

# ROBUSTA BIOMARKÖRER FÖR PREDIKTION AV RISK OCH SJUKDOM

- En utvärdering av reproducerbarheten  
hos de stora kommersiella omik-plattformarna



UMEA UNIVERSITET



# Innehållsförteckning

<b>Sammanfattning</b>	<b>4</b>
<b>Introduktion</b>	<b>5</b>
<b>Utvärdering av kommersiella plattformar</b>	<b>7</b>
Metabolon	7
Nightingale	10
Biocrates	11
Olink	13
Somalogic	14
<b>Slutsatser och rekommendationer</b>	<b>15</b>
<b>Författare</b>	<b>17</b>
<b>Referenslista</b>	<b>17</b>

## Sammanfattning

I och med utveckling inom storskalig analys av blodprover har man idag insett nyttan av att omvandla biobanker med lagrade humanprover till databanker där forskare snabbt kan få tillgång till data för att svara på forskningsfrågor. Problemet är att många av teknikerna för att skapa storskaliga data är semikvantitativa, värdena går inte att relatera till en absolut koncentration och är därmed svåra att slå samman och jämföra över tid. Randomisering, det vill säga att proverna analyseras i slumpvis inbördes ordning, är en av de viktigaste aspekterna för att skapa data som går att slå samman och återanvända för många forskningsfrågor. Detta underlättar korrigeringsvariationer över tid. Utöver detta kan man använda sig av bryggingsprover, QC-prov (kvalitetskontrollprov) eller ankarprover, som analyseras upprepat både inom och mellan analystillfällen, vilket underlättar att lägga samman dataset som analyseras vid olika tillfällen.

Många kommersiella analysplattformar inkluderar ett eget QC-prov i analysen och vissa delar med sig av data för dessa prover. Det vore värdefullt om alla plattformar delade dessa data för kvalitetsutvärdering och eventuell korrigeringsvariationer över tid. För alla semikvantitativa plattformar som undersöktes (Olink, Somalogic, Metabolon och Biocrates) var den tekniska variabiliteten mellan QC-proverna betydligt lägre än variabiliteten mellan analyserade plasmaprover. Detta var tydligast för proteomikplattformarna, vilket antyder att förutsättningarna att upptäcka biologiska skillnader är bättre i proteomikdata. Undantaget från detta är en femte plattform, Nightingale, en kvantitativ men smalare metabolomikmetod som anses generera stabila mätningar.

Vid all utveckling av biomarkörpaneler för att prediktera sjukdom behöver man göra upptäcktsanalyser, sedan valideringsstudier och därefter tester i den situation man tänker att testet ska fungera. De breda omikplattformarna lämpar sig för upptäckt och eventuellt validering, men för det faktiska kliniska testet behövs en kvantitativ analys för att verkligen utvärdera att de proteiner eller metaboliter man vill använda är stabilt uppmätbara och fungerar för att prediktera sjukdom eller risk för sjukdom.

# Introduktion

Frivilligt donerade blodprover för forskning samlas idag in i stor skala och sparas i biobanker världen över. Insamlingarna är ofta befolkningsbaserade och syftet är att de ska kunna användas till många olika forskningsfrågor. I Sverige finns det ett stort antal kohorter där en av de största populationsbaserade kohorterna är Västerbottens hälsoundersökning (VHU). Inom VHU har man samlat in prover från befolkningen vid upprepade hälsokontroller vid 40-60 års ålder ända sedan 1985. Idag har man ett stort antal prover, upp till tre upprepade prover från samma individ (tagna med 10-års intervaller) samt en mogen kohort med många diagnoser.

Det pågår för närvarande stora satsningar med att omvandla biobanker till databanker i många av världens länder. Några exempel är UK biobank, DeCode, Estonia biobank och FinGen. Traditionellt har enskilda forskare fått efterfråga prover för analys från biobankerna. Detta är ofta en tidskrävande process och genererar i allmänhet små dataset. Dessa studier designas ofta som matchade fall/kontroll-studier, vilket innebär att varje individ med en sjukdom (fall) matchas till en individ utan sjukdomen (kontroll). Varje sådan studie utförs vanligen separat och proverna analyseras därför vid olika tillfällen. Tillsammans gör detta att data från sådana studier är svåra att återanvända till nya frågeställningar.

Snabb utveckling inom storskalig analys av molekylära profiler i blod (baserat på olika omik-metoder) har idag möjliggjort att man med en liten volym av blodprovet kan studera ett stort antal molekyler i prover från många individer. Genom att analysera en större mängd prover på en gång, mest optimalt alla prover i en kohort, skapar man data som är återanvändningsbart till många forskningsfrågor. Detta leder i längden till stora kostnads- och tidsbesparingar och är ett effektivt sätt att använda dessa värdefulla biobanksprover. Medan genomik (analys av DNA) genererar förhållandevis stabila data genererar de flesta proteomik- (analys av proteiner) och metabolomik- (analys av metaboliter) metoder relativa data. Dessa data är i motsats till absolut

