

Pregledni znanstveni članek ■

K orodjem bioinformatike za fenomiko in sistemsko biologijo

Towards the bioinformatics tools for phenomics and systems biology

Uroš Petrovič, Mojca Mattiazzi, Tomaž Curk, Blaž Zupan, Igor Križaj

Izvleček. Sistemsko biologija je veda, katere cilj je razumevanje bioloških procesov na sistemski ravni, z upoštevanjem kompleksnih interakcij med geni, proteini in drugimi elementi celice. Področje post-genomske biologije, ki se ukvarja z vplivom celotnega genoma na lastnosti celice, kar je nujen korak na poti k sistemski biologiji, se imenuje fenomika. Zaradi tehničnih omejitev pri genetski manipulaciji človeških celic je zaenkrat pri raziskavah na področju fenomike nujna uporaba modelnih organizmov. Ena glavnih omejitev za razmah systemske biologije je pomanjkanje ustreznih orodij bioinformatike, katerih razvoj zato poteka vzporedno z razvojem novih eksperimentalnih pristopov pri modelnih organizmih, s končnim ciljem aplikacije na biologijo človeka.

Abstract. The aim of systems biology is systems-level understanding of biological processes that takes into account complex interactions of genes, proteins and other cell elements. The area of post-genomic biology that deals with the effect of the whole genome on the cell characteristics – a necessary step towards systems biology – is called phenomics. Because of technical limitations in genetic engineering of human cells, the use of model organisms is currently inevitable in phenomics studies. One of the main limitations for the advancement of systems biology is the lack of appropriate bioinformatics tools. The development of these therefore takes place in parallel with the development of new experimental approaches in model organisms, with the ultimate goal to apply them also to human biology.

Institucije avtorjev: Institut Jožef Stefan, Ljubljana, Slovenija (UP, MM, IK), Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Slovenija (TC, BZ), Baylor College of Medicine, Houston, USA (BZ).

Kontaktna oseba: Uroš Petrovič, Odsek za biokemijo in molekularno biologijo, Institut Jožef Stefan, Jamova 39, 1000 Ljubljana. email: uros.petrovic@ijs.si.

■ **Infor Med Slov:** 2006; 11(1): 66-71

Uvod

Eno ključnih vprašanj biologije je, kako genotip vpliva na fenotip. Klasična genetika je pri iskanju odgovorov na to vprašanje omejena na preučevanje relativno majhnega števila genov, medtem ko je fenotip vedno rezultat delovanja vseh genov, torej genoma. Za razumevanje vpliva celotnega in ne le delnega genotipa na fenotip je potrebno analizirati vplive čim večjega števila genotipov posameznega genoma na čim večje število merljivih in praviloma kvantitativnih lastnosti fenotipa.

Fenomika in modelni organizmi

Za doseg tega cilja je nujna uporaba modelnih organizmov, ki omogočajo uporabo orodij za natančno poseganje v genom. Na njih poteka razvoj pristopov, ki bodo v prihodnosti uporabni tudi za študij sistemske biologije v humani medicini; celovit pristop za študij fenomike pri modelnih organizmih je zatorej načrtovan tako, da bo lahko v čim večji meri prenesen na preučevanje človeka. Najdlje je razvoj tehnik in orodij, potrebnih za globalno analizo fenotipa, fenomiko, prišel pri preučevanju kvasovke *Saccharomyces cerevisiae*. Od določitve zaporedja celotnega genoma kvasovke leta 1996¹ so si sledili:

- izdelava prve DNA mikromreže, ki je vsebovala celoten genom² in omogoča globalno transkriptomsko analizo,
- proteinske mikromreže s celotnim proteomom,³ ki omogočajo določitev fizičnih interakcij endogenih ali eksogenih proteinov, lipidnih molekul in drugih nizkomolekularnih molekul z vsemi kvasnimi proteini,
- priprava zbirke sevov s sistematično izbitimi vsemi neesencialnimi geni,⁴ kar je omogočilo analizo fenotipa vseh mutant s posameznimi delecijami,

- zbirka sevov z vsemi geni s fuzijo z genom za zeleni fluorescirajoči protein,⁵ ki omogoča lokalizacijo vseh proteinov v celici,
- zbirka sevov z možnostjo uravnavanega izražanja vseh kvasnih genov,⁶ kar je omogočilo analizo fenotipa vseh mutant s prekomernim izražanjem posameznih genov, in
- zbirka sevov z vsemi geni s fuzijo z označevalcem "TAP",⁷ ki je omogočila identifikacijo večine proteinskih kompleksov v celici kvasovke.

