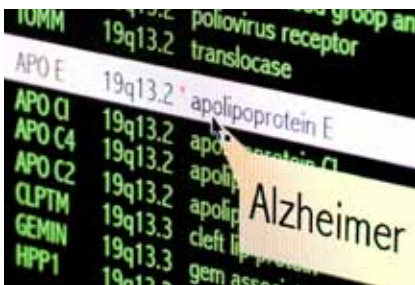


En ny æra for DNA-sekvensering – nye muligheter og nye utfordringer i humanmedisin

I løpet av få år er teknikkene og maskinene man bruker for å lese innholdet i våre DNA-molekyler, betydelig forbedret. Genetikere som ønsker å finne genetiske årsaker til sykdom eller andre egenskaper, får nå oppfylt drømmen om å kartlegge hele genomet til enkeltindivider. Med nye teknologier kommer også nye utfordringer. Hvordan håndterer man utilsiktede funn?

Dag Undlien



Genvarianter kan være koblet til sykdom. Ønsker pasientene å vite hva man finner i genene deres? Foto: iStockphoto.

Teknologien for sekvensering av DNA er i rivende utvikling (se tekstboks s. 18). Med økt effektivitet og lavere priser åpner den for at enkeltpersoner kan få kartlagt hele sitt genom som ledd i sykdomsdiagnostikk eller for å kartlegge mulighetene for forebyggende behandling.

Fra hypotesebasert genetisk forskning ...

Å kartlegge hele genomet til enkeltindivider har lenge vært en drøm for genetikere som ønsker å finne genetiske årsaker til sykdom eller andre egenskaper. Hittil har ikke dette vært praktisk mulig å gjennomføre, og man har derfor vært nødt til å lete målrettet i spesifikke, mindre deler av genomet. Dette fordrer at man har en hypotese om hvilke deler av genomet man vil undersøke. Slike hypoteser kan ha sin basis i studier av nedarving av genetiske markører i familier (genetiske koblingsanalyser) eller konkrete mistanker om hvilke gener som er involvert (gjørne kalt kandidatgentilnærming). Også såkalte helgenomassosiasjonsanalyser, som det har vært

gjort mange av de seneste år, er basert på hypoteser som for eksempel den såkalte «vanlige sykdommer-hypotesen = vanlige genvarianter-hypotesen».

... til en «hypotesefri» genetikk

Når vi nå har en reell mulighet til å kartlegge hele genomet til enkeltindivider, innebærer det at man kan forske på en helt annen måte. Man kan simpelthen lete etter variasjon og se om denne/disse variasjonene kan ha noen sammenheng med forskjellige sykdommer eller egenskaper. Et opplagt eksempel på sykdommer som nå i større grad kan angripes, er sjeldne sykdommer som man ikke kjenner årsaken til, men som man ser følger recessiv arvegang (det vil si at man må arve genvarianten som gir sykdom fra både mor og far). Slike sykdommer er vanskelige å studere med vanlige familieanalyser (koblingsanalyser). Ved å studere den lille andelen av genomet som utgjør de proteinkodende delene (det vil si alle eksonene, også kalt eksomet) til et lite antall pasienter med noen slike sykdommer, har man identifisert den genetiske årsaken til sykdommer i tilfeller hvor tradisjonell metodikk tidligere ikke har ført til målet. Et eksempel er Miller syndrom, hvor man nylig identifiserte den genetiske årsaken med denne tilnærmingen.

Sårbarhetsgener

Høykapasitets-DNA-sekvensering ventes også å få stor betydning for den videre kartlegging av sårbarhetsgener eller risikogener for spesielle sykdommer, og for sykdommer som skyldes mange faktorer. Ved

å kartlegge hele genomet med genetiske markører og sammenligne med forskjellige egenskaper (helgenomassosiasjonsanalyser) er det påvist et stort antall genetiske varianter som er assosiert med ulike multifaktorielle sykdommer. Det å identifisere hvilke av disse mange genetiske variantene som er den egentlige årsaken (kausal rolle), har imidlertid vist seg svært vanskelig i mange tilfeller. Det er ventet at systematisk sekvensering av områder som viser assosiasjon hos flere pasientgrupper, vil gjøre dette vesentlig lettere.

Nye diagnoseverktøy

Høykapasitets-DNA-sekvensering er per i dag først og fremst et forskningsverktøy, men det er også på full fart inn i medisinsk genetisk diagnostikk og vil gjøre det mulig å diagnostisere mange pasienter som per i dag ikke har noen diagnose. For det første vil man kunne forsøke å stille diagnose i tilfeller hvor vi hittil har ansett det for umulig å finne den genetiske årsaken. I tillegg vil den nye teknologien etter hvert gjøre det mulig å sekvensere mye mer kostnadseffektivt, noe som vil være med på å utvide repertoaret av genetiske tester og gjøre det mulig å gjøre slike tester raskere og mer effektivt enn i dag.

