

Рис 2.4. Схема процедури пошуку мотивів

РОЗДІЛ 3

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

3.1 Дослідження мережі взаємодій білків людини між собою та з білками вірусів

3.1.1 Дослідження асиметрії вірусно-людської мережі білкових взаємодій

Щоб краще зрозуміти взаємодію вірусних та людських білків на системному рівні, ми розглянули тенденції щодо кількості взаємодій, що ці білків утворюють. Крім того, Рис 1 підсумовує дані про білкову взаємодію, які ми використовували в нашому дослідженні (кількість білків, кількість взаємодій та їх розподіл між білками).

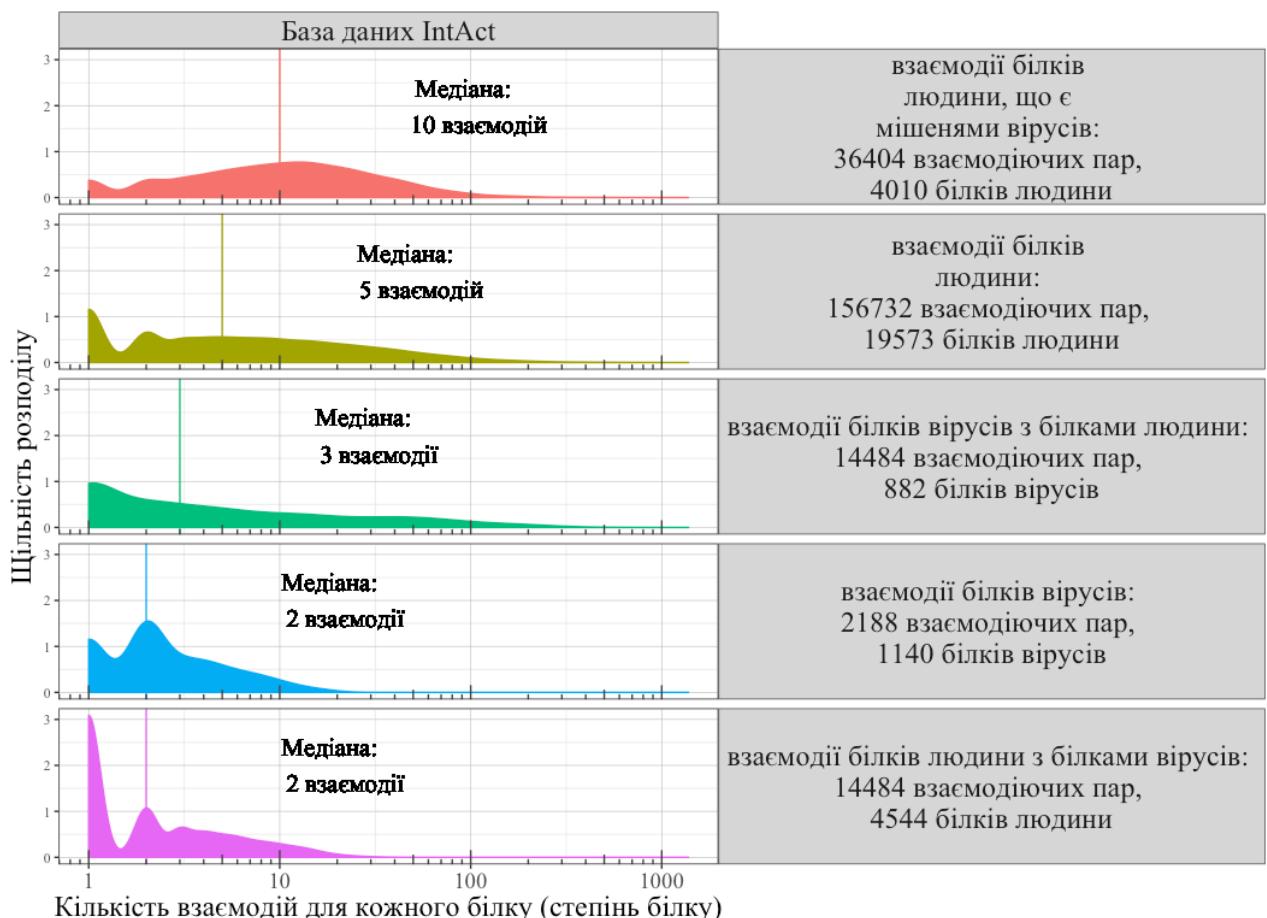


Рис 3.1.1. Графік, що показує щільність розподілу числа взаємодій кожного людського чи вірусного протеїну в кожній мережі, яку ми використовували для нашого аналізу. Для кожного білка вісь X показує

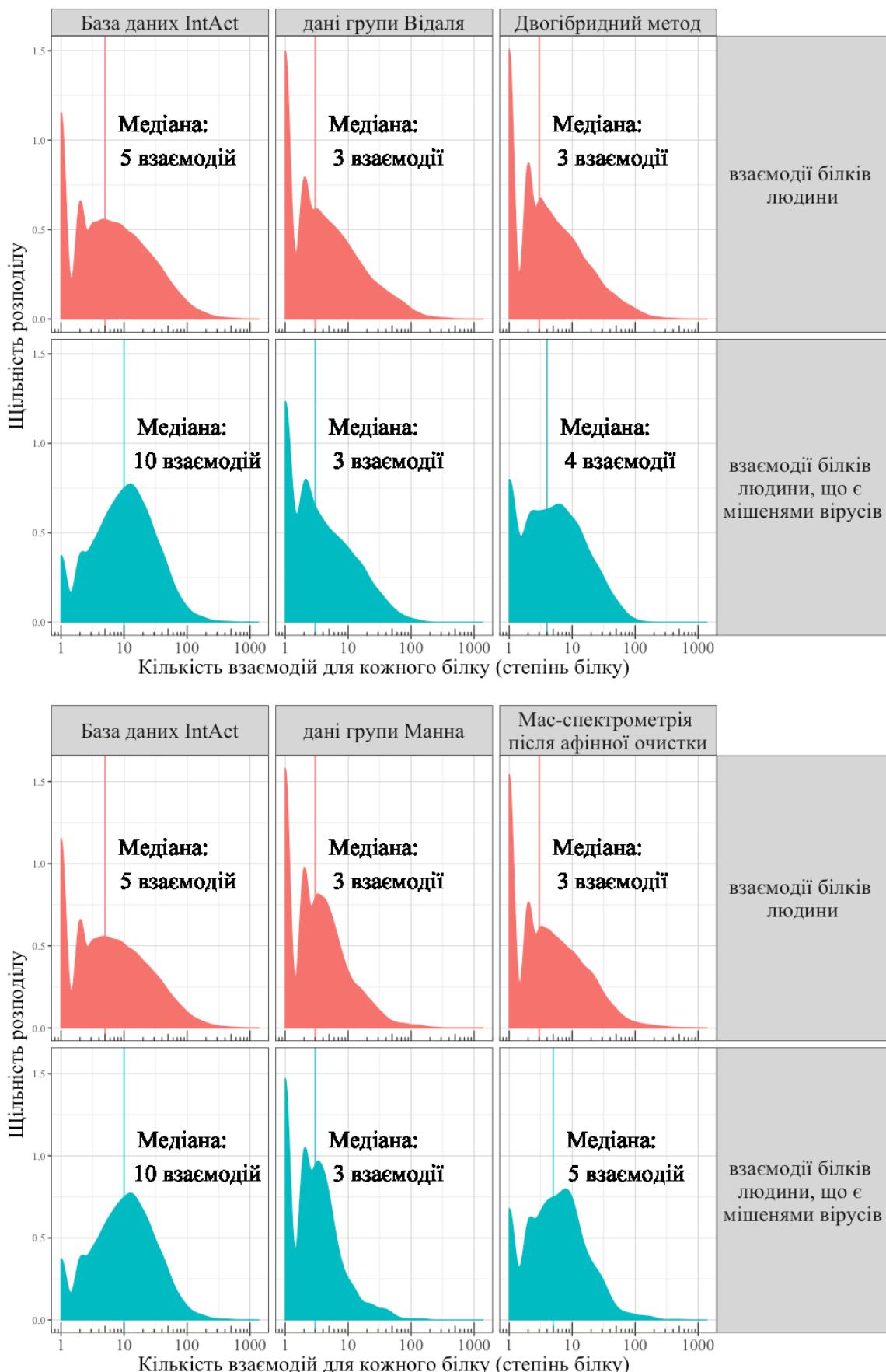


Рис 3.1.2. Графік, що показує щільність розподілу кількості взаємодій для всіх білків людини, білків людини-мішеней вірусів, для всіх даних взаємодій,

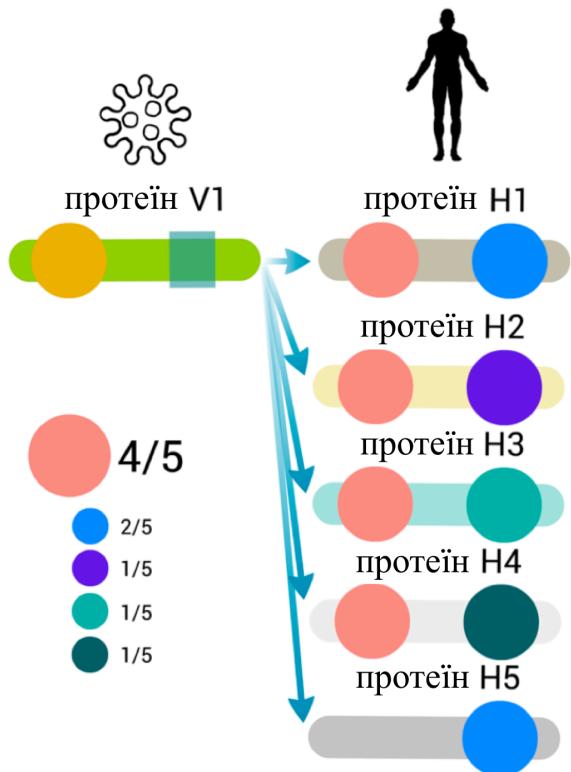


Рис 3.3. Схема методу визначення збагачених доменів.

