

Magnus Barnholt Skulberg
Kristoffer Emil Andreassen

Metodeetablering av long-range PCR for nøyaktig påvisning av klinisk relevante varianter i *PRSS1*-genet.

Establishment of long-range PCR method to accurately detect clinically relevant variants in the *PRSS1*-gene.

Mai 2020

NTNU

Norges teknisk-naturvitenskapelige universitet.
Fakultet for naturvitenskap
Institutt for bioingeniørfag

Bacheloroppgave

2020



Magnus Barnholt Skulberg
Kristoffer Emil Andreassen

Metodeetablering av long-range PCR for nøyaktig påvisning av klinisk relevante varianter i *PRSS1*-genet.

Establishment of long-range PCR method to accurately detect clinically relevant variants in the *PRSS1*-gene.

Bacheloroppgave
Mai 2020

NTNU

Norges teknisk-naturvitenskapelige universitet.
Fakultet for naturvitenskap
Institutt for bioingeniørfag



Kunnskap for en bedre verden

Forord

Bacheloroppgaven er utført i forbindelse med avslutningen av en bachelorgrad i bioingeniørfag ved Norges teknisk- naturvitenskapelige universitet (NTNU). Tema for oppgaven ble gitt av Avdeling for medisinsk genetikk (AMG) ved St. Olavs Hospital i Trondheim, og alt praktisk laboratoriearbeid ble utført i AMG sine lokaler på Laboratoriesenteret.

I forbindelse med arbeidet med denne oppgaven ønsker vi å rette en spesielt stor takk til vår faglige veileder Christa Schmidt, ph.d., CLG, for kompetent faglig veiledning, Eva Svaasand, Molekylærgenetiker, for praktisk veiledning og entusiasme på lab, avdelingsleder Silje Kristin Vean for å ha gjort det mulig for oss å få arbeide med denne oppgaven, og ikke minst professor Geir Bjørkøy for gode tips, rettelser og veiledning i det skriftlige arbeidet.

Vi vil også benytte anledningen til å takke alle de fantastiske ansatte ved AMG for all hjelp og tilrettelegging i en vanskelig periode med covid-19 og overgang til nytt laboratoriedatasystem. Dere har vært helt fantastiske til å skjule frustrasjonen og vi fikk nesten følelsen av å ikke ha vært til noe som helst bry!

Sist, men ikke minst vil vi rette en stor takk til Kristin Bye som uten noen forpliktelser overfor oss har hatt en finger med i både det faglige og praktiske arbeidet, i tillegg til å bidra med kanskje det viktigste; oppmuntring.

Trondheim, 19. Mai 2020


Magnus B. Skulberg


Kristoffer E. Andreassen

Sammendrag

Trypsinogen er et proenzym som omdannes til det viktige fordøyelsesenzymet trypsin i mange arter, deriblant mennesket. Genene som koder for de tre isoformene til trypsinogen er serin protease 1, 2 og 3, forkortet *PRSS1*, *PRSS2*, og *PRSS3*. Av ulike grunner kan det skje sekvensendringer i *PRSS1*-genet som kan øke aktiviteten til, eller hindre nedbrytningen av trypsin. Dette kan gi irritasjon i vevet i bukspyttkjertelen (pankreas) hvor cellene som produserer trypsinogen ligger. Over tid kan irritasjon i pankreas lede til utvikling av kronisk pankreatitt, som pasienter med genfeil i *PRSS1*-genet kan utvikle 20-30 år tidligere enn «normalbefolkningen». Kronisk pankreatitt medfører en forhøyet risiko for utvikling av pankreaskreft (1), og pasienter med påvist genfeil vil få tilbud om tettere oppfølging og kontrollopplegg. I Norge utgjør pankreaskreft 2,6% av alle nye krefttilfeller og 7,1% av alle kreftdødsfall (2). Årlig diagnostiseres rundt 900 personer, oftest nær 70 år, med pankreaskreft. Noen kan ha genetiske predisposisjoner og utvikle kronisk pankreatitt også i svært ung alder (3,4), med flesteparten av *PRSS1*-relaterte pankreaskreft tilfeller rundt 55 år (5). Når diagnosen er stilt er 80-85% av tilfellene ikke reelt mulig å behandle med kirurgisk kurative inngrep. Dette er begrunnet med et aggressivt vekstmønster og anatomisk nærhet til andre livsviktige organer, slik at pasienter kun vil få palliativ behandling (6). I Norge er fem års overlevelse med pankreaskreft 10,9% for menn og 10,5% for kvinner (2). Pankreaskreft har høy dødelighet med en gjennomsnittlig overlevelsestid på under 1 år etter diagnosetidspunkt og under 6 måneder ved metastasert kreft (2,7). Det er både klinisk og etisk sett essensielt med egnet analysemetode for påvisning av genfeil i *PRSS1*-genet.

Hovedoppgaven til laboratoriet ved Avdeling for medisinsk genetikk (AMG) er å påvise årsak til genetisk sykdom og benytter i dag nestegenerasjonssekvensering (NGS) til diagnostikk. *PRSS1*-genet er inkludert i NGS genpanelet: «NGS-bukspyttkjertelkreft-CNV». I genomet tilhører *PRSS1*-genet en genfamilie bestående av pseudogener og gener med områder som har store likheter, kalt homologe sekvenser. NGS teknologien kan forveksle avleste sekvensdata mellom *PRSS1* og lignende sekvenser, noe som gir usikkerhet i påvisningen av varianter i *PRSS1*-genet. En årsak til kronisk pankreatitt er forflyttelse av ett ekson fra *PRSS2*-genet inn i *PRSS1*-genet kjent som genkonversjon (4).

Genkonversjon kan ikke detekteres i NGS da avleste sekvensdata ikke tilordnes *PRSS1*-, men *PRSS2*-genet, og en årsak til pankreatitt kan bli oversett. På grunn av overnevnte problematikk er NGS-analysen ikke lenger vurdert som pålitelig ut ifra et ståsted om medisinsk nytteverdi. Hensikten med bacheloroppgaven var å designe og etablere en målrettet analysemetode for *PRSS1*-genet ved hjelp av long-range PCR. Til utprøving og validering av metoden ble 4 blodbankprøver benyttet som normalkontroller. Vi lyktes med å bestemme long-range PCR program-innstillinger og designe primere som kun oppformerte *PRSS1*-genet. Produktet ble sjekket med agarose gelelektroforese og videre nukleotidsekvensert med designede sekvenseringsprimere for å bestemme DNA-sekvensen. Metoden ble validert med analyseprogrammene Sequence Scanner 2 (vurdering av datakvalitet) og Sequence Pilot (sekvensanalyse og vurdering av kvalitet). Siden patogene sekvensvarianter muligens har blitt oversett, ble alle *PRSS1*-rekvirerte analyser med indikasjonen pankreatitt reanalysert med vår metode for å gi svar på om det ble oversett/ikke påvist sykdomsgivende sekvensvarianter i NGS. Det ble funnet totalt to forskjellige normalvarianter i pasientprøvene med vår metode. Fra NGS dataene ble det funnet totalt 35 ulike varianter. **Vi har dermed lyktes med å etablere en målrettet long-range PCR metode som økte nøyaktigheten i påvisningen av klinisk relevante varianter i *PRSS1*-genet betydelig.**

Abstract

Trypsinogen is a proenzyme that converts to the important digestive enzyme trypsin in many species, including humans. The genes encoding the three isoforms of trypsinogen are serine proteases 1, 2, and 3, abbreviated *PRSS1*, *PRSS2*, and *PRSS3*. For various reasons, sequence changes may occur in the *PRSS1*-gene that increase activity or prevent trypsin degradation. This can cause tissue irritation in the pancreas where the trypsinogen producing cells are located. Given time, this can lead to the development of chronic pancreatitis, which patients with gene defects in the *PRSS1*-gene may develop 20-30 years earlier than the "normal population". Chronic pancreatitis carries an increased risk of developing pancreatic cancer (1). Patients with proven gene defects will be offered closer follow-up and control procedures. In Norway, pancreatic cancer accounts for 2.6% of all new cancer cases and 7.1% of all cancer deaths (2). Each year, approximately 900 people is diagnosed with pancreatic cancer, usually around 70 years of age. Some may have genetic predispositions and develop chronic pancreatitis at a very young age (3,4), with most *PRSS1*-related pancreatic cancer cases around 55 years of age (5). When diagnosed, 80-85% of cases are not possible to treat with surgical curative intervention. This is justified by an aggressive growth pattern and anatomical proximity to other vital organs, so that patients will receive palliative treatment only (6). In Norway, five years of pancreatic cancer survival is 10.9% for men and 10.5% for women (2). Pancreatic cancer has a high mortality rate with an average survival time of less than 1 year after diagnosis and less than 6 months in metastatic cancer (2,7). A suitable analytical method for detecting gene defects in the *PRSS1*-gene is not only clinically and ethically important, but to some; essential.

The main task of the laboratory at the Department of Medical Genetics (Avdeling for Medisinsk Genetikk - AMG) in Trondheim University Hospital, Norway, is to identify the cause of genetic disease and uses next-generation sequencing (NGS) for today's diagnostics. The *PRSS1*-gene is included in the NGS gene panel: "NGS pancreatic cancer CNV". The NGS technology is showing signs of mixing *PRSS1*-reads with reads derived from pseudogenes and similar sequences in the same gene family in the genome, called homologous sequences. This leads to uncertainty in the detection of variants in the *PRSS1*-gene. One cause of chronic pancreatitis is the transfer of one exon from the *PRSS2*-gene into the *PRSS1*-gene known as gene conversion (4). NGS is not able to detect gene conversion as reads cannot be assigned to the *PRSS1*-gene, only *PRSS2*, effectively overlooking a cause of pancreatitis. Due to the aforementioned problems, the NGS analysis is no longer considered reliable from a medical utility point of view. The purpose of the bachelor thesis was to design and establish a targeted assay method for the *PRSS1*-gene using long-range PCR. For testing and validation, 4 blood bank samples were used as controls. We succeeded in determining long-range PCR settings and designing primers that amplified only the *PRSS1*-gene. The product was checked by gel electrophoresis and further Sanger sequenced with designed sequencing primers to determine the DNA-sequence. The method was validated using the Sequence Scanner 2 (data quality assessment) and Sequence Pilot (sequence analysis and quality assessment). Since pathogenic variants may have been overlooked in NGS earlier, all *PRSS1* requisition assays indicated with pancreatitis were reanalyzed with our method. A total of two different normal variants were found in the patient samples using our method. From the NGS data a total of 35 different variants were found. **We have thus succeeded in establishing a targeted long-range PCR method that significantly increased the accuracy of detection of clinically relevant variants in the *PRSS1*-gene.**

Innholdsfortegnelse

Forord	i
Sammendrag	ii
Abstract	iii
Innledning	1
Kreft	1
Pankreaskreft	2
Metodiske forutsetninger	6
Oppgavens hensikt	8
Trypsinogen genfamilien	9
<i>PRSS2</i>	11
Pseudogener	11
Genkonversjon	14
<i>PRSS3P2</i>	19
Primerdesign	23
Materiale og metode	27
Primerdesign	27
Prøvemateriale	28
Amplifisering av <i>PRSS1</i> -genet	29
Gelelektroforese	30
Rensing av PCR-produkt	31
Nukleotidsekvensering	32
Resultater	35
Del 1: Fylogeni og homologisøk	35
Del 2: Design av primere	43
Del 3: PCR innstillinger	44
Del 4: Validering av primere	46
Del 5: Reanalyse av pasientprøver	55
Diskusjon	57
Konklusjon	58
Referanseliste	59
Vedlegg	63

Innledning

Kreft

Denne oppgaven omhandler påvisning av sekvensvarianter i *PRSS1*-genet som kan gi økt risiko for utvikling av pankreaskreft.

Kreft rammer mange i løpet av livet. Innen en person har fylt 75 år er det kumulativt 36% sjanse for menn og 30% sjanse for kvinner å utvikle kreft (2). I kroppen deler celler seg stadig vekk for å erstatte cellene som naturlig dør. Denne celledelingen er nøyaktig regulert i kroppen og har flere sjekkpunkt for å forhindre feil. I en kreftcelle har det oppstått en endring slik at den kan utvikle seg uavhengig av kroppens normale vekstkontroll. Kronisk inflammasjon og irritasjon av vev kan føre til at normale celler utvikler seg unormalt (8). En slik vekstforstyrrelse kan gi opphav til kreftutvikling og dannelse av nytt vev (neoplasi). Neoplasi er en vekstforstyrrelse der det dannes en abnorm celledannelse som er unyttig og noen ganger skadelig for organismen. Etter hvert som disse cellene fortsetter å dele seg ukontrollert vil det skje en opphopning av kreftceller i organet der veksten først startet, omtalt en primærsvulst. En malign svulst vokser hurtig i forhold til omgivende vev og er stadig progredierende. Den har mulighet til infiltrerende vekst og mulighet for å danne metastaser (spredning til andre deler av kroppen enn den utgikk fra). Når svulsten har vokst seg stor kan denne svulsten sees eller kjennes som en kul. En svulst vil kunne forstyrre organets normale funksjon, som tung pust og hoste ved lungekreft eller blokkering av utførselsganger ved leverkreft med økt risiko for gulsott. Smerter, feber og kakeksi er også vanlige symptomer ved langt utviklet kreftsykdom. Kakeksi rammer opptil 50% av alle kreftpasienter (9). Denne tilstanden kjennetegnes med vekttap og tap av muskelmasse der utmattelse og dårlig matlyst også er vanlige symptomer.

Kreftutvikling er en kompleks sykdom der risiko er avhengig av mange faktorer. Livsstil, arv og miljø påvirker sannsynligheten for utvikling av kreft. Eksempelvis har røyking flere ganger blitt påvist å være en viktig risikofaktor for utvikling av lungekreft. Viral infeksjon som humant papillomavirus (HPV) har blitt knyttet opp mot økt risiko for utvikling av livmorhalskreft (10). Genetiske variasjoner i arvematerialet er også vist å kunne gi en økt risiko for utvikling av kreft. Varianter i *BRCA1*- og *BRCA2*-genene er kjent til å kunne gi økt risiko for brystkreft (11) og varianter i *PRSS1*-genet er kjent for å gi økt risiko for utvikling av pankreaskreft (5).

Pankreaskreft

I Norge utgjør pankreaskreft 2,6% av alle nye krefttilfeller og 7,1% av alle kreftdødsfall (2). Rundt 900 personer diagnostiseres med pankreaskreft årlig i Norge. Den opptrer hyppigst hos personer rundt 70 år i en litt større andel menn enn kvinner.

Plassforholdene i pankreas er gode og en svulst vil ikke presse på nærliggende organer før den er blitt relativt stor. Dette gjør at personen med pankreaskreft ofte ikke kjenner smerter eller ubehag før sent i utviklingen hvor svulsten ofte er blitt stor og metastasert. På grunn av vage symptomer blir svulsten ofte oppdaget sent og prognosen til pasienter med pankreaskreft er ofte dårlig. Veksten til pankreaskreft er beskrevet hos de fleste pasienter til å ha et veldig aggressivt vekstmønster som gjør pasienten svært syk. Pankreas ligger også vanskelig til i kroppen og nært andre livsviktige organer. Når diagnosen er stilt er 80-85% av tilfellene ikke reelt mulig å behandle med kirurgisk kurative inngrep. De fleste (>50%) svulstene har metastasert og de færreste (<10%) danner avgrensede, lokaliserte forandringer. På grunn av dette vil de fleste pasienter kun få palliativ behandling (6). I Norge er fem års overlevelse med pankreaskreft, 10,9% for menn og 10,5% for kvinner (2). Gjennomsnittlig overlevelsestid er under 1 år etter diagnositidspunkt og under 6 måneder ved metastasert kreft (2,7).

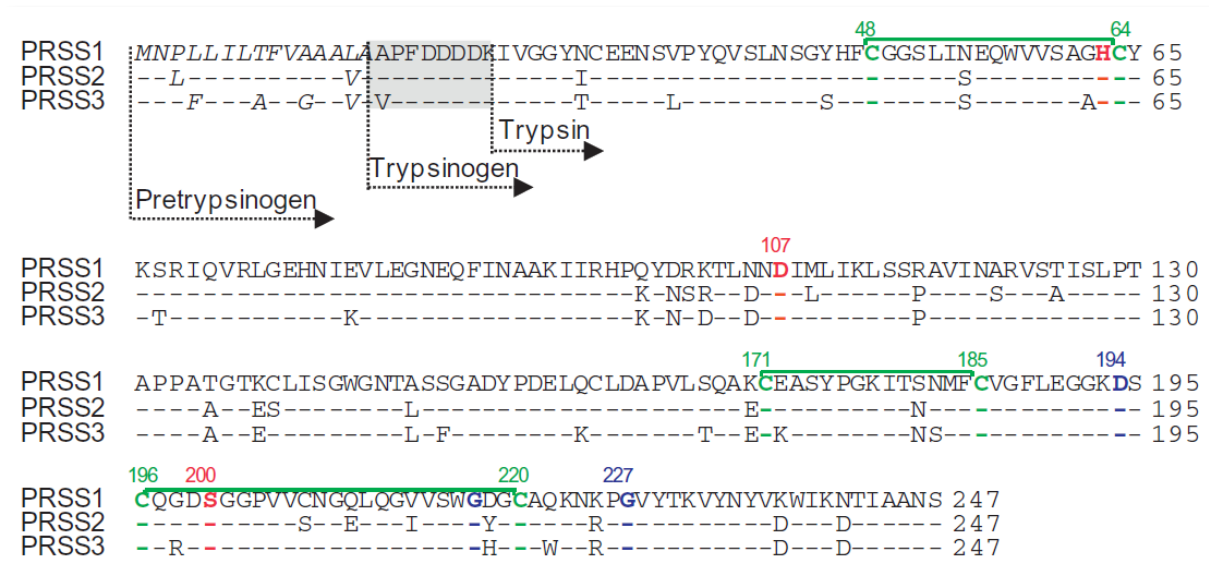
Kakeksi med ufrivillig vekttap er en av de vanligste årsaken som fører til legebesøk der svulsten oppdages (6,7). Andre symptomer som ofte er utløsende årsaker for å oppsøke lege er utilstrekkelig opptak av næring, smerter, og gulsott. For at gulsott skal opptre må svulsten vokse inn i eller trykke på gallegangen. Når gallegangen blir blokkert vil gallen ikke kunne dreneres ut i tarmen, men blir tatt opp i blodet i stedet. Dette fører til en karakteristisk gulfarget hud, mørk urin og en lys avføring. Blokkering av gallegangen vil også kunne forårsake akutt pankreatitt. Ved akutt pankreatitt vil pankreas svulme av at det dannes væskeansamling i kjertelen. Dette kan utvikle seg videre til blødninger, nekrose og pussansamlinger. Akutt pankreatitt oppstår ved at fordøyelsesenzymer som produseres i pankreas gjør skade på vevet i pankreas. Dette skjer oftest som en konsekvens av at enzymene ikke blir drenert ut i tarmen. Normale årsaker er alkoholmisbruk ved alkoholisme, høyt kolesterolnivå som fører til forkalkning ved hyperkolesterolemi, eller en blokkeringen av gallegangen ved gallestein/kreftsvulst (12). I vestlige land er de fleste tilfellene av akutt pankreatitt forårsaket av alkohol. Ved mangel på drenering vil fordøyelsesenzymene starte å degenerere cellene i pankreas. En overaktivering av fordøyelsesenzymene vil også kunne gjøre tilsvarende skade og gi degenerasjon av pankreasvev.

Alvorlighetsgraden til akutt pankreatitt kan variere fra mild til alvorlig der 80% av tilfellene er milde (13). Symptomene starter med magesmerter med utstråling i ryggen. Dette kan utvikle seg til kvalme og oppkast, og eventuelt smerter i hele magen, men også mer alvorlig symptomer som sirkulasjonssvikt. Smertene forverrer seg over de første par timene. Smertene oppleves spesielt etter matinntak hvor pankreas begynner å skille ut fordøyelsesenzym som en forberedelse til maten som snart kommer til tarmen. Siden enzymene ikke forlater pankreas, eller har høyere aktivitet enn vanlig, vil det tære på vevet og gi konstante smerter som når ett platå som kan vare i flere dager. Som regel forbedrer situasjonen seg etter noen dager. Hvis symptomene vedvarer kan det ha skjedd lokale komplikasjoner som akutte væskeansamlinger, nekrose, infeksjon eller pseudocyster (7). Dette kan føre til innleggelse på sykehus i opptil flere måneder i alvorlige tilfeller. Ved mistanke om akutt pankreatitt vil det bli tatt en blodprøve for å analysere mengde amylase og lipase i blodet. Disse enzymene produseres i pankreas og konsentrasjonen øker i blodet ved pankreatitt. Ved mistanke om gallestein blir også ultralyd benyttet. Ved mistanke om lokale komplikasjoner vil CT billedtakning (eventuelt MR eller vanlig røntgen) benyttes for å se etter hevelse i pankreas og væskeansamlinger utenfor kjertelen. Behandling av akutt pankreatitt består av væsketilskudd, faste, smertestillende medikamenter og nøye overvåking. For å forhindre nye anfall og etter hvert kronisk pankreatitt er det gunstig å behandle de underliggende årsakene som gallestein, hyperkolesterolemi eller alkoholisme, i de tilfellene det er mulig.

Dersom vevet i pankreas ikke normaliseres etter et anfall med akutt pankreatitt kan det gå over til å bli kronisk pankreatitt. Dette skjer gjerne etter flere gjentatte anfall med akutt pankreatitt. Pankreas vil på dette tidspunktet være i en tilstand med arrdannelse fra tidligere betennelser, ødelagte gangsystem i kjertelen og generelt nedsatt funksjon. Ettersom kjertelfunksjonen synker vil fordøyelsen til personen bli dårligere og personen kan oppleve vekttap og diaré. Utvikling av diabetes observeres hos noen grunnet nedsatt insulinproduksjon som også er lokalisert i pankreas. Kroniske smerteproblemer i mageområdet er ikke uvanlig. Et totalavhold til alkohol er nødvendig ettersom dette belaster kjertelen. Dette må pasienten gjøre selv om alkoholisme ikke var avgjørende faktor for utviklingen til kronisk pankreatitt. Tilstanden til pankreas vil på dette tidspunktet være spesielt sårbar, hvor nye anfall med akutt pankreatitt oppstår lettere. Dette fører til en ond sirkel der det skal mindre og mindre til for nye anfall. Behandling av kronisk pankreatitt består av erstatning av tapt kjertelfunksjon og passe på at pasienten ikke blir underernært. Den tapte kjertelfunksjon blir korrigert med enzympreparater for å hjelpe fordøyelse av mat. Hvis det oppstår forkalkninger i kjertelgangene i pankreas behandles dette med endoskopi. Har insulinproduksjonen blitt svekket vil diabetesbehandling være nødvendig. Kroniske smerter er vanskelig å behandle annet enn kortvarig.

Noen kan være genetisk disponert for anfall med akutt pankreatitt, en tilstand kalt arvelig pankreatitt. Arvelig pankreatitt oppstår ofte på grunn av en genfeil i ett av genene som kontrollerer fordøyelsesenzymer eller regulerer disse. *PRSS1* er det mest fremtredende i denne sammenhengen, men også *SPINK1*-, og *CTRC*-genet er knyttet til arvelig pankreatitt (14). Akutt pankreatitt tidlig i livet kan være ett tegn på arvelig pankreatitt (1). Anfall med akutt pankreatitt vil i en slik tilstand forekomme hyppigere og dermed gjøre permanente skader på pankreas i en relativt ung alder. Utvikling av kronisk pankreatitt skjer derfor raskere enn hos personer uten en slik genfeil (5), og inntreffer ofte 20-30 år tidligere enn «normalbefolkningen». Tidlig i denne sammenhengen er før 40 år. Kronisk pankreatitt i barndommen er ofte assosiert med en genfeil i *PRSS1* (15,16). Ved arvelig pankreatitt vil det etter hvert forekomme en vedvarende skade på cellene i pankreas. Slik skade og hyppig gjentatt inflammatoriske responser gir økt risiko for kreft (3,5,8,15). Pasienter med påvist genfeil vil derfor få tilbud om tettere oppfølging for kontroll av eventuell kreftutvikling, med mål om å øke overlevelsen.

PRSS1-genet koder for kationisk trypsinogen. *PRSS2*-genet som koder tilsvarende for anionisk trypsinogen. *PRSS3*-genet koder for mesotrypsin, men er ikke knyttet til arvelig pankreatitt. Trypsinogen er et proenzym som må bli aktivert for at det skal være et funksjonelt enzym. Omdannelsen fra trypsinogen til trypsin skjer vanligvis ved hjelp av enterokinase som finnes i tarmsaften.

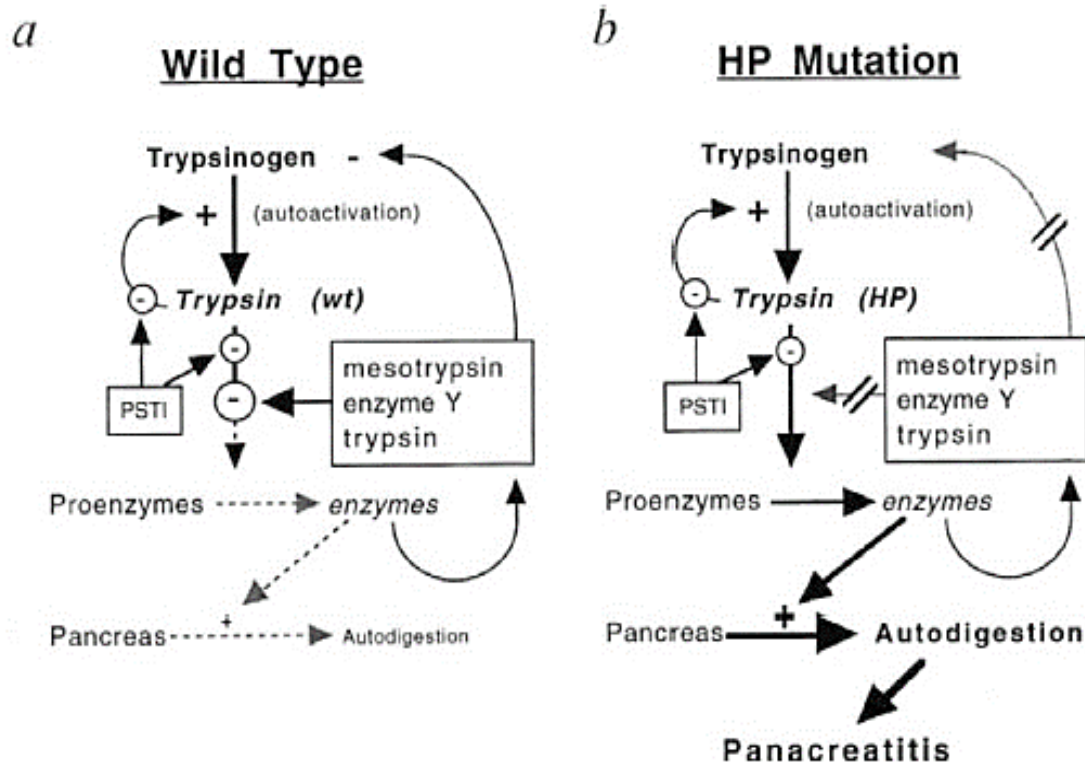


Figur 1: Aminosyresekvenser av humankationisk trypsinogen (*PRSS1*), anionisk trypsinogen (*PRSS2*) og mesotrypsinogen (*PRSS3*) med viktige aminosyrer uthevet. Den katalytiske triaden med His63, Asp107 og Ser200 er uthevet i rødt. De tre aminosyrene som bestemmer trypsinspesifisitet, Asp194, Gly217 og Gly227 er uthevet i blått, og de tre felles disulfidbroene (48–64, 171–185 og 196–220) er uthevet i grønt. Bindestreker indikerer overenstemmelse med *PRSS1*-sekvensen. Figur hentet fra Rawlings og Salvesen (15).

Som Figur 1 illustrerer inneholder hver av de tre trypsinogen kodende genene et N-terminal signal peptid på 15 aminosyrer som er området fra pretrypsinogen til trypsinogen. Trypsinogen syntetiseres i form av et prezymogen (dvs. pretrypsinogen). Fjerning av signalpeptidet resulterer i dannelsen av trypsinogen. Trypsinogen inneholder et kort aktiverings peptid på åtte aminosyrer. Videre fjerning av aktiveringspeptidet (skyggelagt i grått) fører til dannelse av trypsin og skjer enten av serin protease (enterokinase) som er en fysiologisk trypsinogen aktivator med spesifisitet for (Asp)⁴-Lys-Ile sekvensen, eller trypsin(ogen) i seg selv som nevnes senere. Noen aminosyrer som er kritiske for trypsinets struktur og funksjon er bevart gjennom evolusjonen av virveldyr. Selve starten på trypsin er den katalytiske triaden med His63, Asp107 og Ser200 som finnes i alle familier av serin peptidaser. Serin proteaser er også kjent som serin endopeptidaser. De tre aminosyrene som bestemmer trypsinets spesifisitet er Asp194, Gly217 og Gly227. Det finnes også seks absolutt bevarte cysteinsyrer som er nødvendige for å bygge de tre disulfidbroene (48-64, 171-185 og 196-220) observert i trypsiner i alle virveldyr. De primære transkriberte polypeptidene til *PRSS1*, *PRSS2* og *PRSS3* deler derfor betydelig sekvenslikhet.

En liten mengde av trypsinogenet blir selvaktivert til trypsin i pankreas. Trypsin aktiverer også trypsinogen til trypsin. Aktivt trypsin fungerer også som aktivator av andre proenzymmer som finnes i pankreas, som fører til videre selvfordøyelse (17). Dette fører til en ugunstig degenerasjon av vevet. I pankreas finnes det mekanismer for å forhindre selvfordøyelse av sine egne fordøyelsesenzymmer (18). Hoved-mekanismen for hemning av aktivt, selvaktivert trypsin er pankreatisk sekretorisk trypsininhibitor (PSTI). Dersom enten produksjonen av aktivt trypsin overstiger kapasiteten til eller unngår hemning av PSTI vil trypsin-lignende enzymer bli aktivert (Mesotrypsin og Enzym Y). Disse fungerer som en andre forsvarslinje mot start på selvfordøyelse og pankreatitt.

En sykdomsassocisert sekvensvariant som er beskrevet i *PRSS1*-genet er Arg117His substitusjon. Denne er vist å ha en «gain-of-function» effekt på trypsin som reduserer evnen trypsin-lignende enzymene har til å bryte ned aktivt trypsin i pankreas. Mengden aktivt trypsin kan dermed overstige kapasiteten til PSTI slik at aktivt trypsin kan aktivere andre proenzymmer ukontrollert i pankreas og forårsake pankreatitt (Figur 2). Denne varianten er vist å ha 80% penetrans for arvelig pankreatitt (1).



Figur 2: Modell av feedback-løkken for trypsin hos villtype (a) og mutasjon for arvelig pankreatitt (b) der b fører til pankreatitt.

a. Ved selvaktivering og enzymatisk aktivering av trypsinogen dannes sporbare mengder aktiv trypsin i bukspyttkjertelens acinære celler. Aktiv trypsin hemmes av en begrenset tilførsel av pankreatisk sekretorisk trypsininhibitor (PSTI). Hvis trypsinaktiviteten overstiger den hemmende kapasiteten til PSTI, da vil proenzymene mesotrypsin og enzym Y, aktiveres. Aktiveringen av disse enzymene antatt til å være en del av feedback-løkken for inaktivering av villtype (wt) trypsinogen, trypsin og andre proenzymmer.

b. Modell av mutant (HP) trypsinaktivering som overskrider den hemmende kapasiteten til PSTI, som resulterer i ukontrollert aktivering av proenzymmer. Siden Arg 117-spaltningstedet erstattes av His for mutant trypsinogen vil mesotrypsin, enzym Y og trypsin ikke kunne hydrolysere aktivt trypsin. Dette fører til at trypsin fortsetter å aktivere trypsinogen og andre proenzymmer i pankreas uhemmet, noe som fører til selvfordøyelse av vev i bukspyttkjertelen og tilstanden pankreatitt. Figuren er hentet fra Whitcomb, et al. (1).

Metodiske forutsetninger

Avdeling for medisinsk genetikk (AMG) har som en av sine hovedoppgaver å finne årsak til arvelig sykdom. En av sykdommene som AMG kan bidra med å diagnostisere er arvelig pankreaskreft. Det finnes flere gener som har en kjent forbindelse med arvelig pankreaskreft, og ett av disse er *PRSS1*-genet. Sykdomsassosierte sekvensendringer i dette genet er kjent for å kunne forårsake arvelig pankreatitt. Dette fører til skade på pankreas og gir en økt risiko for kreftutvikling (5). Ved AMG benyttes hovedsakelig nestegenerasjonssekvensering (NGS) for analyse av basesekvensen i *PRSS1*-genet for å oppdage slike endringer. NGS genpanelet hvor *PRSS1*-genet er inkludert heter «NGS-bukspyttkjertelkreft-CNV».

NGS teknologien kan brukes til å sekvensere mange gener samtidig og blir mye brukt i diagnostikken. Med NGS blir flere millioner fragmenter analysert parallelt og blir omtalt som «high throughput sequencing» (høykapasitets sekvensering). Ved NGS vil DNA-et bli tilfeldig fragmentert. Dette fører til at sekvenseringsdata ikke vil komme i kronologisk rekkefølge. For å få en sammensatt sekvens vil alle leste sekvenser fra fragmentene sammenlignes med en referansesekvens og alignes (sammenstilles) mot denne. Ett problem oppstår når en lest sekvens passer inn flere plasser i genomet. *PRSS1*-genet er en del av en genfamilie der det finnes flere gener og pseudogener som er veldig like *PRSS1*-genet. Generelt benyttes begrepet homologe sekvenser om områder i genomet med store likheter. Denne homologien fører til at det er liten sikkerhet til at avleste sekvenser blir plassert riktig ved bruk av NGS. Pseudogener har en tendens til å akkumulere flere varianter enn sine beslektede funksjonelle gener siden de sjeldent er under selektivt press. Pseudogenene har ikke analytisk verdi og er ikke ønskelig å sekvensere. En av komplikasjonene dette medfører er at en må finne ut om en variant finnes reelt i *PRSS1*-genet eller skyldes en forveksling av et gen-segment med stor nok likhet til å forstyrre NGS sin sammenstilling. Det er også beskrevet forflyttelse av ett ekson fra *PRSS2*-genet inn i *PRSS1*-genet kjent som genkonversjon (4). Genkonversjon kan ikke detekteres med NGS, da analyserte sekvenseringsdata ikke kan tilordnes *PRSS1*-genet, men tilordnes *PRSS2*-genet, slik at en årsak til pankreatitt kan bli oversett.

Det har derfor oppstått ett behov hos AMG for en tilleggsanalyse når *PRSS1*-genet skal sekvenseres. Denne analysen må kunne sekvensere *PRSS1*-genet med eventuell genkonversjon uten forstyrrelser fra pseudogener eller homologe sekvenser. Ved å benytte Sanger-sekvensering vil vi fortsatt støte på mye av de samme problemene med at primere som skal feste seg i området som skal sekvenseres vil kunne binde seg til lignende områder i andre gener/pseudogener i trypsin-familien. Derfor virker en Sanger-sekvensering som benytter en spesifikk «long-range PCR» (LR PCR) som kun oppformerer *PRSS1*-genet, som den mest logiske analysemetoden. Long-range PCR er en metode som baserer seg på å kopiere lange nukleotidsekvenser, og kan benyttes for å være sikker på at en har fått en oppformering av det spesifikke genet en er interessert i. Metoden baserer seg på at områdene utenfor genene er unike med lite homologe sekvenser. Dermed kan en designe primere som fester seg spesifikt utenfor *PRSS1*-genet og amplifisere kun det aktuelle genet.

