

***English version following the French version.***

**TITRE DU PROJET : Épidémiologie génétique de la COVID-19 (EpiGenCOV)**

**RESPONSABLE :** Hugues Aschard, Institut Pasteur - Département Biologie computationnelle, Paris

**RÉSUMÉ**

**Contexte**

Les études épidémiologiques de la maladie à coronavirus 2019 (COVID-19) ont montré une forte variabilité interindividuelle chez les personnes infectées par le SARS-CoV-2 (coronavirus 2 du syndrome respiratoire aigu sévère), le tableau clinique variant de formes asymptomatiques à des formes létales. Outre le rôle de facteurs cliniques, environnementaux et liés au mode de vie, un nombre croissant d'études souligne la contribution à cette variabilité de facteurs liés à l'hôte, et au patrimoine génétique en particulier. Deux types d'approches sont mises en œuvre pour étudier la génétique de l'hôte dans les phénotypes de la COVID-19 : les études par séquençage (exome ou génome entier) et les études d'association pangénomique (GWAS). Les approches par séquençage visent à identifier de potentielle mutations rares associées à un risque élevé de formes idiopathiques sévères survenant chez de jeunes individus sans comorbidités. A l'inverse, les approches GWAS supposent un modèle polygénique impliquant un grand nombre de variants communs ayant des effets faibles. De même que pour les GWAS de maladies chroniques, l'identification des variants associés repose sur la constitution de cohortes avec de grandes tailles d'échantillons. Pour atteindre cet objectif, la communauté GWAS a rapidement convergé vers la mise en place d'un consortium international ([www.covid19hg.com](http://www.covid19hg.com)) incluant actuellement plus de 150 cohortes à travers le globe. En France, l'Institut Pasteur, en collaboration avec l'Institut Imagine, le CNRGH et l'Université de Bordeaux, a établi le consortium nationale EpiGenCOV pour contribuer à cet effort international.

**Objectifs**

L'objectif de ce projet est de consolider cette initiative et d'assembler une large collection de cohortes françaises pour la mise en œuvre d'études d'association pangénomiques des phénotypes de la COVID-19. Les données de cette collection seront utilisées pour répondre à trois objectifs spécifiques. Premièrement, la collection fournira à la communauté scientifique française une ressource unique pour des recherches portant sur la biologie de l'hôte et la génétique en particulier. Un système d'accès sécurisé sera mis en place en accord avec toutes les cohortes participantes, pour permettre l'accès aux données individuelles génétiques et phénotypiques de la collection pour les chercheurs internes, mais également externes, au consortium. Deuxièmement, la collection conduira les analyses GWAS des phénotypes primaires, tel que la susceptibilité et la sévérité de la COVID-19. Les résultats de ces GWAS seront partagés publiquement, et en particulier avec le consortium international covid19hg susmentionné. Troisièmement, la collection offrira un cadre optimal pour d'éventuelles études complémentaires futures tel que le Covid long ou la réponse au vaccin, en permettant une mise en œuvre rapide de ces analyses.

## Méthodes

Les analyses primaires porteront sur la susceptibilité et la sévérité de la maladie, définies sur la base des données disponibles, et en utilisant le codage proposé par le consortium covid19hg afin de faciliter la méta-analyse au niveau international. Deux modèles seront considérés pour la sévérité de la maladie : 1) les formes les plus sévères seront définies selon plusieurs critères : un test positif à la COVID-19, l'hospitalisation avec la COVID-19 comme motif principal d'admission, et soit le décès soit une assistance respiratoire pendant l'hospitalisation. 2) les personnes hospitalisées avec un test positif à la COVID-19, et dont l'hospitalisation est due à des symptômes liés à la COVID-19. Ces deux types de cas seront comparés 1) aux individus dont l'infection par le SRAS-CoV-2 a été confirmée en laboratoire et qui n'ont pas été hospitalisés 21 jours après le test, et 2) à la population générale (individu sans infection connue). Les analyses de la susceptibilité partielle seront axées sur la comparaison entre participants ayant déclaré une infection par le SRAS-CoV-2 et 1) les personnes dont le test de dépistage de l'infection par le SRAS-CoV-2 est négatif, ou 2) la population générale. Toutes les analyses suivront les standards existants pour la mise en œuvre de GWAS. Dans un second temps, des études d'association génétique seront conduites pour un large éventail de phénotypes supplémentaires liés à la COVID-19, comprenant l'étude des formes asymptomatiques de la maladie. Ces analyses influenceront également des analyses stratifiées de sous-ensemble de participants, afin d'explorer le rôle des comorbidités et d'autres facteurs de risque établis de la COVID-19. Les résultats de GWAS seront finalement utilisés pour des études complémentaires *in silico*, notamment des analyses d'héritabilité et des analyses de fine mapping pour identifier de potentiels variants causaux.

