



Vilnius
universitetas

Ligų prognozavimo algoritmai panaudojant naujos kartos sekoskaitos duomenis

2020 – 2024

Ramunė Vaišnorė, Informatikos inžinerijos pirmo kurso doktorantė

Vadovė – prof. dr. Audronė Jakaitienė

Tyrimo objektai:

- Naujos kartos sekoskaitos genominiai duomenys;
- Ligų prognozavimas panaudojant naujos kartos sekoskaitos duomenis ir atrinktus svarbiausius genominius požymius.

Tyrimo tikslas:

- Sukurti/patobulinti naujos kartos sekoskaitos duomenų analizės algoritmą, kuris gebėtų atrasti reikšmingus ligos genominius požymius, padedančius identifikuoti bei prognozuoti ligą.

Tyrimo uždaviniai

- Atlikti panašių jau egzistuojančių algoritmų analizę;
- Išsiaiškinti naujos kartos sekoskaitos duomenų kokybės problemas ir pritaikyti metodus pradinių duomenų apdorojimui, jų kokybės gerinimui;
- Sukurti/patobulinti pradinį genominių požymių atradimo ir ligų prognozavimo modelį;
- Tobulinti pradinį modelį aukštesniam tikslumui pasiekti;
- Validuoti sukurtą modelį su naujais, algoritmui nematytais duomenimis.

Planuojami pasiekti rezultatai

- Sukurtas/patobulintas naujos kartos sekoskaitos duomenų analizės algoritmas, gebantis atrasti reikšmingus ligos genominius požymius, padedančius identifikuoti bei prognozuoti ligą.

Visų studijų plano santrauka

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2020/2021)	0	2	0	1 (+1)	0	(1)	Priimtas straipsnis publikavimui žurnale be cituojamumo indekso
II (2021/2022)	2		1				
III (2022/2023)	1		2		1		
IV (2023/2024)	1		1		1		
Iš viso:	0	2	0	1	0	0	

Pirmųjų metų darbo planas

Egzaminai		Dalyvavimas konferencijose		Publikacijos	
Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta
	Išlaikyta: Mašininis mokymasis, 7 ECTS, dr. Virginijus Marcinkevičius. 2021 m. vasario 17 d.		Pranešimas „ <i>Functional pathways analysis for COVID-19 modelling: exploring Lithuanian data</i> “ tarptautinėje konferencijoje - 8th Nordic-Baltic Biometrics Virtual Conference, Helsinki		Ramunė Vaišnorė, Audronė Jakaitienė, <i>COVID-19 reproduction number and non-pharmaceutical interventions in Lithuania</i> , 2021. Straipsnis priimtas spausdinimui Lietuvos matematikos rinkinyje (LMD darbai) (be cituojamumo rodiklio)
	Išlaikyta: Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika, 8 ECTS, dr. A. Lupeikienė. 2021 m. birželio 29 d.		Pranešimas „ <i>SARS-CoV-2 viruso plitimo modeliavimas Lietuvos savivaldybėse</i> “ Lietuvos Matematikos Draugijos organizuotoje 62-oje virtualioje konferencijoje, vykusioje 2021 m. birželio 16-17 d.		

Bendrųjų gebėjimų ugdymas

- Praktiniai mokymai apie projektų rengimą – 0,25 ECTS (2020-10-16);
- Mokslinių rezultatų publikavimas pagal formalaus vertinimo reikalavimus – 0,1 ECTS (2020-11-10);
- Atvirosios prieigos kompetencijų tobulinimas – 0,2 ECTS (2020-11-10);
- Mokslinių tyrimų duomenų valdymo dirbtuvės – 0,25 ECTS (2020-12-03);
- Mokslo komunikacija (angl. *science communication*) – 0,3 ECTS (2021-06-30).
- **Viso:** 1,1 ECTS.

Mokslinių tyrimų ir disertacijos rengimo etapai (I)