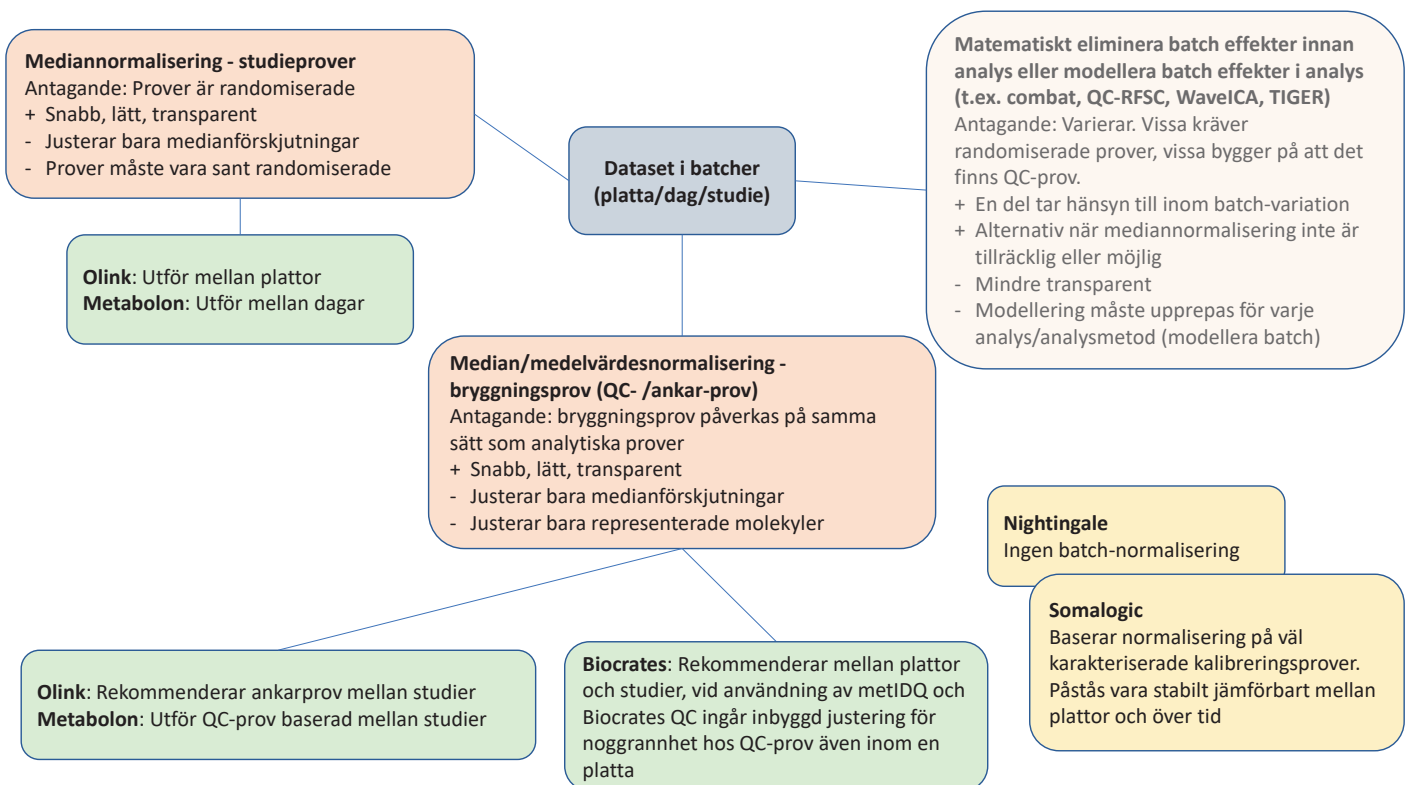
kvantitativa data inte jämförbara mellan studier eller ens mellan molekyler inom en studie vilket gör det svårt att jämföra data som genererats över tid, man får så kallade tekniska batch-effekter. Tekniska batch-effekter kan vara på flera nivåer; allt från platta-till-platta variationer, dag-till-dag variationer till studie-till-studie variationer. För att kunna hantera dessa effekter är det fördelaktigt att utan avbrott analysera stora batcher av prover i slumpmässig ordning (randomiserat), på så sätt hålls den tekniska och biologiska variationen isär och den tekniska variationen kan avlägsnas nedströms om provanalysen. Det är dock inte alltid möjligt att utforma helt balanserade kontinuerliga studier, dels är det både kostsamt och tidskrävande att analysera stora mängder prover, dels kan provinsamlingen ske över tid. Det är därför av praktiska skäl ofta nödvändigt att dela upp analyserna och sedan försöka brygga samman data från olika provanalyser. Ibland vill man även använda sig av redan existerande data. Många av de stora biobankerna världen över har löst den ekonomiska frågan genom stora samarbeten mellan industrin och akademien. UK biobank har t.ex. format ett storskaligt proteomikprojekt tillsammans med 13 biofarmaceutiska företag och FinGen är ett samarbete mellan landets biobanker, sjukhusen, universiteten och farmaindustrin. Samtidigt påstår de större plattformarna för omikanalys att de blir bättre och bättre på att brygga data och det tas fram fler och fler statistiska algoritmer för att avlägsna oönskad teknisk variation (*Han S et al. 2020, Han W et al 2022*).

När prover randomiserats inom en analys är det vanligaste sättet att normalisera data för batch-effekter genom att man antar att medianen är densamma mellan batcher (figur 1). Det är en enkel beräkning som påverkar data minimalt, men som samtidigt endast justerar för medianförskjutningar i data över hela batcher, inte variation inom en batch eller enhetsförskjutningar (där en enhet, t.ex. 1 till 2, inte representerar samma skillnad i två batcher). Att ta hänsyn till variation inom en batch är speciellt viktigt för spektrometrimetoder som man vet lider av kontinuerlig signaldrift. Driften orsakas av en rad fakto-

rer såsom, åldrande kolonner, överföring av prov mellan analyser och temperaturskillnader. Många av de kommersiella analysplattformarna förutsätter att man har randomiserat sina prover och levererar data som är normaliserat mot medianvärdet för varje analyt. När prover ej randomiserats kan man istället använda sig av medianen/medelvärdet på antingen upprepat analyserade kvalitetskontroll-prover (QC-prover) eller ankarprover (ett antal olika studieprover från en batch som analyseras igen i nästa batch) (figur 1). Svårigheten kan vara att få dessa att representera hela datarymden och man kommer inte kunna normalisera molekyler som inte uppmäts i dessa prover. Utöver median/medelvärdesnormalisering finns en rad algoritmer för att eliminera batch-effekter innan statistiskanalys (figur 1) eller för att modellera den under statistiskanalys. Dessa påverkar dock data på ett mer oförutsägbart sätt och man måste vara uppmärksam på att man inte överanpassar data. Det finns en påtaglig risk att man avlägsnar biologiskt intressant variation och även en risk att man skapar artificiella skillnader. Det har de senaste åren kommit ut ett antal artiklar som beskriver, jämför och utvärderar

riskerna med olika algoritmer för att ta bort och ta hänsyn till batch-varianter. Ett idag ofta använt alternativ till att brygga data är metaanalys där man slår samman studieresultat istället för att slå samman data. Den statistiska styrkan begränsas då av delstudiernas mindre storlekar och det finns en kraftig vinst i att istället slå samman data.

I den här rapporten sammanställer och utvärderar vi prestanda vad gäller reproducerbarhet hos fem av de största kommersiella plattformarna för storskalig analys av metaboliter (Nightingale, Metabolon och Biocrates) och proteiner (Olink, Somalogic) i blod. Vi sammanställer de befintliga storskaliga studierna som använt sig av dessa kommersiella plattformar samt utmanar plattformarna genom att vid två upprepade tillfällen ha skickat samma set prover för analys. Detta med avsikten att studera variabilitet inom och mellan studier för de olika metoderna och deras rekommenderade standardmetoder för normalisering. Vi har använt variationskoefficient (CV) som ett mått på hur stabila mätvärdena är för metoderna inom och mellan studie-batcher.



**Figur 1**

Alternativ för att avlägsna batch-effekter (rosa/beige) och vad som görs alternativt rekommenderas av de olika kommersiella plattformarna (grönt/gult).

# Utvärdering av kommersiella plattformar

Fem utvalda kommersiella plattformar för metabolomik- och proteomikanalys utvärderades för prestanda vad gäller reproducerbarhet. Dessa representerar de för närvarande största på marknaden och de har de senaste åren använts av ett flertal stora biobanks för att omvandla prover till data. Vi skickade ett set biobanksprover vid två separata tillfällen (två studie-batcher) till fyra av dessa (Metabolon, Biocrates, Olink och Somalogic) för storskalig analys av metaboliter eller proteiner. Proverna som skickades bestod av 40 biobanksprover samt ett QC-prov. Biobanksproverna härstammade från provsamlingen VHU och spände över olika kön, ålder (40-60 år) och tid i frysförvaring (8-22 år). QC-provet bestod av sammanslagen blodplasma från blodcentralen vid Norrlands Universitetssjukhus (tabell 1). Den femte plattformen, Nightingale, hade av olika praktiska skäl inte möjlighet att analysera proverna inom ramen för projektet. Vi har använt oss av plattformarnas bredaste paneler vars primära syfte är nya upptäckter. En del av plattformarna har även mindre, kvantitativa paneler för riktad analys som ofta används för validering och ytterligare karakterisering av tidigare fynd.

