

Clinical application of a universal NGS-based protocol for diagnosis of β -thalassaemia and sickle cell anaemia in preimplantation embryos

Kubikova, N¹., Babariya, D²., Sarasa, J³., Spath, K²., Alfarawati, S.^{2*}, Wells, D¹.

¹University of Oxford, Nuffield Department of Women's and Reproductive Health, Level 3, Women's Centre, John Radcliffe Hospital, Oxford, OX3 9DU, United Kingdom

²Reprogenetics UK, Institute of Reproductive Sciences, Oxford Business Park North, Alec Issigonis Way, Oxford OX4 2HW, United Kingdom

^{2*}Reprogenetics UK, current address: Oxford Science Park, Magdalen Centre, Robert Robinson Ave, Oxford OX4 4GA, United Kingdom

³iGLS, C/Britania 7, 03540, Alicante, Spain

Abstract

Research question: Mutations of the beta-globin gene (*HBB*) cause beta-thalassaemia and sickle cell anaemia and are the most common cause of severe inherited disease in humans. Traditional PGD protocols for the detection of *HBB* mutations frequently involve labour intensive, patient-specific test design due to the wide diversity of disease associated *HBB* mutations. For this reason, we aimed to develop and clinically apply a universal PGD method to test for mutations in the *HBB* gene.

Design: A multiplex PCR protocol was designed, allowing simultaneous amplification of multiple overlapping DNA fragments encompassing the entire *HBB* gene sequence in addition to 17 characterized, closely linked single nucleotide polymorphisms (SNPs). Amplicons were then analyzed using a next generation sequencing (NGS) method, revealing mutations and SNP genotypes. The protocol was extensively validated, optimized and eventually clinically applied on whole-genome amplified (WGA) DNA derived from embryos

of three couples carrying different combinations of β -thalassemia mutations.

Results: The *HBB* mutation status and associated SNP haplotypes were successfully determined in all 21 embryos. Interestingly, analysis of 141 heterozygous sites showed no instances of allele dropout (ADO) and the test displayed 100% concordance when compared to the results obtained from karyomapping. This suggests that the combination of trophectoderm biopsy and highly sensitive NGS may provide superior accuracy than typically achieved using traditional PGD methods. Importantly, no patient-specific test design or optimization was needed.

Conclusions: It is hoped that protocols that deliver almost universally-applicable low-cost tests, without compromising diagnostic accuracy, will improve patient access to PGD, especially in less affluent parts of the world.

39 **Keywords**

40 *Preimplantation genetic diagnosis, next generation sequencing, beta-thalassaemia, sickle-cell*
41 *anaemia*

42

43 **Key message**

44

45 For monogenic diseases displaying high degree of mutation diversity like beta-thalassaemia,
46 traditional PGD protocols require the creation of numerous distinct tests specific to individual
47 couple, and require time-consuming and expensive customization. These limitations can be
48 overcome using NGS technology, sequencing the entire disease-causing gene along with
49 closely linked polymorphic markers.

50 Introduction

51 Preimplantation genetic diagnosis (PGD) was developed for couples at high risk of transmitting
52 an inherited genetic disorder to their offspring, providing them with an alternative to prenatal
53 diagnosis and termination of affected pregnancy (Handyside et al., 1990). The most common
54 reason for referral for PGD of a single gene disorder is the desire to avoid conditions affecting
55 haemoglobin synthesis, caused by mutations in the *HBB* gene (Moutou et al., 2014). These
56 disorders represent the most common class of gene defect worldwide. Sickle cell anaemia
57 (SCA), a consequence of the Glu6Val point mutation in the *HBB* gene, which results in
58 production of abnormal haemoglobin S (HbS), affects approximately 1 in 500 African
59 Americans and in some parts of West Africa carrier frequencies within the population are
60 estimated to be as high as 1 in 5 (Agasa et al., 2010; Tshilolo et al., 2008). Another disorder
61 caused by *HBB* mutation, β -thalassaemia, is an autosomal recessive disorder common across a
62 broad swathe of the world, from West Africa and the Mediterranean, through the Middle East
63 and across Central and South East Asia (Angastiniotis and Modell, 1998). A wide spectrum of
64 *HBB* mutations are associated with β -thalassaemia, giving rise to significant phenotypic
65 heterogeneity (Cao and Galanello 2010).

66 There is increasing demand for PGD of inherited *HBB* defects, but the vast diversity of
67 potential affected genotypes presents a problem for traditional PGD methods. These typically
68 involve creation of individual protocols, each customized for the detection of a specific
69 combination of mutations. Such strategies have mostly relied on PCR methods capable of
70 simultaneously amplifying DNA fragments encompassing the mutation site(s) as well as DNA
71 polymorphisms located in close proximity to the mutant gene. Identification of informative
72 polymorphic markers, PCR primer design, optimizations of single-cell multiplex PCR to
73 ensure all DNA fragments are sufficiently amplified for subsequent analysis and protocol

validation, require significant investments of time and resources. Frequently, combinations of mutations and informative polymorphisms are unique to an individual family, meaning that the protocol eventually developed, following extensive work in the laboratory, may only be of use for a single couple.

Genetic methodologies are evolving rapidly and offer the possibility of delivering PGD with higher accuracy and lower cost than was possible using earlier techniques. Chief amongst the new wave of methods is next generation sequencing (NGS). This technology has already been adapted for use in assisted reproductive treatment, providing a robust platform for the detection of aneuploidy in embryo biopsy specimens (Wells et al., 2014) (Fiorentino et al., 2014b) (Zheng et al., 2015) (Fiorentino et al., 2014a). Recently, NGS has also been used for the diagnosis of single gene mutations in embryos, although the protocols employed clinically have still required significant customization in terms of designing and optimizing multiplex-PCR for the amplification of mutation sites and polymorphisms (Chen et al., 2016; Ren et al., 2016; Treff et al., 2013; Yan et al., 2015). Here, we describe a novel PGD protocol, based upon NGS technology, for the detection of virtually all mutations responsible for β -thalassemia and sickle cell anaemia. This provides a single cost-effective method applicable to the vast majority of couples seeking PGD for these conditions.

