



HAL
open science

Identification de virus impliqués dans les pathologies inexplicées du patient immunodéprimé

Jacques Fourgeaud

► **To cite this version:**

Jacques Fourgeaud. Identification de virus impliqués dans les pathologies inexplicées du patient immunodéprimé. Virologie. Université Paris Cité, 2023. Français. NNT: 2023UNIP5144. tel-04816865

HAL Id: tel-04816865

<https://theses.hal.science/tel-04816865v1>

Submitted on 3 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris Cité

École doctorale 563 Médicament, Toxicologie, Chimie, Imageries
*URP 7328 Fédération pour la recherche en explorations
thérapeutiques innovantes in utero (FETUS), Université Paris Cité*

Identification de virus impliqués dans les pathologies inexplicées du patient immunodéprimé

Par **Jacques FOURGEAUD**

Thèse de doctorat de **Virologie**

Dirigée par **Marianne LERUEZ-VILLE**

Présentée et soutenue publiquement
le 18 octobre 2023

Devant un jury composé de :

Sonia BURREL, PUPH, Université de Bordeaux, Rapporteur

Sébastien HANTZ, PUPH, Université de Limoges, Rapporteur

Jérôme LE GOFF, PUPH, Université Paris Cité, Examineur

Philippe SOGNI, PUPH, Université Paris Cité, Examineur

Christelle VAULOUP-FELLOUS, MCUPH-HDR, Université Paris Saclay, Examineur

Marianne LERUEZ-VILLE, PH-HDR, Université Paris Cité, Directeur de thèse

Titre : Identification de virus impliqués dans les pathologies inexpliquées du patient immunodéprimé

Résumé : Les patients présentant un déficit immunitaire primitif ou secondaire peuvent être infectés par un large panel de pathogènes entraînant des symptômes parfois atypiques rendant le diagnostic difficile. Alors que les méthodes de culture ou de PCR universelle pour le diagnostic des infections bactériennes ont un spectre large, la détection des virus est ciblée, ce qui restreint la liste des virus détectables aux virus pathogènes les plus courants. Dans ce contexte, des pathologies qui pourraient avoir une étiologie virale restent inexpliquées. La métagénomique (mNGS) permet l'identification non ciblée d'un large éventail d'agents pathogènes, y compris les virus rares ou nouveaux qui ne sont pas détectés avec les techniques conventionnelles. Cependant les performances de cette analyse dans le cadre du soin et ses indications les plus pertinentes sont mal définies. L'objectif de cette étude est d'identifier des virus impliqués dans certaines pathologies inexpliquées du patient immunodéprimé par mNGS et d'identifier des indications de l'analyse mNGS pour le diagnostic des infections virales. Dans la première partie de ce travail, nous avons (i) mis en évidence une nouvelle pathologie associée à *Aichi virus* chez 6 patients présentant un déficit immunitaire primitif et (ii) identifié un nouveau *Circovirus* chez une patiente adulte transplantée cœur-poumon présentant une hépatite aigue inexpliquée. Dans les 2 cas nous avons présenté un ensemble d'arguments en faveur du lien de causalité entre l'infection virale et la pathologie hépatique, grâce à la mise au point de techniques de PCR, d'hybridation *in situ* et de génotypage à façon. Dans la deuxième partie, nous avons analysé les résultats de trois années d'activité de mNGS réalisée dans le cadre du diagnostic. Cette étude nous a permis d'identifier un intérêt majeur du mNGS chez les patients immunodéprimés (i) pour la détection de pathogènes opportunistes, (ii) en cas de symptômes atypiques ou (iii) en cas de confusion avec une pathologie inflammatoire ou dysimmunitaire. Cette étude nous a aussi permis de classer les prélèvements en fonction de leur rentabilité diagnostique, élevée pour les selles et faible pour le liquide cébrospinal (LCS). Nous avons montré l'intérêt de l'analyse des tissus cérébraux par rapport au LCS en cas de symptômes neurologiques inexpliqués. Nos résultats suggèrent que l'analyse des selles pourrait avoir un intérêt en cas de symptômes hépatique, hématologique ou neurologique. L'ensemble de ce travail a permis de préciser le positionnement du mNGS pour le diagnostic des infections virales.

Mots clefs : Métagénomique, *Aichi virus*, *Circovirus*, Déficit immunitaire, NGS

Title: Virus identification in immunocompromised patients with unexplained diseases

Abstract: Patients with primary or secondary immune deficiencies may be infected by a wide range of pathogens, resulting in atypical symptoms that make diagnosis difficult. While culture or universal PCR assays for bacterial detection are broad-spectrum, virus detection is targeted, limiting the list of detectable viruses to the most common pathogens. In this context, diseases that could have a viral etiology remain unexplained. Metagenomics (mNGS) enables the untargeted identification of a wide range of pathogens, including rare or novel viruses not detected by conventional tests. Despite mNGS recognition as a valuable diagnostic tool for infections, the most relevant indications for this innovative strategy remain poorly defined. The aim of this study is to identify viruses involved in unexplained diseases of the immunocompromised patient, and to identify mNGS indications in the diagnosis of viral infections. First, (i) we highlighted a novel disease associated with Aichi virus in six patients with primary immune deficiency and (ii) identified a novel Circovirus in a heart-lung transplant adult patient with unexplained acute hepatitis. In both cases, we provided evidence for a causal link between viral infection and hepatitis, using PCR, *in situ* hybridization and custom genotyping techniques. Second, we analyzed the results of three years of mNGS activity as part of our diagnostic workflow. This study enabled us to identify a major benefit of mNGS in immunocompromised patients (i) for the detection of opportunistic pathogens, (ii) in case of atypical symptoms or (iii) in case of confusion with an inflammatory or dysimmune disorder. This study also allowed us to classify samples according to the proportion of pathogens detected: high for stool and low for cerebrospinal fluid (CSF). We showed that mNGS had higher odds of detecting a pathogenic virus on cerebral tissue compared to CSF in patients with unexplained neurological symptoms. Our results suggest that stool analysis could be of interest in cases of liver, hematological or neurological symptoms. In conclusion, the results of this work contribute to refining the positioning of mNGS in the diagnostic workflow.

Keywords : Metagenomics, *Aichi virus*, *Circovirus*, Immune deficiency, NGS

Remerciements

Au Professeur Sonia BURREL, je vous remercie de me faire l'honneur d'être le rapporteur de ce travail et d'apporter votre expertise sur celui-ci. J'espère avoir l'opportunité d'échanger dans le cadre de la recherche et du soin avec le CHU de Bordeaux que vous représentez.

Au Professeur Sébastien HANTZ, je vous remercie de me faire l'honneur d'être le rapporteur de ce travail et d'apporter votre expertise sur celui-ci. Nous aurons probablement l'occasion de travailler ensemble autour des projets du CNR que vous coordonnez, et j'en suis ravi.

Au Professeur Jérôme LE GOFF, je vous remercie d'avoir accepté de juger ce travail et de m'apporter votre expertise sur celui-ci. C'est un plaisir et un honneur de vous compter parmi les membres de ce jury.

Au Professeur Philippe SOGNI, je vous remercie d'avoir accepté de juger ce travail et de m'apporter votre expertise sur celui-ci. J'espère avoir l'opportunité de travailler avec vous et votre équipe dans la recherche et le diagnostic au sein de notre groupement hospitalier APHP.Centre Université Paris Cité.

Au Docteur Christelle VAULOUP-FELLOUS, je te remercie d'avoir accepté de juger ce travail, mais aussi d'avoir participé à maintenir mon intérêt pour la virologie médicale lorsque j'étais interne à Mayotte et que tu nous encadrais, grâce à ta pédagogie, ton expertise, tes conseils et ta bienveillance.

Au Docteur Marianne LERUEZ-VILLE, je vous remercie de m'avoir encadré pour ce travail de thèse, pour votre soutien à chaque nouveau projet, pour la confiance que vous m'avez accordée dès le début de mon assistantat, pour votre expertise autant scientifique que médicale, pour vos précieux conseils et votre bienveillance au cours de ces 4 années. C'est une chance de travailler dans votre équipe.

A toute l'équipe du laboratoire de Microbiologie de l'Hôpital Necker, je vous remercie pour le cadre de travail privilégié dans lequel nous évoluons au quotidien, c'est un honneur de travailler à vos côtés.

A nos collaborateurs du laboratoire de Découverte de Pathogènes de l'Institut Pasteur de Paris, je vous remercie pour votre expertise sur le plan technique et virologique, ainsi que pour votre rigueur scientifique. J'espère et je pense que nous aurons l'opportunité de développer d'autres projets de recherche.

A l'équipe du service d'Histologie Embryologie Cytogénétique de l'Hôpital Necker, je vous remercie d'avoir mis à disposition votre plateforme d'Hybridation *in situ* pour la réalisation de notre recherche.

A l'équipe du service d'Anatomie pathologie de l'Hôpital Necker, je vous remercie pour vos conseils techniques et pour la mise à disposition de votre scanner de lame pour la réalisation de notre recherche.

Au Professeur Danielle SEILHEAN et au Docteur Franck BIELLE du service de Neuropathologie de l'Hôpital Pitié-Salpêtrière, je vous remercie d'avoir mis à disposition votre plateforme d'Hybridation *in situ* pour la réalisation de notre recherche.

Au Docteur Chantal BROUZES du service d'hématologie biologique de l'Hôpital Necker, je vous remercie d'avoir mis à disposition votre microscope pour la réalisation de notre recherche.

Sommaire

REMERCIEMENTS	3
SOMMAIRE	5
LISTE DES PRINCIPALES ABREVIATIONS.....	7
INTRODUCTION.....	9
I. CLASSIFICATION DES VIRUS	11
II. GENOMES VIRAUX.....	15
A. Génomes à ADN et ARN.....	15
B. Stratégies des virus à ADN.....	16
C. Stratégies des virus à ARN.....	17
D. Implications pour la recherche de cible de détection moléculaire	20
III. PATHOGENESE VIRALE ET ORGANES CIBLES	20
IV. BIOLOGIE MOLECULAIRE EN VIROLOGIE MEDICALE	23
A. Détection et quantification du génome par biologie moléculaire.....	23
B. Caractérisation du génome par séquençage à haut débit	24
1. Les techniques de séquençages à haut débit	24
2. Analyses ciblées	25
a) Séquençage d'amplicons de PCR	25
b) Séquençage après enrichissement par capture	26
3. Analyse non ciblée : métagénomique NGS.....	27
C. Application des tests moléculaire à la virologie	30
V. VIROME HUMAIN.....	31
VI. CRITERES D'IMPUTABILITE	35
VII. INFECTIONS VIRALES OPPORTUNISTES	37
VIII. VIRUS EMERGENTS ET DEFICITS IMMUNITAIRES.....	39
CONTEXTE ET OBJECTIFS	42
CONTEXTE	43
OBJECTIFS	44
MATERIEL ET METHODES	46
I. METAGENOMIQUE ET METATRANSCRIPTOMIQUE	47
A. Technique	47
B. Contrôles de qualité.....	50
C. Interprétation des résultats.....	50
II. PCR ET RT-PCR SPECIFIQUES	52
A. RT-PCR Aichi virus.....	52
B. PCR Circovirus.....	52
III. HYBRIDATION <i>IN SITU</i>	52
A. Hybridation <i>in situ</i> Aichi virus.....	52
B. Hybridation <i>in situ</i> Human circovirus 1	54
IV. ANALYSES PHYLOGENETIQUES	54
A. Aichi virus.....	54
B. Human circovirus 1.....	55
RESULTATS.....	56
PARTIE 1 : IDENTIFICATION DE VIRUS IMPLIQUES DANS LES PATHOLOGIES INEXPLIQUEES DU PATIENT IMMUNODEPRIME	57
Article 1: <i>Chronic Aichi virus infection as a cause of long-lasting multi-organ involvement in patients with primary immune deficiencies</i>	57
Article 2: <i>Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France</i>	58
PARTIE 2 : IDENTIFIER DES INDICATIONS POUR LE DIAGNOSTIC PAR METAGENOMIQUE CLINIQUE.....	59
Article 3: <i>Performance of clinical metagenomics: an observational prospective study in France</i>	59

DISCUSSION ET PERSPECTIVE	60
DISCUSSION.....	61
CONTRIBUTION ET CONCLUSION GENERALE	67
PERSPECTIVE.....	68
REFERENCES.....	69
TABLE DES MATIERES.....	79
TABLE DES ILLUSTRATIONS.....	81
TABLE DES TABLEAUX	82
ANNEXES.....	83
ANNEXE 1 : MATERIEL SUPPLEMENTAIRE ARTICLE 1	84
ANNEXE 2 : MATERIEL SUPPLEMENTAIRE ARTICLE 2	94
ANNEXE 3 : MATÉRIEL SUPPLÉMENTAIRE ARTICLE 3	97

Liste des principales abréviations

A

ADNc : ADN complémentaire

ADN/ARNdb : ADN/ARN bicaténaire ou double brin

ADN/ARNsb : ADN/ARN monocaténaire ou simple brin

ADV : Adénovirus

AiV : *Aichi virus*

ARNm : ARN messenger

B

BHE : Barrière hémato-encéphalique

BKV : *Polyomavirus BK*

C

CID : Déficit immunitaire combiné

CMV : Cytomégalovirus

D

DICV : Déficit immunitaire commun variable

dNTP : Désoxynucléotides triphosphates

E

EBV : Epstein-Barr virus

EEEV : Virus de l'encéphalite équine de l'Est

H

HCirV-1: *Human circovirus 1*

HHV-8: *Human herpes virus 8*

HMPV: *Human metapneumovirus*

HPyV : *Human polyomavirus*

HSCT : Transplantation de cellules souches hématopoïétiques

I

ICF1: type 1 immunodeficiency, centromeric region instability and facial anomalies

ICTV: International Committee on Taxonomy of Viruses

HIS: Hybridation *in situ*

ITS: Internal transcribed spacer

J

JCV : *Polyomavirus JC*

L

LCS : Liquide cérébrospinal

LEMP : Leucoencéphalopathie multifocale progressive

M

MERS-CoV : Middle East Respiratory Syndrome Coronavirus

mNGS : Métagénomique Clinique

N

NCBI : National Center for Biotechnology Information

NGS : Séquençage de nouvelle génération ou séquençage à haut débit

P

PCV : *Porcine circovirus*

R

RdRp : ARN polymérase ARN dépendante

RVFV : Virus de la fièvre de la vallée du Rift

S

SARS-CoV: Severe Acute Respiratory Syndrome Coronavirus

SCID : Déficit immunitaire combiné sévère

SNC : Système nerveux central

T

TTV : *Torque teno virus*

V

VEEV : Virus de l'encéphalite équine Vénézuélienne

VHB : Virus de l'hépatite B

VHC : Virus de l'hépatite C

VHE : Virus de l'hépatite E

VIH : Virus de l'immunodéficience humaine

VRS : Virus respiratoire syncytial

X

XLA : Agammaglobulinémie liée à l'X

Y

YFV : Virus de la fièvre jaune

INTRODUCTION

Le diagnostic en virologie clinique continue d'évoluer rapidement. Dans le passé, les méthodes d'identification des agents pathogènes étaient lentes et fastidieuses. Les technologies de séquençage et de bio-informatique permettent désormais d'identifier des communautés virales entières à partir de leur environnement naturel sans l'intervention de techniques d'isolement chronophages.

Les tests viraux sont essentiels pour la prise en charge des patients hospitalisés en particulier les patients immunodéprimés. En effet, les patients présentant un déficit immunitaire primaire ou secondaire peuvent être infectés par un large éventail de pathogènes opportunistes ou non. En raison de la réponse immunitaire imparfaite, ils peuvent développer des symptômes atypiques qui rendent parfois le diagnostic de la maladie difficile. De plus, ces patients peuvent être affectés par des complications majeures telles que des manifestations auto-immunes et inflammatoires, qui se présentent sous des formes variées, et induisent des symptômes pouvant être confondus avec une infection. Ces difficultés de diagnostic nécessitent également des outils de détection avec un spectre large.

Le diagnostic virologique constitue également la première ligne d'investigation contre l'émergence de nouveaux virus dans la population humaine. Dans le contexte actuel de densification de la population, de l'accès à des moyens de transport rapide et de perturbation de l'environnement favorisant les émergences virales, l'identification des agents pathogènes nécessite des outils de diagnostic à large spectre facilement accessibles au laboratoire.

Plusieurs méthodes diagnostiques peuvent être utiles, mais les tests moléculaires se sont imposés comme la technologie de diagnostic dominante. Parmi les techniques moléculaires disponibles, l'approche métagénomique est celle qui présente aujourd'hui le plus grand potentiel pour le diagnostic virologique et la découverte de nouveau virus.

Dans la première partie de ce travail de thèse nous avons identifié des virus impliqués dans des pathologies inexplicées du patient immunodéprimé et étudié le lien de causalité entre le virus et la maladie. Dans la seconde partie nous avons étudié les performances de la métagénomique pour le diagnostic des infections virales.

I. Classification des virus

Les critères utilisés pour la classification des virus et la nomenclature des taxons ont été, à l'origine et par nécessité, basés sur les propriétés d'observation des isolats de virus, notamment la morphologie des particules virales, le type d'acide nucléique dans leur génome et les attributs physiques tels que la susceptibilité à l'inactivation par des températures élevées, des solvants organiques et un faible pH.¹ La taxonomie est une classification biologique basée sur les relations évolutives dans laquelle les virus sont classés dans une série de taxons hiérarchiques (classification) avec une dénomination réglementée des taxons constitutifs (nomenclature).¹ La division Virologie de l'Union internationale des sociétés de microbiologie (IUMS) a chargé le Comité international de taxonomie des virus (ICTV) d'élaborer, d'affiner et de tenir à jour une taxonomie universelle des virus.

Au cours des dernières années, la taxonomie des virus est passée d'une niche spécialisée à un domaine interdisciplinaire qui englobe la virologie, la bio-informatique, la biologie évolutive et, parfois, des considérations philosophiques sur la signification de la taxonomie, en particulier lorsqu'il s'agit de déduire des relations évolutives plus profondes et moins convaincantes ou la nature des espèces de virus et les principes de catégorisation qui y sont associés. Ce changement de portée est largement dû à l'introduction de la technologie de séquençage à haut débit, à la reconnaissance de la véritable diversité de la virosphère et aux progrès de la phylogénétique informatique.² En 2016 l'ICTV a déclaré que les virus connus seulement par leurs séquences pouvaient être incorporés à la taxonomie virale.¹ Cette politique permet des attributions taxonomiques sans nécessiter de connaissance préalable des propriétés phénotypiques d'un virus, telles que la gamme d'hôtes ou la pathogénicité, ni l'isolement des virus dans des cultures cellulaires, ni la visualisation des virions.³ Des discussions ultérieures ont abouti à la publication de lignes directrices concernant les normes minimales applicables aux données sur les séquences virales, afin de garantir que les virus classés dans la taxonomie ICTV soient représentés par des séquences génomiques complètes assemblées avec précision et exemptes d'artefacts.^{4,5} Cette évolution a permis d'intégrer un grand nombre de nouveaux taxons dans la taxonomie officielle, principalement à partir de données génomiques issues d'études métagénomiques à grande échelle.⁶⁻⁸ Elle a également relancé le débat sur le bien-fondé de l'utilisation de critères différents pour l'attribution de taxons à différents groupes de virus. Ainsi, il a été proposé quatre principes pour guider la construction d'une taxonomie des virus cohérente et complète :¹

« i) La taxonomie des virus doit refléter l'histoire évolutive des virus. La plupart des virus peuvent être classés dans des domaines viraux indépendants, chacun ayant une origine évolutive distincte présumée. Les membres de chaque domaine possèdent des ensembles de gènes orthologues ancestraux, appelés gènes caractéristiques, qui correspondent généralement à des modules de réplication ou de formation de virions au sein de leur génome.

Leurs relations évolutives définissent des assignations taxonomiques monophylétiques au sein de chacun de ces groupes de virus.

ii) Les propriétés des virus peuvent guider l'attribution des rangs afin de maximiser leur utilité. Si les relations évolutives déterminent la topologie des taxonomies virales, les rangs qui leur sont attribués sont des constructions humaines, avec jusqu'à 15 rangs disponibles d'un règne à l'autre. Le classement des virus doit suivre des modèles de propriétés évolutives, génomiques et phénotypiques ; par exemple, les assignations d'espèces peuvent être basées sur la gamme d'hôtes, les associations de maladies ou l'épidémiologie, à condition que ces catégories aboutissent à des groupes monophylétiques.

iii) La taxonomie n'est qu'un des nombreux moyens possibles de classer les virus. La taxonomie établie par l'ICTV fournit un cadre général pour la classification des virus sur la base des relations évolutives. Toutefois, d'autres classifications fondées, par exemple, sur des propriétés cliniques ou épidémiologiques ou sur des exigences réglementaires ont leur propre utilité dans des circonstances spécifiques. Elles peuvent ne pas suivre les relations évolutives (comme la classification de Baltimore) ou inclure des catégories polyphylétiques, telles que les arbovirus ou les virus de l'immunodéficience humaine, qui ont une valeur épidémiologique ou clinique mais ne peuvent pas être représentées dans une taxonomie évolutive.

iv) L'attribution taxonomique des virus déduite des séquences métagénomiques nécessite un contrôle strict de la qualité des séquences. L'attribution d'un nouveau taxon à partir de la séquence en l'absence d'autres caractérisations du virus exige qu'elle soit à la fois précise et complète. Des lignes directrices ont été publiées concernant les informations minimales à fournir sur un génome de virus non cultivé pour l'attribution taxonomique.^{1,4} »

Les virus sont classés en taxons qui sont placés dans 15 rangs hiérarchiques (**Figure 1**); les rangs primaires sont nommés règne, sous-règne, royaume, sous-royaume, phylum, sous-phylum, classe, sous-classe, ordre, sous-ordre, famille, sous-famille, genre et sous-genre et les noms de ces rangs se terminent respectivement par les suffixes "-*viria*", "-*vira*", "-*virae*", "-*virites*", "-*viricota*", "-*viricotina*", "-*viricetes*", "-*viricetidae*", "-*virales*", "-*virinae*", "-*virinae*", "-*virus*" et "-*virus*".² A noter qu'il n'y a pas de rang inférieur à celui de l'espèce, en effet l'ICTV n'est pas responsable de la classification et de la nomenclature des taxons de virus inférieurs au rang d'espèce. La classification et la dénomination des sérotypes, génotypes, souches, variantes et isolats des espèces de virus relèvent de la responsabilité de groupes internationaux de spécialistes reconnus. Par exemple, pour les virus de la grippe A (espèce *Influenza A virus*, famille *Orthomyxoviridae*), la forme de désignation de l'isolat est la suivante: type/hôte/pays/numéro de souche/année/type d'hémagglutinine/ type de neuraminidase, ce qui donne "virus de la grippe A, isolat A/swine/Italie/81220/2017(H1N1)". Ces noms peuvent simplifier l'analyse informatique des séquences virales, mais ils peuvent aussi devenir encombrants.² La taxonomie à 15 rangs est conçue pour prendre en compte l'ensemble du

spectre de la diversité génétique et des relations évolutives dans la virosphère. À l'heure actuelle, 80% des 11 273 taxons d'espèces virales sont rattachés à un taxon familial et 94% des taxons d'espèces ont été rattachés à l'un des six règnes actuellement établis. L'attribution à des rangs autres que le genre et l'espèce n'est pas obligatoire.² En effet, il n'est pas nécessaire d'utiliser tous les rangs taxonomiques. Par exemple, dans le règne *Adnaviria*, les sous-rangs ne sont pas utilisés, et il existe de nombreux autres cas dans lesquels plusieurs rangs primaires et secondaires n'ont pas encore été attribués.² La taxonomie est une science dynamique et se développe au fil de l'évolution des techniques d'analyse. Par exemple, à partir de 1975, le genre *Rubivirus* a été inclus dans la famille des *Togaviridae* avec les membres du genre *Alphavirus*. Cependant, les *Rubivirus* (y compris le virus de la rubéole) diffèrent considérablement des *Alphavirus* tels que le virus chikungunya dans leur mode de transmission, la structure du virion et les relations de séquence. Ainsi, en 2019, le genre a été déplacé dans une famille nouvellement créée, *Matonaviridae*, qui comprend désormais des virus humains et animaux².

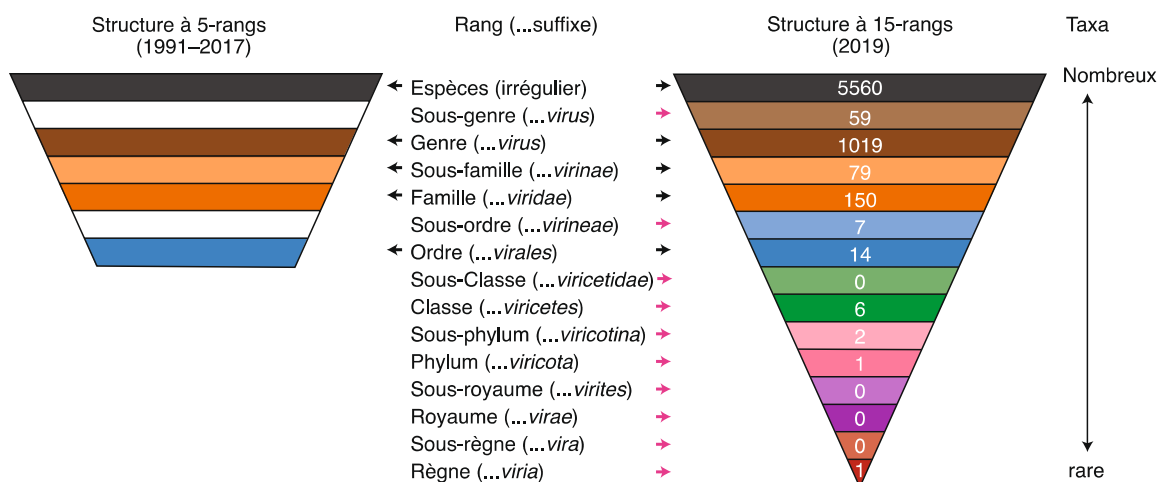


Figure 1. Comparaison de la hiérarchie des rangs taxonomiques de l'ICTV en 1991-2017 et 2019.

Les rangs taxonomiques sont indiqués par rapport au schéma de distribution des taxons. Le nombre de taxons assignés à chaque rang est indiqué en caractères blancs sur la structure à 15 rangs. Le rang de l'espèce est le rang le plus bas et le rang du règne le rang le plus élevé. Les flèches noires indiquent les rangs communs aux structures à cinq et quinze rangs ; les flèches roses indiquent les rangs introduits dans la structure à quinze rangs. Figure adaptée de Gorbalenya et al, 2020.⁹

Pour faciliter la présentation ou la mise en tableau de la taxonomie des virus, des regroupements informels de taxons par catégories sont souvent utilisés. Les critères comprennent généralement la nature du génome viral (ADN ou ARN), le brin du génome viral (simple brin ou double brin), la polarité du génome (sens positif, sens négatif ou ambisens) et la transcription inverse. Le système de classification de Baltimore est un schéma largement utilisé, basé sur la nature du génome des virions et sur la voie de synthèse de l'acide nucléique

que chaque groupe emprunte pour réaliser la synthèse de l'ARN messager (ARNm).¹⁰ Cette classification divise les virus en sept catégories (**Tableau 1**).

Tableau 1. Les classes de virus de Baltimore et leurs principales caractéristiques moléculaires et biologiques

BC	Acide nucléique du virus	Structure du génome	Taille du génome, kb ^a	Segmentation du génome	Diversité d'hôte
I	ADNdb	Principalement linéaire	5–2 500	Non segmenté	Bactéries et archées, protistes, animaux ; aucun chez les plantes, rare chez les champignons
II	ADNsb	Principalement circulaire	1,7–25	Principalement non segmenté	Bactéries, rares dans les archées ; la plupart des eucaryotes
III	ARNdb	Linéaire	4–30	Principalement segmenté	Protistes, animaux, plantes ; une famille chez les bactéries, aucune chez les archées
IV	ARN(+)	Linéaire	3,5-40	Principalement non segmentés, mais souvent segmentés	Tous chez les eucaryotes ; une classe avec six familles chez les bactéries, aucune chez les archées
V	ARN(-)	Principalement linéaire	1,7-20	Environ la moitié segmentée	Animaux, plantes, rare chez les champignons, les protistes ; aucun chez les bactéries ou les archées
VI	ARN(+), RT	Linéaire	5-13	Non segmenté	Tous chez les eucaryotes ; aucun chez les bactéries ou les archées
VII	ADNdb, RT	Circulaire	3-10	Non segmenté	Animaux, plantes ; inconnu chez les protistes ; aucun chez les bactéries ou les archées

^aLes informations sur la taille des génomes proviennent du dixième rapport de l'ICTV. BC, Classification de Baltimore ; db, bicaténaire ; sb, monocaténaire ; (+), polarité positive ; (-), polarité négative. RT, Reverse transcription. Tableau adapté de Koonin et al, 2021.¹¹

II. Génomes viraux

A. Génomes à ADN et ARN

L'aspect le plus frappant des virus au niveau moléculaire est la diversité des structures de leur génome et de leurs stratégies de réplication. Contrairement aux génomes cellulaires, qui sont uniformément constitués d'ADN bicaténaire (ADNdb), les génomes viraux offrent des exemples de toutes les variations structurelles. Les génomes des différentes familles de virus sont constitués d'ADN ou d'ARN bicaténaire (db) ou monocaténaire (sb), de polarité positive (+), négative (-) ou ambisens, de topologie linéaire ou circulaire, et comprennent un ou plusieurs segments. Chaque variation a des conséquences sur les voies de réplication du génome, l'expression des gènes viraux et l'assemblage des virions (**Figure 2**). La taxonomie virale prend en compte ces variations structurelles, ainsi les virus sont classés dans l'un des six règnes suivants : *Adnaviria* (virus filamenteux à ADNdb des archées), *Duplodnaviria* (virus à ADNdb), *Monodnaviria* (virus à ADNdb), *Riboviria* (virus à ARN), *Ribozyviria* (génomes circulaires à ARN négatif) et *Varidnaviria* (ADNdb). Parmi les familles de virus qui infectent les vertébrés, ceux dont le génome est constitué d'ARN sont environ deux fois plus nombreux que ceux dont le génome est constitué d'ADN. Les virus à ADNdb ne sont pas connus chez les plantes, à l'exception de ceux qui, comme les *Hepadnaviridae* des vertébrés, se répliquent par transcription inverse. Cette observation reste inexplicée, mais elle atteste le succès évolutif et de la polyvalence de l'ARN en tant que matériel génétique pour les petits génomes. Les taux d'erreur élevés de la réplication de l'ARN limitent la taille des génomes d'ARN à 30 kb ou moins, alors que la relecture et la réparation des erreurs assurent une réplication suffisamment précise des génomes de virus à ADN aussi grands que ceux des megavirus de 1200 kb et des pandoravirus de 2400 kb. En outre, le fait que l'ADN soit chimiquement plus stable que l'ARN explique probablement pourquoi tous les virus connus des hôtes thermophiles ont des génomes d'ADNdb.¹²

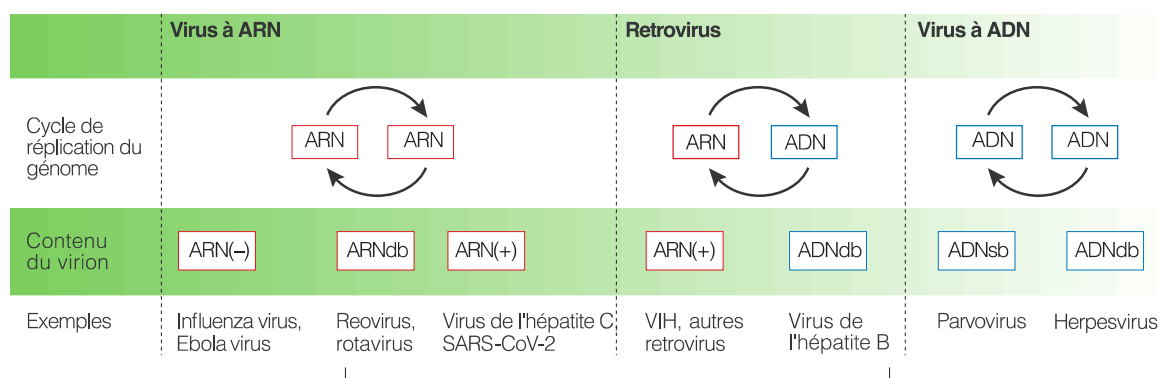


Figure 2. Sept catégories de virus se distinguant par leurs stratégies de réplication du génome

ARN (+), ARN à brin positif, qui est un ARN monocaténaire de la même polarité que l'ARNm viral. ARN (-), ARN à brin négatif, qui est un ARN monocaténaire de polarité opposée à l'ARNm viral ; ARNdb, ARN bicaténaire ; SARS-CoV-2, Coronavirus du Syndrome respiratoire aigu sévère 2. Figure adaptée de Paul Ahlquist, 2006.¹³

B. Stratégies des virus à ADN

La taille des génomes des virus à ADN varie de 1,8 kb pour les circovirus à 2400 kb pour les pandoravirus.^{12,14} Cette grande différence signifie que les virus de différentes familles varient considérablement dans le nombre de fonctions nécessaires à la réplication virale qu'ils codent eux-mêmes. Par exemple, les virus à ADN dotés de petits génomes tels que les polyomavirus, les papillomavirus et les parvovirus utilisent les enzymes de la cellule hôte pour la transcription et la réplication. Ceux dont le génome est de taille intermédiaire (jusqu'à 35 kb), comme les adénovirus, codent une grande partie de leur machinerie de réplication de l'ADN, y compris une ADN polymérase, une protéine terminale et une protéine de liaison à l'ADN monocaténaire, mais ils utilisent les ARN polymérases II et III cellulaires pour la transcription. Ceux dont le génome est plus grand (150-350 kb), comme les herpesvirus et les poxvirus, codent pour des ADN polymérases et des protéines de liaison. Dans le cas des herpesvirus, de multiples facteurs de transcription spécifiques modifient en série la spécificité du promoteur de l'ARN polymérase II (**Figure 3**), ou dans le cas des poxvirus, des complexes de transcriptase à plusieurs sous-unités assurent toutes les fonctions de coiffe et de polyadénylation de l'ARNm ainsi que la transcription de l'ARN. Les hepadnavirus vont à l'encontre de cette tendance générale en ce sens que leur génome est petit (3 kb) mais qu'il code pour l'ADN polymérase/transcriptase inverse qui exécute leur mécanisme unique de réplication de l'ADN par l'intermédiaire d'un ARN monocaténaire.

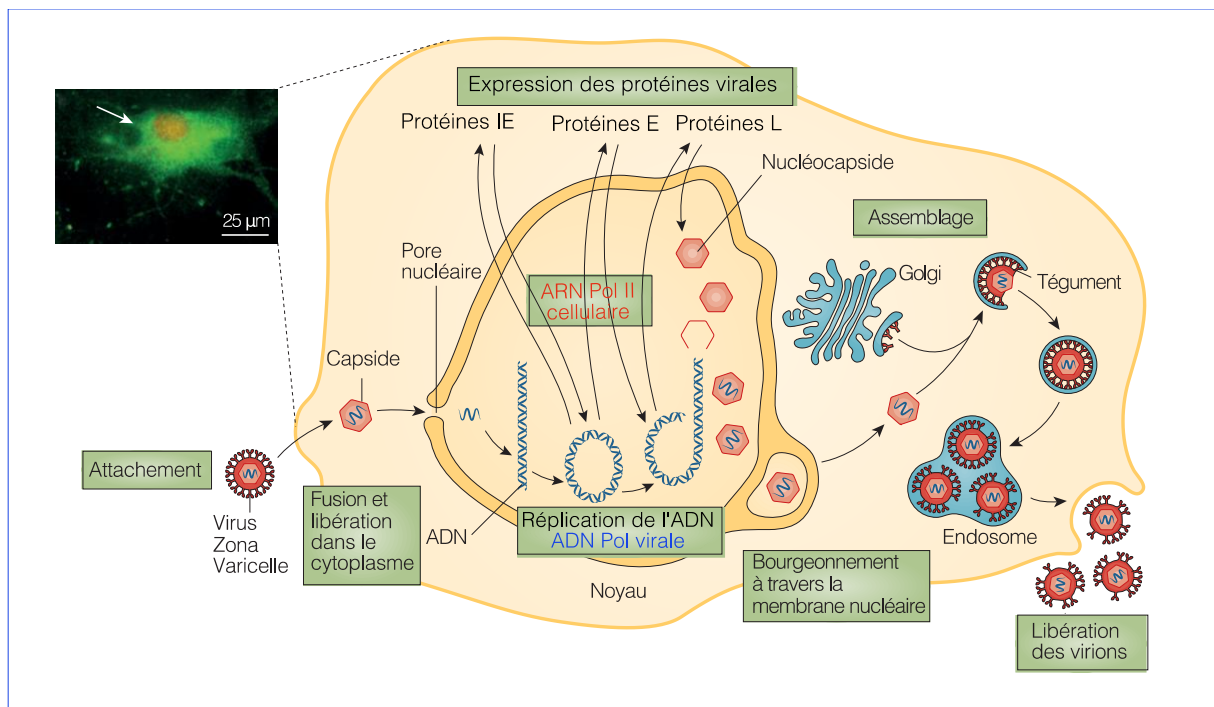


Figure 3. Cycle lytique du virus Zona Varicelle (Herpesviridae) lors de l'infection d'un neurone.

IE, Immediate early; E, Early; L, Late. Figure adaptée de Gershon et al, 2015¹⁵ et de Gatherer, 2020.¹⁶

La synthèse de l'ADN cellulaire a lieu pendant la phase S du cycle cellulaire et pas du tout dans les cellules G0 en phase terminale de différenciation, les virus qui dépendent de l'ADN

polymérase de l'hôte peuvent établir deux stratégies. Soit attendre que la cellule infectée entre spontanément en phase S, comme dans le cas des parvovirus. Soit exprimer un ou plusieurs oncogènes viraux pour outrepasser la régulation des protéines de contrôle du cycle cellulaire p53 ou pRb et stimuler ainsi les cellules infectées à entrer en phase S, comme dans le cas des polyomavirus et des papillomavirus. Les virus dotés de génomes d'ADN de grande taille (par exemple, les herpesvirus et les poxvirus) codent eux-mêmes les enzymes et peuvent donc se répliquer dans des cellules qui ne se divisent pas et dans d'autres environnements qui ne permettraient normalement pas la réplication de l'ADN, tels que les cellules différenciées du système nerveux (*Herpesviridae*, **Figure 3**) ou même le cytoplasme (poxvirus). Bien que ces gènes viraux soient souvent inutiles pour la réplication du virus dans les cellules en culture qui se divisent activement, ils peuvent exercer une influence profonde sur la virulence virale dans les organismes infectés et constituer ainsi des cibles pour l'intervention chimiothérapeutique. Par exemple, le gène de la thymidine kinase de certains herpesvirus (mais pas l'enzyme hôte) phosphoryle la prodrogue acyclovir pour générer un analogue dNTP qui met fin aux brins naissants pendant la synthèse de l'ADN.¹⁷

Outre les cycles typiques de réplication, de nombreux virus à ADN établissent des infections latentes ou persistantes de leurs hôtes. Plusieurs mécanismes distincts de persistance ont été identifiés avec différents virus, mais ils impliquent tous la suppression des effets cytopathiques viraux, le maintien à long terme du génome viral et l'évasion des défenses cellulaires et de l'organisme. Par exemple, les herpesvirus établissent généralement des infections latentes dans lesquelles le génome viral est maintenu sous forme d'épisome circulaire dans le noyau, n'exprimant tout au plus que quelques gènes viraux et ne produisant pas de virus infectieux.^{18,19} Ces infections herpétiques latentes persistent tout au long de la vie de l'hôte, échappant avec succès à la surveillance immunitaire de l'hôte, tout en étant capables de réapparaître à intervalles réguliers sous forme d'infections lytiques productives.^{15,17,19} Le génome d'autres virus peuvent persister en s'intégrant au génome humain dans des régions télomériques (HHV-6) ou de façon aléatoire (HIV, HTLV).¹⁸

C. Stratégies des virus à ARN

Le type de génome ARN dicte les premières étapes de biosynthèse après l'infection. Par exemple, les virus à ARN à polarité positive (ARN+), à l'exception des rétrovirus, livrent tous initialement leur génome aux ribosomes pour assurer la synthèse des protéines essentielles à la réplication virale. Par conséquent, seul l'ARN génomique viral est infectieux une fois introduit dans une cellule hôte, ce qui a grandement facilité la manipulation génétique de ces virus. Contrairement aux génomes des virus à ARN (+), ceux des virus à ARN (-), des rétrovirus et des virus à ARN à double brin doivent introduire dans la cellule un complexe protéine-ARN qui contient en son sein la polymérase virale (**Figure 4**).

Les virus à ARN positif se répartissent en deux groupes, ceux qui transcrivent des ARN sous-génomiques et ceux qui ne le font pas.

Les virus qui ne transcrivent pas d'ARN sous-génomique, comme les virus de la famille des *Picornaviridae* ou des *Flaviviridae*, sont traduits par les ribosomes de l'hôte pour produire un ou plusieurs précurseurs polyprotéiques qui subissent une série de réactions de clivage *in cis* et *in trans* par des protéases codées par le virus.²⁰ Ces réactions de clivage aboutissent à la production des différentes protéines non structurales essentielles à la réplication du génome viral. Pour les virus qui produisent une seule polyprotéine, les précurseurs des composants catalytiques sont générés en quantités équimolaires aux protéines structurales. Dans certains cas, le contrôle de l'abondance des composants catalytiques est assuré par l'accumulation d'intermédiaires dans la cascade de traitement dans lesquels une ou plusieurs activités catalytiques sont absentes.²¹ Les virus qui produisent deux précurseurs polyprotéiques utilisent toujours un site d'entrée interne du ribosome (IRES) pour conduire l'expression de la seconde polyprotéine. La traduction de la seconde polyprotéine étant indépendante de la première, les protéines ne sont pas produites en quantités équimolaires, ce qui permet une certaine régulation de l'abondance des différentes protéines.²²

Pour les virus qui produisent des ARN sous-génomiques, comme les virus de la famille des *Astroviridae* ou des *Coronaviridae*, le génome d'entrée est d'abord traduit en un précurseur des protéines non structurales qui comprend l'ARN polymérase ARN-dépendante (RdRp). Le génome est ensuite copié dans un ou plusieurs ARNm sous-génomiques qui codent pour les protéines structurales. La production d'ARNm sous-génomiques facilite la synthèse de quantités distinctes de protéines structurales et non structurales, de sorte que les protéines structurales sont souvent produites en excès molaire par rapport aux protéines non structurales essentielles sur le plan catalytique.¹⁷

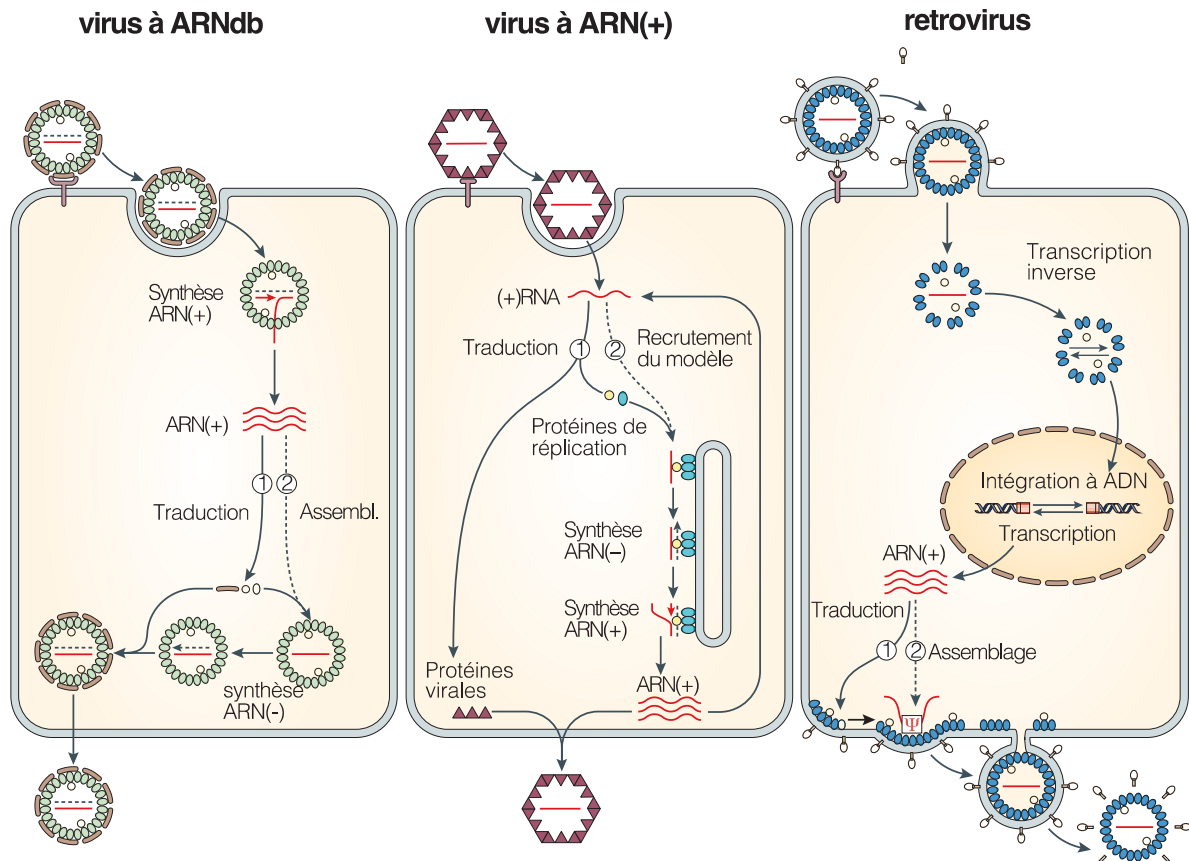


Figure 4. Résumé simplifié des cycles des virus à ARNdb, des virus à ARN (+) et des rétrovirus.
 ARNdb, ARN bicaténaire ; ARN(+), ARN de polarité positive ; ARN(-), ARN de polarité négative. Figure adaptée de Paul Ahlquist, 2006.¹³

Contrairement à l'ARN seul des virus à ARN(+), celui des virus à ARN(-), ambisens et ARNdb n'est pas infectieux. En effet, l'unité infectieuse est un complexe ribonucléoprotéique comprenant l'ARN génomique associé aux composants de la polymérase virale nécessaires à la synthèse de l'ARNm.^{23,24} Dans le cas des virus à ARN(-) et ambisens, l'ARN génomique d'entrée est copié par le complexe de la polymérase virale en ARNm. Les génomes d'entrée sont ensuite répliqués pour donner des antigénomes, un processus qui nécessite une protéine de nucléocapside virale nouvellement synthétisée pour recouvrir le brin d'ARN naissant. Pour les virus à ARN(-), les antigénomes ont un sens positif, mais ils ne servent pas de modèles pour la traduction, ils fonctionnent exclusivement comme modèles pour la répllication génomique. Les orthomyxovirus et bornavirus à polarité négative se répliquent dans le noyau de la cellule hôte, tandis que les rhabdovirus infectant les mammifères ainsi que les paramyxovirus et les filovirus se répliquent dans le cytoplasme.^{25,26} Une autre distinction importante est que les rhabdovirus, filovirus, Paramyxovirus et bornavirus transcrivent séquentiellement une série de 5 à 10 ARNm monocistroniques à partir d'un seul modèle génomique, alors que les orthomyxovirus ne synthétisent qu'un seul ARNm à partir de chaque segment. Contrairement à tous les autres virus à ARN à brin négatif, le génome à ARNs circulaire du VHD ne nécessite pas de polymérase virale spécialisée pour être copié. Au contraire, le génome du VHD est transcrit et répliqué par l'ARN polymérase II dépendante de l'ADN de l'hôte dans une réaction unique basée sur l'ARN.²⁷

Dans le cas des virus à ARNdb comme les rotavirus, les génomes segmentés sont introduits dans la cellule sous la forme d'une particule sous-virale qui reste intacte pour la transcription de l'ARNm. La polymérase et les enzymes modifiant l'ARN forment un élément structural à travers lequel le brin naissant d'ARNm passe et est modifié par cotranscription. L'ARNsb et ses protéines codées se réassemblent en de nouvelles particules sous-virales qui peuvent diriger la synthèse de l'ARN de l'antigénome et former un génome d'ARNdb. Ces particules sous-virales descendantes peuvent servir de modèles pour la production d'autres ARNm et la protéine virale ou peuvent être emballées dans de nouveaux virions (**Figure 4**).^{13,28}

Les génomes d'ARN positif des rétrovirus utilisent un mécanisme distinct pour la réplication. Ici, l'ARNsb entrant est copié en un provirus à ADNdb à l'aide de la transcriptase inverse codée par le virus. Cet ADNdb est intégré dans le génome de l'hôte, où il est copié en transcrits épissés de façon différentielle qui servent de modèle pour la production de protéines virales ou pour la production de nouveaux virions (**Figure 4**).^{13,29}

D. Implications pour la recherche de cible de détection moléculaire

Ces variations structurales des génomes des virus et les multiples stratégies de réplication et de transcription qui en découlent, impliquent de faire des choix judicieux quant aux cibles des techniques de détection moléculaire en virologie. Pour les techniques de détection par PCR spécifique ou pour les techniques de séquençage ciblé, l'ADN ou l'ARN génomique est une cible privilégiée. Pour les techniques non ciblées comme la métagénomique ou la métatranscriptomique, il peut être pertinent de prétraiter les échantillons cliniques afin de ne conserver que les ARN totaux, ce qui permet de s'affranchir de l'ADN de l'hôte, des bactéries et des champignons, pour ne garder que leurs transcrits. De plus, cette approche permet d'apporter un argument en faveur de la réplication active de l'agent pathogène (en plus de la détection et de la description des séquences). En effet, quelles que soient la structure et la stratégie de réplication de leur génome, tous les virus doivent exprimer leurs gènes sous forme d'ARNm.

III. Pathogénèse virale et organes cibles

L'étude de la pathogénèse virale dans l'organisme consiste à diviser le processus pathologique induit par le virus en une série d'étapes, dont l'infection initiale, la dissémination virale, l'invasion du ou des tissus cibles et l'induction de la maladie, suivie de résultats tels que la résolution de la maladie, la clairance virale, la persistance virale, la maladie chronique, voire la mort de l'hôte. En divisant ces processus en étapes distinctes, il est possible de comprendre comment les différentes variables du virus et de l'hôte affectent chaque étape du processus

d'infection et de maladie, et donc dans quel(s) tissu(s) se trouve le virus pour faire le diagnostic de la maladie. Par exemple le poliovirus est transmis par voie fécale-orale, le virus pénètre dans les plaques de Peyer de l'intestin où il se réplique dans les cellules lymphoïdes. Le virus se propage ensuite aux sites de réplication secondaires *via* le sang, où il se réplique et déclenche un deuxième cycle de virémie qui peut conduire à une infection du système nerveux central (SNC) chez certains individus, soit par le passage du virus à travers la barrière hémato-encéphalique (BHE), soit par l'infection des nerfs périphériques qui assurent la médiation de l'entrée du virus dans le SNC.³⁰⁻³³ En pénétrant dans le SNC, le virus se réplique au sein des motoneurones, provoquant leur destruction et entraînant ainsi la paralysie de l'hôte.

Il est important de noter que des virus non apparentés qui provoquent des manifestations cliniques similaires peuvent passer par un ensemble similaire de stades de la maladie. Par exemple, les virus encéphalitiques, qu'il s'agisse de virus à ADN comme le HSV ou de virus à ARN, y compris les poliovirus, le virus de la rage, ou le virus de l'encéphalite équine vénézuélienne (VEEV), doivent tous envahir le SNC pour provoquer une maladie neurologique. Si la neuroinvasion est une étape commune du processus pathologique pour tous les virus encéphalitiques, les mécanismes spécifiques ou même les voies de neuroinvasion varient considérablement d'un virus à l'autre. Le HSV pénètre dans le SNC par transport neuronal rétrograde *via* les nerfs périphériques tels que les nerfs olfactifs.^{34,35} Le virus de la rage pénètre également dans le SNC par transport rétrograde dans les nerfs périphériques après absorption du virus au niveau des jonctions neuromusculaires.³⁶ Le VEEV pénètre initialement dans le SNC *via* les nerfs olfactifs ou le ganglion trigéminal ;³⁷ cependant, une deuxième série de neuroinvasion se produit suite à la rupture de la BHE ; l'inhibition de la rupture de la BHE retarde l'apparition de la maladie induite par le VEEV.³⁸ Le poliovirus peut également pénétrer dans le SNC soit en traversant directement la BHE, soit en infectant les nerfs périphériques.³² Par conséquent, bien que les virus qui provoquent des états pathologiques similaires puissent passer par des étapes identiques dans le processus pathologique, les mécanismes moléculaires spécifiques utilisés par chaque virus pour achever chaque étape peuvent être très différents. Ces mécanismes permettent de comprendre pourquoi dans certains cas la détection du virus sera possible seulement dans les tissus cérébraux tandis qu'elle sera négative dans le liquide cébrospinal (LCS), mais également pourquoi certains virus comme les poliovirus seront facilement détectés dans les selles des patients malgré une symptomatologie neurologique.

Les arbovirus transmis par les moustiques, qui comprennent des membres de l'ordre des *Bunyavirales* et des familles *Flaviviridae* et *Togaviridae*, illustrent également le fait que l'infection virale peut être divisée en étapes distinctes présentant des caractéristiques communes et provoquant des maladies très similaires mais par le biais de mécanismes différents. Ces arbovirus déclenchent tous une infection par voie transcutanée dans la salive des arthropodes et, bien que leurs cellules cibles initiales diffèrent, l'infection virale initiale et l'amplification se produisent dans les cellules résidentes de la peau (**Figure 5**).³⁹ La réplication

virale locale entraîne une virémie sérique qui facilite la transmission à de nouveaux moustiques vecteurs si une charge virale suffisamment élevée est atteinte dans le sang.^{33,40-42} Cette virémie facilite également la dissémination virale vers des sites secondaires de réplication virale, y compris les cibles en aval de la maladie induite par le virus. Les états pathologiques et les tissus cibles associés aux différents arbovirus peuvent également être très similaires. Une grande variété d'arbovirus des trois familles virales provoque des maladies neurologiques et, comme indiqué ci-dessus, la neuroinvasion représente donc une étape clé de la maladie pour chacun de ces virus. De même, des virus non apparentés, tels que le virus de la fièvre jaune (YFV) et le virus de la fièvre de la vallée du Rift (RVFV), peuvent cibler des tissus similaires (par exemple, le foie) et provoquer des manifestations similaires, telles que l'hépatite et l'hémorragie.^{43,44} Dans le cas des arbovirus la recherche directe de l'ARN sera donc plus pertinente dans le plasma des patients au moment de la phase virémique commune à la plupart de ces virus.⁴⁵

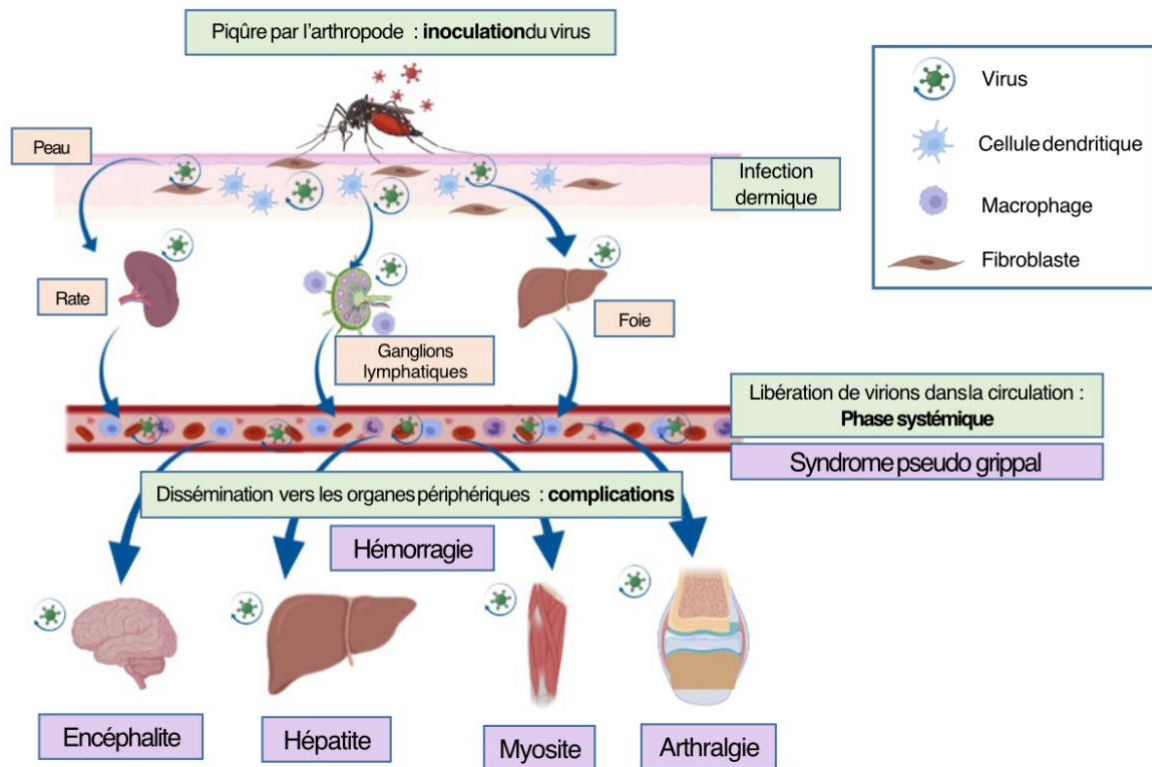


Figure 5. Illustration du cycle des Arbovirus dans l'organisme et des différentes phases de l'infection
Figure adaptée du Traité de Virologie Médicale, 2e édition.

