Identifizierung seltener genetischer Varianten als Ursache für schweres COVID-19

Dissertation

zur Erlangung des Doktorgrades (Dr. med.)

der Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Jannik Boos

aus Marl

2025

Angefertigt mit der Genehmigung
der Medizinischen Fakultät der Universität Bonn

1. Gutachterin: Prof. Dr. Kerstin Urte Ludwig
2. Gutachter: Prof. Dr. Rayk Behrendt

Tag der mündlichen Prüfung: 15.09.2025

Aus dem Institut für Humangenetik

Inhaltsverzeichnis

Abkürzungsverzeichnis	4
1. Deutsche Zusammenfassung	5
1.1 Einleitung	5
1.1.1 COVID-19	5
1.1.2 COVID-19 Pathomechanismus	6
1.1.3 Humangenetische Studien zu COVID-19	6
1.1.4 Seltene Varianten und COVID-19	8
1.1.5 Fragestellung der Arbeit	9
1.2 Material und Methoden	10
1.2.1 Rekrutierung und Sequenzierung	10
1.2.2 Datenverarbeitung und Qualitätskontrollen	11
1.2.3 Studienpopulation und Subgruppen-Definition	12
1.2.4 Assoziationsanalyse der Einzelvarianten	12
1.2.5 Gen-basierte Analyse seltener Varianten	13
1.3 Ergebnisse	14
1.3.1 Assoziationsanalyse der Einzelvarianten	14
1.3.2 Gen-basierte Analyse seltener Varianten	15
1.3.3 Verfeinerte Pathogenitäts-Definitionen für <i>TLR7</i> -Varianten	16
1.3.4 Einfluss der Risikofaktoren in der Subgruppenanalyse	16
1.3.5 Verteilung der <i>TLR7</i> -Varianten innerhalb der <i>TLR7</i> -Proteinsequenz in Frauen	17
1.4 Diskussion	17
1.5 Zusammenfassung	20
1.6 Literaturverzeichnis der deutschen Zusammenfassung	20
2. Veröffentlichung	26
3. Erklärung zum Eigenanteil	27
4. Danksagung	29

Abkürzungsverzeichnis

3D-P	3D-Proteinstruktur (siehe Methoden)
bLoF	Biochemically Loss-of-Function (siehe Methoden)
C10+M1	CADD > 10 oder VEP-Impact high (siehe Methoden)
CAD	Coronary Artery Disease (Koronare Herzerkrankung)
CADD	Combined Annotation Dependent Depletion
CI ₉₅	95 % Confidence Interval (95 % Konfidenzintervall)
COVID-19	Coronavirus Disease of 2019
COVID-19 HGI	COVID-19 Host Genetics Initiative
DNA	Deoxyribonucleic Acid (Desoxyribonukleinsäure)
ECMO	Extrakorporale Membranoxygenierung
GATK	Genome Analysis Toolkit
GenOMICC	Genetics Of Mortality In Critical Care (Studiename)
GWAS	Genome-wide Association Study (Genomweite Assoziationsstudie)
hg19	Human reference genome version 19
I5AN	Interface 5 Ångström Neighborhood (siehe Methoden)
MAF	Minor Allele Frequency (Häufigkeit des seltenen Allels)
NIV	Noninvasive Ventilation (Nicht-invasive Beatmung)
OR	Odds Ratio (Quotenverhältnis)
p	P-Value (P-Wert)
PC	Principal Component (Hauptkomponente)
PCR	Polymerase Chain Reaction (Polymerase-Kettenreaktion)
RNA	Ribonucleic Acid (Ribonukleinsäure)
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
smMIPs	Single Molecule Molecular Inversion Probes
TLR7	Toll-like-receptor 7
VEP	Ensembl Variant Effect Predictor

1. Deutsche Zusammenfassung

1.1 Einleitung

1.1.1 COVID-19

COVID-19 (Coronavirus Disease of 2019) ist eine durch das Virus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) ausgelöste Infektionskrankheit. Diese kann zu individuell sehr unterschiedlichen klinischen Manifestationen führen, von asymptomatischem Verlauf über Fieber, Halsschmerzen, Husten oder Magen-Darm-Beschwerden bis hin zu schwerer Pneumonie, Multiorganversagen und Tod (Elrobaa und New, 2021). Nachdem 2019 erste Fälle in Wuhan, China, bekannt wurden (Zhu et al., 2020), entwickelte sich der lokale Ausbruch zu einer globalen Pandemie mit weltweit über 770 Mio. gemeldeten Infektionen und über 7 Mio. (ca. 0,1 % der Weltbevölkerung) assoziierten Todesfällen (World Health Organization, 2025, Stand 02.03.2025). Die Pandemie sowie die dazugehörigen Maßnahmen zur Eindämmung des Infektionsgeschehens führten zu schwerwiegenden gesellschaftlichen Herausforderungen (Su et al., 2022). Daher wurden die Erkrankung COVID-19 sowie das dazugehörige Virus SARS-CoV-2 zum Gegenstand intensiver weltweiter Forschung, zu deren Errungenschaften mittlerweile neue Therapie- und Präventionsmöglichkeiten gehören (Yuan et al., 2023).

Ein Bereich dieser Forschung ist die Frage nach den Ursachen für die individuell sehr unterschiedlichen Krankheitsverläufe nach Infektion mit SARS-CoV-2. Neben den demographischen Risikofaktoren Alter und Geschlecht sowie Vorerkrankungen wie u.a. Diabetes, Bluthochdruck, Übergewicht (Williamson et al., 2020; Yang et al., 2020; Zhou et al., 2020), wurden früh auch humangenetische Risikofaktoren für einen schweren Verlauf von COVID-19 identifiziert (Severe Covid-19 GWAS Group, 2020). Kenntnisse über Risikofaktoren können zu einem besseren Verständnis der Krankheit führen, insbesondere können humangenetische Studien Informationen darüber gewinnen, welche Gene im Pathomechanismus von COVID-19 eine Rolle spielen (Niemi et al., 2022).

1.1.2 COVID-19 Pathomechanismus

Durch die intensive Forschung der letzten Jahre ist bereits einiges zum Pathomechanismus der SARS-CoV-2 Infektion sowie schwerem COVID-19 bekannt. Steiner et al. (2024) fassen zusammen, dass das Virus nach Bindung an den Rezeptor ACE2 mit Hilfe von TMPRSS2 (direkt) oder Cathepsin L/B (endosomaler Weg) in die Zelle aufgenommen wird. Darauf folgen Replikation, Transkription und Translation, bis schließlich neu zusammengesetzte Viren die Zelle über den Golgi-Apparat verlassen. Nach Merad et al. (2022) wird die angeborene Immunantwort aktiviert durch Pattern Recognition Rezeptoren entweder (i) in dendritischen Zellen durch endosomale Toll-like Rezeptoren (TLRs) oder (ii) in infizierten Zellen durch zytosolische RNA-Rezeptoren wie z. B. RIG-I oder MDA5. Beide Wege führen zur Ausschüttung von Interferon Typ I und III sowie NF- κ B-abhängigen proinflammatorischen Zytokinen, und damit zur Aktivierung der angeborenen Immunabwehr. Die Bildung von Interferon Typ I und III wird jedoch durch das Virus teilweise blockiert (Steiner et al., 2024). Die adaptive Immunantwort wird aktiviert durch Aufeinandertreffen von Viruspartikeln und Antigen-präsentierenden Zellen oder B-Lymphozyten (Primorac et al., 2022) und führt zur Bildung eines Immun-Gedächtnisses gegen SARS-CoV-2, wobei eine erfolgreiche angeborene Immunantwort die adaptive Immunantwort beschleunigt (Diamond und Kanneganti, 2022). Die Schäden im menschlichen Körper aufgrund der Infektion entstehen nach Merad et al. (2022) dabei (i) durch unmittelbar Virus-bedingte Gewebeschäden und, abhängig vom Ausmaß dessen, (ii) durch die lokale und systemische Entzündungsreaktion. Bei schwerem COVID-19 kann hohe Viruslast und Hyperinflammation über Lungenversagen, systemische Koagulopathie, Gefäßschäden und Multiorganversagen bis zum Tod führen, wobei als Auslöser eines schweren Verlaufes eine unzureichende antivirale Immunantwort (insb. eine mangelnde Interferon Typ-I Antwort) und eine dysregulierte Entzündungsreaktion diskutiert werden (Fraser et al., 2023; Merad et al., 2022).

