1287W

Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2,000 Sardinians using tailored sequencing analysis tools. *J. Ding[1], C. Sidore[2,3], T. J. Butler[1], M. K. Wing[2], Y. Qian[1], O. Meirelles[1], F. Busonero[2,3], L. C. Tsoi[2], A. Maschio[2,3], A. Angius[3], H. M. Kang[2], R. Nagaraja[1], F. Cucca[3], G. R. Abecasis[2], D. Schlessinger[1].* 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

DNA sequencing identifies common and rare genetic variants for association studies, but studies typically focus on variants in nuclear DNA and ignore the mitochondrial genome. In fact, analyzing variants in mitochondrial DNA (mtDNA) sequences presents special problems, which we resolve here with a general solution for the analysis of mtDNA in next-generation sequencing studies. The new program package comprises 1) an algorithm designed to identify mtDNA variants (i. e. , homoplasmies and heteroplasmies), incorporating sequencing error rates at each base in a likelihood calculation and allowing allele fractions at a variant site to differ across individuals; and 2) an estimation of mtDNA copy number in a cell directly from whole-genome sequencing data. We also apply the methods to DNA sequence from lymphocytes of ~2,000 SardiNIA Project participants. As expected, mothers and offspring share all homoplasmies but a lesser proportion of heteroplasmies. Both homoplasmies and heteroplasmies show 5-fold higher transition/transversion ratios than variants in nuclear DNA. Also, heteroplasmy increases with age, though on average only ~1 heteroplasmy reaches the 4% level between ages 20 and 90. In addition, we find that mtDNA copy number averages ~110 copies/lymphocyte and is ~54% heritable, implying substantial genetic regulation of the level of mtDNA. Copy numbers also decrease modestly but significantly with age, and females on average have significantly more copies than males. The mtDNA copy numbers are significantly associated with waist circumference (p-value = 0. 0031) and waist-hip ratio (p-value = 2. 4×10-5), but not with body mass index, indicating an association with central fat distribution. To our knowledge, this is the largest population analysis to date of mtDNA dynamics, revealing the age-imposed increase in heteroplasmy, the relatively high heritability of copy number, and the association of copy number with metabolic traits.

1288T

PODKAT: a software package implementing the position-dependent kernel association test. *U. Bodenhofer, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University, Linz, Austria.

High-throughput sequencing technologies have facilitated the identification of large numbers of single-nucleotide variations (SNVs), many of which have already been proven to be associated with diseases or other complex traits. Several large sequencing studies, such as, the 1000 Genomes Project, the UK10K project, or the NHLBI-Exome Sequencing Project, have consistently reported a large proportion of private SNVs, that is, variants that are unique to a family or even a single individual. The role of private SNVs in diseases is poorly understood, largely due to the fact that it is statistically very challenging to consider private SNVs in association testing. While it is generally impossible to make use of private SNVs in single-marker tests or in correlation-based tests like the popular SNP-set (Sequence) Kernel Association Test (SKAT), also burden tests are facing serious statistical issues.

We have proposed the *Position-Dependent Kernel Association Test,* which is designed for detecting associations of very rare and private SNVs with the trait under consideration even if the burden scores are not correlated with the trait. The test assumes that, the closer two SNVs are on the genome, the more likely they have similar effects on the trait under consideration. This assumption is fulfilled as long as deleterious, neutral, and protective variants are grouped sufficiently well along the genome.

This contribution highlights a recently released software package, PODKAT, that implements the position-dependent kernel association test along with the popular SKAT test and all necessary tools for defining regions of interest, multiple testing correction, filtering, and visualization of results. The package is based on the widely used scientific computing platform R. It is publicly and freely available via the Bioconductor project. It is able to read data directly from VCF files and facilitates easy parallelization on multi-processor systems or computing clusters. Due to the special memory management strategies, analyses of large studies can even be performed on regular desktop computers (at the cost of much longer computation times). We will present the PODKAT package in detail along with examples of analyses performed on the real-world data sets, such as, the UK10K whole-genome cohorts.