

В.Ф. Чехун
Л.Г. Бучинська
Н.М. Глущенко
Г.І. Соляник
С.П. Сидоренко
Ю.М. Бабич

Інститут експериментальної
патології, онкології
і радіобіології
ім. Р.Є. Кавецького
НАН України, Київ, Україна

Ключові слова: медико-біологічні, інформаційно-пошукові ресурси Internet, база даних, геном, протеом, діагностика, терапія.

БАЗА ДАНИХ «ОСОБЛИВОСТІ ФУНКЦІОНАЛЬНОГО ОНКОГЕНОМУ» ЯК ОСНОВА СУЧАСНОЇ ДІАГНОСТИКИ І ЛІКУВАННЯ ХВОРИХ ЗІ ЗЛОЯКІСНИМИ ПУХЛИНАМИ. ПОВІДОМЛЕННЯ 1

Резюме. Представлено огляд основних інформаційно-пошукових ресурсів Internet, що використовуються в галузі біології та медицини, у тому числі онкології. Обґрунтована необхідність розробки структури та створення бази даних на засаді інтеграції характеристики генів, особливостей їх експресії на рівні мРНК та білка в нормальних та злоякісно трансформованих клітинах. Створення такої бази даних сприятиме удосконаленню диференційної діагностики пухлин різного генезу та оптимізації терапії хворих зі злоякісними пухлинами.

На сьогодні системи інформаційного забезпечення із застосуванням сучасних засобів обчислювальної техніки мають на увазі не тільки електронний збір і обробку інформації, але і представлення останньої у компактному вигляді, що створює підґрунтя для оперативного її використання як науковцями, так і спеціалістами профільних установ [1, 2].

Сучасна інтенсифікація досліджень у галузі біології і медицини забезпечує отримання великого об'єму нових даних, у результаті чого виникає необхідність їх аналізу, систематизації та створення баз даних (БД), що інтегрує цю інформацію [3, 4].

Під БД слід розуміти сукупність даних великого об'єму і складної структури, комплексу засобів їх обробки, збереження і обслуговування, що забезпечує здійснення заданого набору вимог і обмежень. Ці вимоги охоплюють різні аспекти даних, програм функціонування БД і, зокрема, включають: різноманітність даних та можливість їх зв'язування в довільні структури; уникнення дублювання даних; незалежність даних; можливість захисту від випадкового або навмисного руйнування інформації і відповідно забезпечення доцільного корегування даних; забезпечення секретності; технологічність обробки даних та ін. Обробка даних повинна включати їх оновлення, підтримку, перетворення для запитів, що змінюються в дуже широких межах. Таким чином, створюється можливість динамічного розвитку інформаційних систем, що не потребує зміни в програмному забезпеченні, яке обслуговує дану систему [5].

У загальному вигляді схема типової БД зображена на рисунку.

Існуючі БД по суті є проблемно-орієнтованими автоматизованими інформаційно-пошуковими системами. Накопичені в профільних БД дані основані на конкретних знаннях, а програми, що призначені для вирішення деякої сукупності схожих по змісту задач, дозволяють для їх інформаційного забезпе-

чення створювати комплексну близькотематичну БД.

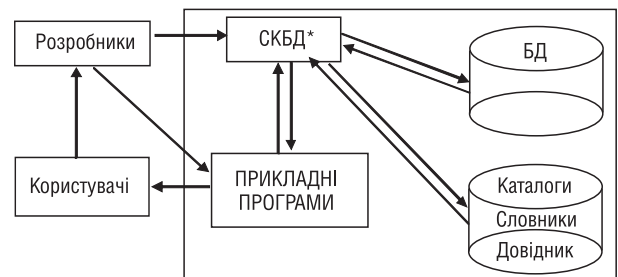


Рисунок. Схема типової БД

*СКБД — система керування БД.

