



Norges miljø- og  
biovitenskapelige  
universitet

**Masteroppgave 2020 60 stp**

Fakultet for kjemi, bioteknologi og matvitenskap (KBM)

# **Kan degradering i mRNA predikere alder på biologiske spor?**

**Celine Therese Gjengedal**

Bioteknologi - Molekylærbiologi



## Forord

Denne oppgaven markerer min avslutning av masterstudiet i bioteknologi ved Norges miljø- og biovitenskapelige universitet. Oppgaven har foregått på Oslo Universitetssykehus, Avdeling for rettsmedisinske fag – Seksjon for rettsgenetikk i straffesaker (RESP) i perioden august 2019 til mai 2020.

Jeg vil først og fremst takke min veileder ved RESP, Ane Elida Fonnelop, for svært god veiledning gjennom hele året. Jeg er så takknemlig for alle råd og tilbakemeldinger, samt tilgjengeligheten gjennom hele prosjektet.

Jeg vil takke Arne Roseth og Marianne Øygarden for god støtte og oppfølging på laboratoriet. Jeg vil også takke resten av gjengen på RESP for å inkludere meg i sosiale aktiviteter gjennom året.

Videre vil jeg takke min interne veileder ved NMBU, Thore Egeland, for oppfølging og verdifull veiledning gjennom prosjektet. Jeg vil også takke med-veileder ved NMBU, Hilde Brustad.

Avslutningsvis vil jeg takke min kjære samboer, familie og venner for støtte og oppmuntring under hele prosjektet.

Oslo, mai 2020

Celine Therese Gjengedal

## Forkortelser

µL	Mikroliter
bp	Basepar
cDNA	Komplementært DNA (engelsk: <i>complementary DNA</i> )
DNA	Deoksyribonukleinsyre
DNase	Deoksyribonuklease
dNTP	Nukleosidtrifosfat
ds	Dobbeltrådet (engelsk: <i>double stranded</i> )
EDTA	Etylendiamintetraacetat
mM	Millimolar
MPS	Massiv parallell sekvensering
mRNA	Budbringer RNA (engelsk: <i>messenger RNA</i> )
NaOH	Natriumhydroksid
ng	Nanogram
nM	Nanomolar
PCR	Polymerase kjedereaksjon
pg	Pikogram
PR2-buffer	Vaskebuffer
Read	Nukleotidsekvens som tilsvarer hele eller deler av et DNA-fragment
RESP	Avdeling for rettsmedisinske fag – Seksjon for rettsgenetikk i straffesaker
rfu	Relativ fluorescens enhet (engelsk: <i>relative fluorescens unit</i> )
RNA	Ribonukleinsyre
RNase	Ribonuklease
rpm	Omdreiningstall (engelsk: <i>rounds per minute</i> )
RTA	Real-time analysis
SBS	Sekvensering ved syntese (engelsk: <i>sequencing by synthesis</i> )
TE-buffer	Tris-EDTA buffer
UTR	Utranslatert region
UV	Ultrafiolett stråling

## Sammendrag

DNA-analyser er et viktig verktøy for å knytte et individ til et biologisk spor funnet på et åsted. DNA gir i hovedsak kun informasjon om identitet, og sier ikke noe om når eller hvordan et spor ble avsatt. RNA har tidligere blitt ansett som for følsomt for bruk i rettsgenetiske analyser. Senere studier har dokumentert at RNA er mer stabilt enn først antatt. Foreløpig er RNA forbundet med identifisering av celletype, men forskning på tidsbestemmelse foregår på området. Målet er at RNA kan supplere DNA-analyser ved å gi informasjon om celletype og til hvilken tid det biologiske materialet ble avsatt. Per dags dato er det ikke funnet en pålitelig metode for å predikere alder på et biologisk spor, men det er gjort flere lovende analyser på området. Målet med studien har vært å undersøke om det er sammenheng mellom tid og degraderingsrate i spesifikke mRNA-markører for spytt og blod. Det var også ønsket å undersøke om degraderingsraten mellom ulike mRNA varierte, samt om 5'-enden degraderte raskere enn 3'-enden.

I denne studien ble det samlet inn spytt- og blodprøver fra 16 frivillige deltakere. Det var til sammen 80 prøver som bestod av 35 spyttprøver og 45 blodprøver. Prøvene ble deretter plassert til lagring over en tidsperiode på 0 til 143 dager. RNA-ekstraksjon med fenol-kloroform ble utført med korte - og lengre tidsintervall for å observere endringer over tid. RNA-konsentrasjonen ble målt i hver prøve med Qubit™ Fluorometer. Videre ble 48 prøver inkludert for videre analyse. Det ble valgt ut 28 spesifikke mRNA-markører for spytt og blod. Deretter ble det forberedt et cDNA-bibliotek ved bruk av TruSeq® Targeted Expression kit. Avslutningsvis ble prøvene sekvensert på Illumina MiSeq® ved bruk av målrettet RNA-sekvensering. Sekvenseringsdataene ble benyttet for å undersøke antall «reads» for hver mRNA-markør og hvordan antallet forandret seg etter antall dager. Programvare for primær - og sekundær analyse av sekvenseringsdataene var forhåndsinstallert på maskinen, og ble benyttet for kvalitetskontroll og sammenstilling av sekvenser. Videre ble programvarene R og RStudio anvendt for datavisualisering og statistiske beregninger. Statistiske metoder som regresjonsanalyse og variansanalyse (ANOVA) ble benyttet i studien.

Resultatene for spyttprøvene viste en signifikant degradering over tid, som prinsipielt betyr at markørene i studien kan benyttes for å predikere tid. Markørene viste lignende degraderingsmønster, samt hastighet over tid. Dermed kunne det ikke utføres en normalisering, som er nødvendig for å kunne benyttes i rutinemessig arbeid. Det ble ikke funnet en forskjell i degraderingsrate mellom 3'- og 5'-enden som kunne benyttes for å

predikere tid. For blodprøvene var genet *hemoglobin-beta* (HBB) den eneste markøren som viste degradering over tid. Resterende blodmarkører i studien var for stabile. Det ble observert forskjell i degradering mellom to gen, *hemoglobin-beta* (HBB) og *hemoglobin-delta* (HBD), hvor ratioen korrelerte med tid. Dermed kan denne ratioen anvendes for å predikere tid. Det ble ikke observert ulik degraderingsrate mellom 3'- og 5'- enden for blodmarkørene.

Konklusjonen i denne studien ble at ratioen mellom to gener (HBB og HBD) som degraderer jevnt, men ulikt, kan benyttes for prediksjon av tid. Det var en signifikant degradering over tid for spyttmarkørene, men ettersom normalisering ikke ble foretatt, kan ikke modellen benyttes for andre datasett. Markørene bør evalueres i videre studier med andre primere, kortere tidsintervall og flere paralleller for hver deltaker. Det var ingen forskjell i degraderingsrate mellom 3'- og 5'- enden for markørene.

## Abstract

DNA-analysis is a valuable tool to link an individual to a biological stain detected at a crime scene. DNA provide information about identity and say nothing about when or how a stain was deposited. Previously, it has been assumed that RNA is too unstable to be used in forensic genetic analysis. More recently, it has been shown that RNA is more stable than initially thought. Currently, RNA is associated with body fluid identification, but research on estimation time since deposition is ongoing. The goal is to use RNA as a supplement for DNA analysis by providing information about cellular origin and at what time the biological material was deposited. To date, no reliable method has been found for predicting age on a biological stain, but several promising methods has been reported in this area. The aim of the study has been to demonstrate whether there is a relationship between time and degradation rate in specific salivary and blood mRNA-markers. The study will also examine whether the degradation rate between different mRNAs varied, as well as whether the 5' - end of a transcript degrade faster than the 3' - end.

In this project, saliva and blood samples were collected from 16 volunteer participants. There was a total of 80 samples consisting of 35 saliva samples and 45 blood samples. The samples were then placed for storage over a time period of 0 to 143 days. RNA-extraction with phenol-chloroform was performed at short - and longer time intervals to observe changes over time. The RNA-concentration was quantified in each sample with the Qubit <sup>TM</sup> Fluorometer. Furthermore, 48 samples were included for further analysis. A total of 28 specific mRNA-markers for salivary and blood were selected. A cDNA library was then prepared using the TruSeq® Targeted Expression kit. Finally, the samples were sequenced on Illumina MiSeq® using targeted RNA sequencing. The sequencing data were used to examine the number of reads for each mRNA-marker and how the amount changed by number of days. Software for primary - and secondary analysis of the sequencing data was pre-installed on the machine and was used for quality control and alignment of sequences. Furthermore, software as R and RStudio were used for visualization of data and statistical calculations. Statistical methods as regression analysis and variance analysis (ANOVA) were used in the project.

The results for saliva samples showed a significant degradation over time, which in principle means that the markers in the study can be used to predict time. The markers showed similar pattern and rate in degradation over time. Therefore, a normalization could not be performed, which is necessary for use in routine work. There was not observed at different rate in

degradation between the 3' - and 5' - end, that could be used for prediction of time. For the blood samples, the *hemoglobin-beta* (HBB) gene was the only marker that showed degradation over time. The remaining blood markers was too stable to predict time in this project. It was observed differences in degradation between two genes, *hemoglobin-beta* (HBB) and *hemoglobin-delta* (HBD), where the ratio correlated with time. No different degradation rates were observed between the 3'- and 5'- end of the blood markers.

The conclusion in this study was that the ratio of two genes (HBB and HBD) that degrade evenly, but different, can be used to predict time. There was a significant degradation over time for the salivary markers, but since normalization was not performed, the model cannot be used for other datasets. The markers should be evaluated in further studies with different primers, shorter time intervals and multiple parallels for each participant. There was not observed a difference in degradation rate between the 3'- and 5'- end which could be used to predict time.



# Innholdsfortegnelse

Forord .....	i
Forkortelser .....	ii
Sammendrag .....	iii
Abstract .....	v
1 Introduksjon .....	1
1.1 RNA .....	2
1.1.1 Stabilitet og degraderingsmekanismer .....	4
1.2 mRNA som biomarkør .....	6
1.2.1 Cellespesifisitet .....	6
1.3 mRNA i rettsgenetiske analyser .....	7
1.3.1 Identifisering av kroppsvæske .....	7
1.3.2 Degradering og tidsbestemmelse .....	8
1.4 RNA ekstraksjon .....	9
1.5 Polymerase kjedereaksjon .....	11
1.5.1 Revers Transkripsjon .....	12
1.6 Kvalitativ og kvantitativ metode .....	13
1.6.1 Gel-elektroforese .....	13
1.6.2 RNA kvantitering .....	14
1.7 Massiv parallell sekvensering .....	14
1.7.1 Illumina MiSeq® .....	15
1.6.2 Signalanalyse .....	17
1.9 Formålet med oppgaven .....	19
2 Materialer og metoder .....	20
2.1 Prøveinnsamling .....	21
2.1.1 Etikk og godkjenninger .....	22
2.2 Forberedelse av prøver .....	23
2.2.1 RNA ekstraksjon .....	23
2.2.2 DNase behandling .....	24
2.2.3 Kvantitativ måling av RNA .....	25
2.3 Forberedelse av cDNA-bibliotek og sekvensering på Illumina MiSeq® .....	25
2.3.1 Valg av mRNA biomarkører .....	25
2.3.2 Tillaging av primere .....	26
2.3.3 cDNA-bibliotek .....	26

2.3.4	Kvalitativ måling av cDNA-bibliotek med gel-elektroforese .....	32
2.3.5	Sekvensering på Illumina MiSeq® .....	32
2.3.6	Signalanalyse.....	33
2.4	Dataanalyse.....	34
2.4.1	Statistisk programvare.....	34
2.4.2	Regresjonsanalyse .....	34
2.4.3	Variansanalyse .....	34
2.4.4	Normalisering.....	34
3	Resultater .....	35
3.1	RNA kvantitering .....	35
3.1.1	Spyttprøvene.....	35
3.1.2	Blodprøvene .....	37
3.2	cDNA kvantitering .....	39
3.3	Fragmentbestemmelse med gel-elektroforese .....	39
3.4	Sekvensering.....	40
3.4.1	Spyttprøvene.....	40
3.4.2	Blodprøvene .....	47
4	Diskusjon.....	54
4.1	RNA-kvantitering .....	54
4.1.1	Spyttprøvene.....	55
4.1.2	Blodprøvene .....	56
4.3	cDNA-bibliotek og gel-elektroforese .....	57
4.5	Sekvensering.....	58
4.5.1	Spyttprøvene.....	59
4.5.2	Blodprøvene .....	61
5	Konklusjon og videre arbeid .....	64
	Referanseliste .....	64
	Appendiks A: Reagenser og komponenter .....	70
	Appendiks B: mRNA-markører for spytt og blod.....	71
	Appendiks C: Resultater fra RNA-kvantitering .....	73
	Appendiks D: Utvalgte prøver for sekvensering.....	76
	Appendiks E: Gjennomsnittlig antall reads for hvert amplikon.....	77
	Appendiks F: R-koder .....	79
	Appendiks G-A: Reads per amplikon i hver blodprøve.....	82
	Appendiks G-B: Reads per amplikon i hver spyttprøve .....	84

# 1 Introduksjon

I straffesaker er DNA-analyser et betydningsfullt verktøy for å koble sammen et individ mot en sporprøve innsamlet i forbindelse med en kriminell hendelse. En DNA-analyse kan frembringe en unik DNA-profil (sett bort i fra eneggede tvillinger) som gir informasjon om hvem et biologisk spor stammer fra. En DNA-profil gir i hovedsak kun informasjon om identitet og sier ikke noe om når eller hvordan et spor ble avsatt. Det har tidligere vært antatt at RNA er for ustabil og følsomt til å kunne benyttes i rettsgenetiske analyser (1). I senere tid er dette blitt nærmere undersøkt og bevist at RNA er mer stabilt enn først antatt (2). Målet er at RNA kan supplere DNA-analyser ved å gi informasjon om celletype og til hvilken tid det biologiske materialet ble avsatt. Det gir mulighet for å ekskludere spor som ikke er innenfor aktuelt tidsrom. Dette er svært verdifullt for politiet og kan bidra til at flere kriminalsaker oppklares. Foreløpig er RNA forbundet med identifisering av celletype, men forskning på tidsbestemmelse foregår på området. Hensikten med denne studien er derfor å undersøke om avsatt biologisk materiale kan assosieres med tid. Dette utføres ved å observere degraderingsraten til ulike spesifikke mRNA-markører ved hjelp av massiv parallell sekvensering.

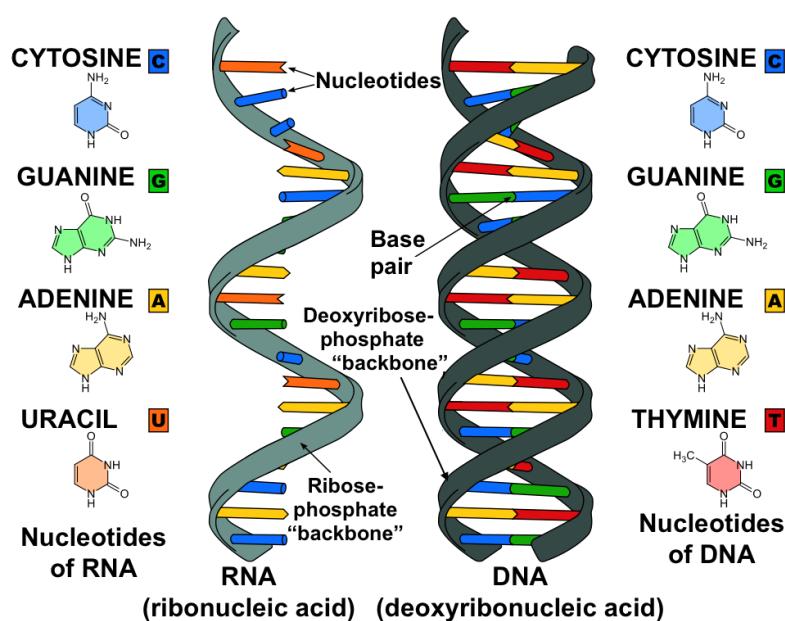
Denne studien har foregått på Oslo Universitetssykehus, Avdeling for rettsmedisinske fag – Seksjon for rettsgenetikk i straffesaker (RESP). Seksjonen har en funksjon innen rettsgenetisk sakkyndig virksomhet for rettsvesenet og politiet i Norge (3). Oppgavene innebærer å undersøke mulig biologisk materiale, utføre DNA-analyser, tolkning og sammenligning av DNA-profiler, samt overføring av profiler til politiets DNA-register (identitetsregisteret, etterforskningsregisteret og sporregisteret). Seksjonen gir også politiet veiledning for riktig innhenting og oppbevaring av materiale som skal undersøkes.

Oppgaven er inndelt i fem kapitler. Kapittel 1 inneholder en introduksjon av temaet, hvor biologisk teori og prinsippene for analysene beskrives, samt formålet med oppgaven. Kapittel 2 inneholder materialer og metoder benyttet i oppgaven. Kapittel 3 inneholder oversikt over resultatene, samt statistiske beregninger. Kapittel 4 tar for seg diskusjon av resultater sammenlignet med litteratur. Kapittel 5 inneholder en oppsummerende konklusjon og videre arbeid på området. Appendiks med relevant informasjon befinner seg til slutt i oppgaven.

## 1.1 RNA

RNA (ribonukleinsyre) er en nukleinsyre som har en essensiell rolle i cellens proteinproduksjon. RNA og DNA er kjemisk svært like. Begge er bygd opp av en kjede med fire forskjellige nukleotider bundet sammen av fosfodiesterbindinger (se **figur 1**). DNA inneholder nukleotidene; adenin, guanin, cytosin og tymin. RNA består av de samme nukleotidene, bortsett fra at tymin er erstattet med uracil. Alle nukleotidene har lik struktur, en fosfatgruppe bundet til et suktermolekyl (pentose). I RNA er suktermolekylet ribose, mens i DNA er det deoksyribose (4). I motsetning til DNA som er dobbeltrådet, er RNA vanligvis enkelttrådet. Denne strukturen gjør RNA mindre stabilt enn DNA.

Både RNA og DNA har ytterpunkter i strukturen som betegnes som 5'- og 3'-enden (4). Navnene defineres av posisjonen de har på suktermolekylet. Det refereres til femte og tredje karbonatom i deoksyribosen/ribose. En enkelt tråd av DNA/RNA vil alltid ha en ubundet 5'-fosfat på karbonet i den ene enden og en ubundet 3'-hydroksylgruppe på karbonet på motsatt ende.



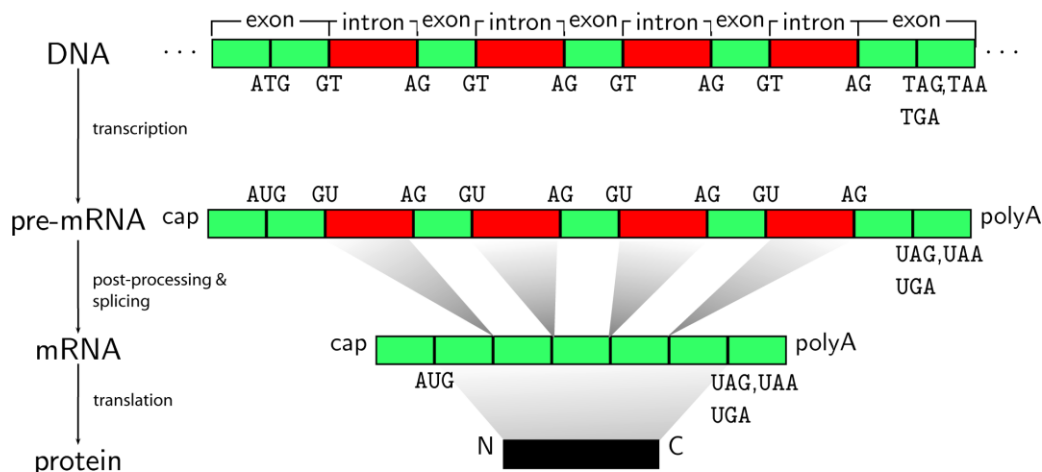
Figur 1: Oppbygning av RNA og DNA. Figur hentet fra: Wikimedia. Tilgjengelig fra:

[https://commons.wikimedia.org/wiki/File:Difference\\_DNA\\_RNA-EN.svg](https://commons.wikimedia.org/wiki/File:Difference_DNA_RNA-EN.svg) (Hentet dato: 03/01-2020)

DNA er det genetiske materialet som inneholder informasjon om aminosyresekvensen til proteinet. DNA transkriberes til flere typer RNA som spiller en viktig rolle i proteinsyntesen. Den totale mengden av RNA i en celle kalles transkriptomet. Transkriptomet kan deles inn i

grupper som har ulike funksjoner i cellen: ribosomalt RNA (rRNA), «budbringer» RNA (mRNA), transport RNA (tRNA), små nukleære RNA (snRNA) og mikro RNA (miRNA). Transkriptomet er unikt for hver celle ettersom det omfatter aktive gener til et gitt tidspunkt (5). I en celle utgjør mRNA 3 - 5 % av det totale RNA-innholdet og inneholder genetisk informasjon for proteinsyntese (5). mRNA betegnes som transkriptet (avskrift) av genet.

Livssyklusen til mRNA begynner med transkripsjon. Transkripsjon er prosessen hvor enzymet RNA-polymerase danner en komplementær kopi av DNA. Kopien kalles pre-mRNA og består av kodende (ekson) og ikke-kodende (intron) sekvenser. Deretter forekommer en prosessering av pre-mRNA, slik at det kan transporteres ut av cellekjernen for proteinsyntese på ribosomene. Under prosesseringen blir intron fjernet ved spleising, slik at ekson kan settes sammen (se **figur 2**). Deretter blir det satt på en kappe i 5'-enden av transkriptet. Kappen kalles 5'-cap og er et modifisert guanin nukleotid som beskytter transkriptet mot degradering av ribonukleaser (RNase). RNaser er enzymer som katalyserer nedbrytningen av RNA ved hydrolyse av fosfodiesterbindingene som binder ribonukleinsyrene sammen. Ribosomer gjenkjenner og binder 5'-cap på mRNA, slik at translasjon kan forekomme senere. Deretter utføres en polyadenylering av 3'-enden på transkriptet. Polyadenylering vil si at det tilsettes en hale av adenin nukleotider på omtrent 200 bp, som kalles poly(A) - hale. Halen bidrar til at transkriptet stabiliseres og kan transporteres ut av cellekjernen. Etter pre-mRNA er ferdig prosessert betegnes det for modent mRNA som kan sendes ut i cytosol for translasjon. Et modent mRNA består av ekson, stabiliserende 5' - cap, 3'- og 5' UTR (uttranslatert region), samt en poly(A) - hale. De uttranslaterte regionene befinner seg på hver side av den kodende sekvensen og har en funksjon i regulering av translasjonen. Etter translasjon fra modent mRNA til protein, blir transkriptet raskt degradert. Denne funksjonen er viktig for å regulere mengden genprodukt.



*Figur 2: Hovedsteg i proteinsyntesen. DNA transkriberes til pre-mRNA, etterfulgt av prosessering og spleising. Ikke-kodende sekvenser (intron) fjernes, slik at de kodende sekvensene (ekson) kan settes sammen. Deretter blir mRNA translateret til protein (6). (Hentet dato: 03/01-2020)*

### 1.1.1 Stabilitet og degraderingsmekanismer

Regulering av stabilitet hos mRNA er essensielt for kontroll av genuttrykk og produksjon av proteiner. Regulering av disse prosessene medfører at en celle raskt kan respondere til endringer i intra- og ekstracellulære stimuli (7). Stabilitet vil si hvor hurtig mRNA degraderes, som samsvarer med levetiden til transkriptet. Desto høyere degraderingsrate, jo lavere er stabiliteten til transkriptet. Det er stabiliteten til transkriptet som avgjør mengden protein som dannes ved translasjon (8). En høy stabilitet resulterer i større mengde protein.

Når modne mRNA transkripter transporteres ut i cytoplasma, er levetiden bestemt av spesifikke sekvenser. Levetiden relateres til funksjonen til proteinet som transkriptet koder for (5, 9). Sekvenser inneholder variasjon i affinitet for RNaser og proteiner som kan promotere eller inhibere degradering. Sekvenser med kjent effekt på stabiliteten til mRNA er elementer med et høyt innhold av basene uracil og adenin (AU-rike elementer) i 3' UTR. Disse elementene destabiliserer mRNA og forkorter levetiden (5).

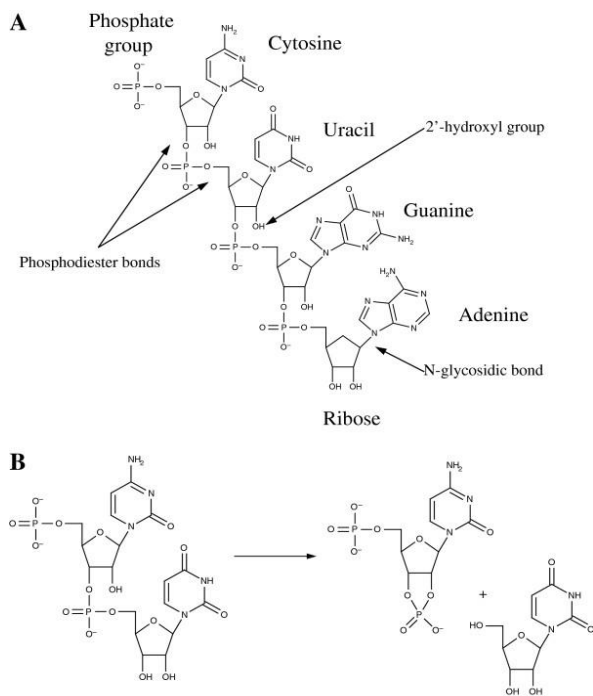
I en levende celle er degraderingsprosessen effektiv (5). Prosessen initieres på flere ulike måter, men hoveddelen av cytoplasmatiske mRNA degraderes via deadenylering (9, 10). Deadenylering innebærer gradvis forkorting av poly(A) - halen i 3'-enden. Dette forårsaker kløyving av 5'-cap. Dermed blir 5' -enden på mRNA eksponert og transkriptet degraderes i

5' - 3' retning av RNaser (endo - og eksoribonukleaser) som er til stede i cellen. Alternativt kan deadenylering føre til degradering fra 3' - 5' retning av et eksosom. Hvilken metode som foretrekkes for degradering avhenger av type mRNA. Degradering fungerer også som en kontrollmekanisme. Transkripter med premature stopp-kodon, mutasjon eller transkripsjonsfeil vil degraderes for å unngå dannelse av feillagde og potensielt farlige proteiner. Nonsense-mediert degradering er et eksempel på dette (9). Hovedfunksjonen til denne typen degradering er å forhindre feil i genuttrykk ved å eliminere transkripter som inneholder premature stopp-kodon. Deteksjon av slike kodon trigger mRNA degradering ved kløyving av 5'- cap strukturen (5'- decapping), fjerning av 3' poly(A) - hale eller kløyving av endonukleaser (11).

Mekanismer for degradering av RNA har blitt godt beskrevet *in vivo*, men degradering i avsatt biologisk materiale gjenstår fortsatt å undersøkes grundigere (12). Det er et dokumentert fenomen at RNA degraderes raskere enn DNA (13). Ustabilitet er hovedgrunnen til at RNA ikke har vært ansett som egnet for rettsgenetiske analyser. RNA er følsomt og dermed utsatt for rask degradering av RNaser i cellen og i miljøet rundt (1). Degraderingsprosessen avhenger av type biologisk spor og hvilke miljø det finnes i. I senere tider er det bevist at RNA er mer stabilt *ex vivo* enn først antatt (14). Dette skyldes inaktivering av RNaser. Inaktivering gjør at RNA *ex vivo* er mer stabilt enn *in vivo*. Dette gjelder for eksempel prøver som er tørket eller frosne (13). I tørket biologisk materiale er prøvene dehydrert. Under dehydrerte forhold, reduseres RNase aktiviteten signifikant. Stabiliteten *ex vivo* påvirkes av ulike faktorer som struktur, pH, tilstedeværelse av metall kationer, UV-lys, fuktighet og tilgang på oksygen (13, 15). Derfor vil RNA degraderingen under slike forhold forekomme hovedsakelig av fysiske og kjemiske faktorer (13). Ved fuktig materiale vil degraderingen forekomme raskt, ettersom RNasene forblir aktive.

En viktig faktor som påvirker RNA *ex vivo*, er strukturen (se **figur 3A**). Den enkelttrådig strukturen er nødvendig for å regulere proteinproduksjonen i cellen, ettersom mRNA både må raskt produseres og degraderes. Strukturen gjør at basene ligger ubeskyttet fra miljøet rundt, i motsetning til DNA hvor basene ligger beskyttet i dobbeltspiralen. Dette gjør at DNA er stabilt og tåler mer før det ødelegges. En annen strukturell grunn til degradering er hydroksylgruppen (2'- OH) på 2' posisjonen til ribosen som gjør RNA-molekylet mer følsomt for spontan hydrolyse enn DNA. Hydroksylgruppen er til stede i fleksible regioner i RNA som gjør at den kjemisk kan angripe den nærliggende fosfodiesterbindingen og kløve ryggraden i molekylet (se **figur 3B**) (13). En studie gjort av Fu og Allen viste at 5'- enden i

mRNA degraderte raskere enn 3'-enden (15). For et husholdningsgen foregikk degraderingen raskt i 5'-enden i en periode på 6-8 uker og deretter flatet kurven ut etter 1 års lagring. Det er mulig at sekundær struktur eller mulige kjemiske modifiseringer av nukleotider i 3' UTR i mRNA er grunnen til forskjellen i degraderingskinetikken (15). Det er vist at degraderingsprosessen varierer mellom ulike mRNA transkripter. Variasjonen i denne prosessen kan forårsakes av faktorer som ikke er kjent (16).



*Figur 3: A) Den kjemiske strukturen til RNA, hvor 2-hydroksyl gruppen i ryggraden gjør at RNA lettere degraderes enn DNA. B) Illustrerer hydrolyse av fosfodiesterbindingen i RNA (13). (Hentet dato: 03/10-2019)*

## 1.2 mRNA som biomarkør

### 1.2.1 Cellespesifisitet

En biomarkør er et molekyl som kan påvises i en celle. Etersom hver celletype inneholder et unikt genuttrykk, er mRNA en godt egnet biomarkør (17). En celle inneholder både husholdningsgener som uttrykkes i omtrent alle celler og har essensielle oppgaver, samt spesifikke gener for den enkelte celletypen. For eksempel genet HBB, koder for proteinet beta-globin som er en subenhet i hemoglobin. Dette proteinet befinner seg spesifikt i blodet. Etersom det uttrykkes spesifikke gener for hver celletype, kan mRNA være et godt verktøy for identifisering av kroppsvæske i kriminalsaker.



Den siste tiden har interessen for kroppsvæskeidentifisering ved bruk av mRNA økt. Flere spesifikke markører er identifisert for forensisk relevante kroppsvæsker som blod, spytt, vaginalt sekret og sæd (18-20). Blod og spytt representerer de mest vanlige kilder til biologisk materiale som finnes på et åsted. Slike materialer kan også befinne seg som blandingsspor, hvor det finnes materiale fra flere individer. I slike tilfeller kunne mRNA blitt benyttet for å skille mellom de ulike celletypene (21).

