Brug af en translationel database i en klinisk afdeling

Bioinformatiker Anne Dirks Højfeldt, professor Hans E. Johnsen, statistiker Martin Bøgsted, post.doc. Alexander Schmitz, post.doc. Kirsten Fogd, post.doc. Linda Pilgaard, seniorforsker Mette Nyegaard, lektor Karen Dybkær & ledende overlæge Anne Bukh

STATUSARTIKEL

Århus Universitetshospital, Aalborg Sygehus, Hæmatologisk Afdeling, Medicinsk Center

"We can only see a short distance ahead, but we can see plenty there that needs to be done"

Alan Turing, 1950

I 1950 definerede computervidenskabens grundlægger *Alan Turing* den såkaldte Turingtest, som afslørede, at forskellen på en maskine og et menneske ligger i evnen til at træffe beslutninger i en sammenhæng, idet en maskine alene baserer sine beslutninger på objektiv information, der er indeholdt i en algoritme, mens et menneskes beslutninger også er baseret på subjektive vurderinger [1]. I daglig klinisk praksis influerer subjektive vurderinger på sikkerheden i diagnostikken – specielt når analyseresultater fra forskellige test skal integreres og kræver klinikerens subjektive ekspertvurdering. Den enkelte analyses resultat vurderes bedst i sammenhæng med andre betydende resultater og burde derfor standardiseres via maskinel, diagnostisk beslutningsstøtte.

Inden for det hæmatologiske speciale er diagnostik baseret på analyseresultater og vurderinger, som afhænger af faglig kompetence og erfaringer. Dette kan eksempelvis være patologens rutine i at stille en diagnose ud fra en undersøgelse af patientens væv i et lysmikroskop, billeddiagnostikerens referenceramme til at vurdere en skanning eller den enkelte klinikers erfaring i at bestille de korrekte analyser og tolke resultaterne. Usikkerheden i samspillet mellem de nuværende analytiske resultater og afhængigheden af klinikerens varierende grader af erfaring gør, at vi i dag må erkende nødvendigheden af at optimere udnyttelsen af objektive, diagnostiske data via etablering af »translationelle databaser«, der på standardiseret måde integrerer informationerne i beslutningsprocessen.

Nye teknologier gør det i dag muligt at analysere det samlede indhold af gener, celler eller proteiner i en vævsprøve eller vævs væske [2-4]. Denne type analyse resulterer i en såkaldt global profil – f.eks. en global genprofil, som indeholder information om op- og nedreguleringen af gener i en vævsprøve. Det forekommer nærliggende at opstille den hypotese, at integrative analyser (se faktaboks) af sådanne globale profiler vil bidrage med nye informationer om patogenesen af de hæmatologiske kræftsygdomme, og at disse informationer kan omsættes til objektive, maskinbaserede algoritmer med det mål: 1) at øge sikkerheden i den diagnostiske beslutningsproces, 2) at estimere prognosen samt 3) at forudsige behandlingseffekt og bivirkninger for den enkelte patient.

I denne statusartikel beskrives etablering og integrativ anvendelse af den translationelle database, der på standardiseret og kvalitetssikret måde opbevarer patientrelaterede data fra klinikken, biobanken og laboratoriet. Disse data forventes retrospektivt at kunne anvendes til opbygning af et dokumenteret erfaringsgrundlag, som har potentiale til udvikling af nye diagnostiske, prognostiske og prædiktive algoritmer, der efter prospektiv validering kan implementeres i klinikken.

DEN TRANSLATIONELLE DATABASES STRUKTUR OG BRUG

Ved udredningen af den hæmatologiske kræftpatient er det erfaringen, at symptomer og kliniske manifestationer ofte varierer meget – selv inden for samme diagnosegruppe. Denne heterogenitet fremgår bedst af den seneste WHO-klassifikation, der grupperer sygdommene på basis af kliniske, morfologiske, immunfænotypiske og genetiske informationer [5]. Den nyeste klassifikation udnytter til fulde de etablerede specialelaboratoriers kompetencer, men der fremkommer løbende overbevisende dokumentation for at klassifikationerne er utilstrækkelige. Der er endnu ikke er taget hensyn til ny viden fra studier baseret på f.eks. globale genprofiler, som har identificeret nye diagnostiske inddelinger, prognostiske og prædiktive biomarkører.

