



Plantilla Protocolo Trabajo Final de Carrera

Título

Marcadores genéticos y perfiles de expresión génica del huésped asociados con susceptibilidad a infección por SARS-CoV-2.

Apellido y Nombre del Alumno: Butti Matías Daniel

Apellido y Nombre del Tutor: Grzona Esteban

Resumen

Introducción:

La finalización del proyecto genoma humano en el año 2000, la innovación de la biotecnología y las ciencias de datos biomédicas han permitido progresar rápidamente en la identificación de biomarcadores de predisposición, diagnóstico, pronóstico y de respuesta a tratamiento en diferentes patologías de interés. La susceptibilidad a enfermedades infecciosas también fue objeto de estudio en este campo.

Objetivos:

En el contexto de la pandemia COVID-19, es clave estudiar todas las aristas que influyen en el contagio y proceso viral. La genética del huésped es una arista que está tomando cada vez mayor relevancia. El objetivo del presente trabajo es realizar una revisión sistemática y un meta-análisis para identificar variantes genéticas y/o perfiles de expresión génica en el huésped, asociados con la susceptibilidad a SARS-CoV-2. Este hallazgo será importante para

adoptar las estrategias de prevención y tratamiento adecuadas, no solo a nivel poblacional sino individual.

Material y métodos:

Se realiza una revisión sistemática de publicaciones de estudios de asociación genotipo-fenotipo y se realiza un metaanálisis que integre toda la información pública disponible utilizando bash, R y bioconductor con el preprocesamiento bioinformático previo correspondiente.

Palabras Clave:

"Risk Factors"

"COVID-19"

"Human genetics"

"Gene Expression profile"

"Genetic Predisposition to Disease"

Contenidos

INTRODUCCION	5
Marco teórico	5
Justificación	6
Fundamentos Teóricos	7
La pregunta de Investigación	8
Estrategia de Búsqueda Bibliográfica	9
OBJETIVOS	9
Objetivo General	9
Objetivos Específicos	9
METODOLOGIA	10
Diseño del Estudio	10
Población en Estudio	10
Criterios de Inclusión	10
Criterios de Exclusión	10
Planificación para Recolección de los Datos	11
Descripción operacional de las variables	11
Instrumento/s para recolección de los datos	11
Plan de Análisis de los Datos	12
Sesgos y Limitaciones del Estudio	12
Recursos necesarios	12
Cronograma de actividades	13
Anexo/s	13
Tablas y/o cuadros	14
Referencias	14

INTRODUCCION

Marco teórico

En diciembre de 2019 un elevado número de pacientes contrajeron neumonía en la ciudad de Wuhan, lo que atrajo la atención de China, y del mundo en general¹. Luego de determinar que la causa de este brote era un coronavirus, y que este tenía la capacidad de causar un síndrome agudo respiratorio grave², el CSG (Coronavirus Study Group) lo reconoció taxonómicamente como hermano de SARS-CoV y lo nombró SARS-CoV-2³. El 11 de febrero de 2020 la Organización Mundial de la salud definió el nombre de la enfermedad causada por este virus, COVID-19⁴, y el 11 de marzo la caracterizó como pandemia⁵.

La pandemia ha afectado, al 17 de mayo de 2020, a 4.525.497 personas a nivel mundial ⁶ y producido 307.395 muertes⁶.

Comprender los determinantes del riesgo de infección y de formas graves de la enfermedad es clave para adoptar las estrategias de prevención y tratamiento adecuadas, no solo a nivel poblacional sino individual.

Se ha detectado que los adultos tienen mayor riesgo de infección respecto a los niños^{7,8} y el sexo masculino respecto al femenino⁹. También se postularon como determinantes del riesgo de infección el grupo sanguíneo A¹⁰ y niveles altos de LDH¹¹.

