

Helse Sør-Øst RHF
Torunn Berge
Postboks 404
2303 HAMAR

Deres ref.: 16/724
Vår ref.: 17/34742-2
Saksbehandler: Tonje Borch
Dato: 28.11.2017

Etablering av nasjonal anonym sekvensvariantdatabase - vurdering av anonymitet

HelseDirektoratet viser til oppdrag fra Helse- og omsorgsdepartementet av 21. november 2017, der departementet ber oss om å svare på Helse Sør-Østs (HSØ) henvendelse om avklaringer knyttet til etablering av anonym sekvensdatabase. Videre viser direktoratet til notat som ble forelagt arbeidsgruppen for etablering av nasjonal anonym sekvensdatabase 31. januar 2018, samt mail av 6. februar fra Torunn Berge med vedlegg til henvendelsen av 21. november.

HSØ oppstiller i sin henvendelse tre problemstillinger de ønsker en juridisk avklaring knyttet til:

1. Vil en sekvensvariant fortsatt anses anonym dersom den kun er observert svært få ganger (<5)?
2. Kan kobling til tekniske metadata som tidsperiode og laboratorium understøtte kravet om anonymitet?
3. Er databasen å anse som anonym dersom sekvensvarianter kobles til en klinisk fortolkning/indikasjon?

Sakens faktum

Utgangspunktet for henvendelsen er HODs oppdrag til HSØ om å etablere en anonym frekvensdatabase over sekvensvarianter. I brevet fra HSØ omtales sekvensvariasjon primært som de endringer i DNA-sekvensen som omfatter ett til noen få nukleotider.

En norsk frekvensdatabase skal være et verktøy til bruk i kvalitetssikring av genetisk diagnostikk. I henvendelsen fra HSØ er innholdet i frekvensdatabasen beskrevet på følgende måte: «*Datasettet i den norske variantdatabasen vil være aggregert fra sekvensinformasjon innhentet fra norske pasienter og pårørende (>5000 individer og stadig voksende) med allel-frekvenser estimert fra dette aggregatet. Dette innebærer at innholdet i databasen vil være av dynamisk karakter så lenge ny informasjon blir lagt til.*

Den enkleste versjonen av databasen vil bygge på såkalte frekvensstudier der man enkelt forklart teller antall forekomster av en bestemt sekvensvariant i et innsamlet gendatamateriale. En populasjonsspesifikk variantdatabase slik den er beskrevet i oppdraget vil derfor kunne si noe om

HelseDirektoratet

Avdeling bioteknologi og helserett
Tonje Borch, tlf.: +4724163946

Postboks 7000 St. Olavs plass, 0130 Oslo • Besøksadresse: Universitetsgata 2, Oslo • Tlf.: 810 20 050
Faks: 24 16 30 01 • Org.nr.: 983 544 622 • postmottak@helsedir.no • www.helseDirektoratet.no

utbredelsen av bestemte genvarianter i den norske befolkningen».

Det framgår av henvendelsen at det i tillegg er ønskelig og nødvendig å koble tekniske data til sekvensvariantene for å understøtte kvalitetssikring og vedlikehold av databasen. Dette vil f.eks. være informasjon om type sekvenseringsplattform, dato for analyse og lokalisering av sekvenseringslaboratorium.

Innhold og bruk av databasen er nærmere beskrevet i notat til møte RHFenes arbeidsgruppe 31.01.18., samt i vedlegg til henvendelse mottatt i mail 6. februar. Disse dokumentene bekrefter at den anonyme frekvensdatabasen skal være en referanse med norske variantfrekvenser. Ved trio-sekvensering (foreldre+ barn) er det bare data fra foreldre som vil inngå.

