



Universiteit
Leiden

The Netherlands

Metagenomic sequencing in clinical virology: advances in pathogen detection and future prospects

Carbo, E.C.

Citation

Carbo, E. C. (2023, May 17). *Metagenomic sequencing in clinical virology: advances in pathogen detection and future prospects*. Retrieved from <https://hdl.handle.net/1887/3618319>

Version: Publisher's Version

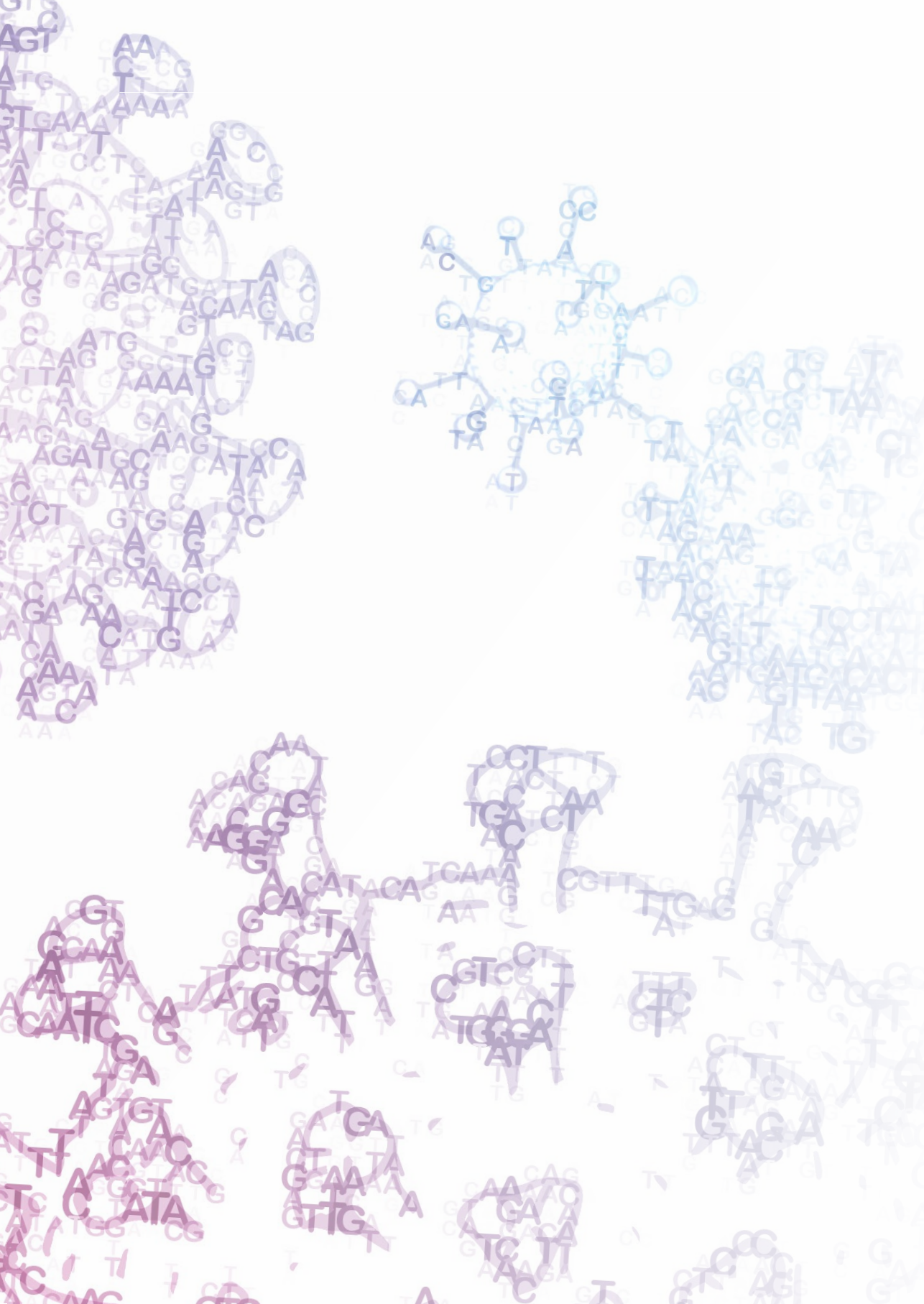
License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3618319>

Note: To cite this publication please use the final published version (if applicable).

