Single Molecule Sequencing of Free DNA from Maternal Plasma for Noninvasive Trisomy 21 Detection

Jessica M.E. van den Oever,1† Sahila Balkassmi,1† E. Joanne Verweij,2 Maarten van Iterson,3 Phebe N. Adama van Scheltema,2 Dick Oepkes,2 Jan M.M. van Lith,2 Mariëtte J.V. Hoffer,1 Johan T. den Dunnen,4 Egbert Bakker,3 and Elles M.J. Boon1*

BACKGROUND: Noninvasive fetal aneuploidy detection by use of free DNA from maternal plasma has recently been shown to be achievable by whole genome shotgun sequencing. The high-throughput next-generation sequencing platforms previously tested use a PCR step during sample preparation, which results in amplification bias in GC-rich areas of the human genome. To eliminate this bias, and thereby experimental noise, we have used single molecule sequencing as an alternative method.

METHODS: For noninvasive trisomy 21 detection, we performed single molecule sequencing on the Helicos platform using free DNA isolated from maternal plasma from 9 weeks of gestation onwards. Relative sequence tag density ratios were calculated and results were directly compared to the previously described Illumina GAII platform.

RESULTS: Sequence data generated without an amplification step show no GC bias. Therefore, with the use of single molecule sequencing all trisomy 21 fetuses could be distinguished more clearly from euploid fetuses.

CONCLUSIONS: This study shows for the first time that single molecule sequencing is an attractive and easy to use alternative for reliable noninvasive fetal aneuploidy detection in diagnostics. With this approach, previously described experimental noise associated with PCR amplification, such as GC bias, can be overcome.

© 2011 American Association for Clinical Chemistry

Trisomy 21 (T21)5 is the most common chromosomal abnormality in live-born children. The diagnosis can be made early in pregnancy by use of invasive testing [e.g., chorionic villus sampling (CVS) or amniocentesis]. These invasive procedures, however, are associated with a risk of miscarriage. Therefore, these tests are commonly offered only to women at increased risk for fetal trisomy. Risk assessment used to be performed on the basis of maternal age. More recently, this criterion was refined by adding serum markers for trisomy and ultrasound measurement of the fetal nuchal translucency (1). Current screening programs have detection rates for T21 of around 80% with a false-positive rate of 5%, meaning that 1 in every 20 women screened is offered invasive testing with its inherent risks, while carrying a healthy fetus (2, 3).

The discovery of cell-free fetal (cff)-DNA and cffRNA in maternal plasma opened possibilities for noninvasive prenatal diagnosis (4). Although cffRNA has been used for noninvasive T21 detection (5–8), the majority of approaches use cffDNA for noninvasive prenatal diagnosis of T21. In the first trimester, the percentage of cffDNA in maternal plasma is on average 1%–10% and differs quite extensively in range depending on gestational age and between individuals (9–14). Therefore, it remains challenging to detect fetal sequences in a large pool of maternal DNA. Previously, several reported studies have shown that noninvasive T21 detection is possible by use of single nucleotide polymorphisms (15, 16) and epigenetic analysis (17–20), although these methods have several limitations.

In 2008, noninvasive T21 detection by next-generation sequencing (NGS) was introduced (21, 22), opening a whole new way of analysis. No longer were only fetal-specific sequences analyzed, because this technique sequenced all free DNA in plasma, both fetal and maternal in origin. Two recent reports confirmed the potential value of NGS for noninvasive fetal T21 detection in multiplexed plasma DNA samples in a
clinical setting (23, 24). Both the Illumina genome analyzer (GA) II (21–25) and the Solid platform (26) have been used for noninvasive T21 detection by NGS. These platforms use amplification steps by PCR that are known to introduce preferential amplification of sequences depending on different GC content (21, 27).

In the present study, we tested single molecule sequencing (Helicos Heliscope™ single molecule sequencer) for noninvasive T21 detection. The Helicos platform uses visual imaging across the flow cell for direct DNA measurement by recording the incorporation of fluorescently labeled nucleotides (28, 29). The use of single molecule sequencing has been described previously (30), and this technique should largely overcome the limitations associated with PCR amplification and bias as mentioned above. Although the sequencing time on the Helicos platform is longer compared to the Illuma platform (4 days vs 2 days, respectively), Helicos sample preparation is simple, 3 times faster (1 day compared to 3 days), and therefore relatively inexpensive. Furthermore, this method requires low amounts of DNA, a feature that could be particularly advantageous early in gestation.

