

ОГЛЯДИ

УДК 577.218

СИСТЕМНА БІОЛОГІЯ І ПРОЕКТ «ENCODE»

М. Ю. БОЛЕНСЬКА

Інститут молекулярної біології і генетики НАН України, Київ;
e-mail: m.obolenska@gmail.com

В огляді подається початкове уявлення щодо витоків системної біології, висвітлюються передумови для її виокремлення в самостійну науку, її основні засади, особливості методології та здобутки в з'ясуванні функціональних елементів геному людини і засобів, які забезпечують їх узгоджену і диференційовану дію. Наводиться коротка характеристика функціональних елементів геному: протеїнкодуєчі послідовності і ті, які кодують РНК, що не транскрибується, ділянки гіперчутливості до ДНКаз I і метильовані CpG-острівці, модифіковані гістони і специфічна тривимірна структура хроматину. Наводяться відомості щодо топології мережі транскрипційних факторів, її основних мотивів, їх ієрархії, комбінації і коасоціації, особливостей їх зв'язування з алель-специфічними послідовностями.

Ключові слова: системна біологія, функціональні елементи геному, мережа генної регуляції.

Systems biology is a success if it is widely accepted that there is nothing more practical than a good theory
Wolkenhauer O., Mesarovic M., 2010

Початок ХХІ ст. позначився появою нової галузі в науці – системної біології. Ця дисципліна виникла на засадах біології і теорії систем. Її метою стало всебічне дослідження живих систем і їх математичне моделювання як засіб для глибшого дослідження системи через симулювання моделі та її постійне удосконалення з максимальним наближенням до природного прототипу. Основний науковий підхід нової дисципліни – системний, що передбачає визнання системи як цілісної одиниці, яка об'єднує взаємопов'язані елементи системи і має властивості, відмінні від сукупності властивостей її компонентів (нова властивість, англ. emergence); водночас вона є підсистемою систем вищого порядку; система має модульну структуру (модульність, англ. modularity); є динамічним утворенням, яке змінюється в процесі її утворення і в часі впродовж її функціонування; є надійною, або безпомилковою (надійність, англ. robustness), тобто забезпечує підтримання фенотипу в умовах генетичних варіацій і постійної

мінливості довкілля. В арсеналі системної біології найсучасніші високотехнологічні методи молекулярної біології, математичні, фізичні та інформативні інструменти. Її плани далекоглядні – від створення віртуальних органів до створення віртуальної людини. Системною біологією передбачаються революційні процеси в медицині – перехід від лікування середньостатистичного хворого до лікування конкретної особи з особливостями її гено- і фенотипу. Масштабність досліджень в цій новій галузі сприяє об'єднанню наукової спільноти світу заради загальнолюдських цінностей.

У цьому огляді передбачається висвітлити передумови для виокремлення системної біології в самостійну науку, її основні засади, особливості методології та здобутки у з'ясуванні функціональних елементів геному людини і засобів, які забезпечують їх узгоджену і диференційовану дію.

Трохи історії

Впродовж ХХ ст., особливо його другої половини, поступово створювалися передумови для виокремлення системної біології в самостійну науку. Її засновником, а також одним із авторів теорії систем вважається Людвіг фон

Берталанффі, який у 1950 р. опублікував працю «Загальна теорія систем у фізиці і біології» [1]. У 1952 р. два нейрофізіологи, сер Алан Лойд Ходжкін і сер Ендрю Філдінг Хакслі, майбутні лауреати Нобелівської премії 1963 р., створюють математичну модель розповсюдження біоелектричного потенціалу вздовж аксона нервової клітини і передбачають існування іонних каналів [2]. Конрад Уоддингтон, генетик, палеонтолог, ембріолог, людина ренесансного світогляду, пророчо описує функціонування генів і клітин у вигляді динамічних мереж, привертає увагу до еволюції їх створення і пропонує описувати мережі системою диференціальних рівнянь [3]. Деніс Нобл, британський фізіолог, на основі моделі Ходжкіна і Хакслі створює модель водія серцевого ритму [4]. У 2001 р. за його активною участю започатковано проект «Фізіом» (IUPS Physiome Project) [<http://physiomeproject.org>], мета якого з'ясувати як пов'язане функціонування геному із функціонуванням всіх систем організму, щоб у віддаленішій перспективі створити віртуальну людину. Теоретик у галузі систем Михайло Месарович у 1966 р. доповідає на міжнародному конгресі у Клівленді (Огайо, США) роботу «Systems Theory and Biology» і протиставляє довгими роками існуючий в наукових підходах редукціонізм новому холізму [5]. Розробляється теорія контролю метаболізму (управління метаболізмом), яка використовує математичний апарат для дослідження регуляторних властивостей метаболічних систем (на противагу «вузькому» місцю системи, «bottleneck») і здатності їх координувати змінювати потоки і концентрацію метаболітів у непостійних умовах середовища заради підтримки стаціонарного стану з мінімальним відхиленнями ключових метаболітів від норми. Коефіцієнтами управління характеризується внесок окремих ензимів і зовнішніх метаболітів в управління метаболічними потоками і концентрацією метаболітів [6, 7]. Стюарт Кауфман досліджує структуру мереж генної регуляції, еволюцію їх утворення і оптимізації [8]. Але поступово інтерес до теоретичної біології, зокрема серед біологів, спадає через невідповідність її результатів перебільшеним очікуванням.

На зміну приходить бум у методології молекулярної біології: принципово нові методи (рестрикція і клонування, секвенування, полімеразна ланцюгова реакція) відкривають

шлях до прочитання геному людини. Проект «Геном людини» (Human Genome Project, http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) був офіційно започаткований в жовтні 1990 р., підтриманий Національним інститутом здоров'я і Міністерством енергетики США. Вчені щонайменше з 18 країн приєдналися до виконання проекту. Через 50 років після відкриття Джеймсом Уотсоном і Френсісом Кріком дволанцюгової структури ДНК у квітні 2003 р. міжнародний консорціум із секвенування геному людини оголосив про завершення проекту [9]. Численні заходи відбувалися в Америці і в світі з нагоди 10-ї річниці закінчення проекту «Геном людини». Наслідки виконання проекту торкнулися всіх сфер життя людства. У науковій сфері виконання проекту сприяло стрімкому розвитку новітніх технологій, накопиченню величезної кількості даних, які потребували систематизації, математичної обробки і інформаційних технологій. Таким чином, взаємонеобхідність спеціалістів різного профілю стала безперечним фактом, а взаємопроникнення різних дисциплін для досягнення спільної мети реалізувалось у створенні системної біології як окремої науки.

Завдання, що постали після завершення проекту «Геном людини», та їх розв'язання

Отже, на завершальному етапі виконання проекту «Геном людини» 99% геному було прочитано з високою точністю – 1 помилка на 100 000 нуклеотидів; з'ясовано, що геном кодує 20–25 тис. протеїнів і сформульовано найближчі завдання для подальшого вирішення. Вони включали: систематичне визначення поліморфізму в геномі людини і його асоціації з хворобами; систематичну ідентифікацію всіх функціональних елементів у геномі – протеїнокодуючих і протеїннекодуючих генів, регуляторних елементів, які визначають транскрипцію, структуру і функціонування хроматину; і, нарешті, систематичне визначення регуляторних модулів (окремих функціонально повноцінних регуляторів) експресії генів [9].

