

Framtiden for biomedisinsk forskning

– store endringer gir nye muligheter



Av Astrid Guldbrandsen,
ph.d.-student,
Institutt for biomedisin,
Universitetet i Bergen



Harald Barsnes,
forsker,
Institutt for biomedisin,
Institutt for informatikk,
Universitetet i Bergen



Marc Vaudel,
forsker,
Klinisk institutt 2,
Haukeland universitetssjukehus,
Universitetet i Bergen

Biomedisinsk forskning har de siste tiårene gjennomgått en rivende utvikling som i hovedsak kan spores tilbake til sekvenseringen av det humane genomet; det menneskelige arvematerialet. Prosjektet feiret i fjor 25-årsjubileum.¹ Mye har skjedd siden oppstarten, spesielt i løpet av de siste årene.

Da prosjektet publiserte det første utkastet for det humane genomet i 2001 hadde det tatt 13 år og kostet rundt USD 3 milliarder.^{2,3} I dag er samme oppgave nærmest blitt rutine og kan utføres på tre dager for under USD 1000.⁴ Dette, sammen med en stadig forbedret biomedisinsk forståelse, har gjort individualisert behandling mulig for en lang rekke både vanlige og uvanlige sykdommer. Men genomet er bare begynnelsen.

Det finnes en lang rekke andre tilsvarende biologiske enheter, slik som transkripter, proteiner, metabolitter og lipider, som også kan undersøkes for å ytterligere forbedre forståelsen av sykdommer og individuelle forskjeller mellom pasienter. Denne artikkelen fokuserer på proteiner, som fagfelt kalt proteomikk,⁵ for å framheve de store endringene den biomedisinske forskningen har gjennomgått siden årtusenskiftet. Vi

vil også vise hvilke nye muligheter dette allerede har resultert i, og diskutere det framtidige potensialet.

Fra gener til proteiner

Det menneskelige arvematerialet består av rundt 20-25.000 gener som koder for ulike proteiner. Det er derfor lett å tro at dersom vi kjenner alle genene, og hvilke proteiner de koder for, så trenger vi ikke å studere proteinene. Dette kan stemme for en del sykdommer. Som hovedregel for såkalte monogenetiske sykdommer, hvor en enkelt feil i et bestemt gen medfører at en pasient utvikler en gitt sykdom. Kjente eksempler på dette er cystisk fibrose og sigdcelleanemi, men også visse typer nevrologiske lidelser som arvelig spastisk paraparese og ataksi.

For en del slike sykdommer finnes det allerede gode verktøy for å oppdage, og i noen tilfeller også behandle, sykdommen. Det er derimot fortsatt store utfordringer knyttet til sykdommer som skyldes feil i flere gener samtidig, og som dermed ikke er like lette å oppdage om vi kun studerer en pasients genom. En lang rekke sykdommer vil også først bli synlige i løpet av/etter at genene er translatert til proteiner. Det er med andre ord ikke alltid sammenheng mel-

lom genotypen og fenotypen, og det er her fagfeltet proteomikk kommer inn.

Proteomikk kan defineres som storskalaanalyse av proteiner, med hovedvekt på identifikasjon og kvantifikasjon. Det innebærer også en mer detaljert forståelse av de ulike proteinenes egenskaper og deres rolle i en større biologisk sammenheng. En vesentlig endring de siste årene har vært at vi i hovedsak har gått fra å studere et mindre sett proteiner, til å identifisere og kvantifisere (tilnærmet) alle proteinene i en prøve. Dette skyldes hovedsakelig utviklingen av stadig bedre instrumenter og analysemetoder, som har gjort det mulig å observere mer av det såkalte proteomet, med andre ord samlingen av alle mulige proteiner.

Et annet viktig element som gjør proteomikk avgjørende for en forbedret sykdomsforståelse er de mange ulike former et protein kan forekomme i. Selv om vi ofte sier at ett gen koder for ett protein, så er ikke dette helt korrekt. En vesentlig reguleringsmekanisme for proteiner er såkalte posttranslasjonelle endringer. Dette er endringer som ikke er kodet i selve genet, men som kan påvirke et proteins rolle i cellen. For eksempel ved å aktivere eller deaktivere en gitt funksjon. Proteiner kan også endres på andre måter, for eksempel via strukturelle forandringer eller ved at deler av proteinet fjernes. De fleste proteiner er også avhengig av et samarbeid med andre proteiner for å utføre sin funksjon. Til sammen bidrar dette til at en pasients proteom er i stadig forandring, og det legger grunnlaget for den store variasjonen vi kan observere mellom pasienter og ulike målinger for samme pasient.

Behovet for bioinformatikk

En direkte konsekvens av bedre instrumenter og nye analysemetoder (både innen proteomikk og relaterte fagfelt) har vært at den gjennomsnittlige datamengden som genereres som del av et forskningsprosjekt har økt betraktelig. Dette har medført en dramatisk

endring for den biomedisinske forskningen, hvor vi nå ser at en stadig større andel av forskningstiden brukes foran datamaskinen for å forstå og tolke de genererte dataene. Noe som har resultert i nye utfordringer da de fleste biomedisinske forskere ikke nødvendigvis har bakgrunn fra dataanalyse eller vesentlig erfaring med bruk av bioinformatikkverktøy.