	Darbo pavadinimas	Atlikimo terminai	Pastabos
1	<p>Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):</p> <ol style="list-style-type: none"> 1. Anototos bibliografijos sudarymas. 2. Mokslinės literatūros apžvalga. Egzistuojančių metodų taikymo naujos kartos sekoskaitos duomenims analizavimas. 	2020/10 – 2021/05	<p>Mokslinės literatūros disertacijos tema papildymas genominių duomenų analizės metodų apžvalga.</p>
2	<p>Mokslinio tyrimo vykdymas:</p> <p>2.1. <u>Tyrimo metodikos sudarymas:</u></p> <ol style="list-style-type: none"> 1. Disertacijos tikslo formulavimas. 2. Disertacijos uždavinių formulavimas. <p>2.2. <u>Teorinis tyrimas:</u></p> <ol style="list-style-type: none"> 1. Matematinio modelio sudarymas. 2. Algoritmų konstravimas ir tobulinimas. <p>2.3. <u>Empirinis tyrimas:</u></p> <ol style="list-style-type: none"> 1. Siūlomų algoritmų pritaikymas naujos kartos sekoskaitos duomenims. 2. Siūlomų algoritmų tobulinimas, atsižvelgiant į gautus rezultatus. <p>2.4. <u>Gautų duomenų analizė, apibendrinimas, išvadų parengimas:</u></p> <ol style="list-style-type: none"> 1. Algoritmų tikslumo įvertinimas, palyginimas su kitų autorių metodais, atrinktais remiantis išanalizuota mokslinė literatūra. 2. Gautų rezultatų apibendrinimas. 3. Išvadų parengimas. 	<p>2021/06 – 2021/09</p> <p>2021/10 – 2022/06</p> <p>2022/06 – 2022/09</p> <p>2022/10 – 2023/02</p> <p>2023/02 – 2023/09</p>	<p>Suformuluotas disertacijos tikslas bei uždaviniai.</p>

Mokslinių tyrimų ir disertacijos rengimo etapai (II)

Darbo pavadinimas		Atlikimo terminai
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas: 1. Tyrimų apžvalga ir analizė. 2. Tyrimo metodikos sudarymas. 3. Teorinis tyrimas. 4. Empirinis tyrimas. 5. Gautų duomenų analizė, apibendrinimas. 6. Išvados, įvadas, literatūros sąrašas.	2023 m. spalio mėn. – 2024 m. gegužės mėn.
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2024 m. birželio mėn.
5.	Daktaro disertacijos gynimas	2024 m. rugsėjo mėn.

Per metus gauti moksliniai rezultatai (I)

1. Atliktas Lietuvos savivaldybių Covid-19 epidemiologinės situacijos įvertinimas atsižvelgiant į efektyvaus reprodukcijos skaičiaus įvertį ir taikytų nemedikamentinių intervencijos priemonių efektą.

Rezultatai pristatyti Lietuvos matematikų draugijos organizuotoje 62-oje konferencijoje (2021-06-16).

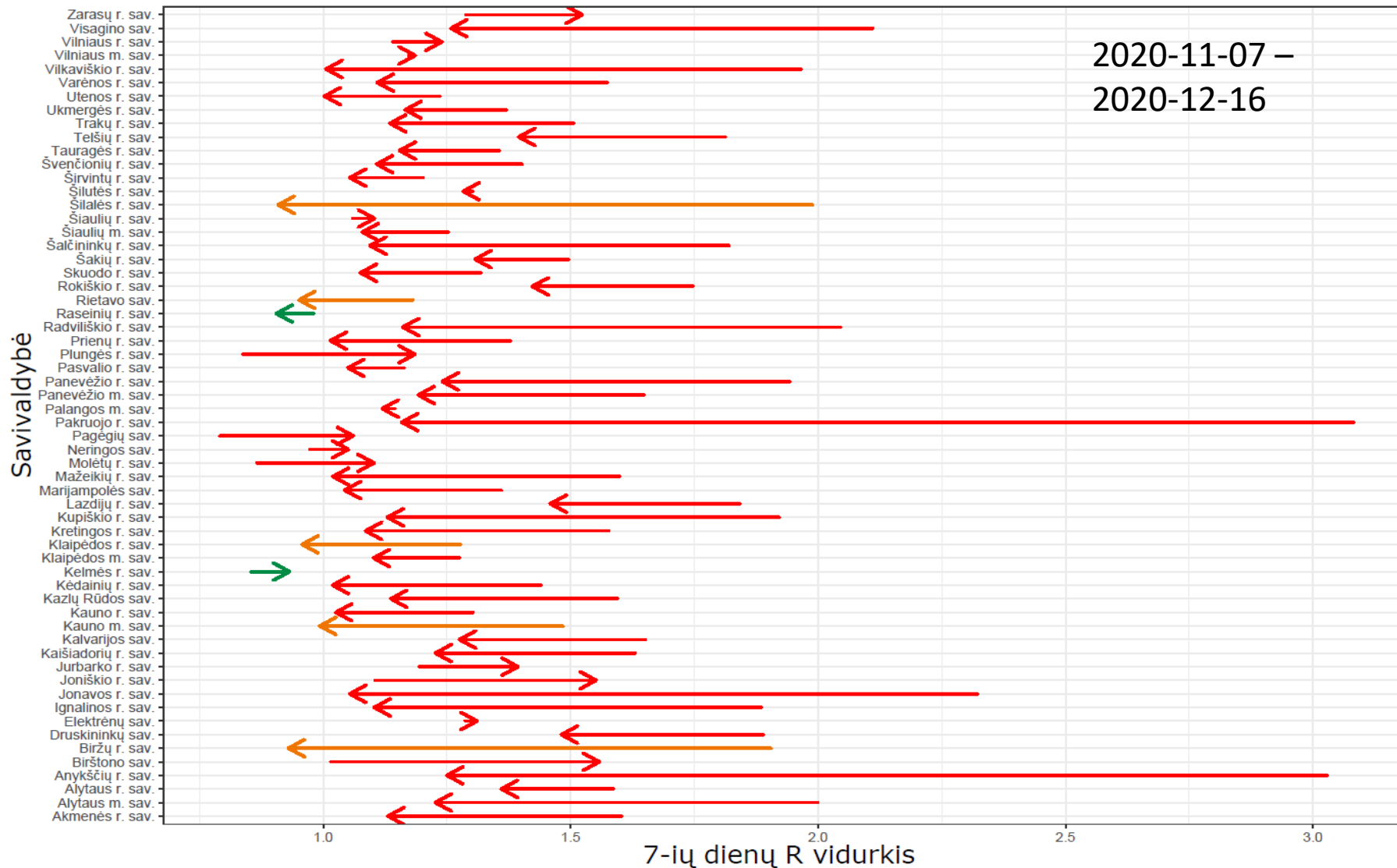
Priimtas straipsnis „*COVID-19 reproduction number and non-pharmaceutical interventions in Lithuania*“ spausdinimui Lietuvos matematikų rinkinyje (LMD darbai).