*Science and everyday life cannot
and should not be separated*

Rosalind Franklin, in a letter to her father



Addendum

Dutch Summary / Nederlandse Samenvatting

List of Publications / Publicaties

Curriculum Vitae

Dutch Summary / Nederlandse Samenvatting

Metagenomische sequentie analyse binnen de klinische virologie: Ontwikkelingen in pathogendetectie en toekomstperspectieven

Om virussen te detecteren zoekt men in de hedendaagse diagnostiek gericht naar een of meerdere specifieke virussen. Zodra vooraf bekend is welk virus men verwacht te vinden bij een patiënt, is het mogelijk om een gerichte *polymerase chain reaction* (PCR) in te zetten. Een andere benadering is om het metagenoom in kaart te brengen, dan is het mogelijk om alle virussen in één keer tegelijk te analyseren. Hierbij wordt al het genetisch materiaal rechtstreeks uit een patiëntenmonster opgewerkt en geanalyseerd op de aanwezigheid van virale micro-organismen. Met deze methode kan ook andere genetische informatie opgespoord worden, zoals aanwezige bacteriën of andere organismen, maar ook bepaalde genetische eigenschappen van de gastheer zelf. Deze mogelijkheid heeft daarom belangrijke toekomstperspectieven. De focus van dit proefschrift ligt op de metagenomische sequentie analyse, die het mogelijk maakt om alle voor de mens pathogene virussen rechtstreeks uit patiëntmateriaal met eenzelfde test te detecteren.

De WHO rapporteert cijfers over de wereldwijd voorkomende doodsoorzaken, en waar vaak wordt aangenomen dat dit hart- en vaatziekten zijn, blijkt als de cijfers van alle verschillende infectieziekten bij elkaar worden opgeteld inclusief lagere luchtweginfecties, dat infectieziekten op nummer een te staan. Sinds de SARS-CoV-2 pandemie is meer dan ooit duidelijk geworden dat infectieziekten ook in onze westerse wereld tot ontwrichting van de samenleving zorgen. In de hedendaagse maatschappij leven wij dicht op elkaar en dicht op andere dierlijke organismen. De overdracht van virusinfecties tussen mensen en dieren vormt een constant en reëel risico. Virussen kunnen evolueren en soms ook overgedragen worden op andere organismen dan de oorspronkelijke gastheer. Dat maakt de ontwikkeling van een test waarmee in één keer naar alle virussen tegelijk kan worden gekeken rechtstreeks in patiëntmateriaal zo belangrijk. Met de bestaande diagnostische testen kunnen patiënten die besmet raakten met hetzij een nieuw, hetzij een dierlijk, hetzij met een niet verwacht virus niet of slechts moeilijk gediagnostiseerd worden.

Bij een metagenomische analyse wordt al het erfelijke materiaal, dus zowel het DNA-als het RNA-materiaal, rechtstreeks uit een patiëntenmonster geïsoleerd. Dit wordt vervolgens opgewerkt met een *library preparation* om zo het genetische

materiaal geschikt te maken voor sequentie analyse. Dit gebeurt door het metagenoom te fragmenteren en door RNA om te zetten in *copy DNA* (cDNA). DNA-fragmenten worden verder behandeld door *adapter linking*, barcodering en amplificatie van het DNA-materiaal. Eventueel kan er voor het daadwerkelijke *sequenzen* nog een verrijking plaatsvinden met gerichte *probe* sequenties, waardoor alle virussen worden geselecteerd door *probe* extractie. Vervolgens worden, met of zonder virale verrijking, de DNA-fragmenten op een *sequencer* geladen. Na amplificatie leest deze de nucleotiden af van de verschillende fragmenten die aanwezig waren in de patiëntenmonsters. Om deze, soms miljoenen, *reads* verder te verwerken ten einde een virus te detecteren, is speciale bioinformatische classificatiesoftware nodig.

Bij aanvang van dit promotietraject waren er nauwelijks *systematic reviews* en meta-analyses bekend die onderzoek hebben gedaan naar de hoeveelheid additionele diagnoses die zouden plaatsvinden bij gebruik van de metagenomische *sequence* test. Er waren tevens vrijwel geen onafhankelijke vergelijkingen van bioinformatische *tools*. En er waren weinig publicaties voor het verbeteren van een diagnostisch toepasbaar protocol. In deze thesis is er verder gegaan met deze openstaande vraagstellingen.

