

УДК 004.8:575.111-047.37

DOI [https://doi.org/10.24144/2616-7700.2026.48\(1\).162-171](https://doi.org/10.24144/2616-7700.2026.48(1).162-171)**О. С. Куруца**

ДВНЗ «Ужгородський національний університет»,
аспірант кафедри інформаційних управляючих систем та технологій
o.kurutsa@uzhnu.edu.ua
ORCID: <https://orcid.org/0000-0001-7020-7996>

ПОВНОГЕНОМНИЙ ПОШУК ГЕНЕТИЧНИХ АСОЦІАЦІЙ ПОШИРЕНИХ ЗАХВОРЮВАНЬ В УКРАЇНІ: СУЧАСНІ МЕТОДИ ТА ІНСТРУМЕНТИ

Комплексні неінфекційні захворювання, зокрема серцево-судинні, онкологічні та аутоімунні, становлять основний тягар для системи охорони здоров'я України. Їх етіологія є результатом складної взаємодії генетичних факторів та впливу навколишнього середовища. Повногеномні асоціативні дослідження (GWAS) є ефективним інструментом для ідентифікації генетичних варіантів, пов'язаних із ризиком розвитку цих хвороб. Однак інтерпретація результатів GWAS залишається складним завданням, що потребує використання сучасних біоінформатичних методів аналізу.

У статті узагальнено сучасні підходи та інструменти для проведення GWAS і подальшої обробки даних з урахуванням особливостей української популяції. Розглянуто основні етапи дослідження — від контролю якості генотипових даних до пріоритетизації причинних варіантів із застосуванням інструментів на основі машинного навчання, зокрема Combined Annotation Dependent Depletion (CADD) та новітніх методів повногеномної регресії, таких як Regenie. Особливу увагу приділено зростаючій ролі штучного інтелекту у виявленні складних нелінійних залежностей у геномних даних. Підкреслено унікальність генофонду населення України, що обґрунтовує необхідність масштабних національних геномних досліджень для розроблення ефективних стратегій персоналізованої медицини та профілактики поширених захворювань.

Ключові слова: повногеномні асоціативні дослідження, геномний аналіз, біоінформатика, машинне навчання, штучний інтелект, генетичне прогнозування, персоналізована медицина, Україна.

1. Вступ. Сучасна система охорони здоров'я України, як і більшості країн світу, стикається з проблемою високої поширеності комплексних (багатофакторних) неінфекційних захворювань. За статистичними даними, серцево-судинні хвороби спричиняють понад 50% випадків смертності в Україні, а онкологічні патології посідають друге місце за цим показником. Значну медико-соціальну проблему становлять також ендокринні захворювання.

Особливе місце серед ендокринної патології займають аутоімунні захворювання щитоподібної залози (АІЗ ЩЗ) є органоспецифічними патологіями, які вражають близько 2-5% загальної популяції, причому поширеність серед жінок (5-15%) є суттєво вищою, ніж серед чоловіків (1-5%) [1]. Такі патології є наслідком складної взаємодії між генетичною схильністю індивіда та факторами навколишнього середовища, зокрема способом життя, харчуванням і станом екосистеми. Розкриття генетичної архітектури цих хвороб має ключове значення для розроблення нових підходів до діагностики, профілактики та персоналізованого лікування.

Револьюційний прогрес у дослідженні генетичних чинників людини забезпечили повногеномні асоціативні дослідження (GWAS), що дають змогу без

попередніх гіпотез аналізувати геном та виявляти варіанти, асоційовані з ризиком розвитку захворювань. Завдяки цим дослідженням ідентифіковано тисячі локусів, пов'язаних із сотнями хвороб та фенотипових ознак. Ефективне застосування GWAS потребує великих обсягів якісних геномних даних, значних обчислювальних ресурсів і високої кваліфікації фахівців. Для України тривалий час це залишалося обмеженням через технологічні та фінансові бар'єри.

Попри значні досягнення, трансляція результатів GWAS у клінічну практику залишається складним завданням. Більшість виявлених асоціацій локалізуються в некодуючих ділянках геному, мають незначний індивідуальний ефект, а їхня функціональна роль часто невідома [2]. Це зумовлює появу проблеми пост-GWAS аналізу: серед тисяч статистично значущих асоціацій необхідно виокремити біологічно релевантні варіанти, встановити причинні гени та механізми.