For at deltage i den faglige udvikling og understøtte den kliniske forskning i specialet har vi etableret den »Translationelle database ved Hæmatologisk Afdeling, Aalborg Sygehus«, som integrerer informationer fra den regionale kræftbiobank, de nationale kliniske databaser og forskningslaboratoriets data- og analyseportaler (**Figur 1**). Udgangspunktet er, at der til *den enkelte patient* kobles fragmenterede informationer, som løbende sammenkobles i én overordnet translationel forsknings- og udviklingsdatabase. Databasen skal også indeholde velkarakteriserede og publicerede data tilgængelig »online«, hvorved der umid-

delbart kan udføres valideringsstudier af observerede resultater på et eller flere matchede tilgængelige datasæt.

Data indeholdt i databasen kan udtrækkes og analyseres i en integrativ strategi med det mål at udvikle ny diagnostik og identificere nye biomarkører samt ikke mindst at forudsige behandlingseffekt og bivirkninger for den enkelte patient. Mange studier har vist, at integrative analyser af data fra globale profiler og kliniske forløb giver ny og mere detaljeret indsigt i kræftsygdommenes patogenetiske forhold. I litteraturen ses denne viden jævnligt anvendt i statistiske multiparameteralgoritmer, som afgiver prædiktive informationer om den enkelte patients diagnose, prognose eller forventet respons på en bestemt behandling. Som eksempel kan nævnes *Wrights* algoritme [6], der på baggrund af en global genprofil kan prædiktere, hvorvidt en patient med DLBCL tilhører én af de prognostisk forskellige undergrupper ABC DLBCL eller GCB DLBCL. *Potti* beskriver på tilsvarende vis en algoritme [7] til forudsigelse af kemosensitivitet, der kan estimere effekten af en given kemoterapeutisk behandling inden denne udføres i praksis. Integrative analyser af cytogenetiske data og globale genprofiler har defineret nye klassifikationssystemer for myelomatose [8] og for akut myeloid leukæmi [9]. Fælles for sådanne algoritmer er, at de via sandsynlighedsvurderinger karakteriserer den enkelte patients diagnostiske, prognostiske og prædiktive sikkerhed.

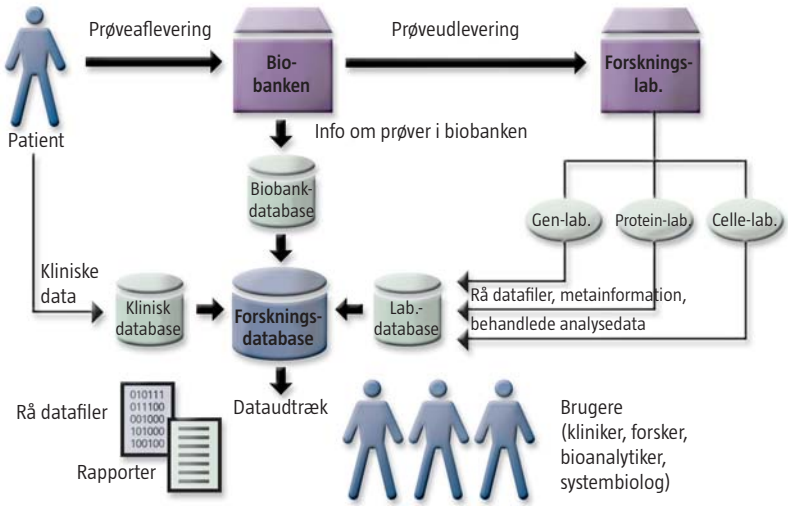
DATAOPBEVARING OG KVALITETSSIKRING

Algoritmer udviklet via integrative analyser som angivet ovenfor kan teoretisk set sikre den individuelle patient en objektiv diagnostik. Det er dog ikke altid tilfældet, at sådanne algoritmer viser sig reproducerbare ved validering på nye patientpopulationer. Dette kan blandt andet skyldes, at den forudgående produktion og registrering af de anvendte data ikke har været udført tilstrækkeligt systematisk og veldokumenteret [7, 10]. Der er derfor behov for en kvalitetssikret indsamling og opbevaring af alle patientrelaterede data i en sådan form, at datagrundlaget kan sikre fremtidig udvikling af effektive og reproducerbare prædiktive algoritmer.

