



Public datasets and tools

Christine Desmedt
(christine.desmedt@bordet.be)

Pre-IMPAKT Training Course 2015



ULB



IMPAKT
BREAST CANCER CONFERENCE



(FREE) Public datasets and tools

Christine Desmedt
(christine.desmedt@bordet.be)

Pre-IMPAKT Training Course 2014



ULB



Outline

1. Clinical tools
2. Research databases
3. Research tools

1. Clinical Tools

Clinical tools

- General info:
 - Cancer.gov
- Clinical trials:
 - Clinicaltrials.gov
- Adjuvant treatment:
 - Adjuvant! Online

Cancer.gov

National Cancer Institute
at the National Institutes of Health

We Can Answer Your Questions
1-800-4-CANCER

 SEARCH[NCI Home](#)[Cancer Topics](#)[Clinical Trials](#)[Cancer Statistics](#)[Research & Funding](#)[News](#)[About NCI](#)

Find a Cancer Type

A B C D E F G H I
J K L M N O P Q R
S T U V W X Y Z

NCI Vision & Priorities



Read Harold Varmus's Interview with Science

[NCI Director Harold E. Varmus, M.D.](#)

[Provocative Questions Project](#)
Explore provocative research questions with the research community

[The National Cancer Program](#)
2013 Annual Plan and Budget Proposal

[NCI Budget Overview](#)
Explore how NCI allocates appropriated funds

Popular Resources

[NCI Dictionary of Cancer Terms](#)
Cancer-related terms

[NCI Drug Dictionary](#)
Definitions, names, and links

[NCI Publications](#)
Order/download free booklets

[Funding Opportunities](#)
Research and training



Better Research Through Team Tools

Researchers are using team-based approaches to address complex scientific challenges like cancer. If you manage, support, or conduct team-based research, NCI's Team Science Toolkit offers a growing collection of more than 900 resources to help maximize your work.

- [What is team science?](#)
- [Team Science Toolkit](#)



[Changes to NCI's Clinical Trials Programs](#)



[Reducing Cancer Health Disparities](#)



[Better Research Through Team Tools](#)

Types of Cancer

Common Cancer Types

[Bladder Cancer](#)
[Breast Cancer](#)
[Colon and Rectal Cancer](#)
[Endometrial Cancer](#)
[Kidney \(Renal Cell\) Cancer](#)
[Leukemia](#)
[Lung Cancer](#)
[Melanoma](#)
[Non-Hodgkin Lymphoma](#)

[Pancreatic Cancer](#)
[Prostate Cancer](#)
[Thyroid Cancer](#)

All Cancer Types

[Cancers by Body Location/System](#)
[Childhood Cancers](#)
[Adolescents and Young Adults](#)
[Women's Cancers](#)

Clinical Trials

Search for Clinical Trials

[What Is a Clinical Trial?](#)
[Clinical Trial Results](#)
[Learn About Clinical Trials](#)

Research & Funding

[Funding Opportunities](#)
[Funding Patterns for R01 and R21 Grants in FY2013](#)
[Training Opportunities](#)

Cancer Topics

[What Is Cancer?](#)
[Treatment](#)
[Prevention, Genetics, Causes](#)
[Screening and Testing](#)
[Coping with Cancer](#)
[Smoking](#)
[Cancer Health Disparities](#)
[NCI Fact Sheets](#)
[Physician Data Query \(PDQ®\)](#)

Cancer Research News

[Cold Spring Harbor researchers find immunogenic mutations in tumor genomes correlate with increased patient survival](#)

[Northwestern U researchers find that vitamin D deficiency could be linked to aggressive prostate cancer](#)

[Hopkins researchers find that brain tumor cells penetrated by tiny, degradable particles can carry genetic instructions](#)

[Visit the NCI News Center](#)

Emergency Preparedness Resources

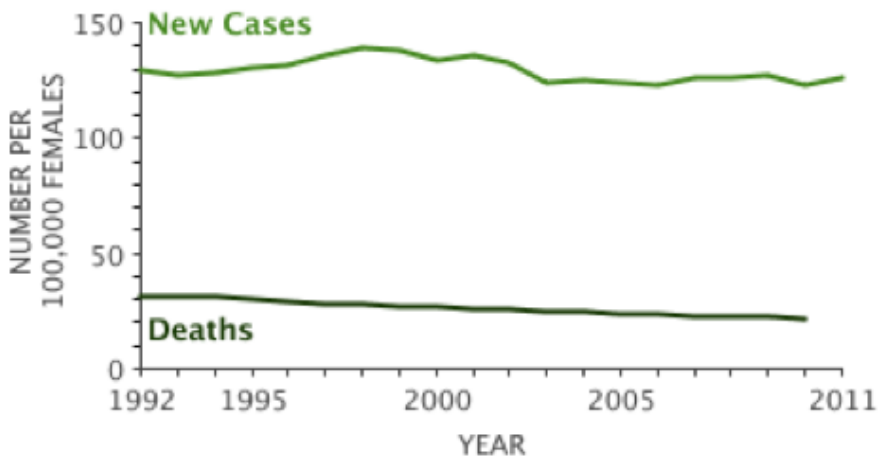
PREVIOUS

NEXT

Statistics at a Glance

At a Glance

Estimated New Cases in 2014	232,670
% of All New Cancer Cases	14.0%
Estimated Deaths in 2014	40,000
% of All Cancer Deaths	6.8%



Percent Surviving 5 Years
89.2%
2004-2010

Number of New Cases and Deaths per 100,000: The number of new cases of breast cancer was 124.5 per 100,000 women per year. The number of deaths was 22.6 per 100,000 women per year. These rates are age-adjusted and based on 2007-2011 cases and 2006-2010 deaths.

Lifetime Risk of Developing Cancer: Approximately 12.3 percent of women will be diagnosed with breast cancer at some point during their lifetime, based on 2008-2010 data.

Prevalence of this cancer: In 2011, there were an estimated 2,899,726 women living with breast cancer in the United States.

Many dictionaries

- NCI dictionary of Cancer Terms
- NCI Drug dictionary
- Glossary of statistical terms
- Glossary of Genetic terms

ClinicalTrials.gov

ClinicalTrials.gov

A service of the U.S. National Institutes of Health

ClinicalTrials.gov is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world. [Learn more about clinical studies](#) and [about this site](#), including relevant [history](#), [policies](#), and [laws](#).

[Find Studies](#) ▾ [About Clinical Studies](#) ▾ [Submit Studies](#) ▾ [Resources](#) ▾ [About This Site](#) ▾

ClinicalTrials.gov currently lists **188,780 studies** with locations in all 50 states and in **190 countries**.