Til en vanlig sekvenseringsanalyse blir prøvematerialet amplifisert ved hjelp av en PCR-reaksjon med en produktlengde på 200-400 basepar (bp). Lengden på produktet er nødt til å være i dette intervallet for å Sanger-sekvensere produktet med vanlige rutinebetingelser. Ved long-range PCR vil lengden på produktet gå utover dette og kan danne produkt på størrelser opp til 20kilobaser (kb). Dette kan benyttes for å amplifisere et helt gen i en og samme PCR, og ikke bare den biten av genet som skal sekvenseres. PCR reaksjonen for *PRSS1* vil nå være spesifikk. Sekvenseringsprimere som

ellers ville vært uspesifikke for *PRSS1* kan benyttes på PCR-produktet ved videre nukleotidsekvensering.

En long-range PCR fungerer på mange måter helt likt som en vanlig PCR-reaksjon. Begge bygger på prinsippet med de følgende fem trinnene der trinn 2-4 gjentas tilstrekkelig antall ganger for å oppnå nok produkt: 1 - Primær denaturering av DNA-et og aktivering av enzym. 2 - Denaturering. 3 - Hybridisering. 4 – Polymerisering. 5 – Forlengelse. Siden det er en lengre sekvens som skal amplifiseres er det nødvendig at noen av stegene i PCR-reaksjonen modifiseres for å tilrettelegge for et lengre PCR-produkt. For at et langt dobbeltrådet DNA-fragment skal denatureres fullstendig, må en høy temperatur opprettholdes over en lengre tid enn ved kortere fragment på grunn av færre hydrogenbindinger som må brytes. Dette fører til at trinnene som inkluderer denaturering blir forlenget. Primere til long-range PCR fungerer likt som ved normal PCR reaksjon. Dette fører til at hybridisering i long-range og normal PCR ikke er spesielt annerledes. Polymerisering i LR PCR vil ta lengre tid enn vanlig PCR. Polymerasen omtalt i denne oppgaven hefter på nukleotider (nt) i et nærmest konstant tempo på ca. 1kb per minutt. Dette fører til at ved en lengre utgangssekvens vil polymeriseringstrinnet ta lengre tid.

Oppgavens hensikt

Hos AMG har det oppstått ett behov for en tilleggsanalyse av *PRSS1*-genet. Ved å benytte long-range PCR før sekvensering av *PRSS1*-genet forventer vi å få et produkt som gjør det mulig å spesifikt sekvensere *PRSS1*-genet uten at resultatet forstyrres av pseudogener eller andre homologe sekvenser.

Hensikten med denne oppgaven er å: **Etablere og validere en målrettet analysemetode for *PRSS1*-genet ved hjelp av long-range PCR.**

Trypsinogen genfamilien

PRSS1

PRSS1 er forkortelsen på protease, serin, 1-genet som koder for et trypsinogen. Trypsinogen er den inaktive formen til trypsin og fungerer ikke som et fordøyelsesenzym før det omdannes til den aktive formen trypsin. Mutasjoner av det humane kationiske trypsinogen-genet (*PRSS1*) er ofte forbundet med arvelig pankreatitt. p.N29I og p.R122H er de hyppigst forekomne variantene som er anerkjent som sykdomsfremkallende mutasjoner.

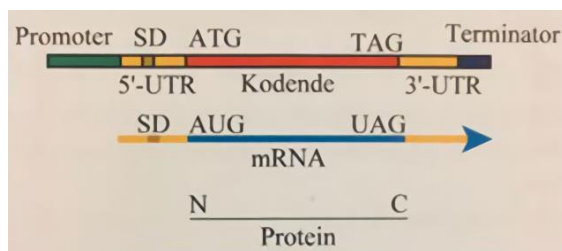
Den humane trypsinogen genfamilien består av 9 svært homologe gener som er fordelt i to grupper. Gruppe 1 består av seks gener totalt. Fem av disse genene; de to funksjonelle genene *PRSS1* og *PRSS2*, de tre pseudogenene *PRSS3P1*, *PRSS3P2* og *TRY7*, befinner seg på 3'-enden i det humane T-celle reseptor β (TCR β) locus på kromosom 7 (7q35). Det siste, *PRSS3*, er translokalisert til kromosom 9 (9p13). Gruppe 2 består av tre pseudogener som finnes i 5'-enden av TCR β locus og har utviklet seg svært mye i forhold til gruppe 1 pseudogener (19), og omtales ikke videre i denne oppgaven.

Trypsin er som nevnt innledningsvis serin proteaser som finnes som 3 ulike isoenzymer. Kationisk trypsin-1 utgjør den vanligste av dem og blir kodet av *PRSS1*-genet. I tillegg har vi anionisk trypsin-2 som *PRSS2*-genet koder for, og mesotrypsin som *PRSS3*-genet koder for. Sistnevnte vil ikke bli diskutert i denne oppgaven. Trypsin-1 og trypsin-2 virker begge to ved å bryte opp peptidbindinger mellom karboksylgruppene til lysin (Lys) og arginin (Arg). Genene som koder for dem har av samme grunn svært høy genetisk likhet. Den genetiske likheten finner vi igjen hos mange andre arter enn kun mennesket. Selv arter som ligger langt unna hverandre i evolusjonen, kan vise stor genetisk likhet innenfor visse gener, eksempelvis *PRSS1*. En kan si at genetisk likhet ikke nødvendigvis er ensbetydende med evolusjonært slektskap. Evolusjonært slektskap betegner den relative tiden mellom to arters felles stamform, og presenteres ofte i form av stamtrær ved fylogenetiske slektskapsanalyser. Genetisk slektskap viser til andelen av gener to individer innenfor eller mellom arter har til felles. Et kjent prinsipp i evolusjon er at noen gener innenfor en art i et stamtre utvikler seg raskere enn andre gener. Hvis en ønsker å se på kun et bestemt gen mellom flere ulike arter, eksempelvis *PRSS1*-genet, istedenfor en helhet i form av genomet til arten, vil det være umulig å sammenstille både et genetisk og evolusjonært slektskap. Det kommer av at en art på en gren kan ha større forandringer innenfor noen gener og en genetisk slektskapsanalyse for et gen vil dermed vise at arten er mindre i slekt med andre arter enn et stamtre vil tilsi. Vi benytter begrepet homoplasi for arvelige egenskaper som ikke har vært tilstede i felles forfedre, men som har utviklet seg parallelt i ulike arter. Ved konvergent evolusjon har det skjedd en uavhengig utvikling av like fenotypiske trekk i

forskjellige arter, som ikke skyldes at organismene nedstammer fra en felles stamform med trekket. Trekket har blitt utviklet via naturlig utvalg i hver av artene uavhengig av hverandre. Konvergent evolusjon finner ofte sted hos arter som lever i samme miljø eller utnytter liknende ressurser, og kan i visse tilfeller også kalles parallell evolusjon siden utviklingen av like trekk med samme funksjoner er et resultat av samme type seleksjonspress.

Noen elementer i DNA-sekvensen har viktige funksjoner, og vil ha større selektivt press og forandres mindre enn områder med lite selektivt press. Noen promotorelementer, dvs. sekvenser som binder regulerende proteiner og tilsvarende transkripsjonsfaktorer, er felles for svært mange gener og kalles generelle. Derimot kan det innenfor visse gener og genfamilier mangle noen områder eller være andre spesifikke promotorelementer (20). Vi ser blant annet ingen CAAT-boks i *PRSS1*-genet (Figur Supplement 1) og det er heller ikke beskrevet i litteraturen vi har undersøkt.

En Coding Sequence (CDS) er en region i DNA eller RNA med en sekvens som koder for aminosyresekvensen som utgjør et protein. All CDS er åpne leserammer (Open Reading Frame (ORF)), men ikke alle ORF er CDS da de kan inneholde stop kodon (21). Transkripsjonen starter et stykke før CDS, ved promotor, og avsluttes nedstrøms ved terminator. Det blir dermed dannet mRNA fra sekvensen som er lengre enn CDS. Endene til mRNA blir med andre ord ikke translert og kalles derfor 5'-UTR og 3'-UTR (un-translated region), evt. UTS (S = sequence) (Figur 3)) (20). Hensikten med slike annoteringer (forklaringer) er å tilordne en funksjon til de ulike delene i DNA-sekvensen slik at det ikke bare er en lang rekke ATCG uten mening for oss.



Figur 3: Figuren viser ulike generelle elementer i et proteinkodende gen som; promotor, 5'-UTR, kodende DNA, 3'-UTR og terminator. SD indikerer området der ribosomet binder seg. Figur hentet fra Sjøberg (20).

Hvis en ønsker å kun se på Trypsinogen genene og slektskapet mellom mennesket og andre arter som lever i dag, vil vi på grunn av homoplasi måtte se bort ifra et stamtre, og heller sammenlikne det kodende området av genene, CDS, som gir det samme produktet, trolig som et resultat av konvergent evolusjon, nemlig trypsin-1, -2 og mesotrypsin. For å få et perspektiv på hvor likt *PRSS1*-genet er mellom mennesket og andre arter, kan en konstruere et fylogenetisk tre basert på produktet mRNA sekvensen til *PRSS1*-genet koder for, istedenfor det evolusjonære slektskapet.

PRSS2

Anionisk serin protease 2 (*PRSS2*) tilhører Trypsin familien av serin proteaser og koder for anionisk trypsinogen. *PRSS2*-genet er svært likt *PRSS1*-genet og er et funksjonelt gen. Det er ingen varianter som er assosiert med «gain-of-function» i *PRSS2*-genet som er knyttet til arvelig pankreatitt. Dette er trolig på grunn av to årsaker: Mengden *PRSS2* som blir syntetisert er omtrent halvparten av det *PRSS1* står for og *PRSS2* selvaktiveres saktere samtidig som det brytes ned hurtigere enn *PRSS1* (14).

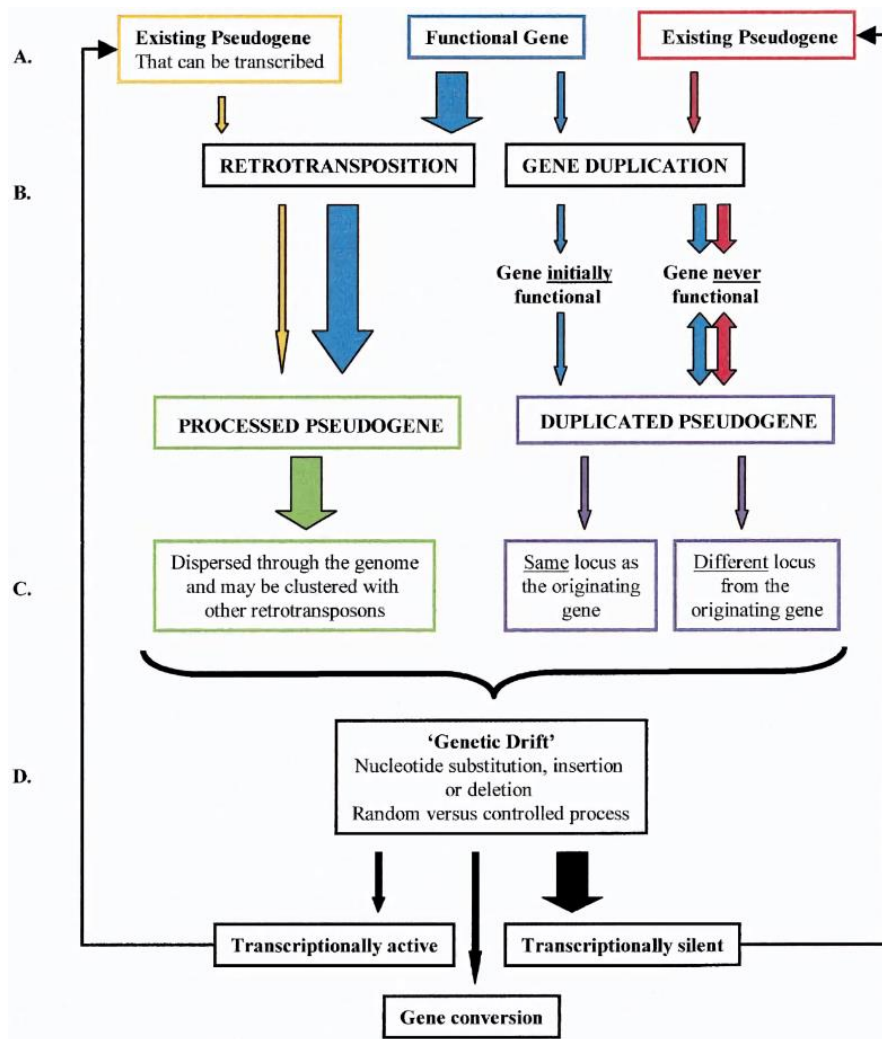
Grunnet homologien mellom *PRSS1* og *PRSS2* er de utsatt for genkonversjon, som omtales senere. Den mest fremtredende varianten som følger av genkonversjonen er der isoleucin er normalvariant på det samme punktet i ekson 2 i *PRSS2* og gir mutasjonen p.N29I i *PRSS1*-genet.

Det finnes kvalitative forskjeller mellom loss-of-function varianter i *PRSS1* og *PRSS2*. En totalt «loss-of-function» genvariant i *PRSS1* er trolig under svært negativt selektivt press, gitt viktigheten av kationisk trypsinogen sin rolle i fordøyelsen. *PRSS2* bistår kun halvparten så mye til trypsinaktivitet enn *PRSS1*. I en studie om arvelig pankreatitt ble det vist at en missense mutasjon, p.Gly191Arg, ble funnet overrepresentert i *PRSS2* hos friske individer i motsetning til individer berørt av pankreatitt. Denne mutasjonen viste et totalt tap av trypsinaktivitet i *PRSS2*. Substitusjonen Gly191Arg lagde et nytt kløyvingsted hvilket gjorde enzymet hypersensitivt for autokatalytisk proteolyse. En slik svak reduksjon i total aktivitet til trypsin i pankreas har vist seg å ha en beskyttende effekt mot pankreatitt og dermed gitt ett selektivt press for varianten (22).

Pseudogener

Et pseudogen er en gen sekvensen som er relatert til en annen funksjonell sekvens, men som har mistet sin funksjon. Disse ikke-funksjonelle pseudogenene har en feil i enten transkripsjon, translasjon eller har produksjon av et protein som ikke har lik funksjon som proteinet som blir produsert av det normale paraloge gen. Den viktige karakteristikken til disse pseudogenene er en endring av gen sekvensen på et kritisk punkt (23) som gjør dem defekt. Et pseudogen blir navngitt på flere måter inkludert greske tegnet « ψ » (psi) som prefiks eller med «P» suffiks. Eks. ψ PGK-1 eller CYP21P. Hvis det er flere pseudogener som stammer fra samme opprinnelige gen vil det kunne merkes med nummer etter P som f.eks *PRSS3P1* og *PRSS3P2*

Pseudogener kan oppstå på flere forskjellige måter gjennom forskjellige mekanismer (Figur 4).

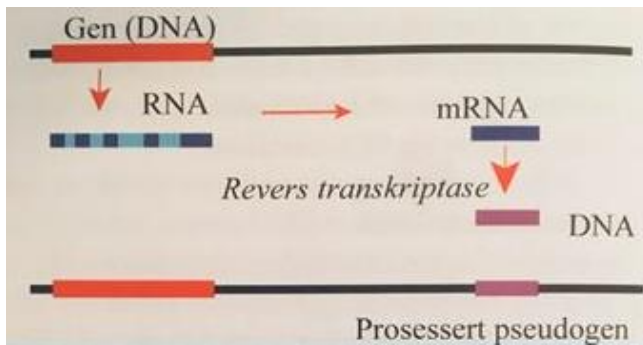


Figur 4: Skjematisering av opprinnelse og utvikling til ulike typer pseudogener. A: Pseudogenets opprinnelse. Størsteparten av pseudogener oppstår fra et funksjonelt gen. B: Mekanisme for dannelse av pseudogen. Den største andelen av pseudogener er et resultat av retrotransposisjon til et kodende gen. C: Pseudogenets plassering i genomet. D: Resultat av pseudogenet. De fleste pseudogener har gjennomgått en genetisk drift og blir aldri transkribert. I noen av pseudogener virker det som at det har vært et selektivt press som forhindrer store endringer i pseudogenekvensen. Noen pseudogener er involvert i genkonversjon og noen få kan transkribes. Figuren er hentet fra Mighell et al. (23).

Pseudogener som oppstår som følger av en retrotransposisjon er kjent som et prosessert pseudogen eller et retro-pseudogen. Retrotransposisjon er når en DNA sekvens blir satt inn i genomet via et RNA intermediet som blir reverstranskribert. Typiske kjennetegn er mangel på 5'-promotor sekvens og introner. Prosesserte pseudogener er retrotransposomer som i likhet med andre retrotransposomer har blitt satt inn i genomet som dobbelttrådet DNA transkribert fra enkelttrådet RNA.

Retrotransposisjon av en hvilken som helst gensekvens kan være truende for normal gen-funksjon. Mekanismene som kontrollerer disse prosessene er fortsatt dårlig forstått (23).

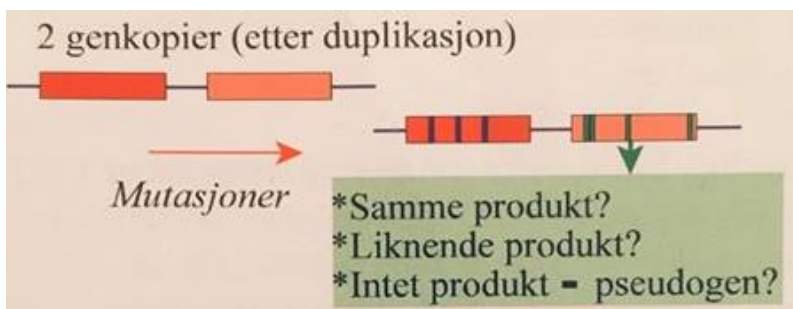
Flesteparten av retrotransposerte gener blir inaktivert til prosesserte pseudogener, men ved noen få anledninger kan det retrotransposerte materiale fungere som et funksjonelt, intronløst gen (Figur 5).



Figur 5: Dannelse av prosessert pseudogen som prosess av retrotransposisjon. Figur hentet fra Sjøberg (20).

Prosesserte pseudogener kan man finne spredt i hele genomet og ikke nødvendigvis/kun i nærheten av genet. På grunn av mekanismene til retrotransposisjon er noen locus i genomet lettere mottakelig for retrotransposomer, inkludert prosesserte pseudogener, og har derfor en tendens til å klynges i nærheten av hverandre. Enkelte isolerte retrotransposerte elementer kan også forekomme. Pseudogener har derfor en tendens til å samle seg i klynger sammen med andre transposomer og vil over tid for det meste finnes på plasser i genomet der de ikke har en skadelig effekt.

Pseudogener kan dannes som et produkt av DNA duplikasjon. En gen duplikasjon fører til at det dannes flere kopier av samme gen. Disse genene vil utvikle seg ulikt med tiden. Genene starter som nært beslektet, og etter hvert som de endrer seg kan de gi samme, liknende eller intet produkt og bli et pseudogen (Figur 6).



Figur 6: Utvikling av to genkopier som kan utvikle seg likt eller ulikt og potensielt bli inaktivt (blir til pseudogen). Figur hentet fra Sjøberg (20).

Duplikasjon av DNA segmenter har vært en drivkraft i utviklingen av kompleks genom og kan være med på å forklare dannelsen av en genfamilie fra ett opprinnelig gen. Over tid vil det uunngåelig forekomme endringer i genomet. Pseudogenene som produkt av duplikasjon representerer gener som ikke ble opprettholdt som funksjonelle elementer enten fordi de ga et negativt selektivt press eller fordi de ikke ga en seleksjonsfordel. Pseudogener som har oppstått som produkt av DNA duplikasjon har en større sannsynlighet til å befinne seg nærliggende det paraloge opprinnelige genet. De kan også ble satt inn i et annet kromosom ved translokasjon. Pseudogener har også blitt funnet på innsiden av andre gener (23).

Utviklingen til pseudogenet varierer svært ut ifra hvor det ble satt inn. Innersjon av noen pseudogener vil kunne ha en uønsket effekt på den normale funksjonen på andre gener og fører til et selektivt press mot pseudogenet. Dette fører til at det mistes. De fleste andre pseudogener vil forbli i genomet og vil utvikle seg over tid. Pseudogener som er etablert og dermed ikke påvirker funksjon til andre gener i genomet er forventet å gjennomgå genetisk drift. Når pseudogenet ikke får et selektivt press til å unngå mutasjoner vil resultatet bli at pseudogener over tid samler opp flere tilfeldige punktmutasjoner, innersjoner og delesjoner. Det kan også bli observert innersjon av retrotransposomer av større repeterte sekvenser, som Alu eller LINE sekvenser, i pseudogener. Innersjon av en slik sekvens er forventet å ha en svært ødeleggende effekt på et funksjonelt gen. Alu sekvenser finnes i noen funksjonelle gener og deres modne mRNA, men er svært sjeldent identifisert i eksoner. Hvis en slik repeterende region blir funnet i et ekson kan det raskt antas å tilhøre et pseudogen. Hvis to ganske sekvenslike gener blir analysert vil slike repetererte regioner i ekson fungere som en god pekepinn for å finne ut hvilket gen som er et pseudogen. Noen pseudogener har vist seg å være svært likt sitt paraloge funksjonelle gen. Andre har vist en hurtigere genetisk drift. Flere faktorer har en signifikant påvirkning på utviklingen av pseudogenene, men den største drivfaktoren til utviklingen er tiden det har gått siden pseudogenet ble dannet (23).

Genkonversjon

I diskusjonen om genetiske sykdommer hos mennesker, er genkonvertering (genkonversjon) som introduserer patogene varianter i funksjonelle gener en velkjent mutasjonsmekanisme. Genkonversjon er en prosess der en DNA-sekvens erstatter en sekvenslik (homolog) gensekvens slik at sekvensene blir identiske etter konverteringshendelsen (24).

Genkonversjon er en form for **genetisk (homolog) rekombinasjon** hvor en bit av en sekvens er blitt kopiert fra et donor gen (annet gen/pseudogen) og erstattet med den korresponderende like sekvensen i det paraloge genet (mottaker gen). Denne forbytningen er ikke gjensidig og sekvensen i mottakergenet blir «overskrevet», mens sekvensen til giveren forblir uendret. På grunn av den store likheten mellom sekvensene vil den effektive sekvensen fortsatt forbli for det meste den samme, med kun noen få ulikheter fra dens egen referansesekvens. Kommer den donerte sekvensen fra et annet funksjonelt gen vil det som er normalssekvens der effektivt gi varianter i mottakergenet, noe som kan være svært uheldig.

Genkonversjon kan være enten allelisk, noe som betyr at et allel av samme gen erstatter et annet allel, eller ektopisk, noe som betyr at en paralog DNA-sekvens konverterer en annen. I eukaryote celler er allelisk genkonversjon en form for meiotisk rekombinasjon som initieres gjennom dannelse

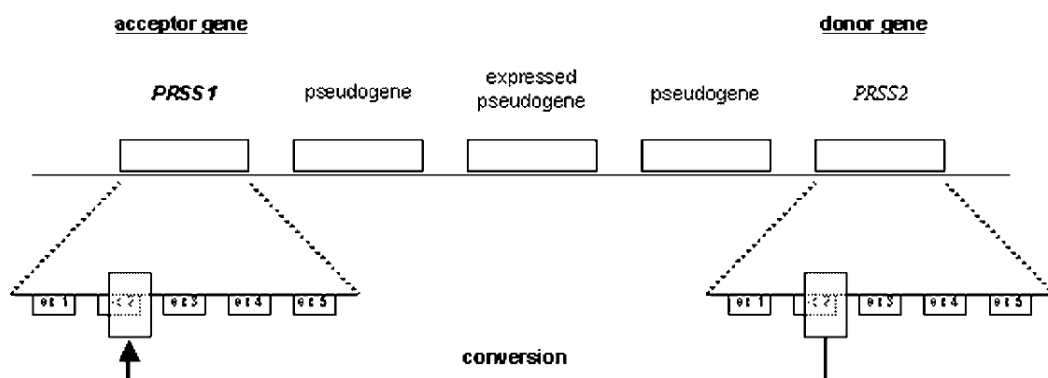
av et dobbeltstrengs brudd (DSB). Ved genkonversjon overføres genetisk informasjon fra intakte homologe sekvenser til regionen som inneholder DSB, og det kan forekomme mellom søsterkromatider, homologe kromosomer eller homologe sekvenser på enten det samme kromatidet eller forskjellige kromosomer (25). Hvis en av de fire kromatidene under meiosen pares opp med et annet kromatid, som kan oppstå på grunn av sekvenshomologi, kan homolog rekombinasjon resultere i en mismatch i baseparring. Dette misforholdet blir deretter gjenkjent og korrigeret av det cellulære maskineriet. Endene av nukleotidsekvensen med DSB blir klippet av 5' → 3'-eksonukleaser, noe som resulterer i dannelse av to 3' ssDNA haler, av og til med en lengde på flere hundre nukleotider. Disse halene 'skanner' aktivt genomet etter homologe sekvenser, oftest i umiddelbar nærhet på kromosomet (26). Som nevnt er både *PRSS2* og *PRSS3P2* gener i samme genfamilie som *PRSS1*, som sitter på samme kromosom som *PRSS1*. Disse sekvensene har områder med stor homologi og er gode kandidater til genkonversjon.

Et av 3'-ende ssDNA trådene invaderer og danner et mellomprodukt med en homolog sekvens. Mellomproduktet kan repareres enten ved overkrysning eller ikke-overkrysning. Ved forskjellige trinn i rekombinasjonsprosessen dannes heteroduplex DNA. Heteroduplex DNA er dobbeltstrengt DNA bestående av enkeltstrenger fra hvert av de to homologe kromosomene som kanskje eller ikke er perfekt komplementære. Når uoverensstemmelser oppstår i heteroduplex DNA, vil sekvensen til den ene strengen repareres for å binde den andre strengen med perfekt komplementaritet, noe som fører til konvertering av en sekvens til en annen (27). Reparasjonsprosessen kan følge en av to alternative veier som ikke vil bli videre utledet i denne oppgaven. Resultatet blir at en av allelene endres til å bli identisk med det andre, og kan forårsake en unormal fordeling av alleler i dattercellene også kjent som ikke-Mendelsk segregering av alleler i kimceller.

Normalt kalles en organisme som har arvet forskjellige kopier av et gen fra hver av foreldrene, heterozygot. Dette er generisk representert som genotype: Aa (dvs. en kopi av varianten (allel) 'A', og en kopi av allelen 'a'). Når en heterozygot skaper gameter ved meiose, dupliseres allelene normalt og ender opp i et 2:2-forhold i de resulterende 4 celler som er de direkte produktene av meiosen. Ved genkonversjon observeres et annet allele forhold etter meiose enn den forventede 2A: 2a, hvor A og a er de to allelene. Eksempler er 3A: 1a og 1A: 3a. Med andre ord kan det for eksempel være tre ganger så mange A-alleler som a alleler uttrykt i dattercellene, som tilfellet er i 3A: 1a. Dette kan påvirke produksjonen og/eller effekten av et eventuelt genprodukt.

PRSS2 har som nevnt blitt rapportert til å undergå en genkonversjon med *PRSS1*. Denne konversjonen erstatter minst 289 nukleotider (fra c.86 til c.200+174) og maks 457 nukleotider (fra c.41-34 og c.200+263) i ekson 2 og det påfølgende intron (Figur 7) i *PRSS1*. Effektivt fører dette til

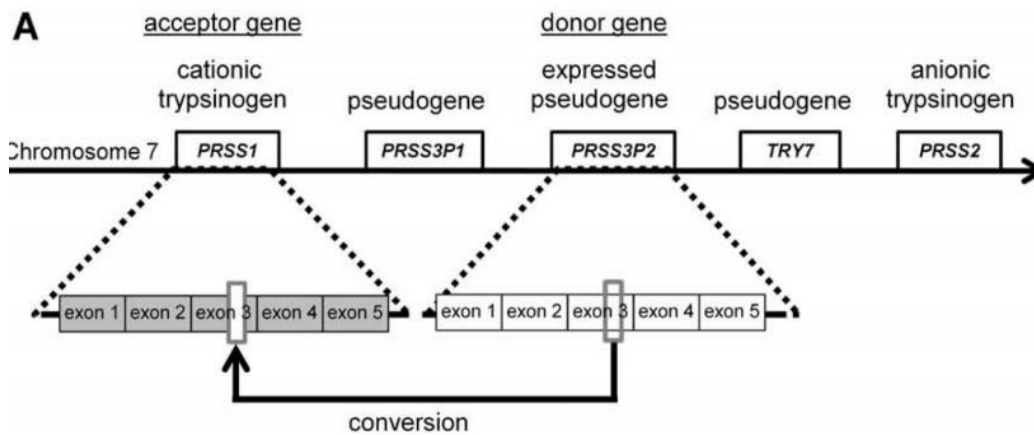
innføring av to punktmutasjoner i *PRSS1*. Variant c.86A>T og c.161A>G (NM_002769.4 RefSeq), som fører til aminosyresubstitusjon Asn29Ile (p.N29I) og Asn54Ser (p.N54S) respektivt (4). Av disse to var kun p.N29I i *PRSS1* påvist å gi en økt selvaktivering av trypsinogen. Dette er et eksempel på genkonversjon mellom to funksjonelle paraloge trypsinogen kodende gener som forårsaker arvelig pankreatitt. En slik «gain-of-function» mutasjon i *PRSS1*-genet er vist til å forårsake kronisk pankreatitt (1). Denne varianten er den nest vanligste varianten knyttet til arvelig pankreatitt og er årsak i ca. 25 % av tilfellene (3,4,28).



Figur 7: Skisse av ikke gjensidig genkonversjon med *PRSS1* som akseptor gen og *PRSS2* som donor gen. Figuren er hentet fra Teich et al. (4).

Varianten R122H i *PRSS1*-genet har også blitt vist å gi en «gain-of-function» effekt. Ved utbytting av Arg med His vil nedbrytningsmekanismene som kløyver Arg122-Val123 binding av trypsin ikke fungere grunnet endring av aminosyren. Det er flere varianter på dette punktet som gir patogen fenotype grunnet endring av kløyvningssete (14). Denne varianten kan komme fra genkonversjon fra *PRSS3P2* (Figur 8).

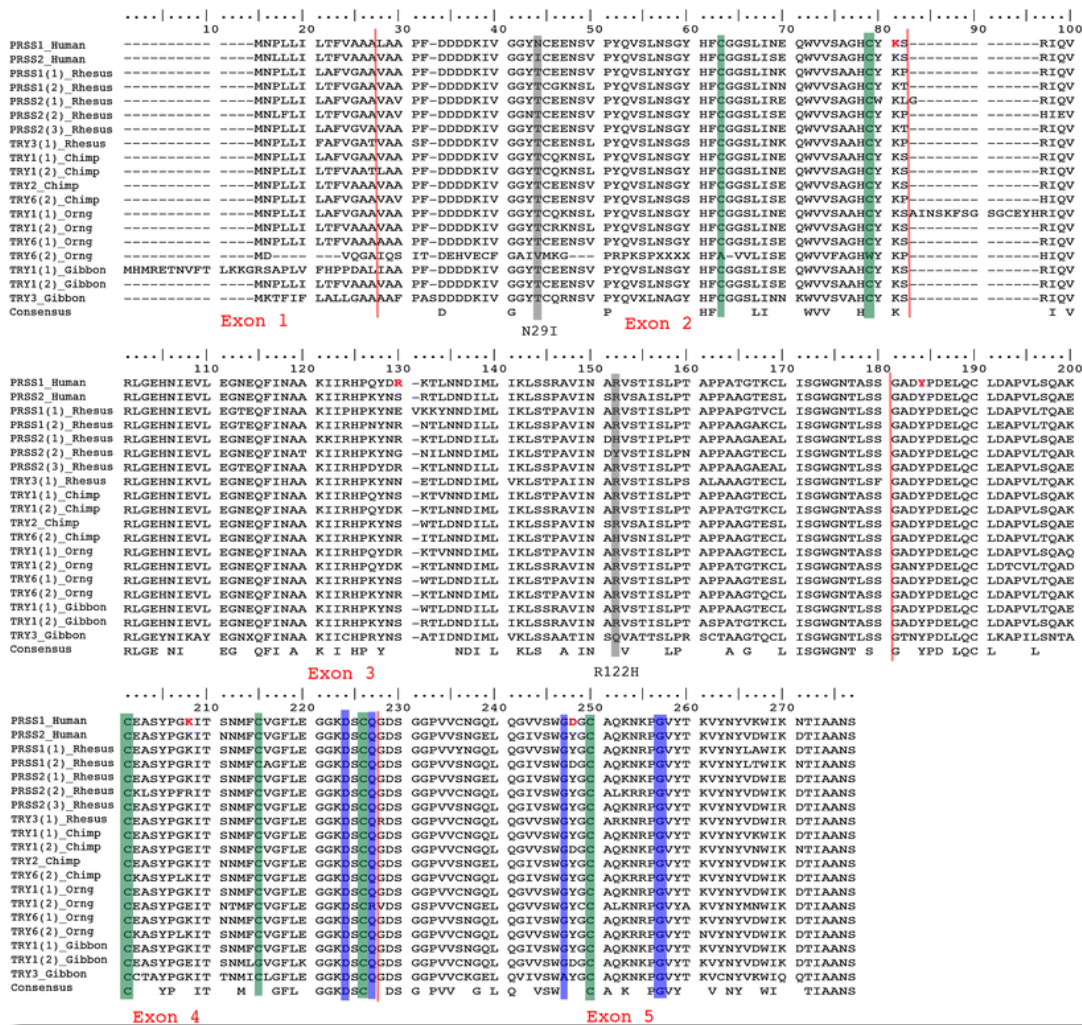
Pseudogener som ikke er drevet av selektivt press får oftere punktmutasjoner og vil derfor også effektivt gi punktmutasjoner til mottaker genet hvis det er et donorgen. En slik konversjon blir ansett som underliggende årsak til et økende antall genetiske sykdommer. Genkonversjon skjer vanligvis mellom et vanlig funksjonelt gen og et av dets pseudogen. Når en analyserer varianter som har oppstått på grunn av genkonversjon må en være sikker på at varianten faktisk befinner seg i det aktuelle genet. Ved uspesifikk sekvensering vil varianten komme frem i resultatene selv om den tilhører det opprinnelige pseudogenet.



Figur 8: Skisse av ikke gjensidig genkonversjon med *PRSS1* som akseptor-gen og *PRSS3P2* som donor-gen. Figuren er hentet fra Rygiel et al. (29).

I rutinemessig genetisk testing av pasienter med kronisk pankreatitt har det blitt indentifisert, hos to ubeslektede individer, lignende genkonversjon på mellom 24-71 nukleotider. Figur 8 illustrerer utvekslingen som skjedde i ekson 3 med *PRSS3P2* som donor-gen og *PRSS1* som akseptor-gen. Denne konversjonen resulterte i tre ikke-synonyme endringer: c.343T>A (p.S115T), c.347G>C (p.R116P) og c.365_366delinsAT (p.R122H). Denne fenotypen viste markant økning av selvaktivering av Trypsin som viste seg identisk med genotypen p.R122H (29). Denne genkonversjonen er årsak til arvelig pankreatitt.

Trypsinogen-genene deler store likheter i peptidsekvensen og finnes blant annet også i nært beslektede arter til mennesket (Figur 9).



Figur 9: Aminosyresekvensen til proteinene av funksjonelle trypsinogen kodende gener i primat-arter. Deler av sekvenser som anses som kritiske for trypsinfunksjon og struktur, og derfor bevart i alle primatartene som vises, er uthevet. Viktige punkt som bestemmer trypsin spesifisitet er vist i blått, de som er ansvarlige for å danne disulfid broer er vist i grønt, punkt som gir aminosyrer som samhandler med *SPINK1*-genet slik at trypsin produksjonen skal reguleres, er vist i rosa (14). Menneskelige *PRSS1*- og *PRSS2*-gener er representert av røde linjer og ytterligere identifisert nedenfor av et eksonnummer. Mutasjonene R122H og N29I som er forbundet med pankreatitt, er fremhevet i grått. Figuren er hentet fra Petronella og Drouin (28).

