

Strømme syndrom - Jakten på et gen

Kristin Brandal

Cellebiologi for medisinsk/teknisk personell
Oppgaven levert: November 2011
Hovedveileder: Åse Krøkje, IBI

Forord

Et, for meg, stort stykke arbeid er gjennomført med god hjelp fra uvurderlige støttespillere. Jeg vil rette en stor og hjertelig takk til min hovedveileder ved Avdeling for medisinsk genetik, Kaja Kristine Selmer, som har vært hjernen bak denne prosessen. Takk for alle konstruktive tilbakemeldinger og for at du er en strålende samarbeidspartner. Jeg kunne ikke gjort dette uten deg! Takk også til Dag Undlien, min biveileder ved avdelingen, for din faglige tyngde som har resultert i mange fruktbare innspill, for organisering av finansiering og for å ha gitt meg muligheten til å gjennomføre dette masterprosjektet ved Avdeling for medisinsk genetik. Videre fortjener min veileder ved NTNU, Åse Krøkje, en takk for å ha bidratt med praktiske innspill der det har vært nødvendig.

En takk rettes til Petter Strømme som har hatt dette prosjektet i tankene gjennom mange år, til Yolande van Bever for godt samarbeid ut over Norges grenser og til familiene som har stilt opp og gitt en del av sitt arvemateriale i søken på løsningen.

Også andre fortjener å bli verdsatt for å ha bidratt i prosessen. Hanne Elisabeth Akselsen: Hjertelig takk for din enorme sjenerøsitet og fleksibilitet! Til alle gode kollegaer ved Avdeling for medisinsk genetik: Hanne Sorte for fruktbare diskusjoner på kontoret, Rune Moe, Kristine Fjelland, Gregor Gilfillan, Christin Thrane Pedersen, Oddveig Røsby, Robert Lyle og sist, men ikke minst, Ying Sheng som har stått for databehandlingen av storskalasekvenseringsresultater.

Takk til mine studievenner, Ingunn Østlie og Ingunn Stav Johannesen, for at dere holdt hodet mitt over vann i regntunge høstucker i Trondheim. Deres vennskap er uten tvil det beste utbyttet masterprosjektet har gitt meg.

Takk mamma og pappa for at dere har vært en ressurs gjennom alle år og lært meg verdien av kunnskap og nysgjerrighet. Og tusen takk kjære Even for at du har gjort hverdagene mine lette og lystbetonte selv når arbeidet har tynget.

November 2011

Kristin Brandal

Sammendrag

I 2005 ble en ny type sekvenseringsteknologi introdusert som åpnet for uante muligheter innen genetikken. Prosjekter som tidligere ble ansett som uoverkommelige, både teknisk og økonomisk, kan nå gjennomføres. Sekvenseringsteknologien som baserer seg på massiv, parallell sekvensering har fått navn som nestegenerasjonssekvensering og storskalasekvensering. Prinsippet for denne metoden og potensielle bruksområder blir diskutert i tillegg til utfordringer knyttet opp mot teknologien.

Et av områdene som har fått en merkbar utvikling etter at denne teknologien kom på markedet er studien av monogene sykdommer; sykdommer som forårsakes av mutasjoner i ett gen alene. Man antar at det finnes rundt 7000 slike sykdommer. Den genetiske årsaken er ukjent for mer enn halvparten av disse.

Strømme syndrom, som først ble beskrevet i 1993, antas å være en monogen sykdom. Syndromet uttrykker seg i pasientene gjennom tarmmisdannelse, øyemisdannelser og liten hodeomkrets. Dette er en av mange sjeldne tilstander hvor man tidligere ikke har hatt metoder for å kunne lokalisere den genetiske årsaken.

Ved bruk av storskalasekvensering av den kodende delen av genomet, har vi i dette prosjektet lett etter de(n) kausale varianten(e) for Strømme syndrom. Tilsvarende prosjekter med lignende design hadde vist seg å være vellykket tidligere, men i dette prosjektet klarte vi ikke finne en sikker årsak til sykdommen. En rekke kandidatgener ble identifisert, men ingen kunne med sikkerhet fastslås å være sykdomsgenet.

Årsakene til manglende konklusjon rundt sykdomsårsaken kan være flere. Disse kan være knyttet både til de biologiske antagelsene som er lagt til grunn for prosjektet og til rent metodiske svakheter knyttet opp mot storskalasekvensering. For å kunne identifisere et eventuelt sykdomsgen for Strømme syndrom vil undersøkelse av flere pasienter trolig være nødvendig.

Innholdsfortegnelse

1	Forkortelser	1
2	Introduksjon	3
2.1	Bakgrunn	3
2.2	Hypotese	4
2.3	Kliniske aspekter	5
2.3.1	“Apple peel” atresi	5
2.3.2	Mikrocefali	6
2.3.3	Øyemisdannelser	7
2.4	Genetiske aspekter	8
2.4.1	Det humane genom.....	8
2.4.2	Varianter.....	8
2.4.3	Monogen og multifaktoriell sykdom.....	10
2.4.4	Arvegang ved monogene sykdommer.....	11
2.5	Genetisk kartlegging av monogene sykdommer	13
2.6	HiSeq 2000	14
2.7	Prinsipp for storskalasekvensering	14
2.7.1	Prøvepreparering	15
2.7.2	“Eksom-fanging”	16
2.7.3	Sekvensering	17
3	Formål	21
4	Materialer og metoder	22
4.1	Materiale	22
4.2	Ekstrahering av DNA	23
4.3	DNA-kvantitering/konsentrasjonsmåling	23
4.4	Array comparative genomic hybridization, aCGH.....	23
4.5	Prøvepreparering for storskalasekvensering.....	24
4.5.1	Fragmentering	25
4.5.2	SPRIWorks.....	25
4.5.3	PCR	28
4.5.4	“Eksom-fanging”	29
4.5.5	PCR av “fanget” bibliotek.....	31

4.5.6	Kvalitetskontroll av bibliotek.....	32
4.5.7	“Clustering”	35
4.5.8	Storskalasekvensering	35
4.5.9	Bioinformatikk	36
4.6	Sangersekvensering	37
4.6.1	Primerdesign.....	37
4.6.2	PCR	37
4.6.3	Rensing av PCR-produkter.....	38
4.6.4	Sekvenseringsreaksjon	38
4.6.5	Rensing av sekvenseringsprodukter	39
4.6.6	ABI3730xl.....	39
4.6.7	SeqScape	39
5	Resultat.....	40
5.1	aCGH.....	40
5.2	Konsentrasjons- og kvalitetsmålinger	40
5.3	qPCR.....	42
5.4	Storskala- og Sangersekvensering.....	44
5.4.1	Homozygote varianter - Kandidatgener	46
5.4.2	Compound heterozygote varianter - Kandidatgener	49
5.4.3	Gener med dårlig sekvensdekning	53
5.4.4	Kandidatgener med en variant hos en pasient og to i den andre	53
6	Diskusjon.....	54
6.1	Hvorfor finner vi ikke årsaken.....	54
6.1.1	Usikkerhet i hypotese og antagelser.....	54
6.1.2	Begrensninger i metode.....	58
6.2	Veien videre i jakten på den genetiske årsaken til Strømme syndrom.....	68
6.3	Storskalasekvensering i forskning og diagnostikk	69
6.3.1	Forskning.....	69
6.3.2	Diagnostikk	70
6.3.3	Utfordringer ved bruk av storskalasekvensering.....	72
	Litteraturliste	76
	Vedlegg	81

1 Forkortelser

aCGH	array comparative genomic hybridization
ATP	adenosintrifosfat
bp	basepar
C	Celsius
CCDS	human konsensus kodende sekvens
ChIP	kromatin immunopresipitasjon
CNV	kopitallsvariasjon
Ct	crossing treshold
Δ Ct	delta crossing treshold
dH ₂ O	destillert vann
DNA	deoksyribonukleinsyre
cDNA	komplementært DNA
gDNA	genomisk DNA
dNTP	deoksynukleosid 5'-trifosfat (N kan være A = adenin, T = tymin, C = cytosin og G = guanin)
EB	elueringsbuffer
EDTA	etylendiamintetraeddiksyre
Fwd	forward
Gb	gigabaser
hg	humant genom
indel	insersjon/delesjon
IQ	intelligenskvotient

kb	kilobaser
Mb	megabaser
min	minutter
mL	milliliter
mM	milli Molar
µg	mikrogram
µL	mikroliter
µM	mikromolar
MLPA	multipleks ligeringsavhengig probe amplifisering
MgCl ₂	magnesiumklorid
NaOH	natriumhydroksid
ng	nanogram
OUS	Oslo Universitetssykehus
PCR	polymerase kjede-reaksjon
pmol	pikomol
qPCR	kvantitativ polymerase kjede-reaksjon
Rev	revers
RNA	ribonukleinsyre
SD	standardavvik
sek	sekunder
SNP	enkeltbasepolymorfisme
ssDNA	enkeltrådet DNA
VUS	variant av ukjent betydning

2 Introduksjon

2.1 Bakgrunn

Man antar at det finnes ca 7000 sykdommer som skyldes feil i ett gen alene, såkalte monogene sykdommer (1). Mange av disse er syndromer, det vil si en samling av symptomer og funn fra flere organsystemer. Per i dag kjenner man den molekylære basisen for over 3000 av disse monogene sykdommene (1). Det betyr at det finnes mange monogene sykdommer som man ennå ikke har funnet årsaken til. Selv om monogene sykdommer hver for seg ofte er sjeldne, har det vist seg at disse til sammen utgjør en stor gruppe pasienter i samfunnet vårt. Med en rivende teknologisk utvikling innenfor genetiske laboratoriemetoder i løpet av de siste tiårene, har man nå muligheten til å utforske disse sjeldne sykdommene. I denne masteroppgaven har man dratt nytte av en av disse metodene, nemlig storskalasekvensering, for å lete etter årsaken til tilstanden Strømme syndrom.

Dette syndromet ble først beskrevet av Petter Strømme i 1993 ved Rikshospitalet i Oslo. Han rapporterte da to søstre av friske foreldre som begge viste det samme sykdomsbildet. De hadde “apple peel” lignende tarmmisdannelse, øyemisdannelser, mikrocefali og har i senere tid blitt rapportert til å ha noe grad av mental retardasjon. Siden den gang har ytterligere åtte tilfeller med tilsvarende symptomer blitt beskrevet i litteraturen på verdensbasis (2-9). I et av tilfellene har man sett en dødelig utgang. Flere syke i en og samme familie er kun rapportert i familien beskrevet av Petter Strømme. Alle øvrig rapporterte tilfeller har vært sporadiske. Totalt ni tilfeller (i åtte familier) er rapportert.

Primære kjennetegn inkluderer øyemisdannelser og lukket tarm. Utover dette ses i enkelte tilfeller mikrocefali og mental retardasjon. Type tarmmisdannelse er oftest “apple peel”, men det har vært rapportert tilfeller hvor den kliniske tilstanden tyder på dette syndromet, men der tarmatresi av annen type enn “apple peel” er observert (5).

Tilstanden ble først omtalt som “Apple Peel syndrom med mikrocefali og øyemisdannelser”, men har i senere tid stadig oftere blitt omtalt som “Strømme syndrom” etter Petter Strømme som først beskrev den.

2.2 Hypotese

Det er ikke sikre data til å underbygge at Strømme syndrom er en arvelig sykdom. Det at sykdommen er medfødt, sykdommens karakter og det at det i en av de svært få familiene som er beskrevet med sykdommen var flere affiserte individer gjør det imidlertid nærliggende å tro at sykdommen skyldes en til nå ukjent mutasjon(er) i ett eller flere gen. Det at sykdommen er genetisk betinget er en underliggende hypotese bak dette prosjektet.

Men det finnes ingen indikasjoner på hvilket gen som kan være knyttet opp mot tilstanden. Mangelen på klare kandidatgener gjør at et hvilket som helst gen kan inneholde de(n) sykdomsgivende varianten(e). Alle gener må derfor undersøkes og en helgenomisk tilnærming, ved bruk av fullsekvensering av genomet, fremtrer som et naturlig metodevalg. Helgenomsekvensering kan oppfattes som å være fullstendig hypotesefri, men dette er ikke tilfelle. For å kunne bearbeide data som produseres ved bruk av metoden, må man ha en hypotese som vil være avgjørende for innfallsvinkel i datahåndteringen.

Familiehistorikken ved Strømme syndrom tyder som nevnt på at tilstanden er genetisk betinget, men det foreligger ikke nok data til å ha noen klar formening om arvegang. Autosomalt recessiv arvegang eller dominant arvegang hvor sykdommen skyldes en nyoppstått mutasjon er nærliggende muligheter. Det at to affiserte søsken av friske foreldre er rapportert, trekker i retning av at førstnevnte mulighet er mest sannsynlig. Dette tyder på at foreldrene er bærere av de(n) sykdomsgivende varianten(e); de har begge en kopi av varianten(e). Dette tyder videre på at en variant alene ikke er nok for å utvikle sykdom. Man forestiller seg derfor at tilstanden mest sannsynlig er recessiv. Det er mindre sannsynlig at en *de novo* (spontan oppstått) variant vil forekomme i to søsken.

Siden sykdommen kan ses hos begge kjønn er den mest sannsynlig autosomal. Autosomal viser til at genet er lokalisert på et av de kromosomene som kalles autosomer. Dette er alle kromosomene utenom kjønnskromosomene.

På bakgrunn av antagelsene som er gjort rundt tilstanden er det mest sannsynlig en compound heterozygot eller en homozygot variant som forårsaker den. Ved compound heterozygote tilstander vil de syke ha to ulike varianter i samme gen. Variantene ligger på hvert sitt allel. Ved homozygote tilstander finnes den samme varianten på begge allelene. Siden det ikke er observert noen form for inngifte i de beskrevne familiene, er det mindre sannsynlig at begge

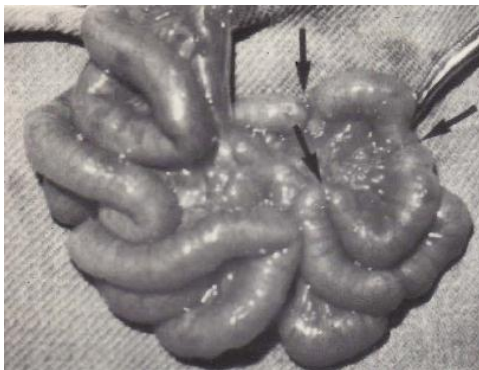
foreldrene er bærere av den samme foundermutasjonen (en spontant oppstått variasjon som nedarves innad familier) enn at de er bærere av hver sin variant. I hypotesen heller man derfor mer mot en compound heterozygot variant enn en homozygot.

Antagelsene om en monogen sykdom som nedarves autosomt recessivt er alle avgjørende for å kunne komme frem til et resultat ved helgenomsekvensering.

2.3 Kliniske aspekter

2.3.1 “Apple peel” atresi

Begrepet “apple peel” ble først brukt av Santulli *et al* i en publikasjon fra 1961 (10), og ordet “epleskrell” henviser til tarmens utseende. Illustrert i Figur 1. Den henvises også til som type IIIb eller type 4 atresi (11).



Figur 1. “Apple peel” atresi.
Bildet viser tarm med “apple peel” atresi. Pilene viser innsnevringar på tarmen.

“Apple peel” atresi er en tilstand hvor misdannelse av tarmen fører til at tynntarmen snurrer seg rundt blodåren som tilfører den blod. Den utgjør rundt 10 % av alle tynntarmmisdannelser (12). Siden tilstanden først ble rapportert i 1961 frem til 2002 har bare 155 tilfeller blitt rapportert i engelskspråklig litteratur (12). Søk på tilfeller etter 2002 resulterte i funn av ytterligere 12 publikasjoner. “Apple peel” misdannelse er ofte assosiert med flere ulike symptomer og funn. For tidlig fødsel, malrotasjon og kort tarm ses ofte i tillegg til “apple peel” misdannelse og oftere her enn i den generelle befolkning eller hos pasienter med andre typer atresier (13). Ytterligere misdannelser har vært rapportert til å utgjøre 15 % ved “apple peel” mens det samme tallet for andre tarmmisdannelser ligger på 7 % (13).

Det kan ses en familiær opphopning av denne type misdannelse og gjentagelsesrisiko hos søsken er rapportert til å være 8/34 (24 %) (13). Man mistenker følgelig en autosomal recessiv form for nedarving. Dette blir støttet av funn gjort i familier med høy forekomst av inngifte.

“Apple peel” atresi kan oppdages prenatalt ved bruk av ultralyd (13). Barna bør umiddelbart etter fødsel følges opp med kirurgisk inngrep for å rette opp i tilstanden. Overlevelsesraten for “apple peel” atresi har økt de senere år, men det er fortsatt en akutt, livstruende tilstand hos nyfødte barn med misdannelsen. I tiden rundt tilstanden først ble rapportert (1961), døde så mange som 70 % kort tid etter fødsel mens andelen som dør kort tid etter fødsel nå er nede i 14 % (12). Dødsårsaken er hovedsakelig sepsis, dysfunksjon etter kirurgisk spleising av tarm og koldbrann i tarmen (12).

2.3.2 Mikrocefali

Mikrocefali er en tilstand hvor hodeomkretsen er mindre normalt. Årsaken til mikrocefali kan være at hjernen ikke har utviklet seg normalt i fosterlivet eller at den har stoppet å vokse etter fødsel. Mikrocefali er en klinisk observasjon og ikke en diagnose.

Når det gjelder definisjon av mikrocefali er det noe diskusjon rundt denne. I enkelte tilfeller blir det referert til en hodeomkrets mindre enn – 2 standardavvik (SD) korrigert for alder og kjønn. Dette kriteriet impliserer at 2 % av den generelle befolkning har tilstanden (14). I andre tilfeller blir definisjonen satt til – 3 SD under gjennomsnittlig hodeomkrets, noe som vil resultere i at bare 0,1 % av befolkningen oppfyller kriteriene til å ha mikrocefali (14). På bakgrunn av dette er det vanskelig å gi et godt estimat på antall rammede av tilstanden. Den har blitt rapportert å være fra 1/10 000 til 1/1360 og observeres oftere i geografiske områder med stor grad av inngifte (14).

Mikrocefali klassifiseres videre i ulike subgrupper. Tilstanden kan være isolert eller ses sammen med andre anomalier. Isolert mikrocefali observert ved fødsel, som ikke kan tilskrives ytre faktorer, kalles primær mens den blir omtalt som sekundær dersom den utvikler seg senere i livet eller i assosiasjon med andre tilstander (15). Dersom den blir observert sammen med andre anomalier blir den også omtalt som syndromtilknyttet (15).

Mest sannsynlig er primær mikrocefali forårsaket av at et redusert antall neuroner blir laget under neurogenesen mens sekundær forårsakes av et redusert antall dendrittiske prosesser og synaptiske forbindelser (15).

Men også rundt klassifiseringen i primær og sekundær er det noe diskusjon; enkelte benytter seg nemlig av ”primær” som henvisning til genetisk påvirket og ”sekundær” til miljøpåvirket (16).

Antall sykdomstilstander/syndromer assosiert med mikrocefali er stort. Søk på mikrocefali i OMIM resulterer i 704 treff per oktober 2011. Dette inkluderer både isolert og syndromisk form.

Graden av mental retardasjon som følge av mikrocefali er relatert til graden av denne. Median IQ faller med fallende hodeomkrets i rapporterte studier (14). Alvorlighetsgraden kan også være knyttet til andre misdannelser i hjernens anatomi.

Årsakene til tilstanden er mange og virker å være påvirket av både genetiske- og miljømessige faktorer. Genetiske årsaker til isolert mikrocefali er hovedsakelig autosomal dominante, recessive eller X-bundet (14). Både store kromosomale avvik så vel som mindre delesjoner/duplikasjoner og enkeltmutasjoner kan være årsak til tilstanden (14). I sjeldne tilfeller kan man se ringkromosomer, mosaikker eller balanserte translokasjoner som forårsaker tilstanden (14).

2.3.3 Øyemisdannelser

Søk på “ocular anomalies” i OMIM gir 284 treff per oktober 2011. Øyemisdannelser er med andre ord hyppig beskrevet og antagelig relativt ofte syndromtilknyttet. En rekke ulike anomalier kan ses hos pasientene med Strømme syndrom: Uklarheter på hornhinnen, uklar grense mellom hornhinnen og det hvite på øyet (sclera), utspiling av øyets fremre vegg på grunn av svekkelse i øyevevet, tap av hornhinnens ytre lag, liten hornhinne, innoverskjeling, irregulær og utvidet pupill, manglende eller grunt forkammer, små øyne, perifer vaskularisering og stort øyeeple. Funnene er altså variable, men misdannelser i øyets forkammer er alltid involvert.

2.4 Genetiske aspekter

2.4.1 Det humane genom

Det humane genomet er omtrent 3000 Mb stort, det vil si 3 milliarder basepar stort. Eksomet, det vil si all proteinkodende sekvens, utgjør rundt en prosent av dette, noe som tilsvarer ca. 30 Mb. Innenfor disse 30 Mb er alle de ca. 20-25 000 genene som finnes i det humane genom lokalisert. Genene består av kodende og ikke-kodende deler kalt henholdsvis eksoner og introner. Eksomet refererer til de kodende delene av genomet; alle eksonene. Siden de fleste funn man har gjort for genetisk årsak til sykdom per i dag ligger i de kodende delene av genomet (17), er undersøkelse av eksomet et godt utgangspunkt for å lete etter sykdomsgivende varianter. En annen grunn til å fokusere på eksomet er at vi per i dag mangler mye kunnskap om funksjonen til de ikke-kodende delene og det vil være mye vanskeligere å identifisere hvilke(n) av de mange varianter i denne delen av genomet som eventuelt forårsaker sykdom.

2.4.2 Varianter

Rundt 99,9 % av DNA-sekvensen er lik mellom to gitte individer (18). Den gjenstående fraksjonen utgjøres av variasjoner som enten er sykdomsgivende eller en del av normalvariasjonen. Omtrent 90 % av variantene er enkeltbasevarianter, også kalt SNPer (Single Nucleotide Polymorphism). Slike polymorfismer er vanlig variasjon hvor den sjeldne varianten er hyppigere enn 1 %. De resterende 10 % utgjøres av mer sjeldne varianter (19) også kalt mutasjoner. Mutasjoner er variasjon med lavere hyppighet enn 1 %.

Ofte brukes uttrykkene mutasjon og sykdomsgivende variant om hverandre for å beskrive sykdomsgivende endringer i DNA-sekvensen. Definisjonen av ordet mutasjon er “endring av genomisk sekvens” og trenger altså ikke å forårsake sykdom. Uttrykket blir derfor ofte feilaktig assosiert med en sykdomsgivende variant. Siden man ikke alltid vet konsekvensen av en endring i DNA-sekvens, brukes fortrinnsvis ordet variant inntil det foreligger tilstrekkelige bevis for at endringen fører til sykdom.

Varianter kan være av ulike typer. Man kan observere enkeltbaseendringer (punktmutasjon), en eller flere baser kan ha kommet til eller blitt fjernet i DNA-sekvensen (insersjon/delesjon), DNA-områder kan ha blitt flyttet på (translokasjon) eller snudd (inversjon) eller DNA-

sekvens kan observere i ulikt antall kopier (kopitallsvariasjon også kalt “copy number variation”/CNV).

Varianter kan videre forekomme i heterozygot eller homozygot form. En variant som forekommer i heterozygot versjon finnes bare på ett av de to allelene til personen, mens den blir karakterisert som homozygot dersom den forekommer på begge allelene.

I de kodende delene av genomet koder tre baser (en triplett) sammen for en aminosyre; de utgjør ett kodon. Rekker av aminosyrene utgjør så proteiner. Varianter i genomet kan gi ulike utfall som påvirker aminosyrene og i neste omgang proteinene.

De kan forårsake endring av hvilken aminosyre tripletten koder for og blir da omtalt som ikke-synonyme varianter. Eventuelt kan varianten være synonym; dvs ikke gi en aminosyreendring. Dette har sammenheng med hvordan den genetiske koden er bygget opp. En del aminosyrer kan nemlig være kodet for av ulike tripletter. I et kodons posisjon en og to, ses oftest den samme basen, mens den siste posisjonen kan være variabel og likevel kode for den samme aminosyren. Som følge av dette vil en baseendring i siste posisjon gjerne være mindre skadelig enn en baseendring i en av de andre posisjonene da en endring her sjeldnere vil forårsake en aminosyreendring. Det er verdt å nevne at selv om det ikke oppstår en aminosyreendring, har man vist at baseendring alene også kan føre til endrede egenskaper hos proteiner (20).

Videre koder enkelte tripletter for stoppkodon istedenfor aminosyrer. Dersom en variant medfører dannelse av et slikt i stedet for en aminosyre vil man få en trunkering av proteinet som skal uttrykkes. Varianter som forårsaker introduksjon av et prematurt stoppkodon i genet blir omtalt som nonsense varianter. Dersom varianten er av typen insersjon eller delesjon, hvor antall baser som har kommet til eller blitt fjernet ikke går opp i tre, vil man få en forskyvning av leserammen for det aktuelle genet (rammeskiftsvariasjon). I slike tilfeller vil ofte premature stoppkodon introduseres.

