

Lidský genom v jinošských letech

10 let se znalostí (ne)kompletního lidského genomu

**MARIAN
NOVOTNÝ**

Když 26. června 2000 předstoupil před novináře v Bílém domě americký prezident Bill Clinton, neměl v úmyslu vyhlásit válku ani prosadit další reformu. Oznamoval „dokončení“ jednoho vědeckého projektu – mapování lidského genomu. Přítomnost amerického prezidenta dávala tušit, že jde o projekt v mnoha ohledech přelomový. Projekt, který mění biologii i medicínu a přináší nové právní i etické otázky. Lidský genom byl představen jako kniha, kterou můžeme díky kompletnímu osekvenování konečně otevřít a přečíst si, co odlišuje člověka od jiných organismů, co odlišuje lidi mezi sebou a možná i to, na co kdo zemře. Bill Clinton ve své řeči mimo jiné předvídal, že znalost lidského genomu přinese revoluci v diagnostice, prevenci a léčbě většiny (ne-li všech) lidských nemocí. Co nového jsme se však skutečně naučili za 10 let od památného vystoupení amerického prezidenta a jak velká část jeho předpovědi se naplnila?

Stojí za připomenutí, že Bill Clinton označoval před deseti lety ukončení „hrubého“ sekvenování lidského genomu. Co to znamená? Lidský genom představuje kompletní genetickou informaci uloženou v pořadí (sekvenci) nukleotidů deoxyribonukleové kyseliny (DNA)¹. Při sekvenování (mapování) je určeno pořadí nukleotidů v DNA. Část genomu (výrazně menší část) kóduje proteiny a RNA, které zajišťují životní funkce buněk. Většina genomu ovšem nekóduje proteiny ani RNA, a dokonce ani nic jiného. Význam nekódujících oblastí zůstával v době dokončení hrubého genomu zahalen tajemstvím. Některými autory byly nekódující oblasti označovány nehezky anglickým slovem „junk“² a vůbec považovány za zbytečné.

Když Bill Clinton mluvil o „ukončení“ mapování lidského genomu, měl ve skutečnosti na mysli, že bylo rozluštěno pořadí nukleotidů u asi 90 % oblastí, které kódují proteiny a RNA, a celkově asi u 83 % celého genomu. Skutečné dokončení „kompletního“ genomu bylo pak opět s velkou slávou oznámeno v roce 2003 a publikováno ještě o rok později. Nicméně ani dnes není lidský genom rozluštěn úplně celý. Odhaduje se, že je známo asi 93 % genomu. Dostupným sekvenáčnickým technikám stále částečně odolávají vysoce repetitivní oblasti³ – centromery⁴ a telomery⁵. Kódující oblasti by však měly být známy už víceméně kompletně. Jisté pochybnosti stále přetrvávají pouze o konkrétní podobě některých multigenových rodin,⁶ které se podílejí především na imunitní odpovědi organismu. Je tedy otázkou, kdy

by se mělo 10. výročí dokončení mapování lidského genomu slavit.

Jedinečný lidský genom?

Lidský genom je zapsán do 3,2 miliardy nukleotidů, což je podobná velikost genomu jako u myši. Nejde o žádný rekord – sladkovodní měňavka *Amoeba dubia* má genom 200krát větší. Velikost genomu tedy není přímo úměrná komplexnosti organismu, ale to se vědělo dříve, než bylo sekvenování člověka zahájeno. Obrovské části genomu *Amoeba dubia* vyplňují repetice DNA, které nekódují proteiny ani RNA. Jedním z překvapení, které sekvenování přineslo, bylo poznání, že lidský genom obsahuje výrazně méně kódujících oblastí (tzv. genů) než původně odhadovaných 100 000. Odhady počtu lidských genů na základě hrubého genomu se pohybovaly mezi 30 000 a 40 000 geny. Dnes víme, že protein-kódujících genů je jen něco málo přes 20 000. Počet RNA-kódujících genů není přesně znám, ale odhaduje se na několik (málo) tisíc. Pro srovnání kvasinka *Saccharomyces cerevisiae* má asi třetinový počet genů, ale jiný známý modelový organismus – hlístice *Caenorhabditis elegans* disponuje prakticky stejným počtem genů jako člověk a rostlina huseníček *Arabidopsis thaliana* nebo parazit bičienka *Trichomonas vaginalis* mají genů dokonce více než člověk. Sekvenování nám ukázalo, že člověk nemá výrazně větší počet genů než jiné organismy, a že tudíž prostým počtem genů nelze vysvětlit, proč člověk staví katedrály a létá do vesmíru, zatímco se hlístice prodírá lesní hrabankou. Dosud nebyla v genomu uspokojivá odpověď na tuto záhadu nalezena. Ukazuje se, že lidské geny by mohly mít více variant a modifikací než geny jiných organismů, a proto by počet proteinů, které z nich mohou vzniknout, mohl být ve skutečnosti vyšší. Jak velkou úlohu tyto varianty hrají, zatím není jasné. Proto ani nelze říct, že vyšší počet proteinů je tím zásadním rozdílem mezi člověkem a ostatními organismy. Rozdílem, který nám umožňuje skládat hudbu a komunikovat na obrovské vzdálenosti.