Les observations cliniques peuvent fournir des informations fondamentales sur la pathogénèse, les sites de réplication et d'excrétion virales. Par exemple, une partie de notre compréhension de la maladie induite par le virus de la rougeole provient d'études d'une épidémie de rougeole dans les îles Féroé en 1886^{33,46,47} à une époque où la cause virale de la rougeole n'avait pas encore été identifiée. De même, l'énorme quantité de données cliniques et épidémiologiques générées par la pandémie de SARS-CoV-2 a permis de comprendre le rôle

des affections sous-jacentes et de l'âge avancé dans la susceptibilité à la maladie, tandis que le succès des traitements anti-inflammatoires dans la réduction de la gravité de la maladie a illustré le rôle de l'inflammation pulmonaire en tant que moteur de la maladie chez certaines personnes infectées par le virus.^{48,49} Il est important de noter qu'au cours des quinze dernières années, la disponibilité de techniques d'analyse plus avancées pour étudier les processus pathologiques et les réponses de l'hôte dans les populations de patients, telles que les méthodes à haut débit pour quantifier l'expression génique à l'échelle du génome (par exemple, la protéomique et le séquençage de l'ARN), permet désormais de générer une image plus détaillée des événements moléculaires qui se produisent chez un patient ou dans un tissu infecté.

La capacité à établir un diagnostic viral spécifique dépend de la qualité de l'échantillon reçu par le laboratoire. Les variables importantes comprennent le moment où le prélèvement est effectué par rapport à la maladie du patient, le type de prélèvement, la qualité et la quantité de matériel prélevé, ainsi que le temps et les conditions de transport jusqu'au laboratoire. L'optimisation des variables énumérées exige des connaissances et une attention particulière aux considérations logistiques. Pour le diagnostic des infections virales aiguës, les meilleurs échantillons sont généralement prélevés sur le site de la maladie. Par exemple, dans le cas d'une suspicion de méningite virale, le LCS est le meilleur échantillon. Dans les infections impliquant la peau ou les muqueuses, les échantillons prélevés sur ces surfaces sont généralement les seuls nécessaires. Les quantités de virus sont souvent les plus élevées au début d'une maladie aiguë, de sorte que les échantillons prélevés dans les premiers jours suivant l'apparition de la maladie ont le plus de chances d'être positifs. Dans les phases tardives de la maladie il faudra se tourner vers des voies d'excrétions des virus ou de leur matériel génétique, par exemple les urines peuvent être considérées à distance de la date de début des symptômes des infections avec le virus de la rougeole ou par le virus zika,^{45,50,51} tandis que les selles peuvent être considérées pour le diagnostic rétrospectif de méningites à entérovirus.⁵²

IV. Biologie moléculaire en virologie médicale

A. Détection et quantification du génome par biologie moléculaire

La virologie diagnostique a été révolutionnée par l'application de techniques de détection des acides nucléiques, qui détectent des séquences spécifiques d'ADN ou d'ARN, et qui peuvent être appliquées à la détection de pratiquement n'importe quel virus. En fonction de la séquence cible, les tests peuvent être spécifiques d'une seule espèce de virus ou d'un groupe de virus apparentés. Cette dernière caractéristique est particulièrement avantageuse, car elle permet d'appliquer les techniques de détection des acides nucléiques à des groupes de virus (par exemple, les entérovirus/rhinovirus) pour lesquels la diversité antigénique a empêché l'application des techniques de détection des antigènes. Par exemple, les entérovirus, pour

lesquels il n'existait pas de méthode de détection non moléculaire rapide, peuvent être détectés à l'aide de tests PCR de transcription inverse qui amplifient une séquence hautement conservée dans la région 5' non traduite du génome.^{53,54} Grâce à la miniaturisation qui a permis de diminuer le temps d'analyse et de multiplexer les tests, ainsi que l'automatisation qui a permis d'augmenter le volume des prélèvements analysés et le rendu des résultats au fil de l'eau, ces techniques restent d'actualité et sont indispensables au diagnostic.

B. Caractérisation du génome par séquençage à haut débit

1. *Les techniques de séquençages à haut débit*

L'évolution de la technologie de séquençage a été l'un des principaux facteurs à l'origine des progrès réalisés dans la découverte des virus. Pendant plus de 30 ans, depuis ses débuts à la fin des années 1970 jusqu'à la fin des années 2000, le séquençage Sanger a été la principale technologie de séquençage utilisée, et tous les efforts déployés pour identifier les virus à l'aide de stratégies basées sur le séquençage reposaient sur cette technologie. À son débit le plus élevé, le séquençage Sanger automatisé à 96 capillaires pouvait générer jusqu'à 100 000 pb de séquence en un seul passage. À partir de 2005, la première des plateformes dites de séquençage de nouvelle génération (NGS) a été introduite.⁵⁵ Ces plateformes ont considérablement augmenté la quantité de séquences pouvant être produites. Par exemple, la plateforme Roche/454, dans son format initial, était capable de générer 100 000 séquences ("lectures" ou "reads") d'une longueur de 100 pb en une seule fois, soit 100 fois plus de séquences qu'une lecture sur l'instrument standard précédent. Au cours des années suivantes, ces plateformes ont rapidement évolué et les séquenceurs les plus performants étaient capables de générer des milliards de lectures en moins d'une journée. Aujourd'hui, selon la plateforme de séquençage à haut débit utilisée, les lectures générées ont des caractéristiques différentes, notamment leur longueur et le nombre d'erreurs qu'elles contiennent. Les séquenceurs à lectures courtes (« short reads ») produisent des lectures entre 75 et 500 nucléotides et les séquenceurs à lectures longues (« long reads ») produisent des lectures de 1 000 à > 30 000 nucléotides. Les plateformes NGS à lectures courtes les plus courantes reposent sur la technologie d'Illumina et les technologies de séquençage à lectures longues reposent sur les technologies de Pacific Biosciences ou d'Oxford Nanopore. Plus les lectures sont longues, plus elles vont permettre de résoudre les problèmes posés par les séquences répétées de l'ADN, et notamment l'assemblage des régions contenant ces séquences répétées. La technologie actuellement l'a plus utilisée pour toutes les analyses de séquençage (génomique, métagénomique, transcriptomique, capture) est celle d'Illumina dont les séquenceurs produisent des lectures courtes (en général 150 nucléotides de long). L'analyse NGS qui repose sur un séquençage Illumina permet une grande profondeur de séquençage en atteignant plusieurs dizaines ou centaines de millions de lectures courtes de 150 nucléotides par échantillon clinique mais nécessite plusieurs jours de temps technique. La technologie d'Oxford Nanopore présente des caractéristiques très différentes car elle peut

générer des lectures de plusieurs kilobases à plusieurs mégabases en seulement quelques heures.⁵⁶ Ceci implique que les techniques d'extraction (longueur des lectures proportionnelle à la qualité des acides nucléiques) et les outils pour l'assignation taxonomique doivent être spécifiques.⁵⁷ Le séquençage nanopore est une technologie plus récente que le séquençage Illumina et permet l'analyse en temps réel des données de séquençage.⁵⁸ En tant que telle, elle est particulièrement applicable au diagnostic car le temps est un facteur essentiel pour le succès du traitement des patients atteints de maladies infectieuses aiguës. À ce jour, les instruments disponibles dans le commerce pour le séquençage par nanopores sont ceux d'Oxford Nanopore Technologies et comprennent le MinION (1 cellule d'écoulement), le GridION (capacité de 5 cellules d'écoulement) et le PromethION (capacité de 48 cellules d'écoulement). À l'instar des plateformes Illumina, la technologie Nanopore a démontré ses performances ces dernières années, notamment dans son application en métagénomique. Grâce à cette technologie la détection par métagénomique de pathogènes a été possible en moins de 6 heures de temps d'exécution de la prise en charge de l'échantillon au résultat. Des études de recherche ont également démontré le potentiel clinique du séquençage par nanopore dans la détection des virus épidémiques et la surveillance microbiologique.⁵⁹⁻⁶¹ Les approches non ciblées telles que la métagénomique ou le RNA-seq du transcriptome entier peuvent toutefois être limitées par le débit de séquençage plus faible de nanopore par rapport au séquençage à lecture courte tel que celui d'un instrument Illumina.⁶²

2. Analyses ciblées

Le séquençage des acides nucléiques viraux, qu'ils proviennent de cultures ou directement d'échantillons cliniques, est compliqué par la présence d'ADN de l'hôte. Actuellement, le séquençage du génome des virus par NGS peut être réalisé par deux méthodes principales : le séquençage d'amplicons de PCR et le séquençage par enrichissement de cible (**Figure 6**).⁶³

a) Séquençage d'amplicons de PCR

L'amplification par PCR du matériel génétique viral à l'aide d'amorces complémentaires d'une séquence de nucléotides connue est l'approche la plus courante pour enrichir en petits génomes viraux (**Figure 6**). Cela est possible pour les virus qui ont des génomes relativement petits qui ne nécessitent que quelques amplicons de PCR pour assembler les séquences en génomes entiers. Toutefois, l'hétérogénéité des virus à ARN, tels que le VHC,⁶⁴ le norovirus,⁶⁵ le virus de la rage⁶⁶ et le VRS,⁶⁷ peut nécessiter l'utilisation de plusieurs séries d'amorces qui se chevauchent pour garantir l'amplification de tous les génotypes. Le séquençage d'amplicons par PCR est plus efficace pour le séquençage du génome entier à partir d'échantillons présentant de faibles concentrations de virus que les méthodes métagénomiques,⁶⁴ bien que d'autres méthodes telles que l'enrichissement de cibles de séquences virales puissent fonctionner aussi bien dans de tels échantillons, comme cela a été démontré pour les échantillons de *Norovirus*.⁶⁸ Des PCR chevauchantes combinées au NGS ont

été utilisées pour séquencer les génomes entiers plus longs, tels que celui du CMV, mais cette méthode est peu évolutive, car elle nécessite de nombreuses amorces et une quantité relativement importante d'ADN de départ.⁶⁹ Cela limite le nombre d'échantillons appropriés disponibles ainsi que les génomes qui peuvent être étudiés à l'aide de cette méthode.

Un exemple d'approche ciblée est l'utilisation d'amorces dans des régions hautement conservées du génome pour la détection de tous les micro-organismes correspondant à un type spécifique à partir d'échantillons cliniques, comme l'amplification du gène de l'ARN ribosomique 16S pour les bactéries⁷⁰ et l'amplification du gène de l'ARNr 18S et de l'espaceur transcrit interne (ITS) pour les champignons.⁷¹

Cette approche ciblée est principalement utilisée en virologie pour séquencer des amplicons obtenus grâce à la conception d'amorces couvrant le génome entier pour la récupération des génomes viraux directement à partir d'échantillons cliniques.^{59,72,73}

b) Séquençage après enrichissement par capture

Les méthodes d'enrichissement des cibles par capture peuvent être utilisées pour séquencer des génomes viraux entiers directement à partir d'échantillons cliniques sans qu'il soit nécessaire de procéder à une culture ou à une PCR préalable (**Figure 6**).⁷⁴ Ces méthodes font généralement appel à de petites sondes d'ARN ou d'ADN complémentaires de la séquence de référence de l'agent pathogène (ou d'un panel de séquences de référence). Lors de la réaction d'hybridation, les sondes, qui sont liées à une phase solide, capturent des séquences d'ADN complémentaires à partir de l'ensemble des acides nucléiques présents dans un échantillon. La capture est suivie de la ligature de l'adaptateur spécifique au séquenceur et d'un petit nombre de cycles de PCR pour enrichir les fragments ligaturés avec succès. Le succès de cette méthode dépend des séquences de référence disponibles pour le virus en question ; la spécificité augmente lorsque les sondes sont conçues par rapport à un panel plus large de séquences de référence, car cela permet de mieux capturer la diversité dans et entre les échantillons. L'enrichissement de la cible est possible malgré de petites discordances entre la matrice et la sonde. Toutefois, si une sonde échoue, les régions internes et chevauchantes peuvent encore être capturées par d'autres sondes. L'enrichissement de cible n'est pas adapté à la caractérisation de nouveaux virus présentant une faible homologie avec des virus connus, pour lesquels la métagénomique et, dans certains cas, la PCR utilisant des amorces dégénérées, qui sont un mélange d'amorces similaires mais variables, peuvent s'avérer plus appropriées. Des techniques d'enrichissement qui utilisent des sondes d'ARN ou d'ADN dégénérées pour capturer des centaines d'espèces virales ont également été mises au point.⁷⁵

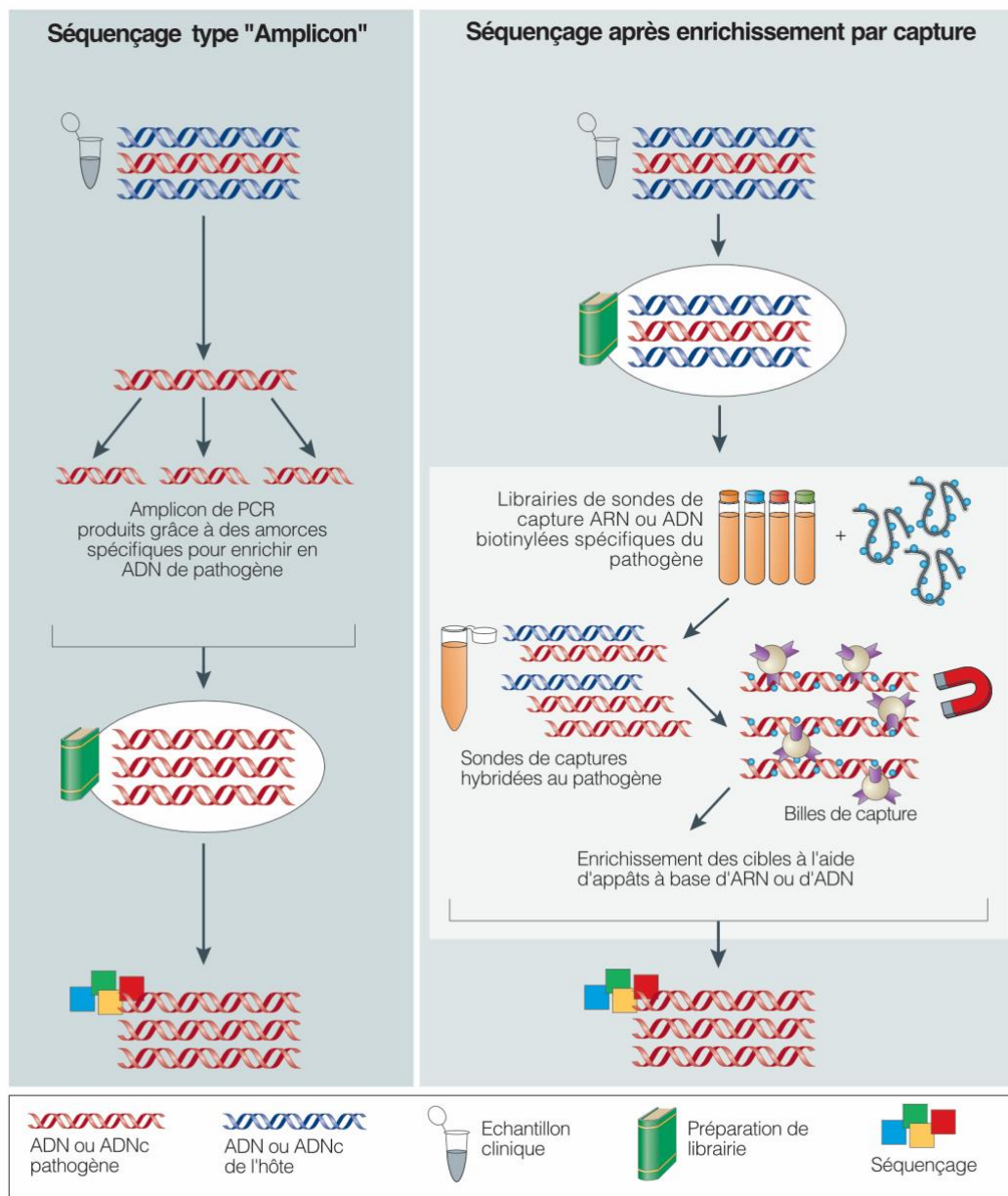


Figure 6. Méthodes de séquençage des génomes viraux à partir d'échantillons cliniques
 Figure adaptée de Houldcroft et al, 2017.⁶³

3. Analyse non ciblée : métagénomique NGS

Contrairement aux approches ciblées de séquençage, les analyses métagénomiques dites « shotgun » renoncent à l'utilisation d'amorces ou de sondes spécifiques. Pour cette technique, après l'extraction des acides nucléique à partir d'échantillon clinique, la totalité de l'ADN et/ou de l'ARN est séquencée. Pour cela l'ADN ou l'ADN complémentaire (ADNc) obtenu après transcription inverse de l'ARN sont fragmentés. La préparation des bibliothèques consiste en la fixation d'amorces et d'index à chaque fragment. Les amorces serviront à amorcer la réaction de séquençage, tandis que les index servent à identifier le fragment, c'est à dire l'associer à un échantillon clinique. Une fois la préparation des bibliothèques terminées l'ensemble de ces fragments sont séquencés (**Figure 7**).

Cette approche est la plus prometteuse pour le diagnostic des infections. En principe, presque tous les agents pathogènes, y compris les virus, les bactéries, les champignons et les parasites, peuvent être identifiés en un seul essai. La métagénomique est une technique de type "aiguille dans une botte de foin", car une petite proportion seulement (généralement <1%) des séquences de l'échantillon clinique sont non humaines, dont seul un sous-ensemble peut correspondre à des agents pathogènes potentiels. Une des limites de la métagénomique est que la sensibilité de l'approche dépend fortement du niveau de bruit de fond. Les tissus, par exemple, ont un bruit de fond d'hôte humain plus important que les fluides corporels exempts de cellules, ce qui entraîne un nombre et une proportion réduits de lectures microbiennes et donc une diminution de la sensibilité de la métagénomique. En outre, il peut être difficile de définir des profils microbiens spécifiques diagnostiques ou prédictifs du développement d'une maladie, en particulier à partir de sites non stériles abritant un microbiome complexe, tels que les sécrétions respiratoires ou les selles. Néanmoins, plusieurs groupes ont validé avec succès la mNGS pour le diagnostic de certaines infections, y compris dans des contextes de méningite ou d'encéphalite,^{67,76} de septicémie⁷⁷ ou de pneumopathie,⁷⁸ et ces tests sont maintenant disponibles dans certains centres en routine pour les patients.⁶²

La sensibilité de la technique est basée sur la profondeur de séquençage, c'est-à-dire la capacité à séquencer une grande quantité de séquences. Néanmoins même si les capacités de séquençage sont de l'ordre du milliard de séquence générée en une seule technique, et que les coûts du séquençage ont baissé, cela reste encore limitatif. Des méthodes supplémentaires pour augmenter efficacement le pourcentage relatif de séquences virales par rapport aux séquences de l'hôte ont été développées. Par exemple, la propagation d'un virus en culture cellulaire est un moyen d'augmenter l'abondance relative des séquences virales. La purification physique des virions, par ultracentrifugation ou traitement à la nucléase pour éliminer les acides nucléiques non encapsidés, est un autre moyen d'enrichir les séquences virales. Les stratégies d'hybridation soustractive atteignent un objectif similaire en réduisant physiquement la quantité de séquences hôtes. Cependant, les capacités de séquençage et les coûts ne seront peut-être bientôt plus limitatifs, et ces méthodes supplémentaires pour augmenter la proportion de séquences virales de l'échantillon ne seront peut-être plus nécessaires.⁶²

Pour les technologies de détection basées sur les séquences, une étape clé est l'analyse informatique requise pour identifier une séquence comme étant de nature virale. Avec les progrès de la capacité de séquençage, le nombre de séquences générées à partir d'un échantillon peut maintenant facilement dépasser 500 millions de séquences. Afin d'analyser efficacement ce volume de données de séquences, des pipelines bio-informatiques automatisés à haut débit ont dû être développés. Bien qu'il existe de nombreuses variations sur ce thème, la plupart des pipelines informatiques utilisent des stratégies générales très similaires (**Figure 8**).

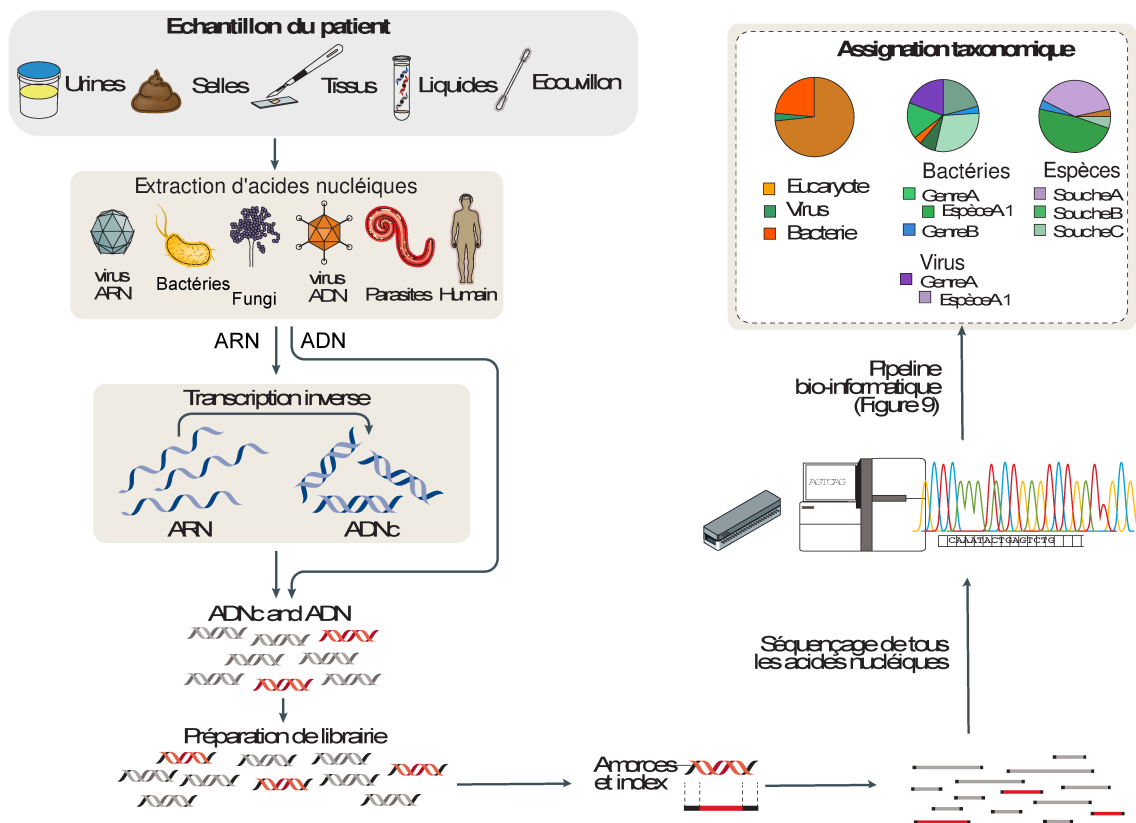


Figure 7. Approche non ciblée de métagénomique NGS dite "shotgun"

Figure adaptée de Chiu et al, 2019.⁶²

Un ensemble de données de séquençage NGS, généralement au format FASTQ ou SAM (sequence alignment map), est analysé sur un serveur informatique, un ordinateur portable ou sur un cloud. Une première étape de prétraitement consiste à la filtration des reads de faible qualité et un élagage des adaptateurs. La soustraction informatique des séquences de l'hôte est réalisée par l'alignement des reads sur un génome humain de référence (**Figure 7**). Les reads non alignés restants sont directement alignés sur de grandes bases de données de référence, telles que la base de données GenBank du National Center for Biotechnology Information (NCBI) ou des collections de séquences ou de génomes microbiens de référence, ou sont d'abord assemblés *de novo* en séquences plus longues appelées « contigs » suivies d'un alignement sur des bases de données de référence. Les séquences présentant des alignements parfaits ou presque parfaits avec des séquences connues (hôtes, virus, bactéries, etc.) peuvent ensuite être classées en fonction de l'identité de la séquence de référence à laquelle une lecture donnée s'est alignée. Après cette étape d'assignement taxonomique des séquences, dans laquelle les reads ou les contigs sont assignés à des taxons spécifiques (par exemple, l'espèce, le genre et la famille), les données peuvent être analysées et visualisées dans un certain nombre de formats différents. Parmi les méthodes utilisées, on peut citer l'utilisation de diagrammes d'identité et de couvertures pour déterminer la part du génome

microbien qui a été récupérée et sa similarité avec les génomes de référence dans la base de données. De plus, des diagrammes de Krona sont employés pour visualiser la diversité taxonomique présente dans l'échantillon. Les séquences dérivées de nouveaux virus sont celles qui n'ont qu'une similarité limitée avec des virus connus dans les bases de données disponibles. Pour une caractérisation plus définitive de la parenté de ces séquences avec les virus connus, des méthodes phylogénétiques sont ensuite utilisées (**Figure 8**).⁶²

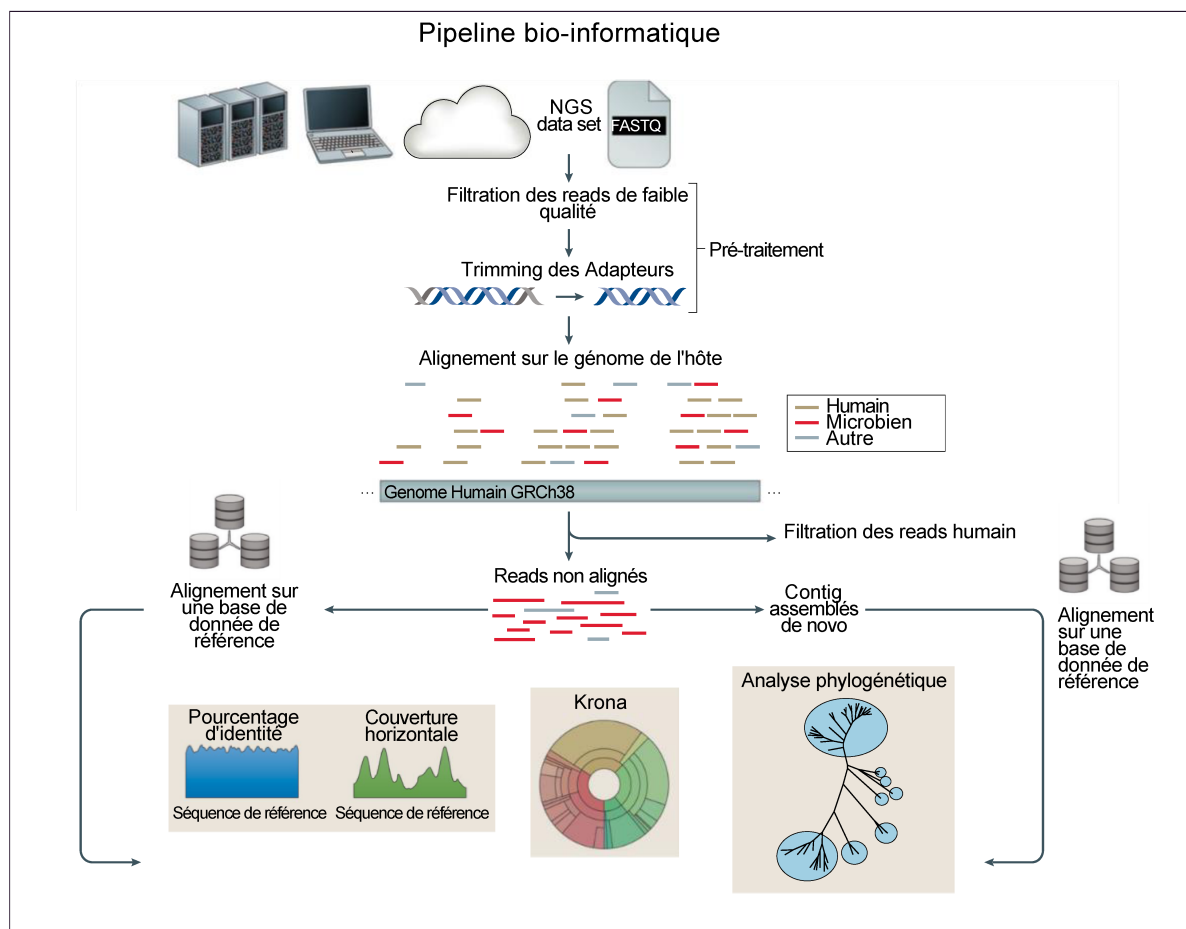


Figure 8. Exemple de pipeline bio-informatique pour l'analyse des séquences de métagénomique.
Figure adaptée de Chiu et al, 2019.⁶²

C. Application des tests moléculaire à la virologie

Les tests moléculaires sont utilisés en routine pour le diagnostic et le suivi des infections virales chroniques par des techniques de PCR ou RT-PCR spécifiques du virus suspecté. L'une des principales applications et raisons du séquençage partiel et ciblé des virus est la détection de mutation de résistance aux médicaments. Par exemple, la gestion de la thérapie

antirétrovirale pour le VIH repose sur le séquençage viral pour la détection des variants résistants aux médicaments. Ces tests sont également nécessaires pour les enquêtes autour d'épidémie pour contrôler la propagation de la maladie. Pour cela, il faut identifier l'agent étiologique, le caractériser par séquençage pour établir les liens épidémiologiques, sa source et son mode de transmission. Enfin la métagénomique est la technique de choix pour l'identification de nouveau virus humain grâce à l'approche non ciblée. En effet, les méthodes génomiques ont supplanté les approches classiques pour la découverte de nouveaux virus. Par exemple, la microscopie électronique nécessite des titres viraux élevés et peut ne pas fournir la résolution nécessaire pour identifier sans ambiguïté le type de virus, même lorsque des particules virales sont observées. De nombreux virus ne peuvent pas être cultivés, et ceux qui peuvent l'être nécessitent souvent des types de cellules très spécifiques. Les méthodes basées sur les anticorps ont une portée limitée, nécessitent le développement de réactifs spécifiques pour un groupe viral donné et exigent la sélection de candidats spécifiques sur la base de connaissances préalables afin d'être utiles pour l'identification d'un virus.

V. Virome humain

Certaines études suggèrent que le nombre total de virus dans le corps humain est du même ordre que le nombre de cellules bactériennes et humaines, soit environ 10^{14} .⁷⁹ La plupart des constituants du virome humain sont supposés être des phages.⁸⁰ Les virus qui infectent les cellules humaines constituent également une partie importante du virome humain. Certains peuvent provoquer des infections aiguës, tandis que d'autres peuvent établir une latence à long terme. Certains virus se livrent à une colonisation bénigne et ne peuvent être associés à aucune maladie particulière, apparaissant comme des "passagers" ou des "commensaux" à long terme. Par exemple, les *Anelloviridae* sont une famille de virus à ADN monocaténaire non enveloppés avec des génomes circulaires assez petits (2-4 kb), comprenant le *Torque teno virus* (TTV), le *Torque teno mini virus* et le *Torque teno midi virus*. Les génomes des virus de cette famille codent ce qui semble être une seule grande protéine, bien que l'expression génétique n'ait pas été bien étudiée car les *Anelloviridae* représentatifs n'ont pas encore été cultivés en culture pure. Les *Anelloviridae* sont extrêmement diversifiés et peuvent être trouvés dans de nombreux sites du corps humain chez une grande partie des humains examinés. Jusqu'à présent, aucun effet pathogène spécifique n'a été associé aux *Anelloviridae*.⁸¹ Une plus grande abondance d'*Anelloviridae* a été trouvée chez les personnes immunodéprimées, y compris les receveurs de greffes de poumons, les personnes séropositives et les personnes sous médicaments immunosuppresseurs en raison d'une maladie inflammatoire de l'intestin, ce qui indique que les *Anelloviridae* sont normalement sous le contrôle immunitaire de l'hôte.^{82,83} Ainsi il a été montré que la charge virale TTV peut être utilisée comme marqueur prédictif du rejet après transplantation d'organes solides et servir à moduler le dosage des immunosuppresseurs.^{84,85} Sur le même principe il a été montré que la détection de picobirnavirus à des niveaux élevés dans les selles pouvaient être prédictif

de la survenue de maladie du greffon contre l'hôte digestive chez des patients receveurs de cellules souches hématopoïétique.⁸⁶ Les *Redondoviridae*, récemment découverts, sont une autre famille de petits virus à ADN circulaire qui semblent être des commensaux largement répandus dans les voies respiratoires.⁸⁷ Chez un adulte en bonne santé, le virome est généralement stable dans le temps. Par exemple, une étude a montré qu'environ 80% des contigs viraux présents persistaient sur une période de 2,5 ans dans l'intestin d'un individu.⁸⁸ Une autre étude récente a suivi le virome intestinal de 10 personnes et a constaté que plus de 90% des contigs viraux reconnaissables persistaient chez chaque individu sur une période d'un an.⁸⁹ Des études sur le virome oral ont révélé une stabilité similaire.⁹⁰ De nombreuses études ont caractérisé le virome humain à différents endroits du corps (**Error! Reference source not found.**).

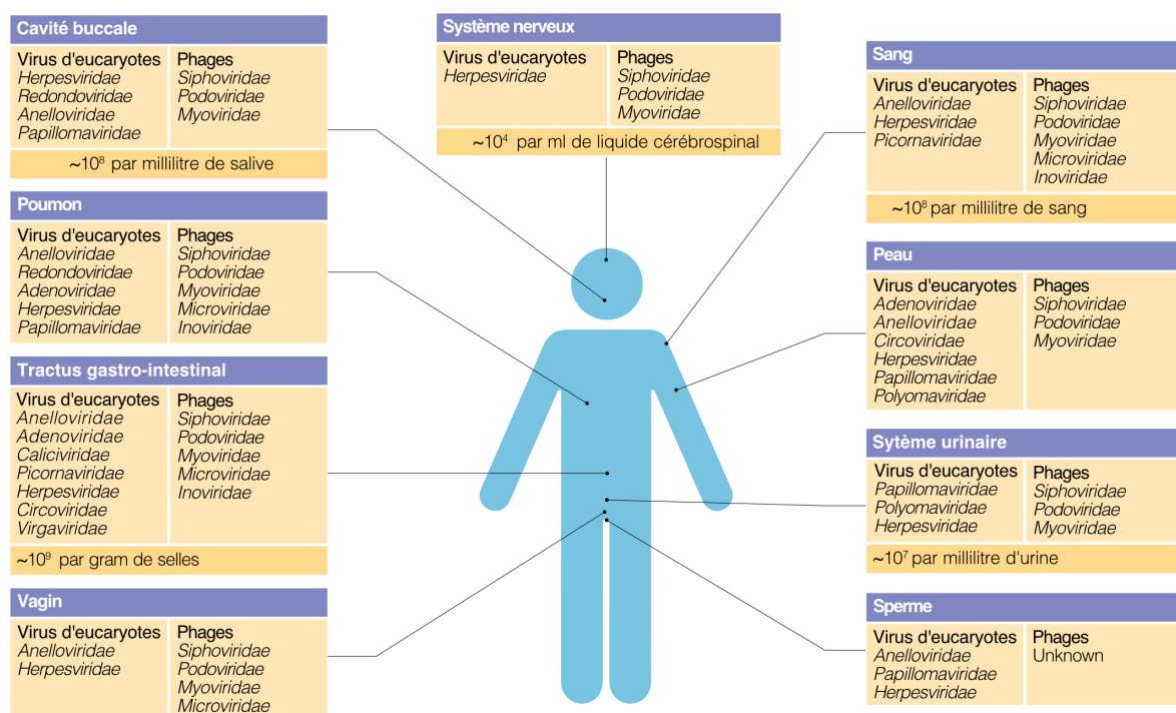


Figure 9. Résumé des virus présents sur chaque site du corps humain
Figure adaptée de Liang et al, 2021.⁸⁰

La déstabilisation du virome est souvent associée à plusieurs facteurs. Les principaux facteurs sont le régime alimentaire, l'allaitement, les médicaments (y compris les antibiotiques et les immunosuppresseurs), la génétique, la cohabitation, la géographie, la présence d'une maladie (**Tableau 2**) et le vieillissement.⁸⁰ Les populations de viromes peuvent influencer leurs hôtes humains de nombreuses façons. Les virus eucaryotes qui infectent les cellules humaines provoquent des infections, déclenchent des réponses immunitaires et, parfois, causent des maladies. Les phages peuvent affecter l'hôte indirectement en modulant la composition bactérienne et l'aptitude bactérienne. Certains phages et virus de cellules humaines peuvent être intégrés dans les cellules de leurs hôtes respectifs, transférant parfois de nouvelles

fonctionnalités aux cellules hôtes. Certains phages peuvent également interagir directement avec les cellules humaines et déclencher des réponses immunitaires.

Bien sûr, de nombreux virus individuels sont bien connus pour être à l'origine de la morbidité et de la mortalité. Des études portant sur des populations virales entières ont également commencé à mettre en évidence des schémas associés à des états pathologiques. Le virome a été considéré comme un déclencheur potentiel de maladies auto-immunes. Dans plusieurs études, des signatures virologiques ont été associées aux maladies inflammatoires de l'intestin de l'enfant et de l'adulte, notamment une expansion reproductible des *Caudovirales* et une réduction des *Microviridae*.⁹¹⁻⁹⁴ Il reste à déterminer s'il s'agit d'une conséquence de l'altération des populations bactériennes dysbiotiques ou d'une implication plus profonde dans le processus de la maladie. Une étude métagénomique a montré que les enfants fréquemment exposés aux entérovirus entre l'âge de 1 et 2 ans présentent un risque plus élevé de maladie cœliaque.⁹⁵ Des études ont révélé que les relations entre les phages et les populations bactériennes peuvent influencer le retard de croissance chez les enfants.^{96,97} Des exemples d'associations proposées entre les maladies et la structure des populations virales sont résumés dans le **Tableau 2**.

Des études utilisant des modèles animaux ont indiqué que certains virus eucaryotes peuvent même être bénéfiques pour l'hôte. Par exemple, une infection persistante par une souche de norovirus murin peut compenser l'absence de bactéries chez des souris, permettant la restauration de la morphologie intestinale et favorisant la différenciation des lymphocytes.⁹⁸ L'astrovirus murin a protégé des souris immunodéficientes contre des infections entériques à *Norovirus* par l'induction d'interférons de type III dans la barrière épithéliale intestinale.⁹⁹

Les virus du virome interagissent activement non seulement avec d'autres micro-organismes, mais aussi avec le système immunitaire des mammifères. Des associations entre les altérations du virome et les états pathologiques sont de plus en plus souvent décrites, mais dans de nombreux cas, la causalité et les mécanismes moléculaires restent à élucider.

Tableau 2. Exemples d'altérations de populations virales dans des pathologies humaines

Maladie Humaine	Échantillons	Altération majeure du virome	Ref.
Malnutrition aigue sévère	Selles	Réduction de la diversité virale	100
Maladie de Crohn et colite ulcéreuse	Selles	Augmentation de la quantité de <i>Caudovirales</i>	92
Maladie de Crohn	Selles et biopsies	Altérations modérées	101
Syndrome de l'immunodéficience acquise	Selles	Augmentation de la quantité d' <i>Adenovirus</i>	82
Déficit en DOCK8	Écouvillons cutanés	Augmentation de l'abondance du virome cutané, en particulier en <i>Papillomavirus</i> humain	102
Cancer colorectal	Selles	Augmentation de la diversité virale	103
Maladie de Crohn et colite ulcéreuse	Selles	Augmentation de l'abondance des <i>Caudovirales</i>	94
Maladie de Crohn et colite ulcéreuse	Biopsies digestives	Augmentation des <i>Hepadnaviridae</i> and <i>Hepeviridae</i> ; et réduction des <i>Polydnaviridae</i> , <i>Tymoviridae</i> and <i>Virgaviridae</i>	104
Maladie cœliaque	Selles	Augmentation des <i>Enterovirus</i>	95
Maladie de Crohn	Selles	Les phages virulents sont remplacés par des phages tempérés	91
Colite ulcéreuse	Biopsies digestives	Augmentation des <i>Caudovirales</i> , des fonctions de virulence des phages et des bactéries, et perte des corrélations entre virus et bactéries	93
Virémie VIH	Liquide Seminal	Augmentation du <i>Cytomegalovirus</i> humain	105
Maladie inflammatoire de l'intestin à un stade très précoce	Selles	Augmentation de la proportion de <i>Caudovirales</i> par rapport aux <i>Microviridae</i>	83
Transplantation de cellules hématopoïétiques	Selles	Augmentation des <i>Picobirnaviruses</i>	86

Tableau adapté de Liang et al, 2021.⁸⁰

VI. Critères d'imputabilité

Il existe de nombreuses maladies pour lesquelles une étiologie microbienne est suspectée. L'hypothèse selon laquelle une maladie a une cause infectieuse est étayée par : des caractéristiques cliniques (similaires à celles de maladies infectieuses connues, par exemple fièvre, leucocytose), l'épidémiologie (regroupement de cas dans le temps ou dans l'espace), l'histologie (inflammation des tissus affectés, par exemple infiltrats lymphocytaires) ou des structures microbiennes caractéristiques, le traitement (réponse clinique à un traitement antimicrobien) et la prévention de la maladie par des vaccins ciblant des antigènes microbiens. La preuve qu'un microbe est à l'origine d'une maladie nécessite des preuves plus rigoureuses.¹⁰⁶

Les postulats de Koch découlent de ses travaux sur des maladies infectieuses telles que l'anthrax et la tuberculose. Ces lignes directrices visaient à établir une norme pour la preuve de la causalité dans les maladies infectieuses. Cette norme avait pour but de convaincre les sceptiques que les microbes pouvaient causer des maladies et de pousser les microbiologistes à utiliser des critères plus rigoureux avant d'affirmer qu'il existe une relation de cause à effet entre un microbe et une maladie. Les postulats de Koch peuvent être résumés ainsi¹⁰⁷ :

- « (i) Le parasite est présent dans tous les cas de la maladie en question et dans des circonstances qui peuvent expliquer les changements pathologiques et l'évolution clinique de la maladie.
- (ii) Le parasite n'est présent dans aucune autre maladie en tant que parasite fortuit et non pathogène.
- (iii) Après avoir été complètement isolé de l'organisme et cultivé de façon répétée en culture pure, le parasite peut induire la maladie de façon nouvelle. »

Ces postulats ont plusieurs limites et ne peuvent plus être appliqués aujourd'hui en particulier quand le micro-organisme ne peut pas être isolé en culture ; qu'il se comporte comme commensal transitoire ou permanent, ou que son pouvoir pathogène est seulement opportuniste ; certains microorganismes ont une gamme d'hôte limitée à l'humain et ne peuvent pas reproduire de maladie typique chez l'animal, et il ne serait pas éthique d'envisager une inoculation chez l'humain ; le micro-organisme peut se trouver sur un site éloigné et provoquer la maladie par la libération d'une toxine ; les méthodes d'amplification de l'acide nucléique permettent de détecter les séquences du micro-organisme en présence d'un traitement anti-infectieux ; les méthodes moléculaires détectent les quelques copies de séquences de micro-organismes pathogènes qui peuvent être présents en l'absence de pathologie, remettant ainsi en question la spécificité de l'association parasite-maladie exigée par les postulats de Koch. Constatant ces limites et au fil des progrès technologiques d'autres auteurs ont proposé de nouvelles interprétations et des révisions des postulats de Koch, comme Thomas Rivers¹⁰⁷ en 1937 soulignant que les postulats n'étaient pas applicables aux virus, ou encore Sir Austin Bradford Hill¹⁰⁸ qui propose en 1965 des critères épidémiologiques

pour l'association causale des maladies d'expositions environnementales. Devant l'impact des méthodes d'identification microbienne basées sur les séquences sur la microbiologie clinique et la fréquence croissante des questions de causalité, Fredricks et Relman ont proposé en 1996 de nouvelles lignes directrices pour l'établissement de relations de causalité entre les microbes et les maladies :

« (i) Une séquence d'acide nucléique appartenant à un agent pathogène présumé doit être présente dans la plupart des cas de maladie infectieuse. Les acides nucléiques microbiens doivent être trouvés de préférence dans les organes ou les sites anatomiques connus pour être malades (c'est-à-dire présentant des signes anatomiques, histologiques, chimiques ou cliniques de pathologie) et non dans les organes dépourvus de pathologie.

(ii) Le nombre de copies de séquences d'acides nucléiques associées à des agents pathogènes devrait être plus faible, voire inexistant, chez les hôtes ou dans les tissus non malades.

(iii) En cas de résolution de la maladie (par exemple, grâce à un traitement cliniquement efficace), le nombre de copies des séquences d'acide nucléique associées à l'agent pathogène devrait diminuer ou devenir indétectable. En cas de rechute clinique, c'est l'inverse qui se produit.

(iv) Lorsque la détection de la séquence précède la maladie ou que le nombre de copies de la séquence est en corrélation avec la gravité de la maladie ou de la pathologie, l'association séquence-maladie est plus susceptible d'être une relation de cause à effet.

(v) La nature du micro-organisme déduite de la séquence disponible doit être cohérente avec les caractéristiques biologiques connues de ce groupe d'organismes. Lorsque les phénotypes (par exemple, la pathologie, la morphologie microbienne et les caractéristiques cliniques) sont prédits par les relations phylogénétiques basées sur la séquence, la signification de la séquence est renforcée.

(vi) Les corrélations entre les séquences tissulaires doivent être recherchées au niveau cellulaire : des efforts doivent être faits pour démontrer par l'hybridation *in situ* spécifique la présence des séquences microbiennes dans les zones de la pathologie tissulaire et des micro-organismes visibles ou dans les zones où l'on présume que les micro-organismes se trouvent.

(vii) Ces formes de preuves de causalité microbienne basées sur les séquences doivent être reproductibles. »

Les auteurs soulignent qu'il n'est peut-être pas approprié de faire des déclarations absolues (exemple ii et iii) concernant les formes de preuve et que le respect strict de chacune de ces lignes directrices n'est peut-être pas nécessaire pour obtenir un argument convaincant concernant le lien de causalité. Néanmoins, la capacité à remplir tout ou partie des critères énumérés ci-dessus pour les recherches de causalité microbienne basées sur des séquences devrait fournir des preuves solides d'un microorganisme cliniquement présomptif dans une maladie.¹⁰⁹

VII. Infections virales opportunistes

Les patients fragilisés par un déficit immunitaire primitif ou secondaire à un traitement immunosuppresseur ou une maladie sous-jacente présentent un risque accru d'infection grave par différents virus.

Parmi les virus persistant à ADN pathogène chez l'individu immunocompétent certains présentent un enjeu majeur chez le patient immunodéprimé.

Les virus de la famille des *Herpesviridae* peuvent se réactiver toute la vie de l'individu après une primo-infection pendant l'enfance. Chez les patients immunodéprimés, les primo-infections sont plus fréquemment symptomatiques, avec des symptômes plus graves lors de la primo-infection ou lors de réactivations (lésions sévères à HSV-1 avec guérison lente, Zona sévère) et des infections disséminées (infection disséminée à HSV-1) ou des présentations cliniques réservées à l'immunodéprimé (rétinite à CMV, maladie de Kaposi associée à l'HHV-8, lymphoprolifération post transplantation induit par l'EBV).¹¹⁰⁻¹¹⁵

D'autres virus à ADN avec un large tropisme cellulaire¹¹⁶ comme l'adénovirus peuvent persister longtemps dans l'organisme après l'infection¹¹⁷⁻¹²¹ en particulier chez les patients immunodéprimés^{121,122} et entraîner des complications sévères en particulier chez les patients transplantés d'organes solides¹²³⁻¹²⁶ et les patients greffés de cellules souches hématopoïétiques.¹²⁷⁻¹²⁹

Les *Polyomavirus Humain* (HPyV) font partie du virome et ne sont pas pathogènes chez l'individu immunocompétent, cependant ils ont été associés à des pathologies chez le patient immunodéprimé. L'une des caractéristiques biologiques de tous les *Polyomavirus* est le maintien d'une infection chronique et à vie chez leurs hôtes. Le rein ou l'urothélium sont les localisations principales du *Polyomavirus* BK (BKV), tandis que le rein et les organes lymphoïdes sont celles du *Polyomavirus* JC (JCV). L'ADN du *Polyomavirus* de Merkel (MCPyV) et du *Polyomavirus humain 7* (HPyV7) peut être facilement isolé à partir de la peau normale.¹³⁰ Le BKV est reconnu comme agent étiologique de néphropathie, de cystite hémorragique et de cancer urothélial.¹³¹ Le BKV a également été associé à de rares cas de sténose urétérale, de pneumonie, d'encéphalite, de rétinite, de vascularite et d'atteinte de plusieurs organes, mais peu d'informations sont connues sur leur prévalence, les facteurs de risque et la pathogenèse.¹³² Le JCV est le principal agent étiologique de la leucoencéphalopathie multifocale progressive (LEMP). Les autres manifestations neurologiques avec lesquelles le virus est associé sont la neuronopathie à cellules granuleuses dans le cervelet, l'encéphalopathie ciblant les neurones pyramidaux corticaux, ainsi que des méningites et des encéphalites.¹³³ Le MCPyV est associé au carcinome à cellules de Merkel.¹³⁴ Enfin le HPyV7 et le *Polyomavirus trichodys-plasia spinulosa* (TSPyV) sont également associés à des pathologies cutanées.^{135,136}

Dans certaines circonstances, des virus à ARN qui provoquent des infections aiguës et bénignes chez l'immunocompétent entraînent des infections chroniques et sévères chez le patient immunodéprimé.

Le virus de l'hépatite E (VHE) infecte environ 20 millions de personnes par an (OMS, 24 juin 2022) et est responsable chez les individus immunocompétents d'une infection le plus souvent asymptomatique et en cas d'infection symptomatique d'une létalité faible. En revanche, chez les adultes ou les enfants qui reçoivent un traitement immunosuppresseur à la suite d'une transplantation d'organe solide ou de cellules souches hématopoïétiques, et chez les patients présentant une infection concomitante par le VIH, la réplication du VHE peut persister au-delà de 6 mois, entraînant une hépatite chronique avec un risque de fibrose hépatique progressive et de cirrhose. L'hépatite E chronique est le plus souvent signalée chez les patients transplantés d'organe solide, avec un tiers des patients capables d'éliminer spontanément le virus et deux tiers développant une hépatite chronique.¹³⁷⁻¹⁴⁰

Les infections virales respiratoires sont très fréquentes chez les enfants et les adultes, les symptômes quand ils sont présents sont généralement aspécifiques et bénins chez l'immunocompétent (rhinorrhée, pharyngite, toux, bronchite, syndrome pseudo-grippal).

L'immunodépression (immunodéficiences congénitales, l'infection par le VIH, le traitement immunosuppresseur pour le cancer ou les hémopathies malignes, et la transplantation d'organes solides ou de cellules souches hématopoïétiques) est associée à une gravité accrue de la maladie en cas d'infection par le Virus Respiratoire Syncytial (VRS), le Metapneumovirus (HMPV) et le Parainfluenza virus (HPIV).¹⁴¹⁻¹⁴⁴

Les patients immunodéprimés ne peuvent pas être vaccinés par les vaccins vivants atténués. Par exemple, ces patients ne peuvent pas mettre en place une réponse immunitaire appropriée à la suite de l'infection par le virus de la rougeole et dans certains cas ne peuvent pas développer les symptômes typiques telle que l'éruption cutanée d'origine immunitaire, ce qui complique le diagnostic clinique. Par ailleurs, ils sont exposés à des complications rares et fatales comme l'encéphalite à inclusion rougeoleuse.^{51,145,146}

Des cas de méningo-encéphalite chronique secondaire à une infection à entérovirus ont été décrits en cas d'agammaglobulinémie liée à l'X (XLA) et ont également été observés occasionnellement en cas de déficit immunitaire commun variable (DICV). Dans plusieurs cohortes de patients, les entérovirus ont été associés à une excrétion plus longue du virus et de manière significative à une évolution sévère et, dans de nombreux cas, au décès.¹⁴⁷⁻¹⁴⁹

L'infection chronique à norovirus dans le cadre d'un déficit immunitaire primaire ou secondaire peut provoquer une diarrhée prolongée nécessitant une alimentation parentérale et peut jouer un rôle central dans la pathogenèse de l'entéropathie chez les patients atteints de déficit immunitaire primaire.^{141,150-152} Des preuves concomitantes de la présence d'autres agents pathogènes viraux tels que les sapovirus et les parechovirus ont également été rapportées dans des contextes d'entéropathie chez des patients présentant un DICV ou post-HSCT.^{141,153,154} Cependant ces entéropathies pourraient être associées à d'autres pathogènes qui pourraient être à l'origine de l'accumulation de lymphocytes intra-épithéliaux et de

l'atrophie villositaire. Néanmoins ce mécanisme pourrait également être indépendant des agents pathogènes chez certains individus.^{150,151,155}

Plus récemment, les techniques de séquençage à haut débit ont permis l'identification de nouveaux agents pathogènes opportunistes comme causes de maladie chez les patients immunodéficients. Des encéphalites chez des patients atteints de XLA ont été rapportées avec un *Astrovirus VA1* neurotrope.^{156,157} Le coronavirus OC43 peut être responsable d'encéphalite particulièrement sévère chez le patient immunodéprimé.^{158,159} Ces techniques de diagnostic continueront probablement à identifier des infections virales importantes dans ces populations au cours des prochaines années. Néanmoins, en raison de la nature ubiquitaire et diverse de ces agents infectieux, la prise en charge efficace de ces infections continuera de représenter un défi thérapeutique.

VIII. Virus émergents et déficits immunitaires

Les deux dernières décennies ont été marquées par un regain d'intérêt pour la virologie, motivé en grande partie par l'émergence de plusieurs menaces virales importantes, notamment le virus Ebola, le virus Zika, le SARS-CoV, le MERS-CoV et, plus particulièrement, le SARS-CoV-2. Ces épidémies et pandémies ont conduit à reconnaître que le développement de traitements et de vaccins sûrs et efficaces contre ces agents pathogènes nécessite une compréhension totale de leur pathogénèse.