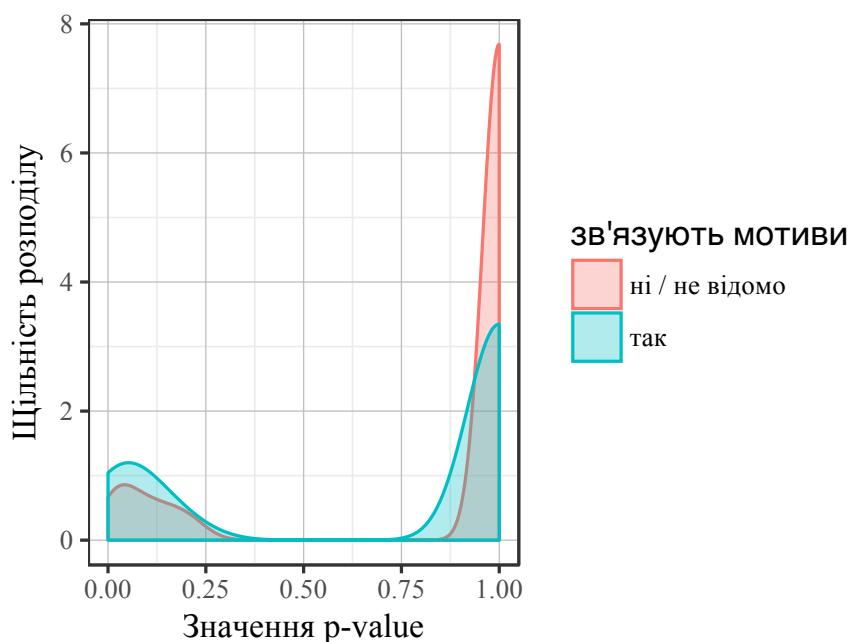


Рис 3.3. Графік щільності розподілу емпіричних значень p-value для відомих доменів, що зв'язують SLIM або для всіх інших доменів. Вісь X показує значення p-value; вісь Y показує щільність розподілу.

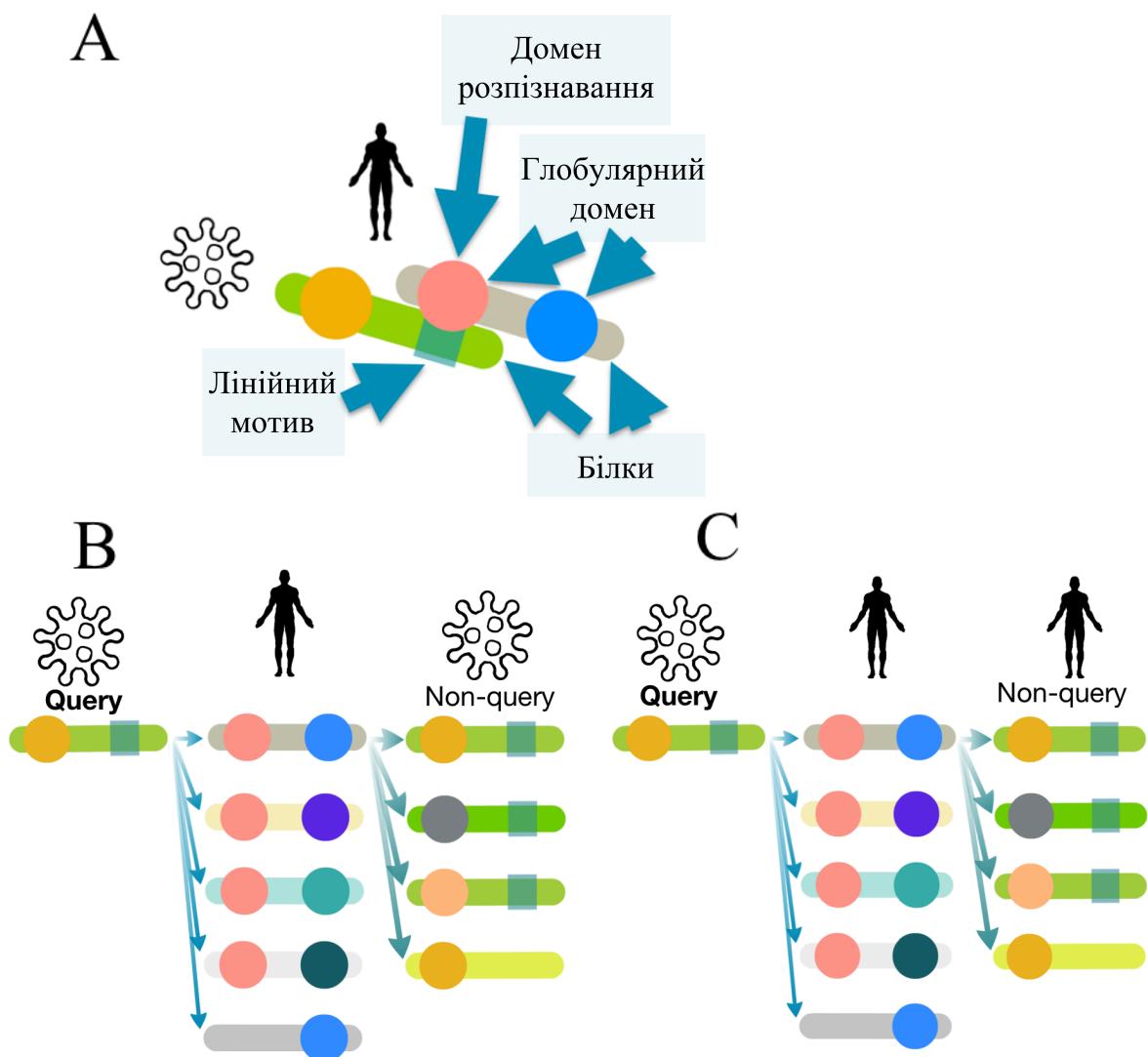


Рис 3.4.1. Схема, яка показує, як будуються набори даних для пошуку мотивів. Білки не запиту (non-query) використовувались для пошуку мотивів, які повинні бути присутніми в білку запиту (query). Кожен набір даних складається з усіх взаємодіючих партнерів одного білка людини та одного білка запиту (query). А. Легенда. В. Набори даних можуть бути побудовані з використанням білків людини, які взаємодіють з декількома вірусними білками. С. Білок-запит може бути вірусним білком, що мімікрує мотив, присутній у білках-не-запиту людини. Додавання цих білків-не запиту може забезпечити більшу потужність і спростити інтерпретацію по відношенню до сухо вірусного набору даних.