Razvoj teh tehnik je napravil iz kvasovke trenutno najbolj primeren modelni organizem za sistemsko biologijo, saj poleg analize fenotipa pri vseh možnih "preprostih" genotipih (hipomorfne mutacije posameznih genov oziroma prekomerno izražanje posameznih genov) lahko dajo podatke tudi o drugi pomembni komponenti sistema, to je o medsebojni povezanosti njegovih gradnikov (proteinske in genetske interakcije, koordinirano izražanje genov). Poleg kvasovke se v zadnjem času kot modelni organizmi uveljavljajo tudi bolj kompleksni organizmi, na primer nematod *Caenorhabditis elegans*.⁸

Uporaba sistemske biologije v medicini bo predvidoma omogočala hiter razvoj novih, natančno usmerjenih zdravil in razvoj posamezniku prilagojene medicine. Za doseg tega cilja bo potrebno razumeti procese v človeških celicah in organizmu na podobni ravni, kot danes razumemo celice kvasovk. Kot primer vzemimo cistično fibrozo, ki je ena najbolj znanih tako imenovanih monogenetskih bolezni. Nastanek cistične fibroze povzroči mutacija v enem genu, imenovanem *CFTR*.⁹ Vendar pa samo iz mesta mutacije v *CFTR* ne moremo sklepati na točen potek bolezni, v kolikor ne upoštevamo tudi aktivnosti drugih genov/proteinov. Na primer, za nastanek bolezenskih simptomov na osnovi mutacije v *CFTR* je nujno potrebna aktivnost proteinov Hsp70. Tako lahko znižana aktivnost Hsp70, bodisi zaradi mutacije bodisi zaradi inhibitorjev, blaži simptome cistične fibroze. Znano je, da lahko kot inhibitorji Hsp70 delujejo

butanojska in druge kratkoverižne maščobne kisline, ki so prisotne v fizioloških razmerah v celicah, kar pomeni, da na simptome cistične fibroze posredno vpliva tudi metabolizem lipidov. Takšnih primerov je v fiziologiji človeških celic še veliko in predstavljajo, zaradi medsebojne povezanosti celičnih procesov, pravilo in ne izjeme.

Bioinformatični pristopi in orodja

Za razumevanje medsebojne povezanosti genov oziroma proteinov v celici ali organizmu je potrebno odgovoriti na nekaj osnovnih vprašanj. Za vse gene v genomu moramo najprej poznati njihovo funkcijo ("Kaj gen/protein počne?"), pri tem pa se moramo zavedati, da je večina genov/proteinov udeleženi v več kot samo enem procesu ("Kaj vse gen/protein počne?"). Naslednja stopnja razumevanja celice kot sistema zahteva poznavanje mehanizma delovanja ("Kako to počne?") ter končno poznavanje vseh genetskih in proteinskih interakcij, ki nastopajo v celičnih procesih ("S kom geni/proteini sodelujejo in kako?"). Odgovore na ta v bistvu zelo preprosta vprašanja lahko na ravni celotnega genoma/proteoma da le več različnih eksperimentalnih tehnik, katerih združena interpretacija šele lahko predstavi celotno sliko. Orodja bioinformatike za doseg tovrstne interpretacije trenutno še niso razvita v zadostni meri.

Orodja bioinformatike so nepogrešljiva pri prevajanju genoma z metodami funkcijske genomike na raven sistemske biologije in fenomike. Pri analizi fenotipov dvojnih mutant se na primer izkaže, da je število možnih genetskih mrež že pri manj kot desetih genih tako veliko, da odkrivanje mrež iz podatkov zahteva računalniško obdelavo in razvoj formaliziranih postopkov.¹⁰ Pri fenomiki pa imamo opraviti na primer z analizo fenotipov vseh enojnih in dvojnih mutant, kar v primeru preprostega organizma kot je kvasovka pomeni približno 6.000 enojnih in $6.000 \times 5.999 /$