На сьогодні основні інформаційні ресурси в біологічній та медичній галузях, у тому числі онкології, які представлені в Internet можна розподілити на 4 умовних категорії:

1. *Генетичні та протеомні (білкові):* **National Center for Biotechnology Information (NCBI)** [6, 7] — пошукова система, що надає доступ до БД, таких як **GenBank** — БД послідовностей ДНК, РНК та білків, що має літературні посилання щодо їх біологічної інформації [8, 9]; **BLAST (The Basic Local Alignment Tool)** — [10, 11]; також БД, які можна знайти через систему NCBI Entrez [12]; **dbGaP** — БД генотипу та фенотипу [13, 14]; **Protein Clusters database** — БД зв'язаних послідовностей (кластерів) білків [15]; **The Cancer Genom Anatomy Project (CGAP)** — електронний ресурс, що надає доступ до БД, які містять інформацію про гени, хромосомні аберації і визначення генного профілю нормальних та пухлинних клітин [16, 17]; **Cancer Chromosomes database** — БД, що надає дані цитогенетичних досліджень про аберації, що виникають у хромосомах при раку [18, 19] та інші, що стосуються інформації про нуклеотидні послідовності генів, структури білків [7]; **The Protein Data Bank (PDB)** — банк даних тривимірних структур біологічних макромолекул, визначених експериментальними методами [20, 21]; **The Universal Protein Resource (UniProt)** — БД

послідовностей білків [22, 23]; **InterPro** — інтегрований ресурс, який надає доступ до інформації щодо сімейств, областей, сайтів білків [24, 25]; **BioLit** — пошукова система біологічної літератури і надає доступ до БД щодо інформації про білки [26, 27]; **Enzyme** — номенклатурна БД ферментів [28]; **The Restriction Enzyme Database (REBASE)** — БД надає інформацію про рестрикційні ферменти [29, 30]; **Database of Protein Disorder (DisProt)** — БД, яка забезпечує інформацію про 3D структуру білків [31, 32]; **Familial Cancer Database (FaCD)** — сімейна БД, яка допомагає у вирішенні питань генетичної диференційної діагностики [33]; **IARC (The International Agency for Research on Cancer) TP53 Mutation Database** — БД TP53 соматичних мутацій, що виникають у новоутвореннях людини і клітинних лініях [34, 35]; **Pancreatic Expression database** — БД, що містить інформацію про гени, експресія яких змінюється у пухлинних клітинах підшлункової залози [36, 37]; **The Tumor Gene Family of Databases** — БД містить інформацію про гени, мутації яких викликають рак, і включає понад 2600 фактів відносно 300 генів [38]; **HGNC (HUGO Gene Nomenclature Committee)** — містить скорочені та повні назви 24 000 генів людини [39, 40]; **Transcription Regulatory Regions Database (TRRD)** — БД містить опис регулювання транскрипції гена [41, 42]; **Breast Cancer Mutation Database** — БД надає інформацію відносно мутацій та поліморфізму у генах при раку молочної залози [43].

2. *Клінічні дослідження: PDQ (Physician Data Query) database* — різнобічна БД Національного інституту раку США містить інформацію по профілактиці, ранній діагностиці злоякісних пухлин та лікуванню і догляду за хворими. Крім того, БД містить довідники для лікарів, спеціалістів по генетиці, інформацію щодо організацій, що спеціалізуються у лікуванні раку [44, 45]; **National Cancer Data Base (NCDB)** — ресурс клінічних спостережень раку в США [46]; **CenterWatch (Clinical Trials Listing Service)** — довідковий ресурс призначений для лікарів і пацієнтів, у якому представлена медична інформація спостережень клінічних досліджень та випробувань препаратів, у тому числі і протипухлинних [47]; **eCancerTrials** — БД надає сучасну інформацію щодо клінічних випробувань препаратів в онкології, які можуть бути використані в оцінці нових методів лікування хворих [48], а також **DynaMed (Dynamic Medical Information Systems)** — медична інформаційна БД надає короткі клінічні звіти і має список гіперпосилань на медичні журнали [49, 50].

