POSTER LIST
ORDERED ALPHABETICALLY BY POSTER TITLE
GROUPED BY THEME/TRACK

THEME/TRACK: GENES

Poster numbers: P_Ge001 - 062 Application posters: *P_Ge001 - 004*

Poster number	EasyChair number	Author list	Presenting author	Title	Abstract	Theme/track	Topics
APPLICATION POSTERS WITHIN GENES THEME							
P_Ge001	*803*	*Haruka Ozaki and Itoshi Nikaido*	*Haruka Ozaki*	*ATAC2NET: A pipeline for reconstructing gene regulatory network based on ATAC-Seq data*	*Reconstruction of gene regulatory networks are important for understanding cell differentiation, cellular functions, and disease progression. Digital genomic footprinting using DNase I-Seq and ATAC-Seq can profile genomic occupancies of several hundreds of transcription factors in the same biological context at once. Thanks to the convenience of performing ATAC-Seq experiments, genome-wide chromatin accessibility data have been accumulating in public repositories, providing resource for reconstructing gene regulatory networks. However, although several studies evaluated performance of footprint detection programs for predicting TF binding, no systematic evaluations have been performed on computational methods for reconstructing gene regulatory network based on detected footprints. Moreover, it is unclear whether footprint detection programs designed for DNase I-Seq data is also effective for ATAC-Seq data. Here, we systematically evaluated the performance of computational methods for detecting footprints as well as for reconstructing gene regulatory networks using ATAC-Seq data. We showed that prediction performance was affected by properties of transcription factors, DNA amounts, and library sizes, rather than experimental methods used. We found that the reconstructed networks showed cell-type specificities properties, suggesting their biological significance. Based of these results, we developed ATAC2NET, a pipeline for reconstructing gene regulatory network based on ATAC-Seq data. We applied it several ATAC-Seq datasets and we are analyzing in detail the properties of the resulted networks. In addition, we are currently evaluating an alternative network reconstruction approach combining ATAC-Seq and gene expression data.*	*Genes/ Application poster*	*Application Fundamental*
P_Ge003	*802*	*Shishir Gupta, Roy Gross and Thomas Dandekar*	*Shishir Gupta*	*Re-annotation of the ant Camponotus floridanus genome, comprehensive analysis of its immune transcriptome and general reconstruction of ant interactomes*	*The sequencing of several ant genomes within the last six years open new research avenues for understanding not only the genetic basis of social insect species but also the complex systems such as immune responses. To form a better view of the immune repository and study the carpenter ant Camponotus floridanus immune responses against the bacteria, experimental data from Illumina sequencing and mass-spectrometry (MS) data in normal and infectious conditions for larvae and adults are analysed and integrated with bioinformatics approaches such as interactomics. Besides infection induced transcriptome profiling the data generated from Illumina sequencing was used for improving existing annotations, identifying alternative transcripts and functional modules in interactomes. We present our latest pipeline for gene prediction and alternative transcripts. The pipeline uncovered 1928 genes affected by alternative splicing events coding for 4666 alternative transcripts in C. floridanus. Current results allow better structural and functional annotation of C. floridanus genome, annotation of alternative transcripts, characterization of the immune system, transcriptome profiling and infection induced subnetworks of C. floridanus. Moreover, we analyze the protein-protein interactions (PPIs) of C. floridanus immune system with pathogenic bacteria such as Serratia marcescens and with the endosymbiont Blochmannia floridanus. We found that the immune system of C. floridanus is equally rich as in other insects, including diverse antimicrobial peptides and does not rely more on 'social immunity' as other social insects. Furthermore, our results indicate strong activation of immune defense in larvae while protection in adults depends mostly on ROS-mediated immunity.*	*Genes/ Application poster*	*Application Biotechnology*
P_Ge004	*864*	*Antonio Colaprico, Tiago Silva, Catharina Olsen, Luciano Garofano, Claudia Cava, Davide Garolini, Thais S. Sabedot, Tathiane M. Malta, Stefano M. Pagnotta, Isabella Castiglioni, Michele Ceccarelli, Gianluca Bontempi and Houtan Noushmehr*	*Antonio Colaprico*	*TCGAbiolinks: An R/Bioconductor package for integrative analysis with TCGA data*	*The Cancer Genome Atlas (TCGA) research network has made public a large collection of clinical and molecular phenotypes of more than 10 000 tumor patients across 33 different tumor types. Using this cohort, TCGA has published over 20 marker papers detailing the genomic and epigenomic alterations associated with these tumor types. Although many important discoveries have been made by TCGA's research network, opportunities still exist to implement novel methods, thereby elucidating new biological pathways and diagnostic markers. However, mining the TCGA data presents several bioinformatics challenges, such as data retrieval and integration with clinical data and other molecular data types (e.g. RNA and DNA methylation). We developed an R/Bioconductor package called TCGAbiolinks to address these challenges and offer bioinformatics solutions by using a guided workflow to allow users to query, download and perform integrative analyses of TCGA data. We combined methods from computer science and statistics into the pipeline and incorporated methodologies developed in previous TCGA marker studies and in our own group. TCGAbiolinks downstream analysis can be divided into 1) supervised analysis, comprising differential expression analysis, enrichment analysis, and master regulator analysis or 2) unsupervised analysis,: comprising inference of gene regulatory network, cluster, classification, ROC, AUC, feature selection, and survival analysis. Using four different TCGA tumor types (Kidney, Brain, Breast and Colon) as examples, we provide case studies to illustrate examples of reproducibility, integrative analysis and utilization of different Bioconductor packages to advance and accelerate novel discoveries.*	*Genes/ Application poster*	*Application*
OTHER POSTERS WITHIN GENES THEME							
P_Ge005	733	Aubin Samacoits, Florian Mueller and Thomas Walter	Aubin Samacoits	3D FISH image simulation framework to develop analysis method for mRNA localization	Many studies have characterized gene expression at the genome-wide level, but focused mostly on expression levels. However, only few studies focus on another key parameter: sub-cellular mRNAs localization. With single molecule FISH (smFISH) it is now possible to visualize individual mRNA molecules and hence investigate their spatial distribution in individual cells. However, to perform these analyses, several computational tools are necessary. First, cells need to be segmented and individual mRNA molecules be detected. While these image analysis tools are already well developed, there currently exists no validated statistical framework for the analysis of the mRNA localization. In such an analysis, the spatial coordinates of mRNAs are mapped into a carefully designed feature space. From this representation, machine-learning analysis will be performed to identify different mRNA localization classes, and eventually group genes according to their mRNA localization. In order to carefully develop and validate these feature sets and the subsequent machine-learning pipeline, an annotated image database with different localization classes is needed. Such validation databases exist already for cell segmentation or protein localization, but not for mRNA localization. Here, we present a virtual cell environment to simulate smFISH with non-random 3D mRNA localization. We base these simulations on experimental data, providing accurate 3D contours for cells and nuclei. Further, mRNAs are simulated considering realistic variations in their intensity and different experimentally observed localization patterns. Taken together, our approach yields realistic smFISH images, which can provide the basis for the development of a machine learning approach for mRNA localization classification.	Genes poster	Biotechnology
P_Ge006	814	Tine Goovaerts, Sandra Steyaert, Jeroen Galle, Tim De Meyer and Wim Van Criekinge	Tine Goovaerts	A mixture model for the omics based identification of monoallelically expressed loci and their deregulation in cancer	Imprinting is an epigenetic phenomenon leading to the expression of a single allele in a parent-of-origin specific manner. Inadequate computational techniques restrict insight in imprinting and diseases associated with imprinting deregulation, such as cancer. Hence, we introduce a mixture model for the identification of monoallelically expressed loci based on large scale omics data and a method to identify samples and loci featured by loss of imprinting.Our rationale is that RNA-seq (or similar omics data) for monoallelically expressed loci will exhibit apparent deviation from the Hardy-Weinberg equilibrium (HWE). As only one allele is expressed or epigenetically modified, heterozygous samples will ideally be recognized as homozygous. The model hence detects those loci in which the observed heterozygous fraction is shifted towards the homozygous fractions. Furthermore, it does not rely on prior genotyping and takes into account sequencing errors and possible partial imprinting. Once imprinted loci have been identified in control data, loci featured by loss of imprinting in the pathology under study can be identified.The model enabled the identification of 140 imprinted SNPs in 113 healthy control samples of the TCGA breast cancer RNA-seq data, corresponding to 53 genes. Some well-known imprinted loci, such as IGF2 and PEG3, were detected. Deregulation of these loci was investigated in 506 breast cancer samples from TCGA. Loss of imprinting - i.e. re-expression of the silenced allele - was observed in 8 of the imprinted SNPs.	Genes poster	Fundamental
P_Ge007	753	Lisa Barros de Andrade E Sousa and Annalisa Marsico	Lisa Barros de Andrade E Sousa	A statistical model for epigenetic regulation of miRNAs	microRNAs are small, non-coding RNAs involved in post-transcriptional gene regulation. Since the dysregulation of only a few miRNAs can affect many biological pathways, miRNAs are thought to play a key role in cancer development and can be used as biomarkers for cancer diagnosis and prognosis. In order to understand how miRNA dysregulation leads to a cancer phenotype it is important to determine the basic regulatory mechanisms that drive miRNA expression. Although much is known about miRNA-mediated post-transcriptional regulation, little is known about the epigenetic control of miRNAs. Here, we performed cell-line specific miRNA promoter predictions and built a classification model for expressed and non-expressed miRNAs. The classification model is based on several epigenetic features, e.g. histone marks and DNA methylation at both, miRNA promoters and around miRNA hairpins. We were able to classify intragenic and intergenic miRNAs with an accuracy of 79% and 85%, respectively, and identified the most important features for classification via feature selection. Surprisingly, we found that DNA methylation seems to have a dual role in regulating miRNA expression at transcriptional level: at promoters, high levels of DNA methylation correlate with transcriptional repression, while around miRNA hairpins high levels of DNA methylation have a positive impact on the expression level of the mature miRNA.	Genes poster	Fundamental
P_Ge008	709	Virag Sharma, Bjoern Langer, Leo Foerster, Pradeep Kiruvale and Michael Hiller	Virag Sharma	A Systematic Approach to Identify Gene Losses using Genome Alignments	Inactivation of protein-coding genes in different species is an important type of genomic change that can explain phenotypic differences among these species. For example, the loss of the Gulo gene in some mammals explains their inability to synthesize Vitamin C. While mutations in gene sequences can be detected from genome alignments, there is no method to systematically detect gene losses in an automated fashion.We have developed a computational pipeline that systematically searches for gene losses across different species without requiring any manual curation. Given a reference species and a genome alignment of the reference species with other species, our pipeline is able to identify the different types of gene inactivating mutations such as frameshifts and in-frame stop codons. To avoid mistaking artifacts for inactivating mutations, we strictly control for assembly gaps, low quality genomic sequences, alignment issues such as processed pseudogene misalignments and changes in gene structures.In order to associate gene losses with phenotypic changes, we applied the pipeline on a multiple genome alignment of 29 species with mouse as the reference and focused on gene losses in mammals which are either completely or partially blind. Our pipeline reports the inactivation of several known and potentially novel genes involved in vision-related functions. We conclude that the pipeline is a valuable tool for the compilation of a gene-loss catalogue for genomes that will be sequenced in the future. Furthermore, the pipeline will provide the basis to systematically link phenotypic changes to genomic changes using approaches like Forward Genomics.	Genes poster	Fundamental