Los factores determinantes de peor evolución una vez adquirida la enfermedad que se conocen hasta el momento son edad > 60 años⁷, sexo masculino^{8,9} y la presencia de comorbilidades¹², principalmente diabetes^{7,13,14,15}, enfermedad cardiovascular^{7,9}, hipertensión⁷ y obesidad^{9,16,17}; también hay mayor riesgo en inmunocomprometidos y pacientes EPOC. Con menor importancia y mayor controversia pacientes con asma y el antecedente personal de ACV¹⁸. Los determinantes bioquímicos son altos niveles de LDH y leucocitosis¹¹.

Desde el punto de vista molecular, se ha demostrado que ACE2 -proteína presente principalmente en las células AT2 alveolares - funciona como sitio de unión para el spike del virus ^{19,20,21,22}. También se había demostrado que ACE2 es el sitio de unión para los coronavirus surgidos previamente, SARS-CoV y NL63^{23,24}. Otro hallazgo importante es que el producto del gen *STMP3* humano es utilizado por el virus para realizar el clivaje que permite la fusión de membranas y posterior penetración del virus a la célula²¹.

Los mismos tipos celulares que producen ACE2 expresan otros genes claves para el ingreso del virus: *ITGB6*, *CAV2*²² y otros para permitir la salida de la célula, a las nuevas partículas virales formadas: *CHMP3*, *CHMP5*, *CHMP1A*, *VPS37B*²².

Los niveles de ACE2 en tejido pulmonar varían en población sana. La población de Asia del Este presenta mayor expresión de ACE2. Esto lo explica en parte, el hecho de que existen variantes genéticas con diferencias entre las poblaciones en los sitios eQTL (loci que alteran la expresión del gen)^{22,25,26}.

Aún no se demostró que existan variantes genéticas en la región codificante del gen *ACE2* que la conviertan en no funcional (lo que podría implicar falta de expresión y por lo tanto inmunidad o mayor resistencia natural al virus).

Respecto a la inmunidad, el sistema de antígeno leucocitario humano (HLA) cumple un rol importante en la susceptibilidad a varias infecciones virales²⁸ y a la gravedad de las enfermedades que producen²⁹. Incluso se han reportado variantes en el gen *IFITM3* (Interferon-induced transmembrane protein 3) asociadas con evolución desfavorable en pacientes infectados con virus de la influenza H749 y H1N1^{30,31}

También se ha postulado que pacientes que presentan ciertas variantes deletéreas en el gen Myxovirus resistance A (*MxA*) tienen peor evolución, debido a que este gen es una proteína antiviral estimulada por interferon α y β ³².

Las tecnologías han sido un complemento clave para los avances en diagnóstico, prevención y tratamiento. En particular el uso de técnicas de Machine learning para inferir modelos basados en grandes volúmenes de datos³³.

Una vez conocida la secuencia de ARN del SARS-CoV2, la biotecnología y la bioinformática permitieron avanzar en identificar las variantes diferenciales respecto a otros virus conocidos e inferir consecuencias funcionales de las mismas¹⁹. Iniciativas como GISAID^{34,35}, que promueven compartir secuencias del SARS-CoV-2, permiten identificar y visualizar las mutaciones del virus en las distintas regiones del planeta³⁶ e inferir la patogenicidad de las distintas cepas y la respuesta al tratamiento.

Ambas disciplinas son fundamentales, también, para estudiar la genética del huésped y su asociación con la susceptibilidad/mala evolución COVID-19. También es necesaria una iniciativa análoga a GISAID para compartir datos genéticos/de expresión/clínicos del huésped. En este caso el rol lo cumple la Host Genetics Initiative³⁷, consorcio internacional que promueve compartir los datos genéticos de huéspedes con COVID-19.

Justificación

Aún con todos los hallazgos e hipótesis descriptos respecto a los factores de susceptibilidad a la infección y de riesgo de aparición de una enfermedad pulmonar grave, no se considera completa la lista de factores moduladores¹¹.

Ciertas variantes presentes en el gen *ACE2* podrían alterar la susceptibilidad si disminuyen la capacidad de asociación del virus a la proteína que codifica²⁷. Se han registrado más de 1700 variantes en el gen.

