DISCOVERY NOTE

Genome analysis
Exploring spatially adjacent TFBS-clustered regions with Hi-C data
H.Chen, S.Jiang, Z.Zhang, H.Li, Y.Lu and X.Bo ..................................................... 2611

ORIGINAL PAPERS

Genome analysis
SigSeeker: a peak-calling ensemble approach for constructing epigenetic signatures
J.Lichtenberg, L.Elnitski and D.M.Bodine ......................................................... 2615
Annotating function to differentially expressed LincRNAs in myelodysplastic syndrome using a network-based method
K.Liu, D.Beck, J.A.I.Thoms, L.Liu, W.Zhao, J.E.Pimanda and X.Zhou ...................... 2622
EBT: a statistic test identifying moderate size of significant features with balanced power and precision for genome-wide rate comparisons
X.Hui, Y.Hu, M.-A.Sun, X.Shu, R.Han, Q.Ge and Y.Wang .................................. 2631
An efficient graph kernel method for non-coding RNA functional prediction
N.Navarin and F.Costa ......................................................................................... 2642
Accounting for tumor purity improves cancer subtype classification from DNA methylation data
W.Zhang, H.Feng, H.Wu and X.Zheng ............................................................... 2651

Sequence analysis
HLA class I binding prediction via convolutional neural networks
Y.S.Vang and X.Xie ............................................................................................. 2658
Riblast: an ultrafast RNA–RNA interaction prediction system based on a seed-and-extension approach
T.Fukunaga and M.Hamada ............................................................................... 2666

Structural bioinformatics
A deep learning framework for improving long-range residue–residue contact prediction using a hierarchical strategy
D.Xiong, J.Zeng and H.Gong ............................................................................. 2675
EigenTHREADER: analogous protein fold recognition by efficient contact map threading
D.W.A.Buchan and D.T.Jones .......................................................................... 2684

Gene expression
Entropy-based consensus clustering for patient stratification
H.Liu, R.Zhao, H.Fang, F.Cheng, Y.Fu and Y.-Y.Liu ............................................ 2691
Network module-based model in the differential expression analysis for RNA-seq
M.Lei, J.Xu, L.-C.Huang, L.Wang and J.Li ....................................................... 2699

Systems biology
Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data

Data and text mining
PIXiE: an algorithm for automated ion mobility arrival time extraction and collision cross section calculation using global data association
J.Ma, C.P.Casey, X.Zheng, Y.M.Ibrahim, C.S.Wilkins, R.S.Renslow, D.G.Thomas, S.H.Payne,
M.E.Monroe, R.D.Smith, J.G.Teeuwarden, E.S.Baker and T.O.Metz .................... 2715
# Databases and ontologies

Neuro-symbolic representation learning on biological knowledge graphs  
M.A.Lshahrani, M.A.Khan, O.Maddouri, A.R.Kinjo, N.Queralt-Rosinach and R.Hoehndorf .......................... 2723

On patterns and re-use in bioinformatics databases  
M.J.Bell and P.Lord .................................................................................................................. 2731

# APPLICATIONS NOTES

## Genome analysis

Assembling draft genomes using contiBAIT  
K.O’Neill, M.Hills, M.Gottlieb, M.Borkowski, A.Karsan and P.M.Lansdorp ............................................. 2737

OMSim: a simulator for optical map data  
G.Miclotte, S.Plasance, S.Rombauts, Y.Van de Peer, P.Audenaert and J.Fostier ............................ 2740

twodpocr: an R/Bioconductor package and Shiny app for Droplet Digital PCR analysis  
A.Chiu, M.Ayub, C.Dive, G.Brady and C.J.Miller ................................................................................. 2743

## Sequence analysis

emMAW: computing minimal absent words in external memory  
A.Héliou, S.P.Pissis and S.J.Puglisi ........................................................................................................ 2746

rMFilter: acceleration of long read-based structure variation calling by chimeric read filtering  
B.Liu, T.Jiang, S.M.Yiu, J.Li and Y.Wang ....................................................................................... 2750

modlAMP: Python for antimicrobial peptides  
A.T.Müller, G.Gabernet, J.A.Hiss and G.Schneider ................................................................................ 2753

POSSUM: a bioinformatics toolkit for generating numerical sequence feature descriptors based on  
PSSM profiles  
J.Wang, B.Yang, J.Revote, A.Leier, T.T.Marquez-Lago, G.Webb, J.Song, K.-C.Chou and T.Lithgow ...... 2756

KMC 3: counting and manipulating k-mer statistics  
M.Kokot, M.Długosz and S.Deorowicz ................................................................................................. 2759

MarDRe: efficient MapReduce-based removal of duplicate DNA reads in the cloud  
R.R.Exposito, J.Veiga, J.González-Domínguez and J.Tourinó .................................................................. 2762

## Structural bioinformatics

Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta  
R.Kleffner, J.Flatten, A.Leaver-Fay, D.Baker, J.B.Siegel, F.Khatib and S.Cooper .............................. 2765

MD-TASK: a software suite for analyzing molecular dynamics trajectories  
D.K.Brown, D.L.Penkler, O.Sheik Amamuddy, C.Ross, A.R.Atilgan, C.Atilgan and  
Ö.Tastan Bishop ...................................................................................................................................... 2768

## Gene expression

veqtl-mapper: variance association mapping for molecular phenotypes  
A.A.Brown ............................................................................................................................................... 2772

RankProd 2.0: a refactored Bioconductor package for detecting differentially expressed features in  
molecular profiling datasets  
F.Del Carratore, A.Jankevics, R.Eisinga, T.Heskes, F.Hong and R.Breitling ............................................ 2774

## Genetics and population analysis

FlashPCA2: principal component analysis of Biobank-scale genotype datasets  
G.Abraham, Y.Qiu and M.Inouye ........................................................................................................... 2776

PopFly: the Drosophila population genomics browser  
S.Hervas, E.Sanz, S.Casillas, J.E.Pool and A.Barbadilla ....................................................................... 2779
Systems biology
MAgentA: a Galaxy implemented tool for complete Tn-Seq analysis and data visualization
K.M.McCoy, M.L.Antonio and T.van Opijnen ................................................................. 2781

Databases and ontologies
STOPGAP: a database for systematic target opportunity assessment by genetic association predictions

PROXIMATE: a database of mutant protein–protein complex thermodynamics and kinetics
S.Jemimah, K.Yugandhar and M.Michael Gromiha .......................................................... 2787

CORRIGENDUM
Kollector: transcript-informed, targeted de novo assembly of gene loci

ERRATUM
Molecular signatures that can be transferred across different omics platforms