Plusieurs tendances ont renforcé la probabilité d'émergence de nouvelles maladies virales.^{160,161} Tout d'abord, la population mondiale a continué à augmenter et l'urbanisation a concentré une plus grande proportion de personnes dans des zones densément peuplées, facilitant ainsi la transmission de tout nouveau virus. Deuxièmement, les moyens de transport modernes permettent de véhiculer des infections dans le monde entier en une seule période d'incubation. Troisièmement, les perturbations de l'environnement causées par l'homme se produisent à un rythme de plus en plus rapide, augmentant ainsi la probabilité de transmission des zoonoses et arboviroses à l'Homme. Enfin, les progrès de la virologie fondamentale ont permis de disposer d'un nouvel arsenal de méthodes pour la détection des virus et l'exploration de leurs pouvoirs pathogènes jusqu'alors inconnus. Parmi les facteurs qui définissent la sensibilité de l'être humain à une infection par un virus zoonotique potentiellement pathogène (**Figure 10**), la réponse immunitaire est l'une des dernières barrières et détermine la probabilité de l'infection ainsi que sa sévérité.¹⁶²

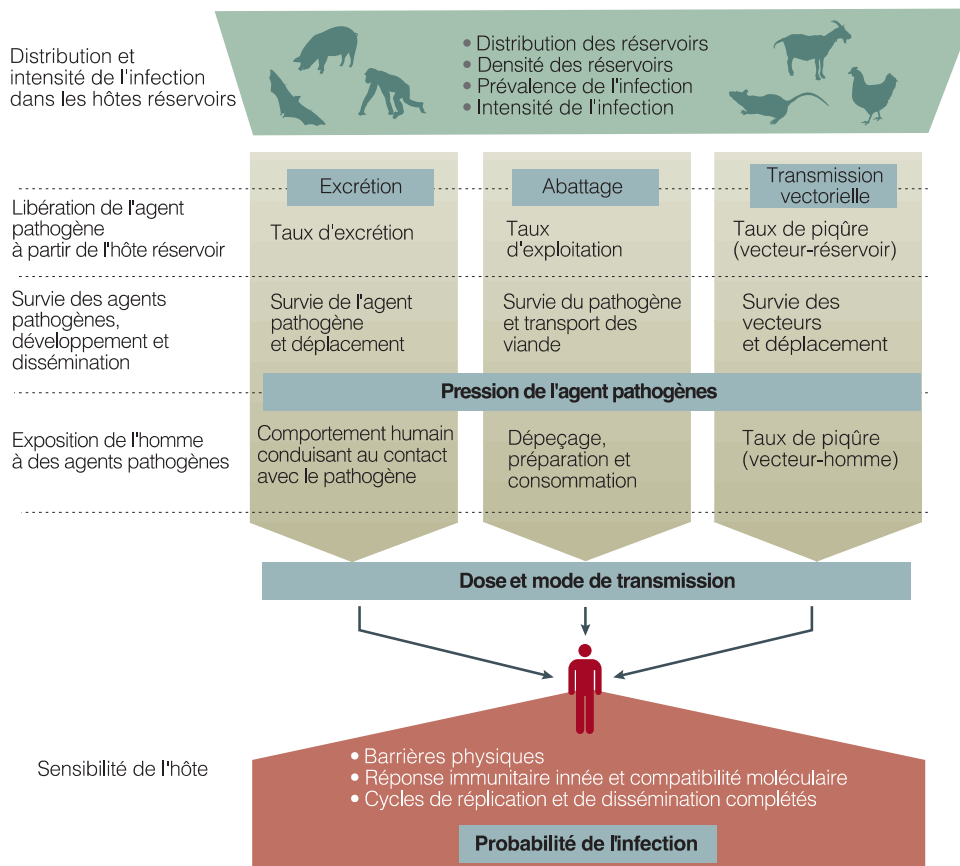


Figure 10. Facteurs influençant la propagation d'un agent pathogène zoonotique jusqu'à l'infection de l'être humain

Le risque de propagation d'un agent pathogène est déterminé par un ensemble de processus qui relie la dynamique de l'infection dans les hôtes réservoirs à la probabilité de l'infection de l'être humain (hôte récepteur). Ces processus comprennent les déterminants microbiologiques de la survie et de la dissémination en dehors des hôtes réservoirs, les déterminants épidémiologiques et comportementaux de l'exposition et les facteurs biologiques internes de l'hôte récepteur. La distribution et l'intensité de l'infection chez les hôtes réservoirs, suivies de la libération de l'agent pathogène, de sa survie et de son développement éventuel jusqu'au stade infectieux, déterminent la pression de l'agent pathogène. Ce sont ensuite les comportements humains de l'hôte récepteur qui vont déterminer la nature de l'exposition avec l'agent pathogène (dose et mode de transmission). Enfin une série de barrières internes de l'hôte (physique, immunitaire ou génétique) détermine sa sensibilité et, par conséquent, la probabilité et la gravité de l'infection. Figure adaptée de Plowright et al, 2017.¹⁶²

Le SARS-CoV-2 est l'exemple le plus récent d'un nouveau virus ayant entraîné une pandémie mondiale. Les revues de la littérature ont montré que les patients présentant un déficit immunitaire infectés par le SARS-CoV-2 ont un risque plus élevé d'hospitalisation et d'admission en soins intensifs.^{163,164} Alors que la grippe saisonnière est considérée comme une maladie endémique dans le monde, le génome segmenté des virus de la grippe permet l'échange de segments entiers de gènes lorsque les cellules sont co-infectées par deux virus de la grippe. Le réassortiment peut entraîner un changement majeur dans l'antigénicité du virus si le nouveau virus porte une hémagglutinine nouvelle pour la population humaine. Ces réassortiments se sont produits en 1957 et en 1968, avec l'apparition des virus de la grippe pandémique H2N2 et H3N2, respectivement. La pandémie de 2009 a été causée par un virus

de la grippe porcine H1N1 complexe à triple réassortiment.¹⁶⁵ Les patients immunodéprimés sont également plus sensibles aux complications de la grippe.¹⁶⁶ Des séries de cas de patients transplantés d'organes solides ou de cellules souches hématopoïétiques atteints de la grippe H1N1 ont montré que 12-16 % d'entre eux ont dû être hospitalisés en soins intensifs et qu'environ 4-6 % sont décédés.^{167,168}

L'émergence de nouveaux virus peut donc également constituer une menace particulière pour cette cohorte de patients et justifie une vigilance constante. En cas d'émergence d'un nouveau virus respiratoire, les patients immunodéprimés seraient particulièrement exposés aux risques de complications sévères et rares^{141,145,158,169} et le phénomène de persistance et d'évolution au sein de l'hôte au cours de l'infection chronique, avec le développement potentiel de nouvelles souches, de facteurs de virulence ou de résistance est particulièrement préoccupant pour la santé publique.¹⁷⁰⁻¹⁷³ Il est donc logique de concentrer nos efforts de surveillance sur cette population afin de détecter l'apparition de nouveaux virus ou de virus résistants aux traitements.

CONTEXTE ET OBJECTIFS

Contexte

Ces deux dernières décennies des progrès considérables ont été obtenus dans tous les domaines de la virologie grâce à la rupture technologique du séquençage à haut débit. Cette technologie a contribué à accélérer la constitution du corpus des connaissances en virologie fondamentale et médicale. La classification des virus a dû évoluer, de nouveaux agents viraux ont été identifiés et leurs séquences génomiques ont été rapidement élucidées, permettant ainsi une meilleure caractérisation du virome humain.

Dans nos établissements de soins de l'Assistance Publique des Hôpitaux de Paris, la patientèle immunodéprimée représente une part importante des consultations et des hospitalisations. Les contextes d'immunodépression sont variés. Les déficits immunitaires peuvent être primitifs ou secondaires à une thérapie immunosuppressive dans le cadre de transplantation d'organe solide, de cellules souches hématopoïétiques, ou du traitement d'une maladie auto-immune. La prise en charge de ces patients pédiatriques ou adultes est très souvent multidisciplinaire, dans le cadre de leur suivi ou du diagnostic de symptômes inexpliqués. Nos centres accueillent fréquemment des patients présentant des déficits immunitaires (parfois rares et complexes) qui présentent des symptômes pour lesquels une maladie infectieuse est suspectée sans qu'aucune documentation microbiologique ni une autre étiologie ne soit retrouvée.

En réponse à cette demande et dans une démarche de progrès technologique, le laboratoire de microbiologie de l'Hôpital Universitaire Necker Enfants malades a mis au service des patients présentant des pathologies complexes, le diagnostic virologique par métagénomique clinique depuis novembre 2019.

Objectifs

Les deux objectifs principaux de la thèse étaient (i) d'identifier au sein de populations immunodéprimés particulièrement sensibles aux infections virales connues mais aussi émergentes, des virus impliqués dans des pathologies jusqu'alors inexpliquées et de faire la preuve de leur pathogénicité (ii) d'identifier de façon plus large les indications pour le diagnostic microbiologique par métagénomique clinique.

Plus spécifiquement, les objectifs pour chaque article sont décrits ci-dessous :

Article 1

L'objectif était d'établir un lien de causalité entre l'infection à Aichi virus (AiV) et le tableau d'hépatite chronique présentée par 6 patients immunodéprimés infectés en s'appuyant sur les critères de Fredricks et Relman.

Les questions de recherche posées pour atteindre cet objectif ont été les suivantes :

- Quelle est la fréquence respective de l'infection à AiV au sein d'une population de patients immunodéprimés asymptomatiques et d'une population de patients immunodéprimés présentant une hépatite chronique d'origine inconnue ?
- Quelles sont les particularités cliniques, histologiques et d'évolution des atteintes d'organes chez les 6 patients immunodéprimés infectés par le AiV virus par rapport à celles d'autres patients immunodéprimés non infectés mais présentant aussi une hépatite virale d'origine inconnue ?
- Le AiV se multiplie-t-il dans les hépatocytes des tissus infectés ?
- Les AiV associés à ces 6 cas d'hépatites chroniques appartiennent-ils au même génotype ?
- La résolution de la maladie s'accompagne-t-elle d'une négativation de détection du virus ?

Hypothèse : La description antérieure d'un cas d'infection à AiV très proche cliniquement des 6 cas de notre étude suggérerait qu'il pouvait s'agir d'une nouvelle entité clinique associée ou causée par l'infection à AiV.

Article 2

L'objectif était d'établir un lien de causalité entre l'infection par ce nouveau circovirus (HCirV-1) et le tableau d'hépatite chronique présenté par la patiente immunodéprimée en s'appuyant sur les critères de Fredricks et Relman

Les questions de recherche posées pour atteindre cet objectif ont été les suivantes :

- Comment le génome complet de ce circovirus se situe par analyse phylogénétique par rapport aux autres circovirus connus ?

- Quelle est la temporalité du niveau de détection (charge virale) du circovirus par rapport aux symptômes cliniques, à leurs intensités et à leurs évolutions ?
- Quelle est la fréquence de détection de ce virus dans une population contrôle de sujets immunodéprimés et immunocompétents asymptomatiques et celle dans une population de patients avec des hépatites chroniques ?
- Le circovirus se multiplie-t-il dans les hépatocytes des tissus infectés ?

Hypothèse : Les circovirus ne sont pas décrits chez l'homme, mais certains types de circovirus sont rapportés comme agents responsables d'hépatite chez le porc et le cheval. Cette association suggérait qu'une causalité entre ce nouveau circovirus découvert chez notre patiente et son hépatite clinique était une possibilité méritant d'être explorée.

Article 3

L'objectif était d'identifier les déterminants de la positivité d'une analyse par NGS sur la base des résultats obtenus chez 523 patients immunocompétents et immunodéprimés pour lesquels 742 échantillons (tissus, LCS, selles, sang, respiratoires) ont été analysés par NGS dans notre laboratoire de 2019 à 2022.

Les questions de recherche posées ont été les suivantes :

- Quelle était la probabilité de détecter un agent causal ou probablement causal dans cette population par mNGS ?
- Quelle était la description des agents causals ou probablement causals ?
- Pour quelle sous population la probabilité de détecter un agent causal ou probablement causal était-elle la plus importante ?
- Quels étaient les prélèvements les plus susceptibles de permettre de détecter un agent causal ou probablement causal ?

Hypothèse : Il y a peu d'études prospectives sur le sujet incluant plusieurs types de patients, d'échantillons et de symptômes. Les études qui ont été réalisées avec plus de deux types de matrices d'échantillons différentes s'attachaient à comparer la sensibilité de l'analyse mNGS par rapport aux techniques microbiologiques conventionnelles dans une utilisation de première ligne. Dans notre étude nous avons positionné l'analyse mNGS en deuxième ligne lorsque les analyses microbiologiques conventionnelles s'étaient avérées négatives. Notre hypothèse était qu'avec ce positionnement de deuxième ligne, notre étude permettrait d'identifier une sous population de patients et/ou de types de prélèvements pour lesquels la probabilité de détecter un agent causal par mNGS serait significativement plus élevée que pour l'ensemble des patients ou des prélèvements.

MATERIEL ET METHODES

I. Métagénomique et métatranscriptomique

A. Technique

Les échantillons ont été transportés au laboratoire de microbiologie de l'hôpital Necker-Enfants malades (Paris, France) sur de la carboglace et conservés congelés à -80°C avant d'être analysés.

Pour les tissus, l'ARN total est utilisé comme signature de l'expression génétique virale, bactérienne ou fongique, qui est une caractéristique de l'infection active. Même dans le cas d'une infection par un virus à ADN, les transcrits restent des intermédiaires obligatoires de l'expression des protéines virales, ce qui signifie qu'une approche transcriptomique (ARN total) est appropriée pour la détection des virus à ARN ou à ADN en cours de réplication dans un contexte tissulaire et, plus largement, pour tout pathogène bactérien ou fongique ayant une activité transcriptionnelle. Un traitement supplémentaire à la DNase a été appliqué à l'extrait d'ARN pour éliminer autant d'ADN résiduel que possible, avant la construction de la librairie, afin d'éviter le séquençage de l'ADN génomique humain qui réduirait considérablement la sensibilité.

Pour les fluides (plasma, LCS, urine...), la logique est différente et vise à détecter des particules virales libres (virus ADN ou ARN), ou des corps bactériens, typiquement dans un contexte de virémie, d'infection pulmonaire, ou de sepsis. On utilise donc une approche ADN/ARN qui, pour certains échantillons, peut être couplée à un traitement en amont par des nucléases pour éliminer les acides nucléiques libres non protégés afin d'augmenter la sensibilité.

En bref, l'ARN total traité à la DNase provenant d'échantillons de tissus et de sang a été isolé après une étape de broyage par des billes et soumis au séquençage du transcriptome. Les acides nucléiques totaux ont été isolés à partir d'échantillons de liquides avec (LBA, plasma, surnageant de selles) ou sans (LCS, urines) prédigestion des acides nucléiques libres à l'aide de nucléases.

Toutes les procédures de traitement des échantillons, y compris les protocoles d'extraction nucléique, les contrôles de qualité et le kit utilisé pour la préparation des librairies sont résumées dans la **Figure 11**.

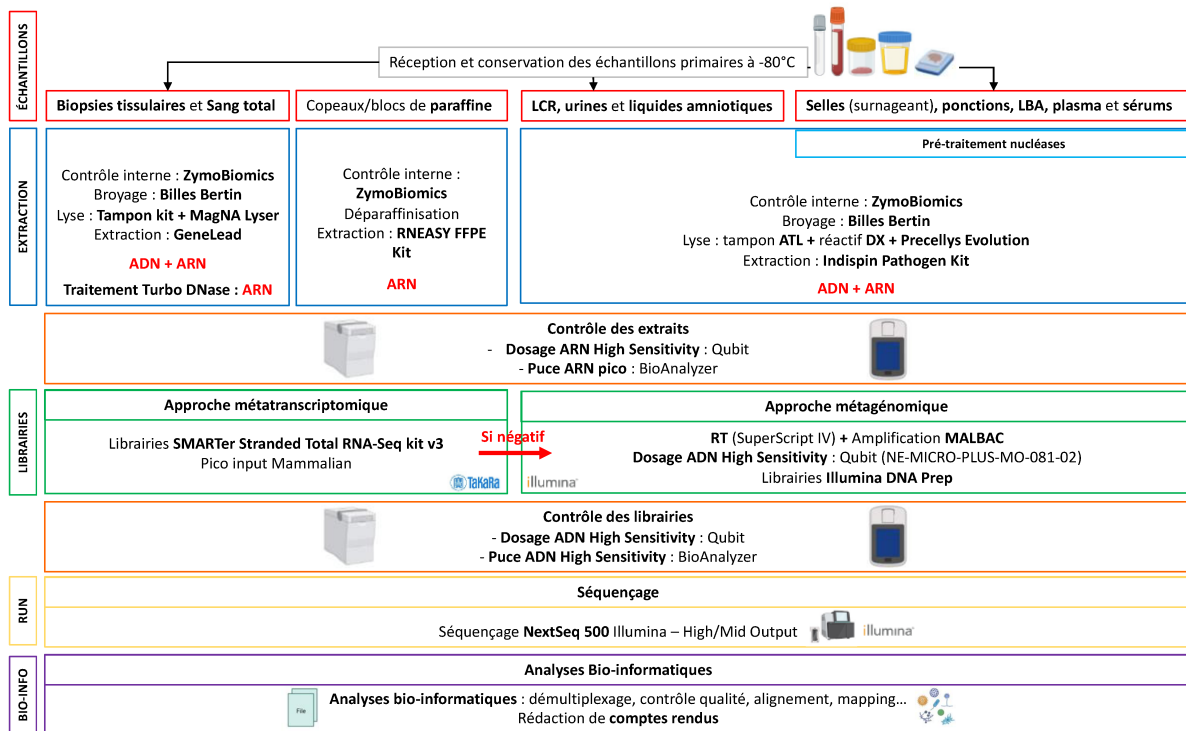


Figure 11. Logigramme résumant les procédures d'analyse de métagénomique et métatranscriptomique

Le séquençage des librairies a été effectué sur un instrument NextSeq500 (Illumina, Evry, France) avec une moyenne de 50 millions de reads (1x150 bp) par échantillon. Les tests ont été effectués par séries de 6 ou 8 échantillons, sur une base hebdomadaire.

Le pipeline d'analyse bio-informatique est décrit dans la **Figure 12**. Pour les librairies générées par l'approche Smarter V3, un premier module permet de démultiplexer les reads grâce à l'information des index UMI (Unique Molecular Identifiers). Après filtration sur critère de qualité (trimming), les séquences humaines sont retirées sur la base de leur composition. Il existe ensuite deux branches spécifiques : une branche virale, et une autre pour les autres agents pathogènes.

La branche virale, assemble les reads après les avoir normalisés pour homogénéiser la couverture. Les contigs résultants, ainsi que les reads n'ayant pas participé à un contig (singletons) sont ensuite utilisés pour inférer toutes les séquences protéiques possibles. La recherche des séquences est faite en trois volets : B1 interroge la base de données RVDB-prot à l'aide de Diamond blastp, une banque ne contenant que des séquences virales. B2, les séquences ayant un résultat en B1 sont ensuite recherchées dans NCBI NR (contenant tout type de séquences protéiques), et les séquences n'ayant pas de résultat viral sont éliminées. B3, les séquences nucléotidiques correspondant aux séquences restantes après B2 sont utilisées pour effectuer une recherche dans la base de données NCBI NT pour vérifier que le meilleur résultat est bien viral (cette dernière étape sert à éliminer l'hypothèse d'une séquence non-codante qui aurait une identité moyenne avec des séquences virales mais une

identité bien plus forte avec des séquences non-codantes eucaryotes par exemple). Les séquences ayant passées ces trois étapes sont considérées comme validées.

La branche bactérienne combine trois outils qui sont utilisés en parallèle. Kraken2 et Centrifuge sont des outils basés sur l'identification de séquence par k-mers. Ils utilisent tous deux une base de séquence composée de NCBI RefSeq bacteria, fungi, ainsi que RVDB. MetaPhlan3 quant à lui utilise une approche par alignement à l'aide de bowtie2 sur une base de marqueurs développés par MetaPhlan.

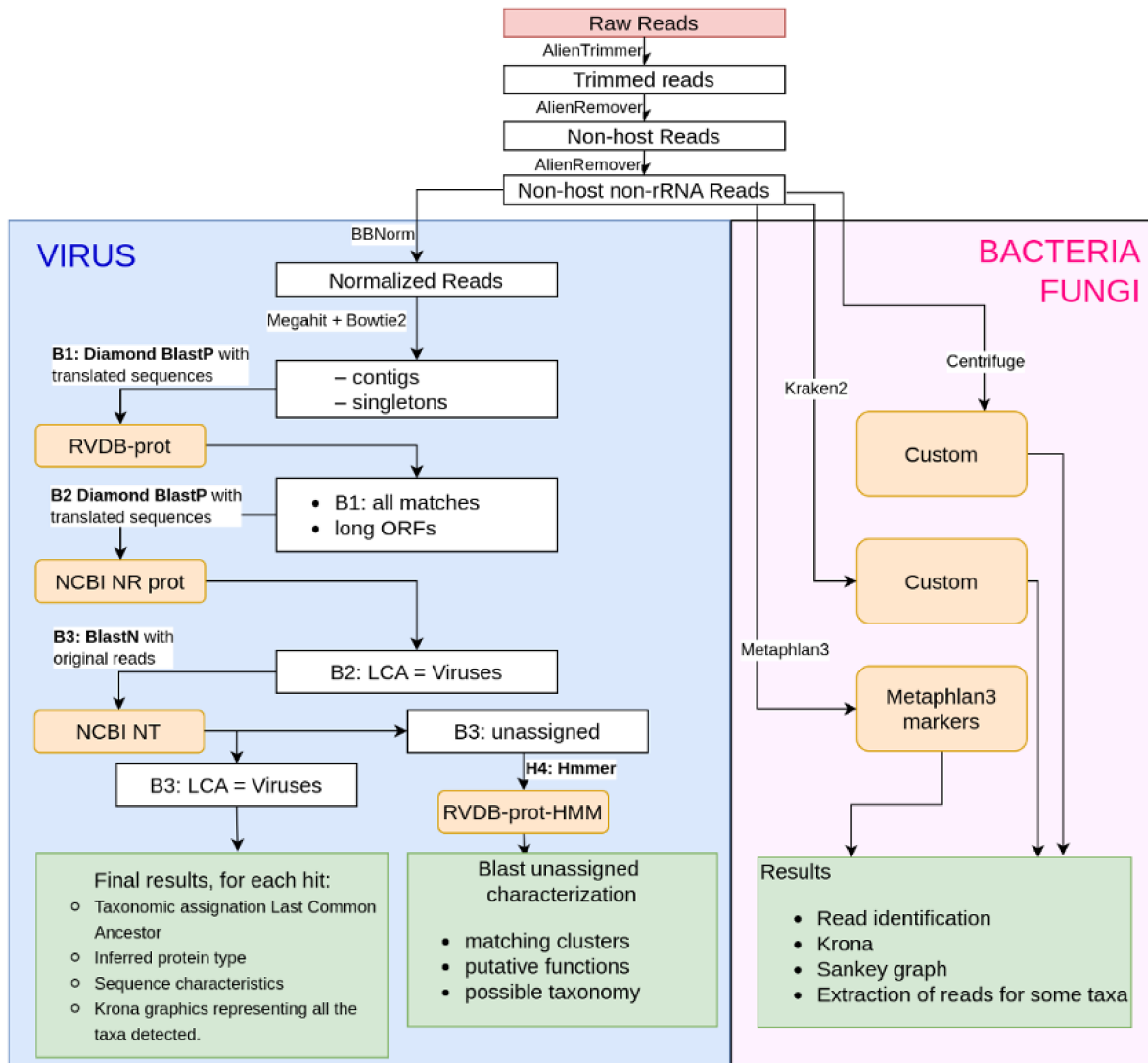


Figure 12. Logigramme résumant les étapes du pipeline bio-informatique Microseek.

Pour chaque opération, les noms des outils (Alien Trimmer, BBNorm, Megahit, Bowtie2, Diamond, Kraken2, Centrifuge, Metaphlan3) et des bases de données (NCBI NR prot, NCBI NT, RVDB-prot) sont indiqués. Figure adaptée de Pérot et al, 2022.¹⁷⁴

B. Contrôles de qualité

Les contrôles de qualité comprenaient l'ajout de 20 µl de ZymoBIOMICS Spike-In Control II (Zymo Research, Irvine, Californie) par échantillon avant l'étape d'extraction en tant que contrôle interne. La qualité a été jugée adéquate lorsque le dosage de la librairie était > 0,5 µM, que > 500 séquences de contrôle spike-in étaient détectées et que > 15 millions de lectures par échantillon étaient produites. La limite de détection et la spécificité ont été évaluées expérimentalement avec des contrôles en aveugle provenant de l'organisation Externe de qualité External Quality Assessment (EQA) / Proficiency Testing (PT) afin d'évaluer la performance globale des protocoles basés sur l'identification qualitative des micro-organismes présents dans les échantillons, au niveau de la famille, du genre, de l'espèce et du sous-type. Nous avons utilisé le panel de l'AQE de la métagénomique virale NGS (QAV204213QCMD), dans lequel des échantillons imitant le LCS ont été infectés avec divers virus humains à ADN ou ARN, à différentes concentrations. Nous avons réussi à détecter des espèces de virus jusqu'à 1,4 log copies/ml et avons été capables de détecter tous les virus et sous-types, sans erreur de spécificité dans le cas de contrôles polymicrobiens.

La limite de détection du pipeline bio-informatique appelé "Microseek" a été évalué en utilisant des ensembles de données injectées, montrant que dans des matrices complexes avec un arrière-plan élevé de séquences non microbiennes, il est possible de détecter aussi peu qu'une lecture injectée dans un ensemble de données de 50 millions de lectures.¹⁷⁴

Nous avons récemment publié un cas d'infection fœtale par le virus Zika diagnostiquée par séquençage métagénomique de la prochaine génération à partir de seulement 3 reads, après une PCR Zika initialement négative, puis confirmée ultérieurement positive faiblement par une réextraction de l'échantillon et une nouvelle PCR.¹⁷⁵

C. Interprétation des résultats

Pour les techniques de diagnostic moléculaire conventionnelles le résultat est souvent binaire (positif ou négatif) et il ne concerne que le virus ciblé. Le plus souvent ces recherches ciblées sont mises en œuvre seulement si le cas clinique est compatible avec la pathologie que provoque le virus. Dans le cadre de la métagénomique clinique, tous les virus présents dans les bases de données peuvent être détectés, seul ou en association, et pas uniquement dans le contexte de la maladie qu'ils sont connus pour provoquer. L'interprétation est donc parfois plus difficile. Pour pouvoir décrire nos résultats de métagénomique nous les avons donc classés en fonction du contexte clinique. Dans le cadre de ce travail les termes « causal », « imputable » ou « responsable » font référence à la détection d'un agent pathogène qui est bien décrit dans la littérature médicale comme provoquant des présentations cliniques similaires. Le terme « possiblement » ou « probablement causal » désigne soit (i) la détection d'un micro-organisme pour lequel une association avec les symptômes a été décrite dans la littérature sans preuve formelle de causalité, soit (ii) la détection d'un agent pathogène connu mais dans une présentation clinique atypique. Pour étayer les nouvelles associations

pathogènes supposées, nous avons été amenés à mettre œuvre des explorations complémentaires, comme l'hybridation *in situ*, l'immunohistochimie ou des PCR quantitatives supplémentaires. Dans certains cas, ces explorations nous ont permis de reclasser les virus détectés dans une catégorie différente. Notre démarche d'interprétation des résultats de mNGS est résumée dans la **Figure 13**.

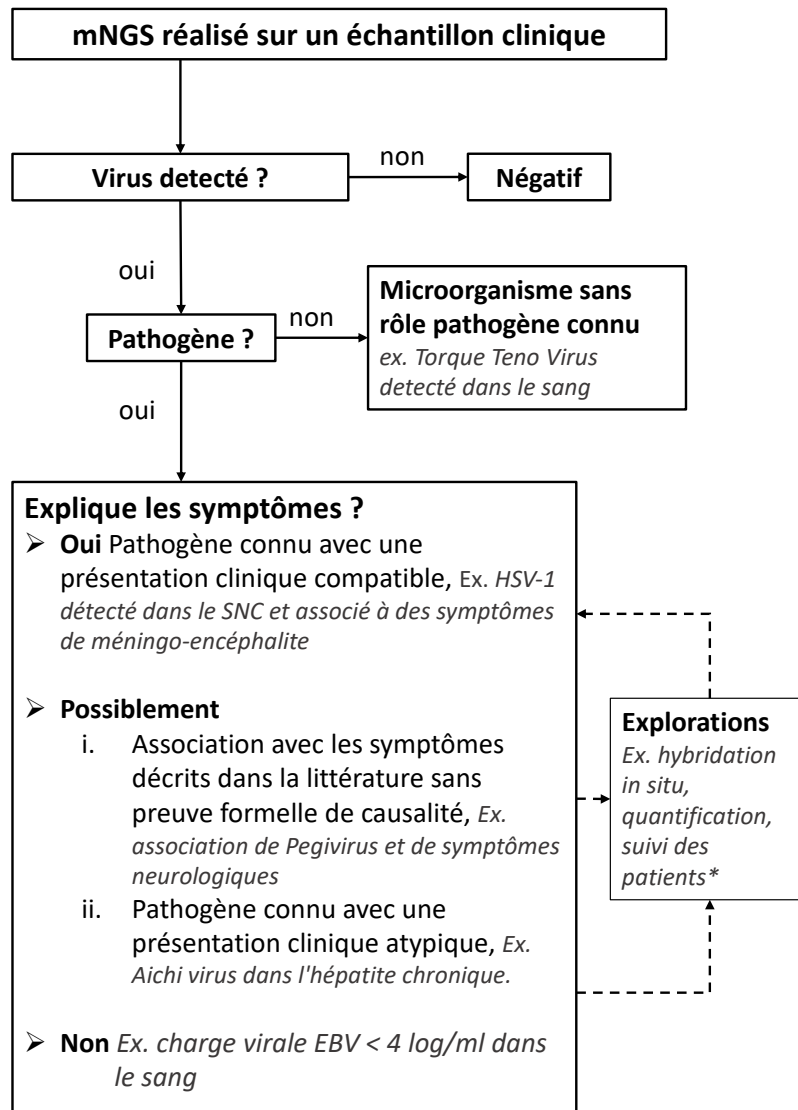


Figure 13. Logigramme d'interprétation des résultats de métagénomique

*Par exemple une technique d'immunohistochimie (IHC) du polyomavirus simien 40 (SV40) a été utilisée dans un cas de détection d'un faible signal de polyomavirus BK dans une biopsie rénale, l'IHC a été négative et nous avons donc classé ce signal comme non causal. Nous avons utilisé l'hybridation *in situ* dans les tissus hépatiques pour deux virus soupçonnés de jouer un rôle pathogène dans une série d'hépatites chroniques (Article 1)¹⁷⁶ et un cas d'hépatite aiguë inexpliquée (Circovirus humain 1)¹⁷⁷ chez des patients immunodéprimés. Nous avons donc reclassé le Aichi virus et le Circovirus humain 1 (initialement « *possiblement* » en cause) dans la catégorie "causale" pour ces cas cliniques.

II. PCR et RT-PCR spécifiques

A. RT-PCR Aichi virus

Dans l'article 1, la RT-PCR utilisée pour détecter l'AiV cible le gène VP0 comme décrit par Kitajima et al.¹⁷⁸ Des contrôles négatifs et positifs ont été utilisés dans chaque lot de RT-PCR. Les acides nucléiques ont été extraits à partir de 200 µl de plasma ou d'urine et de 400 µl de selles prétraitées à l'aide de l'instrument Emag (BioMérieux). L'ARN extrait (5µl) a été ajouté à 20µl de mélange réactionnel contenant le réactif SuperScript III (Invitrogen, Asnières-sur-Seine, France), de l'eau de qualité PCR, des sondes à une concentration finale de 200nM et des amorces à une concentration finale de 400nM. L'amplification par PCR a été réalisée avec un système de détection LC480 (Roche Diagnostic, Meylan, France), avec une transcription inverse à 50°C pendant 15 min, une dénaturation initiale à 95°C pendant 2 min, suivie de 45 cycles d'amplification avec dénaturation à 95°C pendant 15 s et hybridation/élongation à 55°C pendant 1 min.

Les tests RT-PCR dans les groupes asymptomatiques et symptomatiques ont été réalisés dans les mêmes conditions et dans les mêmes séries au moment de l'inclusion des patients. Les personnes effectuant l'essai RT-PCR n'ont pas été informées des groupes et des symptômes. De l'eau de qualité PCR a été utilisée dans chaque série de RT-PCR comme contrôle négatif. Des selles positives à l'AiV fournies par le Centre national de référence pour les virus de la gastro-entérite (Dijon, France) ont été utilisées comme contrôle positif dans chaque série de RT-PCR.

B. PCR Circovirus

Nous avons extrait les acides nucléiques de la biopsie hépatique du patient à l'aide de l'instrument geneLEAD VIII (Diagenode). Nous avons extrait des acides nucléiques de 200 µL de plasma, de sérum, de sang total, de lavage broncho-alvéolaire et de selles à l'aide de l'instrument Emag de (BioMérieux). Nous avons conçu les amorces PCR HCirV1-Fw1, 5'-ACCTGGATGGACCCTGGAAT-3', et HCirV1-Rv1, 5'-AGTTCACCAGGTTCTGC-3' (194 pb), dans le gène de la capsid de l'HCirV-1. Nous avons réalisé des PCR SYBR green avec 45 cycles d'amplification à une température d'élongation de 58°C et purifié les amplicons positifs après 45 cycles sur gel, les avons confirmés par séquençage Sanger, et nous les avons utilisés pour des dilutions en série afin de générer des courbes standards pour le calcul des charges virales.

III. Hybridation *in situ*

A. Hybridation *in situ* Aichi virus

La trousse ViewRNA *in situ* hybridization (HIS) Tissue Assay Kit 2-plex (Invitrogen) a été utilisée pour détecter les séquences d'ARN de l'AiV. Quarante sondes ADN dessinées sur la base des séquences du patient 1 ont été conçues pour cibler la région 846-2884 (peptides VP0 et VP3)

du génome de l'AiV, ces sondes étaient marquées par un signal rouge (sonde de type 1 dans le kit). Un mélange de sondes de contrôle ciblant la glycéraldéhyde-3-phosphate déshydrogénase humaine, la bêta-actine et les transcrits de pyrophosphate était marqué par un signal bleu (sonde de type 6 dans le kit). L'hématoxyline de Gill a été utilisée pour la contre-coloration.

Pour démasquer les cibles ARN, les coupes de tissus inclus en paraffine ont été déparaffinées au xylène puis traitées 20 minutes à 95°C et 15 minutes avec une protéase. La suite de l'expérimentation a été menée selon les recommandations du fabricant.

Trois expériences ont été réalisées sur 3 échantillons (n=1 foie ; n=2 rate) de 2 patients (P1 et P5). Dans ces 3 expériences successives, deux contrôles négatifs ont été utilisés dans chaque lot d'HIS : le premier contrôle négatif était une biopsie hépatique positive à l'AiV traitée sans sonde spécifique de l'AiV, le second contrôle négatif était une biopsie hépatique négative à l'AiV (mNGS négatif). Il est à noter que la biopsie hépatique négative provenait de patients différents dans chaque expérience. Des cellules vero E6 non infectées ont également été utilisées comme contrôle négatif. Des cellules Vero E6 infectées par l'AiV (génotype A), fournies par le centre national de référence des virus de la gastro-entérite (Dijon, France), ont été utilisées comme contrôle positif dans chaque expérience (**Figure 14**).

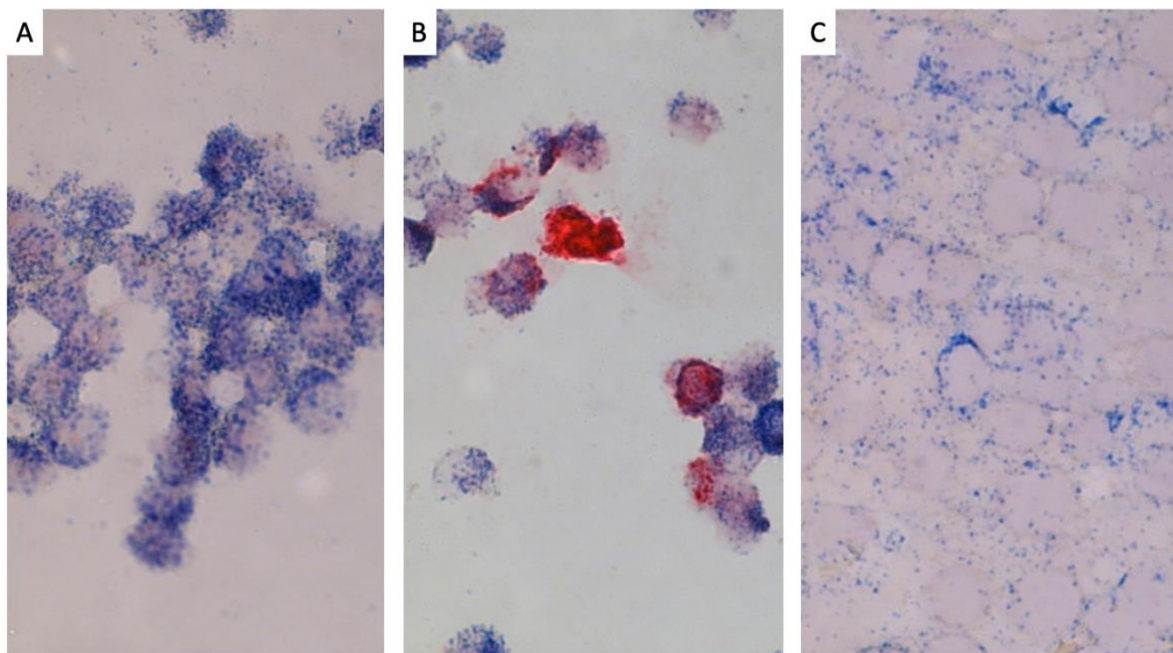


Figure 14. Contrôles négatifs et positif de l'hybridation in situ Aichi virus.

Détection par hybridation in situ chromogénique de l'ARN du virus d'Aichi (marqué en rouge) dans des cellules vero E6 de contrôle négatif (A), dans des cellules vero E6 infectées par le virus d'Aichi (B) et dans une biopsie de foie de contrôle négatif (C). Les transcrits humains de la glycéraldéhyde-3-phosphate déshydrogénase (GAPDH), de la bêta-actine (ACTB) et du pyrophosphate (PPI) ont été marqués en bleu (Gx400).

B. Hybridation *in situ* Human circovirus 1

Pour l'article 2 nous nous sommes servis de l'expérience acquise lors de l'expérimentation HIS de AiV pour faire l'expérimentation selon la même méthode. La différence principale entre les deux expérimentations était la nature du génome du virus HCirV-1 (génome à ADN monocaténaire circulaire de 2kb). Nous avons également utilisé la trousse ThermoFisher 2-plex ViewRNA ISH Tissue Assay afin de détecter des séquences d'ARN de l'HCirV-1 comme marqueur de la réplication du virus. Nous avons conçu un cocktail de 12 sondes ADN à façon pour cibler les gènes de la capsidie du HCirV-1 et 20 sondes pour cibler les gènes Rep, toutes marquées par le signal rouge (sonde de type 1 dans le kit) ; le signal bleu (sonde de type 6 dans le kit) marquait le mélange de sondes de contrôle ciblant des transcriptions les gènes de humains (GAPDH, ACTB, PPI). Nous avons appliqué un prétraitement thermique de 20 minutes et une digestion protéasique de 15 minutes aux coupes de tissus après déparaffinisation au xylène pour démasquer les cibles d'ARN et nous avons utilisé de l'hématoxyline de Harris pour la contre-coloration. Nous avons utilisé des contrôles négatifs qui comprenaient un contrôle négatif sans sonde provenant du patient infecté par le HCirV-1, comme recommandé par le fabricant, et un échantillon de biopsie de contrôle négatif provenant d'un autre patient qui a été testé négatif pour le HCirV-1 par mNGS.

IV. Analyses phylogénétiques

A. *Aichi virus*

Un génome partiel de l'AiV a été obtenu en séquençant l'AiV dans des tissus positifs ou des échantillons d'urine avec le séquençage à façon Ampliseq (Illumina). Les séquences de faible qualité ont été filtrées avec alienTrimmer,¹⁷⁹ et les séquences d'amorces ont été supprimées. Une approche par alignement utilisant Geneious Prime 2022.2.2 (Biomatters, Ltd) a été appliquée pour réaliser l'assemblage. Les séquences partielles de VP1 ont été déposées dans GenBank sous les numéros d'accès OQ134766 (patient 1 [P1]), OQ150754 (P3), OQ150752 (P4), OQ150753 (P5) et OQ161051 (P6).

Les régions nucléotidiques VP1 ont été alignées avec d'autres séquences représentatives de l'AiV provenant de GenBank en utilisant MAFFT v7.490 sous le paramètre L-INS-I, et la reconstruction phylogénétique par maximum de vraisemblance a été réalisée avec PhyML (100 bootstraps) dans le logiciel Geneious Prime 2022.2.2 (Biomatters, Ltd). Les numéros d'accès GenBank des souches de références utilisées sont détaillés dans l'arbre phylogénétique (Article 1, Figure 1).

B. *Human circovirus 1*

Après préamplification avec la polymérase QIAGEN whole transcriptome analysis ϕ 29, nous avons obtenu le génome complet de l'HCirV-1 en séquençant le plasma de la patiente, prélevé le 12 janvier 2022, qui contenait 10^{10} copies virales/mL. Nous avons appliqué une approche d'alignement en utilisant Geneious Prime 2022.2.2 pour réaliser l'assemblage final et avons déposé la séquence complète du génome dans GenBank (numéro d'accès ON677309).

Nous avons effectué des reconstructions phylogénétiques sur la séquence de la protéine de la capsid (Article 2, figure 3), sur la séquence de la protéine de la réplicase (Rep) (Annexe 2, figure 1) et sur la séquence complète des nucléotides du HCirV-1 (Annexe 2, figure 2). Nous avons aligné les cadres de lecture ouverts complets des gènes de la capsid et de la Rep ainsi que d'autres séquences représentatives de circovirus en utilisant l'outil d'alignement multiple Fast Fourier Transform (<https://www.ebi.ac.uk/Tools/msa/mafftExternal Link>) avec le paramètre L-INS-i. Nous avons effectué une reconstruction phylogénétique par maximum de vraisemblance avec PhyML mis en œuvre par le portail NGPhylogeny et évalué le soutien nodal en utilisant les soutiens de branche approximatifs de Bayes.¹⁸⁰

RESULTATS

Partie 1 : Identification de virus impliqués dans les pathologies inexpliquées du patient immunodéprimé

[Article 1: Chronic Aichi virus infection as a cause of long-lasting multi-organ involvement in patients with primary immune deficiencies](#)

À la suite de la détection par métagénomique des séquences du virus *Aichi* (AiV) dans les tissus hépatiques de trois enfants présentant un déficit immunitaire primitif et une hépatite chronique, nous avons conduit une étude prospective chez des patients atteints de déficits immunitaires primaires ou secondaires avec l'objectif d'étudier la causalité d'une infection chronique par AiV dans la pathologie hépatique de ces patients. En effet, avant notre étude la seule pathogénie décrite associée à ce virus était des gastro-entérites aiguës fébriles chez des patients immunocompétents.¹⁸¹ Un seul cas clinique d'hépatite associée à l'AiV sans preuve de causalité avait été rapporté chez un enfant présentant une atteinte chronique sévère de plusieurs organes et une agammaglobulinémie liée à l'X.¹⁸²

La participation de l'étudiant à cette étude a été la suivante :

Participation à la conception de l'étude, coordination de l'étude, interprétation des résultats de métagénomique, inclusions des témoins, mise au point et réalisation de la PCR temps réel AiV, réalisation du séquençage Ampliseq NGS du génome des AiV, analyse phylogénétique, réalisation des techniques d'hybridation *in situ*, écriture du manuscrit.

Chronic Aichi Virus Infection As a Cause of Long-Lasting Multiorgan Involvement in Patients With Primary Immune Deficiencies

Jacques Fourgeaud,^{1,2,3,a} Mathilde M. Lecuit,^{4,a} Philippe Pérot,^{3,a} Julie Bruneau,^{5,6} Beatrice Regnault,³ Nicolas Da Rocha,³ Mael Bessaud,⁷ Capucine Picard,^{8,9} Éric Jeziorski,¹⁰ Benjamin Fournier,^{4,11} Romain Levy,^{4,12} Ambroise Marçais,^{5,13} Stéphane Blanche,⁴ Pierre Frange,^{1,2} Alain Fischer,^{4,11,14} Marina Cavazzana,^{15,16} Agnès Ferroni,² Anne Jamet,^{2,17} Marianne Leruez-Ville,^{1,2,b} Marc Eloit,^{3,18,b} and Bénédicte Neven^{4,11,b}

¹Université Paris Cité, Fédération pour l'Étude et évaluation des Thérapeutiques intra-Utérines, Paris, France; ²Microbiology Department, AP-HP, Hôpital Necker Paris, France; ³Pathogen Discovery Laboratory, Institut Pasteur, Université Paris Cité, Paris, France; ⁴Pediatric Hematology Immunology and Rheumatology Unit, AP-HP, Hôpital Necker Paris, France; ⁵Laboratory of Molecular Mechanisms of Hematologic Disorders and Therapeutic Implications, Université Paris Cité, Inserm, Institut Imagine Paris, France; ⁶Department of Pathology, AP-HP, Hôpital Necker Paris, France; ⁷Laboratoire signalisation antivirale, Institut Pasteur, Université Paris Cité, Paris, France; ⁸Laboratory of Lymphocyte Activation and Susceptibility to EBV Infection, Université Paris Cité, Inserm, Institut Imagine Paris, France; ⁹Study Center for Primary Immunodeficiencies, Necker-Children's hospital, APHP Paris, France; ¹⁰Pediatric Hematology Immunology Unit, Centre Hospitalier Universitaire de Montpellier, Montpellier, France; ¹¹Laboratory of Immunogenetics of Pediatric Autoimmune Diseases, Université Paris Cité, Inserm, Institut Imagine Paris, France; ¹²Laboratory of Human Genetics of Infectious Diseases, Université Paris Cité, Inserm, Institut Imagine Paris, France; ¹³Hepatology Unit, AP-HP, Hôpital Necker Paris, France; ¹⁴Médecine expérimentale, Collège de France, Paris, France; ¹⁵Laboratory of Human Lympho-Hematopoiesis, Université Paris Cité, Inserm, Institut Imagine Paris, France; ¹⁶Department of Biotherapy, Hôpital Necker, AP-HP Paris, France; ¹⁷Department of Pathogenesis of systemic infections, Université Paris Cité, CNRS, Inserm, Institut Necker-Enfants Malades, Paris, France; and ¹⁸Département des Sciences biologiques et Pharmaceutiques, Ecole Nationale Vétérinaire d'Alfort, Maisons-Alfort, France

Background. Metagenomic next-generation sequencing (mNGS) was used to assess patients with primary or secondary immune deficiencies (PIDs and SIDs) who presented with immunopathological conditions related to immunodysregulation.

Methods. Thirty patients with PIDs or SIDs who presented with symptoms related to immunodysregulation and 59 asymptomatic patients with similar PIDs or SIDs were enrolled. mNGS was performed on organ biopsy. Specific Aichi virus (AiV) reverse-transcription polymerase chain reaction (RT-PCR) was used to confirm AiV infection and screen the other patients. In situ hybridization (ISH) assay was done on AiV-infected organs to identify infected cells. Virus genotype was determined by phylogenetic analysis.

Results. AiV sequences were detected using mNGS in tissue samples of 5 patients and by RT-PCR in peripheral samples of another patient, all of whom presented with PID and long-lasting multiorgan involvement, including hepatitis, splenomegaly, and nephritis in 4 patients. CD8+ T-cell infiltration was a hallmark of the disease. RT-PCR detected intermittent low viral loads in urine and plasma from infected patients but not from uninfected patients. Viral detection stopped after immune reconstitution obtained by hematopoietic stem cell transplantation. ISH demonstrated the presence of AiV RNA in hepatocytes (n = 1) and spleen tissue (n = 2). AiV belonged to genotype A (n = 2) or B (n = 3).

Conclusions. The similarity of the clinical presentation, the detection of AiV in a subgroup of patients suffering from immunodysregulation, the absence of AiV in asymptomatic patients, the detection of viral genome in infected organs by ISH, and the reversibility of symptoms after treatment argue for AiV causality.

Keywords. Aichi virus; chronic hepatitis; metagenomic next-generation sequencing; X-linked agammaglobulinemia; primary immune deficiency.

Primary immune deficiencies (PIDs) represent a heterogeneous group of genetically defined conditions characterized by a qualitative and/or quantitative defect of immune responses [1, 2]. Predisposition to infections is a major feature of most PIDs,

but autoimmune manifestations and inflammation are also major complications that affect all categories of PIDs [3]. These manifestations are diverse, from well-characterized autoimmunity to ill-defined immunopathological conditions related to immunodysregulation. These could be chronic benign lymphoproliferation (persistent splenomegaly, lymph node enlargement) or organ-specific inflammation, such as inflammatory bowel disease, granulomatous-lymphocytic interstitial lung disease, or hepatitis [4, 5]. Underlying triggering infection is frequently suspected but rarely documented; if present, it is often difficult to prove its causality in the onset of immunodysregulation manifestations.

Human Aichi virus (AiV) belongs to the *Kobuvirus* genus in the *Picornaviridae* family and is a small nonenveloped RNA

Received 20 January 2023; editorial decision 11 April 2023; published online 20 April 2023

^aJ. F., M. M. L., and P. P. contributed equally to this work.

^bM. L.-V., M. E., and B. N. share senior authorship.

Correspondence: B. Neven, Hôpital Necker, Service d'immuno-hématologie et de rhumatologie pédiatrique, Batiment Hamburger, 149 rue de Sèvres, 75015 Paris, France (benedicte.neven@aphp.fr).

Clinical Infectious Diseases®

© The Author(s) 2023. Published by Oxford University Press on behalf of Infectious Diseases Society of America. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

<https://doi.org/10.1093/cid/ciad237>

virus. In immunocompetent patients, AiV is associated with acute gastroenteritis infection [6–8]. As it relates to immunocompromised patients, recent observations in 4 patients with B-cell immunodeficiency suggested that AiV could be an underdiagnosed opportunistic pathogen [9–11]. In the last decade, metagenomic next-generation sequencing (mNGS) has facilitated detection of microorganisms in various samples [12–15], particularly in immunocompromised patients [16–18]. Currently, mNGS is the only available pan-microbial technique that allows the detection of rare and even novel pathogens, and it appears to be a particularly compelling technique in this context.

We implemented mNGS in our routine diagnostic workflow to investigate patients with PIDs or secondary immune deficiencies (SIDs) with poorly understood immunopathological conditions. Here, we describe a common presentation that appeared in 6 PID patients who presented with severe, chronic, and sometimes lethal multiorgan involvement, subsequently related to AiV infection. Our aim in this study was to investigate the causality of chronic AiV in these PID patients.

METHODS

Population

Between February 2020 and October 2022, 24 patients with PIDs or SIDs who presented with immunopathological conditions of unknown origin were prospectively evaluated at Hôpital Necker. In case of negative first-line microbiology testing, mNGS was performed on tissues from organ biopsies. AiV reverse-transcription polymerase chain reaction (RT-PCR) was performed on stool, plasma, and urine samples when available. In addition, mNGS was performed retrospectively on tissues of 6 additional patients with immunopathological conditions of unknown origin and available materials.

Between June 2021 and February 2023, stool, plasma, and urine samples from 59 patients with similar PIDs and SIDs were prospectively tested with AiV RT-PCR for the absence of AiV carriage in asymptomatic patients.

mNGS Sequencing on Biopsy Samples

Nucleic acids were extracted from tissues using the geneLEAD VIII instrument (Diagenode). Nucleic acids from formalin-fixed paraffin-embedded (FFPE) tissue were extracted using the RNeasy FFPE kit (Qiagen). Libraries were generated with the SMARTer Stranded Total RNA Prep Kit (Takara Bio). The libraries were sequenced in 1×151 bp on a NextSeq500 instrument (Illumina). Raw reads were processed with Microseek, based on the RVDB-prot viral database [19, 20].

In Situ Hybridization

The ViewRNA in situ hybridization (ISH) Tissue Assay Kit 2-plex (Invitrogen) was used to detect AiV RNA sequences.

Forty custom bDNA probes were designed to target the 846–2884 region (VP0 and VP3 peptides) of the AiV genome, which was indicated by a red signal. A mix of control probes targeting human glyceraldehyde-3-phosphate dehydrogenase, beta actin, and pyrophosphate transcripts was indicated by a blue signal. Gill's hematoxylin was used for counterstaining. Negative and positive controls are detailed in the [Supplementary Material \(Supplementary Figure 2\)](#).

AiV RT-PCR

AiV RT-PCR targeted the VP0 gene as previously described [21]. Negative and positive controls were used in each RT-PCR batch ([Supplementary Material](#)).

Phylogeny

A partial AiV genome was obtained by sequencing AiV in positive tissues or urine samples with custom Ampliseq sequencing (Illumina). Nonconfident nucleotides were filtered out with alienTrimmer [22], and primer sequences were removed. A mapping approach using Geneious Prime 2022.2.2 (Biomatters, Ltd) was applied to perform the assembly. The partial VP1 sequences were deposited to GenBank under accession numbers OQ134766 (patient 1 [P1]), OQ150754 (P3), OQ150752 (P4), OQ150753 (P5), and OQ161051 (P6).

VP1 nucleotide regions were aligned along with other representative sequences of AiV using MAFFT v7.490 under the L-INS-I parameter, and maximum-likelihood phylogenetic reconstruction was performed with PhyML (100 bootstraps).

Ethical Consideration

The study was approved by the Hôpital Necker Institutional Research Review Board.

RESULTS

Detection of AiV in 6 Patients

Among 30 patients with immunopathological conditions ([Table 1](#)), AiV was detected in 6 patients (20%) by mNGS ($n = 5$) on tissues (liver, spleen, kidney, lymph node) or RT-PCR on peripheral samples ($n = 1$; [Table 2](#)). Detection was made prospectively for P1, P2, and P5 and retrospectively for P3, P4, and P6. AiV sequences were identified by mNGS performed on 8 frozen biopsies from 5 patients (P1, P3, P4, P5, and P6) with a number of reads comprised between 1 and 169 (1.7 to 41.8 coverage of AB010145) and confirmed by specific AiV RT-PCR with a cycle threshold (CT) from 27 to 41 (median, 35.6; [Supplementary Table 1](#)). For P2, mNGS and AiV RT-PCR performed on FFPE liver biopsy were negative, but RT-PCR performed in urine and plasma were positive.