Text Size ▾

Search for Studies

Example: "Heart attack" AND "Los Angeles"

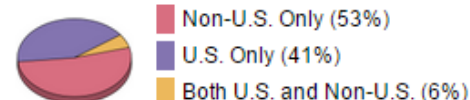
[Advanced Search](#) | [See Studies by Topic](#)

[See Studies on a Map](#)

Search Help

- [How to search](#)
- [How to find results of studies](#)
- [How to read a study record](#)

Locations of Recruiting Studies



Total N = 35,128 studies
Data as of April 21, 2015

- [See more trends, charts, and maps](#)

For Patients & Families

- [How to find studies](#)
- [See studies by topic](#)
- [Learn about clinical studies](#)
- [Learn more...](#)

For Researchers


- [How to submit studies](#)
- [Download content for analysis](#)
- [About the results database](#)
- [Learn more...](#)

For Study Record Managers

- [Why register?](#)
- [How to register study records](#)
- [FDAAA 801 Requirements](#)
- [Learn more...](#)

Learn More

- [ClinicalTrials.gov Online Training](#)
- [Glossary of common site terms](#)

 [For the Press](#)

 [Using our RSS Feeds](#)

Trial record **1 of 50** for: HER2 mutations breast cancer

[Previous Study](#) | [Return to List](#) | [Next Study](#) ▶

Neratinib in Metastatic HER2 Non-amplified But HER2 Mutant Breast Cancer

This study is currently recruiting participants. (see [Contacts and Locations](#))

Verified April 2015 by Washington University School of Medicine

Sponsor:

Washington University School of Medicine

Information provided by (Responsible Party):

Washington University School of Medicine

ClinicalTrials.gov Identifier:

NCT01670877

First received: August 17, 2012

Last updated: April 7, 2015

Last verified: April 2015

[History of Changes](#)

Full Text View

Tabular View

No Study Results Posted

[Disclaimer](#)

[? How to Read a Study Record](#)

► Purpose

This phase II study will test **cancer** to see if it has a **HER2 mutation** and, if so, see how **HER2 mutated cancer** responds to treatment with neratinib.

<u>Condition</u>	<u>Intervention</u>	<u>Phase</u>
Breast Neoplasms	Drug: Neratinib	Phase 2

Study Type: **Interventional**

Study Design: **Endpoint Classification: Efficacy Study**

Intervention Model: Single Group Assignment

Masking: Open Label

Primary Purpose: Treatment

Official Title: **A Phase II Study of Neratinib in Metastatic HER2 Non-amplified But HER2 Mutant Breast Cancer**

Adjuvant! Online

[Adjuvant! Home](#)[Messages](#)[Breast Cancer](#)[Colon Cancer](#)[Lung Cancer](#)[MetResect](#)[Downloads](#)[Online Resources](#)[Personal Info](#)[Logout](#)[Intended Use](#)[FAQs](#)[Contact Us](#)

Adjuvant! Online

Decision making tools for health care professionals

Adjuvant! for Breast Cancer (Version 8.0)

Patient Information

Age:

Comorbidity:

ER Status:

Tumor Grade:

Tumor Size:

Positive Nodes:

Calculate For:

10 Year Risk:

Adjuvant Therapy Effectiveness

Horm:

Chemo:

Hormonal Therapy:

Chemotherapy:

Combined Therapy:

No additional therapy:



87.8 alive in 10 years.

3.8 die of cancer.

8.4 die of other causes.

With hormonal therapy: Benefit = 0.8 alive.



With chemotherapy: Benefit = 0.3 alive.



With combined therapy: Benefit = 1.1 alive.

[Print Results PDF](#)[Access Help and Clinical Evidence](#)[Images for Consultations](#)

Note of caution!



Validity of Adjuvant! Online program in older patients with breast cancer: a population-based study

Nienke A de Glas, Willemien van de Water, Ellen G Engelhardt, Esther Bastiaannet, Anton J M de Craen, Judith R Kroep, Hein Putter, Anne M Stiggelbout, Nir I Weijl, Cornelis J H van de Velde, Johanneke E A Portielje, Gerrit-Jan Liefers

Summary

Lancet Oncol 2014; 15: 722-29

Published Online

May 14, 2014

[http://dx.doi.org/10.1016/S1473-0245\(14\)70200-1](http://dx.doi.org/10.1016/S1473-0245(14)70200-1)

See Comment page 672

Department of Surgery

(N A de Glas MD,

W van de Water MD,

E Bastiaannet PhD,

Prof C J H van de Velde PhD,

G J Liefers PhD), Department of

Gerontology and Geriatrics

(N A de Glas, W van de Water,

E Bastiaannet,

A J M de Craen PhD),

Department of Medical

Decision Making

(E G Engelhardt MSc,

Prof A M Stiggelbout PhD),

Department of Medical

Oncology (J R Kroep PhD),

Department of Medical

Statistics (Prof Hein Putter PhD),

Leiden University Medical

Centre, Leiden, Netherlands;

Department of Medical

Oncology, Bronovo Hospital

The Hague, The Hague,

Netherlands (N I Weijl PhD);

and Department of Medical

Background Adjuvant! Online is a prediction tool that can be used to aid clinical decision making in patients with breast cancer. It was developed in a patient population aged 69 years or younger, and subsequent validation studies included small numbers of older patients. Since older patients with breast cancer differ from younger patients in many aspects, the aim of this study was to investigate the validity of Adjuvant! Online in a large cohort of unselected older patients.

Methods We included patients from the population-based FOCUS cohort, which included all consecutive patients aged 65 years or older who were diagnosed with invasive or in-situ breast cancer between Jan 1, 1997, and Dec 31, 2004, in the southwestern part of the Netherlands. We included all patients who fulfilled the criteria as stated by Adjuvant! Online: patients with unilateral, unicentric, invasive adenocarcinoma; no evidence of metastatic or residual disease; no evidence of T4 features; and no evidence of inflammatory breast cancer. We entered data from all patients with the "average for age" comorbidity status (model 1) and with an individualised comorbidity status (model 2).

Findings We included 2012 patients. Median age of patients in the cohort was 74.0 years (IQR 69.0–79.0). 904 (45%) of 2012 patients died during follow-up, whereas 326 (16%) patients had recurrence. Median follow-up for overall survival was 9.0 years (IQR 7.4–10.7), and 6.6 years (4.4–6.6) for patients without recurrence. Using model 1, Adjuvant! Online overestimated 10-year overall survival by 9.8% [95% CI 5.9–13.7], $p < 0.0001$ and 10-year cumulative recurrence survival by 8.7% [6.7–10.7], $p < 0.0001$. By contrast, when using model 2, Adjuvant! Online underestimated the 10-year overall survival by –17.1% [95% CI –21.0 to –13.2], $p < 0.0001$. However, when using model 2, Adjuvant! Online predicted cumulative recurrence accurately in all patients (–0.7% [95% CI –2.7–1.3], $p = 0.48$).

Interpretation Adjuvant! Online does not accurately predict overall survival and recurrence in older patients with early breast cancer.

Funding Dutch Cancer Foundation.