## Metabolon

### Generell beskrivning

Metabolon är en kommersiell plattform för global metabolomikanalys som använder sig av vätskekromatografi masspektrometri (LC-MS). Plattformen har det största metabolitbiblioteket på marknaden, vilket

medför att man kan analysera ca 1400 metaboliter i ett blodplasma-prov. Som kund skickar man prover till plattformen för analys och får tillbaka semikvantitativa, relativa data. I analysen inkluderar Metabolon även ett eget QC-prov (ca vart 8:e prov) samt blankprover för kvalitetskontroll och normalisering av data. Proverna analyseras i små batcher om 36 prover (platt-batch) och fyra sådana batcher analyseras under en dag, det vill säga 144 prover per dag (dag-batch). Den standardiserade normaliseringsmetoden består av att varje metabolits medianvärde för varje dag sätts till värdet 1 (mediannormalisering) (figur 1). Det vill säga, man antar att medianvärdet för varje metabolit för varje dag är detsamma. Detta gör att det är viktigt att randomisera sina prover över dagar. Det är även känt att LC-MS som metod lider av systematisk drift under en analysdag vilket ökar den tekniska variabiliteten i analyserna. Många studier med kopplade prover (t.ex. fallkontroll- eller longitudinella studier) använder sig därför av begränsad randomisering (constrained randomization) för att enklare hålla isär biologisk variation från teknisk variation orsakad av drift. Det innebär att man håller samman prover som hör till samma set men sedan randomiserar körordningen inom varje set och ordningen på seten.

### Utmärkande studier som använt plattformen

En av de större publicerade studierna som baserats på data från Metabolon är en studie på 11,966 prov från EPIC-Norfolk kohorten (Pietzner, M, 2021).

Tabell 1: Provsammanfattning

Analys Plattform	Plasmaprov	QC-prov (Studie batch 1/2)	Plattor eller dagar (studie-batch 1/2)	Ungerfärlig tid mellan analys
Metabolon*	40	8/7	5/4	20 månader
Biocrates	40	3/3	1/1	2 månader
Olink*	40	5/5	5/5	1 månad
Somalogic	40	3/3	1/1	1 månad

\* Analyserades tillsammans med andra studiers prover

I denna studie delades datasetet i två randomiserade sub-kohorter, där de statistiska analysresultaten slutligen slogs ihop med hjälp av metaanalys. Inom varje sub-kohort bryggades data med hjälp av Metabolons QC-prover. Median CV hos det interna upprepade kvalitetsprovet låg på 10%. Flertalet ytterligare storskaliga analyser baserade på Metabolon-plattformen är pågående. Exempelvis har plattformen nyligen använts för att analysera en stor kanadensisk kohort med fokus på åldrande, samt slutit avtal om storskalig analys med bland annat FinGen och TwinsUK.

### Reproducerbarhetsanalys

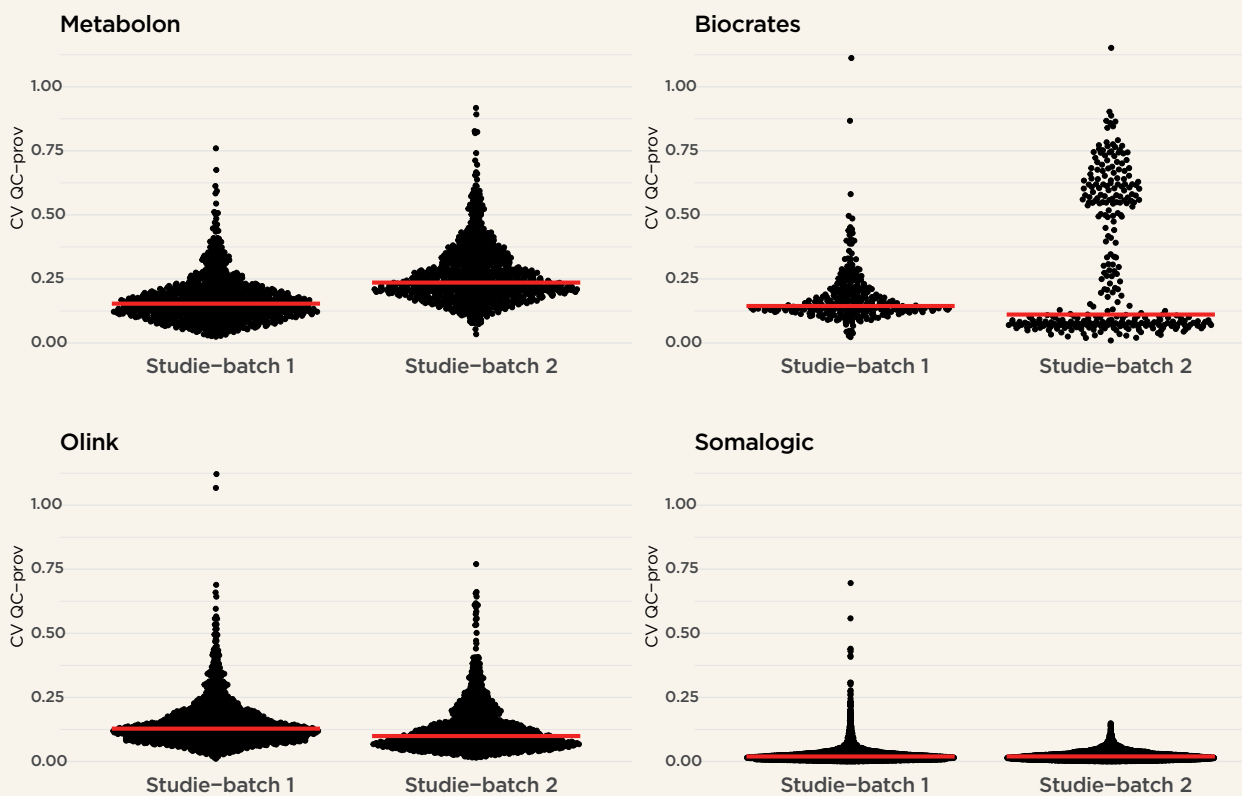
Proverna i vår analys skickades tillsammans med prover tillhörande andra studier. Dessa andra studier innehöll alla prover som var kopplade (antingen fallkontroll- eller longitudinella prover) och var därför begränsat randomiserade. De två analystillfällena utfördes med ca två års mellanrum och själva LC-MS analysen av våra prover var spridd över fem respektive fyra dagar vid de två olika analystillfällena. Standardnormalisering för dag-till-dag variation var som ovan nämnt mediannormalisering baserad på studieprover. För att brygga mellan studier använder Metabolon sig av sitt QC-prov. Antalet detekterade metaboliter varierade mellan studie-batch 1 (1410 st) och studie-batch 2 (1225 st). Av dessa kunde 1083 st respektive 925 st detekteras i minst 75 % av de 40 biobanksprowerna och ansågs därmed stabilt mätbara. Metabolon bryggade även ihop de två studie-batcherna,