2. Atlikta literatūros analizė apie naujos kartos sekoskaitą, šiuo metu taikomus algoritmus naujos kartos sekoskaitos genominių duomenų analizei, genominių žymenų ir fenotipo asociacijų bei priežastinio ryšio nustatymo metodiką.

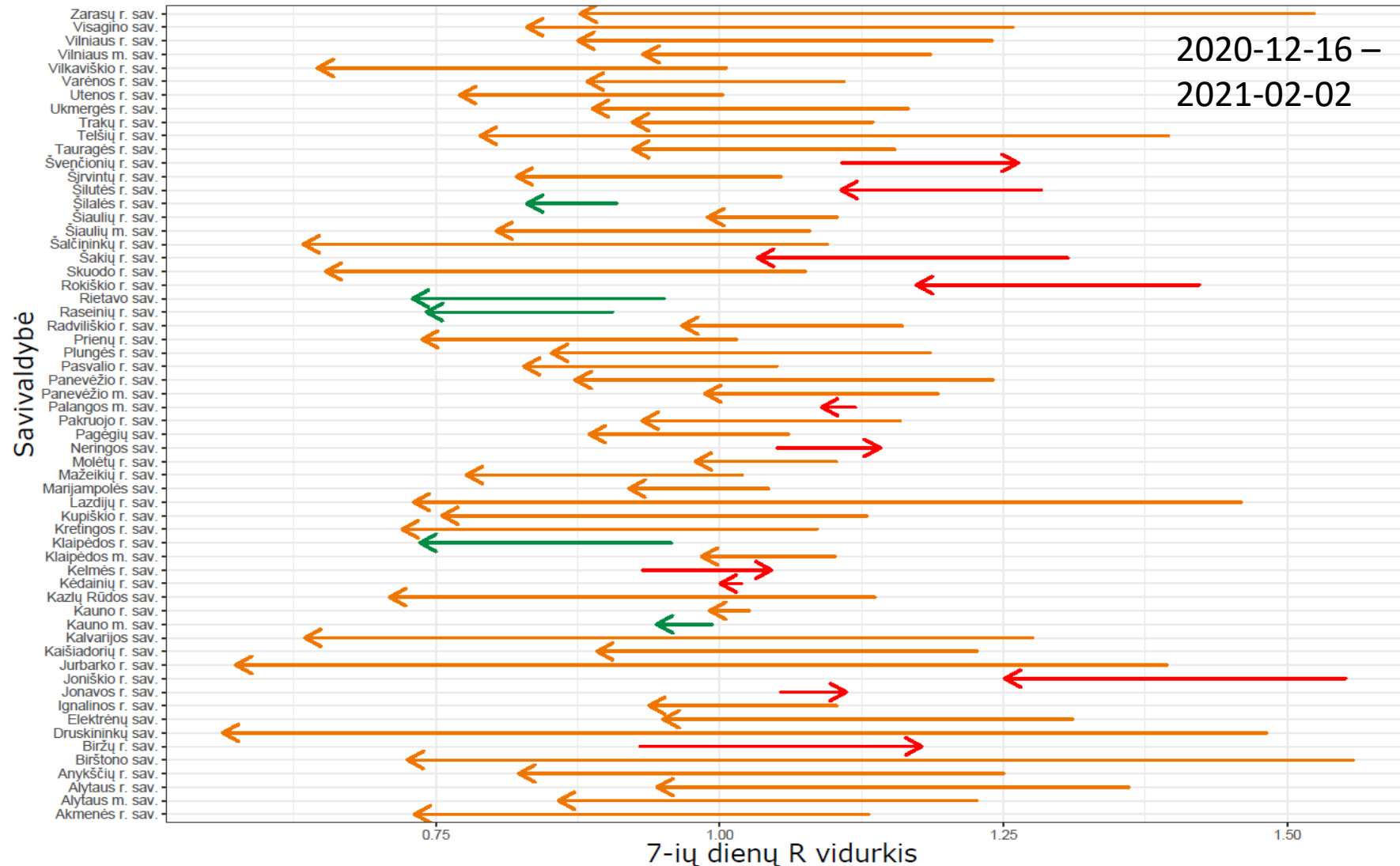
Per metus gauti moksliniai rezultatai (II)