European Journal of Cancer (2012) 48, 982–989



ELSEVIER

Available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.ejconline.com



Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients

Nirmala Bhoo-Pathy^{a,b}, Cheng-Har Yip^c, Mikael Hartman^{d,e}, Nakul Saxena^d, Nur Aishah Taib^c, Gwo-Fuang Ho^f, Lai-Meng Looi^g, Awang M. Bulgiba^h, Yolanda van der Graaf^b, Helena M. Verkooijen^{d,i,*}

^a National Clinical Research Centre, Level 3, Dermatology Block, Kuala Lumpur Hospital, 50586 Jalan Pahang, Kuala Lumpur, Malaysia

^b Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

^c Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia

^d Saw Swee Hock School of Public Health, National University of Singapore, Block MD3, 16 Medical Drive, Singapore 117597, Singapore

^e Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Block MD3, 16 Medical Drive, Singapore 117597, Singapore

^f Department of Oncology, Faculty of Medicine, University of Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia

^g Department of Pathology, Faculty of Medicine, University of Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia

^h Julius Centre University of Malaya, Faculty of Medicine, University of Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia

ⁱ Imaging Division, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

2. Research databases

Research databases

- Gene Expression Omnibus: www.ncbi.nlm.nih.gov/geo/
- Array Express: www.ebi.ac.uk/arrayexpress/
- EGA (European Genome-phenome Archive)
- METABRIC
- TCGA
- ICGC
- The Cancer Cell Line Encyclopedia

ARTICLE

doi:10.1038/nature10983

The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

Christina Curtis^{1,2*}, Sohrab P. Shah^{3,4*}, Suet-Feung Chin^{1,2*}, Gulista Turashvili^{3,4*}, Oscar M. Rueda^{1,2}, Mark J. Dunning², Doug Speed^{2,3*}, Andy G. Lynch¹, Shamith Samarajay^{1,2}, Yinyun Yuan^{1,2}, Stefan Gräf¹, Gavin Ha¹, Gholamreza Hafari¹, Ali Bashashati¹, Roslin Russell¹, Steven McKinney^{3,4}, METABRIC Group^{1,2}, Anita Langerød⁵, Andrew Green⁶, Elena Provenzano⁶, Gordon Wishart⁷, Sarah Prince², Peter Watson^{3,4,10}, Florian Markowetz^{1,2}, Leigh Murphy¹⁰, Ian Ellis¹, Arnie Purushotham^{1,11}, Anne-Lise Børresen-Dale^{6,12}, James D. Brenton^{2,13}, Simon Tavaré^{1,2,5,14}, Carlos Caldas^{1,2,15,16} and Samuel Aparicio^{1,4}

The elucidation of breast cancer subgroups and their molecular drivers requires integrated views of the genome and transcriptome from representative numbers of patients. We present an integrated analysis of copy number and gene expression in a discovery and validation set of 997 and 995 primary breast tumours, respectively, with long-term clinical follow-up. Inherited variants (copy number variants and single nucleotide polymorphisms) and acquired somatic copy number aberrations (CNAs) were associated with expression in ~40% of genes, with the landscape dominated by *cis*- and *trans*-acting CNAs. By delineating expression outlier genes driven in *cis* by CNAs, we identified putative cancer genes, including deletions in *PPP2R2A*, *MTAP* and *MAP2K4*. Unsupervised analysis of paired DNA-RNA profiles revealed novel subgroups with distinct clinical outcomes, which reproduced in the validation cohort. These include a high-risk, oestrogen-receptor-positive 1q13/14 *cis*-acting subgroup and a favourable prognosis subgroup devoid of CNAs. Transcriptome-wide analysis of copy number-specific gene networks, including a copy number-dependent gene expression deletion-mediated adaptive immune response in the 'CNA-devoid' subgroup and a basal-specific chromosome 5 deletion-associated mitotic network. Our results provide a novel molecular stratification of the breast cancer population, derived from the impact of somatic CNAs on the transcriptome.

Inherited genetic variation and acquired genomic aberrations contribute to breast cancer initiation and progression. Although somatically acquired CNAs are the dominant feature of sporadic breast cancers, the driver events that are selected for during tumorigenesis are difficult to elucidate as they co-occur alongside a much larger landscape of random non-pathogenic passenger alterations¹² and germline copy number variants (CNVs). Attempts to define subtypes of breast cancer and to discern possible somatic drivers are still in their relative infancy¹³⁻¹⁶, in part because breast cancer represents multiple diseases, implying that large numbers (many hundreds or thousands) of patients must be studied. However, we can use an integrated genomic/transcriptomic analysis of breast cancers with long-term clinical outcomes composed of a discovery set of 997 primary tumours and a validation set of 995 tumours from METABRIC (Molecular Taxonomy of Breast Cancer International Consortium).

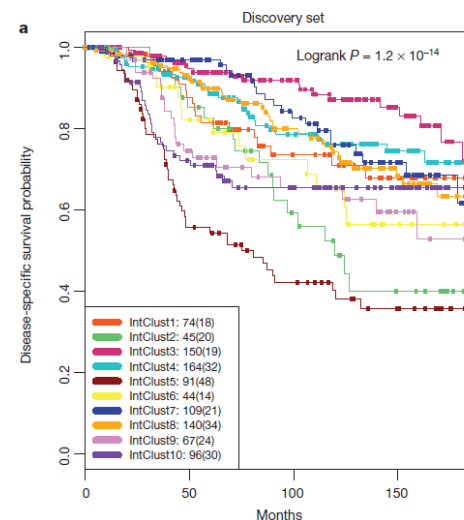
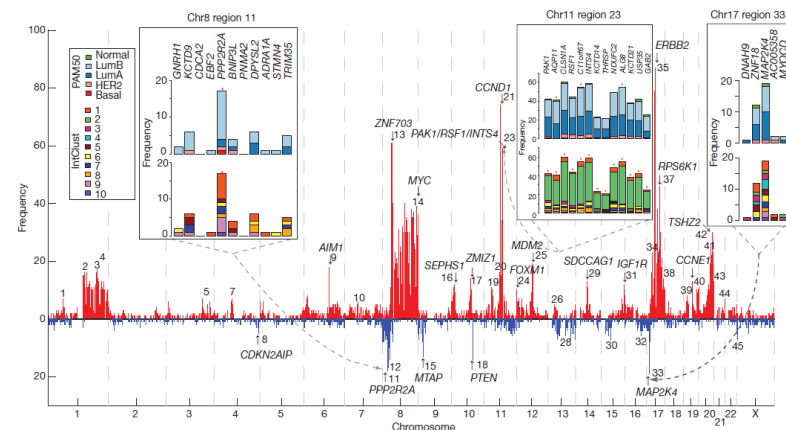
A breast cancer population genomic resource

We assembled a collection of over 2,000 clinically annotated primary fresh-frozen breast cancer specimens from tumour banks in the UK

and Canada (Supplementary Tables 1–3). Nearly all oestrogen receptor (ER)-positive and/or lymph node (LN)-negative patients did not receive chemotherapy, whereas ER-negative and LN-positive patients did. Additionally, none of the HER2⁺ patients received trastuzumab. As such, the treatments were homogeneous with respect to clinically relevant groupings. An initial set of 997 tumours was analysed as a discovery group and a further set of 995 tumours, for which complete data later became available, was used to test the reproducibility of the integrative clusters (described below). An overview of the main analytical approaches is provided in Supplementary Fig. 1. Details concerning expression and copy number profiling, including sample segmentation and copy number calling, are described in Supplementary Methods 1–2. The number analysis (Supplementary Tables 4–8) and validation (Supplementary Figs 3 and 4 and Supplementary Tables 9–11), and TP53 mutational profiling (Supplementary Fig. 5) are described in the Supplementary Information.