P_Ge009	404	Patrick van den Berg, Stefan Semrau and Nikolai Slavov	Patrick van den Berg	An integrated transcriptomics and proteomics study of embryonic stem cell differentiation	Embryonic stem cells (ESCs) can be differentiated into all cell types of the adult body. In vitro differentiation of ESCs has therefore been used extensively as a model for embryonic development and it is critical for applications of ESCs in regenerative medicine and disease modeling. To differentiate ESCs into well-defined cell types, precise manipulation of gene expression is necessary. The majority of existing work has focused on transcriptional regulation of expression. Here, we study gene regulation at the level of protein turnover (translation and degradation) to discover novel ways to control ESC differentiation. In particular, we extracted mRNA and protein during retinoic acid induced differentiation of mouse ESCs. mRNA and protein abundance were then quantified by RNA sequencing and mass spectrometry, respectively. The measurement of 10 samples during a 96 h differentiation time course allowed us to follow the expression dynamics with unprecedented temporal resolution. We have developed a statistical model that identifies genes that are differentially regulated at the mRNA and protein level. After validation of the identified candidate genes we will unravel the general mechanisms that underlie their regulation.	Genes poster	Fundamental
P_Ge010	549	Peter-Bram 't Hoen, Eleonora de Klerk, Martijn Vermaat, Yavuz Ariyurek, Johan den Dunnen, Stephen Turner and Seyed Yahya Anvar	Peter-Bram 't Hoen	Analysis of PacBio full-length mRNA sequencing data uncovers widespread coupling between alternative transcription start sites, exons and polyadenylation sites	Short read sequencing technologies typically fall short in resolving complete transcript structures. The single molecule long read technology offered by the PacBio SMRT® technology provides reads that are well over the average size of an mRNA molecule and therefore generates complete cDNA sequences from the transcription start site until the polyadenylation site. The analysis of millions of these single-molecule long sequencing reads representing full-length mRNA molecules in MCF-7 human breast cancer cells and three human tissues provides the first opportunity to study coordination of transcription initiation, splicing and polyadenylation. To this end, we tested which alternative RNA features (transcription start sites, exon, polyadenylation site) were present more frequently or less frequently in the same transcript than expected by chance (mutually inclusive or mutually exclusive, respectively). Doing this, we found evidence for mutually dependent selection of alternative transcription initiation, splicing and/or polyadenylation sites in thousands of genes. The coordinated selection of mRNA features was often tissue-specific. Moreover, these events occurred across the entire mRNA molecule, where the selection of a particular transcription start site determined the selection of alternative exons or polyadenylation sites far downstream. A selection of events were subsequently validated by classical RT-PCR followed by Sanger sequencing. We conclude that there is an unprecedented degree of coordination between transcription, splicing and polyadenylation contributing to the transcript diversity observed in different tissues.	Genes poster	Fundamental
P_Ge011	467	Polewko-Klim Aneta, Lesiński Wojciech, Kitlas Golińska Agnieszka, Siwek Maria and Rudnicki Witold	Polewko-Klim Aneta	Application of the random forest method in identification of candidate genes in quantitative trait loci regions for adaptive immune responses of chicken	Current study aims at identification of the genetic markers associatedwith the variation of the variation of the adaptive immune traits in chicken. We have used machine learning methods to construct predictive models for the strength of response for three antibodies: KLH, LPS and LTA. The set of descriptive variables consisted of 384 SNPs preselected as candidates, based on the earlier work. Two procedures based on the Random Forest (RF) classifier were applied. To this end the predictive RF models were built and the relevance was assigned to variables using RF's perturbation importance as a measureof the relevance. The features that consistently show high relevance wereconsidered relevant. The entire procedure was performed within cross-validation loop.The predictive RF models based on these variables explain 11.6% of variance for KLH data, and roughly 3.5% of variance for LPS and LTA data. The procedure applied to a control run where antibody samples were collected before immunisation leads to a model with no predictive power.The number of SNPs identified as relevant in all 300 repeats was 10, 12 and 15 for KLH, LPS and LTA respectively. The respective numbers for 90% threshold are 17, 19 and 19.When the threshold is set at 50% of the numbers are 31, 27 and 30 for KLH, LPS and LTA respectively. Many SNPs identified in the study are common for more than one antigenicresponse. The SNPs identified in the study correspond to the several previouslyidentified genetic markers for immune response.	Genes poster	Agro-Food
P_Ge012	474	Brandon Malone, Ilian Atanassov and Christoph Dieterich	Brandon Malone	Bayesian Identification of Translation from Ribosome Profiling	Motivation:Ribosome profiling via high-throughput sequencing, riboseq, is a promising new technique for characterizing the occupancy of ribosomes on messenger RNA (mRNA) at base-pair resolution. The ribosome is responsible for translating mRNA into proteins, so information about its occupancy offers a detailed view of ribosome density and position which could be used to discover new translated open reading frames, alternative start codons and new isoforms.Contributions:We propose Rp-Bp, a Bayesian approach to predict the translation of open reading frames (ORFs) from riboseq data. In particular, Rp-Bp is useful for identifying novel translated short ORFs (micropeptides) and isoforms with high confidence. We use state-of-the-art Markov chain Monte Carlo techniques to estimate posterior distributions of the likelihood of ORF translation. A second novel contribution is automatic selection of periodic read lengths and ribosome P-site offsets via Bayesian model selection. Furthermore, we develop a competitive reference implementation for prediction based on the chi^2 test, Rp-chi.Results:We empirically demonstrate that our read length selection technique significantly improves sensitivity by resulting in up to an order of magnitude more predictions for Rp-Bp. Proteomics- and QTI-seq validation verifies the high quality of all of the predictions. Experimental comparison shows that Rp-Bp compares favorably to another recent tool for translation prediction. Qualitatively, we show that the method effectively identifies novel micropeptides and isoforms.Availability:The source code for Rp-Bp and Rp-chi is available at https://github.com/dieterich-lab/rp-bp.	Genes poster	Health

P_Ge013	349	Karl Koechert, Jie Cheng, Li Liu, Jose Garcia-Vargas, Barry Childs and Carol Pena	Karl Koechert	Biomarker identification in early clinical development – effective combination of hypothesis driven and data driven approaches in a clinical phase II trial assessing copanlisib activity in non-Hodgkin Lymphoma	Copanlisib, a novel pan-class I PI3K inhibitor with predominant activity against α and δ isoforms, has shown promising single agent activity in a phase 2 study in patients with indolent or aggressive NHL. Tumor gene expression profiling of 24 patients was used with both hypothesis- and data-driven approaches to identify genes or gene-signatures that may be associated with copanlisib treatment efficacy.The hypothesis-driven approach focused on pathways directly associated with copanlisib's mechanism of action, namely the B cell receptor (BCR)- and PI3K-signaling pathways, as well as disease-context pathways associated with e.g. tumor microenvironment. Gene expression of candidate pathways was integrated in a weighted manner to a patient-wise pathway score based on logistic or Cox regression models. Response rates were increased in patients with increased BCR and PI3K score (p=0.06 and 0.07; AUC=0.81 and 0.75, respectively). In addition, progression-free survival (PFS) was longer in copanlisib-treated patients with increased BCR score (HR=0.035, p<0.0001) and increased PI3K score (HR=0.24, p=0.02).The data driven approach used adaptive two way filtering (Cheng et al. 2012) combined with permutation-based cross validation to infer single genes predictive for best response or PFS and identified candidate genes with potential prognostic and/or predictive value, most prominently gene GPR18 (AUC=0.95; HR=5.8, p=0.01).In summary, using two complementary and robust analysis approaches, we have identified genes and gene signatures that are associated with objective response and PFS in this population of copanlisib-treated patients. Durable response to single-agent copanlisib is associated with tumors with activated PI3K/BCR pathways.	Genes poster	Health
P_Ge014	362	Irina A. Eliseeva, Ilya E. Vorontsov and Ivan V. Kulakovskiy	Ivan V. Kulakovskiy	Can transcription determine mRNA translation in mammals? Digging evidence with sequence analysis.	Transcriptional regulation of gene expression can determine mRNA stability and localization in yeast. It is an open question whether there is similar machinery in higher eukaryotes, e.g., whether translational state of a particular transcript can be defined at the transcriptional stage.In higher eukaryotes, the translation of many ribosomal and translational factors genes is controlled by the mTOR pathway that is directly involved in cell proliferation, aging, and oncogenesis. The 5' terminal oligopyrimidine sequence motif (TOP) is the specific feature of many mTOR translational targets. However, many mTOR targets carry improperly positioned non-terminal TOP or lack TOP completely.It is tempting to apply sequence analysis methods to identify transcriptional regulators that may leave imprints on transcribed mRNAs and thus determine forthcoming translational control. We utilized public CAGE and Ribo-Seq data to identify robust mTOR targets in human and mouse and performed sequence motif analysis of the respective promoter regions.Binding sites of several transcription factors were significantly enriched in promoters of the mTOR targets; among those transcription factors there were proteins having RNA-binding activity or direct interactions with other RNA-binding proteins. This suggests a principal role of transcription in mTOR translational control in higher eukaryotes.	Genes poster	Fundamental
P_Ge015	843	Sabrina Krakau, Hugues Richard and Annalisa Marsico	Sabrina Krakau	Capturing protein-RNA interaction footprints from iCLIP-seq data	RNA bindings sites for a protein of interest can now be detected genome-wide and at a high resolution thanks to the development of CLIP-seq technologies. Among these methods, iCLIP provides individual-nucleotide resolution and is particularly powerful for the characterization of protein-RNA interaction landscapes. However, existing methods for the analysis of iCLIP sequencing data suffer from several drawbacks: they do not account for the influence of transcript abundances nor do they model possible sources of technical or computational biases.To improve the analysis of such data we are developing an approach based on a non-homogeneous Hidden Markov model. Individual binding sites are called, taking into account regions enriched in protein bound fragments and the specifics of iCLIP truncation patterns. The underlying statistical framework enables us to simultaneously normalize for RNA abundances and to include as well other external data as covariates (e.g. nucleotide compositions, read lengths, mappability information). We devised a realistic iCLIP read simulation setup, that starts from real RNA-seq data and RNA binding sites, in order to evaluate our methods performance. Additionally we validate our approach using published iCLIP datasets from proteins with known predominant binding regions.Preliminary results on simulated data show that our tool is able to recover binding sites with a good accuracy. Further, on a real iCLIP dataset from the eIF4A3 protein our approach is in general more precise in determining the known binding regions than existing methods.	Genes poster	Fundamental