91 Materials and methods

92 *Patients, IVF and embryo biopsy*

93 Prior to clinical implementation of the protocol on embryo biopsies, the technique was
94 validated on genomic DNA obtained from five family trios (each composed of the mother,
95 father and a child or prenatal sample) and two couples, together carrying 14 different *HBB*
96 mutations. Following protocol optimization, three of the families, all healthy carriers of β -
97 thalassaemia, utilized the test clinically for the purpose of PGD. In all three cases, the patients
98 underwent controlled ovarian stimulation and oocytes were collected and fertilized using ICSI,
99 the resulting embryos were biopsied and vitrified at cleavage or blastocyst stages. While in two
100 cases there were no other family members available for testing, one couple had a previous
101 affected pregnancy and a sample of amniotic fluid was included in the study along with the
102 DNA samples extracted from the parents. Analysis of this additional sample allowed the phase
103 of linked polymorphisms to be determined (i.e. revealed which alleles were associated with
104 parental mutations). All patients underwent PGD for β -thalassaemia in different IVF centers
105 and gave consent for NGS-based PGD test to be carried out in parallel with karyomapping (a
106 validated single nucleotide polymorphism SNP-array based technique) (Ben-Nagi et al., 2017;
107 Giménez et al., 2015; Konstantinidis et al., 2015; Natesan et al., 2014). DNA obtained from
108 parents and embryo biopsies was tested at Reprogenetics UK. The diagnosis of β -thalassaemia
109 in embryos has been previously licensed by the Human Fertilisation and Embryology Authority
110 (HFEA). The study was approved by Aspire IRB on 26th August 2015 (reference number
111 PGSP-2015).

112

113 *Genomic DNA extraction and whole genome amplification*

114 Genomic DNA was extracted from 4 ml of parental blood and from an amniotic fluid sample
115 using the QIAamp DNA Blood Mini Kit (Qiagen, Germany), using the standard protocols

recommended by the manufacturer. Extracted DNA and cell(s) obtained from embryo biopsies were lysed as previous described (Thornhill et al., 2015) and subjected to whole-genome amplification with using a REPLI-g Single Cell kit (Qiagen) according to the manufacturer's instructions.

Multiplex PCR, library construction and NGS

An aliquot of each MDA product was further amplified to enrich the regions of interest in a multiplex PCR reaction comprising 8 pairs of primers designed to encompass the entire *HBB* gene locus (Supplementary table 1). As well as these amplicons, which covered all the amino acid coding regions of the gene and all splice donor and acceptor sites, an additional 12 PCR fragments were also amplified, encompassing 17 SNPs flanking the *HBB* gene (Supplementary table 2). All of the targeted SNPs were known to have a high degree of heterozygosity (minor allele frequency [MAF] values higher than 0.24), meaning a high likelihood of informativity for genetic linkage analysis in individual families. For this aspect of the design we employed the NCBI dbVAR database (<https://www.ncbi.nlm.nih.gov/dbvar>) in combination with the ENSEMBL genome browser (<https://www.ensembl.org/index.html>). The multiplex PCR amplification included 0.5 µl of MDA product in a total reaction volume of 50µl, containing 22 µl of nuclease-free water, 25 µl of 2 x Qiagen master mix (Qiagen, Germany), and 2.5 µl of primer mix adjusted to the concentration of 2.5 µM (Eurogentec, UK). Amplifications were performed with the conditions: initial denaturation step at 95°C for 15 min, followed by 55 cycles of denaturation at 94°C for 30 sec, annealing at 56°C for 90 sec, and extension at 72°C for 1 min, with a final extension step at 60° for 10 min. Alongside the clinical samples, we assessed three samples each consisting of a single buccal cell (of normal genotype) and three further samples each composed of five isolated cells from the same source. These were subjected to targeted PCR using the same amplification conditions, but without carrying out of

WGA prior to the multiplex PCR. This confirmed that the protocol could be employed on embryo biopsy specimens directly, avoiding the additional cost of WGA.

The DNA sequencing libraries were prepared using Truseq PCR Free Kit (Illumina) according to the manufacturer's standard work flow. The steps included size selection and clean-up, end repair, adenylation of 3' ends, ligation of indexed adapters and normalization and pooling of the libraries. Cluster generation was carried out on a MiSeq (Illumina) instrument using paired end synthesis 500 cycles MiSeq Reagent kit v2 (Illumina). In parallel, MDA products were processed using karyomapping according to the standard manufacturer's protocol (Infinium Karyomapping Assay Protocol Guide, Illumina).

Data analysis

Sequence alignment files for all indexed libraries were obtained from the Illumina MiSeq Reporter software in BAM format. BAM files were loaded into the Integrative Genomic Viewer (IGV, Broad Institute) to visualize the sequence alignment using the h19 human genome assembly as a reference. Regions of interest were inspected after creating region tracks and importing all of the SNP and mutation positions into the software. Based on the total number of reads for that position, the percentage of bases that did not correspond to the reference nucleotide was determined. The sequences obtained from embryonic DNA were screened for heterozygous loci and compared with the parental DNA, in order to assess allele dropout (ADO) and informativity of selected SNPs. For mutations, the aligned sequences were screened for genotype call at every mutation position in each family. For both categories, the analysis included the determination of sequence coverage at positions of interest and the identification of unexpected variants/polymorphisms within the sequenced DNA, which could serve as additional linkage markers.

164 Results

165 *Validation study using the new NGS protocol - Genotyping of familial β -thalassaemia* 166 *mutations and linkage analysis*

167 All 14 different mutations and all 19 patient genotypes were successfully detected in the
168 *HBB* gene in the five family trios (mother, father, child/prenatal sample) and the two couples.
169 The protocol successfully delivered the complete sequence of the *HBB* gene for all samples
170 tested. The parental genotypes obtained, as well as the mutation status of existing
171 children/prenatal samples (when available), were concordant with the previously determined
172 genotypes in all seven families (Figure 1).

173

174 In addition to the direct mutation detection, all 17 targeted SNPs outside the *HBB* locus (no
175 more than 14 kb away) were examined in addition to patient-specific SNPs found within the
176 *HBB* locus. The polymorphisms informative in each couple were determined - looking for
177 SNPs where one parent was heterozygous and the other was homozygous. For the five families
178 with existing children or prenatal samples, informative SNPs were examined in the offspring's
179 DNA and used to deduce which parental alleles were associated with the mutant and normal
180 copies of the gene. An example of a family-specific haplotype analysis based on the SNP
181 genotyping data produced by the NGS protocol is given (Figure 2).

