

Nomenklatur, tolkning og klassifisering av sekvensvarianter

NPU-kodeverket

Bjørn Ivar Haukanes (Haukeland)

Liss Anne Solberg Lavik (St. Olav)

Hilde Monica Stensland/Torill Fagerheim (UNN)

Kristian Tveten (Telemark)

Helge Rootwelt (MB-OUS)

Mari Ann Kulseth (EEG-OUS)

Lars Retterstøl (OUS)

Torunn Fiskerstrand (Haukeland)

Nomenklatur/variant vurdering

NPU kodeverket

- ✓ Nomenklatur
- ✓ Klassifiseringssystemet
- ✓ Bruk av frekvensdata
- ✓ Huskeliste for variantvurdering
- ✓ Klasse 2
- ✓ Klasse 3 (VUS)
 - ✓ Rapportering
 - ✓ Videre utredning
- ✓ Klasse 4
- ✓ Klasse 5
- ✓ HTS og variantvurdering
- ✓ Fremtiden i genetisk diagnostiske labber

Nomenklatur

GUIDELINES FOR GENE NOMENCLATURE

HGNC - Guidelines for human gene nomenclature



www.genenames.org

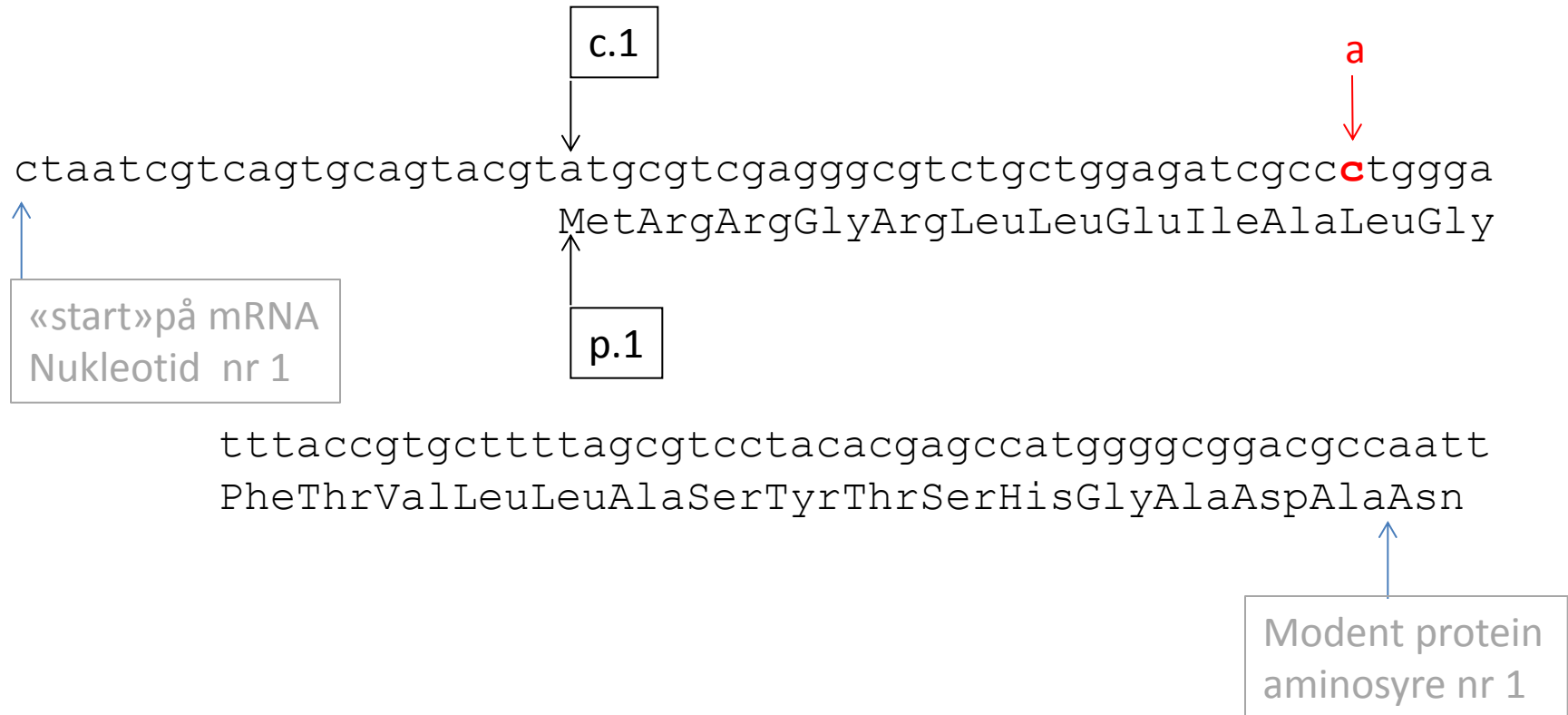
GUIDELINES FOR MUTATION NOMENCLATURE

HGVS - Nomenclature for the description of sequence variations



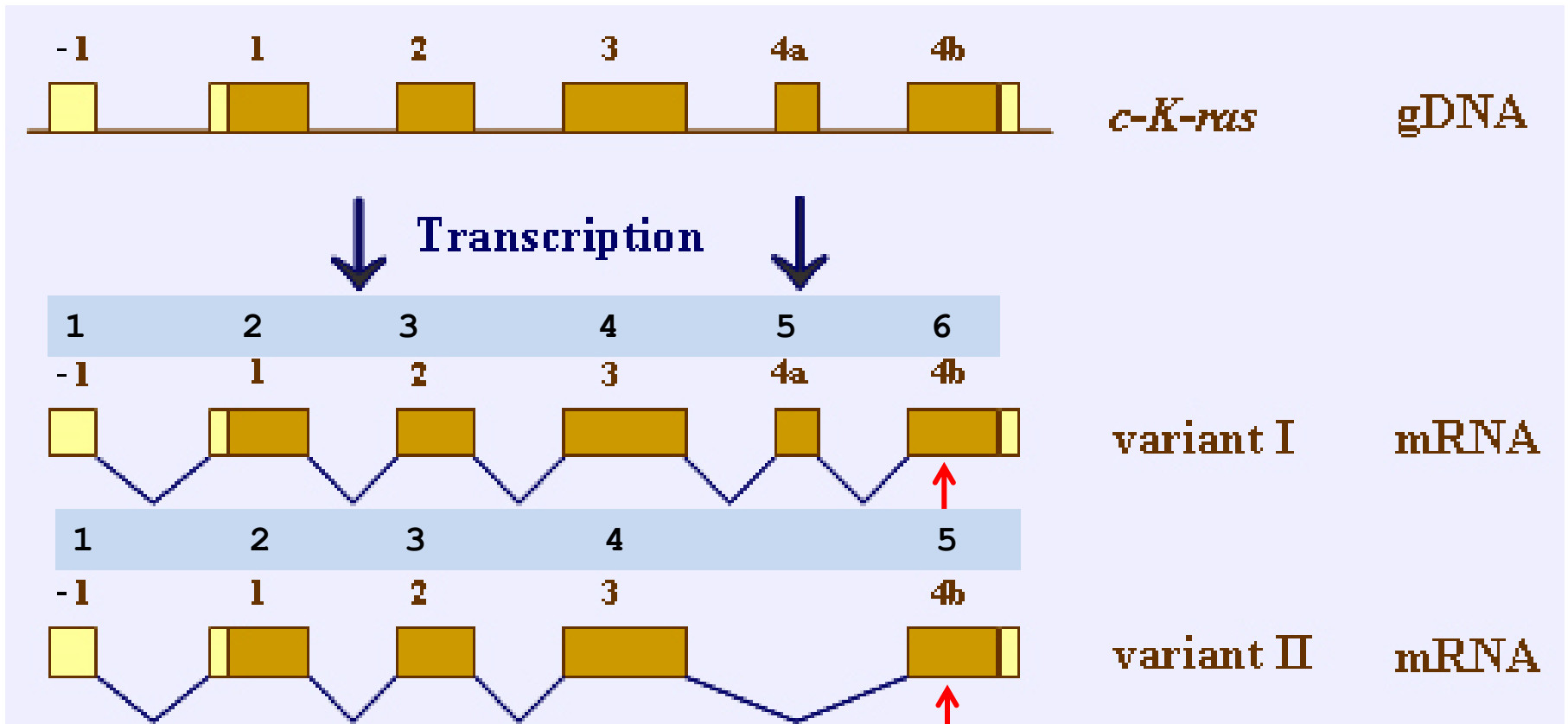
www.hgvs.org/mutnomen/

HGVS nomenklatur versus «eldre» nomenklatur



Bruk HGVS nomenklatur, men inkluder i svarrapporten «eldre» nomenklatur

Transkript ID (Ref Seq) + variant = forståelig



