Noninvasive Prenatal Screening at Low Fetal Fraction: Comparing Whole-Genome Sequencing and Single-Nucleotide Polymorphism Methods

Carlo G. Artieri[1], Carrie Haverty[1], Eric A. Evans[1], James D. Goldberg[1], Imran S. Haque[1,2], Yuval Yaron[3*], and Dale Muzzey[1*†]

[1]Counsyl Inc., South San Francisco, CA 94080, USA. [2]Present address: Freenome, South San Francisco, CA USA. [3]Genetic Institute, Tel Aviv Sourasky Medical Center and Sackler Faculty of Medicine Tel Aviv University, Israel.

[†]Corresponding author

***Both authors contributed equally to this study**

CONTACT INFORMATION:

Dale Muzzey
Counsyl Inc.
180 Kimball Way
South San Francisco, 94080, CA, USA
dale@counsyl.com
Tel: (888) 268-6795

RUNNING HEAD: NIPS at low fetal fraction

NUMBER OF FIGURES: 4

NUMBER OF TABLES: 1

NUMBER OF SUPPLEMENTS: 1 Document

NUMBER OF WORDS: 3,403

FUNDING SOURCES: This study was funded by Counsyl Inc.

CONFLICT OF INTERESTS: All authors other than YY are employees of Counsyl Inc., a company that performs noninvasive prenatal screening. YY is a clinical expert panel member for Illumina Inc., a company that performs noninvasive prenatal screening, and a consultant for Teva Pharmaceuticals, a local distributor of noninvasive prenatal screening.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pd.5036

BULLETED STATEMENTS:

What's already known about this topic?

- The two most popular noninvasive prenatal screening (NIPS) methodologies, the single nucleotide polymorphism (SNP) and the whole-genome sequencing (WGS) methods, report comparable performance.
- However, failure rates vary by an order of magnitude between methodologies.
- A large component of failure is insufficient fetal fraction, creating “no-call” test results.
- The American Congress of Obstetricians and Gynecologists (ACOG) indicates that reported sensitivity is often inflated due to exclusion of failed samples from calculations.

What does this study add?

- The SNP method performs poorly at low fetal fraction on the most common fetal trisomy (maternal M1 nondisjunction).
- Offering invasive testing to all cases of "no-call" results will increase the rate of procedure-related loss.
- Compared to the SNP method, the WGS method maintains high specificity and can detect a higher proportion of aneuploidies in low fetal fraction samples without unnecessary invasive tests.

Abstract

Objective Performance of noninvasive prenatal screening (NIPS) methodologies when applied to low fetal fraction samples is not well established. The single-nucleotide polymorphism (SNP) method fails samples below a predetermined fetal fraction threshold, whereas some laboratories employing the whole-genome sequencing (WGS) method report aneuploidy calls for all samples. Here, the performance of the two methods was compared to determine which approach actually detects more fetal aneuploidies.

Methods Computational models were parameterized with up-to-date published data and used to compare the performance of the two methods at calling common fetal trisomies (T21, T18, T13) at low fetal fractions. Furthermore, clinical experience data were reviewed to determine aneuploidy detection rates based on compliance with recent invasive screening recommendations.

Results The SNP method's performance is dependent on the origin of the trisomy, and is lowest for the most common trisomies (maternal M1 nondisjunction). Consequently, the SNP method cannot maintain acceptable performance at fetal fractions below ~3%. In contrast, the WGS method maintains high specificity independent of fetal fraction and has >80% sensitivity for trisomies in low fetal fraction samples.

Conclusion The WGS method will detect more aneuploidies below the fetal fraction threshold at which many labs issue a no-call result, avoiding unnecessary invasive procedures.

Introduction

Noninvasive prenatal screening (NIPS) for fetal aneuploidies using cell-free DNA (cfDNA) has been widely adopted in clinical practice due to its improved accuracy as compared to traditional screening approaches (1). Consequently, both the American Congress of Obstetricians and Gynecologists (ACOG) and the American College of Medical Genetics and Genomics (ACMG) recommend NIPS as a routine screening option (2,3).

Such performance improvements have been enabled by developments in next-generation sequencing (NGS) technologies, which are employed by most clinical NIPS laboratories. NGS involves the generation of millions of short sequences ("reads"), each originating from a specific chromosomal segment, that provide information about both the genotype and relative abundance of the site in the genome (e.g., the X chromosome receives more reads in XX females than in XY males) (4). The two most widely offered NIPS methodologies differ based on which data they use to detect fetal aneuploidies.

The single-nucleotide polymorphism approach ("SNP method"), measures the relative proportion of maternal and fetal genotypes among cfDNA fragments, and tests whether the observed patterns on specific chromosomes are more consistent with disomic or aneuploid fetal expectations (5). Alternatively, the whole-genome sequencing approach ("WGS method"), measures the relative abundance of cfDNA from whole chromosomes in the maternal blood, testing whether certain chromosomes show elevated or reduced numbers of reads, consistent with fetal aneuploidies (6). Despite differences in the underlying signals, meta-analyses have found that both approaches share comparable clinical sensitivities for detecting common aneuploidies: trisomy 21 (Down Syndrome, T21), trisomy 18 (Edwards Syndrome, T18), trisomy 13 (Patau Syndrome, T13), and monosomy X (7,8).

A key concern has been how these methods perform on the minority of patients with low abundance of circulating fetal DNA (i.e., low fetal fraction samples), where the threshold between signal and noise blurs (9). Low fetal fraction is associated with high maternal body-mass index and certain fetal aneuploidies (1,10,11). In order to maintain high per-patient sensitivity, some tests avoid reporting results to patients with fetal fractions below a preset threshold - referred to as a "no-call" result. Patients who receive a "no-call" may submit a second blood draw or be

offered invasive testing as higher rates of aneuploidy have been reported in such samples (12).

However, it is not clear that the strategy of combining “no-calling” with invasive diagnostic follow-up improves the detection rate of fetal aneuploidies as compared to simply calling low fetal fraction samples with reduced sensitivity: clinical experience shows that only 56.5% of patients submit a second sample and at least 25% of these fail again due to low fetal fraction (10-12). Consequently, both ACOG and ACMG recommend that patients receiving a “no-call” be offered invasive diagnostic testing and that a second blood draw is not appropriate (2,3). Yet compliance with this recommendation is far from perfect: even among women who screen positive for aneuploidies using either NIPS or conventional screening methods, only ~55% seek confirmatory invasive testing (10,13). Finally, an additional consideration is that invasive tests cause procedure-related pregnancy loss in approximately one in 500 cases, which could affect a substantial number of patients in the context of routine NIPS (14).