3. 2021 m. liepos 7-10 d. sudalyvauta tarptautinėje konferencijoje „8th Nordic-Baltic Biometrics Virtual Conference, Helsinki“, pristatyta COVID-19 funkcinių kelių bei COVID-19 modeliavimui taikomų metodų apžvalga.
4. Šiuo metu rašomas COVID-19 funkcinių kelių bei modeliavimo metodų apžvalginis straipsnis „*Functional pathways analysis for COVID-19 outcome modelling*“.

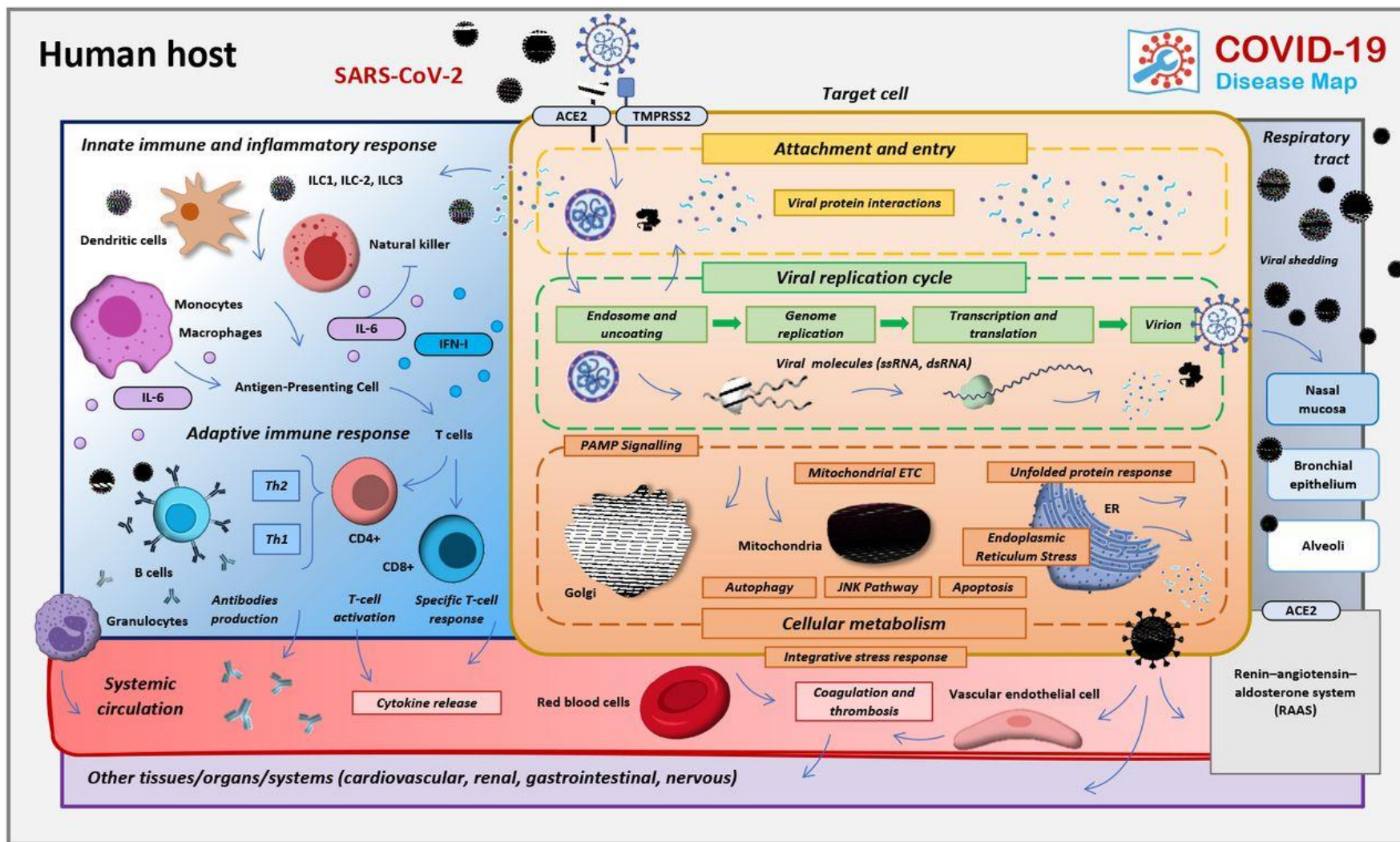
COVID-19 epidemiologinės situacijos Lietuvos savivaldybėse įvertinimas (I)