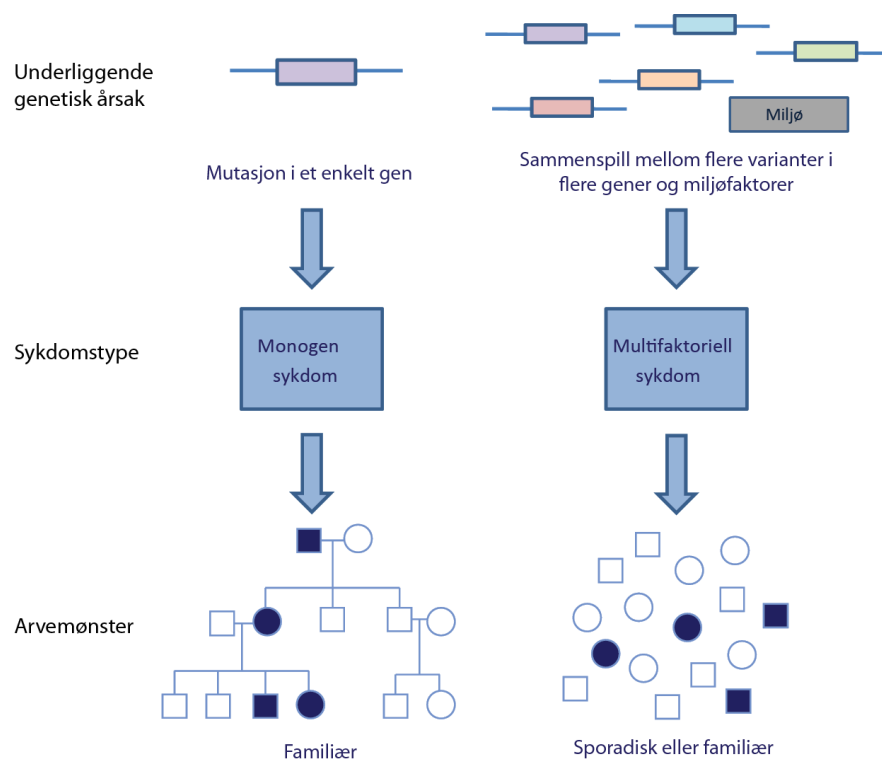
Forekomst av ulike varianter i genomet vil altså kunne føre til trunkering av proteiner og manglende dannelse av disse eller produksjon av misdannede proteiner. Hvilke(n) aminosyre(r) som kommer inn som resultat av en variant kan være avgjørende for utfallet. Dersom den opprinnelige aminosyren blir erstattet med en som innehar mange av de samme

egenskapene, kan det være mindre skadelig enn om den skulle blitt erstattet med en aminosyre med totalt forskjellige egenskaper.

Videre kan det være avgjørende hvor, i den kodende delen av genet, varianten forekommer hvorvidt den er sykdomsgivende eller ikke. Forekommer f.eks. en nonsense mutasjon tidlig i sekvensen er det stor sannsynlighet for at den vil være avgjørende for proteinuttrykket, det samme gjelder om den ligger i bestemte domener som for eksempel kan påvirke foldingen av proteinet.

2.4.3 Monogen og multifaktoriell sykdom

Sykdommer nedarves forskjellig. Dette avhenger av om de er påvirket av flere gener, av flere gener og miljø, eller av kun ett gen alene. Se Figur 2.



Figur 2. Arveformer.

Figuren viser sammenheng mellom genetisk variasjon og arvemønster for monogen og multifaktoriell sykdom.

Sistnevnte gruppe sykdommer kalles monogene eller Mendelske sykdommer. Disse karakteriseres av at feil (eller variasjon) i ett gen alene er nok til å forårsake sykdom. Det betyr at varianten(e) har en avgjørende betydning i sykdomsutviklingen. I slike tilfeller vil man kunne observere hvordan sykdommen går i arv i familier og bli nedarvet direkte fra foreldre til barn.

Trolig finnes det rundt 7000 monogene sykdommer. Pr nå har man klart å identifisere de ansvarlige genene og tilhørende genvarianter for rundt 3000 av disse (1).

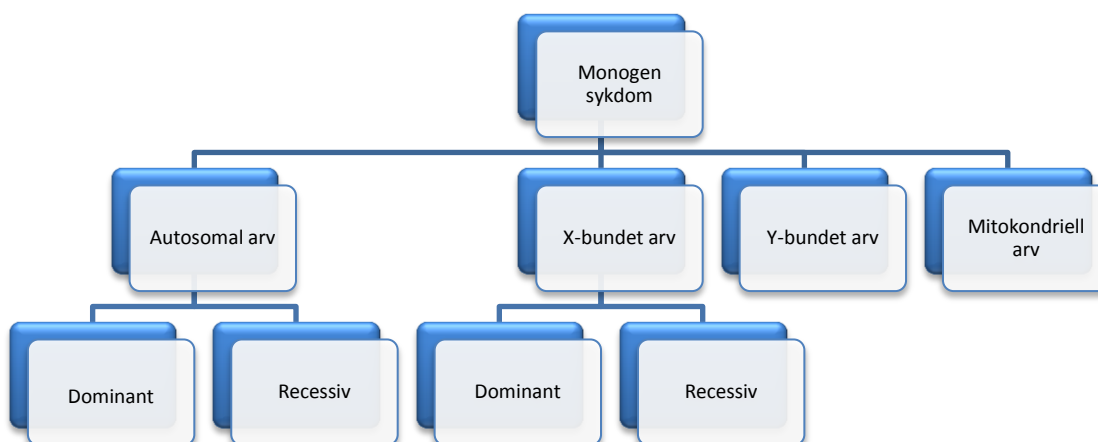
I motsetning til hva som er tilfelle i monogen sykdom, er ikke én bestemt genetisk variant nødvendig eller tilstrekkelig for at sykdom skal utvikle seg i multifaktorielle sykdommer. Her vil derimot flere gener være av betydning, det vil si sykdommen er multifaktoriell, også kalt polygen. Ofte kan det være normalvarianter av gener som sammen fører til sykdomsutvikling. Videre vil miljøfaktorer ofte ha en stor innvirkning på hvorvidt sykdom utvikler seg.

De fleste av våre vanlige folkesykdommer kommer inn under kategorien multifaktorielle sykdommer.

2.4.4 Arvegang ved monogene sykdommer

Monogene sykdommer kan nedarves på ulike måter, avhengig av om begge alleler av et gen må ha sykdomsgivende variant for å forårsake sykdom (recessiv arv), eller om kun ett allel trenger det (dominant arv). Videre skilles det mellom autosomal, kjønnsbundet og mitokondriell arvegang. Se Figur 3.

Dersom den sykdomsgivende varianten forårsaker så alvorlig sykdom at pasienten sjelden eller aldri får egne barn, vil disse kun observeres hos sporadiske pasienter og ikke i familier. Det vil si at varianten(e) ikke er arvet fra en eller begge foreldre, men har oppstått spontant (*de novo*).

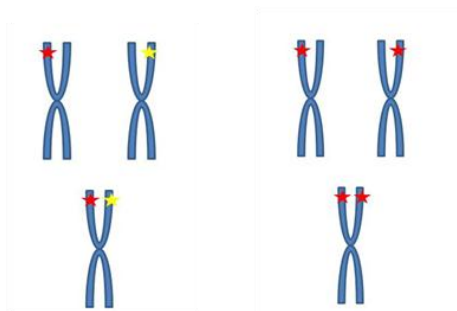


Figur 3. Arvegang, monogen sykdom. Figuren viser ulike typer nedarving av monogen sykdom.

Ved recessiv arv har de affiserte to kopier av en sykdomsgivende variant; en arvet fra far og en arvet fra mor. Foreldrene vil være bærere av sykdommen, men har i tillegg til den sykdomsgivende varianten også en frisk kopi av varianten. Siden bærere har en ikke-affisert variant av genet vil de kunne uttrykke proteinet som det sykdomsassosierte genet koder for. Så lenge det ikke foreligger noen doseavhengighet knyttet til produksjon av proteinet, vil de derfor fremstå som friske. Sannsynligheten for å få barn som er syke vil være på 25 % og sykdommen ses gjerne i søskenflokker.

Recessive varianter er relativt sjeldne, men hver av oss er likevel bærere av et titalls ulike, slike sjeldne varianter (21). Siden disse variantene blir nedarvet mellom generasjonene, vil familier og individer innenfor en familie være bærere av de samme, sjeldne variantene. Recessiv sykdom oppstår oftere når personer med tilknytning til samme familie får barn, siden det da er større sannsynlighet for at barna vil få to kopier av den samme sykdomsgivende varianten.

Ved recessiv arvegang kan man observere en enkelt variant som foreligger i homozygot tilstand eller to ulike heterozygote varianter i samme gen. Sistnevnte kalles compound heterozygot. De to variantene må ligge på hvert sitt allel for å kunne karakteriseres som compound heterozygote. Illustrert i Figur 4.



Figur 4. Compound heterozygot versus homozygot.
I compound heterozygote er to ulike varianter (gul og rød stjerne) lokalisert på hvert sitt allel innenfor samme gen, mens den samme varianten (rød og rød stjerne) finnes på to ulike alleler i homozygote.

Ved dominant arvegang vil en kopi alene av en sykdomsgivende variant være nok til å forårsake sykdom. Så lenge det ikke forekommer redusert penetrans vil det derfor ikke forekomme bærere av denne typen sykdommer. Uttrykket penetrans brukes i genetikken for å beskrive andelen individer som har en bestemt variant og i tillegg uttrykker en assosiert fenotype. Sykdommer med dominant nedarving kan ses i alle generasjoner i en familie. Halvparten av barna til en affisert person vil få sykdommen så lenge partneren er frisk.

Som sagt skilles det også mellom autosomal, kjønnsbundet og mitokondriell arv. Ved autosomal arv er de(n) sykdomsgivende varianten(e) lokalisert til et av autosomene. Autosom refererer til alle kromosomene som ikke er kjønnskromosomer. Ved kjønnsbundet arv er derimot de(n) sykdomsgivende varianten(e) lokalisert til enten X-kromosomet eller Y-kromosomet. Både menn og kvinner kan være affiserte av X-bundne sykdommer, men menn vil oftere være hardere rammet. Sykdommer knyttet til Y-kromosomet kan bare ses hos menn siden kvinner ikke har Y-kromosom. Mitokondriell nedarving skjer dersom de(n) sykdomsgivende variant(en) er lokalisert til gener i mitokondrielt DNA. Mitokondrier overføres kun fra eggceller og nedarving kan derfor bare skje fra mor til barn.

2.5 Genetisk kartlegging av monogene sykdommer

Genetisk kartlegging av monogene sykdommer har tradisjonelt blitt gjort ved bruk av genetiske markører i en koblingsanalyse. Ved monogene sykdommer observeres oftest et tydelig nedarvingsmønster. Genetiske markører (gjerne SNPer eller mikrosatelitter) som er jevnt distribuert utover genomet, vil kunne detektere rekombinasjoner som finner sted innenfor en familie. Følgelig kan markørene benyttes til å se hvilke regioner som kosegreger hos de affiserte individene i en familie. Innenfor disse regionene kan man videre plukke ut kandidatgener som Sangersekvenseres for å identifisere den kausale varianten. For familier med inngifte hvor rescessiv sykdom observeres, har man gjerne brukt en variant av denne metoden kalt homozygositetsmapping.

Begrensninger i bruk av koblingsanalyse ses i tilfeller hvor man har begrenset tilgang på store familier. Problemet er at man da ikke vil ha tilstrekkelig styrke for koblingsanalyse. Ofte er situasjonen slik at man enten har familier med få individer, flere affiserte hvor det ikke er noe slektskap eller sporadiske tilfeller. Alle disse situasjonene vil gi begrenset styrke. Kobling kan heller ikke plukke opp *de novo* varianter. Metoden er videre ikke robust nok til å undersøke sykdommer som er genetisk heterogene; dvs hvor den kausale varianten er tilstede i ulike gener, eller sykdommer som er fenotypisk heterogene; hvor det er vanskelig å stille en klinisk diagnose.

Storskalasekvensering har åpnet muligheten for å finne den genetiske årsaken i slike tilfeller og metoden blir stadig mer teknisk tilrettelagt og kostnadseffektiv. Ved bruk av “eksom-fanging” kan man isolere alle eksonene i genomet (ca 180 000) i én reaksjon i motsetning til

ved tradisjonell sekvensering hvor man ville være nødt til å sette opp et tilsvarende antall PCR-reaksjoner. Deretter kan man ved bruk av storskalasekvensering sekvensere hele eksomet, såkalt eksomsekvensering. Opptil flere mendelske sykdommer har allerede blitt oppdaget med bruk av denne innfallsvinkelen.

2.6 HiSeq 2000

HiSeq 2000 fra Illumina, som er sekvensatoren benyttet i dette prosjektet, har kapasitet til å sekvensere to humane genomer med en dekningsgrad på over 3x i løpet av en kjøring (22). Sekvensatoren er den første kommersielt tilgjengelige platformen som kan sekvensere to humane genomer med nevnte sekvenseringsdybde i en enkelt kjøring til en reagenskostnad på under 10,000 amerikanske dollar per prøve (22). Kjøring av to ulike “flowceller” på samme tid er mulig, noe som øker kapasiteten og fleksibiliteten til platformen. Begrepet “flowcelle” beskrives nærmere under punkt 2.7.3.

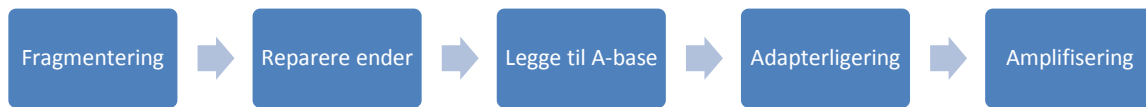
HiSeq 2000 produserer opptil 55 Gb sekvens per dag ved analysering av en 2 x 100 bp kjøring, som er type applikasjon benyttet i dette prosjektet. Ved en slik kjøring vil sekvensatoren lese 100 bp av prøvefragmentene fra to ulike retninger i to omganger. En slik kjøring tar omtrent åtte dager dersom en “flowcelle” benyttes og rundt 11 dager dersom to “flowceller” kjøres samtidig. I sistnevnte tilfelle vil 540 – 600 Gb produseres i løpet av en kjøring. Ved 2 x 100 bp kjøring vil mer enn 85 % av detekterte baser ha en kvalitet høyere enn Q30. Q30 beskriver kvaliteten av “basecalling” (22). Ved Q30 vil 1 av 1000 leste baser være feil, noe som tilsvarer en nøyaktighet på 99,9 %.

2.7 Prinsipp for storskalasekvensering

Før storskalasekvensering kreves en omfattende preparering av prøvene forut for selve sekvenseringen. Hvordan denne prepareringen gjøres avhenger av type sekvensering som skal utføres; om det skal gjøres helgenomsekvensering, eksomsekvensering, mikroRNA-sekvensering, ChIP-seq (Chromatin Immunoprecipitation sequencing) eller annet. Også hvilket instrument som velges for å gjennomføre sekvenseringen er avgjørende for type preparering. Alle disse krever ulik forberedelse. Her beskrives videre preparering for eksomsekvensering og sekvensering ved bruk av HiSeq 2000 fra Illumina.

2.7.1 Prøvepreparering

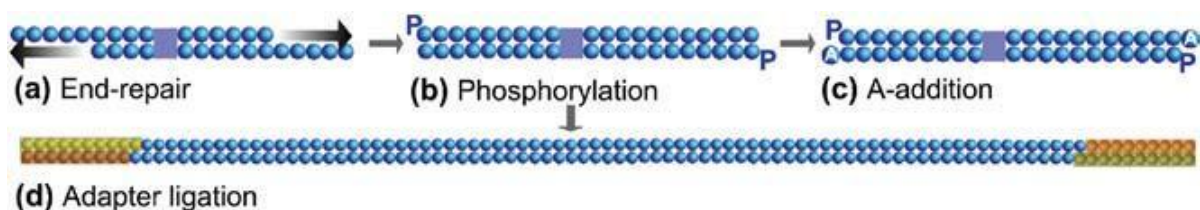
For preparering til eksomsekvensering, som er benyttet i dette prosjektet, er prosessen som vist i Figur 5, før man går videre med selve “eksom-fanging”-prosedyren (exome capture) og sekvenseringen.



Figur 5. Prøvepreparering forut for “eksom-fanging”.

Først blir gDNA fragmentert til fragmenter av en viss størrelse. Størrelsen avhenger av hvilken type protokoll som skal gjennomføres, men vil ofte være i størrelsesorden 200-300 bp. Eventuelle enkeltrådede DNA overheng som er resultat av fragmenteringen blir så fjernet fra fragmentene ved bruk av enzymer. Overheng i 3'-ende blir fjernet ved bruk av enzym med 3' til 5' eksonukleaseaktivitet mens overheng i 5'-ende blir fylt inn ved hjelp av en polymerase som katalyserer syntetisering av ny tråd komplementær til overhengen (Figur 6(a)).

I neste om gang blir en a-base festet til 3'-ende av DNA-fragmentene ved bruk av enzym med 5' til 3' polymeraseaktivitet samt 3' til 5' eksonukleaseaktivitet (Figur 6 (c)). I en etterfølgende ligeringsreaksjon vil denne basen fungerer som templat for adaptere med et enkelt t-base overheng i sin 3'-ende. DNA-ligase driver ligeringsreaksjonen ved å danne en binding mellom 3'hydroksylenden i en nukleotid med 5'fosfatenden i en annen nukleotid (Figur 6 (d)).



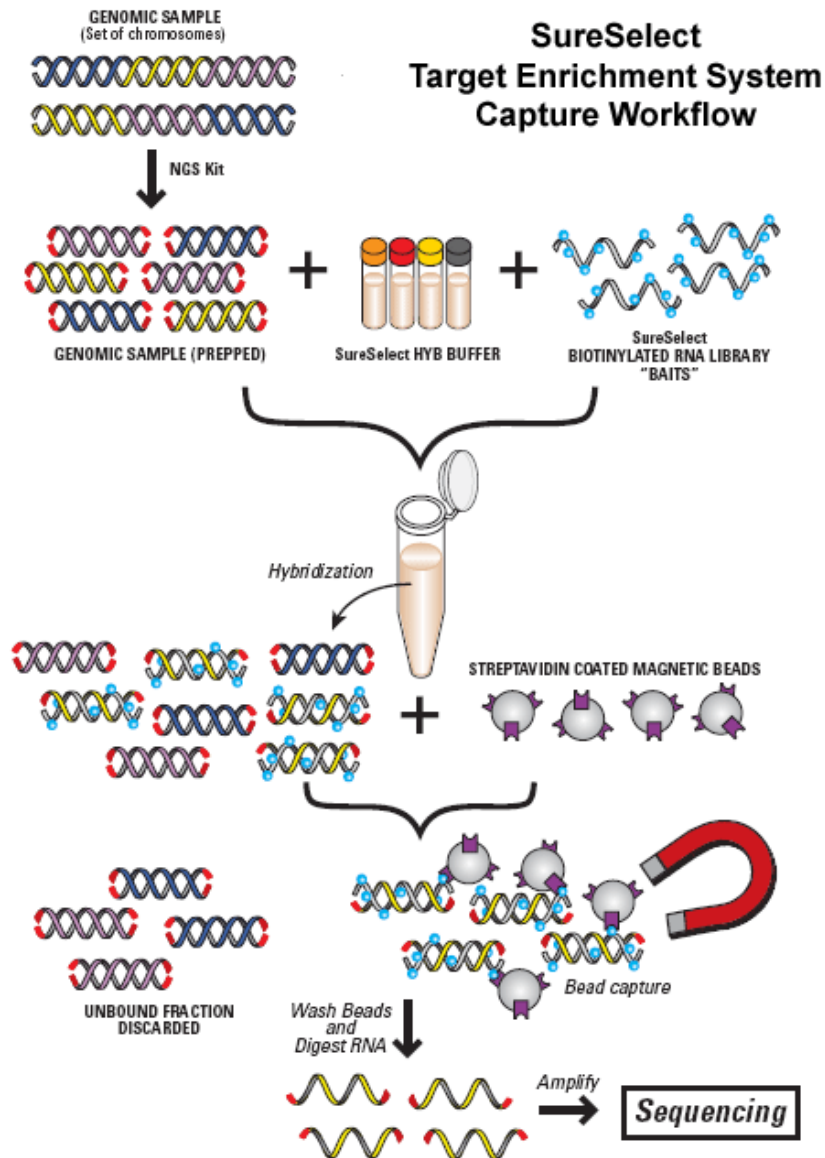
Figur 6.
Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reprodusert med tillatelse (23).

DNA-fragmenter med festet adapter blir så amplifisert ved bruk av PCR for å produsere nok utgangsmateriale for “eksom-fanging”. Man har nå fått laget et representativt utvalg fragmenter av en viss størrelse fra hele genomet.

2.7.2 “Eksom-fanging”

For kun å se på genomets kodende del må man hente ut disse DNA sekvensene spesifikt, dvs. gjøre en såkalt “eksom-fanging”. Man benytter seg da av prober som er spesifikt rettet mot kodende deler av genomet, dvs. eksonene. Amplifiserte DNA-fragmenter blandes med et bibliotek bestående av RNA oligonukleotider merket med biotin. Et bibliotek beskriver en samling segmenter av et genom. Biblioteket vil bestå av kopier av alle disse segmentene.

Oligonuklotidene i biblioteket er komplementære til sekvenser i kodende deler av DNA-fragmentene og vil binde seg til disse områdene under hybridisering. Streptavidinmerkede magnetiske kuler trekker så fragmentene som har festet seg til biotinmerkede oligonukleotider ut ved at det oppstår en streptavidin-biotin binding mellom de to komponentene. Dette er grafisk fremstilt i Figur 7.



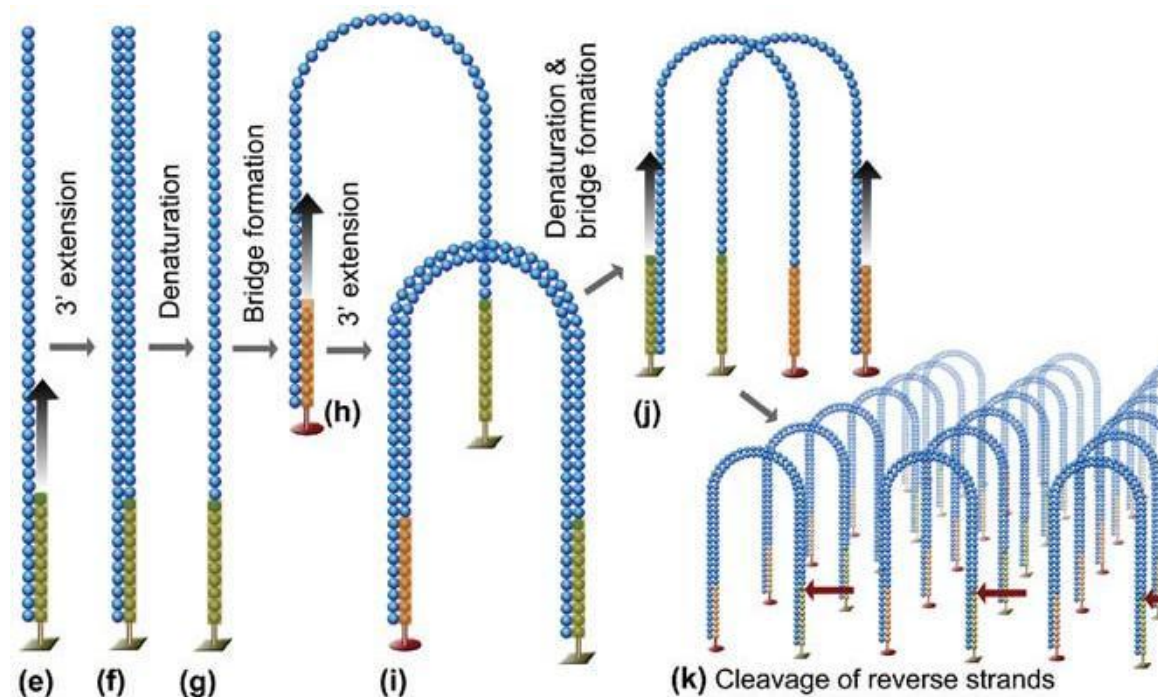
Figur 7. Prinsipp for “eksom-fanging”.
Copyright Agilent Technologies. Reproduert med tillatelse.

2.7.3 Sekvensering

“Clustering”

Adapterene som ble inkorporert i DNA-fragmentene under prøveprepareringen er komplementære til oligonukleotider festet til overflaten av en “flowcelle” og hybridiseres til disse. En “flowcelle” er en liten glassbrikke som er dekket med et tett teppe av oligonukleotider. Den er delt inn i åtte kolonner som hver fungerer som “prøveholder” for ulike prøver under sekvenseringen. Oligonukleotidene fungerer som primere for en polymerase som kopierer det festede DNA-fragmentet. Etter kopiering vil det nå

dobbeltrådig DNAet bli denaturert til to enkeltrådede. Det originale fragmentet vil bli fjernet mens kopien som ble laget blir tilbake og vil være immobilisert til “flowcelle”-overflaten som illustrert i Figur 8 (e-g).



Figur 8.
Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduert med tillatelse (23).

Videre amplifisering baserer seg på “Bridge-PCR” hvor templatene bøyer seg over slik at adapteren i den “frie” enden av fragmentet kan binde sin komplementære oligo et annet sted på “flowcelle”-overflaten. Denne adapteren vil så fungere som primer for en polymerase som lager en ny kopi ved inkorporering av nukleotider. Etter kopiering blir dobbeltrådig DNA denaturert ved bruk av formamid til enkeltrådet DNA. Dette skjer gjentatte ganger for hvert templat og vil resultere i millioner av individuelle, tette klynger (“clustere”) som hver inneholder omtrent 2000 molekylar. Denne prosessen blir kalt “clustering” og illustreres i Figur 8 (h-j).