Andre bruksområder

Bruken av høykapasitets-DNA-sekvensering går langt utover det å identifisere arvelige årsaker til sykdom. Også ikke-arvelige mutasjoner, som for eksempel de som oppstår i kreftceller, vil i økende grad kunne identifiseres med denne teknologien. Andre forskningsfelt, som epigenetikk



Høykapasitets-DNA-sekvensering gjør det blant annet mulig å finne årsaken til sykdommer hos barn med sjeldne sykdommer. Foto: iStockphoto.

(blant annet kjemisk modifisering av basene i DNA-tråden, såkalt DNA-metylering), studier av hvilke gener som er aktive (genekspresjon) og så videre, vil ha stor nytte av høykapasitets-DNA-sekvensering. Det er i det hele tatt ventet at det meste av forskning på genomer vil benytte nettopp høykapasitetssekvensering.

Utsiktede funn

Med de nye maskinene som sekvenserer «alt» DNA som puttes inn i maskinen effektivt og rimelig, kommer nye muligheter og utfordringer. Denne metoden øker muligheten for at man finner ting man i utgangspunktet ikke tok sikte på å finne. Hvordan skal man håndtere denne typen overskuddsinformasjon? Det er i mange tilfeller informasjon det kan være vanskelig å vite på forhånd om man virkelig ønsker kunnskap om. Hvordan skal man håndtere det dersom man gjør funn som sier noe om risikoen for fremtidig sykdom hos den som sekvenseres? Mange vil vel mene at dersom en sykdom kan forebygges, vil man ønske beskjed om det, mens menningene gjerne vil være mer delte hvis det ikke er mulig å gjøre noe for å forebygge. Dette er utfordrende problemstillinger som de medisinske miljøer vil bli nødt til å finne gode løsninger for, løsninger som ivaretar pasienters og forsøkspersoners rett til å ta egne valg i kjølvannet av denne nye teknologien. Dette momentet bør også tas med i

betraktning når man i disse dager skal revidere bioteknologiloven.

Analytiske utfordringer – bioinformatikk
Sekvensering av hele genomer genererer store datamengder. En kjøring med disse nye sekvensatorene kan generere datamengder på flere terabyte (10^{12}). Dette representerer utfordringer både hva gjelder datalagringskapasitet og analyse. Når det gjelder sekvensering av humane genomer, må man også ta hensyn til personvern, noe som krever ekstra sikker datalagring.

Analysen av så store datamengder setter store krav til bioinformatisk kompetanse. Når man sekvenserer et individs genom, vil man kunne finne flere tusen genetiske varianter som aldri før er observert. Hvis formålet med sekvenseringen er å identifisere den arvelige årsaken til en sjelden sykdom, innebærer dette at man må lete blant alle disse variantene, og det er ikke opplagt hvordan man skal plukke ut den riktige. En rekke dataprogrammer er utviklet for å hjelpe til i denne form for analyser. Å gå nærmere inn på disse ligger utenfor temaet for denne artikkelen. Det som er det viktige å være klar over i denne sammenheng, er at med høykapasitets-DNA-sekvensering er de største utfordringene ikke knyttet til de molekylærbiologiske metodene, men derimot til den bioinformatiske analysen.

FAKTABOKS

Arvestoffet – DNA

Alle levende organismer har et arvestoff – et **genom** – som består av DNA. DNA er lange, trådformede molekyler. Et DNA-molekyl består av to DNA-tråder som er tvunnet om hverandre og danner en dobbelspiral, en DNA-dobbelheliks. Trådene er bygget opp av fire byggesteiner: Basene adenin (A), cytosin (C), guanin (G) og tymin (T). Hver av basene på den ene tråden er koblet til en base på den andre tråden og danner et basepar (se bilde i tekstsaks s. 18). Det er alltid slik at C og G danner basepar sammen og A og T danner basepar sammen.

Et **gen** er en del av arvestoffet (DNA) som forenklet sagt inneholder informasjon om hvordan et bestemt protein skal bygges opp. Genene til en organisme utgjør **eksomet** til organismen, og dette er bare en liten andel av hele genomet.

DNA-sekvensering er en metode som brukes for å bestemme rekkefølgen av byggesteinene (A, T, C og G) i DNA-molekyler.

FAKTABOKS



DNA-dobbelheliks. Foto: YAY MICRO.

Utvikling av DNA-sekvenseringsmetoder

Dagens maskiner for høykapasitets-DNA-sekvensering er flere millioner ganger mer effektive enn dem vi hadde for kun få år siden. Måten disse maskinene fungerer på, gjør at de ikke bare kan brukes til å kartlegge genomer mer effektivt, men også til å studere kvantitative og dynamiske aspekter ved genomet, som for eksempel genuttrykk (genekspresjon) og kopitallsvariasjon (antall kopier av genene). Det er mye som tyder på at mange av teknikkene som i dag brukes til å studere genomer, vil bli erstattet av høykapasitets-DNA-sekvensering, som gir en oppløselighet på enkeltbasenivå og kan gi svar på mange aspekter i ett og samme eksperiment.

