποιεσδήποτε μεταξύ τους συγκρίσεις να γίνουν στο επίπεδο της λειτουργικής ανάλυσης, καθώς αλλιώς η μεγάλη μεταξύ τους διαφορά θα υπερκάλυπτε κάθε διαφορεική έκφραση μεταξύ των δύο πόλεων, την οποία άλλωστε θέλουμε να μελετήσουμε.

Έτσι, το αρχικό dataset χωρίζεται σε δύο νέα, το ένα με τα δεδομένα του 2009 και το άλλο με του 2010. Τα τρία πλέον datasets υφίστανται το καθένα χωριστή κανονικοποίηση και φιλτράρισμα των probes, ενώ οι αναλύσεις τους προχωρούν παράλληλα.

### Επιλογή προσέγγισης

Σε αυτό το μέρος της ανάλυσης γίνεται κεντρικής σημασίας το ερώτημα ποιές προσεγγίσεις αντιμετωπίζουν το batch effect διατηρώντας μεγαλύτερο μέρος του βιολογικού σήματος. Προσοχή πρέπει να δοθεί στο γεγονός ότι σε αυτό το σημείο, η σύγκριση μπορεί να γίνει αποκλειστικά με ποσοτικό τρόπο, ενώ το αν τα αποτελέσματα στέκουν από πλευράς βιολογικού περιεχομένου θα πρέπει να κριθεί στο επίπεδο της λειτουργικής ανάλυσης. Έχοντας θέσει αυτόν τον προβληματισμό, οι σχετικές επιδόσεις των προσεγγίσεων παρουσιάζονται στον Πίνακα 3.1:

	No array weights batch correction	No array weights non-parametric batch correction	Array Weights batch in lm	Array weights non-parametric batch correction	Array weights & batch-correction
Full – by year	284	260	128	264	290
2009 – by season	58	3	0	4	61
Winter 2009 – by city	0	0	5	0	0
Summer 2009 – by city	3	3	5	3	3
Prague 2009 – by season	20	11	5	26	42
Ostrava 2009 – by season	63	42	5	35	60
2010 – by city	1	1	5	1	1

**Πίνακας 3.1**

Βάση τόσο της ορθότητας της προσέγγισης, όσο και επιδόσεων, επιλέξαμε την προσέγγιση «Array weights & batch correction».

### Γραμμικά μοντέλα και επιλογή συγκρίσεων

Πρώτα απ' όλα, ο διαχωρισμός σε δύο dataset, ένα για το 2009 κι ένα για το 2010 εξασφαλίζει πως η επίδραση της χρονιάς δε θα δημιουργήσει τεχνητά υψηλές λίστες στατιστικά σημαντικώς διαφορεικά εκφρασμένων γονιδίων. Παρ' όλα αυτά, για λόγους συγκρίσεως, θα μελετηθεί και αυτή στα αδρά της χαρακτηριστικά.

Από εκεί και πέρα, η λογική με την οποία εργαστήκαμε ήταν η εξής:

- 1) Παίρνοντας υποσύνολα του 2009 dataset ανά εποχή (το 2010 είχε δειγματοληψία μόνο το χειμώνα), μπορούμε απ' ευθείας να συγκρίνουμε τα προφίλ έκφρασης των δύο πόλεων, ώστε να απομονώσουμε προφίλ διαφορετικής έκφρασης με κριτήριο την πόλη. Κατ' επέκταση δηλαδή, με κριτήριο τις συνθήκες ζωής, μεγαλύτερη διαφοροποίηση των οποίων παρατηρείται στο

επίπεδο της βιομηχανικού τύπου ρύπανσης. Στο πλαίσιο αυτό, αναμένουμε εντονότερες διαφορές το καλοκαίρι του 2009 απ' όσο το χειμώνα, καθώς τότε σημειώνονται και οι μεγαλύτερες διαφορές των ατμοσφαιρικών ρύπων, ειδικά όσον αφορά τους μικροσωματιδιακούς ρύπους PM<sub>10</sub> και PM<sub>2.5</sub>.

- 2) Παίρνοντας υποσύνολα του 2009 dataset ανά πόλη και συγκρίνοντας με βάση την εποχή στο καθένα, μπορούμε να βρούμε το διαφορικό προφίλ έκφρασης στην αλλαγή της εποχής στην κάθε πόλη. Συγκρίνοντας αυτά τα δύο διαφορικά προφίλ αναμεταξύ τους και με το διαφορικό προφίλ ανά εποχή του συνόλου του 2009, μπορούμε να δούμε αν σημειώνονται διαφορές στον τρόπο με τον οποίο ο οργανισμός ανταποκρίνεται στη μεταβολή των εποχών παρουσία και απουσία βαριάς βιομηχανικής ρύπανσης. Από αυτή τη διαδικασία μπορούν να εξαχθούν συμπεράσματα για την κυκλική ενεργοποίηση μηχανισμών με τους οποίους ο οργανισμός προσπαθεί να προσαρμοστεί και να ανταπεξέλθει στις συνθήκες μόλυνσεως.
- 3) Παρά τα όσα ειπώθηκαν πρωτύτερα, η μελέτη της διαφοράς 2009-2010, μπορεί να μας δώσει χρήσιμα συμπεράσματα για την ποιότητα του καθαρά πειραματικού μέρους της διαδικασίας που προηγήθηκε. Ακόμα περισσότερο, δε μπορεί να μείνει εκτός μελέτης μία μεταβλητή που χωρίζει τόσο έντονα το dataset σε δύο υποομάδες, σε σημείο που να γίνεται δυνητικά προβληματική η μελέτη συμπεριλαμβανοντάς τη.

Εδώ πρέπει να γίνει μια ειδική αναφορά στο data set του 2010. Όπως έχει ειπωθεί και στο δεύτερο κεφάλαιο, κάθε batch-corrected υποσύνολο δεδομένων ελέγχθη με τη χρήση της num.sv του πακέτου SVA για την πιθανότητα ύπαρξης «κρυφών μεταβλητών», η συνεισφορά των οποίων θα έπρεπε να υπολογιστεί κατά την ανάλυσή μας. Στα δεδομένα του 2010 και μόνο, αναγνωρίστηκε η ύπαρξη δύο επιπλέον principal components τα οποία θα έπρεπε να υπολογιστούν. Τα αποτελέσματα της διερεύνησης αυτής για το 2010 παρουσιάζονται στον ακόλουθο πίνακα:

**Πίνακας 3.2**

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	7.9888	2.49738	1.98372	1.74073	1.53100	1.23961	1.15829
Proportion of Variance	0.3526	0.03446	0.02174	0.01674	0.01295	0.00849	0.00741
Cumulative Proportion	0.3526	0.38706	0.40880	0.42554	0.43849	0.44698	0.45440
	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Standard deviation	1.14295	1.12510	1.00383	1.00186	0.98743	0.97763	0.96884
Proportion of Variance	0.00722	0.00699	0.00557	0.00555	0.00539	0.00528	0.00519
Cumulative Proportion	0.46161	0.46861	0.47417	0.47972	0.48511	0.49039	0.49557
	PC15	PC16	PC17	PC18	PC19	PC20	PC21
Standard deviation	0.96129	0.95685	0.93829	0.93439	0.93016	0.92621	0.9228
Proportion of Variance	0.00511	0.00506	0.00486	0.00482	0.00478	0.00474	0.0047
Cumulative Proportion	0.50068	0.50574	0.51060	0.51542	0.52020	0.52494	0.5296
	PC22	PC23	PC24	PC25	PC26	PC27	PC28
Standard deviation	0.91901	0.91798	0.90834	0.90581	0.90380	0.90188	0.89843
Proportion of Variance	0.00467	0.00466	0.00456	0.00453	0.00451	0.00449	0.00446
Cumulative Proportion	0.53431	0.53897	0.54353	0.54806	0.55258	0.55707	0.56153
	PC29	PC30	PC31	PC32	PC33	PC34	PC35

Standard deviation	0.89533	0.8919	0.88834	0.88563	0.87981	0.87694	0.8722
Proportion of Variance	0.00443	0.0044	0.00436	0.00433	0.00428	0.00425	0.0042
Cumulative Proportion	0.56596	0.5704	0.57471	0.57905	0.58332	0.58757	0.5918
	PC36	PC37	PC38	PC39	PC40	PC41	PC42
Standard deviation	0.87066	0.86916	0.86578	0.8619	0.85873	0.85724	0.85336
Proportion of Variance	0.00419	0.00417	0.00414	0.0041	0.00407	0.00406	0.00402
Cumulative Proportion	0.59596	0.60014	0.60428	0.6084	0.61246	0.61652	0.62054
	PC43	PC44	PC45	PC46	PC47	PC48	PC49
Standard deviation	0.8510	0.8507	0.84909	0.84692	0.84550	0.84312	0.84096
Proportion of Variance	0.0040	0.0040	0.00398	0.00396	0.00395	0.00393	0.00391
Cumulative Proportion	0.6245	0.6285	0.63252	0.63648	0.64043	0.64436	0.64827
	PC50	PC51	PC52	PC53	PC54	PC55	PC56
Standard deviation	0.83930	0.83798	0.83517	0.83274	0.83093	0.82721	0.82610
Proportion of Variance	0.00389	0.00388	0.00385	0.00383	0.00381	0.00378	0.00377
Cumulative Proportion	0.65216	0.65604	0.65989	0.66373	0.66754	0.67132	0.67509
	PC57	PC58	PC59	PC60	PC61	PC62	PC63
Standard deviation	0.82338	0.82160	0.81924	0.81737	0.81522	0.81177	0.80990
Proportion of Variance	0.00375	0.00373	0.00371	0.00369	0.00367	0.00364	0.00362
Cumulative Proportion	0.67884	0.68257	0.68627	0.68996	0.69364	0.69728	0.70090
	PC64	PC65	PC66	PC67	PC68	PC69	PC70
Standard deviation	0.8076	0.80581	0.80439	0.80255	0.79984	0.79737	0.7959
Proportion of Variance	0.0036	0.00359	0.00357	0.00356	0.00353	0.00351	0.0035
Cumulative Proportion	0.7045	0.70809	0.71167	0.71523	0.71876	0.72227	0.7258
	PC71	PC72	PC73	PC74	PC75	PC76	PC77
Standard deviation	0.79514	0.79463	0.79224	0.78934	0.78699	0.78338	0.78153
Proportion of Variance	0.00349	0.00349	0.00347	0.00344	0.00342	0.00339	0.00337
Cumulative Proportion	0.72927	0.73275	0.73622	0.73966	0.74309	0.74648	0.74985
	PC78	PC79	PC80	PC81	PC82	PC83	PC84
Standard deviation	0.78096	0.77983	0.77867	0.77693	0.77431	0.7729	0.77088
Proportion of Variance	0.00337	0.00336	0.00335	0.00333	0.00331	0.0033	0.00328
Cumulative Proportion	0.75322	0.75658	0.75993	0.76327	0.76658	0.7699	0.77316
	PC85	PC86	PC87	PC88	PC89	PC90	PC91
Standard deviation	0.76960	0.76738	0.76615	0.76508	0.76416	0.7605	0.75981
Proportion of Variance	0.00327	0.00325	0.00324	0.00323	0.00323	0.0032	0.00319
Cumulative Proportion	0.77643	0.77969	0.78293	0.78616	0.78939	0.7926	0.79578
	PC92	PC93	PC94	PC95	PC96	PC97	PC98
Standard deviation	0.75787	0.75454	0.75380	0.75218	0.7490	0.74827	0.74616
Proportion of Variance	0.00317	0.00315	0.00314	0.00313	0.0031	0.00309	0.00308
Cumulative Proportion	0.79895	0.80209	0.80523	0.80836	0.8115	0.81455	0.81763
	PC99	PC100	PC101	PC102	PC103	PC104	PC105
Standard deviation	0.74339	0.74198	0.73948	0.73753	0.73594	0.73519	0.73356
Proportion of Variance	0.00305	0.00304	0.00302	0.00301	0.00299	0.00299	0.00297
Cumulative Proportion	0.82068	0.82372	0.82674	0.82975	0.83274	0.83573	0.83870
	PC106	PC107	PC108	PC109	PC110	PC111	PC112
Standard deviation	0.73262	0.72981	0.72836	0.72656	0.7247	0.7240	0.72170
Proportion of Variance	0.00297	0.00294	0.00293	0.00292	0.0029	0.0029	0.00288

Cumulative Proportion	0.84167	0.84461	0.84754	0.85046	0.8534	0.8562	0.85913
	PC113	PC114	PC115	PC116	PC117	PC118	PC119
Standard deviation	0.71929	0.71846	0.71682	0.71493	0.71325	0.7119	0.71099
Proportion of Variance	0.00286	0.00285	0.00284	0.00282	0.00281	0.0028	0.00279
Cumulative Proportion	0.86199	0.86484	0.86768	0.87050	0.87332	0.8761	0.87891
	PC120	PC121	PC122	PC123	PC124	PC125	PC126
Standard deviation	0.70955	0.70714	0.70482	0.70290	0.70147	0.6987	0.69733
Proportion of Variance	0.00278	0.00276	0.00274	0.00273	0.00272	0.0027	0.00269
Cumulative Proportion	0.88169	0.88445	0.88720	0.88993	0.89264	0.8953	0.89803
	PC127	PC128	PC129	PC130	PC131	PC132	PC133
Standard deviation	0.69537	0.69410	0.69051	0.68951	0.68836	0.68460	0.68343
Proportion of Variance	0.00267	0.00266	0.00263	0.00263	0.00262	0.00259	0.00258
Cumulative Proportion	0.90070	0.90336	0.90600	0.90862	0.91124	0.91383	0.91641
	PC134	PC135	PC136	PC137	PC138	PC139	PC140
Standard deviation	0.68246	0.68001	0.67803	0.67740	0.67404	0.67040	0.66882
Proportion of Variance	0.00257	0.00255	0.00254	0.00254	0.00251	0.00248	0.00247
Cumulative Proportion	0.91898	0.92154	0.92408	0.92661	0.92912	0.93161	0.93408
	PC141	PC142	PC143	PC144	PC145	PC146	PC147
Standard deviation	0.66846	0.66721	0.66606	0.66383	0.66130	0.65799	0.65640
Proportion of Variance	0.00247	0.00246	0.00245	0.00243	0.00242	0.00239	0.00238
Cumulative Proportion	0.93655	0.93901	0.94146	0.94389	0.94631	0.94870	0.95108
	PC148	PC149	PC150	PC151	PC152	PC153	PC154
Standard deviation	0.65627	0.65266	0.64881	0.64344	0.64259	0.63836	0.63777
Proportion of Variance	0.00238	0.00235	0.00233	0.00229	0.00228	0.00225	0.00225
Cumulative Proportion	0.95346	0.95581	0.95814	0.96043	0.96271	0.96496	0.96721
	PC155	PC156	PC157	PC158	PC159	PC160	PC161
Standard deviation	0.6313	0.62777	0.62595	0.62283	0.61919	0.61470	0.61085
Proportion of Variance	0.0022	0.00218	0.00216	0.00214	0.00212	0.00209	0.00206
Cumulative Proportion	0.9694	0.97159	0.97375	0.97589	0.97801	0.98010	0.98216
	PC162	PC163	PC164	PC165	PC166	PC167	PC168
Standard deviation	0.60966	0.60022	0.58801	0.58271	0.56794	0.4455	0.44159
Proportion of Variance	0.00205	0.00199	0.00191	0.00188	0.00178	0.0011	0.00108
Cumulative Proportion	0.98421	0.98620	0.98811	0.98999	0.99177	0.9929	0.99395
	PC169	PC170	PC171	PC172	PC173	PC174	PC175
Standard deviation	0.36380	0.34554	0.31962	0.31534	0.30705	0.30235	0.28586
Proportion of Variance	0.00073	0.00066	0.00056	0.00055	0.00052	0.00051	0.00045
Cumulative Proportion	0.99468	0.99534	0.99590	0.99645	0.99697	0.99748	0.99793
	PC176	PC177	PC178	PC179	PC180	PC181	
Standard deviation	0.28250	0.27689	0.26224	0.24569	0.21531	0.20714	
Proportion of Variance	0.00044	0.00042	0.00038	0.00033	0.00026	0.00024	
Cumulative Proportion	0.99837	0.99879	0.99917	0.99951	0.99976	1.00000	

Με τη χρήση της μεθόδου svatou ομώνυμου πακέτου, η συνεισφορές των δύο βοηθητικών μεταβλητών υπολογίστηκαν και εισήχθησαν στο γραμμικό μοντέλο.

### Εξαγωγή λιστών στατιστικά σημαντικώς διαφορικώς εκφρασμένων γονιδίων

Στα πλαίσια της εξαγωγής των λιστών με τη μέθοδο topTable, είναι δυνατό να τεθούν κριτήρια τα οποία απαιτούμε να πληρούν τα γονίδια που θα περιληφθούν στην περαιτέρω ανάλυση.

Κατ' αρχάς, ως πρώτο φίλτρο, θέσαμε τη βιολογική σημασία της σχετικής διαφοράς έκφρασης, θεωρώντας βιολογικά σημαντικό κάθε γονίδιο που εκφράζεται διαφορικά με  $|\log \text{fold change}| \geq 0.1$ . Πρόκειται για αρκετά χαμηλή τιμή, αλλά η επιλογή αυτή δικαιολογείται υπό τις εξής συνθήκες στις οποίες εργαζόμαστε:

- Μελετάμε την αντανάκλαση καθαρά περιβαλλοντολογικών αιτίων στην έκφραση γονιδίων, οπότε και πρέπει να περιμένουμε ένα αρκετά πιο αδύναμο σήμα σε σχέση με άλλες προσεγγίσεις ανάλυσης (επί παραδείγματι, μία έρευνα με δεδομένα μεθυλιώσεων θα μπορούσε να δώσει αρκετά πιο καθαρό σήμα).
- Το γεγονός ότι, όπως έχει αναλυθεί νωρίτερα, το βιολογικό σήμα είναι ισχυρά συνδεδεμένο με το batch effect που έχει εισαχθεί από την κακή κατανομή των δειγμάτων στα beadchips, σημαίνει πως ένα μέρος της ισχύος του σήματος θα χαθεί κατά τη διαδικασία του batch effect correction, ανεξαρτήτως της μεθοδολογίας με την οποία αυτό θα γίνει, συμβάλλοντας σε χαμηλά log fold changes.
- Το μεγάλο μέγεθος του δείγματος μπορεί να αυξήσει τη στατιστική ισχύ του πληθυσμού και κατ' επέκταση της ανάλυσής μας, μειώνοντας τις πιθανότητες στατιστικών λαθών.

Δεύτερον, ως δεύτερο φίλτρο, θέσαμε τη στατιστική σημαντικότητα της διαφορικής έκφρασης, θεωρώντας στατιστικά σημαντικό ένα διαφορικώς εκφρασμένο γονίδιο μόνο εάν πληρούσε τη συνθήκη  $\text{false discovery rate} < 0.05$ . Γονίδια πάνω από αυτό το όριο απορρίφθηκαν από την περαιτέρω ανάλυση.

Σε αυτό το σημείο υπήρξαν δύο σημαντικές εξαιρέσεις:

Πρώτη εξαίρεση είναι οι συγκρίσεις που αναφέρονται στο πλήρες data set. Εδώ, κάθε σύγκριση, λόγω της εξαιρετικής ανομοιομορφίας μεταξύ των εποχών, δίνει εξαιρετικά υψηλά log fold changes, με αποτέλεσμα να θεωρείται βιολογικά σημαντικώς διαφορικώς εκφρασμένως ένας εξαιρετικά υψηλός αριθμός γονιδίων, περί τις 4000-5500. Ένας τέτοιος αριθμός, σε συνδυασμό με τα εξαιρετικά υψηλά log fold changes που εμφανίζονται (της τάξης του 1-3 μετά το batch effect correction, το οποίο όπως προαναφέρθηκε έχει εξομαλύνει τις διαφορές), σημαίνει πως μάλλον δε μπορούν να είναι αξιόπιστες, όπως θα αναλυθεί πληρέστερα μετά την παρουσίαση των αποτελεσμάτων. Ως εκ τούτου, πέραν των λιστών με όριο βιολογικής σημαντικότητας  $|lfc| \geq 0.1$ , υπολογίστηκαν και οι λίστες με όριο  $|lfc| \geq 0.5$ , καθώς σε εκείνη την περιοχή, η σχέση μεταξύ lfcορίου και αριθμού γονιδίων της λίστας είναι κοντύτερα σε γραμμική.

Μοναδική εξαίρεση σε αυτή τη διαδικασία αποτέλεσε η σύγκριση του υποσυνόλου της Πράγας 2009, ανά εποχή, με την προσέγγιση Array weights & batch correction. Σε αυτή τη σύγκριση και μόνο, υπήρχε ένα σημαντικό ποσοστό γονιδίων βιολογικώς σημαντικώς εκφρασμένων, με  $\text{FDR} \sim 0.07$ . Κατ' εξαίρεση, για αυτή τη σύγκριση το όριο της στατιστικής σημαντικότητας αυξήθηκε από 0.05 σε 0.1. Αξίζει να

αναφερθεί πως πέραν αυτής, καμμία άλλη σύγκριση δεν παρουσίασε ευαισθησία σε αυτό το εύρος στατιστικής σημαντικότητας, ενισχύοντας την επιλογή του  $FDR < 0.05$  ως ορίου για αυτήν.

### Έλεγχος των gene symbols

Σε αυτό το μέρος της ανάλυσης προέκυψε η ανάγκη να ανανεωθούν τα gene symbols που παρείχε το annotation του IlluminaBeadStudio κατά τη δημιουργία του data set, καθώς με την πάροδο του χρόνου, αρκετά από αυτά έχουν αλλάξει [1]. Αυτή η διαδικασία έγινε περνώντας την πλήρη λίστα με τα gene symbols του Illumina BeadStudio στη βάση δεδομένων Hugo [31], η οποία παρείχε μία λίστα με όσα εξ' αυτών έχουν υποστεί αλλαγές, καταργηθεί, αποκτήσει συνώνυμα gene symbols και ούτω καθ' εξής. Η λίστα αυτή αντιπαρεβλήθη με τις τελικές λίστες των στατιστικά σημαντικώς διαφορικώς εκφρασμένων γονιδίων και τα gene symbols που είχαν αλλάξει, αλλά χωρίς να έχουν αλλάξει χρωμόσωμα (οπότε και μπορούμε να είμαστε αρκετά σίγουροι ότι δεν είχε γίνει κάποιο πολύ μεγάλο λάθος στο αρχικό annotation), ανανεώθηκαν. Επειδή στο annotation της IlluminaHumanv3 ήταν μεγάλη πλειοψηφία τα gene symbols δίχως καταγεγραμμένη θέση στο γονιδίωμα, ο τελευταίος αυτός έλεγχος και η αντικατάσταση ήταν πρακτικό να γίνουν μόνο σε όσα gene symbols είχαν και τότε και τώρα καταγεγραμμένη θέση.

### Σύγκριση με οντολογίες

Οι λίστες που προέκυψαν αναλύθηκαν με βάση τις οντολογίες Gene Ontology, Human Phenotype Ontology, MGI Mammalian Phenotype Ontology, Reactome Pathways Ontology. Η μελέτη αυτή διενεργήθηκε με βάση τη διαδικτυακή πλατφόρμα BioInfoMiner, ορίζοντας ως απαιτούμενο hypergeometric pvalue και corrected pvalue την τιμή 0.1, επιλογή η οποία καθοδηγήθηκε από το σχετικά μικρό μέγεθος των λιστών μας. Τα αποτελέσματα των συγκρίσεων αυτών παρουσιάζονται στο τέταρτο κεφάλαιο της εργασίας.

### Διαχωρισμός με τη χρήση γονιδίων-κόμβων

Μία προσπάθεια επιστροφής, από το γενικό (την πληροφορία που παρείχε η σύγκριση με τις οντολογίες) πίσω στο ειδικό (τη σύνθεση του data set και τις σχέσεις των δειγμάτων μεταξύ τους) είναι η προσπάθεια αναγνώρισης του αν μπορεί κάποια ομάδα βιολογικά σημαντικών γονιδίων να χρησιμοποιηθεί ως δείκτης για την ανταπόκριση του πληθυσμού στην περιβαλλοντική ρύπανση. Αυτή η προσπάθεια πραγματοποιήθηκε ως εξής:

- Μετά από τη μελέτη των hub genes που αναδείχθηκαν μέσω των οντολογιών, δημιουργήθηκε μία λίστα από 13 hub genes τα οποία εκφράζονται διαφορικά με βάση την εποχή στην Οστράβα, αλλά όχι στην Πράγα, τα οποία ήταν τα εξής: PDGFA, HGF, NGFR, ZFYVE27, MINK1, LTK, RAPGEF1, PAX3, AMMECR1, SH3PXD2B, LZTS1, COL9A1, TLX2. Ύστερα μελετήθηκε, με βάση αυτή την ομάδα γονιδίων και μόνο, πώς ομαδοποιούνται ιεραρχικά τα δείγματα μέσω δενδρογράμματος.
- Επαναλήφθηκε αυτή η διαδικασία άλλες 13 φορές, κάθε φορά αφαιρώντας ένα διαφορετικό γονίδιο από τη λίστα.
- Με κριτήριο τη μέγιστη βελτίωση του διαχωρισμού μεταξύ των πόλεων Πράγας και Οστράβας, επελέγη το γονίδιο του οποίου η απάλειψη βελτίωνε περισσότερο το διαχωρισμό.

- Η όλη διαδικασία επαναλήφθηκε, αυτή τη φορά με νέα αρχική λίστα 12 γονιδίων, έχοντας αφαιρέσει το επιλεγμένο γονίδιο, ώστε με διαδοχικές προσεγγίσεις να μπορέσουμε να προσεγγίσουμε μία ολοένα καλύτερη επίλυση του προβλήματος.
- Μετά από τέσσερις επαναλήψεις της πλήρους διαδικασίας, παρατηρήθηκε πως δε μπορούσε πλέον να βελτιωθεί ο διαχωρισμός με την απαλειφή και άλλου γονιδίου, οπότε θεωρήθηκε πως είχε βρεθεί ο βέλτιστος (δεδομένων των συνθηκών και της πολυπλοκότητας του προβλήματος) συνδυασμός που αναδείκνυε την απόκριση του οργανισμού στις διαφορετικές περιβαλλοντολογικές συνθήκες των δύο πόλεων, η οποία απαρτίζεται από τα εξής γονίδια: PDGFA, HGF, NGFR, ZFYVE27, MINK1, RAPGEF1, PAX3, SH3PXD2B, COL9A1.

Για να μπορεί να κριθεί η επιτυχία ή όχι της μεθόδου, η ίδια προσέγγιση εφαρμόστηκε και στη λίστα των τριών γονιδίων που είχαν αναγνωριστεί ως στατιστικώς σημαντικά διαφορικώς εκφρασμένα μεταξύ των δύο πόλεων το καλοκαίρι του 2009, η οποία θεωρητικώς μπορεί να δώσει τον καλύτερο δυνατό διαχωρισμό που θα μπορούσαμε να παρατηρήσουμε υπό τις δεδομένες συνθήκες. Τα αποτελέσματα δίνονται στα επόμενα δύο διαγράμματα, στα οποία φαίνεται πως με τη λίστα που προήλθε από τα hub genes έχουμε μια σαφή υπόδειξη διαχωρισμού μεταξύ των δύο πόλεων, ο οποίος όμως είναι αρκετά πιο αδύναμος απ' ότι μπορούμε να επιτύχουμε χρησιμοποιώντας τα διαφορικώς εκφρασμένα γονίδια.

Για λόγους διεκόνυσης, στα τελικά διαγράμματα των δύο διαχωρισμών σημειώνονται στο ύψος που ορίζει η οριζόντια γραμμή, το ποσοστό των δειγμάτων του κάθε κλάδου που προέρχονται από την Πράγα. Λόγω της σύνθεσης του δειγματικού πληθυσμού, το ποσοστό των δειγμάτων του 2009 που προέρχονται από την Πράγα είναι  $^{105}/_{288} \sim 0.37$ . Έτσι, τιμές μεγαλύτερες αυτής υποδεικνύουν υπερεκπροσώπηση της Πράγας στο συγκεκριμένο κλάδο, και τιμές μικρότερες αυτής υποδεικνύουν υπερεκπροσώπηση της Οστράβα.



## Κεφάλαιο Τέταρτο: Αποτελέσματα

### 4.1 – Λίστες στατιστικώς σημαντικά διαφορικώς εκφρασμένων γονιδίων

Εδώ θα παρατεθούν με τη μορφή πινάκων οι λίστες που προέκυψαν με την προσέγγιση που επιλέξαμε (Array weights & batch correction) για κάθε ένα από τα τρία υπο-data sets που δημιουργήσαμε (πλήρες, 2009 και 2010) στους πίνακες 4.1 - 4.7.

Πίνακας 4.1.1 – Πλήρες data set, σύγκριση με βάση τη χρονιά [ $|lcf| \geq 0.5$ ] (year-season-city)

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
SLC39A8	3.08806315347814	6.0010580980028 2	14.400830369958 1	3.52424600249238 e-39	1.54276906902209e- 37	78.0678031972314
XRN2	2.73792964597438	5.7979995014230 6	14.475451270508 1	1.6667823428407e- 39	7.42449187451323e- 38	78.8126987416829
NBEA	1.70662387789256	4.8899816414848	19.776434252605 5	1.03667643402484 e-63	1.27077495826998e- 61	134.313635142635
PTPN13	1.34257602834608	4.9606856649704 9	25.628740139852	3.82593692283586 e-91	1.43911908845633e- 88	197.252345453696
PACRG	1.12819883817476	4.8316758786788 4	31.641666311427 5	2.61328307490477 e-118	3.79150041553327e- 115	259.550538222158
AKR1C2	1.1121494910707	4.8729368774684 7	12.736981746507 1	4.12696588130549 e-32	1.20440992788904e- 30	61.884790881682
HS.364519	1.10473876837257	4.7704409417313 5	36.104918558803	2.45088082102084 e-137	8.29704853942921e- 134	303.147677279549
ANKRD20A8P	1.07910257597535	4.7262806642155 5	51.246924883008 7	5.12732933943141 e-194	5.20731567712654e- 190	432.680428763194
OAS1	1.02178770053476	5.1914037276448 9	9.0444852574950 3	4.02734696586915 e-18	4.06982445625543e- 17	29.9455509669165
HS.434660	1.01181772598881	5.0199951904104 8	17.819510630231 3	1.37195183874113 e-54	1.23305689152699e- 52	113.390498529638
CYTL1	-1.00905979862373	5.0733187782632 5	12.575741449574 8	1.90485645993583 e-31	5.24274314555779e- 30	60.3651080172307
AOX1	- 0.967448206047179	4.6982699800953 9	- 35.309890065325	5.02148039723947 e-134	1.01996309828728e- 130	295.563270953483
HS.583993	0.966627580257399	4.8657893840153 1	25.450121828457 7	2.57853677500908 e-90	9.03021361620421e- 88	195.351378892415
LIPC	0.94316962436868	5.4042028288766 5	7.9700740529923 2	1.2227262224234e- 14	8.73892154463905e- 14	22.0330711275414
CST4	0.94139896051401	4.9637069484348 9	18.957961539052 3	7.01308912930059 e-60	7.74184056491052e- 58	125.527115676047
LOC645708	0.935468010651712	5.0046955003034 9	16.947106747701 9	1.41639035605428 e-50	1.12176023561557e- 48	104.186094138116
C14ORF11	0.932875709510054	5.6631627579082 5	12.384131163765 8	1.15842646117173 e-30	3.03221111846909e- 29	58.5716719700854
PNLDC1	0.93087075679996	4.7921737423049 3	24.054904941619 1	8.20869283107968 e-84	2.25317525384987e- 81	180.432342537587
MRPL11	- 0.926963414090373	5.1535551488544 5	11.756526288574 6	3.88227334617744 e-28	8.70383401849405e- 27	52.7974276965263
FAM3B	-0.9197412754663	5.3782532139034 2	4.0525398679247	5.93231660084918 e-05	0.0001381531928416 06	0.318151444076202
DMXL1	0.916970131438708	5.3277462650655 1	14.550525015153 6	7.83589046880145 e-40	3.5686832292141e- 38	79.5635871677695
ZFYVE27	- 0.914938249564314	5.1700828894485 1	12.184911259083 7	7.4590494423765e- 30	1.88443050091482e- 28	56.7218182987102
ATOX8	- 0.913407568221857	5.4144223361552 4	6.5102815392319 9	1.93711483692143 e-10	8.89793680858165e- 10	12.5362396510016
HIAT1	0.904123685073661	5.9658535450460 5	11.840086590301 7	1.80633699169069 e-28	4.14111929742904e- 27	53.5570337689715
RPL10	0.901154519815912	4.9063377800314	16.817583688184	5.53748666121747	4.19691899487497e-	102.828415966163

		6	9	e-50	48	
PGA3	0.895329897401341	4.8637382528181 3	24.705754109372 3	7.49079442025415 e-84	2.53588360440337e- 84	187.406320086016
HS.577416	0.876919058213801	4.8424069814644 6	28.443422051386 5	4.71259347906198 e-104	3.41864995523953e- 101	226.86360290706
COL17A1	0.861388710723007	5.1233204659090 5	9.4180003578454	2.10827722473714 e-19	2.41122336648991e- 18	32.8620519163326
LRRC4C	0.848781587788402	4.8318480970173	28.424217214428	5.75816075030898 e-104	3.89865870534253e- 101	226.664023458658
LOC653321	0.846267655241601	5.4352095809159 7	6.7795809955435 7	3.64491691378679 e-11	1.83528885356562e- 10	14.1715734231508
HS.385684	0.838212259467747	4.7847203994332 6	36.911453743546 4	1.16136169912589 e-140	5.89739470816127e- 137	310.760073768815
COL1A1	0.826758040316522	5.2329503029893 2	4.8794895157145 9	1.46109378167002 e-06	4.18467807293873e- 06	3.86000947354809
FDXR	0.820416810414456	5.1326149432437 2	11.770133581348 7	3.4281502609832e- 28	7.71979912428943e- 27	52.9209265935136
SBF2	0.819050238588166	5.0170345105063 1	14.274328123329	1.24995138323522 e-38	5.2893776039039e- 37	76.8084021771143
CEP63	0.816565581985213	5.4365348372414 3	13.396431252867 7	7.22131781186841 e-35	2.47769269247755e- 33	68.1945966329315
LOC651337	0.810413062484062	4.8999699608314 5	19.516293145050 2	1.71725665687705 e-62	2.02796030316782e- 60	131.516750962612
DHH	0.810189794223569	4.9486843330269 7	18.866903498624 5	1.86594149986576 e-59	2.01601083751454e- 57	124.552261761291
SPIN2B	0.809662857604364	5.4066245802171 7	8.1850487587846 4	2.60439645432676 e-15	1.97832837622607e- 14	23.557063252346
ATG4C	0.80477271575584	4.8255716295042 7	21.315245300252 8	6.07094229804367 e-71	1.01076213080216e- 68	150.906326212752
H2AFJ	0.8040283612201	5.3614772001379 5	10.528519953660 2	2.08952993168956 e-23	3.35779525098722e- 22	41.9910576347579
VCAM1	0.80336240422975	5.0379158490785 4	11.983010939763 4	4.84833088916003 e-29	1.13717433049213e- 27	54.8629643240499
GJB1	0.800516753763587	5.0039917551873 6	16.703405576704 1	1.83817882665097 e-49	1.37268707084318e- 47	101.63369813308
FSHB	-0.79727870551046	5.0783076426222 1	- 10.80400821811	1.92661931363253 e-24	3.34474286312e-23	44.3542261959145
SOX5	0.789288787134733	4.8363614050126 7	25.892602946235 3	2.29328157063624 e-92	9.41076379867783e- 90	200.05632021936
C12ORF25	0.788166224870254	4.8092434315121 7	27.846985435687	2.41657834907769 e-101	1.22713848566165e- 98	220.64842272973
HS.579544	0.777587576506371	4.8207557013256	26.510187040963 5	3.22564530581303 e-95	1.48907516935623e- 92	206.598339299803
CACNB2	0.777568914406756	4.7521478818600 8	33.961091469694 5	2.49435779679527 e-128	4.22211629737547e- 125	282.514208112881
UBE2V1	0.775149231168108	4.9580650333632 6	16.886074078723 9	2.69360778373747 e-50	2.07244550391195e- 48	103.546029417384
ARRDC2	- 0.772582799760704	5.2864702878741 7	- 8.0925114159004	5.08631243980005 e-15	3.75957708432382e- 14	22.8973001348863
PAK4	0.768867390648708	4.954145423948	16.512832066013 8	1.35564918556912 e-48	9.76451994939009e- 47	99.6441931363724
OVOL2	0.765997177898685	4.7973253984642 3	27.906924354225 5	1.28882796940262 e-101	7.27185380958501e- 99	221.274575968812
SEPT3	0.765742643293335	4.8950643326088 7	18.329727052504	5.9187288643933e- 57	5.67081229686588e- 55	118.814902577549
RFX3	0.764782794354782	4.9105285567263 8	19.675002139077	3.09880188651399 e-63	3.70252140699248e- 61	133.222690231614
LOC731002	0.760839982411504	4.7849457295348 6	27.259277839754 2	1.16873092889406 e-98	5.65220538754671e- 96	214.491114121423
PPP6R2	-0.75787160361013	5.2904870807890 1	- 8.9495321285919 7	8.415863160262e- 18	8.19477528817075e- 17	29.217276315502
KCNK9	0.757623261471786	4.9229649950977 7	19.851941674871 2	4.58672716713659 e-64	5.75096309993077e- 62	135.1260641424
HS.567705	0.75718207066606	4.9268954773065 4	16.910990049983 9	2.07205890454673 e-50	1.60639925454783e- 48	103.807262783867
KBTBD6	0.746864308515302	4.9841139173291 1	12.619916728426 5	1.2539624149018e- 31	3.47009326586995e- 30	60.780515503356
ALDH3B1	- 0.736693142030213	5.0769684407400 4	- 9.3025375963289 8	5.29226556859719 e-19	5.80434655666015e- 18	31.9518037286335
NBR2	0.729692334691266	5.2844400347572 3	9.3909387188512 2	2.61760299644049 e-19	2.97364385143732e- 18	32.6480240196655

TPTE2	0.729633328787541	4.8137139761692 7	21.604091339592 4	2.65390658206943 e-72	4.64708193922365e- 70	154.025180234218
HIF3A	- 0.728391795042536	5.0128846676121	- 11.213959352528 7	5.19498394907586 e-26	1.00880032479569e- 24	47.9380016650342
LGI2	- 0.727856234965505	5.1289722080840 1	- 9.5934210255146 5	5.13392459575735 e-20	6.1125601634832e-19	34.2595369740536
WFDC8	0.726835822308721	5.0086463982727 7	13.262805892106 9	2.64436064010174 e-34	8.80528742979453e- 33	66.9042273867526
HS.334300	0.722745040840671	4.7358318904484 6	29.997709939148 5	4.83810347720853 e-111	5.45953099050332e- 108	242.889379255238
SOCS5	0.72200665338917	4.7584862821081 9	24.216058499455 9	1.44779008866544 e-84	4.08437670569062e- 82	182.161247758593
BLOC1S3	0.71341095827886	4.8172982398707 8	25.891655507203 4	2.31655272712629 e-92	9.41076379867783e- 90	200.046261350493
KLHL23	0.712801167139552	4.8478659083073 4	23.568927501483 8	1.54871727736072 e-81	3.93219316721888e- 79	175.211160917999
OR10A6	0.712680192909353	4.8455104324534 8	22.954358136790 3	1.18500292711554 e-78	2.56061483569903e- 76	168.594813669319
ZMYM3	0.712461336864112	5.3295626251622 1	7.4540008435689 2	4.42125066319928 e-13	2.74631325599094e- 12	18.5020943133777
MAMLD1	0.710446405892224	5.2358349655271 1	8.8008864535706 7	2.6394531334489e- 17	2.45225586909237e- 16	28.0881393752223
OPA1	0.706403193902862	4.9275255527362 7	15.822253900311 3	1.79560040335991 e-45	1.15418466433691e- 43	92.4872821302645
MUC4	0.706150344030974	5.3943680380569 3	7.5023893242820 5	3.18241835275792 e-13	2.01249319991341e- 12	18.8253464534613
NRCAM	0.704934834848202	5.2404319108660 4	7.4307859193927 7	5.17370591088673 e-13	3.18449437763428e- 12	18.3475948460099
LCE1E	0.702366568384972	5.0235819072509 7	13.002334952487 7	3.26489154168147 e-33	1.03296693138059e- 31	64.4059389845075
NRCAM	0.702364806138411	5.1398189492779 8	8.1316501078887 7	3.83482143934481 e-15	2.87427649726833e- 14	23.1756591064395
LOC441869	- 0.699503059798186	4.7973889929260 4	18.659799044786 6	1.72381517579456 e-58	1.78643540054791e- 56	122.337400766012
TBC1D21	0.690042362890516	4.9931450230128 3	11.839818541575 5	1.81078414668655 e-28	4.14196481841185e- 27	53.5545924313742
MFAP3L	0.688172663468219	4.8370128344798 8	16.918850935157 1	1.90743069667355 e-50	1.49014355041666e- 48	103.889699941444
TMPRSS7	- 0.687514816490605	5.3489222509164 5	- 5.9716669701544 5	4.65092396126271 e-09	1.80976182952429e- 08	9.43299315953638
PPM1G	-0.68435769211634	5.0858114341463 7	8.5991943057219 7	1.21855766016895 e-16	1.05504446689479e- 15	26.5777757923817
HS.573904	0.680592667238852	4.8361152241188 3	23.378176253928 7	1.21450869869813 e-80	3.00842691316541e- 78	173.159032193091
ITIH5	0.674374571659348	4.8538625988717 3	20.835364634401 4	1.09934817671832 e-68	1.64190883569871e- 66	145.726074046882
NOL1	- 0.673826994145662	4.8994549645237 8	- 14.266499612145 1	1.35163046487245 e-38	5.6959165980268e-37	76.7306071893055
LOC653338	0.672365415621768	4.9127538406784	20.164977848640 5	1.55676685311961 e-65	2.04051552553648e- 63	138.496782382278
CPNE9	0.67183520251698	5.1504437731528 3	8.7457427352690 6	4.01985133490482 e-17	3.65820879545639e- 16	27.6726949165965
LOC342900	0.66981295119645	4.7977418520009 7	24.581506178125 1	2.84470313552545 e-86	9.29791266628109e- 84	186.076792188749
LIM2	- 0.666723068294757	5.0960496745714 2	- 7.5183777646863 8	2.85380759294972 e-13	1.81485722692532e- 12	18.9325157730343
ARL4A	0.660439005889257	4.9357775665725 1	14.338787984629 7	6.56079266132931 e-39	2.82336484188392e- 37	77.4495979468505
TXNDC6	0.657128983217274	5.0179480304033 4	9.3395108147106 1	3.94459307162586 e-19	4.3942764798561e-18	32.2424449658333
PTGER3	0.65576129216113	4.7434757398551 9	27.883395241855	1.64947699639607 e-101	8.8168886186308e-99	221.028819434952
BARX1	0.655450996711565	4.8842166821333 3	15.446034665374 1	8.68299137478646 e-44	4.92650616772801e- 42	88.6266285397672
HS.571822	0.6547331578834	4.7571794843181 4	29.892718365125 1	1.42318086547708 e-110	1.44538248697852e- 107	241.814965979296

LOC650144	- 0.653955116306383	5.0343311789545 2	- 10.324170454416 9	1.1956686246793e- 22	1.83709690654205e- 21	40.2624129936808
DKFZP779B1634	0.652162206484758	4.8312167707413 2	20.641270764837 4	8.99388619162205 e-68	1.30488440231591e- 65	143.631847253594
TCF7	0.651933860089172	4.8358494298193 9	18.930538778240 3	9.41727162526202 e-60	1.0284065658727e-57	125.233467330644
LOC649817	0.651310009404781	4.8486722787757 9	20.206129096622 1	9.97598841193913 e-66	1.33310708304808e- 63	138.940170512477
FZD6	0.648185833871523	5.0368188733182 6	12.315857599227 4	2.19679503060615 e-30	5.69149243133573e- 29	57.9359950878571
LOC646174	0.646301169898293	4.8012143755303 9	29.567786322900 1	4.0432415599178e- 109	3.73301466204774e- 106	238.48221105519
HS.543777	0.645102417534538	4.7767638345645 3	24.481402105596 7	8.3409358515456e- 86	2.49148660318521e- 83	185.004971420195
C14ORF45	0.64328038993979	4.9662075700677 9	14.627422841471 3	3.611505982532e- 40	1.69025137136382e- 38	80.3342282679446
MAGEA11	0.64228495581134	5.0328811463370 6	8.8809288113556 5	1.42861767007688 e-17	1.35725360685695e- 16	28.6944835171327
WBCSR19	0.642054984432141	4.7404203042158 7	22.347454521392 3	8.43890589801973 e-76	1.71411056600577e- 73	162.049953031696
FIP1L1	0.640048111532953	4.7646052765343 9	25.982241093429 2	8.82547873343317 e-93	3.89702443551075e- 90	201.007693306109
CRSP9	0.639574667377039	5.2435810711662	9.4816629974881 8	1.26516221276647 e-19	1.46677938731236e- 18	33.3672031625825
DIXDC1	0.635212593035309	4.9786980679905 7	11.743187473867 8	4.38545095499178 e-28	9.74587306321586e- 27	52.6764407630193
ORC5	0.634954819861105	5.0796860126984 9	11.683522938618 4	7.55736940381942 e-28	1.64352556028244e- 26	52.1361822013703
HIST1H1E	- 0.633052671103951	4.8385433017342 9	9.6377398624964 6	3.58336092465523 e-20	4.33245399414267e- 19	34.6153442406114
BAAT	0.632111261044489	4.7014551496536 8	29.048214337283 9	8.72891626905596 e-107	7.3875728023777e- 104	233.129564429203
HOXA4	- 0.630538993237866	4.8622218309003 8	- 16.276363452731 6	1.60438502928098 e-47	1.10095502414714e- 45	97.1839275692641
LOC642093	- 0.628230136505639	5.0738866858488 2	- 5.0113974379416 7	7.67776029573244 e-07	2.27532341883451e- 06	4.48008043948397
PDLIM2	- 0.627370591942889	4.9079030251981 4	11.470352879994 3	5.21829252796337 e-27	1.08378279987722e- 25	50.218348498033
ZDHC9	0.624678044720732	4.7515918917996 5	23.054869772059 5	3.99672641211048 e-79	9.22517123668045e- 77	169.67790397318
NEK3	0.623814732559723	4.8703554228617 5	14.616410167468 8	4.0355660215228e- 40	1.87147070842856e- 38	80.2237706324059
VCX2	0.623504418225576	4.7829025853817 2	20.362258761591 2	1.84277070311761 e-66	2.49535723478166e- 64	140.622919877384
SYNC1	- 0.623360928209972	4.8426388683003 9	16.322679550642 4	9.89544831630246 e-48	6.836610414991e-46	97.665055540166
ASB10	- 0.622752579674207	5.0087978440097 3	11.843795214943 2	1.74591175974126 e-28	4.01164702079915e- 27	53.5908142772537
HS.543222	0.621463160007237	4.8535544966917 2	20.680420213323 1	5.88648686759705 e-68	8.66422617787183e- 66	144.054198002127
C1ORF91	- 0.620592658089187	5.3364623454176 1	6.8333575031318 4	2.59454753177706 e-11	1.33351339740525e- 10	14.5046658481227
HS.537839	0.619930598461151	5.0615094767546 9	8.5201102810986	2.20488195332603 e-16	1.86451133372016e- 15	25.9924918993962
TOPORS	0.619725261265272	5.4567084505009	9.1306044822835 9	2.05464172495992 e-18	2.12655748201727e- 17	30.6107290635635
PAK3	0.619551040845325	4.9556398306858	14.619870135814 8	3.89724143148826 e-40	1.81561394395389e- 38	80.2584709516385
CDKN2C	- 0.619006859492604	5.0343224785804 8	9.6451279070492 9	3.37452868677472 e-20	4.09458940775198e- 19	34.6747646643434
TRPC2	- 0.618371792487085	4.9531409119720 9	11.172515694035 8	7.51153177847644 e-26	1.43937956117371e- 24	47.5721738493538
CDR1	0.617215002976534	4.7030244662676	35.600442351585	3.06641014539812	7.78561535916583e-	298.344337890536