Unngå bruk av ekson nummer i svaret!

Eventuelt benytt systematisk nummerering og angi tellemåte i svaret.

Nonsens variant

~~X~~ * Ter

HGVS:

The three-letter amino acid code is preferred, with "**Ter**" or "*****" designating a translation termination codon.

- ✓ Bruk HGVS nomenklatur
- ✓ Inkluder RefSeqID i svarrapporten
- ✓ Vær forsiktig ved bruk av ekson nummer
- ✓ Bruk tre-bokstav-koden på aminosyrer
- ✓ Bruk «*» eller «Ter» for nonsensvarianter



Mutasjon versus sekvensvariant

HGVS:

Mutation and polymorphism

In some disciplines the term "**mutation**" is used to indicate "*a change*" while in other disciplines it is used to indicate "*a disease-causing change*". Similarly, the term "**polymorphism**" is used both to indicate "*a non disease-causing change*" or "*a change found at a frequency of 1% or higher in the population*". To prevent this confusion we do not use the terms mutation and polymorphism (including SNP or Single Nucleotide Polymorphism) but use neutral terms like "**sequence variant**", "**alteration**" and "**allelic variant**".

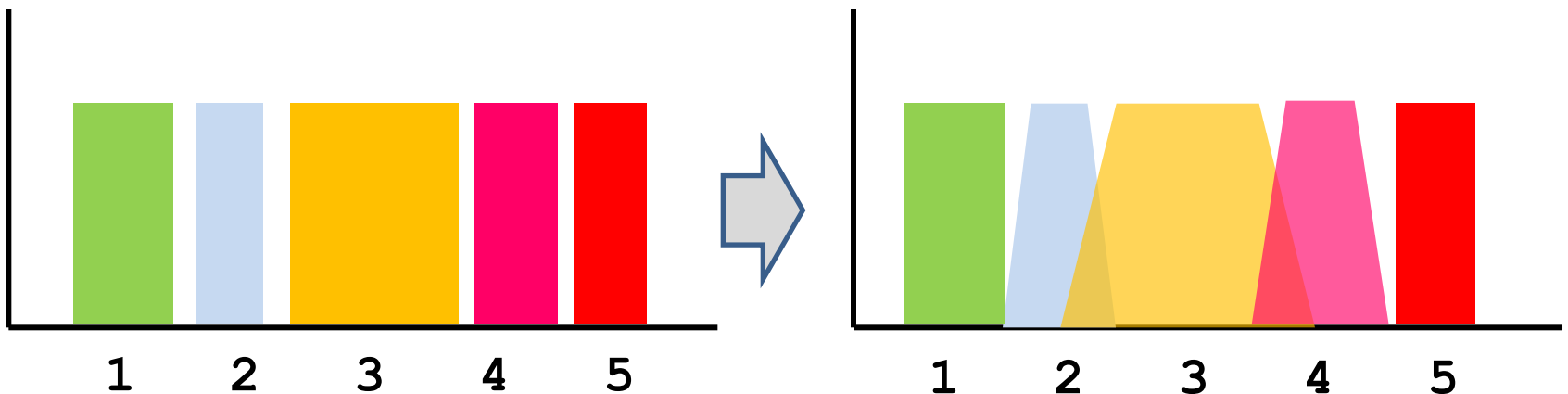
Haukeland:

Bruker «mutasjon» så lenge Gunnar er sjefen!!!



