

## **Θέμα Διπλωματικής:** Αντίστροφη Μηχανική Ανάπτυξης Εμβολίων – Reverse Vaccinology

### **Περιγραφή του προβλήματος**

Πρόσφατα, μια νέα μέθοδος ονόματι Αντίστροφη Μηχανική Ανάπτυξης Εμβολίων (ΑΜΑΕ), η οποία χρησιμοποιεί μια γονιδιωματική (αντί κυτταρική) προσέγγιση εφαρμόστηκε με επιτυχία, από τη φαρμακευτική εταιρία Novartis Vaccines & Diagnostics, για την ανάπτυξη εμβολίων κατά παθογόνων μικροοργανισμών που προηγουμένως ήταν ανθεκτικοί:

<http://www.novartis.com/newsroom/media-releases/en/2010/1443940.shtml>

Η ΑΜΑΕ, εφαρμόζει αρχές και μεθόδους γενομικής και βιοπληροφορικής για να επιτύχει ταχύτατη (1-2 χρόνια) και αποτελεσματική ανάπτυξη εμβολίων ακολουθώντας μια "προς-τα-πίσω" οδό (κάτι αντίστοιχο με Reverse Engineering) ξεκινώντας αντί από το επίπεδο του κυττάρου όπως γίνεται στη συμβατική οδό (5-15 χρόνια) ανάπτυξης εμβολίων, από το επίπεδο του γονιδιώματος του μικροοργανισμού για τον οποίο απαιτείται ανοσοποιητική ανάπτυξη.

Παρότι η ΑΜΑΕ είναι πολλά υποσχόμενη, δεν υπάρχει συστηματική, βιοπληροφορική μέθοδος που να είναι ευρέως εφαρμόσιμη στην *in silico* πρόγνωση εμβολίων για διαφορετικά είδη βακτηρίων και ιών. Οι μέχρι τώρα προσεγγίσεις είναι αρκετά περιγραφικές και παρότι ολιστικές στη θεωρία τους, στην πράξη δεν καταφέρνουν να συγχωνέσουν τις διαφορετικές παραμέτρους και να προτείνουν μια μοναδική και ποσοτική «συνταγή» «καλού» εμβολίου.

### **Current state-of-the-art**

1. <http://www.violinet.org/vaxign/>
2. <http://miracle.igib.res.in/dynavac/>
3. <http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>
4. <http://www.bio.unipd.it/molbinfo>

### **Στόχος**

Η κατανόηση της γενομικής και πρωτεομικής υπογραφής υποψήφιων αντιγόνων για την ανάπτυξη εμβολίων.

Η εργασία αυτή έχει δύο σκέλη. Το πρώτο αφορά στο λειτουργικό χαρακτηρισμό υποψήφιων γονιδίων που μπορεί να λειτουργήσουν ως στόχοι για την ανάπτυξη εμβολίων και στο σχεδιασμό της απαραίτητης εργαλιοθήκης (toolkit) και workflow που θα υποστηρίξει το χαρακτηρισμό αυτό. Το δεύτερο σκέλος, αφορά στην εκπαίδευση μίας μεθόδου Μηχανικής Μάθησης που βασίζεται σε Sparse Bayesian Learning, με σκοπό τόσο την πρόβλεψη νέων υποψήφιων πρωτεϊνών για την ανάπτυξη εμβολίων ("Forward" Machine Learning) όσο και την βιολογική μελέτη των παραμέτρων (weights) του μοντέλου ("Reverse" Machine Learning). Ανάλογα με την εξοικείωση και τη σχετική πρόοδο του φοιτητή, τα δύο σκέλη της εργασίας αυτής μπορεί να αποτελέσουν είτε μία εργασία ή δύο ξεχωριστές εργασίες, όπου βασική προϋπόθεση για να γίνει η δεύτερη εργασία είναι να έχει ολοκληρωθεί η πρώτη.

Αναλυτικά οι στόχοι είναι οι εξής:

- A. Συλλογή του κατάλληλου training dataset (με γνωστά παραδείγματα για το positive dataset και μέσω random sampling για το negative dataset).

- B. Μελέτη διαθέσιμων εργαλείων λειτουργικού χαρακτηρισμού γονιδίων και πρωτεϊνών και συγκριτικής γονιδιωματικής που μπορούν να χρησιμοποιηθούν για AMAE.
- C. Σχεδιασμός του αλγοριθμικού workflow & toolkit με επιλογή των κατάλληλων εργαλείων του στόχου B.
- D. Υλοποίηση του RVM.
- E. Εκπαίδευση του RVM.
- F. Cross-validation του RVM.
- G. Βιολογική μελέτη των παραμέτρων (weights) του RVM.
- H. Πρόταση «δομικής συνταγής καλού» εμβολίου με τη μορφή Generalized Linear Model.