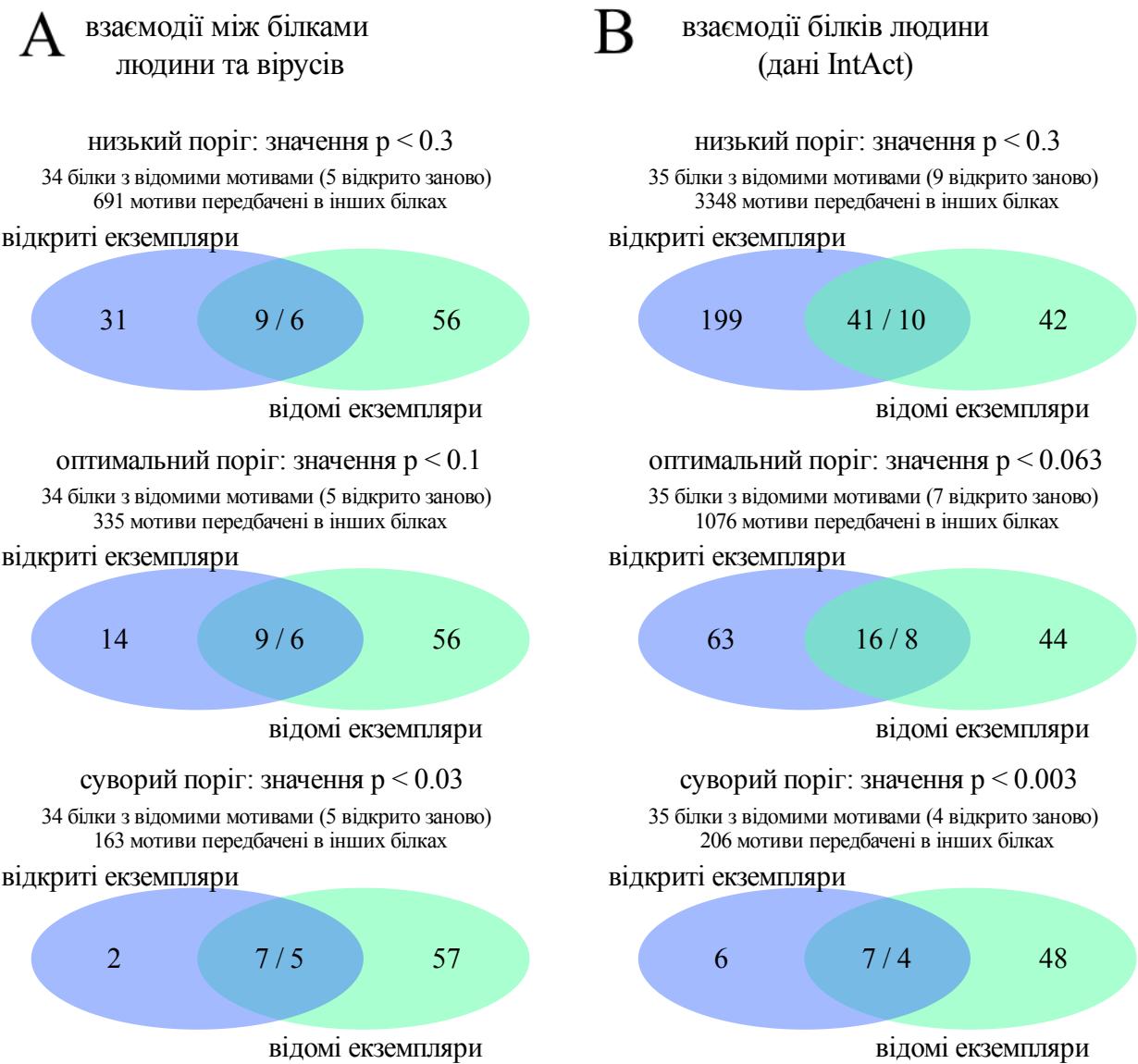


Рис 3.4.2. Діаграми Вена, що показують кількість виявлених мотивів-кандидатів та відомі мотиви, які відкрили заново. Наведено дві стратегії побудови наборів даних (A та B) та 3 пороги значень p -value. Блакитне коло показує кількість екземплярів мотивів, передбачених, але невідомих. Зелене коло показує кількість екземплярів мотивів, відомих, але не відкритих заново. Накладання показує кількість виявлених екземплярів, які відповідають відомим екземплярам (передбачені / відомі). Для кожного набору даних і порогу ми показуємо, скільки мотивів було виявлено у білках, які не містять відомих мотивів.

мотивів кандидатів, які ми виявили, використовуючи дані про взаємодію білків людини. По-перше, ми будемо розглядати подібність мотивів, потім детально розглядати конкретні випадки.

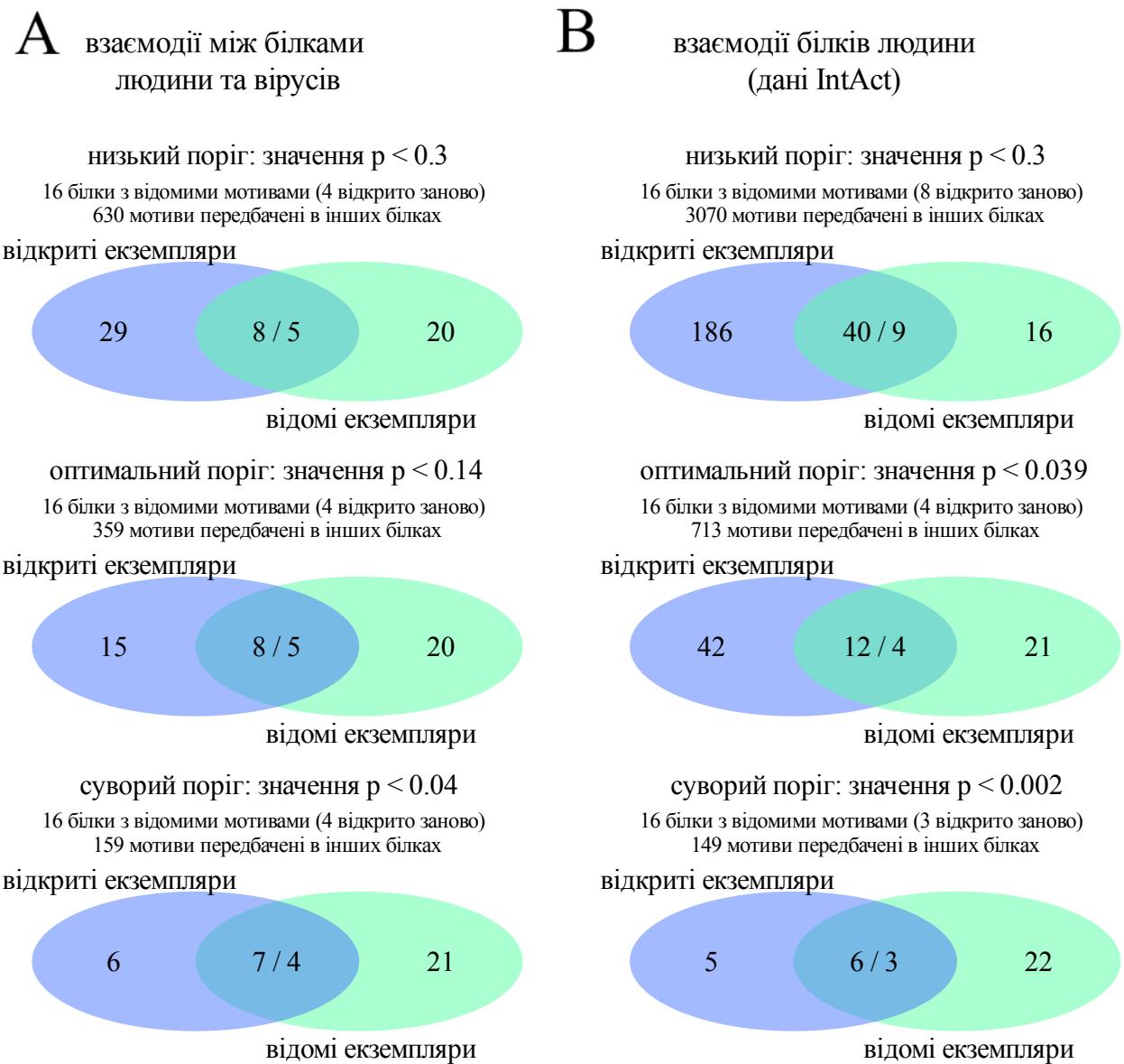


Рис 3.5. Діаграми Вена, що показують кількість знайдених мотивів кандидатів та відкритих відомих мотивів при фільтруванні по домену. Наведено дві стратегії побудови наборів даних (A та B) та пороги 3 значень p -value. Синє коло показує кількість екземплярів мотивів, передбачених, але невідомих. Зелене коло показує кількість екземплярів мотивів, відомих, але не відкритих заново. Накладання показує кількість виявлених екземплярів, які відповідають відомим (передбачених / відомих). Для кожного набору даних і

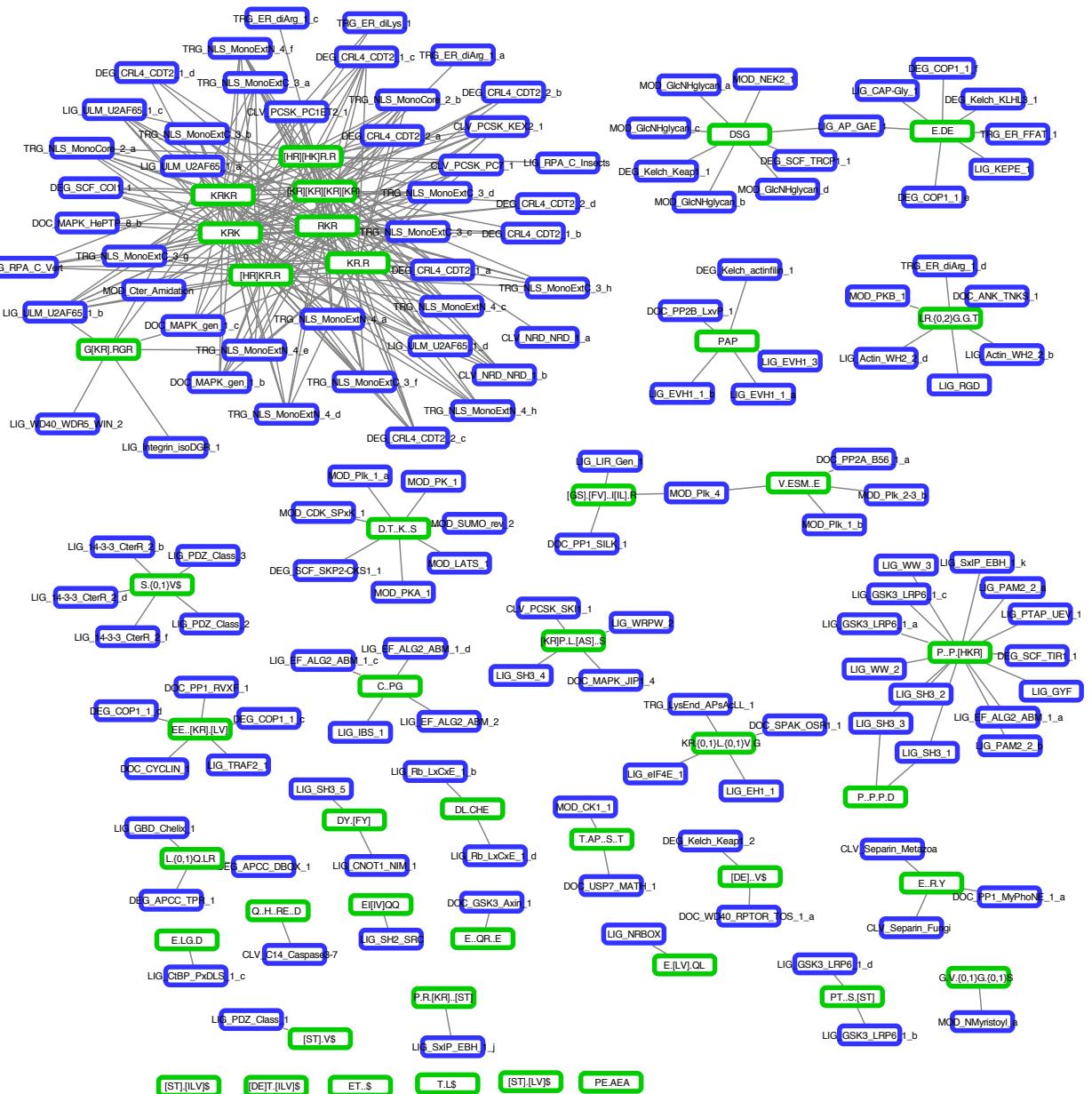


Рис 3.6.1. Схема, що показує подібність виявлених коротких лінійних мотивів до відомих мотивів. Зелені вузли - це послідовності виявлених мотивів, сині вузли - це відомі мотиви в базі даних ELM, ребра показують наявність подібності між мотивами вище порога 1,162 (Score, Comparomtif3).