3. *Бібліографічні або бібліотечні: PubMed* — літературний пошук резюме та повних текстів статей у найбільшій світовій бібліографічній БД публікацій по біології та медицині **MEDLINE (MEDlars onLINE, MEDLARS -MEDical Literature Analysis and Retrieval System)** [51]; **PubMed Central (PMC)** — надає доступ до електронного архіву біомедичної літератури [7, 52]; **Elsevier Bibliographic Databases (EBD)** — надає доступ до міжнародних медико-біологічних

журналів [53]; **Amedeo** (медико-літературний гід) — надає можливість пошуку рефератів опублікованих робіт за бібліографічними списками стосовно нових наукових публікацій і коротких оглядів медичної літератури [54]; **Hardin Library of Electronic Journals (Hardin MD, The University of Iowa)** — у БД містяться усі доступні в мережі Internet журнали по профілю онкології з можливістю доступу до повних текстів статей [55]; **Пантелеймон** — БД публікацій медико-біологічного, хіміко-фармацевтичного та хімічного характеру, надрукованих у періодичних наукових і науково-практичних виданнях України та Російської Федерації [56, 57]; **BioMed Central** — інформаційний ресурс надає доступ до наукових журналів [58], **BioMed Central Databases** — надає доступ до БД [59]; **Cancer Management Guidelines (British Columbia Cancer Agency (BCCA))** — online-підручник, який включає інформацію про сучасні методи діагностики та підходи до лікування раку [60]; **Library of Congress Online Catalog** — інформаційний ресурс бібліотеки конгресу США (Washington), який надає доступ до пошуку наукових статей, у тому числі і в медико-біологічних журналах [61]; **The Merck Manuals Online Medical Library** — інформаційний ресурс по діагностиці і терапії, що включає розділ з гематології і онкології [62]; **Spencer S. Eccles Health Science Library (University of Utah)** — наукова бібліотека здоров'я, що є освітницькою і призначена надавати інформацію в галузі клінічних досліджень медичним працівникам, науковцям та студентам [63]; **HighWire Press (HighWire Library of the Sciences and Medicine, Stanford University Libraries)** — портал, що дозволяє доступ до великої кількості опублікованих робіт [64].

4. *Інші електронні ресурси: @Cancer* — пошуковий сайт медичної інформації [65]; **CancerBACUP** — інформаційний ресурс по онкології у Великобританії, містить рекомендації по лікуванню онкологічних хворих [66]; **MEDICINE OnLine** — інформаційний медичний ресурс по онкології [67]; **OncoLink (Abramson Cancer Center of the University of Pennsylvania)** — сервер Онкологічного центру університету Пенсильванії, містить широке коло питань з експериментальної та клінічної онкології [68]; **Database of Cancer Research in Australia (CARA)** — наведені результати досліджень новоутворень в Австралії [69]; **Carcinogenic Potency Database (CPDB)** — представлені результати довготривалих досліджень новоутворень у тварин [70].

Таким чином, існує багато електронних інформаційних ресурсів, що включають дані фундаментальних досліджень у галузях біології і медицини та надають інформацію щодо методів діагностики і лікування різних захворювань, і в тому числі онкологічних. Але інформація, що є в різноманітних банках і БД представлена в різних форматах та різним ступенем деталізації, що обґрунтовує доцільність створення інтегрованої БД, спрямованої на розширення можливостей диференційної та уточнюючої діагностики передпухлинних і пухлинних процесів

та оцінці ефективності і коректності лікування хворих із пухлинами різного генезу.

В Інституті експериментальної патології, онкології і радіобіології ім. Р.Є. Кавецького НАН України розроблена БД, що відповідає наступним критеріям:

- Висвітлює фундаментальні аспекти та клінічні проблеми канцерогенезу.
- Проблемно-орієнтована автоматизована інформаційно-пошукова система фактографічного типу, яка дає можливість знайти не тільки документи чи довідки, а безпосередньо факти, які містяться у цих документах. Середовищем програмування системи керування БД є Borland Delphi v.7.0. Система орієнтована на роботу як у автономній (мережевій), так і локальній версії без обмеження кількості користувачів за умови, що платформою є Windows 95.../NT/XP.

- Сервіс дозволяє розширювати і поповнювати БД новою інформацією і розрахований на наукових працівників, лікарів, аспірантів та студентів, що не є фахівцями в галузі інформаційних технологій.

На цьому етапі продовжується робота по наповненню БД та створенню додаткових модулів. Принципи системи керування та структури, а також основні властивості створеної БД буде представлено у наступному повідомленні.