Klassifisering av varianter

- 1) Normalvariant
- 2) Sannsynlig normalvariant
- 3) Usikker variant (VUS)
- 4) Sannsynlig patogen variant
- 5) Patogen variant



Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing.

Bell et al., 2011 Sci Transl Med. 3; 65ra4

Bærertesting for 448 alvorlige pediatrike recessive sykdommer

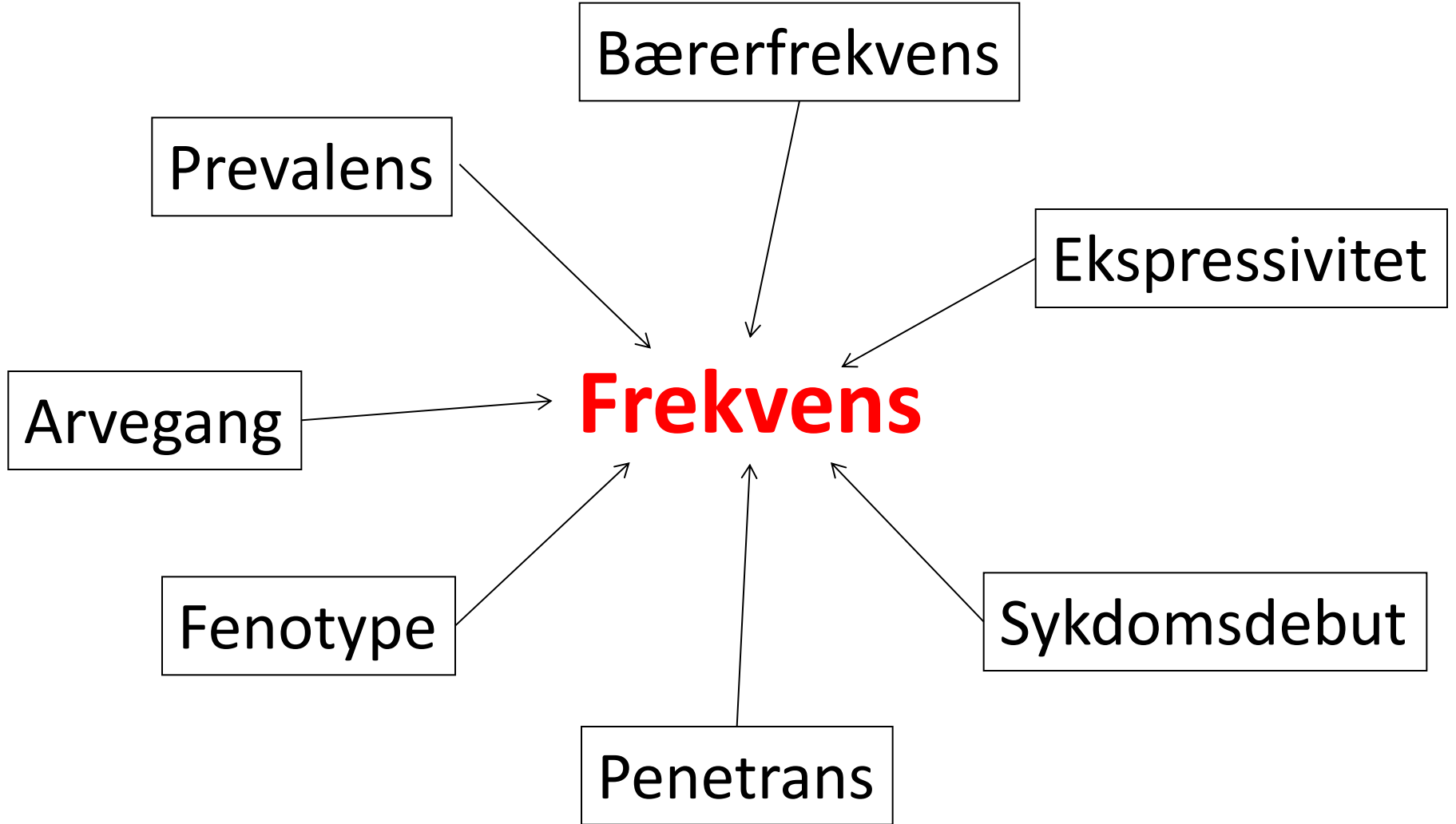
I 104 urelaterte DNA prøver ble det funnet en gjennomsnittlig bærerstatus på 2.8 sykdomsgivende varianter (0 til 7) (totalt 406).

27% av variantene som var angitt som sykdomsgivende i litteraturen, ble funnet å være normalvarianter eller å være feil annotert.

«Severe, orphan disease mutations should be uncommon (<1% incidence) and should not be found in the homozygous state in unaffected individuals.

Unexpectedly, we found that 74% of “disease mutation” calls were accounted for by substitutions with incidences of $\geq 5\%$, of which almost one-half were homozygous in samples unaffected by the corresponding disease.