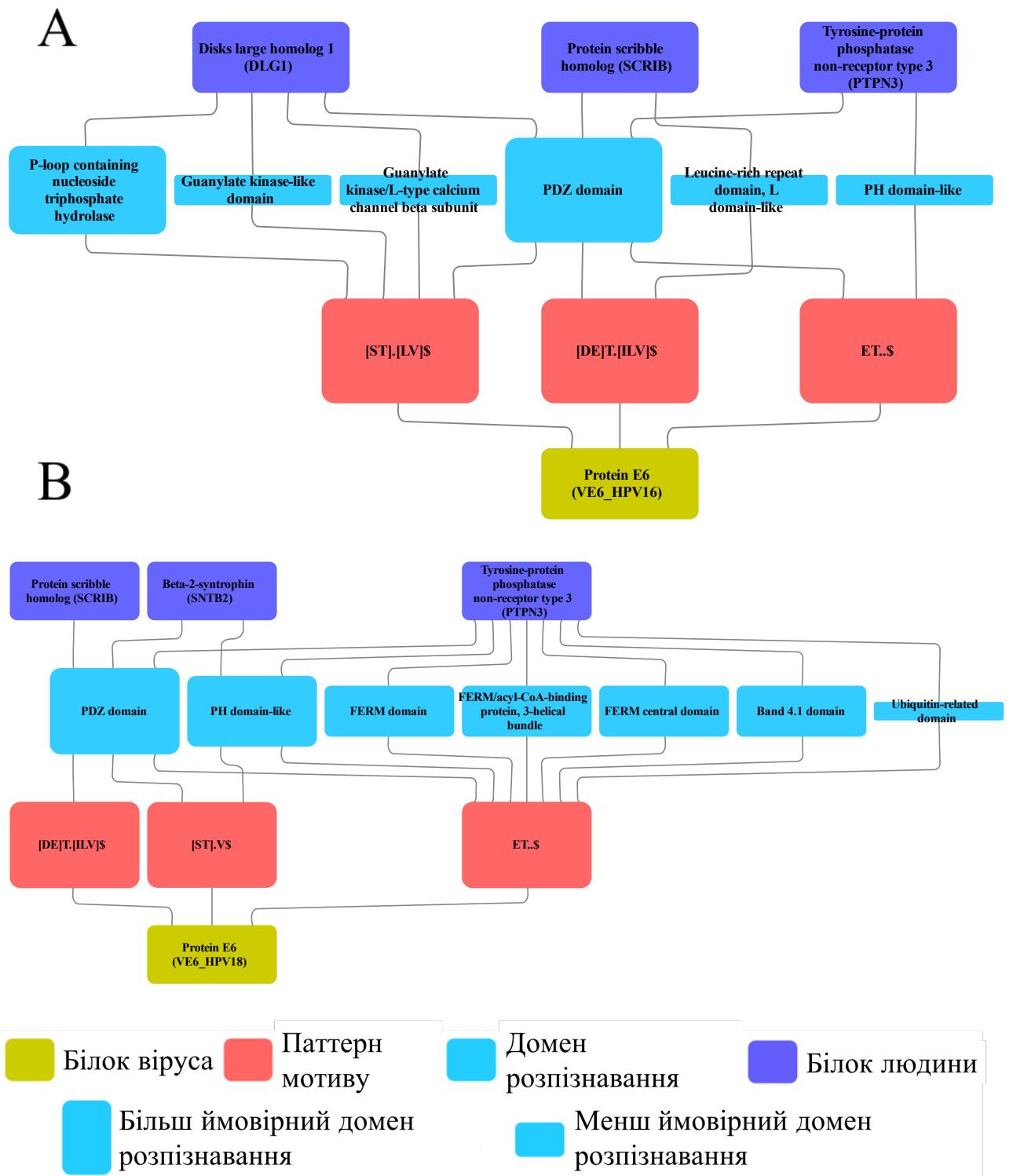


Рис 3.7.2. Схема мережі відомих мотивів PDZ в білку Е6 людського папіломавірусу 16 та 18. Ми показуємо варіанти мотиву та ті домени в білках людини, які можуть бути відповідальними за зв'язування протеїну Е6. Три варіанти цього мотиву були передбачені білком Е6 та білками людини, які

З точки зору функції, ERBIN служить як адаптерний протеїн, який зв'язує нефосфорильований receptor ERBB2, тим самим стабілізуючи цей стан [97]. Він є важливим для локалізації ERBB2 на базолатеральну сторону епітеліальних клітин [98]. З огляду на те, що HPV також зв'язує й інші білки, пов'язані з апікально-базальною полярністю клітини, такими як DLG1 та SCRIB, що було обговорено раніше, ERBIN може представляти реальну мішень.

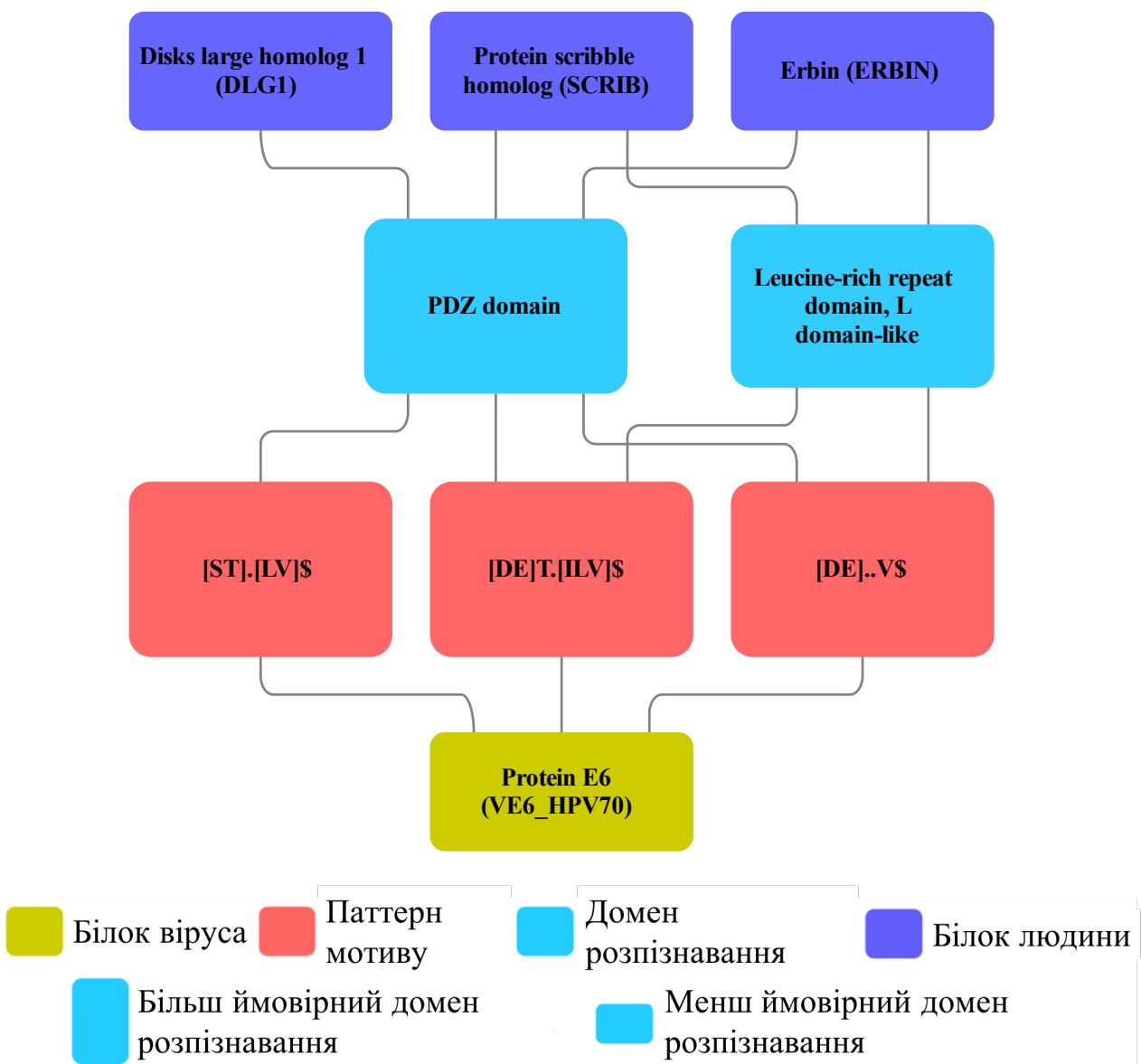


Рис 3.7.3.1. Схема мережі мотиву-кандидату PDZ в білках Е6 людського папіломавірусу 70 були підтвердженні в попередньому дослідженні, але не анотовані в ELM. Ми показуємо варіанти мотивів і ті домени в білках людини, які можуть бути відповідальними за зв'язування протеїну Е6. Три варіанти