		3		e-135	132	
RTKN	-0.61638236990773	5.3809673689524 8	- 6.6204520581434 4	9.84382479227762 e-11	4.6870081852026e-10	13.1986248035062
LOC654000	0.615870864215014	4.9329519297829 3	12.570084343603 9	2.00952318122596 e-31	5.51586957527861e- 30	60.3119626015808
IPMK	0.61519905500381	5.2805835194480 9	7.7486917867963 5	5.8243553168983e- 14	3.91736109923306e- 13	20.4959994017898
HTR3E	0.61484240053661	4.8152759843733 8	21.713954841852 5	8.06950047990905 e-73	1.46346155132065e- 70	155.211449362512
DDAH1	0.612468379670489	4.7554197413830 7	24.494214283585 8	7.26787946229023 e-86	2.23674496421271e- 83	185.14218385307
HS.543495	- 0.612308439977636	4.9597962836368 5	- 11.225592768245 3	4.68350708904792 e-26	9.12969251369879e- 25	48.0408312741394
C10ORF113	0.611318146288705	4.7987255658414 3	15.141901135326 7	1.95423622827665 e-42	1.02305273888545e- 40	85.5276726292832
HS.500311	0.6104347336196	4.8255386538162 1	19.897881140562 4	2.79246053735732 e-64	3.54502865217511e- 62	135.620479444138
CACNG4	0.609953777473786	4.9290523404016 1	14.422471117808 1	2.83676719335177 e-39	1.2526177224209e-37	78.2836782652536
PABPC5	0.609480850563754	4.7819344121192 7	21.955062449102 9	5.9188923848024e- 74	1.09295038291006e- 71	157.814645786812
NOXO1	0.608916538825615	5.1312148023126 5	9.5829543820431 3	5.58805589803501 e-20	6.62220486586272e- 19	34.1756674685302
CHAT	0.608753078829486	4.8438437325437 7	17.974593708397 7	2.6283107023217e- 55	2.38331459756957e- 53	115.036406592536
IGFBP5	0.607251370221037	5.0668518604049 6	3.9791667784986 3	8.01363173535043 e-05	0.0001827676710177 83	0.0332401761372436
EVX1	- 0.606277558326988	5.0538401287163 5	- 8.5587551385634 3	1.65099783712116 e-16	1.41259764395978e- 15	26.2780019566516
CALB1	0.605070308324053	4.8052304407327 1	19.177106554691 7	6.63864184850959 e-61	7.57551085544532e- 59	127.875664261648
KCNJ12	0.604556473843265	4.6564597537163 5	30.046728125936 1	2.92470373881202 e-111	3.71291139642186e- 108	243.390583330851
RAD51B	- 0.603914596335537	4.8482035059958 8	- 14.179825548255	3.20938757045574 e-38	1.31961701075095e- 36	75.8704169211119
CNOT7	- 0.603528243986751	5.0321875899479 1	9.9396043812371 7	3.00874403230705 e-21	4.06340483937639e- 20	37.0676340109328
HS.580783	- 0.602334926720566	4.9092874663421 3	- 14.236879775076 2	1.81677862938337 e-38	7.56196875410552e- 37	76.4364148847437
LOC389118	- 0.601406954541107	5.1312455956054	8.0975120607458 6	4.9063134460929e- 15	3.63711820135179e- 14	22.9328089436032
LOC652610	0.59960061092057	4.8215802413977 1	18.681214475583 5	1.36997388360474 e-58	1.43437677957626e- 56	122.566267601274
ROCK1	0.599176762276029	6.0016135220475	7.3345376800965 5	9.88713714052473 e-13	5.91016861678453e- 12	17.7111092808931
NR2F2	0.598789171087425	4.7245859075341 4	18.420657649993 9	2.23648122496365 e-57	2.22683365889518e- 55	119.784339003654
LOC124685	0.597207932174155	4.7851166819775 6	22.327179916196 2	1.05109219641516 e-75	2.09311614642988e- 73	161.831173500887
HS.581320	0.596994920163445	4.8146844731859 9	22.195053191114 4	4.39669340011157 e-75	8.42506003236474e- 73	160.405252665898
BFSP2	- 0.596604037654993	4.9016423746177 1	- 14.026563811648 2	1.47340542230244 e-37	5.82253131085742e- 36	74.3545091507061
HS.196042	0.595900065151612	4.819247730803	20.556110525788 4	2.26121628165628 e-67	3.18957118840295e- 65	142.713246534339
RANBP17	0.595704395039063	4.7632786445865 3	24.429986237634 5	1.44964964609061 e-85	4.20646908734179e- 83	184.454239435933
FLRT2	0.595374126049763	4.8762405981210 8	18.566783199299	4.67390095518652 e-58	4.79476142433074e- 56	121.343782596314
WBSCR27	- 0.595018199587935	5.0701193730182 4	- 7.4564357400075 8	4.34887309904019 e-13	2.70466351462658e- 12	18.5183209735594
HLA-DPB2	0.594698818113095	4.8481689799970 4	15.638988997928 8	1.19205134719937 e-44	7.20623421556955e- 43	90.603054298
TSHR	0.593625094877404	4.7433927947306 1	23.762761980258 8	1.91314870840511 e-82	5.11314165330587e- 80	177.29492783806

ZGPAT	0.593521156920844	5.0480003404481	9.96436354660615	2.45020697498009e-21	3.34916581936713e-20	37.2709691159595
HS.544611	0.593489351196769	4.9443310712556	11.12547966996	1.14045303789449e-25	2.15287008417406e-24	47.1579306435311
TSPAN4	- 0.591809821080471	4.93809361808524	- 10.8533420389201	1.25231902414796e-24	2.20044152409112e-23	44.7813767545403
HS.582332	0.591288956639083	4.75385914495591	23.6846248514851	4.44407426957563e-82	1.15728252004641e-79	176.455128829917
DUSP27	0.590295435566859	5.23314352148669	7.35685306628679	8.51344428359368e-13	5.1282645399868e-12	17.8580951554954
TMC5	0.590255246673521	5.05255021538782	9.14703669315497	1.80614941797734e-18	1.89496420340681e-17	30.7381497879041
HS.512359	0.590149063372789	4.79326847891877	20.48737264997	4.75848805323074e-67	6.62016502309745e-65	141.971920514366
PPARGC1A	- 0.588732471864773	4.93696358442618	- 8.91051657212999	1.13748956888636e-17	1.09708870480626e-16	28.9196077656823
GPM6B	0.587376536806243	4.90278185512051	13.4702522253891	3.51678365235191e-35	1.231601888734e-33	68.9099226174375
KLF4	-0.58715682428389	5.56335136419744	- 5.59825611396952	3.69569339881476e-08	1.28936661485272e-07	7.41677829559507
KIFC3	0.586138782284819	5.26261560271939	7.6896316490534	8.78440227694151e-14	5.81201234688065e-13	20.0915777846276
PDE2A	0.585739105476027	4.8935925585338	13.7511079877746	2.2421961094847e-36	8.46210899019216e-35	71.646994589219
HS.537638	0.585730309110674	5.07617601135886	9.31022846465848	4.97873325359867e-19	5.47229598739698e-18	32.0121953810567
CCL27	0.583935070755282	4.82042807968987	19.4587775452407	3.19290529726575e-62	3.72725818379666e-60	130.898858317567
FSD1	- 0.583654123735921	4.85319541271625	- 11.7504686410184	4.10323776194591e-28	9.17896094941028e-27	52.7424738230712
HS.543210	- 0.582809791273787	5.01624889363957	9.13871095271356	1.92808815475472e-18	2.00837572304502e-17	30.6735693410593
TPSAB1	- 0.582711148816382	5.0442709803319	- 5.1230697259342	4.40379347267242e-07	1.35447990637375e-06	5.01662517332097
CCDC7	0.582706009018555	4.78277431630881	20.9614172241408	2.80648677020201e-69	4.38502763664179e-67	147.086518304607
PCDH20	0.580781943401344	4.75273190859809	21.6492423223583	1.62702634048869e-72	2.89896131824616e-70	154.512708939742
ITGB8	0.580278393850969	4.88157343968349	14.4755807507862	1.66461609426693e-39	7.42449187451323e-38	78.8139925397778
TSN	0.579666639968517	5.30155776399903	9.47702638427463	1.31320325076296e-19	1.52074027534192e-18	33.3303345969732
MGC42105	0.579431400776252	4.75803680271906	20.9418805586578	3.46795548293809e-69	5.3364478613211e-67	146.875649258124
LINS1	- 0.578881951085101	5.24839568767424	- 7.3908856379812	6.7723668361037e-13	4.13094039564379e-12	18.08293980837
HS.542360	- 0.577893575761077	4.88016333368514	- 11.6927762513559	6.94634600128775e-28	1.52041142217841e-26	52.2198719330595
RFX4	- 0.576250878771141	4.94844496866105	- 10.0099029874505	1.67804997800058e-21	2.31552657290406e-20	37.6458188052817
RFK	0.574839037954944	5.13211253732328	11.8605304313918	1.49731446869797e-28	3.45607403274923e-27	53.7433199287342
TADA2L	- 0.573478924171269	5.08387987434016	- 10.4722446715056	3.38506212008385e-23	5.36329031069759e-22	41.5129122351453
BCL7A	-0.57223353265917	5.09629487781983	8.34872651692331	7.86526618986385e-16	6.2996564214714e-15	24.7377394960617
ZNRF2	0.571749953507387	5.03398060524859	12.4400064854499	6.85256342850185e-31	1.81709227623668e-29	59.0932316249021
OR11G2	0.56998654379816	5.0204511106965	10.9779512040014	4.19810415177127e-25	7.53521795847483e-24	45.8652754959015
SLC39A4	0.569809099429387	4.96556636439974	10.8198402348443	1.67800692775723e-24	2.92815092067053e-23	44.4912222240345
COL12A1	-	5.1315077139211	-	4.21066260100292	3.44588955485783e-	25.354127063288

	0.569652146854241	2	8.4332375757973 9	e-16	15	
NR1I2	- 0.567833263112461	5.1688693704285 1	- 7.9293020461195 7	1.63394043239436 e-14	1.1563971450451e-13	21.7475018384494
ZNF501	- 0.567638063709172	4.7961751693096 6	- 16.341743627253 2	8.10968011802721 e-48	5.68013181232306e- 46	97.8631998921074
DNAJB4	0.567079470831323	4.9687305035503 8	12.510143071435 5	3.53939828549191 e-31	9.61126443514863e- 30	59.7495746969101
HERC2P4	0.566851222685886	4.7931221998559 8	16.19119408052	3.89777865834318 e-47	2.63905600360889e- 45	96.3001796399227
PRTFDC1	- 0.564362520893412	5.3756406623423 1	- 6.1081053314333 8	2.12284878906549 e-09	8.6204127555974e-09	10.1975844062008
GPRC5C	-0.56316274147059	4.9285033807682 4	10.351548910618 9	9.47625627288797 e-23	1.46262703202812e- 21	40.4927877900748
HS.538914	- 0.562040700540106	5.0114121560027 5	- 8.4004496288290 1	5.36861998649344 e-16	4.35145287971488e- 15	25.1144421769765
FGF13	- 0.560789003992814	5.0554501780089	5.4862257419605 4	6.7371202810212e- 08	2.28607395837124e- 07	6.83401912178628
PDE4DIP	0.560753936980497	4.6943210726091 6	25.862150721368 6	3.17253736284435 e-92	1.23924190219412e- 89	199.7329781872
LOC653352	0.560451695508772	4.9697029662285 1	10.774493817526 1	2.49135703146836 e-24	4.28125583952499e- 23	44.0993395828267
F7	0.560135025314727	4.7007422865169 7	28.622205526102 5	7.3095954302308e- 105	5.71048086072492e- 102	228.719771050375
FLJ90650	0.560059886113454	4.8034546645954 8	17.158173430820 4	1.5276700500299e- 51	1.26138349821981e- 49	106.403755811058
MAX	0.559716011938826	4.9283257555289 2	10.228936054286 8	2.67649455791079 e-22	3.99742334266793e- 21	39.4640620011172
H2AFB1	-0.55846326325765	5.0633276432674 5	7.6542970942020 5	1.12201318427973 e-13	7.37074120281043e- 13	19.8507668195162
TIMM13	- 0.558282332360973	5.2177275317725 2	- 6.8937984589865 8	1.76626624923258 e-11	9.27518098614584e- 11	14.8816053802343
ETV2	- 0.557679205630797	5.2238268002742 5	- 6.6339158644093 3	9.05667713339984 e-11	4.32845237490865e- 10	13.2802062122917
SIL1	0.554618809375326	4.8921051184815 7	12.774899258598	2.87637744120998 e-32	8.51676072680132e- 31	62.2435266656777
EPS8L2	0.554501438566973	4.9415629106526 7	10.761344343387 1	2.79337688878305 e-24	4.77601610816173e- 23	43.9858871646456
TIRAP	- 0.554488311230957	5.0670412106702 9	- 8.4775602617556 5	3.02869755991755 e-16	2.51302715837603e- 15	25.6792232425872
HS.564723	0.553756627064964	4.7821421373230 5	22.997406243575 8	7.43949971773013 e-79	1.67901242518371e- 76	169.05867956248
GHRH	0.553123581784385	4.7880311701466 1	20.910104652793 6	4.89281276109246 e-69	7.41662782114254e- 67	146.532688133761
LOC642112	0.55150986564887	4.9269869870648 5	12.075535936303 7	2.06025917846465 e-29	5.04192583529808e- 28	55.7128052390867
GAMT	- 0.551402957937094	5.0323736403479 1	8.0442166435481 6	7.1971850604441e- 15	5.24728007709047e- 14	22.5552187330673
TCOF1	0.550532509248082	4.9478328636500 2	10.182120728796	3.9706814497125e- 22	5.81069752208648e- 21	39.0733265356272
RNF38	0.550014848273236	5.1252115493944 4	6.1860702590859 2	1.34741188378826 e-09	5.60602830469218e- 09	10.6411003103982
EVX2	0.547267267459381	4.9339913682224 5	9.7043988351784 5	2.08227572343022 e-20	2.58528022581385e- 19	35.1525652882328
HS.566771	- 0.546218695783032	4.9917690606482 2	- 10.272348232329 5	1.85475919253823 e-22	2.80311523205628e- 21	39.8274078339219
IL17RD	0.545642075005792	4.7467350444702 2	19.005840432495 6	4.19125569717601 e-60	4.67762558906808e- 58	126.039942487348
WDTC1	0.545387388509401	5.1162851148121 6	7.0306702775172 5	7.32234193756692 e-12	3.99600777635302e- 11	15.7451769360624
LDHC	0.543714260524404	4.7847493374620	21.262006825204	1.08092228047149	1.71528854382319e-	150.33151038528

		6	1	e-70	68	
PLK2	0.543671682545905	4.8693882102273 4	12.967265021830 6	4.57176705966317 e-33	1.41990416690946e- 31	64.0713254848878
HNT	0.541949683679878	4.7015253797384 8	27.969074945111 7	6.71848735225012 e-102	4.0137033852619e-99	221.923470675974
HS.544326	- 0.541512022333715	4.9922601792527 8	- 10.299592126880 2	1.47271441251135 e-22	2.25254330925682e- 21	40.0559262534913
LOC126147	- 0.540619424040348	5.0187077469101 2	- 8.0861843446252 3	5.32342697394649 e-15	3.92057464448155e- 14	22.8523962766291
PRG2	0.540423406401691	4.7772054231511	21.458680315934 8	1.28300597509184 e-71	2.20850994627673e- 69	152.455062754004
ZFAND4	0.540124361262538	4.9505959730610 7	12.083709062211 4	1.90994199483078 e-29	4.68535528973463e- 28	55.7880402160798
LOC644701	- 0.539202905059257	5.3014589898828 9	- 5.5886860786363 7	3.89171777376636 e-08	1.35403513910144e- 07	7.36659402127577
PF4	0.539028531188705	5.5208349617537 6	8.6003833397049 3	1.2077059485554e- 16	1.04743480901184e- 15	26.5866055970056
ARPP-21	- 0.538759720747791	4.9121860492998 7	- 9.5814392292590 9	5.65701560341471 e-20	6.69611310819111e- 19	34.1635315968276
WDR93	0.538616557511831	4.7270562608470 3	28.374047814349 5	9.72060795185143 e-104	6.1701558974377e- 101	226.142482283054
HS.545811	0.538184517348529	4.7851866117730 3	20.164362833524 6	1.56715449972278 e-65	2.04051552553648e- 63	138.490156295245
SIGLEC12	- 0.537996582559558	4.9587907951154 8	- 8.6956268180341	5.88242324221426 e-17	5.27289412603072e- 16	27.2967572705131
ZNF14	- 0.537270304838969	5.0644747026587 7	- 9.1924478701308 5	1.2638376675402e- 18	1.34685575567034e- 17	31.0911108443675
TBL1Y	0.53570677381034	4.8728146380380 3	14.805386298402 9	5.98017904599298 e-41	2.86484426373135e- 39	82.1234360663992
OIP5	- 0.534482830707843	4.9597994626148 7	- 11.212800486809 4	5.2488814125747e- 26	1.01732136691047e- 24	47.927761625195
SLC4A3	- 0.534293566015038	4.9379259848120 9	- 9.9603288603230 4	2.5336420302283e- 21	3.44928531621966e- 20	37.2378117709801
SERINC2	-0.533484807392	5.0632766683480 5	7.2330901413413 2	1.94317842789877 e-12	1.12706568325185e- 11	17.0473774159528
TBX21	- 0.532311419133914	5.2185065688055 4	5.5294137173040 8	5.35125954198705 e-08	1.83544045621143e- 07	7.05745082918289
PIKFYVE	0.531483862146175	4.9632170440274 2	12.022188803304 1	3.37594467513544 e-29	8.01076965436344e- 28	55.2223923710638
HS.538076	- 0.530709312722094	5.0193906899594 3	- 8.3861966942911 2	5.96537428589412 e-16	4.82743754960484e- 15	25.0104656869244
NECTIN2	- 0.529964682767459	4.8481780521007 5	9.7864573012667 1	1.06386546039827 e-20	1.35736402208604e- 19	35.8172653671688
PCDHGA5	0.52908857807048	4.8400496653014 8	17.741891723498 4	3.13399235987229 e-54	2.76772403537939e- 52	112.567729561653
RNF186	0.529080093555435	5.0300392442277 9	8.6849863322427 6	6.37642116361283 e-17	5.69559659961758e- 16	27.2171395025115
CCDC144A	- 0.528804926437083	5.0104015772976 1	- 7.3993687357343 2	6.3961450317793e- 13	3.91085183279654e- 12	18.1391132649087
MAGI1	0.527746412499368	4.7568792798052 6	15.464431922644 8	7.18770935995005 e-44	4.10103237413779e- 42	88.8147336876625
PLXND1	-0.5276400312058	5.2667809786398 4	5.7424359811256 2	1.68168036842572 e-08	6.1413685083537e-08	8.18189624456489
MAPK8IP1	0.52631293423851	5.0070106526094 3	4.6318704984345 4	4.70344563323792 e-06	1.26538261857389e- 05	2.73648325096682
HS.543860	0.523559597179617	4.8441562112049 2	17.166511842783 6	1.39883364257195 e-51	1.1644716781935e-49	106.491495683391
LTB4DH	0.521827551560999	4.8860627598292 3	13.869815426032 3	6.95518989776167 e-37	2.67564047733589e- 35	72.8110628391355
LOC646098	0.521459407615629	4.7405796978330	18.122303768479	5.4339302164612e-	5.06302708975963e-	116.606450704691



		1	2	56	54	
PCDHB1	0.520014814744215	4.7219570396926 9	24.578767954804 8	2.92962982789479 e-86	9.29791266628109e- 84	186.047481448153
OR51I1	0.519919405773805	4.8449043946760 8	15.623212003337 1	1.40259253392085 e-44	8.42883418609476e- 43	90.4411607773975
SORD	- 0.519801461838266	5.0894876938415 6	- 7.2615078525634 3	1.60926345509206 e-12	9.40914199764825e- 12	17.2325611727095
LMO3	0.518223299603351	4.8255158581569 8	14.549363434283 4	7.92802242759984 e-40	3.59450873994214e- 38	79.5519578850308
HS.543727	0.518193705188695	4.7945280264532 8	18.526632140736 6	7.18740392557658 e-58	7.29952742681557e- 56	120.915106201737
RAB27B	0.518129278450216	4.8958382790924 7	11.479994418636 5	4.78360119749557 e-27	9.95537986921414e- 26	50.3046667595713
HS.577994	0.517674371914614	4.7826448545813 6	19.109119495065 6	1.37993928869291 e-60	1.55718482399614e- 58	127.146704279992
ANXA7	- 0.517571364254955	4.8084801214814 5	- 13.415357618175	6.00586333005761 e-35	2.0746785027233e-33	68.3778275742766
IFLTD1	0.517314083694882	4.6922820288422 1	25.555196534788 6	8.39071723835299 e-91	3.04343300973975e- 88	196.469922510885
HS.581083	0.516141353160109	4.7751934224558 3	22.388672589490 8	5.40058657756847 e-76	1.11935423024052e- 73	162.494707039382
CDRT15P1	0.515866148106828	5.1065294954228 4	8.8800667010842 8	1.43812392898279 e-17	1.36373357822121e- 16	28.6879319582959
HS.541249	0.515701961098002	4.8825566610981 8	15.581743559336 9	2.15025107827835 e-44	1.26809571082776e- 42	90.0158818016871
LPAL2	- 0.515226085533755	4.8882357675231 5	- 10.637342608993 2	8.18442942661463 e-24	1.36264041404423e- 22	42.9201365830119
HS.485576	-0.51424871264507	4.823892543592	16.877723848277 5	2.94108972138692 e-50	2.24584264739891e- 48	103.458501491487
SDC1	0.51379770995727	5.3067806131520 5	5.8573213510933 8	8.87569359910217 e-09	3.32622069561117e- 08	8.80363506620489
HS.575565	0.513711427919458	4.8152543716000 3	17.056235402879 3	4.48198329037084 e-51	3.58417498401624e- 49	105.331909126861
PCDH10	- 0.513688998272908	5.2057443473620 2	6.0547556539339 4	2.88965653315149 e-09	1.15223210642664e- 08	9.89685885536156
HS.549678	-0.51220456172744	4.7959366661755 4	15.053010951969 1	4.83661297300567 e-42	2.45603206769228e- 40	84.6258560061774
LOC202134	0.511694332369665	4.8300190170143 5	13.368540856405 9	9.47277968535182 e-35	3.23924412405499e- 33	67.9247921770698
SLC26A6	- 0.510981679693722	4.9984210958654	9.4417509580516 7	1.74302176309826 e-19	2.00024056791254e- 18	33.0502383363761
CYP4X1	- 0.510424904695123	5.1157751633612 6	6.6921401522307 9	6.3063494361119e- 11	3.08810438154062e- 10	13.6345824424452
TSPYL3	0.51008975145877	4.7344745010124	21.365054428109 2	3.5387535815422e- 71	5.98993022902377e- 69	151.444129795068
FOXB2	- 0.509781268152903	5.0791725841483 6	- 7.3685898710526	7.86822284922288 e-13	4.765037045719e-12	17.9355445014472
GSTTP1	0.508773942115576	4.7163139490952 9	21.292013613239 3	7.80880309041057 e-71	1.27913232558403e- 68	150.655491519138
LOC284861	0.508532333546181	4.8181893893384 7	16.946542005210 8	1.42484315079174 e-50	1.12176023561557e- 48	104.18016902482
COL25A1	- 0.506715585148686	4.8179075305064 2	- 14.629388090802 8	3.54064691973611 e-40	1.66475972763148e- 38	80.3539430539682
CEACAM5	0.506459536202591	4.8269302701675 5	15.423959859448 6	1.08921773724947 e-43	6.11165488370473e- 42	88.401017219215
ESRP1	- 0.506440108467536	4.8434799014716 1	12.455809477609 6	5.90572194143076 e-31	1.5742391610806e-29	59.2409561047741
GSTA2	0.506259562183291	4.8221830889411 7	13.304477201784 9	1.76520351355073 e-34	5.91663593518851e- 33	67.3060085636014
ECE2	- 0.505584483510533	4.8979377962314 9	- 11.263271809825 4	3.34657749991286 e-26	6.57405050079594e- 25	48.3743022188963
ARVCF	- 0.505237486981841	4.9839427884424 4	- 8.6907636251024	6.10331894159736 e-17	5.46126054368836e- 16	27.2603595516853

			2			
STEAP2	0.504913353653438	4.7261838875308 1	19.907688030104 6	2.51174122331196 e-64	3.22901821062738e- 62	135.72603628285
ZBED3	0.504369035641035	4.9395720429116 1	10.737083443494	3.44922181259733 e-24	5.85790915196295e- 23	43.7767855488656
CEP112	0.50420288120136	4.7937769462649 3	16.803371687122	6.43011576948373 e-50	4.83735227813902e- 48	102.679597521642
LOC401442	- 0.503906113576167	5.2085677478629 5	- 6.1558496183658 4	1.60791194207505 e-09	6.63280003400252e- 09	10.4686178839617
PDE7A	0.50337925316653	4.9453509710305 4	12.863116729941 2	1.2394516846303e- 32	3.75757352510607e- 31	63.0801275320354
SEPT4	- 0.503240582351942	5.0932941139404 2	- 7.3040777445166	1.21200481108625 e-12	7.17314735512352e- 12	17.5110493565226
PPP2R4	- 0.502938189918122	5.1305122872499 6	- 6.8618280727513 8	2.1653931570888e- 11	1.12317328413656e- 10	14.681884607189
CKLF	- 0.501145806837483	4.8499042809291 3	- 13.265852118685 7	2.56742492307696 e-34	8.57722615749001e- 33	66.9335788401288
DLL3	0.500727589824882	4.8350372901097 3	12.778074291925 9	2.79066349251392 e-32	8.31143062462503e- 31	62.273588825395
VCY	0.500485172652205	4.8193712794779 6	15.148968921586 6	1.81823867818731 e-42	9.56789223609859e- 41	85.5994552284866

**Πίνακας 4.1.2 – Πλήρες data set, σύγκριση με βάση την πόλη [ $|lcf| \geq 0.5$ ] (city-year-season)**

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
XRN2	-2.86455595174407	5.79799950142306	23.1785060013648	1.052023131405246e-79	6.31133090267268e- 78	170.895045529443
SLC39A8	-2.75184371116234	6.00105809800282	-19.640087975829	4.5168343283904e-63	1.68650622937989e- 61	132.66217653392
NBEA	-1.6886940042436	4.8899816414848	-	7.99642293965529e- 111	1.14382635739633e- 108	242.442127657536
PTPN13	-1.31848538185675	4.96068566497049	-38.519619702894	3.48499481144679e- 147	1.76968036525268e- 144	326.017751185515
C14ORF11	-1.13896135320867	5.66316275790825	-	1.58710823480983e-79	9.42612352791147e- 78	170.482791596775
LOC653321	-1.10253673624841	5.43520958091597	13.5178145152524	2.21020062235403e-35	2.58009166903765e- 34	69.0918119504298
LIPC	-1.0864960480396	5.40420282887665	-14.051380839447	1.15169174769175e-37	1.53498443432512e- 36	74.3264904864209
HS.364519	-1.08138544084245	4.77044094173135	-	2.12387049227443e- 203	7.19000957317969e- 200	455.047916211882
PACRG	-1.0771654078352	4.83167587867884	46.2353765239374	1.4863106867677e-176	2.51582855580213e- 173	393.468336158327
AKR1C2	-1.05968120179923	4.87293687746847	-	4.34293412772291e-58	1.30108669619923e- 56	121.212808331689
HS.434660	-1.03621510482407	5.01999519041048	27.9294067427036	1.01819669700173e- 101	1.05518425048465e- 99	221.508048008127
ANKRD20A8P	-1.00965088934835	4.72628066421555	-	8.86588976845376e- 259	9.00419764884164e- 255	581.743450160426
AOX1	0.997845908445604	4.69826998009539	55.7378773175816	1.10520180519171e- 208	5.61221476676352e- 205	467.157776151234
DMXL1	-	5.32774626506551	23.5364755574575	2.19833778222781e-81	1.3953949072691e-79	174.755498381192
HIAT1	-	5.96585354504605	19.3793066446622	7.52010582909393e-62	2.64270570243176e- 60	129.855544384539
SPIN2B	0.943466393562901	5.40662458021717	14.5969389520538	4.91045499297005e-40	7.38823420868204e- 39	79.7628177110562
ATOH8	0.927369378226219	5.41442233615524	10.1159393875223	6.92139485706684e-22	3.98942600274522e- 21	38.2054278835855
RPL10	-	4.90633778003146	26.4340353689037	7.23699419655066e-95	6.74301954680445e- 93	205.754148655756
NRCAM	-	5.24043191086604	14.8686667874435	3.14925479932316e-41	4.96643350029907e- 40	82.4995392923792
CST4	0.920931691444466	4.96370694843489	28.3833494650275	8.82109374614917e- 104	9.63301377267645e- 102	226.249797131794
MUC4	-	5.39436803805693	-14.921998865816	1.8330057250587e-41	2.9316545108183e-40	83.0388129202482

KCNK9	- 0.910331536148219	4.92296499509777	- 36.5062800280286	5.37746232773402e- 139	1.95048240715953e- 136	307.20182537176
ALDH3B1	0.90285219096408	5.07696844074004	17.4481652547089	7.09523567639089e-53	1.79698786856424e- 51	109.237572713043
CYTL1	0.89816410542961	5.07331877826325	17.1313147083021	2.02885741785161e-51	4.84825316134141e- 50	105.893026642992
SBF2	-0.8956458915279	5.01703451050631	- 23.8890992727508	4.8997329658525e-83	3.31744586674653e- 81	178.553347736441
H2AFJ	- 0.886270237903848	5.36147720013795	- 17.7615320685321	2.54300307504351e-54	6.74327395042869e- 53	112.557944326166
PGA3	- 0.882886266630016	4.86373825281813	- -37.285338628717	3.43488849479745e- 142	1.39538910212652e- 139	314.543243083003
LOC645708	- 0.863858450346057	5.00469550030349	- 23.9512129831307	2.50865409122976e-83	1.72147911827902e- 81	179.221762053521
PNLDC1	- 0.858390740606085	4.79217374230493	- 33.9482650887786	2.82874463588829e- 128	7.56019224265303e- 126	282.560958079689
HS.583993	- 0.837980792101116	4.86578938401531	- 33.7662669658487	1.68955568604424e- 125	4.28978188686633e- 125	280.776814505224
PAK4	- 0.831853402957148	4.954145423948	- 27.3423112326663	4.87089421936938e-99	4.89790115761539e- 97	215.34655242591
RNF38	- 0.830475352397817	5.12521154939444	- 14.2950436088819	1.01620723362165e-38	1.42549732937314e- 37	76.7445346427298
GJB1	- 0.822186466057817	5.00399175518736	- 26.2556773673564	4.81176724774589e-94	4.24941810157454e- 92	203.862444981203
HS.577416	- 0.813508137172568	4.84240698146446	- 40.3833683899257	1.44053983428621e- 154	8.12784586500597e- 152	342.981045675425
DHH	- 0.810220684466063	4.94868433302697	- 28.8758973213897	5.22160227014734e- 106	6.31316579233528e- 104	231.371817363035
LCE1E	- 0.808947487116938	5.02358190725097	- 22.9190256680533	1.73650692158664e-78	9.96382163595139e- 77	168.094029973246
FDXR	- 0.806969208926639	5.13261494324372	- 17.7183023278939	4.02782878351166e-54	1.05701884044818e- 52	112.099190006664
LOC731682	0.796086949151888	5.87581752812046	3.31329705243997	0.000993565599924324	0.00157297774479056	- 2.68790275825046
OAS1	- 0.790484785977621	5.19140372764489	- 10.7086568672432	4.41453080292851e-24	2.88136085054897e- 23	43.2204351036104
HS.567705	- 0.788637698714066	4.92689547730654	- 26.9565648747769	2.8555952597825e-97	2.81567237459719e- 95	211.281253508881
MFAP3L	-0.78410426025085	4.83701283447988	- 29.5029581491477	7.89410457854133e- 109	1.041201637658e-106	237.856684332713
RFX3	- 0.778520363988992	4.91052855672638	- 30.6524335099217	5.94931747525969e- 114	9.5906775045615e- 112	249.634700549283
CEP63	- 0.770006466289822	5.43653483724143	- 19.3335107431596	1.23180752922497e-61	4.24075839552839e- 60	129.363075613524
C12ORF25	- 0.769743539670064	4.80924343151217	- 41.6221703890261	2.26061453236219e- 159	1.5305867460447e- 156	354.016444316262
HS.385684	- 0.766205339098881	4.78472039943326	- 51.6381432781723	2.49230469107484e- 192	6.32796161063901e- 192	436.54616293
TPTE2	- 0.762209316695356	4.81371397616927	- 34.5401243326813	8.70306255807148e- 131	2.38887306323713e- 128	288.33475315029
PPM1G	0.761684536888792	5.08581143414637	14.6476532634498	2.94496664873624e-40	4.48412013261848e- 39	80.2721820720837
LOC651337	- 0.760233077260131	4.89996996083145	- 28.0192150524857	3.97318559777794e- 102	4.15996628154977e- 100	222.447734506156
ROCK1	- 0.756281067572185	6.0016135220475	- 14.1683511979843	3.59811224155553e-38	4.89189128851914e- 37	75.485205569775
LOC731002	- 0.751442479180296	4.78494572953486	- -41.203590676027	9.29764339335553e- 158	5.90167914393243e- 155	350.308963444849
NRCAM	- 0.749569583074419	5.13981894927798	- 13.2814716311139	2.20662121953547e-34	2.428000553153e-33	66.8012616546286
NBR2	- 0.744213822516768	5.28444003475723	- 14.6583557362574	2.64354251172745e-40	4.03726582693293e- 39	80.3797592845617
HPCAL1	0.743045067619212	5.44136814306527	10.4495107171572	4.11164284605349e-23	2.55399662045989e- 22	41.0059311592638
COL1A1	- 0.739906891845717	5.23295030298932	- 6.68330586416787	6.66342977711668e-11	1.85968103370148e- 10	13.2362124217203
OVOL2	- 0.732949243395502	4.79732539846423	- -40.867400044555	1.86888609020817e- 156	1.11649453718554e- 153	347.315522319033
PDE2A	- 0.732357245957279	4.8935925585338	- 26.3132661478491	2.60935732133607e-94	2.34518875712293e- 92	204.473519993099
SEPT3	- 0.727115772063716	4.89506433260887	- 26.6375962904309	8.35442633124672e-96	7.92967792711604e- 94	207.910029415748
IPMK	-	5.28058351944809	-	3.78248063481187e-37	4.93130594700249e-	73.1422298263632

	0.722710593381855		13.9314220796509		36	
OPA1	- 0.717919556398901	4.92752555273627	- 24.6098961859788	2.09694943492618e-86	1.52118703293645e-84	186.298111268802
CACNB2	- 0.714799835363148	4.75214788186008	- 47.7799285413332	4.58160509537753e-182	9.30615626973083e-179	406.11666345526
TBC1D21	- 0.714334788157763	4.99314502301283	- 18.7581237316135	6.00132559575873e-59	1.87536808463156e-57	123.187583272256
LTK	- 0.713845812451853	5.43049777960216	- 10.3428522850866	1.02030111081242e-22	6.18637497397667e-22	40.104261507602
ATG4C	- 0.713245156143024	4.82557162950427	- 28.9117396537387	3.59840196164935e-106	4.45675247835498e-104	231.743585822374
KIFC3	- 0.711911744922086	5.26261560271939	- 14.2938605021468	1.02829714644352e-38	1.43933562761381e-37	76.7327541046553
LRRC4C	- 0.710557552059689	4.8318480970173	- 36.4174997200645	1.24949245357689e-138	4.37580874431962e-136	306.360354386347
COL17A1	- 0.709678559386604	5.12332046590905	-11.87516121546	1.30903705121446e-28	1.088827217475299e-27	53.5773469611524
NOXO1	-0.70660671338551	5.13121480231265	- 17.0191439885508	6.62824400862369e-51	1.54042210873186e-49	104.712303106917
SOX5	- 0.703588038509822	4.83636140501267	- 35.3245411023796	4.36010110207736e-134	1.34185414523326e-131	295.919976636408
UBE2V1	- 0.702377507827551	4.95806503336326	- 23.4170545677321	7.9808870096484e-81	4.94231027256032e-79	173.468149477487
MAMLD1	- 0.700100642578108	5.23583496552711	-13.273146115083	2.39215196630289e-34	2.62078698703044e-33	66.7209036761389
EVX2	-0.69807141074492	4.93399136822245	- 18.9446819375729	8.08918010314513e-60	2.59980104833994e-58	125.187270936089
ORC5	-0.69785165419538	5.07968601269849	- 19.6522557035592	3.9610270769379e-63	1.48993299975486e-61	132.793220814295
KLHL23	- 0.696230013456187	4.84786590830734	- 35.2324190009402	1.0601080145283e-133	3.16660499869101e-131	295.033144792629
HS.334300	- 0.694922790644846	4.73583189044846	-44.142592797332	6.76724690503428e-169	7.63646217416979e-166	375.887224417869
SOCS5	-0.69088406647398	4.75848628210819	-35.463838205515	1.14004523822946e-134	3.61821857483074e-132	297.258931420898
MRPL11	0.6893488863931	5.15355514885445	13.3805451188201	8.42873087526179e-35	9.46926889039366e-34	67.7592573763666
TOPORS	-0.6868389531699	5.4567084505009	-15.487222991303	5.68677570395228e-44	1.01324375525157e-42	88.7946428483907
FLRT2	- 0.683790324388803	4.87624059812108	-32.635326590137	1.23074407134207e-122	2.49988735771002e-120	269.599473151578
PTGER3	- 0.675013106576765	4.74347573985519	-43.926882778406	4.29062697122052e-168	4.35756075197156e-165	374.045525881141
DNAJB4	- 0.674096523181064	4.96873050355038	- 22.7592916364072	9.77554873445004e-78	5.48510900260081e-76	166.368781104102
C10ORF113	- 0.672414418994024	4.79872556584143	- 25.4899180427512	1.68526865943005e-90	1.35838004009299e-88	195.713106138577
KBTBD6	- 0.669494385686174	4.98411391732911	- 17.3133068256025	2.9613401020402e-52	7.28217193131243e-51	107.812424215241
BLOC1S3	- 0.669169124378686	4.81729823987078	- 37.1684320680212	1.03086124088002e-141	4.0267026009144e-139	313.446460076534
OR10A6	- 0.663751897921916	4.84551043245348	- 32.7185872401103	5.36919974751436e-123	1.13603317991158e-120	270.427639132773
ITIH5	- 0.663170979439247	4.85386259887173	- 31.3576663544794	4.60398718607352e-117	8.06174032099356e-115	256.787744094227
CPNE9	- 0.658121745537796	5.15044377315283	- 13.1116843494736	1.13974916557243e-33	1.21845184479512e-32	65.1670190298117
TADA2L	0.656420883293385	5.08387987434016	18.3452236390093	5.01441636989727e-57	1.45504036150505e-55	118.77199041172
TCF7	- 0.656090590732021	4.83584942981939	- 29.1569135106391	2.82881114505728e-107	3.54684024558046e-105	234.283065360271
LOC646174	- 0.652014073808144	4.80121437553039	- 45.6519299047413	1.92672045096873e-174	2.4459716125048e-171	388.618742791546
TMPRSS7	0.65121023394997	5.34892225091645	8.65671336099412	7.89742648748243e-17	3.31020484551678e-16	26.6756452601606
NOL1	0.650247673367677	4.89945496452378	21.0700775112217	8.64827629637666e-70	4.04755272193555e-68	148.101486687558
TXNDC6	- 0.647502191083145	5.01794803040334	- 14.0842281576434	8.31029685411154e-38	1.11198122332486e-36	74.6514872033678
DKFZP779B1634	- 0.642958317867232	4.83121677074132	- 31.1445403155727	3.98843593674704e-116	6.8655178599327e-114	254.632043288314
HS.573904	-0.64242536446755	4.83611522411883	- 33.7725882449315	1.58775775877183e-127	4.13468405079146e-125	280.838850800389

ARL4A	- 0.642384074705253	4.93577756657251	- 21.3448347424827	4.40561344117511e-71	2.13063857659878e-69	151.073295619059
GPM6B	- 0.638696318885809	4.90278185512051	-22.416708749135	3.98654783184582e-76	2.13091472527506e-74	162.666570557729
NEK3	- 0.634551596966221	4.87035542286175	- 22.7546593674849	1.02780093093436e-77	5.73535508492822e-76	166.318739136736
ZFYVE27	0.632321946050453	5.17008288944851	12.8880479889884	9.76538954405675e-33	9.97759519209662e-32	63.0293830320129
LOC441869	0.629478836306744	4.79738899292604	25.699032861884	1.80679600848267e-91	1.50408362804508e-89	197.942803577898
WBSCR19	- 0.627665006040721	4.74042030421587	- 33.4350563284752	4.41383842635221e-126	1.06730816804841e-123	277.519501501731
PLK2	- 0.627621665844696	4.86938821022734	- 22.9101484740822	1.91150446981538e-78	1.09063142671039e-76	167.998167703513
TRPV1	- 0.626162412292483	5.08455815063748	- 11.6134952237571	1.42875688577326e-27	1.11963386820318e-26	51.201741387202
HS.579544	- 0.625562262242399	4.8207557013256	- 32.6401632938623	1.17281657756061e-122	2.43084186973583e-120	269.647605015511
LOC649817	- 0.620966972507672	4.84867227877579	- 29.4837130230767	9.62943470716517e-109	1.25380178058935e-106	237.658270826426
MAGEA11	- 0.620458868545743	5.03288114633706	- 13.1299157176792	9.55954442059797e-34	1.02520309541281e-32	65.3420382440552
CACNG4	- 0.620409452735132	4.92905234040161	- 22.4511949781855	2.74428800976156e-76	1.47465550408139e-74	163.039365167487
WDTC1	- 0.619044741540658	5.11628511481216	- 12.2132706510233	5.72727278175274e-30	5.14290257059954e-29	56.6886200004909
HLA-C	- 0.617781371782675	6.20715102548955	- 2.96667788719885	0.00316452493101805	0.00474585280558466	-
HS.571822	- 0.617700859792554	4.75717948431814	- 43.1616047420208	3.14528277803414e-165	2.90395380851952e-162	367.466518206072
HIF3A	0.617015467103462	5.0128846676121	14.5381212316105	8.87747604960491e-40	1.31428056501148e-38	79.1728915904066
EPS8L2	- 0.616803707448938	4.94156291065267	- 18.3201543855663	6.55699710322328e-57	1.88648335921631e-55	118.504389808865
SEC16B	-0.61574086784885	4.93139151842679	-18.991452890091	4.89251933546101e-60	1.59257776829942e-58	125.689012342256
ZDHC9	- 0.615318247910284	4.75159189179965	- 34.7555790501333	1.07131300440303e-131	3.10864424934778e-129	290.425812103237
SIL1	- 0.612469818588125	4.89210511848157	- 21.5906583304234	3.06976277370145e-72	1.54339162028277e-70	153.732542010199
CRSP9	- 0.611338988599886	5.2435810711662	- 13.8705492412156	6.90495528556237e-37	8.83208134510975e-36	72.5428802044794
LOC653338	- 0.610889395076169	4.9127538406784	- 28.0396940925895	3.20620437945059e-102	3.39189704976044e-100	222.661903967686
RFK	- 0.608529499344045	5.13211253732328	- 19.2157702289138	4.37826704591388e-61	1.47726512020935e-59	128.097544056455
BAAT	- 0.608334185780426	4.70145514965368	- 42.7844977735473	8.33227531465222e-164	6.5094298535083e-161	364.198496783886
WNT9B	- 0.605638467991957	4.99932532750603	- 15.7820010281257	2.72282926204503e-45	5.07395485969345e-44	91.823935085188
WFDC8	- 0.60342888880902	5.00864639827277	- 16.8516911390095	3.86803809651346e-50	8.71037581112875e-49	102.953077055338
HS.580290	- 0.600617846976801	4.92090357462728	- 19.0243238196427	3.43573291371511e-60	1.12197117272317e-58	126.041734243994
PABPC5	- 0.600037275123857	4.78193441211927	- 33.0804298838264	1.47387971665238e-124	3.25407008746122e-122	274.01706185128
ZBED3	- 0.598153954041776	4.93957204291161	- 19.4880862230543	2.32777289481809e-62	8.41311797856674e-61	131.025809660126
HS.196042	- 0.597251875552411	4.819247730803	- 31.5314059663949	7.95118721847288e-118	1.44200459626448e-115	258.541179485138
LIM2	0.596734436440831	5.09604967457142	10.2985970282242	1.48518353748137e-22	8.88311190027138e-22	39.7318383434437
LOC653352	- 0.595331514246056	4.96970296622851	- 17.5160374230787	3.45376716486739e-53	8.79109256300582e-52	109.955711231214
FZD6	- 0.594261221097185	5.03681887331826	- 17.2806795648774	4.1827602927487e-52	1.02361719357002e-50	107.46798708223
PATZ1	0.592340421857622	5.21526539844298	10.9637886687034	4.75508822486948e-25	3.27630095059528e-24	45.4325917652083
WBSCR27	0.591194620057724	5.07011937301824	11.338348612837	1.70977142811939e-26	1.26840311351209e-25	48.7354370265898
SLC39A4	- 0.590346045513663	4.96556636439974	- 17.1560154550475	1.56290098436715e-51	3.7435901880266e-50	106.153268712945
ZNRF2	-	5.03398060524859	-	4.62618589414244e-63	1.72100893556449e-	132.638303245156