1.1.3 Humangenetische Studien zu COVID-19

Humangenetische Studien zu Infektionskrankheiten beschäftigen sich damit, welchen Einfluss Varianten im Genom des Menschen (des Wirts) auf die Erkrankung haben

(Wirtsgenetik, englisch "host genetics"). Im Gegensatz zur Wirtsgenetik befasst sich die Virusgenetik mit dem Virusgenom und dessen Variabilität. Die Virusgenetik stellt im Bezug auf COVID-19 ebenfalls ein wichtiges Forschungsfeld dar, wird aber in dieser Arbeit nicht näher diskutiert. Humangenetische Studien zu Krankheiten mit komplexen, multifaktoriellen Ursachen, wie COVID-19, können die Forschung auf verschiedenen Ebenen weiterbringen. Statistische Zusammenhänge zwischen der Ausprägung einer Krankheit und genetischen Varianten in einem Gen oder an einem Locus weisen auf eine Rolle des Gens oder Locus im Pathomechanismus hin (Niemi et al., 2022). Daraus können therapeutische Ansätze entwickelt werden, wie die Verwendung bereits bekannter Medikamente oder die Neuentwicklung medikamentöser Therapien (Carss et al., 2023; Ghossaini et al., 2023; Nelson et al., 2015; Niemi et al., 2022). Außerdem könnten Patient:innen mit genetischen Risikofaktoren auf diese getestet und damit schneller als Risikopatient:innen erkannt und gezielt behandelt werden (van der Made et al., 2022). Eine weitere Möglichkeit zur Nutzung genetischer Daten ist die statistische Differenzierung zwischen Korrelation und Kausalität von Risikofaktoren, welche zB mit Methoden der *Mendelian Randomisation* durchgeführt werden kann (Niemi et al., 2022, Richmond und Davey Smith, 2022).

Bei genetischen Studien kann zwischen der Untersuchung von seltenen Varianten, meist definiert als *Minor Allele Frequency* (MAF) $< 1\%$, und häufigen Varianten, meist definiert als $MAF > 1\%$, unterschieden werden (Frazer et al., 2009). Diese Einteilung reflektiert (i) die unterschiedlichen Techniken, die zur Datenerhebung nötig sind (array-basierte Genotypisierung vs. Sequenzierung) und (ii) die unterschiedlichen statistischen Herausforderungen bei der Datenanalyse (Timpson et al., 2018). Genomweite Assoziationsstudien (GWAS) mit häufigen genetischen Varianten können auf Basis von array-basierten Genotypdaten effizient in großen Kohorten durchgeführt werden. Auf dieser Grundlage umfasst die aktuelle GWAS-Metaanalyse der weltweiten und multizentrischen COVID-19 HGI mittlerweile 219.692 Patient:innen (davon 21.194 schwer erkrankt) und über 3 Millionen Kontrollen (COVID-19 Host Genetics Initiative, 2023). Diese Studie fand hochsignifikante Assoziationen mit moderaten Effektstärken zwischen COVID-19 (speziell: Suszeptibilität und Krankheitsschwere) und häufigen Varianten in 51 verschiedenen genetischen Loci, darunter Loci mit Genen, die in Verbindung mit dem

Eintritt des Virus in die Wirtszelle stehen, z. B. *SLC6A20* (3p21.31), *ABO* (9q34.2), *ACE2* (Xp22.2) und *TMPRSS2* (21q22.3); Loci mit Genen in Verbindung mit Schleimproduktion in Atemwegen, z. B. *MUC1/THBS3* (1q22); sowie Loci mit Genen in Verbindung zur Type-1-Interferon-Signalkaskade, z. B. *IFNAR2* (21q22.11) und *OAS1* (12q24.13).

1.1.4 Seltene Varianten und COVID-19

Im Gegensatz zu GWAS auf Basis von array-basierten Genotypdaten für häufige genetische Varianten werden für die Analyse von seltenen genetischen Varianten Daten aus Sequenzierungen benötigt, die mit einem höheren Ressourcen-Aufwand sowie einer erweiterten Aufklärung bei möglichen genetischen Zusatzbefunden erhoben werden müssen. Die zurzeit größte Sequenzierungs-Studie zu seltenen genetischen Varianten bei schwer erkrankten COVID-19-Patient:innen (Kousathanas et al., 2022) beinhaltet daher nur 7.491 Fälle. Trotz des größeren Aufwandes haben seltene Varianten für das Verständnis von COVID-19 eine große Bedeutung, denn nach Hindorff et al. (2011) sowie Timpson et al. (2018) sind Varianten mit hohen Effektstärken eher unter seltenen Varianten zu finden und nach Povysil et al. (2019) führen Untersuchungen zu seltene Varianten eher zu direkten Einsichten in biologische Mechanismen.

Durch Exom-Sequenzierung bei vier schwer an COVID-19 erkrankten jungen Männern wurden schon früh zwei seltene Varianten im Gen *TLR7* mit schwerem COVID-19 in Verbindung gebracht (van der Made et al., 2020). Das von diesem Gen kodierte Protein TLR7 erkennt einzelsträngige RNA, u.a. RNA von Coronaviren, und ist ein Aktivator der Interferon-Signalkaskade (van der Made et al., 2022). Weitere Studien bestärkten den Zusammenhang zwischen pathogenen Varianten in *TLR7* und schwerem COVID-19 (Mantovani et al., 2022; Matuozzo et al., 2023; Solanich et al., 2021). In Kohortenstudien wurden seltene *TLR7*-Varianten in 2,1 % (Fallerini et al., 2021) bzw. ca. 1 % (Asano et al., 2021) schwer an COVID-19 erkrankter Männer unter 60 Jahren gefunden. Im Verlauf gab es auch durch Studien in Modellsystemen Evidenz für *Tlr7* im COVID-19 Pathomechanismus, z. B. fanden Wang et al. (2024) in Mäusen einen Zusammenhang zwischen genetisch induzierter TLR7-Defizienz und COVID-19-Schwere bei verringerter Interferon-Bildung und erhöhter Viruslast.

In den ersten Studien zu *TLR7* wurden vor allem männliche Patienten untersucht und eine X-chromosomal-rezessive *TLR7*-Defizienz als Pathomechanismus vermutet (Asano et al., 2021). Später zeigten Butler-Laporte et al. (2022) jedoch auch bei Frauen eine Assoziation von heterozygoten seltenen *TLR7*-Varianten zu schwerem COVID-19. Dies ist mit einem X-chromosomal-rezessiven Pathomechanismus nicht gut zu erklären, da sich *TLR7* der X-Inaktivierung entzieht (Souyris et al., 2018). Es stellt sich daher die Frage, ob bei Frauen möglicherweise ein anderer biologischer Mechanismus zum Tragen kommt.