182 *HBB gene mutation detection in clinical samples*

183 Following the validation study, three of the families have so far proceeded to clinical
184 implementation of the newly designed protocol. In the first couple, the mother carried a
185 substitution c.92+6T>C while the father carried a c.118C>T substitution. In the second couple,
186 the mother carried c.93-21G>A substitution and the father carried a c.316-106C>G

substitution. The second couple have had a previous affected pregnancy and a sample of amniotic fluid was analyzed and confirmed as compound heterozygote for the two parental mutations tested. In the third couple, the mother carried an insertion in codons c.27_28insG while the father carried an c.92+5G>C substitution. In case 1, a single blastomere was biopsied from cleavage stage embryos while in cases 2 and 3, approximately 5 cells from the trophectoderm were biopsied from blastocyst stage embryos. All samples were whole genome amplified and the tests were run in parallel with a standard karyomapping protocol (Konstantinidis et al., 2015). All 21 embryo samples amplified after MDA and NGS analysis was successful for each of these. The read depth for the sequenced amplicons was >1000x in all cases except for embryo 9 where no reads were obtained and karyomapping analysis confirmed nullisomy of chromosome 11. The obtained genetic status at the mutation sites were concordant with the results acquired from karyomapping. Of note, application of the targeted amplification protocol to single buccal cells and clumps, followed by NGS, generated a mean coverage of β -globin sequences of 888x (1061x for single cells samples and 716x for clumps of cells), which was equivalent to the coverage obtained from the clinical samples subjected to WGA prior to targeted PCR. These results suggest that it is possible to use this protocol on embryo biopsies directly, without compromising the read depth, eliminating WGA and thus further reducing costs.

Linkage analysis in clinical samples

In case 1, ten informative SNPs were identified and utilized to assist determination of embryo mutation status - five associated with the maternal mutation and five associated with the paternal mutation (Figure 3a). All SNPs assessed by the protocol described here were located outside the *HBB* gene but within 14 kb of upstream and downstream distance. It is worth noting

that for case 1, although the diagnoses of individual embryos obtained from targeted NGS were concordant with those obtained from the standard PGD procedure, the results from karyomapping were suboptimal - associated with low call rates for individual SNPs (in 50% of the embryos) and high ADO rates (>40%) in all embryo samples tested, likely a consequence of suboptimal WGA following blastomere biopsy and/or degradation of the biopsy specimens. In embryo 9 no diagnosis was possible due to complete absence of chromosome 11 and in embryo 13 only paternal alleles linked to *HBB* were detected. The karyomapping confirmed the presence of chromosome 11 monosomy, explaining this observation. Despite using the same low-quality WGA template, the targeted NGS showed excellent performance in all 13 embryo samples that had provided a karyomapping and NGS result, with no incidence of ADO observed in any of the informative SNP and mutation loci and >1000x sequencing coverage for the entire *HBB* gene and flanking polymorphisms. In case 2, thirteen informative SNPs were identified, out of which two were located within the *HBB* locus. Two of these were inherited with the paternal mutation, while the rest were inherited with the maternal mutation. For one embryo (trophectoderm sample 6) only paternal alleles for SNPs linked to *HBB* were detected. Karyomapping analysis confirmed the presence of chromosome 11 monosomy (Figure 3b). In case 3, ten informative SNPs were analyzed, two located within the *HBB* gene sequence itself (Figure 3c). Six of these were associated with the paternal mutation and four with the maternal mutation. Interestingly, three of the maternal SNPs were not amongst those deliberately targeted by the protocol and appear to be rare variants, which are not present within the dbVAR database.

Surprisingly, of 141 heterozygous sites sequenced there were no instances of ADO in any of the clinical samples tested using the NGS protocol, including the poor-quality WGA products from case 1. This contrasts to an ADO rate of ~12.5% for SNPs genotyped using karyomapping in case 3 and exceeding 40% in case 1, and suggests that the targeted NGS approach is highly

236 sensitive and of excellent diagnostic accuracy.

237

Discussion

Next generation sequencing technology has developed rapidly in recent years and has found numerous research and clinical applications. In the context of preimplantation genetic diagnosis, NGS has principally been employed for the detection of chromosome abnormality (Fiorentino et al., 2014a, 2014b; Wells et al., 2014; Zheng et al., 2015), although there has also been growing interest in its application for the diagnosis of inherited single gene disorders (Chen et al., 2016; Ren et al., 2016; Treff et al., 2013; Yan et al., 2015). In this study, we designed, optimized and implemented an NGS protocol for PGD of β -thalassaemia and sickle-cell anaemia that is applicable to the vast majority of couples at risk of having children affected by these disorders.

Preimplantation genetic testing strategies involving whole genome amplification of embryo biopsy specimens, followed by targeted re-amplification and NGS have recently been reported, allowing successful diagnosis of monogenic disorders. Ren et al. (2016) and Yan et al. (2015) developed strategies for the preimplantation diagnosis of selected inherited disorders, combining direct mutation detection and linkage analysis, and reported live births following the clinical utilization of their methods. While NGS-based approaches are expected to eventually replace conventional PGD methods, one of the major challenges with the current technology lies in the extensive test customization required prior to clinical application. In some cases, the effort required to create a custom-NGS protocol, involving targeted mutation detection and simultaneous analysis of linked polymorphisms, are substantial. Indeed, they may necessitate as much, or more, work than required for a conventional PGD protocol. This makes such approaches impractical for the diagnosis of rare disorders or conditions caused by a wide spectrum of mutations, as in such cases the protocols created following extensive laboratory work may only be applicable to a single family.

In a previous study, Chen et al. (2016) described an NGS-based PGD approach requiring less customization. This was achieved by utilizing a specially designed sequence capture array followed by NGS in order to provide data on the genotype of >24,000 SNPs. The information gathered was subsequently used to assemble haplotypes, allowing diagnosis of embryos that inherited mutant copies of the *PKD2* gene from their parents based upon linkage analysis. This sort of NGS strategy is particularly useful because it delivers high fidelity, sensitivity, and throughput and, since it focuses on the inheritance of common polymorphisms rather than family-specific mutations, requires little work-up for each case. However, one of the limitations of this approach is the cost associated with obtaining an adequate depth of sequencing when assessing a large number of genomic regions. In the current study, although the entire sequence of the *HBB* gene was sequenced at high depth, as well as additional multiple closely linked polymorphic sites, the total proportion of the genome investigated remained very small, equivalent to <10 kb of DNA sequence. This allows simultaneous analysis of large numbers of samples in a single sequencing run, reducing costs. The potential throughput is also much greater compared with PGD methods currently in routine use such as karyomapping.