	0.589741426792803		19.6378712157685		61	
BARX1	- 0.589562243768253	4.88421668213333	- 21.2630099618577	1.06923658963578e-70	5.12224849261368e-69	150.188199817139
NR2F2	- 0.589454118976748	4.72458590753414	-27.752333322152	6.52552202202413e-101	6.62732016556771e-99	219.653048832082
CALB1	- 0.588922976886088	4.80523044073271	- 28.5663031505095	1.30863647022295e-104	1.52764505650393e-102	228.155176744059
HS.541249	- 0.588036404513056	4.88255666109818	- 27.1919144828587	2.3783885474271e-98	2.36812883212448e-96	213.763109998202
KIAA1244	-0.58711625620121	4.93999511039951	-19.975272312977	1.21013663854317e-64	4.74523077260405e-63	136.274675459649
SYNC1	0.584487869082371	4.84263886830039	23.4231706015903	7.47080906279479e-81	4.65481821114993e-79	173.53409308403
LOC124685	- 0.583632931877199	4.78511668197756	- 33.3938426524457	6.63055666260326e-126	1.5660449643116e-123	277.113244383857
CHAT	- 0.580775086627017	4.84384373254377	- 26.2448559086274	5.39831352173762e-94	4.72631656265235e-92	203.747589289889
LOC654000	- 0.580397216671522	4.93295192978293	- 18.1297642433175	5.0177961081981e-56	1.41951914414652e-54	116.474069235461
DDAH1	- 0.580173988570266	4.75541974138307	- 35.5104681100362	7.28012835145695e-135	2.46456611791323e-132	297.706605682588
DIXDC1	- 0.579714814476873	4.97869806799057	- 16.4021053234925	4.31680798428067e-48	9.02088516221286e-47	98.2513915666843
CCDC7	- 0.576001620289985	4.78277431630881	- 31.7112348143362	1.29598468709876e-118	2.39309463312272e-116	260.352371021357
HS.543777	- 0.574089593694704	4.77676383456453	- 33.3430708379916	1.094949810286008e-125	2.52734324386394e-123	276.612484640612
CCL27	- 0.573490162740964	4.82042807968987	- 29.2479390537102	1.102105621901e-107	1.39912308700333e-105	235.224309201176
FIP1L1	- 0.571103891235191	4.76460527653439	- 35.4811273011091	9.65357370523944e-135	3.1626353080778e-132	297.424948586034
VCAM1	- 0.570318508085819	5.03791584907854	- 13.0193775199802	2.77171619557964e-33	2.91705178054993e-32	64.282615004949
HS.537638	- 0.570218901159193	5.07617601135886	- 13.8714720859924	6.84229354086462e-37	8.76296761677441e-36	72.5519583306341
LOC342900	- 0.569831763804137	4.79774185200097	- 32.0051528812548	6.73708999231204e-117	1.29097898041361e-117	263.30448513423
VCX2	-0.56962628941224	4.78290258538172	- 28.4704770445562	3.55378601914239e-104	3.96618140773738e-102	227.157605571531
ENTHD1	- 0.568653200813346	5.07074379404967	- 13.2040650758914	4.67015681939478e-34	5.07273932168699e-33	66.0550147188215
RNF186	- 0.567778529590276	5.03003924422779	- 14.2641200749897	1.38414132733179e-38	1.92566292060023e-37	76.4367455617529
SDC1	- 0.567368979254755	5.30678061315205	- 9.89898752686208	4.21097863400081e-21	2.32175347485951e-20	36.4152736506914
HS.537839	- 0.566398652132079	5.06150947675469	- 11.9135901986988	9.19353828473352e-29	7.7164937867565e-28	53.9286395695516
HS.544611	- 0.561034386952451	4.9443310712556	- 16.0958381740247	1.05141448004805e-46	2.05349335757077e-45	95.0680317904132
KCNJ12	- 0.560284900343383	4.65645975371635	- 42.6174570619344	3.57698811950882e-163	2.59484938155225e-160	362.745385669891
HS.543222	- 0.560226411156631	4.85355449669172	- 28.5315819464316	1.87923244280001e-104	2.16880507830419e-102	227.793826911641
CEP112	- 0.559043441700698	4.79377694626493	- 28.5137915642399	2.26217816179899e-104	2.58142487766635e-102	227.608632621199
MINA	-0.55513933638943	5.23358715634974	- 10.5587172255964	1.61192110326102e-23	1.01744379892597e-22	41.9350823736004
LDHC	- 0.553412294768246	4.78474933746206	- 33.1207860583095	9.87959347429448e-125	2.22971447388744e-122	274.416407194795
TPSAB1	- 0.552987292321139	5.0442709803319	- 7.44064172556392	4.83997706250085e-13	1.58054041950993e-12	18.0737055487072
MGC42105	- 0.551623503413803	4.75803680271906	- 30.5122838691571	2.48706117849659e-113	3.94665520762677e-111	248.206471476586
C14ORF45	- 0.549823603973253	4.96620757006779	- 19.1341505359563	1.05408771941996e-60	3.50993930440297e-59	127.220782431338
HS.200774	- 0.549202496860498	4.79987452491943	- 26.2141168230272	7.4847302845002e-94	6.49700177516103e-92	203.421283939938
ITGB8	- 0.549136987631867	4.88157343968349	- 20.9651800120464	2.69439116220143e-69	1.23262327222152e-67	146.967124055528
HTR3E	- 0.547715981311474	4.81527598437338	- 29.6038937224921	2.78594886203595e-109	3.72290745300488e-107	238.896664865138
OR11G2	- 0.546264203967803	5.0204511106965	- 16.1019227350807	9.86952324229731e-47	1.93503625576779e-45	95.1311110468761

TRPC2	0.546028514877762	4.95314091197209	15.098539118444	3.04131322544462e-42	5.08856295183123e-41	84.8287513822604
DGKA	-0.54575501792231	5.03096256363813	13.4183323026082	5.83432778643651e-35	6.58371477767213e-34	68.1254795558174
HNT	-	4.70152537973848	-	1.19937708642249e-164	1.01507280747556e-161	366.131647598925
CLCA1	-0.5426789096411	4.79776249034193	19.7674022877975	1.14287013354672e-63	4.37999587784923e-62	134.033699407192
HNRNPA3P1	-	5.60121065816814	-8.8365325456051	2.009059284195e-17	8.76084417788082e-17	28.0288729954633
PF4	0.541393184278695	5.52083496175376	13.2201731273788	3.99617089217223e-34	4.3499583687997e-33	66.210142898826
FSHB	0.540615948792653	5.07830764262221	11.2119615632277	5.28824536366145e-26	3.81715848708925e-25	47.6137974860476
RC3H1	-	5.06587609206214	-	4.14293186853686e-27	3.17791662060878e-26	50.1438325327154
IGF2AS	0.538999453846609	4.88269896687515	17.3750991028444	1.5391802270365e-52	3.86436504034964e-51	108.465134817298
PRDM1	-	5.42169209586836	-	2.40327479758626e-20	1.26727200645307e-19	34.6892217649195
HS.512359	0.536041449000324	4.79326847891877	-28.480080987048	3.21506375458129e-104	3.62802083239195e-102	227.257625944012
DDO	-	4.83619624980517	-	2.21033963684563e-95	2.07853790294484e-93	206.938494689238
FLJ90650	0.532310146089512	4.80345466459548	24.9585661741355	4.97290586670918e-88	3.76901731211182e-86	190.034209935128
IL17RD	0.531169630830586	4.74673504447022	28.3159258388969	1.78354995508331e-103	1.90670877303433e-101	225.546776418112
PDE4DIP	0.528441050783144	4.69432107260916	37.2998561670572	2.99705498776131e-142	1.26825376898766e-139	314.679322046446
CRYGC	-	4.77520951110202	-	5.57765166822439e-120	1.08935827581706e-117	263.493043564478
HS.500311	0.527868129270897	4.82553865381621	26.3336670169649	2.10092873005649e-94	1.92225515157241e-92	204.689930958617
TSN	0.527288955992612	5.30155776399903	13.1935224458358	5.17145295834126e-34	5.61124746206344e-33	65.9535309126456
ZNF14	0.526894347208519	5.06447470265877	13.79685754063	1.42877470961834e-36	1.79143653714615e-35	71.8187823109076
GHRH	0.525056088072312	4.78803117014661	30.3779199320206	9.82096578321106e-113	1.53448813068141e-110	246.835140010101
WDR47	0.524869883009106	4.89467075037197	11.0337100588999	2.56831139189217e-25	1.80136536574979e-24	46.0442462766512
HS.434595	0.524652367453437	4.68210553257931	26.7809154152849	1.8308487345128e-96	1.77086664263924e-94	209.425850076502
HS.563113	0.524520433854971	4.919159491776	16.1560807791426	5.61813375048728e-47	1.11659033992072e-45	95.6928719364144
HS.543860	0.523370207304217	4.84415621120492	26.2629333869901	4.454633385473e-94	3.96853128621612e-92	203.939453035185
CLUL1	0.521238984957665	4.99744293830519	11.9429132115943	7.01786833808525e-29	5.91972349182673e-28	54.1970988074882
HS.580540	0.520998860161695	4.68449710936636	38.3729396193467	1.35302322363349e-146	6.54347802820081e-144	324.664192079758
PCDHB1	-0.52097619200447	4.72195703969269	37.6860458940763	8.04670929922793e-144	3.55314694099821e-141	318.289480949926
HS.545811	0.519882728562694	4.78518661177303	29.8110309912945	3.29758262780759e-110	4.58770536548136e-108	241.027467770343
AMPD2	0.519401626089825	5.06419605853038	11.4721953009418	5.13229434100075e-27	3.91318178132159e-26	49.9310474255641
CDR1	0.517232399089451	4.70302446626763	45.6586371031817	1.82152005254157e-174	2.4459716125048e-171	388.674718878828
LOC284861	0.516205481353294	4.81818938933847	-26.327124997285	2.25212384740234e-94	2.04219373162662e-92	204.620537272027
HS.577366	0.514166080669567	5.07623692223461	11.1398315933854	1.00412764781484e-25	7.18163407831515e-25	46.9769146773324
PRRX2	0.513065597273433	4.80579196723484	22.9459948218647	1.29718101438526e-78	7.52809736119811e-77	168.385248426722
OR5D18	0.512294344362911	4.96353049318013	17.0422701192402	5.19349557764038e-51	1.22095233996564e-49	104.955584713871
HIG2	0.510753134431236	5.14758566566481	10.0790889187928	9.42200636868506e-22	5.37222947240204e-21	37.899595542389
SORD	0.510192552857296	5.08948769384156	10.9079146849523	7.76644283055708e-25	5.29369083135152e-24	44.9454724148413
SLC26A9	-	5.20545653367901	-	5.37400747915119e-20	2.76907254988633e-	33.8919720047593

	0.509546122788712		9.58777829219433		19	
HS.538361	- 0.509070251566288	4.8407884276332	- 21.9235700789652	8.32581656525546e-74	4.33625605316587e-72	157.333817557605
HS.564723	- 0.508535509565506	4.78214213732305	- 32.3220953291507	2.80995104574554e-121	5.59565937658661e-119	266.476426247089
DLC1	- 0.507240200247653	5.2029879908077	- 9.50210679877338	1.07329216292716e-19	5.4176715739007e-19	33.2068033987178
PCDH20	- 0.506567722906122	4.75273190859809	- 28.8991728699055	4.10010732765163e-106	5.01695060477469e-104	231.613253968151
RANBP17	- 0.506248326175292	4.76327864458653	-31.774177768732	6.87436618204373e-119	1.292890054534e-116	260.985423964098
TNK2	- 0.506079145211082	5.05213736605174	-14.24238059603	1.71971277967962e-38	2.37301671065573e-37	76.2205243090534
FGG	0.505364038656396	4.80124806411838	- 26.1056691377872	2.37202565841456e-93	2.0243943350301e-91	202.269490319395
OVCH1	- 0.504964743787645	4.89875193641112	- 14.2618884771231	1.41533920279427e-38	1.9663727692994e-37	76.4145440734682
BFSP2	0.504921461924023	4.90164237461771	18.1680048527927	3.33517608003897e-56	9.46146599689268e-55	116.881578682421
HS.577994	- 0.502864817253809	4.78264485458136	- 28.4088442014155	6.76013965818347e-104	7.4626063444034e-102	226.515512100245
BCL7A	0.501986292658921	5.09629487781983	11.2087120136115	5.44351099471793e-26	3.92644159533773e-25	47.5850540508222
ASB10	0.501851145939424	5.00879784400973	14.607253868652	4.42570906425828e-40	6.67867774986733e-39	79.8663654007006

### Λίστα 3<sup>η</sup> – Χειμώνας 2009, σύγκριση των δύο πόλεων (Πράγα – Οστράβα)

Η λίστα αυτή είναι κενή, καθώς δεν αναγνωρίστηκε κανένα στατιστικά αξιόπιστο, βιολογικώς σημαντικά διαφορικά εκφραζόμενο γονίδιο.

### Πίνακας 4.1.3 – Καλοκαίρι 2009, σύγκριση των δύο πόλεων (Πράγα – Οστράβα)

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
SERPINB2	-0.339042160373903	4.9929981452134	-6.74880804516338	2.66951824901663e-10	2.77362946072828e-06	13.0563234043082
RBM3	-0.302894701004251	5.19421213086112	-5.57535093496763	1.04379673494352e-07	0.00036150160253544	7.45996896550086
RBBP6	-0.214339079337348	4.85061908725166	-5.61994510454515	8.42236581020589e-08	0.00036150160253544	7.66050706701731

### Πίνακας 4.1.4 – 2009 data set, σύγκριση χειμώνα-καλοκαιριού

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
LOC731682	- 0.469966363461541	5.5309092161410 2	- 2.9836689115342 4	0.0030919982376539 9	0.042382403283938	-2.26009199826377
RAPGEF1	- 0.269710205694027	5.0366656117232 2	- 5.6879826056994 1	3.15525370591995e-08	2.73192383370902e-05	8.5006621806797
WVOX	- 0.259423534780625	5.1812548989106 5	- 4.9671930920873 7	1.16595327418428e-06	0.0003274122842912 08	5.06758859051323
USP6	0.176061967317691	5.2275843347398	4.2910892880925 4	2.43040480308104e-05	0.0024049434194297 1	2.20527388280081
HIST1H1E	0.172690709204083	4.6983406091547 1	5.1327662890455 7	5.2594352155115e-07	0.0001884328685833 26	5.82238332649287
AMMECR1	-0.16397695856163	4.864832746116	- 5.1090953164779 7	5.90068469042596e-07	0.0001977681094629 86	5.71321318988129
RBM3	-0.15811873776402	5.1193689030564 5	- 3.9807046676938	8.70331850562552e-05	0.0050237488485249 5	1.01397226225392
NSUN5B	- 0.158080646301609	5.0478030496010 1	- 4.3792707250004	1.66832985103836e-05	0.0017687701175804 6	2.55806868600556



			7			
B4GALT2	0.152473575163038	4.9359734808857 3	4.4305541413196	1.3367422734446e-05	0.0015544987709715 1	2.76612207428939
PGPEP1	- 0.150279259678467	4.9125170437452 1	- 5.1010231338943 7	6.13620807231167e- 07	0.0001988654342935 66	5.6760802186706
EHD3	0.145018503377358	5.1007522570681 1	3.6948042718104	0.0002632877553186 28	0.0098756670677276 1	-0.0124168610608759
LGI2	0.143303887087314	4.8471155293102 8	4.8135899819931 9	2.39636194420003e- 06	0.0005297489489412 4	4.38596958016466
HS.565151	0.142889771137545	4.7764935339236	6.2497053966214 9	1.47580175537267e- 09	5.04441951958026e- 06	11.4301326824996
HS.565035	0.142347602748925	4.7877267093687 7	5.9537820073218 8	7.60501678099421e- 09	1.12880177649328e- 05	9.86018750922421
ZMYM3	- 0.137870092835005	4.9708385742540 8	- 3.9436173936975 9	0.0001008461823323 57	0.0054572491376728 6	0.876951139794443
KIAA1033	0.137135376299596	5.1830846063770 5	3.5405723201578 4	0.0004653981273627 08	0.0141362997593276	-0.537286559756434
DKFZP434L187	0.132473792334592	4.8614838494437 4	5.8879959558894 4	1.08636298311401e- 08	1.41091392431932e- 05	9.51917493947102
ANKRD34B	0.130450076713402	4.9224104952574 3	3.6849158738059	0.0002732402460159 05	0.0100672558727137	-0.0466790554251224
BCL6	- 0.128764909613195	4.9963667623160 5	- 3.2035018638543 3	0.0015099188610172 4	0.0268492046594325	-1.61266012802541
LTK	0.127442005291471	5.0210247648319 8	3.3123047782419 8	0.0010433341723126 5	0.0216372096812942	-1.27639556384415
HNRNPA3P1	0.126435936089238	5.1485187641818	3.4630827017718 3	0.0006150785829044 61	0.0165991337048762	-0.793265958836519
ARMC8	-0.12584509800446	5.0076414489808 1	- 4.0597480330139 5	6.33498879106877e- 05	0.0042740606194288 6	1.3098208137799
HS.147829	0.125143996245987	4.8825568099332 1	4.5030578891527 8	9.73834659277989e- 06	0.0014250904380138 5	3.06385492901765
MGC35440	- 0.124134277914368	5.1497053725503 6	-3.395309025381	0.0007817970694796 42	0.0186732679353873	-1.01288168126582
FBXL7	0.123187253423331	4.8053316508229 1	4.5616944284892 6	7.51531877119378e- 06	0.0011830933641318 7	3.30770121732694
LOC654114	0.122494208101261	4.8417152159483 6	3.6489267548911 3	0.0003125362449269 91	0.0106817486341824	-0.170672728626801
LOC731196	0.121193947830334	4.7280529754403 7	5.2155828904980 4	3.50543369234527e- 07	0.0001466155558221 15	6.20761590751382
TNFRSF10D	0.120881351109293	5.1415282424670 5	3.1855100741505 7	0.0016035691011006 1	0.0278211580818644	-1.6672599636648
TMEM151A	0.120835891466912	4.6804742971999 2	5.1205029997844 7	5.58272073404522e- 07	0.0001933482280890 99	5.76577296963449
HS.538076	0.120801858486448	4.7422053459068 7	5.7387226248448 8	2.41371712267215e- 08	2.50785209045636e- 05	8.75641345423497
LOC648342	- 0.120701525631426	4.9207390727299 6	- 3.7373251025742 5	0.0002242571071762 29	0.0088932494029046 7	0.135861903606119
KIAA1604	0.120172004935685	5.1447303261396	3.2346748327500 9	0.0013595540521397 3	0.0253065701034464	-1.51738267849825
ERN1	- 0.120009989015094	4.9266056551814 5	- 3.4316097669357 1	0.0006878704560925 72	0.0176034691727914	-0.895748339697566
HIG2	0.116860012921728	4.7588011403783 8	4.4263255376114 8	1.36149555494136e- 05	0.0015544987709715 1	2.74888713287308
LOC728518	- 0.116305229506652	4.9496210421460 7	- 3.3370442125869	0.0009579166463521 18	0.0211301420101855	-1.1984853946776
NEDD9	0.116112050031197	4.8225743377454 1	4.1050777307458	5.26825541911861e- 05	0.0039817653697246	1.4818181357212
CXCR5	- 0.115773029148099	4.9185621508088	- 3.1632453003032	0.0017268974477577 6	0.0289862107951585	-1.73443074887869
PSD3	- 0.115574952477361	4.8754022600792 6	3.5897576282131 4	0.0003889068108653 18	0.0124714252002798	-0.372119463335972
OR11H12	-0.11553685916707	4.9465466094687 5	- 3.4430259404254 3	0.0006605833061064 33	0.0171332992103135	-0.858674330731477
TMEM81	0.113809877271371	4.7052426413496	6.4605950468380 4	4.43131951385805e- 10	2.30207048744926e- 06	12.5840146374443

TLX2	- 0.113779805589024	4.6878383828143	- 6.1427425798966 5	2.68713511974251e- 09	5.40696846879662e- 06	10.8559478433462
QPCTL	- 0.113625086777375	4.7675195888586 2	- 4.8753702164566 4	1.79734630932997e- 06	0.0004383279006953 1	4.65794921336618
HS.572887	0.111275219959183	4.8225083966901 7	4.0618856410309 8	6.28037379772188e- 05	0.0042740606194288 6	1.31789353200801
TMC05	0.110566871920229	4.7724901172969 7	3.9864655402984 2	8.50559429025807e- 05	0.0049869373473040 5	1.03535903079671
ERI3	0.107771418238172	4.8931059014510 8	3.4305329985275 9	0.0006904978028407 24	0.0176034691727914	-0.899239298760386
HGF	0.10714621985264	4.6860238976762 3	5.4327038745355 7	1.18201982085194e- 07	7.22422702273627e- 05	7.24154937399052
MINK1	0.106651706135796	4.7363327221313 8	5.1450291224134 5	4.95433860814645e- 07	0.0001838413504951 49	5.87910380566434
LOC646562	0.106247972820226	4.7920141268820 3	3.8672346740061 2	0.0001361159687826 46	0.0063419054513528 7	0.598375065013987
RUNX2	0.106189495702741	4.9446609700654 5	3.3393176418351 6	0.0009504029575591 61	0.0210774410051829	-1.19129895831498
HS.153349	0.105996691836658	4.8801239857558 4	3.2825105365462 9	0.0011555972245876 8	0.0230626800068892	-1.36951224673796
PIKFYVE	- 0.105227197257665	4.7754834100026 5	- 5.1645995375997 8	4.50264627367275e- 07	0.0001799326722440 76	5.96985685469262
RAB13	- 0.104766923303409	5.0031236305764	- 3.5820922523318	0.0003999952931856 06	0.0127875418344567	-0.397997150540884
SREK1	-0.10441738914137	4.7182545909485 7	- 4.7731404577188 2	2.88847612068041e- 06	0.0006252347269556 13	4.20948954616471
HS.527387	0.104000949225616	4.9066950678206 3	2.9904505035721 5	0.0030262067919047 7	0.041811553946663	-2.24076357226151
CYP4X1	0.103250720520925	4.7632322974834 1	3.8524832593045 3	0.0001441547390630 53	0.0065120336472396 5	0.545139204964252
RNF38	0.103165400917564	4.7492739690773 1	3.4501678608772 2	0.0006440305932604 19	0.0169842619292309	-0.835423390686852
HS.544379	- 0.102192093708661	4.9265683087821	- 4.2155218804781 8	3.33887960247902e- 05	0.0030559250197565 5	1.9079594735358
SLC27A4	0.102027957023107	4.7139095031816 3	3.5398176996321 6	0.0004666747658757 83	0.0141362997593276	-0.53980438995182
COL9A1	0.101163092223989	4.7615848852844 2	4.4278027874484 7	1.35279853444189e- 05	0.0015544987709715 1	2.75490648001394
SEC14L4	0.101056914555455	4.6859478282135 5	5.3175861102050 8	2.11235020907864e- 07	9.97605394196687e- 05	6.68906471574757
UBE2G2	- 0.100584031694971	4.6784686757264 8	- 5.7520964054499 1	2.24849700336621e- 08	2.50785209045636e- 05	8.82412215375955

**Πίνακας 4.1.5 – Πράγα 2009, σύγκριση χειμώνα-καλοκαιριού [|lcf| >= 0.1]**

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
HPCAL1	-0.281997534843765	5.16302384229305	-3.67589245794739	0.000367920136233317	0.0723655805638416	-0.00572649345666409
RBM3	-0.274098147678455	5.25689177787836	-3.84968021538276	0.000198824189234934	0.0558319817878639	0.546872503968841
BCL6	-0.254334284767625	5.06404597496027	-3.64323131886982	0.00041215619489165	0.0723655805638416	-0.107397256559405
ARMC8	-0.232649443229547	5.04691015626177	-4.665693951417	8.70506961047284e-06	0.0129208104646875	3.38403466898881
RBPJ	-0.186849525412905	4.89788126744373	-4.06187492912127	9.14481223188938e-05	0.0395894162872211	1.24732566332387
MRPL11	0.174087332295672	4.82318025355244	3.66190359313449	0.000386285717600878	0.0723655805638416	-0.0493580719846118
HS.572887	0.160580916134611	4.79917352226864	3.57296500396193	0.000524970043080848	0.0823281134663294	-0.323742049012164
ZNF460	-0.155739276535258	4.83090730196059	-4.20121676509603	5.41203224148918e-05	0.0295952710468803	1.72208673419402
SEPT4	0.149151959588803	4.7653054117119	4.48085480204899	1.82680503618812e-05	0.0210894492511051	2.70853173776009
LOC731196	0.143014558503431	4.71747949701999	4.26901402676643	4.17594478867961e-05	0.0271175414714882	1.95718847696527

HS.580783	0.142415421891616	4.75511470236568	4.33572118673864	3.22742174501812e-05	0.0257945476390295	2.19107570253459
HS.122051	0.141976767169688	4.76505854277806	4.24513494723549	4.57662988327354e-05	0.0274625270282114	1.8740804875787
HS.575268	0.140958122978113	4.69132745005291	5.15506514935986	1.12535081653925e-06	0.00292309874596071	5.25536989108568
MSR1	-0.138173945408279	4.69245598225207	-4.03972721470987	9.92953606125613e-05	0.0412671518705805	1.172929198857
HS.582893	0.138068842185771	4.74694019690129	3.76556576319801	0.00026844975010405	0.063576824790381	0.276989921435963
HS.565035	0.13477011011687	4.78924900261288	3.43104913432149	0.000847408806354167	0.0957019293263021	-0.750585673114202
PIKFYVE	-0.133979072878578	4.76307731284097	-4.01719766547326	0.000107936826174604	0.0431332163059284	1.097553990689
LRR73	0.127820035550877	4.64683140183842	5.40363728854124	3.81411090204961e-07	0.00204490446605542	6.2481188978905
CGB5	0.127357633707479	4.73542161322927	3.95395095894073	0.000136207865348464	0.0463196906973333	0.887606201115155
LOC130576	-0.125725756879103	4.74928016699111	-3.49406558385957	0.000686213786152374	0.0902501422547236	-0.562734798944622
TMEM81	0.123624106794761	4.72044040175307	4.27407896440332	4.09540351711312e-05	0.0271175414714882	1.97485830531801
HS.538779	0.122706128531321	4.70606330667646	4.44091787754154	2.13897221156029e-05	0.0222329212781114	2.56498432393099
ARPP-21	0.122217839049583	4.6596732702028	3.89614119283339	0.000168124627178798	0.0499089964682203	0.697859823087419
NOX5	0.119534658237243	4.65753571602572	5.39647983391106	3.93629348615094e-07	0.00204490446605542	6.21916422368111
SEC14L4	0.1171957565727	4.69366831869009	3.48424058380437	0.000709285391307145	0.090550359316136	-0.592201109432462
HS.538285	0.115174597472153	4.67671928210112	4.234996494069	4.75770439372287e-05	0.0274625270282114	1.83889401238355
HS.364166	0.114971810553903	4.81736558560348	3.97125258741977	0.00012784028761252	0.0463196906973333	0.944795821210339
PLCXD2	0.114943944534237	4.59438506484259	4.68572555730637	8.02448857181865e-06	0.0129208104646875	3.45831671965072
MYH16	0.113411347402858	4.78833580993919	3.7152980604829	0.000320531545092602	0.07085793092579	0.117867191840559
NEU4	-0.112884778885862	4.69763271828905	-3.65008325377013	0.000402477910927119	0.0723655805638416	-0.0861259459612906
HIST1H1E	0.112282832325043	4.61841270124366	3.63923962440166	0.00041789555667281	0.0723655805638416	-0.119774922107756
ZC2HC1B	0.110177830277061	4.60269106251928	5.15866921643083	1.10805290387065e-06	0.00292309874596071	5.26957049866417
HS.542360	0.110072398869509	4.63648685341638	3.94888186130755	0.000138756859716699	0.0463196906973333	0.870885443206022
FAM69B	0.109871363444606	4.7490628438342	3.42759878110186	0.000857190325442086	0.0957656718424008	-0.760792637906858
CNOT6L	-0.107724629299541	4.72935270143294	-3.45704998225081	0.000777002875293474	0.0927937916586115	-0.673406323194847
DAZAP1	-0.107327236837048	4.67162766172366	-3.56835374821798	0.000533312231557765	0.0823281134663294	-0.33782496548644
HS.574298	-0.106994555482419	4.65550570063214	-3.82640984537631	0.00021613928295854	0.0590970302615588	0.471759591723156
UBE2G2	-0.105652333645798	4.68327082849254	-3.80759330444029	0.000231180325624947	0.0615888098267488	0.411273587802599
TLX2	-0.103034134280126	4.71831777563361	-3.46918743853587	0.000746055700599054	0.0927910418120264	-0.637219878559121
HS.128236	-0.103021396221312	4.68248386905092	-3.68634511501431	0.00035474058407902	0.0722696993839416	0.0269591292856779
HS.564048	0.101103190493283	4.67684428064588	3.48200261578399	0.00071464191183091	0.090550359316136	-0.598903861259182
NUF2	-0.100815908911384	4.63000879528177	-3.93007259316573	0.00014861887679748	0.0467924281795701	0.808980374805315

Πίνακας 4.1.6 – Οστράβα 2009, σύγκριση χειμώνα-καλοκαιριού

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
RAPGEF1	0.294833187703539	5.02846072607911	5.12781699482329	7.38359026315258e-07	0.000920863194128474	5.61163277620608
WWOX	0.2933938737815532	5.22436628930749	4.19333974186297	4.26645303343746e-05	0.00809766192940888	1.82920280141182
AMMECR1	0.212489710082555	4.89105161493117	4.88201023893155	2.26926700807565e-06	0.0014843415528598	4.55983251854222
HIST1H1E	0.209421100837114	4.74522569480086	4.19217240924868	4.28653903866688e-05	0.00809766192940888	1.82485783755629
B4GALT2	0.192231275323316	4.96112898925492	4.4466741594244	1.50522475749432e-05	0.00579232786309853	2.79533506047772
NSUN5B	0.173365576797825	5.06693067550617	3.62534416561246	0.000373380208646743	0.0287364471691827	0.164741131503257
LTK	0.173166988478894	5.00505686587313	3.64202142315883	0.000351527479680601	0.028496323753313	0.109711277114495
KIAA1604	0.171545116038242	5.13050474842427	3.72577528191488	0.000258854932220183	0.0233638065284436	0.169930263357503
USP6	0.169644566481548	5.1996475589292	3.29945209936086	0.00116307613730122	0.049427024744943	-1.19581893109696

LOC648342	0.166255894487896	4.94101796523955	4.03165262337885	8.10208155867531e-05	0.0126578383128149	1.23689866619185
EHD3	0.163328435787898	5.08162126072461	3.49728256315602	0.000589174770829311	0.0362220465616363	0.580012386629104
ANKRD34B	0.160721237837567	4.90888672326086	3.53070954765291	0.000523668498091353	0.0337945074234109	0.472867288597552
SERPINB2	0.159942792437182	4.93882770946452	3.87827276008869	0.000146308454189462	0.0187672202349198	0.692997572115688
HS.147829	0.154619954562675	4.86976883489806	4.43454067110668	1.58384258666008e-05	0.00586790768220187	2.74802152302457
LGI2	0.153793640990855	4.86172119368268	4.05834585951381	7.29747343219433e-05	0.012429630977131	1.33335398841375
HS.565151	0.153260099095918	4.75523639182346	5.43437233590753	1.72729662967821e-07	0.000406109308947942	6.97653824537383
PGPEP1	0.151077746032186	4.8977253933541	4.21691973991192	3.87948022390252e-05	0.00791694617342121	1.9171827701954
NEDD9	0.150383716216115	4.80555925701725	4.21283094895596	3.9440830810023e-05	0.00791694617342121	1.90189802585981
HS.565035	0.14695501963479	4.78683374405797	4.92816183573147	1.8433341366817e-06	0.0014843415528598	4.75433778182399
HS.153349	0.143500758232398	4.89492784772452	3.4686016791895	0.000651433865507851	0.0381085591098822	0.671236310248171
TLE6	0.141566493812519	4.9532540612599	3.60001008491255	0.00040903980409179	0.0301413018759837	0.247918801851757
ZNF121	0.140874247623032	4.7251317666297	3.91696759892899	0.000126245480223899	0.0170349420717703	0.828542965826077
PSD3	0.138619431659677	4.85272061458194	3.54679010912016	0.000494653276568337	0.032945176561186	0.421007918070546
DKFZP434L187	0.138092866614181	4.87022046900743	4.90605020743671	2.03672729006591e-06	0.0014843415528598	4.6609748523588
TMEM151A	0.135444013787596	4.68992584625746	4.35548205057073	2.20142203404142e-05	0.00676775730819439	2.44228215871488
HS.538076	0.134202696322947	4.74469013084274	4.80304201280324	3.22832765135668e-06	0.00186346246097755	4.23027016720926
PDGFA	0.125877769857113	4.89272264537366	3.79693198163254	0.000198775893702631	0.0215207444340542	0.411779290091111
HS.544379	0.124782962275904	4.93634996811098	4.04389972528484	7.72294994098973e-05	0.0125664850123388	1.28108754671885
C9ORF152	0.124350561209624	4.70588889223711	4.99894774167413	1.33654475427939e-06	0.00138866999969629	5.05535647509191
QPCTL	0.123994424160988	4.75583303345493	4.30160678231722	2.74834154211084e-05	0.00700583432678479	2.23647728107938
HGF	0.123808096111372	4.6897586881825	4.84833287907021	2.63869376846118e-06	0.00161270754437128	4.41878140355625
SYNPO	0.12175352165906	4.80069091705124	3.97856099362722	9.96048946763442e-05	0.0143735396623224	1.04663117557481
ZWINT	0.121646830616121	4.81503404983081	3.51387047296039	0.000555761669032015	0.0349552967371765	0.526953241870412
FBXL7	0.120911479881132	4.80484435921337	3.55912078422432	0.000473435327015354	0.0327932869845968	0.381103289907021
TLX2	0.120313601112512	4.6699594093728	5.11116528944378	7.97667829370189e-07	0.000920863194128474	5.53916054502692
CYP4X1	0.116573504450413	4.7612223998348	3.7669086711627	0.00022307856833657	0.0220494855685323	0.309259719073096
SREK1	0.116524419053009	4.71959212760768	4.35483098745148	2.20735962623112e-05	0.00676775730819439	2.43978272402471
HIG2	0.115244416646519	4.75094124897257	3.61000642736567	0.000394602245812366	0.0299264038977408	-0.21515864435032
ZFYVE27	0.115187243191872	4.8401228208805	3.54991965110871	0.000489185052995629	0.032945176561186	0.410891448155824
PCDH12	0.114819002625973	4.87534135442974	3.31237802660167	0.00111354548292498	0.0485895988466579	-1.15654751088786
MINK1	0.114634097449251	4.73969870154871	4.31302306535232	2.62254774523337e-05	0.00700583432678479	2.27991456782277
HS.583103	0.114289002865708	4.72625176779336	4.07418894741235	6.8565747431505e-05	0.0121485016950264	1.3908525212371
TMCO5	0.113571055249843	4.78088807776125	3.32890546496758	0.00105307705186162	0.0475490306634879	-1.10613636459777
LPAL2	0.112860318151681	4.71849314721733	3.64047492115943	0.000353502025189513	0.028496323753313	0.114823415569135
MLLT10	0.111735158022871	4.67340456607935	3.9208327439869	0.000124391758206564	0.0170056627337658	0.842144530902659
UBE2D3	0.110912411386238	4.69625181971815	4.32029124267258	2.54535738764742e-05	0.00700583432678479	2.30761752531046
SH3PXD2B	0.109992009100671	4.73314142450533	3.99399787732863	9.38199759173591e-05	0.0137294302786107	1.10173608839856
LOC731196	0.107926148710365	4.73425529518418	3.48277856900368	0.000619926156652382	0.0376296784781645	0.626226460606324
CCDC144A	0.107845633378681	4.69317284227557	3.72369433825995	0.000260847118123143	0.0233638065284436	0.162916308743034
TMEM81	0.10784243539798	4.69632775396209	4.88039308371525	2.28580027389381e-06	0.0014843415528598	4.55304232995274
ANKFN1	0.107282204053247	4.6130075625896	5.28477541647306	3.53455022536204e-07	0.00061206628069186	6.30327056159864
NGFR	0.106825631756881	4.69480667878183	3.46191129468999	0.000666818918175417	0.0381085591098822	0.692421753581711

CCDC148	0.105411102334627	4.75405596094179	3.29520869596294	0.00117977883895533	0.049427024744943	-1.20868157743857
HS.539440	-0.1020416230545	4.68402720008809	- 4.75709489669798	3.95586826449819e-06	0.00205507356340681	4.04042074157472
COL9A1	0.101910925688214	4.76786178970633	3.53665085864667	0.000512766765379987	0.033719282862646	- 0.453730566452829
DUSP5P	0.101456146617979	4.67640356324863	4.22759181443922	3.71559720358788e-05	0.00789119356706303	1.95713431123139
LZTS1	0.10124493075821	4.71070684796156	3.7520422244337	0.000234913600620683	0.0222220102663039	0.258752526180685
LSM11	-0.10108075630623	4.72390854483711	- 4.29416741731966	2.83340526301291e-05	0.00700930492445337	2.20822185478916
PAX3	0.101014213696955	4.66519628861093	4.36941104710665	2.0780027963662e-05	0.00676775730819439	2.49582783499113
LOC654109	0.100452359170599	4.72231991556301	3.34101722335777	0.00101072419447957	0.0468813588421552	-1.0690530735563

**Πίνακας 4.1.7 – 2010 data set, σύγκριση με βάση την πόλη (Πράγα – Οστράβα)**

TargetID	logFC	AveExpr	t	P.Value	adj.P.Val	B
LOC642093	-0.514675410151857	5.35437032536959	-5.48849393349231	1.33843071224854e-07	0.00108118432935437	7.21960068560869

## 4.2 – Αποτελέσματα της σύγκρισης των λιστών με τη Gene Ontology

Λόγω χαμηλών αριθμών γονιδίων, οι λίστες για το Χειμώνα 2009 ανά πόλη, το καλοκαίρι 2009 ανά πόλη και το Χειμώνα 2010 ανά πόλη δεν έδωσαν κανένα μετρήσιμο αποτέλεσμα στις οντολογίες που χρησιμοποιήθηκαν. Ως εκ τούτου, θα παρουσιαστούν τα αποτελέσματα που εξήχθησαν από τις λίστες του 2009 ανά εποχή, της Πράγας 2009 ανά εποχή και της Οστράβας 2009 ανά εποχή.