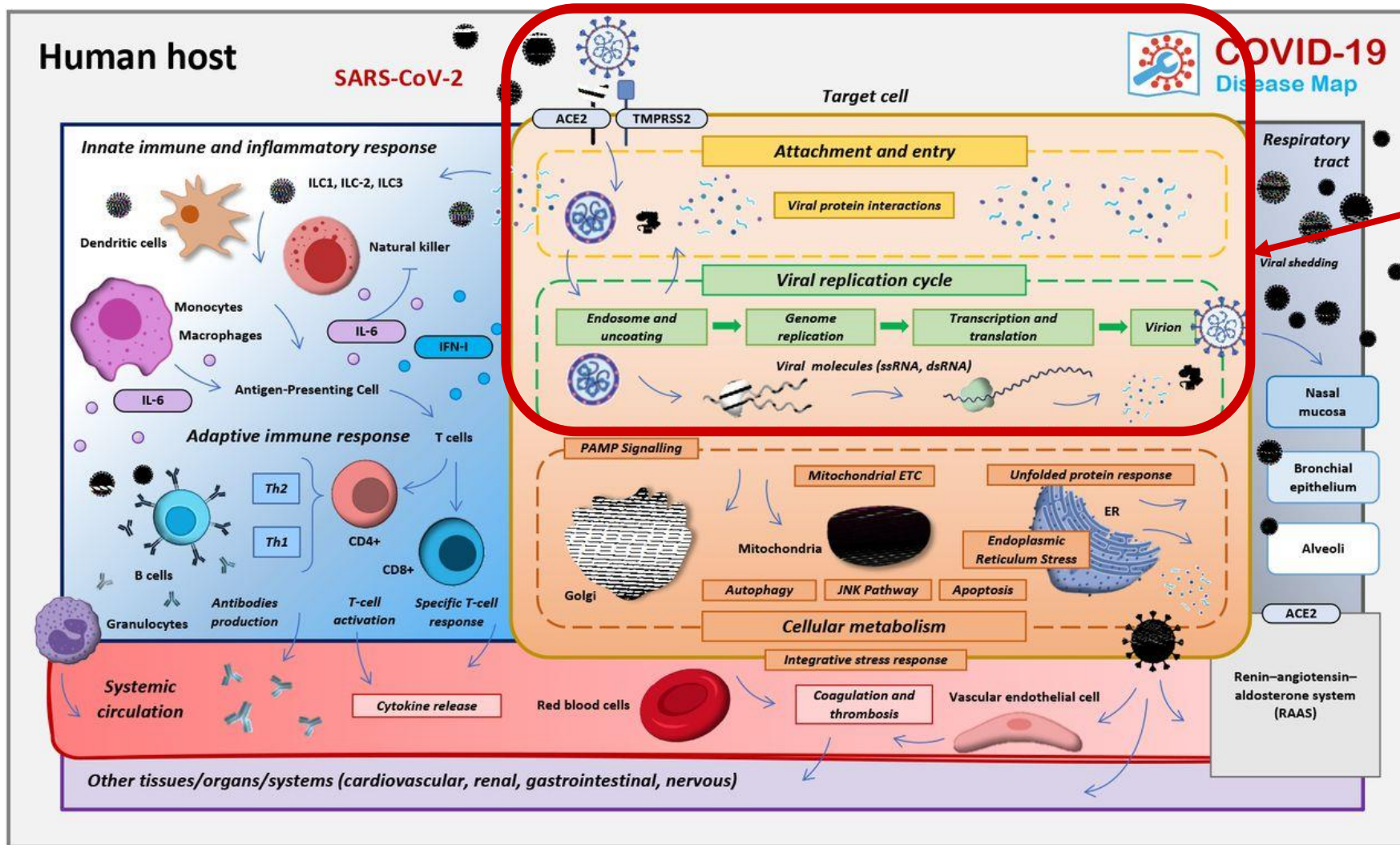
COVID-19 epidemiologinės situacijos Lietuvos savivaldybėse įvertinimas (II)