While PGD methods based upon linkage analysis have the advantage of providing a more generic approach with less patient-specific work-up, a challenge for such strategies is that they cannot be employed in cases where the phase of the SNP alleles cannot be determined (i.e. in cases where it is unclear which alleles are located on the same chromosome as the mutant gene). Deduction of phase not only requires DNA samples from the patients requesting PGD, but also from additional family members who have been previously tested and are of known mutation status (close relatives such as the parents or children of the couple are ideal). Testing of polymorphisms in these extra samples allows the inheritance of specific alleles to be traced through the family, revealing those consistently associated with normal and mutant gene

copies. A lack of DNA samples from close relatives is a common occurrence in PGD, in our experience this is a consideration in about a quarter of all referred cases. Sometimes the couple are reluctant to discuss the fact they are undergoing PGD with other members of the family, other times key relatives may be deceased or unavailable for other reasons, and on some occasions a patient may carry a *de novo* mutation, not present in any relative. Another issue for PGD strategies based entirely on linkage analysis is the possibility that informative alleles, permitting the two chromosomal copies to be distinguished, may be difficult to find. This is a particular problem when offering PGD to consanguineous families as many parts on the genome may be identical due to shared ancestry. In instances such as those outlined, PGD cannot be carried out without direct detection of the causative mutation(s) in the embryos produced.

In the current study, rather than targeting specific mutations, of which there are a very large number in *HBB*, we created a method that provides information on the entirety of the coding region and splice junctions of the gene, as well as selected flanking sequences containing sites of common polymorphism. This approach permits direct detection of virtually all *HBB* gene mutations, effectively eliminating the requirement for patient-tailored test design. The ability to trace the inheritance of defective *HBB* genes by using multiple linked polymorphisms, as well as direct mutation detection, results in a highly redundant test, greatly increasing diagnostic accuracy. In the context of a PGD case, the mutation sites and polymorphisms can each provide their own independent diagnosis. Linkage analysis alone can potentially provide a reliable diagnosis in circumstances where parental mutations have not been identified prior to PGD being undertaken, in cases where mutations are refractory to detection, or in instances where one of the mutation sites fails to amplify appropriately from the embryo biopsy sample as a result of ADO or other technical problems. The extremely close proximity to the *HBB*

gene of the SNPs tested ($\leq 14\text{kb}$ in all cases) makes it highly unlikely that meiotic recombination would ever occur between the polymorphic sites and the sites of mutations (indeed this would be impossible for the intragenic SNPs analyzed). The diagnoses should be highly resistant to errors due to problems caused by failure of individual loci to amplify, preferential amplification and ADO. However, it is noteworthy that during this investigation, considering the embryo biopsy samples with a normal number of chromosome 11 copies, no ADO was detected at any of the 141 heterozygous sites sequenced (0%). Rates of ADO following MDA are influenced by the number of cells within the biopsy specimen and other technical factors. Hou et al. (2015) used the same type of MDA protocol as used in this study and detected a 12.5% ADO rate on single cells re-sequenced at 30x sequence depth. In the current study, the incidence of ADO observed following karyomapping was generally of a similar level, although more than 50% of SNP loci were affected in some samples. The fact that ADO was so low in the current study could be attributable to the high sequencing depth utilized ($>1000\times$). In theory, this should increase the sensitivity for the detection of alleles which are substantially under-represented due to extreme preferential amplification of the alternate allele in a heterozygous sample. Indeed, this was shown to be the case in some of the cleavage stage biopsied embryos analysed in case 1 (embryos 1, 2, 3 and 5), where preferential amplification of several alleles generated over 97% of total reads (Supplementary table 3). In addition to the 17 well-characterized polymorphisms specifically targeted by the PGD protocol described here, an extra three previously uncharacterized intragenic sequence variants/polymorphisms were detected (at positions chr11:5243559, chr11:5243613 and chr11:105236740). These provided a useful additional source of linkage data, further supplementing the diagnosis. The ability of NGS to detect novel polymorphisms/variants unique to individual couples reduces the risk of encountering low informativity when using the method.

338

339 The use of whole genome amplification prior to targeted PCR and NGS, also meant that there
340 was a resource of material available for further testing if desired (e.g. repeat of the original
341 PGD analysis). Furthermore, if desired, low-pass next generation sequencing of the WGA
342 templates generated using multiple displacement amplification can be employed in order to
343 establish the cytogenetic status of the embryo (Wells et al., 2014). The combination of minimal
344 work-up and high throughput provided by this protocol resulted in an extremely economical
345 test. The issue of cost is of great relevance in this particular case, given the fact that many
346 regions of the world where *HBB* mutations are of high prevalence are relatively resource-poor.
347 The experience from the current study confirms that NGS can provide a rapid, streamlined and
348 potentially cost-effective solution for couples seeking to use PGD to avoid genetic disease
349 transmission. It is expected that in the future additional protocols similar to the one described
350 here will be developed for the testing of other single gene disorders where mutation
351 heterogeneity leads to problems for conventional PGD methods. Strategies for PGD based
352 upon NGS technology have the potential to provide the lowest costs without compromising
353 accuracy and are therefore likely to become increasingly popular.

Acknowledgments

DW was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre Programme. NK was supported by the University of Oxford Clarendon Fund.

References

- Agasa, B., Bosunga, K., Opara, A., Tshilumba, K., Dupont, E., Vertongen, F., Cotton, F., Gulbis, B., 2010. Prevalence of sickle cell disease in a northeastern region of the Democratic Republic of Congo: What impact on transfusion policy? *Transfus. Med.* 20, 62–65. doi:10.1111/j.1365-3148.2009.00943.x
- Angastiniotis, M., Modell, B., 1998. Global epidemiology of hemoglobin disorders. *Ann. N. Y. Acad. Sci.* 850, 251–269. doi:10.1111/j.1749-6632.1998.tb10482.x
- Ben-Nagi, J., Wells, D., Doye, K., Loutradi, K., Exeter, H., Drew, E., Alfarawati, S., Naja, R., Serhal, P., 2017. Karyomapping: a single centre's experience from application of methodology to ongoing pregnancy and live-birth rates. *Reprod. Biomed. Online* 35, 264–271. doi:10.1016/j.rbmo.2017.06.004
- Cao, A., Galanello, R., 2010. Beta-thalassemia. *Genet. Med.* 12, 61–76. doi:10.1097/GIM.0b013e3181cd68ed
- Chen, S.-C., Xu, X.-L., Zhang, J.-Y., Ding, G.-L., Jin, L., Liu, B., Sun, D.-M., Mei, C.-L., Yang, X.-N., Huang, H.-F., Xu, C.-M., 2016. Identification of PKD2 mutations in human preimplantation embryos in vitro using a combination of targeted next-generation sequencing and targeted haplotyping. *Sci. Rep.* 6, 25488. doi:10.1038/srep25488
- Fiorentino, F., Biricik, A., Bono, S., Spizzichino, L., Cotroneo, E., Cottone, G., Kokocinski, F., Michel, C.E., 2014a. Development and validation of a next-generation sequencing-based protocol for 24-chromosome aneuploidy screening of embryos. *Fertil. Steril.* 101,