Розв'язання цієї проблеми потребує інтеграції геномних даних із функціональною анотацією, що здійснюється за допомогою сучасних біоінформатичних інструментів, методів машинного навчання та штучного інтелекту [3].

Метою цієї статті є огляд сучасних методів проведення GWAS та подальшого аналізу з акцентом на їх застосуванні для вивчення поширених в Україні захворювань. Розглянемо ключові обчислювальні інструменти, зокрема Regenie для ефективного регресійного моделювання та CADD для функціональної пріоритетизації варіантів. Окремо обґрунтуємо необхідність проведення повногеномних досліджень на українській популяції, зважаючи на її унікальну генетичну різноманітність.

2. Основний результат.

2.1. Генетичні дослідження в Україні. Однією з ключових проблем сучасної біомедичної науки в Україні є відсутність великомасштабних геномних досліджень поширених захворювань. Більшість проведених у світі повногеномних асоціативних досліджень (GWAS) базуються на вибірках осіб переважно європейського походження, що обмежує можливість прямої екстраполяції їхніх результатів на українську популяцію.

Нещодавні дослідження геномної різноманітності населення України, виконані під керівництвом Тараса Олексика, засвідчили унікальність українського генофонду. У ході секвенування геномів 97 осіб було виявлено понад 13 мільйонів генетичних варіантів, серед яких близько 500 тисяч раніше не були описані. Встановлено також, що частоти багатьох клінічно значущих мутацій в українців істотно відрізняються від інших європейських популяцій [4].

Ці дані підкреслюють необхідність проведення системних досліджень генетичних детермінант поширених в Україні патологій. Зокрема, існує висока ймовірність, що фактори генетичного ризику для серцево-судинних, ендокринних та онкологічних захворювань у представників української популяції мають специфічні відмінності від тих, що виявлені в інших етнічних групах.

У цьому контексті актуальними стратегічними завданнями є:

- створення національної інфраструктури для проведення геномних досліджень, включно з формуванням біобанків для збору, зберігання та аналізу біологічних зразків;
- реалізація повномасштабних GWAS, спрямованих на виявлення популяційно специфічних генетичних маркерів ризику поширених захворювань.

Важливим кроком у цьому напрямі стало ухвалення Кабінетом Міністрів України «Порядку збору, зберігання та використання біологічних зразків людини з дослідницькою метою» [5], що створює необхідне правове підґрунтя для розвитку біобанкінгу та геномних ініціатив у країні.

Попри обмежену кількість масштабних проєктів, в Україні вже реалізуються перспективні наукові ініціативи, які демонструють значний потенціал у галузі геноміки. Зокрема, міжнародний проєкт «Геноміка цукрового діабету 1 типу в Україні» передбачає залучення 20000 учасників (10000 пацієнтів та 10000 осіб контрольної групи) для проведення повноекзомного секвенування та генотипування. Незважаючи на обмеження, пов'язані з воєнними діями, у межах проєкту вже зібрано понад 13500 зразків і отримано високоякісні дані майже для 10000 учасників, що заклало основу для формування першого у Східній Європі каталогу генетичних варіантів ЦД1 та дозволило ідентифікувати випадки моногенного діабету [6].

2.2. GWAS і штучний інтелект у сучасних геномних дослідженнях.

Методологія GWAS: від даних до асоціацій. Проведення GWAS є багатапним процесом, що потребує ретельного планування та суворого контролю якості на всіх стадіях дослідження [7]. Вихідні генотипні дані зазвичай формуються у форматі VCF (Variant Call Format), однак з огляду на значні обсяги інформації та обчислювальні вимоги їх часто конвертують у бінарні формати, сумісні зі спеціалізованими інструментами аналізу. Найбільш поширеним серед таких інструментів є програмний пакет PLINK, який фактично став стандартом для опрацювання GWAS-даних.

Однією з ключових складових аналітичного конвеєра є контроль якості даних, спрямований на мінімізацію впливу технічних артефактів і помилок генотипування, що можуть зумовлювати хибнопозитивні асоціації. Типовий протокол контролю якості, реалізований у PLINK, передбачає фільтрацію як зразків, так і генетичних варіантів за показниками частки пропущених генотипів, частоти мінорного алеля (MAF) та відхилення від рівноваги Харді–Вайнберга (HWE), а також ідентифікацію та усунення прихованої спорідненості між учасниками дослідження [8].

