Institutionen för medicinsk epidemiologi och biostatistik

Molecular epidemiology of complex heritable disease: applications in genomics and metabolomics

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Hörsal Petrén, Nobels väg 12B

Fredagen den 25 januari, 2013, kl 09.00

av

Robert Karlsson

Huvudhandledare:
Dr. Fredrik Wiklund
Karolinska Institutet
Institutionen för medicinsk epidemiologi och biostatistik

Bihandledare:
Dr. Silvia Paddock
Rose Li and Associates, Inc.
Bethesda, MD, USA

Professor Juni Palmgren
Karolinska Institutet
Institutionen för medicinsk epidemiologi och biostatistik

Docent Jonathan Prince
Karolinska Institutet
Institutionen för medicinsk epidemiologi och biostatistik

Fakultetsopponent:
Dr. Antonis Antoniou
University of Cambridge, UK
School of Clinical Medicine
Centre for Cancer Genetic Epidemiology

Betygsnämnd:
Professor Lars Alfredsson
Karolinska Institutet
Institutionen för miljömedicin

Docent Henrik Larsson
Karolinska Institutet
Institutionen för medicinsk epidemiologi och biostatistik

Docent Patrik Rydén
Umeå universitet
Institutionen för matematik och matematisk statistik

Stockholm 2013

Abstract

Modern high-throughput molecular technologies (collectively referred to as "omic" platforms) are generating unprecedented amounts of data on human variation. The four papers in this thesis each investigate and characterize associations between common, complex, heritable disease, and genetic or metabolomic markers from omic platforms.

In paper I, we searched bipolar affective disorder (BPAD) pedigrees for genomic copy-number variants (CNVs, segmental deletions or duplications) segregating with disease. In one pedigree, a deletion in the gene *MAGI1* was observed in six out of six affected members. Upon further inspection, another pedigree was found with two out of three affected members carrying a duplication in the same gene. A pooled association analysis was subsequently carried out using in-house and public data sets on CNVs in control subjects and cases of BPAD, schizophrenia (SZ), or schizoaffective disorder (SA). *MAGI1* CNVs greater than 100 kb were found to be rare, nonsignificantly more common in BPAD cases than in controls, and significantly more common in the pooled case sample of BPAD, SZ, and SA than in controls.

In paper II, we studied a rare single nucleotide polymorphism (SNP) in the gene *HOXB13*, which had been recently reported to be strongly associated with prostate cancer (PC) risk. We genotyped and analyzed the variant G84E (rs138213197) in the two large Swedish PC case-control samples CAPS and Stockholm-1 (in total 4,903 cases and 4,589 controls). G84E was less rare in the Swedish samples than in the United States population previously studied, with a carrier rate over 1% in Swedish population controls. The variant was associated with a more than threefold increased relative risk of PC in both Swedish samples. G84E carriers' absolute lifetime risk to age 80 of PC was estimated to 33%. For G84E carriers in the uppermost quartile of a genetic risk score based on common risk SNPs, the same lifetime risk was estimated to 48%.

In paper III, a replication study of previously reported genetic associations with testicular germ cell tumor (TGCT) risk was performed. SNPs in six genes (*ATF7IP*, *BAK1*, *DMRT1*, *KITLG*, *SPRY4*, and *TERT*) were genotyped and analyzed in a combined case-parent, case-control sample from Sweden and Norway. In total, 831 case-parent triads, 474 dyads, 712 singleton cases, and 3,919 control subjects were analyzed. Our results supported the previously reported association with TGCT risk for SNPs in all six genes. Tests of interaction effects revealed no allelic effect differences for the two major TGCT histological subtypes seminoma and non-seminoma. However, a variant in the gene *SPRY4* was found to differ significantly in effect depending on the sex of the parent from which it was inherited. Only maternally inherited alleles were associated with TGCT risk.

In paper IV, a large range of small molecules in human serum, collectively called the metabolome, were studied for association with PC risk and aggressiveness. Samples from 188 controls, 188 PC patients with indolent disease, and 99 PC patients with aggressive disease were analyzed by ultra-performance liquid chromatography coupled with mass spectrometry, generating 6,138 quantitative molecular features. All features were tested for association with PC status, adjusted for patient age and sample storage time. Two features were significantly associated after correction for multiple testing, but none of them could be identified as specific molecules. Testing the PC-associated features for association with 1.4 million SNPs genome-wide produced the strongest associations in variants in annotated genes, which may aid future molecular identification efforts.

In conclusion, we have used omics platforms and modern computational tools to increase our knowledge about specific genetic risk factors and metabolomic markers for complex heritable disease. Our results may come of use in future etiological research as well as in genetic and molecular risk assessment.

ISBN 978-91-7549-013-7