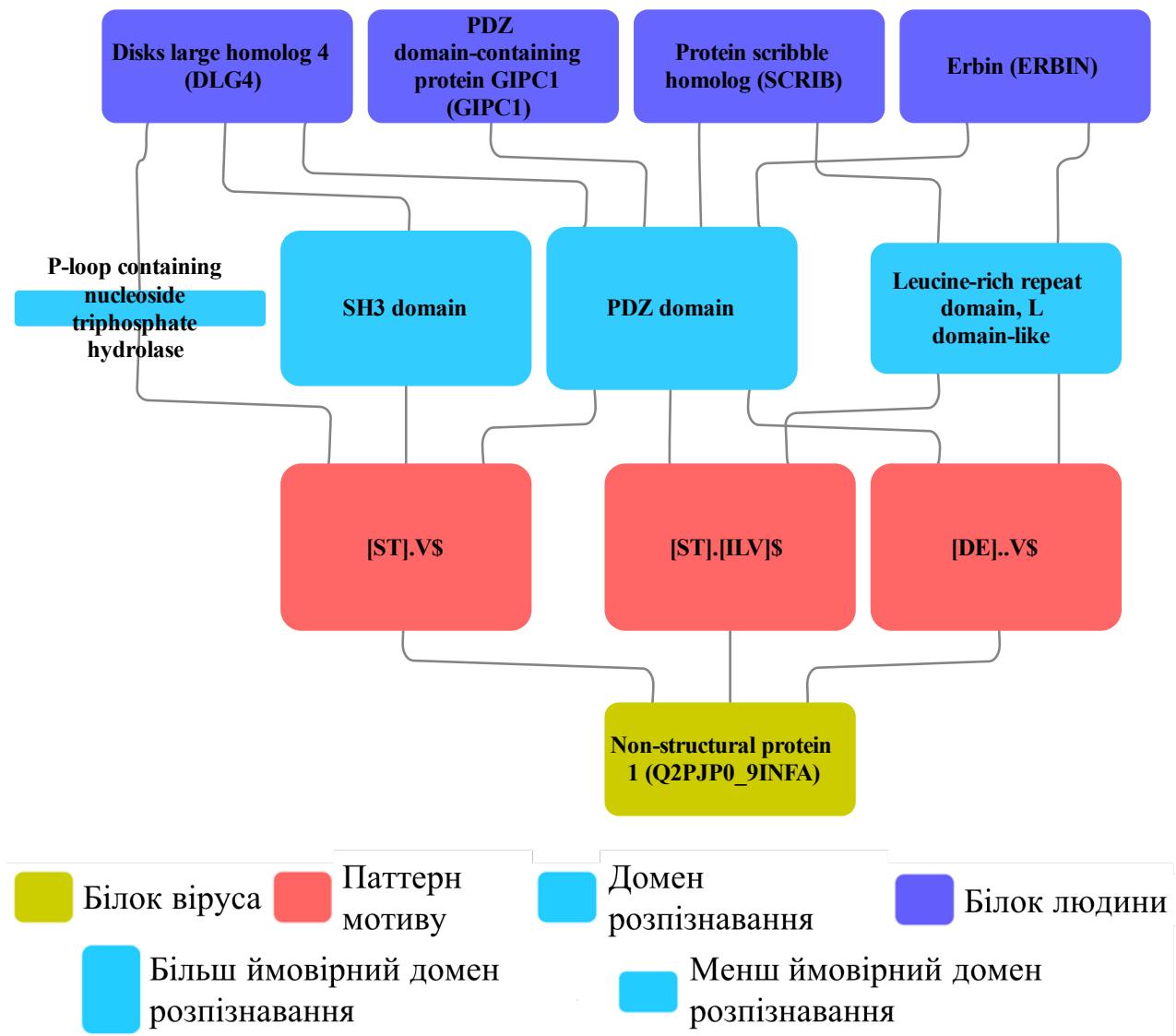


Рис 3.7.3.2. Схема мережі мотиву-кандидату PDZ в неструктурному білку 1 (NS1) вірусу H5N1 грипу були підтвердженні в попередньому дослідженні, але не були анотовані в ELM. Ми показуємо варіанти мотивів і ті домени в білках людини, які можуть бути відповідальними за зв'язування NS1. Три варіанти цього мотиву були передбачені у білку NS1 і у 86 білками людини, які взаємодіють з білками DLG4, GIPC1, SCRIB та ERBIN, що є мішенями вірусів. Домен PDZ є найбільш збагаченим серед мішеней білка E6. Високе збагачення домену SH3 і домену лейцін-багатого повтору може відображати функціональну перевагу NS1.

3.7.4 Мотив-кандидат, що зв'язує домен SH3

Ми виявили екземпляр домен SH3-зв'язуючого мотиву *de novo* в білку Nef віруса імунодефіциту людини типу 1 (рис. 3.7.4). Хоча ми не змогли ідентифікувати єдиного домену взаємодії, ми бачили, що послідовність, що містить Р..Р, нагадує канонічний ліганд домену SH3 [102]. Відомо, що Nef взаємодіє лише з 5 білками людини, з яких 4 поділяють доменну архітектуру SRC-кінази. Цей мотив дійсно є ще одним відомим прикладом, не зазначеним в базі даних ELM. Як підтверджено дослідженнями мутагенезу, мотив Р..Р.[HKR] дозволяє Nef зв'язати домен SH3 з сімейства SRC-кіназ для їх активації та сприяння вірусній патогенності [103, 104].

Тому мотив, що зв'язує домен SH3, в білку Nef є ще одним підтвердженим мотивом, який не був включений в наші навчальні дані.

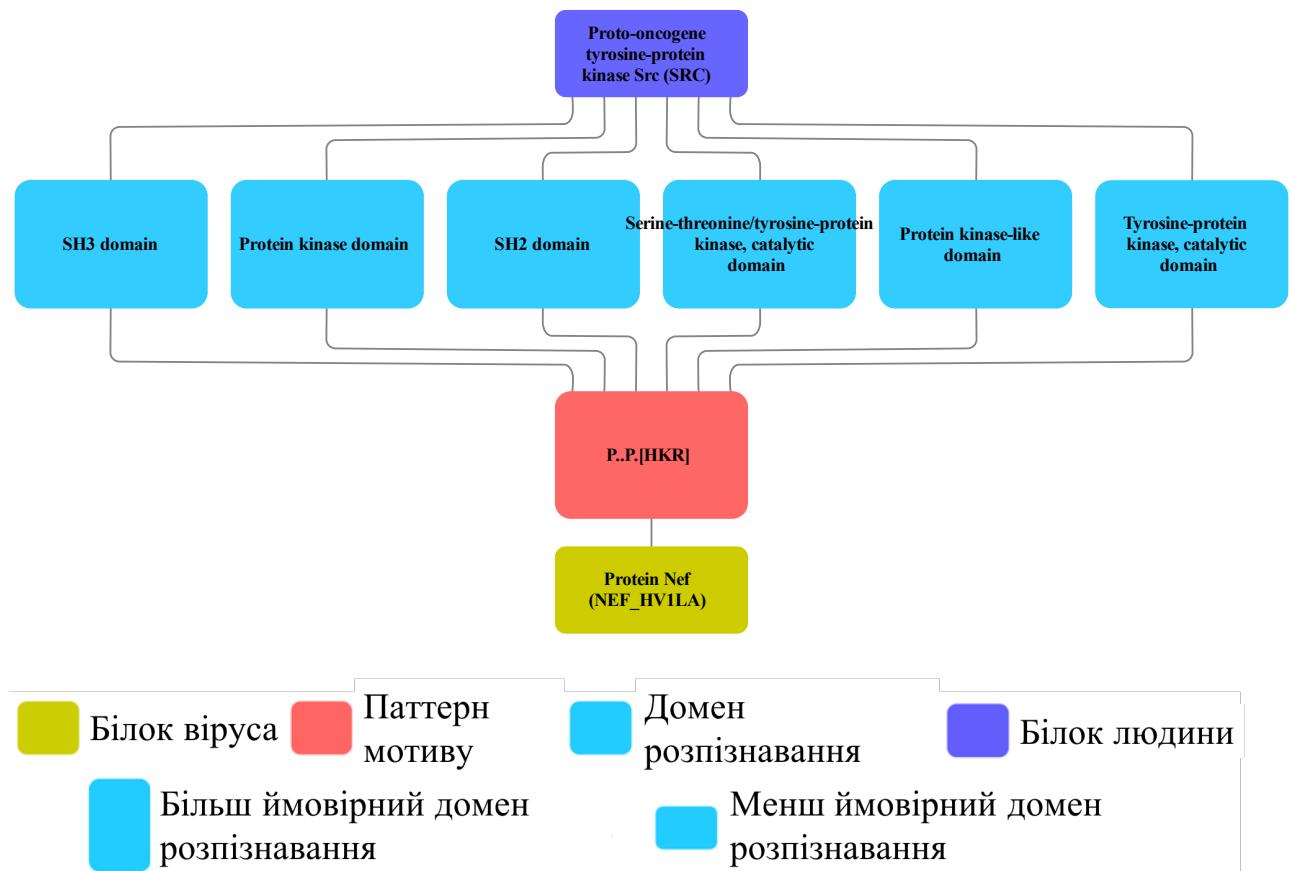


Рис 3.7.4. Схема мережі мотиву-кандидату SH3 в білку Nef вірусу Імунодефіциту людини типу 1 був підтверджений у попередньому дослідженні, але не був анотований в ELM. Ми показуємо ті домени в білках людини, які можуть бути відповідальними за зв'язування Nef. Один із варіантів