Важливим аспектом є також урахування популяційної стратифікації, оскільки ігнорування популяційної структури може призводити до систематичних зміщень оцінок асоціації. З огляду на складну демографічну історію української популяції, доцільним є застосування аналізу головних компонент (PCA), який дає змогу виявити основні напрями генетичної варіабельності між індивідами та включити відповідні компоненти як коваріати у регресійні моделі.

Статистичний аналіз у GWAS ґрунтується на тестуванні асоціації між кожним окремим генетичним варіантом (SNP) і досліджуваною ознакою, зокрема за допомогою лінійної або логістичної регресії залежно від типу фенотипу. Отримані результати зазвичай візуалізують у вигляді мангеттенських графіків, що відображають геномний розподіл сигналів асоціації. Оскільки кількість проведених тестів сягає мільйонів, обов'язковим є застосування корекцій на множинні порівняння, серед яких одним із найконсервативніших підходів є корекція Бонферроні.

2.3. Приклад застосування GWAS: аутоімунні захворювання щитоподібної залози. Аутоімунні захворювання щитоподібної залози, зокрема

аутоімунний тиреоїдит (тиреоїдит Хашимото), є показовою моделлю комплексних патологій, генетична архітектура яких інтенсивно вивчається за допомогою GWAS. Результати таких досліджень засвідчили наявність численних генетичних локусів, що відображають складну взаємодію між системною імунною схильністю та тиреоїд-специфічними біологічними механізмами [9].

Ідентифіковані гени-асоціати умовно групуються за їхньою біологічною роллю у формуванні ризику захворювання. Одна частина з них пов'язана із загальною аутоімунною схильністю та підвищує ймовірність розвитку різних аутоімунних патологій, включно з цукровим діабетом 1 типу та ревматоїдним артритом. До цієї категорії належить регіон HLA на 6-й хромосомі, який кодує білки головного комплексу гістосумісності та відіграє ключову роль у презентації антигенів Т-лімфоцитам. Вагомий внесок також мають варіанти гена RTPN22, що кодує тирозинфосфатазу – негативний регулятор активації Т-клітин, порушення функції якого може знижувати контроль імунної відповіді. Подібно до цього, ген CTLA4 залучений до підтримання імунної толерантності через інгібування активації Т-лімфоцитів, а його поліморфізми асоційовані з підвищеним ризиком аутоімунних реакцій.

Інша група генів має більш безпосередній зв'язок із біологією щитоподібної залози. Зокрема, FOXE1 кодує транскрипційний фактор, критично важливий для ембріонального розвитку та морфогенезу щитоподібної залози, що зумовлює його значення у формуванні органоспецифічної схильності. Водночас ген TSHR, який кодує рецептор тиреотропного гормону, є важливим функціональним елементом регуляції тиреоїдної активності; його генетичні варіанти найчастіше демонструють асоціації з хворобою Грейвса. У сукупності ці дані підкреслюють полігенну природу АІЗ ШЗ і свідчать про поєднання загальноімунних і органоспецифічних генетичних чинників у патогенезі цих захворювань.

Цей приклад демонструє, як GWAS допомагає виявляти складні біологічні шляхи, що лежать в основі патології, вказуючи на конкретні гени та механізми, які можуть стати цілями для подальших досліджень та розробки терапії.

2.4. Regenie: тестування на основі повногеномної регресії. Традиційні GWAS-моделі, які аналізують кожен варіант окремо, є обчислювально затратними та не завжди враховують полігенну природу складних захворювань. Метод Regenie, розроблений дослідницькою групою корпорації Regeneron, базується на машинному навчанні для побудови повногеномної регресійної моделі.

Підхід Regenie складається з двох етапів. На першому етапі будується регресійна модель на основі підмножини високоякісних генетичних маркерів, що дозволяє ефективно врахувати полігенний ефект, спорідненість та популяційну структуру без необхідності обчислення громіздкої матриці генетичної спорідненості (GRM).

