

Open

A comprehensive strategy for exome-based preconception carrier screening

Suzanne C.E.H. Sallevelt, MD¹, Bart de Koning, PhD^{1,2}, Radek Szklarczyk, PhD²,
Aimee D.C. Paulussen, PhD¹, Christine E.M. de Die-Smulders, MD, PhD^{1,3}
and Hubert J.M. Smeets, PhD^{1,2,3,4}

Purpose: Whole-exome sequencing (WES) provides the possibility of genome-wide preconception carrier screening (PCS). Here, we propose a filter strategy to rapidly identify the majority of relevant pathogenic mutations.

Methods: Our strategy was developed using WES data from eight consanguineous and five fictive nonconsanguineous couples and was subsequently applied to 20 other fictive nonconsanguineous couples. Presumably pathogenic variants based on frequency and database annotations or generic characteristics and mutation type were selected in genes shared by the couple and in the female's X-chromosome. Unclassified variants were not included.

Results: This yielded an average of 29 (19–51) variants in genes shared by the consanguineous couples and 15 (6–30) shared by the nonconsanguineous couples. For X-linked variants, the numbers per female

were 3 (1–5) and 1 (0–3), respectively. Remaining variants were verified manually. The majority were able to be quickly discarded, effectively leaving true pathogenic variants.

Conclusion: We conclude that WES is applicable for PCS, both for consanguineous and nonconsanguineous couples, with the remaining number of variants being manageable in a clinical setting. The addition of gene panels for filtering was not favorable because it resulted in missing pathogenic variants. It is important to develop and continuously curate databases with pathogenic mutations to further increase the sensitivity of WES-based PCS.

Genet Med advance online publication 27 October 2016

Key Words: preconception carrier screening; whole-exome sequencing

INTRODUCTION

Preconception carrier screening (PCS) refers to genetic testing of couples who do not have an a priori increased risk of having a child with a recessive genetic disease before they attempt to conceive.

More than 1,300 (autosomal and X-linked) recessive disorders have been identified so far,¹ and these vary greatly in severity and age of onset. Although individually uncommon in general populations, Mendelian diseases are collectively reported to account for ~20% of infant mortality and ~10% of pediatric hospitalizations.² Detection of carrier status enables identification of couples with a 25% risk of affected offspring. The primary aim of PCS is to provide such couples with informed reproductive choices, including prenatal diagnosis, preimplantation genetic diagnosis, accepting the genetic risk and preparing themselves for (the possibility of) having a child with a certain disease, sperm/egg donation, adoption, or refraining from having children. In certain communities, particularly those with a high incidence of specific severe diseases, disease prevention may be viewed as the primary goal.³ Another potentially favorable consequence of carrier screening

is enabling early perinatal diagnosis and treatment that can profoundly reduce morbidity and mortality. The preconception period is considered the optimal timing for carrier screening because only then are all aforementioned reproductive options still applicable.

The prevalence of offspring with major anomalies is higher among consanguineous couples than among nonconsanguineous couples,⁴ mostly because of autosomal recessive disorders. Increased genetic risks are also applicable to subpopulations with high carrier frequencies of specific mutations.⁵ The concept of PCS has been applied for decades in some of these populations.⁶ Typically, one or a few diseases with a relatively high incidence in a population, or subpopulation, are tested in a targeted manner.⁵ However, with the introduction of next-generation sequencing techniques, simultaneous testing of much larger gene numbers has become possible and cost-effective. Until now, particularly targeted, gene panel-based, next-generation sequencing approaches have been described.^{7–9} An approach using targeted analysis after untargeted whole-exome sequencing (WES) has been explored in one study that filtered for identical mutations in approximately 500 genes in both

¹Department of Clinical Genetics, Maastricht University Medical Center+ (MUMC+), Maastricht, The Netherlands; ²Department of Genetics and Cell Biology, Maastricht University, Maastricht, The Netherlands; ³Research School for Developmental Biology (GROW), Maastricht University, Maastricht, The Netherlands; ⁴Research School for Cardiovascular Diseases in Maastricht, CARIM, Maastricht University, Maastricht, The Netherlands. Correspondence: Suzanne C.E.H. Sallevelt (suzanne.sallevelt@mumc.nl)

Submitted 16 June 2016; accepted 15 August 2016; advance online publication 27 October 2016. doi:10.1038/gim.2016.153

partners of four consanguineous couples.¹⁰ However, systematic assessment of the use of WES for PCS has not been performed so far. In our study, we aimed to investigate WES for PCS in a broader setting and to develop a filter strategy that could be used for consanguineous and nonconsanguineous couples.

MATERIALS AND METHODS

Different filter strategies for PCS were explored in WES data from eight consanguineous couples and five fictive nonconsanguineous couples. The resulting proposed filter strategy was applied to another 20 nonconsanguineous fictive couples.

Subjects

The eight consanguineous couples all have a child with suspected mitochondrial or mitochondrial-like disease. They gave consent for WES, including for their own DNA, in a research setting. The genetic cause of the child's disease was established in four couples (1, 5, 6, and 7) (**Table 1**); the affected child was homozygous for the causative mutation(s) and the parents were identified as heterozygous carriers. Couples 1, 3, 5, 6, and 7 were of Moroccan origin, couples 4 and 8 were from Turkey, and couple 2 was from Iraq (**Supplementary Table S1** online). For more details on the degrees of consanguinity of the eight couples, see **Supplementary Table S1** online.

The nonconsanguineous couples, originating from the Dutch population, were anonymized male and female exomes from existing diagnostic data sets that were randomly mixed. For the current proof-of-concept study, it was not necessary to use actual couples. The study was approved by the local Medical Ethical Committee.