Figur 9 illustrerer aminosyresekvensen til trypsinogen fra *PRSS1* og *PRSS2* hos mennesker sammen med tilsvarende trypsinogen kodende gener fra primater, og vi kan se at genene har store likheter med hverandre. De to variantene N29I og R122H til *PRSS1* er markert i grått. I *PRSS1* ser vi i posisjonen for N29I at normalekvensen er N mens i *PRSS2* er normalekvensen på dette stedet I. Dette viser at ved en genkonversjon mellom disse genene vil *PRSS1* få varianten N29I. Ved varianten R122H ser vi normalvarianten R i *PRSS1* mens patogen variant er ikke å se i humane trypsinogen kodende gener. Varianten ser vi riktignok i det mesotrypsin-kodende TRY6 (29) hos orangutang. TRY6 (*PRSS3P2*) finnes også i menneske og er svært likt og inneholder den samme varianten c.365_366delGCinsAT på plassering 122. Ved genkonversjon kan også dette være årsak til patogen variant i *PRSS1*-genet.

PRSS3P2

Det er rapportert at *PRSS1* og *PRSS3P2* er mer enn 91 % identiske på nukleotidnivå og 89 % identiske på aminosyrenivå (29). *PRSS3P2* inneholder flere varianter som tilsvarer pankreatitt-assosierte mutasjoner i *PRSS1*. *PRSS3P2* blir ansett som et uttrykt pseudogen på grunnlag av en tilsynelatende intakt ekson/intron struktur som skal muliggjøre en normal transkripsjon (30,31). Selv om locus til *PRSS3P2* kan se ut til å kode for et protein som ligner trypsinogen, er det antatt å være et transkribert pseudogen. Søk i den menneskelige expressed sequence tag (EST) databasen for det antatt kodende DNA for trypsinogen gir lite bevis for at faktisk mRNA uttrykkes, og en regner derfor med at *PRSS3P2*-genet translateres til et ikke-kodende RNA (NR_001296.3). På grunn av den høye graden av homologi blant trypsinogen-genene er tolkning av publisert arbeid som hevder å kvantitere TRY6 (*PRSS3P2*), problematisk. Selv om genet blir transkribert og deretter oversatt til et proteinprodukt, forventes nivåene å være svært lave og biologisk ubetydelige. Noe som støttes av det faktum at en vanlig delesjon polymorfisme resulterer i fravær av *PRSS3P2* i et stort segment av befolkningen (30).

Transkripsjon er som kjent syntese av RNA som blir komplementær til templat-DNA (antisense-tråd). RNA vil da være lik den andre DNA-tråden (bortsett fra T/U), og denne DNA-tråden kalles den kodende tråd (sense-tråd). Hvilken DNA-tråd som blir brukt som templat-tråd kan være ulikt fra gen til gen, men innen ett bestemt område av DNA (ett gen) er hovedregelen at bare den ene tråden blir brukt. (20). Det er den kodende trådens baser som blir oppgitt når man skriver DNA-sekvensen av et gen, og baserekkefølgen vil tilsvare den vi finner i mRNA.

Når en DNA sekvens oppgis i databaser skrives den alltid i sense retning. Delen av en nukleotidsekvens som starter med et initierings/start kodon (ATG) og slutter med et stopp kodon (TAA/TGA) kalles som kjent en åpen leseramme (Open Reading Frame (ORF)). De fleste ORF-er som vi finner langs en tilfeldig DNA sekvens er ganske korte og ofte for korte til å kode for et protein. Når vi finner lengre ORF-er vil det være usannsynlig ut ifra en statistisk fordeling med 3 stoppkodon for hver 64. triplett. Det kan forklares med at genene som finnes i DNA er et resultat av en biologisk utvikling og vil ikke være tilfeldige sekvenser, men et resultat av evolusjon og selektivt press. Når en leter etter gener i kromosom er en av metodene å finne lange ORF-er, noe som øker sjansen for at det er et proteinkodende gen (20).

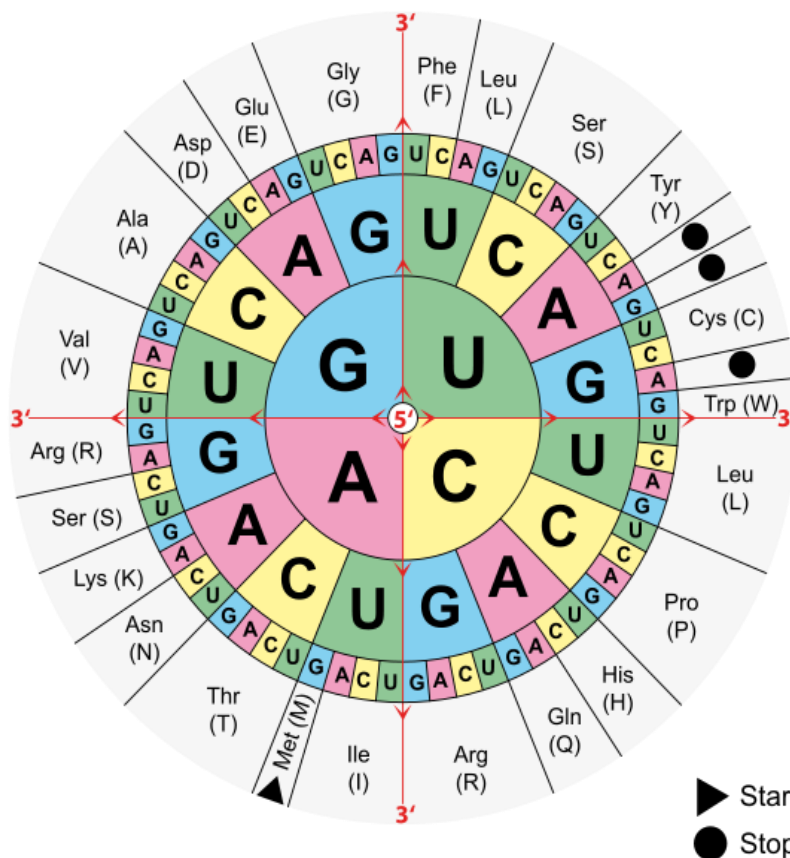
En leseramme referer til en av tre mulige måter en nukleotid sekvens blir lest på og påvirker hvilket protein som lages (32). Siden en positiv sense tråd alltid vil ha en komplementær antisense tråd, vil det være 3 leserammer på sense tråden og 3 på antisense tråden som leses i revers retning. Til

sammen 6 mulige leserammer. Siden sekvensen NCBI oppgir er i sense retning, ser vi kun på leserammer på denne tråden (Figur 10).

Leseramme:	123
	acttaccgga
Første leseramme:	T Y P G L
Andre leseramme:	L T R D
Tredje leseramme:	L P G T

Figur 10: Inndeling av en basesekvens i tre forskjellige leserammer. Store bokstaver representerer aminosyrene som er kodet av de tre bokstavene over og til venstre for dem.

Vi kan starte translasjonen fra den første bokstaven C, som blir den første leserammen. Eller vi kan starte fra den andre bokstaven T som gir den andre leserammen, eller den tredje bokstaven C som gir den tredje leserammen. Når sekvensen er bestemt kan man finne aminosyrene kodonene koder for (Figur 11) og samle alle aminosyreforkortelsene til en peptidsekvens i tekstformat.



Figur 11: Hjul som viser hvilken aminosyre baserekkefølgen i mRNA-templatet oversettes til i translasjonen. Hjulet viser hvilke tre nukleotider som koder for hvilken aminosyre ved å starte fra de fire innerste bokstavene og jobbe seg ut til den ytterste ringen. T er erstattet med U i en mRNA sekvens. (33).

Fylogeni og Homologisøk

Gener som har en beslektet funksjon eller er i samme familie kan ha en viss grad av homologi som nevnt. Et homologisøk vil si å søke i databaser etter bestemte DNA- eller aminosyresekvenser som har stor homologi, f.eks. 70%, eller er helt like, dvs. 100% like i sekvensene. Slike homologisøk kan gjøres for hele gener, eller for kortere sekvenser.

BLAST (Basic Local Alignment Search Tool) er en samling av dataprogram i NCBI som fungerer som et søkeprogram for DNA- og proteinsekvenser. Det finnes flere ulike måter (formater) å fremstille en DNA-sekvens på. På NCBI ligger GenBank og FASTA inne som standard formater, og i tillegg finnes EMBL. GenBank er en database over alle kjente DNA-sekvenser, og den blir oppdatert av forskere fra hele verden. DNA sekvenser på FASTA format blir oppgitt uten mellomrom eller nummerering. GenBank og EMBL deler opp (grupperer) sekvensen i 10 og 10 nukleotider skrevet med små bokstaver, med nukleotid nummerering for hver linje. De øverste linjene inneholder opplysninger om DNA-sekvensen, skrevet på en standardisert måte som gjør at også dataprogram kan lese dem. Dette gjør det mulig å kunne oppgi en bestemt sekvens, og programmet søker i databasene etter sekvenser som er homologe med denne. RefSeq databasen inneholder en godt dokumentert og annotert sekvens for alle eksisterende biomolekyler, slik som DNA. Referansesekvensen for ett bestemt gen vil representere en bestemt variant der det finnes flere varianter/alleler, hvor den best dokumenterte varianten benyttes som en standard sekvens.

Ved å sette inn parameteren med ønsket organisme i søkefeltet kan man sammenligne både mellom samme og andre arter. BLAST tilbyr flere muligheter for å gjøre homologisøk (Tabell 1).

Tabell 1: Oversikt over ulike BLAST-programmer som kan benyttes til homologisøk. For å velge riktig program leses tabellen fra høyre til venstre. Først må en se på resultatet som en ønsker at søket skal gi, deretter kan en velge program ut ifra hvilke sekvenser en har.

Oppgis til BLAST:	BLAST finner DNA- eller aminosyresekvens:	Programmets navn:	Resultat:
DNA	DNA	Blastn	Homologe gener
Protein	Protein	Blastp	Homologe proteiner
DNA	Protein	Blastx	Homologe proteiner med hensyn til 6 RF (leserammer)
Protein	DNA	tblastn	Homologe DNA sekvenser med hensyn til 6 RF + wobble*
DNA	DNA	tblastx	Homologe DNA sekvenser med hensyn til 6 RF og økt antall resultater ved å gå via protein.

*Wobble eller mer korrekt; et wobble base par vil si at to nukleotider i RNA ikke følger Watson og Cricks baseparings regler og gjør at hver aminosyre normalt tilsvarer 2 eller 4 mulige baser. Vi vil derved få svært mange resultater fra en oppgitt DNA-sekvens.

I den siste varianten (tblastx) oppgir du en DNA sekvens som blir oversatt til en eller flere aminosyresekvenser for alle 6 leserammene. Disse 6 proteinene/aminosyresekvensene vil bli omvendt translert til DNA. På grunn av wobble vil hver aminosyre normalt tilsvare 2 eller 4 mulige baser, som gjør at vi kan få svært mange resultater fra en oppgitt DNA-sekvens. Da kan vi finne DNA-sekvenser som kanskje har fjernt slektskap fordi de koder for samme aminosyresekvens (20). Slike homologisøk kan være nyttig for å undersøke slektskap mellom gener, men også for å finne gener. Hvis vi har et gitt gen kan vi finne ut hvor mange forskjeller det er mellom dette genet i ulike arter, ved å legge sammenstille sekvensene under hverandre og telle ulikhetene.

Dataanalyser har i senere tid gjort det mulig å sammenligne homologi mellom sekvenser fra mange ulike arter og på grunnlag av dette bygge et fylogenetisk stamtre for disse artene. Det er mange forskjellige metoder å lage ett fylogenetisk tre på, kort sagt er maximum likelihood og Bayesian metoder de to mest robuste og vanligste metodene (34). Neighbour joining, UPGMA, Maksimal parsimony og minimum evolusjon er metoder saom benyttes i mindre grad.

Maksimal sannsynlighet (maximum likelihood) og Bayesiansk metoder tar hensyn til mange faktorer i sekvensevolusjon slik som homoplasi, konvergent evolusjon og revers evolusjon, i motsetning til de andre nevnte metodene. Dette gjør de ideelle for å bygge et fylogenetisk tre for å kartlegge

viktigheten av enkelte gener med homoplasi i ulike arter. For å teste treet sin støtte til grupperinger, kan det opprinnelige datasettet «bootstrappes» for å lage et nytt datasett. Vanligvis blir antall bootstrap-replikasjoner satt til 500. Bootstrapping skjer ved gjentatte tilfeldige uttrekk og tilbakelegging av data, slik at det oppnås mer robuste estimater. I et fylogenetisk tre betyr dette at de sammenstilte sekvensene treet bygger på, blir plassert på 500 forskjellige måter av programmet. Hvis sekvensene grupperes på samme måte i 70% av tilfellene eller over, anses det som en god støtte, men noen anser så lavt som 50% som sannsynligvis signifikant. Programmet har en nedre grense på 50% som gjør at de grenene som har mindre enn 50% støtte automatisk blir automatisk kollapset. Desto mer informasjon som inkluderes i den fylogenetiske analysen, det vil si flere loci eller et lengre fragment av det samme genet, jo høyere støtte er det mulig å oppnå fordi det begrenser antall grupper sekvensen kan bli plassert i.

Primerdesign

Det er laget flere gratis dataprogrammer som kan designe oligonukleotider (primere) ut ifra en referansesekvens hentet fra for eksempel Ensembl eller GenBank. Spesifiserte krav som en kan stille er at primerne må ha klart definerte smeltepunkt, hybridisere til DNA-sekvensen med en viss avstand fra hverandre, og være spesifikke for målsekvensen. Vi har valgt å benytte NCBI sitt program kalt primer-BLAST til dette. Ved å legge inn DNA-sekvensen fra genet en ønsker å amplifisere og et område på hver side (flankerende DNA) kan dataprogrammet komme med ett eller flere forslag til primer-sett som kan hybridisere til flankerende DNA på hver side av mål-sekvensen.

Når programmet skal velge ut primerne som skal amplifisere sense-tråd/forward-strand benyttes den kodende trådens baserekkefølge, siden den alltid vil være komplementær til templat-tråden. Tilsvarende vil den velge ut primere til antisense-tråden/reverse-strand ved å benytte templat trådens baserekkefølge. Primer I og II i primer-paret skal hybridisere til hver sin tråd, slik at deres 3'-ender peker mot mål-DNA. Lengden av PCR-produktet (unntaksvis long-range) skal være 200-450 bp.

I denne oppgaven benytter vi ikke allele spesifikke oligonukleotider (ASO-primere), men isteden genspesifikke oligonukleotider. En genspesifikk oligo skal hybridisere til ett bestemt gen, uansett hvilken genotype (allel) det er. Den bør da ligge i et område av genet der det ikke er kjent noen polymorfismer. En polymorfisme er en generell endring (mutasjon) i DNA-sekvensen som er enten benign eller sykdomsgivende.

Betegnelsen «snipp», eller single nucleotidepolymorphism (SNP) benyttes dersom polymorfismen kun omfatter en base, eksempelvis ved en enkel punktmutasjon. I et ekson kan en SNP i et kodon innebære en endring i aminosyren den opprinnelig koder for. Noen ganger kan en slik variant være skadelig. Det er derfor mindre tetthet av SNP-er i kodende DNA enn ikke-kodende. Data fra det humane genomprosjektet har gjort at forskere har kunnet sammenlikne DNA-sekvenser fra ulike individer for å finne SNP-er, og alle rapporterte SNP-er fra mennesker og andre arter lagres i databaser slik som databasen dbSNP Short Variants. dbSNP gir informasjon om hvilken populasjon/befolkningsgruppe som er undersøkt og hvor mange individer som ble undersøkt. Basert på disse tallene oppgis også frekvensen for de ulike allelene m.m. (20). En annen database er GnomAD som er den mest brukte og den databasen vi i hovedsak benyttet oss av i denne oppgaven.

Alamut er et program som henter opplysninger fra Swiss Prot variants (dbSNP Short Variants) og Genome Aggregation Database (gnomAD) som samler alle rapporterte genvarianter. Vi benytter oss av Alamut for å sjekke primerens bindingssete opp mot databaser med polymorfismer og velge ut de primerne hvor allelefrekvensen til polymorfierne er lave, slik at det er minst mulig sjanse for at primerne ikke skal kunne binde seg optimalt.

Når primer-kandidater er valgt kan vi sjekke spesifisiteten til primerne ved å legge inn primersekvensen i primer-BLAST. Når vi skal legge inn antisense-primeren vil ikke templat-trådens baserekkefølge være oppgitt som en DNA-sekvens av genet. Vi må manuelt lage en komplementær tråd til den kodende sense-tråden (5' til 3' retning), dvs. en templat tråd (3' til 5' retning) som speilvendes slik at den går fra 5' til 3' retning.

Som kjent skal lengden på en primer ligge i området 15-25 nt. En kort primer vil ha større risiko for å være uspesifikk. Generelt er det slik at sterkere binding mellom to DNA-tråder gir høyere T_m (smeltepunkt). Lengden og GC-prosentsen til primeren, i tillegg til saltkonsentrasjonen i reaksjonen og mulige mismatch, bestemmer smeltetemperaturen til primeren. Hvis ett eller flere nukleotider i primerens bindingssete ikke stemmer med primer-sekvensen vil dette svekke bindingen mellom trådene, altså gi lavere smeltepunkt. Kationkonsentrasjon har betydning for hydrogenbindingene, og en høyere saltkonsentrasjon vil ofte gi lavere smeltepunkt. Smeltepunktet blir høyere med økt innhold av GC. Dette skyldes at basene G og C danner 3 hydrogenbindinger, mot A og T som danner 2. Et oligonukleotid med økt lengde vil gi høyere smeltepunktet grunnet flere H-bindinger som totalt kan binde seg til DNA. For at en primer skal hybridiseres til sitt templat, må temperaturen være under smeltepunktet. Primerpar bør ha smeltepunkt med en maksimal forskjell på 5 °C (20).

Reaksjonsløsningen som det er valgt å bruke har en primerkonsentrasjon på 300 nM og saltkonsentrasjon på 50 mM. Smeltepunktet kan estimeres ut ifra mange ulike formler.

ThermoFisher Scientific har utviklet en kalkulator (35) som regner ut optimal hybridiserings temperatur for PCR med forskjellige polymeraser. Kalkuleringen baserer seg på termodynamiske data fra Santa Lucia (36), hvor hybridiserings temperaturen varierer ut ifra hvilken polymerase som benyttes. ThermoFischer sin T_m kalkulator baserer seg derfor på empiriske observasjoner av effektivitet for hver av polymerasene for å gi en mest mulig korrekt T_m verdi.

Primer-BLAST har også en innebygd T_m kalkulator, men tar ikke hensyn til bufferforholdene i amplifiseringsreaksjonen basert på polymerasen som skal brukes. Primer-BLAST baserer seg isteden kun på saltkonsentrasjon satt til 1M (37).

NEW ENGLAND BioLabs inc. (NEB) har også utviklet en kalkulator (NEB T_m calculator) som regner ut optimal hybridiserings temperatur for PCR med forskjellige polymeraser (38). Kalkuleringen baserer seg på både termodynamiske data fra Santa Lucia (39), men også Owczarzy sine funn for hvordan saltkonsentrasjon påvirker hybridiserings temperatur (40). Denne gir en korrigeret formel:

Owczarzy et al. T_m salt correction:

$$\frac{1}{T_m(2)} = \frac{1}{T_m(1)} + (4,29f(G \cdot C) - 3,95) \cdot 10^{-5} \ln \left[\frac{[Na^+]_2}{[Na^+]_1} \right] + 9,40 \cdot 10^{-6} \left(\ln^2 [Na^+]_2 - \ln^2 [Na^+]_1 \right)$$

Dette gir en enda bedre beregning av T_m, men NEB sin T_m kalkulator gir kun mulighet til å velge NEB sine egne polymeraser.

Formelen over korrigerer for reaksjonens bufferforhold hvor f_{GC} er GC fraksjonen. NEB har omgjort den til en formel som ser på den monovalente kation konsentrasjon kalt m.

$$T_m(\text{corrected}) = \frac{1}{\frac{1}{T_m} + \left[(4,29 \cdot f_{gc} - 3,95) \cdot \ln(m) + 0,94 \cdot (\ln(m))^2 \right] \cdot 10^{-5}}$$

Den korrigerete formelen over er universell og kan benyttes utenfor NEB sin kalkulator, men forutsetter en allerede beregnet ukorrigeret T_m verdi.

Fidelity er et engelsk ord for gjengivelse, og brukes om DNA polymerase sin evne til ikke å gjøre feil. Produsenter av varrestabile DNA polymeraser konkurrerer om å ha størst mulig pålitelighet (high fidelity). Selv om de fleste fortsatt benytter Taq polymeraser, tas det også i bruk enzymer fra andre

bakterier i ulike naturlige varme kilder som blir forbedret ved hjelp genteknologi. Disse polymerasene blir gitt egne navn som «Easy-A», og «Platinum» osv.

På laboratoriet benytter vi Platinum *Taq* DNA Polymerase High Fidelity fra Invitrogen (41). Dette er en *Taq*-basert polymerase og i ThermoFisher sin T_m kalkulator legges dette inn sammen med primer konsentrasjon og oligosekvensene til primerparet. *Taq* polymeraser er avhengig av Mg^{2+} -ioner, og det er viktig å tilpasse konsentrasjonen av disse ionene til hver enkelt PCR-reaksjon (20). T_m verdien som beregnes av ThermoFisher sin kalkulator er ikke justert til forskjellige bufferforhold i PCR, men en antatt konstant monovalent kation konsentrasjon. Ved andre ionekonsentrasjoner vil den bli unøyaktig og må korrigeres med NEB sin korrigerende formel.

Når en primer er laget og bindes komplimentert til DNA i PCR reaksjonen kan DNA syntesen begynne, og dette katalyseres av polymerasen. Den optimale arbeidstemperaturen for Platinum *Taq* DNA polymerase High Fidelity er 68 °C ved polymerisering av lange PCR-produkt som ved long-range PCR. DNA-syntesen foregår da fra den frie 3'-OH-enden på begge primere. Byggesteinene er trifosfater: dATP, dTTP, dGTP og dCTP, samlet kalt dNTP. Denne forlengelsesreaksjonen er i prinsippet lik RNA-syntesen. En ny nukleotid henges på ved at det dannes en fosfoesterbinding mellom α -fosfat i en ny dNTP og den fri 3'-OH i den eksisterende tråden. Templattråden bestemmer hvilken base som skal henges på. Forlengelsen av den nye DNA-tråden skjer altså i retningen 5'→3', og vi sier at platinum *Taq* DNA polymerasen har 5'→3'-polymeraseaktivitet.

Vi ønsker å kontrollere at PCR betingelsene er optimale for å unngå feilkilder. Manglende PCR-produkt kan skyldes feil temperatur, feil konsentrasjon av Mg^{2+} , primer-primer-hybridisering m.m. Dette vil si en falsk negativ og at primeren ikke nødvendigvis passer noe sted i templat-DNA. Hvis en ikke hadde utført «specificity check» i BLAST kan en ende opp med å velge primere som ikke er spesifikk for ett bestemt DNA-område, og primeren vil kunne hybridiseres flere steder i genomet og gi flere PCR-produkt. For lav hybridiseringstemperatur kan gi samme resultat, hvor man tenker at en har fått det PCR-produkt en er ute etter, mens det egentlig er noe annet. Det vil si falskt positive (20).

Materiale og metode

Primerdesign

I NCBI sin database fant vi den genomiske DNA-sekvensen til *PRSS1*-genet under RefSeqGene (NG) NG_008307.3 som samsvarer med genomreferansesamlingen GRCh38 (hg 38). Transkriptet som vi forholdt oss til for å rapportere varianter var RefSeq transcript (NM) NM_002769.4.

I NCBI ligger det et dataprogram kalt Primer-BLAST. Primer-BLAST ble benyttet for å finne ett eller flere forslag til primer-par som kan hybridisere til flankerende DNA på hver side av mål-sekvensen. Dette ble gjort for long-range primerne hvor mål-sekvensen var hele *PRSS1*-genet, og for de sekvenseringsprimerne hvor mål-sekvensen var eksonene.

Primerne (Tabell Supplement 2) ble kontrollert til å være utformet i 5'til 3' retning. 3'-enden av primeren ble kontrollert til å stemme overens med DNA-malen siden polymeriseringstrinnet avhenger av en god match på 3'-enden.

Bare primere på mellom 18 til 25 nt i lengde ble valgt (med unntak av 1-A og 2A-S som var på 17 nt). Primerne hadde et GC-innhold på mellom 40 og 60%, og der det var mulig med C eller G som avsluttende nukleotid på 3'-enden av primeren, for å fremme binding. Når det var aktuelt ble primere med GC-klemmer valgt, dvs. de siste 5 basene på 3'-enden inneholdt minst 2 G- eller C-baser (GC-klemme), for sterkere binding.

Primernes spesifisitet ble kontrollert gjennom NCBI Primer BLAST som har Primer Pair Specificity Checking hvor vi velger *Homo sapiens* som organisme. Dette gjøres for å være sikker på at de ikke vil binde seg til andre genomiske regioner enn *PRSS1*. Det ble samtidig kontrollert at primerparene ikke hadde inter-primer homologi hvor sense- og antisense-primer hadde komplementære sekvenser og dermed kan danne «hårnåler» eller primer-dimere.

Vi benyttet oss av dataprogrammet Alamut visual (interactive biosoftware) for å sjekke primernes bindingssete opp mot databaser med polymorfismer og DNA-varianter og velge ut de primerne hvor allelefrekvensen til variantene var lave, slik at det er størst mulig sjans for at primerne skal binde seg optimalt uten mismatch som vil svekke bindingen. DNA-varianter med en frekvens på $> 0,01\%$ ble forkastet. Primere ble godtatt dersom DNA-variantene hadde en frekvens $\leq 0,1\%$ hos ikke-europeiske folkegrupper (Tabell Supplement 2).

Prøvemateriale

Det ble benyttet ferdig isolert og normalisert DNA for å etablere long-range PCR metoden for å sekvensere *PRSS1*-genet. DNA-et som ble brukt var fra 4 friske blodgivere ved St. Olavs med gitt samtykke. DNA prøvene var fra før anonymisert, og ble nummerert som Blodbank DNA (BB) 1-4 under oppsettet; FD47-20-->BB 1, FD48-20-->BB 2, FD49-20-->BB 3 og FD50-20-->BB 4.

Til reanalyse av tidligere pasientprøver fra NGS genpanel ble ferdig isolert og normalisert DNA fra 15 pasienter med kjent pankreatitt benyttet. Siden pasientene allerede er undersøkt for varianter i *PRSS1*-genet, er det ikke krav om nytt samtykke ved reanalyse. Pasientene er rangert etter anonyme remissenummer (Tabell 2).

Tabell 2: Oversikt over pasientprøver aktuelle for reanalyse. Alle pasientprøvene er tidligere analysert på NGS genpanel for utredning av arvelig pankreatitt.

Remisstype	Remissnr	Remisseår	Født	PC i familien	NGS panel
D	293	2020	1998	nei	95
D	264	2020	1968	nei	95
D	2181	2019	1963	nei	88d
D	2072	2019	1952	nei	87d
D	2031	2019	1975	nei	87c
D	1907	2019	1986	nei	85a
*D	1857	2019	1997	nei	84b
D	1801	2019	1983	nei	84b
D	1593	2019	1999	nei	80d
D	1345	2019	1984	ja	79c
D	1107	2019	1973	nei	77a
D	803	2019	1991	nei	74a
D	331	2019	1991	nei	69a
D	2534	2018	1979	nei	65a
D	2360	2018	1972	nei	64a
*normalt svar fra Blueprint					

Av de 15 pasientene er det kun en pasient med kjent pankreaskreft i familien; D1345-19. Pasienten med remisse ID; D1857-19 har tidligere blitt analysert på laboratoriet til Blueprint Genetics (42), uten klinisk relevante funn. Det er notert hvilket NGS panel pasientprøvene tidligere er blitt analysert på, for senere å kunne sammenligne med data fra Sanger-sekvensering.

I motsetning til oppsettet av long-range PCR hvor NTC ble sjekket ved gelelektroforese, må NTC kontroll være gjennomgående i hele oppsettet til pasientprøvene for å sjekke at resultat ikke skyldes kontaminasjon. Analysen ble ellers utført identisk med metoden som ble etablert med blodbankprøvene.

Amplifisering av *PRSS1*-genet

PRSS1-genet ble amplifisert med 4 forskjellige LR primerpar ved identiske PCR betingelser for å finne det primerparet som var mest spesifikt og ga sterkest intensitet.

Nødvendig volum av PCR reagensene Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), 10X High Fidelity PCR Buffer (pH 8,9) (Invitrogen), 50 mM MgSO₄ (Magnesiumsulfat), 10 mM GeneAmp dNTP miks (Applied Biosystems), ferdig løst opp sense- og antisense primere (Eurogentec) fortynnet til en bruksløsning på 10µM, og deionisert, nukleasefritt vann, ble beregnet ut ifra antall reaksjoner i oppsettet. Reagensene ble blandet sammen (Tabell 3). Blandingen ble alikvotert til 0,2 ml PCR rør med lokk.

Tabell 3: Oversikt over reagenser til long-range PCR for de fire ulike LR primerparene. Hvilke volum som ble brukt av reagenser og DNA per reaksjon, står oppført under «Volum µl n=1».

Reagenser til (1)PRSS1-LR løsning	Bruksløsning	Sluttkonsentrasjon	Volum µL n=1	Reagenser til (3)PRSS1-LR løsning	Bruksløsning	Sluttkonsentrasjon	Volum µL n=1
dH ₂ O	-	-	14,90	dH ₂ O	-	-	14,90
High Fidelity PCR buffer	10X*	1X	2,50	High Fidelity PCR buffer	10X*	1X	2,50
MgSO ₄	50 mM	2,0 mM	1,00	MgSO ₄	50 mM	2,0 mM	1,00
dNTP miks	10 mM	400 µM totalt	0,50	dNTP miks	10 mM	400 µM totalt	0,50
(1)PRSS1-LR-S primer	10 µM	0,3 µM	0,50	(3)PRSS1-LR-S primer	10 µM	0,3 µM	0,50
(1)PRSS1-LR-A primer	10 µM	0,3 µM	0,50	(3)PRSS1-LR-A primer	10 µM	0,3 µM	0,50
Platinum Taq DNA polymerase High Fidelity	5 U/µL	0,5 U	0,10	Platinum Taq DNA polymerase High Fidelity	5 U/µL	0,5 U	0,10
Volum til fordeling: 20,00				Volum til fordeling: 20,00			
Prøver:				Prøver:			
Templat DNA 10 ng/µl			5 µl	Templat DNA 10 ng/µl			5 µl
NTC:			0 µl	NTC:			0 µl
Reagenser til (2)PRSS1-LR løsning				Reagenser til (4)PRSS1-LR løsning			
dH ₂ O	-	-	14,90	dH ₂ O	-	-	14,90
High Fidelity PCR buffer	10X*	1X	2,50	High Fidelity PCR buffer	10X*	1X	2,50
MgSO ₄	50 mM	2,0 mM	1,00	MgSO ₄	50 mM	2,0 mM	1,00
dNTP miks	10 mM	400 µM totalt	0,50	dNTP miks	10 mM	400 µM totalt	0,50
(2)PRSS1-LR-S primer	10 µM	0,3 µM	0,50	(4)PRSS1-LR-S primer	10 µM	0,3 µM	0,50
(2)PRSS1-LR-A primer	10 µM	0,3 µM	0,50	(4)PRSS1-LR-A primer	10 µM	0,3 µM	0,50
Platinum Taq DNA polymerase High Fidelity	5 U/µL	0,5 U	0,10	Platinum Taq DNA polymerase High Fidelity	5 U/µL	0,5 U	0,10
Volum til fordeling: 20,00				Volum til fordeling: 20,00			
Prøver:				Prøver:			
Templat DNA 10 ng/µl			5 µl	Templat DNA 10 ng/µl			5 µl
NTC:			0 µl	NTC:			0 µl

*600 mM TRIS-sulfat og 180 mM ammoniumsulfat

Tabell 3 viser konsentrasjon til stamløsningene sammen med sluttkonsentrasjon på reagensene brukt til PCR. DNA ble tilsatt til hvert av rørene slik at totalvolumet i alle PCR-rør var 25 µl. Til NTC kontroll til primerparene ble det ikke tilsatt DNA, da formålet er å sjekke kontaminasjon i primerne eller reaksjonsmiksen som kan påvirke analysen av prøvene. PCR-rørene ble blandet raskt på IKA MS3 basic og deretter sentrifugert ved 6000 rpm på Galaxy mini Microcentrifuge (VWR international) i noen få sekunder. Deretter ble de plassert i 2720 Thermal Cycler (Applied Biosystems).

Tabell 4: PCR-program innstillinger på 2720 Thermal Cycler brukt til oppsett med long-range primere.

Long-range PCR innstillinger for oppsettet			
Triinn		Temperatur (°C)	Tid
Innledende denaturering		94	05:00
38 sykler	Denaturering	94	00:20
	Hybridisering	61	00:25
	Polymerisering	68	04:30
Avsluttende polymerisering		68	07:00
Avkjøling		4	Ubegrenset

Tabell 4 viser innstillinger som ble benyttet på instrumentet. Hybridiseringstemperaturene til long-range primerparene for *PRSS1* kan ikke være høyere enn korrigert T_m verdi som ble beregnet til å ligge i temperaturområdet 56,5°C til 61,0°C (Tabell Supplement 4). Den ukorrigerede T_m verdien tar høyde for smeltetemperaturen til primer-mal DNA-komplekset for optimal hybridisering. Lavere temperatur gir større fare for at primerne bindes uspesifikt. Derfor ble hybridiseringstemperaturen satt til å være 61°C, som er den høyeste hybridiseringstemperaturen til primerne våre, med mulighet for å nedjusteres. Polymeriseringstemperaturen er satt til 68°C da dette er optimal arbeidstemperatur for Platinum *Taq* polymerasen.

Gelelektroforese

DNA Molecular Weight Marker VII (0,37-8,0 kbp) 0,25 µg/µl med Tris-HCl (10 mM) og EDTA (1 mM) (Roche Diagnostics) blir omtalt videre som størrelsesmarkør. Størrelsesmarkøren ble fortynnet 1:10 med sterilt, nukleasefritt vann og blandet på IKA MS1 Mini shaker og deretter sentrifugert raskt på Galaxy mini Microcentrifuge (VWR international) i noen få sekunder.

Det ble applisert 5 µl long-range PCR-produkt i brønnene til E-Gel General Purpose Agarose Gel 1,2% med etidiumbromid (Invitrogen). I samme gelen ble det applisert 10 µl fortynnet størrelsesmarkør til egne brønner. E-Gel iBase Power System (Invitrogen) ble benyttet som strømkilde til elektroforese av gel. Programmet som ble benyttet var Egel 0,8-2% i 40 minutter. GeneFlash gel doc (Syngene) ble benyttet for å visualisere fragmentene som bånd på gelen med UV lys etter 40 minutter. DNA produktets størrelse (Tabell 5) ble sammenlignet med størrelsesmarkørene.

Tabell 5: Oversikt over long-range primere og forventet produktlengde.

Hvilke long-range primere som ble benyttet:	Produktlengde:
(1)PRSS1_LR-S	3991
(1)PRSS1-LR-A	
(2)PRSS1-LR-S	3910
(2)PRSS1-LR-A	
(3)PRSS1-LR-S	4087
(3)PRSS1-LR-A	
(4)PRSS1-LR-S	4004
(4)PRSS1-LR-A	

Hvilke primere som ble benyttet til å amplifisere hele *PRSS1* og forventet produktlengde er gjengitt i Tabell 5.