P_Ge016	792	Gwenneg Kerdivel and Valentina Boeva	Gwenneg Kerdivel	CIMP in adrenocortical carcinomas is associated with high expression of DNMT1 and increased Wnt and Notch signaling pathways activities.	Adrenocortical carcinomas (ACCs) are rare and aggressive endocrine cancer of the adrenal gland that exhibit recurrent genomic aberrations, negatively correlated with overall survival. Recently, a subtype of ACC characterized by a CpG island methylator phenotype (CIMP) has been discovered. CIMP is associated with especially poor diagnosis and one reason for this could be the promoter silencing through hypermethylation of tumor suppressor genes. By now, no drivers of CIMP in ACC have been identified. Using publicly available gene expression dataset of human ACCs from the TCGA and the Cochin Institute (Assié et al. 2014), we showed that DNMT1 expression is significantly increased in High-CIMP patients as compared to Low-CIMP patients, suggesting that DNMT1, rather than DNMT3A/B, could be responsible for the hypermethylation in CIMP tumors. Interestingly, expression of DNMT1 negatively correlates with overall survival. In addition, in patients with low or intermediate CIMP, the expression of DNMT1 allows a better discrimination of patients with good or poor survival. Together these results suggest that DNMT1 expression provides a reliable prognostic value. Moreover, we observed an inverse correlation between DNMT1 and APC expression, associated with an increased activation of Wnt signaling pathway in High-CIMP versus Low-CIMP samples (pathway analysis with ROMA). Not surprisingly, an increased activation of Notch signaling pathway is also observed as it is known to integrate Wnt signaling. Thus, the high aggressiveness of ACCs exhibiting high-CIMP as compared to low-CIMP could be due to the overactivation of these two pathways, known to cooperate in tumorigenesis of several cancer types.	Genes poster	Health
P_Ge017	573	Oren Tzfadia, Tim Diels, Klaas Vandepoele, Yves Van de Peer and Asaph Aharoni	Oren Tzfadia	CoExpNetViz: the Construction and Vizualisation of Co-expression Networks	Motivation:Comparative transcriptomics is a common approach in functional gene discovery efforts. It allows for finding conserved co-expression patterns between orthologous genes in closely related plant species, suggesting that these genes potentially share similar function and regulation.Existing co-expression tools are limited to data from model systems, which greatly limit their utility. Moreover, in addition, none of the existing pipelines allow plant researchers to make use of their own unpublished gene expression data for performing a comparative co-expression analysis and generate multi-species co-expression networks.Results:We introduce CoExpNetViz, a computational tool that uses a set of query or 'bait' genes as an input (chosen by the user) and a minimum of one pre-processed gene expression dataset	Genes poster	Biotechnology
P_Ge018	423	Josef Panek	Josef Panek	Computational modeling of RNA secondary structure using a novel approach	Information about evolutionary conservation of RNAs is employed for RNA secondary structure prediction in pairwise manner. For evolutionarily related RNAs, conserved structural segments are identified using pairwise sequence alignment and their structure is copied from known, experimentally resolved RNA structure into predicted structure. The remaining structural segments, showing week or no conservation, are predicted de novo using a standard prediction algorithm and merged with structure of conserved segments according to their position in the alignment. The presented approach is demonstrated here by modeling of secondary structure of mammalian ribosomal ribonucleic acids, one of the most essential biological molecules, whose structure is extremely large and complex.	Genes poster	Fundamental
P_Ge019	456	Lukasz Kreft, Pieter De Bleser, Paco Hulpiau, Arne Soete, Alexander Botzki and Yvan Saeys	Lukasz Kreft	ConTra v3: a tool to identify transcription factor binding sites across species, update 2016	Transcription factors are important gene regulators with distinctive roles in development, cell signaling and cell cycling, and they have been associated with many diseases. The ConTra v3 web server allows easy visualization and exploration of predicted transcription factor binding sites in any genomic region surrounding coding or non-coding genes. In this updated version, users can choose from nine reference organisms ranging from human to yeast. ConTra v3 can analyze promoter regions, 5'-UTRs, 3'-UTRs and introns or any other genomic region of interest. Thousands of position weight matrices are available to choose from, but the user can also upload any other matrices for detecting specific binding sites. Besides this visualization option, additional new exploration functionality is added to the tool that will automatically detect transcription factor binding sites (TFBSs) having both the highest regulatory potential and the highest conservation scores of the genomic regions covered by the predicted transcription factor binding sites. The regulatory potential is calculated based on the number of predicted TFBSs weighed by their distances to the reported transcription start site of the gene of interest. A typical analysis is run in four simple steps of choosing the gene, the transcript, the region of interest and then selecting one or more transcription factor binding sites for visualization or, alternatively, let ConTra v3 explore the transcription factors most likely regulating your gene of interest. The ConTra v3 web server is freely available at http://bioit.irc.ugent.be/contrav3/index.php	Genes poster	Biotechnology
P_Ge020	338	Maarten van Iterson, Erik van Zwet, Bastiaan Heijmans and Eline Slagboom	Maarten van Iterson	Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution	Association studies on omic-level data other then genotypes (GWAS) are becoming increasingly common, i.e., epigenome- and transcriptome-wide association studies (EWAS/TWAS). However, a tool box for the analysis of EWAS and TWAS studies is largely lacking and often approaches from GWAS are applied despite the fact that epigenome and transcriptome data have very different characteristics than genotypes.Here, we show that EWASs and TWASs are prone not only to significant inflation but also bias of the test statistics and that these are not properly addressed by GWAS-based methodology (i.e. genomic control) and state-of-the-art approaches to control for unmeasured confounding (i.e. RUV and cate).We developed a novel approach that is based on the estimation of the empirical null distribution using Bayesian statistics. Using simulation studies and empirical data, we demonstrate that our approach maximizes power while properly controlling the false positive rate. Finally, we illustrate the utility of our method in the application of meta-analysis by performing EWASs and TWASs on age and smoking which highlighted an overlap in differential methylation and expression of associated genes. We implemented our new method to control for bias and inflation of test statistics in the software bacon available from http://bioconductor.org/packages/bacon/.	Genes poster	Fundamental
P_Ge021	457	Petr Nazarov, Matthieu Gobin, Andrei Zinovyev, Eric van Dyck and Laurent Vallar	Petr Nazarov	Decomposition of transcriptional signal from tumours using independent component analysis	Tumour samples have complex cellular composition and show a high level of heterogeneity. The presence of stromal and immune cells, as well as polyclonality of cancer cells, limits interpretability of collected high-throughput data. Here we investigated and applied Independent Component Analysis (ICA) to decompose mixed signals in RNAseq data.First, we validated ICA approach in silico. Five cancers presented at TCGA repositories were selected: two brain cancers (GMB, LGG), melanoma (SKCM), lung squamous cell carcinoma (LUSC) and breast cancer (BRAC). Synthetic mixtures of their gene expression profiles were generated and then decomposed by ICA. We showed that, in order to obtain a robust separation, special attention to data transformation was needed and multiple runs of ICA were required.Next, we performed an in-depth analysis of 169 GBM and 473 SKCM samples. Gene signatures specific to each independent component were determined and associated to gene ontology categories. We identified components originated from different cell types and biological processes – some common and some specific to each tumour. Strong immune signals, neural tissue development and cell proliferation components were seen in both cancers, whereas components linked to melanin and keratin production – only in SKCM. Involvement of each component in samples was linked to clinical factors by ANOVA. We found a strong statistical connection between some of the components and methylation status. In GBM, many components were linked to Verhaak's tumour subclasses.Therefore, we conclude that ICA can detect cell subpopulations in bulk tissues, and help identifying gene signatures with diagnostic potential.	Genes poster	Fundamental
P_Ge022	637	Konstantina Dimitrakopoulou, Elisabeth Wik, Lars Akslen and Inge Jonassen	Konstantina Dimitrakopoulou	Deconvolution of transcriptome data from heterogeneous tissue samples	Microarray and RNA-sequencing technologies are key components in systems medicine approaches towards our comprehension of disease mechanisms. However, classical approaches for the analysis of expression data from complex tissue samples are highly biased by the heterogeneity and the variability in cell type composition. To facilitate transcriptome-based predictive and prognostic models for human diseases, it is necessary to deconvolve the tissue expression into the component expression profiles of each cell type. Experimental techniques such as cell sorting and laser-capture microdissection can physically separate the defined cell types before gene expression analysis, but they are time and resource demanding and can add additional stress on cells and thus affect their gene expression profiles.In silico deconvolution methods represent an appealing alternative to physical cell separation methods. We employ the linearity assumption in which the expression levels measured from a mixed sample are modeled as the weighted average of expression of the different cell types. In particular, we developed a method to estimate both the cell type proportions and the cell type-specific gene expression profiles directly from the mixed expression data based on representative profiles without requiring prior information on cell type-specific expression signatures or cell type proportions. We assess the performance of our approach on benchmark expression datasets and compare it with state-of-the-art existing methods.	Genes poster	Health
P_Ge023	608	Kristoffer Niss, Lasse Folkersen, Claus Berthelsen, Kirstine Belling and Søren Brunak	Kristoffer Niss	Decreased immune gene expression variation along the colon in non-inflamed mucosa of ulcerative colitis patients	Ulcerative colitis (UC) is an inflammatory disease of the colon believed to occur in genetically susceptible individuals exposed to a combination of environmental and microbial factors. The inflammation typically begins in the rectum and over time transitions along the colon in a proximal direction. This migratory progression suggests that UC-induced inflammation can not take hold of the entire colon at disease onset, but is limited to certain susceptible colonic segments. A comparative analysis of the colonic segments may provide knowledge of the etiology of UC, which is still limited.Mucosal biopsies of healthy donors (n=28) and UC patients (n=53) were taken from 1-6 colonic segments and microarray gene expression analyses were performed, yielding 217 samples. By applying segment-specific scaling to the expression levels of each gene, we constructed patterns that emphasize the gene expression fluctuations along the colon. This was done for each condition: healthy, inflamed UC (iUC) and non-inflamed UC (nUC). KMeans clustering of all genes using the three condition patterns together revealed major expression tendencies and a functional analysis of each cluster demonstrated clear intra-cluster gene relationships.A single cluster was strongly enriched for immune response related genes and had significantly (Q=0.05) different patterns in healthy, nUC and iUC for 324/2013 genes. In this cluster, we observed a clear state-change between inflamed and non-inflamed UC and that the gene expression variations between colonic segments in healthy individuals are reduced in nUC samples, possibly because the expression profile of proximal colon samples have changed.	Genes poster	Health