Rettslig grunnlag

Pasientjournalloven, helseregisterloven og personopplysningsloven oppstiller ingen krav til behandling av anonyme opplysninger, og en korrekt identifisering av hva som kan defineres som anonyme opplysninger blir derfor viktig for å ivareta den registrertes personvern. Å anonymisere data er utfordrende, og det er mer utfordrende i dag enn det var tidligere. Det enorme tilfanget av offentlig tilgjengelige data, kombinert med tilgang til stadig billigere og mer kraftfull analyseteknologi, har bidratt til å øke faren for reidentifisering.¹

Ved behandling av anonyme opplysninger skal det ikke, direkte eller indirekte, være mulig å spore opplysningene tilbake til de enkeltpersoner opplysningene knytter seg til. Det fremgår av Helseregisterlovens forarbeider og av fortalet til ny personvernforordning (avsnitt 26) at det bør tas hensyn til alle midler som det med rimelighet kan tenkes at den behandlingsansvarlige eller en annen person kan ta i bruk for å identifisere vedkommende direkte eller indirekte. For å fastslå om midler med rimelighet kan tenkes å bli tatt bruk for å identifisere den fysiske personen bør det tas hensyn til alle objektive faktorer, f.eks. kostnadene for og tiden som er nødvendig for å foreta identifikasjonen. Det må tas hensyn til teknologien som er tilgjengelig på behandlingstidspunktet, samt den teknologiske utvikling.

I praksis er det i Norge lagt til grunn at opplysningene først kan anses som anonyme dersom de kan knyttes til en av fire eller fem individer. Dette er en streng fortolkning og samsvarer med de anonymiseringsprinsippene SSB anvender i sin statistikkfremstilling.

Det er ifølge Datatilsynets rapport: "Personvernutfordringer ved genetiske undersøkelser" gjort rede for genetiske undersøkelser status i nasjonalt og internasjonalt regelverk. Med utgangspunkt i et sammensatt rettskildebilde har Datatilsynet konkludert med at genetiske undersøkelser har en særskilt status som spesielt sensitive opplysninger.²

¹ <https://www.datatilsynet.no/globalassets/global/regelverk-skiema/veiledere/anonymisering-veileder-041115.pdf>

² https://www.datatilsynet.no/globalassets/global/om-personvern/rapporter/genetiske_undersokelser_rapport_v2.pdf

Hesledirektoratets vurdering

Til spm 1

Vil en sekvensvariant fortsatt anses anonym dersom den kun er observert svært få ganger (<5)?

Som nevnt ovenfor tilsier dagens praksis at opplysningene som et utgangspunkt først kan anses som anonyme dersom de kan knyttes til en av fire eller fem individer.

Teoretisk kan det være mulig å identifisere et individ ut fra beskrivelse av en sjelden variant. Det kan for eksempel være mulig hvis den samme varianten finnes i en annen database med *personidentifiserbare* opplysninger, eller man har tilgang til personidentifiserbare kliniske data som utvetydig kan knyttes til den aktuelle varianten. Dette medfører at en sekvensvariant ikke vil anses for å være anonym dersom den kun er observert svært få ganger (<5).

Det fremgår imidlertid av notatet til møtet i arbeidsgruppa 31.01.18 at det er mulig å innrette databasen slik at den ikke oppgir eksakt frekvens for eller inneholder de sjeldne variantene. Cut-off for hva som anses som sjeldent kan for eksempel settes til 4/5 alleler eller færre i den aktuelle sykdomsgruppen.

Slik vi forstår dette, vil en slik database ikke ha informasjon om eksakt hvor mange ganger en sjelden variant er observert og heller ikke nærmere beskrivelse av den sjeldne varianten (jf. begrepet «non disclosed, rare variant»). Et søk i, eller filtrering mot, databasen vil kunne oppgi at det finnes en sjelden variant i den aktuelle posisjonen, og at det er observert 4/5 alleler eller færre. Vi kan ikke se at det vil være mulig å identifisere enkeltindivider på bakgrunn av dette, og mener at et register som er innrettet på denne måten må kunne sies å være anonymt. Det er vår oppfatning at en slik innretning vil være i tråd med dagens prinsipper om anonymisering.

Til spm 2

Kan kobling til tekniske metadata som tidsperiode og laboratorium understøtte kravet om anonymitet?

