# ROCK: a resource for integrative breast cancer data analysis

**Saif Ur-Rehman[1], Costas Mitsopoulos[1], Marketa Zvelebil[1]**✉

[1]Breakthrough Breast Cancer Research, Institute of Cancer Research, London, United Kingdom

## Motivation and Objectives

There is currently an ongoing research effort to understand the molecular mechanisms underpinning breast cancer being undertaken worldwide. The generated data measure different molecular characteristics that underlie the cancer phenotype. We present the updated ROCK (ROCK Online Cancer Knowledgebase) database, which now integrates these diverse data types allowing unique analyses of published breast cancer experimental data. The primary usage of the updated version of ROCK is to allow bench scientists to query the state of a given gene/genomic locus within a subset of samples defined by clinical annotation terms.

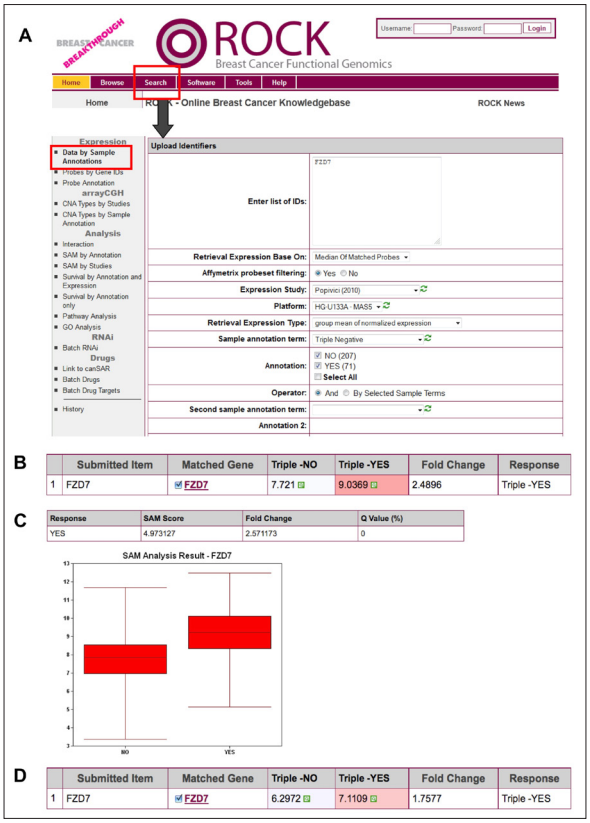


Figure 1: FZD7 differential expression in triple negative breast cancers. (A) ROCK interface for exploring gene expression fold changes. (B) Triple negative response for FZD7 in the Popovici dataset (Popovici et al 2010). (C) Significance of microarrays (SAM) (Tusher et al 2010) analysis of (B) performed in ROCK. The "SAM by annotation" option is under the Analysis header in the ROCK Search menu. (D) Triple negative response for FZD7 across amalgamated gene expression studies.

## Methods

ROCK is implemented through the manual curation of publically available datasets containing data types such as gene expression, genomic copy number aberrations, microRNA expression, RNA interference, protein-protein interactions, and signalling protein networks. These data types are integrated using standard gene identifiers from Ensembl (Flicek et al 2011). It also utilises clinical annotation of samples in order to facilitate comparison of these data types between disease subtypes as illustrated in Figure 1. The new version of ROCK allows sub-setting of data in a multitude of clinical categories.

ROCK now contains an amalgamation of gene expression studies as measured by microarray across multiple platforms. This study is a meta-analysis of breast carcinoma samples from a number of public datasets, which have been normalised together (Boersma et al 2008; Desmedt et al 2007; Farmer et al 2005; Graham et al 2010; Loi et al 2008; Minn et al 2005; Pawitan et al 2005; Popovici et al 2010; Schmidt et al 2008; Sotiriou et al 2006; Wang et al 2005). It also implements a novel confidence score (CS) for binary protein-protein interactions, which is similar to the score, used by IntAct (Kerrien et al 2012), but also utilises structural parameters where available.