P_Ge024	409	Nicolas Nahuel Moreyra, Julian Mensch, Juan Hurtado and Esteban Hasson	Nicolas Nahuel Moreyra	Differential expression analysis of cold tolerance adaptation by RNA-seq de novo approach.	Over the last years, the role of temperature-related gene expression in ecological adaptation has been receiving increasing attention. Previous findings of our group identified specific cold adaptations involving energy metabolism and arrest of reproduction in females of the fly Drosophila buzzatii in response to winter conditions. We performed a RNA-seq analysis to investigate changes in gene expression profiles in order to identify the genetic basis of such cold adaptations.The study was conducted by exposing sets of females to three thermal conditions: one involving the cold tolerant flies and two control treatments. We used the Trinity software to generate a de novo assembly from RNA-seq reads. To analyze expression levels of the reconstructed transcripts, we mapped the reads against the transcriptome and then estimated the number of RNA-seq fragments (counts) that mapped to each contig. Transcripts were filtered based on a mean count value cutoff and normalized by the TMM method to get the differences in RNA composition. Thereby, we extracted transcripts that were at least 5.6-fold differentially expressed at a significance of 1E-3 in any comparison. This step allowed us to compare over and under-expression transcript clusters. Based on a well-annotated genome, we made a blastx against D. melanogaster genome. The results showed an over-representation of specific genetic pathways and cellular processes, highlighting the relevance of sugar metabolism, glycolysis and oxidative phosphorylation in the expression of cold tolerance. Future studies will assess the role of natural selection shaping the evolution of differentially expressed genes identified.	Genes poster	Ecosystems
P_Ge025	581	Ole Eigenbrod, Jane Reznick, Damir Omerbasic and Gary R. Lewin	Ole Eigenbrod	Discovering molecular signatures of extreme physiology using African mole-rats	The African mole-rats (Bathyergidae) are a family of subterranean rodents with very unusual physiological traits for mammals. The most famous member of African mole-rats is the naked mole-rat (Heterocephalus glaber), which shows several extraordinary phenotypes like poikilothermy, extreme longevity, cancer resistance and extreme adaptation to low oxygen environments. Additionally, the naked mole-rat and some other Bathyergidae species are insensitive to several noxious substances or algogens (e.g. acid, capsaicin, or mustard oil) [Park et al., PLOS Biology 2008].This study focuses on understanding the sensory phenotypes of at least 8 African mole-rat species, as these closely related species show different patterns of insensitivity to noxious substances. Recently, a sequence motif in the NaV1.7 ion channel of the naked mole-rat was found to be directly connected to its acid insensitivity [Smith et al., Science 2011].We sequenced poly-A selected mRNA from multiple tissues of 8 African mole-rat species. As there are no annotated genomes available for most of the species, we performed de-novo transcriptome assembly to obtain the protein-coding sequences. We developed a bioinformatic workflow to annotate putatively coding transcripts and exclude contaminating or falsely assembled sequences and chimeras. Using this approach, we were able to identify more than 10,000 unique protein-coding transcripts per species. We also directly compared the protein-coding sequences and transcript levels across species boundaries. This approach allows a multivariate analysis of the relationship between gene expression level, sequence variation and extreme phenotypes across this rodent family.	Genes poster	Fundamental

P_Ge026	708	Foivos Gypas, Andreas Gruber, Alexander Kanitz and Mihaela Zavolan	Foivos Gypas	Discovery, annotation and abundance estimation of transcript isoforms from high-throughput sequencing data	Mammalian genes typically have multiple transcription initiation and termination sites and exon forms that are used in a cell type specific manner to generate distinct transcript isoforms. In recent years it has become clear that an improved accuracy of transcript isoform abundance leads to a better understanding of cellular processes, such as, for example, miRNA-dependent gene regulation. A variety of methods have been proposed for the estimation of transcript isoform abundance from RNA-Seq data. We recently showed that many of them have comparable accuracy, but some excel in their efficiency [1]. A main bottleneck in estimating transcript isoform abundance is the availability of a complete and accurate set of transcript sequences. However, methods for transcript reconstruction based on mRNA-seq data are in their infancy and do not seem sufficiently accurate [2], even when a reference annotation is provided. In this work we use heterogeneous sequence data sets to expand the set of annotated transcript forms and thereby improve the estimation of transcript abundance across cell types. Our results have implications for the analysis of gene expression and for the analysis of protein variants in different cell types.1. Kanitz A, et al. Genome Biol. 20152. Hayer KE, et al. Bioinformatics. 2015	Genes poster	Fundamental
P_Ge027	434	Yao-Ming Chang, Arthur Chun-Chieh Shih, Ling Li, Ya-Ting Chang and Chien-Chang Chen	Yao-Ming Chang	Dynamically Genetic Program by Co-regulated TF Groups during the Pressure Overload-induced Cardiac Hypertrophy in Mice	Many heart diseases, such as hypertension, heart failure, and valvular heart disease, are accompanied by the cardiac hypertrophy. Understanding comprehensively what transcription factors (TFs) induce the hypertrophic process and when this process begins after pressure overload will be important in providing potential therapeutic targets in treating cardiovascular diseases. In this study, we collected the whole transcriptome data, including gene and miRNA expression data, isolated from hypertrophic murine hearts subjected to transverse aorta banding surgery (TAB) and without TAB surgery (sham) at five time points among the four weeks, respectively. From three analytical perspectives, we analyzed the whole transcriptome differences, functional distributions of differentially expressed genes, and clustered transcription factor (TF) coexpression network. In the result, we found that the globally genetic change began in the early stage, after cardiac pressure-overloaded, earlier than morphological change; moreover, the globally genetic change returned to a normal level within few days while the cardiac size kept on enlarging. It reveals the inconsistent timing between genetic and morphological changes. In addition, we also identified quite a few of TFs and miRNAs that differentially expressed in different stages and many of them have been also found in literature. Interesting, some miRNAs, verified with cardiac functions previously, were expressed not in the early stage but in dozens of days after TAB that indicates their suppression role in the hypertrophic process. In short, using and analyzing a time course transcriptome data our results can enhance the understanding in the dynamically genetic regulation in cardiac hypertrophy.	Genes poster	Fundamental
P_Ge028	485	Aslihan Gerhold-Ay, Johanna Mazur and Harald Binder	Aslihan Gerhold-Ay	Enhancing prediction performance by using mapping approaches for data integration of RNA-Seq and methylation data	High-dimensional data of next-generation sequencing platforms enable the development of molecular signatures for prediction of clinical endpoints like death or case-control status. The integration of heterogeneous data types can help for better prognostic modelling and to understand the underlying biological mechanism. The challenge for the integration of studies is to connect entities present in RNA-Seq data on gene-expression and methylation data on CpG sites. The optimal allocation problem has not been solved yet. Our aim is to investigate how prediction performance can be used as a measure for finding the optimal mapping of CpG sites to their related genes. For evaluation of our mapping approaches two different data sets were used. To obtain the optimal mapping, we define a window of nucleotides around genes. In a two-step approach, first we use regularized regression approaches to estimate a gene-signature based purely on the RNA-Seq data. In the following step, methylation data of the CpG sites that are falling within a window around the signature genes are used to estimate a new signature. The performance of the resulting signature is used to judge mapping quality. We compare different window sizes and show the distinct effect on prediction performance with respect to the endpoint. Mapping approaches for the integration of high-dimensional data can be powerful tools in medical research for more effective treatment of patients. Thus, prediction performance in the proposed approach could be recommended for the allocation of CpG sites and genes, if biological knowledge provides no clear guidance.	Genes poster	Health
P_Ge029	560	Inken Wohlers, Andriy Mashychev, Marcel Schilling, Christina M. Lill and Lars Bertram	Inken Wohlers	Evaluating the prediction of SNPs with effects on miRNA-mediated mRNA expression using transcriptome sequencing data	MicroRNAs (miRNAs) are short 19-22 base pair RNAs that post-transcriptionally alter the expression of mRNAs. This is achieved by binding to specific regions predominantly located in the 3' UTR within the target mRNA, which decreases protein output. We hypothesize that single nucleotide polymorphisms (SNPs) located in or near the miRNA-to-mRNA binding sites interfere with this process. To assess this hypothesis, we previously developed a bioinformatics pipeline that predicts the putative effect of all variants in dbSNP on miRNA-mediated changes in transcript expression. To this end, it uses miRNA-target sites predicted to reside in 3' UTRs of human transcripts and scores the effects of SNPs nearby. In this work, we utilize transcriptome sequencing data to generate a set of reference SNPs for benchmarking our predictions. These are SNPs for which we know – under the given physiological conditions – whether they are linked to a miRNA-mediated effect on transcript expression. This reference dataset is created from public mRNA and small RNA sequencing data generated from 345 lymphoblast cell lines as part of the Geuvadis project (www.geuvadis.org), which we re-processed and re-analyzed using updated protocols and reference panels. Expression profiles are then probed for those miRNA-mRNA combinations which are significantly correlated, followed by cis transcript eQTL analyses to identify SNPs with allele-specific effects on mRNA expression. A subset of the derived transcriptome-based reference data is used for optimizing the accuracy of our predictions using ROC analysis. The optimal prediction model is subsequently assessed in the remainder of the reference data.	Genes poster	Health
P_Ge030	592	Michaela Bayerlova, Annalen Bleckmann and Tim Beissbarth	Michaela Bayerlova	Evaluation of gene signatures applied to expression data of cancer patient cohorts	A gene signature is a collection of gene markers whose mRNA expression is associated with clinical outcome or can guide treatment decisions. With the advance in large-scale gene expression profiling technologies, multiple gene signatures have been established for further classification of cancer diseases into molecular subtypes. We examined two approaches of signature integration with patient data: 1) We applied a newly derived signature to public patient data to test its prognostic power. 2) We utilized a published gene signature for the classification of newly sequenced metastasis samples of patients.1) We derived a new pathway-based gene signature based on a pro-invasive perturbation of a pathway receptor in breast cancer cells. Subsequently, we tested the pro-invasive effect of the signature genes in expression data of breast cancer primaries with clinical annotation of metastasis events. The tumours were analysed by hierarchical clustering of the signature expression patterns and the identified patient clusters were subjected to Kaplan-Meier analysis of metastasis-free survival. The patient sub-groups showed significant differences in prognosis of metastasis development in breast cancer.2) In the second application, we first generated RNA-seq data of liver metastasis samples originating from colon cancer primaries. A published signature for defining molecular subtypes of colorectal cancer was applied using a nearest shrunken centroid-based method. We investigated whether the identified metastasis subgroups reflect characteristics of the primaries subtypes and evaluated the usage of the primary tumour signature extended to metastatic tissue setting.Furthermore, we critically evaluated signature-based classifications in respect of interpretability and clinical relevance.	Genes poster	Fundamental
