

## БАЗА ДАНИХ ROUA

Христина ЩУБЕЛКА<sup>1,2</sup>, Валтер ВОЛФСБЕРГЕР<sup>1</sup>, Ольга Т. ОЛЕКСИК<sup>4</sup>, Ярослава ГАСИНЕЦЬ<sup>2</sup>, Сільвія ПАЦКУН<sup>3</sup>, Михайло ВАКЕРИЧ<sup>2,6</sup>, Роман КІШ<sup>2</sup>, Віолета МІРУТЕНКО<sup>2</sup>, Владислав МІРУТЕНКО<sup>2</sup>, Коралія Адіна КОТОРАЧІ<sup>5</sup>, Калін ПОП<sup>5</sup>, Олімпія НЕАГУ<sup>5</sup>, Корнель БАЛТЕ<sup>5</sup>, Хільдегарда ГЕРМАН<sup>5</sup>, Паула МАРЕ<sup>5</sup>, Сімона ДУМІТРА<sup>5</sup>, Гораціу ПАПІУ<sup>5</sup>, Анка ГЕРМЕНЕАН<sup>5</sup>, Тарас К. ОЛЕКСИК<sup>1,2</sup>

*В статті представлений багаторівневий ресурс даних, що надає результати двох популяцій з регіону Карпатських гір, зокрема Закарпаття в Україні та провінцій Satu Mare та Băia Mare в Румунії, малодосліджених раніше в популяційній геноміці. База даних містить як сирі, так і анотовані файли повногеномних послідовностей від 300 осіб із цих регіонів, включаючи анотації загальних та унікальних генетичних варіантів, згідно з протоколом вибірки, розробленим для захоплення генетичної різноманітності українців та румунів, включаючи меншинні групи, як-от волохи та роми. Дані розміщено на спеціалізованому веб-ресурсі. Надано інформацію про те, як отримати доступ до результатів первинного та вторинного аналізу даних, включаючи порівняльний аналіз з раніше опублікованими популяціями з України, а також популяціями з Міжнародного Ресурсу Геномних Зразків (IGSR) та Проекту Різноманітності Людського Геному (HGDP). Вільний доступ до цієї бази даних для досліджень сприяє щораз більшому розумінню генетичної різноманітності людей у Центральній Європі. Ця робота підкреслює потенціал для повторного використання отриманих даних, виступаючи за відкритий доступ для підтримки майбутніх досліджень у геноміці, біоінформатиці та персоналізованій медицині.*

**Ключові слова:** повногеномне секвенування, Карпати, Україна, Румунія, біоінформатика, геноми.

<sup>1</sup>Відділ біологічних наук, Оклендський університет, Рочестер, Мічиган 48309, США;

<sup>2</sup>Біологічний факультет, Ужгородський національний університет, Ужгород, 88000, Україна; e-mail: [khrystyna.shchubelka@uzhnu.edu.ua](mailto:khrystyna.shchubelka@uzhnu.edu.ua);

<sup>3</sup>Медичний факультет № 2, Ужгородський національний університет, Ужгород, 88000, Україна;

<sup>4</sup>Закарпатська обласна клінічна лікарня імені А. Новака, Ужгород, 88000, Україна;

<sup>5</sup>«Vasile Goldiș» Західний університет Арада, 94–96, Revoluției Bld., Арад, 310025, Румунія;

<sup>6</sup>Закарпатський науково-дослідний експертно-криміналістичний центр МВС України, Слов'янська наб., 25, Ужгород, 88000, Україна.

**ROUA Database. Shchubelka K.<sup>1,2</sup>, Wolfsberger W.<sup>1</sup>, Oleksyk O.T.<sup>4</sup>, Hasynets Ya.<sup>2</sup>, Patskun S.<sup>3</sup>, Vakerych M.<sup>2,6</sup>, Kish R.<sup>2</sup>, Mirutenko V.<sup>2</sup>, Mirutenko V.I.<sup>2</sup>, Cotoraci C.A.<sup>5</sup>, Pop C.<sup>5</sup>, Neagu O.<sup>5</sup>, Baltă C.<sup>5</sup>, Herman H.<sup>5</sup>, Mare P.<sup>5</sup>, Dumitra S.<sup>5</sup>, Papiu H.<sup>5</sup>, Hermenean A.<sup>5</sup>, Oleksyk T.<sup>1,2</sup>**