Analysis

WES for the consanguineous couples and the nonconsanguineous couples was performed as previously described.^{11,12} Standard filter steps were location (exonic and canonical splice, the latter defined as ≤ 2 nucleotides intronic), sequencing depth/reads (≥ 8), and zygosity (homozygous variants removed). Variably applied filter steps were amino acid alteration (synonymous variants removed), annotation in databases (HGMD and/or ClinVar), allele frequency (dbSNP $< 5\%$ versus $< 1\%$), mutation type (frameshift, nonsense, splice site, missense), presence of identical variants in both partners of the couple, presence of variants in the same genes in both partners of the couple, and presence of variants in genes included in gene panels. To establish an adequate cutoff for the allele frequency, disease prevalence was reviewed. Allele frequencies were calculated by taking the square root of the prevalence. The ClinVar database on pathogenicity of known variants was downloaded from <ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar> on 2 March 2016 in variant call format. These variants were matched to the patient exome data using SNP identifiers, accounting for multiple SNPs at a given position. For multiple pathogenicity classes assigned to a genetic variant, only classes 4 ("likely pathogenic") and 5 ("pathogenic") were taken into consideration for further analysis.

Gene panel

The gene list described by Bell *et al.* and Kingsmore,^{2,7} which consists of 508 items comprising 437 genes that supposedly cause 448 severe recessive diseases, was studied. A curated list of 459 genes after the removal of some (particularly nonrecessive) genes/diseases was created, with the addition of the genes/diseases for which preimplantation genetic diagnosis was performed in our center as well as those included in a PCS pilot in Groningen, the Netherlands;¹³ this list was used in our analysis.

RESULTS

A first observation was the larger number of variants per individual in the consanguineous individuals compared to the nonconsanguineous. Considering the ethnicity of the consanguineous couples (**Supplementary Table S1** online), this probably reflects population differences related to human evolutionary origin and migration.¹⁴

Proposed filter strategy

Variant filtering in all genes interrogated by WES. As a first and separate step, variants annotated as pathogenic or likely pathogenic in ClinVar and with a $< 1\%$ or unknown dbSNP frequency were selected from the exome data, yielding an average of eight (5–13) variants per individual ($n = 16$) for the consanguineous couples and five ((0–11), $n = 50$) for the nonconsanguineous (**Figures 1** and **2** and **Tables 1** and **2**, **Supplementary Tables S2–S4** online). From the remainder, HGMD-annotated variants with SNP frequency $< 1\%$ /unknown were selected, resulting in 20 (11–33) and 11 (1–21) variants, respectively. The residual data were initially filtered for location, sequencing depth, zygosity, amino acid alteration, and SNP frequency. Subsequently, frameshift, nonsense, and splice site mutations were selected and added to the ClinVar-selected and HGMD-selected variants. This total of presumably pathogenic mutations yielded an average of 100 (84–151) variants for the consanguineous individuals and 80 (59–98) variants for the 25 nonconsanguineous. These were compared between both partners of the couple; the variants in genes where the partner has at least one presumably pathogenic variant as well, and vice versa, were selected, resulting in an average of 29 (19–51) variants for the consanguineous couples and 15 (6–30) for the nonconsanguineous. X-linked variants were separately selected for the females, yielding three (1–5) variants and one (0–3) variant, respectively. All six previously identified disease-causing variants were present in the expected variant pools in our data set of the consanguineous carriers (**Table 1**).

For the first two consanguineous and nonconsanguineous couples, the presumably pathogenic variants in shared genes are displayed in **Supplementary Table S3** online as examples. The majority were able to be discarded based on high in-house and/or ExAC (<http://exac.broadinstitute.org/>) frequency in an absolute manner or relative to disease prevalence and/or based on the absence of a monogenic disease association and/or because they appeared to be a sequencing artifact (**Supplementary Table S3** online). The remaining

Table 1 Previously established disease-causing variants in consanguineous couples 1, 5, 6, and 7 and their identification by our filter strategy

Couple	Gene	Known human disease gene?	Variant	Previously described?	Pathogenicity	Present in expected variant pool in our data?	Selected in our data by:	Gene present in panel of 459 recessive genes?
C1	<i>ADD3</i>	Yes	NM_016824.3: c.1100G>A, p.(Gly367Asp)	Yes ³³	Pathogenic	Yes (pool: total presumably pathogenic variants, in genes shared by the couple)	Presence in HGMD (not present in ClinVar)	No
C5	<i>AP4M1</i>	Yes	NM_004722.2: c.1137+1G>T	Yes, Verkerk et al. 2009 (PMID 19559397) and Langouet et al. 2015 (PMID 26029708)	Pathogenic	Yes (pool: total presumably pathogenic variants, in genes shared by the couple)	Presence in HGMD (not present in ClinVar)	No
C6	<i>ANTXR2</i>	Yes	NM_001145794.1: c.1142A>G, p.(Tyr381Cys)	Yes ³⁴	Pathogenic	Yes (pool: total presumably pathogenic variants, in genes shared by the couple)	Pathogenic annotation in ClinVar (also present in HGMD)	Yes
C6	<i>ACY1</i>	Yes	NM_001198895.1: c.811G>A, p.(Ala271Thr)	No	Possibly pathogenic	Yes (pool: missense variants not annotated (likely) pathogenic, in genes shared by the couple)	Generic filter steps for variants not annotated (likely) pathogenic	No
C6	<i>SERAC1</i>	Yes	NM_032861.3: c.1347_1349dupATC, p.(Ser450dup)	No	Pathogenic	Yes (pool: in-frame rearrangements not annotated (likely) pathogenic, in genes shared by the couple)	Generic filter steps for variants not annotated (likely) pathogenic	No
C7	<i>TMEM126A</i>	Yes	NM_032273.3: c.163C>T, p.(Arg55*)	Yes ³⁵	Pathogenic	Yes (pool: total presumably pathogenic variants, in genes shared by the couple)	Pathogenic annotation in ClinVar (also present in HGMD)	No