P_Ge032	852	Lorena de La Fuente Lorente, Ana Conesa, Manuel Tardaguila, Hector Del Risco, Cristina Marti, Victoria Moreno and Susana Rodriguez	Lorena de La Fuente Lorente	FAIR, Functional Analysis at Isoform Resolution by using long reads technologies	Based on the claimed role transcript variants in conferring functional meaning and the lack of methods to study the functional implications of alternative splicing (AS) and alternative polyadenylation (APA), we have developed a new methodology called FAIR. This methodology will let to address the functional profiling of transcript and protein isoforms at a genome-wide level by using long-reads technologies. Moreover, we have implemented it in a software called Transcript2GO.Therefore, using PacBio and Illumina data, FAIR can generate functional hypothesis about the role of alternative isoforms in our system. First, FAIR allows the functional annotation of each PacBio-resolved isoform which involves the ORF prediction and the annotation of several functional layers: miRNA binding sites, PFAM domains, post-translational modifications, UTR motifs, NMD prediction, repetitive elements, etc. Finally, it applies different statistical methods which combine both expression data and functional annotation over each PacBio-resolved isoform. Among the several included statistical methods, we can highlight the Feature Differential Splicing (FDS) which is able to point out functional elements affected by AS/APA.Using our rich annotation pipeline over a neural differentiation system, we found that nearly all genes expressing several isoforms have them annotated with at least one differential functional label, suggesting that functional profiling at isoform resolution is meaningful. We identified several functions enriched in genes regulated by differential splicing, as well as specific features as miRNAs regulated by DS. Other functional insights of the relationship between function and differential splicing are easily revealed by the tools implemented in Transcript2GO.	Genes poster	Fundamental
P_Ge033	734	Rianne Beukhof, Madelon Engels, Sanne Abeln, Bas Stringer, Maurits Dijkstra, Ted Meeds and Jaap Heringa	Madelon Engels	First among Equals -- Discriminating Driver and Passenger Mutations	Carcinogenesis is typically driven by the accumulation of deleterious mutations. Combined with other clinical observations, these driver mutations allow experts to discriminate between different types of cancer, which is essential to accurately predict prognoses of available treatments, and also to develop new ones. However, many types of cancer cause genetic instability, introducing a multitude of passenger mutations in afflicted cells. Typical passenger mutations have no direct clinical relevance, but their abundance complicates the identification of driver mutations.Our study assesses which features can improve the methods we use to distinguish between driver and passenger mutations. Preliminary results were gathered using exome sequencing data from The Cancer Genome Atlas (TCGA) for four different types of cancer. Mutations in known driver genes occur in regions of the genome with a high evolutionary conservation score more often than expected by chance. Certain types of mutation are also correlated. For example, mutations causing a frameshift are more common in known driver genes, whereas silent mutations are statistically underrepresented. Frequency of mutation, however, appears to have no predictive value when considering the types of cancer separately. We further investigate these trends in a handful of case studies.	Genes poster	Health
P_Ge034	681	Anna Feldmann and Nico Pfeifer	Anna Feldmann	From Predicting to Analyzing HIV-1 Resistance Towards Broadly Neutralizing Antibodies	Recently, combination therapy with broadly neutralizing antibodies (bNAbs) was introduced as a viable new option in antiretroviral treatment against HIV-1, that is capable of reducing viral load under detectable levels for up to 60 days in humanized mice and non-human primates. First clinical trials showed that already a single infusion of one bNAb, 3BNC117, is able to suppress successfully viremia in HIV-1 infected humans and even enhance the antibody responses of the individuals. However, the efficacy of this treatment is also affected by the emergence of resistant strains. Prior to the administration of an antiretroviral bNAb combination therapy to a patient, it has to be ensured that the patient's viral strains are susceptible to the particular bNAbs of the combination. So far, resistance to bNAbs can only be tested in expensive and time-consuming neutralization assays. We propose a non-linear SVM-based model to predict the neutralization susceptibility of unseen viral strains to bNAbs based on the viral envelope sequence. Because non-linear SVM classification results are often difficult to interpret, we offer different visualization techniques to improve the biological interpretability of the results using feature space visualization and motif logos. Learning the important binding sites of the bNAbs, the models are also biologically meaningful and useful for epitope recognition. Moreover, we confirmed a trend towards antibody resistance for the subtype B HIV-1 population and extended the analysis to the global HIV-1 population by predicting the neutralization sensitivity for around 36,000 HIV-1 sequences from the Los Alamos National Laboratory HIV Sequence Database.	Genes poster	Health
P_Ge035	757	Arlin Keo	Arlin Keo	Functional analysis of polyQ genes by examining spatial co-expression across the human brain.	Polyglutamine (polyQ) diseases are inheritable, neurodegenerative disorders caused by an expansion of a CAG repeat tract in the coding region of one of the polyQ disease-associated genes. There are nine polyQ diseases which include Huntington's disease (HD) and multiple spinocerebellar ataxias (SCAs), each with their own causative gene. It is known that a longer CAG repeat tract leads to an earlier onset of the disease, but not all differences in age of onset can be explained by repeat length. Recent studies have shown that interaction among the polyQ genes affects the age of onset in HD and SCAs. In this study we aim to find the functional relations among the nine polyQ genes by analyzing their co-expression patterns across the human brain using the Allen Human Brain Atlas data. This high-resolution spatial microarray data allows the construction of gene-gene networks on a whole brain level as well as on a region-specific level. Genes that co-express with multiple polyQ genes are indicators of interaction between the polyQ genes and potentially play a role in the age of disease onset. Moreover, sets of genes co-expressed with each of the polyQ genes may give rise to the functional relatednes when examining the common functional pathways in which they are involved.	Genes poster	Fundamental
P_Ge036	736	Ahmed Mahfouz, Boudewijn P.F. Lelieveldt, Aldo Grefhorst, Lisa T.C.M. van Weert, Isabel M. Mol, Hetty C.M. Sips, Jose K. van den Heuvel, Nicole A. Datson, Jenny A. Visser, Marcel J.T. Reinders and Onno. C. Meijer	Ahmed Mahfouz	Genome-wide co-expression of steroid receptors in the mouse brain: identifying signaling pathways and functionally coordinated regions	Steroid receptors are pleiotropic transcription factors that coordinate adaptation to different physiological states. An important target organ is the brain, but even though their effects are well studied in specific regions, brain-wide steroid receptor targets and mediators remain largely unknown due to the brain complexity. Here, we tested the idea that novel aspects of steroid action can be identified through spatial correlation of steroid receptors with genome-wide mRNA expression across different regions in the mouse brain. First, we observed significant co-expression of six nuclear receptors (Estrogen Receptor alpha, Esr1; and beta, Esr2; Androgen Receptor, Ar; Progesterone Receptor, Pgr; Glucocorticoid Receptor, Gr; and Mineralocorticoid Receptor, Mr) with sets of steroid target genes that were identified in single brain regions. These co-expression relationships were also present in distinct other brain regions, suggestive of as yet unidentified coordinate regulation of brain regions by e.g. glucocorticoids and estrogens. Second, co-expression of a set of 62 known nuclear receptor co-regulators and the six steroid receptors in 12 non-overlapping mouse brain regions revealed selective downstream pathways, such as Pak6 as a mediator for androgen and glucocorticoid receptor's effects on dopaminergic transmission. Third, Magel2 and Irs4 were identified and validated as strongly responsive to the estrogen diethylstilbesterol in the mouse hypothalamus. The brain- and genome-wide correlations of mRNA expression levels of six steroid receptors that we provide constitute a rich resource for further prediction and understanding brain modulation by steroid hormones.	Genes poster	Health
P_Ge037	598	Ge Tan and Boris Lenhard	Ge Tan	Genome-wide prediction of regulatory territories and target genes under complex long distance cis-regulation	Comparative genomics and high-throughput experimental methods like ChIP-Seq have enabled efficient detection of regulatory elements in metazoan genomes. Nevertheless, the assignment of those elements to their target genes has remained a difficult task. Traditional assignment to the nearest gene, or a manual and semi-intuitive process is far from complete, since regulatory regions can be located hundreds of kilobases away from their target genes, sometimes beyond neighboring genes. We previously showed that arrays of conserved noncoding elements span the loci of developmental regulatory genes ("targets") and several other genes ("bystanders"), and define the edges of genomic regulatory blocks (GRBs). We found that the target genes that respond to distal regulatory elements in those regions have specific features that distinguish them from bystander genes in the locus and the genome. In this study, we proposed a robust approach for the automated determination of GRB spans and a machine learning based method for genome-wide detection of target genes. The result is a comprehensive catalog of nearly one thousand human genes likely to be regulated by long-range interactions and the regions harboring their corresponding cis-regulatory elements. The catalog comprises a large number of genes involved in development, transcription and axon guidance. Furthermore, these genes are enriched for genes involved in complex diseases, including cancer and diabetes. The GRB spans and target genes identified in this study provide a rich resource for studying developmental regulation and disease-associated genomic variation.	Genes poster	Fundamental
P_Ge038	450	Charles-Henri Lecellier, Wyeth W. Wasserman and Anthony Mathelier	Charles-Henri Lecellier	Human enhancers associated with immune response harbor specific sequence composition, activity, and genome organization	Enhancers are distal DNA regions involved in the transcriptional regulation of gene expression.The Cap Analysis of Gene Expression (CAGE) technology allows for a precise identificationof active enhancer regions in biological samples by capturing bidirectional RNA transcriptsat enhancer boundaries. Using this technology, the FANTOM consortium recently characterized> 38, 000 human enhancers from about 800 cell and tissue types. This mapping provides us withan unprecedented opportunity to examine enhancers at large scale for specific DNA sequence featuresand functions. We used the distribution of guanine and cytosine nucleotides at enhancer regionsto distinguish two classes of enhancers harboring distinct DNA shape patterns. A functionalanalysis of their predicted protein-coding gene targets highlighted that one class of enhancers wassignificantly enriched for associations with immune response genes. Confirming this result, wefound that this class of enhancers was specifically enriched for regulatory motifs recognized byTFs involved in immune response (e.g. NF-kB). While these enhancers were generally repressedor lowly active, we observed that they were cell type specific and preferentially activated upon bacterialinfection, reinforcing their potential role in immune response. Looking at chromatin capturedata, we found that the two classes of enhancers were lying in distinct topologically-associateddomains and chromatin loops. Taken together, these results suggest that specific DNA sequencepatterns encode for classes of enhancers that are functionally distinct and specifically organized inthe human genome.	Genes poster	Fundamental