Мовою теоретичної системної біології завдання формулювалися в поняттях: реалізація – створення моделі, яка б характеризувала вхідні стимули і відповіді на них; ідентифікація системи – визначення параметрів моделі на підставі

експериментальних даних і даних зі симуляції; аналізу контролю – визначення характеру взаємодії між компонентами системи і передбачення наслідків змін, внесених у певні шляхи і параметри системи [10].

Виконанню першого завдання був присвячений міжнародний проект із картування гаплотипів людини (Haplotype Map, проект «НарМар»), який мав на меті генотипувати загальні SNPs (однонуклеотидний поліморфізм, single nucleotide polymorphism) у представників різних людських популяцій. Проект офіційно розпочато на симпозиумі 27–29 жовтня 2002 р., а результати його третього етапу були опубліковані навесні 2009 р. і містили дані про загальні і рідкі алелі і поліморфізм за SNP і числом копій [11]. Дещо пізніше, в січні 2008 р., з урахуванням досвіду, який стрімко накопичувався в перші роки нового сторіччя, було розпочато проект «1000 геномів». Метою проекту було створити повний і детальний каталог всіх генетичних варіацій в геномі людини. У 2010 р. скінчився пілотний етап проекту, результати було опубліковано у 2010 р. [12] і доповнені у 2012 р. ще 92 геномами [13]. Загальне число проаналізованих індивідуальних геномів планується найближчим часом збільшити до 2500. Інший проект – «Всегеномне дослідження асоціацій» (Genome-wide association study, GWAS) – передбачає порівняння загальних генетичних варіантів у практично здорових людей і у хворих на розповсюджені хвороби з метою з'ясувати чи існує асоціація між хворобою і певними варіантами в структурі геному [14].

Щодо вирішення другого і третього завдання, то в 2002 р. Національний інститут дослідження геному людини (NHGRI, Бетесда, США) запропонував створити інтерактивний консорціум для проведення пілотного проекту з метою з'ясування придатності існуючих обчислювальних і експериментальних методів, розробки і перевірки нових технологій для анотації функціональних елементів геному.

Проект «ENCODE» (Encyclopedia of DNA elements)

Пропозиція проекту «Encode» була ухвалена на робочій нараді Comprehensive Extraction of Biological Information from Genomic Sequence 23–24 липня 2002 р., а 7 березня 2003 р. п'ятирічний пілотний проект було офіційно роз-

почато. Інформація про пілотний проект і його результати доступні в репозиторії (<http://genome.ucsc.edu/ENCODE>), на сайті Національного інституту дослідження геному людини (<http://www.genome.gov/26525202>) й інших порталах, а також в публікаціях журналів «Nature» [2007; 447(7146)] і «Genome Research» [2007, 17(6)]. Пілотний проект коштував приблизно 55 млн. доларів.

У пілотному проекті 35 груп науковців провели дослідження 29 998 тисяч основ або близько 1% геному на клітинах двох типів – HeLa S3 і GM06990, використовуючи великий обсяг експериментальних і обчислювальних методик, які найчастіше доповнювали одна одну, створюючи підґрунтя для високої надійності результатів, які порівнювали з існуючими на той час даними інших проектів (наприклад, «НарМар Project», результатами еволюційних досліджень). Проведена масштабна робота мала свої наслідки. Дуже коротко, з огляду на повноту інформації, отриману в проекті другого етапу, можна підсумувати результати першого. У межах дослідженого геному переважна його частина транскрибується, і найбільша частка первинних транскриптів зчитана із протеїннекодуючих послідовностей. Виявлено неочікувано велику кількість (приблизно на порядок вище за раніше відому) стартових сайтів транскрипції і зв'язок їхньої активності із властивостями хроматину і модифікацією гістонів; взаємозв'язок між структурою хроматину і сайтами реплікації. За доступністю хроматину («open» vs. «close») і модифікацією гістонів можна прогнозувати наявність і активність стартової точки транскрипції. Певному часовому проміжку процесу реплікації ДНК відповідає певна структура хроматину. Біля 5% послідовностей геному знаходяться під еволюційним тиском, і з них 60% – це функціональні елементи. В той самий час не всі функціональні елементи знаходяться під тиском еволюції [15].

Одержані результати стали підґрунтям для амбітнішого проекту – складання енциклопедії функціональних елементів для всього геному людини, який був започаткований у вересні 2007 р., а закінчився у вересні 2012 р. звітом науковій спільноті у вигляді 30 статей в «Nature», «Genome Research» і «Genome Biology» [<http://www.nature.com/encode/#/threads>; <http://www.encodeproject.org/ENCODE/pubs.html>], огляду

[16] і витраченими коштами близько 185 млн. доларів. На другому етапі проекту спектр досліджень було значно розширено – експерименти проводили на клітинах 147 типів. Серед них: первинні диференційовані клітини, первинні імуорталізовані, пухлинні лінії, мульти- і плюрипотентні клітини-попередники. Досліджували 119 транскрипційних факторів (ТФ) із 1800 відомих на сьогодні, 13 модифікацій гістонів і ДНК із 60 відомих [17]. Метою проекту було визначити тепер уже в межах всього геному функціональні елементи ДНК і механізми регуляції функціонування геному в різних клітинах і за різних умов. Виконання проекту є показовим прикладом ролі системної біології в сучасних дослідженнях, адже реалізація проекту була би неможливою без принципово нового мислення й інтеграції методів математичної статистики, інформатики, моделювання і молекулярної біології.

Внаслідок виконання проекту створені нові і удосконалені існуючі експериментальні методи, налаштовані на одержання кількісних даних з найширшим використанням різних видів секвенування; розширено арсенал біоінформативних підходів із залученням моделювання і розроблено системи для зберігання інформації у вільному доступі; отримано величезний обсяг даних, які систематизовано з урахуванням результатів інших великих проектів типу «НарМар», «1000 геномів», «GWAS» тощо. В огляді надаються лише деякі висновки проекту стосовно функціональних елементів геному, клітинспецифічної регуляції і топології транскрипційної регуляторної мережі. За повною інформацією щодо результатів проекту «Encode» читач переадресується на відповідні сайти (див. вище).

Функціональні елементи геному та їх взаємодія

Функціональним елементом геному визначають його дискретний сегмент, який або кодує певний продукт – протеїн чи РНК, або містить стартову точку транскрипції, або несе певний відтворюваний розпізнавальний знак – є місцем метилування ДНК, зв'язування ТФ, гіперчутливості до ДНКаз 1 або належить специфічній структурі хроматину чи гістонів [17].

За вищенаведеним визначенням у 80,4% геному людини знаходяться функціональні

елементи, більшість з яких розташована за межами протеїнокодуючих послідовностей. У цих послідовностях містяться сайти, з якими зв'язуються ТФ і включають або виключають гени; послідовності, які зчитуються і транскрибуються у вигляді РНК, що не транслюється; контролюють транскрипцію найближчих генів (промотори, яких більше 70 тис.) або впливають на активність генів, знаходячись на великій відстані від них (енхансери, біля 400 тис); регулюють транскрипцію через метилування ДНК і численні модифікації гістонів; впливають на упаковку ДНК. Тобто виконують якусь функцію і не є «непотребом» (junk) [17].

Геном людини кодує 20 687 протеїнів, відповідні екзони складають біля 1,2% всіх послідовностей ДНК. Протеїни можуть варіювати через сплайсинг, що зустрічається в середньому з частотою 6,3 альтернативно сплайсованих транскриптів або 3,9 варіантів протеїнокодуючих транскриптів на локус. Проектом «Encode» анотовано із розрахунку на одну клітину 9640 локусів довгих некодуючих РНК (long noncoding RNA, lncRNA) і 8801 малих (< 200 н.) РНК, а також 11 224 псевдогени, з яких більшість дійсно «мертві», але деякі виявляють ознаки функціональної активності і транскрибуються [18].