Pris og effektivitet for DNA-sekvenseringen er i ferd med å endre seg. Mens kartleggingen av det humane genom i det humane genomprosjektet tok 13 år og kostet 2,8 milliarder dollar, brukte man kun 4,5 måneder på kartleggingen av genomet til genforskningspionéren James Watson i 2008, og kostnadene var falt til 1,5 millioner dollar. Utviklingen har ikke stoppet med det: Kommerielle aktører tilbyr helgenomsekvensering for 50 000 dollar og varsler at vi i løpet av 2010 kanskje vil være under 5 000 dollar. Man nærmer seg med stormskritt

DNA-sekvenseringssuperstjernen Craig Venters gamle visjon om «the 1000 dollar genome». Venters visjon går ut på at dersom man får en pris som er så lav som dette, ser man for seg at et stort antall individer kan få kartlagt sine genomer og nyttiggjøre seg informasjonen til å få individuelt tilpasset medisinsk behandling og forebyggende tiltak, såkalt skreddersydd medisin.

Parallellsekvensering

Per i dag er det en håndfull firmaer som har teknologier for høykapasitets-DNA-sekvensering. Teknologiene bygger på litt ulike prinsipper og har ulike styrker og svakheter. Felles for dem alle er at de, til forskjell fra tradisjonell sekvensering, sekvenserer mange DNA-molekyler (eller egentlig fragmenter av DNA-molekyler) samtidig i en og samme reaksjon. I tradisjonell sekvensering sekvenserer man ett og ett molekyl som normalt ikke er større enn cirka 500–1000 basepar. I typiske tilfeller isolerer man den delen av et genom som man er interessert i å studere (ved hjelp av polymerasekjederaksjon (PCR) eller kloning). De tradisjonelle sekvensatorene sekvenserer derfor ett slikt fragment på 500–1000 basepar om gangen. Skal man sekvensere flere slike DNA-fragmenter, må man kjøre maskinen flere ganger.

Det de nye sekvensatorene har til felles, er at de behandler hundretusener til millioner DNA-fragmenter samtidig i en og samme kjøring på maskinene. Selv om maskinene har noen ulikheter i måten de fungerer på, er prinsippet om parallell sekvensering felles og avgjørende for disse maskinenes effektivitet.

En typisk sekvensering på en av disse maskinene fungerer slik: DNA-molekylene deles opp i mange millioner små fragmenter (på for eksempel noen hundre basepar) som festes på en glassplate. Deretter gjøres alle de kjemiske reaksjonene (amplifisering – oppkopiering – og sekvensering) på denne glassplaten på alle fragmentene som er festet der samtidig. Kjemikaliene som brukes i sekvenseringsreaksjonen er merket, gjerne med en form for fargemerkning. Sekvensa-

torene har avanserte lasere som bruker dette til å lese baserekkefølgen i sekvenseringsreaksjonen. Hovedforskjellen i forhold til «gammeldags» sekvensering er altså at dette gjøres på millioner av fragmenter samtidig.

Mikromatrisebasert seleksjon

Det er sjelden man har behov for å sekvensere hele genomet til individer. Man har derfor sett et behov for metoder som kan brukes til å velge ut de delene av genomet man er interessert i å studere. For eksempel vil det i noen sammenhenger være et bestemt kromosom man er interessert i å studere, mens det andre ganger kun er de proteinkodende delene av genomet («eksomet»). Mens PCR-reaksjonen er vel egnet til å selektere enkeltgener og små DNA-fragmenter, har det inntil nylig vært en mangel på metoder til å selektere større deler av genomet for målrettede analyser. Med den nye sekvenseringsteknologien har vi parallelt sett en utvikling av slike metoder. En av de mest suksessfulle metodene har vært det som på engelsk går under betegnelsen «capture arrays». Denne metoden innebærer at man lager en mikromatrise (liten plate med tusenvis av punkter) hvor de deler av genomet man er interessert i å studere, er festet på mikromatrisen. For eksempel kan man feste DNA fra alle proteinkodende deler av genomet (eksone) på mikromatrisen. DNA fra genomet til personen som skal undersøkes festes (hybridiseres) til mikromatrisen, og DNA som ikke bindes til mikromatrisen vaskes bort. Til slutt løsner man DNA-et som er bundet til mikromatrisen, og dette DNA-et brukes i den videre analysen.

Dag Undlien er professor ved Institutt for medisinsk genetik, Universitetet i Oslo, og avdelingsleder ved Avdeling for medisinsk genetik, Oslo universitetssykehus. I samarbeid med Institutt for biologi, Universitetet i Oslo, driver hans gruppe den nasjonale sekvenseringsplattformen «Norwegian high-throughput Sequencing Centre».