Rensing av PCR-produkt

For å defosforylere dNTP og degradere ubrukne primere ble henholdsvis reagensene Recombinant Shrimp Alkaline Phosphatase og Eksonuclease I, tilhørende settet A'SAP PCR Clean-up (ArcticZymes) benyttet. Disse ble blandet og holdt på kjølebakk under oppsettet. PCR-produktet som ble igjen etter gelelektroforesen ble tilsatt 1 µl A'SAP miks per 2,5 µl PCR-produkt og blandet raskt på IKA MS1 Mini shaker og deretter sentrifugert i noen sekunder på Galaxy mini Microcentrifuge (VWR international). PCR-rørene ble plassert i 2720 Thermal Cycler (Applied Biosystems).

Tabell 6: Innstillinger benyttet for 2720 Thermal Cycler til rensing av PCR-produkt. Hensikten med de ulike innstillingene står til venstre i tabellen.

Hensikt:	Innstilling:
Defosforylere dNTP og degraderer ubrukne primere:	37°C i 5 minutter.
Inaktivere A'SAP enzymene:	80°C i 10 minutter.
Nedkjøling:	4°C - ubegrenset.

Tabell 6 viser PCR innstillingene som ble benyttet på 2720 Thermal Cycler til rensingen av PCR-produkt.

Nukleotidsekvensering

Til nukleotidsekvensering er det for hvert ekson en sense- og en antisense-primer. Det ble laget en sekvenseringsmiks for henholdsvis sense- og antisense-primeren i hvert sekvenseringsprimerpar vi designet (Tabell Supplement 2). Begge miksene inneholder deionisert, nukleasefritt vann og reagensene som finnes i BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Settet består av BigDye Terminator v1.1 v3.1 5X Sequencing Buffer (pH 9,0) bestående av TRIS-hydrogenklorid og magnesiumklorid, og BigDye Terminator v3.1 Cycle Sequencing Ready Reaction mix 100 som inneholder DNA polymerase, dNTP, fluorescensmerkede ddNTP og 2,5X Sequencing Buffer.

Ut ifra antall reaksjoner (1rx=1 pasientprøve) ble nødvendig volum utregnet for sekvenserings-miks (Tabell 7).

Tabell 7: Oversikt over sekvenseringsreagenser brukt til nukleotidsekvensering. Hvilke volum som ble brukt av reagenser per reaksjon står oppført under « μl 1rx».

Sekvenserings-reagens	S	AS
	μl 1rx	μl 1rx
dH ₂ O	3,8	3,8
5x Buffer	2,0	2,0
BigDye v3	1,0	1,0
Primer (S/AS)	0,2	0,2
Til fordeling	7,0 μl	7,0 μl
A'SAP prod.	3 μl	3 μl

Tabell 7 viser rekkefølgen reagensene ble tilsatt i. De ble deretter blandet raskt på IKA MS3 basic og sentrifugert raskt på Galaxy mini Microcentrifuge (VWR international) i noen få sekunder. Miksen inneholder fluoriserende ddNTP som er lysfølsomme, og ble behandlet som lyst-følsomt materiale ved å bli pakket i aluminiumsfolie og oppbevart mørkt frem til neste trinn.

Det ble alikvotert 7,0 μl S-miks og AS-miks i en 96 brønners mikroplate med sterilt gummilokk. Deretter ble 3,0 μl A'SAP rensed PCR-produktet tilsatt i de respektive brønnene i rekkefølgen som ble lagt inn i arbeidsarket til ABI3730. Brønnene hadde da et totalvolum på 10,0 μl hver. 96 brønners mikroplaten ble deretter plassert i 2720 Thermal Cycler (Applied Biosystems).

Tabell 8: PCR-program innstillinger på 2720 Thermal Cycler brukt til sekvenserings-PCR.

PCR innstillinger for sekvensering		
Trinn	Temperatur (°C)	Tid
Innledende denaturering	96	01:00
26 sykler	Denaturering	96
	Hybridisering	54
	Polymerisering	60
Avkjøling	4	Ubegrenset

Tabell 8 viser innstillingene som ble brukt til sekvenserings-PCR for polymerisering med fluoriserende ddNTP. Disse fungerer som en terminerende etter prinsippet i Frederick Sangers sekvenseringsmetode. Sekvenseringsproduktet ble rensset med BigDye XTerminator Purification Kit (Applied Biosystems) bestående av XTerminator-løsning som fjerner ubundne ddNTP som kan gi uønsket fluorescens under kapillærelektroforese og SAM-løsning som stabiliserer rensset sekvenseringsprodukt. Ut ifra antall reaksjoner (1rx=1 pasientprøve) ble nødvendig volum utregnet for brønnene med sekvenseringsprodukt (Tabell 9)

Tabell 9: Oversikt over BigDye XTerminator reagens. Hvilke volum som ble brukt av reagenser per reaksjon står oppført under «µl 1rx».

BIGDYE XTerminator reagens	S	AS
	µl 1rx	µl 1rx
SAM løsning	45	45
Xterminator	10	10
Til fordeling	55 µl	55 µl

Tabell 9 viser forholdet de to reagensene ble blandet i. Blandingen ble vortexet på IKA MS1 Mini shaker. Deretter ble 55 µl alikvotert til brønnene i mikroplaten under konstant bevegelse da XTerminator løsningen består av porøse kuler som synker raskt. I rutinen til dette laboratoriet blir de tomme brønnene tilsatt 55µl 1X ABI buffer for å unngå at kapillærene i ABI3730 blir stående å tørke slik at polymeren som er fylt i kapillærene størkner og skaper blokkering ved fremtidige oppsett.

Den fylte mikroplaten ble dekket med MicroAmp selvklebende film og plassert på MixMate PCR 96 (Eppendorf) der det ble mikset ved 2000 rpm i 30 minutter. Etter 30 minutter ble mikroplaten sentrifugert med Centrifuge 5810R (Eppendorf) ved 1000 rpm i 2 minutter. Rensingen med kulene har som hensikt å binde uønskede rester for deretter å sedimenteres slik at supernatanten en er

interessert i inneholder rensed produkt som kan benyttes til kapillærelektroforese på ABI3730. Etter rensing av sekvenseringsproduktet ble mikroplaten plassert i et automatisert kapillærelektroforese instrument med 48 kapillærer. Instrumentet som ble benyttet var ABI3730 DNA Analyzer (Applied Biosystems). Dette er et automatisert kapillærelektroforese instrument med 48 polymerfylte kapillærer som separer DNA-fragmenter med ulik lengde, ut ifra prinsippet til elektroforese. Instrumentet er utstyrt med laser som sendes gjennom kapillærene og treffer de fluoriserende endenukleotidene på hvert DNA fragment. Fluorescensen detekteres av et kamera som registrerer signalstyrken og bølgelengden. Signalene plasseres i rekkefølgen de ble avlest og utgjør rådataen. Instrument protokoll som ble benyttet var en konfigurert versjon av BDX_RapidSeq_POP7_6sek_DNAlongR som ble laget i forbindelse med utprøving av metoden, hvor navnet ikke ble endret selv om vi gikk ned til 4 sekunder innsugingstid på ABI3730, istedenfor 6 sekunder som navnet tilsier. Analyse protokoll som ble benyttet var 3730POP7_BDTv3_KB_BDXT. Protokollene justerer avlesningen etter innholdet i brønnene. POP7 henviser til separasjonsmatriksen som ble benyttet som var POP-7 Polymer (Applied Biosystems). Rådataene fra ABI3730 ble overført og visuelt vurdert i dataprogrammet Sequence Scanner 2 (43).

Resultatet fra nukleotidsekvenseringen ble overført til analyseprogrammet SEQUENCE PILOT Software For Genetic Analysis (JSI medical systems) (44) omtalt videre som SeqPilot. SeqPilot ekstraherer sekvenser fra rådataen ABI3730 produserer og sammenstiller de analyserte sekvensdataene mot en referansesekvens. Referansesekvensen vi brukte for *PRSS1*-genet var NM_002769.5.

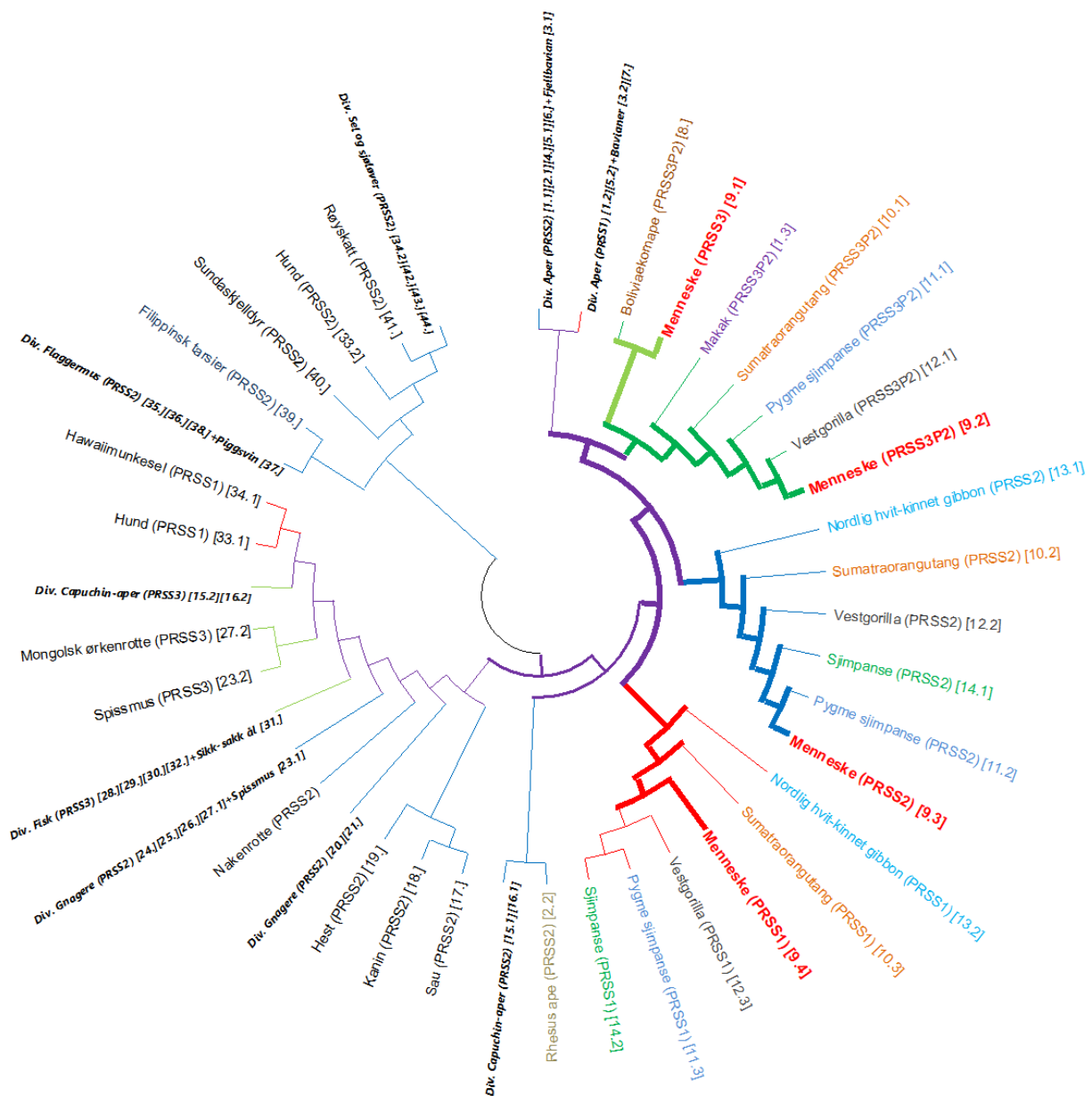
NGS data fra tidligere genpanelundersøkelser av pasientprøvene ble lastet opp som VCF-fil (Variant Calling File) i dataprogrammet FILTUS (OmicX) (45), som filtrerer varianter ut ifra gitte betingelser og databaser med dokumenterte varianter og gir ut en liste over varianter som må sjekkes og annoteres, og eventuelt tolkes. NGS produserer SAM-filer (Sequence Alignment Map) som inneholder sekvensdata, og filene blir videre binært formatert til BAM-filer (Binary Alignment Map). BAM-filen tilhørende hver pasientprøve ble lastet opp i Alamut visual (46) for å undersøke eventuelle varianter.

Resultater

Del 1: Fylogeni og homologisøk

I denne oppgaven ønsket vi å se på om noen sekvenser i genomet kan ha homologer med *PRSS1*-genet. Dette kan forklare hvorfor NGS sitt sammenstillingsverktøy ikke klarer å skille sekvensene fra hverandre. Referanse-nomenklaturen fra HUGO Gene Nomenclature Committee (HGNC) sine godkjente gen-symbol ble brukt til å søke i NCBI sine databaser (47). For *Homo sapiens serine protease 1* fant vi at dette er *PRSS1*. For å få en oversikt over *PRSS1*-genet ble NCBI brukt til å finne den genomiske DNA sekvensen til *PRSS1*-genet under RefSeqGene (NG) NG_008307.3 som samsvarer med genomreferansesamlingen GRCh Build 38 (Genome Reference Consortium Human Reference 38) fra 2013. GRCh Build 38 har ID hg38 i kontekst av UCSC Genome Browser og er en rettet og forbedret versjon av hg19 (GRCh37) fra 2009 som fortsatt er i bruk i diagnostikken. Transkriptet som vi forholdt oss til for å rapportere varianter er RefSeq transcript (NM) NM_002769.4. Dette transkriptet ble oppdatert til NM_002769.5 mens vi skrev oppgaven, og derfor benyttes ulike versjoner i løpet av oppgaven. Transkript NM_002769.4 venter på en godkjenning til å bli valgt til Locus Reference Genomic (LRG) som er en stabil referansesekvens spesielt designet for å rapportere sekvens varianter til klinisk bruk (48).

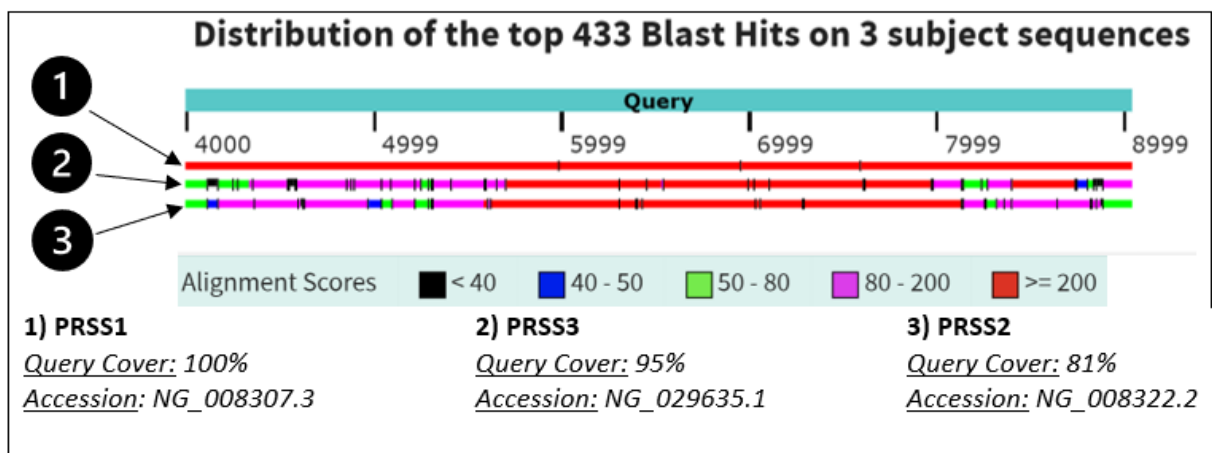
For å finne sekvenser med stor likhet til *PRSS1*-genet ble mRNA sekvensen fra transkript NM_002769.5 benyttet til å utføre et homologisøk i NCBI sitt dataprogram tblastx som tar hensyn til 6 leserammer og wobble. Databasen «Nucleotide collection (nr/nt)» ble benyttet til søket for å se de artene med signifikant alignment til *PRSS1*-genet. Signifikant alignment blir i denne sammenhengen omtalt som homologer. De artene med mellom 100% til 86% homologer til det menneskelige *PRSS1*-genet ble benyttet som grunnlag til et fylogenetisk tre (Figur 12) for å forstå viktigheten av fordøyelsesenzymer trypsin sett i et evolusjon perspektiv. Av artene som ble valgt ut ble det i tillegg funnet stor homologer for *PRSS2*-, *PRSS3*-, og *PRSS3P2*-genene. Søket viste nær 100% homologer med det menneskelige *PRSS2*-genet, som dermed ikke lot seg skille fra *PRSS1*-genet når kun den kodende sekvensen ble benyttet. Noen arter innenfor gruppen beinfisk hadde mellom 86% og 83% homologer for kun *PRSS3*-genet, og belyser et større evolusjonsmessig språk enn *PRSS1* og *PRSS2*. Det menneskelige *PRSS3*-genet ble derfor lagt inn manuelt. For å sammenstille den kodende nukleotidsekvensen til de valgte artene ble dataprogrammet MUSCLE (Multiple Sequence Comparison by Log-Expectation) benyttet (49). For å plassere artene etter likhet i genene som koder for trypsin-1 og trypsin-2, istedenfor evolusjonsmessig adskillelse som i et tids-tre, ble det konstruert et fylogenetisk tre (Figur 12) i visualiseringsprogrammet MEGA (Molecular Evolutionary Genetics Analysis) versjon 6.0 (50) fra sammenstillingen til MUSCLE.



Figur 12: Fylogenetisk tre som representerer genetisk likhet mellom PRSS1, PRSS2, PRSS3 og PRSS3P2 på tvers av ulike arter. Greinene er farget rødt for PRSS1, blått for PRSS2, lyse-grønt for PRSS3 og mørke-grønt for PRSS3P2. Lilla greiner indikerer overgangen mellom dem. Sekvenser fra Tabell Supplement 1 ble innhentet fra tblastx søk i tillegg til PRSS3 og PRSS3P2 for Homo sapiens og sammenstilt med MUSCLE. Evolusjonære analyser ble utført i MEGA X (50) og det fylogenetiske treet ble konstruert med maksimal sannsynlighet-metoden (maximum likelihood method) og Tamura-Nei-modellen (51) og testet med Bootstrap. For å representere den evolusjonære historien til taxa analysert ble Bootstrap konsensus-treet utledet fra 500 replikasjoner (52). Grener som er gjengitt i mindre enn 50% av bootstrap replikasjonene ble kollapset. Første treet for heuristisk søk ble innhentet automatisk ved å bruke Neighbor-Join- og BioNJ-algoritmer på en matrise med parvise avstander estimert ved hjelp av Tamura-Nei-modellen, og deretter ble topologien med superior log sannsynlighetsverdi valgt. Denne analysen involverte 66 nukleotidsekvenser. Kodon-stillinger inkludert var 1+2+3+Noncoding. Det var totalt 1064 stillinger i det endelige datasettet.

Som Figur 12 visualiserer er det en stor homologi mellom de trypsinogen-kodende genene innad i arter, men også homoplasi på tvers av ulike arter. Figur 12 tydeliggjør noen åpenbare slektskap som det mellom menneske, vestgorilla og pygme sjimpanse, som viste størst genetisk slektskap hva gjelder *PRSS1*-, *PRSS2*- og *PRSS3P2*-genene, som et resultat av kort tid siden divergens. Noe overraskende viste treet at arter som er adskilt i evolusjonen, for eksempel hest og kanin, deler flere av de samme genetiske variantene i *PRSS2*-genet. Det at to ganske så ulike arter utvikler identiske varianter innenfor et spesifikt gen vitner om parallell evolusjon og et sterkt selektivt press. Dette understreker viktigheten til fordøyelsesenzymet trypsin i utvalgte pattedyr og beinfisk.

For å undersøke nøyaktig hvor stor homologi det er mellom *PRSS1*- og *PRSS2*-genene ble det utført et nytt automatisk homologisøk med dataprogrammet tblastx. Dette ble gjort ved å legge inn sekvensen i FASTA-format fra NG_008307.3 i søkemotoren til tblastx med «Query subrange» fra 4000 til 9000 baser for å dekke den kodende sekvensen med introner og for å spare tid. Programmet sammenstiller sekvensene under hverandre og teller ulikhetene. Databasen som ble benyttet til søket var Human RefSeqGene sequences (RefSeq_Gene). Det resulterte i 3 treff på over 80% signifikant sekvenslikhet (Figur 13) i det valgte området.



Figur 13: Resultat fra homologisøk som viser de humane trypsinogen-kodende genene og deres homologi.

Som illustrert i Figur 13 har *PRSS2*- og *PRSS3*-genene et klart slektskap med *PRSS1*, som stemmer med homologisøk på nukleotid-nivå, og proteinnivå for *PRSS1* og *PRSS2* fra litteraturen (15), men vi ser ikke slektskapet mellom *PRSS1* og *PRSS3P2*. *PRSS3*-genet ligger på kromosom 9 og er ikke relevant for oppgaven.

Siden vi ikke får opp *PRSS3P2* i homologisøket har vi valgt å utføre et manuelt homologisøk på peptidsekvens mellom *PRSS1*, *PRSS3P2* og *PRSS2* for å undersøke likheten mellom dem.

For å finne en tenkt aminosyresekvens fra en ikke-kodende DNA sekvens fra et pseudogen, ble NCBI brukt til å finne DNA-sekvensen til *PRSS3P2* (Figur 14).

Homo sapiens PRSS3 pseudogene 2 (PRSS3P2), non-coding RNA

NCBI Reference Sequence: NR_001296.3

>NR_001296.3 Homo sapiens PRSS3 pseudogene 2 (PRSS3P2), non-coding RNA

```
CTCACCTCACAGACTCCTCTCTGGATCCTCCAGAGCTATAAAGACGGGCCTCCACCACCAGACAGGC
GCACTCTACCACCATGAATCCACTCCTGATCCTTGCCCTTTGTGGGAGCTGCTGTTGCTGTCCCTTTGAT
GATGATGACAAGATCGTTGGGGGCTACACCTGTGAGGAGAATTCTGTCCCTACCAGGTGCCTTGAATT
CTGGCTCCCACTTCTGCGGTGGCTCCCTCATCAGCGAACAGTGGGTGGTGTGTCAGCAGGTCCTGCTACAA
GCCCCACATCCAGGTGAGACTGGGAGAGCACAATATCGAAGTCCTGGAGGGGAATGAGCAGTTCATCAAT
GCAGCCAAGATCATCCGCCACCCCAAATACAACAGGATTATTCTGAACAATGACATCATGCTGATCAAGC
TCTCCACACCTGCCGTCATCAATGCCATGTGTCCACCATCTCTGCCCCTGCCCCTCCAGCTGCTGG
CACCGAGTGCCTTATCTCCGGCTGGGGCAATACCTGAGCTCTGGGGCCGACTACCCAGATGAGCTGCAG
TGCCTGGACGCTCCTGTGCTGACCCAGGCTAAGTGTAAGCCTCCTACCCTTTAAAGATTACCAGCAACA
TGTTCTGTGTGGCTTCTTGAGGGAGGCAAGGATTCTGCCAGGGTGACTCTGTGGCCCTGTGGTCTG
CAATGGACAGTTC AAGGAATTGTCTCCTGGGGCTATGCTGTGCCAGAAAGAGAAGGCCTGGAGTCTAC
ACCAAGTCTACA AACTATGTGGACTGGATTAAGGACACCATAGCTGCCAACAGCTAAAGCCCCTGGTCAC
TCTGCAGTCTCTATACCAATAAA
```

Figur 14: Referansesekvensen til PRSS3P2 hentet fra NCBI. Det er satt 2 understreker under henholdsvis ATG og TAA som ikke er merket av i sekvensen hentet fra NCBI.

For å finne starten på den åpne leserammen ble en del av sekvensen fra Figur 14 delt i tre og tre nukleotider (kodon) (Figur 15).

Leseramme +1: CTC ACC TCA CAG ACA CTC CTC TCT GGA TCC TCC AGA GCT ATA AAG ACG GGC CTT CCA
CCA CCA GAC AGG CGC ACT CTA CCA CCA **TGA**

Leseramme +2: C TCA CCT CAC AGA CAC TCC TCT CTG GAT CCT CCA GAG CTA **TAA** AGA CGG GCC TTC CAC
CAC CAG ACA GGC GCA CTC TAC CAC CAT GA

Leseramme +3: CT CAC CTC ACA GAC ACT CCT CTC TGG ATC CTC CAG AGC TAT AAA GAC GGG CCT TCC
ACC ACC AGA CAG GCG CAC TCT ACC ACC **ATG** A

Figur 15: Tre mulige leserammer for sense-tråden til PRSS3P2-genet. Den åpne leserammen starter med ATG som vist i leseramme +3.

Som illustrert i Figur 15 starter de to første leserammene med ett stoppkodon og vil ikke gi en åpen leseramme. Leseramme +3 ser vi starter med et startkodon og lengre ut i sekvensen finner vi også et stoppkodon som utgjør den åpne leserammen. Ved å translaterer alle kodon mellom start og stopp i leserammen +3 til aminosyrer, får vi en komplett peptidsekvensen for PRSS3P2 (Figur 16).

Basert på aminosyre tabellen over kan vi finne peptidsekvensen:

ATG AAT CCA CTC CTG ATC CTT GCC TTT GTG GGA GCT GCT
met (M) asn (N) pro (P) leu (L) leu (L) ile (I) leu (L) ala (A) phe (F) val (V) gly (G) ala (A) ala (A)

GTT GCT GTC CCC TTT GAT GAT GAT GAC AAG ATC GTT GGG GGC TAC ACC TGT GAG GAG AAT TCT GTC CCC TAC CAG
V A V P F D D D D K I V G G Y T C E E N S V P Y Q

GTG TCC TTG AAT TCT GGC TCC CAC TTC TGC GGT GGC TCC CTC ATC AGC GAA CAG TGG GTG GTG TCA GCA GGT CAC
V S L N S G S H F C G G S L I S E Q W V V S A G H

TGC TAC AAG CCC CAC ATC CAG GTG AGA CTG GGA GAG CAC AAT ATC GAA GTC CTG GAG GGG AAT GAG CAG TTC
C Y K P H I Q V R L G E H N I E V L E G N E Q F

ATC AAT GCA GCC AAG ATC ATC CGC CAC CCC AAA TAC AAC AGG ATT ATT CTG AAC AAT GAC ATC ATG CTG ATC AAG
I N A A K I I R H P K Y N R I I L N N D I M L I K

CTC TCC ACA CCT GCC GTC ATC AAT GCC CAT GTG TCC ACC ATC TCT CTG CCC ACT GCC CCT CCA GCT GCT GGC ACC
L S T P A V I N A H V S T I S L P T A P P A A G T

GAG TGC CTT ATC TCC GGC TGG GGC AAT ACC CTG AGC TCT GGG GCC GAC TAC CCA GAT GAG CTG CAG TGC CTG
E C L I S G W G N T L S S G A D Y P D E L Q C L

GAC GCT CCT GTG CTG ACC CAG GCT AAG TGT AAA GCC TCC TAC CCT TTA AAG ATT ACC AGC AAC ATG TTC TGT GTG
D A P V L T Q A K C K A S Y P L K I T S N M F C V

GGC TTC CTT GAG GGA GGC AAG GAT TCC TGC CAG GGT GAC TCT GGT GGC CCT GTG GTC TGC AAT GGA CAG CTT
G F L E G G K D S C Q G D S G G P V V C N G Q L

CAA GGA ATT GTC TCC TGG GGC TAT GGC TGT GCC CAG AAG AGA AGG CCT GGA GTC TAC ACC AAG GTC TAC AAC
Q G I V S W G Y G C A Q K R R P G V Y T K V Y N

TAT GTG GAC TGG ATT AAG GAC ACC ATA GCT GCC AAC AGC **TAA**
Y V D W I K D T I A A N S *

Figur 16: Åpen leseramme til PRSS3P2-genet fordelt som kodon med tilhørende ett-bokstavs aminosyre-forkortelse plassert under.

I NCBI finner vi peptidsekvensen til de kodende genene *PRSS1* og *PRSS2* ved å søke på de respektive genene. Peptidsekvensen til disse 3 genene ble lagt over hverandre eksonvis og ulikheter fra *PRSS1* ble uthevet i rødt (Figur 17). Transkriptene som ble benyttet har 5 Eksoner, men vi har valgt å ta med den kodende sekvensen til Ekson 3 fra Ensembl transkriptet (GENCODE transcript (ENST)) ENST00000486171.1 (GRCh38), som vi nummererte som Ekson 2A. Vi vil forholde oss til transkript NM_002769.4 for ekson-nummerering og primernavn, men fordi AMG ønsker å utføre analyser på Eksonet 2A, selv om det ikke er inkludert i LRG transkriptet, har vi valgt å ta med eksonet som er tilstede i transkript ENST00000486171.1. Vi har dermed eksonene: Ekson 1, Ekson 2, Ekson 2A, Ekson 3, Ekson 4 og Ekson 5.

ENST00000486171.1 er på 786 baser (6 eksoner), NM_002769.4 (5 eksoner) er på 744 baser. Ekson 2A utgjør nøyaktig 42 baser. Ekson 2 slutter på c.200 og Ekson 2A slutter på c.242. Vi vil ikke finne

ekson 2A i *PRSS3P2* da *PRSS3P2* er et transkribert pseudogen som ikke har introner hvor ekson 2A ligger i vårt transkript. Vi har valgt å inkludere homologi mellom ekson 2A i *PRSS1* og *PRSS2* (Figur 18).

Gen:	Transkript:	Exon 1:
<i>PRSS1</i>	NM_002769.4	MNPLLLI TFVAAAL
<i>PRSS3P2</i>	NR_001296.3	MNPLLLI AFVGA AV
<i>PRSS2</i>	NM_002770.3	MNPLLLI TFVAAAV
		Exon 2:
<i>PRSS1</i>	NM_002769.4	AAP FDDDDKIVGG YNCEENSVPY QVSLNSGYHF CGGSLINEQW VVSAGHCYKS
<i>PRSS3P2</i>	NR_001296.3	AVP FDDDDKIVGG YTCCEENSVPY QVSLNSGSHF CGGSLISEQW VVSAGHCYKP
<i>PRSS2</i>	NM_002770.3	AAP FDDDDKIVGG YICEENSVPY QVSLNSGYHF CGGSLISEQW VVSAGHCYKS
		Exon 3:
<i>PRSS1</i>	NM_002769.4	RIQVRLGEHN IEVLEGNEQF INAAKIIRHP QYDRKTLNND IMLIKLSSRA VINARVSTIS LPTAPPATGT KCLISGWGNT ASSGA
<i>PRSS3P2</i>	NR_001296.3	HIQVRLGEHN IEVLEGNEQF INAAKIIRHP KYNRILNND IMLIKLSTPA VINAHVSTIS LPTAPPAAGT ECLISGWGNT LSSGA
<i>PRSS2</i>	NM_002770.3	RIQVRLGEHN IEVLEGNEQF INAAKIIRHP KYNSRTLND ILLIKLSSPA VINSRVSAIS LPTAPPAAGT ESLISGWGNT LSSGA
		Exon 4:
<i>PRSS1</i>	NM_002769.4	DYPDE LQCLDAPVLS QAKCEASYPG KITSNMFCVG FLEGGKDCSQ
<i>PRSS3P2</i>	NR_001296.3	DYPDE LQCLDAPVLT QAKCKASYPL KITSNMFCVG FLEGGKDCSQ
<i>PRSS2</i>	NM_002770.3	DYPDE LQCLDAPVLS QA ECEASYPG KITNMFCVG FLEGGKDCSQ
		Exon 5:
<i>PRSS1</i>	NM_002769.4	GDSGGPVVNC GQLQGIVSWG DGCAQKNKPG VYTKVYNYVK WIKNTIAANS*
<i>PRSS3P2</i>	NR_001296.3	GDSGGPVVNC GQLQGI VSWG YGCAQKRRPG VYTKVYNYVD WIKDTIAANS*
<i>PRSS2</i>	NM_002770.3	GDSGGPVVSN GELQGI VSWG YGCAQKNRPG VYTKVYNYVD WIKDTIAANS*
		Peptidsekvens homologi med <i>PRSS1</i> -genet:
<i>PRSS3P2:</i>		88.7%
<i>PRSS2:</i>		89.5%
		Grunnlag:
<i>PRSS3P2:</i>		(1-[28/247])*100
<i>PRSS2:</i>		(1-[26/247])*100
		Formel:
<i>PRSS3P2:</i>		(1-[a/n])*100, der a=avvik fra <i>PRSS1</i> ,
<i>PRSS2:</i>		n=totalt antall aminosyrer

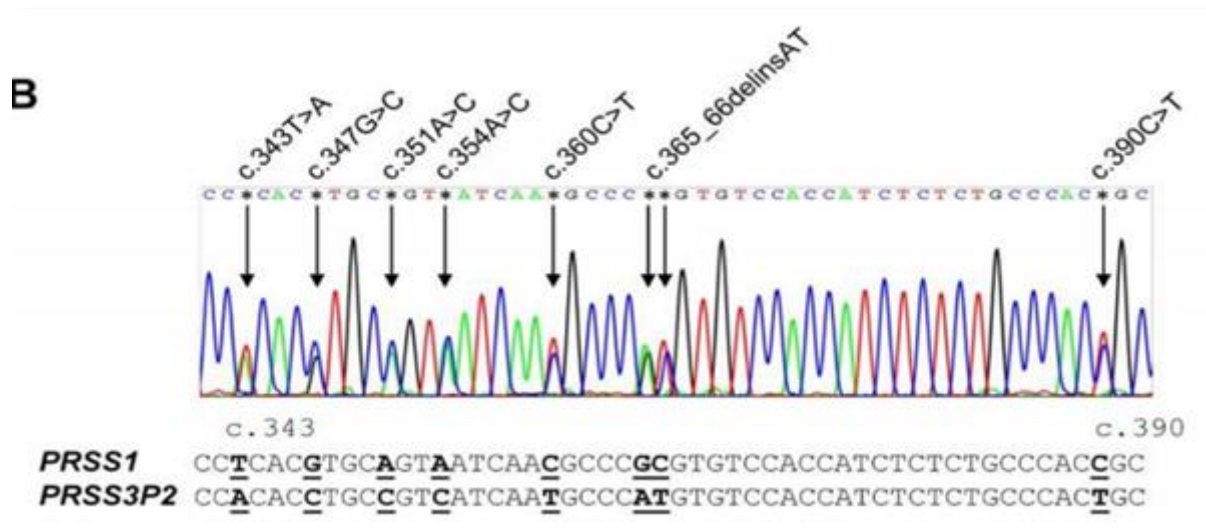
Figur 17: Sammenligning av peptidsekvensen mellom *PRSS1*-, *PRSS3P2*- og *PRSS2*-genet. Beregnet homologi i prosent (%) mellom *PRSS1* og henholdsvis *PRSS3P2* og *PRSS2* står nederst.

Gen:	Transkript:	Exon 2A:
<i>PRSS1</i>	ENST00000486171.1*	AIKSKFSGRGCEYH
<i>PRSS2</i>	NM_001303414.1*	AISKLSGRGCEY
*Denne varianten representerer det lengste transkriptet og koder for den lengste isoformen.		
		Peptidsekvens homologi med <i>PRSS1</i> -genet:
<i>PRSS2:</i>		89.3%
		Grunnlag:
<i>PRSS2:</i>		(1-[28/261])*100

Figur 18: Sammenligning av peptidsekvensen i ekson 2A mellom *PRSS1*- og *PRSS2*-genet. Beregnet homologi i prosent (%) mellom *PRSS1* og *PRSS2* for ekson 2A står nederst.