Here, we present a comparison of the application of single molecule sequencing with the previously described PCR-based Illumina NGS platform for noninvasive T21 detection by use of cfDNA from maternal plasma.

Materials and Methods

STUDY PARTICIPANTS
Pregnant women undergoing prenatal diagnosis were recruited at the Department of Obstetrics of the Leiden University Medical Center, Leiden, the Netherlands. Informed consent was obtained from all study participants, and this study was approved by the institution’s medical ethics committee.

SAMPLE PROCESSING AND ISOLATION
Maternal peripheral blood samples (10–20 mL) were collected in EDTA-coated tubes at the Leiden University Medical Center and were processed within 24 h after collection. All blood samples were drawn at a median gestational age of 12 + 2 weeks (range 9 + 3 to 16 + 6 weeks). Preferably blood samples were drawn before an invasive procedure. If this was not possible samples were drawn at least 5 days after the invasive procedure to minimize any disturbance with fetal material owing to the procedure.

Blood was centrifuged at 1200g (without brake) for 10 min at room temperature. Plasma was transferred to 15-mL microcentrifuge tubes and centrifuged at 2400g for 20 min (with brake) at room temperature to remove residual cells. Cell-free plasma was divided into 800-µL aliquots and stored at −80°C until further processing.

Because each sequencing platform requires different amounts of input DNA, cell-free DNA was isolated from plasma with the EZ1 Virus Mini Kit v2.0 on the EZ1 Advanced (QIAGEN www.qiagen.com) for Helicos sample preparation or manually with the QiaAmp MinElute Virus Spin Kit (QIAGEN) for Illumina sample preparation, according to the manufacturer’s instructions.

To verify fetal sex and to measure the total quantity of cell-free DNA, we performed a pyrophosphorylation-activated polymerization assay on the Y chromosome (Y-PAP) and a real-time Taqman PCR assay on chemokine (C-C motif) receptor 5 (CCR5)6 for QC purposes, as described previously (31). In addition, for male fetuses we estimated the percentage of cfDNA on the basis of sequencing data of chromosome X (21) and by real-time Taqman PCR assay on sex determining region Y (SRY), for which we used a standard curve from male genomic DNA to determine the range of cfDNA percentages in maternal plasma. Percentages were estimated by dividing the amount of SRY (pg/µL) by the maternal fraction of CCR5 from 1 allele [SRY/(0.5*CCR5_total – SRY)], taking into account that the PCR efficiency of both genes is similar.

LIBRARY PREPARATION AND SEQUENCING
A total of 24 plasma samples was included in this retrospective study, including 20 samples from singleton pregnancies, consisting of 11 cases (5 female and 6 male fetuses) of T21, 9 cases of disomy (D21) pregnancies (1 female and 8 male fetuses), and 4 plasma control samples from anonymous adult male blood donors. All samples were deidentified to the investigators before sample preparation and data analysis. These results were not revealed to the investigators until after data analysis. Material from the invasive procedure was sent to the cytogenetics laboratory for full karyotyping. Fetal sex was confirmed by karyotype or after birth.

All cell-free plasma DNA samples were sequenced on both the Helicos (Helicos BioSciences www.helicosbio.com) and the Illumina (Illumina www.illumina.com) GA II platform. Owing to the relatively short length (25) and fragmented nature of free DNA in plasma, no additional shearing step was performed during library preparation.

Helicos sample preparation was performed according to the manufacturer’s ChiP-Seq direct tailing procedure with an input of 400 µL plasma for DNA

---

6 Human genes: CCR5, chemokine (C-C motif) receptor 5; SRY, sex determining region Y.
isolation with the EZ1 (QIAGEN) and the maximum amount of input for tailing. As a QC, size of the fragments and template size distribution were determined by running a high-sensitivity DNA chip on the Agilent Technologies 2100 Bioanalyzer. A standard 120-cycle run was performed on the Heliscope single molecule sequencer, which resulted in a mean read length of 35 nucleotides.