Πίνακας 4.2.1

Gene Ontology data 2009, ανά εποχή	Hub genes 2009, ανά εποχή																																																																																																																												
<p><b>BioInfoMiner</b></p> <p><b>Subject:</b> Enrichment Analysis Report (extended version)  <b>Job tag:</b> 2009 data set - by season - Gene Ontology  <b>Database:</b> Gene Ontology Biological Process (GO_P)  <b>Hypergeometric p-value threshold:</b> 0.1  <b>Corrected p-value threshold:</b> 0.1</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to Gene Ontology Biological Process, BioInfoMiner analyzed 35 genes and reveals 14 ontological terms as statistically significant. This enriched set of terms corresponds to 16 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.</p> <p><b>Table 1: Statistically significant terms ranking according to corrected p-value</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Term id</th> <th>Definition</th> <th>Enrichment</th> <th>Hyp/metric pvalue</th> <th>Corrected pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>GO:0032456</td> <td>endocytic recycling</td> <td>2/23</td> <td>1.783E-3</td> <td>0.0076</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RAB13 EHD3</td> <td>RAB13, member RAS oncogene family EH domain containing 3</td> <td>-0.11 0.15</td> <td>0.013 9.88E-3</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Rank</th> <th>Term id</th> <th>Definition</th> <th>Enrichment</th> <th>Hyp/metric pvalue</th> <th>Corrected pvalue</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>GO:0048705</td> <td>skeletal system morphogenesis</td> <td>2/43</td> <td>6.142E-3</td> <td>0.0125</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2 WWOX</td> <td>runt related transcription factor 2 WW domain containing oxidoreductase</td> <td>0.12 -0.27</td> <td>0.022 3.27E-4</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Rank</th> <th>Term id</th> <th>Definition</th> <th>Enrichment</th> <th>Hyp/metric pvalue</th> <th>Corrected pvalue</th> </tr> </thead> <tbody> <tr> <td>3</td> <td>GO:0006511</td> <td>ubiquitin-dependent protein catabolic process</td> <td>3/169</td> <td>1.119E-2</td> <td>0.0197</td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue	1	GO:0032456	endocytic recycling	2/23	1.783E-3	0.0076	Gene Symbol	Definition	Fold Change	Pvalue	RAB13 EHD3	RAB13, member RAS oncogene family EH domain containing 3	-0.11 0.15	0.013 9.88E-3	Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue	2	GO:0048705	skeletal system morphogenesis	2/43	6.142E-3	0.0125	Gene Symbol	Definition	Fold Change	Pvalue	RUNX2 WWOX	runt related transcription factor 2 WW domain containing oxidoreductase	0.12 -0.27	0.022 3.27E-4	Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue	3	GO:0006511	ubiquitin-dependent protein catabolic process	3/169	1.119E-2	0.0197	<p><b>BioInfoMiner</b></p> <p><b>Subject:</b> Genes Prioritization (extended version)  <b>Job tag:</b> 2009 data set - by season - Gene Ontology  <b>Database:</b> Gene Ontology Biological Process (GO_P)</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to Gene Ontology Biological Process, BioInfoMiner reveals 7 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.</p> <p><b>Table 1: Hub genes ranking according to ontological clusters amount</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Gene Symbol</th> <th>Definition</th> <th>Clusters</th> <th>Interactors</th> <th>Drugs</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>HGF</td> <td>hepatocyte growth factor</td> <td>4</td> <td>0</td> <td>6</td> <td>0.11</td> <td>7.22E-5</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>GO:0007067</td> <td>mitotic nuclear division</td> </tr> <tr> <td>GO:0045669</td> <td>positive regulation of osteoblast differentiation</td> </tr> <tr> <td>GO:0000902</td> <td>cell morphogenesis</td> </tr> <tr> <td>GO:0010976</td> <td>positive regulation of neuron projection development</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Rank</th> <th>Gene Symbol</th> <th>Definition</th> <th>Clusters</th> <th>Interactors</th> <th>Drugs</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>BCL6</td> <td>B-cell CLL/lymphoma 6</td> <td>3</td> <td>1</td> <td>0</td> <td>-0.13</td> <td>0.027</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>GO:0007067</td> <td>mitotic nuclear division</td> </tr> <tr> <td>GO:0000902</td> <td>cell morphogenesis</td> </tr> <tr> <td>GO:0010976</td> <td>positive regulation of neuron projection development</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Rank</th> <th>Gene Symbol</th> <th>Definition</th> <th>Clusters</th> <th>Interactors</th> <th>Drugs</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>3</td> <td>ZMYM3</td> <td>zinc finger MYM-type containing 3</td> <td>2</td> <td>0</td> <td>0</td> <td>-0.14</td> <td>5.46E-3</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>GO:0007067</td> <td>mitotic nuclear division</td> </tr> <tr> <td>GO:0010976</td> <td>positive regulation of neuron projection development</td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue	1	HGF	hepatocyte growth factor	4	0	6	0.11	7.22E-5	Term id	Definition	GO:0007067	mitotic nuclear division	GO:0045669	positive regulation of osteoblast differentiation	GO:0000902	cell morphogenesis	GO:0010976	positive regulation of neuron projection development	Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue	2	BCL6	B-cell CLL/lymphoma 6	3	1	0	-0.13	0.027	Term id	Definition	GO:0007067	mitotic nuclear division	GO:0000902	cell morphogenesis	GO:0010976	positive regulation of neuron projection development	Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue	3	ZMYM3	zinc finger MYM-type containing 3	2	0	0	-0.14	5.46E-3	Term id	Definition	GO:0007067	mitotic nuclear division	GO:0010976	positive regulation of neuron projection development
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																																																																																																																								
1	GO:0032456	endocytic recycling	2/23	1.783E-3	0.0076																																																																																																																								
Gene Symbol	Definition	Fold Change	Pvalue																																																																																																																										
RAB13 EHD3	RAB13, member RAS oncogene family EH domain containing 3	-0.11 0.15	0.013 9.88E-3																																																																																																																										
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																																																																																																																								
2	GO:0048705	skeletal system morphogenesis	2/43	6.142E-3	0.0125																																																																																																																								
Gene Symbol	Definition	Fold Change	Pvalue																																																																																																																										
RUNX2 WWOX	runt related transcription factor 2 WW domain containing oxidoreductase	0.12 -0.27	0.022 3.27E-4																																																																																																																										
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																																																																																																																								
3	GO:0006511	ubiquitin-dependent protein catabolic process	3/169	1.119E-2	0.0197																																																																																																																								
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue																																																																																																																						
1	HGF	hepatocyte growth factor	4	0	6	0.11	7.22E-5																																																																																																																						
Term id	Definition																																																																																																																												
GO:0007067	mitotic nuclear division																																																																																																																												
GO:0045669	positive regulation of osteoblast differentiation																																																																																																																												
GO:0000902	cell morphogenesis																																																																																																																												
GO:0010976	positive regulation of neuron projection development																																																																																																																												
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue																																																																																																																						
2	BCL6	B-cell CLL/lymphoma 6	3	1	0	-0.13	0.027																																																																																																																						
Term id	Definition																																																																																																																												
GO:0007067	mitotic nuclear division																																																																																																																												
GO:0000902	cell morphogenesis																																																																																																																												
GO:0010976	positive regulation of neuron projection development																																																																																																																												
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue																																																																																																																						
3	ZMYM3	zinc finger MYM-type containing 3	2	0	0	-0.14	5.46E-3																																																																																																																						
Term id	Definition																																																																																																																												
GO:0007067	mitotic nuclear division																																																																																																																												
GO:0010976	positive regulation of neuron projection development																																																																																																																												

## Πράγα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Prague 2009 - by season - Gene Ontology

**Database:** Gene Ontology Biological Process (GO\_P)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to Gene Ontology Biological Process, BioInfoMiner analyzed 18 genes and reveals 10 ontological terms as statistically significant. This enriched set of terms corresponds to 6 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue												
1	GO:0045596	negative regulation of cell differentiation	2/55	3.473E-3	0.0095												
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RBPJ</td> <td>recombination signal binding protein for immunoglobulin kappa J region</td> <td>-0.20</td> <td>0.040</td> </tr> <tr> <td>BCL6</td> <td>B-cell CLL/lymphoma 6</td> <td>-0.25</td> <td>0.073</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040	BCL6	B-cell CLL/lymphoma 6	-0.25	0.073
Gene Symbol	Definition	Fold Change	Pvalue														
RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040														
BCL6	B-cell CLL/lymphoma 6	-0.25	0.073														
2	GO:0030183	B cell differentiation	2/68	5.256E-3	0.0215												
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RBPJ</td> <td>recombination signal binding protein for immunoglobulin kappa J region</td> <td>-0.20</td> <td>0.040</td> </tr> <tr> <td>BCL6</td> <td>B-cell CLL/lymphoma 6</td> <td>-0.25</td> <td>0.073</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040	BCL6	B-cell CLL/lymphoma 6	-0.25	0.073
Gene Symbol	Definition	Fold Change	Pvalue														
RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040														
BCL6	B-cell CLL/lymphoma 6	-0.25	0.073														
3	GO:0042953	lipoprotein transport	1/13	2.029E-2	0.0339												
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> </table>						Gene Symbol	Definition	Fold Change	Pvalue								
Gene Symbol	Definition	Fold Change	Pvalue														

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

## Πράγα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Prague 2009 - by season - Gene Ontology

**Database:** Gene Ontology Biological Process (GO\_P)

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to Gene Ontology Biological Process, BioInfoMiner reveals 3 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue								
1	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	3	0	0	-0.19	0.040								
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>GO:0045596</td> <td>negative regulation of cell differentiation</td> </tr> <tr> <td>GO:0048844</td> <td>artery morphogenesis</td> </tr> <tr> <td>GO:0030183</td> <td>B cell differentiation</td> </tr> </tbody> </table>								Term id	Definition	GO:0045596	negative regulation of cell differentiation	GO:0048844	artery morphogenesis	GO:0030183	B cell differentiation
Term id	Definition														
GO:0045596	negative regulation of cell differentiation														
GO:0048844	artery morphogenesis														
GO:0030183	B cell differentiation														
2	BCL6	B-cell CLL/lymphoma 6	2	1	0	-0.26	0.073								
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>GO:0045596</td> <td>negative regulation of cell differentiation</td> </tr> <tr> <td>GO:0030183</td> <td>B cell differentiation</td> </tr> </tbody> </table>								Term id	Definition	GO:0045596	negative regulation of cell differentiation	GO:0030183	B cell differentiation		
Term id	Definition														
GO:0045596	negative regulation of cell differentiation														
GO:0030183	B cell differentiation														
3	MSR1	macrophage scavenger receptor 1	2	0	0	-0.14	0.042								
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>GO:0044872</td> <td>lipoprotein localization</td> </tr> <tr> <td>GO:0030301</td> <td>cholesterol transport</td> </tr> </tbody> </table>								Term id	Definition	GO:0044872	lipoprotein localization	GO:0030301	cholesterol transport		
Term id	Definition														
GO:0044872	lipoprotein localization														
GO:0030301	cholesterol transport														

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

BioInfoMiner					
<b>Subject:</b> Enrichment Analysis Report (extended version)					
<b>Job tag:</b> Ostrava 2009 - by season - Gene Ontology					
<b>Database:</b> Gene Ontology Biological Process (GO_P)					
<b>Hypergeometric p-value threshold:</b> 0.1					
<b>Corrected p-value threshold:</b> 0.1					
<b>Short Description:</b> Input genes list contained 43 genes. Excluding genes without annotation or mapping to Gene Ontology Biological Process, BioInfoMiner analyzed 34 genes and reveals 15 ontological terms as statistically significant. This enriched set of terms corresponds to 14 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.					
<b>Table 1: Statistically significant terms ranking according to corrected p-value</b>					
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue
1	GO:0048011	neurotrophin TRK receptor signaling pathway	2/15	5.500E-4	0.0082
Gene Symbol	Definition	Fold Change	Pvalue		
ZFYVE27	zinc finger FYVE-type containing 27	0.12	0.033		
NGFR	nerve growth factor receptor	0.12	0.039		
2	GO:0048008	platelet-derived growth factor receptor signaling pathway	2/31	2.377E-3	0.0118
Gene Symbol	Definition	Fold Change	Pvalue		
PDGFA	platelet derived growth factor subunit A	-0.14	0.022		
RAPGEF1	Rap guanine nucleotide exchange factor 1	-0.29	9.21E-4		
3	GO:0009887	organ morphogenesis	3/123	3.029E-3	0.0195
e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1					

BioInfoMiner							
<b>Subject:</b> Genes Prioritization (extended version)							
<b>Job tag:</b> Ostrava 2009 - by season - Gene Ontology							
<b>Database:</b> Gene Ontology Biological Process (GO_P)							
<b>Short Description:</b> Input genes list contained 43 genes. Excluding genes without annotation or mapping to Gene Ontology Biological Process, BioInfoMiner reveals 8 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.							
<b>Table 1: Hub genes ranking according to ontological clusters amount</b>							
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue
1	PDGFA	platelet derived growth factor subunit A	5	0	0	-0.13	0.022
Term id	Definition						
GO:0048011	neurotrophin TRK receptor signaling pathway						
GO:0002576	platelet degranulation						
GO:0007067	mitotic nuclear division						
GO:0042060	wound healing						
GO:0010976	positive regulation of neuron projection development						
2	HGF	hepatocyte growth factor	5	0	6	0.13	1.61E-3
Term id	Definition						
GO:0048011	neurotrophin TRK receptor signaling pathway						
GO:0043154	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process						
GO:0002576	platelet degranulation						
GO:0007067	mitotic nuclear division						
GO:0010976	positive regulation of neuron projection development						
3	NGFR	nerve growth factor receptor	3	0	2	0.11	0.039
Term id	Definition						
e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1							

Και σε περιληπτική μορφή πινάκων:

**Πίνακας 4.2.2 – 2009, ανά εποχή – Gene Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	GO:0032456	endocytic recycling	2/23	1.7830e-3	7.6000e-3
2	GO:0048705	skeletal system morphogenesis	2/43	6.1420e-3	0.0125
3	GO:0006511	ubiquitin-dependent protein catabolic process	3/169	0.0112	0.0197
4	GO:0045669	positive regulation of osteoblast differentiation	2/60	0.0117	0.0236
5	GO:0016197	endosomal transport	2/62	0.0124	0.0297
6	GO:0000902	cell morphogenesis	2/63	0.0128	0.0389
7	GO:0010976	positive regulation of neuron projection development	2/89	0.0245	0.0477
8	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	2/91	0.0256	0.0533
9	GO:0007067	mitotic nuclear division	3/241	0.0283	0.0549
10	GO:0001649	osteoblast differentiation	2/101	0.0310	0.0657
11	GO:0007010	cytoskeleton organization	2/112	0.0374	0.0732
12	GO:0006357	regulation of transcription from RNA polymerase II promoter	4/470	0.0399	0.0839



13	GO:0043066	negative regulation of apoptotic process	4/475	0.0413	0.0861
14	GO:0044257	cellular protein catabolic process	1/18	0.0477	0.0947

**Πίνακας 4.2.3 – Πράγα 2009, ανά εποχή – Gene Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	GO:0045596	negative regulation of cell differentiation	2/55	3.4730e-3	9.5000e-3
2	GO:0030183	B cell differentiation	2/68	5.2560e-3	0.0215
3	GO:0042953	lipoprotein transport	1/13	0.0203	0.0339
4	GO:0044257	cellular protein catabolic process	1/18	0.0280	0.0404
5	GO:0030301	cholesterol transport	1/21	0.0326	0.0519
6	GO:0048844	artery morphogenesis	1/23	0.0356	0.0642
7	GO:0006357	regulation of transcription from RNA polymerase II promoter	3/470	0.0388	0.0751
8	GO:0042981	regulation of apoptotic process	2/197	0.0390	0.0829
9	GO:0050767	regulation of neurogenesis	1/26	0.0402	0.0884
10	GO:0042127	regulation of cell proliferation	2/203	0.0412	0.0990

**Πίνακας 4.2.4 – Οστράβα 2009, ανά εποχή – Gene Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	GO:0048011	neurotrophin TRK receptor signaling pathway	2/15	5.5000e-4	8.2000e-3
2	GO:0048008	platelet-derived growth factor receptor signaling pathway	2/31	2.3770e-3	0.0118
3	GO:0009887	organ morphogenesis	3/123	3.0290e-3	0.0195
4	GO:0006511	ubiquitin-dependent protein catabolic process	3/169	7.3170e-3	0.0256
5	GO:0014068	positive regulation of phosphatidylinositol 3-kinase signaling	2/64	9.8000e-3	0.0340
6	GO:0043154	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	2/67	0.0107	0.0370
7	GO:1903358	regulation of Golgi organization	1/7	0.0161	0.0445
8	GO:0042060	wound healing	2/85	0.0168	0.0498
9	GO:0043410	positive regulation of MAPK cascade	2/88	0.0179	0.0584
10	GO:0010976	positive regulation of neuron projection development	2/89	0.0183	0.0636
11	GO:0007067	mitotic nuclear division	3/241	0.0189	0.0744

12	GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	2/91	0.0191	0.0808
13	GO:0006464	cellular protein modification process	2/93	0.0199	0.0869
14	GO:0002576	platelet degranulation	2/101	0.0232	0.0921
15	GO:0043066	negative regulation of apoptotic process	4/475	0.0253	0.0969

**Πίνακας 4.2.5 – 2009, ανά εποχή – Gene Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	HGF	hepatocyte growth factor	4	4	0	6	0.1071	7.2242e-5
2	BCL6	B-cell CLL/lymphoma 6	3	3	1	0	-0.1288	0.0268
3	ZMYM3	zinc finger MYM-type containing 3	2	2	0	0	-0.1379	5.4572e-3
4	RUNX2	runt related transcription factor 2	2	2	0	0	0.1062	0.0211
5	UBE2G2	ubiquitin conjugating enzyme E2 G2	2	2	0	0	-0.1006	2.5079e-5
6	WWOX	WW domain containing oxidoreductase	2	2	0	1	-0.2594	3.2741e-4
7	RAB13	RAB13, member RAS oncogene family	2	2	0	0	-0.1048	0.0128

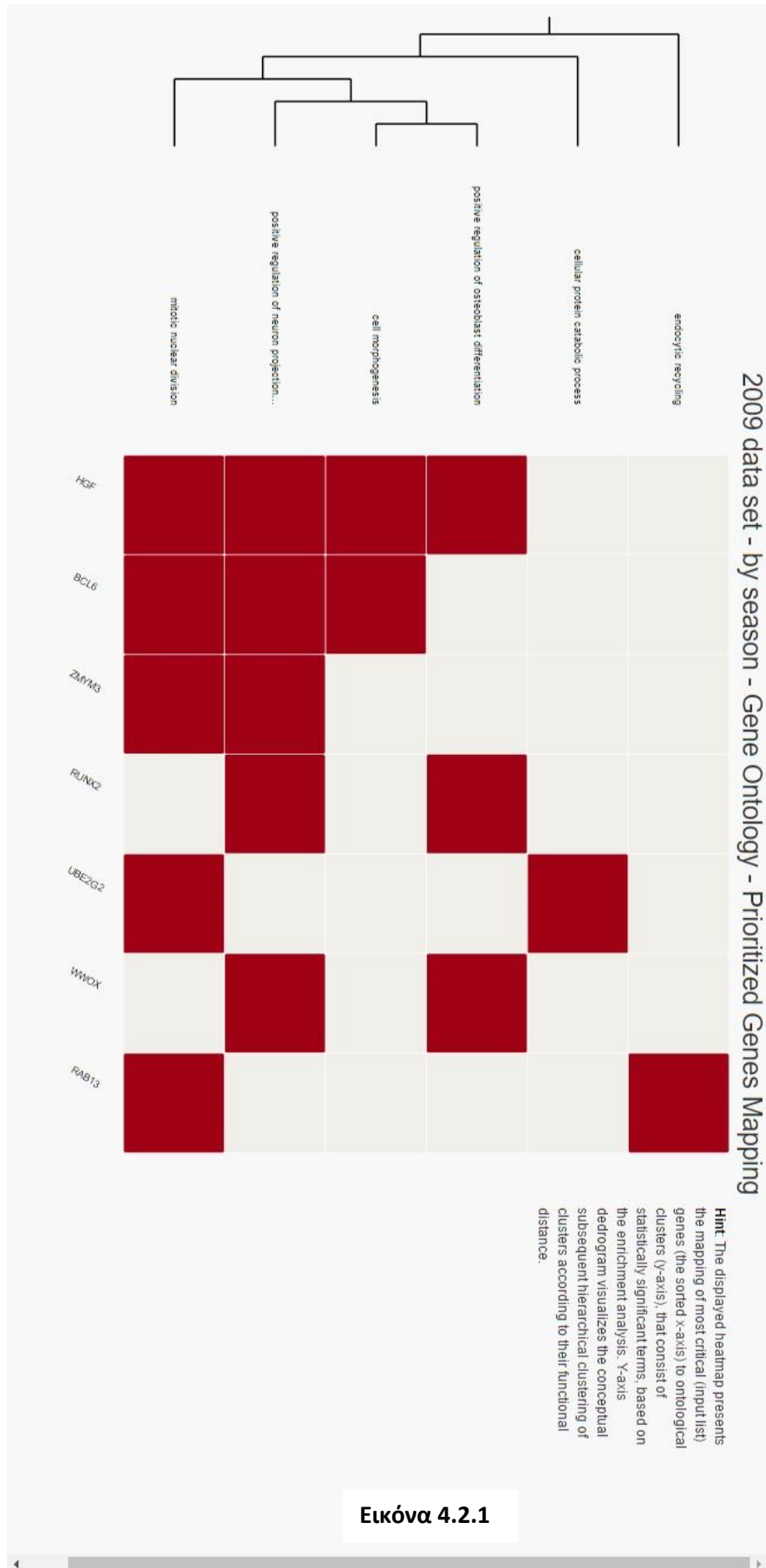
**Πίνακας 4.2.6 – Πράγα 2009, ανά εποχή – Gene Ontology hub genes**

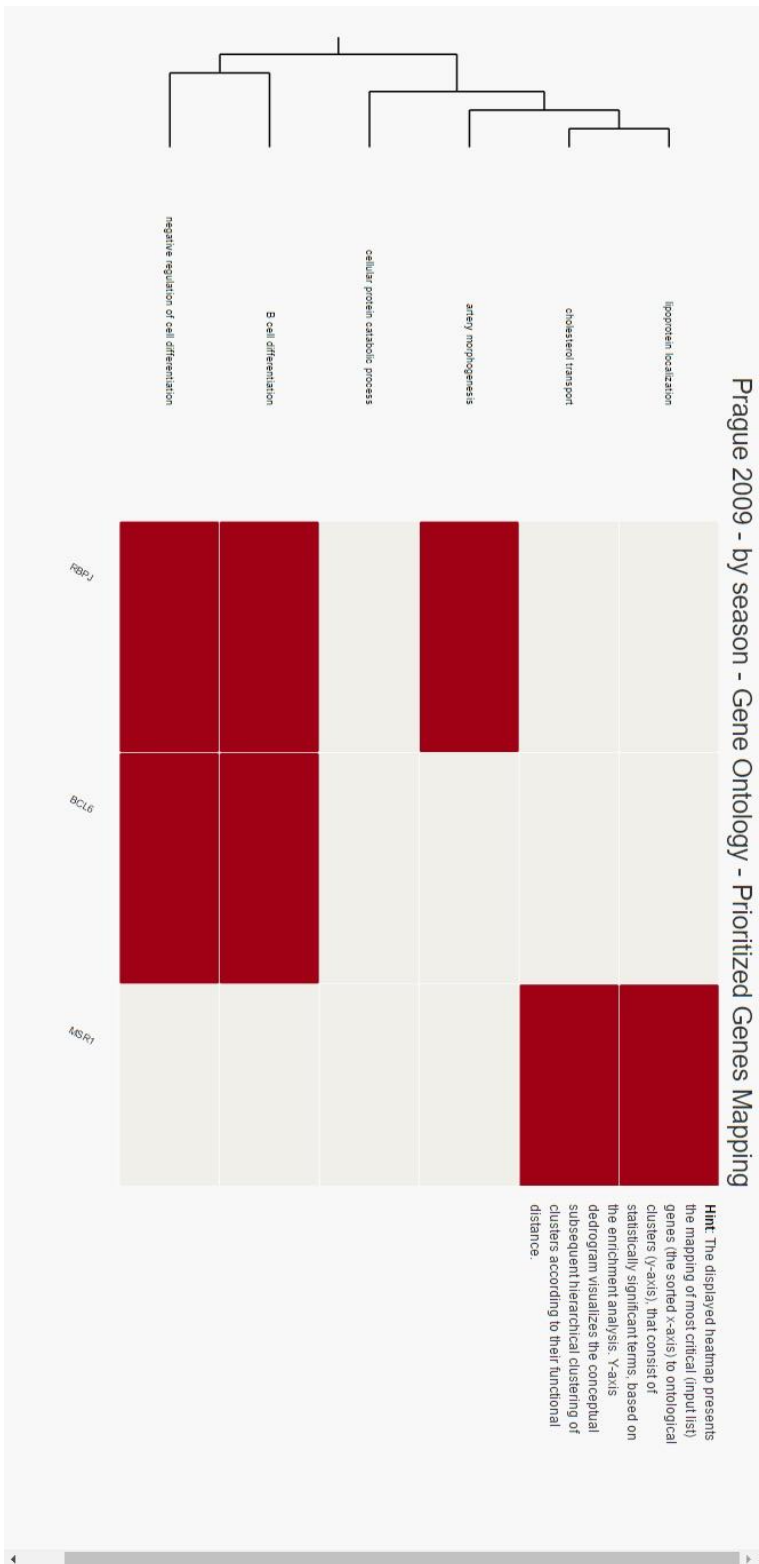
Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	3	3	0	0	-0.1868	0.0396
2	BCL6	B-cell CLL/lymphoma 6	2	2	1	0	-0.2543	0.0724
3	MSR1	macrophage scavenger receptor 1	2	1	0	0	-0.1382	0.0413

**Πίνακας 4.2.7 – Οστράβα 2009, ανά εποχή – Gene Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	PDGFA	platelet derived growth factor subunit A	5	5	0	0	-0.1259	0.0215
2	HGF	hepatocyte growth factor	5	5	0	6	0.1238	1.6127e-3
3	NGFR	nerve growth factor receptor	3	3	0	2	0.1068	0.0381
4	ZFYVE27	zinc finger FYVE-type containing 27	2	2	0	0	0.1152	0.0329
5	MINK1	misshapen like kinase 1	2	2	0	0	0.1146	7.0058e-3
6	LTK	leukocyte receptor tyrosine kinase	2	2	0	0	0.1732	0.0285
7	SERPINB2	serpin family B member 2	2	2	0	2	0.1599	0.0188
8	RAPGEF1	Rap guanine nucleotide exchange factor 1	2	2	1	0	-0.2948	9.2086e-4

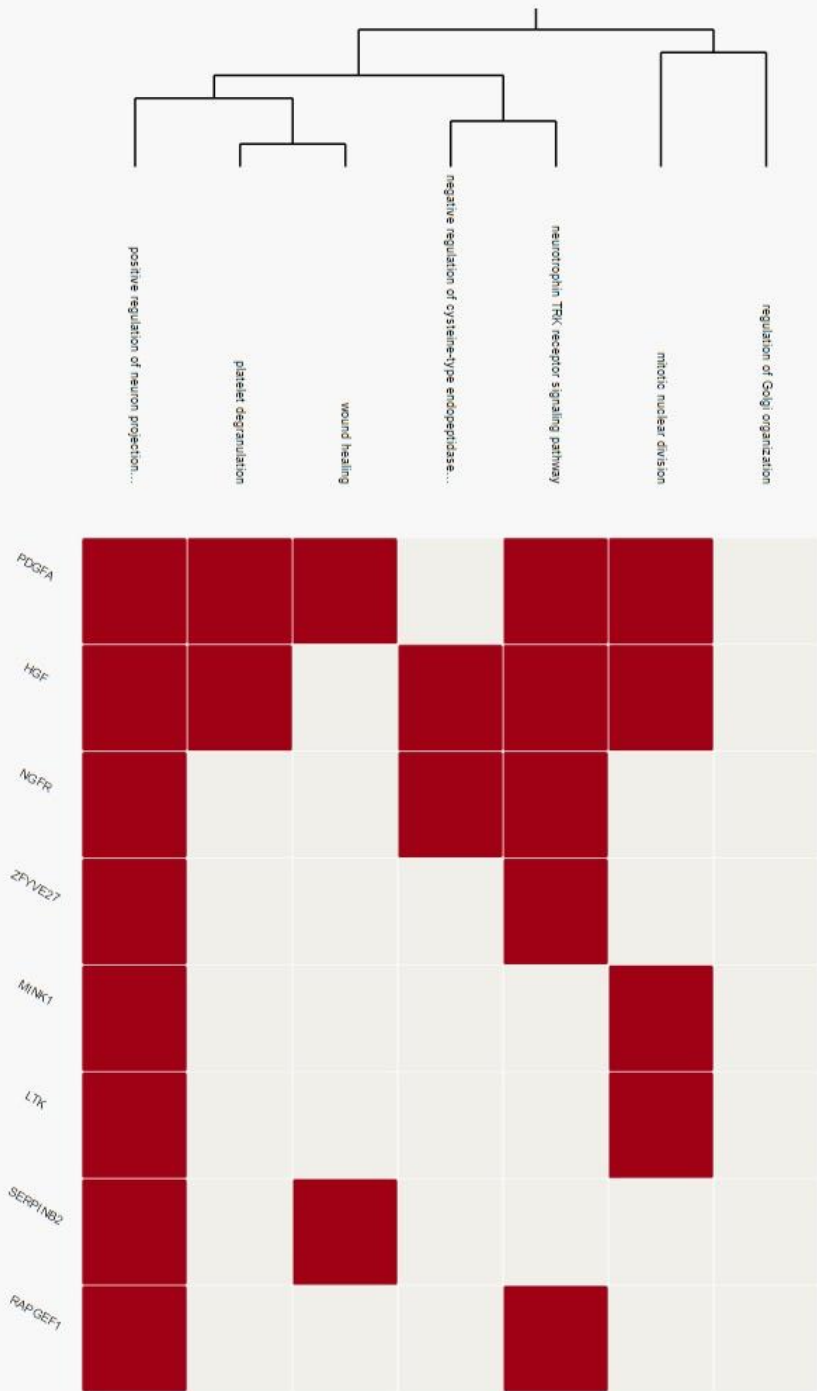
## Παρουσίαση των hub genes με τη χρήση heatmaps





Εικόνα 4.2.2

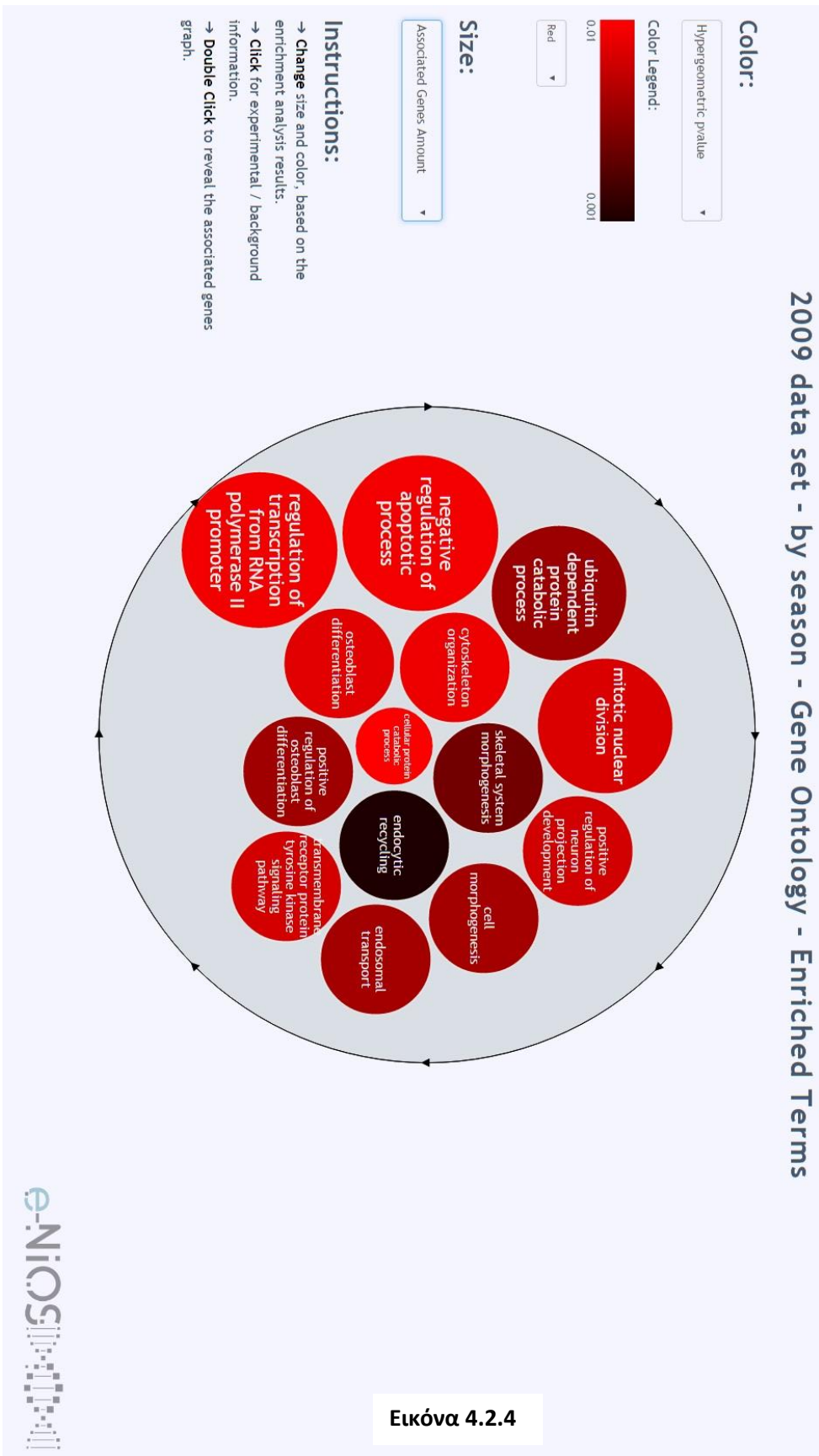
## Ostrava 2009 - by season - Gene Ontology - Prioritized Genes Mapping



**Hint:** The displayed heatmap presents the mapping of most critical (input list) genes (the sorted x-axis) to ontological clusters (y-axis), that consist of statistically significant terms, based on the enrichment analysis. Y-axis dendrogram visualizes the conceptual subsequent hierarchical clustering of clusters according to their functional distance.

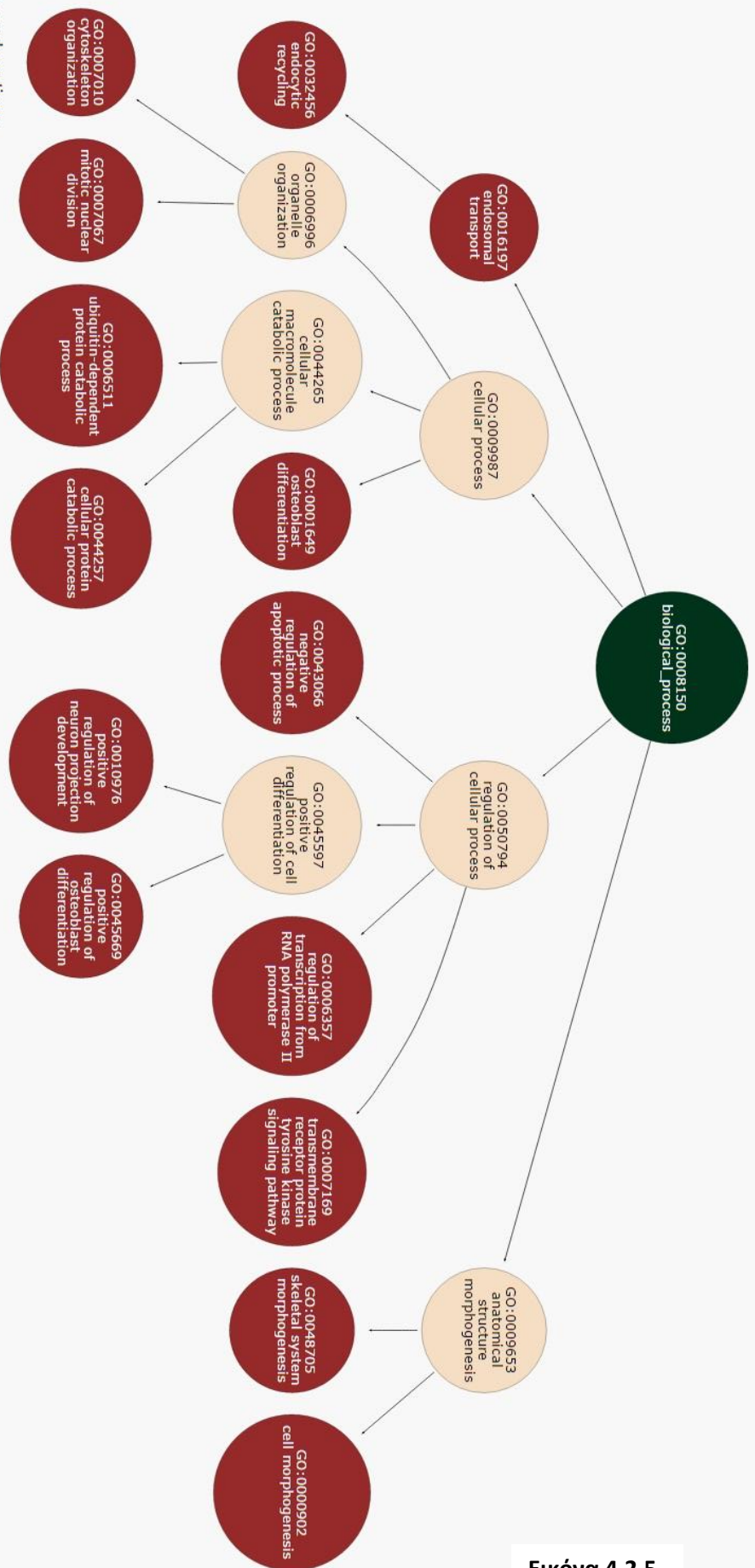
Εικόνα 4.2.3

Ακολουθούν διαγράμματα για τις βιολογικές εργασίες που αναγνωρίστηκαν ως διαφορετικώς εκφραζόμενες ανά εποχή για το 2009



Εικόνα 4.2.4

# Ontological Tree of Significant Terms

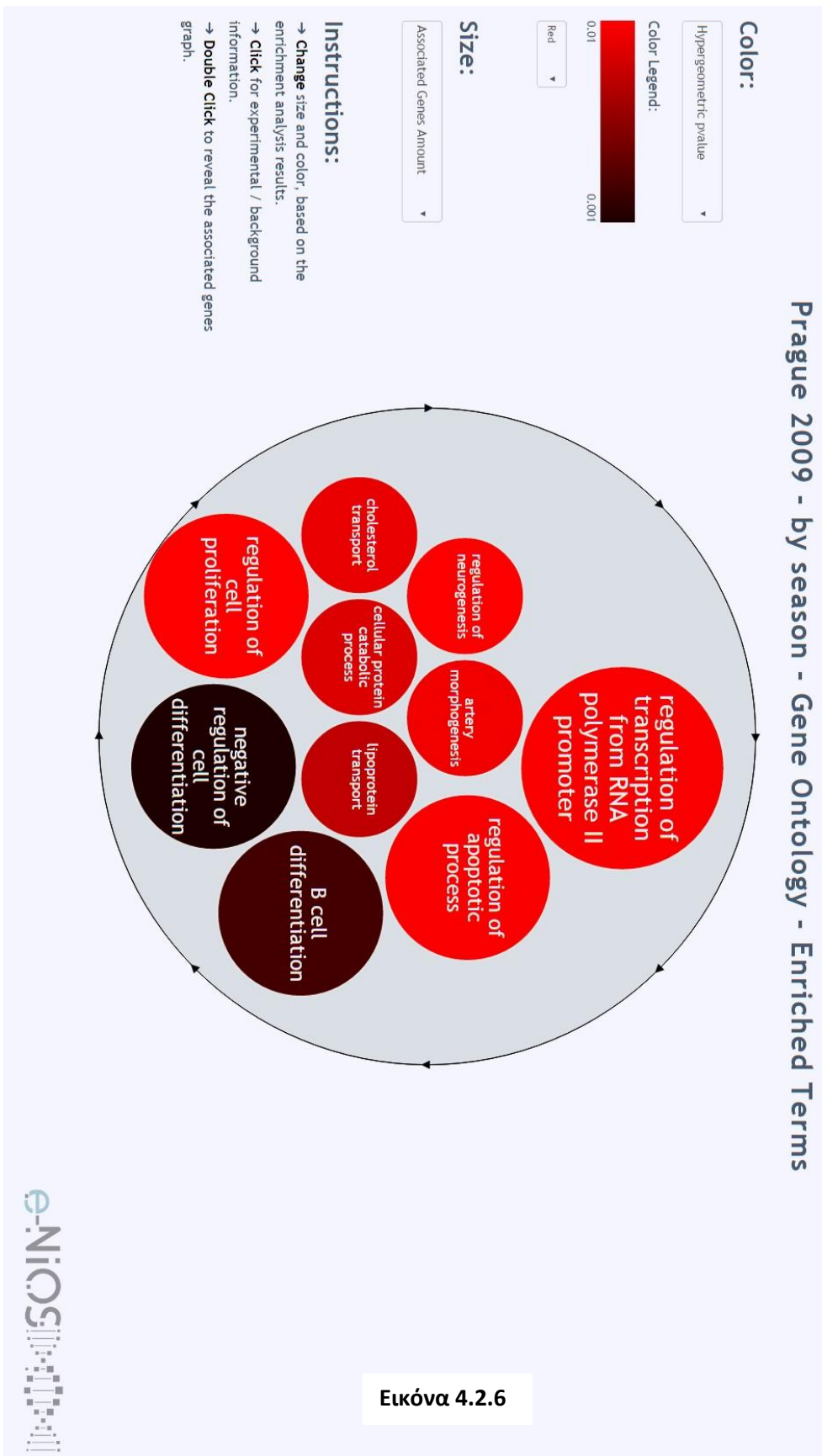


Graph explanation:  
 Significant terms are presented with ● color.  
 Linkage depicts the hierarchical clustering of terms, based on Resnik semantic similarity metric.

Εικόνα 4.2.5

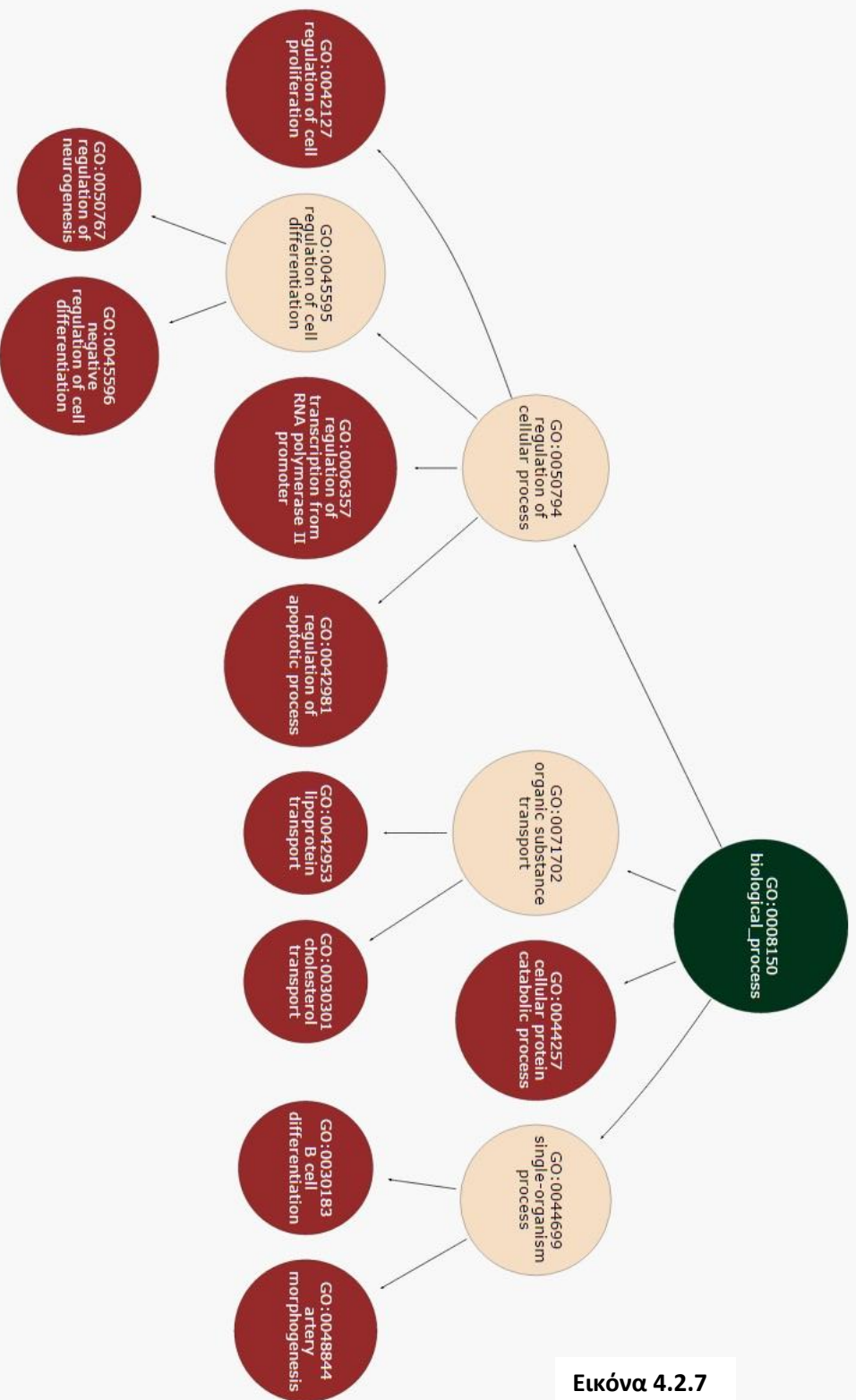


Ακολουθούν διαγράμματα για τις βιολογικές εργασίες που αναγνωρίστηκαν ως διαφορεκώς εκφραζόμενες ανά εποχή στην Πράγα για το 2009



Εικόνα 4.2.6

# Ontological Tree of Significant Terms

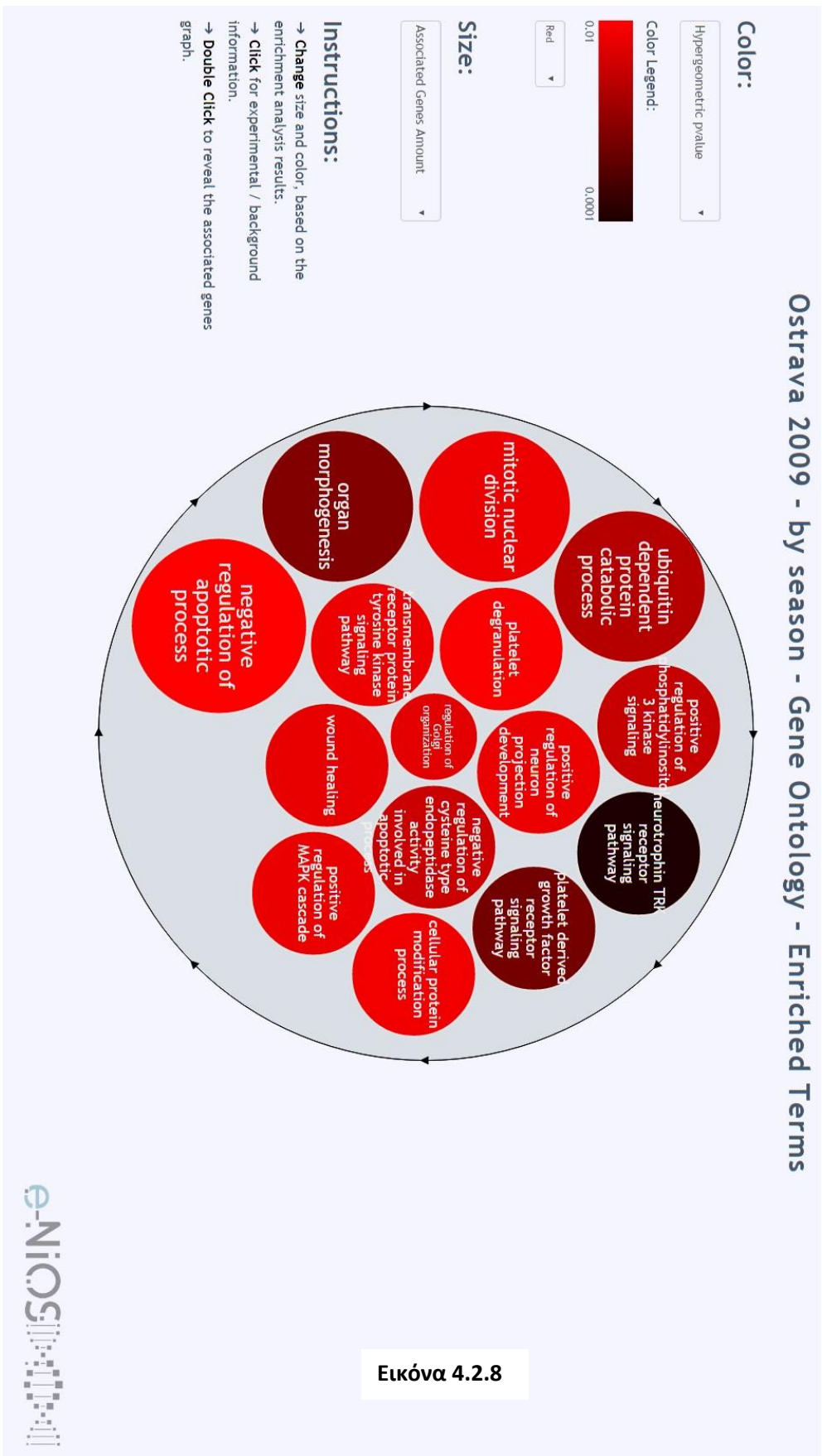


Εικόνα 4.2.7

### Graph explanation:

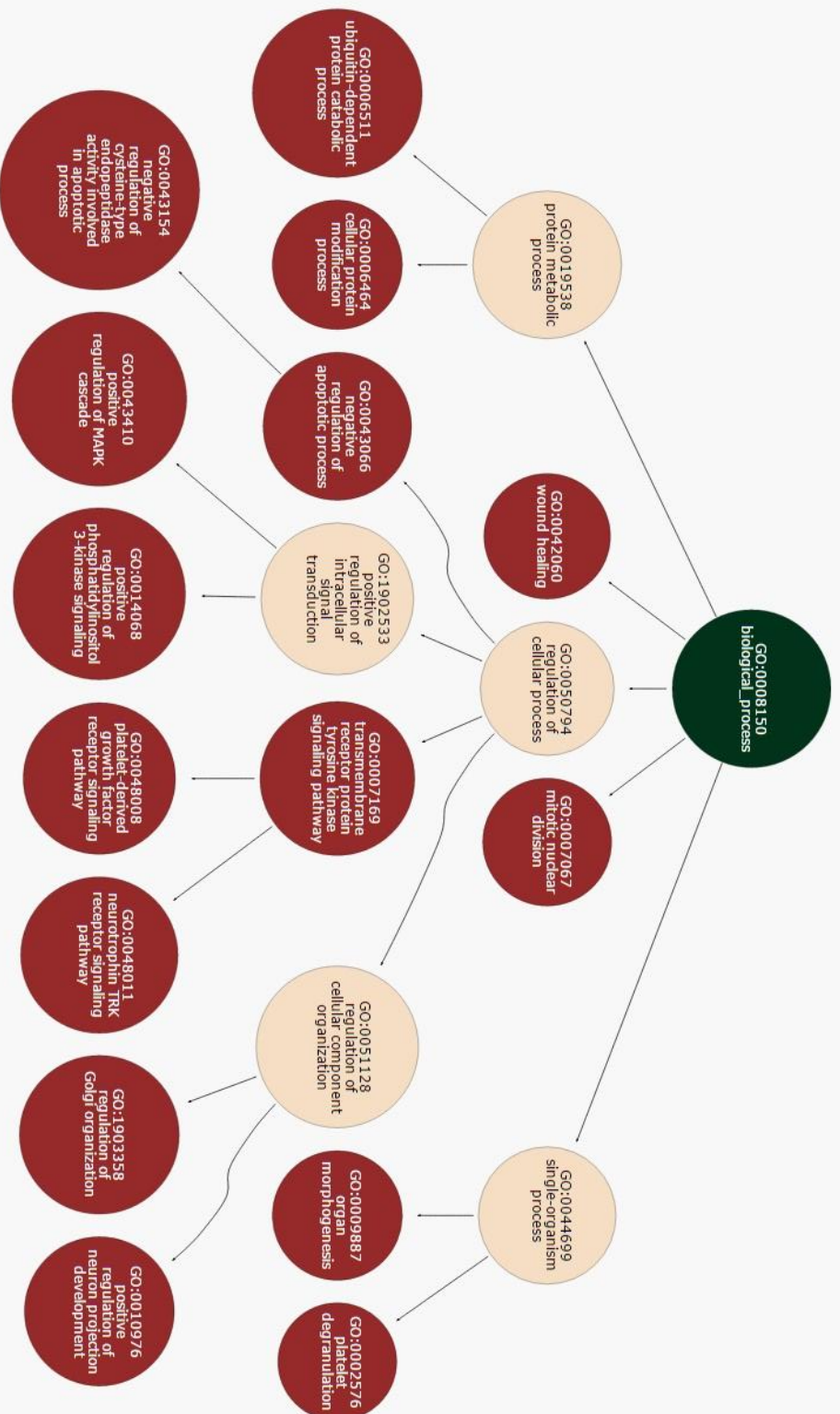
Significant terms are presented with ● color.  
Linkage depicts the hierarchical clustering of terms, based on Resnik semantic similarity metric.