Figur 17 og Figur 18 viser en homologi mellom *PRSS1* og *PRSS3P2* på 89% som stemmer med tidligere funn (29) for homologi på aminosyrenivå. I litteraturen er det beskrevet at den høye homologien mellom *PRSS1*, *PRSS2*, og *PRSS3P2* kan gi problemer med tolkning av analysedata (4,29). Årsak til dette kan være genkonversjon, uspesifikke primere som bindes til *PRSS3P2* og/eller *PRSS2* ved Sanger-sekvensering, eller sekvenser som ikke blir sammenstilt til *PRSS1* ved NGS. Område på *PRSS3P2* som starter mellom c. 331 og c.342 og slutter mellom c. 391 og c.401, har ført til en

mulighet for mistolkning av SNP ved c.347G>C, c.354A>C, c.361G>T og c.365_66delinsAT i *PRSS1*-genet (Figur 19) (29).



Figur 19: Elektroferogram av ekson 3 i *PRSS1*-genet til en polsk index pasient. Heterozygote variasjoner i *PRSS1*-genet er indikert med piler. Sekvensen til *PRSS1* er sammenstilt med tilsvarende sekvens i *PRSS3P2* for å vise opphavet til variantene. Figuren er hentet fra Rygiel et al. (29).

Vi ønsket å sjekke om variantene i Figur 19 kunne utgjøre en mulighet for mistolkning hos AMG. Vi sammenliknet disse variantene med funn i databasene gnomAD og dbSNP som er de hyppigst brukte databasene som sjekkes i Alamut og som AMG bruker for å tolke sekvensvarianter påvist hos pasienter.

For å gjøre dette benyttet vi området i *PRSS1* som startet mellom c. 331 og c.342 og sluttet mellom c. 391 og c.401 som utgangspunkt for å vise problemet med pseudogenet *PRSS3P2* og tolkning av *PRSS1* (Figur 20).

5' TCA CGT GCA GTA ATC AAC GCC CGC GTG TCC ACC ATC TCT CTG CCC ACC 3'

Figur 20: Sekvens mellom c.343 og c.390 i *PRSS1*-genet som skal være identisk hos *PRSS3P2*-genet i det manuelle homologisøket.

En 100% baseidentisk sekvens er ikke mulig å finne i *PRSS3P2* (Protease, serin, 3 pseudogene 2) da pseudogener har en tendens til å akkumulere flere mutasjoner enn kodende sekvenser siden de sjeldent er under selektivt press. Vi benyttet derfor Alamut visual sin «Approximate matching» til å gjøre en lokal sammenstilling, og fikk opp en sekvens mellom n.426 og n.474 i Ekson 3 i *PRSS3P2*-genet (Figur 21).

5'....ACA CCT GCC GTC ATC AAT GCC CAT GTG TCC ACC ATC TCT CTG CCC ACT.....3'

Figur 21: Sekvens mellom n.426 og n.474 i Ekson 3 i *PRSS3P2*-genet med avvik fra *PRSS1*-genet fremhevet i rødt.

Figur 21 illustrerer at det er flere identiske baser enn forskjellige i den aktuelle sekvensen. Den samme fremgangsmåten ble benyttet for å illustrere homologien til *PRSS2* i aktuelt område (Figur 22).

5'....TCA CCT GCC GTC ATC AAT TCC CGC GTG TCC GCC ATC TCT CTG CCC ACT.....3'

Figur 22: Sekvens fra *PRSS2*-genet med avvik fra *PRSS1*-genet fremhevet i rødt.

I Alamut visual kan vi sammenligne dokumenterte varianter fra ulike databaser. Ved å se på registrerte SNPer i GnomAD og dbSNP for *PRSS1*-genet kan vi kartlegge de dokumenterte variantene i det aktuelle området (Figur 23).

5'....TCA CGT GCA GTA ATC AAC GCC CGC GTG TCC ACC ATC TCT CTG CCC ACC.....3'

Figur 23: Sekvens mellom c.343 og c. 390 i ekson 3 i *PRSS1*-genet med trolig benigne varianter fremhevet i grønt, trolig maligne varianter i rødt og de av usikker signifikans i gult.

Sekvensene fra Figur 21, Figur 22 og Figur 23 ble lagt over hverandre for å se om de hadde en sammenheng (Figur 24).

PRSS1: (Registrerte dbSNP er uthevet. Rødt=malign, Gul=usikker signifikans, Grønt=benign)

5'....TCA CGT GCA GTA ATC AAC GCC CGC GTG TCC ACC ATC TCT CTG CCC ACC.....3'

PRSS3P2: (Avvik fra PRSS1 er uthevet i rødt)

5'....ACA CCT GCC GTC ATC AAT GCC CAT GTG TCC ACC ATC TCT CTG CCC ACT.....3'

PRSS2: (Avvik fra PRSS1 er uthevet i rødt)

5'....TCA CCT GCC GTC ATC AAT TCC CGC GTG TCC GCC ATC TCT CTG CCC ACT.....3'

PRSS1 med uthevet SNPer lik PRSS3P2/PRSS2 (trolig malignt med gul skriftfarge):

5'....ACA CGT GCC GTC ATC AAT TCC CAT GTG TCC GCC ATC TCT CTG CCC ACT.....3'

Figur 24: Sammenligning av sekvens i *PRSS1*-genet og tilsvarende sekvenser i *PRSS3P2*- og *PRSS2*-genet. Registrerte SNPer i *PRSS1*-genet er fremhevet, og avvik fra *PRSS1*-genet er fremhevet i både *PRSS3P2* og *PRSS2*. Varianter i *PRSS1* som kan skyldes genkonversjon med *PRSS3P2* eller uspesifikke primere som bindes til *PRSS3P2* og/eller *PRSS2* er fremhevet nederst med rødt hvis varianten kan være sykdomsassosiert hvis den er reel.

Som Figur 24 illustrerer ser vi blant de registrerte SNPene at det er *to* maligne SNP'er ved c.347G>C og c.365_66delinsAT, og *en* med usikker signifikans ved c.354A>C, som kan stamme fra *PRSS3P2*. Fra *PRSS2* kan *en* malign SNP ved c.347G>C og *to* med usikker signifikans ved c.354A>C og c.361G>T, føre til mistolkning ved bruk av uspesifikk Sanger-sekvensering eller NGS. Disse fire SNPene er de samme SNPene vi finner igjen i litteraturen vist i Figur 19 som kan danne en mulighet for mistolkning også hos AMG, og vi anser problemet som reelt.

Del 2: Design av primere

For å forhindre en mulig mistolkning av analysedata fra NGS og vanlig Sanger-sekvensering av enkeltteksoner med trolig uspesifikke primere, ble det designet spesifikke long-range primere for å amplifisere hele den transkriberte *PRSS1*-sekvensen, i tillegg til sekvenseringsprimere til hvert ekson. For å gi en visuell oversikt over områder med spesiell betydning i den transkriberte sekvensen til *PRSS1*-genet ble GenBank sitt format benyttet, men med store bokstaver og uten nukleotid nummerering (Figur Supplement 1). Kodende sekvenser er ordnet i tripletter og fremhevet med farge for å vise målsekvensen for de indre sekvenseringsprimerne.

Design av primerne ble utført i primer-BLAST med «specificity check» på long-range primerne, og betydning av varianter i bindingssetet ble sjekket i Alamut (Tabell Supplement 2). Tabell Supplement 2 viser at alle primerparene som primer-BLAST anså som spesifikke, ikke hadde polymorfismer med høy allelefrequens i primerbindingssetet. Sjansen for at de vil svekke hybridiseringen ble ansett som liten nok til å tolereres. Merknad: I Tabell Supplement 2 benyttes nummerering til GRCh37 (hg19) selv om vi ser på referansesekvensen matchet mot GRCh38. Variantene som ble rapportert mot GRCh37 vil plasseres på riktig genom posisjon i GRCh38, men med annen nummerering.

En viktig del av oppgaven var å kunne oppdage genkonversjon som resulterer i rundt 25% av tilfellene med patogene sekvens varianter. Antisense-primeren til ekson 2 sitter mellom c.200+136 og c.200+153 i *PRSS1*. Ved en genkonversjon mellom *PRSS2* og *PRSS1* vil primeren ligge i området av *PRSS1* (c. 86 til c.200+174) som kan erstattes med tilsvarende område fra *PRSS2*. Metoden vår bygger på at vi spesifikt får amplifisert *PRSS1*-genet med long-range primerparene. På grunn av stor indre homologi i *PRSS1* til andre sekvenser i genomet, ble ikke sekvenseringsprimerne sjekket spesifisitet til med primer BLAST sin «specificity check». Ved genkonversjon kan det være varianter i intronet mellom ekson 2 (2A) og 3 til *PRSS2* som kan svekke bindingen til antisense-primeren. Derfor er vi helt avhengige av at sense-primeren bindes godt for å kunne oppdage genkonversjon og variantene det tilfører *PRSS1*-genet. Antisense-primer til ekson 2 vil kunne godkjennes til bruk i diagnostikken på lik linje som de andre sekvenseringsprimerne gjennom valideringen av metoden.

Del 3: PCR innstillinger

Følgebladet til Platinum polymerasen (41) ble benyttet for å finne kation konsentrasjon og konsentrasjonen til PCR reagensene til oppsettet (Tabell Supplement 3).

Vi ser fra Tabell Supplement 3 at kationet Mg^{2+} som dissosieres fra $MgSO_4$ i en vandig løsning bidrar som eneste kation i reaksjonsblandingen. Kation konsentrasjonen vil dermed tilsvare konsentrasjonen til $MgSO_4$ som er på 2,0 mM i sluttkonsentrasjonen til PCR reaksjonen. Vi benyttet dette til å beregne den korrigerede basert på ukorrigert T_m verdi som tar høyde for til primeren, for å bestemme hybridiseringstemperaturen til long-range primerparene ved første oppsett. Dette ble gjort for å være sikker på at alle primerparene blir dekket av hybridiseringstemperaturen (Tabell Supplement 4).

Som Tabell Supplement 4 illustrerer er long-range primerpar 4 med navnet «(4)PRSS1-LR-A/S» det primerparet med høyest korrigeret T_m på 61,0 °C. Det var ikke hensiktsmessig å regne ut korrigeret T_m verdi for sekvenseringsprimerne, da nukleotidsekvenseringen ikke er like følsom og i tillegg benytter BigDye Terminator v1.1 v3.1 5X Sequencing Buffer som bl.a. inneholder magnesiumklorid i ukjent konsentrasjon.

For å finne utgangsbetingelser til LR PCR ble Datablad fra Thermofisher for High Fidelity enzyme (53) benyttet som utgangspunkt. Vi foretok egne vurderinger ut ifra disse (Tabell 10).

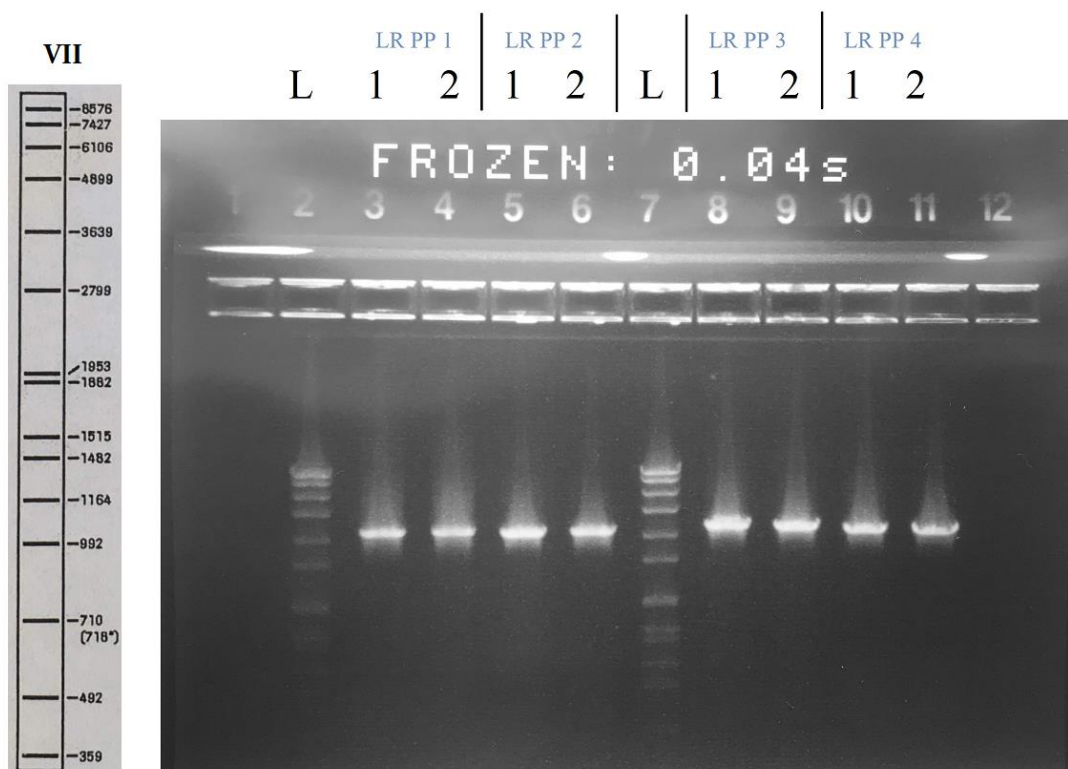
Tabell 10: Vurderinger av utgangsbetingelser hentet fra datablad fra Thermofisher for Long-range PCR.

Trinn i LR PCR:	Vurderinger av utgangsbetingelser:	Bestemte utgangsbetingelser:
Innledende denaturering	Hvis GC-innholdet i templatet er 50 % eller mindre, vil en innledende 1-3 min denaturering ved 94 °C være tilstrekkelig. Den innledende denatureringen ble økt til 5 minutter, da vi ikke kjente GC innholdet i templatet, og behandler det som et GC-rikt templat.	94°C i 5 minutter
Denaturering	Denaturering ved 94-95 °C i 20-60 sekunder er vanligvis tilstrekkelig for å fullstendig denaturere templat DNA i begynnelsen av PCR for å sikre effektiv utnyttelse av templatet under den første amplifiserings syklusen. Det ble valgt denaturering ved 94°C i kortest mulig tid, dvs. 20 sekunder, for å få mest mulig spesifikke PCR-produkt.	94°C i 20 sekunder per syklus
Hybridisering	Hybridiseringstemperaturen til PCR oppsettet ble valgt etter primerparet med høyest korrigert T _m som var long-range primerpar 4 ((4)PRSS1-LR-A/S) med T _m på 61,0 °C. Tankegangen er at lavere T _m enn 61°C ikke vil påvirke binding nevneverdig, mens høyere T _m kan resultere i ingen binding. Hybridiseringstemperaturen ble derfor ikke justert underveis i utprøvingen av long-range primere. Hybridisering i 30 sekunder er vanligvis tilstrekkelig, men vi kutte denne med 5 sekunder, til 25 sekunder.	61°C i 25 sekunder
Polymerisering	Anbefalt forlengelsestemperatur er 72 °C for polymerisering av PCR-produkter opptil 3 kb, og 68 °C når en polymeriserer fragmenter >3 kb. Produktet vårt var på ca. 4kb som vist i Tabell 5 og vi benyttet derfor 68 °C. Den anbefalte tidsinnstillingen er 1 minutt per 1 kb. Produktet vårt var på ca. 4kb så vi benyttet 4 minutter.	68°C i 4 minutter
Antall sykler	Optimalt antall sykluser er mellom 25-35, men avhenger av mengden og kompleksiteten til templat-DNA og lengden på produktet. For å øke konsentrasjonen på PCR-produktet ble antall sykler økt til 38 sykler.	38 sykler
Avsluttende polymerisering og avkjøling	Siste forlengelsestrinn ble satt til å vare i 7 minutter for å fylle ut de utstikkende endene av PCR-produktet. Avkjøling til 4°C ble benyttet for å stabilisere produktet.	68°C i 7 minutter og avkjøling ved 4°C til avbrutt.

Del 4: Validering av primere

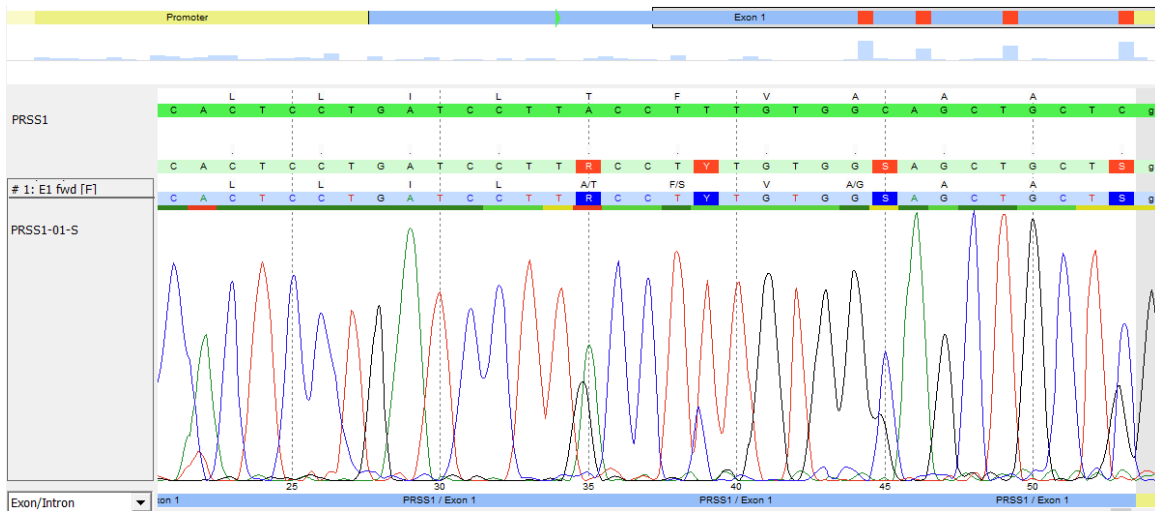
For å være sikker på at long-range primerne var spesifikke for *PRSS1*-genet og ga PCR-produkt i forventet størrelse, ble det utført gelelektroforese av PCR-produkt fra blodbank (BB) prøve 1 og 2 amplifisert med long-range primerpar (LR PP) 1-4 (Figur 25).

Produktet ble satt opp på to geler der størrelsesmarkør VII (Figur 25) og NTC for hver LR PP ble tatt med. Ved å se på resultatene fra dette var målet å kunne se om våre primere ga produkt med rett størrelse. LR PP 1 skulle gi et produkt på 3991 nt, LR PP 2 skulle gi 3910 nt, LR PP 3 skulle gi 4087 nt og LR PP 4 skulle gi 4004 nt, hvor alle skal havne mellom de samme to båndene produsert fra størrelsesmarkørens fjerde og femte bånd telt ovenfra og nedover. Hvis prøvene gir kun ett bånd i dette området vil det si at produktet er spesifikt med en størrelse på mellom 4899- og 3639 nt.

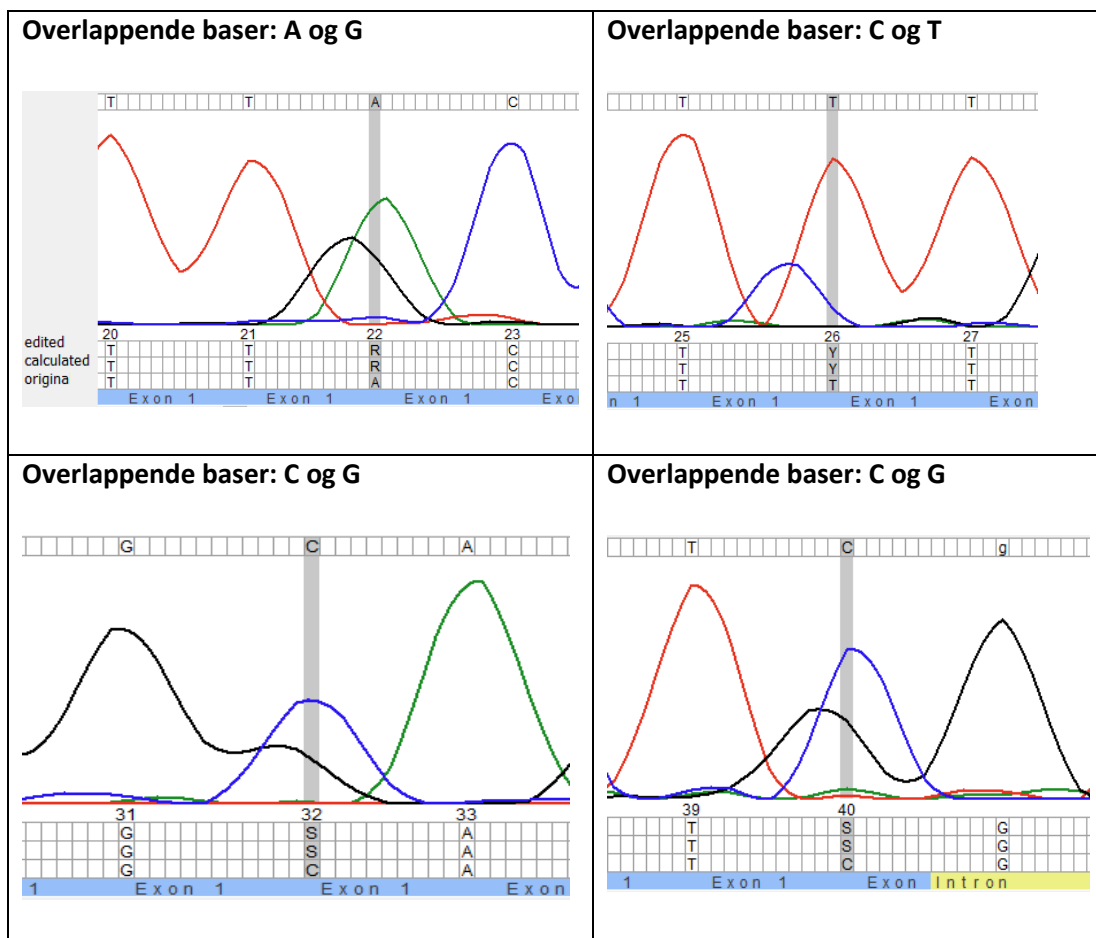


Figur 25: Størrelsesmarkør VII til venstre med resultat fra oppsett med gel for long-range primerpar 1-4 til høyre. Nr. 1 og 2= Blodbankprøve 1 og 2. L= Størrelsesmarkør.

Figur 25 bekreftet at PCR-produktet hadde en størrelse på mellom 4899 og 3639 nt, men i brønn 8 og 9 vises det et svakt dobbelt bånd for long-range primerpar 3 som kan skyldes et uspesifikt produkt. Båndet til LR PP 1 og LR PP 2 så skarpt ut og det ble utførte en nukleotidsekvensering med de designede sekvenseringsprimerne for å se etter bakgrunn, og etter basesekvenser som kan komme fra *PRSS2* eller pseudogen. LR PP 4 ble beholdt som en reserve. Sekvensene ble analysert i SeqPilot og viste at produktet til LR PP 1 var uspesifikt (Figur 26).



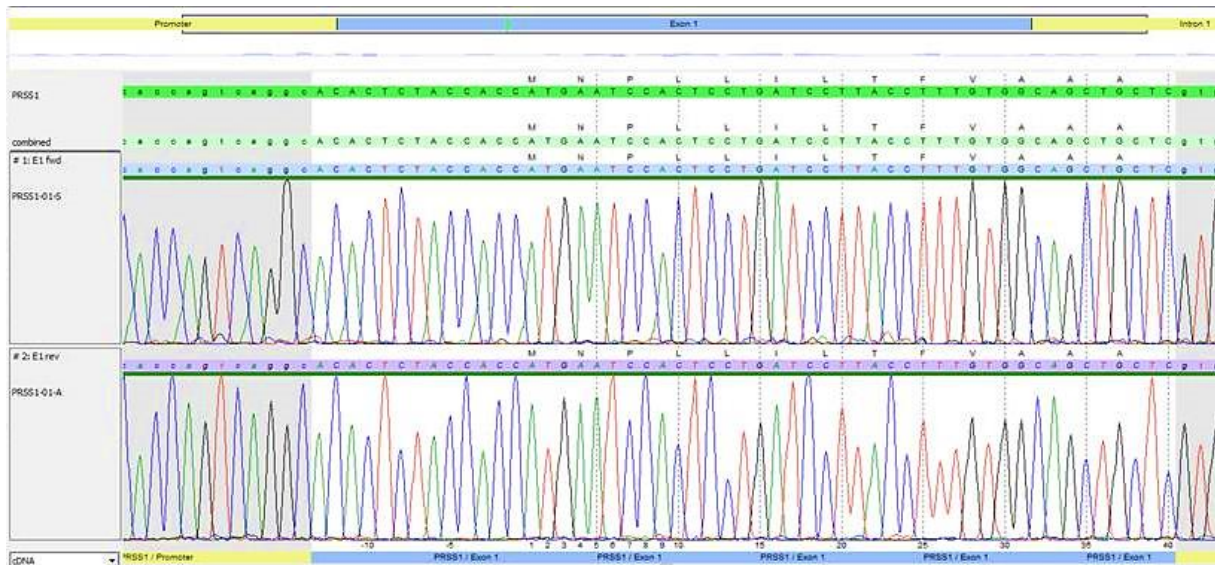
Figur 26: Analysedata fra ekson 1 i blodbankprøve FD47 amplifisert med LR PP 1 med uspesifikt produkt som vist i SeqPilot. Uoverensstemmelse mellom referansesekvens til *PRSS1*-genet og blodbankprøven FD47 er markert med røde firkanter.



Figur 27: Uoverensstemmelse mellom referansesekvens til *PRSS1*-genet og blodbankprøven FD47. De markerte posisjonene viser overlappende sekvenser hentet fra Figur 26. Utklipp fra SeqPilot.

Som vist i Figur 26 og

Figur 27 er det tydelig overlappende signaler i noen posisjoner i sekvensen sammenlignet med long-range primerpar 2 for samme pasientprøve (Figur 28).



Figur 28: Analysedata fra ekson 1 i blodbankprøve FD47 amplifisert med LR PP 2 som vist i SeqPilot. Ingen uoverensstemmelse mellom referansesekvens til *PRSS1* og blodbankprøven FD47 er markert.

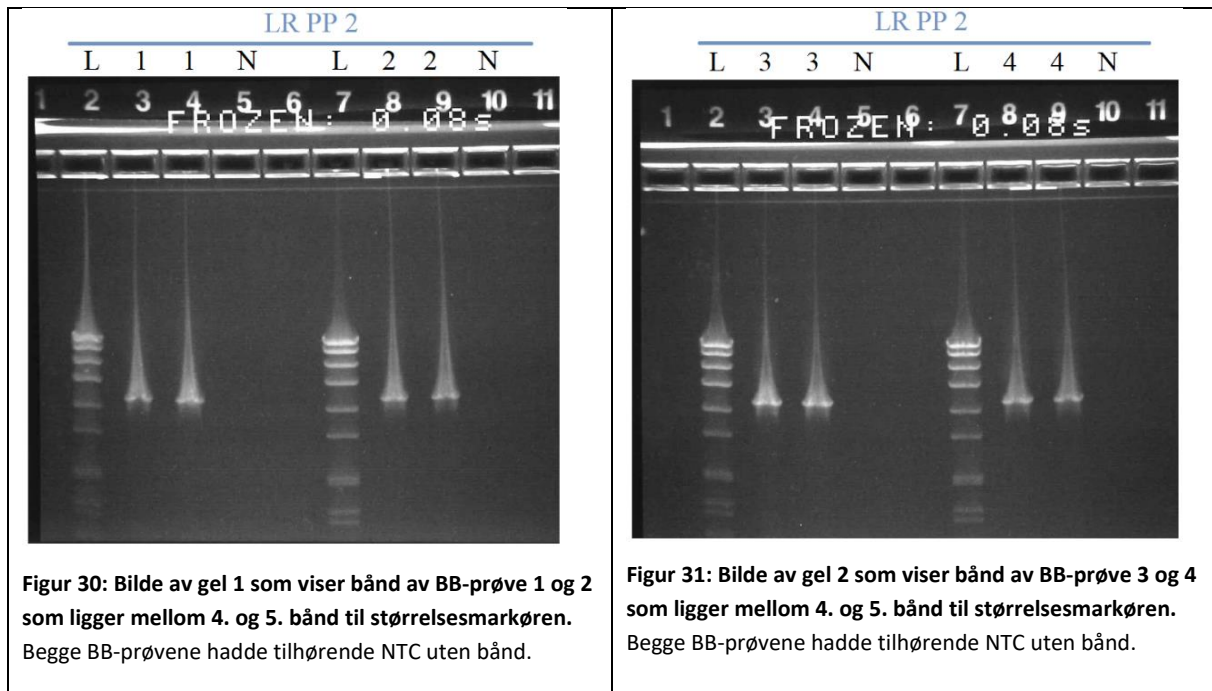
Hvis vi sammenligner posisjonene med overlappende signaler i Figur 26, ser vi en overlapp mellom sekvensen fra ekson 1 i *PRSS1* og pseudogenet *PRSS3P2* (Figur 29)

Sekvens <i>PRSS1</i>: ATGAATCCA CT CCTGATCCTTACCTTTGTGGCAGCTGCTC
Sekvens <i>PRSS2</i>: ATGAATC TA CT T CTGATCCTTACCTTTGT T GCAGCTGCT G
Sekvens <i>PRSS3P2</i>: ATGAATCCA CT CCTGATCCTT G CCTTTGTGG G AGCTGCT G
SeqPilot sekvens ekson 1: ATGAATCCA CT CCTGATCCTT R CCT T GTGG S AGCTGCT S

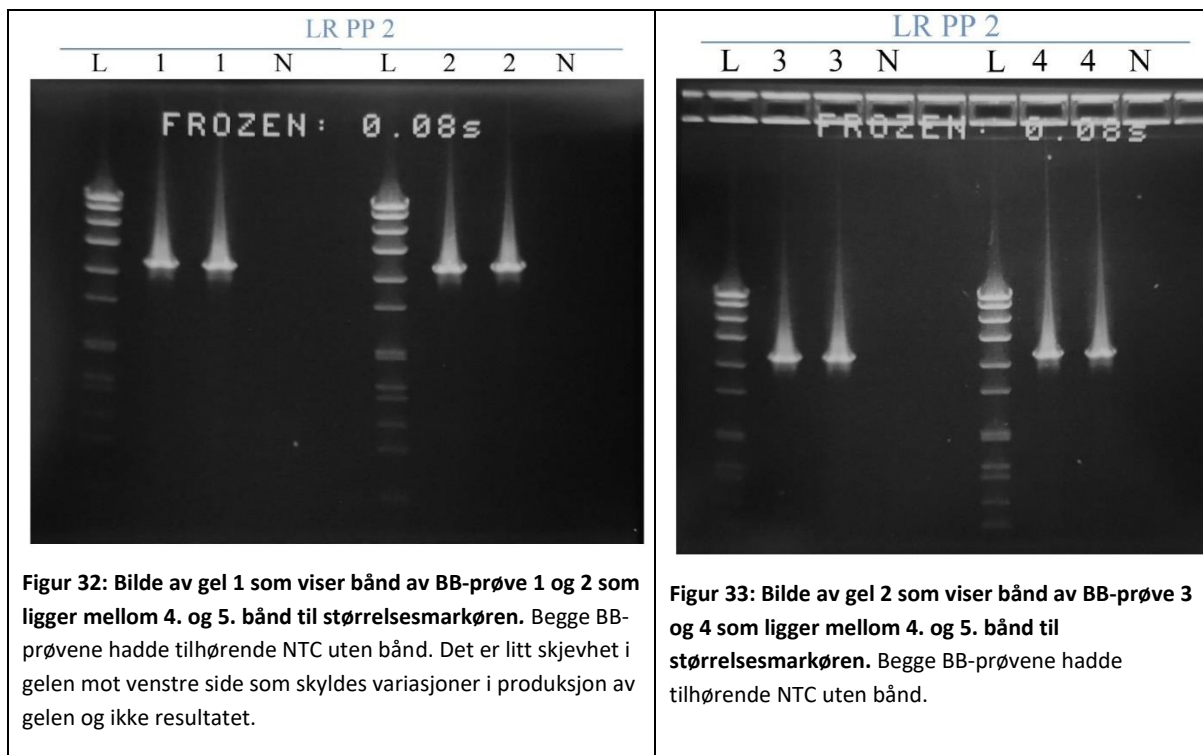
Figur 29: Sammenligning av referansesekvensen til *PRSS1* og basesekvensen i SeqPilot fra analysedataene. Avvik mellom referansesekvensen til *PRSS1* og basesekvensen i SeqPilot, samt tilsvarende sekvens i *PRSS2* og *PRSS3P2* er fremhevet i rødt. Mulig overlapp mellom *PRSS1* og *PRSS2*=1. Overlapp mellom *PRSS1* og *PRSS3P2*=3.

Som Figur 29 illustrerer har det med stor sannsynlighet skjedd en overlapp av sekvenseringsdata mellom *PRSS1* og *PRSS3P2*. Selv om vi prøvde å velge primere med størst mulig spesifisitet til *PRSS1* ligger antisense-primere i et område vi ikke har oversikt over SNPer, siden det manglet data fra GnomAD og dbSNP (Tabell Supplement 2). Det kan derfor være variasjoner i bindingssetet som svekker hybridiseringen. Grunnen til at vi kun så overlapp for ekson 1, kommer trolig av at de øvrige sekvenseringsprimere har bundet seg mer spesifikt til *PRSS1*. Vi valgte derfor å gå videre med long-

range primerpar 2 til validering. Alle 4 BB-prøver ble satt opp med en parallell, NTC og størrelsesmarkør fordelt på to geler (Figur 30 og Figur 31).



Måling av eksakt styrke og størrelse på LR PCR-produkt er ikke mulig ved bruk av gelelektroforese. Gel-resultatene ble derfor vurdert kvalitativt. Figur 30 og Figur 31 viser klare og skarpe bånd, men styrken på båndene er relativt svak til at kameraet hadde en lukketid på 0,08 sekunder. Målet ved oppsettet var å oppnå PCR betingelser som ga sterkest produkt som fortsatt var spesifikt. Utgangsbetingelsene for PCR reaksjonen ifra følgeskriv til Platinum *Taq* DNA polymerasen ble derfor endret i flere forsøk. Kun de endringene som ga ønsket effekt ble tatt med videre. Blant endringene som ble gjort var det å øke mengden initialt DNA fra 50ng til 75ng og endre forlengelsestrinnet i LR PCR fra 4 minutter til 4 minutter og 30 sekunder, vist å gi sterkere bånd (Figur 32 og Figur 33). Det ble senere vist at båndene holdt seg like sterke ved 50ng DNA, slik at volumet ble redusert tilbake igjen.



Figur 32 og Figur 33 viser blankt i NTC, klar og separert størrelsesmarkør, og klare bånd i brønnene til prøvene med ønsket styrke ved 0,08s lukketid på kameraet.

For å godkjenne long-range PCR metoden måtte vi validere primerne etter AMG sitt interne kvalitetssikringssystem for metodeetablering av fullscreening/mutasjonscreening. Hver analyse genererer mye data, og for å vurdere dataene er man avhengig av parametre innen kategoriene; visuell, kvalitet og intensitet. De tre lettest tilgjengelige parametrene i vårt tilfelle er rådata/analysert elektroferogram (visuelt), gjennomsnittlig kvalitets verdi kjent som QVa (average quality value) og gjennomsnittlig signalstyrke (intensitet) (54).

For å visualisere og vurdere rådatene og sjekke eventuelle feilkilder ble analyseprogrammet Sequence Scanner 2 benyttet. Filene fra ABI3730 lastes opp som .abi1 filer til Sequence Scanner Software 2 og åpnes i Trace Manager. Her skalerer vi signalene for å se etter store kvalitetsforstyrrelser og vurdere relativ signal styrke. Vi undersøkte sekvensdataene kalt «trace» for de vanligste potensielle kvalitetsproblemene og rådata variasjon mellom prøvene (Figur Supplement 2) (55).