Genome variation affects tumour expression architecture

Genomic variants are considered to act in *cis* when a variant at a locus has an impact on its own expression, or in *trans* when it is associated



¹Department of Oncology, University of Cambridge Hills Road, Cambridge CB2 2XU, UK; ²Cancer Research UK, Cambridge Research Institute, Li Shing Cheung, Robinson Way, Cambridge CB2 0RE, UK; ³Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada; ⁴Molecular Oncology, British Columbia Cancer Centre, Vancouver, British Columbia V6S 1L3, Canada; ⁵Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge for Mathematical Sciences, Cambridge CB3 0WA, UK; ⁶Department of Genetics, Institute for Genome Sciences and Policy, Duke University Medical Center, Durham, North Carolina, NC 27710, USA; ⁷Department of Histopathology, School of Molecular Medical Sciences, University College London, Gower Street, London WC1E 6BT, UK; ⁸Department of Biomedical Sciences, University of Alberta, Edmonton, Alberta T6G 2G4, Canada; ⁹Department of Pathology, University of Toronto, Toronto, Ontario M5S 1A5, Canada; ¹⁰Department of Pathology, University of Cambridge, Cambridge CB2 2DQ, UK; ¹¹King's College London, Breakthrough Breast Cancer Research Unit, London W2C 2LS, UK; ¹²Manitoba Institute of Cell Biology, University of Manitoba, Manitoba R3E 0V9, Canada; ¹³NHR Comprehensive Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, London WC2R 2LS, UK; ¹⁴Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, 0316 Oslo, Norway; ¹⁵Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK; ¹⁶Molecular and Computational Biology Program, University of California, Santa Barbara, CA 93106-5080, USA; ¹⁷Department of Preventive Medicine, Keio School of Medicine, Yokohama City, Japan; ¹⁸School of Southern California, Los Angeles, California 90095, USA (Ch.C.); ¹⁹University College London, University Centre, WC1E 6BT, UK (D.S.).

†Lists of participants and affiliations appear at the end of the paper.

METABRIC

(www.ebi.ac.uk/ega/studies/EGAS000000000083)

EMBL-EBI

ServicesResearchTrainingAbout Us

European Genome-phenome Archive

All

Search

Examples: EGAS000000000001, cancer

EGA home

About

Studies

Datasets

Data access committees

Data providers

Submit to EGA

Contact Us

STUDY: METABRIC

Study Description

The elucidation of breast cancer subgroups and their molecular drivers requires integrated views of the genome and transcriptome from... [Show More](#)

Study ID	Alternative Stable ID	type
EGAS000000000083		Genotype/Expression

Data provider(s)

- [University of Cambridge & Cancer Research UK Cambridge Research Institute](#)
- [The Cancer Research UK Cambridge Research Institute](#)

Who archives the data?



This study includes 10 datasets:

Click on a Dataset ID in the table below to learn more, and to find out who to contact about access to these data

Dataset ID	Technology	Type	Samples	Description
EGAD00010000162	Illumina HT 12		2136	Illumina HT 12 IDATS
EGAD00010000164	Affymetrix SNP 6.0		1992	Affymetrix 6.0 CEL files
EGAD00010000210	Illumina HT 12		997	Normalized expression data; discovery set
EGAD00010000211	Illumina HT 12		995	Normalized expression data; validation set
EGAD00010000212	Illumina HT 12		144	Normalized expression data; normals
EGAD00010000213	Affymetrix SNP 6.0		997	Segmented (CBS) copy number aberrations (CNA); discovery set
EGAD00010000214	Affymetrix SNP 6.0		997	Segmented (CBS) copy number variants (CNV); discovery set
EGAD00010000215	Affymetrix SNP 6.0		995	Segmented (CBS) copy number aberrations (CNA); validation set

TCGA/ICGC

(The Cancer Genome Atlas/ International Cancer Genomics Consortium)

The Cancer Genome Atlas Data Portal

Understanding genomics to improve cancer care

TCGA Home | Contact Us | For the Media

Home Download Data Tools About the Data Publication Guidelines

TCGA Data Portal Overview

The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high level sequence analysis of the tumor genomes.

Please note some data on the TCGA Data Portal are in controlled-access. Please visit the [Access Tiers](#) page for more information.

The TCGA Data Portal does not host lower levels of sequence data. NCI's [Cancer Genomics Hub \(CGHub\)](#) is the new secure repository for storing, cataloging, and accessing BAM files and metadata for sequencing data.

[Download Data](#)

Choose from four ways to download data

Available Cancer Types	# Cases Shipped by BCR	# Cases with Data	Date Last Updated (mm/dd/yy)
Acute Myeloid Leukemia [LAML]	200	200	01/28/15
Adrenocortical carcinoma [ACC]	80	80	04/10/15
Bladder Urothelial Carcinoma [BLCA]	412	412	04/13/15
Brain Lower Grade Glioma [LGG]	516	516	04/21/15
Breast invasive carcinoma [BRCA]	1100	1098	04/21/15
Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]	308	308	04/21/15
Cholangiocarcinoma [CHOL]	36	36	04/21/15
Colon adenocarcinoma [COAD]	461	461	04/10/15
Esophageal carcinoma [ESCA]	185	185	04/22/15

Announcements

03/11/2015 - TCGA DCC Downtime

TCGA will be having emergency maintenance tomorrow, March 12th, starting at 5:30AM EDT and lasting for approximately 30 minutes. During this time the Data Portal will not be available.

01/08/2015 - Software release

The DCC has successfully completed the software release scheduled for today. Details about this release can be found on the TCGA Wiki: <https://wiki.nci.nih.gov/x/hgExDw>.

[See all announcements](#)

More TCGA Information

More information about The Cancer Genome Atlas program can be found by following the links below:

[TCGA website](#)

[TCGA Publications](#)

[Publications using TCGA Data](#)

[TCGA publication guidelines](#)

N=11,079
N BC= 1,098

ICGC Data Portal

[Cancer Projects](#) [Advanced Search](#) [Data Analysis](#) [Data Repository](#)

eg. BRAF, KRAS G12D, D035100, MU7870, apoptosis, Cancer Gene Census, GO:0016049

About Us

The ICGC Data Portal provides tools for visualizing, querying and downloading the data released quarterly by the consortium's member projects.

To access ICGC controlled tier data, please read these instructions.

New features will be regularly added by the DCC development team. [Feedback](#) is welcome.

Subscribe to our Twitter [feed](#) to get updates.

Data Release 18

February 11th, 2015

Donor Distribution by Primary Site

Information

[Access Controlled Data](#)
[Methods](#)
[Submitter Tools](#)

Tutorial

EXAMPLE QUERIES

1. BRAF missense mutations in colorectal cancer
2. Most frequently mutated genes by high impact mutations in stage III malignant lymphoma
3. Brain cancer donors with frameshift mutations and having methylation data available

Tweets

[Follow](#)

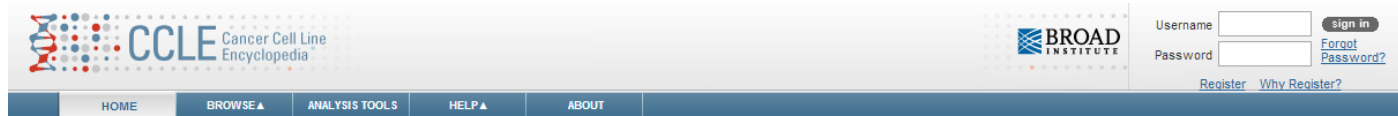
ICGC DCC @icgc_dcc 13h
The DCC Portal is currently down for maintenance/upgrades but will be back online shortly

BF Francis Ouellette @bfco 17 Apr
From @PLOSmedicine | Interview with me on "Cancer Genomics: Data, Data and more Data" blogs.plos.org/speakingofmedicine.../CC/ @icgc_dcc @TCGAupdates

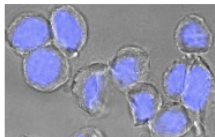
Cancer projects	55
Cancer primary sites	18
Donors	12,807
Simple somatic mutations	12,942,642
Mutated genes	57,517

N=12,807
N BC= 1,186

The Cancer Cell Line Encyclopedia (www.broadinstitute.org/ccle/home)



Broad-Novartis Cancer Cell Line Encyclopedia (CCLE)



The Cancer Cell Line Encyclopedia (CCLE) project is a collaboration between the [Broad Institute](#), and the [Novartis Institutes for Biomedical Research](#) and its [Genomics Institute of the Novartis Research Foundation](#) to conduct a detailed genetic and pharmacologic characterization of a large panel of human cancer models, to develop integrated computational analyses that link distinct pharmacologic vulnerabilities to genomic patterns and to translate cell line integrative genomics into cancer patient stratification. The CCLE provides public access to genomic data, analysis and visualization for about 1000 cell lines.

