



УДК 575.113:577.2

В. М. Запорожан, Ю. І. Бажора

## РОЗШИФРОВКА ГЕНОМУ ЛЮДИНИ І МИШІ — ВАЖЛИВИЙ ЕТАП У РОЗВИТКУ МОЛЕКУЛЯРНОЇ МЕДИЦИНИ

Одеський державний медичний університет

Наприкінці 80-х років ХХ ст. група вчених на чолі з Д. Уотсоном (один з авторів моделі ДНК) склали програму розшифровки геному людини, роботи над якою почалися у 1990 р. Усього на її виконання витрачено близько 6 млрд доларів. У лютому 2001 р. було опубліковано дві попередні версії геному людини. Це результат багаторічної праці багатьох вчених, що склали дві групи. Перша з них — Міжнародний некомерційний проект «Геном людини» — Human Genom Project (HGP) — об'єднав 20 лабораторій, сотні вчених з різних країн світу. Ця група поставила перед собою за мету розшифровку геному людини й одержання даних, які могли б стати загальнодоступними. Приватна ж компанія «Целера Геномікс» (Celera Genomics) також поставила перед собою завдання розшифровки геному людини, але планувала надавати отриману інформацію на комерційній основі.

Обидві версії містять ще багато білих плям і неточностей, тому робота триває. Проте отримані результати дозволили зіставити геном людини із геномами інших еукаріотів (дріжджів, черв'яка, мухи дрозофіли та рослини). Установлено, що послідовність геному людини, як і інших еукаріотів, складається з ділянок, що кодують білки (? 2 %), ділянок,

що кодують РНК (? 20 %), а понад 50 % становлять повтворювані послідовності, які важко клонуються і тому створюється багато пробілів.

Отже, велика частина геному людини не кодує білки. До цієї частини входять фрагменти, які кодують тільки РНК і ділянки ДНК повторів.

Тисячі генів у людини тільки транскрибуються і продукують РНК, які не кодують білок (нкРНК). Ідентифіковано також близько півтисячі генів для транспортних РНК. Наразі немає повних послідовностей для рибосомальних РНК (рРНК), хоча інтерес до них дуже великий з огляду на їхню роль в утворенні пептидних зв'язків при трансляції.

Крім того, ідентифіковано близько 80 малих ядерних РНК, які беруть участь у сплайсингу незрілої РНК, а також майже сотню генів малих ядерцевих РНК, які беруть участь у процесингу.

Гени нкРНК і псевдогени, які утворилися з них, за своїми розмірами є малими і не поділяються на групи — це специфічні структурні особливості, пошук їх за допомогою комп'ютерних методів є важким, хоча вони дуже поширені в геномі людини.

Давно відомо, що прямої кореляції між кількістю ДНК у клітин і рівнем організації живих істот немає. Наприклад, геном людини у 200 разів мен-

ше геному амеби. Причиною цього є те, що в геномі еукаріотів дуже багато повторів (понад 50 % ДНК). Їх поділяють на кілька груп:

1. Довгі та короткі повтори — виникли із рухливих елементів і розсіяні по всьому геному. Вони становлять приблизно 45 % геному людини.

2. Неактивні копії генів (псевдогени) — утворилися в результаті зворотної транскрипції.

3. Прості повтори послідовностей — складаються з 2–8 основ («мікросателіти») або з більш довгих повторів (14–500 основ), іменованих «мінісателітами». У популяціях людини прості повтори характеризуються неабияким поліморфізмом за довжиною, що має важливе значення для генетичних досліджень.

4. Повтори великих (від 1 до 200 кпн) сегментів ДНК — копіюються з одного району геному в інший у межах однієї хромосоми чи між різними хромосомами.

5. Тандемно повторювані послідовності — часто спостерігаються біля централіт хромосом і в теломірних ділянках хромосом людини.

Раніше вважалося, що ДНК-повтори не є предметом наукового інтересу, і це в подальшому призвело до деяких помилкових висновків. Наразі припускають, що «егоїстична ДНК» (термін запропоновано



Ф. Криком) створює нові регуляторні елементи чи генерує нові гени. У геномі людини нарахувано 47 генів — похідних від звичайних транспозонів.

Важливим досягненням при розшифровці геному людини виявилось встановлення кількості генів, які кодують білки. На момент оголошення про практичне завершення розшифровки геному вчені групи HGP нарахували 32 000 генів, які кодують білки, а група Celera Genomics — близько 40 000, що виявилось набагато меншим, ніж постулювалося раніше (до 150 000). Є припущення, що протеом людини може містити до 250 000 різних білків. Настільки велика розбіжність між очікуваними і фактично отриманими даними можна пояснити так. По-перше, повну розшифровку не завершено і багато проблем ідентифікації генів ще треба буде розв'язати. По-друге, не виключено, що людина якось незрозумілим чином ощадливо використовує мінімальну кількість генів для шифрування великої кількості білків. Подальші дослідження у цьому напрямку допоможуть визначити точну кількість людських генів, які кодують білки.

