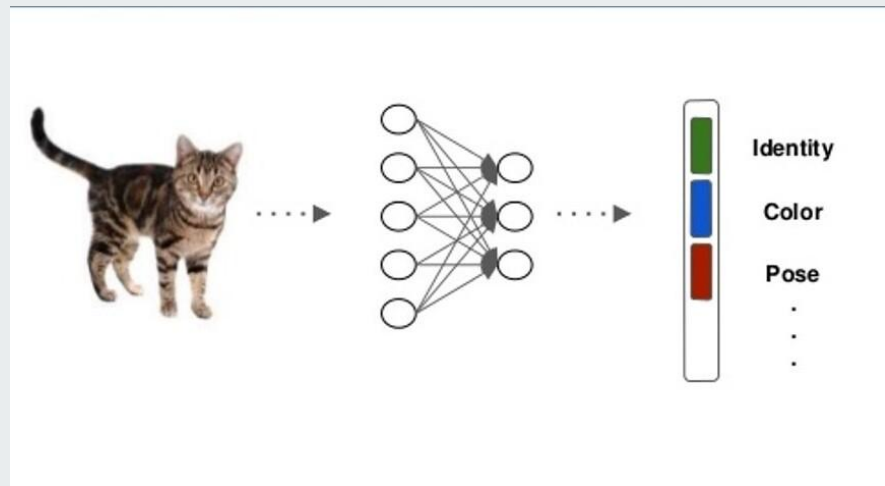
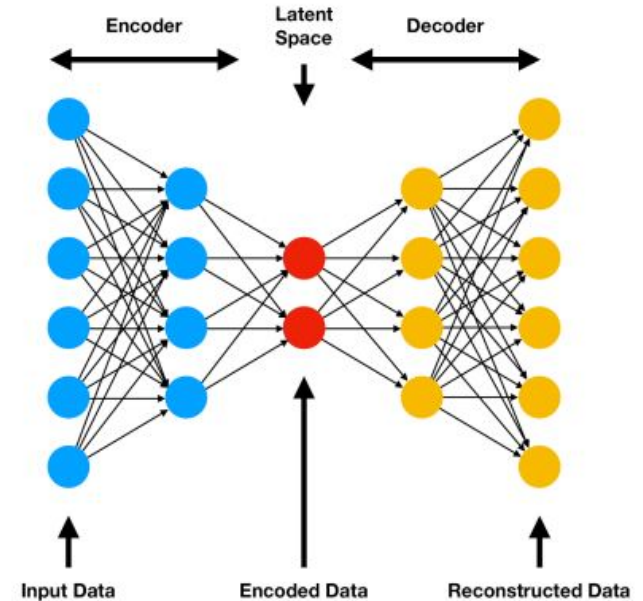