La variabilidad en la expresión de *ACE2* entre poblaciones, debido en gran parte a las variantes en los eQTL del gen, también estimula el estudio de una posible asociación entre la genética/expresión génica y la susceptibilidad/mala evolución. De la misma forma, podrían ser relevantes las variantes en los genes que son funcionales al ciclo viral: *ITGB6*, *CAV2*, *CHMP3*, *CHMP5*, *CHMP1A*, *VPS37B*. Resulta especialmente interesante analizar el papel de *TMPRSS2*, que cumple una función clave junto a *ACE2* para el ingreso del virus a la célula.

En este marco, y considerando el avance en el conocimiento molecular sobre la forma de ingreso y del proceso viral del SARS-CoV-2, estudiar variantes genéticas y perfiles de expresión génica que puedan formar parte del algoritmo de estimación de la susceptibilidad a la infección y el riesgo de evolución desfavorable, resulta conveniente para una prevención y tratamientos más precisos y efectivos.

Disponer de variantes genéticas y patrones de expresión génica diferencial en los genes que participan del ciclo viral facilita la identificación de: i) poblaciones con mayor riesgo de infección; ii) marcadores genéticos de riesgo de enfermedad pulmonar grave que complementan los factores ya conocidos; iii) Nuevos blancos terapéuticos basados en el inhibición/estimulación de expresión génica.

De la misma forma que es importante estudiar las variantes y los perfiles de expresión en genes relacionados con el ciclo viral, también es importante estudiar los relacionados con la inmunidad (inmunomoduladores). Por esta razón es clave incluir los genes del HLA en el análisis.

Fundamentos Teóricos

La finalización del proyecto genoma humano^{38,39} y el avance de la biotecnología ha permitido disponer de diferentes soluciones (microarrays, NGS, RNASeq) para secuenciar las bases

nucelotídicas del ADN germinal o somático de un individuo y conocer el transcriptoma (niveles de expresión génica) del tejido en estudio; incluso existen soluciones para poder conocerlo a nivel de célula única (scRNA-seq).

Por su parte mediante la bioinformática, la bioestadística y técnicas de Machine learning y deep learning es posible encontrar marcadores genéticos asociados a diferentes fenotipos, a partir de los grandes volúmenes de datos generados por los dispositivos biotecnológicos.

En el campo de las enfermedades infecciosas, resulta de interés encontrar variantes genéticas o perfiles de expresión génica que aumenten la susceptibilidad a contraer enfermedades infecciosas o que predispongan a una peor evolución o menor respuesta al tratamiento. Mediante estudios GWAS⁴⁰, de exoma completo o de genoma completo se han postulado diferentes variantes genéticas en el huésped relacionadas con enfermedades infecciosas^{41,42,43,44}, con importantes resultados accionables en el caso de la malaria, HIV e infecciones por priones⁴⁵.

Debido a que la obtención de estos datos es costosa, una limitación frecuente es no disponer de un tamaño muestral suficiente para poder arribar a conclusiones válidas. Pero esta limitante se resuelve con iniciativas que promueven compartir los datos de los estudios y tener acceso a una base de datos más representativa de la población. En el caso de COVID-19 existe tal iniciativa (Host Genetics Initiative³⁷), lo cual me permite postular mi pregunta de investigación.:

Enviamos un proyecto propuesta a la convocatoria del Ministerio de Ciencia y Técnica COVID-19 para estudiar 900 pacientes utilizando un microarray que permite obtener 700.000 posiciones polimórficas (que tienen variabilidad a lo largo de la población) para luego con bioinformática procesarlos y encontrar asociaciones genotipo-fenotipo. Si bien no nos fue otorgado el subsidio, los 3 ejes de evaluación fueron calificados con puntaje ALTO.

- ¿La solución propuesta para dar respuesta a la pandemia es relevante y los mecanismos de implementación resultan adecuados? ALTO
- ¿El proyecto es de corta duración y es factible completarlo en el plazo propuesto? ALTO
- ¿Grado de desarrollo tecnológico? ALTO

La pregunta de Investigación

(P) En individuos sanos no infectados por SARS-CoV-2 con variantes genéticas respecto al genoma de referencia o perfiles de expresión génica alterados, (C) comparado con individuos que no presentan estas características moleculares, (O) tienen una mayor susceptibilidad a la infección?

