Development of an evidence-based algorithm that optimizes sensitivity and specificity in WES-based diagnostics of a clinically heterogeneous patient population

Peter Bauer1, Maximilian E.R. Weiss, Omid Paknia1, Martin Werber1, Aida M. Bertoli-Avellà1, Zafar Yüksel2, Krishna Kumar Kandaswamy1, Malgorzata Bochinska1, Gabriela E. Opera1, Shivendra Kishore1, Volkmar Weckesser1, Ellen Karges1, Arndt Rolfs1,2

Sensitivity of whole exome sequencing (WES) is not well-defined. We applied very low thresholds in WES-associated variant calling to also enable investigation of candidate variants that are commonly neglected. As Sanger sequencing revealed ~5% of these to be true positives (Figure 1), we considered numerous variant-specific features (Tables 1 and 2) for the development of a robust predictor for true and false positives. Iterative rounds of receiver operating characteristic (ROC) curve generation identified features and corresponding thresholds with high predictive value (Figure 2). In a corresponding workflow for our data, 91.3% of variants can be pre-classified with 100% specificity and 99.8% sensitivity, while the remaining 8.7% of variants require confirmatory Sanger sequencing (Figure 3).

1-3 ROC analyses indicate parameters with highest predictive value for true positives (round 1) and true negatives (rounds 2 and 3).

Figure 1: Impact of filtering parameters used for variant calling.

Table 1: Candidate analogous features.

Table 2: Candidate digital features.

References
1. in press as Bauer et al. in Genet Med
3. Mu et al., 2016, J Mol Diagn 18:923-32

Disclosure of conflict of interest
This study was sustained in part by Centogene AG, Rostock. All authors of this presentation are employees of CENTOGENE AG, Rostock, Germany.