Encefalitis is een voorbeeld van een ziektebeeld waarbij een PCR-test niet in alle gevallen uitsluitend geeft. **Hoofdstuk 2** geeft een overzicht van verschillende wetenschappelijke artikelen over onderzoek naar encefalitis met metagenomische sequentie analyse. De focus lag op het aantal additionele virussen die werden gevonden in vergelijking met de initiële traditionele diagnostische testen die waren uitgevoerd. Met een meta-analyse is er aangetoond dat er in de verschillende onderzoeken 10,9 procent extra virale diagnoses konden worden gesteld indien men naar alle virussen tegelijk zou kijken i.p.v. gericht naar enkele vooraf verdachte virussen. Omdat de sensitiviteit van deze test lager was dan een reguliere diagnostische test, is er in **hoofdstuk 3** verder gekeken naar hoe de techniek verbeterd kon worden wat betreft sensitiviteit en specificiteit door het vergelijken van patiëntenmonsters waarbij zowel metagenomische sequentie analyse werd gebruikt alsook de reguliere diagnostische tests. Daarnaast is er een vergelijking gemaakt tussen *shotgun* metagenomische *sequencing*, en metagenomische sequentie analyse met een *probe* set waarbij alle bekende virussen waarmee gewervelden organismen geïnfecteerd kunnen worden, aanwezig zijn. De sensitiviteit met het gebruik van *probes* was 100 procent t.o.v. de initiële PCR-test, en de hoeveelheid *sequence reads* nam 100 tot 10.000x toe, waardoor ook de genomesequenties van

de virussen beter konden worden bekeken. De verbeterde dekking van het genoom zorgde ook voor nauwkeuriger diagnostiek met betrekking tot de specificiteit van de virale metagenomische test. Na de technische vergelijkingen is er in **hoofdstuk 3** een aantal patiëntenmonsters nader bekeken van hematologische patiënten met een verdenking op encefalitis waar geen causaal virus of bacterie uit de initiële diagnostische testen naar voren kwam. Daar kwamen in 12,2 procent van de gevallen niet eerder bij de patiënt aangetoonde virussen uit naar voren, zoals: BK polyomavirus, hepatitis E-virus, humaan herpes virus-6 en Epstein-Barr-virus. De populatie patiënten met een onverklaarde encefalitis zou een goede patiëntenpopulatie zijn om de metagenomische sequentie analyse test in te zetten.

Reizigers uit het buitenland die terugkomen met onverklaarbare koorts behoren eveneens tot een doelgroep die bij het stellen van een diagnose mogelijk baat hebben bij de inzet van virale metagenomische sequentie analyse. In **hoofdstuk 4** is het onderzoek bij deze groep patiënten beschreven. Bij verschillende patiënten werden virussen gevonden waar eerder niet op was getest of waar de initiële antigeentest negatief was. Pathogene virussen die in 5 procent additioneel werden gevonden waren: denguevirus en hepatitis C virus. Tevens werden met deze brede benadering alle initieel gevonden virussen teruggevonden en enkele additionele niet-pathogene virussen. Verder maakt de test het mogelijk om de verschillende virussen verder te typeren, bijvoorbeeld als *subspecies*.

In deze eerdere hoofdstukken is gekeken naar de diagnostische waarde, de verbetering van het laboratoriumprotocol en de toepasbaarheid bij enkele verschillende ziektebeelden van de metagenomische test. In **hoofdstuk 5** is er gekeken naar wat de meest betrouwbare bioinformatische software is voor het accuraat detecteren van virussen bij metagenomische sequentie analyse. Hiervoor is er luchtwegmateriaal van 88 patiënten met een chronische longaandoening getest op verschillende luchtweginfectievirussen met in totaal 1120 PCR-testen. Hetzelfde luchtwegmateriaal van deze patiënten is ook bestudeerd met metagenomische sequentie analyse, om vervolgens met vijf verschillende softwareprogramma's de sensitiviteit en specificiteit van de metagenomische software te bepalen. De bioinformatische programma's hadden een sensitiviteit van 83 tot 100 procent.

Als extra onderzoek is er in **hoofdstuk 5** gekeken of het aantal *sequence reads* correleert met de ct-waarde van de PCR-test, een mate van kwantiteit. Hierbij kwam naar voren dat de meest accurate *tool* de *reads* op basis van aminozuren

classificeert, al waren er grote verschillen tussen bepaalde soorten virussen. Vooral bij divergente virussen zoals rhinovirussen, bleek de kwantiteit in een patiëntenmonster het minst goed lineair te modelleren.