У геномі виявлено 62 403 стартові точки транскрипції, з яких 44% розташовані в межах 100 п.о. від 5'-кінця транскриптів, а інші – в межах екзонів і 3'-нетрансльованих ділянок.

Найбільшу частину основ геному (62%) представлено в складі довгих (> 200 нукл.) РНК, які включають інформаційні РНК і інформаційноподібні lncRNA. ДНК-матриці останніх знаходяться в межах протеїнокодуючих генів, їхніх інтронів, екзонів, 3'-нетрансльованих ділянок, промоторів і в міжгенному просторі (рис. 1). lncRNA транскрибуються за участю РНК-полімерази II у двох напрямках (прямий і зворотний) [19]. На підставі цих даних звичне уявлення про ген як про лінійну структуру певного напрямку змінилось [20]. Довгі некодуючі РНК піддаються сплайсингу, поліаденилюються, але не транслюються. Їхні функції і механізм дії знаходяться на початковій стадії дослідження. Вважається, що вони беруть участь у регуляції експресії генів, у структурній організації ядра і компартменталізації. Порушення з боку довгих некодуючих РНК асоційовані з різними хворобами [21].

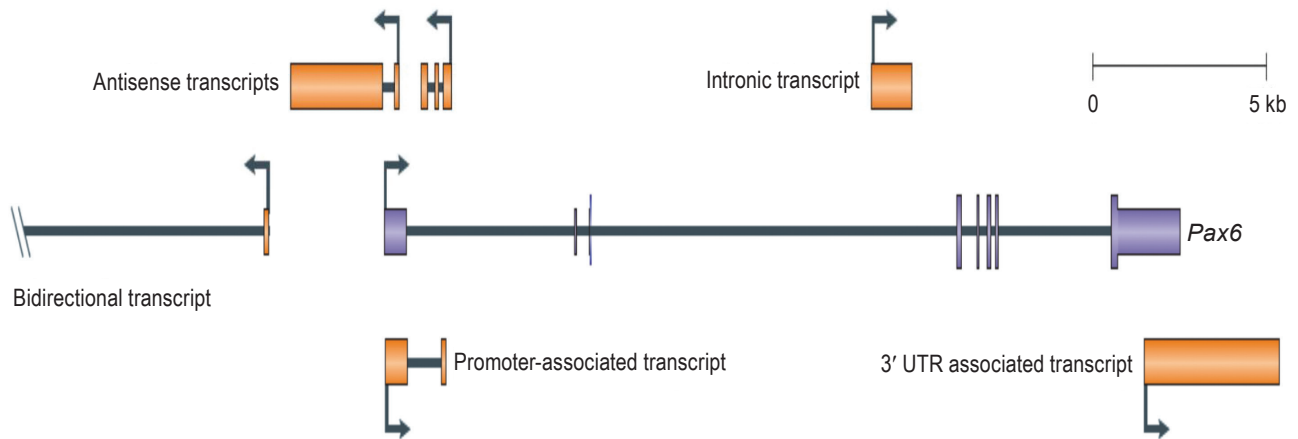


Рис. 1. Взаєморозташування транскриптів, які кодують і не кодують протеїни [19]. Схема ілюструє як транскрипція довгої некодуючої РНК (жовтогарячий колір) перемежується з транскрипцією гена *Pax6* (фіолетовий колір)

Проект надає першу карту розташування 2 890 742 гіперчутливих до ДНКазі 1 ділянок ДНК (DNA hypersensitive site, DHS), найчисленніших функціональних елементів. Кооперативне зв'язування ТФ зі своїми сайтами зсуває канонічні нуклеосоми, ремоделює хроматин і зумовлює в експерименті його гіперчутливість до нуклеаз, зокрема ДНКазі 1 (рис. 2). 95% DHSs розташовано в дистальних районах по відношенню до стартових точок транскрипції, майже половина з них – в інтронах і міжгенних проміжках, і лише 5% знаходяться в районах стартових точок транскрипції. Дистальні DHSs вміщують майже всі (97,4%) *цис*-регуляторні елементи (енхансери, інсулятори, сайленсери, райони локусного контролю). Майже всі ТФ (94,4%) зв'язуються в межах DHSs. Одна третина DHSs зустрічається тільки в клітинах певного типу, ~1,9 млн. – в клітинах двох і більше типів і тільки невелика кількість DHSs (3 692) міститься в усіх досліджених клітинах і переважно локалізується в промоторах. Оскільки доступність хроматину для дії нуклеаз досягається завдяки сумісній дії численних факторів, то на підставі характерного патерну DHSs у певному районі можна передбачити клітинспецифічну активність експресії гена [22].

Наступними за чисельністю функціональними елементами геному є CpG острівці (~1,2 млн). Їх метилювання в складі промоторів частіше пов'язане з репресією транскрипції, а в складі самого гена – з її активацією і корелює з

ацетилюванням гістонів. Показано, що існує зворотна залежність між інтенсивністю експресії транскрипційного фактора і рівнем метилювання сайту зв'язування, а саме чим інтенсивніше експресується ТФ, тим менш метильований його сайт зв'язування, що сприяє зв'язуванню фактора, і навпаки. Метилювання CpG острівців частіше спостерігається і виявляється найваріабельнішим в послідовностях самих генів і міжгенних ділянок, аніж у промоторах. Метилювання CpG острівців варіює залежно від типу клітин і є алель-специфічним [23].

Райони, збагачені на модифіковані гістони, займають 56,1% основ геному і за цим показником поступаються тільки послідовностям, які кодують довгі РНК. Глобальна модифікація гістонів через метилювання і ацетилювання в різних положеннях лізину впливає на доступність хроматину, інтенсивність транскрипції і сплайсингу настільки, що останні можуть бути спрогнозовані на підставі визначених модифікацій. Наприклад, модифікації H3K27ac, H3K9ac, H3K4me3 і H3K4me2 активують транскрипцію, тоді як H3K27me3 і H3K9me3 інгібують її. Модифікований гістон H3K4me3 є стабільним маркером стартової точки транскрипції і навіть використовується для всегеномного пошуку нових стартових точок. Баланс між модифікаціями H3K79me2 і H3K36me3 у структурі самого гена регулює сплайсинг. Метилювання гістонів має відносно стабільний характер. За яким механізмом метильовані гістони впливають на транскрипцію поки що достеменно