vilket resulterade i 911 bryggade metaboliter. CV-värdet är ett mått på variation i en upprepade analys och beräknas genom att man delar standardavvikelsen med medelvärdet och gånger med 100 (det vill säga det är standardavvikelsen relativt medelvärdet presenterat som procent). Ett lågt CV innebär att de upprepade analyserna är lika varandra. Median CV-värdet för alla metaboliter som stabilt kunde detekteras i vårt eget QC-prov i studie-batch 1 och studie-batch 2 var 15.4 % respektive 23.6 % (figur 2, tabell 2). Det höga CV värdet för studie-batch 2 är förmodligen ett resultat av normaliseringen. Prover från studie-batch 2 analyserades tillsammans med prover från tre olika studier (olika provsamlingar och olika design). Inom varje studie är proverna begränsat randomiserade men inte mellan studierna. Detta leder till att antagandet att medianvärdet är samma varje dag inte uppfylls och därför misslyckas normaliseringen. För det bryggade datasetet görs ingen normalisering baserad på studieproverna, utan man använder sig av det upprepade analyserade interna QC-provet. Om man extraherar data för studie-batch 1 och studie-batch 2 från det bryggade datasetet ligger CV-värdena för de enskilda studie-batcherna på 13.6 % respektive 11.3 % (figur 3). Vi ser alltså som väntat ett betydligt lägre CV-värde för studie-batch 2, men även för studie-batch 1. Det antyder att även studie-batch 1, som analyserades tillsammans med prover från enbart en studie, inte uppfyller kravet på randomisering av prover då man även vid den studien använde sig av begränsad randomi-

**Tabell 2:** Median CV inom och mellan studie-batcherna

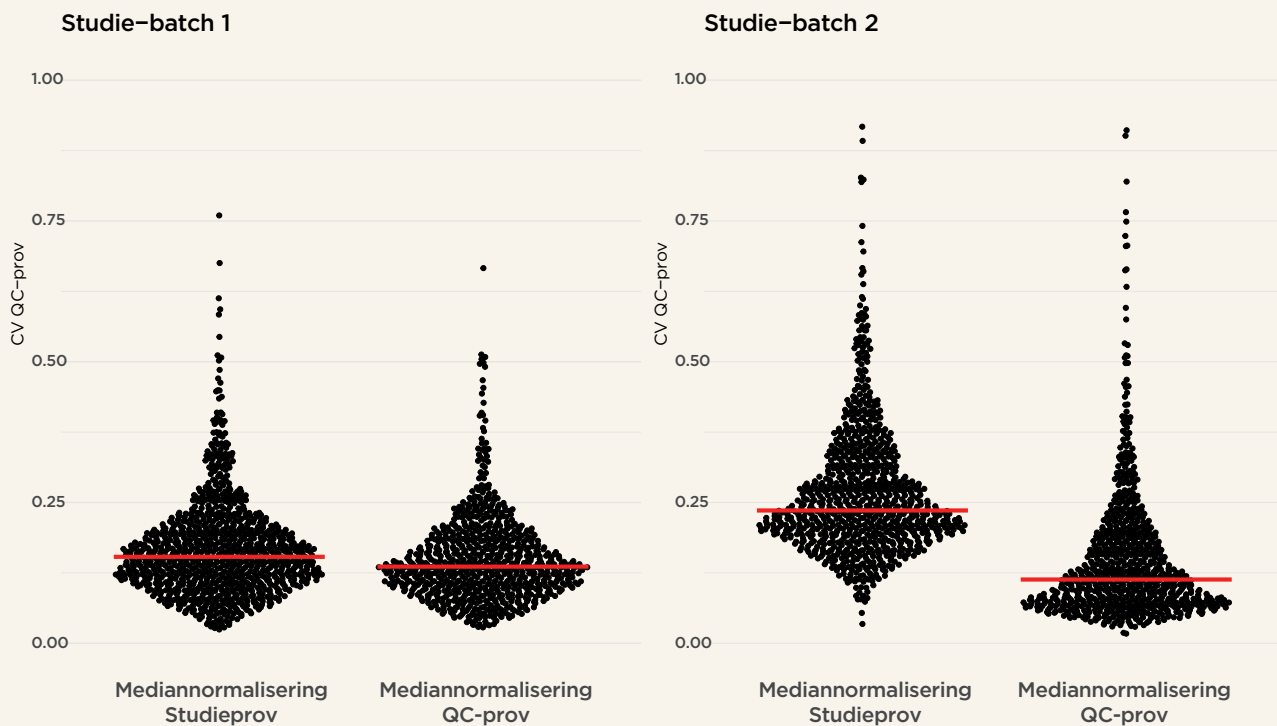
Analys Plattform	Studie-batch 1 (inom studie)		Studie-batch 2 (inom studie)		Bryggat data (mellan studier)		
	CV (%) Plasmaprov	CV (%) QC-prov	CV (%) Plasmaprov	CV (%) QC-prov	CV (%) Plasmaprov	CV (%) QC-prov	Kvot*
Metabolon (QC-norm)	41,6 (40,8)	15,4 (13,6)	44,4 (42,3)	23,6 (11,3)	42,6	14,2	3
Biocrates	34,4	14,4	33,9	11,1	34,0	16,4	2.1
Olink	41,8	12,8	39,4	9,9	41,8	12,3	3.4
Somalogic	29,8	2,5	28,5	2	29,4	3,3	8.9

\* Kvot mellan CV för plasmaprov och CV för QC prov som mått på hur stor den tekniska variationen är jämfört med den totala provvariationen i plasmaprover.



**Figur 2**

CV för QC-provet inom studie-batcher. Inkluderar alla stabilt detekterade metaboliter eller proteiner för varje plattform. Median för CV-värdena är markerat med ett rött streck



**Figur 3**

CV för QC-provet inom Metabolons studie-batcher. Inkluderar alla stabilt detekterade metaboliter. Data är mediannormaliserat antingen baserat på studieproverna eller QC-proverna.