Estrategia de Búsqueda Bibliográfica

1. ("Risk Factors"[Mesh]) AND "COVID-19" [Supplementary Concept]: con filtro clinical trial y systematic review. Dio 126 resultados. Los proceso para armar el marco teórico.
2. "COVID-19" [Supplementary Concept] AND "Genetic Predisposition to Disease"[Mesh]. Si bien actualmente esta búsqueda arroja solo 2 resultados, es esperable que en los próximos meses haya muchos más para poder realizar una revisión sistemática. Para afirmarlo me baso en la iniciativa Host Genetics Initiative³⁷ que ya publicó los estudios que compartieron datos de genotipo y fenotipo. Esta iniciativa claramente dará las bases para futuras publicaciones de asociación genotipo-susceptibilidad. Ver anexo 1.
3. "CORONAVIRUS" [Mesh] AND "Genetic Predisposition to Disease"[Mesh].
Acordé con el director que en caso que no haya publicaciones suficientes para poder estudiar específicamente Sars-Cov-2, utilizaremos esta búsqueda que arroja 19 resultados para generalizar la revisión sistemática.

OBJETIVOS

Objetivo General

Identificar variantes genéticas o perfiles de expresión génica en el huésped, asociados con aumento de susceptibilidad o resistencia a la infección por SARS-Cov-2.

Objetivos Específicos

Se analizarán los papers que ya postulan resultados sobre potenciales marcadores genéticos asociados con susceptibilidad a SARS-CoV-2. Se espera que haya mayor cantidad de resultados debido a iniciativas para compartir resultados de estudios de exoma y microarrays a lo largo del mundo.

Se realizará un meta-análisis tomando como datos de entrada la información que publique Host Genetics Initiative³⁷ utilizando R y Bioconductor como lenguajes de programación para el análisis de datos.

METODOLOGIA

Diseño del Estudio

Se realizará una síntesis de la evidencia disponible sobre la asociación entre variantes genéticas o perfiles de expresión génica y la susceptibilidad al SARS-CoV-2, mediante un estudio de tipo **revisión sistemática** que incluirá un **meta-análisis** de los datos crudos que dieron lugar a las publicaciones y cruzamiento con bases de datos bioinformáticas secundarias como las bases del NCBI y EBI. El meta-análisis pondrá el foco en los genes ACE2 y STMP3 y genes del HLA.

Población en Estudio

Conjunto de papers que postulen o rechacen asociaciones entre variantes genéticas germinales o perfiles de expresión génica del huésped y susceptibilidad a SARS-CoV-2.

Criterios de Inclusión

Se incluirán publicaciones que hayan correlacionado genotipo con susceptibilidad a SARS-CoV2 mediante estudios genéticos GWAS⁴⁵, o de exomas/genoma completo en pacientes COVID-19 de cualquier edad y género.

También se incluirán revisiones que postulen hipótesis a partir del estudio del ciclo viral del SARS-CoV-2 y las proteínas humanas necesarias para perpetuar el mismo.

También se incorporarán revisiones sistemáticas.

Criterios de Exclusión

Quedarán excluidos los reportes de casos que no tengan evidencia estadística suficiente ni aporten datos para el metaanálisis.

Planificación para Recolección de los Datos

- Se realizará la búsqueda utilizando la siguiente expresión basada en términos Mesh: "COVID-19" [Supplementary Concept] AND "Genetic Predisposition to Disease"[Mesh].
- Se realizará un meta-análisis de los datos curdos de las publicaciones. La principal fuente para realizar el meta-análisis será la host genetic initiative ³⁷ que ya dispone de datasets agrupados preliminares (ver anexo 2)
- Para el metaanálisis se tomarán los archivos crudos originales de los estudios y los agrupados, junto con los datos clínicos y se realizarán estudios bioinformáticos para encontrar correlación genotipo-fenotipo utilizando el lenguaje de programación R y Bioconductor. Las variantes identificadas se anotarán bioinformáticamente para incorporar mayor cantidad de información de cada una y claridad para su interpretación.