Illumina sample preparation was performed according to the manufacturer’s ChiP-Seq protocol with an input of 1600 μL plasma per sample per column for manual DNA isolation and a maximum amount of input for this protocol. Sixteen of the 24 samples were sequenced in a duplex assay (T21, n = 7; D21, n = 7; and male plasma controls, n = 2). For this procedure, unique synthetic 6-nucleotide barcodes (indexes) were used. The barcode was ligated to the plasma DNA molecule before the PCR amplification step. Indexed samples were additionally purified on a 3% TAE Agarose gel before the QC run on the Bioanalyzer, as mentioned above. A 36-cycle run was performed on the Illumina GA II.

DATA ANALYSIS
Helicos sequencing data were analyzed with the Helicos Helisphere resequencing pipeline by use of the default settings. Data were aligned against the hg19 reference genome and gaps and repeats were filtered out. Filtered data were sorted and binned per 50 kb.

Illumina raw data from duplexed samples were preanalyzed by splitting the data per indexed barcode with inhouse Linux command lines. Sequencing data were analyzed with NextGENe software (SoftGenetics www.softgenetics.com). Data were mapped to the annotated Human Genome GFCch37-dbSNP 131(4/14/2010) (hg19) for Illumina data compatible with NextGENe software. Expression reports per 50 kb were created. Only unique reads with at most 1 mismatch, which could be aligned to the reference genome, were used for calculations.

For all samples (both T21 and D21) used for non-invasive fetal T21 detection in maternal plasma, ratios of relative sequence tag density (RSTD) were calculated. First, for each sample the total number of reads was calculated per chromosome, by summing the read counts of all 50-kb bins belonging to a particular chromosome. Second, for each sample, the total summed number of reads was normalized by the median value of the autosomes. Finally, ratios of RSTD were calculated by dividing these normalized values by the averaged normalized value of the disomy samples (21) or, in addition, by the normalized mean of male plasma control samples. Because the data were obtained by 2 separate runs for both sequencing technologies, ratios were determined for each run separately.

STATISTICAL ANALYSIS
Statistical analysis was conducted with PASW Statistics version 17.0 (SPSS www.spss.com), Prism 5 (version 5.00, GraphPad Software www.graphpad.com), and R version 2.13 [R Development Core Team (2011) www.R-project.org]. Differences between the numbers of uniquely mapped reads between groups were determined by independent samples t-test. Correlation between the number of reads and RSTD were determined by nonparametric Spearman correlation. P values of <0.05 were considered statistically significant.

Results

INCLUDED SAMPLES
A total of 20 maternal plasma samples were included in this study and were taken at a median gestational age of 12 + 2 weeks (range 9 + 3 to 16 + 6 weeks). In 4 of 20 cases blood samples were drawn after the invasive procedure (mean sampling time >1 week afterward). No correlation between the time of sampling (before or after the invasive procedure) and the ratios was observed. All details on the included samples are depicted in Table 1. For the noninvasive detection of fetal T21, DNA isolated from 20 maternal plasma samples and 4 anonymous male plasma controls was sequenced on both the Helicos and the Illumina GA II platform. One D21 sample, which did pass the QCs before sequencing, failed the QCs after sequencing on both platforms. For this sample very few reads were obtained for the Helicos platform, and sequencing results from the Illumina platform showed preferential amplification of only a few regions. This sample was therefore excluded from further analysis.

SAMPLE STATISTICS
For each NGS platform, the mean number of raw reads, the percentage of filtered reads, and the mean and median number of uniquely mapped reads are depicted in Table 2. For Helicos, our data showed 1 D21 sample, with the overall lowest amount of reads, to have the lowest RSTD ratio, but overall we observed no correlation between RSTD ratio and the number of uniquely mapped reads on either platform [Helicos, Spearman r = −0.088 (95% CI −0.532 to 0.394) P = 0.7210 and Illumina, Spearman r = −0.232, 95% CI −0.629 to 0.263] P = 0.3401]. Furthermore, the numbers of uniquely mapped reads were not found to differ between T21 and D21 (Helicos P = 0.128 and Illumina P = 0.810). When we looked at the duplexed Illumina samples (n = 16), we observed no bias in read counts toward any specific barcode after splitting (P = 0.9551).