Brugen af den translationelle database i forbindelse med kræftforskning er tidligere behandlet i artikler af *Ambrosone* og *Ennis* [11, 12], hvor der sættes fokus på etablering af biobanken med kobling til kliniske og epidemiologiske data. Som de vigtigste emner i forhold til opnåelse af en brugbar og kvalitetssikret database nævnes grundig planlægning af indhold samt oplæring af personalet, der skal varetage produktion og registrering. Derudover er det afgørende for senere anvendelse og reproducerbarhed,

FIGUR 1

Illustration af dataflow fra den kliniske forskningsenhed, biobanken og forskningslaboratoriets teknologier til den translationelle forskningsdatabase, som udnyttes af kompetente brugere.



at der pågår en løbende kvalitetssikring af materialet i biobanken ligesom en kvalitetssikret registrering af de tilknyttede data er nødvendig.

Kvaliteten af denne database sikres via et tilhørende dokumentationssystem, hvori alle procedurer for dataopsamling og -produktion er velbeskrevne, så resultater utvetydigt kan reproduceres og eventuelle fejl og mangler identificeres. Dokumentationssystemet sikrer: 1) at der foreligger standardiserede, diagnostiske udredningsprocedurer, behandlingsinstrukser og forløbsvurderinger [13-15], 2) at data fra globale profiler produceres og dokumenteres efter bedste praksis i litteraturen [16-18], 3) at kvalitetsprojekter løbende optimerer kvaliteten af de producerede data samt, 4) at arbejdet i biobanken altid udføres efter gældende standarder og procedurer [19].

Endelig er det allervigtigste, at procedurerne indeholdt i dokumentationssystemet er let tilgængelige og gennemsigtige, så alle interesserede klinikere har mulighed for at vurdere kvaliteten af data og selv udføre beregninger, før et givent forslag til en ny prædiktiv algoritme valideres og implementeres i klinikken.

PERSPEKTIVER

Det forventes, at indsamling, opbevaring og kvalitetssikring af alle data relateret til patienten vil give unikke muligheder i forbindelse med udvikling og implementering af nye faglige tiltag i sundhedsvæsnet. En sådan udvikling vil udover kliniske færdigheder kræve brug af nye faglige kompetencer inden

! DEN INTEGRATIVE ANALYSESTRATEGI

En hypotese opstilles med udgangspunkt i en velfunderet biologisk og klinisk antagelse.

Et antal patienter med relevante karakteristika i forhold til hypotesen identificeres i databasen med henblik på at besvare spørgsmålene.

Disse patienters retrospektive kliniske forløbsdata udtrækkes sammen med de rå datafiler for hver enkelt global profil og dertilhørende laboratorie- og biobankinformationer.

Det sikres, at alle data er genereret og håndteret under de samme forudsætninger i laboratoriet og biobanken, så de er sammenlignelige og ikke risikerer at afspejle tekniske variationer frem for reelle biologiske forskelle.

Data analyseres ved hjælp af bioinformatiske og biostatistiske analyseteknikker, og resultaterne giver dernæst anledning til en klinisk fortolkning.

Konklusionerne fra den integrative analyse omsættes til en automatiseret algoritme, som objektivt kan bestemme den enkelte patients diagnose, prognose eller behandlingseffekt ud fra en vævsprøve.

Algoritmens brugbarhed afprøves prospektivt i klinikken via et fase I-IV-studium.

for molekylær medicin, datalogi, bioinformatik og statistik. Prioriteres disse kompetencer, er der på længere sigt potentiale for udvikling af nye algortime-baserede, diagnostiske, prognostiske og prædiktive metoder.

Vi forventer i løbet af 2-3 år at have indsamlet detaljeret information omkring subgrupper af patienter med maligne lymfomer, leukæmier og myelomatose, som vil tillade etablering af nye algoritmer. Disse algoritmer vil blive valideret gennem kliniske fase I-IV studier [20] og implementeret i klinisk praksis. Lykkes dette, vil vi komme til at opleve et sundhedsfagligt paradigme skift.

KORRESPONDANCE: *Hans E. Johnsen,* Hæmatologisk Afdeling, Forskningens Hus, Sdr. Skovvej 15, 9000 Aalborg. E-mail: haej@rn.dk.