P_Ge039	513	Konrad Zych, Chris Maliepaard, Roeland E. Voorrips, Gerrit Gort, Nick de Vetten, Johan C.P. Hopman, Jan M. de Haas, Michiel A. Noback, Ronald Wedema, Jan-Peter H. Nap and Ritsert C. Jansen	Konrad Zych	Improving potato breeding with computational and functional genomics	Potato is one of the most important food crops. Potato is an outbred tetraploid plant making its breeding time-consuming and cumbersome. Including genetic markers in the selection process could greatly improve potato breeding. This approach was successfully used in selection for few monogenic traits (e.g. resistance to Phytophtora infestans). In our study we developed markers for reliable screening for multigenic quality traits like color after frying.We created a large potato population, consisting of two experimental crosses and a panel of cultivars and breeding clones. We performed RNA-Seq on the parents of the crosses in order to extract SNPs, from which we created a 60,000 SNP array. We used this array to genotype our population. We extended the mixture models based genotype calling of fitTetra (Voorrips et al, 2011). We used RNA-Seq data to obtain starting values for the algorithm increasing accuracy of the calling. The resulting genotypes were used together with multi-year high quality phenotypes in association studies. Using multiple levels of correction for population structure and environmental variance and multiple-marker association analysis we elucidated new markers for complex potato quality phenotypes.With our improved algorithm we were able to salvage 20% more high quality SNPs and filter out the lower quality SNPs. As a result, we created one of the most comprehensive genomic resources for potato with more than 30,000 SNPs measured in more than 1,500 samples. Association analysis resulted in a set of markers that could be used by the companies to extend their breeding scheme.	Genes poster	Agro-Food

P_Ge040	406	Saskia Trescher, Jannes Münchmeyer, Christopher Schiefer and Ulf Leser	Saskia Trescher	In-silico Approaches for Estimating Transcription Factor Activity from Transcriptome Data	The regulation of gene expression is indispensable for the adaptability of all organisms. It is predominantly controlled by a complex network of transcription factors (TFs). In order to elucidate regulatory principles between TFs and their putative target genes at different scales, numerous algorithms have been presented. Assessing their performance is an important task and facilitated by the availability of a growing number of transcriptome and TF binding datasets.We report on our result from comparing three different in-silico approaches for identifying the most influential regulators of genes using transcriptome data. Specifically, we compare our re-implementation of the work by Schacht et al. [1] and tools provided by ISMARA [2] and RACER [3]. All of them can integrate information about TF binding (from i.e. ENCODE, TRANSFAC) with sample-specific expression data (e.g. mRNA, DNA methylation, CNV) either in each sample or across phenotypes. The resulting most active regulators vary considerably among the investigated methods. Using different underlying TF-gene networks unveils a notable dependence of TF activity on the number of target genes. In order to resolve these discrepancies we plan to study systematically the influence of network topology and data structure using synthetic input.[1] doi:10.1093/bioinformatics/btu446[2] doi:10.1101/gr.169508.113[3] doi:10.1371/journal.pcbi.1003908	Genes poster	Fundamental
P_Ge041	487	Hyojin Kang, Chul Kim, Boseok Seong and Seokjong Yu	Hyojin Kang	Integrated approach to combine RNA-seq- and Microarray-derived gene co-expression networks in Alzheimer's disease	Gene co-expression networks (GCNs) are graphic representations of genes showing similar expression pattern across tissues and experimental conditions. They can be used to identify functional modules and biologically relevant genes based on guilt-by-association framework. GCNs usually have been constructed using gene expression datasets generated by DNA microarrays, however the recent RNA-seq technology is rapidly replacing microarrays and allows more complete characterization of RNA transcripts. Since very few analyses have been performed on co-expression networks based on RNA-seq, it is important to infer GCNs from RNA-seq data. Moreover, GCNs from RNA-seq data can be combined with microarray-based networks to increase the robustness in meta-analysis. In this study, we collected many different datasets from NCBI GEO including 25 RNA-seq and 2,102 microarray samples derived from human brain in Alzheimer's disease and performed meta-analysis to identify functional modules responsible for the characterization of Alzheimer's disease. First, we established the GCN pipelines using in-house data performed on the same samples by both RNA-seq and microarray to reduce the artificial bias between two platforms. The GCNs were generated using Pearson Correlation Coefficient and meta-analysis was conducted using rank-based method. Then the same pipelines were applied to infer GCNs from Alzheimer's disease samples. The preliminary results show that the GCNs from microarray data provide rich molecular information to gain insight into biological processes and disease mechanism. There is low size overlap between microarray- and RNA-seq-derived GCNs however, GCNs from RNA-seq would complement ones from microarray due to the higher coverage and dynamic range of RNA-seq.	Genes poster	Health
P_Ge042	493	Yi-Wei Lee, Ting-Yu Chang, Hsei-Wei Wang, Oscar Kuang-Sheng Lee, I-Fang Chung and Shung-Haur Yang	Yi-Wei Lee	Integrated database for long non-coding RNA discovery, profiling, and annotation from RNA-sequencing data sets across cancers	Long non-coding RNAs (lncRNAs) are non-protein coding transcripts longer than 200 nucleotides. Recently, with the rapid growth of deep-sequencing technology and the development of computational prediction algorithms, a lot of lncRNAs have been identified in cancers. Therefore, the aim of this research is to identify lncRNAs by analyzing RNA-sequencing data in a clinically meaningful way, as well as to provide a cancer genomics database.We developed a user-friendly database to systematically collect a comprehensive list of lncRNAs from public databases including Ensemble, GENCODE, NONCODE, and lncRNAdb. In addition, there were >32,000 novel lncRNAs assembled from different cancer RNA-Seq data. These novel lncRNAs were filtered by considering a series of steps, such as transcripts length and coding potential score. Furthermore, we provided analysis results for the related genomic information of lncRNAs, such as cellular function and expression profiles. To investigate the association between diseases and de-regulation of lncRNAs, a straightforward query interface enables a user to find a set of potential biomarkers defined by the expression profiles and clinical outcomes.Using our database, an individual can observe the significant aberrant lncRNAs across cancers without knowing the underlying software that supports analyzing RNA-sequencing data and predicting novel transcripts. Our database simplifies visualization of the lncRNA-disease associations and was adapted for accurate selection of biomarkers by considering multiple data sources simultaneously. We believe that it will allow more efficient translation of laboratory discoveries into the clinical context, and will assist in reinterpreting the function of lncRNAs in cancer research.	Genes poster	Biotechnology
P_Ge043	480	Ping-Han Hsieh, Wen-Ting Wang, He Wang, Wei-Jhen Huang and Chien-Yu Chen	Ping-Han Hsieh	Investigating the effect of similar subsequences present in assembled transcripts on RNA-seq quantification for non-model organisms	Transcript abundance analysis based on RNA-seq has been widely adopted to study transcript expression in different physiological conditions or diseases for non-model organisms. Without reference genome or transcriptome, researchers have to perform de novo transcriptome assembly prior to expression quantification. In such situation, accurate quantification is challenging because the assemblers might produce incomplete sequences or incorrect splicing forms, which may mislead the estimation of expression quantities. This study aims to reveal the effect of similar subsequences present in the assembled sequences on the accuracy of expression quantification. To mimic the transcriptome analysis in non-model organisms, we used synthetic data of RNA-seq data generated by Bioconductor polyester. The expression intensities preset in the simulated data were used as the expected answers for performance evaluation. In this study, Bowtie2 and eXpress were used for read mapping and expression quantification, respectively. We observed that the accuracy of quantification decreases as the number of transcripts that share subsequences increases. Similar results were observed on real data where the expression abundances from RNA-seq were compared with that from microarrays for model organisms. On the other hand, for non-model organisms, qPCR data was used to evaluate the quantification accuracy. The results suggested that similar subsequences present in transcripts indeed have a strong influence on quantification accuracy. In the end, we provided practical suggestions on how the reference can be prepared in order to reduce the influence of similar subsequences on RNA-seq quantification for non-model organisms.	Genes poster	Fundamental
P_Ge044	468	Stefan Tomiuk, Jutta Kollet, Michail Knauel, Lena Willnow, Stefan Wild, Silvia Rüberg, Claudius Fridrich, Peter Mallmann, Frauke Alves, Philipp Ströbel, Dominik Eckardt, Andreas Bosio and Olaf Hardt	Stefan Tomiuk	Isolation of primary human tumor cells improves culture of target cells and reduces bias in molecular analysis	Solid tumors are infiltrated by cells of non-tumor origin, including heterogeneous lymphocyte subpopulations, fibroblasts, and endothelial cells. The amount and composition of infiltrating cells is highly variable and patient dependent, which makes analyses of primary tumor samples difficult. We have developed a fast and easy method to isolate untouched human tumor cells from primary tissue. This procedure is based on the comprehensive depletion of cells of non-tumor origin by combining automated tissue dissociation and magnetic cell sorting (MACS® Technology). Here, we have applied the method to isolate human tumor cells from primary and metastatic ovarian carcinoma, as well as thymoma specimens. The purified human ovarian carcinoma tumor cell fraction was further used for the isolation of CD133+ cancer stem cells (CSCs). We performed Whole Exome Sequencing (WES) and gene expression profiling to i) compare genomic characteristics of isolated tumor cells and unpurified samples, ii) identify tumor cell- and CSC-specific expression signatures, and iii) compare the latter expression data with that of ovarian cancer cells (GSE29450), which had been collected by laser capture microdissection (LCM).	Genes poster	Biotechnology
P_Ge045	788	Tareq Malas	Tareq Malas	Meta-analysis of Polycystic Kidney Disease expression profiles defines strong involvement of injury repair processes	Expression profiling experiments are becoming very popular in human disease study and drug discovery. Although they are useful in revealing novel insights about the disease etiology, there are several pitfalls and limitations to their use that need to be addressed. Among these limitations are the experimental and technology-related biases in the data, and the use of general gene annotation databases such as KEGG and Gene Ontology, which jeopardize the functional interpretation of the data. To overcome these limitations in the context of a study of Polycystic Kidney Disease (PKD), we completed a meta-analysis of published PKD expression profiles in combination with our in-house RNASeq study of a Pkd1-mutant mouse model. We included samples from mice, rats and patients, and from microarray and RNASeq platforms to limit experimental and technology based biases. Comparing these datasets we generated a PKD signature that consists of 960 genes, including several known PKD genes. We show the robustness of our signature by significantly distinguishing PKD from WT samples in independent datasets. To define the tissue injury and repair component of PKD, we also integrated experimental data, namely expression profiles of ischemia reperfusion injury samples, and literature data by mining PubMed abstracts for injury repair and gene/protein associations. We discovered that at least 22% of the PKD Signature genes and 40% of functions are implicated in injury repair processes, supporting the hypothesis that PKD is a state of chaotic renal repair.	Genes poster	Health
P_Ge046	691	Alexandra Poos, Andre Maicher, Anna Dieckmann, Marcus Oswald, Roland Eils, Martin Kupiec, Brian Luke and Rainer König	Rainer König	Mixed Integer Linear Programming based machine learning approach identifies regulators of telomerase in yeast	Understanding telomere length maintenance mechanisms is central in cancer biology as their dysregulation is one of the hallmarks for immortalization of cancer cells. Important for this well-balanced control is the transcriptional regulation of the telomerase genes. We integrated mixed integer linear programming models into a comparative machine learning based approach to identify regulatory interactions that best explain the discrepancy of telomerase transcript levels in yeast mutants with deleted regulators showing aberrant telomere length, when compared to mutants with normal telomere length. We uncover novel regulators of telomerase expression, several of which affect histone levels or modifications. In particular, our results point to the transcription factors Sum1, Hst1 and Srb2 as being important for the regulation of EST1 transcription, and we validated the effect of Sum1 experimentally. We compiled our machine learning method leading to a user friendly package for R which can straightforwardly be applied to similar problems integrating gene regulator binding information and expression profiles of samples of, e.g. different phenotypes, diseases or treatments.	Genes poster	Health