In this study we use up-to-date published validation reports to model the performance of the WGS and SNP methods at low fetal fractions in order to determine whether returning a "no-call" result leads to a higher rate of detection of common aneuploidies as opposed to providing results for all cases regardless of fetal fraction.

Methods

Simulating the WGS method

A detailed description of the equations and procedures underlying both methods is available in the Supplementary Material accompanying this manuscript and all code used to generate and analyze the data is available at https://github.com/counsylresearch/artieri_et_al_nips_at_low_ff. Here we briefly outline the key references and assumptions used to assess their performance. The WGS method was simulated according to the parameters in Jensen et al. (15): 16 million reads per sample counted in 50 kb bins along chromosomes 21, 18, and 13 (we show in Supplemental Figure S1 that results remain unchanged when both methods are simulated at the same sequencing depth). Chromosomal sizes,

(GRCh37/hg19, Feb. 2009) (16) were reduced by 10% to account for exclusion of poorly performing bins (resulting from high-GC content or the presence of repetitive elements) (15). As proper normalization leads to bin counts being distributed according to Poisson expectations (Supplemental Figure S2) (17), bin depths were sampled from a Poisson distribution with mean based on chromosomal ploidy and fetal fraction (see Supplemental Methods).

Sample-specific z-scores for each chromosome were generated by simulating batches of 100 samples (as could be run on a single Illumina v4 High Output sequencing flowcell on a HiSeq 2500 instrument [Illumina, San Diego, CA, USA]), randomly drawn from a population with trisomy prevalence of 3.3%, T21; 1.5%, T18, 0.5%, T13 (8). A single z-score from each of a disomic and trisomic sample (if present) were chosen at random from each batch until 10,000 of each were sampled at a given fetal fraction. The number of permutations was chosen such that repeated iterations of the same analysis would not produce significant differences in conclusions due to stochastic variation during *in silico* sampling.

Samples were called trisomic if their z-score was greater or equal to three. Sensitivity was calculated as the fraction of trisomic samples correctly called, while specificity was the fraction of disomic samples not called trisomic.

Simulating the SNP method

For the SNP method, parameters were obtained from Ryan et al. (12). 13,392 SNP sites were equally divided among chromosomes 21, 18, 13, and X. The mean number of NGS reads-per-SNP, 859, was obtained by dividing 11.5 million, the average sequencing depth of samples with <7% fetal fraction, by the total number of SNPs. As the variance in counts per SNP and allelic balance are not published, we modeled these parameters to generate data consistent with published figures (18) and show that all conclusions drawn are robust to the specific parameter values (Supplemental Figures S3-6). Samples were simulated in accordance with allele distributions as expected based on the parent- and meiotic-stage-of-origin of the trisomy and classified according to the approach outlined in Rabinowitz et al. (19) (see Supplemental Material).

Sensitivity of the SNP method at each fetal fraction was calculated by simulating 10,000 samples for each of the four types of meiotic nondisjunction and determining the proportion of trisomic samples for which the trisomy hypothesis had log-odds ratio (LOR) below the threshold at which 99.87% of disomic samples would be called disomic (corresponding to the same specificity as the WGS test). The aggregate sensitivity of the SNP method was determined by calculating the weighted expectation of the sensitivity of detection of each fetal ploidy hypothesis multiplied by its prevalence (20).

Determining clinical outcomes

The proportion of patients who submit redraws, 56.5%, was obtained from Dar et al. (10) and is consistent with the value of 55% reported by Yared et al. (11). We conservatively used the highest reported probability of obtaining a successful result upon redraw for samples below 3% fetal fraction: 74% (12). As it is unlikely that "no-calls" lead to termination of pregnancy, we used the proportion of remaining patients receiving a positive NIPS call in Dar et al. (10) who elected invasive testing, 55%, to estimate the maximal rate at which patients receiving a "no-call" would seek invasive testing. This value agrees with the value of 57% reported for conventional first trimester screening (13). The rate of procedure-related loss, 0.002, was obtained from Yaron (21).

To calculate the detection sensitivity for the WGS method for samples < 2.8% fetal fraction (12), we determined the frequency of observing samples in bins of 0.1% fetal fraction from 0 to 2.7% by fitting a beta distribution to the parameters reported in Nicolaides et al. (22) (median: 0.1; 25/75 percentiles: 0.078 and 0.13) using the 'optimize.brute' function in SciPy beginning with priors of zero and minimizing the sum of squares deviation from the 25th and 75th percentiles (23). We then obtained the dot product of the sensitivity calculated at each fetal fraction bin multiplied by its relative prevalence among samples in the 0 to 2.7% range.

Statistics and Data

All statistics were calculated using SciPy in Python (version 0.17) (23). The output of all analyses as well as the Python code required to reproduce all results is available under the Creative Commons Attribution-NonCommercial 4.0 International Public License at https://github.com/counsylresearch/artieri_et_al_nips_at_low_ff.

Results

Comparing NIPS methods

To compare the WGS and SNP methods, we developed a computational framework with two steps: (1) simulation models that mimic the raw data generated by each method, and (2) aneuploidy-calling algorithms that process the simulated data and yield ploidy calls for trisomies in chromosomes 21, 18, and 13 (see Methods). The simulations allowed us to model an arbitrary number of pregnancies over a precise range of fetal fractions and calculate analytical performance in terms of both sensitivity and specificity. To ensure a fair comparison between methods, simulation parameters and calling algorithms were drawn directly from the most up-to-date published reports, such as peer-reviewed manuscripts and patents (see Methods).

The WGS method partitions each chromosome into equal sized bins and tallies the number of reads per bin, normalized for GC content and repetitive regions (15). Trisomies manifest as a higher number of reads per bin relative to the disomic background, with read excess proportional to the fetal fraction (e.g., 55 reads in trisomic chr21 bin vs 50 reads in disomic chr3, Figure 1, top). In contrast, the SNP method measures relative counts among alleles at pre-selected, polymorphic sites on the chromosomes of interest. Trisomies manifest as a global shift in allelic counts relative to disomic chromosomes (Figure 1, bottom).

Influence of parent-of-origin on SNP and WGS methods

The expected deviation in allele frequencies caused by aneuploidies in the SNP method differs based on the parental- and meiotic-stage of origin of the aneuploidy (Figure 2; Supplemental Figure S7), the rates of which vary substantially: 70% of nondisjunction events occur during maternal meiosis phase I (M1), leading to

fetal inheritance of both maternal chromosomes, while 20% occur during meiosis phase II (M2), causing fetal inheritance of two copies of a single maternal chromosome (24). The remaining nondisjunctions are paternal in origin, with 3% originating from M1, and 7% from M2 (20) (Figure 2A).