Jak moc se liší dva lidé?

Ne každý člověk je však schopen skládat hudbu, někdo má obličej kulatý, někdo protáhlý, někdo hází daleko oštěpem, jiný radši vaří – lidé jsou velmi rozdílní. Jak moc se tyto rozdíly projevují na úrovni DNA? Z výsledků sekvenování vyplývá, že pokud se sledují bodové

1) Nukleotidy jsou základní stavební kameny DNA. DNA je složena ze čtyř různých nukleotidů. Genetická informace je zakódována v pořadí nukleotidů. Nukleotidy si lze představit jako písmena v knize. Chceme-li knize porozumět, musíme vědět nejenom kolik písmen v knize je, ale i v jakém pořadí se v knize vyskytují.

2) Anglické „junk“ znamená neúčinný, bezvýznamný nebo taky odpad či brak.

3) Repetitivní oblasti – oblasti DNA složené z opakujících se motivů v pořadí nukleotidů (např. dlouhé úseky DNA obsahující pouze opakující se nukleotid adenin).

4) Centromera – oblast (obvykle uprostřed) chromozomu, kde se setkávají dvě sesterské chromatidy.

5) Telomery – koncové oblasti chromozomu. Telomery jsou složeny především z repetitivních úseků.

6) Multigenové rodiny – rodiny blízké příbuzných genů, které vznikly duplikací původního genu.

7) Bodová mutace – záměna jednoho nukleotidu za jiný (např. záměna adeninu za thymin).

8) Delece, inserce, duplikace – vyjmutí, vložení, zdvojení úseku DNA.

9) Crohnova choroba – závažné onemocnění, které postihuje především trávicí trakt. Nemoc je pravděpodobně vyvolána vlastním imunitním systémem nemocných.

mutace⁷ nukleotidů, tak se dva lidé v průměru liší asi v jednom z 1300 nukleotidů, což činí asi 2,5 milionu rozdílů mezi dvěma lidmi. Většina těchto rozdílů je však lokalizována v oblastech, které nekódují proteiny ani RNA. Navíc vzhledem k mechanismu přepisu genetické informace do proteinů se nijak neprojevuje ani řada bodových mutací v genech.

Kromě bodových mutací však dochází v genomu i k delecím, inzercím, duplikacím⁸ a jiným úpravám, které mohou mít mnohem větší dopad než bodové mutace. Mohou například vést ke změně počtu kopií genu v genomu. Nejnovější výzkumy ukazují, že se tyto změny objevují mnohem častěji, než se původně očekávalo. Po započtení těchto modifikací se odhaduje, že dva lidé v průměru sdílejí 99,5 % genetické informace. Je však 0,5 % dost na vysvětlení všech rozdílů mezi lidmi?

„Junk“ DNA

Vědci v poslední době věnují čím dál tím větší pozornost studiu nekódujících oblastí lidského genomu. Např. projekt ENCODE si klade za cíl co nejdetailněji popsat alespoň část lidského genomu za pomoci nejmodernějších metod. Zatím byla publikována analýza 1 % genomu, části, která zahrnovala kódující i nekódující úseky. Výsledky byly poměrně překvapivé. Doposud se předpokládalo, že k transkripci (přepisu do RNA) dochází téměř výhradně jen u kódujících oblastí genomu, z výsledků projektu ENCODE však vyplývá, že do RNA je přepisováno nejméně 74 % studované oblasti. Toto pozorování bylo upřesněno v letošním roce, kdy tým vědců z univerzity v Torontu ukázal, že k transkripci mimo geny dochází především v regulačních oblastech genů a mnohem méně v oblastech bez známých genů. Dalším pozoruhodným nálezem, vyplývajícím z projektu ENCODE je, že k transkripci obvykle dochází z obou řetězců DNA a ne jenom z řetězce, který kóduje gen. Detailní popis také ukázal, že regulační oblasti genů, které především ovlivňují kdy a v jakém množství bude gen přepisován, jsou rovnoměrně zastoupeny před i za geny. Tím opět naboural dřívější představy, totiž že regulační oblasti se vyskytují především před geny. Všechny tyto nálezy naznačují, že by „junk“ DNA mohla hrát významnou úlohu například v regulaci kódujících oblastí. Další důkaz přinesla studie, v níž byly srovnávány bodové mutace u 3000 zdravých jedinců a 2000 lidí trpících závažnou Crohnovou chorobou.⁹ Kromě mutací v genech (např. autofágní dráhy, která degraduje vlastní buněčné komponenty), byl u pacientů zaznamenán i zvýšený výskyt dvou mutací, které se vyskytují v tzv. genové poušti, v rozsáhlé oblasti bez jakýchkoliv genů. V letošním roce bylo také ukázáno, že další nekódující oblasti, transpo-

zóny se vyskytují častěji v buňkách nádorů plic než ve zdravých lidských buňkách. Zdá se tedy, že se nekódující oblasti mohou podílet i na vzniku velmi závažných onemocnění.