ANTAGET: 29. januar 2010.

FØRST PÅ NETTET: 5. april 2010

INTERESSEKONFLIKTER: Ingen

LITTERATUR

1. Turing AM. Computing machinery and intelligence. Mind 1950;59:433-60.
2. Margalit O, Somech R, Amariglio N et al. Microarray-based gene expression profiling of hematologic malignancies: basic concepts and clinical applications. Blood Rev 2005;19:223-34.
3. Orfao A, Ruiz-Arguelles A, Lacombe F et al. Flow cytometry: its applications in hematology. Haematologica 1995;80:69-81.
4. Unwin RD, Whetton AD. How will haematologists use proteomics? Blood Rev 2007;21:315-26.
5. Swerdlow SH, Campo E, Harris NL et al. WHO Classification of tumours of haematopoietic and lymphoid tissues. 4th ed. Lyon: IARC Press, 2008.
6. Wright G, Tan B, Rosenwald A et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci USA 2003;100:9991-6.
7. Potti A, Dressman HK, Bild A et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med 2006;12:1294-300.
8. Bergsagel PL, Kuehl WM. Molecular pathogenesis and a consequent classification of multiple myeloma. J Clin Oncol 2005;23:6333-8.
9. Marcucci G, Radmacher MD, Maharry K et al. MicroRNA expression in cytogenetically normal acute myeloid leukemia. N Engl J Med 2008;358:1919-28.
10. Coombes KR, Wang J, Baggerly KA. Microarrays: retracing steps. Nat Med 2007;13:1276-7.
11. Ambrosone CB, Nesline MK, Davis W. Establishing a cancer center data bank and biorepository for multidisciplinary research. Cancer Epidemiol Biomarkers Prev 2006;15:1575-7.
12. Ennis DP, Pidgeon GP, Millar N et al. Building a bioresource for esophageal research: lessons from the early experience of an academic medical center. Dis Esophagus 2009, 15. april (Epub ahead of print).
13. Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003;326:41-4.
14. Magid E, Jørgensen PE, Jørgensen LG et al. Klaringsrapporterne 1996-2001. Ugeskr Læger 2004;166:1664-6.
15. Rud B, Matzen P, Hilden J. Mål for diagnostiske tests ydeevne. Ugeskr Læger 2005;167:3018-22.
16. Minimum Information about a flow cytometry experiment MIFlowCyt 1.0. http://flowcyt.sourceforge.net/miflowcyt/ (1. marts 2010).
17. MIAPE: The Minimum Information About a Proteomics Experiment. http://www.psidev.info/index.php?q=node/91 (1. marts 2010).
18. Brazma A, Hingamp P, Quackenbush J et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001;29:365-71.
19. Steen Loiborg, Karen Brøndum-Nielsen, Camilla Daasnes et al. Redegørelse om biobanker – Forslag til retlig regulering af biobanker inden for sundhedsområdet, Indenrigs- og Sundhedsministeriet, Betænkning 1414, 2002.
20. Johnsen HE, Lanza F. Clinical validation – the final step in translational medicine! From CD34 enumeration to probability estimation of autograft quality. J Biol Regul Homeost Agents 2002;16:270-1.

Purpura fulminans

Reservelæge Karina Jordan & afdelingslæge Kim Kristensen

KASUISTIK

Rigshospitalet, Pædiatrisk Klinik

Purpura fulminans (PF) er en tilstand med hastigt progredierende hæmoragisk hudnekrose grundet småkarstrombose, der er udløst af dissemineret intravaskulær koagulation [1]. Tilstanden involverer oftest ekstremiteterne symmetrisk, og kun sjældent ses trombosering, der involverer viscera og centralnervesystemet [1, 2]. PF kan ses ved medfødt mangel på protein S eller C, men hyppigere er den udløsende årsag sepsis forårsaget af meningokokker eller gruppe A-streptokokker. Medvirkende faktorer kan være faktor V Leiden-mutation [2], protrombin G 20210A-polymorfisme, antitrombin III-mangel og antikardiolipinantistoffer.

SYGEHISTORIE

En tidligere rask femårig dreng, som debuterede med variceller otte dage forinden, overflyttedes fra et lokalt sygehus på grund af hastigt progredierende pur-