Ακολουθούν διαγράμματα για τις βιολογικές εργασίες που αναγνωρίστηκαν ως διαφορετικώς εκφραζόμενες ανά εποχή στην Οστράβα για το 2009



Εικόνα 4.2.8

# Ontological Tree of Significant Terms



## Graph explanation:

Significant terms are presented with ● color.

Linkage depicts the hierarchical clustering of terms, based on Resnik semantic similarity metric.

Εικόνα 4.2.9

## 4.3 – Αποτελέσματα της σύγκρισης των λιστών με την Human Phenotype Ontology

Πίνακας 4.3.1

Human Phenotype Ontology data	Hub genes																																																																																																																																																		
2009, ανά εποχή	2009, ανά εποχή																																																																																																																																																		
<h3>BioInfoMiner</h3> <p><b>Subject:</b> Enrichment Analysis Report (extended version)  <b>Job tag:</b> 2009 data set - by season - Human Phenotype  <b>Database:</b> Human Phenotype Ontology (HPO)  <b>Hypergeometric p-value threshold:</b> 0.1  <b>Corrected p-value threshold:</b> 0.1</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to Human Phenotype Ontology, BioInfoMiner analyzed 9 genes and reveals 22 ontological terms as statistically significant. This enriched set of terms corresponds to 6 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.</p> <p><b>Table 1: Statistically significant terms ranking according to corrected p-value</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Term id</th> <th>Definition</th> <th>Enrichment</th> <th>Hyp/metric pvalue</th> <th>Corrected pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>HP:0001182</td> <td>Tapered finger</td> <td>3/80</td> <td>1.507E-3</td> <td>0.0043</td> </tr> <tr> <td colspan="6"> <table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td>WWOX</td> <td>WW domain containing oxidoreductase</td> <td>-0.27</td> <td>3.27E-4</td> </tr> <tr> <td>AMMECR1</td> <td>Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1</td> <td>-0.17</td> <td>1.98E-4</td> </tr> </tbody> </table> </td> </tr> <tr> <td>2</td> <td>HP:0011069</td> <td>Increased number of teeth</td> <td>2/25</td> <td>2.228E-3</td> <td>0.0087</td> </tr> <tr> <td colspan="6"> <table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td>AMMECR1</td> <td>Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1</td> <td>-0.17</td> <td>1.98E-4</td> </tr> </tbody> </table> </td> </tr> <tr> <td>3</td> <td>HP:0000233</td> <td>Thin vermilion border</td> <td>3/139</td> <td>7.117E-3</td> <td>0.0126</td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue	1	HP:0001182	Tapered finger	3/80	1.507E-3	0.0043	<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td>WWOX</td> <td>WW domain containing oxidoreductase</td> <td>-0.27</td> <td>3.27E-4</td> </tr> <tr> <td>AMMECR1</td> <td>Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1</td> <td>-0.17</td> <td>1.98E-4</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RUNX2	runx related transcription factor 2	0.12	0.022	WWOX	WW domain containing oxidoreductase	-0.27	3.27E-4	AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	-0.17	1.98E-4	2	HP:0011069	Increased number of teeth	2/25	2.228E-3	0.0087	<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td>AMMECR1</td> <td>Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1</td> <td>-0.17</td> <td>1.98E-4</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RUNX2	runx related transcription factor 2	0.12	0.022	AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	-0.17	1.98E-4	3	HP:0000233	Thin vermilion border	3/139	7.117E-3	0.0126	<h3>BioInfoMiner</h3> <p><b>Subject:</b> Genes Prioritization (extended version)  <b>Job tag:</b> 2009 data set - by season - Human Phenotype  <b>Database:</b> Human Phenotype Ontology (HPO)</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to Human Phenotype Ontology, BioInfoMiner reveals 5 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.</p> <p><b>Table 1: Hub genes ranking according to ontological clusters amount</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Gene Symbol</th> <th>Definition</th> <th>Clusters</th> <th>Interactors</th> <th>Drugs</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>5</td> <td>0</td> <td>0</td> <td>0.11</td> <td>0.022</td> </tr> <tr> <td colspan="8"> <table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>HP:0000322</td> <td>Short philtrum</td> </tr> <tr> <td>HP:0001151</td> <td>Impaired horizontal smooth pursuit</td> </tr> <tr> <td>HP:0008936</td> <td>Muscular hypotonia of the trunk</td> </tr> <tr> <td>HP:0002866</td> <td>Hypoplastic iliac wing</td> </tr> <tr> <td>HP:0011069</td> <td>Increased number of teeth</td> </tr> </tbody> </table> </td> </tr> <tr> <td>2</td> <td>COL9A1</td> <td>collagen type IX alpha 1</td> <td>5</td> <td>0</td> <td>0</td> <td>0.11</td> <td>1.55E-3</td> </tr> <tr> <td colspan="8"> <table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>HP:0001151</td> <td>Impaired horizontal smooth pursuit</td> </tr> <tr> <td>HP:0005041</td> <td>Irregular capital femoral epiphysis</td> </tr> <tr> <td>HP:0003236</td> <td>Elevated serum creatine phosphokinase</td> </tr> <tr> <td>HP:0002866</td> <td>Hypoplastic iliac wing</td> </tr> <tr> <td>HP:0000322</td> <td>Short philtrum</td> </tr> </tbody> </table> </td> </tr> <tr> <td>3</td> <td>WWOX</td> <td>WW domain containing oxidoreductase</td> <td>4</td> <td>0</td> <td>1</td> <td>-0.26</td> <td>3.27E-4</td> </tr> <tr> <td colspan="8"> <table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> </tbody> </table> </td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue	1	RUNX2	runx related transcription factor 2	5	0	0	0.11	0.022	<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>HP:0000322</td> <td>Short philtrum</td> </tr> <tr> <td>HP:0001151</td> <td>Impaired horizontal smooth pursuit</td> </tr> <tr> <td>HP:0008936</td> <td>Muscular hypotonia of the trunk</td> </tr> <tr> <td>HP:0002866</td> <td>Hypoplastic iliac wing</td> </tr> <tr> <td>HP:0011069</td> <td>Increased number of teeth</td> </tr> </tbody> </table>								Term id	Definition	HP:0000322	Short philtrum	HP:0001151	Impaired horizontal smooth pursuit	HP:0008936	Muscular hypotonia of the trunk	HP:0002866	Hypoplastic iliac wing	HP:0011069	Increased number of teeth	2	COL9A1	collagen type IX alpha 1	5	0	0	0.11	1.55E-3	<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>HP:0001151</td> <td>Impaired horizontal smooth pursuit</td> </tr> <tr> <td>HP:0005041</td> <td>Irregular capital femoral epiphysis</td> </tr> <tr> <td>HP:0003236</td> <td>Elevated serum creatine phosphokinase</td> </tr> <tr> <td>HP:0002866</td> <td>Hypoplastic iliac wing</td> </tr> <tr> <td>HP:0000322</td> <td>Short philtrum</td> </tr> </tbody> </table>								Term id	Definition	HP:0001151	Impaired horizontal smooth pursuit	HP:0005041	Irregular capital femoral epiphysis	HP:0003236	Elevated serum creatine phosphokinase	HP:0002866	Hypoplastic iliac wing	HP:0000322	Short philtrum	3	WWOX	WW domain containing oxidoreductase	4	0	1	-0.26	3.27E-4	<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> </tbody> </table>								Term id	Definition
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																																																																																																																																														
1	HP:0001182	Tapered finger	3/80	1.507E-3	0.0043																																																																																																																																														
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td>WWOX</td> <td>WW domain containing oxidoreductase</td> <td>-0.27</td> <td>3.27E-4</td> </tr> <tr> <td>AMMECR1</td> <td>Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1</td> <td>-0.17</td> <td>1.98E-4</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RUNX2	runx related transcription factor 2	0.12	0.022	WWOX	WW domain containing oxidoreductase	-0.27	3.27E-4	AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	-0.17	1.98E-4																																																																																																																														
Gene Symbol	Definition	Fold Change	Pvalue																																																																																																																																																
RUNX2	runx related transcription factor 2	0.12	0.022																																																																																																																																																
WWOX	WW domain containing oxidoreductase	-0.27	3.27E-4																																																																																																																																																
AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	-0.17	1.98E-4																																																																																																																																																
2	HP:0011069	Increased number of teeth	2/25	2.228E-3	0.0087																																																																																																																																														
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RUNX2</td> <td>runx related transcription factor 2</td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td>AMMECR1</td> <td>Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1</td> <td>-0.17</td> <td>1.98E-4</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RUNX2	runx related transcription factor 2	0.12	0.022	AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	-0.17	1.98E-4																																																																																																																																		
Gene Symbol	Definition	Fold Change	Pvalue																																																																																																																																																
RUNX2	runx related transcription factor 2	0.12	0.022																																																																																																																																																
AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	-0.17	1.98E-4																																																																																																																																																
3	HP:0000233	Thin vermilion border	3/139	7.117E-3	0.0126																																																																																																																																														
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue																																																																																																																																												
1	RUNX2	runx related transcription factor 2	5	0	0	0.11	0.022																																																																																																																																												
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>HP:0000322</td> <td>Short philtrum</td> </tr> <tr> <td>HP:0001151</td> <td>Impaired horizontal smooth pursuit</td> </tr> <tr> <td>HP:0008936</td> <td>Muscular hypotonia of the trunk</td> </tr> <tr> <td>HP:0002866</td> <td>Hypoplastic iliac wing</td> </tr> <tr> <td>HP:0011069</td> <td>Increased number of teeth</td> </tr> </tbody> </table>								Term id	Definition	HP:0000322	Short philtrum	HP:0001151	Impaired horizontal smooth pursuit	HP:0008936	Muscular hypotonia of the trunk	HP:0002866	Hypoplastic iliac wing	HP:0011069	Increased number of teeth																																																																																																																																
Term id	Definition																																																																																																																																																		
HP:0000322	Short philtrum																																																																																																																																																		
HP:0001151	Impaired horizontal smooth pursuit																																																																																																																																																		
HP:0008936	Muscular hypotonia of the trunk																																																																																																																																																		
HP:0002866	Hypoplastic iliac wing																																																																																																																																																		
HP:0011069	Increased number of teeth																																																																																																																																																		
2	COL9A1	collagen type IX alpha 1	5	0	0	0.11	1.55E-3																																																																																																																																												
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>HP:0001151</td> <td>Impaired horizontal smooth pursuit</td> </tr> <tr> <td>HP:0005041</td> <td>Irregular capital femoral epiphysis</td> </tr> <tr> <td>HP:0003236</td> <td>Elevated serum creatine phosphokinase</td> </tr> <tr> <td>HP:0002866</td> <td>Hypoplastic iliac wing</td> </tr> <tr> <td>HP:0000322</td> <td>Short philtrum</td> </tr> </tbody> </table>								Term id	Definition	HP:0001151	Impaired horizontal smooth pursuit	HP:0005041	Irregular capital femoral epiphysis	HP:0003236	Elevated serum creatine phosphokinase	HP:0002866	Hypoplastic iliac wing	HP:0000322	Short philtrum																																																																																																																																
Term id	Definition																																																																																																																																																		
HP:0001151	Impaired horizontal smooth pursuit																																																																																																																																																		
HP:0005041	Irregular capital femoral epiphysis																																																																																																																																																		
HP:0003236	Elevated serum creatine phosphokinase																																																																																																																																																		
HP:0002866	Hypoplastic iliac wing																																																																																																																																																		
HP:0000322	Short philtrum																																																																																																																																																		
3	WWOX	WW domain containing oxidoreductase	4	0	1	-0.26	3.27E-4																																																																																																																																												
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> </tbody> </table>								Term id	Definition																																																																																																																																										
Term id	Definition																																																																																																																																																		

## Πράγα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Prague 2009 - by season - Human Phenotype

**Database:** Human Phenotype Ontology (HPO)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to Human Phenotype Ontology, BioInfoMiner analyzed 4 genes and reveals 10 ontological terms as statistically significant. This enriched set of terms corresponds to 4 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue
1	HP:0007962	Speckled corneal dystrophy	1/1	8.726E-4	0.0088

Gene Symbol	Definition	Fold Change	Pvalue
PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing	-0.14	0.044

2	HP:0100580	Barrett esophagus	1/3	2.616E-3	0.018
---	------------	-------------------	-----	----------	-------

Gene Symbol	Definition	Fold Change	Pvalue
MSR1	macrophage scavenger receptor 1	-0.15	0.042

3	HP:0011459	Esophageal carcinoma	1/4	3.486E-3	0.0257
---	------------	----------------------	-----	----------	--------

Gene Symbol	Definition	Fold Change	Pvalue
MSR1	macrophage scavenger receptor 1	-0.15	0.042

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Πράγα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Prague 2009 - by season - Human Phenotype

**Database:** Human Phenotype Ontology (HPO)

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to Human Phenotype Ontology, BioInfoMiner reveals 2 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue
1	MSR1	macrophage scavenger receptor 1	3	0	0	-0.14	0.042

Term id	Definition
HP:0100066	Neglect of the central nervous system
HP:0100580	Barrett esophagus
HP:0000119	Abnormality of the genitourinary system

2	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	2	0	0	-0.19	0.040
---	------	--	---	---	---	-------	-------

Term id	Definition
HP:0001971	Hypersplenism
HP:0000119	Abnormality of the genitourinary system

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Ostrava 2009 - by season - Human Phenotype

**Database:** Human Phenotype Ontology (HPO)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 43 genes. Excluding genes without annotation or mapping to Human Phenotype Ontology, BioInfoMiner analyzed 9 genes and reveals 22 ontological terms as statistically significant. This enriched set of terms corresponds to 7 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue
1	HP:0100751	Esophageal neoplasm	2/15	7.018E-4	0.0046

Gene Symbol	Definition	Fold Change	Pvalue
WWOX	WW domain containing oxidoreductase	-0.29	8.10E-3
LZTS1	leucine zipper, putative tumor suppressor 1	0.11	0.023

2	HP:0001258	Spastic paraplegia	3/87	1.604E-3	0.0087
---	------------	--------------------	------	----------	--------

Gene Symbol	Definition	Fold Change	Pvalue
ZFYVE27	zinc finger FYVE-type containing 27	0.12	0.033
PAX3	paired box 3	0.11	6.77E-3
WWOX	WW domain containing oxidoreductase	-0.29	8.10E-3

3	HP:0100749	Chest pain	2/35	3.842E-3	0.0149
---	------------	------------	------	----------	--------

Gene Symbol	Definition	Fold Change	Pvalue
-------------	------------	-------------	--------

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Ostrava 2009 - by season - Human Phenotype

**Database:** Human Phenotype Ontology (HPO)

**Short Description:** Input genes list contained 43 genes. Excluding genes without annotation or mapping to Human Phenotype Ontology, BioInfoMiner reveals 6 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue
1	WWOX	WW domain containing oxidoreductase	6	0	1	-0.30	8.10E-3

Term id	Definition
HP:0002716	Lymphadenopathy
HP:0008715	Testicular dysgenesis
HP:0012368	Flat face
HP:0001508	Abnormality of the voice
HP:0007105	Infantile encephalopathy
HP:0100751	Esophageal neoplasm

2	PAX3	paired box 3	3	0	0	0.11	6.77E-3
---	------	--------------	---	---	---	------	---------

Term id	Definition
HP:0007105	Infantile encephalopathy
HP:0002719	Trachomalacia
HP:0012368	Flat face

3	AMMECR1	Alport syndrome, mental retardation, mid-face hypoplasia and elliptocytosis chromosomal region gene 1	3	0	0	-0.22	1.48E-3
---	---------	---	---	---	---	-------	---------

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

Και σε περιληπτική μορφή πινάκων:

**Πίνακας 4.3.2 – 2009, ανά εποχή – Human Phenotype Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	HP:0001182	Tapered finger	3/80	1.5070e-3	4.3000e-3
2	HP:0011069	Increased number of teeth	2/25	2.2280e-3	8.7000e-3
3	HP:0000233	Thin vermilion border	3/139	7.1170e-3	0.0126
4	HP:0000272	Malar flattening	4/337	0.0152	0.0187
5	HP:0011800	Midface retrusion	3/195	0.0176	0.0239
6	HP:0005041	Irregular capital femoral epiphysis	1/8	0.0221	0.0270
7	HP:0008715	Testicular dysgenesis	1/11	0.0302	0.0300
8	HP:0004279	Short palm	2/99	0.0315	0.0379
9	HP:0002857	Genu valgum	2/106	0.0357	0.0445
10	HP:0002652	Skeletal dysplasia	2/115	0.0413	0.0477
11	HP:0008936	Muscular hypotonia of the trunk	2/137	0.0565	0.0501
12	HP:0001824	Weight loss	2/139	0.0579	0.0555
13	HP:0011097	Epileptic spasms	1/29	0.0778	0.0607
14	HP:0000322	Short philtrum	2/154	0.0693	0.0648
15	HP:0000752	Hyperactivity	2/167	0.0796	0.0657
16	HP:0000179	Thick lower lip vermilion	2/168	0.0804	0.0687
17	HP:0001151	Impaired horizontal smooth pursuit	1/31	0.0829	0.0764
18	HP:0002866	Hypoplastic iliac wing	1/33	0.0880	0.0845
19	HP:0005280	Depressed nasal bridge	3/382	0.0921	0.0870
20	HP:0003236	Elevated serum creatine phosphokinase	2/185	0.0947	0.0921
21	HP:0001336	Myoclonus	2/182	0.0921	0.0930
22	HP:0030260	Microphallus	1/37	0.0981	0.0996

**Πίνακας 4.3.3 – Πράγα 2009, ανά εποχή – Human Phenotype Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	HP:0007962	Speckled corneal dystrophy	1/1	8.7260e-4	8.8000e-3
2	HP:0100580	Barrett esophagus	1/3	2.6160e-3	0.0180
3	HP:0011459	Esophageal carcinoma	1/4	3.4860e-3	0.0257
4	HP:0001824	Weight loss	2/139	6.7040e-3	0.0329
5	HP:0001971	Hypersplenism	1/12	0.0104	0.0457
6	HP:0100006	Neoplasm of the central nervous system	1/14	0.0121	0.0545
7	HP:0000119	Abnormality of the genitourinary system	1/15	0.0130	0.0631
8	HP:0001849	Foot oligodactyly	1/16	0.0139	0.0728

9	HP:0002132	Porencephaly	1/25	0.0216	0.0852
10	HP:0002558	Supernumerary nipple	1/28	0.0242	0.0961

**Πίνακας 4.3.4 – Οστράβα 2009, ανά εποχή – Human Phenotype Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	HP:0100751	Esophageal neoplasm	2/15	7.0180e-4	4.6000e-3
2	HP:0001258	Spastic paraplegia	3/87	1.6040e-3	8.7000e-3
3	HP:0100749	Chest pain	2/35	3.8420e-3	0.0149
4	HP:0002860	Squamous cell carcinoma	2/48	7.1210e-3	0.0197
5	HP:0000179	Thick lower lip vermilion	3/168	0.0100	0.0213
6	HP:0000916	Broad clavicles	1/7	0.0182	0.0256
7	HP:0001376	Limitation of joint mobility	3/212	0.0186	0.0302
8	HP:0001182	Tapered finger	2/80	0.0189	0.0353
9	HP:0012368	Flat face	2/83	0.0202	0.0407
10	HP:0004279	Short palm	2/99	0.0281	0.0434
11	HP:0008715	Testicular dysgenesis	1/11	0.0284	0.0527
12	HP:0002716	Lymphadenopathy	2/113	0.0357	0.0546
13	HP:0001608	Abnormality of the voice	2/114	0.0363	0.0593
14	HP:0002017	Nausea and vomiting	2/123	0.0417	0.0632
15	HP:0100490	Camptodactyly of finger	2/132	0.0473	0.0681
16	HP:0001824	Weight loss	2/139	0.0519	0.0693
17	HP:0002779	Tracheomalacia	1/21	0.0536	0.0777
18	HP:0000494	Downslanted palpebral fissures	3/332	0.0574	0.0805
19	HP:0000272	Malar flattening	3/337	0.0595	0.0845
20	HP:0011069	Increased number of teeth	1/25	0.0634	0.0910
21	HP:0004568	Beaking of vertebral bodies	1/26	0.0659	0.0953
22	HP:0007105	Infantile encephalopathy	1/28	0.0708	0.0969

**Πίνακας 4.3.5 – 2009, ανά εποχή – Human Phenotype Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	RUNX2	runt related transcription factor 2	5	5	0	0	0.1062	0.0211
2	COL9A1	collagen type IX alpha 1	5	5	0	0	0.1012	1.5545e-3
3	WWOX	WW domain containing oxidoreductase	4	4	0	1	-0.2594	3.2741e-4
4	AMMECR1	Alport syndrome, mental	4	4	0	0	-0.1640	1.9777e-4



		retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1						
5	KIAA1033	KIAA1033	3	3	0	0	0.1371	0.0141

**Πίνακας 4.3.6 – Πράγα 2009, ανά εποχή – Human Phenotype Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	MSR1	macrophage scavenger receptor 1	3	3	0	0	-0.1382	0.0413
2	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	2	2	0	0	-0.1868	0.0396

**Πίνακας 4.3.7 – Οστράβα 2009, ανά εποχή – Human Phenotype Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	WWOX	WW domain containing oxidoreductase	6	6	0	1	-0.2939	8.0977e-3
2	PAX3	paired box 3	3	3	0	0	0.1010	6.7678e-3
3	AMMECR1	Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1	3	3	0	0	-0.2125	1.4843e-3
4	SH3PXD2B	SH3 and PX domains 2B	3	3	0	0	0.1100	0.0137
5	LZTS1	leucine zipper, putative tumor suppressor 1	3	3	0	0	0.1012	0.0222
6	COL9A1	collagen type IX alpha 1	2	2	0	0	0.1019	0.0337

## 4.4 – Αποτελέσματα της σύγκρισης των λιστών με την MGI Mammalian Phenotype Ontology

Πίνακας 4.4.1

MGI Mammalian Phenotype Ontology data	Hub genes																																																																																																																																																																																								
2009, ανά εποχή	2009, ανά εποχή																																																																																																																																																																																								
<p><b>BioInfoMiner</b></p> <p><b>Subject:</b> Enrichment Analysis Report (extended version)</p> <p><b>Job tag:</b> 2009 data set - by season - MGI Mammalian Phenotype</p> <p><b>Database:</b> MGI Mammalian Phenotype (MGIMP)</p> <p><b>Hypergeometric p-value threshold:</b> 0.1</p> <p><b>Corrected p-value threshold:</b> 0.1</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to MGI Mammalian Phenotype, BioInfoMiner analyzed 16 genes and reveals 16 ontological terms as statistically significant. This enriched set of terms corresponds to 9 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.</p> <p><b>Table 1: Statistically significant terms ranking according to corrected p-value</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Term id</th> <th>Definition</th> <th>Enrichment</th> <th>Hyp/metric pvalue</th> <th>Corrected pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>MP:0000600</td> <td>liver hypoplasia</td> <td>3/50</td> <td>6.681E-4</td> <td>0.0058</td> </tr> <tr> <td></td> <td><b>Gene Symbol</b></td> <td><b>Definition</b></td> <td></td> <td><b>Fold Change</b></td> <td><b>Pvalue</b></td> </tr> <tr> <td></td> <td>HGF</td> <td>hepatocyte growth factor</td> <td></td> <td>0.12</td> <td>7.22E-5</td> </tr> <tr> <td></td> <td>RUNX2</td> <td>runt related transcription factor 2</td> <td></td> <td>0.12</td> <td>0.022</td> </tr> <tr> <td></td> <td>ERN1</td> <td>endoplasmic reticulum to nucleus signaling 1</td> <td></td> <td>-0.12</td> <td>0.018</td> </tr> <tr> <td>2</td> <td>MP:0008474</td> <td>absent spleen germinal center</td> <td>2/22</td> <td>2.518E-3</td> <td>0.013</td> </tr> <tr> <td></td> <td><b>Gene Symbol</b></td> <td><b>Definition</b></td> <td></td> <td><b>Fold Change</b></td> <td><b>Pvalue</b></td> </tr> <tr> <td></td> <td>BCL6</td> <td>B-cell CLL/lymphoma 6</td> <td></td> <td>-0.14</td> <td>0.027</td> </tr> <tr> <td></td> <td>CXCR5</td> <td>C-X-C motif chemokine receptor 5</td> <td></td> <td>-0.12</td> <td>0.029</td> </tr> <tr> <td>3</td> <td>MP:0010018</td> <td>pulmonary vascular congestion</td> <td>2/23</td> <td>2.751E-3</td> <td>0.018</td> </tr> <tr> <td></td> <td><b>Gene Symbol</b></td> <td><b>Definition</b></td> <td></td> <td><b>Fold Change</b></td> <td><b>Pvalue</b></td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue	1	MP:0000600	liver hypoplasia	3/50	6.681E-4	0.0058		<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>		HGF	hepatocyte growth factor		0.12	7.22E-5		RUNX2	runt related transcription factor 2		0.12	0.022		ERN1	endoplasmic reticulum to nucleus signaling 1		-0.12	0.018	2	MP:0008474	absent spleen germinal center	2/22	2.518E-3	0.013		<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>		BCL6	B-cell CLL/lymphoma 6		-0.14	0.027		CXCR5	C-X-C motif chemokine receptor 5		-0.12	0.029	3	MP:0010018	pulmonary vascular congestion	2/23	2.751E-3	0.018		<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>	<p><b>BioInfoMiner</b></p> <p><b>Subject:</b> Genes Prioritization (extended version)</p> <p><b>Job tag:</b> 2009 data set - by season - MGI Mammalian Phenotype</p> <p><b>Database:</b> MGI Mammalian Phenotype (MGIMP)</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to MGI Mammalian Phenotype, BioInfoMiner reveals 8 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.</p> <p><b>Table 1: Hub genes ranking according to ontological clusters amount</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Gene Symbol</th> <th>Definition</th> <th>Clusters</th> <th>Interactors</th> <th>Drugs</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>RUNX2</td> <td>runt related transcription factor 2</td> <td>4</td> <td>0</td> <td>0</td> <td>0.11</td> <td>0.022</td> </tr> <tr> <td></td> <td><b>Term id</b></td> <td><b>Definition</b></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0000600</td> <td>liver hypoplasia</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0010018</td> <td>pulmonary vascular congestion</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0000694</td> <td>spleen hypoplasia</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0009346</td> <td>decreased trabecular bone thickness</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>2</td> <td>ERN1</td> <td>endoplasmic reticulum to nucleus signaling 1</td> <td>3</td> <td>0</td> <td>1</td> <td>-0.13</td> <td>0.018</td> </tr> <tr> <td></td> <td><b>Term id</b></td> <td><b>Definition</b></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0000600</td> <td>liver hypoplasia</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0008903</td> <td>abnormal placental labyrinth vasculature morphology</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>MP:0000694</td> <td>spleen hypoplasia</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>3</td> <td>NEDD9</td> <td>neural precursor cell expressed, developmentally down-regulated 9</td> <td>2</td> <td>1</td> <td>0</td> <td>0.12</td> <td>3.98E-3</td> </tr> <tr> <td></td> <td><b>Term id</b></td> <td><b>Definition</b></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue	1	RUNX2	runt related transcription factor 2	4	0	0	0.11	0.022		<b>Term id</b>	<b>Definition</b>							MP:0000600	liver hypoplasia							MP:0010018	pulmonary vascular congestion							MP:0000694	spleen hypoplasia							MP:0009346	decreased trabecular bone thickness						2	ERN1	endoplasmic reticulum to nucleus signaling 1	3	0	1	-0.13	0.018		<b>Term id</b>	<b>Definition</b>							MP:0000600	liver hypoplasia							MP:0008903	abnormal placental labyrinth vasculature morphology							MP:0000694	spleen hypoplasia						3	NEDD9	neural precursor cell expressed, developmentally down-regulated 9	2	1	0	0.12	3.98E-3		<b>Term id</b>	<b>Definition</b>					
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																																																																																																																																																																																				
1	MP:0000600	liver hypoplasia	3/50	6.681E-4	0.0058																																																																																																																																																																																				
	<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>																																																																																																																																																																																				
	HGF	hepatocyte growth factor		0.12	7.22E-5																																																																																																																																																																																				
	RUNX2	runt related transcription factor 2		0.12	0.022																																																																																																																																																																																				
	ERN1	endoplasmic reticulum to nucleus signaling 1		-0.12	0.018																																																																																																																																																																																				
2	MP:0008474	absent spleen germinal center	2/22	2.518E-3	0.013																																																																																																																																																																																				
	<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>																																																																																																																																																																																				
	BCL6	B-cell CLL/lymphoma 6		-0.14	0.027																																																																																																																																																																																				
	CXCR5	C-X-C motif chemokine receptor 5		-0.12	0.029																																																																																																																																																																																				
3	MP:0010018	pulmonary vascular congestion	2/23	2.751E-3	0.018																																																																																																																																																																																				
	<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>																																																																																																																																																																																				
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue																																																																																																																																																																																		
1	RUNX2	runt related transcription factor 2	4	0	0	0.11	0.022																																																																																																																																																																																		
	<b>Term id</b>	<b>Definition</b>																																																																																																																																																																																							
	MP:0000600	liver hypoplasia																																																																																																																																																																																							
	MP:0010018	pulmonary vascular congestion																																																																																																																																																																																							
	MP:0000694	spleen hypoplasia																																																																																																																																																																																							
	MP:0009346	decreased trabecular bone thickness																																																																																																																																																																																							
2	ERN1	endoplasmic reticulum to nucleus signaling 1	3	0	1	-0.13	0.018																																																																																																																																																																																		
	<b>Term id</b>	<b>Definition</b>																																																																																																																																																																																							
	MP:0000600	liver hypoplasia																																																																																																																																																																																							
	MP:0008903	abnormal placental labyrinth vasculature morphology																																																																																																																																																																																							
	MP:0000694	spleen hypoplasia																																																																																																																																																																																							
3	NEDD9	neural precursor cell expressed, developmentally down-regulated 9	2	1	0	0.12	3.98E-3																																																																																																																																																																																		
	<b>Term id</b>	<b>Definition</b>																																																																																																																																																																																							

## Πράγα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Prague 2009 - by season - MGI Mammalian Phenotype

**Database:** MGI Mammalian Phenotype (MGIMP)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to MGI Mammalian Phenotype, BioInfoMiner analyzed 8 genes and reveals 9 ontological terms as statistically significant. This enriched set of terms corresponds to 5 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue
1	MP:001202b	abnormal visceral endoderm physiology	1/1	1.200E-3	0.0094
<b>Gene Symbol</b>					
	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing		-0.14	0.044
2	MP:0011097	embryonic lethality between somite formation and embryo turning, complete penetrance	2/71	3.367E-3	0.0213
<b>Gene Symbol</b>					
	RBPJ	recombination signal binding protein for immunoglobulin kappa J region		-0.20	0.040
	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing		-0.14	0.044
3	MP:0001547	abnormal lipid level	2/83	4.568E-3	0.0276

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Πράγα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Prague 2009 - by season - MGI Mammalian Phenotype

**Database:** MGI Mammalian Phenotype (MGIMP)

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to MGI Mammalian Phenotype, BioInfoMiner reveals 3 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue
1	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing	4	1	0	-0.14	0.044
<b>Term id</b>		<b>Definition</b>					
	MP:0011255	abnormal anterior visceral endoderm cell migration					
	MP:0000944	spleen hypoplasia					
	MP:0001547	abnormal lipid level					
	MP:0011097	embryonic lethality between somite formation and embryo turning, complete penetrance					
2	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	3	0	0	-0.19	0.040
<b>Term id</b>		<b>Definition</b>					
	MP:0011097	embryonic lethality between somite formation and embryo turning, complete penetrance					
	MP:0012099	decreased spongiotrophoblast size					
	MP:0000944	spleen hypoplasia					
3	BCL6	B-cell CLL/Lymphoma 6	2	1	0	-0.26	0.073

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Ostrava 2009 - by season - MGI Mammalian Phenotype

**Database:** MGI Mammalian Phenotype (MGIMP)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 43 genes. Excluding genes without annotation or mapping to MGI Mammalian Phenotype, BioInfoMiner analyzed 18 genes and reveals 8 ontological terms as statistically significant. This enriched set of terms corresponds to 10 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue
1	MP:0000496	abnormal small intestine morphology	2/53	1.396E-2	0.0061
<b>Gene Symbol</b>					
	PDGFA	platelet derived growth factor subunit A		-0.14	0.022
	TLX2	T-cell leukemia homeobox 2		-0.12	9.21E-4
2	MP:0000961	abnormal dorsal root ganglion morphology	2/54	1.447E-2	0.0137
<b>Gene Symbol</b>					
	NGFR	nerve growth factor receptor		0.12	0.039
	PAX3	paired box 3		0.11	6.77E-3
3	MP:0002100	abnormal tooth morphology	2/62	1.878E-2	0.0179

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Ostrava 2009 - by season - MGI Mammalian Phenotype

**Database:** MGI Mammalian Phenotype (MGIMP)

**Short Description:** Input genes list contained 43 genes. Excluding genes without annotation or mapping to MGI Mammalian Phenotype, BioInfoMiner reveals 5 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue
1	NGFR	nerve growth factor receptor	4	0	2	0.11	0.039
<b>Term id</b>		<b>Definition</b>					
	MP:0011967	increased or absent threshold for auditory brainstem response					
	MP:0002100	abnormal tooth morphology					
	MP:0000961	abnormal dorsal root ganglion morphology					
	MP:0001463	abnormal spatial learning					
2	TLX2	T-cell leukemia homeobox 2	3	0	0	-0.13	9.21E-4
<b>Term id</b>		<b>Definition</b>					
	MP:0011096	embryonic lethality between implantation and somite formation, complete penetrance					
	MP:0000496	abnormal small intestine morphology					
	MP:0002100	abnormal tooth morphology					
3	PAX3	paired box 3	3	0	0	0.11	6.77E-3
<b>Term id</b>		<b>Definition</b>					
	MP:0011096	embryonic lethality between implantation and somite formation, complete penetrance					
	MP:0002100	abnormal tooth morphology					
	MP:0000961	abnormal dorsal root ganglion morphology					

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

Και σε περιληπτική μορφή πινάκων:

**Πίνακας 4.4.2 – 2009, ανά εποχή – MGI Mammalian Phenotype Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	MP:0000600	liver hypoplasia	3/50	6.6810e-4	5.8000e-3
2	MP:0008474	absent spleen germinal center	2/22	2.5180e-3	0.0130
3	MP:0010018	pulmonary vascular congestion	2/23	2.7510e-3	0.0180
4	MP:0000696	abnormal Peyer's patch morphology	2/27	3.7830e-3	0.0226
5	MP:0000694	spleen hypoplasia	3/105	5.5720e-3	0.0311
6	MP:0008395	abnormal osteoblast differentiation	2/34	5.9530e-3	0.0350
7	MP:0009346	decreased trabecular bone thickness	2/35	6.3000e-3	0.0408
8	MP:0000558	abnormal tibia morphology	2/43	9.3920e-3	0.0485
9	MP:0003156	abnormal leukocyte migration	2/52	0.0135	0.0531
10	MP:0005006	abnormal osteoblast physiology	2/54	0.0145	0.0536
11	MP:0000559	abnormal femur morphology	2/59	0.0172	0.0641
12	MP:0000135	decreased compact bone thickness	2/62	0.0189	0.0687
13	MP:0008803	abnormal placental labyrinth vasculature morphology	2/72	0.0249	0.0747
14	MP:0000601	small liver	2/74	0.0262	0.0787
15	MP:0002113	abnormal skeleton development	2/75	0.0269	0.0897
16	MP:0002123	abnormal definitive hematopoiesis	2/93	0.0398	0.0937

**Πίνακας 4.4.3 – Πράγα 2009, ανά εποχή – MGI Mammalian Phenotype Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	MP:0012028	abnormal visceral endoderm physiology	1/1	1.2000e-3	9.4000e-3
2	MP:0011097	embryonic lethality between somite formation and embryo turning, complete penetrance	2/71	3.3670e-3	0.0213
3	MP:0001547	abnormal lipid level	2/83	4.5680e-3	0.0276
4	MP:0010187	decreased T follicular helper cell number	1/4	4.7900e-3	0.0384
5	MP:0001800	abnormal humoral immune response	2/100	6.5560e-3	0.0504
6	MP:0000694	spleen hypoplasia	2/105	7.2040e-3	0.0601
7	MP:0003666	impaired sperm capacitation	1/13	0.0155	0.0699
8	MP:0000689	abnormal spleen morphology	2/165	0.0170	0.0743

9	MP:0012099	decreased spongiotrophoblast size	1/17	0.0202	0.0847
---	------------	-----------------------------------	------	--------	--------

**Πίνακας 4.4.4 – Οστράβα 2009, ανά εποχή – MGI Mammalian Phenotype Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	MP:0000496	abnormal small intestine morphology	2/53	0.0140	6.1000e-3
2	MP:0000961	abnormal dorsal root ganglion morphology	2/54	0.0145	0.0137
3	MP:0002100	abnormal tooth morphology	2/62	0.0188	0.0179
4	MP:0001463	abnormal spatial learning	3/167	0.0193	0.0238
5	MP:0002113	abnormal skeleton development	2/75	0.0268	0.0302
6	MP:0001732	postnatal growth retardation	5/587	0.0502	0.0337
7	MP:0011096	embryonic lethality between implantation and somite formation, complete penetrance	3/245	0.0508	0.0442
8	MP:0011967	increased or absent threshold for auditory brainstem response	3/246	0.0513	0.0484

**Πίνακας 4.4.5 – 2009, ανά εποχή – MGI Mammalian Phenotype Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	RUNX2	runt related transcription factor 2	4	4	0	0	0.1062	0.0211
2	ERN1	endoplasmic reticulum to nucleus signaling 1	3	3	0	1	-0.1200	0.0176
3	NEDD9	neural precursor cell expressed, developmentally down-regulated 9	2	2	1	0	0.1161	3.9818e-3
4	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing	2	2	1	0	-0.1052	1.7993e-4
5	CXCR5	C-X-C motif chemokine receptor 5	2	2	0	0	-0.1158	0.0290
6	WWOX	WW domain containing oxidoreductase	2	2	0	1	-0.2594	3.2741e-4
7	HGF	hepatocyte	2	2	0	6	0.1071	7.2242e-

		growth factor						5
8	BCL6	B-cell CLL/lymphoma 6	2	2	1	0	- 0.1288	0.0268

**Πίνακας 4.4.6 – Πράγα 2009, ανά εποχή – MGI Mammalian Phenotype Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing	4	3	1	0	- 0.1340	0.0431
2	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	3	3	0	0	- 0.1868	0.0396
3	BCL6	B-cell CLL/lymphoma 6	2	2	1	0	- 0.2543	0.0724

**Πίνακας 4.4.7 – Οστράβα 2009, ανά εποχή – MGI Mammalian Phenotype Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	NGFR	nerve growth factor receptor	4	4	0	2	0.1068	0.0381
2	TLX2	T-cell leukemia homeobox 2	3	3	0	0	- 0.1203	9.2086e-4
3	PAX3	paired box 3	3	3	0	0	0.1010	6.7678e-3
4	SH3PXD2B	SH3 and PX domains 2B	2	2	0	0	0.1100	0.0137
5	PDGFA	platelet derived growth factor subunit A	2	2	0	0	- 0.1259	0.0215

## 4.5 – Αποτελέσματα της σύγκρισης των λιστών με τη Reactome Pathways Ontology

Πίνακας 4.5.1

Reactome Pathways Ontology data	Hub genes																																																																																														
2009, ανά εποχή	2009, ανά εποχή																																																																																														
<p style="text-align: center;"><b>BioInfoMiner</b></p> <p><b>Subject:</b> Enrichment Analysis Report (extended version)  <b>Job tag:</b> 2009 data set - by season - Reactome pathways  <b>Database:</b> Reactome Pathways (Reactome)  <b>Hypergeometric p-value threshold:</b> 0.1  <b>Corrected p-value threshold:</b> 0.1</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to Reactome Pathways, BioInfoMiner analyzed 18 genes and reveals 8 ontological terms as statistically significant. This enriched set of terms corresponds to 7 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.</p> <p style="text-align: center;"><b>Table 1: Statistically significant terms ranking according to corrected p-value</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Term id</th> <th>Definition</th> <th>Enrichment</th> <th>Hyp/metric pvalue</th> <th>Corrected pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>R-HSA-5633008</td> <td>TP53 Regulates Transcription of Cell Death Genes</td> <td>2/43</td> <td>1.085E-3</td> <td>0.0135</td> </tr> <tr> <td></td> <td><b>Gene Symbol</b></td> <td><b>Definition</b></td> <td></td> <td><b>Fold Change</b></td> <td><b>Pvalue</b></td> </tr> <tr> <td></td> <td>BCL6</td> <td>B-cell CLL/lymphoma 6</td> <td></td> <td>-0.14</td> <td>0.027</td> </tr> <tr> <td></td> <td>TNFRSF10D</td> <td>tumor necrosis factor receptor superfamily member 10d</td> <td></td> <td>0.12</td> <td>0.028</td> </tr> <tr> <td>2</td> <td>R-HSA-5619108</td> <td>Defective SLC27A4 causes ichthyosis prematurity syndrome (IPS)</td> <td>1/1</td> <td>1.118E-3</td> <td>0.0253</td> </tr> <tr> <td></td> <td><b>Gene Symbol</b></td> <td><b>Definition</b></td> <td></td> <td><b>Fold Change</b></td> <td><b>Pvalue</b></td> </tr> <tr> <td></td> <td>SLC27A4</td> <td>solute carrier family 27 member 4</td> <td></td> <td>0.11</td> <td>0.015</td> </tr> <tr> <td>3</td> <td>R-HSA-804914</td> <td>Transport of fatty acids</td> <td>1/8</td> <td>8.906E-3</td> <td>0.0386</td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue	1	R-HSA-5633008	TP53 Regulates Transcription of Cell Death Genes	2/43	1.085E-3	0.0135		<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>		BCL6	B-cell CLL/lymphoma 6		-0.14	0.027		TNFRSF10D	tumor necrosis factor receptor superfamily member 10d		0.12	0.028	2	R-HSA-5619108	Defective SLC27A4 causes ichthyosis prematurity syndrome (IPS)	1/1	1.118E-3	0.0253		<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>		SLC27A4	solute carrier family 27 member 4		0.11	0.015	3	R-HSA-804914	Transport of fatty acids	1/8	8.906E-3	0.0386	<p style="text-align: center;"><b>BioInfoMiner</b></p> <p><b>Subject:</b> Genes Prioritization (extended version)  <b>Job tag:</b> 2009 data set - by season - Reactome pathways  <b>Database:</b> Reactome Pathways (Reactome)</p> <p><b>Short Description:</b> Input genes list contained 42 genes. Excluding genes without annotation or mapping to Reactome Pathways, BioInfoMiner reveals 1 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.</p> <p style="text-align: center;"><b>Table 1: Hub genes ranking according to ontological clusters amount</b></p> <table border="1"> <thead> <tr> <th>Rank</th> <th>Gene Symbol</th> <th>Definition</th> <th>Clusters</th> <th>Interactors</th> <th>Drugs</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>HIST1H1E</td> <td>histone cluster 1, H1e</td> <td>2</td> <td>0</td> <td>0</td> <td>0.18</td> <td>1.88E-4</td> </tr> <tr> <td></td> <td><b>Term id</b></td> <td><b>Definition</b></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>R-HSA-2559584</td> <td>Formation of Senescence-Associated Heterochromatin Foci (SAHF)</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td></td> <td>R-HSA-2559583</td> <td>Cellular Senescence</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>e-NIOS Applications Private Company    telephone: +30 211 999 75 67,    e-mail: info@e-nios.com    1</p>	Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue	1	HIST1H1E	histone cluster 1, H1e	2	0	0	0.18	1.88E-4		<b>Term id</b>	<b>Definition</b>							R-HSA-2559584	Formation of Senescence-Associated Heterochromatin Foci (SAHF)							R-HSA-2559583	Cellular Senescence					
Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																																																																																										
1	R-HSA-5633008	TP53 Regulates Transcription of Cell Death Genes	2/43	1.085E-3	0.0135																																																																																										
	<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>																																																																																										
	BCL6	B-cell CLL/lymphoma 6		-0.14	0.027																																																																																										
	TNFRSF10D	tumor necrosis factor receptor superfamily member 10d		0.12	0.028																																																																																										
2	R-HSA-5619108	Defective SLC27A4 causes ichthyosis prematurity syndrome (IPS)	1/1	1.118E-3	0.0253																																																																																										
	<b>Gene Symbol</b>	<b>Definition</b>		<b>Fold Change</b>	<b>Pvalue</b>																																																																																										
	SLC27A4	solute carrier family 27 member 4		0.11	0.015																																																																																										
3	R-HSA-804914	Transport of fatty acids	1/8	8.906E-3	0.0386																																																																																										
Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue																																																																																								
1	HIST1H1E	histone cluster 1, H1e	2	0	0	0.18	1.88E-4																																																																																								
	<b>Term id</b>	<b>Definition</b>																																																																																													
	R-HSA-2559584	Formation of Senescence-Associated Heterochromatin Foci (SAHF)																																																																																													
	R-HSA-2559583	Cellular Senescence																																																																																													

## Πράγα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Prague 2009 - by season - Reactome pathways

**Database:** Reactome Pathways (Reactome)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to Reactome Pathways, BioInfoMiner analyzed 12 genes and reveals 6 ontological terms as statistically significant. This enriched set of terms corresponds to 6 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue																				
1	R-HSA-212436	Generic Transcription Pathway	4/787	4.525E-3	0.0168																				
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>ZNF460</td> <td>zinc finger protein 460</td> <td>-0.17</td> <td>0.030</td> </tr> <tr> <td>CNOT6L</td> <td>CCR4-NOT transcription complex subunit 6 like</td> <td>-0.12</td> <td>0.093</td> </tr> <tr> <td>BCL6</td> <td>B-cell CLL/lymphoma 6</td> <td>-0.25</td> <td>0.073</td> </tr> <tr> <td>RBPJ</td> <td>recombination signal binding protein for immunoglobulin kappa J region</td> <td>-0.20</td> <td>0.040</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	ZNF460	zinc finger protein 460	-0.17	0.030	CNOT6L	CCR4-NOT transcription complex subunit 6 like	-0.12	0.093	BCL6	B-cell CLL/lymphoma 6	-0.25	0.073	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040
Gene Symbol	Definition	Fold Change	Pvalue																						
ZNF460	zinc finger protein 460	-0.17	0.030																						
CNOT6L	CCR4-NOT transcription complex subunit 6 like	-0.12	0.093																						
BCL6	B-cell CLL/lymphoma 6	-0.25	0.073																						
RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040																						
2	R-HSA-210744	Regulation of gene expression in late stage (branching morphogenesis) pancreatic bud precursor cells	1/7	5.888E-3	0.0334																				
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>RBPJ</td> <td>recombination signal binding protein for immunoglobulin kappa J region</td> <td>-0.20</td> <td>0.040</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040												
Gene Symbol	Definition	Fold Change	Pvalue																						
RBPJ	recombination signal binding protein for immunoglobulin kappa J region	-0.20	0.040																						

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Πράγα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Prague 2009 - by season - Reactome pathways

**Database:** Reactome Pathways (Reactome)

**Short Description:** Input genes list contained 26 genes. Excluding genes without annotation or mapping to Reactome Pathways, BioInfoMiner reveals 1 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue						
1	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	2	0	0	-0.19	0.040						
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>R-HSA-212436</td> <td>Generic Transcription Pathway</td> </tr> <tr> <td>R-HSA-210744</td> <td>Regulation of gene expression in late stage (branching morphogenesis) pancreatic bud precursor cells</td> </tr> </tbody> </table>								Term id	Definition	R-HSA-212436	Generic Transcription Pathway	R-HSA-210744	Regulation of gene expression in late stage (branching morphogenesis) pancreatic bud precursor cells
Term id	Definition												
R-HSA-212436	Generic Transcription Pathway												
R-HSA-210744	Regulation of gene expression in late stage (branching morphogenesis) pancreatic bud precursor cells												

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

### BioInfoMiner

**Subject:** Enrichment Analysis Report (extended version)

**Job tag:** Ostrava 2009 - by season - Reactome pathways

**Database:** Reactome Pathways (Reactome)

**Hypergeometric p-value threshold:** 0.1

**Corrected p-value threshold:** 0.1

**Short Description:** Input genes list contained 43 genes. Excluding genes without annotation or mapping to Reactome Pathways, BioInfoMiner analyzed 16 genes and reveals 9 ontological terms as statistically significant. This enriched set of terms corresponds to 10 input genes. Significant terms, in conjunction with their correlated genes, are displayed in Table 1. A brief delineation of results is visualized in Figure 1.