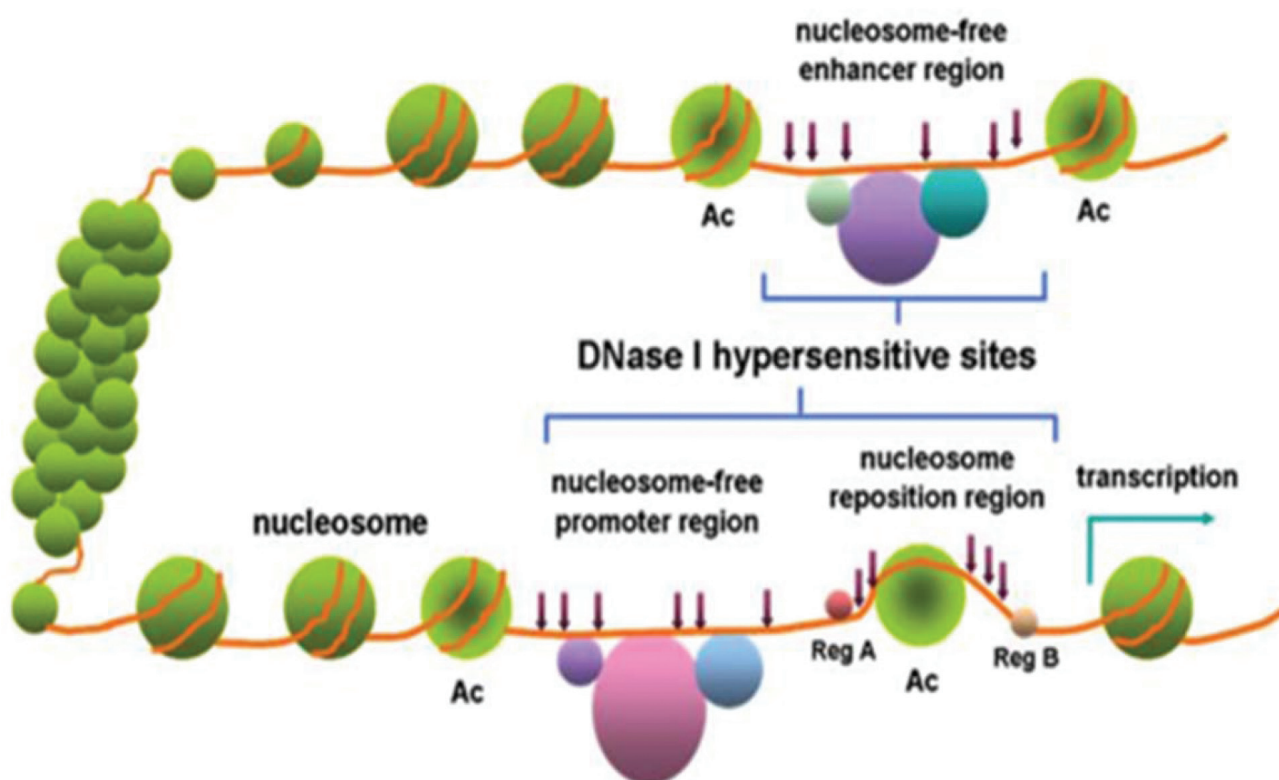


Рис. 2. Функціональне значення ділянок хроматину, гіперчутливих до ДНКазу 1 [22]. Із ділянками хроматину, вільними від нуклеосом, зв'язуються транскрипційні фактори, позначені овалами і маленькими кульками, Ac – ацетильований гістон, який сприяє релаксації структури хроматину як передумови транскрипції

невідомо, наразі вважають, що вони діють опосередковано через інші функціональні елементи. Модифікація гістонів має клітинспецифічний і алель-специфічний характер [17, 24].

Ацетилювання гістонів – це високодинамічний процес із періодом напівжиття ацетильованого і деацетильованого лізину протягом декількох хвилин. Відповідні ацетилази і деацетилази асоційовані із сайтами активної транскрипції і забезпечують зсув та відновлення нуклеосом під час проходження полімерази.

Взаємодія ділянок геному, віддалених одна від одної сотнями, навіть тисячами кілобаз, є також функціонально значовою і створює неповторний ландшафт хроматину в клітинах різного типу. Взаємодія між віддаленими участками відбувається переважно в межах однієї хромосоми. Середня кількість дистальних DHSs, які взаємодіють з DHS промотора, дорівнює 3,9, а середня кількість промоторних DHSs різних генів, що взаємодіють з одним дистальним DHS, дорівнює 2,5 (рис. 3). Частота мультикомплексів,

які включають стартові точки декількох генів, із дистальними елементами вища за таку комплексів типу «один промотор – один енхансер». Віддалені DHSs і споріднений з ними промотор, з яким вони взаємодіють, регулюються специфічними для обох елементів ТФ, а також мають спільні мотиви в своїх послідовностях. Ще одне питання було поставлене і вирішене – які гени мають найскладнішу регуляцію, тобто з якою кількістю дистальних регуляторних елементів зв'язуються їхні промотори. Виявилося, що це гени клітин імунної системи. Встановлення взаємодії віддалених елементів геному і кількісна характеристика їх стали можливими завдяки застосуванню 3С-методології (chromatin conformation capture) і її похідних (4С, 5С та ін.), що вперше дало можливість безпосередньо досліджувати тривимірну структуру хроматину [25].

Ділянками «відкритого» хроматину, які доступні для взаємодії з регуляторними факторами зайнято 15,2% геному, а

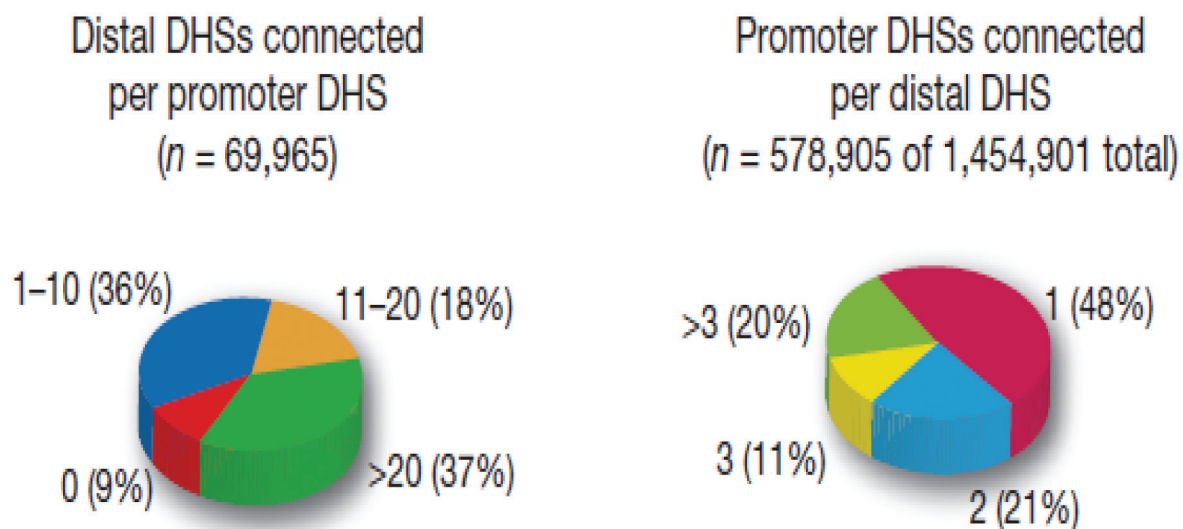


Рис. 3. Співвідношення між кількістю гіперчутливих до ДНКазі 1 ділянок у промоторах і в дистальних регуляторних ділянках генів [23]. Зліва: розподіл за кількістю гіперчутливих до ДНКазі 1 дистальних регуляторних ділянок, що зв'язані з відповідними ділянками промоторів, із розрахунку на один промотор ($r > 0,7$). Справа: розподіл за кількістю промоторних районів гіперчутливості до ДНКазі 1, що зв'язані з відповідними дистальними регуляторними ділянками, з розрахунку на одну дистальну гуперчутливу до ДНКазі 1 ділянок промотору, з якими вони корелюють із розрахунку на один непромоторний район гіперчутливості

8,1% геному вкрито протеїнами, більшість з яких є сиквенс-специфічними ТФ. Сайти зв'язування ТФ розташовані як у промоторах – із ТАТА-послідовністю і в (G+C)-збагачених послідовностях без ТАТА-ділянки, так і в міжгенних районах. Сайти в кожному з районів селективно специфічні до певних ТФ. За тим, які ТФ зв'язані в промоторі, можна прогнозувати рівень експресії відповідного гена.