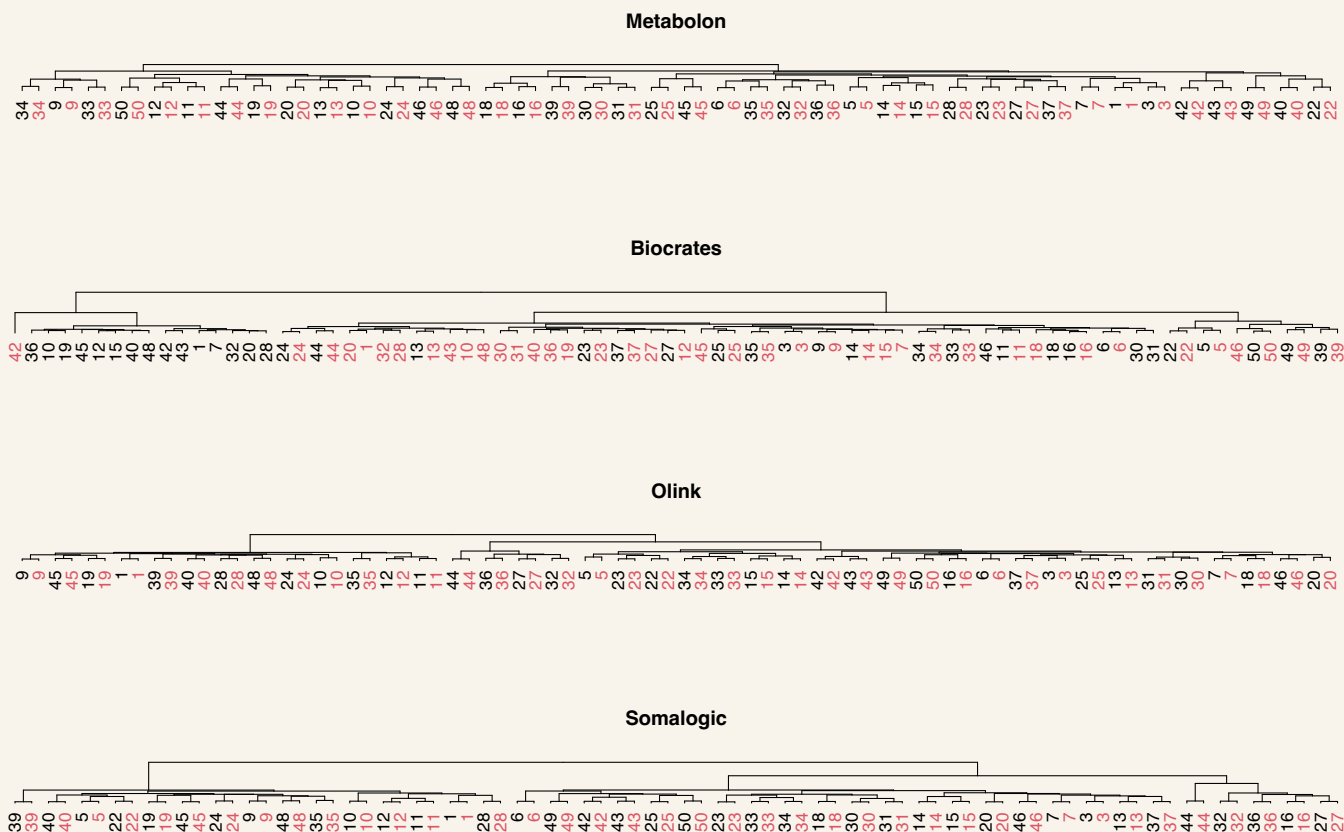
sering. Detta visar på vikten av att sant randomisera sina prover om man ska förlita sig på mediannormalisering baserat på studieprover (oavsett plattform). För hela det bryggade datasetet, bestående av både studie-batch 1 och studie-batch 2, ligger CV-värdet på 14.2%. Detta visar att CV-värdet inte är avsevärt högre mellan studie-batcherna än det är inom studie-batcherna, trots att proverna analyserades med nästan två års mellanrum. Hierarkisk klusteranalys, där prov grupperas efter hur lika de är varandra, visade vidare att de upprepade proverna från studie-batch 1 och studie-batch 2 klustrar närmast varandra (figur 4). Mediannormalisering över dagar tar dock inte hänsyn till inom dag-variation. I ett förenklat försök att skatta inom dag-variation antar vi (pågrund av att proverna är randomiserade) att majoriteten av variationen som orsakar en linjär ökning eller minskning av en metabolit under en dag beror på körordning. Vi observerar då att drygt 30% av metaboliterna fort-

farande är signifikant påverkade av drift inom varje dag efter mediannormaliseringen (figur 5.)

## Nightingale

### Generell beskrivning

Nightingale är en kommersiell plattform för riktad metabolomik som använder sig av NMR (nuklär magnetisk resonans) för att utföra 249 metaboliska mätningar, 168 av dessa är absoluta nivåer medans 81 är kvoter. Metoden är kvantitativ och enligt plattformen finns inget behov av att ha en designad körordning, utan mätningarna är stabila över tid. Plattformen inkluderar två egna separata QC-prov per platta (per 96 prov) för att identifiera eventuella problem med en hel provanalys. Data på dessa prover lämnas dock inte ut. NMR har tidigare varit den tekniska metod med krav på störst volym av blod för analys men Nightingale använder idag endast ca 100 ul,



**Figur 4**

Hierarkisk klusteranalys av de 40 plasma-proverna. Prover från studie-batch 1 indikeras med svart och prover från studie-batch 2 med rött. Som distansmått användes Manhattan och för klustring Wards.D.



analysera upp till 6000 prover per år, vilket innebär att tillgång till flera instrument som används parallellt är nödvändigt för att uppnå hög genomströmning jämförbar med andra plattformar. Metoden är likt Metabolon baserad på masspektrometri (MS) kopplad till LC (vätskekromatografi) och FIA (flödesinjektionsanalys). Biocrates kit är en riktad analys som delvis är kvantitativ (102 metaboliter är klassade som kvantitativa). För ett mindre antal metaboliter har de med en sju-punkts spädningsserie (42 metaboliter), i övrigt använder de en-punkts-kalibrering. Eftersom det är en riktad analys detekteras färre metaboliter än t.ex. Metabolon men med förhoppningen att precisionen är mycket högre. Metoden använder dessutom små provvolymmer vid analys, endast 10 µl. Vid användning av Biocrates största kit (MxP Quant 500) kan upp till 630 metaboliter detekteras. Proverna analyseras på plattor om 96 prov. Utöver spädningsserien inkluderar Biocrates även blank- och nollprov samt ett QC-prov i tre spädningar, varav det ena analyseras 4-5 gånger över en provplatta. Data för dessa prov levereras tillsammans med data från studieprover, vilket ger möjligheten att själv använda dessa vid utvärdering av data. Data visualiseras och analyseras idag i mjukvaran MetIDQ och den rekommenderade normaliseringsmetoden är att medelvärdesnormalisera data på valfritt QC-prov som analyserats minst tre gånger på en platta. För att även innefatta normalisering inom en platta rekommenderas man använda specifikt Biocrates egna QC-prov samt en funktion i MetIDQ för att justera för avvikelser i noggrannhet hos QC-provet under provanalysen.