Linearisering, blokkering og hybridisering

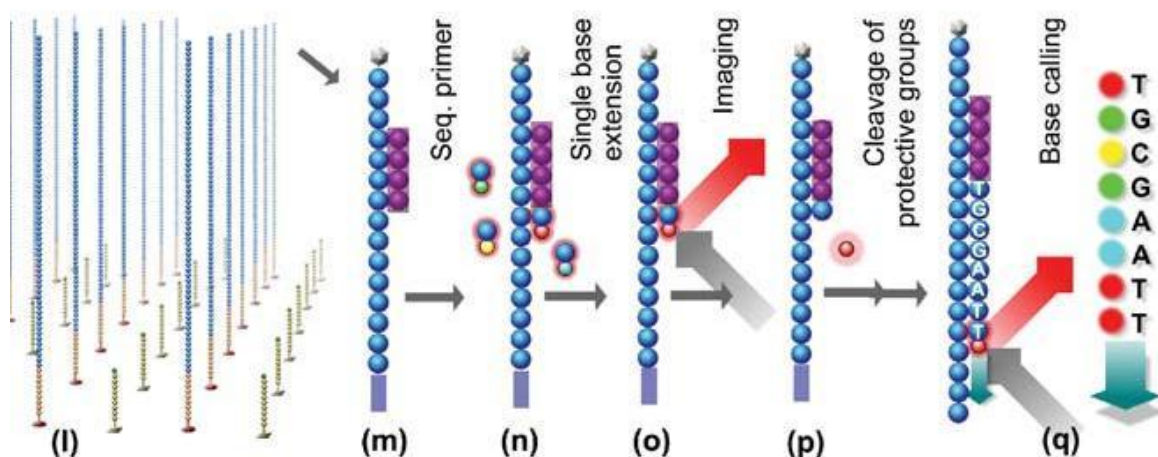
Etter endt “clustering” blir reverstråden kuttet ved primersetet, denaturert og selektivt fjernet fra overflaten av “flowcellen” (Figur 8 (k)). Kun “forward”-tråd blir igjen på “flowcellen” for sekvensering. Denne prosessen kalles linearisering. Videre skjer en blokkering av fri 3’hydroksylende av fragmentene for å forhindre uspesifikk priming.

Sekvenseringsprimer komplementær til adapterne på DNA-fragmentene, hybridiseres så til adapteren som sitter i den enden av fragmentet som ikke er festet til “flowcelle”-overflaten. Se Figur 8 (k) og Figur 9 (l-m).

Sekvensering

“Flowcellen” overføres til HiSeq 2000 hvor alle “clusterene” på “flowcelle”-overflaten vil bli sekvensert parallelt. I en enzymreaksjon bygges en tråd komplementær til DNA-templatene. Fluorescensemerkede nukleotider blir inkorporert i riktig rekkefølge avhengig av templatet. De fire ulike nukleotidetyperne (A, C, G, T) er merket med hver sin spesifikke farge og et kamera kan registrere hvilken base som er festet utifra fluorescensemerkingen. Etter at avlesningen er ferdig, blir fluoroforen enzymatisk fjernet fra nukleotiden og en ny nukleotid kan binde seg. Dette gjentas i et valgfritt antall repeterte syklar. Prosessen er illustrert i Figur 9 (n-q).

I tillegg til å være fluorescensemerket, er nukleotidene også reversible terminatorer. De har et modifisert motiv i 3’hydroksylenden som fungerer som terminator for polymeriseringen og forhindrer at tråden bygges videre etter at en nukleotide har blitt inkorporert. Det modifiserende motivet vil bli fjernet gjennom en enzymatisk reaksjon etter avlesning i hver syklus slik at neste nukleotide kan binde seg.

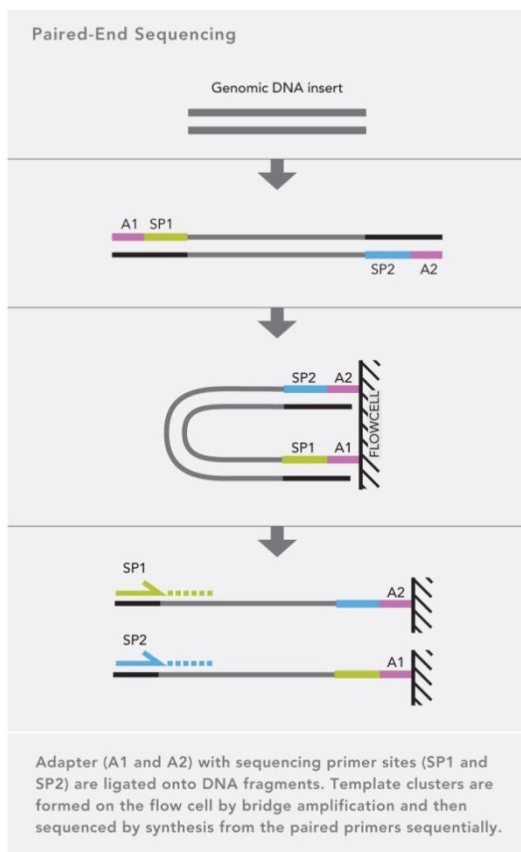


Figur 9.
Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduert med tillatelse (23).

“Paired-End” sekvensering

For å lese fragmentene fra motsatt ende, må to forskjellige adaptore være ligert til DNA-fragmentene. Hver av disse er komplementære til hver sin primer; en som benyttes i den første sekvenseringen og en annen som tilsettes ved sekvensering fra motsatt ende. Dette blir kalt “Paired-End” sekvensering. Illustrert i Figur 10.

Etter at “forward”-tråd er ferdig sekvensert vil den nylagede tråden fra sekvenseringen bli fjernet ved denaturering. Modifiserende motiv i 3’ ende blir fjernet og en ny “Bridge-PCR” blir gjennomført for dannelse av nye dobbeltrådede “clustere”. Denne gangen vil “forward”-tråd bli selektivt fjernet slik at det i neste runde er reverstråden som vil bli sekvensert.



©2011, Illumina Inc. All rights reserved.

Figur 10. “Paired-End” sekvensering.

3 Formål

Formålet med prosjektet er å lokalisere en sykdomsgivende mutasjon i et gen eller et område av det humane genomet som kan knyttes opp mot utviklingen av en tilstand som viser såkalt “apple peel” atresi (endringer i tarm), øyemisdannelser og mikrocefali (unormalt liten hjerne).

Ved å finne en genetisk årsak til denne tilstanden, vil pasienter som har levd lenge uten diagnose kunne diagnostiseres. Av erfaring har man sett at personer med ukjent sykdom vil kunne slå seg mer til ro med tilstanden sin når de får kjennskap til årsaken til denne. Man får svar på spørsmål man kanskje har stilt seg gjennom lang tid om hvordan og hvorfor.

Dersom disse symptomene skulle vise seg å stamme fra en felles genetisk årsak, vil man kunne gruppere individer med genetiske endringer tilsvarende denne. En slik klassifisering og sammenligning av individer vil i fremtiden kanskje kunne si mer om prognosen for personer som rammes slik at man på et tidlig stadium kan gjøre tilrettelegginger knyttet til deres behov.

En kjent genetisk årsak vil også forenkle diagnostiseringen for klinikere som i fremtiden vil møte pasienter med et lignende sykdomsbilde. Videre vil kjent genetisk årsak muliggjøre både bærertestning og prenataltesting i rammede familier.

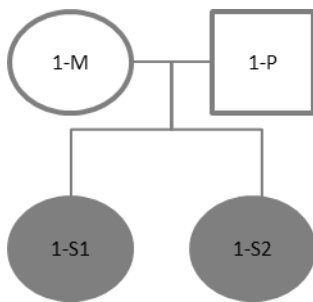
Kunnskap om hvilket gen som forårsaker sykdom kan også bidra til økt forståelse av mekanismene rundt sykdomsutviklingen og i beste fall lede til ny forskning for om mulig å identifisere nye muligheter for behandling.

4 Materialer og metoder

4.1 Materiale

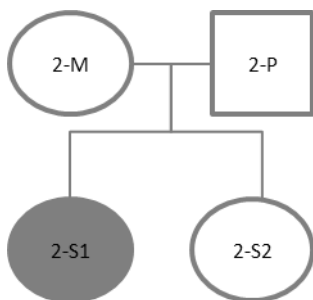
Familie 1 ble identifisert av Petter Strømme ved Rikshospitalet i Oslo i 1993. Han henviste pasientene til Avdeling for medisinsk genetikk for videre utredning. Familie 2 ble først beskrevet i litteraturen i 2007 av Yolande van Bever ved Erasmus Medical Centre i Rotterdam. Petter Strømme opprettet kontakt med van Bever og videreformidlet denne kontakten til Avdeling for medisinsk genetikk, OUS i forbindelse med oppstart av prosjektet. DNA-materiale fra de to kjernefamiliene ble samlet inn.

Familie 1 består av mor, far og to affiserte barn som illustrert i Figur 11. Sirkel angir kvinne/jente, firkant mann/gutt. Skravert symbol angir sykdom.



Figur 11. Familie 1.

Familie 2 består av mor, far, et affisert og et friskt barn som illustrert av familietre i Figur 12. Prøvemateriale fra det friske barnet fikk vi ikke tilgang til.



Figur 12. Familie 2.

Det ble søkt om tillatelse for gjennomføring av prosjektet fra Regional Etisk Komité og Helsedirektoratet. Begge instanser ga sin godkjennelse. Familiene som ble inkludert ble forespurt om deltagelse via sin behandlende lege. De inkluderte pasientene fra den norske

familien har fått genetisk veiledning ved Avdeling for medisinsk genetikk, OUS og skriftlige samtykker er innhentet. Når det gjelder den nederlandske familien, er informasjon om prosjektet samt innhenting av samtykker håndtert av prosjektdeltagernes behandlende lege i Nederland.

4.2 Ekstrahering av DNA

DNA ble ekstrahert fra EDTA-fullblod for alle individene inkludert i studien. Roboten Autopure LS (Qiagen) som baserer seg på automatisert utvinning av genomisk DNA gjennom saltfelling, samt robotsystemet MagnaPure LC (Roche Applied Science) basert på magnetiske kuler ble benyttet. Protokollene “Comp Lysate Prep” og “Comp Lysate Finish” utarbeidet av leverandør ble brukt for å gi optimalt utbytte ved bruk av Autopure LS mens leverandørens standardprotokoll DNA LV Blood_1000 ble benyttet ved bruk av MagnaPure LC.

4.3 DNA-kvantitering/konsentrasjonsmåling

For å bestemme utbyttet og kvaliteten av ekstrahert DNA ble prøvene målt på NanoDrop 8000 spektrofotometer fra Thermo Scientific hvor en spesiallampe kalt “Xenon flash lamp” måler absorbanse i prøvematerialet. 1,5 uL DNA ble brukt under måling.

Til estimering av DNA-konsentrasjon for input i prøvepreparering til storskalasekvensering samt av DNA-bibliotek etter prøvepreparering ble Qubit Quantitation Platform benyttet. Plattformen baserer seg på fluoroforer som blir fluoriserende etter binding av DNA, RNA eller proteiner. Kitene Quant-IT™ dsDNA BR Assay Kit og Quant-IT™ dsDNA HS Assay Kit ble brukt.

4.4 Array comparative genomic hybridization, aCGH

Begge de affiserte i familie 1 (Fam1_II-2 og Fam1_II-1) ble analysert på aCGH, 105k fra Agilent Technologies for å sjekke hvorvidt pasientene kunne ha submikroskopiske kromosomavvik som årsak til sykdommen. Ved aCGH blir pasientens DNA og et normalt, referanse DNA merket med hver sin fluorofor. Disse blir så hybridisert mot et stort antall prober rettet mot det humane genom. Deretter beregnes ratioen i fluorescensintensitet mellom

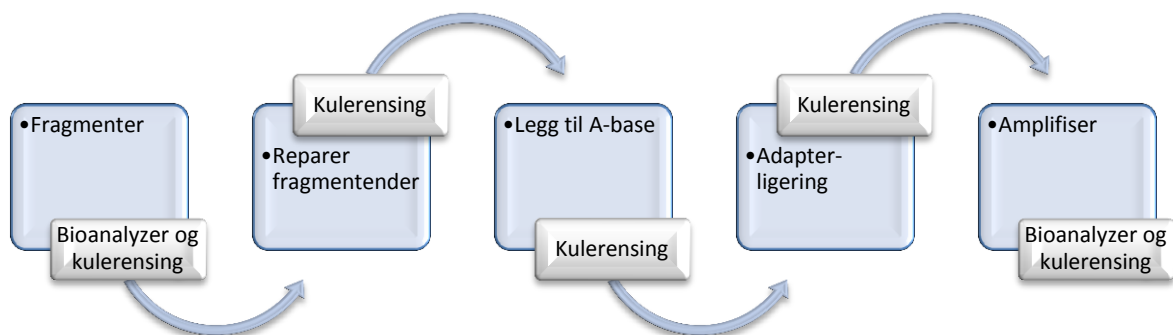
prøven og referansen for å sjekke for eventuelle doseavvik. Sterkere intensitet hos pasientens DNA kan tyde på duplikasjon mens svakere intensitet kan tyde på delesjon.

aCGH analysen ble utført ved bruk av standard protokollen “Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis” v4.0. Direkte metode for aCGH workflow ble benyttet. Microarray Hybridization Oven fra Agilent Technologies ble brukt i hybridiseringen og GenePix4000B til scanning av arrayene. I resultatbehandlingen ble software Feature Extraction (bearbeidelse av rådata) og DNA Analytics v4.0 brukt.

Pasientene i familie 1 hadde i tillegg blitt undersøkt ved karyotyping forut for oppstart av dette prosjektet.

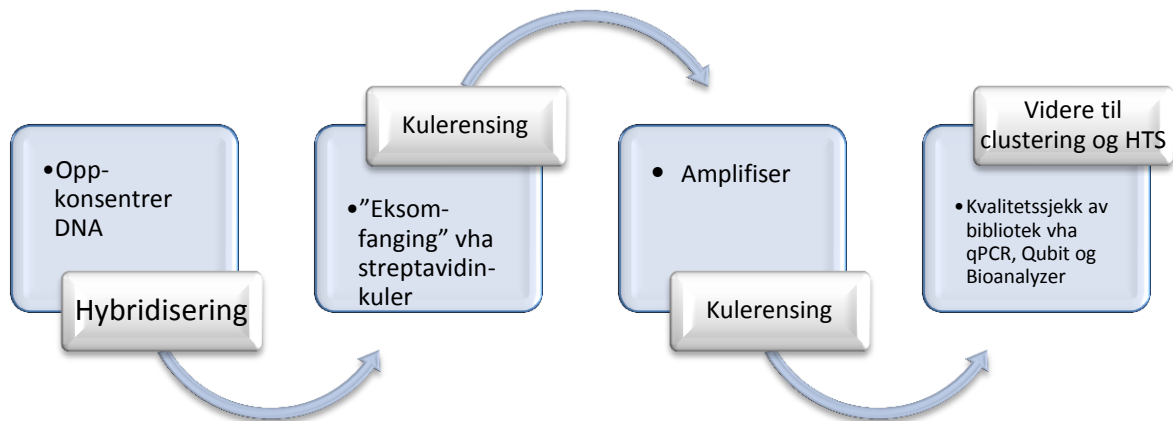
4.5 Prøvepreparering for storskalasekvensering

Forut for storskalasekvensering ble prøvematerialet preparert. Prepareringen består av en rekke reaksjoner som vil bli beskrevet i punkt 4.5.1 til 4.5.3 og som illustrert i Figur 13



Figur 13. Prøvepreparering forut for “eksom-fanging”.

For videre preparering, inkludert isolering av de kodende delene av genomet, ble preparert prøvemateriale fra forrige prosess behandlet videre som beskrevet i punkt 4.5.4 og 4.5.5 og som illustrert i Figur 14.



Figur 14. "Eksom-fanging" og videre prøvepreparering.

4.5.1 Fragmentering

3 μg gDNA i et volum på 120 μL ble fragmentert til fragmenter av størrelsesorden 150-200 bp ved hjelp av Covaris S2 fra Covaris Inc ved bruk av følgende innstillinger: "Duty cycle" = 10 %, "Intensity" = 5, "Cycles per burst" = 200, "Time" = 6 cycles a 60 sec, "Set mode" = frequency sweeping og "Temperature" = 4 – 7 °C.

Covaris S2 baserer seg på Adaptive Focused Acoustic (AFA) teknologi hvor fokuserte, energibølger blir sendt mot prøvematerialet. Når denne energien beveger seg gjennom prøven vil det dannes bobler fra løste gasser som finnes naturlig forekommende i prøven. Fjerning av energien medfører kollaps av disse boblene noe som skaper et sjokkbølgemiljø hvor sterke krefter vil forårsake fragmentering av DNAet.

For å sjekke for vellykket fragmentering av DNAet ble 1 μL av prøvematerialet kjørt på Agilent 2100 Bioanalyser fra Agilent Technologies med bruk av High sensitivity DNA kit.

4.5.2 SPRIWorks

Til den videre prepareringen ble roboten SPRIWorks fra Beckman Coulter benyttet.

Kulerensing

Her ble fragmentene først rensset ved bruk av Agencourt AMPure XP kuler. 120 μL fragmentert DNA ble blandet med 180 μL AMPure kuler og satt til inkubering i fem minutter ved romtemperatur. Løsningen ble så overført til magnet for videre inkubering i ytterligere

fem minutter før supernatanten ble fjernet. Deretter ble kulene vasket to ganger med 500 μL 70 % etanol. Etter siste vask ble all etanol fjernet før 30 μL nukleasefritt vann ble tilsatt og prøvene satt til inkubering i to minutter ved romtemperatur. Prøvene ble deretter satt på magnet i tre minutter hvorpå supernatanten ble overført til et nytt rør for videre preparering.

Kulerensingen baserer seg på at DNA er løselig i vandige løsninger. I de første stegene, hvor etanolholdig væske blir benyttet, vil DNA være bundet til kulene. Vann som tilsettes i det siste steget av rensingen vil fungere som elueringsbuffer for DNA som da løsner fra kulene. DNAet må være av en viss størrelse (~100 bp) for å binde seg til kulene. Alle ikke-inkorporerte dNTPer, primere, primerdimerer og salter kan derfor effektivt fjernes uten at man mister DNAet i løsningen.

Reparer fragmentender

Videre ble det gjort “end-repair” av fragmentene. 29 μL prøvemateriale ble blandet med følgende reaksjonsmiks:

Reagens	Volum (μL)
Nukleasefritt vann	46
T4 DNA ligasebuffer med 10 mM ATP	10
10 mM dNTP mix	4
T4 DNA Polymerase	5
Klenow DNA Polymerase	1
T4 PNK	5

og satt til inkubering i 30 minutter ved 20 °C.

I denne reaksjonen fungerer enkelttrådede overheng på DNA-fragmentene som templat for nukleotider i reaksjonen. Klenow DNA polymerase bidrar i denne reaksjonen og gir dannelse av “blunt-ends” på DNA-fragmentene.

Etter “end-repair” ble det gjort en ny kulerensing med AMPure XP kuler etter samme prosedyre som beskrevet i starten av avsnittet. Siste eluering ble gjort i et volum på 32 μ L nukleasefritt vann.

Legg til A-base

I neste omgang ble en A-base lagt til fragmentene ved å blande 32 μ L prøvemateriale med følgende reaksjonsmiks:

Reagens	Volum (μ L)
10x Klenowbuffer	5
10 mM dATP	10
Klenow fragment (3' to 5' exo minus)	3

og inkubert ved 37 °C i 30 minutter.

I denne reaksjonen vil Klenow fragment (enzym med 5-3 polymeraseaktivitet + 3-5 exonukleaseaktivitet) bidra til å legge til en A-base til 3'enden av alle DNA-fragmentene.

Etter fullført inkubering ble det gjort en ny AMPure XP kulerensing etter samme prosedyre som beskrevet over bortsett fra siste eluering som ble gjort i et volum på 15 μ L nukleasefritt vann.

Adapterligering

For videre adapterligering ble 14 μ L av prøvematerialet inkubert i 15 minutter ved 20 °C blandet med følgende reaksjonsmiks:

Reagens	Volum (μ L)
2x Ligasebuffer	25
Paired-end adapter oligomix	6
DNA Ligase	5

DNA-ligase vil i denne reaksjonen ligere adaptere av kjent sekvens (komplementære til oligoer på “flowcelle”) til a-enden som ble lagt på fragmentene i forrige steg. Ligasen gjør dette ved å danne binding mellom 3’hydroksylende i en nukleotid med 5’fosfatende i en annen.

En ny AMPure kulerensing ble gjort etter fullført inkubering i henhold til prosedyren beskrevet i starten av avsnittet. Siste eluering ble gjort i et volum på 32 µL nukleasefritt vann. Bruk av SPRIWorks pipetteringsrobot ble avsluttet her og videre prøvepreparering ble gjort manuelt.

Preparert prøvemateriale fra SPRIworks ble ytterligere fortynnet med 20 µL nukleasefritt vann.

4.5.3 PCR

Det ble gjort PCR på 25 µL av prøvematerialet ved bruk av følgende reaksjonsmiks og PCR-betingelser:

Reagens	Volum (µL)
Nukleasefritt vann	11,5
Illumina PCR primer PE 1.0	1
Illumina PCR primer PE 2.0	1
Herculase II 5x Reaksjonsbuffer	10
dNTP	0,5
Herculase II polymerase	1

98 °C 30 sek

98 °C 10 sek

65 °C 30 sek |x4

72 °C 30 sek

72 °C 5 min

4 °C ∞

PCR-primere er benyttet i reaksjonen er komplementære til adaptersekvensen som ble ligert på prøvematerialet som beskrevet under forrige steg. dNTPer fungerer som byggesteiner i reaksjonen hvor polymerase katalyserer dannelsen av templatkopier.

PCR-produktene ble så rensert med AMPure XP kuler i henhold til samme prosedyre som beskrevet i punkt 4.5.2 bortsett fra siste eluering som ble gjort i et volum på 30 µL nukleasefritt vann.

For å sjekke hvorvidt de foregående reaksjonene hadde fungert som forventet ble 1 µL av prøvematerialet kjørt på Agilent 2100 Bioanalyser fra Agilent Technologies med bruk av DNA 1000 assay. I tillegg ble 1 µL PCR-produkt brukt til konsentrasjonsmåling på Qubit Fluorometer ved bruk av dsDNA BR Assay Kit.

4.5.4 “Eksom-fanging”

Forut for hybridisering ble preparert DNA oppkonsentrert ved bruk av Savant Automatic Environmental SpeedVac System AES1010 fra Thermo Scientific. 25,5 µL prøvemateriale ble kjørt i totalt 2 timer og 45 minutter hvorav 15 minutt med infrarødt lys. Sentrifugen ble kjørt med innstillingene “medium drying rate”, “full vacuum”. Tørket prøvemateriale ble så resuspendert i 3,4 µl nukleasefritt vann.

Etter resuspensjon ble 3,4 µL prøvemateriale blandet med “SureSelect Block Mix” tilberedt på is etter følgende oppsett:

Reagens	Volum (µL)
SureSelect Block #1	2,5
SureSelect Block #3	2,5
SureSelect Block #3	0,6

Blandingen ble kjørt i PCR-maskin med følgende temperaturinnstillinger:

95 °C 5 min

65 °C ∞

Hybridiseringsbuffer ble preparert etter følgende mix:

Reagens	Volum (μL)
SureSelect Hyb #1	25
SureSelect Hyb #3	1
SureSelect Hyb #3	10
SureSelect Hyb #4	13

og inkubert ved 65 °C i fem minutter.

SureSelect capture librarymix ble tilberedt på is etter følgende oppsett:

Reagens	Volum (μL)
SureSelect capture library	5
Rnase Block (fortynnet 1 uL + 2 uL vann)	2

og inkubert ved 65 °C i to minutter.

13 μL varm hybridiseringsbuffer ble overført til SureSelect library mix ved 65 °C og blandet godt før prøvematerialet ble tilsatt. Hybridisering ble så gjort ved 65 °C i 24 timer.

For å “fiske ut” målregionene etter hybridisering ble 50 μL streptavidinmerkede magnetiske kuler (DynaL MyOne Streptavidin T1) blandet med 200 μL SureSelect Binding buffer og satt på magnet for så å fjerne supernatanten. Dette ble gjentatt tre ganger for ordentlig vask av kulene. Etter siste vask ble kulene resuspendert i 200 μL SureSelect Binding buffer.

Prøvematerialet fra hybridisering (25 μL) ble overført til de vaskede kulene og inkubert ved romtemperatur i 30 minutter på en rister. Prøvemiksen ble så satt på magnet for fjerning av supernatant. Kulene ble deretter resuspendert i 500 μL SureSelect Wash Buffer #1 og inkubert i 15 minutter ved romtemperatur. Deretter ble prøvene overført til magnet og supernatant fjernet. Kulene ble så vasket tre ganger ved å tilsette 500 μL SureSelect Wash Buffer #2 forvarmet til 65 °C, inkubert i 10 minutter ved 65 °C og satt på magnet for fjerning av

supernatant. Videre ble kulene resuspendert i 50 μ L SureSelect Elution Buffer og inkubert ved romtemperatur i 10 minutter. Ved bruk av magnet ble så kulene holdt igjen mens supernatanten (mål-DNA) ble separert for videre bruk. 50 μ L SureSelect Neutralization Buffer ble tilsatt.

