

- Clauser KR, Wang P, et al. Proteogenomics connects somatic mutations to signaling in breast cancer. *Nature* 2016;534(7605):55-62.
15. Şenbabaoğlu Y, Sümer SO, Sánchez-Vega F, Bemis D, Ciriello G, N. Schultz N, et al. A Multi-Method Approach for Proteomic Network Inference in 11 Human Cancers. *PLoS Comput Biol* 2016;12(2):e1004765.
16. Zhao J, Cheng F, Wang Y, Arteaga CL, Zhao Z. Systematic Prioritization of Druggable Mutations in ~5000 Genomes Across 16 Cancer Types Using a Structural Genomics-based Approach. *Mol Cell Proteomics* 2016;15(2):642-56.
17. Lu Q, Yao X, Hu Y, Zhao H. GenoWAP: GWAS signal prioritization through integrated analysis of genomic functional annotation. *Bioinformatics* 2016;32(4):542-8.
18. Li JR, Sun CH, Li W, Chao RF, Huang CC, Zhou XJ, et al. Cancer RNA-Seq Nexus: a database of phenotype-specific transcriptome profiling in cancer cells. *Nucleic Acids Res* 2016;44(D1):D944-D951.
19. Weitzel KW, Alexander M, Bernhardt BA, Calman N, Carey DJ, Cavallari LH, et al. The IGNITE network: a model for genomic medicine implementation and research. *BMC Med Genomics* 2016;9:1.
20. Global Alliance for Genomics and Health. GENOMICS. A federated ecosystem for sharing genomic, clinical data. *Science* 2016;352(6291):1278-80.

Correspondance to:

Dr Hélène Dauchel
 Normandy University, UNIROUEN, LITIS EA 4108
 76821 Mont-Saint-Aignan Cedex, France
 Tel : +33 235 146 389
 E-mail: Helene.Dauchel@univ-rouen.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2017, Section Bioinformatics and Translational Informatics

Chen J, Rozowsky J, Galeev TR, Harmanci A, Kitchen R, Bedford J, Abyzov A, Kong Y, Regan L, Gerstein M

A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals

Nat Commun 2016 Apr 18;7:11101

In this article, existing large and diverse collections of individual genomes from the 1000 genomes Project, RNA-seq and ChIP-

seq data sets were unified to build a comprehensive data corpus used to detect and functionally annotate allele-specific single nucleotide variants (SNVs) with allelic functional imbalance. The authors considered 1,263 functional genomics data sets from eight different studies to annotate variants associated with allele-specific binding and expression in 382 individuals consisting of 993 RNA-seq and 287 ChIP-seq data for coding and non-coding regions respectively. For each individual, the authors first built a diploid personal genome using the variants from the 1000 Genomes Project. Then expression data was mapped onto each of the haplotypes of the diploid genome, instead of the human reference genome. Results were then filtered to correct overdispersion and mapping bias, and finally enrichment analyses were performed thanks to a beta-binomial test to identify genomic regions that were enriched or depleted in allelic activity. Inheritance of allele-specific behavior was detected in autosomal protein-coding genes, untranslated regions (UTRs), introns and enhancers, and transcription factor (TF)-binding regions. Furthermore, considering the enrichment of rare variants, the authors examined selective constraints in allele-specific SNVs in coding DNA sections regions and TF motifs. The final data and results were organized into a distributed resource called AlleleDB that can be directly visualized as a UCSC (University of California, Santa Cruz) track in the UCSC Genome browser.

Marbach D, Lamarter D, Quon G, Kellis M, Kutalik Z, Bergmann S

Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases

Nat Methods 2016 Apr;13(4):366-70

Re-using previous 37 genome-wide association studies (GWASs) data for complex traits and diseases in the light of information supplied by diverse sequence data (CAGE-defined enhancers and promoters, ChIP-seq, and RNA-seq data) and expression quantitative trait loci (eQTL), the authors inferred and validated transcriptional regulatory circuits and the connectivity between trait-associated genes for 394 cell types

or tissue-specific regulatory networks for human. Their integrative pipeline and network connectivity enrichment revealed that GWASs variants associated with specific diseases have impact on regulatory modules that are specific to disease-relevant cell types or tissues. All networks are freely available and they allow the systematic analysis of regulatory programs across hundreds of human cell types and tissues.

Zhang D, Chen P, Zheng CH, Xia J

Identification of ovarian cancer subtype-specific network modules and candidate drivers through an integrative genomics approach

Oncotarget 2016 Jan 26;7(4):4298-309

The identification of cancer subtypes is required to understand cancer heterogeneity and to propose the personalized therapy treatment appropriate to the different subtypes. In this study, the authors re-used large-scale ovarian cancer genomic data, including micro-array data (mRNA and microRNA expressions), SNP-array (copy number variations) and protein-protein interactions data in order to build a novel integrative procedure for defining ovarian cancer subtypes, identifying core pathways and candidate driver genes for each subtype. By applying a similarity network fusion approach to a patient cohort with 379 ovarian cancers from The Cancer Genome Atlas (TCGA) cancer samples, the authors were able to discover subnetworks enriched with genetic alterations. They identified two clinically relevant ovarian cancer subtypes with distinct molecular and clinical phenotypes and different survival profiles. Enrichment analysis of pathways associated with the two ovarian cancer subtype-specific networks revealed distinct molecular mechanisms of the tumorigenesis that could explain the different clinical outcomes.

Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, Wilson RK, Maher CA

INTEGRATE: gene fusion discovery using whole genome and transcriptome data

Genome Res 2016;26(1):108-18

Among somatic aberrations in cancer genome, gene fusions are the most prevalent chromosomal rearrangements. Especially in solid tumors, their detection can serve as specific diagnostic markers, prognostic indicators, and therapeutic targets. Mono-modal data tools (structural variations with whole genome sequencing (WGS) or RNA-seq expression data) suffer from variability between fusion callers and from a poor sensitivity and specificity of fusion detection. In this article, the authors developed a new gene fusion discovery method that integrates both whole genome and transcriptome sequencing data from the same patient to reconstruct

gene fusion junctions and genomic breakpoints by split-read mapping. INTEGRATE first utilizes mapped and unmapped RNA-seq reads, then analyzes WGS reads from tumors, and if available, from normal samples. INTEGRATE uses discordant RNA-seq reads to construct a gene fusion graph connecting genes involved in a putative fusion event. It finally proposes a prioritization of gene fusion candidates. INTEGRATE was evaluated by comparison to eight other gene fusion discovery tools by reusing data from a previously studied breast cancer cell line and peripheral blood lymphocytes derived from the same patient leading. INTEGRATE was

also applied to a cohort of 62 breast cancer patients from The Cancer Genome Atlas (TCGA) and enabled the identification of novel gene fusions, a subset of which were recurrent. All together, by combining WGS and RNA-seq NGS data from a same patient, the authors demonstrated both high sensitivity and accuracy of INTEGRATE to detect novel causative mutations. Furthermore, unlike many gene fusion prediction tools that ignore read-through or trans-splicing events, INTEGRATE was able to provide valuable insight into RNA chimeras. The tool is freely available for academic use.