Таким чином, транскрипція і процесинг РНК кількісно залежать від перелічених вище розпізнавальних знаків на хроматині, від патернів модифікації гістонів і метилування цитозинів, розташування ТФ, зв'язаних із промотором, тобто комбінація всіх цих знаків є провісником інтенсивності і спрямованості транскрипції і характеру сплайсингу. Клітинспецифічний характер експресії генів зумовлений також певною комбінацією різних функціональних елементів.

Найбільша частина послідовностей в геномі людини належить не генам, а їхнім регуляторам і протеїнокодуєчим РНК, вивчення функцій яких ще чекає на дослідників, як і функції 20% геному з невиявленими поки що функціональними елементами. Важливо, що асоційований із хворобами однонуклеотид-

ний поліморфізм у більшості випадків розташований також за межами протеїнокодуєчих послідовностей, і фенотип хвороби асоціюється з клітинами певного типу або специфічними ТФ.

Одержані результати є у вільному доступі і незамінні для інтеграції з іншими даними з дослідження функціонування геному людини.

Топологія мережі транскрипційних факторів

Протягом останніх десяти років ХХ ст. численні системні дослідження принципів регуляції транскрипції геному, проведені на одноклітинних організмах (*E. coli*, дріжджі), показали, що регуляторні фактори утворюють мережу, яка має певні властивості, взаємодіє з іншими мережами клітини – протеомом, фосфо-риломом тощо. Під час виконання другого етапу проекту «Encode» вперше звернулися до аналізу мережі ТФ людини, яка наразі складається з 119 факторів із відомих близько 1800 [26]. До них входили канонічні сиквенс-специфічні фактори (88%), загальні ТФ і фактори, що модифікують і ремодельнують хроматин. Проведено більш ніж 450 експериментів із використанням від однієї до п'яти клітинних ліній і широкого набору

методів – імунопреципітації хроматину за допомогою антитіл проти ТФ, виділення і створення бібліотек сайтів зв'язування з наступним секвенуванням ділянок (ChIP-seq), knock-down-експериментів зі штучним зменшенням експресії ТФ, визначення профілю експресії генів (RNA-seq) тощо. У роботі було широко використано дані з різних баз (1000 геномів, «Gencode» як підпроект «Encode», TargetScan та інші), а також потужний арсенал математичних інструментів – теорія графів, моделювання і використання тренувальних моделей, карти самоорганізації великих наборів даних (self-organizing models, SOM) тощо. За профілем сайтів зв'язування ТФ визначали характер їх асоціації, їх патерни в проксимальних (в межах 2,5 тис. п.о.) і дистальних (в межах 8,0 тис. п.о.) ділянках від стартової точки транскрипції анованих генів, здатність зв'язуватися з алель-специфічними (материнськими чи батьківськими) ділянками геному, залежність від еволюційного тиску, взаємозв'язок між ТФ у вигляді ієрархічної структури і основних мотивів їх взаємодії. Мережу ТФ аналізували у зв'язку з експресією і регуляторними властивостями мікроРНК, а також із мережами протеїн–протеїнових взаємодій і фосфорилування, таким чином створивши і проаналізувавши мета-мережу. Нижче наведено підсумки одержаних результатів.

Комбінації ТФ й їх функціональне значення

ТФ комбінуються по-різному залежно від контексту в місці зв'язування; проксимального чи дистального району відносно стартової точки транскрипції генів; регуляторної ділянки протеїнкодуєчих і протеїннекодуєчих генів; функції регуляторної ділянки (промотор, енхансер, сайленсер). Гени, біля яких зв'язуються ТФ в одній і тій самій комбінації, частіше виявляють спільні функції.

Коасоціація ТФ

ТФ відрізняються за схильністю до певних факторів-партнерів залежно від ділянки геному, з якою вони зв'язуються. Деякі є майже постійними у своєму виборі партнера (GATA і його партнери), інші зв'язуються з одними партнерами в проксимальних районах і з іншими – в дистальних районах (FOS і його партнери);

Глобальний аналіз ТФ та їхніх партнерів виявив як відомі фактори, які утворюють димери, так і раніше невідомі, але надалі підтвержені

імунопреципітацією і мас-спектрометрією. За характером коасоціації ТФ розподілено на дев'ять класів, які утворюють чотири групи: фактори, які зв'язуються тільки із проксимальними або тільки з дистальними районами, з обома районами і з ділянками, які відповідають за репресію генів.

Ієрархія ТФ

Для повноти уявлення про взаємозв'язок між ТФ і про їх місце та важливість у системі створено мережу ТФ, які зв'язуються тільки в промоторних ділянках генів. Взаємозалежність ТФ, тобто який фактор є регулятором для експресії іншого, визначено на підставі профілю сайтів зв'язування ТФ відносно генів, що їх кодують, і перевірено в knock-down-експериментах з усіма 119 обраними ТФ. Для кожного з них розраховано показник центральності, тобто з кількома іншими транскрипційними факторами він зв'язаний – скільки ТФ регулюють його експресію (вхідна ступінь зв'язаності вузла графу, англ. in-degree centrality) або експресію якої кількості ТФ він регулює (вихідна ступінь зв'язаності вузла графу, англ. out-degree centrality). Кожний з факторів у мережі наводиться у вигляді вузла графу, а ребра цього графу відтворюють взаємодію й її напрямок між вузлами (рис. 4). Цю мережу проаналізовано разом з іншими мережами, а саме з мережею ТФ, що зв'язуються дистально, мережами мікроРНК і фосфокіназ, які задіяні в їх регуляції, а також з даними щодо їх протеїн–протеїнової взаємодії.

Транскрипційні фактори утворюють трирівневу ієрархію – вищий або управлінський рівень, проміжний, який регулює і регулюється, і нижчий, переважно регульований. Вищий рівень збагачений на транскрипційні фактори, які модулюють хроматин, його представники активніше взаємодіють між собою (connectivity), мають більшу кількість мішеней як протеїнкодуєчих, так і протеїннекодуєчих генів порівняно з нижчими рівнями. Гени, які є регульованими майже виключно факторами вищого рівня ієрархії, беруть участь у таких загальних біологічних процесах, як розвиток, диференціювання, морфогенез і т.п. Серед транскрипційних факторів проміжного рівня частіше, ніж у вищому рівні, представлено фактори, що зв'язуються в промоторах у межах стартових точок транскрипції. Гени, які регулюються ТФ середнього і нижчого рівнів ієрархії,

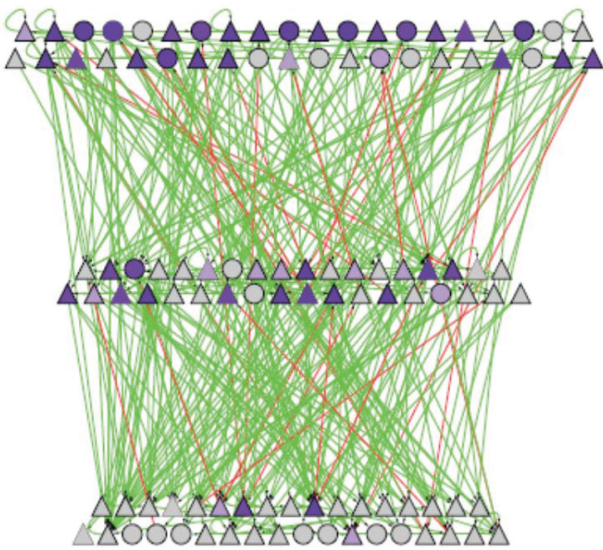


Рис. 4. Ієрархія проксимальних транскрипційних факторів [26]. Вузли графу є транскрипційними факторами. Канонічні сиквенс-специфічні фактори позначені трикутниками, а інші – кружечками. Ребра графу відображають взаємодію між факторами, що регулюють, і регульованими. Низхідний і висхідний напрямки дії позначені відповідно зеленим і червоним кольором. Рівень затінювання вузлів відповідає ступеню їх вихідної зв'язаності. Чим цей показник вищий, тим інтенсивніше забарвлення

відповідають за менш масштабні процеси, їх експресія має виражений клітинспецифічний характер. ТФ проміжного рівня утворюють вузьке місце (bottle neck) у системі, оскільки через них проходить найбільше число найкоротших шляхів між різними факторами вищого і нижчого рівнів. Це означає, що саме цей рівень регуляції є найбільш лімітуючим, а відтак і найвірогіднішим для цілеспрямованих втручань.