Strukturartefakter som ble vurdert hos rådataene til blodbankprøvene (Figur 34) var unormale topper (peaks), tilblending av andre PCR-produkt (krysskontaminering), og primerrester.



Figur 34: Miniaturbildevisning av rådata fra sekvensering med long-range primerpar 2 som vist i Seq Scanner 2.

Som rådatene i Figur 34 viser, er det en høy topp på slutten av rådataen til samtlige sekvenseringsprimere til long-range primerpar 2. Vi kontrollerte at instrumentet, polymeren og riktig BigDye var blitt brukt i forhold til det eksperimentelle oppsettet. Piggen vi ser stemmer overens med en normal reptasjonstopp, «reptation peak», som vist i punkt 2 i Figur Supplement 2, og kan forekomme ved nesten all kapillærelektroforese (56). På slutten av kjøring med stor datainnsamlingstid i forhold til det som er nødvendig for sekvenseringsproduktet vil det kunne ses en slik reptasjonstopp. Området rett før og etter piggen er uten sekvens informasjon, så det trengtes ikke å gjøre endringer i oppsettet og vi kunne gå videre med å vurdere sekvensene.

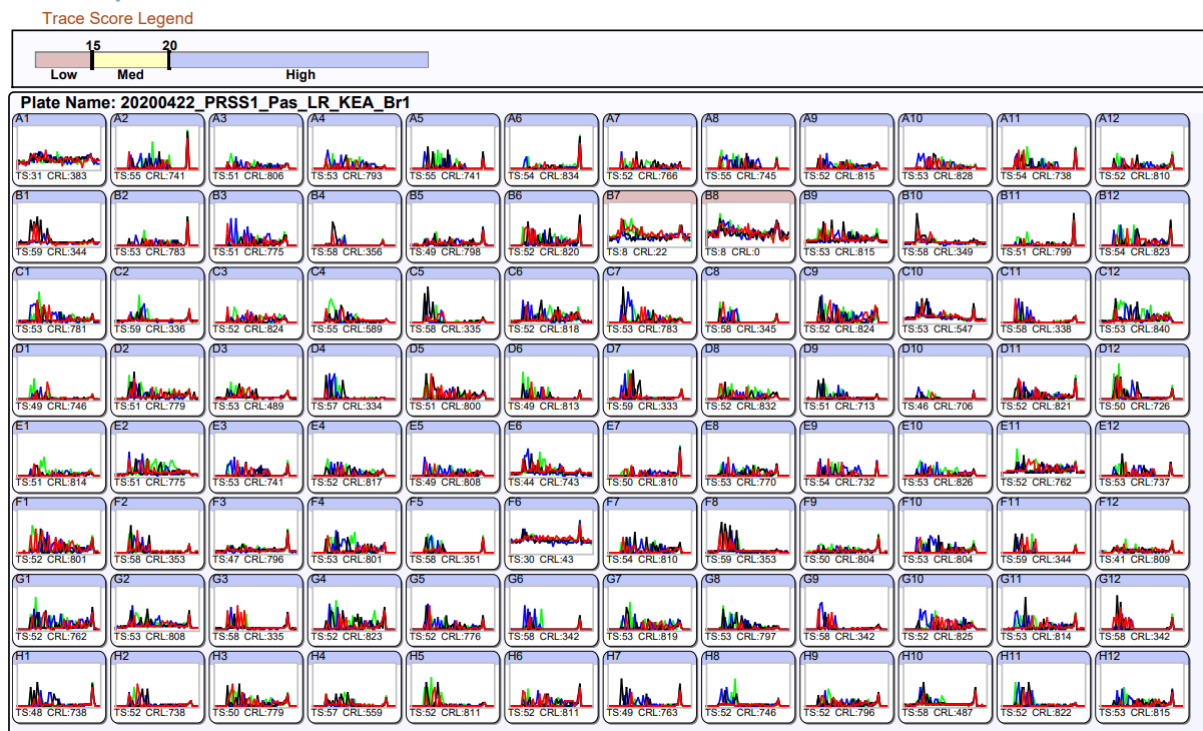
Sammen med den visuelle vurderingen av rådataene fra ABI3730 ble intensiteten av fluorescenssignalet målt i relative fluorescence unit (RFU) vurdert. AMG har satt krav om at samtlige primere som skal godkjennes til bruk i diagnostikken må ha en signalstyrke på minimum 500 rfu og

maksimum 10 000 rfu. På grunn av lave signaler og lave QVa verdier ved de første oppsettene ble det gjort endringer i nukleotidsekvenseringstrinnet.

Utgangsbetingelsene i dette trinnet ble hentet fra følgeblad for BigDye® Direct Cycle Sequencing Kit Protocol (57).

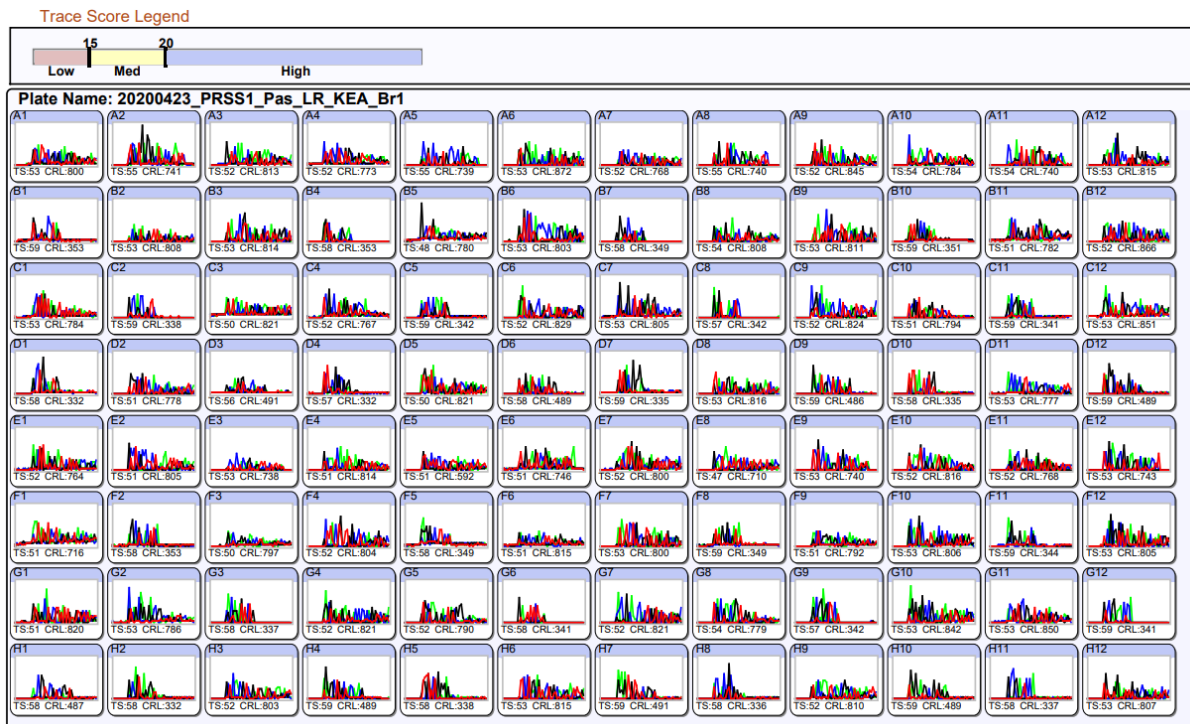
Av erfaring har AMG benyttet 53°C i hybridiseringstrinnet og 26 sykler. Dette ble benyttet som utgangsbetingelser for metoden. Under testing for å øke signalstyrken gav en økning i hybridiseringstemperaturen fra 53°C til 54°C sterkere signaler, trolig på grunn av mer spesifikk binding av sekvenseringsprimerne. Forskjellig innsugingstid i instrument-protokollen ble også testet på ABI3730 (Figur 35 og Figur 36). Instrument-protokollen vi tok utgangspunkt i var for lange cDNA produkt ved RNA analyse. Run_time ble derfor satt ned til 2780 sekunder som brukes for DNA analyser. Injection_time ble økt fra 3 sekunder til 6 sekunder, men senere satt ned til 4 sekunder.

Plate Report



Figur 35: Rådata fra Sangersekvensering av pasientprøver med long-range primerpar 2 og innsugingstid satt til 3 sekunder. Importerte rådata fra oppsettet er presentert i miniatyrbildevisning som vist i Seq Scanner 2.

Figur 35 viser at flere sekvenser lav trace score (<50) og noen så lave at de ikke regnes for å gi noe signal, eksempelvis brønn B7 og B8. Samme brettet ble satt på ABI3730 med endret innsugingstid til 6 sekunder (Figur 36)

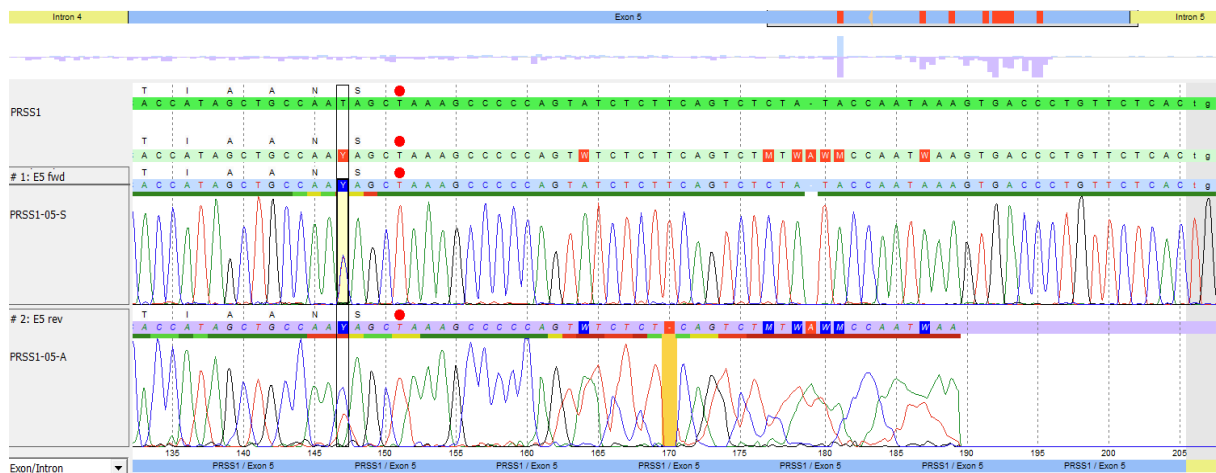


Figur 36: Rådata fra Sangersekvensering av pasientprøver med long-range primerpar 2 og innsugingstid satt til 6 sekunder. Importerte rådata fra oppsettet er presentert i miniatyrbildevisning som vist i Seq Scanner 2.

Figur 36 viser en generell økning i trace score, de fleste over 50 ts. AMG har satt krav til at kvalitetsverdiene – QVa - (gjennomsnitt for et definert område) må være ≥ 20 hvis en har to sekvenserte tråder. Samtlige blodbankprøver hadde over 50 ts i QV og signalstyrken kunne dermed vurderes (Tabell Supplement 6). Alle sekvenseringsprimerne hadde en signalstyrke på over 500 rfu og sekvensene ble visuelt vurdert.

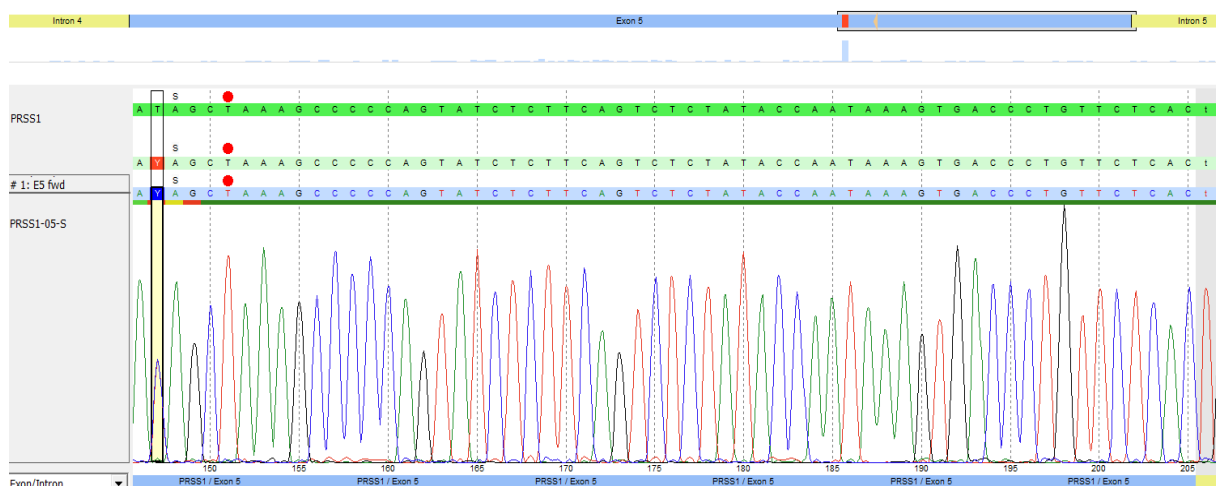
AMG har også satt krav til kvalitetsverdiene om at flankerende sekvens må være $\geq 20-30$ nt før og etter transkribert området i målsekvensen med unntak av de fragmentene hvor ønskede analysekriterier ikke er oppfylt. Når vi designet indre sekvenseringsprimere ønsket vi å plassere antisense-primere til ekson 5 i området som var dokumentert i gnomAD og dbSNP. Primerne ble plassert mellom det transkriberte området etter stoppkodon til ekson 5, og området som er dokumentert i GnomAD og dbSNP. Det resulterte i at primeren havnet for nært det transkriberte området, men $>20-30$ nt fra stoppkodonet slik at primeren likevel kunne godkjennes. Base calling tilbake mot stoppkodonet ble påvirket, men hadde ikke en innvirkning på videre variantundersøkelser (Figur 37). Siste trinn i valideringen av metoden var en visuell vurdering av «base calling» kvaliteten. Vi vurderte sekvensene til alle sekvenseringsprimerne til de fire blodbankprøvene i SeqPilot. De ble vurdert i forhold til regelmessig «base spacing» (G-rike regioner

kan være tettere), jevn høyde på toppene (Peak height) og signal-to-noise som må være mindre enn 1/4 av totale signalstyrken.



Figur 37: Analysedata fra ekson 5 i blodbankprøve FD47 amplifisert med LR PP 2 som vist i SeqPilot. Tydelig dårlig signal i starten av antisense-tråden gir uoverensstemmelse mellom referansesekvensen til *PRSS1* og blodbankprøven som markert med røde bokser. Den første røde boksen etter stoppkodonet, markert med rød sirkel, er utenfor det dårlige området og markerer en homozygot benign variant.

Som Figur 37 illustrerer har starten på antisense-tråden til ekson 5 for dårlig kvalitet på signalene til å kunne skille mellom ulike baser og gir usikkerhet i basene som vist med røde bokser i SeqPilot programmet. Dette selv om signalstyrken er høy. Signalene er dårlige fordi primeren ligger for tett på målsekvensen slik at de første basene etter primerbindesetet blir dårlige. SeqPilot har muligheten til å ignorere slutten på en sekvens eller en hel sekvens hvis signalet på den andre sekvensen gir tydelige topper med god datakvalitet (Figur 38).



Figur 38: Analysedata fra ekson 5 i blodbankprøve FD47 amplifisert med LR PP 2 med hele antisense tråden ignorert. Kun en heterozygot variant er markert.

Som Figur 38 illustrerer er datakvaliteten god nok for sensetråden til at vi kan ignorere det dårlige området av antisense tråden, siden resten av sekvensen er fin. Da står vi igjen med kun en reel benign variant. Oppsummert ble alle primerne godkjent for bruk i lab og vi kunne starte med reanalyse av de 15 pasientprøvene med tilsvarende vurdering av sekvenseringsdataene.

Del 5: Reanalyse av pasientprøver

Når pasientprøvene skulle bli reanalysert med den validerte LR PCR-metoden ble det sjekket at det hadde blitt oppformering av PCR-produkt etter LR PCR med bånd med god intensitet og spesifisitet. Dette ble gjort ved å kjøre en agarosegel. Alle båndene i gelen kom frem tydelige og NTC kom ut blank, samtlige av pasientprøvene gikk videre til sekvensering på ABI3730.

Selv om det er et krav til minst 500 rfu i signalstyrke for alle analyser, finner vi fra litteraturen (54) at signalstyrken kan ligge ned mot 200 rfu, og kan vurderes visuelt ved signalstyrke ned mot 20 rfu (SeqPilot krever en minimumsverdi på >30), forutsatt en tilstrekkelig høy QV verdi (Tabell 11).

Merknad: dette gjelder ikke validering av analysen, men kan benyttes ved rutineanalyser.

Tabell 11: Oversikt over kvalitetsparametre som benyttes til vurdering av analysedata fra ABI3730. Ruter markert i grønt indikerer godkjente parametre, ruter markert i gult kan godkjennes ved en visuell vurdering i henhold til AMG sine krav.

QV*:	Signalintensitet [rfu]:	Anbefalt handling:	Begrunnelse:
>50	<20	finnes ikke	finnes ikke
	20-200	visuell vurdering	lav kvalitet
	200-15,000	analyser	ok for analyse
	>15,000	finnes ikke	finnes ikke
49-36	<20	ikke analyser	for dårlig
	20-200	visuell vurdering	lav kvalitet og/eller leserammeskift
	200-15,000	analyser	ok for analyse
	>15,000	visuell vurdering	gått i metning?
37-17	<20	ikke analyser	for dårlig
	20-200	visuell vurdering	lav kvalitet og/eller leserammeskift
	200-15,000	analyser	ok for analyse
	>15,000	visuell vurdering	gått i metning?
<16	<20	ikke analyser	for dårlig
	20-200	visuell vurdering	lav kvalitet og/eller leserammeskift
	200-15,000	finnes ikke	finnes ikke
	>15,000	finnes ikke	finnes ikke

*QV er oppgitt i SeqPilot i tabellen "Positions / Result files" i kolonne average Phred

Tabell Supplement 7 og Tabell Supplement 8 viser at alle pasientprøver hadde en trace score (QV) på >50 (med unntak av primer 5A for D1593-19 som var på 49). Alle primerne hadde signalstyrke på over 200 rfu, med unntak av primer 2A for D293-20, primer 4A for D264-20, D2181-19, D2072-19, D1907-19, og D1107-19, samt primer 2A-S for D2360-18. Disse sekvensene ble vurdert visuelt med de andre parameterne på samme måte som under valideringen.

For hver pasientprøve ble i tillegg tidligere NGS data, der det foreligger uspesifisitet, filtrert i FILTERUS og varianter sjekket i Alamut. Ved å sammenligne disse med varianter funnet i SeqPilot med metoden vi etablerte (Tabell Supplement 9 - Tabell Supplement 20) vil man se om den nye analysen er selektiv for *PRSS1*-genet, eller om det finnes interferens med pseudogener eller homologe gener. Det ble funnet totalt to forskjellige normalvarianter, c.486T>C og c.738T>C, i pasientprøvene med vår metode. Begge variantene opptrer som homo- eller heterozygot i SeqPilot. Det var ikke tilgjengelige positivkontroller for analysen. Fra NGS dataene ble det funnet totalt 35 ulike varianter, hvorav 4 (merket CLIN_pathogenic i Alamut) kunne blitt tolket som sykdomsassosiert hvis de ble vurdert som troverdige funn og ikke tilblending fra andre gener. De to variantene vi fant i SeqPilot fant vi også igjen i NGS, noe som underbygger metodens nøyaktighet. Første variant finner vi igjen i NGS med henholdsvis allelefraksjon på 40:60 og 20:80 (Tabell Supplement 14). Den siste varianten finner vi igjen i NGS med fordeling på 100% for homozygot og 50:50 for heterozygot (Tabell Supplement 20). **Vi har dermed lyktes i å øke nøyaktigheten i påvisningen av relevante varianter i *PRSS1*-genet.**

Diskusjon

I denne oppgaven undersøkte vi de trypsinogen-kodende genene *PRSS1* og *PRSS2*, samt pseudogenet *PRSS3P2*. Med disse resultatene designet vi long-range primere som var spesifikke for *PRSS1*. Analyse av gener med pseudogener og homologe sekvenser er beskrevet i litteraturen som vanskelig. Om Sanger-sekvensering uten spesifikk LR PCR hadde blitt benyttet, ville resultatene blitt omtrent det samme som ved NGS analyse av disse genene. Uten spesifikk amplifisering av hele *PRSS1*-genet vil sekvenseringsprimere ikke være spesifikke nok og kunne feste seg i pseudogener slik som vi så med første sekvenseringsforsøk i Figur 26. Blant de 4 primerparene som ble designet ble LR PP2 valgt som best egnet. Dette primerparet og sekvenseringsprimere ble validert med 4 blodbankprøver. Valideringen viste at metoden nøyaktig sekvenserte *PRSS1*-genet. Etter metoden ble validert ble tidligere pasientprøver reanalysert med vår metode. Sammenlignet med gammel metode som benyttet NGS ble resultatene våre betydelig mer nøyaktige. Kun reelle varianter ble funnet i prøvemateriale. To normalvarianter, c.486T>C og c.738T>C, gikk igjen i mange av prøvene analysert med vår metode. Ingen varianter knyttet til pankreatitt ble funnet. I analysedata fra NGS ble det funnet flere varianter i alle pasientprøvene. Dette inkluderer reelle varianter i genet og varianter som dukker opp i analysen på grunn av pseudogener og homologe sekvenser. Siden det kommer frem flere varianter ved NGS analyse gir dette også rom for feiltolkning. Vi har vist med metoden vår at ved å benytte LR PCR kan vi få spesifikke resultater kun fra det aktuelle *PRSS1*-genet.

Med høykapasitets sekvenseringsdata genereres det mer data, og mye arbeid går med på å analysere og tolke resultatene. Ressursene som brukes for å tolke ett svar blir derfor mer verdifull. Vi ser at vår metode gir større sikkerhet i prøvesvarene. Med den økte nøyaktigheten og gode spesifisiteten vil tolking av prøvene gå vesentlig raskere med vår metode og åpner for en stor besparelse av ressurser som ellers ville gått med på å vurdere flere varianter unødvendig. Selve oppsettet til analysen av *PRSS1* med LR PCR gjøres realistisk over to dager. NGS oppsettet går over flere dager, og åpner dermed for muligheten til å sette opp metoden vår parallelt.

Denne metoden benytter noe spesielle programinnstillinger til PCR og Sanger-sekvensering som vil begrense mengden genanalyser som kan settes opp parallelt. Dette påvirker rutinene på lab i liten grad. Vi anslår at tiden som blir spart ved analyse av dataene, med høyde for begrensninger, gir en raskere metode hvis analysen blir utført mens NGS-oppsettet forberedes.

På grunn av det lille antallet varianter som kan forventes å bli funnet og den økte sikkerheten med vår metode for å finne reelle varianter, vurderer vi denne metoden som en nødvendig tilleggsanalyse når *PRSS1*-genet skal bli analysert. I prøvematerialet vi undersøkte fant vi ingen risiko-varianter. Til videre arbeid kan metoden testes mot kjente risiko-variant for å få bekreftet metodens spesifisitet.

På grunn av høyst 80% penetrans vil ikke genfeil i *PRSS1*-genet gi grunnlag for å ta i bruk metoden vår i fosterdiagnostikk. Det tilbys generelt ikke i Norge for kreft-relaterte gener, og spesielt med høyst andel sykdomsdebut i voksenalder. Av de samme årsakene kan heller ikke metoden benyttes til preimplantasjonsdiagnostikk. Siden penetransen ikke er 100% er det en sjanse for at fosteret kan utvikle seg uten å bli påvirket og leve et normalt liv. Dette veier mot at metoden vår heller kan benyttes til diagnostikk i tidlig alder av barn av foreldre med kjent arvelig pankreatitt.

Pasienter med påvist genfeil vil få tilbud om tettere oppfølging og kontrollopplegg. Mulige tiltak er å begrense store fettrike måltider, og unngå alkohol i senere alder, noe som kan senke frekvensen av anfall med akutt pankreatitt. Selv om det ikke finnes noen direkte tiltak for å helbrede tilstanden, kan en tettere oppfølging gjennom oppveksten gi økt livskvalitet for pasienten. I senere alder vil en jevnlig undersøkelse hos lege for tidlig oppdagelse av pankreaskreft være aktuelt for pasienter som får påvist risikovarianter i *PRSS1*-genet. Ved tidlig oppdagelse av kreft gir det en økt mulighet for kurative inngrep.

Konklusjon

I dette bachelorprosjektet skulle vi finne en egnet metode for å spesifikt analysere *PRSS1*-genet. Analysen skulle gi svar på om det ble oversett/ikke detektert sykdomsgivende sekvensvarianter. Vi lyktes med å bestemme long-range PCR program-innstillinger og designe primere som kun oppformerte *PRSS1*-genet med eventuell genkonversjon og ikke pseudogener eller homologe gener. Produktet ble sjekket med agaose gelelektroforese og videre nukleotidsekvensert med egne designede sekvenseringsprimere for å bestemme DNA-sekvensen og påvise mulige sekvensendringer. Metoden ble validert med analyseprogrammene Sequence Scanner 2 (vurdering av datakvalitet) og Sequence Pilot (sekvensanalyse og vurdering av kvalitet). Det ble funnet totalt to forskjellige normalvarianter i pasientprøvene med vår metode. Fra NGS dataene ble det funnet totalt 35 ulike varianter som trolig stammer fra tilblending fra andre gener og som kunne blitt tolket som sykdomsassosiert hvis de ble vurdert som troverdige funn. Vi har dermed lyktes med å etablere en målrettet long-range PCR metode som økte nøyaktigheten i påvisningen av klinisk relevante varianter i *PRSS1*-genet betydelig. **Fra dette arbeidet kan det konkluderes at den etablerte long-range PCR metoden kan benyttes som en tilleggs undersøkelse for å kunne oppdage genfeil i *PRSS1*-genet på en pålitelig måte hvis panelet «NGS-bukspyttkjertelkreft-CNV» er rekvirert eller ved indikasjon «pankreatitt».**

Referanseliste

1. Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, mfl. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet.* oktober 1996;14(2):141–5.
2. Larsen IK. Cancer in Norway 2018 [Internett]. 2019 okt [sitert 15. mars 2020]. Tilgjengelig på: <https://www.kreftregisteret.no/Generelt/Rapporter/Cancer-in-Norway/cancer-in-norway-2018/>
3. Howes N, Lerch MM, Greenhalf W, Stocken DD, Ellis I, Simon P, mfl. Clinical and genetic characteristics of hereditary pancreatitis in Europe. *Clin Gastroenterol Hepatol.* 1. mars 2004;2(3):252–61.
4. Teich N, Nemoda Z, Köhler H, Heinritz W, Mössner J, Keim V, mfl. Gene conversion between functional trypsinogen genes PRSS1 and PRSS2 associated with chronic pancreatitis in a six-year-old girl. *Hum Mutat.* 2005;25(4):343–7.
5. Lowenfels AB, Maisonneuve P, DiMagno EP, Elitsur Y, Gates LK, Perrault J, mfl. Hereditary Pancreatitis and the Risk of Pancreatic Cancer. *JNCI J Natl Cancer Inst.* 19. mars 1997;89(6):442–6.
6. Hidalgo M. Pancreatic Cancer. *N Engl J Med.* 29. april 2010;362(17):1605–17.
7. Bukspyttkjertelkreft [Internett]. NHI.no. [sitert 13. mars 2020]. Tilgjengelig på: <https://nhi.no/sykdommer/kreft/magetarm-kreft/bukspyttkjertelkreft/>
8. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell.* 4. mars 2011;144(5):646–74.
9. Tisdale MJ. Mechanisms of Cancer Cachexia. *Physiol Rev.* 1. april 2009;89(2):381–410.
10. Muñoz N, Castellsagué X, de González AB, Gissmann L. Chapter 1: HPV in the etiology of human cancer. *Vaccine.* 21. august 2006;24:S1–10.
11. Chen S, Parmigiani G. Meta-Analysis of BRCA1 and BRCA2 Penetrance. *J Clin Oncol Off J Am Soc Clin Oncol.* 10. april 2007;25(11):1329–33.
12. Aabakken L. bukspyttkjertelbetennelse. I: Store medisinske leksikon [Internett]. 2020 [sitert 13. mars 2020]. Tilgjengelig på: <http://sml.snl.no/bukspyttkjertelbetennelse>
13. Bukspyttkjertelbetennelse, akutt [Internett]. NHI.no. [sitert 3. mai 2020]. Tilgjengelig på: <https://nhi.no/sykdommer/magetarm/bukspyttkjertel/bukspyttkjertelbetennelse-akutt/>
14. Chen J-M, Férec C. Chronic Pancreatitis: Genetics and Pathogenesis. *Annu Rev Genomics Hum Genet.* 28. august 2009;10(1):63–87.
15. Salvesen G, Rawlings ND. Handbook of Proteolytic Enzymes [Internett]. Bd. 3rd ed. Oxford: Academic Press; 2012 [sitert 1. april 2020]. Tilgjengelig på: <https://www.sciencedirect.com/book/9780123822192/handbook-of-proteolytic-enzymes>

16. Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, mfl. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. *Nat Genet.* desember 2012;44(12):1349–54.
17. Chen J-M, Le Maréchal C, Lucas D, Raguénès O, Férec C. “Loss of function” mutations in the cationic trypsinogen gene (PRSS1) may act as a protective factor against pancreatitis. *Mol Genet Metab.* 1. mai 2003;79(1):67–70.
18. Mitchell R, Byrne M, Baillie J. Pancreatitis. *The Lancet.* 26. april 2003;361(9367):1447–55.
19. Chen J-M, Férec C. Gene Conversion-like Missense Mutations in the Human Cationic Trypsinogen Gene and Insights into the Molecular Evolution of the Human Trypsinogen Family. *Mol Genet Metab.* 1. november 2000;71(3):463–9.
20. Sjøberg NO. Molekylær genetik: genteknologi, humant DNA. 4. utg. Nesbru: Vett & viten; 2006. 333 s.
21. What are UniProtKB’s criteria for defining a CDS as a protein? [Internett]. [sitert 2. april 2020]. Tilgjengelig på: https://www.uniprot.org/help/cds_protein_definition
22. Witt H, Sahin-Tóth M, Landt O, Chen J-M, Kähne T, Drenth JP, mfl. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. *Nat Genet.* juni 2006;38(6):668–73.
23. Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. *FEBS Lett.* 2000;468(2–3):109–14.
24. Chen J-M, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet.* oktober 2007;8(10):762–75.
25. Kobayashi I. Mechanisms for gene conversion and homologous recombination: The double-strand break repair model and the successive half crossing-over model. *Adv Biophys.* 1. januar 1992;28:81–133.
26. Schildkraut E, Miller CA, Nickoloff JA. Gene conversion and deletion frequencies during double-strand break repair in human cells are controlled by the distance between direct repeats. *Nucleic Acids Res.* 1. mars 2005;33(5):1574–80.
27. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. *Cell.* 1. mai 1983;33(1):25–35.
28. Petronella N, Drouin G. Strong purifying selection against gene conversions in the trypsin genes of primates. *Hum Genet.* 1. november 2012;131(11):1739–49.
29. Rygiel AM, Beer S, Simon P, Wertheim-Tysarowska K, Oracz G, Kucharzik T, mfl. Gene Conversion Between Cationic Trypsinogen (PRSS1) and the Pseudogene Trypsinogen 6 (PRSS3P2) in Patients with Chronic Pancreatitis. *Hum Mutat.* 2015;36(3):350–6.
30. Rowen L, Koop BF, Hood L. The Complete 685-Kilobase DNA Sequence of the Human β T Cell Receptor Locus. *Science.* 21. juni 1996;272(5269):1755–62.

31. Chen J-M, Montier T, Férec C. Molecular pathology and evolutionary and physiological implications of pancreatitis-associated cationic trypsinogen mutations. *Hum Genet.* 1. september 2001;109(3):245–52.
32. MLA CE Course Manual: Molecular Biology Information Resources (Genetics Review: Reading Frames) [Internett]. [sitert 2. april 2020]. Tilgjengelig på: <https://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Genetics/readingframe.html>
33. Mouagip. Aminoacids table (Codons sun) [Internett]. 2009 [sitert 2. april 2020]. Tilgjengelig på: https://commons.wikimedia.org/wiki/File:Aminoacids_table.svg
34. Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP. Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. *Mol Biol Evol.* 1. februar 2003;20(2):248–54.
35. Tm Calculator - NO [Internett]. [sitert 2. april 2020]. Tilgjengelig på: <https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html>
36. Allawi HT, SantaLucia J. Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry.* 26. august 1997;36(34):10581–94.
37. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* 18. juni 2012;13:134.
38. NEB Tm Calculator [Internett]. [sitert 2. april 2020]. Tilgjengelig på: <https://tmcalculator.neb.com/#!/main>
39. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci.* 17. februar 1998;95(4):1460–5.
40. Owczarzy R, You Y, Moreira BG, Manthey JA, Huang L, Behlke MA, mfl. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry.* 1. mars 2004;43(12):3537–54.
41. Life Technologies. Platinum Taq DNA Polymerase High Fidelity [Internett]. Tilgjengelig på: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/0948_PlatinumTaqDNA_Polymerase_High_Fidelity.pdf
42. Blueprint Genetics | Global Genetic Tests and Genetic Diagnostics [Internett]. Blueprint Genetics. [sitert 2. april 2020]. Tilgjengelig på: <https://blueprintgenetics.com/>
43. Life Technologies. Sequence Scanner Software. Get the software safe and easy. [Internett]. Software Informer. [sitert 16. mai 2020]. Tilgjengelig på: <https://sequence-scanner-software.software.informer.com/2.0/>
44. SEQUENCE PILOT [Internett]. JSI medical systems. [sitert 7. mai 2020]. Tilgjengelig på: <https://www.jsi-medisys.de/products/sequence-pilot/>
45. FILTUS [Internett]. omicX. [sitert 7. mai 2020]. Tilgjengelig på: <https://omictools.com/filtus-tool>

46. Alamut® Visual: a mutation analysis software [Internett]. Interactive Biosoftware. [sitert 7. mai 2020]. Tilgjengelig på: <https://www.interactive-biosoftware.com/alamut-visual/>
47. Home | HUGO Gene Nomenclature Committee [Internett]. [sitert 2. april 2020]. Tilgjengelig på: <https://www.genenames.org/>
48. LRG_1013 - PRSS1 [PENDING APPROVAL] [Internett]. [sitert 2. april 2020]. Tilgjengelig på: http://ftp.ebi.ac.uk/pub/databases/lrgex/pending/LRG_1013.xml
49. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 19. august 2004;5:113.
50. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*. desember 2013;30(12):2725–9.
51. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 1. mai 1993;10(3):512–26.
52. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. 1985;39(4):783–91.
53. Product Information: High Fidelity PCR Enzyme Mix, #K0191. :13.
54. Noguchi T, Bourdon V, Sobol H. About Sequence Quality: Impact on Clinical Applications. *Genet Test Mol Biomark*. 12. mars 2014;18(5):299–305.
55. Applied Biosystems. Troubleshooting Sanger sequencing data [Internett]. [sitert 7. mai 2020]. Tilgjengelig på: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0014435_Trbleshoot_Sanger_seq_data_UB.pdf
56. Applied Biosystems. DNA Sequencing by Capillary Electrophoresis [Internett]. [sitert 7. mai 2020]. Tilgjengelig på: https://tools.thermofisher.com/content/sfs/manuals/cms_041003.pdf
57. Applied Biosystems. BigDye® Direct Cycle Sequencing Kit Protocol [Internett]. [sitert 13. mai 2020]. Tilgjengelig på: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_091370.pdf

Vedlegg

Tabell Supplement 1: Liste over arter med sekvenser innhentet fra tblastx søk med homologi til det menneskelige PRSS1-genet.