Also unexpected was the finding that 14 of 113 literature annotated disease mutations were incorrect. Thus, for many recessive diseases, HGMD, dbSNP, OMIM, and the literature are insufficient arbiters of whether variants are disease mutations»



Den ideelle kontrollpopulasjonen

- Ikke samme fenotype som pasienten
- Samme etnisitet, geografisk tilhørighet som pasienten
- Ved X-bundet arvegang er kjønn viktig
- For sykdommer med sein debut bør kontrollgruppen være eldre enn forventet debut alder
- Tilstrekkelig antall

Følgende kilder kan benyttes til å finne frekvensen i den generelle befolkningen:

1000Genome

Phase 1: represents low coverage and exome data analysis available for the first 1092 samples. Available on (<http://browser.1000genomes.org>) (AlaMut)

Phase 3: 2535 individuals from 26 different populations around the world. Available in a ftp file for downloading.

NHLBI Exome Sequencing Project (ESP).

4300 EA og 2203 AA (<http://evs.gs.washington.edu>).(AlaMut)

Bør benyttes med forsiktighet for hjerterelaterte sykdommer og cystisk fibrose

CSAgilent

The ClinSeq Project Biesecker LG et al., 2009

Ca 660 individuals available at <http://www.ncbi.nlm.nih.gov/projects/SNP/>

Goal is 1500 participants

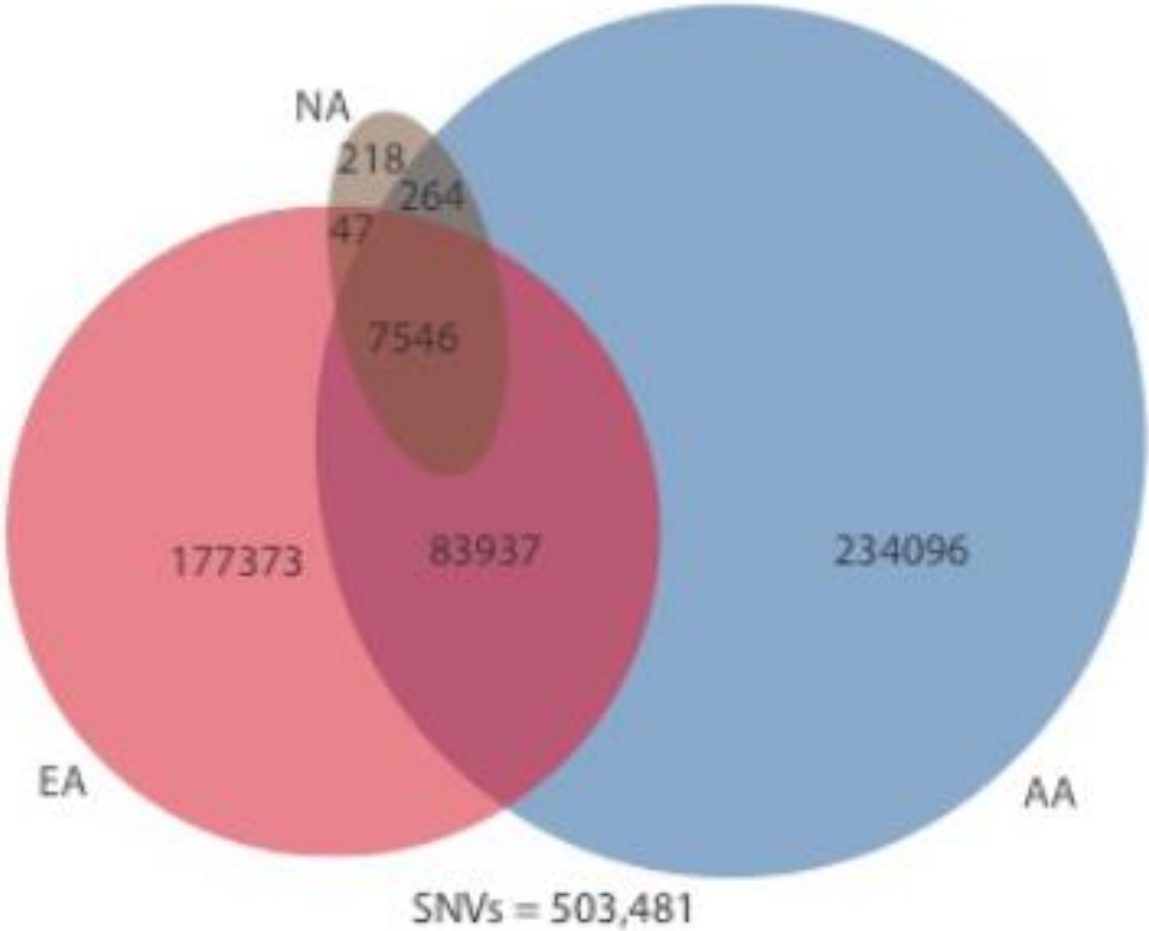
dbSNP

Med unntak av CSAgilent og ESP er populasjonene i dbSNP små og ikke alle representerer friske kontroller

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. (ESP)

Tennessen JA et al., 2012 Science 337; 64

A.



Detecting Polymorphisms and Mutations in Candidate Genes

Table 1

Sample Sizes Needed to Detect Polymorphisms

N^a	Polymorphism Frequency	Alpha	Power
40	.05	.05	.80
65	.05	.05	.95
210	.01	.05	.80
340	.01	.05	.95
2,400	.001	.05	.80
3,910	.001	.05	.95

^a N signifies the number of chromosomes, and this applies to either X-linked or autosomal diseases.

Bedre frekvensdata i fremtiden

- Exome Aggregation Consortium (ExAC), Cambridge, MA . Sequence of 63 000 unrelated individuals

"We have removed individuals affected by severe pediatric disease, so this data set should serve as a useful reference set of allele frequencies for severe disease studies."