COVID-19 funkciniai keliai ir ligos modeliavimas



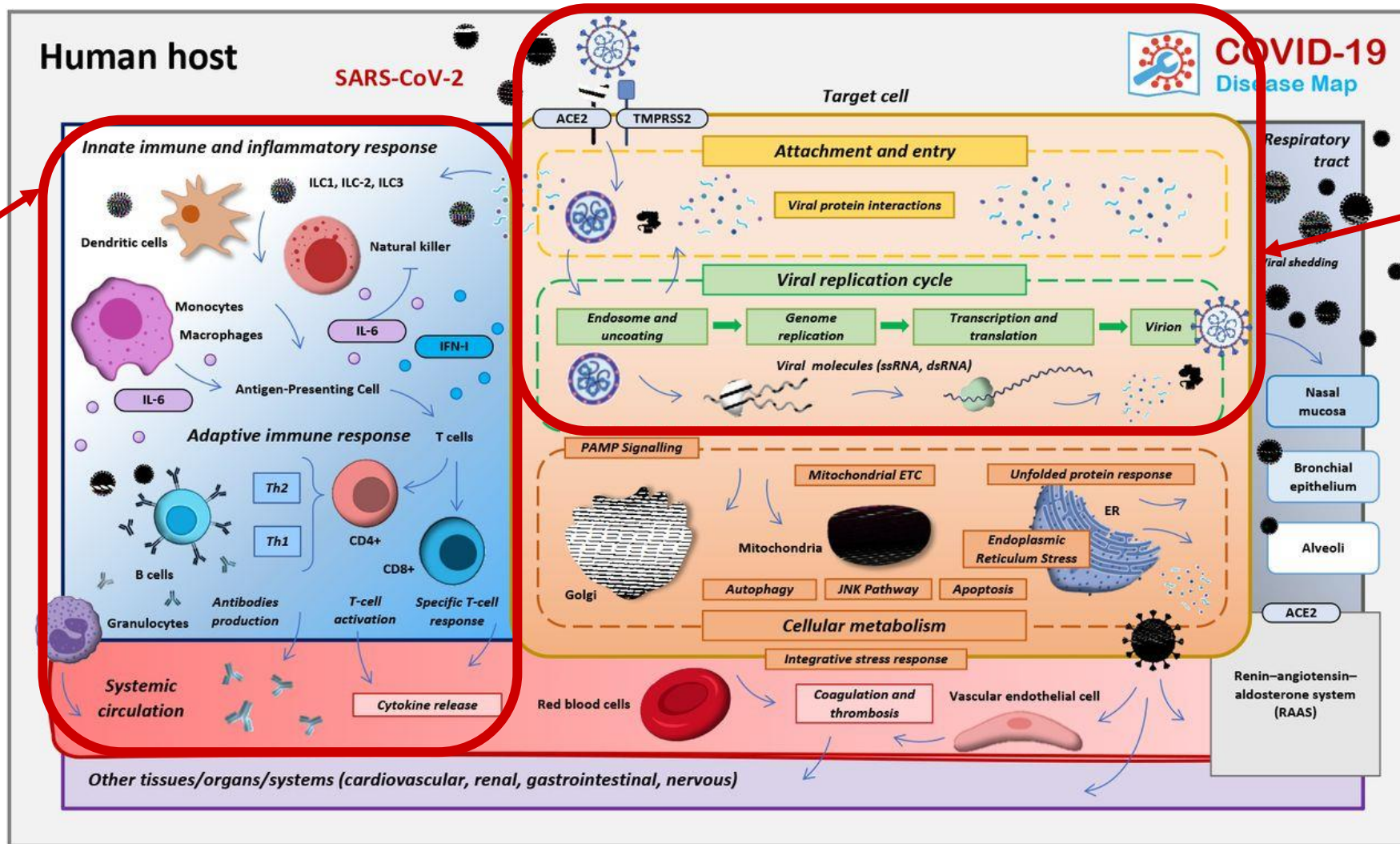
COVID-19 funkciniai keliai ir ligos modeliavimas