### **Utmärkande studier som använt plattformen**

Tidigare studier har visat på relativt låg mellan-batch variabilitet för Biocrates på mellan 7-15%. Det har även utförts två ringstudier, där samma prover skickas till en rad olika analyslaboratorier världen över. Studierna använde de mindre Biocrates-kiten (p.180 och p.400) och resultaten visade på relativt god överensstämmelse mellan analyser utförda vid olika analyslaboratorier (*Siskos A P et al. 2017, Thompson J W et al. 2019*). Den tekniska variationen varierade dock mellan metabolitklasser, där de absolut kvantifierade aminosyrorna hade lägst median CV på 10.2%,

medans acylcarnitiner hade en median CV på 38.2% (*Thompson J W et al. 2019*). En ringstudie på Biocrates senast lanserade kit, Mxp Quant 500, pågår men har ännu inte publicerats. Exempel på kohorter där Biocrates använts för provanalys är EPIC, Fenland study och ToMMo (Tohoku Medical Megabank Organization). EPIC har använt Biocrates kit i många år och har analyserat 10 000-tals prover. De har framförallt använt fall/kontroll-design och kitet p.180. Fenland study har över 9000 prover karakteriserade med kit p.180 och ToMMo har data från p.180 men även data från det nyare kitet Mxp Quant 500 från närmare 9000 individer. Även Biocrates själva har analyserat över 1000 friska individer med syftet att definiera normala koncentrationer av upp till 620 metaboliter. De mätningarna har samlats i en avgiftsbelagd databas vars syfte är att kunna användas som kontroll i studier eller för att utvärdera kontrollprover. Förhoppningen är att databasen ska växa och analyskunder som bidrar med data får därför ett reducerat pris.

### **Reproducerbarhetsanalys**

Proverna skickades på analys till ett Biocrates certifierat forskningslaboratorium som använde sig av Biocrates MxP® Quant 500 kit. Med det kittet kan upp till 630 metaboliter detekteras men som namnet antyder förväntas man dock bara kunna detektera närmare 500. I analysen detekterades 355 och 371 av metaboliterna i minst 75% av de 40 biobanksproverna i studie-batch 1 respektive studie-batch 2, vilket var en lägre detektionsnivå än förväntat. Det finns dock indikationer på att utförandet av analyserna varit bristfälligt. Dels observerades hög bakgrundssignal i framförallt studie-batch 1 och dels var det i båda studie-batcherna en hög andel prover (5%) där analysen misslyckats och nästan inga metaboliter detekterades. Vilka prover som misslyckades var olika mellan studiebatch-1 och -2, vilket visar att detta inte var ett problem med provets kvalitet utan med analysutförandet. Trehundrasju metaboliter kunde detekteras i mer än 75% av de 40 proverna i båda batcherna. Median CV-värden för QC-provet baserat på de stabilt detekterade metaboliterna för de två studie-batcherna beräknades till 14.4% respektive 11.1% och till 16.4% mellan de två studie-batcherna

(tabell 2). Även om median CV-värdet för studie-batch 2 var relativt lågt kan man i figur 2 observera ett antal metaboliter i studie-batch 2 med ett förhöjt CV-värde som inte finns i studie-batch 1. Klusteranalys visar även att de upprepade proverna (tekniska replikaten) inte klustrar med varandra (figur 4). Efter en representant för företaget Biocrates granskat den utförda provanalysen visade det sig att provplattan för studie-batch 2 analyserats över två dagar, vilket inte rekommenderas. Konsekvensen är att stora delar av data därför inte kan anses tillförlitligt. Detta visar på vikten av att det laboratorium som anlitas för att utföra provanalysen har genomgått nödvändig träning och har full teknisk förståelse för Biocrates metod.

## Olink

### Generell beskrivning

Olink-plattformen utför analys av proteomet och tekniken bygger på matchad dubbel antikroppsbindning tillsammans med oligonukletid-sekvensering. Plattformens första storskaliga analyspanel, Explore, vilket är den som använts i vår reproducerbarhetsanalys, kan detektera upp till 1463 proteiner (i 1472 analyser). Tillgänglig på marknaden har de även en nyutvecklade analyspanel med kapacitet att detektera upp till ca 3000 proteiner. Metoden använder sig av ytterst små provvolymmer om endast 1  $\mu$ l. Man kan skicka prover för analys hos Olink men de säljer även kit så att metoden kan sättas upp lokalt. Olink rapporterar tillbaka relativa semikvantitativa data. Utöver studieproverna inkluderas även två QC-prov om två och tre replikat och tre blankprov i en provanalys om 96 prover. Data från alla dessa prover levereras vid utförd provanalys. Den rekommenderade normaliseringsmetoden av data mellan provplattor inom ett projekt är mediannormalisering baserat på studieprover. Mellan två studie-batcher rekommenderar Olink att ett antal prover (ankarprover) från den första studie-batchen analyseras igen i den andra studie-batchen och sedan används för normalisering mellan batcher. Om det sedan utförs analys av en tredje studie-batch så väljs nya ankarprover från andra studie-batchen. Plattformen är tydlig på sin hemsida att ankarproverna är nödvändiga för att

kunna jämföra resultat mellan studier. Det är även möjligt att använda QC-prover för detta ändamål men syftet är att ankarproverna ska spänna över en större datarymd och därför utgöra det primära valet. Proverna analyseras i fyra spädningar för att få bättre dynamiskt omfång och över ett flertal sub-paneler.

### Utmärkande studier som använt plattformen

UK-biobank har nyligen analyserat över 54 000 prover med Olink Explore. Proverna analyserades randomiserat inom 2 dataset och i studien observerades en väldigt låg övergripande batch-variation, med ett CV-värde på 6.3 % (Sun B B et al. 2022). De är nu på väg att analysera proverna vidare med den nyutvecklade 3000 protein-panelen. I andra mindre studier har CV-värden kring 8-14% rapporterats. SCALOP (Systematic and Combined AnaLysis of Olink Proteins) är ett samarbetsinitiativ där man samlat data från individer som har både genotypdata och Olink data. Inom initiativet har man idag data från närmare 70 000 individer och 45 kohorter och man har bland annat genomfört en omfattande metaanalys innefattande över 30 000 individer och 90 kardiovaskulära proteiner (Folkersen L. et al. 2020).

### Reproducerbarhetsanalys

Vi skickade prover för Olink-analys till den nationella analysfaciliteten SciLifeLab. Våra prover analyserades vid båda studie-batcherna tillsammans med andra prover. Totalt utgjordes provsetet vid båda tillfällena av fem plattor på 96 prov vardera, där våra plasma-prover och QC-prover var spridda över alla plattor. En större andel av proteinerna i analyspanelen detekterades stabilt i blodplasmaproverna jämfört med metabolomik metoderna, 1221 och 1212 av 1472 analyser detekterades i minst 75 % av proverna i de två studie-batcherna. Median CV-värdet beräknades till 12.8 % och 10.0 % för studie-batch 1 respektive studie-batch 2 (tabell 2). Median CV-värdet mellan de två batcherna var 12.3 %. Klusteranalys visar att motsvarande prover från de två analyserade batcherna klustrar närmast varandra (figur 4). Vi ser med andra ord ingen större ökad variation mellan två studier jämfört med inom en studie. Tiden mellan de två analyserna var dock bara ca 2 månader.

## Somalogic

### Generell beskrivning

Somalogic-plattformen använder sig av en aptamer-baserad proteomikmetod. Aptamerer är korta oligonucleotider som binder med hög affinitet till målproteiner. Liknande Olink analyserar Somalogic prover på plattor om 96 prover i tre spädningar där de inkluderar egna QC-prover, kalibreringsprov och blankprover. Data från alla dessa prover levereras också vid utförd provanalys. Somalogic baserar sin datanormalisering på kalibreringsprov och referensvärden. Referensvärdena har genererats genom upprepad analys av kalibrerings proverna. De skalar även data i förhållande till en amerikansk referenspopulation. Data ska i och med denna normalisering inte vara beroende av övriga prover som analyserats i plattan och resultaten påstås därför vara direkt jämförbara mellan studie-batcher. Den senast tillgängliga versionen av Somalogics SomaScan assay detekterar enligt specifikation 6628 proteiner. Med detta erbjuder Somalogic-plattformen den bredaste proteinpanelen på marknaden, men de kräver också en större provvolym (130  $\mu$ l) och har i dagsläget ett högre pris.

