

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE
A
L'UNIVERSITE PARIS-SUD

ECOLE DOCTORALE N° 582
CBMS CANCEROLOGIE : BIOLOGIE - MEDECINE - SANTE
SPECIALITE DE DOCTORAT : Recherche clinique, innovation technologique, santé publique

Par

Yu FU

**Analyse intégrative de données génomiques et pharmacologiques
pour améliorer la prédiction de la réponse aux thérapies ciblées**

**Thèse présentée et soutenue à Gustave Roussy, le 19 Décembre 2016 :
Composition du Jury :**

Professeur Ambroise Christophe	UMR CNRS 8071, INRA	Rapporteur
Professeur Vert Jean-Philippe	Inserm U900, Institut Curie	Rapporteur
Docteur Barillot Emmanuel	Inserm U900, Institut Curie	Examineur
Professeur D'Alché-Buc Florence	Université Paris-Saclay	Examinatrice
Professeur André Fabrice	Inserm U981, Université Paris-Saclay	Directeur de thèse
Docteur Lefebvre Céline	Inserm U981, Gustave Roussy	Directrice de thèse

Analyse intégrative de données génomiques et pharmacologiques pour améliorer la prédiction de la réponse aux thérapies ciblées

Mot cles: génomique,modelisation,prédiction,médecine personnalisée

L'utilisation de thérapies ciblées dans le contexte de la médecine personnalisée du cancer a permis d'améliorer le traitement des patients dans différents types de cancer. Cependant, alors que la décision thérapeutique est basée sur une unique altération moléculaire (par exemple une mutation ou un changement du nombre de copies d'un gène), les tumeurs montrent différents degrés de réponse. Dans cette thèse, nous démontrons que la décision thérapeutique basée sur une unique altération n'est pas optimale et nous proposons un modèle mathématique intégrant des données génomiques et pharmacologiques pour identifier de nouveaux biomarqueurs prédictifs de la réponse thérapeutique. Le modèle a été construit à partir de deux bases de données de lignées cellulaires (the Genomics of Drug Sensitivity in Cancer, GDSC and the Cancer Cell Line Encyclopedia, CCLE) et validé avec des données de lignées et des données cliniques. De plus, nous avons également développé une nouvelle méthode pour améliorer la détection des mutations somatiques à partir de données de séquençage d'exomes complets et proposons un nouvel outil, cmDetect, disponible gratuitement pour la communauté scientifique.

Integrated analysis of genomic and pharmacological data to better predict the response to targeted therapies

Key words: genomics, predictive modeling, personalized medicine

The use of targeted therapies in the context of cancer personalized medicine has shown great improvement of patients' treatment in different cancer types. However, while the therapeutic decision is based on a single molecular alteration (for example a mutation or a gene copy number change), tumors will show different degrees of response. In this thesis, we demonstrate that a therapeutic decision based on a unique alteration is not optimal and we propose a mathematical model integrating genomic and pharmacological data to identify new single predictive biomarkers as well as combinations of biomarkers of therapy response. The model was trained using two public large-scale cell line data sets (the Genomics of Drug Sensitivity in Cancer, GDSC and the Cancer Cell Line Encyclopedia, CCLE) and validated with cell line and clinical data. Additionally, we also developed a new method for improving the detection of somatic mutations using whole exome sequencing data and propose a new tool, cmDetect, freely available to the scientific community.