Genom a medicína

V oblasti medicíny se již objevují první náznaky využití znalosti lidského genomu. V klinické praxi se například využívá test na přítomnost funkční thioipurin methyltransferázy, která odbourává thioipuriny, používané při léčbě leukémie. U pacientů s nefunkční thioipurin methyltransferázou je následně indikována jiná léčba. Obdobně existuje test na zjištění variant cytochromu P450, který se podílí na degradaci velkého množství léků. Na základě výsledků testu lze upravit výběr léků tak, aby je pacient dobře snášel.

Objevily se už i první kompletní genomy nádorových buněk (např. akutní myeloidní leukémie). V rámci těchto studií byly objevené nové mutace a vytipovány nové slibné terapeutické cíle. S rozvojem sekvenačních technik je navíc možné sledovat i různá stadia nádorových buněk a porovnávat, k jakým mutacím dochází při přerodu buňky nemetastazující v buňku metastazující.

V posledních pěti letech jsme svědky skutečně bouřlivého rozvoje sekvenačních technik a prudkého poklesu ceny sekvenování. Zatímco první lidský genom stál 3 milardy dolarů, dnes se cena odhaduje na desítky tisíc dolarů. První genom byl složen z více jedinců různých ras, proto nikdy fotku prvního osekvenovaného člověka neuvídíte, ale dnes už několik lidí svůj genom zná. Nejznámějším člověkem s kompletně přečteným genomem je asi James D. Watson, spoluobjevitel struktury DNA. Dalším známým je Craig Venter, zakladatel společnosti Celera, které patří dělené první místo v závodu, kdo první dokončí mapování lidského genomu. Osekvenování byli už i dva Asiaté a pravděpodobně není daleko ani doba, kdy každý bude znát svůj genom. Sekvenační technologie, které se zatím stále jen vyvíjejí (především nanoporové sekvenování), mají potenciál zvládnout sekvenování genomu za několik hodin a cenu okolo 1000 dolarů.

Aby nám však byla znalost genomu užitečná, musíme se naučit v této knize pořádně číst. U velkého množství genů ještě stále není známa jejich funkce a nekódující oblasti jsou prakticky zcela neprobádané. Již zmiňovaný Craig Venter nedávno v rozhovoru neskrýval zklamání, že se ze svého genomu nedozvěděl ani vlastní barvu očí.

Přesto nám znalost genomu již mnohé přinesla – podnítila prudký rozvoj sekvenačních technik, obrátila pozornost i na nekódující oblasti a pomalu proniká také do medicíny. Další využití lze jistě očekávat v blízké budoucnosti. Asi se však nelze domnívat, že nám znalost genomu pomůže vyléčit většinu chorob. Uvidíme, co budou psát o genomu u příležitosti 20. výročí ukončení mapování. Třeba bude tou dobou známa i 3D struktura lidského genomu, která by odhalila mnohé o interakcích uvnitř chromozómů.

<http://www.spiegel.de/international/world/0,1518,709174,-00.html>

<http://clinton5.nara.gov/WH/New/html/genome-20000-626.html> http://expasy.org/sprot/hpi/hpi_stat.html

International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 409, 860–921, 2001.

Venter et al.: The Sequence of the Human Genome. *Science* 291(5507), 1304–1351, 2001.

International Human Genome Sequencing Consortium: Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945, 2004.

ENCODE Project Consortium: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816, 2007.

Ley et al.: DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72, 2008.

Wellcome Trust Case-Control Consortium: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464, 713–720, 2010.

van Bakel et al.: Most “Dark Matter” Transcripts Are Associated With Known Genes. *PLOS Biology* 8(5), e1000371, 2010.

Mgr. Marian Novotný, PhD. (*1979), viz *Vesmír* 89, 444, 2010/7.

Abstract: Human genome in its adolescence by Marian Novotný. The sequencing of human genome was finished 10 years ago. It was welcomed with tremendous expectations. But have these expectations been met? A decade of human genome studies has brought a number of surprises and first applications in diagnosis and treatment of illnesses. However, the detailed understanding of the information deposited in the genome sequence has not been reached.