На другому етапі проводиться швидке тестування асоціацій для кожного варіанта окремо, враховуючи вже розраховані на першому етапі полігенні предиктори. Такий підхід істотно прискорює аналіз і дозволяє одночасно працювати з багатьма фенотипами [10].

Ключовою перевагою Regenie є можливість проведення аналізу на рівні генів із використанням агрегованих тестів (burden, SKAT, ACAT), що забезпечує виявлення сукупного ефекту рідкісних варіантів у межах одного гена. Це особливо актуально для виявлення генів, у яких накопичення рідкісних мутацій

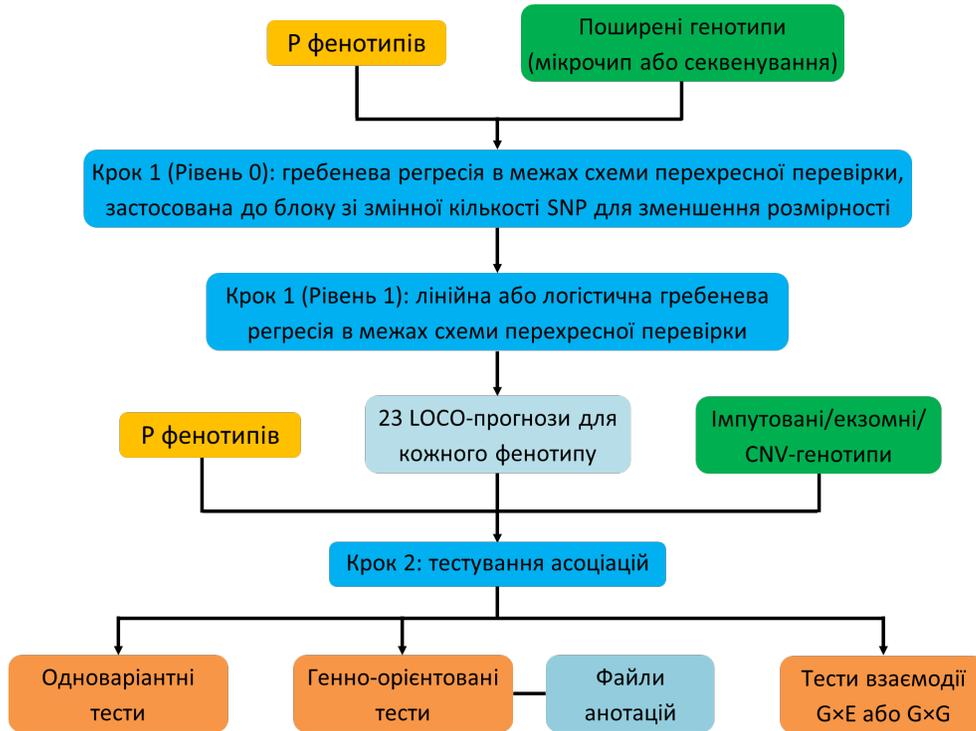


Рис. 1. Схема повногеномного асоціативного аналізу за допомогою Regenie.

може суттєво впливати на ризик захворювання, що часто пропускається стандартними GWAS.

У порівнянні з BOLT-LMM та SAIGE, алгоритм regenie демонструє суттєві переваги з точки зору обчислювальної ефективності, що є критично важливим для аналізу великих когорт.

Швидкодія regenie суттєво перевищує BOLT-LMM, забезпечуючи понад 30-разове загальне прискорення на етапах побудови нульової моделі та асоціативного тестування. Порівняно з SAIGE, regenie приблизно у 350 разів швидшим на кроці 1, використовував лише 40% пам'яті, а на кроці 2 демонстрував у десятки разів менший час виконання завдяки ефективній паралелізації. Максимальне використання оперативної пам'яті для regenie становило 12,9 ГБ, натомість BOLT-LMM потребував до 50 ГБ.

Поєднання високої швидкодії та помірною використання пам'яті дозволяє regenie забезпечувати стабільну масштабованість для аналізу як кількісних, так і бінарних фенотипів, включно з ознаками з вираженим дисбалансом «випадок-контроль». Це робить regenie оптимальним вибором для високопродуктивних GWAS-аналізів у великих популяційних дослідженнях.