The CCLE is an ongoing project and some data are not complete yet. The CCLE website is subject to periodic changes and improvements. Please visit regularly!

This project is funded by [Novartis](#).

News / Events



Nov 12, 2013: The Gene Set Enrichment Analysis (GSEA) is not currently able to run due to server issues. Engineers are working on the problem.

Sep 30, 2013: The CCLE can now upload data directly to GenomeSpace. To learn more, see the GenomeSpace blog [more...](#)

Sep 18, 2013: The portal will be unavailable from Thursday October 17 through Sunday October 20th for server maintenance.

Aug 6, 2013: The problems with the GeneNeighbors and Differential Expression (CMS) analyses have been corrected. All systems are functioning normally once again.

Aug 5, 2013: Due to a database issue, the GeneNeighbors and Differential Expression (CMS) analyses are not currently working. We have engineers trying to resolve the problem and hope to have this fixed shortly.

May 20, 2013: The server problem has been resolved. Differential Expression, Gene Set Enrichment Analysis (GSEA) and sample browsing are working once again.

New User?



Please register for full access to the data and analyses tools CCLE provides.

[Terms of Access](#) [Register](#)

Tutorials / Manuals



Tutorials, analysis descriptions and other documentation is available at: [HELP > DOCUMENTS](#)

Frequently Altered Genes



Tag clouds summarizing genes frequently altered in the datasets of this portal are available in the [Mutation tag cloud](#)

Publications



The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Barretina Caponigro Stransky et al., *Nature* 483, 603-307, 2012

What you can do on this portal

Search *Cell line, Annotation, Gene*

Search CCLE

Search for information

Enter a keyword to search for genes, news items and publications. Search results for a gene include links to annotations and analyses.

Data Sets



Browse, analyze and download studies and data sets.

[BROWSE > DATA](#)

Analysis Tools



The portal provides the following analysis tools:

Integrative Genomics Viewer (IGV)

Visualize a data set in the Integrative Genomics Viewer (IGV), a high-performance visualization tool for interactive exploration of large integrated data sets.

Differential Expression Analysis

Find genes that are significantly differentially expressed between two user-defined classes of samples from an expression data set available on this portal.

Gene co-expression

View the top 20 genes in a data set that are co-expressed with a gene of interest.

Gene Set Enrichment Analysis (GSEA)

Find [pathway gene sets](#) correlated with a gene of interest. Domain experts curate the pathway gene sets based on data from several online pathway databases.

[ANALYSIS TOOLS](#)

Sample Sets

3. Research Tools

Research tools

- R: a door to many research avenues...!
(www.bioconductor.org)
- GSEA (www.broadinstitute.org/gsea/)
- MSigDB (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>)
- DAVID (<http://david.abcc.ncifcrf.gov/>)
- Oncomine (www.oncomine.org)

GSEA & MSigDB

- **Gene Set Enrichment Analysis (GSEA)** is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).
- The **Molecular Signatures Database (MSigDB)** is a collection of annotated gene sets for use with GSEA software.

Annotation of mutations

- COSMIC (<http://cancer.sanger.ac.uk/cancergenome/projects/cosmic>)
- Oncotator (www.broadinstitute.org/oncotator)
- Annovar (<http://www.openbioinformatics.org/annovar>)
- DGIdb (dgidb.genome.wustl.edu)
- ...!

COSMIC



COSMIC v72

eg: *Braf*, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell

SEARCH

R Resources

Key COSMIC resources

[Cell Lines Project](#)
[COSMIC Whole Genomes](#)
[Cancer Gene Census](#)
[Drug Sensitivity](#)
[Mutational Signatures](#) **New**
[GRCh37 Cancer Archive](#) **New**

T Tools

Additional tools to explore COSMIC

[Cancer Browser](#)
[Genome Browser](#)
[CONAN](#)
[COSMIC Mart](#)

C Expert Curation

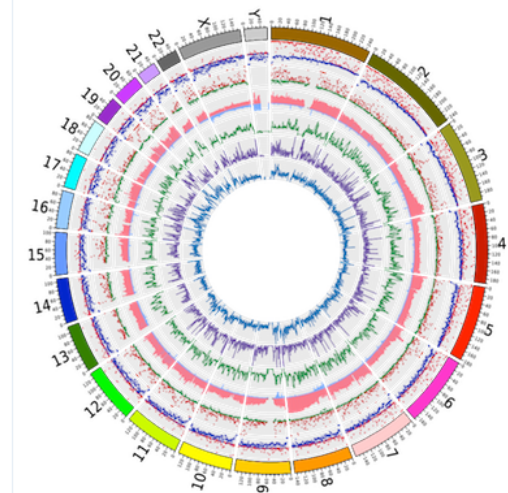
High quality curation by expert postdoctoral scientists

[Cancer Gene Census](#)
[Curated Genes](#)
[Gene Fusions](#)
[Genome-Wide Screens](#)

D Data

Further details on using COSMIC's content

[Downloads](#)
[License](#)
[Submission](#)
[Genome Annotations](#)
[Datasheet V72](#)
[Help](#)
[FAQ](#)



Genomic Landscape of Cancer

Statistics

Domain	Counts
Samples	1103964
Coding Mutations	3158657
Papers	21086
Fusions	10890
Genomic Rearrangements	61232
Whole Genomes	19672
Copy Number	842651
Gene Expression	8228797

Cancer Gene Census



[Census](#) |
 [Breakdown](#) |
 [Abbreviations](#)

The cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in [Nature Reviews Cancer](#) and [supplemental analysis information](#) related to the paper is also available.