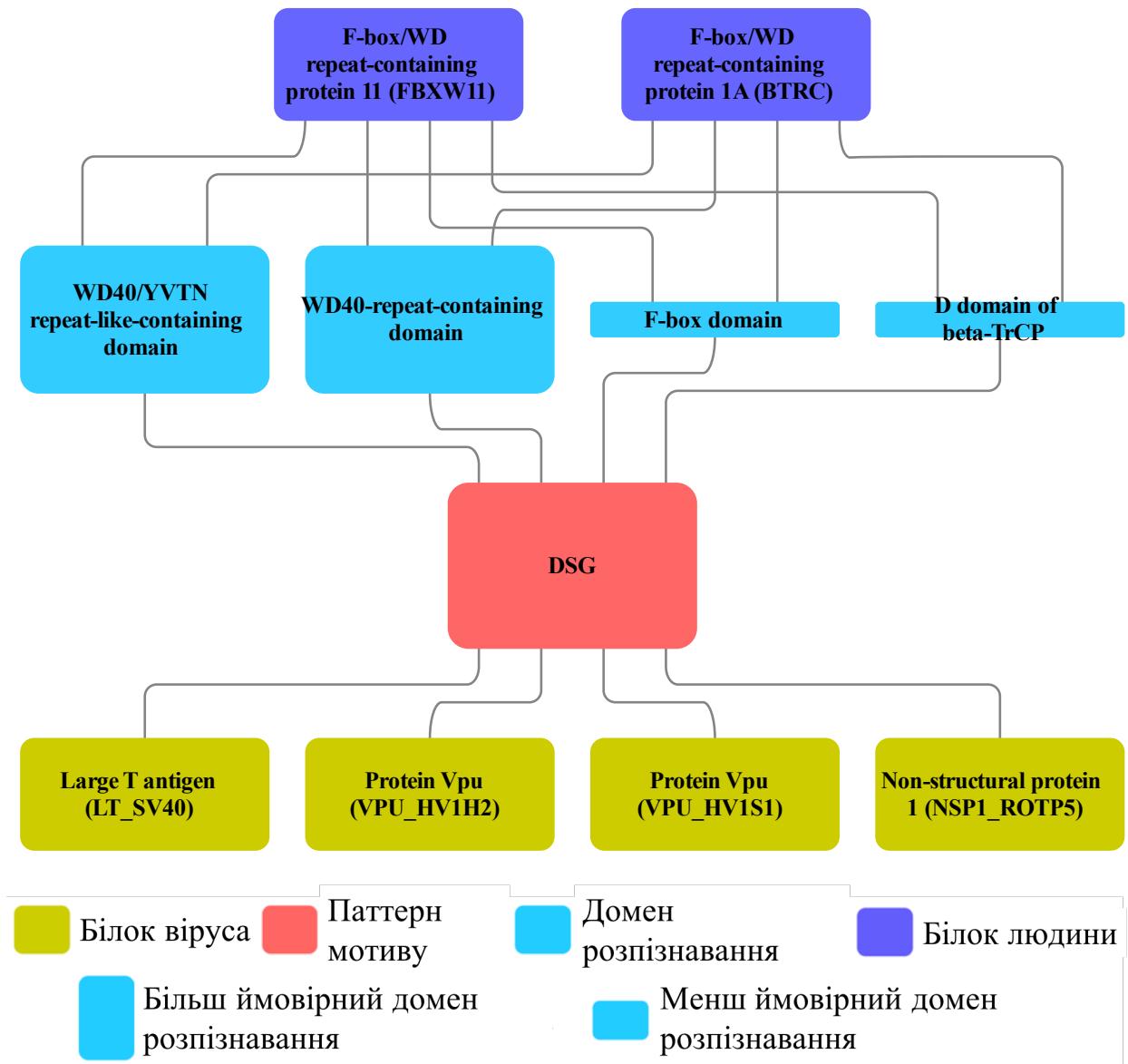


Рис 3.7.5.1. Схема мережі мотиву-кандидату DSG. Ці мотиви були передбачені в 4 вірусних білках. Всі вони обрали як мішень 2 субодиниці розпізнавання субстрату у комплексі SCF Е3 убіквітин лігази людини, FBXW11 та BTRC. 3 екземпляри мотиву були підтвердженні в попередньому дослідженні, але не були анотовані в ELM. Виняток становить LT у SV40. Домен WD40 найбільш збагачений серед мішеней вірусних білків VPU_HV1H2 та LT_SV40.

Другий мотив-кандидат, що зв'язує WD40 (рис 3.7.5.2), був передбачений в полімеразному лужному білку 2 (PB2) РНК-полімерази у 2 штамів вірусу грипу А (білок B4URF7 у штамі A/WS/1933 H1N1, білок C5E527 в A/New

York/1682/2009 H1N1). Ми також передбачаємо цей мотив у 4 людських білках, які всі зв'язують білок 1 елонгаторного комплексу людини (ELP1). ELP1 бере участь у елонгації транскрипції РНК-полімеразою 2 у складі комплексу, який відіграє роль в ремоделюваннях хроматину та ацетилює гістон H3 [22854966]. WD40 передбачено як найбільш імовірний домен, що підтримується 9/210 білками або 5/140 білками, що містять цей домен (для кожного штаму віруса відповідно). З огляду на РНК-полімеразну функцію PB2, ми можемо припустити, що він також викрадає фактори елонгації хазяїна використовуючи мотив E.V..G.{0,2}N.{0,1}Q для полегшення цього процесу.



Рис 3.7.5.2. Схема мережі мотиву-кандидату E.V..G.{0,2}N.{0,1}Q передбачений в полімеразному лужному білку 2 у 2-х штамів грипу А. Ми передбачаємо, що цей мотив розпізнається доменом WD40 в людському протеїні ELP1.

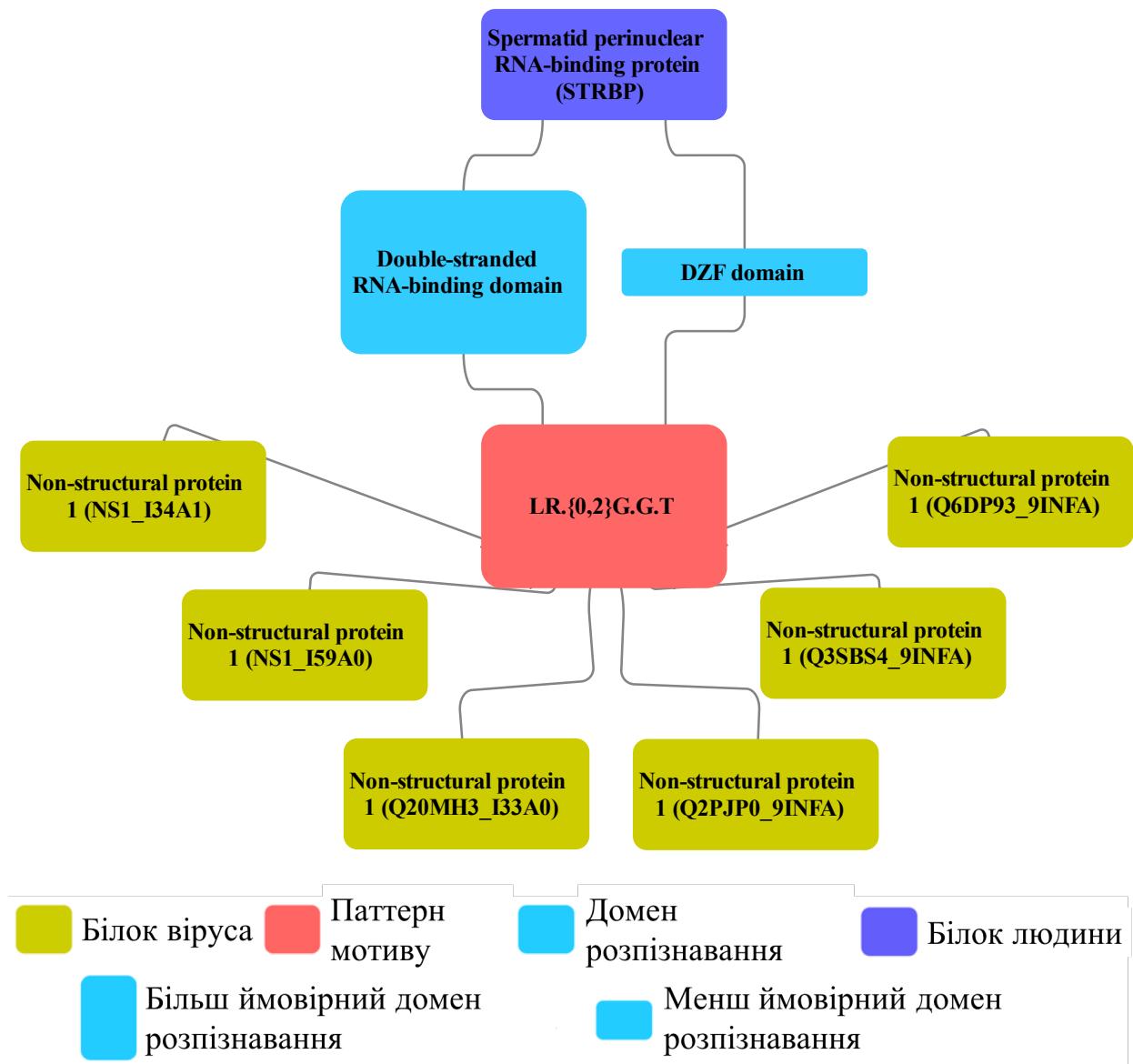


Рис 3.7.6.1. Схема мережі мотиву-кандидату LR.{0,2}G.G.T передбачений в 6 неструктурних білках різних штамів грипу А, що включають як пташину, так і людську лінію. Ми передбачаємо, що цей мотив розпізається дволанцюговим РНК-зв'язуючим доменом людського білка STRBP.

