**Introduction.** Since the release of the first assembly of the human genome in 2001, our knowledge of the genetic architecture of complex traits and diseases has grown steadily. Genome-wide association studies (GWAS), which investigate the association between single-nucleotide polymorphisms (SNPs) and traits or disease risks, have played a major role in this development. Moreover, the emergence of next-generation sequencing (NGS) technology led to an ever-increasing amount of genetic data available to research. These developments have made sequencing and GWAS common approaches for research studies.

Recently, Mendelian Randomization (MR) has been popularized as a method to investigate the relationship between traits using readily available GWAS results.

Despite these successes, GWAS have not been able to completely explain the heritability that is observed in complex traits. In part, this is due to GWAS being mostly restricted to SNPs. However, large copy number variations (CNVs) have been shown to be associated with diseases but have largely been absent from GWAS. CNVs are difficult to sequence using the short reads generated by NGS. Novel long-read sequencing technologies have emerged and promise to make CNV detection easier.

**Project I.** In **project I**, we aimed to quantify the effect of SNPs on circulating estradiol levels and the effect of estradiol on bone mineral density.

We performed a GWAS using genotyping and estradiol measurements from UK Biobank (UKB) in males (N = 147,690) and females (N = 163,985). Estradiol was transformed into a binary phenotype (above/below detection limit of 175 pmol/L). We then quantified the effect of estradiol on bone mineral density (BMD) using MR.

We found 14 independent loci with genome-wide significant associations (P < 5E-8) with estradiol levels in males. One locus was also significant in females and we identified an additional locus specific to females. We found a significant effect of detectable estradiol levels on BMD in males (P=1.58x10-11) and females (P=7.48x10-6).

These results confirm previous research into the effect of estrogens on skeletal health.

**Project II.** In **project II**, we aimed to identify the effect of endogenous estradiol on breast, endometrial and ovarian cancer. We performed a MR analysis using the GWAS results from project I, estimating the effect of estradiol on the aforementioned cancers.

We found an association between estradiol levels and breast (P= 0.0074) and endometrial (P = 0.0065) but not ovarian cancer.

These findings highlight the effect of the body’s own estrogen production on cancer risk. Our results for endometrial cancer confirm previous MR studies.

**Project III.** In **project III**, we aimed to identify associations between CNVs and blood plasma protein levels and validate our findings using long-read sequencing.

We performed whole-genome sequencing in a cohort from Northern Sweden (N= 1,021) and called CNVs using CNVnator. We measured 438 plasma protein biomarkers in 892 participants using Olink Protein Extension Assay. Among the 872 participants with genotyping and protein data, we tested for association between copy numbers and protein levels. We validated 5 polymorphic CNVs in 15 individuals by WGS using Pacific Biosciences SMRT sequencing.

A total of 243,987 polymorphic non-overlapping CNV loci were identified. After merging of adjacent loci with consistent copy numbers, 23,381 remained. Significant associations ( ) were found between the copy numbers of 30 independent CNVs and the expression level of 17 protein biomarkers. Out of these CNVs, the breakpoints for two CNVs were validated by PacBio and two CNVs were identified to be clusters of many short repetitive elements. The fifth CNV could not be validated conclusively. Two CNVs could be validated by PacBio and two CNVs were identified to be clusters of many short repetitive elements. The fifth CNV could not be validated conclusively.

Our findings provide insight into the effects of CNVs on protein biomarkers and highlight the application of high-throughput sequencing for CNV detection, including issues with comparability between these technologies.