Hoofdstuk 6 rapporteert over een verdieping waarbij de metagenomische analyse wordt toegepast voor het kwantificeren van virussen op basis van een test die gebruikt maakt van kalibratiemonsters. Dit is gedaan in een cohort van transplantatiepatiënten, de *virusloads* in het plasma van invloed waren op de behandeling, waarbij alle patiëntenmonsters succesvol werden gekwantificeerd met metagenomische analyse.

Om na te gaan hoe metagenomische sequentie analyse functioneert bij de detectie van nieuwe of geëvolueerde virussen in patiëntenmonsters, is er in **hoofdstuk 7** onderzocht of SAR-CoV, MERS-CoV en SARS-CoV-2 rechtstreeks in patiënten materiaal gedetecteerd kon worden. Al deze corona virussen konden worden gedetecteerd door het gebruik van *capture probes*. SARS-CoV-2 monsters gaven een zeer hoge genoom dekking met gebruik van *probes* die waren ontworpen jaren vóór de ontdekking van het virus. Monsters met de desbetreffende virussen werden geanalyseerd met gebruikmaking van software waarvan de database uitsluitend virussen bevatte van alvorens de initiële ontdekking. Met de reguliere software bleken er steeds enkele *sequence reads* in te delen in de juiste virus familie. Middels *de novo assembly* werd een nieuw (deel) virusgenoom gebouwd op basis van de sequenties in het patiëntenmonster. Dat nieuwe (deel) virusgenoom werd vergeleken met de op dat moment al bekende virusgenomen, waardoor het homologie percentage berekend kon worden. Zolang een nieuw virus een zekere mate van homologie vertoont met een al bekend virus, zal het met metagenomische *sequencing* gedetecteerd kunnen worden.

Niet lang na het uitbreken van de SARS-CoV-2 pandemie zijn de specificaties van verschillende *sequence* technieken voor het SARS-CoV-2 genoom onderzocht. De resultaten van dat onderzoek staan beschreven in **hoofdstuk 8**. Een panel van 26 respiratoire patiëntenmonsters werden met vijf verschillende technieken en *sequencing* platformen behandeld. Zo is het metagenomische sequentie analyse protocol vergeleken met vier amplicon *sequence* technieken van drie verschillende fabrikanten. Metagenomische sequentie analyse was bij een hoge ct-waarde minder gevoelig in vergelijking met de amplicon methodes, maar de amplicon methodes leidden soms tot aspecifieke *sequence reads* en leverden om deze reden minder informatie over het SARS-CoV-2 genoom op. Wel is metagenomische sequentie

analyse in vergelijking met amplicon methodes geschikter om onontdekte virussen te detecteren, aangezien er geen nieuw protocol voor hoeft te worden opgezet. Dit is daardoor van nut in de vroege fase van een eventuele pandemie.

Metagenomische sequentie analyse kent voordelen: met één test kan tegelijkertijd gekeken worden naar alle bekende virussen, maar eveneens naar onverwachte, en nog niet ontdekte virussen. De test kan in de toekomst eventueel nog verder verbeterd worden, zo zal een verbeterde sensitiviteit van shotgun metagenomische *sequencing* een nog betere dekking geven, en indien een real-time sequentie analyse test wordt gebruikt zal dit de doorlooptijd aanzienlijk bekorten. Metagenomische sequentie analyse zou in ieder geval uitstekend als tweede test ingezet kunnen worden in het geval bij een patiënt wel een sterke verdenking bestaat op een virus-infectie, maar de initiële testen negatief blijken.

In de toekomst kan de test mogelijk worden gecombineerd worden met virus transcriptie factor analyse om zo te bepalen of een virus actief is of slechts latent aanwezig. Samenwerking met verschillende aanverwante medische disciplines biedt eveneens waardevolle perspectieven. Zo zou het wellicht mogelijk zijn om aan de hand van metagenomische sequentie analyse genetische aandoeningen in het immuunsysteem te detecteren of om simultaan te screenen op pathologische markers of farmacogenetische informatie.

Metagenomische sequentie analyse heeft duidelijk potentie binnen en buiten de microbiologie, waarbij de ontwikkeling en potentieel multidisciplinaire toepassing nog in de kinderschoenen staat.