## Perspectives

Comprendre le rôle de la génétique de l'hôte dans la COVID-19 peut avoir de nombreuses implications dans la lutte contre la pandémie. Les connaissances acquises pourraient aider à distinguer les facteurs liés au mode de vie de la prédisposition clinique, et mettre en évidence les marqueurs génétiques liés à la progression de COVID-19, permettant de mieux stratifier les patients en fonction de leur risque de développer des complications graves. La compréhension des mécanismes génétiques impliqués dans la maladie peut également fournir des connaissances biologiques importantes susceptibles d'accélérer la découverte de cibles thérapeutiques à la fois pour le traitement des individus infectés, et pour le développement de médicaments prophylactiques. Enfin, plusieurs études récentes suggèrent des effets délétères à long terme sur la santé de l'infection au SARS-CoV-2, incluant des séquelles chroniques sur les poumons, le cerveau et le système cardiovasculaire. Le consortium EpiGenCOV offre un cadre idéal pour l'étude des facteurs génétiques sous-jacents à ces complications, et pour toute autre étude future de phénotypes liés à la COVID-19.

## **TITLE OF THE PROJECT: Host genetic and COVID-19 phenotypes (EpiGenCOV)**

**HEAD OF THE TEAM:** Hugues Aschard, Institut Pasteur - Departement of Computational Biology, Paris (France)

### **SUMMARY**

#### **Context**

Since the emergence of the novel coronavirus (SARS-CoV-2), epidemiological studies showed a stunning inter-individual variability among infected individuals, ranging from asymptomatic infection to lethal coronavirus infectious disease-19 (COVID19). Besides strong evidence that clinical, environmental and lifestyle factors contribute to these different infection outcomes, it seems increasingly likely that host-related factors, of genetic nature in particular, explain part of this variability. Two types of approaches have been initiated to study the host genetic component of COVID19: sequencing studies (whole-exome or whole-genome) and genome-wide association studies (GWAS). Sequencing approaches mostly aim at investigating rare mutations in outlier individuals, focusing on rare idiopathic severe cases occurring in young individuals without co-morbidities, with the hypothesis that such cases might represent patients with monogenic causes. On the other hand, the GWAS approach is built on the hypothesis of polygenicity, assuming the disease outcomes depends on multiple common genetic variants with modest to low effect. As for GWAS of chronic diseases, the detection of such variants relies on the aggregation of many cohorts to achieve the largest possible sample size. Thanks to well-established practice within the GWAS community, most existing COVID GWAS efforts have quickly converged into an international consortium which currently includes over 150 cohorts worldwide ([www.covid19hg.com](http://www.covid19hg.com)). In France, the Pasteur Institute, together with the Imagine Institute, the CNRGH, and the University of Bordeaux, have initiated the EpiGenCOV national collection to contribute to this international GWAS effort.

#### **Objectives**

The objective of this proposal is to strengthen this ongoing effort to assemble the largest possible French GWAS collection, in order to study how common host genetic variants affect COVID-19 phenotypes. Data from this collection will be used to address three specific aims. First, the collection will provide the French community a single and extensive accessible resource for projects addressing host biology questions involving host-biology, and host genetics in particular. In this regard, we will implement a controlled-access framework in agreement with all participating cohorts, providing access to individual-level genetic and phenotypic data for investigators internal and external to the EpiGenCOV consortium. Second, the collection will conduct a number of GWAS analyses on primary outcomes, including disease susceptibility and disease severity. Summary statistics from these analyses will be shared with the community, and especially the aforementioned covid19hg international efforts. Third and final, the collection will provide a unique resource for follow-up studies of additional COVID19 phenotypes such as long Covid and vaccination efficacy.

#### **Methods**

The primary analyses will address the susceptibility and severity of the disease defined based on available data, and using coding choices proposed by the covid19hg consortium to facilitate meta-analysis at the international level. The severity of the disease will be treated at least two ways:

1) The most severe form will be defined as individuals with laboratory confirmed SARS-CoV-2 infection that have been hospitalized with COVID19 as primary reason for admission, and are either deceased or have received respiratory support. 2) Hospitalized individuals with laboratory confirmed SARS-CoV-2 infection, whom hospitalization is due to COVID19-related symptoms. Both type of cases will be contrasted against 1) individuals with laboratory confirmed SARS-CoV-2 infection without hospitalization 21 days after the test, and 2) the general population (any individual without known infection). Partial susceptibility analyses will focus on contrasting individuals with reported SARS-CoV-2 infection against either 1) individuals with negative test for SARS-CoV-2 infection, or 2) the general population. All analyses will follow existing standard guideline for GWAS. We will also perform series of genetic association studies on a wide range of additional COVID19-related phenotypes, including the study of asymptomatic forms of the disease, but also stratified analysis in subset of participants, to explore the role of comorbidities and other COVID19 risk factors. When relevant, GWAS summary statistics will be used to conduct follow-up analyses including e.g. heritability analysis and fine-mapping analyses to identify potential causal variants.

## **Perspectives**

Understanding potential host contribution to susceptibility can provide critical insights to fight the disease. It can first help delineate lifestyle factors from clinical predisposition and point to genetic markers for COVID-19 risk and progression, enabling better stratification of patients into high or low risk of developing severe complications. This knowledge can be used to develop preventative measures, such as prophylactic medication and/or vaccination, when such opportunities arise in the later phases of the current pandemic or during future similar pandemic. Understanding the genetic mechanisms involved in the disease can also provide important biological knowledge that might accelerate the discovery of drug targets and suggest drug repurposing. Finally, there are now increasing concerns for long-term deleterious effect of COVID19 infection on individual's health, with potential chronic sequelae on the lung, brain, and cardiovascular system. Genetic cohorts can reveal essential to help deciphering the biological basis of these complications and our collection will be very well positioned to help answering such questions.