1 – viruso patekimas į ląstelę ir dauginimasis

COVID-19 funkciniai keliai ir ligos modeliavimas

2 –
šeiminių
imuninis
atsakas



1 – viruso
patekimas į
ląstelę ir
dauginimasis

Šeimininko genetinės savybės

Funkcinis kelias	Genas	Variantas	Potencialus efektas	Šaltiniai
Viruso prisijungimas ir patekimas į ląstelę	ACE2	rs73635825, rs778030746, rs4646116, rs1199100713	Padidėjęs jautrumas viruso prisijungimui	Stawiski et al. 2020, Procko et al. 2020
		rs1348114695, rs146676783, rs755691167	Sumažina receptoriaus specifiškumą viruso S baltymui	
	TMPRSS2	rs12329760, rs75603675	Padidėjęs COVID-19 imlumas	Hou et al. 2020
		rs12329760	Apsauginis efektas	Monticelli et al. 2021
		rs12329760	Sumažėjusi sunkios formos COVID19 rizika	David et al. 2021
	TLL-1	rs17047200	Padidėjęs COVID-19 imlumas	Grimaudo et al., 2021

Šeiminingo genetinės savybės

Funkcinis kelias	Genas	Variantas	Potencialus efektas	Šaltiniai
Imuninis atsakas	<i>HLA</i>	A*25:01, B*15:27, DRB1*15:01, DQB1*06:02, B*27:07, C*01:02, C*07:29	Padidėjusi COVID-19 rizika	Secolin et al., 2021; Fricke-Galindo and Falfán-Valencia, 2021
		A*02:02, B*15:03, C*12:03	Sumažėjusi COVID-19 rizika	
	<i>IFNAR2</i>	rs13050728, rs2236757, rs9976829	Imlumas COVID-19	COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Ma et al., 2021
	<i>ABO</i>	rs657152, rs912805253, rs9411378, rs8176719, rs550057, rs505922, rs657152		Severe Covid-19 GWAS Group, 2020; The COVID-19 Host Genetics Initiative, 2021; Shelton et al., 2021
	<i>TLR7</i>	rs189681811, rs147244662, rs149314023, rs200146658		Fallerini et al., 2021
	<i>LZTFL1</i>	rs71325088, rs11385942, rs73064425, rs11385942		Severe Covid-19 GWAS Group, 2020; Rescenko et al., 2021
	<i>APOE</i>	rs429358, rs7412		Kuo et al., 2020
	<i>TYK2</i>	rs74956615		Pairo-Castineira et al., 2021

Metodai, naudojami COVID-19 modeliavimui: įprasti metodai

1. Viso genomo asociacijos tyrimas (angl. *genome-wide association study*, GWAS) (COVID-19 Host Genetics Initiative, 2021; Severe Covid-19 GWAS Group, 2020; Pairo-Castineira et al., 2021; Dey et al., 2021; Hu et al., 2021)
2. Viso egzomo asociacijos tyrimas (angl. *exome-wide association study*, EWAS) (Kosmicki et al., 2020)
3. Fenotipo asociacijos tyrimas (angl. *phenotype-wide association study*, PheWas) (COVID-19 Host Genetics Initiative, 2021, Verma et al., 2021)
4. Regresija: tiesinė (Ma et al., 2021), logistinė (Severe Covid-19 GWAS Group, 2020; Grimaudo et al., 2021; Dite et al., 2021 ; Horowitz et al., 2020; Kuo et al., 2020)

Metodai, naudojami COVID-19 modeliavimui: nauji pažangūs metodai

1. Sherlock (gene-disease associations' detection using eQTL and GWAS data) (Ma et al., 2021)
2. Mašininio mokymosi metodai: atsitiktiniai miškai, XGBoost (Wang R. Y., 2020, Toh, C., 2020, Wang, F., 2020)
3. Dirbtiniai neuroniniai tinklai (Wang R. Y., 2020)
4. LASSO logistinė regresija (Fallerini et al., 2021)

Įprasti vs. nauji pažangūs metodai

- Mokslininkai yra linkę analizuoti genus, kurie, jau įrodyta, yra susiję su analizuojamu objektu, dėl ko kiti potencialiai svarbūs genai yra nustumiami į šoną (Stoeger and Amaral, 2020).
- Asociacijoms nusatyti naudojami įprasti, gerai žinomi standartiniai metodai.
- GWAS metodo du pagrindiniai trūkumai: atliekami daugybiniai vieno VNP statistiniai testai, kurie neįskaito koreliacijų tarp VNP; galimai atmetami svarbūs VNP, turintys mažą efektą; nėra tiriama VNP ir fenotipo priežastiniai ryšiai; nepaaiškinamas kompleksinių ligų paveldimumas.
- Pažangesnių ir sudėtingesnių šiuolaikinių skaičiavimo metodų (pvz. mašininio mokymosi) taikymas COVID-19 modeliavimui potencialiai galėtų pasiūlyti naujų įžvalgų.