Descripción operacional de las variables

Para poder hacer el metaanálisis será necesario disponer de los siguientes atributos en pacientes COVID-19 para contrastar con el genoma de referencia (Genome Reference Consortium Human Build 37 (GRCh37)) e identificar posibles variantes que tengan una fuerte asociación con la susceptibilidad.

- #CHR: Cromosoma de la base nucleotídica. Variable discreta.
- POS: Posición en el cromosoma. Variable discreta.
- REF: alelo de referencia. Categórica no ordinal (A, C, T, G por adenina, citosina, tirosina, guanina)
- ALT: alelo alternativo según frecuencia poblacional. Categórica no ordinal (A, C, T, G por adenina, citosina, tirosina, guanina)
- SNP: Single nucleotide polymorphism. Posición con variabilidad a lo largo de la población.
- Alelo encontrado en el paciente

Se deberá incorporar también información clínica de relevancia como hospitalización y necesidad o no de respirador.

Instrumento/s para recolección de los datos

Para la revisión sistemática se utilizará pubmed.

Para el metaanálisis, la fuente principal de datos será el Host Genetics Initiative³⁷ en su sección de resultados. Pero también se utilizarán otros datasets públicos que no fueran depositados en la Host Genetics Initiative.

Plan de Análisis de los Datos

Para la revisión sistemática, se hará una tabla excel con las variantes y las consecuencias clínicas postuladas por los distintos estudios para encontrar patrones y diferencias.

Para el metaanálisis, se utilizarán archivos de texto plano que registren para cada, locus (posición en el genoma), el alelo de referencia y todos los alelos encontrados en los diferentes estudios.

Debido a que son archivos de gran tamaño (>5 gb) serán preprocesados utilizando scripts de bash y luego analizados con R y bioconductor cruzando con bases de datos y anotadores bioinformáticos.

Se utilizarán técnicas estadísticas de comparación de las frecuencias alélicas en todas las muestra COVID-19 obtenidas de los diferentes estudios contra las frecuencias alélicas en el genoma humano de referencia. Se utilizará como genoma de referencia el Build 37 (GRCh37).

Sesgos y Limitaciones del Estudio

La limitación del estudio es que es un tema muy actual del cual aún no hay evidencia suficiente para realizar ni la revisión sistemática ni el metaanálisis. Pero siguiendo a diario como se mueve la iniciativa Host, es un hecho que en pocos meses habrá información suficiente.

Por ser un tema tan novedoso y ser la iniciativa host abierta, también está el riesgo de perder la originalidad del tema porque otro grupo lo haga primero.