**Table 1: Statistically significant terms ranking according to corrected p-value**

Rank	Term id	Definition	Enrichment	Hyp/metric pvalue	Corrected pvalue								
1	R-HSA-205017	NFG and proNGF binds to p75NTR	1/2	2.721E-3	0.0116								
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>NGFR</td> <td>nerve growth factor receptor</td> <td>0.12</td> <td>0.039</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	NGFR	nerve growth factor receptor	0.12	0.039
Gene Symbol	Definition	Fold Change	Pvalue										
NGFR	nerve growth factor receptor	0.12	0.039										
2	R-HSA-193681	Ceramide signalling	1/3	4.079E-3	0.0231								
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>NGFR</td> <td>nerve growth factor receptor</td> <td>0.12</td> <td>0.039</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	NGFR	nerve growth factor receptor	0.12	0.039
Gene Symbol	Definition	Fold Change	Pvalue										
NGFR	nerve growth factor receptor	0.12	0.039										
3	R-HSA-209563	Axonal growth stimulation	1/4	5.434E-3	0.0328								
<table border="1"> <thead> <tr> <th>Gene Symbol</th> <th>Definition</th> <th>Fold Change</th> <th>Pvalue</th> </tr> </thead> <tbody> <tr> <td>NGFR</td> <td>nerve growth factor receptor</td> <td>0.12</td> <td>0.039</td> </tr> </tbody> </table>						Gene Symbol	Definition	Fold Change	Pvalue	NGFR	nerve growth factor receptor	0.12	0.039
Gene Symbol	Definition	Fold Change	Pvalue										
NGFR	nerve growth factor receptor	0.12	0.039										

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1

## Οστράβα 2009

### BioInfoMiner

**Subject:** Genes Prioritization (extended version)

**Job tag:** Ostrava 2009 - by season - Reactome pathways

**Database:** Reactome Pathways (Reactome)

**Short Description:** Input genes list contained 43 genes. Excluding genes without annotation or mapping to Reactome Pathways, BioInfoMiner reveals 1 genes as hub nodes on the enriched ontological graph. Central role genes, in conjunction with their related ontological clusters, are displayed in Table 1.

**Table 1: Hub genes ranking according to ontological clusters amount**

Rank	Gene Symbol	Definition	Clusters	Interactors	Drugs	Fold Change	Pvalue						
1	NGFR	nerve growth factor receptor	2	0	2	0.11	0.039						
<table border="1"> <thead> <tr> <th>Term id</th> <th>Definition</th> </tr> </thead> <tbody> <tr> <td>R-HSA-193634</td> <td>Axonal growth inhibition (RHOA activation)</td> </tr> <tr> <td>R-HSA-205017</td> <td>NFG and proNGF binds to p75NTR</td> </tr> </tbody> </table>								Term id	Definition	R-HSA-193634	Axonal growth inhibition (RHOA activation)	R-HSA-205017	NFG and proNGF binds to p75NTR
Term id	Definition												
R-HSA-193634	Axonal growth inhibition (RHOA activation)												
R-HSA-205017	NFG and proNGF binds to p75NTR												

e-NIOS Applications Private Company telephone: +30 211 999 75 67, e-mail: info@e-nios.com 1



Και σε περιληπτική μορφή πινάκων:

**Πίνακας 4.5.2 – 2009, ανά εποχή – Reactome pathways Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	R-HSA-5633008	TP53 Regulates Transcription of Cell Death Genes	2/43	1.0850e-3	0.0135
2	R-HSA-5619108	Defective SLC27A4 causes ichthyosis prematurity syndrome (IPS)	1/1	1.1180e-3	0.0253
3	R-HSA-804914	Transport of fatty acids	1/8	8.9060e-3	0.0386
4	R-HSA-1660517	Synthesis of PIPs at the late endosome membrane	1/10	0.0111	0.0474
5	R-HSA-975577	N-Glycan antennae elongation	1/15	0.0166	0.0601
6	R-HSA-2559584	Formation of Senescence-Associated Heterochromatin Foci (SAHF)	1/16	0.0177	0.0751
7	R-HSA-2559583	Cellular Senescence	2/189	0.0192	0.0848
8	R-HSA-1660514	Synthesis of PIPs at the Golgi membrane	1/18	0.0199	0.0984

**Πίνακας 4.5.3 – Πράγα 2009, ανά εποχή – Reactome pathways Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	R-HSA-212436	Generic Transcription Pathway	4/787	4.5250e-3	0.0168
2	R-HSA-210744	Regulation of gene expression in late stage (branching morphogenesis) pancreatic bud precursor cells	1/7	5.8880e-3	0.0334
3	R-HSA-1660517	Synthesis of PIPs at the late endosome membrane	1/10	8.4010e-3	0.0461
4	R-HSA-2197563	NOTCH2 intracellular domain regulates transcription	1/12	0.0101	0.0684
5	R-HSA-2559584	Formation of Senescence-Associated Heterochromatin Foci (SAHF)	1/16	0.0134	0.0825
6	R-HSA-1660514	Synthesis of PIPs at the Golgi membrane	1/18	0.0151	0.0974

**Πίνακας 4.5.4 – Οστράβα 2009, ανά εποχή – Reactome pathways Ontology**

Rank	Term id	Term Definition	Enrichment	Hypergeometric pvalue	Corrected pvalue
1	R-HSA-205017	NFG and proNGF binds to p75NTR	1/2	2.7210e-3	0.0116

2	R-HSA-193681	Ceramide signalling	1/3	4.0790e-3	0.0231
3	R-HSA-209563	Axonal growth stimulation	1/4	5.4340e-3	0.0328
4	R-HSA-193634	Axonal growth inhibition (RHOA activation)	1/9	0.0122	0.0463
5	R-HSA-186797	Signaling by PDGF	3/357	0.0131	0.0570
6	R-HSA-77588	SLBP Dependent Processing of Replication-Dependent Histone Pre-mRNAs	1/11	0.0149	0.0658
7	R-HSA-2262752	Cellular responses to stress	3/383	0.0158	0.0729
8	R-HSA-1266695	Interleukin-7 signaling	1/12	0.0162	0.0919
9	R-HSA-5682910	LGI-ADAM interactions	1/14	0.0189	0.0975

**Πίνακας 4.5.5 – 2009, ανά εποχή – Reactome pathways Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	HIST1H1E	histone cluster 1, H1e	2	2	0	0	0.1727	1.8843e-4

**Πίνακας 4.5.6 – Πράγα 2009, ανά εποχή – Reactome pathways Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	2	2	0	0	-0.1868	0.0396

**Πίνακας 4.5.7 – Οστράβα 2009, ανά εποχή – Reactome pathways Ontology hub genes**

Rank	Gene Symbol	Definition	Clusters	Enriched Clusters	Interactors	Associated Drugs	Fold Change	Pvalue
1	NGFR	nerve growth factor receptor	2	2	0	2	0.1068	0.0381

#### 4.6 – Αναζήτηση υποσυνόλου των γονιδίων-κόμβων που να αναδεικνύει το διαχωρισμό μεταξύ των πόλεων.

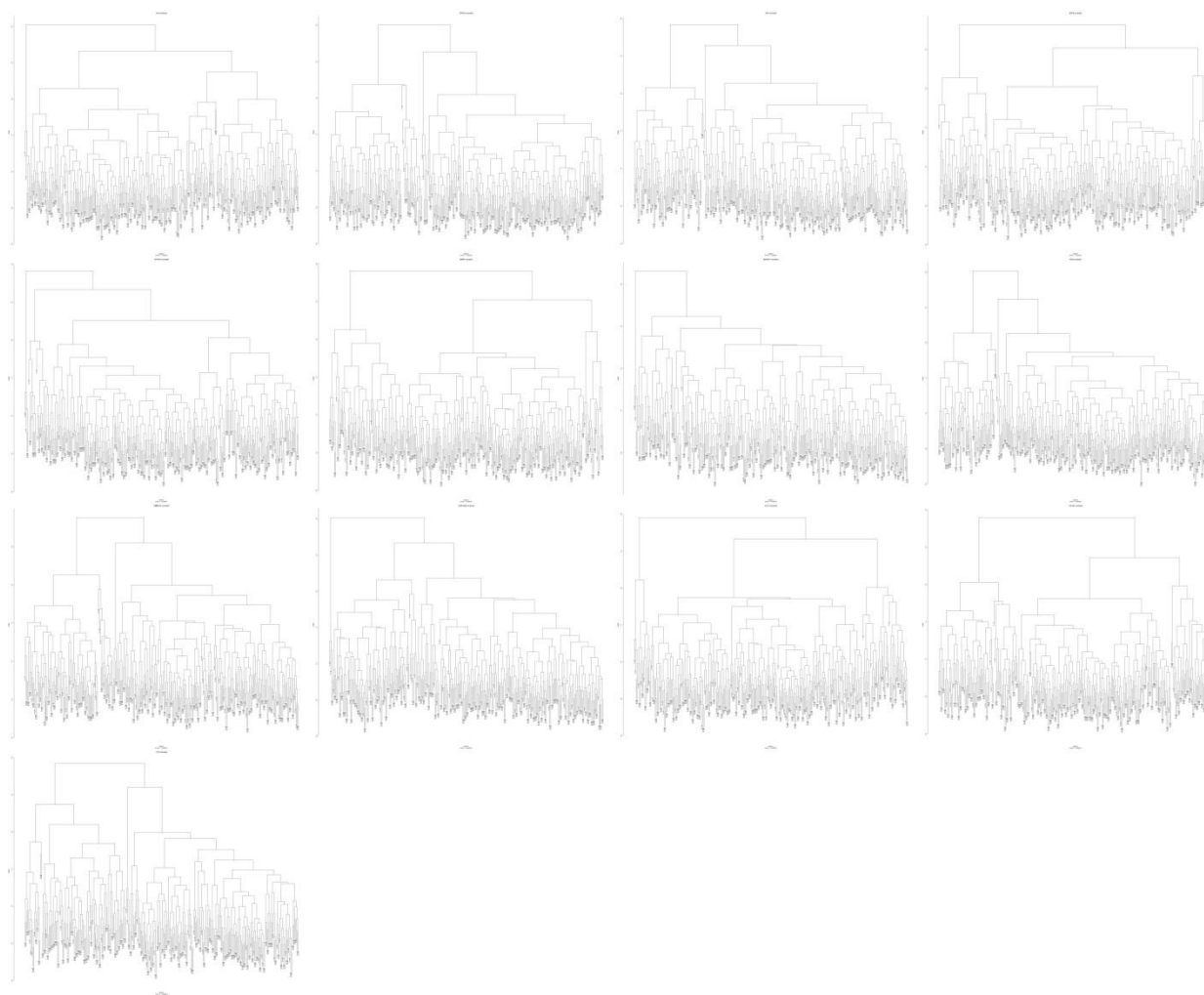
Ξεκινώντας από τη λίστα των επιλεγμένων hub genes PDGFA, HGF, NGFR, ZFYVE27, MINK1, LTK, RAPGEF1, PAX3, AMMECR1, SH3PXD2B, LZTS1, COL9A1, TLX2, με κριτήριο τη βελτίωση του διαχωρισμού, όπως παρουσιάζεται στα επόμενα τέσσερα διαδοχικά σύνολα διαγραμμάτων, αποκλείστηκαν κατά σειρά τα γονίδια LTK, TLX2, AMMECR1 και LZTS1, ώστε να φτάσουμε στην τελική λίστα PDGFA, HGF, NGFR, ZFYVE27, MINK1, RAPGEF1, PAX3, SH3PXD2B, COL9A1, καθώς η περεταίρω αφαίρεση γονιδίων δε φάνηκε να βελτιώνει την ποιότητα του διαχωρισμού. Λόγω του μεγάλου αριθμού δειγμάτων, θα πρέπει να παραπέμψουμε στα supplementary αρχεία, στο link που υπάρχει στην κάθε εικόνα, ώστε να μπορεί να επεικονιστεί σε καλή ανάλυση το κάθε σύνολο γραφημάτων.

Με την πλήρη λίστα:



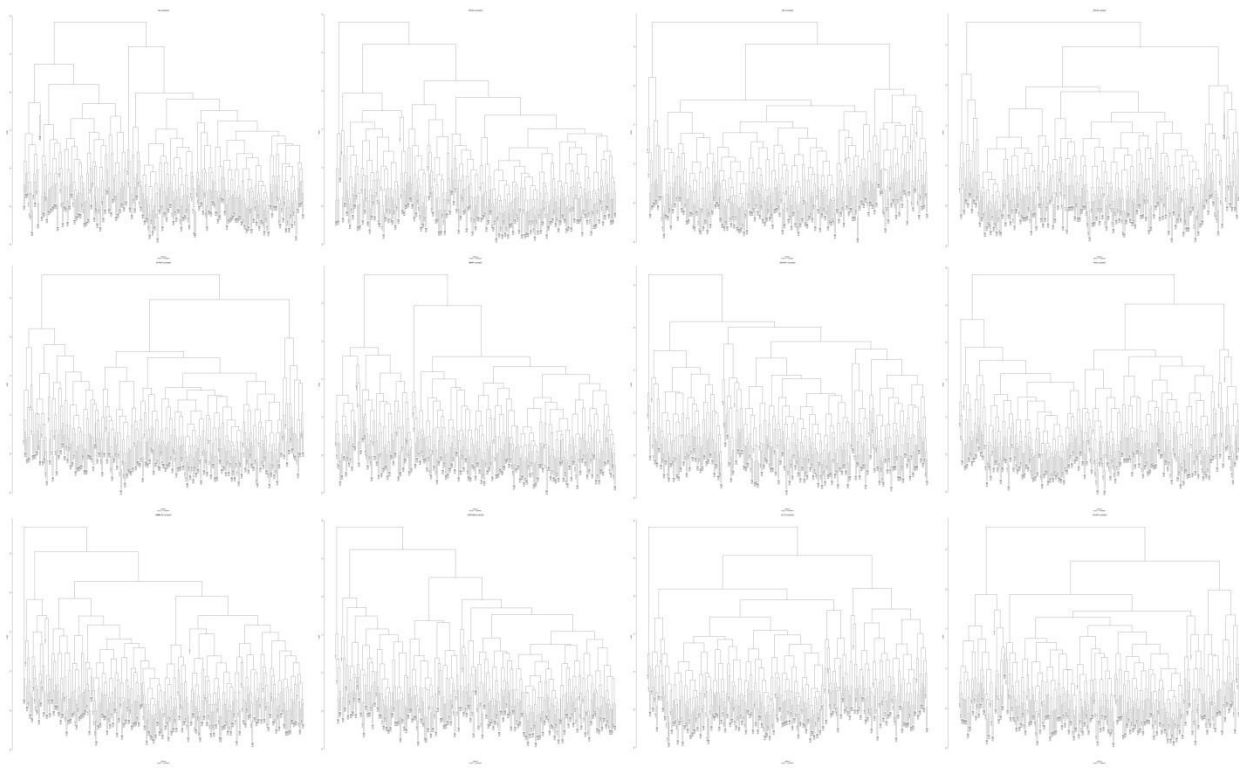
Μετά την αφαίρεση του LTK.

**Εικόνα 4.6.1**



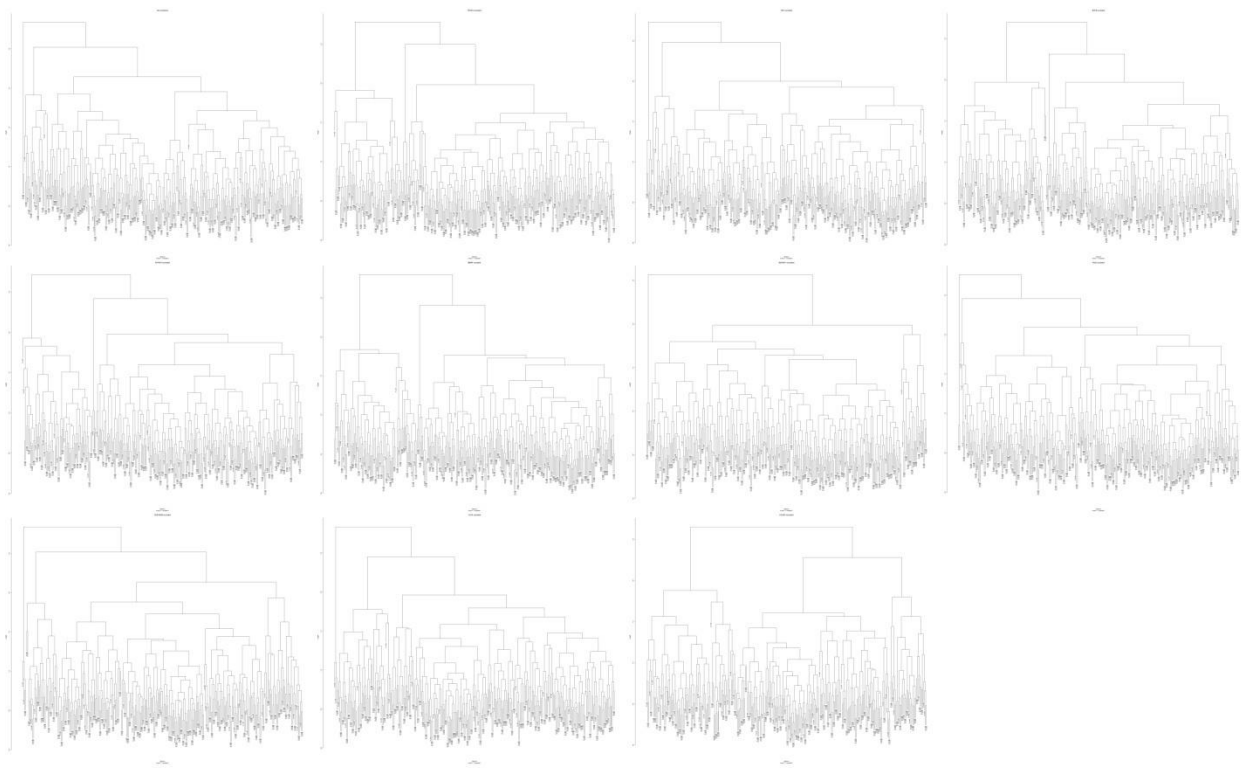
**Εικόνα 4.6.2**

Μετά την αφαίρεση του TLX2.



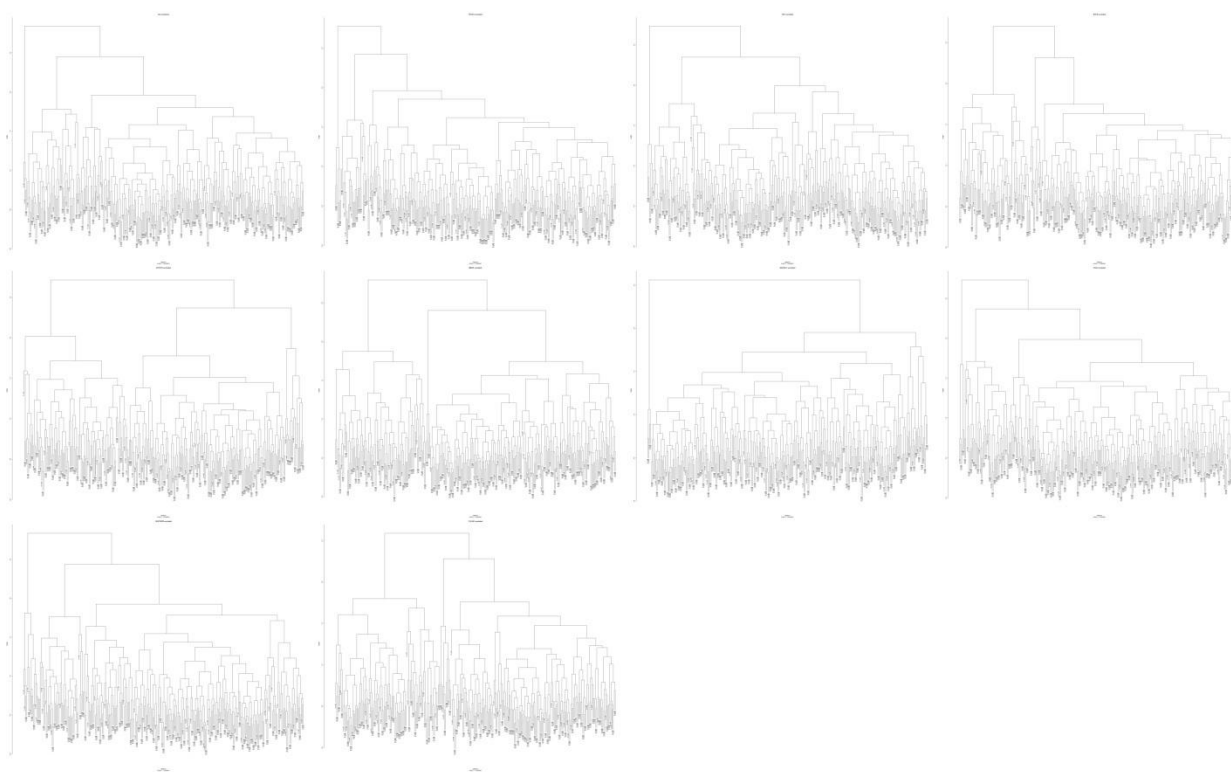
**Εικόνα 4.6.3**

Μετά την αφαίρεση του AMMECR1.



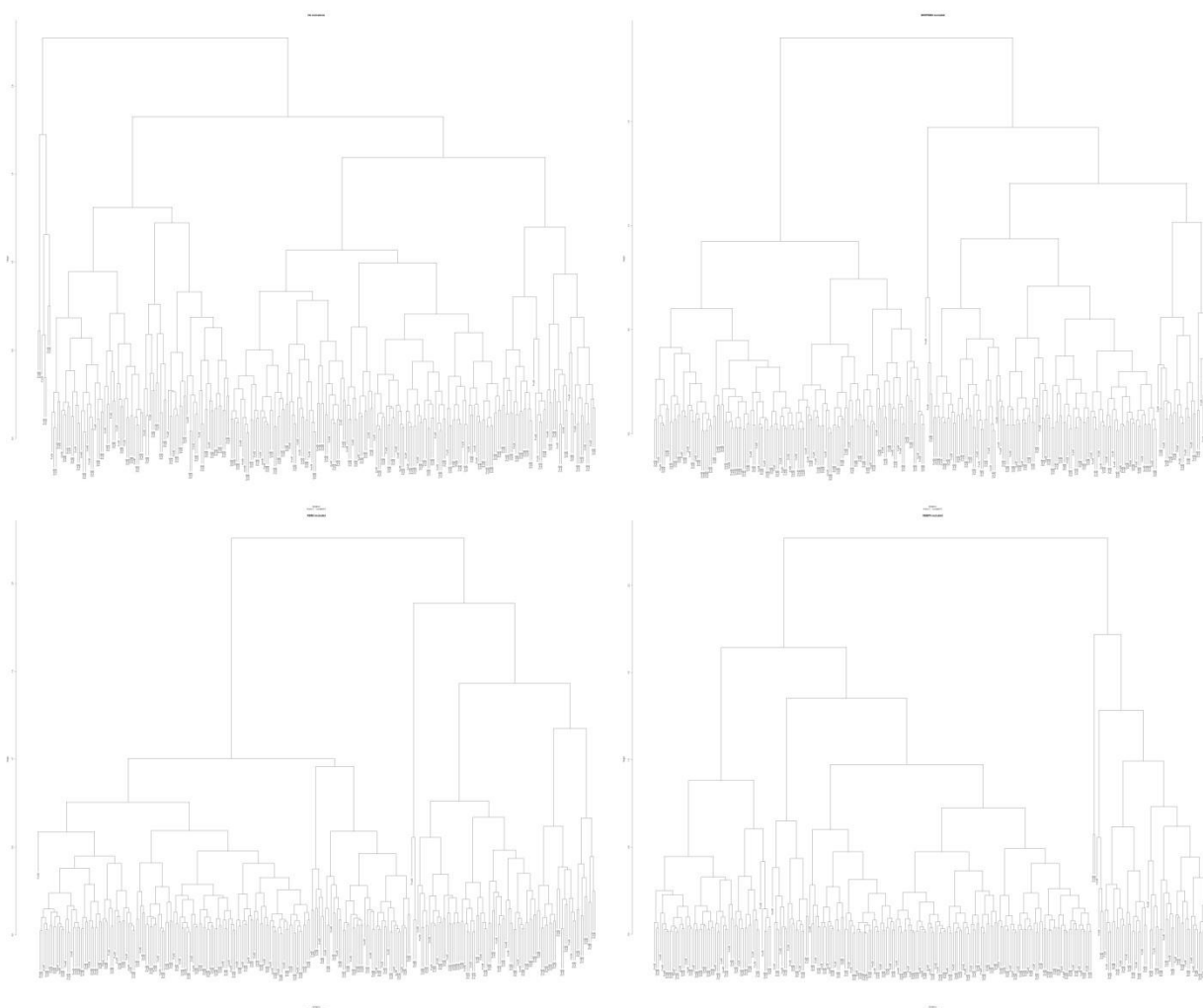
**Εικόνα 4.6.4**

Μετά την αφαίρεση του LZTS1.



Εικόνα 4.6.5

Αναζήτηση βέλτιστου συνδυασμού μεταξύ των γονιδίων που αναγνωρίστηκαν ως διαφορετικώς εκφρασμένα το καλοκαίρι του 2009, ώστε να χρησιμοποιηθεί ως μέτρο σύγκρισης.

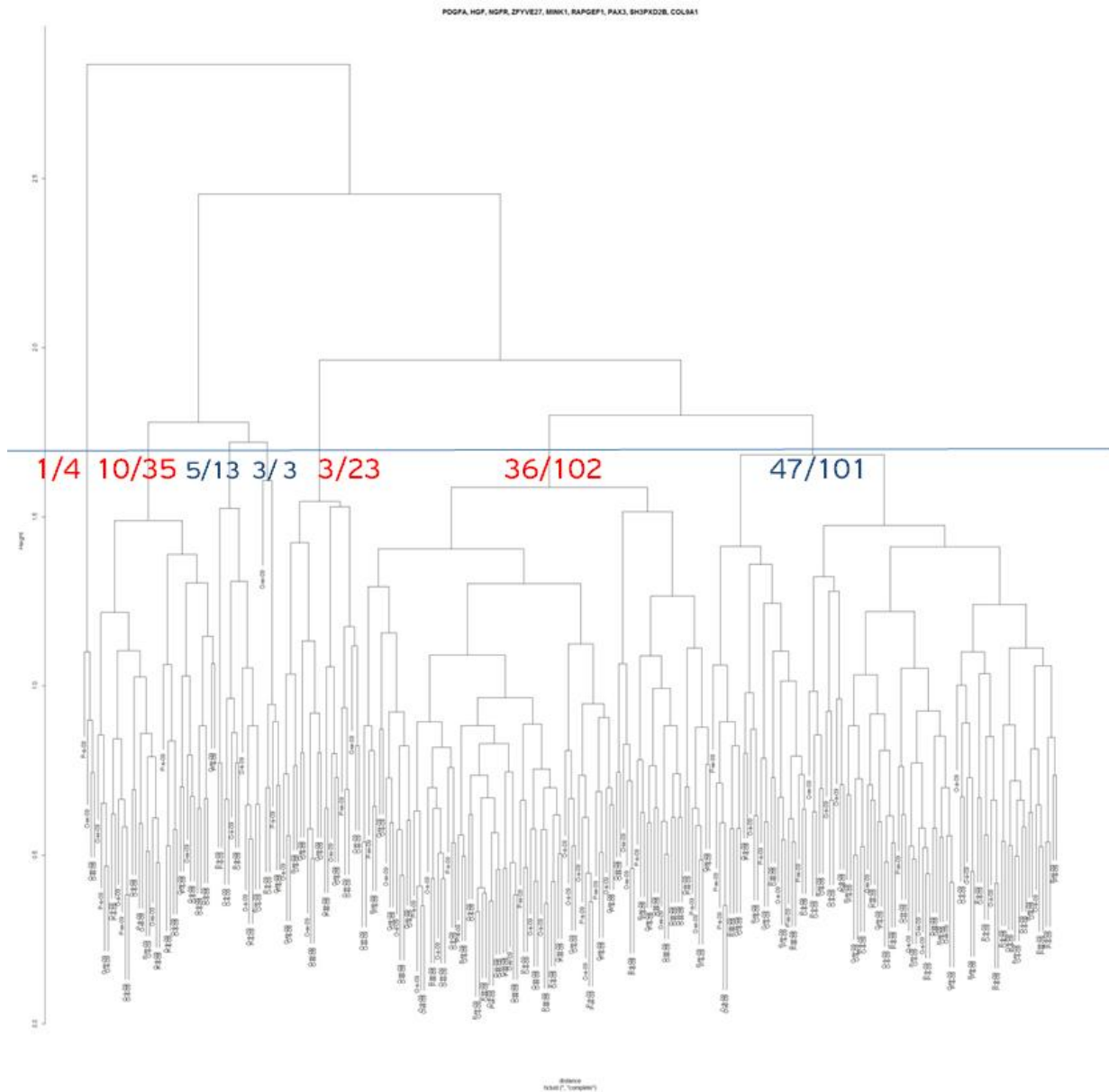


**Εικόνα 4.6.6**



Δενδρόγραμμα με βάση την τελική λίστα, προερχόμενη από τα hub genes: PDGFA, HGF, NGFR, ZFYVE27, MINK1, RAPGEF1, PAX3, SH3PXD2B, COL9A1

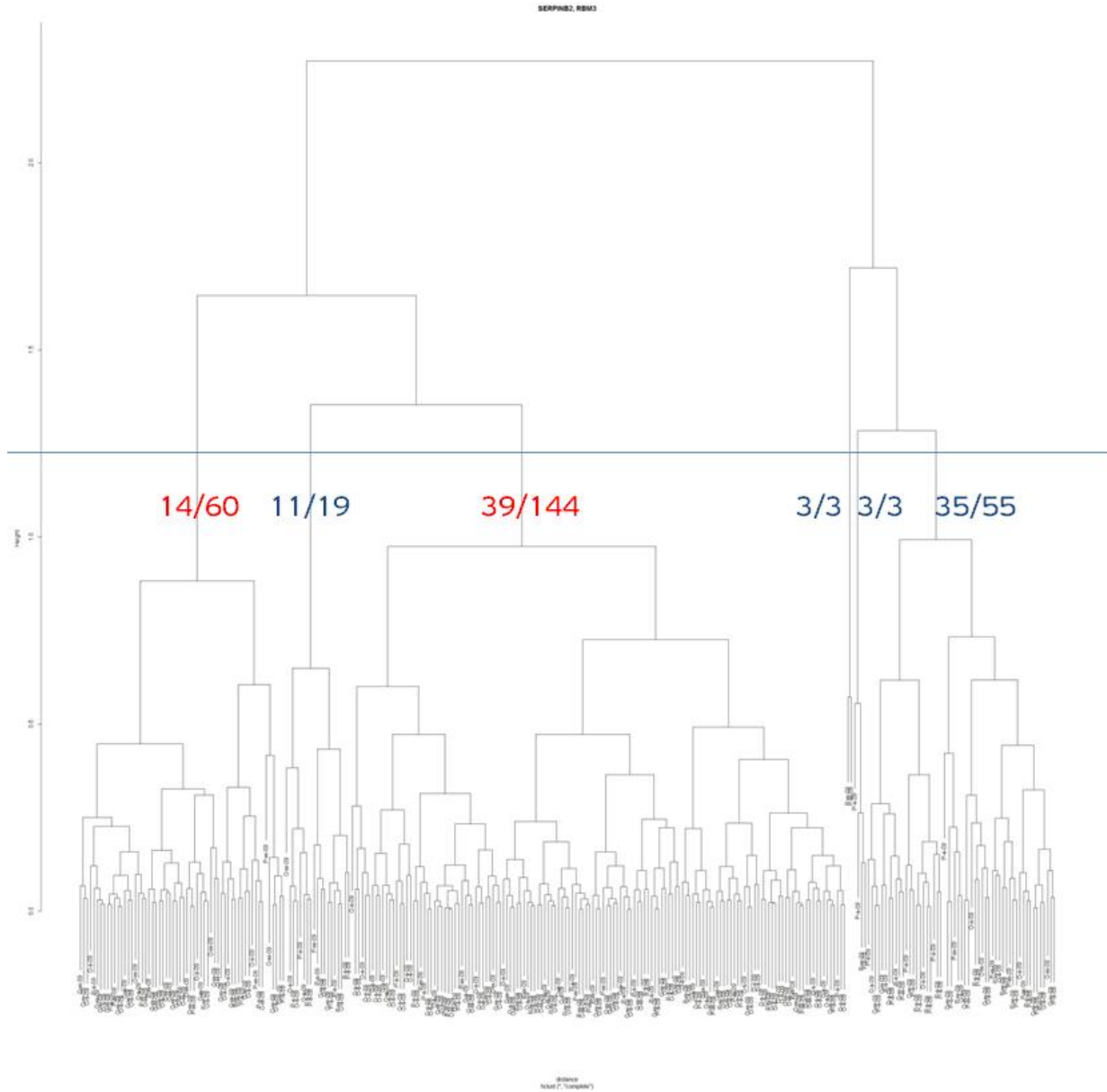
Η μέση τιμή της εκπροσώπησης της Πράγας στον πληθυσμό δειγμάτων του 2009 είναι ~37%. Έτσι, τιμές χαμηλότερες από αυτήν σε ένα κλάδο του δενδροδιαγράμματος δείχνουν υπερεκπροσώπηση της Οστράβα και υψηλότερες από αυτήν, υπερεκπροσώπηση της Πράγας.



Εικόνα 4.6.7

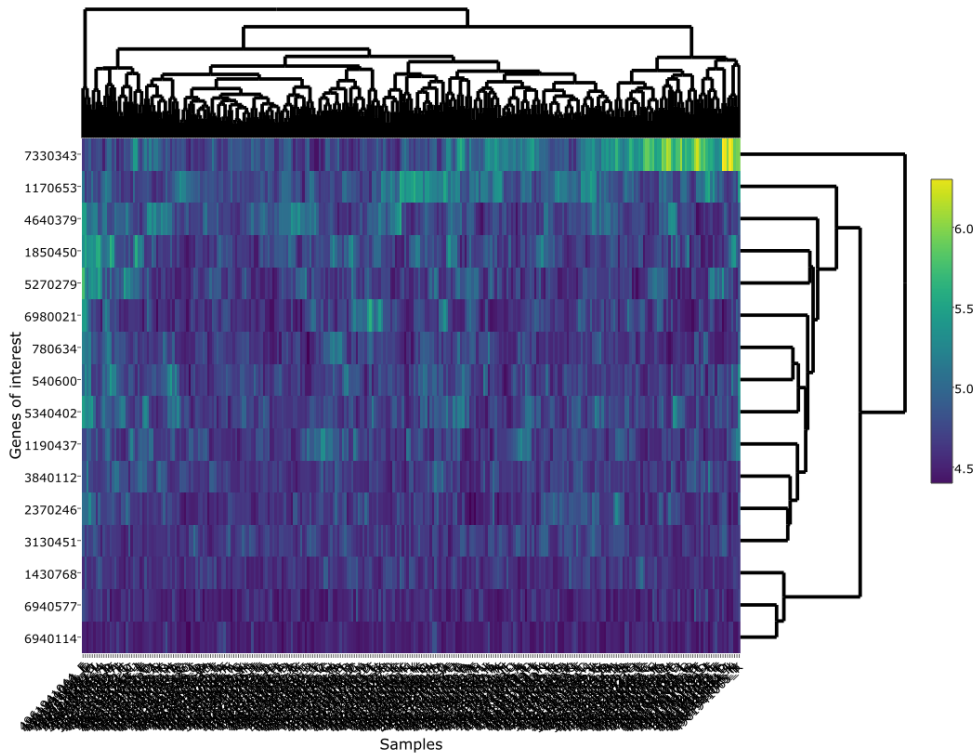
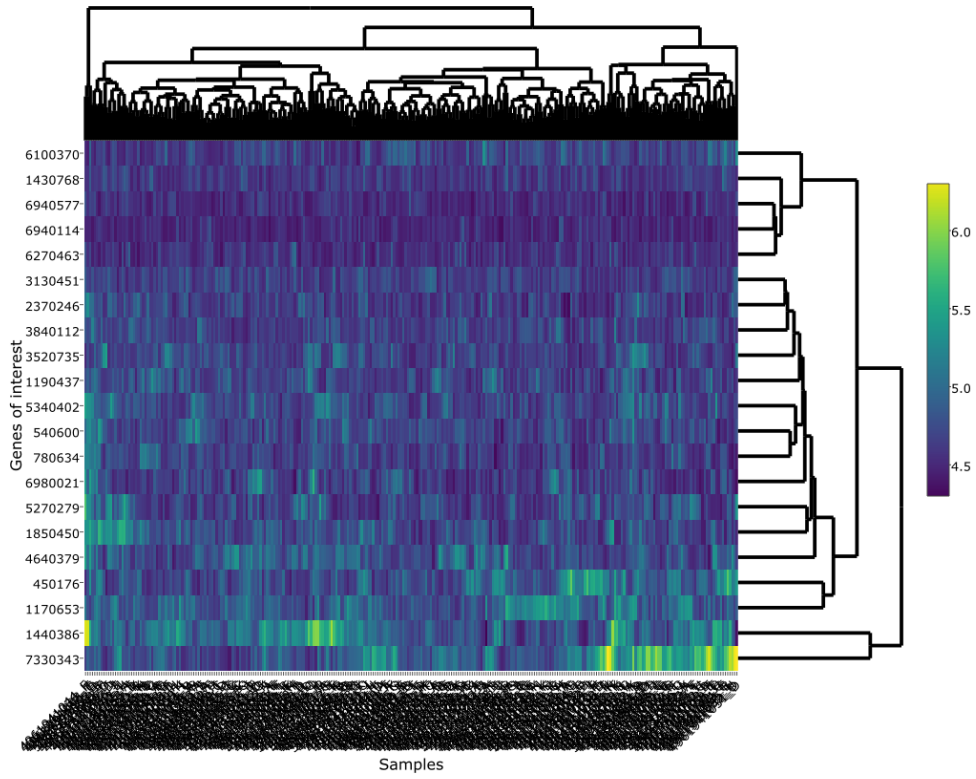
Δενδρόγραμμα με βάση το ζεύγος SERPINB2 και RBM3, το οποίο αναγνωρίστηκε ως βέλτιστο μεταξύ των διαφορετικώς εκφρασμένων μεταξύ των δύο πόλεων για το καλοκαίρι του 2009.

Η μέση τιμή της εκπροσώπησης της Πράγας στον πληθυσμό δειγμάτων του 2009 είναι ~37%. Έτσι, τιμές χαμηλότερες από αυτήν σε ένα κλάδο του δενδροδιαγράμματος δείχνουν υπερεκπροσώπηση της Οστράβα και υψηλότερες από αυτήν, υπερεκπροσώπηση της Πράγας.



Εικόνα 4.6.8

Μπορούμε να δούμε τη συμβολή των γονιδίων στο διαχωρισμό και με τη μορφή heatmaps.



Εικόνα 4.6.9

Τα δείγματα και τα γονίδια έχουν αμφότερα κατανεμηθεί με βάση ιεραρχικό clustering, εξασφαλίζοντας πως αντικείμενα με αντίστοιχη συμπεριφορά θα βρίσκονται σε κοντινούς μεταξύ τους κλάδους. Η αντιστοίχιση των αριθμητικών διευθύνσεων με τα γονίδια τα οποία χειριζόμαστε δίνεται στον πίνακα 4.5.8.

Αριθμητική διεύθυνση	Γονίδιο
450176	AMMECR1
5270279	COL9A1
5340402	COL9A1
6940114	COL9A1
780634	HGF
3130451	HGF
1440386	LTK
3520735	LZTS1
540600	MINK1
1850450	MINK1
6980021	NGFR
1430768	PAX3
2370246	PAX3
6940577	PAX3
1170653	PDGFA
3840112	PDGFA
7330343	RAPGEF1
1190437	SH3PXD2B
6100370	TLX2
6270463	TLX2
4640379	ZFYVE27

**Πίνακας 4.6.1**

Παρατηρούμε πως το μοτίβο διαχωρισμού που παρατηρήσαμε με βάση τα διαφορετικώς εκφρασμένα γονίδια SERPINB2 και RBM3 επαναλαμβάνεται στα πολύ αδρά του χαρακτηριστικά όταν χρησιμοποιήσουμε ως βάση το υποσύνολο των hub genes που έχουν προκύψει από τη σύγκριση των λειτουργικών αναλύσεων που κάναμε για την επίδραση της εποχής σε κάθε πόλη ξεχωριστά, αν και αισθητά πιο αδύναμο, όπως θα περιμέναμε.

Σε αυτό το σημείο πρέπει να σημειωθεί πως η αξία αυτής της διαδικασίας είναι πως από μια υψηλού επιπέδου πληροφορία για τα γονίδια με βιολογικά κεντρικούς ρόλους σε αρκετούς μηχανισμούς οι οποίοι ενεργοποιούνται για να προσαρμοστεί ο οργανισμός εντός μολυσμένου περιβάλλοντος, μπορέσαμε να δημιουργήσουμε ένα δείκτη ο οποίος μπορεί να ξεχωρίσει τα δείγματα με επιτυχία αντίστοιχης τάξης μεγέθους με τη χρήση των γονιδίων που εξ' αρχής εκφράζονται διαφορετικώς μεταξύ των πόλεων, οπότε και είναι και ένα κατ' εξοχήν ισχυρό κριτήριο διαχωρισμού.