ПОДЯКА

Автори щиро вдячні д-ру мед. наук, проф. Н.М. Бережній; д-ру біол. наук, проф. М.Ф. Гамалії; д-ру мед. наук, проф. Д.Ф. Глузману; д-ру мед. наук, проф. Г.І. Кулик; д-ру мед. наук, проф. В.Г. Ніколаєву; д-ру мед. наук, проф. С.П. Осинському; д-ру біол. наук, проф. П.В. Погрібному; д-ру мед. наук, проф. Є.П. Сидорику; д-ру мед. наук, проф. В.О. Шляховенку; д-ру біол. наук М.О. Дружині; д-ру біол. наук С.П. Залеток; д-ру біол. наук Ю.Й. Кудрявцю; д-ру мед. наук Г.П. Потебні; канд. біол. наук В.М. Михайленку; канд. біол. наук О.О. Фільченкову за плідні дискусії, цінні поради та інформаційне наповнення БД. Робота виконана за підтримки гранта НАН України «Особливості функціонування онкогену» П-01-02, № держреєстрації 0102U003228.

ЛІТЕРАТУРА

1. Кальниш ВВ. Роль информационных технологий в интеллектуализации здравоохранения. *Клин информ телемедицина* 2004; **1**: 28–34.
2. Brazhnik O, Jones JF. Anatomy of data integration. *J Biomed Inform* 2007; **40** (3): 252–69.
3. Горобец СВ, Горобец ОЮ, Дереча ДО. Биоинформатика как основной инструмент нанобиотехнологии и наномедицины. *Клин информ телемедицина* 2008; **4** (5): 41–9.
4. Гельфанд МС. Апология биоинформатики 2005; **50** (4): 752–66.
5. Система автоматизации библиотек ИРБИС. АРМ «Каталогизатор». Руководство пользователя. Москва: ГПНТБ России, 2000. 76 с.
6. <http://www.ncbi.nlm.nih.gov>

7. Wheeler DL, Barrett T, Benson DA, *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2008; **36** (27): 13–21.
8. <http://www.ncbi.nlm.nih.gov/Genbank/>
9. Benson DA, Karsch-Mizrachi I, Lipman DJ, *et al.* GenBank. *Nucleic Acids Res* 2008; **36**: 25–30.
10. <http://ncbi.nlm.nih.gov/blast>
11. Johnson M, Zaretskaya I, Raytselis Y, *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008, **36** (2): 5–9.
12. <http://www.ncbi.nlm.nih.gov/sites/gquery?itool=toolbar>
13. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>
14. Mailman MD, Feolo M, Jin Y, *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007; **39** (10): 1181–86.
15. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=proteinclusters>
16. <http://cgap.nci.nih.gov>
17. Riggins GJ, Strausberg RL. Genome and genetic resources from the Cancer Genome Anatomy Project. *Hum Mol Genet* 2001; **10** (7): 663–7.
18. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=cancerchromosomes>
19. Knutsen T, Gobu V, Knaus R, *et al.* The Interactive Online SKY/M-FISH & CGH Database and the Entrez Cancer Chromosomes Search Database: Linkage of Chromosomal Aberrations with the Genome Sequence. *Gen Chromosom Cancer* 2005; **44** (1): 52–64.
20. <http://www.wwpdb.org/index.html>
21. Dutta S, Burkhardt K, Swaminathan GJ, *et al.* Data deposition and annotation at the worldwide protein data bank. *Methods Mol Biol* 2008; **426**: 81–101.
22. <http://www.uniprot.org/>
23. UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res* 2008; **36**: 190–5.
24. <http://www.ebi.ac.uk/interpro>
25. Mulder NJ, Apweiler R, Attwood TK, *et al.* New developments in the InterPro database. *Nucleic Acids Res* 2007; **35**: 224–8.
26. <http://biol.its.ucsd.edu>
27. Fink JL, Kusch S, Williams PR, Bourne PE. *Nucleic Acids Res* 2008; **36**: 385–9.
28. <http://au.expasy.org/enzyme/>
29. <http://rebase.neb.com/rebase/rebase.html>
30. Roberts RJ, Vincze T, Postai J, Macelis D. REBASE — enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* 2007; **35**: 269–70.
31. <http://www.disprot.org/>
32. Sickmeier M, Hamilton JA, LeGall T, *et al.* DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* 2007; **35**: 786–93.
33. <http://www.familialcancerdatabase.nl/>
34. <http://www-p53.iarc.fr>
35. Olivier M, Eeles R, Hollstein M, *et al.* The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat* 2002; **19** (6): 607–14.
36. <http://www.pancreasexpression.org/>
37. Chelala C, Hahn SA, Whiteman HJ, *et al.* Pancreatic Expression database: a generic model for the organization, integration and mining of complex cancer datasets. *BMC Genomics* 2007; **28** (8): 1471–2164.
38. <http://www.tumor-gene.org>
39. <http://www.genenames.org/>
40. Bruford EA, Lush MJ, Wright MW, *et al.* NC Database in 2008: a resource for the human genome. *Nucleic Acids Res* 2008; **36**: 445–8.
41. <http://www.mgs.bionet.nsc.ru/mgs/gnw/trrd/>
42. Kolchanov NA, Ignatieva EY, Ananko EA, *et al.* Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucleic Acids Res* 2002; **30** (1): 312–7.
43. <http://www.genome.gov/10001504>