### **Απαιτούμενες γνώσεις**

**Απολύτως απαραίτητες:** Γνώσεις προγραμματισμού (Java, Biojava or Matlab) & Μηχανικής Μάθησης.

**Μερικώς απαραίτητες/επιθυμητές:** Γνώσεις Γενωμικής, Πρωτεομικής, & Μικροβιολογίας.

### **Τι θα μάθει ο φοιτητής που θα εμπλακεί**

Ο φοιτητής θα αποκτήσει ειδικές γνώσεις πάνω σε συγκριτική γονιδιωματική, λειτουργικό και δομικό χαρακτηρισμό γονιδίων και πρωτεϊνών, ανάλυση της νουκλεοτιδικής σύστασης γονιδιωμάτων, πηγές αναζήτησης πρωτογενών δεδομένων, χρήση εργαλείων χαρακτηρισμού γονιδιωμάτων (Artemis Genome Browser), εργαλείων (TAVERNA) για διαχείριση ροής εργασίας (workflow) και σύγκρισης γονιδιωμάτων (MAUVE), βασικές γνώσεις σχετικά με τους τύπους εμβολίων και τέλος απόκτηση εμπειρίας σε Sparse Bayesian Learning, και συγκεκριμένα σε Relevance Vector Machine. Κατεύθυνση: Η πρόταση αυτή απευθύνεται στους φοιτητές και των δύο κατευθύνσεων του ΠΜΣ ΤΠΙΒ.

### **Αναφορές**

1. Vernikos, G. S. Genome watch: overtake in reverse gear. *Nat Rev Microbiol* 6, 334-5 (2008).
2. He Y, Xiang Z, Mobley HLT. Vaxign: the first web-based vaccine design program for reverse vaccinology and an application for vaccine development. *J Biomed Biotechnol* 2010;2010:297505.
3. Pizza, M. et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287, 1816-20 (2000).
4. Rappuoli, R. Reverse vaccinology. *Curr Opin Microbiol* 3, 445-50 (2000).
5. Giuliani, M. M. et al. A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci U S A* 103, 10834-9 (2006).
6. Scarselli, M., Giuliani, M. M., Adu-Bobie, J., Pizza, M. & Rappuoli, R. The impact of genomics on vaccine design. *Trends Biotechnol* 23, 84-91 (2005).

7. Mora, M., Donati, C., Medini, D., Covacci, A. & Rappuoli, R. Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. *Curr Opin Microbiol* 9, 532-6 (2006).
8. Musser, J. M. The next chapter in reverse vaccinology. *Nat Biotechnol* 24, 157-8 (2006).
9. Rappuoli, R. Reverse vaccinology. *Curr Opin Microbiol* 3, 445-50 (2000).
10. Rappuoli, R. & Covacci, A. Reverse vaccinology and genomics. *Science* 302, 602 (2003).
11. Vernikos, G. S. & Parkhill, J. Resolving the structural features of genomic islands: A machine learning approach. *Genome Res* 18, 331-42 (2008).
12. Tipping, M. E. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research* 1, 211-244 (2001).
13. Down, T. A. & Hubbard, T. J. Computational detection and location of transcription start sites in mammalian genomic DNA. *Genome Res* 12, 458-61 (2002).
14. Down, T. A. & Hubbard, T. J. Relevance vector machines for classifying points and regions in biological sequences. *Quantitative Biology Archive*, [<http://arxiv.org/abs/q-bio.GN/0312006>] (2003).
15. Harris, S. R. et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327, 469-74.
16. Maione, D. et al. Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 309, 148-50 (2005).
17. Tettelin, H. et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102, 13950-5 (2005).
18. Schölkopf, B., Burges, C. J. C. & Smola, A. J. (eds.) *Advances in Kernel Methods - Support Vector Learning* (MIT Press, Cambridge, Massachusetts, 1999).
19. Down, T., Leong, B. & Hubbard, T. J. A machine learning strategy to identify candidate binding sites in human protein-coding sequence. *BMC Bioinformatics* 7, 419 (2006).
20. Johri, A. K. et al. Group B Streptococcus: global incidence and vaccine development. *Nat Rev Microbiol* (2006).
21. Guss, B. et al. Getting to grips with strangles: an effective multi-component recombinant vaccine for the protection of horses from *Streptococcus equi* infection. *PLoS Pathog* 5, e1000584 (2009).
22. Serruto, D., Serino, L., Massignani, V. & Pizza, M. Genome-based approaches to develop vaccines against bacterial pathogens. *Vaccine* 27, 3245-50 (2009).

### **Επιβλέποντες**

Ηλίας Σ. Μανωλάκος, Av. Καθηγητής, ΕΚΠΑ, eliasm@di.uoa.gr

Γιώργος Σ. Βερνίκος, Διδάκτωρ, Cambridge University; The Wellcome Trust Sanger Institute, gvernikos@gmail.com.