Bei der Untersuchung von seltenen genetischen Varianten im Zusammenhang mit schwerem COVID-19 gibt es bislang nur wenige systematische Exom/Genom-weite Sequenzierungs-Studien in größeren Kohorten. Die Ergebnisse sind insgesamt uneindeutig, so fanden Kosmicki et al. (2021) bei sehr konservativer Korrektur für multiples Testen keine klare Assoziation zu seltenen Varianten, jedoch bei etwas weniger konservativer Korrektur u.a. eine Assoziation von seltenen Varianten in *TLR7* zu schwerem COVID-19; Kousathanas et al. (2022) fanden einige wenige signifikante Assoziationen zu seltenen Varianten, jedoch nicht in *TLR7*; Butler-Laporte et al. (2022) fanden klare Assoziationen zu *MARK1* und *TLR7*, letztere bemerkenswerterweise sowohl bei Männern als auch bei Frauen; und Schmidt et al. (2024) fanden nominale Evidenz für Gene der Interferon-Signalkaskade, jedoch nicht für *TLR7*. Trotz dieser ersten Befunde weisen die genannten Studien individuelle Limitationen (Anzahl an teilnehmenden Patient:innen, Sequenzieretechnik (Exom vs. Genom), einheitliche Datenerhebung und -verarbeitung von Fällen und Kontrollen, oder Umfang der statistischen Aufarbeitung) auf, die ein vollständiges Verständnis der genetischen Architektur von schwerem COVID-19 bisher nur eingeschränkt ermöglichen.

1.1.5 Fragestellung der Arbeit

Diese Studie soll die Frage beantworten, ob seltene humangenetische Varianten mit schwerem COVID-19 assoziiert sind, und in welcher Subgruppe der Patient:innen diese den höchsten Anteil in der Ätiologie haben. In Erweiterung der zuvor genannten, in etwa zeitgleich durchgeführten Studien sollen dabei neben Alter auch andere, nicht primär genetische Risikofaktoren für schweres COVID-19 in der statistischen Auswertung

berücksichtigt werden und eine differenzierte Analyse für Frauen und Männer durchgeführt werden. Dafür wurde eine Sequenzierung von 55 Kandidatengenomen in der Kohorte der Severe Covid-19 GWAS Group (Severe Covid-19 GWAS Group, 2020; Degenhardt et al., 2022) durchgeführt. Diese große, ethnisch homogene Kohorte wurde 2020 in Italien und Spanien vor Verfügbarkeit von Impfungen rekrutiert und enthält insbesondere auch Daten zu Vorerkrankungen, die in dieser Studie zur statistischen Verfeinerung der Analyse genutzt wurden.

Diese Zusammenfassung beschränkt sich aus Platzgründen nur auf die wesentlichen Inhalte der Studie, Details sowie weitere Ergebnisse können der Originalpublikation entnommen werden (Boos et al., 2024). Dieser Text wurde soweit wie möglich in genderneutraler Sprache verfasst, an einigen Stellen wird jedoch aus Platz- und Formatierungsgründen (insb. in den Tabellen) das generische Maskulinum verwendet.

1.2 Material und Methoden

1.2.1 Rekrutierung und Sequenzierung

Die Untersuchungen dieser Arbeit sind auf der Grundlage der revidierten Deklaration von Helsinki des Weltärztebundes 1983 und den entsprechenden gesetzlichen Grundlagen durchgeführt worden, die Referenznummern der Zustimmung der entsprechenden Ethikräte dieser multizentrischen Studie sind im Supplement der Originalpublikation nachzulesen. Die Rekrutierung der Studienteilnehmer:innen sowie Probengewinnung wurde im Jahr 2020 zu Beginn der COVID-19 Pandemie durch kooperierende Forschungsgruppen in mehreren Kliniken in Italien und Spanien durchgeführt, wie beschrieben von der Severe Covid-19 GWAS Group (2020) und Degenhardt et al. (2022). Die Proben wurden am Institut für Klinische Molekularbiologie der Universität Kiel gesammelt und für die oben genannte GWAS-Studie (Severe Covid-19 GWAS Group, 2020; Degenhardt et al., 2022) genotypisiert. Dabei wurden bereits Qualitätskontrollen für Probenqualität, Verwandtschaft sowie Analysen zur Zuordnung zu genetisch-basierten Populationsgruppen durchgeführt. Für 9.104 Individuen wurde ein Teil der extrahierten DNA an das Institut für Humangenetik der Uni Bonn zur Sequenzierung weitergeleitet.

Basierend auf der im August 2020, zum Zeitpunkt des Studienstarts, vorliegenden Evidenzlage wurden 55 Gene mit vermuteter Rolle im COVID-19-Pathomechanismus ausgewählt und diese in Einklang mit den vorliegenden Patienteneinwilligungen in den Proben sequenziert. Für die Sequenzierung wurden *Single Molecule Molecular Inversion Probes* (smMIPs) verwendet, welche die Sequenzierung von gezielten Regionen, die Zuordnung zu den Proband:innen sowie durch die Verwendung von *Unique Molecular Identifiern* die Vermeidung von PCR-Artefakten ermöglichen (Hiatt et al., 2013). Die laborseitige Durchführung der Sequenzierung wurde von Julia Fazaal, Mitarbeiterin der AG Ludwig, betreut. Die technische und inhaltliche Auswertung der Daten, insbesondere durch Organisation der 9.104 Probanddaten und Auswertung der durchgeführten Testläufe mit dem Ziel der Konzentrations-Balancierung der Proband:innen-DNA und der verwendeten smMIPs für die eigentlichen Sequenzierläufe, sowie die gesamte weitere Datenverarbeitung und statistische Aufarbeitung, erfolgte durch den Autor dieser Arbeit. Die 3D-Struktur-basierten Analysen der *TLR7*-Varianten wurden von Gayatri Ramakrishnan (Radboud University Medical Center, Niederlande) und die unabhängige Replikation in der GenOMICC-Kohorte von Eamon Coughlan und Kolleg:innen (University of Edinburgh, Großbritannien) durchgeführt.

1.2.2 Datenverarbeitung und Qualitätskontrollen

Nach Abschluss der Datenerhebung mittels Sequenzierung wurde über mehrere Schritte die Datenverarbeitung inklusive ausführlicher Qualitätskontrollen von den Sequenzier-Rohdaten ("Reads") hin zu einer Variantentabelle ("Variant Call Format"), mit Variantendaten pro Person, durchgeführt. Zusammenfassend gehörte dazu das Alignment der *Reads* an eine Referenzsequenz mittels *bwa-mem2* Version 2.1 (Vasimuddin et al., 2019), das Trimmen der Arme der smMIPs, das Deduplizieren der *Reads* mittels *umi_tools* Version 0.2.3 (Smith et al., 2017) zur Reduktion von PCR-Artefakten, sowie das Generieren der Varianteninformationen aus den *Reads* ("Variant Calling") mit dem Genome Analysis Toolkit (GATK) Version 4.2.4.1 (McKenna et al., 2010). Die Qualitätskontrollen enthielten insbesondere das Filtern nach (i) der Abdeckung der genetischen Abschnitte durch die Sequenzierung ("Coverage"), (ii) Qualität der

einzelnen Varianten nach dem *Variant Calling* anhand verschiedener GATK-Qualitätsmarker sowie anhand der Verteilung der alternativen Allele in den *Reads* (erlaubte Verteilung: 0 - 5 % bzw. 95 - 100 % für homozygote Varianten, 20 - 80 % für heterozygote Varianten) und (iii) das Entfernen aller Varianten mit validem *Variant Calling* in weniger als 95 % aller Individuen. Zuletzt wurden alle Individuen mit diskordantem phänotypischen und genotypischen Geschlecht ausgeschlossen. Genauere Informationen zu Datenverarbeitung und Qualitätskontrollen finden sich im Supplement der Originalpublikation (Boos et al., 2024).