Ми можемо припустити, що Cab45 служить для полегшення цього процесу. Незважаючи на те, що це можливо, довіра до цього мотиву зменшується, оскільки на відміну від мотиву DSG, описаного раніше, цей мотив-кандидат, що зв'язує EF-hand, був виявлений лише в 4 з 33 білкових послідовностей (партнери Cab45), і релевантність домену підтримується тільки 2 з 45 РВ1-зв'язуючих білків (рис. 13). Дослідження 3 білків людини, в

яких передбачено цей мотив, може прояснити, наскільки цей мотив може бути справжнім хітом.

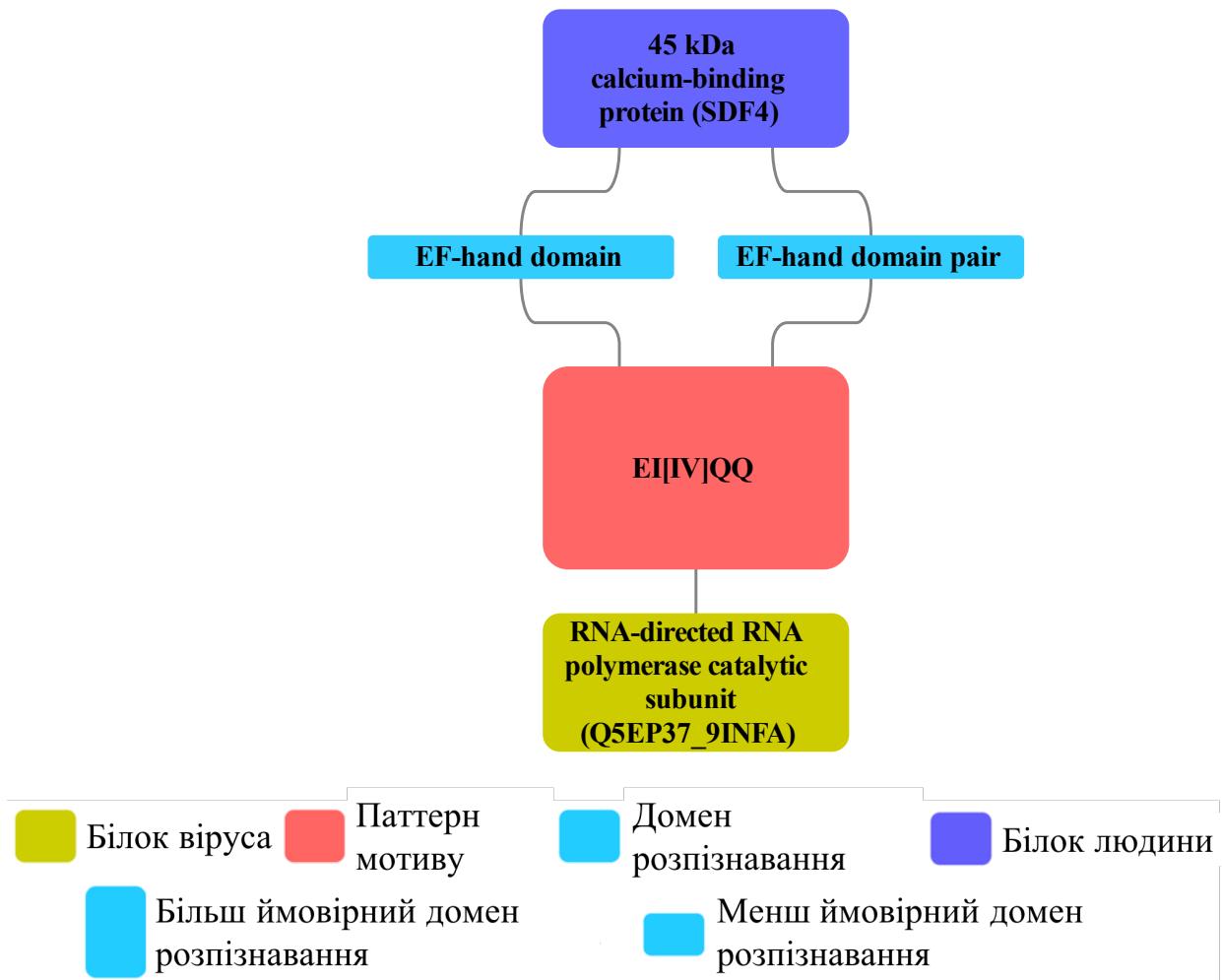


Рис 3.7.6.2. Схема мережі мотиву-кандидату EI[IV]QQ розташований в РНК-залежній РНК -полімеразі грипу А і потенційно розпізнаного доменом EF-hand білка людини SDF4.

3.7.7 Мотив-кандидат, що зв'язує BAG-домен

Ми знайшли мотив-кандидат ($L.\{0,1\}Q.LR$), який потенційно розпізнається доменом BAG у семи повтореннях у Епштейн-Барр ядерному білку антиген-лідеру (рис. 3.7.7). Цей мотив міститься в 13 інших білках, які зв'язуються з ко-шапероном людини BAG2, і також є передбаченим як посередник взаємодії зі спорідненим білком BAG3, але за низького порогу значущості. Епштейн-Барр ядерний білок антиген-лідер 5 (EBNA5), є одним з

перших білків, виявлених під час інфікування EBV, і є необхідним для трансформації В-клітин, діючи як транскрипційний ко-активуючий агент [123, 124]. BAG2 і BAG3 є ко-шапероновими білками HSP70 та HSC70 і працюють як фактор обміну нуклеотидів [125]. Таким чином, EBNA5 може посилювати діяльність ко-шаперонів HSP70 та HSC70 або впливати на проліферацію клітин або апоптоз через функції BAG2 або BAG3. Досліджуючи ще 13 білків, в яких цей мотив був передбачений, може надати більше доказів, чи BAG-домен дійсно може розпізнати мотив L.{0,1}Q.LR. Проте передбачення зв'язування вірусного пептиду (LGQLLR) з PDB структурою BAG3 миші (1uk5) або домену людського BAG1 (3fzf) з використанням PepSite 2 [126] не свідчить про наявність сильного сайту зв'язування у цьому домені.

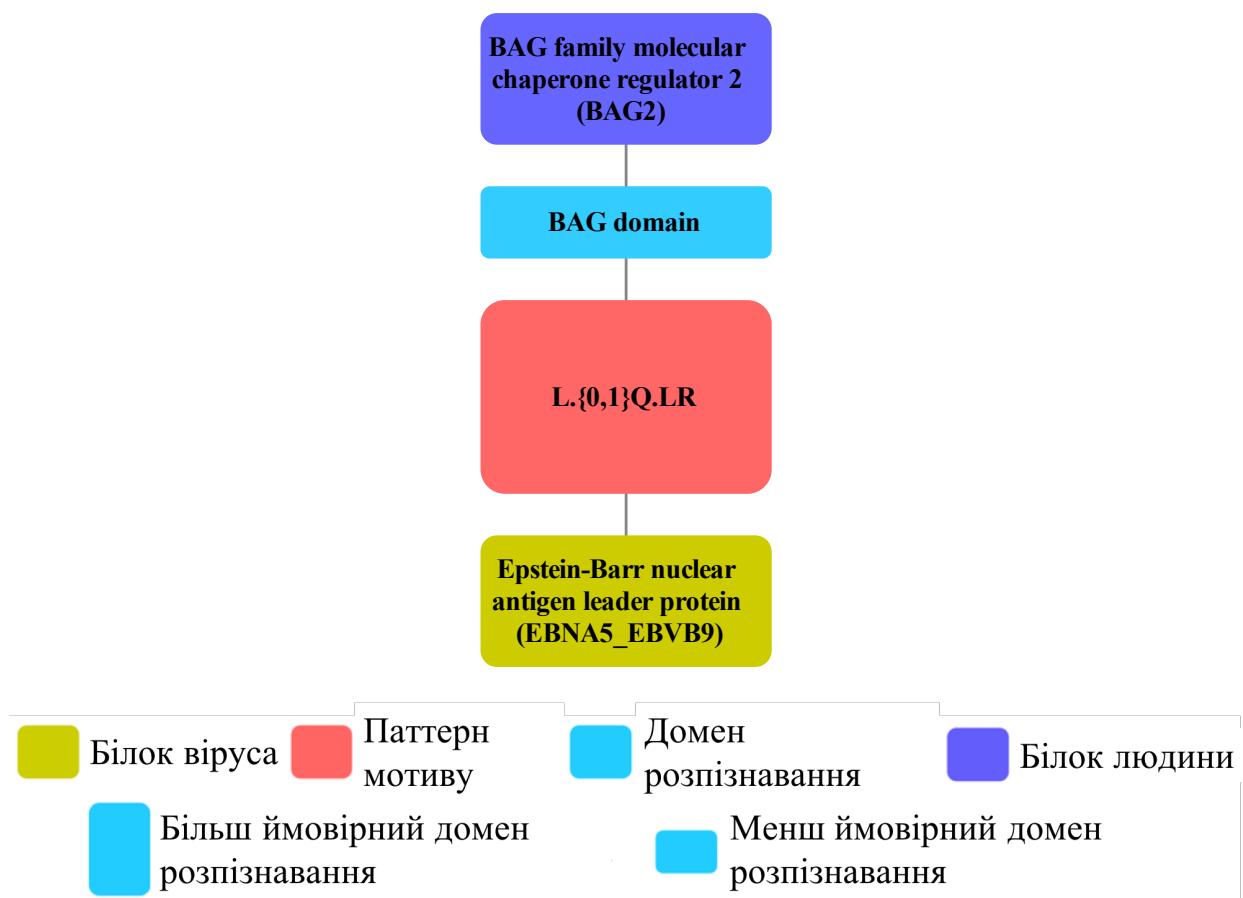


Рис 3.7.7. Схема мережі мотиву-кандидату L.{0,1}Q.LR в білку EBNA5 вірусу Епштейна-Барр потенційно розпізнається доменом BAG у ко-шапероновому білку людини BAG2.