### Utmärkande studier som använt plattformen

Multipla storskaliga studier har använt sig av Somalogics SomaScan assay och de flesta rapporterar låga CV värden på runt 6% över stort antal analyserade plattor. Två av de större studierna som gjorts med SomaScan är analysen av drygt 35 000 isländare inom DeCode och drygt 12,000 individer från Cambridge-området inom The Fenland Study (Ferkingstad E et al. 2021, Pietzner M et al. 2021). I en studie från tidigare i år jämfördes plattformarna Olink och Somalogic (Katz D H et al. 2022). Där noterad man högre precision för data från Somalogic men fler fenotypassociationer för data från Olink. Man visade på vinsten med den högre precisionen hos Somalogic genom att illustrera hur stora studiegrupper man behövde för att hitta en signifikant medelvärdeskillnad beroende på hur hög batch-variationen var. Det krävdes runt tio gånger så många fall vid ett CV på 14% som de såg för Olink jämfört med ett CV på 4% som de såg för Somalogic. Ett område som behöver

redas ut mer är specificiteten hos plattformarna då korrelationen mellan plattformarna är låg för majoriteten av proteinerna.

### Reproducerbarhetsanalys

Somalogics assay innehåller 7596 analyser, varav 7288 utgör humana proteinanalyser av 6628 proteiner. Resterande analyser utgörs av icke-humana proteiner och kontrollanalyser. Biobanksproverna i vår studie analyserades med Somalogic enskilt, dvs inte tillsammans med andra studier. Detta innebär att alla prover fick plats på en platta vid analys av de båda batcherna. Det var extremt få av de 7288 analyserna som var under detektionsnivå i mer än 25% av proverna (1 analys för studie-batch 1 och 12 analyser för studie-batch 2). Ett och samma av de 40 biobanksproverna flaggades för låg kvalitet i båda batcherna. Detta indikerar att provet antingen är av dålig kvalitet eller att det är något i provets karaktär som gör att det är svårt att analysera med Somalogics assay. Ingen av de övriga utvärderade plattformarna har tydligt flaggat för problem med det specifika provet. Median CV-värdet beräknades till 2.5% och 2.0% för de två studie-batcherna vilket var mycket lägre än för någon annan utvärderad plattform (tabell 2). Median CV-värdet mellan de två studie-batcherna beräknades till ett något högre värde av 3.3%, dock fortfarande i den låga nivå som plattformslieferantören själva och andra studier indikerat. Igen var tiden mellan de två studie-batcherna dock kort, ca 1 månad. Klusteranalys visar att motsvarande prover från de två analyserade studie-batcherna klustrar närmast varandra (figur 4).

## Slutsatser och rekommendationer

Data från de utvärderade kommersiella plattformarna innehåller alla en viss batch-variation (CV för QC-Prov) även efter rekommenderad batch-normalisering (tabell 2). Dock är denna variation betydligt lägre än variationen mellan plasmaproverna (CV för plasmaprov), vilket tydliggörs av klusteranalysen (figur 4) där de båda analyserna av ett plasmaprov (studie-batch 1 och 2) hamnar intill varandra, för mätningar från samtliga plattformar utom Biocrates där själva provanalysen utförts undermåligt. Detta betyder att korrelationen mellan samma prov i två olika studie-batcher är högre än korrelationen mellan olika prover i samma studie-batch. Magnituden på batch-variationen för QC-provet inom studie-batch 1 och studie-batch 2 skiljer sig något. Oftast då en högre batch-variation för QC-proven observerats är även variationen mellan plasmaproverna högre. Detta visar på små variationer i hur mycket brus som skapas vid varje provanalystillfälle. Vid bryggning av de två batcherna ökar variationen ytterligare för QC-provet, men bara marginellt och i storleksordning med differensen i batch-variation mellan batch 1 och batch 2. Skillnaden mellan variationen inom QC-proven och inom plasmaproven är en indikation på hur stor andel batch-variationen utgör av den totala variationen och denna skillnad är betydligt större för proteomikmetoderna än metabolomikmetoderna. Detta indikerar att det är lättare att upptäcka relevanta biologiska skillnader i proteomikdata på grund av större differens mellan brus och biologisk variation.

Även om den tekniska batch-variationen är betydligt mindre än variationen mellan plasmaproverna bidrar den till ökad normalvariation, som i sin tur leder till att det är svårare att detektera signifikanta avvikelser mot normalvariationen, vilket ofta är målet i populationsbaserade kohortstudier. Det är därför viktigt att minimera effekten av batch-variation. De flesta kommersiella plattformar rekommenderar att man randomiserar sina prover inom en studie eller provset. Om en kohort analyseras ska alltså alla prover analyseras i slumpmässig ordning. Antagandet att medianvärdet för varje analys är konstant över

batcher (platta/dag/studie) bör därmed vara uppfyllt och man kan använda sig av mediannormalisering baserat på studieprover. Om man behöver dela upp studien/kohorten i delanalyser är det fortfarande fördelaktigt att proverna randomiseras över alla analyser. Det rekommenderas dock att man även inkluderar bryggingsprover i form av QC-prover eller ankarprover. QC-prov kan även komma till användning när inom-batch-effekter behöver åtgärdas, medan ankarprover kan spänna över en större analysbredd. Metabolon har med QC-prov i sin analys som används för bryggning mellan studier. Bryggningen görs av företaget vilket innebär att de måste hålla de data som ska bryggas över tid och en ny bryggning måste göras varje gång ett nytt dataset analyseras, dessutom förlitar man sig på att företagets QC-prov finns tillgängligt över tid. Det går även att utveckla och skicka sitt eget QC prov för analys. Detta är dock, utöver att det ökar totala antalet prover som ska analyseras vilket bidrar till mer brus, en kostnadsfråga. Om vart 8:e prov i en LC-MS analys ska utgöras av ett QC-prov kommer dessa att motsvara över 12% av den totala kostnaden och provantalet. Olink rekommenderar istället ankarprover. Det innebär att inom en studie måste proverna vara randomiserade men mellan dessa studier bryggas man med hjälp av ankarprover. Vid muntlig kommunikation rekommenderade Olink att man inkluderar 30-40 ankarprover och detta oberoende av storleken på vardera delanalys. Biocrates och SomaLogic inkluderar båda egna QC-prov som ingår i batch-normalisering och data för dessa ingår i leveransen medan Nightingale påstår sig vara fri från batch-effekter. Vi noterade att redan vid en liten påverkan på randomiseringen av prover är det bättre att använda sig av medianen för ett bryggingsprov än medianen av studieproverna (från den begränsade randomiseringen som gjordes vid metabolon-analyserna, figur 3).