The colors correspond to those in Figure 2.

relevant variants are summarized in **Table 2**. If necessary, low coverage variants may be confirmed by Sanger sequencing. Consanguineous couple 1 (C1) was identified as a clear carrier couple only for the pathogenic mutation in the *ADD3* gene, which was causative of their child's cerebral palsy. For the *ABCA4* gene and the *PRPH2* gene, both implicated in ophthalmologic phenotypes, only in one of the partners did a pathogenic mutation remain (**Supplementary Table S3** online); therefore, the finding is not relevant in the context of PCS. Consanguineous couple 2 (C2) was identified as a carrier couple of a known pathogenic *ABCA4* mutation and of a *FGG* mutation that has been reported once in association with delayed clot formation/dysfibrinogenemia in heterozygous state.¹⁵ It is unknown whether homozygosity for this particular mutation would result in a more severe phenotype (congenital afibrinogenemia, OMIM 616004). Furthermore, in both partners a rare *PRRX1* missense substitution was detected (**Supplementary Table S3** online). The latter has been reported in one association study of atrial fibrillation,¹⁶ whereas mutations in this gene are otherwise implicated in the agnathia-otocephaly complex (OMIM 202650), mostly in

the heterozygous state, although one homozygous mutation has been reported. In nonconsanguineous couple 1 (NC1), no relevant presumably pathogenic variants were found. Nonconsanguineous couple 2 (NC2) was identified as a *CFTR* (cystic fibrosis, OMIM 219700) carrier couple, with the female carrying the classic DF508 mutation and the male carrying the mild/nonclassical *CFTR* mutation L997F.^{17,18}

In X-linked genes, a pathogenic *G6PD* mutation (*G6PD* deficiency, OMIM 300908) was identified in consanguineous female 2 and a canonical splice site substitution in the *PHF8* gene (siderius X-linked mental retardation syndrome, OMIM 300263) was found in nonconsanguineous female 2 (**Table 2** and **Supplementary Table S4** online).

To compare our results to those after adding an extensive gene panel, the same four couples were filtered for 459 autosomal and X-linked recessive genes. This failed to identify the pathogenic *ADD3* variant in C1, the pathogenic *ABCA4* variant in consanguineous couple 2, and the *PHF8* splice site substitution in NC2. The identification of mutations that pose some uncertainties, for example, regarding predicting homozygous phenotype, was not prevented (e.g., *FGG*). Only one (*ANTXR2*)

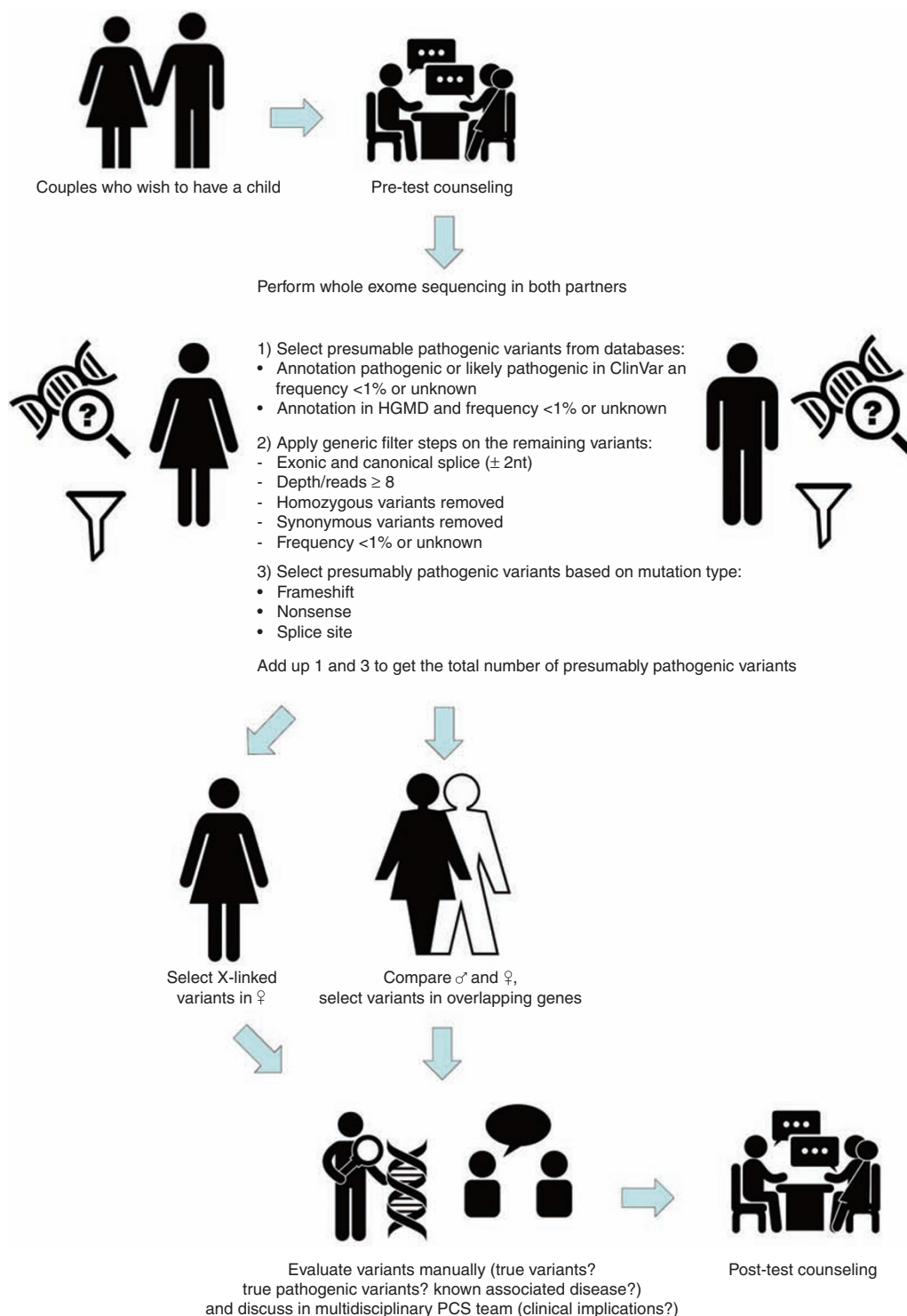


Figure 1 Strategy for whole-exome sequencing-based preconception carrier screening.

of the six genes that were disease-causing in the consanguineous families was included in the gene panel.