Samlet sett medfører dette et økende krav til de bioinformatiske miljøene for utvikling av programvare som kan håndtere slike data på en brukervennlige og forståelig måte, med det mål at forskerne selv skal kunne analysere sine egne data uten krav til avansert programmeringskompetanse. I tillegg har utviklingen av nye bioinformatiske og statistiske framgangsmåter vært avgjørende for å kunne håndtere de store datamengdene, og behovet vil neppe bli mindre i tiden framover.

Hvordan utnytte andres kunnskap og data

En annen viktig endringen innen biomedisin de siste årene er hvordan data og kunnskap gjøres tilgjengelig. Feltet har gjennomgått et vesentlig skifte fra å hovedsakelig dele forskningsresultater via vitenskapelig publikasjoner, til at det nå også er vanlig å dele dataene som ligger til grunn for resultatene. Den nye kunnskapen blir samlet og gjort tilgjengelig i åpne webbaserte databaser, en prosess ofte referert til som crowdsourcing.⁶

Innen proteomikk har man hatt store globale forskningsprosjekter knyttet til The Human Proteome Project,⁷ deling av rådata har blitt forenklet via ProteomeXchange,⁸ og vi har fått de første utkastene til det menneskelige proteomet.^{9,10} Dette har resultert i nye muligheter, hvor vi lettere kan basere ny forskning på eksisterende biologisk og medisinsk kunnskap.¹¹ Vi får dermed et bedre grunnlag for å planlegge nye forsøk, samt en forbedret mulighet til å tolke egne resultater i en større biomedisinsk kontekst. Et eksempel innen nevrologi er CSF proteome resource (CSF-PR),¹² en database som samler



og visualiserer kunnskap om hvilke proteiner som kan observeres i human cerebrospinalvæske. En utvidelse som også inkluderer kvantitative data fra pasienter med multippel sklerose, Alzheimers og Parkinsons er under utvikling.

Muligheten for å generere ny kunnskap uten nye eksperimentelle forsøk, men ved å heller reanalysere eksisterende data er et annet felt med stort framtidig potensial. En sammenkobling av kunnskap på tvers av fagfelt, for eksempel ved å se på både genomikk- og proteomikkdata for pasienter med en gitt sykdom, kan også gi ny innsikt. Dette krever at informasjonen fra de ulike fagfeltene blir gjort tilgjengelig på en måte slik at den lett kan integreres, noe som igjen bringer oss tilbake til behovet for bioinformatikk.

Til slutt bør det nevnes at fri deling av data og kunnskap også gjør det mulig for andre enn bare forskere å bidra. Et eksempel er Foldit (<https://fold.it>), hvor helt vanlige internettbrukere kan bidra til å løse utfordringer relatert til hvordan proteiner oppnår sin tredimensjonale struktur i en spillbasert kontekst. Dette har allerede resultert i både bedre undervisningsmetoder¹³ og ny kunnskap.^{14,15} ■

KONKLUSJON

En vesentlig utfordring videre blir å ha kunnskapen til å utnytte de store mulighetene som ligger i stadig mer og bedre data. Dette inkluderer at vi må kunne jobbe tverrfaglig og ofte på tvers av gamle skillelinjer mellom fagfelt og institutter. For å oppnå dette vil det være avgjørende at vi forbedrer de bioinformatiske verktøyene ved å gjøre disse enda mer interaktive og brukervennlige. Like viktig er det at vi fokuserer mer på undervisning i dataanalyse for biomedisinske studenter og forskere. Kun gjennom et slikt tverrfaglig samarbeid vil det bli mulig å nærme seg det enorme potensialet som ligger i biomedisinsk forskning.

Forfatterens forskning er støttet av: Norges forskningsråd, Bergen forskningsstiftelse og Stiftelsen Kristian Gerhard Jebsen

Referanser

1. Green ED, Watson JD, Collins FS. Human Genome Project: Twenty-five years of big biology. *Nature* 2015;526:29-31.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304-1351.
4. Hayden EC. Technology: The \$1,000 genome. *Nature* 2014;507:294-295.
5. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198-207.
6. Barsnes H, Martens L. Crowdsourcing in proteomics: public resources lead to better experiments. *Amino Acids* 2013;44:1129-1137.
7. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, et al. The human proteome project: current state and future direction. *Mol Cell Proteomics* 2011;10:111.
8. Juan A Vizcaino, Eric W Deutsch, Rui Wang, Attila Csordas, Florian Reisinger, Daniel Rios, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 2014;32:223-226.
9. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. *Nature* 2014;509:575-581.
10. Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509:582-587.
11. Vaudel M, Verheggen K, Csordas A, Raeder H, Berven FS, Martens L, et al. Exploring the potential of public proteomics data. *Proteomics* 2016;16(2):214-225.
12. Guldbrandsen A, Vetthe H, Farag Y, Oveland E, Garberg H, Berle M, et al. In-depth characterization of the cerebrospinal fluid proteome displayed through the CSF Proteome Resource (CSF-PR). *Mol Cell Proteomics* 2014;13(11):3152-3163.
13. Farley PC. Using the computer game "Foldit" to entice students to explore external representations of protein structure in a biochemistry course for nonmajors. *Biochem Mol Biol Educ* 2013;41:56-57.
14. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, Shen BW, et al. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat Biotechnol* 2012;30:190-192.
15. Khatib F, DiMaio F, Cooper S, Kazmierczyk M, Gilski M, Krzywda S, et al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol* 2011;18:1175-1177.