44. <http://www.cancer.gov/cancertopics/pdq>
45. **Fare C, Ugolini D.** The PDQ (Physician Data Query), the cancer database, in oncological clinical practice. *Cancer Treat Rev* 1991; **18** (2): 137–43.
46. <http://www.facs.org/cancer/ncdb>
47. <http://www.centerwatch.com>
48. <http://www.ecancertrials.com>
49. <http://www.ebscohost.com/dynamed/default.php>
50. **Barton H.** *DynaMed. J Med Libr Associat* 2005; **93** (3): 412–4.
51. <http://www.ncbi.nlm.nih.gov/pubmed/>
52. <http://www.pubmedcentral.nih.gov>
53. http://www.elsevier.com/wps/find/bibliographic_browse.cws_home
54. <http://amedeo.com/index.htm>
55. <http://www.lib.uiowa.edu/hardin/md/ej.html>
56. <http://www.panteleimon.org>
57. **Kumar A, Wright J, Adams CE.** Searching a biomedical bibliographic database from the Ukraine: the Panteleimon database. *Health Info Libr J* 2005; **22** (3): 223–7.
58. <http://www.biomedcentral.com/browse/journals>
59. <http://databases.biomedcentral.com/>
60. <http://www.bccancer.bc.ca/HPI/CancerManagementGuidelines/default.htm>
61. <http://catalog.loc.gov/>
62. <http://www.merck.com/mmpe/index.html>
63. <http://library.med.utah.edu/index.php>
64. <http://highwire.stanford.edu>
65. <http://www.mymedline.com/cancer/>
66. <http://www.cancerbackup.org.uk/Home>
67. <http://www.meds.com>
68. <http://oncolink.upenn.edu>
69. <http://www.ludwig.edu.au>
70. <http://potency.berkeley.edu/cpdb.html>

DATABASE «FEATURES OF FUNCTIONAL ONCOGENOMICS» AS THE FOUNDATION OF MODERN DIAGNOSTICS AND TREATMENT OF PATIENTS WITH CANCER (MALIGNANT TUMORS). REPORT 1

V.F. Chekhun, L.G. Buchynska, N.M. Glushchenko, G.I. Solyanik, S.P. Sidorenko, Yu.M. Babich

Summary. *The report includes the review of basic Internet Life science informatics resources, used in the bio-medical field, especially in oncology. The database formation and development of its structure is based on gene features, their expression on the level of mRNA and protein in normal and malignant cells. This database will promote the improvement of tumor differential diagnostics and optimize therapy of patients with malignant tumors.*

Key Words: Internet, bio-medical informatics resources, database, genome, proteome, diagnostics, therapy.

Адреса для листування:

Чехун В.Ф.

03022, Київ, вул. Васильківська, 45

Інститут експериментальної патології,
онкології і радіобіології ім. Р.Є. Кавецького
НАН України