*We present a multi-layered data source, providing the results of Whole Genome Sequencing of two human populations in the Carpathian Mountains region, specifically Ukraine's Transcarpathia and Romania's Satu Mare and Băia Mare provinces, areas previously underexplored in population genomics. The database contains the raw and annotated files of the whole genome sequences from 300 individuals from these regions, including annotations of common and unique genetic variants following a sampling protocol designed to capture the genetic diversity of Ukrainians and Romanians, including minority groups like Wallachians and Roma. The data is hosted on a dedicated web resource. We provide information on how to access to results of primary and secondary analysis of the data, including comparative analysis with previously published populations from Ukraine, and populations from International Genome Sample Resource and Human Genome Diversity Project. The free research access to this database is contributing to growing understanding of human genetic diversity in Central Europe. This effort emphasizes the potential for reuse of the generated data, advocating for open access to support future research in genomics, bioinformatics, and personalized medicine.*

**Key words:** Whole Genome Sequencing, Carpathians, Ukraine, Romania, bioinformatics, genomes

<sup>1</sup>Department of Biological Sciences, Oakland University, Rochester, MI 48309, USA.

<sup>2</sup>Faculty of Biology, Uzhhorod National University, Uzhhorod, 88000, Ukraine; e-mail: [khrystyna.shchubelka@uzhnu.edu.ua](mailto:khrystyna.shchubelka@uzhnu.edu.ua);

<sup>3</sup>Faculty of Medicine #2, Uzhhorod National University, Uzhhorod, 88000, Ukraine;

<sup>4</sup>A. Novak Transcarpathian Regional Clinical Hospital, Uzhhorod, 88000, Ukraine;

<sup>5</sup>"Vasile Goldiș" Western University of Arad, 94-96, Revoluției Bld., Arad 310025, Romania;

<sup>6</sup>Transcarpathian scientific research expert and forensic center of the Ministry of Internal Affairs of Ukraine, Slovianska nab., 25, Uzhhorod, 88000, Ukraine.

## Вступ

База даних містить послідовності цілого геному 300 осіб із двох сучасних популяцій, які проживають у регіоні Карпатських гір на міжнародному кордоні між Україною та Румунією, зокрема в Закарпатській області України та в провінціях Сату Маре та Бая Маре Румунії, це райони, які раніше були недостатньо представлені в геномних базах даних. База даних знаходиться на спеціалізованому вебпорталі, який містить інші матеріали, пов'язані з проектом «Партнерство в геномних дослідженнях в Україні та Румунії», виконаним українськими та румунськими партнерами за підтримки Спільної операційної програми Румунія–Україна через Європейський інструмент сусідства (ENI). Портал також надає детальні описи проекту, типи доступних даних та форму для запиту доступу до даних на різних етапах аналітичного процесу.

Процес відбору зразків для дослідження було схвалено Інституційною радою з етики Ужгородського національного університету, а медичні фахівці з певних регіонів наглядали за збором крові в лікарнях. Здорових добровольців було запрошено до участі через оголошення для інтерв'ю в амбулаторних установах, де вони надавали інформовану згоду на публічну доступність своїх даних і заповнювали анкету про своє походження та медичну історію. Усі особисті ідентифікатори було анонімізовано після збору, а зразки маркувалися унікальними кодами. Зразки крові збирали в пробірки з ЕДТА, відправляли в сертифіковану лабораторію для екстракції ДНК, а будь-який надлишок зберігався в біобанку Ужгородського національного університету.

Загалом 300 учасників надали зразки, які було проаналізовано на платформі DNBSEQ-G50. Дані секвенування оброблялися за допомогою розробленого нами високопродуктивного обчислювального конвеєра *PopGen Playground (PGP)* (Wolfsberger 2023), зчитування вирівнювалися з референтним геномом людини GRCh38 і підготовлювалися для виклику варіантів та їх аналізу. Отримана база даних поєднує результати секвенування цілого геному (WGS) та результати первинного та вторинного аналізу. Сирі послідовності зчитувань для кожного зразка в популяціях зберігаються в архівованому форматі FASTQ. Підсумкова статистика секвенування представлена в Таблиці 1.