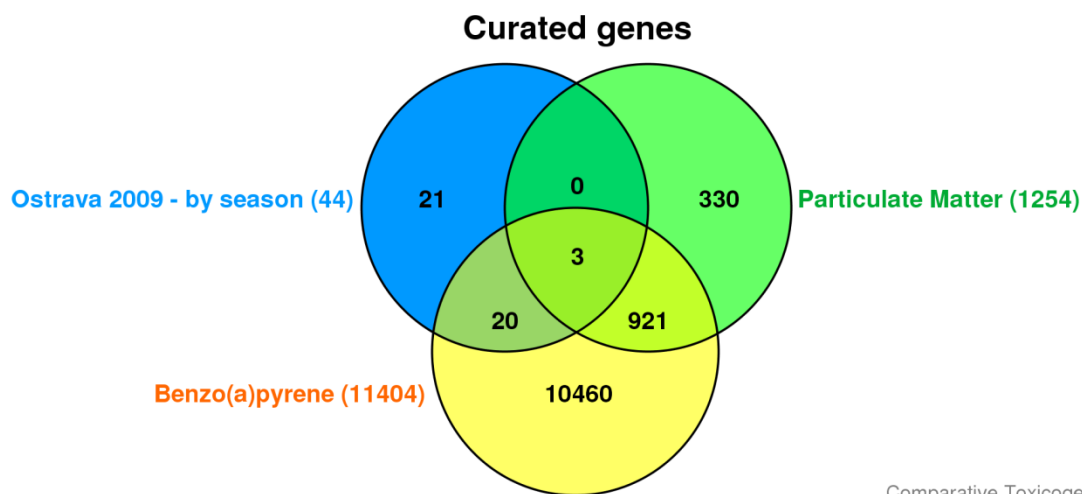
Το γεγονός δε πως οι δύο αυτές λίστες δεν αλληλεπικαλύπτονται κατά κανέναν τρόπο, ούτε συνηπήρξαν σε οποιαδήποτε στιγμή κατά τη διάρκεια της ανάλυσης, δείχνει πως η πληροφορία που έχει εξαχθεί έχει βιολογικό βάρος, παρ' όλη την ύπαρξη θορύβου και παρ' όλη τη διαδικασία αντιμετώπισής του, η οποία όπως αναλύσαμε νωρίτερα, έσβησε και μέρος του βιολογικού σήματος που θα θέλαμε να μετρήσουμε.

Τέλος, ισχυροποιεί τις επιλογές που έχουν γίνει όσον αφορά τα εργαλεία και τους τρόπους ανάλυσης το γεγονός ότι η πορεία από το ειδικό (διαφορικώς εκφρασμένα γονίδια) μπορέσαμε να περάσουμε στο γενικό (δεδομένα από τις οντολογίες που χρησιμοποιήθηκαν) και ξανά στο ειδικό (με τη διαμόρφωση μιας λίστας hub genes που έχει μετρήσιμη ικανότητα να ξεχωρίζει τα δείγματα) έχοντας μια σχετική επιτυχία.

#### 4.7 Σύγκριση με υπάρχουσα βιβλιογραφία

Σε αυτό το σημείο συγκρίναμε τις λίστες διαφορετικώς εκφρασμένων γονιδίων που αναγνωρίστηκαν στην Οστράβα με την υπάρχουσα καταγεγραμμένη γνώση που συνδέει τη ρύπανση τύπου αιωρούμενων μικροσωματιδίων και τη ρύπανση από Β(α)Ρ με τη διαφορετική έκφραση γονιδίων. Έτσι, μπορούμε να εκτιμήσουμε την αλληλεπικάλυψη των αποτελεσμάτων μας με την υπάρχουσα βιβλιογραφία, βλέποντας αν τα αποτελέσματά μας μπορούν να τοποθετηθούν εντός της υπάρχουσας εξελισσόμενης γνώσης.

Κατ' αρχάς, μπορούμε να δούμε την ποσοτική αλληλεπικάλυψη των διαφορετικώς εκφραζόμενων γονιδίων στις εικόνες 4.7.1.-4.7.3.



Εικόνα 4.7.1

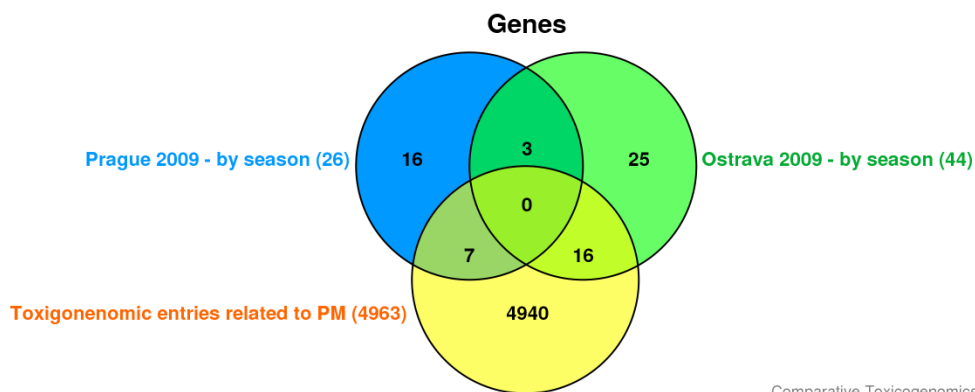
Λεπτομέρειες για το ποιά γονίδια εκφράζονται διαφορετικώς βρίσκονται στον πίνακα 4.7.1.

Πίνακας 4.7.1

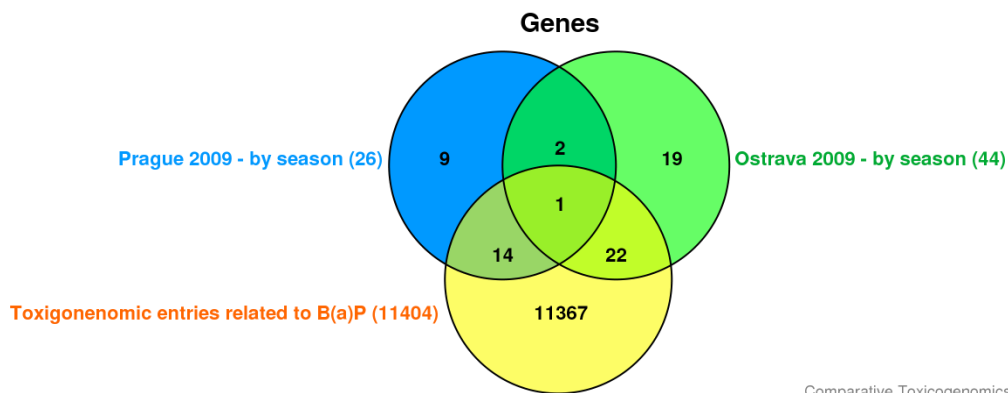
Αλληλεπικάλυψη Οστράβας – Β(α)Ρ	Αλληλεπικαλυπτόμενα Οστράβας - ΡΜ
AMMECR1	HGF
CYP4X1	NEDD9
EHD3	SERPINB2
FBXL7	
HIST1H1E	
LGI2	
MINK1	
NGFR	
PDGFA	
PGPEP1	
PSD3	
RAPGEF1	
SH3PXD2B	

SREK1	
SYNPO	
TLE6	
TMEM151A	
UBE2D3	
WVOX	
ZWINT	

Τέλος, για λόγους εποπτείας μπορούμε να δούμε και την ποσοτική αλληλεπικάλυψη μεταξύ των διαφορετικώς εκφρασμένων γονιδίων της κάθε μίας από τις δύο πόλεις, των γονιδίων συνδεδεμένων με αιωρούμενους μικροσωματιδιακούς ρύπους και των γονιδίων συνδεδεμένων με B(a)P, στα σχήματα.



**Εικόνα 4.7.2**



**Εικόνα 4.7.3**

## **Κεφάλαιο Πέμπτο: Συζήτηση επί των αποτελεσμάτων και συμπεράσματα**

### **Περί του data set**

Πρώτα απ' όλα, πρέπει να αναγνωρισθεί πως το data set που επεξεργαστήκαμε εμφάνισε αρκετές δυσκολίες, τόσο όσον αφορά την εμπλοκή του βιολογικού σήματος με το batch effect που εισήγαγε η απόδοση των beadchips (η οποία έχει αναλυθεί ήδη), όσο και όσον αφορά το χαμηλό αριθμό δειγματοληπτικών περιόδων, πράγμα που περιορίσει σημαντικά τον αριθμό συγκρίσεων που θα μπορούσαν να γίνουν, περιορίζοντας τη δυνατότητα ερμηνείας των αποτελεσμάτων. Παρ' όλα αυτά, η ποιότητά του όσον αφορά το μεγάλο αριθμό δειγμάτων, που εξασφάλιζε τη στατιστική σημαντικότητα των αποτελεσμάτων αλλά και τη μελέτη ενός όσο το δυνατόν ομοιογενούς πληθυσμού με κριτήριο την έκθεση στην περιβαλλοντική ρύπανση που δέχεται ένας πολίτης στην καθημερινή του ζωή, δε μπορεί σε καμία περίπτωση να υποτιμηθεί. Πρόκειται για έναν τομέα μελέτης ο οποίος κατά βάση δεν έχει υποβληθεί στα εργαλεία της βιοπληροφορικής, καθώς σχεδόν το σύνολο των μελετών που έχουν μέχρι στιγμής γίνει αναφέρεται είτε σε μικρού μεγέθους (συνήθως in vitro) πληθυσμούς, είτε στις επιδράσεις αποκλειστικά του καπνού, στα πλαίσια της οικονομικής περιθωριοποίησης της καπνοβιομηχανίας.

### **Περί των προσεγγίσεων για την αντιμετώπιση του batch effect**

Η διαδικασία σύγκρισης των διαφορετικών προσεγγίσεων αποτέλεσε μια καλή δοκιμή, σε πραγματικές συνθήκες ανάλυσης ενός dataset με υψηλό ποσοστό θορύβου. Στα πλαίσια αυτά, τα συμπεράσματα μπορούν να φανούν χρήσιμα όσον αφορά την προστασία του βιολογικού σήματος υπό παρόμοιες συνθήκες. Από ποσοτικής πλευράς, είναι σαφές το γεγονός ότι η προσέγγιση που συνδύαζε τον υπολογισμό συντελεστών βάρους για τα δείγματα και με τον υπολογισμό και αφαίρεση του batch effect, διατήρησε μεγαλύτερο ποσό βιολογικού σήματος σε σχέση με τις άλλες δύο προσεγγίσεις (batch correction χωρίς συντελεστές βάρους και υπολογισμό συντελεστών βάρους με το batch effect περασμένο στο γραμμικό μοντέλο, τακτική η οποία ακολουθήθηκε στο τελικό στάδιο της αρχικής έρευνας [28]).

Από ποιοτικής πλευράς, το βιολογικό σήμα που απομονώθηκε φαίνεται να αλληλεπικαλύπτεται με υπάρχουσα επιδημιολογική βιβλιογραφία, δίνοντας μία όχι μόνο στατιστική συνύπαρξη βαριάς βιομηχανικής ρύπανσης και συγκεκριμένων συνδρόμων αναπτυξιακού, ανοσολογικού και νευρολογικού χαρακτήρα, αλλά και φωτίζοντας γονίδια με ρόλο σε αυτά τα σύνδρομα τα οποία φαίνεται να αλλάζουν την έκφρασή τους παρουσία της ρύπανσης. Αυτό το γεγονός δρα ως παράγοντας ενίσχυσης της έρευνας που διενεργήθηκε, καθώς τα αποτελέσματά μας τοποθετούνται σε ένα γενικότερο επιδημιολογικό προφίλ και το ενισχύουν με βαθύτερη πληροφορία. Ταυτόχρονα όμως, ενισχύει και την ορθότητα των επιλογών που έγιναν κατά τη διάρκεια της ανάλυσης, καθώς το τελικό γονομικό προφίλ φαίνεται πως έχει πολύ μεγαλύτερη επικάλυψη με τη στοιχειοθετημένη επιδημιολογία απ' όσο θα μπορούσε ένα τελικό σήμα αποτελούμενο κυρίως από θόρυβο.





απόδοση της διαδικασίας υβριδισμού, είτε στο επίπεδο της διατήρησης και προετοιμασίας των δειγμάτων, είτε στο επίπεδο του τρόπου με τον οποίο τοποθετήθηκαν στα arrays. Ακόμα περισσότερο δε, καθώς μετά το batch effect correction τα δεδομένα διαχωρίστηκαν εντονότατα με βάση αποκλειστικά τη χρονιά δειγματοληψίας, παρ' όλου που υπήρχαν beadchips με δείγματα και από τις δύο χρονιές, στοιχείο που υποδεικνύει πως η έρευνά μας για την πηγή αυτών των διαφορών θα πρέπει να ξεκινήσει από αυτό το σημείο. Δεν έχουμε αρκετά στοιχεία για να αποφανθούμε για ο,τιδήποτε περισσότερο.

Όσον αφορά τη δική μας ανάλυση, κατ' αρχάς πρέπει να σημειωθεί το γεγονός ότι σε κανέναν από τους δύο χειμώνες στους οποίους πραγματοποιήθηκαν δειγματοληψίες δε μπορεί να αναγνωριστεί κάποιο στατιστικώς σημαντικά διαφορικώς εκφρασμένο γονίδιο. Αυτό εκτιμούμε πως οφείλεται σε δύο παράγοντες: Πρώτον, τα δεδομένα για τα επίπεδα ατμοσφαιρικής ρύπανσης δείχνουν σαφώς μικρότερες διαφορές μεταξύ των προφίλ μόλυνσεως για τις δύο πόλεις όσον αφορά τους χειμερινούς μήνες, σε σχέση με τους θερινούς, για τις περιόδους δειγματοληψίας, όπως παρουσιάζεται στον Πίνακα 6.1. Αυτά τα στοιχεία συνεπάγονται πως ένα βιολογικό σήμα προερχόμενο από προσπάθεια προσαρμογής του οργανισμού στις συνθήκες ρύπανσης θα είναι εντονότερο τους θερινούς μήνες, όπου αυτές οι διαφορές είναι εντονότερες. Δεύτερον και καθόλου λιγότερο σημαντικό, για τους λόγους που αναφέρθηκαν παραπάνω, η ίδια η διαδικασία του batch correction μείωσε την ένταση του βιολογικού σήματος που είχαμε τη δυνατότητα να μετρήσουμε. Ως εκ τούτου, μένει ανοιχτό και ιδιαίτερα πιθανό το ενδεχόμενο με νέες μετρήσεις και ορθότερο πρωτόκολλο να μπορούμε να απομονώσουμε ένα αισθητό βιολογικό σήμα και μεταξύ δειγμάτων που έχουν ληφθεί το χειμώνα, όσο και να εξάγουμε μια πολύ ακριβέστερη και πληρέστερη εικόνα του διαφορικού προφίλ εστιάζοντας σε θερινές περιόδους.

### **Περί της ερμηνείας των αποτελεσμάτων οντολογιών και hub genes**

Κατά τη σύγκριση των οντολογικών αποτελεσμάτων Πράγας και Οστράβας, πρέπει να έχουμε υπ' όψιν το γεγονός ότι τα enrichment scores στις οντολογίες που χρησιμοποιήθηκαν ήταν χαμηλά, γεγονός που υπαγορεύεται άλλωστε από το αποδυναμωμένο βιολογικό σήμα το οποίο μπορέσαμε να εξάγουμε. Ως εκ τούτου, τα συμπεράσματα που μπορούμε να εξάγουμε από αυτές έχουν το χαρακτήρα περισσότερο σημείου εκκίνησης για τη διατύπωση ερωτημάτων νέων ερευνών. Παρ' όλα αυτά, αξίζει να δούμε εν συντομία μερικά από τα πιο ενδιαφέροντα ευρήματα.

Κατ' αρχάς, παρατηρούμε πως στην Οστράβα εμφανίζονται διαφορικώς εκφρασμένοι όροι περί της επούλωσης τραυμάτων, αποπτώσεων και της μιτωτικής διαίρεσης, με το τελευταίο μάλιστα να καθοδηγείται από υπερεκφραζόμενα γονίδια. Θα μπορούσε να διερευνηθεί αν πρόκειται για μέρος της προσπάθειας του οργανισμού να επουλώσει τις βλάβες που προκαλούνται από την εντονότερη έκθεση σε ρύπους, ειδικά μικροσωματιδιακού χαρακτήρα, οι οποίοι προκαλούν βλάβες σε μια σειρά επίπεδα [13] [14] [15]. Επίσης σημαντική αρχή μπορεί να θεωρηθεί η εκπροσώπηση του οντολογικού όρου «platelet-derived growth factor receptor signaling pathway», καθώς υπάρχει βιβλιογραφία που συνδέει την ενεργοποίηση αιμοπεταλίων με την ατμοσφαιρική ρύπανση [32]. Αντιστοίχως σημαντικό εύρημα είναι η παρατήρηση πλείστων όρων αναπτυξιακού χαρακτήρα στη Phenotype Ontology, μεταξύ των οποίων και η infantile encephalopathy.

Ένα επίσης σημαντικό αποτέλεσμα της έρευνας που κάναμε είναι η δημιουργία, μέσω των hub genes που αναγνωρίστηκαν από τις οντολογίες που χρησιμοποιήσαμε, μίας λίστας γονιδίων τα οποία μπορούν να επιτύχουν έναν σχετικό διαχωρισμό του πληθυσμού ανάλογα με τις συνθήκες ρύπανσης, σε βαθμό συγκρίσιμο, αν και φυσικά χαμηλότερο από τον αντίστοιχο που μπορεί να επιτευχθεί με τη χρήση των διαφορετικώς εκφρασμένων γονιδίων. Μπορούμε να θεωρήσουμε, και είναι ζήτημα μελλοντικής μελέτης, πως τα δείγματα της Οστράβα που εμφανίζονται σε περιοχές δειγμάτων της Πράγας έχουν μια παρόμοια έκφραση, υποδεικνύοντας χαμηλή ευπάθεια αυτών στη ρύπανση. Αντιθέτως, η παρουσία δειγμάτων της Πράγας σε περιοχές της Οστράβα θα μπορούσε να αποδοθεί σε μεγαλύτερη ευπάθεια αυτών.

Στον παρακάτω πίνακα εκθέεται η συσχέτιση μεταξύ των genes της λίστας και του ρόλου τους στις οντολογίες που χρησιμοποιήθηκαν:

**Πίνακας 5.1**

	Gene Ontology	Hyman Phenotype	MGI Mammalian Phenotype	Reactome Pathways
PDGFA	platelet derived growth factor subunit A	-	platelet derived growth factor subunit A	-
HGF	hepatocyte growth factor	-	-	-
NGFR	nerve growth factor receptor	-	nerve growth factor receptor	nerve growth factor receptor
ZFYVE27	zinc finger FYVE-type containing 27	-	-	-
MINK1	misshapen like kinase 1	-	-	-
RAPGEF1	Rap guanine nucleotide exchange factor 1	-	-	-
PAX3	-	paired box 3	paired box 3	-
SH3PXD2B	-	SH3 and PX domains 2B	SH3 and PX domains 2B	-
COL9A1	-	collagen type IX alpha 1	-	-

### Σύγκριση με την υπάρχουσα βιβλιογραφία

Μπορούμε να παρατηρήσουμε μια υπαρκτή αλληλεπικάλυψη με την υπάρχουσα βιβλιογραφία, η οποία ενισχύει τα αποτελέσματά μας. Επίσης, γονίδια τα οποία αναδείχθηκαν μέσω των οντολογιών (AMMECR1, HGF, MINK1, NGFR, PDGFA, RAPGEF1, SH3PXD2B) ως έχοντα πιθανώς ρυθμιστικό ρόλο, εμφανίζονται ως συνδεόμενα με της λειτουργίες που διέπει κυρίως η ρύπανση τύπου B(a)P. Αυτό το γεγονός αναδεικνύει μια άμεση σχέση της ομοιόστασης υγιούς ανθρώπινου οργανισμού υπό συνθήκες ρύπανσης με την επίδραση ισχυρών ρύπων και την επαγωγή συστημικής απόκρισης, επιτρέποντας να αναγνωρίσουμε το αποτύπωμά τους στα δεδομένα μας. Τέλος, η αλληλεπικάλυψη με τα υποδεικνυόμενα από τους ρύπους γονίδια δείχνει απ' ευθείας τη βαρύτερη ρύπανση της Οστράβας.

## Κεφάλαιο Έβδομο: Βιβλιογραφία

- [1] Barbosa-Morais, N. L.; Dunning, M. J.; Samarajiwa, S. A.; Darot, J. F. J.; Ritchie, M. E.; Lynch, A. G.; Tavaré, S. A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Research*. 38 (3): e17–e17 (2009).
- [2] Shi, W, Oshlack, A, and Smyth, GK. Optimizing the noise versus bias trade-off for Illumina Whole Genome Expression BeadChips. *Nucleic Acids Research* 38, e204 (2010).
- [3] Xie Y, Wang X, Story M. Statistical methods of background correction for Illumina beadarray data. *Bioinformatics* 25:751-757 (2009).
- [4] Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 19 (2): 185–193 (2003).
- [5] Mark Dunning, Andy Lynch and Matthew Eldridge. illuminaHumanv3.db: Illumina HumanHT12v3 annotation data (chip illuminaHumanv3).
- [6] Ritchie, M. E., Diyagama, D., Neilson, van Laar, R., J., Dobrovic, A., Holloway, A., and Smyth, G. K. Empirical array quality weights in the analysis of microarray data. *BMC Bioinformatics* 7, 261 (2006).
- [7] Jeffrey T. Leek, Robert B. Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W. Evan Johnson, Donald Geman, Keith Baggerly & Rafael A. Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics* 11, 733-739 (2010).
- [8] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8 (1), 118-127 (2007).
- [9] Leek JT, Storey JD. A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences* , 105: 18718-18723 (2008).
- [10] Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by Surrogate Variable Analysis. *PLoS Genetics*, 3: e161 (2007).
- [11] Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 3 (2004).
- [12] Koutsandreas Theodoros, Ilona Binenbaum, Eleftherios Pilalis, Ioannis Valavanis, Olga Papadodima and Aristotelis Chatziioannou. Analyzing and Visualizing Genomic Complexity for the Derivation of the Emergent Molecular Networks. *IJMSTR* 4.2: 30-49 (2016).
- [13] B. Brunekreef, S.T. Holgate. Air pollution and health. *Lancet* 360 1233–1242 (2002).
- [14] A. Valavanidis, K. Fiotakis, T. Vlachogianni. Airborne particulate matter and human health: toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* 26 339–362 (2008).
- [15] G. Oberdorster, E. Oberdorster, J. Oberdorster. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. *Environ. Health Perspect.* 113 823–839 (2005).
- [16] IARC, IARC monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, in: Monograph on Some Non-Heterocyclic Polycyclic Aromatic Hydrocarbons and Some Related Exposures. IARC, Publications, Lyon, France (2010).

- [17] IARC, IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Chemical Agents and Related Occupations. IARC, Publications, Lyon, France (2012).
- [18] P. Gerde, B.A. Muggenburg, M. Lundborg, A.R. Dahl. The rapid alveolar absorption of diesel soot-adsorbed benzo[a]pyrene: bioavailability, metabolism and dosimetry of an inhaled particle-borne carcinogen. *Carcinogenesis* 22 741–749 (2001).
- [19] Xu X, Ha SU, Basnet R. A Review of Epidemiological Research on Adverse Neurological Effects of Exposure to Ambient Air Pollution. *Frontiers in Public Health* 4:157 (2016).
- [20] Calderón-Garcidueñas L, Mora-Tiscareño A, Ontiveros E, Gómez-Garza G, Barragán-Mejía G, Broadway J, Chapman S, Valencia-Salazar G, Jewells V, Maronpot RR, Henríquez-Roldán C, Pérez-Guillé B, Torres-Jardón R, Herriot L, Brooks D, Osnaya-Brizuela N, Monroy ME, González-Maciel A, Reynoso-Robles R, Villarreal-Calderon R, Solt AC, Engle RW. Air pollution, cognitive deficits and brain abnormalities: a pilot study with children and dogs. *Brain Cogn* 68(2):117-27 (2008).
- [21] Calderón-Garcidueñas L, Engle R, Mora-Tiscareño A, Styner M, Gómez-Garza G, Zhu H, Jewells V, Torres-Jardón R, Romero L, Monroy-Acosta ME, Bryant C, González-González LO, Medina-Cortina H, D'Angiulli A. Exposure to severe urban air pollution influences cognitive outcomes, brain volume and systemic inflammation in clinically healthy children. *Brain Cogn*. 77(3):345-55 (2011).
- [22] Luke Curtis, William Rea, Patricia Smith-Willis, Ervin Fenyves, Yaqin Pan. Adverse health effects of outdoor air pollutants. *Environment International*, Volume 32, Issue 6 Pages 815-830 (2006).
- [23] E. Mantas, E. Remoundaki, I. Halari, P. Kassomenos, C. Theodosi, A. Hatzikioseyan, N. Mihalopoulos. Mass closure and source apportionment of PM<sub>2.5</sub> by Positive Matrix Factorization analysis in urban Mediterranean environment. *Atmospheric Environment*, Volume 94,, Pages 154-163 (2014).
- [24] E. Remoundaki, A. Papayannis, P. Kassomenos, E. Mantas, P. Kokkalis, M. Tsezos. Influence of Saharan Dust Transport Events on PM<sub>2.5</sub> Concentrations and Composition over Athens. *Water, Air, & Soil Pollution* 224:1373 (2013).
- [25] C. Theodosi, G. Grivas, P. Zarmpas, A. Chaloulakou, and N. Mihalopoulos. Mass and chemical composition of size-segregated aerosols (PM<sub>1</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>) over Athens, Greece: local versus regional sources. *Atmos. Chem. Phys.*, 11895-11911, (2011).
- [26] Maria Kanakidou, Nikolaos Mihalopoulos, Tayfun Kindap, Ulas Im, Mihalis Vrekoussis, Evangelos Gerasopoulos, Eirini Dermizaki, Alper Unal, Mustafa Koçak, Kostas Markakis, Dimitris Melas, Georgios Kouvarakis, Ahmed F. Youssef, Andreas Richter, Nikolaos Hatzianastassiou, Andreas Hilboll, Felix Ebojje, Folkard Wittrock, Christian von Savigny, John P. Burrows, Annette Ladstaetter-Weissenmayer, Hani Moubasher. Megacities as hot spots of air pollution in the East Mediterranean. *Atmospheric Environment*, Volume 45, Issue 6, Pages 1223-1235 (2011).
- [27] S. Fuzzi, U. Baltensperger, K. Carslaw, S. Decesari, H. Denier van der Gon, M. C. Facchini, D. Fowler, I. Koren, B. Langford, U. Lohmann, E. Nemitz, S. Pandis, I. Riipinen, Y. Rudich, M. Schaap, J. G. Slowik, D. V. Spracklen, E. Vignati, M. Wild, M. Williams, and S. Gilardoni. Particulate matter, air quality and climate: lessons learned and future needs. *Atmos. Chem. Phys.*, 8217-8299 (2015).
- [28] Rossner P Jr, Tulupova E, Rossnerova A, Libalova H, Honkova K, Gmuender H, Pastorkova A, Svecova V, Topinka J, Sram RJ. Reduced gene expression levels after chronic exposure to high concentrations of air pollutants. *Mutat Res*. 780:60-70 (2015).

- [29] Rossner P Jr, Svecova V, Schmuczerova J, Milcova A, Tabashidze N, Topinka J, Pastorkova A, Sram RJ. Analysis of biomarkers in a Czech population exposed to heavy air pollution. Part I: bulky DNA adducts. *Mutagenesis* 28(1):89-95 (2013).
- [30] Rossner P Jr, Rossnerova A, Spatova M, Beskid O, Uhlirova K, Libalova H, Solansky I, Topinka J, Sram RJ. Analysis of biomarkers in a Czech population exposed to heavy air pollution. Part II: chromosomal aberrations and oxidative stress. *Mutagenesis* 28(1):97-106 (2013).
- [31] [http://www.genenames.org/cgi-bin/symbol\\_checker](http://www.genenames.org/cgi-bin/symbol_checker)
- [32] Frampton MW<sup>1</sup>, Bausch J, Chalupa D, Hopke PK, Little EL, Oakes D, Stewart JC, Utell MJ. Effects of outdoor air pollutants on platelet activation in people with type 2 diabetes. *Inhal Toxicol* 24(12):831-8. (2012).

**Παράρτημα: Ο κώδικας βάση του οποίου πραγματοποιήθηκε η ανάλυση**

```

source("http://bioconductor.org/biocLite.R")
biocLite("GEOquery")
biocLite("limma")
biocLite("illuminaHumanv3.db")
biocLite("sva")
biocLite("d3heatmap")
biocLite("mvtnorm")
biocLite("heatmaply")

library(GEOquery)
library(limma)
library(illuminaHumanv3.db)
library(sva)
library(d3heatmap)
library(heatmaply)

#point to your favorite directory
directory <- "C:/Users/gr/Desktop/Final run"

setwd(directory)

#getting the data
file_list <- getGEOSuppFiles('GSE60767')
czech_data_expression_set <- getGEO("GSE60767",GSEMatrix=TRUE)
gunzip("GSE60767/GSE60767_raw.txt.gz",overwrite=TRUE,remove=FALSE)

raw_data <- read.ilmn(files="GSE60767/GSE60767_raw.txt",probeid="ProbeID",annotation="TargetID",other.columns=c("Detection","Avg_NBEADS"))

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Initial data.RData")

normalized_czech_data <- neqc(raw_data)
dim(normalized_czech_data)

par(mfrow=c(2,1))
boxplot(log2(raw_data$E),range=0,ylab=expression(log[3](intensity)),las=2,xlab="",main="log2-
transformed intensity of raw data")
boxplot(normalized_czech_data$E,range=0,ylab=expression(log[3](intensity)),las=2,xlab="",main="log2-
transformed intensity of NEQC normalized data")
par(mfrow=c(1,1))

```



```

plotMDS(normalized_czech_data$E,main="Raw data")

message("Removing technical outliers 4613710042_F and 4609518292_F")
raw_data <-
raw_data[,colnames(raw_data@.Data[[3]])!="4613710042_F"&colnames(raw_data@.Data[[3]])!="4609
518292_F"]
normalized_czech_data <- neqc(raw_data)
dim(normalized_czech_data)
boxplot(normalized_czech_data$E,range=0,ylab=expression(log[3](intensity)),las=2,xlab="",main="Inten
sity of outlier-free & NEQC normalized data")
plotMDS(normalized_czech_data$E,main="Normalized data without outliers")

#labelling
sample_origin <-
as.character(czech_data_expression_set$GSE60767_series_matrix.txt.gz@phenoData@data$title)
for (i in 1:length(sample_origin)){
  for (j in
1:length(as.character(czech_data_expression_set$GSE60767_series_matrix.txt.gz@phenoData@data$ti
tle))){
    if
(strsplit(as.character(czech_data_expression_set$GSE60767_series_matrix.txt.gz@phenoData@data$tit
le)[[j]],split=" ")[[2]][5]==colnames(normalized_czech_data@.Data[[3]])[i]){
      sample_origin[i] <-
as.character(czech_data_expression_set$GSE60767_series_matrix.txt.gz@phenoData@data$title)[[j]]
    }
  }
}
labelling <- strsplit(sample_origin,split=" ")
coloring <- vector(length=length(labelling))
Origin=c("Prague","Ostrava","winter","summer","2009","2010")
for (i in 1:length(labelling)){
  labelling[[i]] <-
paste(strsplit(labelling[[i]][2],split=" ")[[2]][2],strsplit(labelling[[i]][3],split=" ")[[2]][2],paste(strsplit(labelli
ng[[i]][4],split=" ")[[2]][4],strsplit(labelling[[i]][4],split=" ")[[2]][5],sep=""),sep="-")
  coloring[[i]] <-
c("darkblue","darkred","darkturquoise","red","lightblue","deeppink")[labelling[[i]]==c("P-w-09","O-w-
09","P-s-09","O-s-09","P-w-10","O-w-10")]
}

#chips
chip_id <- character(length=length(sample_origin))
for (i in 1:length(chip_id)){

```

```

chip_id[i] <- strsplit(strsplit(sample_origin,split=" ")[[i]][5],split="_")[[2]][2]
}
chip <- integer(length=length(sample_origin))
for (i in 1:length(chip)){
  for (j in 1:length(levels(as.factor(chip_id)))){
    if (chip_id[i]==levels(as.factor(chip_id))[j]){
      chip[i] <- j
    }
  }
}

#data overview
samples_overview_matrix <- matrix(nrow=dim(normalized_czech_data)[3],ncol=6)
ProbeID_sample=c(colnames(normalized_czech_data))
dimnames(samples_overview_matrix) <- list(ProbeID_sample,Origin)
for (i in 1:dim(normalized_czech_data)[3]){
  for (k in 1:6){
    samples_overview_matrix[i,k] <- 0
    for (j in 1:4){
      if (Origin[k] == strsplit(sample_origin,split=" ")[[i]][j]){
        samples_overview_matrix[[i,k]] <- 1
      }
    }
  }
}

#design matrix
design_matrix <- matrix(nrow=dim(normalized_czech_data)[3],ncol=4)
#city_and_time=c("(Intercept)","Prague","winter","y2010")
city_and_time=c("Intercept","City","season","year")
dimnames(design_matrix) <- list(ProbeID_sample,city_and_time)
for (i in 1:dim(normalized_czech_data)[3]){
  design_matrix[[i,1]] <- 1
  for (j in 2:4){
    design_matrix[[i,j]] <- 0
    if (j==2&samples_overview_matrix[[i,2]]==1){
      design_matrix[[i,j]] <- 1
    }
    if (j==3&samples_overview_matrix[[i,3]]==1){
      design_matrix[[i,j]] <- 1
    }
    if (j==4&samples_overview_matrix[[i,6]]==1){

```

```

    design_matrix[[i,j]] <- 1
  }
}

save.image("Initial data [full].RData")

setwd(paste(directory,"Gene symbol update",sep="/"))
write.table(file="Gene symbol update
query.txt",normalized_czech_data$genes$TargetID,quote=FALSE,row.names=FALSE,col.names="Gene
Symbol")
setwd(paste(directory,"Workspaces",sep="/"))

plotMDS(normalized_czech_data$E,labels=labelling,main="Initial data - labeled")
plotMDS(normalized_czech_data$E,pch=19,col=coloring,main="Initial data, color-coded")
for (i in 1:2){
  for (j in 1:3){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark blue","Dark
red","Turquoise","Red","Pale blue","Pink")[[i+2*(j-1)]])),sep=" ")
  }
}

message("Testing probe filtering options")
graph <- matrix(nrow=20,ncol=3)
for (i in 1:20){
  graph[i,1] <- i
  normalized_and_filtered_data <- normalized_czech_data
  normalized_and_filtered_data <-
normalized_and_filtered_data[rowSums(normalized_and_filtered_data$other$Detection<0.05)>=i,]
  graph[i,2] <- dim(normalized_and_filtered_data)[[2]]

  ids <- as.character(rownames(normalized_and_filtered_data))
  ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
  quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
  normalized_and_filtered_data <- normalized_and_filtered_data[quality_value!="No
match"&quality_value!="Bad",]
  graph[i,3] <- dim(normalized_and_filtered_data)[[2]]

  plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main=paste("Cutoff at",i,"reliable
signals"))
}

```

```

plot(graph[,1],graph[,2],main="Number-of-reliable-signals filtering only",xlab="Reliable values
required",ylab="Probes left to work with",type="l")
plot(graph[,1],graph[,3],main="Number-of-reliable-signals and annotation-based
filtering",xlab="Reliable values required",ylab="Probes left to work with",type="l")

normalized_and_filtered_data <- normalized_czech_data
dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Initial")

#number-of-reliable-signals filtering
normalized_and_filtered_data <-
normalized_and_filtered_data[rowSums(normalized_and_filtered_data$other$Detection<0.05)>=5,]
dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Filtered by number-of reliable-
signals")

#annotation-based filtering
normalized_and_filtered_data <- normalized_czech_data
ids <- as.character(rownames(normalized_and_filtered_data))
ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
table(quality_value)
normalized_and_filtered_data <- normalized_and_filtered_data[quality_value!="No
match"&quality_value!="Bad",]
dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Filtered by annotation")

#Probe filtering
message("Probe filtering both by number-of-reliable-values and by annotation")
normalized_and_filtered_data <- normalized_czech_data
dim(normalized_and_filtered_data)

#number-of-reliable-values filtering
normalized_and_filtered_data <-
normalized_and_filtered_data[rowSums(normalized_and_filtered_data$other$Detection<0.05)>=5,]
dim(normalized_and_filtered_data)

#annotation-based filtering
ids <- as.character(rownames(normalized_and_filtered_data))
ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
table(quality_value)

```

```

data <- normalized_and_filtered_data[quality_value!="No match"&quality_value!="Bad",]
dim(data)
rm(normalized_and_filtered_data)

save.image("Filtered data [full].RData")

plotMDS(data$E,pch=19,col=coloring,main="Filtered by both methods")
plotMDS(data$E,labels=chip,col=coloring,main="Filtered by both methods - possible chip-induced batch
effect")
interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of full data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect")

plotMDS(removeBatchEffect(data,batch=chip,design=design_matrix),pch=19,col=coloring,main="Batch-
corrected with 'removeBatchEffect' [limma] with design matrix")
plotMDS(removeBatchEffect(data,batch=chip),pch=19,col=coloring,main="Batch-corrected          with
'removeBatchEffect' [limma] without design matrix")

#Unweighted batch effect correction
message("Batch-correcting full unweighted data")
batch_corrected_data_non_parametric                                     <-
ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=FALSE,prior.plots=TRUE)
data$E <- ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=TRUE,prior.plots=TRUE)
par(mfrow=c(1,1))
save.image("Unweighted batch-corrected data [full].RData")

#Array performance weighting & batch correction
message("Weighting and batch-correcting full data")
rm(raw_data,normalized_czech_data,batch_corrected_data_non_parametric)
load("Filtered data [full].RData")
array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights,main="Full data set", xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)
batch_corrected_data_non_parametric                                     <-
ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=FALSE,prior.plots=TRUE)
data$E <- ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=TRUE,prior.plots=TRUE)
par(mfrow=c(1,1))
save.image("Weighted & batch-corrected data [full].RData")

#Data butchering

```

```

rm(raw_data,normalized_czech_data,array_weights,batch_corrected_data_non_parametric)
load("Initial data [full].RData")
butcher <- logical(length=length(sample_origin))
for (i in 1:length(butcher)){
  butcher[i] <- strsplit(sample_origin,split=" ")[[i]][4]=="2009"
}

coloring <- coloring[butcher]
chip <- chip[butcher]
chip_id <- chip_id[butcher]
labelling <- labelling[butcher]

dim(design_matrix)
design_matrix <- design_matrix[butcher,1:3]
dim(design_matrix)

normalized_czech_data <- neqc(raw_data[,butcher])
normalized_and_filtered_data <- normalized_czech_data
dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Initial")

message("Testing probe filtering options")
graph <- matrix(nrow=20,ncol=3)
for (i in 1:20){
  graph[i,1] <- i
  normalized_and_filtered_data <- normalized_czech_data
  normalized_and_filtered_data <- normalized_and_filtered_data[
    rowSums(normalized_and_filtered_data$other$Detection<0.05)>=i,]
  graph[i,2] <- dim(normalized_and_filtered_data)[[2]]

  ids <- as.character(rownames(normalized_and_filtered_data))
  ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
  quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
  normalized_and_filtered_data <- normalized_and_filtered_data[quality_value!="No
  match"&quality_value!="Bad",]
  graph[i,3] <- dim(normalized_and_filtered_data)[[2]]

  plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main=paste("Cutoff",i,"reliable
  signals"))
}
plot(graph[,1],graph[,2],main="Number-of-reliable-signals filtering only",xlab="Reliable values
  required",ylab="Probes left to work with",type="l")

```

```
plot(graph[,1],graph[,3],main="Number-of-reliable-signals and annotation-based filtering",xlab="Reliable values required",ylab="Probes left to work with",type="l")
```

```
#number-of-reliable-signals filtering
```

```
plotMDS(normalized_czech_data$E,pch=19,col=coloring,main="Initial")
```

```
normalized_and_filtered_data <-
```

```
normalized_czech_data[rowSums(normalized_czech_data$other$Detection<0.05)>=3,]
```

```
dim(normalized_and_filtered_data)
```

```
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Filtered by number-of reliable-signals")
```

```
#annotation-based filtering
```

```
normalized_and_filtered_data <- normalized_czech_data
```

```
ids <- as.character(rownames(normalized_and_filtered_data))
```

```
ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
```

```
quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
```

```
table(quality_value)
```

```
normalized_and_filtered_data <- normalized_and_filtered_data[quality_value!="No match"&quality_value!="Bad",]
```

```
dim(normalized_and_filtered_data)
```

```
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Filtered by annotation")
```

```
#Probe filtering
```

```
message("Probe filtering both by number-of-reliable-values and by annotation")
```

```
normalized_and_filtered_data <- normalized_czech_data
```

```
dim(normalized_and_filtered_data)
```

```
#number-of-reliable-values filtering
```

```
normalized_and_filtered_data <-
```

```
normalized_and_filtered_data[rowSums(normalized_and_filtered_data$other$Detection<0.05)>=3,]
```

```
dim(normalized_and_filtered_data)
```

```
#annotation-based filtering
```

```
ids <- as.character(rownames(normalized_and_filtered_data))
```

```
ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
```

```
quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
```

```
table(quality_value)
```

```
data <- normalized_and_filtered_data[quality_value!="No match"&quality_value!="Bad",]
```

```
dim(data)
```

```
rm(normalized_and_filtered_data)
```

```
plotMDS(data$E,pch=19,col=coloring,main="Filtered by both methods [2009 data]")
```

```

save.image("Filtered data [2009].RData")

#Unweighted batch effect correction
message("Batch-correcting 2009 unweighted data")
batch_corrected_data_non_parametric <-
ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=FALSE,prior.plots=TRUE)
data$E <- ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=TRUE,prior.plots=TRUE)
par(mfrow=c(1,1))
save.image("Unweighted batch-corrected data [2009].RData")

#Array performance weighting & batch correction
message("Weighting and batch-correcting 2009 data")
rm(raw_data,normalized_czech_data,batch_corrected_data_non_parametric)
load("Filtered data [2009].RData")
array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights,main="2009 data set", xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)
batch_corrected_data_non_parametric <-
ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=FALSE,prior.plots=TRUE)
data$E <- ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=TRUE,prior.plots=TRUE)
par(mfrow=c(1,1))
save.image("Weighted & batch-corrected data [2009].RData")

rm(raw_data,normalized_czech_data,array_weights,batch_corrected_data_non_parametric)
load("Initial data [full].RData")
butcher <- logical(length=length(sample_origin))
for (i in 1:length(butcher)){
  butcher[i] <- strsplit(sample_origin,split=" ")[[i]][4]=="2010"
}

coloring <- coloring[butcher]
chip <- chip[butcher]
chip_id <- chip_id[butcher]
labelling <- labelling[butcher]

dim(design_matrix)
design_matrix <- design_matrix[butcher,1:2]
dim(design_matrix)

normalized_czech_data <- neqc(raw_data[,butcher])
normalized_and_filtered_data <- normalized_czech_data

```



```

dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Initial")

message("Testing probe filtering options")
graph <- matrix(nrow=20,ncol=3)
for (i in 1:20){
  graph[i,1] <- i
  normalized_and_filtered_data <- normalized_czech_data
  normalized_and_filtered_data
  normalized_and_filtered_data[rowSums(normalized_and_filtered_data$other$Detection<0.05)>=i,]
  graph[i,2] <- dim(normalized_and_filtered_data)[[2]]

  ids <- as.character(rownames(normalized_and_filtered_data))
  ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
  quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
  normalized_and_filtered_data <- normalized_and_filtered_data[quality_value!="No
match"&quality_value!="Bad",]
  graph[i,3] <- dim(normalized_and_filtered_data)[[2]]

  plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main=paste("Cutoff      at",i,"reliable
signals"))
}
plot(graph[,1],graph[,2],main="Number-of-reliable-signals      filtering      only",xlab="Reliable      values
required",ylab="Probes left to work with",type="l")
plot(graph[,1],graph[,3],main="Number-of-reliable-signals      and      annotation-based
filtering",xlab="Reliable values required",ylab="Probes left to work with",type="l")

#number-of-reliable-signals filtering
plotMDS(normalized_czech_data$E,pch=19,col=coloring,main="Initial")
normalized_and_filtered_data
normalized_czech_data[rowSums(normalized_czech_data$other$Detection<0.05)>=3,]
dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Filtered by number-of reliable-
signals")

#annotation-based filtering
normalized_and_filtered_data <- normalized_czech_data
ids <- as.character(rownames(normalized_and_filtered_data))
ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
table(quality_value)

```

```

normalized_and_filtered_data      <-      normalized_and_filtered_data[quality_value!="No
match"&quality_value!="Bad",]
dim(normalized_and_filtered_data)
plotMDS(normalized_and_filtered_data$E,pch=19,col=coloring,main="Filtered by annotation")

#Probe filtering
message("Probe filtering both by number-of-reliable-values and by annotation")
normalized_and_filtered_data <- normalized_czech_data
dim(normalized_and_filtered_data)

#number-of-reliable-values filtering
normalized_and_filtered_data      <-
normalized_and_filtered_data[rowSums(normalized_and_filtered_data$other$Detection<0.05)>=3,]
dim(normalized_and_filtered_data)

#annotation-based filtering
ids <- as.character(rownames(normalized_and_filtered_data))
ids_from_annotation <- unlist(mget(ids,revmap(illuminaHumanv3ARRAYADDRESS),ifnotfound=NA))
quality_value <- unlist(mget(ids_from_annotation,illuminaHumanv3PROBEQUALITY,ifnotfound=NA))
table(quality_value)
data <- normalized_and_filtered_data[quality_value!="No match"&quality_value!="Bad",]
dim(data)
rm(normalized_and_filtered_data)

plotMDS(data$E,pch=19,col=coloring,main="Filtered by both methods [2010 data]")

save.image("Filtered data [2010].RData")

#Unweighted batch effect correction
message("Batch-correcting 2010 unweighted data")
dim(data$E)
message("Removing confounding sample 4961941011_D")
design_matrix <- design_matrix[colnames(data$E)!="4961941011_D",]
chip <- chip[colnames(data$E)!="4961941011_D"]
chip_id <- chip_id[colnames(data$E)!="4961941011_D"]
coloring <- coloring[colnames(data$E)!="4961941011_D"]
labelling <- labelling[colnames(data$E)!="4961941011_D"]
data <- data[,colnames(data$E)!="4961941011_D"]
dim(data$E)
batch_corrected_data_non_parametric      <-
ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=FALSE,prior.plots=TRUE)
data$E <- ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=TRUE,prior.plots=TRUE)