In the SNP method, paternally-derived trisomies produce a much stronger signal than do the more common maternal trisomies, illustrated by the shift in the red and blue distributions in Figure 2B. Because the majority of allelic counts in cfDNA are maternal in origin, paternally inherited trisomies nearly double the presence of paternal-specific alleles, whereas maternally-derived trisomies only slightly increase the signal of one of the two maternal alleles (~4%) (Table 1).

Conversely, in the WGS method, all four origins of trisomy produce the same signal: an elevated number of NGS reads mapping to the trisomic chromosome (Figure 2C). Consequently, the sensitivity of the method is independent of the origin.

Specificity and sensitivity of both methods as a function of fetal fraction

We compared the performance of the two methods at low fetal fraction by simulating 10,000 samples of each possible origin of trisomy for chromosomes 21, 18, and 13 at fetal fractions ranging from 0.1% to 4% (see Methods). Per-sample sequencing depths were obtained from published validation reports (12,15).

We determined calling performance by generating receiver-operating characteristic (ROC) curves and calculating the area under the curve (AUC) for each method as a function of fetal fraction (Figure 3A, B) (25). A test with high sensitivity and specificity will have AUC near 1. The SNP method calculates a likelihood that each chromosome is disomic or trisomic on a per-sample basis, returning a log-odds ratio (LOR) where disomic samples have positive LOR and trisomic samples have negative LOR. The SNP algorithm shows poor differentiation between disomic and trisomic LORs at low fetal fractions as an AUC of 0.99 is not achieved below fetal fractions of 3% (Figure 3A). Setting a fetal fraction threshold below which all samples are “no-called” is thus appropriate to maintain SNP-method performance (12).

In contrast, the WGS method processes ~100 samples per batch and, for each sample and chromosome, assigns a z-score indicating the extent of which the counts in this sample deviate from the distribution of all other samples in the batch (26). Because the expected distribution of disomic samples is independent of fetal

fraction, its specificity is a function of the z-score threshold (z ≥ 3) (9). This is confirmed by the AUCs, which are greater than 0.99 at fetal fractions above 1.5% for all three common trisomies (Figure 3B).

To translate AUCs into sensitivities, we calculated the maximum attainable sensitivity of the SNP method at each fetal fraction while maintaining specificity ≥ 99.87% (z ≥ 3 in the WGS method, see Methods). As expected, rare paternal trisomies could, in principle, be confidently detected at extremely low fetal fractions (Figure 3C). However, the most common fetal trisomies, maternal M1 nondisjunctions (70%), show the lowest sensitivity, with maternal M2 nondisjunctions (20%) falling in between the two extremes. We obtained the aggregate sensitivity of the SNP method at each fetal fraction by taking a prevalence-weighted sum of the sensitivities for each trisomic origin.

The sensitivity of the WGS method is dependent on the number of bins used to count reads and increases with chromosome size (Figure 1A) (17). Therefore, we calculated the sensitivity of the WGS method for each of the three common trisomies separately. In all cases, the SNP method shows lower aggregate sensitivity than the WGS method at low fetal fractions (Figure 3D).

Clinical outcomes of low fetal fraction samples

Patients receiving a "no-call" have the option of sample redraw or invasive testing; however, not performing a follow-up test may lead to undiagnosed aneuploidies. Therefore, we assessed the T21 detection rate under two scenarios: 1) "no-calling" all samples below a fetal fraction of 2.8% (the most-recent no-call threshold reported by clinical laboratories using the SNP method (12)), and 2) calling all such samples using the sensitivity parameters established from simulations of the WGS method. We first calculated the sensitivity of the WGS method for all samples below 2.8% fetal fraction by summing the prevalence-weighted sensitivity at each

fetal fraction (Supplemental Figure S8) (22). Applied to the simulated data, the WGS method shows a sensitivity of 86% for samples with fetal fraction < 2.8%.

Among patients initially receiving a “no-call” by the SNP method, approximately 42% will submit a redraw and receive a result (Figure 4A) (10-12). Subsequently, the rate of invasive procedures among remaining patients will determine the proportion of aneuploidies that are detected (we assume that redraws and invasive testing are 100% sensitive to establish an upper limit on the detection rate). The invasive test rate among patients receiving a “no-call” would have to exceed 76% to equal the sensitivity of the WGS method, which is unlikely given the upper-limit estimate of 55% (Figure 4B; see methods).

To illustrate the clinical consequences of systematically “no-calling” low fetal fraction samples, we assessed T21 screening outcomes for both methods on a simulated cohort of 10,000 initially low fetal fraction samples, assuming a trisomy 21 incidence of 3.3% (8), and a rate of invasive testing of 55% after a “no-call” result (Figure 4C; see Methods). The WGS method would successfully detect 285 of the expected 330 cases (86%) but result in 45 false negatives. In comparison, the SNP method would detect 138 cases upon redraw, while 3,200 invasive procedures would be required to detect an additional 106 cases (74% detected), totaling 86 false negatives and six procedure-related pregnancy losses.

Discussion

The vast majority of samples submitted for NIPS have sufficient fetal fraction to enable screening by both methods with excellent sensitivity, likely explaining why meta-analyses have noted comparable test performance (2,3,7,8). Nevertheless, we demonstrate that there are substantial performance differences between the two methodologies in low fetal fraction samples, leading to important clinical consequences.

Modeling NIPS at low fetal fraction

We implemented a modeling approach to analyze the performance of the two NIPS methods due to a paucity of clinical data in samples with low fetal fraction. The key parameters of the WGS model - the distribution of counts along chromosomes as well as total reads per sample - have been well characterized, indicating that the results of our simulated model likely reflect biological reality for most patients (17). A limitation of the SNP model is that two key parameters - the number of reads per SNP and the accuracy with which allele fractions are estimated - were inferred from published reports (see Supplemental Methods, Supplemental Figure S3) (10,18,27). Importantly, we show that even when these parameters are set to biologically unrealistic, theoretical ideals, the conclusions of the analysis remain unaltered (Supplemental Material, Supplemental Figures S4-6).