Отримані дані дають змогу охарактеризувати типовий ген людини. Він складається приблизно з 28 000 основ і містить у собі вісім екзонів. Кодуюча послідовність такого гена має близько 1300 пн і містить інформацію про білок, який складається з 450 амінокислот. Найбільший із виявлених генів людини — ген дистрофіну ( $2,4 \cdot 10^6$  пн). Фібрилярний білок титин охоплює 27 000 амінокислотних залишків, а його ген містить найбільшу (234) кількість екзонів серед усіх виявлених генів людини.

Структура генів людини є набагато складнішою, ніж в інших структурах еукаріотів. Часто вони перериваються великими інтронами, приблизно 35 % генів можуть зчитувати-

ся з різними рамками, 40 % іРНК можуть піддаватися альтернативному сплайсингу. Отже, одна послідовність ДНК може кодувати більше ніж один тип іРНК.

У людській популяції широко розповсюджені поодинокі заміни нуклеотидів — поліморфізм поодиноких нуклеотидів — single nucleotide polymorphism (SNPS). У геномі людини точно картовано 1,42 млн поодиноких нуклеотидних заміни. Вони частіше зустрічаються в ділянках кодуючих генів, ніж у ділянках повторів ДНК. У середині генів виявлено 60 000 поодиноких заміни нуклеотидів. Практично кожен ген людини відрізняється варіабельністю. В геномі людини є ділянки з підвищеною та зниженою варіабельністю. Наприклад, ділянки головного комплексу гістосумісності (MHC), які кодують білки гістосумісності, відрізняються значною варіабельністю, визначаючи імунологічну індивідуальність людини.

Генетичний поліморфізм має велике біологічне значення для людини. Так, поліморфізм гена апоЕ4 сприяє збільшенню щільності бляшок при хворобі Альцгеймера, а делеція гена, що кодує хемокіновий рецептор CCR 5, збільшує стійкість імунодефіциту людини щодо вірусу. Цей корецептор, разом з рецептором CD4, є необхідним для зв'язування та проникнення вірусу в Т-лімфоцит. При порівнянні розташування та частоти поодиноких заміни у хворих і здорових людей виявляються ті заміни, які пов'язані з тією чи іншою хворобою. Такі зіставлення дають можливість з'ясувати роль певних генів у розвитку мультифакторіальних захворювань. Дослідження в цьому напрямку є дуже перспективними й наразі інтенсивно розвиваються.

У 1994 р. в молекулярній біології виник новий термін — протеом. Він, фактично, покли-

каний описати усі сукупності білків, що синтезуються протягом життя клітин організму. Галузь дослідження структури та функції білків — продуктів функціонування генів — дістала назву протеоміка. Її значення в медицині є вкрай важливим, тому що будуть ідентифіковані білкові маркери різних хвороб. Перспективним також є вивчення ефектів взаємодії лікарських речовин з геномом людини (фармакогеноміка).

Слід зазначити, що розшифровка первинної структури білків на основі вивчених генів, які кодують білки, ще не вказує на розкриття функцій тих чи інших продуктів генів. За цим йтиме тривалий систематичний аналіз протеому людини. Велике значення в розшифровці ролі певних генів, які синтезують білок, має порівняння первинної структури білків з відомими та невідомими функціями, які отримані від представників видів різного рівня еволюційного розвитку. Сьогодні така робота розпочинається. На основі виявлення тільки первинної структури білка не можна говорити про точну його функцію. Проте вивчення геному дає важливу інформацію про виникнення білкових доменів, про розширення їхніх родин, родин самих білків і т. ін.

У геномі людини виявлено гени, гомологічні таким у геномі мухи (61 %), у геномі черв'яка (43 %), у геномі дріжджів (46 %). Це основний набір генів, які кодують головні життєві процеси в клітині: основний метаболізм, реплікація та репарація ДНК, біосинтез білка.

Виявлено також понад 220 генів, продукти яких схожі на білки бактерій, але не схожі на білки дріжджів, рослин, безхребетних. Швидше за все, ці гени потрапили до геному людини від бактерій шляхом переносу.

Зіставлення геному людини і досліджених безхребетних дало можливість виявити значно більшу кількість генів, які



відповідають за різні регуляторні функції в організмі: захист та імунність; розвиток, структура і функції центральної нервової системи, білків, що беруть участь у побудові цитоскелета і русі везикул; внутрішньо- та міжклітинна сигналізація в розвитку та гомеостазі; транскрипція і трансляція; гемостаз; апоптоз та ін.