Remaining variants. Beyond the annotated variants and presumed pathogenic variants, two variant categories remained (**Figure 2**, **Supplementary Figure S1** and **Supplementary Table S2** online).

The first consists of missense variants that are not present in HGMD or annotated as at least likely pathogenic in ClinVar, with an average of 1,190 (961–2,269; $n = 16$) variants per individual in the consanguineous couples and 650 (513–924; $n = 50$) in the nonconsanguineous. Second, nonannotated in-frame deletions, insertions, and complex rearrangements remained; an

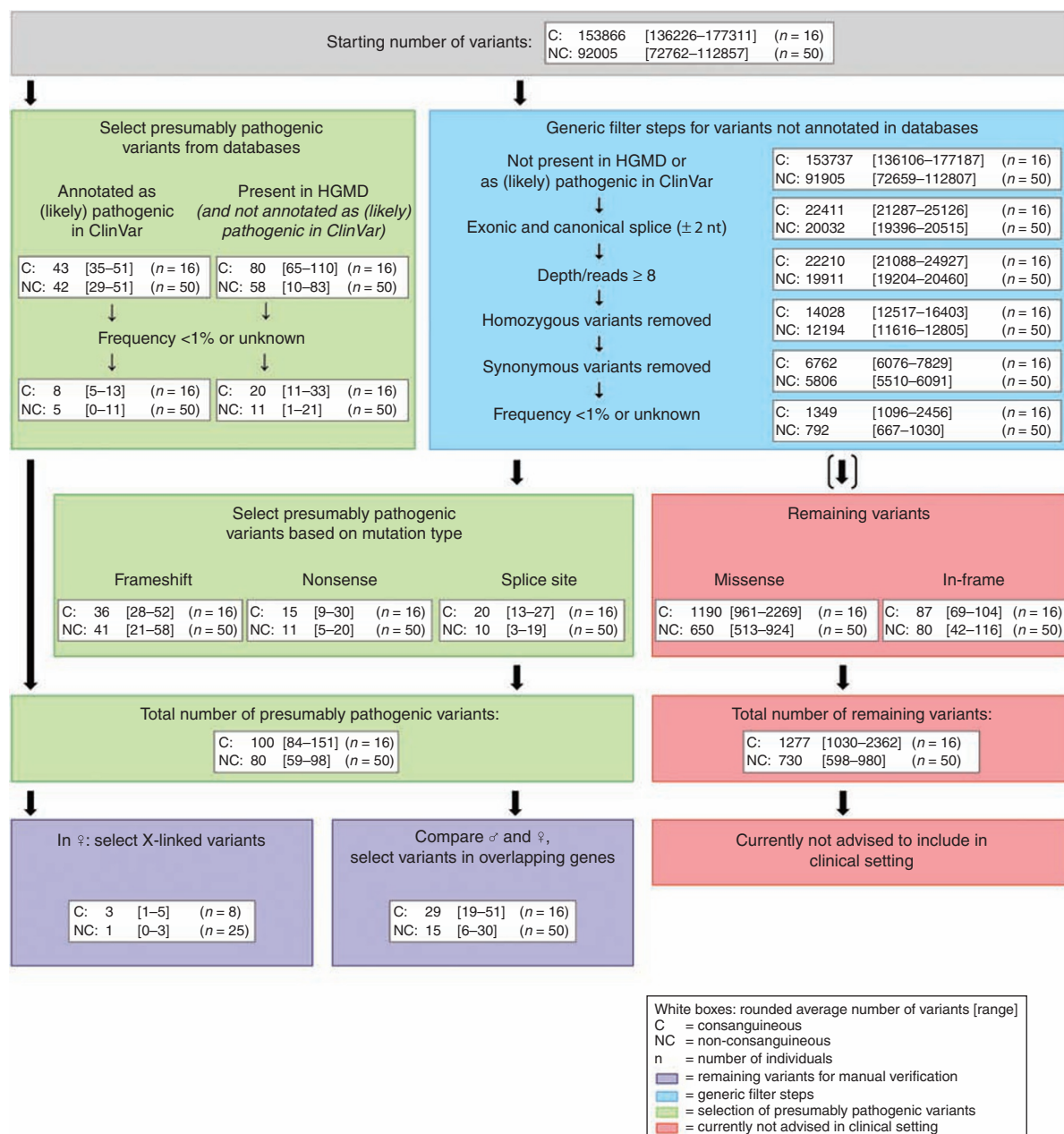


Figure 2 Average variant numbers resulting from our filter strategy for whole-exome sequencing-based preconception carrier screening.

average of 87 (69–104) in the consanguineous and 80 (42–116) in the nonconsanguineous individuals. These added up to a total of 1,277 (1,030–2,362) and 730 (598–980) remaining variants, respectively. The selection of variants present in the same genes in both partners reduced the numbers to 534 (385–1,150) and 178 (91–329). A further reduction was achieved by applying the gene list: 15 (8–35) and 5 (0–10) variants, respectively.