In the 3 patients who were prospectively followed, AiV RT-PCR was repeatedly performed in various biological fluids ([Supplementary Tables 2–4](#)). Before hematopoietic stem cell transplantation (HSCT), low viral loads (CT range, 32.9–42.2)

Table 1. Characteristics of 30 Symptomatic (Group 1) and 59 Asymptomatic (Group 2) Primary or Secondary Immune Deficiency Patients

Characteristic	Group 1			Group 2 All n = 59
	All n = 30	AiV-Positive (+) n = 6	AiV-Negative n = 24	
Median age (range), y	12.4 (1–34)	10.5 (6–13)	13.2 (1–34)	11.1 (0–40)
Gender, n (%), male	25 (83)	6 (100)	19 (79)	43 (73)
Inclusion, prospective, n (%)	24 (80)	3 (50)	21 (88)	59 (100)
Secondary immune deficiency, n (%)	3 (10)	0 (0)	3 (13)	4 (7)
Primary antibody deficiency, ^a n (%)	7 (23)	2 (33)	5 (21)	11 (19)
Combined immunodeficiency, severe combined immunodeficiency, n (%)	9 (30)	1 (17)	8 (33)	23 (39)
Post hematopoietic stem cell transplantation, ^b n (%)	11 (37)	3 (50)	8 (33)	21 (35)
Persistent hepatitis, n (%)	28 (93)	6 (100)	22 (92)	0 (0)
Digestive symptoms, ^c n (%)	12 (40)	3 (50)	9 (38)	0 (0)
Growth retardation, n (%)	17 (85)	5 (83)	12 (50)	0 (0)
Kidney disease, ^d n (%)	4 (13)	4 (67)	0 (0)	0 (0)
Splenomegaly, n (%)	17 (57)	6 (100)	11 (46)	0 (0)
AiV RT-PCR stool, positive/tested (%)	1/23 (4)	1/4 (25)	0/19 (0)	0/46 (0)
AiV RT-PCR urine, positive/tested (%)	2/22 (9)	2/3 (67)	0/19 (0)	0/28 (0)
AiV RT-PCR plasma, positive/tested (%)	3/24 (13)	3/3 (100)	0/21 (0)	0/31 (0)
Metagenomic next-generation sequencing biopsy, positive/tested (%)	5/24 (21)	5/6 (83)	0/18 (0)	0/0 (0)

Abbreviations: AiV, Aichi virus; RT-PCR, reverse-transcriptase polymerase chain reaction.

^aIncluding X-linked agammaglobulinemia and other B-cell deficiency.

^bIncluding severe combined immunodeficiency with persistent B-cell deficiency and other primary immune deficiency.

^cDiarrhea, abdominal bloating.

^dNephromegaly and/or glomerular filtration rate alteration.

were intermittently detected in plasma (3 of 3 patients), urine (2 of 3), and stool (1 of 3). The AiV RT-PCR was positive in 55%, 50%, and 100% of plasma samples and 67%, 100%, and 0% of urine samples for P1, P2, and P5, respectively (Supplementary Tables 2–4). For RT-PCR performed in cerebrospinal fluid of P1 and P2, the bronchoalveolar lavage and bone marrow aspirate of P1 were negative.

Furthermore, AiV was not detected by RT-PCR in stool (n = 46), urine (n = 28), and plasma (n = 31) from 59 asymptomatic PID or SID patients (Table 1; group 2).

Genotyping of AiV

To characterize AiV, we sequenced the VP1 gene (Figure 1) from tissue samples (P1, P3, P4, P5, and P6) and a urine sample (P2). AiV genotype could be identified in 5 of 6 patients (Table 1). Two AiV genotypes were identified: AiV genotype A (P3 and P6) and AiV genotype B (P1, P4, and P5). The VP1 region could not be covered in P2, possibly due to low viral load in the urine sample (CT, 37.7).

Clinical Description of the 6 Patients Infected With AiV

All patients suffered from PIDs; P1 and P2 from X-linked agammaglobulinemia (XLA) and P3 from type 1 immunodeficiency, centromeric region instability and facial anomalies syndrome. P4, P5, and P6 had severe combined immunodeficiency and received haploidentical bone marrow transplantation (P4 and P6) or gene therapy (P5) performed without a myeloablative conditioning regimen (CR) [23]. As a result,

T-cell immunity was partially restored in P6; humoral deficiency remained in all, requiring immunoglobulin replacement therapy (IgRT). AiV was detected 1.3–8 years after symptom onset in P1, P2, and P5 and retrospectively in P3, P4, and P6.

Hepatic disease was present in all patients, with hepatitis prevailing on aspartate aminotransferase and alanine aminotransferase (1.5–4N), anicteric cholestasis (gamma-glutamyl transferase (GGT) 1.5–20N), and hepatomegaly (Table 2). Portal hypertension was noted during the latter stages of their evolution, either mild (P2, P4, and P5) or severe (P1, P3, and P6). Liver biopsies showed marked inflammation (Metavir scoring system, A1–A3), mainly CD8⁺ T-cell infiltration (Figure 2A and 2B); fibrosis was variable (F1–F3). Regenerative nodular hyperplasia (NRH) was observed in 5 of 6 patients and peliosis was observed in 1. Hepatic granulomas were documented in 4 of 6 patients.

Impaired glomerular filtration rate (GFR) was noted in 3 patients (P2, P3, and P6) and started 6–14 years after the first symptoms. Alteration of GFR progressively worsened, reaching stage 2 to stage 3 chronic renal failure over a period of 2–5 years. Proteinuria was noted in 3 patients (P1, P3, and P6). Nephromegaly was found in 4 patients, isolated or concomitant with renal biochemical anomalies. A renal biopsy performed in P3 revealed a major destructive CD8⁺ T-cell infiltrate with lesions of tubulitis (T3). Marked fibrosis (F3) was also noted, as well as arteriolar lesions evocative of thrombotic microangiopathy.

Table 2. Prospective (P1, P2, and P5) and Retrospective (P3, P4, and P6) Clinical and Laboratory Data for 6 Male Primary Immunodeficient Patients With Identification of Aichi Virus in Clinical Samples

Characteristic	P1	P2	P3	P4	P5	P6
Gender	M	M	M	M	M	M
PID/gene defect	XLA/ <i>BTK</i>	XLA/ <i>BTK</i>	ICF1/ <i>DNMT3B</i>	SCID-X1/ <i>IL2RG</i>	SCID-X1/ <i>IL2RG</i>	SCID/ <i>DCLRE1C</i>
Treatment of PID	IgRT	IgRT	IgRT	Haplo-id HSCT, IgRT	Haplo-id HSCT, IgRT	Haplo-id HSCT, IgRT
Age at first symptoms, y	10	10	11	5	8	6
Age at first medical visit, y	12	12	13	6	9	6
Age at AiV diagnostic, y	12	18	Post mortem	14	9	Post mortem
First symptoms	Hepatitis, fatigue, recurrent fever, GF	Hepatitis	Hepatitis, GF	Hepatitis, psoriasiform skin rash	Splenomegaly, hepatitis	Hepatitis
Recurrent fever	+	-	-	-	-	-
Growth retardation	+	-	+	+	+	+
Chronic hepatitis >6 mo	+	+	+	+	+	+
Renal abnormalities	+	+	+	-	-	+
Altered glomerular filtration rate	-	+	+	-	-	+
Proteinuria	+	-	+	-	-	+
Nephromegaly	+	+	+	-	-	+
Splenomegaly	+	+	+	+	+	+
Adenomegaly	+	+	-	-	+	+
Digestive symptoms	-	-	-/+	-/+	-	-/+
Thrombopenia	+	+	+	+	+	+
Hemorrhagic syndrome	+	-	-	+	-	-
Portal space inflammation	++/mainly CD8	+/-	+/mainly CD8	+/mainly CD8T cells	+/mainly CD8	++/mainly CD8
Fibrosis	F3	F2	F3	F1-F2	F2	F3
Lobular inflammation	+++/mainly CD8	++/mainly CD8	++/mainly CD8	++/mainly CD8	++/mainly CD8	+++/mainly CD8
Granuloma	-	+	-	+	-	+
Nodular regenerative hyperplasia	+	+	+	+(mild)	+(mild)	+
AiV detection by metagenomic next-generation sequencing	Liver (+)	Liver ^a (-)	Liver (+), kidney (+)	Liver (+)	Spleen (+), lymph node (+)	Liver (+)
AiV Reverse-transcription polymerase chain reaction	Liver (+), spleen ^a (+), plasma (+), urine (+), stools (+)	Plasma (+), urine (+), stools (-)	Liver (+), kidney (+)	liver (+)	Spleen (+), lymph node (+), plasma (+), urine (-), stools (-), liver (+)	NA
AiV by in situ hybridization	liver (+), spleen (+)	NA	NA	NA	Spleen (+)	NA
AiV VP1 genotype	B	undetermined	A	B	B	A

Abbreviations: AiV, Aichi virus; CD8, CD8T cells; GF, growth failure; haplo-id HSCT, haploidentical hematopoietic stem cell transplantation; ICF1, type 1 immunodeficiency, centromeric region instability and facial anomalies; IgRT, immunoglobulin replacement therapy; M, male; NA, samples not available; PID, primary immune deficiency; SCID, severe combined immunodeficiency; XLA, X-linked agammaglobulinemia.

^aFormalin-fixed paraffin-embedded.

Splenomegaly was noted in all patients, significant and early stage in P4, P5, and P6 and modest and late in P1, P2, and P3. Of note, portal hypertension in the context of chronic hepatic disease could have contributed to splenomegaly. Centimetric intraabdominal lymphadenopathies were noted in 4 patients. Histological analysis of lymph nodes and spleen, performed in P1 and P2, respectively, displayed marked CD8 T-cell infiltration. All patients displayed thrombocytopenia that progressively worsened over the course of the disease. Bone marrow biopsies performed in 3 patients displayed nodular infiltration with CD8⁺ T cells with additional medullar granulomas in 1 case.

Digestive symptoms manifested during the course of the disease in 3 patients (P3, P4, and P6), with intermittent diarrhea and abdominal bloating. In 2 of the patients, other pathogens were concomitantly detected in stool samples (sapovirus in P4, enterovirus and *Campylobacter jejuni* in P6), possibly contributing to these digestive symptoms.

P3 and P6 ultimately died 7 and 18 years, respectively, after initial manifestations. In P3, death was precipitated by disseminated thrombotic microangiopathy that occurred in the context of chronic renal failure; P6 died from severe exudative enteropathy and portal hypertension.

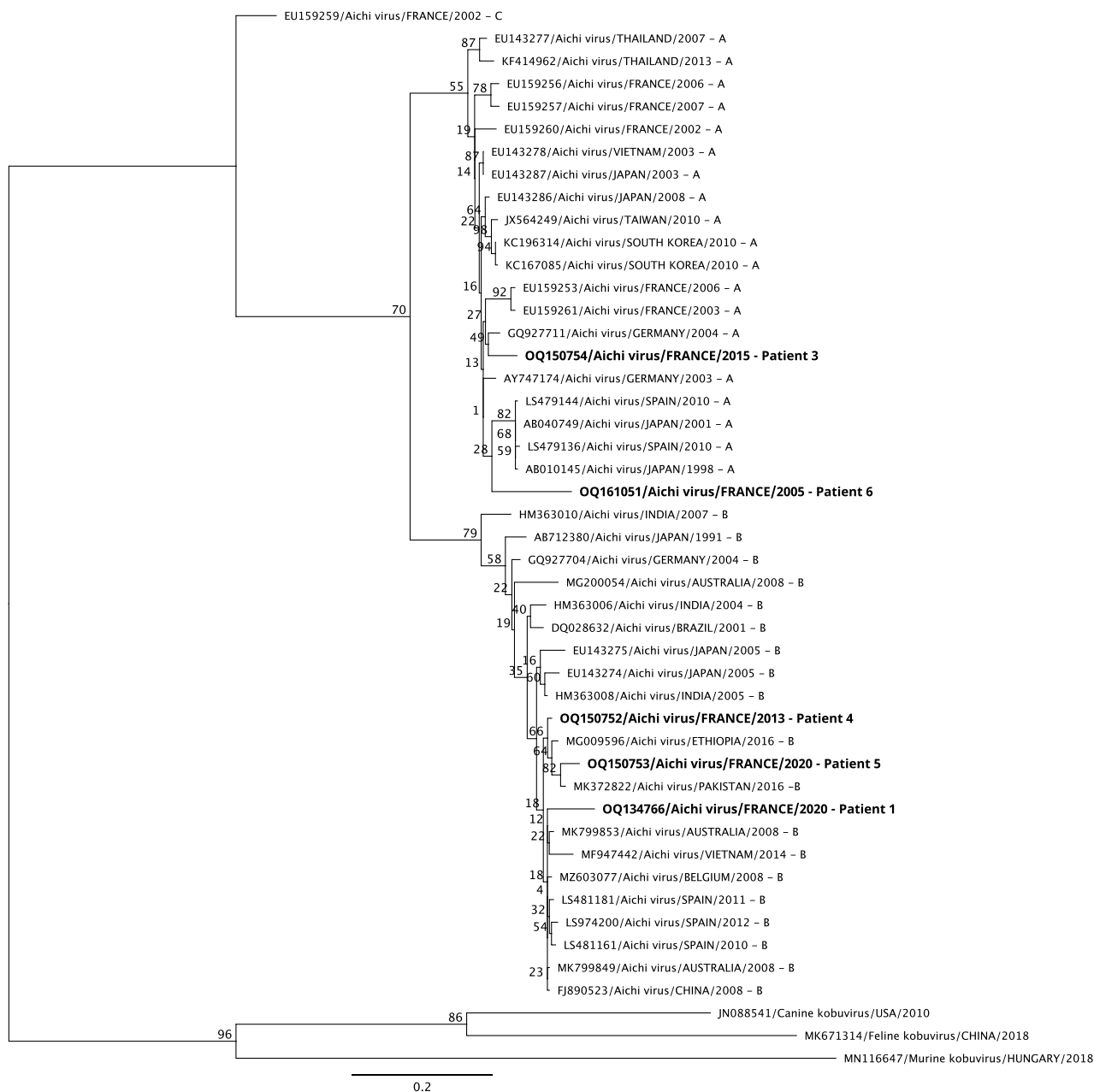


Figure 1. Phylogenetic analysis of Aichi virus. Phylogenetic analysis of VP1 sequences of Aichi virus from patients (in bold) and representative Aichi virus sequences from GenBank. Sequences were aligned with MAFFT v7.490 under the L-INS-I parameter, and maximum-likelihood phylogenetic reconstruction was performed with PhyML implemented through Geneious 2.2.1 software. GenBank accession numbers of the reference strains used are detailed in the tree. Abbreviations: A, genotype A; B, genotype B; C, genotype C.

Causality of AIV Infection in the Patient's Disease

Additional investigations were performed for P1 and P5 to study the causality of chronic AiV in PID patients who presented with a severe and chronic multi-organ involvement.

Clinical Recovery Associated With Viral Clearance

Because of the severity of P1's condition, the absence of improvement under high doses of polyvalent immunoglobulins,

the lack of available potent antiviral treatment, and the only transient benefit from steroids and ruxolitinib, HSCT was considered. The patient received a TCRab/CD19 depleted graft from his haplo-identical father preceded by a myeloablative reduced-toxicity CR ([Supplementary Material](#)). Primary graft failure required a second transplant on day 43 from his haplo-identical mother preceded by a nonmyeloablative CR. Complete and stable donor engraftment was reached at day 19. T-cell immune

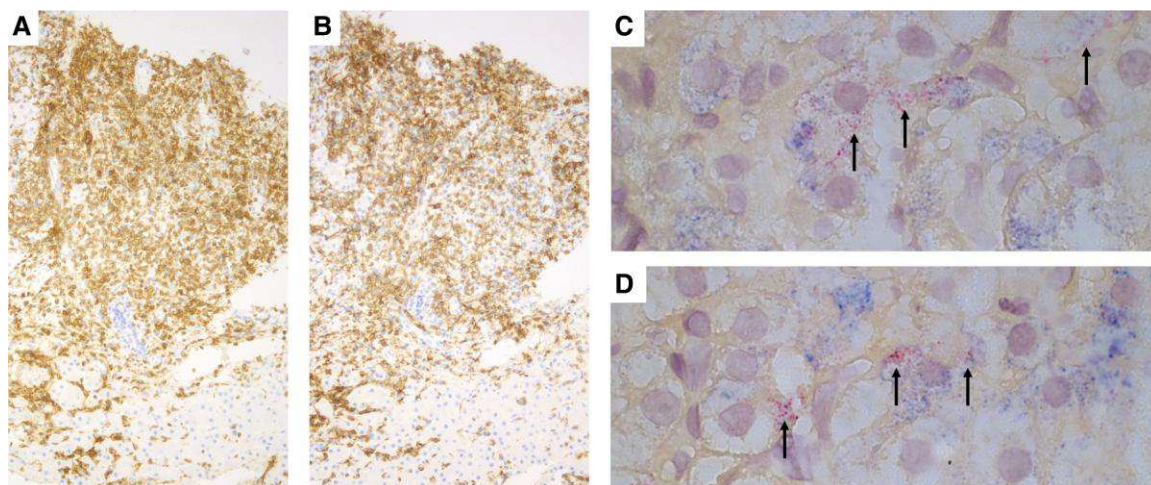


Figure 2. Lymphocytic infiltrate and Aichi virus RNA detection in patient 1 (P1) liver biopsy. Immunohistochemistry showing an infiltrate (label with brown) of CD3⁺ (A) and CD8⁺ (B) lymphocytes in P1 formalin-fixed paraffin-embedded liver section (G x 100). Chromogenic in situ hybridization (C, D) detection of Aichi virus RNA (label in red) in the hepatocytes cytoplasm (G x 400). GAPDH, ACTB, and PPI transcripts are labeled in blue.

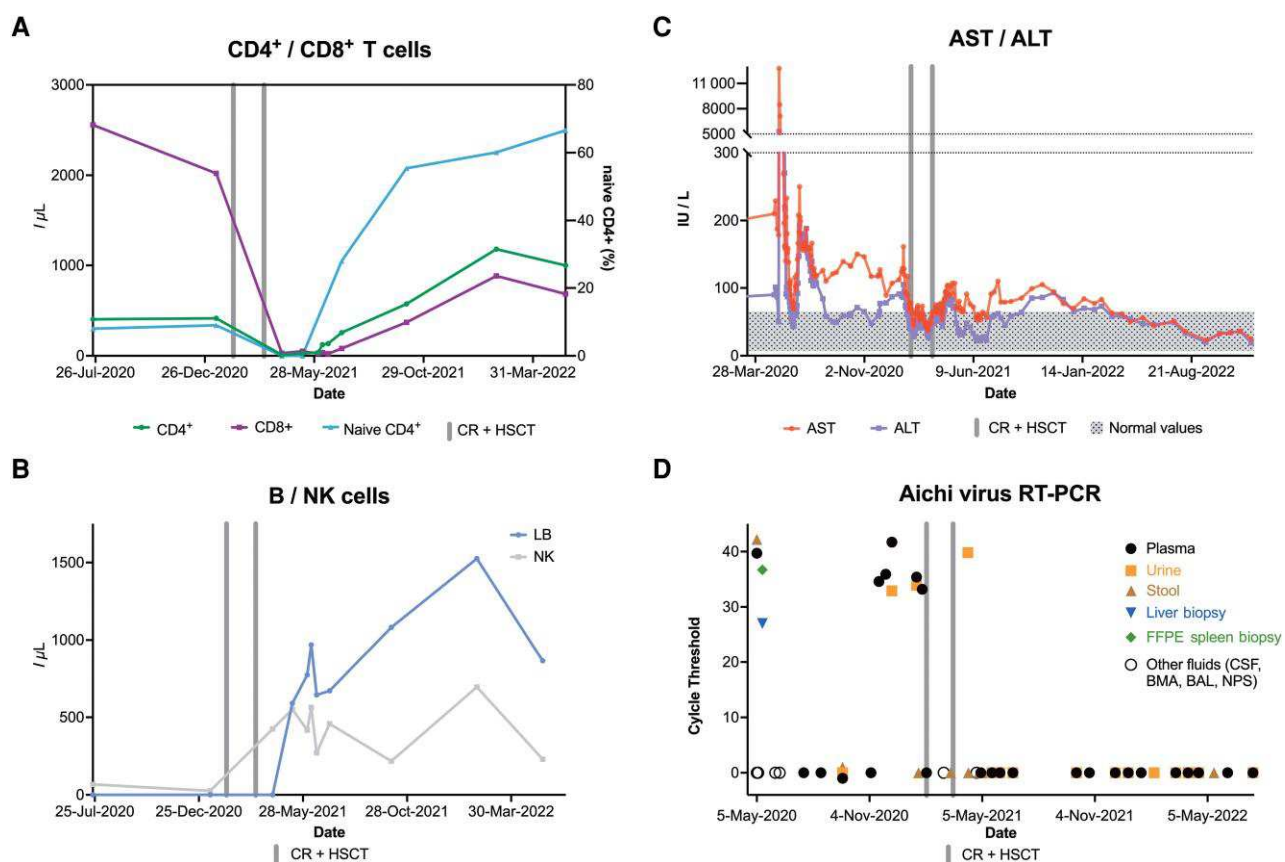


Figure 3. Patient 1 laboratory findings. Monitoring over time of CD4⁺ and CD8⁺ T-cell count (A), B and NK cells count (B), liver transaminases (C), Aichi virus RNA load in plasma, urine, stool, and other samples (D). Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BAL, bronchoalveolar lavage; BMA, bone marrow aspirate; CR, conditioning regimen; CSF, cerebrospinal fluid; FFPE, formalin-fixed paraffin-embedded; HSCT, hematopoietic stem cell transplantation; B, B cells; NK, natural killer; NPS, nasopharyngeal swab; RT-PCR, reverse-transcriptase polymerase chain reaction.

reconstitution progressively improved between 3 and 12 months post HSCT (Figure 3A). Peripheral B-cell count increased as soon as 1 month post second HSCT (Figure 3B). IgA and IgM normalized at 2 months. Interestingly, liver enzymes transiently improved during and after CR and durably normalized after 12 months (Figure 3C). Following the second HSCT, AiV RT-PCR was performed on 26 samples (plasma and urine) collected over a 15-month period; only 1 urine sample was positive on D + 24, with a low viral load showing a rapid clearance from plasma and urine (Figure 3D, Supplementary Table 2).

P4, diagnosed retrospectively, was treated with several immunosuppressive treatments without long-lasting improvement. Gene therapy preceded by low myeloablation allowed definitive remission of hepatitis and complete T-cell and B-cell immune reconstitution in parallel with the clearance of AiV, as confirmed by the negative RT-PCR done on the liver biopsy performed 12 months after HSCT [23].

Twelve months after diagnosis of AiV infection, P5 received a geno-identical transplant preceded by myeloablative reduced-toxicity CR (Supplementary Material) that allowed a full donor engraftment. Circulating B cells started to increase from month 2 post HSCT. As shown in Supplementary Table 4, most AiV RT-PCR tests were negative after HSCT. Importantly, RT-PCR performed on the liver, positive before HSCT, was negative at month 6 post HSCT.

AiV Infect the Hepatocytes Cytoplasm in ISH

To investigate the causality of AiV in hepatic disorder and splenomegaly, the liver biopsy of P1 and the splenic biopsy of P1 and P5 were tested using ISH (biopsies from patients P2, P3, P4, and P6 were no longer available). ISH showed infected (red signal) hepatocytes in the P1 biopsy (Figure 2C and 2D). Red labeling was only positive in the cytoplasm and negative in the nuclei as expected. Because the pretreatment digestion was heterogeneous across the slide, the proportion of infected hepatocytes could not be determined in the section. ISH also showed red signals in P1 (Supplementary Figure 1A) and P5 (Supplementary Figure 1B) spleen biopsies; the red labeling was intense and homogeneous over the whole spleen sections for both. However, the infected spleen cells were not identified because the tissue was overdigested. Pretreatment digestion did not challenge the specificity of the technique but only limit the possibility of assessing the type (spleen) or number (spleen, liver) of infected cells. No red signal was recorded in the negative controls (Supplementary Figure 2).

DISCUSSION

Here, we describe chronic and severe multiorgan infection with AiV in immunocompromised hosts that manifests as persistent immunopathological conditions previously of unknown origin. The similarity of the clinical presentation, the detection of AiV

in patients suffering from immunodysregulation, the absence of AiV in asymptomatic matched immunodeficient patients, the detection of the viral genome in infected organs using ISH, and the reversibility of symptoms associated with AiV clearance after treatment argue for causality of AiV in these patients' clinical presentation as they fulfill a number of postulates established by Hill [24] and reviewed by Fredricks and Relman [25].

Similarity of Clinical Presentation and Association With B-Cell Deficiency

The 6 patients in this series (including P4, P5, and P6 previously reported [23]) all carried B-cell deficiency (associated with abnormalities of the T-cell compartment in 2 of 6 cases and low numbers of natural killer cells numbers in 3 of 6). All patients had a similar presentation characterized by long-lasting multiorgan inflammation with CD8+ T-cell infiltration. The liver was consistently affected and resulted in an increase in liver transaminases and/or cholestatic hepatitis. NRH and peliosis were also documented. The kidney was involved in 2 of 3 of our patients. The latter gives rise to nephromegaly and interstitial nephritis, as well as histologic features of thrombotic microangiopathy in 1 case. Significant lymphoproliferation characterized by splenomegaly and intraabdominal lymph nodes was also constant. Four cases of suspected AiV-disseminated infection have been previously described in children with primary or secondary B-cell deficiency and with very similar features including the unusual renal disease [9–11]. The strength of the association of AiV with these symptoms supports the causality of AiV. The occurrence of chronic AiV infection in B cell-deficient patients is reminiscent of chronic infections with enterovirus, another *Picornaviridae* that can occur in humoral-deficient patients [26, 27]. This emphasizes the plausibility of such association. AiV was documented in 32% of patients with PIDs and immunopathological conditions that we screened but in none of the asymptomatic matched immunodeficient patients. Among the symptomatic group, features of AiV-positive and AiV-negative patients largely overlapped except for the kidney abnormalities only documented in AiV-positive patients. Interestingly, immunodysregulation-related manifestations such as liver disease and splenomegaly are generally less common in XLA patients compared with other conditions such as common variable or combined immunodeficiency [3, 28]. This may support the concept that these symptoms are driven by the viral infection rather than the disease itself. Based on our 6 observations and the 4 reported cases, B-cell immunity seems critical in immunity against AiV. Infections occurred despite IgRT, while serological studies suggested that most adults worldwide have been exposed [29–32]. A neutralization assay of the commercial immunoglobulin preparations could not be conducted but would be informative to better understand the reason for this immune evasion. CD8+

T-cell infiltration of affected organs is also a key feature, and testing for anti-AiV specificity of the cells would be of interest.

AiV infection can be easily overlooked. Indeed, based on our observation, diagnosis of AiV infection can be challenging. The RNA loads in plasma and urine were low and intermittent, and the shedding in stool was rare. The high sequence diversity between the different strains of the same species among *Picornaviridae* members due to the high error rates of RNA-dependent RNA polymerase [33] and the absence of a commercial AiV detection test are some of the pitfalls that highlight the value of mNGS in this context. False-negative results could be avoided by mNGS with deep sequencing and by repeating RT-PCR on stool, plasma, and urine over time.

Phylogenetic analysis showed that chronic AiV infections in these immunocompromised patients can be caused by either genotype A (2 of 5) or genotype B (3 of 5), probably depending on the circulation of the virus in the community at the time of contamination.

Finally, using specific ISH, AiV RNA sequences could be documented at the cellular level in the cytoplasm of P1's hepatocytes, concomitantly to an increase in liver transaminases. Given the lytic cycle of kobuviruses [34], it reinforces the suspected pathogenicity of AiV. Technical issues did not allow the evaluation of the percentage of infected hepatocytes nor the host cells in the spleen of P5, which also showed a strong signal in ISH. Due to the documentation of peliosis in P1 and thrombotic microangiopathy in P3, both suggesting vascular damages one can only wonder about AiV tropism for endothelial cells.

Reversibility of the disease concomitantly to viral clearance of AiV, the only pathogen common to all of the patients, also argues for its causality. Indeed, clearance of the AiV genome from plasma and urine in P1 and P5 was associated with clinical and biological improvement after successful HSCT. Clearance from the liver was also documented in P3 and P5 in whom liver biopsies were performed 6 and 12 months after cellular therapy, respectively. Observations from Bekassy et al and Bucciol et al [11, 35] are in line with this temporality.

Our preliminary study has limitations. Even if the resolution of the disease coincides with viral clearance, arguing for a causal relationship between AiV infection and the disease, we cannot exclude that the conditioning regimen, the engraftment, and the subsequent correction of the immune deficiency may have cured an autoimmune process that could not be previously identified. In addition, material was not available to demonstrate the involvement of AiV at the cellular level by ISH in renal or digestive biopsies.

In conclusion, we bring pieces of evidence in favor of the causality of human AiV infection in patients with B-cell deficiency who develop multiorgan involvement. The symptoms such as hepatitis, splenomegaly, and nephromegaly with CD8 + T-cell infiltration should be promptly and properly investigated, keeping in mind that diagnosis of AiV infection can be challenging. Restoration of B-cell immunity allows AiV

clearance and cure of symptoms. Prevalence and burden of chronic AiV infection in such patients needs to be further evaluated in a larger prospective, controlled study.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. The authors thank the Neuropathology Department of La Pitié-Salpêtrière Hospital for providing access to the in situ hybridization platform. They thank the National Reference Center for Gastroenteritis Viruses (Dijon, France) for providing an Aichi virus strain used as a positive control.

Financial support. This work was supported by the Assistance Publique - Hôpitaux de Paris (Hôpital Universitaire Necker-Enfants malades and Centre de Référence Déficits Immunitaires Hérités), and the Institut Pasteur.

Potential conflicts of interest. The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Tangye SG, Al-Herz W, Bousfiha A, et al. Human inborn errors of immunity: 2022 update on the Classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol* **2022**; 42:1473–507.
2. Long A, Kleiner A, Looney RJ. Immune dysregulation. *J Allergy Clin Immunol* **2023**; 151:70–80.
3. Fischer A, Provot J, Jais J-P, Alcais A, Mahlaoui N, Members of the CEREDIH French PID Study Group. Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. *J Allergy Clin Immunol* **2017**; 140:1388–1393.e8.
4. Agarwal S, Cunningham-Rundles C. Gastrointestinal manifestations and complications of primary immunodeficiency disorders. *Immunol Allergy Clin North Am* **2019**; 39:81–94.
5. Lopes JP, Ho H-E, Cunningham-Rundles C. Interstitial lung disease in common variable immunodeficiency. *Front Immunol* **2021**; 12:605945.
6. Ambert-Balay K, Lorrot M, Bon F, et al. Prevalence and genetic diversity of Aichi virus strains in stool samples from community and hospitalized patients. *J Clin Microbiol* **2008**; 46:1252–8.
7. Rivadulla E, Varela MF, Romalde JL. Epidemiology of Aichi virus in fecal samples from outpatients with acute gastroenteritis in northwestern Spain. *J Clin Virol* **2019**; 118:14–9.
8. Rivadulla E, Romalde JL. A comprehensive review on human Aichi virus. *Viral Sin* **2020**; 35:501–16.
9. Bucciol G, Moens L, Payne K, et al. Chronic Aichi virus infection in a patient with X-linked agammaglobulinemia. *J Clin Immunol* **2018**; 38:748–52.
10. Meyts I, Bucciol G, Jansen K, Wollants E, Breuer J. Aichivirus: an emerging pathogen in patients with primary and secondary B-cell deficiency. *J Clin Immunol* **2022**; 43:532–5.
11. Bekassy Z, Ehinger M, Pronk LN, Pronk CJ. Immunologic control of disseminated Aichi virus infection in X-linked agammaglobulinemia by transplantation of TcR $\alpha\beta$ -depleted haploidentical hematopoietic cells. *J Clin Immunol* **2022**; 42:1401–4.
12. Somasekar S, Lee D, Rule J, et al. Viral surveillance in serum samples from patients with acute liver failure by metagenomic next-generation sequencing. *Clin Infect Dis* **2017**; 65:1477–85.
13. Schlager R. Clinical metagenomics—from proof-of-concept to routine use. *Clin Chem* **2022**; 68:997–9.
14. Wilson MR, Sample HA, Zorn KC, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N Engl J Med* **2019**; 380:2327–40.
15. Chen S, Kang Y, Li D, Li Z. Diagnostic performance of metagenomic next-generation sequencing for the detection of pathogens in bronchoalveolar lavage fluid in patients with pulmonary infections: systematic review and meta-analysis. *Int J Infect Dis* **2022**; 122:867–73.
16. Bodemer C, Sauvage V, Mahlaoui N, et al. Live rubella virus vaccine long-term persistence as an antigenic trigger of cutaneous granulomas in patients with primary immunodeficiency. *Clin Microbiol Infect* **2014**; 20:O656–663.

17. Frémond M-L, Pérot P, Muth E, et al. Next-generation sequencing for diagnosis and tailored therapy: a case report of astrovirus-associated progressive encephalitis. *J Pediatric Infect Dis Soc* **2015**; 4:e53–57.
18. Pérot P, Bielle F, Bigot T, et al. Identification of umbre orthobunyavirus as a novel zoonotic virus responsible for lethal encephalitis in 2 French patients with hypogammaglobulinemia. *Clin Infect Dis* **2021**; 72:1701–8.
19. Bigot T, Temmam S, Pérot P, Eloit M. RVDB-prot, a reference viral protein database and its HMM profiles. *F1000Res* **2019**; 8:530.
20. Pérot P, Bigot T, Temmam S, Regnault B, Eloit M. Microseek: a protein-based metagenomic pipeline for virus diagnostic and discovery. *Viruses* **2022**; 14:1990.
21. Kitajima M, Hata A, Yamashita T, Haramoto E, Minagawa H, Katayama H. Development of a reverse transcription-quantitative PCR system for detection and genotyping of Aichi viruses in clinical and environmental samples. *Appl Environ Microbiol* **2013**; 79:3952–8.
22. Criscuolo A, Brisse S. Alientrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **2013**; 102:500–6.
23. Riller Q, Fourgeaud J, Bruneau J, et al. Late-onset enteric virus infection associated with hepatitis (EVAH) in transplanted SCID patients [manuscript published online ahead of print 8 March 2013]. *J Allergy Clin Immunol* **2023**; doi:10.1016/j.jaci.2022.12.822.
24. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* **1965**; 58:295–300.
25. Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* **1996**; 9:18–33.
26. Bearden D, Collett M, Quan PL, Costa-Carvalho BT, Sullivan KE. Enteroviruses in X-linked agammaglobulinemia: update on epidemiology and therapy. *J Allergy Clin Immunol Pract* **2016**; 4:1059–65.
27. Jones TPW, Buckland M, Breuer J, Lowe DM. Viral infection in primary antibody deficiency syndromes. *Rev Med Virol* **2019**; 29:e2049.
28. Cunningham-Rundles C. Common variable immune deficiency: dissection of the variable. *Immunol Rev* **2019**; 287:145–61.
29. Oh D-Y, Silva PA, Hauroeder B, Diedrich S, Cardoso DDP, Schreier E. Molecular characterization of the first Aichi viruses isolated in Europe and in South America. *Arch Virol* **2006**; 151:1199–206.
30. Goyer M, Aho L-S, Bour J-B, Ambert-Balay K, Pothier P. Seroprevalence distribution of Aichi virus among a French population in 2006–2007. *Arch Virol* **2008**; 153:1171–4.
31. Ribes JM, Montava R, Téllez-Castillo CJ, Fernández-Jiménez M, Buesa J. Seroprevalence of Aichi virus in a Spanish population from 2007 to 2008. *Clin Vaccine Immunol* **2010**; 17:545–9.
32. Yamashita T, Sakae K, Ishihara Y, Isomura S, Utagawa E. Prevalence of newly isolated, cytopathic small round virus (Aichi strain) in Japan. *J Clin Microbiol* **1993**; 31:2938–43.
33. Kok CC, McMinn PC. Picornavirus RNA-dependent RNA polymerase. *Int J Biochem Cell Biol* **2009**; 41:498–502.
34. Zell R. Picornaviridae—the ever-growing virus family. *Arch Virol* **2018**; 163:299–317.
35. Bucciol G, Tousseyn T, Jansen K, et al. Hematopoietic stem cell transplantation cures chronic Aichi virus infection in a patient with X-linked agammaglobulinemia. *J Clin Immunol* **2021**; 41:1403–5.

Article 2: Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France

Dans cette étude, nous rapportons l'identification d'un nouveau circovirus infectant une patiente de 61 ans transplantée cardiaque et pulmonaire qui présentait une hépatite aigue d'origine indéterminée. L'objectif de ce travail était également d'étudier la responsabilité de ce virus dont les séquences n'étaient pas connues dans les bases de données de références. En plus de la description du génome entier du virus et de son analyse phylogénétique, notre démarche scientifique s'est appuyée sur la recherche d'arguments en faveur de la responsabilité du virus pour satisfaire les critères de Fredricks et Relman (VI).

La participation de l'étudiant à cette étude a été la suivante :

Participation à la conception de l'étude, coordination de l'étude, interprétation des résultats de métagénomique, inclusions des témoins, réalisation des PCR HCirV-1 et des qPCR Torque teno virus, réalisation des techniques d'hybridation in situ ; analyse et mise en forme de l'ensemble des résultats, écriture du matériel et méthode et des résultats du manuscrit.

Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France

Philippe Pérot,¹ Jacques Fourgeaud,¹ Claire Rouzaud,¹ Béatrice Regnault, Nicolas Da Rocha, Hélène Fontaine, Jérôme Le Pavec, Samuel Dolidon, Margaux Garzaro, Delphine Chrétien, Guillaume Morcrette, Thierry Jo Molina, Agnès Ferroni, Marianne Leruez-Ville, Olivier Lortholary,² Anne Jamet,² Marc Eloit²

In March 2022, a 61-year-old woman in France who had received a heart-lung transplant sought treatment with chronic hepatitis mainly characterized by increased liver enzymes. After ruling out common etiologies, we used metagenomic next-generation sequencing to analyze a liver biopsy sample and identified an unknown species of circovirus, tentatively named human circovirus 1 (HCirV-1). We found no other viral or bacterial sequences. HCirV-1 shared 70% amino acid identity with the closest known viral sequences. The viral genome was undetectable in blood samples from 2017–2019, then became detectable at low levels in September 2020 and peaked at very high titers (10^{10} genome copies/mL) in January 2022. In March 2022, we found $>10^8$ genome copies/g or mL in the liver and blood, concomitant with hepatic cytolysis. We detected HCirV-1 transcripts in 2% of hepatocytes, demonstrating viral replication and supporting the role of HCirV-1 in liver damage.

Solid organ transplant recipients are highly susceptible to infections that sometimes have uncommon clinical manifestations (1–3). Etiologic diagnosis of hepatitis is even more problematic in organ transplant recipients because a wide range of possible drug toxicities induced by immunosuppressive

therapies must be considered (4). Diagnosing unexplained hepatitis remains a challenge, as exemplified by 74 cases of acute hepatitis of unknown etiology reported in children in the United Kingdom, 3 cases in Spain, and up to 5 possible cases in Ireland (5). Among transplant recipients, it is especially critical to detect infection, often viral, to prevent unnecessary interruption of treatment. Cases of direct viral transmission from the graft itself have been reported (6,7). Agnostic metagenomic analysis of clinical samples based on next-generation sequencing (NGS) combined with specifically designed PCRs constitute promising tools for diagnosing infection with unforeseen or novel microorganisms (8). We report identifying a novel circovirus infecting humans, tentatively named human circovirus type 1 (HCirV-1), and its role in causing liver damage in a solid organ transplant recipient.

Case Report

A woman, 61 years of age, sought treatment in France with cytolytic hepatitis of unknown etiology. The patient had received a heart-lung transplant 17 years earlier because of Eisenmenger syndrome related to

Author affiliations: Institut Pasteur Pathogen Discovery Laboratory, Paris, France (P. Pérot, B. Regnault, N. Da Rocha, D. Chrétien, M. Eloit); The OIE Collaborating Center for the Detection and Identification in Humans of Emerging Animal Pathogens, Paris (P. Pérot, B. Regnault, N. Da Rocha, D. Chrétien, M. Eloit); Institut Imagine, Paris (J. Fourgeaud, M. Leruez-Ville); Université Paris Cité, Paris (J. Fourgeaud, A. Jamet); Necker-Enfants Malades Hospital, Paris (J. Fourgeaud, G. Morcrette, T.J. Molina, A. Ferroni, M. Leruez-Ville, A. Jamet); Hôpital Necker Enfants-Malades Centre d'Infectiologie Necker-Pasteur, Paris (C. Rouzaud, M. Garzaro, O. Lortholary); Groupe Hospitalier Paris Saint Joseph-Marie Lannelongue, Équipe Mobile de

Microbiologie Clinique, Paris (C. Rouzaud); Hôpital Cochin Département d'Hépatologie-Addictologie, Paris (H. Fontaine); Université Paris-Sud, Paris (J. Le Pavec); Hôpital Marie Lannelongue Service de Pneumologie et Transplantation Pulmonaire, Le Plessis-Robinson, France (J. Le Pavec, S. Dolidon); Institut Necker Enfants Malades, Paris (A. Jamet); Ecole Nationale Vétérinaire d'Alfort, Maisons-Alfort, France (M. Eloit)

DOI: <https://doi.org/10.3201/eid2902.221468>

¹These authors contributed equally to this article.

²These senior authors contributed equally to this article.

ventricular septal defect. Her maintenance immunosuppressive regimen included tacrolimus, mycophenolate mofetil, and prednisone. She had received pulse steroids 16 months before the hepatitis occurred to treat acute lung allograft dysfunction. The patient had had a localized left upper lobe pulmonary adenocarcinoma treated by surgery 2.5 years before this episode. She also had a history of chronic renal failure from calcineurin inhibitor (Cockcroft-Gault up to 23 mL/min), *Aspergillus* spp. and *Scedosporium* spp. bronchopulmonary infections, and parvovirus B19 and COVID-19 infections. In November 2021, 14 months after receiving pulsed corticosteroid therapy, the patient was hospitalized for cytomegalovirus (CMV) colitis with ganciclovir-resistant virus requiring treatment with foscavir, then maribavir, followed by letermovir for long-term maintenance therapy. At

that time ultrasound examination detected cytolysis without symptoms or liver abnormalities (Figure 1).

Four months after that hospitalization, hepatitis worsened; transaminase serum levels increased to up to 40 times the normal upper limit (Figure 1), leading to hospital readmission. The patient was asymptomatic except for recent weight loss but without biological severity because the prothrombin time level was normal. We ruled out common infections, including hepatitis A, B, C, and E, as well as HIV, CMV, herpes simplex virus, varicella zoster virus (VZV), human herpes virus 6 (HHV6), adenovirus, enterovirus in blood and feces, parvovirus B19, toxoplasmosis, syphilis, and leptospirosis. Only Epstein-Barr virus (EBV) PCR was positive ($10^{4.4}$ genome copies/mL of blood). There was no evidence for a timeline relationship between levels of liver enzymes and use of various medications by the patient.

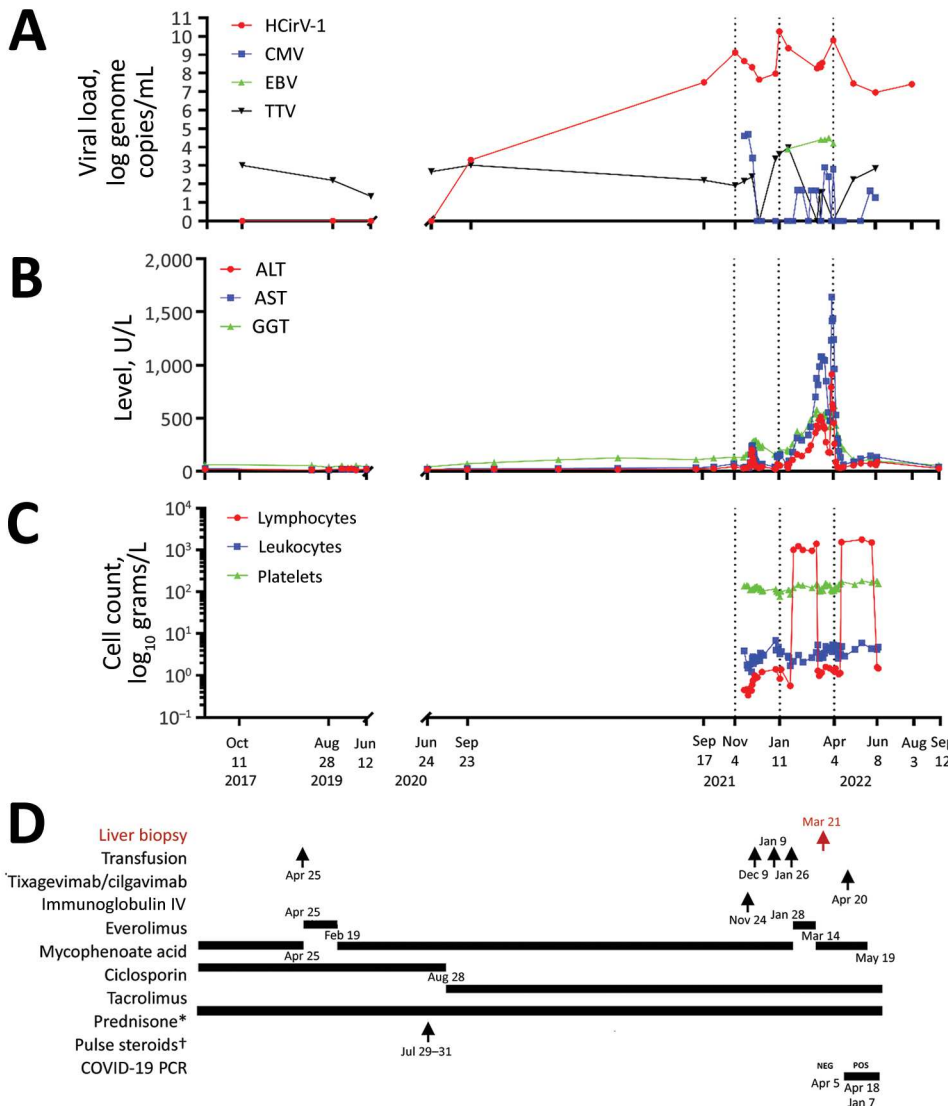


Figure 1. Clinical and laboratory data over time for a heart-lung transplant patient in France who had cytolytic hepatitis caused by HCirV-1 develop. A–C) Monitoring of patient over time: A) viral loads; B) liver cytolysis markers; C) blood cell counts. D) Patient’s treatment history; red indicates timing of liver biopsy (March 21, 2022). Average values are depicted for HCirV-1 and TTV viral loads in plasma and blood on June 8 and August 3, 2022. *Dose 5 mg/day; †, solumedrol, dose 500 mg/day. ALT, alanine transaminase; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; GGT, gamma-glutamyl transferase; HCirV-1, human circovirus type 1; IV, intravenous; TTV, Torque teno virus.

She reported no use of paracetamol (acetaminophen), alcohol, illicit drugs, or phytotherapy (use of herbal/plant supplements). A cardiac ultrasound did not show cardiac failure. We found no cardiac arrhythmia or vascular or biliary abnormalities from liver ultrasound and Doppler examinations, computerized tomography scan, or biliary magnetic resonance imaging results. Test results for autoimmune abnormalities were negative. Finally, aceruleoplasminemia test results were normal.

Pathological examination of a liver biopsy showed lobular hepatitis, slightly inflammatory, without epithelioid granuloma, substantial portal inflammation and fibrosis, or lymphoproliferation (Figure 2). Results of Ziehl, EBV, CMV, and adenovirus stains and bacteriological examinations of the biopsy sample, including mycobacterial cultures and PCR, broad-range 16S ribosomal RNA gene PCR, were negative. Herpes simplex virus 1 and 2, CMV, adenovirus, and enterovirus PCR results were negative; EBV PCR was positive (135 genome copies/ μg DNA, cycle threshold = 35) but not ascribed to acute hepatitis. Overall, the lymphocytic pattern of inflammation in contrast to the lack of eosinophil infiltration, vascular or biliary injuries, and autoimmune liver disease (no plasmacytes), strongly suggested viral hepatitis. We considered the extent of both centrilobular necrosis and resorption by macrophages, which suggested recent liver cytolysis. After these first-line negative results from the liver biopsy, we conducted metagenomic NGS (mNGS), which revealed presence of the novel circovirus, HCirV-1. We further explored and quantified HCirV-1 using quantitative PCR (qPCR) on liver tissue, historical blood samples, feces, bronchoalveolar lavage, urine, and saliva samples from the patient (Table).

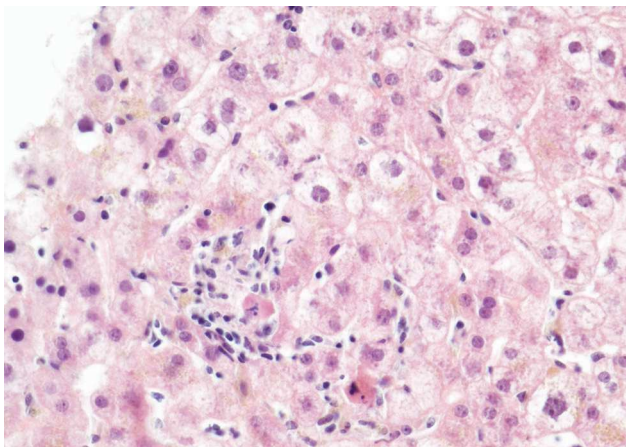


Figure 2. Lobular hepatitis in a heart-lung transplant patient in France. Apoptotic bodies, hepatocyte swelling and ballooning were present surrounded by a slight inflammatory infiltrate made of lymphocytes and histiocytes. Hematoxylin and eosin stain; original magnification $\times 40$.

In November 2022, the patient's clinical condition was stable, and we noted complete recovery from cytotoxicity. In the context of her recent infectious diseases, we minimized the patient's immunosuppressive treatment, discontinuing mycophenolate mofetil (Figure 1). When we tried to ascertain possible sources of her HCirV-1 infection, the patient reported contact with 2 cats, both of which had contact with birds and rodents. HCirV-1 PCR was negative in the feces of the cats 2 months after her diagnosis. The patient reported no international travel. She did report receiving a blood transfusion 17 months before HCirV-1 was first detected (Figure 1).

Methods

Description of Controls

We performed retrospective HCirV-1 qPCRs on feces and blood samples from immunocompromised or immunocompetent control patients with or without hepatitis (Appendix Table 1, <https://wwwnc.cdc.gov/EID/article/29/2/22-1468-App1.pdf>). We also reviewed the results of mNGS routinely performed since 2019 on all liver biopsies in our laboratory to detect pathogens (Appendix Table 2).

Nucleic Acid Extraction

We extracted nucleic acids from the patient's liver biopsy using the geneLEAD VIII instrument (Diagenode, <https://www.diagenode.com>). We extracted nucleic acids from 200 μL of plasma, serum, whole blood, bronchoalveolar lavage (BAL), and feces using the bioMérieux Emag instrument (<https://www.biomerieux.com>).

mNGS

We generated complementary DNA from a biopsy from the patient's liver using the Invitrogen SuperScript IV First-Strand Synthesis System kit (ThermoFisher Scientific, <https://www.thermo-fisher.com>) and used MALBAC amplification (Yikon Genomics, <https://en.yikongenomics.com>) as described elsewhere (9) to use as input for Illumina DNA Prep (<https://www.illumina.com>). We sequenced the library in 1×151 bp on an Illumina NextSeq500 instrument generating ≈ 80 million raw reads. We processed raw reads with Microseek (10) using the RVDB-prot reference viral database (11).

qPCR HCirV-1

We designed PCR primers HCirV1-Fw1, 5'-ACCTG-GATGGACCCTGGAAT-3', and HCirV1-Rv1, 5'-AGAGTTCCACCAGGTTCTGC-3' (194 bp), in the capsid

gene of HCirV-1. We performed qPCRs in SYBR green format with 45 cycles of amplification at an annealing temperature of 58°C and purified positive amplicons after 45 cycles on gel, confirmed them by Sanger sequencing, and used them for serial-dilutions to generate standard curves for the calculation of viral loads.

Acquisition of Complete Genome

After preamplification with the QIAGEN whole transcriptome analysis ϕ 29 polymerase (<https://www.qiagen.com>), we obtained the full genome of HCirV-1 by sequencing plasma from the patient, sampled on January 12, 2022, which harbored 10^{10} viral copies/mL. We applied an iterative mapping approach using Geneious Prime 2022.1.1 (<https://www.geneious.com>) to perform the final assembly and deposited the complete genome sequence at GenBank (accession no. ON677309).

Phylogeny

We performed phylogenetic reconstructions on the capsid protein sequence (Figure 3), on the replicase (Rep) protein sequence (Appendix Figure 1), and on the complete nucleotide sequence of HCirV-1 (Appendix Figure 2). We aligned complete open reading frames of the capsid and Rep genes along with other representative sequences of circoviruses using Multiple Alignment using Fast Fourier Transform (<https://www.ebi.ac.uk/Tools/msa/mafft>) under the L-INS-i parameter. We performed maximum-likelihood phylogenetic reconstruction with PhyML implemented through the NGPhylogeny portal (12) and evaluated nodal support using the approximate Bayes branch supports.

TTV and EBV qPCR

We performed TTV (Torque teno virus) DNA load using the TTV R-GENE kit (bioMérieux, <https://www.biomerieux.com>) as described elsewhere (13) and EBV DNA load using the bioMérieux EBV R-GENE kit. We performed real-time PCR amplification on an Applied Biosystems AB7500 platform (ThermoFisher).

In Situ Hybridization

We used a ThermoFisher 2-plex ViewRNA ISH Tissue Assay Kit for in situ hybridization (ISH) to detect RNA sequences of HCirV-1 as a marker of virus replication. We designed a cocktail of 12 custom-branched DNA probes to target HCirV-1 capsid genes and 20 probes to target Rep genes, which were revealed by a red signal (probe type 1 in the

Table. Samples tested by qPCR to detect and quantify human circovirus type 1 infection in a heart-lung transplant patient in France*

Sample type	Sampling date	Ct values	Virus copies†
Serum	2017 Oct 11	Negative	Negative
Serum	2019 Aug 28	Negative	Negative
Serum	2020 Jun 12	Negative	Negative
Serum	2020 Jul 24	Negative	Negative
Plasma	2020 Sep 17	17.9	3.24×10^7
Serum	2020 Sep 23	32.4	2.00×10^3
Serum	2021 Nov 05	12.4	1.25×10^9
Serum	2021 Nov 18	13.9	4.38×10^8
Feces	2021 Nov 19	23.5	3.12×10^5
Serum	2021 Nov 30	15.1	2.07×10^8
Serum	2021 Dec 10	17.4	4.49×10^7
Serum	2022 Jan 05	16.3	9.19×10^7
BAL	2022 Jan 06	24.8	3.24×10^5
Serum	2022 Jan 12	8.4	1.72×10^{10}
Serum	2022 Jan 25	11.5	2.19×10^9
Feces	2022 Jan 27	25.5	7.98×10^4
Plasma	2022 Mar 10	15.3	1.77×10^8
Serum	2022 Mar 14	14.6	2.77×10^8
Plasma	2022 Mar 16	15.1	2.08×10^8
Serum	2022 Mar 18	14.3	3.52×10^8
Liver biopsy	2022 Mar 21	16.4	3.57×10^8
Feces	2022 Apr 01	23.7	2.68×10^5
Blood	2022 Apr 04	10.0	5.81×10^9
Serum	2022 May 05	18.1	2.73×10^7
Serum	2022 May 05	17.2	4.92×10^7
Feces	2022 May 11	28.6	1.06×10^4
Urine	2022 Jun 08	26.4	1.13×10^5
Saliva	2022 Jun 08	26.0	1.45×10^5
Plasma	2022 Jun 08	19.2	1.36×10^7
Blood	2022 Jun 08	20.3	6.36×10^6
Feces	2022 Jun 09	27.2	2.60×10^4
Blood	2022 Aug 03	18.8	1.70×10^7
Plasma	2022 Aug 03	17.5	4.01×10^7

*Virus copies are per gram for liver biopsy and feces samples, per milliliter for all other samples. BAL, bronchoalveolar lavage; qPCR, quantitative PCR; Ct, cycle threshold.

assay kit); a blue signal (probe type 6 in the kit) revealed a mix of control probes targeting human housekeeping transcripts. We applied 20-min heat pretreatment and 15-min protease digestion to tissue sections after deparaffinization using xylene to unmask RNA targets and used Harris hematoxylin for counterstaining. We used negative controls for ISH that included a no-probe negative control from the patient infected with HCirV-1, as recommended by the manufacturer, and a negative control biopsy sample from a different patient that tested negative for HCirV-1 with mNGS.

We obtained informed consent for publication of our findings from the patient. The samples we analyzed were from control patients in a retrospective noninterventional study with no additional standard care procedures. Our study was performed in accordance with ethical guidelines for medical research and approved by the institutional review boards of Necker Hospital.