379 1375–1382.e2. doi:10.1016/j.fertnstert.2014.01.051

380 Fiorentino, F., Bono, S., Biricik, A., Nuccitelli, A., Cotroneo, E., Cottone, G., Kokocinski, F.,
381 Michel, C.-E., Minasi, M.G., Greco, E., 2014b. Application of next-generation
382 sequencing technology for comprehensive aneuploidy screening of blastocysts in clinical
383 preimplantation genetic screening cycles. *Hum. Reprod.* 29, 2802–2813.
384 doi:10.1093/humrep/deu277

385 Giménez, C., Sarasa, J., Arjona, C., Vilamajó, E., Martínez-Pasarell, O., Wheeler, K., Valls,
386 G., Garcia-Guixé, E., Wells, D., 2015. Karyomapping allows preimplantation genetic
387 diagnosis of a de-novo deletion undetectable using conventional PGD technology.
388 *Reprod. Biomed. Online* 31, 770–775. doi:10.1016/j.rbmo.2015.08.017

389 Handyside, A.H., Kontogianni, E.H., Hardy, K., Winston, R.M., 1990. Pregnancies from
390 biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature*
391 344, 768–770. doi:10.1097/00006254-199107000-00024

392 Hou, Y., Wu, K., Shi, X., Li, F., Song, L., Wu, H., Dean, M., Li, G., Tsang, S., Jiang, R., Zhang,
393 X., Li, B., Liu, G., Bedekar, N., Lu, N., Xie, G., Liang, H., Chang, L., Wang, T., Chen, J.,
394 Li, Y., Zhang, X., Yang, H., Xu, X., Wang, L., Wang, J., 2015. Comparison of variations
395 detection between whole-genome amplification methods used in single-cell resequencing.
396 doi:10.1186/s13742-015-0068-3

397 Konstantinidis, M., Prates, R., Goodall, N.-N., Fischer, J., Tecson, V., Lemma, T., Chu, B.,
398 Jordan, A., Armenti, E., Wells, D., Munné, S., 2015. Live births following Karyomapping
399 of human blastocysts: experience from clinical application of the method. *Reprod.*
400 *Biomed. Online* 31, 394–403. doi:10.1016/j.rbmo.2015.05.018

401 Moutou, C., Goossens, V., Coonen, E., De Rycke, M., Kokkali, G., Renwick, P., Sengupta,
402 S.B., Vesela, K., Traeger-Synodinos, J., 2014. ESHRE PGD Consortium data collection
403 XII: Cycles from January to December 2009 with pregnancy follow-up to October 2010.

404 Hum. Reprod. 29, 880–903. doi:10.1093/humrep/deu012

405 Natesan, S. a., Handyside, A.H., Thornhill, A.R., Ottolini, C.S., Sage, K., Summers, M.C.,
 406 Konstantinidis, M., Wells, D., Griffin, D.K., 2014. Live birth after PGD with confirmation
 407 by a comprehensive approach (karyomapping) for simultaneous detection of monogenic
 408 and chromosomal disorders. *Reprod. Biomed. Online* 29, 600–605.
 409 doi:10.1016/j.rbmo.2014.07.007

410 Ren, Y., Zhi, X., Zhu, X., Huang, J., Lian, Y., Li, R., Jin, H., Zhang, Y., Zhang, W., Nie, Y.,
 411 Wei, Y., Liu, Z., Song, D., Liu, P., Qiao, J., Yan, L., 2016. Clinical applications of
 412 MARSALA for preimplantation genetic diagnosis of spinal muscular atrophy. *J. Genet.*
 413 *Genomics* 43, 541–547. doi:10.1016/j.jgg.2016.03.011

414 Thornhill, A.R., Handyside, A.H., Ottolini, C., Natesan, S.A., Taylor, J., Sage, K., Harton, G.,
 415 Cliffe, K., Affara, N., Konstantinidis, M., Wells, D., Griffin, D.K., 2015. Karyomapping-
 416 a comprehensive means of simultaneous monogenic and cytogenetic PGD: comparison
 417 with standard approaches in real time for Marfan syndrome. *J. Assist. Reprod. Genet.* 32,
 418 347–56. doi:10.1007/s10815-014-0405-y

419 Treff, N.R., Fedick, A., Tao, X., Devkota, B., Taylor, D., Scott, R.T., 2013. Evaluation of
 420 targeted next-generation sequencing-based preimplantation genetic diagnosis of
 421 monogenic disease. *Fertil. Steril.* 99, 1377–1384.e6. doi:10.1016/j.fertnstert.2012.12.018

422 Tshilolo, L., Kafando, E., Sawadogo, M., Cotton, F., Vertongen, F., Ferster, A., Gulbis, B.,
 423 2008. Neonatal screening and clinical care programmes for sickle cell disorders in sub-
 424 Saharan Africa: Lessons from pilot studies. *Public Health* 122, 933–941.
 425 doi:10.1016/j.puhe.2007.12.005

426 Wells, D., Kaur, K., Grifo, J., Glassner, M., Taylor, J.C., Fragouli, E., Munne, S., 2014. Clinical
 427 utilisation of a rapid low-pass whole genome sequencing technique for the diagnosis of
 428 aneuploidy in human embryos prior to implantation. *J. Med. Genet.* 51, 553–62.

429 doi:10.1136/jmedgenet-2014-102497

430 Yan, L., Huang, L., Xu, L., Huang, J., Ma, F., Zhu, X., Tang, Y., Liu, M., Lian, Y., Liu, P., Li,

431 R., Lu, S., Tang, F., Qiao, J., Xie, X.S., 2015. Live births after simultaneous avoidance of

432 monogenic diseases and chromosome abnormality by next-generation sequencing with

433 linkage analyses. *Proc. Natl. Acad. Sci.* 112, 15964–15969.

434 doi:10.1073/pnas.1523297113

435 Zheng, H., Jin, H., Liu, L., Liu, J., Wang, W.-H., 2015. Application of next-generation

436 sequencing for 24-chromosome aneuploidy screening of human preimplantation embryos.