Noen markører som uttrykkes i én spesifikk kroppsvæske, kan også uttrykkes i en annen. Dermed kan det forekomme kryssreaksjoner mellom kroppsvæsker. Naturlig forekommende blodceller i vaginalt sekret forklarer høy uttrykkelse av blod-spesifikke markører (22). Markører for vaginalt sekret kan observeres også i prøver fra spytt ettersom begge kroppsvæskene produseres av slimdannende epitelceller (23). I tillegg er flere transkripter ikke fullt vevsspesifikke, men kan ha høye nivåer av uttrykkelse i én type kroppsvæske og veldig lav uttrykkelse i en annen (5).

## 1.3 mRNA i rettsgenetiske analyser

DNA-analyser har hatt en dominerende rolle i rettsgenetikk, men i senere tid har interessen for analyse av mRNA økt. RNA er godt egnet i rettsgenetiske analyser ettersom det finnes i store mengder i cellen, og kan gi verdifull informasjon som er uoppnåelig med kun DNA-analyser (13). Dermed kan prøver med små mengder biologisk materiale analyseres, noe som er gunstig innenfor det rettsmedisinske feltet.

### 1.3.1 Identifisering av kroppsvæske

Nylig er det foretatt flere studier med hensikt på benyttelse av mRNA i rutinemessige rettsgenetiske undersøkelser (1, 20, 23). Slike undersøkelser inkluderer identifisering av kroppsvæske og analyse av degradering for en mulig indikator på når et spor ble avsatt. Per dags dato benyttes serologiske analyser for identifisering av kroppsvæske i biologiske spor. Dette er indikasjonsanalyser som ikke er human-spesifikke, grunnet kryssreaksjoner med andre arter eller vev (24). Analysene utføres basert på enzymer, antigener og proteiner som befinner seg i ulike kroppsvæsker. De tradisjonelle metodene som benyttes for identifisering av kroppsvæske består av mikroskopering, immunologiske, kjemiske og enzymatiske

analyser. Mikroskopering benyttes for påvisning av spermier, mens de resterende analysene benyttes på spytt, blod og sæd (25). Immunologiske, kjemiske og enzymatiske analyser av disse kroppsvæskene er ikke optimale og kan gi både falsk positive- og negative resultater (25). Foreløpig finnes det ingen indikasjonsanalyser for å påvise vaginalt sekret (26).

Ettersom mRNA har potensialet til å skille mellom ulike celletyper, undersøkes det om molekylet kan benyttes for identifisering kroppsvæsker (27, 28). Identifisering av kroppsvæske kan i større grad bidra til å belyse en kriminalsak ettersom det gir et innblikk i hendelsesforløpet. De viktigste fordelene med mRNA-profilering sammenlignet med ordinære protein-baserte metoder er en signifikant forbedring i spesifisitet, flere kroppsvæsker og markører kan analyseres samtidig, samt en mulighet for simultan isolering av RNA og DNA fra samme prøve (23).

### 1.3.2 Degradering og tidsbestemmelse

Identifisering av kroppsvæske gir ingen informasjon om når materialet ble avsatt. Det pågår forskning om degraderingsmekanismen kan si noe om dette (21). DNA kan forbli stabilt i biologiske prøver i lang tid. I motsetning begynner RNA raskt å degradere etter avsettelse av biologisk materiale eller ved død (29). Grunnet den umiddelbare degraderingen kan endringer i transkriptomet i en prøve observeres fra minutter til år etter en biologisk prøve er avsatt eller døden inntreffer. Tidligere studier har vist at mengden RNA i en prøve reduseres over tid og kan derfor være godt egnet for å tidsbestemme et spor (21, 30).

Per dags dato har ingen studie beskrevet en metode som lover for fremtidig bruk i rettsgenetiske laboratorier (15). Flere metoder har vist en sammenheng mellom degradering i RNA og tid etter avsettelse (21). Dette kan mulig predikeres av de ulike degraderingsratene hos mRNA (13). I en studie gjort av Anderson et al (31) ble det demonstrert at degraderingsraten til to husholdningsgener (18S rRNA og beta-aktin mRNA) kunne benyttes for å estimere alderen på avsatt blod. Bauer et al (29) undersøkte degradering av mRNA i tørket blod. Det ble observert en korrelasjon mellom degradering i mRNA og alder på det tørkede blodet. Prøvene var lagret i opptil 15 år. Forskerne fant ut at tørkede blodflekker eller dehydrerte prøver inneholdt redusert RNase aktivitet grunnet redusert mengde vann. I disse studiene ble det ikke benyttet markører som var spesifikke for kroppsvæskene i studien (29, 31). Derfor vil blandet prøvemateriale gi et problem ved analyse, ettersom kroppsvæskene kan være avsatt på forskjellig tidspunkt. I senere tid er det blitt gjort analyse på mRNA-

markører som er spesifikke for kroppsvæske og en rekke markører er blitt identifisert (21). Den største bekymringen for bruk av mRNA i rettsgenetisk analyse, er den antatte høye følsomheten for degradering. I en studie gjort av Zubakov et al (22) ble det identifisert stabile markører for blod og spytt. Disse cellespesifikke markørene ble detektert i 180 dager gammelt avsatt spor. Kohlmeier og Schneider demonstrerte senere at spesifikke mRNA-markører kunne isoleres og detekteres i 23 år gammelt avsatt blod (14). De siste årene har flere studier vist at RNA-molekyler har en ulik degraderingsrate i biologisk materiale som blod, spytt og sæd (28, 32). I en studie nylig gjennomført av Fu og Allen (15) ble det benyttet kvantitativ PCR for estimering av alder på blod. Analysen demonstrerte at det er mulig å estimere alder på avsatt blod ettersom degraderingsmønsteret var forskjellig for hver markør. Studien konkluderte med at 5'-enden i transkriptet degraderte raskere enn 3'-enden i avsatt tørket blod. En stor andel av studiene gjort på mRNA og tidsbestemmelse er utført under kontrollerte lagringsforhold. Eksponering for sollys, fuktighet, høy temperatur, bakterier og andre ugunstige påvirkninger kan føre til degradering av RNA (13). Påvirkning av miljøfaktorer som fuktighet og temperatur er under evaluering (15). Setzer et al (33) demonstrerte at lagringsmiljøet hadde en signifikant effekt på degradering. Regn hadde en særdeles negativ effekt på degradering, hvor blod-spesifikke markører ikke kunne påvises etter 3 dager. Flere studier har sett på sammenhengen mellom degradering og tid, men en modell som kan benyttes på data, påvirket av ulike faktorer, er ikke blitt beregnet (34).

## 1.4 RNA ekstraksjon

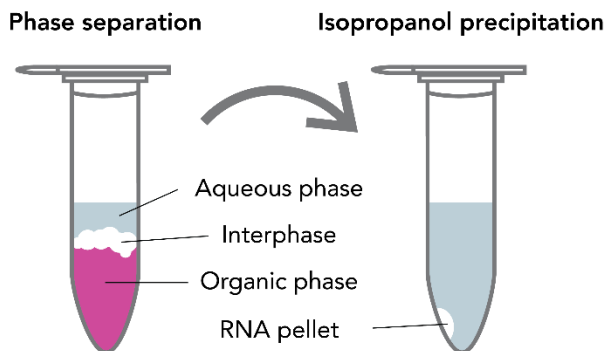
RNA ekstraksjon er en metode for å isolere RNA fra biologiske prøver. Både ekstrahering og kvantitering er viktige steg for videre analyser. Dette gjelder spesielt for sekvensering, hvor en viss mengde RNA er essensielt for optimal analysering (23). Derfor er ekstraksjon med maksimalt utbytte av RNA, samt en spesifikk og sensitiv kvantiteringsmetode viktig for senere analyse (23). Ekstraksjon av RNA kompliseres ettersom molekylet raskt kan brytes ned av RNaser. RNaser befinner seg i celler og vev, samt i miljøet rundt. På laboratoriet kan potensielle kilder være mikroorganismer i luften, på overflater, pipettespisser og i reaksjonsrør. Det er derfor viktig at arbeidsområdet er godt rengjort før bruk.

Det finnes flere ulike metoder for RNA ekstraksjon. En godt kjent metode er ekstrahering med fenol-kloroform, som er metoden som benyttes i denne oppgaven. Det benyttes organiske

løsemidler for å separere nukleinsyrer fra proteiner og resterende organisk materiale. Metoden er foretrukket ettersom en større mengde RNA ekstraheres sammenlignet med andre metoder (23). Prinsippet baserer seg på separering av faser ved sentrifugering av en prøve tilsatt fenol-kloroform (se figur 4). Guanidinium thiocyanate tilsettes for å denaturere proteiner, inkludert RNaser. Separeringen gir en øvre vannholdig fase, midtre interfase og en nedre organisk fase. RNA vil vandre til den vannholdige fasen, mens DNA og andre proteiner beholdes i den organiske fasen.

Hvilke nukleinsyrer som isoleres bestemmes av pH-nivået til løsningen (35). Under syrlige forhold (pH 4 - 6), vil kun RNA forbli i den vannholdige fasen. Dette skyldes at DNA har en høyere pKa-verdi enn RNA. De negative fosfatgruppene i DNA blir nøytralisert av  $H^+$  ved sur pH, dermed vil uladet DNA vandre ned i den organiske fasen. RNA blir deretter utfelt fra den vannholdige fasen ved å tilsette isopropanol. Isopropanolen blir videre vasket vekk med etanol og RNA-pelleten lufttørkes før den blir løst opp i RNase-fritt vann.

Til slutt blir det isolerte RNAet behandlet med DNase for å fjerne eventuelle rester av DNA. DNase er en nuklease som kan bryte ned dobbelt- og enkelttrådet DNA til oligonukleotider og enkle baser. Enzymet er spesifikt for DNA og har dermed ingen negativ effekt på innholdet av RNA i prøven (36).



**Figur 4: RNA ekstraksjon.** Illustrerer separasjonen av lagene. Den øvre vannholdige fasen inneholder RNA. Interfasen inneholder proteiner og DNA-fragmenter. Den organiske fasen inneholder DNA-fragmenter og resterende organisk materiale. Væsken med RNA blir deretter overført til et nytt rør. Etter sentrifugering kan det observeres en RNA-pellet. Figur hentet fra: Zymo Research. Tilgjengelig fra: <https://www.zymoresearch.com/pages/what-is-trizol> (Hentet dato: 05/01/20)

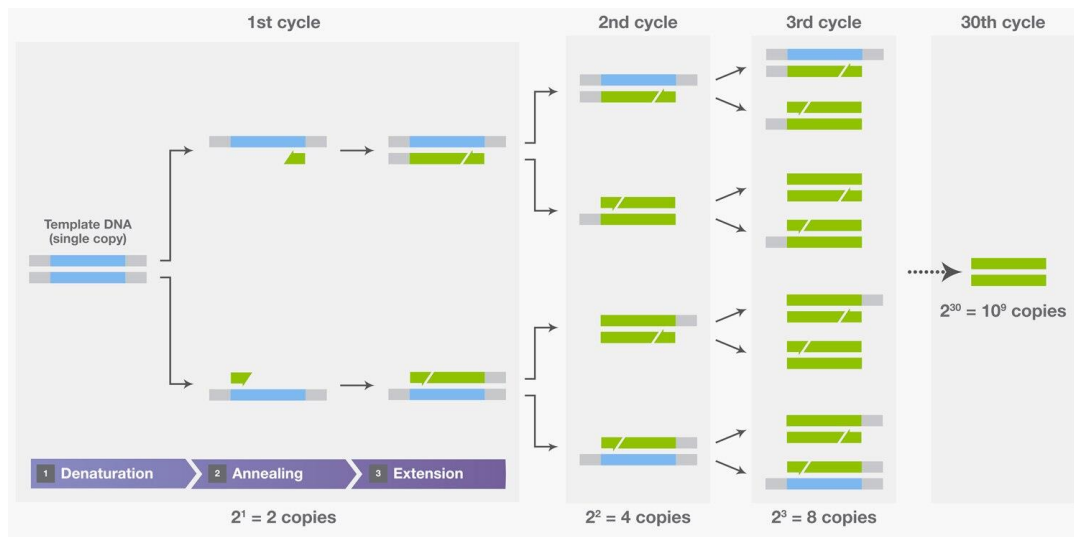
## 1.5 Polymerase kjedereaksjon

Polymerase kjedereaksjon (PCR) er en metode for å amplifisere opp tusenvis av kopier av en bestemt DNA-sekvens. PCR kan kun benyttes på DNA, dermed må RNA omgjøres til komplementært DNA (cDNA) for å kunne analyseres (se **1.5.1 revers transkripsjon**).

Metoden ble utviklet på tidlig 1980-tallet av amerikaneren Kary B. Mullis (37). Han benyttet en termostabil DNA-polymerase for å amplifisere DNA (38). Prinsippet baserer seg på evnen DNA-polymerase har til å syntetisere en komplementær DNA-tråd fra templat DNA. En PCR-analyse kan kun benyttes på DNA og krever tilstedeværelse av varmestabil DNA-polymerase, templat DNA, primere som er komplementær til starten (forward) – og slutten (revers) av genet, samt nukleotider. Prinsippet går ut på at de nevnte komponentene blandes sammen, tilsettes i en PCR-plate og deretter plasseres i en termosyklus (PCR-maskin) som tillater at repeterte sykluser av DNA-amplifisering forekommer i tre steg (se **figur 5**). En termosyklus inneholder en varmeblokk som automatisk justerer temperatur etter et forhåndslaget program.

Første steget i en PCR-analyse er å denaturere det dobbeltrådede DNAet. Ved 95 ° C brytes hydrogenbindingene mellom nukleotidene i dobbeltspiralen og DNAet blir enkelttrådet. Deretter senkes temperaturen til 50 - 65 ° C. Ved dette steget forekommer hybridisering av primere. Spesifikke primere bindes til hver sin ende av mål-sekvensen på DNA. Primerne er korte DNA-fragmenter med en definert sekvens som er komplementær til mål-DNAet som skal detekteres og amplifiseres. Hybridisering vil kun forekomme hvis primerne og DNA-sekvensen er komplementære. Deretter økes temperaturen til 75 - 80 ° C. I dette steget kan DNA-polymerase forlenge primerne ved å tilsette nukleotider til den voksende DNA-tråden. Et enzym (DNA-polymerase) kopierer sekvensen. DNA syntesen går fra den frie 3' OH-enden på begge primerne. Primerne fungerer som et punkt som DNA-polymerase kan bygge videre på (38). DNA-polymerase kobler nukleotider sammen og danner et PCR-produkt. Ved hver repetisjon av disse stegene, øker antall DNA-kopier eksponentielt.

En termosyklus kan benyttes for flere prøver per kjøring. Når PCR utføres for flere målregioner i én kjøring, kalles det multipleks PCR. For å oppnå dette blir flere primer-par tilsatt i én enkelt reaksjonsmikse.



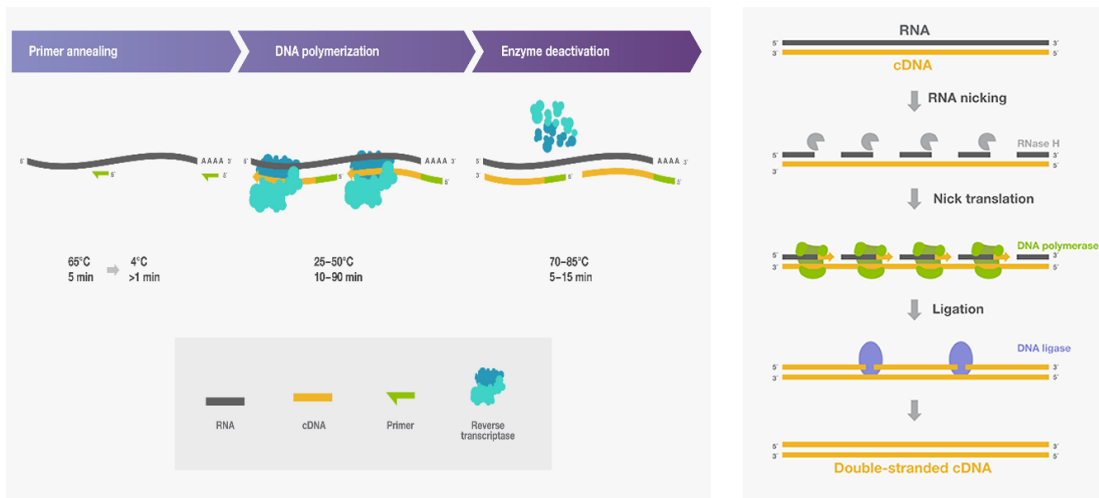
*Figur 5: PCR reaksjon. Illustrerer prinsippet for en PCR-reaksjon. Reaksjonen består av tre steg: 1) denaturering, 2) hybridisering og 3) forlenging. De tre stegene repeteres for hver syklus og antall DNA-kopier øker eksponentielt (39). (Hentet dato: 07/01-2020)*

### 1.5.1 Revers Transkripsjon

På 1970-tallet oppdaget den amerikanske biokjemikeren Howard Martin Temin et enzym som kunne danne DNA-kopier fra RNA (40). Enzymet ble kalt revers transkriptase og er naturlig forekommende i flere organismer inkludert virus, bakterier, planter og dyr (41). Revers transkriptase har blitt tatt i bruk i laboratorier og benyttes spesielt for å omdanne mRNA til cDNA for å studere kodende regioner i genene.

Revers transkripsjon (RT) er syntese av DNA fra enkelttrådet RNA ved bruk av PCR-teknikk (40). For å kunne analysere RNA må det transkribes til DNA, ettersom den konvensjonelle PCR-teknikken kun kan benyttes på DNA. For å katalysere prosessen benyttes enzymet revers transkriptase. Revers transkripsjon består av tre hovedsteg: hybridisering av primer, DNA-polymerisering og deaktivering av enzym (se **figur 6**) (41). Primer bindes til RNA-templatet ved omtrent 65 ° C. Deretter senkes temperaturen for å begynne polymerisering av cDNA. Revers transkriptase har evnen til å produsere cDNA fra RNA ved å forlenge en komplementær primer til RNA. Primeren er spesifikk for en bestemt sekvens på RNA. Dette må foregå under gunstige forhold med riktig pH og saltkonsentrasjon, samt tilstedeværelse av nukleotider. Denne syntesen genererer en cDNA:RNA hybrid. RNA blir kløyyet av en RNase, slik at cDNA forblir enkelttrådet. Dermed kan en termostabil DNA-polymerase syntetisere en

komplementær tråd til cDNA. Avslutningsvis deaktiveres revers transkriptase ved 70 – 85 °C. Temperaturen avhenger av termostabiliteten til enzymet.

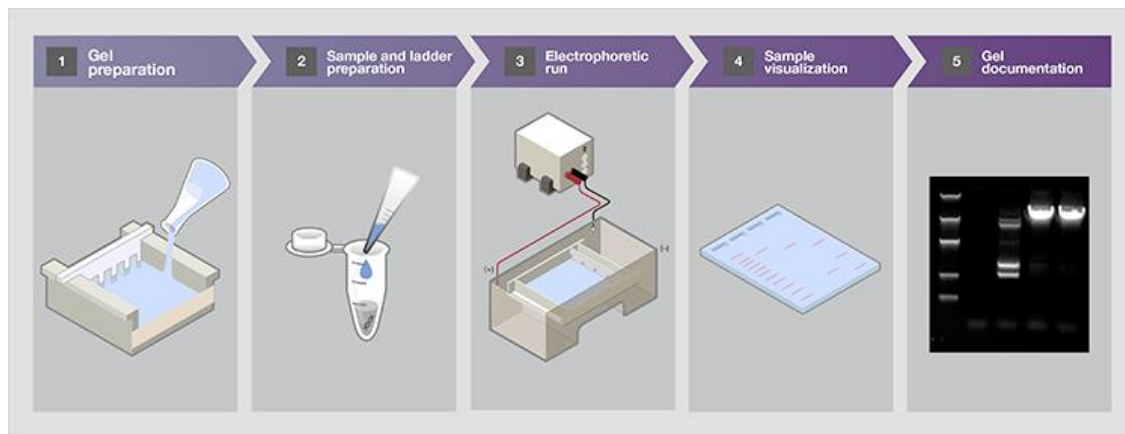


*Figur 6: Prinsippet for revers transkripsjon. Figur t.v) Hybridisering av spesifikk primer ved 65 °C, revers transkriptase polymeriserer cDNA fra RNA ved 25-50 °C og deaktivering av revers transkriptase ved 70-85 °C. Figur t.h) Illustrerer ødeleggelse av RNA-tråd med RNase H, slik at dobbeltrådet cDNA kan syntetiseres (41). (Hentet dato: 07/01-2020)*

## 1.6 Kvalitativ og kvantitativ metode

### 1.6.1 Gel-elektroforese

Gel-elektroforese er beskrevet som den enkleste og mest anvendte metoden for visualisering og analysing av PCR-produkter (38). Gel-elektroforese er en teknikk som separerer DNA-fragmenter basert på størrelse (se figur 7). Teknikken muliggjør visualisering av PCR-produkt, samt størrelsen på fragmentene. Fragmentene separeres ved påvirkning av elektrisk spenning. DNA er negativt ladet og vil derfor vandre mot den positive polen i gelen. Fragmentene separeres etter størrelse, ettersom små DNA-molekyler vandrer raskere enn store. Størrelsen kan bestemmes ved å sammenligne båndene fra PCR-produktet med en bestemt størrelsesmarkør. Størrelsesmarkøren inneholder fragmenter med kjent antall basepar. PCR-produktet farges med et fargestoff som bindes til DNA og avgir lys ved UV-stråling (38).



*Figur 7: Prinsipp for gel-elektroforese. Hentet fra Thermo Fisher Scientific. Tilgjengelig fra: <https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/na-electrophoresis-education/na-electrophoresis-workflow.html> (Hentet dato: 10/02-2020)*

## 1.6.2 RNA kvantitering

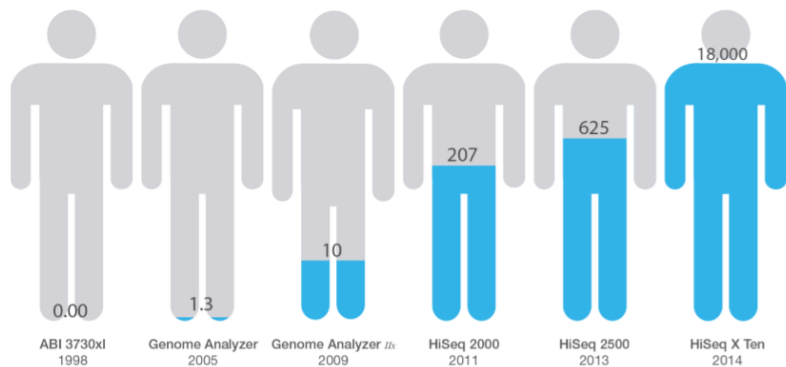
Kvantitering av RNA vil si å mengdebestemme RNA i en prøve. I denne studien utføres kvantitering ved hjelp av et fluorometer. Prøvematerialet blandes med et fargestoff med fluorescens. Fargestoffet bindes spesifikt til RNA, og vil kun emittere fluorescens ved binding. Ettersom dette ikke er human-spesifikt, blir den totale mengden RNA i prøven målt. Fluorescensen detekteres av fluorometeret (42). I denne studien benyttes Qubit™ 4 Fluorometer med et kit for RNA.

## 1.7 Massiv parallell sekvensering

DNA-sekvensering ble først beskrevet av Maxam, Gilbert og Sanger et al i 1977 (43, 44). Siden den tid har metoder for sekvensering blitt betydelig utviklet. På starten av 2000-tallet ble det utviklet en sekvenseringsteknologi som har revolusjonert genomisk forskning (se figur 8). Massiv parallell sekvensering (MPS), også referert til som neste generasjonssekvensering og dypsekvensering, gir mulighet for en mer effektiv og sensitiv sekvensering til en lavere kostnad. Sammenlignet med Sanger sekvensering ble kostnaden redusert til en fraksjon og sekvenseringstiden ble betydelig lavere. Ved benyttelse av MPS kan et human genom sekvenseres innen én enkelt dag, noe som ikke var mulig før. Det humane genom-prosjektet brukte 13 år på å fullstendig sekvensere ett humant genom med konvensjonell teknologi og



kostnadene var på omtrent 1,8 milliarder kroner (45). MPS gir mulighet for å sekvensere flere gener på samme tid, ved hjelp av svært høy hastighet og gjennomstrømmingsevne. Introduksjonen av MPS har også revolusjonert transkriptomikk. Med målrettet RNA-sekvensering er det mulig å analysere flere markører i samme analyse ved langt kortere tid enn tidligere. I tillegg er MPS spesielt tilegnet degraderte prøver, da mRNA vanligvis er fragmentert etter steget med revers transkripsjon i MPS protokollen (23).



*Figur 8: Sekvensering av humant genom over tid. Kapasiteten for å sekvensere et fullstendig genom (3,2 milliarder baser) har økt eksponentielt siden 1990-tallet. Hentet fra Illumina Inc. Tilgjengelig fra: [https://www.illumina.com/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf) (46) (Hentet dato: 07/12-2019)*

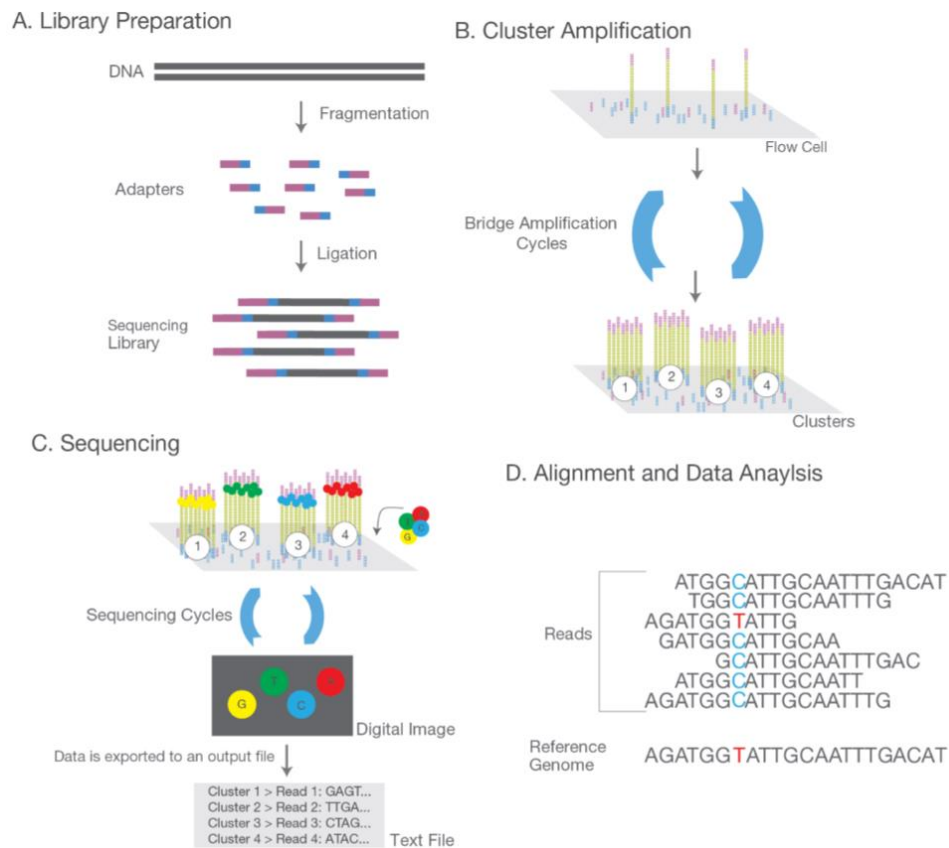
### 1.7.1 Illumina MiSeq®

Illumina MiSeq® ligner de større sekvenseringssystemene til Illumina (for eksempel Genome Analyzer og HiSeq), men skiller seg ut ved at den kun kan sekvensere et bibliotek ( gjerne multiplekset) av gangen (47). Maskinen tilbyr flere ulike analyser, inkludert målrettet RNA-sekvensering. Målrettet RNA-sekvensering benyttes for å velge og sekvensere spesifikke transkripter av interesse. Metoden gir både kvantitativ og kvalitativ informasjon. Illumina har en sekvensering ved syntese (SBS) teknologi som ligner på konvensjonell Sanger sekvensering. Forskjellen er at SBS benytter modifiserte nukleotider som inneholder en terminator som hindrer videre polymerisering. Dermed blir kun én base tilført den voksende DNA-tråden av gangen. Sekvenseringssystemet til Illumina er dominerende, ettersom det produseres resultater med høy kvalitet til en relativt lav kostnad (47).

Arbeidsflyten ved SBS-teknologi består av forberedelse av et DNA-bibliotek, klyngedannelse på flowcelle, sekvensering og signalanalyse (46). Et DNA-bibliotek er en samling av alle

målsekvensene, og forberedes ved bruk av rensset og isolert DNA. Det blir tilsatt spesifikke primere, adaptere og unike sekvenser for identifisering på hver ende av fragmentet (se **figur 9A**). Etter DNA-fragmentet er ferdig modifisert, blir det amplifisert med PCR. Det dobbeltrådige DNAet denatureres til enkeltrådet, slik at det er mottakelig for binding til flowcellen. Det ferdig modifiserte biblioteket tilsettes på en flowcelle. På flowcellen befinner det seg forankrede oligonukleotider. Oligonukleotidene er komplementære til adaptersekvensene, og dermed forekommer det en binding. Bindingen mellom oligonukleotidene og adaptersekvensene fasiliterer syntese av ny komplementær tråd. Målet med dette er å danne klynger av kopier med de ulike målsekvensene. Dette foregår gjennom bro-amplifisering (se **figur 9B**). Det dannes omtrentlig 1000 kopier av hvert DNA-fragment. Etter generering av klynger er templatene klare for sekvensering.

SBS-teknologien anvender fire deoksynukleotid trifosfater (dNTP) merket med fluorescens for å sekvensere millioner av klynger med DNA-sekvenser på overflaten til flowcellen (se **figur 9C**). Det benyttes en metode basert på reversible terminatorer, som detekterer enkle baser ved inkorporering til DNA-templattråden. Hver enkelt av de fire nukleotidene identifiseres med en karakteristisk bølgelengde. Under hver syklus i sekvenseringen, blir én enkelt dNTP tilført DNA-tråden. Nukleotidene inneholder reversible 3'- blokkere som tvinger polymerasen til å kun tilsette én base av gangen. Det merkede nukleotidet fungerer som en terminator for polymerisering. Etter hver dNTP er tilført kjeden, blir fargen avbildet for å identifisere basen og deretter enzymatisk kløvet slik at neste nukleotid skal kunne tilsettes. Grunnet de unaturlige nukleotidene, blir sekvenseringen terminert inntil en fri 3'- OH gruppe kan regenereres. Nye reversible nukleotider inkorporeres (48). Et kamera registrerer digitalt hvilken base som ble tilført ved å detektere bølgelengden på fluorescensen. Dette detekteres for hver klynge av DNA-kopier. Maskinen omdanner intensiteten til signalet til basebestemmelse under hver syklus. Bildende prosesseres videre av en programvare for bildeanalyse.