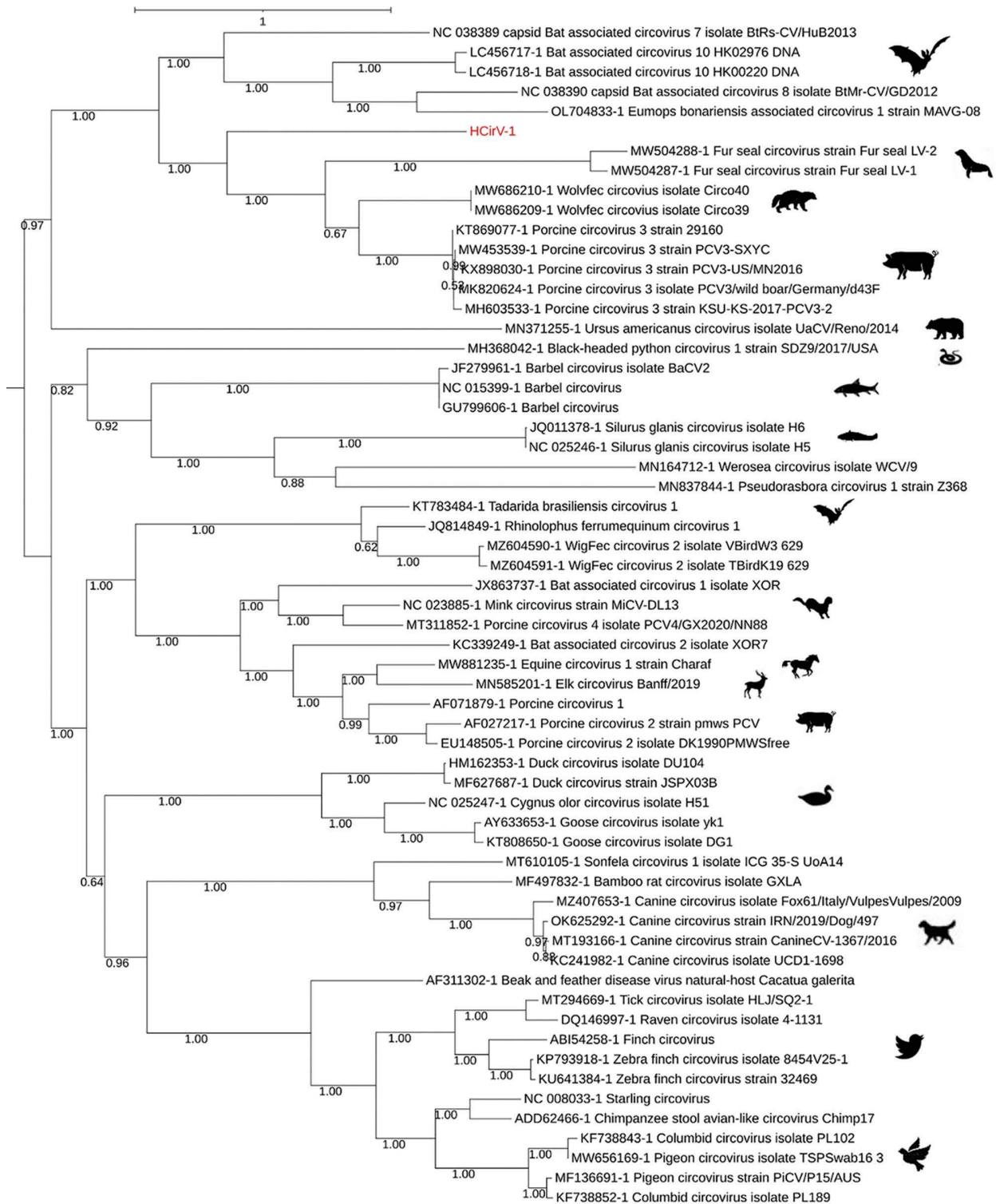


Figure 3. Phylogenetic analysis of capsid protein sequences of human circovirus type 1 (HcIrV-1) from a heart-lung transplant patient in France (red) and representative circovirus strains. Sequences were aligned with Multiple Alignment using Fast Fourier Transform (<https://www.ebi.ac.uk/Tools/msa/mafft>) under the L-INS-I parameter, and maximum-likelihood phylogenetic reconstruction was performed with PhyML implemented through the NGPhylogeny portal (<https://ngphylogeny.fr>). GenBank accession numbers for reference sequences are indicated, along with graphic representations of the animal of origin. Scale bar indicates the number of amino acid substitutions per site.

Results

Identifying Distant Circovirus Sequences by Liver Biopsy

Using NGS to analyze viral sequences from the liver biopsy specimen, we identified 1,011 sequences corresponding to the capsid and Rep genes from the *Circoviridae* family. From a plasma sample taken at the peak of viremia, we derived a consensus full-genomic sequence (2,021 nt) of the HCirV-1 virus (Genbank accession no. ON677309) and confirmed that the sequence was identical to partial sequences of the capsid gene found in the liver. We conducted phylogenetic analyses on the 2 HCirV-1 viral proteins, capsid and Rep, and on representative animal circoviruses; congruent analyses showed that HCirV-1 defines a new clade at a basal position relative to its closest viral species, Wolvfec circovirus (host: *Gulo gulo*) and porcine circovirus 3.

Kinetics of Infection

We designed an assay and conducted qPCR to assess changes in the viral load in plasma and other biologic samples taken from the case-patient since 2017. We also evaluated the viral load for TTV, a commensal virus whose level of viremia is well correlated with the immune status of graft recipients (14). TTV DNA load varied (0–3.99 log copies/mL), indicating a low level of immune suppression (15); peaks in TTV viremia and cytolysis were not temporally correlated. In contrast, HCirV-1 was first detected in patient blood samples ($10^{3.3}$ genome copies/mL) in September 2020, 2 months after steroids pulse therapy, but a July 2020 sample tested negative. However, it is noteworthy that the finding of a low level of HCirV-1 in the September 2020 sample was questionable because the tubes were opened at the same time as others harboring up to $10^{9.1}$ genome copies/mL, possibly resulting in cross-contamination. The next available sample, taken in September 2021, was strongly positive ($10^{7.5}$ genome copies/mL); thereafter, loads varied ($10^{8.3}$ – $10^{10.2}$ genome copies/mL). HCirV-1 loads were high only during the period of pronounced cytolysis.

Assessing HCirV-1 Presence in Controls

We tested 36 blood and 20 feces samples from 56 control patients using the same specific qPCR primers for assessing HCirV-1 presence, but none tested positive (Appendix Table 1). Among those patients, 19 (34%) were primitively immunocompromised, 17 (30%) immunocompetent, 13 (23%) solid organ transplant recipients, 5 (9%) hematopoietic stem cell transplant recipients or receiving

immunosuppressive treatment, and 2 (4%) HIV positive. Among 16 patients with hepatic cytolysis (Appendix Table 1), 8 cases were of unknown etiology. Control biopsies routinely screened by mNGS since 2019 did not show HCirV-1 sequences among 57 patients with hepatitis of suspected infectious origin, among whom 25 (44%) were primitively immunocompromised, 20 (35%) hematopoietic stem cell transplant recipients or receiving immunosuppressive treatment, 5 (9%) immunocompetent, 5 (9%) solid organ transplant recipients, and 2 (4%) HIV positive (Appendix Table 2).

ISH

ISH of the liver biopsy specimen showed 2% infected hepatocytes (Figure 4); red labeling was strong in the nuclei and mild in the cytoplasm. Some hepatocytes were also poorly labeled in the nuclei, and, in this case, labeling was negative in the cytoplasm. Endothelial cells and lymphocytes were not infected. No red signal was recorded in the ISH-negative controls.

Discussion

Our findings show a temporal association between hepatitis and a high load of a novel virus, HCirV-1, in the liver and blood of a 61 year-old female heart-lung transplant recipient. The patient received standard immunosuppressive therapy of tacrolimus, mycophenolate mofetil, and corticosteroids, but CMV colitis developed a year after pulsed corticosteroid therapy, suggesting a contemporary drug-induced high level of immunosuppression. Our results should be evaluated with Hill's criteria for causality (16).

We demonstrated the specificity of association using agnostic mNGS in the liver biopsy sample, showing absence of any additional virus except PCR-detected EBV. Given the low concentration of EBV in the liver biopsy, the concomitant higher replication in the blood ($10^{4.4}$ genome copies/mL), and the negative result from staining on liver biopsy, we did not consider EBV to be responsible for the cytolysis. In total, we did not detect HCirV-1 in 113 control participants: 40 tested by qPCR without hepatitis, 16 with hepatitis of known (50%) or unknown (50%) etiology (Appendix Table 1), and 57 mNGS-tested case-patients with hepatitis of unknown etiology (unpub. data, Appendix Table 2).

Circoviruses are DNA viruses. Of note, transcripts were detected in the nucleus of 2% of liver cells by HCirV-1-specific ISH, therefore consistent with viral expression and probably with virus replication. Given the lytic cycle of circoviruses, the results strongly support the role of HCirV-1 as the

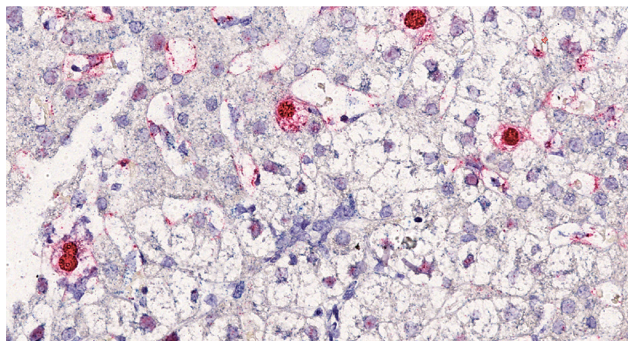


Figure 4. In situ hybridization in liver section from a heart-lung transplant patient in France. Chromogenic in situ hybridization detected human circovirus type 1 mRNA (red staining) in hepatocytes nuclei and cytoplasm. Nuclei were counterstained with Harris hematoxylin and eosin stain; original magnification $\times 40$.

causative agent of hepatitis by targeting and lysing hepatocytes and also likely as a trigger of immune cytotoxic responses.

Animal circoviruses, such as porcine circoviruses (PCV) 1–4, have been responsible for reproductive failure, dermatitis, nephropathy, and respiratory diseases in animals (17,18), manifestations not found in the case we report. However, PCV2 associated with cases of jaundice in pigs (19–21) and hepatitis in a horse have been demonstrated (22). After PCV3 is inoculated into piglets, viremia progressively increases up to $10^{8.9}$ copies/mL, reflective of findings in our study, but generally peaks around 21 days after infection, contrary to the more delayed peak we identified. Pathological lesions and PCV3 antigens have been detected in the liver, including white-gray nodules and necrosis with PCV3 antigen detected in liver lobular stroma, the hepatic sinus antral wall, and the cytoplasm in the sinus (23) and in other organs (lung, heart, kidney, lymph nodes, spleen, and small intestine) (24). PCV3 viremia could still be observed 140 days after infection (25), which is indicative of the ability of circoviruses to sustain long-term infection, as described in our patient. PCV3 was also able to infect nonhuman primate recipients of porcine cardiac xenografts (26,27).

Supporting the plausibility of HCirV-1 as disease agent, we found no temporal link for drug etiology or pathological evidence as alternative explanations. We ruled out common hepatitis virus infections, including hepatitis E virus. EBV test results were positive, but we ruled out a pathogenic role because EBV viral load remained stable around 10^4 genome copies/mL (Figure 1) and Epstein-Barr encoding region (EBER) staining was negative.

We first detected HCirV-1 before the liver pathology testing. Viral loads peaked 3 times, before

the first and second cytolysis peaks and a third time concurrently with cytolysis peak (Figure 1). The high viral load of HCirV-1 (10^{10} genome copies/mL) also favors interpretation of its causative role in hepatitis. Indeed, a strong relationship between high virus load and disease severity has been described for hepatitis A (23) and B (28) viruses and adenovirus in hematopoietic stem cell transplant recipients (29), further illustrating that viral load often influences organ pathology (30).

As of November 2022, we had not identified the source of infection for this patient; specifically, we were unable to determine whether it was because of spillover of an animal virus or if it might be a previously unknown human virus. Although phylogenetic analysis placed this virus closest to known viruses hosted by a wild exotic animal that was not a possible source for our patient, we speculate that that HCirV-1 could be of animal origin, possibly from a dietary source in a similar manner as hepatitis E virus. Developing a specific antibody test will be instrumental in deciphering the ecology of the virus and better understanding its mechanisms of transmission. HCirV-1 was shed in saliva, urine, and feces (Table); whether the virus can possibly be transmitted between persons remains to be investigated.

In conclusion, detection of HCirV-1 expression in liver cells by ISH, the exclusion of other hepatitis etiologies, and the temporal association of severe cytolysis with very high HCirV-1 loads in the liver and blood strongly suggest a causal link. Furthermore, our experience using mNGS in immunocompromised patients, including transplant recipients, as well as the specific qPCR we performed on a representative cohort of these patients prove that, unlike TTV, HCirV-1 is not part of the commensal viral flora. Agnostic NGS was key to elucidate the etiology of this hepatitis case. HCirV-1 could now be tested for in hepatitis cases of unknown origin in solid organ transplant recipients.

Acknowledgments

We thank the Biomics platform of Institut Pasteur for providing access to the sequencers. We thank the histology embryology cytogenetic department of Necker Hospital for providing access to the in situ hybridization platform.

This work was supported by Institut Pasteur and the Necker-Enfants malades University Hospital.

About the Author

Dr. Pérot works at Institut Pasteur, Virology Department, Paris. His research focuses on pathogen discovery, clinical metagenomics, molecular virology, and bioinformatics.

References

1. Lease ED, Budev MM. Infectious complications in lung transplant recipients. *Thorac Surg Clin*. 2022;32:211–20. <https://doi.org/10.1016/j.thorsurg.2021.12.002>
2. Kunnathu Puthanveedu NDK, Lum J. Central nervous system infections after solid organ transplantation. *Curr Opin Infect Dis*. 2021;34:207–16. <https://doi.org/10.1097/QCO.0000000000000722>
3. Ivana T, Robert P, Pavel S, Lenka T, Irena K. Cytomegalovirus and other herpesviruses after hematopoietic cell and solid organ transplantation: from antiviral drugs to virus-specific T cells. *Transpl Immunol*. 2022;71:101539. <https://doi.org/10.1016/j.trim.2022.101539>
4. Hirschi S, Biondini D, Ohana M, Solis M, D'Urso A, Rosner V, et al. Herpes simplex virus 2 hepatitis in a lung transplant recipient: a diagnostic challenge. *Transpl Infect Dis*. 2015;17:904–8. <https://doi.org/10.1111/tid.12459>
5. World Health Organization. Acute hepatitis of unknown aetiology – the United Kingdom of Great Britain and Northern Ireland [cited 2022 Dec 15]. <https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON368>
6. Mrzljak A, Novak R, Pandak N, Tabain I, Franusic L, Barbic L, et al. Emerging and neglected zoonoses in transplant population. *World J Transplant*. 2020;10:47–63. <https://doi.org/10.5500/wjt.v10.i3.47>
7. Soto RA, McDonald E, Annambhotla P, Velez JO, Laven J, Panella AJ, et al. West Nile virus transmission by solid organ transplantation and considerations for organ donor screening practices, United States. *Emerg Infect Dis*. 2022;28:403–6. <https://doi.org/10.3201/eid2802.211697>
8. Casto AM, Fredricks DN, Hill JA. Diagnosis of infectious diseases in immunocompromised hosts using metagenomic next generation sequencing-based diagnostics. *Blood Rev*. 2022;53:100906. <https://doi.org/10.1016/j.blre.2021.100906>
9. Regnault B, Bigot T, Ma L, Pérot P, Temmam S, Eloit M. Deep impact of random amplification and library construction methods on viral metagenomics results. *Viruses*. 2021;13:253. <https://doi.org/10.3390/v13020253>
10. Pérot P, Bigot T, Temmam S, Regnault B, Eloit M. Microseek: a protein-based metagenomic pipeline for virus diagnostic and discovery. *Viruses*. 2022;14:1990. <https://doi.org/10.3390/v14091990>
11. Bigot T, Temmam S, Pérot P, Eloit M. RVDB-prot, a reference viral protein database and its HMM profiles. *F1000 Res*. 2019;8:530. <https://doi.org/10.12688/f1000research.18776.1>
12. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res*. 2019;47(W1):W260–5. <https://doi.org/10.1093/nar/gkz303>
13. Kulifaj D, Durgueil-Larivière B, Meynier F, Munteanu E, Pichon N, Dubé M, et al. Development of a standardized real time PCR for Torque teno viruses (TTV) viral load detection and quantification: A new tool for immune monitoring. *J Clin Virol*. 2018;105:118–27. <https://doi.org/10.1016/j.jcv.2018.06.010>
14. Jaksch P, Görzer I, Puchhammer-Stöckl E, Bond G. Integrated immunologic monitoring in solid organ transplantation: the road toward Torque teno virus-guided immunosuppression. *Transplantation*. 2022;106:1940–51. <https://doi.org/10.1097/TP.0000000000004153>
15. Redondo N, Navarro D, Aguado JM, Fernández-Ruiz M. Viruses, friends, and foes: The case of Torque Teno Virus and the net state of immunosuppression. *Transpl Infect Dis*. 2022;24:e13778. <https://doi.org/10.1111/tid.13778>
16. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295–300. <https://doi.org/10.1177/003591576505800503>
17. Sirisereewan C, Thanawongnuwech R, Kedkovid R. Current understanding of the pathogenesis of porcine circovirus 3. *Pathogens*. 2022;11:64. <https://doi.org/10.3390/pathogens11010064>
18. Wang D, Mai J, Yang Y, Xiao CT, Wang N. Current knowledge on epidemiology and evolution of novel porcine circovirus 4. *Vet Res (Faisalabad)*. 2022;53:38. <https://doi.org/10.1186/s13567-022-01053-w>
19. Gillespie J, Opriessnig T, Meng XJ, Pelzer K, Buechner-Maxwell V. Porcine circovirus type 2 and porcine circovirus-associated disease. *J Vet Intern Med*. 2009;23:1151–63. <https://doi.org/10.1111/j.1939-1676.2009.0389.x>
20. Staebler S, Buergi E, Litzemberger B, McCullough K, McNair I, McNeilly F, et al. Porcine circovirus as a possible cause of postweaning wasting in pigs in Switzerland. *Schweiz Arch Tierheilkd*. 2004;146:461–8, discussion 469. <https://doi.org/10.1024/0036-7281.146.10.461>
21. Harding JCS. Postweaning multisystemic wasting syndrome: epidemiology and clinical presentation. *J Swine Health Prod*. 1998;6:249–54.
22. Hui A, Altan E, Slovis N, Fletcher C, Deng X, Delwart E. Circovirus in blood of a febrile horse with hepatitis. *Viruses*. 2021;13:944. <https://doi.org/10.3390/v13050944>
23. Fujiwara K, Kojima H, Yasui S, Okitsu K, Yonemitsu Y, Omata M, et al. Hepatitis A viral load in relation to severity of the infection. *J Med Virol*. 2011;83:201–7. <https://doi.org/10.1002/jmv.21958>
24. Jiang H, Wang D, Wang J, Zhu S, She R, Ren X, et al. Induction of porcine dermatitis and nephropathy syndrome in piglets by infection with porcine circovirus type 3. *J Virol*. 2019;93:e02045–18. <https://doi.org/10.1128/JVI.02045-18>
25. Opriessnig T, Prickett JR, Madson DM, Shen HG, Juhan NM, Pogranichniy RR, et al. Porcine circovirus type 2 (PCV2)-infection and re-inoculation with homologous or heterologous strains: virological, serological, pathological and clinical effects in growing pigs. *Vet Res*. 2010;41:31. <https://doi.org/10.1051/vetres/2010003>
26. Denner J, Mankertz A. Porcine circoviruses and xenotransplantation. *Viruses*. 2017;9:83. <https://doi.org/10.3390/v9040083>
27. Krüger L, Längin M, Reichart B, Fiebig U, Kristiansen Y, Prinz C, et al. Transmission of porcine circovirus 3 (PCV3) by xenotransplantation of pig hearts into baboons. *Viruses*. 2019;11:650. <https://doi.org/10.3390/v11070650>
28. Yu SJ, Kim YJ. Hepatitis B viral load affects prognosis of hepatocellular carcinoma. *World J Gastroenterol*. 2014;20:12039–44. <https://doi.org/10.3748/wjg.v20.i34.12039>
29. Lion T, Baumgartinger R, Watzinger F, Matthes-Martin S, Suda M, Preuner S, et al. Molecular monitoring of adenovirus in peripheral blood after allogeneic bone marrow transplantation permits early diagnosis of disseminated disease. *Blood*. 2003;102:1114–20. <https://doi.org/10.1182/blood-2002-07-2152>
30. Funk GA, Gosert R, Hirsch HH. Viral dynamics in transplant patients: implications for disease. *Lancet Infect Dis*. 2007;7:460–72. [https://doi.org/10.1016/S1473-3099\(07\)70159-7](https://doi.org/10.1016/S1473-3099(07)70159-7)

Address for correspondence: Marc Eloit, Institut Pasteur, Pathogen Discovery Laboratory, 25-28 rue du Dr. Roux, 75015, Paris, France; email: marc.eloit@pasteur.fr

Partie 2 : Identifier des indications pour le diagnostic par métagénomique clinique

[Article 3: Performance of clinical metagenomics: an observational prospective study in France](#)

L'objectif de la métagénomique clinique (mNGS) est de détecter des agents pathogènes non reconnus par les méthodes microbiologiques conventionnelles afin d'améliorer les soins prodigués aux patients. Bien que l'analyse par mNGS soit reconnue comme un outil de diagnostic valable pour la découverte de pathogène, les indications les plus pertinentes pour cette stratégie novatrice restent mal définies. En particulier, l'identification des populations ou des situations cliniques qui bénéficieraient le plus de cette technologie de pointe, ainsi que les types d'échantillons optimaux pour l'analyse par métagénomique clinique, n'ont pas encore été pleinement identifiés. Dans cette étude prospective de trois ans, nous avons cherché à évaluer les déterminants de la positivité et de la performance clinique des résultats de la mNGS sur un large panel d'échantillons et d'indications.

La participation de l'étudiant à cette étude a été la suivante :

Conception et coordination de l'étude, interprétation des résultats de métagénomique, nettoyage de la base de données, analyse de l'ensemble des données dont les analyses statistiques, mise en forme de l'ensemble des résultats, écriture du manuscrit.

Performance of clinical metagenomics: an observational prospective study in France

Jacques Fourgeaud^{1,2+}, Beatrice Regnault^{3,4+}, Vichita Ok², Nicolas Da Rocha³, Émilie Sitterlé², Meryem Mekouar², Hélène Faury², Catherine Milliancourt-Seels², Florence Jagorel³, Delphine Chrétien³, Thomas Bigot⁴, Éric Troadec³, Isabelle Marques², Alexandra Serris⁵, Danielle Seilhean^{6,15}, Benedicte Neven^{7,8}, Pierre Frange^{1,2}, Agnès Ferroni², Marc Lecuit^{5,9,10}, Xavier Nassif¹¹, Olivier Lortholary^{5,12†}, Marianne Leruez-Ville^{1,2†}, Philippe Pérot^{3,13†}, Marc Eloit^{3,13,14\$}, Anne Jamet^{2,11*}

1 Université Paris Cité, FETUS, F-75015 Paris, France. 2 Microbiology department, AP-HP, Hôpital Necker, F-75015 Paris, France. 3 Institut Pasteur, Université Paris Cité, Pathogen Discovery Laboratory, 25-28 rue du Dr. Roux, F-75015, Paris, France. 4 Bioinformatics and Biostatistics Hub, Computational Biology Department, Institut Pasteur, F-75015 Paris, France. 5 Université Paris Cité, Centre d'Infectiologie Necker-Pasteur, IHU Imagine, Hôpital Necker, F-75015 Paris, France. 6 Département de Neuropathologie Raymond Escourolle, AP-HP-Sorbonne, Groupe Hospitalier Pitié-Salpêtrière, F- 75013 Paris, France. 7 Pediatric Hematology Immunology and Rheumatology Unit, AP-HP, Hôpital Necker, F-75015 Paris, France. 8 Université Paris Cité, Inserm, Institut Imagine, Laboratory of Immunogenetics of Pediatric Autoimmune Diseases, F-75015 Paris, France. 9 Institut Pasteur, Université de Paris, Inserm U1117, Biology of Infection Unit, Paris, France. 10 Institut Pasteur, National Reference Center and WHO Collaborating Center Listeria, Paris, France. 11 Université Paris Cité, CNRS, INSERM, Institut Necker-Enfants Malades, Team Pathogenesis of systemic infection, F- 75015 Paris, France. 12 Institut Pasteur, Centre National de Référence Mycoses Invasives et Antifongiques, Département de Mycologie, Labex IBEID, Paris, France. 13 Institut Pasteur, Université Paris Cité, The WOA (OIE) Collaborating Center for the detection and identification in humans of emerging animal pathogens, 25-28 rue du Dr Roux, F-75015, Paris, France. 14 École Nationale Vétérinaire d'Alfort, Université Paris-Est, F-94704 Maisons-Alfort, France. 15 Institut du Cerveau - Paris Brain Institute - ICM, Inserm U1127, CNRS UMR7225, AP-HP, Sorbonne University, Pitié-Salpêtrière Hospital, Paris, France. + contributed equally as first authors. † contributed equally. * contributed equally as last authors. \$ Part of STROBE-metagenomics working group

Summary

Background

Metagenomic next-generation sequencing (mNGS) allows untargeted identification of a broad range of pathogens, including rare or novel microorganisms. We aimed to assess the determinants of positivity and clinical utility of mNGS.

Methods

At the Necker-Enfants malades Hospital (Paris, France), we prospectively performed short-read shotgun metagenomics analysis as a second-line technique (in case of negativity of first-line techniques or when the symptoms were not fully explained by initial positive results) or as a first-line test in life-threatening situations requiring urgent non-targeted pathogen identification. We describe variables associated with the detection of “causative” (i.e., known to cause similar clinical presentations) or “possibly causative” (i.e., known to be associated with similar clinical presentations without formal evidence of causality) pathogens by mNGS.

Findings

Between Oct 29, 2019, and Nov 7, 2022, we analyzed 742 specimens collected from 523 patients. The initial suspicion of infection was either high (n=470, 63%) or low (n=272, 37%). Causative/possibly causative pathogens were detected in 117 (25%) samples from patients with high initial suspicion of infection, versus 9 (3%) samples analyzed to rule infection out (OR=9.1, p<0.0001). We showed that mNGS had higher odds of detecting a causative/possibly causative pathogenic virus on cerebral biopsies compared to CSF (OR=4.1, p=0.0025) and in samples from immunodeficient compared to immunocompetent individuals (OR=2.4, p=0.0013). The highest positivity rate was found in stool samples, which detected causative/possibly causative pathogens in 46% (24/52) of cases with digestive (12/24), hepatic (7/24), hematological (3/24), or neurological (2/24) indications.

Concordance with conventional confirmatory tests results was 97% (103/106), when mNGS detected causative/possibly causative pathogens.

Interpretation

The clinical performance of mNGS is excellent compared to conventional microbiology and major determinants of positivity are immune status and initial level of suspicion of infection.

Funding

This work was supported by the Necker Hospital and Institut Pasteur.

Research in context

Evidence before this study

The aim of clinical metagenomics (mNGS) is to detect pathogens not-recognized by conventional microbiological methods to improve patient care. Despite mNGS recognition as a valuable diagnostic tool for infections, the most relevant indications for this innovative strategy remain poorly defined. In particular, the identification of the populations or clinical situations that would benefit most from this cutting-edge technology, as well as the optimal sample types for mNGS analysis, have not yet been fully identified.

We searched PubMed for studies published between January 1, 2015, and February 6, 2023, using the terms "metagenomics" AND "clinical" AND "hospital" AND "diagnostic" without language restriction and filtering for reviews and case reports. We retrieved 853 articles, from which 47 reported on the prospective use of mNGS for the diagnosis of infections. The most frequently studied medical conditions were lower respiratory tract infections (12/47), central nervous system disorders (11/47) and sepsis (7/47). We identified only 6 prospective studies (conducted in China, France, and Switzerland), which used multiple clinical sample types and clinical conditions. Half of them involved only adults (including a previous study from our group in immunocompromised adults), while the other half involved only pediatric patients. The number of patients studied ranged from 10 to 168. One of these studies found no viruses and another used a pipeline limiting the search to viruses only. Of the six studies, three excluded analysis of tissue samples.

Although the results of these studies suggest the clinical relevance of mNGS in complementing current routine tests, the limited size and heterogeneity of populations included, and the limited number of sample types tested do not allow to draw definitive conclusions.

Added value of this study

To our knowledge, this is the first national

prospective study of the clinical application of mNGS to be carried out without any restrictions on sample types, medical indications or the patient population studied. In this study, 742 samples collected from 523 patients were analyzed. The inclusion of both immunocompetent and immunocompromised patients in the same study allowed to reveal that immunocompromised status was associated with higher odds of detecting a causative pathogen by mNGS than immunocompetent status (OR=2.4, p=0.0013). Our results also highlight the high potential for discovery of novel viruses or medical entities in the immunocompromised patient population as we have demonstrated in the following cases: the role of *Aichi virus* in multi-organ involvement and a human pathogenic *circovirus* which we have described in detail in published case reports.

By not limiting the type of sample analyzed to sterile sites, and by guiding the physician to send stool, beyond the sole investigation of diarrhea symptoms, to also investigate liver or neurological disorders, we showed that mNGS on stool was more likely to detect a causative pathogen than mNGS on sterile fluids (OR=5.7, p=0.0013).

Implications of all the available evidence

In combination with previous evidence, our results highlight the strength of the non-targeted mNGS approach, particularly in the immunocompromised population. We also confirm the high positivity rate of CNS tissue analysis and reveal the contribution of stool sample analysis in etiological diagnosis of hepatic or neurological conditions. In this context of broad development and availability of mNGS techniques in high-income countries' microbiology laboratories, awareness of the best target populations and samples for high mNGS performance should assist clinicians to appropriately prescribe this innovative test.

However, further research is needed to assess the position of mNGS, whether routinely performed by the local polyvalent microbiology laboratories or subsequently by distant reference centers.

INTRODUCTION

Pathogen identification at the time of international travel and emerging microorganism outbreaks,

including rapidly evolving viruses (coronaviruses, arboviruses, *influenza* or *Ebola virus*), requires broad-spectrum diagnostic tools readily available to clinicians. In addition, immunocompromised

patients can potentially be infected with a wide range of opportunistic pathogens or atypical organisms.¹⁻³ Although cultures and universal PCR allow for broad-spectrum analysis of bacterial and fungal microorganisms, conventional tests in virology consist of targeted pathogen-specific nucleic acids, antigens, or antibodies detection, which limits the list of detectable viruses to the most common and/or foreseen pathogens. The diagnosis of parasitic diseases is also based on specific molecular or serological tests, in addition to morphological analysis by microscopy.

Microorganisms detectable by direct high-throughput nucleic acid sequencing of clinical samples (clinical metagenomics or mNGS) include viruses (DNA or RNA), bacteria, fungi and parasites. Thus, mNGS on patient fluids and tissues has enabled diagnosis that could not be achieved with conventional laboratory tests.^{4,5} In case of detection of novel or unexpected microorganisms, mNGS can be combined with light or electron microscopy, organism-specific histological immunostaining, and/or in situ hybridization to provide evidence of causality. Noteworthy, designing specific probes based on the nucleotide sequences of the organism detected by mNGS is of great value.⁶

On Oct 29, 2019, the clinical microbiology laboratory at Necker-Enfants malades Hospital and the Institut Pasteur (both based in Paris, France) established a mNGS collaborative diagnostic platform for infection diagnosis. Although the diagnostic value of mNGS compared retrospectively to positive conventional tests is well described in the literature,^{7,8} only a limited number of prospective studies covered multiple clinical sample types and clinical conditions (supplemental table S1 in appendix p8).^{2,9-13} Thus, mNGS positioning in the arsenal of diagnostic tools still warrant clarification. This study aimed to evaluate, following the STROBE-metagenomics guidelines,¹⁴ the determinants of positivity and clinical utility of mNGS findings on a large panel of samples and indications in a tertiary care center.

METHODS

Patients

The samples for mNGS analysis were received from October 29, 2019 at the French Necker-Enfants malades Hospital belonging to the Greater Paris University Hospitals (AP-HP). Our inclusion

criteria for mNGS indication were either negativity of first-line standard microbiology techniques (SMT), complex medical condition (e.g., immunodepression) in which the symptoms may not be fully explained by initial SMT positive results, or life-threatening situations without specific symptoms which required urgent non-targeted pathogen identification. First-line SMT required before mNGS analysis are described in supplemental table S2 (appendix p9). In severe situations with organ biopsy collection, mNGS was used in parallel with SMT without waiting for SMT results. Based on the expertise gained from our findings, we suggested to clinicians that central nervous system (CNS) biopsies be preferred over cerebrospinal fluid (CSF) when available, and recommended stool or urine analysis to investigate hepatic, neurological, dysimmune, or nephrological disorders (prolonged viral shedding observed in immunocompromised patients).

Samples were accompanied by a mandatory prescription form completed by the senior clinician or pathologist. The clinical level of suspected infection was indicated directly on the prescription form by the senior clinicians. Additional information requested included current antimicrobial and immunosuppressive therapies, immune status, microbiological test results and histological findings. On the basis of this information, categorization as high or low level of suspected infection was carried out independently by two microbiologists from the Necker laboratory.

We categorized the identified microorganisms into two groups: "pathogens" and "microorganisms without known pathogenic role". Pathogens were further sub-divided into causative, possibly causative, or unlikely to be responsible for the symptoms (Figure 1).

In case of microorganism detection, causality was established based on the consistency of symptoms with the literature. The terms "causative" and "possibly causative" are described and illustrated in supplementary materials (Figure S1 in appendix p5). The confirmation requirements for each pathogen are detailed in supplementary table S3 (appendix pp 10-12).

The study was a prospective observational study with no addition to standard care procedure. Patients and/or their legal guardians were informed by a written letter at the time of collection, and were given the opportunity to object to the use of the data at any time. The study was carried out in

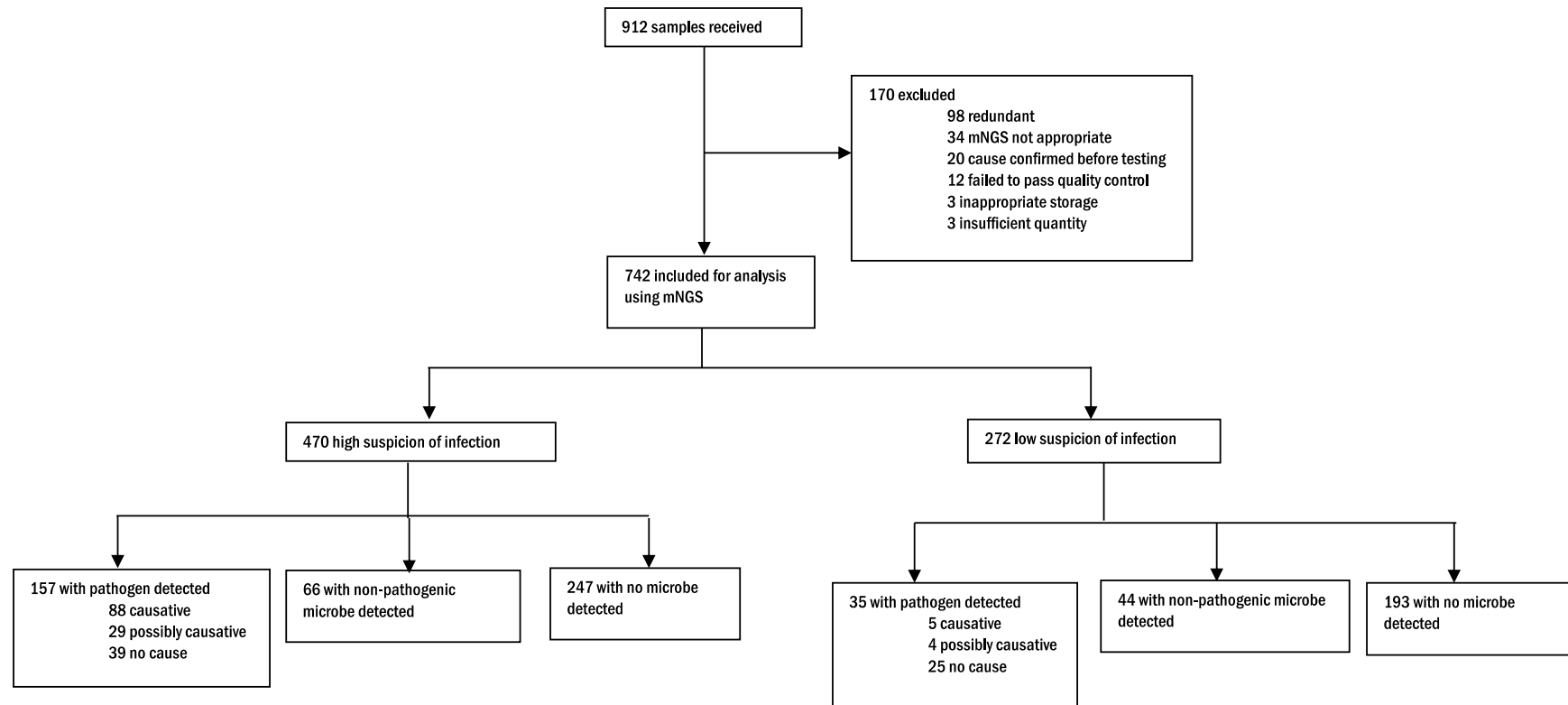


Figure 1. Flowchart of samples analyzed by mNGS and overall findings.

Out of 912 received samples, 742 were selected for analysis in a routine mNGS pipeline. The detected microorganisms were first classified according to their known pathogenicity (known pathogenic role or no known pathogenic role). The pathogenic microorganisms were then classified according to their causality in the symptoms of the studied patients. mNGS, metagenomic next-generation sequencing; SMT, Standard microbiology techniques including culture of non-fastidious bacteria, broad range PCR, routine syndromic PCR panel and specific viral PCR depending on the symptoms.

accordance with the ethical guidelines for medical research and was approved by “The AP-HP.Centre Institutional Research Review Board” (IRB #00011928, April 9th, 2022).

Sample Processing

Samples were transported to the microbiology laboratory of the Necker-Enfants malades Hospital (Paris, France) on dry ice and kept frozen at -80°C before testing. Nucleic extraction protocols, quality controls and kit used for library preparation are described in supplementary materials (Supplementary figure S2 in appendix p6, supplementary table S4 in appendix pp 13-14). Library sequencing was performed on a NextSeq500 instrument (Illumina, Evry, France) with an average of 50 million reads (1x150 bp) per sample. Testing was done in batches, on a weekly basis. Bioinformatic analysis pipeline is described in the supplementary materials (Supplementary figure S3 in appendix p7).

We excluded samples that were redundant (as in the case of several concomitant prescriptions for the same patient) or inappropriate (as in the case of a very strong suspicion of a specific pathogen, we suggested a specific PCR analysis instead) (Figure 1). We also excluded samples if a cause was identified between the time of prescription and sample processing, in cases of inappropriate storage, insufficient quantity or quality control failure (Figure 1).

Outcomes

Our study showed that the greatest clinical benefits of mNGS were seen in immunocompromised patients, in cases where the primary diagnostic hypothesis was infection, and in subjects who had undergone brain biopsies and/or stool sampling.

Statistical Analysis

Categorical variables are reported as numbers and percentage. Multivariate logistic regression was used to compute odds ratio that gender, age, immune status, initial suspicion of infection, indication, and sample type has on the causative pathogen detection (for the model used see supplementary materials and supplementary table S5 in appendix p15); only p values <0.05 were considered as significant. All analyses were conducted using GraphPad Prism 9 software (GraphPad Software, San Diego, California).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

RESULTS

Between Oct 29, 2019, and Nov 7, 2022, we received 912 samples with a prescription to perform mNGS analysis (Figure 1). We excluded 170 samples, mainly because of redundancy: when multiple specimens were collected from the same patient or from the same tissue only certain samples were analyzed according to the clinical situation (n=98, 58%). The excluded samples were blood (n=43, 25%), CSF (n=27, 16%), other biological fluids or stools (n=57, 34%), and tissue samples (n=43, 25%). Eighty-seven percent (643/742) of the analyzed samples were obtained from health facilities located in the Paris area (breakdown of sample provenance is given in the appendix p1). The median time for mNGS analysis reporting was nine days from sample extraction to data analysis (IQR=5-12). It is worth noting that in 9 emergency cases, we were able to reduce the turnaround time to 3 days. The 742 analyzed samples were collected from 523 patients for whom no microbiological etiology of their symptomatology had been found by conventional techniques at the time of mNGS prescription (Table 1 and supplementary table S2 appendix p9). For 129 patients, we analyzed several samples from different dates and/or types (range 2 to 8). The median age was 18 years (range, 0 to 89, IQR, 9 to 48) (Table 1). In 59% (306/523) of cases, infection was the main suspected diagnostic and mNGS was prescribed to identify causative pathogens, whereas in the remaining cases, the likelihood of infection was lower and mNGS was prescribed to rule out a possible infection (e.g., before initiating immunosuppressive therapy or stopping anti-infective therapies). The number of samples received per patient was higher in the former cases (high probability of infection) than in the latter (1.5 [95% CI, 1.4 to 1.6] versus 1.3 samples/patient [95% CI, 1.2 to 1.4], respectively (p=0.0401)). The characteristics of the analyzed samples are summarized in Figure 2. All sample types were acceptable, with liver and CNS biopsies being the primary sample type analyzed among tissue samples, and urines and bone marrow aspirates being the primary sample types analyzed among

Table 1. Demographic and baseline characteristics of 523 patients.

Characteristic	Total (n=523)
Age, median (IQR), years	18 (9-48)
Children <15 years old, n (%)	228 (44)
Adults >15 years old, n (%)	295 (56)
Female, n (%)	230 (44)
Male, n (%)	293 (56)
Disorder types, n (%)	
Neurological	204 (39)
Hepatic	64 (12)
Nephrological	42 (8)
Gastro intestinal	37 (7)
Osteo-muscular	39 (7)
Respiratory	34 (7)
Inflammatory	34 (7)
Cutaneous	18 (3)
Otorhinolaryngological - ocular	17 (3)
Hematologic	17 (3)
Fetal	9 (2)
Cardiac	8 (2)
Immune status, n (%)	
Immunocompetent	277 (53)
Primary immune deficiency	101 (19)
Immunosuppressive or corticosteroid therapy	57 (11)
Solid organ transplant	35 (7)
Hematopoietic stem cell transplant	31 (6)
Patients living with HIV	22 (4)
Microorganism detection, n (%)	
Causative or possibly causative pathogen	98 (19)
Negative or not causative	221 (42)

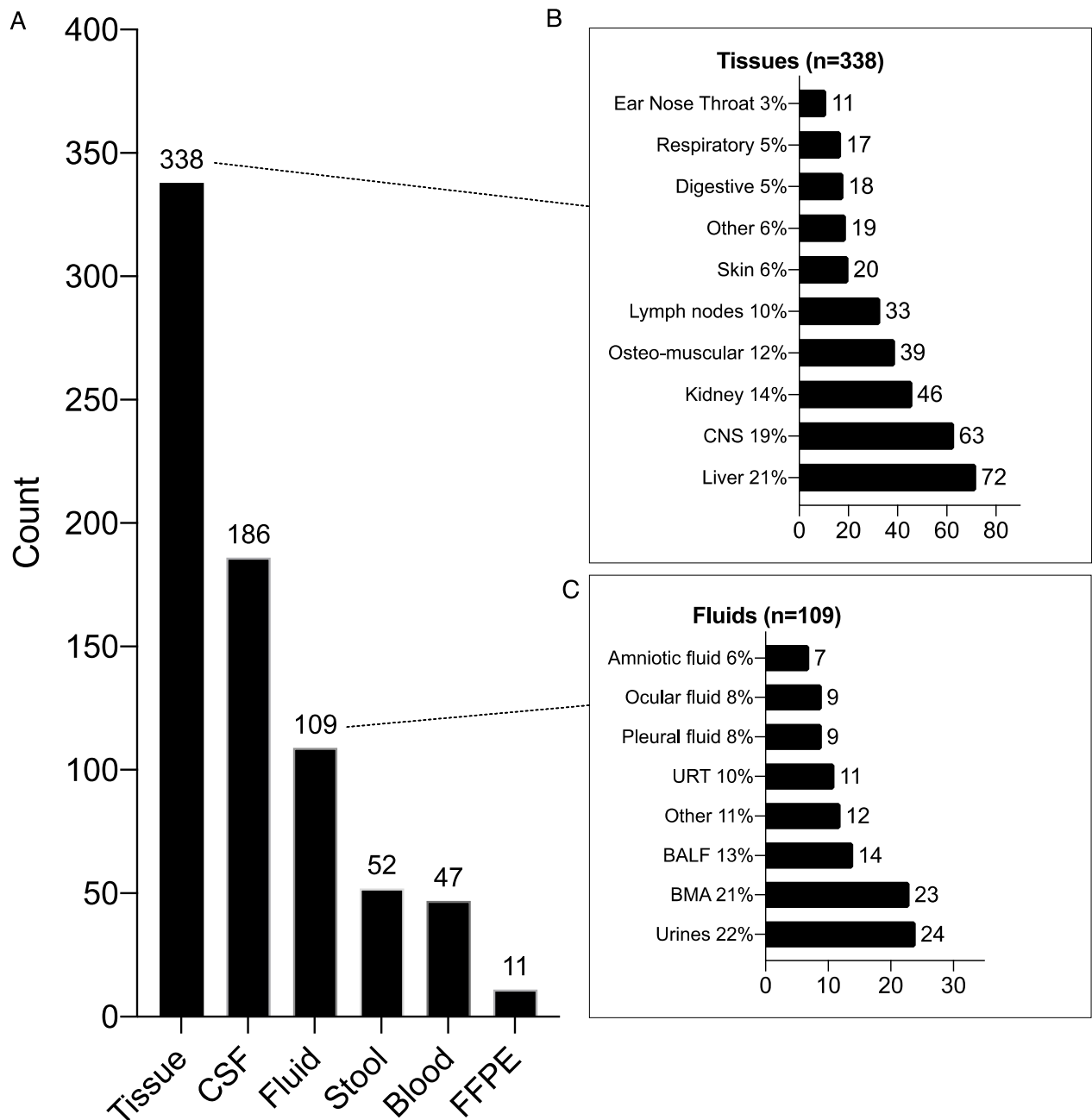


Figure 2. Characteristics of samples analyzed by mNGS.

A) Distribution of sample types among the 742 samples analyzed by mNGS in the present study. The number at the end of each bar indicates the number of samples. CSF, Cerebrospinal Fluid; FFPE, Formalin-fixed Paraffin-Embedded.

B) Distribution of the sample origins in the tissue category illustrating the predominance of brain and liver biopsies. CNS, central nervous system.

C) Distribution of the sample origins in the fluid category illustrating the predominance of bone marrow aspiration samples and urines. BMA, Bone marrow aspiration; URT, Upper Respiratory Tract; BALF, Bronchoalveolar Lavage Fluid.

body fluids (Figure 2).

Overall, at least one microorganism (i.e., viruses, bacteria, or fungi independent of their involvement in the symptoms) has been detected in 41% (301/742) of mNGS samples and 63% (190/301) of the detected microorganisms belonged to the category of pathogens (interpreted as causative or not). Finally, mNGS allowed etiological diagnosis by identification of a causative pathogen in 19% (98/523) of the patients.

We classified the detected microorganisms according to sample types and clinical/biological signs. The highest proportions of pathogens were detected in stool and fluids (Figure 3A). Indeed, 46% (24/52) of stool samples allowed detection of causative/possibly causative pathogens (see supplementary Table S6 in appendix pp 16-17), whereas the proportion of pathogens detected in CSF was only 7% (13/185). The highest proportions of samples in which pathogens were identified were collected from patients with gastrointestinal, hematologic, respiratory, or hepatic symptoms (Figure 3B). The percentage of positivity for different sample types and clinical indications is indicated in supplemental table S6 (appendix pp 16-17).

Identified microorganisms classified as causative/possibly causative pathogens were viruses (95/111, 85%), bacteria (12/111, 11%), and fungi (4/111, 4%) (Figure 3C) in cases with only one pathogen was detected (111/126). Viruses were mainly single stranded RNA viruses (52/95, 55%) and double stranded DNA viruses (31/95, 33%) (Figure 3C). The most frequently detected microorganisms were human herpesviruses, noroviruses and enteroviruses, streptococci and *Staphylococcus aureus* (Figure 3D), *Candida parapsilosis* and *Aspergillus fumigatus*, respectively. The complete list of causative/possibly causative microorganisms identified is provided in Table S7 (appendix pp 18-19). Causative/possibly causative pathogens were detected alone or associated with another pathogen (causative or not) in 73% (92/126) and 27% (34/126) of the samples, respectively. The most common microorganisms detected in case of co-detection of pathogens were *Cytomegalovirus* (n=6), noroviruses (n=4) and adenoviruses (n=4) (Table S7 in appendix pp 18-19).

To assess the variables (gender, age, immune status, initial suspected diagnostic of infection, indication of mNGS, and sample type) associated

with the detection of causative/possibly causative pathogens by mNGS, we calculated odds ratios by multivariate logistic regression. Compared to CSF, biopsies of the CNS were associated to a significantly higher rate of detection of causative/possibly causative pathogens (14/63, 22% versus 13/185, 7%, OR=4.1, [95% CI, 1.7 to 11], p=0.0025) (Figure 4). For 15 patients, we received both CNS (n=16) and CSF (n=21) samples, of which 6 and 2 were positive, respectively. In 2 cases, both CSF and CNS were positive, whereas in 4 cases only CNS was positive. *Human herpes simplex virus* (HSV), *human herpesvirus 6* (HHV6), *cytomegalovirus*, *Aggregatibacter aphrophilus* and *Enterococcus faecium* detected by mNGS in CSF were concomitantly detected in CSF by mNGS and routine standard PCR techniques. In contrast, mNGS could identify *human pegivirus* (HPgV) in two CSF samples in which standard PCR techniques (not targeting HPgV) were negative.¹⁵ No category of clinical/biological signs underlying mNGS prescriptions were associated with a higher rate of detection of causative/possibly causative pathogens (rates ranging from 0/12, 0% to 9/25, 36%) (Figure 3B; Figure 4). Conversely, the initial high probability of infection was found to be a major determinant of the positivity rate of mNGS: causative/possibly causative pathogens were detected in 25% (117/470) of samples from patients with initial suspected diagnostic of infection, versus 3% (9/272) of samples analyzed to rule out the presence of a pathogen (OR=9.1, [95% CI, 4.6 to 5.5], p<0.0001). Compared to immunocompetent individuals, causative/possibly causative pathogens were identified more frequently in samples from immunodeficient patients (25% (95/386) versus 9% (31/356); OR=2.4, [95% CI, 1.4 to 4.1], p=0.0013). Of note, the nature of the analyzed samples and the proportion of patients with a high suspicion of infection were similar between immunocompetent and immunodeficient subjects. Among immunodeficient patients, the highest rate of detection of causative/possibly causative pathogens was found in those with primary immune deficiencies (OR=2.0, [95% CI, 1.1 to 3.8], p=0.0292) (Figure 4).

Confirmatory tests include existing conventional laboratory techniques that were not performed in the first instance due to lack of initial suspicion (e.g., detection of *Epstein-Barr virus*, *cytomegalovirus*, enteroviruses) or dedicated tests developed based

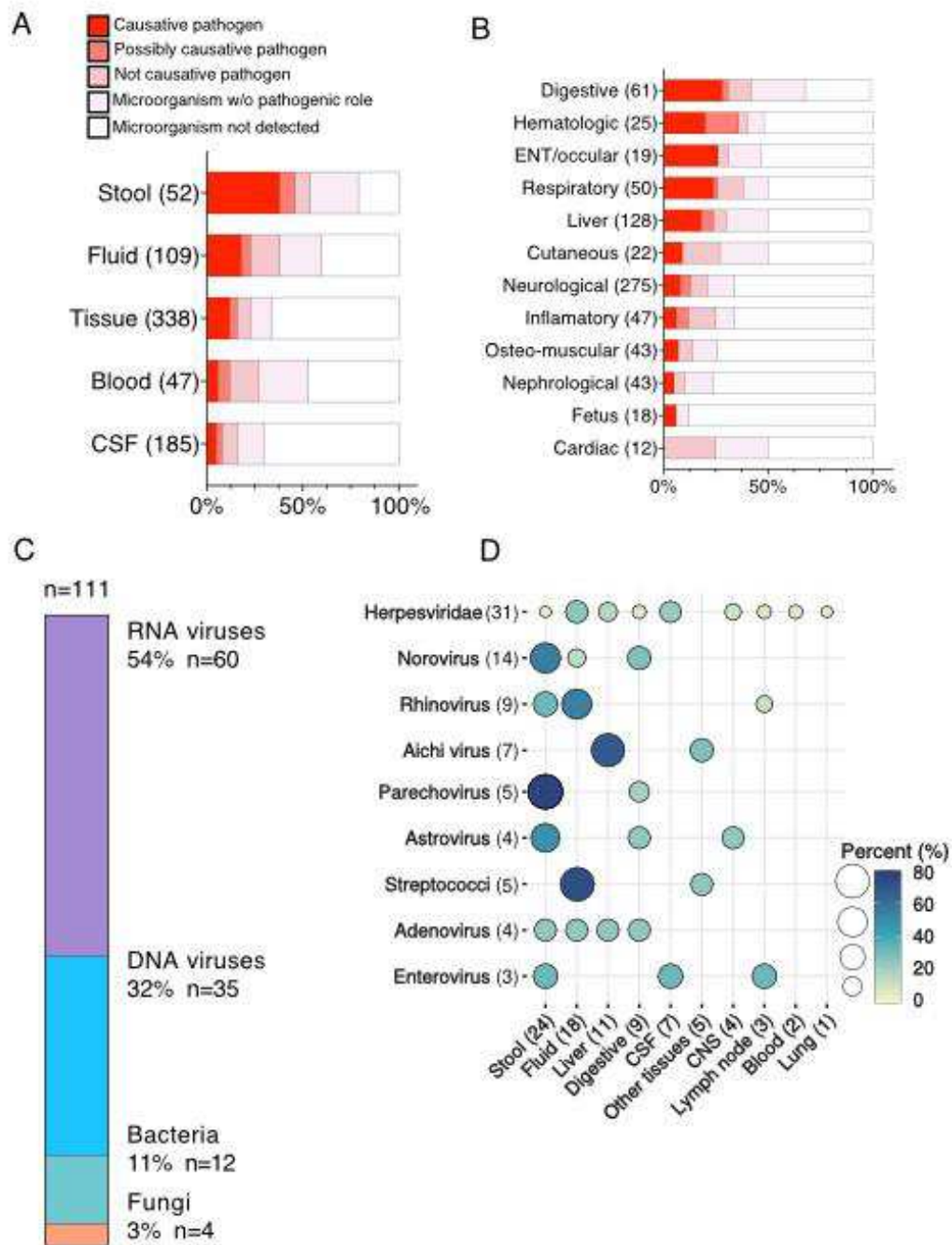


Figure 3. Classification of detected microorganisms.

A) Classification of detected microorganisms according to sample types.

B) Classification of detected microorganisms according to clinical/biological signs leading to mNGS prescription.

C) Classification of causative/possibly causative pathogens detected in mono-infection.

D) The most prevalent families of causative/possibly causative pathogens identified and their distribution among the different sample types (the total for each row adding up to 100%).