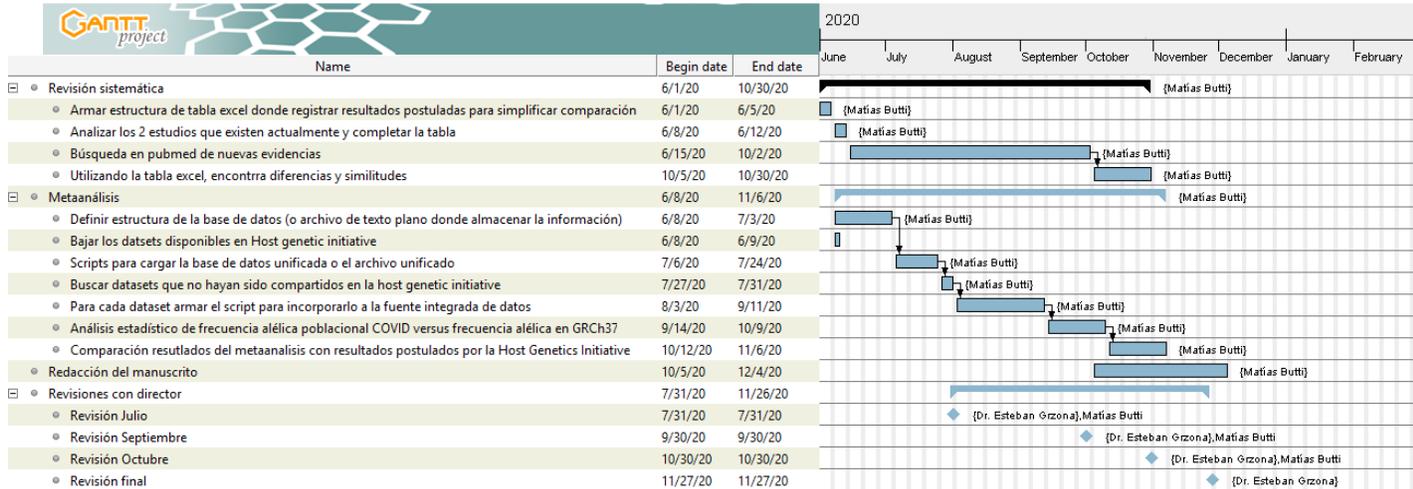
Si ocurriera cualquiera de los casos, según lo acordado con el director, se generalizará la revisión sistemática a susceptibilidad genética a coronavirus que ya tiene 19 publicaciones en pubmed.

Recursos necesarios

Revisión sistemática: Nuevas publicaciones en el tema.

Metaanálisis: acceso a los datos de la host initiative e idealmente a datasets que no estén publicados en la host.

Cronograma de actividades



Anexo/s

Anexo 1 - Estudios genotipo-fenotipo pacientes COVID, ya incorporados a la host genetics initiative ³⁷

Results		
COVID19-hg GWAS meta-analyses round 2		
Release Date	May 15, 2020	
Study Abbreviations	Full Name	Abbreviation
	BioMe	BioMe_RGN
	FinnGen	FinnGen
	Genes & Health	GNH
	Helix Exome+ COVID-19 Phenotypes	Helix
	LifeLines Global Screening Array	LifelinesCyto
	LifeLines CytosNP	LifelinesGsa
	Netherlands Twin Register	NTR
	Partners Healthcare Biobank	PHBB
	UK Biobank	UKBB

Anexo 2 - Fragmento de tabla integrada por host genetics initiative ³⁷ de los diferentes genotipos de pacientes COVID

NCRR	POS	REF	ALT	SNP	BioMe_RSN_EUR_AF_Allele2	FinnGen_FIN_AF_Allele2	GNM_S&S_AF_Allele2	LifelinesCyto_EUR_AF_Allele2	LifelinesGsa_EUR_AF_Allele2	NTR_CEU_AF_Allele2	PHBB_AFR_AF_Allele2	PH	
1	58296	T	C	1:58296:1:C	NA	NA	NA	2	-1.18e+00	1.29e+00	3.59e-01	8.59e-01	
1	523573	C	A	1:523573:1:C	NA	NA	NA	2	2.4e-01	8.09e-01	8.85e-01	6.85e-01	
1	541944	T	C	1:541944:1:C	NA	NA	NA	2	-1.08e+00	6.82e-01	1.12e-01	9.44e-01	
1	457788	C	G	1:457788:1:C	NA	NA	NA	2	-2.16e-02	1.16e-01	8.53e-01	6.28e-01	
1	458228	C	T	1:458228:1:C	NA	NA	NA	2	0.8218095497784999	0.0005457161105343	NA	NA	
1	462613	G	A	1:462613:1:C	0.8194672419265095	2.597e-02	NA	NA	2	-8.38e-02	3.21e-01	7.94e-01	
1	462622	G	A	1:462622:1:C	NA	1.757e-01	NA	NA	2	0.173246515805629	0.180974731147209	NA	NA
1	462897	C	C	1:462897:1:C	NA	7.744e-01	NA	NA	2	-6.21e-02	1.14e-01	5.48e-01	
1	466249	C	T	1:466249:1:C	NA	1.178e-03	NA	NA	2	6.18e-01	1.84e-00	5.57e-01	
1	492794	GA	C	1:492794:1:GA	0.8959637489532685	NA	NA	NA	3	1.153e-01	0.13e-02	2.38e-02	
1	493553	AT	A	1:493553:1:AT	NA	3.876e-03	NA	NA	2	-8.24e-02	7.87e-01	9.17e-01	
1	493731	A	G	1:493731:1:A	NA	1.858e-01	NA	NA	2	0.171788273291969	0.099118114216885	2.4332e-01	
1	493823	G	C	1:493823:1:C	NA	1.843e-01	NA	NA	2	0.183918886758984	0.0995744894252586	NA	
1	497487	A	C	1:497487:1:A	NA	5.845e-03	NA	NA	2	-6.48e-01	7.75e-01	4.82e-01	
1	498477	C	T	1:498477:1:C	NA	NA	NA	NA	2	-2.89e-01	5.46e-01	5.97e-01	
1	498487	A	G	1:498487:1:A	NA	NA	NA	NA	2	-2.89e-01	5.46e-01	5.97e-01	