Prøvematerialet ble så rensert ved bruk av AMPure XP kuler i henhold til samme prosedyre som beskrevet i punkt 4.5.2. Eluering ble gjort i et volum på 30 μ L nukleasefritt vann.

4.5.5 PCR av “fanget” bibliotek

Det ble videre gjort amplifisering av 15 μ L “fanget” DNA ved bruk av følgende reaksjonsoppsett og PCR-betingelser:

Reagens	Volum (μ L)
Nukleasefritt vann	22,5
5x Herculase II Reaction Buffer	10
dNTPmix (25 mM av hver)	0,5
Herculase II Fusion DNA Polymerase	1
SureSelect GA PCR primers	1

98 °C 2 min

98 °C 20 sek

60 °C 30 sek |x12

72 °C 30 sek

72 °C 5 min

4 °C ∞

PCR-produktene ble så rensert med Ampure XP kuler i henhold til samme prosedyre som beskrevet under punkt 4.5.2. Eluering ble gjort i et volum på 30 μ L nukleasefritt vann.

4.5.6 Kvalitetskontroll av bibliotek

Bioanalyser og Qubit

For å sjekke hvorvidt de foregående reaksjonene hadde fungert som forventet ble 1 μL av prøvematerialet kjørt på Agilent 2100 Bioanalyser fra Agilent Technologies med bruk av High Sensitivity DNA assay. I tillegg ble det gjort en konsentrasjonsmåling av PCR-produktene ved bruk av Qubit Fluorometer, dsDNA BR Assay Kit. 1 μL materiale ble brukt i målingen.

Basert på målt konsentrasjon i $\text{ng}/\mu\text{L}$ gjort på Qubit, samt fragmentstørrelse i bp fra Bioanalyser, ble 25 μL av et 10 nM bibliotek laget. Beregning ble gjort ved bruk av et excelbasert kalkuleringsprogram gitt av Illumina. For prøve Fam1_II-2 ble 6,7 μL preparert bibliotek blandet med 18,3 μL EB-buffer mens 4,4 μL preparert bibliotek ble blandet med 20,6 μL EB-buffer for prøve Fam1_II-1.

qPCR

For en ytterligere kvalitetssjekk av biblioteket forut for sekvensering, ble ferdig preparert, 10 nM DNA-bibliotek sjekket ved bruk av qPCR.

Sjekk av “eksom-fanging”

Prøvemateriale, både pre og post “eksom-fanging”, ble sjekket mot fire primersett fra Nimblegen (NSC-0237, NSC-0247, NSC-0268 og NSC-0272). For primersekvenser se Tabell 6. Et normalt gDNA ble fortynnet til konsentrasjoner på 10, 1, 0,1 og 0,01 $\text{ng}/\mu\text{L}$ og brukt som standardkurve.

En aliquot av prøvemateriale fra henholdsvis pre og post “eksom-fanging” ble fortynnet til 0,5 $\text{ng}/\mu\text{L}$ med dH_2O . Fire parallelle forsøk ble satt opp (en for hver primermix) hvor 2 μL prøvemateriale ble blandet med følgende reaksjonsmiks:

Reagens	Volum (μL)
SYBR Green Master (2x)	5
Primermix (10x)	1
dH ₂ O	2

og kjørt på 7900HT Fast Real-Time PCR System fra Applied Biosystems med følgende temperaturinnstillinger:

50 °C 2 min

95 °C 10 min

95 °C 15 sek

60 °C 1 min |x40

HTS SYBR ble brukt som detektor. Alle assay ble kjørt i tre paralleller og en negativ kontroll ble inkludert for hver av assaymiksene.

qPCR assayene er rettet mot loci som skal bli fanget under “eksom-fanging”-prosedyren. Ved sammenligning av DNA konsentrasjon for prøven i amplifisert prøvebibliotek før og etter “eksom-fanging”, får man et estimat på hvorvidt “eksom-fangings”-prosessen har fungert.

Sjekk mot referanse

Videre ble ferdig 10 nM bibliotek sjekket mot et kjent, velfungerende DNA-bibliotek, WEF1-2.

Ferdig klargjort bibliotek samt kontroll bibliotek (WEF1-2) ble fortynnet 1/1000 med dH₂O. 2 μL av hver fortynning ble blandet med følgende mix:

Reagens	Volum (μL)
SYBR Green Master (2x)	5
Primermix (SYB_FP6 / SYB_RP8)	1
dH ₂ O	2

og kjørt på 7900HT Fast Real-Time PCR System fra Applied Biosystems med følgende temperaturinnstillinger:

50 °C 2 min

95 °C 10 min

95 °C 15 sek

60 °C 1 min |x40

HTS SYBR ble brukt som detektor. I denne testen ble PhiX, fortynnet til 10, 1 og 0,1 pmol, brukt som standardkurve. SYB_FP6 / SYB_RP8 ble brukt som primersett. For primersekvenser se Tabell 6. Det ble tilberedt og kjørt tre paralleller av prøvene og kontrollbiblioteket samt tre paralleller av hver PhiX-fortynning. I tillegg ble det kjørt en negativ kontroll for oppsettet.

Primermiksen har sekvens som er komplementær til 3'adapter og Illumina PCR primer 1 som er lagt til prøvefragmentene under prøveprepareringen til storskalasekvensering. Ved bruk av denne primermiksen vil kun fragmenter som inneholder hele adapteren og PCR-forlengelsen (primeren) bli kvantifisert i qPCRen. Fragmenter med adapter- eller primersekvens som har blitt degradert vil ikke binde seg til "flowcellen" og kan bidra til dårlig utbytte av sekvenseringen. Disse vil gi en falsk for høy konsentrasjon om man kun baserer seg på konsentrasjonsmålinger til beregning av mengde bibliotek som skal benyttes i sekvenseringen. Ved bruk av qPCR blir slike "falsk positive" fragmenter eliminert i beregningen.

Til analysering av alle qPCR-resultater ble programvaren SDS 2.4 fra Applied Biosystems by Life Technologies benyttet.

På bakgrunn av resultatene fra qPCR kunne optimal mengde prøvemateriale til bruk i "clustering" beregnes som vist i punkt 5.3 ved å sammenligne Ct-verdi (crossing treshold) for preparert prøvebibliotek mot Ct-verdien for kontrollbibliotek.

4.5.7 “Clustering”

“Clustering” ble gjort på cBot “cluster”-stasjon. Prøvene ble preparert ved å fortynne 2 µL 10 nM prøvebibliotek med 8 µL Buffer-EB. Videre ble de denaturert ved å tilsette 10 µL NaOH (2M fortynnet 1:10 med dH₂O) og hensatt i romtemperatur i 5 minutter. 6,8 µL denaturert prøvemateriale fra prøvenummer Fam1_II-2 ble blandet med 993 µL hybridiseringsbuffer mens 7,7 µL prøvemateriale og 992 µL hybridiseringsbuffer ble benyttet for prøve Fam1_II-1.

“cBot Cluster Generation Reagent Plate” og “HiSeq PE Flow Cell v1.5” ble satt på cBot sammen med 140 µL prøvemateriale for “clustering”. Fam1_II-2 ble avsatt i rad seks på “flowcellen” og Fam1_II-1 i rad syv. I tillegg ble kontrollen PhiX tilsatt i rad 5. Protokollen PE_Amp_Lin_Block_Hyb_v7.0 ble brukt og eksperimentet fikk navnet “Berge-Brandal-Sorte-excap-20110609”.

4.5.8 Storskalasekvensering

Etter fullført “clustering” ble “flowcellen” overført til HiSeq 2000 for sekvensering. Reagenskitene “TrueSeq SBS Kit - HTS” boks en og to ble også satt i maskinen. Her ble det gjort noe reagenspreparering i forkant. Innholdet fra to like 50 syklers kit ble blandet sammen. LFN36 ble overført til ICR og deretter ble 620 µL LRP tilsatt. 175 µL av reagenset HP3 (NaOH) ble fortynnet i 3325 µL PW1 (dH₂O).

Kjøring på HiSeq 2000 ble så startet med innstillingene “Align to PhiX, lane 5”, “Read 1 = 100 cycles”, “Read 2 = 100 cycles”, “HiSeq Sequencing Kit 200 cycles”, “1st run = 115 cycles” ved bruk av prøveoppsettet “Berge-Brandal-Sorte-excap-20110609”.

Eksperimentnavn ble satt til “Berge-Brandal-Sorte-excap-20110609”.

“Paired-end”-kjøring ble startet fem dager senere etter at maskinen hadde fullført “read 1”. “TrueSeq SBS kit -HTS” ble preparert på tilsvarende måte som forut for første kjøring. I tillegg ble “Illumina Read 2 Cluster Resynthesis Kit” for kjøring av “paired-end” kjemi satt på maskinen.

4.5.9 Bioinformatikk

Etter fullført sekvensering ble sekvensene sammenstilt med kjent genomsekvens, versjon hg19, ved bruk av dataprogrammet Burrows-Wheeler Alignment tool, BWA (24). Sekvenser som ikke kunne sammenstilles mot sekvensregioner inkludert i “eksom-fanging”-kit ble ekskludert. Resultatene fra sammenstillingen ble ved hjelp av dataprogrammet SAMtools (25) komprimert til BAM-filer for videre håndtering. Til “calling” og recalibrering av varianter ble dataprogrammet GATK (Genomic Analysis Tool Kit) benyttet (26). Alle funn ble sjekket mot de fritt tilgjengelige internettdatabasene dbSNP132 (27) og 1000 Genomes (versjon av november2010) (28). Varianter som var rapportert her ble ekskludert fra videre resultathåndtering.

For å lokalisere homozygote varianter som var felles hos begge de affiserte ble programmet VCFtool brukt (29). Et egetprogrammert script ble brukt sammen med VCFtool for å lokalisere felles gener hvor pasientene hadde to eller flere felles mulige sykdomsfremkallende varianter (compound heterozygote). Dette ble også gjort for å lokalisere gener hvor en variant kunne ses hos en av pasientene, men to varianter kunne ses hos den andre. Programmet annoVAR (30) ble brukt til annotering av funn.

Sekvenseringsdybde for alle sekvenserte regioner ble sjekket ved hjelp av programmet BEDTools (31). I tillegg ble gjennomsnittlig sekvenseringsdybde for alle gener i genomet, uavhengig om de var inkludert i “eksom-fangings”-kit eller ikke, sjekket. Til dette ble BEDTools brukt i kombinasjon med et egetprogrammert script.

For manuell gjennomgang av regioner med aktuelle funn ble softwaren IGV 2.0 benyttet (32).

For å vurdere hvorvidt gener som ikke var dekket eller var dårlig dekket av storskalasekvenseringen kunne være aktuelle kandidatgener og prioritere disse, ble programmet Endeavour fra K. U. Leuven og VIB benyttet (33). Gener med dekningsgrad mindre enn 20x ble tatt med.

4.6 Sangersekvensering

4.6.1 Primerdesign

Til design av primere for kjøring av PCR ble programmet Primer3 (34) med standard innstillinger benyttet. Primerne ble lagt minst 30 bp oppstrøms eller nedstrøms for ekson/intron grenser. Ved behov for overlappende primersett i kodende deler, ble overlappende sekvens satt å være minst 50 bp. Settene ble også testet i programmet NCBI nucleotide BLAST (35) for å forsikre om at primerene ikke matchet andre deler av genomet enn det ønskede området.

I primerenes 5'ende ble det inkludert en M13-sekvens for bruk av universelle M13-primere i sekvenseringsreaksjonen. For primersekvenser se vedlegg, Tabell 7.

4.6.2 PCR

PCR ble kjørt med bruk av AmpliTaq Gold 360 DNA Polymerase kit fra Applied Biosystems by Life Technologies. 0,5 μ L DNA med konsentrasjon \sim 50 ng/ μ L ble blandet med følgende reaksjonsmiks:

Reagens	Volum (μ L)
Primer 5 μ M Fwd + 5 μ M Rev	2
GeneAmp dNTP mix with dTTP, 10 mM	1,6
dH ₂ O	10,6
AmpliTaQ Gold 360 Buffer, 10x	2
MgCl ₂ , 25 mM	1,2
360 GC Enhancer	2
AmpliTaQ Gold 360 DNA Polymerase	0,1

Og kjørt på PCR-maskin med følgende innstillinger:

95 °C 10 min

95 °C 30 sek

60 °C 30 sek | x30

72 °C 1 min

72 °C 7 min

10 °C ∞

4.6.3 Rensing av PCR-produkter

Til rensing av PCR produkter ble Agencourt AMPure XP kuler fra Beckman Coulter Genomics benyttet (36). Metoden baserer seg på rensing ved hjelp av kuler hvor man får en effektiv fjerning av ikke-inkorporerte dNTPer, primere og salter (36). Rensingen ble utført på Biomek Fx pipetteringsrobot fra Beckman Coulter Genomics ved bruk av programmet AMPure 96 2.45 – 3.1. Ferdig rensede produkter ble eluert i dH₂O.

4.6.4 Sekvenseringsreaksjon

Sekvenseringsreaksjon ble satt opp med ferdig kit, BigDye Terminator v3.1 Cycle Sequencing Kit fra Applied Biosystems by Life Technologies. 2 µL rensset PCR-produkt ble blandet med reaksjonsmiksen:

Reagens	Volum (µL)
BigDye Terminator v1.1, v3.1 Sequencing Buffer, 5x	2
BigDye Terminator v3.1 Ready Reaction Mix	0,25
M13-primer Fwd/Rev, 3,2 µM	1
dH ₂ O	4,75

Og kjørt på PCR-maskin med følgende innstillinger:

96 °C 10 sek

50 °C 5 sek | x25

60 °C 4 min

10 °C ∞

4.6.5 Rensing av sekvenseringsprodukter

Til rensing av sekvenseringsprodukter ble Agencourt CleanSEQ kuler fra Beckman Coulter Genomics benyttet (37). Her blir dye-terminatorer effektivt fjernet fra reaksjonen ved hjelp av kuler, SPRI (“Solid Phase Reversible Immobilization”) teknologi (37). Rensing ble utført på Biomek Fx pipetteringsrobot fra Beckman Coulter Genomics ved bruk av programmet CleanSEQ 96 2.39 – 3.1. Ferdig rensede produkter ble eluert i 0,05 mM EDTA.

4.6.6 ABI3730xl

Analysing av prøvene ble gjort på ABI3730xl fra Applied Biosystems by Life Technologies installert med 36 cm, 96 kapillær og polymer POP-7. Analyseprotokoll “3730BDTv3 – KB – DeNovo_v5.2” og instrumentprotokoll “StdSeq_140404” med “DyeSet” “Z-BigDye v3” ble brukt under kjøring av instrumentet.

4.6.7 SeqScape

Vurdering av sekvenser etter analyse ble gjort i softwaren SeqScape v2.6 fra Applied Biosystems by Life Technologies.

For bekreftelse av funn etter storskalasekvensering for genene *LTBP4*, *AOAH*, *DIXDC1*, *PKDIL3*, *CENPF*, *GPRC6A*, *GPRINI* og *HAVCRI*, ble posisjonen hvor de aktuelle funnene var lokalisert samt 20 basepar oppstrøms og nedstrøms vurdert. Referansesekvens ble hentet fra UCSC, hg19 ved å søke på aktuell posisjon i genet og legge til 20 basepar oppstrøms og nedstrøms. Tillegg av ekstra baser ble gjort for å gjøre tolkingen lettere. Referansesekvens ble så importert til SeqScape hvor sammenligning mot forsøksindividenes sekvens ble utført. Default innstillinger for analysing i SeqScape ble benyttet.

Til fullsekvensering av *CENPF* ble referansesekvens hentet fra UCSC, hg19. Alle kodende deler samt 20 basepar oppstrøms og nedstrøms for disse ble vurdert.

5 Resultat

5.1 aCGH

Som resultat av aCGH ble det detektert en 150 kb stor duplikasjon på X-kromosomet hos begge de syke i familie 1. I forhold til NCBI Build 36 ble duplikasjonen lokalisert til posisjon Xq22.2, g.(103054429_103072782)_(103190036_103232959).

Karyotyping gjort på de to barna i familie 1 forut for dette prosjektet, viste ingen funn.

5.2 Konsentrasjons- og kvalitetsmålinger

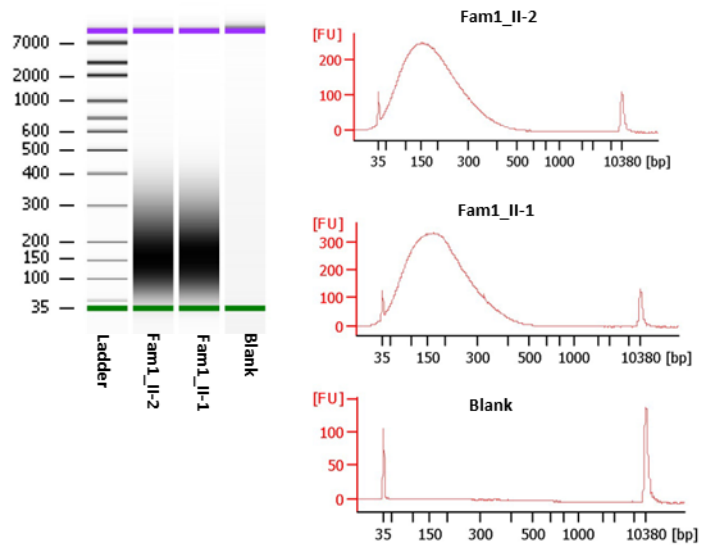
Konsentrasjoner ble målt en rekke ganger under prøveprepareringen for å sjekke hvorvidt prøvematerialet var av god kvalitet, samt for å sjekke at de ulike stegene i prøveprepareringen til storskalasekvensering var vellykket.

Målinger etter DNA-ekstraksjon for familie 1 vises i Tabell 1. På prøvene fra familie 2 ble det ikke gjort noen målinger, da ferdig preparert DNA-materiale ble levert fra laboratorium i Nederland.

Tabell 1. Konsentrasjons- og kvalitetsmålinger.

Prøvenummer	Ekstraktor	Blodvolum ekstrahert (mL)	Elueringsvolum (µL)	Konsentrasjon Qubit (ng/µL)	A260/280 Nanodrop 8000
Fam1_I-1	MagnaPure	1	200	130	1,92
Fam1_I-2	Autopure	4	200	300	1,92
Fam1_II-1	MagnaPure	1	200	78,5	1,88
Fam1_II-2	MagnaPure	1	200	44,6	1,90

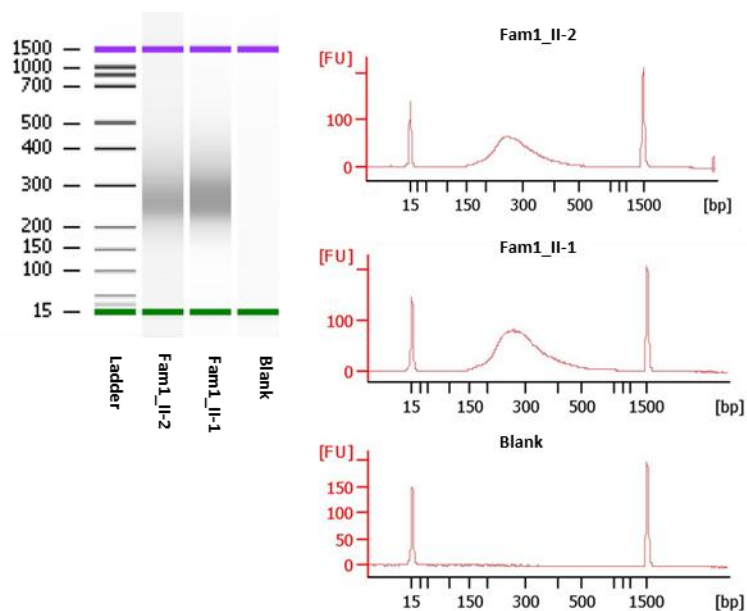
Videre målinger viser kun resultater for prøvene Fam1_II-1 og Fam1_II-2 siden bare disse to ble preparert for storskalasekvensering. Test av fragmentstørrelse på Bioanalyser 2100, i etterkant av “DNA-shearing” på Covaris, viste fragmenter av størrelsesorden 150 – 200 bp. Se Figur 15.



Figur 15. Fragmentstørrelse etter fragmentering av DNA.

Etter preparering på SpriWorks robot og påfølgende kjøring av PCR, ble DNA-konsentrasjon målt på Qubit til 21,4 ng/μL for prøve Fam1_II-1 og 21,6 ng/μL for prøve Fam1_II-2.

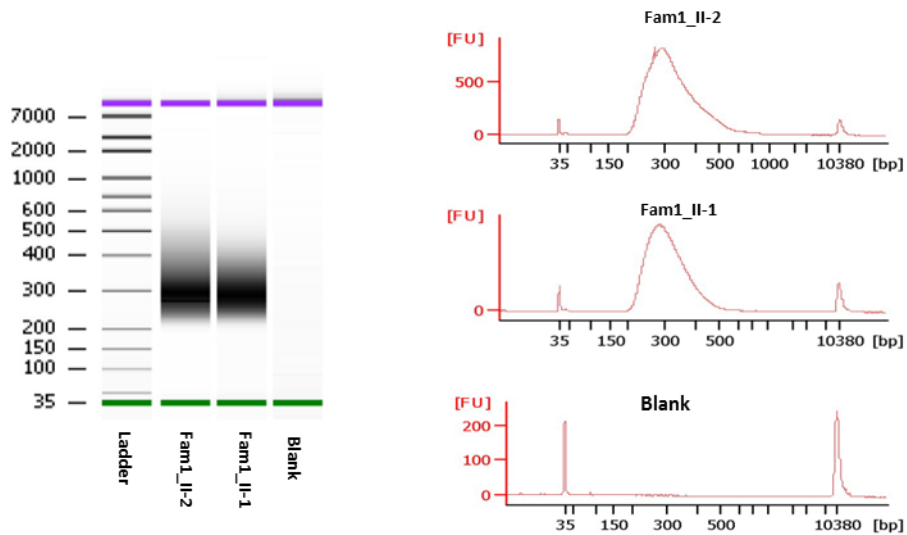
Fragmentstørrelsen, målt på Bioanalyzer 2100, hadde for begge prøver et maksimum rundt 250 bp som vist i Figur 16.



Figur 16. Fragmentstørrelse forut for ”eksom-fanging”.

Målinger før HiSeq

Etter hybridisering og påfølgende PCR ble konsentrasjonen målt på Qubit til 11,36 ng/μL for prøve Fam1_II-1 og 7,46 ng/μL for prøve Fam1_II-2. Målinger på Bioanalyser 2100 viste en hovedvekt av fragmenter med størrelsesorden 310 bp for begge prøvene. Figur 17.



Figur 17. Fragmentstørrelse før storskalasekvensering.

5.3 qPCR

qPCR for sjekk av “eksom-fanging”

qPCRen kjørt for å sjekke hvorvidt “eksom-fanging”-prosedyren hadde fungert optimalt gav resultat som vist i Tabell 2:

Tabell 2. qPCR- resultater for sjekk av “eksom-fanging”.

Prøvenummer	Primersett	Ct, Før “eksom-fanging”	Ct, Etter “eksom-fanging”	ΔCt	E-verdi	Anrikelse, E ^{ΔCt}
Fam1_II-2	NSC-0237	26,83	21,67	5,16	2,09	45,07
Fam1_II-1	NSC-0237	27,03	21,41	5,62	2,09	63,28
Fam1_II-2	NSC-0247	26,84	24,57	2,27	2,13	5,58
Fam1_II-1	NSC-0247	27,26	24,37	2,89	2,13	8,93
Fam1_II-2	NSC-0268	27,15	21,9	5,25	2,26	72,73
Fam1_II-1	NSC-0268	26,6	21,97	4,63	2,26	43,84
Fam1_II-2	NSC-0272	27,68	22,46	5,22	2,14	52,82
Fam1_II-1	NSC-0272	27,75	22,27	5,48	2,14	64,35

Ct = Crossing Threshold. Refererer til den sykel i PCRen hvor fluorescenssignalet overstiger bakgrunnssignalet; ΔCt = differanse i Ct-verdi før og etter “eksom-fanging”; E-verdi = PCR effektivitet. Beregnet utifra skråningen på kurven fra qPCR ved bruk av formel $10^{(-1/slope)}$; E^{ΔCt} = graden av anrikelse før og etter “eksom-fanging”.

Tabellen viser en anrikelse av målregionene i størrelsesorden ~ 45 – 70 ganger for tre av primersettene (NSC-0237, NSC-0268, NSC-0272) og i underkant av 10 ganger for det gjenstående primersettet, NSC-0247.

qPCR for sjekk mot referansebibliotek

Referansebiblioteket som ble brukt i qPCR-kjøringen heter WEF1-2. Dette ga ved bruk av 6 pmol prøvemateriale til storskalasekvensering ett “cluster”-antall på 526,415.

Tabell 3. qPCR-resultater for sammenligning mot referansebibliotek.

	Ct	Δ Ct
WEF1-2	9,48	-
Fam1_II-1	9,99	0,51
Fam1_II-2	10,06	0,58

Ct = Crossing Threshold. Refererer til den sykel i PCRen hvor fluorescenssignalet overstiger bakgrunnssignalet; ΔCt = differanse i Ct-verdi mellom pasientprøve og referanse.