The percentage of cffDNA in maternal plasma was calculated by 2 different methods. When we used the
method of Fan et al. based on Illumina sequence data from chromosome X (21), we estimated the percentage of fetal DNA for male pregnancies (n = 6) to be on average approximately 7% (range 1%–18%). Concordant results were obtained by real-time PCR on the SRY gene (mean approximately 9%, range 3%–18%).

NONINVASIVE T21 DETECTION

For the noninvasive detection of fetal T21, RSTD ratios for all 19 maternal plasma samples are shown per chromosome for each NGS platform (Fig. 1). The autosomes were ordered by increasing GC content (21).

The overall distribution of reads across the genome was similar between both platforms and seemed independent of GC content (data not shown). However, our data showed a clear difference in read coverage between platforms. For Helicos, the RSTD ratios for all chromosomes (Fig. 1), the normalized total number of reads per chromosome (Fig. 2), and the mean number of reads per bin (Fig. 3) were quite uniform between samples and virtually independent of GC content of the chromosome, whereas as reported before (21), Illu-

| Table 1. Overview of included maternal plasma samples. |
|-----------|----------------|-------------------|-----------------|----------------|----------------|-----------------|
| Sample number | Karyotype | Maternal age, years | Indication | AD blood<sup>a</sup> | CVS/amnio | AD procedure |
| 1 | 46,XY | 35 | Family history of mental retardation/ICSI pregnancy | 12 + 2 | Amnio | 16 + 1 |
| 2 | 46,XY | 39 | Advanced maternal age | 9 + 6 | Amnio | 17 + 0 |
| 3 | 47,XX,+21 | 41 | Ultrasound abnormality: hygroma colli and generalized edema | 15 + 3 | CVS | 14 + 2 |
| 4 | 47,XY,+21 | 43 | Advanced maternal age | 10 + 2 | Amnio | 16 + 0 |
| 5 | 46,XX | 40 | Advanced maternal age | 10 + 1 | Amnio | 16 + 0 |
| 6 | 46,XY | 38 | Advanced maternal age | 11 + 2 | Amnio | 17 + 1 |
| 7 | Male | 39 | Increased NT, advanced maternal age | 9 + 3 | - | - |
| 8 | Male | 37 | Increased NT, advanced maternal age | 13 + 2 | - | - |
| 9 | 47,XY,+21 | 34 | Ultrasound abnormality | 13 + 2 | CVS | 12 + 0 |
| 10 | Male | 38 | Increased NT, advanced maternal age | 11 + 0 | - | - |
| 11 | 47,XX,+21 | 41 | Advanced maternal age | 16 + 6 | Amnio | 16 + 1 |
| 12 | 47,XY,+21 | 41 | Advanced maternal age | 14 + 4 | CVS | 12 + 3 |
| 13 | 47,XX,+21 | 41 | Ultrasound abnormality: hydrothorax, cystic hygroma | 11 + 2 | CVS | 11 + 3 |
| 14 | 46,XY | 40 | Advanced maternal age | 11 + 5 | CVS | 11 + 5 |
| 15 | 46,XY | 39 | Advanced maternal age | 16 + 6 | Amnio | 16 + 6 |
| 16 | 47,XY,+21 | 38 | Ultrasound abnormality, | 12 + 2 | CVS | 12 + 2 |
| 17 | 47,XX,+21 | 39 | Ultrasound abnormality: hygroma colli | 12 + 1 | CVS | 12 + 1 |
| 18 | 47,XX,+21 | 36 | Ultrasound abnormality: hygroma colli | 14 + 2 | CVS | 14 + 2 |
| 19 | 47,XY,+21 | 35 | Increased NT | 12 + 4 | CVS | 12 + 4 |
| 20 | 47,XY,+21 | 34 | Ultrasound abnormality: hygroma colli | 13 + 1 | CVS | 13 + 1 |

<sup>a</sup> AD blood, gestational age at the time of blood collection depicted as weeks + days; amnio, amniocentesis; AD procedure, gestational age at the time of the invasive procedure depicted as weeks + days; ICSI, intracytoplasmic sperm injection; increased NT, increased nuchal translucency thickness.