Performance of NIPS methods

By comparing each chromosome to a baseline disomic distribution, the WGS method maintains high specificity (≥ 99.87%), independent of fetal fraction (Figure 3A) (9). In contrast, by evaluating the likelihood of aneuploidy on a sample-by-sample basis without a baseline expectation, the SNP method is unable to clearly distinguish disomic from trisomic samples at low fetal fractions. This justifies establishment of a threshold below which all samples are "no-called" (Figure 3B). In fact, the threshold of 2.8% recently reported by Ryan et al. (12) agrees well with the value below which our simulations show a substantial drop in sensitivity (i.e., 3%, Figure 3C,D).

A key determinant of the sensitivity of the SNP method in detecting trisomies is the origin of the nondisjunction event (Figure 2). The low sensitivity of detection of the most common type of nondisjunction (maternal M1, 70% of cases) overwhelms the improved sensitivity of the remaining types of nondisjunction, leading to overall poor test performance at low fetal fractions. Crucially, this performance deficit is not limited to the SNP method algorithm, but rather by the underlying biology of nondisjunction (Figure 2B, Table 1). Indeed, any method measuring shifts in allele balances within the maternal-fetal mixture of cfDNA will have relatively reduced sensitivity in samples with the most common type of nondisjunction.

We also note that the results presented here model analytical sensitivities and specificities. Reported clinical sensitivities of NIPS have been higher for T21 when compared to T18 and T13 (7,8). This is potentially due to T21 samples showing slightly elevated fetal fractions relative to disomic pregnancies, while those of T18 - and in some cases T13 – have reportedly lower fetal fractions (28,29). In addition, fetal fraction is negatively correlated with maternal weight, further illustrating the importance of maximizing detection performance in low fetal fraction samples (9,11). Other factors that reduce clinical sensitivity, such as sample contamination and confined placental mosaicism (29), should impact all NIPS methods as currently practiced (7).

Clinical consequences of no-calling low fetal fraction samples

Our model demonstrates that while setting a minimum fetal fraction threshold may help to maintain high per-patient analytical sensitivity (e.g., 1,11,31,32), it may ultimately prove to be counter-productive. Based upon published clinical data, the probability of detecting a trisomy 21 case after an initial "no-call" using the SNP method is only ~74% (Figure 4B). This is almost certainly an overestimate as it is unlikely that patients who receive a "no-call" due to low fetal fraction will seek confirmatory invasive testing at the same rate as those who screen positive for T21 (i.e., 10,13). In contrast, the WGS model identifies a larger fraction of T21 fetuses, all noninvasively, obviating the need for invasive testing on failed samples (Figure 4C).

A major advantage of cfDNA-based NIPS over previous screening modalities is a tenfold reduction in the false-positive rate, which has vastly reduced the number of needless invasive procedures (1,21,33). However, this benefit is undermined by tests with high failure rates, most commonly due to low fetal fraction. For example, nearly three-quarters of the projected overall test failure rate of the SNP method reported in Ryan et al. (12) was due to insufficient fetal fraction (3.8% low fetal fraction; 5.2% total failure rate).

It is also notable that test failures are routinely excluded from claimed test sensitivity in published validation reports (2,21). This is especially important given that multiple studies have noted an increased rate of aneuploidies among patients who receive a "no-call" (1,11,34). For instance, the 100% sensitivity of T21 screening reported by Pergament et al. (34) drops to 86.5% if T21 positive fetuses among "no-

calls" are counted among false negatives (21). Barring an increase in the rate of invasive procedures - and the concomitant iatrogenic pregnancy loss that this entails - the findings of this study suggest that the most effective approach to improving fetal screening efforts is to implement methods that improve the performance of NIPS methods at low fetal fractions.

Conclusion

Via empirically informed simulation, we show that unlike the WGS approach, the SNP method cannot maintain high specificity and sensitivity at low fetal fractions, justifying its reliance on a minimal fetal fraction threshold for calling (12). Finally, using published clinical data, we find that the WGS method detects a higher proportion of common aneuploidies in low fetal fraction samples than setting a "no-call" threshold, avoiding large numbers of invasive tests and associated complications.

References

1. Norton ME, Jacobsson B, Swamy GK, et al. Cell-free DNA analysis for noninvasive examination of trisomy. N Engl J Med. 2015 Apr 23;372(17):1589–97.
2. ACOG Committee on Genetics. Committee Opinion No. 640: Cell-free DNA Screening for Fetal Aneuploidy. Obstet Gynecol. 2015 Jun 29;163:e31–7.
3. Gregg AR, Skotko BG, Benkendorf JL, et al. Noninvasive prenatal screening for fetal aneuploidy, 2016 update: a position statement of the American College of Medical Genetics and Genomics. Genet Med. 2016 Jul 28;18(10):1056-65.
4. Goodwin S, McPherson JD, Richard McCombie W. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016 May 17;17(6):333–51.
5. Zimmermann B, Hill M, Gemelos G, et al. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. Prenat Diagn. 2012 Dec;32(13):1233–41.
6. Fan HC, Blumenfeld YJ, Chitkara U, et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A. 2008 Oct 21;105(42):16266–71.
7. Gil MM, Quezada MS, Revello R, et al. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. Ultrasound

Obstet Gynecol. 2015 Mar;45(3):249–66.

8. Taylor-Phillips S, Freeman K, Geppert J, et al. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. BMJ Open. 2016 Jan 18;6(1):e010002.

9. Canick JA, Palomaki GE, Kloza EM, et al. The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies. Prenat Diagn. 2013 Jul;33(7):667–74.

10. Dar P 'er, Curnow KJ, Gross SJ, et al. Clinical experience and follow-up with large scale single-nucleotide polymorphism-based noninvasive prenatal aneuploidy testing. Am J Obstet Gynecol. 2014 Nov;211(5):527.e1–527.e17.

11. Yared E, Dinsmoor MJ, Endres LK, et al. Obesity increases the risk of failure of noninvasive prenatal screening regardless of gestational age. Am J Obstet Gynecol. 2016 Sep;215(3):370.e1-6.

12. Ryan A, Hunkapiller N, Banjevic M, et al. Validation of an Enhanced Version of a Single-Nucleotide Polymorphism-Based Noninvasive Prenatal Test for Detection of Fetal Aneuploidies. Fetal Diagn Ther. 2016 Mar 31;40(3):219-223.

13. Currier R, Wu N, Van Meter K, et al. Integrated and first trimester prenatal screening in California: program implementation and patient choice for follow-up services. Prenat Diagn. 2012 Nov;32(11):1077–83.

14. Akolekar R, Beta J, Picciarelli G. Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2015 Jan;45(1):16-26.

15. Jensen TJ, Zwiefelhofer T, Tim RC, et al. High-throughput massively parallel sequencing for fetal aneuploidy detection from maternal plasma. PLoS One. 2013 Mar 6;8(3):e57381.

16. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860–921.

17. Fan HC, Quake SR. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. PLoS One. 2010 May 3;5(5):e10439.

18. Hall MP, Hill M, Zimmermann B, et al. Non-invasive prenatal detection of trisomy 13 using a single nucleotide polymorphism- and informatics-based approach. PLoS One. 2014 May 7;9(5):e96677.

19. Rabinowitz M, Gemelos G, Banjevic M, et al. Methods for non-invasive prenatal ploidy calling. US Patent. 14179399, 2014.

20. Oliver TR, Bhise A, Feingold E, et al. Investigation of factors associated with paternal nondisjunction of chromosome 21. Am J Med Genet A. 2009 Aug;149A(8):1685–90.

21. Yaron Y. The implications of non-invasive prenatal testing failures: a review of an under-discussed phenomenon. Prenat Diagn. 2016 May;36(5):391–6.

22. Nicolaides KH, Syngelaki A, Ashoor G, et al. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. Am J Obstet Gynecol. 2012 Nov;207(5):374.e1–6.

23. Jones E, Oliphant T, Peterson P. SciPy: Open source scientific tools for Python. SciPy.org. 2001. Available from: http://www.scipy.org/

24. Antonarakis SE, Petersen MB, McInnis MG, et al. The meiotic stage of nondisjunction in trisomy 21: determination by using DNA polymorphisms. Am J Hum Genet. 1992 Mar;50(3):544–50.

25. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev. 2008 Aug;29 Suppl 1:S83–7.

26. Sokal RR, Rohlf FJ. Biometry: The Principles and Practice of Statistics in Biological Research, 3rd edn Freeman. New York, NY. 1995;

27. Curnow KJ, Wilkins-Haug L, Ryan A, et al. Detection of triploid, molar, and vanishing twin pregnancies by a single-nucleotide polymorphism-based noninvasive prenatal test. Am J Obstet Gynecol. 2015 Jan;212(1):79.e1–9.

28. Rava RP, Srinivasan A, Sehnert AJ, Bianchi DW. Circulating fetal cell-free DNA fractions differ in autosomal aneuploidies and monosomy X. Clin Chem. 2014 Jan;60(1):243–50.

29. Palomaki GE, Kloza EM, Lambert-Messerlian GM, et al. Circulating cell free DNA testing: are some test failures informative? Prenat Diagn. 2015 Mar;35(3):289–93.

30. Grati FR, Malvestiti F, Ferreira JCPB, et al. Fetoplacental mosaicism: potential implications for false-positive and false-negative noninvasive prenatal screening results. Genet Med. 2014 Aug;16(8):620–4.

31. Sparks AB, Struble CA, Wang ET, et al. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. Am J Obstet Gynecol. 2012 Apr;206(4):319.e1–9.

32. Lefkowitz RB, Tynan JA, Liu T, et al. Clinical validation of a noninvasive prenatal test for genomewide detection of fetal copy number variants. Am J Obstet Gynecol. 2016 Aug;215(2):227.e1–227.e16.

33. Bianchi DW, Parker RL, Wentworth J, et al. DNA sequencing versus standard prenatal aneuploidy screening. N Engl J Med. 2014 Feb 27;370(9):799–808.

34. Pergament E, Cuckle H, Zimmermann B, et al. Single-nucleotide polymorphism-based noninvasive prenatal screening in a high-risk and low-risk cohort. Obstet Gynecol. 2014 Aug;124(2 Pt 1):210–8.

35. Karp G. Cell and Molecular Biology: Concepts and Experiments. John Wiley & Sons; 2009. 832 p.

36. Hand DJ. Evaluating diagnostic tests: The area under the ROC curve and the balance of errors. Stat Med. 2010 Jun 30;29(14):1502–10.

Tables

Table 1. Rare paternally inherited trisomies produce a stronger SNP count signal than do more common maternally inherited trisomies. Values are shown for a 10% fetal fraction, a maternal genotype, AA, and a fetal genotype, AB. Paternal trisomies will increase the frequency of the paternally inherited B allele almost two-fold over euploid expectations, while the increase of the frequency of the A allele only shifts the expected abundance of the B allele by 0.2%.

	Disomy		Maternally-derived trisomy		Paternally-derived trisomy	
	Counts	%B	Counts	%B	Counts	%B
Maternal Genotype	AA × 9	5%	AA × 9	4.8%	AA × 9	9.5%
Fetal Genotype	AB × 1		AAB × 1		ABB × 1	