Ct-verdien avhenger av utgangskonsentrasjonen av templatet i qPCRen. Høyere Ct-verdi korresponderer til en lavere utgangskonsentrasjon. For hver Ct-verdi prøvene kommer opp senere enn referansen, må en dobling av mengde prøvemateriale til “clustering” til for å oppnå samme mengde “cluster” som referansen.

Fra qPCR-resultatene (Tabell 3) kunne man lese at prøve nummer Fam1_II-2 kom opp rundt en halv sykel (Ct-verdi) senere enn referansen. Fam1_II-1 kom opp omtrent to tredels sykler (Ct-verdier) senere.

Med ønske om et “cluster”-antall på 450 000 (optimalt antall for maks utbytte ved sekvensering) ble mengde prøvebibliotek for bruk i “clustering” beregnet som følger:

Dersom prøvene kommer opp 1 sykel etter referansen i qPCR, kreves en prøvemengde på 12 pmol:

$$526\,415 \text{ cluster}/12 \text{ pmol} = 43\,868 \text{ clusters}/\text{pmol}.$$

Prøvemengde for prøve Fam1_II-2, som kom opp omtrent en halv sykel etter referansen, blir da:

$$43\,868 + (0,5 \text{ sykler} \times 43\,868) = 65\,802 \text{ clusters}/\text{pmol}.$$

Målet for sekvenseringen var 450 000 “cluster”:

$450\,000 \text{ cluster} / 65\,802 \text{ clusters}/\text{pmol} = 6,8 \text{ pmol}$ bibliotek må brukes for å få dannet omtrent 450 000 “cluster” under storskalasekvenseringen.

For prøve Fam1_II-1, som kom opp omtrent to tredels sykler senere enn referansen, blir mengden:

$$43\,868 + (0,33 \text{ sykler} \times 43\,868) = 58\,491 \text{ clusters}/\text{pmol}.$$

Målet for sekvenseringen var 450 000 “cluster”:

$450\,000 \text{ cluster} / 58\,491 \text{ clusters}/\text{pmol} = 7,7 \text{ pmol}$ bibliotek må brukes for å få dannet omtrent 450 000 “cluster” under storskalasekvenseringen.

5.4 Storskala- og Sangersekvensering

Storskalasekvensering resulterte i dannelselse av 92,717,615 sekvensfragmenter for prøve Fam1_II-2 og 101,238,848 fragmenter for prøve Fam1_II-1. Av disse kunne henholdsvis 80,5 % og 83,6 % sammenstilles mot referansegenom (hg19) gjennom bioinformatisk analyse. Totalt var gjennomsnittlig sekvenseringsdybde på 182 for førstnevnte prøve og 197 for

sistnevnte. Sekvenseringsdybde viser til antall ganger en enkelt posisjon i DNAet har blitt sekvensert/avlest.

Antall detekterte SNPer var for prøve Fam1_II-2 totalt 21,947 mens tilsvarende tall for prøve Fam1_II-1 var 21,618. Antall indels var på henholdsvis 425 og 442. Etter bioinformatisk filtrering (variant rekalkuleringsmetode) ble antall SNPer for prøve Fam1_II-2 redusert til 18,530 hvorav 7,563 var homozygote. For prøve Fam1_II-1 ble tallet 18,265 SNPer hvor 7,597 var homozygote. Antall indeler ble redusert til 348 for Fam1_II-2 hvorav 177 var til stede i homozygot form og til 345 hvorav 182 homozygote for Fam1_II-1.

En oppsummering av resultatene fra eksomsekvensering vises i Tabell 4.

Tabell 4. Resultater eksomsekvensering.

	Fam1_II-2	Fam1_II-1
Totalt antall sekvensfragmenter	92 717 615	101 238 848
Prosent antall sekvensfragmenter riktig sammenstilt mot referansegenom	80,5	83,6
Gjennomsnittlig sekvenseringsdybde for alle sekvenserte regioner	182	197
Totalt antall SNPer i kodende område	21 947	2 1618
Totalt antall indels i kodende område	425	442
SNPer (etter bioinformatisk filtrering, “ variant quality score recalibration”)	18 530 (7563 homoz)	18 265 (7957 homoz)
Indels (etter bioinformatisk filtrering, “ variant quality score recalibration”)	348 (177 homoz)	345 (182 homoz)
Homozygot		
Delte, homozygote SNPer	5380	
Delte homozygote indels	127	
Ukjente, delte, homozygote SNPer	0	
Ukjente, delte, homozygote indels (passert alle kvalitetssettinger)	9	
Compound heterozygot		

Delte, compound heterozygote		4	
En + to varianter			
Gener med en variant hos en pasient og to varianter i den andre		22	
Gener med en variant hos en pasient og to varianter i den andre (passert alle kvalitetssettinger)		4	
Sekvenseringsdybde			
Gener med manglende eller med lav sekvenseringsdybde	0	97	91
	<10x	317	317
	<20x	638	668
	<30x	983	1062

5.4.1 Homozygote varianter - Kandidatgener

Antall homozygote SNPer som var felles mellom begge søstrene (utenfor repeterte regioner) var 5380 mens tallet på delte indels var 127. En videre sortering av disse viste at alle de 5380 SNPene var rapportert i en eller flere databaser tidligere; enten i 1000 Genome Project eller i dbSNP.

Ni av de 127 indelene som passerte alle kvalitetskriterier var ikke rapportert tidligere. Fire av disse førte til et skift i leserammen og var lokalisert til genene *LTBP4*, *AOAH*, *DIXDC1* og *PKDIL3*. Alle fire indelene var av typen insersjon med innskudd av en ekstra base i de tre førstnevnte tilfellene og fire ekstra baser i den øvrige. Tre av disse var lokalisert til konserverte posisjoner i genomet. Alle disse funnene ble Sangersekvensert for alle inkluderte individer for nærmere vurdering.

Av de resterende fem indelene, var to av funnene lokalisert til konserverte posisjoner i genomet. Begge disse var spleisevarianter og var lokalisert til genene *SLC38A3* og *SEC63*. Etter litteratursøk ble begge disse ekskludert fra videre undersøkelse.

LTBP4

For *LTBP4* ble det observert innskudd av en G i posisjon 41123093 på kromosom 19 etter storskalasekvenseringen. I forhold til referanse NM_001042545 kunne den annoteres c.3032_3033insG på cDNA-nivå og p.Q1011fs på proteinnivå.

Sangersekvensering av området hvor insersjon ble påvist, viste at alle inkluderte forsøksindivider hadde aktuell insersjon i homozygot form som illustrert i Figur 18. Både foreldre og barn fra begge familier ble sjekket.

Reference	c t t c c a g t g c c t c t g t g a c c a g g - c t a c g a g g g g g c a c
Reference-AA	S F Q C L C D Q V T R G H
▶ Fam2_II-1	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (
▶ Fam2_I-2	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (
▶ Fam2_I-1	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (
▶ Fam1_I-1	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (
▶ Fam1_II-1	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (
▶ Fam1_II-2	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (
▶ Fam1_I-2	C T T C C A G T G C C T C T G T G A C C A G G G T T A C G A G G G G G C A C (

Figur 18. Insersjon i *LTBP4*.

Figuren viser insersjon av G (rosa font) hos alle pasientene i forhold til referansen som er merket med grønn bakgrunn.

AOAH

I genet *AOAH* kunne det observeres innskudd av T i posisjon 36552790 på kromosom 7 etter storskalasekvensering. Bruk av NM_001177506 gav annotasjon c.1915_1916insA p.P639fs på henholdsvis cDNA og proteinnivå.

Både foreldre og barn fra begge familier ble sjekket med Sangersekvensering for området hvor insersjon ble påvist i storskalasekvensering. Som vist i Figur 19, hadde alle inkluderte forsøksindivider aktuell insersjon i homozygot form. Figuren viser revers tråd.

Reference	g g g t a a a c t c a t t c c a c a a - c c c t a t g g g g g c t g c c a c g t c
Reference-AA	L G K L I P Q P Y G G C H V
▶ Fam2_II-1	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,
▶ Fam2_I-2	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,
▶ Fam2_I-1	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,
▶ Fam1_I-1	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,
▶ Fam1_II-1	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,
▶ Fam1_II-2	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,
▶ Fam1_I-2	G G G T A A A C T C A T T C C A C A A A C C C T A T G G G G G C T G C C A C G T C ,

Figur 19. Insersjon i *AOAH*.

Figuren viser insersjon av A (rosa font) hos alle pasientene i forhold til referansen som er merket med brun bakgrunn.

Det ble også rapportert en insersjon G i posisjon 36552787, navngitt c.1918_1919insC, p.S60fs i forhold til referanse NM_001177506, fra storskalasekvenseringen. Denne kunne ikke reproduseres hos de to fullsekvenserte individene ved bruk av Sangersekvensering. Insersjonen kunne heller ikke ses hos noen av de øvrige prosjektprøvene.

DIXDC1

Innskudd av C i posisjon 11853106 på kromosom 11 ble funnet i gen *DIXDC1*. Dette resulterte i annotasjon for cDNA og proteinnivå på henholdsvis c.179_180insC og p.S60fs i forhold til referanse NM_001177506.

Området hvor insersjon ble påvist ved storskalasekvensering, ble sjekket ved bruk av Sangersekvensering for både foreldre og barn i de to inkluderte familiene. Resultatet viste insersjonen i homozygot form hos alle individene. Se Figur 20.

Reference	C G G C G G C C A G G G A G C C - T G G A A C C T A T C T G G A G A C																																
Reference-AA	W	R	P	G	S		L	E	P	I	W	R																					
▶ Fam2_II-1	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C
▶ Fam2_I-2	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C
▶ Fam2_I-1	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C
▶ Fam1_I-1	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C
▶ Fam1_II-1	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C
▶ Fam1_II-2	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C
▶ Fam1_I-2	T	G	G	C	G	G	C	A	G	G	G	A	G	C	C	T	G	G	A	A	C	C	T	A	T	C	T	G	G	A	G	A	C

Figur 20. Insersjon i *DIXDC1*.
Figuren viser insersjon av C (rosa font) hos alle pasientene i forhold til referansen som er merket med brun bakgrunn.

Ins G i posisjon 111853105 med annotasjon c.178_179insG og p.S60fs funnet i dette genet ved storskalasekvensering, kunne ikke reproduseres hos forsøksindividene ved bruk av Sangersekvensering. Insersjonen kunne heller ikke ses hos noen av de øvrige personene inkludert i studien.

PKD1L3

Som følge av storskalasekvensering ble det funnet innskudd av TTTG i genet *PKD1L3* i posisjon 71981414 på kromosom 16. NM_181536 ble brukt i annotasjon av insersjonen som ble c.3695_3696insCAAA og p.R1232fs på henholdsvis cDNA og proteinnivå.

Sangersekvensering av alle inkluderte forsøkspersoner viste insersjonen i homozygot form hos alle disse som illustrert i Figur 21. Figuren viser revers tråd.

Reference	agaaatgaacaacaaa---gaggatccttggcactccttgg
Reference-AA	E I E Q Q R G S W H S W
▶ Fam2_II-1	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG
▶ Fam2_I-2	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG
▶ Fam2_I-1	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG
▶ Fam1_I-1	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG
▶ Fam1_II-1	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG
▶ Fam1_II-2	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG
▶ Fam1_I-2	AGAAAGAACAAACAAA CAAA GAGGATCTTGGCACTCTTGG

Figur 21. Inersjon i *PKDIL3*.

Figuren viser inersjon av CAAA (rosa font) hos alle pasientene i forhold til referansen som er merket med brun bakgrunn.

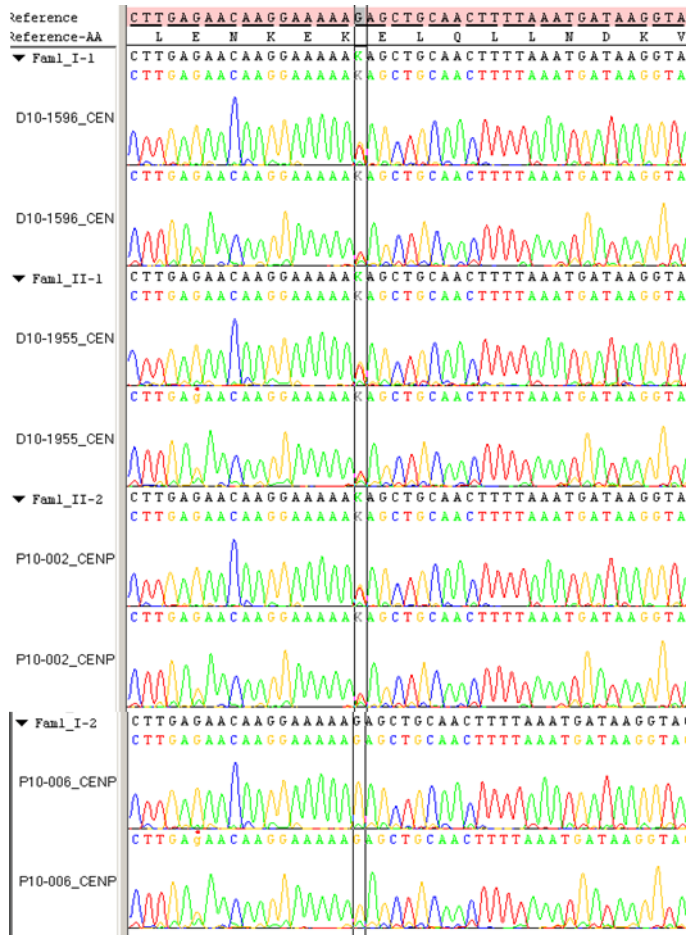
5.4.2 Compound heterozygote varianter - Kandidatgener

Antall gener hvor søstrene var compound heterozygote var fire. Kandidatgenene var *CENPF*, *GPRC6A*, *GPRINI* og *HAVCRI*. Førstnevnte viste to funn som resulterte i henholdsvis introduksjon av et stoppkodon og rammeskiftsendring. Funnene i *GPRC6A* viste to ulike utfall; en rammeskiftsendring og en ikke-rammeskiftsendring. Begge funnene i *GPRINI* ville gi rammeskift mens begge funnene i *HAVCRI* ville resultere i ikke-rammeskiftsendringer. For videre vurdering av disse funnene ble det gjort Sangersekvensering av alle inkluderte pasienter.

CENPF

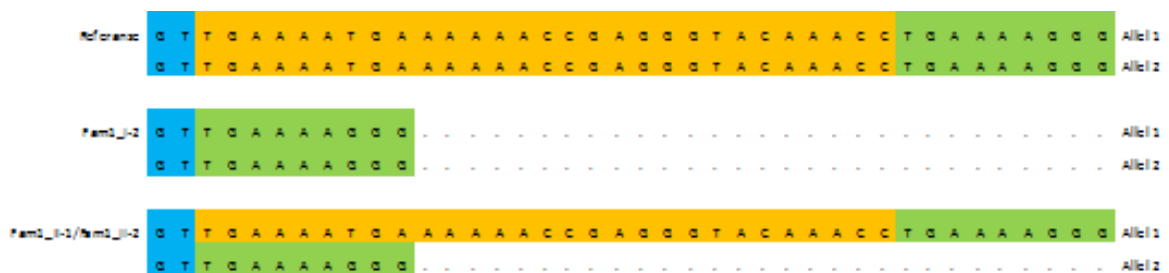
I genet *CENPF* ble det observert en enkeltbaseendring (G>T) i posisjon 214814415 på kromosom 1 og en 29 basers delesjon (TGAAAATGAAAAACCGAGGGTACAAACC) i posisjon 214788177 på kromosom 1. I forhold til referanse NM_016343 ble de annotert c.G2734T, p. E912X og c.165_193del, p.55_65del.

For *CENPF* viste Sangersekvensering at enkeltbaseendringen funnet hos barna i familie 1 også kunne ses hos mor, mens delesjonen som ble funnet i barna også kunne ses hos far. Henholdsvis Figur 22 og Figur 23. Allelene til individene med delesjon, er også illustrert i Figur 23.



Figur 22. Enkeltbaseendring i *CENPF*.

Figuren viser heterozygot enkeltbaseendring G>T hos mor og de to barna i familie 1. Man kan se både en gul (G) og en rød (T) topp i elektroferogrammet hos disse individene. Referansesekvens er merket med rosa bakgrunn.



Figur 23. Delesjon i *CENPF*.

Figuren illustrerer hvordan de to allelene vises i området med delesjon. Blå bakgrunn viser basesekvens oppstrøms for delesjon, gul bakgrunn viser området hvor delesjon forekommer og grønn bakgrunn viser basesekvens nedstrøms for deletert område. Hos referansen ses fullstendig sekvens på begge alleler. Hos far ses delesjon på begge alleler (gul sekvens mangler) og hos barna ses delesjon på et av allelene (gul sekvens mangler) mens det andre allelet viser sekvens tilsvarende referansen.

Ingen av disse to funnene kunne observeres i medlemmer av familie 2. Fullsekvensering av alle kodende deler av *CENPF* genot kan ikke påvise andre sannsynlig sykdomsgivende varianter hos noen av medlemmene i familie 2.

GPRC6A

I genet *GPRC6A* ble det observert to heterozygote insersjoner etter storskalasekvensering. Insersjon av A i posisjon 117113762 på kromosom 6 og insersjon av TCC i posisjon 117113765 på kromosom 6. Ved bruk av referanse NM_148963 ble annoteringen av delesjonene c.2324_2325insT, p.Y775fs og c.2321_2322insGGA, p.K774delinsRK.

Sjekk av *GPRC6A* ved Sangersekvensering viste at begge de påviste insersjonene funnet hos de syke barna, også kunne observeres hos far i familie 1. Mor hadde ingen av disse. Ingen av de to insersjonene kunne ses hos medlemmer av familie 2. Resultat fra Sangersekvensering med referansens og forsøkspersonenes to alleler er fremstilt skjematisk i Figur 24.

Sangersekvenseringen viste videre at far og barna i familie 1 hadde insersjon av GG rett oppstrøms for de to insersjonene påvist ved storskalasekvensering. Far hadde insersjonen i homozygot form mens barna hadde den i heterozygot form.



Figur 24. Insersjoner i *GPRC6A*.

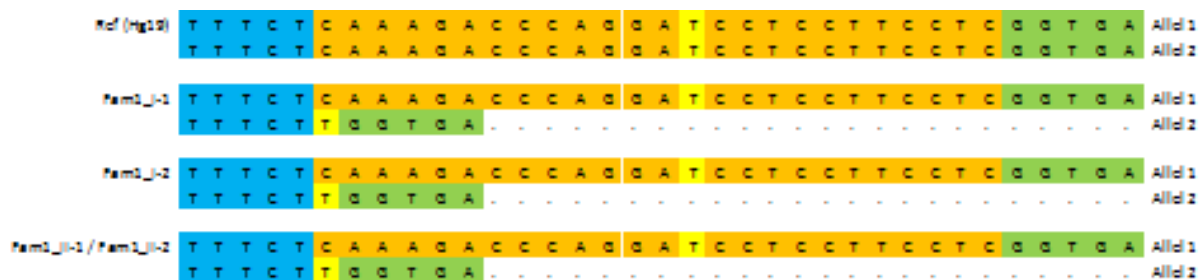
Figuren illustrerer hvordan de to allelene vises i området med insersjoner. Blå bakgrunn viser basesekvens oppstrøms for delesjon, grønn bakgrunn viser området hvor insersjonene forekommer og gul bakgrunn viser basesekvens nedstrøms for området med insersjoner. Insersjonene er merket ut med rød font. Hos referansen og mor ses villtype sekvens på begge alleler. Hos far ses insersjoner på begge alleler (rød sekvens) og hos barna ses insersjonene på et av allelene mens det andre allelet viser sekvens tilsvarende referansen.

GPRIN1

I *GPRIN1* ble det ved storskalasekvensering funnet en heterozygot delesjon av CAAAGACCCAGGA i posisjon 176026134 og en heterozygot delesjon av CCTCCTTCCTG i posisjon 176026146, begge på kromosom 5. Annotasjon ble, i forhold til referanse NM_052899, c.702_714del og p.234_238del for førstnevnte og c.690_700del og p.230_234del for sistnevnte.

De to delesjonene som ble funnet hos barna i familie 1 ved storskalasekvensering ble også påvist hos mor og far i familie 1 ved bruk av Sangersekvensering. Hverken mor eller far i

familie 2 hadde noen av disse deleksjonene. Sekvens for referanse samt forsøkspersonenes (familie 1) to alleler er vist skjematisk i Figur 25.



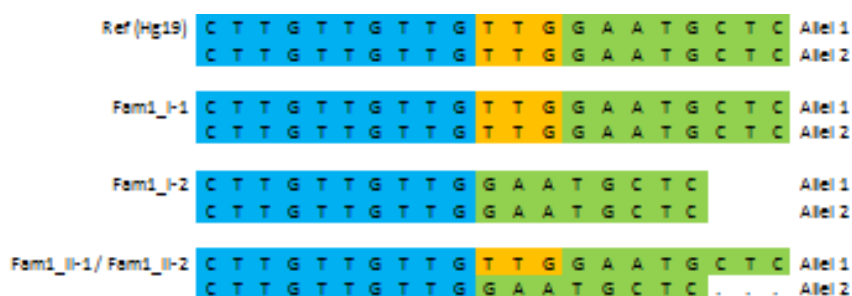
Figur 25. Deleksjoner i *GPRIN1*.

Figuren illustrerer hvordan de to allelene vises i området med deleksjon. Blå bakgrunn viser basesekvens oppstrøms for deleksjon, oransje bakgrunn viser området hvor deleksjonene forekommer, gul sekvens viser området lokalisert mellom deleksjonene og grønn bakgrunn viser basesekvens nedstrøms for deletert område. Hos referansen ses fullstendig sekvens på begge alleler. Hos alle medlemmene av familie 1 ses begge deleksjonene på ett av allelene (orange sekvens mangler). Hvilket av allelene de to deleksjonene er lokalisert på kan ikke avgjøres utifra denne metoden.

HAVCR1

For *HAVCR1* kunne en deleksjon og en insersjon observeres etter storskalasekvensering. Deleksjon av TTG kunne ses i posisjon 156479444 på kromosom 5 og insersjon av GTT kunne ses i posisjon 156479568. I forhold til referanse NM_012206 ble de navngitt c.599_601del p.200_201del og c.477_478insACC p.T159delinsTT.

Deleksjonen som ble funnet hos de syke under storskalasekvensering kunne også observeres hos far i familie 1 etter Sangersekvensering. Deleksjonen var til stede i homozygot form. Hverken mor i familie 1, eller noen av familiemedlemmene i familie 2, hadde deleksjonen. Tolkningen av Sangersekvenseringen er skjematisk fremstilt for familie 1 i Figur 26.



Figur 26. Deleksjon i *HAVCR1*.

Figuren illustrerer hvordan de to allelene vises i området med deleksjon. Blå bakgrunn viser basesekvens oppstrøms for deleksjon, orange bakgrunn viser området hvor deleksjon forekommer og grønn bakgrunn viser basesekvens nedstrøms for deletert område. Hos referansen og mor ses fullstendig sekvens på begge alleler. Hos far ses deleksjon på begge alleler (orange sekvens mangler) og hos barna ses deleksjon på et av allelene mens det andre allelet viser sekvens tilsvarende referansen.

Sangersekvensering av *HAVCR1* viste at far i familie 1 hadde insersjonen, som ble påvist hos de syke barna ved storskalasekvensering, i homozygot form. Mor i familie 1 hadde ikke insersjon. Insersjonen kunne også ses i homozygot form hos alle familiemedlemmene i familie 2. Se Figur 27 for resultat fra Sangersekvensering med visuell fremstilling av familie 1's alleler.



Figur 27. Insersjon i *HAVCR1*.

Figuren illustrerer hvordan de to allelene vises i området med insersjoner. Blå bakgrunn viser basesekvens oppstrøms for deleksjon og grønn bakgrunn viser basesekvens nedstrøms for området med insersjoner. Insersjon er merket ut med rød font. Hos referansen og mor ses villtype sekvens på begge alleler. Hos far ses insersjoner på begge alleler (rød sekvens) og hos barna ses insersjonene på et av allelene mens det andre allelet viser sekvens tilsvarende referansen.

5.4.3 Gener med dårlig sekvensdekning

Bioinformatisk sjekk av dekningsgraden til alle gener i genomet, uavhengig av om de var inkludert i “eksom-fangings”-kit eller ikke, viste at 97 gener ikke hadde noen dekning etter storskalasekvensering for prøve Fam1_II-2. For prøve Fam1_II-1 var tallet 91. 317 gener hadde en dekningsgrad lavere enn 10x for begge prøvene mens dekningsgraden for 638 gener var lavere enn 20x for prøve Fam1_II-2 og for 668 gener i Fam1_II-1.

Dataprioritering av alle gener med dekningsgrad under 20x ved hjelp av programmet “Endeavour” resulterte i følgende prioriterte gener: *PHOX2A*, *ARX*, *LHX3*, *TWIST1*, *SOX8*, *GP9*, *NEUROD2*, *AATK*, *HCRT*, *WFIKKN1*, *NPCC*, *APC2*, *PITX3* og *HMX3*.

5.4.4 Kandidatgener med en variant hos en pasient og to i den andre

I 22 gener kunne det ses en variant i den ene pasienten og to varianter i den andre. Etter ytterligere vurdering av kvaliteten på disse viste det seg at bare fire av genene hadde tre varianter som alle passerte gitte kvalitetskriterier i den bioinformatiske analysen. Dette var genene *KRTAP9-1*, *MIK67*, *NEB* og *PRSS48*.