**Figur 9: Illumina sekvenseringsteknologi.** Illustrerer fire steg: (A) Forberedelse av bibliotek, (B) Generering av klynger ved bro amplifisering, (C) Sekvensering og (D) Sammenstilling og dataanalyse. Hentet fra Illumina Inc. Tilgjengelig fra: [https://www.illumina.com/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf) (46). (Hentet dato: 07/12-2019)

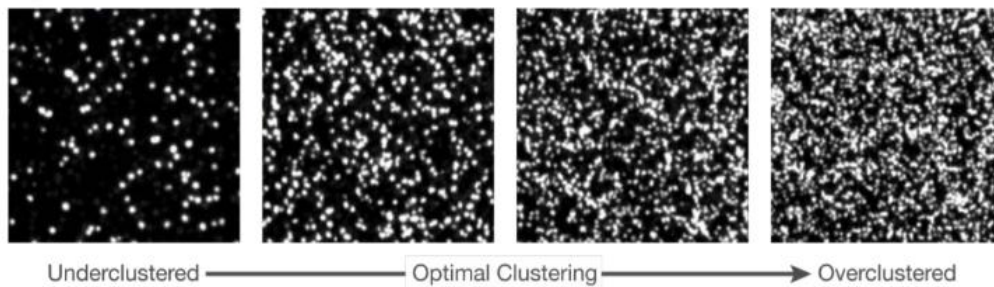
## 1.6.2 Signalanalyse

Real-time Analysis (RTA) er en forhåndsinstallert programvare på sekvenseringsmaskinen som foretar en primær analyse av rådata. Dataene blir behandlet underveis i sekvenseringssyklusene. Signalene (intensiteten) fra klyngene måles og dataene filtreres basert på kvalitet. Ved kvalitet over en viss grense blir dataene inkludert i en sekundær analyse i MiSeq Reporter (Illumina Inc). Den målbare intensiteten som hver klynge emitterer, danner grunnlag for basebestemmelse. Klyngetettheten er særdeles viktig ettersom dette påvirker basebestemmelse og kvaliteten på hele sekvenseringskjøringen (se **figur 10**). For å utføre en kjøring med optimal klyngetetthet, må det være en balanse mellom under- og overgruppering. Målet er å sekvensere med høy nok tetthet slik at den totale mengden data blir maksimert, men lav nok tetthet til at overgruppering ikke forekommer. Overgruppering fører til økt styrke på intensiteten. Den økte intensiteten gjør det problematisk for

sekvenseringsmaskinen å detektere svakere signal. Hvis overgrupperingen er svært høy, kan dette føre til at sekvenseringen blir terminert på hvilken som helst syklus.

Undergruppering vil gi data med høy kvalitet, men den totale mengden av data blir lavere.

Sammenlignet med overgruppering er undergruppering foretrukket, som skyldes at effektene er mindre alvorlig. Tetthet måles som 1000 klynger per kvadratmillimeter ( $\text{mm}^2$ ) (49).



*Figur 10: Illustrerer under-, optimal og overgruppering. Tettheten på klynger er essensielt for en optimal sekvenseringskjøring. Hentet fra Cluster Optimization Overview Guide (Illumina)(49). (Hentet dato: 27/03-2020)*

Den sekundære analysen utføres av programvaren MiSeq Reporter (Illumina Inc).

Beregninger gjort av MiSeq Reporter inkluderer kvalitetsscore på basebestemmelsen fra bilder, demultipleksing, prosentvis passeringsfilter for klynger og verdier for innfasing og pre-innfasing. Demultipleksing går ut på å sortere hvilke sekvenser som tilhører hvilken prøve, basert på indekssekvensene. Passeringsfilteret for klynger vil si en renhet på farger  $\geq 80\%$ . Innfasing forekommer når en base henger etter den nåværende syklusen. Det befinner seg en grense på  $\leq 0,25\%$ . Pre-innfasing er det motsatte av innfasing og vil si at en base hopper foran den nåværende syklusen. Grensen er på  $\leq 0,15\%$ . En overgruppering av klynger kan føre til lavere kvalitetsscore, færre klynger som passerer passeringsfilteret og en ukorrekt demultipleksing.

Som følge av begrensninger i teknologien blir DNAet sekvensert i flere tusen små enheter. Disse små enhetene betegnes for «reads», og er korte gensekvenser som korresponderer til deler av et DNA-fragment. Et høyt antall reads tilsier at genet er høyt uttrykt. Høyt uttrykte gener produserer flere mRNA, som fører til et høyere antall cDNA. Dette vil igjen gi en større mengde reads. Readsene må sammenstilles til referansesekvenser. MiSeq Reporter sammenstiller reads med referansesekvensene og genererer en tekstfil. Filen inneholder godkjente reads per amplikon, for hver prøve.

## 1.9 Formålet med oppgaven

Utvikling av en nøyaktig metode for estimering av alder på avsatt biologisk materiale kan spille en stor rolle for oppklaring av kriminalsaker. Selv om flere metoder har vist en sammenheng mellom degradering i RNA og tid (12, 15, 31, 34), er det fortsatt ingen metode som så langt er bekreftet for fremtidig bruk i rutinemessig rettsgenetisk arbeid.

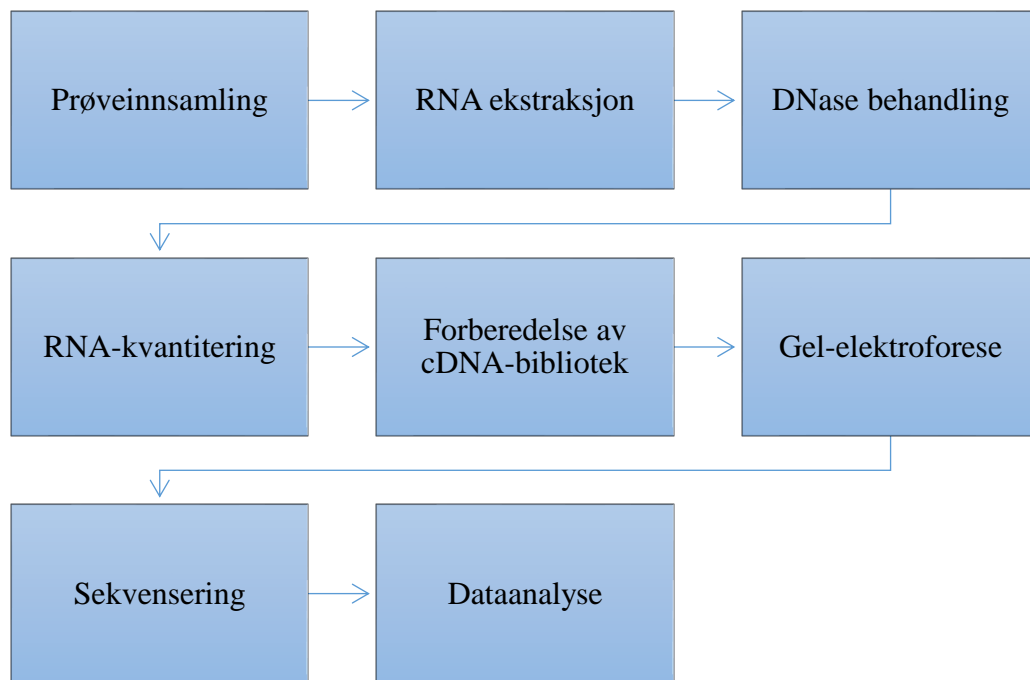
Hovedmålet med oppgaven er å undersøke sammenhengen mellom degradering i mRNA og antall dager etter avsettelse av det biologiske materialet. Det er hovedsakelig tre hypoteser som er ønsket å besvare i denne studien; 1) degradering av mRNA i et spor skjer korrelert med tid, 2) degradering av ulike mRNA skjer med ulik hastighet og 3) analyse av to markører på samme gen, plassert i hver sin ende (3' og 5'), vil vise ulik hastighet i degradering over tid.

Hypotesene skal besvares ved å utføre målrettet RNA-sekvensering på Illumina MiSeq® for å undersøke antall reads over en tidsperiode. Ved å benytte massiv parallell sekvensering er det mulig å utvikle en metode som har større kapasitet, slik at et stort antall markører kan analyseres samtidig. Statistiske metoder som regresjonsanalyse og variansanalyse anvendes for å undersøke eventuelle sammenhenger.

## 2 Materialer og metoder

Materialer benyttet i denne studien er oppført under Appendiks A med tilhørende informasjon. Alle markører som er brukt i studien er oppført under Appendiks B med tilhørende informasjon.

Arbeidsflyten for studien er illustrert i **figur 11**. Innledningsvis ble spytt- og blodprøvene innsamlet. Videre ble det utført RNA-ekstraksjon av prøvene på ulike tidspunkt, samt DNase behandling av ekstraktet for å fjerne mulig rester av DNA. Deretter ble mengde RNA målt i prøvene. RNA i utvalgte prøver ble revers-transkribert til cDNA og samlet til et cDNA-bibliotek for videre analyse. Gel-elektroforese ble utført for å måle størrelse på PCR-produktene. Avslutningsvis ble det foretatt sekvensering og analyse av sekvenseringsdataene.

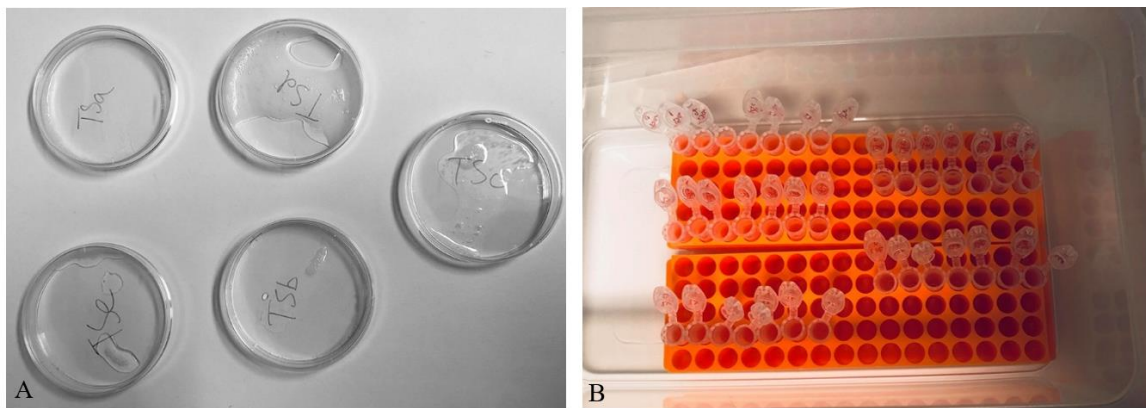


**Figur 11: Arbeidsflyt for studien.** Studien begynte med innsamling av prøver. Deretter ble det utført RNA-ekstraksjon på ulike tidspunkt etter innsamling. Videre ble ekstraktene behandlet for å fjerne mulig rester av DNA i prøven. Kvantitering ble utført for å mengdebestemme RNA i prøvene. RNA ble så revers-transkribert til cDNA og deretter samlet til et cDNA-bibliotek for videre analyse. Gel-elektroforese ble utført for å måle størrelse på PCR-produktene. Avslutningsvis ble det foretatt sekvensering og analyse av sekvenseringsdataene.

## 2.1 Prøveinnsamling

I denne studien ble det samlet inn 80 prøver fra 16 frivillige personer. Prøvematerialet bestod av spytt (n = 35) og blod (n = 45). Blodprøvene bestod av venøst EDTA-blod (n = 18) og kapillært blod (n = 27).

Spyttprøvene ble avgitt fra 5 frivillige deltakere. Prøvene ble avsatt i sterile petriskåler som deretter ble overført til sterile 1,5 ml Eppendorf® rør (se **figur 12**). Videre ble 20 µL spytt pipettert over i nye sterile 1,5 ml Eppendorf® rør. Det ble laget 6 paralleller for hver prøve. Deretter ble prøvene plassert med åpne lokk for tørking i en ren ventilert plastboks. Videre ble prøvene oppbevart i et mørkt rom med romtemperatur. Spyttprøvene ble lagret til ønsket tidspunkt for ekstrahering var oppnådd (dag 0, 3, 7, 14, 29, 56 og 143 etter avsetning av prøve) (se **tabell 1**). Tidspunktene ble valgt for å studere endring over kort – og langt tidsperspektiv. Det ble dermed valgt dager med korte mellomrom, samt dager med lengre mellomrom.



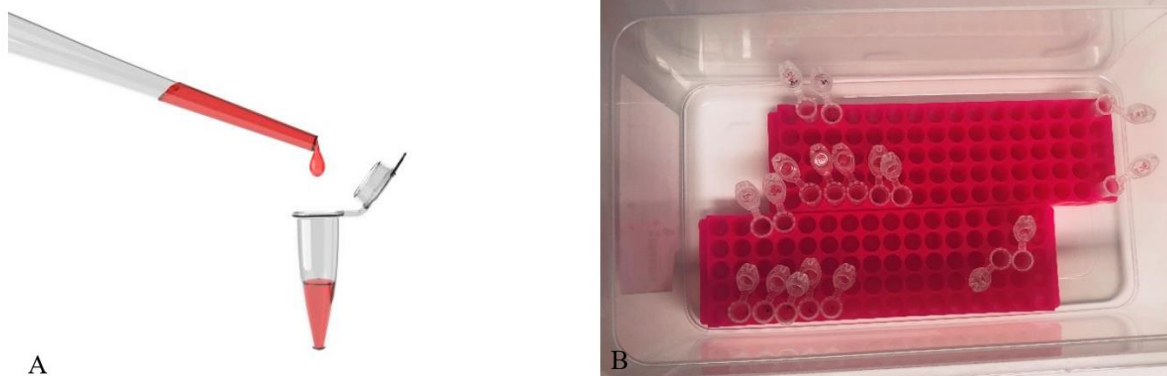
*Figur 12: Spytprøvene. A) Illustrerer spyttprøver dirkete avsatt i petriskåler. B) Illustrerer hvordan prøvene ble oppbevart med åpne lokk i en ventilert plast boks.*

Blodprøvene ble avgitt av til sammen 11 frivillige deltakere. Kapillærprøver ved bruk av Accu-Chek® Fastclix blodprøvetaker (Roche) ble innsamlet av 7 deltakere. Blodet ble avsatt i et sterilt og DNA-fritt prøverør (se **figur 13A**). Deretter ble det pipettert 5 µL blod i 1,5 ml Eppendorf® rør for videre analyse. Blodet koagulerte raskt. Derfor ble det kun pipettert blod i 3 paralleller for person A, 2 paralleller for person B, 2 paralleller for person C, 6 paralleller for person D, 3 paralleller for person E, 3 paralleller for person F og 6 paralleller for person G.

Ettersom det ble for lite blod for ønsket antall paralleller av kapillærprøvene, ble det supplert med venøse EDTA-blodprøver fra 3 frivillige deltakere. Det ble pipettert 20 µL fra hver prøve

i 1,5 ml Eppendorf® rør i 7 paralleller. Det ble også innsamlet en ekstra kapillærprøve av én frivillig deltaker, hvor 10 µL ble pipettert over i 2 paralleller.

Deretter ble prøvene plassert med åpne lokk i en ren ventilert plastboks (se **figur 13B**). Boksen ble plassert i et mørkt rom med romtemperatur. Blodprøvene ble lagret til ønsket tidspunkt for ekstrahering var oppnådd (se **tabell 1**). EDTA-prøvene ble ekstrahert på dag 1, 7, 16, 31, 42 og 114. Kapillærprøvene med 5 µL ble ekstrahert på dag 0, 6, 10, 19, 34, 45 og 118. Kapillærprøven med 10 µL ble ekstrahert på dag 7 og 80. Blodprøvene ble avsatt på forskjellig tidspunkt, dermed varierte ekstraksjonstidspunktet mellom prøvene. Tidspunktene for ekstraksjon ble valgt for kortere og lengre tidsintervall, for å studere endringer over ulike tidsperspektiv.



**Figur 13: Blodprøvene.** A) Illustrerer hvordan blodet ble avsatt i sterile rør. Figur hentet fra: iStock. Tilgjengelig fra: <https://www.istockphoto.com/no/photo/pipette-dropping-blood-gm664916074-121104887> (Hentet dato: 06/02-2020) B) Viser hvordan prøvene ble oppbevart med åpne lokk i en ventilert plast boks. Noen prøver er blitt ekstrahert, dermed er det færre i boksen.

### 2.1.1 Etikk og godkjenninger

Studien er godkjent av personvernombudet ved Oslo Universitetssykehus. Retningslinjer når det gjelder godkjenning er fulgt. De frivillige deltakerne har på forhånd fylt ut et samtykkeskjema.



## 2.2 Forberedelse av prøver

### 2.2.1 RNA ekstraksjon

Totalt RNA ble manuelt ekstrahert fra blod og spytt med fenol-kloroform metoden. Metoden ble utført i henhold til protokoll (18). Ekstraheringen ble utført under sterile forhold, hvor alt utstyr og avtrekksskap ble behandlet med RNaseZap® (Thermo Fisher Scientific) i forkant av ekstraksjonen.

Innledningsvis ble 3000 µL av en forhåndspreparert denatureringsløsning (bestående av 4 M guanidin thiocyanat, 0,02 M natriumsitrat, 0,5 % sarkosyl og 0,1 M β-mercaptoel) forvarmet på 56 °C i 10 minutter. Deretter ble 500 µL av løsningen tilsatt i hvert prøverør. Prøvene ble videre vortexet og inkubert i 56 °C i 30 minutter.

Til hvert av prøverørene ble det tilsatt 50 µL 2M natriumacetat (pH 4) for å holde pH stabil, samt 600 µL acid Phenol:Chloroform (Thermo Fisher Scientific) (pH 4,5) for å skille DNA fra RNA. Deretter ble prøvene plassert i kjøleskap med 4 °C i 30 minutter. Videre ble prøvene sentrifugert på 13,000 rpm i 20 minutter for å separere lagene. Den RNA-holdige fasen ble overført til et nytt 1,5 ml Eppendorf® rør. Deretter ble det tilsatt 2 µL GlycoBlue™ (Thermo Fisher Scientific), som bindes til nukleinsyrer og synliggjør pelleten. Påfølgende ble 500 µL isopropanol tilsatt for å lettere felle ut RNA. Avslutningsvis ble prøvene plassert i – 20 °C over natten.

Etter sentrifugering på 13,000 rpm i 20 minutter ble supernatanten fjernet. Pelleten ble vasket for salter som kan kontaminere RNA med 900 µL 75% etanol løsning (absolutt etanol og DEPC-behandlet vann). Videre ble det utført sentrifugering i 10 minutter på 13,000 rpm og deretter fjerning av supernatant, slik at pelleten kunne lufttørke i omtrent 30 minutter. Deretter ble det tilsatt 20 µL av en forvarmet (60 °C i 5 min) RNAsecure™ Resuspension Solution (Thermo Fisher Scientific) i hver prøve for å resuspendere RNA pelleten.

*Tabell 1: Tidspunkt for ekstraksjon av prøvene. Tabellen viser hvor mange prøver som ble ekstrahert ved de forskjellige tidspunktene. Hver rad oppgir antall dager etter avsettelse som ekstraksjon ble utført. Prøvene ble avsatt på forskjellig tidspunkt, og er ekstrahert etter ulikt antall dager.*

Tid (dag etter avsettelse)	Venøst EDTA-blod (20 µL)	Kapillært blod (5 µL)	Kapillært blod (10 µL)	Spytt (20 µL)
0		7		5
1	3			
3				5
6		7		
7	3		1	5
10		5		
14				5
16	3			
19		2		
29				5
31	3			
34		2		
42	3			
45		2		
56				5
80			1	
114	3			
118		2		
143				5

### 2.2.2 DNase behandling

Behandling med DNase ble utført ved bruk av TURBO DNA-free™ kit (Thermo Fisher Scientific) for å fjerne mulige mengder genomisk DNA i de ekstraherte prøvene. Dette ble utført i henhold til produsentens protokoll. Ekstraktet ble tilført 2 µL TURBO DNase buffer og 1 µL TURBO DNase enzym for å kløve fosfodiesterbindingen i DNA-ryggraden, etterfulgt av 30 minutter inkubasjon i romtemperatur. Videre ble 2 µL resuspendert DNase Inactivation Reagent tilsatt for å inaktivere DNase enzymet. Prøvene inkuberte i 5 minutter og ble samtidig lett blandet to-tre ganger. Deretter ble prøverørene sentrifugert på 10,000 x g i 1,5 minutt. Avslutningsvis ble supernatanten pipettert over i nye 1,5 ml Eppendorf® rør og lagret i – 80 °C inntil videre analyse.

### 2.2.3 Kvantitativ måling av RNA

Qubit™ 4 Fluorometer (Thermo Fisher Scientific) og Qubit® RNA HS Assay kit (Thermo Fisher Scientific) ble anvendt for å måle mengde RNA i prøven. Dette ble utført i henhold til protokoll. Arbeidsløsningen ble forberedt ved å fortynne Qubit® RNA HS Reagent 1:200 i Qubit® RNA HS buffer.

Det ble tilsatt 10 µL Qubit® RNA HS Standard 1 sammen med 190 µL arbeidsløsning i et 500 µL Qubit™ - kvantiteringsrør (Thermo Fisher Scientific). Dette ble repetert for Qubit® RNA HS Standard 2. Deretter ble 195 µL arbeidsløsning tilsatt 5 µL prøve i et 500 µL Qubit™ - kvantiteringsrør. Dette ble utført for alle prøvene. Standardene og prøvene ble så vortexet og spunnet ned før 2 minutter inkubasjon i romtemperatur. Avslutningsvis ble standardene avlest på fluorometeret for en kalibrering av apparatet, og deretter ble prøvene avlest. Fluorometeret beregnet automatisk konsentrasjonen av RNA i prøven ved bruk av denne formelen:

$$\text{Konsentrasjon av prøve (ng/ } \mu\text{L)} = \text{QF verdi} * \frac{200}{x}$$

*(QF-verdi = verdi gitt av fluorometeret, x = antall µL prøve tilsatt i analyserøret)*

## 2.3 Forberedelse av cDNA-bibliotek og sekvensering på Illumina MiSeq®

### 2.3.1 Valg av mRNA biomarkører

I denne studien ble det valgt ut til sammen 28 spesifikke mRNA-markører for blod og spytt. De bestod av 12 markører for blod og 16 markører for spytt (se Appendiks B). Markørene ble først evaluert i «The Human Protein Atlas» databasen (tilgjengelig fra <https://www.proteinatlas.org>) basert på spesifisitet og høy/lav uttrykkelse av mRNA. Deretter ble de undersøkt i forhold til tidligere forskning. Stabile markører som var lite kryssreaktive

med andre kroppsvæsker ble betraktet i studien (18, 50-52). Standard markører som i tidligere forskning har vært kryssreaktive med andre kroppsvæsker, ble ekskludert.

Ettersom studien skal studere markører som kan knyttes opp mot celletype, ble ikke markører for husholdningsgener inkludert. Husholdningsgener er høyt uttrykt i en celle, og vil kunne overdominere analysen (50).

### 2.3.2 Tillaging av primere

Primere, referert til som TruSeq® Targeted RNA Custom Oligonucleotide pool (TOP (Illumina Inc)) for de spesifikke humane markørene ble bestilt i Illumina DesignStudio™ (Illumina Inc). Søkemotoren ble benyttet for å finne primere for de utvalgte markørene. Det ble valgt primere for hver av de 28 markørene. Der det var mulig ble det valgt primer for 3'- og 5'- enden av sekvensen (se Appendiks B). Det ble valgt primerne som dekket intron for å kunne skille mellom genomisk DNA og cDNA. Ved amplifisering av genomisk DNA vil et lengre produkt genereres ettersom det inneholder intron.

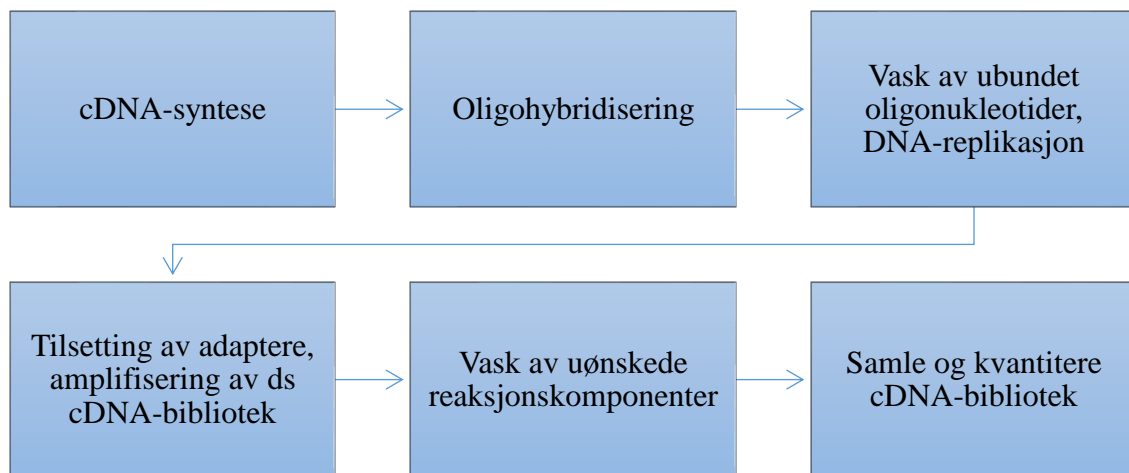
### 2.3.3 cDNA-bibliotek

Forberedelse av cDNA-bibliotek for sekvensering på Illumina MiSeq® ble utført ved bruk av protokoll for TruSeq® Targeted RNA Expression. Arbeidsflyten er illustrert i **figur 14**.

TruSeq® Targeted RNA Index Kit (Illumina Inc) og TruSeq® Targeted RNA Custom Panel kit (Illumina Inc) ble benyttet. Ettersom kitet var for 48 prøver, ble flere prøver ekskludert. De inkluderte prøvene ble valgt ut ifra målt RNA-konsentrasjon (se Appendiks C).

Alle termosykliske reaksjoner for PCR ble utført på Veriti™ termosykler (Applied Biosystems™ (Thermo Fisher Scientific)) ved bruk av 96-brønn PCR-plater forseglet med Microseal® A eller B (Bio-rad). Alle vaskesteg (ved bruk av magnetiske kuler) ble utført på

96-brønn midi-plater med magnet og forseglet med Microseal® B (Bio-rad).



*Figur 14: Arbeidsflyt for cDNA-bibliotek. Illustrerer forberedelse av et cDNA- bibliotek ved bruk av protokoll for TruSeq® Targeted RNA Expression.*

Totalt RNA ble revers transkribert til cDNA ved bruk av TruSeq® Targeted RNA Custom Panel kit (Illumina Inc) med protokoll for intakt RNA. Protokollen ble valgt ettersom RNA-konsentrasjonen var for lav til å benytte protokoll for degradert RNA. En mastermiks bestående av komponenter for cDNA-syntese (Reverse Transcription cDNA Synthesis Master Mix 1(Illumina Inc)) ble først vortexet og sentrifugert på 600 x g i 5 sekunder. Deretter ble det preparert en reaksjonsmiks for 48 prøver bestående av 211 µL mastermiks og 53 µL enzym (ProtoScript® II Revers Transcriptase (Illumina Inc)). Grunnet lav konsentrasjon av RNA, ble det ikke foretatt en fortykning med nuklease-fritt vann. Dermed ble 5 µL prøve og 5,5 µL reaksjonsmiks tilsatt i hver brønn på en 96-brønn PCR-plate. Etter tilsetting ble prøvene ristet i 20 sekunder på 1600 rpm og deretter sentrifugert i 1 minutt på 280 x g. Så ble platen med prøvene satt på Veriti™ termosyklus (Applied Biosystems™ (Thermo Fisher Scientific)) med følgende program for intakt RNA: 25 °C i 5 min, 42 °C i 15 min, 95 °C i 10 min og hold på 4 °C.

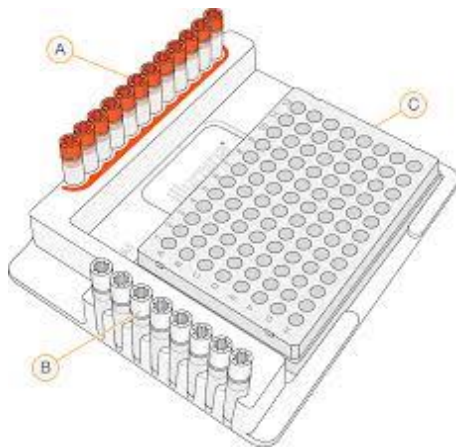
For videre hybridisering av oligonukleotider ble en reaksjonsblanding forberedt for 48 prøver. Reaksjonsblandingen bestod av 264 µL forhåndsdesignet målrettede oligonukleotider (TOP) og 264 µL TE-buffer. Blandingen ble vortexet og sentrifugert på 600 x g i 5 sekunder. Videre ble 10 µL pipettert i hver prøvebrønn på en 96-brønn PCR-plate, slik at en binding mellom cDNA og oligonukleotidene kunne forekomme. Deretter ble platen ristet på 1600 rpm i 20

sekunder etterfulgt av inkubasjon i romtemperatur i 1 minutt. Hver prøvebrønn ble så tilsatt 30  $\mu$ L av paramagnetiske streptavidin kuler (Illumina Inc) som bindes til DNA. Deretter ble prøvene satt på Veriti™ termosykler (Applied Biosystems™ (Thermo Fisher Scientific)) med følgende program: 70 °C i 5 min, 68 °C i 1 min, 65 °C i 2,5 min, 60 °C i 2 min, 55 °C i 4 min, 50 °C i 4 min, 45 °C i 4 min, 40 °C i 4 min, 35 °C i 4 min, 30 °C i 4 min og hold på 30 °C.

Etter oligohybridiseringen ble prøvene overført til en ny 96-brønn midi-plate (hybridiseringsplate) og plassert på magnet i 2 minutter. Deretter ble supernatanten fjernet fra hver brønn uten å forstyrre magnetkulene. Platen ble så fjernet fra magneten. Det ble så tilsatt 100  $\mu$ L av vaskebuffer 1 i hver brønn for å resuspendere pelleten og fjerne ubundet oligonukleotider. Videre ble platen ristet på 1800 rpm i 2 minutter, sentrifugert på 280 x g i 5 sekunder og plassert på magnet. All supernatant ble fjernet og 175  $\mu$ L av vaskebuffer 2 ble tilsatt for å fjerne rester av vaskebuffer 1. Deretter ble platen ristet på 1800 rpm, sentrifugert på 280 x g i 5 sekunder og plassert på magnet. All supernatant ble fjernet fra hver brønn.

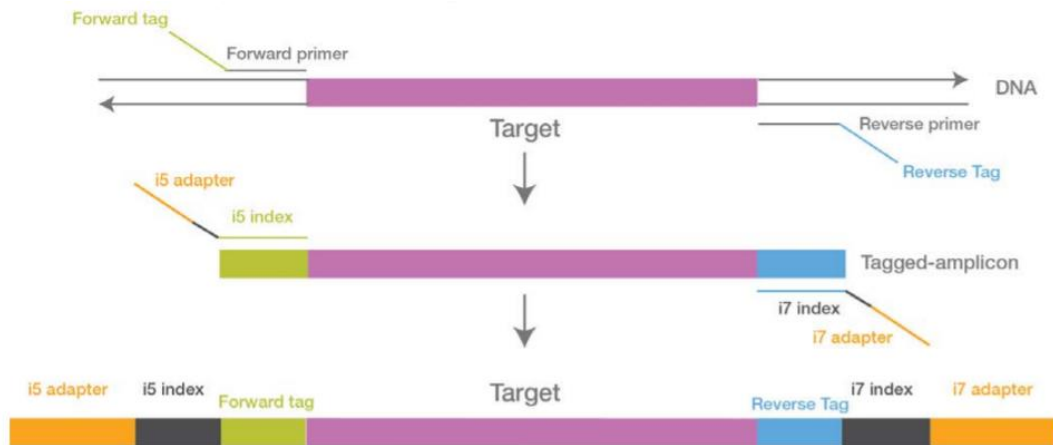
Platen ble så tatt av magneten og 40  $\mu$ L av reaksjonsmiksen for DNA-replikasjon (Extension and Ligation Mix 4 (Illumina Inc)) ble tilsatt. Videre ble platen ristet på 1800 rpm i 2 minutter, sentrifugert på 280 x g i 5 sekunder og inkubert i 45 minutter på varmeblokk med 37 °C. Deretter ble platen plassert på magnet og all supernatant ble fjernet. Avslutningsvis ble 50  $\mu$ L av vaskebuffer 2 tilsatt i hver brønn.