<p>[1.1] Makak (<i>Macaca nemestrina</i>) [Primater] serine protease 2 (PRSS2): XM 024789591.1:302-1110</p> <p>[1.2] Makak (<i>Macaca nemestrina</i>) [Primater] trypsin-1-like (LOC105471239): XM 011723600.2:53-861</p> <p>[1.3] Makak (<i>Macaca nemestrina</i>) [Primater] putative trypsin-6 (LOC105471238): XM 011723599.2:704-1512</p> <p>[2.1] Rhesus ape (<i>Macaca mulatta</i>) [Primater] serine protease 2 (PRSS2): XM 015135091.2:219-1027</p> <p>[2.2] Rhesus ape (<i>Saimiri boliviensis boliviensis</i>) [Primater] anionic trypsin (LOC101033484): XM 003929786.2:71-878</p> <p>[3.1] Fjellbavian (<i>Theropithecus gelada</i>) [Primater] trypsin-2 (LOC112620197): XM 025378482.1:66-874</p> <p>[3.2] Fjellbavian (<i>Theropithecus gelada</i>) [Primater] trypsin-1-like (LOC112620199): XM 025378491.1:1-809</p> <p>[4.] Ugandisk rød colobus-ape (<i>Ptilocolobus tephrosceles</i>) [Primater] serine protease 2 (PRSS2): XM 023184897.2:20-827</p> <p>[5.1] Stumpneseape (<i>Rhinopithecus roxellana</i>) [Primater] serine protease 2 (PRSS2): XM 010379380.2:67-875</p> <p>[5.2] Stumpneseape (<i>Rhinopithecus roxellana</i>) [Primater] trypsin-1-like (LOC104670402): XM 030932093.1:26-834</p> <p>[6.] Sort stumpneseape (<i>Rhinopithecus bieti</i>) [Primater] trypsin-2 (LOC108526176): XM 017866824.1:66-874</p> <p>[7.] Oliven bavian (<i>Papio Anubis</i>) [Primater] trypsin-1 (LOC100998585): XM 031664950.1:1-826</p> <p>[8.] Boliviaekornape (<i>Macaca mulatta</i>) [Primater] putative trypsin-6 (LOC716882): XM 001087821.4:1-802</p> <p>[9.1] Menneske (<i>Homo sapiens</i>) [Primater] serine protease 3 (PRSS3): NM 007343.3</p> <p>[9.2] Menneske (<i>Homo sapiens</i>) [Primater] PRSS3 pseudogene 2 (PRSS3P2): NR 001296.3</p> <p>[9.3] Menneske (<i>Homo sapiens</i>) [Primater] serine protease 2 (PRSS2): NM 002770.4:1-809</p> <p>[9.4] Mennesket (<i>Homo sapiens</i>) [Primater] serine protease 1 (PRSS1): NM 002769.5</p> <p>[10.1] Sumatraorangutang (<i>Pongo abelii</i>) [Primater] putative trypsin-6 (LOC100452346): XM 009243333.2:66-858</p> <p>[10.2] Sumatraorangutang (<i>Pongo abelii</i>) [Primater] trypsin-2 (LOC100443647): XM 024249595.1:22-830</p> <p>[10.3] Sumatraorangutang (<i>Pongo abelii</i>) [Primater] trypsin-1 (LOC100444247): XM 024249598.1:1-807</p> <p>[11.1] Pygme sjimpanse (<i>Pan paniscus</i>) [Primater] putative trypsin-6 (LOC100993827): XM 003820568.2:66-858</p> <p>[11.2] Pygme sjimpanse (<i>Pan paniscus</i>) [Primater] serine protease 2 (PRSS2): XM 008974013.1:22-829</p> <p>[11.3] Pygme sjimpanse (<i>Pan paniscus</i>) [Primater] trypsin-1 (LOC100992428): XM 003820563.2:1-807</p> <p>[12.1] Vestgorilla (<i>Gorilla gorilla gorilla</i>) [Primater] putative trypsin-6 (LOC101145449): XM 004046368.2:66-858</p> <p>[12.2] Vestgorilla (<i>Gorilla gorilla gorilla</i>) [Primater] serine protease 2 (PRSS2): XM 031012890.1:1-805</p> <p>[12.3] Vestgorilla (<i>Gorilla gorilla gorilla</i>) [Primater] trypsin-1 (LOC101147134): XM 019030569.2:1-803</p> <p>[13.1] Nordlig hvit-kinnet gibbon (<i>Nomascus leucogenys</i>) [Primater] serine protease 2 (LOC100591887): XM 030825929.1:72-879</p> <p>[13.2] Nordlig hvit-kinnet gibbon (<i>Nomascus leucogenys</i>) [Primater] trypsin-1 (LOC100592562): XM 003270841.4:1-803</p> <p>[14.1] Sjimpanse (<i>Pan troglodytes</i>) [Primater] serine protease 2 (PRSS2): XM 003318843.3:3945-4753</p> <p>[14.2] Sjimpanse (<i>Pan troglodytes</i>) [Primater] trypsin-1 (LOC742453): XM 016956966.1:32-787</p> <p>[15.1] Panamansk hvitkledd capuchin-ape (<i>Cebus capucinus</i>) [Primater] imitator protease serine 2 (PRSS2): XM 017511340.1:66-874</p> <p>[15.2] Panamansk hvitkledd capuchin-ape (<i>Cebus capucinus</i>) [Primater] imitator cationic trypsin-3 (LOC108291165): XM 017511346.1:29-744</p> <p>[16.1] Robust capuchin-ape (<i>Sapajus apella</i>) [Primater] serine protease 2 (PRSS2): XM 032245199.1:66-874</p> <p>[16.2] Robust capuchin-ape (<i>Sapajus apella</i>) [Primater] trypsin (LOC116528011): XM 032245203.1:84-829</p>	<p>[17.] Sau (<i>Ovis aries</i>) [Klovdyr] anionic trypsin (LOC101112559): XM 027968886.1:1-768</p> <p>[18.] Kanin (<i>Oryctolagus cuniculus</i>) [Kanin- og haretyr] protease serine 2 (PRSS2): XM 002721845.2:1-738</p> <p>[19.] Hest (<i>Equus caballus</i>) [Upartåede hovdyr] anionic trypsin (LOC100055297): XM 023639857.1:30-828</p> <p>[20.] Arktisk jordekorn (<i>Urocitellus parryii</i>) [Gnagere] anionic trypsin-2 (LOC113199668): XM 026412712.1:1-738</p> <p>[21.] Gulbukmurdedyr (<i>Marmota flaviventris</i>) [Gnagere] serine protease 2 (PRSS2): XM 027943399.1:20-767</p> <p>[22.] Nakenrotte (<i>Heterocephalus glaber</i>) [Gnagere] trypsin (LOC101719198): XM 021240846.1:218-918</p> <p>[23.1] Spissmus (<i>Mus pahari</i>) [Insektetere] anionic trypsin-2-like (LOC110316506): XM 021190859.1:88-838</p> <p>[23.2] Spissmus (<i>Mus Pahari</i>) [Insektetere] cationic trypsin-3 (LOC110316467): XM 021190803.1:31-773</p> <p>[24.] Ryukyu-mus (<i>Mus caroli</i>) [Gnagere] anionic trypsin-2-like (LOC110289230): XM 021155373.1:75-825</p> <p>[25.] Brunrotte (<i>Rattus norvegicus</i>) [Gnagere] protease serine 2 (PRSS2): XM 012729.2:1-768</p> <p>[26.] Hjortemus (<i>Peromyscus maniculatus bairdii</i>) [Gnagere] anionic trypsin-2 (LOC102905363): XM 006998117.2:1-738</p> <p>[27.1] Mongolsk ørkenrotte (<i>Meriones unguiculatus</i>) [Gnagere] anionic trypsin-2 (LOC110539819): XM 021626642.1:90-837</p> <p>[27.2] Mongolsk ørkenrotte (<i>Meriones unguiculatus</i>) [Gnagere] cationic trypsin-3 (LOC110543493): XM 021629803.1:38-741</p> <p>[28.] Indisk glassfisk (<i>Parambassis ranga</i>) [Beinfisk] trypsin-3-like (LOC114448060): XM 028424706.1:138-814</p> <p>[29.] Mummichog (<i>Fundulus heteroclitus</i>) [Beinfisk] trypsin-3 (LOC105932609): XM 012871847.2:109-787</p> <p>[30.] Gulabor (<i>Perca flavescens</i>) [Beinfisk] trypsin-3 (LOC114556580): XM 028579572.1:68-744</p> <p>[31.] Sikk-sakk ål (<i>Mastacembelus armatus</i>) [Beinfisk] trypsin-3 (LOC113122268): XM 026293475.2:117-814</p> <p>[32.] Stumpneset klengefisk (<i>Gouania willdenowii</i>) [Beinfisk] trypsin-3 (LOC114477702): XM 028470196.1:100-800</p> <p>[33.1] Hund (<i>Canis lupus familiaris</i>) [Dyr] cationic trypsin-like (LOC475521): XM 532744.4:117-867</p> <p>[33.2] Hund (<i>Canis lupus familiaris</i>) [Dyr] protease serine 2 (PRSS2): XM 003432056.3:57-816</p> <p>[34.1] Hawaiiimunkesel (<i>Neomonachus schauinslandi</i>) [Dyr] cationic trypsin (LOC110582989): XM 021693046.1:114-855</p> <p>[34.2] Hawaiiimunkesel (<i>Neomonachus schauinslandi</i>) [Dyr] protease serine 2 (PRSS2): XM 021693011.1:3-769</p> <p>[35.] Nordamerikansk serotine (<i>Eptesicus fuscus</i>) [Flaggermus] anionic trypsin (LOC103297283): XM 008153974.2:81-842</p> <p>[36.] Natal lang-fingret flaggermus (<i>Miniopterus natalensis</i>) [Flaggermus] protease serine 2 (PRSS2): XM 016195499.1:1-744</p> <p>[37.] Piggsvin (<i>Erinaceus europaeus</i>) [Insektetere] anionic trypsin (LOC103112822): XM 007522304.2:1-738</p> <p>[38.] Blek spyd-nese flaggermus (<i>Phyllostomus discolor</i>) [Flaggermus] anionic trypsin (LOC114507497): XM 028525291.1:26-747</p> <p>[39.] Filippinsk tarsier (<i>Carlito syrichta</i>) [Primater] protease serine 2 (PRSS2): XM 008067747.2:2-755</p> <p>[40.] Sundaskjelldyr (<i>Manis javanica</i>) [Placentale pattedyr] anionic trypsin (LOC108403392): XM 017670529.1:1-744</p> <p>[41.] Røyskatt (<i>Mustela erminea</i>) [Dyr] anionic trypsin (LOC116569686): XM 032306129.1:60-821</p> <p>[42.] Steinkobbe-sel (<i>Phoca vitulina</i>) [Dyr] anionic trypsin (LOC116630705): XM 032401429.1:3-769</p> <p>[43.] Stellers' sjøløve (<i>Eumetopias jubatus</i>) [Dyr] anionic trypsin (LOC114211590): XM 028106185.1:1-766</p> <p>[44.] Californiasjøløven (<i>Zalophus californianus</i>) [Dyr] serine protease 2 (PRSS2): XM 027573040.1:160-880</p>
---	---

5'..... tgggaggagt ggtgaccctc acctcacagt cacctcctct ctggatcctc gtgagg**tata aa**gacgagtc
ctccaccacc agtc**AGGCAC ACTCTACCAC C ATG AAT CCA CTC CTG ATC CTT ACC TTT GTG GCA GCT GCT**
C GTGAGTATC ATGCCCTGCC TCAGGCCCCA ACCACCCCCG CGTTCCTGGC CGACAAATGC CCTTCCATC
TTACGACCTC TCCTCTTTTG ACTGTGCTCT GATATTCCGT TTCCTCCATC TGGCATATCT CCTTCCATC CTCCTTGGGC
TCTTTTAAAG TCTCACCTGT TCACCTTCTC CTTGACTTCA CTCCCACCAC TGTCAATCAT CCATATCCGA GTTGTGGTTG
GAGAAGCTGG GAAGGGGGCC AGGTGGGGCT GTCCACGAA ATGAAGCAGC AGGCTTCAGG CTGGCTCCA ACAGCAACAG
AATAGCACCA CTATAGCTGC TCCTAACCTC GAATGCACCT GGGGAGGTTG AAAAAATACT CATGCCAGAG ACTCAGCTCT
AGAAATTCTA ACTTCAATTG TCTGGGGTAG AGCTTGTGTT CTGGGCTTTT AAGCTTCCA GGTGATTTTT AATATTTCCA
GCCATGCAGC CAAGGTTAAG AATTGTGTC CTATTGGCCA ATAACAAGT CTACCTTGT TCTGCCAAG TGAGCCTGGG
GGTGCCCTC AACTCTGCC TGACTGCACA GATCTGAGCT ATGGGGGAAG GTGGTCATGG CCAGGTCTAT GCAGACAGGG
GGTTTTCCTA GCTTGGCAA AGAATCTGA CAATCCAGG CCCAAATAGC CAGGGGAAGT ACACAGGTGA TGAATAAAG
AGAGAAGCAC TCAGTGGGAG AGACAACCAC ATCCCAACTC CTATCCACT GGAAGCATTG TGAGGACATT CCTTGGGACT
TCAGCCTGGT GACCCAGGA GAGCTCGGAT CCTCCGAGG GTACCTAGCT ATGTGCCCTG CAGGCACAGA GACTTGGGAG
CCACAGGCAG TGATGATCAC CAGGGTGGC AGAGCTCCCT CCCTTGCCTA GCCTACTGT GCTTGTAAAG GATTTCATAT
TAGCAGAAAG CAATCACAGG CTGGGAGCG CACCCCTAAC ATGCTATTGA CCTTGCCTTCT CCTTCCCAT
CTCCACTCCAG **TT GCT GCC CCC TTT GAT GAT GAT GAC AAG ATC GTT GGG GGC TAC AAC TGT GAG GAG**
AAT TCT GTC CCC TAC CAG GTG TCC CTG AAT TCT GGC TAC CAC TTC TGT GGT GGC TCC CTC ATC AAC
GAA CAG TGG GTG GTA TCA GCA GGC CAC TGC TAC AAG TC GTAAGTGTG GGGCCCCGA CTGCAAAGCT
CCCCGCCAGT CTGCCTGGGA GAGCTTGGCT TCAGCCCAGG GAACTACTGA GGTGGGTAA GATGGATGGG AGAGGTGGTG
GAGAAGAAA CTTGTTGGCA GCTGCGGACT CTCAGAGCA GAGAGTGAAC ACAAGACAGG AACCTCTCAC ACCCAGGCAA
ATCCATGAAA CAGCAAGGT TGTGGTCATA AAAGCAGGCA GGGATGATCT TGGGGTGGT AGAGCTAGTG AGAAAAGCAG
GCAAGTATCT TTTGCTGGTT AGCTACACAT TAAAGCCAAC TAAGAAAGAC TTTTAAAAA TACAGATGCC TTTGCTCTAT
CCCAG G GCA ATT AAG TCA AAA TTT TCA GGA AGA GGG TGT GAA TAT CA GTG AGAATTTTAC
ACTCTACCTC TGCTAACTGT AGAGTGATA GACAGAGCTG AGAAGTGTG CCTACACAA GAACTCTCAA ACCTGAGTAT
GCATCAGAAC GCCCTGCAGG CTTGTTAAG CACAAATCAC TGGGCCCTT CCCCAGGTT CTGATCAGTA GGTGGGGTA
AGGACCAAGA ATTCACATTT CTAACAAGTT CCCAGGAGAT GCTAATGCTA TGGCTACCCT TGGATTAGAT TACACAGAAG
GGTGGTTCTC ACCAGGCCAA GAATGGAGG AGGAACAGGC ACTGTGCACA GTTGGCAAAG GCCTGGGGTG AAGAACGCTG
GGAAACTTC AAGGAGCTCC TTGTGCCAC AGTGCTAGTG ACTGTGGAGA TTGTGGGAAA GAGTCTGGGG AGGCAGTTG
AGGAGCAGCC TCTGGTGGGA TCCCTTTGAC TCTTCCCCAC CCCACTACCA CCAACCTCTG GAGCAGATAG GTCCTGGGTC
TCATACCTTC ACTGACCAC ATCCCTCTGC TGCCCATGCG ATATGGCCAC ACACCCACC CCATGCCTCC AGAGCTGTCC
ATGAGCAGAG AGCTTGAGGA ACCTGGGAA GGTGGATAG GTGCCCTGGC TGTGGGAGAA GGTCTTACC ATGCCTGCC
TGCCCATCAG **C CGC ATC CAG GTG AGA CTG GGA GAG CAC AAC ATC GAA GTC CTG GAG GGG AAT GAG**
CAG TTC ATC AAT GCA GCC AAG ATC ATC CGC CAC CCC CAA TAC GAC AGG AAG ACT CTG AAC AAT GAC
ATC ATG TTA ATC AAG CTC TCC TCA CGT GCA GTA ATC AAC GCC CGC GTG TCC ACC ATC TCT CTG CCC
ACC GCC CCT CCA GCC ACT GGC ACG AAG TGC CTC ATC TCT GGC TGG GGC AAC ACT GCG AGC TCT GGC
G GTGAGT GGGACCTTA GTCCTTCTAC TTCCCTCCAT CCTCACAATT TCCAGAACAA ACCATGCCCC TTAAGTTGAA
TCCTCTACC TCCAGGCTTA AGACACATTT CGAGTGCCCA TTACACACAG ACTCTGCACT GGGCACCAGA GAGATGCAAA
CTATCAAGGA CTTGGCTCCT AAAATCAAGA GACAGGACAA ATGGAGAACT TGATATGATC ACATCTTGGG AGGGGTTCAA
CAATGATCAT TCTGGGAAT AAAAGCAGA GTCTTTGCC AGGACTTATG TTCTGGAGTC CTCTCCAGGG GCTGTGTCC
TCTTCAGTTT TCCATCCAAG ATTATTGTCT CCTTCTCTGG CCTGACCAC ATTTCTACTT CCTTGTATCT CTTCTGATC
CTCACAG **CC GAC TAC CCA GAC GAG CTG CAG TGC CTG GAT GCT CCT GTG CTG AGC CAG GCT AAG TGT**
GAA GCC TCC TAC CCT GGA AAG ATT ACC AGC AAC ATG TTC TGT GTG GGC TTC CTT GAG GCA GGC AAG
GAT TCA TGT CAG GTGATT TGACCAACCC TTCCATGCT GAGGCTCCCA CTGATACCTA GGCCCCACCA
GGGAAAAGGA TTTGAATCA AAAGGTGGTG GGGCTGAGGA GGCTCCCTGC AGTGCCACA TGGAGAAGTG AGGAAGACTC
CCTTGGGCTG CATCTGTCT GCTTAGGAAG AACAGAGAAT GGGCCACCAT GAGAAGGACA TGGAGCCACA GAGCTGGCTG
GAAAGGGGTC TTTAAGGTT CAGAGTAAAT GTAGTATAT TCCTCTCCA TCTCTCCATA CAACTTGTCC CTTCTTCCC
CCAG GGT GAT TCT GGT GGC CCT GTG GTC TGC AAT GGA CAG CTC CAA GGA GTT GTC TCC TGG GGT
GAT GGC TGT GCC CAG AAG AAC AAG CCT GGA GTC TAC ACC AAG GTC TAC AAC TAT GTG AAA TGG ATT
AAG AAC ACC ATA GCT GCC AAT AGC TAAAGC CCCAGTATC TCTTCACTCT CTATACC**TAT AAG**GTGACC
TGTTCTCACT GTCtgtgtct gtgctgtct cctctcactc cttcacactg gaagcatcc tccaatttca ggt.....3'

Figur Supplement 1: Oversikt over PRSS1-genet med eksonene 1, 2, 2A, 3, 4 og 5 fremhvert i magenta. Basesekvensen er ordnet som tripletter (kodoner) med start og stopp kodon uthevet i gult. Viktige sekvenser i promotor er farget blått, spleisesteder i grønt og poly-A-signal i orange. Sekvenser før og etter transkripsjon er med små bokstaver. Store bokstaver er da transkribert "Coding Sequence" (CDS). (Ved behov kan nettsiden: <http://www.bioinformatics.org/sms2/index.html> benyttes til å omdanne sekvensen tilbake til en av de standardiserte formatene for videre arbeid)

Tabell Supplement 2: Oversikt over long-range primerpar og indresekvenserings primere, med søkeresultat fra gnomAD database etter genetiske SNP som kan redusere binding til DNA-tråden. Etnisitet er oppgitt som superpopulasjonene; europeisk (kaukasiske), afrikansk og asiatick (orientalske).

Primernavn	Sekvens	Variant (GRCh37)	Allele frekvens	Etnisitet	Allele count/number
(1)PRSS1_LR-S	GGGGCTTGACTTCCCTGTC	7-142457055-G-C	0.0003116	african	2/6418
	GGGGCTTGACTTCCCTGTC	7-142457055-G-C	0.00007298	europaean non-finnish	1/13702
	GGGGCTTGACTTCCCTGTC	7-142457061-G-C	0.0002302	african	2/8688
	GGGGCTTGACTTCCCTGTC	7-142457066-T-C	0.0002238	europaean non-finnish	3/13402
	GGGGCTTGACTTCCCTGTC	7-142457066-T-C	0.001002	african	6/5988
(1)PRSS1-LR-A	GAACACATTTGGGAGATGGGG	none		ingen i gnomAD (slutter på c.*54)	
(2)PRSS1-LR-S	TATCCTTGACGAAACAGCCG	7-142457087-T-A	0.0009423	african	5/5306
	TATCCTTGACGAAACAGCCG	7-142457087-T-A	0.0001590	europaean non-finnish	2/12582
	TATCCTTGACGAAACAGCCG	7-142457087-T-C	0.0005654	african	3/5306
	TATCCTTGACGAAACAGCCG	7-142457087-T-C	0.00007948	europaean non-finnish	1/12582
	TATCCTTGACGAAACAGCCG	7-142457094-G-C	0.003501	african	17/4856
	TATCCTTGACGAAACAGCCG	7-142457094-G-C	0.001225	other	1/816
	TATCCTTGACGAAACAGCCG	7-142457094-G-C	0.0008696	east asian	1/1150
	TATCCTTGACGAAACAGCCG	7-142457094-G-C	0.0007519	europaean non-finnish	9/11970
	TATCCTTGACGAAACAGCCG	7-142457106-C-T	0.0001947	europaean non-finnish	3/15410
	TATCCTTGACGAAACAGCCG	7-142457107-G-A	0.001157	europaean finnish	4/3458
	TATCCTTGACGAAACAGCCG	7-142457107-G-A	0.0009311	other	1/1074
	TATCCTTGACGAAACAGCCG	7-142457107-G-A	0.0002633	europaean non-finnish	4/15192
(2)PRSS1-LR-A	TCGTGTCTAACCTGAAATTTGGAG	none		ingen i gnomAD (slutter på c.*54)	
(3)PRSS1-LR-S	AGCCAGCTCTGTTATCTGG	7-142457043-C-G	0.00006489	europaean non-finnish	1/15410
	AGCCAGCTCTGTTATCTGG	7-142457045-C-G	0.0003125	african	2/6400
	AGCCAGCTCTGTTATCTGG	7-142457045-C-G	0.00007293	europaean non-finnish	1/13712
	AGCCAGCTCTGTTATCTGG	7-142457049-T-G	0.0001623	african	1/6162
	AGCCAGCTCTGTTATCTGG	7-142457049-T-G	0.00007377	europaean non-finnish	1/13556
	AGCCAGCTCTGTTATCTGG	7-142457051-T-C	0.0006435	east asian	1/1554
	AGCCAGCTCTGTTATCTGG	7-142457055-G-C	0.0003116	african	2/6418
	AGCCAGCTCTGTTATCTGG	7-142457055-G-C	0.00007298	europaean non-finnish	1/13702
(3)PRSS1-LR-A	ACAGTACTCCCCTTTCTCCCAAG	none		ingen i gnomAD (slutter på c.*54)	
(4)PRSS1-LR-S	CCAGCTCTGTTATCTGGGGCTT	Lik som (3)PRSS1-LR-S			
(4)PRSS1-LR-A	ACACATTTGGGAGATGGGGG	none		ingen i gnomAD (slutter på c.*54)	
PRSS1_1-S	GCAGGTGTGTTGTGCTGGGAG	7-142457222-G-A	0.0002296	african	2/8712
	GCAGGTGTGTTGTGCTGGGAG	7-142457225-G-C	0.0006418	east asian	1/1588
	GCAGGTGTGTTGTGCTGGGAG	7-142457225-G-C	0.0002594	europaean non-finnish	4/15422
PRSS1_1-A	ATGCCAGATGGAGGAAA	none			
PRSS1_2-S	CCTTGCCTAGCCTCACTG	7-142458291-C-T	0.0001152	african	1/8682
PRSS1_2-A	TCTCTGCTCTGGAGAGTC	7-142458702-A-T	0.00001257	europaean non-finnish	1/79568
	TCTCTGCTCTGGAGAGTC	7-142458704-T-G	0.00007596	south asian	2/26328
	TCTCTGCTCTGGAGAGTC	7-142458707-C-G	0.0002172	african	4/18418
	TCTCTGCTCTGGAGAGTC	7-142458710-G-C	0.0001948	south asian	5/25668
	TCTCTGCTCTGGAGAGTC	7-142458711-A-T	0.00001378	europaean non-finnish	1/72558
	TCTCTGCTCTGGAGAGTC	7-142458713-C-T	0.00003857	latino	1/25926
PRSS1_2A-S	TGTGGTCATAAAGCAG	7-142458776-G-T	0.00004372	south asian	1/22872
	TGTGGTCATAAAGCAG	7-142458778-G-T	0.00003907	latino	1/25598
	TGTGGTCATAAAGCAG	7-142458782-A-C	0.00004379	south asian	1/22836
	TGTGGTCATAAAGCAG	7-142458786-A-G	0.00001791	europaean non-finnish	1/55844
	TGTGGTCATAAAGCAG	7-142458789-C-G	0.0001266	african	2/15792
	TGTGGTCATAAAGCAG	7-142458790-A-C	0.00009050	east asian	1/11050
PRSS1_2A-A	GTTCTCAGCTCTGTCTATAC	7-142459000-G-A	0.00001916	europaean non-finnish	1/52186
	GTTCTCAGCTCTGTCTATAC	7-142459002-A-G	0.0003325	east asian	4/12030
	GTTCTCAGCTCTGTCTATAC	7-142459003-T-C	0.00009580	east asian	1/10438
	GTTCTCAGCTCTGTCTATAC	7-142459019-C-T	0.0001147	african	1/8716
PRSS1_3-S	ATATGGCCACACACCCACC	7-142459500-G-A	0.001182	latino	1/846
	ATATGGCCACACACCCACC	7-142459508-C-T	0.0003447	african	3/8704
PRSS1_3-A	GGATTCAAGTTAAGGGCATGG	none		ingen i gnomAD (slutter på c.454+54)	
PRSS1_4-S	GTTCTCTTACAGTTTCCATCC	none		ingen i gnomAD (starter på c.455-54)	
PRSS1_4-A	GAGTTCAAATCCTTTTCCC	none		ingen i gnomAD (slutter på c.591+52)	
PRSS1_5-S	CACCATGAGAAGGACATGGAG	none		ingen i gnomAD (starter på c.592-51)	
PRSS1_5-A	AGGCACAGACACAGACATG	7-142460925-G-A	0.00006486	europaean non-finnish	1/15418

Tabell Supplement 3: Oversikt over PCR reagenser med konsentrasjon hentet fra følgebladet til Platinum *Taq* DNA polymerase High Fidelity (41).

Component	25- μ L rxn	50- μ L rxn	Custom	Final Conc.
Autoclaved, distilled water	to 25 μ L	to 50 μ L	to μ L	–
10X High Fidelity PCR Buffer	2.5 μ L	5 μ L	μ L	1X
50 mM MgSO ₄	1 μ L	2 μ L	μ L	2.0 mM
10 mM dNTP Mix	0.5 μ L	1 μ L	μ L	0.2 mM each
10 μ M forward primer	0.5 μ L	1 μ L	μ L	0.2 μ M
10 μ M reverse primer	0.5 μ L	1 μ L	μ L	0.2 μ M
Template DNA	varies	varies		< 500 ng
Platinum® <i>Taq</i> DNA Polymerase High Fidelity (5 U/ μ L)	0.1 μ L	0.2 μ L	μ L	1 U/rxn

Tabell Supplement 4: Oversikt over long-range primerparene med beregnet og korrigert T_m verdier.

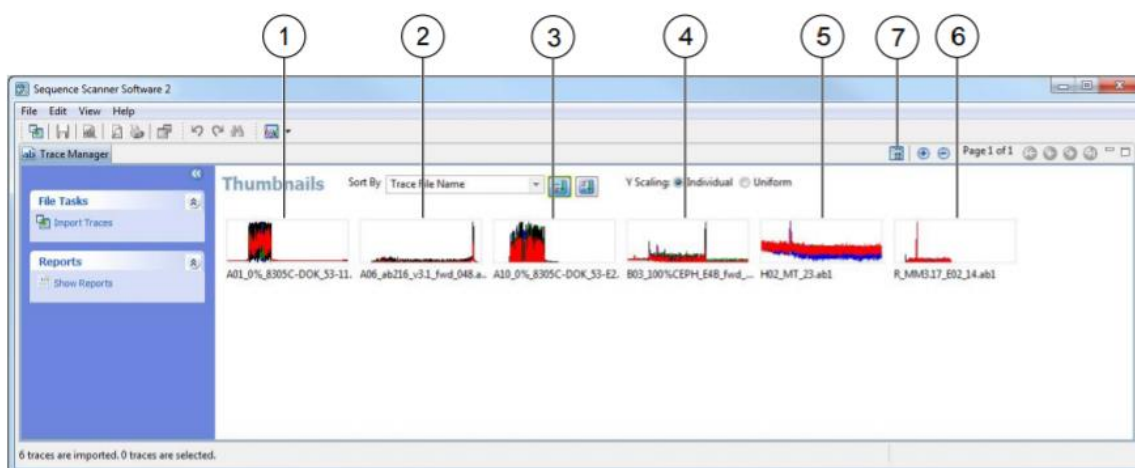
Primer:	Oligosekvens:	Lengde:	%GC:	fGC:	T _m BLAST (°C):	ThermoFisher T _m (°C):	T _m korrigert med Ckation (°C):
(1)PRSS1_LR-S	GGGGCTTGACTTTCCCTGTC	20	60,00		60.61		
(1)PRSS1_LR-A	GAACACATTTGGGAGATGGGG	22	50,00	0,55	59.50	61,2	59,5
(2)PRSS1_LR-S	TATCCTTGACAGAAACAGCCG	21	52,38		60.68		
(2)PRSS1_LR-A	TCGTGTCTAACCTGAAATTGGAG	23	43,48	0,48	58.43	59,6	57,9
(3)PRSS1_LR-S	AGCCCAGCTCTGTTATCTGG	20	55,00		59.16		
(3)PRSS1_LR-A	ACAGTACTCCCCTTTCTCCAAG	23	52,17	0,54	61.33	58,0	56,5
(4)PRSS1_LR-S	CCAGTCTGTTATCTGGGGCTT	22	54,55		62.08		
(4)PRSS1_LR-A	ACACATTTGGGAGATGGGGG	21	52,38	0,53	60.27	62,8	61,0

Tabell Supplement 5: Utgangsbetingelser til nukleotidsekvensering hentet fra protokoll til BigDye® Direct Cycle Sequencing Kit.

Stage	Veriti® thermal cyclers		9700 thermal cyclers	
	Temp	Time	Temp	Time
Hold	37°C	15 min	37°C	15 min
Hold	80°C	2 min	80°C	2 min
Hold	96°C	1 min	96°C	1 min
Cycle (25 cycles)	96°C	10 sec	96°C	10 sec
	50°C	5 sec	50°C	5 sec
	60°C	75 sec	60°C	4 min
Hold	4°C	∞	4°C	∞

Tabell Supplement 6: Signalstyrke til de indre sekvenseringsprimerne for blodbankprøve 1-4. S=sense, A=antisense. Første tallet referer til signalstyrken til Guanine (G) og er den basen som benyttes til vurdering av signalstyrken (resten er i rekkefølgen GATC).

Blodbankprøver: / Signal faktor (RFU) for ekson:	FD47 (BB-1)	FD48 (BB-2)	FD49 (BB-3)	FD50 (BB-4)
1S	<u>2931</u> , 2511, 2738, 3625	<u>830</u> , 790, 818, 1051	<u>1734</u> , 1835, 1910, 1984	<u>1584</u> , 1623, 1716, 1854
1A	<u>5417</u> , 4266, 3245, 3338	<u>5417</u> , 4266, 3245, 3338	<u>4934</u> , 4169, 3173, 2986	<u>2495</u> , 2050, 1553, 1491
2S	<u>1474</u> , 1548, 1379, 1217	<u>600</u> , 517, 475, 525	<u>1149</u> , 1173, 1051, 901	<u>715</u> , 677, 580, 566
2A	<u>4111</u> , 3376, 2522, 3572	<u>3291</u> , 2952, 2553, 2826	<u>2061</u> , 2272, 1907, 1847	<u>2823</u> , 2657, 2203, 2387
2A-S	<u>3695</u> , 3620, 2538, 3055	<u>1645</u> , 1686, 1562, 1658	<u>2080</u> , 2735, 2097, 1834	<u>2735</u> , 3119, 2322, 2205
2A-A	<u>2265</u> , 2147, 2196, 2173	<u>2265</u> , 2147, 2196, 2173	<u>2114</u> , 2368, 2337, 2134	<u>2939</u> , 2761, 2799, 2774
3S	<u>2428</u> , 2345, 1764, 2569	<u>2268</u> , 2652, 1879, 2615	<u>2317</u> , 2783, 2262, 2628	<u>2815</u> , 3225, 2669, 3075
3A	<u>1569</u> , 1156, 979, 1119	<u>1282</u> , 1152, 1018, 938	<u>2655</u> , 2066, 1630, 1797	<u>580</u> , 427, 356, 392
4S	<u>2156</u> , 1726, 1438, 1842	<u>1956</u> , 1932, 1795, 1675	<u>2145</u> , 1836, 1506, 1891	<u>1197</u> , 1093, 991, 989
4A	<u>1771</u> , 1823, 1619, 1585	<u>1816</u> , 1970, 1712, 1553	<u>827</u> , 706, 508, 542	<u>2300</u> , 2231, 2010, 2053
5S	<u>2483</u> , 2445, 2915, 2456	<u>3818</u> , 3612, 4020, 3805	<u>3723</u> , 3184, 3162, 3826	<u>2307</u> , 2013, 1910, 2146
5A	<u>2385</u> , 2126, 1670, 2156	<u>877</u> , 857, 669, 795	<u>1890</u> , 1712, 1343, 1708	<u>2254</u> , 2280, 1855, 2208



Figur Supplement 2: Oversikt over de vanligste kvalitetsproblemene ved Sangersekvensering på ABI3730. Miniaturbildevisning av importerte .ab1-filer: 1) Stor signalmetning (de fleste topper er over det anbefalte signalområdet). 2) Signalet er innenfor det anbefalte signalområdet og viser en normal reagentstopp. Reagentstopp er store topper observert på slutten av lange kjøringar. 3) Mindre signalmetning (noen topper er over det anbefalte signalområdet). 4) Innenfor det anbefalte signalområdet. 5) Lav signalintensitet. 6) Dye blob. 7) Vis detaljer-ikonet.

Tabell Supplement 7: Oversikt over signalstyrke og trace score for sekvenseringsprimere fra pasientprøvene D293-20 til D1801-19. S=Signal faktor. TS= Trace Score (Average Pread value / Datakvalitet). PAR= Parametre. PAS= Pasienter. Ex=Ekson. Ruter markert med beige indikerer sekvenser med for lav signalstyrke/trace score som krever visuell vurdering.