The numbers between brackets indicate the number of samples in A, B and D.

on mNGS findings or literature (*Aichi virus*, HPgV, *Astrovirus MLB1* or VA1). It was possible to perform specific confirmatory tests in 84% (106/126) of the cases in which causative/possibly causative pathogens were identified by mNGS. Concordance with specific confirmatory tests was 100% (60/60), 94% (29/31), and 91% (10/11) for RNA viruses, DNA viruses and bacteria, respectively. The confirmation requirements for each pathogen are detailed in the supplementary Table S3 (appendix pp 10-12).

Besides, in 125 instances where the possibility of infection was deemed particularly high and the mNGS analysis did not identify any causative/possibly causative pathogens, additional investigations were conducted using specific PCR techniques (searching for *Epstein-Barr virus*, *cytomegalovirus*, HSV-1, HHV-6, *hepatitis B virus*, polyomaviruses, HPgV, *rhinovirus*, HIV-1, SARS-CoV-2, enteroviruses, *Aichi virus*, *Tropheryma whipplei*, *Mycobacterium spp.*, 16S, *Candida spp.*, 23S), serology (for *cytomegalovirus*, HSV-1) or prolonged culture (for mycobacteria). Altogether, among 231 samples investigated by both mNGS and subsequent specific tests, discordant results were found in 69 samples, with 58 being mNGS positive/specific tests negative and 11 being mNGS negative/specific tests positive. The 11 cases with lack of findings with mNGS and positive results of conventional tests are described in supplementary Table S8. Among the 58 samples with results considered as “false positive” (specific tests negative), the number of reads detected with mNGS for the suspected pathogen was below 5 in 97% (56/58) of cases, including contamination in the same sequencing run by a high viral load of *Norovirus GII* (n=3) and unconfirmed SARS-CoV-2 sequences (n=11). The two remaining cases (2/58) included 23 reads of *Epstein-Barr virus* and 15 reads of *BK virus* detected in CSF and liver biopsy sample, respectively.

Notable cases where direct genotyping was provided to physicians are described below. Metagenomic NGS performed in amniotic fluid of a pregnant woman referred for fetal brain abnormalities allowed us to detect three *Zika virus* (ZIKV) specific sequences (2 from the *envelope* gene and 1 from the NS1 region) of the Asian genetic lineage sharing 96.1% nucleotide sequence identity with ZIKV French Polynesia/HPF/2013/KJ776791, which is consistent with the patient’s travel in Thailand.¹⁶ We

also detected 40 sequences of *Dengue virus type 1* in the urine sample of a 7-year-old hematopoietic stem cell transplant recipient from French Guiana who presented with neurological symptoms. Vaccine *Rubella* strain was identified in a skin granuloma in a 20 years-old patient with primary immunodeficiency, a previously reported but rare disease.³ We also typed 29 gastroenteritis viruses (*Human Adenovirus 1*, *Aichi virus genotype B*, *Human Astrovirus 1*, *Astrovirus MLB1* or VA1, *Norovirus GI.6* or *GII.4*, *Sapovirus G.I* or *G.V*), and 14 respiratory viruses (*human respiratory syncytial virus A*, *Rhinovirus A, B* or *C*, and *Enterovirus D68* or *Coxsackievirus A4* or *B5*). Finally, our pipeline allowed us to identify a lethal case of encephalitis with *European Bat Lyssavirus 1* in cerebral tissue⁴ and a novel human hepatitis virus (a *circovirus*) in a liver sample (HCirV-1).⁶

Regarding bacterial pathogens, we were able to genotype the *S. aureus* strain present in respiratory fluid from a patient with cystic fibrosis who died from acute pneumonia and showed that the strain belonged to complex clonal 30 (CC30) and harbored virulence-associated superantigenic toxin genes (*sea* and *tst-1*).¹⁷ In a case of chronic diarrhea in an immunocompromised patient, we were able to infer the serotype of the *Salmonella* strain present in the patient’s stool (i.e., serotype Typhimurium).

DISCUSSION

In this study, we report the prospective use of mNGS during the first three years of its implementation in our center.

The addition of mNGS to our diagnostic workflow resulted in an etiologic diagnosis in 25% (117/470) of samples when the main diagnostic hypothesis at the time of sampling was infection, and did not detect any pathogen in 97% (263/272) of samples for which mNGS was prescribed to rule out infection. Overall, results of this study show the potential benefit of adding mNGS to routine diagnosis workflows both for detection and discovery of rare or novel pathogens.^{4,6,16,18} We identified major determinants of clinical benefits especially in immunocompromised patients and in subjects with available brain biopsies and/or stool samples. Finally, mNGS seems to have not only the highest sensitivity in cases where the primary diagnostic hypothesis is an infection, but also a good negative predictive value in subjects in whom

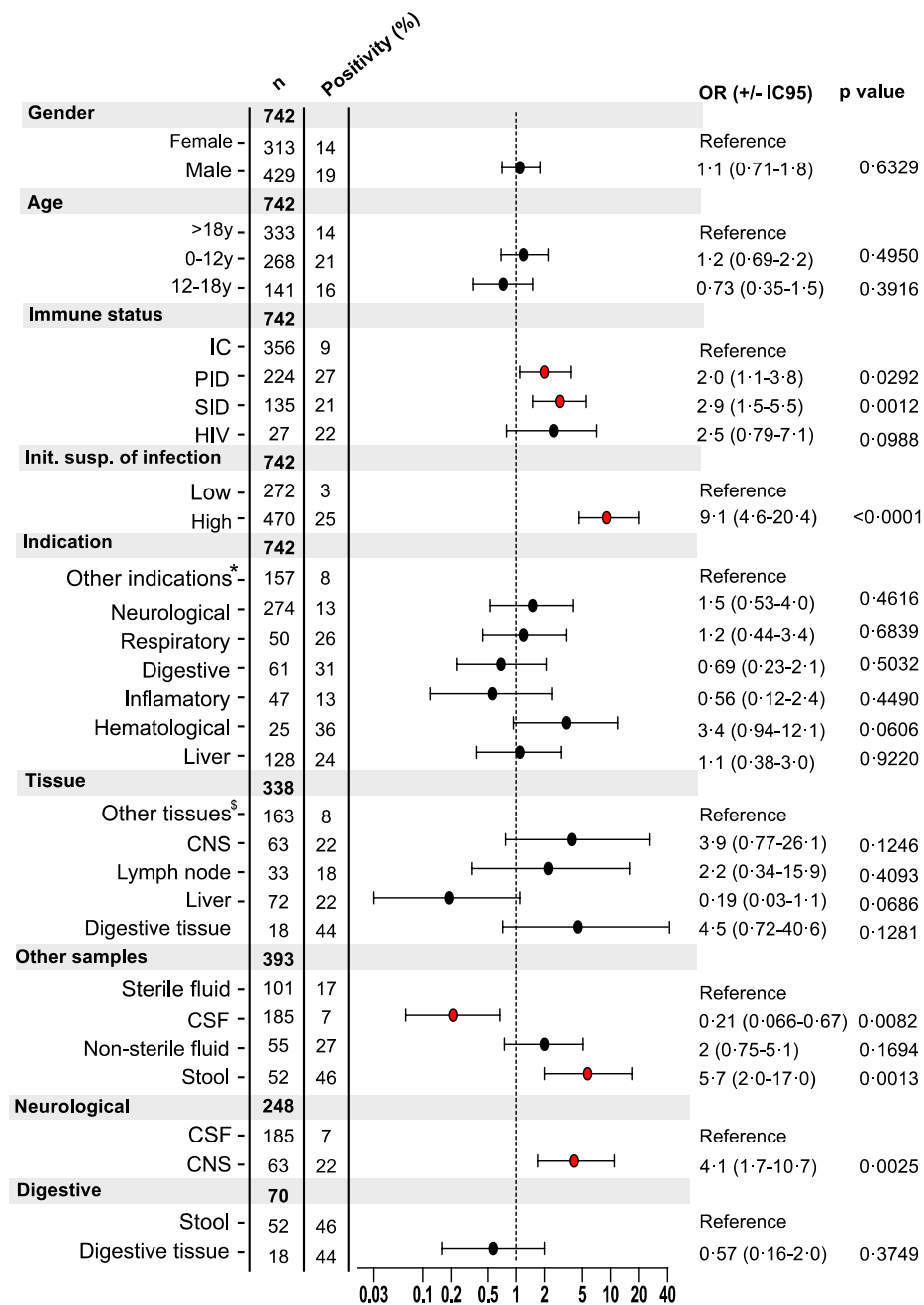


Figure 4. Odds Ratios for causative/possibly causative pathogen detection (95% CI)

Association among categorical variables (gender, age, initial suspicion of infection, indication, immune status, and sample type) and causative pathogen detection was assessed by multivariate logistic-regression analysis. Results are reported as odds ratios with 95% confidence intervals (IC95). Only p values <0.0500 were considered as significant and indicated by red dots.

IC, Immunocompetent; PID, Primary immune deficiency, including agammaglobulinemia, severe combined immunodeficiency, combined immunodeficiency; SID, Secondary immune deficiency, including dysimmune condition post hematopoietic stem cell transplantation, or immunosuppressive therapy; HIV, Patients living with HIV; CNS, Central Nervous System; CSF, Cerebrospinal Fluid; *=nephrological, osteo-muscular, cardiac, fetal, dermatological, otorhinolaryngological and ocular disorders; §=lung, kidney, osteo muscular, skin, ear, nose, throat, eye, spleen and placenta.

clinicians aim to exclude the presence of a pathogen.

Our study primarily identified viruses (95/111, 86% of infections diagnosed by mNGS in our cohort, including 57/95, 60% of RNA viruses), consistent with the second-line diagnostic approach where most causative bacteria, parasites and fungi have been eliminated through microscopic examination, prior culture and broad range PCR (16S, 16-23S, and 18S). Several drugs with broad spectrum antiviral activities – some of which are still in development, such as ribavirin, favipiravir, cidofovir, and brincidofovir may be helpful to treat severe viral infections for whom no specific treatment has been validated to date.¹⁹ Finally, the management of immunocompromised patients and the identification with mNGS of new clinical entities in this particularly infection-prone population are of great value.^{18,20} In immunocompromised patient population, we observed three situations in which metagenomics had a benefit as an untargeted and sensitive diagnostic tool: first, in case of opportunistic infections restricted to immunodeficient patients (e.g., encephalitis due to *Astrovirus VA1* in X-linked agammaglobulinemia patients).¹ Second, in case of infection with microorganisms known to be pathogenic in immunocompetent patients but with atypical symptoms in terms of chronicity, severity and/or clinical presentation (e.g., skin granuloma triggered by *Rubella virus*).²¹ Finally, in cases where infectious symptoms may be confused with inflammatory or dysimmune manifestations (e.g., enteric viruses causing symptoms versus gastrointestinal graft versus host disease).²² Overall, in 263 cases, the lack of pathogen detection by mNGS contributed to rule out infection before clinicians started immunosuppressive therapy.

In our setting, we observed that false-positive mNGS results for pathogen detection were associated with a very low number of reads (<5), prompting us to confirm the mNGS results with specific techniques before reporting results to the clinician. “False positive” mNGS results probably included contamination in the same series by high load of microorganism in another sample or contamination from other series as observed for SARS-CoV-2 sequence detection during COVID-19 outbreak. Alternatively, it could be true positive because of a greater sensitivity of mNGS, as demonstrated in a case of congenital Zika virus

infection (see also supplementary materials).¹⁶ “False negatives” correspond to low load of pathogens. Indeed, loads of viruses detected only by qPCR were very low and/or non-pathogenic reactivation of *herpesviruses*. Two microorganisms were detected only in culture (not detected by PCR), in line with the detection limit by molecular approach (Supplementary Table S8).

Most studies have evaluated the diagnostic performance of mNGS by comparing it with conventional microbiological tests based on retrospective sample analysis.²³ Those studies suggest positioning mNGS as a first-line test would increase the diagnosis rate compared to SMT. Owing to the current cost of mNGS compared to standard microbiology tests, we have initially decided to use it mostly as a second-line microbiological test. In our center, clinical microbiologists help in the prescription of the mNGS and can thus postpone it by performing other conventional tests in the first place, or on the contrary propose it as first line in case of precious samples. Interestingly, in 9% (65/742) of cases, the prescriptions came from pathologists who suspected an infectious process based on histological findings (lymphocytic infiltrate, placental intervillitis). Samples addressed by pathologists, allowed us to diagnose a case of *European Bat Lyssavirus 1* encephalitis in Western Europe,⁴ and a series of unexplained multiorgan involvement due to *Aichi virus*.²⁰

Our study has some limitations. We focused on the determinants of mNGS positivity and we did not assess the impact of mNGS on the management of patients nor did we include a cost analysis. Besides, we revealed the potential contribution of mNGS stool analysis to the diagnosis of hepatic,¹⁸ neurological disorders, or unexplained inflammatory diseases in immunocompromised hosts. Since pathogen excretion in the stool can be prolonged, and many individuals have low levels of pathogens in the stool, it is difficult to demonstrate causality beyond digestive symptoms. Further studies, including patient follow-up with monitoring of pathogen load kinetics and organ biopsies, are therefore warranted to establish guidelines based on how mNGS fits into diagnostic and treatment decision-making algorithms. Another limitation is the calculation of turnaround time from nucleic acid extraction. The distribution of time intervals between sampling and mNGS prescription, which will be highly variable (ranging from immediate to

several weeks when prescribed by a surgeon or a pathologist, respectively), would be interesting information. Finally, a delay of 3 to 9 days to obtain mNGS results may seem long compared with rapid, targeted PCR tests available in a matter of hours. However, non-targeted tests such as 16S PCR or uncommon specific PCRs or serologies performed by distant reference centers can take at least several days.

In the coming years, mNGS will likely allow the discovery of under-diagnosed or emerging infectious agents. Indeed, the increasing number of immunocompromised patients will most likely correlate with the emergence of new pathogen variants. Two scenarios may then occur. The mNGS could remain a second-line test for referral laboratories, while the most common newly discovered agents are progressively added to the first-line detection panels. For instance, following our mNGS findings, we have implemented a specific PCR for routine detection of *Aichi virus*, *Astrovirus VA1*, and *circovirus HCirV-1* in our center and currently perform it as a first-line test before mNGS in evocative cases. The development of very broad range first-line tests targeting hundreds of pathogens, by combining specific amplification or capture systems and NGS could be useful in this scenario. Alternatively, the democratization of mNGS, through its implementation in an increasing number of laboratories using high-throughput sequencing, could allow its use as a first-line test progressively supplanting the targeted approaches in all laboratories.

In conclusion, the results of this prospective observational study will contribute, along with future studies, to refining the positioning of mNGS in diagnostic and treatment decision-making algorithms.

Contributors

AJ, JF, PP, BR and ME conceptualized the study. AJ, JF, BR, TB, PP, and ME curated the data. JF, BR, NDR, FJ, DC, TB, ET, PP, ME, and AJ did the formal analysis. All authors contributed to the methods. AJ and JF wrote the original draft. All authors reviewed and edited the manuscript. AJ, JF, BR, PP, and ME have accessed and verified all the data in the study. All authors had full access to all the data of the study and accepted responsibility to submit for publication.

Declaration of interests

All other authors declare no competing interests.

Data sharing

Requests for data sharing should be submitted to the corresponding authors for consideration. Access to anonymized data might be granted following review via the AP-HP.Centre Institutional Research Review Board.

Acknowledgements

We thank the sequencing platform of Necker-Enfants Malades Hospital and Biomics platform of Institut Pasteur for providing access to the sequencers. We thank Prof. Fanny Lanternier and Dr Claire Rouzaud for providing clinical material from immunocompromised patients. We thank the technicians, biologists, and administrative staff of the microbiology department at Necker-Enfants Malades Hospital.

References

1. Brown JR, Morfopoulou S, Hubb J, et al. Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 2015; **60**(6): 881-8.
2. Parize P, Muth E, Richaud C, et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2017; **23**(8): 574 e1- e6.
3. Neven B, Pérot P, Bruneau J, et al. Cutaneous and Visceral Chronic Granulomatous Disease Triggered by a Rubella Virus Vaccine Strain in Children With Primary Immunodeficiencies. *Clinical Infectious Diseases* 2017; **64**(1): 83-6.
4. Regnault B, Evrard B, Plu I, et al. First Case of Lethal Encephalitis in Western Europe Due to European Bat Lyssavirus Type 1. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 2022; **74**(3): 461-6.
5. Pérot P, Bielle F, Bigot T, et al. Identification of Umbre Orthobunyavirus as a Novel Zoonotic Virus Responsible for Lethal Encephalitis in 2 French Patients with Hypogammaglobulinemia. *Clinical Infectious Diseases: An Official Publication*

- of the Infectious Diseases Society of America 2021; **72**(10): 1701-8.
6. Perot P, Fourgeaud J, Rouzaud C, et al. Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France. *Emerging infectious diseases* 2023; **29**(2): 286-93.
 7. Gu W, Deng X, Lee M, et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat Med* 2021; **27**(1): 115-24.
 8. Tao Y, Yan H, Liu Y, et al. Diagnostic Performance of Metagenomic Next-Generation Sequencing in Pediatric Patients: A Retrospective Study in a Large Children's Medical Center. *Clinical Chemistry* 2022; **68**(8): 1031-41.
 9. Chen L, Zhao Y, Wei J, et al. Metagenomic Next-Generation Sequencing for the Diagnosis of Neonatal Infectious Diseases. *Microbiol Spectr* 2022; **10**(6): e0119522.
 10. Lamoureux C, Surgers L, Fihman V, et al. Prospective Comparison Between Shotgun Metagenomics and Sanger Sequencing of the 16S rRNA Gene for the Etiological Diagnosis of Infections. *Frontiers in microbiology* 2022; **13**: 761873.
 11. Zhang R, Zhuang Y, Xiao ZH, et al. Diagnosis and Surveillance of Neonatal Infections by Metagenomic Next-Generation Sequencing. *Frontiers in microbiology* 2022; **13**: 855988.
 12. Shen H, Shen D, Song H, et al. Clinical assessment of the utility of metagenomic next-generation sequencing in pediatric patients of hematology department. *Int J Lab Hematol* 2021; **43**(2): 244-9.
 13. Kufner V, Plate A, Schmutz S, et al. Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases. *Genes (Basel)* 2019; **10**(9).
 14. Bharucha T, Oeser C, Balloux F, et al. STROBE-metagenomics: a STROBE extension statement to guide the reporting of metagenomics studies. *Lancet Infect Dis* 2020; **20**(10): e251-e60.
 15. Bukowska-Oško I, Perlejewski K, Pawełczyk A, et al. Human Pegivirus in Patients with Encephalitis of Unclear Etiology, Poland. *Emerg Infect Dis* 2018; **24**(10): 1785-94.
 16. Fourgeaud J, Regnault B, Faury H, et al. Fetal Zika virus infection diagnosed by metagenomic next-generation sequencing of amniotic fluid. *Ultrasound in Obstetrics & Gynecology* 2023; **61**(1): 116-7.
 17. Tam K, Torres VJ. Staphylococcus aureus Secreted Toxins and Extracellular Enzymes. *Microbiol Spectr* 2019; **7**(2).
 18. Riller Q, Fourgeaud J, Bruneau J, et al. Late-onset enteric virus infection associated with hepatitis (EVAH) in transplanted SCID patients. *J Allergy Clin Immunol* 2023: S0091-6749(23)00009-X.
 19. Stevaert A, Groaz E, Naesens L. Nucleoside analogs for management of respiratory virus infections: mechanism of action and clinical efficacy. *Current Opinion in Virology* 2022; **57**: 101279.
 20. Fourgeaud J, Lecuit MM, Perot P, et al. Chronic Aichi virus infection as a cause of long-lasting multi-organ involvement in patients with primary immune deficiencies. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2023.
 21. Wanat KA, Perelygina L, Chen M-h, et al. Association of Persistent Rubella Virus With Idiopathic Skin Granulomas in Clinically Immunocompetent Adults. *JAMA Dermatology* 2022; **158**(6): 626-33.
 22. Naymagon S, Naymagon L, Wong S-Y, et al. Acute graft-versus-host disease of the gut: considerations for the gastroenterologist. *Nat Rev Gastroenterol Hepatol* 2017; **14**(12): 711-26.
 23. Liu J, Zhang Q, Dong YQ, Yin J, Qiu YQ. Diagnostic accuracy of metagenomic next-generation sequencing in diagnosing infectious diseases: a meta-analysis. *Sci Rep* 2022; **12**(1): 21032.

DISCUSSION ET PERSPECTIVE

Discussion

Objectif 1

Discussion Article 1

Dans le premier article,¹⁷⁶ nous avons décrit une infection chronique par l'AiV chez six enfants présentant un déficit primitif lymphocytes B (XLA, ICF1 et SCID).

Ces enfants présentaient des symptômes similaires marqués par : une atteinte hépatique se traduisant par une augmentation des transaminases et une hépatomégalie ; une lymphoprolifération se traduisant par une splénomégalie et des adénopathies ; une atteinte rénale (4/6 patients) caractérisée par une néphromégalie et une néphropathie interstitielle ; et un infiltrat de cellules T CD8+ important dans les tissus hépatique et splénique. Ces caractéristiques incluant l'atteinte rénale ont aussi été décrites par d'autres auteurs chez 4 enfants présentant un déficit primaire ou secondaire en lymphocytes B.¹⁸²⁻¹⁸⁴

AiV n'a pas été retrouvé dans les selles, les urines et le plasma de 59 patients asymptomatiques présentant des déficits primaires en anticorps, des CID, SCID ou chez des patients post-HSCT.

L'analyse phylogénétique a montré que les infections chroniques à AiV chez ces patients pouvaient être causées soit par des AiV de génotype A, soit par des virus de génotype B probablement en fonction de la circulation du virus dans la population au moment de la contamination.

Nous avons mis en évidence l'ARN de AiV au niveau cellulaire dans le cytoplasme des hépatocytes d'un patient et dans des cellules des tissus spléniques pour deux patients, ce qui est en faveur d'une responsabilité directe du virus dans les lésions tissulaires et de la réaction immunitaire locale.

Enfin nous avons observé la clairance de l'ARN des AiV des compartiments sanguins, urinaires et hépatique chez deux patients jusqu'à un an après le traitement par HSCT. En parallèle de la clairance de AiV nous avons constaté une amélioration clinique significative, marquée par la reconstitution immunitaire et la normalisation des transaminases.

L'ensemble de ces résultats sont donc en faveur d'une nouvelle entité clinique survenant chez des patients ayant un déficit de l'immunité humorale et associant hépatite chronique, lymphoprolifération, néphromégalie et dysimmunité et causée par une infection chronique à AiV.

Bien que cette étude ait satisfait la majorité des postulats de Fredricks et Relman,¹⁰⁹ il existe plusieurs limites :

Tout d'abord, même si la résolution de la maladie coïncide avec l'élimination du virus, nous ne pouvons pas exclure que la restauration immunitaire du patient 1 grâce à la HCST aient pu guérir un processus auto-immun pathologique non identifié et indépendant de l'infection à AiV. Cette question pourrait être étudiée par l'expérimentation animale en infectant des

souris juvéniles SCID ou agammaglobulinémiques par une souche de *Murine kobuvirus* ou d'AiV. Le décès prématuré et la présence du virus dans les organes des souris infectées, comparés à un groupe témoin de souris SCID ou agammaglobulinémiques non infectées, plaideraient en défaveur d'un processus auto-immun pathologique non identifié chez les patients et en faveur d'une causalité directe du virus dans l'étiologie de l'hépatite chronique et de la dysimmunité.

Ensuite, les tissus rénaux et digestifs des patients n'étaient pas disponibles pour réaliser l'HIS et démontrer ou non l'implication de l'AiV dans les atteintes rénales ou digestives. La présence de séquences du virus dans des tissus rénaux avait été démontrée par *Buccioli et al*¹⁸² par mNGS chez un enfant présentant des symptômes similaires aux patients de notre étude. Dans cette dernière, 4 patients infectés présentaient une maladie rénale, les urines de 2 de ces 4 patients ont été testées et l'ARN de l'AiV a été détecté dans les 2 cas. L'implication du virus dans l'atteinte rénale est donc probable. Cependant nous n'avons pas pu démontrer la présence du virus, de son génome ou de ses transcrits dans les tissus digestifs. Seules les selles de 3/6 patients infectés ont pu être analysées, le AiV a été détecté dans 1 de ces 3 échantillons de selles mais les patients testés ne présentaient pas de signes digestifs francs. Même si l'AiV a déjà été incriminé dans des épidémies de gastroentérites, les résultats de notre étude n'apportent pas d'arguments supplémentaires à ce sujet.¹⁸¹ L'expérimentation animale pourrait également répondre à la question de la responsabilité du virus dans l'atteinte digestive.

Enfin, dans le groupe de 30 patients symptomatiques présentant les mêmes déficits immunitaires nous avons mis en évidence l'ARN de AiV chez seulement 6 patients (32%). Dans ce groupe, les caractéristiques des patients positifs pour AiV et ceux négatifs pour AiV montrent une grande superposition, sauf en ce qui concerne les anomalies rénales qui ont été observées exclusivement chez les patients positifs pour AiV. Il est important de noter que les signes de déséquilibre immunitaire, comme les affections hépatiques et la splénomégalie, sont généralement moins fréquents chez les patients atteints de XLA par rapport à ceux souffrant d'autres maladies telles qu'un déficit immunitaire commun variable ou combiné,^{185,186} ce qui est en faveur de l'hypothèse que ces symptômes sont dus à l'infection virale plutôt qu'à la maladie elle-même, néanmoins nous n'avons pas retrouvé l'AiV chez 24 patients symptomatiques. Un autre virus pourrait être le déclencheur d'une maladie immunitaire et pourrait persister dans un compartiment à distance des organes atteints. Une étude de cohorte à la recherche de traces d'autres étiologies virales par mNGS dans les selles et les urines de ces patients pourrait explorer cette hypothèse.

Discussion Article 2

Dans le deuxième article,¹⁷⁷ nous avons exploré le cas d'une patiente de 61 ans greffée cœur-poumon présentant une hépatite aigüe chez qui nous avons détecté un nouveau circovirus : *Human circovirus 1* (HCirV-1).

Nous avons d'abord caractérisé entièrement le génome de ce virus et l'analyse phylogénétique a révélé que les virus les plus proches étaient ceux du porc, *Porcine circovirus 3* (PC3), mais aussi ceux de mammifères plus exotique, l'otarie et le glouton.

L'étude des échantillons sanguins a ensuite montré que la temporalité de l'infection était en faveur d'une relation de cause à effet du virus dans la pathologie hépatique. En effet, le virus a été détecté pour la première fois dans un sérum prélevé avant la date de début des symptômes et les charges virales consécutives ont mis en évidence une corrélation entre une réplication virale élevée et une cytolysse hépatique importante.

Nous n'avons pas détecté le HCirV-1 chez 113 sujets contrôles : 56 testés par PCR dans le sang ou les selles, avec ou sans hépatite, d'étiologie connue (50%) ou indéterminée (50%), et 57 patients testés par l'analyse en mNGS du tissu hépatique avec une hépatite d'étiologie inconnue. Nous n'avons donc pas d'argument en faveur d'une prévalence importante de l'ADN de HCirV-1 dans le sang, les selles ou les tissus hépatiques.

Nous avons détecté les transcrits viraux au niveau cellulaire par hybridation *in situ*, avec un fort signal dans le noyau de 2% des hépatocytes, ce qui est compatible avec le cycle viral des virus de la famille des *Circoviridae*, et est fortement en faveur du rôle du HCirV-1 en tant qu'agent causal de l'hépatite en ciblant et en induisant la lyse des hépatocytes directement ou via la réponse immunitaire de l'hôte.

Étant donné que les circovirus qui infectent l'animal, tels que PCV2 et PCV3, peuvent être responsables d'hépatites chez le porc et le cheval,^{187,188} l'ensemble des résultats sont en faveur du rôle pathogène du HCirV-1 dans l'hépatite aigüe de la patiente.

Bien que cette étude ait satisfait la majorité des critères d'imputabilités de Fredricks et Relman,¹⁰⁹ il existe plusieurs limites :

Nous n'avons pas pu déduire le mode et la source de transmission du HCirV-1. Les porcs peuvent être infectés expérimentalement par inoculation intramusculaire, orale, oro-nasale et intra-utérine, l'exposition oro-nasale est probablement la voie naturelle de transmission du PCV2.¹⁸⁹⁻¹⁹¹ Nous avons exploré l'hypothèse d'une contamination interhumaine ou par un animal domestique en analysant les prélèvements de salive et de selles du conjoint ainsi que des prélèvements de selles des deux chats de la patiente, ces prélèvements étaient négatifs. De plus la patiente n'avait jamais été en contact direct avec un porc ou un animal exotique dont les circovirus étaient phylogénétiquement proches. Une autre hypothèse serait une contamination par la transplantation d'organe, 17 ans avant l'étude, mais les prélèvements sanguins du donneur comme ceux de la patiente avant les symptômes étaient négatifs en PCR

HCirV-1. Pour explorer cette hypothèse, le développement d'une technique de sérologie pourrait révéler si la patiente était déjà immunisée 4 années avant l'étude ou si la séroconversion s'est produite quelques jours ou semaines avant la date de début des symptômes.

Le nombre de témoins inclus dans l'étude est faible. Une enquête épidémiologique sur un grand nombre de sujets en population générale (donneurs de sang) pour déterminer la prévalence de l'ADN du virus et des anticorps anti-HCirV-1 permettrait probablement d'identifier la source du virus, son mode de transmission et de mieux comprendre les maladies qu'il est susceptible de provoquer. Les virus qui provoquent des maladies graves ou le décès de l'hôte sont plus facilement identifiés en raison de leur présentation clinique facilement perceptible, tandis qu'un virus qui entraîne peu ou pas de symptômes lors de l'infection de l'hôte est plus difficile à identifier. Le HCirV-1 a donc pu circuler un certain temps à bas bruit en infectant des hôtes immunocompétents de façon asymptomatique et provoquer une maladie symptomatique uniquement chez quelques patients immunodéprimés, permettant ainsi sa détection.

Objectif 2

Discussion Article 3

Dans cette étude, nous rapportons l'utilisation prospective de l'analyse mNGS au cours des trois premières années de sa mise en place dans notre centre. Au cours de cette période, nous avons analysé 742 échantillons chez 523 patients dont 47% présentaient un déficit immunitaire. L'ajout de l'analyse mNGS à notre panel diagnostique a permis de trouver une étiologie microbienne dans 25% (117/470) des échantillons lorsque l'hypothèse diagnostique principale au moment du prélèvement était l'infection. Les résultats de cette étude montrent les avantages potentiels de l'ajout de la mNGS aux processus de diagnostic de routine, tant pour la détection que pour la découverte de pathogènes rares ou nouveaux.^{175-177,192,193} Nous avons identifié que i) dans la sous population des patients immunodéprimés la rentabilité diagnostique du mNGS était significativement augmentée par rapport à l'ensemble de la population (OR=2,4, p=0,0013), ii) que les échantillons de selles étaient ceux dans lesquels un agent causal était le plus souvent retrouvé (OR=5,7 ; p=0,0013), tandis que les échantillons de LCS étaient ceux dans lesquels un agent causal était le moins souvent retrouvé (OR=0,21 ; p=0,0082).

Dans la population des patients immunodéprimés, nous avons observé trois situations dans lesquelles l'analyse mNGS présentait un avantage: premièrement, en cas d'infections opportunistes rares retrouvées uniquement en cas d'immunodépression (par exemple, une encéphalite due à l'*Astrovirus VA1* chez un patient atteint d'agammaglobulinémie liée au chromosome X).^{156,157} Deuxièmement, en cas d'infection par des micro-organismes connus pour être pathogènes chez les patients immunocompétents, mais présentant des symptômes atypiques en termes de chronicité, de gravité et/ou de présentation clinique (par exemple, granulome cutané déclenché par la rubéole ou l'encéphalite à inclusion de la rougeole).^{145,194} Enfin, dans les cas où les symptômes infectieux peuvent être confondus avec des manifestations inflammatoires ou dysimmunitaires (par exemple, les virus entériques provoquant des symptômes comparables à la maladie gastro-intestinale du greffon contre l'hôte).¹⁹⁵

Dans cette étude le taux de positivité le plus élevé a été observé dans les échantillons de selles, chez des patients avec des symptômes digestifs, hépatiques, hématologiques ou neurologiques. Étant donné que l'excrétion de pathogènes dans les selles peut être prolongée¹²¹ et que de nombreux individus présentent de faibles niveaux de pathogènes dans les selles sans conséquence clinique, il est difficile de démontrer la causalité au-delà des symptômes digestifs, sauf si le virus est connu pour être responsable du tableau clinique, mais à défaut de faire le diagnostic de certitude, cette démarche pourrait encourager les explorations complémentaires.

Sur la base des résultats de cette étude et étant donné les coûts encore élevés du séquençage, nous avons décidé de maintenir la technique mNGS en seconde ligne après un bilan de première ligne bien conduit et de maintenir la vérification des prescriptions à la réception des

prélèvements. Nous pensons que l'analyse mNGS devrait être prescrite dans les situations cliniques graves et inexplicées suivantes :

- Dans les cas où les prélèvements de tissus sont disponibles (biopsies d'organes) avec une forte suspicion d'infection virale.
- En cas d'arguments histologiques en faveur d'une infection le pathologiste peut prescrire l'analyse mNGS.
- En l'absence de tissus, l'analyse d'un prélèvement de selles peut être considérée notamment en cas de symptômes digestifs, hépatiques, hématologiques ou neurologiques.
- En l'absence de tissus, l'analyse d'un prélèvement d'urine peut être considérée en cas de suspicion d'arboviroses ou en cas d'atteinte rénale.
- En cas d'encéphalite inexplicée et en l'absence de tissus cérébraux, il nous semble pertinent d'analyser par mNGS des prélèvements de selles ou d'urines selon le contexte.

Nous pensons que l'analyse mNGS ne devrait pas être prescrite dans les situations suivantes :

- En cas de suspicion d'infection bactérienne ou fongiques les tests conventionnels doivent être priorités, y compris les techniques de séquençage à large spectre 16S et ITS.
- L'analyse du LCS par mNGS en cas d'encéphalite inexplicée par un bilan microbiologique de première ligne complet, dans ce cas d'autres prélèvements devraient être priorités (tissus cérébraux, selles ou urines).

Dans cette étude prospective, 742 échantillons prélevés chez 523 patients ont été analysés sans aucune restriction quant aux types d'échantillons, aux indications médicales ou à la population de patients étudiée. Nous avons révélé les déterminants majeurs de la positivité de l'analyse et précisé certaines indications. Cependant nous n'avons pas évalué l'impact de la détection ou non d'un agent pathogène sur la prise en charge du patient, pas plus que nous n'avons inclus une analyse des coûts. D'autres études, comprenant une analyse globale des coûts dans le parcours de soins, un suivi des patients et de l'issue de leurs maladies, sont donc nécessaires pour affiner les algorithmes incluant le mNGS pour une prise de décision en matière de diagnostic et de traitement.

Contribution et conclusion générale

Ces travaux ont contribué à démontrer le rôle pathogène suspecté du *Aichi virus* chez des enfants présentant un déficit immunitaire primitif, à la découverte d'un nouveau virus plausiblement pathogène chez une patiente transplantée d'organe solide, et à préciser les indications de la métagénomique clinique pour le diagnostic microbiologique. Grâce à la première partie de cette thèse, nous avons ajouté la recherche par RT-PCR spécifique AiV dans les selles, le plasma et les urines des patients présentant un déficit immunitaire à notre bilan de première ligne et nous avons développé une technique de PCR en temps réel pour la détection du HCirV-1 au laboratoire. Comme dans d'autres études,¹⁹⁶ nous avons confirmé que l'analyse mNGS est un outil précieux pour l'exploration de maladies inexplicables. Grâce à la seconde partie de ce travail de thèse, nous avons pu préciser les indications de l'analyse mNGS et les partager avec les cliniciens de notre centre dans une démarche institutionnelle d'amélioration de la prescription.

Perspective

Au sein des laboratoires de virologie des centres hospitaliers la définition d'un groupe d'agents pathogènes viraux de première ligne adapté à la patientèle du centre est nécessaire pour le diagnostic et le suivi des infections virales. Par exemple les charges virales CMV, EBV, ADV et BKV chez les patients immunodéprimés ou les charges virales et génotypages de résistances aux antiviraux pour le VIH, le VHC et le VHB. Ces tests moléculaires de première ligne sont de plus en plus automatisés et peuvent produire des résultats rapidement. La mise en place de l'analyse mNGS dans certains centres va probablement permettre à court terme d'enrichir le panel de virus recherchés en première ligne diagnostique en fonction de la maladie ou du terrain. Malgré la démocratisation du séquençage à haut débit et la diminution du coût des séquences, l'analyse des centaines de millions de séquences générées en mNGS nécessite des compétences spécialisées et limite encore l'accès de cette technique de diagnostic à tous les centres de soins.

À la suite de la détection par mNGS d'un virus nouveau ou rare, plusieurs médicaments à large spectre d'activité antivirale tels que la ribavirine, le favipiravir, le cidofovir et le brincidofovir pourraient être utilisés pour traiter les infections graves pour lesquelles aucun traitement spécifique n'a été validé à ce jour.¹⁹⁷ Les options de traitements sont donc limitées, mais la disponibilité des traitements antiviraux a augmenté à la suite de la pandémie de COVID-19. Par exemple les traitements anti-SARS-CoV-2 comme le remdesivir ou le nirmatrelvir/ritonavir (Paxlovid®) ont montré leurs efficacités *in vitro* contre certains coronavirus saisonniers^{198,199} et pourraient être utilisés de manière expérimentale pour traiter les encéphalites rares et sévères à coronavirus OC43 du patient immunodéprimé.^{158,159} La disponibilité d'anticorps monoclonaux pour le traitement des maladies virales a également augmenté ce qui contribuera probablement à proposer d'avantages de traitements expérimentaux, et peut être même des anticorps monoclonaux synthétisés à façon sur la base des séquences des protéines de surface des virus. Dans cette dernière approche hypothétique, l'accès aux séquences virales par l'analyse mNGS serait particulièrement intéressante.

Outre la minorité de nouveaux virus qui ont été clairement associés à une pathogénicité chez l'Homme, tels que le MERS-CoV²⁰⁰ et le SARS-CoV-2, de nombreux virus nouvellement identifiés attendent une caractérisation plus poussée et une évaluation de leur pertinence clinique.²⁰¹ Par exemple le *Human Pegivirus* est généralement considéré comme non pathogène. Cependant, il a été détecté par mNGS dans des tissus cérébraux de patients et associé à des déficits neurologiques et visuels, bien que sans preuve formelle à ce jour.²⁰²⁻²⁰⁴ Pour ces virus qui sont de plus en plus nombreux dans ce contexte de séquençage de masse, des efforts doivent être déployés en virologie fondamentale afin de mieux les caractériser et en virologie clinique dans le but d'évaluer leur intérêt médical.

Références

- 1 Simmonds P, Adriaenssens EM, Zerbini FM, *et al.* Four principles to establish a universal virus taxonomy. *PLOS Biology* 2023; **21**: e3001922.
- 2 Siddell SG, Smith DB, Adriaenssens E, *et al.* Virus taxonomy and the role of the International Committee on Taxonomy of Viruses (ICTV). *Journal of General Virology* 2023; **104**: 001840.
- 3 Simmonds P, Adams MJ, Benkő M, *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* 2017; **15**: 161–8.
- 4 Roux S, Adriaenssens EM, Dutilh BE, *et al.* Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat Biotechnol* 2019; **37**: 29–37.
- 5 Dutilh BE, Varsani A, Tong Y, *et al.* Perspective on taxonomic classification of uncultivated viruses. *Current Opinion in Virology* 2021; **51**: 207–15.
- 6 Koonin EV, Dolja VV, Krupovic M, *et al.* Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol Mol Biol Rev* 2020; **84**: e00061-19.
- 7 Wolf YI, Silas S, Wang Y, *et al.* Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* 2020; **5**: 1262–70.
- 8 Shi M, Lin X-D, Tian J-H, *et al.* Redefining the invertebrate RNA virosphere. *Nature* 2016; **540**: 539–43.
- 9 Gorbalenya AE, Krupovic M, Mushegian A, *et al.* The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol* 2020; **5**: 668–74.
- 10 Baltimore D. Expression of animal virus genomes. *Bacteriol Rev* 1971; **35**: 235–41.
- 11 Koonin EV, Krupovic M, Agol VI. The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? *Microbiology and Molecular Biology Reviews* 2021; **85**: 10.1128/mmbr.00053-21.
- 12 Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci USA* 2011; **108**: 17486–91.
- 13 Ahlquist P. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nat Rev Microbiol* 2006; **4**: 371–82.
- 14 Philippe N, Legendre M, Doutre G, *et al.* Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 2013; **341**: 281–6.
- 15 Gershon AA, Breuer J, Cohen JI, *et al.* Varicella zoster virus infection. *Nat Rev Dis Primers* 2015; **1**: 1–18.
- 16 Gatherer D, Depledge DP, Hartley CA, *et al.* ICTV Virus Taxonomy Profile: Herpesviridae 2021. *Journal of General Virology* 2021; **102**: 001673.
- 17 Sean P. J. Whelan. Fields virology. Volume 4, Fundamentals, chapter 5, Viral Replication Strategies. <https://eds.s.ebscohost.com/eds/detail/detail?vid=5&sid=72563363-03f5-437f-962b-72470bb0fcac%40redis&bdata=Jmxhbm9ZnImc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=msip.270489282&db=cat03336a> (accessed Aug 12, 2023).
- 18 Lieberman PM. Epigenetics and Genetics of Viral Latency. *Cell Host & Microbe* 2016; **19**: 619–28.
- 19 Cohen JI. Herpesvirus latency. *J Clin Invest* 2020; **130**: 3361–9.
- 20 Dougherty WG, Semler BL. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol Rev* 1993; **57**: 781–822.

- 21 Strauss JH, Strauss EG. The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev* 1994; **58**: 491–562.
- 22 Hertz MI, Thompson SR. Mechanism of translation initiation by Dicistroviridae IGR IRESs. *Virology* 2011; **411**: 355–61.
- 23 Neumann G, Brownlee GG, Fodor E, Kawaoka Y. Orthomyxovirus replication, transcription, and polyadenylation. *Curr Top Microbiol Immunol* 2004; **283**: 121–43.
- 24 Whelan SPJ, Barr JN, Wertz GW. Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr Top Microbiol Immunol* 2004; **283**: 61–119.
- 25 Nikolic J, Le Bars R, Lama Z, *et al.* Negri bodies are viral factories with properties of liquid organelles. *Nat Commun* 2017; **8**. DOI:10.1038/s41467-017-00102-9.
- 26 Nevers Q, Albertini AA, Lagaudrière-Gesbert C, Gaudin Y. Negri bodies and other virus membrane-less replication compartments. *Biochim Biophys Acta Mol Cell Res* 2020; **1867**: 118831.
- 27 Taylor JM. Chapter 3. Replication of the hepatitis delta virus RNA genome. *Adv Virus Res* 2009; **74**: 103–21.
- 28 Sm M, Jt P. Assortment and packaging of the segmented rotavirus genome. *Trends in microbiology* 2011; **19**. DOI:10.1016/j.tim.2010.12.002.
- 29 Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 1970; **226**: 1209–11.
- 30 Minor PD. An Introduction to Poliovirus: Pathogenesis, Vaccination, and the Endgame for Global Eradication. In: Martín J, ed. *Poliovirus: Methods and Protocols*. New York, NY: Springer, 2016: 1–10.
- 31 Ohka S, Nomoto A. The molecular basis of poliovirus neurovirulence. *Dev Biol (Basel)* 2001; **105**: 51–8.
- 32 Racaniello VR. One hundred years of poliovirus pathogenesis. *Virology* 2006; **344**: 9–16.
- 33 Mark T. Heise, Peter M. Howley, David M. Knipe. *Fields virology*. Volume 4, Fundamentals, chapter 8, Pathogenesis of Viral Infection. <https://eds.s.ebscohost.com/eds/detail/detail?vid=5&sid=72563363-03f5-437f-962b-72470bb0fcac%40redis&bdata=Jmxhbm9ZnImc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=msip.270489282&db=cat03336a> (accessed Aug 12, 2023).
- 34 Becker Y. HSV-1 brain infection by the olfactory nerve route and virus latency and reactivation may cause learning and behavioral deficiencies and violence in children and adults: a point of view. *Virus Genes* 1995; **10**: 217–26.
- 35 Mori I, Goshima F, Ito H, *et al.* The vomeronasal chemosensory system as a route of neuroinvasion by herpes simplex virus. *Virology* 2005; **334**: 51–8.
- 36 Klingen Y, Conzelmann K-K, Finke S. Double-labeled rabies virus: live tracking of enveloped virus transport. *J Virol* 2008; **82**: 237–45.
- 37 Cain MD, Rubin Klein N, Jiang X, Klein RS. Post-exposure intranasal IFN α suppresses replication and neuroinvasion of Venezuelan Equine Encephalitis virus within olfactory sensory neurons. *bioRxiv* 2023; : 2023.06.30.547169.
- 38 Schäfer A, Brooke CB, Whitmore AC, Johnston RE. The role of the blood-brain barrier during Venezuelan equine encephalitis virus infection. *J Virol* 2011; **85**: 10682–90.
- 39 Bernard E, Hamel R, Neyret A, *et al.* Human keratinocytes restrict chikungunya virus replication at a post-fusion step. *Virology* 2015; **476**: 1–10.
- 40 Pesko K, Westbrook CJ, Mores CN, Lounibos LP, Reiskind MH. Effects of infectious virus dose and bloodmeal delivery method on susceptibility of *Aedes aegypti* and *Aedes albopictus* to chikungunya virus. *J Med Entomol* 2009; **46**: 395–9.
- 41 Reiskind MH, Pesko K, Westbrook CJ, Mores CN. Susceptibility of Florida mosquitoes to infection with chikungunya virus. *Am J Trop Med Hyg* 2008; **78**: 422–5.

- 42 Tesla B, Demakovskiy LR, Packiam HS, *et al.* Estimating the effects of variation in viremia on mosquito susceptibility, infectiousness, and R0 of Zika in *Aedes aegypti*. *PLoS Negl Trop Dis* 2018; **12**: e0006733.
- 43 Ikegami T, Makino S. The pathogenesis of Rift Valley fever. *Viruses* 2011; **3**: 493–519.
- 44 Monath TP. Treatment of yellow fever. *Antiviral Res* 2008; **78**: 116–24.
- 45 Ioos S, Mallet H-P, Leparac Goffart I, Gauthier V, Cardoso T, Herida M. Current Zika virus epidemiology and recent epidemics. *Médecine et Maladies Infectieuses* 2014; **44**: 302–7.
- 46 Griffin DE. Immune responses during measles virus infection. *Curr Top Microbiol Immunol* 1995; **191**: 117–34.
- 47 Griffin DE. Measles virus-induced suppression of immune responses. *Immunological Reviews* 2010; **236**: 176–89.
- 48 Ko JY, Danielson ML, Town M, *et al.* Risk Factors for Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization: COVID-19-Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System. *Clin Infect Dis* 2021; **72**: e695–703.
- 49 RECOVERY Collaborative Group, Horby P, Lim WS, *et al.* Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* 2021; **384**: 693–704.
- 50 Gu SH, Song DH, Lee D, *et al.* Whole-genome sequence analysis of Zika virus, amplified from urine of traveler from the Philippines. *Virus Genes* 2017; **53**: 918–21.
- 51 Hübschen JM, Gouandjika-Vasilache I, Dina J. Measles. *The Lancet* 2022; **399**: 678–90.
- 52 Tyler KL. Acute Viral Encephalitis. *N Engl J Med* 2018; **379**: 557–66.
- 53 Rotbart HA. Enzymatic RNA amplification of the enteroviruses. *J Clin Microbiol* 1990; **28**: 438–42.
- 54 Chapman NM, Tracy S, Gauntt CJ, Fortmueller U. Molecular detection and identification of enteroviruses using enzymatic amplification and nucleic acid hybridization. *J Clin Microbiol* 1990; **28**: 843–50.
- 55 Margulies M, Egholm M, Altman WE, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; **437**: 376–80.
- 56 Sun X, Song L, Yang W, *et al.* Nanopore Sequencing and Its Clinical Applications. In: Huang T, ed. Precision Medicine. New York, NY: Springer US, 2020: 13–32.
- 57 Maghini DG, Moss EL, Vance SE, Bhatt AS. Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nat Protoc* 2021; **16**: 458–71.
- 58 Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016; **17**: 239.
- 59 Quick J, Loman NJ, Duraffour S, *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016; **530**: 228–32.
- 60 Grubaugh ND, Ladner JT, Kraemer MUG, *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 2017; **546**: 401–5.
- 61 Ko KKK, Chng KR, Nagarajan N. Metagenomics-enabled microbial surveillance. *Nat Microbiol* 2022; **7**: 486–96.
- 62 Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet* 2019; **20**: 341–55.
- 63 Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol* 2017; **15**: 183–92.
- 64 Thomson E, Ip CLC, Badhan A, *et al.* Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. *J Clin Microbiol* 2016; **54**: 2470–84.
- 65 Cotten M, Petrova V, Phan MVT, *et al.* Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. *J Virol* 2014; **88**: 11056–69.
- 66 Jakava-Viljanen M, Nokireki T, Sironen T, Vapalahti O, Sihvonen L, Huovilainen A.

- Evolutionary trends of European bat lyssavirus type 2 including genetic characterization of Finnish strains of human and bat origin 24 years apart. *Arch Virol* 2015; **160**: 1489–98.
- 67 Agoti CN, Otieno JR, Munywoki PK, *et al.* Local evolutionary patterns of human respiratory syncytial virus derived from whole-genome sequencing. *J Virol* 2015; **89**: 3444–54.
- 68 Brown JR, Roy S, Ruis C, *et al.* Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: a Robust and Sensitive Method. *J Clin Microbiol* 2016; **54**: 2530–7.
- 69 Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog* 2011; **7**: e1001344.
- 70 Salipante SJ, Sengupta DJ, Rosenthal C, *et al.* Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLoS One* 2013; **8**: e65226.
- 71 Wagner K, Springer B, Pires VP, Keller PM. Molecular detection of fungal pathogens in clinical specimens by 18S rDNA high-throughput screening in comparison to ITS PCR and culture. *Sci Rep* 2018; **8**: 6964.
- 72 Quick J, Grubaugh ND, Pullan ST, *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017; **12**: 1261–76.
- 73 Faria NR, Quick J, Claro IM, *et al.* Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 2017; **546**: 406–10.
- 74 Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. *Genome Res* 2015; **25**: 1910–20.
- 75 Briese T, Kapoor A, Mishra N, *et al.* Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *mBio* 2015; **6**: e01491-01415.
- 76 Miller S, Naccache SN, Samayoa E, *et al.* Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res* 2019; **29**: 831–42.
- 77 Blauwkamp TA, Thair S, Rosen MJ, *et al.* Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* 2019; **4**: 663–74.
- 78 Schlaberg R, Queen K, Simmon K, *et al.* Viral Pathogen Detection by Metagenomics and Pan-Viral Group Polymerase Chain Reaction in Children With Pneumonia Lacking Identifiable Etiology. *J Infect Dis* 2017; **215**: 1407–15.
- 79 Shkoporov AN, Hill C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host & Microbe* 2019; **25**: 195–209.
- 80 Liang G, Bushman FD. The human virome: assembly, composition and host interactions. *Nat Rev Microbiol* 2021; **19**: 514–27.
- 81 Spandole S, Cimponeriu D, Berca LM, Mihăescu G. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol* 2015; **160**: 893–908.
- 82 Monaco CL, Gootenberg DB, Zhao G, *et al.* Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* 2016; **19**: 311–22.
- 83 Liang G, Conrad MA, Kelsen JR, *et al.* Dynamics of the Stool Virome in Very Early-Onset Inflammatory Bowel Disease. *Journal of Crohn's and Colitis* 2020; **14**: 1600–10.
- 84 van Rijn A, Roos R, Dekker F, Rotmans J, Feltkamp M. Torque teno virus load as marker of rejection and infection in solid organ transplantation – A systematic review and meta-analysis. *Reviews in Medical Virology* 2023; **33**: e2393.
- 85 van Rijn AL, Wunderink HF, Sidorov IA, *et al.* Torque teno virus loads after kidney transplantation predict allograft rejection but not viral infection. *J Clin Virol* 2021; **140**: 104871.
- 86 Legoff J, Resche-Rigon M, Bouquet J, *et al.* The eukaryotic gut virome in hematopoietic

- stem cell transplantation: new clues in enteric graft-versus-host disease. *Nat Med* 2017; **23**: 1080–5.
- 87 Abbas AA, Taylor LJ, Dothard MI, *et al.* Redondoviridae, a Family of Small, Circular DNA Viruses of the Human Oro-Respiratory Tract Associated with Periodontitis and Critical Illness. *Cell Host & Microbe* 2019; **25**: 719-729.e4.
- 88 Minot S, Sinha R, Chen J, *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011; **21**: 1616–25.
- 89 Shkoporov AN, Clooney AG, Sutton TDS, *et al.* The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* 2019; **26**: 527-541.e5.
- 90 Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLoS One* 2015; **10**: e0134941.
- 91 Clooney AG, Sutton TDS, Shkoporov AN, *et al.* Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe* 2019; **26**: 764-778.e5.
- 92 Norman JM, Handley SA, Baldrige MT, *et al.* Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 2015; **160**: 447–60.
- 93 Zuo T, Lu X-J, Zhang Y, *et al.* Gut mucosal virome alterations in ulcerative colitis. *Gut* 2019; **68**: 1169–79.
- 94 Fernandes MA, Verstraete SG, Phan TG, *et al.* Enteric Virome and Bacterial Microbiota in Children With Ulcerative Colitis and Crohn Disease. *J Pediatr Gastroenterol Nutr* 2019; **68**: 30–6.
- 95 Lindfors K, Lin J, Lee H-S, *et al.* Metagenomics of the faecal virome indicate a cumulative effect of enterovirus and gluten amount on the risk of coeliac disease autoimmunity in genetically at risk children: the TEDDY study. *Gut* 2020; **69**: 1416–22.
- 96 Khan Mirzaei M, Khan MAA, Ghosh P, *et al.* Bacteriophages Isolated from Stunted Children Can Regulate Gut Bacterial Communities in an Age-Specific Manner. *Cell Host Microbe* 2020; **27**: 199-212.e5.
- 97 Desai C, Handley SA, Rodgers R, *et al.* Growth velocity in children with Environmental Enteric Dysfunction is associated with specific bacterial and viral taxa of the gastrointestinal tract in Malawian children. *PLoS Negl Trop Dis* 2020; **14**: e0008387.
- 98 Kernbauer E, Ding Y, Cadwell K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature* 2014; **516**: 94–8.
- 99 Ingle H, Lee S, Ai T, *et al.* Viral complementation of immunodeficiency confers protection against enteric pathogens via interferon- λ . *Nat Microbiol* 2019; **4**: 1120–8.
- 100 Reyes A, Blanton LV, Cao S, *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci U S A* 2015; **112**: 11941–6.
- 101 Pérez-Brocal V, García-López R, Nos P, Beltrán B, Moret I, Moya A. Metagenomic Analysis of Crohn's Disease Patients Identifies Changes in the Virome and Microbiome Related to Disease Status and Therapy, and Detects Potential Interactions and Biomarkers. *Inflamm Bowel Dis* 2015; **21**: 2515–32.
- 102 Tirosh O, Conlan S, Deming C, *et al.* Expanded skin virome in DOCK8-deficient patients. *Nat Med* 2018; **24**: 1815–21.
- 103 Nakatsu G, Zhou H, Wu WKK, *et al.* Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology* 2018; **155**: 529-541.e5.
- 104 Ungaro F, Massimino L, Furfaro F, *et al.* Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes* 2019; **10**: 149–58.
- 105 Li Y, Altan E, Pilcher C, *et al.* Semen virome of men with HIV on or off antiretroviral treatment. *AIDS* 2020; **34**: 827–32.
- 106 Fredricks DN, Relman DA. The infectious aetiology of disease: the search for new agents. *Medicine* 2005; **33**: 37–8.