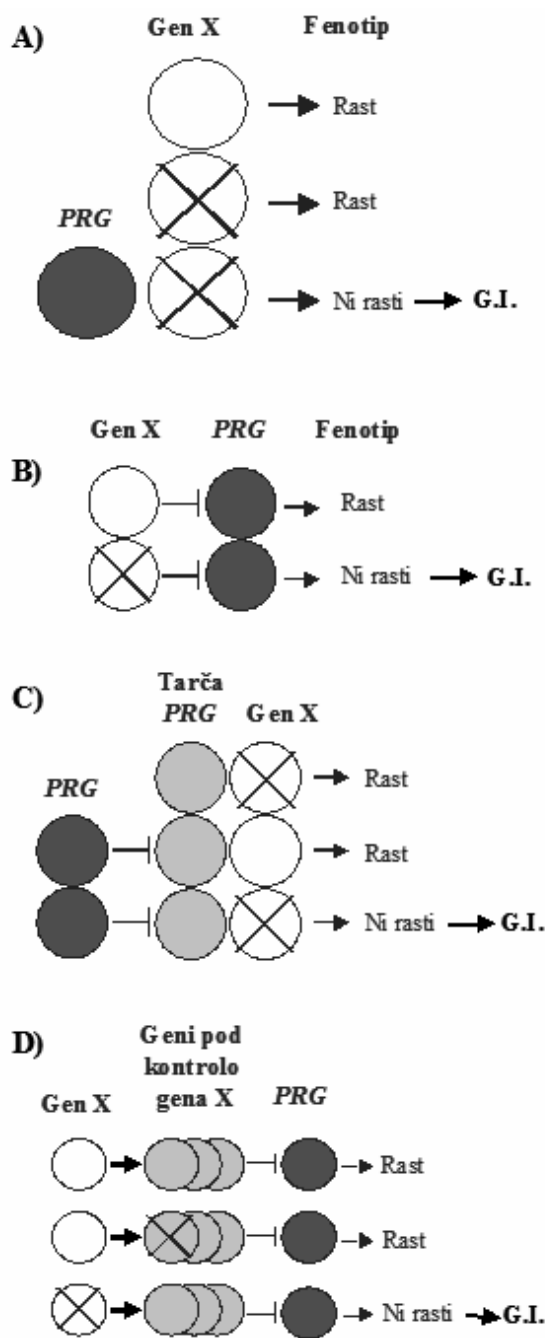
$2 = 17.997.000$ dvojnih mutant. Zato so potrebna nova orodja, ki so sposobna analize takšne količine podatkov in ki so hkrati sposobna integracije dodatnih podatkov o mutantih za reševanje konfliktov, do katerih neizogibno pride zaradi kopičenja napak pri eksperimentalnih pristopih na ravni celotnega genoma oziroma proteoma. Dodatni podatki so lahko raznovrstni, pomembno je le, da so relativno zanesljivi (na primer iz objavljenih ciljanih študij).

Za ponazoritev lastnosti orodij bioinformatike, ki so potrebna za analizo na ravni fenomike in sistemske biologije, vzemimo primer nadzorovanega oziroma prekomernega izražanja preiskovanega gena v zbirki sevov s sistematično izbitimi vsemi neesencialnimi geni. Nedavno je bila razvita tehnika, ki omogoča sistematično uvajanje dvojnih mutacij v genom kvasovke, z namenom določitve genetskih interakcij.¹¹ Genetska interakcija je definirana kot interakcija med dvema genoma, kjer ima mutacija v obeh genih za posledico fenotipsko lastnost, ki se ne pojavlja pri nobeni od posameznih enojnih mutacij. V opisanem primeru dobimo dvojne mutante, kjer je posledica ene mutacije prekomerno izražanje preiskovanega gena, posledica druge mutacije pa odsotnost izražanja drugega gena. Takšne dvojne mutante so uporaben model za določitev molekulske osnove delovanja preiskovanega gena v celici kvasovke kot modelu.¹²

Orodja bioinformatike za tovrstno analizo delovanja prekomerno izraženih genov še niso razvita, pričujoči primer pa nakazuje eno od smeri razvoja algoritmov, ki bodo predvidoma uporabljeni v teh orodjih. Z analizo ene same fenotipske lastnosti, zmanjšane hitrosti rasti, lahko določimo nabor genov, ki so v genetski interakciji s preiskovanim genom (slika 1A). Možnih razlag, zakaj obstaja genetska interakcija med preiskovanim genom in nekim drugim genom v genomu kvasovke, je več (slike 1B-D). Gen, ki je v genetski interakciji s preiskovanim genom, lahko inhibira delovanje preiskovanega gena, ki potencialno negativno vpliva na rast celice¹² (slika 1B); gen je lahko funkcijski homolog gena, ki je