437 *Mol. Cytogenet.* 8, 38. doi:10.1186/s13039-015-0143-6

438

Mutation analysis													
Mutation	HBB:c.17_18delC T	HBB:c.135delC	HBB:c.92+5G>C	HBB:c.126_129del ICTTT	HBB:c.93-21G>A	HBB:c.92G>A	HBB:c.316- 106C>G	HBB:c.118C>T	HBB:c.92+5G>C	HBB:c.27_28insG	HBB:c.118C>T	HBB:c.92+6T>C	HBB:c.316-106C>G
Position	5248235	5247987	5248155	5247996	5248050	5248159	5247062	5248004	5248155	5248224	5248004	5248154	5247062
Sample													
F1 Male	G/-												
F1 Female		G/-											
F1 Son	G/G	G/G											
F2 Male			C/G										
F2 Female				A/-									
F2 Son			C/G	A/-									
F3 Male					C/T								
F3 Female						C/T							
F3 Daughter					C/C	C/T							
F4 Male							C/G						
F4 Female								A/G					
F4 Daughter							C/G	A/G					
F5 Male									C/G				
F5 Female										A+C			
F5 Prenatal Sample									C/G	A+C			
F6 Male											G/A		
F6 Female												G/A	
F7 Male													G/C
F7 Female													T/C

Figure 1: Summary of the genotypes for β -thalassaemia mutations in families referred for PGD. Each pair of mutations for which the parents are carriers is indicated in the top row with their respective annotations. The seven families are separated by different shades of grey fill. Carrier genotypes are indicated in white and the affected genotypes in black. Five of the seven families already had a child/prenatal sample available, which could be analyzed in parallel with parental DNA samples, assisting in the validation of mutation detection and determination of the phase of linked SNPs.

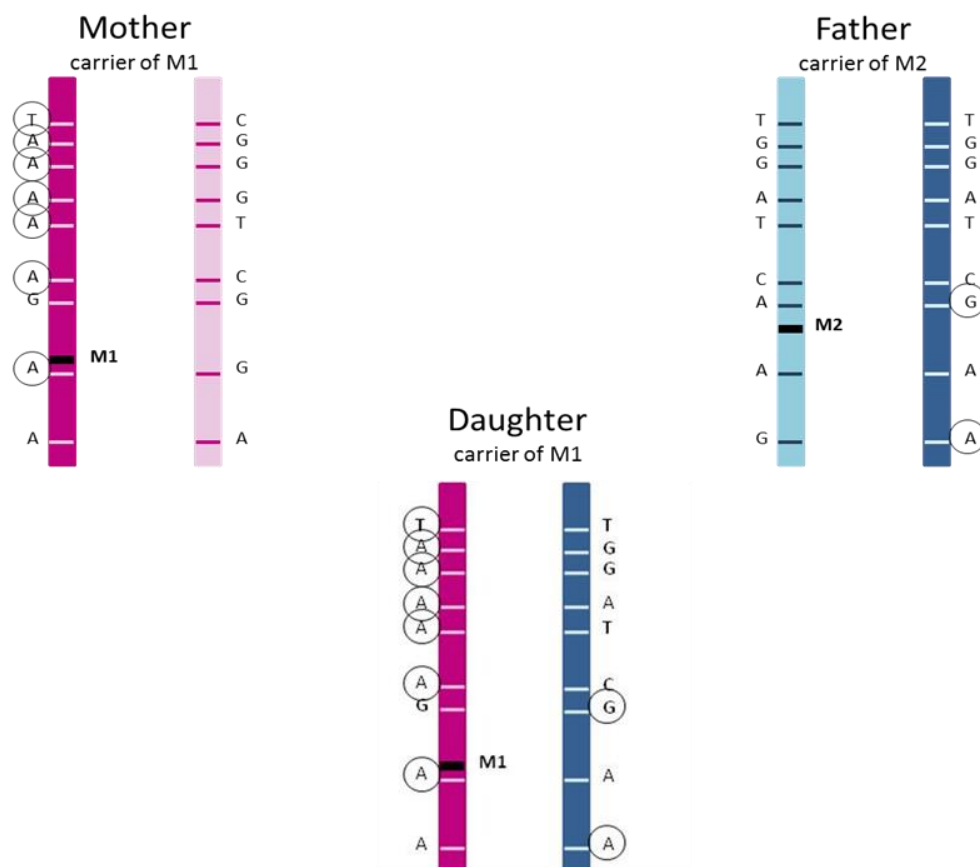


Figure 2: Demonstration of the principle of linkage analysis and targeted mutation detection For both parents, carrier states were confirmed by direct mutation detection. Linkage analysis identified nine informative SNPs in the vicinity of the *HBB* gene in this family. The father is heterozygous for two SNPs for which the mother is homozygous (blue in circle) while the mother is heterozygous for seven SNPs for which the father is homozygous (pink in circle). The inheritance of parental chromosomes in the daughter is thus inferred from nine SNPs as well as direct mutation detection. The daughter has inherited the affected maternal haplotype, associated with the M1 mutation (c.92G>A), and unaffected paternal haplotype, and is therefore a carrier of M1. This was further confirmed by the detection of the maternal c.92G>A mutation in the daughter.

A: Case 1	SNP ID	Position	Mother	Father	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
Linked SNPs	rs3813727	5255912	G/G	A/G	G/G	A/G	A/G	A/G	A/G	A/G	A/G	G/G	no reads / copy of chromosome 11 detected	G/G	G/G	G/G	G	A/G
	rs4910736	5258989	A/A	A/C	A/A	A/C	A/C	A/C	A/C	A/C	A/C	A/A		A/A	A/A	A/A	C	A/C
	rs2105819	5259727	C/C	C/G	C/C	C/G	C/G	C/G	C/G	C/G	C/G	C/C		C/C	C/C	C/C	G	C/G
	rs7936823	5250168	A/A	A/G	A/A	A/G	A/G	A/G	A/G	A/G	A/G	A/A		A/A	A/A	A/A	G	A/G
	rs6578588	5252251	C/C	C/T	C/C	C/T	C/T	C/T	C/T	C/T	C/T	C/C		C/C	C/C	C/C	T	C/T
	rs7945118	5236417	C/G	G/G	C/G	C/G	C/G	G/G	C/G	C/G	C/G	C/G		C/G	G/G	G/G	G	C/G
	rs34220818	5236851	C/T	T/T	C/T	C/T	C/T	T/T	C/T	C/T	C/T	C/T		C/T	T/T	T/T	T	C/T
	rs12364872	5244144	A/G	G/G	A/G	A/G	A/G	G/G	A/G	A/G	A/G	A/G		A/G	G/G	G/G	G	A/G
	rs7110263	5246512	T/G	G/G	T/G	T/G	T/G	G/G	T/G	T/G	T/G	T/G		T/G	G/G	G/G	G	T/G
	rs10837626	5244299	T/A	A/A	T/A	T/A	T/A	A/A	T/A	T/A	T/A	T/A		T/A	A/A	A/A	A	T/A
♀ Mutation	c.92+6T>C	5248154	G/A	A/A	G/A	G/A	G/A	A/A	G/A	G/A	G/A	G/A	no reads / copy of chromosome 11 detected	G/A	A/A	A/A	A	G/A
♂ Mutation	c.118C>T	5248004	G/G	G/A	G/G	G/A	G/A	G/A	G/A	G/A	G/A	G/G		G/G	G/G	G/G	A	G/A
Diagnosis			C	C	MC	A	A	PC	A	A	A	MC	CA	MC	U	U	CA	A