# XA4C: eXplainable representation learning via Autoencoders revealing Critical genes



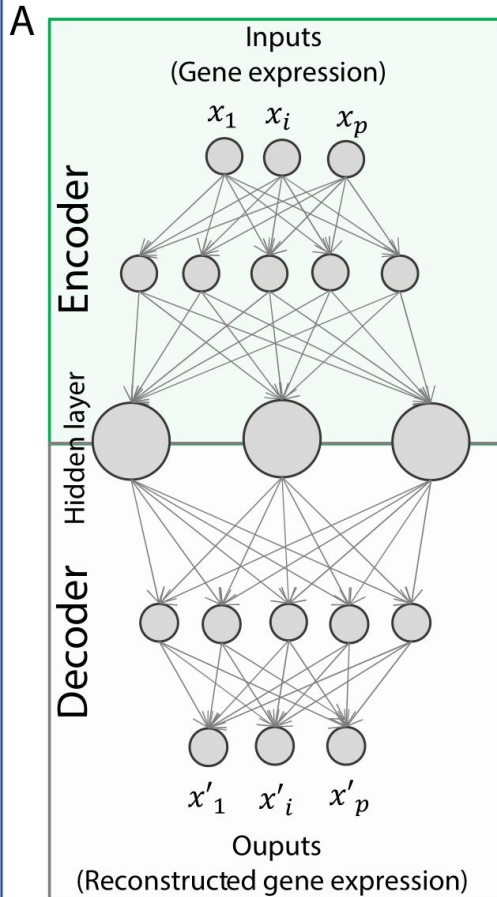
# Εισαγωγή

1. XA4C = eXplainable Autoencoder for Critical genes.  
Αποτελεί εργαλείο ανάλυσης δεδομένων γονιδιακής έκφρασης, το οποίο δημιουργεί έναν επεξηγήσιμο Autoencoder για την ανίχνευση Κρίσιμων γονιδίων.
2. Κρίσιμα γονίδια = γονίδια που συμβάλλουν σε μεγάλο βαθμό στην εκμάθηση των αναπαραστάσεων (οι αναπαραστάσεις σε έναν Autoencoder ονομάζονται latent variables ή αλλιώς λανθάνουσες μεταβλητές).
3. Οι **Autoencoders** ([τύπος νευρωνικού δικτύου](#)) είναι σε θέση να μαθαίνουν τις κρυφές αναπαραστάσεις (representations) των δεδομένων εισόδου, οδηγώντας στην δημιουργία των latent variables. Αυτές μπορούν να χρησιμοποιηθούν στα επόμενα στάδια ανάλυσης. Η διαδικασία αυτή ονομάζεται **Representation Learning (RL)**.
4. Μια λανθάνουσα μεταβλητή είναι μια μεταβλητή που δεν μπορεί να παρατηρηθεί ή να μετρηθεί άμεσα, αλλά θεωρείται ότι υπάρχει με βάση τα παρατηρήσιμα δεδομένα.
5. Επειδή οι αναπαραστάσεις που μαθαίνονται σε έναν autoencoder είναι δύσκολο να ερμηνευθούν, το XA4C παρέχει για κάθε γονίδιο τη συμβολή του στις λανθάνουσες μεταβλητές του Autoencoder, με βάση την οποία τα Κρίσιμα γονίδια ταξινομούνται.



# Μέθοδοι/Αλγόριθμοι

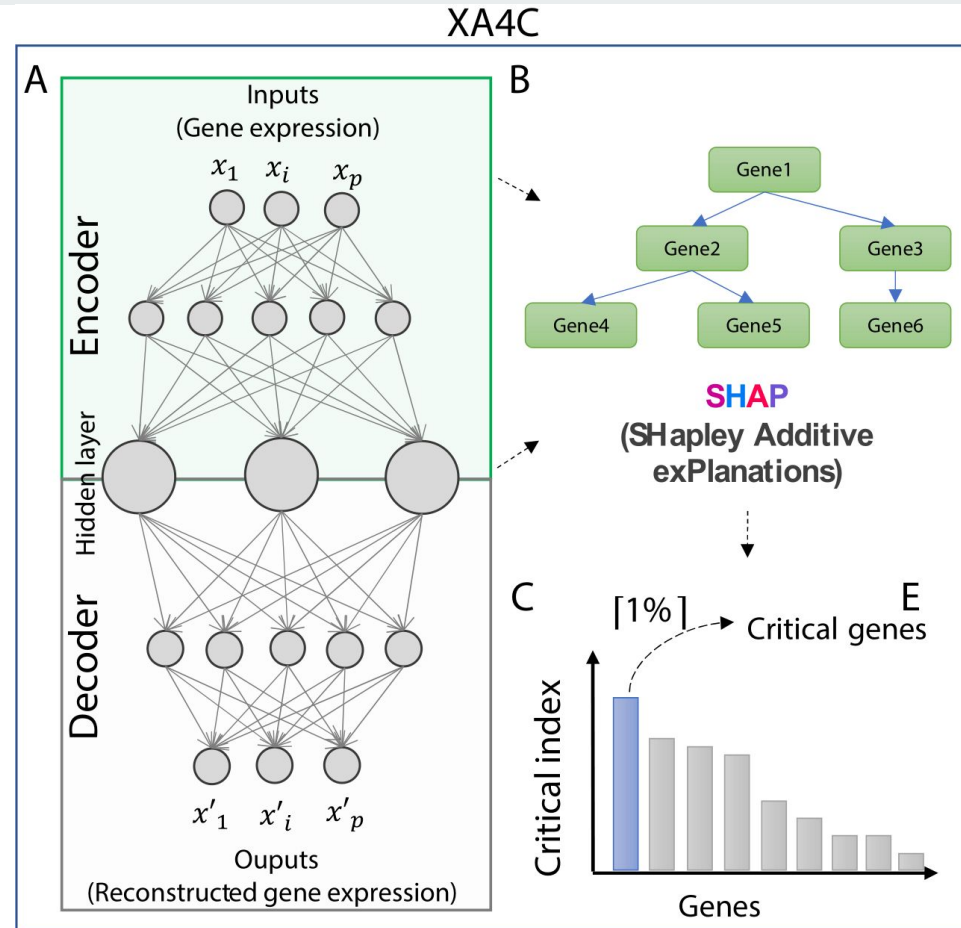
- A. **Autoencoder:** Οι Autoencoders μπορούν να εκπαιδευτούν για την επίτευξη του Representation Learning (RL). Στόχος: Κατασκευάζεται ένας Autoencoder για την εκμάθηση αναπαραστάσεων των προφίλ της έκφρασης των γονιδίων εισόδου.