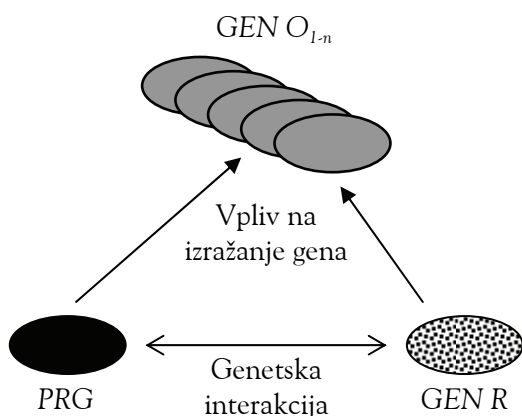
tarča neposrednega inhibitornega delovanja preiskovanega gena (slika 1C); gen je lahko aktivator skupine genov, ki inhibira delovanje preiskovanega gena, ki sicer negativno vpliva na rast celice (slika 1D). Že če upoštevamo samo eno dodatno fenotipsko lastnost, na primer povečanje hitrosti rasti kot posledico genetske interakcije, se število možnih mehanizmov delovanja preiskovanega gena bistveno poveča. Da bi ugotovili, kateri mehanizem je v danem primeru ustrezen, je potrebno poznati še druge lastnosti fenotipa kot samo hitrost rasti. Kadar v literaturi ni na voljo dovolj podatkov o določenem genu, je najhitrejša pot za določitev kompleksnega fenotipa, povezanega s preučevanim genom, določitev vpliva mutacije tega gena na transkriptom, saj ta način zahteva v teoriji le en eksperiment z uporabo DNA mikromreže s celotnim genomom. Na profil izražanja genoma mutiranega seva lahko gledamo kot na poseben fenotip, iz katerega je moč razbrati dinamično komponento genoma povezano s preučevanim genom. Podobno lahko postopamo tudi tedaj, ko želimo analizirati fenotipski odziv na izražanje eksogenega gena.

Kot primer določitve molekulske osnove delovanja eksogenega gena v kvasovki smo uporabili amoditoksin, večfunkcijsko fosfolipazo A₂, ki ima na sesalcih raznovrstne patofiziološke učinke, med njimi tudi nevrotoksičnega. Najprej smo določili vpliv nadzorovanega izražanja amoditoksina v celici kvasovke na izražanje celotnega transkriptoma, s čimer smo dobili kompleksno sliko fenotipa. Spremembe v izražanju smo zasledili predvsem pri genih, ki so neposredno udeleženi v odzivu celice na izražanje amoditoksina. Nato smo gen za amoditoksin s križanjem prenesli in nato nadzorovano izrazili v zbirki sevov z izbitimi vsemi posamičnimi neesencialnimi geni ter analizirali rast tako dobljenih dvojnih mutant.



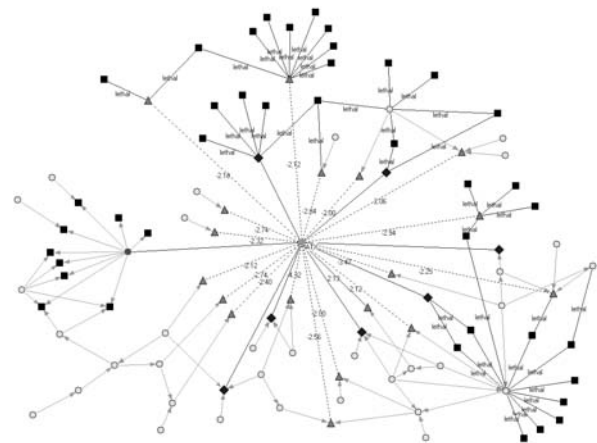
Slika 1 Genetske interakcije. A: Osnovna definicija genetske interakcije med prekomerno izraženim preiskovanim genom in kvasnim genom (Gen X). B-D: Nekatere možne relacije med geni, ki privedejo do genetske interakcije (za podrobnosti glej tekst). Okrajšave: PRG – preiskovani gen; G.I. – genetska interakcija.