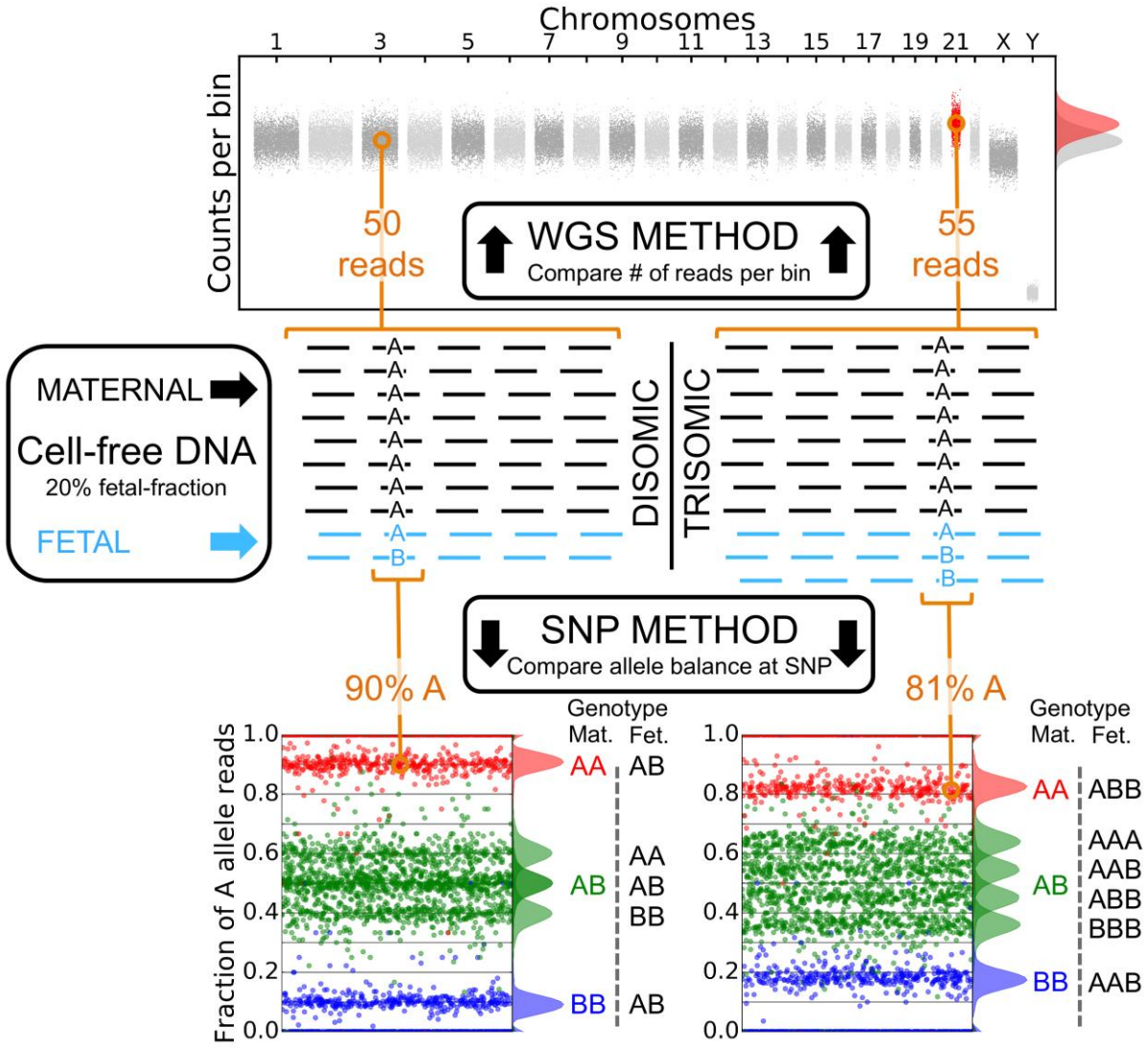

Figure 1. Overview of the WGS and SNP methods. Both methods take advantage of NGS reads originating from the mixture of maternal and fetal cfDNA (in this example, fetal cfDNA constitutes 20% of the pool). The WGS method (top) divides the genome into equally-sized bins and counts the number of reads mapped to each bin. As illustrated, the presence of a fetus affected with trisomy 21 leads to an increase in the distribution of counts-per-bin originating from chromosome 21 (55 reads, red) relative to the euploid background (50 reads, grey). Detection of this increase forms the basis of the test. Alternatively, the SNP method (bottom) measures the relative abundance of alleles at polymorphic sites in the cfDNA. Fetal aneuploidies lead to predictable shifts in the frequency of allelic counts based on the possible combinations of maternal and fetal genotypes (see below). By aggregating the signal across many SNPs on a given chromosome, the algorithm calculates a likelihood that the overall pattern of allele frequencies is more consistent with a

normal versus an aneuploid fetus. The strength of the aneuploid signal in either method is proportional to the fetal fraction, and so too is the sensitivity of detection. Note that the number of reads illustrated in the figure are substantially lower than those published in validation reports (see Methods).

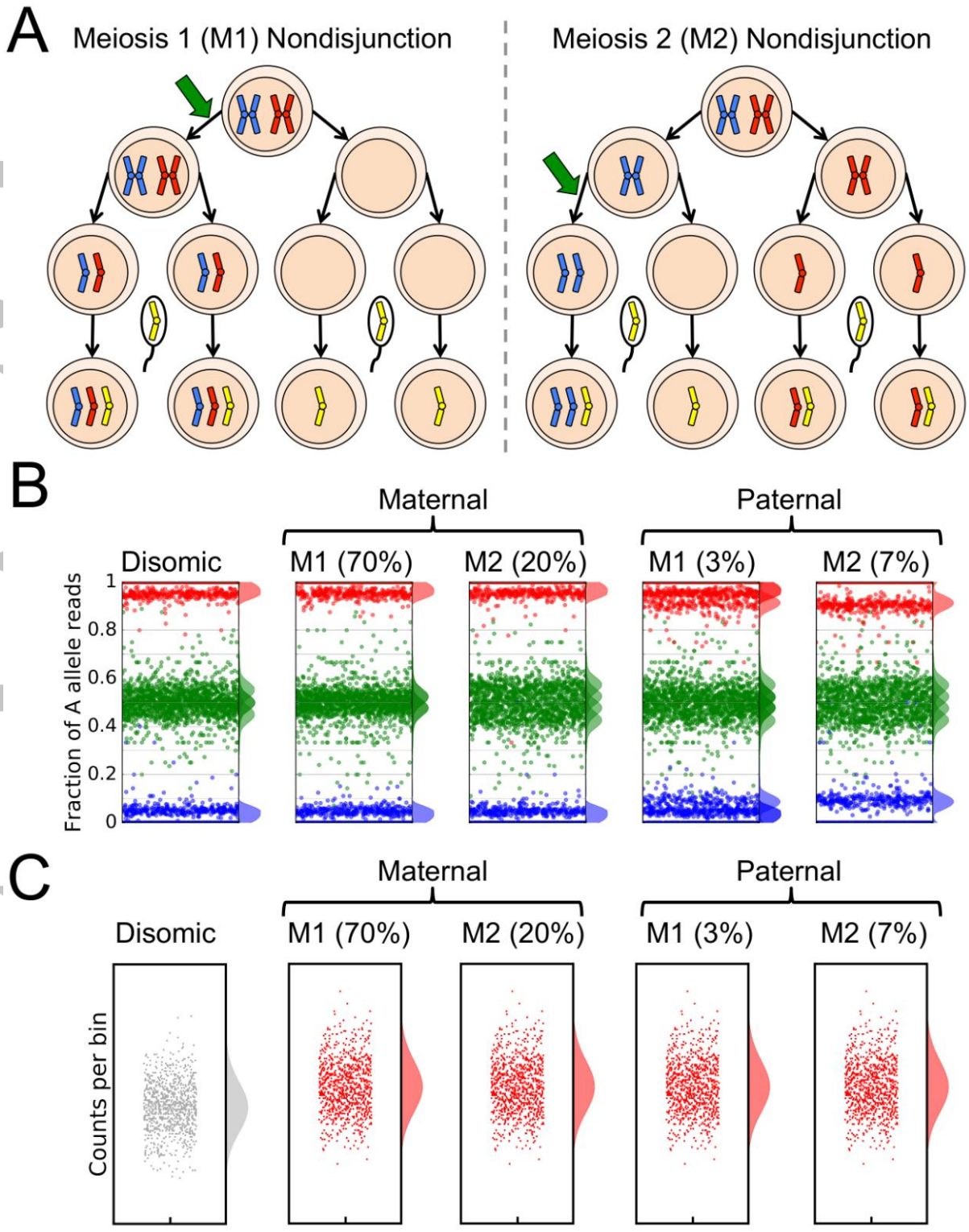

Figure 2. The parent-of-origin and meiotic-stage-of-origin of the aneuploidy produce different expected distributions of allele-frequencies with the SNP method. A) Illustration of the two origins of maternal trisomies. Trisomies can originate either from meiosis stage 1 (M1; left) or stage 2 (M2; right) nondisjunction. In M1 nondisjunctions, the oocyte inherits one copy of each of the different maternal