У базі також представлено результати порівняльних аналізів даних української та румунської популяції зі зразками інших євразійських популяцій, отриманими від Міжнародного ресурсу геномних зразків та Проекту різноманітності людського геному (International Genome Sample Resource and Human Genome Diversity Project) (Bergström et al. 2020, Fairley et al. 2020). Ці бази даних містять табличні та графічні результати аналізу принципних компонентів (PCA), аналізу генетичного змішування та попарного точного тесту Фішера (FET) щодо частот алелів для кожної пари включених популяцій. Крім того, ми надаємо доступ до оброблених та відфільтрованих файлів формату виклику варіантів, що містять потенційно клінічно значущі алелі, з інтегрованими анотаціями бази даних *ClinVar* (Landrum et al. 2018). Дані аналізу включають функціонально анотовані файли формату виклику варіантів (VCF), які були оброблені з використанням адаптованого

Таблиця 1. Зведена статистика секвенування та властивості сирих файлів послідовностей від 300 осіб із двох сучасних людських популяцій, що проживають у Карпатському регіоні в прикордонні між Україною та Румунією: Закарпаття в Україні та провінції Сату Маре та Бая Маре в Румунії

Table 1. Summary of sequencing statistics and properties of raw sequence files from 300 individuals from two modern human populations living in the Carpathian region on the border between Ukraine and Romania: Transcarpathia in Ukraine and the provinces of Satu Mare and Baia Mare in Romania

Популяція	К-сть зразків	Довжина зчитувань, bp	Всього зчитувань (середнє)	Всього нуклеотидів (середнє)	Якість Q20 (середнє %)
Українці на Закарпатті	150	150	666,990,234	100,046,395,519	97.75
Румуни в Сату Маре та Бая Маре	150	150	674,134,737	101,120,210,556	96.28

конвеєра найкращих практик GATK. Зведену статистику файлів VCF наведено в Таблиці 2. Набір даних включає геноми 300 осіб із двох сучасних людських популяцій, що проживають у регіоні Карпатських гір на міжнародному кордоні між Україною та Румунією: 150 зразків із Закарпаття України (UA150) та 150 зразків – із провінцій Сату Маре та Бая Маре Румунії (RO150).

#### Контекст

Регіон Карпатських гір у Східній Європі, включаючи Закарпаття в Україні, та провінції Сату Маре та Бая Маре в Румунії, був малодосліджений з точки зору популяційної геноміки, незважаючи на свою багату історію та етнічну різноманітність (Oleksyuk et al. 2022). Соціоекономічна залежність регіону від сільського господарства та нижчі рівні індустріалізації порівняно з іншими районами України та Румунії диктують специфіку структури цих популяцій і створили потенціал для збереження цінних геномних варіантів. Попередні дослідження в регіоні в основному зосереджу-

валися на мітохондріальних гаплогрупах, але не включали всебічний аналіз на рівні всього геному. Виявлення унікальних генетичних варіантів у цих культурно та географічно значущих регіонах сприятиме розумінню генетичної різноманітності людей у Центральній Європі та розширить попередньо опубліковані 97 зразків WGS, отриманих в Україні (Oleksyuk et al. 2021).

Нова база даних містить індивідуальний та популяційний аналіз 300 повністю просеквенованих осіб із Закарпаття, Сату Маре та Бая Маре. Це дослідження, фінансоване Європейським Союзом у рамках Спільної операційної програми Румунія-Україна 2014–2020 (Грант ENI CBC-2SOFT/1.2/48), було проведено у співпраці між дослідниками прикордонних регіонів між Румунією та Україною.

#### Матеріал та методи

Процедура відбору зразків отримала схвалення від Інституційної комісії з перегляду (IRB) Ужгородського національного університету