We used the data from the gene panel to assess the number of carrier couples in the entire cohort (**Supplementary Figure S1** online, **Supplementary Tables S2 and S5** online). Extrapolating this to the expected number of carrier couples gives an idea of the proportion of pathogenic variants that is missed by our strategy (see Discussion). A total of 5 unique

variants in X-linked genes and 21 in the overlapping genes were found in the 33 couples (**Supplementary Table S5** online). Again, not all variants remain as (likely) pathogenic. In C2, the aforementioned *FGG* variant was found. Additionally, as anticipated, consanguineous couple 6 (C6) was identified as a carrier couple of the pathogenic *ANTXR2* mutation (**Table 1**). In the nonconsanguineous couples, no couple other than the previously mentioned *CFTR* carrier couple was identified (NC2). In nonconsanguineous 3 (NC3), only one partner carried a pathogenic *PKHD1* mutation (AR polycystic kidney disease, OMIM 263200), whereas the *PKHD1* variant in the other partner could be discarded. In the X-linked analysis, only one relevant variant other than the *G6PD* (C2) variant was identified: a missense

Table 2 Pathogenic variants detected by our PCS strategy, filtering all genes interrogated by whole-exome sequencing, in two consanguineous and two nonconsanguineous couples

Couple (variant present in ♀ and/or ♂)	Gene	Associated disease in OMIM	Present in gene panel?	Variant	Variant type	Selected in our data by:	Relevant information on variant	(Potentially) Relevant in PCS context for current couple?
Pathogenic variants in autosomal genes shared by the couple								
C1 (♀+ ♂)	<i>ADD3</i>	Spastic quadriplegic cerebral palsy (OMIM 617008)	No	NM_016824.3: c.1100G>A, p.(Gly367Asp)	Missense substitution	Presence in HGMD (not present in ClinVar)	Disease-causing annotation in HGMD. 1 citation: reported to cause cerebral palsy in four siblings, functional evidence included ³³	Yes
C2 (♀+ ♂)	<i>ABCA4</i>	Stargardt disease 1 (OMIM 248200), retinitis pigmentosa 19 (OMIM 601718), cone-rod dystrophy 3 (OMIM 604116). Of note: heterozygous mutations are associated with susceptibility to age-related macular degeneration 2 (OMIM 153800)	Yes	NM_000350.2: c.5882G>A, p.(Gly1961Glu)	Missense substitution	Annotation pathogenic in ClinVar (also present in HGMD)	Multiple pathogenic annotations in ClinVar. Disease-causing annotation in HGMD (1 citation).	Yes
C2 (♀+ ♂)	<i>FGG</i>	Congenital afibrinogenemia (OMIM 202400). Of note: the milder or even clinically silent congenital dysfibrinogenemia (OMIM 616004) is mostly caused by heterozygous mutations	Yes	NM_021870.2: c.124G>A, p.(Gly42Ser)	Missense substitution	Presence in HGMD (not present in ClinVar)	Disease-causing annotation in HGMD. 1 citation: associated with delayed clot formation/ dysfibrinogenemia in heterozygous state in a single individual ¹⁵	Yes?
NC2 (♀)	<i>CFTR</i>	Cystic fibrosis (OMIM 219700)	Yes	NM_000492: c.1521_1523del, p.(Phe508del)	In-frame deletion	Annotation pathogenic in ClinVar (also present in HGMD)	Well-known pathogenic mutation	Yes
NC2 (♂)	<i>CFTR</i>	Cystic fibrosis (OMIM 219700)	Yes	NM_000492.3: c.2991G>C, p.(Leu997Phe)	Missense substitution	Annotation pathogenic in ClinVar (also present in HGMD)	Two pathogenic annotations in ClinVar, 2 risk factor annotations in ClinVar. Likely disease-causing annotation in HGMD (10 citations)	Yes
Pathogenic variants in X-linked genes in female								
C2 (♀)	<i>G6PD</i>	G6PD deficiency (OMIM 300908)	Yes	NM_000402.3: c.653C>T, p.(Ser218Phe)	Missense substitution	Annotation pathogenic/ likely pathogenic in ClinVar (also present in HGMD)	Three pathogenic annotations in ClinVar, 1 likely pathogenic annotation in ClinVar. Disease-causing annotation in HGMD (6 citations)	Yes
NC2	<i>PHF8</i>	Siderius X-linked mental retardation syndrome (OMIM 300263)	No	NM_001184896.1: c.16+1G>A	Canonical splice site substitution	Mutation type	No rs number, dbSNP frequency unknown, not present in in-house* database or ExAC database. Note: low coverage variant (15 reads)	Yes

C, consanguineous; NC, non-consanguineous; PCS, preconception carrier screening. * > 5,000 exomes, Department of Human Genetics, Radboud UMC Nijmegen.

substitution in the *AR* gene (NC21), which has been reported in two boys with severe virilization defects due to partial androgen insensitivity^{19,20} and also in a boy with normal phenotype.²¹ Whether the mutation has a phenotypic consequence appears to depend on the size of polymorphic repeats.²⁰

Altogether, four of the eight consanguineous couples and one of the 25 fictive nonconsanguineous couples were identified as autosomal recessive carrier couples (Tables 1 and 2; **Supplementary Table S5** online). X-linked carriership was established in one consanguineous and two nonconsanguineous females (Table 2 and **Supplementary Table S5** online).

DISCUSSION

We explored WES as a method for PCS and provided a filter strategy to rapidly identify the majority of pathogenic mutations in genes shared by the couple and in the female's X-linked genes. A major advantage of our PCS approach is that WES, used increasingly as a standard cost-effective diagnostic technique with a short turn-around time of a few months, could easily be implemented in the same flow. Whether WES-PCS will be cost-effective will require a proper HTA (health technology assessment) study, but it is clear that in genetic testing the prevention of disease occurrence is (cost-wise) most rewarding and it is likely that, given the high number of unexpected recessive disease cases occurring every year, this is or will become cost-effective soon. Furthermore, using WES enables flexible adjustment of the genes included in the PCS test and of the exact filter strategies used. The most important benefit of PCS based on WES from a clinical point of view is the possibility to screen all known disease genes instead of a selected and (by definition) limited/arbitrary gene panel. Very rare recessive disease alleles, mainly relevant for consanguineous couples, are not likely to be included in such panels.