1.2.3 Studienpopulation und Subgruppen-Definition

Nach Abschluss der Qualitätskontrollen blieben zur weiteren Analyse 52 Gene (drei Gene wurden aufgrund geringer *Coverage* ausgeschlossen) sowie 1.772 schwer an COVID-19 erkrankte Patient:innen (definiert als aufgrund von COVID-19 hospitalisierte Patient:innen mit Bedarf nach Atemunterstützung mittels O₂-Nasenbrille, NIV, Intubation oder ECMO) und 5.347 Populations-basierte Kontroll-Individuen. Von den insgesamt 7.119 Individuen waren 38 % weiblich. Neben weiteren klinischen Daten wurde für 1.706 der 1.772 schwer an COVID-19 Erkrankten während der Rekrutierung insbesondere erhoben, ob eine Koronare Herzerkrankung (CAD), Diabetes oder Hypertonie als Vorerkrankung bestand. Die Subgruppe der schwer an COVID-19 Erkrankten ohne Risikofaktoren wurde definiert als schwer an COVID-19 Erkrankte unter 60 Jahren ohne CAD, Diabetes oder Hypertonie. In der Originalpublikation (Boos et al., 2024) wurden weitere Subgruppen basierend auf vorbekannten Risikofaktoren sowie basierend auf der Krankheitsschwere definiert, die hier aus Platzgründen nicht weiter ausgeführt werden. Alle Analysen wurden zusätzlich auch für Männer und Frauen separat durchgeführt.

1.2.4 Assoziationsanalyse der Einzelvarianten

Für alle Varianten, die in mehr als einem Individuum gefunden wurden (sogenannte "nicht-Singleton-Varianten") wurde eine logistische Regression mit Firth-Korrektur und additivem Modell in der Implementierung von Plink v2.0 (Chang et al., 2015) mit den Kovariaten

Alter, Geschlecht, Alter², Alter*Geschlecht und den ersten 10 PCs aus der vorherigen Genotypisierung (Degenhardt et al., 2022) durchgeführt. Die Analyse wurde aufgrund des unterschiedlichen Populations-genetischen Hintergrundes sowie der unterschiedlichen Fall-Kontroll-Verhältnisse zunächst getrennt für die italienische und spanische Subkohorte durchgeführt und die Ergebnisse dann mittels METAL (Willer et al., 2010) meta-analysiert. Zwei p-Grenzwerte für multiples Testen wurden festgelegt mit $6,7 \times 10^{-6}$ als strikter Grenzwert, welcher für alle Testungen (ca. 1.211 Varianten für 12 Subgruppen-Analysen) korrigiert und $4,1 \times 10^{-5}$ als weniger strikter Grenzwert unter Berücksichtigung wahrscheinlicher Korrelationen zwischen unterschiedlichen Subgruppen-Analysen.

1.2.5 Gen-basierte Analyse seltener Varianten

Bei der Untersuchung von seltenen genetischen Varianten wurden, zur Erhöhung der statistischen Power, Gen-basierte zusammenfassende Analysen durchgeführt ("Rare Variant Collapsing Analysis"), wie beschrieben in Povysil et al. (2019). Dazu wird das Tragen mindestens einer seltenen und pathogenen Variante in einem bestimmten Gen als Varianten-Trägerschaft (unabhängige Variable) definiert. Dann wird in der statistischen Analyse die Gruppe der Varianten-Träger:innen den Nicht-Varianten-Träger:innen gegenübergestellt und die Verteilung der schwer an COVID-19 Erkrankten (abhängige Variable) innerhalb dieser Gruppen untersucht. Als Definitionen für seltene Varianten zum Einbezug in *Rare Variant Collapsing* Analysen werden in dieser Arbeit $MAF < 1 \%$ (selten) und $MAF < 0,1 \%$ (sehr selten) verwendet, wie auch in Butler-Laporte et al. (2022). Die Idee hinter der Verwendung eines MAF-Filters ist die negative Korrelation zwischen Häufigkeit und Schädlichkeit von Varianten durch Selektion (Povysil et al., 2019). Die Definition der Varianten-Schädlichkeit anhand von anderen Merkmalen, neben der Häufigkeit, ist deutlich komplexer. Möglich sind (i) die Verwendung von vorberechneten Varianten-Scores aus der Literatur (hier verwendet: CADD, siehe Rentzsch et al., 2021), (ii) Auswirkung der Variante auf die Transkription oder Translation (hier verwendet: VEP-Impact, siehe McLaren et al., 2016), (iii) die Verwendung von Proteinstruktur-Daten mit Berechnung verschiedener biochemischer Struktur-Parameter oder (iv) die Nutzung von funktionellen Daten zu Auswirkungen von genetischen

Varianten auf unterschiedliche *in vitro* Parameter wie Enzymaktivität oder Produktion von Botenstoffen. Die hier verwendeten Definitionen sind: (i) C10+M1, mit einem CADD-Score größer als 10 oder einem VEP-Impact von high (starke Beeinträchtigung der Transkription/Translation); (ii) bLoF (*biochemically Loss-of-Function*) mit *TLR7*-Varianten, die in funktionellen *in vitro* Studien (Asano et al., 2021; Fallerini et al., 2021; Mantovani et al., 2022) als *Loss-of-Function*-Varianten beschrieben wurden; (iii) 3D-P (3D-Proteinstruktur) mit *TLR7*-Varianten, die in 3D-Modellierung der Proteinstruktur zu einer eingeschränkten Integrität und Proteinstabilität führen; sowie I5AN als *TLR7*-Varianten in 5 Ångström-Nachbarschaft zum Dimerisierungs-Interface. Details sowie weitere Definitionen sind in der Originalpublikation zu finden (Boos et al., 2024). Um die Assoziation zwischen Varianten-Trägerschaft in einem Gen und schwerem COVID-19 zu untersuchen, wurde der Cochran-Mantel-Haenszel-Test (Mantel und Haenszel, 1959) im dominanten Modell in der Implementation von Plink v1.9 genutzt. Dieser Test hat den Vorteil, Subgruppen (hier: italienische/spanische Kohorte sowie Männer/Frauen) differenziert zu analysieren ("Stratifizierung") und mit seltenen Ereignissen umgehen zu können (Efthimiou, 2018). Die p-Grenzwerte für multiples Testen für die Gen-basierte Analyse seltener Varianten wurden festgelegt mit $8,7 \times 10^{-6}$ als strikter Grenzwert, der für alle Testungen (ca. 2×5 Varianten-Kategorien in 52 Genen in 12 Subgruppen-Analysen) korrigiert und $9,6 \times 10^{-4}$ als weniger strikter Grenzwert unter Berücksichtigung wahrscheinlicher Korrelationen zwischen unterschiedlichen Subgruppen-Analysen und Varianten-Definitionen.