- 107 Rivers TM. Viruses and Koch's Postulates. *J Bacteriol* 1937; **33**: 1–12.
- 108 Hill AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? *Proc R Soc Med* 1965; **58**: 295–300.
- 109 Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* 1996; **9**: 18–33.
- 110 Boutolleau D, Burrel S. Prise en charge thérapeutique des infections à herpèsvirus : traitements actuels et futurs. *Revue Francophone des Laboratoires* 2016; **2016**: 55–64.
- 111 Gershon AA, Gershon MD. Pathogenesis and Current Approaches to Control of Varicella-Zoster Virus Infections. *Clinical Microbiology Reviews* 2013; **26**: 728–43.
- 112 Agut H, Bonnafous P, Gautheret-Dejean A. Laboratory and Clinical Aspects of Human Herpesvirus 6 Infections. *Clinical Microbiology Reviews* 2015; **28**: 313–35.
- 113 Kotton CN, Kumar D, Caliendo AM, *et al.* The Third International Consensus Guidelines on the Management of Cytomegalovirus in Solid-organ Transplantation. *Transplantation* 2018; **102**: 900.
- 114 Styczynski J, van der Velden W, Fox CP, *et al.* Management of Epstein-Barr Virus infections and post-transplant lymphoproliferative disorders in patients after allogeneic hematopoietic stem cell transplantation: Sixth European Conference on Infections in Leukemia (ECIL-6) guidelines. *Haematologica* 2016; **101**: 803–11.
- 115 Strahan R, Uppal T, Verma SC. Next-Generation Sequencing in the Understanding of Kaposi's Sarcoma-Associated Herpesvirus (KSHV) Biology. *Viruses* 2016; **8**: 92.
- 116 Lion T. Adenovirus infections in immunocompetent and immunocompromised patients. *Clin Microbiol Rev* 2014; **27**: 441–62.
- 117 Kosulin K, Haberler C, Hainfellner JA, Amann G, Lang S, Lion T. Investigation of adenovirus occurrence in pediatric tumor entities. *J Virol* 2007; **81**: 7629–35.
- 118 Lion T. Adenovirus persistence, reactivation, and clinical management. *FEBS Lett* 2019; **593**: 3571–82.
- 119 Roy S, Calcedo R, Medina-Jaszek A, Keough M, Peng H, Wilson JM. Adenoviruses in Lymphocytes of the Human Gastro-Intestinal Tract. *PLOS ONE* 2011; **6**: e24859.
- 120 Garnett CT, Talekar G, Mahr JA, *et al.* Latent Species C Adenoviruses in Human Tonsil Tissues. *Journal of Virology* 2009; **83**: 2417–28.
- 121 Kosulin K, Geiger E, Vécsei A, *et al.* Persistence and reactivation of human adenoviruses in the gastrointestinal tract. *Clin Microbiol Infect* 2016; **22**: 381.e1-381.e8.
- 122 D M, E L, G H, S D, M R, A H. Type dependent patterns of human adenovirus persistence in human T-lymphocyte cell lines. *Journal of medical virology* 2014; **86**. DOI:10.1002/jmv.23736.
- 123 Michaels MG, Green M, Wald ER, Starzl TE. Adenovirus Infection in Pediatric Liver Transplant Recipients. *The Journal of Infectious Diseases* 1992; **165**: 170–4.
- 124 Ito M, Hirabayashi N, Uno Y, Nakayama A, Asai J. Necrotizing tubulointerstitial nephritis associated with adenovirus infection. *Hum Pathol* 1991; **22**: 1225–31.
- 125 Otori NP, Michaels MG, Jaffe R, Williams P, Yousem SA. Adenovirus pneumonia in lung transplant recipients. *Human Pathology* 1995; **26**: 1073–9.
- 126 Shirali GS, Ni J, Chinnock RE, *et al.* Association of viral genome with graft loss in children after cardiac transplantation. *N Engl J Med* 2001; **344**: 1498–503.
- 127 Gustafson I, Lindblom A, Yun Z, *et al.* Quantification of adenovirus DNA in unrelated donor hematopoietic stem cell transplant recipients. *Journal of Clinical Virology* 2008; **43**: 79–85.
- 128 Avivi I, Chakrabarti S, Milligan DW, *et al.* Incidence and outcome of adenovirus disease in transplant recipients after reduced-intensity conditioning with alemtuzumab. *Biol Blood Marrow Transplant* 2004; **10**: 186–94.
- 129 Lion T, Kosulin K, Landlinger C, *et al.* Monitoring of adenovirus load in stool by real-

- time PCR permits early detection of impending invasive infection in patients after allogeneic stem cell transplantation. *Leukemia* 2010; **24**: 706–14.
- 130 Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel Cell Polyomavirus and Two Previously Unknown Polyomaviruses Are Chronically Shed from Human Skin. *Cell Host & Microbe* 2010; **7**: 509–15.
- 131 Kuypers DRJ. Management of polyomavirus-associated nephropathy in renal transplant recipients. *Nat Rev Nephrol* 2012; **8**: 390–402.
- 132 Babel N, Volk H-D, Reinke P. BK polyomavirus infection and nephropathy: the virus–immune system interplay. *Nat Rev Nephrol* 2011; **7**: 399–406.
- 133 Cortese I, Reich DS, Nath A. Progressive multifocal leukoencephalopathy and the spectrum of JC virus-related disease. *Nat Rev Neurol* 2021; **17**: 37–51.
- 134 Harms PW, Harms KL, Moore PS, *et al.* The biology and treatment of Merkel cell carcinoma: current understanding and research priorities. *Nat Rev Clin Oncol* 2018; **15**: 763–76.
- 135 Meijden E van der, Janssens RWA, Lauber C, Bavinck JNB, Gorbalenya AE, Feltkamp MCW. Discovery of a New Human Polyomavirus Associated with Trichodysplasia Spinulosa in an Immunocompromized Patient. *PLOS Pathogens* 2010; **6**: e1001024.
- 136 Ho J, Jedrych JJ, Feng H, *et al.* Human Polyomavirus 7–Associated Pruritic Rash and Viremia in Transplant Recipients. *The Journal of Infectious Diseases* 2015; **211**: 1560–5.
- 137 Wang Y, Metselaar HJ, Peppelenbosch MP, Pan Q. Chronic hepatitis E in solid-organ transplantation: the key implications of immunosuppressants. *Curr Opin Infect Dis* 2014; **27**: 303–8.
- 138 Dalton HR, Bendall RP, Keane FE, Tedder RS, Ijaz S. Persistent Carriage of Hepatitis E Virus in Patients with HIV Infection. *N Engl J Med* 2009; **361**: 1025–7.
- 139 Colson P, Dhiver C, Poizot-Martin I, Tamalet C, Gérolami R. Acute and chronic hepatitis E in patients infected with human immunodeficiency virus. *Journal of Viral Hepatitis* 2011; **18**: 227–8.
- 140 Kamar N, Izopet J, Pavio N, *et al.* Hepatitis E virus infection. *Nat Rev Dis Primers* 2017; **3**: 1–16.
- 141 Jones TPW, Buckland M, Breuer J, Lowe DM. Viral infection in primary antibody deficiency syndromes. *Reviews in Medical Virology* 2019; **29**: e2049.
- 142 Erard V, Chien JW, Kim HW, *et al.* Airflow Decline after Myeloablative Allogeneic Hematopoietic Cell Transplantation: The Role of Community Respiratory Viruses. *The Journal of Infectious Diseases* 2006; **193**: 1619–25.
- 143 Ison MG, Hirsch HH. Community-Acquired Respiratory Viruses in Transplant Patients: Diversity, Impact, Unmet Clinical Needs. *Clinical Microbiology Reviews* 2019; **32**: 10.1128/cmr.00042-19.
- 144 Moyes J, Walaza S, Pretorius M, *et al.* Respiratory syncytial virus in adults with severe acute respiratory illness in a high HIV prevalence setting. *J Infect* 2017; **75**: 346–55.
- 145 Baldolli A, Dargère S, Cardineau E, *et al.* Measles inclusion-body encephalitis (MIBE) in a immunocompromised patient. *J Clin Virol* 2016; **81**: 43–6.
- 146 Rodriguez C, Gouilh MA, Weiss N, *et al.* Fatal Measles Inclusion-Body Encephalitis in Adult with Untreated AIDS, France - Volume 26, Number 9—September 2020 - Emerging Infectious Diseases journal - CDC. DOI:10.3201/eid2609.200366.
- 147 Bearden D, Collett M, Quan PL, Costa-Carvalho BT, Sullivan KE. Enteroviruses in X-Linked Agammaglobulinemia: Update on Epidemiology and Therapy*. *The Journal of Allergy and Clinical Immunology: In Practice* 2016; **4**: 1059–65.
- 148 Khetsuriani N, Prevots DR, Quick L, *et al.* Persistence of vaccine-derived polioviruses among immunodeficient persons with vaccine-associated paralytic poliomyelitis. *J Infect Dis* 2003; **188**: 1845–52.

- 149 Rudge P, Webster AD, Revesz T, *et al.* Encephalomyelitis in primary hypogammaglobulinaemia. *Brain* 1996; **119** (Pt 1): 1–15.
- 150 Brown L-AK, Ruis C, Clark I, *et al.* A comprehensive characterization of chronic norovirus infection in immunodeficient hosts. *Journal of Allergy and Clinical Immunology* 2019; **144**: 1450–3.
- 151 Brown L-AK, Clark I, Brown JR, Breuer J, Lowe DM. Norovirus infection in primary immune deficiency. *Rev Med Virol* 2017; **27**: e1926.
- 152 Gras J, Abdel-Nabey M, Dupont A, Le Goff J, Molina J-M, Peraldi MN. Clinical characteristics, risk factors and outcome of severe Norovirus infection in kidney transplant patients: a case-control study. *BMC Infect Dis* 2021; **21**: 351.
- 153 van de Ven AAJM, Douma JW, Rademaker C, *et al.* Pleconaril-resistant chronic parechovirus-associated enteropathy in agammaglobulinaemia. *Antivir Ther* 2011; **16**: 611–4.
- 154 Mageau A, Ambert-Balay K, Boutolleau D, *et al.* Norovirus and sapovirus infections after allogeneic hematopoietic stem cell transplantation: is it worth it to look for them? *Leuk Lymphoma* 2023; **64**: 1295–303.
- 155 Jørgensen SF, Reims HM, Frydenlund D, *et al.* A Cross-Sectional Study of the Prevalence of Gastrointestinal Symptoms and Pathology in Patients With Common Variable Immunodeficiency. *Am J Gastroenterol* 2016; **111**: 1467–75.
- 156 Brown JR, Morfopoulou S, Hubb J, *et al.* Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. *Clin Infect Dis* 2015; **60**: 881–8.
- 157 Frémond M-L, Pérot P, Muth E, *et al.* Next-Generation Sequencing for Diagnosis and Tailored Therapy: A Case Report of Astrovirus-Associated Progressive Encephalitis. *J Pediatric Infect Dis Soc* 2015; **4**: e53–57.
- 158 Morfopoulou S, Brown JR, Davies EG, *et al.* Human Coronavirus OC43 Associated with Fatal Encephalitis. *N Engl J Med* 2016; **375**: 497–8.
- 159 Nilsson A, Edner N, Albert J, Ternhag A. Fatal encephalitis associated with coronavirus OC43 in an immunocompromised child. *Infect Dis (Lond)* 2020; **52**: 419–22.
- 160 Weiss RA, McMichael AJ. Social and environmental risk factors in the emergence of infectious diseases. *Nat Med* 2004; **10**: S70–6.
- 161 Baker RE, Mahmud AS, Miller IF, *et al.* Infectious disease in an era of global change. *Nat Rev Microbiol* 2022; **20**: 193–205.
- 162 Plowright RK, Parrish CR, McCallum H, *et al.* Pathways to zoonotic spillover. *Nat Rev Microbiol* 2017; **15**: 502–10.
- 163 Drzymalla E, Green RF, Knuth M, Khoury MJ, Dotson WD, Gundlapalli A. COVID-19-related health outcomes in people with primary immunodeficiency: A systematic review. *Clinical Immunology* 2022; **243**: 109097.
- 164 Buccioli G, Tangye SG, Meyts I. Coronavirus disease 2019 in patients with inborn errors of immunity: lessons learned. *Current Opinion in Pediatrics* 2021; **33**: 648.
- 165 Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. *N Engl J Med* 2009; **360**: 2605–15.
- 166 Waggoner JJ, Soda EA, Deresinski S. Rare and Emerging Viral Infections in Transplant Recipients. *Clinical Infectious Diseases* 2013; **57**: 1182–8.
- 167 Kumar D, Michaels MG, Morris MI, *et al.* Outcomes from pandemic influenza A H1N1 infection in recipients of solid-organ transplants: a multicentre cohort study. *The Lancet Infectious Diseases* 2010; **10**: 521–6.
- 168 Ljungman P, Camara R de la, Perez-Bercoff L, *et al.* Outcome of pandemic H1N1 infections in hematopoietic stem cell transplant recipients. *Haematologica* 2011; **96**: 1231–5.
- 169 Singer TG, Evankovich KD, Fisher K, Demmler-Harrison GJ, Risen SR. Coronavirus Infections in the Nervous System of Children: A Scoping Review Making the Case for Long-

- Term Neurodevelopmental Surveillance. *Pediatric Neurology* 2021; **117**: 47–63.
- 170 Collett MS, Hincks JR, Benschop K, *et al.* Antiviral Activity of Pocopavir in a Randomized, Blinded, Placebo-Controlled Human Oral Poliovirus Vaccine Challenge Model. *The Journal of Infectious Diseases* 2017; **215**: 335–43.
- 171 Lee N, Hurt AC. Neuraminidase inhibitor resistance in influenza: a clinical perspective. *Current Opinion in Infectious Diseases* 2018; **31**: 520.
- 172 Iketani S, Mohri H, Culbertson B, *et al.* Multiple pathways for SARS-CoV-2 resistance to nirmatrelvir. *Nature* 2023; **613**: 558–64.
- 173 Gandhi S, Klein J, Robertson AJ, *et al.* De novo emergence of a remdesivir resistance mutation during treatment of persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. *Nat Commun* 2022; **13**: 1547.
- 174 Pérot P, Bigot T, Temmam S, Regnault B, Eloit M. Microseek: A Protein-Based Metagenomic Pipeline for Virus Diagnostic and Discovery. *Viruses* 2022; **14**: 1990.
- 175 Fourgeaud J, Regnault B, Faury H, *et al.* Fetal Zika virus infection diagnosed by metagenomic next-generation sequencing of amniotic fluid. *Ultrasound in Obstetrics & Gynecology* 2023; **61**: 116–7.
- 176 Fourgeaud J, Lecuit MM, Pérot P, *et al.* Chronic Aichi Virus Infection As a Cause of Long-Lasting Multiorgan Involvement in Patients With Primary Immune Deficiencies. *Clinical Infectious Diseases* 2023; : ciad237.
- 177 Pérot P, Fourgeaud J, Rouzaud C, *et al.* Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France - Volume 29, Number 2—February 2023 - Emerging Infectious Diseases journal - CDC. DOI:10.3201/eid2902.221468.
- 178 Kitajima M, Hata A, Yamashita T, Haramoto E, Minagawa H, Katayama H. Development of a Reverse Transcription-Quantitative PCR System for Detection and Genotyping of Aichi Viruses in Clinical and Environmental Samples. *Appl Environ Microbiol* 2013; **79**: 3952–8.
- 179 Criscuolo A, Brisse S. AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* 2013; **102**: 500–6.
- 180 Lemoine F, Correia D, Lefort V, *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* 2019; **47**: W260–5.
- 181 Rivadulla E, Romalde JL. A Comprehensive Review on Human Aichi Virus. *Virol Sin* 2020; published online April 27. DOI:10.1007/s12250-020-00222-5.
- 182 Buccioli G, Moens L, Payne K, *et al.* Chronic Aichi Virus Infection in a Patient with X-Linked Agammaglobulinemia. *J Clin Immunol* 2018; **38**: 748–52.
- 183 Bekassy Z, Ehinger M, Pronk LN, Pronk CJ. Immunologic Control of Disseminated Aichi Virus Infection in X-Linked Agammaglobulinemia by Transplantation of TcR $\alpha\beta$ -Depleted Haploidentical Hematopoietic Cells. *J Clin Immunol* 2022; **42**: 1401–4.
- 184 Meyts I, Buccioli G, Jansen K, Wollants E, Breuer J. Aichivirus: an Emerging Pathogen in Patients with Primary and Secondary B-Cell Deficiency. *J Clin Immunol* 2023; **43**: 532–5.
- 185 Cunningham-Rundles C. Common variable immune deficiency: Dissection of the variable. *Immunol Rev* 2019; **287**: 145–61.
- 186 Fischer A, Provot J, Jais J-P, Alcais A, Mahlaoui N, members of the CEREDIH French PID study group. Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. *J Allergy Clin Immunol* 2017; **140**: 1388-1393.e8.
- 187 Sirisereewan C, Thanawongnuwech R, Kedkovid R. Current Understanding of the Pathogenesis of Porcine Circovirus 3. *Pathogens* 2022; **11**: 64.
- 188 Gillespie J, Opriessnig T, Meng XJ, Pelzer K, Buechner-Maxwell V. Porcine circovirus type 2 and porcine circovirus-associated disease. *J Vet Intern Med* 2009; **23**: 1151–63.
- 189 Madson DM, Patterson AR, Ramamoorthy S, Pal N, Meng XJ, Opriessnig T.

- Reproductive failure experimentally induced in sows via artificial insemination with semen spiked with porcine circovirus type 2. *Vet Pathol* 2009; **46**: 707–16.
- 190 Opriessnig T, Patterson AR, Elsener J, Meng XJ, Halbur PG. Influence of maternal antibodies on efficacy of porcine circovirus type 2 (PCV2) vaccination to protect pigs from experimental infection with PCV2. *Clin Vaccine Immunol* 2008; **15**: 397–401.
- 191 Opriessnig T, Ramamoorthy S, Madson DM, *et al.* Differences in virulence among porcine circovirus type 2 isolates are unrelated to cluster type 2a or 2b and prior infection provides heterologous protection. *J Gen Virol* 2008; **89**: 2482–91.
- 192 Regnault B, Evrard B, Plu I, *et al.* First Case of Lethal Encephalitis in Western Europe Due to European Bat Lyssavirus Type 1. *Clin Infect Dis* 2022; **74**: 461–6.
- 193 Riller Q, Fourgeaud J, Bruneau J, *et al.* Late-onset enteric virus infection associated with hepatitis (EVAH) in transplanted SCID patients. *Journal of Allergy and Clinical Immunology* 2023; **0**. DOI:10.1016/j.jaci.2022.12.822.
- 194 Wanat KA, Perelygina L, Chen M, *et al.* Association of Persistent Rubella Virus With Idiopathic Skin Granulomas in Clinically Immunocompetent Adults. *JAMA Dermatology* 2022; **158**: 626–33.
- 195 Naymagon S, Naymagon L, Wong S-Y, *et al.* Acute graft-versus-host disease of the gut: considerations for the gastroenterologist. *Nat Rev Gastroenterol Hepatol* 2017; **14**: 711–26.
- 196 Morfopoulou S, Buddle S, Torres Montaguth OE, *et al.* Genomic investigations of unexplained acute hepatitis in children. *Nature* 2023; **617**: 564–73.
- 197 Stevaert A, Groaz E, Naesens L. Nucleoside analogs for management of respiratory virus infections: mechanism of action and clinical efficacy. *Current Opinion in Virology* 2022; **57**: 101279.
- 198 Brown AJ, Won JJ, Graham RL, *et al.* Broad spectrum antiviral remdesivir inhibits human endemic and zoonotic deltacoronaviruses with a highly divergent RNA dependent RNA polymerase. *Antiviral Research* 2019; **169**: 104541.
- 199 Li J, Wang Y, Solanki K, *et al.* Nirmatrelvir exerts distinct antiviral potency against different human coronaviruses. *Antiviral Research* 2023; **211**: 105555.
- 200 van Boheemen S, de Graaf M, Lauber C, *et al.* Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* 2012; **3**: e00473-12.
- 201 Greninger AL. A decade of RNA virus metagenomics is (not) enough. *Virus Res* 2018; **244**: 218–29.
- 202 Valyraki N, Maillart E, Pourcher V, *et al.* Human pegivirus identified in severe myelitis and optic neuritis in immunocompromised patients: A pathogenic role for a forgotten virus? *Rev Neurol (Paris)* 2023; **179**: 361–7.
- 203 Doan MAL, Roczakowsky A, Smith M, *et al.* Infection of Glia by Human Pegivirus Suppresses Peroxisomal and Antiviral Signaling Pathways. *Journal of Virology* 2021; **95**: 10.1128/jvi.01074-21.
- 204 Balcom EF, Doan MAL, Branton WG, *et al.* Human pegivirus-1 associated leukoencephalitis: Clinical and molecular features. *Ann Neurol* 2018; **84**: 781–7.

Table des matières

REMERCIEMENTS	3
SOMMAIRE	5
LISTE DES PRINCIPALES ABREVIATIONS.....	7
INTRODUCTION.....	9
I. CLASSIFICATION DES VIRUS	11
II. GENOMES VIRAUX.....	15
A. Génomomes à ADN et ARN.....	15
B. Stratégies des virus à ADN.....	16
C. Stratégies des virus à ARN.....	17
D. Implications pour la recherche de cible de détection moléculaire	20
III. PATHOGENESE VIRALE ET ORGANES CIBLES	20
IV. BIOLOGIE MOLECULAIRE EN VIROLOGIE MEDICALE	23
A. Détection et quantification du génome par biologie moléculaire.....	23
B. Caractérisation du génome par séquençage à haut débit	24
1. Les techniques de séquençages à haut débit	24
2. Analyses ciblées	25
a) Séquençage d'amplicons de PCR	25
b) Séquençage après enrichissement par capture	26
3. Analyse non ciblée : métagénomique NGS.....	27
C. Application des tests moléculaire à la virologie	30
V. VIROME HUMAIN.....	31
VI. CRITERES D'IMPUTABILITE	35
VII. INFECTIONS VIRALES OPPORTUNISTES	37
VIII. VIRUS EMERGENTS ET DEFICITS IMMUNITAIRES.....	39
CONTEXTE ET OBJECTIFS	42
CONTEXTE	43
OBJECTIFS	44
MATERIEL ET METHODES	46
I. METAGENOMIQUE ET METATRANSCRIPTOMIQUE	47
A. Technique	47
B. Contrôles de qualité.....	50
C. Interprétation des résultats.....	50
II. PCR ET RT-PCR SPECIFIQUES	52
A. RT-PCR Aichi virus.....	52
B. PCR Circovirus.....	52
III. HYBRIDATION <i>IN SITU</i>	52
A. Hybridation <i>in situ</i> Aichi virus.....	52
B. Hybridation <i>in situ</i> Human circovirus 1	54
IV. ANALYSES PHYLOGENETIQUES	54
A. Aichi virus.....	54
B. Human circovirus 1.....	55
RESULTATS.....	56
PARTIE 1 : IDENTIFICATION DE VIRUS IMPLIQUES DANS LES PATHOLOGIES INEXPLIQUEES DU PATIENT IMMUNODEPRIME	57
Article 1: <i>Chronic Aichi virus infection as a cause of long-lasting multi-organ involvement in patients with primary immune deficiencies</i>	57
Article 2: <i>Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France</i>	58
PARTIE 2 : IDENTIFIER DES INDICATIONS POUR LE DIAGNOSTIC PAR METAGENOMIQUE CLINIQUE.....	59
Article 3: <i>Performance of clinical metagenomics: an observational prospective study in France</i>	59

DISCUSSION ET PERSPECTIVE	60
DISCUSSION.....	61
CONTRIBUTION ET CONCLUSION GENERALE	67
PERSPECTIVE.....	68
REFERENCES.....	69
TABLE DES MATIERES.....	79
TABLE DES ILLUSTRATIONS.....	81
TABLE DES TABLEAUX	82
ANNEXES.....	83
ANNEXE 1 : MATERIEL SUPPLEMENTAIRE ARTICLE 1	84
ANNEXE 2 : MATERIEL SUPPLEMENTAIRE ARTICLE 2	94
ANNEXE 3 : MATÉRIEL SUPPLÉMENTAIRE ARTICLE 3	97

Table des illustrations

FIGURE 1. COMPARAISON DE LA HIERARCHIE DES RANGS TAXONOMIQUES DE L'ICTV EN 1991-2017 ET 2019.....	13
FIGURE 2. SEPT CATEGORIES DE VIRUS SE DISTINGUANT PAR LEURS STRATEGIES DE REPLICATION DU GENOME.....	15
FIGURE 3. CYCLE LYTIQUE DU VIRUS ZONA VARICELLE (HERPESVIRIDAE) LORS DE L'INFECTION D'UN NEURONE.	16
FIGURE 4. RESUME SIMPLIFIE DES CYCLES DES VIRUS A ARNDB, DES VIRUS A ARN (+) ET DES RETROVIRUS.....	19
FIGURE 5. ILLUSTRATION DU CYCLE DES ARBOVIRUS DANS L'ORGANISME ET DES DIFFERENTES PHASES DE L'INFECTION	22
FIGURE 6. METHODES DE SEQUENÇAGE DES GENOMES VIRAUX A PARTIR D'ECHANTILLONS CLINIQUES	27
FIGURE 7. APPROCHE NON CIBLEE DE METAGENOMIQUE NGS DITE "SHOTGUN"	29
FIGURE 8. EXEMPLE DE PIPELINE BIO-INFORMATIQUE POUR L'ANALYSE DES SEQUENCES DE METAGENOMIQUE.....	30
FIGURE 9. RESUME DES VIRUS PRESENTS SUR CHAQUE SITE DU CORPS HUMAIN	32
FIGURE 10. FACTEURS INFLUENÇANT LA PROPAGATION D'UN AGENT PATHOGENE ZOOTIQUE JUSQU'A L'INFECTION DE L'ETRE HUMAIN	40
FIGURE 11. LOGIGRAMME RESUMANT LES PROCEDURES D'ANALYSE DE METAGENOMIQUE ET METATRANSCRIPTOMIQUE	48
FIGURE 12. LOGIGRAMME RESUMANT LES ETAPES DU PIPELINE BIO-INFORMATIQUE MICROSEEK.	49
FIGURE 13. LOGIGRAMME D'INTERPRETATION DES RESULTATS DE METAGENOMIQUE	51
FIGURE 14. CONTROLES NEGATIFS ET POSITIF DE L'HYBRIDATION IN SITU AICHI VIRUS.....	53

Table des tableaux

TABLEAU 1. LES CLASSES DE VIRUS DE BALTIMORE ET LEURS PRINCIPALES CARACTERISTIQUES MOLECULAIRES ET BIOLOGIQUES	14
TABLEAU 2. EXEMPLES D'ALTERATIONS DE POPULATIONS VIRALES DANS DES PATHOLOGIES HUMAINES.....	34

Annexes

Annexe 1 : Matériel supplémentaire Article 1

Title: Chronic *Aichi virus* infection as a cause of long-lasting multi-organ involvement in patients with primary immune deficiencies

Supplementary Data

Hematopoietic stem cell transplantation (HSCT) and conditioning regimen (CR) of patient 1 (P1) and patient 5 (P5)

P1 received a TCRab/CD19 depleted graft from his haplo-identical father preceded by a CR consisting in alemtuzumab (0.5 mg/kg/day on D-16 to D-15 and D-8), treosufan 42 g/kg, fludarabine 160 mg/m² and thiotepa 10 mg/kg. Primary graft failure required a second transplant on D+43 from his haplo-identical mother preceded by alemtuzumab and fludarabine. Neutrophil and thrombocyte recoveries were reached at D+15 and D+19 respectively with a full stable donor engraftment.

Twelve months after diagnosis of AiV infection, P5 received a geno-identical transplant preceded by CR consisting in alemtuzumab (0.5 mg/kg/day on D-9 to D-10), treosufan 42 g/kg, fludarabine 160 mg/m² and thiotepa 10 mg/kg that allowed a full donor engraftment.

Table S1. Characteristics of the six tissue samples of primary immunodeficient patients analyzed by metagenomic Next Generation Sequencing and/or *Aichi virus* RT-PCR. FFPE= Formalin-Fixed Paraffin-Embedded

Patients	Date	Type of biopsy	Million raw reads	Number of reads assigned to <i>Aichi virus</i>	Coverage of AB010145 (%)	Percent identity with AB010145 (%)	<i>Aichi virus</i> RT-PCR cycle threshold
P1	5/14/2020	liver	64	169	41.8	98.5	27.0
P1	5/14/2020	FFPE spleen	not tested	not tested	not tested	not tested	36.7
P2	4/4/2014	FFPE liver	65	0	NA	NA	negative
P3	6/28/2012	liver	106	18	8.5	93.1	36.5
P3	8/20/2015	kidney	111	26	18.2	88.5	32.7
P4	10/25/2013	liver	133	4	3.0	92.8	35.1
P5	9/16/2020	spleen	102	29	31.7	90.5	27.8
P5	9/16/2020	lymph node	104	4	2.4	94.9	36.1
P6	12/9/2003	liver	26	1	1.7	90.1	negative
P6	1/3/2005	liver	80	2	3.3	86.2	41.3

Table S2. Patient 1 results over time, following the diagnosis of Aichi virus infection in May 2020 and the hematopoietic stem cell transplantation (HSCT) in March 2021. FFPE= Formalin-Fixed Paraffin-Embedded

Collection date	Sample type	<i>Aichi virus</i> RT-PCR cycle threshold	<i>Aichi virus</i> RT-PCR results	Number of reads assigned to <i>Aichi virus</i>
5/5/20	stool	42.2	Positive	
5/5/20	plasma	39.7	Positive	
5/5/20	bone marrow aspiration	0	Negative	
5/7/20	cerebrospinal fluid	0	Negative	
5/14/20	liver	27.0	Positive	169
5/14/20	FFPE spleen	36.7	Positive	
6/4/20	pleural fluid	0	Negative	
6/11/20	plasma	0	Negative	
6/11/20	bronchoalveolar fluid	0	Negative	
7/20/20	plasma	0	Negative	
8/17/20	plasma	0	Negative	
9/21/20	urine	0	Negative	
9/21/20	plasma	0	Negative	
9/21/20	stool	0	Negative	
11/19/20	plasma	34.6	Positive	
11/30/20	plasma	35.9	Positive	
12/10/20	urine	32.9	Positive	
12/10/20	plasma	41.7	Positive	
1/19/21	plasma	35.4	Positive	
1/19/21	urine	33.9	Positive	
1/22/21	stool	0	Negative	
1/28/21	plasma	33.2	Positive	
2/4/21	HSCT N°1			
2/4/21	plasma	0	Negative	
3/4/21	Bone marrow aspiration	0	Negative	
3/15/21	stool	0	Negative	
3/19/21	HSCT N°2			
4/12/21	urine	39.8	Positive	
4/13/21	stool	0	Negative	

4/26/21	nasopharyngeal swab	0	Negative	
4/30/21	stool	0	Negative	
5/3/21	plasma	0	Negative	
5/4/21	stool	0	Negative	
5/7/21	stool	0	Negative	
5/21/21	stool	0	Negative	
5/21/21	urine	0	Negative	
5/21/21	plasma	0	Negative	
6/3/21	urine	0	Negative	
6/3/21	plasma	0	Negative	
6/24/21	plasma	0	Negative	
6/24/21	urine	0	Negative	
10/5/21	stool	0	Negative	
10/5/21	urine	0	Negative	
10/5/21	plasma	0	Negative	
10/26/21	plasma	0	Negative	
12/7/21	urine	0	Negative	
12/7/21	plasma	0	Negative	
12/28/21	urine	0	Negative	
12/28/21	plasma	0	Negative	
1/18/22	plasma	0	Negative	
2/8/22	urine	0	Negative	
3/15/22	urine	0	Negative	
3/15/22	plasma	0	Negative	
4/4/22	urine	0	Negative	
4/4/22	plasma	0	Negative	
4/21/22	urine	0	Negative	
4/21/22	plasma	0	Negative	
5/16/22	stool	0	Negative	
6/7/22	plasma	0	Negative	
7/18/22	plasma	0	Negative	
7/18/22	urine	0	Negative	

Table S3. Patient 2 results over time, following the diagnosis of *Aichi virus* infection with metagenomic Next Generation Sequencing. FFPE= Formalin-Fixed Paraffin-Embedded

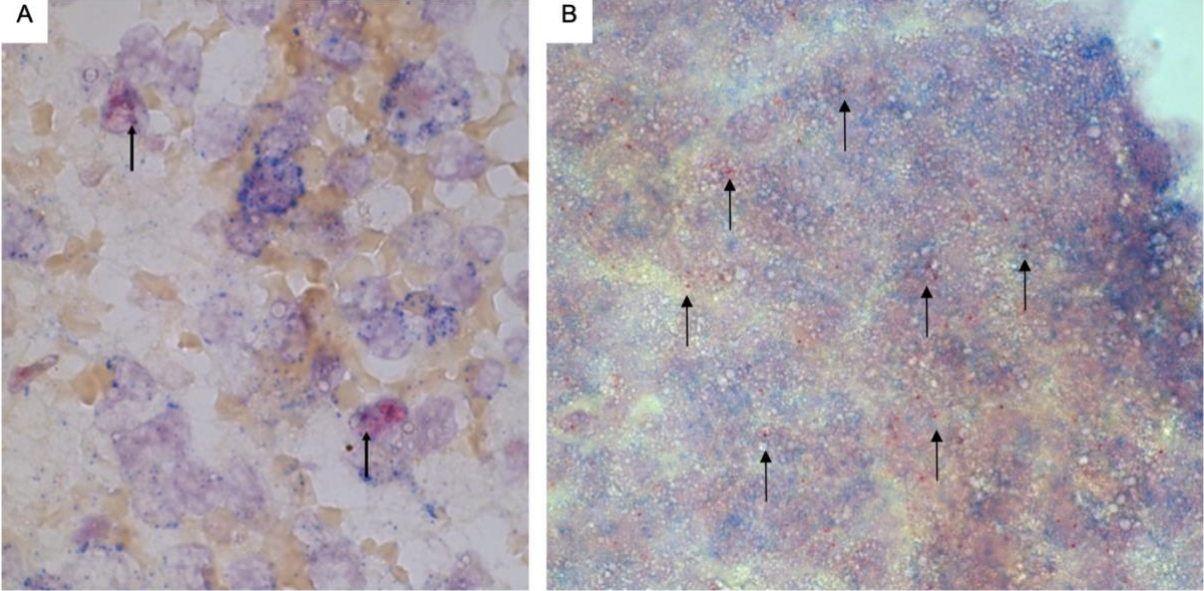
Collection date	Sample type	Aichi virus RT-PCR cycle threshold	Aichi virus RT-PCR results	Number of reads assigned to <i>Aichi virus</i>
4/4/14	FFPE liver	0	Negative	0
8/25/21	urine	37.7	Positive	
8/25/21	plasma	0	Negative	
8/25/21	stool	0	Negative	
11/4/21	urine	37.9	Positive	
11/4/21	stool	0	Negative	
11/4/21	cerebrospinal fluid	0	Negative	
11/5/21	plasma	40.0	Positive	
8/30/22	stool	0	Negative	

Table S4. Patient 5 RT-PCR results over time, following the diagnosis of *Aichi virus* infection with metagenomic Next Generation Sequencing in September 2020 and the hematopoietic stem cell transplantation (HSCT) in September 2021.

Collection date	Sample type	<i>Aichi virus</i> RTPCR cycle threshold	<i>Aichi virus</i> RTPCR result	Number of mNGS reads assigned to <i>Aichi virus</i>
9/16/20	spleen	27.8	Positive	29
9/16/20	lymph node	36.1	Positive	4
8/17/21	plasma	38.8	Positive	
8/17/21	urine	0	Negative	
8/24/21	stool	0	Negative	
8/25/21	liver	40	Positive	
9/7/21	HSCT			
9/20/21	plasma	0	Negative	
9/28/21	stool	0	Negative	
10/7/21	plasma	0	Negative	
10/7/21	urine	0	Negative	
10/8/21	stool	0	Negative	
10/13/21	stool	0	Negative	
10/13/21	plasma	0	Negative	
10/13/21	urine	0	Negative	
10/18/21	plasma	0	Negative	
10/18/21	urine	0	Negative	
10/20/21	stool	0	Negative	
10/25/21	plasma	37.8	Positive	
10/25/21	urine	0	Negative	
11/4/21	stool	0	Negative	
11/7/21	stool	0	Negative	
11/19/21	stool	0	Negative	
12/7/21	stool	0	Negative	
1/11/22	urine	0	Negative	
1/13/22	plasma	41	Positive	
1/13/22	stool	0	Negative	
1/22/22	stool	0	Negative	
1/24/22	plasma	0	Negative	
1/24/22	bone marrow aspiration	0	Negative	

2/5/22	stool	0	Negative	
2/10/22	plasma	0	Negative	
2/17/22	stool	0	Negative	
2/17/22	plasma	0	Negative	
2/17/22	urine	0	Negative	
2/21/22	stool	0	Negative	
2/21/22	plasma	0	Negative	
2/24/22	plasma	0	Negative	
2/28/22	stool	0	Negative	
3/3/22	plasma	0	Negative	
3/8/22	plasma	0	Negative	
3/10/22	stool	0	Negative	
3/10/22	urine	0	Negative	
3/21/22	liver	0	Negative	
4/11/22	stool	0	Negative	

Figure S1. Chromogenic In situ Hybridization detection of *Aichi virus* RNA (labeled with red) in spleen biopsy of patient 1 (A) and patient 5 (B) (Gx400). Human Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Beta Actin (ACTB) and pyrophosphate (PPI) transcripts were labeled with blue (Gx400).



Supplementary methods

AiV RT-PCR

Nucleic acids were extracted from 200 μ L of plasma or urine and from 400 μ L of preprocessing stool using the Emag instrument (Biomérieux). Extracted RNA (5 μ l) was added to 20 μ l of reaction mixture containing SuperScript III reagent (Invitrogen, Asnières-sur-Seine, France), PCR grade water, probes at a final concentration of 200nM and primers at a final concentration of 400nM. PCR amplification was performed with an LC480 detection system (Roche Diagnostic, Meylan, France), reverse transcription at 50°C for 15 min, initial denaturation at 95°C for 2 min, followed by 45 cycles of amplification with denaturation at 95°C for 15 s and hybridization/elongation at 55°C for 1 min.

The RT-PCR assays in asymptomatic and symptomatic groups were performed under the same conditions and in the same batches at the time of inclusion. The individuals performing the RT-PCR assay were blinded to the groups and symptoms.

PCR grade water was used in each RT-PCR batch as negative control. Positive AiV stool provided by the National Reference Center for Gastroenteritis Viruses (Dijon, France) was used as a positive control in each RT-PCR batch.

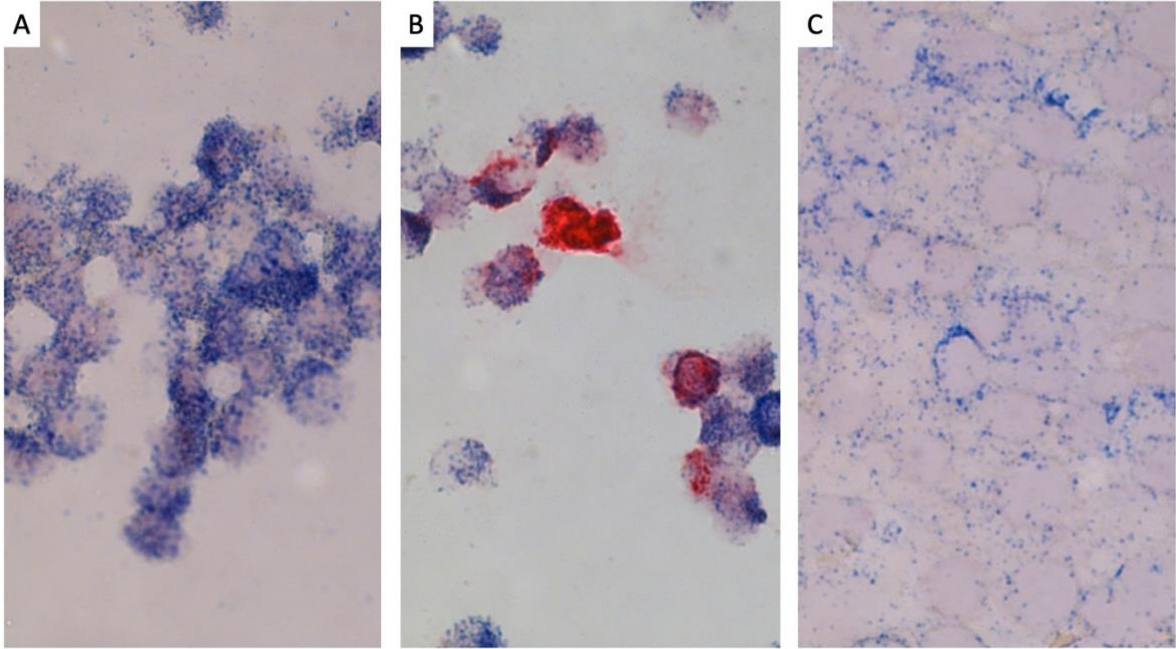
In situ Hybridization (ISH)

To unmask RNA targets, the tissue sections (after deparaffinization using xylene) were treated 20 min at 95°C and 15 min with protease.

ISH Negative and positive controls

Three experiments were performed on 3 samples (n=1 liver; n=2 spleen) from 2 patients (P1 and P5). In these 3 successive experiments, two negative controls were used in each ISH batch: the first negative control consisted of AiV positive liver biopsy treated without AiV specific probe, the second negative control was an AiV negative liver biopsy in mNGS. Of note the negative liver biopsy was collected from different patients in each experiment. Uninfected vero E6 cells were also used as a negative control. AiV (genotype A) infected Vero E6 cells provided by the National Reference Center for Gastroenteritis Viruses (Dijon, France) was used as a positive control in each experiment (Figure S2).

Figure S2. Chromogenic In situ Hybridization detection of Aichi virus RNA (labeled with red) in negative control vero E6 cells (A), in AiV infected vero E6 cells (B) and in negative control liver biopsy (C). Human Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Beta Actin (ACTB) and pyrophosphate (PPI) transcripts were labeled with blue (Gx400).



Annexe 2 : Matériel supplémentaire Article 2

Title : Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France

Appendix Table 1. Control characteristics for assessment of human circovirus type 1 (HCirV-1) presence in 36 blood and 20 stool samples performed with specific HCirV-1 qPCR*

Controls	Hepatitis†(n=16)	Immunocompetent, no hepatitis (n=12)	Immunocompromised, no hepatitis (n=28)	Total (n = 56)
Age, mean (range), y	17 (3–53)	30 (0.2–67)	44 (1–84)	33 (0.2–84)
Adults >15 years old, n (%)	5 (31)	7 (58)	22 (78)	35 (63)
Sex, female, n (%)	4 (25)	7 (58)	11 (39)	22 (39)
Samples tested, n (%)				
Blood	10 (62)	8 (66)	18 (64)	36 (64)
Stool	6 (38)	NA	NA	NA
Known hepatitis etiology‡	8 (50)	NA	NA	NA
Immunological status, n (%)				
Immunocompetent	5 (31)	12 (100)	0	17 (30)
HIV seropositive	0	0	2 (7)	2 (4)
Solid organ transplant	1 (6)	0	12 (43)	13 (23)
Primary immune deficiency	10 (63)	0	9 (32)	19 (34)
HSCT or immunosuppressive therapy	0	0	5 (18)	5 (9)

*HSCT, hematopoietic stem cell transplant; NA, not applicable

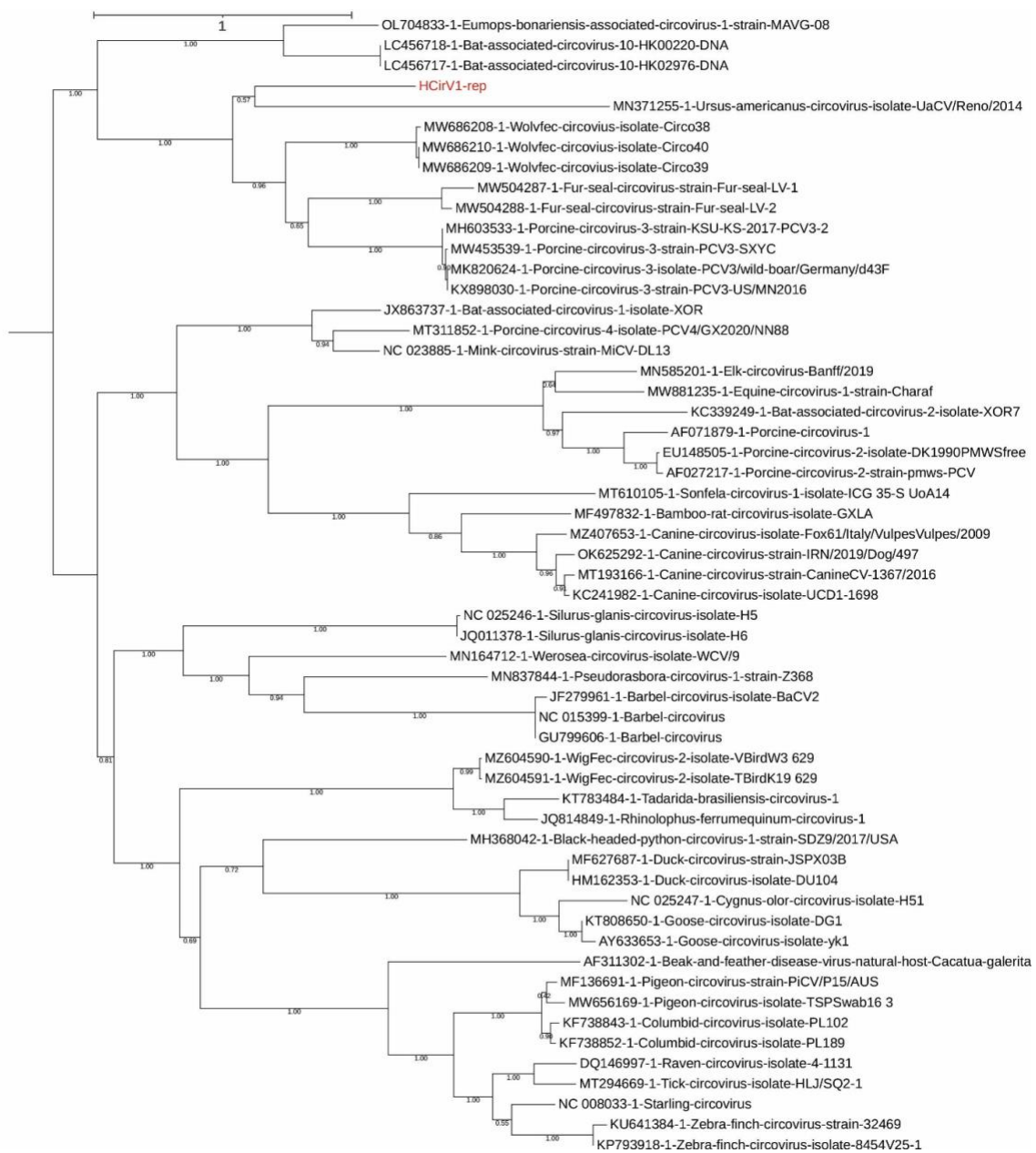
†hepatomegaly or liver cytolysis

‡Infectious (4 viral, 1 bacterial), dysimmune (1), metabolic (1) and toxic (1) etiologies

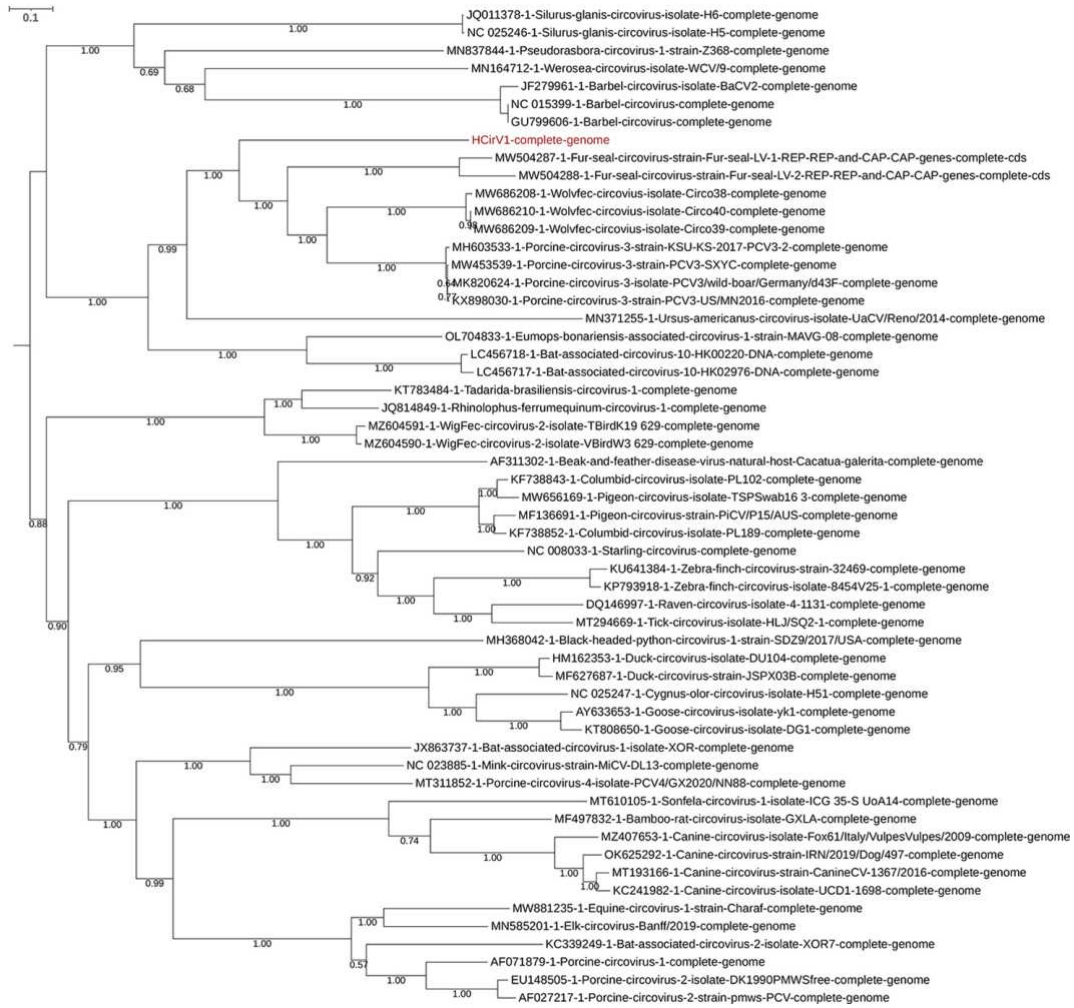
Appendix Table 2. Control characteristics for assessment of human circovirus type 1 (HCirV-1) presence in 57 liver biopsies samples screened with metagenomic next generation sequencing

Controls	Total (n=57)
Age, y, mean (range)	13 (0.3–68)
Adults >15 y, n (%)	10 (18)
Sex, female, n (%)	20 (35)
Immunological status, n (%)	
Immunocompetent	5 (9)
HIV seropositive	2 (4)
Solid organ transplant	5 (9)
Primary immune deficiency	25 (44)
HSCT or immunosuppressive therapy	20 (35)

*HSCT, hematopoietic stem cell transplant



Appendix Figure 1. Phylogenetic analysis of human circovirus type 1 (HCrV-1). Phylogenetic analysis of replicase (Rep) protein sequences of HCrV-1 and representative of circovirus strains. Sequences were aligned with MAFFT under the L-INS-I parameter, and maximum-likelihood phylogenetic reconstruction was performed with PhyML implemented through the NGPhylogeny portal. HCrV-1 protein sequence is depicted in red. The tree scale indicates the number of amino acid substitutions per site.



Appendix Figure 2. Phylogenetic analysis of human circovirus type 1 (HcIrV-1). Phylogenetic analysis of complete nucleotide sequences of HcIrV-1 and representative of circovirus strains. Sequences were aligned with MAFFT under the L-INS-I parameter, and maximum-likelihood phylogenetic reconstruction was performed with PhyML implemented through the NGPhylogeny portal. HcIrV-1 complete genome is depicted in red. The tree scale indicates the number of amino acid substitutions per site.

Annexe 3 : Matériel supplémentaire Article 3

Title: Performance of clinical metagenomics: an observational prospective study in France

Patients

First-line SMT include culture of fast-growing bacteria, broad range PCR (16S, 16-23S, and 18S ribosomal RNA genes for bacteria, mycobacteria, and fungi respectively) as well as routine syndromic multiplex PCR panels (meningo-encephalitis, respiratory, gastrointestinal) and specific viral PCR depending on the symptoms and immune status (such as *cytomegalovirus*, *Epstein-Barr virus*, *adenovirus*, BK virus in immunocompromised patients). All the required first-line tests are listed in Table S2.

The location of healthcare facilities sending samples was distributed as follows: 643 samples were sent from six healthcare facilities in the Paris region. Of these, 97% (622/643 samples) came from 12 hospitals belonging to the Greater Paris University Hospitals (AP-HP). Ninety-nine samples were obtained from 25 healthcare facilities located in other regions of France or in other countries.

"Life-threatening situations" involved patients hospitalized in intensive care units. The level of urgency was defined on the basis of the prescribing physician's assessment.

Sample processing and limit of detection (LOD)

For tissues, total RNAs is used as a signature of viral, bacterial or fungal gene expression, which is a hallmark of active infection. Even in the case of a DNA virus infection, transcripts remain mandatory intermediates of viral protein expression, meaning that a transcriptomic approach (total RNA) is appropriate for the detection of both replicating RNA or DNA viruses in a tissular context, and more largely also for any bacterial or fungal pathogen having a transcriptional activity. Additional DNase treatment is applied to the RNA extract to remove as much residual DNA as possible, before library construction, to avoid human genomic DNA sequencing which would drastically reduce the sensitivity.

For fluids (plasma, CSF, urine...), the logic is different and aims at detecting free viral particles (DNA or RNA viruses), or bacterial bodies, typically in a context of viremia, pulmonary infection, or sepsis. Thus, a DNA/RNA approach is used, which for certain samples can be coupled with an upstream treatment with nucleases to remove free unprotected nucleic acids to increase the sensitivity.

Briefly, DNase-treated total RNA from tissue and blood samples was isolated after a bead-beating step and subjected to transcriptome sequencing. Total nucleic acids were isolated from fluid samples with (BALF, plasma, stool supernatant) or without (CSF, urines) a pre-digestion of free-nucleic acids using a cocktail of nucleases.

All the procedures for sample processing are summarized in Figure S2 and Figure S3. The list of reagents is provided in supplementary table S4. The kits were used in accordance with the manufacturers' protocols.

Quality assurance included the addition of 20 µl of ZymoBIOMICS Spike-In Control II (Zymo Research, Irvine, California) per sample prior to the extraction step as an internal control. Quality was considered adequate when the dosage of the library was >0.5 µM, >500 spike-in control sequences were detected and >15 million reads per sample were produced. The limit of detection and specificity were evaluated experimentally with blind controls coming from the External Quality Assessment (EQA) / Proficiency Testing (PT) organization to assess overall performance of protocols based on the qualitative identification of microorganisms present in the samples, at the family, genus, species, and subtype levels. We used the Viral Metagenomics NGS EQA Pilot Study panel (QAV204213QCMD), in which samples mimicking CSF were spiked with various human viruses harboring a DNA or RNA genome, at various concentrations. We succeeded in detecting virus species down to 1.4 log copies/ml and were among the few laboratories able to detect all viruses and subtypes, with no specificity error in the case of polymicrobial controls.

The LOD of non-viral microbial sequences was also challenged using the ZymoBIOMICS Microbial Community Standard (Cat# D6300) consisting of eight bacteria (three Gram-negative and five Gram-positive) and two yeasts. All 10 species were successfully detected in serial dilution down to 10³ cells/ml.