Tablas y/o cuadros

Se iniciará el trabajo partiendo de las tablas de los archivos crudos publicados por la host genetics initiative.

- COVID19_HGI_ANA5_20200513.txt.gz: https://storage.googleapis.com/covid19-hg-public/20200508/results/COVID19_HGI_ANA5_20200513.txt.gz
- COVID19_HGI_ANA5_20200513.txt.gz.tbi: https://storage.googleapis.com/covid19-hg-public/20200508/results/COVID19_HGI_ANA5_20200513.txt.gz.tbi

Referencias

1. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. 2020;395(10223):514–23.
2. Zhou P, Yang X Lou, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270–3.
3. Gorbalenya AE, Baker SC, Baric RS, Groot RJ De, Gulyaeva AA, Haagmans BL, et al. The species and its viruses – a statement of the Coronavirus Study Group. *Biorxiv (Cold Spring Harb Lab)*. 2020;1–15.
4. WHO. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020> (WHO, 11 February 2020).
5. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (WHO, 11 March 2020).
6. WHO Coronavirus disease 2019 (COVID-19), Situation Report 118 https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200517-covid-19-sitrep-118.pdf?sfvrsn=21c0d4fe_6 (WHO, 17 May 2020)
7. Surveillances V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2020;41(2):145–51.
8. Dhochak N, Singhal T, Kabra SK, Lodha R. Pathophysiology of COVID-19 : Why Children Fare

- Better than Adults ? 2020;416.
9. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol.* 2020;(0123456789).
 10. Zhao J, Yang Y, Huang H-P, Li D, Gu D-F, Lu X-F, et al. Relationship between the ABO Blood Group and the COVID-19 Susceptibility. *medRxiv.* 2020;2020.03.11.20031096.
 11. Romera-Liebana L, Orfila F, Segura JM, Real J, Fabra ML, Möller M, et al. Risk factors of Severe Disease and Efficacy of Treatment in patients infected with COVID-19: A systematic Review, Meta-Analysis and Meta-Regression Analysis . *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 2018;2019:1–11.
 12. Gold JAW, Wong KK, Szablewski CM, Patel PR, Rossow J, Silva J, et al. Characteristics and clinical outcomes of adult patients hospitalized with COVID-19 — Georgia, March 2020. *Morb Mortal Wkly Rpt.* 2020;69(18):1–6.
 13. Li Z, Nie Y, Liu B, Kuai Z, Zhao M, Liu F. COVID-19 in diabetic patients: related risks and specifics of management. *Annales d'Endocrinologie* 2020;108709.
 14. Yan Y, Yang Y, Wang F, Ren H, Zhang S, Shi X, et al. Clinical characteristics and outcomes of patients with severe covid-19 with diabetes. *BMJ Open Diabetes Res Care.* 2020;8(1):e001343.
 15. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol.* 2020;(0123456789).
 16. Petrakis D, Margină D, Tsarouhas K, Tekos F, Stan M, Nikitovic D, et al. Obesity - a risk factor for increased COVID-19 prevalence, severity and lethality (Review). *Mol Med Rep.* 2020;22(1):9–19.
 17. Gao F, Zheng KI, Wang X, Sun Q, Pan K, Wang T, et al. Obesity Is a Risk Factor for Greater COVID-19 Severity. 2020;(April):10–2.
 18. Aggarwal G, Lippi G, Michael Henry B. Cerebrovascular disease is associated with an increased disease severity in patients with Coronavirus Disease 2019 (COVID-19): A pooled analysis of published literature. *Int J Stroke.* 2020;0(0):1–5.
 19. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020;26(4):450–2.
 20. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol.* 2020;(March).
 21. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 2020;181(2):271-280.e8.
 22. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. *bioRxiv.* 2020;2020.01.26.919985.
 23. Li W, Sui J, Huang IC, Kuhn JH, Radoshitzky SR, Marasco WA, et al. The S proteins of human coronavirus NL63 and severe acute respiratory syndrome coronavirus bind overlapping regions

