Non-invasive prenatal testing or NIPT is currently among the top researched topic in obstetric care. While the performance of the current state-of-the-art NIPT solutions achieve high sensitivity and specificity, they still struggle with a considerable number of samples that cannot be concluded with certainty. Such uninformative results are often subject to repeated blood sampling and re-analysis, usually after two weeks, and this period may cause a stress to the future mothers as well as increase the overall cost of the test. The great source of such uninformative samples is in the nature of the statistical testing. Other results are presented by maternal DNA aberrations such as maternal mosaicism, unidentified maternal tumors, or copy number variations.

There is a growing body of studies addressing this issue. A particularly interesting venue of research focuses on qualitative differences between fetal and maternal cfDNA fragments, namely their lengths [2]. Extending this work, we present a method to further boost the elimination of uninformative results. Even if the proposed method is not as accurate as the standard z-score, our results suggest that combination of these two independent methods correctly resolves a substantial portion of healthy samples with an uninformative result. Additionally, our method can be also used to detect false positive or negative results; however, we do not yet have samples to verify this.

Reference z-scores $Z_{NCV}$ and $Z_{SZ}$

The reference z-scores of samples were calculated as normalized chromosome values (NCV) according to [1]. Samples scoring 4 and higher were considered trisomic, while samples scoring 2.5 or lower were considered euploid. The range (2.5, 4) was considered uninformative. Alternative z-scores $Z_{SZ}$ were calculated in the same manner but using only reads of length at most 150bp.

Length-based z-scores $Z_{FL}$

For each sample and chromosome, we defined a series of z-scores as

$$Z_{FL} = \frac{\chi^2(FL)}{\chi^2(1)}$$

where $\chi^2(FL)$ is the number of reads mapped to this chromosome of any size and of size at most $l$, and $\chi^2(1)$ is the sample’s number of reads of any size and of size at most $l$. Furthermore, we defined for each sample and chromosome a FL-score as

$$FL = \max_{\lambda_i} \chi^2(FL) - \lambda_i(1).$$

Finally, the FL-scores were normalized into z-scores which approximately follow standard normal distribution for euploid samples. Then, for any sample, its normalized FL-score value $Z_{FL}$ was used as an alternative method for the prediction of aneuploidy.

Combining the scores $Z_{NCV}$ and $Z_{FL}$ were found to have no statistically significant correlation for euploid samples, and a chi-squared distribution with 2 degrees of freedom was used for derivation of the combined score $Z_{NCV} + Z_{FL}$

$$Z_{NCV} + Z_{FL} \sim \chi^2(2).$$

$Z_{SZ}$ and $Z_{FL}$ were found to be correlated (Pearson R = 0.55, p < 0.001), and we used confidence ellipses to transform them to independent standard normals which were combined into $Z_{SZ + FL}$ as above.

$$Z_{SZ + FL} = \frac{Z_{SZ} + Z_{FL}}{\sqrt{\lambda_1 + 1}},$$

where $\lambda_1$ is the largest eigenvalue of the covariance matrix as

$$\alpha = \arctan \left( \frac{\sqrt{\lambda_1} Z_{FL}}{\sqrt{\lambda_2} Z_{SZ}} \right).$$

Improved evaluation method for reducing uninform. calls

Figure 2: Comparison of z-score calculation methods on chromosome 21. Reference method based on chromosomal fragment counts (NCV) performs better than proposed FL method, albeit their combination (NCV + FL) increased accurate results of trisomic samples (black marks). Size-selection (SZ) further improves z-scores of trisomic samples. The combination with FL method (SZ + FL) most markedly reduces z-scores of samples evaluated as uninformative using traditional NCV score alone (white marks). Empty shapes represent results of the two analyses of an IVF sample.

References