P_Ge047	716	Luca Santuari, Gabino F. Sanchez-Perez, Bas Rutjens, Lidija Berke, Viola Willemsen, Berend Snel, Kenzo Nakamura, Dick de Ridder, Ben Scheres and Renze Heidstra	Luca Santuari	Partitioning of PLETHORA target expression domains guides cell differentiation	Organ formation in animals and plants relies on precise control of cell state transitions to turn stem cell daughters into fully differentiated cells. In plants, cells cannot rearrange due to shared cell walls. Thus, differentiation progression and the accompanying cell expansion must be tightly coordinated. PLETHORA (PLT) transcription factor gradients were shown to guide the progression of cell differentiation at different positions in the Arabidopsis root. While well-describedtranscription factor gradients in animals specify distinct cell fates within an essentially static context, the PLT gradient is unique in its ability to control cell differentiation in a growing organ during continuous production and expansion of cells. To understand the output of their gradients we studied the gene set transcriptionally controlled by PLTs. Our work reveals how the PLT gradient regulates cell state by region-specific induction of cell proliferation genes and repression ofdifferentiation. Moreover, PLT targets include major patterning genes and autoregulatory feedback components, enforcing their role as master regulators of organ development.	Genes poster	Fundamental
P_Ge048	563	Tuomo Hartonen, Biswajyoti Sahu, Kashyap Dave, Teemu Kivioja and Jussi Taipale	Tuomo Hartonen	PeakXus: A Comprehensive Peak Calling Software for ChIP-Nexus and ChIP-exo	Novel chromatin immunoprecipitation (ChIP) experiments ChIP-Nexus [1] and ChIP-exo [2] allow studying transcription factor (TF) binding with unprecedented accuracy. True TF binding locations are separated from noise by peak calling softwares.Most peak calling softwares search binding events by creating a model of "true" peaks from the sites with highest enrichment in the ChIP-experiments and then accepting only the peaks resembling this model. It is however known that most TFs bind cooperatively with other TFs, form dimers or interact with other proteins. These different types of binding create different ChIP-Nexus/exo fingerprints. Fitting the peaks to just one model may lead to missing important binding events.PeakXus is a peak caller specifically designed to leverage the increased resolution of ChIP-Nexus/exo experiments. PeakXus is developed with the aim of making as few assumptions of the data as possible to allow novel discoveries. PeakXus supports use of Unique Molecular Identifiers (UMI) [3] to remove PCR-duplicates that can create artefacts closely resembling true ChIP-Nexus/exo binding events. We show that PeakXus consistently finds more peaks overlapping with TF-specific recognition sequences than published methods. PeakXus is available at https://github.com/hartonen/PeakXus.[1] He et al. (2015). ChIP-Nexus enables improved detection of in vivo transcription factor binding footprints. Nature biotechnology.[2] Rhee & Pugh (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell.[3] Kivioja et al. (2012). Counting absolute numbers of molecules using unique molecular identifiers. Nature methods.	Genes poster	Fundamental
P_Ge050	496	Mei-Ju May Chen, Yu-Rui Su, Ping Chang, Tai-Rong Hong, Bor-Wei Cherng,Yi-An Tung and Chien-Yu Chen	Chien-Yu Chen	Potential of lncRNA to regulate gene expression through promoter binding in Drosophila Melanogaster	Recent studies have revealed that a novel factor, long non-coding RNA (lncRNA), may also be a key player in gene regulation. However, it remains unclear for most of lncRNAs on how they regulate gene expression. In this regard, this study aims at investigating whether lncRNAs affect gene expression through binding to gene promoters by exploiting sequence reverse complementary. Here, we examined the possibility of this scenario in Drosophila melanogaster. A set of 4,599 fly lncRNAs was collected from FlyBase and recent studies. To identify promoters that might be bound by lncRNAs, we first adopted BLASTn to align lncRNA sequences to the promoter sequences of mRNAs. An lncRNA was reported to have potential of binding promoters if the number of the qualified alignments in promoter regions was significantly higher than that in the whole genome. We proposed that a high binding enrichment score indicates that a lncRNA might regulate some genes through binding to their promoters. The results revealed that 1,070 lncRNA-gene pairs (involving 82 lncRNAs and 410 promoters) were shown with binding potential owing to sequence reverse complementary. We further utilized the developmental transcriptome of D. melanogaster (Nature 471:473-9, 2011) to see whether the expression of these lncRNA-gene pairs were correlated. The analyses showed that the identified lncRNA-gene pairs have significantly higher correlated expression than random pairs. In summary, this study presented that lncRNAs might regulate gene expression through sequence reverse complementary with promoters, and suggested the potential of lncRNA to regulate gene expression through promoter binding.	Genes poster	Fundamental
P_Ge051	388	Christian Groß, Marcel Reinders, Dick De Ridder, Martijn Derks, Mirte Bosse, Hendrik-Jan Megens and Martien Groenen	Christian Groß	Predicting the impact of genetic variation in livestock	In recent years, advancements in functional effect prediction of variants in human genomes have led to several new discoveries and insights in heritable diseases. Methods such as CADD or Eigen incorporate various forms of variant annotation information to compute one generic score of deleteriousness for every DNA sequence variant. Currently, these methods are solely available for research of human genomes. The goal of this project is to develop methods for gene variant evaluation for livestock i.e. poultry, pig and cattle. This would open up the possibility for new approaches to adjust breeding schedules with the aim to achieve breeding goals without accumulating negative inbreeding side-effects. This would increase the overall health of livestock populations and help to reduce unnecessary suffering in animal farming.Numerous research groups are working with livestock genome data but epigenetic information and annotation lag behind, compared to data which is available for human or model organisms like mouse. With this in mind we first conducted a feasibility study by developing a method for sequence variant evaluation in mouse, based on human epigenetic data. By focusing on mouse data we are able to validate the possibility of transferring annotations from highly conserved regions in the human genome to non-human species.	Genes poster	Agro-Food
P_Ge052	646	Jairo Rocha, Jaume Sastre Tomas and Emidio Capriotti	Jairo Rocha	Ranking Putative Cancer Driver Gene Subsets	We develop a score for some subsets of genes that represents the possibility that this subset be associated with a specific type of cancer. The score depends on the correlation of SNP appearance on normal samples with respect to the same correlation on tumour samples. The normal samples include the genomic data from the 1000Genome Project. The tumour samples could be from different types of cancer (lung, colon and prostate cancers) from the TCGA (The Cancer Genome Atlas Consortium). This is the first time that all possible gene pairs (around 20 million) would be considered. The list of pairs most likely related to each type of cancer would be published. Each pair could be a target to be studied deeply by animal models and future therapeutic targets. The genes in a pair with high score should be treated simultaneously as possible cancer drivers. The score can be used to evaluate patients individually.The work carried out by Dr. Emidio Capriotti and other authors who have published it in September 2014 (Bioinformatics) describes a method to assign a score to each gene in the entire human genome and represents the possibility that the gene is associated with a type of cancer (this study used samples of lung, colon and prostate). There are multiple gene candidates but candidate pairs and subsets could be fewer and revealing.Some results are shown as promising.	Genes poster	Health
P_Ge053	481	Audrey Michel, James P. A. Mullan, Stephen Kiniry, Vimalkumar Velayudhan, Patrick B. F. O'Connor and Pavel Baranov	Audrey Michel	RiboSeq.Org for ribosome profiling data analysis and visualisation.	The ribosome profiling (ribo-seq) technique uses high-throughput sequencing to provide Genome Wide Information on Protein Synthesis (GWIPS) by revealing the locations and densities of actively translating ribosomes at a genome-wide level. On RiboSeq.Org (http://riboseq.org/) we provide freely available resources to help researchers analyse and explore ribo-seq data without having to use command-line tools.GWIPS-viz is an online genome browser which hosts over 1000 pre-populated ribo-seq and corresponding mRNA-seq tracks across 20 genomes generated from data from over 70 published studies, thereby enabling cross-study and cross-species comparisons.RiboGalaxy is a Galaxy-based web server where researchers can pre-process, align, analyse and visualize their ribo-seq data. GUI-based tools are provided to determine the strength of the triplet periodicity signal in ribo-seq data, generate metagene and ribosome profiles and carry out differential translation expression analysis using riboSeqR. The RUST suite of tools can be used to quickly characterise ribosome profiling datasets to assess their quality as well as analyze the relative impact of mRNA sequence features on local decoding rates. The RiboTools suite provides functionality for exploring translation in alternative reading frames and stop codon readthrough events. As well as help pages, we provide forums on both GWIPS-viz and RiboGalaxy usage (http://gwips.ucc.ie/Forum/).	Genes poster	Fundamental

P_Ge054	378	Kerem Wainer Katsir and Michal Linial	Kerem Wainer Katsir	Single Cell Expression Data as a Direct Measure for Identifying Human Genes that Escape X-Inactivation	Sex chromosomes pose an inherent genetic imbalance between genders. In mammals, one of the female's X-chromosomes undergoes inactivation (Xi). Indirect measurements estimate 15-25% of Xi genes to completely or partially escape inactivation. The identity of these escaper genes, and their propensity for escape remain unsolved. We applied a direct method to identify escapees based on RNA-Seq from 25 single-cell lymphoblasts and a pooled version. We quantified the differential allelic expression by assigning reads from expressed genes to SNPs with distinct maternal or paternal identities. We confirmed that X-inactivation occurs and is maintained in single cells. Using strict and relaxed protocols, we confidently identified 27 and 35 escaper genes, respectively. Using 30 published datasets, we compiled a genes' catalogue characterized as escapees or inhibited along with a confidence value. The nature of most reported genes (454 in total) as escapees and inhibited is mixed across many biological contexts. We report a strong statistical overlap between escapees identified from single cells and those reported in the literature-based catalogue. We confirmed the usefulness of single cells' expression data for studying allelic bias phenomena. We conclude that escaping X-inactivation is less deterministic than previously reported with only few genes acting as exclusive escapees.	Genes poster	Fundamental
P_Ge055	479	Volodimir Olexiouk, Steven Verbruggen, Jeroen Crappé, Kenneth Verheggen,Lennart Martens and Gerben Menschaert	Volodimir Olexiouk	sORFs.org: a repository of small ORFs identified by ribosome profiling	Micropeptides, defined as translation products from small open reading frames sORFs (<300nt) are becoming widely recognized. This is also demonstrated by recent characterisation of several members of this new group of bio-active players: Toddler, Pri-peptides, Sarcolipin and Myoregulin (Pauli et. al., Science, 2014; Chanut-Delalande et. al., Nat. Cell Biol., 2014; Magny et. al., Science, 2013; Anderson et. al., Cell, 2015). Ribosome profiling, a NGS-technique measuring translation synthesis, enabled the identification of numerous sORFs demonstrating ribosomal occupancy (Ingolia et. al., Science, 2009 and Cell, 2014). Historically, sORFs have been neglected and their discovery could thus potentially provide new and important biological insights. Means to distribute this knowledge are necessary. Our public repository, sORFs.org (Olexiouk et. al., Nucleic Acids Research, 2015), currently holds 263.496 sORFs identified using ribosome profiling, from three model species (human, mouse, fruit fly). Furthermore, sORFs.org includes various tool and metrics to assess the coding potential of sORFs at multiple levels: ribosome profiling analytics, genomic information, experimental information, visualization of data, dataset information and sORF-specific calculated metrics trying to determine their coding potential in different ways (FLOSS, ORFscore, variation analysis, PhyloCSF conservation score) as described by recent literature. sORFs.org provides researchers an easy way to inspect and query sORFs information, facilitating the integration of sORFs into in-house research projects. Moreover, the PRIDE-respin pipeline enables the automatic rescan of MS fragmentation spectra stored in PRIDE using Pladipus (Verheggen et. al., journal of proteome research, 2015), acquiring proteomic evidence for sORFs.	Genes poster	Fundamental