455

B: Case 2	SNP ID	Position	Mother	Father	E4	E6	E7
Linked SNPs	rs4910543	5258827	C/G	C/C	C/G	C	C/C
	rs4910735	5258852	A/G	A/A	A/G	A	A/A
	rs4910544	5258856	A/T	A/A	A/T	A	A/A
	rs4910736	5258989	A/C	A/A	A/C	A	A/A
	rs2105819	5259727	C/G	C/C	C/G	C	C/C
	rs11036364	5249004	A/G	A/A	A/G	A	A/A
	rs6578588	5252251	C/T	C/C	C/T	C	C/C
	rs10837620	5243559	G/A	G/G	A/G	G	G/G
	rs12364872	5244144	G/A	A/A	A/A	A	G/A
	rs10837628	5244404	A/G	A/A	A/G	A	A/A
	rs10837631	5246356	T/A	T/T	A/T	T	T/T
	rs7480526	5247733	C/A	A/A	A/A	A	C/A
	rs63750628	5248281	G/G	G/A	G/A	A	G/G
♀ Mutation	c.93-21G>A	5248050	T/C	C/C	C/C	C	T/C
♂ Mutation	c.316-106C>G	5247062	G/G	G/C	G/C	C	G/G
Diagnosis			C	C	PC	CA	MC

456

C: Case 3	SNP ID	Position	Mother	Father	TE1	TE2	TE4	TE5
Linked SNPs	rs713040	5248243	G/G	G/A	G/G	G/A	G/A	G/A
	rs1609812	5247141	A/A	A/G	A/A	A/G	A/G	A/G
	rs10837631	5246356	A/T	A/A	A/A	A/A	A/A	T/A
	rs7110263	5246512	G/G	G/T	G/G	G/T	G/T	G/T
	rs12574989	5246514	C/C	C/T	C/C	C/T	C/T	C/T
	rs10837626	5244299	A/A	A/T	A/A	A/T	A/T	A/T
	N/A	5243559	G/A	G/G	G/G	G/G	G/G	G/A
	N/A	5243613	C/T	C/C	C/C	C/C	C/C	C/T
	N/A	5236740	G/A	A/A	G/A	G/A	G/A	A/A
	rs7945118	5236417	G/G	G/C	G/G	G/C	G/C	G/C

♀ Mutation	c.27_28insG	5248224	A+C	A	A+C	A+C	A+C	A
♂ Mutation	c.92+5G>C	5248155	C/C	C/G	C/C	C/G	C/G	C/G
Diagnosis			C	C	MC	A	A	PC

457

458 **Figure 3: Results of mutation and linkage analysis: a) Case 1, b) Case 2, c) Case 3.** SNPs where
459 heterozygosity was detected in one parent but not the other were selected for identification of disease-associated
460 alleles in the embryo WGA products subjected to targeted multiplex amplification in three clinical cases. In green
461 are the alleles associated with the paternal *HBB* mutation and in red are the alleles inherited together with the
462 maternal *HBB* mutation. Black alleles represent the disease-free genotype. C: Carrier, MC: carrier of maternal
463 mutation, PC: carrier of paternal mutation, A: affected embryo, U: unaffected embryo, CA: chromosomally
464 abnormal, N/A: not applicable, the SNP IDs do not exist for these variants as they were not present within the
465 dbVAR database

466 Supplementary table 1: The list of HBB gene primer sequences

467

HBB gene			
Amplicon ID	Forward Primer	Reverse Primer	Amplicon Size
1	AGTCAGGGCAGAGCCATCTA	GTCTCCACATGCCCAGTTTC	229
2	CAAGACAGGTTTAAGGAGACCAA	ACTTAACCATAGAAAAGAAGGGG	391
3	ATGGGACGCTTGATGTTTTTC	TGTACTAGGCAGACTGTGTAAAG	323
4	TGTGTATAACAAAAGGAAATATCTCTG	GCCCTGAAAGAAAGAGATTAGG	328
5	CACATATTGACCAAATCAGGGTA	TGCTATTGCCTTAACCCAGAA	200
6	ATGCCTCTTTGCACCATTCT	CCAGCCTTATCCCAACCATA	183
7	TCCAGCTACCATTCTGCTTTT	GGACTTAGGGAACAAAGGAACC	294
8	CTCGCTTTCTTGCTGTCCAA	ATGCACTGACCTCCCACATT	194

468 Supplementary table 2: The list of extragenic SNP primer sequences

#	SNP ID	Position	MAF	Alelles	Amplicon ID	Forward Primer	Reverse Primer	Amplicon Size
Centromeric SNPs								
1	rs3813727	5,234,682	0.4918	T/C	9	TGGCTGTTCTGTCATGTGTG	CAACCTCTCAAAATTCCTTGG	288
3	rs4910543	5,237,597	0.4064	G/C	11	TGAGTCTGAGGTGCCTATA	ATCTCCTACCTGCTCTGAA	363
4	rs4910735	5,237,622	0.406	T/C				
5	rs4910544	5,237,626	0.406	A/T				
6	rs4910736	5,237,759	0.4099	G/T	12	TCCAAGAGTGTGATGAATAC	CCAGCCAAGAATGTGAAT	380
7	rs2105819	5,238,497	0.4071	G/C	13	GAACAATGCCTAGAGACA	GAATGGTAATTGACAGAAGG	200
10	rs11036364	5,227,774	0.4038	T/C	15	CTGCGTCTCCAGAATATG	TTGACACCACTGATTACC	394
11	rs7936823	5,228,938	0.3776	C/T	16	TGGTTCTTCTATGGCTATCT	GTGAAACAGGGTCTTGAAA	322
12	rs6578588	5,231,021	0.355	G/A	17	TCACTGGGTCTTGATGTACAGA	GCCGAGCACACACAATTACT	308
Telomeric SNPs								
13	rs7945118	5,215,187	0.3299	G/C	18	TCTGTGATGCCTCCTTTG	TTCTCCAGTGGATTCTTG	189
14	rs34220818	5,215,621	0.498	G/A	19	GGATCTCAGTCACCAAGGCT	TGGAATCAACAAGCTAGGGGA	277
15	rs10837620	5,222,329	0.246	C/T	20	AATTGCTGGGATTACACATGC	CAACCCAAAGTAGAACTATCAAGG	244
16	rs12364872	5,222,914	0.3253	C/T	21	TGAAGCCATTTTATAGATAAACCAA	TGCATCTTGATGATTAGAATTGC	361
17	rs10837626	5,223,069	0.3293	A/T				
18	rs10837628	5,223,174	0.2794	C/T				
22	rs10837631	5,225,126	0.248	A/T	24	AGACAACAGAGACAATAAG	ACAGCTAATGCACATTGG	309
23	rs7110263	5,225,282	0.2897	C/A				