- The 100 000 genome project UK (planlagt ferdig i 2017) 35 000 kontrollere
- HUNT
- **Nasjonale databaser !!**

COL5A2 c.2011C>T rs139189200

- 2 i esp-ea
- 5 i 1000genome

Exome Aggregation Consortium				
Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (Finnish)	96	6740	0	0.01424
Other	7	928	0	0.007543
European (Non-Finnish)	89	67572	0	0.001317
African	0	10570	0	0
East Asian	0	8724	0	0
Latino	0	11530	0	0
South Asian	0	16612	0	0
Total	192	122676	0	0.001565

Bedre frekvensdata i fremtiden

- Exome Aggregation Consortium (ExAC), Cambridge, MA . Sequence of 62 000 unrelated individuals

"We have removed individuals affected by severe pediatric disease, so this data set should serve as a useful reference set of allele frequencies for severe disease studies."

- The 100 000 genome project UK (planlagt ferdig i 2017) 35 000 kontrollere
- HUNT
- **Nasjonale databaser !!**

Sjeldne varianter er vanlige

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes.

Tennessen JA et al., 2012 Science 337; 64

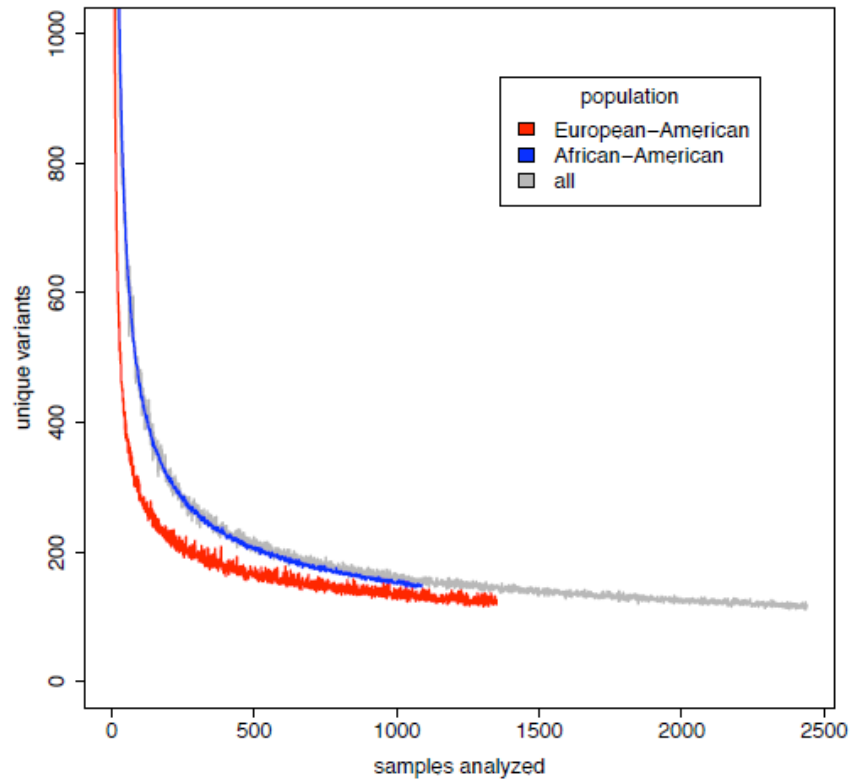


Figure S11. Number of unique variants identified as a function of sample size. Each point is the mean across 100 random samples of the desired number of individuals. Calculation of the number of unique variants was done for EA and AA populations individually and with both populations combined.

Nasjonal database for sekvensvarianter!



Huskeliste for variantvurdering

- Bruk av frekvensdata
- Synonyme varianter
- Missens varianter
- In-frame indels
- Nonsens varianter
- Out-of-frame indels
- Stop-loss
- Spleisesete varianter
- Intron varianter

Huskeliste for variantvurdering

- Bruk av frekvensdata
- Synonyme varianter
- Missens varianter
- In-frame indels
- Nonsens varianter
- Out-of-frame indels
- Stop-loss
- Spleisesete varianter
- Intron varianter

- ✓ frekvens i 1000genome, esp, CSAgent, In-house databaser
- ✓ predikert nytt spleisesete (prediksjonsprogram i AlaMut)
- ✓ rapportert i databaser/litteratur
- kan påvirke bindings seter for spleiseregulatoriske faktorer, men foreløpig mangler vi gode prediksjonsverktøy
- kan ha en rekke andre effekter på RNA stabilitet, lokalisasjon og translasjonshastighet

Huskeliste for variantvurdering

- Bruk av frekvensdata
- Synonyme varianter
- Missens varianter
- In-frame indels
- Nonsens varianter
- Out-of-frame indels
- Stop-loss
- Spleisesete varianter
- Intron varianter

Nonsens som sannsynlig aktiverer nonsense mediated decay (NMD) →. lokalisert oppstrøms for de siste 50 basene i nest siste ekson.

- ✓ er varianten rapportert tidligere hos pasienter med samme fenotype?
- ✓ Ved dominant arvegang: er haploinsuffisiens årsak til sykdom?
- ✓ er fenotypen forskjellig fra missense varianter?
- ✓ Ved recessiv arvegang: er andre nonsens varianter forbundet med sykdom?)

Nonsens som ikke aktiverer NMD. Lokalisert nedstrøms for de siste 50 basene i nest siste ekson. Det vil si at det sannsynligvis lages trunkert protein

- ✓ er varianten rapportert tidligere hos pasienter med samme fenotype?
- ✓ er det andre kjente sykdomsgivende nonsens varianter i samme område (som ikke aktiverer NMD)?
- ✓ er det en kjent proteinfunksjon knyttet til C-terminalen?
- ✓ er det sannsynlig at C-terminalen er viktig for proteinfolding?