chromosomes (marked in blue and red). Upon fertilization, the paternal chromosome (yellow) is added. In such cases, SNP-based analysis for each locus will yield one of six possibilities, resulting from the admixture of maternal and fetal cfDNA (AA|A, AA|B, AB|A, AB|B, BB|A, and BB|B, where the maternally inherited genotype is to the left and the paternal is to the right). In M2 nondisjunctions, the oocyte inherits two copies of the same maternal chromosome, leading to only four possibilities (AA|A, AA|B, BB|A, and BB|B). Cases of paternal nondisjunctions follow the same lines of reasoning with the parental origins reversed. Adapted from Karp (35). B) The four possible parental origins of trisomies produce distinct signals with the SNP method. Simulated allele-frequency distributions for samples with 10% fetal fraction for each of the different origins of trisomies are shown along with their relative frequencies. The density plots to the right of each panel indicate the shape of the distributions (each corresponding to the admixture of maternal and fetal cfDNA). Note that the shifts in heterozygous fetal SNPs on a homozygous maternal background (blue and red dots above) are significantly more pronounced in the rare paternally-derived aneuploidies in comparison to the much more common maternally-derived aneuploidies when compared to a euploid sample. The allelic basis of all possible maternal-fetal cfDNA genotypic combinations is shown in Supplemental Fig S7. C) With the WGS method, all four trisomic origins lead to the same increase in the number of reads per bin for the trisomic chromosome.

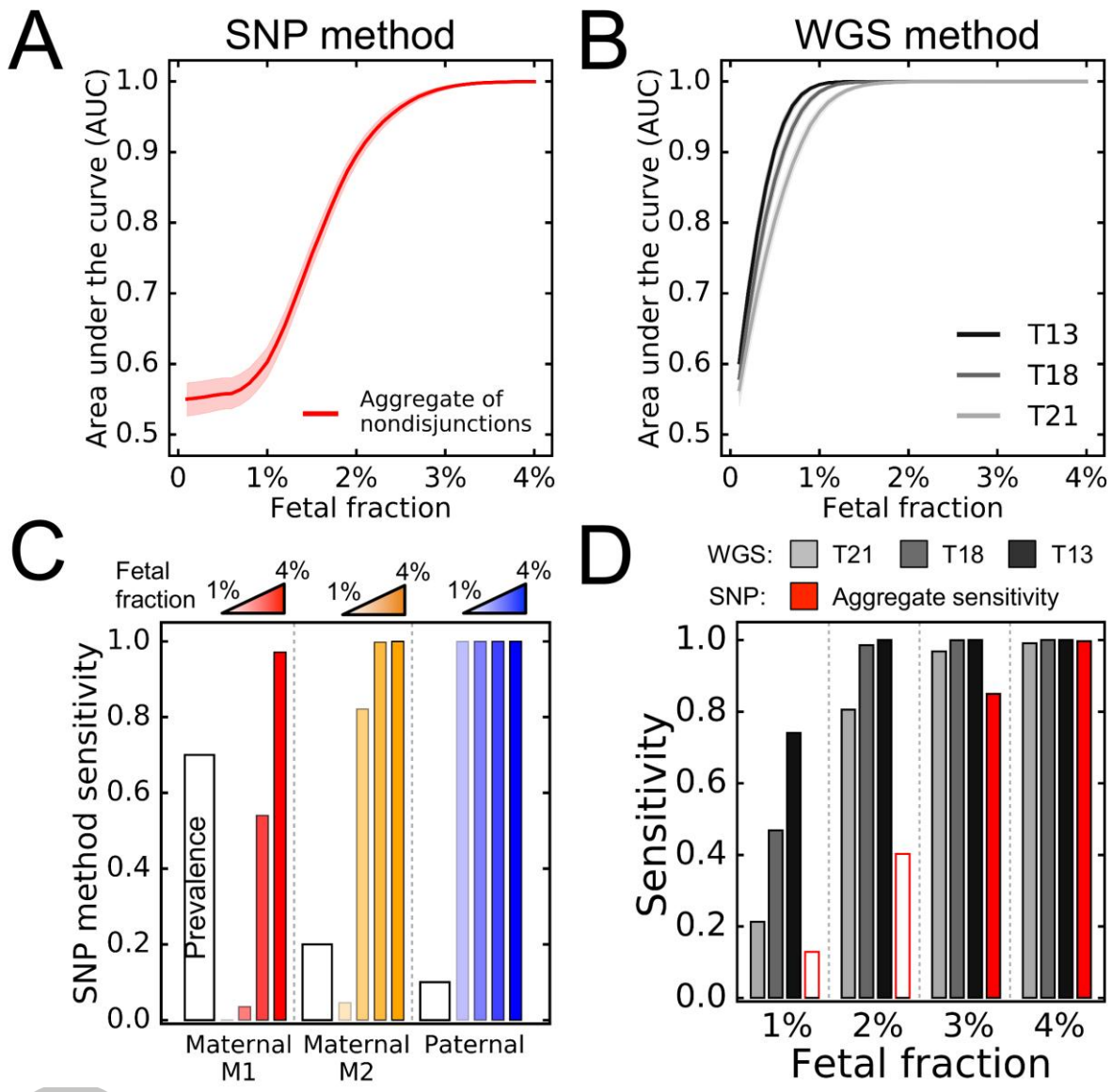

Figure 3. Comparison of performance characteristics of the two methods at low fetal fractions. Area-under the curve (AUC) values as a function of fetal fraction for the SNP (A) and WGS method (B) calling algorithms. The AUC captures the overall performance of disease classification and represents the probability that a random trisomic fetal sample would have a higher likelihood of aneuploidy in the SNP method (or a higher z-score than a random disomic sample in the WGS method) (25,36). The WGS method achieves AUC ≥ 95% for all three common trisomies at ≥ 1% fetal fraction. In contrast, the SNP method only achieves AUC ≥ 95% at fetal fractions ≥ 2.4%. The shaded areas above and below the curves represent the 95% confidence intervals on the estimates of AUC. C) Sensitivity of the