# Μέθοδοι/Αλγόριθμοι

B. Τα **XGBoost** και **TreeSHAP** χρησιμοποιούνται για την αξιολόγηση των τιμών SHAP και των Κρίσιμων δεικτών για όλα τα γονίδια. Οι **τιμές SHAP** αποδίδουν μια τιμή σπουδαιότητας σε κάθε χαρακτηριστικό. Τα χαρακτηριστικά με θετικές τιμές SHAP επηρεάζουν θετικά το αποτέλεσμα, ενώ εκείνα με αρνητικές τιμές έχουν αρνητικό αντίκτυπο.

C. Το XA4C εξάγει τιμές SHAP για την έκφραση κάθε γονιδίου εισόδου ξεχωριστά. Έπειτα ποσοτικοποιεί τη συμβολή του σε κάθε αναπαράσταση και τις αθροίζει για να σχηματίσει τον "Κρίσιμο δείκτη" για κάθε γονίδιο. Αυτοί οι Κρίσιμοι δείκτες θα χρησιμοποιηθούν για την ιεράρχηση των Κρίσιμων γονιδίων με βάση ένα καθορισμένο από τον χρήστη όριο, π.χ. 1%. Τα Κρίσιμα γονίδια είναι αυτά με τους κορυφαίους Κρίσιμους δείκτες 1%.