ДОДАТКИ

Додаток А

ГРАФІК, ЩО ПОКАЗУЄ ЩІЛЬНІСТЬ РОЗПОДІЛУ ЧИСЛА ВЗАЄМОДІЙ ЛЮДСЬКИХ ЧИ ВІРУСНИХ ПРОТЕЇНІВ

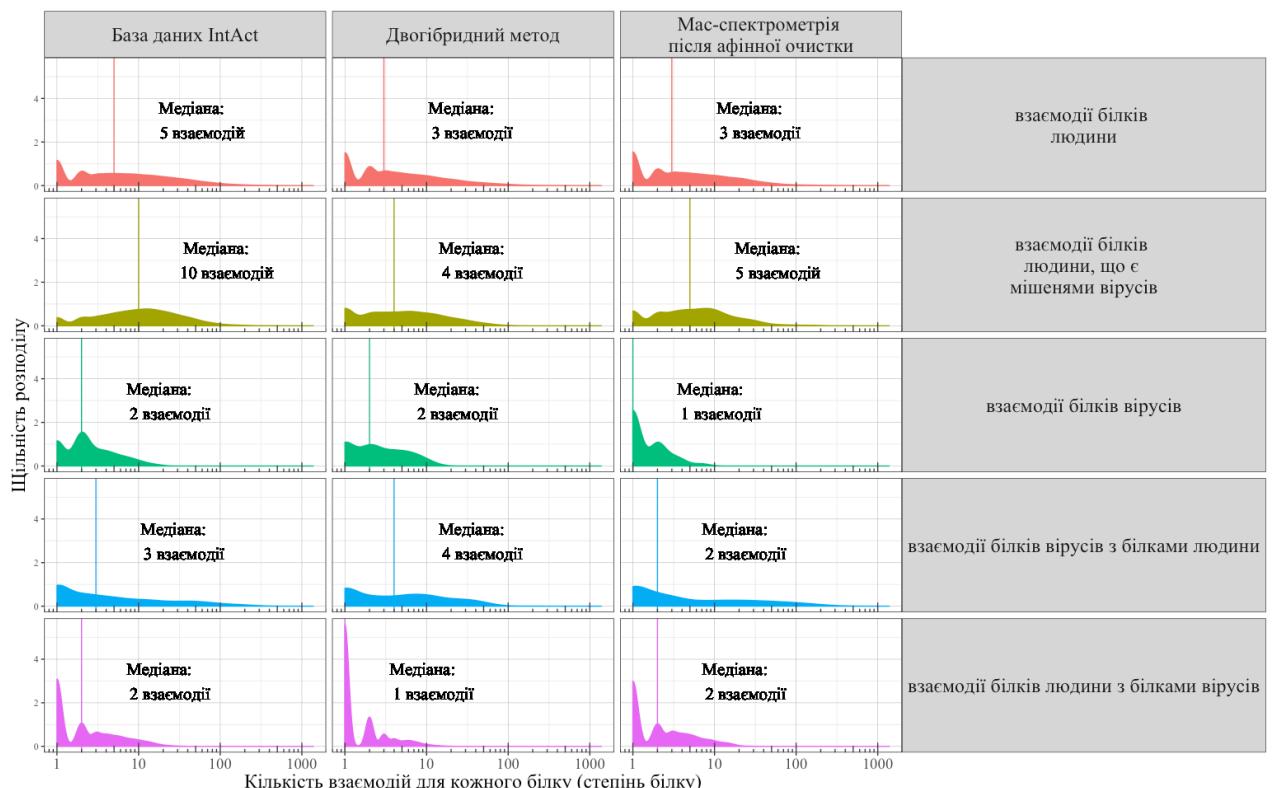


Рис 4.1. Графік, що показує щільність розподілу числа взаємодій кожного людського чи вірусного протеїну в кожній мережі, яку ми використовували для нашого аналізу. Для кожного білка вісь X показує кількість білків що з ним взаємодіють, вісь Y показує щільність розподілу. Різні мережі та різні білки (вірусні чи людські) показані в рядках. Різні методи визначення білкових взаємодій, або всі наявні дані у базі даних IntAct, зазначені у стовбцях.

ДІАГРАМА ВЕНА ЩО ПОКАЗУЄ

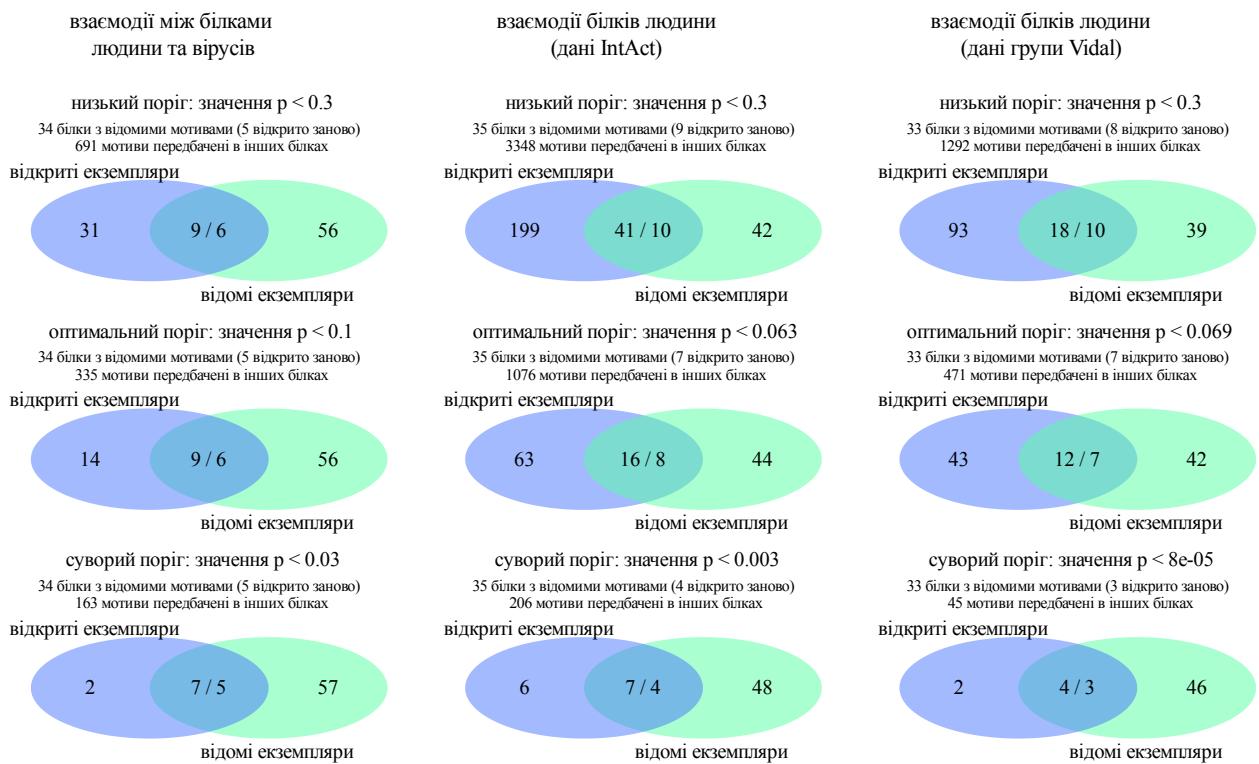


Рис 3.4.2. Діаграми Вена, що показують кількість виявлених мотивів-кандидатів та відомі мотиви ми відкрили заново. Наведено три стратегії побудови наборів даних (стовбці) та 3 пороги значень p -value (рядки). Блакитне коло показує кількість екземплярів мотивів, передбачених, але невідомих. Зелене коло показує кількість екземплярів мотивів, відомих, але не відкритих заново. Накладання показує кількість виявлених екземплярів, які відповідають відомим екземплярам (передбачені / відомі). Ці цифри відрізняються, оскільки кілька схожих передбачених екземплярів мотивів можуть співпадати з одним відомим мотивом у тому ж місці в послідовності білка (наприклад, див. розділ 3.7.2). Коректоване значення p -value - це QSLIMFinder Sig, який є вірогідністю спостереження N числа мотивів у випадковій послідовності, скоригованої для числа тестів всіх можливих мотивів. Низький поріг відображає ймовірність такого високого помилкового відкриття, як ми максимально готові допустити. При оптимальному порозі,