469 Supplementary table 3: The list of genotypes at SNP and HBB mutation positions in the tested embryos

Case 1	rs4910736		rs7945118		rs2105819		rs12364872		rs7936823		rs34220818		rs6578588		rs7110263		rs10837626		c.92+6T>C		c.118C>T	
	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %
Mother	A/A	100	C/G	47/53	C/C	100	A/G	43/57	A/A	100	C/T	48/52	C/C	100	T/G	71/29	T/A	44/56	G/A	52/48	G/G	100
Father	A/C	51/49	G/G	100	C/G	50/50	G/G	100	A/G	53/47	T/T	100	C/T	53/47	G/G	100	A/A	100	A/A	100	G/A	49/51
E1	A/A	100	C/G	91/9	C/C	100	A/G	93/7	A/A	100	C/T	90/10	C/C	100	T/G	90/10	T/A	96/4	G/A	97/3	G/G	97/3
E2	A/C	20/80	C/G	90/10	C/G	7/93	A/G	91/9	A/G	97/3	C/T	88/12	C/T	98/2	T/G	88/12	T/A	88/12	G/A	96/4	G/A	65/35
E3	A/C	84/16	C/G	99/1	C/G	85/15	A/G	70/30	A/G	54/46	C/T	2/98	C/T	42/58	T/G	47/53	T/A	71/29	G/A	61/39	G/A	69/31
E4	A/C	55/45	G/G	100	C/G	56/44	G/G	100	A/G	35/65	T/T	100	C/T	38/62	G/G	100	A/A	100	A/A	100	G/A	86/14
E5	A/C	66/34	C/G	3/97	C/G	64/36	A/G	2/98	A/G	84/16	C/T	4/96	C/T	82/18	T/G	1/99	T/A	3/97	G/A	2/98	G/A	4/96
E8	A/A	100	C/G	66/34	C/C	100	A/G	29/71	A/A	100	C/T	65/35	C/C	100	T/G	14/86	T/A	29/71	G/A	23/76	G/G	100
E10	A/A	100	C/G	51/49	C/C	100	A/G	20/80	A/A	100	C/T	49/51	C/C	100	T/G	9/91	T/A	17/83	G/A	84/16	G/G	100
E11	A/A	100	G/G	100	C/C	100	G/G	100	A/A	100	T/T	100	C/C	100	G/G	100	A/A	100	A/A	100	G/G	100
E12	A/A	100	G/G	100	C/C	100	G/G	100	A/A	100	T/T	100	C/C	100	G/G	100	A/A	100	A/A	100	G/G	100
E14	A/C	20/80	C/G	44/56	C/G	19/81	A/G	21/79	A/G	29/71	C/T	42/58	C/T	24/76	T/G	16/84	T/A	21/79	G/A	30/69	G/A	38/62

470

Case 2	rs4910543		rs4910735		rs4910544		rs4910736		rs2105819		rs11036364		rs6578588		rs10837620		rs12364872		rs10837628		rs10837631		rs7480526		rs63750628		c.93-21G>A		c.316-106C>G	
	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %
Mother	C/G	48/51	A/G	49/51	A/T	49/51	A/C	55/45	C/G	49/51	A/G	62/38	C/T	51/49	G/A	53/47	G/A	51/49	A/G	50/50	T/A	52/48	C/A	50/50	G/G	100	T/C	45/55	G/G	100
Father	C/C	100	A/A	100	A/A	100	A/A	100	C/C	100	A/A	100	C/C	100	G/G	100	A/A	100	A/A	100	T/T	100	A/A	100	G/A	50/50	C/C	100	G/C	75/24
E4	C/G	32/68	A/G	29/71	A/T	30/70	A/C	35/65	C/G	35/65	A/G	46/54	C/T	38/61	A/G	37/63	A/A	100	A/G	61/39	A/T	100	A/A	100	G/A	66/34	C/C	100	G/C	83/17
E7	C/C	100	A/A	100	A/A	100	A/A	100	C/C	100	A/A	100	C/C	100	G/G	100	G/A	54/46	A/A	100	T/T	100	C/A	51/49	G/G	100	T/C	48/51	G/G	100

471

Case 3	rs713040		rs1609812		rs10837631		rs7110263		rs12574989		rs10837626		N/A		N/A		N/A		rs7945118		c.27_28insG		c.92+5G>C	
	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %	Genotype	Read %
Mother	G/G	100	A/A	100	A/T	46/54	G/G	100	C/C	100	A/A	100	G/A	50/50	C/T	50/50	G/A	50/50	G/G	100	A+C	69/31	C/C	100
Father	G/A	51/49	A/G	49/51	A/A	100	G/T	37/63	C/T	37/63	A/T	54/47	G/G	100	C/C	100	A/A	100	G/C	51/49	A	100	C/G	51/49
TE1	G/G	100	A/A	100	A/A	100	G/G	100	C/C	100	A/A	100	G/G	100	C/C	100	G/A	25/75	G/G	100	A+C	80/20	C/C	100
TE2	G/A	23/77	A/G	25/75	A/A	100	G/T	18/82	C/T	18/82	A/T	30/70	G/G	100	C/C	100	G/A	35/65	G/C	34/66	A+C	82/18	C/G	24/76
TE4	G/A	65/35	A/G	66/34	A/A	100	G/T	57/43	C/T	57/43	A/T	73/27	G/G	100	C/C	100	G/A	69/31	G/C	70/30	A+C	61/39	C/G	65/35
TE5	G/A	83/17	A/G	83/17	T/A	24/76	G/T	76/24	C/T	76/24	A/T	89/11	G/A	15/85	C/T	15/85	A/A	100	G/C	85/15	A	100	C/G	84/16

472

473 Summary of the genotypes at informative SNP and mutation loci in the analysed parental and embryonic DNA in cases 1,2

474 and 3 where heterozygosity was detected with the corresponding read percentage per allele.