

ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΠΡΟΒΛΕΨΗΣ ΓΟΝΟΤΥΠΩΝ ΣΕ ΟΛΟ ΤΟ ΓΟΝΙΔΙΩΜΑ ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΔΕΔΟΜΕΝΑ ΤΟΥ 1000GENOMES

Βασιλάρου Μ¹., Τοπαλούδη Α¹., Τσέτσος Φ¹., Δρινέας Π.², Γούναρη Ε.³, Μεταλλινού Χ¹, Γαλάνης, Γιαννάκη Ε.³, Σταματογιαννόπουλος Γ.⁴, Πάσχου Π.¹

¹ Τμήμα Μοριακής Βιολογίας και Γενετικής, Δημοκρίτειο Πανεπιστήμιο Θράκης, Αλεξανδρούπολη, Ελλάδα

² Τμήμα Πληροφορικής, Rensselaer Polytechnic Institute, Troy, NY, United States of America

³ Τμήμα Αιματολογίας, Νοσοκομείο Γεώργιος Παπανικολάου, Θεσσαλονίκη, Ελλάδα

⁴ Τμήμα Ιατρικής και Επιστημών Γονιδιώματος, University of Washington, Seattle, WA, United States of America

Με την αλματώδη τεχνολογική εξέλιξη των τελευταίων χρόνων, έχουν αυξηθεί θεαματικά οι γενετικές μελέτες μεγάλης κλίμακας. Η πιο διαδεδομένη μέθοδος είναι η μελέτη συσχέτισης σε επίπεδο γονιδιώματος, στην οποία γίνεται προσπάθεια συσχέτισης κοινών πολυμορφισμών με ένα συγκεκριμένο φαινότυπο. Οι μελέτες αυτές διεκπεραιώνονται μέσω γονοτυπικών πλατφορμών, οι οποίες καλύπτουν μεν πολυμορφισμούς σε όλο το γονιδίωμα, αλλά όχι σε μεγάλη πυκνότητα. Για να αυξηθεί η έκταση της κάλυψης των πολυμορφισμών, χρησιμοποιούνται μέθοδοι πρόβλεψης πολυμορφισμών (imputation) που βρίσκονται σε ανισορροπία σύνδεσης με τους πολυμορφισμούς που γονοτυπούνται. Η πρόβλεψη γίνεται με πλαίσιο αναφοράς συγκεκριμένες βάσεις δεδομένων που περιέχουν γονοτυπικά δεδομένα υψηλής πυκνότητας, με κύρια βάση το 1000Genomes. Στην έρευνά μας δοκιμάζουμε την ακρίβεια της πρόβλεψης των γονοτύπων δειγμάτων Νοτιοευρωπαϊκής καταγωγής. Ως μάρτυρα χρησιμοποιήσαμε γονοτυπικά δεδομένα 555 δειγμάτων Νοτιοευρωπαϊκής καταγωγής, με την πλειοψηφία να κατάγεται από την ηπειρωτική και νησιωτική Ελλάδα, σε εύρος 2.300.000 πολυμορφισμών. Συγκεντρώσαμε πληροφορίες από μια ευρεία ποικιλία από εμπορικές πλατφόρμες γονοτύπησης που χρησιμοποιούνται συχνά στις μελέτες συσχέτισης και απομονώσαμε τους αλληλεπικαλυπτόμενους πολυμορφισμούς. Χρησιμοποιώντας τους αλληλεπικαλυπτόμενους πολυμορφισμούς κάθε πλατφόρμας και με πλαίσιο αναφοράς τους Ευρωπαϊκούς πληθυσμούς IBS, TSI και CEU του 1000Genomes, κάναμε πρόβλεψη των γονοτύπων και τους συγκρίναμε με τους γονοτύπους βρίσκονταν στα αρχικά δεδομένα μας. Για τις αναλύσεις αυτές χρησιμοποιήθηκαν τα λογισμικά BEAGLE, SHAPEIT, IMPUTE και PLINK. Τα αποτελέσματά μας αξιολογούν την ικανότητα των Ευρωπαϊκών πληθυσμών του 1000Genomes να προβλέψουν τους γονοτύπους δειγμάτων Νοτιοευρωπαϊκής καταγωγής, και ειδικά Ελληνικής. Τα αποτελέσματά μας μπορούν να καθοδηγήσουν και να αυξήσουν την αξιοπιστία γονιδιωματικών αναλύσεων. [Με συγχρηματοδότηση από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και εθνικούς πόρους στο πλαίσιο της πράξης ΑΡΙΣΤΕΙΑ II (4386: GENOMAP.GR) του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Διά Βίου Μάθηση»]

EVALUATING GENOME-WIDE IMPUTATION ACCURACY ACROSS SOUTHERN EUROPEAN POPULATIONS USING 1000 GENOMES DATA AS REFERENCE

Vasilarou M¹, Topaloudi A¹, Tsetsos F¹, Drineas P.², Gounari E.³, Metallinou C.¹, Galanis, Yannaki E.³, Stamatoyannopoulos G⁴, Paschou P.¹

¹ *Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece*

² *Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY, United States of America*

³ *Department of Hematology, George Papanicolaou Hospital, Thessaloniki, Greece*

⁴ *Departments of Medicine and Genome Sciences, University of Washington, Seattle, WA, United States*

With the leaping technological advances of the new genomic era, we have witnessed an impressive increase in the numbers of large-scale genomic studies. The most common analytical method is the genome wide association study, in which a multitude of common polymorphisms that span the entire genome are tested for association with a specific phenotype. These studies are conducted by means of genotypic platforms, that cover genome wide polymorphisms, although not in full density. To increase the density of the polymorphisms, state-of-the-art methods of imputing the missing genotypes, through means of linkage disequilibrium, are employed. The imputation process is based on a reference panel that contains high-density genotypic data, with the most complete available one being the 1000Genomes project. In our study we test the accuracy of the imputation processes on samples of South-European ancestry. As comparison base we used genotypic data from 555 samples of South-European origin, with the majority being from mainland and insular Greece, genotyped on 2.3 million polymorphisms. We gathered data from a variety of commercial genotyping platforms that are used commonly on genome wide association studies and we extracted the genotypes that intersected with our original data. Using the overlapping data, we imputed polymorphisms using the IBS, TSI and CEU populations from the 1000Genomes project and we compared them with the full original dataset. For our analyses we used the software BEAGLE, SHAPEIT, IMPUTE and PLINK. Our results evaluate the suitability of the European populations of 1000Genomes as reference panels for imputation on South European samples, and especially of Greek origin. Our results can guide and increase the fidelity of genomic analyses based on samples of South European origin. [Supported by ARISTEIA II grant to PP (4386 - GENOMAP.GR). Operational Programme "Education and Lifelong Learning", co-funded by the European Union (European Social Fund) and national funds]