| Table 2. Overview of mean and median number of uniquely mapped reads.<sup>a</sup> |
|----------------|----------------|----------------|-----------------|----------------|----------------|-----------------|
| Platform | Mean no. of raw reads | % Filtered reads | Mean no. of aligned reads | Median no. of aligned reads |
| Helicos | 1.06 × 10<sup>8</sup> (0.46 × 10<sup>6</sup>)<sup>b</sup> | 35.09 (1.52) | 4.65 × 10<sup>6</sup> (3.6 × 10<sup>6</sup>)<sup>b</sup> | 2.65 × 10<sup>6</sup> |
| Illumina | 2.47 × 10<sup>7</sup> (0.38 × 10<sup>6</sup>)<sup>b</sup> | 76.07 (5.14)<sup>b</sup> | 1.26 × 10<sup>7</sup> (0.40 × 10<sup>6</sup>)<sup>b</sup> | 1.26 × 10<sup>7</sup> |

<sup>a</sup> Results are indicated by NGS platform. Data for each platform are represented as mean (SD) for n = 23 samples.
<sup>b</sup> For the Illumina platform, the mean number of raw reads and % filtered reads is depicted for the duplexed samples, n = 15 (T21, n = 7; D21, n = 6; male plasma control, n = 2).
mina results showed increased read density in GC-rich areas of the genome (Figs. 1–3).

Our data show RSTD ratios for T21 samples in a range of 1.04–1.11 for Helicos and a range of 1.03–1.12 for Illumina. For D21 samples we obtained RSTD ratios of 0.98–1.01 and 0.99–1.01, respectively (Fig. 4). Our data showed a clear distinction between plasma samples from women carrying a T21 fetus and women carrying a D21 fetus for both platforms when we looked at the overrepresentation of the affected chromosome (Fig. 4). All maternal plasma samples of women carrying a fetus with Down syndrome were correctly classified as T21 (n = 11). In addition, all euploid samples (n = 8) were correctly identified as D21, resulting in a sensitivity and specificity of both 100% (95% CI [87.0–100]). When constructing a 99% CI of the distribution of RSTD from all D21 samples; all T21 samples were outside the upper boundary of 1.01 and all D21 samples on or below this boundary. Overall, we showed that noninvasive detection of T21 can be performed on both NGS platforms, although Helicos results show a better distinction between T21 and D21 samples (Fig. 4).

CALCULATION METHODS
We have based our calculations on the method of Fan et al. (21), which uses read counts to calculate ratios of RSTD. Samples can be normalized against averaged normalized RSTD from both adult male plasma controls (Fig. 1) or disomy samples (Fig. 4). Our results looked similar when we applied either one of these methods to data from both NGS platforms. Recently, a new calculation method for the detection of fetal chromosomal abnormalities was published by the group of Sehnert et al. (32). With this method, samples can be classified as affected (i.e., aneuploid for that chromosome) or unaffected by calculation of a normalized chromosome value by using data from a previously analyzed training set consisting of unaffected samples (i.e., maternal plasma samples from women carrying a euploid fetus). When applying this new calculation method to our Illumina data, we found that all Illumina samples were again correctly classified as either T21 or D21, within the criteria as described (see Supplemental Fig. S1 1 in the Data Supplement that accompanies the online version of this article at http://
Because these criteria are determined only for Illumina data, they are not applicable on our Helicos results and thus first need to be established.

Fig. 2. Normalized total number of reads per chromosome against GC content.
Normalized reads are shown in order of GC content per chromosome (upper panel) and by GC percentage (lower panel) for both the Helicos (left) and Illumina (right) platform. Chromosomes subject to possible GC bias because of high GC content are depicted within dashed lines (upper and lower panel).

Fig. 3. Mean read count per 50-kb bin against GC content.
For both platforms, the mean number of reads per bin was determined by dividing the summed total number of reads by the number of bins. Chromosomes are ordered by GC content.