ТФ вищого і проміжного рівнів частіше фізично взаємодіють або асоціюються між собою на відміну від факторів нижчого рівня; вони також більше взаємодіють з іншими мережами, а саме з мережею протеїн–протеїнових взаємодій, з некодуючими РНК, зокрема з мікроРНК. Останні поділяються на ті, що регулюють ТФ, і ті, що регулюються ними. Взаємодія з кіназами, що належать до різних рівнів ієрархії фосфорилому, засвідчила, що кінази вищого рівня регулюють ТФ, а кінази нижчого рівня переважно регулюються ТФ.

Збагачені мотиви в мережі ТФ

Мережа ТФ вирізняється наявністю певних мотивів – статистично значущих структур у мережі, що повторюються і відображають специфічний характер взаємодії між вузлами мережі. Виявлено типовий ланцюг прямого зв'язку (feed-forward-loop, FFL) – структуру із трьох транскрипційних факторів, в якій один ТФ регулює інший і обидва регулюють третій – мішень. Інший мотив – це різновидність ланцюга прямого зв'язку, в якому два ТФ взаєморегулюються (toggle-switches). Його вважають важливим для визначення долі клітини [27]. Наступний досить поширений мотив у мережі – протеїн–протеїнова взаємодія між ТФ, які зв'язуються з віддаленими між собою ділянками ДНК, наприклад, дистальними і проксимальними районами. Це співзвучне ідеї щодо випетлювання ДНК для фізичного зближення дистального і проксимального районів і утворення загального транскрипційного регуляторного комплексу. Суттєво, що і мікроРНК переважно зв'язуються з димерами ТФ, а також із проксимальними і дистальними регуляторами одного і того самого гена, і, таким чином, одразу за різними напрямками виключають його експресію. І, нарешті, мотив авторегуляції (у 28 з 119 факторів), який частіше пов'язаний з репресивною функцією і має більше мішеней серед протеїннекодуючих РНК, ніж ТФ, позбавлені мотивів авторегуляції.

Алель-специфічне зв'язування ТФ і алель-специфічна експресія генів-мішеней

Для того, щоб визначити як регулюється експресія генів із гетерозиготним одонуклеотидним поліморфізмом, для кожного ТФ окреслили набір регульованих ним гетерозиготних генів, алель, з якою він зв'язується, і алель, з якою відбувається транскрипція. Виявилось, що деякі ТФ зв'язуються переважно з материнським або з батьківським алелем. Відповідно є гетерозиготні гени, які регулюються ТФ, що зв'язуються з однаковою частотою як із батьківським, так і з материнським алелем. Чим більше ТФ, задіяних в алель-специфічній регуляції певного гену, тим координація між експресією генів і зв'язуванням факторів із материнським чи батьківським алелем зростає.

Селекція в контексті мережі ТФ

Гени, які регулюються більшим числом ТФ, знаходяться під сильнішою негативною селекцією. У той самий час, чим більше мішеней має той чи інший ТФ, тим він більшою мірою підпадає під негативну селекцію. Факторами, за якими оцінюється вплив селекції, є співвідношення несинонімічних до синонімічних одонуклеотидних поліморфізмів і густина несинонімічних одонуклеотидних поліморфізмів. Чим менші ці величини, тим більше тиск негативної селекції в еволюції.

Цікаво, що місця зв'язування ТФ, що виявляють алейну специфічність, знаходяться під меншим негативним тиском порівняно з тими, що її не виявляють, тобто густина одонуклеотидних поліморфізмів у них вища за інші.

Таким чином, другий етап виконання проекту «Encode» закінчено. Одержано безліч нових даних, розроблено численні нові методичні підходи і сформульовано принципи функціонування мережі ТФ разом із мережами мікроРНК, протеїн–протеїнових взаємодій і фосфорилуму. Виявлено загальні характеристики мереж, які є типовими для еукаріотів і для людини, зокрема. Третій етап проекту триває.

Паралельно з проектом «Encode» виконувався проект «modEncode». Його метою було створити енциклопедію функціональних елементів для *C. elegans* [28] і *D. melanogaster* [29]. Його вичерпні результати опубліковано у 2010 р. [див. блог <http://blog.modencode.org/papers/>].

Виконання проекту «Encode» є яскравим прикладом співробітництва молекулярних біологів, математиків, біоінформатиків, які об'єднали свої методологічні підходи для одержання якісно нових результатів і надання вільного доступу для користування ними.

Post Scriptum

Системна біологія в Україні як новий напрям дослідження тільки-но починає виокремлюватися. У червні 2009 р. у відділі механізмів трансляції генетичної інформації (зав. віділом акад. НАНУ Г. В. Єльська) Інституту молекулярної біології і генетики Національної академії наук України було організовано першу в Україні лабораторію системної біології. За час існування колективу створено програмний продукт всегеномного пошуку сайтів

зв'язування ТФ Cotrasif (<http://biomed.org.ua/COTRASIF/>), вдосконалений Ensembl-метод реконструкції мереж генної регуляції, реконструйовано стехіометричну модель фолатного циклу в плаценті людини, визначено профіль генної експресії в первинних гепатоцитах щура під впливом квазіфізіологічних доз інтерферону альфа методом гібридизації з мікроареєм олігонуклеотидів. Продовжується робота з експериментальної перевірки теоретичних передбачень [30]. Істотний поштовх для розвитку в Україні системної біології, біоінформатики і її запровадження в медицину дало виконання Державної цільової науково-технічної програми із впровадження і застосування грид-технологій на 2009–2013 роки. Програма сприяла удосконаленню методів реконструкції мереж генної регуляції, обробки і зберігання в електронному вигляді індивідуальних даних медичного обстеження.

Незважаючи на відсутність в Україні приладів для проведення масштабних експериментів, існує можливість працювати з численними базами даних, проводити такі експерименти в співпраці з іноземними колегами, вчитись аналізувати і обробляти такі результати і, нарешті, вивчити специфічну мову системної біології, без знання якої навіть світові здобутки не є зрозумілими. Все це потребує часу, зусиль і активної співпраці з представниками точних наук. У Києві функціонують Інститут високих технологій і факультет біомедичного обладнання в Політехнічному інституті, які можуть бути залученими до таких починань. Потрібно торувати той шлях, який пройшли інші країни – виховувати кадри, вводити дисципліну у вищі наукові заклади, аспірантуру (див. Доповідь Королівської академії технічних наук і Академії медичних наук, Велика Британія, 2007; <http://www3.imperial.ac.uk/pls/portallive/docs/1/23357696.PDF>), аби не залишитися навік на далекому узбіччі.