Для геному людини характерним є збільшення білкових родин, які беруть участь у розвитку нервової системи: нейротрофічні фактори, фактори росту нервів, мієлінові, синаптичні білки, сигнальні молекули. За допомогою традиційних методів виявлено ряд генів для білків цитоскелета клітини людини і вищих хребетних.

Порівняно з безхребетними, у людини істотно побільшало родин білків, які беруть участь у процесах розвитку і диференціювання: фактор росту фібробластів, фактор росту нервів та інші фактори росту: гормони, рецептори, сигнальні молекули, фактори транскрипції. Кількість генів, які кодують транскрипційні білки, а також білки, які беруть участь у посттранскрипційних процесах, у людини виявилася значно більшою, ніж у безхребетних.

Секвенування геному стало поштовхом до дослідження генів, які «відповідають» за хвороби людини. Потрібним є проведення функціональної класифікації самих генів та їхніх продуктів — білків. Усі гени (923), які викликають моногенні захворювання чи підвищують імовірність виникнення хвороби, характеризувалися за функцією їхніх продуктів щодо патологічного процесу та клінічних проявів. Найбільшу функціональну групу становили ферменти (31 %). Друга за величиною група — білки-активатори і стабілізатори, білки, які беруть участь у правильному згортанні поліпептидних ланцюгів (14 %). Кожна із решти груп (рецептори,

фактори транскрипції, трансмембранні переносники та ін.) становили менше 10 % від усіх генів, які викликають хвороби. Кореляційний аналіз між функцією продуктів генних хвороб і віком хворих показав, що хвороби, пов'язані із розладами функції ферментів, виявляються на всіх етапах розвитку. Тим же часом хвороби, пов'язані з генами, що кодують транскрипційні фактори, виявляються на етапі внутрішньоутробного розвитку.

Таким чином, секвенування геному людини свідчить про ускладнення геному в ході еволюції — від дріжджів до людини. Однак кількість генів збільшилася лише у 5 разів. Ускладнення полягає у виникненні великої кількості білків, а не генів, які синтезують білок. Організм людини, використовуючи відомі структурні конструкції, збирає нові білки з новими функціями. Можливо, це досягається за допомогою складних механізмів посттранскрипційних процесів.

Слід зазначити, що описані досягнення ще не є розшифрованою генетичною «тексту». Справжнє читання генетичного тексту, переклад його з мови молекул на мову характеристик морфологічних і функціональних особливостей людини, її хвороб тільки розпочинаються.

Перспектива одержати відповідь на ці та інші питання щодо геному людини з'явилася в грудні 2002 р., коли було завершено секвенування геному миші. Це шостий відсеквенований еукаріотичний геном і другий після людини — ссавця.

Порівняння геномів людини і миші дозволить краще розуміти еволюцію і функціонування геному. За попередніми даними, у миші кількість генів, які кодують білок, орієнтовно становить 27–35,5 тис., а в людини, як вже було зазначено вище, — 30–40 тис. Таким чином, у ссавців кількість генів, які кодують білок, усього 30–40 тис., а не 100 тис., як

прогнозувалося раніше. Імовірно, що дивергенція вихідного геному загального предка почалася 75–65 млн років тому. При цьому частота перебудов виявилася досить низькою, а деякі гени залишилися практично незмінними, завдяки чому стало можливим розпізнати синтенні райони, які відносно збереженими передалися від загального предка.

У миші 99 % генів мають подібні до геному людини послідовності, 96 % з них знаходяться усередині синтенних ділянок хромосом миші і людини. У геномі миші визначено 558 ортологосних маркерів для виявлення консервативних синтенних ділянок. Обидва геноми можна розподілити на 342 синтенні сегменти, тобто максимально довгі ділянки, в яких послідовності маркерів у хромосомі миші і людини мають однаковий порядок. Близько 90,2 % геному людини і 93,3 % геному миші містять консервативні синтенні сегменти.

Крім того, було виявлено 217 синтенних блоків. Синтенний блок — один чи більше синтенних сегментів, розташованих на однаковій хромосомі у людини та миші, але які можуть відрізнитися порядком розташування чи орієнтації. Наприклад, X-хромосома — цілий синтенний блок, а хромосома 20 людини, крім невеликого центрального сегмента, практично цілком відповідає хромосомі 2 миші. Хромосома 17 людини відповідає частині хромосоми 11 миші. Інші хромосоми людини та миші відрізняються між собою значно більше. Карта консервативних синтенних сегментів геномів миші та людини дала змогу припустити мінімальну кількість перебудов, необхідну для трансформації одного геному в інший. З урахуванням 342 синтенних сегментів, це, за вельми скромними підрахунками, — 295 перебудов. Причому у деяких ділянках хромосом перебудови могли



відбуватися повторно. На підставі знань про геном тільки двох видів ссавців неможливо визначити порядок генів у хромосомах загального попередника чи відновити точну послідовність перебудов.