SNP method at 99% specificity with respect to each of the trisomic origins (as both paternal non-disjunctions have identical sensitivity in this range, they are combined into a single category). The prevalence of each of the origins is shown as the white bar, while the colored bars indicate increasing fetal fraction in 1% intervals. Note that the sensitivity of the maternal M1 trisomy at 1% fetal fraction is zero, thus the leftmost bar is missing. Importantly, sensitivity is lowest for the most prevalent trisomies (maternal M1, 70%). D) Comparison of the sensitivities of the WGS method for each of the three common autosomal trisomies (grey bars) to the aggregate sensitivity of the SNP method (solid red bars). The aggregate sensitivity of the SNP method was obtained by summing the sensitivities scaled by prevalence across the three categories of trisomic origin. As illustrated by the simulations, the sensitivity of the WGS method improves with increasing chromosomal size. The aggregate sensitivity of the SNP method is shown as hollow bars at 1% and 2% fetal fraction as these are below the threshold at which all samples are “no-called”. In panels C and D, groups of bars are separated by dotted lines as a visual aid.

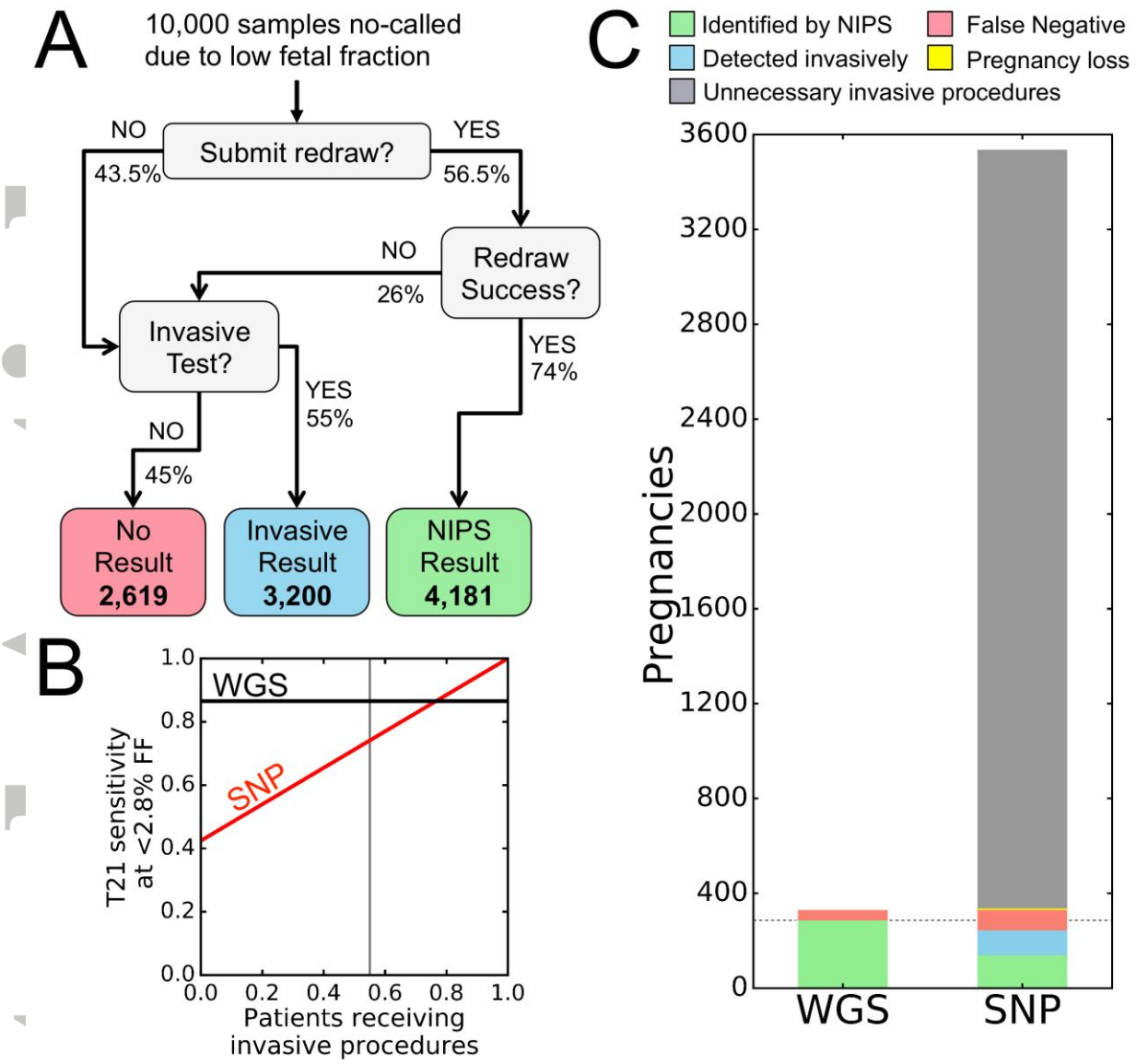

Figure 4. Consequences of "no-calling" samples at low fetal fraction vs. reporting at reduced sensitivity. A) Clinical decision flowchart for samples receiving a "no-call" from the SNP method. Frequencies associated with each decision branch-point were obtained from published literature (see Methods). The outcomes of 10,000 samples that receive a "no-call" result are shown in the colored boxes. B) By summing the T21 sensitivity over the frequency of fetal fractions in the range of under which the SNP method reports a "no-call", the aggregate sensitivity of T21 detection of the WGS method is a constant 86% (black line). In contrast the SNP method (red line) is expected to detect 41% of aneuploid cases by re-draw (the intercept), while all further cases must be detected by invasive procedures. The grey line indicates a maximal estimate of the rate of patients consenting to invasive procedure given a "no-call" result (55%), which results in a total detection rate of

74%. FF, fetal fraction. C) Clinical consequences for 10,000 patients with fetal fraction below the 2.8% “no-call” threshold (12). While the WGS method would detect 285 out of 330 expected cases of T21 (86%), the SNP method would detect 138 by NIPS (all due to redraw). In addition, 3,200 invasive procedures would be required to detect an additional 106 cases for a net sensitivity of 74%. This would also result in six procedure-related pregnancy loss (yellow line). The dotted line indicates the number of cases of T21 detected by the WGS method for comparison.