Kito pusmečio darbo planas

Egzaminai	Dalyvavimas konferencijose	Publikacijos
Fundamentalieji informatikos ir informatikos inžinerijos metodai, 8 kreditai	Dalyvavimas tarptautinėje konferencijoje (IBC2022 ar panašioje)	Publikacija indeksuojamame žurnale (BJMC ar panašiam)

<u>2.2. Teorinis tyrimas:</u> 1. Matematinio modelio sudarymas. 2. Algoritmų konstravimas ir tobulinimas.	2021 m. spalio mėn. – 2022 m. birželio mėn.
---	--

Šaltiniai

- COVID-19 Host Genetics Initiative, 2021. Mapping the human genetic architecture of COVID-19. *Nature*. <https://doi.org/10.1038/s41586-021-03767-x>
- David, Alessia, et al. 2021. "A Common *TMPRSS2* Variant Protects against Severe COVID-19." Preprint. *Genetic and Genomic Medicine*.
<https://doi.org/10.1101/2021.03.04.21252931>
- Fallerini, C., Daga, S., Mantovani, S. Et al., 2021. Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: findings from a nested case-control study. *Elife* 10, p.e67569. <https://doi.org/10.7554/eLife.67569>
- Fricke-Galindo, Ingrid, and Ramcés Falfán-Valencia. 2021. "Genetics Insight for COVID-19 Susceptibility and Severity: A Review." *Frontiers in Immunology* 12 (April): 622176. <https://doi.org/10.3389/fimmu.2021.622176>
- Grimaudo, S., Amodio, E., Pipitone, R.M. et al., 2021. PNPLA3 and TLL-1 Polymorphisms as Potential Predictors of Disease Severity in Patients With COVID-19. *Frontiers in Cell and Developmental Biology* 9, p.1589. <https://doi.org/10.3389/fcell.2021.627914>
- Hou, Yuan, et al. 2020. "New Insights into Genetic Susceptibility of COVID-19: An ACE2 and *TMPRSS2* Polymorphism Analysis." *BMC Medicine* 18 (1): 216.
<https://doi.org/10.1186/s12916-020-01673-z>
- Kuo, C.L., Pilling, L.C., Atkins, J.L. et al., 2020. APOE e4 genotype predicts severe COVID-19 in the UK Biobank community cohort. *The Journals of Gerontology: Series A*, 75(11), pp.2231-2232. <https://doi.org/10.1093/gerona/glaa131>
- Monticelli, Maria, et al. 2021. "Protective Role of a *TMPRSS2* Variant on Severe COVID-19 Outcome in Young Males and Elderly Women." *Genes* 12 (4): 596.
<https://doi.org/10.3390/genes12040596>
- Pairo-Castineira, E., Clohisey, S., Klaric, L. et al., 2021. Genetic mechanisms of critical illness in Covid-19. *Nature* 591(7848), pp.92-98.
<https://doi.org/10.1101/2020.09.24.20200048>
- Procko, Erik. 2020. "The Sequence of Human ACE2 Is Suboptimal for Binding the S Spike Protein of SARS Coronavirus 2." Preprint. *Biochemistry*.
<https://doi.org/10.1101/2020.03.16.994236>
- Secolin, Rodrigo, et al. 2021. "Genetic Variability in COVID-19-Related Genes in the Brazilian Population." *Human Genome Variation* 8 (1): 15.
<https://doi.org/10.1038/s41439-021-00146-w>
- Severe Covid-19 GWAS Group, 2020. Genomewide association study of severe Covid-19 with respiratory failure. *New England Journal of Medicine* 383 (16), pp.1522-1534. <https://doi.org/10.1056/NEJMoa2020283>
- Shelton JF, Shastri AJ, Ye C, Weldon CH, 23andMe COVID-19 Team, et al, 2021. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet* 53(6):801-808. PMID: 33888907. <https://doi.org/10.1038/s41588-021-00854-7>
- Stawiski, Eric W., et al. 2020. "Human ACE2 Receptor Polymorphisms Predict SARS-CoV-2 Susceptibility." Preprint. *Genetics*.
<https://doi.org/10.1101/2020.04.07.024752>
- Stoeger, T. and Amaral, L.A.N., 2020. Meta-Research: COVID-19 research risks ignoring important host genes due to pre-established research patterns. *Elife*, 9, p.e61981. <https://doi.org/10.7554/eLife.61981>
- Wang, F., Huang, S., Gao, R. et al., 2020. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell discovery* 6(1), pp.1-16. <https://doi.org/10.1038/s41421-020-00231-4>
- Wang, R.Y., Guo, T.Q., Li, L.G. et al., 2020. Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data. 2020, ICCSNT (pp. 92-96). IEEE. <https://doi.org/10.1109/ICCSNT50940.2020.9304990>