One of 25 nonconsanguineous couples was identified as a carrier couple. The expected number of nonconsanguineous carrier couples can be derived from the cumulative risk of being a carrier couple of any autosomal recessive disease in the population concerned. The latter was calculated/estimated to be ~2% (based on prevalence numbers of rare diseases provided by the study by Orphanet,²² in which only diseases that can technically be identified by WES were considered and prevalence applicable to our northern European population was applied), including disorders ranging from severe to mild. Although our numbers are small, finding 1 nonconsanguineous carrier couple out of 25 may be grossly in line with this. The one couple identified as being a *CFTR* carrier is also in agreement with the relatively common *CFTR* carriership in northern Europe.²³ For females, a cumulative X-linked carrier risk of ~1% was estimated,²² again exclusively based on diseases that can be picked up by WES and including conditions with limited clinical relevance, whereby a de novo rate of one-third for lethal/nonreproducing X-linked disorders was taken into account.²⁴ Our findings of 3 (presumed) carriers out of 33 exceed this estimation, indicating, although in a limited data set, that probably not many variants are missed. For X-linked carriership, no differences were

expected between consanguineous and nonconsanguineous couples; this was illustrated by our data.

Selection of pathogenic variants. We selected presumably pathogenic variants based on published data (databases) and on predictions by the nature of the variant. Although databases are known to be imperfect, we initially hypothesized that by selecting variants associated with human disease, annotated in the Human Genome Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/ac/index.php>),²⁵ additional filter steps such as frequency would not be necessary. However, the remaining numbers were much higher than expected and included high-frequency benign variants or variants associated with multifactorial disease. The same was true for variants that are annotated as (likely) pathogenic in ClinVar. Therefore, we added frequency as an additional step, which largely reduced the numbers without removing presumably pathogenic variants.

In addition, we selected nonannotated (likely) pathogenic variants in our strategy. Synonymous variants as well as homozygous variants in healthy adults were discarded from the variant pool. Other criteria for determining a variant's pathogenicity are location and frequency. An allele frequency cutoff of <1% (corresponding with a disease prevalence of <1/10,000 for autosomal recessive diseases) was considered justifiable based on the aforementioned Orphanet disease frequency data.²² Most diseases with a prevalence above this cutoff have limited clinical relevance (e.g., congenital isolated thyroxine-binding globulin deficiency), are caused by mutation types that are not detected by sequencing analysis and therefore require a separate test anyway (e.g., SMA, fragile X syndrome), or are more frequent only in specific populations (e.g., thalassemia). The latter is generally well known for such populations, thereby enabling the addition of population-specific tests. Moreover, an allele frequency of <1% in the filter algorithm represents the frequency of a single disease allele, whereas for most diseases multiple disease alleles exist. Exceptions may be disease alleles that are responsible for the majority of disease cases (e.g., p.Phe508del mutation in cystic fibrosis or p.Lys304Glu mutation in MCAD deficiency) or (again population-specific) certain founder mutations. These can be assessed in or added to the data separately.

A frequency cutoff of <5% was also considered but resulted in a 50% increase in remaining variants (data not shown), which is expected to be particularly attributable to nonpathogenic variants. This will significantly increase the interpretation load at the end and, in our opinion, does not outweigh the effort of separate mutation analysis/testing that may be necessary in specific populations. After filtering for generic filter steps, we selected frameshift mutations, nonsense mutations, and canonical splice site mutations, which are often considered pathogenic due to predicted effects on the protein. Obviously, as for the annotated variants, all resulting variants will have to be evaluated because, for various reasons (annotation errors, database errors), not every one of these is actually pathogenic. Also, if coverage is borderline, then validation by Sanger sequencing is advisable. The number of variants resulting from our filter strategy for PCS

was small enough to enable such follow-up in a diagnostic setting. With our strategy, only between 0 and ~20 variants needed manual verification, varying from fast checks (e.g., HGMD/ClinVar, OMIM, Orphanet prevalence list) for most to more extensive evaluation for a few. Of note, in-house variant frequency databases as well as the ExAC database appeared very useful for relatively quickly discarding a first set of variants, even though only variants with low or unknown dbSNP frequency were selected in our algorithm. Benign variants or variants of uncertain significance (VUS) may be present in the final results, with the latter posing the greatest clinical challenge. However, the majority of variants, even when analyzing the entire exome, could be clearly assigned as being either benign or pathogenic. A complicating factor in PCS compared to genetic testing in affected individuals is the lack of a phenotype to which to relate the genetic findings. This is not a WES-specific issue; it also applies to, for example, targeted sequencing of specific genes. The only way to circumvent this issue is to exclusively test for known pathogenic mutations. However, collecting these mutations is very time-consuming and, more importantly, once a certain mutation-based array is designed, mutations are not easily added. Given the daily basis on which new pathogenic mutations are identified, such a method would be outdated in no time.

Pathogenic mutations that may be missed by WES-based PCS. Pathogenic mutations may be missed either because they are not recognized as being pathogenic in the filtering process or because they are not detected by the sequencing technique. The first category consists mainly of amino acid substitutions (missense variants) and, to a lesser extent, in-frame rearrangements not yet known/described to be pathogenic. These variants are often difficult to interpret regarding pathogenicity. As illustrated by our data, their amount in genes shared by the couple can be reduced to manageable numbers by applying gene panels. However, reducing the variant numbers will not solve the interpretation issues. Therefore, we feel this category is not recommended in a diagnostic setting at this point. It is necessary to obtain more experience (in a research setting) with the nature of the resulting variants to further evaluate their use in clinical practice. More importantly, increasing knowledge of pathogenic mutations and collection of these mutations in an aforementioned database will eventually decrease this variant category. Also, newer filtering pipelines that are able to identify other presumably pathogenic subcategories (e.g., stop-loss, start-loss, and/or glycosylation site missense variants) will move variants from the unknown to the presumably pathogenic variant pool. In time, the same will be true for intronic variants, microRNA variants, and others.