1.3 Ergebnisse

1.3.1 Assoziationsanalyse der Einzelvarianten

In der Studienkohorte fanden sich innerhalb der sequenzierten Region 3.218 unterschiedliche genetische Varianten, wovon 2.007 Varianten Singletons waren, also nur in einem Individuum gefunden wurden. In der logistischen Regression über die 1.211 Nicht-Singleton-Varianten zeigten, nach Korrektur für multiples Testen, sieben Varianten eine signifikante Assoziation mit schwerem COVID-19. Alle diese Varianten waren häufig (MAF > 1 %) und sind in der unabhängigen Kohorte der COVID-19 HGI Release 7

ebenfalls signifikant und mit gleicher Effektrichtung assoziiert mit schwerem COVID-19. Für Varianten mit hoher Effektstärke und nominaler Signifikanz (odds ratio (OR) > 5 und $p < 0,05$) wurde eine follow-up Metaanalyse mit Daten der Regeneron-Studie (Kosmicki et al., 2021) durchgeführt. Dabei zeigte eine seltene Variante in *TBK1* (p.Arg358His, chr12:64878163:G:A [hg19]) eine starke Assoziation zu schwerem COVID-19 ($p = 1,67 \times 10^{-5}$; OR = 26,3; 95 % Konfidenzintervall (CI₉₅) = [5,93; 116,2]).

1.3.2 Gen-basierte Analyse seltener Varianten

In der Gen-basierten *Rare Variant Collapsing* Analyse zeigten sich pathogene (Kategorie C10+M1, MAF < 0,1 %) Varianten im Gen *TLR7* in der Subgruppe der schwer an COVID-19 Erkrankten ohne Risikofaktoren signifikant häufiger als in der Populations-basierten Kontrollgruppe (siehe Tab. 1). Diese Assoziation war in dieser Konstellation und mit noch höherem OR auch in der Subgruppe der Frauen signifikant (siehe Tab. 1). Unter Verwendung des weniger strikten Signifikanzniveaus (siehe Methoden) zeigten sich neben *TLR7* auch Hinweise auf eine Assoziation von seltenen pathogenen Varianten in den Genen *IFNAR2* und *IFIH1* mit schwerem COVID-19.

Tab. 1: Ergebnisse der Gen-basierten Analyse seltener Varianten (MAF < 0,1 %) in schwer COVID-19 Erkrankten ohne Risikofaktoren vs. Populations-basierte Kontrollen. Abkürzungen: OR: *Odds Ratio*, CI₉₅: 95 % Konfidenzintervall, C10+M1: Varianten mit CADD > 10 oder VEP-Impact *high*; bLoF: *biochemically Loss-of-Function*, 3D-P: 3D-Proteinstruktur (siehe Methoden)

Varianten-Kategorie (Subgruppe)	P-Wert	OR [CI ₉₅]	Trägerschaft Fallgruppe: Träger / Alle	Trägerschaft Kontrollen: Träger / Alle
C10+M1	1.27×10^{-10}	12,3 [4,7; 32,2]	9 / 378	12 / 5.347
C10+M1 (nur Frauen)	1.75×10^{-9}	24,8 [5,9; 105,2]	4 / 126	5 / 2.102
bLoF	1.73×10^{-10}	34,6 [6,8; 177,2]	4 / 378	3 / 5.347
3D-P	1.74×10^{-15}	46,5 [10,9; 198,7]	7 / 378	4 / 5.347

1.3.3 Verfeinerte Pathogenitäts-Definitionen für *TLR7*-Varianten

Basierend auf *in vitro* Studien wurde eine neue Definition (bLoF, siehe Methoden) für schädliche *TLR7*-Varianten erstellt. Die auf dieser Grundlage erneut durchgeführte *Rare Variant Collapsing* Analyse brachte für die schwer an COVID-19 Erkrankten ohne Risikofaktoren ein ähnliches Ergebnis, allerdings mit deutlich höherem OR (siehe Tab. 1). Die Definition für schädliche *TLR7*-Varianten unter Einbeziehung von 3D-Proteinstruktur-Daten (3D-P, siehe Methoden) führte zu einer weiteren Verstärkung der zuvor gefundenen Assoziation (siehe Tab. 1).

1.3.4 Einfluss der Risikofaktoren in der Subgruppenanalyse

Bei Stratifizierung der Betroffenen basierend auf den oben genannten Risikofaktoren (Diabetes, Hypertonie, CAD sowie Alter) zeigte sich für schädliche *TLR7*-Varianten (3D-P, MAF < 0,1 %) innerhalb der 378 schwer an COVID-19 Erkrankten ohne Risikofaktoren eine deutliche Anreicherung der Variantenträger im Vergleich zu den Subgruppen ohne Berücksichtigung der Risikofaktoren (siehe Tab. 2). Eine graphische Darstellung findet sich in Abbildung 4B der Originalpublikation (Boos et al., 2024).

Tab. 2: Anteil der Variantenträger:innen (Varianten-Kategorie: 3D-P, MAF < 0,1 %) in unterschiedlichen Subgruppen.

Subgruppe	Trägerschaft: Anteil	Trägerschaft: Träger / Alle
Kontrollen (Populations-basiert)	0,07 %	4 / 5.347
Schwer COVID-19 Erkrankte	0,45 %	8 / 1772
davon Schwer COVID-19 Erkrankte unter 60 Jahren	1,25 %	7 / 558
davon Schwer COVID-19 Erkrankte ohne Risikofaktoren	1,85 %	7 / 378

1.3.5 Verteilung der *TLR7*-Varianten innerhalb der *TLR7*-Proteinsequenz in Frauen

In der Subgruppe der schwer an COVID-19 erkrankten Frauen fiel unter den *TLR7*-Varianten eine Überrepräsentation von Varianten in den *Leucine-Rich Repeat* Regionen auf, welche maßgeblich an der Dimerisierung von *TLR7* beteiligt sind (Zhang et al., 2016). Um eine Beeinträchtigung der Dimerisierung von *TLR7* als möglichen Pathomechanismus genauer zu untersuchen, wurde nach seltenen Varianten in 5 Ångström Nachbarschaft des Dimerisierungs-Interfaces (I5AN-Varianten) gesucht und die Verteilung dieser in der Kohorte analysiert. Zwei solche Varianten wurden unter den schwer an COVID-19 erkrankten Frauen ohne Risikofaktoren gefunden, zwei weitere in den männlichen Kontrollpersonen. Damit ergibt sich bei den Frauen eine signifikante ($p = 2.1 \times 10^{-6}$) Überrepräsentation in den schwer an COVID-19 Erkrankten ohne Risikofaktoren (2 von 126) im Vergleich zu den weiblichen Kontrollen (0 von 2.101). Ein unabhängiger Replikations-Versuch dieser I5AN-Varianten in der Kohorte der GenOMICC-Studie (Kousathanas et al., 2022) brachte keine neuen Erkenntnisse, da dort keine I5AN-Variante unter den weiblichen Studienteilnehmerinnen gefunden wurde.

1.4 Diskussion

Diese Arbeit untersucht die Frage, ob seltene genetische Varianten in bestimmten Genen mit einem schweren Verlauf der COVID-19 Erkrankung zusammenhängen. Insgesamt konnte *TLR7* als relevantes Gen im COVID-19 Pathomechanismus bestätigt werden. Die Assoziation konnte durch Stratifizierung nach Risikofaktoren und Hinzunahme von Daten zum Einfluss von Varianten auf die 3D-Proteinstruktur deutlich verstärkt herausgearbeitet werden. Außerdem zeigten sich in Frauen seltene *TLR7*-Varianten in Nachbarschaft zum Dimerisierungs-Interface zu schwerem COVID-19 assoziiert.