2.5. CADD: інструмент пріоритезації варіантів на основі машинного навчання. Інтеграція алгоритмів передбачення патогенності варіантів дозволяє ідентифікувати мутації, що потенційно впливають на ризик розвитку захворювання. Одним із найефективніших таких інструментів є Combined Annotation Dependent Depletion (CADD), який поєднує десятки анотацій у єдину оцінку шкідливості (C-score) за допомогою моделі машинного навчання.

CADD навчається розрізняти нейтральні варіанти, що збереглися в еволю-

ції, та симульовані *de novo* мутації, використовуючи різноманітні геномні характеристики (рівень консервативності, дані ENCODE, функціональні наслідки тощо).

PHRED-шкальована оцінка CADD забезпечує інтуїтивну інтерпретацію результатів, оскільки її значення відображають відносний ранг потенційної шкідливості варіанта в масштабі всього геному: показник 10 відповідає верхнім 10% імовірно найбільш шкідливих варіантів, 20 – 1%, а 30 – 0,1%. Така шкала спрощує порівняння варіантів між собою та створює зручну основу для їх пріоритезації у подальших функціональних дослідженнях [11].

Водночас, попри широку застосовність, підхід CADD має низку концептуальних і методологічних обмежень, які не завжди належно враховуються під час інтерпретації результатів. Одним із них є проблема циклічності, що зумовлена використанням у моделі як предикторів виходів інших інструментів, зокрема SIFT і PolyPhen, унаслідок чого CADD частково успадковує їхні систематичні похибки. Додатковим джерелом невизначеності є упередженість навчальних міток, адже припущення про шкідливість усіх симульованих мутацій є спрощенням, яке не відображає значної частки нейтральних варіантів у геномі. Це призводить до зашумлення тренувальної вибірки та зниження точності, особливо в некодуючих регіонах. Крім того, включення частоти алелів як однієї з ознак формує схильність моделі автоматично присвоювати вищі бали рідкісним варіантам, що може зміщувати оцінку їхнього функціонального значення. У сукупності ці чинники зумовлюють необхідність обережної інтерпретації CADD-скорю та його використання у поєднанні з іншими джерелами доказів.

2.6. Пост-GWAS аналіз та роль штучного інтелекту. Окрім CADD та Regenie, які самі по собі є прикладами застосування машинного навчання, штучний інтелект (ШІ) все глибше інтегрується в геномні дослідження. Моделі глибокого навчання (Deep Learning), зокрема згорткові нейронні мережі (CNN), демонструють високу ефективність у прогнозуванні функціонального впливу некодуючих варіантів, аналізуючи послідовність ДНК безпосередньо та виявляючи складні регуляторні мотиви.⁷

Ідентифікація статистично значущого SNP не обов'язково означає виявлення причинної мутації. Через явище нерівноважного зчеплення (LD), знайдений SNP часто є лише маркером для цілого геномного регіону, що містить десятки або сотні варіантів, будь-який з яких може бути функціональним [12]. Для визначення найбільш ймовірних причинних варіантів застосовують методи пост-GWAS аналізу, що інтегрують дані функціональної геноміки [13].

Застосування методів штучного інтелекту у пост-GWAS аналізі суттєво розширює аналітичні можливості дослідження генетичних асоціацій та їхнього біологічного тлумачення. Зокрема, алгоритми глибокого навчання використовуються для підвищення точності пріоритезації генетичних варіантів і генів-кандидатів. Такі моделі, як DeepSEA, здатні прогнозувати функціональні наслідки генетичних варіантів, зокрема їхній вплив на епігенетичні маркери у різних типах клітин, що дає змогу оцінювати регуляторний потенціал варіантів поза межами кодувальних ділянок геному.

Крім того, підходи на основі ШІ відкривають можливість виявлення нелінійних міжгенних взаємодій, або епістазу, коли сукупний ефект кількох варіантів не зводиться до простої суми їхніх індивідуальних внесків. Такі складні

взаємозв'язки часто залишаються поза межами класичних статистичних моделей, але можуть мати вагоме значення для формування ризику складних захворювань. Важливою перевагою є також здатність ШІ інтегрувати різноманітні мультиоміксні дані, включно з геномікою, транскриптомікою та протеомікою, у межах єдиних аналітичних моделей, що сприяє більш цілісному розумінню молекулярних механізмів хвороби.