Because conservation is an important factor in pathogenicity, one may argue for its inclusion in the filter strategy. However, conservation scores predicting (potential) variant pathogenicity either isolated (e.g., GERP²⁶, PhyloP²⁷) or incorporated in more comprehensive software tools (e.g., SIFT²⁸, PolyPhen²⁹, CADD³⁰) are merely predictions and may give conflicting

results, if at all accurate, to distinguish pathogenic from non-pathogenic rare alleles.³¹ This was also illustrated by a PolyPhen-based spot check in our own data, whereby 134 homozygous variants in a healthy individual were assigned as being pathogenic (data not shown). Taken together, conservation and pathogenicity predictions are more suited for manual variant verification than for the standard filter step. Notably, the issues of discarding unknown pathogenic mutations and encountering interpretation problems are not WES-specific; they are applicable to other sequencing methods as well. Another reason why disease-causing variants may be discarded in the filter process is that allele frequencies exceed the filter cutoff. If relevant for the population concerned, then a complementary test and/or panel can be offered.

Second, there is a category of variants not detected by WES. One of the potential causes is insufficient callable coverage. Although WES coverage has improved considerably over the past years (~95% capturing of the exome³²), coverage is not complete. In time, whole-genome sequencing techniques will replace WES in clinical genetic practice, thereby further improving completeness of the data.³² To ensure that the most relevant mutations in the population concerned are covered sufficiently, the presence (and expected frequencies) of these mutations should be evaluated in WES data available to the laboratory. If needed, a specific complementary mutation panel could be offered. Technical characteristics of the test are another non-WES-specific cause of missing mutations by sequencing analysis. The ability to detect exon deletions and repeat expansions varies between sequencing techniques. Including these in PCS may require additional test methods. Some relevant examples are indicated elsewhere in the discussion section.

Gene panels. Our data show that the number of presumably pathogenic variants resulting from WES-based PCS is small enough not to need a gene panel. In fact, by using an extensive gene panel, three of the four known pathogenic mutations in the consanguineous couples were discarded, whereas all four were identified in our algorithm without a gene panel. Also, our other data illustrate that several relevant disease-associated genes are not included in the panel (**Supplementary Tables S3 and S4** online). The occasional encounter of a variant with interpretation uncertainties was not prevented by applying gene panels. Therefore, taken together, our results argue against the use of gene panels, especially when aiming a generic PCS test applicable to a broad population with varying ethnicities. The main rationale behind gene panels is to prefilter out certain variants or genes that are deemed undesirable to report for preconception carrier status. Although we do not consider this necessary, if preferred, a gene panel could easily be incorporated in our pipeline before manual evaluation.

Also, creating a gene panel that is complete, given certain agreed criteria, is hardly possible. When omitting gene panels and thereby not restricting the analysis to autosomal-recessive genes, one might encounter autosomal-dominant mutations with potential clinical consequences for the tested individual.

In that case, PCS transforms into presymptomatic testing, which is not deemed desirable in this context. However, when combining the data of both partners, the risk for identifying dominant disease genes is very small. Furthermore, if offspring are at risk for inheriting two dominant disease alleles, then this is of particular relevance in the context of PCS.

Another effect of leaving gene panels out is that carrier couples of, for example, relatively mild and/or treatable disease may be identified. We feel this does not differ significantly from issues encountered in current clinical genetics practice. Adequate counseling is the key for responsible consideration of consequences. Moreover, the manual evaluation of remaining variants includes not only pathogenicity but also clinical implications. With such a post hoc approach, the chances of missing highly relevant pathogenic mutations or genes are smallest. Defining the exact criteria of the findings to be reported is a difficult/arbitrary discussion and can be more easily accomplished with the small number of remaining variants in case-by-case decisions (**Supplementary Tables S3–S5** online) of clinicians and laboratory geneticists. This stresses the importance of performing WES-PCS in a specialized, academic setting.

In conclusion, our results show that WES is eligible for PCS, both for consanguineous and nonconsanguineous couples, with the remaining number of variants being manageable in a clinical setting. Although pathogenic mutations may be missed due to either technical limitations or current lack of knowledge regarding new mutations, respectively, only the first is WES-specific and, with current WES-performance, not a major issue. Specific mutations or genes may be tested complementarily. Overall, we think that WES is able to identify a higher proportion of relevant carrier couples compared to other PCS tests. The omission of gene panels convincingly added value in our experience. Adequate counseling of couples opting for WES-based PCS is critical. PCS may clearly identify a carrier couple or may pose some additional questions, for example, concerning disease severity, the difficulty to predict the phenotype, and the issue of possibly manifesting carriership of X-linked disease, of which discussion in a multidisciplinary PCS team is crucial. Each couple should also be informed that the test is not 100% sensitive; therefore, residual carrier risks remain. It is important to develop and continuously curate databases with pathogenic mutations to reduce manual work in variant analyses and to increase sensitivity of PCS.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