Klasse 2 – sannsynlig normalvariant

Enighet om følgende kriterier:

- Frekvens i kontroller er betydelig, men ikke bra nok for klasse 1 (høyere enn forventet prevalens)
- Ko-segregerer ikke med sykdom (syk uten varianten)
- Påvist hos frisk forelder (dominant arvegang med $\approx 100\%$ penetrans)

Variierende bruk av følgende kriterier:

- Synonyme varianter og introniske varianter (utenfor konsensus spleisesete) som i følge prediksjonsprogrammer ikke introduserer ett nytt spleisesete.
- Grad av konservering på DNA nivå (phyloP og phastCons): Dersom den aktuelle nukleotiden ikke er godt konservert eller ikke befinner seg i en godt konservert region på DNA nivå settes varianten i klasse 2



Klasse 3 – usikker variant



Practice guidelines for the Interpretation and Reporting of Unclassified Variants (UVs) in Clinical Molecular Genetics.

Prepared and edited by Jennie Bell¹, Danielle Bodmer², Erik Sistermans³ and Simon C Ramsden⁴

2007

Svarrapportering av VUS'er; essensielt og *ikke* essensielt *Bell et al 2007*

- Det er essensielt at laboratoriet har databaser med oversikt over varianter og pasienter, og har mulighet til oppdatering av svarrapporter ved reklassifisering på bakgrunn av nye opplysninger.
- Det er essensielt at betegnelsene "polymorfi" og "mutasjon" ikke benyttes i en svarrapporten
Terminologi; Sekvensvariant/DNA-variant av ukjent klinisk betydning
- Det er essensielt å inkludere hvilke analyser som er utført
Om noen av analysene for screening ikke er utført (f.eks mangler exon 6,7,8), bør dette oppgis
- Det er *ikke* essensielt å dokumentere *hele* prediksjons- og klassifiseringsarbeidet som fører fram til klasse 3 varianten (VUS'en). Dette arbeidet bør lagres i laboratoriets databaser.
Ved reklassifisering bør imidlertid begrunnelsen dokumenteres grundig.

Svarrapportering; hvem trenger VUS-svar? *Bell et al 2007*

- Klasse 3 varianter bør i utgangspunktet gis ut til klinikere som er trent i faget, genetikere og genetisk veiledere.
- Varsomhet kreves ved utgivelse av "klasse 3-svar" til spesialiteter som ikke forstår kompleksiteten i informasjonen:
 - Vurder om varianten ikke skal gis ut
 - Forklar mer omstendelig betydningen (annen ordlyd)
 - Foreslå at pasienten får svaret under genetisk veiledning

Svarrapportering; hvilke VUS'er skal gis ut? *Bell et al 2007*

- Det er generelt ikke nødvendig å rapportere alle ikke-patogene varianter (klasse 2 og 3), dette kan i enkelte tilfeller føre til mistolkning av svaret. Det kan inkluderes i svarrapporten at ikke-patogene varianter ikke rapporteres.
- Det kan være viktig å rapportere VUS'er når den kliniske sammenhengen er usikker, eller når det er sterk klinikk og ingen patogene funn.
- Det er i utgangspunktet ikke anbefalt å utføre prediktiv testing basert på funn av en VUS. Det kan gjøres i spesielle tilfeller under god genetisk veiledning og ved segresjonsanalyse.

POLICY

Recommendations for reporting results of diagnostic genetic testing (biochemical, cytogenetic and molecular genetic)

Mireille Claustres¹, Viktor Kožich², Els Dequeker³, Brain Fowler⁴, Jayne Y Hehir-Kwa⁵, Konstantin Miller⁶, Cor Oosterwijk⁷, Borut Peterlin⁸, Conny van Ravenswaaij-Arts⁹, Uwe Zimmermann¹⁰, Orsetta Zuffardi¹¹, Ros J Hastings^{*,12} and David E Barton¹³, on behalf of [the ESHG Quality committee](#)

“It is important that findings of uncertain significance are included in reports, as their significance may become clear at a later date. Attempts should be made to explain significance of the finding using literature, additional databases, prediction programs and other means, and to report the likelihood of the finding in respect to the clinical question asked. “

Klasse 3, usikker variant (VUS)



- ✓ Påvist i friske kontroll grupper
- ✓ Endring til lignende aminosyre
- ✓ Samme aminosyreendring funnet i andre arter

- ✓ Ingen predikert effekt på spleisesete

- ✓ Ingen frekvensdata
- ✓ Vesentlig endring av aminosyre
- ✓ Høy konservering mellom arter
- ✓ Sannsynlig effekt på proteinstabilitet
- ✓ Lokalisert i funksjonelt viktig område

- ✓ Predikert effekt på spleisesete
- ✓ Predikert nytt spleisesete



Ved rapportering av usikre funn er det viktig å formidle hva **usikker betyr:**

«Vurdering av variantens betydning baseres på tidligere rapporter og molekylær biologisk kunnskap om varianten. I dette tilfelle mangler vi opplysninger som kan benyttes til å evaluere om varianten representerer en sjelden normalvariant eller om den kan være sykdomsgivende.»