6 Diskusjon

6.1 Hvorfor finner vi ikke årsaken

Formålet med dette prosjektet var å finne årsaken til Strømme syndrom som er en svært sjelden tilstand. Til tross for høye forventninger basert på suksess i andre, lignende prosjekter, har analysene utført ikke gjort det mulig å identifisere genvariant(er) som forårsaker sykdommen hos de undersøkte pasientene.

Det er flere mulige forklaringer på hvorfor vi ikke har lyktes med dette, men disse faller hovedsakelig inn under to hovedgrupper. Dette kan ha bakgrunn i det rent biologiske aspektet noe som kan gjøre at hypotesene og antagelsene vi gjorde i forkant av analysene ikke stemmer, eller det kan være mer av teknisk art siden metodene vi har benyttet har begrensninger som kan være medvirkende til at vi ikke lykkes. I det følgende diskuteres disse ulike mulighetene.

6.1.1 Usikkerhet i hypotese og antagelser

Alle mennesker er bærere av flere hundre sjeldne, ukjente sekvensvarianter som forårsaker endring i aminosyresekvens. Utfordringen ved storskalasekvensering er å detektere de variantene som er forårsaker sykdom i hvert enkelt tilfelle.

Siden storskalasekvensering resulterer i en svært stor mengde sekvensvarianter er bioinformatisk filtrering av disse et avgjørende verktøy. For å gjennomføre denne filtreringen kreves det en innfallsvinkel for databehandlingen. På bakgrunn av valgt hypotese kan data filtreres. Etter filtrering kan antall potensielt sykdomsgivende varianter reduseres til et overkommelig nivå for videre bekreftelse/avkreftelse om hvorvidt de er sykdomsgivende.

Hypotesen baserer seg både på betraktninger rundt arvemønster for aktuell sykdom og på forventninger rundt hvilken type sekvensendringer som forårsaker sykdom. Hvorvidt de riktige antagelsene er lagt til grunn for databehandling vil derfor være avgjørende for utfallet. Følgelig finnes det i denne filtreringen momenter som kan være bidragsytende for et negativt resultat så langt i jakten på Strømme syndrom.

Genetisk betinget. Monogen tilstand

Man hører stadig diskusjoner rundt hvorvidt sykdomstilstander er påvirket av arv eller miljø. Mulighetene er til stede for at hver av faktorene alene kan forårsake sykdom eller at en kombinasjon av disse gjør det. Slik som Strømme syndrom fremstår kan man anta at den er genetisk betinget. Det at en så distinkt og spesiell fenotype har vært rapportert hos flere individer innen en familie, er en indikasjon på at det er en arvelig oppstått tilstand.

Med et så lite pasientgrunnlag kan man likevel ikke med sikkerhet si at Strømme syndrom er genetisk betinget. Tilstander hvor man i utgangspunktet kan tenke seg et genetisk opphav, kan i ettertid vise seg å være forårsaket av ytre faktorer. Et slikt tilfelle så man for eksempel på 1950-60 tallet hvor en rekke barn ble født med store misdannelser hvor det viste seg at årsaken var medikamentbruk (Talidomid) av mor under svangerskapet (38).

Om det likevel antas en genetisk betinget årsak til utvikling av sykdommen må det også vurderes hvorvidt sykdommen er monogen eller multifaktoriell. Selv om det at et entydig fenotypisk bilde kommer til uttrykk i flere pasienter kan tale for at Strømme syndrom er forårsaket av én enkelt variant, finnes det ikke absolutte holdepunkter for at tilstanden er monogent nedarvet. Man kan også tenke seg en polygen/multifaktoriell nedarving hvor flere varianter sammen kan føre til sykdom. Et slikt tilfelle vil ikke kunne fanges opp med den studiedesignen vi har benyttet siden vi i den bioinformatiske filtreringen kun har fokusert på homozygote og compound heterozygote felles varianter i enkeltgener. Med tilgang til så få pasienter og følgelig et så lite datagrunnlag vil det være nærmest umulig å detektere årsaken om det viser seg at den har et multifaktorielt opphav.

Recessiv arv

I den bioinformatiske behandlingen av storskalasekvenseringsdata må dessuten sykdommens nedarvingsmønster tas i betraktning. Dette er avgjørende for å velge ut varianter som det skal fokuseres på i den bioinformatiske behandlingen.

Nedarving av sykdommer kan skje på ulike måter; den kan være dominant eller recessiv. I recessiv sykdom må to varianter til for at sykdom skal komme til uttrykk mens én variant er nok ved dominant sykdom. Strømme syndrom har vært observert i søsken av friske foreldre. Siden foreldrene ikke viser noen tegn til sykdom ble det i vår hypotese antatt en recessiv form

for nedarving der begge foreldrene er bærere av tilstanden og der de affiserte barna har arvet ett sykt allel fra hver av sine foreldre slik at sykdom oppstår.

Selv om det antas en recessiv form for nedarving av Strømme syndrom er det er imidlertid verdt å merke seg muligheten for at en *de novo*, dominant variant kan forårsake syndromet. Dette ble i første omgang ansett som lite sannsynlig siden syndromet kan ses hos to søsken. Muligheten er likevel til stede for at en *de novo* dominant variant kan observeres hos søsken dersom mor eller far er mosaikk for den sykdomsgivende varianten. Varianten forekommer da bare i enkelte celletyper hos forelderen og sykdomsfenotypen behøver derfor ikke komme til uttrykk. Om en av foreldrene har varianten i sine kjønnseller er muligheten til stede for at denne overføres til barna hvor den fremstår som en *de novo* mutasjon. I et slikt tilfelle, kan også en dominant variant tenkes å forårsake syndromet.

Siden vi med utgangspunkt i vår hypotese har fokusert på recessive varianter vil en eventuell dominant, sykdomsgivende variant ikke ha blitt fanget opp i den bioinformatiske behandlingen av rådata fra storskalasekvensering. I denne prosessen ble kun varianter som kunne ses i homozygot eller compound heterozygot form hos begge de to sekvenserte individene valgt ut. Alle andre varianter ble forkastet. Om det skulle vise seg at syndromet er forårsaket av en dominant variant har denne blitt ekskludert i databehandlingen.

Kodende, ikke-synonym, ukjent variant

For at en genetisk variant skal kunne antas å være kausal er det flere forutsetninger som må legges til grunn (21). I tillegg til at varianten må følge sykdomsfenotypen i familier hvor flere er affisert, forventes det at flertallet av sykdomsgivende varianter er lokalisert til proteinkodende deler av genomet. Dette baseres på at tradisjonelle koblingsanalyser, hvor man kun har fokusert på proteinkodende fraksjon av genomet, har vist seg å være suksessfulle for å detektere varianter som forårsaker monogen sykdom (17). Videre forårsaker det store flertallet av kjente varianter (85 % (39)) som er kjent for å forårsake monogen sykdom endring i kodende sekvens (17).

I dette prosjektet ble derfor sekvensering av genomets proteinkodende deler gjennomført. Ytterligere bakgrunnen for valg av eksomsekvensering blir nærmere diskutert under punkt 6.1.2, Storskalasekvensering. Ved eksomsekvensering vil rundt 99 % av genomet ikke bli sjekket; nemlig de ikke-kodende delene. Disse delene ble lenge ansett som å være “søppel-

DNA”, men etter hvert har man funnet ut at viktige DNA-sekvenser kan være lokalisert til disse områdene. Høyt konserverte og regulatoriske områder som kan være avgjørende for genuttrykk kan være lokalisert her. Manglende funn i tilknytning til Strømme syndrom kan derfor ha sammenheng med at deler av biologien blir ekskludert gjennom metodevalget. Dersom Strømme syndrom er forårsaket av en regulatorisk variant som ligger utenfor de kodende delene av genomet, har det ikke vært mulig å fange den opp med metoden benyttet i dette prosjektet.

Videre forventes det som sagt at det mest sannsynlig er ikke-synonyme varianter som forårsaker denne type sykdommer og det fokuserer derfor på disse i databehandlingen (1). Dette i lys av at en stor del av sjeldne, ikke-synonyme varianter er predikert til å være skadelige (17). Om en synonym variant mot formodning er avgjørende, vil den ikke bli tatt hensyn til i databehandlingen. Tilfeller hvor synonyme forårsaker sykdom har blitt rapportert. I Hutchinson-Gilford progeria syndrom resulterer en synonym variant i aktivering av et kryptisk spleisesete og har vist seg å være årsaken til utvikling av sykdom (40).

I tillegg forventes det at varianter som forårsaker monogen sykdom er svært sjeldne og derfor ikke vil dukke opp i databaser med oversikt over kjente varianter (1). Med dette som utgangspunkt, kan alle varianter som tidligere er rapportert filtreres bort og på den måten redusere antall funn betraktelig. Faren her er at man regner med at det i de mest brukte databasene finnes et antall “falske” SNPer; dvs varianter man antar er normalvarianter, men som mest sannsynlig er sykdomsgivende (41). Det er derfor en risiko for å filtrere bort den sykdomsgivende varianten, særlig når man tenker seg at varianten ikke forårsaker sykdom i heterozygot tilstand. Så mye som 15 – 17 % er rapportert til å være falsk positive SNPer i dbSNP (1). Som følge av dette kan det være en mulighet for at aktuell, kausal variant har blitt forkastet i filtreringen mot normalvarianter i databasene. Det er derfor en hårfin balansegang under filtreringen mellom å ha en streng “cutoff” som vil kunne føre til at vi forkaster reelle varianter og en lite streng “cutoff” som vil kunne medføre et uoverkommelig antall potensielle varianter som må sjekkes ut.

Felles genetisk opphav i ulike familier

I hypotesen har vi antatt at det er et felles genetisk opphav til Strømme syndrom hos alle de rammede. Men man kan ikke utelukke det faktum at genetisk heterogenitet spiller en rolle. Genetisk heterogenitet er tilstedeværelsen av flere distinkte sykdomsgener/loci for en enkelt

klinisk fenotype (21) og er et ikke helt uvanlig fenomen. I enkelte nevrologiske sykdommer er det for eksempel påvist en stor grad av heterogenitet. Opp til >40 loci for samme sykdom er rapportert (21). En slik heterogenitet vil komplisere bruken av bioinformatiske filter i håndtering av rådata og det vil i slike tilfeller være vanskelig å bevise sammenheng mellom en mistenkt variant og sykdom.

Med det begrensede prøvematerialet som vi har hatt tilgjengelig i dette prosjektet, hvor bare individer fra en familie er storskalasekvensert, vil utvelgelse av eventuelle sykdomsgivend(e) variant(er) ikke være mulig dersom det finnes en stor grad av locus heterogenitet.

6.1.2 Begrensninger i metode

I tillegg til de rent biologiske aspektene knyttet til sykdommen, som er avgjørende for valg av riktig hypotese, kan også faktorer knyttet til metodene benyttet være avgjørende for manglende mulighet til å trekke en konklusjon for Strømme syndrom.

Tidligere undersøkelser

En del kliniske tester har blitt gjennomført på pasienter med Strømme syndrom, men så langt har lite vært gjort for å lokalisere den genetiske årsaken til syndromet. Manglende genetiske undersøkelser kan nok i stor grad tilskrives begrensningene i molekylærbiologiske metoder som har vært tilgjengelig for undersøkelse av sjeldne, monogene tilstander inntil de siste årene.

Av tidligere genetiske undersøkelser har det blitt gjort karyotyping på pasienter med Strømme syndrom. I de fleste tilfeller har en normal karyotype blitt beskrevet (2-5). I en av pasientene som er antatt å ha syndromet kunne det observeres en avvikende karyotype (6) uten at denne kunne gi noen forklaring på årsaken til utvikling av Strømme syndrom. Videre er det observert karyotype 46XY i en annen, kvinnelig pasient som mistenkes å ha syndromet, men heller ikke denne kan forklare årsaken til sykdomsutvikling (8).

I tillegg har det vært gjennomført en kandidatstudie hvor fire gener, funnet i tilstander med delvis overlappende fenotype, ble undersøkt på en pasient med Strømme syndrom (3). Genene var *PAX6*, *FOXC1*, *PITX2* og *MYCN*. Det ble gjennomført både fullsekvensering og MLPA (Multiplex Ligation-dependent Probe Amplification) analyse av disse genene uten at funn ble gjort.

Karyotyping

Karyotyping har en begrenset oppløslighet og kan bare detektere endringer større enn 5-15 Mb avhengig av lokalisasjonen til endringen (42). Ut i fra de manglende resultatene på karyotyping er det naturlig å tro at en mindre genetisk endring oftest er årsaken til Strømme syndrom. Ved videre undersøkelser for å lokalisere den genetiske varianten var det derfor avgjørende å benytte seg av metoder som kan fange opp mindre endringer i arvematerialet.

aCGH er en slik metode. Vi valgte derfor å benytte oss av denne veletablerte metoden som en screening av pasientmaterialet vi hadde tilgjengelig.

aCGH

Genomiske strukturelle endringer, spesielt store duplikasjoner eller delesjoner, er en hyppig årsak til monogen sykdom (21). Når karyotyping ikke resulterte i funn, ble det gjennomført aCGH-testing av pasientene slik at eventuelle mindre genetiske endringer i DNA-mengde kunne detekteres. Slike endringer blir gjerne kalt kopitallsvariasjoner (CNV) og kan være av typen duplikasjon eller delesjon. Selv om disse strukturelle endringene etter hvert også kan observeres ved storskalasekvensering, er analysemetodene for CNV deteksjon ved sekvensering ennå ikke godt nok innarbeidet eller validert. aCGH er derfor et godt alternativ som kan brukes som screening før man går i gang med storskalasekvensering.

aCGH mikromatriser benyttet i prosjektet, består av rundt 105 000 oligonukleotider og har mulighet til å oppdage ubalanserte genomiske forandringer større enn ca 22 kb (43).

Ved aCGH ble det gjort et funn på kromosom X hos de to syke barna i familie 1. Siden Strømme syndrom er antatt til å ha en autosomal nedarving, ble dette funnet tolket til mest sannsynlig å være en normalvariant og ikke årsaksgivende til Strømme syndrom. Videre var funnet lokalisert til et område uten klare kandidatgener for sykdommen.

Det har også vært gjort aCGH-testing av andre pasienter med Strømme syndrom, enn de som er inkludert i denne studien, uten at det har blitt gjort funn som kan knyttes opp mot sykdomsutviklingen (7).

På bakgrunn av resultatene fra aCGH samt negativt funn ved karyotyping, er det naturlig å mistenke at en liten indel, som ikke kan oppdages ved de to nevnte metodene på grunn av begrenset oppløselighet, eller at en enkeltbaseendring er årsaken til Strømme syndrom. For å

kunne observere endringer på enkeltbasenivå må sekvensering av DNA gjennomføres. Siden det ikke har vært mulig å peke på noen kandidatgener i forbindelse med Strømme syndrom var behovet tilstede for en helgenomsekvensering.

Storskalasekvensering

Når det ikke har vært mulig å påvise nye, sykdomsgivende varianter ved bruk av “lavoppløselighetsmetoder”, var det tidligere vanlig å gå videre med koblingsanalyse. I koblede regioner ble det valgt ut kandidatgener som kunne passe med pasientenes fenotype og Sangersekvensert disse. En slik tilnærming forutsetter tilgang på et visst antall pasienter, noe som kan være problematisk i flere antatt monogene sykdommer, deriblant Strømme syndrom. Man har gjerne familier med få individer, affiserte i familier hvor det ikke er noe slektskap eller sporadiske tilfeller. Alle disse situasjonene vil gi begrenset styrke slik at en koblingsanalyse ikke kan gjennomføres.

Siden det viste seg at styrken ikke var god nok til å gjennomføre en tradisjonell koblingsanalyse, valgte vi å gå videre med storskalasekvensering. Denne metoden har etter at den ble introdusert i 2005 åpnet for muligheten til å detektere den genetiske årsaken til sjeldne, monogene tilstander som Strømme syndrom (1). For autosomt recessive tilstander har det vist seg at kun to affiserte individer kan være nok for identifikasjon av sykdomsgenet (39).

Istedenfor å gjennomføre en helgenomsekvensering valgte vi å se kun på den kodende delen av genomet gjennom en eksomsekvensering.

Dette valget ble gjort på bakgrunn av flere faktorer. Foruten faktorene nevnt innledningsvis knyttet til forventninger rundt type varianter som forårsaker sykdom (punkt 6.1.1, Kodende, ikke-synonym, ukjent variant) vil det med et så lite materiale som var tilgjengelig i dette prosjektet, vært svært vanskelig å sortere ut den biologisk viktige varianten i en enorm mengde ulike, unike varianter som helt sikkert hadde vært resultat av en helgenomsekvensering.

Ved bruk av eksomsekvensering vil sekvenseringsdybden (antall ganger et gitt område blir lest) av genomet kodende deler videre bli betraktelig større en ved en helgenomsekvensering hvor det vil være mange flere fragmenter som skal dekkes under sekvenseringen. Økt sekvenseringsdybde øker spesifisiteten av sekvenseringen og det endelige resultatet av

sekvenseringen vil følgelig være mer sikkert (se punkt 6.1.2 Sekvenseringsdybde og “feilcalling”). Eksomet utgjør rundt 1 % av genomet. For å oppnå en tilsvarende sekvenseringsdybde som eksomsekvensering vil kunne gi, kreves det for en helgenomsekvensering at flere eksperiment gjennomføres i parallell. I motsetning til mengde sekvens som må til for å gjennomføre en helgenomsekvensering, kreves bare rundt 5 % så mye sekvensering, for å gjennomføre en eksomsekvensering (17).

I tillegg til de faglige argumentene for å gjennomføre en eksomsekvensering var også det økonomiske aspektet medvirkende. En eksomsekvensering kan gjennomføres til en 1/10 – 1/20 av prisen av en helgenomsekvensering (44).

Korte avlesninger (“reads”) og fasebestemmelse

På bakgrunn av hypotesen om en recessiv tilstand, samt det faktum at foreldrene i familie 1 ikke var i slekt, søkte vi først og fremst etter en compound heterozygot variant.

Av de fire compound heterozygote kandidatgenene, kunne vi i *HAVCRI* se at begge varianter observert hos de to syke barna i familie 1, var nedarvet fra far. Far hadde begge variantene i homozygot form mens ingen av variantene kunne påvises hos mor. Dette viser at variantene ikke var i compound heterozygot form i barna, men at begge var nedarvet fra fars allel. Dette, samt det faktum at far ikke viser noen tegn til sykdom gjør at *HAVCRI* kan ekskluderes fra å være genet knyttet til sykdommen.

Den samme konklusjonen kunne trekkes for *GPRC6A* hvor Sangersekvensering viste at far hadde de to variantene, påvist hos de syke barna, i homozygot form mens mor ikke hadde noen av variantene.

Begge variantene som ble observert i *GPRINI* i de to syke barna etter storskalasekvensering, kunne også ses hos både mor og far etter Sangersekvensering. Både mor og far hadde begge variantene i heterozygot form. Siden ingen av foreldrene viste tegn til sykdom er dette mest sannsynlig en normalvariant og ikke årsaksgivende til sykdom.

I jakten på compound heterozygote varianter endte vi opp med å Sangersekvensere flere kandidatgener fra storskalasekvenseringen som viste seg ikke å være aktuelle sykdomsgener siden de ikke var nedarvet i compound heterozygot form. Bruken av korte avlesninger (“reads”) i storskalasekvensering hindrer muligheten for fasebestemmelse av varianter, noe

som kan være avgjørende i deteksjon av compound heterozygote varianter som forårsaker recessive tilstander (45). Manglende mulighet til å finne denne informasjonen i storskalasekvenseringsdata ble tydelig illustrert i de tilfellene hvor Sangersekvensering i etterkant viste at begge variantene var nedarvet fra far.

Det siste kandidatgenet som viste compound heterozygot funn hos barna etter storskalasekvensering var *CENPF*. Sangersekvensering viste at delesjonen som ble funnet hos barna også kunne ses hos far og da i heterozygot form. Baseendringen som ble funnet kunne ses i heterozygot form hos mor. Barna har altså nedarvet en variant av hver av sine foreldre og det var mulig å konkludere med at denne fantes i compound heterozygot form i barna. Dette funnet ble gått videre med siden det fortsatt var en klar kandidat for sykdom. Ingen av de to variantene funnet i familie 1 kunne ses hos familie 2 og genet ble derfor fullsekvensert i medlemmene av familie 2. Siden ingen flere ukjente varianter kunne observeres i dette genet for familie 2, ble det konkludert med at *CENPF* mest sannsynlig ikke er knyttet opp mot utvikling av Strømme syndrom.

Med tanken på diskusjonen knyttet til genetisk heterogenitet bør det gjøres testing av dette genet i ytterligere pasienter for å kunne utelukke genet med høyere sikkerhet.

Det ukjente genom

Siden compound heterozygote varianter ikke gav noen forklaring, ble homozygote kandidater vurdert.

I fire gener ble det funnet en ukjent homozygot rammeskiftsvariant hos de to syke i familie 1 etter storskalasekvensering. Disse ble også funnet i homozygot form hos alle de øvrige, friske personene inkludert i studien ved Sangersekvensering. Ut i fra dette ble det konkludert med at alle disse mest sannsynlig er normalvarianter og ikke årsaksgivende for sykdommen. Genene *LTBP4*, *AOAH*, *DIXDC1* og *PKDIL3* ble altså utelukket som kandidater.

Mistenkelige funn etter storskalasekvensering, som ble avkreftet som følge av Sangersekvensering, sier noe om hvor ukjent det humane genom fortsatt er for oss. Selv om man nå har teknologien til å studere genomet mye mer inngående enn man har hatt anledning til tidligere, gjenstår det fortsatt mye. Kun et begrenset antall humane genomer er fullsekvensert så langt noe som gjør at mye av normalvariasjonen fortsatt ikke er kartlagt.

Mange funn som resultat av storskalasekvensering viser seg derfor nettopp å være sjelden normalvariasjon.

Arbeidet som blir gjort i disse tider med helgenomsekvensering av et stadig større antall genomer, er derfor avgjørende ikke bare for å detektere sykdomsalleler, men også for å bli bedre kjent med normalvarianter i det humane genom. Dette vil dermed bidra til å forenkle tolkingen av resultater fra storskalasekvensering.

Også manglende informasjon om enkeltgener i litteraturen er avgjørende i denne sammenheng. Som følge av dette kan det ofte være vanskelig å utelukke gener som kandidater. Istedenfor å ha muligheten til å utelukke gener basert på litteraturfunn må man derfor gå videre med ytterligere analyser før disse kan avskrives, noe som er både tid og ressurskrevende. Forhåpentligvis vil økt kunnskap om alle gener bidra til at vi lettere kan ekskludere kandidater basert på informasjon hentet fra litteraturen og igjen redusere arbeidsmengden. Ufullstendig informasjon i litteraturen kan også medføre at man feilaktig ekskluderer reelle kandidater.

Sekvenseringsdybde og “feilcalling”

En laboratorieteknisk svakhet ved metoden kunne tydelig ses ved sjekk av sekvenseringsdybde i etterkant av storskalasekvenseringen. Sekvenseringsdybde er ett avgjørende moment ved storskalasekvensering da den har betydning både for kvaliteten på “basecallingen” og på hvilke regioner av genomet som blir undersøkt.

For å oppnå nøyaktighet i storskalasekvensering avleses et bestemt sekvensområde gjentatte ganger (massiv parallell sekvensering). Hver avlesning bidrar til økt sekvenseringsdybde og en konsensussekvens kan bestemmes ut i fra alle avlesningene i et gitt område. Nødvendig sekvenseringsdybde for å unngå “feilcalling” vil avhenge noe av type prosjekt (46), men det er vist at sensitiviteten øker betraktelig ved å øke sekvenseringsdybden fra rundt 5x til 20x (44). En ytterligere økning i sekvenseringsdybden vil resultere i gradvis økning av sensitivitet før den når et platå rundt 50x dybde (44). Økning av sekvenseringsdybden utover 30x resulterer i mindre enn 5 % økning av detekterte sekvensvarianter (47). Det er ikke lineær sammenheng mellom økt sekvenseringsdybde og graden av variantdeteksjon (47).

For Sangersekvensering ligger nøyaktigheten på 99.999 % “basecalling” for rådata (48). Man ser en høyere feilrate i “basecalling” ved storskalasekvensering enn det man gjør ved

tradisjonell Sangersekvensering (45;46), noe som kan øke funn av falsk positive varianter (1). Men total feilrate kan komme lavere enn Sangersekvensering ved bruk av økt sekvenseringsdybde (45). Den dominerende typen “feilcallinger” er substitusjoner og gjennomsnittlig feil i rådata ligger på 1 – 1,5 %, men denne verdien kan forbedres gjennom kvalitetsparametere knyttet til “basecallingen” (48).

Foruten sekvenseringsdybden avhenger feilraten også av andre faktorer som det er verdt å nevne, som ratioen mellom styrke på signal og bakgrunn, faseforskyvning og “crosstalk” mellom “clustere” liggende ved siden av hverandre i “flowcellen” (46). “Crosstalk” kan oppstå som følge av at “clustere” blir liggende for tett på “flowcellen”. Fluoroforen på ett “cluster” kan da påvirke signalet til fluoroforen i ett annet og feil base blir registrert. Faseforskyvningen viser til at et uriktig antall nukleotider blir satt inn før avlesning under sekvenseringen. Denne faseforskyvningen kan oppstå som følge av at flere nukleotider blir satt inn forut for en enkelt avlesning, at nukleotider ikke blir satt inn i det hele tatt eller at blokkeringen som skal skje mellom hver avlesning under sekvenseringen feiler. Ved kjøring av gjentatte sykler vil slike feilinkorporeringer akkumuleres og resultere i dannelsen av heterogene “clustere” som inneholder tråder av ulik lengde. Som følge av disse faktorene faller kvaliteten på de sist inkorporerte basene tydelig med økt fragmentlengde og vil derfor være begrensende for lengden/antall baser som kan avleses i hver kjøring. Om for mange baser leses av, er faren til stede for både falsk positive og falsk negative varianter.