Світ рухається вперед велетенськими кроками. Так, у Німеччині за підтримки Міністерства освіти і науки виконується проект «Virtual Liver Network». Його мета – створити динамічну модель печінки, яка б відтворювала морфологію, функціонування органа шляхом інтеграції кількісних даних з усіх рівнів його організації [<http://www.virtual-liver.de/wordpress/en/2010/06/>]. В Америці в штаті Огайо виконується пілотний

проект «4P medicine» – Predictive, Preventive, Personalized and Participatory, що запроваджений з метою прогнозувати і попереджувати захворювання у конкретної людини за співучасті самої людини у збереженні власного здоров'я. Проект реалізується за використання гено- і фенотипування кожного індивідуума. Ця програма запропонована доктором Лероем Худом (Leroy Hood, MD, PhD), президентом і співзасновником Інституту системної біології (Сіетл, США), винахідником автоматизованого секвенування і синтезу ДНК. Вартість генотипування стрімко падає і вже нині можна визначити численні ризики розповсюджених захворювань, наприклад, раку молочної залози, хвороби Альцгеймера тощо лише за 100\$ (компанія 23andMe - Mountain View, California). Сумісний аналіз результатів кількох всегеномних проектів і досліджень, що проводяться на сьогодні, надає великі надії для ідентифікації функціональних елементів геному, на які припадають варіабельні ділянки, пов'язані із загальними хворобами і визначені у всегеномному дослідженні асоціацій (Genome Wide Association Study, GWAS). Це тільки короткий, довільний перелік деяких перспектив розвитку системної біології, і українські науковці не мають залишатися осторонь.

СИСТЕМНАЯ БИОЛОГИЯ И ПРОЕКТ «ENCODE»

М. Ю. Оболенская

Институт молекулярной биологии
и генетики НАН Украины, Киев;
e-mail: m.obolenska@gmail.com

В обзоре дается первоначальное представление относительно истоков системной биологии, предпосылок для ее выделения в отдельную науку, освещаются основные принципы системной биологии, особенности методологии и достижения в исследовании функциональных элементов генома человека и способов их координированной и дифференцированной регуляции. Приводится краткая характеристика функциональных элементов: протеинкодирующие последовательности и те, которые кодируют РНК, но не транслируются, гиперчувствительные к ДНКазе 1 участки, метилированные CpG-островки, модифицированные гистоны и специфическая трехмерная структура хроматина.

Приводятся сведения относительно топологии сети транскрипционных факторов, ее основных мотивов, иерархии, комбинации, их коассоциации и особенностей их связывания с аллель-специфическими последовательностями.

Ключевые слова: системная биология, функциональные элементы генома, сеть генной регуляции.

SYSTEM BIOLOGY AND THE PROJECT ENCODE

M. Yu. Obolenskaya

Institute of Molecular Biology and Genetics,
National Academy of Sciences of Ukraine, Kyiv;
e-mail: m.obolenska@gmail.com

The goal of this review is to give an incipient knowledge on the background of system biology, the premises to its assignment as a new branch of biology, its principles, methodology and its great achievements in identification of functional elements of human genome and regulation of their concordant and differential activity. The short characteristics of functional elements including the protein-coding sequences and those coding noncoding RNAs, the DNase 1 hypersensitivity sites and methylated CpG islets, modified histones and specific 3D structure of chromatin, are represented. The topology of transcription factors network with its main motifs, hierarchy, combination and association of transcription factors and their allelic specificity are highlighted.

Key words: system biology, functional elements of genome, network of transcription factors.

References

1. *Von Bertalanffy L.* The theory of open systems in physics and biology // *Science*. – 1950. – **111**, N 2872. – P. 23–29.
2. *Hodgkin A. L., Huxley A. F.* A quantitative description of membrane current and its application to conduction and excitation in nerve // *J. Physiol.* – 1952. – **117**, N 4. – P. 500–544.
3. *Waddington C. H.* The Nature of Life. – London: George, Allen, & Unwin, 1961. – 131 p.
4. *Noble D.* Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations // *Nature*. – 1960. – **188**, N 4749. – P. 495–497.

5. Mesarovic' M. D. Systems Theory and Biology—View of a Theoretician / In: System Theory and Biology, ed. M. D. Mesarovic, Vienna: Springer-Verlag. — 1968. — P. 59–87.
6. Kacser H., Burns J. A. The control of flux // Symp. Soc. Exp. Biol. — 1973. — **27**. — P. 65–104.
7. Heinrich R., Rapoport T. A. A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector // Eur. J. Biochem. — 1974. — **42**, N 1. — P. 97–105.
8. Kauffman S. A. Metabolic stability and epigenesis in randomly constructed genetic nets // J. Theor. Biol. — 1969. — **22**, N 3. — P. 437–467.
9. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome // Nature. — 2004. — **431**, N 7011. — P. 931–945.
10. Wolkenhauer O., Mesarovic M. Feedback dynamics and cell function: Why systems biology is called Systems Biology // Mol. Biosyst. — 2005. — **1**, N 1. — P. 14–16.
11. International HapMap 3 Consortium, Altshuler D. M. et al. Integrating common and rare genetic variation in diverse human populations // Nature. — 2010. — **467**, N 7311. — P. 52–58. doi: 10.1038/nature09298.
12. 1000 Genomes Project Consortium, Abecasis G. R., Altshuler D., Auton A., Brooks L. D., Durbin R. M., Gibbs R. A., Hurles M. E., McVean G. A. A map of human genome variation from population-scale sequencing // Nature. — 2010. — **467**, N 7319. — P. 1061–1073. doi:10.1038/nature09534.
13. 1000 Genomes Project Consortium, Abecasis G. R., Auton A., Brooks L. D., DePristo M. A., Durbin R. M., Handsaker R. E., Kang H. M., Marth G. T., McVean G. A. An integrated map of genetic variation from 1,092 human genomes // Nature. — 2012. — **491**, N 7422. — P. 56–65.
14. Witte J. S. Genome-wide association studies and beyond // Annu. Rev. Public Health. — 2010. — **31**. — P. 9–24.
15. ENCODE Project Consortium, Birney E. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project // Nature. — 2007. — **447**, N 7146. — P. 799–816.
16. Maher B. ENCODE: The human encyclopedia // Nature. — 2012. — **489**, N 7414. — P. 46–48.
17. ENCODE Project Consortium. Bernstein B. E., Birney E., Dunham I., Green E. D., Gunter C., Snyder M. An integrated encyclopedia of DNA elements in the human genome // Nature. — 2012. — **489**, N 7414. — P. 57–74.
18. Pei B., Sisu C., Frankish A., Howald C., Habegger L., Mu X. J., Harte R., Balasubramanian S., Tanzer A., Diekhans M., Reymond A., Hubbard T. J., Harrow J., Gerstein M. B. The GENCODE pseudogene resource // Genome Biol. — 2012. — **13**, N 9. — R51.
19. Mercer T. R., Dinger M. E., Mattick J. S. Long non-coding RNAs: insights into functions // Nat. Rev. Genet. — 2009. — **10**, N 3. — P. 155–159. doi: 10.1038/nrg2521. Review.
20. Gerstein M. B., Bruce C., Rozowsky J. S., Zheng D., Du J., Korbel J. O., Emanuelsson O., Zhang Z. D., Weissman S., Snyder M. What is a gene, post-ENCODE? History and updated definition // Genome Res. — 2007. — **17**, N 6. — P. 669–681.
21. Moran V. A., Perera R. J., Khalil A. M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs // Nucleic Acids Res. — 2012. — **40**, N 14. — P. 6391–6400.
22. Wang Y. M., Zhou P., Wang L. Y., Li Z. H., Zhang Y. N., Zhang Y. X. Correlation between Dnase I hypersensitive site distribution and gene expression in HeLa S3 cells // PLoS One. — 2012. — **7**, N 8. — e42414.
23. Thurman R. E., Rynes E., Humbert R., Vierstra J., Maurano M. T., Haugen E., Sheffield N. C., Stergachis A. B., Wang H., Vernot B., Garg K., John S., Sandstrom R., Bates D., Boatman L., Canfield T. K., Diegel M., Dunn D., Ebersol A. K., Frum T., Giste E., Johnson A. K., Johnson E. M., Kutayvin T., Lajoie B., Lee B. K., Lee K., London D., Lotakis D., Neph S., Neri F., Nguyen E. D., Qu H., Reynolds A. P., Roach V., Safi A., Sanchez M. E., Sanyal A., Shafer A., Simon J. M., Song L., Vong S., Weaver M., Yan Y., Zhang Z., Zhang Z., Lenhard B., Tewari M., Dorschner M. O., Hansen R. S., Navas P. A., Stamatoyannopoulos G., Iyer V. R., Lieb J. D., Sunyaev S. R., Akey J. M., Sabo P. J., Kaul R., Furey T. S., Dekker J., Crawford G. E., Stamatoyannopoulos J. A. The accessible chromatin landscape of the human genome // Nature. — 2012. — **489**, N 7414. — P. 75–82.
24. Zentner G. E., Henikoff S. Regulation of nucleosome dynamics by histone modifications