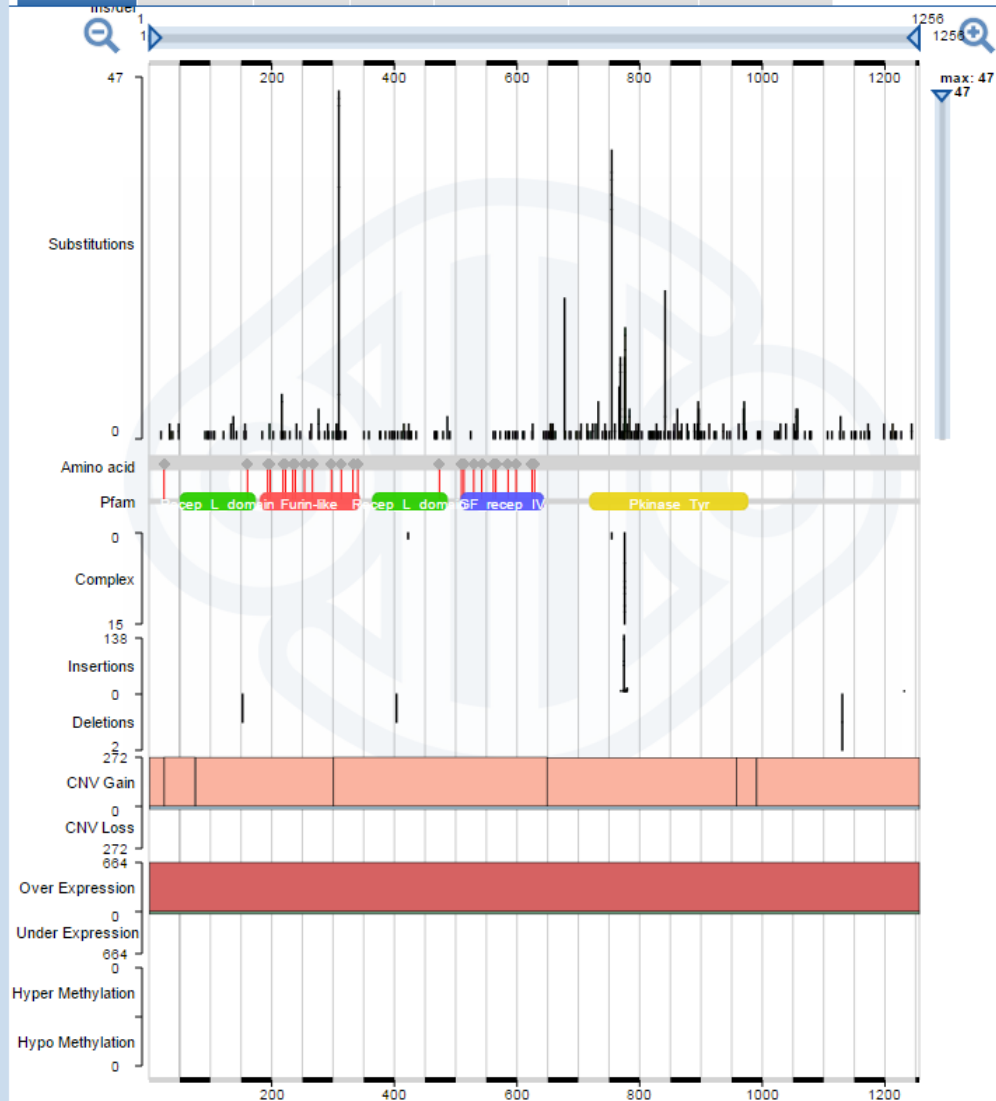
The census is not static but rather is updated regularly/as needed. In particular we are grateful to Felix Mitelman and his colleagues in providing information on more genes involved in uncommon translocations in leukaemias and lymphomas. Currently, more than 1% of all human genes are implicated via mutation in cancer. Of these, approximately 90% have somatic mutations in cancer, 20% bear germline mutations that predispose to cancer and 10% show both somatic and germline mutations.

Show 10 entries

Search: Export: [CSV](#) [TSV](#)



Gene Symbol ▾	Name ▲	Entrez GeneId ▾	Genome Location ▾	Chr Band ▾	Somatic ▾	Germline ▾	Tumour Types(Somatic) ▾	Tumour Types(Germline) ▾	Cancer Syndrome ▾	Tissue Type ▾	M
EP300	300 kd E1A-Binding protein gene	2033	22:41093005-41178956	22q13	yes		colorectal; breast; pancreatic; AML; ALL; DLBCL			L; E	R
NT5C2	5'-nucleotidase; cytosolic II	22978	10:103089672-103174958	10q24.32	yes		relapse ALL			L	C
ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	471	2:215312144-215349656	2q35	yes		ALCL			L	C
AKAP9	A kinase (PRKA) anchor protein (yotiao) 9	10142	7:91941100-92110159	7q21-q22	yes		papillary thyroid			E	C
AF15Q14	AF15q14 protein	57082	15:40602933-40654977	15q14	yes		AML			L	C
AF5q31	ALL1 fused gene from 5q31	27125	5:132881059-132937189	5q31	yes		ALL			L	C

Histogram Mutations Fusions Tissue Distribution CNV & Expr Methylation

Filters

Gene

Position

Start End

Sequence Type:

cDNA ☐Amino Acid ☒Cancer type [Select](#)

- ☒ Systematic Screen
- ☒ Somatic Status
- ☒ Tumour Source
- ☒ Mutation Type
- ☒ Copy Number Variation
- ☒ Gene Expression
- ☒ Methylation

[Apply](#)[Reset](#)

cBioPortal



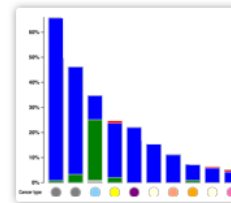
Visualize, analyze, discover.



The cBioPortal for Cancer Genomics provides **visualization, analysis** and **download** of large-scale **cancer genomics** data sets.

Please adhere to [the TCGA publication guidelines](#) when using TCGA data in your publications.

Please cite Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012 when publishing results based on cBioPortal.



Query

Download Data

Select Cancer Study: All Cancer Studies

Select Data Type Priority: ☒ Mutation and CNA ☐ Only Mutation ☐ Only CNA

Enter Gene Set: Advanced: Onco Query Language (OQL)

User-defined List

Enter HUGO Gene Symbols or Gene Aliases

Submit

What's New

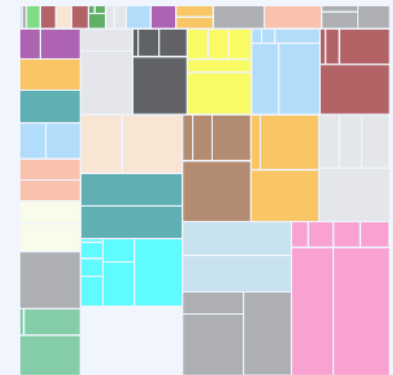
- New data and features released
- New tools released

Sign up for low-volume email news alerts:

Or follow us @cbioportal on Twitter

Data Sets

The Portal contains data for **20958 tumor samples from 89 cancer studies**. [\[Details\]](#)



Example Queries

RAS/RAF alterations in colorectal cancer

BRCA1 and BRCA2 mutations in ovarian cancer

Modify Query

Breast Invasive Carcinoma (TCGA, Provisional)

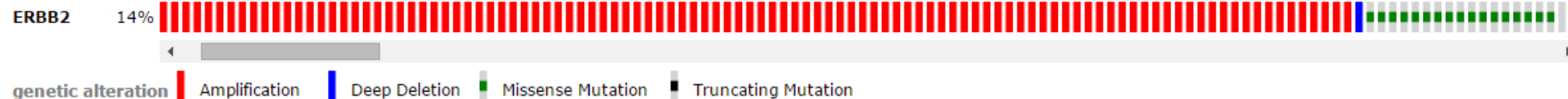
Tumors with sequencing and CNA data (962 samples) / 1 Gene

Gene Set / Pathway is altered in 138 (14.3%) of queried samples

OncoPrint Plots Mutations Co-Expression Protein Changes Survival Network IGV Download Bookmark

Case Set: Tumors with sequencing and CNA data: All tumor samples that have CNA and sequencing data (962 samples)

Altered in 138 (14%) of cases



Modify Query

Breast Invasive Carcinoma (TCGA, Provisional)