Таблиця 2. Огляд анотації набору даних із 300 геномів

Table 2. Overview of annotation of a dataset of 300 genomes

Функціональна Анотація (кількість ефектів за регіоном геному)				
Type (alphabetical order)	RO150		UA150	
	Count	Percent	Count	Percent
Downstream regions	7,576,139	5.55%	8,694,629	9.51%
Exons	1,796,755	1.32%	1,403,258	1.54%
Genes	943	0.00%	2,352	0.00%
Intergenic Regions	11,380,842	8.34%	10,108,582	11.06%
Introns	105,645,185	77.37%	61,663,007	67.47%
Splice site acceptors	7,386	0.01%	5,326	0.01%
Splice site donors	6,336	0.01%	5,338	0.01%
Splice site regions	169,779	0.12%	107,995	0.12%
Transcripts	334,150	0.25%	4,810	0.01%
Upstream regions	7,495,661	5.49%	8,744,166	9.57%
3' UTR regions	1,662,611	1.22%	508,075	0.56%
5' UTR regions	461,771	0.34%	142,714	0.16%
Функціональна Анотація (кількість ефектів за ступенем впливу)				
Type	RO150		UA150	
	Count	Percent	Count	Percent
High	41,966	0.03%	23,203	0.03%
Low	663,955	0.49%	300,938	0.33%
Moderate	575,400	0.42%	224,767	0.25%
Modifier	135,256,237	99.06%	90,841,344	99.40%
Тип і кількість варіантів				
Type	RO150		UA150	
	Total		Total	
SNP	17,884,931		18,704,768	
Ins	2,607,813		2,938,616	
Del	2,573,906		3,429,057	
Total	23,066,650		25,072,441	

(Протокол № 1 від 18.09.2018 р.). Відбір учасників проводився з метою захоплення генетичної різноманітності українців у всіх районах Закарпаття (Закарпатська область) та румунів у провінціях Сату Маре та Бая Маре. Крім того, були включені генетичні зразки від груп національних меншин, зокрема волохів та ромів.

Медичні фахівці Закарпаття та провінцій Сату Маре та Бая Маре в Румунії були залучені для нагляду за збором крові в лікарнях. Здорових осіб, які на той момент не перебували в лікарнях, залучали через оголошення та запрошували на співбесіди в амбулаторні установи. Під час цих сесій учасникам роз'яснювали цілі дослідження та процес відбору проб, надаючи інформовану згоду на публічну доступність їх геномної та фенотипічної інформації. Учасники також заповнювали анкету, деталізуючи своє походження, місця народження бабусь та дідусів (коли відомо), стать та деякі фенотипічні ознаки, включаючи коротку історію здоров'я. Документація про згоди та співбесіди зберігається на біологічному факультеті Ужгородського Національного Університету. Після співбесіди та збору зразків особисті ідентифікатори були видалені, а самі зразки потім маркувалися унікальним буквено-цифровим кодом та штрих-кодом, забезпечуючи анонімність у всіх подальших аналізах та публікаціях.

Після завершення співбесіди сертифікований медичний працівник забирав зразок цільної крові у дві пробірки по 5 мл з ЕДТА, кожна з яких маркувалася штрих-кодом, і відправлялася в сертифіковану біомедичну лабораторію на сухому льоду для негайного виділення ДНК після отримання. Будь-яка зайва кров та зразки ДНК після генетичного аналізу зберігаються замороженими в біобанку біологічного факультету Ужгородського національного університету (Україна) та в Західному університеті «Василе Голдіш» в Арадї (Румунія).

Для всіх 300 зразків ДНК була виділена в лабораторії молекулярної генетики Ужгородського національного університету за допомогою набору для очищення ДНК Monarch (New England Biolabs, Inc., Rowley, MA, США) для екстракції геномної ДНК з оригінальних заморожених зразків цільної крові. Приблизно 1 мкг геномної ДНК фрагментувалося за допомогою Covaris (Woburn, Massachusetts) і підготовлювалося для секвенування на платформі DNBSEQ-G50 в BGI-Copenhagen (Данія).

Усі особи в цьому дослідженні були секвенувані за допомогою DNBSEQ-G50. Дані секвенування, отримані на цій платформі для 300 зразків, були вирівняні з референтним геномом людини

GRCh38 за допомогою BWA-MEM (Версія: 0.7.16a-r1181). Визначення варіантів проводилося відповідно до найкращих практик GATK (Depristo et al. 2011), використовуючи конвеєр *PopGen Playground (PGP)*, адаптований нами з каталогу робочих процесів *Snakemake* (Köster et al. 2021) і розміщений на GitHub (<https://github.com/valerpok/dna-seq-gatk-variant-calling>). Визначення варіантів проводилося у двох окремих партіях (150 українців і 150 румунів) і потім об'єднувалося для подальшого аналізу.