```

```

par(mfrow=c(1,1))
save.image("Unweighted batch-corrected data [2010].RData")

#Array performance weighting & batch correction
message("Weighting and batch-correcting 2010 data")
rm(raw_data,normalized_czech_data,batch_corrected_data_non_parametric)
load("Filtered data [2010].RData")
dim(data$E)
message("Removing confounding sample 4961941011_D")
design_matrix <- design_matrix[colnames(data$E)!="4961941011_D",]
chip <- chip[colnames(data$E)!="4961941011_D"]
chip_id <- chip_id[colnames(data$E)!="4961941011_D"]
coloring <- coloring[colnames(data$E)!="4961941011_D"]
labelling <- labelling[colnames(data$E)!="4961941011_D"]
data <- data[,colnames(data$E)!="4961941011_D"]
dim(data$E)
array_weights <- arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights,main="2010 data set", xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)
batch_corrected_data_non_parametric <-
ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=FALSE,prior.plots=TRUE)
data$E <- ComBat(dat=data$E,batch=chip,mod=design_matrix,par.prior=TRUE,prior.plots=TRUE)
par(mfrow=c(1,1))
save.image("Weighted & batch-corrected data [2010].RData")

message("Moving on to data analysis")

gene_list_sizes <- matrix(nrow=7,ncol=3)
dimnames(gene_list_sizes) <- list(c("Full by year","2009 by season","Winter 2009","Summer
2009","Prague 2009","Ostrava 2009","Winter 2010"),c("Unweighted & batch -corrected","Weighted &
batch in lm","Weighted & batch-corrected"))

gene_list_sizes_non_parametric <- matrix(nrow=7,ncol=2)
dimnames(gene_list_sizes_non_parametric) <- list(c("Full by year","2009 by season","Winter
2009","Summer 2009","Prague 2009","Ostrava 2009","Winter 2010"),c("Unweighted & batch-
corrected","Weighted & batch-corrected"))

```

```

message("Trying the 'Unweighted & batch corrected' approach")
load("Unweighted batch-corrected data [Full].RData")

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,season,y
ear,City-season-year,year-season-City,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value")
for (i in 1:5){
  print(c("City","season","year","City-season-year","year-season-City")[i])
  print(head(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
  for (j in 1:15){
    print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"
))
  }
}

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {

```

```

if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
{
  message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
  if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.5 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.5 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.5 - StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
      |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)

```

```

write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected full - by city - lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]),split="p")[[2]][2])==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
      |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as

```

```
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
  {
    message(c("Updating gene symbol ",j))
    gene_list[j,1] <- as.character(gene_symbol_update[i,3])
  }
}
}
```

```
setwd(paste(directory,"Gene lists/Unweighted arrays/Full/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.5 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.5 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.5 - StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes[1,1] <- dim(gene_list)[2]
```

```
gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)
```

```
message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
```



```

message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL"
|
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3)
|
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected full - by year - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [full - Unweighted & batch-corrected].RData")

data$E <- batch_corrected_data_non_parametric

```

```

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,season,y
ear,City-season-year,year-season-City,levels=design_matrix)))

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

```

```

}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc
0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc 0.5 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc 0.5 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc 0.5 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[

```

```

as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by city - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{

```

```

for (i in 1:dim(gene_symbol_update)[2])
{
  if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
  {
    message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
    if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL"
|
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3)
|
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3))
&
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.5 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.5 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.5 - StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[1,1] <- dim(gene_list)[2]

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)

```

```

head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

setwd(paste(directory,"Gene lists/Unweighted arrays/Full/Non-parametric ComBat",sep="/"))

```

```

write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected full - by year - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [full - Unweighted & batch-corrected - non-parametric].RData")

#####
#
#Focusing on 2009
message("Analysing the 2009 data by themselves")
setwd(paste(directory,"Workspaces",sep="/"))
load("Unweighted batch-corrected data [2009].RData")
dim(data)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,season,s
eason-City,City-season,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of 2009 data")
for (i in 1:4){
  print(c("City","season","season-City","City-season")[i])
  print(head(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
  for (j in 1:15){
    print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"
))
  }
}

gene_list <- topTable(linear_model,coef=2,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
      |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(as.character(gene_symbol_update[i,3]))
      }
    }
  }
}

setwd(paste(directory,"Gene lists/Unweighted arrays/2009 - by season/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)

```



```

write.table(file="Gene list [Unweighted & batch-corrected 2009 - by season - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected 2009 - by season - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected 2009 - by season - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[2,1] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2009 - by season - Unweighted & batch-corrected].RData")

data$E <- batch_corrected_data_non_parametric

linear_model <- lmFit(data,design_matrix,weights=NULL,contrasts=makeContrasts(City,season,l
evels=design_matrix))

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL"
      |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]

```

```

)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="p")[[2]][2],split="")[[2]]<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]]<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="q")[[2]][2],split="")[[2]]<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]]<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/2009 - by season/Non-parametric
ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2009 - by season - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2009 - by season - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2009 - by season - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[2,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2009 - by season - Unweighted & batch-corrected - non-parametric].RData")

```

```

#Winter is coming
message("Isolating winter 2009 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Unweighted batch-corrected data [2009].RData")
dim(data)

```

```

normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1,1:2]
dim(data)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of Winter 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{

```

```

for (i in 1:dim(gene_symbol_update)[2])
{
  if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
  {
    message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
    if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL"
|
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3)
|
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3))
&
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Winter 2009 - by city/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected Winter 2009 - by city - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected Winter 2009 - by city - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected Winter 2009 - by city - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected Winter 2009 - by city - lfc 0.1 - StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes[3,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))

```

```

save.image("Final data [Winter 2009 - by city - Unweighted & batch-corrected].RData")

data$E <- batch_corrected_data_non_parametric

linear_model <- lmFit(data, design_matrix, weights=NULL)
eBayes(contrasts.fit(fit=lmFit(data, design_matrix, weights=NULL), contrasts=makeContrasts(City, levels=design_matrix)))

gene_list <- topTable(linear_model, coef=1, number=20000, sort.by="logFC", p.value=0.05, lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory, "Gene symbol update", sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt", header=TRUE, sep="\t", na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ", j, " out of ", dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])], split="p")[[2]][2], split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]), split="p")[[2]][2], split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])], split="q")[[2]][2], split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]), split="q")[[2]][2], split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])], split="p")[[2]][2])=="strsplit(as.character(gene_symbol_update[i,6]), split="p")" |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])], split="q")[[2]][2])=="strsplit(as.character(gene_symbol_update[i,6]), split="q")"))
    }
  }
}

```

```

{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Winter 2009 - by city/Non-parametric
ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Winter 2009 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Winter 2009 - by city - lfc
0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Winter 2009 - by city - lfc
0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Winter 2009 - by city - lfc
0.1
StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[3,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Winter 2009 - by city - Unweighted & batch-corrected - non-parametric].RData")

```

```

message("Isolating summer 2009 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Unweighted batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,
4]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1,1:2]

```

```

dim(data)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,levels=d
esign_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of Summer 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2], "genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[

```

```

as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]]<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]]<3) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Summer 2009 - by city/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected Summer 2009 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected Summer 2009 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected Summer 2009 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected Summer 2009 - by city - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[4,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Summer 2009 - by city - Unweighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,levels=d
esign_matrix)))

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```



```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
      |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}
}

setwd(paste(directory,"Gene lists/Unweighted arrays/Summer 2009 - by city/Non-parametric
ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Summer 2009 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)

```

```

write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Summer 2009 - by city - lfc
0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Summer 2009 - by city - lfc
0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Summer 2009 - by city - lfc
0.1
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[4,1] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Summer 2009 - by city - Unweighted & batch-corrected - non-
parametric].RData")

message("Isolating Prague 2009 data for season-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Unweighted batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,
1]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1,c(1,3)]
dim(data)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(season,levels
=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of Prague 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){

```

```

print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {
      message(c("Updating gene symbol ",j))
      gene_list[j,1] <- as.character(gene_symbol_update[i,3])

```

```

}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Prague 2009 - by season/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes[5,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Prague 2009 - by season - Unweighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(season,levels=design_matrix)))

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])

```

```

{
  if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
  {
    message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
    if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
  {
    message(c("Updating gene symbol ",j))
    gene_list[j,1] <- as.character(gene_symbol_update[i,3])
  }
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Prague 2009 - by season/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Prague 2009 - by season - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[5,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Prague 2009 - by season - Unweighted & batch-corrected - non-
parametric].RData")

message("Isolating Ostrava 2009 data for season-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Unweighted batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1,][,
2]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1,c(1,3)]
dim(data)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(season,levels
=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of Prague 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))

```

```

gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))
][as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {
      message(c("Updating gene symbol ",j))
      gene_list[j,1] <- as.character(gene_symbol_update[i,3])
    }
  }
}
}

setwd(paste(directory,"Gene lists/Unweighted arrays/Ostrava 2009 - by season/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected Ostrava 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected Ostrava 2009 - by season - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))

```

```

write.table(file="Gene list [Unweighted & batch-corrected Ostrava 2009 - by season - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected Ostrava 2009 - by season - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[6,1] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Ostrava 2009 - by season - Unweighted & batch-corrected].RData")

data$E <- batch_corrected_data_non_parametric

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(season,levels
=design_matrix)))

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[

```



```

as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]]<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]]<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/Ostrava 2009 - by season/Non-parametric
ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Ostrava 2009 - by season -
lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Ostrava 2009 - by season -
lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Ostrava 2009 - by season -
lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected Ostrava 2009 - by season -
lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[6,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Ostrava 2009 - by season - Unweighted & batch-corrected - non-
parametric].RData")

```

```

#####
#
#Focusing on 2010
message("Isolating the 2010 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))

```

```

load("Unweighted batch-corrected data [2010].RData")
dim(data)

design_matrix <- cbind(design_matrix,surrogate_variables$sv)
colnames(design_matrix) <- c("Intercept","City","PC2","PC3")
linear_model <- lmFit(data,design_matrix,weights=NULL)
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=NULL),contrasts=makeContrasts(City,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of 2010 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"-->",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      ((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]

```

```

)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Unweighted arrays/2010 - by city/ComBat",sep="/"))
write.table(file="Gene list [Unweighted & batch-corrected 2010 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Unweighted & batch-corrected 2010 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Unweighted & batch-corrected 2010 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Unweighted & batch-corrected 2010 - by city - lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes [7,1] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2010 - by city - Unweighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix[,1:2],weights=array_weights),contrasts=makeContrast
s(City,levels=design_matrix[,1:2])))

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)

```

```

head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

setwd(paste(directory,"Gene lists/Unweighted arrays/2010 - by city/Non-parametric ComBat",sep="/"))

```

```

write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2010 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2010 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2010 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Unweighted & batch-corrected 2010 - by city - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[7,1] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2010 - by city - Unweighted & batch-corrected - non-parametric].RData")

message("Trying the 'Weighted & batch in linear model' approach")
load("Filtered data [Full].RData")

design_matrix <-
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
dimnames(design_matrix) <-
list(rownames(design_matrix),c("Intercept","City","season","year",paste("Chip",levels(as.factor(chip)))[1:
(length(levels(as.factor(chip_id)))-1)],sep="")))
for (i in 1:dim(design_matrix)[2])
{
  for (j in 5:dim(design_matrix)[3])
  {
    design_matrix[i,j] <-
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
  }
}

array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,season,year,City-season-year,year-season-City,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value")

```

```

for (i in 1:5){
  print(c("City","season","year","City-season-year","year-season-City")[i])
  print(head(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
  for (j in 1:15){
    print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"
))
  }
}

```

```

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as

```

```
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
```

```
setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Full",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.5 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.5 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.5 - StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
```

```
gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)
```

```
message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
    }
  }
}
```

```

if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Full",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model full - by city - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))

```



```

gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))
][as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Full",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc
0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.5 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))

```

```

write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.5 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.5 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[1,2] <- dim(gene_list)[2]

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))

```

```

{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Full",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model full - by year - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [full - Weighted & batch in linear model].RData")

```

```
#####
```

```
#
```

```
#Focusing on 2009
```

```
message("Analysing the 2009 data by themselves")
```

```
setwd(paste(directory,"Workspaces",sep="/"))
```

```
load("Filtered data [2009].RData")
```

```
dim(data)
```

```
design_matrix
```

```
<-
```

```
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
```

```
dimnames(design_matrix)
```

```
<-
```

```
list(rownames(design_matrix),c("Intercept","City","season",paste("Chip",levels(as.factor(chip)))[1:(length(levels(as.factor(chip_id)))-1)],sep=""))
```

```
for (i in 1:dim(design_matrix)[2])
```

```
{
```

```
  for (j in 5:dim(design_matrix)[3])
```

```
  {
```

```
    design_matrix[i,j]
```

```
<-
```

```
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
```

```
}  
}
```

```
array_weights <- arrayWeights(data, design=design_matrix, trace=FALSE)  
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)  
abline(h=1, lwd=1, lty=2)
```

```
linear_model <- eBayes(contrasts.fit(fit=lmFit(data, design_matrix, weights=array_weights),  
contrasts=makeContrasts(City  
, season, season-City, City-season, levels=design_matrix)))
```

```
#LFC cutoff  
message("Choosing a log fold change cutoff value for season-based comparison of 2009 data")  
for (i in 1:4){  
  print(c("City", "season", "season-City", "City-season")[i])  
  print(head(topTable(linear_model, coef=i, number=20000, sort.by="logFC", p.value=0.05, lfc=0)))  
  for (j in 1:15){  
    print(paste("lfc=", 0.1*j, "-->", dim(topTable(linear_model, coef=i, number=20000, sort.by="logFC",  
p.value=0.05, lfc=0.1*j))[2], "genes"  
))  
  }  
}
```

```
gene_list <- topTable(linear_model, coef=2, number=20000, sort.by="logFC", p.value=0.05, lfc=0.1)  
head(gene_list)  
dim(gene_list)
```

```
message("Updating gene symbols")  
setwd(paste(directory, "Gene symbol update", sep="/"))  
gene_symbol_update <- read.table(file="Gene symbol  
update.txt", header=TRUE, sep="\t", na.strings="NA")  
dim(gene_symbol_update)  
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous  
symbol",]  
dim(gene_symbol_update)  
for (j in 1:dim(gene_list)[2])  
{  
  for (i in 1:dim(gene_symbol_update)[2])  
  {  
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))  
    {  
      message(c("Checking gene: ", j, " out of ", dim(gene_list)[2]))  
    }  
  }  
}
```

```

if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(as.character(gene_symbol_update[i,3]))
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/2009 - by season",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model 2009 - by season - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model 2009 - by season - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model 2009 - by season - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model 2009 - by season - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes[2,2] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2009 - by season - Weighted & batch in linear model].RData")

```

```

#Winter is coming
message("Isolating winter 2009 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))

```

```

load("Filtered data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1,1:2]
dim(data)

design_matrix <-
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
dimnames(design_matrix) <-
list(rownames(design_matrix),c("Intercept","City",paste("Chip",levels(as.factor(chip)))[1:(length(levels(as
.factor(chip_id)))-1)],sep="")))
for (i in 1:dim(design_matrix)[2])
{
  for (j in 5:dim(design_matrix)[3])
  {
    design_matrix[i,j] <-
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
  }
}

array_weights <- arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of Winter 2009 data")
print("City")

```

```

print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {

```

```

    message(c("Updating gene symbol ",j))
    gene_list[j,1] <- as.character(gene_symbol_update[i,3])
  }
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Winter 2009 - by city",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model Winter 2009 - by city - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model Winter 2009 - by city - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model Winter 2009 - by city - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model Winter 2009 - by city - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[3,2] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Winter 2009 - by city - Weighted & batch in linear model].RData")

```

```

message("Isolating summer 2009 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Filtered data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,4]==1,1:2]
dim(data)

```



```

num.sv(dat=data$E,mod=design_matrix,method="leek")

design_matrix <-
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
dimnames(design_matrix) <-
list(rownames(design_matrix),c("Intercept","City",paste("Chip",levels(as.factor(chip)))[1:(length(levels(as
.factor(chip_id)))-1)],sep="")))
for (i in 1:dim(design_matrix)[2])
{
  for (j in 5:dim(design_matrix)[3])
  {
    design_matrix[i,j] <-
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
  }
}

array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of Summer 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene
symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")

```

```

dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Summer 2009 - by city",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model Summer 2009 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model Summer 2009 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model Summer 2009 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))

```

```

write.table(file="Gene list [Weighted & batch in linear model Summer 2009 - by city - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes[4,2] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Summer 2009 - by city - Weighted & batch in linear model].RData")

message("Isolating Prague 2009 data for season-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Filtered data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1,c(1,3)]
dim(data)

design_matrix <-
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
dimnames(design_matrix) <-
list(rownames(design_matrix),c("Intercept","season",paste("Chip",levels(as.factor(chip))[1:(length(levels
(as.factor(chip_id)))-1)],sep="")))
for (i in 1:dim(design_matrix)[2])
{
  for (j in 5:dim(design_matrix)[3])
  {
    design_matrix[i,j] <-
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
  }
}

array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)

```

```
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)
```

```
linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(season,levels=design_matrix)))
```

```
#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of Prague 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"-->",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"))
}
```

```
gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)
```

```
message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
```

```

(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]]<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]]<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Prague 2009 - by season",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model Prague 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model Prague 2009 - by season - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model Prague 2009 - by season - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model Prague 2009 - by season - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[5,2] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Prague 2009 - by season - Weighted & batch in linear model].RData")

```

```

message("Isolating Ostrava 2009 data for season-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Filtered data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]

```

```

batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[,samples_overview_matrix[samples_overview_matrix[,5]==1],[,
2]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1,c(1,3)]
dim(data)

design_matrix <-
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
dimnames(design_matrix) <-
list(rownames(design_matrix),c("Intercept","season",paste("Chip",levels(as.factor(chip))[1:(length(levels
(as.factor(chip_id)))-1)],sep="")))
for (i in 1:dim(design_matrix)[2])
{
  for (j in 5:dim(design_matrix)[3])
  {
    design_matrix[i,j] <-
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
  }
}

array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(seas
on,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of Prague 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

```

```

}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

```

```

}

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/Ostrava 2009 - by
season",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model Ostrava 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model Ostrava 2009 - by season - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model Ostrava 2009 - by season - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model Ostrava 2009 - by season - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[6,2] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Ostrava 2009 - by season - Weighted & batch in linear model].RData")

#####
#
#Focusing on 2010
message("Isolating the 2010 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Filtered data [2010].RData")
dim(data)

design_matrix <- cbind(design_matrix,surrogate_variables$sv)
colnames(design_matrix) <- c("Intercept","City","PC2","PC3")

design_matrix
cbind(design_matrix,matrix(nrow=dim(design_matrix)[2],ncol=length(levels(as.factor(chip_id)))-1))
dimnames(design_matrix)
list(rownames(design_matrix),c("Intercept","City","PC2","PC3",paste("Chip",levels(as.factor(chip)))[1:(len
gth(levels(as.factor(chip_id)))-1)],sep="")))
for (i in 1:dim(design_matrix)[2])
{
  for (j in 5:dim(design_matrix)[3])
  {
    design_matrix[i,j]
as.integer(strsplit(rownames(design_matrix)[i],split="_")[[2]][2]==levels(as.factor(chip_id))[j-4])
  }
}

```



```

}

array_weights <-arrayWeights(data,design=design_matrix,trace=FALSE)
barplot(array_weights, xlab="Array", ylab="Weight", col="white", las=2)
abline(h=1, lwd=1, lty=2)
linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of 2010 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]

```

```

)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch in linear model/2010 - by city",sep="/"))
write.table(file="Gene list [Weighted & batch in linear model 2010 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch in linear model 2010 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch in linear model 2010 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch in linear model 2010 - by city - lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[7,2] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2010 - by city - Weighted & batch in linear model].RData")

```

```

message("Trying the 'Weighted & batch-corrected' approach")
load("Weighted & batch-corrected data [Full].RData")

```

```

num.sv(dat=data$E,mod=design_matrix,method="leek")

```

```

for (i in 1:2){
  for (j in 1:3){

```

```

    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark
red","Turquoise","Red","Pale blue","Pink"))[[i+2*(j-1)]]),sep=" ")
  }
}

```

```

plotMDS(data$E,pch=19,col=coloring,main="Final data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering")
plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="Final data, non-
parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking for chip-
related clustering")

```

```

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of full data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect")

```

```

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,season,year,City-season-year,year-season-City,levels=design_matrix)))

```

```

#LFC cutoff
message("Choosing a log fold change cutoff value")
for (i in 1:5){
  print(c("City","season","year","City-season-year","year-season-City")[i])
  print(head(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
  for (j in 1:15){
    print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"
))
  }
}

```

```

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")

```

```

dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {
      message(c("Updating gene symbol ",j))
      gene_list[j,1] <- as.character(gene_symbol_update[i,3])
    }
  }
}
}
}

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc
0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc 0.5 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc 0.5 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))

```

```
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc 0.5 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
```

```
gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)
```

```
message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}
```

```

}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected full - by city - lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[

```

```

as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]]<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]]<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc
0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.5 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.5 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.5 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[1,3] <- dim(gene_list)[2]

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])

```

```

{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected full - by year - lfc 0.1 - StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

setwd(paste(directory,"Workspaces",sep="/"))

```



```

save.image("Final data [full - weighted & batch-corrected].RData")

data$E <- batch_corrected_data_non_parametric

num.sv(dat=data$E,mod=design_matrix,method="leek")

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,season,year,City-season-year,year-season-City,levels=design_matrix)))

gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]),split="p")[[2]][2])==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as

```

```
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
```

```
setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.5 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.5 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.5 - StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
```

```
gene_list <- topTable(linear_model,coef=4,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)
```

```
message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
    }
  }
}
```

```

if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by city - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.5)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))

```

```

gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))
][as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP))][as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {
      message(c("Updating gene symbol ",j))
      gene_list[j,1] <- as.character(gene_symbol_update[i,3])
    }
  }
}
}

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc
0.5].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.5 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))

```

```

write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.5 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.5 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[1,2] <- dim(gene_list)[2]

gene_list <- topTable(linear_model,coef=5,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {

```

```

    message(c("Updating gene symbol ",j))
    gene_list[j,1] <- as.character(gene_symbol_update[i,3])
  }
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Full/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected full - by year - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [full - weighted & batch-corrected - non-parametric].RData")

```

```
#####
```

```

#
#Focusing on 2009
message("Analysing the 2009 data by themselves")
setwd(paste(directory,"Workspaces",sep="/"))
load("Weighted & batch-corrected data [2009].RData")
dim(data)

num.sv(dat=data$E,mod=design_matrix,method="leek")

for (i in 1:2){
  for (j in 1:2){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark blue","Dark red","Turquoise","Red","Pale blue","Pink")[[i+2*(j-1)]])),sep=" ")
  }
}

```

```

plotMDS(data$E,pch=19,col=coloring,main="2009 data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering [2009]")

```

```

plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="2009 data, non-
parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking for chip-
related clustering [2009]")

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of 2009 data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect [2009]")

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,season,season-City,City-season,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of 2009 data")
for (i in 1:4){
  print(c("City","season","season-City","City-season")[i])
  print(head(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
  for (j in 1:15){
    print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=i,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes"
))
  }
}

gene_list <- topTable(linear_model,coef=2,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])

```

```

{
  if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
  {
    message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
    if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
  {
    message(c("Updating gene symbol ",j))
    gene_list[j,1] <- as.character(as.character(gene_symbol_update[i,3]))
  }
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/2009 - by season/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected 2009 - by season - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected 2009 - by season - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected 2009 - by season - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected 2009 - by season - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes[2,3] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2009 - by season - weighted & batch-corrected].RData")

```



```

data$E <- batch_corrected_data_non_parametric

num.sv(dat=data$E,mod=design_matrix,method="leek")

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,season,levels=design_matrix)))

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
      |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as

```

```
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
```

```
setwd(paste(directory,"Gene lists/Weighted & batch-corrected/2009 - by season/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2009 - by season - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2009 - by season - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2009 - by season - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2009 - by season - lfc 0.1 - StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[2,2] <- dim(gene_list)[2]
```

```
setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2009 - by season - weighted & batch-corrected - non-parametric].RData")
```

```
#Winter is coming
message("Isolating winter 2009 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Weighted & batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
```

```

labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,3]==1,1:2]
dim(data)

num.sv(dat=data$E,mod=design_matrix,method="leek")

for (i in 1:2){
  for (j in 1:1){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark blue", "Dark
red", "Turquoise", "Red", "Pale blue", "Pink")[[i+2*(j-1)]])),sep=" ")
  }
}

plotMDS(data$E,pch=19,col=coloring,main="Winter 2009 data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering [Winter 2009]")
plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="Winter 2009 data, non-
parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking for chip-
related clustering [Winter 2009]")

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of Winter 2009 data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect [Winter
2009]")

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of Winter 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Winter 2009 - by city/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected Winter 2009 - by city - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected Winter 2009 - by city - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected Winter 2009 - by city - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected Winter 2009 - by city - lfc 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[3,3] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Winter 2009 - by city - weighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

num.sv(dat=data$E,mod=design_matrix,method="leek")

```

```

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {

```

```

message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Winter 2009 - by city/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Winter 2009 - by city - lfc 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Winter 2009 - by city - lfc 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Winter 2009 - by city - lfc 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Winter 2009 - by city - lfc 0.1 - StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[3,2] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Winter 2009 - by city - weighted & batch-corrected - non-parametric].RData")

```

```

message("Isolating summer 2009 data for city-based comparison")

```

```

setwd(paste(directory,"Workspaces",sep="/"))
load("Weighted & batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[4]==1,1:2]
dim(data)

num.sv(dat=data$E,mod=design_matrix,method="leek")

for (i in 1:2){
  for (j in 2:2){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark blue","Dark red","Turquoise","Red","Pale blue","Pink"))[[i+2*(j-1)])),sep=" ")
  }
}

plotMDS(data$E,pch=19,col=coloring,main="Summer 2009 data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering [Summer 2009]")
plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="Summer 2009 data, non-parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking for chip-related clustering [Summer 2009]")

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of Summer 2009 data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect [Summer 2009]")

```

```

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of Summer 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3))

```



```

&(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))[as.character(
as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sy
mbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]))[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Summer 2009 - by
city/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected Summer 2009 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected Summer 2009 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected Summer 2009 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected Summer 2009 - by city - lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[4,3] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Summer 2009 - by city - weighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

num.sv(dat=data$E,mod=design_matrix,method="leek")

```

```

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update      <-      read.table(file="Gene      symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update      <-      gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3))      &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}
}
}

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Summer 2009 - by city/Non-parametric
ComBat",sep="/"))

```

```

write.table(file="Gene list [non-parametric - Weighted & batch-corrected Summer 2009 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Summer 2009 - by city - lfc 0.1
- BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Summer 2009 - by city - lfc 0.1
- EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Summer 2009 - by city - lfc 0.1
-
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[4,2] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Summer 2009 - by city - weighted & batch-corrected - non-parametric].RData")

message("Isolating Prague 2009 data for season-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Weighted & batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[samples_overview_matrix[samples_overview_matrix[,5]==1],[,
1]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,1]==1,c(1,3)]
dim(data)

num.sv(dat=data$E,mod=design_matrix,method="leek")

for (i in 1:1){
  for (j in 1:2){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark
red","Turquoise","Red","Pale blue","Pink")[[i+2*(j-1)]])),sep=" ")
  }
}

```

```

}

plotMDS(data$E,pch=19,col=coloring,main="Prague 2009 data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering [Prague 2009]")
plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="Prague2009 data, non-
parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking for chip-
related clustering [Prague 2009]")

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of Prague 2009 data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect [Prague
2009]")

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(seas
on,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of Prague 2009 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.1,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.1,lfc=0.1*j))[2],"genes"
))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.1,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)

```

```

for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
      {
        message(c("Updating gene symbol ",j))
        gene_list[j,1] <- as.character(gene_symbol_update[i,3])
      }
    }
  }
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Prague 2009 - by season/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))

```

```

gene_list_sizes[5,3] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Prague 2009 - by season - weighted & batch-corrected].RData")

data$E <- batch_corrected_data_non_parametric

num.sv(dat=data$E,mod=design_matrix,method="leek")

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(season,levels=design_matrix)))

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.1,lf=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])],split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a

```

```

s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]),split="p")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Prague 2009 - by season/Non-parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Prague 2009 - by season - lfc 0.1 - pvalue 0.1 - StRAnGER2].txt",cbind(gene_list[,1],2*gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[5,2] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Prague 2009 - by season - weighted & batch-corrected - non-parametric].RData")

```

```

message("Isolating Ostrava 2009 data for season-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Weighted & batch-corrected data [2009].RData")
dim(data)
normalized_czech_data <-
normalized_czech_data[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]
data <- data[samples_overview_matrix[samples_overview_matrix[,5]==1,][,2]==1]

```

```

batch_corrected_data_non_parametric <-
batch_corrected_data_non_parametric[,samples_overview_matrix[samples_overview_matrix[,5]==1],[,
2]==1]
array_weights <- array_weights[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
coloring <- coloring[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
chip <- chip[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
chip_id <- chip_id[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
labelling <- labelling[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1]
design_matrix <-
design_matrix[samples_overview_matrix[samples_overview_matrix[,5]==1],[,2]==1,c(1,3)]
dim(data)

num.sv(dat=data$E,mod=design_matrix,method="leek")

for (i in 2:2){
  for (j in 1:2){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark blue","Dark
red","Turquoise","Red","Pale blue","Pink")[[i+2*(j-1)]])),sep=" ")
  }
}

plotMDS(data$E,pch=19,col=coloring,main="Ostrava 2009 data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering [Ostrava 2009]")
plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="Ostrava 2009 data, non-
parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking for chip-
related clustering [Ostrava 2009]")

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of Ostrava 2009 data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect [Ostrava
2009]")

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(seas
on,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for season-based comparison of Prague 2009 data")
print("City")

```



```

print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {

```

```

    message(c("Updating gene symbol ",j))
    gene_list[j,1] <- as.character(gene_symbol_update[i,3])
  }
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Ostrava 2009 - by
season/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected Ostrava 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected Ostrava 2009 - by season - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected Ostrava 2009 - by season - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected Ostrava 2009 - by season - lfc 0.1 -
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[6,3] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Ostrava 2009 - by season - Weighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

num.sv(dat=data$E,mod=design_matrix,method="leek")

```

```

linear_model <- eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(seas
on,levels=design_matrix)))

```

```

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

```

```

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)

```

```

gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
      (as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])])=="NULL" |
      (((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2],split="")[[2]])<3
      & length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
      (length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[
as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2],split="")[[2]])<3 &
      length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
      (strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p") |
      strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])]),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
    {
      message(c("Updating gene symbol ",j))
      gene_list[j,1] <- as.character(gene_symbol_update[i,3])
    }
  }
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/Ostrava 2009 - by season/Non-
parametric ComBat",sep="/"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Ostrava 2009 - by season - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Ostrava 2009 - by season - lfc
0.1 - BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [non-parametric - Weighted & batch-corrected Ostrava 2009 - by season - lfc
0.1 - EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))

```

```

write.table(file="Gene list [non-parametric - Weighted & batch-corrected Ostrava 2009 - by season - lfc
0.1
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes_non_parametric[6,2] <- dim(gene_list)[2]

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [Ostrava 2009 - by season - weighted & batch-corrected - non-
parametric].RData")

#####
#
#Focusing on 2010
message("Isolating the 2010 data for city-based comparison")
setwd(paste(directory,"Workspaces",sep="/"))
load("Weighted & batch-corrected data [2010].RData")
dim(data)

for (i in 1:2){
  for (j in 3:3){
    message(c(paste(Origin[[i]],Origin[[4-abs(j-2)]],Origin[[5+max(j-2,0)]],":",c("Dark
blue","Dark
red","Turquoise","Red","Pale blue","Pink")[[i+2*(j-1)]])),sep=" ")
  }
}

plotMDS(data$E,pch=19,col=coloring,main="2010 data")
plotMDS(data$E,labels=chip,col=coloring,main="Checking for chip-related clustering [2010]")
plotMDS(batch_corrected_data_non_parametric,pch=19,col=coloring,main="2010
data,
non-
parametric")
plotMDS(batch_corrected_data_non_parametric,labels=chip,col=coloring,main="Checking
for
chip-
related clustering [2010]")

interquartile_range <- apply(data$E,1,IQR,na.rm=TRUE)
top_variable <- order(interquartile_range,decreasing=TRUE)[1:500]
distance <- dist(t(data$E[top_variable,]))
plot(hclust(distance),labels=labelling,main="Dendrogram of 2010 data")
plot(hclust(distance),labels=chip,main="Dendrogram check for chip-induced batch effect [2010]")

num.sv(dat=data$E,mod=design_matrix,method="leek")
summary(prcomp(data$E,scale.=TRUE))

```

```

surrogate_variables <-
sva(dat=data$E,mod=design_matrix,mod0=design_matrix[,1],n.sv=num.sv(dat=data$E,mod=design_mat
rix,method="leek"))

design_matrix <- cbind(design_matrix,surrogate_variables$sv)
colnames(design_matrix) <- c("Intercept","City","PC2","PC3")
linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix,weights=array_weights),contrasts=makeContrasts(City
,levels=design_matrix)))

#LFC cutoff
message("Choosing a log fold change cutoff value for city-based comparison of 2010 data")
print("City")
print(head(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0)))
for (j in 1:15){
  print(paste("lfc=",0.1*j,"--
>",dim(topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1*j))[2],"genes
"))
}

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene
symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])))[as.character(as.list(illu
minaHumanv3ALIAS2PROBE)[gene_list[j,1]])]=="NULL" |

```

```

(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]]<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]]<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)]
)[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]]<3) &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]]<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(a
s.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2]==strsplit(as.character(gene_sym
bol_update[i,6]),split="p")
|
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as
.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2]==strsplit(as.character(gene_symb
ol_update[i,6]),split="q"))))
{
  message(c("Updating gene symbol ",j))
  gene_list[j,1] <- as.character(gene_symbol_update[i,3])
}
}
}
}
}

```

```

setwd(paste(directory,"Gene lists/Weighted & batch-corrected/2010 - by city/ComBat",sep="/"))
write.table(file="Gene list [Weighted & batch-corrected 2010 - by city - lfc
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)
write.table(file="Gene list [Weighted & batch-corrected 2010 - by city - lfc 0.1 -
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene
symbol","LFC - logarithmic","adj. p-value"))
write.table(file="Gene list [Weighted & batch-corrected 2010 - by city - lfc 0.1 -
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))
write.table(file="Gene list [Weighted & batch-corrected 2010 - by city - lfc 0.1 -
StRAnGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name
s=c("Gene symbol","LFC - natural"))
gene_list_sizes[7,3] <- dim(gene_list)[2]

```

```

setwd(paste(directory,"Workspaces",sep="/"))
save.image("Final data [2010 - by city - weighted & batch-corrected].RData")

```

```

data$E <- batch_corrected_data_non_parametric

```

```

num.sv(dat=data$E,mod=design_matrix,method="leek")

```

```

linear_model <-
eBayes(contrasts.fit(fit=lmFit(data,design_matrix[,1:2],weights=array_weights),contrasts=makeContrasts(City,levels=design_matrix[,1:2])))

gene_list <- topTable(linear_model,coef=1,number=20000,sort.by="logFC",p.value=0.05,lfc=0.1)
head(gene_list)
dim(gene_list)

message("Updating gene symbols")
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
for (j in 1:dim(gene_list)[2])
{
  for (i in 1:dim(gene_symbol_update)[2])
  {
    if (gene_list[j,1]==as.character(gene_symbol_update[i,1]))
    {
      message(c("Checking gene: ",j," out of ",dim(gene_list)[2]))
      if
(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]]]))=="NULL" |
(((length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2],split="")[[2]])<3
& length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="p")[[2]][2],split="")[[2]])<3) |
(length(strsplit(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2],split="")[[2]])<3 &
length(strsplit(strsplit(as.character(gene_symbol_update[i,6]),split="q")[[2]][2],split="")[[2]])<3)) &
(strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="p")[[2]][2])!=strsplit(as.character(gene_symbol_update[i,6]),split="p") |
strsplit(as.character(as.list(illuminaHumanv3MAP[mappedkeys(illuminaHumanv3MAP)])[as.character(as.list(illuminaHumanv3ALIAS2PROBE)[gene_list[j,1]])),split="q")[[2]][2])!=strsplit(as.character(gene_symbol_update[i,6]),split="q"))))
    {
      message(c("Updating gene symbol ",j))
      gene_list[j,1] <- as.character(gene_symbol_update[i,3])
    }
  }
}

```

```
}  
}  
}
```

```
setwd(paste(directory,"Gene lists/Weighted & batch-corrected/2010 - by city/Non-parametric  
ComBat",sep="/"))  
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2010 - by city - lfc  
0.1].txt",gene_list,sep="\t",quote=FALSE,row.names=FALSE,col.names=TRUE)  
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2010 - by city - lfc 0.1 -  
BioInfoMiner].txt",gene_list[,c(1,2,6)],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene  
symbol","LFC - logarithmic","adj. p-value"))  
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2010 - by city - lfc 0.1 -  
EnrichR].txt",gene_list[,1],sep="\t",quote=FALSE,row.names=FALSE,col.names=c("Gene symbol"))  
write.table(file="Gene list [non-parametric - Weighted & batch-corrected 2010 - by city - lfc 0.1 -  
StRANGER2].txt",cbind(gene_list[,1],2**gene_list[,2]),sep="\t",quote=FALSE,row.names=FALSE,col.name  
s=c("Gene symbol","LFC - natural"))  
gene_list_sizes_non_parametric[7,2] <- dim(gene_list)[2]  
  
setwd(paste(directory,"Workspaces",sep="/"))  
save.image("Final data [2010 - by city - weighted & batch-corrected - non-parametric].RData")
```

```
setwd(paste(directory,"Workspaces",sep="/"))  
load("Weighted & batch-corrected data [2009].RData")  
par(mfrow=c(1,1))  
plotMDS(data$E,pch=19,col=coloring,main="Original data")
```



```

#Enter the gene symbols of classifier candidates
genes_of_interest_original <-
c("PDGFA","HGF","NGFR","ZFYVE27","MINK1","LTK","RAPGEF1","PAX3","AMMECR1","SH3PXD2B","LZTS
1","COL9A1","TLX2")

setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
print(c("Original: ",genes_of_interest_original))
for(i in 1:length(genes_of_interest_original))
{
  if (sum(as.integer(as.character(gene_symbol_update[,3]) == genes_of_interest_original[i])) > 0)
  {
    message("Changing gene symbol ",genes_of_interest_original[i], " back to
",as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i,1]),".")
    genes_of_interest_original[i] <-
as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i,1])
    print(genes_of_interest_original)
  }
}
print(c("Final list: ",genes_of_interest_original))

par(mfrow=c(4,4))
for (i in 1:length(genes_of_interest_original))
{
  if (i==1)
  {
    inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
    for(j in 1:length(genes_of_interest_original))
    {
      inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest_original[j])
    }
    plotMDS(data$E[inspector_gadget,],pch=19,col=coloring,main="No exclusions")
  }
}

genes_of_interest <- genes_of_interest_original[-i]

```

```
message(paste("Attempting exclusion of gene ",genes_of_interest_original[i]))
print(genes_of_interest)
```

```
inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
for(j in 1:length(genes_of_interest))
{
  inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest[j])
}
message(sum(as.integer(inspector_gadget))," instances of the genes of interest located.")
```

```
plotMDS(data$E[inspector_gadget,],pch=19,col=coloring,main=paste(genes_of_interest_original[i],"excl
uded",sep=" "))
}
par(mfrow=c(1,1))
distance <- dist(t(data$E))
plot(hclust(distance),labels=labelling,main="Original data")
par(mfrow=c(4,4))
for (i in 1:length(genes_of_interest_original))
{
  if (i==1)
  {
    inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
    for(j in 1:length(genes_of_interest_original))
    {
      inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest_original[j])
    }
    distance <- dist(t(data$E[inspector_gadget,]))
    plot(hclust(distance),labels=labelling,main="No exclusions")
  }
}
```

```
genes_of_interest <- genes_of_interest_original[-i]
message(paste("Attempting exclusion of gene ",genes_of_interest_original[i]))
print(genes_of_interest)
```

```
inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
for(j in 1:length(genes_of_interest))
{
  inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest[j])
}
message(sum(as.integer(inspector_gadget))," instances of the genes of interest located.")
```

```
distance <- dist(t(data$E[inspector_gadget,]))
```

```

plot(hclust(distance),labels=labelling,main=paste(genes_of_interest_original[i],"excluded",sep=" "))
}

message("Final gene list")
genes_of_interest_original <-
c("PDGFA","HGF","NGFR","ZFYVE27","MINK1","RAPGEF1","PAX3","SH3PXD2B","COL9A1")

setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
print(c("Original: ",genes_of_interest_original))
for(i in 1:length(genes_of_interest_original))
{
  if (sum(as.integer(as.character(gene_symbol_update[,3]) == genes_of_interest_original[i])) > 0)
  {
    message("Changing gene symbol ",genes_of_interest_original[i]," back to
",as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i],1]),".")
    genes_of_interest_original[i] <-
as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i],1])
    print(genes_of_interest_original)
  }
}
print(c("Final list: ",genes_of_interest_original))

par(mfrow=c(1,1))
genes_of_interest <- genes_of_interest_original
inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
for(j in 1:length(genes_of_interest))
{
  inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest[j])
}
message(sum(as.integer(inspector_gadget))," instances of the genes of interest located.")
plotMDS(data$E[inspector_gadget,],pch=19,col=coloring,main="PDGFA, HGF, NGFR, ZFYVE27, MINK1,
RAPGEF1, PAX3, SH3PXD2B, COL9A1")
distance <- dist(t(data$E[inspector_gadget,]))

```

```
plot(hclust(distance),labels=labelling,main="PDGFA, HGF, NGFR, ZFYVE27, MINK1, RAPGEF1, PAX3, SH3PXD2B, COL9A1")
```

```
message("Attempting to create a gene list from the summer 2009 differentially expressed genes")
genes_of_interest_original <- c("SERPINB2","RBM3","RBBP6")
```

```
setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])!="Previous
symbol",]
dim(gene_symbol_update)
print(c("Original: ",genes_of_interest_original))
for(i in 1:length(genes_of_interest_original))
{
  if (sum(as.integer(as.character(gene_symbol_update[,3]) == genes_of_interest_original[i])) > 0)
  {
    message("Changing gene symbol ",genes_of_interest_original[i], " back to
",as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i],1]),".")
    genes_of_interest_original[i] <-
as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i],1])
    print(genes_of_interest_original)
  }
}
print(c("Final list: ",genes_of_interest_original))
```

```
par(mfrow=c(1,1))
inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
for(j in 1:length(genes_of_interest_original))
{
  inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest_original[j])
}
plotMDS(data$E[inspector_gadget,],pch=19,col=coloring,main="No exclusions")
```

```
distance <- dist(t(data$E))
plot(hclust(distance),labels=labelling,main="Original data")
par(mfrow=c(2,2))
for (i in 1:length(genes_of_interest_original))
{
```

```

if (i==1)
{
  inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
  for(j in 1:length(genes_of_interest_original))
  {
    inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest_original[j])
  }
  distance <- dist(t(data$E[inspector_gadget,]))
  plot(hclust(distance),labels=labelling,main="No exclusions")
}

genes_of_interest <- genes_of_interest_original[-i]
message(paste("Attempting exclusion of gene ",genes_of_interest_original[i]))
print(genes_of_interest)

inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
for(j in 1:length(genes_of_interest))
{
  inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest[j])
}
message(sum(as.integer(inspector_gadget))," instances of the genes of interest located.")

distance <- dist(t(data$E[inspector_gadget,]))
plot(hclust(distance),labels=labelling,main=paste(genes_of_interest_original[i],"excluded",sep=" "))
}

message("Final gene list")
genes_of_interest_original <- c("SERPINB2","RBM3")

setwd(paste(directory,"Gene symbol update",sep="/"))
gene_symbol_update <- read.table(file="Gene symbol
update.txt",header=TRUE,sep="\t",na.strings="NA")
dim(gene_symbol_update)
gene_symbol_update <- gene_symbol_update[as.character(gene_symbol_update[,2])=="Previous
symbol",]
dim(gene_symbol_update)
print(c("Original: ",genes_of_interest_original))
for(i in 1:length(genes_of_interest_original))
{
  if (sum(as.integer(as.character(gene_symbol_update[,3]) == genes_of_interest_original[i])) > 0)
  {

```

```

    message("Changing gene symbol ",genes_of_interest_original[i," back to
",as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i,1]),".")
    genes_of_interest_original[i] <-
as.character(gene_symbol_update[as.character(gene_symbol_update[,3])
genes_of_interest_original[i,1])
    print(genes_of_interest_original)
  }
}
print(c("Final list: ",genes_of_interest_original))

par(mfrow=c(1,1))
genes_of_interest <- genes_of_interest_original
inspector_gadget <- matrix(data=FALSE,ncol=length(data$genes$TargetID))
for(j in 1:length(genes_of_interest))
{
  inspector_gadget <- inspector_gadget | (data$genes$TargetID == genes_of_interest[j])
}
message(sum(as.integer(inspector_gadget))," instances of the genes of interest located.")
distance <- dist(t(data$E[inspector_gadget,]))
plot(hclust(distance),labels=labelling,main="SERPINB2, RBM3")

```