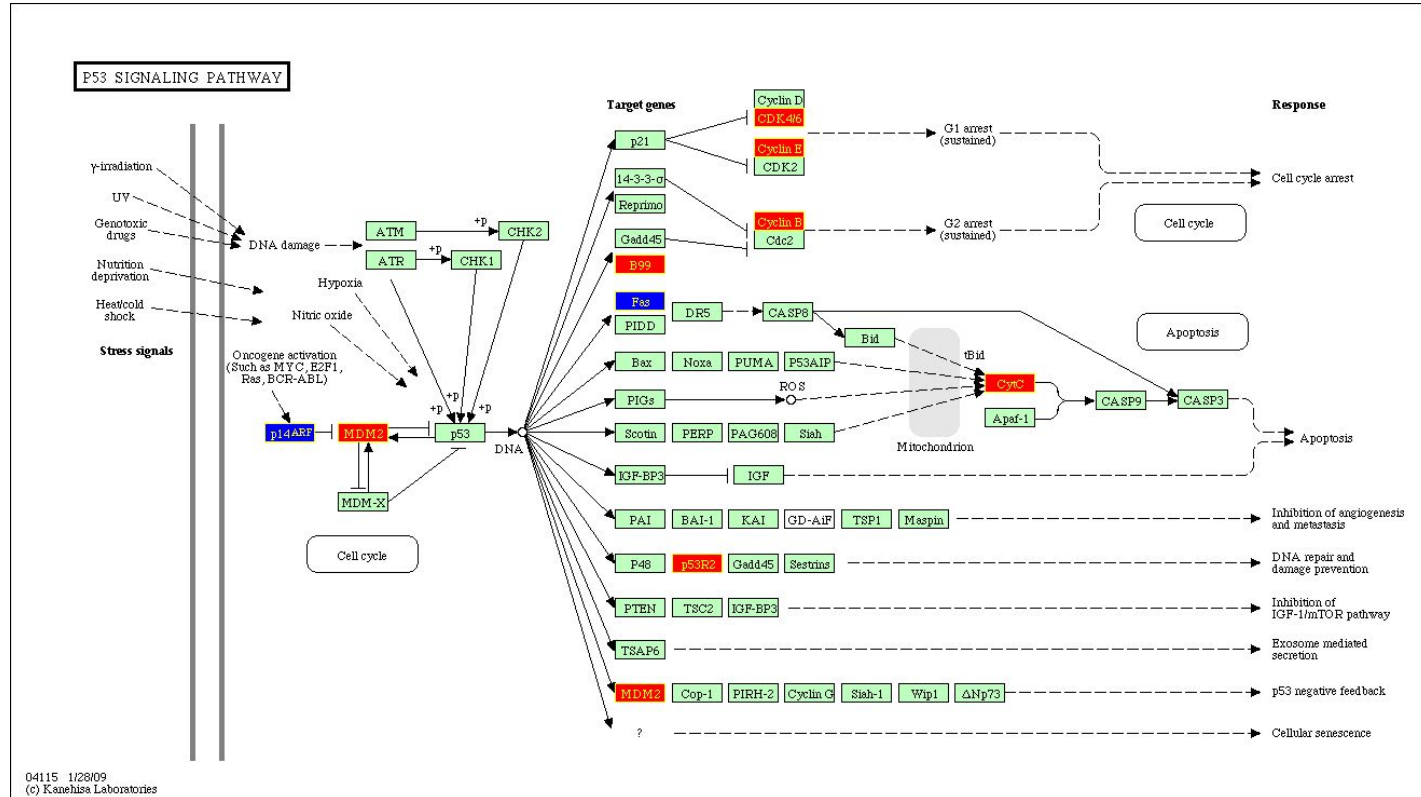


# Μέθοδοι/Αλγόριθμοι

D. Τα μονοπάτια KEGG είναι μια συλλογή χαρτών διαδρομών που αντιπροσωπεύουν τις γνώσεις μας για τα δίκτυα μοριακών αλληλεπιδράσεων, αντιδράσεων και σχέσεων.

Το **KEGG Pathway Enrichment analysis** είναι μια υπολογιστική μέθοδος που χρησιμοποιείται για τον εντοπισμό των μονοπατιών KEGG που υπερεκπροσωπούνται από γονίδια, με βάση τις τιμές SHAP (δηλαδή αυτά τα μονοπάτια εκπροσωπούνται περισσότερο στη λίστα εισόδου από ό,τι θα αναμενόταν τυχαία).

(<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0126492>).