Файли варіантів послідовностей анотувалися за допомогою програмного забезпечення SNPEff (SNPEff, RRID:SCR\_005191) (Cingolani et al. 2012) із застосуванням бази даних анотацій довідників GRCh38. Ми використали базу даних *Clivar* (Landrum et al. 2016) та каталог GWAS (Sollis et al. 2023) для анотації медично значимих та функціональних варіантів за допомогою інструменту *snpSift*.

База даних містить інформацію про генетичну різноманітність та змішування популяцій. Для проведення аналізу основних компонентів (PCA) ми об'єднали геномні дані з нашого дослідження з європейськими зразками з проекту 1000 геномів (1KG) та Проекту різноманітності людського геному (Human Genome Diversity Project або HGDP (Fairley et al. 2020). Аналіз проводився за допомогою *Eigensoft* (Price et al. 2006). Після генотипування, фільтрації за швидкістю та обрізання за зв'язком з нерівновагою зчеплення залишилося 677 зразків з 208 945 варіантами. PCA візуалізувалася за допомогою *Python* з бібліотеками *pandas*, *matplotlib* та *seaborn*, за винятком двох викидів. Для проведення аналізу структури популяції на основі моделі з використанням того ж набору даних ми використовували програмне забезпечення ADMIXTURE (Alexander et al. 2009). Ми визначили оптимальний параметр K як 3 шляхом 10-кратної перехресної перевірки, що відповідає нашим попереднім результатам з проекту WGS в Україні. Графіки структури популяції створювалися за допомогою мови *Python*, включаючи зразки з баз даних IGSR та HGDP.

**Потенціал повторного використання та доступність даних.** База даних надає вільний доступ до геномних даних для дослідників у Європі та за її межами й заповнює важливу прогалину в нашому розумінні геномної різноманітності та історії популяцій. Ця база даних дотримується філософії відкритого доступу, з даними, що надаються за запитом для досліджень. Дані, розміщені на вебресурсі Ужгородського національного університету (<https://genomes.uzhnu.edu>).



ua/), стануть унікальним джерелом для майбутніх досліджень у галузі геноміки, біоінформатики та персоналізованої медицини.

Дані з повними інструкціями щодо використання бази даних ROUA, додаткові матеріали, включаючи протоколи збору, інформовані згоди та рішення Інституційної комісії з перегляду (IRB) Ужгородського національного університету (Протокол № 1 від 18.09.2018), доступні через веб-портал: <https://genomes.uzhnu.edu.ua/>.

#### Фінансування

Фінансування проекту було надано проектом 2SOFT/1.2/48 «Партнерство для геномних дослі-

джень в Україні та Румунії» Спільної операційної програми Румунія-Україна через Європейський інструмент сусідства («Partnership for Genomic Research in Ukraine and Romania» by the Joint Operational Programme Romania-Ukraine, through the European Neighbourhood Instrument (ENI)).

#### Подяки

База даних ROUA є частиною розвивальної інфраструктури для біоінформатики в Україні. Ми дякуємо всім учасникам BioinformaticsForUkraine.com та Консорціуму різноманітності геномів в Україні, які працювали з нами над розробленням інструментів для цього проекту.

ALEXANDER, D.H., NOVEMBRE, J., LANGE, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. DOI: 10.1101/gr.094052.109

BERGSTRÖM, A., MCCARTHY, S.A., HUI, R., ALMARRI, M.A., AYUB, Q., DANECSEK, P., CHEN, Y., FELKEL, S., HALLAST, P., KAMM, J., BLANCHÉ, H., DELEUZE, J.F., CANN, H., MALLICK, S., REICH, D., SANDHU, M.S., SKOGLUND, P., SCALLY, A., XUE, Y., DURBIN R., TYLER-SMITH, C. (2020) Insights into human genetic variation and population history from 929 diverse genomes. *Science*, 367(6484):eaay5012. DOI: 10.1126/science.aay5012

CINGOLANI, P., PLATTS, A., WANG, L.L., COON, M., NGUYEN, T., WANG, L., LAND, S.J., LU, X., RUDEN, D. M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80. DOI: 10.4161/FLY.19695

DEPRISTO, M.A., BANKS, E., POPLIN, R., GARIMELLA, K.V., MAGUIRE, J.R., HARTL, C., PHILIPPAKIS, A.A., DELANGEL, G., RIVAS, M.A., HANNA, M., MCKENNA, A., FENNEL, T.J., KERNYTSKY, A.M., SIVACHENKO, A.Y., CIBULSKIS, K., GABRIEL, S.B., ALTSHULER, D., DALY, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. DOI: 10.1038/ng.806