In addition to this ROCK allows survival analysis (using Kaplan-Meier plots) on both gene expression levels as well as the clinical features of a given tumour.

The system is implemented using a three-tier web architecture. Web browsers communicate with the interface using Javascript via Apache Tomcat. Server-side data requests are then primarily handled using a set of core Java classes that communicate with the backend relational database. This database is implemented using Oracle.t

## Results and Discussion

ROCK provides a unique breast cancer analysis platform of integrated experimental datasets at the genomic, transcriptomic, and proteomic level. ROCK is available on rock.icr.ac.uk.

The data integration within ROCK allows for bespoke analyses to be carried out on a given data type, the results of which can be projected onto a different data type, e.g. a set of genes detected as differentially expressed between cancer subtypes can then be subjected to an analysis of the interactions of their resultant proteins. This allows a user to carry out an iterative analysis where the results of a preliminary investigation can be used as the input to a series of successive investigations. ROCK is used by an average of 2329 unique users annually.

We will present the recent and major functional updates and enhancements to the ROCK resource, including new analysis modules and microRNA and NGS data integration, and illustrate how ROCK can be used to confirm known experimental results as well as generate novel leads and new experimental hypotheses using the Wnt signalling cell surface receptor FZD7 and the Myc oncogene.

## References


1. Boersma BJ, Reimers M, Yi M et al (2008) A Stromal Gene Signature Associated with Inflammatory Breast Cancer. Int J Cancer 122:1324
2. Desmedt C, Piette F, Loi S et al (2007) Strong Time Dependence of the 76-Gene Prognostic Signature for Node-Negative Breast Cancer Patients in the Transbig Multicenter Independent Validation Series. Clin Cancer Res 13:3207
3. Farmer P, Bonnefoi H, Becette V et al (2005) Identification of Molecular Apocrine Breast Tumours by Microarray Analysis. Oncogene 24:4660
4. Flicek P, Amode MR, Barrell D et al (2011) Ensembl 2011. Nucleic Acids Research 39:D800
5. Graham K, de las Morenas A, Tripathi A et al (2010) Gene Expression in Histologically Normal Epithelium from Breast Cancer Patients and from Cancer-Free Prophylactic Mastectomy Patients Shares a Similar Profile. Br J Cancer 102:1284
6. Kerrien S, Aranda B, Breuza L et al (2012) The Intact Molecular Interaction Database in 2012. Nucleic Acids Research 40:D841
7. Loi S, Haibe-Kains B, Desmedt C, Wirapati P et al (2008) Predicting Prognosis Using Molecular Profiling in Estrogen Receptor-Positive Breast Cancer Treated with Tamoxifen. BMC Genomics 9:239
8. Minn AJ, Gupta GP, Siegel PM et al (2005) Genes That Mediate Breast Cancer Metastasis to Lung. Nature 436:518
9. Pawitan Y, Bjohle J, Amler L et al (2005) Gene Expression Profiling Spares Early Breast Cancer Patients from Adjuvant Therapy: Derived and Validated in Two Population-Based Cohorts. Breast Cancer Res 7:R953
10. Popovici V, Chen W, Gallas BG et al (2010) Effect of Training-Sample Size and Classification Difficulty on the Accuracy of Genomic Predictors. Breast Cancer Res 12:R5
11. Schmidt M, Bohm D, von Torne C, Steiner E et al (2008) The Humoral Immune System Has a Key Prognostic Impact in Node-Negative Breast Cancer. Cancer Res 68:5405
12. Sotiriou C, Wirapati P, Loi S et al (2006) Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis. J Natl Cancer Inst 98:262
13. Tusher VG, Tibshirani R, Chu G (2001) Significance Analysis of Microarrays Applied to the Ionizing Radiation Response. Proc Natl Acad Sci U S A 98:5116
14. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM et al (2005) Gene-Expression Profiles to Predict Distant Metastasis of Lymph-Node-Negative Primary Breast Cancer. Lancet 365:671