Videre ble hver prøve merket med en unik kombinasjon av adaptere ved bruk av TruSeq® Targeted RNA Index kit (Illumina Inc). I alle kolonnene på en 96-brønn PCR-plate (indeks-adapter plate) ble det tilsatt 4  $\mu$ L av ulike Index 1 (i7) adaptere og i alle radene ble det tilsatt 4  $\mu$ L ulike Index 2 (i5) (se **figur 15**). Dermed kan prøvene identifiseres ved samling av alle bibliotekene.



**Figur 15: Tilsetning av adaptere.** A = i7 adaptere. B = i5 adaptere. C = PCR plate. Hentet fra: Illumina TruSeq® targeted RNA expression reference guide. Tilgjengelig fra: <https://support.illumina.com/custom-protocol-selector.html> (Hentet dato: 03/11-2019)

Hybridiseringsplaten ble plassert på magnet og supernatanten ble fjernet. 2N NaOH (Illumina Inc) ble fortynnet 1:40 med nuklease-fritt vann. Deretter ble 22,5 µL av fortynnet nylaget 2N NaOH tilsatt i hver brønn, etterfulgt av risting på 1800 rpm og 5 minutter inkubasjon i romtemperatur. En amplifiseringsmiks ble forberedt for 48 bibliotek og bestod av 28 µL TruSeq DNA Polymerase 1 (Illumina Inc) og 1,4 ml PCR Master Mix 2 (Illumina Inc). Deretter ble 22 µL av amplifiseringsmiksen tilsatt hver brønn på indeks-adapler platen som inneholdt adapterne. Etter inkubasjon ble hybridiseringsplaten på nytt plassert på magneten og 20 µL av supernatanten ble overført til indeks-adapler platen. Indeks-adapler platen ble deretter ristet på 1600 rpm i 30 sekunder og sentrifugert i 1 minutt på 280 x g før den ble satt på GeneAmp PCR System 9700 (Applied Biosystems™ (Thermo Fisher Scientific)) med følgende program: 95 °C i 2 minutter. Deretter 32 sykluser av 98 °C i 30 sekunder, 62 °C i 30 sekunder og 72 °C i 60 sekunder. Til slutt 72 °C i 5 minutter og hold på 10 °C.



**Figur 16: Modifisert målsekvens.** Illustrerer målsekvensen med påsatte merkede (blå, grønn) oligosekvenser (forward/reverse primere), i5 og i7 indekser (grå) og i5 og i7 adaptore (oransje). Hentet fra: Illumina TruSeq® targeted RNA expression reference guide. Tilgjengelig fra: <https://support.illumina.com/custom-protocol-selector.html> (Hentet dato: 03/11-2019)

I henhold til protokoll ble bibliotekene rensert med AMPure XP magnetkuler (Beckman Coulter Inc) for å fjerne uønskede reaksjonskomponenter fra PCR-produktene. Det ble tilsatt 85 µL AMPure XP i alle brønner i en ny 96-brønn midi-plate (vaskeplate). AMPure XP bindes til DNA via negativt ladede karboksylgrupper på magnetkulene (47). Deretter ble vaskeplaten sentrifugert på 280 x g i 1 minutt.

Supernatanten fra indeks-adapter platen med prøvene ble overført til vaskeplaten. Deretter ble vaskeplaten ristet på 1800 rpm, sentrifugert på 280 x g i 5 sekunder og inkubert 15 minutter i romtemperatur. Videre ble platen plassert på en magnet i 5 minutter for å isolere DNA fra resterende væske. Etter observasjon av klar væske, ble 135 µL supernatant med uønskede reaksjonskomponenter fjernet fra hver brønn. Prøvene ble deretter vasket to ganger uten å forstyrre magnetkulene (tilsatte 200 µL 80 % etanol i hver brønn, inkuberte 30 sekunder i romtemperatur og fjernet supernatanten), etterfulgt av 15 minutter lufttørking på magnet.

Vaskeplaten ble fjernet fra magneten og den tørkede pelleten ble resuspendert med 15 µL resuspensjonsbuffer (RSB (Illumina Inc)). Platen ble deretter ristet på 1800 rpm, sentrifugert i 5 sekunder på 280 x g og inkubert 2 minutter i romtemperatur. Videre ble platen plassert på magnet og 12,5 µL supernatant ble overført til en ny 96-brønn PCR-plate.

Samling av cDNA-bibliotek ble utført ved å tilsette 5 µL bibliotek fra hver prøvebrønn i et 2 ml Eppendorf® rør. Deretter ble det samlede biblioteket vortexet og sentrifugert på 600 x g i 5 sekunder. Konsentrasjon av cDNA i det samlede biblioteket ble målt ved bruk av Qubit™ 1X dsDNA HS Assay Kit (Thermo Fisher Scientific) på Qubit™ 4 Fluorometeret. Dette ble



gjort i henhold til protokoll. Det ble blandet 197  $\mu\text{L}$  Qubit™ 1X dsDNA HS arbeidsløsning med 3  $\mu\text{L}$  av det samlede biblioteket. Videre ble 10  $\mu\text{L}$  av Qubit™ 1X dsDNA HS Standard 1 blandet med 190  $\mu\text{L}$  arbeidsløsning. Dette ble repetert for Qubit™ 1X dsDNA HS Standard 2. Standardene ble først avlest for en kalibrering av fluorometeret, deretter prøven.

Etter kvantitering ble konsentrasjonen ( $\text{ng}/\mu\text{L}$ ) av det samlede cDNA-biblioteket konvertert til nM ved bruk av denne formelen (Illumina Inc (53)):

$$\text{Konsentrasjon i nM} = \frac{\text{Kons. i ng}/\mu\text{L}}{660 \frac{\text{g}}{\text{mol}} * \text{Gjennomsnittlig bibliotek str i basepar}} * 10^6$$

Deretter ble biblioteket fortynnet med resuspensjonsbuffer (Illumina Inc) til en konsentrasjon på 4 nM. Fortynningen ble utregnet ved bruk av denne formelen:

$$c1 * v1 = c2 * v2$$
$$v2 = \frac{100 \mu\text{L} * 4 \text{nM}}{204,54 \text{nM}} \approx 2 \mu\text{L}$$

Det ble tilsatt 2  $\mu\text{L}$  bibliotek og 98  $\mu\text{L}$  resuspensjonsbuffer for å lage en konsentrasjon på 4 nM. Denne konsentrasjonen kreves for videre fortynning og denaturering av prøven med MiSeq Reagent kit v3 (Illumina Inc) etter protokoll (Denature and Dilute Libraries Guide (Illumina Inc)).

Denaturering av 4 nM bibliotek ble utført i henhold til protokoll A: Standard Normaliserings Metode med MiSeq Reagent v3 kit (54). Det ble overført 5  $\mu\text{L}$  fra det samlede 4 nM biblioteket til et nytt 1,5 ml Eppendorf® rør. Videre ble det tilsatt 5  $\mu\text{L}$  av nylaget 0,2M NaOH (bestående av 10  $\mu\text{L}$  2N NaOH og 90  $\mu\text{L}$  nuklease-fritt vann). Dette resulterte i et volum på 10  $\mu\text{L}$ . Løsningen ble deretter vortexet, sentrifugert på 280 x g i 1 minutt og inkubert 5 minutter i romtemperatur. Videre ble 990  $\mu\text{L}$  hybridiseringsbuffer (Illumina Inc) tilsatt. Dette resulterte i 1 mL av 20 pM denaturert bibliotek.

For å hindre overgruppering av klynger under sekvenseringen, ble biblioteket fortynnet til 15 pM etter anbefaling av produsenten. Det ble blandet 450  $\mu\text{L}$  av 20 pM bibliotek med 150  $\mu\text{L}$  hybridiseringsbuffer (Illumina Inc). Det resulterte i 600  $\mu\text{L}$  15 pM cDNA-bibliotek.

#### 2.3.4 Kvalitativ måling av cDNA-bibliotek med gel-elektroforese

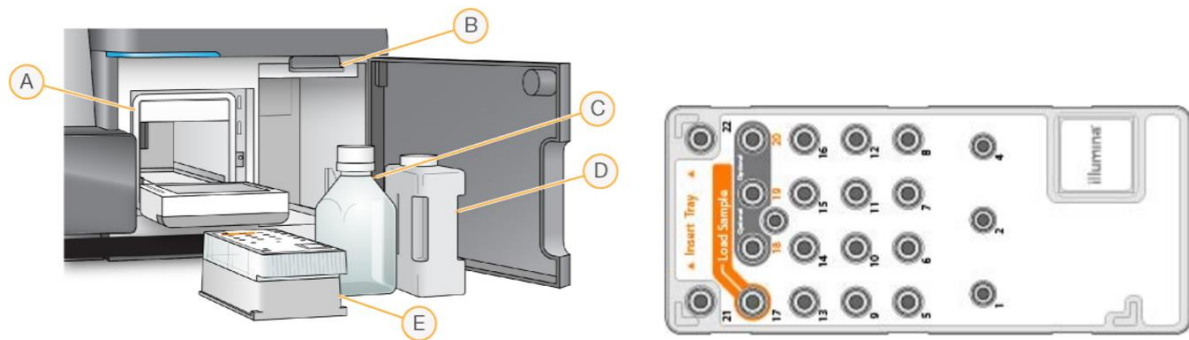
Gel-elektroforese ble benyttet som kvalitativ kontroll for undersøkelse av fragmentlengder. En agarosegel på 3% ble tillaget etter protokoll ved tilsetning av 3 gram agarose (Sigma-Aldrich) og 100 ml 1X TBE-buffer. Løsningen ble kokt opp under tilsyn i omtrent 5 minutter. Deretter ble 5  $\mu\text{L}$  konsentrert GelRed® (Biotium) tilsatt. Videre ble løsningen avkjølt til ca 70 °C under omrøring. En støpeform med kam til å danne 20 brønner ble forberedt. Deretter ble løsningen overført til støpeformen og avkjølt til den stivnet. Det ble blandet 5  $\mu\text{L}$  gel-elektroforese buffer (Sigma-Aldrich) og 10  $\mu\text{L}$  prøve. Det ble blandet 5  $\mu\text{L}$  stige (50 – 1000 bp (Lonza)) med 1  $\mu\text{L}$  loading-buffer fra samme produsent (Lonza). Etter gelen var stivnet, ble kammen trukket ut. Deretter ble gelen plassert i et elektroforesekar med 1X TBE-buffer. Det ble tilsatt 4  $\mu\text{L}$  stige (størrelsesmarkør) i første brønn. Videre ble 10  $\mu\text{L}$  prøve med gel-elektroforese buffer tilsatt i en brønn i nærheten av stige. Gelen ble kjørt på 70V i omtrent 90 minutter. Deretter ble den plassert i en maskin med UV-lys for avbildning. Prøven ble avlest ved å sammenligne resultatet med stigen.

#### 2.3.5 Sekvensering på Illumina MiSeq®

Kjøreark med oversikt over prøvene og indekssekvenser ble laget ved benyttelse av programvaren Illumina Experiment Manager Software (Illumina Inc). Dette ble utført i henhold til Illumina Experiment Manager Software Guide for MiSeq®. Kjørearket ble lastet opp på maskinen før sekvensering. Manifest-filen inneholdt referansesekvenser som de leste sekvensene skulle sammenstilles mot. Filen ble lastet ned fra MyIllumina og ble laset opp på MiSeq®. I henhold til protokoll ble manifest-filen spesifisert i kjørearket, samt kopiert til en egen mappe for manifest-filer på MiSeq®.

Sekvensering av cDNA-biblioteket ble utført på Illumina MiSeq® ved bruk av MiSeq Reagent kit v3 (150 sykluser (Illumina Inc)) i henhold til produsentens protokoll. Det ble pipettert 600  $\mu\text{L}$  15 pM av det samlede biblioteket over i reagenspatronen for sekvensering på

MiSeq® instrumentet (Illumina Inc) ved bruk av v3 flowcelle (se **figur 17**). Flowcellen ble vasket med nuklease-fritt vann, tørket over med 70 % etanol og plassert i maskinen. Avslutningsvis ble PR2-bufferen (vaskebuffer) plassert i maskinen og riktig kjøreak ble valgt. Sekvenseringen bestod av 51 «single-end» sykluser, som var standard parametere for kjøringen.



**Figur 17:** Figur t.v) Illustrerer A: Reagensavkjøler (holder temperatur mellom 2 °C og 11 °C), B: Håndtak for å føre ned et sugerør til PR2 buffer, C: PR2 buffer(vaskebuffer), D: Avfallsboks og E: Patronen med prøve og reagenser for sekvensering. Figur t.h) Illustrerer MiSeq v3 Reagent Patron. Inneholder reagenser for sekvenseringen og selve prøven. Figurer hentet fra: Illumina MiSeq System guide (Hentet dato: 12/12-2019)

### 2.3.6 Signalanalyse

Underveis i sekvenseringen ble signalintensiteten målt for hver klynge. Dette ble utført av den forhåndsinstallerte programvaren Real-Time Analysis (RTA). Det ble benyttet standard innstillinger for programvaren. Intensiteten ble målt ut ifra bildene som ble tatt av flowcellen under hver syklus. RTA utfører en basebestemmelse basert på intensiteten, samt en kvalitetsscore. Kun klynger over en viss kvalitetsscore ble videreført for sekundær analyse. Dette ble utført av den forhåndsinstallerte programvaren MiSeq Reporter (Illumina Inc). MiSeq Reporter sammenstiller de leste readsene mot referansesekvensene i manifest-filen og produserer en tekstfil. Tekstfilen inneholder godkjente reads per amplicon for hver prøve (se Appendiks G). Denne filen ble benyttet for videre dataanalyse og visualisering.

## 2.4 Dataanalyse

### 2.4.1 Statistisk programvare

Programmet R <https://cran.r-project.org/bin/windows/base/> ble benyttet og kjørt fra RStudio <https://rstudio.com/>. Dette ble anvendt for datavisualisering og statistiske beregninger. Se R-koder i Appendix F. Statistisk signifikans ble akseptert når p-verdi  $< 0,05$ .

### 2.4.2 Regresjonsanalyse

Lineær regresjonsanalyse ble benyttet for å undersøke hvordan variablene  $x_1, \dots, x_n$  potensielt predikerer eller forklarer en kontinuerlig respons  $y$ .

Regresjonsmodell:  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \epsilon$

Antatte antagelser for feilleddene  $\epsilon$  er:

- Normalfordelt
- Konstant varians
- Uavhengig

### 2.4.3 Variansanalyse

Variansanalysen ANOVA (*analysis of variance*) ble benyttet for å undersøke om det var variasjon mellom deltakere. ANOVA ble valgt ettersom variabelen «deltaker» var en faktor. Analysen baserer seg på samme antagelser som regresjonsanalyse og benyttes typisk når prediktorvariablene,  $x$ -variablene, er faktorer.

### 2.4.4 Normalisering

Normalisering ble utført for å ta høyde for variasjoner mellom individer, samt ulike prøvevolum. Dette er nødvendig for å lage en prediksjonsmodell som er gjeldene uavhengig av disse faktorene.

## 3 Resultater

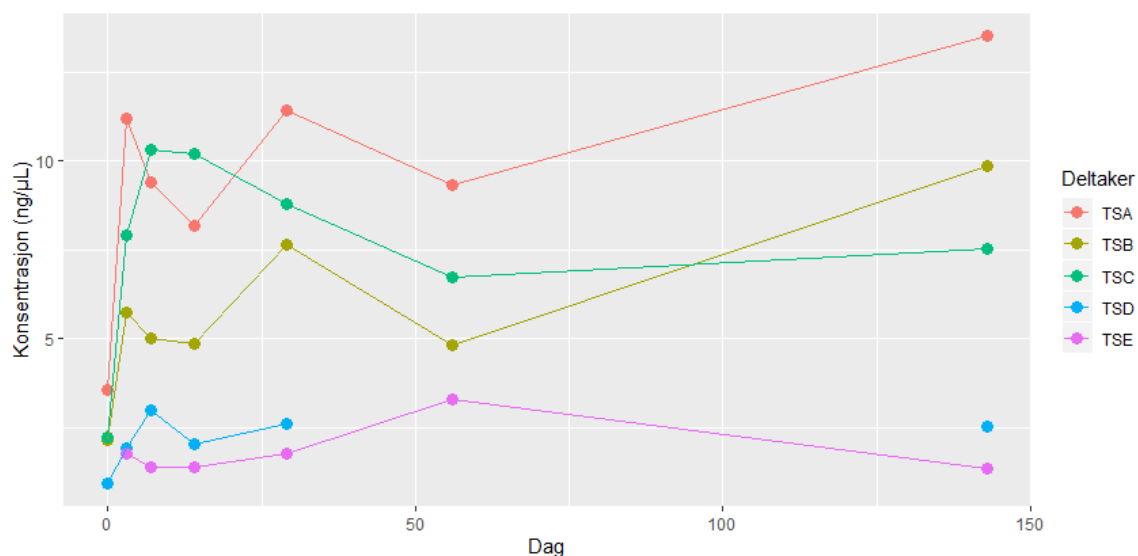
Tabell med resultater for RNA-kvantitering er i Appendiks C. Prøver utvalgt for sekvenseringen er i Appendiks D. Tabell med resultater fra sekvenseringen er under Appendiks G. R-koder for statistisk beregning er i Appendiks F.

### 3.1 RNA kvantitering

Prøver med lavere RNA-konsentrasjon enn nedre deteksjonsgrense for fluorometeret, ble oppført som «ikke detekterbar». Disse prøvene ble ekskludert fra diagrammene nedenfor. Fullstendige resultater er i Appendiks C. Negative kontroller var godkjent.

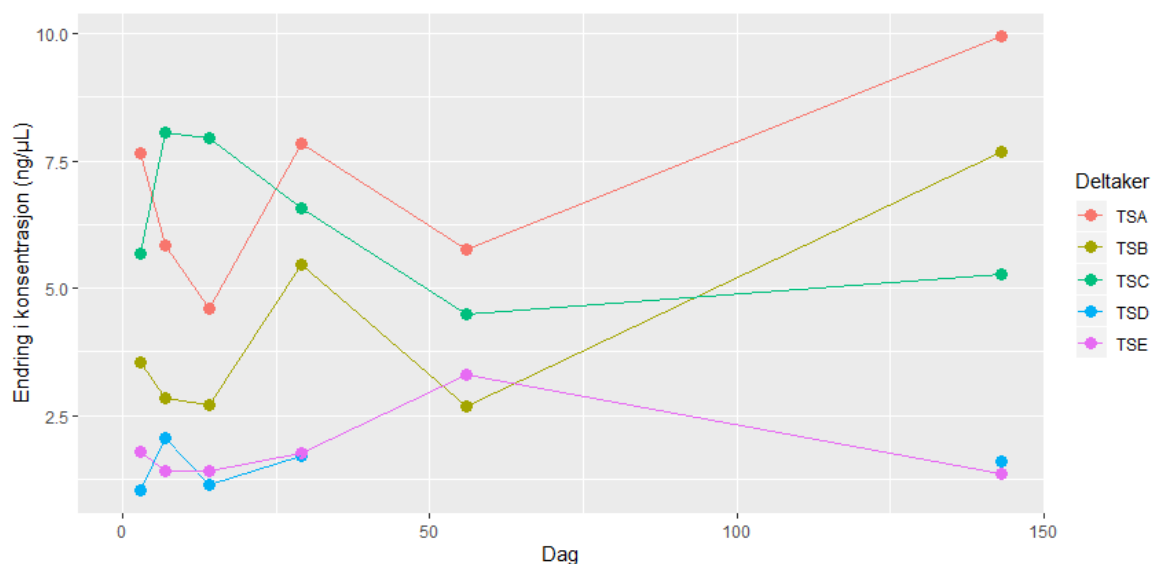
#### 3.1.1 Spyttprøvene

Spyttprøvene hadde et volum på 20 µL per prøve. RNA-konsentrasjoner for disse prøvene ble målt fra «ikke detekterbar» opp til 13,5 ng/µL. Visuelt observeres ingen lineær sammenheng mellom konsentrasjon og tid, men det ble observert en variasjon mellom deltakere (se **figur 18**). Deltakerne TSA, TSB og TSC hadde en høyere konsentrasjon enn TSD og TSE. Konsentrasjonen hadde økt fra første ekstraksjon (dag 0) til siste (dag 143). Dag 14 og 56 hadde en nedgang i konsentrasjon.



**Figur 18: Linjediagram for RNA-konsentrasjon over tid.** Diagrammet viser RNA-konsentrasjon i spytt over tid for hver deltaker. TSA, TSB og TSC hadde enn høyere konsentrasjon enn deltaker TSD og TSE. Konsentrasjon for TSE på dag 0 og TSD på dag 56 er «ikke detekterbar» og derfor fjernet fra diagrammet.

Endring i konsentrasjon over tid for hver av deltakerne er illustrert i **figur 19**. Ettersom det er interessant å observere endringer, ble det justert etter ulike startverdi. Endringen ble beregnet ved å trekke verdien til dag 0 fra resterende dager, for hver deltaker ( $y_t - y_0$ ). Ettersom dag 0 for deltaker TSE hadde en «ikke detekterbar» konsentrasjon, ble konsentrasjonen satt til 0 ng/μL. Det observeres ingen sammenheng mellom endring og tid.



**Figur 19: Linjediagram for endring i RNA-konsentrasjon over tid.** Diagrammet viser endring i RNA-konsentrasjon over tid ( $y_t - y_0$ ) i spytt for alle deltakerne. Dag 56 for deltaker TSD er ekskludert.

En variansanalyse viste en signifikant forskjell i konsentrasjon mellom deltakere (p-verdi =  $2,03 \cdot 10^{-6}$ ) (se **tabell 2**). Analysen gir  $R^2 = 292,57 / (292,57 + 142,29) = 0,67$ . Det vil si at 67 % av variasjonen i konsentrasjon kan forklares av deltaker.

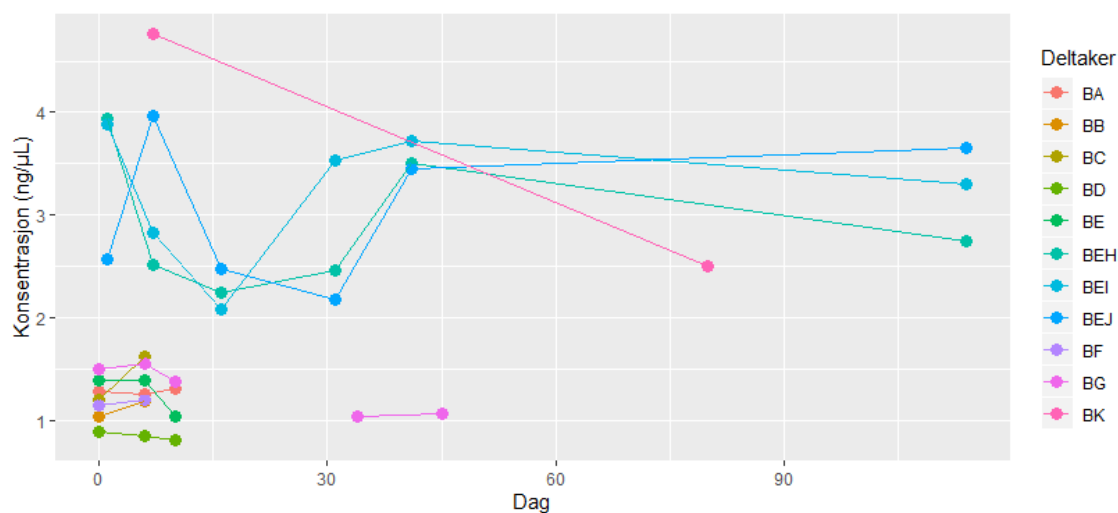
*Tabell 2: Variansanalyse (ANOVA) av RNA-konsentrasjon i spytt mellom deltakerne. Analysen viser en signifikant variasjon i konsentrasjon mellom deltakere (p-verdi =  $2,03 \cdot 10^{-6}$ ).*

<b>ANOVA-TABELL</b>					
<b>Konsentrasjon ~ Deltaker</b>					
	<i>Df</i>	<i>Sum sq</i>	<i>Mean sq</i>	<i>F-verdi</i>	<i>P-verdi</i>
<i>Deltaker</i>	4	292,57	73,143	14,096	$2,03 \cdot 10^{-6}$
<i>Residualer</i>	28	145,29	5,189		

### 3.1.2 Blodprøvene

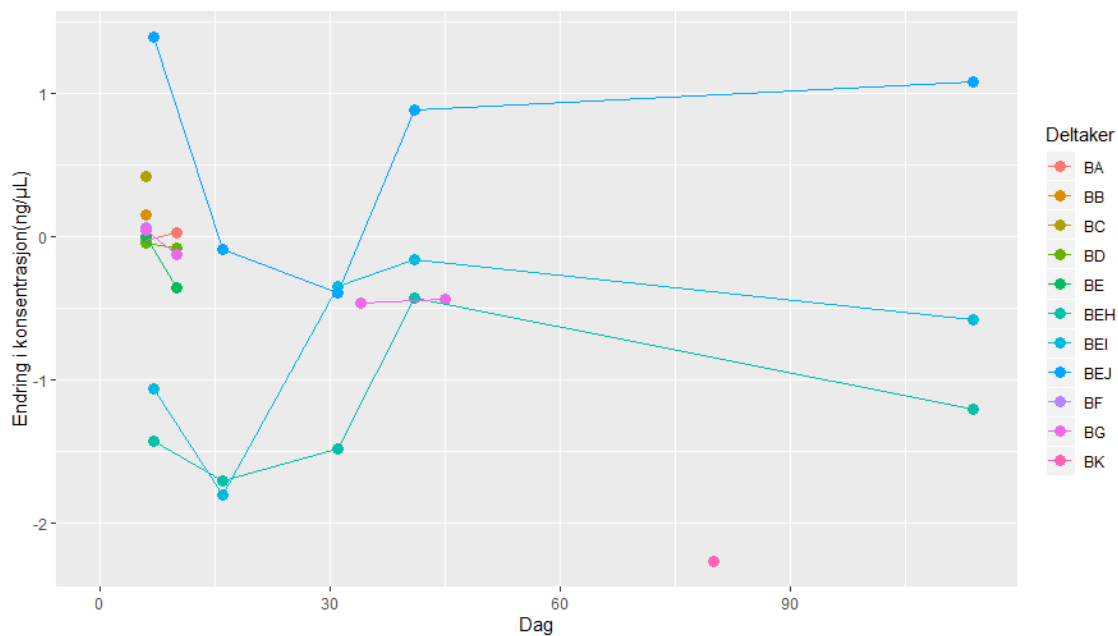
Blodprøvene hadde volum på 5-, 10-, og 20  $\mu\text{L}$ . RNA-konsentrasjonene ble målt fra «ikke detekterbar» til høyeste konsentrasjon på 4,76  $\text{ng}/\mu\text{L}$  (se **figur 20**). EDTA-prøvene med volum på 20  $\mu\text{L}$  hadde som forventet høyere konsentrasjon enn kapillærprøvene med 5  $\mu\text{L}$ .

Av **figur 20** ser det ut til at det er liten forskjell mellom deltagerne innen de to gruppene. ANOVA-analyse av EDTA-prøvene viste ingen signifikant variasjon mellom deltakere (p-verdi = 0,73). For kapillærprøvene ble det observert en signifikant variasjon (p-verdi = 0,05). Deltaker BK ble ekskludert fra denne analysen, ettersom prøvevolumet var større. Prøven med den høyeste målte konsentrasjonen var kapillærprøven med 10  $\mu\text{L}$  blod.



**Figur 20: Linjediagram for RNA-konsentrasjon over tid.** Diagrammet viser RNA-konsentrasjon i blod over tid, for hver deltaker. De EDTA-inneholdende prøvene fra deltaker BEH, BEI og BEJ (20 µL) viste en generelt høyere konsentrasjon enn hos kapillærprøvene (5 µL). Kapillærprøven fra deltaker BK (10 µL) hadde den høyeste målte konsentrasjon av blodprøvene.

Endring i konsentrasjon over tid for hver av deltakerne er illustrert i **figur 21**. Flere av prøvene viste en negativ endring i konsentrasjon, ettersom første måling hadde en høyere konsentrasjon enn resterende dager. Størsteparten av kapillærprøvene lå omkring 0, som vil si at det er lite endring over tid.



**Figur 21: Linjediagram for endring i RNA-konsentrasjon for alle deltakere.** Diagrammet viser endring i RNA-konsentrasjon over tid ( $y_t - y_0$ ) i blod for hver deltaker. BEH, BEI og BEJ er EDTA-prøver. Resterende er kapillærprøver.



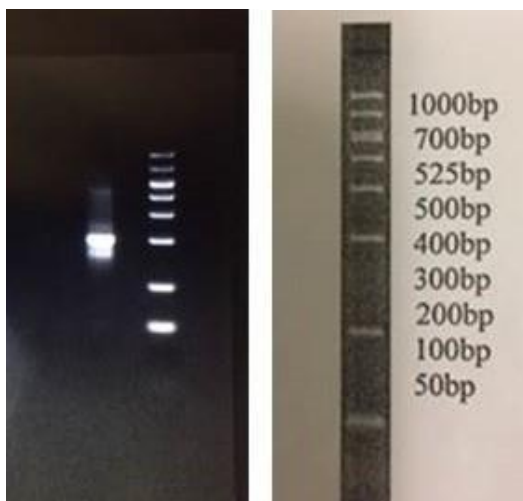
Separate regresjonsanalyser ble utført for EDTA- og kapillærprøvene. Grunnen var at prøvene hadde ulikt volum, samt at de ble ekstrahert på forskjellig tidspunkt. Regresjonsanalysen for EDTA-prøvene viste ingen signifikant endring ( $p$ -verdi = 0,625) i konsentrasjon over tid. Deltaker BK ble ekskludert i regresjonsanalysen for kapillærprøvene, ettersom volumet var ulikt fra resterende prøver. Prøver med for lav konsentrasjon ble også ekskludert. Regresjonsanalysen for kapillærprøvene viste ingen signifikant endring ( $p$ -verdi = 0,344) i konsentrasjon over tid.

### 3.2 cDNA-kvantitering

Ved kvantitering av det samlede biblioteket ble konsentrasjonen 27 ng/ $\mu$ L. Konvertering fra ng/ $\mu$ L til nM = 204,54 nM (se formel under kapittel 2, Materiale og metode).

### 3.3 Fragmentbestemmelse med gel-elektroforese

Resultatet fra gel-elektroforesen viste et tydelig bånd på 200 bp (se **figur 22**). Dette stemmer med størrelsen på fragmentene i cDNA-biblioteket. Svake bånd over og under hovedbåndet er forventet. Dette skyldes variasjon i amplikonlengde. Det svake båndet på 500 bp kan være kontaminering med DNA.