Für eine Assoziation pathogener Varianten im Gen *TLR7* mit schwerem COVID-19 gab es bereits vor dieser Studie deutliche Evidenz, wie in der Einleitung dargestellt. Einige Studien bezogen sich insbesondere auf junge Erkrankte, zB. fanden Asano et al. (2021) *TLR7*-Defizienz in ca. 1 % schwer an COVID-19 erkrankter Männer unter 60 Jahren. In Erweiterung dazu wurden in der hier vorliegenden Studie neben dem Alter zusätzlich die

als Risikofaktoren beschriebenen Vorerkrankungen CAD, Diabetes und Hypertonie in die stratifizierte Analyse eingebunden. Während in den schwer Erkrankten ohne Stratifizierung nur 0,45 % Variantenträger:innen (3D-P, MAF < 0,1 %) gefunden wurden, waren es in der Subgruppe der schwer Erkrankten ohne Risikofaktoren 1,85 % (in den Kontrollen 0,07 %). Dies weist darauf hin, dass eine TLR7-Defizienz als Ursache einer schweren Erkrankung insbesondere bei Patient:innen ohne andere Risikofaktoren in Frage kommt. Mit mehr Informationen zu weiteren Risikofaktoren und unter Einbeziehung von polygenen Risikoscores zur Abschätzung des Risikos durch häufige genetische Varianten könnten zukünftige Studien diese Ergebnisse noch weiter verfeinern.

In dieser, sowie in einer weiteren Studie (Butler-Laporte et al., 2022), waren heterozygote schädliche Varianten in *TLR7* auch in der Subgruppe der Frauen signifikant mit schwerem COVID-19 assoziiert. Der von Asano et al. (2021) vorgeschlagene Pathomechanismus für Männer als rezessive X-chromosomale *TLR7*-Defizienz trifft bei Frauen wahrscheinlich nicht zu, da sich *TLR7* der X-Inaktivierung entzieht (Souyris et al., 2018). Eine andere Möglichkeit wäre ein dominant-negativer Effekt, der die Funktionalität von TLR7 auch bei heterozygoten Variantenträgerinnen deutlich einschränkt. Seltene heterozygote *TLR7*-Varianten könnten bei Frauen durch Beeinträchtigung der Dimerisierung von TLR7 einen Effekt ähnlich zu *Loss-of-Function*-Varianten in Männern haben und damit das Risiko für schweres COVID-19 deutlich erhöhen. Als Hinweis darauf fand sich in dieser Arbeit eine signifikante Anreicherung von *TLR7*-Varianten nahe dem Dimerisierungs-Interface in schwer an COVID-19 erkrankten jungen Frauen ohne Risikofaktoren. Folgeuntersuchungen in größeren Studien sind nötig, um diesen Mechanismus zu bestätigen, ein Replikations-Versuch in der Kohorte der GenOMICC-Studie war aufgrund des Fehlens von I5AN-Varianten innerhalb der Frauen der Kohorte nicht informativ. *In vitro* Untersuchungen der funktionellen Effekte der hier gefundenen I5AN-Varianten könnten weitere Informationen zu Plausibilität und Funktion des vorgeschlagenen Mechanismus liefern.

Auch für einzelne Varianten konnten signifikante Assoziationen gefunden werden, insbesondere für sieben häufige Varianten, die die Ergebnisse vorheriger Studien bestätigten. Des Weiteren konnte eine einzelne seltene Variante im Gen *TBK1* durch Metaanalyse mit Ergebnissen der Regeneron-Studie (Kosmicki et al., 2021) mit schwerem

COVID-19 in Verbindung gebracht werden. *TBK1* wurde bereits von Schmidt et al. (2021) im Zusammenhang mit schwerem COVID-19 bei einem Kind beschrieben.

Diese Studie enthält einige Limitationen. Zunächst konnten viele Gene, die nach August 2020 Evidenz für eine Beteiligung im COVID-19 Pathomechanismus erhielten, nicht untersucht werden, da diese Studie auf 55 im August 2020 ausgewählte Gene beschränkt war. Auch Gene, für die bislang noch keine Rolle im COVID-19-Pathomechanismus bekannt ist, konnten durch diesen Ansatz nicht gefunden werden. Umfassendere Ergebnisse würden eine Sequenzierung des Exoms oder Genoms bringen, diese waren jedoch durch die vorliegenden Einwilligungen der Studienteilnehmer:innen nicht abgedeckt. Ein weiteres Problem dieser Arbeit ist die Definition der Kontrollgruppe als Populations-basierte Kohorte. Dadurch sind auch Individuen in der Kontrollgruppe, welche eventuell im weiteren Verlauf der Pandemie einen schweren COVID-19 Verlauf entwickelten. Eine Kontroll-Definition als SARS-CoV-19 positiv ohne schweren Verlauf würde vermutlich die Power der Analysen erhöhen, dies war allerdings während der Rekrutierung dieser Studie organisatorisch nicht möglich. Eine weitere Einschränkung war der Umfang der zur Verfügung stehenden klinischen Daten. Möglicherweise hätten Daten zu weiteren Risikofaktoren, wie Übergewicht (Williamson et al., 2020), die Ergebnisse noch weiter verfeinert. Eine große Herausforderung dieser Arbeit war die Einschätzung der Pathogenität von Varianten für die Gen-basierte *Rare Variant Collapsing Analyse*. *In silico* Scores bieten hierfür einen Ansatz, sind aber teils fehlerbehaftet (Rentzsch et al., 2021). Daher nutzte diese Arbeit einerseits Definitionen basierend auf solchen Scores, wie in vorherigen, ähnlichen Studien (Kosmicki et al., 2021; Kousathanas et al., 2022), erweiterte die Varianten-Definitionen aber für *TLR7* um zwei Definitionen basierend auf (i) in anderen Arbeiten publizierten funktionellen *in vitro* Ergebnissen und (ii) 3D-Proteinstruktur-Daten. Diese Ansätze führten zu höheren Effektstärken und scheinen daher die tatsächlich pathogenen Varianten besser herauszufiltern. Damit zeigten sie sich als vielversprechende Ansätze für zukünftige Studien zu seltenen Varianten, waren aber auf diese Studie (bzw. *TLR7*) maßgeschneidert und wenig standardisiert. Neuere Methoden, welche den strukturbasierten Ansatz systematisieren (z. B. Schmidt et al., 2023) oder funktionelle *in vitro* Untersuchungen skalieren ("Multiplexed assays of variant

effects”, z. B. Tabet et al., 2022), könnten zu noch besseren Schädlichkeits-Definitionen und damit zu erhöhter Power der Analyse von seltenen Varianten führen.