Fig. 4. Ratios of normalized RSTD from chromosome 21.
Ratios were calculated against averaged normalized read counts from disomy samples. Data are shown for each NGS platform (T21 n = 11 and D21 n = 8).
Discussion

Noninvasive fetal aneuploidy detection by use of free DNA from maternal plasma has evolved dramatically the past few years with the introduction of NGS. The majority of studies use the Illumina GA II platform for whole genome shotgun detection of T21. Data obtained in these studies have shown that limitations owing to low percentages of cfDNA in maternal plasma no longer seem to be a major problem. However, the Illumina platform is PCR based and the amplification step could initiate several negative side effects, such as read density bias in GC-rich areas of the genome.

In this study, we have demonstrated successful fetal T21 detection using free DNA from maternal plasma by single molecule sequencing on the Helicos platform and compared it to the Illumina GA II platform (21–24). For Illumina, we could confirm previously described findings (21). Moreover, we found a more distinct separation between T21 and D21 samples in Helicos data vs Illumina. We showed that as early as 9 + 3 weeks of gestational age, cfDNA samples from maternal plasma can be classified correctly with high sensitivity and specificity. For single molecule sequencing only small amounts of free DNA were required as input for sample preparation and performance of direct sequencing. We hypothesize, therefore, that this method might be more suitable for early noninvasive aneuploidy detection.

Our study also provides confirmation that data obtained on the Helicos platform is not biased in GC-rich areas, thereby leading to increased accuracy of analysis. Previously, a strong correlation between GC-rich areas and read coverage has been observed on the Illumina platform, with increased numbers of reads in areas containing increased GC content (21, 33, 34). There has been discussion as whether this is a biological effect relating to chromatin structure or originates from PCR artifacts introduced during sample preparation, cluster formation, or the sequencing process itself. Because GC bias is not observed in single molecule sequencing it is less likely a true biological effect but more likely a consequence of the sequencing process. We hypothesize that GC bias is introduced in the preamplification step for DNA enrichment or during local amplification for cluster formation on the flow cell. The exact reason, however, remains to be elucidated.

Before the implementation of noninvasive trisomy detection into routine diagnostics several QC criteria must be determined and validated. The Quality Assessment of Diagnostic Accuracy Studies criteria that take into account the experimental bias and variation can be applied (35). Equally important are the QCs before and during sample preparation. Because the percentage of cell-free DNA in maternal plasma differs between samples and at different times of gestation (11, 12), it is difficult to determine the most appropriate time of gestation for testing. However, for diagnostic use inclusion criteria including time of gestation will need to be determined. Measurement of the amount of cfDNA and its relationship to the reliability of diagnosis and the time of gestation must be studied more thoroughly in large validation studies. Before sequencing isolated free DNA, combined real-time PCR results on CCR5 and SRY could help in the estimation of the ratio of maternal and fetal DNA in maternal plasma, the percentage of fetal DNA, and the quality of DNA, as shown in our data. After sequencing, percentages of cfDNA can then be verified by using data from chromosome X as described previously (21). Both methods, however, are limited to male pregnancies only. When samples containing low percentages of fetal sequences or large amounts of contaminating maternal sequences are encountered, restrictions for the detection limit should be taken into account. For female pregnancies a sex- and polymorphism-independent method based on epigenetic differences could be used for quantification (36), although it remains to be established whether differences in methylation are sufficiently stable and comparable between individuals to be used reliably in diagnostics.

Another issue that should be taken into account is maternal copy number variations. These can be of particular interest for the interpretation of trisomy detection by use of NGS data in diagnostics. Predetermination of copy number variations in the maternal genome could be a useful control in diagnostics, because these findings may influence the interpretation of data when looking at the overrepresentation of a specific chromosome, regardless of the NGS platform used.

In summary, this study shows for the first time that single molecule sequencing can be a reliable and easy-to-use alternative for noninvasive T21 detection in diagnostics. By using single molecule sequencing, previously described experimental noise associated with PCR amplification, such as GC bias, can be overcome. This method is promising not only for noninvasive T21 detection, but also may be potentially useful for the detection of other aneuploidies.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.
Authors' Disclosures or Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: None declared.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honorary: None declared.

Research Funding: TECHGENE-HEALTH-F5--2009-223143, EuroGentest2-HEALTH-F4--2010-261469, and NWO.

Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We thank Jenny Verdoes for including pregnant women, Michel van Galen for bioinformatics, Yavuz Ariyurek and Henk Buermans for technical support, and BIOKE (the Netherlands) for NextGENe software assistance.

References