*Figur 22: Resultat fra gel-elektroforesen sammenlignet med stigen. Hovedbåndet som viser det samlede biblioteket ligger på 200 bp. Variasjon i amplikonlengde observeres ved svake bånd under og over hovedbåndet. Et svakt bånd på 500 bp kan være kontaminering av DNA.*

## 3.4 Sekvensering

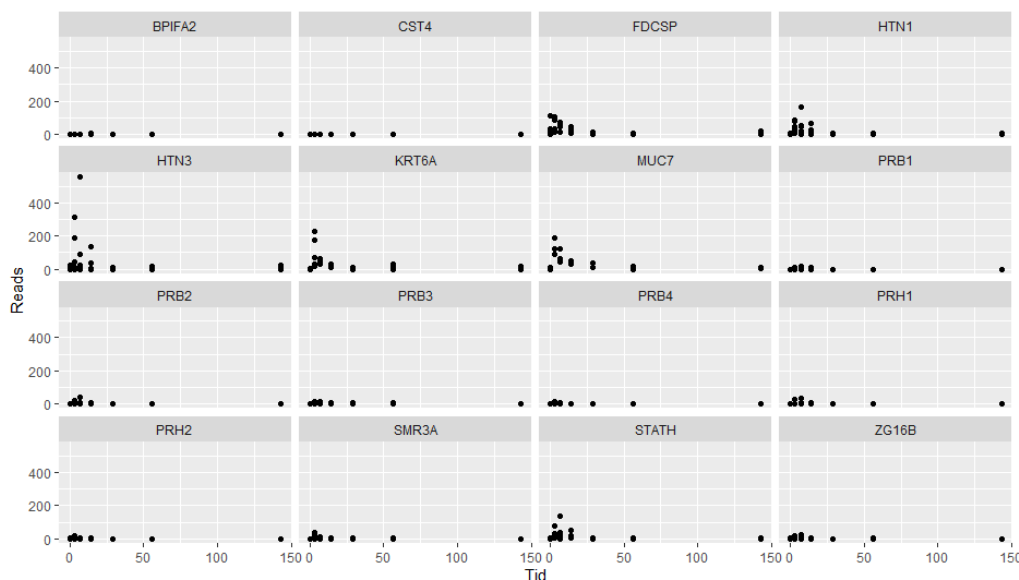
Sekvenseringen resulterte i en tabell med godkjente reads for hvert amplicon, i hver prøve. Tabellen er i Appendix G og ble benyttet for beregninger og plott. Tabell over gjennomsnittlig antall reads per amplicon er i Appendix E.

Det ble observert reads i negativ kontroll for markørene HBB og HBD. Antall reads var lavt sammenlignet med gjennomsnittlig observerte reads for reelle blodprøver (4 % for 5'-enden på HBB, 1,2 % for 3'-enden til HBB og 1,5 % for 3'-enden til HBD). Negativ kontroll ble derfor godkjent, ettersom nivået ble ansett som tilstrekkelig lavt.

### 3.4.1 Spyttprøvene

#### Evaluering av markører

Alle spyttmarkørene benyttet i analysen er illustrert i **figur 23**, her inndelt etter gener hvor amplicon for 3'- og 5'-enden er vist i samme plott. Plottene viser antall reads for tidspunkt etter avsettelse (dag 0, 3, 7, 14, 29, 56 og 143) for hver markør.



*Figur 23: Punktplott for alle spyttmarkører. Plottene viser totalt antall reads (amplicon for 3'- og 5'-ende) for alle spyttmarkørene over en tidsperiode på 0 til 143 dager.*

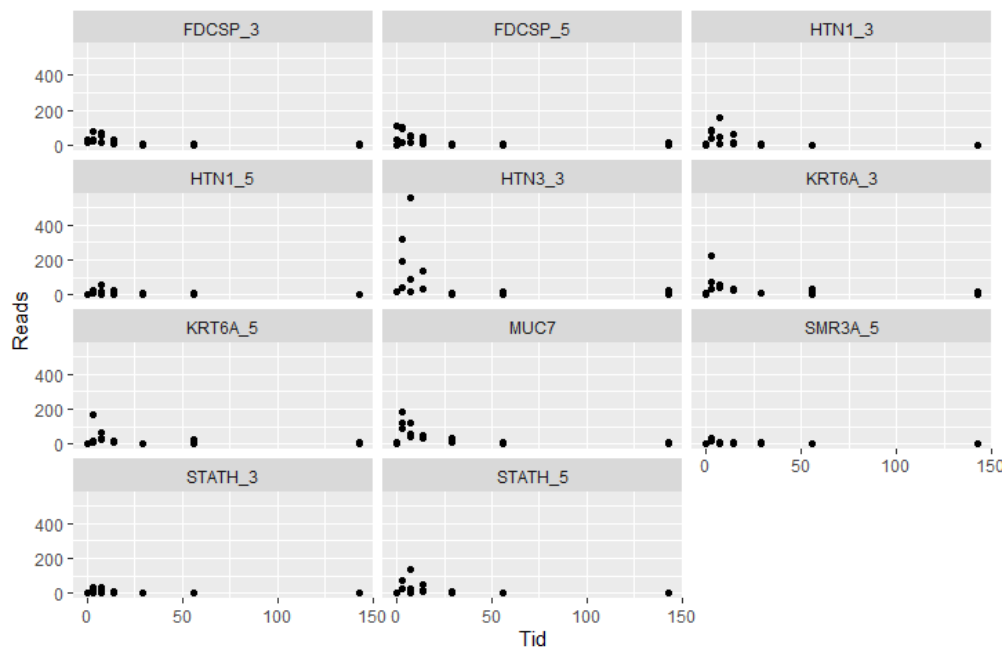
Alle spyttmarkørene ble visuelt studert hver for seg. Antall reads for hvert amplicon ble evaluert. Spyttmarkøren med høyest antall gjennomsnittlige reads var 3'-enden for HTN3. For PRH2 og CST4 ble ingen reads observert. Amplicon med et gjennomsnittlig antall reads

< 5 ble ansett som mulig bakgrunnsstøy og ekskludert fra videre analyse. De ekskluderte amplikon var: BPIFA2, CST4, 5' for HTN3, 3' og 5' for PRB1, 3' og 5' for PRB2, 3' og 5' for PRB3, 3' og 5' for PRB4, PRH1, 3' og 5' for PRH2, 3' for SMR3A, samt 3' og 5' for ZG16B.

Det var 11 amplikon som hadde gjennomsnittlig antall reads > 5 og ble benyttet i videre analyse: 3' og 5' for FDCSP, 3' og 5' for HTN1, MUC7, 3' og 5' for STATH, 3' for HTN3, 5' for SMR3A, samt 3' og 5' for KRT6A. Antall reads over tid for hvert amplikon er illustrert i figur 24.

Det totale antall observerte reads for de utvalgte markørene var lavere ved dag 0 enn ved dag 3. Dag 0 viste et samlet antall reads for markørene på 306, mens dag 3 etter avsettelse hadde flest reads med et samlet antall på 2419. Etter dag 3 sank antallet korrelert med tid. Det siste målepunktet (dag 143) etter avsettelse hadde færrest reads med et samlet antall på 122.

Mønsteret kan observeres i figur 24



Figur 24: Punktplott av reads over tid for utvalgte amplikon. Plottene viser antall reads over tid for hvert amplikon med gjennomsnittlig antall reads > 5.

### Sammenheng mellom tid og reads

Det ble foretatt en regresjonsanalyse for å undersøke om antall reads hadde en sammenheng med tid (se tabell 3). Ettersom plottene viste at dataene var ikke-lineære, ble et kvadratisk ledd ( $\text{reads}^2$ ) inkludert i analysen. Analysen viste et signifikant forhold mellom tid og reads ( $p\text{-verdi} = 2,16 * 10^{-7}$ ).  $R^2 = 0,1211$  som vil si at 12,1 % av variasjonen i tid kan forklares av reads. Basert på regresjonsanalysen kan  $y = \text{tid}$  predikeres fra ligningen:  $y = 47,32 - 0,58 * \text{Reads} + 0,001 * \text{Reads}^2 + \epsilon$ . Imidlertid er  $R^2$  bare 0,1211, slik at nøyaktige prediksjoner ikke kan forventes.

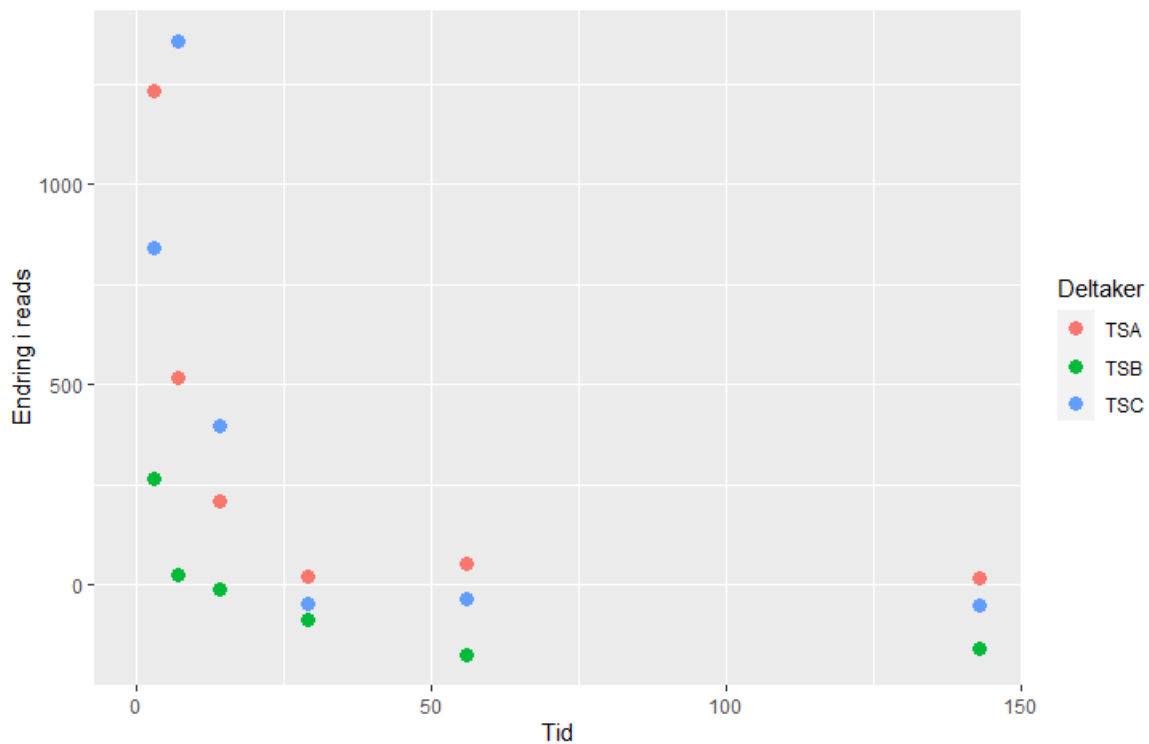
Lavere antall reads ble observert ved dag 0. Ettersom dette kan skyldes en ekstraksjonseffekt, ble det foretatt en regresjonsanalyse uten dag 0. Regresjonsanalysen uten dag 0 viste et signifikant forhold mellom tid og reads ( $p\text{-verdi} = 9,12 * 10^{-10}$ ). Variasjon i tid kan beskrives med 18,9 % av reads. Modellen ble dermed bedre ved å ekskludere dag 0.

*Tabell 3: Lineær regresjonsanalyse av de utvalgte spyttmarkørene, basert på alle tidspunkt. Reads har en signifikant påvirkning på tid ( $p\text{-verdi} = 2,16 * 10^{-7}$ ). Modellen passet bedre til dataene ved å inkludere et kvadratisk ledd.*

<b>Tid ~ Reads + Reads<sup>2</sup></b>		
	<i>Estimat</i>	<i>P-verdi</i>
<i><math>\beta_0</math></i>	47,32	$< 2 * 10^{-16}$
<i>Reads</i>	-0,578	$2,16 * 10^{-7}$
<i>Reads<sup>2</sup></i>	0,001	0,000202
<i>Residual standardfeil</i>	44,56 på 228 frihetsgrader	
<i>R<sup>2</sup></i>	0,1211	

### Sammenheng mellom tid og endring i reads

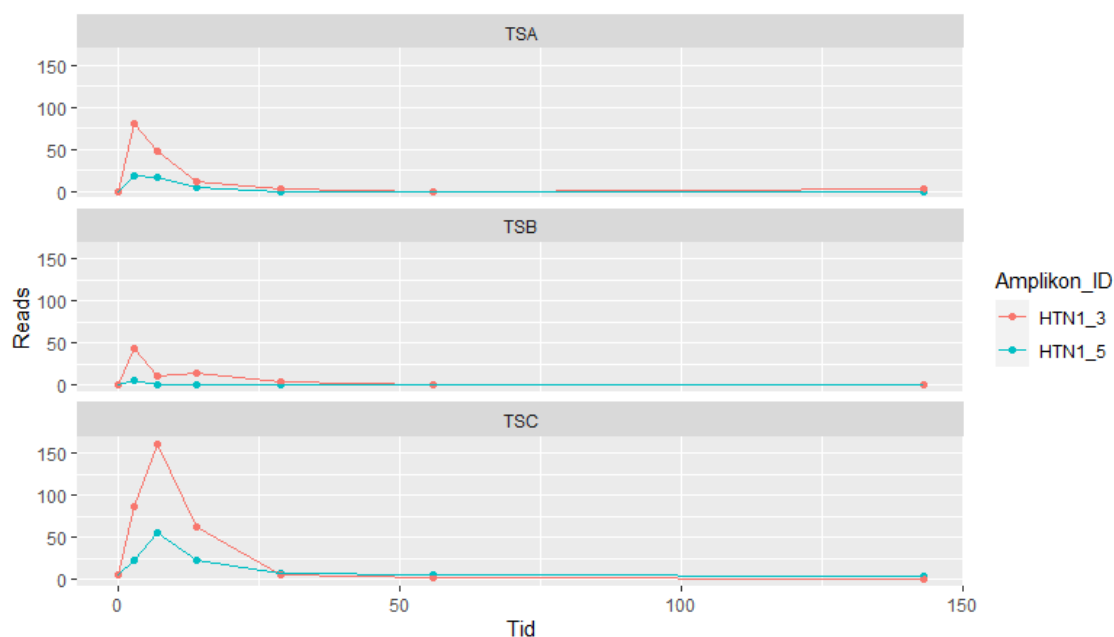
Endring i reads ble beregnet som beskrevet tidligere. Endringen var størst ved dag 3 og 7, mens de senere tidspunktene viste mindre endring fra dag 0. Dette kan forklares av et uventet lavt antall reads ved dag 0. En regresjonsanalyse viste en signifikant sammenheng mellom tid og endring i reads ( $p\text{-verdi} = 9,71 * 10^{-14}$ ).  $R^2 = 0,247$ , som vil si at 24,7 % av variasjonen i tid kan forklares av endringen i reads.



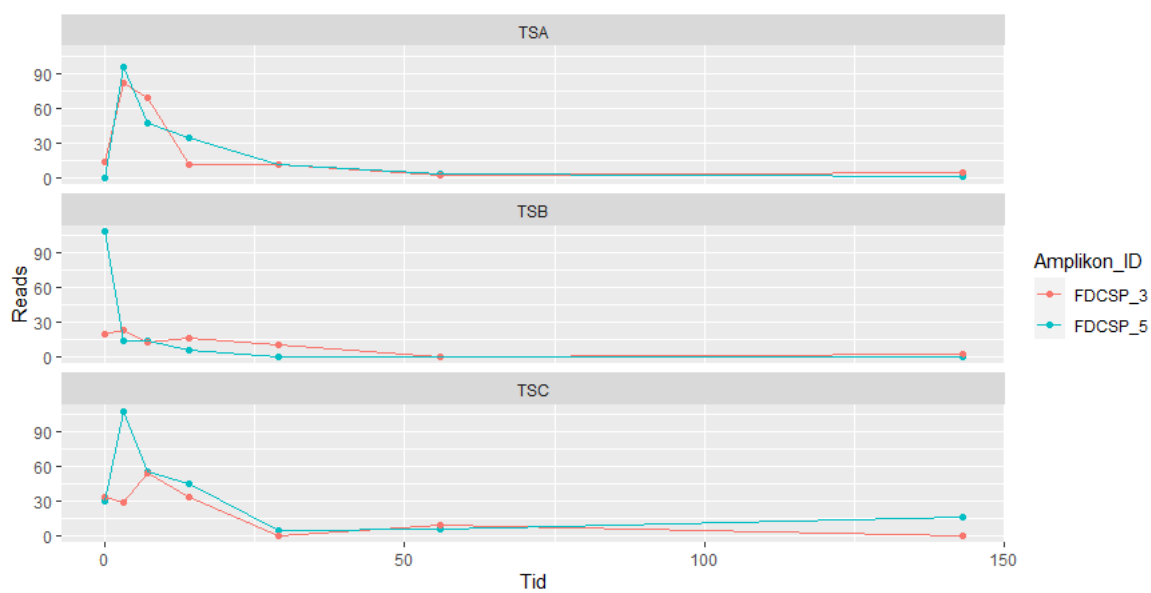
*Figur 25: Punktplott for endring i reads over tid. Plottet viser endring i reads over tid basert på signifikante spyttmarkører. Totalt antall reads for dag 0 ble trukket fra resterende dager ( $yr - y_0$ ) for hver deltaker. Punktene er farget etter deltaker (TSA, TSB og TSC).*

### Reads for 3'- og 5'- enden av markører

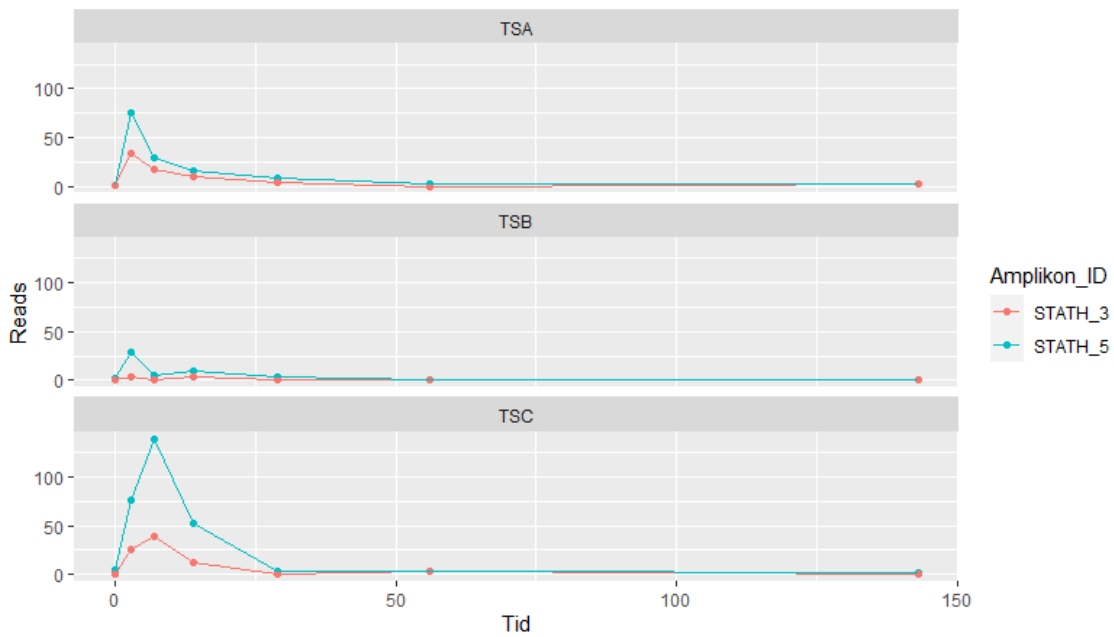
Endring i reads over tid for markører plassert i 3'- og 5'- enden for de ulike genene ble sammenlignet for å se etter ulike degraderingsmønstre. Sammenligningen ble gjort for markørene hvor både 3'- og 5'- enden hadde gjennomsnittlig  $> 5$  reads (FDCSP, HTN1, KRT6A og STATH). Det ble observert lignende degraderingsmønstre for de ulike genene, samt for markørene i 3' - og 5' - enden i samme gen (se **figur 26**, **figur 27**, **figur 28** og **figur 29**). Det var derfor ikke mulig å foreta en normalisering mellom to markører basert på dette datasettet.



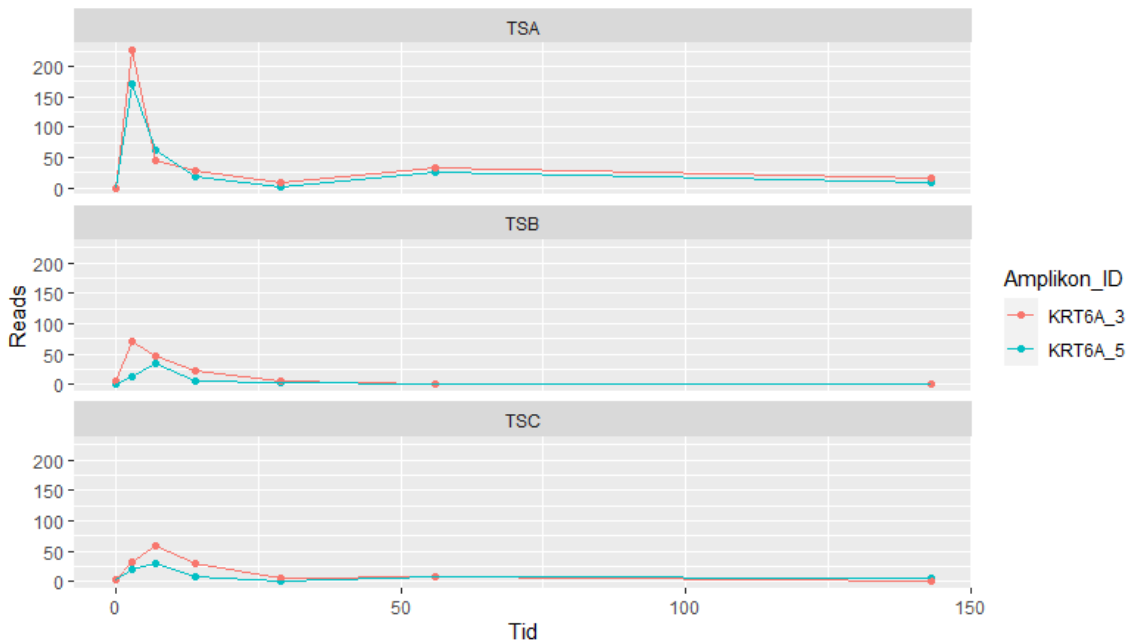
**Figur 26: Linjediagram for HTN1.** Diagrammet viser reads over tid for 3' (rød) – og 5'-enden (blå) til spyttmarkøren HTN1. Det er flest reads for 3'-enden for alle deltakere. Readsene stiger raskt etter dag 0. Etter dag 7 synker antall reads med tiden. Kurven flater ut ved dag 29.



**Figur 27: Linjediagram for FDCSP.** Diagrammet viser reads over tid for 3' (rød) – og 5'-enden (blå) til spyttmarkøren FDCSP. Det er generelt flest reads for 5'-enden for alle deltakere. Readsene stiger raskt etter dag 0, bortsett fra TSB som viser et motsatt mønster.



**Figur 28: Linjediagram for STATH.** Diagrammet viser reads over tid for 3' (rød) - og 5'-enden (blå) til spyttmarkøren STATH. Det er flest reads for 5'-enden for alle deltakere. Det er en økning i reads etter dag 0. Antallet synker igjen etter dag 3 for TSA og TSB. For TSC synker antallet etter dag 7.



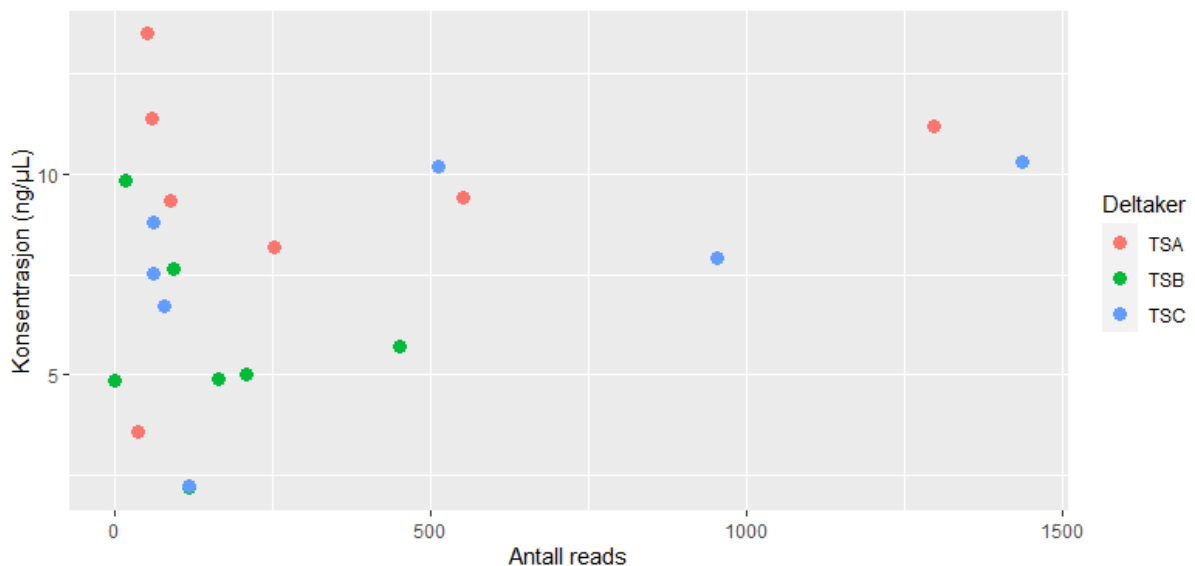
**Figur 29: Linjediagram for KRT6A.** Diagrammet viser reads over tid for 3'(rød) - og 5'-enden (blå) til spyttmarkøren KRT6A. Det er flest reads for 3'-enden. Det er en økning i reads etter dag 0. Antallet synker igjen etter dag 3 for TSA og TSB. For TSC synker antallet etter dag 7.

### Variasjon i reads mellom deltakere

Gjennomsnittlig antall reads varierte for hver deltaker. TSC hadde høyest gjennomsnittlig antall reads på 461, etterfulgt av TSA med 329. TSB hadde lavest gjennomsnittlig antall reads på 155. Det ble utført en variansanalyse for å undersøke om forskjellen i reads mellom deltakere var signifikant. Analysen viste en signifikant forskjell i antall reads (ANOVA, p-verdi = 0,02).  $R^2 = 0,033$ , som vil si at 3,3 % av variasjonen i reads kan forklares av deltaker.

### Sammenheng mellom RNA-konsentrasjon og reads

Spredningsplottet (se **figur 30**) antyder at det ikke er en sammenheng mellom RNA-konsentrasjon og antall reads for spyttprøvene.



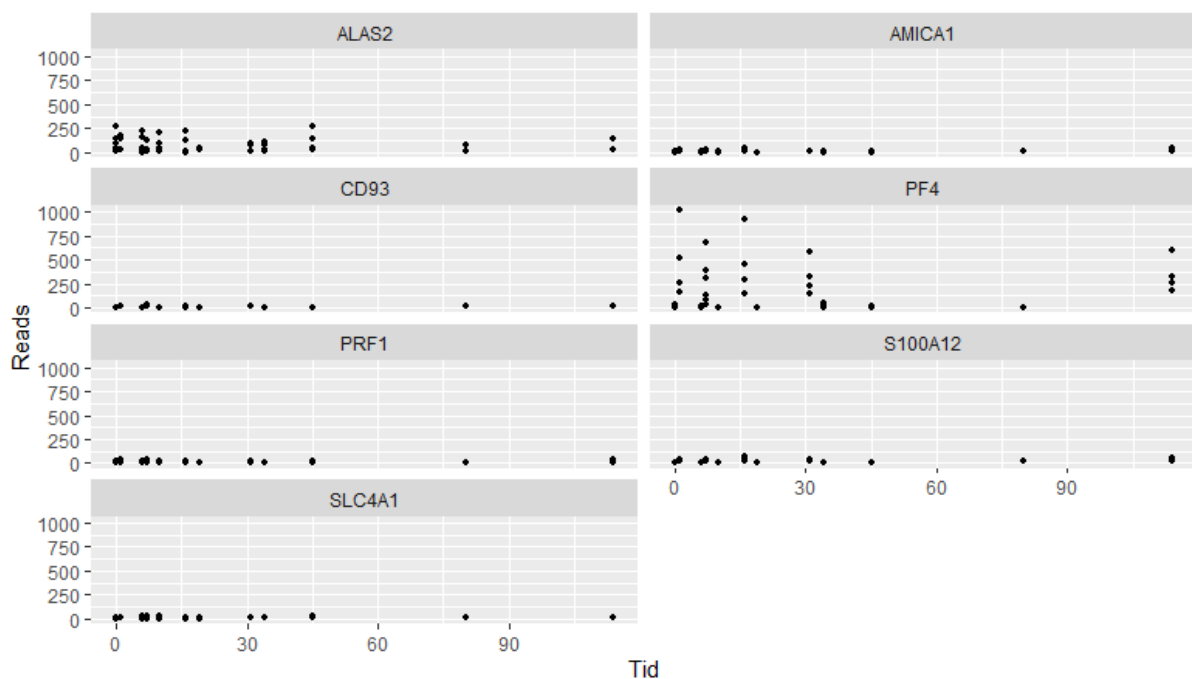
*Figur 30: Spredningsplott for RNA-konsentrasjon og reads. Plottet viser sammenhengen mellom RNA-konsentrasjon og reads i spyttprøvene. Det er ingen sammenheng mellom konsentrasjon og reads. Viser at ved høy RNA-konsentrasjon er det få reads. Punktene er farget etter deltaker (TSA, TSB og TSC).*



### 3.4.2 Blodprøvene

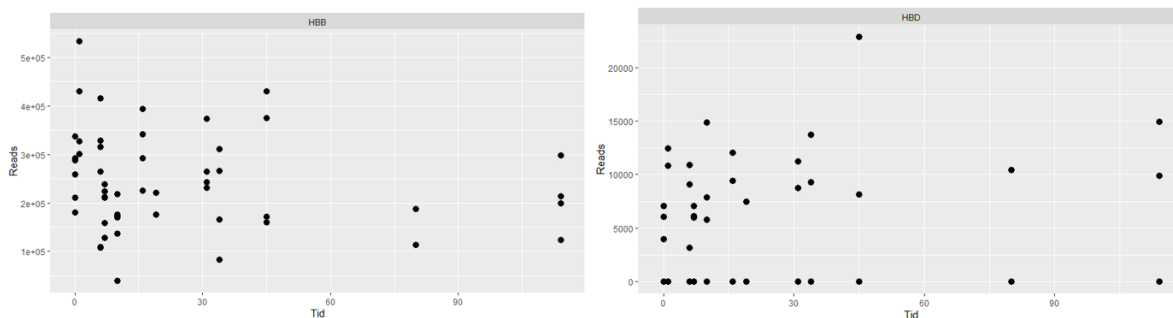
#### Evaluering av markører

Blodmarkørene ble analysert for å se etter sammenheng mellom reads og tid (se **figur 31**). HBB og HBD vises å være svært mye høyere uttrykt enn de resterende markørene. For beregning og visualisering håndteres disse separat (se **figur 32**). Markørene som gav flest reads etter HBB og HBD var ALAS2 og PF4. Markørene LGALS2, CLC, CD93 og SPTB gav færre reads (< 5 reads) og ble ekskludert fra videre analyse. Ingen trend til nedgang i reads ble observert (se **figur 31**).