Počasnejša rast dvojne mutante kot posameznih enojnih mutant (genetska interakcija) kaže na funkcijsko povezavo med preiskovanim eksogenim genom in genom, ki je mutiran.¹¹⁻¹³ S to metodo smo identificirali predvsem regulatorne gene, ki posredno ali neposredno uravnavajo izražanje genov, identificiranih z zgoraj opisano analizo vpliva izražanja amoditoksina na transkriptom (slika 2). Naši eksperimenti so torej pokazali, da dobimo z določitvijo vseh genetskih interakcij preiskovanega gena ter vseh genov, ki se jim raven izražanja značilno spremeni kot posledica prekomernega izražanja preiskovanega gena, komplementarne rezultate. Za analizo tovrstnih rezultatov so potrebna nova orodja bioinformatike, ki kombinirajo analizo transkriptomskih podatkov z analizo mutant ter vključujejo tudi vse ostale razpoložljive podatke, kar pa se lahko hitro sprevrže v zelo kompleksen kombinatorični problem. Zato je nujno potrebno razviti in implementirati različne hevrstične in optimizacijske pristope k reševanju tovrstnih kompleksnih problemov. Drugi pristop k obvladovanju kompleksnosti je večnivojsko in postopno odkrivanje in opisovanje rezultatov, od splošnega, kjer pokažemo le najbolj očitne lastnosti, do zelo podrobnega nivoja.¹⁴



Slika 2 Shematski prikaz tipičnega rezultata, dobljenega s kombinacijo identifikacije genetskih interakcij ter vpliva preiskovanega gena na transkriptom. Okrajšave: PRG – preiskovani gen; GEN R – regulatorni gen; GEN O_{l-n} – geni, ki so pod kontrolo regulatornega gena GEN R.

Nazorni prikaz dobljenih rezultatov, denimo v obliki genetskih mrež ali podobnih oblik hkratnega prikazovanja večje množice odkritih relacij med geni oziroma opazovanimi funkcionalni deli, je lahko odločilnega pomena za uspešno interpretacijo rezultatov in načrtovanje dodatnih analiz, za kar je spet potrebno razviti dodatna orodja. Pomembno orodje za vizualizacijo kompleksnih mrež je programski paket Pajek¹⁵ (slika 3).



Slika 3 Mreža interakcij med amoditoksinom (v sredini grafa) in kvasnimi geni ter proteini, prikazana s programom Pajek. Različne oblike prikazujejo anotacije genov/proteinov: karo – gen v genetski interakciji z amoditoksinom; trikotnik – gen, na katerega izražanje vpliva amoditoksin; krog – transkripcijski faktor; kvadrat – ostali geni/proteini. Puščice prikazujejo vpliv transkripcijskih faktorjev na izražanje genov, polne črte genetske interakcije in prekinjene črte vpliv amoditoksina na izražanje genov. Številke označujejo kvantitativen vpliv amoditoksina na izražanje genov, beseda "lethal" pa podtip genetske interakcije, kjer je inhibicija rasti dvojne mutante popolna.

Pri razvoju orodij bioinformatike moramo še posebej stremeti k njihovi interaktivnosti, ki omogoča uporabniku, da enostavno in hitro preveri različne hipoteze. Čeprav so mnogokrat osnova numerični, kvantitativni podatki, je potrebno razviti in implementirati metode, ki generirajo simbolne in zato potencialno bolj

razumljive modele. Primer takih kvalitativnih modelov so tudi genske mreže.

Zaključek

V današnji poplavi javno dostopnih genetskih podatkov na spletu morajo orodja bioinformatike ob navajanju rezultatov ponuditi povezave do relevantnih podatkov in objavljenih izsledkov, ki jih lahko raziskovalec uporabi za dodatno podkrepitev ali ovržbo dobljenih rezultatov in tako bistveno pospeši proces odkrivanja novih zakonitosti.

Zadnje in morda najbolj pomembno področje, na katero lahko posežejo orodja bioinformatike, pa je razvoj pristopov, ki pomagajo genetiki pri načrtovanju in izvajanju novih eksperimentov. Čeprav je razvoj tovrstnih tehnik za podporo raziskavam v sistemski biologiji še v povojih, nekaj nedavnih objav na tem področju priča, da je z uporabo pristopov umetne inteligence moč vsaj do določene mere avtomatizirati načrtovanje bioloških eksperimentov za potrebe odkrivanja novih bioloških znanj.^{16,17}

Orodja bioinformatike na področju fenomike bodo omogočila razmeroma hitro in natančno določitev molekulske funkcije človeških genov na genomski ravni, kakor hitro bo razvoj eksperimentalnih tehnik dovolj napredoval, da bo omogočal analize, primerljive z modelnimi organizmi. Zato predstavljajo takšna orodja pomemben korak k sistemski biologiji, tako pri modelnih organizmih kot pri človeku.