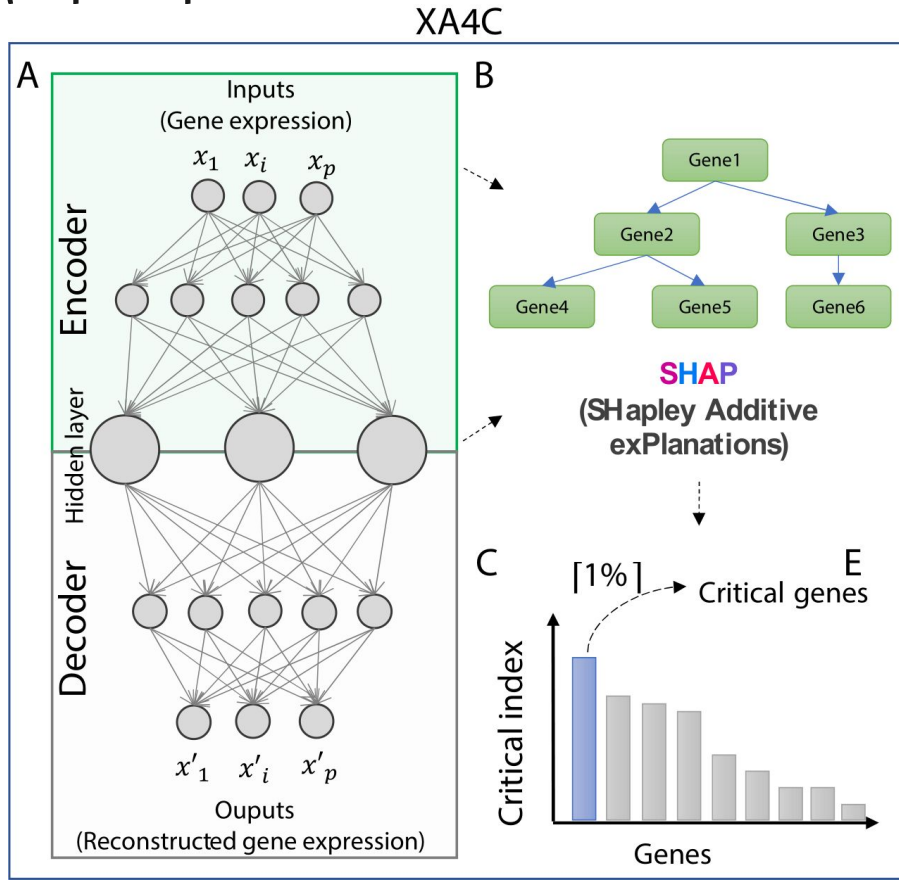


# Μέθοδοι/Αλγόριθμοι

E.

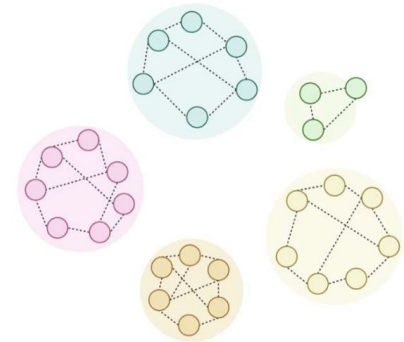
## Connectivity analysis:

αποκαλύπτει μοτίβα αλληλεπίδρασης μεταξύ γονιδίων, με επίκεντρο τα Κρίσιμα γονίδια σε μονοπάτια. Συγκρίνονται καρκινικά κύτταρα (όγκοι) και φυσιολογικά κύτταρα.

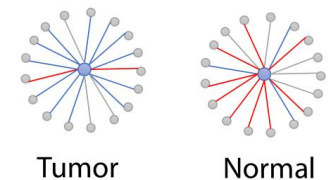


## Analyses


### D Pathways enrichment



### Connectivity pattern



# Αποτελέσματα



1. Η εφαρμογή του ΧΑ4C σε δεδομένα γονιδιακής έκφρασης σε έξι καρκίνους, έδειξε ότι τα κρίσιμα γονίδια καταγράφουν βασικές διαδρομές στις υποκείμενες μορφές καρκίνου.
2. Ανακαλύφθηκε ότι τα Κρίσιμα γονίδια έχουν τις εξής δύο ιδιότητες:
  - 1) παρέχουν νέες γνώσεις στα δεδομένα μεταγραφώματος που δεν μπορούν να αποτυπωθούν με την παραδοσιακή ανάλυση και
  - 2) τα Κρίσιμα γονίδια έχουν ισχυρή σχέση με την παθολογία της νόσου.Ως εκ τούτου, το ΧΑ4C μπορεί να αποκαλύψει ένα πρόσθετο περιβάλλον δεδομένων γονιδιακής έκφρασης.
3. Για παράδειγμα, ανακαλύφθηκαν πέντε κρίσιμα γονίδια που βρίσκονται στο κέντρο της διαδρομής της αποικοδόμησης της λυσίνης (hsa00310), εμφανίζοντας διακριτά μοτίβα αλληλεπίδρασης σε όγκους και φυσιολογικούς ιστούς.