*Figur 31: Punktplott for antall reads over tid per blodmarkør. Plottet viser antall reads for utvalgte blodmarkør (3'- og 5'-ende) over en tid. Ettersom HBB og HBD har vært svært sterke, er disse fremstilt i egne plott for å lettere observere antall reads hos de andre markørene.*

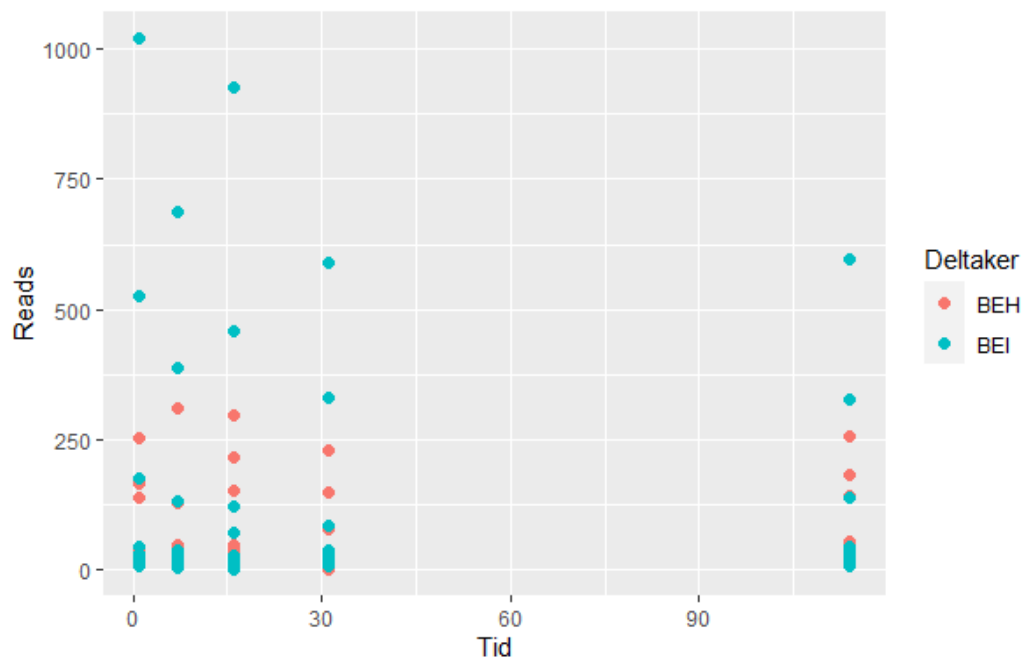
For HBB kan en observere en tendens til minking i antall reads over tid (se **figur 32**). Denne trenden observeres ikke for HBD. Det var tilnærmet ingen observerte reads for 5'-enden til HBD.



*Figur 32: Punktplott for HBB og HBD. Plottet viser antall reads for hvert tidspunkt for markørene HBB (t.v) og HBD (t.h).*

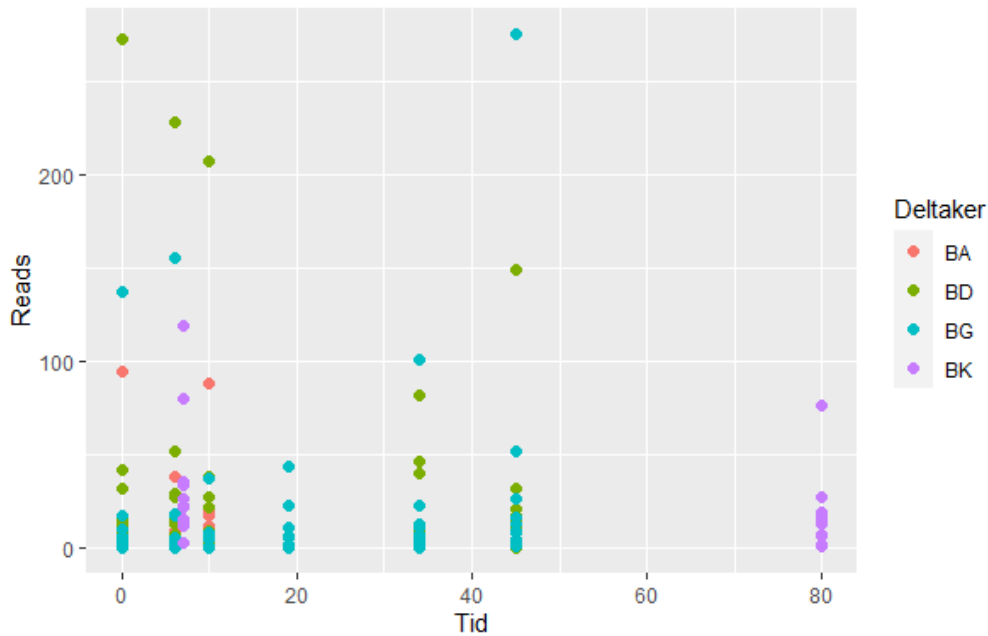
### Sammenheng mellom tid og reads

Siden EDTA- og kapillærprøvene hadde ulike volum og ekstraksjonstidspunkt ble de delt i to grupper for videre analyser. Ut fra **figur 33** for EDTA-prøvene ble det observert en nedgang i reads over tid. Generelt observeres noe variasjon i reads mellom tidspunktene (spesielt ved ekstraksjon nr 2), men regresjonsanalyse av EDTA-prøvene (HBB og HBD var ekskludert) viste ingen signifikant endring ( $p\text{-verdi} = 0,734$ ).



*Figur 33: Punktplott for reads over tid. Plottet viser antall reads for hvert ekstraksjonstidspunkt for EDTA-prøvene basert på utvalgte blodmarkører. HBB og HBD er ekskludert. Det er et lavere antall reads for begge deltakere ved ekstraksjon nr 2. Punktene er farget etter deltaker (BEH og BEI).*

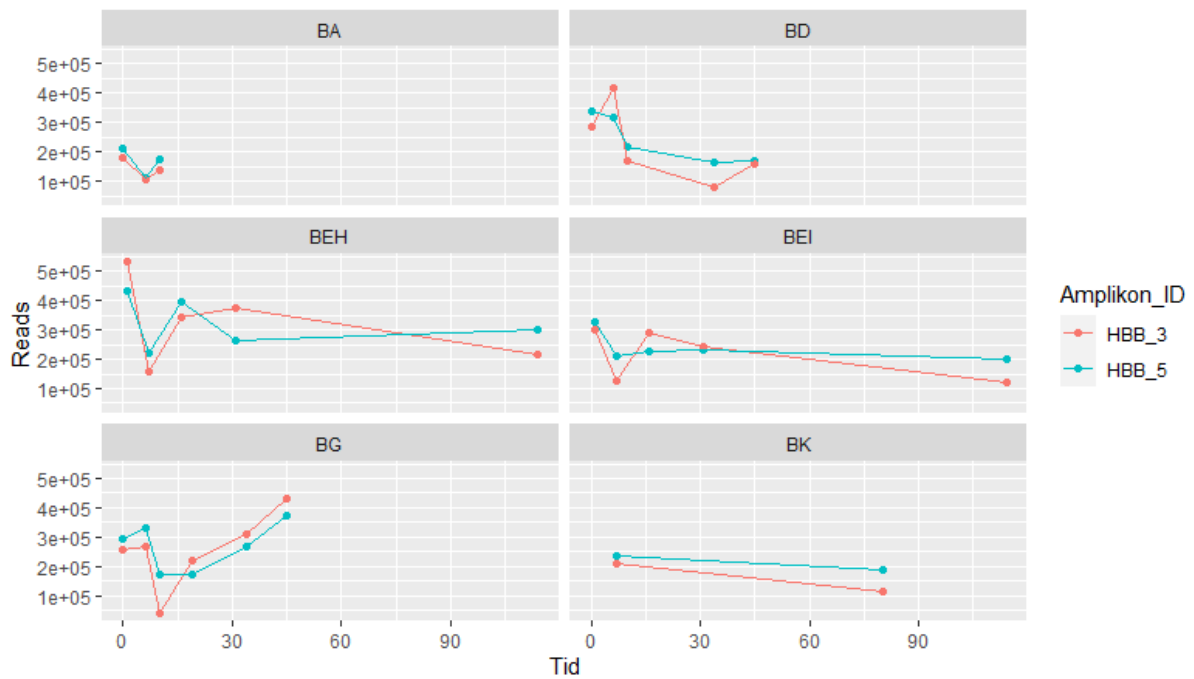
Kapillærprøvene er fremstilt i **figur 34**. Ingen signifikant endring i reads over tid ble funnet for kapillærprøvene (regresjonsanalyse, p-verdi = 0,867). Deltaker BK ble ekskludert fra regresjonsanalysen. Deltaker BG viser en annen trend enn for resterende deltakere. Det er først en nedgang, etterfulgt av en stigning på dag 34 og 45 etter avsettelse.



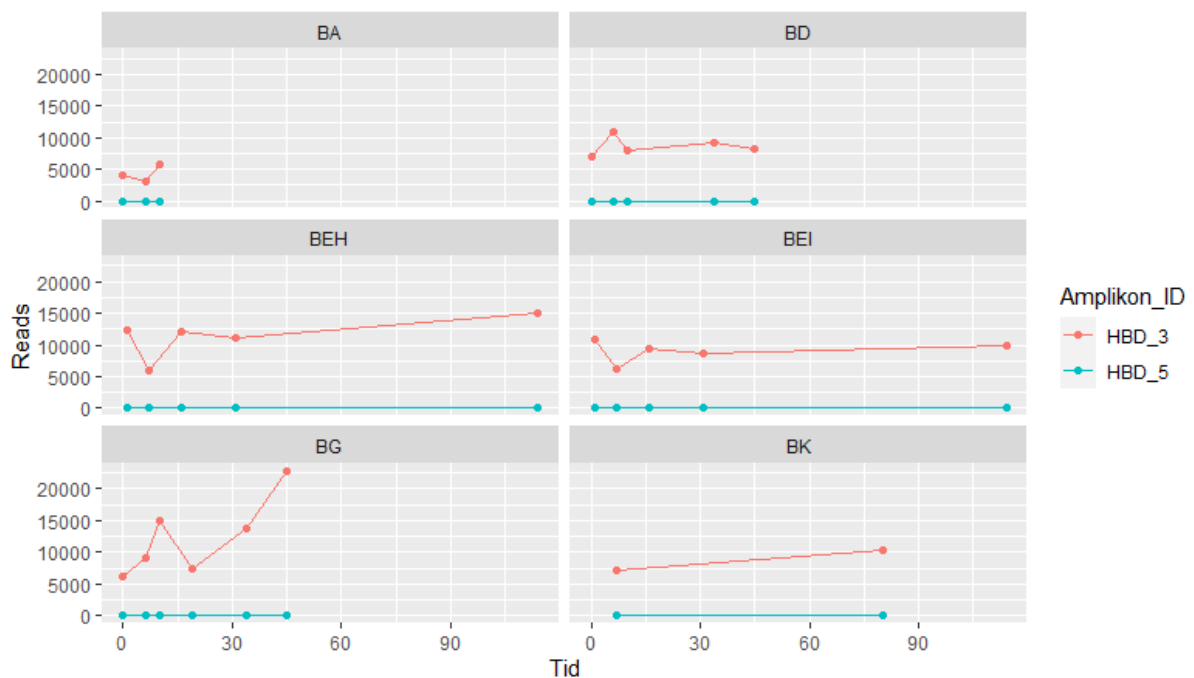
*Figur 34: Punktplott for reads over tid. Plottet viser antall reads ved hvert ekstraksjonstidspunkt for kapillærprøvene basert på utvalgte blodmarkører. HBB og HBD er ekskludert. BG viser en annen trend enn de andre prøvene. Punktene er farget etter deltaker (BA, BD, BG og BK).*

### Sammenheng mellom tid og ratio for HBB/HBD

HBB var den eneste blodmarkøren som viste nedgang i reads over tid. Dette mønsteret var likt for markørene i 3'- og 5'- enden, samt for alle prøver bortsett fra BG, som viste en økning (se **figur 35**). Markøren i 3'- enden til HBD observeres til å være en stabil markør, med liten nedgang over tid (se **figur 36**). Det ble derfor undersøkt om en ratio (3' HBB/3' HBD) korrelerte med tid og kunne benyttes som en metode for estimering av alder på et blodspor. Dermed kan modellen benyttes til tross for stor variasjon i andre faktorer.



**Figur 35: Linjediagram for HBB.** Diagrammet viser reads for hvert tidspunkt for genet HBB, for alle deltakere. Readsene er farget etter 3'(rød)- og 5'-ende (blå). Det er et lignende mønster for 3' – og 5' – enden av markøren. Deltaker BG viser et avvikende mønster, ettersom antall reads stiger over tid.

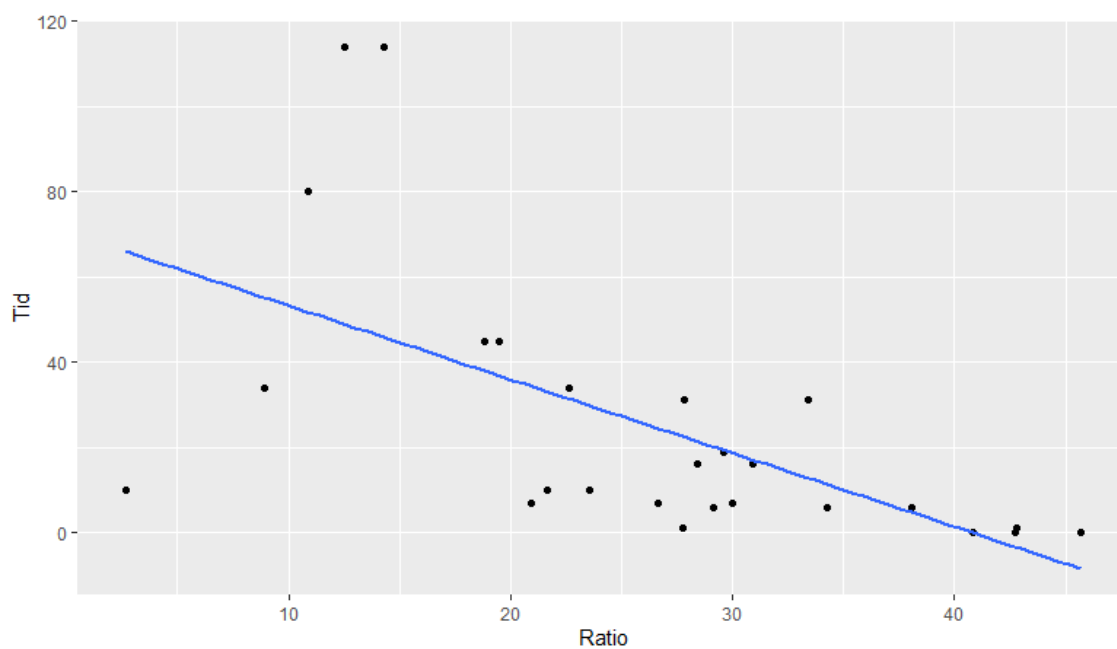


**Figur 36: Linjediagram for HBD.** Diagrammet viser reads for hvert tidspunkt for genet HBD, for alle deltakere. Readsene er farget etter 3'(rød)- og 5'- ende (blå). Det er flest reads for 3' - enden av markøren for alle deltakere. Deltaker BG viser et avvikende mønster, ettersom antall reads stiger over tid.

En regresjonsanalyse på ratioen 3' HBB/ 3' HBD viste at det var en signifikant endring ( $p$ -verdi = 0,00137) i ratio som korrelerer med tid. Dette kan forklares med regresjonsligningen  $y = 70,51 - 1,73 * \text{ratio} + \epsilon$ , hvor  $y = \text{tid}$ . Variasjonen i ratio kan forklares av tid med 35,3 %.

*Tabell 4: Regresjonsanalyse av ratio mellom 3' - ende for HBB og 3' - ende for HBD. Det er en signifikant endring i ratio som korrelerer med tid ( $p$ -verdi = 0,00137).*

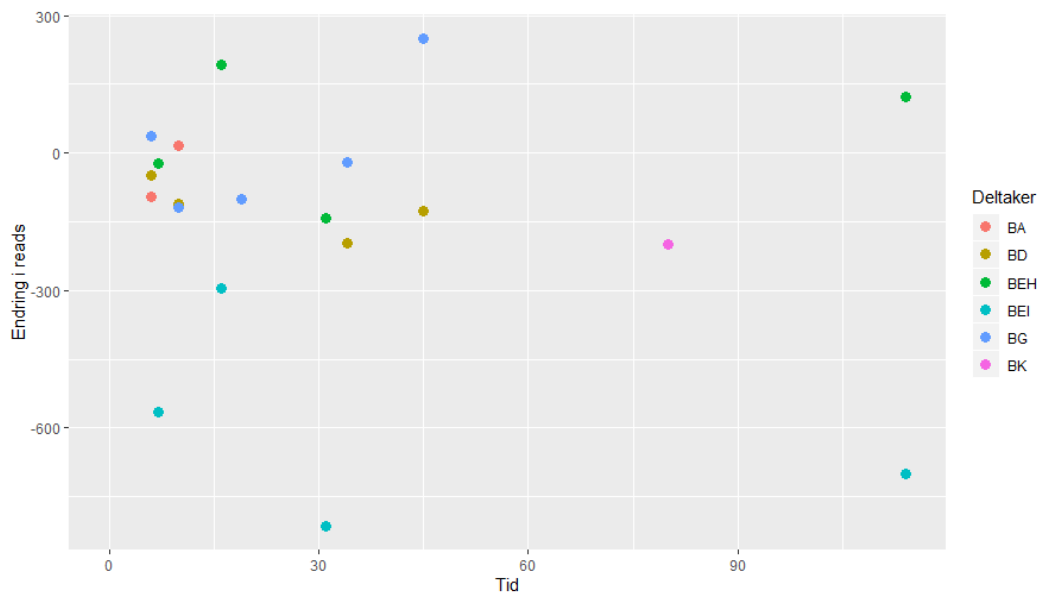
<b>Tid ~ Ratio</b>		
	<i>Estimat</i>	<i>P-verdi</i>
$\beta_0$	70,51	$2,61 * 10^{-5}$
<i>Ratio</i>	-1,73	0,00137
<i>Residual standardfeil</i>	26,4 på 24 frihetsgrader	
$R^2$	0,353	



*Figur 37: Lineær regresjonsanalyse av ratio mellom 3' - ende for HBB og 3' - ende for HBD. Det er en signifikant endring i ratio mellom HBB og HBD som korrelerer med tid ( $p$ -verdi = 0,00137)*

### Sammenheng mellom tid og endring i reads

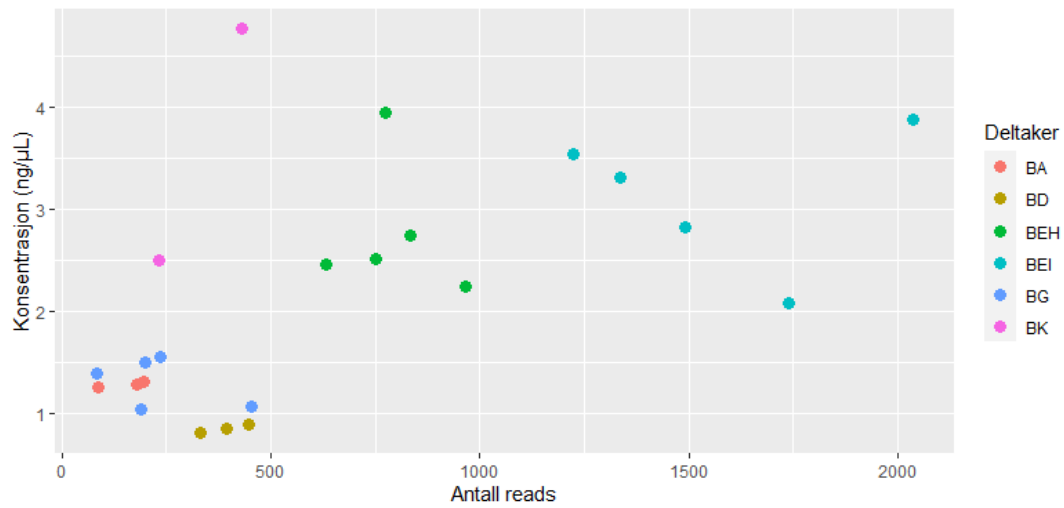
Endring i reads over tid for blodprøvene er illustrert i **figur 38**. Dette ble utført for alle deltakere. Flesteparten av deltakerne viste en negativ endring i reads, som vil si at antall reads for målt tidspunkt var lavere enn antallet for dag 0. Regresjonsanalyse for EDTA-prøvene viste ingen signifikant sammenheng (p-verdi = 0,52) mellom endring i reads og tid. Analysen for kapillærprøvene viste ingen signifikant sammenheng (p-verdi = 0,08) mellom endring i reads og tid.



*Figur 38: Punktplott for endring i reads over tid. Plottet viser endring i reads over tid ( $y_t - y_0$ ), for alle deltakere. Antall reads fra første måling er trukket fra resterende tidspunkt for alle deltakere. Dette er utført for prøver med samme volum. Punktene er farget etter deltaker (BA, BD, BEH, BEI, BG og BK). HBB og HBD er ekskludert.*

### Sammenheng mellom RNA-konsentrasjon og reads

Spredningsplottet viser sammenhengen mellom RNA-konsentrasjon og antall reads for alle deltakere. Plottet viste en antydning til sammenheng mellom reads og konsentrasjon (se **figur 39**). En regresjonsanalyse viser en signifikant sammenheng mellom konsentrasjon og antall reads (p-verdi = 0,0037) ved ekskludering av HBB og HBD. For HBB og HBD er det ingen signifikant sammenheng.



*Figur 39: Spredningsplott for RNA-konsentrasjon og antall reads. Plottet viser sammenhengen mellom RNA-konsentrasjon og antall reads for blodprøvene. Plottet viser en antydning til sammenheng. HBB og HBD er ekskludert. Punktene er farget etter deltaker (BA, BD, BEH, BEI, BG og BK).*

## 4 Diskusjon

Hovedmålet med denne studien var å undersøke om degraderingsraten i ulike mRNA korrelerte med tid. Videre var hensikten å undersøke om forskjellige mRNA degraderte i ulik hastighet, samt om hastigheten var forskjellig for 3'- og 5'- enden av transkriptet. Markørene benyttet i studien var spesifikke for spytt og blod.

RNA fra avsatt spytt og blod fra 16 deltakere ble ekstrahert med fenol-kloroform på ulike tidspunkt over en periode på 0 – 143 dager. Markørene ble valgt basert på spesifisitet for spytt og blod, samt resultater fra tidligere studier. Prøvene for sekvensering ble valgt ut ifra målt RNA-konsentrasjon. Det ble utført sekvensering av 48 prøver på Illumina MiSeq®. Markørene ble evaluert etter antall reads. Regresjonsanalyse ble benyttet for å undersøke sammenheng mellom tid og antall reads. Variansanalyse ble anvendt for å undersøke om det var signifikant forskjell mellom deltakere. Datavisualisering og statistiske beregninger ble utført i RStudio.

### 4.1 RNA-kvantitering

Flere studier har konkludert med at RNA-ekstraksjon med fenol-kloroform er en foretrukket metode (20, 55). Den er mindre kostbar sammenlignet med andre metoder, og en større mengde RNA blir ekstrahert. Ekstraksjon med fenol-kloroform er egnet for prøver med små mengder celler eller vev ettersom den gir 2,4 – 93 ganger mer RNA enn silika-baserte metoder (55). Ettersom prøver innen rettsmedisin kan være svært små og degraderte, er dette gunstig. Derfor ble metoden benyttet i denne studien. Baksiden er at metoden krever stor nøyaktighet da overflødig fenol, kloroform eller salter kan ha signifikante effekter på videre analyse (55). Metoden inneholder mange steg, noe som også øker sjansen for kontaminering. Det kan ikke utelukkes at noen av de observerte endringene, spesielt den lavere målte mengden RNA ved dag 0 for spyttprøvene, kan skyldes ekstraksjonen.

Qubit™ RNA HS Assay kit ble benyttet for kvantitering på Qubit™ fluorometer ettersom det er sensitivt og svært spesifikt for RNA, sammenlignet med andre metoder (42). Metoden er også tolerant for kontaminanter som salt, frie nukleotider, proteiner og løsemidler. Kitet er designet for å være korrekt for RNA-konsentrasjoner mellom 250 pg/μL til 100 ng/μL (56). I denne studien har RNA-konsentrasjoner i flere av spytt- og blodprøvene ikke vært



detekterbare. Dette har også forekommet i en studie fra 2019, hvor flere prøver hadde RNA-konsentrasjon lavere enn nedre deteksjonsgrense (20). Qubit™ fluorometeret var ikke sensitivt nok til å detektere disse konsentrasjonene. Små mengder RNA kan være til stede i prøven, selv om dette ikke detekteres, ettersom kun en liten del av prøven måles. Ingold et al konkluderte med at Qubit™ fluorometer verken var spesifikk eller sensitiv (23). Den nedre deteksjonsgrensen ble nådd for flere prøver i studien, spesielt for blod, sæd og hud. Resultatene ble sammenlignet med andre apparater, som detekterte lave konsentrasjoner som ikke kunne måles av Qubit™ fluorometer. En studie av Haas et al belyste ulike kvantiteringsmetoder, hvor ingen av metodene var human-spesifikke (50). De konkluderte med at en korrekt måling av humant totalt RNA fortsatt venter på en utvikling av et human-spesifikt kvantiteringssystem. Qubit™ fluorometer er ikke human-spesifikk ettersom den måler total mengde RNA i en prøve. Dermed kan også samlet RNA-konsentrasjon inneholde bakterie-RNA. I fremtiden kan det være ideelt å skaffe seg et mer sensitivt deteksjonsapparat, ettersom dette er avgjørende for videre analyse.

#### 4.1.1 Spyttprøvene

Det er tidligere observert at RNA-konsentrasjonen i spyttprøver har en høyere kvantitet sammenlignet med for eksempel blod. I en studie ble spytt lagret i romtemperatur over flere dager (57). Prøvene hadde et blakket utseende, samt høy RNA-konsentrasjon. Økningen i konsentrasjon kan skyldes vekst av bakterier. Disse funnene samsvarer med resultater i denne studien. RNA-konsentrasjonen i spyttprøvene var betydelig høyere enn hos blodprøvene.

Kvantitering av spyttprøvene viste en økning i RNA-konsentrasjon fra dag 0 til dag 143 etter avsettelse. Konsentrasjonen steg raskt fra dag 0 til dag 3. Ettersom kitet som ble benyttet for RNA-kvantitering ikke er human-spesifikt, er det mulig at økningen i konsentrasjonen skyldes RNA fra bakterier. Dag 0 viste en lav RNA-konsentrasjon, samt et lavere antall reads for spyttmarkørene. Dette kan skyldes at ekstraksjonen ikke har vært optimal. Etter dag 3 stabiliserte konsentrasjonen seg. Dette kan skyldes at prøvene er tørket og økning i bakterier har stabilisert seg. Dag 14 og dag 56 etter avsettelse viste en nedgang i konsentrasjon. Ved ekstraksjon på disse dagene ble det observert en større, blakket interfase i fase-separeringssteget. Den RNA-holdige væsken var dermed mindre enn hos resterende prøver. Det er uvisst hvorfor dette skjer, og har foreløpig kun forekommet i spyttprøvene. Det kan

skyldes innhold av proteiner i spytt, ettersom interfasen består av store DNA-fragmenter og proteiner (35). Det kan også skyldes kvaliteten på reagenset eller andre ekstraksjonseffekter.

Flere studier beskriver at analyse av spytt er problematisk, ettersom det er variasjoner mellom individer (12). Det vises til stor variasjon i oral bakterieflora hos mennesker. Dette ble også observert i denne studien. Variansanalysen viste at det var en signifikant forskjell i konsentrasjon mellom deltakerne. Det er kjent at munnen inneholder kolonier av mange forskjellige mikroorganismer (58). Spyttet kan inneholde ulike bakterier ettersom bevegelse av muskler i munnen ved spytting kan gi en større mengde mikroorganismer sammenlignet med passiv sikling. Mengden av humant-RNA i spytt varierer også mellom individer. I denne studien ble prøvene innsamlet med hensyn på dette. Deltakerne ble bedt om å ikke spise eller drikke én time før avsetting av prøve. Dermed hadde deltakerne mest mulig likt utgangspunkt, samtidig som mat og drikke ikke bidro til tap av RNA. Variasjon i RNA-konsentrasjon kan også skyldes mangel på homogenitet grunnet viskositeten i spyttet. Det er prøvd å unngå dette ved å samle inn spyttprøver i petriskåler. Deltakerne ble bedt om å spytte to-tre ganger i skålen. Videre ble spyttet pipettert over i rør og blandet lett før det ble fordelt.

#### 4.1.2 Blodprøvene

Sammensetningen av blodprøvene i denne studien var ikke optimal ettersom de ble innsamlet med ulikt volum på forskjellig tidspunkt. Dette kompliserte sammenligningen av prøvene. Dermed ble prøver med samme volum sammenlignet. Det var ønsket å kun benytte kapillærprøver uten tilsatt EDTA, ettersom dette kunne påvirke degradering i mRNA. Prøvetaking testet med én testperson viste at kapillærprøve med 1-2 stikk gav tilstrekkelig mengde blod. Under prøvetaking viste det seg at mengden blod innsamlet fra kapillærprøvene varierte mellom deltakerne, samtidig som blodet koagulerte hurtig og var vanskelig å fordele. Dermed ble ikke ønsket antall paralleller for blodprøvene oppnådd. Det ble derfor besluttet å inkludere venøse EDTA-prøver, samt ekstra kapillærprøver som kunne benyttes til langtidsanalyse. Separate regresjonsanalyser for EDTA-prøvene og kapillærprøvene viste ingen signifikant endring i konsentrasjon over tid.

Kvantitering av blodprøvene viste en stabil RNA-konsentrasjon over hele tidsperioden. Blod er homogent og sterilt, og derfor lettere å analysere enn spytt. EDTA-prøvene hadde en høyere konsentrasjon. Det er naturlig ettersom volumet var større for disse prøvene.

Koagulering kan ha hatt en innvirkning på prøvevolum ved pipettering av kapillærprøver over i prøveglass. Det var flere av kapillærprøvene som hadde lavere konsentrasjon enn nedre deteksjonsgrense for fluorometeret. Noen av disse prøvene ble likevel inkludert til videre analyse, ettersom fluorometeret ikke er sensitivt nok og RNA kan befinne seg i prøven. Visuelt så det ut til at det var liten variasjon mellom deltakerne, samt liten variasjon innad hos deltakerne. ANOVA-analyse av EDTA-prøvene viste ingen signifikant variasjon i konsentrasjon mellom deltakere, men analyse for kapillærprøvene viste en signifikant variasjon. Dette kan skyldes koaguleringen ved innsamling av prøvene.

### 4.3 cDNA-bibliotek og gel-elektroforese

Til tross for ekstraksjon med fenol-kloroform, ble ikke tilstrekkelig mengde RNA innfridd for protokoll til cDNA-bibliotek. Mengden for degradert RNA protokoll var  $\geq 200$  ng i hver prøve. For intakt protokoll var kravet 50 ng i hver prøve. Intakt protokoll ble derfor anvendt selv om ønsket mengde ikke ble oppnådd for alle prøvene. I en studie hvor også ekstraksjon med fenol-kloroform ble utført, kunne heller ikke ønsket RNA-mengde for intakt protokoll innfris. Prøvene ga likevel resultater ved målrettet RNA-sekvensering (23).