Literatura

- Goffeau A, Barrell BG, Bussey H, et al.: Life with 6000 genes. *Science* 1996; 274(5287): 546, 563-567.
- DeRisi JL, Iyer VR, Brown PO: Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; 278(5338): 680-686.
- Zhu H, Bilgin M, Bangham R, et al.: Global analysis of protein activities using proteome chips. *Science* 2001; 293(5537): 2101-2105.
- Winzeler EA, Shoemaker DD, Astromoff A, et al.: Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999; 285(5429): 901-906.
- Huh WK, Falvo JV, Gerke LC, et al.: Global analysis of protein localization in budding yeast. *Nature* 2003; 425(6959): 686-691.
- Sopko R, Huang D, Preston N, et al.: Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* 2006; 21(3):319-330.
- Gavin AC, Aloy P, Grandi P, et al.: Proteome survey reveals modularity of the yeast cell machinery. *Nature* 2006; v tisku.
- Kamath RS, Fraser AG, Dong Y, et al.: Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 2003; 421(6920):231-237.
- Kerem B, Rommens JM, Buchanan JA, et al.: Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989; 245(4922):1073-1080.
- Zupan B, Demšar J, Bratko I, et al.: GenePath: a system for automated construction of genetic networks from mutant data. *Bioinformatics* 2003; 19(3): 383-389.
- Tong AH, Lesage G, Bader GD, et al.: Global mapping of the yeast genetic interaction network. *Science* 2004; 303(5659):808-813.
- Sopko R, Huang D, Preston N, et al.: Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* 2006; 21(3):319-330.
- Ye P, Peyser BD, Pan X, et al.: Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol Systems Biol* 2005; msb4100034-E1.
- Bornholdt S: Systems biology. Less is more in modeling large genetic networks. *Science* 2005; 310(5747): 449-451.
- Batagelj V, Mrvar A: Pajek: program for large network analysis. <http://vlado.fmf.uni-lj.si/pub/networks/pajek>, 2006.
- King RD, Whelan KE, Jones FM, et al.: Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 2004; 427(6971): 247-252.
- Zupan B, Bratko I, Demšar J, et al.: GenePath: a system for inference of genetic networks and proposal of genetic experiments. *Artif Intell Med* 2003; 29(1-2): 107-130.

Bilten SDMI ■

Zaključki kongresa "Zdravje na informacijski poti" (MI 2006), Zreče, 9.-11.4.2006

O kongresu

Slovensko društvo za medicinsko informatiko je med 9. in 11. aprilom 2006 v Zrečah organiziralo tradicionalni kongres medicinske informatike MI 2006 z naslovom "Zdravje na informacijski poti". Dogodek je privabil 177 udeležencev iz Slovenije in tujine, kar je največja udeležba na kongresih in strokovnih srečanjih društva doslej. Kongres je bil namenjen predstavitvi aktualnega dogajanja na področju zdravstvene informatike s poudarkom na uresničevanju nacionalne strategije eZdravje.

V okviru kongresa sta bili izvedeni delavnici na temo prenove zakona o zbirkah podatkov s področja zdravstva in na temo uporabe informacijskih tehnologij pri informiranju ter izobraževanju bolnikov in zdravstvenih delavcev.

Obravnavana so bila strateška področja zdravstvene informatike - standardizacija, baze podatkov, portali, poročanje, kartica zdravstvenega zavarovanja, komunikacije, elektronski zdravstveni zapisi, informacijska podpora čakalnim vrstam in upravljanju kakovosti v zdravstvu.

Predstavljene so bile nekatere izkušnje iz tujine (Danska, Velika Britanija, Srbija).

Pod pokroviteljstvom Ministrstva za zdravje je bila organizirana delavnica na temo uresničevanja

strategije eZdravje, v okviru katere so bile obravnavane načrtovane zagonske naloge - konsolidacija infrastrukture, vzpostavitev zdravstvenega portala, informacijska podpora čakalnim dobam. Obravnavani so bili tudi organizacijski in finančni aspekti uresničevanja.

Na kongresu so bili predstavljeni aktualni projekti, obravnavana je bila informatika v zdravstveni negi. Krajša delavnica je bila namenjena obravnavanju področja varnosti informacij v zdravstvu. Predstavljenih je bilo tudi nekaj rešitev iz najsodobnejših strokovnih področij (bioinformatika, inteligentni sistemi).



Slika 1 Sekcija Informatika v zdravstveni negi uspeva mnogo bolje, kot je to napovedoval urednik te revije.