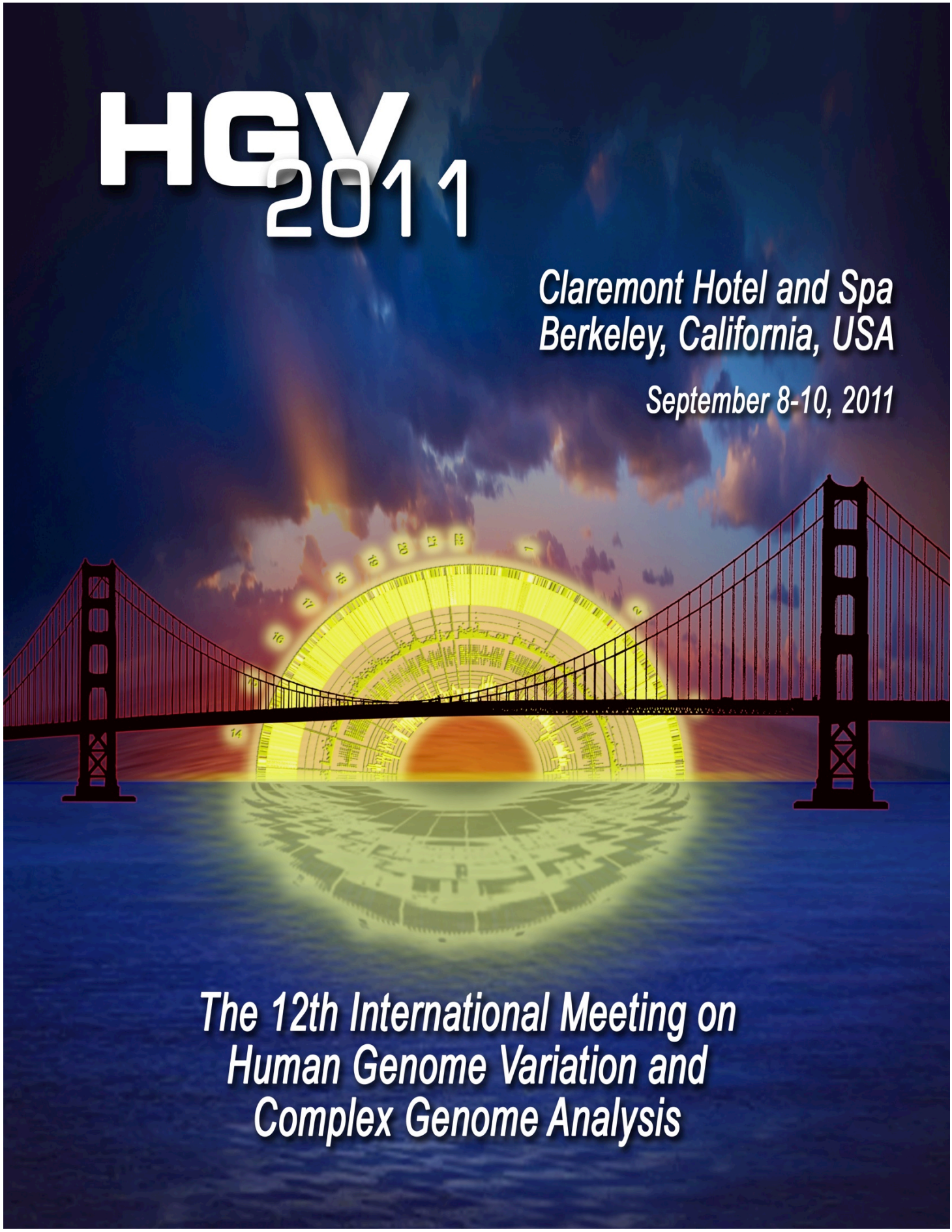
HGV
2011
Claremont Hotel and Spa
Berkeley, California, USA
September 8-10, 2011
The 12th International Meeting on
Human Genome Variation and
Complex Genome Analysis

A low false discovery rate at detection of copy-number aberrations in microarray data
Djork-Arné Clevert, Andreas Mayr, Andreas Mitterecker, Günter Klambauer, and Sepp Hochreiter

Motivation: A low false discovery rate (FDR) at the detection of copy-number aberrations (CNAs) in microarray data ensures sufficient detection power and prevents failures in CNA-disease association studies. A high FDR means many falsely discovered aberrations, which are not associated with the disease, though correction for multiple testing must take them into account. Thus, a high FDR not only decreases the discovery power of studies but also the significance level of the remaining discoveries after correction for multiple testing.

Methods: We obtain a low FDR at the detection of CNAs in microarray data by a probabilistic latent variable model, called 'cn.FARMS'. The model is optimized by Bayesian maximum a posteriori approach, where a Laplace prior prefers models, which represent the null hypothesis of observing a constant copy number 2 for all samples. The posterior can only deviate from this prior by strong (deviation from copy number 2 intensities) and consistent signals in the data, which hints at a CNA - the alternative hypothesis. The information gain of the posterior over the prior gives the informative/non-informative (I/NI) call that serves as a filter for CNA candidate regions. I/NI call filtering reduces the FDR, because a region with a large I/NI call is unlikely to be a falsely detected CNA, which would neither have strong nor consistent measurements. It can be shown that the I/NI call filter applied to null hypotheses of the association study is independent of the test statistic which in turn guarantees that a type I error rate control by correction for multiple testing is still possible after filtering. I/NI-calls perform well for the usually rare CNAs that are seen at few samples only, where variance-based filtering approaches fail.

Results: cn.FARMS clearly outperformed prevalent methods for CNA detection with respect to sensitivity and especially with respect to FDR on different HapMap benchmark data sets.
Availability: The software cn.FARMS is publicly available as an R package at Bioconductor and at http://www.bioinf.jku.at/software/cnfarms/cnfarms.html.

Djork-Arné Clevert
Johannes Kepler University
Gleimstrasse 13a
Berlin, 10437
Germany
Tel: +49-3044324702
Email: okko@clevert.de