Tumors with sequencing and CNA data (962 samples) / 1 Gene

Gene Set / Pathway is altered in 138 (14.3%) of queried samples

OncoPrint Plots Mutations Co-Expression Protein Changes Survival Network IGV Download Bookmark

One Gene Custom

Plot Parameters

Gene

ERBB2

Plot Type

mRNA vs. Copy Number

Data Type

- mRNA -

mRNA expression (RNA Seq V2 RSEM)

☒ log scale

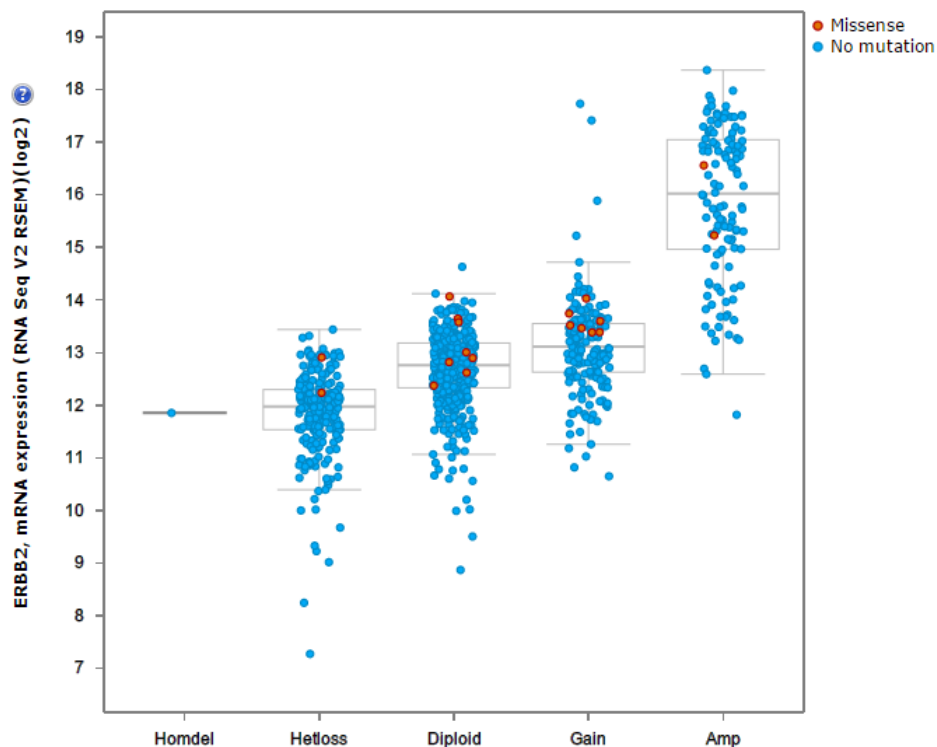
- Copy Number -

Putative copy-number alterations from

Search case(s)

Case ID...

ERBB2: mRNA Expression v. CNA SVG PDF



Modify Query

Breast Invasive Carcinoma (TCGA, Provisional)
Tumors with sequencing and CNA data (962 samples) / 1 Gene

Gene Set / Pathway is altered in 138 (14.3%) of queried samples

OncoPrint Plots Mutations Co-Expression Protein Changes Survival Network IGV Download Bookmark

ERBB2

ERBB2: [Somatic Mutation Rate: 2.0%]

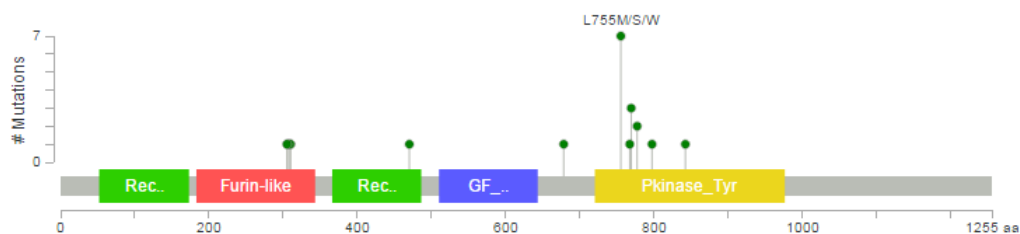
ERBB2_HUMAN

PDF

SVG

Customize

Color Codes



Show / hide columns

Showing 21 mutation(s)

Search:

Sample ID	AA change	Type	Copy #	COSMIC	Mutation Assessor	#Mut in Sample
TCGA-AO-A128-01	V797A	3D Missense	Diploid	1	Medium	703
TCGA-A8-A082-01	V842I	3D Missense	Diploid	12	Neutral	24
TCGA-AC-A3YI-01	L755S	3D Missense	Diploid	35	Medium	11
TCGA-BH-A18P-01	L755S	3D Missense	Gain	35	Medium	221
TCGA-A8-A0AB-01	L755S	3D Missense	Gain	35	Medium	21
TCGA-A8-A0A6-01	L755S	3D Missense	Gain	35	Medium	2052
TCGA-D8-A1XM-01	L755S	3D Missense	Gain	35	Medium	42
TCGA-A2-A0T6-01	L755W	3D Missense	Diploid	35	High	91
TCGA-C8-A135-01	D769H	3D Missense	AMP	9	Neutral	37
TCGA-C8-A274-01	I767M	3D Missense	Diploid	5	Neutral	136
TCGA-OL-A5D6-01	V777L	3D Missense	Diploid	12	Neutral	30

Data Set mRNA expression (RNA Seq V2 RSEM)

This table lists the genes with the highest expression correlation with the query genes. Click on a row to see the corresponding correlation plot. [?](#)