In addition, we have benchmarked the LOD of our bioinformatic pipeline called "Microseek" using spiked datasets,¹ showing that in complex matrices with a high background of non-microbial sequences, as low as one read spiked in a 50 million reads dataset can be detected. Microseek also succeeded in detecting viruses starting from 5 spiked sequences in a 50 million reads dataset where other software did not. Microseek also succeeded in detecting viruses starting from 5 spiked sequences in a 50 million reads dataset where other software did not.

We recently published a case of Fetal Zika virus infection diagnosed by metagenomic next-generation sequencing from only 3 reads, after initial negative Zika PCR, then later confirmed with a more sensitive Zika PCR.² In earlier works, we detected very few sequences (<10 reads) of rubella virus vaccine strain sequences from granulomatous lesions, then confirmed by sensitive nested-PCR.^{3,4} These examples indicate that the LOD of our NGS approach is at least comparable to that of sensitive specific PCRs.

Microorganism classification, confirmatory tests and reports

The term "causative" refers to the detection of a pathogen that is well described in the medical literature as causing similar clinical presentations. The term "possibly causative" refers to either (i) the detection of a microorganism for which an association with the symptoms has been described in the literature without formal proof of causality, or (ii) the detection of a known pathogen in an atypical clinical presentation. To support putative novel pathogenic associations, we confirmed microorganism detection in tissue by *in situ* hybridization for Aichi virus and circovirus in liver biopsies. When quantitative techniques were available, we also used

quantification to support causation (comparison of blood load compared to CSF for human pegivirus).

In our study, stools were collected from 6 immunocompetent patients and 46 immunocompromised patients. Of the stools identified with a causative or possibly causative pathogen, 96% (23/24) were collected from patients with immune deficiency. We classified stool pathogens in the “causative” category when we identified viruses known to be digestive pathogens (noroviruses, *sapovirus*, *C. jejuni*, *S. enterica*) in the stool of patients with acute or chronic digestive disease (50%; 12/24) with no other etiology found. We also classified as “causative” certain viruses known to be pathogens (enteroviruses, astroviruses, parechoviruses, noroviruses) in patients with unexplained hepatitis and primary immune deficiencies.⁵ We classified as “possibly causative” two co-infections with adenovirus and norovirus in the stools of a patient with a post-infectious encephalitis picture and in those of another patient presenting with a primary immune deficiency and an aseptic lymphocytic meningitis, with no other microorganisms detected in blood or CSF. We classified as “possibly causative” 2 parechoviruses and 1 *rhinovirus* in poorly understood inflammatory and hematological disorders in 3 patients with primary immune deficiencies. Finally, as the pathogenic role of *picobirnavirus* is uncertain, we classified it as “possibly causative” in 4 patients with immune deficiency and digestive (n=2) or hepatic (n=2) symptoms with no other microorganisms or etiology found.

Human pegivirus is usually regarded as non-pathogenic. However, it has been associated with neurological and visual deficits albeit without formal proof to date.⁶⁻⁸ When the clinical presentation was similar to the presentations described in the mentioned literature and in the absence of any other etiologies, we classified Human pegivirus as “possibly causative”. Taking into account this literature, the severe and poorly understood neurological cases, the activity of some direct active antivirals against human pegivirus,⁹ and the uncertain role of this virus in the course of the disease, we reported the presence of a Human pegivirus as “possibly causative” and referred the clinician to the published literature.

Confirmatory tests were systematically carried out when a pathogenic microorganism was identified, regardless of the number of reads (Table S3). Exceptions were insufficient remaining sample or concomitant detection of the same pathogen in another sample from the same patient. The reports were prepared in a collegial manner during weekly meetings by a multidisciplinary team of virologists, bacteriologists and mycologists who compared the results of the mNGS, the results of conventional and confirmatory tests, and the medical records of each patient. In case of positive concordance between mNGS and confirmatory tests, results were provided accordingly, while in case of discordance, arbitration was performed on the basis of all the available evidence including clinical outcome if known. If necessary, physicians were guided to perform additional tests, such as other confirmatory tests (typically qPCR targeting the pathogen identified to confirm its presence and evaluate its load). A flowchart summarizing the categorization is presented in supplementary figure S1.

The reports provided to the physicians contained the following information: viral species identified, bacterial species or type of commensal flora identified, fungi/parasites identified and a conclusion highlighting only the pathogenic microorganisms.

Statistical analysis

A multivariable logistic regression model was developed to compare the rate of causative pathogen detection in each group. First, the candidate variables were discussed with the scientific committee in order to agree a consistent hypothesis to be tested in a multivariate analysis. Gender (male, female), age (infants, adolescents, adults), immune status (immunocompetent, primary immune deficiency, secondary immune deficiency, HIV), initial suspicion of infection (high, low), indications (all indications types), and sample type (all sample types) were examined for input into a multivariable model to determine the determinant of the positivity. To analyze age data according to the National institutes of Health (US Department of Health), we decided to set up three age groups with sufficient size: a first group with newborns (birth to 1 month), infants (1 month to 1 year) and children (1 year through 12 years), a second group of adolescents (13 years through 17 years) and a third group of adults (18 years or older). To analyze the symptom data, we analyzed the symptoms that prompted the analysis (inflammatory, cardiac, skin, digestive, fetal, hematological, liver, nephrological, neurological, ocular, otorhinological, osteo muscular, respiratory) and included in the model the 6 indications with the highest positivity rate and size greater than n=20 patients (neurological, respiratory, digestive, inflammatory, hematological, liver). The other indication groups were pooled and used as a reference. Given the great diversity of sample types, we first analyzed tissue samples, then fluids, and finally compared the positivity in central nervous system tissue versus cerebrospinal fluids, and in digestive tissue versus stool. To analyze the tissue samples, we selected the tissue types relevant to the major indications (CNS, lymph nodes, liver, digestive tract). The other tissue types were pooled and used as a reference for the analysis. To analyze biological fluids, and given their great diversity, we chose to differentiate between sterile and non-sterile samples, and also to differentiate between stools, which had a high positivity rate, and CSF, with a low positivity rate and high number of samples. Sterile fluids were used as a reference for the analysis. For all these analyses, we used the same covariates. The association between each of the variables and the occurrence of causative pathogen was estimated by an OR and 95% confidence interval. The goodness-of-fit of the model was assessed by the Hosmer–Lemeshow goodness-of-fit test and Akaike’s corrected Information Criterion (AICc). The Hosmer–Lemeshow goodness-of-fit P values and AICc values reported in supplementary table S5 indicated the models are correct.

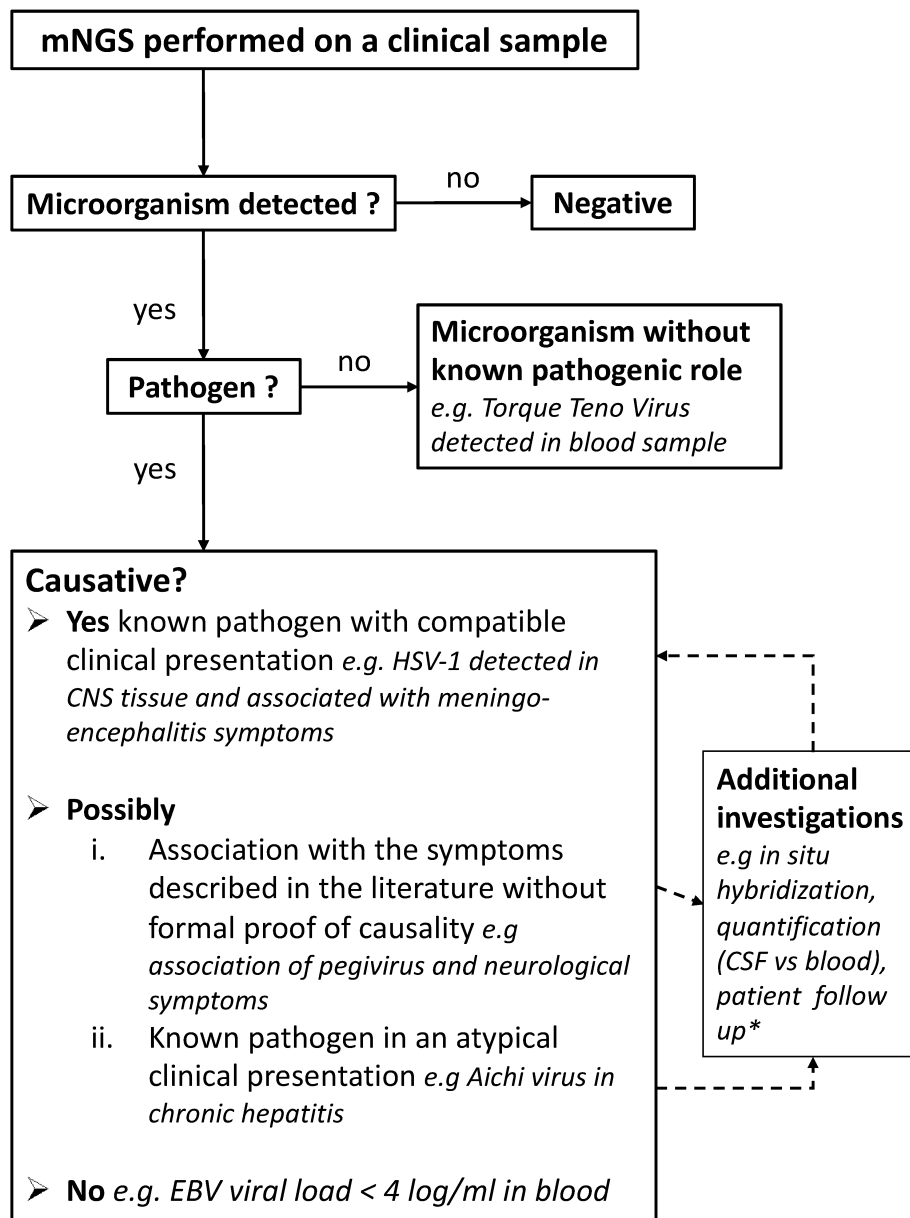


Figure S1. Categorization of metagenomic NGS findings.

*Tissue-based polyomavirus simian virus 40 (SV40) IHC was used in a case of detection of a low BK polyomavirus signal in a kidney biopsy, the IHC was negative and we therefore classified this signal as non-causative. We used in situ hybridization in liver tissues for 2 viruses suspected of a pathogenic role in a series of chronic hepatitis (Aichi virus)¹⁰ and a case of unexplained acute hepatitis (Human circovirus 1)¹¹ in immunocompromised patients. We have therefore classified Aichi virus and Human circovirus 1 in "causative" category for these clinical cases.

Quantification included comparison of viral load by qPCR in CSF and blood for instance in the case of Human pegivirus detection.

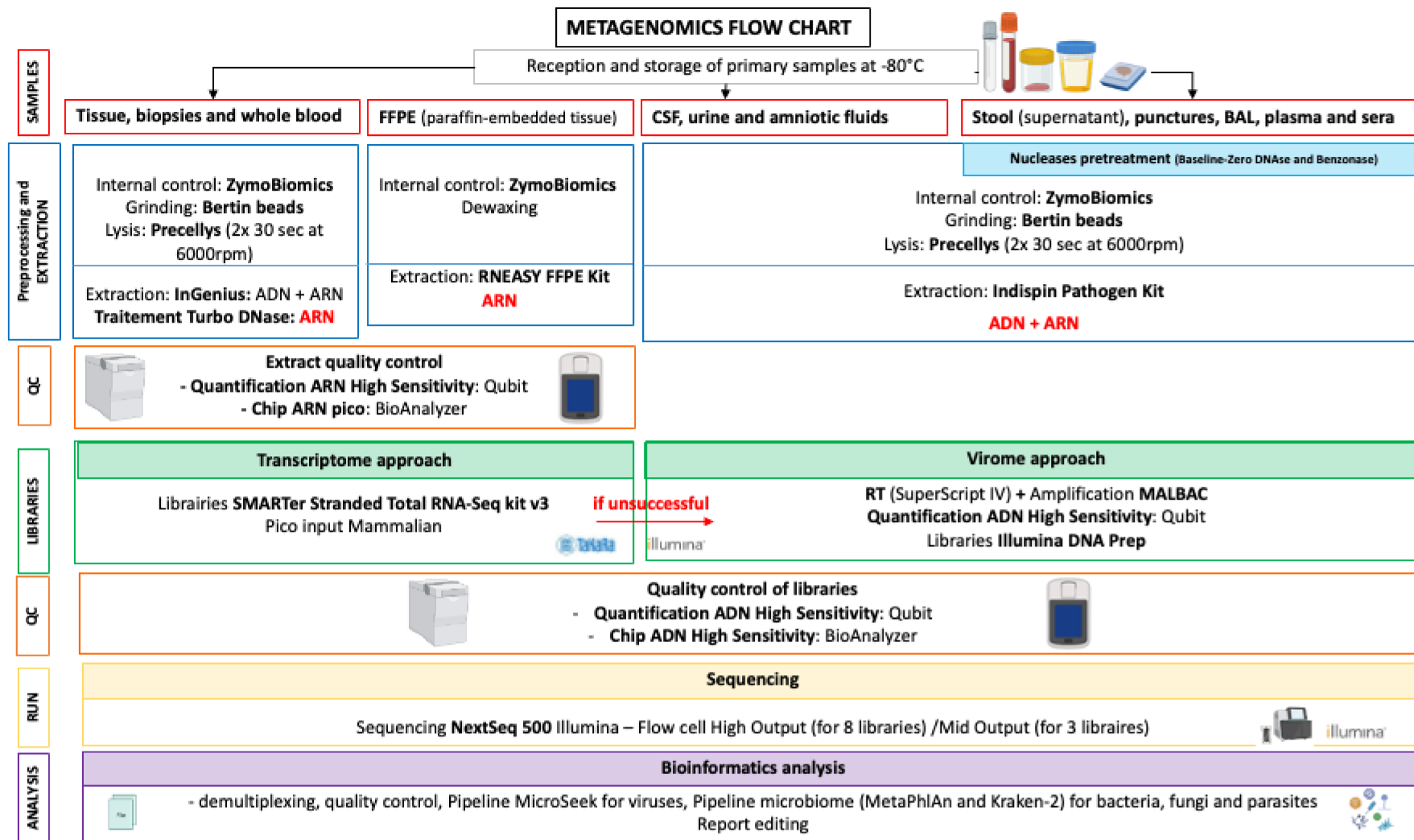


Figure S2. Flowchart summarizing the procedures for mNGS analysis performed in this study.

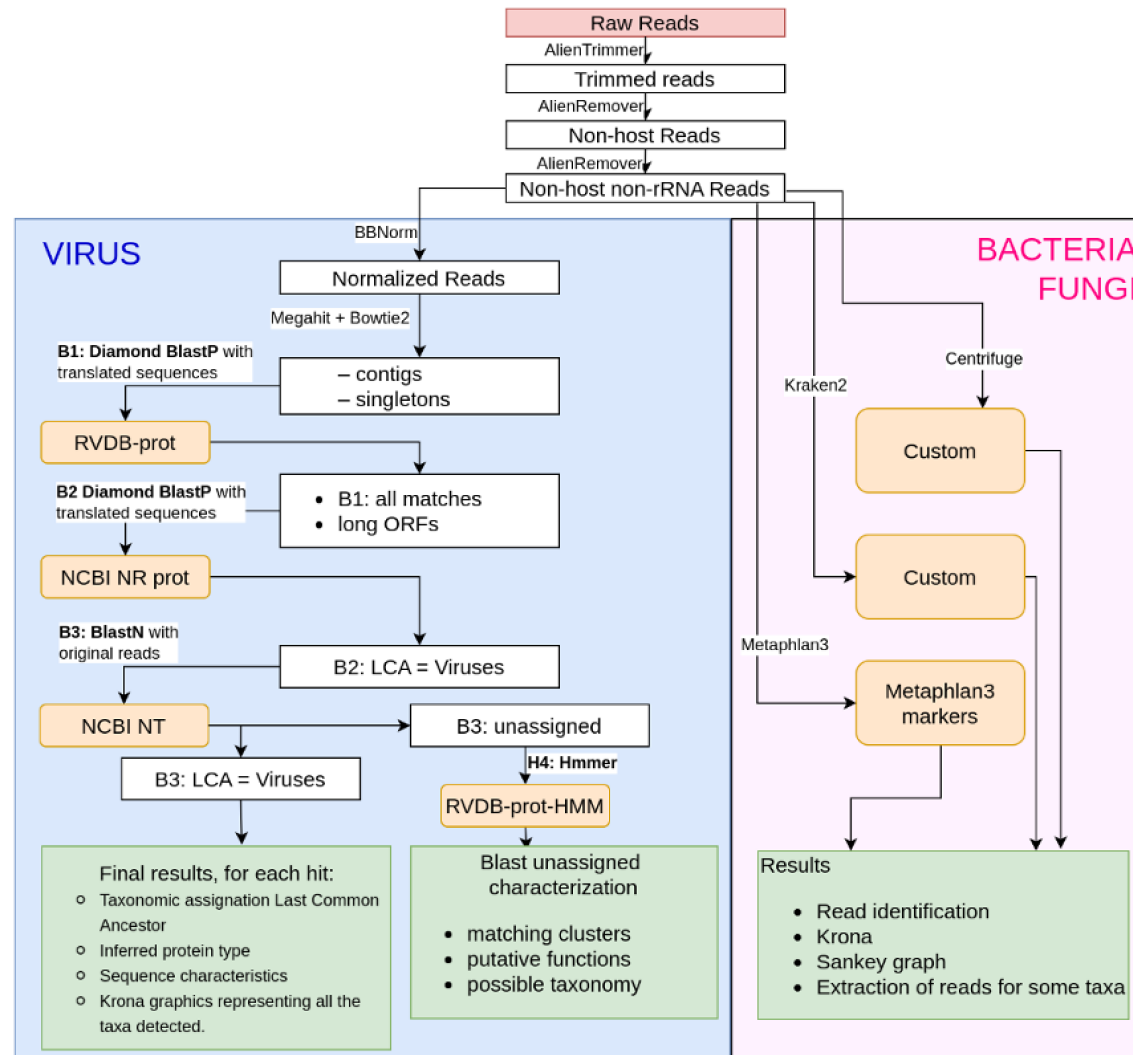


Figure S3. Flowchart summarizing the bioinformatic analysis performed in this study.

Table S1. Summary of prospective studies detecting pathogens with mNGS across various sample types and clinical indications.

Study	Patient population	No patients	Location	Sample types	Pathogens detected by mNGS	CT used for comparison	Agreement with CT (%)
Parize <i>et al.</i> , 2017 ¹²	Immunocompromised adults	101	France	Biological fluids, blood, nasopharyngeal swabs	Bacteria Viruses	Antigen detection, culture, PCR, serology	71%
Kufner <i>et al.</i> , 2019 ¹³	Adults	105	Switzerland	Biological fluids, biopsies, blood, respiratory tract samples, stool, throat and other swabs	Viruses	Culture, immunofluorescence, intrathecal antibody synthesis testing, PCR, serology	81%
Shen <i>et al.</i> , 2021 ^{*14}	Pediatric patients with hematological disorders	70	China	Biological fluids, blood, respiratory samples, throat swabs	Bacteria Fungi Viruses	Culture, serology, PCR	NA
Chen <i>et al.</i> , 2022 ¹⁵	Neonates	168	China	Biological fluids, blood, respiratory samples	Bacteria Fungi Viruses	Antigen detection, culture, PCR, serology	68·3%
Lamoureux <i>et al.</i> , 2022 ^{†16}	Adults	64	France	Abscesses, biological fluids, biopsies	Bacteria	16S rRNA gene sequencing	68·8% / 84·4%
Zhang <i>et al.</i> , 2022 ^{‡17}	Neonates	10	China	Biological fluids, blood, respiratory samples	Bacteria Viruses	Antigen detection, culture, serology	NA

No=number of; CT=conventional testing methods; NA=not applicable. *mNGS agreement with CT was not indicated in the study of Shen *et al.* †mNGS agreement with CT for the study of Lamoureux *et al.* depends on the level of bacterial identification: 68·8% at the species level and 84·4% at the genus level. ‡mNGS agreement with CT is not available for the study of Zhang *et al.* because mNGS was only performed when CT failed to elucidate the etiology.

Table S2. List of first-line tests required by sample type collected at Necker hospital *

Sample type	Always required	Required by context
CSF	standard bacterial culture, multiplex PCR targeting bacteria, viruses: BioFire® FilmArray® Meningitis/Encephalitis (ME) Panel (Biomérieux), HSV, VZV, <i>Enterovirus</i> PCR	fungus, mycobacterial cultures, cryptococcal Ag, 16S
Stool	<i>Adenovirus</i> PCR, multiplex PCR targeting bacteria, viruses, parasites: QIAstat-Dx® Gastrointestinal Panel (Qiagen)	Immunodeficient patients (<i>Enterovirus</i> , <i>Parvovirus</i> , <i>Aichi virus</i> PCR)
Plasma	<i>Enterovirus</i> RT-PCR	<i>Aichi virus</i> RT-PCR, HCV/HBV/HIV viral load
Blood	EBV, CMV, <i>Adenovirus</i> PCR, blood culture	<i>Human herpesvirus 6</i> PCR
Urines	BK virus PCR, standard bacterial culture	RT-PCR <i>Aichi virus</i> , Measles, <i>L. pneumophila</i> and <i>S. pneumoniae</i> antigen
Bone marrow aspiration	EBV, CMV, <i>Parvovirus B19</i> , <i>Adenovirus</i> PCR, standard bacterial culture	specific PCRs, fungus, mycobacterial cultures
Nasopharyngeal swab	multiplex PCR targeting bacteria, viruses: QIAstat-Dx Respiratory SARS-CoV-2 Panel	<i>Enterovirus</i> , Measles RT-PCR
Respiratory fluids	standard bacterial culture	specific PCRs, fungus, mycobacterial cultures
Pleural fluids	standard bacterial culture	specific PCRs, fungus, mycobacterial cultures
Ocular fluids	none for anterior chamber paracentesis (precious material with small volume)	<i>Adenovirus</i> , HSV, VZV, CMV PCR, standard bacterial culture
Amniotic fluids	PCRs and serology (IgG/IgM): CMV, <i>Parvovirus B19</i> , <i>Toxoplasmosis</i>	<i>Zika virus</i> , <i>Enterovirus</i> , HSV PCR
CNS tissues	none (CSF has always been explored prior to CNS biopsies)	JC virus PCR, <i>Astrovirus VA1</i> RT-PCT
Liver tissues	CMV, EBV, <i>Adenovirus</i> PCR, Hepatitis A virus IgG/IgM, <i>Hepatitis B virus</i> Antigen HBs, <i>Hepatitis C virus</i> PCR in plasma	HSV, VZV, EBV, <i>Human herpesvirus 6</i> , <i>hepatitis E virus</i> PCR, standard bacterial culture, fungus, mycobacterial cultures
Kidney tissues	BK virus PCR, standard bacterial culture	<i>Aichi virus</i> RT-PCR
Bone tissues	standard bacterial culture, mycobacteria detection by culture and PCR, 16S PCR	specific PCRs
Lymph nodes	standard bacterial culture, mycobacteria detection by culture, <i>Bartonella</i> PCR	specific PCRs
Skin tissues	standard bacterial/fungal culture	specific PCRs
Digestive tissues	CMV, <i>Adenovirus</i> PCR	specific PCRs
Respiratory tissues	standard bacterial culture	specific PCRs, fungus, mycobacterial cultures
Ear nose throat tissues	standard bacterial culture	specific PCRs, fungus cultures

* For samples collected at other hospitals: prescribers complete a form indicating the tests already performed, usually including the first-line tests recommended in their own health care facility. CMV=cytomegalovirus; EBV=Epstein-Barr virus; HSV=Herpes simplex virus

Table S3. Confirmatory tests used per pathogen with the required results for concordance with metagenomic NGS.

Causative pathogen detected with mNGS	Confirmatory tests used per detected pathogen	Results required for concordance	Limit of detection (LoD) for molecular detection according to manufacturer or authors	Quality controls used
VIRUS - Linear ssRNA(+) genome				
<i>Norovirus GI/GII</i>	RIDA@GENE Norovirus I & II (r-biopharm@)	Cycle threshold calculated	50 copies/reaction	Positive/negative control + RNA Internal control
<i>Rhinovirus</i>	Rhino&EV/Cc r-gene@ (Biomérieux)	Cycle threshold < 40	0.43-6.8 TID50/ml (LoD 95%)	Positive/negative control + HPRT1 gene control
<i>Aichi virus</i>	Described by Kitajima et al. Appl Environ Microbiol, 2013	Cycle threshold calculated	10 copies/reaction	Positive/negative control + RNP gene control
<i>Human pegivirus</i>	Described by Liu et al, Epidemiol Infect, 2015	Cycle threshold calculated	Not reported	Positive/negative control + Beta-Actine gene control
<i>Human parechovirus</i>	Parechovirus r-gene@ (Biomérieux)	Cycle threshold calculated	LoD 95% = 0.62-9.9 TID50/ml	Positive/negative control + Internal control
<i>Human astrovirus</i>	QIAstat-Dx@ Gastrointestinal Panel (Qiagen)	Cycle threshold calculated	1.0E-04 dilution from a 1.0E05 TCID50/mL stock	Internal control (IC)
<i>Astrovirus VAI</i>	Described by Souza et al, JCM, 2006	Cycle threshold calculated	500 copies/ml	Positive/negative control + RNP Gene control
<i>Enterovirus</i>	Enterovirus R-GENE@ (Biomérieux)	Cycle threshold calculated	0.03-0.4 TCID50/mL	Positive/negative control + Internal Control
<i>Sapovirus</i>	QIAstat-Dx@ Gastrointestinal Panel (Qiagen)	Cycle threshold calculated	3.2E-05 dilution from a 1.0E05 TCID50/mL stock	Internal control (IC)
<i>Dengue virus 1</i>	RealStar® Dengue Virus RT-PCR Kit (Altona Diagnostics)	Cycle threshold calculated	7,04 copies/μL	Positive/negative control + Internal Control
<i>Rubella virus</i>	Described by Neven et al, CID, 2016	Cycle threshold calculated	Not reported	Positive/negative control + Beta-Actine gene control
<i>Zika virus</i>	RealStar® Zika Virus RT-PCR Kit (Altona Diagnostics)	Cycle threshold calculated	80 genome copies/ml	Positive/negative control + Internal Control
VIRUS - Linear ssRNA(-) genome				

<i>Human respiratory syncytial virus A</i>	RSV/hMPV r-gene® (Biomérieux)	Cycle threshold < 40	0.06 TCID50/mL	Positive/negative control + HPRT1 gene cellular control
<i>European bat lyssavirus 1</i>	Described by Regnault et al, CID, 2022	Cycle threshold calculated	Not reported	Positive/negative control + Beta-Actine gene control
VIRUS - Circular ssDNA				
<i>Human circovirus 1</i>	Described by Perot et al, EID, 2023	Cycle threshold calculated	Not reported	Positive/negative control + Beta-Actine gene control
VIRUS - Circular dsDNA				
<i>BK polyomavirus</i>	BK Virus R-GENE® (Biomérieux)	Cycle threshold calculated	2.3-2-6 log copies/mL	Positive/negative control + Internal Control
<i>JC polyomavirus</i>	RealStar® JC virus PCR Kit (Altona Diagnostics)	Cycle threshold calculated	61 UI/ml	Positive/negative control + Internal Control
VIRUS - Linear ssDNA				
<i>Parvovirus B19</i>	Parvovirus B19 R-GENE® (Biomérieux)	Cycle threshold calculated	200-400 copies/ml	Positive/negative control + Internal Control
VIRUS - Linear dsDNA				
<i>Cytomegalovirus</i>	CMV B19 R-GENE® (Biomérieux)	Cycle threshold calculated	500 copies/mL	Positive/negative control + Internal Control
<i>Epstein-Barr virus</i>	EBV B19 R-GENE® (Biomérieux)	Cycle threshold calculated	182 copies/mL	Positive/negative control + Internal Control
<i>Human herpes virus 6</i>	RealStar® HHV-6 PCR Kit (Altona Diagnostics)	Cycle threshold calculated	0.89-2.47 copies/μL	Positive/negative control + Internal Control
<i>Human adenovirus</i>	ADENOVIRUS R-GENE® (Biomérieux)	Cycle threshold calculated	200-670 copies/ml	Positive/negative control + Internal Control
<i>Herpes simplex virus 1</i>	RealStar® HSV PCR Kit (Altona Diagnostics)	Cycle threshold calculated	0.16-1.3 copies/ml	Positive/negative control + Internal Control
BACTERIA				
<i>Streptococcus agalactiae</i>	Non selective culture up to 10 days in different media - 37°C	Positive < 10 days	Not applicable	Strains culture per month: E coli ATCC25922, K. pneumoniae ATCC700603, S aureus ATCC29213, S. pneumoniae ATCC49619, H. influenzae ATCC33560, C. jejuni ATCC33560, P. aeruginosa ATCC27853, E. faecalis ATCC51299, B. fragilis ATCC25285

<i>Streptococcus pneumoniae</i>	Non selective culture up to 10 days in different media - 37°C	Positive < 10 days	Not applicable	Strains culture per month: E coli ATCC25922, K. pneumoniae ATCC700603, S aureus ATCC29213, S. pneumoniae ATCC49619, H. influenzae ATCC33560, C. jejuni ATCC33560, P. aeruginosa ATCC27853, E. faecalis ATCC51299, B. fragilis ATCC25285
<i>Campylobacter jejuni</i>	QIAstat-Dx® Gastrointestinal Panel (Qiagen)	Cycle threshold calculated	Not applicable	Internal control (IC)
<i>Fusobacterium necrophorum</i>	Non selective culture up to 10 days in different media - 37°C anaerobic conditions	Positive < 10 days	Not applicable	Strains culture per month: E coli ATCC25922, K. pneumoniae ATCC700603, S aureus ATCC29213, S. pneumoniae ATCC49619, H. influenzae ATCC33560, C. jejuni ATCC33560, P. aeruginosa ATCC27853, E. faecalis ATCC51299, B. fragilis ATCC25285
<i>Mycobacterium avium</i>	Lowenstein-Jensen, Colestos, BD BBL MGIT media - 30°C	Positive < 42 days	Not applicable	1 <i>Mycobacterium tuberculosis</i> strain cultured per month
<i>Salmonella enterica</i>	QIAstat-Dx® Gastrointestinal Panel (Qiagen)	Cycle threshold calculated	Not applicable	Internal control (IC)
<i>Staphylococcus aureus</i>	Non selective culture up to 10 days in different media - 37°C	Positive < 10 days	Not applicable	Strains culture per month: E coli ATCC25922, K. pneumoniae ATCC700603, S aureus ATCC29213, S. pneumoniae ATCC49619, H. influenzae ATCC33560, C. jejuni ATCC33560, P. aeruginosa ATCC27853, E. faecalis ATCC51299, B. fragilis ATCC25285
FUNGI				
<i>Aspergillus fumigatus</i>	MycogenIE Aspergillus fumigatus PCR (Ademtech)	Cycle threshold calculated	6 copies/reaction	Positive/negative control + Internal control
<i>Candida Orthopsilosis</i>	Sabouraud medium - 27°C and 35°C	Positive < 21 days	Not applicable	1 <i>Candida albicans</i> ATCC strain cultured per month
<i>Candida Tropicalis</i>	Sabouraud medium - 27°C and 35°C	Positive < 21 days	Not applicable	1 <i>Candida albicans</i> ATCC strain cultured per month

Footnotes: In blood samples, thresholds used for significance based on quantitative level were: > 1.E04 copies/ml for CMV, Adenovirus, BK virus and, > 1.E05 copies/ml for EBV.

Table S4. List of the kits used for metagenomic NGS in this study.

Supplier	Part no.	Article	Category
ZYMO RESEARCH	D6321-10	ZymoBIOMICS Spike-in Control II	3 bacteria at different concentrations
Fisher	10646175	Turbo DNase 1000u	extraction
Indical	SP54104	IndiSpin Pathogen kit	extraction
Precision System Science	E13000	Lysis buffer MagDEA Dx SV	extraction
Qiagen	19076	Buffer ATL	extraction
Qiagen	19088	Reagent DX	extraction
QIAGEN	73504	RNEasy FFPE kit	extraction
Bertin / Ozyme	P000915-LYSK0-A.0	Kit SK38	extraction: glass and ceramic beads
Beckman Coulter	A63880	Agencourt ampure XP 5 mL (billes magnétiques)	library
Fischer	AM9850G	Invitrogen Tris (1M), pH 7.0 100mL	library
Takara Bio	634485	SMARTer® Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian 24 Rxns	library
Illumina	20024904	NextSeq 500/550 Mid Output Kit V2.5 150 cycles	NGS
Illumina	20024907	NextSeq 500/550 High Output Kit V2.5 150 cycles	NGS
Illumina	15017872	PhiX Control v3	NGS
Illumina	20018704	Illumina® DNA Prep, (M) Tagmentation (24 Samples)	NGS
Illumina	20018707	Nextera™ DNA CD Indexes (24 Indexes, 24 Samples)	NGS
Sigma	S2770-100ML	Sodium hydroxide solution (NaOH) 1N	NGS
Lucigen/Epicentre	DB0715K	Buffer Baseline-ZERO™ 10X	nuclease treatment
Lucigen/Epicentre	DB0715K	Baseline-Zero DNase 5000 U	nuclease treatment
Merck	70746-4	Benzonase Nuclease	nuclease treatment
YIKON GENOMICS	YK001A	MALBAC Single Cell WGA	pre-amplification
Agilent	5067-4626	Kit DNA High Sensitivity	QC
Agilent	5067-4627	Reagent DNA High Sensitivity	QC
Agilent	5067-1513	Kit RNA 6000 Pico	QC
Agilent	5067-1514	Reagent RNA 6000 Pico	QC
FISHER	10606433	Qubit dsDNA HS Assay Kit 100	QC
FISHER	10034622	Qubit RNA HS Assay Kit 100	QC
FISHER	Q32856	QUBIT assay tubes set (x500)	QC

Supplier	Part no.	Article	Category
ZYMO RESEARCH	D6321-10	ZymoBIOMICS Spike-in Control II	3 bacteria at different concentrations
Fisher	10646175	Turbo DNase 1000u	extraction
Indical	SP54104	IndiSpin Pathogen kit	extraction
Precision System Science	E13000	Lysis buffer MagDEA Dx SV	extraction
Qiagen	19076	Buffer ATL	extraction
Qiagen	19088	Reagent DX	extraction
QIAGEN	73504	RNEasy FFPE kit	extraction
Bertin / Ozyme	P000915-LYSK0-A.0	Kit SK38	extraction: glass and ceramic beads
Beckman Coulter	A63880	Agencourt ampure XP 5 mL (billes magnétiques)	library
Fischer	AM9850G	Invitrogen Tris (1M), pH 7.0 100mL	library
FISHER	15327696	SuperScript IV First-Strand Synthesis (x50)	RT
Devices			
ELITE InGenius	ELITech Group		
Precellys Evolution	Bertin Technologies		
NextSeq500	Illumina		

Table S5: Goodness-of-fit of the multivariable logistic regression model: Hosmer–Lemeshow goodness-of-fit test and Akaike’s corrected Information Criterion (AICc) results.

multivariable logistic regression model	n	Hosmer-Lemeshow goodness-of-fit P values	Null hypothesis	Reject Null Hypothesis?	Akaike’s corrected Information Criterion (AICc)	
					Intercept-only model	Selected model
Gender, age, immune status, initial suspicion of infection, indication	742	0.08	Selected model is correct	No	678	575
Tissue	338	0.52	Selected model is correct	No	302	266
Other samples	393	0.97	Selected model is correct	No	367	305
Neurological	248	0.82	Selected model is correct	No	173	168
Digestive	70	0.96	Selected model is correct	No	99	99

Table S6. Results by sample type and clinical indications.

Sample type	Indication	Result	Number of samples	Total sample type per indication	Percent per sample type per indication
Blood	Digestive	Negative	8	8	100.0
Blood	Hematological	Positive	1	1	100.0
Blood	Hepatic	Negative	13	14	92.9
Blood	Hepatic	Positive	1	14	7.1
Blood	Inflammatory	Negative	2	2	100.0
Blood	Neurological	Negative	12	15	80.0
Blood	Neurological	Positive	3	15	20.0
Blood	Respiratory	Positive	1	1	100.0
Blood	Other Indication	Negative	6	6	100.0
CNS	Neurological	Negative	45	58	77.6
CNS	Neurological	Positive	13	58	22.4
CNS	Respiratory	Negative	3	4	75.0
CNS	Respiratory	Positive	1	4	25.0
CNS	Other Indication	Negative	1	1	100.0
CSF	Hematological	Negative	1	1	100.0
CSF	Hepatic	Negative	2	2	100.0
CSF	Inflammatory	Negative	2	2	100.0
CSF	Neurological	Negative	163	176	92.6
CSF	Neurological	Positive	13	176	7.4
CSF	Respiratory	Negative	2	2	100.0
CSF	Other Indication	Negative	2	2	100.0
Digestive	Digestive	Negative	8	14	57.1
Digestive	Digestive	Positive	6	14	42.9
Digestive	Hematological	Negative	1	1	100.0
Digestive	Hepatic	Negative	1	1	100.0
Digestive	Respiratory	Positive	2	2	100.0
Eart Nose Throat	Inflammatory	Negative	2	2	100.0
Eart Nose Throat	Neurological	Negative	1	1	100.0
Eart Nose Throat	Other Indication	Negative	7	8	87.5
Eart Nose Throat	Other Indication	Positive	1	8	12.5
FFPE	Digestive	Negative	3	3	100.0
FFPE	Hepatic	Negative	2	2	100.0
FFPE	Inflammatory	Negative	1	1	100.0
FFPE	Neurological	Negative	1	3	33.3
FFPE	Neurological	Positive	2	3	66.7
FFPE	Other Indication	Negative	2	2	100.0
Kidney	Digestive	Negative	1	1	100.0
Kidney	Hematological	Negative	1	1	100.0
Kidney	Hepatic	Negative	2	3	66.7
Kidney	Hepatic	Positive	1	3	33.3
Kidney	Neurological	Negative	2	2	100.0
Kidney	Other Indication	Negative	39	39	100.0
Liver	Digestive	Negative	2	2	100.0
Liver	Hematological	Negative	3	3	100.0
Liver	Hepatic	Negative	50	66	75.8
Liver	Hepatic	Positive	16	66	24.2
Liver	Other Indication	Negative	1	1	100.0
Lung	Respiratory	Negative	13	15	86.7
Lung	Respiratory	Positive	2	15	13.3
Lung	Other Indication	Negative	2	2	100.0
Lymph Node	Hematological	Positive	1	1	100.0
Lymph Node	Hepatic	Negative	1	3	33.3
Lymph Node	Hepatic	Positive	2	3	66.7
Lymph Node	Inflammatory	Negative	24	26	92.3
Lymph Node	Inflammatory	Positive	2	26	7.7
Lymph Node	Neurological	Negative	1	1	100.0
Lymph Node	Respiratory	Positive	1	1	100.0
Lymph Node	Other Indication	Negative	1	1	100.0
Non-Sterile Fluid	Digestive	Negative	8	9	88.9
Non-Sterile Fluid	Digestive	Positive	1	9	11.1
Non-Sterile Fluid	Hematological	Negative	2	3	66.7
Non-Sterile Fluid	Hematological	Positive	1	3	33.3

Non-Sterile Fluid	Hepatic	Negative	6	8	75.0
Non-Sterile Fluid	Hepatic	Positive	2	8	25.0
Non-Sterile Fluid	Inflammatory	Negative	2	4	50.0
Non-Sterile Fluid	Inflammatory	Positive	2	4	50.0
Non-Sterile Fluid	Neurological	Negative	4	7	57.1
Non-Sterile Fluid	Neurological	Positive	3	7	42.9
Non-Sterile Fluid	Respiratory	Negative	8	13	61.5
Non-Sterile Fluid	Respiratory	Positive	5	13	38.5
Non-Sterile Fluid	Other Indication	Negative	10	11	90.9
Non-Sterile Fluid	Other Indication	Positive	1	11	9.1
Osteo Muscular	Neurological	Negative	2	2	100.0
Osteo Muscular	Respiratory	Negative	1	1	100.0
Osteo Muscular	Other Indication	Negative	34	36	94.4
Osteo Muscular	Other Indication	Positive	2	36	5.6
Other Tissue	Digestive	Negative	1	1	100.0
Other Tissue	Hematological	Negative	1	2	50.0
Other Tissue	Hematological	Positive	1	2	50.0
Other Tissue	Hepatic	Positive	1	1	100.0
Other Tissue	Inflammatory	Negative	1	2	50.0
Other Tissue	Inflammatory	Positive	1	2	50.0
Other Tissue	Neurological	Negative	3	3	100.0
Other Tissue	Respiratory	Negative	1	1	100.0
Other Tissue	Other Indication	Negative	9	9	100.0
Skin	Digestive	Negative	1	1	100.0
Skin	Neurological	Negative	1	1	100.0
Skin	Other Indication	Negative	16	18	88.9
Skin	Other Indication	Positive	2	18	11.1
Sterile Fluid	Hematological	Negative	7	10	70.0
Sterile Fluid	Hematological	Positive	3	10	30.0
Sterile Fluid	Hepatic	Negative	10	11	90.9
Sterile Fluid	Hepatic	Positive	1	11	9.1
Sterile Fluid	Inflammatory	Negative	3	3	100.0
Sterile Fluid	Neurological	Negative	2	2	100.0
Sterile Fluid	Respiratory	Negative	7	8	87.5
Sterile Fluid	Respiratory	Positive	1	8	12.5
Sterile Fluid	Other Indication	Negative	14	20	70.0
Sterile Fluid	Other Indication	Positive	6	20	30.0
Stool	Digestive	Negative	10	22	45.5
Stool	Digestive	Positive	12	22	54.5
Stool	Hematological	Positive	2	2	100.0
Stool	Hepatic	Negative	10	17	58.8
Stool	Hepatic	Positive	7	17	41.2
Stool	Inflammatory	Negative	4	5	80.0
Stool	Inflammatory	Positive	1	5	20.0
Stool	Neurological	Negative	1	3	33.3
Stool	Neurological	Positive	2	3	66.7
Stool	Respiratory	Negative	2	2	100.0
Stool	Other Indication	Negative	1	1	100.0

Table S7. List of confirmed causative or possibly causative microorganisms identified.

Causative pathogen or microorganism associated with a pathogen in the same sample (n)	n	1 pathogen per sample	>1 pathogen per sample	CSF (n=)	Fluid (n=)	Stool (n=)	Blood (n=)	Tissues (n=)	Tissue type (n)
VIRUS - Linear ssRNA(+) genome									
<i>Norovirus GI (2) GII (11)</i>	13	9	4	0	2	8	0	3	digestive (3)
<i>Rhinovirus A(2), B(1), C (5)</i>	8	6	2	0	5	3	0	1	
<i>Aichi virus A (4), B (3)</i>	7	7	0	0	0	0	0	7	liver (5), spleen (1), kidney (1)
<i>Human pegivirus</i>	6	6	0	2	0	0	0	4	CNS(4)
<i>Human parechovirus 1(3)</i>	3	2	1	0	0	2	0	1	digestive (1)
<i>Human astrovirus 1 (2), Astrovirus VA1 (1) MLB1 (1)</i>	4	4	0	0	0	2	0	2	CNS (1), Digestive (1)
<i>Coxsackievirus A4 (1), CV-B5 (1), Enterovirus D68 (1)</i>	3	1	2	0	0	1	0	2	CNS (1), Lymph node (1)
<i>Sapovirus GI (1) GV (1)</i>	2	0	2	0	0	2	0	0	
<i>Dengue virus 1</i>	1	0	1	0	1	0	0	0	
<i>Rubella virus</i>	1	1	0	0	0	0	0	1	skin (1)
<i>Zika virus (asiatic lineage)</i>	1	1	0	0	1	0	0	0	
VIRUS - Linear ssRNA(-) genome									
<i>Human respiratory syncytial virus A</i>	2	1	1	0	1	0	0	1	lung (1)
<i>European bat lyssavirus 1</i>	1	1	0	0	0	0	0	1	CNS (1)
VIRUS - Circular ssDNA									
<i>Human circovirus 1</i>	1	1	0	0	0	0	0	1	liver (1)
VIRUS - Circular dsDNA									
<i>BK polyomavirus</i>	1	0	1	0	1	0	0	0	
<i>JC polyomavirus</i>	1	1	0	0	0	0	0	1	CNS (1)
VIRUS - Linear ssDNA									
<i>Parvovirus B19</i>	1	1	0	0	0	0	1	0	
VIRUS - Linear dsDNA									
<i>Cytomegalovirus</i>	11	5	6	1	3	1	2	4	liver (1), Digestive (2), lymph node (1)
<i>Epstein-Barr virus</i>	11	11	0	3	1	0	1	6	liver (4), CNS (2)
<i>Human herpes virus 6</i>	7	4	3	3	3	0	0	1	CNS (1)
<i>Human adenovirus 1</i>	4	0	4	0	1	1	0	2	liver (1), digestive (1)
<i>Herpes simplex virus 1</i>	3	2	1	0	1	0	0	2	CNS (1), lymph node (1)
BACTERIA									
<i>Streptococcus agalactiae (2), pneumoniae (3)</i>	5	4	1	0	3	0	0	2	osteo-muscular (2)

<i>Campylobacter jejuni</i>	2	0	2	0	0	2	0	0	
<i>Fusobacterium necrophorum</i>	1	1	0	0	0	0	0	1	retroauricular abscess (1)
<i>Mycobacterium avium</i>	1	1	0	0	1	0	0	0	
<i>Salmonella enterica</i>	1	0	1	0	0	1	0	0	
<i>Staphylococcus aureus</i>	1	0	1	0	0	0	0	1	lymph node (1)
FUNGI									
<i>Aspergillus fumigatus</i>	2	2	0	0	1	0	0	1	CNS (1)
<i>Candida Orthopsilosis (1) C. Tropicalis (1)</i>	2	2	0	0	1	0	0	1	lung (1)

Footnotes : CNS, central nervous system

Table S8. Cases with mNGS negative findings and positive results of conventional tests.

Microorganism	Sample type	Conventional test	Conventional test results
VIRUS			
<i>Aichi virus</i> , n=1	liver	Aichi virus RT-PCR (Kitajima et al, 2013)	40 (cycle threshold)
<i>Cytomegalovirus</i> , n=1	blood	ARGENE Cytomegalovirus qPCR (Biomérieux, Marcy L'Etoile, France)	72 copies/ml
<i>Epstein-Barr virus</i> , n=2	blood	ARGENE Epstein-Barr virus qPCR (Biomérieux, Marcy L'Etoile, France)	<500 copies/ml
<i>Human Immunodeficiency Virus 1</i> , n=1	cerebrospinal fluid	Alinity m HIV-1 assay (Abbott molecular, Rungis, France)	166 copies/ml
<i>Sapovirus</i> , n=1	stool	QIAstat-Dx Gastrointestinal Panel (Qiagen, Courtaboeuf, France)	36 (cycle threshold)
BACTERIA			
<i>Mycobacterium tuberculosis</i> , n=1	lymph node	culture	positive in 2 weeks
FUNGI			
<i>Aspergillus fumigatus</i> , n=1	cerebrospinal fluid	MycoGENIE <i>Aspergillus fumigatus</i> PCR (Ademtech, Pessac, France)	37 (cycle threshold)
<i>Aspergillus fumigatus</i> , n=1	Formalin-fixed Paraffin-Embedded brain biopsy	MycoGENIE <i>Aspergillus fumigatus</i> PCR (Ademtech, Pessac, France)	27 (cycle threshold)
<i>Histoplasma</i> sp, n=1	Bronchoalveolar fluid	<i>Histoplasma</i> PCR	positive
<i>Paecilomyces lilacinus</i> , n=1	bone	culture	positive

References

1. Pérot P, Bigot T, Temmam S, Regnault B, Eloit M. Microseek: A Protein-Based Metagenomic Pipeline for Virus Diagnostic and Discovery. *Viruses* 2022; **14**(9): 1990.
2. Fourgeaud J, Regnault B, Faury H, et al. Fetal Zika virus infection diagnosed by metagenomic next-generation sequencing of amniotic fluid. *Ultrasound in Obstetrics & Gynecology* 2023; **61**(1): 116-7.
3. Bodemer C, Sauvage V, Mahlaoui N, et al. Live rubella virus vaccine long-term persistence as an antigenic trigger of cutaneous granulomas in patients with primary immunodeficiency. *Clin Microbiol Infect* 2014; **20**(10): O656-63.
4. Neven B, Pérot P, Bruneau J, et al. Cutaneous and Visceral Chronic Granulomatous Disease Triggered by a Rubella Virus Vaccine Strain in Children With Primary Immunodeficiencies. *Clinical Infectious Diseases* 2017; **64**(1): 83-6.
5. Riller Q, Fourgeaud J, Bruneau J, et al. Late-onset enteric virus infection associated with hepatitis (EVAH) in transplanted SCID patients. *J Allergy Clin Immunol* 2023: S0091-6749(23)00009-X.
6. Balcom EF, Doan MAL, Branton WG, et al. Human pegivirus-1 associated leukoencephalitis: Clinical and molecular features. *Ann Neurol* 2018; **84**(5): 781-7.
7. Doan MAL, Roczowski A, Smith M, et al. Infection of Glia by Human Pegivirus Suppresses Peroxisomal and Antiviral Signaling Pathways. *J Virol* 2021; **95**(23): e0107421.
8. Valyraki N, Maillart E, Pourcher V, et al. Human pegivirus identified in severe myelitis and optic neuritis in immunocompromised patients: A pathogenic role for a forgotten virus? *Rev Neurol (Paris)* 2023; **179**(4): 361-7.
9. Hlavay BA, Zhuo R, Ogando N, et al. Human pegivirus viremia in HCV/HIV co-infected patients: Direct acting antivirals exert anti-pegivirus effects. *J Clin Virol* 2023; **162**: 105445.
10. Fourgeaud J, Lecuit MM, Perot P, et al. Chronic Aichi virus infection as a cause of long-lasting multi-organ involvement in patients with primary immune deficiencies. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2023.
11. Perot P, Fourgeaud J, Rouzaud C, et al. Circovirus Hepatitis Infection in Heart-Lung Transplant Patient, France. *Emerging infectious diseases* 2023; **29**(2): 286-93.
12. Parize P, Muth E, Richaud C, et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2017; **23**(8): 574 e1- e6.
13. Kufner V, Plate A, Schmutz S, et al. Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases. *Genes (Basel)* 2019; **10**(9).
14. Shen H, Shen D, Song H, et al. Clinical assessment of the utility of metagenomic next-generation sequencing in pediatric patients of hematology department. *Int J Lab Hematol* 2021; **43**(2): 244-9.
15. Chen L, Zhao Y, Wei J, et al. Metagenomic Next-Generation Sequencing for the Diagnosis of Neonatal Infectious Diseases. *Microbiol Spectr* 2022; **10**(6): e0119522.
16. Lamoureux C, Surgers L, Fihman V, et al. Prospective Comparison Between Shotgun Metagenomics and Sanger Sequencing of the 16S rRNA Gene for the Etiological Diagnosis of Infections. *Frontiers in microbiology* 2022; **13**: 761873.
17. Zhang R, Zhuang Y, Xiao ZH, et al. Diagnosis and Surveillance of Neonatal Infections by Metagenomic Next-Generation Sequencing. *Frontiers in microbiology* 2022; **13**: 855988.

Titre : Identification de virus impliqués dans les pathologies inexpliquées du patient immunodéprimé

Résumé : Les patients présentant un déficit immunitaire primitif ou secondaire peuvent être infectés par un large panel de pathogènes entraînant des symptômes parfois atypiques rendant le diagnostic difficile. Alors que les méthodes de culture ou de PCR universelle pour le diagnostic des infections bactériennes ont un spectre large, la détection des virus est ciblée, ce qui restreint la liste des virus détectables aux virus pathogènes les plus courants. Dans ce contexte, des pathologies qui pourraient avoir une étiologie virale restent inexpliquées. La métagénomique (mNGS) permet l'identification non ciblée d'un large éventail d'agents pathogènes, y compris les virus rares ou nouveaux qui ne sont pas détectés avec les techniques conventionnelles. Cependant les performances de cette analyse dans le cadre du soin et ses indications les plus pertinentes sont mal définies. L'objectif de cette étude est d'identifier des virus impliqués dans certaines pathologies inexpliquées du patient immunodéprimé par mNGS et d'identifier des indications de l'analyse mNGS pour le diagnostic des infections virales. Dans la première partie de ce travail, nous avons (i) mis en évidence une nouvelle pathologie associée à *Aichi virus* chez 6 patients présentant un déficit immunitaire primitif et (ii) identifié un nouveau *Circovirus* chez une patiente adulte transplantée cœur-poumon présentant une hépatite aigue inexpliquée. Dans les 2 cas nous avons présenté un ensemble d'arguments en faveur du lien de causalité entre l'infection virale et la pathologie hépatique, grâce à la mise au point de techniques de PCR, d'hybridation *in situ* et de génotypage à façon. Dans la deuxième partie, nous avons analysé les résultats de trois années d'activité de mNGS réalisée dans le cadre du diagnostic. Cette étude nous a permis d'identifier un intérêt majeur du mNGS chez les patients immunodéprimés (i) pour la détection de pathogènes opportunistes, (ii) en cas de symptômes atypiques ou (iii) en cas de confusion avec une pathologie inflammatoire ou dysimmunitaire. Cette étude nous a aussi permis de classer les prélèvements en fonction de leur rentabilité diagnostique, élevée pour les selles et faible pour le liquide cébrospinal (LCS). Nous avons montré l'intérêt de l'analyse des tissus cérébraux par rapport au LCS en cas de symptômes neurologiques inexpliqués. Nos résultats suggèrent que l'analyse des selles pourrait avoir un intérêt en cas de symptômes hépatique, hématologique ou neurologique. L'ensemble de ce travail a permis de préciser le positionnement du mNGS pour le diagnostic des infections virales.

Mots clefs : Métagénomique, *Aichi virus*, *Circovirus*, Déficit immunitaire, NGS

Title: Virus identification in immunocompromised patients with unexplained diseases

Abstract: Patients with primary or secondary immune deficiencies may be infected by a wide range of pathogens, resulting in atypical symptoms that make diagnosis difficult. While culture or universal PCR assays for bacterial detection are broad-spectrum, virus detection is targeted, limiting the list of detectable viruses to the most common pathogens. In this context, diseases that could have a viral etiology remain unexplained. Metagenomics (mNGS) enables the untargeted identification of a wide range of pathogens, including rare or novel viruses not detected by conventional tests. Despite mNGS recognition as a valuable diagnostic tool for infections, the most relevant indications for this innovative strategy remain poorly defined. The aim of this study is to identify viruses involved in unexplained diseases of the immunocompromised patient, and to identify mNGS indications in the diagnosis of viral infections. First, (i) we highlighted a novel disease associated with Aichi virus in six patients with primary immune deficiency and (ii) identified a novel Circovirus in a heart-lung transplant adult patient with unexplained acute hepatitis. In both cases, we provided evidence for a causal link between viral infection and hepatitis, using PCR, *in situ* hybridization and custom genotyping techniques. Second, we analyzed the results of three years of mNGS activity as part of our diagnostic workflow. This study enabled us to identify a major benefit of mNGS in immunocompromised patients (i) for the detection of opportunistic pathogens, (ii) in case of atypical symptoms or (iii) in case of confusion with an inflammatory or dysimmune disorder. This study also allowed us to classify samples according to the proportion of pathogens detected: high for stool and low for cerebrospinal fluid (CSF). We showed that mNGS had higher odds of detecting a pathogenic virus on cerebral tissue compared to CSF in patients with unexplained neurological symptoms. Our results suggest that stool analysis could be of interest in cases of liver, hematological or neurological symptoms. In conclusion, the results of this work contribute to refining the positioning of mNGS in the diagnostic workflow.

Keywords : Metagenomics, *Aichi virus*, *Circovirus*, Immune deficiency, NGS