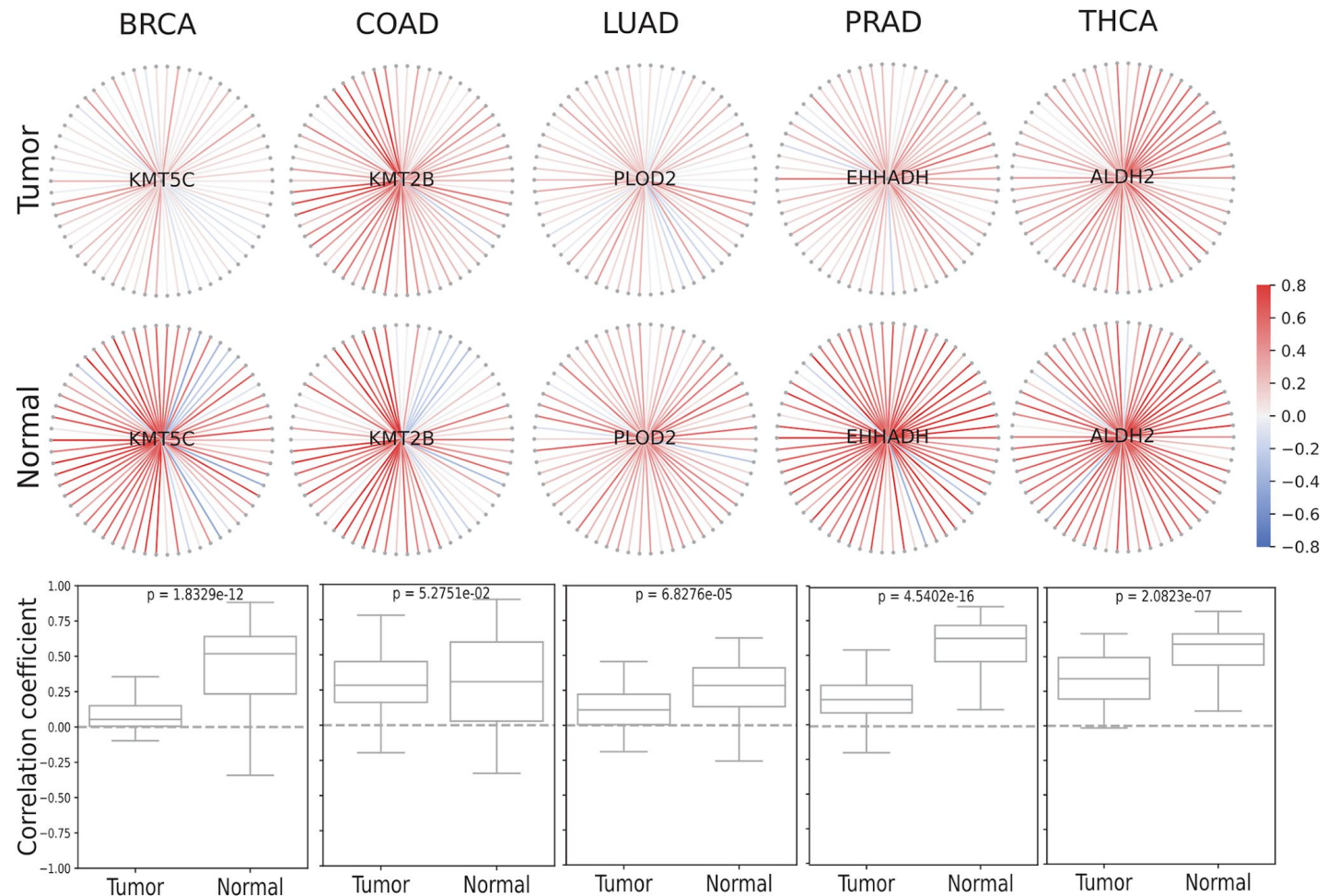


# Αποτελέσματα (συνέχεια)

Μονοπάτι αποικοδόμησης της  
λυσίνης (I00310).

Τα Κρίσιμα γονίδια έχουν γαλάζιο  
χρώμα και περιβάλλονται από  
πρόσθετα γονίδια του ίδιου  
μονοπατιού με γκρι χρώμα.

Τα Κρίσιμα γονίδια που  
παρουσιάζονται σε αυτό το σχήμα  
είναι νέα, καθώς δεν έχουν  
εντοπιστεί ούτε από την  
παραδοσιακή ανάλυση  
αναζήτησης για τα γονίδια Hub  
ούτε DiffEx.







***The End!***