We acknowledge the contribution of the Department of Human Genetics, Radboud UMC, Nijmegen, The Netherlands (involved colleagues: T. Rinne, H. Scheffer, and C. Gillissen, among others), for providing the nonconsanguineous exome data. We also thank I.F.M. de Coö, Department of Neurology, Erasmus MC-Sophia Children's Hospital Rotterdam, Rotterdam, The Netherlands, for enrolling the

consanguineous families in the Rotterdam-Maastricht collaboration for diagnosing mitochondrial disorders, and D.M.E.I. Hellebrekers, Department of Clinical Genetics, Maastricht UMC+, for her diagnostic work in establishing the disease-causing variants in the affected children of the consanguineous couples. This work was supported by Stichting Metakids and Prinses Beatrix Spierfonds.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Henneman L, Borry P, Chokoshvili D, et al. Responsible implementation of expanded carrier screening. *Eur J Hum Genet* 2016;24:e1–e12.
2. Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011;3:65ra4.
3. De Wert GM, Dondorp WJ, Knoppers BM. Preconception care and genetic risk: ethical issues. *J Community Genet* 2012;3:221–228.
4. Becker R, Keller T, Wegner RD, et al. Consanguinity and pregnancy outcomes in a multi-ethnic, metropolitan European population. *Prenat Diagn* 2015;35:81–89.
5. Mathijssen IB, Henneman L, van Eeten-Nijman JM, et al. Targeted carrier screening for four recessive disorders: high detection rate within a founder population. *Eur J Med Genet* 2015;58:123–128.
6. Bajaj K. Carrier screening: past, present, and future. *J Clin Med* 2014;3:1033–1042.
7. Kingsmore S. Comprehensive carrier screening and molecular diagnostic testing for recessive childhood diseases. *PLoS Curr* 2012:e4f9877ab9878ffa9879.
8. Martin J, Asan, Yi Y, et al. Comprehensive carrier genetic test using next-generation deoxyribonucleic acid sequencing in infertile couples wishing to conceive through assisted reproductive technology. *Fertil Steril* 2015;104:1286–1293.
9. Abulí A, Boada M, Rodríguez-Santiago B, et al. NGS-Based Assay for the Identification of Individuals Carrying Recessive Genetic Mutations in Reproductive Medicine. *Hum Mutat* 2016;37:516–523.
10. Teeuw M, Waisfisz Q, Zwijnenburg PJ, et al. First steps in exploring prospective exome sequencing of consanguineous couples. *Eur J Med Genet* 2014;57:613–616.
11. Nguyen M, Boesten I, Hellebrekers DM, et al. Pathogenic CWF19L1 variants as a novel cause of autosomal recessive cerebellar ataxia and atrophy. *Eur J Hum Genet* 2016;24:619–622.
12. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344–347.
13. Plantinga M, Birnie E, Abbott KM, et al. Population-based preconception carrier screening: how potential users from the general population view a test for 50 serious diseases. *Eur J Hum Genet* 2016;24:1417–1423.
14. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 2008;9:403–433.
15. Pietrys D, Balwiercz W, Iwaniec T, Vorjohann S, Neerman-Arbez M, Undas A. Two different fibrinogen gene mutations associated with bleeding in the same family (A α Gly13Glu and γ Gly16Ser) and their impact on fibrin clot properties: fibrinogen Krakow II and Krakow III. *Thromb Haemost* 2011;106:558–560.
16. Lin H, Sinner MF, Brody JA, et al.; CHARGE Atrial Fibrillation Working Group. Targeted sequencing in candidate genes for atrial fibrillation: the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Targeted Sequencing Study. *Heart Rhythm* 2014;11:452–457.
17. Strom CM, Redman JB, Peng M. The dangers of including nonclassical cystic fibrosis variants in population-based screening panels: p.L997F, further genotype/phenotype correlation data. *Genet Med* 2011;13:1042–1044.
18. Schippa S, Iebba V, Santangelo F, et al. Cystic fibrosis transmembrane conductance regulator (CFTR) allelic variants relate to shifts in faecal microbiota of cystic fibrosis patients. *PLoS One* 2013;8:e61176.
19. Hiort O, Sinnecker GH, Holterhus PM, Nitsche EM, Kruse K. The clinical and molecular spectrum of androgen insensitivity syndromes. *Am J Med Genet* 1996;63:218–222.
20. Werner R, Holterhus PM, Binder G, et al. The A645D mutation in the hinge region of the human androgen receptor (AR) gene modulates AR activity, depending on the context of the polymorphic glutamine and glycine repeats. *J Clin Endocrinol Metab* 2006;91:3515–3520.

21. Nordenskjöld A, Söderhäll S. An androgen receptor gene mutation (A645D) in a boy with a normal phenotype. *Hum Mutat* 1998;11:339.
22. Orphanet. *Orphanet Report Series—prevalence of rare diseases: bibliographic data*. http://www.orpha.net/orpha.com/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf. Accessed March 2016.
23. Moskowitz SM, Chmiel JF, Sternen DL, Cheng E, Cutting GR. CFTR-Related Disorders. In: Pagon RA, Adam MP, Ardinger HH, et al. (eds.). *GeneReviews*. University of Washington: Seattle, WA, 1993.
24. Nussbaum RL, McInnes RR, Willard HF. *Thompson & Thompson Genetics in Medicine*. 8th edn. Saunders: Philadelphia, PA, 2015.
25. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;133:1–9.
26. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglu S, Sidow A; NISC Comparative Sequencing Program. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 2005;15:901–913.
27. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;20:110–121.
28. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–3814.
29. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–249.
30. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–315.
31. Li MX, Kwan JS, Bao SY, et al. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS Genet* 2013;9:e1003143.
32. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat* 2015;36:815–822.
33. Kruer MC, Jepperson T, Dutta S, et al. Mutations in γ adducin are associated with inherited cerebral palsy. *Ann Neurol* 2013;74:805–814.
34. Hanks S, Adams S, Douglas J, et al. Mutations in the gene encoding capillary morphogenesis protein 2 cause juvenile hyaline fibromatosis and infantile systemic hyalinosis. *Am J Hum Genet* 2003;73:791–800.
35. Hanein S, Perrault I, Roche O, et al. TMEM126A, encoding a mitochondrial protein, is mutated in autosomal-recessive nonsyndromic optic atrophy. *Am J Hum Genet* 2009;84:493–498.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) (2016)