I dette prosjektet ble det oppnådd en gjennomsnittlig sekvenseringsdybde i overkant av 180x for begge pasientprøvene, noe som er godt innenfor kravene for å få en nøyaktig “basecalling”. Dersom enkeltområder har lav sekvenseringsdybde kan likevel feil i “basecalling” forekomme. Om aktuell variant ligger i et slikt område er risikoen for at den blir feiltolket til stede. Falsk negative er potensielt et stort problem siden filtreringen baserer seg på deteksjon av ukjente varianter i samme gen hos ulike pasienter. “Feilcalling” vil følgelig kunne medføre eksklusjon av reelle kandidater.

Med faren for eventuelle falsk negative “basecallinger” i mente, valgte vi å ta en nærmere sjekk av alle gener hvor kun én ukjent variant kunne observeres i en pasient, men hvor to varianter kunne ses i den andre. Basert på det vi kunne finne i litteraturen om de fire genene hvor dette var tilfelle (*KRTAP9-1*, *MIK67*, *NEB* og *PRSS48*), var det ikke mulig å finne informasjon som kan knytte noen av dem opp mot fenotypen ved Strømme syndrom.

Sekvenseringsdybde og undersøkte regioner

I tillegg til sekvenseringsdybdens avgjørende betydning knyttet opp mot “basecalling” og kvaliteten av denne, er den som nevnt også avgjørende i lys av hvilke regioner av genomet som har blitt undersøkt.

Gjennomgang av rådata viste at en rekke gener enten var dårlig dekket eller ikke dekket overhode etter storskalasekvensering; dette til tross for en veldig høy gjennomsnittlig sekvenseringsdybde. Rundt 100 gener hadde ingen dekning og rundt 650 gener hadde en sekvenseringsdybde mindre enn 20x hos hver av de to pasientene. Dersom den kausale varianten for Strømme syndrom ligger i et av områdene som ikke er dekket, vil den ikke kunne gjenfinnes i resultatene fra storskalasekvenseringen.

Vi valgte derfor å se nærmere på genene som var lokalisert i områder med ingen eller lav sekvenseringsdybde. Etter gjennomgang av alle de topp prioriterte genene med sekvenseringsdybde mindre enn 20x, var det ikke mulig å peke på noen klare kandidater knyttet opp mot Strømme syndrom. Disse genene ble derfor ekskludert som aktuelle kandidater. I prioriteringen av gener ble det benyttet et genprioriteringsprogram (Endeavour). En svakhet ved bruk av slike programmer er valg av hvilke kriterier som skal benyttes under søket og vektlegging av de ulike kriteriene. Avhengig av disse faktorene kan det oppstå en forskyvning mot utvelgelse av enkelte gener fremfor andre.

Årsak til manglende eller lav sekvenseringsdybde

Årsaken til manglende eller lav sekvenseringsdybde av målregioner kan oppstå som følge av laboratorietekniske årsaker; både knyttet til prøveprepareringen og til selve sekvenseringen.

I sammenheng med prøvepreparering er en viktig årsak knyttet opp mot “eksom-fanging”. Ved bruk av “eksom-fanging” kan man risikere at enkelte av sekvensområdene man er ute etter ikke blir fanget opp i prosedyren (21).

Manglende fanging av målområder kan for det første være resultat av dårlig hybridisering av enkelte prober mot målregionene slik at disse ikke fanges under prøveprepareringen. Som følge av risikoen for dårlig hybridisering, bør det inkorporeres kvalitetssjekk av “eksom-fangingen” i etterkant av hybridisering for å teste at aktuelle målregioner har kommet med. I dette prosjektet ble dette gjort ved å kjøre qPCR med primere rettet mot områder som skal

fanges i “eksom-fangingen”. Resultatet av qPCR viste en amplifisering av alle de fire, aktuelle målregionene, noe som tyder på en godt fungerende fangingsprosedyre. Disse fire testområdene er likevel bare stikkprøver som gir en indikasjon på at prosessen har fungert. Det er fortsatt ingen garanti for at alle målregioner har blitt fanget som de skal. For Agilent SureSelect Human All Exome 50 Mb Kit benyttet i dette prosjektet vises det til en spesifisitet på i overkant av 77 % for fanging av målregioner inkludert i kit’et (49).

For det andre kan manglende fanging av målregioner være resultat av at områder av eksomet ikke har hybridiseringsoligoer spesifikt rettet mot seg og av den grunn ikke blir fanget opp i prosedyren (50). Kit’et brukt i “eksom-fangings”-prosedyren baserer seg på kodende regioner rapportert i “the human consensus coding sequence” (CCDS) (51) og the GENCODE project (52) og har prober rettet mot disse områdene (53). Hva som karakteriseres som del av eksomfraksjonen av genomet vil variere mellom de ulike genomdatabaser tilgjengelig (47). CCDS er et relativt konservativt gensett som kun inneholder proteinkodende sekvens (50). CCDS inneholder bare 83 % av kodende eksonbaser registrert i RefSeq (50). En “eksom-fanging”-studie hvor målregionene var CCDS ble det vist til en 86 % sjans for å lokalisere den kausale varianten (54).

Foruten de nevnte faktorene knyttet til prøveprepareringens betydning for sekvenseringsdybde etter storskalasekvensering, finnes det også rent sekvenstekniske faktorer som har betydning i denne sammenheng.

Storskalasekvensering baserer seg på sekvensering av en rekke små fragmenter i parallell. Selv om lengden på fragmentene har blitt stadig lenger siden metoden først ble introdusert, er de fortsatt korte sammenlignet med fragmentstørrelsen man jobber med i tradisjonell Sangersekvensering. Korte avlesninger er vanskeligere å sammenstille mot en referansesekvens enn de lange sekvenser siden det finnes en viss grad av sekvenslikhet i genomet (45). Desto kortere en avlesning er, desto vanskeligere vil det derfor være å sammenstille denne eksakt mot et spesifikt sted i genomet.

Spesielt vil sammenstilling av komplekse, repetitive deler av genomer være vanskelig siden slike sekvenser forekommer mange steder i genomet. Det er en utfordring å bestemme eksakt hvor den aktuelle avlesningen stammer fra i genomet. For eksempel vil bare 90 % sammenstilles mot et unikt sted i det humane genomet ved bruk av 30 bp avlesninger (46). Det samme gjelder homologe områder/genfamilier og pseudogener som også har en stor grad

av sekvenslikhet (46). Repetitive områder kan altså få en tilfeldig matching av avlest fragment i forhold til referanse. Feil “mapping” kan også på virke dekningsgraden av genomet ved at den bli dårlig i regioner hvor unik mapping er vanskelig (46).

Som følge av de nevnte begrensningene, både knyttet til prøvepreparering og det sekvenstekniske aspektet, er det derfor avgjørende å sjekke dekningsgraden av genomet i etterkant av storskalasekvensering. Ved å gjøre dette kan man se hvorvidt eventuelle regioner av interesse kan ha falt ut.

Andre laboratorietekniske svakheter ved metoden

En annen laboratorieteknisk faktor i prøveprepareringen som ikke er knyttet til sekvenseringsdybde, men som likevel kan være avgjørende for resultatet av storskalasekvensering, er risikoen for å introdusere en forskyvning i hvilke fragmenter som ender opp i det endelige biblioteket. Forskyvning kan oppstå som følge av PCR brukt under prepareringen. Små og lett amplifiserbare sekvenser vil lettere kopieres og utgjøre en større fraksjon av biblioteket etter PCR (55). Spesielt har man sett at GC-rike regioner blir utsatt for negativ “seleksjon” (47). Under prøveprepareringen kan man som følge av dette risikere å miste sjeldne varianter som er underrepresentert i genomet (55). For i størst mulig grad å unngå denne problematikken er det viktig å holde antall sykler brukt under preparering på et minimum; det kreves altså en avveining mellom å få laget nok bibliotek til bruk i sekvenseringen og samtidig sørge for at dette er representativt for DNA-sekvensen man ønsker å sekvensere. Forskyvning kan også forekomme som følge av ulik liggering av adaptere til templatsekvensen (46).

Disse faktorene ville i stor grad være unngått om man kunne sekvensere direkte på DNA-molekylene uten en omfattende prøvepreparering. Slike metoder ser man nå at er i ferd med å utvikles (55).

6.2 Veien videre i jakten på den genetiske årsaken til Strømme syndrom

Positive resultater som følge av storskalasekvensering blir stadig publisert (56). Dessverre har det ikke vært mulig å trekke en endelig konklusjon for Strømme syndrom så langt. Ulike begrensninger knyttet til metoden og hypotesen som er benyttet i prosjektet kan altså være en medvirkende årsak til manglende funn av den genetiske årsaken til Strømme syndrom

Blant annet er det en mulighet for at årsaken til sykdom kan være lokalisert i regulatoriske områder som ligger utenfor eksom-fraksjonen av genomet. Disse har ikke blitt sjekket ved bruk av den metoden som har blitt benyttet i dette prosjektet. Et alternativ kan derfor være å gå videre med en helgenomsekvensering av pasientene. Problemet her er at sekvens utenom de kodende delene av DNA er enda mindre kjent enn den kodende og resultater vil foreløpig være enda vanskeligere å tolke.

Et annet alternativ vil være å eksomsekvensere flere pasienter så man har et større datagrunnlag for sammenligning i den bioinformatiske analysen. Ved å ha data fra flere pasienter åpnes også muligheten for å se etter en eventuell dominant variant.

Som følge av at enkelte deler av genomet kan mangle dekning, ikke bare hele gener som det har blitt gjort en vurdering av, men også enkelt eksoner i ellers godt dekkede gener, kunne det være aktuelt å gå dypere inn i rådata for å lokalisere slike regioner. Nylig har man klart å finne årsaken til sykdom ved en slik gjennomgang (57).

Videre kunne det på bakgrunn av informasjon om at referansedatabasene inneholder et visst antall falske SNPer, være interessant å gå gjennom data for å lokalisere alle compound heterozygote gener felles mellom de to pasientene; uavhengig av om noen av variantene er rapportert tidligere.

6.3 Storskalasekvensering i forskning og diagnostikk

Metoden storskalasekvensering som har vært benyttet i dette prosjektet har allerede vist seg å være et revolusjonerende verktøy i gjennomføringen av prosjekter knyttet til genetikk. Det er derfor også verdt å diskutere bakgrunnen og nytteverdien av denne metoden innen ulike deler av faget samt se på eventuelle utfordringer den kan ha.

6.3.1 Forskning

Tradisjonell Sangersekvensering har så langt vært gullstandarden for å kunne gjengi DNA-sekvenser. Frederick Sanger beskrev metoden første gang i 1977 og det humane genomet ble publisert i 2003 basert på denne metoden. Da hadde mer enn 3000 forskere fra seks ulike land jobbet med prosjektet i 13 år. Kostnaden for prosjektet, som ble kalt “the Human Genome Project”, ble nesten tre milliarder amerikanske dollar (58).

Storskalasekvensering ble først lansert i 2005 (46). Ulike teknologier innenfor denne kategorien sekvensering har i de senere årene kommet på markedet. Felles for storskalasekvenseringsteknologiene er at de baserer seg på massiv, parallell sekvensering av amplifiserte eller enkelt DNA-molekyler som er festet til en “flowcelle” (46). Dette er en avgjørende forskjell fra tradisjonell sekvensering hvor man baserer seg på deteksjon av produkter fra individuelle sekvenseringsreaksjoner (46). Ved storskalasekvensering kan man derfor produsere en betraktelig større datamengde per kjøring (45). Massiv parallell sekvensering resulterer dessuten i vesentlig lavere kostnad per base sekvensert (45). Ved bruk av storskalasekvenseringsteknologi kan man nå sekvensere et humant genom i løpet av uker til en pris i underkant av 50 000 amerikanske dollar (58). I motsetning til dette, som tilsvarer rundt 0,02 US dollar per kilobase, ligger kostnaden for Sangersekvensering på rundt 0,50 US dollar per kilobase (48).

Stadig utvikling innen storskalateknologien, både med tanke på effektivitet og redusert kostnad, endrer landskapet for sekvensering og hvilke prosjekter som kan gjennomføres. Mange applikasjoner er allerede på markedet; blant annet *de novo* genomsekvensering, transkriptomsekvensering, resekvensering, og CHIP-sekvensering. Innen feltet human genetikk har så langt resekvensering vært den mest brukte applikasjonen (59). Fokuset har i stor grad vært å identifisere enkeltbaseendringer, indeler, kopitallsvariasjoner og strukturelle

varianter (59). I forskningen er storskalasekvensering allerede godt etablert og har ført til en revolusjon med tanke på gjennomføring av prosjekter som for få år siden ikke var gjennomførbare. Antall publikasjoner basert på metoden har hatt en eksponensiell vekst fra den først ble lansert og frem til i dag (56).

Innen medisinsk genetikk er det spesielt verdt å legge merke til den store suksessraten i deteksjon av den genetiske årsaken til monogene sykdommer (17;39;60). Her vil metoden trolig ha stor betydning også i fortsettelsen.

Enkelte kan spørre seg om verdien av å bruke store ressurser på lokalisering av genetiske varianter som forårsaker sykdom hos et fåtall mennesker. Det finnes imidlertid mange gode grunner til å forske på nettopp disse sjeldne sykdommene. Det viser seg ofte at en del monogene sykdommer er ikke så sjeldne som man skulle tro (21) og deteksjon av sykdomsvarianter vil kunne gjøre det lettere for klinikere å stille en diagnose. Sjeldne, monogene varianter kan videre bidra i forståelsen av komplekse sykdommer (1;21).

Monogene sykdommer gir stort sett en sterk, fenotypisk effekt. De kan derfor lettere benyttes i forsøksdyr eller celledinjer for å undersøke de patogenetiske mekanismene. I multifaktorielle sykdommer derimot, ser man gjerne en marginal fenotypisk effekt eller det kan være vanskelig å teste direkte på en aktuell variant på grunn av koblingsulikevekt til andre varianter (21). I enkelte tilfeller kan det dessuten tenkes at sykdommer med antatt kompleks bakgrunn vil vise seg å være forårsaket av sjeldne monogene varianter. Kjennskap til monogene varianter vil også kunne bidra til informasjon om funksjonen til affiserte gener (21). Forhåpentligvis vil også identifisering av monogene sykdommer bidra til å identifisere farmakologiske mål ved behandling av sykdom (1).

6.3.2 Diagnostikk

Storskalasekvenseringsteknologien er ikke etablert på lik linje innen diagnostikk som den er innen forskning (61). Med et stadig økende fokus på rask prøvehåndtering samt ønske om reduserte kostnader, vil dette trolig endre seg betraktelig i de nærmeste årene.

Storskalasekvensering krever en rekke, komplekse prepareringssteg i forkant av selve sekvenseringen. Prepareringen er tidkrevende og det har til dels vært lite automatiseringsløsninger for denne typen prøvepreparering. Gode resultater på gjentatte kontrollmålinger underveis i prøveprepareringen i dette prosjektet tyder på en god, stabil

prosedyre. Siden storskalasekvensering fortsatt baserer seg på relativt ny teknologi, utviklinger metoden seg stadig, noe som også gjenspeiler seg i prøveprepareringen. Økende erfaring med kritiske steg i prosedyren har åpnet for revisjon av de første prosedyrene som inneholdt kolonne og/eller gelrensing av DNA til isteden å benytte seg av kulerensing. Dette har vært en avgjørende faktor som har gjort det mulig å gå over til en mer automatisk prøvepreparering. Denne delen av storskalasekvensering er tidvis en flaskehals på grunn av behovet for manuelt arbeid (56). Automatisering av dette trinnet effektiviserer metoden og gjør at metoden i fremtiden vil kunne tilpasses en diagnostisk setting.

Muligheten for fanging av målregioner i prøveprepareringen forut for selve sekvenseringen er en annen viktig faktor for innlemming av storskalasekvensering i diagnostikk. Ved å benytte seg av denne muligheten kan man lage “analysepakker” for effektiv testing av kandidatgener. Målregionene for flere gener med kjente sykdomsvarianter kan da fanges under prepareringen for så å kjøres sammen i sekvenseringen. Det vil være hensiktsmessig å kombinere gener som forårsaker til dels lignende fenotyper. På den måten kan enkeltindivider testes for en rekke gener samtidig, i motsetning til hva som er fremgangsmåten per i dag hvor hvert enkelt gen blir satt opp i separate forsøk og enkeltgener blir ekskludert i en kjedereaksjon inntil man identifiserer sykdomsgenet. Tidsaspektet fra pasientens fenotype er beskrevet frem til en diagnose vil således kunne reduseres betraktelig.

Flere slike analysepakker inneholdende ulike målregioner er allerede lansert som HALOplex, Raindance og Fluidigm Acces array (62). Uttesting viser lovende resultater med tanke på bruk i en diagnostisk setting (61).

Ved å kombinere flere gener i hver kjøring som beskrevet ved bruk av “analysepakker” vil dekningsgraden på en “flowcelle” bli godt utnyttet. En “flowcelle” brukt i storskalasekvensering kan nemlig dekket av en bestemt mengde sekvens. Om det kun er ønskelig å teste for et enkelt gen, og da fange kun denne regionen under prøveprepareringen, kan man fort tenke seg at storskalasekvensering blir lite hensiktsmessig. En unødig stor mengde sekvens vil da produseres per pasient for det ene genet; det blir som å skyte spurv med kanoner. Med lanseringen av indeksering av prøver under prepareringen og multipleksing av opptil 96 prøver med ulik indeks åpner muligheter seg også med tanke på testing av enkeltgener. Kapasiteten per kjøring utnyttes da ved å øke antall prøver som skal testes i hver kjøring istedenfor å teste store regioner på enkeltprøver.

Med innføring av storskalasekvensering i diagnostikk vil det være svært viktig med fokus på kvalitetskontroll og ansvarlig håndtering av resultater. Siden bioinformatisk databehandling er avgjørende for resultatet av storskalasekvensering, har det blitt fremmet forslag om at alle prosjekter som benytter seg av denne teknologien bør oppgi alle kvalitetsscorer gjort for rådata. Spesielt vil dette være viktig i utsvaringen av diagnostiske resultater slik at en sammenligning og standardisering mellom laboratorier kan gjennomføres (45). På grunn av den høye feilraten i basecalling ved denne metoden sammenlignet med tradisjonelle metoder, vil behovet for bekreftelse av funn ved Sangersekvensering fortsatt være til stede inntil videre for å kvalitetssikre funn etter storskalasekvensering (61).

Innføring av storskalasekvensering kan potensielt endre hele innfallsvinkelen til hvordan vi jobber med genetiske sykdommer. Etter hvert vil man trolig i stor grad ha mulighet til å starte i “motsatt ende av skalaen”. Mens man tidligere vurderte pasienters kliniske symptomer i forsøk på å stille en diagnose for deretter å lokalisere aktuelle kjente, kandidater for resekvensering, kan man nå se for seg en screening av hele genomet. Funn kan så vurderes utifra pasientens fenotype og en klinisk diagnose kan stilles i etterkant av testing.

6.3.3 utfordringer ved bruk av storskalasekvensering

Ny teknologi vil ofte medføre store endringer; både i måten vi tenker og arbeider på. I kjølevannet av utvikling dukker det gjerne opp en rekke problemstillinger man tidligere ikke har trengt å forholde seg til.

I sammenheng med den nye sekvenseringsteknologien har de etiske problemstillingene blitt veldig synlige. Dette har fått konsekvenser også for dette prosjektet siden Regionale Etiske Komiteer har vært tilbakeholdne med å gi tillatelse for prosjekter basert på storskalasekvensering; og da spesielt helgenom- og eksomsekvensering av det humane genom. Årsaken er at det blir stilt spørsmålstegn rundt hvordan man skal håndtere all den informasjonen som denne type metode resulterer i.

Storskalasekvensering vil resultere i funn av ulik art. Mange vil man ikke forstå betydningen av, andre kan ha en behandlingsmessig konsekvens for pasienten, mens andre igjen kan gi informasjon om fremtidig sykdom hos pasienten uten at noe kan gjøres for å forhindre utvikling av sykdom. Funn som ikke er direkte relatert til den sykdommen man ønsker å teste for, men som kan være avgjørende for pasientenes helse senere i livet, har vært en stor kilde

til diskusjon knyttet opp mot storskalasekvenseringsteknologien. Slike “biprodukt” av storskalasekvensering er allerede beskrevet (17). Et viktig moment ved storskalasekvensering er altså hvordan de ulike typene funnene skal håndteres; hvorvidt pasientene skal få tilbakemelding om slike funn eller ikke.

Når det gjelder funn som er av ukjent betydning (VUSer) er det per i dag vanskelig å trekke konklusjoner. En stor mengde slike varianter vil dukke opp som følge av storskalasekvensering. Problemstillingen ser man allerede per i dag ved bruk av tradisjonell Sangersekvensering der man i økende grad finner varianter av denne typen. Prediktive dataverktøy kan benyttes til å gi en indikasjon på hvorvidt disse er skadelige eller ikke, men de gir ikke noe fasitsvar. Det fremstår derfor som vanskelig å gi tilbakemelding på slike funn til pasientene per nå. Informasjon vil trolig skape mer forvirring enn klarhet hos pasientene når ikke genetikerne selv kan forklare dem. Forhåpentligvis vil sekvensering av et stort antall genomer gjøre tolking av DNA-sekvenser lettere, ved at man da får bedre kjennskap til alle normalvarianter som finnes i det humane genom. Man vil da ha et bedre datagrunnlag å sammenligne pasienters sekvenser med noe som forhåpentligvis vil gjøre det lettere å trekke klare konklusjoner.

Etter hvert som genomet blir stadig mer kjent, kan man komme til å finne betydningen av varianter som hittil har vært ukjent i ettertid av gjennomføringen av storskalasekvenseringsprosjekter. Rent logistisk vil det være vanskelig å sørge for tilbakemelding på et langt senere tidspunkt enn oppstart for prosjektet. Videre kan pasientens situasjon ha endret seg slik at tilbakemelding ikke lenger er ønskelig. Det har blitt fremmet forslag om at pasienten er den som bør initiere kontakt på et senere tidspunkt om det skulle være ønskelig/nødvendig (63).

Funn som kan være av en behandlingsmessig konsekvens for pasienten vil det være naturlig å gi tilbakemelding om. Det er viktig å presisere risikoen for slike funn under en genetisk veiledning av pasientene forut for storskalasekvensering.

En slik presisering er også avgjørende når det gjelder funn av varianter som kan være prediktive for sykdom. Pasienten må få anledning til å ta standpunkt til hvorvidt tilbakemelding om slike er ønskelig eller ikke forut for gjennomføring av storskalasekvensering. Så langt har man vært tilbakeholden med å gjøre prediktiv testing av barn siden de ikke selv kan gjøre en vurdering av hva som er ønskelig å vite om egen genetisk

sykdom. Det er ikke lett å se hvordan storskalasekvensering skiller seg fra andre metoder i så måte. Informasjon av slik karakter bør fortrinnsvis holdes tilbake inntil barnet selv kan ta en informert avgjørelse.

Genetisk testing blir tradisjonelt innledet med en genetisk veiledning hvor fordeler og ulemper knyttet opp mot testing blir diskutert. På bakgrunn av informasjon gitt under veiledning kan individer ta en informert avgjørelse hvorvidt testing er ønskelig eller ikke. Det fremtrer et tydelig behov for informert samtykke og genetisk veiledning også i forbindelse med gjennomføring av prosjekter basert på storskalasekvensering. På den måten kan man på forhånd avklare etiske dilemmaer som eventuelt kan oppstå.

Som følge av den enorme datamengden som produseres ved storskalasekvensering oppstår også praktiske problemer rundt sikker lagring av data. DNA-sekvenser er identifiserbare og det vil være avgjørende at disse håndteres etter strenge retningslinjer for datasikkerhet. Sikker håndtering av data skaper også problemer i rådatadatabehandlingen etter storskalasekvensering hvor man er avhengig av fritt tilgjengelige genom databaser som ligger på nett. Tilgang til disse er avgjørende for å sammenligne egne data mot kjente genomer. I denne sammenheng er det viktig å ikke laste opp identifiserbar informasjon. Samtidig er det avgjørende at resultater kan kobles tilbake til riktig pasientprøve siden feilkobling mellom sekvens og prøve-ID kan få fatale konsekvenser. For å kunne benytte seg av alle de positive aspektene ved metoden må man altså klare å finne en balansegang mellom faren for at pasienter skal kunne identifiseres og faren for at data kobles til feil pasient.

En annen utfordring, som er av en mer praktisk art, er at det kan se ut til at introduksjon av storskalasekvensering i stor grad vil medføre et skift i arbeidsmengde fra laboratoriet og over på databehandling. Den enorme datamengden som produseres stiller vesentlig større krav til bioinformatisk kompetanse enn den tradisjonelle Sangersekvenseringen slik den fremstår per i dag. Mens teknisk personale i lab tidligere kunne bruke enkle programvarer for håndtering av Sangersekvenseringsrådata, kreves omfattende bioinformatisk kunnskap for å håndtere data fra storskalasekvensering. Siden metoden fortsatt kan karakteriseres som å være i sin barndom har man ikke standardiserte programvarer/fremgangsmåter for datahåndtering tilgjengelig så langt. Dette er spesielt utfordrende i en diagnostisk setting og behovet for standardisert datahåndtering er absolutt tilstedeværende.