Pas:	D293-20		D264-20		D2181-19		D2072-19		D2031-19		D1907-19		D1857-19		D1801-19	
PAR:	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS
Ex1S	833, 951, 990, 1058	57	877, 1127, 1280, 1341	58	1600, 1694, 1853, 1983	58	1567, 1602, 1817, 2003	57	1787, 1783, 1991, 2218	57	2088, 1953, 2261, 2641	58	1619, 1661, 1802, 1938	57	1658, 1713, 1937, 2029	57
Ex1A	637, 633, 470, 412	60	288, 308, 240, 225	60	6058, 5401, 4240, 3865	60	1501, 1307, 1094, 1004	59	4364, 3645, 2831, 2626	60	6163, 4881, 3893, 4083	60	271, 225, 180, 180	60	1189, 1071, 833, 807	60
Ex2S	852, 854, 899, 773	61	1083, 884, 879, 897	60	1735, 1601, 1582, 1488	60	1745, 1381, 1461, 1543	60	5862, 4520, 5031, 4688	61	2304, 1926, 1778, 1838	60	2363, 1996, 1978, 1818	60	2892, 2556, 2348, 2358	60
Ex2A	149, 167, 151, 159	61	1071, 1064, 916, 944	61	1756, 1963, 1716, 1648	58	3113, 3028, 2539, 2800	61	1303, 1374, 1187, 1180	61	3134, 2822, 2457, 2795	61	2533, 2717, 2391, 2493	61	213, 214, 187, 207	61
Ex2A-S	511, 823, 674, 532	57	3342, 3891, 3093, 2996	58	1371, 1836, 1463, 1223	57	456, 533, 406, 403	58	3058, 3833, 2990, 2632	57	940, 1046, 836, 789	57	2036, 2560, 2134, 1820	57	1351, 1670, 1237, 1164	57
Ex2A-A	268, 349, 368, 316	54	3825, 3713, 3951, 4005	55	1485, 1689, 1760, 1556	55	3379, 3089, 3264, 3445	56	3198, 3487, 3660, 3464	55	2408, 2272, 2375, 2451	55	2522, 2735, 2904, 2701	55	1138, 1107, 1121, 1167	56
Ex3S	755, 907, 732, 923	60	357, 423, 378, 450	60	2428, 3188, 2650, 2994	61	3421, 3749, 3161, 4166	60	2460, 2921, 2465, 2913	60	1911, 2052, 1748, 2222	60	910, 1075, 924, 1103	60	3584, 3985, 3395, 4239	60
Ex3A	1596, 1312, 1150, 1101	57	393, 352, 306, 285	57	969, 885, 778, 649	59	708, 598, 571, 540	57	3474, 2739, 2520, 2384	58	2393, 2073, 1949, 1853	58	2496, 1938, 1836, 1713	58	3793, 3300, 3215, 2954	57
Ex4S	437, 468, 418, 437	59	651, 687, 664, 656	58	4309, 3903, 3791, 3780	57	4487, 4263, 4172, 4092	58	1887, 1684, 1602, 1624	58	4613, 4189, 3995, 3935	57	2474, 2306, 2097, 2143	59	4258, 4076, 3844, 3799	59
Ex4A	617, 649, 519, 469	56	116, 119, 99, 92	54	103, 99, 86, 79	51	101, 103, 86, 81	56	206, 190, 153, 151	56	147, 150, 126, 113	53	526, 518, 416, 373	55	310, 320, 261, 233	56
Ex5S	1702, 1731, 1992, 1871	61	1754, 1583, 1873, 1816	60	3786, 3383, 4095, 3844	60	3583, 3044, 3612, 3464	61	2393, 2079, 2476, 2467	60	6441, 5137, 6261, 6446	60	5169, 4464, 5308, 5283	61	3846, 3268, 3966, 3918	61
Ex5A	657, 810, 747, 719	52	1808, 2030, 1728, 1807	50	320, 344, 297, 323	50	1060, 1066, 969, 1025	50	442, 450, 393, 432	50	359, 350, 298, 330	51	321, 344, 309, 309	51	598, 615, 546, 547	51

Tabell Supplement 8: Oversikt over signalstyrke og trace score for sekvenseringsprimere fra pasientprøvene D1593-19 til D2360-18. S=Signal faktor. TS= Trace Score (Average Pread value / Datakvalitet). PAR= Parametre. PAS= Pasienter. Ex=Eksjon. Ruter markert med beige indikerer sekvenser med for lav signalstyrke/trace score som krever visuell vurdering

Pas:	D1593-19		D1345-19		D1107-19		D803-19		D331-19		D2534-18		D2360-18	
PAR:	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS	S (RFU)	TS
Ex1S	619, 680, 771, 859	58	1530, 1691, 1915, 2066	57	807, 805, 875, 957	58	1308, 1500, 1648, 1864	57	757, 729, 804, 886	58	1562, 1495, 1720, 1968	58	1495, 1583, 1782, 2027	59
Ex1A	5721, 5159, 4012, 3701	60	1193, 1265, 1030, 944	60	6395, 5415, 4293, 4055	60	4893, 4059, 3226, 3138	60	4514, 3746, 2935, 2811	60	6383, 5214, 4099, 4123	61	2335, 2021, 1523, 1558	60
Ex2S	4016, 3626, 3834, 3398	60	1213, 976, 964, 995	60	1606, 1493, 1437, 1389	60	3054, 2645, 2521, 2408	60	2321, 2045, 1928, 1876	61	2077, 1878, 1718, 1671	60	1482, 1322, 1287, 1138	61
Ex2A	1801, 2261, 1986, 1930	60	1374, 1361, 1164, 1219	61	2290, 2541, 2107, 2174	59	2503, 2416, 2043, 2369	61	811, 806, 697, 724	61	3481, 3133, 2697, 3018	61	377, 351, 295, 347	61
Ex2A-S	717, 1110, 900, 704	57	2350, 2982, 2351, 2190	57	1756, 2385, 1893, 1702	57	3305, 3599, 2744, 2774	57	1849, 2394, 1771, 1646	57	1260, 1460, 1150, 1074	57	182, 230, 190, 172	57
Ex2A-A	851, 1128, 1184, 1002	55	3143, 3109, 3275, 3290	56	1872, 2096, 2160, 2057	55	2951, 2703, 2810, 2862	56	1466, 1458, 1565, 1648	56	3341, 3237, 3419, 3399	56	1035, 1098, 1132, 1079	55
Ex3S	532, 656, 545, 579	60	2385, 2765, 2435, 3006	60	2285, 2878, 2440, 2772	60	2812, 3066, 2674, 3365	61	2614, 2900, 2484, 3386	60	1190, 1358, 1148, 1458	60	2400, 2812, 2441, 2855	60
Ex3A	3045, 2572, 2306, 2072	58	1548, 1267, 1219, 1127	58	2380, 2058, 1803, 1723	58	2065, 1797, 1621, 1537	58	1466, 1458, 1565, 1648	59	1974, 1728, 1701, 1572	59	1517, 1155, 1054, 1022	58
Ex4S	3190, 3254, 2999, 3050	58	2673, 2732, 2667, 2738	58	2920, 2537, 2367, 2504	58	2681, 2644, 2389, 2526	58	2760, 2365, 2253, 2295	59	2661, 2660, 2555, 2503	59	3935, 3676, 3376, 3525	57
Ex4A	469, 481, 389, 361	56	963, 942, 794, 724	56	142, 135, 116, 107	56	1575, 1547, 1260, 1122	56	2173, 2149, 1958, 2002	50	1350, 1312, 1042, 992	56	2164, 1988, 1578, 1471	54
Ex5S	1509, 1505, 1896, 1623	61	4226, 3977, 4589, 4471	61	2317, 2064, 2453, 2347	61	3634, 3063, 3692, 3465	61	4910, 4075, 5004, 4622	60	5069, 4161, 5162, 4871	61	3359, 3206, 3652, 3293	60
Ex5A	1208, 1573, 1377, 1357	49	1568, 1686, 1461, 1604	51	1563, 1658, 1466, 1583	52	2000, 2050, 1778, 1941	51	973, 994, 869, 965	53	1093, 1055, 956, 986	52	289, 323, 280, 290	50


Tabell Supplement 9: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser.

Ekson 1:	chr7:g.142457343 (hg19) PRSS1:NM_002769:exon1:c.C8T: p.P3L <i>Missense=substusjon. Pseudogen/ homolog sekvens: Basen T er kun tilstede i reads med lavere MappingQ enn 60 og ikke i det hele tatt i reads med MappingQ på 60.</i>	chr7:g.142457347 (hg19) PRSS1:NM_002769:exon1:c.C12T:p.L4 L <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen T er kun tilstede i reads med lavere MappingQ enn 60 og ikke i det hele tatt i reads med MappingQ på 60.</i>	chr7:g.142457365 (hg19) PRSS1:NM_002769:exon1: c.G30T:p.V10V <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen T er kun tilstede i reads med lavere MappingQ enn 60 og ikke i det hele tatt i reads med MappingQ på 60.</i>	chr7:g.142457375 (hg19) PRSS1:NM_002769:exon1:c.C40 G:p.L14V <i>Missense. Fra Alamut: CLIN_ uncertain_significance. Pseudogen/ homolog sekvens: Basen G er kun tilstede i reads med lavere MappingQ enn 60 og ikke i det hele tatt i reads med MappingQ på 60.</i>	
Pasientpr øver:					
	D2072-19	88% C og 12% T av 178.	89% C og 11% T av 193.	87% G og 12% T av 192.	88% C og 12% G av 191.
	D1857-19	77% C og 23% T av 61.	79% C og 21% T av 68.	81% G og 19% T av 72.	83% C og 17% G av 72.
	D1593-19	76% C og 24% T av 149.	77% C og 23% T av 158.	80% G og 20% T av 178.	80% C og 20% G av 179.
	D1345-19	69% C og 31% T av 171.	72% C og 28% T av 190.	73% G og 26% T av 205.	74% C og 26% G av 200.
	D1107-19	90% C og 10% T av 117.	91% C og 9% T av 136.		

Tabell Supplement 10: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Sykdomsassosiert variant er merket med CLIN_pathogenic i Alamut.

Ekson 2:	chr7:g.142458409 (hg19) PRSS1:NM_002769:exon2:c.C44 G:p.A15G	chr7:g.142458412 (hg19) PRSS1:NM_002769:exon2:c.C47T :p.A16V	chr7:g.142458414 (hg19) PRSS1:NM_002769:exon2:c.C49T :p.P17S	chr7:g.142458437 (hg19) PRSS1:NM_002769:exon2:c.C72T :p.I24I
	<p>Missense=substitusjon. Pseudogen/homolog sekvens: Basen G er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</p> <p>✓ Coverage ✓ Reads</p> <p>TCCACTCCAGTTGCTGCCCCCTT</p> <p>G T T</p>	<p>Missense=substitusjon. Fra Alamut: CLIN_pathogenic. Pseudogen/homolog sekvens: Basen T er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</p> <p>✓ Reads</p> <p>TTGCTGCCCCCTTTGATGATGA</p> <p>G T T</p>	<p>Missense=substitusjon. Pseudogen/homolog sekvens: Basen T er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</p> <p>✓ Coverage ✓ Reads</p> <p>TTGCTGCCCCCTTTGATGATGA</p> <p>G T T</p>	<p>Missense=substitusjon. Pseudogen/homolog sekvens: Basen T er tilstede i reads med en lav allelefraksjon i fordeling 85/15 - 80/20.</p> <p>✓ Coverage ✓ Reads</p> <p>TGATGACAAGATCGTTGGGGGCT</p> <p>T T T T</p>
Pasientprøver:				
D2031-19	91% C og 9% G av 359.	91% C og 9% T av 359.	92% C og 8% T av 365.	90% C og 10% T av 373.
D1857-19	89% C og 11% G av 105.	89% C og 11% T av 105.	88% C og 11% T av 107.	86% C og 14% T av 115.
D1801-19	83% C og 17% G av 241.	84% C og 16% T av 243.	81% C og 19% T av 253.	75% C og 25% T av 270.
D1345-19				79% C og 20% T av 266.

Tabell Supplement 11: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For varianten er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Sykdomsassosiert variant er merket med CLIN_pathogenic i Alamut.

<p>Ekson 2:</p>	<p>chr7:g.142458451 (hg19) PRSS1:NM_002769:exon2:c.A86T:p.N29I <i>Missense=substitusjon. Fra Alamut: CLIN_pathogenic.</i> <i>ExAC har sekvenseringsfeil her. Pseudogen/ homolog sekvens: Basen T er kun til stede i reads med lavere MappingQ enn 60, mens de med A har 60. Allelefraksjon i fordeling 33/67.</i></p> 																						
<p>Pasientprøver:</p>	<table border="1"> <tbody> <tr> <td>D2181-19</td> <td>79% A og 21% T av 320.</td> </tr> <tr> <td>D2072-19</td> <td>71% A og 29% T av 229.</td> </tr> <tr> <td>D2031-19</td> <td>82% A og 18% T av 389.</td> </tr> <tr> <td>D1857-19</td> <td>73% A og 27% T av 120.</td> </tr> <tr> <td>D1593-19</td> <td>74% A og 26% T av 222.</td> </tr> <tr> <td>D1345-19</td> <td>74% A og 26% T av 302.</td> </tr> <tr> <td>D1107-19</td> <td>81% A og 19% T av 68.</td> </tr> <tr> <td>D803-19</td> <td>82% A og 18% T av 73.</td> </tr> <tr> <td>D331-19</td> <td>75% A og 25% T av 79.</td> </tr> <tr> <td>D2534-18</td> <td>73% A og 27% T av 67.</td> </tr> <tr> <td>D2360-18</td> <td>71% A og 29% T av 95.</td> </tr> </tbody> </table>	D2181-19	79% A og 21% T av 320.	D2072-19	71% A og 29% T av 229.	D2031-19	82% A og 18% T av 389.	D1857-19	73% A og 27% T av 120.	D1593-19	74% A og 26% T av 222.	D1345-19	74% A og 26% T av 302.	D1107-19	81% A og 19% T av 68.	D803-19	82% A og 18% T av 73.	D331-19	75% A og 25% T av 79.	D2534-18	73% A og 27% T av 67.	D2360-18	71% A og 29% T av 95.
D2181-19	79% A og 21% T av 320.																						
D2072-19	71% A og 29% T av 229.																						
D2031-19	82% A og 18% T av 389.																						
D1857-19	73% A og 27% T av 120.																						
D1593-19	74% A og 26% T av 222.																						
D1345-19	74% A og 26% T av 302.																						
D1107-19	81% A og 19% T av 68.																						
D803-19	82% A og 18% T av 73.																						
D331-19	75% A og 25% T av 79.																						
D2534-18	73% A og 27% T av 67.																						
D2360-18	71% A og 29% T av 95.																						

Tabell Supplement 12: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Sykdomsassosiert variant er merket med CLIN_pathogenic, og benign variant er merket med CLIN_benign i Alamut.

Ekson 2:	chr7:g.142458511 (hg19) PRSS1:NM_002769:exon2:c.G146 T:p.G49V <i>Missense=substitusjon. Fra Alamut: CLIN_benign. Pseudogen/ homolog sekvens: Basen T er tilstede i reads med lavere MappingQ enn 60 med lav allelefraksjon i fordeling 90/10.</i>	chr7:g.142458523 (hg19) PRSS1:NM_002769:exon2:c.T158A :p.I53N <i>Missense=substitusjon. Pseudogen/ homolog sekvens: Basen A er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>	chr7:g.142458526 (hg19) PRSS1:NM_002769 :exon2:c.A161G;p.N54S <i>Missense=substitusjon. Fra Alamut: CLIN_pathogenic. Pseudogen/ homolog sekvens: Basen G er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>	chr7:g.142458527 (hg19) PRSS1:NM_002769 :exon2:c.C162G;p.N54K <i>Missense=substitusjon. Pseudogen/ homolog sekvens: Basen G er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>
Pasientprøver:				
D2181-19	91% G og 9% T av 291.	5% A og 95% T av 251.	94% A og 6% G av 248.	95% C og 5% G av 244.
D2031-19	87% G og 13% T av 352.	9% A og 91% T av 301.	90% A og 10% G av 296.	91% C og 9% G av 289.
D1857-19	81% G og 19% T av 95.	8% A og 92% T av 78.	Lavt antall. 93% A og 7% G av 76.	Lavt antall. 95% C og 5% G av 76.
D1801-19	76% G og 24% T av 250.	16% A og 84% T av 191.	85% A og 15% G av 181.	84% C og 16% G av 178.
D1345-19	76% G og 24% T av 262.	16% A og 84% T av 221.	80% A og 19% G av 217.	83% C og 16% G av 214.


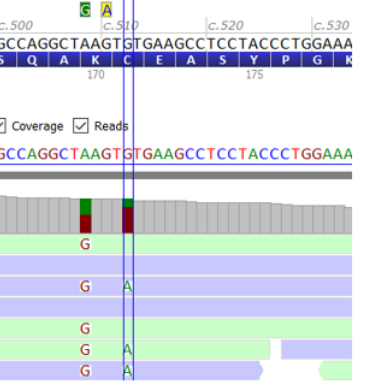

Tabell Supplement 13: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Benign variant er merket med CLIN_benign i Alamut.

Ekson 2:	chr7:g.142458531 (hg19) PRSS1:NM_002769 :exon2:c.C166T:p.Q56X <i>Missense - STOPGAIN. Pseudogen/ homolog sekvens: Basen T er tilstede i reads med en lav allelefraksjon i fordeling 90/10 - 80/20.</i>	chr7:g.142458542 (hg19) PRSS1:NM_002769 :exon2:c.A177G:p.V59V <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen G er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>	chr7:g.142458550 (hg19) PRSS1:NM_002769 :exon2:c.G185C:p.G62A <i>Missense=substitusjon. Pseudogen/ homolog sekvens: Basen C er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>	chr7:g.142458551 (hg19) PRSS1:NM_002769 :exon2:c.C186T:p.G62G <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen T er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>	chr7:g.142458566 (hg19) NM_002769:exon2:c.200+1 G>A <i>Spleising. Fra Alamut: CLIN_benign. Pseudogen/ homolog sekvens: Basen A er tilstede i reads med lavere MappingQ enn 60 med en lav allelefraksjon i fordeling 90/10.</i>	
Pasientprøver:						
	D2031-19	91% C og 9% T av 279.	89% A og 11% G av 244.	9% C og 91% G av 206.	92% C og 8% T av 203.	
	D1857-19	Lavt ant. 95% C og 5% T av 73.				
	D1801-19	84% C og 16% T av 176.	84% A og 16% G av 157.	13% C og 87% G av 124.	89% C og 11% T av 122.	14% A og 86% G av 86.

Tabell Supplement 14: Sekvensvariant funnet i pasientprøver med LR PCR metode i analyseprogrammet SeqPilot, sammenlignet med variant funnet i NGS. For varianten er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Vi finner varianten igjen i SeqPilot som homo- eller heterozygot, og i NGS med henholdsvis allelefraksjon på 40% og 20%.

Ekson 4:	chr7:g.142460313 (hg19) PRSS1:NM_002769:exon4:c.T486C:p.D162D <i>Silent=Synonym. Fra Alamut: CLIN_benign. Dette ser ut som kan være en SNP som opptrer både i PRSS1 og pseudogen/homolog sekvens. Fordeling 50/50 og 25/75 finnes både i sekvenser med MappingQ 0 og 60.</i> c.470 c.480 c.490 c.500 AGCTGCAGTGCCTGGATGCTCCTGTGCTGAGCCA E L Q C L D A P V L S Q 160 165 rgets <input checked="" type="checkbox"/> Coverage <input checked="" type="checkbox"/> Reads AGCTGCAGTGCCTGGATGCTCCTGTGCTGAGCCA Pasientprøver:	chr7:g.142460313 (hg19) PRSS1:NM_002769:exon4:c.T486C:p.D162D <i>Silent=Synonym.</i> Intron 3 Exon 4 L D A P L D A P L D A P C C T G G A T G C T C C C C C T G G A T G C T C C C C C T G G A C G C T C C C C C T G G A Y G C T C C C C C T G G A C G C T C C C C C T G G A Y G C T C C C 480 485 490 480 485 490 PRSS1 / Exon 4 PRSS1 / Exon 4 (Homozygot) (Heterozygot)
D293-20	NGS: 40% C og 60% T av 2108.	Homozygot i SeqPilot.
D264-20	23% C og 77% T av 1794.	Heterozygot i SeqPilot.
D2181-19	21% C og 79% T av 619.	Heterozygot i SeqPilot.
D2031-19	22% C og 78% T av 600.	Heterozygot i SeqPilot.
D1907-19	21% C og 79% T av 445.	Heterozygot i SeqPilot.
D1857-19	39% C og 61% T av 222.	Homozygot i SeqPilot.
D1801-19	33% C og 67% T av 396.	Homozygot i SeqPilot.
D1345-19	38% C og 62% T av 552.	Homozygot i SeqPilot.
D1107-19	29% C og 71% T av 272.	Homozygot i SeqPilot.
D2534-18	39% C og 60% T av 323.	Homozygot i SeqPilot.
D2360-18	25% C og 74% T av 292.	Heterozygot i SeqPilot.

Tabell Supplement 15: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Benigne varianter er merket med CLIN_benign i Alamut.

Ekson 4:	<p>chr7:g.142460335 (hg19) PRSS1:NM_002769:exon4:c.A508G:p.K170E Missense=substitusjon. Fra Alamut: <i>CLIN_benign</i>. Pseudogen/ homolog sekv: Basen G er tilstede i reads med MappingQ 0 og 60 i fordeling 50/50 - 25/75 G/A.</p>	<p>chr7:g.142460339 (hg19) PRSS1:NM_002769:exon4:c.G512A:p.C171Y Missense=substitusjon. Pseudogen/ homolog sekvns: Basen A er tilstede i reads med lav allelefraksjon i fordeling 25/75 A/G.</p>	<p>chr7:g.142460369 (hg19) PRSS1:NM_002769:exon4:c.G542A:p.S181N Missense=substitusjon. Fra Alamut: <i>CLIN_benign</i>. Pseudogen/ homolog sekvns: Basen A er tilstede i reads med MappingQ 0 og 60 i fordeling 25/75 A/G.</p>
Pasientprøver:			
D293-20	46% A og 54% G av 1866.	25% A og 75% G av 1854.	42% A og 58% G av 1413.
D264-20	53% A og 47% G av 1561.		25% A og 75% G av 1501
D2181-19	49% A og 51% G av 490.		37% A og 63% G av 428.
D2072-19	59% A og 41% G av 256.		29% A og 71% G av 242.
D2031-19	47% A og 53% G av 502.		40% A og 60% G av 429.
D1907-19	51% A og 49% G av 382.		41% A og 58% G av 349.
D1857-19	36% A og 64% G av 151.	21% A og 79% G av 161.	51% A og 49% G av 146.
D1801-19	41% A og 59% G av 269.	24% A og 76% G av 271.	48% A og 52% G av 235.
D1593-19	46% A og 54% G av 273.		39% A og 61% G av 241.
D1345-19	38% A og 62% G av 406.	23% A og 77% G av 413.	47% A og 53% G av 348.
D1107-19	31% A og 69% G av 156.	38% A og 62% G av 154.	48% A og 52% G av 95.
D803-19	58% A og 42% G av 106.		29% A og 71% G av 95.
D331-19	56% A og 44% G av 165.		23% A og 77% G av 119.
D2534-18	44% A og 56% G av 187.	30% A og 70% G av 181.	39% A og 61% G av 149.
D2360-18	43% A og 57% G av 175		49% A og 51% G av 141.

Tabell Supplement 16: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser.

Ekson 4:	chr7:g.142460374 (hg19) PRSS1:NM_002769:exon4:c.A547G:p.M183V <i>Missense=substitusjon. Pseudogen/ homolog sekvens: Basen G er tilstede i reads med en lav allelefraksjon i fordeling 20/80.</i>	chr7:g.142460379 (hg19) PRSS1:NM_002769:exon4:c.C552T:p.F184F <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen T er tilstede i reads med en lav allelefraksjon i fordeling 20/80.</i>	chr7:g.142460388 (hg19) PRSS1:NM_002769:exon4:c.C561T:p.G187G <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen T er tilstede i reads med en lav allelefraksjon i fordeling 20/80.</i>
Pasientprøver:			
D293-20	81% A og 19% G av 1371.	82% C og 18% T av 1313	84% C og 16% T av 1181
D2181-19	90% A og 10% G av 424.	90% C og 10% T av 415.	90% C og 10% T av 362.
D2031-19	89% A og 11% G av 422.	90% C og 10% T av 414.	89% C og 11% T av 368.
D1857-19	88% A og 12% G av 147.	88% C og 12% T av 142.	87% C og 13% T av 124.
D1801-19	85% A og 15% G av 233.	85% C og 15% T av 229.	85% C og 15% T av 194.
D1345-19	83% A og 17% G av 350.	84% C og 16% T av 340.	83% C og 17% T av 308.
D1107-19	69% A og 31% G av 95.	74% C og 26% T av 95.	82% C og 18% T av 98.




Tabell Supplement 17: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Benign variant er merket med CLIN_benign i Alamut.

Ekson 4:	chr7:g.142460394 (hg19) PRSS1:NM_002769:exon4:c.T567C:p.L189L <i>Silent=synonym. Fra Alamut: CLIN_likely_benign.</i> <i>Pseudogen/ homolog sekvens: Basen C er tilstede i reads med både MappingQ 0 og 60 i fordeling 25/75 eller 10/90 C/T.</i>	chr7:g.142460412 (hg19) PRSS1:NM_002769:exon4:c.A585C:p.S195S <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen C er tilstede i hovedsak i reads med MQ<60 i fordeling 25/75 eller 10/90 C/A.</i>	chr7:g.142460415 (hg19) PRSS1:NM_002769:exon4:c.T588C:p.C196C <i>Silent=synonym. Pseudogen/ homolog sekvens: Basen C er tilstede i hovedsak i reads med MQ<60 i fordeling 25/75 eller 10/90 C/A.</i>	
Pasientprøver:				
	D293-20	19% C og 81% T av 1091.	73% A og 27% C av 863.	26% C og 74% T av 836.
	D264-20	16% C og 84% T av 961.	82% A og 18% C av 800.	17% C og 83% T av 780
	D2181-19	15% C og 85% T av 306.	82% A og 18% C av 213.	18% C og 81% T av 209.
	D2072-19	21% C og 79% T av 184.	83% A og 17% C av 148.	16% C og 84% T av 146.
	D2031-19	20% C og 80% T av 305.	75% A og 25% C av 241.	25% C og 75% T av 240.
	D1907-19	23% C og 77% T av 265	72% A og 27% C av 219.	28% C og 72% T av 212.
	D1857-19	23% C og 77% T av 91.	75% A og 25% C av 64.	23% C og 77% T av 60.
	D1801-19	26% C og 74% T av 148.	81% A og 19% C av 96.	21% C og 79% T av 94.
	D1593-19	26% C og 74% T av 175.	76% A og 24% C av 145.	24% C og 76% T av 143.
	D1345-19	19% C og 81% T av 262.	74% A og 26% C av 195.	28% C og 72% T av 193.
	D803-19	12% C og 88% T av 114.	88% A og 12% C av 111.	11% C og 88% T av 109.
D2360-18	15% C og 85% T av 172.	89% A og 11% C av 175.		


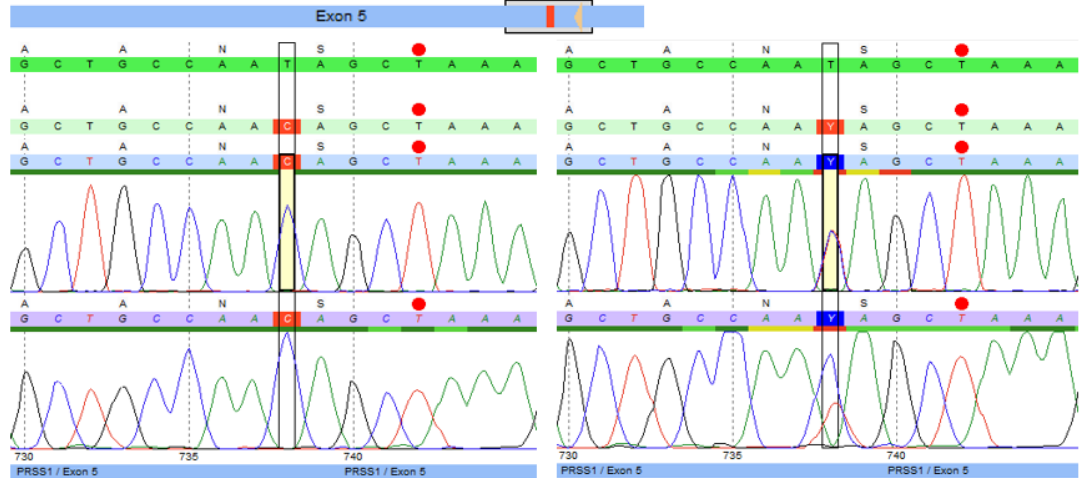
Tabell Supplement 18: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Benign variant er merket med CLIN_benign i Alamut.

Ekson 5:	chr7:g.142460744 (hg19) PRSS1:NM_002769:exon5:c.G617C:p.C206S Missense=Substitusjon. Pseudogen/ homolog sekvens: Basen C er tilstede i reads med MappingQ <60 og =60 i fordeling 50/50 C/G.	chr7:g.142460752 (hg19) PRSS1:NM_002769:exon5:c.C625G:p.Q209E Missense=Substitusjon. Pseudogen/ homolog sekvens: Basen C er tilstede i reads med MappingQ <60 og =60 i fordeling 50/50 C/G.	chr7:g.142460764 (hg19) PRSS1:NM_002769:exon5:c.G637A:p.V213I Missense=Substitusjon. Fra Alamut: CLIN_benign. Pseudogen/ homolog sekvens: Basen A er tilstede i reads med MappingQ <60 og =60 i fordeling 40/60 A/G.	
Pasientpr øver:				
	D2181-19	47% C og 53% G av 393.	55% C og 45% G av 348.	43% A og 57% G av 288.
	D2031-19	45% C og 54% G av 415.	56% C og 44% G av 374.	42% A og 58% G av 299.
	D1907-19	40% C og 60% G av 354.	63% C og 37% G av 322.	37% A og 63% G av 270.
	D1857-19	45% C og 55% G av 139.	55% C og 45% G av 127.	46% A og 54% G av 95.
	D1801-19	37% C og 63% G av 273.	65% C og 35% G av 237.	35% A og 65% G av 179.
	D1593-19	51% C og 49% G av 319.	55% C og 45% G av 272.	46% A og 54% G av 200.
	D1345-19	36% C og 64% G av 325.	66% C og 34% G av 298.	32% A og 68% G av 241.
	D1107-19	43% C og 57% G av 146.	59% C og 41% G av 131.	41% A og 59% G av 91.
	D803-19	50% C og 50% G av 151.	55% C og 45% G av 128.	41% A og 59% G av 100.
	D331-19	59% C og 41% G av 259.	41% C og 58% G av 227.	58% A og 42% G av 170.
	D2534-18	45% C og 55% G av 201.	55% C og 45% G av 166.	41% A og 59% G av 126.
D2360-18	43% C og 57% G av 212.	60% C og 40% G av 169.	48% A og 52% G av 127.	

Tabell Supplement 19: Varianter funnet i PRSS1-genet ved NGS analyse av pasientprøver er markert i rødt. For variantene er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser.

Ekson 5:	c.651_652delinsCT Usikker signifikans, men linket mot arvelig pankreatitt. Fordeling 35/65 heterozygot fra Pseudogen/homolog sekvens: Basen er i hovedsak tilstede i reads med MappingQ <60.	chr7:g.142460778 (hg19) PRSS1:NM_002769:exon5:c.T651C:p.G217G Silent=synonym.	chr7:g.142460801 (hg19) PRSS1:NM_002769:exon5:c.A674G:p.K225R Missense=Substitusjon. Fra Alamut: CLIN_uncertain_significance. Pseudogen/homolog sekvens: Basen G er i hovedsak tilstede i reads med MappingQ <60 i fordeling 35/65.
Pasientprøver:			
D2181-19	c.651T>C: 40% C og 60% T av 246.	c.652G>T: 60% G og 40% T av 243.	74% A og 26% G av 175.
D2031-19	c.651T>C: 38% C og 62% T av 240.	c.652G>T: 63% G og 37% T av 235.	73% A og 23% G av 168.
D1907-19	c.651T>C: 34% C og 65% T av 244.	c.652G>T: 65% G og 35% T av 243.	75% A og 25% G av 201.
D1857-19	c.651T>C: 43% C og 57% T av 75.	c.652G>T: 57% G og 43% T av 75	82% A og 18% G av 57.
D1801-19	c.651T>C: 28% C og 72% T av 160.	c.652G>T: 73% G og 27% T av 158.	80% A og 20% G av 117.
D1593-19	c.651T>C: 42% C og 58% T av 173.	c.652G>T: 60% G og 40% T av 168.	75% A og 25% G av 135.
D1345-19	c.651T>C: 28% C og 72% T av 200.	c.652G>T: 73% G og 27% T av 196.	83% A og 17% G av 144.
D1107-19	c.651T>C: 38% C og 62% T av 73.	c.652G>T: 62% G og 38% T av 73.	
D803-19	c.651T>C: 38% C og 62% T av 79.	c.652G>T: 62% G og 38% T av 79.	80% A og 20% G av 60.
D331-19	c.651T>C: 57% C og 43% T av 122.	c.652G>T: 44% G og 56% T av 118.	61% A og 39% G av 92.
D2534-18	c.651T>C: 40% C og 60% T av 100.	c.652G>T: 60% G og 40% T av 100.	77% A og 23% G av 77.
D2360-18	c.651T>C: 41% C og 58% T av 101.	c.652G>T: 63% G og 37% T av 95.	81% A og 19% G av 73.

Tabell Supplement 20: Sekvensvariant funnet i pasientprøver med LR PCR metode i analyseprogrammet SeqPilot, sammenlignet med variant funnet i NGS. For varianten er allelefraksjon oppgitt i prosent (%) av dybden på sammenstilte sekvenser. Vi finner varianten igjen i SeqPilot som homo- eller heterozygot, og i NGS med henholdsvis allelefraksjon på 100% og 50%.

Ekson 5:	<p>chr7:g.142460865 (hg19) PRSS1:NM_002769:exon5:c.T738C:p.N246N</p> <p><i>Silent=synonym. Fra Alamut: CLIN_benign. Klasse 1. Dette ser ut som kan være en SNP som opptrer både i PRSS1 og pseudogen/homolog sekvens. Fordeling 50/50 og finnes både i sekvenser med MappingQ 0 og 60.</i></p> 	<p>Chr7:g.142460865 (hg19) PRSS1:NM_002769:exon5:c.T738C:p.N246N</p> <p><i>Silent=synonym.</i></p>  <p>(Homozygot) (Heterozygot)</p>
D293-20	NGS: 99% C av 996.	<u>Homozygot i SeqPilot.</u>
D264-20	50% C og 50% T av 883.	<u>Heterozygot i SeqPilot.</u>
D2181-19	53% C og 47% T av 113.	<u>Heterozygot i SeqPilot.</u>
D2031-19	67% C og 33% T av 91.	<u>Heterozygot i SeqPilot.</u>
D1907-19	59% C og 41% T av 109.	<u>Heterozygot i SeqPilot.</u>
D1857-19	100% C av 23.	<u>Homozygot i SeqPilot.</u>
D1801-19	100% C av 86.	<u>Homozygot i SeqPilot.</u>
D1345-19	100% C av 90.	<u>Homozygot i SeqPilot.</u>
D1107-19	100% C av 73.	<u>Homozygot i SeqPilot.</u>
D2534-18	100% C av 128.	<u>Homozygot i SeqPilot.</u>
D2360-18	43% C og 57% T av 125.	<u>Heterozygot i SeqPilot</u>