Titel:

Two years of whole exome sequencing analyses: A bioinformatician’s memoirs

Abstract:

Following the widespread adoption of high-throughput sequencing, whole exome sequencing (WES) has proven to be highly effective in clinical and research settings. Within the scope of DFG-funded clinical research unit 326 (“Male Germ Cells”) more than 1000 exomes have now been sequenced to decipher the genetics of male infertility. While many best-practice recommendations exist for detection of single nucleotide variants (SNVs) and short indels, linking these variants to specific phenotypes is still a challenging task. Subsequently, population sampling probability (PSAP) scores were computed to aid variant prioritisation. In brief, PSAP scores enable estimation of variant pathogenicity by comparing pathogenicity scores in the general population to those in patient exomes. Multiple variants were successfully associated with male infertility using PSAP scores (e.g. STAG3 & M1AP variants). Although many disease-causing SNVs and indels could be detected, larger variants, such as copy number variants (CNVs), are currently mostly assessed by microarray techniques. In theory, such variants can also be inferred by analysing read depth from WES experiments: First experimental screenings of zero-coverage target regions successfully detected known X-linked TEX11 deletions. However, generalized detection of CNVs including heterozygous deletions and duplications is hindered by many sources of noise (e.g. GC bias and batch effects). Thus, CNV calls based on WES data should be treated with caution. In contrast to CNV calling, runs of homozygosity (ROHs) associated with homozygous SNVs/indels could be detected robustly in first experiments. ROH detection also enables estimation of parental consanguinity in patients, with first test runs indicating high concordance between estimated and patient-reported kinship.