FAIRLEY, S., LOWY-GALLEGU, E., PERRY, E., FLICEK, P. (2020) The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research*, 48(D1), D941–D947. DOI: 10.1093/NAR/GKZ836

KÖSTER, J., MÖLDER, F., JABLONSKI, K.P., LETCHER, B., HALL, M.B., TOMKINS-TINCH, C.H., SOCHAT, V., FORSTER, J.,

LEE, S., TWARDZIOK, S.O., KANITZ, A., WILM, A., HOLTGREWE, M., RAHMANN, S., NAHNSEN, S. (2021) Sustainable data analysis with Snakemake. *F1000Research*, 10, 33. DOI: 10.12688/f1000research.29032.2

LANDRUM, M.J., LEE, J.M., BENSON, M., BROWN, G., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., HOOVER, J., JANG, W., KATZ, K., OVETSKY, M., RILEY, G., SETHI, A., TULLY, R., VILLAMARIN-SALOMON, R., RUBINSTEIN, W., MAGLOTT, D.R. (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, 44(D1), D862–D868. DOI: 10.1093/nar/gkv1222

LANDRUM, M.J., LEE, J.M., BENSON, M., BROWN, G.R., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., JANG, W., KARAPETYAN, K., KATZ, K., LIU, C., MADDIPATLA, Z., MALHEIRO, A., MCDANIEL, K., OVETSKY, M., RILEY, G., ZHOU, G., HOLMES, J.B., KATTMAN, B.L., MAGLOTT, D.R. (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062–D1067. DOI: 10.1093/nar/gkx1153

OLEKSYK, T.K., WOLFSBERGER, W.W., SCHUBELKA, K., MANGUL, S., O'BRIEN, S.J. (2022) The Pioneer Advantage: Filling the blank spots on the map of genome diversity in Europe. *GigaScience*, 11, 1–7. DOI: 10.1093/GIGASCIENCE/GIAC081

OLEKSYK, T.K., WOLFSBERGER, W.W., WEBER, A.M., SHCHUBELKA, K., OLEKSYK, O.T., LEVCHUK, O., PATRUS, A., LAZAR, N., CASTRO-MARQUEZ, S.O., HASYNETS, Y., BOLDYZHAR, P., NEYMET, M., URBANOVYCH, A., STAKHOVSKA, V., MALYAR, K., CHERVYAKOVA, S., PODOROHA, O., KOVALCHUK, N., RODRIGUEZ-FLORES, J.L., ZHOU, W., MEDLEY, S., BATTISTUZZI, F., LIU, R.,

- HOU, Y., CHEN, S., YANG, H., YEAGER, M., DEAN, M., MILLS, R.E., SMOLANKA, V. (2021) Genome diversity in Ukraine. *GigaScience*, 10(1), 1–14. DOI: 10.1093/GIGASCIENCE/GIAA159
- PRICE, A.L., PATTERSON, N.J., PLENGE, R.M., WEINBLATT, M.E., SHADICK, N.A., REICH, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909. DOI: 10.1038/ng1847
- SOLLIS, E., MOSAKU, A., ABID, A., BUNIELLO, A., CEREZO, M., GIL, L., GROZA, T., GÜNEŞ, O., HALL, P., HAYHURST, J., IBRAHIM, A., JI, Y., JOHN, S., LEWIS, E., MACARTHUR, J.A.L., MCMAHON, A., OSUMI-SUTHERLAND, D., PANOUTSOPOULOU, K., PENDLINGTON, Z., RAMACHANDRAN, S., STEFANCSIK, R., STEWART, J., WHETZEL, P., WILSON, R., HINDORFF, L., CUNNINGHAM, F., LAMBERT, S.A., INOUE, M., PARKINSON, H., HARRIS, L.W. (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, 51(D1), D977–D985. DOI: 10.1093/NAR/GKAC1010
- WOLFSBERGER, W.W. (2023) *PopGen Playground* (0.1). Available from: [https://github.com/wwolfsberger/OU\\_popgen\\_playground](https://github.com/wwolfsberger/OU_popgen_playground) (accessed 10.11.2023).