Det er ingen tvil om at storskalasekvensering allerede har endret, og fortsatt vil komme til å endre, det genetiske fagfeltet. Vi står overfor mange utfordringer, både tekniske, faglige og ikke minst etiske. Det er imidlertid med dette som med mange andre tekniske nyvinninger i historien: man er nødt til å forholde seg til det uansett om man liker det eller ikke. Ulike fagretninger som finnes i dette sammensatte feltet må derfor samarbeide om å finne gode løsninger på utfordringene. Om vi lykkes i dette, vil de mange mulighetene som ligger i storskalasekvensering forhåpentligvis kunne utnyttes til å hjelpe mange pasienter, gjennom bedring av både diagnostikk, oppfølging og behandling.

Litteraturliste

- (1) Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011 Apr;129(4):351-70.
- (2) Slee J, Goldblatt J. Further evidence for a syndrome of "apple peel" intestinal atresia, ocular anomalies and microcephaly. *Clin Genet* 1996 Oct;50(4):260-2.
- (3) van BY, van HL, Wolfs R, Tibboel D, van den Hoonard TL, Gischler SJ. Exclusion of a PAX6, FOXC1, PITX2, and MYCN mutation in another patient with apple peel intestinal atresia, ocular anomalies and microcephaly and review of the literature. *Am J Med Genet A* 2008 Feb 15;146A(4):500-4.
- (4) Bellini C, Mazzella M, Arioni C, Fondelli MP, Serra G. "Apple-peel" intestinal atresia, ocular anomalies, and microcephaly syndrome: brain magnetic resonance imaging study. *Am J Med Genet* 2002 Jun 15;110(2):176-8.
- (5) Shanske AL, Gurland JE, Mbekeani JN, Bello JA, Campbell D, Kleinhaus S. Possible new syndrome of microcephaly with cortical migration defects, Peters anomaly and multiple intestinal atresias: a multiple vascular disruption syndrome. *Clin Dysmorphol* 2002 Jan;11(1):67-9.
- (6) Bower KL, Dennis NR, Wellesley D, Williams CP, Hodgkins P, Tyreman C, et al. New case of "apple-peel" intestinal atresia and ocular anomalies with mosaic variegated aneuploidy. *Am J Med Genet A* 2003 Mar 1;117A(2):200-1.
- (7) Castori M, Laino L, Briganti V, Pedace L, Zampini A, Marconi M, et al. Jejunal atresia and anterior chamber anomalies: Further delineation of the Stromme syndrome. *Eur J Med Genet* 2010 May;53(3):149-52.
- (8) Keegan CE, Vilain E, Mohammed M, Lehoczky J, Dobyns WB, Archer SM, et al. Microcephaly, jejunal atresia, aberrant right bronchus, ocular anomalies, and XY sex reversal. *Am J Med Genet A* 2004 Mar 15;125A(3):293-8.
- (9) Waters FM, Lloyd IC, Clayton-Smith J. Apple peel atresia in association with bilateral colobomatous malformation of the optic nerve heads, dysmorphic features, and learning disability - a new syndrome? *Ophthalmic Genet* 2000 Jun;21(2):117-21.
- (10) SANTULLI TV, BLANC WA. Congenital atresia of the intestine: pathogenesis and treatment. *Ann Surg* 1961 Dec;154:939-48.
- (11) Wigglesworth JS, Singer DB. *Textbook of fetal and perinatal pathology*. Boston: Blackwell; 1991.
- (12) Festen S, Brevoord JC, Goldhoorn GA, Festen C, Hazebroek FW, van Heurn LW, et al. Excellent long-term outcome for survivors of apple peel atresia. *J Pediatr Surg* 2002 Jan;37(1):61-5.

- (13) Seashore JH, Collins FS, Markowitz RI, Seashore MR. Familial apple peel jejunal atresia: surgical, genetic, and radiographic aspects. *Pediatrics* 1987 Oct;80(4):540-4.
- (14) Abuelo D. Microcephaly syndromes. *Semin Pediatr Neurol* 2007 Sep;14(3):118-27.
- (15) Woods CG. Human microcephaly. *Curr Opin Neurobiol* 2004 Feb;14(1):112-7.
- (16) Opitz JM, Holt MC. Microcephaly: general considerations and aids to nosology. *J Craniofac Genet Dev Biol* 1990;10(2):175-204.
- (17) Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010 Jan;42(1):30-5.
- (18) Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nat Genet* 2004 Nov;36(11 Suppl):S28-S33.
- (19) The International HapMap Project. *Nature* 2003 Dec 18;426(6968):789-96.
- (20) Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011 Jan;12(1):32-42.
- (21) Kuhlenbaumer G, Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat* 2011 Feb;32(2):144-51.
- (22) Illumina. HiSeq 2000. http://www.illumina.com/systems/hiseq_2000 ilmn 2011
- (23) Janitz M. In: Janitz M, editor. *Next-Generation Genome Sequencing. Towards Personalized Medicine*. Wiley-VCH Verlag GmbH & Co KGaA; 2008. p. 18.
- (24) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009 Jul 15;25(14):1754-60.
- (25) Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15;25(16):2078-9.
- (26) McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010 Sep;20(9):1297-303.
- (27) Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001 Jan 1;29(1):308-11.
- (28) 1000 Genomes. 1000 Genomes. <http://www.1000genomes.org/> 2011
- (29) Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* 2011 Aug 1;27(15):2156-8.
- (30) Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010 Sep;38(16):e164.

- (31) Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010 Mar 15;26(6):841-2.
- (32) Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* 2011 Jan;29(1):24-6.
- (33) Aerts S, Lambrechts D, Maity S, Van LP, Coessens B, De SF, et al. Gene prioritization through genomic data fusion. *Nat Biotechnol* 2006 May;24(5):537-44.
- (34) Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000;132:365-86.
- (35) Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990 Oct 5;215(3):403-10.
- (36) Beckman Coulter Genomics. Agencourt AmPure XP. http://www.beckmangenomics.com/products/dna_purification_and_cleanup/agencourt_ampure_xp.html 2011
- (37) Beckman Coulter Genomics. Agencourt CleanSEQ. http://www.beckmangenomics.com/products/dna_purification_and_cleanup/agencourt_cleanseq.html 2011
- (38) SPEIRS AL. Thalidomide and congenital abnormalities. *Lancet* 1962 Feb 10;1(7224):303-5.
- (39) Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, et al. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. *Hum Mutat* 2010 Aug;31(8):918-23.
- (40) Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, et al. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* 2003 May 15;423(6937):293-8.
- (41) Reich DE, Gabriel SB, Altshuler D. Quality and completeness of SNP databases. *Nat Genet* 2003 Apr;33(4):457-8.
- (42) Ness GO, Houge G. Diagnostikk av medfødte kryptiske kromosomavvik. *Tidsskr Nor Laegeforen* 2003 Sep 11.
- (43) Agilent Technologies. Human Genome CGH Microarrays - Details & Specifications. <http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductData&PageID=1464> 2006
- (44) Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009 Nov 10;106(45):19096-101.
- (45) Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet* 2009 Aug;85(2):142-54.
- (46) Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009 Apr;55(4):641-58.

- (47) Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 2011;29(10):908-14.
- (48) Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008 Oct;26(10):1135-45.
- (49) Agilent Technologies. SureSelect Human All Exon 50 Mb Kit - Details and Specifications. <http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductData&PageID=2318> 2010
- (50) Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet* 2010 Oct 15;19(R2):R145-R151.
- (51) Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* 2009 Jul;19(7):1316-23.
- (52) Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, et al. GENCODE: producing a reference annotation for ENCODE. *Genome Biol* 2006;7 Suppl 1:S4-S9.
- (53) Agilent Technologies. Agilent Technologies. <http://www.home.agilent.com/agilent/home.jsp?cc=NO&lc=eng> 2011
- (54) Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009 Sep 10;461(7261):272-6.
- (55) Singer E. Sequencing a Single Molecule of DNA. *Technol Rev* 2008.
- (56) Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, et al. Next-generation sequencing and its applications in molecular diagnostics. *Expert Rev Mol Diagn* 2011 Apr;11(3):333-43.
- (57) Selmer KK, Gilfillan GD, Stromme P, Lyle R, Hughes T, Hjorthaug HS, et al. A mild form of Mucopolysaccharidosis IIIB diagnosed with targeted next-generation sequencing of linked genomic regions. *Eur J Hum Genet* 2011 Jun 29.
- (58) Diamandis EP. Next-generation sequencing: a new revolution in molecular diagnostics? *Clin Chem* 2009 Dec;55(12):2088-92.
- (59) Zhou X, Ren L, Meng Q, Li Y, Yu Y, Yu J. The next-generation sequencing technology and application. *Protein Cell* 2010 Jun;1(6):520-36.
- (60) Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010 Sep;42(9):790-3.
- (61) De LK, Helleman J, De SJ, Baetens M, Poppe B, Van CW, et al. Massive parallel amplicon sequencing of the breast cancer genes BRCA1 and BRCA2: opportunities, challenges, and limitations. *Hum Mutat* 2011 Mar;32(3):335-44.

- (62) De LK, De SJ, Clement L, Baetens M, Lefever S, De KS, et al. Practical tools to implement massive parallel pyrosequencing of PCR products in next generation molecular diagnostics. *PLoS One* 2011;6(9):e25531.
- (63) Sijmons RH, Van Langen IM, Sijmons JG. A clinical perspective on ethical issues in genetic testing. *Account Res* 2011 May;18(3):148-62.

Vedlegg

Tabell 5. Utstyr.

	Instrument	Produsent	Prosess
DNA ekstraksjon	Magnapure	Roche Applied Science	Ekstraksjon av DNA fra blodprøver
	Autopure	Qiagen	Ekstraksjon av DNA fra blodprøver
DNA målinger	Nanodrop 8000	Thermo Scientific	Renhets- og konsentrasjonsmålinger
	Qubit Quantitation Platform	Invitrogen	Konsentrasjonsmålinger
aCGH	GenePix 4000B	Molecular devices Inc	Scanning av array
	Microarray Hybridization Oven	Agilent Technologies	Hybridisering av array
Storskalasekvensering	Covaris S2	Covaris, Inc	Fragmentering av DNA
	Agilent 2100 Bioanalyser	Agilent Technologies	Sjekk av fragmentstørrelse
	SPRIworks fragment library system I	Beckman Coulter	Prøvepreparering for storskalasekvensering
	ABI 2720 Thermal Cycler	Applied Biosystems by Life Technologies	PCR og inkuberinger
	Savant Automatic Enviromental SpeedVac System AES1010	Thermo Scientific	Oppkonsentrering av DNA
	7900HT Fast Real-Time PCR System	Applied Biosystems by Life Technologies	qPCR
	cBot	Illumina	“Clustering” til

			storskalasekvensering
	HiSeq 2000	Illumina	Storskalasekvensering
Diverse	Sentrifuger (5415R og 5810R)	Eppendorf	
	Pipetter, Reference	Eppendorf	
	Thermomixer Comfort	Eppendorf	
	Termaks varmeskap	Termaks	
Sangersekvensering	ABI3730	Applied Biosystems by Life Technologies	Sangersekvensering
	Biomek Fx	Beckman Coulter	Kulerensing

Tabell 6. Kit og reagenser.

	Kit/reagens	Kommentar	Produsent	Produkt-nummer
aCGH	Human genomic DNA – female	Kutting av DNA	Promega	G1521
	Alu I		Promega	R6281
	Rsa I		Promega	R6371
	Genomic DNA Labeling Kit Plus	Merking av DNA	Agilent Technologies	5190-0449
	Amicon Ultracel-30 Membrane	Rensing av merket DNA	Millipore	UFC5030B K
	Oligo sCGH/Chip-on-Chip Hybridization Kit	Hybridisering	Agilent Technologies	5188-5380
	Human Cot-1 DNA		Invitrogen	15279-011
	Human Genome CGH 2x105k array ID		Agilent Technologies	G4412A
	2x1 Gasket slide		Agilent Technologies	G2534-60002
	Oligo aCGH Wash Buffer 1 og 2	Vask av array	Agilent Technologies	5188-5226
	Stabilization and Drying Solution		Agilent Technologies	5185-5979
DNA-ekstraksjon	MagnaPure LC DNA isolation kit – large volume	MagnaPure	Roche	0331051500 1
	70 % Ethanol	Autopure	Puregene	AE-3800
	Qiagen Hydration Solution		Qiagen	1045702
	Qiagen Cell Lysis Solution		Qiagen	949006
	Qiagen Precipitaion Solution		Qiagen	949008

	Isopropanol 100 %		Puregene	AI-3800
	Qiagen RBC Lysis Solution		Qiagen	949004
Prøvepreparering til storskalasekvensering	dsDNA BR Assay	Qubit Quantitation Platform	Invitrogen	Q32850
	dsDNA HS Assay		Invitrogen	Q32851
	Agilent DNA 1000 Kit	Agilent 2100 Bioanalyzer	Agilent Technologies	5067-1504
	Agilent High Sensitivity DNA Kit		Agilent Technologies	5067-4626
	SPRIworks Fragment Library Cartridges I	SPRIworks		A87320B
	Illumina PE Adapter Oligo Mix			1003455
	Illumina PCR primer PE 1.0	Amplifisering	Illumina	1003450
	Illumina PCR primer PE 2.0		Illumina	1003454
	Herculase II Fusion DNA Polymerase (inkl dNTPmix og 5x buffer)		Agilent Technologies	600677
	SureSelect Human All Exon 50 Mb Kit	Hybridisering	Agilent Technologies	G3370B
	Dynabeads MyOne Streptavidin T1	Eksom-fanging	Invitrogen	656-01
	AMPure XP	Kulerensing	Beckman Coulter Genomics	A63882
	Absolutt alkohol prima		Kemetyl	E41332
qPCR	SybFP6	“Forward” primer: 5’-AAT GAT ACG GCG ACC ACC GA - 3’	Eurofins	

	SybRP8	Reverse primer: 5' - CAA GCA GAA GAC GGC ATA CGA - 3'	Eurofins	
	Power SYBR Green PCR mastermix 2x		Applied Biosystems by Life Technologies	4367659
	WEF1-2	Referanse- bibliotek		
	gDNA	Standardkurve		
	Illumina PhiX control kit v2.0	Standardkurve	Illumina	CT-901- 2001
	5 mM Tris pH8.0; 0,1% Tween-20	Fortynningsbuffer til bibliotek		
	NSC-0237 – “Forward”	5' - CGC ATT CCT CAT CCC AGT ATG - 3'	Nimblegen	
	NSC-0237 – “Reverse”	5' - AAA GGA CTT GGT GCA GAG TTC AG - 3'	Nimblegen	
	NSC-0247 – “Forward”	5' - CCC ACC GCC TTC GAC AT - 3'	Nimblegen	
	NSC-0247 – “Reverse”	5' - CCT GCT TAC TGT GGG CTC TTG - 3'	Nimblegen	
	NSC-0268 – “Forward”	5' - CTC GCT TAA CCA GAC TCA TCT ACT GT - 3'	Nimblegen	
	NSC-0268 – “Reverse”	5' - ACT TGG CTC AGC TGT ATG AAG GT - 3'	Nimblegen	
	NSC-0272 – “Forward”	5' - CAG CCC CAG CTC AGG TAC AG - 3'	Nimblegen	

	NSC-0272 – “Reverse”	5' – ATG ATG CGA GTG CTG ATG ATG - 3'	Nimblegen	
Storskalasekvensering	Paired End Cluster Plate	cBot (protokoll: PE all in one recipe)	Illumina	15005589
	Buffer-EB		Qiagen	19086
	HiSeq PE Flow Cell v1.5		Illumina	15020949
	TruSeq SBS Kit – HTS (50 cycles) Box 1 of 2	HiSeq 2000	Illumina	15014583
	TruSeq SBS Kit – HTS (50 cycles) Box 2 of 2		Illumina	15014584
	PhiX		Illumina	1006471
	Illumina Read 2 Cluster Resynthesis Kit		Illumina	15011162
Sangersekvensering	AmpliTaq Gold 360 DNA Polymerase kit	PCR	Applied Biosystems by Life Technologies	4398853
	GeneAmp dNTP mix with dTTP, 10 mM		Applied Biosystems by Life Technologies	362275
	BigDye Terminator v3.1 Cycle Sequencing Kit	Sekvensering	Applied Biosystems by Life Technologies	4337458
	POP-7	ABI3730xl	Applied Biosystems by Life Technologies	4335615
	3730buffer		Applied Biosystems by Life	4335613

			Technologies	
	CleanSEQ	Kulerensing	Beckman Coulter Genomics	A29154
	Ethylenediaminetetraa cetic acid (EDTA), pH 8.0, 0.5M		Fluka	03690

Tabell 7. Primersekvenser for Sangersekvensering.

Primernavn	Primersekvens (5' – 3')
CENPF_ex1_F	tgtaaacgacggccagtTAAAATCCATCCGAAAGGGC
CENPF_ex2_F	tgtaaacgacggccagtTCATTAATTTCTGAGACTTTGTTTCC
CENPF_ex3_F	tgtaaacgacggccagtTGTTTCATATGGCTTATTGCAGC
CENPF_ex4_F	tgtaaacgacggccagtTCTGGGAATGTAAGGCATTG
CENPF_ex5_F	tgtaaacgacggccagtTGATCTGTGAATTCCTTCATGG
CENPF_ex6_F	tgtaaacgacggccagtTGTTAACTTCTTGGGATTATGGC
CENPF_ex7_F	tgtaaacgacggccagtTTTGTTCGTCAAAGATTCAG
CENPF_ex8_F	tgtaaacgacggccagtATAGTTCCTAACATAAATTGTGTGC
CENPF_ex9_F	tgtaaacgacggccagtTAGGATGCTCATGCCTCACC
CENPF_ex10_F	tgtaaacgacggccagtTTGATATTCATGACCATTTTATTG
CENPF_ex11_F	tgtaaacgacggccagtGGGACCTTTATTAAGGCAGAAG
CENPF_ex12_1_F	tgtaaacgacggccagtTTTGTTCGGATTTTGCCAGG
CENPF_ex12_2_F	tgtaaacgacggccagtGAGTAAGAACGCTGGAGATGG
CENPF_ex12_3_F	tgtaaacgacggccagtGCCATCCTACAAAATAGAGTTGATTC
CENPF_ex12_4_F	tgtaaacgacggccagtTCCAGAAAAGTGAGAGTTTTGC
CENPF_ex12_5_F	tgtaaacgacggccagtTGAACAGCTGATGAAGGTAATG
CENPF_ex12_6_F	tgtaaacgacggccagtGAAAATATGTGAAATACTGCAGGC
CENPF_ex12_7_F	tgtaaacgacggccagtTCTTGTGTGCCTGACAGCTC
CENPF_ex13_1_F	tgtaaacgacggccagtGTGTTTAGCAGAGGCCACG
CENPF_ex13_2_F	tgtaaacgacggccagtGCTAAATGAAATGAAAGAATTAGACTC
CENPF_ex13_3_F	tgtaaacgacggccagtGAGAGAAACCAGCTTCGTGG
CENPF_ex13_4_F	tgtaaacgacggccagtTCGAGAAACTGAGAGTTCGC
CENPF_ex13_5_F	tgtaaacgacggccagtATGAGGCAGTAGCAGCCTTG
CENPF_ex13_6_F	tgtaaacgacggccagtGAGGTCTGGAATTAGATGTTGTTAC

CENPF_ex14_F	tgtaaaacgacggccagtGGCTAAAATAAATGGCATGAATATG
CENPF_ex15_F	tgtaaaacgacggccagtTGAATACATTGCTAAGGATTATAGCAC
CENPF_ex16_F	tgtaaaacgacggccagtTTTTAGCTGTGCCTCCTGG
CENPF_ex17_F	tgtaaaacgacggccagtAACCACATTGTCAGTTGGGG
CENPF_ex18_F	tgtaaaacgacggccagtTTATCTGCTTCACGATGCCC
CENPF_ex19_F	tgtaaaacgacggccagtTCTTTTCCACTGTAGATAGAATTGG
CENPF_ex20_1_F	tgtaaaacgacggccagtTCTCTGTTAGATGGGGCCTG
CENPF_ex20_2_F	tgtaaaacgacggccagtCATTGCCATTCTCTACTGC
CENPF_ex1_R	caggaaacagctatgaccGGAGAGTGATCCCAGAGGC
CENPF_ex2_R	caggaaacagctatgaccCAGTTTTAAATACCAGCACTTCTCTG
CENPF_ex3_R	caggaaacagctatgaccTCTTCCACTCTTTCAAGTTACGG
CENPF_ex4_R	caggaaacagctatgaccTGCAGAATTAATAATCTCAAGCTAAAC
CENPF_ex5_R	caggaaacagctatgaccTGAGCCCAAACCTTTTCTC
CENPF_ex6_R	caggaaacagctatgaccCACCTGTGAAATTACCTCAAGC
CENPF_ex7_R	caggaaacagctatgaccTGCCCAACTTAAAACAGCTTTC
CENPF_ex8_R	caggaaacagctatgaccACAACCTCCCTTTCCAAGTG
CENPF_ex9_R	caggaaacagctatgaccACAAGTACAAGGAACAAGAGGG
CENPF_ex10_R	caggaaacagctatgaccCTTTTCCCCTTGCACTTGG
CENPF_ex11_R	caggaaacagctatgaccATCCCCTCTCACGGGAATAG
CENPF_ex12_1_R	caggaaacagctatgaccCTGAGAACTCAGCTTTCTGCTG
CENPF_ex12_2_R	caggaaacagctatgaccTGACTTGTTTCAGCCACAAAAC
CENPF_ex12_3_R	caggaaacagctatgaccTCTTTTCCTGTGCTGCTTTG
CENPF_ex12_4_R	caggaaacagctatgaccCAAGATCCATCTGAGGTTTAAAATTAC
CENPF_ex12_5_R	caggaaacagctatgaccCACCATGGAGAAGACCACTG
CENPF_ex12_6_R	caggaaacagctatgaccCTACACTGCACTGGTTTGCTG
CENPF_ex12_7_R	caggaaacagctatgaccCATTTTCCTCAAGTTCTGGTTTG

CENPF_ex13_1_R	caggaaacagctatgaccTGAGGCCTTCAGAAGTTTCTAC
CENPF_ex13_2_R	caggaaacagctatgaccACTCACTTTGATGAGACTCAAGC
CENPF_ex13_3_R	caggaaacagctatgaccAGTGAATCATTCTCCCGCTC
CENPF_ex13_4_R	caggaaacagctatgaccTTTCATCAGCTTCTAGGCGG
CENPF_ex13_5_R	caggaaacagctatgaccGCCTTCTCAAGTTCAAGACACTC
CENPF_ex13_6_R	caggaaacagctatgaccCATGTCCCAAAGAATCAGCC
CENPF_ex14_R	caggaaacagctatgaccCAACAGAAGGTCTTACAGGAACC
CENPF_ex15_R	caggaaacagctatgaccACCTGGGGAAAAGATCGTG
CENPF_ex16_R	caggaaacagctatgaccACCATTAAAATCACTACCACCTG
CENPF_ex17_R	caggaaacagctatgaccTCCCAAGATAAGCACAAAATTC
CENPF_ex18_R	caggaaacagctatgaccGCAGACTATTGCTTATGAACGTC
CENPF_ex19_R	caggaaacagctatgaccTCCTCACCCAGGTACTACTGC
CENPF_ex20_1_R	caggaaacagctatgaccTCAGCTCTGAGAAATTCCCG
CENPF_ex20_2_R	caggaaacagctatgaccTTCACGTGTATAAGAGGCTGTTG
CENPF-14F	tgtaaaacgacggccagtAGGAAGCATGCCTTCAGAGA
CENPF-14R	caggaaacagctatgaccTGCAAACTCTCACTTTTCTGG
CENPF-5F	tgtaaaacgacggccagtTGGCTTATTGCAGCTGTATCTC
CENPF-5R	caggaaacagctatgaccACGGTACAGAGACCGAATCA
GPRC6A-1F	tgtaaaacgacggccagtGCTGCTCACACTATGGGAAGA
GPRC6A-1R	caggaaacagctatgaccTTTGGAGTGAGCTTTACTCTTTG
GPRIN1-1F	tgtaaaacgacggccagtATGAGCCTGGAATCAGCTTG
GPRIN1-1R	caggaaacagctatgaccTCACCCTCCAAGGAGACATT
HAVCR1-6F	tgtaaaacgacggccagtACCCTGAGGAGACCCTGATT
HAVCR1-6R	caggaaacagctatgaccGGGCAATGACCAAGATTGAC
AOAH-22F	tgtaaaacgacggccagtTGGGAAATCTTGTTGGTAAGTG
AOAH-22R	caggaaacagctatgaccCTCATGAAGCCTGCTCTGC

DIXDC-7og8F	tgtaaaacgacggccagtACCCTGTGAACGCTGTCTG
DIXDC-7og8R	caggaaacagctatgaccTCCCTATATCTAAACTCCTCATTCAG
LTBP4-23-25F	tgtaaaacgacggccagtCTTCTTAGTCTTGGGCACCG
LTBP4-23-25R	caggaaacagctatgaccATATATGGGGAGGGACAGGG
PKD1L3-22F	tgtaaaacgacggccagtATGAGTGCCACCTCTTCC
PKD1L3-22R	caggaaacagctatgaccTATCGAGTCAGCAAGGTGGC