P_Ge056	387	Sivan Gershanov, Shalom Michowiz, Helen Toledano, Orit Barinfeld, Albert Pinhasov, Nitza Goldenberg-Cohen and Mali Salmon-Divon	Sivan Gershanov	Subgrouping of pediatric medulloblastoma using an integrated analysis of MicroRNA-mRNA expression profile	Medulloblastoma (MB), the commonest malignant brain tumor of childhood, is divided into four tumor subgroups representing distinct molecular entities. Subsequently, treatment should be designed according to the specific subgroup. MicroRNAs (miRNAs) are involved in carcinogenesis and tumor progression by regulating post-transcriptional gene expression. However, the miRNA-mRNA regulatory network in MB is far from being fully understood. The aim of the study is to identify novel miRNA subgroup biomarkers for specific diagnosis by analyzing integrated mRNA-miRNA MB transcriptome sequencing. With this aim, integrated whole transcriptome mRNA and miRNA expression analysis was performed on primary tumor samples collected from 10 MB patients. 867 mature miRNAs were identified in at least a single sample, of them 462 were common to all 4 subgroups. 25 (2.5%) of all expressed miRNAs appeared to be significantly differentially expressed between the subtypes. Namely, upregulation of hsa-miR-224-5p and hsa-miR-449c-5p was found exclusively among WNT, while downregulation of hsa-miR-135b-5p characterized SHH. Among groups 3 and 4, hsa-miR-20a-5p was upregulated or downregulated, respectively. RNA-seq from the same tumor samples identified 500 genes that vary between the four subtypes, among which 69 (13.8%) have anti-correlated miRNA-mRNA interactions with the 25 detected miRNA biomarkers. The predicted mRNAs targets of these miRNAs are associated with different signaling pathways, known to have a role in MB biology. Our study demonstrates that miRNAs are highly specific to distinct MB subgroups. Understanding the involvement of miRNAs and their targets in MB related signaling pathways may improve diagnosis and advance the development of targeted treatment for MB.	Genes poster	Health
P_Ge057	825	Joana P. Gonçalves, Jeroen de Ridder and Lodewyk F.A. Wessels	Joana P. Gonçalves	Temporally-aware discovery of regulatory cascades	Temporal transcriptomes expose dynamics of gene regulation and disruptions leading to disease. Many studies uncover functional units by grouping genes with similar transcriptional responses and linking them to transcriptional regulators. Gene grouping is typically obtained through differential expression or (bi)clustering, while regulators are predicted by direct target enrichment based on protein-DNA binding or regulator-target co-expression.Differential expression scores camouflage distinct variations over time. Clustering maintains chronology, focusing on global patterns often associated with broad biological functions. Biclustering achieves increased granularity from locality, but also generates patterns with arbitrary time gaps. Target enrichment ignores joint regulatory effects and co-expression likely captures targets with upstream co-regulation rather than regulator-target relationships.We propose a method that groups coordinated genes based on temporal phenomena: biological tasks may span shorter periods than the experiment, and participating genes likely coordinate mostly in that time; genes are involved in multiple tasks with different partners; some genes exhibit correlated profiles with delays induced by different response times and/or transcriptional cascades. Additionally, we predict regulators from curated regulator-target interactions exploring multi-layered paths without co-expression assumptions.We analysed androgen response in LNCaP cells. Our method recovered prostate cancer genes more effectively than traditional approaches. Identified regulatory units accurately characterised known pathways affected by androgen response, including cell proliferation, lipid metabolism, and unfolded protein response. Notably, delays and co-expression-free regulator prediction enabled discovery of time-shifted targets and inhibitory regulations, respectively, which would be missed otherwise. We validated predictions on the regulation of gene groups using public and in-house experimental data.	Genes poster	Health
P_Ge058	250	David Holloway and Alexander Spirov	David Holloway	TRANSCRIPTIONAL BURSTING IN DROSOPHILA DEVELOPMENT: STOCHASTIC DYNAMICS OF PAIR-RULE EXPRESSION	Segmentation of the anterior-posterior (AP) axis of the fruit fly (Drosophila) is first seen in the striped expression patterns of the pair-rule genes, well before the physical appearance of body segments. even-skipped (eve) is one of the best-studied pair-rule genes, forming 7 expression stripes orthogonal to the AP axis, which in turn regulate downstream genes involved in determining unique cell fates for each segment. Transcriptional control specific to particular stripe locations was first shown with eve: a 1.7 kb enhancer upstream of the coding region is sufficient to drive reporter expression in the 2nd eve stripe position (42 %EL, percent egg length). Recent live imaging of an eve stripe 2 reporter has demonstrated the stochastic nature of pair-rule gene expression. We have developed a stochastic model of eve stripe 2 expression, including binding of the enhancer by upstream transcriptional regulators and the initiation and completion of transcript elongation. All parameters in the model are constrained by experimental data. Simulations allow us to test different regulatory possibilities. A simple on-off model for transcriptional initiation does not fit the experimental time series for the stripe centre, indicating that eve has multiple 'on' rates for transcriptional initiation.	Genes poster	Fundamental
P_Ge059	658	Nick Dimonaco, Robert Hoehndorf and Amanda Clare	Nick Dimonaco	Using Gene Ontology annotations to understand lethality phenotypes	Online databases such as FlyBase provide information regarding the genes of model organisms such as Drosophila melanogaster, including a near complete set of gene disruption phenotypes. In most cases, genes contained in these databases are annotated using the Gene Ontology (GO), which provides information about the molecular function, cellular component and biological process. Here, we use these annotations to train a machine learning algorithm that can be used to identify combinations of GO features that lead to accurate and informative predictions for gene disruption phenotypes.The databases of C. elegans, D. melanogaster, M. musculus, S. cerevisiae and D. rerio were queried for genes associated with lethal or viable phenotype classifications. The available annotated genes were then filtered to remove those with phenotypes corresponding to: conditionally lethal, produced by multiple disruptions, allele-specific or not fully characterised. The remaining genes were then categorised into two subsets per organism; a subset of genes characterised as lethal and a subset characterised as viable. The GO terms associated with these two subsets were used to train a decision tree machine learning algorithm within Weka. We also investigated the over-representation of GO terms within these lethal and viable classes.Our results clearly demonstrate that GO terms can be used to successfully describe and predict lethal and viable phenotypes in model organisms. We discuss the causes of lethality and the variation that we found across the five species.	Genes poster	Fundamental
P_Ge060	401	Deepak Karthik, Gil Stielzer, Sivan Gershanov, Danny Baranes and Mali Salmon-Divon	Deepak Karthik	Utilizing the Benford law for unravelling tissue specificity	The reduction in sequencing costs has led to an unprecedented trove of gene expression data from diverse biological systems. Subsequently, principles from other disciplines such as the Benford law, which can be properly judged only in data-rich systems, can now be examined on this high-throughput transcriptomic information. The Benford law states that in numerical data, the proportion of numbers beginning in any given digit is not uniform but rather skewed, with 1 being the most common digit and 9 the rarest. Here we demonstrate that digital gene expression data has a Benford-like distribution when observing an entire gene set. This phenomenon was conserved in a wide range of biological tissues and developmental conditions. However, when obedience to the Benford law is calculated for individual expressed genes across thousands of cells, genes that best and least adhere to the law are enriched with tissue specific or cell maintenance descriptors, respectively. Surprisingly, a positive correlation was found between the obedience a gene exhibits to the Benford law and its expression level, despite the former being calculated solely according to first digit frequency while totally ignoring the expression value itself. These results demonstrate the applicability and potential predictability of the Benford law for gleaning biological insight from simple count data.	Genes poster	Fundamental
P_Ge061	665	Djordje Djordjevic, Kenro Kusumi and Joshua Ho	Djordje Djordjevic	XGSA: A statistical method for cross-species gene set analysis	Gene set analysis is a powerful tool for determining whether an experimentally derived set of genes is statistically significantly enriched for genes in other pre-defined gene sets, such as known pathways, gene ontology terms, or other experimentally derived gene sets. Current gene set analysis methods do not facilitate comparing gene sets from different organisms as they do not explicitly deal with homology mapping between species. There lacks a systematic investigation about the effect of complex gene homology on cross-species gene set analysis.In this work, we show that not accounting for the complex homology structure when comparing gene sets from two species can lead to false positive discoveries, especially when comparing gene sets that have complex gene homology relationships. To overcome this bias, we propose a straightforward statistical approach, called XGSA, that explicitly takes the cross-species homology mapping into consideration when doing gene set analysis. Simulation experiments confirm that XGSA can avoid false positive discoveries, while maintaining good statistical power compared to other ad hoc approaches forcross-species gene set analysis. We further demonstrate the effectiveness of XGSA with two real-life case studies that aim to discover conserved or species-specific molecular pathways involved in social challenge and vertebrate appendage regeneration.	Genes poster	Application Fundamental
P_Ge062	783	Joske Ubels, Erik van Beers, Pieter Sonneveld, Martin van Vliet and Jeroen de Ridder	Joske Ubels	zPFS: a method to identify gene expression signatures to predict treatment specific survival in cancer	Cancer treatments may have heterogeneous response rates. Patient perspectives such as adverse treatment-related events and survival may be improved by selecting the right treatment at diagnosis. This is a major challenge that requires identification of biomarkers, such as a gene expression signature, based on which the best treatment regime can be determined.Here, we propose a new computational method to identify gene expression signatures that predict if a patient is likely to survive longer when receiving a specific treatment as compared to an alternative treatment. Our algorithm exploits tumor cell gene expression data from phase 3 clinical trials in which patients were randomly assigned to the treatment of interest or another treatment.Our method hinges on the notion that potential signature genes and gene sets can be identified by searching for patients receiving different treatments that have a large difference in survival while exhibiting similar gene expression profiles. To identify these we introduce zPFS, a measure for how much larger than expected this survival difference is for a specific patient. This zPFS measure enables identification of signature gene sets and exemplar patients that can be used to predict treatment specific survival.We demonstrate the utility of our method in a multiple myeloma dataset, where patients either received the proteasome inhibitor bortezomib or not. We find fourteen GO categories that can identify patient groups, comprising at minimum 20% of the patients, that have at least a 2-fold lower risk of experiencing an event (p-value < 0.05) when receiving bortezomib.	Genes poster	Health