Forberedelse av cDNA-bibliotek ble utført ved bruk av PCR. Metoden har både fordeler og ulemper (38). Den er svært sensitiv og har mulighet til å danne tusenvis av kopier av en bestemt sekvens. En ulempe er at det kan forekomme feil inkorporering av nukleotider av DNA-polymerase under amplifisering. PCR er også svært følsom for inhibitorer ettersom reaksjonen er enzymatisk (59). Slike inhibitorer kan for eksempel være salter, kalsiumioner, etanol eller fenol. Tilstedeværelse av inhibitorer kan ha en negativ effekt og kan komme fra selve prøven, ekstraksjon eller miljøet rundt. Konsekvensen er redusert sensitivitet eller falsk-negative resultater. Slike faktorer kan være grunnen til det varierende resultatet fra sekvenseringen i denne studien. Det er derfor viktig at vaskestegene i ekstraksjon og videre forberedelse av bibliotek blir utført nøyaktig.

Før amplifisering er prøvene med det enkeltrådigc cDNA-biblioteket svært svake og kan dermed lett kontamineres. Ettersom PCR er en særdeles sensitiv metode vil det si at en liten kontaminering av prøven, for eksempel DNA-kontaminering, kan gi misvisende resultater. Det er mange steg under forberedelsen av et cDNA-bibliotek som øker sjansen for kontaminering med eget DNA, for eksempel ved å snakke over prøvene. En annen kilde kan

være arbeidsområdet, ettersom prøvene ble behandlet på en amplifiseringszone. For å forhindre kontaminering med genomisk DNA ble det benyttet primere som dekker intron. Det vil si at hvis DNA blir amplifisert, vil sekvensene være lengre ettersom de inneholder intron. Dermed blir ikke DNA-sekvensene sammenstilt mot referansesekvensene i manifest-filen.

Visualiseringen av cDNA-biblioteket med gel-elektroforese viste et tydelig bånd på 200 bp. Dette var forventet, ettersom lengden var forhåndsbestemt. Det ble også observert svake bånd rundt 200 bp. Dette er sekvenser med noe lengre eller kortere lengde. Det ble også observert et bånd på 500 bp. Dette kan skyldes kontaminering med DNA underveis i tillaging av cDNA-biblioteket. Det kan også skyldes at DNA ikke er blitt godt nok fjernet under ekstraksjonen.

Ved forberedelse av cDNA-biblioteket er det viktig med en sensitiv kvantiteringsmetode. Feil fortytning av cDNA-bibliotek, samt feil beregning ved konvertering fra ng/ $\mu$ L til nM kan gi konsekvenser ved sekvensering. For høy eller lav konsentrasjon kan gi over- eller undergruppering i sekvenseringsmaskinen som i ytterste konsekvens kan føre til ingen resultater. Biblioteket ble først fortynnet etter protokoll. Dette viste seg å bli en for sterk fortytning som førte til overgruppering ved sekvensering. Dette kan skyldes kvantiteringsmetoden. Biblioteket ble deretter fortynnet på nytt og sekvensert etter anbefaling fra Illumina. Det er også essensielt at konsentrasjonen til NaOH ikke er mer enn 1 mM i den ferdige løsningen etter fortytning med hybridiseringsbuffer. Høyere konsentrasjon av NaOH i biblioteket inhiberer hybridisering av bibliotek til flowcellen og senker tettheten på klynger under sekvensering.

## 4.5 Sekvensering

Illumina MiSeq har gitt gode resultater i tidligere analyser, spesielt for små og degraderte prøver (23). RNA-sekvensering er fortsatt ikke optimalt for rettsgenetiske prøver, men det er en stor fordel at flere markører kan studeres i én analyse. Sekvensering ved syntese (SBS) – teknologi er en av de mest lovende teknologiene for fremtiden (60).

Forskjellen på kapillær-elektroforese som per dags dato benyttes på rettsgenetiske laboratorier og SBS-teknologi, er at flere prøver og markører kan analyseres samtidig til en lav kostnad (61). Ved få prøver eller markører (~20 eller færre) er ulempen at dette ikke er kostnadseffektivt ved bruk av SBS-teknologi. I dette tilfellet er kapillær-elektroforese et bedre

alternativ. En annen ulempe ved SBS-teknologi er den omfattende og tidkrevende arbeidsflyten, hvor det lett kan forekomme kontaminering. Det kan ta flere dager å forberede et cDNA-bibliotek, mens analyse ved kapillær-elektroforese kan gjøres i løpet av en arbeidsdag.

#### 4.5.1 Spyttprøvene

Spytt kan påvirkes av flere faktorer. For eksempel stimulering og sammensetning kan gi stor variasjon mellom individer og celletyper. Uttrykkelsen av gener kan dermed være personavhengig. Sammenlignet med den høye RNA-konsentrasjonen i spyttprøvene, var antall reads lavere. Dette skyldes antageligvis mengden bakterier i prøvene, ettersom instrumentet for kvantitering ikke er spesifikt for humant RNA. Under amplifisering av cDNA-biblioteket ble det benyttet primere som var spesifikke for de utvalgte markørene, slik at kun humane transkripter ble amplifisert. Readsene var derfor kun fra de utvalgte markørene og ikke bakterier. I en tidligere studie har det også blitt observert færre reads for spyttmarkørene, sammenlignet med andre kroppsvæsker (19).

På dag 0 ble det observert færre reads for alle deltakere sammenlignet med dag 3. Ettersom en økning i reads over tid ikke skal forekomme, kan en mulig årsak være ekstraksjonseffekter. Det kan for eksempel være ufullstendig fjerning av DNase inaktiveringsreagent, som ble anvendt under behandling med DNase. Ved rester i prøven kan PCR inhiberes, som igjen gir færre reads ved sekvensering. Samme effekt kan også forekomme ved overflødig fenol i prøven.

Det ble observert variasjoner i reads mellom deltakerne i studien. Variansanalysen viste at det var en signifikant forskjell i antall reads mellom deltakerne. Dette kan skyldes den generelle variasjonen av genuttrykkelse i spytt, ettersom dette er personavhengig. Dette illustrerer nødvendigheten av å kunne normalisere for å lage en prediksjonsmodell som er gjeldene uavhengig av individ. Sammenlignet med RNA-konsentrasjonen, hvor 67 % av variasjonen kunne forklares av deltaker, er dette betydelig mindre for reads. Kun 3,3 % av variasjonen i reads kan forklares av deltaker. Dette kan også observeres i figurene over utvalgte markører (se **figur 26**, **figur 27**, **figur 28** og **figur 29**), hvor det ser ut til at det er liten forskjell mellom deltakerne.

I samsvar med tidligere studier viser markørene en degradering over tid (28). Plottene viser en jevn degradering, som foregår en kort periode før kurven flater ut. Regresjonsanalysen for alle utvalgte markører viste en signifikant endring i reads over tid, med og uten dag 0. Markørene kan dermed samlet predikere tid i denne studien. Ettersom  $R^2$  var 12,1 % kunne det i videre studier være interessant å inkludere eller ekskludere markører for å observere om modellen hadde blitt bedre. Det ble også observert en signifikant sammenheng mellom endring i reads og tid. Modellen for spyttmarkørene samsvarer med tid i denne studien, men kan ikke benyttes for et annet datasett ettersom normalisering ikke er foretatt. For å kunne normalisere mot hverandre, måtte en av markørene vist en jevn og stabil degradering over tid. Ettersom det ikke var forskjell i degradering mellom markørene og alle viste lignende mønster, kunne ikke dette utføres. En normalisering må forekomme for at modellen skal kunne benyttes i rutinemessig arbeid. En beregning må kunne anvendes for forskjellige personer, samt forskjellig prøvevolum. For å kunne si noe om tid siden avsettelse av spytt, må det være estimeringsmønstre som er repeterbare og ikke påvirket av faktorer som mengde, aktivitet før avsettelse eller individuelle forskjeller.

Fu og Allen rapporterte et funn om at 5'-enden degraderte raskere enn 3'-enden i transkriptet (15). Dette resultatet samsvarte kun delvis med funn i denne studien. Fire utvalgte markører: FDCSP, STATH, HTN1 og KRT6A ble studert. Hos HTN1 og KRT6A var det flest reads for 3'-enden, noe som kan bety at 5'-enden har degradert raskere. Det kan også bety at primeren i 5'-enden for alle markørene har fungert dårlig. For FDCSP og STATH var det flest reads for 5'-enden. I videre studier kan man konsentrere seg om et kortere tidsintervall, samt flere prøver for hver deltaker for å studere degraderingsraten. Ved å studere flere aktuelle markører kan man finne noen som er mer stabile og kan benyttes over en lengre tidsperiode. Andre primer-sett og analysemetoder kan gi annet utfall og flere tester bør gjennomføres for å bekrefte resultatene.

Ettersom markørsettet i denne studien var unikt, var det interessant å se hvilke markører som gav et godt resultat. En god og egnet biomarkør for identifisering av spytt, evalueres etter et høyt antall reads for spyttmarkører i spyttprøvene. For å kunne benyttes i rettsgenetisk analyse, må markøren ha en jevn og målbar degradering over tid. I denne studien var noen markører høyere uttrykt enn andre. Markørene som gav reads  $< 5$  og ble ekskludert var ZG16B, SMR3A, PRH2, PRH1, PRB4, PRB3, PRB2 og PRB1. Markørene BPIFA2 og CST4 gav ingen reads i denne studien. Dette kan skyldes lav uttrykkelse i spytt eller at primerne har fungert dårlig i denne analysen. For de resterende markørene HTN3, MUC7, STATH,

KRT6A, FDCSP og HTN1 ble det observert et høyere antall reads. Dette samsvarer med tidligere forskning hvor markørene har blitt ansett som gode for å identifisere spytt (19, 22, 62). Tidligere forskning har konkludert med at HTN3 og STATH er svært gode markører for identifisering av spytt (20, 28). Hanson et al rapporterte at HTN3 var høyest uttrykt av spyttmarkørene i studien (63). Dette samsvarer med resultatet i denne studien, ettersom HTN3 var høyest uttrykt. I tidligere forskning har STATH vært høyere uttrykt enn HTN3 (64). I motsetning til dette var STATH lavere uttrykt enn HTN3 i denne studien. Dette funnet har også forekommet i en studie fra 2014 (64). KRT6A har også blitt ansett som en god markør for identifisering av spytt i nye og gamle prøver (2, 22). Dette støttes av funn i denne studien ettersom KRT6A var en av markørene som var høyest uttrykt i spyttprøvene. En studie viste at markørene fra slim KRT4, KRT13 og SPR2A oppførte seg bedre i gamle spyttprøver enn STATH og HTN3 (27). Disse markørene var ikke inkludert i denne studien ettersom de dannet kryssreaksjoner med andre celletyper. FDCSP har også i tidligere studie blitt ansett som en god markør for å identifisere spytt (62).

Det ble observert reads fra HBB og HBD i spyttprøvene. Dette er naturlig forekommende ettersom det kan være små sår eller rifter i munn. Dette er også observert i tidligere studier (64). Det observerte nivået av reads fra blodmarkører i spytt var derimot svært lavt sammenlignet med reelle blodprøver og ble ansett å være på nivå med bakgrunnsstøy. Det kan ikke utelukkes at reads fra blodmarkørene skyldes en kontaminering, ettersom det også ble observert i negativ kontroll. Tidligere studier har benyttet en nedre grense for godkjenning av reads (19). Hvordan dette kan implementeres for markører bør studeres nærmere i fremtidige studier.

#### 4.5.2 Blodprøvene

I denne studien var det ønsket å analysere blod uten tilsetninger. Ettersom kapillærprøvene koagulerte raskt, ble blodprøver tilsatt EDTA inkludert. I en studie ble det konkludert med at EDTA ikke hadde en signifikant effekt på RNA-degradering ettersom påvirkning av bivalente kationer på RNaser er mye mindre enn på DNaser eller andre DNA/RNA-modifiserte enzymer (29). Det ble ikke observert noen forskjeller mellom EDTA- og kapillærprøvene i denne studien med hensyn på degradering. Ved å ha kortere tidsintervall, samt flere prøver fra samme deltaker, kan mulige forskjeller avdekkes. Ettersom EDTA-prøvene inneholder stoffer

som hindrer koagulering, kan dette ha en effekt på inntørkingen av det avsatte blodet og dermed kunne ha innvirkning på degradering i kortere tidsintervall.

Det ble observert et generelt høyere antall reads for blodprøvene enn for spyttprøvene. Sammenligning mellom antall reads for blodprøvene og RNA-konsentrasjon viser en antydning til sammenheng. Økt konsentrasjon av RNA samsvarer med et høyere antall reads for prøvene i studien og sammenhengen var signifikant. Dette er ulikt fra spyttprøvene hvor slik sammenheng ikke er observert og skyldes mest sannsynlig innhold av bakterier i spytt. Blodprøvene er dermed bedre egnet for kvantiteringsmetoden benyttet i studien, enn spyttprøvene.

Det ble observert en nedgang i reads ved ekstraksjon nr 2. Dette var dag 7 for EDTA-prøvene og dag 10 for kapillærprøvene. Dette kan muligens skyldes feil under ekstraksjon eller ekstraksjonseffekter som har inhibert PCR, slik at det ble dannet færre reads. Deltaker BG viste en annen trend enn resten av deltakerne, ved at antall reads øker drastisk etter dag 10. Dette kan muligens skyldes koagulering under prøveinnsamling, hvor ulik mengde blod kan ha blitt tilført prøverørene. RNA-konsentrasjon for kapillærprøvene var svært lav etter dag 10. For dag 19 kunne ikke konsentrasjonen detekteres. Dette har ikke gitt noe utslag på antall reads. Dette bekrefter at metoden for kvantitering ikke er sensitiv nok.

Resultatet fra sekvenseringen viste at markørene AMICA1, CD93, CLC, LGALS2, PRF1, S100A12, SLC4A1 og SPTB var lavt uttrykt sammenlignet med HBB, som var svært høyt uttrykt. HBB har i tidligere studier blitt benyttet som markør for blod, samt for å studere degradering over tid (28). Det har derfor vært naturlig å inkludere HBB i denne studien. Utfordringen er at den er høyt uttrykt og derfor dominerer analysen. Når et gen er høyt uttrykt, kan den ta overhånd i analysen ved å bruke opp de tilgjengelige ressursene. Dermed blir det færre reads av de andre markørene. Det er mulig at det hadde vært bedre resultater for de andre markørene hvis ikke HBB ble inkludert. Samtidig viser det seg at HBB er en svært viktig markør i analyse av blod. I en studie ble HBB ekskludert grunnet svært høy uttrykkelse, som førte til en reduksjon i den generelle effektiviteten i multipleksen (63). Samme studie viste også at ALAS2 var høyt uttrykt, som samsvarer med denne studien. ALAS2 og PF4 var de høyst uttrykte blodmarkørene etter ekskludering av HBB og HBD. Det er tidligere konkludert med at markøren PF4 fungerer godt for identifisering av blod (65).

Plottene av HBB og HBD (**figur 35** og **figur 36**) som viste antall reads over tid for 3'- og 5'-enden, hadde varierende resultat. For HBB var det omtrent lik degradering i begge endene



over tid. For HBD var det svært få reads for 5'-enden. Dette skyldes mest sannsynlig at primeren har fungert dårlig i denne analysen, siden lav uttrykkelse observeres for alle målinger. En tidligere studie konkluderte med at ulike markører ekstrahert fra blodflekker, degraderte med ulik rate over tid (32). Det ble også konkludert med at ratio mellom blodspesifikke markører representerer et potensielt verdifullt verktøy for å estimere alder på en blodflekk. En studie fra 2019 benyttet to transkripter for undersøkelse av degradering (15). Det ene transkriptet var forventet å være stabilt, mens det var forventet en endring i det andre transkriptet. Dette er også blitt utført i denne studien. 3'-enden for HBD viste seg å være stabil over tid, mens 3'-enden for HBB viste en endring i reads over tid. Ettersom det var ulik degraderingsrate mellom HBB og HBD, ble det beregnet en ratio mellom disse markørene. Det ble utført en regresjonsanalyse på ratioen over tid, som viste en signifikant korrelasjon. Det vil si at ratioen kan benyttes for å predikere tid. Det gjenstår fortsatt å undersøke hvor presis denne prediksjonen er. Ettersom 35 % av variasjon beskrives av tid, kan det undersøkes om denne modellen kunne vært forbedret. Ved å studere plottet kan det se ut til at det kan fungere fint for å skille mellom nye og gamle prøver, men kortere tidsintervall må muligens undersøkes videre.

Det var kun HBB som viste degradering over tid i denne studien. Resterende markører var for stabile til å kunne observere degradering. Det kunne vært interessant å finne markører som har en raskere degradering, slik at de kan benyttes for kortere tidsintervall. Ved bruk av færre markører kan det være mer gunstig å benytte kapillær-elektroforese, av den grunn at primerkonsentrasjonen kan endres. Det gir mulighet for å endre primerkonsentrasjon for HBB, slik at den ikke dominerer analysen i like stor grad. Dette er ikke mulig ved Illuminas målrettede RNA-sekvensering, ettersom det bestilles ferdige primere. Kapillær-elektroforese er mer sensitiv og kan derfor være en bedre metode ved benyttelse av færre markører.

## 5 Konklusjon og videre arbeid

Målet med denne studien var å undersøke om degradering i avsatt biologisk materiale korrelerer med tid. Det skulle også undersøkes om ulike mRNA har forskjellig hastighet i degradering og om to markører fra samme gen, plassert i hver sin ende (3' og 5') vil vise at 5'-enden degraderer raskere enn 3'-enden.

Samlet viste spyttmarkørene en signifikant nedgang i reads over tid. Dermed kan disse utvalgte markørene prinsipielt benyttes for å predikere tid. Utfordringen med spyttmarkørene er at de degraderer raskt, og derfor bør det undersøkes videre hvilke markører som skal benyttes. For spyttprøvene må man i fremtidige studier se etter andre markører som er mer stabile og muligens konsentrere seg om kortere tidsintervaller, ettersom markørene degraderte hurtig. De utvalgte markørene (HTN1, FDCSP, STATH og KRT6A) viste lignende degraderingsmønster, samt hastighet over tid. Dermed kunne det ikke foretas en normalisering, som er nødvendig for å foreta en beregning som kan benyttes i rutinemessig arbeid. Det ideelle hadde vært å finne en ratio mellom 3'- og 5'-enden i samme gen, for å beregne en ratio. Dette ble ikke funnet i denne studien.

I denne studien var det kun HBB av blodmarkørene som viste en nedgang i reads over tid. Resterende markører var for stabile. Andre markører som degraderer hurtigere, kan derfor undersøkes for benyttelse i rettsgenetikk. Det ble ikke observert noen forskjell mellom 3'- og 5'-enden av transkriptene for blodprøvene. Hypotesen om at ulike mRNA degraderer i ulik hastighet ble det funnet bevis for. HBB degraderte over tid, mens HBD var stabil. Ratioen mellom disse kan dermed benyttes for å finne en degraderingsrate. Hvor presis denne predikeringen vil være, må undersøkes grundigere.

Fremtidige studier kan sette søkelys på de samme markørene, samt inkludere flere målinger fra samme person slik at modellen kan forbedres. Miljøfaktorer som for eksempel fuktighet, temperatur eller substratet som det biologiske materialet er avsatt på er også interessant å inkludere i videre studier. For å kunne gjøre riktige prediksjoner er det nødvendig å ha kunnskap om miljøfaktorer som materialet er eksponert for og statistiske modeller må kunne ta høyde for disse forskjellene.

## Referanseliste

1. An JH, Shin KJ, Yang WI, Lee HY. "Body fluid identification in forensics". *BMB reports*. 2012;45(10):545-53.
2. Zubakov D, Kokshoorn M, Kloosterman A, Kayser M. "New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains". *International journal of legal medicine*. 2009;123(1):71-4.
3. Seksjon for rettsgenetikk i straffesaker. Oslo Universitetssykehus; [03.11.2019]. Tilgjengelig fra: <https://oslo-universitetssykehus.no/avdelinger/klinikk-for-laboratoriemedisin/avdeling-for-rettsmedisinske-fag/seksjon-for-rettsgenetikk-i-straffesaker>.
4. Lodish H BA, Zipursky SL, et al. "Structure of Nucleic Acids". *Molecular Cell Biology*. 4 ed. New York: W. H. Freeman; 2000.
5. Vennemann M, Koppelkamm A. "mRNA profiling in forensic genetics I: Possibilities and limitations". *Forensic science international*. 2010;203(1-3):71-5.
6. Ben-Hur A, Ong CS, Sonnenburg S, Schölkopf B, Rätsch G. "Support Vector Machines and Kernels for Computational Biology". *PLoS Computational Biology*. 2008;4(10):e1000173.
7. Barreau C, Paillard L, Osborne HB. "AU-rich elements and associated factors: are there unifying principles?" *Nucleic acids research*. 2005;33(22):7138-50.
8. Aarnes H. "Proteinsyntese". Institutt for biovitenskap: Universitetet i Oslo 2011 [oppdatert 11. nov. 2019] Tilgjengelig fra: <https://www.mn.uio.no/ibv/tjenester/kunnskap/plantefys/genetikk/protein.html>.
9. Siwaszek A, Ukleja M, Dziembowski A. "Proteins involved in the degradation of cytoplasmic mRNA in the major eukaryotic model systems". *RNA biology*. 2014;11:e34406.
10. Beelman CA, Parker R. "Degradation of mRNA in eukaryotes". *Cell*. 1995;81(2):179-83.
11. Isken O, Maquat LE. "Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function". *Genes & development*. 2007;21(15):1833-56.
12. Simard A-M, DesGroseillers L, Sarafian V. "Assessment of RNA Stability for Age Determination of Body Fluid Stains". *Canadian Society of Forensic Science Journal*. 2012;45:179-94.
13. Fordyce SL, Kampmann ML, van Doorn NL, Gilbert MT. "Long-term RNA persistence in postmortem contexts". *Investigative genetics*. 2013;4(1):7.
14. Kohlmeier F, Schneider PM. "Successful mRNA profiling of 23 years old blood stains". *Forensic science international Genetics*. 2012;6(2):274-6.
15. Fu J, Allen RW. "A method to estimate the age of bloodstains using quantitative PCR". *Forensic science international Genetics*. 2019;39:103-8.

16. Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MSH. "Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells". *DNA Res.* 2009;16(1):45-58.
17. Roeder ADH, C. "mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification". *International journal of legal medicine.* 2013;127(4):707 - 21.
18. Juusola J, Ballantyne J. "Multiplex mRNA profiling for the identification of body fluids". *Forensic science international.* 2005;152(1):1-12.
19. Hanson E, Ingold S, Haas C, Ballantyne J. "Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing". *Forensic science international Genetics.* 2018;34:206-21.
20. Ingold S, Dørum G, Hanson E, Ballard D, Berti A, Gettings KB, et al. "Body fluid identification and assignment to donors using a targeted mRNA massively parallel sequencing approach – results of a second EUROFORGEN / EDNAP collaborative exercise". *Forensic Science International: Genetics.* 2019:102208.
21. Bird TAG, Walton-Williams L, Williams G. "Time since deposition of biological fluids using RNA degradation". *Forensic Science International: Genetics Supplement Series.* 2019.
22. Zubakov D, Hanekamp E, Kokshoorn M, van Ijcken W, Kayser M. "Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples". *International journal of legal medicine.* 2008;122(2):135-42.
23. Ingold S, Dorum G, Hanson E, Berti A, Branicki W, Brito P, et al. "Body fluid identification using a targeted mRNA massively parallel sequencing approach - results of a EUROFORGEN/EDNAP collaborative exercise". *Forensic science international Genetics.* 2018;34:105-15.
24. Haas C, Hanson E, Bar W, Banemann R, Bento AM, Berti A, et al. "mRNA profiling for the identification of blood--results of a collaborative EDNAP exercise". *Forensic science international Genetics.* 2011;5(1):21-6.
25. Roeder AD, Haas C. "mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification". *International journal of legal medicine.* 2013;127(4):707-21.
26. "Biologiske spor". Rettsgenetikk i straffesaker: Oslo Universitetssykehus; [oppdatert 06.12.2019] [07.02.2020]. Tilgjengelig fra: <https://oslo-universitetssykehus.no/fag-og-forskning/nasjonale-og-regionale-tjenester/rettsmedisinske-fag/rettsgenetikk/straffesaker/biologiske-spor>.
27. Lindenbergh A, de Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M, et al. "A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces". *Forensic science international Genetics.* 2012;6(5):565-77.

28. Alshehhi S, Haddrill PR. "Estimating time since deposition using quantification of RNA degradation in body fluid-specific markers". *Forensic science international*. 2019;298:58-63.
29. Bauer M, Polzin S, Patzelt D. "Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains?" *Forensic science international*. 2003;138(1-3):94-103.
30. Asaghiar FE. "Application of forensic RNA analysis as a method for body fluid stain age prediction": *University of Huddersfield*; 2019.
31. Anderson S, Howard B, Hobbs GR, Bishop CP. "A method for determining the age of a bloodstain". *Forensic science international*. 2005;148(1):37-45.
32. Alshehhi S, McCallum NA, Haddrill PR. "Quantification of RNA degradation of blood-specific markers to indicate the age of bloodstains". *Forensic Science International: Genetics Supplement Series*. 2017;6:e453-e5.
33. Setzer M, Juusola J, Ballantyne J. "Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains". *Journal of forensic sciences*. 2008;53(2):296-305.
34. Weinbrecht K, Fu J, Payton M, Allen R. "Time-dependent loss of mRNA transcripts from forensic stains". *Research and Reports in Forensic Medical Science*. 2017;Volume 7:1-12.
35. Kanani P, Shukla YM, Modi AR, Subhash N, Kumar S. "Standardization of an efficient protocol for isolation of RNA from *Cuminum cyminum*". *Journal of King Saud University - Science*. 2019;31(4):1202-7.
36. "DNA-free RNA Isolation". Roche Life Science. Roche Molecular Systems Inc [opdatert 22.12.2017] [03.12.2019]. Tilgjengelig fra: [https://lifescience.roche.com/en\\_no/articles/dna-free-rna-isolation-rna-isolation.html](https://lifescience.roche.com/en_no/articles/dna-free-rna-isolation-rna-isolation.html).
37. Mullis KB. "The unusual origin of the polymerase chain reaction". *Scientific American*. 1990;262(4):56-61, 4-5.
38. Garibyan L, Avashia N. "Polymerase chain reaction". *J Invest Dermatol*. 2013;133(3):1-4.
39. "PCR Basics". Molecular Biology Education: Thermo Fisher Scientific; [05.12.2019]. Tilgjengelig fra: <https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-basics.html>.
40. Haddad F, Baldwin KM. "Reverse transcription of the ribonucleic acid: the first step in RT-PCR assay". *Methods in molecular biology* (Clifton, NJ). 2010;630:261-70.
41. "Reverse Transcription Reaction Setup - Seven Important Considerations". Molecular Biology Education: Thermo Fisher Scientific [15.12.2019]. Tilgjengelig fra: <https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcription-setup.html>.
42. Qubit Fluorometric Quantification: Thermo Fisher Scientific; [03.02.2020]. Tilgjengelig fra: <https://www.thermofisher.com/no/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html>.

43. Maxam AM, Gilbert W. "A new method for sequencing DNA". *Proceedings of the National Academy of Sciences of the United States of America*. 1977;74(2):560-4.
44. Sanger F, Nicklen S, Coulson AR. "DNA sequencing with chain-terminating inhibitors". *Proceedings of the National Academy of Sciences of the United States of America*. 1977;74(12):5463-7.
45. Vidar Beisvåg AKS. "Genomikk: Fra teknologisk utvikling - via forskning - til klinikk": Bioingeniøren; 2013 [oppdatert 04.11.2016]. Tilgjengelig fra: <https://www.bioingenioren.no/fag/fag-i-praksis/2013/genomikk-fra-teknologisk-utvikling---via-forskning---til-klinikk/>.
46. "An introduction to Next-Generation Sequencing Technology". Illumina 2017. [07.12.2019]. Tilgjengelig fra: <https://www.illumina.com/science/technology/next-generation-sequencing.html>.
47. Bronner IF, Quail MA, Turner DJ, Swerdlow H. "Improved Protocols for Illumina Sequencing". *Current protocols in human genetics*. 2014;80:18.2.1-42.
48. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. "Accurate whole human genome sequencing using reversible terminator chemistry". *Nature*. 2008;456(7218):53-9.
49. "Cluster Optimization Overview Guide" .Illumina; 2019 [20.03.2020]. Tilgjengelig fra: [https://support.illumina.com/content/dam/illumina-support/documents/documentation/system\\_documentation/cluster-optimization-overview-guide-1000000071511-00.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cluster-optimization-overview-guide-1000000071511-00.pdf).
50. Haas C, Hanson E, Kratzer A, Bar W, Ballantyne J. "Selection of highly specific and sensitive mRNA biomarkers for the identification of blood". *Forensic science international Genetics*. 2011;5(5):449-58.
51. Juusola J, Ballantyne J. "mRNA profiling for body fluid identification by multiplex quantitative RT-PCR". *Journal of forensic sciences*. 2007;52(6):1252-62.
52. Richard ML, Harper KA, Craig RL, Onorato AJ, Robertson JM, Donfack J. "Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis". *Forensic science international Genetics*. 2012;6(4):452-60.
53. "Converting ng/μl to nM when calculating dsDNA library concentration" Illumina Inc; [oppdatert 19.02.20] [20.05.20]. Tilgjengelig fra: <https://emea.support.illumina.com/bulletins/2016/11/converting-ngl-to-nm-when-calculating-dsdna-library-concentration-.html?langsel=/it>.
54. "MiSeq System - Denature and Dilute Libraries Guide" Illumina 2019 [10.02.20]. Tilgjengelig fra: [https://support.illumina.com/downloads/prepare\\_libraries\\_for\\_sequencing\\_miseq\\_15039740.html](https://support.illumina.com/downloads/prepare_libraries_for_sequencing_miseq_15039740.html).
55. Toni LS, Garcia AM, Jeffrey DA, Jiang X, Stauffer BL, Miyamoto SD, et al. "Optimization of phenol-chloroform RNA extraction". *MethodsX*. 2018;5:599-608.
56. "Qubit RNA HS Assay Kit" Thermo Fisher Scientific [21.04.2020]. Tilgjengelig fra: <https://www.thermofisher.com/order/catalog/product/Q32852#/Q32852>.

57. Sullivan R, Heavey S, Graham DG, Wellman R, Khan S, Thrumurthy S, et al. "An optimised saliva collection method to produce high-yield, high-quality RNA for translational research". *PLOS ONE*. 2020;15(3):e0229791.
58. Lim Y, Totsika M, Morrison M, Punyadeera C. "The saliva microbiome profiles are minimally affected by collection method or DNA extraction protocols". *Scientific Reports*. 2017;7(1):8523.
59. Schrader C, Schielke A, Ellerbreek L, Johne R. "PCR inhibitors – occurrence, properties and removal". 2012;113(5):1014-26.
60. Borsting C, Morling N. "Next generation sequencing and its applications in forensic genetics". *Forensic science international Genetics*. 2015;18:78-89.
61. "Key differences between next-generation sequencing and Sanger sequencing" Illumina Inc; 2020 [01.05.2020]. Tilgjengelig fra: <https://emea.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html?langsel=it>.
62. Park SM, Park SY, Kim JH, Kang TW, Park JL, Woo KM, et al. "Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification". *Forensic science international Genetics*. 2013;7(1):143-50.
63. Hanson E, Ingold S, Haas C, Ballantyne J. "Targeted multiplexed next generation RNA sequencing assay for tissue source determination of forensic samples". *Forensic Science International: Genetics Supplement Series*. 2015;5:e441-e3.
64. Xu Y, Xie J, Cao Y, Zhou H, Ping Y, Chen L, et al. "Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification". *PLOS ONE*. 2014;9(7):e100123.
65. Chong KWY, Wong Y, Ng BK, Thong Z, Syn CK-C. "Development of a RNA profiling assay for biological tissue and body fluid identification". *Forensic Science International: Genetics Supplement Series*. 2015;5:e196-e8.