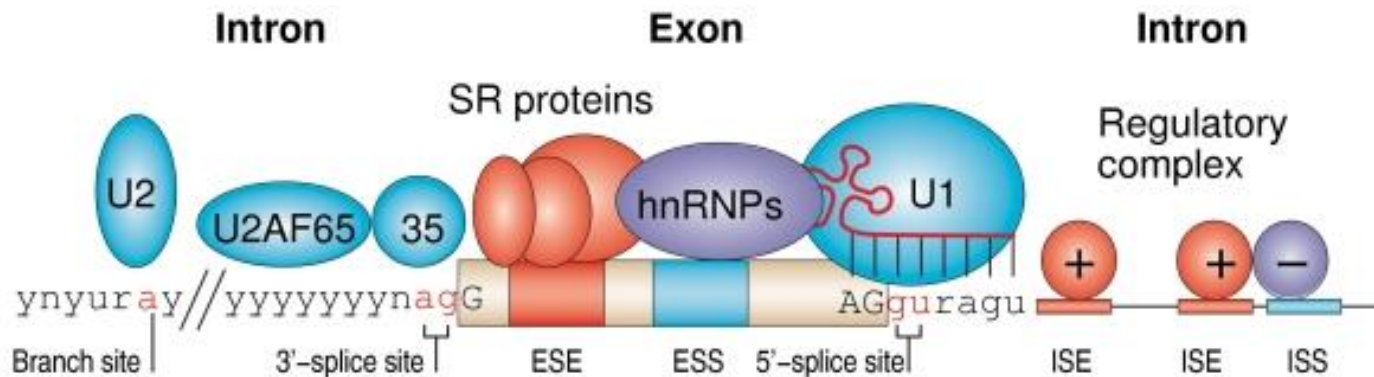
Videre utredning av VUSer

- ✓ RNA analyser
- ✓ Ko-segregering med sykdom i familien (dominant nedarving)
- ✓ Biokjemiske analyser
- ✓ Funksjonelle studier i celler (celler fra pasienten eller transfekterte celler)

RNA analyser

- På varianter med predikert effekt på spleising (reduisert styrke på normalt spleisesete eller dannelse av et nytt spleisesete)
- Uten funn av varianter, men ved sterk klinisk mistanke om at et gitt gen (2-3) er årsak til sykdom.
- Som diagnostisk verktøy på gener med mange små ekson og hyppig funn av spleisemutasjoner (NF1 , NF2– St.Olav)

Behov for RNA analyser vil øke når vi får bedre prediksjonsverktøy for å påvise mulig endring av bindingssete for spleiseregulatoriske faktorer



RNA analyser

Tilgang på vev hvor genet uttrykkes

Blodprøve i PAXGene rør /Tempur rør (RNA stabiliserende væske)

Heparinblod → korttidsdyrking av lymfocytter (mulighet for hemming av NMD)

Biopsi → dyrking av fibroblaster (mulighet for hemming av NMD)

Grad av feilspleising

Spesifikk amplifisering av normalt transkript og bruk av SNP for verifisering av biallelisk eller monoallelisk opprinnelse.

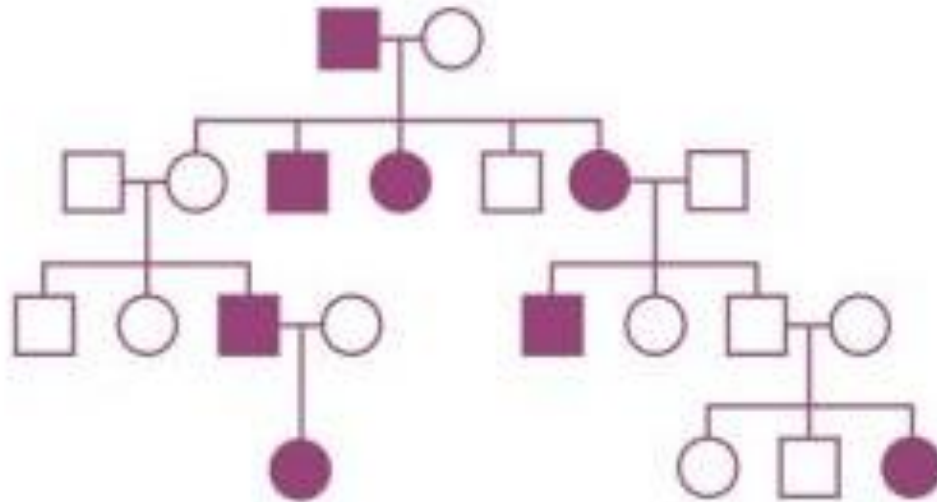
qRT-PCR – spesifikk amplifisering og kvantitering av ulike transkripter

- ✓ Alle labbene tilbyr RNA analyser ved behov (unntatt EEG-OUS)
- ✓ Behovet avgjøres i hvert enkelt tilfelle og da i samarbeid med kliniker.



Ko-segregerings analyser

Viktig at klinikerne kommuniserer resultatet tilbake til lab



Klasse 4 – sannsynlig patogen

- Publisert dokumentasjon, men som ikke holder til klasse 5
- Nonsens (NMD).
 - Dominant arvegang: Haploinsuffisiens er kjent årsak til sykdom.
 - Recessive arvegang: Nonsens er kjent årsak til sykdom
- Spleisesetevarianter -1,-2,+1,+2
- *De novo* i et gen med kjent forbindelse med fenotypen (med noe forbehold)
- mRNA analyser viser fullstendig feilspleising
- Endring av funksjonelt viktige aminosyrer i godt karakteriserte protein

Klasse 5 - patogen

- Publisert dokumentasjon (helst flere uavhengige artikler)
- Beskrevet funnet i **x** uavhengige pasienter med samme fenotype
- Eventuelle funn i normale kontroll grupper kan forklares utfra arvegang, penetrans etc.
- Funksjonelle studier (kritisk gjennomgang)
- Kosegregering med sykdom

$$P = 1 - (1/2)^n \quad (n = \text{affekteerte individer} - 1)$$

(Møller P, et al., 2011)