Окремий напрям становить пояснюваний штучний інтелект (ХАІ), який орієнтований на підвищення інтерпретованості моделей. Використання таких підходів дає змогу ідентифікувати SNP або гени, що роблять найбільший внесок у прогноз, тим самим поєднуючи високу прогностичну здатність моделей із можливістю їхнього біологічного тлумачення [14].

2.7. Результати досліджень в Україні. В Україні вже здійснюються важливі кроки у сфері медичної геноміки. Одним із прикладів є перше системне дослідження генетичних причин затримки розвитку та інтелектуальної недостатності у дітей, у якому було проаналізовано 417 випадків із використанням секвенування екзому та таргетних панелей. Це дозволило встановити молекулярний діагноз у 15,8% випадків і підтвердило ефективність використання локальних геномних даних для інтерпретації рідкісних варіантів [15].

На конференції ANSG 2024 було представлено попередні результати GWAS у рамках проєкту «Геноміка цукрового діабету 1 типу в Україні» на основі 9 000 зразків, що охоплюють усі регіони країни. Ідентифіковано низку генів-кандидатів на хромосомах 1, 6 та 11, а розширений аналіз наразі готується до публікації.

Ці ініціативи демонструють здатність української наукової спільноти інтегруватися у світовий геномний простір і формують підґрунтя для майбутніх національних проєктів у сфері персоналізованої медицини.

3. Висновки. Повногеномні асоціативні дослідження, у поєднанні з сучасними біоінформатичними інструментами та методами машинного навчання, становлять потужний підхід до розкриття генетичних механізмів поширених захворювань. Однак ефективність цього підходу безпосередньо залежить від належної репрезентації генетичної різноманітності популяцій, що досліджуються.

Аналіз останніх досліджень чітко демонструє, що населення України має унікальну генетичну структуру, яка наразі недостатньо представлена в міжнародних геномних базах даних. Це створює «сліпу зону» у розумінні генетичних ризиків найпоширеніших в Україні захворювань. Використання результатів GWAS, отриманих на інших популяціях, може бути неточним і навіть помилковим.

Таким чином, проведення масштабних повногеномних асоціативних досліджень, орієнтованих на українську популяцію, є не лише науковим пріоритетом, а й стратегічною передумовою для розвитку національної системи охорони здоров'я. Це відкриття можливість:

- ідентифікувати генетичні маркери ризику, специфічні для української популяції;
- поглибити розуміння патогенезу поширених захворювань у місцевому контексті;
- створити основу для розробки програм генетичного скринінгу та профі-

лактики;

- прокласти шлях до впровадження принципів персоналізованої медицини, де лікування та профілактичні заходи підбираються з урахуванням індивідуального генетичного профілю пацієнта.

Створення національних біобанків та розвиток обчислювальної інфраструктури для аналізу геномних даних із використанням передових інструментів, таких як Regenie, CADD та моделей штучного інтелекту, є критично важливими кроками на цьому шляху. Інтеграція зусиль наукової спільноти, медичної галузі та державних інституцій дозволить Україні стати повноцінним учасником глобальної геномної революції та зробити вагомий внесок у покращення громадського здоров'я.

Конфлікт інтересів

Автор заявляє, що немає конфлікту інтересів щодо даного дослідження, включаючи фінансовий, особистий, авторський або будь-який інший, який міг би вплинути на дослідження, а також на результати, представлені в даній статті.

Фінансування

Дослідження здійснено в рамках міжнародного проекту «Геноміка цукрового діабету 1 типу в Україні».

Доступність даних

Усі дані доступні в цифровій або графічній формі в основному тексті рукопису.

Використання штучного інтелекту

Автор підтверджує, що при створенні даної роботи не використовував технології штучного інтелекту.

Авторські права ©



(2026). Куруца О. С. Ця робота ліцензується відповідно до Creative Commons Attribution 4.0 International License.