- // Nat. Struct. Mol. Biol. – 2013. – **20**, N 3. – P. 259–266. doi: 10.1038/nsmb.2470.
25. De Laat W., Dekker J. 3C-based technologies to study the shape of the genome// Methods. – 2012. – **58**, N 3. – P. 189–191. doi: 10.1016/j.ymeth.2012.11.005.
 26. Gerstein M. B., Kundaje A., Hariharan M., Landt S. G., Yan K. K., Cheng C., Mu X. J., Khurana E., Rozowsky J., Alexander R., Min R., Alves P., Abyzov A., Addleman N., Bhardwaj N., Boyle A. P., Cayting P., Charos A., Chen D. Z., Cheng Y., Clarke D., Eastman C., Euskirchen G., Frietze S., Fu Y., Gertz J., Grubert F., Harmanci A., Jain P., Kasowski M., Lacroute P., Leng J., Lian J., Monahan H., O'Geen H., Ouyang Z., Partridge E. C., Patacsil D., Pauli F., Raha D., Ramirez L., Reddy T. E., Reed B., Shi M., Slifer T., Wang J., Wu L., Yang X., Yip K. Y., Zilberman-Schapira G., Batzoglou S., Sidow A., Farnham P. J., Myers R. M., Weissman S. M., Snyder M. Architecture of the human regulatory network derived from ENCODE data // Nature. – 2012. – **489**, N 7414. – P. 91–100.
 27. Zhou J. X., Huang S. Understanding gene circuits at cell-fate branch points for rational cell-reprogramming // Trends Genet. 2011. – **27**, N 2. – P. 55–62.
 28. Gerstein M. B., Lu Z. J., Van Nostrand E. L., Cheng C., Arshinoff B. I., Liu T., Yip K. Y., Robilotto R., Rechtsteiner A., Ikegami K., Alves P., Chateigner A., Perry M., Morris M., Auerbach R. K., Feng X., Leng J., Vielle A., Niu W., Rhrissorrakrai K., Agarwal A., Alexander R. P., Barber G., Brdlik C. M., Brennan J., Brouillet J. J., Carr A., Cheung M. S., Clawson H., Contrino S., Dannenberg L. O., Dernburg A. F., Desai A., Dick L., Dosé A. C., Du J., Egelhofer T., Ercan S., Euskirchen G., Ewing B., Feingold E. A., Gassmann R., Good P. J., Green P., Gullier F., Gutwein M., Guyer M. S., Habegger L., Han T., Henikoff J. G., Henz S. R., Hinrichs A., Holster H., Hyman T., Iniguez A. L., Janettex J., Jensen M., Kato M., Kent W. J., Kephart E., Khivansara V., Khurana E., Kim J. K., Kolasinska-Zwierzx P., Lai E. C., Latorre I., Leahey A., Lewis S., Lloyd P., Lochovsky L., Lowdon R. F., Lubling Y., Lyne R., MacCoss M., Mackowiak S. D., Mangone M., McKay S., Mecnas D., Merrihew G., Miller D. M. 3rd, Muroyama A., Murray J. I., Ooi S. L., Pham H., Phippen T., Preston E. A., Rajewsky N., Räscht G., Rosenbaum H., Rozowsky J., Rutherford K., Ruzanov P., Sarov M., Sasidharan R., Sboner A., Scheid P., Segal E., Shin H., Shou C., Slack F. J., Slightam C., Smith R., Spencer W. C., Stinson E. O., Taing S., Takasaki T., Vafeados D., Voronina K., Wang G., Washington N. L., Whittle C. M., Wu B., Yan K. K., Zeller G., Zha Z., Zhong M., Zhou X. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project // Science. – 2010. – **330**, N 6012. – P. 1775–1787. doi: 10.1126/science.1196914.
 29. modENCODE Consortium, Roy S., Ernst J., Kharchenko P. V., Kheradpour P., Negre N., Eaton M. L., Landolin J. M., Bristow C. A., Ma L., Lin M. F., Washietl S., Arshinoff B. I., Ay F., Meyer P. E., Robine N., Washington N. L., Di Stefano L., Berezikov E., Brown C. D., Candéias R., Carlson J. W., Carr A., Jungreis I., Marbach D., Sealfon R., Tolstorukov M. Y., Will S., Alekseyenko A. A., Artieri C., Booth B. W., Brooks A. N., Dai Q., Davis C. A., Duff M. O., Feng X., Gorchakov A. A., Gu T., Henikoff J. G., Kapranov P., Li R., MacAlpine H. K., Malone J., Minoda A., Nordman J., Okamura K., Perry M., Powell S. K., Riddle N. C., Sakai A., Samsonova A., Sandler J. E., Schwartz Y. B., Sher N., Spokony R., Sturgill D., van Baren M., Wan K. H., Yang L., Yu C., Feingold E., Good P., Guyer M., Lowdon R., Ahmad K., Andrews J., Berger B., Brenner S. E., Brent M. R., Cherbas L., Elgin S. C., Gingeras T. R., Grossman R., Hoskins R. A., Kaufman T. C., Kent W., Kuroda M. I., Orr-Weaver T., Perrimon N., Pirrotta V., Posakony J. W., Ren B., Russell S., Cherbas P., Graveley B. R., Lewis S., Micklem G., Oliver B., Park P. J., Celniker S. E., Henikoff S., Karpen G. H., Lai E. C., MacAlpine D. M., Stein L. D., White K. P., Kellis M. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE // Science. – 2010. – **330**, N 6012. – P. 1787–1797. doi: 10.1126/science.1198374.
 30. Obolenskaya M. Yu., Tokovenko B. T., Kuklin A. V., Frolova A. A., Rodriguez R. R., Dotsenko V. A., Dragushchenko O. O. The start of systems biology in Ukraine // Biopolym. Cell. – **30**, N 1. – P. 16–24.

Отримано 24.09.2013