1.5 Zusammenfassung

Diese Arbeit untersucht den bisher nur zum Teil verstandenen Zusammenhang zwischen seltenen Wirts-genetischen Varianten und schwerem COVID-19 in 52 Kandidatengenen in einer spanisch/italienischen Kohorte mit 1.772 schwer an COVID-19 Erkrankten und 5.347 Populations-basierten Kontrollen. Insbesondere unter Stratifizierung nach Risikofaktoren wie Alter und Vorerkrankungen zeigten seltene, schädliche Varianten im Gen *TLR7* eine signifikante Assoziation zu schwerem COVID-19. Zusätzliche funktionelle sowie 3D-Proteinstruktur-basierte Daten zu Varianten-Pathogenität führten zu höheren Effektstärken. Während bei Männern in Vorarbeiten insbesondere ein X-chromosomal rezessiver Mechanismus diskutiert wurde, stellt diese Arbeit eine neue Hypothese für einen Pathomechanismus bei Frauen auf. Demnach könnten bei Frauen Varianten in Nähe des Dimerisierungs-Interface einen dominant-negativen Effekt durch Beeinträchtigung der Dimerisierung von *TLR7* zeigen. Weiterhin wurden in den Genen *IFIH1*, *IFNAR2* und *TBK1* Hinweise auf Assoziationen zwischen seltenen Varianten und schwerem COVID-19 gefunden. Insgesamt bietet diese Arbeit deutliche Evidenz für eine Beteiligung von seltenen pathogenen *TLR7*-Varianten in der Ätiologie von schwerem COVID-19 und zeigt, dass Risikostratifizierung sowie differenzierte Variantenbewertung durch funktionelle *in vitro* Untersuchungen oder 3D-Proteinstruktur-Analysen die statistische Power von Studien zu seltenen genetischen Varianten erhöhen können.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

Asano T, Boisson B, Onodi F, ..., Casanova J-L. X-linked recessive *TLR7* deficiency in ~1% of men under 60 years old with life-threatening COVID-19. *Sci Immunol* 2021; 6: eabl4348

Boos J, van der Made CI, Ramakrishnan G, ..., Ludwig KU. Stratified analyses refine association between TLR7 rare variants and severe COVID-19. *HGG Adv* 2024; 5: 100323

Butler-Laporte G, Povysil G, Kosmicki JA, ..., Richards JB. Exome-wide association study to identify rare variants influencing COVID-19 outcomes: Results from the Host Genetics Initiative. *PLoS Genet* 2022; 18: e1010367

Carss KJ, Deaton AM, Del Rio-Espinola A, ..., Yuan J. Using human genetics to improve safety assessment of therapeutics. *Nat Rev Drug Discov* 2023; 22: 145–162

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7

COVID-19 Host Genetics Initiative. A second update on mapping the human genetic architecture of COVID-19. *Nature* 2023; 621: E7–E26

Degenhardt F, Ellinghaus D, Juzenas S, ..., Franke A. Detailed stratified GWAS analysis for severe COVID-19 in four European populations. *Hum Mol Genet* 2022; 31: 3945–3966

Diamond MS, Kanneganti T-D. Innate immunity: the first line of defense against SARS-CoV-2. *Nat Immunol* 2022; 23: 165–176

Efthimiou O. Practical guide to the meta-analysis of rare events. *Evid Based Ment Health* 2018; 21: 72–76

Elrobaa IH, New KJ. COVID-19: Pulmonary and Extra Pulmonary Manifestations. *Front Public Health* 2021; 9: 711616

Fallerini C, Daga S, Mantovani S, ..., Frullanti E. Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: findings from a nested case-control study. *Elife* 2021; 10: e67569

Fraser R, Orta-Resendiz A, Dockrell D, Müller-Trutwin M, Mazein A. Severe COVID-19 versus multisystem inflammatory syndrome: comparing two critical outcomes of SARS-CoV-2 infection. *Eur Respir Rev* 2023; 32: 220197

Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 2009; 10: 241–251

Ghousaini M, Nelson MR, Dunham I. Future prospects for human genetics and genomics in drug discovery. *Curr Opin Struct Biol* 2023; 80: 102568

Hiatt JB, Pritchard CC, Salipante SJ, O’Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res* 2013; 23: 843–854

Hindorff LA, Gillanders EM, Manolio TA. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. *Carcinogenesis* 2011; 32: 945–954

Kosmicki JA, Horowitz JE, Banerjee N, ..., Ferreira MAR. Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. *Am J Hum Genet* 2021; 108: 1350–1355

Kousathanas A, Pairo-Castineira E, Rawlik K, ..., Baillie JK. Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature* 2022; 607: 97–103

Mantel N, Haenszel W. Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *J Natl Cancer Inst* 1959; 22: 719–748

Mantovani S, Daga S, Fallerini C, ..., Renieri A. Rare variants in Toll-like receptor 7 results in functional impairment and downregulation of cytokine-mediated signaling in COVID-19 patients. *Genes Immun* 2022; 23: 51–56

Matuozzo D, Talouarn E, Marchal A, ..., Cobat A. Rare predicted loss-of-function variants of type I IFN immunity genes are associated with life-threatening COVID-19. *Genome Med* 2023; 15: 22

McKenna A, Hanna M, Banks E, ..., DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20: 1297–1303

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17: 122

Merad M, Blish CA, Sallusto F, Iwasaki A. The immunology and immunopathology of COVID-19. *Science* 2022; 375: 1122–1127

Nelson MR, Tipney H, Painter JL, ..., Sanseau P. The support of human genetic evidence for approved drug indications. *Nat Genet* 2015; 47: 856–860

Niemi MEK, Daly MJ, Ganna A. The human genetic epidemiology of COVID-19. *Nat Rev Genet* 2022; 23: 533–546

Povysil G, Petrovski S, Hostyk J, Aggarwal V, Allen AS, Goldstein DB. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet* 2019; 20: 747–759

Primorac D, Vrdoljak K, Brlek P, Pavelić E, Molnar V, Matišić V, Erceg Ivkošić I, Parčina M. Adaptive Immune Responses and Immunity to SARS-CoV-2. *Front Immunol* 2022; 13: 848582

Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* 2021; 13: 31

Richmond RC, Davey Smith G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harb Perspect Med* 2022; 12: a040501

Schmidt A, Casadei N, Brand F, ..., Ludwig KU. Systematic assessment of COVID-19 host genetics using whole genome sequencing data. *PLoS Pathog* 2024; 20: e1012786

Schmidt A, Peters S, Knaus A, ..., Bagci S. TBK1 and TNFRSF13B mutations and an autoinflammatory disease in a child with lethal COVID-19. *NPJ Genom Med* 2021; 6: 55

Schmidt A, Röner S, Mai K, Klinkhammer H, Kircher M, Ludwig KU. Predicting the pathogenicity of missense variants using features derived from AlphaFold2. *Bioinformatics* 2023; 39: btad280

Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, ..., Karlsen TH. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med* 2020; 383: 1522–1534

Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 2017; 27: 491–499

Solanich X, Vargas-Parra G, van der Made CI, ..., Lázaro C. Genetic Screening for TLR7 Variants in Young and Previously Healthy Men With Severe COVID-19. *Front Immunol* 2021; 12: 719115

Souyris M, Cenac C, Azar P, Daviaud D, Canivet A, Grunenwald S, Pienkowski C, Chaumeil J, Mejía JE, Guéry J-C. TLR7 escapes X chromosome inactivation in immune cells. *Sci Immunol* 2018; 3: eaap8855

Steiner S, Kratzel A, Barut GT, Lang RM, Aguiar Moreira E, Thomann L, Kelly JN, Thiel V. SARS-CoV-2 biology and host interactions. *Nat Rev Microbiol* 2024; 22: 206–225

Su R, Obrenovic B, Du J, Godinic D, Khudaykulov A. COVID-19 Pandemic Implications for Corporate Sustainability and Society: A Literature Review. *Int J Environ Res Public Health* 2022; 19: 1592

Tabet D, Parikh V, Mali P, Roth FP, Claussnitzer M. Scalable Functional Assays for the Interpretation of Human Genetic Variation. *Annu Rev Genet* 2022; 56: 441–465

Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet* 2018; 19: 110–124

van der Made CI, Netea MG, van der Veerdonk FL, Hoischen A. Clinical implications of host genetic variation and susceptibility to severe or critical COVID-19. *Genome Medicine* 2022; 14: 96

van der Made CI, Simons A, Schuurs-Hoeijmakers J, ..., Hoischen A. Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA* 2020; 324: 663–673

Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *Proc IPDPS (Conf) 2019*; 2019: 314–324

Wang C, Khatun MS, Ellsworth CR, ..., Qin X. Deficiency of Tlr7 and Irf7 in mice increases the severity of COVID-19 through the reduced interferon production. *Commun Biol* 2024; 7: 1162

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26: 2190–2191

Williamson EJ, Walker AJ, Bhaskaran K, ..., Goldacre B. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020; 584: 430–436

World Health Organization, 2025: WHO COVID-19 dashboard. <https://data.who.int/dashboards/covid19/> (Zugriffsdatum: 22.03.2025)

Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Ji R, Wang H, Wang Y, Zhou Y. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis* 2020; 94: 91–95

Yuan Y, Jiao B, Qu L, Yang D, Liu R. The development of COVID-19 treatment. *Front Immunol* 2023; 14: 1125246

Zhang Z, Ohto U, Shibata T, ..., Shimizu T. Structural Analysis Reveals that Toll-like Receptor 7 Is a Dual Receptor for Guanosine and Single-Stranded RNA. *Immunity* 2016; 45: 737–748

Zhou F, Yu T, Du R, ..., Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020; 395: 1054–1062

Zhu N, Zhang D, Wang W, ..., China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; 382: 727–733

2. Veröffentlichung

Dieser Publikationsdissertation liegt die folgende, unabhängig begutachtete Veröffentlichung zugrunde:

Boos J, van der Made CI, Ramakrishnan G, ..., Ludwig KU. Stratified analyses refine association between TLR7 rare variants and severe COVID-19. HGG Adv 2024; 5: 100323

<https://doi.org/10.1016/j.xhgg.2024.100323>

3. Erklärung zum Eigenanteil

Die Arbeit wurde am Institut für Humangenetik unter Betreuung von meiner Doktormutter Prof. Dr. rer. nat. Kerstin U. Ludwig und unterstützender Betreuung von Dr. med. Axel Schmidt durchgeführt. Die zentralen konzeptionellen Elemente wurden von mir erstellt, ebenso führte ich federführend die Analysen und Interpretationen durch. Dies ist u.a. auch dadurch repräsentiert, dass ich alleiniger Erstautor der resultierenden Publikation bin.

Die Konzeption der Studie erfolgte in Zusammenarbeit mit meinen Betreuer:innen. Für spezifische Teile der Studie, welche sich aus den Ergebnissen der Hauptuntersuchung ergaben, wurden weitere Expert:innen hinzugezogen, welche spezifische Aspekte der Studie mitgestalteten. Diese sind im Folgenden genauer beschrieben.

Die Rekrutierung der Studienteilnehmer:innen, die DNA-Extraktion, die Genotypisierung sowie darauf beruhende Qualitätskontrollen wurden durch Mitglieder des Konsortiums "Severe Covid-19 GWAS Group" durchgeführt und von Prof. Dr. Andre Franke koordiniert. Die laborseitige Probenvorbereitung für die Sequenzierung wurde von einer Labormitarbeiterin der AG Ludwig, Julia Fazaal, durchgeführt, die eigentliche Sequenzierung von Mitarbeiter:innen der NGS Core Facility des Universitätsklinikums Bonn.

Das zur Auswertung verwendete Datenmaterial wurde durch mich aus den Sequenzier-Rohdaten eigenständig generiert. Alle anschließenden Schritte der Datenverarbeitung und Analyse erfolgten durch mich. Ebenso führte ich die statistische Auswertung eigenständig durch, in Absprache mit meinen Betreuer:innen.

Die Interpretation der Ergebnisse erfolgte größtenteils durch mich in Absprache mit meinen Betreuer:innen. Die Konzeption und Interpretation der spezifischen Analysen zu TLR7-Varianten erfolgten durch mich in Absprache mit meinen Betreuer:innen und in Zusammenarbeit mit Caspar I. van der Made und Prof. Dr. Alexander Hoischen (Radboud UMC). Darüber hinaus wurden spezifische Aspekte der Hauptergebnisse mit weiteren Personen, die Spezialwissen in diesen Bereichen haben, diskutiert. Dies waren Dr. Gayatri Ramakrishnan (Radboud UMC), Eamon Coughlan (University of Edinburgh), Dr.

Erola Pairo-Castineira (University of Edinburgh), Berina Zametica (Institut für Humangenetik, Uni Bonn), Dr. Julia Heggemann (Institut für Humangenetik, Uni Bonn) und Dr. Konrad Rawlik (University of Edinburgh).

Die 3D-Struktur-Analysen von TLR7 wurden von Dr. Gayatri Ramakrishnan (Radboud UMC) durchgeführt. Die unabhängige Replikation in der GenOMICC-Kohorte wurde von Eamon Coughlan, PhD Erola Pairo-Castineira und Dr. Konrad Rawlik (University of Edinburgh) durchgeführt.

Bei der Formulierung dieser Arbeit verwendete ich keine LLMs (large language models).

Ich versichere, die Dissertationsschrift selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

4. Danksagung

Mein wichtigster Dank gilt meiner Doktormutter Prof. Dr. Kerstin U. Ludwig für die fantastische Betreuung. Du bist fachlich und menschlich ein großes Vorbild für mich. Ich bin beeindruckt von Deinem Wissen, Deinem Überblick über die aktuelle Forschung, Deiner Fähigkeit, den springenden Punkt zu erkennen und bei Diskussionen dabei zu bleiben. Danke für die vielen guten und produktiven Meetings, Deine konstruktive Kritik und Deine Ideen, die mich jedes Mal weiter gebracht haben. Nicht zuletzt danke für Deine herzliche Art, die das gemeinsame Arbeiten sehr angenehm gemacht hat.

Außerdem möchte ich Dr. Axel Schmidt für die tolle Zusammenarbeit und Unterstützung danken, insbesondere dafür, dass ich von ihm so viel über statistische Tools und Methoden in der Genetik lernen durfte sowie für die geteilte und gemeinsam ausgelebte Begeisterung für informatische und statistische Fragen. Vielen Dank auch an das ganze Team der Humangenetik der Uni Bonn unter Leitung von Prof. Dr. Markus M. Nöthen, insbesondere an Julia Fazaal für die hervorragende Laborarbeit bei der Sequenzierung sowie an die NGS Core Facility für die Unterstützung bei der Datenerhebung. Danke an alle beteiligten Arbeitsgruppen für die gute Zusammenarbeit, die mir einen besonderen Einblick in die internationale Forschungswelt gegeben hat. Danke insbesondere an das Team des Instituts für Klinische Molekularbiologie der Uni Kiel sowie an die beteiligten Forschungsgruppen aus Spanien und Italien, ohne die es dieses Projekt nicht gegeben hätte. Ein großer Dank geht auch an Prof. Dr. Alexander Hoischen, Caspar van der Made und PhD Gayatri Ramakrishnan vom Radboud University Medical Center für die zahlreichen produktiven Meetings sowie die Begeisterung und das Fachwissen, die unser Projekt so viel weiter gebracht haben.

Darüber hinaus möchte ich meinen Freund:innen für die Unterstützung und die schöne gemeinsame Zeit abseits der Doktorarbeit danken. Mein letzter Dank gilt meiner Familie, Mira, Frauke und Jochen, die mich und damit auch meine naturwissenschaftlichen Interessen geformt und unterstützt haben und die immer an mich glauben. Ich danke Euch für eure Unterstützung und dafür, dass Ihr nach wie vor so tolle Gesprächspartner:innen, sowohl für fachliche als auch menschliche Themen, seid.