- of ACE2. *Virology*. 2007;367(2):367–74.
24. Wu K, Li W, Peng G, Li F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc Natl Acad Sci U S A*. 2009;106(47):19970–4.
 25. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discov*. 2020;6(1):4–7.
 26. Liu Y, Yang Y, Zhang C, Huang F, Wang F, Yuan J, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci*. 2020;63(3):364–74.
 27. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*. 2005;24(8):1634–43.
 28. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun*. 2017;8(1). <https://doi.org/10.1038/s41467-017-00257-5>
 29. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PIW, Walker BD, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* (80-). 2010;330(6010):1551–7.
 30. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature*. 2012;484(7395):519–23.
 31. Wang Z, Zhang A, Wan Y, Liu X, Qiu C, Xi X, et al. Early hypercytokinemia is associated with interferon-induced transmembrane protein-3 dysfunction and predictive of fatal H7N9 infection. *Proc Natl Acad Sci U S A*. 2014;111(2):769–74.
 32. Ching JC, Chan KYK, Lee EHL, Xu M, Ting CKP, So TMK, et al. Significance of the Myxovirus Resistance A (MxA) Gene –123C>A Single-Nucleotide Polymorphism in Suppressed Interferon β Induction of Severe Acute Respiratory Syndrome Coronavirus Infection. *J Infect Dis*. 2010;201(12):1899–908.
 33. Kumar A, Gupta PK, Srivastava A. Diabetes & Metabolic Syndrome : Clinical Research & Reviews A review of modern technologies for tackling COVID-19 pandemic. *Diabetes Metab Syndr Clin Res Rev*. 2020;14(4):569–73.
 34. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob Challenges*. 2017;1(1):33–46.
 35. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*. 2017;22(13):2–4.
 36. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121–3.
 37. Ganna A, Unit TG, General M. The COVID-19 Host Genetics Initiative , a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. 2020;

38. Craig Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* (80-). 2001;291(5507):1304–51.
39. Correction: Initial sequencing and analysis of the human genome. *Nature*. 2001;412(6846):565–6.
40. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017 Jan;45(D1):D896–901.
41. Chapman SJ, Hill AVS. Human genetic susceptibility to infectious disease. *Nat Rev Genet*. 2012 Feb;13(3):175–88.
42. Hill A V. The genomics and genetics of human infectious disease susceptibility. *Annu Rev Genomics Hum Genet*. 2001;2:373–400.
43. Vannberg FO, Chapman SJ, Hill AVS. Human genetic susceptibility to intracellular pathogens. *Immunol Rev*. 2011 Mar;240(1):105–16.
44. Mozzi A, Pontremoli C, Sironi M. Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis*. 2018 Dec;66:286–307.
45. Murray MF, Kenny EE, Ritchie MD, Rader DJ, Bale AE, Giovanni MA, et al. COVID-19 outcomes and the human genome. *Genet Med*. 2020;0(0):18–20.