Список використаної літератури

1. Ovcariček, P. P., Görgeš, R., & Giovanella, L. (2024). Autoimmune thyroid diseases. *Seminars in Nuclear Medicine*, 54(2), 219–236. <https://doi.org/10.1053/j.semnuclmed.2023.11.002>
2. Pilalis, E., Zisis, D., Andrinopoulou, C., Karamanidou, T., Antonara, M., Stavropoulos, T. G., & Chatziioannou, A. (2025). Genome-wide functional annotation of variants: a systematic review of state-of-the-art tools, techniques and resources. *Frontiers in pharmacology*, 16, 1474026. <https://doi.org/10.3389/fphar.2025.1474026>

3. Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics*, 101(1), 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
4. 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., Podoroza, O., Kovalchuk, N., Rodriguez-Flores, J. L., . . . , & Smolanka, V. (2021). Genome diversity in Ukraine. *GigaScience*, 10(1), g1aa159. <https://doi.org/10.1093/gigascience/g1aa159>
5. Kaminet Ministriv Ukrainy. (2024). Postanova No. 999 "Pro zatverdzhennia Poriadku zboru, zberihannia ta vykorystannia biolohichnykh zrazkiv liudyny z doslidnytskoiu metoiu". Retrieved from <https://zakon.rada.gov.ua/laws/show/999-2024-%D0%BF#Text> [in Ukrainian].
6. Shchubelka, K., Wolfsberger, W. W., Oleksyk, O., & et. al. (2025). Genomics of Type 1 Diabetes in Ukraine Initiative. *International Journal of Endocrinology (Ukraine)*, 21(4), 343–350. <https://doi.org/10.22141/2224-0721.21.4.2025.1557>
7. Uffelmann, E., Huang, Q. Q., Munung, N. S., & et al. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1, 59. <https://doi.org/10.1038/s43586-021-00056-9>
8. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
9. Hwangbo, Y., & Park, Y. J. (2018). Genome-Wide Association Studies of Autoimmune Thyroid Diseases, Thyroid Function, and Thyroid Cancer. *Endocrinology and Metabolism (Seoul, Korea)*, 33(2), 175–184. <https://doi.org/10.3803/EnM.2018.33.2.175>
10. Mbatchou, J., Barnard, L., Backman, J., & et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, 53, 1097–1103. <https://doi.org/10.1038/s41588-021-00870-7>
11. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886–D894. <https://doi.org/10.1093/nar/gky1016>
12. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., & Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*, 42 (Database issue), D1001–D1006. <https://doi.org/10.1093/nar/gkt1229>
13. Gallagher, M. D., & Chen-Plotkin, A. S. (2018). The Post-GWAS Era: From Association to Function. *American journal of human genetics*, 102(5), 717–730. <https://doi.org/10.1016/j.ajhg.2018.04.002>
14. Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. *Nature reviews. Genetics*, 20(7), 389–403. <https://doi.org/10.1038/s41576-019-0122-6>
15. Shchubelka, K., Turova, L., Wolfsberger, W., & et al. (2024). Genetic determinants of global developmental delay and intellectual disability in Ukrainian children. *Journal of Neurodevelopmental Disorders*, 16, 13. <https://doi.org/10.1186/s11689-024-09528-x>

Kurutsa O. O. Genome-Wide Analysis for Genetic Associations of Common Diseases in Ukraine: Current Methods and Tools.

Complex non-communicable diseases, including cardiovascular, oncological, and autoimmune disorders, represent a major burden on the healthcare system of Ukraine. Their etiology results from a complex interplay between genetic determinants and environmental factors. Genome-wide association studies (GWAS) are an effective tool for identifying genetic variants associated with disease susceptibility. However, the interpretation of GWAS findings remains a challenging task that requires advanced bioinformatic analysis.

This article summarizes modern approaches and tools for conducting GWAS and subsequent data processing, with a particular focus on the Ukrainian population. The key stages of analysis are outlined — from genotype data quality control to causal variant prioritization using machine learning-based tools, such as Combined Annotation Depen-

dent Depletion (CADD) and advanced genome-wide regression methods, such as Regenie. Special attention is given to the growing role of artificial intelligence in detecting complex nonlinear patterns in genomic data. The uniqueness of the Ukrainian gene pool is emphasized, highlighting the need for large-scale national genomic studies to support the development of effective strategies for personalized medicine and prevention of common diseases.

Keywords: genome-wide association studies (GWAS), genomic analysis, bioinformatics, machine learning, artificial intelligence, genetic risk prediction, personalized medicine, Ukraine.

Отримано: 05.10.2025

Прийнято: 25.11.2025

Опубліковано: 29.01.2026