## Appendiks A: Reagenser og komponenter

Komponent	Produsent	LOT
Acid Phenol:Chlorofom	Invitogen™	Ikke tilgængelig
Acid Phenol:Chloroform	Sigma	MKCL55725
Qubit™ RNA HS Assay kit	Thermo Fisher Scientific	2113240
Qubit™ 1X dsDNA HS Assay Kit	Thermo Fisher Scientific	2054387
GlycoBlue™ Coprecipitant	Thermo Fisher Scientific	00730920
RNAsecure™ Resuspention Solution	Thermo Fisher Scientific	00735452
TURBO DNA-free™ kit	Thermo Fisher Scientific	00812255
MiSeq® Reagent kit v3 (150 cycle) (cartridge + HT1)	Illumina	20426543
MiSeq® Reagent kit v3 (PR2 + flowcelle)	Illumina	20427084
TruSeq Targeted RNA Index Kit (48 indices)	Illumina	20077658
TruSeq Targeted RNA Custom Panel Kit (48 sample)	Illumina	20416823
TruSeq® Targeted RNA CSTM (TOP)	Illumina	20415718
Agarose	Sigma-Aldrich	SLBW0321
Gel Loading Buffer for NA electrophoresis	Sigma-Aldrich	SLBT2484
GelRed® Nucleic Acid Gel Stain	Biotium	18G0327
50-1000 bp DNA markers	Lonza	10-05-16
6X Loading Buffer	Lonza	14-05-28
AMPure XP beads	Beckman Coulter	17746600



## Appendiks B: mRNA-markører for spytt og blod

Biologisk materiale	Navn (engelsk)	Gennavn	Transkript ID	Analyse ID	Kromosom	Start posisjon	Slutt posisjon	Primer-posisjon
Spytt	Keratin 6A	KRT6A	NM_005554	6687637	12	52881874	52881709	5
		KRT6A	NM_005554	6687636	12	52885342	52884917	3
	Mucin 7	MUC7	NM_152291	6928255	4	71346673	71346730	
	Cystatin-S	CST4	NM_001899	6989959	20	23666542	23666488	
	Statherin	STATH	NM_003154	6820238	4	70861710	70864174	5
		STATH	NM_003154	6820236	4	70866669	70867937	3
	Zymogen granule protein 16B	ZG16B	NM_145252	6851652	16	2880454	2880727	5
		ZG16B	NM_145252	6851651	16	2880768	2881822	3
	Proline rich protein 1	PRB1	NM_199354	6724570	12	11506044	11504826	5
		PRB1	NM_199354	6724573	12	11507479	11506908	3
	Histatin 3	HTN3	NM_000200	6974100	4	70894155	70894216	5
		HTN3	NM_000200	6973356	4	70902189	70902246	3
	Histatin 1	HTN1	NM_002159	6827099	4	70921285	70924332	5
		HTN1	NM_002159	6986358	4	70924390	70924454	3
	Proline rich protein 4	PRB4	NM_002723	6982136	12	11461280	11461221	5
		PRB4	NM_002723	6810773	12	11463303	11462311	3
	Proline rich protein 2	PRB2	NM_006248	6652811	12	11545758	11544550	5
		PRB2	NM_006248	6871918	12	11548462	11548407	3
	Proline rich protein 3	PRB3	NM_006249	6653331	12	11420141	11418925	5
		PRB3	NM_006249	6653053	12	11422578	11421571	3
	BPI fold containing family A	BPIFA2	NM_080574	6747581	20	31763333	31765992	
	Proline rich protein haell subfamily 2	PRH2	NM_001110213	6683726	12	11081904	11082898	5
		PRH2	NM_005042	6870401	12	11086355	11086412	3
	Proline rich protein haell subfamily 1	PRH1	NM_006250	6617751	12	11034850	11033614	
	Submaxillary gland androgen regulated protein 3A	SMR3A	NM_012390	6973929	4	71232764	71232819	5
		SMR3A	NM_012390	6985709	4	71227824	71227885	3
	Follicular dendritic cell secreted protein	FDCSP	NM_152997	6676592	4	71091865	71096939	5
FDCSP		NM_152997	6676591	4	71099900	71100843	3	
Hemoglobin-beta	HBB	NM_000518	6976205	11	5247999	5247943	5	
	HBB	NM_000518	6984389	11	5246849	5246789	3	
Junctional adhesion molecule-like	AMICA1	NM_001098526	6645100	11	118067480	118065089	5	
	AMICA1	NM_001098526	6645104	11	118083145	118081400	3	
Cluster of differentiation 93	CD93	NM_012072	6974701	20	23061730	23061675	5	
	CD93	NM_012072	6973146	20	23064962	23064906	3	
5-aminolevulinic acid synthase	ALAS2	NM_001037968	6678577	X	55052283	55051231	5	
	ALAS2	NM_001037968	6678588	X	55041206	55040056	3	
Perforin-1	PRF1	NM_005041	6769798	10	72360156	72358899	5	

		PRF1	NA	12000739	10	72357530	72357511	3
Platelet factor 4	PF4	NM_002619	6990527	4	74847646	74847588	5	
	PF4	NM_002619	6697078	4	74847163	74846983	3	
Spectrin Beta	SPTB	NM_001024858	6758523	14	65289690	65271781	5	
	SPTB	NM_001024858	6758525	14	65234047	65233476	3	
Hemoglobin-delta	HBD	NM_000519	6986645	11	5255787	5255726	5	
	HBD	NM_000519	6984571	11	5254260	5254201	3	
Solute carrier family 4 member 1	SLC4A1	NM_000342	6828569	17	42345449	42340280	5	
	SLC4A1	NM_000342	6828289	17	42328554	42327882	3	
Galectin-2	LGALS2	NM_006498	6738790	22	37966613	37966378	5	
	LGALS2	NM_006498	6738791	22	37975934	37967913	3	
Charcot-Leyden Crystal Galectin	CLC	NM_001828	6665769	19	40224955	40222107	5	
	CLC	NM_001828	6665770	19	40228610	40225683	3	
S100 Calcium Binding Protein A12	S100A12	NM_005621	6810540	1	153346963	153346407	5	
	S100A12	NM_005621	6810539	1	153348058	153347053	3	

## Appendiks C: Resultater fra RNA-quantitering

Resultatene fra RNA-quantitering er oppgitt i ng/μL. Alle negative kontroller var ok.

Deltaker	Prøvenavn	Materiale	Dag	Volum (μL)	RNA-konsentrasjon (ng/μL)
TSA	TSA0	Spytt	0	20	3,6
TSA	TSA1	Spytt	3	20	11,2
TSA	TSA2	Spytt	7	20	9,4
TSA	TSA3	Spytt	14	20	8,2
TSA	TSA4	Spytt	29	20	11,4
TSA	TSA5	Spytt	56	20	9,3
TSA	TSA6	Spytt	143	20	13,5
TSB	TSB0	Spytt	0	20	2,2
TSB	TSB1	Spytt	3	20	5,7
TSB	TSB2	Spytt	7	20	5,0
TSB	TSB3	Spytt	14	20	4,9
TSB	TSB4	Spytt	29	20	7,6
TSB	TSB5	Spytt	56	20	4,8
TSB	TSB6	Spytt	143	20	9,8
TSC	TSC0	Spytt	0	20	2,2
TSC	TSC1	Spytt	3	20	7,9
TSC	TSC2	Spytt	7	20	10,3
TSC	TSC3	Spytt	14	20	10,2
TSC	TSC4	Spytt	29	20	8,8
TSC	TSC5	Spytt	56	20	6,7
TSC	TSC6	Spytt	143	20	7,5
TSD	TSD0	Spytt	0	20	0,9
TSD	TSD1	Spytt	3	20	1,9
TSD	TSD2	Spytt	7	20	3,0
TSD	TSD3	Spytt	14	20	2,1
TSD	TSD4	Spytt	29	20	2,6
TSD	TSD5	Spytt	56	20	Ikke detektert
TSD	TSD6	Spytt	143	20	2,5
TSE	TSE0	Spytt	0	20	Ikke detektert
TSE	TSE1	Spytt	3	20	1,8
TSE	TSE2	Spytt	7	20	1,4
TSE	TSE3	Spytt	14	20	1,4
TSE	TSE4	Spytt	29	20	1,8
TSE	TSE5	Spytt	56	20	3,3
TSE	TSE6	Spytt	143	20	1,4
BA	BA0	Kapillært blod	0	5	1,3
BA	BA1	Kapillært blod	6	5	1,3
BA	BA2	Kapillært blod	10	5	1,3

BB	BB0	Kapillært blod	0	5	1,0
BB	BB1	Kapillært blod	6	5	1,2
BC	BC0	Kapillært blod	0	5	1,2
BC	BC1	Kapillært blod	6	5	1,6
BD	BD0	Kapillært blod	0	5	0,9
BD	BD1	Kapillært blod	6	5	0,8
BD	BD2	Kapillært blod	10	5	0,8
BD	BD3	Kapillært blod	19	5	Ikke detektert
BD	BD4	Kapillært blod	34	5	Ikke detektert
BD	BD5	Kapillært blod	45	5	Ikke detektert
BE	BE0	Kapillært blod	0	5	1,4
BE	BE1	Kapillært blod	6	5	1,4
BE	BE2	Kapillært blod	10	5	1,0
BF	BF0	Kapillært blod	0	5	1,1
BF	BF1	Kapillært blod	6	5	1,2
BF	BF2	Kapillært blod	10	5	Ikke detektert
BG	BG0	Kapillært blod	0	5	1,5
BG	BG1	Kapillært blod	6	5	1,6
BG	BG2	Kapillært blod	10	5	1,4
BG	BG3	Kapillært blod	19	5	Ikke detektert
BG	BG4	Kapillært blod	34	5	1,0
BG	BG5	Kapillært blod	45	5	1,1
BEH	BEH0	Venøst blod m/EDTA	1	20	3,9
BEH	BEH1	Venøst blod m/EDTA	7	20	2,5
BEH	BEH2	Venøst blod m/EDTA	16	20	2,2
BEH	BEH3	Venøst blod m/EDTA	31	20	2,5
BEH	BEH4	Venøst blod m/EDTA	41	20	3,5
BEH	BEH5	Venøst blod m/EDTA	114	20	2,7
BEI	BEI0	Venøst blod m/EDTA	1	20	3,9
BEI	BEI1	Venøst blod m/EDTA	7	20	2,8
BEI	BEI2	Venøst blod m/EDTA	16	20	2,1
BEI	BEI3	Venøst blod m/EDTA	31	20	3,5
BEI	BEI4	Venøst blod m/EDTA	41	20	3,7
BEI	BEI5	Venøst blod m/EDTA	114	20	3,3
BEJ	BEJ0	Venøst blod m/EDTA	1	20	2,6
BEJ	BEJ1	Venøst blod m/EDTA	7	20	4,0
BEJ	BEJ2	Venøst blod m/EDTA	16	20	2,5
BEJ	BEJ3	Venøst blod m/EDTA	31	20	2,2
BEJ	BEJ4	Venøst blod m/EDTA	41	20	3,5
BEJ	BEJ5	Venøst blod m/EDTA	114	20	3,7
BK	BK0	Kapillært blod	7	10	4,8
BK	BK1	Kapillært blod	80	10	2,5

<b>Deltaker</b>	<b>Volum pr prøve (µL)</b>	<b>Gjennomsnittlig RNA-konsentrasjon (ng/µL)</b>
TSA	20	9,51
TSB	20	5,73
TSC	20	7,67
TSD	20	1,75
TSE	20	1,83
BA	5	1,28
BB	5	1,11
BC	5	1,41
BD	5	0,84
BE	5	1,39
BF	5	1,17
BG	5	1,31
BEH	20	2,90
BEI	20	3,22
BEJ	20	3,05
BK	10	3,63

## Appendiks D: Utvalgte prøver for sekvensering

TSA-0	TSB-1	TSC-2	BD-0	BG-3	BEI-0
TSA-1	TSB-2	TSC-3	BD-1	BG-4	BEI-1
TSA-2	TSB-3	TSC-4	BD-2	BG-5	BEI-2
TSA-3	TSB-4	TSC-5	BD-4	BEH-0	BEI-3
TSA-4	TSB-5	TSC-6	BD-5	BEH-1	BEI-5
TSA-5	TSB-6	BA-0	BG-0	BEH-2	BK-1
TSA-6	TSC-0	BA-1	BG-1	BEH-3	BK-2
TSB-0	TSC-1	BA-2	BG-2	BEH-5	NK

## Appendiks E: Gjennomsnittlig antall reads for hvert amplikon

Gennavn	Amplikon ID	Analyse ID	Antall reads (gjenn.snitt)	Posisjon
BPIFA2	BPIFA2.NM_080574.4.5.6747581	6747581	0,14	
CST4	CST4.NM_001899.2.2.6989959	6989959	0,00	
FDCSP	FDCSP.NM_152997.0.1.6676592	6676592	28,62	5
FDCSP	FDCSP.NM_152997.3.4.6676591	6676591	20,95	3
HTN1	HTN1.NM_002159.4.5.6827099	6827099	8,05	5
HTN1	HTN1.NM_002159.5.5.6986358	6986358	25,71	3
HTN3	HTN3.NM_000200.0.0.6974100	6974100	2,38	5
HTN3	HTN3.NM_000200.5.5.6973356	6973356	73,29	3
KRT6A	KRT6A.NM_005554.2.1.6687636	6687636	20,19	5
KRT6A	KRT6A.NM_005554.8.7.6687637	6687637	30,95	3
MUC7	MUC7.NM_152291.2.NA.6928255	6928255	39,90	
PRB1	PRB1.NM_199354.2.1.6724573	6724573	1,43	5
PRB1	PRB1.NM_199354.4.3.6724570	6724570	1,43	3
PRB2	PRB2.NM_006248.0.NA.6871918	6871918	0,95	5
PRB2	PRB2.NM_006248.3.2.6652811	6652811	3,95	3
PRB3	PRB3.NM_006249.1.0.6653053	6653053	2,57	5
PRB3	PRB3.NM_006249.4.3.6653331	6653331	2,10	3
PRB4	PRB4.NM_002723.1.0.6810773	6810773	1,48	5
PRB4	PRB4.NM_002723.2.2.6982136	6982136	1,14	3
PRH1	PRH1.NM_006250.4.3.6617751	6617751	4,24	
PRH2	PRH2.NM_001110213.0.1.6683726	6683726	1,71	5
PRH2	PRH2.NM_005042.2.NA.6870401	6870401	0,00	3
SMR3A	SMR3A.NM_012390.1.0.6985709	6985709	6,52	5
SMR3A	SMR3A.NM_012390.2.2.6973929	6973929	2,43	3
STATH	STATH.NM_003154.0.1.6820238	6820238	22,29	5
STATH	STATH.NM_003154.4.5.6820236	6820236	7,57	3
ZG16B	ZG16B.NM_145252.1.2.6851652	6851652	2,90	5
ZG16B	ZG16B.NM_145252.2.3.6851651	6851651	1,86	3
ALAS2	ALAS2.NM_001037968.3.2.6678577	6678577	133,19	5
ALAS2	ALAS2.NM_001037968.9.8.6678588	6678588	22,12	3
AMICA1	AMICA1.NM_001098526.3.2.6645104	6645104	11,15	5
AMICA1	AMICA1.NM_001098526.9.8.6645100	6645100	8,08	3
CD93	CD93.NM_012072.0.0.6973146	6973146	4,08	5
CD93	CD93.NM_012072.1.1.6974701	6974701	8,88	3
CLC	CLC.NM_001828.1.0.6665770	6665770	5,31	5
CLC	CLC.NM_001828.3.2.6665769	6665769	2,65	3
LGALS2	LGALS2.NM_006498.1.0.6738791	6738791	2,88	5
LGALS2	LGALS2.NM_006498.3.2.6738790	6738790	1,54	3

PF4	PF4.NM_002619.0.0.6990527	6990527	208,27	5
PF4	PF4.NM_002619.2.1.6697078	6697078	113,08	3
PRF1	PRF1.NA.2.2.12000739	12000739	10,12	5
PRF1	PRF1.NM_005041.2.1.6769798	6769798	14,35	3
S100A12	S100A12.NM_005621.1.0.6810539	6810539	21,96	5
S100A12	S100A12.NM_005621.2.1.6810540	6810540	16,00	3
SLC4A1	SLC4A1.NM_000342.1.0.6828569	6828569	10,12	5
SLC4A1	SLC4A1.NM_000342.19.18.6828289	6828289	13,35	3
SPTB	SPTB.NM_001024858.1.0.6758523	6758523	2,88	5
SPTB	SPTB.NM_001024858.30.29.6758525	6758525	2,08	3
HBB	HBB.NM_000518.1.1.6976205	6976205	251799,65	5
HBB	HBB.NM_000518.2.2.6984389	6984389	234675,38	3
HBD	HBD.NM_000519.0.0.6986645	6986645	1,69	5
HBD	HBD.NM_000519.2.2.6984571	6984571	9594,58	3



## Appendiks F: R-koder

Nedenfor er datasett, variabler og benyttet R-kode beskrevet. Navnet på datasettet er oppgitt, samt beskrivelse av variablene.

### Analyse av RNA-kvantiterings resultater

#### **Datasett og variabler:**

- RNAkvant: datasett for totalt antall prøver
  - *Konsentrasjon: kontinuerlig*
  - *Deltaker: kategorisk (navnene til deltakerne)*
- tabspytt: datasett for kun spyttprøvene
- tabedta: datasett for kun EDTA-prøvene
- tabkapillær: datasett for kun kapillærprøvene

#### **ANOVA-analyse: Variasjon i RNA-konsentrasjon mellom deltakere (spytt-, EDTA- og kapillærprøver)**

```
res <- lm (Konsentrasjon ~ Deltaker, data = tabspytt)
anova(res)
```

```
res <- lm (Konsentrasjon ~ Deltaker, data = tabedta)
anova(res)
```

```
res <- lm (Konsentrasjon ~ Deltaker, data = tabkapillær)
anova(res)
```

#### **Regresjonsanalyse: Konsentrasjon ~ Dag (EDTA- og kapillærprøver)**

```
edta.fit = lm (Konsentrasjon ~ Dag, data=tabedta)
summary(edta.fit)
```

```
kap.fit = lm (Konsentrasjon ~ Dag, data=tabkapillær)
summary(kap.fit)
```

### Analyse av sekvenseringsresultater

#### **Datasett og variabler:**

- RNAres2: datasett for alle sekvenseringsresultater for alle prøver
  - *Reads: kontinuerlig*
  - *Tid: kontinuerlig*
  - *Deltaker: kategorisk*
  - *Reads^2: kontinuerlig*
- Kvant\_reads: datasett med alle RNA-konsentrasjonene for alle prøver, sammenlignet med antall reads
  - *RNAkons: kontinuerlig*
  - *Reads\_uten\_HBB\_HBD: kontinuerlig*
- tabS: datasett for kun markører og prøver fra spytt
- tabKap: datasett for kun kapillærprøver
- tabEDTA: datasett for kun EDTA-prøver
- tabS\_nymark: datasett uten spyttmarkører med reads < 5
- tabB\_nymark: datasett uten blodmarkører med reads < 5

- tabBE: datasett for kun EDTA-prøver uten markører med reads < 5
- tabBEN: datasett for kun kapillærprøver uten markører med reads < 5
- tabSRED: datasett for markører og prøver fra spytt, uten dag 0
- tabKRb: datasett over RNA-konsentrasjon og reads for alle blodprøver
- tabKRs: datasett over RNA-konsentrasjon og reads for alle spyttprøver

### Regresjonsanalyse: Tid ~ Reads (spyttprøver, inkludert dag 0)

```
tabS_nymark <- subset(tabS, tabS$Amplikon_ID != "BPIFA2" & tabS$Amplikon_ID != "CST4" &
tabS$Amplikon_ID != "PRB1_5" & tabS$Amplikon_ID != "PRB1_3" & tabS$Amplikon_ID != "PRB2_5" &
tabS$Amplikon_ID != "PRB2_3" & tabS$Amplikon_ID != "PRB3_5" & tabS$Amplikon_ID != "PRB3_3" &
tabS$Amplikon_ID != "PRB4_5" & tabS$Amplikon_ID != "PRB4_3" & tabS$Amplikon_ID != "PRH1" &
tabS$Amplikon_ID != "PRH2_5" & tabS$Amplikon_ID != "PRH2_3" & tabS$Amplikon_ID != "SMR3A_3" &
tabS$Amplikon_ID != "ZG16B_5" & tabS$Amplikon_ID != "ZG16B_3" & tabS$Amplikon_ID != "HTN3_5")
```

```
tabS_nymark <- tabS_nymark %>% mutate (readssq = Reads^2)
```

```
saliva.fit = lm (Tid ~ Reads + readssq, data = tabS_nymark)
summary(saliva.fit)
```

### Regresjonsanalyse: Tid ~ Reads (spyttprøver, uten dag 0)

```
tabSRED <- filter (tabS_nymark, Ftid != "0")
```

```
saliva.fitR = lm(Tid ~ Reads + readssq, data = tabSRED)
summary(saliva.fitR)
```

### Regresjonsanalyse: Tid ~ Endring i reads (spyttprøver)

```
endring.analyse = lm (Tid ~ Endring_reads, data = tabS_nymark)
```

```
summary(endring.analyse)
```

### ANOVA-analyse: Variasjon i antall reads mellom deltakere (spyttprøver)

```
res <- lm (Reads ~ Deltaker, data = tabS_nymark)
anova(res)
```

### Regresjonsanalyse: Tid ~ Reads (EDTA- og kapillærprøver)

```
Kap.blod.fit = lm (Tid ~ Reads, data = tabKap)
summary (Kap.blod.fit)
```

```
EDTAblood.fit = lm (Tid ~ Reads, data = tabEDTA)
summary (EDTAblood.fit)
```

### Regresjonsanalyse: Tid ~ Ratio (HBB/HBD)

```
tabBi <- filter (RNAres2, Materiale=="B", Marker=="B")
```

```

tab.wide <- filter (tabBi, Gennavn=="HBB" | Gennavn=="HBD")
tab.wide <- tab.wide %>% select (Amplikon_ID, Prøve, Reads, Tid)
tab.wide <- spread (tab.wide, Amplikon_ID, Reads)
tab.wide <- tab.wide %>% mutate (rate = HBB_3/HBD_3) #beregner frekvens
ggplot (data = tab.wide, aes (x = Tid, y = rate)) + geom_point() + geom_smooth (method="lm", se = F) +
scale_x_continuous (name = "Ratio")
rate.fit = lm (Tid ~ rate, data = tab.wide)
summary (rate.fit)

```

### **Regresjonsanalyse: Tid ~ Endring i reads (EDTA- og kapillærprøver)**

```

endring.edta = lm (Tid ~ Endring_reads, data = tabBE)
summary(endring.edta)

endring.kap = lm (Tid ~ Endring_reads, data = tabBEN)
summary(endring.kap)

```

### **Regresjonsanalyse: Konsentrasjon ~ Reads (blodprøver)**

```

reg.analyse = lm (Konsentrasjon ~ Reads_uten_HBB_HBD, data=tabKRb)
summary(reg.analyse)

```

## Appendiks G-A: Reads per amplikon i hver blodprøve

Gen-navn	Amplikon ID	Analyse ID	BD0	BG4	BEI 1	BD1	BG5	BEI 2	BD2	BEH 0	BEI 3	BD4	BEH 1	BEI 5	BD5	BEH 2	BG3	BA0	BG0	BEH 3	BK1	BA1	BG1	BEH 5	BK2	BA2	BG2	BEI 0	N K	
ALAS 2	ALAS2.NM_001037968.3.2.6678577	6678577	273	101	131	229	276	122	208	139	85	82	128	139	149	216	44	95	137	78	119	38	156	142	76	88	37	175	0	
ALAS 2	ALAS2.NM_001037968.9.8.6678588	6678588	42	23	29	52	52	2	38	26	8	11	15	23	32	17	23	11	17	9	22	3	17	35	16	20	7	25	1	
AMIC A1	AMICA1.NM_001098526.3.2.6645104	6645104	8	4	21	2	3	10	0	20	14	2	15	17	4	37	1	3	2	17	12	3	3	49	6	7	0	30	0	
AMIC A1	AMICA1.NM_001098526.9.8.6645100	6645100	5	0	17	7	1	23	1	21	10	5	8	8	0	22	1	3	6	12	15	1	0	14	13	3	0	14	0	
BPIF A2	BPIFA2.NM_080574.4.5.6747581	6747581	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD93	CD93.NM_012072.0.0.6973146	6973146	0	0	10	0	0	9	0	8	7	0	6	4	0	12	0	1	0	6	7	3	0	17	7	1	0	8	1	
CD93	CD93.NM_012072.1.1.6974701	6974701	0	4	26	0	4	7	0	19	12	0	15	20	2	16	0	1	0	18	22	1	1	22	15	4	4	18	0	
CLC	CLC.NM_001828.1.0.6665770	6665770	0	1	8	1	0	13	0	3	9	3	5	7	0	17	0	7	0	11	2	2	1	15	9	5	0	19	0	
CLC	CLC.NM_001828.3.2.6665769	6665769	2	0	5	2	0	9	0	4	5	0	0	3	0	6	0	4	0	9	1	1	0	4	0	0	0	14	0	
CST4	CST4.NM_001899.2.2.6989959	6989959	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FDSP P	FDSP.NM_152997.0.1.6676592	6676592	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FDSP P	FDSP.NM_152997.3.4.6676591	6676591	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HBB	HBB.NM_000518.1.1.6976205	6976205	337567	265891	210983	315846	375263	224812	218012	430726	231866	165534	223979	199041	172179	394393	175637	210535	292077	264401	238839	109830	329095	297986	187055	175390	172250	327604	9273	
HBB	HBB.NM_000518.2.2.6984389	6984389	287417	311563	128256	415654	429665	292238	170575	532669	242842	82522	159343	123288	159561	341409	220520	180488	259496	374166	212282	108155	265388	213850	113174	136918	39595	300526	3408	
HBD	HBD.NM_000519.0.0.6986645	6986645	0	0	0	2	3	3	2	0	6	1	3	2	0	7	0	1	0	3	0	0	0	7	0	0	0	4	0	
HBD	HBD.NM_000519.2.2.6984571	6984571	7044	13755	6121	10920	22854	9454	7888	12443	8731	9268	5979	9864	8176	12024	7455	3948	6068	11207	7073	3155	9115	14959	10418	5808	14899	10833	138	
HTN1	HTN1.NM_002159.4.5.6827099	6827099	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HTN1	HTN1.NM_002159.5.5.6986358	6986358	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HTN3	HTN3.NM_000200.0.0.6974100	6974100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HTN3	HTN3.NM_000200.5.5.6973356	6973356	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	





LGAL S2	LGALS2.NM_006498.1.0.6738791	6738791	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LGAL S2	LGALS2.NM_006498.3.2.6738790	6738790	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MUC7	MUC7.NM_152291.2.NA.6928255	6928255	0	125	122	186	42	31	61	35	9	51	34	2	9	0	4	14	8	4	0	12	89	1
PF4	PF4.NM_002619.0.0.6990527	6990527	0	0	1	0	0	0	0	0	1	0	0	2	1	0	0	0	0	0	0	0	0	1
PF4	PF4.NM_002619.2.1.6697078	6697078	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2
PRB1	PRB1.NM_199354.2.1.6724573	6724573	0	5	10	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0
PRB1	PRB1.NM_199354.4.3.6724570	6724570	0	0	16	0	0	11	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0
PRB2	PRB2.NM_006248.0.NA.6871918	6871918	0	2	4	0	3	3	0	0	0	0	4	0	0	1	0	0	0	0	0	0	3	0
PRB2	PRB2.NM_006248.3.2.6652811	6652811	0	6	40	0	0	10	8	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0
PRB3	PRB3.NM_006249.1.0.6653053	6653053	0	0	13	6	1	5	1	0	5	0	0	0	0	3	5	0	0	0	0	0	15	0
PRB3	PRB3.NM_006249.4.3.6653331	6653331	0	0	17	9	0	10	0	0	1	0	0	0	0	0	0	0	0	4	0	3	0	
PRB4	PRB4.NM_002723.1.0.6810773	6810773	0	13	4	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	7	0	
PRB4	PRB4.NM_002723.2.2.6982136	6982136	0	0	4	7	0	4	0	0	0	0	0	0	0	1	0	0	0	1	0	7	0	
PRF1	PRF1.NA.2.2.12000739	12000739	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PRF1	PRF1.NM_005041.2.1.6769798	6769798	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
PRH1	PRH1.NM_006250.4.3.6617751	6617751	0	4	35	4	7	6	0	0	0	3	0	0	0	0	2	0	0	0	0	28	0	
PRH2	PRH2.NM_001110213.0.1.6683726	6683726	0	1	2	15	0	5	0	0	1	2	0	0	0	0	0	0	0	0	0	6	4	0
PRH2	PRH2.NM_005042.2.NA.6870401	6870401	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S100A12	S100A12.NM_005621.1.0.6810539	6810539	0	0	1	17	0	0	0	0	0	0	6	0	0	0	0	0	0	0	4	0	4	0
S100A12	S100A12.NM_005621.2.1.6810540	6810540	0	8	3	4	8	0	0	0	0	6	0	0	0	0	0	0	1	0	0	0	0	0
SLC4A1	SLC4A1.NM_000342.1.0.6828569	6828569	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SLC4A1	SLC4A1.NM_000342.1.9.18.6828289	6828289	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SMR3A	SMR3A.NM_012390.1.0.6985709	6985709	0	21	9	28	2	3	14	7	0	5	8	4	0	0	1	0	0	0	0	0	35	0
SMR3A	SMR3A.NM_012390.2.2.6973929	6973929	0	8	2	16	0	2	12	4	3	0	4	0	0	0	0	0	0	0	0	0	0	0
SPTB	SPTB.NM_001024858.1.0.6758523	6758523	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPTB	SPTB.NM_001024858.30.29.6758525	6758525	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STAT H	STAT.NM_003154.0.1.6820238	6820238	1	29	139	76	5	52	29	10	4	16	4	4	8	1	2	2	0	2	5	2	77	1
STAT H	STAT.NM_003154.4.5.6820236	6820236	1	4	39	33	1	13	18	3	1	10	0	3	4	0	0	0	0	3	1	0	25	0
ZG16B	ZG16B.NM_145252.1.2.6851652	6851652	0	0	5	17	3	0	24	0	0	7	1	0	0	0	0	0	0	0	0	4	0	
ZG16B	ZG16B.NM_145252.2.3.6851651	6851651	0	4	14	0	0	1	4	0	0	5	0	0	0	0	3	0	0	0	5	3	0	







**Norges miljø- og biovitenskapelige universitet**  
Noregs miljø- og biovitenskapelige universitet  
Norwegian University of Life Sciences

Postboks 5003  
NO-1432 Ås  
Norway