Det er i tillegg stilt spørsmål om inkludering av tekniske data. Vi oppfatter at dette er viktig og nødvendig for å sikre god kvalitet i datasettet som til enhver tid ligger i databasen. Notatet til arbeidsgruppa gir flere eksempler på tekniske data som er ønskelig å inkludere i databasen. Så lenge de sjeldne variantene håndteres slik det er beskrevet i notatet til arbeidsgruppa ser vi ikke at rent tekniske opplysninger om sekvenseringsplattform, protokoll for prøvepreparering, dekningsgrad, dato, laboratorium etc. vil bidra til at enkeltindivider kan identifiseres. Det vil imidlertid være avgjørende at inkludering av slike data ikke gjør det mulig å re-identifisere enkeltindivider i datasettet. For å sikre at dette ikke skjer vil direktoratet foreslå at tekniske data knyttes til databasene som den enkelte avdelingen har over sine varianter, og at det settes tekniske krav til datasett som inkluderes i den anonyme frekvensdatabasen som må oppfylles.

Til spm 3

Er databasen å anse som anonym dersom sekvensvarianter kobles til en klinisk fortolkning/indikasjon?

Som ovenfor vil utgangspunktet være at det er faren for reidentifikasjon av enkeltindivider som avgjør om databasen vil være anonym eller ikke. Om man kobler variantene til en klinisk

fortolkning/indikasjon vil dette fortsatt kunne være anonymt dersom samme fortolkning knyttes til varianter som er observert 4-5 ganger eller flere.

Helsedirektoratets videre oppfølging av de nevnte problemstillinger

HSØ sier i vedlegg til henvendelsen; "*for at en nasjonal database skal ha størst mulig klinisk verdi, er det viktig at det kan gjennomføres aggregering av variantfrekvenser slik at databasen får bredest mulig nasjonal representasjon. Dette vil innebære at det må legges til rette for innhenting av data, med godkjent prosedyre, fra alle nasjonale leverandører av genetisk data. Skulle denne leveransen blir begrenset på lokalt nivå, vil også databasen kliniske betydning bli betydelig redusert. Det følger av dette at dersom hver enkelt lokale kilde pålegges en nedre terskelverdi eller rapporteringsgrense, vil det samlet sett være et stort antall varianter som ikke vil fanges opp, fordi de ikke kan summeres nasjonalt. I dette tilfelle vil databasen miste mye av sin nytteverdi for klinisk anvendelse.*" Dette betyr, slik Helsedirektoratet oppfatter det, at en anonymitetsvurdering som baseres på at variantdatabasen først anses som anonym dersom variantene kan knyttes til en av fire eller fem individer, vil påvirke databasens nytteverdi. Etter vår oppfatning vil dette kunne avhenge av hvordan data samles inn. Hvis alt samles inn nasjonalt uten tekniske data og varianter færre enn 4/5 vil databasen kunne være nyttig som verktøy til bruk i kvalitetssikring av diagnostikk.

Vi ser imidlertid at en frekvensdatabase vil kunne ha større potensiale dersom man også kan lagre varianter hvor cut off ikke er 4/5, og hvor man kan knytte flere variabler til dataene. Spørsmålet blir imidlertid da om databasen fortsatt kan sies å være anonym. Det er Helsedirektoratets oppfatning at man bør se nærmere på hvilke krav til anonymisering man skal legge til grunn når det gjelder genetiske varianter. Særlig når det, som her, er snakk om genetiske varianter som i seg selv ikke er personidentifiserbare med mindre de har andre variabler knyttet til seg. Den rettslige vurderingen er, som nevnt ovenfor, knyttet til hva som skal til for å identifisere et enkelt individ ut i fra de opplysninger man har tilgjengelig i registeret og tilfanget av offentlig tilgjengelige data, kombinert med tilgang til stadig billigere og mer kraftfull analyseteknologi. Vurderingen vil videre være knyttet til en vurdering av alle midler som det med rimelighet kan tenkes at den behandlingsansvarlige eller en annen person kan ta i bruk for å identifisere vedkommende direkte eller indirekte. Konklusjonen vil ha betydning for hvordan vi i fremtiden behandler genetiske opplysninger i Norge, og bør etter Helsedirektoratets oppfatning bygges på en grundig utredning. Direktoratet vil prioritere arbeidet med en slik utredning våren 2018.

Vennlig hilsen

Anne Louise Valle e.f.
avdelingsdirektør

Tonje Borch
seniorrådgiver

Dokumentet er godkjent elektronisk

Kopi:
Helse- og omsorgsdepartementet