Om batch-effekter kvarstår, och det inte finns något bryggingsprov eller mer komplexa studier ska bryggas, kan det vara nödvändigt att utvärdera och imple-

mentera vidare normaliseringsmetoder. Detta kan t.ex. göras i den statistiska modelleringen av data, dock med nackdelen att det måste göras om för varje analyssteg och att det kan vara svårt att utföra homogent över flera analyssteg. Alternativt kan försök att matematiskt eliminera batch-effekter innan statistisk analys tillämpas. Det finns ett flertal publicerade och utvärderade algoritmer för båda dessa alternativ som täcker in både de fall då bryggningsprov finns tillgängligt och då det inte finns (*Han S et al. 2020, Han W et al 2022*). Man bör dock vara medveten om att batch-eliminering/modellering i sig påverkar data och kan både minska chanserna att upptäcka sanna skillnader och även skapa konstgjorda skillnader. Man bör därför noga granska effekterna av sin batch-normalisering. Den viktigast åtgärden för att undvika oönskade effekter eller svårigheter med att avlägsna oönskad variation är att randomisera prover. Desto större justeringar och ju mer integrerad den oönskade variationen är med variationen av intresse desto svårare är det att korrekt kompensera för den (*Zhou L. et al. 2019*). Att inkludera upprepade analys av ytterligare ett prov som inte ingår i batch-normaliseringen av data är ett bra verktyg för att utvärdera hur väl normaliseringen fungerat.

Det är även vanligt att man vill slå samman data från olika kohorter. Variabilitet som uppkommer av att proverna tillhör olika provsamlings är ofta större än och skild från den tekniska variabiliteten. Den hanteras vanligtvis i den statistiska analysen. Försök att minska denna typ av variabilitet framöver pågår genom nationella rekommendationer för provhantering och provinsamling.

Nightingale plattformen är undantaget ovan problematik, det är en kvantitativ metod och enligt litteraturen genererar de data i klass med kliniska standardmätningar. De har idag 29 CE-märkta och kliniskt validerade markörer. Somalogic-plattformen uppvisar även betydligt bättre precision i sina mätningar trots att det är en semi-kvantitativ metod. Det finns i litteraturen vissa indikationer på att detta kan vara på viss bekostnad av biologisk variation (*Zhou L et al. 2019*).

Vad gäller överlapp mellan plattformarna är den hög mellan Somalogic och Olink, 85% av proteinerna i Olinks Explore-panel bestående av 1463 proteiner går att återfinna i Somalogics panel på 6628 proteiner. För de proteiner som förekommer på båda plattformarna, är korrelationen mellan plattformarnas mätresultat relativt låg, vilket indikerar att de i många fall mäter olika målprotein och specificiteten för de båda plattformarna behöver utredas mer. För metabolomik-metoderna vi utvärderat är överlappet begränsat och det handlar i detta fall snarare om att plattformarna kompletterar varandra med detektion av olika metaboliter.

Från litteraturen beskrivs även Biocrates, som är en delvis kvantitativ riktad metod för metabolomik, ha relativt hög precision och därmed inte kräva så mycket batch-normalisering. I denna studie observeras dock inte bättre CV-värden för data från Biocrates-plattformen än från Metabolon som är en global semi-kvantitativ metod. Detta är dock den enda metod där provanalyserna inte utfördes av ett standardiserat analysföretag utan proverna har analyserats på ett akademiskt forskningslaboratorium där Biocrates-metoden etablerats. Det har under data-analysarbetet visat sig att handhavandefel påverkat det utvärderade Biocratesdatat och vi kommer därför att fortsätta undersöka Biocrates lämplighet som provanalysmetod. Biocrates utökar stadigt antalet metaboliter i sin plattform varför det också är av värde att upprepa analysen med kommande utvidgad plattform.

De långa ledtiderna som upplevts inom projektet, både avseende att få tillgång till prover och att upprätta alla kontrakt med analysföretagen har belyst den enorma vinsten med att omvandla provbanker (biobanker) till databanker med återanvändningsbart data. Initialt skulle proverna i de två studie-batcherna analyseras med minst 6 månaders mellanrum inom ett 2-årigt projekt. I slutändan kortades detta ner kraftigt för alla utom en plattform (där data för studie-batch 1 redan fanns) och data levererades först i slutet av projekttiden. Att snabbt kunna få tillgång till data är viktigt både för att effektivt kunna

besvara forskningsfrågor men är även en förutsättning för att nytänkande eller högrisk-hypoteser inte ska bortprioriteras på grund av att det är för resurskrävande. Samtidigt är det då viktigt att de storskaliga dataseten genereras på ett adekvat sätt och att användare har förståelse för vad data har för ursprung och därmed möjligheter/begränsningar. Under projektets gång har ett flertal stora studier publicerats som även dom adresserat teknisk variation i data vilket är glädjande. God kunskap om de analytiska plattformarnas möjligheter, men även begränsningar, är oerhört viktigt för att lyckas hitta framtidens biomarkörer för olika sjukdomar.

## Författare

Matilda Rentoft  
Beatrice Melin  
Carl Wibom  
Benny Björkblom

## Referenslista

- Han, S., et al. (2022). "TIGER: Technical variation elimination for metabolomics data using ensemble learning architecture." *Briefings in Bioinformatics* 23(2).
- Han, W. and L. Li (2022). Evaluating and minimizing batch effects in metabolomics. *Mass Spectrometry Reviews*. 41.
- Pietzner, M., et al. (2021). "Plasma metabolites to profile pathways in noncommunicable disease multimorbidity." *Nat Med* 27(3): 471-479.
- Wurtz, P., et al. (2017). "Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies." *Am J Epidemiol* 186(9): 1084-1096.
- Li-Gao, R., et al. (2019). "Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using <sup>1</sup>H NMR spectroscopy." *PLoS ONE* 14(6): e0218549.
- Siskos, A. P., et al. (2017). "Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma." *Anal Chem* 89(1): 656-665.
- Thompson, J. W., et al. (2019). "International Ring Trial of a High Resolution Targeted Metabolomics and Lipidomics Platform for Serum and Plasma Analysis." *Anal Chem* 91(22): 14407-14416.
- Sun, B. B., et al. (2022). "Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants" *BioRxiv*
- Ferkingstad, E., et al. (2021). "Large-scale integration of the plasma proteome with genetics and disease." *Nat Genet* 53(12): 1712-1721.
- Pietzner, M., et al. (2021). "Mapping the proteo-genomic convergence of human diseases." *Science* 374(6569): eabj1541.
- Folkersen, L., et al. (2020). "Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals." *Nat Metab* 2(10): 1135-1148.
- Katz, D. H., et al. (2022). "Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods." *Sci Adv* 8(33): eabm5164.
- Zhou, L., et al. (2019). "Examining the practical limits of batch effect-correction algorithms: When should you care about batch effects?" *J Genet Genomics* 46(9): 433-443.

**SWE**life  
Innovation Programme for Life Science

**VINNOVA**  
Sveriges innovationsmyndighet



UMEÅ UNIVERSITET