Близько 40 % геному людини є спорідненим геному миші на нуклеотидному рівні. Імовірно, це ортологосні послідовності, які збереглися від спільного предка. Менше 1 % геному миші не має гомологічної ділянки в геномі людини і навпаки. Зроблено також висновок про ідентичність амінокислотної послідовності в 78,5 % випадків. Він ґрунтується на подібності близько 13 тис. генів людини та миші.

Аналіз геному миші та зіставлення його з геномом людини дало змогу виявити нові гени людини, як, наприклад, новий ген родини APOA – APOAV, який бере участь у метаболізмі ліпідів, було виявлено при порівнянні нуклеотидної послідовності миші та синтенної ділянки хромосоми 11 людини.

Порівняння геномів двох видів надало можливість також визначити відмінності між механізмами, що формували геном, включаючи різницю між мутаційним і селективним тиском. У геномі миші виявлено експансію генів, які «відповідають» за репродукцію, нюх, захист.

При порівнянні двох геномів постали також і нові питання. Виявилось, що подібні типи повторюваних послідовностей нагромаджуються у відповідних ділянках геному обох організмів, що свідчить про якісь додаткові чинники, які впливають на перебудову геному транспозонами.

Завершення секвенування геному миші має найважливіше практичне значення. Воно дасть можливість більш точно моделювати біологічні процеси і хвороби людини. Лабораторна миша — експериментальний ключ до геному людини, який дозволить маніпу-

лювати кожним геном людини.

Миша використовується як об'єкт лабораторних досліджень протягом усього XX ст. У 1909 р. було виведено першу інбредну лінію мишей DBA. На мишах інбредних ліній було проведено величезну кількість експериментів, але дані, отримані в експерименті на мишах, не завжди можна екстраполювати на людину, особливо щодо медичних аспектів. Проблема аналізу хвороб людини, які моделюються на мишах, була частково розв'язана шляхом створення трансгенних мишей, першу з яких було отримано в 1982 р., коли в ембріон миші вмонтували ген гормону росту щура під контролем Zn-залежного промотора. У 1989 р. створили першу «нокаут» мишу із селективною інактивацією гена в ембріональних стовбурних клітинах.

Знання нуклеотидних послідовностей і синтенних ділянок людини та миші дозволить цілеспрямовано вмонтувати певні гени людини чи інактувати подібні ділянки ДНК миші. Аналіз фенотипів таких мишей дає можливість виявити невідому зараз функцію окремих генів людини. Вивчення мишачих моделей полігенних хвороб людини і перебування подібних послідовностей у геномі людини дасть змогу виявити та локалізувати гени, які відповідають за розвиток цих захворювань. Знання відмінностей геномів також запобігатиме створенню неповноцінних моделей при видаленні одного гена у миші, якщо у її геномі він представлений кількома копіями.

Знання відповідності між послідовностями ДНК миші та людини дозволить точно вмонтувати гени людини в синтенні ділянки геному миші й одержати «олюднену» за певними ознаками мишу. Особливий інтерес такі миші становлять для фармацевтичної промисловості. Так, дослідження геному миші показало експан-

сію в неї чотирьох підродин генів цитохрому P450. Ці гени відповідають за ферменти, які активно беруть участь у метаболізмі ксенобіотиків, у тому числі і лікарських засобів. Генетично модифікована за генами цитохромів миша може стати точною моделлю людини в процесі дослідження біотрансформації лікарських препаратів.

Із завершенням у найближчому майбутньому секвенування геному ще однієї лабораторної тварини — щура — моделювання хвороб людини буде набагато достовірнішим. Плановане секвенування геномів шимпанзе, собаки, корови значно збільшить наші знання про геном ссавців і, відповідно, людини. Досягнення в галузі молекулярної біології на рубежі XX і XXI ст. сприяло виникненню молекулярної медицини — нового напрямку в природознавстві. Це, у свою чергу, ініціювало розвиток прогресивних технологій, які удосконалювали старі й створювали нові методи. Переважна більшість цих методів автоматизовані і поєднанні з комп'ютерними технологіями. Наочними прикладами цього є автоматичні підходи секвенування геному людини і наступний етап — дослідження людського протеому. Нові технології дають змогу досить швидко одержати колосальний обсяг інформації про структуру дотепер невідомих білків. Подальше виявлення та вивчення їхніх функцій зробить можливим досконале дослідження клітинних процесів, механізмів індивідуального розвитку, еволюції живого світу.

Ідентифікація нових білків зміцнить інтеграцію біології та фундаментальної медицини, що приведе до відкриття нових діагностичних маркерів, виявлення білків-мішеней для фармакологічних препаратів. Великі надії покладаються на вивчення протеому пухлин, що дозволить поліпшити їхню діагностику і особливо лікування.