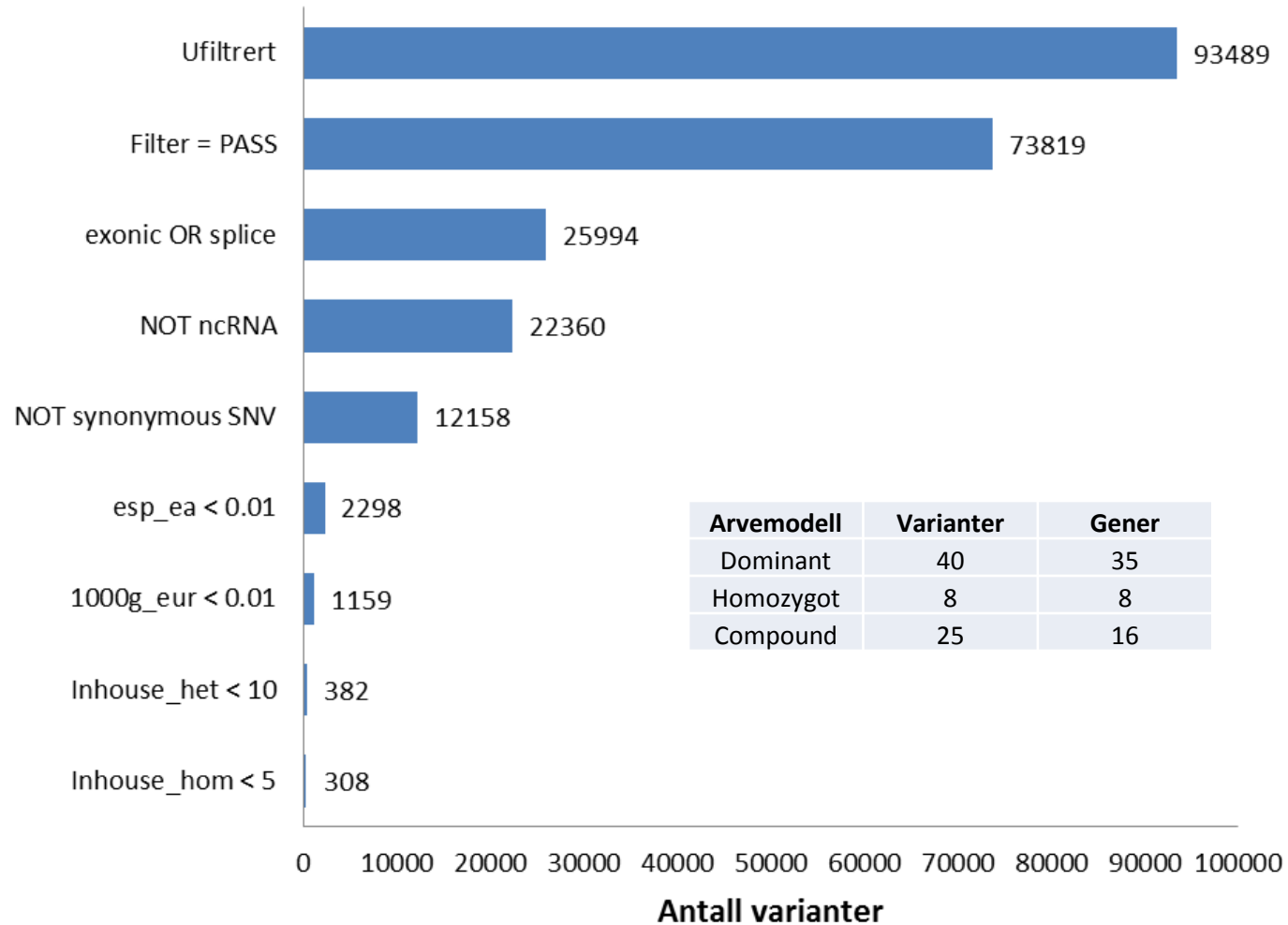
antall meisoser	sannsynlig sammenheng med sykdom (%)
2	50,00
3	75,00
4	87,50
5	93,75
6	96,88
7	98,44
8	99,22

Kosegregering er ikke et bevis på at varianten er årsak til sykdom. Varianten kan være lokalisert i cis til den sykdomsgivende varianten hos den analyserte familien.

Variantvurdering HTS

Kristian Tveten

Filtrering av varianter



Arvemodell	Varianter	Gener
Dominant	40	35
Homozygot	8	8
Compound	25	16

Filtrering av varianter

- De novo / homozygot
- Kjent mutasjon i HGMD
- Gen med kjent sykdomsassosiasjon (HGMD, OMIM ++)
- Ny filtrering
 - synonyme, intron/UTR, frekvens, ikke-PASS

Case

- Pasient tidligere henvist for CMT uten funn
- MR cerebrum forenlig med leukoencefalopati
- Første filtrering gir ingen funn
- Fjerner exonic OR splice

Arvemodell	Varianter	Gener
Dominant	306 (40)	268 (35)
Homozygot	49 (8)	48 (8)
Compound	222 (25)	122 (16)

- To spleisemutasjoner i DARS2 i *trans*
 - c.228-21_228-20delTTinsC (Partiell exon 3 skipping - NMD)
 - Frekvens i Finske kontroller 6 / 560 kromosomer
 - c.492+2T>C (Ekson 5 skipping - NMD)
 - ESP frekvens 0.03%

HTS-Skien

- Rapporterer ikke ut klasse 3 i første runde:

«Det ble ikke gjort sikre patogene funn som kan forklare pasientens tilstand. Det ble imidlertid funnet genetisk variasjon av usikker betydning.»

Disse kan gis ut på forespørsel til medisinsk genetikere.



Hva vi ikke har sagt noe om?

Varianter kan endre

- ekspresjonen av genet (binding av transkripsjonsfaktorer)
- mRNA stabilitet
- translasjonshastighet (kodon-bruk)
- mm

Bør fremtidens diagnostikk-labber ha større fokus på funksjonelle assay?

- **RNA**

RT-PCR, qRT-PCR i celler fra pasienten

Kunstig system med ekson-trapping

- **Protein**

Funksjonelle studier i celler fra pasienten

Funksjonelle studier i transfekterte celler

www.nshg.no



Norsk selskap for Humangenetikk

Veileder for rapportering av MLPA resultater (versjon 1; 11.02.2014)

Utarbeidet av: Bjørn Ivar Haukanes (HUS)
 Wenche Sjursen (St Olavs Hosp)
 Monica Ingebrigtsen og Marijke van Ghelue (UNN)
 Mari Ann Kulseth (OUS)