ERBB2

Search Gene

Show All ▼

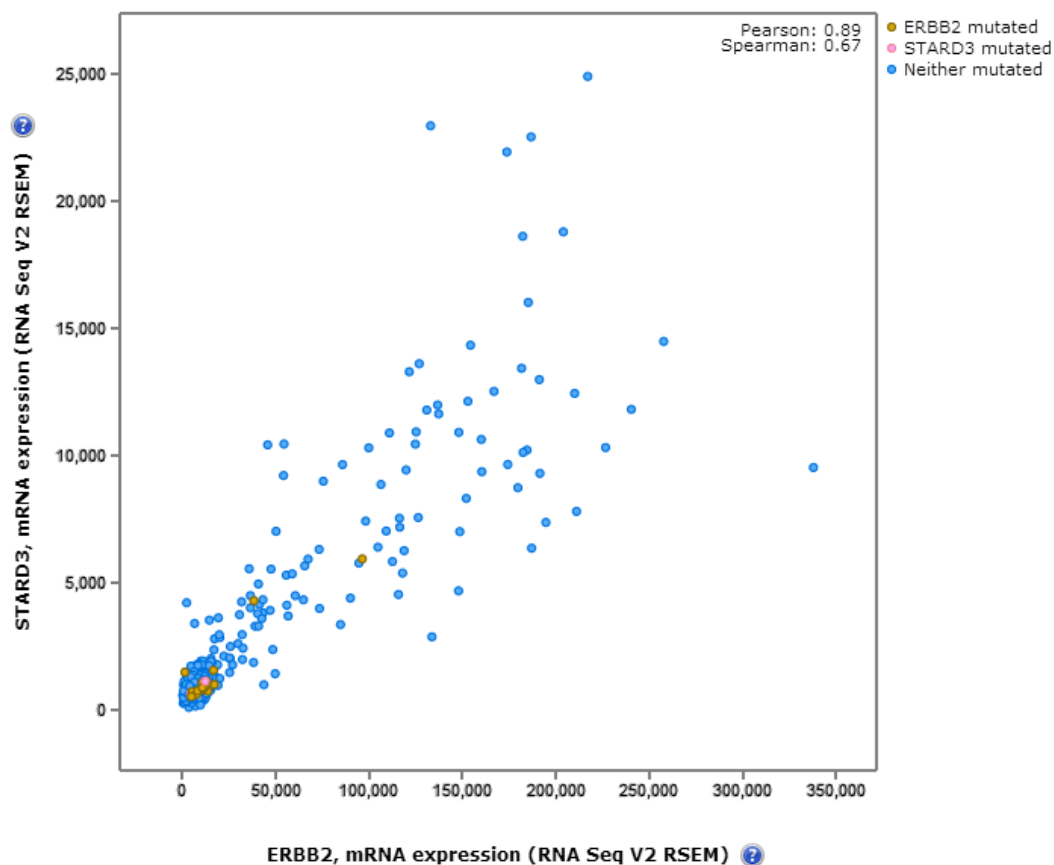
Correlated Gene ↕	Pearson's Correlation ▼	Spearman's Correlation ↕
STARD3	0.89	0.67
GRB7	0.87	0.68
PGAP3	0.87	0.74
MIEN1	0.79	0.53
ORMDL3	0.78	0.67
PSMD3	0.71	0.46
CDK12	0.63	0.39
MED1	0.61	0.39
MED24	0.61	0.52
FBXL20	0.51	0.39
CASC3	0.50	0.43
MSL1	0.50	0.42
RPL19	0.45	0.32
WIPF2	0.45	0.43
PIP4K2B	0.43	0.46
CWC25	0.42	0.36
LASP1	0.40	0.48
TMEM86A	0.40	0.30
THRA	0.40	0.36
SRCIN1	0.40	0.48
MRPL45	0.39	0.38
PCGF2	0.38	0.46
MLLT6	0.36	0.35
CISD3	0.36	0.48
TNFAIP1	0.35	0.35
SEC16A	0.32	0.33
EPN3	0.32	0.41
RAB3D	0.31	0.32
ARHGAP8	0.30	0.37

1 to 29 of 29

1 to 29 of 29


mRNA co-expression: ERBB2 vs. STARD3

[PDF](#)
[SVG](#)
☒ Show Mutations
 ☐ Log Scale X
 ☐ Log Scale Y



DGIdb

(<http://dgidb.genome.wustl.edu/>)

**DGIdb**
THE DRUG GENE INTERACTION DATABASE

Search InteractionsSearch CategoriesBrowse CategoriesHelp

Search Interactions

search for drug-gene interactions by gene name

[Show Tour](#)

Genes

Search Interactions Tour
Start by entering one or more gene names here.
Next

Replace Genes with Demo ListClear All Genes

Source Databases 9 of 9

Source Trust Level 2 of 2

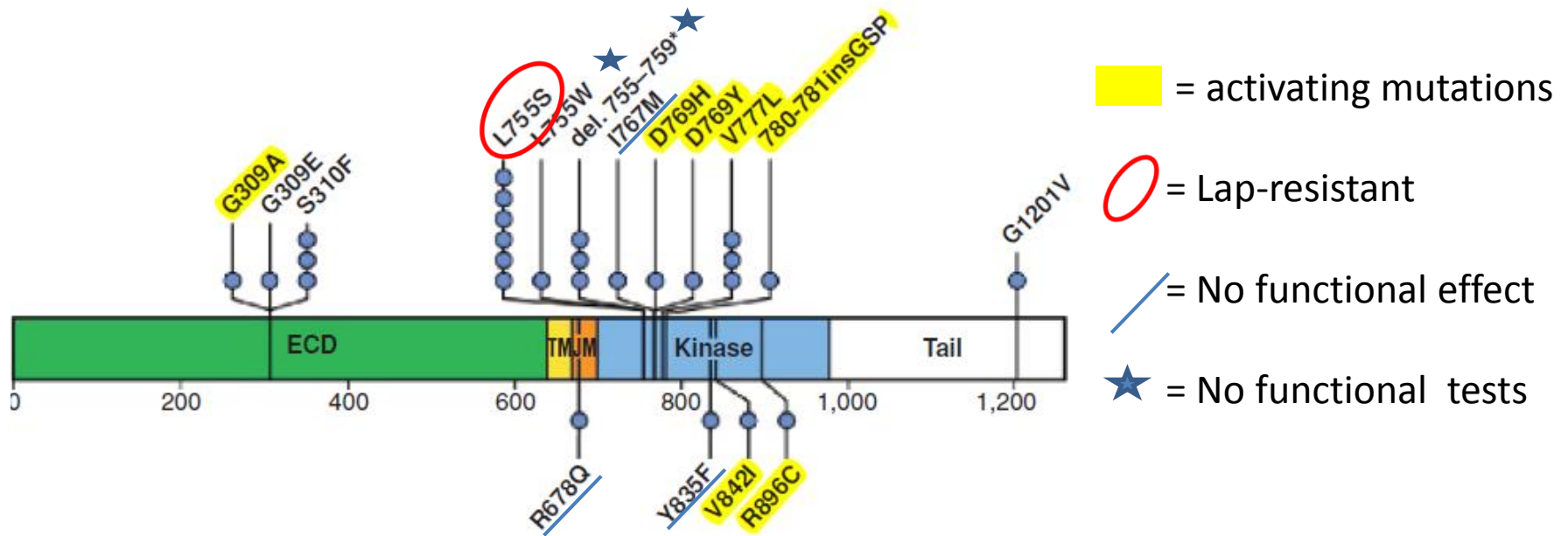
Gene Categories 41 of 41

Interaction Types 33 of 33

Anti-Neoplastic Drugs Only

Find Drug Interactions

A note of caution however!



ERBB2 mutations	Activity	Position	Mutation Assessor	Provean	SIFT	PolyPhen II
L755S	Lapatinib-resistance	chr17:g.37880220T>C	medium	Deleterious	Damaging	Probably damaging
D769H	activating	chr17:g.37880261G>C	neutral	Deleterious	Damaging	Probably damaging
V777L	activating	chr17:g.37881000G>T	neutral	Neutral	Tolerated	Benign
V842I	activating	chr17:g.37881332G>A	neutral	Neutral	Damaging	Probably damaging
G309A	activating	chr17:g.37868205G>C	low	Neutral	Damaging	Probably damaging
G309E	activating	chr17:g.37868205G>A	medium	Deleterious	Damaging	Probably damaging
R678Q	no functional effect	chr17:g.37879658G>A	low	Neutral	Tolerated	Possibly damaging
D769Y	activating	chr17:g.37880261G>T	low	Deleterious	Damaging	Probably damaging
R896C	activating	chr17:g.37881616C>T	low	Neutral	Tolerated	Benign
S310F	activating	chr17:g.37868208C>T	medium	Deleterious	Damaging	Probably damaging
I767M	no functional effect	chr17:g.37880257C>G	neutral	Neutral	Damaging	Probably damaging
Y835F	no functional effect	chr17:g.37881312A>T	low	Deleterious	Damaging	Probably damaging
P780-Y781insertionGSP	activating	chr17:g.37881011A>AGGGCTCCCC	not assessed*	Deleterious	not assessed*	not assessed*

Thank you for your attention!



121

INSTITUT
JULES BORDET
INSTITUUT

ULB

